

PERGUDANGAN DATA (Data Warehousing)

JILID 1

Dr. Budi Raharjo, S.Kom.,M.Kom.,MM.



At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio. Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum hic tenetur a sapiente delectus, ut aut nesciunt voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat.

Pergudangan Data Jilid 1 (Data Warehousing)

Penulis :

Dr. Budi Raharjo, S.Kom., M.Kom., MM.

ISBN : 9 786238 642045

Editor :

Dr. Mars Caroline Wibowo. S.T., M.Mm.Tech

Penyunting :

Dr. Joseph Teguh Santoso, M.Kom.

Desain Sampul dan Tata Letak :

Irdha Yuniyanto, S.Ds., M.Kom.

Penebit :

Yayasan Prima Agus Teknik Bekerja sama dengan
Universitas Sains & Teknologi Komputer (Universitas STEKOM)

Anggota IKAPI No: 279 / ALB / JTE / 2023

Redaksi :

Jl. Majapahit no 605 Semarang

Telp. (024) 6723456

Fax. 024-6710144

Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :

Universitas STEKOM

Jl. Majapahit no 605 Semarang

Telp. (024) 6723456

Fax. 024-6710144

Email : info@stekom.ac.id

Hak cipta dilindungi undang-undang

Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara apapun tanpa ijin dari penulis

KATA PENGANTAR

Segala puji dan Syukur bagi Tuhan Yang Maha Esa atas Rahmat dan karunia-Nya, Sehingga penulis dapat menyelesaikan buku yang berjudul "***Pergudangan Data (Data Warehousing) Jilid 1***" dengan baik. Data warehouse merupakan sistem yang didesain khusus untuk mengintegrasikan dan menyimpan data dari berbagai sumber yang berbeda dalam sebuah organisasi. Fungsinya adalah menyediakan satu sumber data yang terpadu, memungkinkan untuk analisis dan pengambilan keputusan yang lebih efektif. Data warehouse menyimpan data dalam format yang dioptimalkan untuk analisis, termasuk data historis untuk memungkinkan analisis tren dan pola. Sebelum data dimasukkan ke dalamnya, seringkali dilakukan proses pembersihan dan transformasi untuk memastikan kualitas dan konsistensi data. Pengguna dapat mengakses dan menganalisis data melalui antarmuka yang mudah digunakan, melakukan berbagai jenis analisis seperti multidimensional, laporan, dan pemodelan prediktif. Struktur penyimpanan data dalam data warehouse biasanya menggunakan skema bintang atau skema salju, dengan manajemen metadata yang terpusat. Dengan menyediakan akses terpusat dan terstruktur ke data yang relevan, data warehouse membantu organisasi dalam pengambilan keputusan yang lebih baik, respons yang lebih cepat terhadap perubahan pasar, dan meningkatkan efisiensi operasional secara keseluruhan.

Bab 1 ini menyoroti urgensi informasi strategis dalam bisnis yang kompetitif, serta mengidentifikasi krisis informasi yang mungkin terjadi di perusahaan. Perbedaan antara sistem operasional dan sistem informasi disajikan, sambil menyoroti kegagalan upaya sebelumnya dalam menyediakan informasi strategis karena kurangnya integrasi data. Solusi dalam bentuk data warehousing untuk integrasi data ditekankan, bersama dengan pentingnya intelijen bisnis dalam mengubah data menjadi wawasan tindakan. Dalam Bab 2 ini membahas secara komprehensif mengenai gudang data, dimulai dari tinjauan definisi formalnya hingga diskusi tentang ciri-ciri yang menentukannya. Selain itu, pembahasan juga mencakup perbedaan antara gudang data dan data mart, serta tinjauan terhadap tipe arsitektur yang berkembang dalam konteks gudang data. Pembaca juga diajak untuk memahami setiap komponen atau blok bangunan yang membentuk gudang data, serta diperkenalkannya metadata beserta signifikansinya dalam konteks ini. Melalui pembahasan ini, pembaca diberi pemahaman yang lengkap tentang konsep, struktur, dan pentingnya gudang data dalam lingkungan bisnis dan teknologi informasi modern.

Selanjutnya dalam Bab 3 ini akan membahas pertumbuhan berkelanjutan dalam data warehousing, menggambarkan bagaimana data warehousing telah menjadi arus utama dalam dunia teknologi informasi. Diskusi tentang beberapa tren utama dalam data warehousing membantu pembaca memahami perkembangan terkini di bidang ini. Selain itu, bab ini menyoroti perlunya standar dalam data warehousing dan meninjau kemajuan yang telah dicapai dalam hal ini. Terakhir, pembahasan tentang gudang data yang mendukung Web memberikan wawasan tentang bagaimana data warehousing terus beradaptasi dengan perkembangan teknologi untuk mendukung kebutuhan bisnis modern. Dengan demikian,

pembaca diberi pemahaman yang komprehensif tentang evolusi dan pentingnya data warehousing dalam lingkungan bisnis dan teknologi saat ini. Bab 4 ini membahas esensi perencanaan gudang data, membedakan antara proyek gudang data dan proyek sistem OLTP, serta mengajarkan cara mengadaptasi pendekatan siklus hidup untuk proyek gudang data. Selain itu, metodologi pembangunan yang tangkas untuk proyek-proyek PRT diperkenalkan, disertai dengan diskusi tentang organisasi tim proyek, peran, dan tanggung jawab. Bab ini juga menyoroti pentingnya pertimbangan terhadap tanda-tanda peringatan dan faktor keberhasilan dalam implementasi proyek gudang data. Dengan demikian, pembaca diberi pemahaman yang komprehensif tentang proses perencanaan dan implementasi proyek gudang data, serta faktor-faktor kunci yang perlu dipertimbangkan untuk mencapai keberhasilan dalam proyek tersebut.

Dalam Bab 5 ini akan menguraikan pentingnya pendefinisian persyaratan yang berbeda untuk gudang data, menyajikan peran penting dimensi bisnis dalam proses ini. Pembahasan tentang paket informasi dan penggunaannya membantu dalam menentukan persyaratan dengan lebih tepat. Selain itu, metode pengumpulan persyaratan dievaluasi secara mendalam, sementara pentingnya dokumen definisi persyaratan formal disorot sebagai landasan yang krusial. Dengan demikian, pembaca diberi pemahaman yang kokoh tentang proses yang diperlukan untuk mengidentifikasi, merancang, dan mendokumentasikan persyaratan dalam konteks pengembangan gudang data. Selanjutnya Bab 6 ini menyoroti peran penting persyaratan bisnis sebagai kekuatan pendorong dalam pengembangan gudang data, dengan mendiskusikan bagaimana persyaratan mempengaruhi setiap fase pengembangan. Fokus khusus diberikan pada bagaimana persyaratan memengaruhi desain data dan dampaknya terhadap arsitektur gudang data. Selain itu, pembahasan juga mencakup pertimbangan khusus untuk ETL dan metadata serta bagaimana persyaratan membentuk penyampaian informasi dalam konteks gudang data. Dengan demikian, bab ini memberikan pemahaman yang mendalam tentang bagaimana persyaratan bisnis memainkan peran kunci dalam seluruh proses pengembangan gudang data, dari konseptualisasi hingga implementasi.

Dalam bab 7 bertujuan untuk memberikan pemahaman yang mendalam tentang arsitektur data warehouse dengan menguraikan komponen-komponennya serta karakteristik yang membedakan arsitektur tersebut. Melalui pembahasan ini, pembaca diajak untuk memahami bagaimana kerangka arsitektur mendukung aliran data dan arti dari arsitektur teknis. Selain itu, fungsi dan jasa dari komponen arsitektur turut dipelajari untuk memperkaya pemahaman tentang sistem ini. Dengan meninjau kembali lima tipe arsitektur utama, bab ini memberikan gambaran yang komprehensif tentang beragam pendekatan dalam merancang dan mengimplementasikan data warehouse. Sebagai hasilnya, pembaca diberi pemahaman yang kokoh tentang struktur dan fungsi arsitektur data warehouse yang dapat membantu dalam pengembangan sistem yang efektif dan efisien. Bab 8 ini bertujuan untuk memberikan pemahaman yang komprehensif tentang perbedaan antara arsitektur dan infrastruktur dalam konteks data warehouse. Melalui pembahasan, pembaca diajak untuk menemukan bagaimana infrastruktur data warehouse mendukung arsitekturnya, dengan mendapatkan wawasan tentang komponen infrastruktur fisik, termasuk perangkat keras, sistem operasi, dan opsi

pemrosesan paralel yang relevan. Diskusi tentang opsi server, pemilihan DBMS, serta jenis alat yang diperlukan untuk gudang data membantu pembaca memahami secara detail tentang infrastruktur yang diperlukan untuk mendukung operasi gudang data. Dengan mempelajari konsep dan penggunaan peralatan data warehouse, pembaca diberi pemahaman yang kokoh tentang bagaimana infrastruktur memainkan peran kunci dalam keseluruhan arsitektur dan kinerja sistem data warehouse.

Bab 9 ini bertujuan untuk mendalami tentang pentingnya metadata dalam konteks data warehouse. Melalui pembahasan, pembaca diperkenalkan pada kebutuhan akan metadata dan jenis-jenis yang diperlukan oleh berbagai pemangku kepentingan. Tinjauan dilakukan terhadap jenis metadata berdasarkan tiga area fungsional, dengan fokus khusus pada metadata bisnis dan teknis. Bab ini juga mengidentifikasi persyaratan yang harus dipenuhi oleh metadata, sementara juga mempertimbangkan tantangan dalam pengelolaannya. Terakhir, pilihan untuk menyediakan metadata dieksplorasi untuk memberikan wawasan tentang cara mengelola metadata secara efektif dalam konteks data warehouse. Dengan demikian, pembaca diberi pemahaman yang mendalam tentang peran dan pentingnya metadata dalam mendukung operasi dan analisis data warehouse. Bab 10 ini mengarahkan pembaca untuk memahami bagaimana definisi persyaratan mempengaruhi desain data, terutama dalam konteks data warehouse. Melalui pembahasan tentang pemodelan dimensi dan perbedaannya dengan pemodelan hubungan entitas, pembaca diperkenalkan pada dasar-dasar skema STAR. Tinjauan yang mendalam tentang struktur tabel fakta dan dimensi memberikan pemahaman yang jelas tentang komponen-komponen inti dalam skema STAR. Keuntungan dari penggunaan skema STAR untuk data warehouse juga diteliti, sementara contoh-contoh skema STAR memberikan gambaran yang konkret tentang implementasi dalam situasi nyata. Dengan demikian, pembaca diberi pemahaman yang komprehensif tentang pentingnya desain data yang tepat dalam mencapai tujuan efisiensi dan efektivitas dalam pengelolaan data di lingkungan data warehouse.

Selanjutnya Bab 11 ini mengarahkan pembaca untuk memahami dan mendiskusikan dimensi yang berubah secara perlahan serta cara menghadapinya dengan tepat dalam konteks data warehouse. Melalui pembahasan tentang dimensi besar, pembaca diperkenalkan pada strategi untuk mengelola dimensi yang memiliki volume data yang besar. Skema kepingan salju dieksplorasi secara detail untuk memberikan pemahaman yang mendalam tentang struktur ini dalam lingkungan data warehouse. Pembahasan tentang tabel agregat membantu pembaca untuk menentukan kapan penggunaannya tepat dalam analisis data. Terakhir, survei lengkap dilakukan terhadap keluarga STARS dan aplikasinya untuk memberikan gambaran yang komprehensif tentang jenis-jenis skema yang dapat digunakan dalam data warehouse. Dengan demikian, pembaca diberi pemahaman yang kuat tentang berbagai aspek penting dalam desain dan pengelolaan dimensi serta skema data dalam lingkungan data warehouse. Bab 12 ini memberikan gambaran luas tentang fungsi ekstraksi, transformasi, dan pemuatan data (ETL) dalam konteks data warehouse. Melalui survei yang komprehensif, pembaca diajak untuk memahami tantangan, teknik, dan evaluasi dalam ekstraksi data, serta berbagai tugas dan jenis fungsi transformasi data. Pembahasan tentang integrasi dan konsolidasi data memberikan

wawasan tentang pentingnya menyatukan data dari berbagai sumber. Selain itu, pentingnya fungsi pemuatan data ditekankan, dengan penyelidikan terhadap metode utama untuk menerapkan data ke dalam gudang. Bab ini mengakhiri dengan pemahaman yang mendalam tentang mengapa proses ETL penting, serta tantangan dan kompleksitas yang terlibat di dalamnya. Dengan demikian, pembaca diberi pemahaman yang kokoh tentang peran dan signifikansi ETL dalam mengelola dan memanfaatkan data dalam lingkungan data warehouse.

Demikian buku ajar ini kami buat, dengan harapan agar pembaca dapat memahami informasi dan juga mendapatkan wawasan mengenai data warehouse. Dengan kemampuannya dalam mengintegrasikan, menyimpan, dan mengelola data secara holistik, data warehouse memberikan landasan yang kuat bagi perusahaan dalam mengambil keputusan yang lebih tepat dan responsif terhadap perubahan pasar dan kebutuhan bisnis. Terima kasih.

Semarang, Mei 2024

Penulis

Dr. Budi Raharjo, S.Kom., M.Kom., MM.

DAFTAR ISI

Halaman Judul	i
Kata Pengantar	ii
Daftar Isi	vi
BAB 1 KEBTUHAN MENARIK UNTUK DATA WAREHOUSE	1
1.1. Meningkatnya Kebutuhan Informasi Strategis	2
1.2. Kegagalan Sistem Pendukung Keputusan Masa Lalu	8
1.3. Sistem Operasional Versus Pendukung Keputusan	10
1.4. Data Warehousing, Satu-Satunya Solusi	13
1.5. Data Gudang Ditetapkan	14
1.6. Gerakan Pergudangan Data	16
1.7. Evolusi Kecerdasan Bisnis	18
BAB 2 GUDANG DATA	21
2.1. Definisi Fitur	21
2.2. Gudang Data Dan Data Marts	27
2.3. Jenis Arsitektur	30
2.4. Gambaran Umum Komponen	32
2.5. Metadata Di Gudang Data	42
BAB 3 TREN DALAM PERGUDANGAN DATA	44
3.1. Pertumbuhan Terus-Menerus Di Data Warehousing	45
3.2. Tren Yang Signifikan	49
3.3. Munculnya Standar	66
BAB 4 PERENCANAAN DAN MANAJEMEN PROYEK	73
4.1. Merencanakan Gudang Data Anda	74
4.2. Proyek Gudang Data	80
4.3. Tahap Perkembangan	84
4.4. Tim Proyek	85
4.5. Pertimbangan Manajemen Proyek	93
BAB 5 MENDEFIKAN PERSYARATAN BISNIS	100
5.1. Analisis Dimensi	101
5.2. Paket Informasi Konsep Yang Bermanfaat	105
5.3. Metode Pengumpulan Persyaratan	109
5.4. Definisi Persyaratan: Ruang Lingkup Dan Isi	117
BAB 6 PERSYARATAN SEBAGAI KEKUATAN PENGGERAK DATA WAREHOUSING	121
6.1. Desain Data	122
6.2. Rencana Arsitektur	125
6.3. Spesifikasi Penyimpanan Data	131
6.4. Strategi Penyampaian Informasi	134

BAB 7 KOMPONEN ARSITEKTUR	139
7.1. Memahami Arsitektur Data Warehouse	139
7.2. Karakteristik Yang Membedakan	141
7.3. Kerangka Arsitektur	144
7.4. Arsitektur Teknis	146
7.5. Jenis Arsitektur	155
BAB 8 INFRASTRUKTUR SEBAGAI YAYASAN UNTUK DATA WAREHOUSING	161
8.1. Arsitektur Pendukung Infrastruktur	161
8.2. Perangkat Keras Dan Sistem Operasi	164
8.3. Perangkat Lunak Database	180
8.4. Koleksi Alat	183
8.5. Peralatan Gudang Data	187
BAB 9 PERAN METADATA YANG PENTING	191
9.1. Mengapa Metadata Penting	191
9.2. Jenis Metadata Berdasarkan Wilayah Fungsional	202
9.3. Metadata Bisnis	207
9.4. Metadata Teknis	209
9.5. Bagaimana Memberikan Metadata	212
BAB 10 PRINSIP PEMODELAN DIMENSI	222
10.1. Dari Persyaratan Hingga Desain Data	222
10.2. Skema Bintang	228
10.3. Kunci Skema Bintang	237
10.4. Keunggulan Skema Bintang	239
BAB 11 PEMODELAN DIMENSI	246
11.1. Pembaruan Table Dimensi	246
11.2. Dimensi Lain-Lain	252
11.3. Tabel Fakta Agregat	261
11.4. Skema Bintang	271
BAB 12 EKSTRASI, TRASFORMASI DAN PEMUATAN DATA	279
12.1. Gambaran Umum ETL	280
12.2. Persyaratan Dan Langkah-Langkah ETL	282
12.3. Ekstrasi Data	284
12.4. Transformasi Data	295
12.5. Pemuatan Data	303
12.6. Ringkasan ETL	309
12.7. Pendekatan Integritas Lainnya	312
Daftar Pustaka	316

BAB 1

KEBUTUHAN MENARIK UNTUK DATA WAREHOUSING

Tujuan Bab

- ◆ Memahami betapa mendesaknya kebutuhan akan informasi strategis
- ◆ Mengenali krisis informasi di setiap perusahaan
- ◆ Membedakan antara sistem operasional dan sistem informasi
- ◆ Pelajari mengapa semua upaya di masa lalu untuk memberikan informasi strategis gagal
- ◆ Melihat dengan jelas mengapa data warehousing merupakan solusi yang tepat
- ◆ Memahami intelijen bisnis untuk suatu perusahaan

Sebagai seorang profesional teknologi informasi (TI), Anda pernah bekerja pada aplikasi komputer sebagai analis, pemrogram, perancang, pengembang, administrator basis data, atau manajer proyek. Anda telah terlibat dalam desain, implementasi, dan pemeliharaan sistem yang mendukung operasi bisnis sehari-hari. Tergantung pada industri tempat Anda bekerja, Anda pasti pernah terlibat dalam aplikasi seperti pemrosesan pesanan, buku besar, inventaris, sumber daya manusia, penggajian, penagihan rawat inap, rekening giro, klaim asuransi, dan sebagainya.



Gambar 1.1 Penggunaan data warehousing oleh organisasi.

Aplikasi ini adalah sistem penting yang menjalankan bisnis. Mereka memproses pesanan, memelihara inventaris, menyimpan pembukuan akuntansi, melayani klien, menerima pembayaran, dan memproses klaim. Tanpa sistem komputer ini, tidak ada bisnis modern yang dapat bertahan. Perusahaan mulai membangun dan menggunakan sistem ini pada tahun 1960an dan menjadi sangat bergantung pada sistem tersebut. Ketika suatu perusahaan tumbuh lebih besar, ratusan aplikasi komputer dibutuhkan untuk mendukung berbagai proses bisnis. Aplikasi ini efektif sesuai dengan apa yang dirancang untuk mereka

lakukan. Mereka mengumpulkan, menyimpan, dan memproses semua data yang diperlukan agar berhasil melakukan operasi rutin sehari-hari. Mereka menyediakan informasi online dan menghasilkan berbagai laporan untuk memantau dan menjalankan bisnis.

Pada tahun 1990-an, ketika bisnis tumbuh semakin kompleks, korporasi menyebar secara global, dan persaingan menjadi semakin ketat, para eksekutif bisnis menjadi sangat membutuhkan informasi agar tetap kompetitif dan meningkatkan keuntungan. Sistem komputer operasional memang menyediakan informasi untuk menjalankan operasi sehari-hari namun yang dibutuhkan para eksekutif adalah berbagai jenis informasi yang dapat digunakan dengan mudah untuk membuat keputusan strategis. Para pengambil keputusan ingin mengetahui wilayah geografis mana yang menjadi fokus, lini produk mana yang perlu diperluas, dan pasar mana yang perlu diperkuat. Mereka membutuhkan jenis informasi dengan isi dan format yang tepat yang dapat membantu mereka mengambil keputusan strategis. Kita dapat menyebut jenis informasi ini sebagai informasi strategis yang berbeda dengan informasi operasional.

Sistem operasional, meskipun penting, tidak dapat memberikan informasi strategis. Oleh karena itu, dunia usaha terpaksa beralih ke cara-cara baru untuk mendapatkan informasi strategis. Pergudangan data adalah paradigma baru yang secara khusus dimaksudkan untuk menyediakan informasi strategis yang penting. Pada tahun 1990an, organisasi mulai mencapai keunggulan kompetitif dengan membangun sistem data warehouse. Gambar 1-1 menunjukkan contoh area strategis di mana data warehousing telah memberikan hasil di berbagai industri. Pertama-tama, mari kita periksa pertanyaan krusialnya: mengapa perusahaan benar-benar membutuhkan data warehouse? Diskusi ini penting karena kecuali kita memahami pentingnya kebutuhan kritis ini, studi kita tentang data warehousing tidak akan memiliki motivasi. Jadi mohon diperhatikan baik-baik.

1.1 MENINGKATNYA KEBUTUHAN INFORMASI STRATEGIS

Saat kita membahas tuntutan perusahaan akan informasi strategis, kita perlu melihat krisis informasi yang menghambat mereka, serta tren teknologi dalam beberapa tahun terakhir yang menguntungkan kita, sehingga memungkinkan kita menyediakan informasi strategis. Pembahasan kita mengenai perlunya informasi strategis tidak akan lengkap kecuali kita mempelajari peluang yang diberikan oleh informasi strategis dan risiko yang dihadapi perusahaan tanpa informasi tersebut.

Siapa yang membutuhkan informasi strategis dalam suatu perusahaan? Apa sebenarnya yang dimaksud dengan informasi strategis? Para eksekutif dan manajer yang bertanggung jawab menjaga persaingan perusahaan memerlukan informasi untuk membuat keputusan yang tepat. Mereka memerlukan informasi untuk merumuskan strategi bisnis, menetapkan tujuan, menetapkan sasaran, dan memantau hasil.

Berikut adalah beberapa contoh tujuan bisnis:

- ◆ Mempertahankan basis pelanggan yang ada
- ◆ Meningkatkan basis pelanggan sebesar 15% selama 5 tahun ke depan
- ◆ Meningkatkan tingkat kualitas produk pada lima kelompok produk teratas

- ◆ Mendapatkan pangsa pasar sebesar 10% dalam 3 tahun ke depan
- ◆ Meningkatkan tingkat layanan pelanggan dalam pengiriman
- ◆ Membawa tiga produk baru ke pasar dalam 2 tahun
- ◆ Meningkatkan penjualan sebesar 15% di Divisi North East

Untuk mengambil keputusan mengenai tujuan-tujuan ini, para eksekutif dan manajer memerlukan informasi untuk tujuan berikut: untuk mendapatkan pengetahuan mendalam tentang operasi perusahaan mereka, meninjau dan memantau indikator-indikator kinerja utama dan mencatat bagaimana hal-hal ini mempengaruhi satu sama lain, melacak bagaimana faktor-faktor bisnis berubah seiring berjalannya waktu dan membandingkan kinerja perusahaan mereka relatif terhadap persaingan dan tolok ukur industri. Para eksekutif dan manajer perlu memusatkan perhatian mereka pada kebutuhan dan preferensi pelanggan, teknologi baru, hasil penjualan dan pemasaran, serta tingkat kualitas produk dan layanan. Jenis informasi yang diperlukan untuk mengambil keputusan dalam perumusan dan pelaksanaan strategi dan tujuan bisnis bersifat luas dan mencakup seluruh organisasi. Semua jenis informasi penting ini dapat digabungkan dalam klasifikasi luas yang disebut informasi strategis.

TERINTEGRASI	Harus memiliki satu pandangan seluruh perusahaan
INTEGRITAS DATA	Informasi harus akurat dan harus sesuai dengan aturan bisnis
DAPAT DIAKSES	Mudah diakses dengan jalur akses, intuitif, dan responsif terhadap analisis
KREDIBILITAS	Setiap faktor bisnis pasti mempunyai satu dan hanya satu nilai
TEPAT WAKTU	Informasi harus tersedia dalam jangka waktu yang ditentukan

Gambar 1.2 Karakteristik informasi strategis.

Informasi strategis bukan untuk menjalankan operasi bisnis sehari-hari. Hal ini tidak dimaksudkan untuk membuat faktur, melakukan pengiriman, menyelesaikan klaim, atau mengirimkan penarikan dari rekening bank. Informasi strategis jauh lebih penting bagi kelangsungan kesehatan dan kelangsungan hidup perusahaan. Keputusan bisnis yang penting bergantung pada ketersediaan informasi strategis yang tepat dalam suatu perusahaan. Gambar 1.2 mencantumkan karakteristik informasi strategis yang diinginkan.

Krisis Informasi

Anda mungkin bekerja di departemen TI pada konglomerat besar atau Anda mungkin menjadi bagian dari perusahaan menengah. Apa pun ukuran perusahaan Anda, pikirkan berbagai macam aplikasi komputer di perusahaan Anda. Pikirkan semua database dan jumlah data yang mendukung operasional perusahaan Anda. Berapa tahun data pelanggan disimpan dan tersedia? Berapa tahun data keuangan disimpan? Sepuluh tahun? Lima belas tahun? Di mana semua data ini? Di satu platform? Dalam sistem warisan? Dalam aplikasi klien/server?

Kita dihadapkan pada dua fakta mengejutkan: (1) organisasi memiliki banyak data, (2) sumber daya dan sistem teknologi informasi tidak efektif dalam mengubah semua data tersebut menjadi informasi strategis yang berguna. Selama dua dekade terakhir, perusahaan

telah mengumpulkan berton-ton data tentang operasi mereka. Ada banyak sekali data. Informasi dikatakan berlipat ganda setiap 18 bulan.

Jika kita memiliki data dalam jumlah besar di organisasi kita, mengapa para eksekutif dan manajer kita tidak dapat menggunakan data ini untuk membuat keputusan strategis? Banyak sekali informasi yang ada. Lalu mengapa kita berbicara tentang krisis informasi? Kebanyakan perusahaan dihadapkan pada krisis informasi bukan karena kurangnya data yang memadai, namun karena data yang tersedia tidak siap digunakan untuk pengambilan keputusan strategis. Data dalam jumlah besar ini sangat berguna dan baik untuk menjalankan operasi bisnis, namun sulit digunakan dalam pengambilan keputusan mengenai strategi dan tujuan bisnis.

Mengapa demikian? Pertama, data suatu perusahaan tersebar di berbagai jenis struktur dan sistem yang tidak kompatibel. Sistem pemrosesan pesanan Anda mungkin telah dikembangkan 25 tahun yang lalu dan masih berjalan pada mainframe lama. Mungkin sebagian datanya masih ada di file VSAM. Penugasan kredit dan sistem verifikasi Anda nanti mungkin ada pada platform klien/server dan data untuk aplikasi ini mungkin ada dalam tabel relasional. Data di suatu perusahaan berada di berbagai sistem yang berbeda, berbagai platform, dan struktur yang beragam. Semakin banyak teknologi yang digunakan perusahaan Anda di masa lalu, semakin banyak pula perbedaan data di perusahaan Anda. Namun, untuk pengambilan keputusan yang tepat mengenai strategi dan tujuan perusahaan secara keseluruhan, kita memerlukan informasi yang terintegrasi dari semua sistem.

Data yang diperlukan untuk pengambilan keputusan strategis harus dalam format yang sesuai agar analisis mudah mengetahui tren. Para eksekutif dan manajer perlu melihat tren dari waktu ke waktu dan mengarahkan perusahaan mereka ke arah yang tepat. Banyaknya data operasional yang tersedia tidak dapat langsung digunakan untuk melihat tren. Data operasional didasarkan pada peristiwa. Anda mendapatkan gambaran transaksi yang terjadi pada waktu tertentu. Anda memiliki data tentang unit penjualan suatu produk dalam urutan tertentu pada tanggal tertentu kepada pelanggan tertentu. Dalam sistem operasional, Anda tidak mudah mengetahui tren suatu produk selama periode satu bulan, satu kuartal, atau satu tahun.

Untuk pengambilan keputusan strategis, eksekutif dan manajer harus mampu meninjau data dari sudut pandang bisnis yang berbeda. Misalnya, mereka harus mampu meninjau dan menganalisis kuantitas penjualan berdasarkan produk, tenaga penjualan, distrik, wilayah, dan kelompok pelanggan. Dapatkah Anda membayangkan data operasional tersedia untuk analisis semacam itu? Data operasional tidak secara langsung cocok untuk ditinjau dari sudut pandang yang berbeda.

Tren Teknologi

Kita yang telah bekerja di bidang teknologi informasi selama dua atau tiga dekade telah menyaksikan perubahan menakjubkan yang terjadi. Pertama, nama departemen komputer di suatu perusahaan berubah dari “pemrosesan data” menjadi “sistem informasi manajemen”, kemudian menjadi “sistem informasi”, dan yang terbaru menjadi “teknologi informasi”. Seluruh spektrum komputasi telah mengalami perubahan yang luar biasa. Fokus

komputasi itu sendiri telah berubah selama bertahun-tahun. Praktik lama tidak dapat memenuhi kebutuhan baru. Layar dan laporan yang telah diformat sebelumnya tidak lagi memadai untuk memenuhi kebutuhan pengguna.

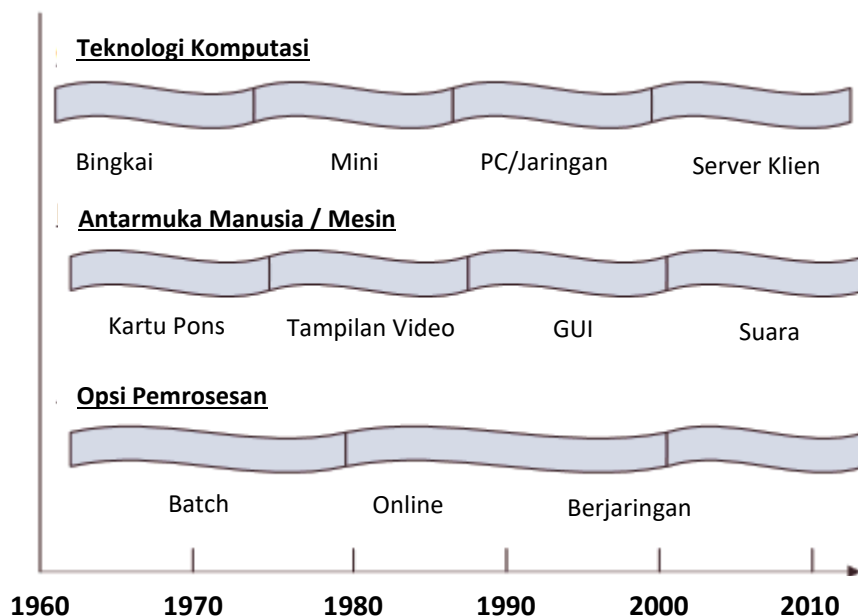
Selama bertahun-tahun, harga MIPS (juta instruksi per detik) terus menurun, biaya penyimpanan digital semakin berkurang, dan bandwidth jaringan meningkat seiring dengan penurunan harganya. Secara khusus, kami telah melihat perubahan besar di bidang-bidang penting berikut:

- ◆ Teknologi komputasi
- ◆ Antarmuka manusia–mesin
- ◆ Opsi pemrosesan

Gambar 1.3 mengilustrasikan gelombang pertumbuhan yang eksplosif ini.

Apa posisi kita saat ini dalam revolusi teknologi? Keekonomian dan miniaturisasi perangkat keras memungkinkan adanya stasiun kerja di setiap meja dan memberikan peningkatan daya sekaligus mengurangi biaya. Perangkat lunak baru menyediakan sistem yang mudah digunakan. Arsitektur sistem terbuka menciptakan kerjasama dan memungkinkan penggunaan perangkat lunak multivendor. Peningkatan konektivitas, jaringan, dan Internet membuka interaksi dengan sejumlah besar sistem dan database. Semua kemajuan teknologi ini bermanfaat. Hal ini membuat komputasi menjadi lebih cepat, lebih murah, dan tersedia secara luas. Namun apa relevansinya dengan meningkatnya kebutuhan akan informasi strategis? Mari kita pahami bagaimana kondisi teknologi saat ini kondusif dalam penyediaan informasi strategis.

Menyediakan informasi strategis memerlukan pengumpulan data perusahaan dalam jumlah besar dan menyimpannya dalam format yang sesuai. Kemajuan teknologi dalam penyimpanan data dan pengurangan biaya penyimpanan siap mengakomodasi kebutuhan penyimpanan data untuk sistem pendukung keputusan strategis. Analis, eksekutif, dan manajer menggunakan informasi strategis secara interaktif untuk menganalisis dan melihat tren bisnis. Pengguna akan mengajukan pertanyaan dan mendapatkan hasilnya, lalu mengajukan pertanyaan lain, melihat hasilnya, dan mengajukan pertanyaan lain lagi. Proses interaktif ini terus berlanjut. Kemajuan luar biasa dalam perangkat lunak antarmuka memungkinkan analisis interaktif seperti itu.



Gambar 1.3 Pertumbuhan teknologi informasi yang eksplosif.

Memproses data dalam jumlah besar dan menyediakan analisis interaktif memerlukan daya komputasi ekstra. Peningkatan daya komputasi yang eksplosif dan biaya yang lebih rendah membuat penyediaan informasi strategis menjadi mungkin dilakukan. Apa yang tidak dapat kita capai beberapa tahun sebelumnya dalam menyediakan informasi strategis kini dapat dilakukan dengan teknologi informasi yang canggih saat ini.

Peluang dan Risiko

Kami telah melihat krisis informasi yang terjadi di setiap perusahaan dan memahami bahwa meskipun terdapat banyak data operasional di perusahaan, data yang sesuai untuk pengambilan keputusan strategis tidak tersedia. Namun, kondisi teknologi saat ini memungkinkan penyediaan informasi strategis. Sementara kita masih mendiskusikan meningkatnya kebutuhan perusahaan akan informasi strategis, mari kita ajukan beberapa pertanyaan mendasar. Peluang apa saja yang tersedia bagi perusahaan akibat kemungkinan penggunaan informasi strategis? Apa saja ancaman dan risiko akibat kurangnya informasi strategis yang tersedia di perusahaan?

Berikut adalah beberapa contoh peluang yang tersedia bagi perusahaan melalui penggunaan informasi strategis:

- ◆ Unit bisnis dari operator telepon jarak jauh terkemuka memberdayakan staf penjualannya untuk membuat keputusan bisnis yang lebih baik dan dengan demikian meraup lebih banyak bisnis di pasar yang sangat kompetitif dan bernilai miliaran dolar. Solusi yang dapat diakses melalui Web mengumpulkan data internal dan eksternal untuk memberikan informasi strategis.
- ◆ Ketersediaan informasi strategis di salah satu bank terbesar di Amerika Serikat dengan aset sekitar \$250 miliar memungkinkan pengguna mengambil keputusan dengan cepat untuk mempertahankan pelanggan berharga mereka.

- ◆ Dalam kasus sebuah organisasi manajemen kesehatan yang besar, terdapat perbaikan yang signifikan dalam program layanan kesehatan, yang mengakibatkan penurunan kunjungan ke ruang gawat darurat sebesar 22%, penurunan jumlah pasien asma yang dirawat di rumah sakit sebesar 29%, dan pemeriksaan yang berpotensi menyelamatkan penglihatan bagi ratusan penderita diabetes, peningkatan tingkat vaksinasi, dan lebih dari 100.000 laporan kinerja dibuat setiap tahun untuk dokter dan apoteker.
- ◆ Di salah satu dari lima pengecer terbesar, informasi strategis yang dikombinasikan dengan alat analisis berbasis Web memungkinkan pedagang mendapatkan wawasan tentang basis pelanggan mereka, mengelola inventaris dengan lebih ketat, dan menyajikan produk yang tepat kepada orang yang tepat di tempat yang tepat di waktu yang tepat.
- ◆ Apotek berbasis komunitas yang bersaing dalam skala nasional dengan lebih dari 800 apotek waralaba dari seluruh wilayah memperoleh pemahaman mendalam tentang apa yang dibeli pelanggan, sehingga mengurangi tingkat persediaan, meningkatkan efektivitas promosi dan kampanye pemasaran, dan meningkatkan profitabilitas bagi perusahaan.
- ◆ Sebuah perusahaan elektronik besar menghemat Miliyaran Rupiah per tahun karena pengelolaan persediaan yang lebih baik.

Di sisi lain, pertimbangkan kasus-kasus berikut di mana risiko dan ancaman kegagalan sudah ada sebelum informasi strategis tersedia untuk analisis dan pengambilan keputusan:

- ◆ Dengan armada rata-rata sekitar 150.000 kendaraan, perusahaan penyewaan mobil berskala nasional dapat dengan mudah mendapat kerugian jika manajemen armada tidak efektif. Armada adalah biaya terbesar dalam bisnis itu. Dengan semakin ketatnya persaingan, potensi kegagalan sangat besar jika armada tidak dikelola secara efektif. Waktu idle mobil harus dijaga seminimal mungkin. Dalam upaya mencapai hal ini, kegagalan dalam menyediakan kelas mobil yang tepat di tempat yang tepat pada waktu yang tepat, dalam kondisi sudah dicuci dan siap, dapat menyebabkan kerugian serius dalam bisnis.
- ◆ Bagi pemasok sistem dan komponen terkemuka di dunia untuk produsen peralatan mobil dan truk ringan, tantangan serius yang dihadapi mencakup komputasi data yang tidak konsisten di hampir 100 pabrik, ketidakmampuan untuk mengukur metrik kualitas, dan pengumpulan data manual yang memakan waktu. Laporan yang dibutuhkan untuk mendukung pengambilan keputusan membutuhkan waktu berminggu-minggu. Mendapatkan informasi terintegrasi di seluruh perusahaan tidak pernah mudah.
- ◆ Bagi perusahaan utilitas besar yang menyediakan listrik kepada sekitar 25 juta konsumen di lima negara bagian Atlantik tengah di Amerika Serikat, deregulasi dapat mengakibatkan sedikit pihak yang diuntungkan dan banyak pihak yang dirugikan. Tetap kompetitif dan bahkan mungkin bertahan bergantung pada pemusatan

informasi strategis dari berbagai sumber, penyederhanaan akses data, dan fasilitasi analisis informasi oleh unit bisnis.

1.2 KEGAGALAN SISTEM PENDUKUNG KEPUTUSAN MASA LALU

Asumsikan skenario tertentu. Departemen pemasaran di perusahaan Anda mengkhawatirkan kinerja wilayah Pantai Barat dan angka penjualan dari laporan bulanan bulan ini sangat rendah. Wakil presiden pemasaran gelisah dan ingin mendapatkan beberapa laporan dari departemen TI untuk menganalisis kinerja selama dua tahun terakhir, produk demi produk, dan dibandingkan dengan target bulanan. Dia ingin mengambil keputusan strategis dengan cepat untuk memperbaiki situasi. CIO ingin atasan Anda menyampaikan laporan sesegera mungkin. Atasan Anda mendatangi Anda dan meminta Anda menghentikan semuanya dan mengerjakan laporan. Tidak ada laporan rutin dari sistem mana pun yang dapat memberikan apa yang diinginkan departemen pemasaran. Anda harus mengumpulkan data dari berbagai aplikasi dan memulai dari awal. Apakah ini terdengar familier?

Pada suatu saat dalam karir Anda di bidang teknologi informasi, Anda pasti pernah dihadapkan pada situasi seperti ini. Terkadang, Anda mungkin bisa mendapatkan informasi yang diperlukan untuk laporan ad hoc tersebut dari database atau file satu aplikasi. Biasanya tidak demikian. Anda mungkin harus membuka beberapa aplikasi, mungkin berjalan pada platform berbeda di lingkungan perusahaan Anda, untuk mendapatkan informasi. Apa yang terjadi selanjutnya? Departemen pemasaran menyukai laporan ad hoc yang Anda buat. Namun kini mereka menginginkan laporan dalam format berbeda, berisi lebih banyak informasi yang tidak terpikirkan oleh mereka pada awalnya. Setelah putaran kedua, mereka menemukan bahwa isi laporan masih belum sesuai dengan keinginan mereka. Mereka mungkin juga menemukan ketidakkonsistenan antara data yang diperoleh dari aplikasi berbeda.

Faktanya adalah bahwa selama hampir dua dekade atau lebih, departemen TI telah berupaya memberikan informasi kepada personel kunci di perusahaan mereka untuk membuat keputusan strategis. Terkadang departemen TI dapat menghasilkan laporan ad hoc dari satu aplikasi. Dalam kebanyakan kasus, laporan memerlukan data dari berbagai sistem, sehingga memerlukan penulisan program ekstrak untuk membuat file perantara yang dapat digunakan untuk menghasilkan laporan ad hoc. Sebagian besar upaya TI di masa lalu berakhir dengan kegagalan. Pengguna tidak dapat dengan jelas mendefinisikan apa yang mereka inginkan. Begitu mereka melihat kumpulan laporan pertama, mereka menginginkan lebih banyak data dalam format berbeda. Rantai itu berlanjut. Hal ini terutama disebabkan oleh sifat proses pengambilan keputusan strategis. Informasi yang diperlukan untuk pengambilan keputusan strategis harus tersedia secara interaktif. Pengguna harus dapat melakukan kueri secara online, mendapatkan hasil, dan melakukan kueri lainnya. Informasi tersebut harus dalam format yang sesuai untuk analisis.

Untuk memahami alasan kegagalan TI dalam menyediakan informasi strategis di masa lalu, kita perlu mempertimbangkan bagaimana TI berupaya melakukan hal ini selama ini. Oleh karena itu, mari kita segera melihat sejarah singkat sistem pendukung keputusan.

Sejarah Sistem Pendukung Keputusan

Tergantung pada ukuran dan sifat bisnisnya, sebagian besar perusahaan telah melalui tahapan upaya berikut untuk menyediakan informasi strategis untuk pengambilan keputusan. Laporan Ad hoc. Ini adalah tahap paling awal. Pengguna, terutama dari pemasaran dan keuangan, akan mengirimkan permintaan laporan khusus ke TI. TI akan menulis program khusus, biasanya satu untuk setiap permintaan, dan menghasilkan laporan ad hoc.

Program Ekstrak Khusus. Tahap ini merupakan upaya TI untuk mengantisipasi jenis-jenis laporan yang akan diminta dari waktu ke waktu. TI akan menulis serangkaian program dan menjalankan program secara berkala untuk mengekstrak data dari berbagai aplikasi. TI akan membuat dan menyimpan file ekstrak untuk memenuhi permintaan laporan khusus. Untuk laporan apa pun yang tidak dapat dijalankan dari file yang diekstraksi, TI akan menulis program khusus individual.

Aplikasi Kecil. Pada tahap ini, TI meresmikan proses ekstrak. TI akan membuat aplikasi sederhana berdasarkan file yang diekstraksi. Pengguna dapat menetapkan parameter untuk setiap laporan khusus. Program pencetakan laporan akan mencetak informasi berdasarkan parameter spesifik pengguna. Beberapa aplikasi tingkat lanjut juga memungkinkan pengguna melihat informasi melalui layar online.

Pusat Informasi. Pada awal tahun 1970-an, beberapa perusahaan besar mendirikan apa yang disebut pusat informasi. Pusat informasi biasanya merupakan tempat di mana pengguna dapat meminta laporan ad hoc atau melihat informasi khusus di layar. Ini adalah laporan atau layar yang telah ditentukan sebelumnya. Personil IT hadir di pusat informasi tersebut untuk membantu pengguna mendapatkan informasi yang diinginkan.

Sistem Pendukung Keputusan. Pada tahap ini, perusahaan mulai membangun sistem yang lebih canggih yang dimaksudkan untuk menyediakan informasi strategis. Sekali lagi, mirip dengan upaya sebelumnya, sistem ini didukung oleh file yang diekstraksi. Sistemnya berbasis menu dan menyediakan informasi online serta kemampuan untuk mencetak laporan khusus. Banyak sistem pendukung keputusan yang ditujukan untuk pemasaran.

Sistem Informasi Eksekutif. Ini merupakan upaya untuk membawa informasi strategis ke desktop eksekutif. Kriteria utamanya adalah kesederhanaan dan kemudahan penggunaan. Sistem akan menampilkan informasi penting setiap hari dan menyediakan kemampuan untuk meminta laporan yang sederhana dan lugas. Namun, hanya layar dan laporan terprogram yang tersedia. Setelah melihat total penjualan di seluruh negeri, jika eksekutif ingin melihat analisis berdasarkan wilayah, produk, atau dimensi lain, hal ini tidak mungkin dilakukan kecuali perincian tersebut sudah diprogram sebelumnya. Keterbatasan ini menyebabkan frustrasi dan sistem informasi eksekutif tidak bertahan lama di banyak perusahaan.

Ketidakmampuan Memberikan Informasi

Setiap upaya yang dilakukan di masa lalu dalam memberikan informasi strategis kepada pengambil keputusan tidak memuaskan. Gambar 1-4 menggambarkan kurangnya upaya TI dalam menyediakan informasi strategis. Sebagai profesional TI, kita semua akrab dengan situasi ini.



Gambar 1.4 Kurangnya upaya TI dalam menyediakan informasi strategis.

Berikut beberapa faktor yang berkaitan dengan ketidakmampuan memberikan informasi strategis:

- ◆ TI menerima terlalu banyak permintaan ad hoc, yang mengakibatkan kelebihan beban yang besar. Dengan sumber daya yang terbatas, TI tidak dapat menanggapi berbagai permintaan secara tepat waktu.
- ◆ Permintaan terlalu banyak; mereka juga terus berubah setiap saat. Pengguna memerlukan lebih banyak laporan untuk memperluas dan memahami laporan sebelumnya.
- ◆ Pengguna mendapati bahwa mereka sering kali meminta lebih banyak laporan tambahan, sehingga terkadang mereka beradaptasi dengan meminta setiap kemungkinan kombinasi, yang hanya akan semakin menambah beban TI.
- ◆ Pengguna harus bergantung pada TI untuk menyediakan informasi. Mereka tidak dapat mengakses informasi itu sendiri secara interaktif.
- ◆ Lingkungan informasi yang cocok untuk pengambilan keputusan strategis harus sangat fleksibel dan kondusif untuk analisis. TI tidak mampu menyediakan lingkungan seperti itu.

1.3 SISTEM OPERASIONAL VERSUS PENDUKUNG KEPUTUSAN

Apakah ada alasan mendasar atas kegagalan semua upaya TI sebelumnya dalam menyediakan informasi strategis? Apa yang telah dilakukan TI selama ini? Alasan mendasar ketidakmampuan menyediakan informasi strategis adalah karena selama ini kami berusaha menyediakan informasi strategis dari sistem operasional. Sistem operasional seperti

pemrosesan pesanan, pengendalian inventaris, pemrosesan klaim, penagihan rawat jalan, dan sebagainya tidak dirancang atau dimaksudkan untuk memberikan informasi strategis. Jika kita memerlukan kemampuan untuk menyediakan informasi strategis, kita harus mendapatkan informasi dari berbagai jenis sistem. Hanya sistem pendukung keputusan atau sistem informasi yang dirancang khusus yang dapat memberikan informasi strategis. Mari kita pahami alasannya.

Masukan Data



Gambar 1.5 Sistem operasional.

Memutar Roda Bisnis

Sistem operasional adalah sistem pemrosesan transaksi online (OLTP). Ini adalah sistem yang digunakan untuk menjalankan bisnis inti perusahaan sehari-hari. Inilah yang disebut sistem roti dan mentega. Sistem operasional membuat roda bisnis berputar (lihat Gambar 1.5). Mereka mendukung proses bisnis dasar perusahaan. Sistem ini biasanya memasukkan data ke dalam database. Setiap transaksi memproses informasi tentang satu entitas seperti satu pesanan, satu faktur, atau satu pelanggan.

Menyaksikan Putaran Roda Bisnis

Di sisi lain, sistem pendukung keputusan yang dirancang dan dibangun secara khusus tidak dimaksudkan untuk menjalankan proses bisnis inti. Mereka terbiasa mengamati bagaimana bisnis berjalan, dan kemudian membuat keputusan strategis untuk meningkatkan bisnis (lihat Gambar 1.6).

Sistem pendukung keputusan dikembangkan untuk mendapatkan informasi strategis dari database, berbeda dengan sistem OLTP yang dirancang untuk memasukkan data ke dalam database. Sistem pendukung keputusan dikembangkan untuk menyediakan informasi strategis.

Beda Ruang Lingkup, Beda Tujuan

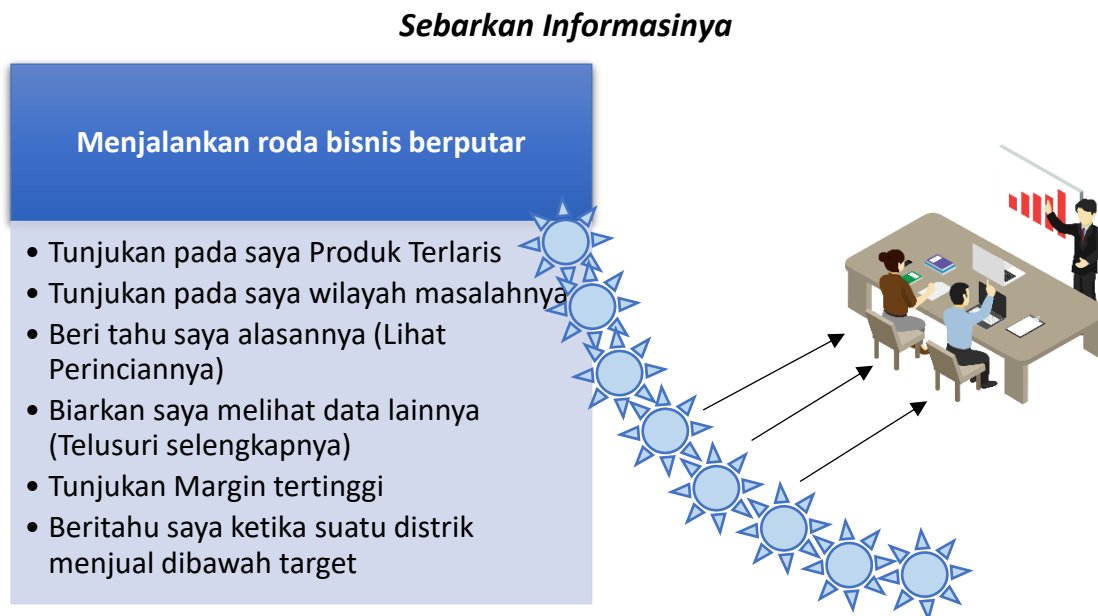
Oleh karena itu, kami menemukan bahwa untuk menyediakan informasi strategis, kami perlu membangun sistem informasi yang berbeda dari sistem operasional yang telah

kami bangun untuk menjalankan bisnis dasar. Tidak ada gunanya terus menyelami sistem operasional untuk mendapatkan informasi strategis seperti yang telah kita lakukan di masa lalu. Ketika perusahaan menghadapi persaingan yang semakin ketat dan bisnis menjadi lebih kompleks, meneruskan praktik-praktik lama hanya akan membawa bencana.

Kita perlu merancang dan membangun sistem informasi

- ◆ Hal tersebut memiliki tujuan yang berbeda
- ◆ Yang cakupannya berbeda
- ◆ Yang isi datanya berbeda
- ◆ Dimana pola penggunaan datanya berbeda-beda
- ◆ Dimana tipe akses datanya berbeda

Gambar 1.7 merangkum perbedaan antara sistem operasional tradisional dan sistem informasi baru yang perlu dibangun.



Gambar 1.6 Sistem pendukung keputusan.

Bagaimana Mereka Berbeda?		
	OPERASIONAL	INFORMATIONAL
Konten Data	Nilai – nilai saat ini	Diarsipkan, diturunkan, diringkas
Struktur Data	Dioptimalkan untuk transaksi	Dioptimalkan untuk kueri yang kompleks
Frekuensi Akses	Tinggi	Sedang hingga rendah
Jenis Akses	Read, Update, delete	Read
Penggunaan	Dapat diprediksi, berulang	Ad Hoc, Acak, Heuristik
Waktu Merespon	Sub-detik	Beberapa menit hingga menit
Pengguna	Jumlah besar	Jumlah yang relatif kecil

Gambar 1.7 Sistem operasional dan informasi.

1.4 DATA WAREHOUSING, SATU-SATUNYA SOLUSI

Pada tahap diskusi ini, kami kini menyadari bahwa kami memerlukan berbagai jenis sistem pendukung keputusan untuk menyediakan informasi strategis. Jenis informasi yang dibutuhkan untuk pengambilan keputusan strategis berbeda dengan yang tersedia dari sistem operasional. Kita memerlukan lingkungan sistem jenis baru yang bertujuan untuk menyediakan informasi strategis untuk analisis, membedakan tren, dan memantau kinerja. Mari kita periksa fitur yang diinginkan dan persyaratan pemrosesan dari lingkungan sistem jenis baru ini. Mari kita pertimbangkan juga keuntungan dari lingkungan sistem jenis ini yang dirancang untuk informasi strategis.

Lingkungan Sistem Tipe Baru

Fitur yang diinginkan dari lingkungan sistem tipe baru adalah:

- ◆ Basis data dirancang untuk tugas analitis
- ◆ Data dari beberapa aplikasi
- ◆ Mudah digunakan dan kondusif untuk sesi interaktif yang panjang oleh pengguna
- ◆ Penggunaan data yang intensif membaca
- ◆ Interaksi langsung dengan sistem oleh pengguna tanpa bantuan TI
- ◆ Konten diperbarui secara berkala dan stabil
- ◆ Konten untuk menyertakan data terkini dan historis
- ◆ Kemampuan bagi pengguna untuk menjalankan kueri dan mendapatkan hasil secara online
- ◆ Kemampuan bagi pengguna untuk memulai laporan

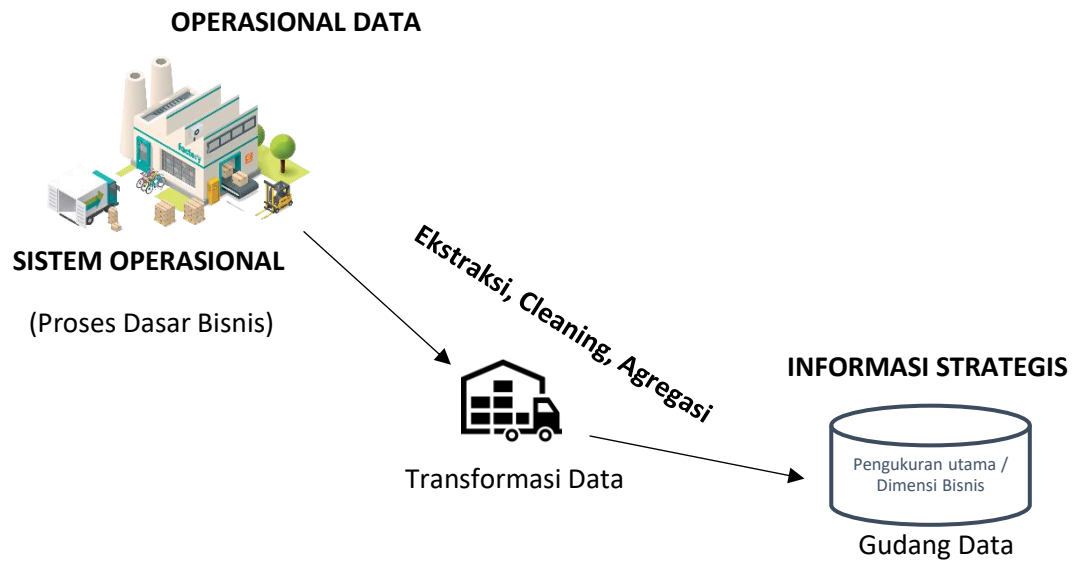
Persyaratan Pemrosesan di Lingkungan Baru

Sebagian besar pemrosesan informasi strategis di lingkungan baru harus bersifat analitis. Setidaknya ada empat tingkat persyaratan pemrosesan analitis:

1. Menjalankan pertanyaan dan laporan sederhana terhadap data terkini dan historis.
2. Kemampuan untuk melakukan analisis “bagaimana jika” dalam berbagai cara.
3. Kemampuan untuk melakukan kueri, mundur, menganalisis, dan kemudian melanjutkan proses hingga durasi yang diinginkan.
4. Kemampuan untuk melihat tren historis dan menerapkannya dalam proses interaktif di masa depan.

Informasi Strategis dari Data Warehouse

Lingkungan sistem baru yang sangat dibutuhkan pengguna untuk memperoleh informasi strategis ini merupakan paradigma baru pergudangan data. Dimulai pada akhir tahun 1980an dan memasuki tahun 1990an perusahaan mulai membangun lingkungan sistem seperti itu. Lingkungan baru ini, yang dikenal sebagai lingkungan data warehouse, dipisahkan dari lingkungan sistem yang mendukung operasi rutin sehari-hari. Gudang data pada dasarnya telah menjadi sumber informasi strategis bagi perusahaan untuk memungkinkan pengambilan keputusan strategis. Gudang data telah terbukti menjadi satu-satunya solusi yang layak. Kami telah melihat dengan jelas bahwa solusi yang didasarkan pada data yang diambil dari sistem operasional semuanya tidak memuaskan. Gambar 1-8 menunjukkan gambaran umum data warehouse sebagai sumber informasi strategis bagi perusahaan.



Gambar 1.8 Gambaran umum gudang data.

Pada interpretasi tingkat tinggi, gudang data berisi metrik penting dari proses bisnis yang disimpan sepanjang dimensi bisnis. Misalnya, gudang data mungkin berisi unit penjualan, berdasarkan produk, hari, kelompok pelanggan, distrik penjualan, wilayah penjualan, dan promosi. Di sini dimensi bisnisnya adalah produk, hari, kelompok pelanggan, distrik penjualan, wilayah penjualan, dan jenis promosi penjualan. Penjualan unit mewakili metrik yang diukur pada produk, hari, kelompok pelanggan, distrik penjualan, wilayah penjualan, dan jenis promosi penjualan.

Dari mana data warehouse mendapatkan datanya? Data tersebut berasal dari sistem operasional yang mendukung proses bisnis dasar organisasi. Di antara sistem operasional dan gudang data, terdapat area pementasan data. Di area pementasan ini, data operasional dibersihkan dan diubah menjadi bentuk yang sesuai untuk ditempatkan di gudang data agar mudah diambil.

1.5 DATA GUDANG DITETAPKAN

Kami telah mencapai kesimpulan kuat bahwa data warehousing adalah satu-satunya solusi yang layak untuk menyediakan informasi strategis. Kami sampai pada kesimpulan ini berdasarkan fungsi lingkungan sistem baru yang disebut gudang data. Jadi, mari kita coba memberikan definisi fungsional dari data warehouse.

Gudang data adalah lingkungan informasi yang:

- ◆ Memberikan pandangan yang terpadu dan total mengenai perusahaan.
- ◆ Membuat informasi terkini dan historis perusahaan tersedia dengan mudah untuk pengambilan keputusan strategis.
- ◆ Memungkinkan transaksi pendukung keputusan tanpa menghambat sistem operasional.
- ◆ Membuat informasi organisasi konsisten.
- ◆ Menyajikan sumber informasi strategis yang fleksibel dan interaktif.

Konsep Sederhana untuk Penyampaian Informasi

Dalam analisis akhir, data warehousing adalah konsep sederhana. Hal ini lahir dari kebutuhan akan informasi strategis dan merupakan hasil pencarian cara baru untuk menyediakan informasi tersebut. Metode dekade sebelumnya yang menggunakan lingkungan komputasi operasional tidak memuaskan. Konsep baru ini bukan untuk menghasilkan data baru, namun untuk memanfaatkan sejumlah besar data yang ada dan mengubahnya menjadi bentuk yang sesuai untuk menyediakan informasi strategis.

Gudang data ada untuk menjawab pertanyaan pengguna tentang bisnis, kinerja berbagai operasi, tren bisnis, dan tentang apa yang dapat dilakukan untuk meningkatkan bisnis. Gudang data ada untuk memberikan pengguna bisnis akses langsung ke data, untuk menyediakan satu versi terpadu dari indikator kinerja utama, untuk mencatat masa lalu secara akurat, dan untuk memberikan kemampuan untuk melihat data dari berbagai perspektif. Singkatnya, gudang data ada untuk mendukung proses pengambilan keputusan. Pergudangan data sebenarnya merupakan konsep sederhana: Ambil semua data yang sudah Anda miliki di organisasi, bersihkan dan ubah, lalu berikan informasi strategis yang berguna. Apa yang lebih sederhana dari itu?

Lingkungan, Bukan Produk

Gudang data bukanlah produk perangkat lunak atau perangkat keras tunggal yang Anda beli untuk menyediakan informasi strategis. Sebaliknya, ini adalah lingkungan komputasi di mana pengguna dapat menemukan informasi strategis, sebuah lingkungan di mana pengguna terhubung langsung dengan data yang mereka perlukan untuk membuat keputusan yang lebih baik. Ini adalah lingkungan yang berpusat pada pengguna.

Mari kita rangkum karakteristik lingkungan komputasi baru yang disebut gudang data:

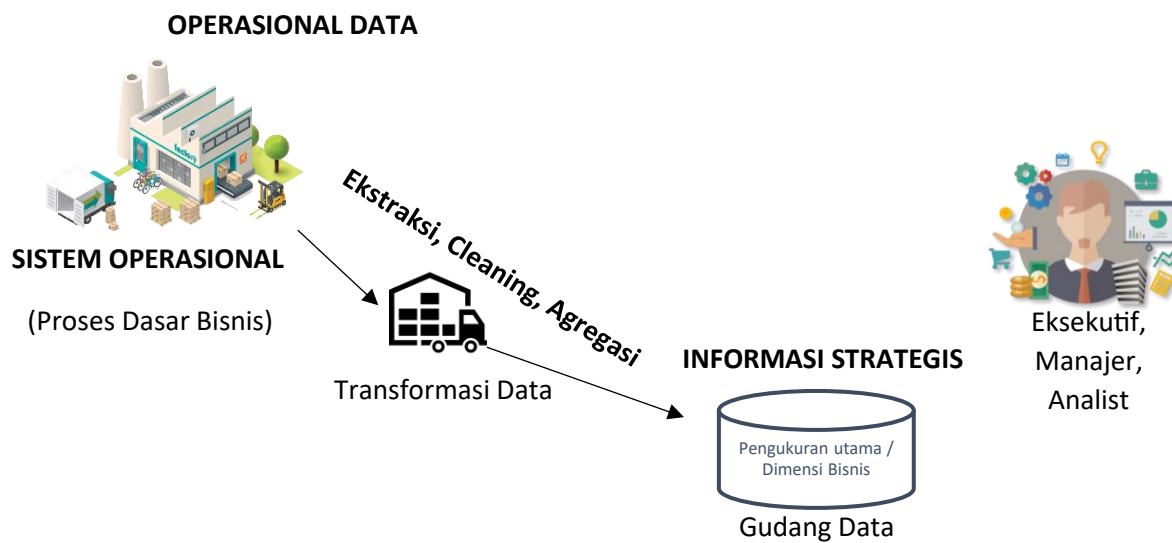
- ◆ Lingkungan yang ideal untuk analisis data dan dukungan keputusan
- ◆ Lancar, fleksibel, dan interaktif
- ◆ 100% didorong oleh pengguna
- ◆ Sangat responsif dan kondusif terhadap pola tanya-jawab-tanya lagi
- ◆ Memberikan kemampuan untuk menemukan jawaban atas pertanyaan yang kompleks dan tidak dapat diprediksi

Perpaduan Banyak Teknologi

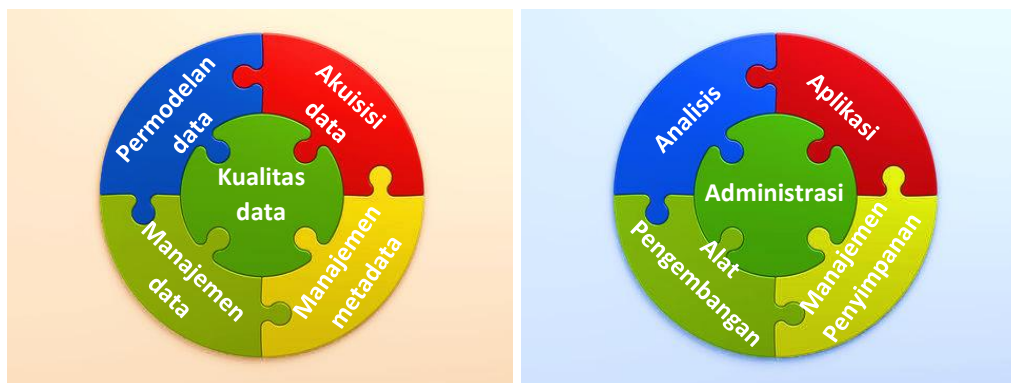
Mari kita periksa kembali konsep dasar data warehousing. Konsep dasar data warehousing adalah:

- ◆ Mengambil semua data dari sistem operasional.
- ◆ Bila perlu, sertakan data relevan dari luar, seperti indikator tolok ukur industri.
- ◆ Integrasikan semua data dari berbagai sumber.
- ◆ Hapus inkonsistensi dan transformasi data.
- ◆ Menyimpan data dalam format yang sesuai untuk kemudahan akses pengambilan keputusan.

Oleh karena itu, diperlukan teknologi yang berbeda untuk mendukung fungsi-fungsi ini. Gambar 1.9 menunjukkan bagaimana data warehouse merupakan perpaduan dari banyak teknologi yang diperlukan untuk berbagai fungsi.



PERCAMPURAN TEKNOLOGI



Gambar 1.9 Gudang data: perpaduan teknologi.

Meskipun banyak teknologi yang digunakan, semuanya bekerja sama dalam gudang data. Hasil akhirnya adalah terciptanya lingkungan komputasi baru yang bertujuan untuk menyediakan informasi strategis yang sangat dibutuhkan setiap perusahaan.

1.6 GERAKAN PERGUDANGAN DATA

Ketika perusahaan mulai menyadari efektivitas data warehousing, semakin banyak organisasi yang ikut serta dan data warehousing mulai menyebar dengan cepat. Pertama, perusahaan besar yang mampu dengan cepat membayar pengeluaran sumber daya mulai meluncurkan proyek data warehousing. Perusahaan skala menengah juga memasuki arena data warehousing. Segera beberapa bisnis mulai merasakan manfaat yang diberikan oleh data warehousing. Banyak penelitian mulai terfokus pada fenomena baru ini. Banyak vendor mulai menawarkan produk perangkat keras dan perangkat lunak untuk mendukung berbagai fungsi dalam gudang data.

Sebelum konsep data warehousing dengan model arsitektur untuk pergerakan data dari sistem operasional ke lingkungan pendukung keputusan, perusahaan mencoba berbagai lingkungan pendukung keputusan dalam organisasi mereka. Hal ini harus dilakukan dengan

biaya yang sangat besar dan disertai dengan redundansi dan inkonsistensi data dalam jumlah besar. Setiap lingkungan pendukung keputusan dimaksudkan untuk melayani kelompok pengguna tertentu untuk tujuan terbatas. Penerapan data warehousing mengubah semua ini. Mirip dengan gudang industri, gudang data ditujukan untuk pengumpulan dan penyimpanan data perusahaan dalam skala besar guna menyediakan informasi strategis untuk kebutuhan keseluruhan. Sama seperti produk yang disimpan di gudang industri didistribusikan ke toko ritel atau pasar, data yang disimpan di gudang data dapat disalurkan ke data mart untuk pengguna tertentu.

Tonggak Pencapaian Pergudangan Data

Ketika data warehousing mulai diterima pada tahun 1980an dan 1990an, kita dapat menelusuri beberapa hal penting dari gerakan ini. Di bawah ini adalah tonggak-tonggak penting selama fase awal gerakan ini:

- ◆ 1983; Teradata memperkenalkan sistem manajemen basis data (DBMS) yang dirancang untuk sistem pendukung keputusan.
- ◆ 1988; Artikel An Architecture for a Business and Information Systems yang memperkenalkan istilah “business data warehouse” diterbitkan oleh Barry Devlin dan Paul Murphy di IBM Systems Journal.
- ◆ 1990; Red Brick Systems memperkenalkan Red Brick Warehouse, sebuah DBMS khusus untuk data warehousing.
- ◆ 1991; Bill Inmon menerbitkan bukunya Building the Data Warehouse (dia umumnya dianggap sebagai bapak data warehousing).
- ◆ 1991; Prism Solutions memperkenalkan perangkat lunak Prism Warehouse Manager untuk mengembangkan gudang data.
- ◆ 1995; Data Warehousing Institute, sebuah institusi terkemuka yang mempromosikan data warehousing didirikan. (Lembaga ini telah muncul sebagai suara terdepan dalam arena data warehousing dan intelijen bisnis yang menyediakan pendidikan, penelitian, dan dukungan.)
- ◆ 1996; Ralph Kimball menerbitkan buku penting The Data Warehousing Toolkit. (Dia adalah salah satu otoritas tertinggi di bidang pergudangan data dan sistem pendukung keputusan.)
- ◆ 1997; Oracle 8, dengan dukungan untuk kueri skema STAR, dirilis.

Tantangan Awal

Seiring dengan berlanjutnya adopsi data warehousing oleh organisasi, bahkan perusahaan yang mengimplementasikan data warehouse pun menghadapi tantangan signifikan yang mendorong terjadinya pergeseran moderat dari implementasi awal. Berikut adalah daftar tantangan utama yang harus diatasi:

- ◆ Pelanggan menjadi lebih canggih dan cerdas, mendesak untuk mendapatkan layanan yang lebih baik, peningkatan kualitas, dan penyesuaian yang inovatif.
- ◆ Deregulasi pemerintah terhadap industri membuat perusahaan menghadapi persaingan yang lebih ketat dan kebutuhan akan operasi yang lebih ramping.

- ◆ Perluasan globalisasi membuka arena bagi para pesaing, yang jumlahnya lebih banyak dan kekuatannya lebih besar.
- ◆ Peraturan privasi baru menciptakan kebutuhan untuk merevisi metode pengumpulan dan penggunaan informasi.
- ◆ Arsitektur yang tidak tepat pada beberapa sistem gudang data awal menghasilkan pandangan data perusahaan yang terfragmentasi dan cenderung menghasilkan silo informasi yang berbeda.
- ◆ Alat kueri, pelaporan, dan analisis yang diberikan kepada pengguna di lingkungan data warehousing awal untuk layanan mandiri terbukti terlalu rumit dan membebani untuk digunakan oleh pengguna itu sendiri.
- ◆ Janji-janji mengenai lingkungan gudang data awal untuk menyediakan alat yang mudah digunakan bagi banyak orang masih belum terpenuhi.

1.7 EVOLUSI KECERDASAN BISNIS

Tantangan awal setelah penerapan sistem data warehousing awal memaksa perusahaan untuk mempertimbangkan kembali penyediaan dukungan pengambilan keputusan. Perusahaan mulai memahami bahwa tujuan sistem pendukung keputusan ada dua: transformasi data menjadi informasi; derivasi pengetahuan dari informasi. Masing-masing dari kedua aspek ini perlu ditekankan dan diperkuat secara tepat untuk memberikan hasil yang diperlukan. Intelijen bisnis untuk suatu organisasi memerlukan dua lingkungan, satu untuk berkonsentrasi pada transformasi data menjadi informasi dan yang lainnya untuk menangani transformasi informasi menjadi pengetahuan. Oleh karena itu, intelijen bisnis (BI) adalah sekelompok aplikasi dan teknologi yang luas.

Pertama, istilah ini mengacu pada sistem dan teknologi untuk mengumpulkan, membersihkan, mengkonsolidasikan, dan menyimpan data perusahaan. Selanjutnya, intelijen bisnis berkaitan dengan alat, teknik, dan aplikasi untuk menganalisis data yang disimpan. Grup Gartner mempopulerkan BI sebagai istilah umum yang mencakup konsep dan metode untuk meningkatkan pengambilan keputusan bisnis melalui sistem pendukung berbasis fakta. Data Warehousing Institute membandingkan BI dengan kilang data. Mirip dengan kilang minyak, setting BI mengambil data sebagai bahan mentah, mengumpulkannya, menyaringnya, dan mengolahnya menjadi beberapa produk informasi.

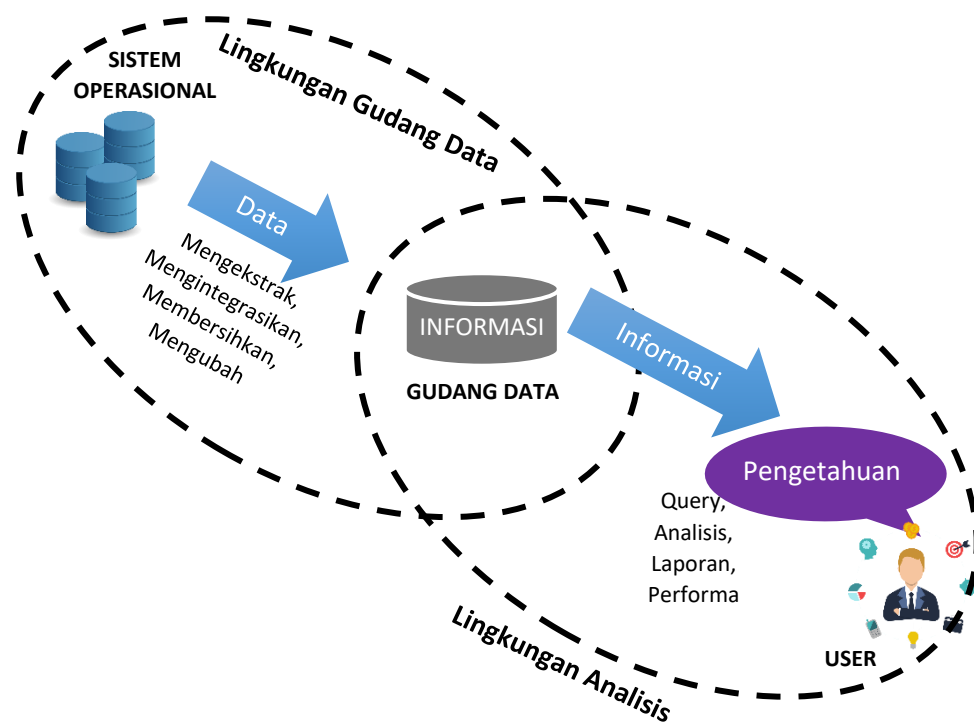
BI: Dua Lingkungan

Jika Anda mempertimbangkan semua hal yang tercakup dalam BI, Anda dapat memandang BI untuk suatu perusahaan terdiri dari dua lingkungan: Data menjadi Informasi. Dalam lingkungan ini, data dari berbagai sistem operasional diekstraksi, diintegrasikan, dibersihkan, diubah, dan disimpan sebagai informasi dalam repositori yang dirancang khusus. Informasi ke Pengetahuan. Dalam lingkungan ini alat analisis tersedia bagi pengguna untuk mengakses dan menganalisis konten informasi dalam repositori yang dirancang khusus dan mengubah informasi menjadi pengetahuan.

BI: Pergudangan Data dan Analisis

Sebagaimana ditunjukkan oleh beberapa tantangan awal, perhatian yang cukup dan terpisah perlu diberikan pada kedua lingkungan yang dicakup oleh BI. Dalam dunia bisnis saat ini, ekstraksi, konsolidasi, transformasi, dan penyimpanan data sebagai informasi strategis adalah tugas yang berat. Sekali lagi, menggunakan informasi ini dengan alat canggih untuk pengambilan keputusan yang tepat juga merupakan tantangan yang sama. Oleh karena itu, tren yang ada adalah mempertimbangkan hal ini sebagai dua lingkungan yang berbeda bagi BI korporasi. Vendor juga cenderung mengkhususkan diri pada alat yang sesuai untuk dua lingkungan berbeda ini. Namun, kedua lingkungan tersebut saling melengkapi dan perlu bekerja sama.

Gambar 1-10 menunjukkan dua lingkungan yang saling melengkapi, lingkungan pergudangan data, yang mengubah data menjadi informasi, dan lingkungan analitis, yang menghasilkan pengetahuan dari informasi. Saat kami melanjutkan dari bab ke bab, kami akan terus memperluas dan mengintensifkan diskusi kami mengenai kedua lingkungan ini.



Gambar 1.10 BI: pergudangan data dan lingkungan analitis.

RINGKASAN BAB

- ◆ Perusahaan sangat membutuhkan informasi strategis untuk menghadapi persaingan yang semakin ketat, memperluas pangsa pasar, dan meningkatkan profitabilitas.
- ◆ Meskipun banyak sekali data yang dikumpulkan oleh perusahaan selama beberapa dekade terakhir, setiap perusahaan terjebak di tengah krisis informasi. Informasi yang dibutuhkan untuk pengambilan keputusan strategis tidak tersedia.

- ◆ Semua upaya TI di masa lalu untuk menyediakan informasi strategis telah gagal. Hal ini terutama karena TI telah berusaha memberikan informasi strategis dari sistem operasional.
- ◆ Sistem informasi berbeda dengan sistem operasional tradisional. Sistem operasional tidak dirancang untuk informasi strategis.
- ◆ Kita memerlukan lingkungan komputasi jenis baru untuk menyediakan informasi strategis. Gudang data menjanjikan lingkungan komputasi baru.
- ◆ Pergudangan data adalah solusi yang tepat. Ada kebutuhan mendesak akan data warehousing di setiap perusahaan.
- ◆ Tantangan yang dihadapi dalam implementasi awal data warehouse mendorong pergerakan menuju kedewasaan.
- ◆ Gagasan intelijen bisnis bagi suatu perusahaan telah berkembang sebagai sebuah konsep umum yang mencakup pergudangan data dan analitik untuk mengubah data menjadi informasi dan informasi menjadi pengetahuan.

Latihan Soal

1. Apa yang dimaksud dengan informasi strategis? Untuk bank komersial, sebutkan lima jenis tujuan strategis.
2. Apakah Anda setuju bahwa toko ritel pada umumnya mengumpulkan data dalam jumlah besar melalui sistem operasionalnya? Sebutkan tiga jenis data transaksi yang mungkin dikumpulkan oleh toko ritel dalam volume besar selama operasi sehari-harinya.
3. Menelaah peluang yang dapat diberikan oleh informasi strategis bagi sebuah pusat kesehatan. Dapatkah Anda menyebutkan lima peluang tersebut?
4. Mengapa semua upaya TI di masa lalu untuk menyediakan informasi strategis gagal? Sebutkan tiga alasan konkrit dan jelaskan.
5. Jelaskan lima perbedaan antara sistem operasional dan sistem informasi.
6. Mengapa sistem operasional tidak cocok untuk menyediakan informasi strategis? Berikan tiga alasan spesifik dan jelaskan.
7. Sebutkan enam karakteristik lingkungan komputasi yang diperlukan untuk menyediakan informasi strategis.
8. Jenis pemrosesan apa yang dilakukan di gudang data? Menggambarkan.
9. Gudang data adalah sebuah lingkungan, bukan produk. Membahas.
10. Pergudangan data adalah satu-satunya cara yang layak untuk menyelesaikan krisis informasi dan menyediakan informasi strategis. Sebutkan empat alasan yang mendukung pernyataan ini dan jelaskan.

BAB 2

GUDANG DATA

TUJUAN BAB

- ◆ Tinjau definisi formal gudang data
- ◆ Diskusikan ciri-ciri yang menentukan
- ◆ Membedakan antara data warehouse dan data mart
- ◆ Tinjau tipe arsitektur yang berkembang
- ◆ Pelajari setiap komponen atau blok bangunan yang membentuk gudang data
- ◆ Memperkenalkan metadata dan menyoroti signifikansinya

Seperti yang telah kita lihat di bab terakhir, data warehousing telah berkembang sebagai bagian dari intelijen bisnis bagi perusahaan. Di gudang data Anda mengintegrasikan dan mengubah data perusahaan menjadi informasi yang sesuai untuk pengambilan keputusan strategis. Anda mengambil semua data historis dari berbagai sistem operasional, menggabungkan data internal ini dengan data relevan apa pun dari sumber luar, dan menggabungkannya. Anda menyelesaikan konflik apa pun dalam cara data berada di sistem yang berbeda dan mengubah konten data terintegrasi ke dalam format yang sesuai untuk menyediakan informasi ke berbagai kelas pengguna. Terakhir, Anda melengkapi dengan metode penyampaian informasi.

Untuk menyiapkan sistem penyampaian informasi ini, Anda memerlukan komponen atau blok bangunan yang berbeda. Blok-blok penyusun ini disusun bersama dengan cara yang paling optimal untuk mencapai tujuan yang dimaksudkan. Mereka diatur dalam arsitektur yang sesuai. Sebelum kita masuk ke masing-masing komponen dan susunannya dalam keseluruhan arsitektur, mari kita lihat dulu beberapa fitur mendasar dari data warehouse. Bill Inmon (1996, p. 33), yang dianggap sebagai bapak data warehousing sebagaimana disebutkan dalam bab sebelumnya, memberikan definisi berikut: “Data Warehouse adalah kumpulan data yang berorientasi subjek, terintegrasi, non-volatil, dan varian waktu dalam dukungan keputusan manajemen.”

Sean Kelly, praktisi data warehousing terkemuka lainnya, mendefinisikan data warehouse dengan cara berikut. Data yang ada di gudang data adalah:

Terpisah Tersedia Waktu Terintegrasi yang dicap
Dapat Diakses Nonvolatil Berorientasi Subjek

2.1 DEFINISI FITUR

Mari kita periksa beberapa fitur utama yang menentukan gudang data berdasarkan definisi ini. Bagaimana dengan sifat data di data warehouse? Apa perbedaan data ini dengan data dalam sistem operasional mana pun? Mengapa harus berbeda? Bagaimana konten data di gudang data digunakan?

Data Berorientasi Subjek

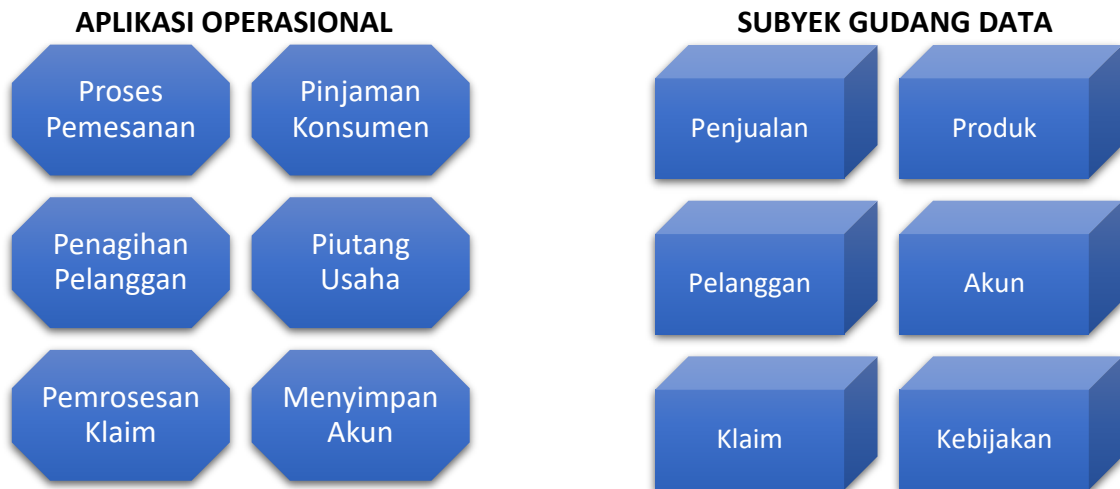
Dalam sistem operasional, kami menyimpan data berdasarkan aplikasi individual. Dalam kumpulan data untuk aplikasi pemrosesan pesanan, kami menyimpan data untuk aplikasi tersebut. Kumpulan data ini menyediakan data untuk semua fungsi memasukkan pesanan, memeriksa stok, memverifikasi kredit pelanggan, dan menetapkan pesanan untuk pengiriman. Namun kumpulan data ini hanya berisi data yang diperlukan untuk fungsi-fungsi yang berkaitan dengan aplikasi khusus ini. Kami akan memiliki beberapa kumpulan data yang berisi data tentang pesanan individu, pelanggan, status stok, dan transaksi terperinci, namun semua ini disusun berdasarkan pemrosesan pesanan.

Demikian pula halnya dengan lembaga perbankan, kumpulan data untuk aplikasi pinjaman konsumen berisi data untuk aplikasi tersebut. Kumpulan data untuk aplikasi berbeda lainnya untuk rekening giro dan rekening tabungan berhubungan dengan aplikasi spesifik tersebut. Sekali lagi, di perusahaan asuransi, kumpulan data berbeda mendukung aplikasi individual seperti asuransi mobil, asuransi jiwa, dan asuransi kompensasi pekerja. Di setiap industri, kumpulan data disusun berdasarkan aplikasi individual untuk mendukung sistem operasional tertentu. Kumpulan data individual ini harus menyediakan data untuk aplikasi spesifik agar dapat menjalankan fungsi spesifik secara efisien. Oleh karena itu, kumpulan data untuk setiap aplikasi perlu diatur berdasarkan aplikasi spesifik tersebut.

Sebaliknya, di gudang data, data disimpan oleh subjek atau peristiwa bisnis di dunia nyata, bukan oleh aplikasi. Data dalam gudang data diatur sedemikian rupa sehingga semua kumpulan data yang berkaitan dengan subjek atau peristiwa bisnis dunia nyata yang sama terikat menjadi satu. Kami telah mengatakan bahwa data dihubungkan dan disimpan oleh pelaku bisnis. Nah, apa itu mata pelajaran bisnis? Subjek bisnis berbeda dari satu perusahaan ke perusahaan lainnya. Ini adalah hal-hal yang penting bagi perusahaan. Bagi perusahaan manufaktur, penjualan, pengiriman, dan inventaris merupakan subjek bisnis yang penting. Untuk toko ritel, penjualan di konter check-out akan menjadi subjek bisnis yang penting.

Gambar 2.1 membedakan antara cara data disimpan dalam sistem operasional dan di gudang data. Dalam sistem operasional yang ditampilkan, data untuk setiap aplikasi diatur secara terpisah berdasarkan aplikasi: pemrosesan pesanan, pinjaman konsumen, penagihan pelanggan, piutang, pemrosesan klaim, dan rekening tabungan. Misalnya, klaim merupakan subjek bisnis yang penting bagi perusahaan asuransi. Klaim berdasarkan polis asuransi mobil diproses dalam aplikasi asuransi mobil. Data klaim asuransi mobil diatur dalam aplikasi itu. Demikian pula, data klaim untuk asuransi kompensasi pekerja diatur dalam aplikasi asuransi kompensasi pekerja. Namun di gudang data perusahaan asuransi, data tentang klaim diatur berdasarkan subjek klaim dan bukan berdasarkan aplikasi individual dari asuransi mobil dan perusahaan pekerja. Di gudang data, tidak ada ragam aplikasi. Data di gudang data melintasi aplikasi.

**Digudang Data, Data Tidak Disimpan Oleh Aplikasi Operasional,
Namun Oleh Subjek Bisnis**



Gambar 2.1 Gudang data berorientasi pada subjek.

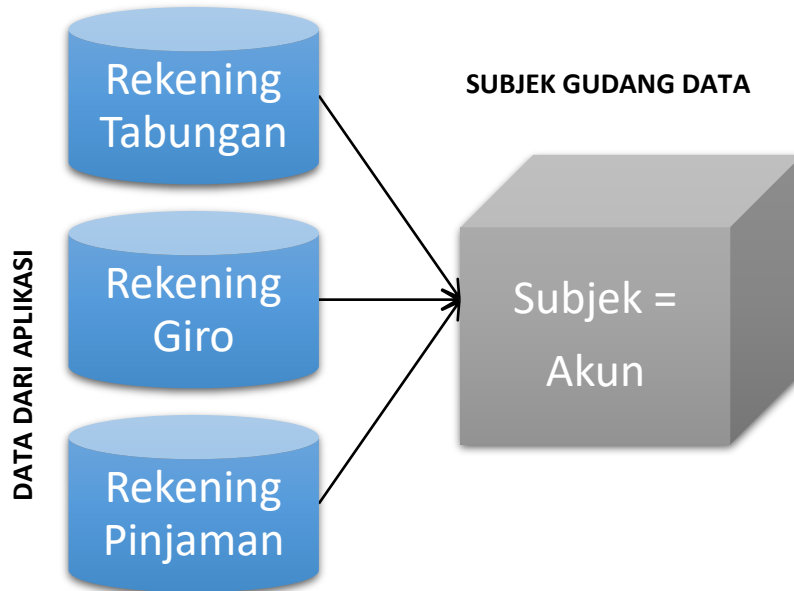
Data Terintegrasi

Untuk pengambilan keputusan yang tepat, Anda perlu mengumpulkan semua data relevan dari berbagai aplikasi. Data yang ada pada data warehouse berasal dari beberapa sistem operasional. Sumber data berada di database, file, dan segmen data yang berbeda. Ini adalah aplikasi yang berbeda, sehingga platform operasional dan sistem operasinya mungkin berbeda. Tata letak file, representasi kode karakter, dan konvensi penamaan field semuanya bisa berbeda.

Selain data dari sistem operasional internal, bagi banyak perusahaan, data dari sumber luar mungkin merupakan hal yang sangat penting. Perusahaan seperti Metro Mail, A.C. Nielsen, dan IRI mengkhususkan diri dalam menyediakan data penting secara teratur. Gudang data Anda mungkin memerlukan data dari sumber tersebut. Ini adalah satu lagi variasi dalam campuran data sumber untuk gudang data. Gambar 2.2 mengilustrasikan proses sederhana integrasi data untuk lembaga perbankan. Di sini data yang dimasukkan ke dalam area subjek akun di gudang data berasal dari tiga aplikasi operasional yang berbeda. Bahkan hanya dalam tiga kali penerapan, mungkin terdapat beberapa variasi. Konvensi penamaan bisa berbeda; atribut untuk item data bisa berbeda. Nomor rekening di aplikasi rekening tabungan bisa panjangnya delapan byte, tapi di aplikasi rekening giro hanya enam byte.

Sebelum data dari berbagai sumber berbeda dapat disimpan dengan berguna di gudang data, Anda harus menghilangkan ketidakkonsistenan. Anda harus membakukan berbagai elemen data dan memastikan arti nama data di setiap aplikasi sumber. Sebelum memindahkan data ke gudang data, Anda harus melalui proses transformasi, konsolidasi, dan integrasi sumber data.

Inkonsistensi Data Dihilangkan; Data Dari Beragam Aplikasi Operasional Terintegrasi



Gambar 2.2 Gudang data terintegrasi.

Berikut adalah beberapa item yang perlu distandarisasi dan dibuat konsisten:

- Konvensi penamaan
- Kode
- Atribut data
- Pengukuran

Data Varian Waktu

Untuk sistem operasional, data yang disimpan berisi nilai saat ini. Dalam sistem piutang, saldo adalah saldo terutang saat ini di rekening pelanggan. Dalam sistem entri pesanan, status pesanan adalah status pesanan saat ini. Dalam aplikasi pinjaman konsumen, jumlah saldo yang terutang oleh pelanggan adalah jumlah saat ini. Tentu saja, kita menyimpan beberapa transaksi masa lalu dalam sistem operasional, namun, pada dasarnya, sistem operasional mencerminkan informasi terkini karena sistem ini mendukung operasi terkini sehari-hari.

Di sisi lain, data di gudang data dimaksudkan untuk analisis dan pengambilan keputusan. Jika pengguna melihat pola pembelian pelanggan tertentu, pengguna memerlukan data tidak hanya tentang pembelian saat ini, namun juga pembelian sebelumnya. Saat pengguna ingin mengetahui alasan penurunan penjualan di divisi North East, pengguna memerlukan semua data penjualan untuk divisi tersebut selama periode waktu yang lampau. Ketika seorang analis di jaringan toko kelontong ingin mempromosikan dua atau lebih produk secara bersamaan, analis tersebut menginginkan penjualan produk yang dipilih selama beberapa kuartal terakhir.

Gudang data, karena sifat tujuannya, harus berisi data historis, bukan hanya nilai saat ini. Data disimpan sebagai snapshot dari periode yang lalu dan saat ini. Perubahan pada data dilacak dan dicatat sehingga, jika perlu, laporan dapat dibuat untuk menunjukkan perubahan dari waktu ke waktu. Setiap struktur data di gudang data mengandung elemen waktu. Anda akan menemukan cuplikan historis data operasional di gudang data. Aspek data warehouse ini cukup signifikan baik untuk tahap desain maupun implementasi.

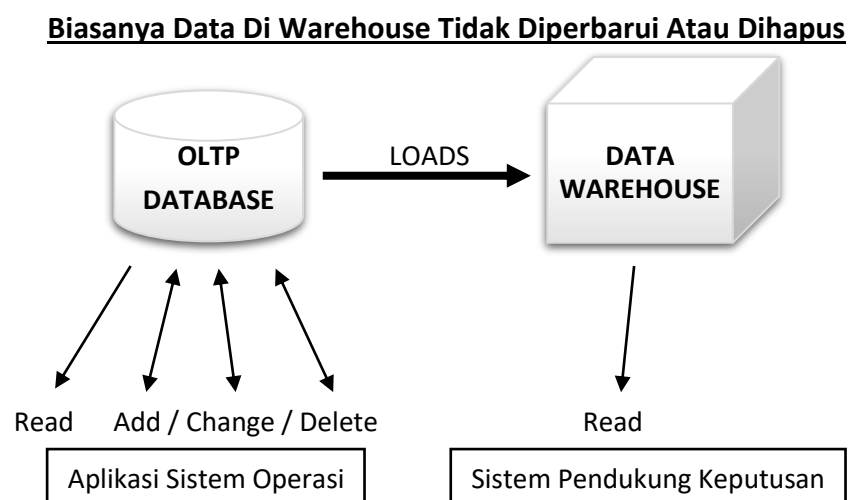
Misalnya, dalam gudang data yang berisi unit penjualan, kuantitas yang disimpan di setiap catatan file atau baris tabel berhubungan dengan elemen waktu tertentu. Bergantung pada tingkat rincian di gudang data, kuantitas penjualan dalam catatan mungkin berhubungan dengan tanggal, minggu, bulan, atau kuartal tertentu.

Sifat data yang bervariasi waktu dalam gudang data

- Memungkinkan analisis masa lalu
- Mengaitkan informasi dengan masa kini
- Memungkinkan perkiraan untuk masa depan

Data Nonvolatil

Data yang diambil dari berbagai sistem operasional dan data terkait yang diperoleh dari sumber luar diubah, diintegrasikan, dan disimpan di gudang data. Data di gudang data tidak dimaksudkan untuk menjalankan bisnis sehari-hari. Saat Anda ingin memproses pesanan berikutnya yang diterima dari pelanggan, Anda tidak perlu melihat ke dalam data gudang untuk mengetahui status stok saat ini. Aplikasi entri pesanan operasional dimaksudkan untuk tujuan itu. Di gudang data, Anda menyimpan data status stok yang diekstraksi sebagai snapshot dari waktu ke waktu. Anda tidak memperbarui gudang data setiap kali Anda memproses satu pesanan.



Gambar 2.3 Gudang data bersifat nonvolatile.

Data dari sistem operasional dipindahkan ke gudang data pada interval tertentu. Bergantung pada kebutuhan bisnis, pergerakan data ini terjadi dua kali sehari, sekali sehari, seminggu sekali, atau mungkin sekali dalam dua minggu. Faktanya, dalam gudang data pada umumnya, perpindahan data ke kumpulan data yang berbeda mungkin terjadi pada frekuensi

yang berbeda. Perubahan atribut produk dapat dipindahkan seminggu sekali. Setiap revisi pada pengaturan geografis dapat dipindahkan sebulan sekali.

Unit penjualan dapat dipindahkan satu kali sehari. Anda merencanakan dan menjadwalkan pergerakan data atau pemuatan data berdasarkan kebutuhan pengguna Anda. Seperti diilustrasikan pada Gambar 2.3, setiap transaksi bisnis tidak memperbarui data di gudang data. Transaksi bisnis memperbarui database sistem operasional secara real time. Kami menambah, mengubah, atau menghapus data dari sistem operasional seiring setiap transaksi terjadi namun biasanya tidak memperbarui data di gudang data. Anda tidak menghapus data di gudang data secara real time. Setelah data diambil dan dimasukkan ke dalam gudang data, Anda tidak menjalankan transaksi individual untuk mengubah data di sana. Pembaruan data merupakan hal yang lumrah dalam database operasional; tidak demikian halnya di gudang data. Data dalam gudang data tidak mudah berubah seperti data dalam database operasional. Data di gudang data terutama untuk kueri dan analisis.

Granularitas Data

Dalam sistem operasional, data biasanya disimpan pada tingkat detail yang paling rendah. Dalam sistem tempat penjualan untuk toko kelontong, unit penjualan ditangkap dan disimpan pada tingkat unit produk per transaksi di konter check-out. Dalam sistem entri pesanan, jumlah yang dipesan ditangkap dan disimpan pada tingkat unit produk per pesanan yang diterima dari pelanggan. Kapan pun Anda membutuhkan data ringkasan, Anda menjumlahkan masing-masing transaksi. Jika Anda mencari unit produk yang dipesan bulan ini, Anda membaca semua pesanan yang dimasukkan selama sebulan penuh untuk produk tersebut dan menjumlahkannya. Anda biasanya tidak menyimpan data ringkasan dalam sistem operasional.

Saat pengguna menanyakan gudang data untuk dianalisis, dia biasanya memulai dengan melihat ringkasan data. Pengguna dapat memulai dengan total unit penjualan suatu produk di seluruh wilayah. Kemudian pengguna mungkin ingin melihat pengelompokan berdasarkan negara bagian di wilayah tersebut. Langkah selanjutnya mungkin adalah pemeriksaan unit penjualan oleh masing-masing toko tingkat berikutnya. Seringkali, analisis dimulai dari tingkat tinggi dan turun ke tingkat detail yang lebih rendah.

Oleh karena itu, di gudang data, Anda merasa efisien untuk menyimpan data yang diringkas pada tingkat yang berbeda. Bergantung pada kueri, Anda kemudian dapat menuju ke tingkat detail tertentu dan memenuhi kueri. Perincian data dalam gudang data mengacu pada tingkat detail. Semakin rendah tingkat detailnya, semakin halus pula granularitas datanya. Tentu saja, jika Anda ingin menyimpan data pada tingkat detail paling rendah, Anda harus menyimpan banyak data di data warehouse. Anda harus memutuskan tingkat perincian berdasarkan tipe data dan kinerja sistem yang diharapkan untuk kueri. Gambar 2-4 menunjukkan contoh granularitas data di gudang data pada umumnya.

TIGA TINGKAT DATA DALAM GUDANG DATA PERBANKAN		
Detail Harian	Ringkasan Bulanan	Ringkasan Triwulan
Akun	Akun	Akun
Tanggal Kegiatan	Bulan	Quarter
Jumlah	Jumlah Transaksi	Jumlah Transaksi
Setoran/Penarikan	Penarikan	Penarikan
	Deposito	Deposito
	Saldo Awal	Saldo Awal
	Saldo Akhir	Saldo Akhir
Perincian data mengacu pada tingkat detail. Tergantung pada persyaratan, beberapa tingkat detail mungkin ada. Banyak gudang data memiliki setidaknya dua tingkat granularitas		

Gambar 2.4 Perincian data.

2.2 GUDANG DATA DAN DATA MARTS

Mereka yang telah mengikuti literatur tentang data warehouse pada tahun-tahun awal topik ini pasti pernah menemukan istilah “*data warehouse*” dan “*data mart*.” Banyak orang yang baru mengenal paradigma ini bingung dengan istilah-istilah ini. Beberapa penulis dan vendor menggunakan kedua istilah tersebut secara sinonim. Beberapa orang membuat perbedaan yang tidak cukup jelas. Pada titik ini, akan bermanfaat bagi kita untuk mengkaji kedua istilah ini dan mengambil posisi kita.

Menulis di majalah perdagangan terkemuka pada tahun 1998, hari-hari awal pergudangan data, Bill Inmon menyatakan, “Masalah paling penting yang dihadapi manajer TI tahun ini adalah apakah akan membangun gudang data terlebih dahulu atau data mart terlebih dahulu.” Pernyataan ini agaknya benar bahkan hingga saat ini. Mari kita periksa pernyataan ini dan membahas definisinya.

Sebelum memutuskan untuk membangun gudang data untuk organisasi Anda, Anda perlu menanyakan pertanyaan dasar dan mendasar berikut dan mengatasi masalah yang relevan:

- Pendekatan top-down atau bottom-up?
- Tingkat perusahaan atau departemen?
- Mana yang pertama—data warehouse atau data mart?
- Membangun proyek percontohan atau menerapkan implementasi penuh?
- Data mart yang bergantung atau independen?

Ini adalah permasalahan kritis yang memerlukan pemeriksaan dan perencanaan yang cermat. Haruskah Anda melihat gambaran besar organisasi Anda, mengambil pendekatan top-down, dan membangun gudang data yang sangat besar? Atau, haruskah Anda mengadopsi pendekatan bottom-up, melihat kebutuhan masing-masing lokal dan departemen, dan membangun data mart departemen yang kecil? Haruskah Anda membangun gudang data yang besar dan kemudian membiarkan repositori tersebut memasukkan data ke dalam data mart departemen lokal? Di sisi lain, haruskah Anda membangun data mart lokal individual,

dan menggabungkannya untuk membentuk gudang data Anda secara keseluruhan? Haruskah data mart lokal ini independen satu sama lain? Atau haruskah mereka bergantung pada gudang data secara keseluruhan untuk umpan data? Haruskah Anda membangun data mart percontohan? Ini adalah pertanyaan-pertanyaan penting.

Bagaimana mereka berbeda?

Mari kita lihat lebih dekat Gambar 2.5. Berikut adalah dua pendekatan dasar yang berbeda: (1) keseluruhan data warehouse yang memberi makan data mart yang bergantung, dan (2) beberapa data mart departemen atau lokal yang digabungkan menjadi sebuah data warehouse. Pada pendekatan pertama, Anda mengekstrak data dari sistem operasional; Anda kemudian mengubah, membersihkan, mengintegrasikan, dan menyimpan data di gudang data. Jadi, pendekatan mana yang terbaik bagi Anda, pendekatan top-down atau bottom-up? Mari kita periksa kedua pendekatan ini dengan cermat.

Pendekatan Top-Down versus Bottom-Up

Pendekatan Top-Down Bill Inmon adalah salah satu pendukung utama pendekatan top-down. Dia telah mendefinisikan gudang data sebagai gudang terpusat untuk seluruh perusahaan. Dalam pendekatan ini data di gudang data disimpan pada tingkat granularitas terendah berdasarkan model data yang dinormalisasi. Dalam visi Inmon, gudang data berada di pusat “Pabrik Informasi Perusahaan” (CIF) yang menyediakan kerangka logis untuk menyampaikan intelijen bisnis ke perusahaan. Operasi bisnis menyediakan data untuk mendorong CIF. Gudang data terpusat akan memberi makan data mart dependen yang dapat dirancang berdasarkan model data dimensional.

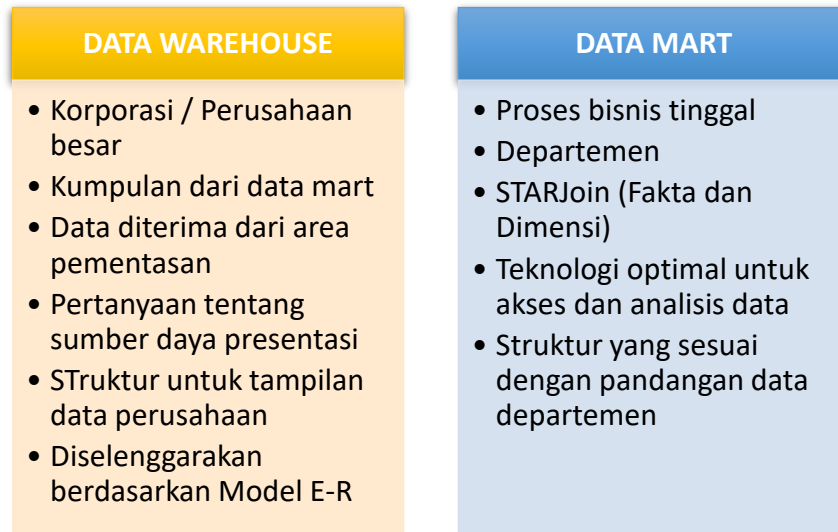
Keuntungan dari pendekatan ini adalah:

- ◆ Upaya yang benar-benar korporat, pandangan perusahaan terhadap data
- ◆ Didesain secara inheren, bukan gabungan dari data mart yang berbeda
- ◆ Penyimpanan data terpusat dan tunggal tentang konten
- ◆ Aturan dan kontrol terpusat
- ◆ Dapat memperoleh hasil yang cepat jika diimplementasikan dengan iterasi. Kerugiannya adalah:
 - ◆ Membutuhkan waktu lebih lama untuk membangun bahkan dengan metode berulang
 - ◆ Paparan risiko kegagalan yang tinggi
 - ◆ Membutuhkan keterampilan lintas fungsi tingkat tinggi
 - ◆ Pengeluaran yang tinggi tanpa bukti konsep

Ini adalah pendekatan gambaran besar di mana Anda membangun gudang data keseluruhan perusahaan yang besar. Di sini Anda tidak memiliki kumpulan pulau informasi yang terfragmentasi. Gudang datanya besar dan terintegrasi. Namun pendekatan ini membutuhkan waktu yang lebih lama untuk dibangun dan memiliki risiko kegagalan yang tinggi. Jika Anda tidak memiliki profesional berpengalaman di tim Anda, pendekatan ini bisa berbahaya. Selain itu, akan sulit untuk menjual pendekatan ini kepada manajemen senior dan sponsor. Mereka mungkin tidak akan melihat hasilnya dalam waktu dekat.

Pendekatan Bottom-Up Ralph Kimball, penulis terkemuka lainnya dan praktisi ahli di bidang data warehousing, adalah pendukung pendekatan yang kemudian dikenal sebagai

pendekatan bottom-up. Kimball (1996) membayangkan gudang data perusahaan sebagai kumpulan data mart yang disesuaikan. Pertimbangan utamanya adalah kesesuaian dimensi antara data mart yang terpisah. Dalam pendekatan ini data mart dibuat pertama kali untuk memberikan kemampuan analitis dan pelaporan untuk subjek bisnis tertentu berdasarkan model data dimensi.



Gambar 2.5 Gudang data versus data mart.

Data mart berisi data pada tingkat granularitas terendah dan juga sebagai ringkasan tergantung pada kebutuhan analisis. Data mart ini digabungkan atau “disatukan” dengan menyesuaikan dimensinya. Kita akan membahas proses penyesuaian dimensi ini dengan sangat rinci di bab berikutnya.

Keuntungan dari pendekatan ini adalah:

- Implementasi bagian-bagian yang dapat dikelola dengan lebih cepat dan mudah
- Pengembalian investasi yang menguntungkan dan bukti konsep
- Lebih sedikit risiko kegagalan
- Secara inheren bersifat inkremental; dapat menjadwalkan data mart penting terlebih dahulu
- Memungkinkan tim proyek untuk belajar dan berkembang. Kerugiannya adalah:
- Setiap data mart mempunyai pandangan sempitnya sendiri terhadap data
- Menyerap data yang berlebihan di setiap data mart
- Melestarikan data yang tidak konsisten dan tidak dapat diselaraskan
- Memperbanyak antarmuka yang tidak dapat dikelola

Dalam pendekatan bottom-up ini, Anda membangun data mart departemen Anda satu per satu. Anda akan menetapkan skema prioritas untuk menentukan data mart mana yang harus Anda bangun terlebih dahulu. Kelemahan paling parah dari pendekatan ini adalah fragmentasi data. Setiap data mart independen tidak akan memperhatikan keseluruhan kebutuhan seluruh organisasi.

Pendekatan Praktis

Untuk merumuskan pendekatan bagi organisasi Anda, Anda perlu memeriksa apa sebenarnya yang diinginkan organisasi Anda. Apakah organisasi Anda saat ini mencari hasil jangka panjang atau data mart yang cepat hanya untuk beberapa subjek? Apakah organisasi Anda menginginkan implementasi yang cepat, bukti konsep, dan mudah digunakan? Atau, apakah Anda ingin mencari pendekatan praktis lainnya? Meskipun pendekatan top-down dan bottom-up masing-masing memiliki kelebihan dan kekurangannya masing-masing, pendekatan kompromi yang mengakomodasi kedua pandangan tersebut tampaknya praktis. Dalam pendekatan ini kita tidak melupakan gambaran besar keseluruhan perusahaan. Kami mendasarkan perencanaan kami pada gambaran besar keseluruhan ini. Aspek ini berasal dari pendekatan top-down.

Kemudian kami mengadopsi prinsip pendekatan bottom-up dan membangun data mart yang sesuai berdasarkan skema prioritas. Langkah-langkah dalam pendekatan praktis ini adalah sebagai berikut:

1. Merencanakan dan menetapkan persyaratan di tingkat perusahaan secara keseluruhan
2. Ciptakan arsitektur sekeliling untuk gudang yang lengkap
3. Menyesuaikan dan membakukan isi data
4. Mengimplementasikan gudang data sebagai rangkaian supermarket, satu per satu

Dalam pendekatan praktis ini, Anda mempelajari dasar-dasarnya dan menentukan apa sebenarnya yang diinginkan organisasi Anda dalam jangka panjang. Kunci dari pendekatan ini adalah Anda terlebih dahulu membuat rencana di tingkat perusahaan. Anda mengumpulkan persyaratan di tingkat keseluruhan. Anda membangun arsitektur untuk gudang yang lengkap. Kemudian Anda menentukan isi data masing-masing supermarket. Supermart adalah pasar data yang dirancang dengan cermat. Anda menerapkan supermarket ini, satu per satu. Sebelum implementasi, Anda memastikan bahwa konten data di antara berbagai supermarket telah disesuaikan dalam hal tipe data, panjang bidang, presisi, dan semantik. Elemen data tertentu harus memiliki arti yang sama di setiap supermarket. Hal ini akan menghindari penyebaran data yang berbeda di beberapa data mart.

Data mart, dalam pendekatan praktis ini, adalah bagian logis dari gudang data yang lengkap, semacam bagian dari seluruh gudang data. Oleh karena itu, gudang data adalah gabungan semua data mart. Data mart individual ditargetkan pada kelompok bisnis tertentu di perusahaan, namun kumpulan semua data mart membentuk keseluruhan yang terintegrasi, yang disebut gudang data perusahaan. Ketika kami mengacu pada data warehouse dan data mart dalam diskusi kami di sini, kami menggunakan makna yang dipahami dalam pendekatan praktis ini. Bagi kami, gudang data berarti kumpulan data mart konstituen.

2.3 JENIS ARSITEKTUR

Kami sekarang telah meninjau definisi dasar dan fitur gudang data dan data mart dan menyelesaikan diskusi signifikan tentangnya. Kami telah menetapkan posisi kami mengenai

arti istilah data warehouse dan data mart bagi kami. Sekarang kita hampir siap untuk memeriksa masing-masing komponen.

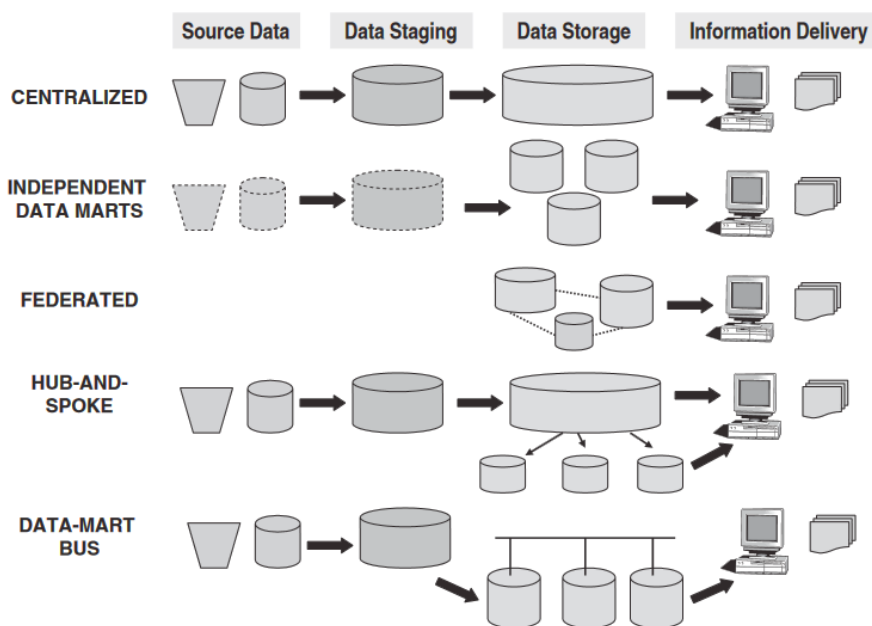
Selama bertahun-tahun, dalam praktiknya, beberapa pengaturan gudang data dan data mart telah berkembang berdasarkan kebutuhan individu organisasi. Sebelum kita melanjutkan mempelajari masing-masing komponen dalam data warehouse, mari kita bahas secara singkat tipe arsitektur terkemuka dan perhatikan bagaimana fitur data mart individual dalam tipe arsitektur ini. Untuk masing-masing tipe arsitektur ini kami akan menunjukkan bagaimana data disimpan di gudang data dan hubungan antara gudang data dan data mart. Gambar 2.6 mengilustrasikan masing-masing tipe arsitektur ini.

Gudang Data Terpusat

Tipe arsitektur ini memperhitungkan kebutuhan informasi tingkat perusahaan. Infrastruktur keseluruhan telah dibangun. Data yang dinormalisasi tingkat atom pada tingkat granularitas terendah disimpan dalam bentuk normal ketiga. Kadang-kadang, beberapa data ringkasan disertakan. Kueri dan aplikasi mengakses data yang dinormalisasi di gudang data pusat. Tidak ada data mart yang terpisah.

Data Mart Independen

Tipe arsitektur ini berkembang di perusahaan dimana unit organisasi mengembangkan data mart mereka sendiri untuk tujuan spesifik mereka sendiri. Meskipun masing-masing data mart melayani unit organisasi tertentu, data mart yang terpisah ini tidak memberikan “satu versi kebenaran.” Data mart tidak bergantung satu sama lain. Akibatnya, data mart yang berbeda ini cenderung memiliki definisi dan standar data yang tidak konsisten. Varians seperti itu menghambat analisis data di seluruh data mart. Misalnya, jika ada dua data mart independen, satu untuk penjualan dan yang lainnya untuk pengiriman, meskipun penjualan dan pengiriman adalah subjek terkait, data mart independen akan menyulitkan analisis data penjualan dan pengiriman secara bersamaan.



Gambar 2.6 Tipe arsitektur gudang data.

Federasi

Beberapa perusahaan masuk ke dalam data warehousing dengan warisan yang sudah ada berupa beragam struktur pendukung keputusan dalam bentuk sistem operasional, kumpulan data yang diekstraksi, data mart primitif, dan sebagainya. Bagi perusahaan seperti itu, mungkin tidak bijaksana untuk membuang seluruh investasi besar dan memulai dari awal. Solusi praktisnya adalah tipe arsitektur gabungan di mana data dapat diintegrasikan secara fisik atau logis melalui bidang kunci bersama, keseluruhan metadata global, kueri terdistribusi, dan metode lainnya. Dalam tipe arsitektur ini, tidak ada satu gudang data secara keseluruhan.

Hub-dan-Spoke

Ini adalah pendekatan Pabrik Informasi Perusahaan Inmon. Mirip dengan arsitektur gudang data terpusat, di sini juga terdapat gudang data seluruh perusahaan. Data atom dalam bentuk normal ketiga disimpan di gudang data terpusat. Perbedaan utama dan berguna adalah kehadiran data mart dependen dalam tipe arsitektur ini. Dependent data mart memperoleh data dari gudang data terpusat. Gudang data terpusat membentuk hub untuk memasukkan data ke data mart di jari-jari. Data mart dependen dapat dikembangkan untuk berbagai tujuan: kebutuhan analitis departemen, pertanyaan khusus, penambangan data, dan sebagainya. Setiap data mart dependen mungkin memiliki struktur data yang dinormalisasi, didenormalisasi, diringkas, atau dimensional berdasarkan kebutuhan individu. Sebagian besar kueri diarahkan ke data mart dependen meskipun gudang data terpusat sendiri dapat digunakan untuk kueri. Tipe arsitektur ini dihasilkan dari penerapan pendekatan top-down dalam pengembangan data warehouse.

Bus Data-Mart

Ini adalah pendekatan supermarket yang disesuaikan dengan Kimbal. Anda mulai dengan menganalisis persyaratan untuk subjek bisnis tertentu seperti pesanan, pengiriman, penagihan, klaim asuransi, persewaan mobil, dan sebagainya. Anda membangun data mart (supermart) pertama menggunakan dimensi dan metrik bisnis. Dimensi bisnis ini akan dibagikan di data mart masa depan. Gagasan utamanya adalah dengan menyesuaikan dimensi di antara berbagai data mart, hasilnya akan menjadi supermart yang terintegrasi secara logis yang akan memberikan pandangan perusahaan terhadap data tersebut. Data mart berisi data atom yang disusun sebagai model data dimensi. Tipe arsitektur ini dihasilkan dari penerapan pendekatan bottom-up yang ditingkatkan pada pengembangan gudang data.

2.4 GAMBARAN UMUM KOMPONEN

Setelah meninjau beberapa cara khas pengorganisasian data warehouse dan data mart, sekarang kita siap untuk meluncurkan diskusi umum tentang komponen data warehouse. Saat kami membangun sistem operasional seperti entri pesanan, pemrosesan klaim, atau rekening tabungan, kami menggabungkan beberapa komponen untuk membentuk sistem tersebut. Komponen front-end terdiri dari GUI (antarmuka pengguna grafis) untuk berinteraksi dengan pengguna untuk input data. Komponen penyimpanan data mencakup sistem manajemen basis data, seperti Oracle, Informix, atau Microsoft SQL Server. Komponen tampilan adalah kumpulan layar dan laporan untuk pengguna. Antarmuka data dan perangkat

lunak jaringan membentuk komponen konektivitas. Bergantung pada kebutuhan informasi dan kerangka organisasi kami, kami mengatur komponen-komponen ini dengan cara yang paling optimal.

Arsitektur adalah penataan komponen-komponen yang tepat. Anda membangun gudang data dengan komponen perangkat lunak dan perangkat keras. Untuk memenuhi kebutuhan organisasi Anda, Anda mengatur blok bangunan ini dengan cara tertentu untuk mendapatkan manfaat maksimal. Anda mungkin ingin memberikan penekanan khusus pada satu komponen; Anda mungkin ingin meningkatkan komponen lain dengan alat dan layanan tambahan. Semua ini tergantung pada keadaan Anda.

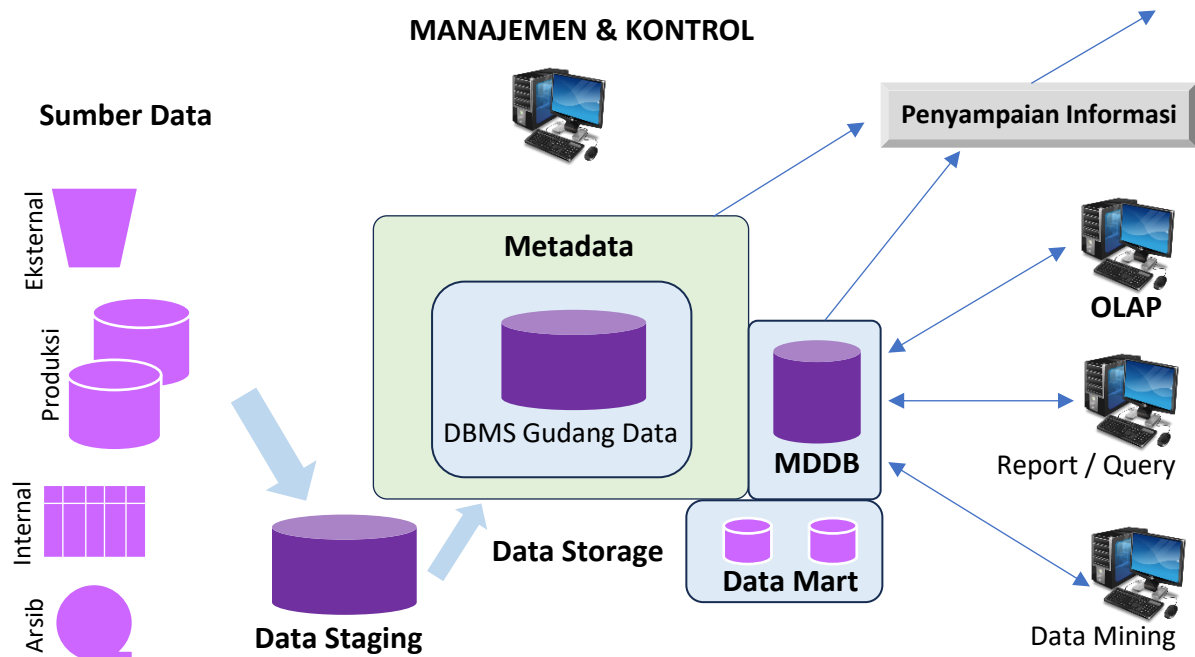
Gambar 2.7 menunjukkan komponen dasar gudang pada umumnya. Anda melihat komponen Sumber Data ditampilkan di sebelah kiri. Komponen Data Staging berfungsi sebagai blok penyusun berikutnya. Di tengah, Anda melihat komponen Penyimpanan Data yang mengelola data gudang data. Komponen ini tidak hanya menyimpan dan mengelola data, namun juga melacak data melalui repositori metadata. Komponen Pengiriman Informasi yang ditunjukkan di sebelah kanan terdiri dari semua cara berbeda untuk membuat informasi dari gudang data tersedia bagi pengguna.

Baik Anda membangun gudang data untuk perusahaan manufaktur besar yang masuk dalam daftar Fortune 500, jaringan toko kelontong terkemuka dengan toko di seluruh negeri, atau lembaga perbankan global, komponen dasarnya sama. Setiap gudang data disatukan dengan blok penyusun yang sama. Perbedaan penting bagi masing-masing organisasi terletak pada cara penyusunan elemen-elemen tersebut. Variasinya terletak pada cara beberapa balok dibuat lebih kuat dibandingkan balok lainnya dalam arsitektur.

Sekarang kita akan melihat lebih dekat masing-masing komponen. Pada tahap ini, kita ingin mengetahui apa saja komponen-komponennya dan bagaimana masing-masing komponen tersebut cocok dengan arsitekturnya. Kami juga ingin meninjau isu-isu spesifik yang berkaitan dengan masing-masing komponen tertentu.

Komponen Data Sumber

Sumber data yang masuk ke gudang data dapat dikelompokkan menjadi empat kategori besar, seperti yang dibahas di sini.



Gambar 2.7 Gudang data: blok atau komponen penyusun.

Data Produksi Kategori data ini berasal dari berbagai sistem operasional perusahaan. Hal ini biasanya mencakup sistem keuangan, sistem manufaktur, sistem sepanjang rantai pasokan, dan sistem manajemen hubungan pelanggan. Berdasarkan kebutuhan informasi di gudang data, Anda memilih segmen data dari sistem operasional yang berbeda. Saat menangani data ini, Anda menemukan banyak variasi dalam format data. Anda juga memperhatikan bahwa data berada pada platform perangkat keras yang berbeda. Selanjutnya, data tersebut didukung oleh sistem database dan sistem operasi yang berbeda. Ini adalah data dari banyak aplikasi vertikal.

Dalam sistem operasional, permintaan informasi bersifat sempit. Anda menanyakan sistem operasional untuk informasi tentang contoh spesifik objek bisnis. Anda mungkin hanya menginginkan nama dan alamat satu pelanggan. Atau, Anda mungkin memerlukan pesanan yang dilakukan oleh satu pelanggan dalam satu minggu. Atau, Anda mungkin hanya perlu melihat satu faktur dan item yang ditagihkan pada satu faktur tersebut. Dalam sistem operasional, Anda tidak memiliki pertanyaan luas. Anda tidak menanyakan sistem operasional dengan cara yang tidak terduga. Semua pertanyaan dapat diprediksi. Sekali lagi, Anda tidak mengharapkan kueri tertentu dijalankan di sistem operasional yang berbeda. Apa maksudnya semua ini? Sederhananya: tidak ada kesesuaian data di antara berbagai sistem operasional suatu perusahaan. Istilah seperti akun mungkin memiliki arti berbeda dalam sistem yang berbeda.

Karakteristik data produksi yang signifikan dan meresahkan adalah disparitas. Tantangan terbesar Anda adalah menstandarisasi dan mengubah data yang berbeda dari berbagai sistem produksi, mengubah data, dan mengintegrasikan potongan-potongan tersebut menjadi data yang berguna untuk disimpan di gudang data. Integrasi berbagai sumber inilah yang memberikan nilai pada data di gudang data. Data Internal Di setiap

organisasi, pengguna menyimpan spreadsheet, dokumen, profil pelanggan, dan terkadang bahkan database departemen “pribadi”. Ini adalah data internal, yang sebagiannya dapat berguna dalam gudang data.

Jika organisasi Anda melakukan bisnis dengan pelanggan secara tatap muka dan kontribusi masing-masing pelanggan terhadap laba adalah signifikan, maka profil pelanggan terperinci dengan demografi yang luas adalah penting dalam gudang data. Profil pelanggan individu menjadi sangat penting untuk dipertimbangkan. Ketika perwakilan akun Anda berbicara dengan pelanggan yang ditugaskan kepada mereka atau ketika departemen pemasaran Anda ingin memberikan penawaran khusus kepada pelanggan individu, Anda memerlukan rinciannya. Meskipun sebagian besar data ini dapat diambil dari sistem produksi, sebagian besar data tersebut disimpan oleh individu dan departemen dalam file pribadi mereka.

Anda tidak dapat mengabaikan data internal yang disimpan dalam file pribadi di organisasi Anda. Ini merupakan keputusan kolektif mengenai berapa banyak data internal yang harus dimasukkan ke dalam gudang data. Departemen TI harus bekerja dengan departemen pengguna untuk mengumpulkan data internal. Data internal menambah kompleksitas tambahan pada proses transformasi dan integrasi data sebelum dapat disimpan di gudang data. Anda harus menentukan strategi untuk mengumpulkan data dari spreadsheet, menemukan cara mengambil data dari dokumen tekstual, dan mengaitkannya dengan database departemen untuk mengumpulkan data terkait dari sumber tersebut. Sekali lagi, Anda mungkin ingin menjadwalkan perolehan data internal. Awalnya, Anda mungkin ingin membatasi diri Anda sendiri hanya pada beberapa porsi yang signifikan sebelum meluncurkan data mart pertama Anda.

Sistem Operasional Data Arsip terutama ditujukan untuk menjalankan bisnis saat ini. Di setiap sistem operasional, Anda secara berkala mengambil data lama dan menyimpannya dalam file arsip. Keadaan di organisasi Anda menentukan seberapa sering dan bagian mana dari database operasional yang diarsipkan untuk disimpan. Beberapa data diarsipkan setelah satu tahun. Terkadang data tertinggal di database sistem operasional selama lima tahun. Sebagian besar data yang diarsipkan berasal dari sistem lama yang sudah mendekati akhir masa manfaatnya dalam organisasi.

Ada banyak metode pengarsipan yang berbeda. Ada metode pengarsipan bertahap. Pada tahap pertama, data terkini diarsipkan ke database arsip terpisah yang mungkin masih online. Pada tahap kedua, data lama diarsipkan ke file datar di penyimpanan disk. Pada tahap berikutnya, data terlama diarsipkan ke tape cartridge atau mikrofilm dan bahkan disimpan di luar lokasi. Seperti disebutkan sebelumnya, gudang data menyimpan cuplikan data historis. Anda pada dasarnya memerlukan data historis untuk analisis dari waktu ke waktu. Untuk mendapatkan informasi historis, Anda melihat kumpulan data yang diarsipkan. Bergantung pada kebutuhan gudang data Anda, Anda harus menyertakan data historis yang memadai. Jenis data ini berguna untuk membedakan pola dan menganalisis tren.

Data Eksternal Kebanyakan eksekutif bergantung pada data dari sumber eksternal untuk sebagian besar informasi yang mereka gunakan. Mereka menggunakan statistik yang

berkaitan dengan industri mereka yang dihasilkan oleh lembaga eksternal dan kantor statistik nasional. Mereka menggunakan data pangsa pasar pesaing. Mereka menggunakan nilai standar indikator keuangan untuk bisnis mereka untuk memeriksa kinerjanya. Misalnya, gudang data perusahaan penyewaan mobil berisi data jadwal produksi terkini dari produsen mobil terkemuka. Data eksternal di gudang data ini membantu perusahaan persewaan mobil merencanakan manajemen armadanya.

Tujuan yang dilayani oleh sumber data eksternal tersebut tidak dapat dipenuhi oleh data yang tersedia dalam organisasi Anda sendiri. Wawasan yang diperoleh dari data produksi dan data arsip Anda agak terbatas. Mereka memberi Anda gambaran berdasarkan apa yang sedang atau telah Anda lakukan di masa lalu. Untuk mengetahui tren industri dan membandingkan kinerja dengan organisasi lain, Anda memerlukan data dari sumber eksternal. Biasanya, data dari sumber luar tidak sesuai dengan format Anda. Anda harus memikirkan cara untuk mengonversi data ke dalam format dan tipe data internal Anda. Anda harus mengatur transmisi data dari sumber eksternal. Beberapa sumber mungkin memberikan informasi secara teratur dan dalam jangka waktu tertentu. Orang lain mungkin memberi Anda data berdasarkan permintaan. Anda perlu mengakomodasi variasinya.

Komponen Pementasan Data

Setelah Anda mengekstrak data dari berbagai sistem operasional dan dari sumber eksternal, Anda harus menyiapkan data untuk disimpan di data warehouse. Data yang diekstraksi yang berasal dari beberapa sumber berbeda perlu diubah, dikonversi, dan disiapkan dalam format yang sesuai untuk disimpan untuk kueri dan analisis. Tiga fungsi utama perlu dilakukan untuk menyiapkan data. Anda harus mengekstrak data, mengubah data, dan kemudian memuat data ke dalam penyimpanan gudang data. Ketiga fungsi utama ekstraksi, transformasi, dan persiapan pemuatan berlangsung di area pementasan. Komponen pementasan data terdiri dari meja kerja untuk fungsi-fungsi ini. Pementasan data menyediakan tempat dan area dengan serangkaian fungsi untuk membersihkan, mengubah, menggabungkan, mengubah, menghapus duplikat, dan menyiapkan data sumber untuk disimpan dan digunakan di gudang data.

Mengapa diperlukan tempat atau komponen tersendiri untuk melakukan persiapan data? Tidak bisakah Anda memindahkan data dari berbagai sumber ke dalam penyimpanan data warehouse itu sendiri dan kemudian menyiapkan datanya? Saat kami menerapkan sistem operasional, kami cenderung mengambil data dari sumber berbeda, memindahkan data ke database sistem operasional baru, dan menjalankan konversi data. Mengapa metode ini tidak dapat berfungsi pada gudang data? Perbedaan penting di sini adalah: di gudang data Anda mengambil data dari banyak sistem operasional sumber. Ingatlah bahwa data dalam gudang data berorientasi pada subjek dan mencakup seluruh aplikasi operasional. Oleh karena itu, area pementasan yang terpisah merupakan kebutuhan untuk menyiapkan data untuk gudang data. Kadang-kadang, terutama dengan sistem lama, Anda mungkin memerlukan area pementasan literal yang terpisah. Apa yang Anda miliki di area pementasan literal ini adalah salinan literal dari konten sistem sumber tetapi dalam lingkungan yang lebih nyaman (seperti

database relasional yang dapat diakses ODBC). Area pementasan literal ini kemudian bertindak sebagai pengganti sistem sumber tertentu.

Sekarang kita telah memperjelas kebutuhan akan komponen pementasan data yang terpisah, mari kita pahami apa yang terjadi dalam pementasan data. Sekarang kita akan membahas secara singkat tiga fungsi utama yang terjadi di area pementasan. Ekstraksi Data Fungsi ini harus menangani berbagai sumber data. Anda harus menggunakan teknik yang sesuai untuk setiap sumber data. Sumber data mungkin berasal dari mesin sumber berbeda dalam format data beragam. Bagian dari data awal mungkin berada dalam sistem database relasional. Beberapa data mungkin berada di jaringan lama dan model data hierarki lainnya. Banyak sumber data yang mungkin masih dalam bentuk file datar. Anda mungkin ingin memasukkan data dari spreadsheet dan kumpulan data departemen setempat. Ekstraksi data mungkin menjadi sangat rumit.

Alat tersedia di pasar untuk ekstraksi data. Anda mungkin ingin mempertimbangkan untuk menggunakan alat luar yang sesuai untuk sumber data tertentu. Untuk sumber data lainnya, Anda mungkin ingin mengembangkan program internal untuk melakukan ekstraksi data. Membeli alat dari luar mungkin memerlukan biaya awal yang tinggi. Sebaliknya, program internal dapat berarti biaya berkelanjutan untuk pengembangan dan pemeliharaan.

Setelah Anda mengekstrak data, di mana Anda menyimpan data tersebut untuk persiapan lebih lanjut? Anda dapat melakukan fungsi ekstraksi di platform lama itu sendiri jika pendekatan tersebut sesuai dengan kerangka kerja Anda. Lebih sering lagi, tim implementasi gudang data mengekstrak sumber ke dalam lingkungan fisik terpisah sehingga memindahkan data ke dalam gudang data akan lebih mudah. Di lingkungan terpisah, Anda dapat mengekstrak data sumber ke dalam grup file datar, atau database relasional pementasan data, atau kombinasi keduanya.

Transformasi Data Dalam setiap implementasi sistem, konversi data merupakan fungsi yang penting. Misalnya, ketika Anda mengimplementasikan sistem operasional seperti aplikasi berlangganan majalah, Anda harus terlebih dahulu mengisi database Anda dengan data dari catatan sistem sebelumnya. Anda mungkin melakukan konversi dari sistem manual. Atau, Anda mungkin berpindah dari sistem berorientasi file ke sistem modern yang didukung dengan tabel database relasional. Apa pun kasusnya, Anda akan mengonversi data dari sistem sebelumnya. Jadi, apa bedanya dengan gudang data? Bagaimana transformasi data untuk gudang data lebih terlibat dibandingkan sistem operasional?

Sekali lagi, seperti yang Anda ketahui, data untuk data warehouse berasal dari banyak sumber berbeda. Jika ekstraksi data untuk gudang data menimbulkan tantangan besar, transformasi data menghadirkan tantangan yang lebih besar lagi. Faktor lain dalam gudang data adalah bahwa data feed bukan sekedar muatan awal. Anda harus terus mengambil perubahan yang sedang berlangsung dari sistem sumber. Tugas transformasi apa pun yang Anda siapkan untuk pemuatan awal juga akan disesuaikan untuk revisi yang sedang berlangsung.

Anda melakukan sejumlah tugas individual sebagai bagian dari transformasi data. Pertama, Anda membersihkan data yang diekstraksi dari setiap sumber. Pembersihan

mungkin hanya berupa koreksi kesalahan ejaan, atau mungkin mencakup penyelesaian konflik antara kode negara bagian dan kode pos dalam data sumber, atau mungkin berhubungan dengan pemberian nilai default untuk elemen data yang hilang, atau penghapusan duplikat saat Anda memasukkan data yang sama dari beberapa sistem sumber.

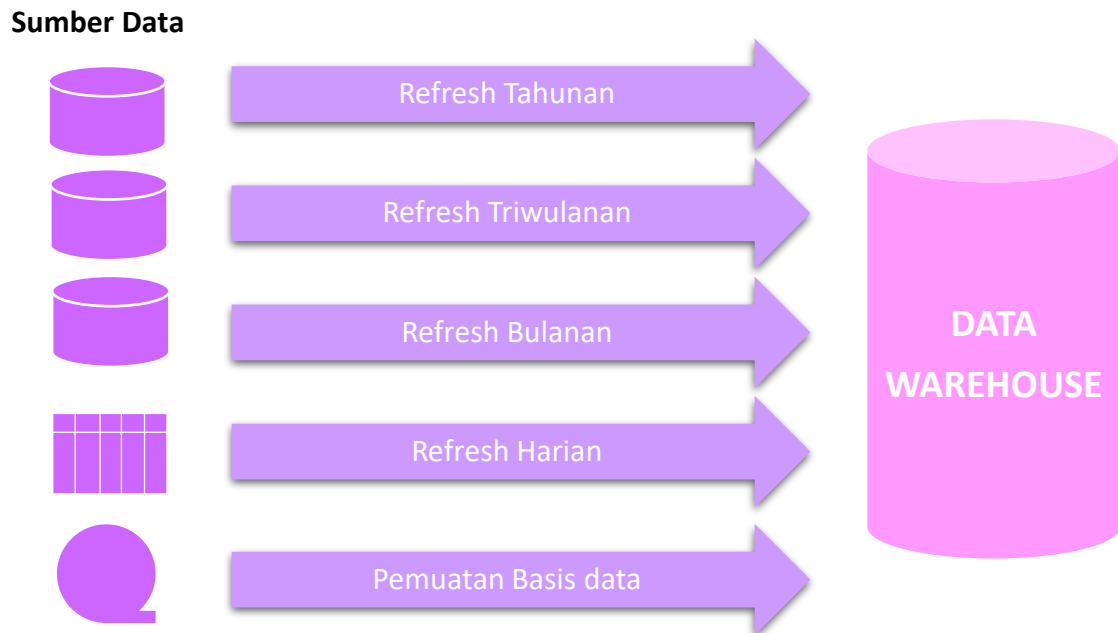
Standarisasi elemen data merupakan bagian besar dari transformasi data. Anda membakukan tipe data dan panjang bidang untuk elemen data yang sama yang diambil dari berbagai sumber. Standardisasi semantik adalah tugas besar lainnya. Anda menyelesaikan sinonim dan homonim. Jika dua istilah atau lebih dari sistem sumber berbeda memiliki arti yang sama, Anda menyelesaikan sinonimnya. Ketika satu istilah memiliki arti yang berbeda dalam sistem sumber yang berbeda, Anda menyelesaikan homonimnya.

Transformasi data melibatkan banyak bentuk penggabungan potongan data dari berbagai sumber. Anda menggabungkan data dari satu rekaman sumber atau elemen data terkait dari banyak rekaman sumber. Di sisi lain, transformasi data juga melibatkan pembersihan data sumber yang tidak berguna dan memisahkan rekaman sumber menjadi kombinasi baru. Penyortiran dan penggabungan data terjadi dalam skala besar di area pementasan data.

Dalam banyak kasus, kunci yang dipilih untuk sistem operasional adalah nilai lapangan dengan makna yang tertanam di dalamnya. Misalnya, nilai kunci produk dapat berupa kombinasi karakter yang menunjukkan kategori produk, kode gudang tempat produk disimpan, dan beberapa kode untuk menunjukkan batch produksi. Kunci primer dalam gudang data tidak dapat mempunyai arti bawaan. Kita akan membahas hal ini lebih lanjut di Bab 10. Transformasi data juga mencakup penetapan kunci pengganti yang berasal dari kunci primer sistem sumber.

Sistem operasional titik penjualan rantai bahan makanan menyimpan penjualan unit dan jumlah pendapatan berdasarkan transaksi individual di konter check-out di setiap toko. Namun di gudang data, mungkin tidak perlu menyimpan data pada tingkat sedetail ini. Anda mungkin ingin meringkas total berdasarkan produk di setiap toko pada hari tertentu dan menyimpan ringkasan total unit penjualan dan pendapatan di penyimpanan gudang data. Dalam kasus seperti ini, fungsi transformasi data akan mencakup peringkasan yang sesuai.

- Fungsi ini memakan waktu
- Pemuatan awal memindahkan data dalam jumlah yang sangat besar
- Kondisi bisnis menentukan siklus penyegaran



Gambar 2.8 Perpindahan data ke data warehouse.

Ketika fungsi transformasi data berakhir, Anda memiliki kumpulan data terintegrasi yang dibersihkan, distandarisasi, dan diringkas. Anda sekarang memiliki data yang siap dimuat ke setiap kumpulan data di gudang data Anda.

Pemuatan Data Dua kelompok tugas berbeda membentuk fungsi pemuatan data. Saat Anda menyelesaikan desain dan konstruksi gudang data dan ditayangkan untuk pertama kalinya, Anda melakukan pemuatan awal data ke dalam penyimpanan gudang data. Pemuatan awal memindahkan data dalam jumlah besar dan menghabiskan banyak waktu. Saat gudang data mulai berfungsi, Anda terus mengekstrak perubahan pada data sumber, mengubah revisi data, dan memasukkan revisi data tambahan secara berkelanjutan. Gambar 2.8 mengilustrasikan tipe umum pergerakan data dari staging area ke penyimpanan data warehouse.

Komponen Penyimpanan Data

Penyimpanan data untuk data warehouse adalah repositori terpisah. Sistem operasional perusahaan Anda mendukung operasi sehari-hari. Ini adalah aplikasi pemrosesan transaksi online. Penyimpanan data untuk sistem operasional biasanya hanya berisi data terkini. Selain itu, repositori data ini berisi data yang terstruktur dalam format yang sangat dinormalisasi untuk pemrosesan yang cepat dan efisien. Sebaliknya, dalam repositori data untuk gudang data, Anda perlu menyimpan data historis dalam jumlah besar untuk dianalisis. Selanjutnya, Anda harus menyimpan data di gudang data dalam struktur yang sesuai untuk analisis, dan bukan untuk pengambilan informasi individual dengan cepat. Oleh karena itu,

penyimpanan data untuk data warehouse disimpan terpisah dari penyimpanan data untuk sistem operasional.

Dalam database Anda yang mendukung sistem operasional, pembaruan data terjadi saat transaksi terjadi. Transaksi-transaksi ini masuk ke database secara acak. Bagaimana dan kapan transaksi mengubah data dalam database tidak sepenuhnya berada dalam kendali Anda. Data dalam database operasional dapat berubah dari waktu ke waktu. Saat analis Anda menggunakan data di gudang data untuk analisis, mereka perlu mengetahui bahwa data tersebut stabil dan mewakili snapshot pada periode tertentu. Saat mereka bekerja dengan data, penyimpanan data tidak boleh terus diperbarui. Oleh karena itu, gudang data merupakan tempat penyimpanan data “*read-only*”.

Umumnya database di gudang data Anda harus terbuka. Tergantung kebutuhan Anda, kemungkinan besar Anda akan menggunakan alat dari beberapa vendor. Gudang data harus terbuka untuk alat yang berbeda. Sebagian besar gudang data menggunakan sistem manajemen basis data relasional. Banyak gudang data juga menggunakan sistem manajemen basis data multidimensi. Data yang diambil dari penyimpanan gudang data dikumpulkan dalam banyak cara dan ringkasan data disimpan dalam database multidimensi (MDDDB). Sistem database multidimensi seperti itu biasanya merupakan produk berpemilik.

Komponen Penyampaian Informasi

Siapa saja pengguna yang membutuhkan informasi dari data warehouse? Kisarannya cukup komprehensif. Pengguna pemula datang ke gudang data tanpa pelatihan dan, oleh karena itu, memerlukan laporan prefabrikasi dan kueri yang telah ditetapkan sebelumnya. Pengguna biasa membutuhkan informasi sesekali, tidak secara teratur. Pengguna jenis ini juga memerlukan informasi yang dikemas sebelumnya. Analis bisnis mencari kemampuan untuk melakukan analisis kompleks menggunakan informasi di gudang data. Pengguna yang mahir ingin dapat menavigasi seluruh gudang data, mengambil data yang menarik, memformat kuerinya sendiri, menelusuri lapisan data, dan membuat laporan khusus dan kueri ad hoc.

Untuk memberikan informasi kepada komunitas luas pengguna gudang data, komponen penyampaian informasi mencakup metode penyampaian informasi yang berbeda. Gambar 2.9 menunjukkan berbagai metode penyampaian informasi. Laporan ad hoc adalah laporan standar yang ditujukan terutama untuk pengguna pemula dan biasa. Penyediaan pertanyaan kompleks, analisis multidimensi (MD), dan analisis statistik memenuhi kebutuhan analis bisnis dan pengguna listrik. Informasi yang dimasukkan ke dalam sistem informasi eksekutif (EIS) ditujukan untuk para eksekutif senior dan manajer tingkat tinggi. Beberapa gudang data juga menyediakan data untuk aplikasi data mining. Aplikasi penambangan data adalah sistem penemuan pengetahuan di mana algoritma penambangan membantu Anda menemukan tren dan pola dari penggunaan data Anda.

Di gudang data Anda, Anda dapat menyertakan beberapa mekanisme pengiriman informasi. Paling umum, Anda menyediakan pertanyaan dan laporan online. Pengguna akan memasukkan permintaan mereka secara online dan akan menerima hasilnya secara online. Anda dapat mengatur pengiriman laporan terjadwal melalui email atau Anda dapat

memanfaatkan intranet organisasi Anda secara memadai untuk pengiriman informasi. Baru-baru ini, penyampaian informasi melalui Internet semakin berkembang.

Komponen Metadata

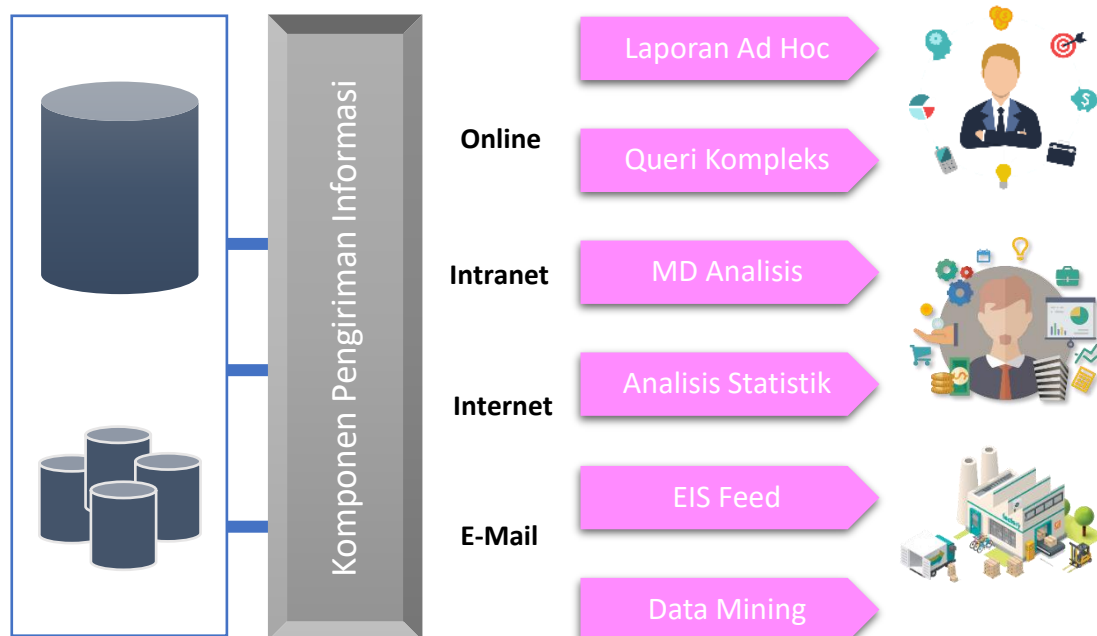
Metadata dalam gudang data mirip dengan kamus data atau katalog data dalam sistem manajemen basis data. Dalam kamus data, Anda menyimpan informasi tentang struktur data logis, informasi tentang file dan alamat, informasi tentang indeks, dan sebagainya. Kamus data berisi data tentang data dalam database.

Demikian pula komponen metadata adalah data tentang data di gudang data. Definisi ini merupakan definisi yang umum digunakan. Kita perlu menguraikan definisi ini. Metadata dalam gudang data mirip dengan kamus data, namun lebih dari sekadar kamus data. Nanti, di bagian terpisah dalam bab ini, kami akan mencurahkan lebih banyak waktu untuk membahas metadata. Di sini, demi kelengkapan, kami hanya ingin mencantumkan metadata sebagai salah satu komponen arsitektur data warehouse.

Komponen Manajemen dan Pengendalian

Komponen arsitektur gudang data ini berada di atas semua komponen lainnya. Komponen manajemen dan pengendalian mengoordinasikan layanan dan aktivitas dalam gudang data. Komponen ini mengontrol transformasi data dan transfer data ke penyimpanan data warehouse. Di sisi lain, ini memoderasi penyampaian informasi kepada pengguna. Ia bekerja dengan sistem manajemen basis data dan memungkinkan data disimpan dengan benar di repositori. Ini memantau pergerakan data ke staging area dan dari sana ke penyimpanan data warehouse itu sendiri.

Komponen manajemen dan pengendalian berinteraksi dengan komponen metadata untuk menjalankan fungsi manajemen dan pengendalian. Karena komponen metadata berisi informasi tentang gudang data itu sendiri, metadata adalah sumber informasi untuk modul manajemen.



Gambar 2.9 Komponen penyampaian informasi.

2.5 METADATA DI GUDANG DATA

Bayangkan metadata sebagai Yellow Pagesw kota Anda. Apakah Anda memerlukan informasi tentang toko-toko di kota Anda, lokasinya, apa namanya, dan produk spesialisasinya? Pergi ke Halaman Kuning. Yellow Pages adalah direktori berisi data tentang institusi di kota Anda. Hampir sama, komponen metadata berfungsi sebagai direktori isi gudang data Anda.

Karena pentingnya metadata dalam gudang data, kami telah memisahkan seluruh Bab 9 untuk topik ini. Pada tahap ini, kami hanya ingin memperkenalkan topik dan menyoroti bahwa metadata adalah komponen arsitektur utama dari gudang data.

Jenis Metadata

Metadata di gudang data terbagi dalam tiga kategori utama:

- ◆ Metadata operasional
- ◆ Ekstraksi dan transformasi metadata
- ◆ Metadata pengguna akhir

Metadata Operasional Seperti yang Anda ketahui, data untuk data warehouse berasal dari beberapa sistem operasional perusahaan. Sistem sumber ini berisi struktur data yang berbeda. Elemen data yang dipilih untuk gudang data memiliki panjang bidang dan tipe data yang berbeda-beda. Dalam memilih data dari sistem sumber untuk gudang data, Anda membagi catatan, menggabungkan bagian-bagian catatan dari file sumber yang berbeda, dan menangani beberapa skema pengkodean dan panjang bidang. Saat Anda menyampaikan informasi kepada pengguna akhir, Anda harus dapat mengaitkannya kembali dengan kumpulan data sumber asli. Metadata operasional berisi semua informasi tentang sumber data operasional.

Metadata Ekstraksi dan Transformasi Metadata ekstraksi dan transformasi berisi data tentang ekstraksi data dari sistem sumber, yaitu frekuensi ekstraksi, metode ekstraksi, dan aturan bisnis untuk ekstraksi data. Selain itu, kategori metadata ini berisi informasi tentang semua transformasi data yang terjadi di area pementasan data.

Metadata Pengguna Akhir Metadata pengguna akhir adalah peta navigasi gudang data. Ini memungkinkan pengguna akhir untuk menemukan informasi dari gudang data. Metadata pengguna akhir memungkinkan pengguna akhir untuk menggunakan terminologi bisnis mereka sendiri dan mencari informasi dengan cara yang biasa mereka pikirkan tentang bisnis.

Signifikansi Khusus

Mengapa metadata sangat penting dalam gudang data?

- Pertama, ia bertindak sebagai perekat yang menghubungkan seluruh bagian data warehouse.
- Selanjutnya, memberikan informasi tentang isi dan struktur kepada pengembang.
- Terakhir, hal ini membuka pintu bagi pengguna akhir dan membuat konten dapat dikenali sesuai dengan keinginan mereka.

RINGKASAN BAB

- Mendefinisikan fitur data warehouse adalah: terpisah, berorientasi subjek, terintegrasi, varian waktu, dan nonvolatile.
- Anda dapat menggunakan pendekatan top-down dan membangun gudang data perusahaan yang besar dan komprehensif, atau Anda dapat menggunakan pendekatan bottom-up dan membangun data mart departemen yang kecil dan independen. Meskipun mempunyai beberapa kelebihan, kedua pendekatan ini mempunyai kelemahan yang serius.
- Pendekatan praktis yang dapat dilakukan adalah dengan membangun data mart yang sesuai, yang bersama-sama membentuk gudang data perusahaan.
- Blok atau komponen penyusun gudang data adalah sumber data, pementasan data, penyimpanan data, pengiriman informasi, metadata, serta manajemen dan pengendalian.
- Dalam gudang data, metadata sangat penting karena berfungsi sebagai perekat yang menyatukan semua komponen dan berfungsi sebagai peta jalan bagi pengguna akhir.

LATIHAN SOAL

1. Sebutkan setidaknya enam karakteristik atau fitur gudang data.
2. Mengapa integrasi data lebih diperlukan dalam gudang data dibandingkan aplikasi operasional?
3. Setiap struktur data di data warehouse mengandung elemen waktu. Mengapa?
4. Menjelaskan granularitas data dan penerapannya pada gudang data.
5. Apa perbedaan pendekatan top-down dan bottom-up dalam membangun data warehouse? Buat daftar jenis arsitektur utama dan soroti fitur dari kedua arsitektur tersebut.
6. Apa sajakah sumber data untuk data warehouse?
7. Mengapa Anda memerlukan komponen pementasan data terpisah?
8. Di bawah transformasi data, sebutkan lima fungsi berbeda yang dapat Anda pikirkan.
9. Sebutkan enam metode penyampaian informasi yang berbeda.
10. Apa tiga tipe utama metadata dalam gudang data? Sebutkan secara singkat tujuan masing-masing jenis.

BAB 3

TREN DALAM PERGUDANGAN DATA

TUJUAN BAB

- Meninjau pertumbuhan berkelanjutan dalam data warehousing
- Pelajari bagaimana data warehousing telah menjadi arus utama
- Diskusikan beberapa tren utama, satu per satu
- Memahami perlunya standar dan meninjau kemajuannya
- Memahami gudang data yang mendukung Web

Pada bab sebelumnya, kita telah melihat mengapa intelijen bisnis dan gudang data yang mendasarinya sangat penting bagi perusahaan dari semua ukuran di semua industri. Kami telah meninjau bagaimana bisnis memperoleh manfaat besar dari data warehousing. Kami juga telah membahas blok bangunan gudang data. Anda sekarang memiliki gagasan yang cukup bagus tentang fitur dan fungsi komponen dasar serta definisi data warehousing yang masuk akal. Anda telah memahami bahwa ini adalah konsep yang pada dasarnya sederhana; pada saat yang sama, Anda tahu bahwa ini juga merupakan perpaduan dari banyak teknologi. Beberapa penggerak bisnis dan teknologi telah memajukan data warehousing dalam beberapa tahun terakhir.

Sebelum melangkah lebih jauh, kita sudah sampai pada titik di mana kita ingin menanyakan beberapa pertanyaan yang relevan. Bagaimana skenario dan keadaan pasar saat ini? Bisnis apa saja yang sudah mengadopsi data warehousing? Apa saja kemajuan teknologi? Singkatnya, apa saja tren signifikannya?. Apakah Anda bertanya-tanya apakah masih terlalu dini dalam pembahasan kita untuk membicarakan tren? Praktek yang biasa dilakukan adalah memasukkan bab tentang tren masa depan menjelang akhir, hampir sebagai sebuah renungan. Pembaca biasanya mengabaikan diskusi tentang tren masa depan. Bab ini tidak seperti melihat ke dalam bola kristal untuk kemungkinan kejadian di masa depan; kami ingin menghadapi tren penting yang sedang terjadi saat ini.

Penting bagi Anda untuk mengingat pengetahuan tentang tren saat ini sebagai latar belakang saat Anda melanjutkan studi mendalam tentang subjek tersebut. Saat Anda mengumpulkan persyaratan informasi untuk gudang data Anda, Anda perlu menyadari tren saat ini. Saat Anda memasuki fase desain, Anda harus menyadari tren. Saat Anda mengimplementasikan gudang data, Anda perlu memastikan bahwa gudang data Anda sejalan dengan tren. Pengetahuan tentang tren penting dan diperlukan bahkan pada tahap awal studi Anda.

Dalam bab ini, kita akan membahas sebagian besar tren utama. Anda akan memahami bagaimana dan mengapa data warehousing terus berkembang dan semakin meluas. Kami akan mendiskusikan tren dalam solusi dan produk vendor. Kami akan menghubungkan data warehousing dengan fenomena teknologi lainnya seperti Internet dan World Wide Web. Jika

diperlukan diskusi yang lebih rinci, kami akan meninjau kembali beberapa tren di bab-bab berikutnya.

3.1 PERTUMBUHAN TERUS-MENERUS DI DATA WAREHOUSING

Pergudangan data bukan lagi ide baru untuk dipelajari dan bereksperimen. Ini telah menjadi arus utama. Benar, gudang data belum ada di semua praktik dokter gigi, namun juga tidak terbatas hanya pada bisnis kelas atas. Lebih dari separuh perusahaan AS telah membuat komitmen terhadap data warehousing. Sekitar 90% perusahaan multinasional memiliki gudang data atau berencana menerapkan gudang data dalam beberapa bulan ke depan.

Di setiap industri, mulai dari toko ritel hingga lembaga keuangan, dari perusahaan manufaktur hingga departemen pemerintah, dari perusahaan penerbangan hingga farmasi, data warehousing telah merevolusi cara orang melakukan analisis bisnis dan membuat keputusan strategis. Setiap perusahaan yang memiliki gudang data terus menyadari manfaat besar yang diterjemahkan ke dalam hasil positif pada keuntungannya. Banyak dari perusahaan-perusahaan ini, yang menggabungkan teknologi berbasis web, telah meningkatkan potensi penyampaian informasi penting yang lebih besar dan lebih mudah.

Bahkan selama beberapa tahun pertama data warehousing di akhir tahun 1990an, ratusan vendor telah membanjiri pasar dengan berbagai produk. Solusi dan produk vendor menjalankan keseluruhan pergudangan data: pemodelan data, akuisisi data, kualitas data, analisis data, metadata, dan sebagainya. Panduan pembeli yang diterbitkan oleh Data Warehousing Institute saat itu menampilkan tidak kurang dari 105 produk unggulan. Pasarnya sangat besar dan pendapatan dolar terus tumbuh.

Pergudangan Data Telah Menjadi Arus Utama

Pada tahap awal, empat faktor penting yang mendorong banyak perusahaan beralih ke data warehousing:

- Persaingan sengit
- Deregulasi pemerintah
- Perlunya pembenahan proses internal
- Penting untuk pemasaran yang disesuaikan

Telekomunikasi, perbankan, dan ritel adalah industri pertama yang mengadopsi gudang data. Hal ini sebagian besar disebabkan oleh deregulasi pemerintah di bidang telekomunikasi dan perbankan. Bisnis ritel berpindah ke data warehousing karena persaingan yang semakin ketat. Perusahaan-perusahaan utilitas bergabung dengan kelompok ini ketika sektor tersebut dideregulasi. Gelombang bisnis berikutnya yang masuk ke data warehousing terdiri dari perusahaan di bidang jasa keuangan, layanan kesehatan, asuransi, manufaktur, farmasi, transportasi, dan distribusi.

Saat ini, jasa telekomunikasi dan keuangan terus memimpin dalam belanja data warehouse. Sebanyak 15% anggaran teknologi di industri ini dihabiskan untuk gudang data. Perusahaan di industri ini mengumpulkan data transaksi dalam jumlah besar. Gudang data mampu mengubah data dalam jumlah besar menjadi informasi strategis yang berguna untuk pengambilan keputusan. Saat ini, gudang data ada di setiap industri. Gambar 3-1

mencantumkan industri dalam urutan gaji rata-rata yang dibayarkan kepada profesional data warehousing. Pada tahun 2008, sektor farmasi berada di urutan teratas dengan gaji rata-rata tertinggi.

Pada tahap awal data warehousing, sebagian besar digunakan secara eksklusif oleh perusahaan global. Membangun gudang data membutuhkan biaya yang mahal dan peralatan yang diperlukan tidak memadai. Hanya perusahaan besar yang mempunyai sumber daya untuk menerapkan paradigma baru ini. Kini kita mulai melihat kehadiran data warehousing yang kuat di perusahaan-perusahaan menengah dan kecil, yang kini mampu membayar biaya pembangunan data warehouse atau membeli data mart turnkey. Lihatlah sistem manajemen basis data (DBMS) yang pernah Anda gunakan sebelumnya. Anda akan menemukan bahwa vendor database kini telah menambahkan fitur untuk membantu Anda dalam membangun gudang data menggunakan DBMS ini. Solusi terpakat juga menjadi lebih murah dan sistem operasi menjadi cukup kuat untuk mendukung fungsi data warehousing.

Perluasan Gudang Data

Meskipun gudang data sebelumnya berkonsentrasi pada penyimpanan data ringkasan untuk analisis tingkat tinggi, kini kita melihat gudang data yang semakin besar dibangun oleh berbagai bisnis. Kini perusahaan memiliki kemampuan untuk menangkap, membersihkan, memelihara, dan menggunakan sejumlah besar data yang dihasilkan oleh transaksi bisnis mereka. Jumlah data yang disimpan di gudang data terus membengkak hingga mencapai kisaran terabyte. Gudang data yang menyimpan beberapa terabyte data bukanlah hal yang aneh di bidang ritel dan telekomunikasi.

Misalnya saja industri telekomunikasi. Sebuah perusahaan telekomunikasi menghasilkan ratusan juta transaksi detail panggilan dalam setahun. Untuk mempromosikan produk dan layanan yang tepat, perusahaan perlu menganalisis transaksi rinci ini. Gudang data untuk perusahaan harus menyimpan data pada tingkat detail paling rendah. Demikian pula, pertimbangkan jaringan ritel dengan ratusan toko. Setiap hari, setiap toko menghasilkan ribuan transaksi point-of-sale. Sekali lagi, contoh lainnya adalah perusahaan di industri farmasi yang memproses ribuan pengujian dan pengukuran untuk mendapatkan persetujuan produk dari pemerintah. Gudang data di industri ini cenderung sangat besar.

Terakhir, mari kita lihat potensi ukuran unit pengendalian penipuan Medicaid di negara bagian yang besar. Organisasi ini secara eksklusif bertanggung jawab untuk menyelidiki dan menuntut penipuan layanan kesehatan yang timbul dari miliaran dolar yang dihabiskan untuk Medicaid di negara bagian tersebut. Unit ini juga harus mengadili kasus-kasus pelecehan pasien di panti jompo dan memantau praktik penagihan palsu yang dilakukan oleh dokter, apoteker, dan penyedia serta vendor layanan kesehatan lainnya. Biasanya terdapat beberapa kantor wilayah. Skema penipuan yang terdeteksi di satu wilayah harus dibandingkan dengan wilayah lainnya. Dapatkah Anda membayangkan ukuran gudang data yang diperlukan untuk mendukung unit pengendalian penipuan? Mungkin ada banyak terabyte data.

Gaji rata-rata tahunan dalam Rp. X0.000.000			
Farmasi	117	Manufaktur Non Komputer	100
Konsultasi / Profesional	112	Telekomunikasi	98
Software Engineer / Internet	107	Kesehatan	98
Pemerintahan	105	Retail/Grosir/Distributor	92
Perbankan	105	Asuransi	90
Media/Hiburan	103	Pendidikan	85
Manufaktur Komputer	103	Pemerintahan (Daerah)	82

Gambar 3.1 Industri yang menggunakan data warehousing.

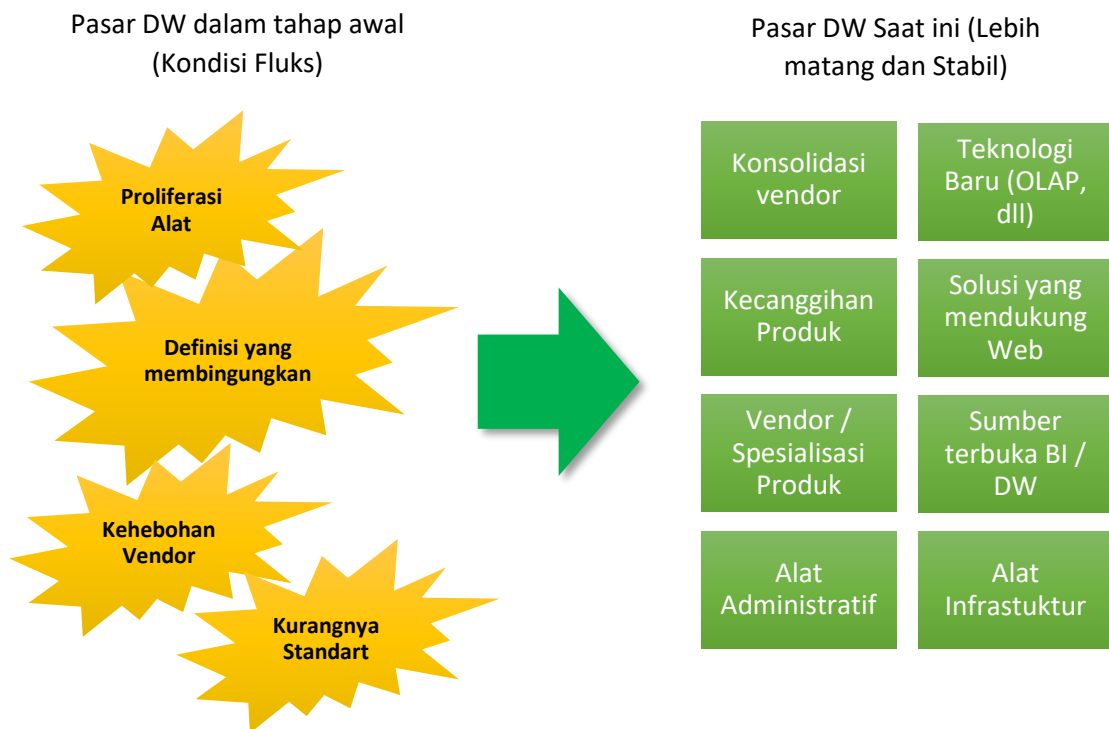
Solusi dan Produk Vendor

Sebagai seorang profesional teknologi informasi, Anda sudah familiar dengan vendor database dan produk database. Dengan cara yang sama, Anda sudah familiar dengan sebagian besar sistem operasi dan vendornya. Berapa banyak vendor database terkemuka yang ada? Berapa banyak vendor sistem operasi terkemuka yang ada? Segenggam? Jumlah vendor basis data dan sistem operasi tidak ada artinya jika dibandingkan dengan produk dan vendor pergudangan data. Ada ratusan vendor data warehousing dan ribuan produk dan solusi data warehousing.

Pada awalnya, pasar dipenuhi dengan kebingungan dan kehebohan vendor. Setiap vendor, kecil atau besar, yang memiliki produk apa pun yang terhubung secara jarak jauh ke data warehousing ikut serta. Pergudangan data berarti apa yang didefinisikan oleh masing-masing vendor. Setiap perusahaan memposisikan produknya sendiri sebagai seperangkat alat pergudangan data yang tepat. Pergudangan data adalah konsep baru bagi banyak bisnis yang mengadopsinya. Bisnis-bisnis ini bergantung pada kehebohan pemasaran para vendor.

Selama dekade terakhir, situasinya telah membaik secara signifikan. Pasar sudah mencapai kematangan, hingga mampu memproduksi paket siap pakai dan menjadi semakin stabil. Gambar 3.2 menunjukkan keadaan pasar pergudangan data saat ini dibandingkan dengan kondisi pasar pada tahap awal.

Apa yang biasanya kita lihat di pasar yang sudah matang? Kami berharap menemukan proses konsolidasi. Dan itulah yang terjadi di pasar data warehousing. Vendor data warehousing bergabung untuk membentuk perusahaan yang lebih kuat dan lebih mampu bertahan. Beberapa pemain besar di industri ini memperluas jangkauan solusi mereka dengan mengakuisisi perusahaan lain. Beberapa vendor memposisikan rangkaian produk, baik milik mereka sendiri atau produk dari kelompok vendor lain, menyatukannya sebagai solusi pergudangan data terintegrasi.



Gambar 3.2 Status pasar data warehousing.

Sekarang perusahaan database tradisional juga berada di pasar data warehousing. Mereka telah mulai menawarkan solusi data warehousing yang dibangun berdasarkan produk database mereka. Di satu sisi, alat ekstraksi dan transformasi data dikemas dengan sistem manajemen basis data. Di sisi lain, alat penyelidikan dan pelaporan ditingkatkan untuk gudang data. Beberapa vendor database mengambil peningkatan lebih lanjut dengan menawarkan produk canggih seperti alat penambangan data.

Dengan banyaknya vendor dan produk, bagaimana kita dapat mengklasifikasikan vendor dan produk, sehingga dapat memahami pasar? Yang terbaik adalah memisahkan pasar menjadi dua kelompok berbeda. Kelompok pertama terdiri dari vendor gudang data dan produk yang memenuhi kebutuhan gudang data perusahaan di mana semua data perusahaan diintegrasikan dan diubah. Segmen ini telah disebut sebagai pasar gudang data strategis. Segmen ini menyumbang sekitar seperempat dari total pasar. Segmen kedua lebih longgar dan tersebar, terdiri dari data mart departemen, sistem pemasaran basis data yang terfragmentasi, dan berbagai sistem pendukung keputusan. Vendor dan produk tertentu mendominasi setiap segmen. Kami juga dapat melihat daftar produk dengan cara lain. Gambar 3-3 menunjukkan daftar jenis produk dan layanan gudang data dan intelijen bisnis, dikategorikan berdasarkan fitur dan fungsi yang dijalankannya.

Administrasi & Operasi: Administrasi sistem/jaringan, pencadangan/pemulihan, pemulihan bencana, manajemen kinerja/penggunaan, perencanaan kapasitas, keamanan, manajemen basis data

Analisis: Pemantauan Aktivitas Bisnis, Kueri/Pelaporan, OLAP, Peramalan, Penambangan Data, Dasbor/Kartu Skor, Visualisasi, Analisis Teks, Pengembangan Aplikasi Analitik

Layanan BI: Penyedia Layanan Analitik (ASP), Perangkat Lunak BI, Konsultan, Integrator Sistem, Asosiasi/Konsorsium Industri, Riset

Integrasi Data: Manajemen Proses Bisnis, Profiling/Pembersihan/Konversi/Pergerakan Data, Pemetaan/Transformasi Data, Manajemen Data Master, Manajemen Metadata

Desain DW: Analisis Data, Pemodelan Data (Pemodelan Dimensi/Pemodelan Hubungan Entitas), Perangkat Pergudangan Data

Pengiriman Informasi: Penyiaran, Kolaborasi, Manajemen Konten, Manajemen Dokumen, Portal Informasi Perusahaan, Analisis Data Nirkabel

Infrastruktur: Akselerator Data, Peralatan Gudang Data, Sistem Manajemen Basis Data, Basis Data Multidimensi, Server, Sistem Manajemen Penyimpanan

BI/DW Sumber Terbuka: Intelijen Bisnis, Integrasi Data, Basis Data, Konsultan, Integrator Sistem

Gambar 3.3 Kategori produk dan jasa.

3.2 TREN YANG SIGNIFIKAN

Beberapa ahli berpendapat bahwa hingga saat ini, teknologilah yang mendorong data warehousing. Para ahli ini menyatakan bahwa kita sekarang mulai melihat kemajuan penting dalam perangkat lunak. Dalam beberapa tahun ke depan, data warehousing diharapkan membuat kemajuan besar dalam perangkat lunak, terutama untuk mengoptimalkan query, mengindeks tabel yang sangat besar, meningkatkan SQL, meningkatkan kompresi data, dan memperluas pemodelan dimensi.

Mari kita pisahkan tren-tren yang signifikan dan membahasnya secara singkat. Bersiaplah untuk mengunjungi setiap tren, satu per satu—masing-masing memiliki dampak serius pada data warehousing. Saat kita menelusuri setiap tren, cobalah memahami signifikansinya dan pastikan Anda memahami relevansinya dengan gudang data perusahaan Anda. Bersiaplah untuk menjawab pertanyaan: Apa yang harus Anda lakukan untuk memanfaatkan tren di gudang data Anda?

Pergudangan Data Waktu Nyata

Sistem intelijen bisnis dan gudang data pendukung telah digunakan terutama untuk pengambilan keputusan strategis. Gudang data disimpan terpisah dari sistem operasional. Baru-baru ini momentum industri beralih ke penggunaan intelijen bisnis untuk pengambilan keputusan taktis dalam operasi bisnis sehari-hari. Pergudangan data mengalami kemajuan pesat hingga pergudangan data real-time menjadi fokus para eksekutif senior.

Pergudangan data tradisional bersifat pasif, memberikan tren historis, sedangkan pergudangan data real-time bersifat dinamis, memberikan pandangan bisnis terkini secara real-time. Gudang data real-time disegarkan terus-menerus, dengan latensi hampir nol. Penyampaian informasi secara real-time meningkatkan produktivitas secara signifikan dengan

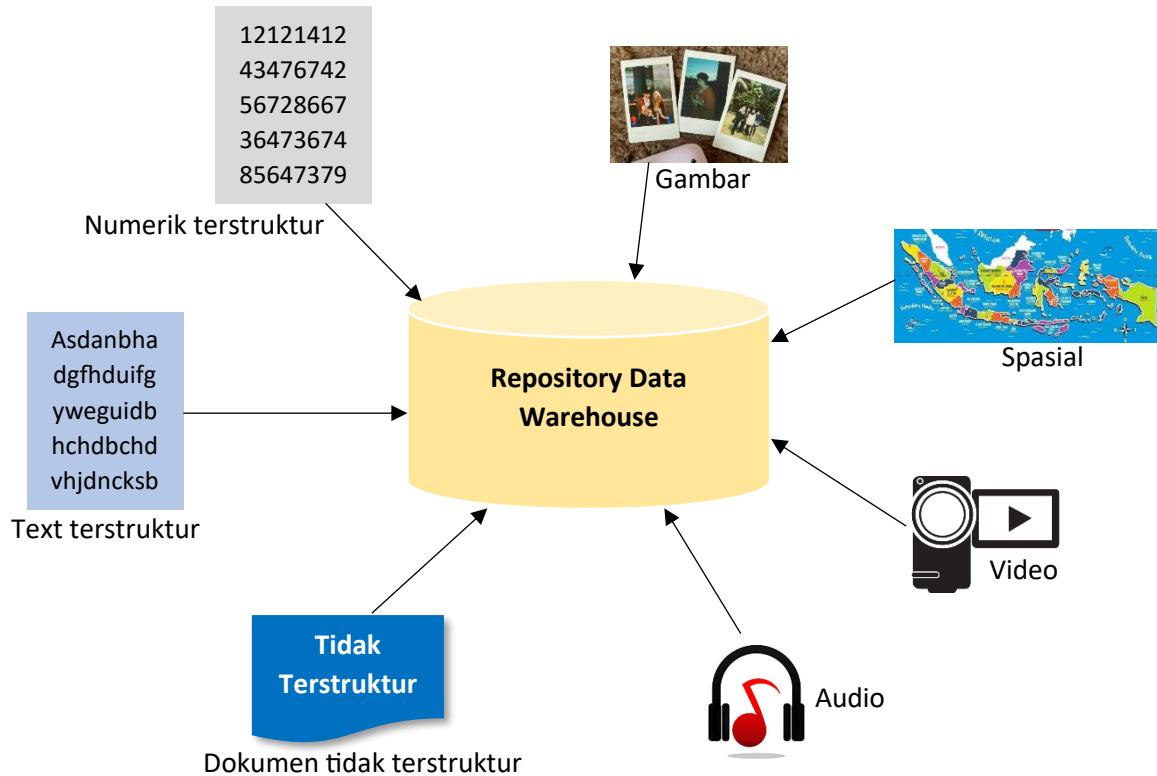
berbagi informasi dengan lebih banyak orang. Oleh karena itu, perusahaan berada di bawah tekanan besar untuk menyediakan informasi, secara real-time, kepada semua orang yang terhubung dengan proses bisnis penting. Namun, ekstraksi, transformasi, dan integrasi data untuk data warehousing real-time memiliki beberapa tantangan. Kita akan membahasnya secara rinci pada bab berikutnya.

Beberapa Tipe Data

Saat Anda membangun iterasi pertama gudang data, Anda hanya boleh menyertakan data numerik. Namun Anda akan segera menyadari bahwa memasukkan data numerik terstruktur saja tidak cukup. Bersiaplah untuk mempertimbangkan tipe data lainnya juga. Secara tradisional, perusahaan memasukkan data terstruktur, sebagian besar berupa data numerik, ke dalam gudang data mereka. Dari sudut pandang ini, sistem pendukung keputusan dibagi menjadi dua kubu: pergudangan data yang menangani data terstruktur, manajemen pengetahuan yang melibatkan data tidak terstruktur. Perbedaan ini menjadi kabur. Misalnya, sebagian besar data pemasaran terdiri dari data terstruktur dalam bentuk nilai numerik. Data pemasaran juga berisi data tidak terstruktur berupa gambar. Katakanlah seorang pengambil keputusan sedang melakukan analisis untuk menemukan jenis produk terlaris. Pengambil keputusan sampai pada jenis produk tertentu selama analisis. Dia sekarang ingin melihat gambar produk dalam jenis tersebut untuk membuat keputusan lebih lanjut. Bagaimana hal ini bisa terwujud? Perusahaan menyadari adanya kebutuhan untuk mengintegrasikan data terstruktur dan tidak terstruktur di gudang data mereka.

Jenis data apa yang kita sebut data tidak terstruktur? Gambar 3-4 menunjukkan berbagai jenis data yang perlu diintegrasikan dalam gudang data untuk mendukung pengambilan keputusan dengan lebih efektif. Sekarang mari kita beralih ke kemajuan yang dicapai dalam industri dalam memasukkan beberapa jenis data tidak terstruktur. Anda akan memperoleh pemahaman tentang apa yang harus dilakukan untuk memasukkan tipe data ini ke dalam gudang data Anda.

Menambah Data Tidak Terstruktur Beberapa vendor mengatasi penyertaan data tidak terstruktur, khususnya teks dan gambar, dengan memperlakukan data multimedia tersebut hanya sebagai tipe data lain. Ini didefinisikan sebagai bagian dari data relasional dan disimpan sebagai objek besar biner (BLOB) hingga berukuran 2 GB. Fungsi yang ditentukan pengguna (UDF) digunakan untuk mendefinisikannya sebagai tipe yang ditentukan pengguna (UDT). Tidak semua BLOB dapat disimpan hanya sebagai tipe data relasional lainnya. Misalnya, klip video memerlukan server yang mendukung pengiriman beberapa aliran video pada kecepatan tertentu dan sinkronisasi dengan bagian audio. Untuk tujuan ini, server khusus disediakan.



Gambar 3.4 Beberapa tipe data di gudang data.

Mencari Data Tidak Terstruktur Anda telah menyempurnakan gudang data Anda dengan menambahkan data tidak terstruktur. Apakah ada hal lain yang perlu Anda lakukan? Tentu saja, tanpa kemampuan untuk mencari data yang tidak terstruktur, integrasi data tersebut tidak akan ada gunanya. Vendor kini menyediakan mesin pencari baru untuk mencari informasi yang dibutuhkan pengguna dari data tidak terstruktur. Kueri berdasarkan konten gambar adalah contoh mekanisme pencarian gambar. Produk ini memungkinkan Anda mengindeks gambar terlebih dahulu berdasarkan bentuk, warna, dan tekstur. Jika lebih dari satu gambar cocok dengan argumen pencarian, gambar yang dipilih akan ditampilkan satu demi satu.

Untuk data teks bentuk bebas, mesin pengambilan mengindeks terlebih dahulu dokumen tekstual untuk memungkinkan pencarian berdasarkan kata, string karakter, frasa, wild card, operator proximity, dan operator Boolean. Beberapa mesin cukup kuat untuk menggantikan kata dan pencarian yang sesuai. Pencarian dengan kata mouse juga akan mengambil dokumen yang mengandung kata mouse. Pencarian data audio dan video secara langsung masih dalam tahap penelitian. Biasanya, hal ini dijelaskan dengan teks bentuk bebas, dan kemudian dicari menggunakan metode pencarian tekstual yang tersedia saat ini.

Data Spasial Bayangkan salah satu pengguna penting Anda, mungkin direktur pemasaran, sedang online dan melakukan analisis menggunakan gudang data Anda. Direktur pemasaran menjalankan pertanyaan: tunjukkan penjualan untuk dua kuartal pertama untuk semua produk dibandingkan tahun lalu di toko XYZ. Setelah meninjau hasilnya, dia memikirkan dua pertanyaan lainnya. Berapa pendapatan rata-rata orang yang tinggal di sekitar toko

tersebut? Berapa jarak berkendara rata-rata bagi orang-orang tersebut untuk datang ke toko? Pertanyaan-pertanyaan ini mungkin terjawab hanya jika Anda memasukkan data spasial ke dalam gudang data Anda.

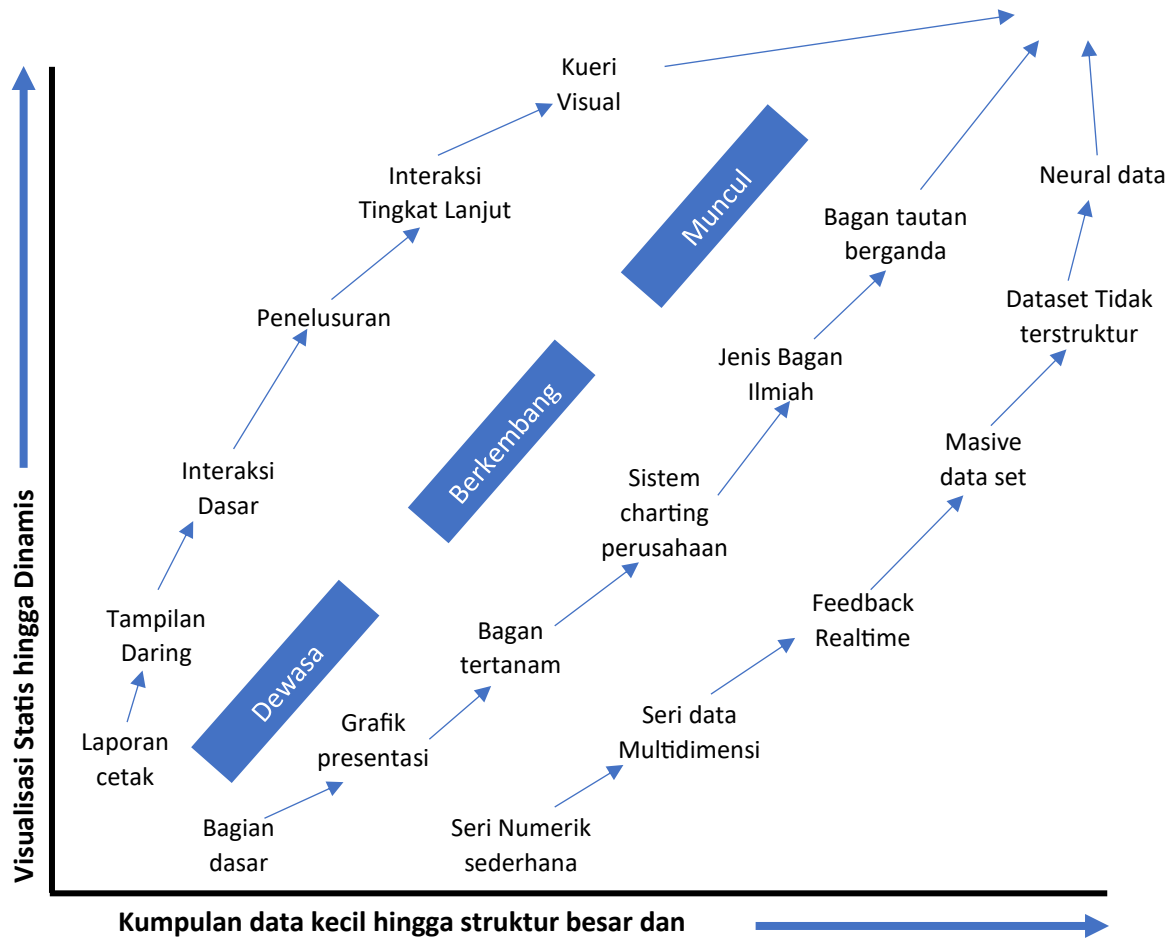
Menambahkan data spasial akan sangat meningkatkan nilai gudang data Anda. Alamat, blok jalan, kuadran kota, kabupaten, negara bagian, dan zona merupakan contoh data spasial. Vendor sudah mulai mengatasi kebutuhan untuk memasukkan data spasial. Beberapa vendor database menyediakan perluasan spasial pada produk mereka menggunakan ekstensi SQL untuk menyatukan data spasial dan bisnis.

Visualisasi data

Saat pengguna menanyakan gudang data Anda dan mengharapkan untuk melihat hasil hanya dalam bentuk daftar output atau spreadsheet, gudang data Anda sudah usang. Anda perlu menampilkan hasil dalam bentuk grafik dan bagan juga. Setiap pengguna sekarang mengharapkan untuk melihat hasil kueri yang ditampilkan sebagai grafik. Visualisasi data dalam kumpulan hasil meningkatkan proses analisis bagi pengguna, terutama saat pengguna mencari tren dari waktu ke waktu. Visualisasi data membantu pengguna menafsirkan hasil kueri dengan cepat dan mudah.

Tren Visualisasi Utama Dalam beberapa tahun terakhir, tiga tren utama telah membentuk arah perangkat lunak visualisasi data. Jenis Bagan Lainnya. Sebagian besar visualisasi data berbentuk beberapa tipe bagan standar. Hasil numerik diubah menjadi diagram lingkaran, plot sebar, atau jenis diagram lainnya. Kini daftar jenis bagan yang didukung oleh perangkat lunak visualisasi data telah bertambah panjang.

Visualisasi Interaktif. Visualisasi tidak lagi statis. Tipe bagan dinamis itu sendiri merupakan antarmuka pengguna. Pengguna Anda dapat meninjau bagan hasil, memanipulasinya, dan kemudian melihat tampilan baru secara online. Visualisasi Kumpulan Hasil yang Kompleks dan Besar. Pengguna Anda dapat melihat serangkaian titik hasil numerik sederhana sebagai diagram lingkaran atau batang yang belum sempurna. Namun perangkat lunak visualisasi yang lebih baru dapat memvisualisasikan ribuan titik hasil dan struktur data yang kompleks.



Gambar 3.5 Tren visualisasi data.

Gambar 3.5 merangkum tren-tren utama ini. Lihat bagaimana teknologi semakin matang, berkembang, dan berkembang. Jenis Visualisasi Perangkat lunak visualisasi sekarang mendukung beragam jenis bagan. Lewatlah sudah zaman grafik garis sederhana. Kebutuhan pengguna saat ini sangat bervariasi. Pengguna bisnis memerlukan diagram lingkaran dan batang. Pengguna teknis dan ilmiah memerlukan plot sebar dan grafik konstelasi. Analisis yang melihat data spasial memerlukan peta dan representasi tiga dimensi lainnya. Dalam beberapa tahun terakhir, tren besar telah membentuk arah perangkat lunak visualisasi data.

Teknik Visualisasi Tingkat Lanjut Kemajuan paling luar biasa dalam teknik visualisasi adalah transisi dari grafik statis ke presentasi interaktif dinamis. Manipulasi Bagan. Pengguna dapat memutar bagan atau mengubah jenis bagan secara dinamis untuk mendapatkan tampilan hasil yang lebih jelas. Dengan tipe visualisasi yang kompleks seperti konstelasi dan plot sebar, pengguna dapat memilih titik data dengan mouse dan kemudian memindahkan titik tersebut untuk memperjelas tampilan.

Visualisasi pertama-tama menyajikan hasil pada tingkat ringkasan. Pengguna kemudian dapat menelusuri visualisasi untuk menampilkan visualisasi lebih lanjut pada tingkat detail berikutnya. Interaksi Tingkat Lanjut. Teknik-teknik ini menyediakan antarmuka pengguna yang minimal invasif. Pengguna cukup mengklik dua kali bagian dari visualisasi dan kemudian menarik dan melepas representasi entitas data. Atau pengguna cukup mengklik

kanan dan memilih opsi dari menu. Kueri visual adalah fitur interaksi pengguna yang paling canggih. Misalnya, pengguna dapat melihat titik-titik data terpencil dalam plot sebar, lalu memilih beberapa di antaranya dengan mouse dan meminta visualisasi baru dari titik-titik yang dipilih tersebut. Perangkat lunak visualisasi data menghasilkan kueri yang sesuai dari pilihan, mengirimkan kueri ke database, dan kemudian menampilkan hasilnya dalam representasi lain.

Dasbor dan Kartu Skor Dalam beberapa tahun terakhir banyak perusahaan telah mengadopsi dasbor dan kartu skor sebagai metode pilihan mereka dalam melihat informasi kinerja. Dalam arti tertentu, cara memperoleh informasi ini merupakan puncak dari intelijen bisnis. Eksekutif dan manajer, yang perlu memantau metrik kinerja, seperti dasbor digital yang memungkinkan mereka memvisualisasikan metrik seperti pada speedometer, termometer, atau lampu lalu lintas. Dasbor memberi tahu pengguna tentang apa yang terjadi; kartu skor memberi tahu pengguna seberapa baik hal itu terjadi.

Dasbor memantau dan mengukur proses. Dasbor memberikan informasi real-time seperti dasbor pada mobil, memungkinkan pengemudi memeriksa kecepatan saat ini, suhu mesin, ketinggian cairan, dan sebagainya. Dashboard terhubung langsung ke sistem real-time yang menangkap peristiwa yang terjadi. Dasbor memperingatkan pengguna dengan peringatan atau kondisi pengecualian. Vendor industri telah meningkatkan sebagian besar dasbor dan menjadikannya interaktif dan komprehensif. Kartu skor melacak kemajuan dibandingkan dengan tujuan. Kartu skor menampilkan gambaran berkala kinerja yang dilihat berdasarkan tujuan dan target strategis organisasi. Pengguna memilih indikator kinerja utama dan menampilkan kemajuannya di kartu skor.

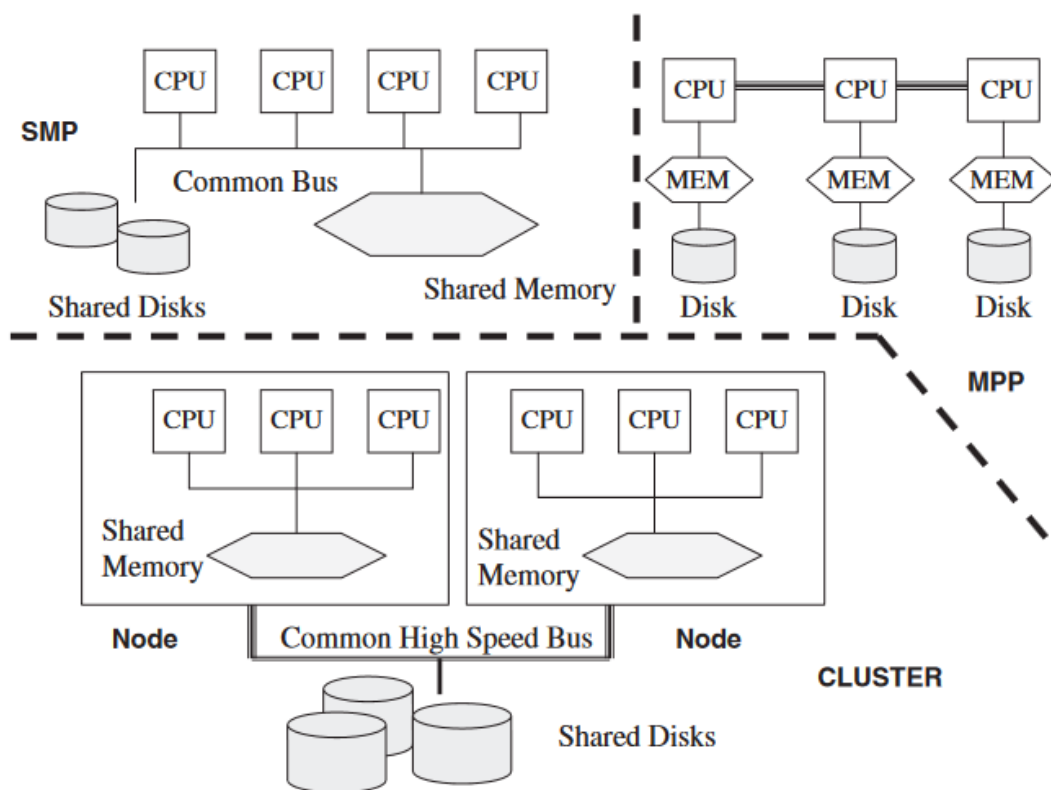
Proses paralel

Anda tahu bahwa gudang data adalah lingkungan yang berpusat pada pengguna dan intensif kueri. Pengguna Anda akan terus-menerus menjalankan kueri kompleks untuk melakukan semua jenis analisis. Setiap kueri perlu membaca data dalam jumlah besar untuk menghasilkan kumpulan hasil. Analisis, biasanya dilakukan secara interaktif, memerlukan eksekusi beberapa query, satu demi satu, oleh setiap pengguna. Jika gudang data tidak disetel dengan benar untuk menangani permintaan yang besar, kompleks, dan simultan secara efisien, nilai gudang data akan hilang. Kinerja adalah hal yang sangat penting. Fungsi lain yang kinerjanya sangat penting adalah fungsi memuat data dan membuat indeks. Karena volumenya yang besar, pemuatan data bisa menjadi lambat. Pengindeksan di gudang data biasanya rumit karena kebutuhan untuk mengakses data dengan berbagai cara. Karena banyaknya indeks, pembuatan indeks juga bisa lambat.

Bagaimana Anda mempercepat pemrosesan kueri, pemuatan data, dan pembuatan indeks? Cara yang sangat efektif untuk melakukan hal ini adalah dengan menggunakan pemrosesan paralel. Konfigurasi perangkat keras dan teknik perangkat lunak berjalan seiring untuk mencapai pemrosesan paralel. Sebuah tugas dibagi menjadi unit-unit yang lebih kecil dan unit-unit yang lebih kecil ini dijalankan secara bersamaan. Opsi Perangkat Keras Pemrosesan Paralel Dalam lingkungan pemrosesan paralel, Anda akan menemukan karakteristik berikut: beberapa CPU, modul memori, satu atau lebih node server, dan tautan

komunikasi berkecepatan tinggi antara node yang saling berhubungan. Pada dasarnya, Anda dapat memilih dari tiga pilihan arsitektur. Gambar 3.6 menunjukkan tiga pilihan dan keunggulan komparatifnya. Catat kelebihan dan kekurangannya sehingga Anda dapat memilih opsi yang tepat untuk gudang data Anda.

Implementasi Perangkat Lunak Pemrosesan Paralel Anda dapat memilih konfigurasi perangkat keras pemrosesan paralel yang sesuai untuk gudang data Anda. Perangkat keras saja tidak akan berguna jika sistem operasi dan perangkat lunak database tidak dapat memanfaatkan fitur paralel perangkat keras. Anda harus memastikan bahwa perangkat lunak dapat mengalokasikan unit tugas yang lebih besar ke komponen perangkat keras dengan tepat.



Gambar 3-6 Pilihan perangkat keras untuk pemrosesan paralel.

Perangkat lunak pemrosesan paralel harus mampu melakukan langkah-langkah berikut:

- Menganalisis tugas besar untuk mengidentifikasi unit independen yang dapat dijalankan secara paralel
- Mengidentifikasi unit mana yang lebih kecil yang harus dijalankan satu demi satu
- Mengeksekusi unit-unit independen secara paralel dan unit-unit dependen dalam urutan yang benar
- Mengumpulkan, menyusun, dan mengkonsolidasikan hasil-hasil yang dihasilkan oleh unit-unit yang lebih kecil

Vendor database biasanya menyediakan dua opsi untuk pemrosesan paralel: opsi server paralel dan opsi kueri paralel. Anda dapat membeli setiap opsi secara terpisah. Tergantung pada ketentuan yang dibuat oleh vendor database, opsi ini dapat digunakan dengan satu atau lebih konfigurasi perangkat keras paralel.

Opsi server paralel memungkinkan setiap node perangkat keras memiliki instance database terpisah, dan memungkinkan semua instance database mengakses sekumpulan file database yang mendasarinya. Opsi kueri paralel mendukung operasi utama seperti pemrosesan kueri, pemuatan data, dan pembuatan indeks. Menerapkan gudang data tanpa opsi pemrosesan paralel hampir tidak terpikirkan dalam kondisi teknologi saat ini. Singkatnya, Anda akan menyadari keuntungan signifikan berikut ketika Anda mengadopsi pemrosesan paralel di gudang data Anda:

- Peningkatan kinerja untuk pemrosesan kueri, pemuatan data, dan pembuatan indeks
- Skalabilitas, memungkinkan penambahan CPU dan modul memori tanpa perubahan pada aplikasi yang sudah ada
- Toleransi kesalahan sehingga database tersedia bahkan ketika beberapa prosesor paralel gagal
- Tampilan logis tunggal dari database meskipun datanya mungkin berada pada disk beberapa node

Peralatan Gudang Data

Pada bagian sebelumnya kita membahas bagaimana adopsi pemrosesan paralel telah menjadi kebutuhan dalam lingkungan data warehouse. Baru-baru ini perusahaan menyadari bahwa penerapan intelijen bisnis secara penuh dapat meningkatkan atau menghancurkan bisnis. Jumlah yang besar perlu diambil untuk proses pengambilan keputusan. Ketepatan waktu dan kedalaman analisis menjadi hal yang sangat penting. Gudang data berukuran terabyte sudah menjadi hal yang umum.

Sampai saat ini vendor dan pengguna telah menangani pertumbuhan pesat dalam ukuran database dan kompleksitas analisis dengan peningkatan perangkat keras dan perangkat lunak yang sangat mahal dan terus-menerus. Kini menjadi jelas bahwa infrastruktur yang ada tidak mampu menangani pertumbuhan yang berkelanjutan. Tujuannya adalah untuk menemukan pendekatan baru terhadap infrastruktur data warehousing yang fleksibel, dapat diperluas, dan secara spesifik dapat diterapkan pada suatu organisasi. Hasilnya adalah munculnya peralatan gudang data.

Peralatan gudang data dirancang khusus untuk menangani beban kerja intelijen bisnis. Hal ini dibangun dengan komponen perangkat keras dan perangkat lunak yang dirancang khusus untuk tujuan ini. Melalui arsitekturnya, alat gudang data mengintegrasikan perangkat keras, penyimpanan, dan DBMS ke dalam satu perangkat terpadu. Ini menggabungkan elemen terbaik dari SMP dan MPP untuk memungkinkan kueri diproses dengan cara yang paling optimal. Peralatan dari sebagian besar vendor dirancang untuk berinteraksi secara lancar dengan aplikasi dan alat standar yang tersedia di pasar intelijen bisnis.

Peralatan gudang data cukup skalabel. Karena semua bagian peralatan, perangkat keras dan perangkat lunak, berasal dari vendor yang sama, maka dirancang agar homogen dan dapat diandalkan. Bagi administrator, alat data warehouse memberikan kemudahan karena sifatnya yang terintegrasi. Peralatan gudang data sudah mendukung penerapan gudang data dan intelijen bisnis di perusahaan-perusahaan besar. Hal ini terutama berlaku di bidang telekomunikasi dan ritel di mana volume data sangat besar dan permintaan serta analisis lebih kompleks dan menuntut.

Alat Kueri

Di gudang data, jika ada satu set alat fungsional yang paling signifikan, itu adalah set alat kueri. Keberhasilan gudang data Anda bergantung pada alat kueri Anda. Oleh karena itu, vendor gudang data telah meningkatkan alat kueri selama beberapa tahun terakhir. Kita akan membahas alat kueri secara lebih rinci di Bab 14. Pada tahap ini, perhatikan saja fungsi-fungsi berikut yang telah ditingkatkan secara signifikan oleh vendor dalam alat kueri mereka.

- ❖ Presentasi yang fleksibel; Mudah digunakan dan mampu menyajikan hasil secara online dan dalam laporan, dalam berbagai format.
- ❖ Kesadaran agregat; Mampu mengenali keberadaan tabel ringkasan atau agregat dan secara otomatis mengarahkan kueri ke tabel ringkasan ketika hasil ringkasan diinginkan.
- ❖ Crossing subject area; Dapat berpindah dari satu data mart subjek ke data mart lainnya secara otomatis.
- ❖ Berbagai sumber heterogen; Mampu mengakses sumber data heterogen pada platform berbeda.
- ❖ Integrasi; Mampu mengintegrasikan alat kueri untuk kueri online, laporan batch, dan ekstraksi data untuk analisis, dan menyediakan antarmuka yang mulus untuk berpindah dari satu jenis keluaran ke jenis keluaran lainnya.
- ❖ Mengatasi keterbatasan SQL; Mampu menyediakan ekstensi SQL untuk menangani permintaan yang biasanya tidak dapat dilakukan melalui SQL standar.

Alat Peramban

Di sini kami menggunakan istilah “browser” dalam arti umum, tidak terbatas pada browser Web. Pengguna Anda akan menjalankan kueri terhadap gudang data Anda. Mereka akan menghasilkan laporan dari gudang data Anda. Mereka akan menjalankan fungsi-fungsi ini secara langsung dan bukan dengan bantuan orang seperti Anda di bidang TI. Hal ini diharapkan menjadi salah satu keuntungan utama dari pendekatan data warehouse.

Jika pengguna harus pergi ke gudang data secara langsung, mereka perlu mengetahui informasi apa yang tersedia di sana. Pengguna memerlukan alat browser yang baik untuk menelusuri metadata informasi dan mencari untuk menemukan informasi spesifik yang ingin mereka terima. Demikian pula, ketika Anda adalah bagian dari tim TI untuk mengembangkan gudang data perusahaan Anda, Anda perlu mengidentifikasi sumber data, struktur data, dan aturan bisnis. Anda juga memerlukan alat browser yang bagus untuk menelusuri informasi tentang sumber data. Berikut adalah beberapa tren terkini dalam penyempurnaan alat browser:

- Alat dapat diperluas untuk memungkinkan definisi jenis data atau objek informasi apa pun.
- Open API (antarmuka program aplikasi) disertakan.
- Beberapa jenis fungsi penjelajahan, termasuk navigasi melalui pengelompokan hierarki, disediakan.
- Pengguna dapat menelusuri katalog (kamus data atau metadata), menemukan objek informasi yang diinginkan, dan melanjutkan lebih jauh untuk meluncurkan alat kueri yang sesuai dengan parameter yang relevan.
- Teknik penjelajahan dan pencarian web tersedia untuk menelusuri katalog informasi.

Penggabungan Data

Gudang data adalah tempat di mana data dari berbagai sumber diintegrasikan untuk memberikan pandangan terpadu tentang perusahaan. Data mungkin berasal dari berbagai sistem operasional yang berjalan pada berbagai platform, di mana data tersebut dapat disimpan dalam file datar atau dalam database yang didukung oleh DBMS berbeda. Selain sumber internal, data dari sumber eksternal juga dimasukkan ke dalam data warehouse. Di dalam gudang data warehouse, Anda juga dapat menemukan berbagai jenis data tidak terstruktur dalam bentuk dokumen, gambar, audio, dan video.

Intinya, berbagai jenis data dari berbagai sumber yang berbeda perlu diintegrasikan atau digabungkan menjadi satu dan disimpan dalam data warehouse. Penggabungan data adalah teknologi yang berhubungan dengan penggabungan data dari sumber berbeda. Ini memiliki cakupan yang lebih luas dan mencakup penggabungan data secara real-time dari instrumen dan sistem pemantauan. Penelitian serius sedang dilakukan dalam teknologi fusi data. Prinsip dan teknik teknologi fusi data memiliki penerapan langsung dalam pergudangan data.

Penggabungan data tidak hanya berkaitan dengan penggabungan data dari berbagai sumber, tetapi juga memiliki penerapan lain dalam pergudangan data. Di gudang masa kini, kita cenderung mengumpulkan data dalam proporsi yang sangat besar. Semakin banyak informasi yang disimpan, semakin sulit menemukan informasi yang tepat pada waktu yang tepat. Teknologi fusi data diharapkan dapat mengatasi masalah ini juga. Secara umum, fusi data masih dalam bidang penelitian. Vendor belum terburu-buru memproduksi alat fusi data. Pada tahap ini, yang perlu Anda lakukan hanyalah tetap membuka mata dan memperhatikan perkembangan.

Integrasi data

Pada bagian sebelumnya kita telah membahas kebutuhan dan tantangan dalam menggabungkan data dari sumber berbeda dalam data warehouse. Mengintegrasikan data yang berbeda telah menjadi tantangan bagi pengembang data warehouse. Tugas ini kini semakin sulit mengingat ledakan data di sebagian besar bisnis. Integrasi data telah menjadi prioritas dan tantangan utama bagi banyak perusahaan. Pada bagian ini kita akan mempertimbangkan tren integrasi yang diadopsi dalam organisasi progresif.

Integrasi data tidak hanya terbatas pada proyek pergudangan data, namun juga penting bagi perusahaan untuk mengembangkan strategi integrasi di seluruh perusahaan.

Data Warehousing Institute melakukan studi tentang pendekatan integrasi data secara menyeluruh. Mereka memperhatikan bahwa organisasi-organisasi besar condong ke arah arsitektur integrasi seluruh perusahaan, sementara perusahaan-perusahaan menengah berurusan dengan integrasi data terutama dari sudut pandang intelijen bisnis.

Integrasi dilihat dalam beberapa cara berbeda: Integrasi Informasi Perusahaan (EII), Integrasi Aplikasi Perusahaan (EAI), dan seterusnya. Secara umum, integrasi dapat dilakukan pada empat tingkat berbeda dalam sistem teknologi informasi: integrasi data, integrasi aplikasi, integrasi proses bisnis, dan integrasi interaksi pengguna. Arsitektur integrasi harus menggabungkan semua tingkatan yang berbeda ini.

Integrasi data, tujuannya adalah untuk memberikan pandangan tunggal dan terpadu mengenai data bisnis yang tersebar di berbagai sistem dalam suatu organisasi. Tampilan terpadu ini mungkin berupa tampilan fisik yang telah digabungkan dan dikonsolidasikan dari berbagai sumber berbeda ke dalam satu repositori seperti gudang data atau penyimpanan data operasional. Di sisi lain, ini mungkin merupakan tampilan gabungan virtual dari data berbeda yang dikumpulkan secara dinamis pada saat akses data. Teknik lain untuk integrasi data adalah pendekatan langsung yang menggabungkan data dari satu database ke database lainnya, misalnya menggabungkan data pelanggan dari database CRM ke dalam database sistem manufaktur.

Integrasi Aplikasi, di sini tujuannya adalah untuk memberikan pandangan terpadu tentang aplikasi bisnis yang digunakan dalam organisasi. Aliran transaksi, pesan, dan data antar aplikasi dikendalikan dan dikelola untuk memperoleh tampilan terpadu. Integrasi Proses Bisnis. Tujuannya adalah untuk memberikan pandangan terpadu tentang proses bisnis organisasi. Jenis integrasi ini dicapai melalui alat desain proses bisnis dan alat manajemen proses bisnis. Integrasi Interaksi Pengguna. Tujuannya adalah untuk menyediakan antarmuka tunggal yang dipersonalisasi kepada pengguna ke proses bisnis dan data untuk menjalankan fungsi organisasi mereka. Jenis integrasi ini mendorong berbagi data dan kolaborasi pengguna. Portal perusahaan adalah contoh teknik integrasi interaksi pengguna.

Analisis

Pergudangan data dan intelijen bisnis adalah tentang analisis data. Selama beberapa tahun terakhir, vendor telah membuat kemajuan luar biasa dalam menyediakan alat analisis yang efektif. Perusahaan lebih berkonsentrasi pada dua bidang analisis dibandingkan yang lain. Analisis Multidimensi Saat ini, setiap lingkungan data warehouse menyediakan analisis multidimensi. Ini menjadi bagian integral dari sistem penyampaian informasi di gudang data. Penyediaan analisis multidimensi kepada pengguna Anda berarti bahwa mereka akan dapat menganalisis pengukuran bisnis dengan berbagai cara. Analisis multidimensi juga identik dengan pemrosesan analitis online (OLAP).

Karena pentingnya OLAP, kita akan membahas topik ini secara lebih rinci di Bab 15. Pada tahap ini, perhatikan bahwa vendor telah membuat kemajuan luar biasa dalam alat OLAP. Sekarang produk vendor sebagian besar dievaluasi berdasarkan kekuatan komponen OLAP mereka. Analisis Prediktif Sudah banyak organisasi yang memperoleh manfaat luar biasa dengan menggunakan analisis prediktif dalam lingkungan intelijen bisnis mereka. Analisis

prediktif sangat membantu organisasi dengan meningkatkan pemahaman mereka tentang perilaku pelanggan, dengan mengoptimalkan proses bisnis mereka, dengan memungkinkan mereka mengantisipasi masalah sebelum masalah itu muncul, dan dengan membantu mereka mengenali peluang jauh sebelumnya.

Teknologi Agen

Agen perangkat lunak adalah program yang mampu melakukan tugas yang dapat diprogram yang telah ditentukan sebelumnya atas nama pengguna. Misalnya, di Internet, agen perangkat lunak dapat digunakan untuk mengurutkan dan memfilter email menurut aturan yang ditentukan oleh pengguna. Di dalam gudang data, agen perangkat lunak mulai digunakan untuk mengingatkan pengguna akan kondisi bisnis yang telah ditentukan sebelumnya. Mereka juga mulai digunakan secara luas bersamaan dengan penambahan data dan teknik pemodelan prediktif. Beberapa vendor mengkhususkan diri pada alat sistem peringatan. Anda pasti harus mempertimbangkan program agen perangkat lunak untuk gudang data Anda.

Seiring dengan bertambahnya ukuran gudang data, teknologi agen semakin banyak diterapkan. Katakanlah analis pemasaran Anda perlu menggunakan gudang data Anda pada interval tertentu dan tidak berubah-ubah untuk mengidentifikasi kondisi ancaman dan peluang yang dapat menawarkan keuntungan bisnis bagi perusahaan. Analis harus menjalankan beberapa query dan melakukan analisis bertingkat untuk menemukan kondisi ini. Kondisi seperti ini merupakan kondisi pengecualian. Jadi analis harus melalui analisis berulang yang sangat intens. Beberapa kondisi ancaman dan peluang mungkin baru diketahui setelah analisis berulang dalam jangka waktu lama. Hal ini menyita banyak waktu analis, mungkin setiap hari.

Setiap kali kondisi ancaman atau peluang ditemukan melalui analisis yang rumit, masuk akal untuk menggambarkan kejadian tersebut ke program agen perangkat lunak. Program ini kemudian akan secara otomatis memberi sinyal kepada analis setiap kali kondisi tersebut ditemui di masa mendatang. Inilah inti dari teknologi agen. Agen perangkat lunak bahkan dapat digunakan untuk pemantauan rutin kinerja bisnis. CEO Anda mungkin ingin diberi tahu setiap kali penjualan seluruh perusahaan berada di bawah target bulanan, tiga bulan berturut-turut. Program agen perangkat lunak dapat digunakannya setiap kali kondisi ini terjadi. Wakil Presiden pemasaran Anda mungkin ingin mengetahui kapan promosi penjualan bulanan di semua toko berhasil. Sekali lagi, program agen perangkat lunak dapat digunakan untuk tujuan ini.

Data Sindikasi

Nilai konten data diperoleh tidak hanya dari sistem operasional internal, namun juga dari data eksternal yang sesuai. Dengan meningkatnya pertumbuhan implementasi data warehouse, pasar untuk data sindikasi berkembang pesat. Contoh pemasok tradisional data sindikasi adalah A.C. Nielsen dan Information Resources, Inc. untuk data ritel dan Dun & Bradstreet dan Reuters untuk data keuangan dan ekonomi. Beberapa gudang data sebelumnya menggabungkan data sindikasi dari pemasok tradisional untuk memperkaya konten data.

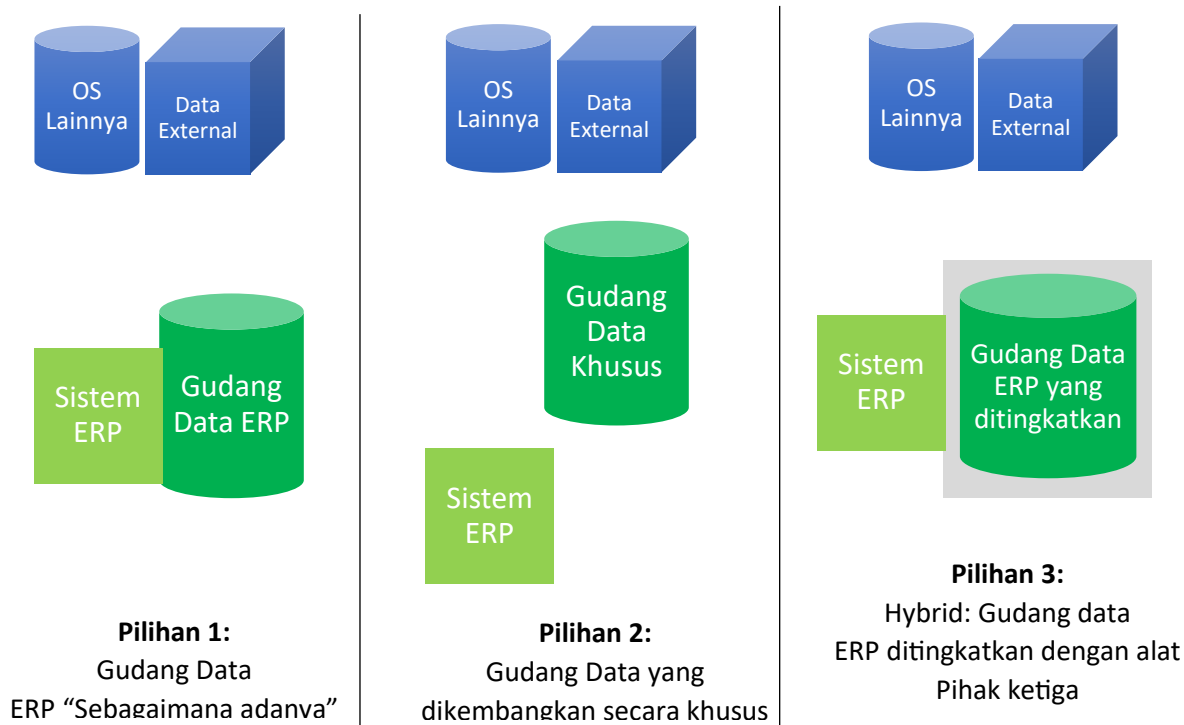
Sekarang pengembang data warehouse sedang mencari sejumlah pemasok baru yang menangani banyak jenis data sindikasi lainnya. Gudang data yang lebih baru menerima data demografis, psikografis, riset pasar, dan jenis data berguna lainnya dari pemasok baru. Data tersindikasi menjadi bisnis besar.

Pergudangan Data dan ERP

Lihatlah ke sekeliling untuk melihat jenis aplikasi apa yang telah diterapkan perusahaan dalam beberapa tahun terakhir. Anda akan mengamati fenomena yang dominan. Banyak bisnis yang mengadopsi paket aplikasi ERP (enterprise resource Planning) yang ditawarkan oleh vendor besar seperti SAP, Baan, JD Edwards, dan PeopleSoft. Pasar ERP sangat besar, melampaui angka Rp.50 miliar. Mengapa perusahaan terburu-buru dalam menerapkan ERP? Kebanyakan perusahaan diganggu oleh banyaknya aplikasi berbeda yang tidak dapat menyajikan satu kesatuan pandangan mengenai informasi perusahaan. Banyak sistem lama yang sudah ketinggalan jaman. Rekonsiliasi data yang diambil dari berbagai sistem untuk menghasilkan informasi yang bermakna dan benar sangatlah sulit dan, di beberapa perusahaan besar, hampir mustahil. Beberapa perusahaan mencari cara alternatif untuk menghindari upaya besar dalam membuat sistem lama yang sesuai dengan Y2K. Vendor ERP tampaknya datang untuk menyelamatkan perusahaan-perusahaan tersebut.

Data dalam Paket ERP Fitur luar biasa dari paket ERP adalah paket ini mendukung hampir setiap fase bisnis sehari-hari suatu perusahaan, mulai dari pengendalian inventaris hingga penagihan pelanggan, dari sumber daya manusia hingga manajemen produksi, dari penetapan biaya produk hingga pengendalian anggaran. . Karena fitur ini, paket ERP menjadi besar dan kompleks. Aplikasi ERP mengumpulkan dan mengintegrasikan banyak data perusahaan. Karena ini adalah aplikasi berpemilik, sejumlah besar data disimpan dalam format berpemilik yang tersedia untuk diakses hanya melalui program yang ditulis dalam bahasa berpemilik. Biasanya, ribuan tabel database relasional diperlukan untuk mendukung semua fungsi yang berbeda.

Mengintegrasikan ERP dan Data Warehouse Pada tahun 1990an dan seterusnya, ketika ERP diperkenalkan, solusi besar ini menjanjikan penyimpanan data korporat terintegrasi yang dicari oleh perusahaan. Karena semua data dibersihkan, diubah, dan diintegrasikan di satu tempat, visi yang menarik adalah bahwa pengambilan keputusan dan pengambilan tindakan dapat dilakukan dari satu lingkungan yang terintegrasi. Segera perusahaan yang menerapkan ERP menyadari bahwa ribuan tabel database relasional, yang dirancang dan dinormalisasi untuk menjalankan operasi bisnis, sama sekali tidak cocok untuk menyediakan informasi strategis. Selain itu, repositori data ERP kekurangan data dari sumber eksternal dan dari sistem operasional lain di perusahaan. Jika perusahaan Anda memiliki ERP atau berencana mengakuisisi ERP, Anda perlu mempertimbangkan integrasi ERP dengan data warehousing.



Gambar 3.7 Opsi integrasi ERP dan gudang data.

Opsi Integrasi Perusahaan yang mengintegrasikan ERP dan inisiatif gudang data biasanya mengadopsi salah satu dari tiga opsi yang ditunjukkan pada Gambar 3.7. Vendor ERP telah mulai melengkapi paket mereka dengan solusi data warehousing. Perusahaan yang mengadopsi opsi 1 mengimplementasikan solusi data warehousing dari vendor ERP dengan fungsionalitas yang tersedia saat ini dan menunggu peningkatan. Kelemahan dari pendekatan ini adalah Anda mungkin menunggu selamanya untuk peningkatannya. Pada opsi 2, perusahaan menerapkan gudang data yang disesuaikan dan menggunakan alat pihak ketiga untuk mengekstrak data dari kumpulan data ERP. Mengambil dan memuat data dari kumpulan data ERP berpemilik tidaklah mudah. Opsi 3 adalah pendekatan hibrid yang menggabungkan fungsionalitas yang disediakan oleh gudang data vendor dengan fungsionalitas tambahan dari alat pihak ketiga.

Pergudangan Data dan KM

Jika tahun 1998 menandai kebangkitan sistem ERP, maka tahun 1999 menandai lahirnya sistem manajemen pengetahuan (KM) di banyak perusahaan. Sejak itu, hingga tahun 2000an, KM berkembang pesat. Sistem operasional berhubungan dengan data; sistem informasi seperti gudang data memberdayakan pengguna dengan menangkap, mengintegrasikan, menyimpan, dan mengubah data menjadi informasi yang berguna untuk analisis dan pengambilan keputusan. Manajemen pengetahuan membawa ini ke tingkat yang lebih tinggi. Ini melengkapi proses dengan memberikan pengguna pengetahuan untuk menggunakan informasi yang tepat, pada waktu yang tepat, dan di tempat yang tepat. Pengetahuan adalah informasi yang dapat ditindaklanjuti. Apa yang dimaksud dengan manajemen pengetahuan? Ini adalah proses sistematis untuk menangkap, mengintegrasikan, mengatur, dan mengkomunikasikan pengetahuan yang dikumpulkan oleh karyawan. Ini

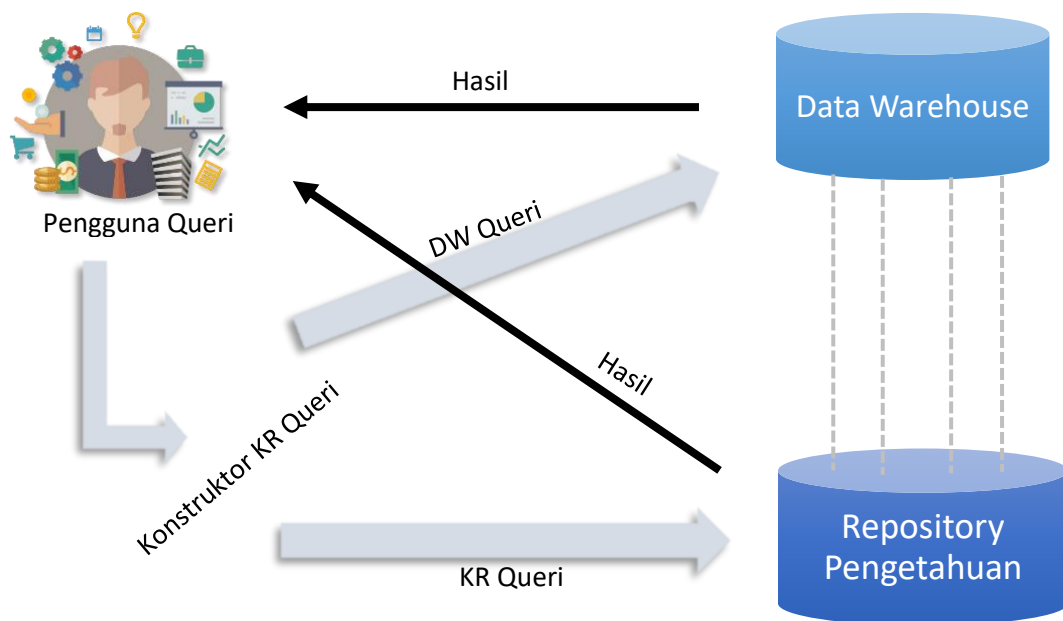
merupakan sarana untuk berbagi pengetahuan perusahaan sehingga karyawan dapat lebih efektif dan produktif dalam bekerja. Di manakah pengetahuan itu ada di sebuah perusahaan? Prosedur perusahaan, dokumen, laporan yang menganalisis kondisi pengecualian, objek, model matematika, kasus bagaimana-jika, aliran teks, klip video semua ini dan masih banyak lagi instrumen serupa yang berisi pengetahuan perusahaan.

Sistem manajemen pengetahuan harus menyimpan semua pengetahuan tersebut dalam repositori pengetahuan, kadang-kadang disebut gudang pengetahuan. Gudang data berisi informasi terstruktur; gudang pengetahuan menyimpan informasi yang tidak terstruktur. Oleh karena itu, kerangka manajemen pengetahuan harus memiliki alat untuk mencari dan mengambil informasi tidak terstruktur. Sebagai pengembang data warehouse, apa kekhawatiran Anda tentang manajemen pengetahuan? Ambil skenario perusahaan tertentu.

Katakanlah penjualan menurun di wilayah tengah selatan. Wakil Presiden pemasaran Anda dapat membedakan hal ini dari gudang data Anda dengan menjalankan beberapa pertanyaan dan melakukan beberapa analisis awal. Wakil presiden tidak mengetahui mengapa penjualan menurun, namun segala sesuatunya akan menjadi jelas jika, pada saat itu, dia mempunyai akses terhadap dokumen yang disiapkan oleh seorang analis yang menjelaskan mengapa penjualan rendah dan menyarankan tindakan perbaikan. Dokumen tersebut berisi pengetahuan yang relevan, meskipun ini adalah contoh sederhana. VP memerlukan informasi numerik, tetapi juga sesuatu yang lebih.

Pengetahuan, yang disimpan dalam format bebas tidak terstruktur, harus dikaitkan dengan hasil penjualan untuk memberikan konteks pada nomor penjualan dari gudang data. Dengan kemajuan teknologi dalam pengorganisasian, pencarian, dan pengambilan data tidak terstruktur, lebih banyak filosofi pengetahuan akan masuk ke dalam data warehousing. Gambar 3-8 menunjukkan bagaimana Anda dapat memperluas gudang data Anda untuk menyertakan pengambilan dari gudang pengetahuan yang merupakan bagian dari kerangka manajemen pengetahuan perusahaan Anda.

Sekarang, dalam skenario di atas, VP bisa mendapatkan informasi tentang penurunan penjualan dari gudang data dan kemudian mengambil dokumen analis yang relevan dari gudang pengetahuan. Pengetahuan yang diperoleh dari sistem manajemen pengetahuan dapat memberikan konteks terhadap informasi yang diterima dari data warehouse untuk memahami cerita di balik angka-angka tersebut.



Integrasi antara Data Warehouse – Repository Pengetahuan

Gambar 3.8 Integrasi KM dan data warehouse.

Pergudangan Data dan CRM

Persaingan yang lebih ketat telah memaksa banyak perusahaan memberikan perhatian lebih besar dalam mempertahankan pelanggan dan memenangkan pelanggan baru. Program loyalitas pelanggan telah menjadi hal biasa. Perusahaan beralih dari pemasaran massal ke pemasaran satu lawan satu. Fokus pelanggan sudah menjadi semboyan. Konsentrasi pada pengalaman pelanggan dan keintiman pelanggan telah menjadi kunci layanan pelanggan yang lebih baik. Semakin banyak perusahaan yang menerapkan sistem manajemen hubungan pelanggan (CRM). Sejumlah vendor terkemuka menawarkan solusi CRM siap pakai yang menjanjikan layanan satu-satu kepada pelanggan.

Ketika perusahaan Anda bersiap untuk lebih menyesuaikan diri dengan layanan pelanggan tingkat tinggi, apa yang dapat Anda, sebagai arsitek gudang data, lakukan? Jika Anda sudah memiliki data warehouse, bagaimana cara menyesuaikannya kembali? Jika Anda sedang membangun gudang data baru, faktor apa saja yang perlu mendapat perhatian khusus? Anda harus membuat gudang data Anda lebih fokus pada pelanggan. Anda harus membuat gudang data Anda siap untuk CRM, bukan tugas yang mudah. Meskipun terdapat kesulitan, manfaat dari gudang data yang mendukung CRM sangatlah besar.

Gudang data Anda harus menyimpan rincian setiap transaksi di setiap titik kontak dengan setiap pelanggan. Artinya setiap unit dari setiap penjualan setiap produk ke setiap pelanggan harus dikumpulkan di gudang data gudang. Anda tidak hanya memerlukan data penjualan secara detail tetapi juga detail setiap jenis pertemuan dengan setiap pelanggan. Selain ringkasan data, Anda harus memuat setiap pertemuan dengan setiap pelanggan ke dalam gudang data. Data atom atau detail memberikan fleksibilitas maksimum untuk gudang data yang siap CRM. Membuat gudang data Anda siap untuk CRM akan meningkatkan volume data secara signifikan. Untungnya, teknologi saat ini memfasilitasi sejumlah besar data atom

untuk ditempatkan di beberapa perangkat manajemen penyimpanan yang dapat diakses melalui alat gudang data umum.

Untuk membuat gudang data Anda siap untuk CRM, Anda juga harus meningkatkan beberapa fungsi lainnya. Untuk data terkait pelanggan, fungsi pembersihan dan transformasi lebih rumit dan rumit. Sebelum membawa catatan nama dan alamat pelanggan ke gudang data, Anda harus mengurai data tidak terstruktur untuk menghilangkan duplikat, menggabungkannya untuk membentuk rumah tangga yang berbeda, dan memperkayanya dengan data demografis dan psikografis eksternal. Ini adalah upaya besar. Alat pergudangan data tradisional tidak cukup sesuai untuk kebutuhan khusus aplikasi yang berfokus pada pelanggan.

Pengembangan Tangkas

Selama tahun 1990an semakin banyak organisasi mulai menyadari masalah yang dihadapi ketika mengadopsi metode tradisional dalam pengembangan perangkat lunak. Metodologi pengembangannya terlalu terstruktur, linier, dan tidak fleksibel. Penundaan proyek dan pembengkakan anggaran menjadi hal biasa dan menyusahkan. Organisasi menderita dampak yang luas. Upaya pembangunan memakan waktu terlalu lama dan menjadi terlalu mahal. Pengguna dan pemangku kepentingan menjadi frustrasi. Perseteruan yang terus berlanjut antar kelompok dalam tim pengembangan mengakibatkan maraknya manuver politik. Isu-isu penting lolos begitu saja. Sebuah cara baru yang radikal dalam mendekati pengembangan perangkat lunak segera diperlukan.

Menjawab seruan tegas untuk reformasi, pada bulan Februari 2001 kelompok awal yang terdiri dari 17 ahli metodologi bertemu di kota Snowbird, Utah, dan membentuk Agile Software Development Alliance (singkatnya dikenal sebagai Agile Alliance). Metodologi pengembangan perangkat lunak tangkas telah lahir. Kelompok ini tidak terdiri dari para profesional dengan keterampilan dan latar belakang yang sama persis; kelompok ini mencakup panel profesional dengan beragam latar belakang dan tingkat pengalaman. Mereka mencapai satu tujuan utama: meningkatkan upaya pengembangan perangkat lunak secara signifikan.

Anggota awal Agile Alliance menyepakati sebuah manifesto berani yang menyebutkan tema-tema preferensial dan mendasar bagi keberhasilan pengembangan perangkat lunak: manusia dan interaksi, perangkat lunak yang dapat digunakan, kolaborasi pengguna, dan penerimaan perubahan. Pengembangan perangkat lunak tangkas didasarkan pada serangkaian prinsip dan praktik yang terdefinisi dengan baik. Ini bukan seperangkat alat dan mekanisme perangkat lunak untuk pengembangan perangkat lunak.

Saat ini, banyak organisasi telah menerapkan prinsip dan praktik pengembangan tangkas dan telah menyadari keberhasilan luar biasa dalam proyek perangkat lunak mereka. Baru-baru ini perusahaan mulai menerapkan pengembangan tangkas pada proyek pergudangan data dan intelijen bisnis. Pada bab selanjutnya, kita akan membahas hal ini lebih lanjut.

Pergudangan Data Aktif

Sejauh ini kita telah membahas sejumlah tren signifikan yang sangat relevan dengan apa yang perlu Anda ingat saat membangun data warehouse Anda. Mengapa tidak mengakhiri diskusi kita tentang tren yang signifikan dengan sebuah kejutan? Mari kita lihat apa yang disebut pergudangan data aktif.

Apa pendapat Anda tentang membuka gudang data Anda kepada 30.000 pengguna di seluruh dunia, yang terdiri dari karyawan, pelanggan, dan mitra bisnis, selain mengizinkan sekitar 15 juta pengguna mengakses informasi secara publik setiap hari? Apa pendapat Anda tentang menjadikannya lingkungan pemberian layanan berkelanjutan 24 jam sehari dan 7 hari seminggu dengan ketersediaan 99,9%? Gudang data Anda dengan cepat menjadi penting bagi misi, bukan hanya bersifat strategis. Anda menggunakan pergudangan data aktif.

Inilah yang dicapai oleh sebuah perusahaan global dengan gudang data yang aktif. Perusahaan ini beroperasi di lebih dari 60 negara, memproduksi di lebih dari 40 negara, melakukan penelitian di hampir 30 negara, dan menjual lebih dari 50.000 produk di 200 negara. Keuntungan membuka data warehouse kepada pihak luar selain karyawan sangatlah besar. Pemasok bekerja sama dengan perusahaan dalam meningkatkan perencanaan permintaan dan manajemen rantai pasokan; perusahaan dan distributornya bekerja sama dalam perencanaan berbagai strategi penjualan; pelanggan membuat keputusan pembelian yang cepat. Gudang data aktif benar-benar menyediakan layanan satu-satu kepada pelanggan dan mitra bisnis.

3.3 MUNCULNYA STANDAR

Pikirkan kembali diskusi kita di Bab 1 tentang lingkungan data warehousing sebagai perpaduan banyak teknologi. Kombinasi beberapa jenis teknologi diperlukan untuk membangun gudang data. Kisarannya luas: pemodelan data, ekstraksi data, transformasi data, sistem manajemen basis data, modul kontrol, agen sistem peringatan, alat kueri, alat analisis, penulis laporan, dan sebagainya.

Saat ini, di industri yang sedang berkembang seperti data warehousing, tidak ada kekurangan vendor dan produk. Di setiap teknologi yang mendukung gudang data, terdapat banyak vendor dan produk. Implikasinya adalah ketika Anda membangun gudang data, banyak pilihan tersedia bagi Anda untuk menciptakan solusi efektif dengan produk terbaik. Itu adalah kabar baik. Namun, kabar buruknya adalah ketika Anda mencoba menggunakan produk multivendor, hasilnya juga bisa berupa kebingungan dan kekacauan total. Produk multivendor ini harus bekerja sama dan bekerja sama di gudang data Anda.

Sayangnya, tidak ada standar yang rumit dan mapan bagi berbagai produk untuk bertukar informasi dan berfungsi secara bersamaan. Saat Anda menggunakan produk database dari satu vendor, alat kueri dan reporter dari vendor lain, dan produk OLAP (pemrosesan analitis online) dari vendor lain, ketiga produk ini tidak memiliki metode standar untuk pertukaran data. Standar sangat penting dalam dua bidang: pertukaran metadata dan fungsi OLAP.

Metadata seperti peta jalan total menuju informasi yang terkandung dalam gudang data. Setiap produk menambah total konten metadata; setiap produk perlu menggunakan metadata yang dibuat oleh produk lainnya. Metadata seperti lem yang menyatukan semua bagian fungsional. Tidak ada gudang data modern yang lengkap tanpa fungsionalitas OLAP. Tanpa OLAP, Anda tidak dapat memberikan kemampuan penuh kepada pengguna Anda untuk melakukan analisis multidimensi, melihat informasi dari berbagai perspektif, dan melakukan penghitungan kompleks. OLAP sangat penting. Pada bagian berikut, kami akan meninjau kemajuan yang dicapai sejauh ini dalam menetapkan standar di dua bidang penting tersebut. Meskipun banyak kemajuan telah dicapai, hingga akhir tahun 2000an, kami belum mencapai standar yang diadopsi sepenuhnya di salah satu bidang tersebut.

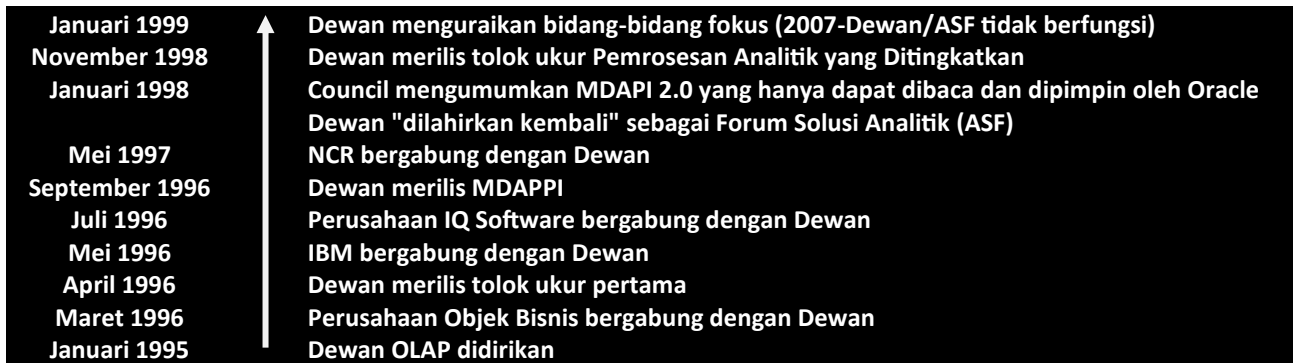
Metadata

Dua badan terpisah telah mengerjakan standar metadata, Koalisi Meta Data dan Grup Manajemen Objek. *Koalisi Meta Data* (MDC) dibentuk sebagai konsorsium yang terdiri dari sekitar 50 vendor dan pihak yang berkepentingan pada bulan Oktober 1995 untuk meluncurkan inisiatif standar metadata. Koalisi ini menerapkan standar yang dikenal sebagai Open Information Model (OIM). Microsoft bergabung dengan koalisi pada bulan Desember 1998 dan menjadi pendukung setia bersama dengan beberapa vendor terkemuka lainnya. Pada bulan Juli 1999, Koalisi Meta Data menerima Model Informasi Terbuka sebagai standar dan mulai mengerjakan perluasannya. Pada bulan November 1999, koalisi tersebut mendorong inisiatif-inisiatif penting yang baru. Namun, pada bulan November 2000, Koalisi Meta Data menghentikan operasi independennya dan bergabung dengan Object Management Group. Hal ini membuka jalan bagi standar terpadu yang menggabungkan hasil upaya dua badan independen.

Object Management Group (OMG) terdiri dari grup vendor lain, termasuk Oracle, IBM, Hewlett-Packard, Sun, dan Unisys, yang mencari standar metadata melalui Object Management Group, sebuah forum yang lebih besar dan mapan yang menangani beragam standar di bidang teknologi objek. Pada bulan Juni 2000, Object Management Group meluncurkan *Common Warehouse Metamodel* (CWM) sebagai standar pertukaran metadata untuk data warehousing. Pada bulan April 2000, Koalisi Meta Data dan Grup Manajemen Objek mengumumkan bahwa mereka akan bekerja sama dalam mencapai konsensus mengenai satu standar. Tujuan ini terwujud pada bulan November 2000 ketika kedua badan tersebut bergabung. Karena sebagian besar data perusahaan dikelola dengan alat dari Oracle, IBM, dan Microsoft, kerja sama antara kedua kubu menjadi semakin penting. Sejak merger, OMG telah berupaya mengintegrasikan OIM Koalisi Meta Data ke dalam CWM-nya.

OLAP

Dewan OLAP didirikan pada bulan Januari 1995 sebagai kelompok advokasi pelanggan yang berfungsi sebagai panduan industri. Keanggotaan dan partisipasi terbuka bagi organisasi yang berminat. Pada awal tahun 2000 dewan tersebut terdiri dari 16 anggota umum, sebagian besar adalah vendor produk OLAP.



Gambar 3.9 Dewan OLAP: jadwal kegiatan.

Selama bertahun-tahun, dewan telah mengerjakan standar OLAP untuk *Antarmuka Pemrogram Aplikasi Multi-Dimensi* (MDAPI) dan telah melakukan revisi. Gambar 3.9 menunjukkan garis waktu kegiatan-kegiatan dewan sebelumnya. Beberapa vendor OLAP, vendor platform, konsultan, dan integrator sistem telah menunjukkan dukungan mereka terhadap MDAPI 2.0. Keanggotaan OLAP menyusut dan dewan tersebut segera terlahir kembali sebagai Analytical Solutions Forum (ASF). Dalam praktiknya, MDAPI 2.0 tidak pernah benar-benar diadopsi oleh kelompok besar vendor mana pun. Tampaknya Dewan OLAP dan penggantinya ASF sudah lama dilupakan.

Gudang Data yang Diaktifkan Web

Kita semua tahu bahwa fenomena paling luar biasa yang berdampak pada komputasi dan komunikasi selama beberapa tahun terakhir adalah Internet. Di setiap konferensi industri besar dan di setiap jurnal perdagangan, sebagian besar diskusi berkaitan dengan Internet dan World Wide Web dalam satu atau lain cara.

Dimulai dengan hanya empat sistem komputer host pada tahun 1969, Internet membengkak hingga mencapai proporsi yang sangat besar, dengan hampir 95 juta host pada tahun 2000. Internet masih terus berkembang secara eksponensial. Jumlah situs World Wide Web meningkat menjadi hampir 26 juta pada tahun 2000. Hampir 150 juta pengguna global menggunakan Internet. Memanfaatkan sepenuhnya teknologi Web yang semakin populer, banyak perusahaan telah membangun Intranet dan Ekstranet untuk menjangkau karyawan, pelanggan, dan mitra bisnis mereka. Web telah menjadi sistem penyampaian informasi universal.

Kami juga menyadari bagaimana Internet telah mendorong pertumbuhan perdagangan elektronik yang luar biasa dalam beberapa tahun terakhir. Volume tahunan e-commerce bisnis-ke-bisnis melebihi Rp. 300 miliar dan total e-commerce akan segera melampaui angka Rp. 1 triliun. Tidak ada bisnis yang dapat bersaing atau bertahan tanpa kehadiran Web. Jumlah perusahaan yang menjalankan bisnis melalui Internet telah berkembang menjadi 400.000 pada tahun 2003.

Sebagai seorang profesional data warehouse, apa implikasinya bagi Anda? Jelasnya, Anda harus memanfaatkan potensi besar teknologi Internet dan Web untuk meningkatkan nilai gudang data Anda. Selain itu, Anda perlu menyadari pentingnya e-commerce dan

meningkatkan gudang Anda untuk mendukung dan memperluas bisnis elektronik perusahaan Anda. Anda harus mengubah gudang data Anda menjadi gudang data yang mendukung Web. Di satu sisi, Anda harus membawa gudang data Anda ke Web, dan di sisi lain, Anda perlu membawa Web ke gudang data Anda. Dalam dua subbagian berikutnya, kita akan membahas dua aspek berbeda dari gudang data yang mendukung Web.

Gudang ke Web

Pada awal implementasi, gudang data perusahaan ditujukan untuk para manajer, eksekutif, analis bisnis, dan beberapa karyawan tingkat tinggi lainnya sebagai alat untuk analisis dan pengambilan keputusan. Informasi dari gudang data dikirimkan ke kelompok pengguna ini dalam lingkungan client-server. Namun gudang data saat ini tidak lagi terbatas pada sekelompok pengguna internal tertentu. Dalam kondisi saat ini, korporasi perlu meningkatkan produktivitas seluruh anggota rantai nilai korporasi. Informasi yang berguna dari gudang data perusahaan harus disediakan tidak hanya kepada karyawan tetapi juga kepada pelanggan, pemasok, dan semua mitra bisnis lainnya.

Jadi dalam iklim bisnis saat ini, Anda perlu membuka gudang data Anda kepada seluruh komunitas pengguna dalam rantai nilai, dan mungkin juga kepada masyarakat umum. Ini adalah tugas yang sulit. Bagaimana Anda dapat memenuhi persyaratan ini untuk menyajikan informasi kepada ribuan pengguna dalam mode 24/7? Bagaimana Anda dapat melakukan hal ini tanpa mengeluarkan biaya selangit untuk penyampaian informasi? Internet, bersama dengan teknologi Web, adalah jawabannya. Web akan menjadi mekanisme penyampaian informasi utama Anda.

Metode pengiriman baru ini akan secara radikal mengubah cara pengguna Anda mengambil, menganalisis, dan berbagi informasi dari gudang data Anda. Komponen penyampaian informasi Anda akan berbeda. Antarmuka Internet akan mencakup browser, mesin pencari, teknologi push, halaman beranda, konten informasi, tautan hypertext, dan applet Java atau ActiveX yang diunduh. Saat Anda membawa gudang data Anda ke Web, dari sudut pandang pengguna, persyaratan utamanya adalah: akses data layanan mandiri, analisis interaktif, ketersediaan dan kinerja tinggi, klien tanpa administrasi (teknologi klien tipis seperti Applet Java), keamanan yang ketat, dan metadata terpadu.

Web ke Gudang

Membawa Web ke gudang pada dasarnya melibatkan menangkap clickstream dari semua pengunjung situs Web perusahaan Anda dan melakukan semua fungsi data warehousing tradisional. Dan Anda harus mencapai hal ini, hampir secara real-time, dalam lingkungan yang kini dikenal sebagai Webhouse data. Upaya Anda akan melibatkan ekstraksi, transformasi, dan pemuatan data clickstream ke repositori Webhouse. Anda harus membangun skema dimensi dari data clickstream dan menerapkan sistem pengiriman informasi dari Webhouse.

Data Clickstream melacak bagaimana orang menelusuri situs Web perusahaan Anda, apa yang memicu pembelian, apa yang menarik orang, dan apa yang membuat mereka datang kembali. Data Clickstream memungkinkan analisis beberapa langkah utama, termasuk:

- ◆ Permintaan pelanggan

- ◆ Efektivitas promosi pemasaran
- ◆ Efektivitas hubungan afiliasi antar produk
- ◆ Pengumpulan data demografi
- ◆ Pola pembelian pelanggan
- ◆ Umpan balik mengenai desain situs Web

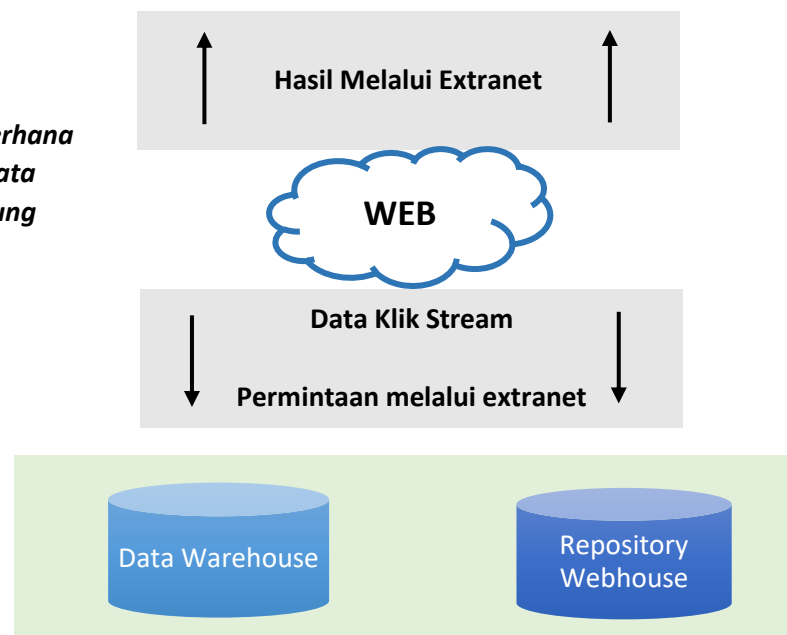
Webhouse clickstream mungkin merupakan satu-satunya alat paling penting untuk mengidentifikasi, memprioritaskan, dan mempertahankan pelanggan e-commerce.

Webhouse dapat menghasilkan informasi berguna berikut:

- ◆ Statistik situs
- ◆ Konversi pengunjung
- ◆ Metrik iklan
- ◆ Merujuk tautan mitra
- ◆ Navigasi situs menghasilkan pesanan
- ◆ Navigasi situs tidak menghasilkan pesanan
- ◆ Halaman yang merupakan session killer
- ◆ Hubungan antara profil pelanggan dan aktivitas halaman
- ◆ Analisis pelanggan terbaik dan pelanggan terburuk



Tampilan sederhana dari gudang data yang mendukung web



Gambar 3.10 Gudang data yang mendukung web.

Konfigurasi yang Diaktifkan Web

Gambar 3.10 menunjukkan konfigurasi arsitektur untuk gudang data yang mendukung Web. Perhatikan kehadiran fitur fungsional penting dari gudang data tradisional. Selain repositori gudang data yang menyimpan jenis informasi biasa, repositori Webhouse berisi data clickstream. Konvergensi Web dan data warehousing merupakan hal yang sangat penting bagi setiap perusahaan yang melakukan bisnis di abad kedua puluh satu. Karena pentingnya hal ini, kita akan membahas topik ini secara lebih rinci di Bab 16.

RINGKASAN BAB

- Data warehousing telah menjadi arus utama seiring dengan tersebarnya data warehouse bervolume tinggi dan pesatnya peningkatan jumlah produk vendor.
- Agar efektif, gudang data modern perlu menyimpan berbagai jenis data: terstruktur dan tidak terstruktur, termasuk dokumen, gambar, audio, dan video.
- Visualisasi data berkaitan dengan menampilkan informasi dalam beberapa jenis bentuk visual: teks, array numerik, spreadsheet, bagan, grafik, dan sebagainya. Kemajuan luar biasa telah dicapai dalam visualisasi data, termasuk informasi melalui dasbor dan kartu skor.
- Kinerja gudang data dapat ditingkatkan dengan menggunakan pemrosesan paralel dengan pilihan perangkat keras dan perangkat lunak yang sesuai. Banyak perusahaan memilih peralatan gudang data.
- Penting untuk mengadaptasi data warehousing agar dapat bekerja dengan paket ERP, manajemen pengetahuan, dan sistem hubungan pelanggan.
- Industri data warehousing secara serius mencari standar yang disepakati untuk metadata dan OLAP. Akhir zaman mungkin sudah di depan mata.
- Web-enabled data warehouse berarti menggunakan Web untuk penyampaian informasi dan mengintegrasikan data clickstream dari situs Web perusahaan untuk analisis. Konvergensi data warehousing dan teknologi Web sangat penting bagi setiap bisnis di abad kedua puluh satu.

LATIHAN SOAL

1. Sebutkan tiga faktor yang menunjukkan pertumbuhan berkelanjutan dalam data warehousing dan intelijen bisnis. Dapatkah Anda memikirkan beberapa contoh?
2. Mengapa data warehouse terus bertambah besar dan menyimpan data dalam jumlah besar? Berikan tiga alasan.
3. Mengapa penting untuk menyimpan berbagai jenis data di gudang data? Berikan contoh beberapa data tidak terstruktur yang mungkin ditemukan di gudang data organisasi manajemen kesehatan (HMO).
4. Apa yang dimaksud dengan fusi data? Di mana letaknya dalam data warehousing?
5. Jelaskan empat jenis grafik yang mungkin Anda lihat dalam penyampaian informasi dari data mart yang mendukung departemen keuangan.

6. Apa pemahaman Anda tentang peralatan gudang data? Sebutkan beberapa manfaat peralatan gudang data.
7. Jelaskan apa yang dimaksud dengan teknologi agen? Bagaimana teknologi ini dapat digunakan dalam data warehouse?
8. Jelaskan salah satu opsi yang tersedia untuk mengintegrasikan ERP dengan data warehousing.
9. Apa itu CRM? Bagaimana Anda bisa membuat gudang data Anda siap untuk CRM?
10. Apa yang dimaksud dengan gudang data yang mendukung Web? Jelaskan tiga fitur fungsionalnya.

BAB 4

PERENCANAAN DAN MANAJEMEN PROYEK

TUJUAN BAB

- Meninjau esensi perencanaan gudang data
- Membedakan antara proyek gudang data dan proyek sistem OLTP
- Pelajari cara mengadaptasi pendekatan siklus hidup untuk proyek gudang data
- Memperkenalkan metodologi pembangunan yang tangkas untuk proyek-proyek PRT
- Diskusikan organisasi tim proyek, peran, dan tanggung jawab
- Pertimbangkan tanda-tanda peringatan dan faktor keberhasilan

Begitu Anda membaca judul bab ini, Anda mungkin langsung menyimpulkan bahwa bab ini ditujukan untuk manajer proyek atau koordinator proyek. Jika Anda belum menjadi manajer proyek atau berencana menjadi manajer proyek dalam waktu dekat, Anda mungkin cenderung membaca sekilas bab ini saja. Itu adalah sebuah kesalahan. Bab ini dirancang untuk semua profesional TI, terlepas dari peran mereka dalam proyek pergudangan data. Ini akan menunjukkan kepada Anda cara terbaik untuk menyesuaikan diri dengan peran spesifik Anda dalam sebuah proyek. Jika Anda ingin menjadi bagian dari tim yang bersemangat membangun gudang data yang sukses, Anda memerlukan detail yang disajikan dalam bab ini. Jadi silakan baca terus.

Baca dulu pengakuannya berikut ini.

- Konsultan** : Jadi, perusahaan Anda bergerak di bidang data warehousing? Berapa banyak data mart yang Anda miliki?
- Manajer Proyek** : Sebelas.
- Konsultan** : Bagus sekali. Tapi kenapa banyak sekali?
- Manajer Proyek** : Sepuluh kesalahan.

Meskipun pembicaraan ini agak berlebihan, menurut pakar industri, lebih dari 50% proyek gudang data dianggap gagal. Dalam banyak kasus, proyek tidak selesai dan sistem tidak terlaksana. Dalam beberapa kasus, proyek entah bagaimana selesai tetapi gudang data ternyata menjadi gudang data. Proyek ini memiliki ukuran dan arsitektur yang tidak tepat. Gudang data tidak selaras dengan bisnis. Proyek ditinggalkan di tengah jalan.

Beberapa faktor berkontribusi terhadap kegagalan tersebut. Saat perusahaan Anda memasuki data warehousing untuk pertama kalinya, proyek tersebut akan melibatkan banyak perubahan organisasi. Saat ini, penekanannya adalah pada analisis informasi di seluruh perusahaan. Hingga saat ini, setiap departemen dan setiap pengguna “memiliki” datanya dan peduli dengan serangkaian sistem komputer “sendiri”. Pergudangan data telah mengubah semua itu dan membuat pengelola, pemilik data, dan pengguna akhir merasa tidak nyaman. Anda kemungkinan besar akan menemukan masalah dengan sistem produksi saat Anda membangun gudang data.

4.1 MERENCANAKAN GUDANG DATA ANDA

Lebih dari faktor lainnya, perencanaan yang tidak tepat dan manajemen proyek yang tidak memadai cenderung mengakibatkan kegagalan. Pertama dan terpenting, tentukan apakah perusahaan Anda benar-benar membutuhkan gudang data. Apakah ini benar-benar siap untuk itu? Anda perlu mengembangkan kriteria untuk menilai nilai yang diharapkan dari gudang data Anda. Perusahaan Anda harus memutuskan jenis gudang data yang akan dibangun dan di mana menyimpannya. Anda harus memastikan dari mana data tersebut berasal dan apakah Anda memiliki semua data yang diperlukan. Anda harus menentukan siapa yang akan menggunakan data warehouse, bagaimana mereka akan menggunakannya, dan pada waktu apa.

Kami akan membahas berbagai masalah terkait perencanaan data warehouse yang tepat. Anda akan mempelajari perbedaan proyek gudang data dari jenis proyek yang pernah Anda tangani sebelumnya. Kami akan mempelajari pedoman untuk membuat proyek gudang data Anda sukses.

Masalah kunci

Perencanaan gudang data Anda dimulai dengan pertimbangan menyeluruh terhadap isu-isu utama. Jawaban atas pertanyaan-pertanyaan kunci sangat penting untuk perencanaan yang tepat dan keberhasilan penyelesaian proyek. Oleh karena itu, mari kita bahas masalah-masalah yang berkaitan satu per satu. Nilai dan Harapan Beberapa perusahaan terjun ke proyek pergudangan data dan intelijen bisnis tanpa menilai nilai yang akan diperoleh dari gudang data yang mereka usulkan. Tentu saja, pertama-tama Anda harus yakin bahwa, mengingat budaya dan kebutuhan perusahaan Anda saat ini, gudang data adalah solusi yang paling tepat.

Setelah Anda menentukan kesesuaian solusi ini, barulah Anda dapat mulai menghitung manfaat dan proposisi nilainya. Akankah gudang data Anda membantu para eksekutif dan manajer melakukan perencanaan yang lebih baik dan membuat keputusan yang lebih baik? Apakah ini akan meningkatkan keuntungan? Apakah ini akan meningkatkan pangsa pasar? Jika ya, berapa banyak? Apa harapannya? Apa yang ingin dicapai manajemen melalui gudang data? Sebagai bagian dari keseluruhan proses perencanaan, buatlah daftar manfaat dan harapan yang realistis. Ini adalah titik awalnya.

Perencana Penilaian Risiko umumnya mengaitkan risiko proyek dengan biaya proyek. Jika proyek tersebut gagal, berapa banyak uang yang akan terbuang percuma? Namun penilaian risiko lebih dari sekadar menghitung kerugian biaya proyek. Apa saja risiko yang dihadapi perusahaan tanpa manfaat yang dapat diperoleh dari data warehouse? Kerugian apa saja yang mungkin terjadi? Peluang apa yang mungkin terlewatkan? Penilaian risiko bersifat luas dan relevan untuk setiap bisnis. Gunakan budaya dan kondisi bisnis perusahaan Anda untuk menilai risikonya. Sertakan penilaian ini sebagai bagian dari dokumen perencanaan Anda.

Top-Down atau Bottom-Up Pada Bab 2, kita membahas pendekatan top-down dan bottom-up untuk membangun data warehouse. Pendekatan top-down adalah memulai dari gudang data seluruh perusahaan, dan mungkin membangunnya secara berulang. Kemudian

data dari gudang data keseluruhan perusahaan besar mengalir ke data mart departemen dan subjek. Di sisi lain, pendekatan bottom-up adalah memulai dengan membangun data mart individual, satu per satu. Konglomerat data mart ini akan membentuk gudang data perusahaan.

Kami melihat pro dan kontra dari kedua metode tersebut. Kami juga membahas pendekatan praktis dari bottom-up, namun memastikan bahwa masing-masing data mart disesuaikan satu sama lain sehingga dapat dilihat secara keseluruhan. Agar pendekatan praktis ini berhasil, Anda harus terlebih dahulu merencanakan dan menentukan persyaratan di tingkat perusahaan secara keseluruhan.

Anda harus mempertimbangkan pilihan-pilihan ini ketika diterapkan pada perusahaan Anda. Apakah Anda memiliki sumber daya besar yang diperlukan untuk membangun gudang data seluruh perusahaan terlebih dahulu dan kemudian menerapkan data mart individual? Opsi ini mungkin memerlukan lebih banyak waktu untuk diterapkan dan menunda realisasi manfaat potensial. Namun opsi ini, berdasarkan pendekatan bawaannya, akan memastikan pandangan yang sepenuhnya terpadu mengenai data perusahaan.

Ada kemungkinan bahwa perusahaan Anda akan puas dengan penerapan cepat beberapa data mart karena alasan tertentu. Pada saat ini, mungkin penting untuk bereaksi dengan cepat terhadap kekuatan pasar atau menangkal pesaing yang kuat. Mungkin tidak ada waktu untuk membangun gudang data secara keseluruhan. Atau, Anda mungkin ingin menguji dan mengadopsi pendekatan praktis dari data mart yang disesuaikan. Pendekatan apa pun yang ingin diadopsi oleh perusahaan Anda, cermati pilihannya dengan cermat dan buatlah pilihan. Dokumentasikan implikasi pilihan dalam dokumen perencanaan.

Membangun atau Membeli Ini adalah masalah besar bagi semua organisasi. Tidak ada seorang pun yang membangun gudang data sepenuhnya dari awal dengan pemrograman internal. Tidak perlu menemukan kembali roda setiap saat. Tersedia beragam alat dan solusi pihak ketiga. Setelah hampir satu dekade pergerakan data warehousing, pasar telah matang, dengan alat yang sesuai untuk data warehousing dan intelijen bisnis. Pertanyaan sebenarnya adalah seberapa banyak data mart yang harus Anda bangun sendiri? Berapa banyak dari solusi ini yang dapat terdiri dari solusi siap pakai? Mix and match seperti apa yang harus dilakukan?

Di gudang data, ada berbagai macam fungsi. Apakah Anda ingin menulis lebih banyak program internal untuk ekstraksi data dan transformasi data? Apakah Anda ingin menggunakan program internal untuk memuat penyimpanan data warehouse? Apakah Anda ingin menggunakan alat vendor sepenuhnya untuk penyampaian informasi dan intelijen bisnis? Anda tetap memegang kendali atas fungsi di mana pun Anda menggunakan perangkat lunak internal. Di sisi lain, opsi beli dapat mempercepat implementasi jika dikelola secara efektif.

Berhati-hatilah dengan mart-in-the-box atau data mart 15 menit. Tidak ada solusi terbaik di luar sana. Intinya adalah mengerjakan pekerjaan rumah Anda dan menemukan keseimbangan yang tepat antara perangkat lunak internal dan vendor. Lakukan ini pada tahap perencanaan itu sendiri. Vendor Tunggal atau Vendor Terbaik hadir dalam berbagai kategori. Ada banyak vendor dan produk yang melayani banyak fungsi gudang data. Jadi apa saja

pilihannya? Bagaimana sebaiknya Anda memutuskan? Dua pilihan utama adalah: menggunakan produk dari satu vendor, atau menggunakan produk dari lebih dari satu vendor, dengan memilih alat yang sesuai. Memilih solusi vendor tunggal memiliki beberapa keuntungan:

- ◆ Tingkat integrasi yang tinggi antar alat
- ◆ Tampilan dan nuansa yang konstan
- ◆ Kerjasama yang lancar antar komponen
- ◆ Pertukaran informasi yang dikelola secara terpusat
- ◆ Harga keseluruhan bisa dinegosiasikan

Pendekatan ini secara alami akan memungkinkan gudang data Anda terintegrasi dengan baik dan berfungsi secara koheren. Namun, hanya sedikit vendor seperti IBM dan NCR yang menawarkan solusi terintegrasi penuh.

Meninjau opsi khusus ini lebih lanjut, berikut adalah keuntungan utama dari solusi terbaik yang menggabungkan produk dari beberapa vendor:

- ◆ Anda dapat membangun lingkungan yang sesuai dengan organisasi Anda.
- ◆ Tidak perlu berkompromi antara database dan alat pendukung.
- ◆ Anda dapat memilih produk yang paling sesuai untuk fungsi tertentu.

Dengan pendekatan yang terbaik, kompatibilitas antar alat dari vendor yang berbeda dapat menjadi masalah yang serius. Jika Anda mengambil rute ini, pastikan alat yang dipilih terbukti kompatibel. Dalam hal ini, daya tahan masing-masing vendor sangatlah penting. Selain itu, Anda akan memiliki daya tawar yang lebih rendah terhadap masing-masing produk dan mungkin mengeluarkan biaya keseluruhan yang lebih tinggi. Catat pendekatan yang direkomendasikan: mintalah satu vendor untuk database dan fungsi pengiriman informasi, dan pilih vendor lain untuk fungsi lainnya. Namun, pendekatan multivendor tidak disarankan jika lingkungan Anda tidak terlalu bersifat teknis.

Persyaratan Bisnis, Bukan Teknologi

Biarkan kebutuhan bisnis yang mengendalikan gudang data Anda, bukan teknologi. Meskipun hal ini tampak jelas, Anda tidak akan percaya betapa banyak proyek data warehouse yang melanggar prinsip ini. Banyak pengembang data warehouse tertarik untuk menampilkan gambar-gambar cantik di layar pengguna dan kurang memperhatikan kebutuhan sebenarnya. Mereka suka membangun sistem cepat dengan memanfaatkan kedalaman teknologi dan sekadar menunjukkan kehebatan mereka dalam memanfaatkan kekuatan teknologi.

Ingat, data warehousing bukan tentang teknologi, melainkan tentang menyelesaikan kebutuhan pengguna akan informasi strategis. Jangan berencana membangun gudang data sebelum memahami persyaratannya. Mulailah dengan berfokus pada informasi apa yang dibutuhkan dan bukan pada bagaimana memberikan informasi tersebut. Jangan menekankan alatnya. Alat dan produk datang dan pergi. Struktur dasar dan arsitektur untuk mendukung kebutuhan pengguna lebih penting.

Jadi sebelum membuat rencana keseluruhan, lakukan survei awal terhadap persyaratan. Bagaimana kamu melakukannya? Tidak ada rincian yang diperlukan pada tahap ini. Tidak diperlukan penyelidikan mendalam. Coba saja pahami kebutuhan pengguna secara

keseluruhan. Niat Anda adalah untuk mendapatkan pemahaman luas tentang bisnis ini. Hasil survei pendahuluan ini akan membantu Anda merumuskan rencana keseluruhan. Penting untuk menetapkan ruang lingkup proyek. Selain itu, ini akan membantu Anda dalam memprioritaskan dan menentukan rencana peluncuran untuk data mart individual. Misalnya, Anda mungkin harus merencanakan untuk meluncurkan data mart pemasaran terlebih dahulu, selanjutnya mart keuangan, dan baru kemudian mempertimbangkan sumber daya manusia.

Jenis informasi apa yang harus Anda kumpulkan dalam survei pendahuluan? Minimal, dapatkan informasi umum berikut ini dari setiap kelompok pengguna:

- ◆ Misi dan fungsi masing-masing kelompok pengguna
- ◆ Sistem komputer yang digunakan oleh kelompok
- ◆ Indikator kinerja utama
- ◆ Faktor-faktor yang mempengaruhi keberhasilan kelompok pengguna
- ◆ Siapa pelanggannya dan bagaimana mereka diklasifikasikan
- ◆ Jenis data yang dilacak untuk pelanggan, secara individu dan kelompok
- ◆ Produk yang diproduksi atau dijual
- ◆ Kategorisasi produk dan layanan
- ◆ Lokasi di mana bisnis dilakukan
- ◆ Tingkat keuntungan yang diukur—per pelanggan, per produk, per distrik
- ◆ Tingkat rincian biaya dan pendapatan
- ◆ Pertanyaan dan laporan terkini untuk informasi strategis

Sebagai bagian dari survei pendahuluan, sertakan audit sistem sumber. Bahkan pada tahap ini, Anda harus mempunyai gagasan yang cukup bagus dari mana data akan diekstraksi untuk gudang data. Tinjau arsitektur sistem sumber. Cari tahu tentang hubungan antar struktur data. Bagaimana kualitas datanya? Dokumentasi apa yang tersedia? Apa mekanisme yang mungkin untuk mengekstraksi data dari sistem sumber? Rencana keseluruhan Anda harus berisi informasi tentang sistem sumber.

Dukungan Manajemen Puncak

Tidak ada inisiatif besar dalam sebuah perusahaan yang dapat berhasil tanpa dukungan dari manajemen senior. Hal ini berlaku dalam kasus proyek gudang data perusahaan. Proyek ini harus mendapat dukungan penuh dari manajemen puncak sejak hari pertama. Tidak ada perusahaan lain yang menyatukan pandangan informasi seluruh perusahaan seperti halnya gudang data perusahaan. Seluruh organisasi dilibatkan dan diposisikan untuk mendapatkan keuntungan strategis. Tidak ada satu departemen atau kelompok yang dapat mensponsori inisiatif data warehousing di sebuah perusahaan.

Pastikan Anda memiliki sponsor dari tingkat manajemen tertinggi untuk tetap fokus. Gudang data sering kali harus memenuhi persyaratan yang saling bertentangan. Sponsor harus menggunakan pengaruhnya untuk melakukan arbitrase dan mediasi. Di sebagian besar perusahaan yang meluncurkan gudang data, CEO juga secara langsung tertarik pada keberhasilannya. Di beberapa perusahaan, eksekutif senior di luar TI menjadi sponsor utama. Orang ini, pada gilirannya, menominasikan beberapa manajer senior untuk terlibat aktif dalam

kemajuan proyek sehari-hari. Setiap kali proyek mengalami kemunduran serius, sponsor akan turun tangan untuk menyelesaikan masalah tersebut.

Membenarkan Gudang Data Anda

Bahkan jika perusahaan Anda adalah perusahaan menengah, ketika semuanya diperhitungkan, total investasi di gudang data Anda bisa mencapai beberapa juta dolar. Rincian kasar biayanya adalah sebagai berikut: perangkat keras, 31%; perangkat lunak, termasuk DBMS, 24%; staf dan integrator sistem, 35%; administrasi, 10%. Bagaimana Anda membenarkan total biaya dengan menyeimbangkan risiko terhadap manfaat, baik yang berwujud maupun tidak berwujud? Bagaimana cara menghitung ROI dan ROA? Bagaimana cara membuat kasus bisnis?

Ini tidak mudah. Manfaat sebenarnya mungkin baru diketahui setelah gudang data Anda dibangun dan digunakan sepenuhnya. Gudang data Anda akan memungkinkan pengguna menjalankan kueri dan menganalisis variabel dengan berbagai cara. Pengguna Anda dapat menjalankan analisis bagaimana-jika dengan beralih ke beberapa skenario hipotetis dan membuat keputusan strategis. Mereka tidak akan dibatasi dalam cara mereka melakukan kueri dan menganalisis. Siapa yang dapat memprediksi pertanyaan dan analisis apa yang mungkin mereka jalankan, keputusan signifikan apa yang dapat mereka ambil, dan seberapa besar dampak keputusan tersebut terhadap keuntungan perusahaan?

Banyak perusahaan yang mampu memperkenalkan data warehousing tanpa analisis pembenaran biaya secara penuh. Di sini pembenarannya terutama didasarkan pada intuisi dan potensi tekanan persaingan. Di perusahaan-perusahaan ini, manajemen puncak dapat dengan mudah mengenali manfaat integrasi data, peningkatan kualitas data, otonomi pengguna dalam menjalankan kueri dan analisis, dan kemudahan aksesibilitas informasi. Jika perusahaan Anda adalah perusahaan seperti itu, semoga Anda beruntung. Lakukan beberapa pembenaran dasar dan terjun ke proyek dengan kedua kaki di dalamnya.

Tidak semua manajemen puncak perusahaan mudah untuk dipuaskan. Di banyak perusahaan, beberapa jenis pembenaran formal diperlukan. Kami ingin menyajikan pendekatan umum yang diambil untuk membenarkan proyek gudang data. Tinjau contoh-contoh ini dan pilih pendekatan yang paling mendekati keberhasilan di organisasi Anda. Berikut adalah beberapa contoh pendekatan untuk mempersiapkan pembenaran:

1. Hitung biaya teknologi saat ini untuk menghasilkan aplikasi dan laporan yang mendukung pengambilan keputusan strategis. Bandingkan dengan perkiraan biaya gudang data dan temukan rasio antara biaya saat ini dan biaya yang diusulkan. Lihat apakah rasio ini dapat diterima oleh manajemen senior.
2. Hitung nilai bisnis dari gudang data yang diusulkan dengan perkiraan nilai dolar untuk keuntungan, dividen, pertumbuhan pendapatan, pertumbuhan pendapatan, dan pertumbuhan pangsa pasar. Tinjau nilai bisnis ini yang dinyatakan dalam dolar terhadap biaya gudang data dan buatlah pembenarannya.
3. Lakukan latihan secara penuh. Identifikasi semua komponen yang akan terpengaruh oleh data warehouse yang diusulkan dan komponen-komponen yang akan mempengaruhi data warehouse. Mulailah dengan item biaya, satu per satu, termasuk

pembelian atau penyewaan perangkat keras, perangkat lunak vendor, perangkat lunak internal, instalasi dan konversi, dukungan berkelanjutan, dan biaya pemeliharaan. Kemudian berikan nilai dolar pada masing-masing manfaat berwujud dan tidak berwujud, termasuk pengurangan biaya, peningkatan pendapatan, dan efektivitas dalam komunitas bisnis. Melangkah lebih jauh untuk melakukan analisis arus kas dan menghitung ROI.

Rencana Keseluruhan

Benih inisiatif data warehousing disebarakan melalui berbagai cara. Inisiatif ini mungkin terpicu hanya karena pesaing memiliki gudang data.

Perencanaan inisiasi data warehouse



- PERKENALAN
- PERNYATAAN MISI
- CAKUPAN
- TUJUAN & TUJUAN
- MASALAH & PILIHAN UTAMA
- NILAI & HARAPAN
- PEMBENARAN
- SPONSOR EKSEKUTIF
- STRATEGI PELAKSANAAN
- JADWAL TENTATIF
- OTORISASI PROYEK

Gambar 4.1 Rencana keseluruhan inisiatif data warehousing.

Atau CIO memberikan rekomendasi kepada CEO atau beberapa eksekutif senior lainnya mengusulkan data warehouse sebagai solusi permasalahan informasi dalam suatu perusahaan. Dalam beberapa kasus, seorang eksekutif senior mengetahui gagasan tersebut di konferensi atau seminar. Apa pun alasan perusahaan Anda memikirkan data warehousing, inisiatif sebenarnya dimulai dengan rencana formal yang dipikirkan dengan matang yang menetapkan arah, nada, dan tujuan inisiatif. Rencana tersebut menetapkan motivasi dan insentif. Ini mempertimbangkan berbagai pilihan dan alasan dalam proses seleksi. Rencana tersebut membahas jenis gudang data dan menyebutkan harapannya. Ini bukanlah rencana proyek yang terperinci. Ini adalah rencana keseluruhan untuk meletakkan fondasi, untuk mengenali kebutuhan, dan untuk mengesahkan proyek formal.

Gambar 4.1 mencantumkan jenis konten yang akan dimasukkan dalam rencana formal keseluruhan. Tinjau daftar tersebut dengan cermat dan sesuaikan dengan inisiatif gudang data Anda.

4.2 PROYEK GUDANG DATA

Sebagai seorang profesional TI, Anda pernah mengerjakan proyek aplikasi sebelumnya. Anda mengetahui apa yang terjadi dalam proyek ini dan mengetahui metode yang diperlukan untuk membangun aplikasi mulai dari perencanaan hingga implementasi. Anda telah menjadi bagian dari analisis, desain, pemrograman, atau fase pengujian. Jika Anda pernah berfungsi sebagai manajer proyek atau pemimpin tim, Anda pasti tahu bagaimana proyek dipantau dan dikendalikan. Sebuah proyek adalah sebuah proyek. Jika Anda pernah melihat satu proyek TI, apakah Anda belum melihat semuanya?

Jawabannya tidak sederhana ya atau tidak. Proyek gudang data berbeda dengan proyek yang membangun sistem pemrosesan transaksi. Jika Anda baru mengenal data warehousing, proyek data warehouse pertama Anda akan mengungkapkan perbedaan besarnya. Kami akan membahas perbedaan-perbedaan ini dan juga mempertimbangkan cara untuk menyikapinya. Kami juga akan menanyakan pertanyaan mendasar tentang kesiapan departemen TI dan pengguna untuk meluncurkan proyek gudang data. Bagaimana dengan pendekatan siklus hidup pengembangan sistem (SDLC) tradisional? Bisakah kita menggunakan pendekatan ini pada proyek gudang data juga? Jika ya, apa saja fase perkembangan dalam siklus hidup?

Apa Bedanya?

Mari kita pahami mengapa proyek gudang data berbeda. Anda sudah familiar dengan proyek aplikasi untuk sistem OLTP. Perbandingan dengan proyek aplikasi OLTP akan membantu kita mengenali perbedaannya. Cobalah untuk mendeskripsikan gudang data dalam bentuk bagian fungsional utama. Pertama, Anda memiliki komponen akuisisi data. Berikutnya adalah komponen penyimpanan data. Terakhir, ada komponen penyampaian informasi. Pada tingkat yang sangat umum, gudang data terdiri dari tiga komponen besar ini. Anda akan melihat bahwa proyek gudang data berbeda dari proyek pada aplikasi OLTP di masing-masing tiga area fungsional. Mari kita bahas perbedaannya. Gambar 4-2 mencantumkan perbedaannya dan juga menjelaskannya.

AKUISISI DATA	PENYIMPANAN DATA	INFORMASI. PENGIRIMAN
Sejumlah besar sumber	Penyimpanan volume data yang besar	Beberapa tipe pengguna
Banyak sumber yang berbeda	Pertumbuhan yang cepat	Pertanyaan meluas hingga batasnya
Platform komputasi yang berbeda	Kebutuhan untuk pemrosesan paralel	Beberapa jenis kueri
Sumber luar	Penyimpanan data di area pementasan	Mendukung web
Beban awal yang besar	Beberapa jenis indeks	Analisis multidimensi
Umpan data yang sedang berlangsung	Beberapa file indeks	fungsionalitas OLAP
Replikasi data pertimbangan	Penyimpanan tipe data yang lebih baru	Manajemen metadata
Integrasi data yang sulit	Arsip data lama	Antarmuka ke aplikasi DSS.

Transformasi Data yang kompleks	Kompatibilitas dengan alat	Masukan ke Data Mining
Pembersihan data	RDBMS & MDDDBMS	Alat multi-vendor

Gambar 4.2 Perbedaan antara proyek data warehouse dan proyek pada aplikasi OLTP.

Pergudangan data adalah paradigma baru. Kami hampir mengharapkan proyek gudang data berbeda dari proyek sistem OLTP. Kita bisa menerima perbedaan yang ada. Namun yang lebih penting adalah pembahasan mengenai konsekuensi dari perbedaan tersebut. Apa yang harus Anda lakukan terhadap perbedaan tersebut? Bagaimana tahapan proyek diubah dan ditingkatkan untuk menghadapinya? Saran berikut mengatasi perbedaan tersebut:

- ◆ Secara sadar menyadari bahwa proyek data warehouse memiliki cakupan yang lebih luas, cenderung lebih kompleks, dan melibatkan banyak teknologi berbeda.
- ◆ Memberikan waktu dan tenaga ekstra untuk jenis aktivitas baru.
- ◆ Jangan ragu untuk mencari dan mempekerjakan spesialis di mana pun talenta internal tidak tersedia. Proyek gudang data memiliki banyak tugas yang tidak biasa.
- ◆ Metadata dalam gudang data sangat penting sehingga memerlukan perlakuan khusus selama proyek berlangsung. Berikan perhatian ekstra untuk membangun kerangka metadata dengan benar.
- ◆ Biasanya, Anda akan menggunakan beberapa alat pihak ketiga selama pengembangan dan untuk memfungsikan gudang data secara berkelanjutan. Dalam jadwal proyek Anda, rencanakan untuk memasukkan waktu untuk evaluasi dan pemilihan alat.
- ◆ Memberikan waktu yang cukup untuk membangun dan menyelesaikan infrastruktur.
- ◆ Menyiapkan waktu yang cukup untuk desain arsitektur.
- ◆ Melibatkan pengguna dalam setiap tahap proyek. Pergudangan data bisa jadi merupakan hal yang benar-benar baru bagi TI dan pengguna di perusahaan Anda. Upaya bersama sangat penting.
- ◆ Berikan waktu yang cukup untuk melatih pengguna dalam alat kueri dan pelaporan.
- ◆ Karena banyaknya tugas dalam proyek data warehouse, jalur pengembangan paralel mutlak diperlukan. Bersiaplah menghadapi tantangan menjalankan jalur paralel dalam siklus hidup proyek.

Penilaian Kesiapan

Katakanlah Anda telah membenarkan proyek gudang data dan menerima persetujuan dan restu dari manajemen puncak. Anda memiliki rencana menyeluruh untuk inisiatif pergudangan data. Anda telah memahami permasalahan utama dan memahami perbedaan proyek gudang data dan apa yang harus Anda lakukan untuk menangani perbedaan tersebut. Apakah Anda siap untuk langsung mempersiapkan rencana proyek dan bergerak cepat?

Belum. Anda perlu melakukan penilaian kesiapan formal. Biasanya, bagi banyak anggota tim proyek dan hampir semua pengguna, data warehousing merupakan sebuah konsep baru. Penilaian dan orientasi kesiapan merupakan hal yang penting. Siapakah orang yang melakukan penilaian kesiapan? Manajer proyek biasanya melakukannya dengan bantuan ahli dari luar. Pada saat ini, manajer proyek sudah terlatih dalam pergudangan data atau dia

mungkin memiliki pengalaman sebelumnya. Terlibat dalam diskusi dengan sponsor eksekutif, pengguna, dan calon anggota tim. Tujuannya adalah untuk menilai pemahaman mereka terhadap data warehousing secara umum, menilai kesiapan mereka, dan mengungkap kesenjangan dalam pengetahuan mereka. Siapkan laporan penilaian kesiapan formal sebelum rencana proyek ditetapkan.

Laporan penilaian kesiapan diharapkan memiliki tujuan sebagai berikut:

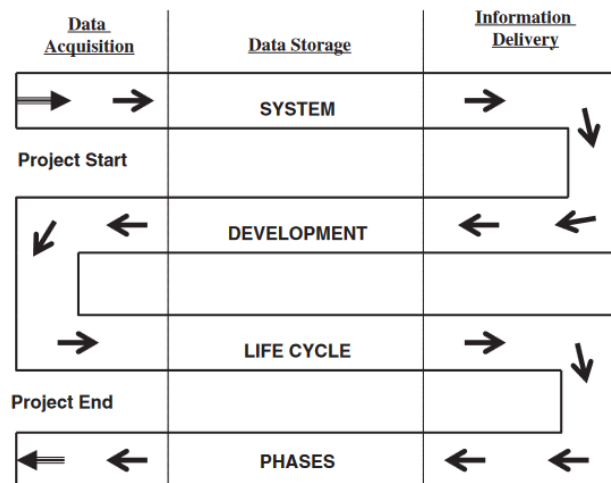
- ❖ Menurunkan risiko terjadinya kejutan besar selama implementasi
- ❖ Memberikan pendekatan proaktif terhadap penyelesaian masalah
- ❖ Menilai kembali komitmen perusahaan
- ❖ Meninjau dan mengidentifikasi kembali ruang lingkup dan ukuran proyek
- ❖ Mengidentifikasi faktor penentu keberhasilan
- ❖ Menyatakan kembali ekspektasi pengguna
- ❖ Pastikan kebutuhan pelatihan

Pendekatan Siklus Hidup

Sebagai seorang profesional TI, Anda pasti sudah familiar dengan siklus hidup pengembangan sistem (SDLC) tradisional. Anda tahu bagaimana memulai dengan rencana proyek, beralih ke tahap analisis persyaratan, kemudian ke tahap desain, konstruksi, dan pengujian, dan akhirnya ke tahap implementasi. Pendekatan siklus hidup menyelesaikan semua tujuan utama dalam proses pengembangan sistem. Ini menegakkan ketertiban dan memungkinkan pendekatan sistematis untuk membangun sistem komputer. Metodologi siklus hidup memecah kompleksitas proyek dan menghilangkan segala ambiguitas sehubungan dengan tanggung jawab anggota tim proyek. Ini menyiratkan serangkaian tugas dan hasil yang dapat diprediksi.

Bahwa pendekatan siklus hidup memecah kompleksitas proyek adalah alasan yang cukup untuk mempertimbangkan pendekatan ini untuk proyek gudang data. Proyek gudang data rumit dalam hal tugas, teknologi, dan peran anggota tim. Namun pendekatan siklus hidup satu ukuran untuk semua tidak akan berhasil untuk proyek gudang data. Sesuaikan pendekatan siklus hidup dengan kebutuhan khusus proyek gudang data Anda. Perlu diperhatikan bahwa siklus hidup pengembangan data warehouse bukanlah metode air terjun yang satu fase berakhir dan mengalir ke fase berikutnya.

Pendekatan untuk proyek gudang data harus mencakup tugas-tugas berulang yang melalui siklus penyempurnaan. Misalnya, jika salah satu tugas Anda dalam proyek adalah mengidentifikasi sumber data, Anda dapat memulai dengan meninjau semua sistem sumber dan membuat daftar semua struktur data sumber. Iterasi tugas berikutnya dimaksudkan untuk meninjau elemen data dengan pengguna. Anda melanjutkan ke iterasi berikutnya dalam meninjau elemen data dengan administrator database dan beberapa staf TI lainnya. Iterasi berikutnya dengan menelusuri elemen data sekali lagi menyelesaikan penyempurnaan dan tugas. Jenis proses berulang ini diperlukan untuk setiap tugas karena kompleksitas dan luasnya cakupan proyek.



Gambar 4.3 Komponen fungsional gudang tanggal dan SDLC.

Ingatlah bahwa komponen fungsional luas dari gudang data adalah akuisisi data, penyimpanan data, dan pengiriman informasi. Pastikan fase siklus hidup pengembangan Anda mencakup komponen fungsional ini. Gambar 4.3 menunjukkan bagaimana menghubungkan komponen fungsional ke SDLC.

Seperti dalam siklus hidup pengembangan sistem apa pun, proyek gudang data dimulai dengan persiapan rencana proyek. Rencana proyek menggambarkan proyek, mengidentifikasi tujuan spesifik, menyebutkan faktor penentu keberhasilan, membuat daftar asumsi, dan menyoroti isu-isu penting. Rencana tersebut mencakup jadwal proyek, daftar tugas dan penugasan, dan menyediakan pemantauan kemajuan. Gambar 4.4 memberikan contoh garis besar rencana proyek gudang data.

Rencana Proyek Gudang Data.

- PERKENALAN
- TUJUAN
- PENILAIAN KESIAPAN
- TUJUAN & TUJUAN
- PEMANGKU KEPENTINGAN
- ASUMSI
- MASALAH KRITIS
- FAKTOR KEBERHASILAN
- TIM PROYEK
- JADWAL PROYEK
- RINCIAN PENERAPAN

Gambar 4.4 Contoh garis besar rencana proyek gudang data.

4.3 TAHAP PERKEMBANGAN

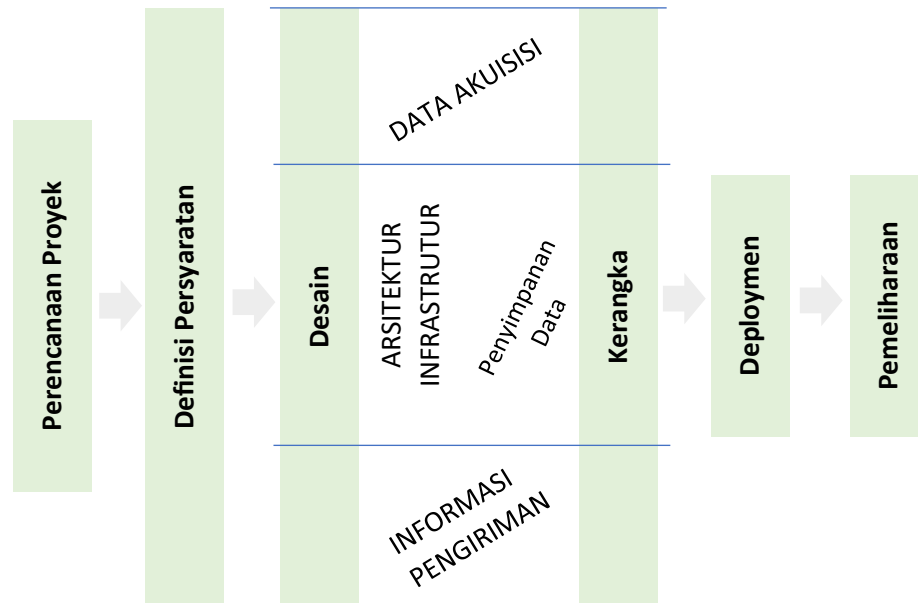
Pada bagian sebelumnya, kita kembali mengacu pada keseluruhan komponen fungsional data warehouse sebagai akuisisi data, penyimpanan data, dan pengiriman informasi. Ketiga komponen fungsional ini membentuk arsitektur umum data warehouse. Harus ada infrastruktur teknis yang memadai untuk mendukung ketiga komponen fungsional tersebut. Oleh karena itu, ketika kita merumuskan fase-fase pengembangan dalam siklus hidup, kita harus memastikan bahwa fase-fase tersebut mencakup tugas-tugas yang berkaitan dengan ketiga komponen tersebut. Fase-fase tersebut juga harus mencakup tugas-tugas untuk mendefinisikan arsitektur yang terdiri dari tiga komponen dan untuk membangun infrastruktur yang mendasari untuk mendukung arsitektur. Tahap desain dan konstruksi untuk ketiga komponen ini mungkin berjalan secara paralel.

Lihat Gambar 4.5 dan perhatikan tiga jalur fase pengembangan. Dalam pengembangan setiap gudang data, jalur ini hadir dengan serangkaian tugas yang berbeda-beda. Anda dapat mengubah dan menyesuaikan tugas agar sesuai dengan kebutuhan spesifik Anda. Anda mungkin ingin menekankan satu lagu lebih banyak daripada yang lain. Jika kualitas data menjadi masalah di perusahaan Anda, Anda perlu memberikan perhatian khusus pada fase terkait. Gambar tersebut menunjukkan pembagian luas siklus hidup proyek ke dalam fase-fase tradisional:

- Rencana proyek
- Definisi persyaratan
- Desain
- Konstruksi
- Penyebaran
- Pertumbuhan dan pemeliharaan

Terjalin dalam fase desain dan konstruksi adalah tiga jalur beserta definisi arsitektur dan pembangunan infrastruktur. Masing-masing kotak yang ditunjukkan dalam diagram mewakili aktivitas utama yang akan dipecah lebih lanjut menjadi tugas-tugas individual dan ditugaskan kepada anggota tim yang sesuai. Gunakan diagram sebagai panduan untuk membuat daftar aktivitas dan tugas proyek gudang data Anda. Meskipun aktivitas utama mungkin tetap sama untuk sebagian besar gudang, tugas individual dalam setiap aktivitas cenderung bervariasi untuk gudang data spesifik Anda.

Pada bab-bab berikutnya, kita akan membahas kegiatan-kegiatan pembangunan ini secara lebih rinci. Ketika Anda sampai pada bab-bab tersebut, Anda mungkin ingin merujuk kembali ke diagram ini.



Gambar 4.5 Tahapan pengembangan gudang data.

Mengadopsi Pengembangan Agile

Pada bab sebelumnya kami memperkenalkan penggunaan prinsip dan praktik pengembangan tangkas untuk pengembangan perangkat lunak. Ini adalah fenomena terkini, lebih baru dalam pengembangan gudang data dibandingkan jenis aplikasi lainnya. Perusahaan mulai mengadopsi pengembangan tangkas untuk proyek gudang data dan intelijen bisnis. Metodologi pengembangan yang tangkas mendorong kolaborasi di antara anggota tim proyek dan pengguna, serta mendorong upaya pengembangan yang berulang dan bertahap. Keseluruhan metodologi bertumpu pada penerapan nilai-nilai inti, prinsip, dan praktik tertentu dalam setiap aspek pengembangan dan implementasi data warehouse.

Nilai Inti. Nilai-nilai yang ditekankan antara lain mengupayakan kesederhanaan dan tidak terjebak dalam kompleksitas, memberikan dan memperoleh umpan balik secara terus-menerus terhadap tugas-tugas perkembangan individu, membina komunikasi yang bebas dan tanpa hambatan, dan menghargai keberanian untuk belajar dari kesalahan. **Prinsip Inti.** Prinsip-prinsip yang diikuti mencakup mendorong kualitas, menerima perubahan, melakukan perubahan secara bertahap, menerapkan kesederhanaan, dan memberikan umpan balik yang cepat.

Praktik Inti. Praktik yang diterapkan termasuk membuat rilis singkat komponen aplikasi, melakukan tugas pengembangan bersama (*"pemrograman berpasangan"*), bekerja secara intensif selama 40 jam seminggu, tidak menambah waktu untuk hal-hal yang tidak efektif, dan memiliki perwakilan pengguna di lokasi dengan tim proyek.

Variabel. Variabel kontrol yang dapat dimanipulasi untuk trade-off guna mencapai hasil adalah waktu, kualitas, ruang lingkup, dan biaya.

4.4 TIM PROYEK

Seperti halnya jenis proyek apa pun, keberhasilan proyek gudang data berada di pundak tim proyek. Tim terbaik menang. Proyek gudang data mirip dengan proyek perangkat

lunak lainnya karena memerlukan banyak manusia. Dibutuhkan beberapa orang yang terlatih dan berketerampilan khusus untuk membentuk tim proyek. Mengorganisir tim proyek untuk proyek gudang data berkaitan dengan mencocokkan beragam peran dengan keterampilan dan tingkat pengalaman yang tepat. Hal tersebut bukanlah hal yang mudah untuk dicapai.

Ada dua hal yang dapat merusak sebuah proyek: kompleksitas yang berlebihan dan ambiguitas tanggung jawab. Dalam pendekatan siklus hidup, tim proyek meminimalkan kompleksitas upaya dengan berbagi dan melakukan. Ketika orang yang tepat dalam tim, dengan jenis keterampilan yang tepat dan tingkat pengalaman yang tepat, melakukan tugas individu, orang tersebut benar-benar menyelesaikan masalah kompleksitas. Dalam tim proyek yang dibentuk dengan baik, setiap orang diberi tanggung jawab khusus atau peran tertentu berdasarkan tingkat keahlian dan pengalamannya. Dalam tim seperti itu, tidak ada kebingungan atau ambiguitas mengenai tanggung jawab.

Pada bagian berikut, kita akan membahas penyesuaian anggota tim ke dalam peran yang sesuai. Kami juga akan membahas tanggung jawab yang terkait dengan peran tersebut. Selanjutnya, kita akan membahas tingkat keterampilan dan pengalaman yang dibutuhkan untuk masing-masing peran tersebut. Perhatikan baik-baik dan pelajari cara menentukan peran proyek untuk gudang data Anda. Selain itu, cobalah mencocokkan peran proyek Anda dengan tanggung jawab dan tugas dalam proyek gudang data Anda.

Mengorganisir Tim Proyek

Mengorganisir tim proyek melibatkan penempatan orang yang tepat pada pekerjaan yang tepat. Jika Anda mengorganisir dan menyusun tim untuk mengerjakan pengembangan sistem OLTP, Anda tahu bahwa keahlian yang diperlukan cukup besar dan dapat dikelola. Anda memerlukan keterampilan khusus di bidang manajemen proyek, analisis persyaratan, desain aplikasi, desain basis data, dan pengujian aplikasi. Namun proyek gudang data memerlukan banyak peran lainnya. Lalu bagaimana Anda mengisi semua peran yang bervariasi ini?

Titik awal yang baik adalah dengan membuat daftar semua tantangan proyek dan keterampilan khusus yang diperlukan. Daftar Anda mungkin berjalan seperti ini: perencanaan, penentuan persyaratan data, penentuan tipe kueri, pemodelan data, pemilihan alat, desain basis data fisik, ekstraksi data sumber, validasi data dan kendali mutu, pengaturan kerangka metadata, dan sebagainya. Sebagai langkah berikutnya, dengan menggunakan daftar keterampilan dan tantangan yang Anda antisipasi, siapkan daftar peran tim yang diperlukan untuk mendukung pekerjaan pengembangan.

Setelah Anda memiliki daftar peran, Anda siap untuk menugaskan individu ke peran tim. Tidak perlu menugaskan satu orang atau lebih untuk masing-masing peran yang teridentifikasi. Jika upaya gudang data Anda tidak besar dan sumber daya perusahaan Anda sedikit, cobalah membuat orang yang sama mempunyai banyak jabatan. Dalam proses alokasi personel ini, perlu diingat bahwa perwakilan pengguna juga harus dianggap sebagai anggota tim proyek. Jangan gagal untuk mengenali pengguna sebagai bagian dari tim dan menugaskan mereka ke peran yang sesuai.

Keterampilan, pengalaman, dan pengetahuan penting bagi anggota tim. Namun, sikap, semangat tim, semangat dalam upaya data warehouse, dan komitmen yang kuat sama pentingnya, atau bahkan lebih penting. Jangan lupa untuk mencari ciri-ciri penting ini.

DEVELOPER PEMELIHARAAN

Sponsor eksekutif	Spesialis Penyediaan Data
Direktur Proyek	Analisis Bisnis
Manajer proyek	Administrator sistem
Manajer Perwakilan Pengguna	Spesialis Migrasi Data
Administrator Gudang Data	Spesialis Perawatan Data
Manajer Perubahan Organisasi	Pemimpin Data Mart
Administrator Basis Data	Spesialis Infrastruktur
Manajer Metadata	Pengguna Listrik
Analisis Persyaratan Bisnis	Pemimpin Pelatihan
Arsitek Gudang Data	Penulis teknis
Pengembang Akuisisi Data	Spesialis Alat
Pengembang Akses Data	Spesialis Hubungan Vendor
Analisis Kualitas Data	Ahli Web
Penguji Gudang Data	Pemodel Data
Pengembang Pemeliharaan	Arsitek Keamanan

Gambar 4.6 Judul pekerjaan dalam proyek gudang data.

Peran dan Tanggung Jawab

Peran tim proyek ditetapkan untuk melakukan satu atau lebih tugas terkait. Dalam banyak proyek gudang data, peran tim identik dengan jabatan yang diberikan kepada anggota tim. Jika Anda meninjau proyek pengembangan sistem OLTP, Anda akan menemukan bahwa jabatan untuk anggota tim kurang lebih terstandarisasi. Dalam proyek sistem OLTP, Anda akan menemukan jabatan manajer proyek, analisis bisnis, analisis sistem, pemrogram, analisis data, administrator basis data, dan sebagainya. Namun, proyek gudang data belum terstandarisasi dalam hal ubin pekerjaan. Masih ada unsur eksperimen dan eksplorasi.

Jadi apa saja jabatan yang umum? Pertama-tama mari kita lihat daftar panjang yang ditunjukkan pada Gambar 4.6. Jangan khawatir dengan panjangnya daftarnya. Kecuali gudang data Anda berukuran besar, Anda tidak memerlukan semua jabatan ini. Daftar ini hanya menunjukkan kemungkinan dan variasinya. Tanggung jawab dengan peran yang sama mungkin melekat pada jabatan yang berbeda dalam proyek yang berbeda. Dalam banyak proyek, anggota tim yang sama akan memenuhi tanggung jawab lebih dari satu peran.

Penulis dan praktisi data warehousing cenderung mengklasifikasikan peran atau jabatan dalam berbagai cara. Mereka pertama-tama membuat klasifikasi luas dan kemudian memasukkan jabatan individu ke dalam klasifikasi ini. Berikut beberapa klasifikasi peran:

1. Penempatan staf untuk pengembangan awal, penempatan staf untuk pengujian, penempatan staf untuk pemeliharaan berkelanjutan, penempatan staf untuk manajemen gudang data

2. TI dan pengguna akhir, kemudian subklasifikasi dalam masing-masing dua klasifikasi besar, diikuti dengan subklasifikasi lebih lanjut
3. Peran kantor depan, peran kantor belakang
4. Pelatih, susunan pemain reguler, tim khusus
5. Manajemen, pengembangan, dukungan
6. Administrasi, perolehan data, penyimpanan data, penyampaian informasi

Dalam proyek gudang data, Anda mungkin ingin membuat klasifikasi luas yang paling sesuai untuk lingkungan Anda. Bagaimana Anda membuat klasifikasi luas? Anda harus memeriksa kembali tujuan dan sasaran. Anda harus menilai bidang-bidang dalam tahap pengembangan yang memerlukan perhatian khusus. Apakah ekstraksi data akan menjadi tantangan terbesar Anda? Kemudian dukung fungsi tersebut dengan peran khusus. Apakah fungsi penyampaian informasi Anda akan menjadi rumit? Kemudian miliki peran tim proyek khusus yang kuat dalam penyampaian informasi. Setelah Anda menentukan klasifikasi luas, kerjakan peran individual dalam setiap klasifikasi. Jika ini adalah proyek gudang data pertama Anda, Anda mungkin tidak mendapatkan semua peran yang diperlukan sebelumnya. Jangan terlalu khawatir. Anda dapat terus mendukung proyek dengan peran tim tambahan di sana-sini seiring berjalannya proyek.

Anda telah membaca daftar panjang kemungkinan peran tim dan cara klasifikasi peran tersebut. Ini mungkin proyek gudang data pertama Anda dan Anda mungkin bertanggung jawab menentukan peran tim untuk proyek tersebut. Anda ingin memulai dan memiliki pertanyaan mendasar: Apakah ada serangkaian peran dasar standar untuk menjalankan proyek? Tidak terlalu. Tidak ada standar yang ditetapkan seperti itu. Jika Anda cenderung mengikuti metodologi tradisional, ikuti klasifikasi manajemen, pengembangan, dan dukungan. Jika Anda ingin menemukan kekuatan untuk tiga area fungsional utama, maka gunakanlah klasifikasi akuisisi data, penyimpanan data, dan penyampaian informasi. Anda mungkin juga menemukan bahwa kombinasi dua cara pengklasifikasian ini akan berhasil untuk gudang data Anda.

Meskipun tidak ada seperangkat peran tim yang standar, kami menyarankan perangkat dasar berikut:

- Sponsor eksekutif
- Manajer proyek
- Manajer penghubung pengguna
- Arsitek utama
- Spesialis infrastruktur
- Analis bisnis
- Pemodel data
- Administrator gudang data
- Spesialis transformasi data
- Analis penjaminan mutu
- Koordinator pengujian
- Spesialis aplikasi pengguna akhir

- Pemrogram pengembangan
- Pelatih utama

Gambar 4.7 mencantumkan tanggung jawab yang biasa melekat pada rangkaian peran yang disarankan. Tinjau deskripsi tanggung jawab. Tambahkan atau ubah deskripsi agar dapat diterapkan pada keadaan khusus gudang data Anda.

Tingkat Keterampilan dan Pengalaman

Kami membahas pedoman untuk menentukan klasifikasi luas peran tim. Setelah Anda mengetahui klasifikasi yang relevan dengan proyek gudang data Anda, Anda akan menemukan serangkaian peran tim yang sesuai dengan situasi Anda. Kami meninjau beberapa contoh peran umum. Peran tersebut juga dapat disebut jabatan dalam suatu proyek. Ke depannya, Anda akan menuliskan tanggung jawab yang terkait dengan peran yang telah Anda tetapkan. Anda telah menetapkan peran dan Anda telah membuat daftar tanggung jawab. Apakah Anda siap untuk mencocokkan orang-orang untuk mengisi peran ini? Ada satu langkah lagi yang diperlukan sebelum Anda dapat melakukan itu.

Untuk menyesuaikan diri dengan peran dan melaksanakan tanggung jawab, orang-orang yang dipilih harus memiliki kemampuan yang tepat. Mereka harus memiliki keterampilan yang sesuai dan pengalaman kerja yang tepat. Jadi, Anda harus membuat daftar keterampilan dan pengalaman yang diperlukan untuk berbagai peran. Gambar 4-8 menggambarkan tingkat keterampilan dan pengalaman untuk sampel peran tim kami. Gunakan deskripsi yang terdapat pada gambar sebagai contoh untuk menyusun deskripsi peran tim dalam proyek gudang data Anda.

Tidak mudah menemukan profesional TI untuk mengisi semua peran yang ditetapkan untuk gudang data Anda. Sistem OLTP ada di mana-mana. Semua profesional TI telah mengambil peran tertentu dalam proyek sistem OLTP. Hal ini tidak terjadi pada proyek gudang data. Tidak banyak profesional yang memiliki pengalaman langsung dalam pengembangan gudang data. Keterampilan dan kemampuan yang luar biasa tidak banyak tersedia.

Jika orang yang memenuhi syarat untuk bekerja pada proyek data warehouse tidak tersedia, apa solusi Anda? Bagaimana Anda bisa mengisi peran dalam proyek Anda? Di sinilah pelatihan menjadi penting. Melatih para profesional yang sesuai dalam konsep dan teknik pergudangan data. Biarkan mereka mempelajari dasar-dasarnya dan berspesialisasi untuk peran tertentu. Selain melatih personel internal Anda, gunakan konsultan eksternal dalam peran tertentu yang tidak dapat Anda temukan orangnya dari dalam. Namun, sebagai aturan umum, konsultan tidak boleh digunakan dalam peran utama. Manajer proyek atau administrator utama harus berasal dari dalam organisasi.

Sponsor eksekutif

Arahan, dukungan, arbitrase.

Manajer proyek

Penugasan, pemantauan, pengendalian.

Manajer Penghubung Pengguna

Koordinasi dengan kelompok pengguna.

Arsitek Utama

Administrator Gudang Data

fungsi DBA.

Spesialis Transformasi Data

Ekstraksi data, integrasi, transformasi.

Analisis Penjaminan Mutu

Kontrol kualitas untuk data gudang.

Koordinator Pengujian

Desain arsitektur.	Pengujian program, sistem, alat.
Spesialis Infrastruktur	Spesialis Aplikasi Pengguna Akhir
Desain/konstruksi infrastruktur.	Konfirmasi makna/hubungan data.
Analisis Bisnis	Pemrogram Pengembangan
Definisi persyaratan.	Program dan skrip internal.
Pemodelan Data	Pelatih Utama
Pemodelan relasional dan dimensi.	Koordinasi pelatihan Pengguna dan Tim.

Gambar 4.7 Peran dan tanggung jawab tim proyek gudang data.

Partisipasi Pengguna

Dalam aplikasi OLTP pada umumnya, pengguna berinteraksi dengan sistem melalui layar GUI. Mereka menggunakan layar untuk memasukkan data dan mengambil informasi. Pengguna menerima informasi tambahan melalui laporan yang dihasilkan oleh sistem secara berkala. Jika pengguna memerlukan laporan khusus, mereka harus melibatkan TI untuk menulis program ad hoc yang bukan bagian dari aplikasi biasa.

Sebaliknya, interaksi pengguna dengan gudang data bersifat langsung dan intim. Biasanya, tidak ada atau hanya ada beberapa laporan atau kueri. Ketika implementasi selesai, pengguna Anda akan mulai menggunakan gudang data secara langsung tanpa mediasi dari TI. Tidak ada kepastian mengenai jenis kueri yang akan dijalankan, jenis laporan yang akan diminta, atau jenis analisis yang akan dilakukan. Perbedaan utama antara sistem OLTP dan sistem data warehousing adalah penggunaan sistem oleh pengguna.

Apa implikasi dari perbedaan besar dalam komposisi tim proyek dan pengembangan data warehouse? Implikasinya sangat penting. Apa artinya ini? Artinya jika pengguna akan menggunakan data warehouse secara langsung dengan cara yang tidak terduga, mereka harus mempunyai suara yang kuat dalam pengembangannya. Mereka harus menjadi bagian dari tim proyek sepenuhnya. Lebih dari sekedar proyek sistem OLTP, proyek gudang data memerlukan teknik pengembangan aplikasi gabungan (JAD) yang serius.

Proyek gudang data Anda akan berhasil hanya jika anggota komunitas pengguna yang sesuai diterima sebagai anggota tim dengan peran tertentu. Manfaatkan keahlian dan pengetahuan mereka tentang bisnis. Manfaatkan pengalaman mereka dalam membuat keputusan bisnis. Libatkan mereka secara aktif dalam pemilihan alat penyampaian informasi. Carilah bantuan mereka dalam menguji sistem sebelum implementasi.

Sponsor eksekutif

Eksekutif tingkat senior, pengetahuan mendalam tentang bisnis, antusiasme dan kemampuan untuk memoderasi dan melakukan arbitrase jika diperlukan.

Manajer proyek

Keterampilan orang, pengalaman manajemen proyek, berorientasi bisnis dan pengguna, kemampuan untuk menjadi praktis dan efektif.

Manajer Penghubung Pengguna

Keterampilan manusia, dihormati dalam komunitas pengguna, keterampilan organisasi, pemain tim, pengetahuan tentang sistem dari sudut pandang pengguna.

<p>Arsitek Utama Keterampilan analitis, kemampuan melihat gambaran besar, keahlian dalam antarmuka, pengetahuan tentang konsep data warehouse.</p> <p>Spesialis Infrastruktur Spesialis dalam perangkat keras, sistem operasi, platform komputasi, pengalaman sebagai staf operasi.</p> <p>Analisis Bisnis Keterampilan analitis, kemampuan berinteraksi dengan pengguna, pengalaman industri yang memadai sebagai analis.</p> <p>Pemodel Data Keahlian dalam pemodelan relasional dan dimensi dengan alat kasus, pengalaman sebagai analis data.</p> <p>Administrator Gudang Data Ahli dalam desain dan implementasi basis data fisik, pengalaman sebagai DBA relasional, pengalaman MDDDBMS merupakan nilai tambah.</p> <p>Spesialis Transformasi Data Pengetahuan tentang struktur data, pengetahuan mendalam tentang sistem sumber, pengalaman sebagai analis.</p> <p>Analisis Penjaminan Mutu Pengetahuan tentang teknik kualitas data, pengetahuan tentang sistem sumber data, pengalaman sebagai analis.</p> <p>Koordinator Pengujian Keakraban dengan metode dan standar pengujian, penggunaan alat pengujian, pengetahuan tentang beberapa alat penyampaian informasi gudang data, pengalaman sebagai programmer/analisis.</p> <p>Spesialis Aplikasi Pengguna Akhir Pengetahuan mendalam tentang aplikasi sumber.</p> <p>Pemrogram Pengembangan Keterampilan pemrograman dan analisis, pengalaman sebagai programmer dalam bahasa tertentu dan DBMS.</p> <p>Pelatih Utama Keterampilan pelatihan, pengalaman dalam pelatihan TI/Pengguna, keterampilan koordinasi dan organisasi.</p>

Gambar 4.8 Tingkat keterampilan dan pengalaman yang diperlukan untuk tim proyek gudang data.

Perencanaan proyek

- Memberikan tujuan, sasaran, harapan, informasi bisnis selama survei pendahuluan; memberikan dukungan manajemen puncak yang aktif; memulai proyek sebagai sponsor eksekutif.

Definisi Persyaratan

- Berpartisipasi aktif dalam pertemuan untuk menentukan persyaratan; mengidentifikasi semua sistem sumber; menentukan metrik untuk mengukur keberhasilan bisnis, dan dimensi bisnis untuk analisis; mendefinisikan informasi yang dibutuhkan dari gudang data.

Desain

- Tinjau model data dimensi, ekstraksi data dan desain transformasi; memberikan antisipasi penggunaan untuk ukuran basis data; meninjau desain arsitektur dan metadata; berpartisipasi dalam pemilihan alat; meninjau desain penyampaian informasi.

Konstruksi

- Berpartisipasi aktif dalam pengujian penerimaan pengguna; menguji alat penyampaian informasi; memvalidasi fungsi ekstraksi dan transformasi data; konfirmasi kualitas data; menguji penggunaan metadata; fungsi kueri patokan; menguji fungsi OLAP; berpartisipasi dalam dokumentasi aplikasi.

Penyebaran

- Verifikasi jejak audit dan konfirmasi pemuatan data awal; mencocokkan hasil dengan harapan yang dinyatakan; mengatur dan berpartisipasi dalam pelatihan pengguna; memberikan penerimaan akhir.

Pemeliharaan

- Memberikan masukan untuk perbaikan; menguji dan menerima penyempurnaan.

Gambar 4.9 Partisipasi pengguna dalam pengembangan data warehouse.

Gambar 4.9 mengilustrasikan bagaimana dan di mana dalam proses pengembangan, pengguna harus dilibatkan. Tinjau setiap fase pengembangan dan putuskan dengan jelas bagaimana dan di mana pengguna Anda perlu berpartisipasi. Angka ini menghubungkan partisipasi pengguna dengan tahapan dalam proses pengembangan. Berikut adalah daftar beberapa peran tim yang dapat diambil oleh pengguna untuk berpartisipasi dalam pengembangan:

- a) Sponsor proyek; bertanggung jawab untuk mendukung upaya proyek secara menyeluruh (harus seorang eksekutif)
- b) Perwakilan penghubung departemen pengguna; membantu TI mengoordinasikan pertemuan dan sesi peninjauan serta memastikan partisipasi aktif departemen pengguna
- c) Pakar bidang studi; memberikan panduan mengenai kebutuhan pengguna dalam bidang studi tertentu dan memperjelas makna semantik istilah bisnis yang digunakan dalam perusahaan
- d) Spesialis peninjauan data; meninjau model data yang disiapkan oleh TI; mengkonfirmasi elemen data dan hubungan data
- e) Konsultan penyampaian informasi; memeriksa dan menguji alat penyampaian informasi; membantu dalam pemilihan alat
- f) Teknisi dukungan pengguna; bertindak sebagai dukungan tingkat pertama dan garis depan bagi pengguna di departemennya masing-masing

4.5 PERTIMBANGAN MANAJEMEN PROYEK

Tim proyek Anda telah terorganisir, fase pengembangan telah selesai, pengujian telah selesai, gudang data telah disebar, dan proyek dinyatakan selesai tepat waktu dan sesuai anggaran. Apakah usahanya berhasil?



Data Basement

Data berkualitas buruk tanpa akses yang tepat



Gudang Data

Dump data yang menyedihkan runtuh bahkan sebelum selesai



Makam Data

Gudang data yang mahal dengan akses dan kinerja yang buruk



Gudang Data

Data Mart pulau yang berdiri, menyendiri, terfragmentasi



Rumah Petak Data

Dibangun oleh vendor sistem lama atau konsultan bodoh yang tidak tahu apa yang diinginkan pengguna



Penjara Data

Sistem data yang terbatas dan tidak terlihat menjaga data tetap terkurung sehingga pengguna tidak dapat memperoleh data tersebut

Gambar 4.10 Kemungkinan skenario kegagalan.

Terlepas dari niat terbaik tim proyek, kemungkinan besar gudang data yang diterapkan ternyata bukan gudang data. Gambar 4.10 menunjukkan kemungkinan skenario kegagalan. Bagaimana hasil gudang data Anda pada akhirnya?

Manajemen proyek yang efektif sangat penting untuk keberhasilan proyek gudang data. Pada bagian ini, kita akan mempertimbangkan isu-isu manajemen proyek yang berlaku khususnya pada proyek gudang data, meninjau beberapa prinsip dasar manajemen proyek, dan membuat daftar faktor-faktor keberhasilan yang mungkin. Kami akan meninjau proyek sukses dalam kehidupan nyata dan memeriksa alasan keberhasilannya. Ketika semuanya sudah dikatakan dan dilakukan, Anda tidak selalu dapat menjalankan proyek Anda sepenuhnya sesuai aturan. Gunakan pendekatan praktis yang membuahkan hasil tanpa terjebak dalam pekerjaan membosankan yang tidak perlu.

Prinsip Panduan

Setelah mengerjakan proyek sistem OLTP, Anda sudah mengetahui beberapa prinsip panduan manajemen proyek jangan menyerah pada kelumpuhan analisis, jangan biarkan ruang lingkup merayap, pantau slippage, jaga agar proyek tetap pada jalurnya, dan sebagainya. Meskipun sebagian besar prinsip panduan tersebut juga berlaku pada manajemen

proyek gudang data, kami tidak ingin mengulanginya di sini. Di sisi lain, kami ingin mempertimbangkan beberapa prinsip panduan yang berkaitan dengan proyek data warehouse secara eksklusif. Pada setiap tahapan proyek, Anda harus menjaga prinsip-prinsip panduan sebagai latar belakang sehingga prinsip-prinsip ini dapat mengkondisikan setiap keputusan dan tindakan manajemen proyek. Prinsip panduan utama adalah:

- a) **Sponsor:** Tidak ada proyek gudang data yang berhasil tanpa sponsor eksekutif yang kuat dan berkomitmen.
- b) **Manajer proyek:** Merupakan kesalahan serius untuk memiliki manajer proyek yang lebih berorientasi pada teknologi daripada berorientasi pada pengguna dan berorientasi bisnis.
- c) **Paradigma Baru:** Pergudangan data adalah hal baru bagi sebagian besar perusahaan; metode manajemen proyek yang inovatif sangat penting untuk menghadapi tantangan yang tidak terduga.
- d) **Peran Tim:** Peran tim tidak boleh diberikan secara sembarangan; perannya harus mencerminkan kebutuhan setiap proyek gudang data.
- e) **Kualitas data:** Tiga aspek penting dari data dalam data warehouse adalah: kualitas, kualitas, dan kualitas.
- f) **Persyaratan Pengguna:** Meskipun jelas, kebutuhan pengguna sendirilah yang menjadi kekuatan pendorong setiap tugas pada jadwal proyek.
- g) **Membangun untuk Pertumbuhan:** Jumlah pengguna dan jumlah pertanyaan meningkat sangat cepat setelah penerapan; gudang data yang tidak dibangun untuk pertumbuhan akan runtuh dengan cepat.
- h) **Politik Proyek:** Proyek gudang data pertama di sebuah perusahaan menimbulkan tantangan dan ancaman bagi pengguna di berbagai tingkatan; mencoba menangani politik proyek adalah seperti berjalan di atas tali, yang harus dilakukan dengan sangat hati-hati.
- i) **Harapan yang Realistis:** Sangat mudah untuk menjanjikan dunia dalam proyek gudang data pertama; menetapkan ekspektasi pada tingkat yang tepat dan dapat dicapai adalah jalan terbaik.
- j) **Pemodelan Data Dimensi:** Model data dimensi yang dirancang dengan baik merupakan landasan dan cetak biru yang diperlukan.
- k) **Data Eksternal:** Gudang data tidak hanya bergantung pada data internal saja; data dari sumber eksternal yang relevan merupakan unsur yang mutlak diperlukan.
- l) **Pelatihan:** Alat pengguna gudang data berbeda dan baru. Jika pengguna tidak mengetahui cara menggunakan alat tersebut, mereka tidak akan menggunakan gudang data. Gudang data yang tidak digunakan adalah gudang data yang gagal.

Tanda peringatan

Saat siklus hidup proyek gudang data Anda berjalan dan fase pengembangan terus berjalan, Anda harus terus mencermati tanda-tanda peringatan yang mungkin menyebabkan bencana. Teruslah mencari indikator apa pun yang menunjukkan malapetaka dan kegagalan. Beberapa tanda peringatan mungkin hanya menunjukkan ketidaknyamanan yang

memerlukan sedikit tindakan. Namun mungkin ada tanda-tanda peringatan lain yang menunjukkan adanya masalah yang lebih luas yang memerlukan tindakan perbaikan untuk memastikan keberhasilan akhir.

Beberapa tanda peringatan mungkin menandakan kelemahan serius yang memerlukan tindakan perbaikan segera. Apa pun sifat dari tanda peringatannya, tetaplah waspada dan awasi dengan cermat. Segera setelah Anda melihat suatu pertanda, kenali potensi masalahnya, dan ambil tindakan perbaikan. Gambar 4.11 menyajikan daftar tanda peringatan umum dan tindakan perbaikan yang disarankan. Daftar pada gambar hanyalah kumpulan contoh. Dalam proyek gudang data Anda, Anda mungkin menemukan jenis tanda peringatan lainnya. Tindakan korektif Anda terhadap potensi masalah mungkin berbeda tergantung pada keadaan Anda.

Faktor Keberhasilan

Anda telah mengikuti prinsip manajemen proyek yang efektif dan gudang data Anda selesai. Bagaimana Anda tahu bahwa gudang data Anda sukses? Apakah Anda memerlukan tiga atau lima tahun untuk melihat apakah Anda mendapatkan ROI (laba atas investasi) yang diusulkan dalam rencana Anda? Berapa lama Anda harus menunggu sebelum Anda dapat menyatakan bahwa upaya gudang data Anda berhasil. Atau, adakah tanda-tanda langsung yang menunjukkan keberhasilan?

Ada beberapa indikasi keberhasilan yang dapat diamati dalam waktu singkat setelah penerapan. Kejadian-kejadian berikut umumnya menunjukkan keberhasilan:

- ❖ Kueri dan laporan: peningkatan pesat dalam jumlah kueri dan laporan yang diminta oleh pengguna langsung dari gudang data
- ❖ Jenis kueri: kueri menjadi lebih canggih
- ❖ Pengguna aktif: peningkatan jumlah pengguna secara terus-menerus
- ❖ Penggunaan: pengguna menghabiskan lebih banyak waktu di gudang data untuk mencari solusi
- ❖ Waktu penyelesaian: penurunan nyata dalam waktu yang diperlukan untuk memperoleh informasi strategis

TANDA PERINGATAN	INDIKASI	TINDAKAN
Fase Definisi Persyaratan telah melewati tanggal target.	Menderita "kelumpuhan analisis".	Hentikan pengambilan informasi yang tidak diinginkan. Hilangkan masalah apa pun dengan bertemu dengan pengguna. Tetapkan tanggal target akhir yang pasti.
Perlu menulis terlalu banyak program internal.	Alat pihak ketiga yang dipilih kehabisan tenaga.	Jika ada waktu dan anggaran, dapatkan alat yang berbeda. Jika tidak, tingkatkan staf pemrograman.
Pengguna tidak mau bekerja sama untuk memberikan rincian data.	Kemungkinan kekhawatiran mengenai kepemilikan data.	Masalah yang sangat rumit. Bekerja sama dengan sponsor eksekutif untuk menyelesaikan masalah ini.
Pengguna tidak nyaman dengan alat kueri.	Pengguna tidak terlatih secara memadai.	Pertama, pastikan alat kueri yang dipilih sudah sesuai. Kemudian berikan pelatihan tambahan.

Melanjutkan masalah dengan data yang dibawa ke area pementasan.	Transformasi dan pemetaan data belum selesai.	Tinjau kembali semua transformasi data dan rutinitas integrasi. Pastikan tidak ada data yang hilang. Sertakan perwakilan pengguna dalam proses verifikasi.
---	---	--

Gambar 4.11 Tanda peringatan untuk proyek gudang data.

Gambar 4.12 memberikan daftar faktor kunci keberhasilan proyek gudang data. Daftar ini sama sekali bukan merupakan kompilasi lengkap dari semua bahan yang memungkinkan untuk sukses. Ini juga bukan tongkat ajaib yang menjamin kesuksesan dalam setiap situasi. Anda tahu betul bahwa sebagian besar keberhasilan bergantung pada proyek spesifik Anda, tujuan pastinya, dan tantangan manajemen proyek yang unik. Oleh karena itu, gunakan daftar tersebut sebagai panduan umum.

Faktor kunci keberhasilan proyek gudang data.
<ul style="list-style-type: none"> • Pastikan dukungan yang berkelanjutan, jangka panjang, dan berkomitmen dari sponsor eksekutif. • Di awal, tetapkan nilai bisnis yang terdefinisi dengan baik, nyata, dan disepakati dari Gudang data Anda. Kelola ekspektasi pengguna secara realistis. Libatkan pengguna dengan antusias sepanjang proyek. • Fungsi ekstraksi, transformasi, dan pemuatan data (ETL) adalah yang paling memakan Waktu kegiatan yang memakan waktu dan padat karya. Jangan meremehkan waktu dan upaya untuk kegiatan ini. • Ingat arsitektur terlebih dahulu, lalu teknologi, lalu peralatan. Pilih arsitektur yang tepat untuk lingkungan Anda. • Alat kueri dan informasi yang tepat bagi pengguna sangatlah penting. Pilih yang paling berguna dan mudah digunakan, bukan yang glamor. Hindari teknologi mutakhir. • Rencanakan pertumbuhan dan evolusi. Perhatikan pertimbangan kinerja. • Tetapkan manajer proyek yang berorientasi pengguna. • Fokuskan desain pada kueri, bukan transaksi. • Tentukan sumber data yang tepat. Hanya memuat data yang diperlukan.

Gambar 4.12 Faktor kunci keberhasilan proyek gudang data.

Anatomi Proyek yang Sukses

Tidak peduli berapa banyak faktor keberhasilan yang Anda ulas, dan tidak peduli berapa banyak pedoman yang Anda pelajari, Anda akan lebih memahami prinsip-prinsip kesuksesan dengan menganalisis detail tentang apa yang sebenarnya membuat proyek dunia nyata sukses. Kami sekarang akan melakukan hal itu. Mari kita tinjau studi kasus bisnis aktual di mana proyek gudang data sukses luar biasa.

Gudang memenuhi tujuan dan memberikan hasil yang diinginkan. Gambar 4-13 menggambarkan gudang data ini, yang menunjukkan faktor keberhasilan dan manfaatnya. Nama fiktif digunakan untuk bisnis tersebut.



Gambar 4.13 Analisis gudang data yang berhasil.

Gunakan Pendekatan Praktis

Setelah seluruh prinsip manajemen proyek diumumkan, sejumlah metode perencanaan dijelaskan, dan beberapa nuansa teoretis dieksplorasi, pendekatan praktis masih merupakan pilihan terbaik untuk mencapai hasil. Jangan terjebak pada ketatnya prinsip, aturan, dan metode. Mengadopsi pendekatan praktis untuk mengelola proyek. Hasil saja yang penting; hanya bersikap aktif dan berlarian mengejar prinsip-prinsip teoretis tidak akan membuahkan hasil yang diinginkan.

Pendekatan praktis hanyalah pendekatan yang masuk akal yang memiliki perpaduan yang bagus antara kebijaksanaan praktis dan teori inti. Saat menggunakan pendekatan praktis, Anda sepenuhnya berorientasi pada hasil. Anda terus-menerus menyeimbangkan aktivitas penting dengan aktivitas kurang penting dan menyesuaikan prioritas. Anda tidak didorong oleh teknologi hanya demi teknologi itu sendiri; Anda termotivasi oleh kebutuhan bisnis.

Dalam konteks proyek gudang data, berikut beberapa tip dalam mengadopsi pendekatan praktis:

- Menjalankan proyek dengan cara yang pragmatis berarti terus-menerus memantau penyimpangan dan kesalahan, serta melakukan koreksi agar tetap berada di jalur yang benar. Susun ulang prioritas jika diperlukan.

- Biarkan jadwal proyek bertindak sebagai panduan untuk kelancaran alur kerja dan mencapai hasil, bukan hanya untuk mengendalikan dan menghambat kreativitas. Mohon jangan mencoba mengontrol setiap tugas hingga detail terkecil. Anda kemudian hanya punya waktu untuk memperbarui jadwal, dengan lebih sedikit waktu untuk melakukan pekerjaan sebenarnya.
- Tinjau ketergantungan tugas proyek secara terus menerus. Minimalkan waktu tunggu untuk tugas-tugas yang bergantung.
- Memang ada yang namanya “terlalu banyak perencanaan”. Jangan menyerah pada godaan. Kadang-kadang, bidikan siap tembak mungkin merupakan prinsip yang berguna untuk pendekatan praktis.
- Demikian pula, terlalu banyak analisis dapat menghasilkan “kelumpuhan analisis.”
- Hindari teknologi yang “paling canggih” dan belum terbukti. Hal ini sangat penting jika proyek tersebut merupakan proyek data warehouse pertama di perusahaan Anda.
- Selalu menghasilkan hasil awal sebagai bagian dari proyek. Hasil kerja ini akan mempertahankan minat pengguna dan juga berfungsi sebagai sistem pembuktian konsep.
- Bangun arsitekturnya terlebih dahulu, baru kemudian alatnya. Jangan memilih alat lalu membangun gudang data Anda berdasarkan alat yang dipilih. Bangun arsitekturnya terlebih dahulu, berdasarkan kebutuhan bisnis, lalu pilih alat untuk mendukung arsitektur tersebut.

Tinjau saran ini dan gunakan dengan tepat dalam proyek gudang data Anda. Terutama jika ini adalah proyek gudang data pertama mereka, pengguna akan tertarik dengan manfaat yang cepat dan mudah dirasakan. Anda akan segera mengetahui bahwa mereka tidak pernah tertarik dengan alat penjadwalan proyek termewah Anda yang memungkinkan mereka melacak setiap tugas per jam atau menit. Mereka puas hanya dengan hasil. Mereka tertarik pada gudang data hanya karena kegunaan dan kemudahan penggunaannya.

RINGKASAN BAB

- Saat merencanakan gudang data Anda, isu-isu utama yang harus dipertimbangkan termasuk menetapkan ekspektasi yang tepat, menilai risiko, memutuskan antara pendekatan top-down atau bottom-up, memilih solusi vendor.
- Persyaratan bisnis, bukan teknologi, yang harus mendorong proyek Anda.
- Sebuah proyek gudang data tanpa dukungan penuh dari manajemen puncak dan tanpa sponsor eksekutif yang kuat dan antusias akan mengalami kegagalan sejak hari pertama.
- Manfaat dari gudang data diperoleh hanya setelah pengguna menggunakannya sepenuhnya. Pembenaran melalui perhitungan ROI yang ketat tidak selalu mudah. Beberapa gudang data dibenarkan dan proyek dimulai hanya dengan meninjau potensi manfaatnya.
- Proyek gudang data jauh berbeda dari proyek sistem OLTP pada umumnya. Pendekatan siklus hidup tradisional pengembangan aplikasi harus diubah dan

disesuaikan untuk proyek gudang data. Juga, pertimbangkan untuk menggunakan metodologi pengembangan tangkas.

- Standar pengorganisasian dan penetapan peran tim masih dalam tahap percobaan di banyak proyek. Ubah peran agar sesuai dengan apa yang penting bagi proyek Anda.
- Partisipasi pengguna adalah wajib untuk keberhasilan proyek gudang data. Pengguna dapat berpartisipasi dalam berbagai cara.
- Pertimbangkan tanda-tanda peringatan dan faktor keberhasilan: pada analisis akhir, terapkan pendekatan praktis untuk membangun gudang data yang sukses.

PERTANYAAN TINJAUAN

1. Sebutkan empat isu utama yang harus dipertimbangkan saat merencanakan gudang data.
2. Jelaskan perbedaan antara pendekatan top-down dan bottom-up dalam membangun data warehouse. Apakah Anda punya preferensi? Jika ya, mengapa?
3. Sebutkan tiga keunggulan masing-masing solusi vendor tunggal dan multivendor.
4. Apa yang dimaksud dengan survei pendahuluan persyaratan? Sebutkan enam jenis informasi yang akan Anda kumpulkan selama survei pendahuluan.
5. Apa perbedaan proyek gudang data dengan proyek sistem OLTP? Jelaskan empat perbedaan tersebut.
6. Sebutkan dan jelaskan empat fase pengembangan dalam siklus hidup proyek gudang data.
7. Apa yang Anda anggap sebagai serangkaian peran tim inti untuk proyek gudang data? Jelaskan tanggung jawab tiga peran dari set Anda.
8. Sebutkan tiga tanda peringatan yang mungkin ditemui dalam proyek gudang data. Tindakan perbaikan apa yang perlu Anda ambil untuk mengatasi potensi masalah yang ditunjukkan oleh ketiga tanda peringatan ini?
9. Sebutkan dan jelaskan lima faktor keberhasilan dalam proyek gudang data.
10. Apa yang dimaksud dengan “mengambil pendekatan praktis” terhadap pengelolaan proyek data warehouse? Berikan dua alasan mengapa menurut Anda pendekatan praktis akan berhasil.

BAB 5

MENDEFINISIKAN PERSYARATAN BISNIS

TUJUAN BAB

- Diskusikan bagaimana dan mengapa pendefinisian persyaratan berbeda untuk gudang data
- Memahami peran dimensi bisnis
- Pelajari tentang paket informasi dan penggunaannya dalam menentukan persyaratan
- Meninjau metode pengumpulan persyaratan
- Memahami pentingnya dokumen definisi persyaratan formal

Gudang data adalah sistem pengiriman informasi untuk intelijen bisnis. Ini bukan tentang teknologi, tapi tentang menyelesaikan masalah pengguna dan memberikan informasi strategis kepada pengguna. Pada tahap menentukan persyaratan, Anda perlu berkonsentrasi pada informasi apa yang dibutuhkan pengguna, bukan pada bagaimana Anda akan memberikan informasi yang diperlukan. Metode sebenarnya untuk memberikan informasi akan dibahas kemudian, bukan saat Anda mengumpulkan persyaratan.

Sebelum kita melanjutkan, mari kita perjelas ruang lingkup dan isi bab ini. Dalam bab ini, kita akan fokus pada penentuan persyaratan untuk apa yang disebut sebagai pendekatan praktis di Bab 2. Seperti yang kita diskusikan di bab tersebut, kami merencanakan solusi di seluruh perusahaan; lalu kami mengumpulkan persyaratan untuk setiap data mart, subjek demi subjek. Persyaratan ini akan memungkinkan kita untuk menerapkan data mart yang sesuai, satu per satu, berdasarkan prioritas; pada akhirnya untuk memenuhi kebutuhan seluruh perusahaan.

Di sisi lain, jika tujuannya adalah untuk mengambil pendekatan yang ketat dari atas ke bawah dan membangun gudang data seluruh perusahaan terlebih dahulu, penentuan persyaratannya akan berbeda dari apa yang dibahas dalam bab ini. Dengan pendekatan murni top-down, gudang data akan dikembangkan berdasarkan model data relasional bentuk normal ketiga. Basis data relasional ini akan membentuk gudang data. Pada gilirannya, gudang data perusahaan ini akan memberi makan data mart yang bergantung. Namun, data mart dependen ini akan dikembangkan berdasarkan penentuan persyaratan yang diuraikan dalam bab ini.

Sekarang, lanjutkan diskusi kita mengenai penentuan kebutuhan data mart—hasil dari pendekatan praktis atau produk turunan dari pendekatan yang murni bersifat top-down. Sebagian besar pengembang data warehouse berasal dari latar belakang pengembangan sistem operasional atau OLTP (pemrosesan transaksi online). Sistem OLTP pada dasarnya adalah sistem penangkapan data. Di sisi lain, sistem gudang data adalah sistem penyampaian informasi. Ketika Anda mulai mengumpulkan persyaratan untuk data warehouse yang Anda usulkan, pola pikir Anda harus berbeda. Anda harus beralih dari model pengambilan data ke model penyampaian informasi. Perbedaan ini harus terlihat melalui semua tahapan proyek gudang data.

Pengguna juga memiliki perspektif berbeda tentang sistem data warehouse. Berbeda dengan sistem OLTP, yang diperlukan untuk menjalankan bisnis sehari-hari, tidak ada pembayaran langsung yang terlihat dalam sistem pendukung keputusan. Para pengguna tidak segera merasakan adanya kebutuhan yang mendesak untuk menggunakan sistem pendukung keputusan, padahal mereka tidak dapat menahan diri untuk tidak menggunakan sistem operasional, yang tanpanya mereka tidak dapat menjalankan bisnis mereka.

5.1 ANALISIS DIMENSI

Dalam beberapa hal, membangun gudang data sangat berbeda dengan membangun sistem operasional. Hal ini menjadi penting terutama dalam fase pengumpulan persyaratan. Karena perbedaan ini, metode pengumpulan persyaratan tradisional yang bekerja dengan baik untuk sistem operasional tidak dapat langsung diterapkan pada gudang data.

Penggunaan Informasi Tidak Dapat Diprediksi

Bayangkan Anda sedang membangun sistem operasional untuk pemrosesan pesanan di perusahaan Anda. Untuk mengumpulkan persyaratan, Anda mewawancarai pengguna di departemen pemrosesan pesanan. Pengguna akan mencantumkan semua fungsi yang perlu dilakukan. Mereka akan memberi tahu Anda bagaimana mereka menerima pesanan, memeriksa stok, memverifikasi pengaturan kredit pelanggan, menentukan harga pesanan, menentukan pengaturan pengiriman, dan mengarahkan pesanan ke gudang yang sesuai. Mereka akan menunjukkan kepada Anda bagaimana mereka ingin berbagai elemen data disajikan pada layar GUI (antarmuka pengguna grafis) untuk aplikasi tersebut. Pengguna juga akan memberi Anda daftar laporan yang mereka perlukan dari aplikasi pemrosesan pesanan. Mereka akan dapat memberi tahu Anda bagaimana, kapan, dan di mana mereka akan menggunakan aplikasi tersebut setiap hari.

Dalam memberikan informasi tentang persyaratan sistem operasional, pengguna dapat memberikan rincian yang tepat tentang fungsi yang diperlukan, konten informasi, dan pola penggunaan. Sebaliknya, untuk sistem data warehousing, pengguna umumnya tidak dapat mendefinisikan kebutuhan mereka dengan jelas. Mereka tidak dapat menentukan dengan tepat informasi apa yang sebenarnya mereka inginkan dari data warehouse, dan mereka juga tidak dapat mengungkapkan bagaimana mereka ingin menggunakan informasi tersebut atau memprosesnya.

Bagi sebagian besar pengguna, ini mungkin merupakan gudang data pertama yang mereka temui. Pengguna sudah familiar dengan sistem operasional karena mereka menggunakannya dalam pekerjaan sehari-hari, sehingga mereka dapat memvisualisasikan persyaratan untuk sistem operasional baru lainnya. Mereka tidak dapat menghubungkan sistem gudang data dengan apa pun yang pernah mereka gunakan sebelumnya.

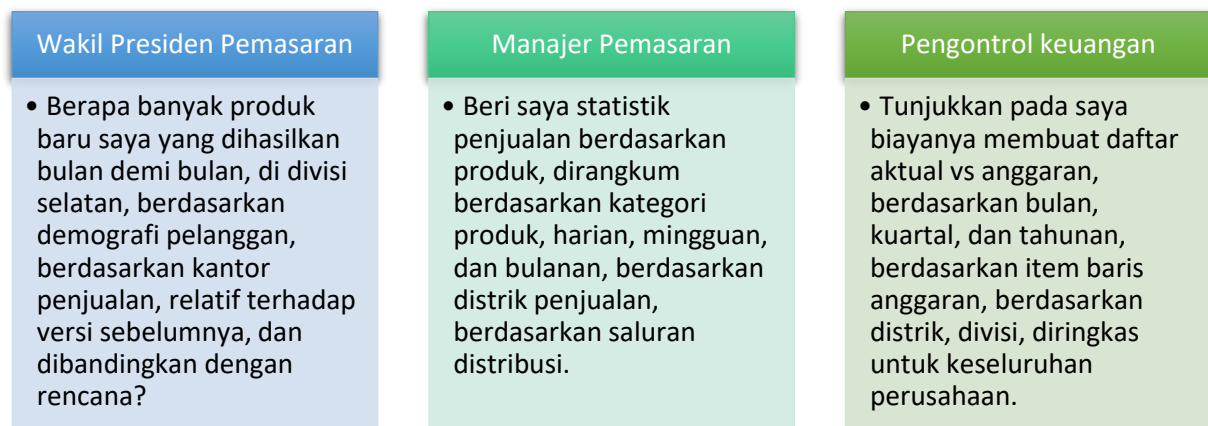
Oleh karena itu, jika seluruh proses pendefinisian kebutuhan data warehouse begitu samar-samar, bagaimana Anda bisa melanjutkan sebagai salah satu analis dalam proyek data warehouse? Anda berada dalam kebingungan. Agar aman, apakah Anda kemudian memasukkan setiap bagian data ke dalam gudang data yang menurut Anda dapat digunakan

oleh pengguna? Bagaimana Anda bisa membangun sesuatu yang pengguna tidak dapat definisikan dengan jelas dan tepat?

Awalnya, Anda dapat mengumpulkan data tentang keseluruhan bisnis organisasi. Anda dapat memeriksa praktik terbaik industri ini. Anda dapat mengumpulkan beberapa aturan bisnis yang memandu pengambilan keputusan sehari-hari. Anda mungkin mengetahui bagaimana produk dikembangkan dan dipasarkan. Namun hal ini bersifat umum dan tidak cukup untuk menentukan persyaratan secara rinci.

Sifat Dimensi Data Bisnis

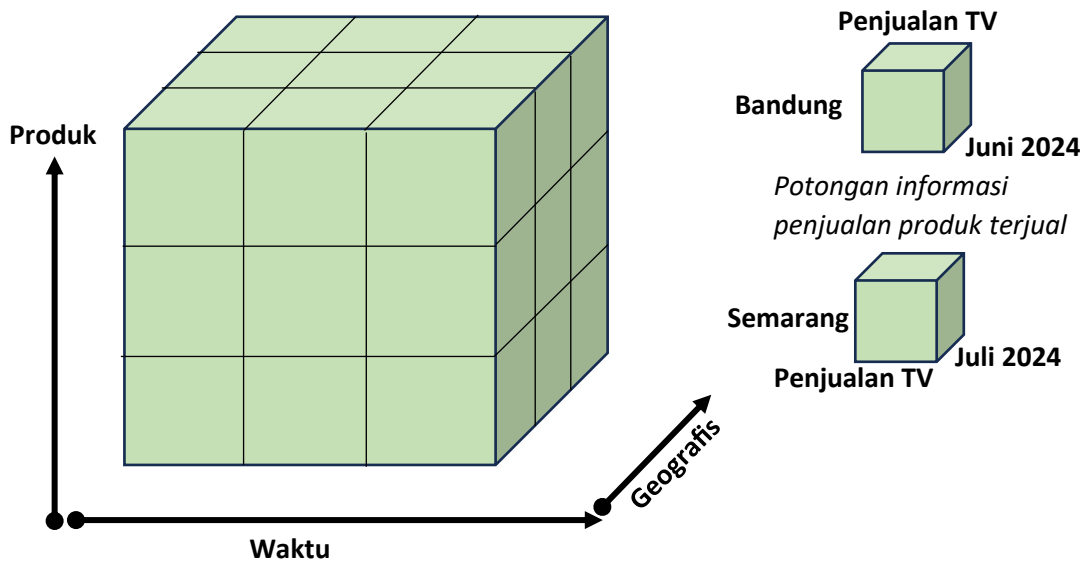
Untungnya, situasinya tidak seburuk kelihatannya. Meskipun pengguna tidak dapat sepenuhnya menjelaskan apa yang mereka inginkan dalam gudang data, mereka dapat memberi Anda wawasan yang sangat penting tentang cara mereka berpikir tentang bisnis. Mereka dapat memberi tahu Anda satuan pengukuran apa yang penting bagi mereka. Setiap departemen pengguna dapat memberi tahu Anda bagaimana mereka mengukur keberhasilan di departemen tersebut. Pengguna dapat memberi Anda wawasan tentang bagaimana mereka menggabungkan berbagai informasi untuk pengambilan keputusan strategis.



Gambar 5.1 Manajer berpikir dalam dimensi bisnis.

Manajer memikirkan bisnis dalam kaitannya dengan dimensi bisnis. Gambar 5.1 menunjukkan jenis pertanyaan yang cenderung diajukan manajer dalam pengambilan keputusan. Gambar tersebut menunjukkan pertanyaan apa yang mungkin ditanyakan oleh wakil presiden pemasaran, manajer pemasaran, dan pengontrol keuangan.

Mari kita periksa secara singkat pertanyaan-pertanyaan ini. Wakil presiden pemasaran tertarik pada pendapatan yang dihasilkan oleh produk barunya, namun dia tidak tertarik pada satu angka pun. Dia tertarik pada angka pendapatan per bulan, pada divisi tertentu, berdasarkan demografi pelanggan, berdasarkan kantor penjualan, relatif terhadap versi produk sebelumnya, dan dibandingkan dengan rencana. Jadi wakil presiden pemasaran ingin agar angka pendapatan dipecah berdasarkan bulan, divisi, demografi pelanggan, kantor penjualan, versi produk, dan rencana. Ini adalah dimensi bisnisnya yang ingin dia analisis angka-angkanya.



Gambar 5.2 Sifat dimensi data bisnis.

Demikian pula bagi manajer pemasaran, dimensi bisnisnya adalah produk, kategori produk, waktu (hari, minggu, bulan), wilayah penjualan, dan saluran distribusi. Untuk pengontrol keuangan, dimensi bisnisnya adalah garis anggaran, waktu (bulan, triwulan, tahun), wilayah, dan divisi.

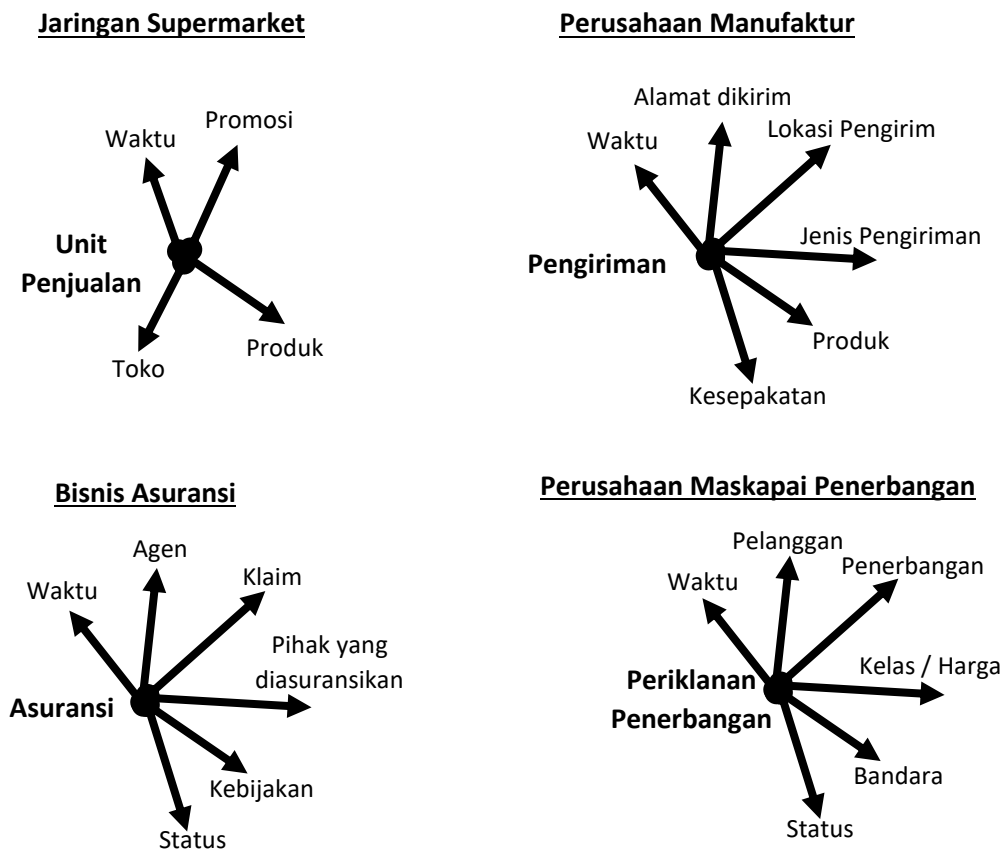
Jika pengguna gudang data Anda memikirkan dimensi bisnis untuk pengambilan keputusan, Anda juga harus memikirkan dimensi bisnis saat mengumpulkan persyaratan. Meskipun usulan penggunaan data warehouse sebenarnya mungkin tidak jelas, dimensi bisnis yang digunakan oleh manajer untuk pengambilan keputusan tidaklah samar-samar sama sekali. Pengguna akan dapat menjelaskan dimensi bisnis ini kepada Anda. Anda tidak sepenuhnya tersesat dalam proses definisi persyaratan. Anda dapat mengetahui tentang dimensi bisnis.

Mari kita mencoba memahami dengan baik sifat dimensional data bisnis. Gambar 5.2 menunjukkan analisis unit penjualan sepanjang tiga dimensi bisnis yaitu produk, waktu, dan geografi. Ketiga dimensi ini diplot terhadap tiga sumbu koordinat. Anda akan melihat bahwa tiga dimensi membentuk kumpulan kubus. Di setiap kubus berdimensi kecil, Anda akan menemukan unit penjualan untuk segmen waktu, produk, dan divisi geografis tertentu. Dalam hal ini data bisnis unit penjualan bersifat tiga dimensi karena hanya tiga dimensi yang digunakan dalam analisis ini. Jika terdapat lebih dari tiga dimensi, kami memperluas konsepnya ke beberapa dimensi dan memvisualisasikan kubus multidimensi, yang juga disebut hiperkubus.

Contoh Dimensi Bisnis

Konsep dimensi bisnis merupakan dasar definisi persyaratan untuk gudang data. Oleh karena itu, kami ingin melihat lebih banyak contoh dimensi bisnis dalam beberapa kasus lainnya. Gambar 5.3 menampilkan dimensi bisnis dalam empat kasus berbeda.

Mari kita segera melihat masing-masing contoh ini. Untuk jaringan supermarket, pengukuran yang dianalisis adalah unit penjualan. Hal ini dianalisis dalam empat dimensi bisnis. Saat Anda mencari hypercubes, sisi dari kubus tersebut adalah waktu, promosi, produk, dan toko. Jika Anda adalah manajer pemasaran jaringan supermarket, Anda ingin penjualan Anda dipecah berdasarkan produk, di setiap toko, berdasarkan urutan waktu, dan dalam kaitannya dengan promosi yang berlangsung.



Gambar 5.3 Contoh dimensi bisnis.

Bagi perusahaan asuransi, dimensi bisnisnya berbeda-beda dan sesuai dengan bisnis tersebut. Di sini Anda ingin menganalisis data klaim berdasarkan agen, klaim individu, waktu, pihak bertanggung, polis individu, dan status klaim. Contoh perusahaan penerbangan menunjukkan dimensi untuk analisis data frequent flyer. Di sini dimensi bisnisnya adalah waktu, pelanggan, penerbangan spesifik, kelas tarif, bandara, dan status frequent flyer.

Contoh analisis pengiriman untuk perusahaan manufaktur menunjukkan beberapa dimensi bisnis lainnya. Dalam hal ini, dimensi bisnis yang digunakan untuk analisis pengiriman adalah dimensi yang relevan dengan bisnis tersebut dan subjek analisisnya. Di sini Anda melihat dimensi waktu, lokasi pengiriman ke dan dari, mode pengiriman, produk, dan penawaran khusus apa pun.

Apa yang kami temukan dari contoh-contoh ini adalah bahwa dimensi bisnis berbeda dan relevan dengan industri dan subjek analisis. Kami juga menemukan dimensi waktu sebagai dimensi umum dalam semua contoh. Hampir semua analisis bisnis dilakukan seiring waktu.

5.2 PAKET INFORMASI KONSEP YANG BERMANFAAT

Kami sekarang akan memperkenalkan ide baru untuk menentukan dan mencatat kebutuhan informasi untuk gudang data. Konsep ini membantu kita memberikan bentuk konkrit terhadap berbagai wawasan, pemikiran samar-samar, dan pendapat yang diungkapkan selama proses pengumpulan persyaratan. Paket informasi, yang dikumpulkan sambil mengumpulkan persyaratan, sangat berguna untuk membawa pengembangan data warehouse ke tahap berikutnya.

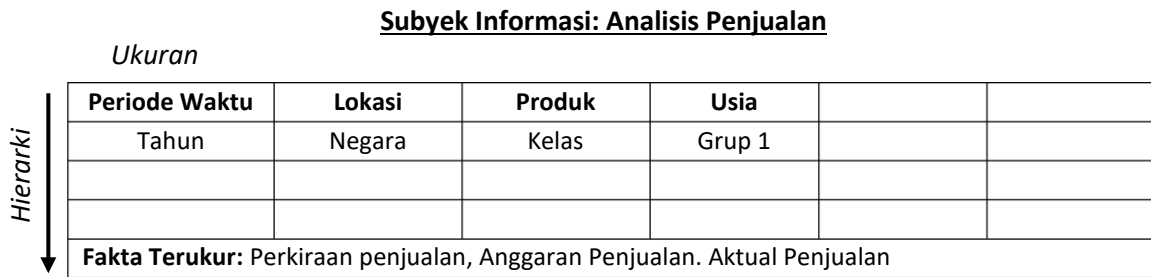
Persyaratan Tidak Sepenuhnya Ditentukan

Seperti yang telah kita diskusikan, pengguna tidak dapat menjelaskan sepenuhnya apa yang mereka harapkan untuk dilihat di gudang data. Anda tidak dapat mengetahui informasi apa yang ingin Anda simpan di gudang data. Anda tidak yakin dengan pola penggunaannya. Anda tidak dapat menentukan bagaimana setiap kelas pengguna akan menggunakan sistem baru. Jadi, ketika persyaratan tidak dapat ditentukan sepenuhnya, kita memerlukan konsep baru dan inovatif untuk mengumpulkan dan mencatat persyaratan tersebut. Metode tradisional yang diterapkan pada sistem operasional tidak memadai dalam konteks ini. Kita tidak bisa memulai dengan fungsi, layar, dan laporan. Kita tidak bisa memulai dengan struktur data. Kami telah mencatat bahwa pengguna cenderung berpikir dalam dimensi bisnis dan menganalisis pengukuran sepanjang dimensi bisnis tersebut. Ini merupakan pengamatan yang signifikan dan dapat menjadi dasar pengumpulan informasi.

Metodologi baru untuk menentukan persyaratan sistem gudang data didasarkan pada dimensi bisnis. Hal ini muncul dari kebutuhan pengguna untuk mendasarkan analisis mereka pada dimensi bisnis. Konsep baru ini menggabungkan pengukuran dasar dan dimensi bisnis yang digunakan pengguna untuk menganalisis pengukuran dasar tersebut. Dengan menggunakan metodologi baru, Anda mendapatkan pengukuran dan dimensi relevan yang harus ditangkap dan disimpan di gudang data. Anda menemukan apa yang dikenal sebagai paket informasi untuk subjek tertentu.

Mari kita lihat paket informasi untuk menganalisis penjualan untuk bisnis tertentu. Gambar 5.4 berisi paket informasi tersebut. Subjek di sini adalah penjualan. Fakta terukur atau pengukuran yang menarik untuk dianalisis ditunjukkan di bagian bawah diagram paket. Dalam hal ini yang diukur adalah penjualan aktual, perkiraan penjualan, dan anggaran penjualan. Dimensi bisnis yang akan dianalisis pengukurannya ditunjukkan di bagian atas diagram sebagai judul kolom. Dalam contoh kita, dimensi tersebut adalah waktu, lokasi, produk, dan kelompok usia demografis. Masing-masing dimensi bisnis ini berisi hierarki atau tingkatan. Misalnya, dimensi waktu memiliki hierarki mulai dari tahun ke bawah hingga tingkat hari tertentu. Tingkat perantara lainnya dalam dimensi waktu dapat berupa kuartal, bulan, dan minggu. Tingkat atau komponen hierarki ini ditunjukkan dalam diagram paket informasi.

Tujuan utama Anda dalam fase definisi persyaratan adalah mengkompilasi paket informasi untuk semua subjek gudang data. Setelah Anda menyiapkan paket informasi, Anda dapat melanjutkan ke fase lainnya.



Gambar 5.4 Paket informasi.

Pada dasarnya, paket informasi memungkinkan Anda untuk:

- ❖ Tentukan bidang subjek yang umum
- ❖ Merancang metrik bisnis utama
- ❖ Memutuskan bagaimana data harus disajikan
- ❖ Menentukan bagaimana pengguna akan menggabungkan atau menggabungkannya
- ❖ Tentukan kuantitas data untuk analisis atau kueri pengguna
- ❖ Memutuskan bagaimana data akan diakses
- ❖ Menetapkan granularitas data
- ❖ Perkirakan ukuran gudang data
- ❖ Menentukan frekuensi penyegaran data
- ❖ Memastikan bagaimana informasi harus dikemas

Dimensi Bisnis

Seperti yang telah kita lihat, dimensi bisnis membentuk dasar metodologi baru untuk definisi persyaratan. Data harus disimpan untuk menyediakan dimensi bisnis. Dimensi bisnis dan tingkat hierarkinya menjadi dasar untuk semua fase pengembangan lebih lanjut. Jadi kami ingin melihat lebih dekat dimensi bisnis. Kita harus mampu mengidentifikasi dimensi bisnis dan tingkat hierarkinya. Kita harus dapat memilih kumpulan dimensi yang tepat dan optimal terkait dengan pengukuran.

Kita mulai dengan memeriksa dimensi bisnis produsen mobil. Katakanlah tujuannya adalah untuk menganalisis penjualan. Kami ingin membangun gudang data yang memungkinkan pengguna menganalisis penjualan mobil dalam beberapa cara. Dimensi pertama yang jelas adalah dimensi produk. Sekali lagi bagi produsen mobil, analisis penjualan harus mencakup analisis dengan mengelompokkan penjualan berdasarkan dealer. Oleh karena itu, dealer adalah dimensi penting lainnya untuk analisis. Sebagai produsen mobil, Anda pasti ingin mengetahui bagaimana penjualan Anda dibagi berdasarkan demografi pelanggan. Anda pasti ingin tahu siapa yang membeli mobil Anda dan berapa jumlahnya. Demografi pelanggan akan menjadi dimensi bisnis lain yang berguna untuk dianalisis. Bagaimana cara pelanggan membayar mobil? Apa dampak pembiayaan pembelian terhadap penjualan? Pertanyaan-pertanyaan ini dapat dijawab dengan memasukkan metode pembayaran sebagai dimensi analisis lainnya. Bagaimana dengan waktu sebagai dimensi bisnis? Hampir setiap kueri atau analisis melibatkan elemen waktu. Singkatnya, kami telah

menghasilkan dimensi berikut untuk subjek penjualan produsen mobil: produk, dealer, demografi pelanggan, metode pembayaran, dan waktu.

Mari kita ambil satu contoh lagi. Dalam hal ini, kami ingin membuat paket informasi untuk jaringan hotel. Subyek dalam hal ini adalah okupansi hotel. Kami ingin menganalisis tingkat hunian kamar di berbagai cabang jaringan hotel. Kami ingin menganalisis tingkat hunian berdasarkan masing-masing hotel dan tipe kamar. Jadi, tipe hotel dan kamar merupakan dimensi bisnis yang penting untuk analisis. Seperti pada kasus lainnya, kita juga perlu memasukkan dimensi waktu. Pada paket informasi okupansi hotel, dimensi yang dicantumkan adalah hotel, tipe kamar, dan waktu.

Hirarki dan Kategori Dimensi

Saat pengguna menganalisis pengukuran sepanjang dimensi bisnis, pengguna biasanya ingin melihat angka-angka terlebih dahulu secara ringkasan dan kemudian pada berbagai tingkat detail. Apa yang dilakukan pengguna di sini adalah menelusuri tingkat hierarki dimensi bisnis untuk mendapatkan detail di berbagai tingkat. Misalnya, pengguna pertama kali melihat total penjualan sepanjang tahun. Kemudian pengguna berpindah ke tingkat kuartal dan melihat penjualan per kuartal. Setelah ini, pengguna berpindah lebih jauh ke tingkat bulan individual untuk melihat angka bulanan. Yang kita perhatikan di sini adalah hierarki dimensi waktu terdiri dari tingkatan tahun, triwulan, dan bulan. Hirarki dimensi adalah jalur untuk menelusuri atau menggulirkan analisis kita.

Dalam setiap dimensi bisnis utama terdapat kategori elemen data yang juga dapat berguna untuk analisis. Dalam dimensi waktu, Anda mungkin memiliki elemen data untuk menunjukkan apakah hari tertentu merupakan hari libur. Elemen data ini memungkinkan Anda menganalisis berdasarkan hari libur dan melihat bagaimana penjualan pada hari libur dibandingkan dengan penjualan pada hari lainnya. Demikian pula dalam dimensi produk, Anda mungkin ingin menganalisis berdasarkan jenis paket. Jenis paket adalah salah satu elemen data dalam dimensi produk. Tanda hari libur dalam dimensi waktu dan jenis paket dalam dimensi produk tidak selalu menunjukkan tingkat hierarki dalam dimensi ini. Elemen data dalam dimensi bisnis dapat disebut kategori.

Hirarki dan kategori disertakan dalam paket informasi untuk setiap dimensi. Mari kita kembali ke dua contoh di bagian sebelumnya dan mencari tahu tingkat hierarki dan kategori mana yang harus disertakan untuk dimensi tersebut. Mari kita periksa dimensi produk. Di sini, produknya adalah mobil dasar. Oleh karena itu, kami menyertakan elemen data yang relevan dengan produk sebagai hierarki dan kategori. Ini akan berupa nama model, tahun model, gaya paket, lini produk, kategori produk, warna eksterior, warna interior, dan tahun model pertama. Melihat dimensi bisnis lain untuk analisis penjualan mobil, kami merangkum hierarki dan kategori untuk setiap dimensi sebagai berikut:

- ❖ **Produk:** Nama model, tahun model, gaya paket, lini produk, kategori produk, warna eksterior, warna interior, tahun model pertama
- ❖ **Dealer:** Nama dealer, kota, negara bagian, bendera merek tunggal, tanggal pengoperasian pertama

- ❖ **Demografi Pelanggan:** Usia, jenis kelamin, kisaran pendapatan, status perkawinan, ukuran rumah tangga, kendaraan yang dimiliki, nilai rumah, dimiliki atau disewa
- ❖ **Metode Pembayaran:** Jenis keuangan, jangka waktu dalam bulan, tingkat bunga, agen
- ❖ **Waktu:** Tanggal, bulan, kuartal, tahun, hari dalam seminggu, hari dalam bulan, musim, bendera hari libur

Mari kita kembali ke analisis hunian hotel. Kami telah memasukkan tiga dimensi bisnis. Mari kita buat daftar kemungkinan hierarki dan kategori untuk tiga dimensi.

- ❖ **Hotel:** Jalur hotel, nama cabang, kode cabang, wilayah, alamat, kota, negara bagian, kode pos, pengelola, tahun konstruksi, tahun renovasi
- ❖ **Tipe Kamar:** Tipe kamar, ukuran kamar, jumlah tempat tidur, jenis tempat tidur, jumlah penghuni maksimal, suite, kulkas, dapur kecil
- ❖ **Waktu:** Tanggal, hari dalam sebulan, hari dalam seminggu, bulan, kuartal, tahun, bendera hari libur

Metrik atau Fakta Bisnis Utama

Sejauh ini kita telah membahas dimensi bisnis dalam dua contoh di atas. Ini adalah dimensi bisnis yang relevan bagi pengguna kedua data mart untuk melakukan analisis. Masing-masing pengguna memikirkan subjek bisnis mereka dalam kaitannya dengan dimensi bisnis ini untuk memperoleh informasi dan untuk melakukan analisis. Namun dengan menggunakan dimensi bisnis ini, apa sebenarnya yang dianalisis oleh pengguna? Angka-angka apa yang mereka analisa? Angka-angka yang dianalisis pengguna adalah ukuran atau metrik yang mengukur keberhasilan departemen mereka. Ini adalah fakta yang menunjukkan kepada pengguna bagaimana kinerja departemen mereka dalam memenuhi tujuan departemen mereka.

Dalam kasus produsen mobil, metrik ini berhubungan dengan penjualan. Ini adalah angka-angka yang memberi tahu pengguna tentang kinerja mereka dalam penjualan. Ini adalah angka penjualan setiap mobil. Kumpulan metrik yang bermakna dan berguna untuk menganalisis penjualan mobil adalah sebagai berikut:

- ◆ Harga Jual Sebenarnya MSRP
- ◆ Harga opsi Harga penuh Pengaya dealer Kredit dealer Faktur dealer
- ◆ Jumlah uang muka Pendapatan produsen Jumlah yang dibiayai

Pada contoh okupansi hotel yang kedua, angka atau metriknya berbeda. Sifat metrik bergantung pada apa yang dianalisis. Untuk okupansi hotel, metriknya akan berhubungan dengan okupansi kamar di setiap cabang jaringan hotel. Berikut daftar metrik untuk menganalisis okupansi hotel:

Kamar yang terisi Kamar kosong Kamar yang tidak tersedia Jumlah penghuni Pendapatan

Sekarang, jika digabungkan, mari kita bahas apa yang ada dalam diagram paket informasi untuk kedua contoh ini. Dalam setiap kasus, metrik atau fakta ditempatkan di bagian bawah paket informasi. Dimensi bisnis akan menjadi judul kolom. Di setiap kolom, Anda akan menyertakan hierarki dan kategori untuk dimensi bisnis.

Gambar 5.5 dan 5.6 menunjukkan paket informasi untuk dua contoh yang baru saja kita diskusikan.

Subjek Informasi: Penjualan Automaker

Hierarki	Waktu	Produk	Metode Pembayaran	Demografi Pelanggan	Pedagang
	Tahun	Nama Model	Tipe Financial	Usia	Nama Dealer
	Triwulan	Tahun Model	Tempo	Jenis Kelamin	Kota
	Bulan	Gaya pengemasan	Rata-rata Minat	Rata-rata Pendapatan	Negara
	Tanggal	Line produk	Agen	Status perkawinan	Merek
	Sehari dalam seminggu	Kategori Produk		Ukuran Rumah Tangga	Operasi Pertama
	seharu dalam sebulan	Warna Eksterior		Kepemilikan Kendaraan	
	Musim	Warna Interior		Nilai Rumah	
	Liburan	Tahun Pertama		Sewa atau Hak Milik	
	Fakta: Harga jual aktual, MSRP, Harga Opso, Harga penuh, pengaya Dealer, Kredit Dealer, Faktur Dealer, Uang Muka, Hasil, Keuangan.				

Gambar 5.5 Paket informasi: penjualan pembuat mobil.

Subjek informasi: Persewaan Hotel						
Dimensi						
Hierarki	Waktu	Hotel	Tipe Ruangan			
	Tahun	Jalur Hotel	Tipe Ruangan			
	Triwulan	Nama Cabang	Ukuran Ruangan			
	Bulan	Kode Cabang	Jumlah Tempat Tidur			
	Tanggal	Wilayah	Jenis tempat Tidur			
	Sehari dalam seminggu	Alamat	Maks. Penghuni			
	Sehari dalam sebulan	Kota/Negara	Kelengkapan			
	Hari Libur	Tahun Kontruksi	Lemari Es			
		Tahun Renovasi	Dapur			
	Fakta: Kamar Terhuni, Kamar Kosong, Kamar tidak tersedia, Jumlah Penghuni, Pendapatan					

Gambar 5.6 Paket informasi: hunian hotel.

5.3 METODE PENGUMPULAN PERSYARATAN

Sekarang kita mempunyai cara memformalkan definisi kebutuhan melalui diagram paket informasi, mari kita bahas metode pengumpulan kebutuhan. Ingatlah bahwa gudang data adalah sistem penyampaian informasi untuk menyediakan informasi untuk pengambilan keputusan strategis. Ini bukan sistem untuk menjalankan bisnis sehari-hari. Siapa pengguna yang dapat memanfaatkan informasi di gudang data? Kemana Anda pergi untuk mendapatkan persyaratannya?

Secara garis besar, kita dapat mengklasifikasikan pengguna data warehouse sebagai berikut:

Eksekutif senior (termasuk sponsor)

Manajer departemen utama

Analisis bisnis

Administrator basis data sistem operasional (DBA)

Lainnya dinominasikan oleh yang di atas

Eksekutif akan memberi Anda arah dan ruang lingkup untuk gudang data Anda. Merekalah yang terlibat erat dalam area fokus. Manajer departemen utama adalah orang-orang yang melapor kepada para eksekutif di bidang fokusnya. Analisis bisnis adalah orang yang menyiapkan laporan dan analisis untuk para eksekutif dan manajer. DBA sistem operasional dan staf aplikasi TI akan memberi Anda informasi tentang sumber data untuk gudang.

Persyaratan apa yang perlu Anda kumpulkan? Berikut daftar luasnya:

Elemen data: kelas fakta, dimensi

Pencatatan data berdasarkan waktu

Ekstrak data dari sistem sumber

Aturan bisnis: atribut, rentang, domain, catatan operasional

Anda harus pergi ke kelompok orang yang berbeda di berbagai departemen untuk mengumpulkan persyaratan. Tiga teknik dasar yang digunakan secara universal untuk memperoleh informasi dari sekelompok orang: (1) wawancara, satu lawan satu atau dalam kelompok kecil; (2) sesi pengembangan aplikasi bersama (JAD); (3) kuesioner. Wawancara satu lawan satu adalah metode yang paling interaktif dari ketiga metode tersebut. JAD atau sesi kelompok sedikit kurang interaktif. Kuesioner tidak interaktif. Metode interaktif mempunyai beberapa keuntungan dalam memperoleh informasi yang terkonfirmasi.

Berikut ini beberapa pemikiran tentang ketiga pendekatan dasar ini.

Wawancara

- ❖ Dua atau tiga orang sekaligus
- ❖ Mudah untuk dijadwalkan
- ❖ Pendekatan yang baik jika detailnya rumit
- ❖ Beberapa pengguna merasa nyaman hanya dengan wawancara satu lawan satu
- ❖ Perlu persiapan yang baik agar efektif
- ❖ Selalu melakukan penelitian sebelum wawancara
- ❖ Tetapkan tujuan untuk setiap wawancara
- ❖ Tentukan jenis pertanyaan
- ❖ Juga mendorong pengguna untuk mempersiapkan wawancara

Sesi Kelompok

- ❖ Kelompok yang terdiri dari 20 orang atau kurang pada satu waktu
- ❖ Gunakan hanya setelah mendapatkan pemahaman dasar tentang persyaratannya
- ❖ Tidak baik untuk pengumpulan data awal
- ❖ Gunakan ketika aliran ide yang bebas sangat penting
- ❖ Berguna untuk mengonfirmasi persyaratan
- ❖ Efisien bila pengguna tersebar di berbagai lokasi
- ❖ Harus terorganisir dengan baik

Kuesioner

- ❖ Dapat mengumpulkan banyak kebutuhan dengan cepat
- ❖ Berguna ketika orang yang akan ditanyai tersebar luas
- ❖ Baik dalam fase eksplorasi untuk mendapatkan reaksi keseluruhan
- ❖ Dapat digunakan untuk orang yang jadwal kerjanya terlalu padat untuk wawancara
- ❖ Namun, kuesioner tidak mengizinkan tanggapan interaktif seperti wawancara

Jenis Pertanyaan

Dalam ketiga metode pengumpulan persyaratan yang berbeda ini, kami menggunakan pertanyaan untuk memperoleh informasi. Oleh karena itu, sangat penting untuk memahami jenis pertanyaan yang dapat digunakan dan efektivitas masing-masing pertanyaan. Pertanyaan terbuka. Hal ini membuka pilihan bagi orang yang diwawancarai untuk memberikan tanggapan. Manfaat menggunakan pertanyaan terbuka adalah membuat orang yang diwawancarai merasa nyaman, memberikan wawasan tentang nilai-nilai dan keyakinan, memberikan paparan terhadap kosa kata orang yang diwawancarai, membuka kesempatan untuk bertanya lebih banyak, dan menarik serta spontan. Kelemahannya adalah dapat menghasilkan terlalu banyak detail yang tidak perlu, berisiko kehilangan kendali dalam wawancara, mungkin memakan waktu terlalu lama, tidak sebanding dengan informasi yang dikumpulkan.

Pertanyaan Tertutup. Hal ini memungkinkan tanggapan yang terbatas terhadap orang yang diwawancarai. Beberapa pertanyaan tertutup bersifat bipolar dalam artian pertanyaan tersebut mencari jawaban tipe “Ya atau Tidak”. Pertanyaan tertutup memungkinkan Anda menghemat waktu dan langsung ke pokok permasalahan dengan cepat dan mudah. Pertanyaan tertutup memungkinkan wawancara untuk dibandingkan, memberikan kendali atas wawancara dan kemampuan untuk membahas banyak hal dengan cepat, dan cenderung hanya mengumpulkan informasi yang relevan. Kelemahannya adalah ketidakmampuan untuk mendapatkan rincian yang kaya, kurangnya kesempatan untuk membangun kepercayaan dan hubungan baik antara pewawancara dan orang yang diwawancarai, dan hal-hal tersebut mungkin menjadi membosankan dan membosankan. Penyelidikan. Ini sebenarnya adalah pertanyaan lanjutan. Probe dapat digunakan setelah pertanyaan terbuka atau tertutup. Tujuannya adalah untuk melampaui pertanyaan dan jawaban awal. Probe berguna untuk menarik sudut pandang orang yang diwawancarai.

Susunan Pertanyaan

Pengumpulan informasi hanya menggunakan jenis pertanyaan yang tepat saja tidak cukup. Anda harus dapat menyusun pertanyaan dalam urutan yang efektif agar sesuai dengan audiens dan tujuan. Struktur penyusunan pertanyaan berikut digunakan dalam praktik:

- ❖ **Struktur Piramida:** Ini adalah metode induktif dalam menyusun pertanyaan. Anda memulai dengan pertanyaan tertutup yang sangat spesifik dan kemudian memperluas topik dengan pertanyaan terbuka. Struktur ini berguna ketika orang yang diwawancarai perlu melakukan pemanasan terhadap topik yang sedang dibahas. Gunakan struktur ini ketika pandangan umum tentang topik ingin diekstraksi di bagian akhir.

- ❖ **Struktur Corong:** Ini adalah metode deduktif. Mulailah dengan pertanyaan umum terbuka dan kemudian persempit topik dengan pertanyaan spesifik dan tertutup. Struktur ini berguna ketika orang yang diwawancarai sedang emosional terhadap topik yang sedang dibahas. Gunakan struktur ini ketika diperlukan tingkat detail yang bertahap di bagian akhir.
- ❖ **Struktur Berbentuk Berlian:** Dalam hal ini, Anda menghangatkan wawancara dengan pertanyaan tertutup yang spesifik. Anda kemudian melanjutkan ke pertanyaan yang luas, umum, dan terbuka. Akhirnya Anda mempersempit wawancara dan mencapai penutupan dengan pertanyaan tertutup yang spesifik. Biasanya struktur ini lebih baik daripada dua lainnya. Namun, struktur ini mungkin memperpanjang wawancara.

Teknik Wawancara

Sesi wawancara dapat menghabiskan sebagian besar waktu proyek. Oleh karena itu, hal ini harus diorganisir dan dikelola dengan baik. Sebelum tim proyek Anda memulai proses wawancara, pastikan tugas-tugas utama berikut telah diselesaikan.

- Memilih dan melatih anggota tim proyek yang melakukan wawancara
- Tetapkan peran spesifik untuk setiap anggota tim (pewawancara utama/juru tulis)
- Menyiapkan daftar pengguna yang akan diwawancarai dan menyiapkan jadwal luas
- Buatlah daftar ekspektasi Anda dari setiap rangkaian wawancara
- Menyelesaikan penelitian sebelum wawancara
- Mempersiapkan kuesioner wawancara
- Mempersiapkan pengguna untuk wawancara
- Melakukan pertemuan awal dengan seluruh pengguna untuk diwawancarai

Sebagian besar pengguna yang akan Anda wawancarai terbagi dalam tiga kategori besar: eksekutif senior, manajer/analisis departemen, dan profesional departemen TI. Apa harapan dari wawancara masing-masing kategori ini? Gambar 5.7 menunjukkan ekspektasi dasar.

Penelitian pra-wawancara penting untuk keberhasilan wawancara. Berikut adalah daftar beberapa topik penelitian utama:

- ❖ Sejarah dan struktur unit bisnis saat ini
- ❖ Jumlah karyawan serta peran dan tanggung jawabnya
- ❖ Lokasi pengguna
- ❖ Tujuan utama unit bisnis dalam perusahaan
- ❖ Hubungan unit bisnis dengan inisiatif strategis perusahaan
- ❖ Tujuan sekunder dari unit bisnis
- ❖ Hubungan unit bisnis dengan unit lain dan dengan organisasi luar
- ❖ Kontribusi unit bisnis terhadap pendapatan dan biaya perusahaan
- ❖ Pasar perusahaan
- ❖ Persaingan di pasar

Berikut ini beberapa tip tentang sifat pertanyaan yang akan diajukan dalam wawancara.

Eksekutif senior	Manajer Departemen / Analisis	Profesional Departemen TI
<ul style="list-style-type: none"> • Tujuan organisasi • Kriteria untuk mengukur keberhasilan • Masalah bisnis utama, saat ini & masa depan • Masalah identifikasi • Visi dan arah organisasi • Antisipasi penggunaan DW 	<ul style="list-style-type: none"> • Tujuan departemen • Metrik keberhasilan • Faktor-faktor yang membatasi keberhasilan • Masalah-masalah bisnis utama • Produk dan Layanan • Dimensi bisnis yang berguna untuk analisis • Antisipasi penggunaan DW 	<ul style="list-style-type: none"> • Sistem sumber operasional utama • Proses penyampaian informasi terkini • Jenis analisis rutin • Masalah kualitas yang diketahui • Dukungan TI terkini untuk permintaan informasi • Kekhawatiran terhadap usulan DW

Gambar 5.7 Harapan dari wawancara.

- ❖ Sumber Informasi Terkini Sistem operasional manakah yang menghasilkan data tentang bidang studi bisnis yang penting?
- ❖ Apa saja jenis sistem komputer yang mendukung bidang studi ini?
- ❖ Informasi apa yang saat ini disampaikan dalam laporan dan pertanyaan online yang ada?
- ❖ Bagaimana dengan tingkat detail sistem penyampaian informasi yang ada?

Area subjek

- ❖ Bidang studi manakah yang paling berharga untuk dianalisis?
- ❖ Apa dimensi bisnisnya? Apakah ini memiliki hierarki alami?
- ❖ Apa saja partisi bisnis untuk pengambilan keputusan?
- ❖ Apakah berbagai lokasi memerlukan informasi global atau hanya informasi lokal untuk pengambilan keputusan?
- ❖ Apa campurannya?
- ❖ Apakah produk dan layanan tertentu hanya ditawarkan di wilayah tertentu?

Metrik Kinerja Utama

- ❖ Bagaimana kinerja unit bisnis saat ini diukur?
- ❖ Apa saja faktor penentu keberhasilan dan bagaimana hal ini dipantau?
- ❖ Bagaimana metrik utama dirangkum?
- ❖ Apakah semua pasar diukur dengan cara yang sama?

Frekuensi Informasi

- ❖ Seberapa sering data harus diperbarui untuk pengambilan keputusan?
- ❖ Apa kerangka waktunya?
- ❖ Bagaimana setiap jenis analisis membandingkan metrik dari waktu ke waktu?
- ❖ Apa persyaratan ketepatan waktu untuk informasi di gudang data?

Sebagai dokumentasi awal untuk definisi persyaratan, siapkan tulisan wawancara menggunakan garis besar umum ini:

1. Profil pengguna
2. Latar belakang dan tujuan
3. Persyaratan informasi
4. Persyaratan analitis
5. Alat yang digunakan saat ini
6. Kriteria keberhasilan
7. Metrik bisnis yang berguna
8. Dimensi bisnis yang relevan

Mengadaptasi Metodologi JAD

Jika Anda dapat mengumpulkan banyak data dasar dari berbagai sumber, sesi kelompok mungkin merupakan pengganti yang baik untuk wawancara individu. Dalam metode ini, Anda bisa mengajak sejumlah pengguna yang tertarik untuk bertemu bersama dalam sesi grup. Secara keseluruhan, metode ini dapat menghasilkan sesi kelompok yang lebih sedikit dibandingkan sesi wawancara individu. Waktu keseluruhan untuk pengumpulan persyaratan mungkin terbukti lebih singkat dan, oleh karena itu, mempersingkat proyek. Selain itu, sesi kelompok mungkin lebih efektif jika pengguna tersebar di lokasi terpencil.

Teknik pengembangan aplikasi bersama (JAD) berhasil digunakan untuk mengumpulkan persyaratan sistem operasional pada tahun 1980an. Pengguna sistem komputer telah berkembang menjadi lebih paham komputer dan partisipasi langsung mereka dalam pengembangan aplikasi terbukti sangat berguna.

Sesuai dengan namanya, JAD adalah sebuah proses bersama, di mana semua kelompok terkait berkumpul untuk mencapai tujuan yang jelas. Ini adalah metodologi untuk mengembangkan aplikasi komputer bersama-sama oleh pengguna dan profesional TI dengan cara yang terstruktur dengan baik. JAD berpusat pada lokakarya diskusi yang berlangsung selama beberapa hari di bawah arahan seorang fasilitator. Dalam kondisi yang sesuai, pendekatan JAD dapat diadaptasi untuk membangun gudang data.

JAD terdiri dari pendekatan lima fase:

- Definisi proyek
- Selesaikan wawancara tingkat tinggi
- Lakukan wawancara manajemen
- Siapkan panduan definisi manajemen
- Riset
- Menjadi akrab dengan area bisnis dan sistem.
- Mendokumentasikan kebutuhan informasi pengguna
- Dokumentasikan proses bisnis
- Mengumpulkan informasi awal
- Mempersiapkan agenda sesi
- Persiapan
- Membuat dokumen kerja dari tahap sebelumnya.

- Latih juru tulis
- Siapkan alat bantu visual
- Melaksanakan pertemuan pre sesi.
- Siapkan tempat untuk sesi
- Siapkan daftar periksa untuk tujuan sesi JAD
- Dibuka dengan review agenda dan tujuan
- Review asumsi
- Tinjau persyaratan data
- Tinjau metrik dan dimensi bisnis
- Diskusikan hierarki dan penggabungan dimensi
- Selesaikan semua masalah yang terbuka
- Tutup sesi dengan daftar item tindakan
- Dokumen akhir
- Konversikan dokumen kerja
- Petakan informasi yang dikumpulkan
- Cantumkan semua sumber data
- Identifikasi semua metrik bisnis
- Buat daftar semua dimensi dan hierarki bisnis.
- Kumpulkan dan edit dokumen
- Lakukan sesi peninjauan
- Dapatkan persetujuan akhir
- Tetapkan prosedur untuk mengubah persyaratan

Keberhasilan suatu proyek yang menggunakan pendekatan JAD sangat bergantung pada komposisi tim JAD. Ukuran dan gabungan tim akan bervariasi berdasarkan sifat dan tujuan gudang data. Namun, komposisi tipikal harus memiliki peran terkait dalam tim. Untuk masing-masing peran berikut, biasanya satu atau lebih orang ditugaskan.

- Sponsor eksekutif: Orang yang mengendalikan pendanaan, memberikan arahan, dan memberdayakan anggota tim
- Fasilitator: Orang yang membimbing tim sepanjang proses JAD Juru Tulis—Orang yang ditunjuk untuk mencatat semua keputusan
- Peserta penuh waktu: Semua orang yang terlibat dalam pengambilan keputusan tentang data warehouse
- Peserta yang siap dipanggil: Orang yang terkena dampak proyek, namun hanya di area tertentu
- Pengamat: Orang yang ingin ikut serta dalam sesi tertentu tanpa berpartisipasi dalam pengambilan keputusan

Menggunakan Kuesioner

Karena penggunaan kuesioner tidak bersifat interaktif, kuesioner harus dirancang dengan sangat hati-hati. Kami mencatat poin-poin penting yang berkaitan dengan aspek-aspek penting dalam penyelenggaraan kuesioner. Jika dilakukan dengan benar, kuesioner dapat menjadi metode penting untuk mengumpulkan kebutuhan gudang data Anda.

Jenis dan Pilihan Soal. Anda dapat menggunakan pertanyaan terbuka dan tertutup dalam kuesioner. Pemilihan bahasa itu penting. Gunakan bahasa responden, bukan jargon teknis yang samar-samar. Bersikaplah spesifik, tidak samar-samar. Jaga agar pertanyaan tetap singkat dan tepat. Hindari bahasa yang tidak pantas dan salah secara politis. Hindari merendahkan responden. Targetkan pertanyaan kepada kelompok responden yang sesuai.

Penerapan Timbangan, Kuesioner biasanya berisi skala nominal dan interval. Ini membuatnya mudah untuk ditanggapi. Skala nominal digunakan untuk mengklasifikasikan sesuatu. Skala interval digunakan untuk analisis kuantitatif. Desain Kuesioner, urutan pertanyaan itu penting. Mulailah kuesioner dengan pertanyaan yang tidak terlalu kontroversial dan sangat penting. Kelompokkan pertanyaan dengan konten serupa. Desainnya harus menarik dan menyenangkan. Berikan ruang putih yang cukup. Sediakan ruang yang cukup untuk memberikan tanggapan. Permudah untuk menandai atau menunjukkan respons saat menggunakan skala. Pertahankan gaya yang konsisten.

Mengelola Kuesioner. Putuskan dengan hati-hati siapa yang mendapat kuesioner. Pastikan tidak ada kelalaian. Beberapa cara penyampaian kuesioner termasuk pada sesi kelompok awal, melalui pengiriman pribadi dan pengumpulan selanjutnya, administrasi mandiri oleh responden, melalui surat ke lokasi responden, dan secara elektronik melalui email.

Tinjauan Dokumentasi yang Ada

Meskipun sebagian besar pengumpulan persyaratan akan dilakukan melalui wawancara, sesi kelompok, dan kuesioner, Anda akan dapat mengumpulkan informasi berguna dari tinjauan dokumentasi yang ada. Tinjauan terhadap dokumentasi yang ada dapat dilakukan oleh tim proyek tanpa terlalu banyak keterlibatan dari pengguna unit bisnis. Penjadwalan peninjauan dokumentasi yang ada hanya melibatkan anggota tim proyek.

Dokumentasi dari Departemen Pengguna Apa yang dapat Anda peroleh dari dokumentasi yang ada? Pertama, mari kita lihat laporan dan layar yang digunakan oleh pengguna di area bisnis yang akan menggunakan data warehouse. Anda perlu mengetahui segala sesuatu tentang fungsi unit bisnis, informasi operasional yang dikumpulkan dan digunakan oleh pengguna, apa yang penting bagi mereka, dan apakah mereka menggunakan laporan yang ada untuk analisis. Anda perlu melihat dokumentasi pengguna untuk semua sistem operasional yang digunakan. Anda perlu memahami apa yang penting bagi pengguna.

Unit bisnis biasanya mempunyai dokumentasi mengenai proses dan prosedur di unit tersebut. Bagaimana cara pengguna menjalankan fungsinya? Tinjau secara rinci semua proses dan prosedur. Anda mencoba mencari tahu jenis analisis apa yang mungkin diminati oleh pengguna di unit bisnis ini. Tinjau dokumentasi dan kemudian tambahkan apa yang telah Anda pelajari dari dokumentasi yang disiapkan dari sesi wawancara. Dokumentasi dari TI Dokumentasi dari pengguna dan wawancara dengan pengguna akan memberi Anda informasi tentang metrik yang digunakan untuk analisis dan dimensi bisnis yang digunakan dalam analisis tersebut. Namun dari mana Anda mendapatkan data untuk metrik dan dimensi bisnis? Hal ini harus berasal dari sistem operasional internal. Anda perlu mengetahui apa yang tersedia di sistem sumber.

Ke mana Anda mencari informasi yang tersedia di sistem sumber? Di sinilah operasional sistem DBA dan pakar aplikasi dari IT menjadi sangat penting dalam pengumpulan data. DBA akan memberi Anda semua struktur data, elemen data individual, atribut, domain nilai, dan hubungan antar bidang dan struktur data. Dari informasi yang Anda kumpulkan dari pengguna, Anda kemudian akan dapat menghubungkan informasi pengguna dengan sistem sumber sebagaimana dipastikan dari personel TI.

Bekerjalah dengan DBA Anda untuk mendapatkan salinan kamus data atau entri katalog data untuk sistem sumber yang relevan. Pelajari struktur data, bidang data, dan hubungan. Pada akhirnya, Anda akan mengisi gudang data dari sistem sumber ini, jadi Anda perlu memahami sepenuhnya data sumber, platform sumber, dan sistem operasi. Sekarang mari kita beralih ke pakar aplikasi TI. Para profesional ini akan memberi Anda aturan bisnis dan membantu Anda memahami dan menghargai berbagai elemen data dari sistem sumber. Anda akan belajar tentang kepemilikan data, tentang orang-orang yang bertanggung jawab atas kualitas data, dan bagaimana data dikumpulkan dan diproses dalam sistem sumber. Tinjau program dan modul yang membentuk sistem sumber. Lihatlah buku salinan di dalam program untuk memahami bagaimana struktur data digunakan dalam program.

5.4 DEFINISI PERSYARATAN: RUANG LINGKUP DAN ISI

Dokumentasi formal sering diabaikan dalam proyek sistem komputer. Tim proyek melewati fase definisi persyaratan. Mereka melakukan wawancara dan sesi kelompok. Mereka meninjau dokumentasi yang ada. Mereka mengumpulkan materi yang cukup untuk mendukung fase selanjutnya dalam siklus hidup pengembangan sistem. Namun mereka melewatkan dokumentasi rinci tentang definisi persyaratan.

Ada beberapa alasan mengapa Anda harus memasukkan hasil tahap definisi kebutuhan Anda ke dalam tulisan. Pertama-tama, dokumen definisi persyaratan adalah dasar untuk tahapan selanjutnya. Jika anggota tim proyek harus meninggalkan proyek karena alasan apa pun, proyek tidak akan menderita karena orang-orang meninggalkan proyek dengan pengetahuan yang telah mereka kumpulkan. Dokumentasi formal juga akan memvalidasi temuan Anda ketika ditinjau bersama pengguna. Kami akan memberikan garis besar yang disarankan untuk dokumen definisi persyaratan formal. Sebelum itu, mari kita lihat jenis informasi yang harus dimuat dalam dokumen ini.

Sumber data

Informasi ini penting dalam dokumen definisi persyaratan. Sertakan semua detail yang telah Anda kumpulkan tentang sistem sumber. Anda akan menggunakan data sistem sumber di gudang data. Anda akan mengumpulkan data dari sistem sumber ini, menggabungkan dan mengintegrasikannya, mengubah data dengan tepat, dan mengisi gudang data.

Biasanya, dokumen definisi persyaratan harus mencakup informasi berikut:

- ❖ Sumber data yang tersedia
- ❖ Struktur data dalam sumber data
- ❖ Lokasi sumber data
- ❖ Sistem operasi, jaringan, protokol, dan arsitektur klien

- ❖ Prosedur ekstraksi data
- ❖ Ketersediaan data historis

Transformasi Data

Tidaklah cukup hanya mencantumkan sumber data yang mungkin. Anda akan mencantumkan struktur data yang relevan sebagai sumber yang mungkin karena hubungan struktur data dengan data potensial di gudang data. Setelah Anda membuat daftar sumber data, Anda perlu menentukan bagaimana sumber data harus diubah dengan tepat menjadi tipe data yang sesuai untuk disimpan di gudang data.

Dalam dokumen definisi persyaratan Anda, sertakan detail transformasi data. Hal ini tentu melibatkan pemetaan data sumber ke data di gudang data. Tunjukkan dari mana data tentang metrik dan dimensi bisnis Anda akan berasal. Jelaskan penggabungan, konversi, dan pemisahan yang perlu dilakukan sebelum memindahkan data ke gudang data.

Penyimpanan data

Dari wawancara Anda dengan pengguna, Anda akan mengetahui tingkat detail data yang perlu Anda simpan di gudang data. Anda akan memiliki gambaran tentang jumlah data mart yang Anda perlukan untuk mendukung pengguna. Selain itu, Anda akan mengetahui detail metrik dan dimensi bisnis.

Saat Anda mengetahui jenis analisis yang biasanya dilakukan pengguna, Anda dapat menentukan jenis agregasi yang harus disimpan di gudang data. Ini akan memberi Anda informasi tentang persyaratan penyimpanan tambahan. Dokumen definisi persyaratan Anda harus menyertakan rincian yang memadai tentang persyaratan penyimpanan. Siapkan perkiraan awal tentang jumlah penyimpanan yang diperlukan untuk data rinci dan ringkasan. Perkirakan berapa banyak data historis dan arsip yang perlu ada di gudang data.

Penyampaian Informasi

Dokumen definisi persyaratan Anda harus berisi persyaratan berikut tentang penyampaian informasi kepada pengguna:

- Analisis penelusuran hingga detail
- Analisis gabungan
- Analisa mendalam
- Analisis mengiris dan memotong
- Laporan ad hoc
- Alat pemantauan online seperti dasbor dan kartu skor

Diagram Paket Informasi

Kehadiran diagram paket informasi dalam dokumen definisi persyaratan merupakan perbedaan utama dan signifikan antara sistem operasional dan sistem gudang data. Ingatlah bahwa diagram paket informasi adalah pendekatan terbaik untuk menentukan persyaratan gudang data. Diagram paket informasi mengkristalkan kebutuhan informasi untuk gudang data. Mereka berisi metrik penting yang mengukur kinerja unit bisnis, dimensi bisnis yang digunakan untuk menganalisis metrik, dan rincian bagaimana analisis penelusuran dan penelusuran dilakukan. Luangkan waktu sebanyak yang diperlukan untuk memastikan

diagram paket informasi lengkap dan akurat. Desain data Anda untuk gudang data akan sepenuhnya bergantung pada keakuratan dan kecukupan diagram paket informasi.

Garis Besar Dokumen Definisi Persyaratan

1. Perkenalan. Nyatakan tujuan dan ruang lingkup proyek. Sertakan justifikasi proyek secara luas. Berikan ringkasan eksekutif untuk setiap bagian berikutnya.
2. Uraian Persyaratan Umum. Jelaskan sistem sumber yang ditinjau. Sertakan ringkasan wawancara. Nyatakan secara luas jenis kebutuhan informasi apa yang diperlukan dalam gudang data.
3. Persyaratan Khusus. Cantumkan rincian sumber data yang dibutuhkan. Buat daftar persyaratan transformasi dan penyimpanan data. Jelaskan jenis metode penyampaian informasi yang dibutuhkan oleh pengguna.
4. Paket Informasi. Berikan sedetail mungkin untuk setiap paket informasi. Sertakan ini dalam bentuk diagram paket.
5. Persyaratan Lainnya. Mencakup berbagai persyaratan seperti frekuensi ekstraksi data, metode pemuatan data, dan lokasi di mana informasi harus dikirimkan.
6. Harapan Pengguna. Nyatakan harapan dalam kaitannya dengan masalah dan peluang. Tunjukkan bagaimana pengguna berharap untuk menggunakan gudang data.
7. Partisipasi dan Sign-Off Pengguna. Buat daftar tugas dan aktivitas di mana pengguna diharapkan berpartisipasi sepanjang siklus hidup pengembangan.
8. Rencana Pelaksanaan Umum. Pada tahap ini, berikan rencana implementasi tingkat tinggi.

RINGKASAN BAB

- Berbeda dengan persyaratan sistem operasional, persyaratan gudang data tidak terlalu jelas.
- Data bisnis bersifat dimensional dan pengguna gudang data berpikir dalam dimensi bisnis.
- Oleh karena itu, definisi persyaratan untuk gudang data dapat didasarkan pada dimensi bisnis seperti produk, geografi, waktu, dan promosi.
- Paket informasi—sebuah konsep baru dan berguna—merupakan tulang punggung definisi kebutuhan. Paket informasi mencatat pengukuran penting atau fakta dan dimensi bisnis yang biasanya digunakan untuk menganalisis fakta.
- Wawancara, sesi kelompok, dan kuesioner adalah metode standar untuk mengumpulkan persyaratan.
- Orang-orang penting yang akan diwawancarai atau dilibatkan dalam sesi kelompok adalah eksekutif senior (termasuk sponsor), manajer departemen, analis bisnis, dan DBA sistem operasional.
- Meninjau seluruh dokumentasi sistem operasional terkait yang ada.
- Ruang lingkup dan isi dokumen definisi persyaratan meliputi sumber data, transformasi data, penyimpanan data, penyampaian informasi, dan diagram paket informasi.

Latihan Soal

1. Apa perbedaan penting antara mendefinisikan persyaratan untuk sistem operasional dan gudang data?
2. Jelaskan dimensi bisnis. Mengapa dan bagaimana dimensi bisnis berguna untuk menentukan persyaratan gudang data?
3. Data apa yang terkandung dalam paket informasi?
4. Apa yang dimaksud dengan hierarki dimensi? Berikan tiga contoh.
5. Jelaskan metrik atau fakta bisnis dengan lima contoh.
6. Buat daftar jenis pengguna yang harus diwawancarai untuk mengumpulkan persyaratan. Informasi apa yang dapat Anda peroleh dari mereka?
7. Dalam situasi apa metodologi JAD berhasil mengumpulkan persyaratan?
8. Mengapa peninjauan dokumen yang ada penting? Apa yang bisa Anda harapkan dari ulasan seperti itu?
9. Berbagai sumber data memberi makan gudang data. Informasi apa saja yang perlu Anda peroleh tentang sumber data?
10. Sebutkan lima komponen utama dokumen definisi persyaratan formal. Jelaskan apa yang masuk ke masing-masing komponen ini.

BAB 6

PERSYARATAN SEBAGAI KEKUATAN PENGGERAK DATA WAREHOUSING

TUJUAN BAB

- Memahami mengapa persyaratan bisnis adalah kekuatan pendorongnya
- Diskusikan bagaimana persyaratan mendorong setiap fase pengembangan
- Pelajari secara khusus bagaimana persyaratan memengaruhi desain data
- Meninjau dampak persyaratan pada arsitektur
- Perhatikan pertimbangan khusus untuk ETL dan metadata
- Periksa bagaimana persyaratan membentuk penyampaian informasi

Pada bab sebelumnya, kita membahas tahap definisi persyaratan secara rinci. Anda telah mempelajari bahwa mengumpulkan persyaratan untuk gudang data tidak sama dengan menentukan persyaratan untuk sistem operasional. Kami sampai pada cara baru untuk membuat paket informasi untuk mengekspresikan kebutuhan. Terakhir, kami menggabungkan semuanya dan menghasilkan dokumen definisi persyaratan.

Saat Anda merancang dan mengembangkan sistem apa pun, jelas bahwa sistem tersebut harus mencerminkan dengan tepat apa yang dibutuhkan pengguna untuk menjalankan proses bisnis mereka. Mereka harus memiliki layar GUI yang tepat, sistem harus memiliki logika yang benar untuk menjalankan fungsinya, dan pengguna harus menerima keluaran yang diperlukan termasuk layar dan laporan. Definisi persyaratan memandu seluruh proses desain dan pengembangan sistem.

Bagaimana dengan definisi persyaratan untuk gudang data? Jika definisi kebutuhan yang akurat penting untuk sistem operasional apa pun, hal ini jauh lebih penting untuk gudang data. Mengapa? Lingkungan data warehouse adalah sistem penyampaian informasi dimana pengguna sendiri akan mengakses repositori data warehouse dan membuat output mereka sendiri. Dalam sistem operasional, Anda memberikan keluaran yang telah ditentukan sebelumnya kepada pengguna.

Oleh karena itu, sangatlah penting bahwa gudang data Anda berisi elemen informasi yang tepat dalam format yang paling optimal. Pengguna Anda harus dapat menemukan semua informasi strategis yang mereka perlukan sesuai dengan keinginan mereka. Mereka harus dapat mengakses gudang data dengan mudah, menjalankan kueri, mendapatkan hasil tanpa kesulitan, dan melakukan berbagai jenis analisis data tanpa masalah.

Dalam gudang data, kebutuhan bisnis pengguna merupakan kekuatan pendorong tunggal dan paling kuat. Setiap tugas yang dilakukan di setiap fase pengembangan data warehouse ditentukan oleh persyaratan. Setiap keputusan yang dibuat selama tahap desain, apakah itu desain data, desain arsitektur, konfigurasi infrastruktur, atau skema metode penyampaian informasi, sepenuhnya dipengaruhi oleh persyaratan. Gambar 6.1 menggambarkan prinsip dasar ini.



Gambar 6.1 Persyaratan bisnis sebagai kekuatan pendorong.

Karena persyaratan merupakan kekuatan pendorong utama untuk setiap fase proses pengembangan, Anda perlu memastikan khususnya bahwa definisi kebutuhan Anda berisi rincian yang memadai untuk mendukung setiap fase. Bab ini secara khusus menyoroti beberapa kegiatan pembangunan yang signifikan dan menentukan bagaimana persyaratan harus memandu, mempengaruhi, dan mengarahkan kegiatan-kegiatan ini. Mengapa perhatian khusus seperti ini perlu dilakukan? Ketika Anda mengumpulkan persyaratan bisnis dan menghasilkan dokumen definisi persyaratan, Anda harus selalu ingat bahwa apa yang Anda lakukan dalam fase proyek ini sangat penting bagi setiap fase lainnya. Definisi kebutuhan Anda akan mendorong setiap fase proyek, jadi harap berikan perhatian khusus.

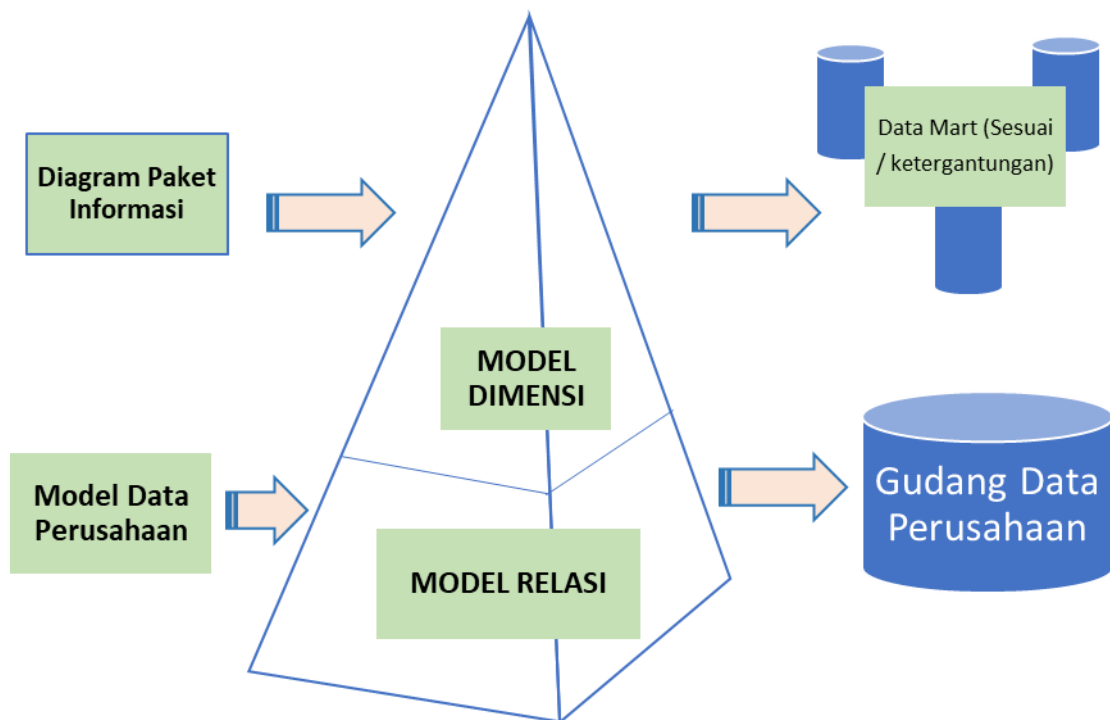
6.1 DESAIN DATA

Pada fase desain data, Anda membuat model data untuk repositori data berikut:

- Area pementasan tempat Anda mentransformasi, membersihkan, dan mengintegrasikan data dari sistem sumber sebagai persiapan untuk memuat ke dalam repositori gudang data
- Repositori gudang data itu sendiri

Jika Anda mengadopsi pendekatan praktis dalam membangun gudang data Anda sebagai konglomerasi data mart yang disesuaikan, model data Anda pada saat ini akan terdiri dari model data dimensional untuk kumpulan data mart pertama Anda. Di sisi lain, perusahaan Anda mungkin memutuskan untuk membangun gudang data besar di seluruh perusahaan terlebih dahulu bersama dengan data mart awal yang diumpungkan oleh gudang data besar.

Dalam hal ini, model data Anda akan mencakup model data untuk gudang data besar dan model data untuk data mart awal.



Gambar 6.2 Persyaratan yang menggerakkan model data.

Model data ini akan membentuk cetak biru untuk desain fisik dan implementasi repositori data. Anda akan menggunakan model ini untuk berkomunikasi di antara anggota tim tentang elemen data apa yang akan tersedia di gudang data dan bagaimana semuanya akan cocok satu sama lain. Anda akan mempelajari model data ini bersama pengguna untuk memberi tahu mereka tentang konten data dan hubungan data. Model data untuk data mart individual memainkan peran yang kuat dan berguna dalam komunikasi dengan pengguna.

Bagian mana dari definisi persyaratan yang mendorong desain data? Untuk memahami dampak persyaratan pada desain data, bayangkan model data sebagai piramida isi data seperti yang ditunjukkan pada Gambar 6.2. Bagian dasar piramida mewakili model data untuk penyimpanan data di seluruh perusahaan dan bagian atas piramida menunjukkan model data dimensional untuk data mart. Apa yang Anda perlukan dalam definisi persyaratan untuk membangun dan menggabungkan dua bagian piramida? Dua informasi dasar diperlukan: model data sistem sumber dan diagram paket informasi.

Model data dari sistem sumber saat ini akan digunakan untuk bagian bawah. Oleh karena itu, pastikan dokumen definisi kebutuhan Anda berisi informasi yang memadai tentang komponen dan hubungan data sistem sumber. Pada bab sebelumnya, kita telah membahas diagram paket informasi dengan cukup detail. Harap berhati-hati agar diagram paket informasi yang merupakan bagian dari dokumen definisi persyaratan benar-benar mencerminkan persyaratan bisnis sebenarnya. Jika tidak, model data Anda tidak akan menunjukkan apa yang sebenarnya ingin dilihat pengguna di gudang data.

Struktur Dimensi Bisnis

Dalam model data untuk data mart, dimensi bisnis yang digunakan pengguna untuk menganalisis metrik bisnis harus ditampilkan secara menonjol. Pada bab terakhir, saat membahas diagram paket informasi, kita meninjau beberapa contoh. Dalam diagram paket informasi, dimensi bisnis dicantumkan sebagai judul kolom. Misalnya, lihat dimensi bisnis Penjualan Produsen Mobil pada Gambar 6.3, yang merupakan reproduksi sebagian dari Gambar 5.5 sebelumnya.

Jika Anda membuat model data untuk data mart ini, dimensi bisnis seperti yang ditunjukkan pada gambar harus disertakan dalam model. Kegunaan data mart berhubungan langsung dengan keakuratan model data. Ke mana hal ini membawa Anda? Hal ini membawa Anda pada pentingnya memiliki dimensi yang sesuai dan konten yang tepat dalam diagram paket informasi.

Diagram Paket Informasi: Penjualan Produsen Mobil



Ukuran

Waktu	Produk	Cara Pembayaran	Demo Pelanggan-Grafik	Pedagang	
Tahun	Nama model	Jenis keuangan	Usia	Nama Dealer	
Triwulan	Model Tahun	Jangka waktu (Bulan)	Jenis kelamin	Kota	
Bulan	Penataan Paket	Suku Bunga	Kisaran pendapatan	Negara	
Tanggal	Lini Produk	Agen	Status pernikahan	Negara perakit	
Hari dalam minggu	Kategori Produk		Ukuran rumah tangga	Tanggal Operasi Pertama	
Hari dalam bulan	Warna Eksterior		Kepemilikan kendaraan		
Musim	Warna interior		Nilai Rumah		
Tanggal Merah	Tahun pertama		Sewa atau HM		
Metrik: Harga jual Aktual, MSRP, Harga Opsi, Harga Penuh, Dealer Tambahan, Kredit dealer, Fatur Dealer, Uang Muka					

Gambar 6.3 Dimensi bisnis dalam model data.

Struktur untuk Pengukuran Kunci

Pengukuran utama adalah metrik atau ukuran yang digunakan untuk analisis dan pemantauan bisnis. Pengguna mengukur kinerja dengan menggunakan dan membandingkan pengukuran utama. Untuk penjualan pembuat mobil, pengukuran utamanya mencakup harga jual aktual, MSRP, harga opsi, harga penuh, dan sebagainya. Pengguna mengukur kesuksesan mereka berdasarkan pengukuran utama. Mereka cenderung membuat perhitungan dan ringkasan berdasarkan metrik tersebut.

Selain mendapatkan hasil kueri berdasarkan kombinasi dimensi apa pun, fakta atau metrik digunakan untuk analisis. Saat pengguna Anda menganalisis penjualan berdasarkan dimensi produk, waktu, dan lokasi, mereka melihat hasil yang ditampilkan dalam metrik

seperti unit penjualan, pendapatan, biaya, dan margin keuntungan. Agar pengguna dapat meninjau hasil pengukuran kunci yang tepat, Anda harus menjamin bahwa diagram paket informasi yang Anda sertakan sebagai bagian dari definisi persyaratan berisi semua pengukuran kunci yang relevan.

Dimensi bisnis dan ukuran-ukuran utama membentuk tulang punggung model data dimensi. Struktur model data berhubungan langsung dengan jumlah dimensi bisnis. Konten data setiap dimensi bisnis merupakan bagian dari model data. Misalnya, jika diagram paket informasi memiliki produk, pelanggan, waktu, dan lokasi sebagai dimensi bisnis, keempat dimensi ini akan menjadi empat komponen berbeda dalam struktur model data. Selain dimensi bisnis, kelompok pengukuran utama juga membentuk komponen lain yang berbeda dari model data.

Tingkat Detil

Apa lagi yang harus tercermin dalam model data? Untuk menjawab pertanyaan ini, mari kita cermati bagaimana pengguna Anda berencana menggunakan gudang data untuk analisis. Mari kita ambil contoh spesifik. Analis senior ingin menganalisis penjualan di berbagai wilayah. Pertama, dia memulai dengan total penjualan nasional berdasarkan produk pada tahun ini. Kemudian langkah berikutnya adalah melihat total penjualan nasional berdasarkan produk di masing-masing wilayah sepanjang tahun tersebut. Selanjutnya, langkah selanjutnya adalah mengelompokkannya menjadi empat bagian. Setelah langkah ini, pengguna mungkin ingin mendapatkan perbandingan dengan anggaran dan kinerja tahun sebelumnya.

Apa yang kami amati adalah bahwa dalam analisis semacam ini Anda perlu menyediakan fasilitas penelusuran dan penelusuran untuk analisis. Apakah Anda ingin menyimpan data pada tingkat detail terendah? Jika demikian, bila pengguna Anda ingin melihat total nasional selama setahun penuh, sistem harus melakukan agregasi selama analisis saat pengguna menunggu di stasiun kerja. Di sisi lain, apakah Anda harus menyimpan detail untuk menampilkan data pada tingkat terendah, dan ringkasan untuk menampilkan data pada tingkat agregasi yang lebih tinggi?

Diskusi ini membawa kita ke aspek spesifik lain dari definisi persyaratan yang berkaitan dengan model data. Jika Anda memerlukan ringkasan di gudang data Anda, maka model data Anda harus menyertakan struktur untuk menyimpan detail serta ringkasan data. Jika Anda mampu membiarkan sistem menyimpulkan dengan cepat selama analisis, maka model data Anda tidak perlu memiliki struktur ringkasan. Cari tahu tentang fungsi penelusuran dan roll-up yang penting dan sertakan rincian yang cukup tentang jenis ringkasan dan tingkat detail data yang harus dimiliki gudang data Anda.

6.2 RENCANA ARSITEKTUR

Anda tahu bahwa arsitektur gudang data mengacu pada pengaturan komponen arsitektur yang tepat untuk mendapatkan manfaat maksimal. Bagaimana Anda merencanakan arsitektur gudang data Anda? Pada dasarnya, setiap data warehouse terdiri dari komponen yang hampir sama. Oleh karena itu, ketika Anda merencanakan arsitektur, Anda tidak menciptakan komponen baru apa pun untuk dimasukkan ke dalam gudang khusus Anda. Anda

benar-benar mengukur setiap komponen untuk lingkungan Anda. Anda merencanakan bagaimana seluruh komponen harus dirangkai sehingga dapat bekerja sebagai suatu sistem yang terintegrasi.

Sebelum kita melangkah lebih jauh, mari kita rekap komponen-komponen arsitektur utama seperti yang dibahas dalam Bab 2:

- ❖ Sumber data
- ❖ Data produksi
- ❖ data dalaman
- ❖ Data yang diarsipkan
- ❖ Data eksternal
- ❖ Pementasan data
- ❖ Ekstraksi data
- ❖ Transformasi data
- ❖ Memuat data
- ❖ Penyimpanan data
- ❖ Penyampaian informasi
- ❖ Metadata
- ❖ Manajemen dan pengendalian

Saat Anda merencanakan arsitektur keseluruhan untuk gudang data Anda, Anda akan mengatur cakupan dan konten dari masing-masing komponen ini. Misalnya, di perusahaan Anda, semua sumber data utamanya mungkin berada pada platform komputasi tunggal dan juga pada database relasional tunggal. Jika hal ini terjadi, maka komponen ekstraksi data dalam arsitektur akan jauh lebih kecil dan mudah. Sekali lagi, jika perusahaan Anda memutuskan untuk hanya menggunakan fasilitas yang disediakan oleh DBMS, seperti definisi alias dan fitur komentar, untuk penyimpanan metadata, maka komponen metadata Anda akan sederhana.

Oleh karena itu, perencanaan arsitektur melibatkan peninjauan setiap komponen berdasarkan konteks khusus Anda, dan menetapkan parameternya. Juga, ini melibatkan antarmuka antara berbagai komponen. Bagaimana modul manajemen dan pengendalian dirancang untuk mengoordinasikan dan mengendalikan fungsi berbagai komponen? Informasi apa yang Anda perlukan untuk melakukan perencanaan? Bagaimana Anda mengetahui cara mengukur setiap komponen dan menyediakan infrastruktur yang sesuai untuk mendukungnya? Tentu saja jawabannya adalah kebutuhan bisnis. Semua informasi yang Anda perlukan untuk merencanakan arsitektur harus berasal dari definisi persyaratan. Pada subbagian berikut, kita akan mengeksplorasi pentingnya persyaratan bisnis untuk rencana arsitektur. Kami akan mengambil setiap komponen dan meninjau bagaimana persyaratan yang tepat mendorong ukuran dan konten gudang data.

Komposisi Komponen

Mari kita tinjau setiap komponen dan memastikan apa sebenarnya yang diperlukan dalam definisi persyaratan untuk merencanakan arsitektur gudang data. Sekali lagi, ingatlah bahwa perencanaan arsitektur melibatkan penentuan ukuran dan isi setiap komponen. Dalam

daftar berikut, poin-poin di bawah masing-masing komponen menunjukkan jenis informasi yang harus terkandung dalam definisi persyaratan untuk menggerakkan rencana arsitektur.

1. Sumber data

Sistem sumber operasional

Platform komputasi, sistem operasi, database, file Data departemen seperti file, dokumen, dan spreadsheet Sumber data eksternal

2. Pementasan Data

Pemetaan data antara sumber data dan struktur data area pementasan

Transformasi data

Pembersihan data Integrasi data

3. Penyimpanan data

Ukuran fitur DBMS data yang diekstraksi dan terintegrasi

Potensi pertumbuhan

Terpusat untuk gudang data perusahaan

Data mart: sesuai, bergantung, independen, terfederasi

4. Penyampaian Informasi

Jenis dan jumlah pengguna

Jenis pertanyaan dan laporan

Kelas analisis

Dasbor/kartu skor

Operasi penambangan data

Aplikasi DSS ujung depan

5. Metadata

Metadata operasional

Metadata ETL (ekstraksi data, transformasi, dan pemuatan).

Metadata pengguna akhir

Penyimpanan metadata

6. Manajemen dan Pengendalian

Memuat data

Sumber eksternal

Sistem peringatan

Pengiriman informasi pengguna akhir

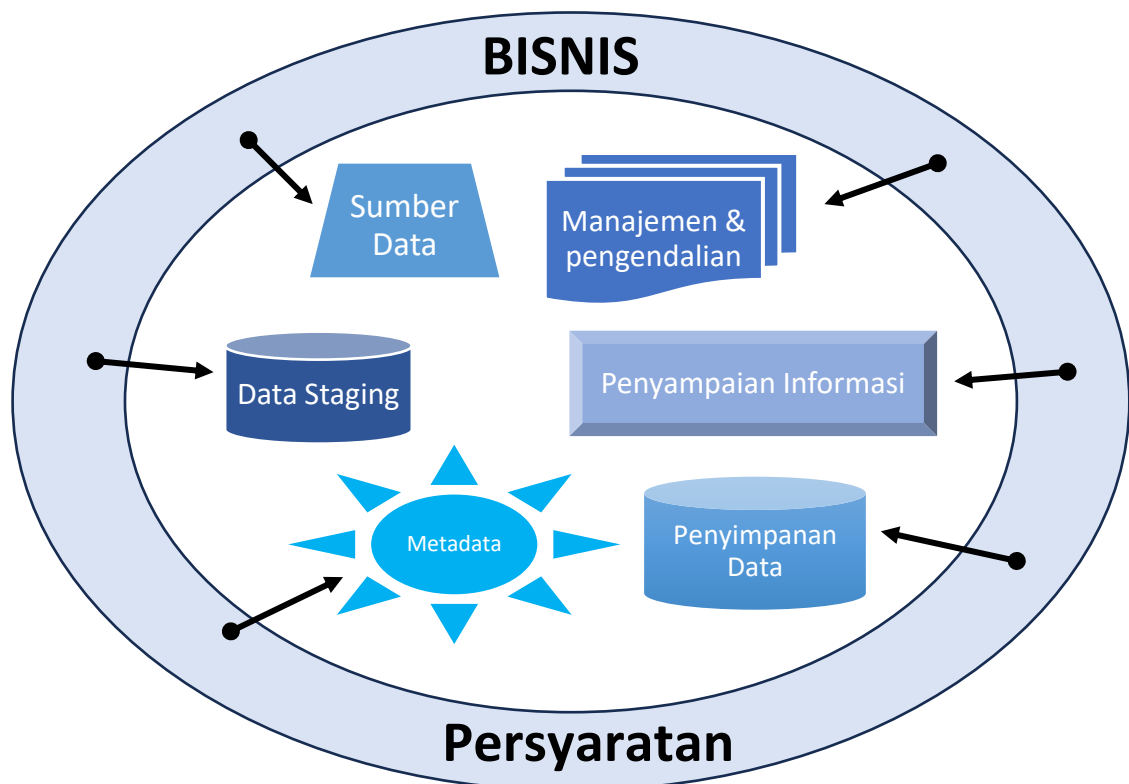
Gambar 6.4 memberikan ringkasan berguna mengenai komponen arsitektural yang didorong oleh kebutuhan. Angka tersebut menunjukkan dampak kebutuhan bisnis pada arsitektur data warehouse.

Pertimbangan Khusus

Setelah meninjau dampak persyaratan pada komponen arsitektur secara rinci, sekarang kita mengalihkan perhatian kita ke beberapa fungsi yang patut mendapat pertimbangan khusus. Kita perlu mengemukakan pertimbangan khusus ini karena jika hal ini tidak dimasukkan dalam definisi persyaratan, konsekuensi serius akan terjadi. Saat Anda

berada dalam fase definisi persyaratan, Anda harus memberikan perhatian khusus pada faktor-faktor ini.

Ekstraksi, Transformasi, dan Pemuatan Data (ETL) Aktivitas yang berkaitan dengan ETL di gudang data sejauh ini merupakan aktivitas yang paling memakan waktu dan intensif manusia. Pengakuan khusus terhadap tingkat dan kompleksitas aktivitas ini dalam persyaratan akan sangat membantu dalam meringankan kesulitan saat menyiapkan arsitektur. Jika Anda berencana untuk menyimpan data secara real-time, maka fungsi ETL memerlukan perhatian khusus. Teknik harus diterapkan untuk ekstraksi data waktu nyata, transformasi cepat dan integrasi data, serta pembaruan gudang data secara waktu nyata dengan latensi mendekati nol. Mari kita pisahkan fungsinya dan nyatakan pertimbangan khusus yang diperlukan dalam definisi persyaratan.



Gambar 6.4 Dampak persyaratan pada arsitektur.

Ekstraksi Data Identifikasi dengan jelas semua sumber data internal. Tentukan semua platform komputasi dan file sumber dari mana data akan diekstraksi. Jika Anda akan menyertakan sumber data eksternal, tentukan kompatibilitas struktur data Anda dengan sumber luar. Tunjukkan juga metode ekstraksi data.

Transformasi Data Banyak jenis fungsi transformasi yang diperlukan sebelum data dapat dipetakan dan disiapkan untuk dimuat ke dalam repositori data warehouse. Fungsi-fungsi ini meliputi pemilihan masukan, pemisahan struktur masukan, normalisasi dan denormalisasi struktur sumber, agregasi, konversi, penyelesaian nilai yang hilang, dan konversi nama dan alamat. Dalam praktiknya, ini ternyata merupakan daftar fungsi yang

panjang dan rumit. Periksa setiap elemen data yang direncanakan untuk disimpan di gudang data terhadap elemen data sumber dan pastikan pemetaan dan transformasinya.

Pemuatan Data Tentukan pemuatan awal. Tentukan seberapa sering setiap kelompok data utama harus selalu diperbarui di gudang data. Berapa banyak pembaruan yang akan dilakukan setiap malam? Apakah lingkungan Anda memerlukan lebih dari satu siklus pembaruan dalam sehari? Bagaimana perubahan akan ditangkap dalam sistem sumber? Tentukan bagaimana pembaruan harian, mingguan, dan bulanan akan dimulai dan dilaksanakan. Jika paket Anda mencakup pergudangan data waktu nyata, tentukan metode untuk pembaruan waktu nyata.

Kualitas Data Data yang buruk menyebabkan keputusan yang buruk. Tidak peduli seberapa baik Anda menyempurnakan gudang data Anda, dan tidak peduli seberapa mahir Anda menyediakan fungsi kueri dan analisis kepada pengguna, jika kualitas data gudang data Anda dicurigai, pengguna akan segera kehilangan kepercayaan dan meninggalkan gudang data. Bahkan perbedaan kecil sekalipun dapat mengakibatkan dampak serius dalam pengambilan keputusan strategis yang memiliki konsekuensi luas. Kualitas data dalam gudang data adalah sesuatu yang sakral. Oleh karena itu, tepat pada tahap awal definisi persyaratan, identifikasi sumber potensi polusi data dalam sistem sumber. Selain itu, waspadai semua kemungkinan jenis masalah kualitas data yang mungkin ditemui dalam sistem operasional Anda. Perhatikan tips berikut ini.

Sumber Polusi Data

Konversi dan migrasi sistem

Integrasi sistem heterogen

Desain database sistem sumber yang tidak memadai

Penuaan data

Informasi yang tidak lengkap dari pelanggan

Kesalahan masukan Internasionalisasi/lokalisasi sistem

Kurangnya kebijakan/prosedur pengelolaan data

Jenis Masalah Kualitas Data

Nilai tiruan di bidang sistem sumber

Tidak adanya data di bidang sistem sumber

Bidang serbaguna

Data samar Data yang bertentangan

Penggunaan baris nama dan alamat yang tidak tepat

Pelanggaran aturan bisnis

Kunci utama yang digunakan kembali

Pengidentifikasi tidak unik

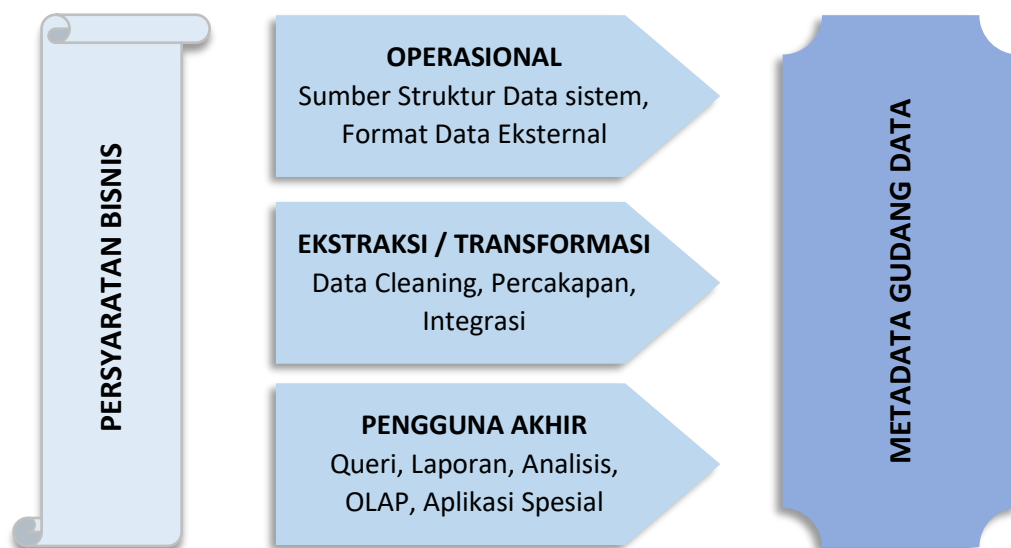
Metadata Anda sudah tahu bahwa metadata dalam data warehouse bukan sekedar entri kamus data. Metadata dalam gudang data lebih dari sekedar rincian yang dapat dibawa

dalam kamus data atau katalog data. Metadata bertindak sebagai perekat untuk mengikat semua komponen menjadi satu. Saat data berpindah dari satu komponen ke komponen lainnya, perpindahan tersebut diatur oleh bagian metadata yang relevan. Saat pengguna menanyakan gudang data, metadata bertindak sebagai sumber informasi untuk menghubungkan parameter kueri dengan komponen database.

Sebelumnya, kami telah mengkategorikan metadata di gudang data menjadi tiga kelompok: operasional, ekstraksi dan transformasi data, dan pengguna akhir. Gambar 6.5 menampilkan dampak kebutuhan bisnis pada komponen arsitektur metadata. Pentingnya komponen metadata tidak perlu ditegaskan kembali. Pelajari gambar tersebut dan terapkan pada proyek gudang data Anda. Untuk setiap jenis metadata, cari tahu seberapa banyak detail yang diperlukan dalam definisi kebutuhan Anda. Memiliki detail yang cukup untuk memungkinkan pengambilan keputusan penting seperti memilih jenis repositori metadata dan memperhitungkan apakah repositori harus dipusatkan atau didistribusikan.

Alat dan Produk

Ketika alat disebutkan dalam konteks pergudangan data, Anda mungkin hanya memikirkan alat pengguna akhir. Banyak orang melakukannya. Namun untuk membangun dan memelihara gudang data, Anda memerlukan berbagai jenis alat untuk mendukung berbagai komponen arsitektur.



Gambar 6.5 Dampak persyaratan pada metadata.

Saat kita membahas dampak persyaratan pada arsitektur data warehouse di bagian ini, kami ingin mengangkat topik alat dan produk karena dua alasan. Pertama, persyaratan tidak berdampak langsung pada pemilihan alat. Jangan memilih alat berdasarkan kebutuhan dan kemudian menyesuaikan arsitektur agar sesuai dengan alat tersebut. Ini seperti meletakkan kereta di depan kudanya. Rancang arsitektur gudang data dan kemudian cari alat yang tepat untuk mendukung arsitektur tersebut. Alat tertentu, yang cocok untuk fungsi di satu gudang data, mungkin tidak cocok di gudang data lain. Itu karena arsitekturnya berbeda. Apa yang kami maksud dengan pernyataan bahwa arsitekturnya berbeda? Meskipun

komponen arsitektur pada kedua gudang data secara umum sama, cakupan, ukuran, konten, dan komposisi masing-masing komponen tidaklah sama.

Alasan kedua untuk menyebutkan alat dan produk adalah ini. Saat mengumpulkan persyaratan untuk merencanakan arsitektur, terkadang Anda mungkin merasa terkendala untuk membuat arsitektur sesuai dengan persyaratan. Anda mungkin berpikir bahwa Anda tidak akan dapat merancang jenis arsitektur yang ditentukan oleh persyaratan karena alat yang sesuai untuk mendukung jenis arsitektur tersebut mungkin tidak tersedia. Harap dicatat bahwa ada banyak alat layak produksi yang tersedia di pasar. Kami ingin menunjukkan bahwa setelah desain arsitektur Anda selesai, Anda bisa mendapatkan alat dan produk pihak ketiga yang paling sesuai.

Secara umum, alat tersedia untuk fungsi-fungsi berikut:

- ❖ Ekstraksi Data dan Transformasi
 - Perangkat Tengah
 - Ekstraksi data
 - Transformasi data
 - Jaminan kualitas data
 - Muat pembuatan gambar
- ❖ Gudang Penyimpanan
 - Pasar data
 - data meta
- ❖ Akses/Penyampaian Informasi
 - Penulis laporan
 - Pemroses kueri
 - Sistem peringatan
 - Dasbor
 - Kartu Skor
 - aplikasi DSS
 - Penambangan data

6.3 SPESIFIKASI PENYIMPANAN DATA

Jika perusahaan Anda mengadopsi pendekatan top-down dalam mengembangkan gudang data, maka Anda harus menentukan spesifikasi penyimpanannya

- Area pementasan data
- Gudang data perusahaan secara keseluruhan
- Masing-masing data mart dependen, dimulai dari yang pertama
- Setiap database multidimensi untuk OLAP

Alternatifnya, jika perusahaan Anda memilih pendekatan bottom-up, Anda memerlukan spesifikasinya

- Area pementasan data
- Setiap data mart yang disesuaikan, dimulai dari yang pertama

- Setiap database multidimensi untuk OLAP

Biasanya, gudang data perusahaan secara keseluruhan akan didasarkan pada model relasional yang didukung oleh sistem manajemen basis data relasional (RDBMS). Data mart biasanya terstruktur pada model dimensi yang diimplementasikan menggunakan RDBMS. Banyak vendor menawarkan sistem database multidimensi berpemilik (MDDDB). Spesifikasi MDDDB Anda akan didasarkan pada vendor pilihan Anda. Luas dan canggihnya area pementasan bergantung pada kompleksitas dan luasnya transformasi, pembersihan, dan konversi data. Area pementasan mungkin hanya sekumpulan file datar atau, di sisi lain, database relasional yang dikembangkan sepenuhnya.

Apa pun pilihan sistem manajemen basis data Anda, sistem tersebut harus berinteraksi dengan alat back-end dan front-end. Alat back-end adalah produk untuk transformasi data, pembersihan data, dan pemuatan data. Alat front-end berhubungan dengan penyampaian informasi kepada pengguna. Jika Anda mencoba mencari alat terbaik yang sesuai dengan lingkungan Anda, kemungkinan besar alat tersebut tidak berasal dari vendor yang sama yang memasok produk database. Oleh karena itu, salah satu kriteria penting untuk sistem manajemen basis data adalah sistemnya harus terbuka. Itu harus kompatibel dengan alat back-end dan front-end yang dipilih. Baru-baru ini banyak produk peralatan gudang data komposit tersedia. Ini adalah bundel perangkat keras/perangkat lunak yang terintegrasi.

Jadi apa yang kami katakan tentang dampak kebutuhan bisnis terhadap spesifikasi penyimpanan data? Persyaratan bisnis menentukan seberapa kuat dan terbukanya sistem database. Saat menentukan persyaratan, ingatlah pengaruhnya terhadap spesifikasi penyimpanan data dan kumpulkan semua detail yang diperlukan tentang komponen arsitektur back-end dan front-end.

Seleksi DBMS

Pada fase definisi persyaratan, ketika Anda mewawancarai pengguna dan mengadakan pertemuan formal dengan mereka, Anda tidak secara khusus mendiskusikan jenis DBMS yang akan dipilih. Namun, banyak persyaratan pengguna mempengaruhi pemilihan DBMS yang tepat. Produk DBMS relasional yang ada di pasaran biasanya digabungkan dengan seperangkat alat untuk memproses pertanyaan, menulis laporan, berinteraksi dengan produk lain, dan sebagainya. Pilihan Anda terhadap DBMS mungkin ditentukan oleh komponen tool kit-nya. Dan kebutuhan bisnis kemungkinan besar menentukan jenis komponen tool kit yang dibutuhkan. Secara umum, elemen persyaratan bisnis berikut mempengaruhi pilihan DBMS:

- Tingkat Pengalaman Pengguna:** Jika pengguna sama sekali tidak berpengalaman dengan sistem basis data, DBMS harus memiliki fitur untuk memantau dan mengendalikan permintaan yang tidak terkendali. Di sisi lain, jika banyak pengguna Anda adalah pengguna yang mahir, maka mereka akan merumuskan pertanyaan mereka sendiri. Dalam hal ini, DBMS harus mendukung antarmuka bahasa tipe SQL yang mudah.
- Jenis Kueri:** DBMS harus memiliki pengoptimal yang kuat jika sebagian besar kueri bersifat kompleks dan menghasilkan kumpulan hasil yang besar. Alternatifnya, jika

- terdapat campuran kueri sederhana dan kompleks, harus ada semacam manajemen kueri dalam perangkat lunak database untuk menyeimbangkan eksekusi kueri.
- C. **Perlunya Keterbukaan:** Tingkat keterbukaan bergantung pada komponen arsitektur back-end dan front-end, yang pada gilirannya bergantung pada kebutuhan bisnis.
 - D. **Beban Data:** Volume data dan frekuensi pemuatan menentukan kekuatan di bidang pemuatan data, pemulihan, dan memulai ulang.
 - E. **Manajemen Metadata:** Jika komponen metadata Anda tidak harus rumit, maka DBMS dengan kamus data aktif mungkin sudah cukup. Biarkan definisi kebutuhan Anda mencerminkan jenis dan cakupan kerangka metadata.
 - F. **Lokasi Penyimpanan Data:** Apakah gudang data Anda akan berada di satu lokasi pusat, atau akan didistribusikan? Jawaban atas pertanyaan ini akan menentukan apakah DBMS yang dipilih harus mendukung database terdistribusi.
 - G. **Pertumbuhan Gudang Data:** Definisi kebutuhan bisnis Anda harus berisi informasi tentang perkiraan pertumbuhan jumlah pengguna, serta jumlah dan kompleksitas kueri. Estimasi pertumbuhan akan mempunyai hubungan langsung dengan bagaimana DBMS yang dipilih mendukung skalabilitas.

Ukuran Penyimpanan

Seberapa besar gudang data Anda? Berapa banyak penyimpanan yang diperlukan untuk semua repositori data? Berapa total ukuran penyimpanannya? Jawaban atas pertanyaan-pertanyaan ini akan mempengaruhi jenis dan ukuran media penyimpanan. Bagaimana Anda menemukan jawaban atas pertanyaan-pertanyaan ini? Sekali lagi, ini kembali ke kebutuhan bisnis. Dalam definisi persyaratan, Anda harus memiliki informasi yang cukup untuk menjawab pertanyaan-pertanyaan ini.

Mari kita rangkum. Anda perlu memperkirakan ukuran penyimpanan berikut ini dalam fase definisi persyaratan:

- 1) **Area Pementasan Data:** Hitung perkiraan penyimpanan untuk area pementasan data gudang data perusahaan secara keseluruhan dari ukuran struktur data sistem sumber untuk setiap subjek bisnis. Bayangkan transformasi data dan pemetaannya ke dalam perhitungan Anda. Untuk data mart, awalnya perkirakan penyimpanan area pementasan berdasarkan dimensi dan metrik bisnis untuk data mart pertama.
- 2) **Gudang Data Perusahaan Secara Keseluruhan:** Perkirakan ukuran penyimpanan berdasarkan struktur data untuk setiap subjek bisnis. Anda tahu bahwa data di gudang data disimpan oleh pelaku bisnis. Untuk setiap subjek bisnis, buat daftar berbagai atribut, perkirakan panjang bidangnya, dan dapatkan perhitungan penyimpanan yang diperlukan untuk subjek tersebut.
- 3) **Data Mart:** Sesuai, Independen, Bergantung, atau Terfederasi. Saat menentukan persyaratan, Anda membuat diagram informasi. Satu set diagram ini merupakan data mart. Setiap diagram informasi berisi dimensi bisnis dan atributnya. Diagram informasi juga menyimpan metrik atau pengukuran bisnis yang dimaksudkan untuk analisis. Gunakan rincian dimensi bisnis dan ukuran bisnis yang ditemukan dalam diagram

informasi untuk memperkirakan ukuran penyimpanan data mart. Mulailah dengan data mart pertama Anda.

- 4) **Basis Data Multidimensi:** Basis data ini mendukung OLAP atau analisis multidimensi. Berapa banyak pemrosesan analitis online (OLAP) yang diperlukan bagi pengguna Anda? Gudang data perusahaan atau data mart individu yang disesuaikan atau bergantung memasok data untuk database multidimensi. Cari tahu detail OLAP yang direncanakan untuk pengguna Anda, lalu gunakan detail tersebut untuk memperkirakan penyimpanan database multidimensi ini.

6.4 STRATEGI PENYAMPAIAN INFORMASI

Dampak persyaratan bisnis pada mekanisme penyampaian informasi di gudang data sangatlah jelas. Selama fase definisi persyaratan, pengguna memberi tahu Anda informasi apa yang ingin mereka ambil dari gudang data. Anda mencatat persyaratan ini dalam dokumen definisi persyaratan. Anda kemudian menyediakan semua fitur dan konten yang diinginkan dalam komponen penyampaian informasi. Apakah ini terdengar sederhana dan lugas? Meskipun dampaknya tampak jelas dan sederhana, ada beberapa permasalahan yang perlu dipertimbangkan. Banyak aspek persyaratan yang berbeda berdampak pada berbagai elemen komponen penyampaian informasi dengan cara yang berbeda.

Komposisi komunitas pengguna yang diharapkan menggunakan intelijen bisnis dari gudang data mempengaruhi strategi penyampaian informasi. Apakah sebagian besar pengguna potensial gudang data adalah pengguna dan analis yang mahir? Maka strategi informasi harus diarahkan untuk menyediakan alat analisis yang ampuh. Apakah banyak pengguna yang mengharapkan untuk menerima laporan yang telah diformat sebelumnya dan menjalankan kueri yang telah disusun sebelumnya? Kemudian fasilitas query dan pelaporan pada komponen penyampaian informasi harus diperkuat.

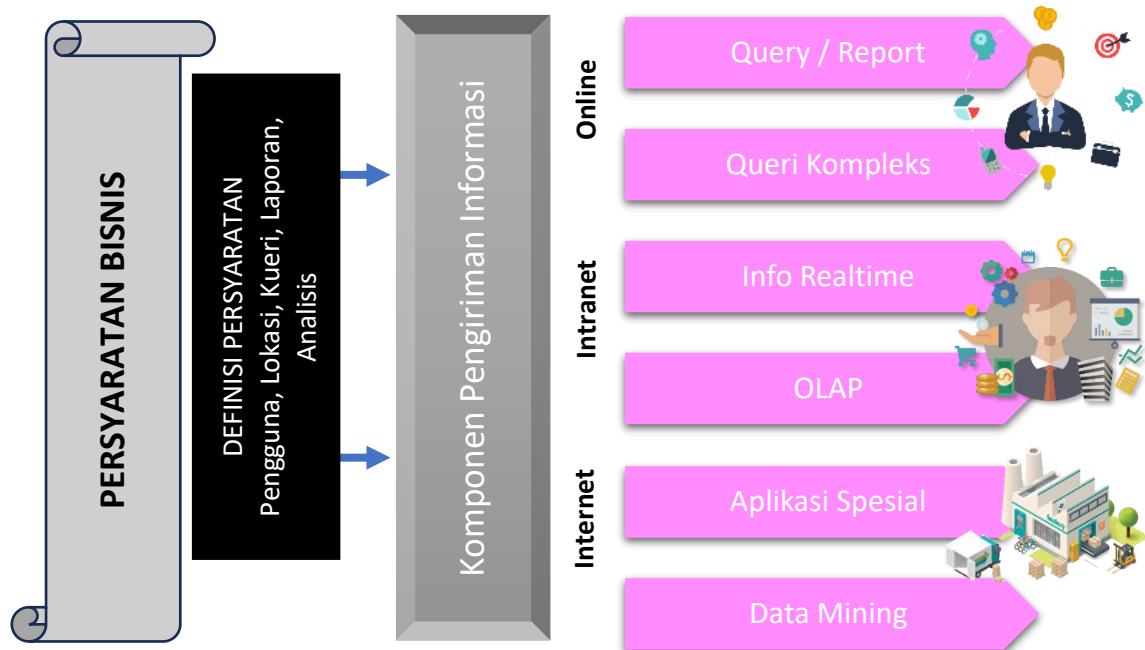
Area luas komponen penyampaian informasi yang terkena dampak langsung oleh kebutuhan bisnis adalah:

- ◆ Pertanyaan dan laporan
- ◆ Jenis analisis
- ◆ Distribusi informasi
- ◆ Pengiriman informasi secara real-time
- ◆ Aplikasi pendukung keputusan
- ◆ Pertumbuhan dan ekspansi

Gambar 6.6 menunjukkan dampak kebutuhan bisnis terhadap penyampaian informasi.

Gudang data ada karena satu alasan dan satu alasan saja—untuk menyediakan informasi strategis kepada pengguna. Penyampaian informasi menempati urutan teratas dalam daftar komponen arsitektur. Sebagian besar komponen lainnya bersifat transparan bagi pengguna, namun mereka melihat dan merasakan apa yang tersedia bagi mereka dalam komponen penyampaian informasi. Pentingnya persyaratan bisnis yang berkaitan dengan penyampaian informasi tidak dapat terlalu ditekankan. Subbagian berikut berisi beberapa tip

berharga untuk definisi persyaratan agar komponen penyampaian informasi yang sangat penting menjadi efektif dan berguna. Pelajarilah ini dengan cermat.



Gambar 6.6 Dampak kebutuhan bisnis terhadap penyampaian informasi.

Pertanyaan dan Laporan

Cari tahu siapa yang akan menggunakan kueri yang telah ditentukan sebelumnya dan laporan yang telah diformat sebelumnya. Dapatkan spesifikasinya. Dapatkan juga spesifikasi produksi dan frekuensi distribusi laporan. Berapa banyak pengguna yang akan menjalankan kueri yang telah ditentukan sebelumnya? Seberapa sering? Jenis kueri kedua bukanlah kumpulan kueri yang telah ditentukan sebelumnya. Dalam hal ini, pengguna merumuskan kueri mereka sendiri dan mereka sendiri yang menjalankan kueri tersebut. Juga di kelas ini adalah kumpulan laporan dimana pengguna menyediakan parameter laporan dan mencetak sendiri laporan yang cukup canggih. Dapatkan sebanyak mungkin detail jenis kueri dan kumpulan laporan jenis ini.

Pengguna tingkat lanjut dapat menjalankan kueri yang kompleks, sebagian besar waktunya sebagai bagian dari sesi analisis interaktif. Selain analisis, apakah pengguna tingkat lanjut Anda memerlukan kemampuan untuk menjalankan satu kueri kompleks? Temukan.

Jenis Analisis

Sebagian besar lingkungan gudang data dan intelijen bisnis menyediakan beberapa fitur untuk menjalankan sesi interaktif dan melakukan analisis data yang kompleks. Analisis yang mencakup metode penelusuran hingga penelusuran (drill-down) dan roll-up (roll-up) merupakan hal yang cukup umum. Tinjau bersama pengguna Anda semua jenis analisis yang ingin mereka lakukan. Dapatkan informasi tentang kompleksitas jenis analisis yang diantisipasi.

Selain analisis yang dilakukan secara langsung pada data mart, sebagian besar gudang data dan lingkungan intelijen bisnis saat ini membekali pengguna dengan OLAP. Dengan menggunakan fasilitas OLAP, pengguna dapat melakukan analisis multidimensi dan memperoleh berbagai tampilan data dari database multidimensi. Jenis analisis ini disebut mengiris dan memotong. Perkirakan sifat dan luas fasilitas penelusuran dan penyiapan yang akan disediakan. Tentukan berapa banyak pemotongan dan pemotongan yang harus tersedia.

Sebagian besar lingkungan intelijen bisnis saat ini menyediakan informasi melalui dasbor dan kartu skor. Jadi, kebutuhan informasi Anda harus mencakup rincian tentang komposisi kedua sarana penyampaian informasi tersebut. Pengguna dapat bekerja dengan dasbor secara interaktif dan melakukan berbagai jenis analisis.

Distribusi Informasi

Di mana pengguna Anda? Apakah mereka berada di satu lokasi? Apakah mereka berada dalam satu situs lokal yang terhubung dengan jaringan area lokal (LAN)? Apakah mereka tersebar di jaringan area luas (WAN)? Faktor-faktor ini menentukan bagaimana informasi harus didistribusikan kepada pengguna Anda. Tunjukkan dengan jelas rincian ini dalam definisi persyaratan.

Di banyak perusahaan, pengguna mendapatkan akses informasi melalui intranet perusahaan. Teknologi berbasis web digunakan. Jika hal ini terjadi di perusahaan Anda, teknologi berbasis web harus dimasukkan ke dalam komponen penyampaian informasi. Biarkan definisi kebutuhan Anda menjelaskan secara eksplisit tentang faktor-faktor ini.

Pengiriman Informasi Waktu Nyata

Jika gudang data Anda dikonfigurasi menjadi gudang data waktu nyata di mana pembaruan data dilakukan secara waktu nyata atau mendekati waktu nyata, maka kebutuhan informasi Anda harus mempunyai pertimbangan khusus untuk penyampaian informasi waktu nyata. Pengiriman informasi real-time secara langsung menyiratkan lebih banyak informasi dengan lebih banyak orang.

Gudang data Anda harus memungkinkan orang-orang dengan informasi yang mereka perlukan untuk membuat keputusan penting sepanjang hari, setiap hari, dan harus menyampaikan informasi ini dengan cara yang hemat biaya. Untuk menghadapi tantangan ini, Anda harus mampu menerapkan sistem penyampaian informasi ke sejumlah besar orang di seluruh perusahaan, dengan cepat dan dengan biaya kepemilikan yang rendah. Anda harus dapat memberikan informasi, secara real-time, kepada semua orang yang mempengaruhi proses bisnis utama: karyawan, manajemen, pemasok, mitra, dan pelanggan. Bagaimana kamu melakukan ini? Kumpulkan detailnya dalam fase pengumpulan persyaratan.

Faktor-faktor berikut memerlukan pertimbangan khusus dalam definisi kebutuhan Anda: penyampaian informasi kepada orang-orang sesuai keinginan mereka—dalam format yang paling umum digunakan, email, halaman Web, Excel, PDF, dan dokumen pengolah kata; penerapan pendekatan yang dipersonalisasi dan terorganisir terhadap intelijen bisnis yang menyampaikan informasi secara luas ke seluruh perusahaan; penggunaan teknologi berbasis web untuk memenuhi permintaan informasi strategis yang terus meningkat.

Aplikasi Pendukung Keputusan

Ini adalah aplikasi khusus yang dirancang untuk mendukung kelompok pengguna individu untuk tujuan tertentu. Sistem informasi eksekutif menyediakan dukungan keputusan kepada eksekutif senior. Aplikasi data mining adalah sistem dengan tujuan khusus untuk menemukan pola hubungan baru dan kemungkinan prediktif. Kita akan membahas data mining secara lebih rinci di Bab 17.

Gudang data memasok data untuk aplikasi pendukung keputusan ini. Terkadang desain dan pengembangan aplikasi ad hoc ini berada di luar lingkup proyek gudang data. Salah satunya koneksi dengan data warehouse adalah pemberian data dari repositori data warehouse. Apa pun strategi pengembangan untuk aplikasi pendukung keputusan khusus di perusahaan Anda, pastikan definisi persyaratan menjelaskan secara rinci. Jika gudang data akan digunakan hanya untuk pengumpulan data, tentukan elemen data dan frekuensi pergerakan data.

Pertumbuhan dan Ekspansi

Katakanlah gudang data Anda telah dikerahkan. Anda telah memberi pengguna Anda kemampuan untuk menjalankan kueri, mencetak laporan, melakukan analisis, menggunakan OLAP untuk analisis kompleks, dan memberi data pada aplikasi khusus. Komponen penyampaian informasi sudah lengkap dan berfungsi dengan baik. Apakah itu akhir dari usahanya? Ya, mungkin hanya untuk iterasi pertama.

Komponen penyampaian informasi terus tumbuh dan berkembang. Jumlah dan kompleksitas kueri dan laporan terus bertambah. Ini memperluas peningkatan pada setiap bagian komponen. Dalam definisi kebutuhan awal Anda, Anda perlu mengantisipasi pertumbuhan dan perluasan. Detail yang cukup tentang pertumbuhan dan perluasan dapat memengaruhi desain komponen penyampaian informasi yang tepat, jadi kumpulkan detail yang cukup untuk memperkirakan pertumbuhan dan peningkatan.

RINGKASAN BAB

- Definisi kebutuhan yang akurat dalam proyek gudang data jauh lebih penting dibandingkan jenis proyek lainnya. Memahami dengan jelas dampak persyaratan bisnis pada setiap fase pengembangan.
- Persyaratan bisnis mengkondisikan hasil tahap desain data.
- Setiap komponen arsitektur data warehouse sangat dipengaruhi oleh kebutuhan bisnis.
- Untuk memberikan kualitas data, identifikasi sumber polusi data, jenis masalah kualitas yang umum terjadi, dan cara untuk menghilangkan kerusakan data sejak awal pada tahap definisi persyaratan itu sendiri.
- Spesifikasi penyimpanan data, khususnya pemilihan DBMS, ditentukan oleh kebutuhan bisnis. Pastikan Anda mengumpulkan detail yang cukup relevan selama fase persyaratan.

- Persyaratan bisnis sangat mempengaruhi mekanisme penyampaian informasi. Persyaratan menentukan bagaimana, kapan, dan di mana pengguna akan menerima informasi dari gudang data.

Latihan soal

1. “Dalam gudang data, kebutuhan bisnis pengguna merupakan kekuatan pendorong tunggal dan paling kuat.” Apa kamu setuju? Jika ya, sebutkan empat alasannya. Jika tidak, apakah ada kekuatan pendorong lain yang serupa?
2. Bagaimana diagram informasi yang akurat berubah menjadi model data yang baik untuk data mart Anda? Jelaskan secara singkat.
3. Sebutkan lima komponen arsitektur yang sangat dipengaruhi oleh kebutuhan bisnis. Jelaskan dampak persyaratan bisnis pada salah satu dari lima komponen tersebut.
4. Apa dampak persyaratan terhadap pemilihan alat dan produk vendor? Apakah persyaratan secara langsung menentukan pilihan alat?
5. Sebutkan empat aspek penyampaian informasi yang secara langsung dipengaruhi oleh kebutuhan bisnis. Untuk dua aspek tersebut, jelaskan dampaknya.
6. Bagaimana kebutuhan bisnis mempengaruhi pilihan DBMS? Jelaskan tiga cara yang mempengaruhi pemilihan DBMS.
7. Apa itu MDDB? Jenis persyaratan bisnis apa yang menentukan penggunaan MDDB di gudang data?
8. Bagaimana persyaratan mempengaruhi pilihan kerangka metadata? Jelaskan dengan sangat singkat.
9. Jenis kebutuhan pengguna apa yang menentukan granularitas atau tingkat detail dalam gudang data?
10. Bagaimana Anda memperkirakan ukuran penyimpanan? Faktor apa saja yang menentukan besar kecilnya?

BAB 7

KOMPONEN ARSITEKTUR

TUJUAN BAB

- Memahami arsitektur data warehouse
- Pelajari tentang komponen arsitektur
- Meninjau karakteristik yang membedakan arsitektur data warehouse
- Periksa bagaimana kerangka arsitektur mendukung aliran data
- Memahami arti arsitektur teknis
- Mempelajari fungsi dan jasa komponen arsitektur
- Tinjau kembali lima tipe arsitektur utama

7.1 MEMAHAMI ARSITEKTUR DATA WAREHOUSE

Di Bab 2, Anda diperkenalkan dengan blok penyusun data warehouse. Pada tahap itu, kami dengan cepat melihat daftar komponen dan meninjau masing-masing komponen secara singkat. Pada Bab 6, kami meninjau kembali arsitektur gudang data dan menetapkan bahwa persyaratan bisnis merupakan kekuatan pendorong utama untuk semua desain dan pengembangan, termasuk rencana arsitektur.

Dalam bab ini, kami ingin meninjau arsitektur data warehouse dari berbagai perspektif. Anda akan mempelajari komponen arsitektur sesuai urutan yang memungkinkan aliran data dari sumber sebagai intelijen bisnis ke pengguna akhir. Kemudian Anda akan dapat melihat setiap area arsitektur dan memeriksa fungsi, prosedur, dan fitur di area tersebut. Diskusi itu akan membawa Anda ke dalam arsitektur teknis di bidang arsitektur tersebut.

Definisi Arsitektur

Struktur yang menyatukan semua komponen gudang data dikenal sebagai arsitektur. Misalnya saja pada kasus arsitektur gedung sekolah. Arsitektur bangunan bukan sekedar gaya visual. Ini mencakup berbagai ruang kelas, kantor, perpustakaan, koridor, gimnasium, pintu, jendela, atap, dan sejumlah besar komponen lainnya. Apabila seluruh komponen tersebut dirangkai dan ditempatkan menjadi satu, maka struktur yang mengikat seluruh komponen tersebut menjadi satu adalah arsitektur bangunan sekolah. Jika Anda dapat memperluas perbandingan ini ke gudang data, berbagai komponen gudang data bersama-sama membentuk arsitektur gudang data.

Saat membangun gedung sekolah, katakanlah para pembangun disuruh membuat ruang kelas menjadi besar. Jadi mereka membuat ruang kelas lebih besar tetapi menghilangkan kantor sama sekali, sehingga membangun gedung sekolah dengan arsitektur yang salah. Apa yang salah dengan arsitekturnya? Salah satu penyebabnya adalah tidak adanya semua komponen yang diperlukan. Mungkin penataan komponen lainnya juga kurang tepat. Arsitektur yang benar sangat penting untuk keberhasilan gudang data Anda. Oleh karena itu, dalam bab ini, kita akan melihat lebih dekat arsitektur data warehouse.

Di gudang data Anda, arsitektur mencakup sejumlah faktor. Terutama, ini mencakup data terintegrasi yang menjadi inti. Arsitektur mencakup segala sesuatu yang diperlukan untuk menyiapkan data dan menyimpannya. Di sisi lain, ini juga mencakup semua sarana untuk menyampaikan informasi dari gudang data Anda. Arsitekturnya selanjutnya terdiri dari aturan, prosedur, dan fungsi yang memungkinkan gudang data Anda berfungsi dan memenuhi kebutuhan bisnis. Terakhir, arsitektur terdiri dari teknologi yang memberdayakan gudang data Anda.

Apa tujuan umum dari arsitektur gudang data? Arsitekturnya menyediakan kerangka kerja keseluruhan untuk mengembangkan dan menerapkan gudang data Anda; ini adalah cetak biru yang komprehensif. Arsitektur mendefinisikan standar, pengukuran, desain umum, dan teknik pendukung.

Arsitektur di Tiga Bidang Utama

Seperti yang telah Anda ketahui, tiga area utama dalam data warehouse adalah:

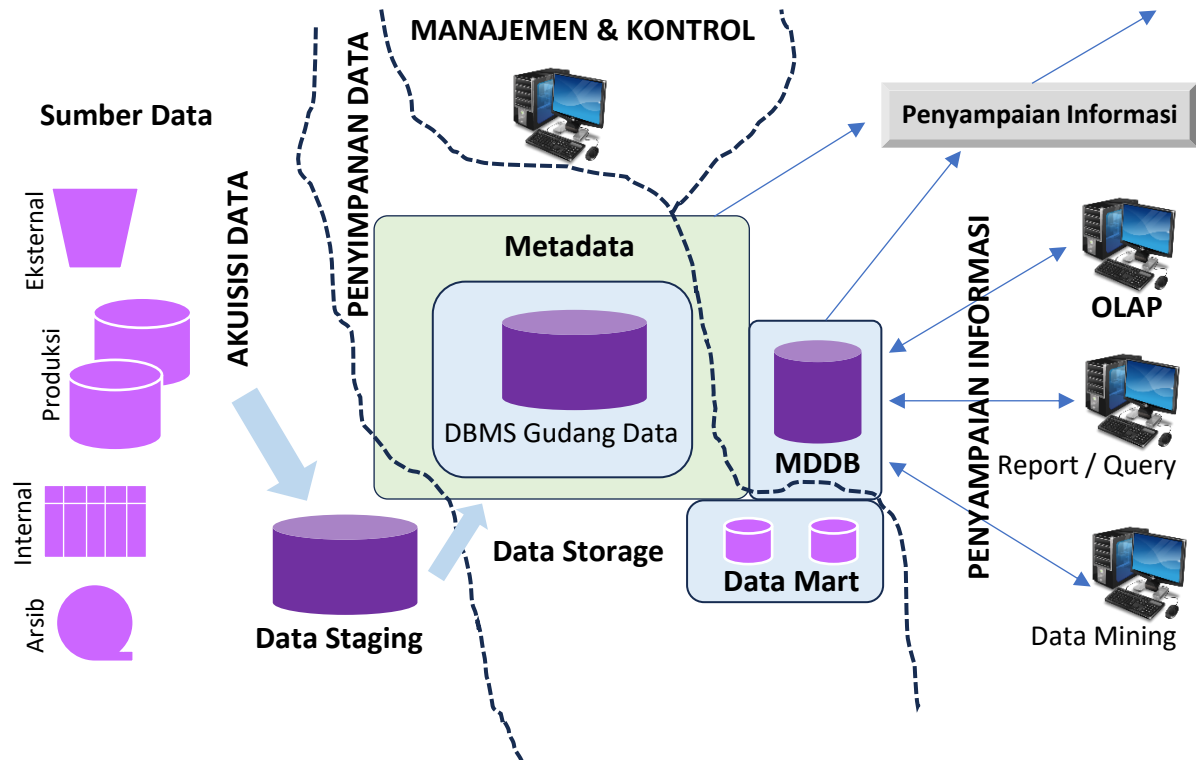
- Akuisisi data
- Penyimpanan data
- Penyampaian informasi

Pada Bab 2, kami mengidentifikasi blok bangunan utama data warehouse berikut ini:

- Sumber data
- Pementasan data
- Penyimpanan data
- Penyampaian informasi
- Metadata
- Manajemen dan pengendalian

Gambar 7.1 mengelompokkan komponen arsitektur utama ini ke dalam tiga area. Dalam bab ini, kita akan mempelajari arsitektur yang berkaitan dengan ketiga bidang ini. Di setiap area, kami akan mempertimbangkan komponen arsitektur pendukungnya. Masing-masing komponen mempunyai fungsi tertentu dan memberikan layanan tertentu. Kami akan menyelidiki fungsi dan layanan ini dan juga memeriksa arsitektur teknis yang mendasari masing-masing komponen di ketiga bidang ini.

Karena pentingnya komponen arsitektur, Anda juga akan menerima rincian tambahan di bab selanjutnya. Untuk saat ini, untuk tiga area data warehouse, mari kita berkonsentrasi pada fungsi, layanan, dan arsitektur teknis di area utama ini seperti yang disorot pada Gambar 7.1.



Gambar 7.1 Komponen arsitektur di tiga bidang utama.

7.2 KARAKTERISTIK YANG MEMBEDAKAN

Sebagai seorang profesional TI, ketika Anda terlibat dalam pengembangan sistem OLTP seperti pemrosesan pesanan atau pengendalian inventaris, atau pelaporan penjualan, apakah Anda mempertimbangkan arsitektur untuk setiap sistem? Meskipun istilah arsitektur biasanya tidak disebutkan dalam konteks sistem operasional, namun arsitektur yang mendasarinya juga ada untuk sistem ini. Misalnya, pertimbangan arsitektural untuk sistem seperti itu akan mencakup konversi file, populasi awal database, metode input data, penyampaian informasi melalui layar online, dan keseluruhan rangkaian pelaporan online dan batch. Namun untuk sistem seperti ini kami tidak membahas pertimbangan arsitektural secara rinci. Gudang data berbeda dan khas sehingga memaksa kita untuk mempertimbangkan arsitektur secara rinci.

Arsitektur gudang data luas, kompleks, dan ekspansif. Dalam gudang data, arsitekturnya terdiri dari komponen-komponen yang berbeda. Arsitekturnya memiliki ciri khas yang patut dipertimbangkan secara mendetail. Sebelum melanjutkan ke pembahasan kerangka arsitektur itu sendiri, mari kita meninjau karakteristik yang membedakan arsitektur data warehouse.

Tujuan dan Ruang Lingkup yang Berbeda

Arsitekturnya harus mendukung persyaratan untuk menyediakan informasi strategis. Informasi strategis sangat berbeda dengan informasi yang diperoleh dari sistem operasional. Ketika Anda memberikan informasi dari aplikasi operasional, konten dan kuantitas informasi per sesi pengguna dibatasi. Sebagai contoh, pada waktu tertentu, pengguna hanya tertarik pada informasi tentang satu pelanggan dan semua pesanan terkait. Namun, dari gudang data, pengguna tertarik untuk mendapatkan kumpulan hasil yang besar. Contoh kumpulan hasil

besar dari gudang data Anda adalah semua penjualan untuk tahun tersebut yang dikelompokkan berdasarkan kuartal, produk, dan wilayah penjualan.

Oleh karena itu, arsitektur gudang data harus memiliki komponen yang berfungsi menyediakan data kepada pengguna dalam volume besar dalam satu sesi. Pada dasarnya, sejauh mana sistem pendukung keputusan berbeda dari sistem operasional secara langsung diterjemahkan ke dalam satu prinsip penting: gudang data harus memiliki arsitektur yang berbeda dan lebih rumit.

Mendefinisikan ruang lingkup data warehouse juga sulit. Bagaimana Anda mencakup sistem operasional? Anda mempertimbangkan kelompok pengguna, rentang fungsi, penyimpanan data, serta layar keluaran dan laporan. Untuk gudang data dengan arsitektur sebagai cetak birunya, faktor apa saja yang harus Anda pertimbangkan untuk menentukan cakupannya?

Ada beberapa faktor yang perlu dipertimbangkan. Pertama, Anda harus mempertimbangkan jumlah dan luasnya sumber data. Berapa banyak sistem lama yang datanya akan Anda ekstrak? Apa saja sumber eksternalnya? Apakah Anda berencana untuk memasukkan file departemen, spreadsheet, dan database pribadi? Bagaimana dengan memasukkan data yang diarsipkan? Cakupan arsitektur dapat diukur lagi berdasarkan transformasi data dan fungsi integrasi. Dalam gudang data, granularitas data dan volume data juga merupakan pertimbangan penting.

Pertimbangan serius lainnya adalah dampak data warehouse terhadap sistem operasional yang ada. Karena ekstraksi data, perbandingan, dan rekonsiliasi, Anda harus menentukan seberapa besar dampak negatif gudang data terhadap kinerja sistem operasional. Kapan ekstrak batch Anda akan dijalankan dan bagaimana pengaruhnya terhadap sistem sumber produksi? Di sisi lain, jika gudang data Anda bersifat real-time, fungsi pengambilan data akan menjadi rumit dan kompleks.

Konten Data

Data “*read-only*” di gudang data berada di tengah sebagai komponen utama dalam arsitektur. Dalam suatu sistem operasional, meskipun database penting, namun tidak sepenting tempat penyimpanan data di data warehouse. Sebelum data dibawa ke gudang data Anda dan disimpan sebagai data baca-saja, sejumlah fungsi harus dilakukan. Fungsi-fungsi yang lengkap dan penting ini tidak dapat dibandingkan dengan konversi data yang terjadi dalam sistem operasional.

Di gudang data, Anda menyimpan data yang terintegrasi dari berbagai sumber. Setelah mengekstraksi data, yang merupakan proses rumit, Anda mengubah data, membersihkannya, dan mengintegrasikannya dalam area pementasan. Baru setelah itu Anda memindahkan data terintegrasi ke dalam repositori gudang data sebagai data hanya-baca. Data operasional bukanlah data “*read-only*”.

Selanjutnya, arsitektur gudang data Anda harus mendukung penyimpanan data yang dikelompokkan berdasarkan subjek bisnis, bukan dikelompokkan berdasarkan aplikasi seperti dalam kasus sistem operasional. Data di gudang data Anda tidak mewakili snapshot yang berisi

nilai variabel sebagaimana adanya saat ini. Ini berbeda dan berbeda dari kebanyakan sistem operasional.

Ketika kita menyebutkan data historis yang disimpan di gudang data, kita berbicara tentang volume data yang sangat tinggi. Kebanyakan perusahaan memilih untuk menyimpan data 10 tahun yang lalu di gudang data. Beberapa perusahaan ingin menyimpan lebih banyak lagi, jika datanya tersedia. Ini adalah alasan lain mengapa arsitektur data warehouse harus mendukung volume data yang tinggi.

Analisis Kompleks dan Respon Cepat

Arsitektur gudang data Anda harus mendukung analisis kompleks atas informasi strategis oleh pengguna. Proses pengambilan informasi dalam sistem operasional berkurang kompleksitasnya jika dibandingkan dengan penggunaan informasi dari gudang data. Sebagian besar pengambilan informasi online selama sesi oleh pengguna di gudang data adalah analisis interaktif. Seorang pengguna tidak menjalankan kueri yang terisolasi, keluar dari gudang data, dan kembali lagi nanti untuk kueri tunggal berikutnya. Sesi oleh pengguna bersifat berkelanjutan dan berlangsung lama karena pengguna biasanya memulai dengan kueri pada tingkat tinggi, meninjau kumpulan hasil, memulai kueri berikutnya dengan melihat data dengan cara yang sedikit berbeda, dan seterusnya.

Oleh karena itu, arsitektur gudang data Anda harus mendukung variasi dalam menyediakan analisis. Pengguna harus mampu menelusuri, menggabungkan, membagi data, dan bermain dengan skenario “bagaimana-jika”. Pengguna harus memiliki kemampuan untuk meninjau kumpulan hasil dalam opsi keluaran yang berbeda. Pengguna tidak lagi puas dengan kumpulan hasil tekstual atau hasil yang ditampilkan dalam format tabel. Setiap hasil yang ditetapkan dalam format tabel harus diterjemahkan ke dalam grafik grafis.

Penyediaan informasi strategis dimaksudkan untuk mengambil keputusan dengan cepat dan menghadapi situasi dengan cepat. Misalnya, wakil presiden pemasaran Anda ingin segera mengetahui alasan penurunan penjualan selama tiga minggu berturut-turut di wilayah tengah dan segera mengambil keputusan untuk memperbaiki situasi tersebut. Gudang data Anda harus memberinya alat dan informasi untuk merespons masalah dengan cepat.

Arsitektur gudang data Anda harus memudahkan pengambilan keputusan strategis dengan cepat. Harus ada komponen yang sesuai dalam arsitektur untuk mendukung respon cepat oleh pengguna untuk menghadapi situasi dengan menggunakan informasi yang disediakan oleh gudang data Anda. Jika gudang data Anda mendukung pengambilan informasi secara real-time, arsitekturnya harus diperluas untuk mengakomodasi pengambilan data secara real-time dan kemampuan untuk memperoleh informasi strategis secara real-time untuk membuat keputusan di tempat. Pergudangan data waktu nyata berarti penyampaian informasi ke sejumlah besar pengguna baik di dalam maupun di luar organisasi.

Fleksibel dan Dinamis

Khususnya dalam hal desain dan pengembangan data warehouse, Anda tidak mengetahui semua kebutuhan bisnis sebelumnya. Dengan menggunakan teknik pembuatan paket informasi, Anda dapat menilai sebagian besar persyaratan dan memodelkan persyaratan data secara dimensional. Namun demikian, bagian persyaratan yang hilang

muncul setelah pengguna Anda mulai menggunakan gudang data. Apa implikasinya? Anda harus memastikan arsitektur gudang data Anda cukup fleksibel untuk mengakomodasi kebutuhan tambahan saat persyaratan tersebut muncul.

Persyaratan tambahan muncul untuk memasukkan item yang terlewat dalam persyaratan bisnis. Terlebih lagi, kondisi bisnis sendiri berubah. Faktanya, mereka terus berubah. Perubahan kondisi bisnis memerlukan persyaratan bisnis tambahan untuk disertakan dalam gudang data. Jika arsitektur gudang data dirancang agar fleksibel dan dinamis, maka gudang data Anda dapat memenuhi kebutuhan tambahan ketika persyaratan tersebut muncul.

Berbasis Metadata

Saat data berpindah dari sistem sumber ke pengguna akhir sebagai informasi yang berguna dan strategis, metadata melingkupi keseluruhan pergerakan. Komponen metadata dari arsitektur menyimpan data tentang setiap fase pergerakan, dan, dalam arti tertentu, memandu pergerakan data. Dalam suatu sistem operasional, tidak ada komponen yang setara dengan metadata dalam data warehouse. Kamus data DBMS sistem operasional hanyalah bayangan samar dari metadata di gudang data. Jadi, dalam arsitektur gudang data Anda, komponen metadata bersinggungan dan menghubungkan komponen lainnya. Metadata dalam gudang data sangat penting sehingga kami telah mendedikasikan Bab 9 untuk membahas metadata.

7.3 KERANGKA ARSITEKTUR

Sebelumnya di bagian sebelumnya dari bab ini, kami mengelompokkan komponen arsitektur sebagai blok bangunan dalam tiga bidang berbeda yaitu akuisisi data, penyimpanan data, dan pengiriman informasi. Di masing-masing area gudang data yang luas ini, komponen arsitektural melayani tujuan tertentu.

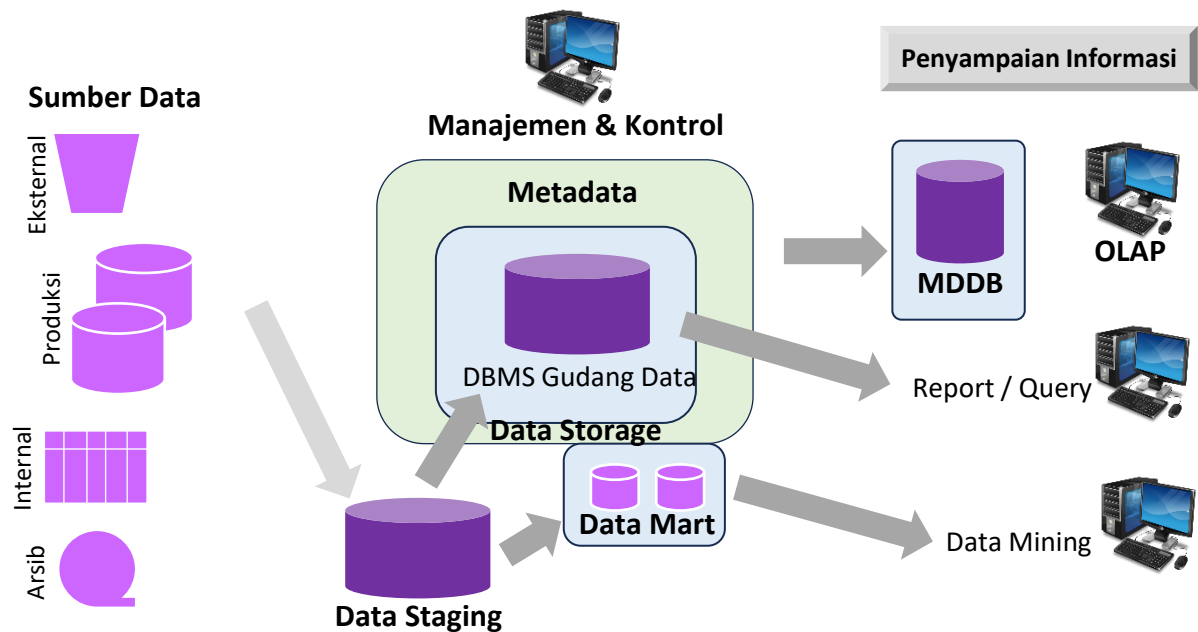
Arsitektur Pendukung Aliran Data

Sekarang kita ingin mengaitkan komponen-komponen tersebut sebagai pembentuk kerangka kerja untuk mengkondisikan dan memungkinkan aliran data dari awal hingga akhir. Seperti yang Anda ketahui dengan baik, data yang akhirnya sampai ke pengguna akhir sebagai informasi strategis yang berguna dimulai sebagai elemen data yang berbeda di berbagai sumber data. Pengumpulan data dari berbagai sumber ini berpindah ke area pementasan. Apa yang terjadi selanjutnya? Data yang diekstraksi melewati proses persiapan terperinci di staging area sebelum dikirim ke gudang data untuk disimpan dengan benar.

Dari penyimpanan gudang data, data yang diubah menjadi informasi berguna diambil oleh pengguna atau dikirimkan ke desktop pengguna sesuai kebutuhan. Pada dasarnya, apa itu data warehousing? Apakah Anda setuju bahwa data warehousing berarti mengambil semua sumber data yang diperlukan, menyiapkannya, menyimpannya dalam format yang sesuai, dan kemudian menyampaikan informasi yang berguna kepada pengguna akhir?. Gambar 7.2 menunjukkan aliran data dari awal hingga akhir dan juga menyoroti komponen arsitektur yang memungkinkan aliran data seiring pergerakan data.

Sekarang mari kita ikuti aliran data dan mengidentifikasi komponen arsitektur. Beberapa komponen arsitektur mengatur aliran data dari awal hingga akhir. Modul manajemen dan kontrol adalah salah satu komponen tersebut. Modul ini menyentuh setiap langkah di sepanjang jalur pergerakan data. Apa yang terjadi pada titik kritis aliran data? Apa saja komponen arsitekturnya, dan bagaimana komponen ini memungkinkan aliran data?

Pada Sumber Data Disini sumber data internal dan eksternal membentuk komponen arsitektur data sumber. Sumber data mengatur ekstraksi data untuk persiapan dan penyimpanan di gudang data. Komponen arsitektur pementasan data mengatur transformasi, pembersihan, dan integrasi dat



Gambar 7.2 Kerangka arsitektur yang mendukung aliran data.

Dalam Repositori Gudang Data Komponen arsitektur penyimpanan data mencakup pemuatan data dari area pementasan dan juga penyimpanan data dalam format yang sesuai untuk penyampaian informasi. Komponen arsitektur metadata juga merupakan mekanisme penyimpanan untuk menampung data tentang data di setiap titik aliran data dari awal hingga akhir.

Di Sisi Pengguna Komponen arsitektur penyampaian informasi mencakup data mart dependen, database multidimensi khusus, dan serangkaian fasilitas kueri dan pelaporan, termasuk dasbor dan kartu skor.

Modul Manajemen dan Pengendalian

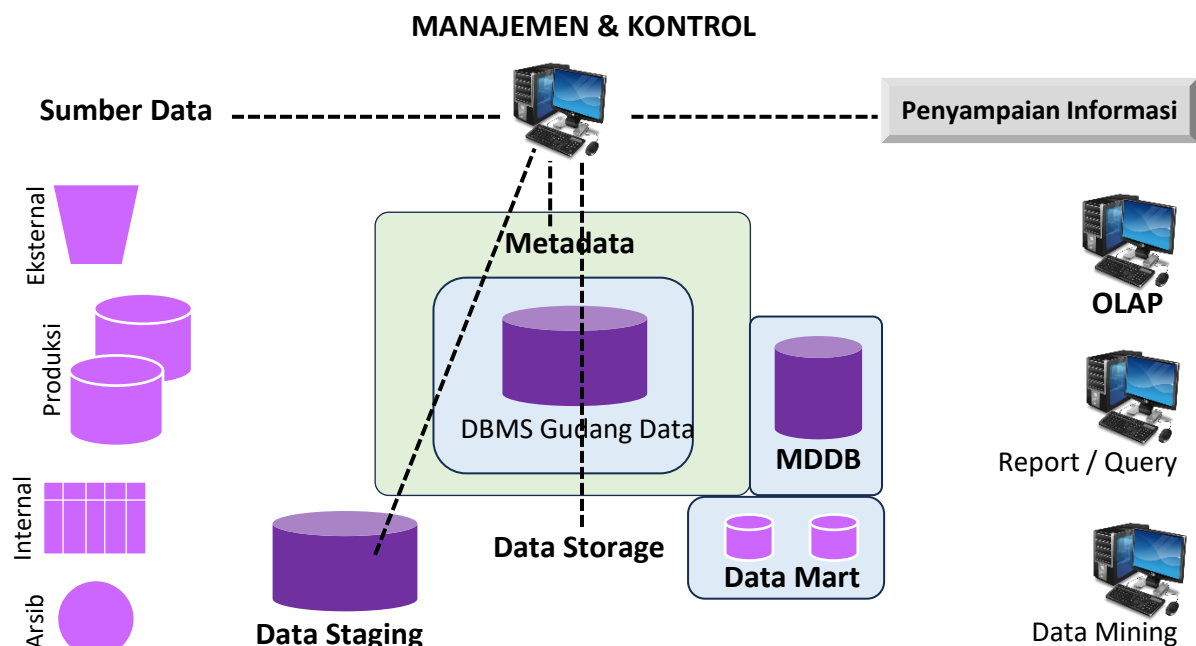
Komponen arsitektur ini merupakan keseluruhan modul yang mengelola dan mengendalikan seluruh lingkungan data warehouse. Ini adalah komponen payung yang bekerja di berbagai tingkatan dan mencakup semua operasi. Komponen ini memiliki dua fungsi utama: pertama untuk terus memantau semua operasi yang sedang berlangsung, dan kedua untuk mengambil tindakan dan memulihkan masalah ketika terjadi kesalahan. Gambar 7.3

menunjukkan bagaimana komponen manajemen berhubungan dan mengelola seluruh operasi data warehouse.

Pada awalnya di gudang data Anda, Anda memiliki operasi yang berkaitan dengan akuisisi data. Hal ini termasuk mengekstraksi data dari sistem sumber baik untuk penyegaran penuh atau untuk pemuatan tambahan. Memindahkan data ke area pementasan dan menjalankan fungsi transformasi data juga merupakan bagian dari akuisisi data. Komponen arsitektur manajemen mengelola dan mengendalikan fungsi akuisisi data ini, memastikan bahwa ekstraksi dan transformasi dilakukan dengan benar dan tepat waktu.

Modul manajemen juga mengelola pencadangan sebagian besar gudang data dan pemulihan dari kegagalan. Layanan manajemen meliputi pemantauan pertumbuhan dan pengarsipan data secara berkala dari gudang data. Komponen arsitektur ini juga mengatur keamanan data dan menyediakan akses resmi ke gudang data. Selain itu, komponen manajemen berinteraksi dengan komponen penyampaian informasi pengguna akhir untuk memastikan bahwa penyampaian informasi dilakukan dengan benar.

Hanya sedikit alat yang dirancang khusus untuk administrasi gudang data yang tersedia saat ini. Umumnya administrator gudang data menjalankan fungsi komponen manajemen dan pengendalian dengan menggunakan alat yang tersedia di DBMS gudang data.



Gambar 7.3 Komponen manajemen dan pengendalian.

7.4 ARSITEKTUR TEKNIS

Kami telah meninjau berbagai komponen arsitektur gudang data dalam beberapa cara berbeda. Pertama, kami mengelompokkan komponen ke dalam tiga bidang utama yaitu akuisisi data, penyimpanan data, dan penyampaian informasi. Kemudian, kami mengeksplorasi karakteristik yang membedakan arsitektur data warehouse. Kami memeriksa arsitekturnya dan menyoroti karakteristik yang membedakan arsitektur gudang data

dibandingkan dengan sistem operasional apa pun. Kami juga menelusuri aliran data melalui gudang data dan menghubungkan masing-masing komponen arsitektur ke stasiun-stasiun di sepanjang aliran data. Anda sekarang telah memahami dengan baik bagaimana kami menerapkan istilah arsitektur pada gudang data. Kami telah menunjukkan bahwa setiap komponen arsitektur ada untuk menjalankan serangkaian fungsi tertentu dan menyediakan sekelompok layanan tertentu. Ketika semua komponen menjalankan fungsi yang telah ditentukan sebelumnya dan menyediakan layanan yang diperlukan, maka keseluruhan arsitektur mendukung gudang data untuk memenuhi tujuan dan kebutuhan bisnis.

Kami menunjukkan arsitektur teknis gudang data sebagai rangkaian lengkap fungsi dan layanan yang disediakan dalam struktur komponennya. Arsitektur teknis juga mencakup prosedur dan aturan yang diperlukan untuk menjalankan fungsi dan menyediakan layanan. Arsitektur teknis juga mencakup penyimpanan data yang dibutuhkan setiap komponen untuk menyediakan layanan.

Mari kita membuat perbedaan penting lainnya. Arsitektur bukanlah seperangkat alat yang diperlukan untuk menjalankan fungsi dan menyediakan layanan. Ketika kita mengacu pada fungsi ekstraksi data dalam salah satu komponen arsitektur, kita hanya menyebutkan fungsi itu sendiri dan berbagai tugas yang terkait dengan fungsi tersebut. Selain itu, kami menghubungkan penyimpanan data untuk staging area dengan fungsi ekstraksi data karena data yang diekstraksi dipindahkan ke staging area. Perhatikan bahwa tidak disebutkan alat apa pun untuk menjalankan fungsi tersebut. Di manakah letak alat-alat tersebut? Apa saja alat untuk mengekstrak data? Apa saja alat yang berhubungan dengan arsitektur? Alat adalah sarana untuk mengimplementasikan arsitektur teknis. Itu sebabnya Anda harus ingat bahwa arsitektur adalah yang utama dan peralatannya menyusul.

Anda akan memilih alat yang paling sesuai untuk arsitektur gudang data Anda. Mari kita ambil contoh yang sangat sederhana, mungkin tidak realistis. Misalkan satu-satunya sumber data untuk gudang data Anda hanyalah empat tabel dari satu database relasional terpusat. Jika ya, apa cakupan dan cakupan komponen sumber data? Berapa besarnya fungsi ekstraksi data? Jumlahnya sangat terbatas. Apakah Anda kemudian memerlukan alat pihak ketiga yang canggih untuk ekstraksi data? Tentu saja tidak. Mengambil posisi ekstrim lainnya, misalkan sumber data Anda terdiri dari database dan file dari 50 atau lebih sistem lama yang berjalan pada berbagai platform di lokasi jarak jauh. Dalam hal ini, komponen arsitektur sumber data dan fungsi ekstraksi data Anda memiliki cakupan yang sangat luas dan kompleks. Anda tentu perlu meningkatkan upaya internal Anda dengan alat ekstraksi data yang tepat dari vendor.

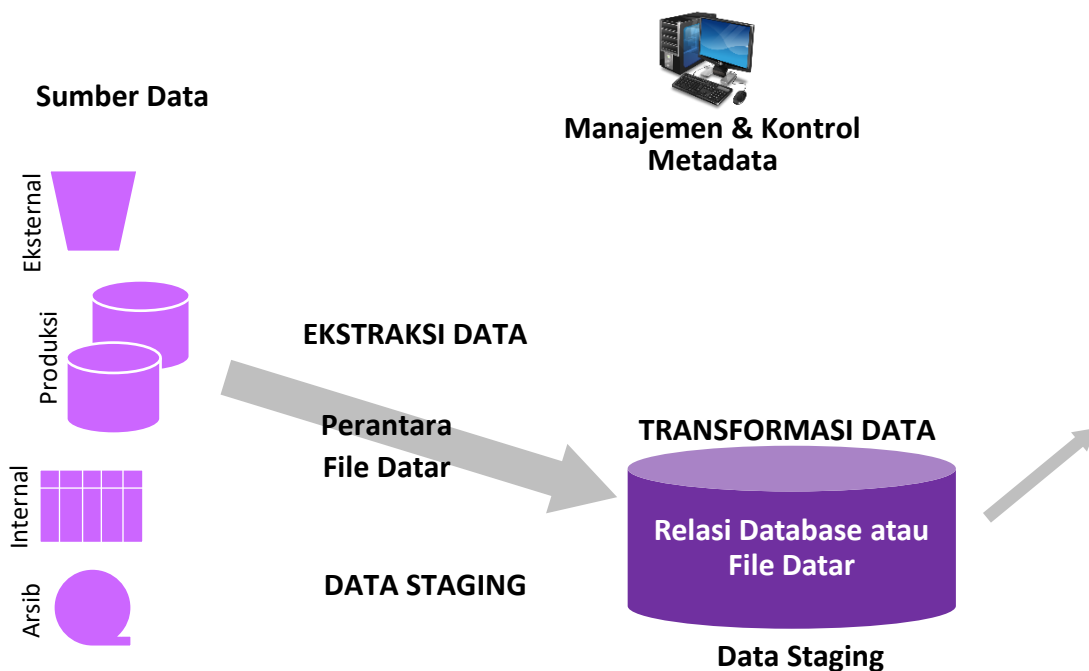
Pada bagian selanjutnya dari bab ini, kita akan membahas arsitektur teknis komponen. Kami akan membahas dan menguraikan jenis fungsi, layanan, prosedur, dan penyimpanan data yang relevan dengan setiap komponen arsitektur. Ini adalah pedoman. Anda harus mengambil pedoman ini dan meninjau serta menyesuaikannya untuk membangun arsitektur gudang data Anda. Saat Anda membangun arsitektur untuk gudang data Anda, Anda akan menyiapkan rencana arsitektur yang akan mencakup semua komponen. Rencana tersebut juga akan menyatakan secara rinci tingkat dan kompleksitas seluruh fungsi, layanan, prosedur,

dan penyimpanan data yang terkait dengan setiap komponen arsitektur. Rencana arsitektur akan berfungsi sebagai cetak biru untuk desain dan pengembangan data warehouse. Ini juga akan berfungsi sebagai daftar periksa utama untuk pemilihan alat Anda.

Akuisisi Data

Area ini mencakup seluruh proses ekstraksi data dari sumber data, memindahkan semua data yang diekstraksi ke staging area, dan menyiapkan data untuk dimuat ke dalam repositori data warehouse. Dua komponen arsitektur utama yang diidentifikasi sebelumnya sebagai bagian dari area ini adalah sumber data dan pementasan data. Fungsi dan pelayanan pada kawasan ini berkaitan dengan kedua komponen arsitektur tersebut. Variasi sumber data berdampak langsung pada cakupan dan cakupan fungsi dan layanan.

Apa yang terjadi di area ini sangat penting dalam pengembangan gudang data Anda. Proses ekstraksi, transformasi, dan pemuatan data memakan waktu, intensif manusia, dan sangat penting. Bab 12 membahas proses-proses ini secara mendalam.



Gambar 7.4 Akuisisi data: arsitektur teknis.

Namun, saat ini, kami akan membahasnya cukup panjang agar Anda dapat menempatkan semua komponen arsitektur dalam perspektif yang tepat. Gambar 7.4 merangkum arsitektur teknis untuk akuisisi data.

Aliran data

Aliran Di area akuisisi data, aliran data dimulai dari sumber data dan berhenti di area pementasan. Setelah transformasi dan integrasi, data siap dimuat ke dalam repositori gudang data. Sumber Data Untuk sebagian besar gudang data, sumber data utama terdiri dari sistem operasional perusahaan. Di banyak organisasi, beberapa sistem operasional masih merupakan sistem warisan. Data lama berada pada database hierarki atau jaringan. Anda harus menggunakan bahasa generasi keempat yang sesuai dari DBMS tertentu untuk mengekstrak

data dari database ini. Namun, sebagian besar sistem operasional terkini berjalan pada arsitektur klien/server. Biasanya, sistem ini didukung oleh DBMS relasional. Di sini Anda dapat menggunakan bahasa berbasis SQL untuk mengekstraksi data.

Sejumlah besar perusahaan telah mengadopsi sistem ERP (perencanaan sumber daya perusahaan). Sumber data ERP memberikan keuntungan karena data dari sumber tersebut sudah terkonsolidasi dan terintegrasi. Namun, mungkin ada beberapa kelemahan dalam menggunakan sumber data ERP. Anda harus menggunakan alat milik vendor ERP untuk ekstraksi data. Selain itu, sebagian besar penawaran ERP berisi tabel data sumber dalam jumlah yang sangat besar. Untuk memasukkan data dari sumber luar, Anda harus membuat file sementara untuk menyimpan data yang diterima dari sumber luar. Setelah memformat ulang dan menata ulang elemen data, Anda harus memindahkan data ke staging area.

Penyimpanan Data Perantara Saat data diekstraksi dari sumber data, data berpindah melalui file sementara. Terkadang, ekstrak data homogen dari beberapa aplikasi sumber ditarik ke dalam file sementara yang terpisah dan kemudian digabungkan ke dalam file sementara lain sebelum memindahkannya ke area pementasan. Proses sebaliknya juga sering terjadi. Dari setiap aplikasi, satu atau dua file datar besar dibuat dan kemudian dibagi menjadi file yang lebih kecil dan digabungkan dengan tepat sebelum memindahkan data ke staging area. Biasanya, praktik umum adalah menggunakan file datar untuk mengekstrak data dari sistem operasional.

Staging Area Ini adalah tempat di mana semua data yang diekstraksi disatukan dan disiapkan untuk dimuat ke dalam gudang data. Area pementasannya seperti pabrik perakitan atau area konstruksi. Di area ini, Anda memeriksa setiap file yang diekstraksi, meninjau aturan bisnis, melakukan berbagai fungsi transformasi data, mengurutkan dan menggabungkan data, mengatasi ketidakkonsistenan, dan membersihkan data. Ketika data akhirnya disiapkan untuk gudang data seluruh perusahaan atau salah satu data mart yang sesuai, data untuk sementara berada di repositori staging area menunggu untuk dimuat ke dalam repositori data warehouse.

Di banyak gudang data, data di staging area disimpan dalam file berurutan atau datar. Namun, file datar ini berisi data yang sepenuhnya terintegrasi dan dibersihkan dalam format yang sesuai dan siap untuk dimuat. Biasanya, file-file ini dalam format yang dapat dimuat oleh alat utilitas RDBMS gudang data. Kini semakin banyak repositori data staging area yang menjadi database relasional. Data di area pementasan tersebut disimpan untuk jangka waktu yang lebih lama. Meskipun ekstrak untuk memuat dapat dengan mudah diperoleh dari database relasional dengan indeks yang tepat, pembuatan dan pemeliharaan database relasional ini memerlukan overhead untuk pembuatan indeks dan migrasi data dari sistem sumber.

Area pementasan mungkin berisi data pada butiran terendah untuk mengisi tabel yang berisi pengukuran bisnis. Data gabungan juga umum disimpan di area pementasan untuk dimuat. Jenis data lain yang disimpan di area pementasan berkaitan dengan dimensi bisnis seperti produk, waktu, wilayah penjualan, pelanggan, dan skema promosi. Fungsi dan Layanan Daftar fungsi dan layanan di bagian ini berkaitan dengan area akuisisi data dan dibagi menjadi

tiga kelompok. Ini adalah daftar umum. Hal ini tidak menunjukkan tingkat atau kompleksitas setiap fungsi atau layanan. Untuk arsitektur teknis gudang data Anda, Anda harus menentukan konten dan kompleksitas setiap fungsi atau layanan.

Daftar Fungsi dan Layanan

Ekstraksi Data

- 1) Pilih sumber data dan tentukan jenis filter yang akan diterapkan ke masing-masing sumber.
- 2) Hasilkan file ekstrak otomatis dari sistem operasional menggunakan replikasi dan teknik lainnya.
- 3) Buat file perantara untuk menyimpan data yang dipilih untuk digabungkan nanti. Transportasi file yang diekstraksi dari berbagai platform.
- 4) Menyediakan layanan kontrol pekerjaan otomatis untuk membuat file ekstrak. Format ulang masukan dari sumber luar.
- 5) Format ulang input dari file data departemen, database, dan spreadsheet.
- 6) Hasilkan kode aplikasi umum untuk ekstraksi data.
- 7) Selesaikan ketidakkonsistenan untuk elemen data umum dari berbagai sumber.

Transformasi Data

- 1) Memetakan data masukan ke data untuk penyimpanan gudang data.
- 2) Bersihkan data, hapus duplikat, dan gabungkan/bersihkan.
- 3) Denormalisasi struktur data yang diekstraksi seperti yang disyaratkan oleh model dimensi gudang data.
- 4) Konversi tipe data.
- 5) Menghitung dan memperoleh nilai atribut.
- 6) Periksa integritas referensial.
- 7) Kumpulkan data sesuai kebutuhan. Selesaikan nilai yang hilang.
- 8) Konsolidasikan dan integrasikan data.
- 9) Pementasan Data
- 10) Menyediakan cadangan dan pemulihan untuk repositori area pementasan.
- 11) Sortir dan gabungkan file.
- 12) Buat file sebagai input untuk melakukan perubahan pada tabel dimensi.
- 13) Jika penyimpanan pementasan data adalah database relasional, buat dan isi database.
- 14) Pertahankan jejak audit untuk menghubungkan setiap item data di gudang data dengan sumber input.
- 15) Selesaikan dan buat kunci utama dan asing untuk memuat tabel.
- 16) Konsolidasikan kumpulan data dan buat file datar untuk dimuat melalui utilitas DBMS.
- 17) Jika penyimpanan area pementasan adalah database relasional, ekstrak file beban.

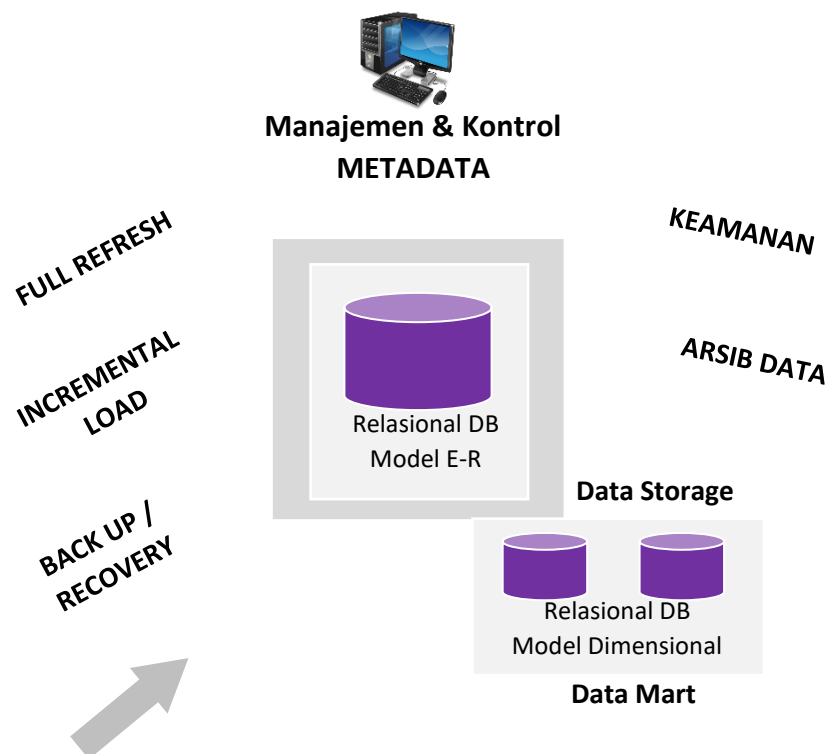
Penyimpanan data

Ini mencakup proses memuat data dari staging area ke dalam repositori data warehouse. Semua fungsi untuk mengubah dan mengintegrasikan data diselesaikan di area pementasan data. Data yang disiapkan di data warehouse ibarat produk jadi yang siap ditumpuk di gudang industri.

Bahkan sebelum memuat data ke dalam gudang data, metadata, yang merupakan komponen lain dari arsitektur, sudah aktif. Selama tahap ekstraksi data dan transformasi data, repositori metadata akan diisi. Gambar 7-5 menunjukkan gambaran ringkasan arsitektur teknis untuk penyimpanan data.

Aliran data

Aliran Untuk penyimpanan data, aliran data dimulai pada area pementasan data. Data yang diubah dan diintegrasikan dipindahkan dari staging area ke gudang data warehouse. Jika gudang data adalah gudang data seluruh perusahaan yang dibangun secara top-down, maka mungkin terdapat perpindahan data dari gudang data gudang perusahaan ke gudang data mart yang bergantung. Alternatifnya, jika gudang data merupakan konglomerasi data mart yang disesuaikan dan dibangun dengan cara bottom-up, maka pergerakan data berhenti pada data mart yang sesuai.



Gambar 7.5 Penyimpanan data: arsitektur teknis.

Kelompok Data Data yang sudah disiapkan dan menunggu di area pementasan data dibagi menjadi dua kelompok. Grup pertama adalah kumpulan file atau tabel yang berisi data untuk penyegaran penuh. Kelompok data ini biasanya dimaksudkan untuk pemuatan awal gudang data. Terkadang, beberapa tabel gudang data mungkin disegarkan sepenuhnya.

Kelompok data lainnya adalah kumpulan file atau tabel yang berisi muatan tambahan yang sedang berlangsung. Sebagian besar berhubungan dengan beban malam hari. Beberapa penambahan data dimensi dapat dilakukan pada interval yang lebih jarang. Repositori Data Hampir semua database gudang data saat ini adalah database relasional. Semua kemampuan, fleksibilitas, dan kemudahan penggunaan RDBMS tersedia untuk pemrosesan data.

Fungsi dan Layanan Daftar umum fungsi dan layanan yang diberikan di bagian ini adalah untuk panduan Anda. Daftar tersebut berkaitan dengan tempat penyimpanan data dan mencakup fungsi dan layanan yang luas. Ini adalah daftar umum. Hal ini tidak menunjukkan tingkat atau kompleksitas setiap fungsi atau layanan. Untuk arsitektur teknis gudang data Anda, Anda harus menentukan konten dan kompleksitas setiap fungsi atau layanan.

Daftar Fungsi dan Layanan

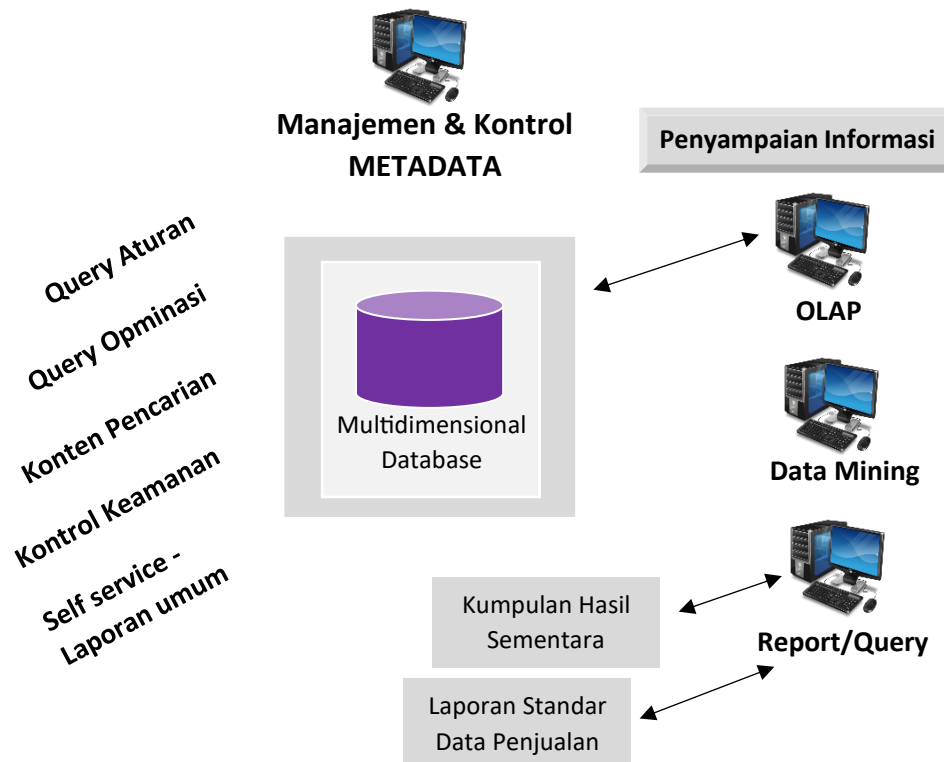
- Memuat data untuk penyegaran penuh tabel gudang data.
- Lakukan beban tambahan pada interval teratur yang ditentukan.
- Mendukung pemuatan ke dalam beberapa tabel pada tingkat terperinci dan ringkasan.
- Optimalkan proses pemuatan.
- Menyediakan layanan kontrol pekerjaan otomatis untuk memuat gudang data.
- Menyediakan cadangan dan pemulihan untuk database data warehouse.
- Memberikan keamanan.
- Memantau dan menyempurnakan database.
- Arsipkan data dari database secara berkala sesuai kondisi yang telah ditetapkan.

Penyampaian Informasi

Area ini mencakup spektrum metode yang luas untuk membuat informasi tersedia bagi pengguna. Bagi pengguna Anda, komponen penyampaian informasi adalah gudang data. Mereka tidak bersentuhan langsung dengan komponen lainnya. Bagi pengguna, kekuatan arsitektur gudang data Anda terutama terkonsentrasi pada ketahanan dan fleksibilitas komponen penyampaian informasi.

Komponen penyampaian informasi memudahkan pengguna untuk mengakses informasi baik secara langsung dari gudang data perusahaan, dari data mart dependen, atau dari kumpulan data mart yang disesuaikan. Sebagian besar akses informasi di gudang data adalah melalui pertanyaan online dan sesi analisis interaktif. Namun demikian, gudang data Anda juga akan menghasilkan laporan rutin dan ad hoc.

Hampir semua gudang data modern menyediakan pemrosesan analitis online (OLAP). Dalam hal ini, gudang data primer memasukkan data ke database multidimensi berpemilik (MDDB) di mana data yang diringkas disimpan sebagai kubus informasi multidimensi. Pengguna melakukan analisis multidimensi yang kompleks menggunakan kubus informasi di MDDB. Lihat Gambar 7.6 untuk gambaran ringkasan arsitektur teknis penyampaian informasi.



Gambar 7.6 Penyampaian informasi: arsitektur teknis.

Aliran data

Aliran untuk penyampaian informasi, aliran data dimulai di gudang data seluruh perusahaan dan data mart dependen ketika desainnya didasarkan pada teknik top-down. Ketika desain mengikuti metode bottom-up, aliran data dimulai pada kumpulan data mart yang sesuai. Umumnya, data diubah menjadi informasi mengalir ke desktop pengguna selama sesi kueri. Selain itu, informasi yang dicetak pada laporan reguler atau laporan ad hoc sampai ke pengguna. Terkadang, kumpulan hasil dari kueri atau laporan individual disimpan di penyimpanan data milik vendor alat kueri atau pelaporan. Informasi yang disimpan dapat digunakan berulang kali dengan lebih cepat. Organisasi progresif baru-baru ini menerapkan dasbor dan kartu skor sebagai bagian dari penyampaian informasi. Dashboard merupakan perangkat tampilan informasi real time atau mendekati real time.

Data mengalir ke dashboard secara real time dari data warehouse. Di banyak gudang data, data juga mengalir ke aplikasi pendukung keputusan hilir khusus seperti sistem informasi eksekutif (EIS) dan penambangan data. Aliran informasi lain yang lebih umum adalah kepemilikan database multidimensi untuk OLAP.

Lokasi Layanan Dalam komponen pengiriman informasi, Anda dapat menyediakan layanan kueri dari desktop pengguna, dari server aplikasi, atau dari database itu sendiri. Ini akan menjadi salah satu keputusan penting untuk desain arsitektur Anda. Untuk menghasilkan laporan reguler atau ad hoc, Anda mungkin ingin menyertakan layanan pelaporan yang komprehensif. Layanan ini akan memungkinkan pengguna untuk membuat dan menjalankan laporan mereka sendiri. Ini juga akan menyediakan laporan standar untuk dijalankan secara berkala.

Penyimpanan Data Untuk penyampaian informasi, Anda dapat mempertimbangkan penyimpanan data perantara berikut:

- ◆ Penyimpanan sementara yang bersifat eksklusif untuk menyimpan hasil pertanyaan dan laporan individual untuk penggunaan berulang
- ◆ Penyimpanan data untuk pelaporan standar
- ◆ Penyimpanan data untuk dasbor
- ◆ Database multidimensi berpemilik

Fungsi dan Layanan Tinjau daftar umum fungsi dan layanan yang diberikan di bawah ini dan gunakan sebagai panduan untuk menetapkan komponen pengiriman informasi arsitektur gudang data Anda. Daftar ini berkaitan dengan penyampaian informasi dan mencakup fungsi dan layanan secara luas. Sekali lagi, ini adalah daftar umum. Hal ini tidak menunjukkan tingkat atau kompleksitas setiap fungsi atau layanan. Untuk arsitektur teknis gudang data Anda, Anda harus menentukan konten dan kompleksitas setiap fungsi atau layanan.

- ▶ Memberikan keamanan untuk mengontrol akses informasi.
- ▶ Memantau akses pengguna untuk meningkatkan layanan dan peningkatan di masa mendatang.
- ▶ Izinkan pengguna menelusuri konten gudang data.
- ▶ Menyederhanakan akses dengan menyembunyikan kompleksitas internal penyimpanan data dari pengguna.
- ▶ Secara otomatis memformat ulang kueri untuk eksekusi optimal.
- ▶ Aktifkan kueri untuk mengetahui tabel agregat untuk hasil yang lebih cepat.
- ▶ Mengatur kueri dan mengontrol kueri yang tidak terkendali.
- ▶ Menyediakan pembuatan laporan mandiri bagi pengguna, yang terdiri dari berbagai pilihan fleksibel untuk membuat, menjadwalkan, dan menjalankan laporan.
- ▶ Menyimpan kumpulan hasil kueri dan laporan untuk digunakan di masa mendatang.
- ▶ Menyediakan berbagai tingkat perincian data.
- ▶ Menyediakan pemacu peristiwa untuk memantau pemuatan data.
- ▶ Membuat ketentuan bagi pengguna untuk melakukan analisis kompleks melalui pemrosesan analitis online (OLAP).
- ▶ Mengaktifkan umpan data ke sistem pendukung keputusan khusus hilir seperti EIS dan penambangan data.

Dalam bab ini kita telah melihat arsitektur data warehouse dan komponennya dari berbagai sudut pandang. Kami membahas arsitektur yang berkaitan dengan tiga bidang utama yaitu akuisisi data, penyimpanan data, dan penyampaian informasi. Kemudian kami mempertimbangkan kerangka arsitektur secara keseluruhan dan fokus pada bagaimana arsitektur mendukung aliran data. Kami meninjau komponen manajemen dan kontrol dalam arsitektur.

Cakupan kami merinci arsitektur teknis di masing-masing dari tiga bidang utama. Di masing-masing dari tiga area ini, kami mempelajari arsitektur yang berkaitan dengan aliran data dan membuat daftar fungsi dan layanan. Seluruh pembahasan ini telah memberikan

Anda pemahaman yang kuat dan jelas tentang komponen arsitektur data warehouse, susunannya, dan fungsinya.

7.5 JENIS ARSITEKTUR

Pada Bab 2, kami memperkenalkan lima jenis arsitektur yang umum dan utama. Seperti disebutkan sebelumnya, jenis-jenis ini pada dasarnya berbeda dalam cara data diintegrasikan dan disimpan dan juga dalam cara “data warehouse” dan “data mart” saling terkait.

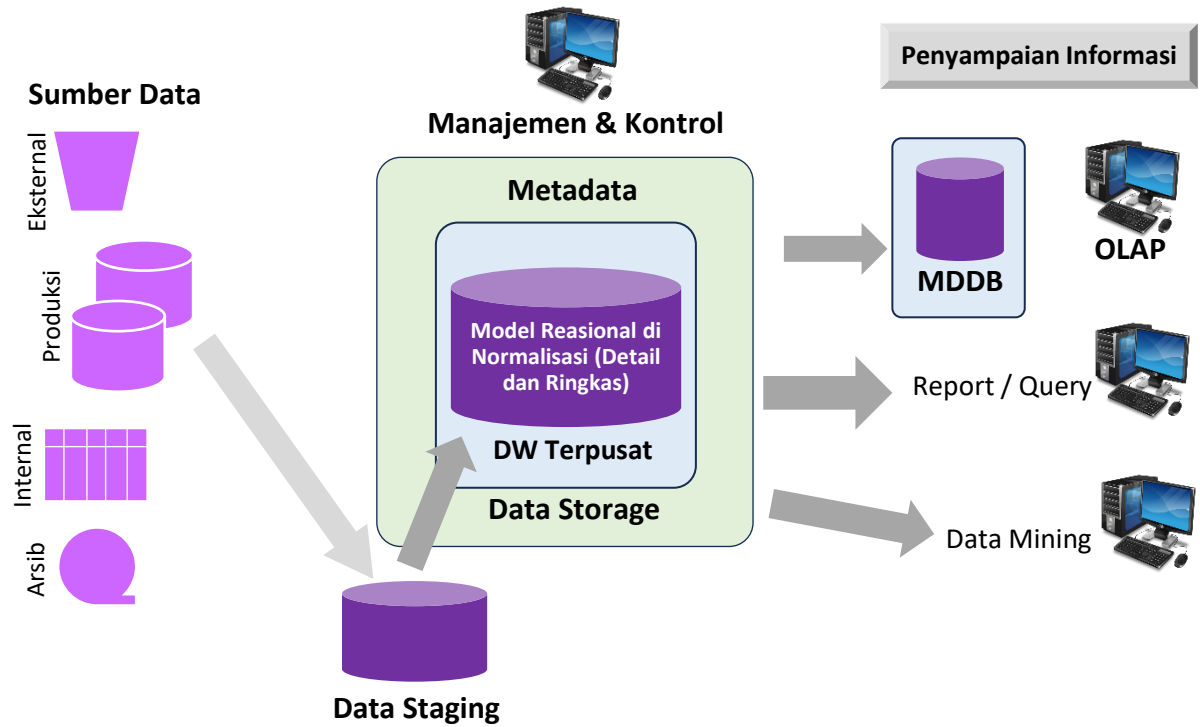
Pada titik ini, kami ingin meninjau kembali tipe-tipe arsitektur ini sehingga Anda dapat melihat keseluruhan pembahasan kami tentang arsitektur dalam bab ini dan melihat bagaimana penerapannya pada masing-masing dari lima tipe arsitektur umum ini. Perhatikan pengaturan dan keterkaitan “gudang data” dan “data mart” dalam setiap kasus jika memungkinkan. Perhatikan juga bagaimana pengaturan arsitektural memfasilitasi aliran data yang diinginkan seperti yang dibahas sebelumnya dalam bab ini.

Gudang Data Perusahaan Terpusat

Dalam tipe arsitektur ini, terdapat gudang data perusahaan terpusat. Tidak ada data mart, baik dependen maupun independen. Oleh karena itu semua penyampaian informasi berasal dari gudang data terpusat. Lihat Gambar 7.7 untuk gambaran umum komponen tingkat tinggi. Catat aliran data dari sistem sumber ke staging area, lalu ke gudang data pusat yang dinormalisasi, dan setelahnya ke pengguna akhir sebagai intelijen bisnis.

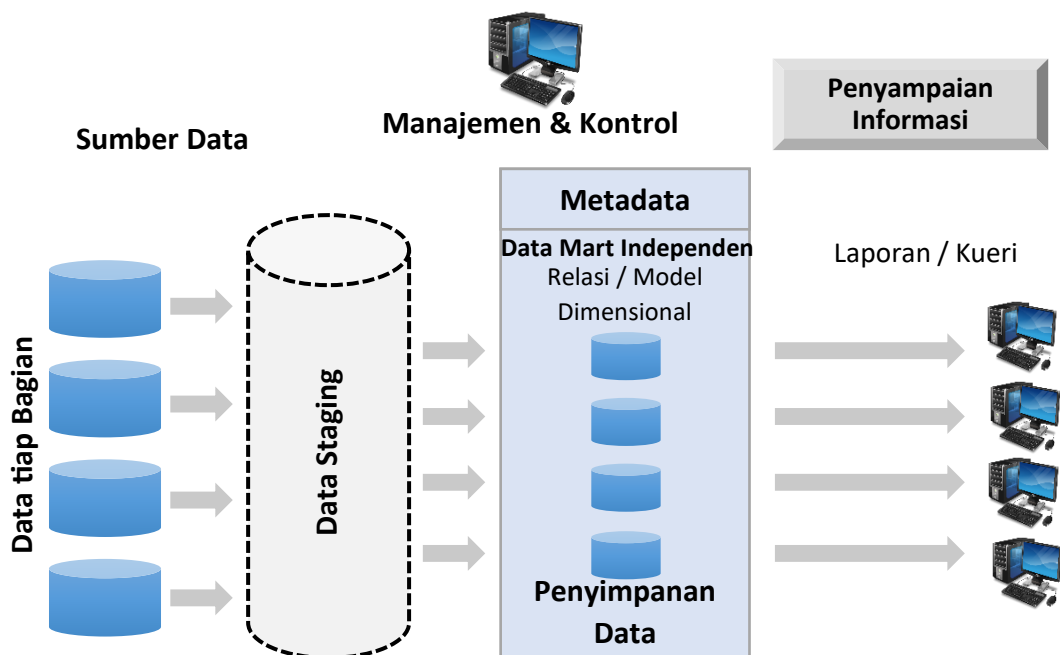
Data Mart Independen

Dalam tipe arsitektur ini, data warehouse sebenarnya merupakan kumpulan data mart yang tidak terhubung dan berbeda, masing-masing melayani departemen atau tujuan tertentu. Data mart di organisasi seperti ini biasanya berkembang seiring berjalannya waktu tanpa adanya perencanaan menyeluruh. Setiap data mart mengirimkan informasi ke kelompok penggunanya sendiri.

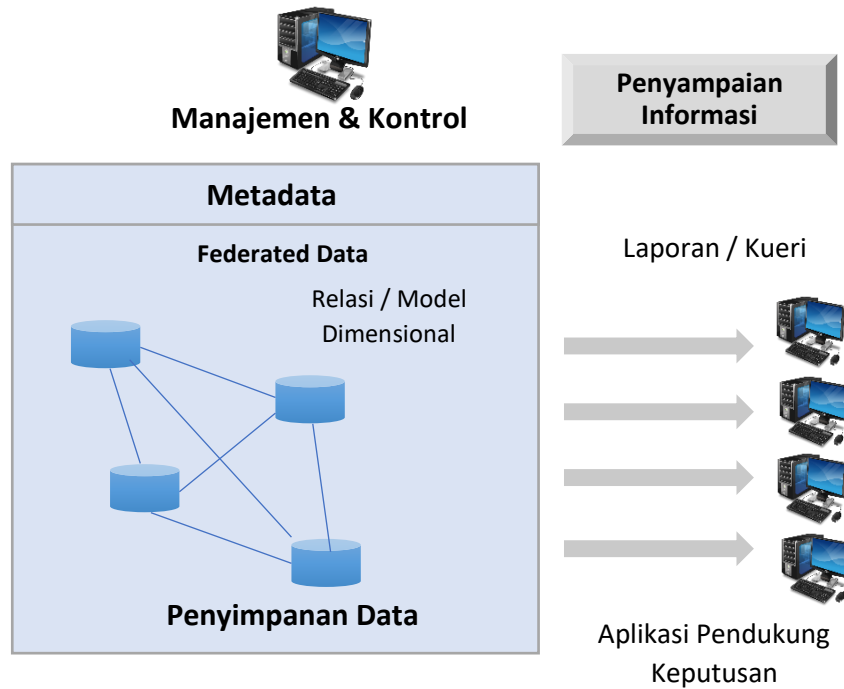


Gambar 7.7 Ikhtisar komponen gudang data terpusat.

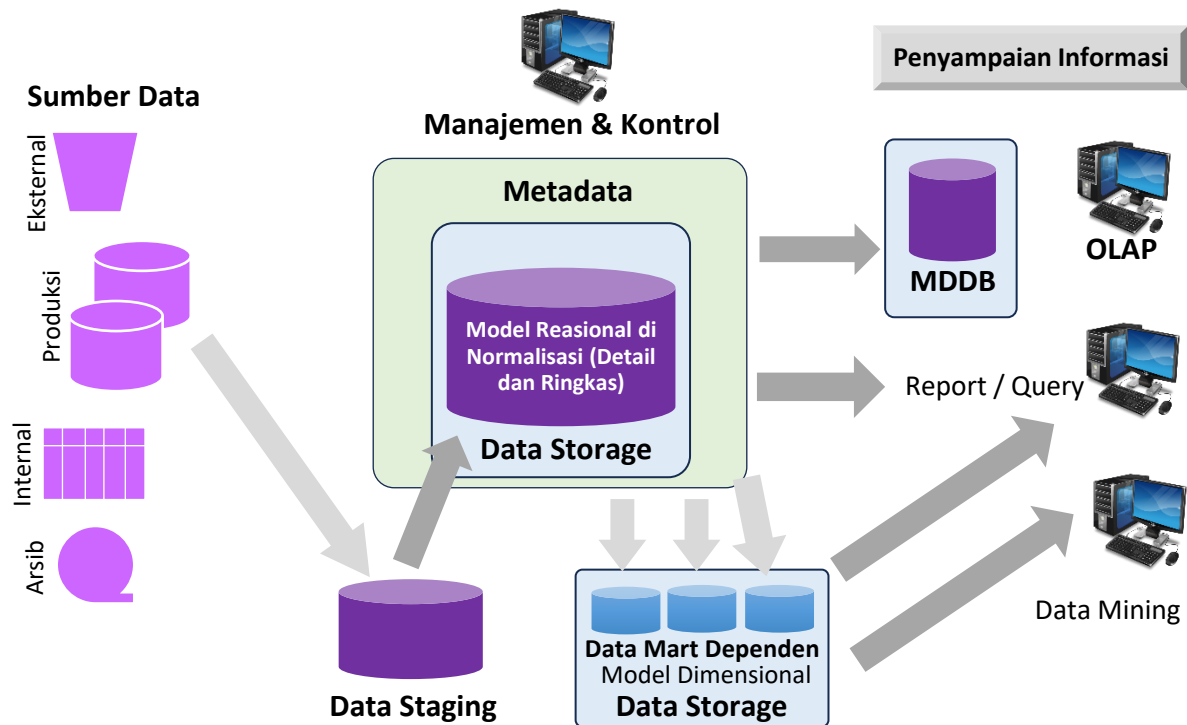
Lihat Gambar 7.8 untuk gambaran umum komponen tingkat tinggi. Catat aliran data dari sistem sumber ke staging area, kemudian ke berbagai data mart independen, dan selanjutnya ke kelompok pengguna akhir sebagai intelijen bisnis. Dalam banyak kasus, fungsi pementasan data dan perpindahan ke setiap data mart dapat dilakukan secara terpisah.



Gambar 7.8 Ikhtisar komponen data mart independen.



Gambar 7.9 Ikhtisar komponen gudang data gabungan.



Gambar 7.10 Ikhtisar komponen gudang data tipe hub-and-spoke.

Federasi

Tipe arsitektur ini tampaknya mirip dengan tipe data mart independen. Namun ada satu perbedaan besar. Dalam tipe arsitektur federasi, elemen data umum di berbagai data mart dan bahkan gudang data yang membentuk federasi diintegrasikan secara fisik atau logis. Tujuannya adalah mengupayakan satu versi kebenaran bagi organisasi; gudang data

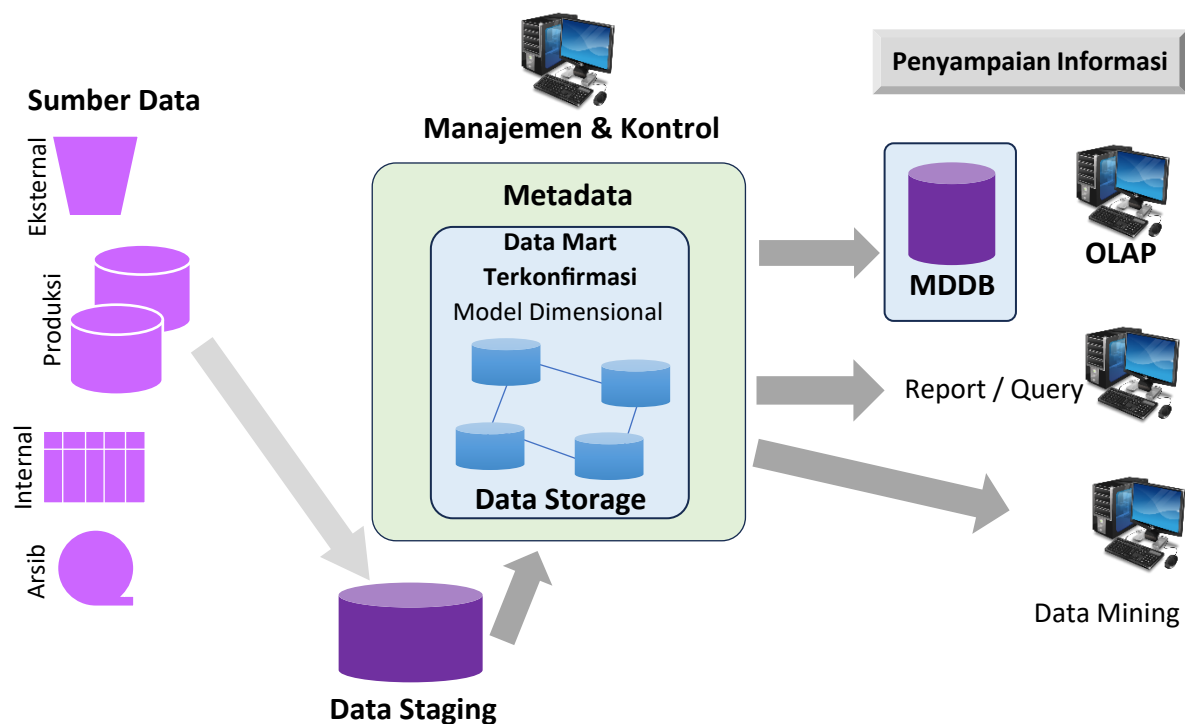
perusahaan terpusat hadir. Tidak ada data mart, baik dependen maupun independen. Oleh karena itu semua penyampaian informasi berasal dari gudang data terpusat.

Lihat Gambar 7.9 untuk gambaran umum komponen tingkat tinggi. Catat aliran data dari federasi data mart, gudang data, dan sumber lain ke pengguna akhir sebagai intelijen bisnis. Di antaranya, terjadi integrasi logis atau fisik dari elemen data umum.

Hub-dan-Spoke

Dalam tipe arsitektur ini, terdapat gudang data perusahaan terpusat. Selain itu, ada data mart yang bergantung pada gudang data perusahaan untuk umpan data. Oleh karena itu, pengiriman informasi dapat dilakukan dari gudang data terpusat dan data mart yang bergantung.

Lihat Gambar 7.10 untuk gambaran umum komponen tingkat tinggi. Catat aliran data dari sistem sumber ke staging area, lalu ke gudang data pusat yang dinormalisasi, dan setelah itu ke pengguna akhir sebagai intelijen bisnis baik dari gudang data pusat maupun data mart dependen.



Gambar 7.11 Ikhtisar komponen bus data-mart.

Bus Data-Mart

Dalam tipe arsitektur ini, tidak ada gudang data tunggal yang berbeda. Pengumpulan seluruh data mart membentuk data warehouse karena data mart tersebut merupakan “*super-mart*” karena dimensi bisnis dan fakta terukur disesuaikan dan dihubungkan di antara data mart. Semua penyampaian informasi berasal dari konglomerasi data mart yang telah disesuaikan. Data mart ini dapat melayani seluruh perusahaan, bukan hanya satu departemen saja.

Lihat Gambar 7.11 untuk gambaran umum komponen tingkat tinggi. Perhatikan aliran data dari sistem sumber ke staging area, lalu ke berbagai data mart yang disesuaikan, kemudian ke pengguna akhir sebagai intelijen bisnis dari data mart yang sesuai.

RINGKASAN BAB

- Arsitektur adalah struktur yang menyatukan semua komponen.
- Arsitektur data warehouse terdiri dari komponen-komponen berbeda dengan penyimpanan data read-only sebagai pusatnya.
- Beberapa tipe arsitektur data warehouse yang umum digunakan di berbagai organisasi. Secara umum, tipe ini mencerminkan bagaimana data disimpan dan tersedia—secara terpusat sebagai database gudang data perusahaan tunggal atau sebagai kumpulan data mart yang kohesif.
- Komponen arsitektur mendukung fungsi gudang data di tiga bidang utama yaitu akuisisi data, penyimpanan data, dan pengiriman informasi.
- Arsitektur data warehouse luas, kompleks, ekspansif, dan memiliki beberapa karakteristik yang membedakan.
- Kerangka arsitektur memungkinkan aliran data dari sumber data di satu sisi ke desktop pengguna di sisi lain.
- Arsitektur teknis gudang data adalah serangkaian fungsi dan layanan lengkap yang disediakan dalam komponen-komponennya. Ini mencakup prosedur dan aturan yang diperlukan untuk menjalankan fungsi dan menyediakan layanan. Ini mencakup penyimpanan data yang dibutuhkan setiap komponen untuk menyediakan layanan.
- Aliran data dari sistem sumber ke pengguna akhir sebagai intelijen bisnis bergantung pada tipe arsitektur.

Latihan Soal

1. Apa pemahaman Anda tentang arsitektur data warehouse? Jelaskan dalam satu atau dua paragraf.
2. Apa tiga area utama dalam data warehouse? Apakah ini pembagian yang logis? Jika ya, mengapa menurut Anda demikian? Hubungkan komponen arsitektur dengan tiga bidang utama.
3. Sebutkan empat karakteristik yang membedakan arsitektur data warehouse. Jelaskan masing-masing secara singkat.
4. Menelusuri aliran data melalui data warehouse dari awal hingga akhir.
5. Untuk penyampaian informasi, apa perbedaan antara pendekatan top-down dan bottom-up dalam implementasi data warehouse?
6. Komponen arsitektur manakah yang cocok dengan OLAP? Apa fungsi OLAP?
7. Mendefinisikan arsitektur teknis data warehouse. Bagaimana kaitannya dengan masing-masing komponen arsitektur?
8. Sebutkan lima fungsi dan layanan utama di area penyimpanan data.
9. Apa saja jenis repositori penyimpanan di area pementasan data?

10. Sebutkan empat fungsi dan layanan utama penyampaian informasi. Jelaskan masing-masing secara singkat.

LATIHAN TAMBAHAN

1. Tunjukkan apakah benar atau salah:
 - A. Arsitektur gudang data hanyalah pedoman keseluruhan. Ini bukan cetak biru untuk gudang data.
 - B. Dalam gudang data, komponen metadata bersifat unik, tanpa ada komponen yang benar-benar cocok dalam sistem operasional,
 - C. Biasanya, data mengalir dari penyimpanan gudang data ke area pementasan data.
 - D. Komponen manajemen dan pengendalian tidak berhubungan dengan seluruh operasi di gudang data.
 - E. Arsitektur teknis berarti alat vendor.
 - F. Bahasa berbasis SQL digunakan untuk mengekstrak data dari database hierarki.
 - G. Pengurutan dan penggabungan file adalah hal biasa di staging area.
 - H. MDDDB umumnya merupakan database relasional.
 - I. Terkadang, hasil kueri individual disimpan di penyimpanan data sementara untuk digunakan berulang kali.
 - J. Aplikasi khusus hilir diumpankan langsung dari komponen data sumber.
2. Anda baru saja dipromosikan menjadi administrator gudang data sebuah perusahaan asuransi mobil nasional. Anda diminta menyiapkan daftar periksa untuk memilih alat vendor yang tepat untuk membantu Anda dalam administrasi gudang data. Buatlah daftar fungsi dalam komponen manajemen dan kontrol arsitektur gudang data Anda. Gunakan daftar ini untuk mendapatkan daftar periksa pemilihan alat.
3. Sebagai analis senior yang bertanggung jawab atas pementasan data, Anda bertanggung jawab atas desain area pementasan data. Jika gudang data Anda mendapat masukan dari beberapa sistem lama di berbagai platform, dan juga masukan reguler dari dua sumber eksternal, bagaimana Anda mengatur area pementasan data Anda? Jelaskan repositori data yang Anda miliki untuk pementasan data.
4. Anda adalah arsitek gudang data untuk jaringan department store nasional terkemuka. Gudang data telah aktif dan berjalan selama hampir satu tahun. Sekarang manajemen telah memutuskan untuk menyediakan fasilitas OLAP kepada power user. Bagaimana Anda mengubah komponen pengiriman informasi arsitektur gudang data Anda? Buat asumsi yang realistis dan lanjutkan.
5. Anda baru saja bergabung sebagai spesialis ekstraksi data di tim proyek gudang data yang mengembangkan data mart yang sesuai untuk apotek lokal namun progresif. Buat daftar detail fungsi dan layanan untuk ekstraksi data, transformasi data, dan pementasan data.

BAB 8

INFRASTRUKTUR SEBAGAI YAYASAN UNTUK DATA WAREHOUSING

TUJUAN BAB

- Memahami perbedaan antara arsitektur dan infrastruktur
- Cari tahu bagaimana infrastruktur data warehouse mendukung arsitekturnya
- Dapatkan wawasan tentang komponen infrastruktur fisik
- Meninjau perangkat keras dan sistem operasi untuk gudang data
- Pelajari opsi pemrosesan paralel yang berlaku pada gudang data
- Diskusikan opsi server secara detail
- Pelajari cara memilih DBMS
- Tinjau jenis alat yang diperlukan untuk gudang data
- Mempelajari konsep dan penggunaan peralatan data warehouse

Apa infrastruktur gudang data dalam kaitannya dengan arsitekturnya? Apa perbedaan antara arsitektur dan infrastruktur? Dalam hal apa mereka berbeda? Mengapa kita harus mempelajari keduanya secara terpisah?

Pada bab sebelumnya, kita membahas arsitektur data warehouse secara rinci. Kami melihat berbagai komponen arsitektur dan mempelajarinya dengan mengelompokkannya ke dalam tiga area utama gudang data, yaitu akuisisi data, penyimpanan data, dan pengiriman informasi. Anda mempelajari elemen-elemen yang menyusun arsitektur teknis dari setiap komponen arsitektur. Anda juga meninjau tipe arsitektur berbeda yang umum dan bagaimana aliran data didukung di setiap tipe.

Dalam bab ini, mari kita cari tahu apa yang dimaksud dengan infrastruktur dan apa saja yang termasuk di dalamnya. Kami akan membahas setiap bagian dari infrastruktur gudang data. Anda akan memahami pentingnya infrastruktur dan menguasai teknik untuk menciptakan infrastruktur yang tepat untuk gudang data Anda.

8.1 ARSITEKTUR PENDUKUNG INFRASTRUKTUR

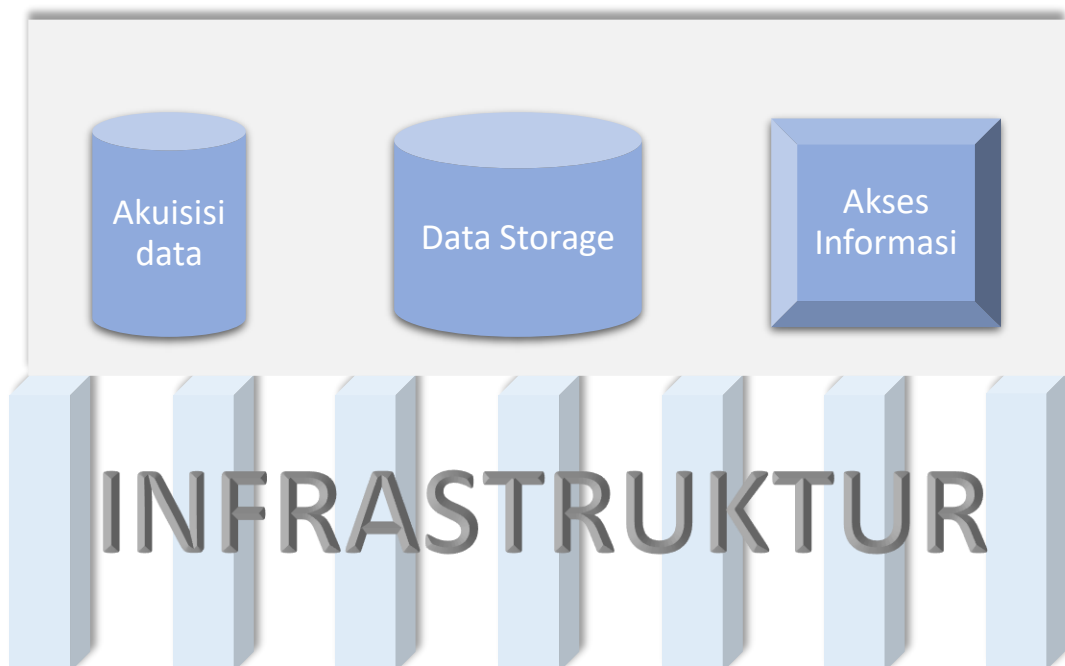
Pertimbangkan komponen arsitektur. Misalnya, mari kita ambil arsitektur teknis komponen pementasan data. Bagian arsitektur teknis gudang data Anda ini melakukan beberapa hal. Pertama-tama, ini menunjukkan bahwa ada bagian arsitektur yang relevan dengan pementasan data. Kemudian dicatat bahwa bagian arsitektur ini berisi area di mana data dipentaskan sebelum dimuat ke dalam repositori gudang data. Selanjutnya, ini menunjukkan bahwa bagian arsitektur ini menjalankan fungsi tertentu dan menyediakan layanan khusus di gudang data. Fungsi dan layanannya antara lain meliputi transformasi data dan pembersihan data.

Sekarang mari kita mengajukan beberapa pertanyaan. Di manakah tepatnya area pementasan data? Apa saja file dan database spesifiknya? Bagaimana fungsi-fungsi tersebut dijalankan? Apa yang memungkinkan layanan disediakan? Apa dasar yang mendasarinya? Apa

struktur dasarnya? Infrastruktur merupakan fondasi pendukung arsitektur. Gambar 8.1 mengungkapkan fakta ini dengan cara yang sederhana.

Apa sajakah berbagai elemen yang dibutuhkan untuk mendukung arsitektur? Infrastruktur dasar mencakup banyak elemen. Pertama, terdiri dari platform komputasi dasar. Platform ini mencakup semua perangkat keras dan sistem operasi yang diperlukan. Berikutnya, sistem manajemen basis data (DBMS) merupakan elemen penting dari infrastruktur. Semua jenis perangkat lunak dan alat lainnya juga merupakan bagian dari infrastruktur. Bagaimana dengan orang-orang dan prosedur yang membuat arsitektur menjadi hidup? Apakah ini juga merupakan bagian dari infrastruktur? Dalam arti tertentu, memang demikian.

Infrastruktur gudang data mencakup semua elemen dasar yang memungkinkan arsitektur diimplementasikan. Singkatnya, infrastruktur mencakup beberapa elemen seperti perangkat keras server, sistem operasi, perangkat lunak jaringan, perangkat lunak database, LAN dan WAN, alat vendor untuk setiap komponen arsitektur, orang, prosedur, dan pelatihan.



Gambar 8.1 Arsitektur pendukung infrastruktur.

Elemen infrastruktur gudang data dapat diklasifikasikan menjadi dua kategori: infrastruktur operasional dan infrastruktur fisik. Perbedaan ini penting karena unsur-unsur dalam setiap kategori mempunyai sifat dan ciri yang berbeda dibandingkan dengan unsur-unsur dalam kategori lainnya. Pertama, kita akan membahas elemen-elemen yang dapat dikelompokkan sebagai infrastruktur operasional. Infrastruktur fisik jauh lebih luas dan mendasar. Setelah memperoleh pemahaman dasar tentang unsur-unsur arsitektur fisik, kita akan menghabiskan sebagian besar bab ini untuk mengkaji unsur-unsur tertentu secara lebih rinci.

Infrastruktur Operasional

Untuk memahami infrastruktur operasional, mari kita ambil contoh pementasan data lagi. Salah satu bagian dari infrastruktur dasar mengacu pada perangkat keras komputasi dan perangkat lunak terkait. Anda memerlukan perangkat keras dan perangkat lunak untuk menjalankan fungsi pementasan data dan memberikan layanan yang sesuai. Anda memerlukan alat perangkat lunak untuk melakukan transformasi data. Anda memerlukan perangkat lunak untuk membuat file keluaran. Anda memerlukan perangkat keras disk untuk menempatkan data di file staging area. Namun bagaimana dengan orang-orang yang terlibat dalam menjalankan fungsi-fungsi ini? Bagaimana dengan aturan bisnis dan prosedur transformasi data? Bagaimana dengan perangkat lunak manajemen untuk memantau dan mengelola tugas transformasi data?

Infrastruktur operasional untuk mendukung setiap komponen arsitektur terdiri dari

- Rakyat
- Prosedur
- Pelatihan
- Perangkat lunak manajemen

Ini bukanlah orang-orang dan prosedur yang diperlukan untuk mengembangkan gudang data. Ini diperlukan untuk menjaga gudang data tetap berjalan. Elemen-elemen ini sama pentingnya dengan perangkat keras dan perangkat lunak yang menjaga agar data warehouse tetap berjalan. Mereka mendukung pengelolaan gudang data dan menjaga efisiensinya.

Pengembang gudang data menaruh banyak perhatian pada elemen perangkat keras dan perangkat lunak sistem dari infrastruktur. Adalah benar untuk melakukan hal tersebut. Namun infrastruktur operasional sering kali diabaikan. Meskipun Anda mungkin memiliki perangkat keras dan perangkat lunak yang tepat, gudang data Anda memerlukan infrastruktur operasional agar dapat berfungsi dengan baik. Tanpa infrastruktur operasional yang tepat, gudang data Anda kemungkinan besar akan berjalan pincang dan tidak lagi efektif. Perhatikan detail infrastruktur operasional Anda.

Infrastruktur Fisik

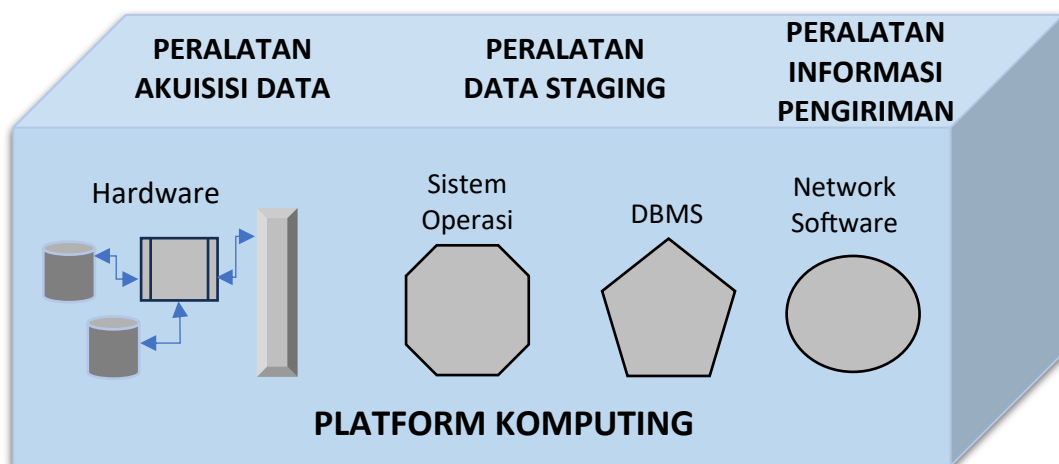
Mari kita mulai dengan diagram. Gambar 8.2 menyoroti elemen utama infrastruktur fisik. Apa yang Anda lihat pada diagram? Seperti yang Anda ketahui, setiap sistem, termasuk gudang data Anda, harus memiliki platform keseluruhan untuk ditempatkan. Pada dasarnya, platform terdiri dari komponen perangkat keras dasar, sistem operasi dengan perangkat lunak utilitasnya, jaringan, dan perangkat lunak jaringan. Bersamaan dengan platform keseluruhan adalah seperangkat alat yang berjalan pada platform yang dipilih untuk melakukan berbagai fungsi dan layanan dari masing-masing komponen arsitektur.

Kami akan memeriksa elemen infrastruktur fisik dalam beberapa bagian berikutnya. Keputusan tentang perangkat keras berada di urutan teratas dalam daftar keputusan yang harus Anda buat tentang infrastruktur gudang data Anda. Keputusan perangkat keras tidaklah mudah. Anda harus mempertimbangkan banyak faktor. Anda harus memastikan bahwa perangkat keras yang dipilih akan mendukung seluruh arsitektur gudang data.

Mungkin kita bisa kembali ke masa mainframe dan mendapatkan beberapa petunjuk bermanfaat. Ketika model mainframe korporat yang lebih baru diumumkan dan ketika kami kehabisan tenaga pada konfigurasi saat ini, kami berpegang pada dua prinsip. Pertama, kami memanfaatkan sebanyak mungkin infrastruktur fisik yang ada. Berikutnya, kami menjaga infrastruktur se-modular mungkin. Ketika kebutuhan muncul dan versi yang lebih baru tersedia dengan harga lebih murah, kami mencabut komponen yang ada dan memasang komponen pengganti.

Di gudang data Anda, cobalah menerapkan dua prinsip ini. Anda sudah memiliki komponen perangkat keras dan sistem operasi di perusahaan Anda yang mendukung operasional saat ini. Berapa banyak yang dapat Anda gunakan untuk gudang data Anda? Berapa kapasitas tambahan yang tersedia? Berapa banyak ruang disk yang dapat disisihkan untuk penyimpanan gudang data? Temukan jawaban atas pertanyaan-pertanyaan ini.

Dengan menerapkan pendekatan modular, dapatkah Anda menambahkan lebih banyak prosesor ke perangkat keras server? Jelajahi apakah Anda dapat mengakomodasi gudang data dengan menambahkan lebih banyak disk atau unit penyimpanan lainnya. Lakukan inventarisasi masing-masing komponen perangkat keras. Periksa komponen mana yang perlu diganti dengan versi yang lebih kuat. Buatlah juga daftar komponen tambahan yang harus dibeli dan dipasang.



Gambar 8.2 Infrastruktur fisik

8.2 PERANGKAT KERAS DAN SISTEM OPERASI

Perangkat keras dan sistem operasi membentuk lingkungan komputasi untuk gudang data Anda. Semua pekerjaan ekstraksi, transformasi, integrasi, dan staging data dijalankan pada perangkat keras yang dipilih dalam sistem operasi yang dipilih. Saat Anda memindahkan data yang terkonsolidasi dan terintegrasi dari staging area ke repositori gudang data, Anda menggunakan perangkat keras server dan perangkat lunak sistem operasi. Ketika query dimulai dari workstation klien, perangkat keras server, bersama dengan perangkat lunak database, mengeksekusi query dan menghasilkan hasilnya.

Berikut adalah beberapa pedoman umum untuk pemilihan perangkat keras, tidak sepenuhnya spesifik untuk perangkat keras untuk data warehouse:

- 1) **Skalabilitas:** Pastikan perangkat keras pilihan Anda dapat ditingkatkan ketika gudang data Anda bertambah dalam hal jumlah pengguna, jumlah kueri, dan kompleksitas kueri.
- 2) **Mendukung:** Dukungan vendor sangat penting untuk pemeliharaan perangkat keras. Pastikan dukungan dari vendor perangkat keras berada pada tingkat setinggi mungkin.
- 3) **Referensi Vendor:** Penting untuk memeriksa referensi vendor dengan situs lain yang menggunakan perangkat keras dari vendor ini. Anda tidak ingin terjebak ketika gudang data Anda sedang down karena kerusakan perangkat keras ketika CEO ingin beberapa analisis penting diselesaikan.
- 4) **Stabilitas Vendor:** Periksa stabilitas dan daya tahan vendor.

Selanjutnya mari kita pertimbangkan dengan cepat beberapa kriteria umum untuk pemilihan sistem operasi. Pertama-tama, sistem operasi harus kompatibel dengan perangkat keras. Daftar kriterianya berikut ini.

- ❖ **Skalabilitas:** Sekali lagi, skalabilitas adalah yang pertama dalam daftar karena ini adalah salah satu fitur umum dari setiap gudang data. Gudang data tumbuh dan berkembang sangat cepat. Seiring dengan perangkat keras dan perangkat lunak database, sistem operasi harus mampu mendukung peningkatan jumlah pengguna dan aplikasi.
- ❖ **Keamanan:** Ketika beberapa stasiun kerja klien mengakses server, sistem operasi harus mampu melindungi setiap klien dan sumber daya terkait. Sistem operasi harus menyediakan lingkungan yang aman bagi setiap klien. Keandalan. Sistem operasi harus mampu melindungi lingkungan dari malfungsi aplikasi. Ketersediaan. Ini adalah akibat wajar dari keandalan. Lingkungan komputasi harus terus tersedia setelah penghentian aplikasi yang tidak normal.
- ❖ **Multitugas Preemptif:** Perangkat keras server harus mampu menyeimbangkan alokasi waktu dan sumber daya di antara banyak tugas. Selain itu, sistem operasi harus dapat membiarkan tugas dengan prioritas lebih tinggi mendahului atau mengganggu tugas lain jika diperlukan. Gunakan Pendekatan Multithread. Sistem operasi harus mampu melayani beberapa permintaan secara bersamaan dengan mendistribusikan thread ke beberapa prosesor dalam konfigurasi perangkat keras multiprosesor. Fitur ini sangat penting karena konfigurasi multiprosesor merupakan arsitektur pilihan dalam lingkungan data warehouse.
- ❖ **Perlindungan Memori:** Sekali lagi, dalam lingkungan gudang data, sejumlah besar pertanyaan adalah hal biasa. Artinya, beberapa kueri akan dijalankan secara bersamaan. Fitur perlindungan memori dalam sistem operasi mencegah satu tugas melanggar ruang memori tugas lainnya.

Setelah meninjau persyaratan perangkat keras dan sistem operasi di lingkungan gudang data, mari kita coba mempersempit pilihan. Apa saja pilihan yang mungkin? Tinjau daftar tiga opsi umum berikut.

Mainframe

- ◆ Perangkat keras sisa dari aplikasi lama
- ◆ Terutama dirancang untuk OLTP dan bukan untuk aplikasi pendukung keputusan
- ◆ Tidak hemat biaya untuk data warehousing
- ◆ Tidak mudah diskalakan
- ◆ Jarang digunakan untuk data warehousing; hanya digunakan ketika terlalu banyak sumber daya cadangan yang tersedia untuk data mart yang lebih kecil

Buka Server Sistem

- ◆ Server UNIX, media pilihan bagi sebagian besar gudang data
- ◆ Secara umum kuat
- ◆ Diadaptasi untuk pemrosesan paralel

Server NT

- ◆ Mendukung gudang data berukuran sedang
- ◆ Kemampuan pemrosesan paralel yang terbatas
- ◆ Hemat biaya untuk gudang data berukuran sedang dan kecil

Opsi Platform

Sekarang mari kita mengalihkan perhatian kita ke platform komputasi yang diperlukan untuk menjalankan beberapa fungsi dari berbagai komponen arsitektur data warehouse. Platform komputasi adalah kumpulan komponen perangkat keras, sistem operasi, jaringan, dan perangkat lunak jaringan. Baik itu fungsi sistem OLTP atau sistem pendukung keputusan seperti gudang data, fungsi tersebut harus dijalankan pada platform komputasi.

Sebelum kita membahas lebih dalam mengenai opsi platform, mari kita kembali ke fungsi dan layanan komponen arsitektur di tiga bidang utama. Berikut ringkasan ringkasan singkatnya:

- **Akuisisi Data:** ekstraksi data, transformasi data, pembersihan data, integrasi data, dan pementasan data.
- **Penyimpanan Data:** pemuatan data, pengarsipan, dan pengelolaan data.
- **Pengiriman Informasi:** pembuatan laporan, pemrosesan kueri, dan analisis kompleks.

Kami sekarang akan membahas opsi platform dalam kaitannya dengan fungsi di ketiga area ini. Di mana setiap fungsi harus dilaksanakan? Di platform mana? Bagaimana Anda bisa mengoptimalkan fungsinya?

Opsi Platform Tunggal Ini adalah opsi paling mudah dan sederhana untuk mengimplementasikan arsitektur gudang data. Dalam opsi ini, semua fungsi mulai dari ekstraksi data back-end hingga pemrosesan kueri front-end dilakukan pada platform komputasi tunggal. Ini mungkin merupakan pendekatan paling awal, ketika para pengembang mengimplementasikan gudang data pada mainframe, komputer mini, atau server tunggal berbasis UNIX yang ada.

Karena semua operasi di area akuisisi data, penyimpanan data, dan pengiriman informasi dilakukan pada platform yang sama, opsi ini hampir tidak pernah menemui masalah kompatibilitas atau antarmuka. Data mengalir dengan lancar dari awal hingga akhir tanpa

konversi platform ke platform. Tidak diperlukan middleware. Semua alat bekerja dalam satu lingkungan komputasi.

Di banyak perusahaan, sistem lama masih berjalan di mainframe atau komputer mini. Beberapa dari perusahaan ini telah bermigrasi ke server berbasis UNIX dan yang lainnya telah pindah ke sistem ERP di lingkungan klien/server. Sebagian besar sistem lawas masih berada di server berbasis mainframe, mini, atau UNIX. Apa hubungan sistem lama dengan gudang data? Ingat, sistem lama menyumbangkan sebagian besar data gudang data. Jika perusahaan-perusahaan ini ingin mengadopsi solusi platform tunggal, platform pilihan tersebut haruslah server berbasis mainframe, mini, atau UNIX.

Jika situasi di perusahaan Anda memerlukan pertimbangan serius terhadap opsi platform tunggal, maka analisislah implikasinya sebelum mengambil keputusan. Solusi platform tunggal nampaknya merupakan pilihan ideal. Jika ya, mengapa saat ini tidak banyak perusahaan yang mengadopsi opsi ini? Mari kita periksa alasannya.

Platform Lama yang Diperluas hingga Kapasitasnya Di banyak perusahaan, lingkungan komputasi lama yang ada mungkin telah ada selama bertahun-tahun dan telah sepenuhnya mencapai kapasitasnya. Lingkungan mungkin berada pada titik di mana tidak dapat lagi ditingkatkan untuk mengakomodasi gudang data Anda.

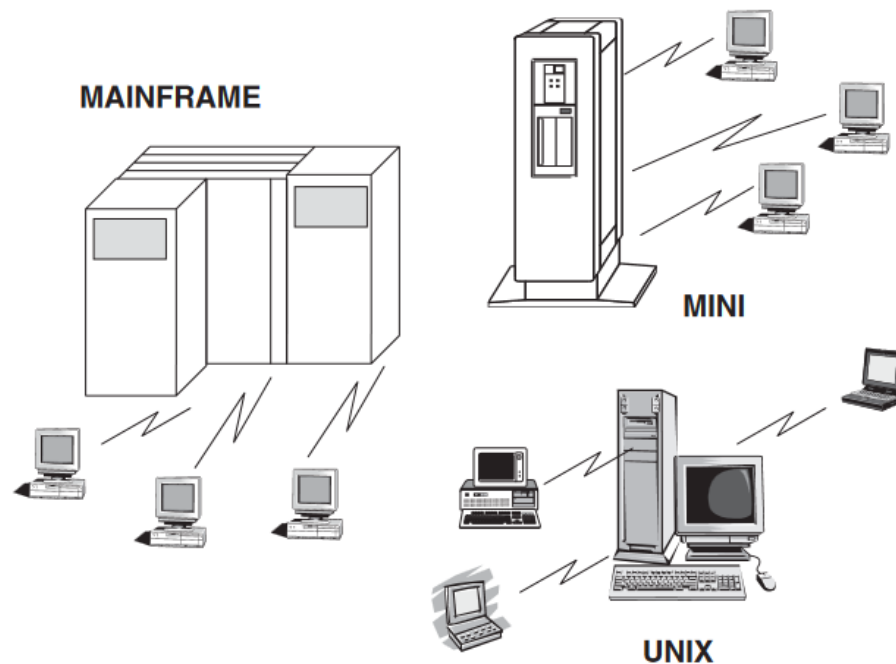
Ketidakterdediaan Alat Alat perangkat lunak merupakan bagian besar dari infrastruktur gudang data. Anda akan dengan jelas memahami fakta ini dari beberapa subbagian terakhir bab ini. Sebagian besar alat yang disediakan oleh banyak vendor gudang data tidak mendukung lingkungan mainframe atau komputer mini. Tanpa alat yang tepat dalam infrastruktur, gudang data Anda akan berantakan.

Berbagai Platform Lama Meskipun kami menduga bahwa lingkungan mainframe atau komputer mini lama dapat diperluas hingga mencakup pergudangan data, fakta praktisnya menunjukkan situasi yang berbeda. Di sebagian besar perusahaan, kombinasi beberapa sistem mainframe, beragam aplikasi komputer mini, dan segelintir sistem berbasis PC yang lebih baru hadir berdampingan. Jalur yang diambil sebagian besar perusahaan adalah dari main-frame ke mini dan kemudian ke PC. Gambar 8.3 menyoroti konfigurasi tipikal.

Jika perusahaan Anda merupakan salah satu perusahaan pada umumnya, apa yang dapat Anda lakukan dengan solusi platform tunggal? Tidak banyak. Dengan konglomerasi platform yang berbeda-beda, opsi platform tunggal yang memiliki gudang data Anda bersama dengan semua aplikasi lainnya tidak dapat dipertahankan.

Kebijakan Migrasi Perusahaan Ini merupakan pertimbangan penting lainnya. Anda mengetahui dengan baik beragam manfaat arsitektur klien/server untuk komputasi. Anda juga menyadari fakta bahwa setiap perusahaan telah berubah untuk merangkul paradigma komputasi ini dengan memindahkan aplikasi dari platform mainframe dan komputer mini. Di sebagian besar perusahaan, kebijakan penggunaan teknologi informasi tidak mengizinkan pelestarian platform lama. Jika perusahaan Anda memiliki kebijakan serupa, maka Anda tidak akan diizinkan untuk menambahkan sistem penting lainnya seperti gudang data Anda pada platform lama.

Opsi Hibrid Setelah memeriksa sistem lama dan aplikasi yang lebih modern di perusahaan Anda, kemungkinan besar Anda akan memutuskan bahwa pendekatan platform tunggal tidak dapat diterapkan untuk gudang data Anda. Ini adalah kesimpulan yang diambil sebagian besar perusahaan. Di sisi lain, jika perusahaan Anda termasuk dalam kategori di mana platform lama akan mengakomodasi gudang data Anda, maka ambillah pendekatan solusi platform tunggal. Sekali lagi, solusi platform tunggal, jika memungkinkan, merupakan solusi yang lebih mudah.



Gambar 8.3 Berbagai platform di suatu perusahaan pada umumnya.

Bagi kita yang tidak seberuntung itu, kita harus mempertimbangkan pilihan lain. Mari kita mulai dengan ekstraksi data, operasi besar pertama, dan ikuti aliran data hingga dikonsolidasikan ke dalam gambar pemuatan dan menunggu di area pementasan. Kami sekarang akan menelusuri aliran data dan memeriksa opsi platform.

Ekstraksi Data Di gudang data mana pun, yang terbaik adalah menjalankan fungsi ekstraksi data dari setiap sistem sumber pada platform komputasinya sendiri. Jika data penjualan telepon Anda berada di lingkungan komputer mini, buat file ekstrak di komputer mini itu sendiri untuk penjualan telepon. Jika aplikasi pesanan lewat pos Anda dijalankan di mainframe menggunakan database IMS, buatlah file ekstrak untuk pesanan lewat pos di platform mainframe. Jarang sekali ada tindakan bijaksana untuk menyalin semua file database pesanan lewat pos ke platform lain dan kemudian melakukan ekstraksi data.

Pemformatan Ulang dan Penggabungan Awal Setelah membuat ekstrak data mentah dari berbagai sumber, file yang diekstrak dari masing-masing sumber diformat ulang dan digabungkan menjadi file ekstrak dalam jumlah yang lebih kecil. Verifikasi data yang diekstraksi terhadap laporan sistem sumber dan rekonsiliasi jumlah catatan input dan output dilakukan pada langkah ini. Sama seperti langkah ekstraksi, yang terbaik adalah melakukan

langkah penggabungan awal setiap kumpulan ekstrak sumber pada platform sumber itu sendiri.

Pembersihan Data Awal Pada langkah ini, Anda memverifikasi data yang diekstrak dari setiap sumber data untuk setiap nilai yang hilang di masing-masing bidang, memberikan nilai default, dan melakukan pengeditan dasar. Ini adalah langkah lain untuk platform komputasi dari sistem sumber itu sendiri. Namun, di beberapa gudang data, pembersihan data jenis ini terjadi setelah data dari semua sumber direkonsiliasi dan dikonsolidasikan. Dalam kedua kasus tersebut, fitur dan kondisi data dari sistem sumber Anda menentukan kapan dan di mana langkah ini harus dilakukan untuk gudang data Anda.

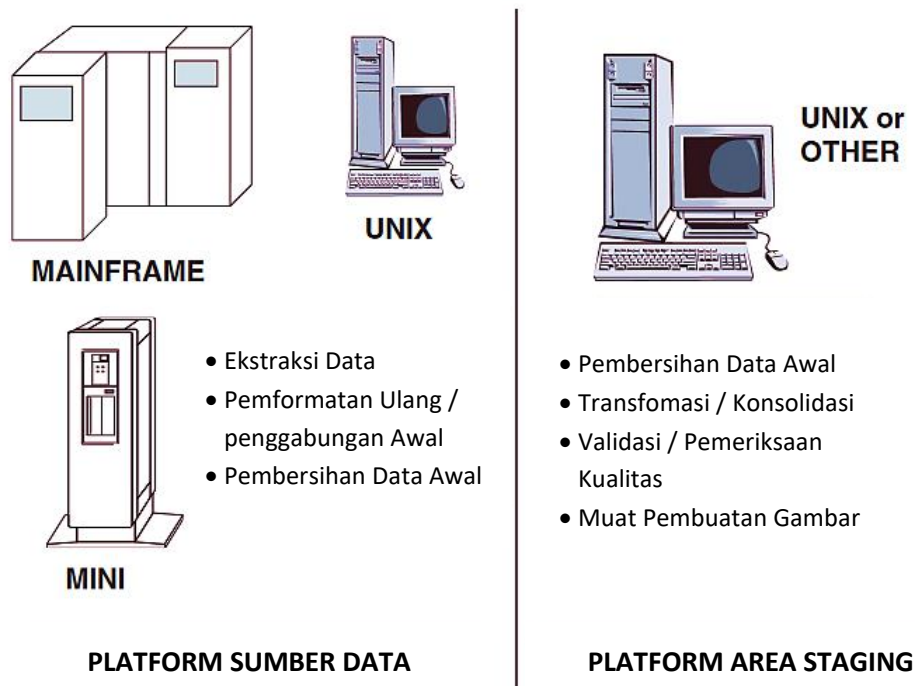
Transformasi dan Konsolidasi Langkah ini mencakup semua fungsi transformasi dan integrasi data utama. Biasanya, Anda akan menggunakan perangkat lunak transformasi untuk tujuan ini. Di manakah tempat terbaik untuk melakukan langkah ini? Tentu saja, tidak di platform lama mana pun. Anda melakukan langkah ini pada platform tempat staging area Anda berada.

Validasi dan Pemeriksaan Kualitas Akhir Langkah validasi akhir dan pemeriksaan kualitas ini merupakan kandidat kuat untuk staging area. Anda akan mengatur agar langkah ini terjadi pada platform tersebut. Pembuatan Gambar Muat Langkah ini membuat gambar pemuatan untuk file database individual dari repositori gudang data. Langkah ini hampir selalu terjadi di area pementasan dan, oleh karena itu, di platform tempat area pementasan berada. Gambar 8.4 merangkum langkah-langkah akuisisi data dan platform terkait. Anda akan melihat opsi langkah-langkahnya. Kaitkan hal ini dengan lingkungan perusahaan Anda dan tentukan di mana langkah akuisisi data harus dilakukan.

Opsi untuk Area Pementasan Dalam pembahasan langkah-langkah akuisisi data, kami telah menyoroti platform komputasi yang optimal untuk setiap langkah. Anda akan melihat bahwa langkah-langkah penting terjadi di area pementasan. Ini adalah tempat di mana semua data untuk data warehouse dikumpulkan dan dipersiapkan. Platform apa yang ideal untuk area pementasan? Mari kami ulangi bahwa platform yang paling cocok untuk area pementasan Anda bergantung pada status platform sumber Anda. Namun demikian, mari kita jelajahi opsi untuk menempatkan area pementasan dan menghasilkan pedoman umum. Ini akan membantu Anda memutuskan. Gambar 8.5 menunjukkan berbagai pilihan untuk staging area. Pelajari gambar tersebut dan ikuti amplifikasi opsi yang diberikan pada subbagian di bawah ini.

Di Salah Satu Platform Lama Jika sebagian besar sumber data lama Anda berada di platform yang sama dan jika kapasitas tambahan sudah tersedia, pertimbangkan untuk mempertahankan area pementasan data Anda di platform lama tersebut. Dalam opsi ini, Anda akan menghemat waktu dan tenaga dalam memindahkan data lintas platform ke staging area. Pada Platform Penyimpanan Data Ini adalah platform di mana DBMS gudang data berjalan dan database berada. Saat Anda menyimpan area pementasan data pada platform ini, Anda akan menyadari semua keuntungan menerapkan gambar pemuatan ke database. Anda bahkan mungkin dapat menghilangkan beberapa sublangkah perantara dan menerapkan data langsung ke database dari beberapa file konsolidasi di staging area.

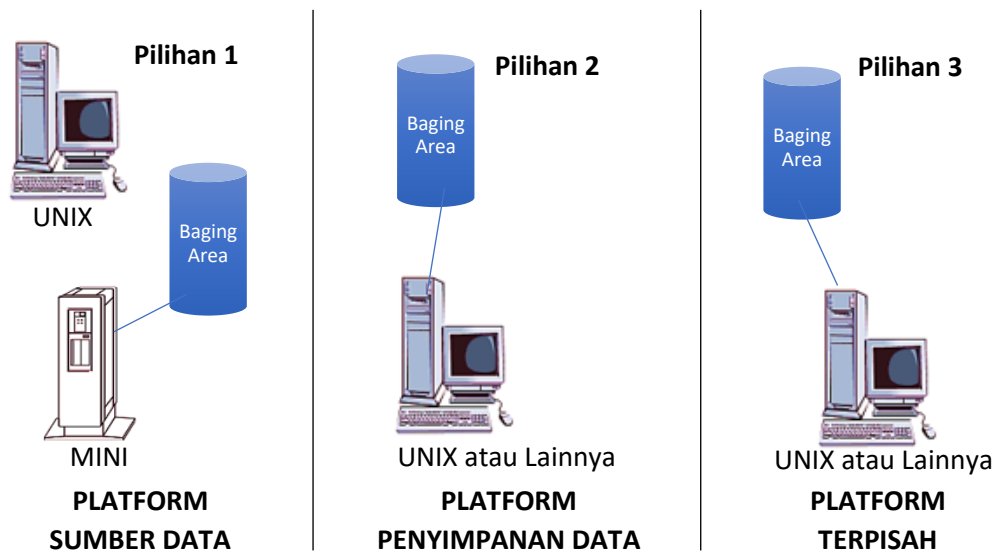
Pada Platform Optimal Terpisah Anda dapat meninjau platform sumber data Anda, memeriksa platform penyimpanan gudang data, dan kemudian memutuskan bahwa tidak satu pun dari platform ini yang benar-benar cocok untuk area pementasan Anda. Kemungkinan besar lingkungan Anda memerlukan transformasi data yang kompleks. Ada kemungkinan Anda perlu mengerjakan data Anda secara menyeluruh untuk membersihkan dan mempersiapkannya untuk gudang data Anda. Dalam keadaan seperti itu, Anda memerlukan platform terpisah untuk menyimpan data Anda sebelum memuat ke database.



Gambar 8.4 Platform untuk akuisisi data.

Berikut adalah beberapa keuntungan berbeda dari platform terpisah untuk pementasan data:

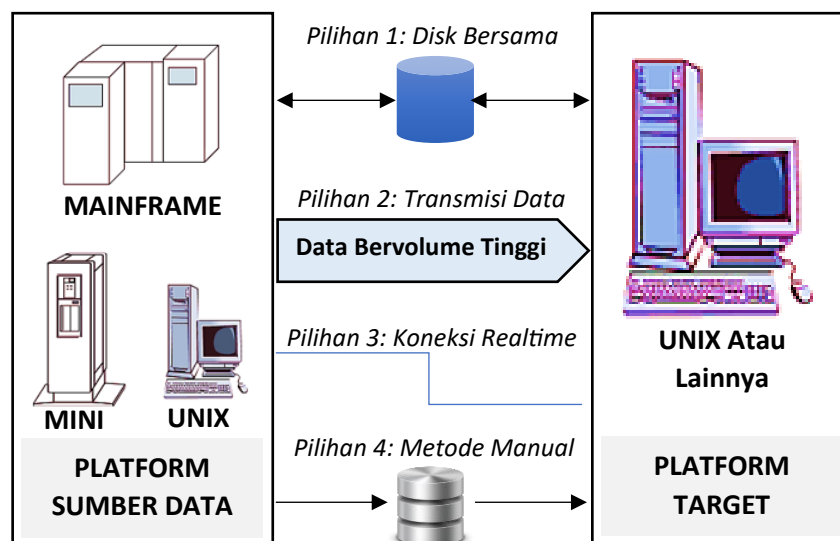
- Anda dapat mengoptimalkan platform terpisah untuk transformasi data yang kompleks dan pembersihan data. Apa yang kita maksud dengan ini? Anda dapat melengkapi platform netral dengan semua alat yang diperlukan untuk transformasi data, pembersihan data, dan pemformatan data.
- Saat data yang diekstraksi diubah dan dibersihkan di area pementasan data, Anda perlu menyimpan seluruh konten data dan memastikan tidak ada yang hilang dalam perjalanan. Anda mungkin ingin memikirkan beberapa file atau tabel pelacakan untuk memuat entri pelacakan. Lingkungan terpisah paling kondusif untuk mengelola pergerakan data.
- Kami membicarakan kemungkinan memiliki alat khusus untuk memanipulasi data di staging area. Jika Anda memiliki lingkungan komputasi terpisah untuk area pementasan, Anda dapat dengan mudah meminta orang yang dilatih khusus tentang alat ini untuk menjalankan peralatan komputasi terpisah.



Gambar 8.5 Opsi platform untuk area pementasan.

Pertimbangan Pergerakan Data Pada platform komputasi mana pun langkah-langkah akuisisi data dan penyimpanan data terjadi, data harus berpindah antar platform. Bergantung pada platform sumber di perusahaan Anda dan pilihan platform untuk pementasan data dan penyimpanan data, Anda harus menyediakan transportasi data di berbagai platform.

Tinjau opsi berikut. Gambar 8-6 merangkum opsi standar. Anda mungkin menyadari bahwa satu pendekatan saja tidak cukup. Jangan ragu untuk memiliki kombinasi yang seimbang dari berbagai pendekatan yang berbeda. Dalam setiap pergerakan data di dua platform komputasi, pilih opsi yang paling sesuai untuk lingkungan tersebut. Penjelasan singkat tentang opsi standar berikut ini.

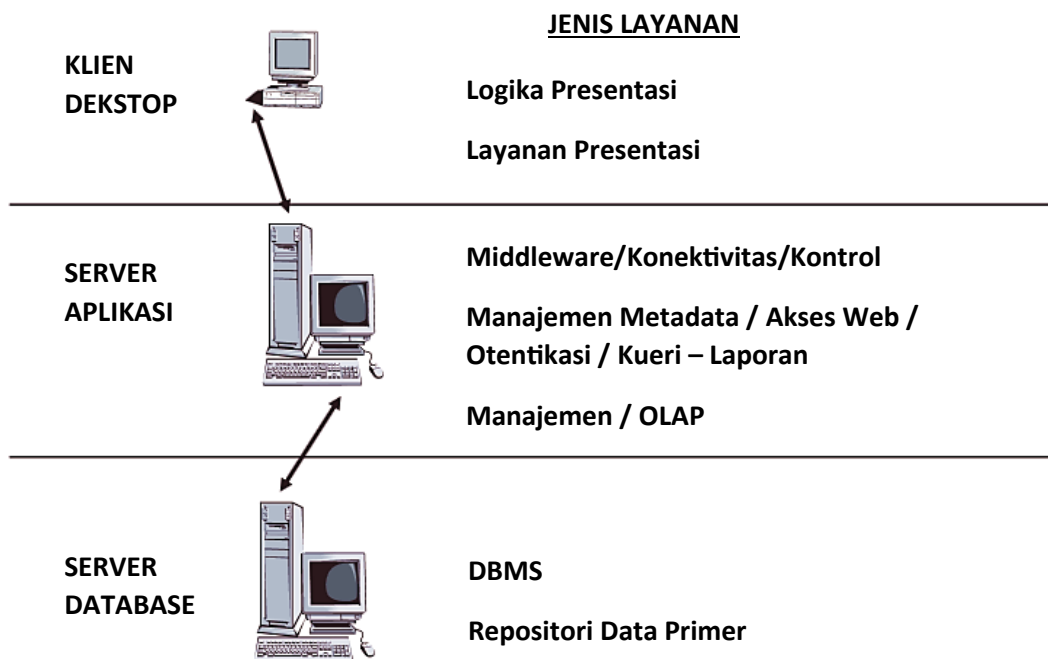


Gambar 8.6 Opsi perpindahan data.

Disk Bersama Metode ini kembali ke masa mainframe. Aplikasi yang berjalan di partisi atau wilayah berbeda diizinkan untuk berbagi data dengan menempatkan data umum pada disk bersama. Anda dapat mengadaptasi metode ini untuk meneruskan data dari satu langkah ke langkah lainnya untuk akuisisi data di gudang data Anda. Anda harus menentukan tempat penyimpanan disk dan mengaturnya sehingga masing-masing platform mengenali tempat penyimpanan disk sebagai miliknya.

Transmisi Data Massal Dalam hal ini, transmisi data lintas platform terjadi melalui port data. Port data hanyalah perangkat antar platform yang memungkinkan sejumlah besar data diangkut dari satu platform ke platform lainnya. Setiap platform harus dikonfigurasi untuk menangani transfer melalui port. Opsi ini memerlukan perangkat keras, perangkat lunak, dan komponen jaringan khusus. Bandwidth jaringan juga harus cukup untuk membawa volume data yang tinggi.

Koneksi Real-Time Dalam opsi ini, dua platform membuat koneksi secara real-time sehingga program yang berjalan pada satu platform dapat menggunakan sumber daya dari platform lainnya. Sebuah program di satu platform dapat menulis ke penyimpanan disk di platform lain. Selain itu, pekerjaan yang berjalan di satu platform dapat menjadwalkan pekerjaan dan acara di platform lain. Dengan meluasnya adopsi TCP/IP, opsi ini sangat cocok untuk gudang data Anda. Jika gudang data Anda dimaksudkan untuk menjadi gudang data waktu nyata atau mendekati waktu nyata, maka ini merupakan pertimbangan penting.



Gambar 8.7 Arsitektur klien/server untuk gudang data.

Metode Manual Mungkin ini adalah pilihan terakhir. Meskipun demikian, pilihan ini mudah dan sederhana. Sebuah program pada satu platform menulis ke media eksternal seperti tape atau disk. Program lain pada platform penerima membaca data dari media eksternal. Arsitektur Klien/Server untuk Gudang Data Meskipun platform mainframe dan

komputer mini digunakan pada awal implementasi gudang data, pada umumnya gudang saat ini dibangun menggunakan arsitektur klien/server. Sebagian besar di antaranya adalah arsitektur klien/server generasi kedua yang bertingkat. Gambar 8-7 menunjukkan arsitektur klien/server tipikal untuk implementasi data warehouse.

DBMS gudang data dijalankan pada komponen server data. Repositori data gudang data berada di mesin ini. Komponen server ini merupakan komponen utama dan dibahas secara rinci pada bagian selanjutnya.

Seiring dengan berkembangnya teknologi data warehousing secara substansial, Anda kini akan melihat proliferasi komponen server aplikasi di tingkat menengah. Anda akan menemukan server aplikasi untuk sejumlah tujuan. Inilah yang penting:

- Untuk menjalankan middleware dan membangun konektivitas
- Untuk menjalankan perangkat lunak manajemen dan pengendalian
- Untuk menangani akses data dari Web
- Untuk mengelola metadata
- Untuk otentikasi
- Sebagai ujung depan
- Untuk mengelola dan menjalankan laporan standar
- Untuk manajemen kueri yang canggih
- Untuk aplikasi OLAP

Umumnya, stasiun kerja klien masih menangani logika presentasi dan menyediakan layanan presentasi. Mari kita bahas secara singkat pertimbangan penting untuk stasiun kerja klien.

Pertimbangan untuk Stasiun Kerja Klien Ketika Anda siap untuk mempertimbangkan konfigurasi mesin stasiun kerja, Anda akan segera menyadari bahwa Anda perlu melayani berbagai jenis pengguna. Kami hanya mempertimbangkan kebutuhan di workstation sehubungan dengan penyampaian informasi dari data warehouse. Pengguna biasa mungkin puas dengan mesin yang dapat menjalankan browser Web untuk mengakses laporan HTML. Sebaliknya, seorang analis yang serius membutuhkan mesin stasiun kerja yang lebih besar dan bertenaga. Tipe pengguna lain di antara kedua ekstrem ini memerlukan beragam layanan.

Apakah Anda kemudian membuat konfigurasi unik untuk setiap pengguna? Itu tidak praktis. Lebih baik menentukan konfigurasi minimum pada platform yang sesuai yang akan mendukung seperangkat alat pengiriman informasi standar di gudang data Anda. Terapkan konfigurasi ini untuk sebagian besar pengguna Anda. Di sana-sini, tambahkan beberapa fungsi lagi jika diperlukan. Untuk pengguna tingkat lanjut, pilih konfigurasi lain yang akan mendukung alat untuk analisis kompleks. Umumnya, konfigurasi untuk power user ini juga mendukung OLAP. Baru-baru ini, pertimbangan diberikan pada kemampuan untuk mendukung presentasi dashboard untuk pemantauan berkelanjutan.

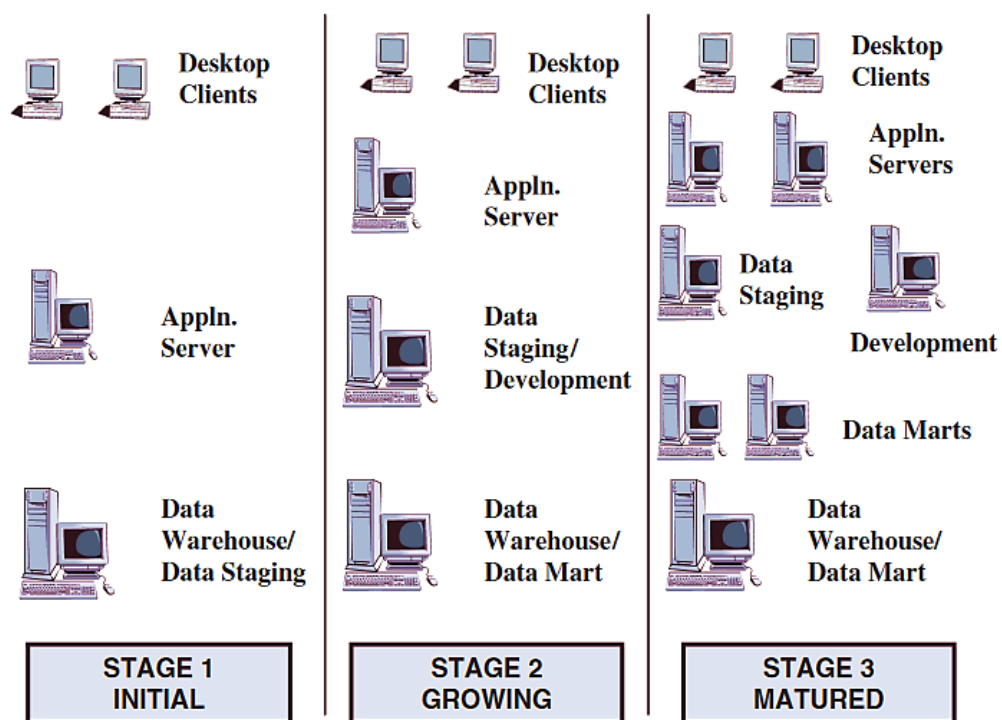
Faktor-faktor yang perlu dipertimbangkan ketika memilih konfigurasi untuk workstation pengguna Anda serupa dengan faktor-faktor yang ada di lingkungan operasi mana pun. Namun, pertimbangan utama bagi stasiun kerja yang mengakses gudang data adalah dukungan untuk seperangkat alat yang dipilih. Inilah alasan utama preferensi satu platform dibandingkan platform lainnya.

Gunakan daftar periksa ini saat mempertimbangkan stasiun kerja:

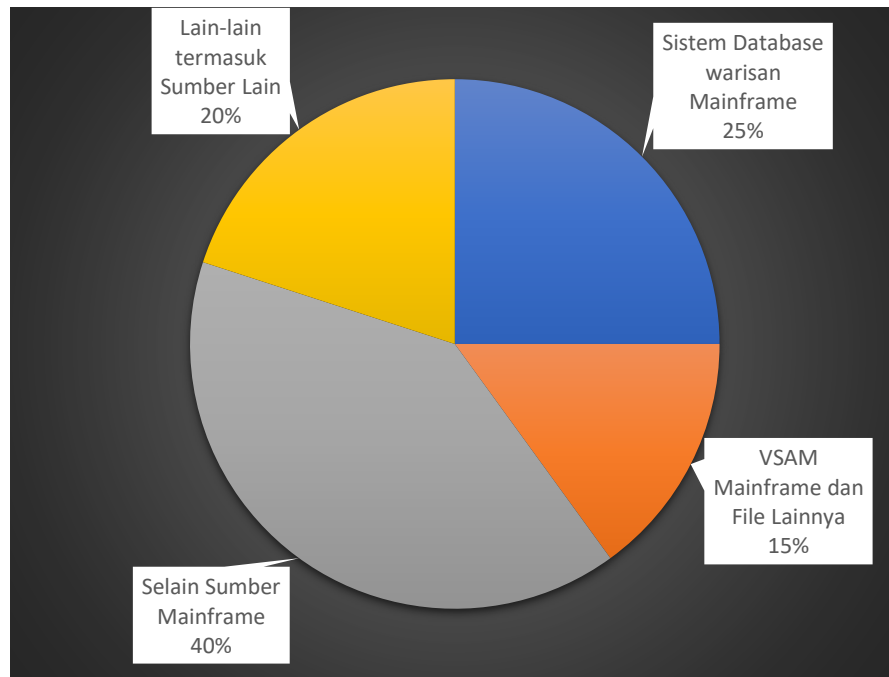
- ◆ Sistem operasi stasiun kerja
- ◆ Kekuatan pemrosesan
- ◆ Penyimpanan
- ◆ Penyimpanan disk
- ◆ Jaringan dan transportasi data
- ◆ Dukungan alat

Pilihan Saat Gudang Data Semakin Matang Setelah semua diskusi tentang platform komputasi untuk gudang data Anda, Anda mungkin mencapai kesimpulan bahwa pilihan platform sudah pasti segera setelah pilihan awal dibuat. Menarik untuk dicatat bahwa seiring dengan semakin matangnya gudang data di setiap perusahaan, susunan platformnya juga berkembang. Pementasan data dan penyimpanan data mungkin dimulai pada platform komputasi yang sama. Seiring berjalannya waktu dan semakin banyak pengguna Anda mulai bergantung pada gudang data Anda untuk pengambilan keputusan strategis, Anda akan menyadari bahwa pilihan platform mungkin harus diubah. Gambar 8.8 menunjukkan apa yang diharapkan seiring matangnya gudang data Anda.

Pilihan dalam Praktek Sebelum kita meninggalkan bagian ini, ada baiknya kita melihat jenis sumber data dan platform target yang digunakan di berbagai perusahaan. Sebuah survei independen telah menghasilkan beberapa temuan menarik. Gambar 8.9 menunjukkan perkiraan distribusi persentase untuk bagian pertama survei mengenai sumber data utama. Gambar 8-10 menunjukkan distribusi jawaban atas pertanyaan tentang platform yang digunakan responden untuk komponen penyimpanan data di gudang data mereka.



Gambar 8.8 Opsi platform saat gudang data sudah matang.



Gambar 8.9 Distribusi persentase untuk bagian pertama survei mengenai sumber data utama

Perangkat Keras Server

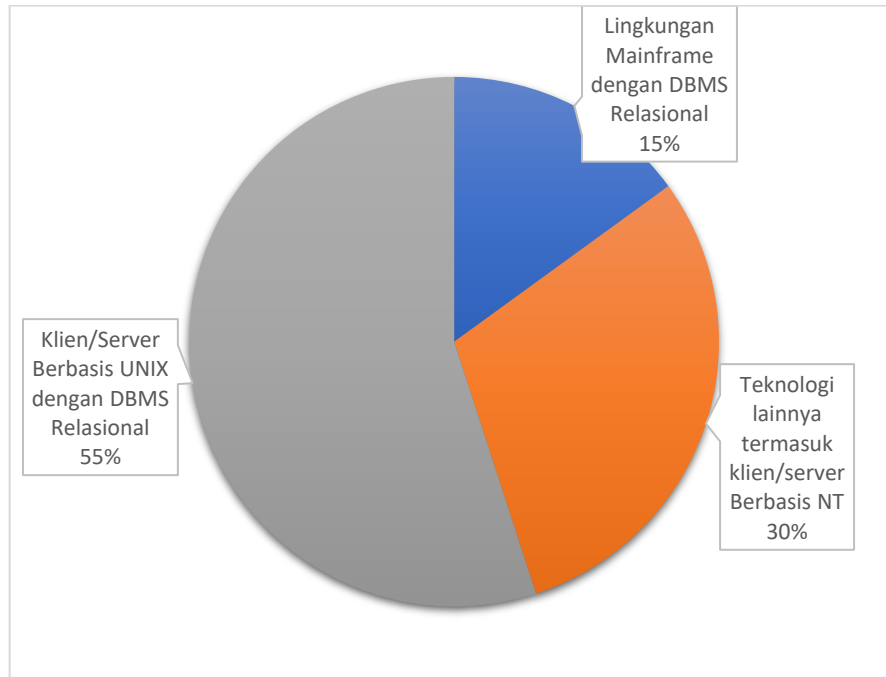
Memilih perangkat keras server adalah salah satu keputusan terpenting yang dihadapi tim proyek gudang data Anda. Mungkin, bagi sebagian besar gudang, pemilihan perangkat keras server bisa menjadi keputusan yang “bertaruh pada uang Anda”. Skalabilitas dan kinerja kueri yang optimal adalah frase kuncinya.

Anda tahu bahwa gudang data Anda ada untuk satu tujuan utama—untuk memberikan informasi kepada pengguna Anda. Kueri gudang data yang bersifat ad hoc, tidak dapat diprediksi, dan rumit adalah metode penyampaian informasi yang paling umum. Jika perangkat keras server Anda tidak mendukung pemrosesan kueri yang lebih cepat, keseluruhan proyek berada dalam bahaya.

Kebutuhan untuk melakukan penskalaan didorong oleh beberapa faktor. Seiring dengan semakin matangnya gudang data Anda, Anda akan melihat peningkatan tajam dalam jumlah pengguna dan jumlah kueri. Bebannya akan melonjak begitu saja. Biasanya, jumlah pengguna aktif meningkat dua kali lipat dalam enam bulan. Sekali lagi, seiring dengan semakin matangnya gudang data Anda, Anda akan meningkatkan konten dengan memasukkan lebih banyak bidang subjek bisnis dan menambahkan lebih banyak data mart. Gudang data perusahaan mulai dari sekitar 200 hingga 300 GB dan beberapa di antaranya meningkat hingga satu terabyte dalam waktu 18 hingga 24 bulan.

Opsi perangkat keras untuk skalabilitas dan pemrosesan kueri kompleks terdiri dari empat jenis arsitektur paralel. Awalnya, arsitektur paralel adalah yang paling masuk akal. Bukankah kueri akan selesai lebih cepat jika Anda menambah jumlah prosesor, masing-masing prosesor mengerjakan bagian kueri secara bersamaan? Tidak bisakah Anda membagi kueri

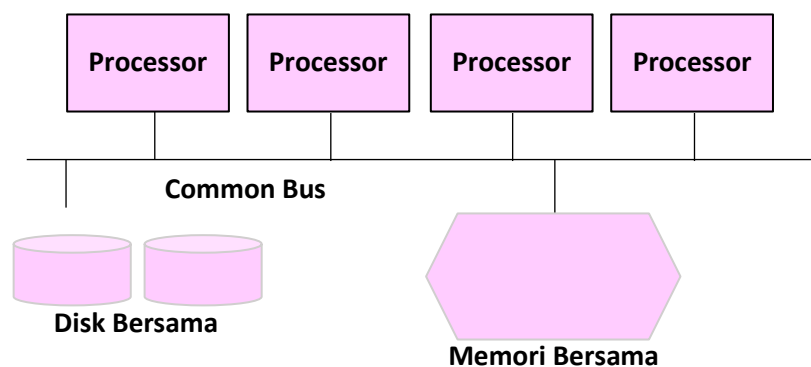
besar menjadi tugas-tugas terpisah dan menyebarkan tugas-tugas tersebut ke banyak prosesor? Pemrosesan paralel dengan beberapa mesin komputasi memang memberikan berbagai manfaat, namun tidak ada opsi pemrosesan tunggal yang dapat melakukan segalanya dengan benar.



Gambar 8.10 Platform target untuk komponen penyimpanan data.

Pada Bab 3, kami meninjau pemrosesan paralel sebagai salah satu tren signifikan dalam gudang data. Kami juga secara singkat melihat tiga arsitektur pemrosesan yang umum. Di bagian ini, mari kita rangkum opsi perangkat keras pemrosesan paralel saat ini. Anda akan mendapatkan wawasan yang cukup mengenai fitur, manfaat, dan keterbatasan masing-masing opsi ini. Dengan melakukan itu, Anda akan dapat menyumbangkan pemahaman Anda kepada tim proyek Anda untuk memilih perangkat keras server yang tepat.

SMP (*Symmetric Multiprocessing*) Lihat Gambar 8.11.



Gambar 8.11 Opsi perangkat keras server: SMP.

Fitur:

- ☞ Ini adalah arsitektur yang berbagi segalanya, mesin pemrosesan paralel yang paling sederhana.
- ☞ Setiap prosesor memiliki akses penuh ke memori bersama melalui bus umum.
- ☞ Komunikasi antar prosesor terjadi melalui memori bersama.
- ☞ Pengontrol disk dapat diakses oleh semua prosesor.

Manfaat:

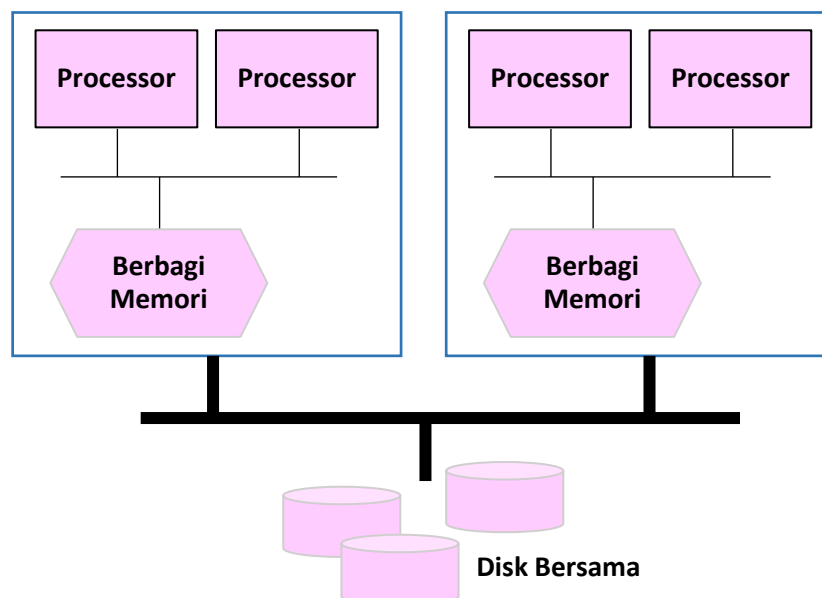
- ☞ Ini adalah teknologi yang telah terbukti digunakan sejak awal tahun 1970an.
- ☞ Ini memberikan konkurensi yang tinggi. Anda dapat menjalankan banyak kueri secara bersamaan.
- ☞ Ini menyeimbangkan beban kerja dengan sangat baik.
- ☞ Memberikan kinerja yang terukur; cukup tambahkan lebih banyak prosesor ke bus sistem.
- ☞ Dengan desain yang sederhana, Anda dapat mengelola server dengan mudah.

Keterbatasan:

- ☞ Memori yang tersedia mungkin terbatas.
- ☞ Kinerja mungkin dibatasi oleh bandwidth untuk komunikasi prosesor-ke-prosesor, I/O, dan komunikasi bus.
- ☞ Ketersediaannya terbatas; seperti satu komputer dengan banyak prosesor.

Anda dapat mempertimbangkan opsi ini jika ukuran gudang data Anda diperkirakan sekitar 200 atau 300 GB dan persyaratan konkurensi masuk akal.

Cluster Lihat Gambar 8.12.



Gambar 8.12 Opsi perangkat keras server: cluster.

Fitur:

- ☞ Setiap node terdiri dari satu atau lebih prosesor dan memori terkait.
- ☞ Memori tidak dibagi antar node; itu hanya dibagikan dalam setiap node.

- ☞ Komunikasi terjadi melalui bus berkecepatan tinggi.
- ☞ Setiap node memiliki akses ke kumpulan disk yang umum.
- ☞ Arsitektur ini adalah sekelompok node.

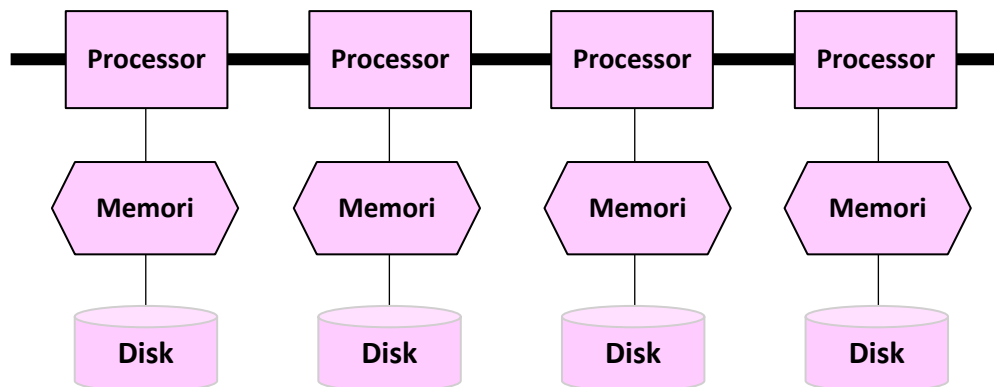
Manfaat:

- ☞ Arsitektur ini menyediakan ketersediaan tinggi; semua data dapat diakses meskipun satu node gagal.
- ☞ Ini mempertahankan konsep satu database.
- ☞ Opsi ini bagus untuk pertumbuhan tambahan.

Keterbatasan:

- ☞ Bandwidth bus dapat membatasi skalabilitas sistem.
- ☞ Opsi ini disertai dengan overhead sistem operasi yang tinggi.
- ☞ Setiap node memiliki cache data; arsitektur perlu menjaga konsistensi cache untuk sinkronisasi internode. Cache adalah “area kerja” yang menyimpan data yang sedang digunakan; memori utama seperti lemari arsip besar yang membentang di seluruh ruangan.

Anda dapat mempertimbangkan opsi ini jika gudang data Anda diharapkan tumbuh secara bertahap.



Gambar 8.13 Opsi perangkat keras server: MPP.

MPP (Massively Parallel Processing) Lihat Gambar 8.13.

Fitur:

- ☞ Ini adalah arsitektur yang tidak berbagi apa pun.
- ☞ Arsitektur ini lebih mementingkan akses disk dibandingkan akses memori.
- ☞ Ia bekerja dengan baik dengan sistem operasi yang mendukung akses disk transparan.
- ☞ Jika tabel database terletak pada disk tertentu, akses ke disk tersebut bergantung sepenuhnya pada prosesor yang memilikinya.
- ☞ Komunikasi antar node dilakukan melalui koneksi prosesor-ke-prosesor.

Manfaat:

- ☞ Arsitektur ini sangat scalable.
- ☞ Opsi ini menyediakan akses cepat antar node.

- ☞ Setiap kegagalan bersifat lokal pada node yang gagal; ini meningkatkan ketersediaan sistem.
- ☞ Secara umum, biaya per node rendah.

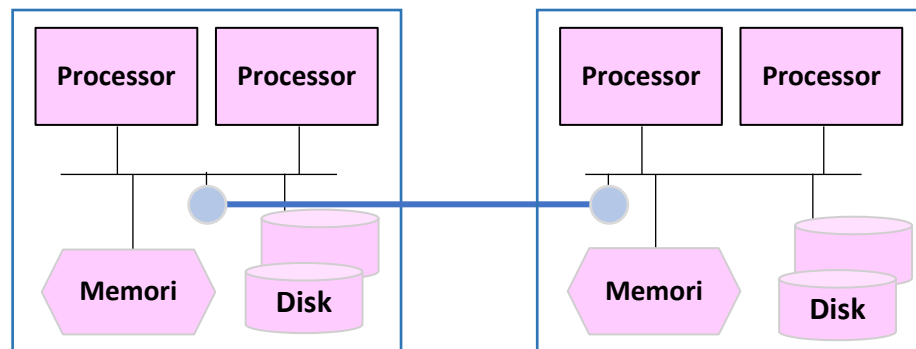
Keterbatasan:

- ☞ Arsitektur memerlukan partisi data yang kaku.
- ☞ Akses data dibatasi.
- ☞ Penyeimbangan beban kerja terbatas.
- ☞ Konsistensi cache harus dijaga.

Pertimbangkan opsi ini jika Anda membangun gudang data berukuran sedang atau besar dalam kisaran 400 hingga 500 GB. Untuk gudang yang lebih besar dalam kisaran terabyte, carilah kombinasi arsitektur khusus.

ccNUMA atau NUMA (Arsitektur Memori Tidak Seragam yang Koheren Cache)

Lihat Gambar 8.14.



Gambar 8.14 Opsi perangkat keras server: NUMA.

Fitur:

- ☞ Ini adalah arsitektur terbaru; itu dikembangkan pada tahun 1990an.
- ☞ Arsitektur NUMA seperti SMP besar yang dipecah menjadi SMP-SMP lebih kecil yang lebih mudah dibangun.
- ☞ Perangkat keras menganggap semua unit memori sebagai satu memori raksasa. Sistem ini memiliki satu ruang alamat memori nyata di seluruh mesin; alamat memori dimulai dengan 1 pada node pertama dan berlanjut pada node berikutnya. Setiap node berisi direktori alamat memori di dalam node tersebut.
- ☞ Dalam arsitektur ini, jumlah waktu yang dibutuhkan untuk mengambil nilai memori bervariasi karena node pertama mungkin memerlukan nilai yang berada di memori node ketiga. Itulah sebabnya arsitektur ini disebut arsitektur akses memori tidak seragam.

Manfaat:

- ☞ Memberikan fleksibilitas maksimum.
- ☞ Mengatasi keterbatasan memori SMP.
- ☞ Skalabilitas yang lebih baik dibandingkan SMP.

- ☞ Jika Anda perlu mempartisi database data warehouse Anda dan menjalankannya menggunakan pendekatan terpusat, Anda mungkin ingin mempertimbangkan arsitektur ini. Anda juga dapat menempatkan data OLAP Anda di server yang sama.

Keterbatasan:

- ☞ Pemrograman arsitektur NUMA lebih kompleks dibandingkan dengan MPP.
- ☞ Dukungan perangkat lunak untuk NUMA cukup terbatas.
- ☞ Teknologi ini masih berkembang.

Opsi ini merupakan pendekatan yang lebih agresif untuk Anda. Anda dapat memilih mesin NUMA yang terdiri dari satu atau dua node SMP, namun jika perusahaan Anda tidak berpengalaman dalam teknologi perangkat keras, opsi ini mungkin bukan untuk Anda.

8.3 PERANGKAT LUNAK DATABASE

Periksa fitur sistem manajemen basis data relasional komersial (RDBMS) terkemuka. Karena data warehousing telah menjadi hal yang lazim, Anda kini melihat fitur data warehouse disertakan dalam produk perangkat lunak. Hal itulah yang telah dilakukan oleh vendor database selama dekade terakhir. Add-on yang berhubungan dengan gudang data telah menjadi bagian dari penawaran database. Perangkat lunak database yang awalnya digunakan dalam sistem operasional OLTP telah ditingkatkan untuk memenuhi sistem pendukung keputusan. DBMS juga telah ditingkatkan untuk mendukung database yang sangat besar.

Beberapa produk RDBMS kini menyertakan dukungan untuk area akuisisi data di gudang data. Pemuatan massal dan pengambilan data dari sistem database lain menjadi lebih mudah. Beberapa vendor memberikan perhatian khusus pada fungsi transformasi data. Fitur replikasi telah diperkuat untuk membantu penyegaran massal dan pemuatan gudang data secara bertahap.

Indeks yang dipetakan bit bisa sangat efektif dalam lingkungan gudang data untuk mengindeks bidang yang memiliki jumlah nilai berbeda yang lebih kecil. Misalnya, dalam tabel database yang berisi wilayah geografis, jumlah kode wilayah yang berbeda sedikit. Namun seringkali, pertanyaan melibatkan pemilihan berdasarkan wilayah. Dalam hal ini, pengambilan dengan indeks yang dipetakan bit pada nilai kode wilayah bisa sangat cepat. Vendor telah memperkuat jenis pengindeksan ini. Kita akan membahas pengindeksan bit-mapped lebih lanjut pada Buku jilid 2 ini.

Selain penyempurnaan ini, hal yang lebih penting berkaitan dengan penyeimbangan beban dan kinerja kueri. Kedua fitur ini sangat penting dalam gudang data. Gudang data Anda berpusat pada kueri. Segala sesuatu yang dapat dilakukan untuk meningkatkan kinerja kueri adalah hal yang paling diinginkan. Vendor DBMS telah menyediakan fitur pemrosesan paralel untuk meningkatkan kinerja kueri. Mari kita tinjau secara singkat opsi pemrosesan paralel dalam DBMS yang dapat memanfaatkan sepenuhnya perangkat keras server paralel.

Opsi Pemrosesan Paralel

Opsi pemrosesan paralel dalam perangkat lunak basis data hanya ditujukan untuk mesin dengan banyak prosesor. Sebagian besar perangkat lunak database saat ini dapat memparalelkan sejumlah besar operasi. Operasi ini meliputi: pemuatan data secara massal,

pemindaian tabel penuh, kueri dengan kondisi pengecualian, kueri dengan pengelompokan, pemilihan dengan nilai berbeda, agregasi, pengurutan, pembuatan tabel menggunakan subkueri, pembuatan dan pembuatan ulang indeks, penyisipan baris ke dalam tabel dari tabel lainnya, batasan yang diaktifkan, transformasi bintang (teknik optimasi saat memproses kueri terhadap skema STAR), dan sebagainya. Perhatikan bahwa ini adalah daftar operasi mengesankan yang dapat diproses oleh RDBMS secara paralel.

Sekarang mari kita periksa apa yang terjadi ketika pengguna memulai query di workstation. Setiap sesi mengakses database melalui proses server. Kueri dikirim ke DBMS dan pengambilan data dilakukan dari database. Data diambil dan hasilnya dikirim kembali, semua di bawah kendali proses server khusus. Perangkat lunak operator kueri bertanggung jawab untuk membagi pekerjaan, mendistribusikan unit yang akan dilakukan di antara kumpulan proses server kueri yang tersedia, dan menyeimbangkan beban. Terakhir, hasil proses kueri dikumpulkan dan dikembalikan sebagai kumpulan hasil tunggal yang terkonsolidasi.

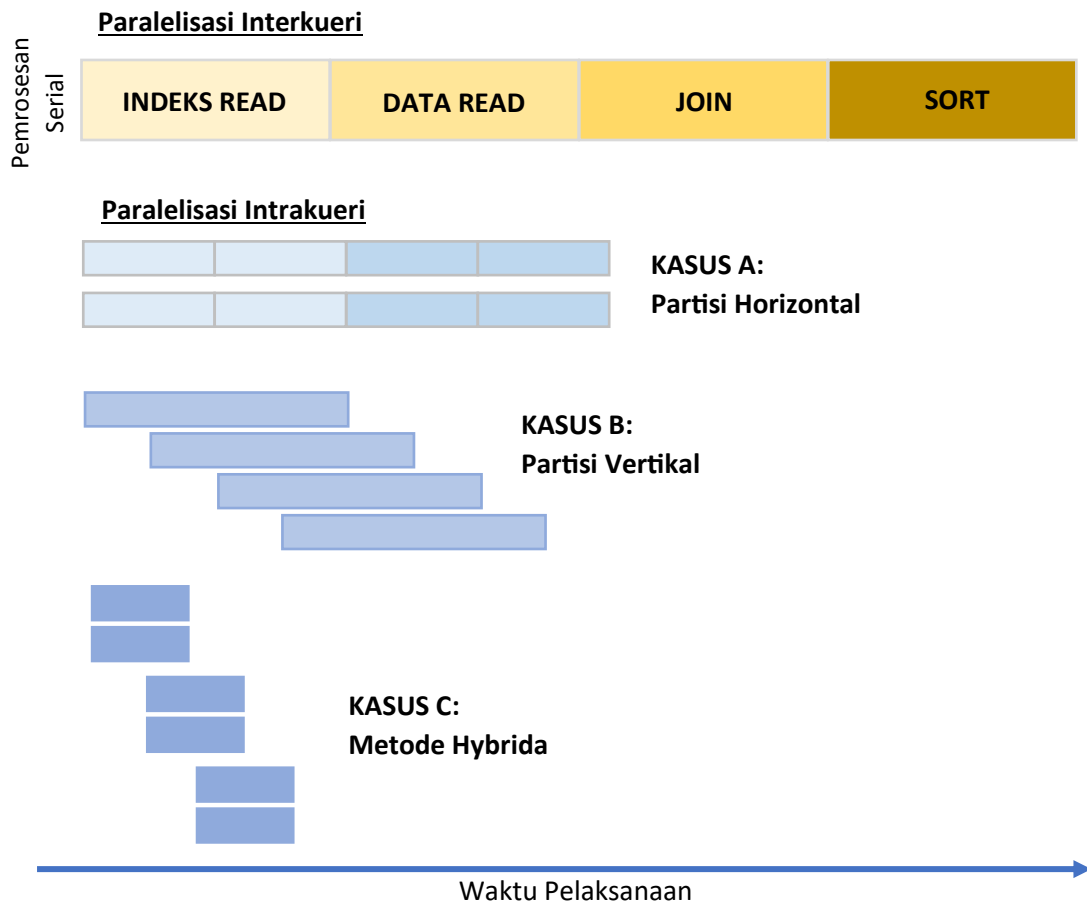
Paralelisasi Interkueri Dalam metode ini, beberapa proses server menangani beberapa permintaan secara bersamaan. Beberapa kueri mungkin dilayani berdasarkan konfigurasi server Anda dan jumlah prosesor yang tersedia. Anda mungkin berhasil memanfaatkan fitur DBMS pada sistem SMP ini, sehingga meningkatkan throughput dan mendukung lebih banyak pengguna secara bersamaan.

Namun, paralelisme interquery terbatas. Mari kita lihat apa yang terjadi di sini. Beberapa kueri diproses secara bersamaan, namun setiap kueri masih diproses secara serial oleh satu proses server. Misalkan kueri terdiri dari operasi pembacaan indeks, pembacaan data, penggabungan, dan pengurutan; operasi ini dilakukan dalam urutan ini. Setiap operasi harus selesai sebelum operasi berikutnya dapat dimulai. Bagian dari kueri yang sama tidak dijalankan secara paralel. Untuk mengatasi keterbatasan ini, banyak vendor DBMS telah membuat versi produk mereka untuk menyediakan paralelisasi intraquery.

Paralelisasi Intraquery Kita akan menggunakan Gambar 8.15 untuk pembahasan kita tentang paralelisasi intraquery. Memahami hal ini akan membantu Anda mencocokkan pilihan perangkat keras server dengan pilihan RDBMS Anda. Katakanlah kueri dari salah satu pengguna Anda terdiri dari pembacaan indeks, pembacaan data, penggabungan data, dan pengurutan data dari database gudang data. DBMS pemrosesan serial akan memproses kueri ini dalam urutan operasi dasar ini dan menghasilkan kumpulan hasil. Namun, saat kueri ini dijalankan pada satu prosesor di sistem SMP, kueri lain dapat dijalankan secara paralel. Metode ini adalah paralelisasi interquery yang dibahas di atas. Kelompok operasi pertama pada Gambar 8-15 mengilustrasikan metode eksekusi ini.

Dengan menggunakan teknik paralelisasi intraquery, DBMS membagi query menjadi operasi tingkat rendah yaitu pembacaan indeks, pembacaan data, penggabungan data, dan pengurutan data. Kemudian masing-masing operasi dasar ini dijalankan secara paralel pada satu prosesor. Kumpulan hasil akhir adalah konsolidasi hasil perantara. Mari kita tinjau tiga cara DBMS dapat menyediakan paralelisasi intraquery, yaitu paralelisasi bagian-bagian operasi dalam query yang sama itu sendiri.

Paralelisme Horizontal Data dipartisi di beberapa disk. Pemrosesan paralel terjadi dalam setiap tugas dalam kueri; misalnya, pembacaan data, yang dilakukan pada beberapa prosesor secara bersamaan pada kumpulan data berbeda untuk dibaca dari beberapa disk. Setelah tugas pertama diselesaikan dari semua bagian relevan dari data yang dipartisi, tugas berikutnya dari kueri tersebut dijalankan, lalu tugas berikutnya setelah tugas tersebut, dan seterusnya. Masalah dengan pendekatan ini adalah menunggu sampai semua data yang diperlukan dibaca. Lihatlah Kasus A pada Gambar 8.15.



Gambar 8.15 Paralelisasi query oleh DBMS.

Paralelisme Vertikal Jenis paralelisme ini terjadi di antara tugas-tugas yang berbeda, bukan hanya satu tugas dalam sebuah kueri seperti dalam kasus paralelisme horizontal. Semua operasi kueri komponen dijalankan secara paralel, namun secara pipeline. Hal ini mengasumsikan bahwa RDBMS memiliki kemampuan untuk menguraikan kueri menjadi subtugas; setiap subtugas memiliki semua operasi pembacaan indeks, pembacaan data, penggabungan, dan pengurutan. Kemudian setiap subtugas dijalankan pada data secara serial. Dalam pendekatan ini, catatan database idealnya diproses dalam satu langkah dan segera diberikan ke langkah berikutnya untuk diproses, sehingga menghindari waktu tunggu. Tentunya dalam metode ini DBMS harus memiliki tingkat kecanggihan yang sangat tinggi dalam menguraikan tugas. Sekarang, lihat Kasus B pada Gambar 8.15.

Metode Hibrid Dalam metode ini, pengurai kueri mempartisi kueri secara horizontal dan vertikal. Tentu saja, pendekatan ini memberikan hasil terbaik. Anda akan menyadari pemanfaatan sumber daya terbesar, kinerja optimal, dan skalabilitas tinggi. Kasus C pada Gambar 8.15 mengilustrasikan metode ini.

Pemilihan DBMS

Diskusi kami mengenai perangkat keras server dan opsi pemrosesan paralel DBMS pasti meyakinkan Anda bahwa pemilihan DBMS adalah hal yang paling penting. Anda harus memilih perangkat keras server dengan arsitektur paralel yang sesuai. DBMS pilihan Anda harus sesuai dengan perangkat keras server yang dipilih. Ini adalah keputusan penting untuk gudang data Anda.

Saat membahas bagaimana persyaratan bisnis mendorong desain dan pengembangan gudang data di Bab 6, kami secara singkat menyebutkan bagaimana persyaratan memengaruhi pemilihan DBMS. Terlepas dari kriteria bahwa DBMS yang dipilih harus memiliki opsi penyeimbangan beban dan pemrosesan paralel, kriteria utama lainnya adalah fitur-fitur yang tercantum di bawah ini harus dipertimbangkan ketika memilih DBMS untuk gudang data Anda.

Pengatur kueri, untuk mengantisipasi dan membatalkan kueri yang tidak berfungsi. Pengoptimal kueri, untuk mengurai dan mengoptimalkan kueri pengguna. Manajemen kueri, untuk menyeimbangkan eksekusi berbagai jenis kueri Memuat utilitas, untuk memuat, memulihkan, dan memulai ulang data berperforma tinggi. Manajemen metadata, dengan katalog atau kamus data yang aktif. Skalabilitas, dalam hal jumlah pengguna dan volume data. Ekstensibilitas, memiliki ekstensi hibrid ke database OLAP Portabilitas, lintas platform. Alat kueri Antarmuka Program Aplikasi (API), untuk alat dari vendor terkemuka Administrasi, menyediakan dukungan untuk semua fungsi DBA

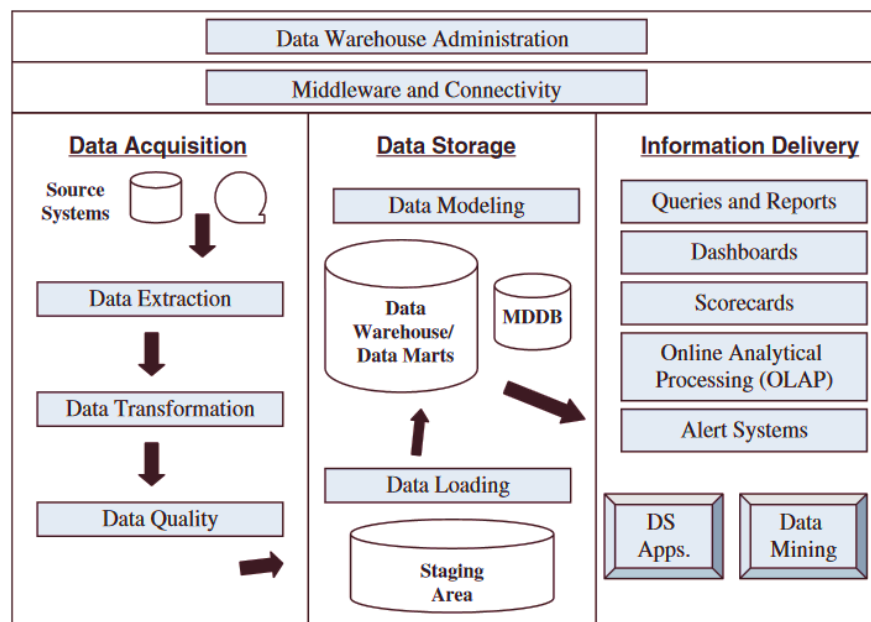
8.4 KOLEKSI ALAT

Pikirkan tentang aplikasi OLTP, mungkin sistem rekening giro di bank komersial. Ketika Anda, sebagai pengembang, merancang dan menyebarkan aplikasi, berapa banyak perangkat lunak pihak ketiga yang Anda gunakan untuk mengembangkan aplikasi tersebut? Tentu saja, jangan menghitung bahasa pemrograman atau perangkat lunak database. Yang kami maksud adalah alat vendor pihak ketiga lainnya untuk pemodelan data, perangkat lunak desain GUI, dan sebagainya. Anda mungkin hanya menggunakan sedikit, jika ada, sama sekali. Demikian pula, ketika teller bank menggunakan aplikasi tersebut, dia mungkin tidak menggunakan perangkat lunak pihak ketiga.

Namun lingkungan gudang data berbeda. Saat Anda, sebagai anggota tim proyek, mengembangkan gudang data, Anda akan menggunakan alat pihak ketiga untuk berbagai fase pengembangan. Anda dapat menggunakan generator kode untuk menyiapkan perangkat lunak internal untuk ekstraksi data. Saat gudang data diterapkan, pengguna Anda akan mengakses informasi melalui alat kueri pihak ketiga dan membuat laporan dengan penulis laporan. Alat perangkat lunak adalah bagian yang sangat penting dari infrastruktur di lingkungan gudang data.

Alat perangkat lunak tersedia untuk setiap komponen arsitektur gudang data. Seiring dengan semakin matangnya data warehousing selama dekade terakhir, vendor telah menghasilkan alat yang lebih canggih dan berguna untuk semua aspek. Gambar 8.16 menunjukkan kelompok alat yang mendukung berbagai fungsi dan layanan di gudang data.

Alat perangkat lunak sangat penting dalam gudang data. Seperti yang Anda lihat dari gambar ini, alat mencakup semua fungsi utama. Tim proyek gudang data hanya menulis sebagian kecil dari perangkat lunak yang diperlukan untuk menjalankan fungsi ini. Karena alat gudang data sangat penting, kita akan membahasnya lagi di bab selanjutnya: alat ekstraksi dan transformasi data, kualitas data dan alat kueri pada buku Jilid 2. Ketika Anda sampai pada tahap memilih alat untuk proyek gudang data Anda, daftar tersebut dapat berfungsi sebagai referensi praktis.



Gambar 8.16 Alat untuk gudang data Anda.

Pada tahap ini, mari kita perkenalkan jenis perangkat lunak yang umumnya diperlukan dalam lingkungan data warehouse. Untuk masing-masing jenisnya akan kita bahas secara singkat mengenai tujuan dan fungsinya.

Sebelum kita membahas jenis-jenis perangkat lunak, mari kita tegaskan kembali pepatah penting yang telah disebutkan sebelumnya di bab sebelumnya. Dalam bab itu kita membahas komponen arsitektural dan mempelajari fungsi serta layanan masing-masing komponen. Sekarang lanjutkan ke subbagian berikutnya di sini dan baca lagi tentang prinsip penting itu.

Arsitektur Pertama, Lalu Peralatan

Judul sub-bagian ini berarti: abaikan alatnya; rancang arsitekturnya terlebih dahulu; kemudian, dan hanya setelah itu, pilih alat yang sesuai dengan fungsi dan layanan yang ditetapkan untuk komponen arsitektur. Kerjakan arsitekturnya terlebih dahulu; pilih alatnya nanti.

Mengapa prinsip ini sakral? Mengapa tidak disarankan untuk hanya membeli seperangkat alat lalu menggunakan alat tersebut untuk membangun dan menerapkan gudang data Anda? Tampaknya ini merupakan solusi yang mudah. Tenaga penjual dari vendor alat menjanjikan kesuksesan. Mengapa hal ini pada akhirnya tidak berhasil? Mari kita ambil contoh.

Mari kita mulai merancang komponen arsitektur penyampaian informasi Anda. Pertama-tama, kebutuhan bisnis adalah kekuatan pendorongnya. Kelompok pengguna terbesar Anda adalah kelompok pengguna tingkat lanjut. Mereka akan membuat laporan sendiri dan menjalankan kueri mereka sendiri. Para pengguna ini akan terus-menerus melakukan analisis kompleks yang terdiri dari menelusuri, mengiris dan memotong data, dan visualisasi ekstensif dari rangkaian hasil. Anda tahu bahwa pengguna ini adalah pengguna yang mahir. Mereka membutuhkan komponen penyampaian informasi yang paling canggih. Fungsi dan layanan komponen penyampaian informasi harus sangat terlibat dan kuat. Namun Anda belum menetapkan komponen arsitektur penyampaian informasi.

Tahan di sana. Katakanlah sekarang staf penjualan dari XYZ Report Writer, Inc. telah meyakinkan Anda bahwa alat pembuat laporan mereka adalah semua yang Anda perlukan untuk penyampaian informasi di gudang data Anda. Dua pesaing Anda menggunakannya di gudang data mereka. Anda membeli alat tersebut dan siap menginstalnya. Bagaimana nasib pengguna listrik Anda? Apa yang salah dengan skenario ini? Alat penyampaian informasi dipilih sebelum komponen arsitektur ditetapkan. Alat tersebut tidak memenuhi persyaratan karena hal ini akan tercermin dalam arsitektur.

Sekarang mari kita beralih ke meninjau jenis perangkat lunak untuk gudang data Anda. Seperti disebutkan sebelumnya, rincian lebih lanjut akan ditambahkan di bab-bab selanjutnya. Bab-bab ini juga akan menguraikan masing-masing jenis alat. Pada subbagian berikut, kami menyebutkan tujuan dasar dan fitur jenis alat yang ditunjukkan oleh judul setiap subbagian.

Pemodelan Data

- ☞ Memungkinkan pengembang untuk membuat dan memelihara model data untuk sistem sumber dan database target gudang data. Jika perlu, model data dapat dibuat untuk staging area.
- ☞ Memberikan kemampuan rekayasa ke depan untuk menghasilkan skema database.
- ☞ Menyediakan kemampuan rekayasa balik untuk menghasilkan model data dari entri kamus data database sumber yang ada.
- ☞ Memberikan kemampuan pemodelan dimensi kepada perancang data untuk membuat skema STAR.

Ekstraksi Data

- ☞ Tersedia dua metode ekstraksi utama: ekstraksi massal untuk penyegaran penuh dan replikasi berbasis perubahan untuk pemuatan tambahan.
- ☞ Pilihan alat bergantung pada faktor-faktor berikut: platform dan basis data sistem sumber, serta fasilitas ekstraksi dan duplikasi bawaan yang tersedia dalam sistem sumber.

Transformasi Data

- ☞ Mengubah data yang diekstraksi ke dalam format dan struktur data yang sesuai.
- ☞ Berikan nilai default seperti yang ditentukan.
- ☞ Fitur utama mencakup pemisahan lapangan, konsolidasi, standardisasi, dan deduplikasi.

Memuat Data

- ☞ Memuat data yang diubah dan dikonsolidasikan dalam bentuk memuat gambar ke dalam gudang gudang data.
- ☞ Beberapa loader menghasilkan kunci utama untuk tabel yang dimuat.
- ☞ Untuk memuat gambar yang tersedia pada mesin RDBMS yang sama dengan gudang data, prosedur yang telah dikodekan sebelumnya dan disimpan dalam database itu sendiri dapat digunakan untuk memuat.

Kualitas data

- ☞ Membantu dalam mencari dan memperbaiki kesalahan data.
- ☞ Dapat digunakan pada data di staging area atau pada sistem sumber secara langsung.
- ☞ Membantu mengatasi ketidakkonsistenan data dalam memuat gambar.

Pertanyaan dan Laporan

- ☞ Memungkinkan pengguna untuk menghasilkan laporan yang terekam, intensif grafis, dan canggih.
- ☞ Membantu pengguna merumuskan dan menjalankan kueri.
- ☞ Dua klasifikasi utama adalah penulis laporan dan server laporan.

Dasbor

- ☞ Memberikan informasi real-time atau mendekati real-time kepada pengguna secara interaktif.
- ☞ Kebanyakan dashboard real-time terhubung langsung ke sistem operasional.
- ☞ Memungkinkan beberapa kemampuan bagi pengguna seperti menelusuri, mengubah parameter dengan cepat, berbagai jenis tampilan, dan sebagainya.

Kartu Skor

- ☞ Memungkinkan pengguna memilih indikator kinerja utama dengan mudah untuk pelaporan.
- ☞ Memberikan perbandingan kinerja saat ini terhadap target dan kinerja masa lalu.
- ☞ Fokus pada kejelasan dan kemudahan penggunaan.

Pemrosesan Analitik Online (OLAP)

- ☞ Memungkinkan pengguna menjalankan kueri dimensi yang kompleks.
- ☞ Memungkinkan pengguna untuk menghasilkan kueri terekam.
- ☞ Dua kategori pemrosesan analitis online adalah pemrosesan analitis online multidimensi (MOLAP) dan pemrosesan analitis online relasional (ROLAP). MOLAP bekerja dengan database multidimensi berpemilik yang menerima umpan data dari gudang data utama. ROLAP menyediakan kemampuan pemrosesan analitis online dari database relasional gudang data itu sendiri.

Sistem Peringatan

- ☞ Soroti dan dapatkan perhatian pengguna berdasarkan pengecualian yang ditentukan.
- ☞ Memberikan peringatan dari database gudang data untuk mendukung keputusan strategis.
- ☞ Tiga jenis peringatan dasar adalah: dari sistem sumber individual, dari gudang data terpadu di seluruh perusahaan, dan dari data mart individual.

Middleware dan Konektivitas

- ☞ Akses transparan ke sistem sumber di lingkungan yang heterogen.
- ☞ Akses transparan ke berbagai jenis database di berbagai platform.
- ☞ Alat-alatnya cukup mahal namun terbukti sangat berharga dalam menyediakan interoperabilitas di antara berbagai komponen gudang data.

Administrasi Gudang Data

- ☞ Membantu administrator gudang data dalam manajemen sehari-hari.
- ☞ Beberapa alat berfokus pada proses pemuatan dan melacak riwayat pemuatan.
- ☞ Alat lain melacak jenis dan jumlah permintaan pengguna.

8.5 PERALATAN GUDANG DATA

Dalam Bab 3 saat membahas tren data warehousing, kami memperkenalkan tren peralatan data warehouse yang baru-baru ini lazim. Pada bab ini kita telah meninjau infrastruktur gudang data. Kita telah melihat bagaimana perangkat keras server dan sistem operasi serta sistem manajemen basis data yang menyertainya merupakan bagian penting dari infrastruktur. Perusahaan yang mendapatkan lebih banyak pengalaman dalam pergudangan data telah menyadari pentingnya opsi pemrosesan paralel untuk perangkat keras server dan perangkat lunak yang menyertainya untuk mendukung pemrosesan paralel. Seiring dengan bertambahnya ukuran dan kompleksitas gudang data mereka, perusahaan-perusahaan ini mengatasi tantangan tersebut dengan terus melakukan peningkatan pada perangkat keras dan perangkat lunak.

Baru-baru ini, dengan penekanan besar pada intelijen bisnis, metode-metode lama yang selalu melakukan peningkatan secara terus-menerus tampaknya tidak berhasil. Memutakhirkan perangkat keras dan perangkat lunak sedikit demi sedikit membuka masalah kompatibilitas. Hal ini karena perangkat keras dan perangkat lunak tidak dirancang dan disesuaikan secara khusus untuk gudang data dengan volume data yang besar dan mekanisme penyampaian intelijen bisnis yang kompleks. Akibatnya, organisasi mulai menggunakan apa yang dikenal sebagai peralatan gudang data.

Peralatan gudang data adalah perangkat dengan komponen perangkat keras dan perangkat lunak yang dirancang dan dirancang khusus untuk gudang data. Melalui arsitekturnya, peralatan gudang data mengintegrasikan perangkat keras prosesor, penyimpanan data, sistem operasi, dan perangkat lunak basis data menjadi satu kesatuan yang utuh. Skalabilitas secara keseluruhan dicapai dengan mudah karena hampir semua komponen berasal dari satu vendor.

Evolusi Peralatan DW

Dimulai dengan penerapan gudang data oleh organisasi, dalam banyak kasus infrastruktur intelijen bisnis masih berupa perangkat keras, perangkat lunak, dan penyimpanan yang digabungkan secara longgar. Biasanya apa yang kita perhatikan sebagai konfigurasi umum terdiri dari yang berikut:

- ◆ Perangkat keras dan sistem operasi, disatukan secara umum, dan terus dioptimalkan dan ditingkatkan
- ◆ Sistem manajemen basis data (DBMS), yang awalnya dirancang untuk pemrosesan transaksi, kini diubah dan diperluas untuk mendukung data dalam jumlah besar yang tidak terduga di gudang data
- ◆ Unit penyimpanan dan manajemen penyimpanan, terus berkembang untuk memenuhi kebutuhan gudang data

Satu-satunya jawaban terhadap meningkatnya permintaan akan intelijen bisnis tampaknya hanya dengan menambahkan perangkat keras, perangkat lunak, dan penyimpanan hingga mencapai titik puncaknya. Di sebagian besar perusahaan, situasinya semakin buruk.

Dalam konteks ini, peralatan gudang data dibuat khusus untuk menyederhanakan beban kerja intelijen bisnis dengan mengintegrasikan komponen perangkat keras, perangkat lunak, dan penyimpanan. Fitur terbaik dari pendekatan SMP dan MPP ditanamkan ke dalam arsitektur peralatan untuk menghasilkan hasil terbaik.

Secara umum, lini produk inti Teradata dianggap sebagai induk dari semua peralatan gudang data, meskipun istilah tersebut tidak dikaitkan dengan penawaran produk. Minat yang lebih besar terhadap peralatan gudang data melonjak dengan munculnya Netezza di awal tahun 2000an. Vendor ini berbuat banyak untuk membuktikan konsep dan mendidik organisasi di arena data warehouse. Istilah alat gudang data diciptakan oleh Foster Hinshaw, pendiri Netezza.

Pada tahun 2009, kami memasuki generasi kedua peralatan gudang data yang dihasilkan dari integrasi vendor utama. Vendor telah menciptakan peralatan DW sepenuhnya dengan produk mereka sendiri. Beberapa vendor lain telah bermitra dengan vendor perangkat keras besar untuk membuat bundel yang merupakan penawaran peralatan DW mereka.

Seiring dengan semakin matangnya pasar data warehousing dan intelijen bisnis, kami telah memperhatikan standarisasi antarmuka, protokol, fitur, dan fungsionalitas komponen. Peralatan gudang data, yang muncul dalam iklim standarisasi ini, dengan mudah dibuat agar berfungsi secara lancar dengan alat dan aplikasi.

Manfaat Peralatan DW

Ketika peralatan gudang data mendapatkan popularitas dan penggunaan, organisasi yang memperolehnya memperoleh sejumlah manfaat. Berikut ini menyoroti manfaat peralatan gudang data.

Pengurangan Biaya Total biaya yang terkait dengan gudang data terutama dipecah menjadi biaya pengaturan awal dan biaya pemeliharaan berkelanjutan. Peralatan gudang data mengeluarkan biaya awal yang lebih sedikit dibandingkan dengan opsi alternatif. Pemeliharaan berkelanjutan juga membutuhkan biaya lebih sedikit. Karena harga komponen

siap pakai yang kuat terus turun, gudang data yang dibangun dengan komponen tersebut akan mencerminkan penurunan harga. Dimanapun alat gudang data disatukan dengan komponen dari satu vendor, total biaya kepemilikan dikaitkan hanya dengan satu vendor, sehingga sangat mengurangi biaya dukungan.

Peningkatan Kinerja Rasio harga/kinerja yang menguntungkan menonjol sebagai manfaat besar dari peralatan gudang data. Sebagian besar peralatan mendukung beban kerja campuran, memungkinkan laporan, kueri, dan analisis kompleks dijalankan secara bersamaan dengan mudah. Vendor peralatan DW menerapkan berbagai metode partisi dan distribusi beban untuk mencapai pemrosesan paralel. Peningkatan kinerja membuka kemungkinan besar bagi pengguna. Laporan yang membutuhkan waktu berhari-hari untuk diselesaikan kini membutuhkan waktu berjam-jam; analisis yang memakan waktu berjam-jam kini hanya membutuhkan waktu beberapa menit; kueri yang berjalan selama beberapa menit sekarang akan memakan waktu beberapa detik.

Ketersediaan Tinggi Dengan membangun redundansi selektif, vendor peralatan DW mampu menyediakan ketersediaan tinggi. Banyak vendor menawarkan server siaga hangat, catu daya ganda, jaringan cadangan, pencerminan disk, dan solusi otomatis untuk kegagalan server.

Mengurangi Administrasi Peralatan DW adalah impian administrator. Sifat komponen yang terintegrasi dalam suatu peralatan berarti sangat mengurangi waktu untuk pemecahan masalah dan pemeliharaan. Karena sebagian besar peralatan DW adalah perangkat vendor tunggal, desain dan tanggung jawab pengoptimalan berada di tangan vendor tunggal. Administrator gudang data perlu menghabiskan lebih sedikit waktu pada tugas-tugas administratif. Peralatan DW mengurangi biaya administrasi melalui fitur-fitur seperti alokasi ruang otomatis, pengurangan pemeliharaan indeks, dan pengurangan penyesuaian kinerja.

Skalabilitas Kebanyakan peralatan DW menerapkan desain modular untuk setiap komponennya. Hal ini memudahkan peralatan untuk mengukur kapasitas dan kinerjanya. Misalnya, dalam arsitektur MPP, penambahan server akan meningkatkan kinerja dan kapasitas. Keandalan Sifat homogen dari alat DW, biasanya dengan komponen dari satu pemilik, berkontribusi terhadap keandalannya. Pengguna tidak perlu melakukan integrasi penyimpanan disk, sistem operasi, DBMS, dan prosesor yang sulit dan tidak pasti serta berusaha membuat semuanya bekerja sama. Peralatan DW adalah perangkat tunggal, terarsitektur, dan kohesif.

Implementasi Lebih Cepat Karena peralatan DW dapat diimplementasikan tanpa pengujian regresi atau integrasi, waktu implementasi keseluruhan menjadi sangat berkurang. Selain itu, pembuatan prototipe cepat dapat dilakukan dengan peralatan DW. Implementasi yang lebih cepat berarti kemampuan untuk memperoleh manfaat dari intelijen bisnis dalam siklus bisnis saat ini.

RINGKASAN BAB

- Infrastruktur berperan sebagai fondasi yang mendukung arsitektur data warehouse.
- Infrastruktur data warehouse terdiri dari infrastruktur operasional dan infrastruktur fisik.
- Perangkat keras dan sistem operasi membentuk lingkungan komputasi untuk gudang data.
- Terdapat beberapa opsi untuk platform komputasi yang diperlukan untuk mengimplementasikan berbagai komponen arsitektur.
- Memilih perangkat keras server adalah keputusan penting. Selalu, pilihannya adalah salah satu dari empat arsitektur server paralel.
- Opsi pemrosesan paralel sangat penting dalam DBMS. Produk perangkat lunak basis data saat ini mampu melakukan paralelisasi interkueri dan intrakueri.
- Alat perangkat lunak digunakan di gudang data untuk pemodelan data, ekstraksi data, transformasi data, pemuatan data, jaminan kualitas data, kueri dan laporan, dan pemrosesan analitis online (OLAP). Alat juga digunakan sebagai middleware, sistem peringatan, dan untuk administrasi gudang data.
- Peralatan gudang data adalah perangkat arsitek yang terdiri dari prosesor perangkat keras, penyimpanan, sistem operasi, dan DBMS terintegrasi, yang secara khusus dioptimalkan untuk pergudangan data. Biasanya, ini adalah solusi vendor tunggal, yang memberikan manfaat besar. Peralatan DW menjadi semakin populer.

PERTANYAAN SOAL

1. Bagaimana komposisi infrastruktur operasional data warehouse? Mengapa infrastruktur operasional sama pentingnya dengan infrastruktur fisik?
2. Sebutkan komponen utama infrastruktur fisik. Tulislah dua atau tiga kalimat untuk mendeskripsikan setiap komponen.
3. Jelaskan secara singkat enam kriteria yang akan Anda gunakan untuk memilih sistem operasi untuk gudang data Anda.
4. Apa saja pilihan platform untuk area pementasan? Bandingkan pilihannya dan sebutkan kelebihan dan kekurangannya.
5. Apa empat metode umum perpindahan data dalam data warehouse? Jelaskan kedua metode ini.
6. Apa yang dimaksud dengan peralatan gudang data? Buat daftar manfaatnya.
7. Apa saja empat pilihan perangkat keras server paralel? Buat daftar fitur, manfaat, dan batasan dari salah satu opsi ini.
8. Bagaimana vendor RDBMS meningkatkan produk mereka untuk pergudangan data? Jelaskan secara singkat dalam satu atau dua paragraf.
9. Apa yang dimaksud dengan paralelisasi intraquery oleh DBMS? Apa saja ketiga metode tersebut?
10. Sebutkan enam jenis perangkat lunak yang digunakan dalam gudang data. Pilih tiga jenis dari daftar Anda dan jelaskan fitur dan tujuannya.

BAB 9

PERAN METADATA YANG PENTING

TUJUAN BAB

- Cari tahu mengapa metadata sangat penting
- Memahami siapa yang membutuhkan metadata dan jenis apa yang mereka perlukan
- Tinjau jenis metadata berdasarkan tiga area fungsional
- Membahas metadata bisnis dan metadata teknis secara detail
- Periksa semua persyaratan yang harus dipenuhi metadata
- Memahami tantangan pengelolaan metadata
- Mempelajari pilihan untuk menyediakan metadata

Kita telah membahas metadata secara singkat di bab-bab sebelumnya. Di Bab 2, kita menganggap metadata sebagai salah satu blok bangunan utama untuk gudang data. Kami mengelompokkan metadata menjadi tiga jenis, yaitu operasional, ekstraksi dan transformasi, serta metadata pengguna akhir. Saat membahas tren utama data warehousing di Bab 3, kami meninjau inisiatif industri untuk menstandarisasi metadata.

Bab ini membahas topik metadata secara lebih mendalam. Kami akan berusaha menghilangkan ketidakpastian yang mungkin Anda miliki tentang arti sebenarnya, konten, dan karakteristik metadata. Kami juga akan mendapat apresiasi mengapa metadata sangat penting. Selanjutnya, kita akan mencari metode praktis untuk menyediakan metadata yang efektif dalam lingkungan data warehouse.

9.1 MENGAPA METADATA PENTING

Mari kita mulai dengan asumsi positif. Asumsikan tim proyek Anda telah berhasil menyelesaikan pengembangan data mart pertama. Semuanya dilakukan sesuai dengan jadwal. Manajemen Anda senang bahwa tim menyelesaikan proyek sesuai anggaran dan nyaman sebelum tenggat waktu. Semua hasil terbukti dalam pengujian komprehensif. Gudang data Anda siap digunakan. Ini adalah hari besarnya.

Salah satu pengguna terkemuka Anda sedang duduk di stasiun kerja siap untuk menulis dan menjalankan kueri pertama. Sebelum dia menyentuh keyboard, beberapa pertanyaan penting muncul di benaknya.

- ◆ Apakah ada pertanyaan standar yang dapat saya lihat?
- ◆ Apa sajakah berbagai elemen data di gudang?
- ◆ Apakah terdapat informasi mengenai penjualan unit dan biaya unit berdasarkan produk?
- ◆ Bagaimana cara menelusuri dan melihat apa yang tersedia?
- ◆ Dari mana mereka mendapatkan data gudang? Dari sistem sumber mana?
- ◆ Bagaimana mereka menggabungkan data dari sistem pemesanan melalui telepon dan sistem pemesanan lewat pos?
- ◆ Berapa umur data di gudang?

- ◆ Kapan terakhir kali data baru dimasukkan?
- ◆ Apakah ada ringkasan berdasarkan bulan dan produk?

Pertanyaan-pertanyaan ini dan beberapa pertanyaan serupa lainnya sangat valid dan relevan. Apa jawabannya? Dimana jawabannya? Bisakah pengguna Anda melihat jawabannya? Seberapa mudah bagi pengguna untuk mendapatkan jawabannya?

Metadata pada data warehouse berisi jawaban atas pertanyaan mengenai data yang ada pada data warehouse. Anda menyimpan jawabannya di tempat yang biasa disebut repositori metadata. Bahkan jika Anda bertanya kepada beberapa praktisi data warehousing atau jika Anda membaca hanya beberapa buku tentang data warehousing, Anda akan menerima definisi metadata yang tampaknya berbeda. Berikut adalah contoh daftar definisi:

- Data tentang data
- Daftar isi data
- Katalog untuk data
- Atlas gudang data
- Peta jalan gudang data
- Direktori gudang data
- Lem yang menyatukan isi data warehouse
- Penjepit untuk menangani data
- Pusat saraf

Jadi, apa sebenarnya metadata itu? Manakah dari definisi berikut yang paling mendekati kebenaran? Mari kita ambil contoh spesifik. Asumsikan pengguna Anda ingin mengetahui tentang tabel atau entitas bernama Pelanggan di gudang data Anda sebelum menjalankan kueri apa pun pada data pelanggan. Apa isi informasi tentang Pelanggan di repositori metadata Anda? Mari kita tinjau elemen metadata untuk entitas Pelanggan seperti yang ditunjukkan pada Gambar 9.1.

Nama kesatuan: Pelanggan	
Nama Alias: Akun, Klien	
Definisi	: Seseorang atau organisasi yang membeli barang atau jasa dari perusahaan.
Perkataan	: Entitas pelanggan mencakup pelanggan tetap, pelanggan saat ini, dan pelanggan lama.
Sistem	: Pesanan Barang Jadi, Kontrak Pemeliharaan, Penjualan Online.
Sumber	
Tanggal Pembuatan	: 15 Januari 2006
Tanggal Pembaruan Terakhir	: 21 Januari 2008
Siklus Pembaruan	: Mingguan
Tanggal Penyegaran Penuh Terakhir	: 29 Desember 2007
Siklus Penyegaran Penuh	: Setiap enam bulan
Kualitas Data Ditinjau	: 25 Januari 2008
Deduplikasi Terakhir	: 10 Januari 2008
Arsip yang Direncanakan	: Setiap enam bulan
Pengguna yang Bertanggung Jawab	: Jane Brown

Gambar 9.1 Elemen metadata untuk entitas Pelanggan.

Apa yang kamu lihat pada gambar tersebut? Elemen metadata menjelaskan entitas bernama Pelanggan yang berada di gudang data. Ini bukan sekedar deskripsi. Ini memberi tahu Anda lebih banyak. Ini memberikan lebih dari sekedar penjelasan semantik dan sintaksis. Metadata menggambarkan semua aspek terkait data di gudang data secara lengkap dan tepat. Berhubungan dengan siapa? Berhubungan terutama dengan pengguna dan juga dengan Anda sebagai pengembang dan bagian dari tim proyek.

Dalam bab ini, kita akan mengeksplorasi mengapa metadata memiliki peran yang sangat penting dalam gudang data. Kami akan mencari tahu mengapa dan bagaimana metadata sangat penting bagi pengguna dan pengembang. Tanpa metadata, gudang data Anda hanya akan menjadi sistem yang terputus-putus. Jika metadata sangat penting, bagaimana cara terbaik Anda menyediakannya? Kami akan mendiskusikan beberapa opsi yang tersedia dan memberikan beberapa saran yang valid.

Kebutuhan Penting dalam Gudang Data

Mari kita periksa dulu kebutuhan metadata dengan cara yang sedikit umum. Kami akan membahas lebih spesifik di bagian selanjutnya. Secara umum, metadata yang tepat mutlak diperlukan untuk menggunakan, membangun, dan mengelola gudang data Anda.

Untuk Menggunakan Gudang Data Ada satu perbedaan besar antara gudang data dan sistem operasional apa pun seperti aplikasi pemrosesan pesanan. Perbedaannya terletak pada penggunaan—akses informasi. Dalam aplikasi pemrosesan pesanan, bagaimana pengguna Anda mendapatkan informasi? Anda memberi mereka layar GUI dan laporan yang telah ditentukan sebelumnya. Mereka mendapatkan informasi tentang pending atau back order melalui layar yang relevan. Mereka mendapatkan informasi tentang total pesanan hari itu dari laporan harian tertentu. Anda membuat layar dan memformat laporan untuk pengguna. Tentu saja, ini dirancang berdasarkan spesifikasi dari pengguna. Meskipun demikian, pengguna sendiri tidak membuat format layar atau menyusun laporan setiap kali mereka membutuhkan informasi.

Sebaliknya, pengguna sendiri yang mengambil informasi dari gudang data. Pada umumnya, pengguna sendiri yang membuat kueri ad hoc dan menjalankannya di gudang data. Mereka memformat laporan mereka sendiri. Karena perbedaan besar ini, sebelum mereka dapat membuat dan menjalankan kueri, pengguna perlu mengetahui tentang data di gudang data. Mereka membutuhkan metadata. Namun, dalam sistem operasional kami, kami tidak mempunyai metode yang mudah dan fleksibel untuk mengetahui sifat isi database. Faktanya, tidak diperlukan antarmuka yang ramah pengguna terhadap konten database. Kamus atau katalog data dimaksudkan untuk penggunaan TI saja.

Situasi untuk gudang data sangat berbeda. Pengguna gudang data Anda perlu menerima nilai maksimal dari gudang data Anda. Mereka memerlukan metode canggih untuk menelusuri dan memeriksa isi gudang data. Mereka perlu mengetahui arti dari item data. Anda harus mencegah mereka menarik kesimpulan yang salah dari analisis mereka karena ketidaktahuan mereka tentang arti sebenarnya.

Implementasi data mart sebelumnya mungkin terbatas pada satu bidang studi. Sebagian besar, data mart tersebut digunakan oleh sekelompok kecil pengguna di satu

departemen. Pengguna data mart tersebut mampu bertahan dengan metadata yang sedikit. Gudang data saat ini memiliki cakupan yang lebih luas dan ukuran yang lebih besar. Tanpa dukungan metadata yang memadai, pengguna gudang data yang lebih besar ini akan mengalami hambatan total.

Untuk Membangun Gudang Data Katakanlah Anda adalah ahli ekstraksi dan transformasi data di tim proyek. Anda mengetahui metode ekstraksi data dengan sangat baik. Anda dapat bekerja dengan alat ekstraksi data. Anda memahami teknik transformasi data secara umum. Namun, untuk menerapkan keahlian Anda, pertama-tama Anda harus mengetahui sistem sumber dan struktur datanya. Anda perlu mengetahui struktur dan konten data di gudang data. Kemudian Anda perlu menentukan pemetaan dan transformasi datanya. Sejauh ini, untuk melakukan tugas Anda dalam membangun komponen ekstraksi data dan transformasi data di gudang data, Anda memerlukan metadata tentang sistem sumber, pemetaan sumber-ke-target, dan aturan transformasi data.

Cobalah untuk memakai topi yang berbeda. Anda sekarang adalah administrator basis data (DBA) untuk basis data gudang data. Anda bertanggung jawab atas desain fisik database dan melakukan pemuatan awal. Anda juga bertanggung jawab atas beban tambahan secara berkala. Ada lebih banyak tanggung jawab untuk Anda. Meskipun mengabaikan semua tanggung jawab lainnya untuk sesaat, untuk melakukan tugas desain fisik dan pemuatan saja, Anda memerlukan metadata tentang beberapa hal. Anda memerlukan tata letak di area pementasan. Anda memerlukan metadata tentang struktur logis database gudang data. Anda memerlukan metadata tentang siklus penyegaran dan pemuatan data. Ini hanyalah informasi minimum yang Anda perlukan.

Jika Anda mempertimbangkan setiap aktivitas dan setiap tugas untuk membangun gudang data, Anda akan menyadari bahwa metadata adalah kebutuhan yang mendesak secara keseluruhan dan merupakan komponen yang sangat signifikan dalam gudang data Anda. Metadata sangat penting untuk membangun gudang data Anda.

Ekstraksi/Transformasi/Pemuatan Data

Bagaimana cara menangani perubahan data?

Bagaimana cara memasukkan sumber baru?

Dimana membersihkan data? Bagaimana cara mengubah metode pembersihan data?

Bagaimana cara membersihkan data setelah mengisi gudang?

Bagaimana cara beralih ke teknik transformasi data baru?

Bagaimana cara mengaudit penerapan perubahan yang sedang berlangsung?

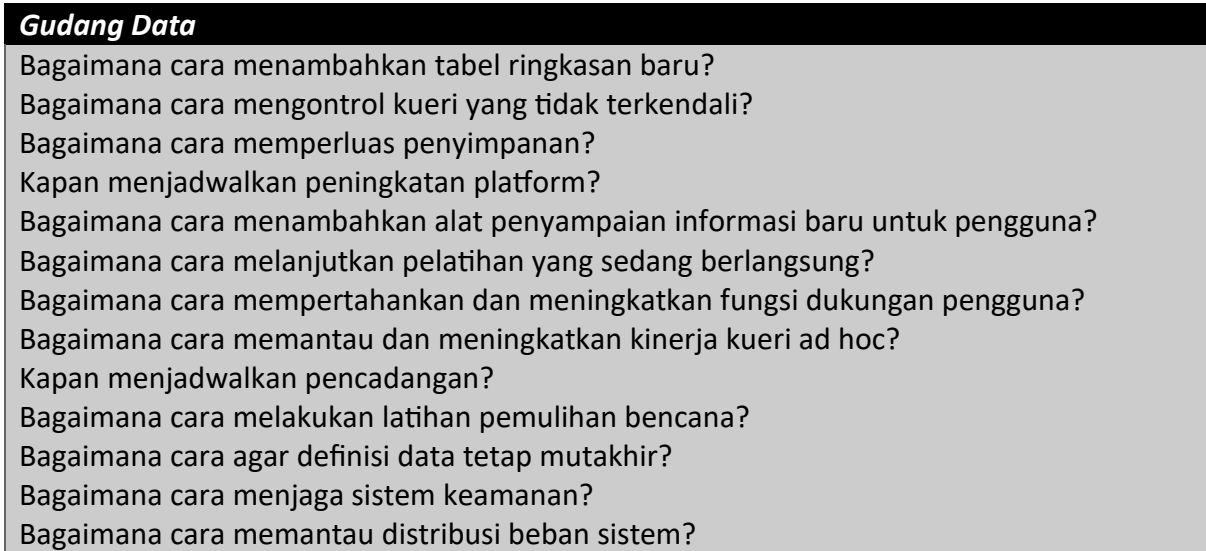
Data dari Sumber Eksternal

Bagaimana cara menambahkan sumber data eksternal baru?

Bagaimana cara menghilangkan beberapa sumber data eksternal?




Ketika merger dan akuisisi terjadi, bagaimana cara memasukkan data baru ke gudang?

Bagaimana cara memverifikasi semua data eksternal secara berkelanjutan?



Gambar 9.2 Administrasi gudang data: pertanyaan dan permasalahan.

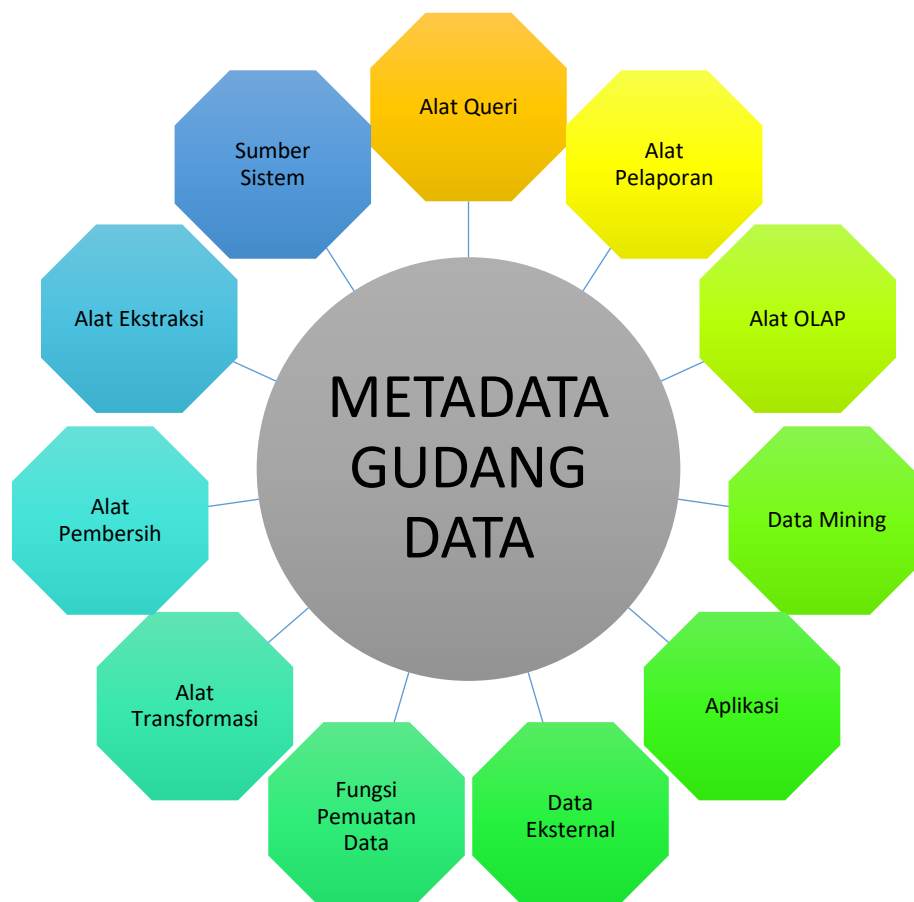
Untuk Mengelola Gudang Data Karena kompleksitas dan ukuran gudang data modern yang sangat besar, mustahil untuk mengelola gudang data tanpa metadata yang substansial. Gambar 9.2 mencantumkan serangkaian pertanyaan yang berkaitan dengan administrasi gudang data. Tinjau setiap pertanyaan dalam daftar dengan cermat. Anda tidak dapat mengelola gudang data Anda tanpa jawaban atas pertanyaan-pertanyaan ini. Metadata gudang data Anda harus mengatasi masalah ini. Gambar 9.3 memberi Anda gambaran tentang siapa yang membutuhkan dan menggunakan metadata. Kami akan menguraikan hal ini di bagian selanjutnya.

	 Profesional TI	 Pengguna Listrik	 Pengguna Biasa
Penemuan Informasi	Basis Data, Tabel, Kolom, Platform Server	Database, Tabel, Kolom	Daftar Kueri dan Laporan Standar, Tampilan Bisnis
Penambahan Data	Struktur Data, Definisi Data, Pemetaan Data, Fungsi Pembersihan, Aturan Transformasi.	Istilah Bisnis, Definisi Data, Pemetaan Data, Fungsi Pembersihan, Aturan Transformasi.	Istilah Bisnis, Definisi Data, Filter, Sumber Data, Konversi, Pemilik Data.
Akses Informasi	Kode Program dalam SQL, 3GL, 4GL, Aplikasi Front-end, Keamanan.	Perangkat Kueri, Akses Basis Data untuk Analisis Kompleks.	Permintaan Otorisasi, Pengambilan Informasi ke dalam Aplikasi Desktop seperti Spreadsheet

Gambar 9.3 Siapa yang memerlukan metadata?

Bayangkan sebuah lemari arsip yang penuh dengan dokumen tanpa folder dan label apa pun. Tanpa metadata, gudang data Anda seperti lemari arsip yang tidak terorganisir. Ini mungkin berisi informasi yang sangat berguna bagi pengguna Anda dan bagi pengembang dan administrator TI. Namun tanpa sarana yang mudah untuk mengetahui apa yang ada di sana, data warehouse hanya mempunyai nilai yang terbatas.

Metadata Seperti Pusat Saraf Berbagai proses selama pembangunan dan pengelolaan gudang data menghasilkan bagian dari metadata gudang data. Bagian dari metadata yang dihasilkan oleh satu proses digunakan oleh proses lainnya. Di gudang data, metadata mengambil posisi kunci dan memungkinkan komunikasi antar berbagai proses. Ini bertindak seperti pusat saraf di gudang data. Gambar 9.4 menunjukkan lokasi metadata dalam data warehouse. Gunakan gambar ini untuk menentukan komponen metadata yang berlaku pada lingkungan gudang data Anda. Dengan memeriksa setiap komponen metadata dengan cermat, Anda juga akan melihat bahwa masing-masing bagian metadata dibutuhkan oleh dua kelompok orang: (1) pengguna akhir, dan (2) pengembang dan administrator TI. Dalam dua subbagian berikutnya, kita akan meninjau mengapa metadata penting bagi kedua kelompok ini.



Gambar 9.4 Lokasi metadata dalam gudang data.

Mengapa Metadata Penting bagi Pengguna Akhir

Berikut ini adalah penggunaan umum gudang data Anda oleh pengguna utama, misalnya, seorang analis bisnis. Wakil Presiden Pemasaran perusahaan Anda telah meminta

analisis bisnis ini untuk melakukan analisis menyeluruh terhadap masalah yang baru-baru ini muncul. Karena potensi penjualan yang sangat besar di wilayah midwest dan timur laut, perusahaan Anda telah membuka lima toko baru di setiap wilayah. Meskipun penjualan secara keseluruhan di seluruh negeri meningkat dengan baik selama dua bulan setelah pembukaan toko, setelah itu penjualan kembali ke tingkat sebelumnya dan tetap datar. Wakil Presiden pemasaran ingin mengetahui alasannya, sehingga dia dapat mengambil tindakan yang tepat.

Sebagai pengguna, analisis bisnis berharap dapat menemukan jawaban dari data warehouse yang baru, namun ia tidak mengetahui secara detail tentang data yang ada di data warehouse tersebut. Secara khusus, dia tidak mengetahui jawaban atas pertanyaan-pertanyaan berikut:

- ◆ Apakah unit penjualan dan dolar disimpan berdasarkan transaksi individual atau sebagai ringkasan total, berdasarkan produk, untuk setiap hari di setiap toko?
- ◆ Bisakah penjualan dianalisis berdasarkan produk, promosi, toko, dan bulan?
- ◆ Dapatkah penjualan bulan ini dibandingkan dengan penjualan bulan yang sama tahun lalu?
- ◆ Apakah penjualan dapat dibandingkan dengan target?
- ◆ Bagaimana margin keuntungan dihitung? Apa aturan bisnisnya?
- ◆ Apa definisi wilayah penjualan? Kabupaten manakah yang termasuk dalam masing-masing wilayah yang dianalisis?
- ◆ Dari mana asal penjualannya? Dari sistem sumber mana?
- ◆ Berapa umur angka penjualannya? Seberapa sering angka-angka ini diperbarui?

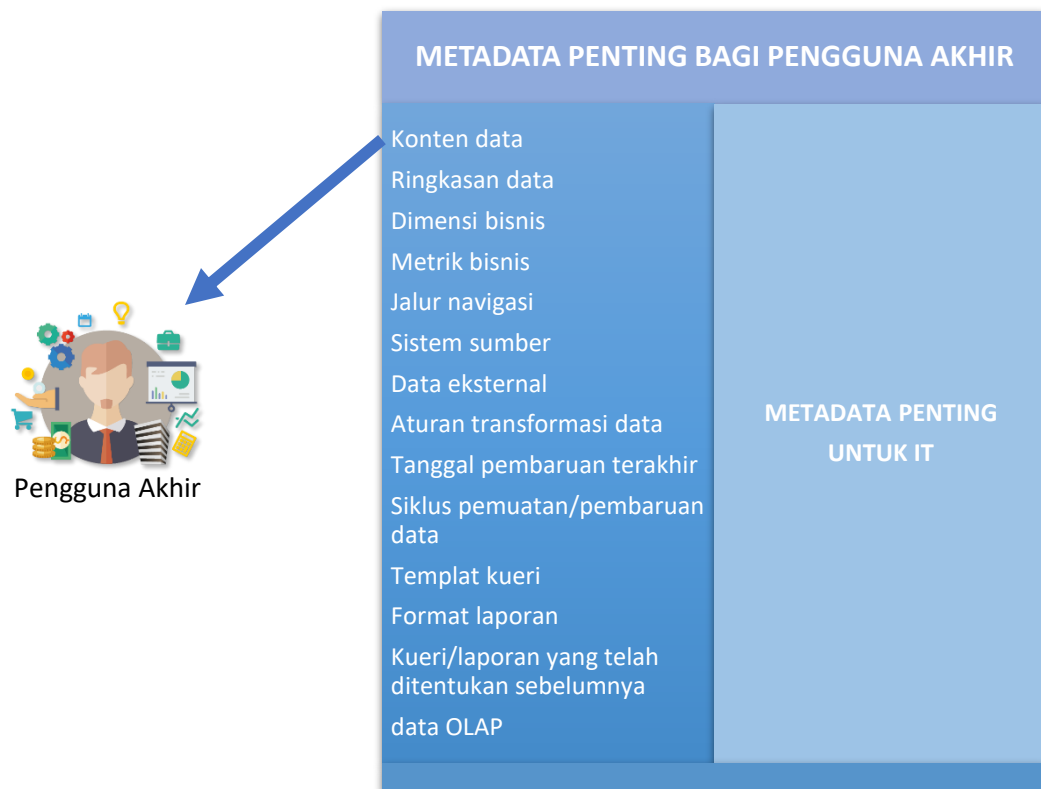
Jika analisis tidak yakin dengan sifat datanya, kemungkinan besar dia akan salah menafsirkan hasil analisisnya. Ada kemungkinan bahwa toko-toko baru tersebut melakukan kanibalisme penjualan dari toko-toko lain di wilayah mereka dan itulah sebabnya penjualan secara keseluruhan tetap datar. Namun analisis mungkin tidak menemukan alasan yang tepat karena kesalahan interpretasi terhadap hasil.

Analisis akan lebih efektif jika Anda menyediakan metadata yang memadai untuk membantu sebagai peta jalan data yang kuat. Jika terdapat metadata yang cukup dan tepat, maka analisis tidak perlu mendapatkan bantuan dari IT setiap kali perlu menjalankan analisis. Metadata yang mudah diakses sangat penting bagi pengguna akhir.

Mari kita analogikan gudang industri yang menyimpan barang dagangan yang dijual melalui katalog. Pelanggan mengacu pada katalog untuk mencari barang dagangan yang akan dipesan. Pelanggan menggunakan nomor item dalam katalog untuk melakukan pemesanan. Katalog juga menunjukkan warna, ukuran, dan bentuk barang dagangan. Pelanggan menghitung jumlah total yang harus dibayar dari rincian harga di katalog. Singkatnya, katalog mencakup semua barang di gudang industri, menjelaskan barang, dan memfasilitasi penempatan pesanan. Dengan cara yang sama, pengguna gudang data Anda seperti pelanggan. Permintaan informasi dari pengguna seperti pemesanan barang dagangan di gudang industri.

Sama seperti pelanggan memerlukan katalog untuk melakukan pemesanan, pengguna Anda juga memerlukan metadata untuk menjalankan kueri di gudang data Anda. Gambar 9.5

merangkum kebutuhan vital metadata bagi pengguna akhir. Gambar tersebut menunjukkan jenis informasi yang disediakan metadata kepada pengguna akhir dan tujuan mereka memerlukan jenis informasi tersebut.



Gambar 9.5 Metadata penting bagi pengguna akhir.

Mengapa Metadata Penting untuk TI

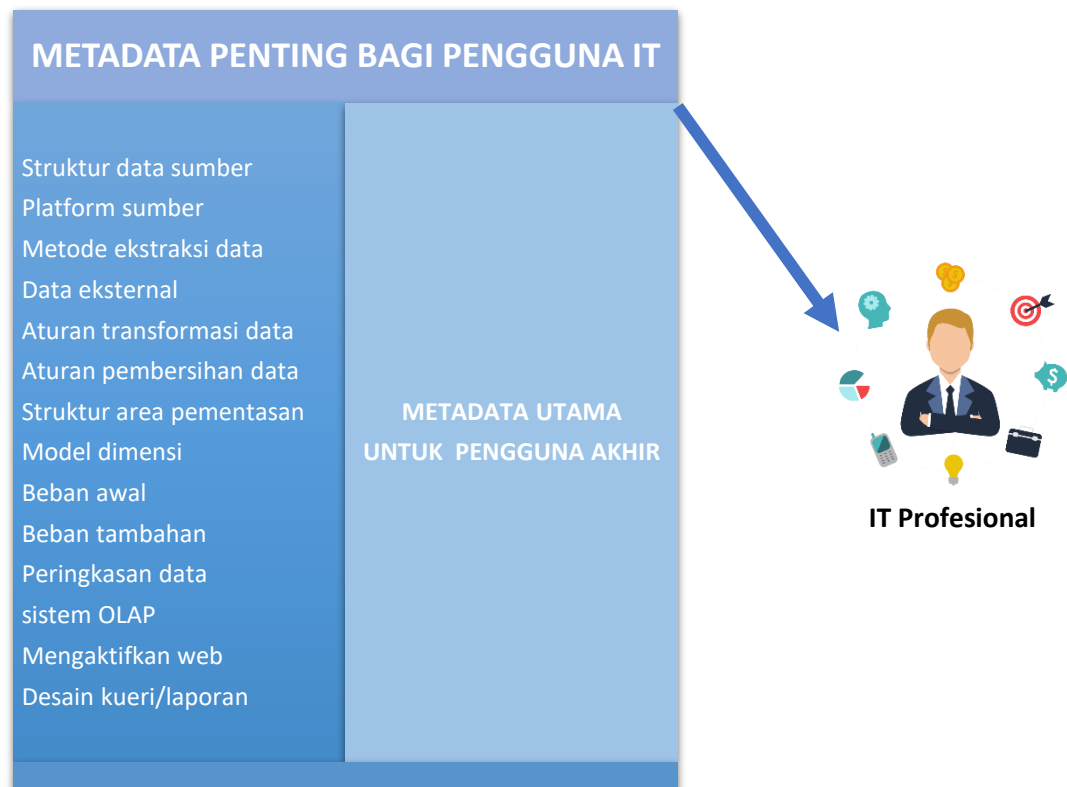
Pengembangan dan penerapan gudang data Anda merupakan upaya bersama antara staf TI dan perwakilan pengguna Anda. Namun demikian, karena masalah teknis, TI terutama bertanggung jawab atas desain dan administrasi gudang data yang berkelanjutan. Untuk melaksanakan tanggung jawab desain dan administrasi, TI harus memiliki akses ke metadata yang tepat. Sepanjang keseluruhan proses pengembangan, metadata sangat penting bagi TI. Dimulai dengan ekstraksi data dan diakhiri dengan penyampaian informasi, metadata sangat penting bagi TI. Ketika proses pengembangan bergerak melalui ekstraksi data, transformasi data, integrasi data, pembersihan data, pementasan data, penyimpanan data, desain kueri dan laporan, serta desain untuk OLAP dan sistem front-end lainnya, metadata sangat penting bagi TI untuk melakukan aktivitas pengembangannya.

Berikut adalah daftar ringkasan proses yang metadatanya penting bagi TI:

- ◆ Ekstraksi data dari sumber
- ◆ Transformasi data
- ◆ Pembersihan data
- ◆ Agregasi dan ringkasan data
- ◆ Pementasan data

- ◆ Penyegaran data
- ◆ Desain basis data
- ◆ Kueri dan desain laporan

Gambar 9.6 merangkum kebutuhan penting akan metadata untuk TI. Gambar tersebut menunjukkan jenis metadata informasi yang disediakan oleh staf TI dan tujuan mereka memerlukan jenis informasi tersebut.



Gambar 9.6 Metadata penting untuk TI.

Otomatisasi Tugas Pergudangan

Mempertahankan metadata bukan lagi suatu bentuk dokumentasi yang dimulihkan. Secara tradisional, metadata telah dibuat dan dipelihara sebagai dokumentasi tentang data untuk setiap proses. Sekarang metadata mengambil peran aktif baru. Mari kita lihat bagaimana hal ini terjadi.

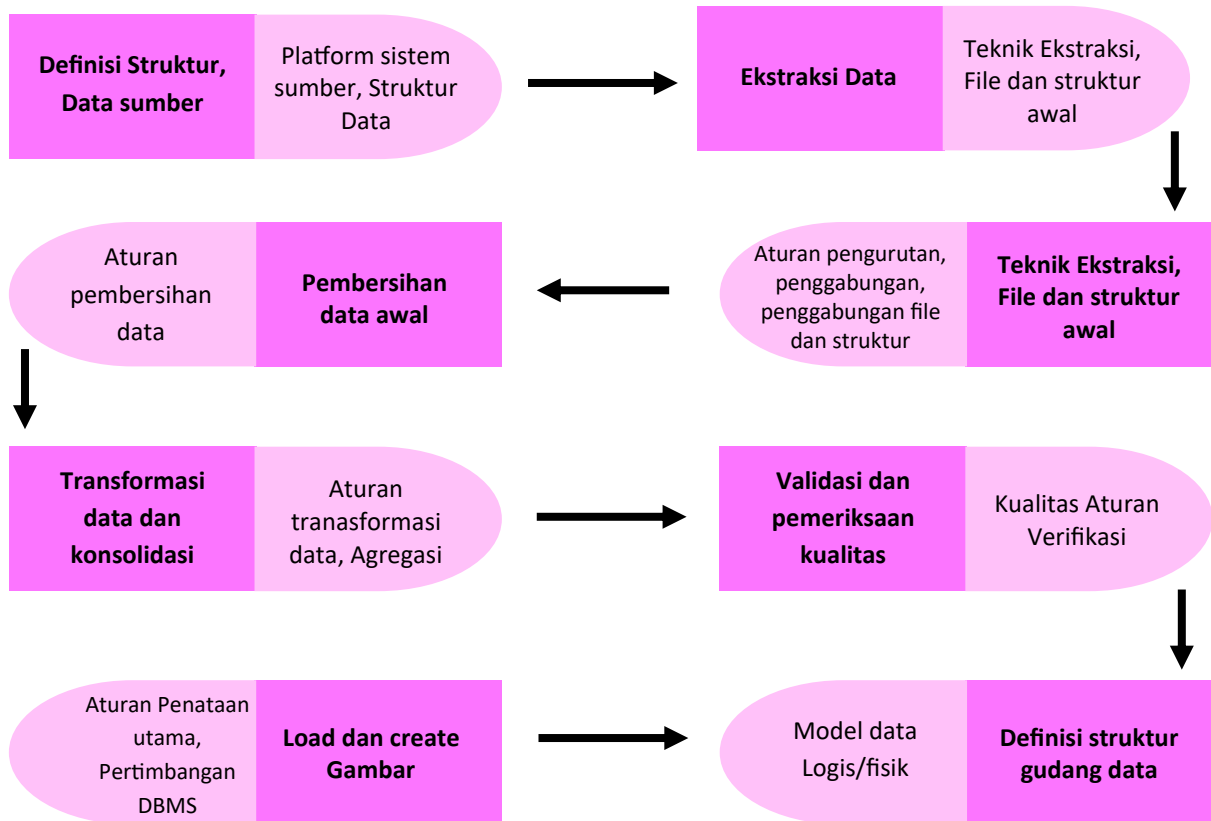
Seperti yang Anda ketahui, alat menjalankan fungsi utama dalam lingkungan gudang data. Misalnya, alat memungkinkan ekstraksi data dari sumber yang ditunjuk. Saat Anda menyediakan algoritme pemetaan, alat transformasi data mengubah elemen data agar sesuai dengan struktur data target. Anda dapat menentukan nilai yang valid untuk elemen data dan alat kualitas data akan menggunakan nilai ini untuk memastikan integritas dan validitas data. Di bagian depan, alat memberdayakan pengguna untuk menelusuri konten data dan mendapatkan akses ke gudang data. Alat-alat ini umumnya terbagi dalam dua kategori: alat pengembangan untuk profesional TI, dan alat akses informasi untuk pengguna akhir.

Saat Anda, sebagai pengembang, menggunakan alat untuk desain dan pengembangan, dalam proses tersebut, alat tersebut memungkinkan Anda membuat dan mencatat bagian metadata gudang data. Saat Anda menggunakan alat lain untuk melakukan proses lain dalam desain dan pengembangan, alat ini menggunakan metadata yang dibuat oleh alat pertama. Saat pengguna akhir Anda menggunakan alat kueri untuk akses informasi di ujung depan, alat kueri tersebut menggunakan metadata yang dibuat oleh beberapa alat ujung belakang. Apa sebenarnya yang terjadi dengan metadata? Metadata bukan lagi dokumentasi pasif. Metadata mengambil bagian dalam proses tersebut. Ini membantu dalam otomatisasi proses gudang data.

Mari kita pertimbangkan proses back-end yang dimulai dengan penentuan sumber data. Saat perpindahan data terjadi dari sumber data ke database gudang data melalui area pementasan data, beberapa proses terjadi. Dalam gudang data pada umumnya, alat yang sesuai membantu dalam proses ini. Setiap alat mencatat metadatanya sendiri saat perpindahan data terjadi. Metadata yang direkam oleh satu alat menggerakkan satu atau lebih proses berikutnya. Inilah cara metadata berperan aktif dan membantu otomatisasi proses gudang data.

Berikut adalah daftar proses back-end yang ditampilkan sesuai urutan kemunculannya secara umum:

1. Definisi struktur data sumber
2. Ekstraksi data
3. Pemformatan ulang/penggabungan awal
4. Pembersihan data awal
5. Transformasi dan konsolidasi data
6. Validasi dan pemeriksaan kualitas
7. Definisi struktur gudang data
8. Muat pembuatan gambar



Gambar 9.7 Metadata menggerakkan proses gudang data.

Gambar 9.7 menunjukkan masing-masing dari delapan proses ini. Gambar tersebut juga menunjukkan metadata yang direkam oleh setiap proses. Lebih lanjut, gambar tersebut menunjukkan bagaimana setiap proses dapat menggunakan metadata yang direkam dalam proses sebelumnya.

Metadata penting dalam gudang data karena menggerakkan proses. Namun, diskusi kami di atas mengarah pada kesadaran bahwa setiap alat dapat merekam metadata dalam format yang dimilikinya masing-masing. Sekali lagi, metadata yang direkam oleh masing-masing alat mungkin berada di platform tempat proses terkait dijalankan. Jika hal ini terjadi, bagaimana metadata yang direkam oleh satu alat dalam format kepemilikan dapat mendorong proses untuk alat berikutnya? Ini adalah pertanyaan kritis. Di sinilah standardisasi metadata berperan. Kita akan membahas standar metadata nanti di bab ini.

Menetapkan Konteks Informasi

Bayangkan skenario ini. Salah satu pengguna Anda ingin menjalankan kueri untuk mengambil data penjualan tiga produk selama tujuh hari pertama bulan April di wilayah selatan. Pengguna ini menulis kueri sebagai berikut:

Produk ¼ Widget-1 atau Widget-2 atau Widget-3, Wilayah ¼ Periode 'SELATAN' ¼ 01-04-2009 hingga 07-04-2009

Hasilnya kembali:

	Unit Terjual	Jumlah
Widget 1	25.355	253.550
Widget 2	16.978	254.670

Widget 3	7.994	271.796
-----------------	-------	---------

Mari kita periksa pertanyaan dan hasilnya. Dalam spesifikasi wilayah, wilayah manakah yang termasuk dalam wilayah “SELATAN”? Apakah ini wilayah yang diminati pengguna Anda? Apa konteks item data “SELATAN” di gudang data Anda? Selanjutnya, apakah data item 01-04-2009 menunjukkan 1 April 2009 atau 4 Januari 2009? Konvensi apa yang digunakan untuk tanggal di gudang data Anda?

Lihatlah hasil yang ditetapkan. Apakah angka yang ditampilkan sebagai unit penjualan diberikan dalam satuan fisik produk, atau dalam beberapa ukuran seperti pound atau kilogram? Bagaimana dengan jumlah yang ditunjukkan pada kumpulan hasil? Apakah jumlah ini dalam dolar atau euro? Ini adalah pertanyaan relevan jika pengguna Anda mengakses gudang data Anda dari Eropa.

Untuk tanggal yang disimpan di gudang data Anda, jika dua digit pertama dari format tanggal menunjukkan bulan dan dua digit berikutnya menunjukkan tanggal, maka 01-04-2009 berarti 1 April 2009. Hanya dalam konteks ini adalah interpretasi yang benar. Demikian pula, konteks juga penting untuk interpretasi elemen data lainnya.

Bagaimana pengguna Anda dapat mengetahui apa sebenarnya setiap elemen data dalam kueri dan apa arti dari kumpulan hasil? Jawabannya adalah metadata. Metadata memberi pengguna Anda arti dari setiap elemen data. Metadata menetapkan konteks untuk elemen data. Pengguna gudang data, pengembang, dan administrator menafsirkan setiap elemen data dalam konteks yang ditetapkan dan dicatat dalam metadata.

9.2 JENIS METADATA BERDASARKAN WILAYAH FUNGSIONAL

Sejauh ini dalam bab ini, kita telah membahas beberapa aspek metadata dalam lingkungan data warehouse. Kami telah melihat mengapa metadata merupakan kebutuhan penting bagi pengguna akhir serta profesional TI yang bertanggung jawab atas pengembangan dan administrasi. Kami telah menetapkan bahwa metadata memainkan peran aktif dalam otomatisasi proses gudang data.

Pada tahap ini, kita dapat meningkatkan pemahaman kita lebih lanjut dengan mengelompokkan berbagai jenis metadata. Saat Anda mengklasifikasikan setiap jenis, apresiasi Anda terhadap setiap jenis akan meningkat dan Anda dapat lebih memahami peran metadata dalam setiap kelompok.

Penulis dan praktisi gudang data yang berbeda mengklasifikasikan dan mengelompokkan metadata dengan berbagai cara: beberapa berdasarkan penggunaan dan beberapa lagi berdasarkan siapa yang menggunakannya. Mari kita lihat beberapa metode berbeda untuk klasifikasi metadata:

- ◆ Administratif/pengguna akhir/optimasi
- ◆ Pengembangan/penggunaan
- ◆ Di data mart/di workstation
- ◆ Membangun/memelihara/mengelola/menggunakan
- ◆ Teknis/bisnis

- ◆ Ruang belakang/ruang depan
- ◆ Intern eksternal

Pada bab sebelumnya, kita membahas cara membagi lingkungan data warehouse berdasarkan fungsi-fungsi utama. Kita dapat membayangkan lingkungan data warehouse secara fungsional dibagi menjadi tiga area yaitu akuisisi data, penyimpanan data, dan pengiriman informasi. Semua proses data warehouse terjadi di tiga area fungsional ini. Sebagai pengembang, Anda merancang proses di masing-masing dari tiga area fungsional. Masing-masing alat yang digunakan untuk proses ini membuat dan mencatat metadata dan juga dapat menggunakan dan digerakkan oleh metadata yang dicatat oleh alat lain.

Pertama, mari kita kelompokkan tipe metadata berdasarkan tiga area fungsional ini. Mengapa? Karena setiap proses data warehouse terjadi di salah satu dari tiga area ini saja. Perhitungkan semua proses yang terjadi di setiap area fungsional dan kemudian gabungkan semua proses di ketiga area fungsional tersebut. Anda akan mendapatkan satu set lengkap proses data warehouse tanpa melewatkan satu pun. Selain itu, Anda juga akan dapat menyusun daftar lengkap tipe metadata.

Mari kita beralih ke klasifikasi tipe metadata berdasarkan area fungsional di gudang data:

1. Akuisisi data
2. Penyimpanan data
3. Penyampaian informasi

Akuisisi Data

Di area ini, proses data warehouse berhubungan dengan fungsi-fungsi berikut:

- Ekstraksi data
- Transformasi data
- Pembersihan data
- Integrasi data
- Pementasan data

Saat proses berlangsung, alat yang sesuai mencatat elemen metadata yang berkaitan dengan proses tersebut. Alat tersebut mencatat elemen metadata selama fase pengembangan serta saat gudang data beroperasi setelah penerapan.

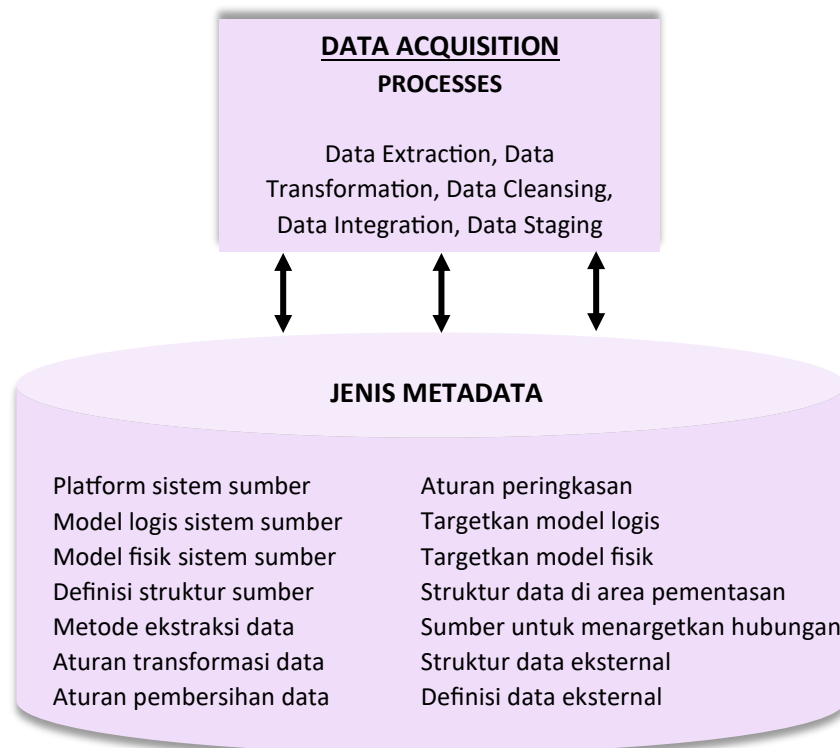
Sebagai profesional TI dan bagian dari tim proyek gudang data, Anda akan menggunakan alat pengembangan yang mencatat metadata yang berkaitan dengan area ini. Selain itu, beberapa alat lain yang akan Anda gunakan untuk proses lain baik di area ini atau di area lain mungkin menggunakan metadata yang direkam oleh alat lain di area ini. Misalnya, saat Anda menggunakan alat kueri untuk membuat kueri standar, Anda akan menggunakan metadata yang direkam oleh proses di area akuisisi data. Seperti yang akan Anda ketahui, alat kueri dimaksudkan untuk proses di area berbeda, yaitu area pengiriman informasi.

Profesional TI juga akan menggunakan metadata yang direkam oleh proses di area akuisisi data untuk mengelola dan memantau fungsi gudang data yang sedang berlangsung setelah penerapan. Anda akan menggunakan metadata dari area ini untuk memantau ekstraksi dan transformasi data yang sedang berlangsung. Anda akan memastikan bahwa

gambar pemuatan yang sedang berlangsung dibuat dengan benar dengan mengacu pada metadata dari area ini.

Pengguna gudang data Anda juga akan menggunakan metadata yang dicatat di area akuisisi data. Saat pengguna ingin menemukan sumber data untuk elemen data dalam kuerinya, dia akan mencari metadata dari area akuisisi data. Sekali lagi, ketika pengguna ingin mengetahui bagaimana margin keuntungan dihitung dan disimpan di gudang data, dia akan mencari aturan derivasi dalam metadata yang dicatat di area akuisisi data.

Untuk tipe metadata yang direkam dan digunakan di area akuisisi data, lihat Gambar 9.8. Gambar ini merangkum tipe metadata dan proses gudang data yang relevan. Cobalah untuk menghubungkan jenis dan proses metadata ini dengan lingkungan gudang data Anda.



Gambar 9.8 Akuisisi data: tipe metadata.

Penyimpanan data

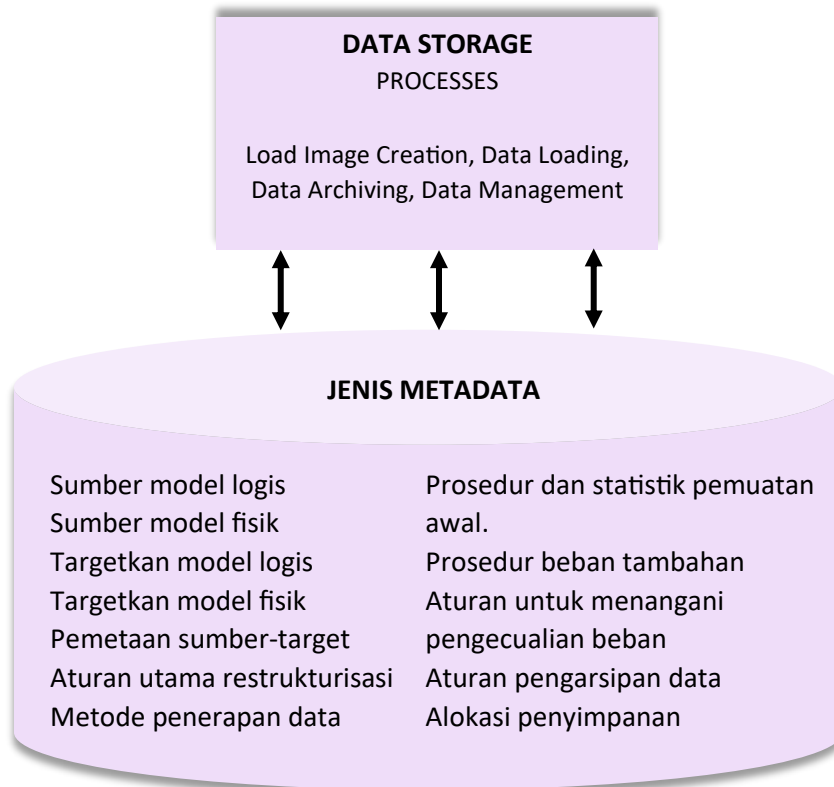
Di area ini, proses data warehouse berhubungan dengan fungsi-fungsi berikut:

- Memuat data
- Pengarsipan data
- Manajemen data

Sama seperti di area lain, saat proses berlangsung di area fungsional penyimpanan data, alat yang sesuai mencatat elemen metadata yang berkaitan dengan proses tersebut. Alat tersebut mencatat elemen metadata selama fase pengembangan serta saat gudang data beroperasi setelah penerapan.

Mirip dengan metadata yang direkam oleh proses di area akuisisi data, metadata yang direkam oleh proses di area penyimpanan data digunakan untuk pengembangan, administrasi,

dan oleh pengguna. Anda akan menggunakan metadata dari area ini untuk merancang penyegaran data lengkap dan pemuatan data tambahan. DBA akan menggunakan metadata untuk proses pencadangan, pemulihan, dan penyetelan database. Untuk membersihkan gudang data dan pengarsipan data secara berkala, metadata dari area ini akan digunakan untuk administrasi gudang data.



Gambar 9.9 Penyimpanan data: tipe metadata.

Apakah pengguna akan menggunakan metadata dari area fungsional penyimpanan data? Sebagai contoh saja, misalkan salah satu pengguna Anda ingin membuat kueri yang mengelompokkan total penjualan triwulanan berdasarkan distrik penjualan. Sebelum pengguna menjalankan kueri, dia ingin mengetahui kapan terakhir kali data penggambaran distrik dimuat. Dari mana pengguna dapat memperoleh informasi tentang tanggal pemuatan batas wilayah? Metadata yang direkam oleh proses pemuatan data di area fungsional penyimpanan data akan memberikan pengguna tanggal pemuatan terbaru untuk penggambaran distrik.

Untuk tipe metadata yang direkam dan digunakan di area penyimpanan data, lihat Gambar 9.9. Gambar ini merangkum tipe metadata dan proses gudang data yang relevan. Lihat bagaimana tipe metadata dan prosesnya berhubungan dengan lingkungan gudang data Anda.

Penyampaian Informasi

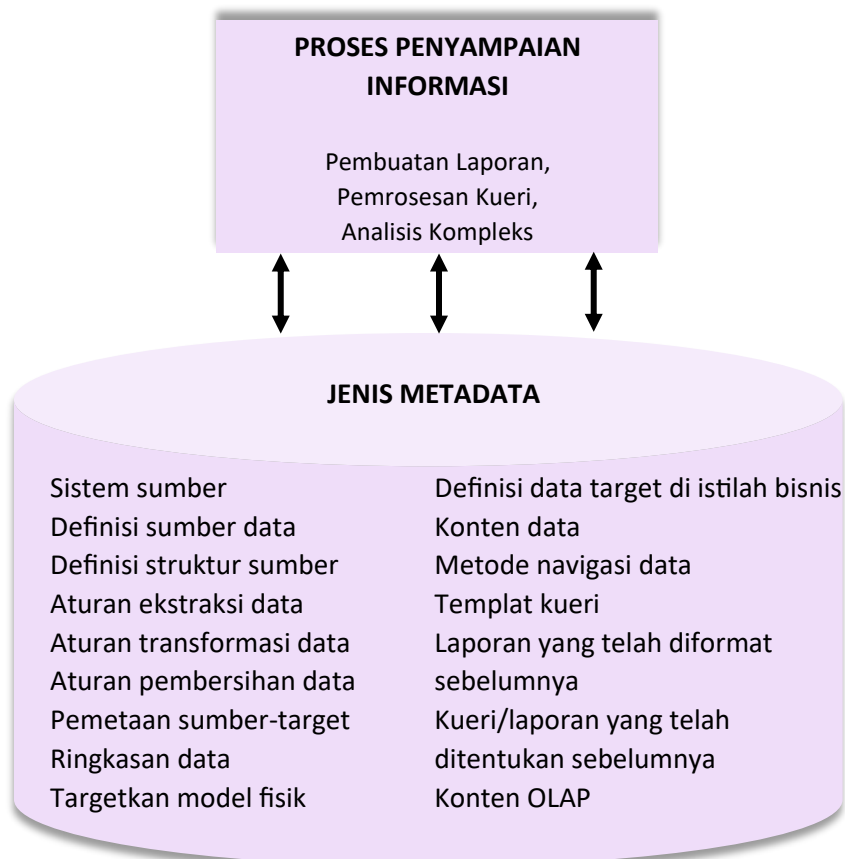
Di area ini, proses data warehouse berhubungan dengan fungsi-fungsi berikut:

- Pembuatan laporan

- Pemrosesan kueri
- Analisis yang kompleks

Sebagian besar, proses di area ini ditujukan untuk pengguna akhir. Saat menggunakan proses tersebut, pengguna akhir umumnya menggunakan metadata yang direkam dalam proses di dua area lainnya yaitu akuisisi data dan penyimpanan data. Saat pengguna membuat kueri dengan bantuan alat pemrosesan kueri, dia bisa merujuk kembali ke metadata yang direkam di area akuisisi data dan penyimpanan data dan bisa mencari konfigurasi data sumber, struktur data, dan transformasi data dari metadata. dicatat di area perolehan data. Dengan cara yang sama, dari metadata yang direkam di area penyimpanan data, pengguna dapat menemukan tanggal penyegaran penuh terakhir dan pemuatan tambahan untuk berbagai tabel di database gudang data.

Secara umum, metadata yang dicatat dalam area fungsional pengiriman informasi berkaitan dengan kueri yang telah ditentukan sebelumnya, laporan yang telah ditentukan sebelumnya, dan definisi parameter masukan untuk kueri dan laporan. Metadata yang direkam di area fungsional ini juga mencakup informasi untuk OLAP. Pengembang dan administrator terlibat dalam proses ini.



Gambar 9.10 Penyampaian informasi: tipe metadata.

Untuk tipe metadata yang direkam dan digunakan di area pengiriman informasi, lihat Gambar 9.10. Gambar ini merangkum tipe metadata dan proses gudang data yang relevan.

Lihat bagaimana tipe dan proses metadata diterapkan pada lingkungan gudang data Anda. Jenis metadata juga dapat diklasifikasikan menjadi metadata bisnis dan metadata teknis. Ini adalah metode lain yang efektif untuk mengklasifikasikan tipe metadata karena sifat dan format metadata dalam satu kelompok sangat berbeda dari kelompok lainnya. Dua bagian berikutnya membahas metode klasifikasi ini.

9.3 METADATA BISNIS

Metadata bisnis menghubungkan pengguna bisnis Anda ke gudang data Anda. Pengguna bisnis perlu mengetahui apa yang tersedia di gudang data dari sudut pandang yang berbeda dari sudut pandang profesional TI seperti Anda. Metadata bisnis seperti peta jalan atau direktori informasi yang mudah digunakan yang menunjukkan konten dan cara mencapainya. Ini seperti pemandu wisata bagi para eksekutif dan peta rute bagi para manajer dan analis bisnis.

Ikhtisar Konten

Pertama-tama, metadata bisnis harus mendeskripsikan konten dalam bahasa yang sederhana, memberikan informasi dalam istilah bisnis. Misalnya, nama tabel data atau elemen data individual tidak boleh samar-samar namun merupakan istilah yang bermakna dan familiar bagi pengguna bisnis. Nama item data `calc_pr_sle` tidak dapat diterima. Anda perlu mengganti namanya menjadi penjualan bulan sebelumnya yang dihitung.

Metadata bisnis kurang terstruktur dibandingkan metadata teknis. Sebagian besar metadata bisnis berasal dari dokumen tekstual, spreadsheet, dan bahkan aturan dan kebijakan bisnis yang tidak dituliskan secara lengkap. Meskipun sebagian besar metadata bisnis berasal dari sumber informal, metadata ini sama pentingnya dengan metadata dari sumber formal seperti entri kamus data. Semua metadata informal harus ditangkap, dimasukkan ke dalam bentuk standar, dan disimpan sebagai metadata bisnis di gudang data.

Sebagian besar pengguna bisnis tidak memiliki keahlian teknis yang cukup untuk membuat pertanyaan mereka sendiri atau memformat laporan mereka sendiri. Mereka perlu mengetahui kueri standar apa yang tersedia dan laporan terformat apa yang dapat dihasilkan. Mereka harus mampu mengidentifikasi tabel dan kolom di gudang data dengan mengacu pada nama bisnis. Oleh karena itu, metadata bisnis harus mengungkapkan semua informasi ini dalam bahasa yang sederhana.

Contoh Metadata Bisnis

Metadata bisnis berfokus pada penyediaan dukungan bagi pengguna akhir di stasiun kerja. Ini harus memudahkan pengguna akhir untuk memahami data apa yang tersedia di gudang data dan bagaimana mereka dapat menggunakannya. Metadata bisnis menggambarkan gudang data murni dari sudut pandang pengguna akhir. Ini seperti tampilan eksternal dari gudang data yang dirancang dan disusun dalam istilah bisnis sederhana yang mudah dipahami pengguna.

Mari kita coba lebih memahami metadata bisnis dengan melihat daftar contoh:

- ◆ Prosedur konektivitas
- ◆ Keamanan dan hak akses

- ◆ Struktur data secara keseluruhan dalam istilah bisnis
- ◆ Sistem sumber
- ◆ Pemetaan sumber-ke-target
- ◆ Aturan bisnis transformasi data
- ◆ Ringkasan dan derivasi
- ◆ Nama tabel dan definisi bisnis
- ◆ Nama atribut dan definisi bisnis
- ◆ Kepemilikan data
- ◆ Alat kueri dan pelaporan
- ◆ Kueri yang telah ditentukan sebelumnya
- ◆ Laporan yang telah ditentukan sebelumnya
- ◆ Laporkan informasi distribusi
- ◆ Rute akses informasi umum
- ◆ Aturan analisis menggunakan OLAP
- ◆ Mata uang data OLAP
- ◆ Jadwal penyegaran gudang data

Daftar ini tidak berarti semuanya inklusif, namun memberikan dasar yang baik bagi Anda untuk membuat daftar serupa untuk gudang data Anda. Gunakan daftar tersebut sebagai panduan untuk memastikan bahwa metadata bisnis disediakan menggunakan nama bisnis dan dibuat mudah dimengerti oleh pengguna Anda.

Sorotan Konten

Dari daftar contoh, mari kita soroti isi metadata bisnis. Apa sajakah jenis pertanyaan yang dapat dijawab oleh metadata bisnis? Jenis informasi apa yang dapat diperoleh pengguna dari metadata bisnis?

Mari kita dapatkan daftar pertanyaan yang dapat dijawab oleh metadata bisnis bagi pengguna akhir. Meskipun daftar berikut tidak mencakup semua pertanyaan yang mungkin diajukan pengguna, daftar ini dapat menjadi referensi yang berguna:

- Bagaimana saya bisa masuk dan terhubung dengan gudang data?
- Bagian gudang data manakah yang dapat saya akses?
- Bisakah saya melihat semua atribut dari tabel tertentu?
- Apa definisi atribut yang saya perlukan dalam kueri saya?
- Apakah ada pertanyaan dan laporan yang sudah ditentukan sebelumnya untuk memberikan hasil yang saya perlukan?
- Dari sistem sumber manakah data yang saya inginkan berasal?
- Nilai default apa yang digunakan untuk item data yang diambil oleh kueri saya?
- Jenis agregasi apa yang tersedia untuk metrik yang diperlukan?
- Bagaimana nilai item data yang saya perlukan diperoleh dari item data lainnya?
- Kapan terakhir kali pembaruan item data dalam kueri saya?
- Pada item data manakah saya dapat melakukan analisis penelusuran?
- Berapa umur data OLAP? Haruskah saya menunggu pembaruan berikutnya?

Siapa yang Diuntungkan?

Metadata bisnis terutama menguntungkan pengguna akhir. Ini adalah pernyataan umum. Siapa yang secara spesifik mendapat manfaat dari metadata bisnis? Bagaimana metadata bisnis melayani anggota komunitas pengguna akhir tertentu? Lihatlah daftar berikut:

- Manajer
- Analisis bisnis
- Pengguna yang mahir
- Pengguna biasa
- Pengguna biasa
- Manajer senior/eksekutif junior

9.4 METADATA TEKNIS

Metadata teknis ditujukan untuk staf TI yang bertanggung jawab atas pengembangan dan administrasi gudang data. Personil teknis memerlukan informasi untuk merancang setiap proses. Ini adalah proses di setiap area fungsional gudang data. Anda, sebagai bagian dari kelompok teknis di tim proyek, harus mengetahui struktur dan konten gudang data yang diusulkan. Anggota tim proyek yang berbeda memerlukan jenis informasi berbeda dari metadata teknis. Jika metadata bisnis seperti peta jalan bagi pengguna untuk menggunakan gudang data, maka metadata teknis seperti panduan dukungan bagi profesional TI untuk membangun, memelihara, dan mengelola gudang data.

Ikhtisar Konten

Staf TI yang bekerja pada proyek gudang data memerlukan metadata teknis untuk tujuan yang berbeda. Jika Anda ahli akuisisi data, kebutuhan Anda akan metadata berbeda dengan kebutuhan pengembang akses informasi di tim. Secara keseluruhan, staf teknis proyek perlu memahami proses ekstraksi data, transformasi data, dan pembersihan data. Mereka harus mengetahui tata letak keluaran dari setiap rutinitas ekstraksi dan harus memahami aturan transformasi data.

Staf TI memerlukan metadata teknis untuk tiga tujuan berbeda. Pertama, personel TI memerlukan metadata teknis untuk pengembangan awal gudang data. Katakanlah Anda bertanggung jawab atas desain dan pengembangan proses transformasi data. Untuk tujuan ini, metadata dari proses ekstraksi data sebelumnya dapat membantu upaya pengembangan Anda.

Kedua, metadata teknis sangat penting untuk pertumbuhan dan pemeliharaan gudang data yang berkelanjutan. Jika Anda bertanggung jawab untuk membuat perubahan pada beberapa struktur data, atau bahkan untuk rilis kedua gudang data, di mana Anda akan menemukan informasi mengenai konten dan berbagai proses? Anda memerlukan metadata teknis.

Metadata teknis juga penting untuk administrasi gudang data produksi yang berkelanjutan. Sebagai administrator, Anda harus memantau ekstraksi data yang sedang berlangsung. Anda harus memastikan bahwa beban tambahan diselesaikan dengan benar dan

tepat waktu. Tanggung jawab Anda mungkin juga mencakup pencadangan basis data dan pengarsipan data lama. Administrasi gudang data hampir tidak mungkin dilakukan tanpa metadata teknis.

Contoh Metadata Teknis

Metadata teknis berkonsentrasi pada dukungan bagi staf TI yang bertanggung jawab atas pengembangan, pemeliharaan, dan administrasi. Metadata teknis lebih terstruktur dibandingkan metadata bisnis. Metadata teknis seperti tampilan internal gudang data yang menunjukkan detail batin dalam istilah teknis. Berikut adalah daftar contoh metadata teknis:

- Model data sistem sumber
- Catat tata letak sumber luar
- Sumber untuk melakukan pemetaan wilayah
- Pementasan area untuk pemetaan gudang data
- Aturan dan jadwal ekstraksi data
- Aturan dan pembuatan versi transformasi data
- Aturan agregasi data
- Aturan pembersihan data
- Ringkasan dan derivasi
- Jadwal dan kontrol pemuatan dan penyegaran data
- Ketergantungan pekerjaan

Nama dan deskripsi program

- Model data gudang data
- Nama basis data
- Nama tabel/tampilan
- Nama kolom dan deskripsinya
- Atribut utama
- Aturan bisnis untuk entitas dan hubungan
- Pemetaan antara model logis dan fisik
- Informasi jaringan/server
- Konektivitas data
- Pengendalian audit pergerakan data
- Aturan pembersihan dan pengarsipan data
- Otoritas/hak istimewa akses
- Penggunaan/waktu data
- Pola akses kueri dan laporan
- Alat kueri dan pelaporan

Tinjau daftar tersebut dan kembangkan daftar yang sebanding untuk lingkungan gudang data Anda.

Sorotan Konten

Daftar contoh memberi Anda gambaran tentang jenis informasi yang harus dimiliki metadata teknis di lingkungan gudang data. Seperti halnya metadata bisnis, mari kita buat

daftar pertanyaan yang dapat dijawab oleh metadata teknis untuk pengembang dan administrator. Tinjau daftar berikut:

- ◆ Database dan tabel apa saja yang ada?
- ◆ Berapakah kolom untuk setiap tabel?
- ◆ Apa saja kunci dan indeksinya?
- ◆ Apa saja file fisiknya?
- ◆ Apakah uraian bisnis sesuai dengan uraian teknis?
- ◆ Kapan terakhir kali pembaruan berhasil?
- ◆ Apa saja sistem sumber dan struktur datanya?
- ◆ Apa aturan ekstraksi data untuk setiap sumber data?
- ◆ Sumber apa yang menjadi target pemetaan untuk setiap item data di gudang data?
- ◆ Apa saja aturan transformasi data?
- ◆ Nilai default apa yang digunakan untuk item data saat membersihkan data yang hilang?
- ◆ Jenis agregasi apa yang tersedia?
- ◆ Apa saja bidang turunan dan aturan turunannya?
- ◆ Kapan terakhir kali pembaruan item data dalam kueri saya?
- ◆ Bagaimana jadwal pemuatan dan penyegarannya?
- ◆ Seberapa sering data dibersihkan atau diarsipkan? Item data yang mana?
- ◆ Bagaimana jadwal pembuatan data untuk OLAP?
- ◆ Alat kueri dan laporan apa yang tersedia?

Siapa yang Diuntungkan?

Daftar berikut menunjukkan tipe personel tertentu yang akan mendapatkan manfaat dari metadata teknis:

- Manajer proyek
- Administrator gudang data
- Administrator basis data
- Manajer metadata
- Arsitek gudang data
- Pengembang akuisisi data
- Analisis kualitas data
- Analisis bisnis
- Administrator sistem
- Spesialis infrastruktur
- Pemodel data
- Arsitek keamanan

9.5 BAGAIMANA MEMBERIKAN METADATA

Saat gudang data Anda dirancang dan dibangun, metadata perlu dikumpulkan dan dicatat. Seperti yang Anda ketahui, metadata menggambarkan gudang data Anda dari berbagai sudut pandang. Anda melihat gudang data melalui metadata untuk menemukan

sumber data, memahami ekstraksi dan transformasi data, menentukan cara menavigasi konten, dan mengambil informasi. Sebagian besar proses data warehouse dilakukan dengan bantuan perangkat lunak. Metadata yang sama atau salinan asli dari subset yang relevan harus tersedia untuk setiap alat.

Dalam studi yang dilakukan oleh Data Warehousing Institute pada tahun-tahun awal data warehousing, 86% responden sepenuhnya menyadari pentingnya memiliki strategi manajemen metadata. Namun, hanya 9% yang telah menerapkan solusi metadata. 16% lainnya mempunyai rencana dan mulai mengerjakan implementasinya. Situasi ini sudah agak membaik di lingkungan saat ini.

Jika sebagian besar perusahaan yang memiliki gudang data menyadari betapa pentingnya manajemen metadata, mengapa hanya sebagian kecil yang melakukan tindakan untuk mengatasinya? Manajemen metadata menghadirkan tantangan besar. Tantangannya bukan pada penangkapan metadata melalui penggunaan alat selama proses gudang data, namun terletak pada integrasi metadata dari berbagai alat yang membuat dan memelihara metadatanya sendiri.

Kami akan mengeksplorasi tantangannya. Bagaimana Anda dapat menemukan opsi untuk mengatasi tantangan dan membangun manajemen metadata yang efektif di lingkungan gudang data Anda? Apa yang terjadi di industri ini? Sementara standar sedang disusun dalam koalisi industri, apakah ada pilihan sementara bagi Anda? Pertama, mari kita tetapkan persyaratan dasar untuk pengelolaan metadata yang baik. Apa saja persyaratannya? Selanjutnya, kami akan mempertimbangkan sumber metadata sebelum kami mengkaji tantangannya.

Persyaratan Metadata

Sederhananya, metadata harus berfungsi sebagai peta jalan menuju gudang data bagi pengguna Anda. Hal ini juga harus mendukung TI dalam pengembangan dan administrasi data warehouse. Mari kita melampaui pernyataan sederhana ini dan melihat persyaratan spesifik untuk manajemen metadata.

Menangkap dan Menyimpan Data Kamus data dalam sistem operasional menyimpan struktur dan aturan bisnis sebagaimana adanya pada saat ini. Untuk sistem operasional, tidak perlu menyimpan riwayat entri kamus data. Namun, riwayat data di gudang data Anda mencakup beberapa tahun, biasanya 5 hingga 15 tahun di sebagian besar gudang data. Selama ini, perubahan memang terjadi pada sistem sumber, metode ekstraksi data, algoritma transformasi data, dan pada struktur serta konten database data warehouse itu sendiri. Oleh karena itu, metadata di lingkungan gudang data harus melacak revisinya. Dengan demikian, manajemen metadata harus menyediakan sarana untuk menangkap dan menyimpan metadata dengan versi yang tepat untuk menunjukkan fitur varian waktu.

Berbagai Sumber Metadata Metadata untuk data warehouse tidak pernah berasal dari satu sumber. Alat CASE, sistem operasional sumber, alat ekstraksi data, alat transformasi data, definisi kamus data, dan sumber lainnya semuanya berkontribusi pada metadata gudang data. Oleh karena itu, pengelolaan metadata harus cukup terbuka untuk menangkap metadata dari berbagai sumber.

Integrasi Metadata Kita telah melihat elemen metadata bisnis dan teknis. Anda harus dapat mengintegrasikan dan menggabungkan semua elemen ini secara terpadu agar dapat bermakna bagi pengguna akhir Anda. Metadata dari model data sistem sumber harus diintegrasikan dengan metadata dari model data database gudang data. Integrasi harus berlanjut lebih jauh ke alat front-end yang digunakan oleh pengguna akhir. Semua ini adalah proposisi yang sulit dan sangat menantang.

Standardisasi Metadata Jika alat ekstraksi data dan alat transformasi data mewakili struktur data, maka kedua alat tersebut harus mencatat metadata tentang struktur data dengan cara standar yang sama. Metadata yang sama di penyimpanan metadata berbeda pada alat berbeda harus direpresentasikan dengan cara yang sama.

Gelombang Revisi Revisi akan terjadi pada metadata seiring dengan perubahan data atau aturan bisnis. Saat revisi metadata dilacak dalam satu proses gudang data, revisi tersebut harus menyebar ke seluruh gudang data ke proses lainnya. Menjaga Metadata Tersinkronisasi Metadata tentang struktur data, elemen data, peristiwa, aturan, dan sebagainya harus tetap tersinkronisasi setiap saat di seluruh gudang data.

Pertukaran Metadata Saat pengguna akhir Anda menggunakan alat front-end untuk akses informasi, mereka harus dapat melihat metadata yang direkam oleh alat back-end seperti alat transformasi data. Pertukaran metadata secara gratis dan mudah dari satu alat ke alat lainnya harus dimungkinkan.

Dukungan untuk Pengguna Akhir Manajemen metadata harus memberikan presentasi grafis dan tabel sederhana kepada pengguna akhir, sehingga memudahkan mereka menelusuri metadata dan memahami data di gudang data murni dari perspektif bisnis.

Persyaratan yang tercantum sangat valid untuk pengelolaan metadata. Integrasi dan standarisasi metadata merupakan tantangan besar. Namun demikian, sebelum mengatasi masalah ini, Anda perlu mengetahui sumber metadata yang umum. Daftar umum sumber metadata akan membantu Anda menetapkan inisiatif manajemen metadata untuk gudang data Anda.

Sumber Metadata

Saat alat digunakan untuk berbagai proses gudang data, metadata dicatat sebagai produk sampingan. Misalnya, ketika alat transformasi data digunakan, metadata pada pemetaan sumber ke target dicatat sebagai produk sampingan dari proses yang dilakukan dengan alat tersebut. Mari kita lihat semua sumber metadata biasa tanpa referensi apa pun ke proses individual.

Sistem Sumber

- ✘ Model data sistem operasional (manual atau dengan alat CASE)
- ✘ Definisi elemen data dari dokumentasi sistem
- ✘ Copybook COBOL dan spesifikasi blok kontrol
- ✘ Tata letak file fisik dan definisi bidang
- ✘ Spesifikasi program
- ✘ Tata letak file dan definisi bidang untuk data dari sumber luar
- ✘ Sumber lain seperti spreadsheet dan daftar manual

Ekstraksi Data

- ☞ Data tentang platform sumber dan konektivitas
- ☞ Tata letak dan definisi sumber data yang dipilih
- ☞ Definisi bidang yang dipilih untuk ekstraksi
- ☞ Kriteria penggabungan ke dalam file ekstrak awal pada setiap platform
- ☞ Aturan untuk membakukan jenis dan panjang lapangan
- ☞ Jadwal ekstraksi data
- ☞ Metode ekstraksi untuk perubahan tambahan
- ☞ Aliran pekerjaan ekstraksi data

Transformasi dan Pembersihan Data

- ✚ Spesifikasi untuk memetakan file yang diekstraksi ke file pementasan data
- ✚ Aturan konversi untuk masing-masing file
- ✚ Nilai default untuk kolom yang nilainya hilang
- ✚ Aturan bisnis untuk pemeriksaan validitas
- ✚ Pengaturan penyortiran dan pengurutan ulang
- ✚ Jejak audit untuk pergerakan dari ekstraksi data ke pementasan data

Memuat Data

- ❖ Spesifikasi untuk memetakan file pementasan data untuk memuat gambar
- ❖ Aturan untuk menetapkan kunci untuk setiap file
- ❖ Jejak audit untuk pergerakan dari pementasan data hingga memuat gambar
- ❖ Jadwal untuk penyegaran penuh
- ❖ Jadwal untuk beban tambahan
- ❖ Aliran pekerjaan pemuatan data

Penyimpanan data

- ▶ Model data untuk gudang data terpusat dan data mart dependen
- ▶ Pengelompokan tabel pada area subjek
- ▶ Model data untuk data mart yang disesuaikan
- ▶ File fisik
- ▶ Definisi tabel dan kolom
- ▶ Aturan bisnis untuk pemeriksaan validitas

Penyampaian Informasi

- ☞ Daftar alat kueri dan laporan
- ☞ Daftar pertanyaan dan laporan yang telah ditentukan sebelumnya
- ☞ Model data untuk database khusus untuk OLAP
- ☞ Jadwal pengambilan data untuk OLAP

Tantangan untuk Manajemen Metadata

Meskipun metadata sangat penting dalam lingkungan gudang data, mengintegrasikan semua bagian metadata dengan lancar adalah tugas yang berat. Standardisasi industri secara luas masih jauh dari kenyataan. Metadata yang dibuat oleh suatu proses di satu ujung tidak dapat dilihat melalui alat yang digunakan di ujung lain tanpa melalui transformasi yang

berbelit-belit. Tantangan-tantangan ini memaksa banyak pengembang gudang data untuk mengabaikan persyaratan manajemen metadata yang tepat.

Berikut adalah tantangan utama yang harus diatasi saat menyediakan metadata:

- ✚ Setiap perangkat lunak mempunyai metadata yang sesuai. Jika Anda menggunakan beberapa alat di gudang data Anda, bagaimana Anda bisa merekonsiliasi formatnya?
- ✚ Tidak ada standar yang diterima secara luas di industri untuk format metadata.
- ✚ Ada klaim yang bertentangan mengenai keunggulan penyimpanan metadata terpusat dibandingkan dengan kumpulan penyimpanan metadata yang terfragmentasi.
- ✚ Tidak ada metode yang mudah dan diterima untuk meneruskan metadata sepanjang proses ketika data berpindah dari sistem sumber ke area pementasan dan selanjutnya ke penyimpanan gudang data.
- ✚ Mempertahankan kontrol versi metadata secara seragam di seluruh gudang data adalah hal yang membosankan dan sulit.
- ✚ Dalam gudang data yang besar dengan banyak sistem sumber, menyatukan metadata yang berkaitan dengan sumber data bisa menjadi tugas yang sangat besar. Anda harus menghadapi standar, format, konvensi penamaan data, definisi data, atribut, nilai, aturan bisnis, dan satuan ukuran yang saling bertentangan. Anda harus menyelesaikan penggunaan alias yang sembarangan dan mengkompensasi aturan validasi data yang tidak memadai.

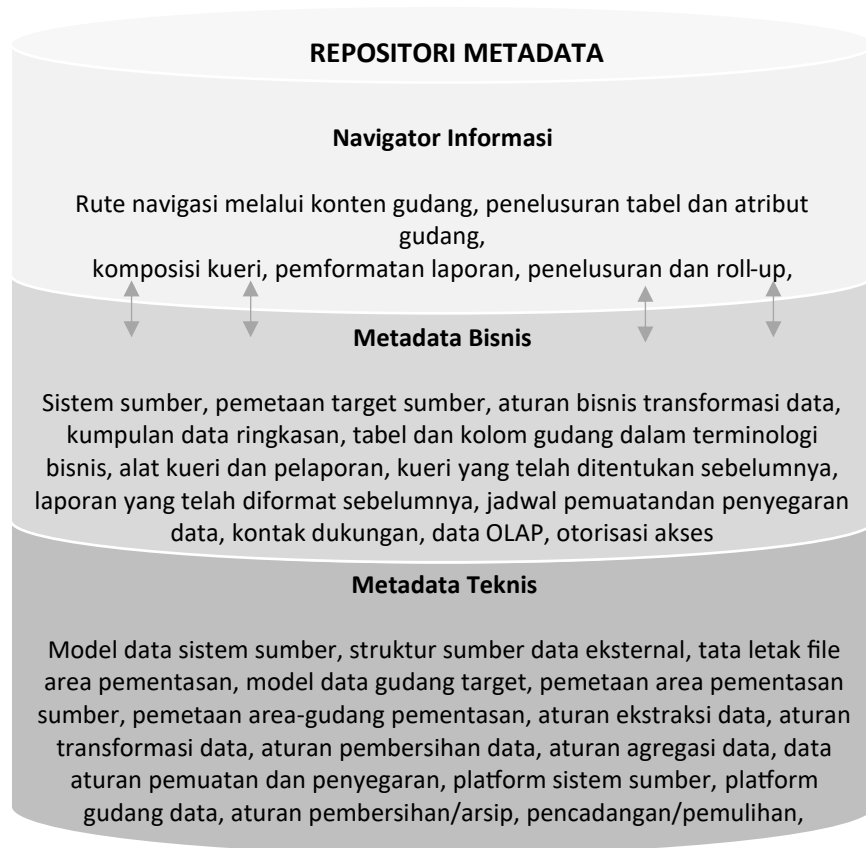
Repositori Metadata

Bayangkan repositori metadata sebagai direktori informasi tujuan umum atau perangkat katalog untuk mengklasifikasikan, menyimpan, dan mengelola metadata. Seperti yang telah kita lihat sebelumnya, metadata bisnis dan metadata teknis memiliki tujuan yang berbeda. Pengguna akhir memerlukan metadata bisnis; pengembang dan administrator gudang data memerlukan metadata teknis. Struktur kedua kategori metadata ini juga berbeda-beda. Oleh karena itu, repositori metadata dapat dianggap sebagai dua direktori informasi yang berbeda, satu untuk menyimpan metadata bisnis dan yang lainnya untuk menyimpan metadata teknis. Pembagian ini mungkin juga logis dalam satu repositori fisik.

Gambar 9.11 menunjukkan konten khas dalam repositori metadata. Perhatikan pembagian antara metadata bisnis dan teknis. Apakah Anda juga memperhatikan komponen lain yang disebut navigator informasi? Komponen ini diimplementasikan dengan cara yang berbeda dalam penawaran komersial. Fungsi navigator informasi antara lain sebagai berikut:

- 1) Antarmuka dari Alat Kueri. Fungsi ini melampirkan data gudang data ke alat kueri pihak ketiga sehingga definisi metadata di dalam metadata teknis dapat dilihat dari alat ini.
- 2) Telusuri untuk Detailnya. Pengguna metadata dapat menelusuri dan melanjutkan dari satu tingkat metadata ke tingkat yang lebih rendah untuk mendapatkan informasi lebih lanjut. Misalnya, Anda bisa mendapatkan definisi tabel data terlebih dahulu, lalu melanjutkan ke tingkat berikutnya untuk melihat semua atribut, dan melangkah lebih jauh untuk mendapatkan detail masing-masing atribut.

- 3) Tinjau Kueri dan Laporan yang Telah Ditentukan Sebelumnya. Pengguna dapat meninjau kueri dan laporan yang telah ditentukan sebelumnya, dan meluncurkan yang dipilih dengan parameter yang tepat.



Gambar 9.11 Repositori metadata.

Repositori metadata terpusat yang dapat diakses dari seluruh bagian gudang data untuk pengguna akhir, pengembang, dan administrator Anda tampaknya menjadi solusi ideal untuk manajemen metadata. Namun agar repositori metadata terpusat menjadi solusi terbaik, repositori tersebut harus memenuhi beberapa persyaratan dasar. Mari kita segera meninjau persyaratan ini. Tidak mudah untuk menemukan alat repositori yang memenuhi semua persyaratan yang tercantum di bawah.

Organisasi yang Fleksibel. Izinkan administrator data untuk mengklasifikasikan dan mengatur metadata ke dalam kategori dan subkategori logis, dan menetapkan komponen metadata tertentu ke klasifikasi tersebut.

- a) Historis; Gunakan pembuatan versi untuk mempertahankan perspektif historis metadata.
- b) Terintegrasi; Simpan metadata bisnis dan teknis dalam format yang dapat dimengerti oleh semua jenis pengguna.
- c) Kompartementalisasi yang Baik; Mampu memisahkan dan menyimpan model database logis dan fisik.

- d) Kemampuan Analisis dan Pencarian; Mampu menelusuri seluruh bagian metadata dan juga menavigasi hubungan.
- e) Dapat disesuaikan; Mampu membuat tampilan metadata yang disesuaikan untuk masing-masing kelompok pengguna dan menyertakan objek metadata baru jika diperlukan.
- f) Pertahankan Deskripsi dan Definisi; Lihat metadata dalam istilah bisnis dan teknis.
- g) Standarisasi Konvensi Penamaan; Fleksibilitas untuk mengadopsi segala jenis konvensi penamaan dan melakukan standarisasi di seluruh repositori metadata.
- h) Sinkronisasi; Jaga agar metadata tetap tersinkronisasi di seluruh bagian lingkungan gudang data dan dengan sistem eksternal terkait.
- i) Membuka; Mendukung pertukaran metadata antar proses melalui antarmuka standar industri dan kompatibel dengan berbagai macam alat.

Pemilihan produk repositori metadata yang sesuai adalah salah satu keputusan penting yang harus diambil oleh tim proyek. Gunakan daftar kriteria di atas sebagai panduan saat mengevaluasi alat repositori untuk gudang data Anda.

Integrasi dan Standar Metadata

Untuk pertukaran metadata secara bebas dalam gudang data antara proses yang dilakukan dengan bantuan perangkat lunak, kebutuhan akan standarisasi sudah jelas. Diskusi kita sejauh ini pasti telah meyakinkan Anda akan hal ini. Seperti disebutkan dalam Bab 3, Koalisi Meta Data dan Grup Manajemen Objek telah mengerjakan standar metadata. Koalisi Meta Data telah menerima standar yang dikenal sebagai Model Informasi Terbuka (OIM). Grup Manajemen Objek telah merilis *Common Warehouse Metamodel (CWM)* sebagai standarnya. Seperti disebutkan sebelumnya, pada bulan November 2000 kedua badan ini bergabung menjadi OMG. Sejak saat itu, beberapa integrasi dari kedua standar tersebut telah muncul dan dapat diterima secara luas oleh industri.

Anda perlu menyadari upaya ini untuk mencapai tujuan standar metadata yang bermanfaat. Selain itu, perhatikan hal-hal penting berikut dari inisiatif ini yang berkaitan dengan metadata gudang data:

- ❖ Model standar menyediakan konsep metadata untuk manajemen skema database, desain, dan penggunaan kembali dalam lingkungan data warehouse. Ini mencakup konsep database logis dan fisik.
- ❖ Model ini mencakup rincian transformasi data yang dapat diterapkan pada pengisian gudang data.
- ❖ Model ini dapat diperluas untuk menyertakan tipe metadata khusus OLAP yang menangkap deskripsi kubus data.
- ❖ Model standar berisi rincian untuk menentukan skema sumber dan target serta transformasi data antara skema yang biasa ditemukan dalam proses akuisisi data di lingkungan gudang data. Jenis metadata ini dapat digunakan untuk mendukung desain transformasi, analisis dampak (transformasi mana yang dipengaruhi oleh perubahan skema tertentu), dan silsilah data (sumber data dan transformasi mana yang digunakan untuk menghasilkan data tertentu di gudang data) .

- ❖ Komponen transformasi model standar menangkap informasi tentang skrip transformasi data gabungan. Transformasi individu mempunyai hubungan dengan sumber dan sasaran transformasi. Beberapa semantik transformasi dapat ditangkap oleh batasan dan kumpulan kode-dekode untuk pemetaan berbasis tabel.

Opsi Implementasi

Sudah cukup banyak yang dikatakan tentang kebutuhan mutlak metadata dalam lingkungan gudang data. Pada saat yang sama, kami telah mencatat perlunya integrasi dan standar metadata. Terkait dengan kedua fakta ini adalah kenyataan kurangnya standar metadata yang diterima secara universal. Oleh karena itu, dalam lingkungan gudang data yang umum di mana beberapa alat dari vendor berbeda digunakan, apa saja pilihan untuk menerapkan manajemen metadata? Di bagian ini, kita akan mengeksplorasi beberapa opsi acak. Namun kita harus berharap bahwa tujuan standar universal akan segera tercapai.

Tinjau opsi berikut dan pertimbangkan opsi yang paling sesuai untuk lingkungan gudang data Anda.

- ❖ Pilih dan gunakan produk repositori metadata dengan komponen direktori informasi bisnisnya. Alat akses informasi dan akuisisi data Anda yang kompatibel dengan produk repositori akan berinteraksi dengan mulus dengannya. Untuk alat lain yang tidak kompatibel, Anda harus mencari metode integrasi lain.
- ❖ Menurut pendapat beberapa konsultan data warehouse, satu repositori terpusat adalah pendekatan terbatas yang membahayakan otonomi proses individual. Meskipun repositori terpusat memungkinkan berbagi metadata, repositori tersebut tidak dapat dikelola dengan mudah di gudang data yang besar. Dalam pendekatan desentralisasi, metadata tersebar di berbagai bagian arsitektur dengan beberapa penyimpanan metadata pribadi dan unik. Pertukaran metadata bisa menjadi masalah.
- ❖ Beberapa pengembang telah menemukan solusinya sendiri. Mereka telah membuat serangkaian prosedur untuk penggunaan standar setiap alat di lingkungan pengembangan dan menyediakan daftar isi.
- ❖ Pengembang lain membuat database mereka sendiri untuk mengumpulkan dan menyimpan metadata dan mempublikasikannya di intranet perusahaan.
- ❖ Beberapa mengadopsi metode cerdas dalam mengintegrasikan akses informasi dan alat analisis. Mereka menyediakan tampilan metadata secara berdampingan dengan satu alat dan tampilan data sebenarnya dengan alat lain. Terkadang, teks bantuan di alat kueri mungkin diisi dengan metadata yang diekspor dari repositori pusat.

Seperti yang Anda ketahui, tren saat ini adalah penggunaan teknologi Web untuk fungsi pelaporan dan OLAP. Intranet perusahaan banyak digunakan sebagai sarana penyampaian informasi. Gambar 9.12 menunjukkan bagaimana perubahan paradigma ini mengubah cara metadata diakses.

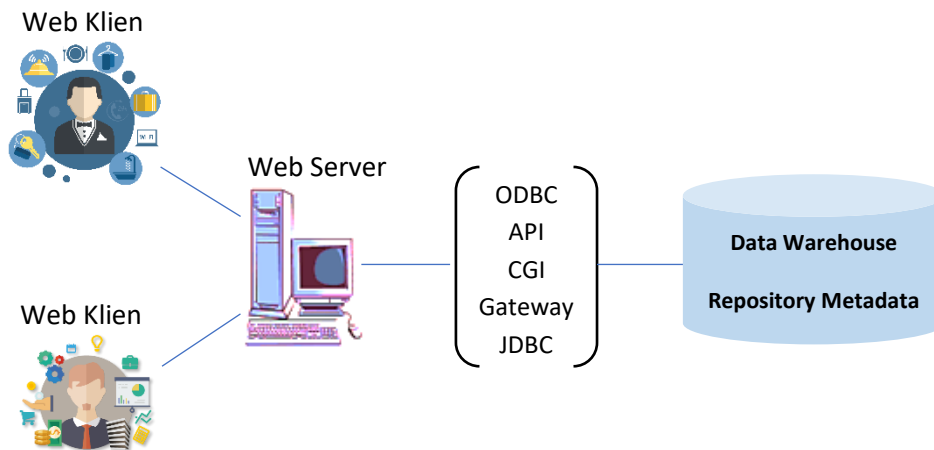
Pengguna bisnis dapat menggunakan browser Web mereka untuk mengakses metadata dan menavigasi gudang data dan data mart apa pun. Sejak awal, berikan perhatian khusus pada metadata untuk lingkungan gudang data Anda.

Siapkan inisiatif metadata untuk menjawab pertanyaan-pertanyaan berikut:

- ✓ Apa tujuan metadata di perusahaan Anda? Metadata apa yang diperlukan untuk mencapai tujuan?
- ✓ Apa saja sumber metadata di lingkungan Anda? Siapa yang akan memeliharanya?
- ✓ Bagaimana mereka mempertahankannya?

Apa standar metadatanya?

Bagaimana metadata akan digunakan? Oleh siapa? Alat metadata apa yang diperlukan? Tetapkan tujuan Anda untuk metadata di lingkungan Anda dan tindak lanjuti.



Gambar 9.12 Metadata: Akses berbasis web.

RINGKASAN BAB

- ❖ Metadata merupakan kebutuhan penting untuk menggunakan, membangun, dan mengelola gudang data.
- ❖ Bagi pengguna akhir, metadata seperti peta jalan menuju isi gudang data.
- ❖ Bagi profesional TI, metadata mendukung fungsi pengembangan dan administrasi.
- ❖ Metadata berperan aktif dalam gudang data dan membantu otomatisasi proses.
- ❖ Jenis metadata dapat diklasifikasikan berdasarkan tiga area fungsional gudang data, yaitu akuisisi data, penyimpanan data, dan pengiriman informasi. Jenis-jenis tersebut terkait dengan proses yang terjadi di ketiga bidang tersebut.
- ❖ Metadata bisnis menghubungkan pengguna bisnis ke gudang data. Metadata teknis ditujukan untuk staf TI yang bertanggung jawab atas pengembangan dan administrasi.
- ❖ Metadata yang efektif harus memenuhi sejumlah persyaratan. Manajemen metadata itu sulit; banyak tantangan yang harus dihadapi.
- ❖ Standardisasi metadata universal masih merupakan tujuan yang sulit dicapai. Kurangnya standarisasi menghambat kelancaran penyampaian metadata dari satu alat ke alat lainnya.
- ❖ Repositori metadata seperti direktori informasi tujuan umum yang mencakup beberapa fungsi peningkatan.
- ❖ Salah satu opsi implementasi metadata mencakup penggunaan repositori metadata komersial. Ada kemungkinan pilihan lain yang bisa ditanam di dalam negeri.

LATIHAN SOAL

- 1) Menurut Anda mengapa metadata penting dalam lingkungan gudang data? Berikan penjelasan umum dalam satu atau dua paragraf.
- 2) Jelaskan bagaimana metadata sangat penting untuk pengembangan dan administrasi data warehouse.
- 3) Mengkaji konsep bahwa metadata itu seperti pusat saraf. Jelaskan bagaimana konsep ini diterapkan pada lingkungan data warehouse.
- 4) Sebutkan dan jelaskan tiga alasan utama mengapa metadata sangat penting bagi pengguna akhir.
- 5) Mengapa metadata penting bagi TI? Sebutkan enam proses yang metadatanya penting bagi TI dan jelaskan alasannya.
- 6) Pilih tiga proses yang metadatanya membantu otomatisasi proses ini. Tunjukkan bagaimana metadata berperan aktif dalam proses ini.
- 7) Apa yang dimaksud dengan menetapkan konteks informasi? Jelaskan secara singkat dengan contoh bagaimana metadata menetapkan konteks informasi dalam gudang data.
- 8) Sebutkan empat tipe metadata yang digunakan di masing-masing dari tiga bidang akuisisi data, penyimpanan data, dan penyampaian informasi.
- 9) Sebutkan sepuluh contoh metadata bisnis.
- 10) Sebutkan empat persyaratan utama yang harus dipenuhi metadata. Jelaskan masing-masing dari empat persyaratan ini.

LATIHAN TAMBAHAN

1. Tunjukkan apakah benar atau salah:
 - A. Pentingnya metadata dalam gudang data sama pentingnya dengan dalam sistem operasional.
 - B. Metadata dibutuhkan oleh IT untuk administrasi data warehouse.
 - C. Metadata teknis biasanya kurang terstruktur dibandingkan metadata bisnis.
 - D. Memelihara metadata di gudang data modern hanya untuk dokumentasi.
 - E. Metadata memberikan informasi tentang kueri yang telah ditentukan sebelumnya.
 - F. Metadata bisnis berasal dari sumber yang lebih bervariasi dibandingkan metadata teknis.
 - G. Metadata teknis dibagikan antara pengguna bisnis dan staf TI.
 - H. Repositori metadata seperti alat direktori tujuan umum.
 - I. Standar metadata memfasilitasi pertukaran metadata antar alat.
 - J. Metadata bisnis hanya untuk pengguna bisnis; metadata bisnis tidak dapat dipahami atau digunakan oleh staf TI.
2. Sebagai project manager pengembangan data warehouse produsen minuman ringan dalam negeri, tugas Anda adalah menulis proposal penyediaan metadata. Pertimbangkan pilihannya dan tentukan apa yang menurut Anda diperlukan dan bagaimana Anda berencana menerapkan strategi metadata.

3. Sebagai administrator gudang data, jelaskan semua jenis metadata yang Anda perlukan untuk melakukan pekerjaan Anda. Jelaskan bagaimana tipe ini akan membantu Anda.
4. Anda bertanggung jawab untuk melatih pengguna akhir gudang data. Tulis prosedur singkat bagi pengguna akhir biasa Anda untuk menggunakan metadata bisnis dan menjalankan kueri. Jelaskan prosedurnya dalam istilah pengguna tanpa menggunakan kata metadata.
5. Sebagai spesialis akuisisi data, jenis metadata apa yang dapat membantu Anda? Pilih salah satu proses akuisisi data dan jelaskan peran metadata dalam proses tersebut.

BAB 10

PRINSIP PEMODELAN DIMENSI

TUJUAN BAB

- Memahami dengan jelas bagaimana definisi persyaratan menentukan desain data
- Memperkenalkan pemodelan dimensi dan membedakannya dengan pemodelan hubungan entitas
- Tinjau dasar-dasar skema STAR
- Cari tahu apa yang ada di dalam tabel fakta dan di dalam tabel dimensi
- Menentukan keuntungan skema STAR untuk data warehouse
- Tinjau contoh skema STAR

10.1 DARI PERSYARATAN HINGGA DESAIN DATA

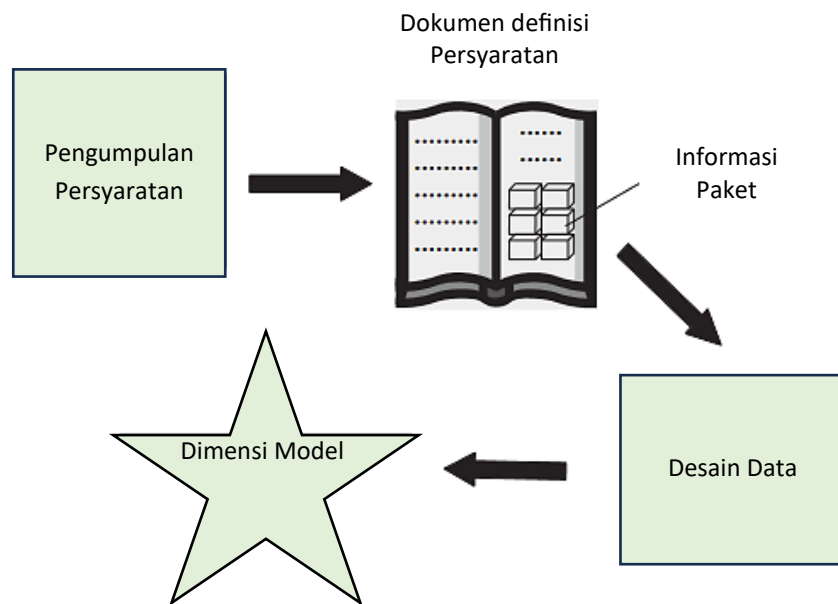
Definisi persyaratan sepenuhnya mendorong desain data untuk gudang data. Desain data terdiri dari menyusun struktur data. Sekelompok elemen data membentuk struktur data. Perancangan data logis mencakup penentuan berbagai elemen data yang diperlukan dan kombinasi elemen data tersebut ke dalam struktur data. Desain data logis juga mencakup membangun hubungan antar struktur data.

Mari kita lihat Gambar 10.1. Perhatikan bagaimana fase dimulai dengan pengumpulan persyaratan. Hasil tahap pengumpulan persyaratan didokumentasikan secara rinci dalam dokumen definisi persyaratan. Komponen penting dari dokumen ini adalah kumpulan diagram paket informasi. Ingatlah bahwa ini adalah matriks informasi yang menunjukkan metrik, dimensi bisnis, dan hierarki dalam dimensi bisnis individual. Diagram paket informasi membentuk dasar desain data logis untuk gudang data yang terdiri dari data mart. Proses desain data menghasilkan model data dimensi.

Keputusan Desain

Sebelum kita melanjutkan merancang model data dimensi, mari kita tinjau beberapa keputusan desain yang harus Anda buat:

- ❖ Memilih Proses. Memilih subjek dari paket informasi untuk rangkaian struktur logis pertama yang akan dirancang.
- ❖ Memilih Gandum. Menentukan tingkat detail data dalam struktur data. Mengidentifikasi dan Menyesuaikan Dimensi. Memilih dimensi bisnis (seperti produk, pasar, waktu, dll.) untuk dimasukkan dalam rangkaian struktur pertama dan memastikan bahwa setiap elemen data tertentu di setiap dimensi bisnis disesuaikan satu sama lain.
- ❖ Memilih Fakta. Memilih metrik atau unit pengukuran (seperti unit penjualan produk, penjualan dolar, pendapatan dolar, dll.) untuk dimasukkan dalam rangkaian struktur pertama.
- ❖ Memilih Durasi Database. Menentukan seberapa jauh Anda harus menelusuri data historis.



Gambar 10.1 Dari persyaratan hingga desain data.

Dasar-dasar Pemodelan Dimensi

Pemodelan dimensi mendapatkan namanya dari dimensi bisnis yang perlu kita masukkan ke dalam model data logis. Ini adalah teknik desain logis untuk menyusun dimensi bisnis dan metrik yang dianalisis sepanjang dimensi ini. Teknik pemodelan ini intuitif untuk tujuan itu. Model ini juga terbukti memberikan performa tinggi untuk kueri dan analisis. Diagram paket informasi multidimensi yang telah kita bahas adalah dasar dari model dimensi. Oleh karena itu, model dimensi terdiri dari struktur data spesifik yang diperlukan untuk mewakili dimensi bisnis. Struktur data ini juga berisi metrik atau fakta.

Pada Bab 5, kita membahas diagram paket informasi dengan cukup rinci. Kami secara khusus melihat diagram paket informasi untuk penjualan pembuat mobil. Silakan kembali dan tinjau Gambar 5.5 di bab itu. Apa yang kamu lihat? Di bagian bawah diagram, Anda melihat daftar pengukuran atau metrik yang ingin digunakan pembuat mobil untuk analisis. Selanjutnya, lihat judul kolom. Ini adalah dimensi bisnis yang ingin digunakan oleh pembuat mobil untuk menganalisis pengukuran atau metriknya. Di bawah setiap judul kolom, Anda melihat hierarki dimensi dan kategori dalam dimensi bisnis tersebut. Apa yang Anda lihat di bawah setiap judul kolom adalah atribut yang berkaitan dengan dimensi bisnis tersebut.

Meninjau diagram paket informasi untuk penjualan produsen mobil, kami melihat tiga jenis entitas data: (1) pengukuran atau metrik, (2) dimensi bisnis, dan (3) atribut untuk setiap dimensi bisnis. Jadi, ketika kita menyusun model dimensi untuk mewakili informasi yang terdapat dalam paket informasi penjualan produsen mobil, kita perlu menghasilkan struktur data untuk mewakili ketiga jenis entitas data ini. Mari kita bahas bagaimana kita dapat melakukan hal ini.

Pertama, mari kita bekerja dengan pengukuran atau metrik yang terlihat di bagian bawah diagram paket informasi. Ini adalah fakta untuk dianalisis. Dalam diagram penjualan pembuat mobil, faktanya adalah sebagai berikut:

- 1) Harga jual sebenarnya
- 2) MSRP
- 3) Harga pilihan
- 4) Harga penuh
- 5) Pengaya dealer
- 6) Kredit dealer
- 7) Faktur dealer
- 8) Jumlah uang muka
- 9) Hasil pabrikan
- 10) Jumlah yang dibiayai

Masing-masing item data tersebut merupakan pengukuran atau fakta. Harga jual sebenarnya adalah fakta mengenai berapa harga sebenarnya penjualan tersebut. Harga penuh adalah fakta mengenai berapa harga penuh yang berkaitan dengan penjualan. Ketika kami meninjau setiap item faktual ini, kami menemukan bahwa kami dapat mengelompokkan semua ini ke dalam satu struktur data. Dalam terminologi database relasional, Anda dapat menyebut struktur data sebagai tabel relasional. Jadi, metrik atau fakta dari diagram paket informasi tersebut akan membentuk tabel fakta. Untuk analisis penjualan produsen mobil, tabel fakta ini akan menjadi tabel fakta penjualan produsen mobil.

Lihatlah Gambar 10.2 yang menunjukkan bagaimana tabel fakta terbentuk. Tabel fakta mendapatkan namanya dari subjek yang akan dianalisis; dalam hal ini adalah penjualan pembuat mobil. Setiap item fakta atau pengukuran dimasukkan ke dalam tabel fakta sebagai atribut penjualan produsen mobil. Kami telah menentukan salah satu struktur data yang akan dimasukkan dalam model dimensi penjualan pembuat mobil dan memperoleh tabel fakta dari diagram paket informasi. Sekarang mari kita beralih ke bagian lain diagram paket informasi, dengan mengambil dimensi bisnis, satu per satu. Lihatlah dimensi bisnis produk pada Gambar 5.5.

Dimensi bisnis produk digunakan ketika kita ingin menganalisis fakta berdasarkan produk. Terkadang analisis kami dapat berupa pengelompokan berdasarkan model individual. Analisis lain mungkin berada pada tingkat yang lebih tinggi berdasarkan lini produk. Namun analisis lain mungkin berada pada tingkat yang lebih tinggi berdasarkan kategori produk. Daftar item data yang berkaitan dengan dimensi produk adalah sebagai berikut:

- 1) Nama model
- 2) Model tahun
- 3) Penataan paket
- 4) Lini produk
- 5) Kategori Produk
- 6) Warna eksterior
- 7) Warna interior

8) Model tahun pertama

Apa yang bisa kita lakukan dengan semua item data ini dalam model dimensi kita? Semua ini berhubungan dengan produk dalam beberapa hal. Oleh karena itu, kita dapat mengelompokkan semua item data ini dalam satu struktur data atau satu tabel relasional. Kita dapat menyebut tabel ini sebagai tabel dimensi produk. Item data dalam daftar di atas semuanya akan menjadi atribut dalam tabel ini.

Melihat lebih jauh ke dalam diagram paket informasi, kami mencatat dimensi bisnis lainnya yang ditampilkan sebagai judul kolom. Dalam kasus diagram paket informasi penjualan pembuat mobil, dimensi bisnis lainnya adalah dealer, demografi pelanggan, metode pembayaran, dan waktu. Sama seperti kita membentuk tabel dimensi produk, kita dapat membentuk tabel dimensi sisa dealer, demografi pelanggan, metode pembayaran, dan waktu. Item data yang ditampilkan dalam setiap kolom kemudian akan menjadi atribut untuk setiap tabel dimensi terkait.



Gambar 10.3 Pembentukan tabel dimensi pembuat mobil.

Gambar 10.3 menyatukan semua ini. Ini menunjukkan bagaimana berbagai tabel dimensi dibentuk dari diagram paket informasi. Perhatikan gambar dengan seksama dan lihat bagaimana setiap tabel dimensi terbentuk. Sejauh ini kita telah membentuk tabel fakta dan tabel dimensi. Bagaimana seharusnya tabel-tabel ini disusun dalam model dimensi? Apa hubungannya dan bagaimana kita menandai hubungan tersebut dalam model? Model dimensi terutama harus memfasilitasi pertanyaan dan analisis. Apa saja jenis kueri dan analisisnya? Ini

akan berupa kueri dan analisis yang mana metrik di dalam tabel fakta dianalisis di satu atau beberapa dimensi menggunakan atribut tabel dimensi.

Mari kita periksa kueri umum terhadap data penjualan pembuat mobil. Berapa banyak hasil penjualan yang dihasilkan Jeep Cherokee, Model Tahun 2007 dengan opsi standar, pada bulan Juli 2007 di dealer Big Sam Auto untuk pembeli yang memiliki rumah dan menyewa 3 tahun, yang dibiayai oleh Daimler-Chrysler Financing? Kami sedang menganalisis harga jual aktual, MSRP, dan harga penuh. Kami menganalisis fakta-fakta ini beserta atribut dalam berbagai tabel dimensi. Atribut dalam tabel dimensi bertindak sebagai batasan dan filter dalam kueri kami. Kami juga menemukan bahwa salah satu atau semua atribut dari setiap tabel dimensi dapat berpartisipasi dalam kueri. Selanjutnya, setiap tabel dimensi memiliki peluang yang sama untuk menjadi bagian dari kueri.

Sebelum kita memutuskan bagaimana menyusun tabel fakta dan dimensi dalam model dimensi dan menandai hubungannya, mari kita bahas apa yang perlu dicapai oleh model dimensi dan apa tujuannya. Berikut beberapa kriteria untuk menggabungkan tabel menjadi model dimensi.

- ❖ Model harus memberikan akses data terbaik.
- ❖ Keseluruhan model harus berpusat pada kueri.
- ❖ Ini harus dioptimalkan untuk kueri dan analisis.
- ❖ Model harus menunjukkan bahwa tabel dimensi berinteraksi dengan tabel fakta.
- ❖ Tabel tersebut juga harus disusun sedemikian rupa sehingga setiap dimensi dapat berinteraksi secara setara dengan tabel fakta.
- ❖ Model harus memungkinkan menelusuri atau menggulirkan hierarki dimensi.

Dengan persyaratan tersebut, kami menemukan bahwa model dimensi dengan tabel fakta di tengah dan tabel dimensi disusun mengelilingi tabel fakta memenuhi kondisi. Dalam susunan ini, masing-masing tabel dimensi mempunyai hubungan langsung dengan tabel fakta yang berada di tengahnya. Hal ini diperlukan karena setiap tabel dimensi beserta atributnya harus mempunyai peluang yang sama untuk berpartisipasi dalam query untuk menganalisis atribut pada tabel fakta. Susunan model dimensi seperti itu tampak seperti formasi bintang, dengan tabel fakta di inti bintang dan tabel dimensi di sepanjang paku bintang. Oleh karena itu model dimensi ini disebut skema STAR.

Mari kita periksa skema STAR untuk penjualan pembuat mobil seperti yang ditunjukkan pada Gambar 10.4. Tabel fakta penjualan ada di tengah. Di sekitar tabel fakta ini terdapat tabel dimensi produk, dealer, demografi pelanggan, metode pembayaran, dan waktu. Setiap tabel dimensi berhubungan dengan tabel fakta dalam hubungan satu-ke-banyak. Dengan kata lain, untuk satu baris pada tabel dimensi produk, terdapat satu atau lebih baris terkait pada tabel fakta.

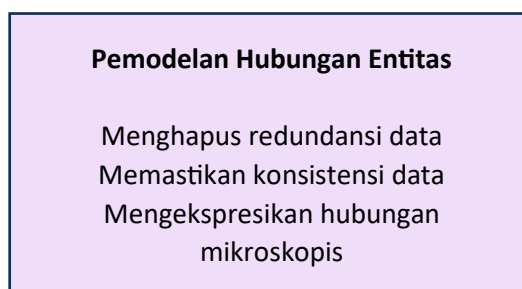
Pemodelan ER versus Pemodelan Dimensi

Kita familiar dengan pemodelan data untuk sistem operasional atau OLTP. Kami mengadopsi teknik pemodelan Entity-Relationship (E-R) untuk membuat model data untuk sistem ini.



Gambar 10-4 Skema STAR untuk penjualan pembuat mobil.

- ☞ Sistem OLTP menangkap rincian kejadian atau transaksi
- ☞ Sistem OLTP fokus pada peristiwa individual
- ☞ Sistem OLTP adalah jendela menuju transaksi tingkat mikro
- ☞ Gambaran pada tingkat detail diperlukan untuk menjalankan bisnis
- ☞ Hanya cocok untuk pertanyaan di tingkat transaksi
- ☞ Konsistensi data, non-redundansi, dan penyimpanan data yang efisien sangat penting

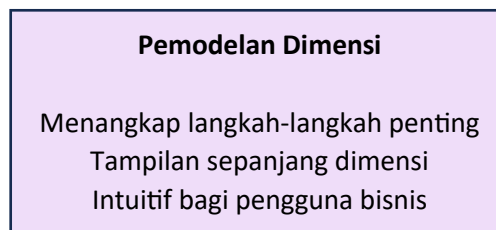


Gambar 10.5 Pemodelan E-R untuk sistem OLTP

Gambar 10.5 mencantumkan karakteristik sistem OLTP dan menunjukkan mengapa pemodelan E-R cocok untuk sistem OLTP. Sejauh ini kita telah membahas dasar-dasar model dimensi dan menemukan bahwa model ini paling cocok untuk memodelkan data untuk data warehouse. Mari kita merekapitulasi karakteristik informasi gudang data dan meninjau

bagaimana pemodelan dimensi cocok untuk tujuan ini. Pelajari fitur pemodelan dimensi yang tercantum pada Gambar 10.6.

- ☞ DW dimaksudkan untuk menjawab pertanyaan tentang keseluruhan proses
- ☞ Fokus DW adalah bagaimana manajer memandang bisnisnya
- ☞ DW mengungkap tren bisnis
- ☞ Informasi berpusat di sekitar proses bisnis
- ☞ Jawaban menunjukkan bagaimana bisnis mengukur proses tersebut
- ☞ Langkah-langkah yang harus dipelajari dalam banyak hal sepanjang beberapa dimensi bisnis



Gambar 10.6 Pemodelan dimensi untuk data warehouse.

Penggunaan Alat KASUS

Banyak alat kasus tersedia untuk pemodelan data. Di Bab 8, kami memperkenalkan alat-alat ini dan fitur-fiturnya. Anda dapat menggunakan alat ini untuk membuat skema logis dan skema fisik untuk sistem manajemen basis data target (DBMS) tertentu. Anda dapat menggunakan alat kasus untuk menentukan tabel, atribut, dan hubungan. Anda dapat menetapkan kunci utama dan menunjukkan kunci asing. Anda dapat membentuk diagram hubungan entitas. Semua ini dilakukan dengan sangat mudah menggunakan antarmuka pengguna grafis dan fasilitas drag-and-drop yang kuat. Setelah membuat model awal, Anda dapat menambahkan kolom, menghapus kolom, mengubah karakteristik kolom, membuat hubungan baru, dan membuat sejumlah revisi dengan sangat mudah.

Fungsi lain yang sangat berguna yang ditemukan dalam alat kasus ini adalah kemampuan untuk merekayasa model ke depan dan menghasilkan skema untuk sistem database target yang perlu Anda gunakan. Rekayasa ke depan mudah dilakukan dengan alat kasus ini. Untuk pemodelan data warehouse, kami tertarik pada teknik pemodelan dimensi. Sebagian besar vendor yang ada telah memperluas alat kasus pemodelan mereka untuk menyertakan pemodelan dimensi. Anda dapat membuat tabel fakta, tabel dimensi, dan menetapkan hubungan antara setiap tabel dimensi dan tabel fakta. Hasilnya adalah skema STAR untuk model Anda. Sekali lagi, Anda dapat merekayasa model STAR dimensi ke dalam skema relasional untuk sistem manajemen basis data pilihan Anda.

10.2 SKEMA BINTANG

Sekarang Anda telah mengenal skema STAR, mari kita ambil contoh sederhana dan memeriksa karakteristiknya secara detail. Membuat skema STAR adalah teknik desain data dasar untuk gudang data. Penting untuk memahami teknik ini dengan baik.

Tinjauan Skema STAR Sederhana

Kami akan mengambil skema STAR sederhana yang dirancang untuk analisis pesanan. Asumsikan ini adalah skema untuk perusahaan manufaktur dan departemen pemasaran tertarik untuk menentukan bagaimana kinerja mereka terhadap pesanan yang diterima oleh perusahaan.

Gambar 10.7 menunjukkan skema STAR sederhana ini. Ini terdiri dari tabel fakta pesanan yang ditunjukkan di tengah diagram skema. Di sekeliling tabel fakta terdapat tabel empat dimensi pelanggan, tenaga penjualan, tanggal pemesanan, dan produk. Mari kita mulai mengkaji skema STAR ini. Lihatlah struktur dari sudut pandang departemen pemasaran. Pengguna di departemen ini akan menganalisis pesanan menggunakan jumlah dolar, biaya, margin keuntungan, dan kuantitas penjualan. Informasi ini terdapat pada tabel fakta struktur. Pengguna akan menganalisis pengukuran ini dengan mengelompokkan angka-angka dalam kombinasi berdasarkan pelanggan, tenaga penjualan, tanggal, dan produk. Semua dimensi yang akan dianalisis pengguna ditemukan dalam struktur.

Struktur skema STAR adalah struktur yang mudah dipahami oleh pengguna dan dapat digunakan dengan nyaman. Struktur tersebut mencerminkan bagaimana pengguna biasanya memandang ukuran-ukuran penting mereka sepanjang dimensi bisnis mereka. Ketika Anda melihat dolar pesanan, struktur skema STAR secara intuitif menjawab pertanyaan tentang apa, kapan, oleh siapa, dan kepada siapa. Dari skema STAR, pengguna dapat dengan mudah memvisualisasikan jawaban atas pertanyaan berikut: Untuk jumlah dolar tertentu, produk apa yang dijual? Siapa pelanggannya? Penjual mana yang membawakan pesanan? Kapan pesanan dilakukan?



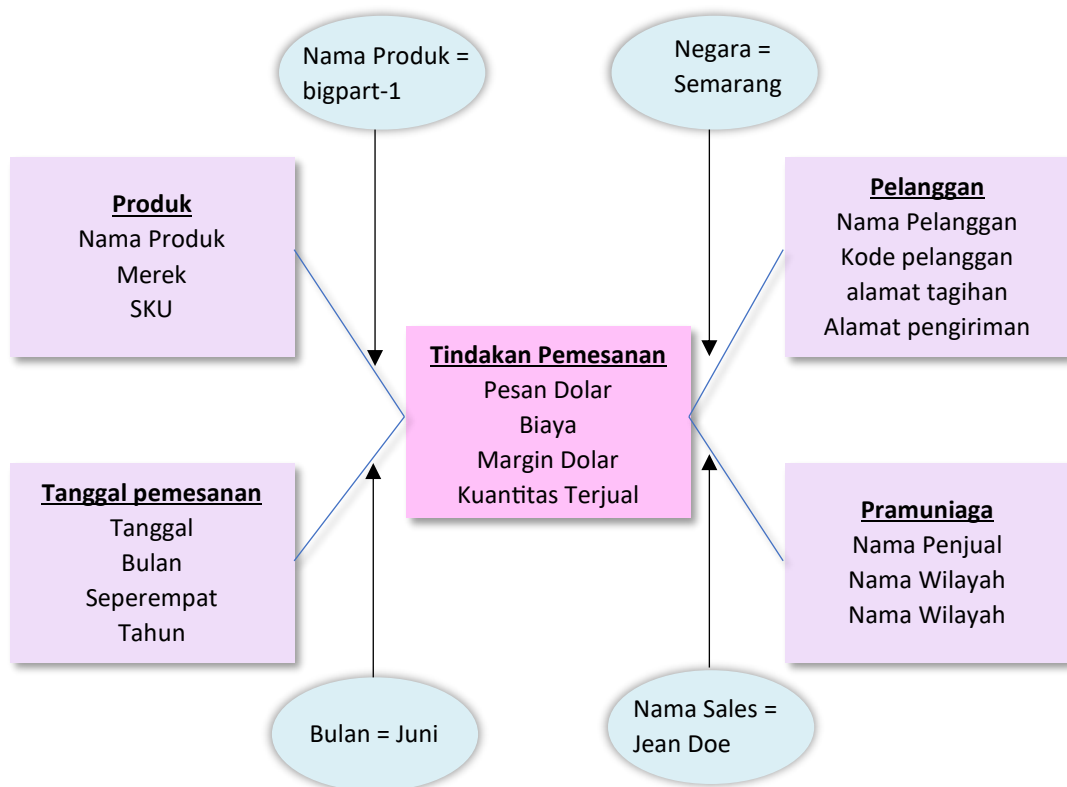
Gambar 10.7 Skema STAR sederhana untuk analisis pesanan.

Ketika kueri dibuat terhadap gudang data, hasil kueri dihasilkan dengan menggabungkan atau menggabungkan salah satu tabel dimensi dengan tabel fakta. Gabungannya adalah antara tabel fakta dan tabel dimensi individual. Hubungan baris tertentu

pada tabel fakta adalah dengan baris-baris pada setiap tabel dimensi. Hubungan individu ini jelas terlihat sebagai puncak dari skema STAR.

Ambil kueri sederhana terhadap skema STAR. Katakanlah departemen pemasaran menginginkan kuantitas yang terjual dan dolar pesanan untuk produk bigpart-1, berkaitan dengan pelanggan di negara bagian Maine, yang diperoleh oleh penjual Jane Doe, selama bulan Juni. Gambar 10.8 menunjukkan bagaimana query ini dirumuskan dari skema STAR. Batasan dan filter untuk kueri mudah dipahami dengan melihat skema STAR.

Jenis analisis yang umum adalah menelusuri angka-angka ringkasan untuk mendapatkan rincian di tingkat yang lebih rendah. Katakanlah departemen pemasaran telah memulai analisis spesifik dengan mengajukan pertanyaan berikut: Tunjukkan pada saya jumlah total penjualan produk suku cadang merek besar kepada pelanggan di wilayah timur laut untuk tahun 2008. Pada langkah analisis berikutnya, pasar- departemen sekarang ingin menelusuri hingga tingkat kuartal pada tahun 2008 untuk wilayah timur laut untuk merek produk yang sama, suku cadang besar. Selanjutnya, analisis turun ke tingkat masing-masing produk dalam merek tersebut. Terakhir, analisis ini mengarah pada tingkat rincian masing-masing negara bagian di wilayah timur laut. Pengguna dapat dengan mudah melihat semua analisis penelusuran ini dengan meninjau skema STAR. Lihat Gambar 10-9 untuk melihat bagaimana penelusuran diperoleh dari skema STAR.



Gambar 10.8 Memahami query dari skema STAR.

Di dalam Tabel Dimensi

Kita telah melihat bahwa komponen kunci dari skema STAR adalah kumpulan tabel dimensi. Tabel dimensi ini mewakili dimensi bisnis yang metriknya dianalisis. Mari kita melihat ke dalam tabel dimensi dan mempelajari karakteristiknya. Lihat Gambar 10.10 dan tinjau pengamatan berikut:

- ❖ Kunci Tabel Dimensi; Kunci utama tabel dimensi secara unik mengidentifikasi setiap baris dalam tabel.
- ❖ Mejanya Lebar; Biasanya, tabel dimensi memiliki banyak kolom atau atribut. Tidak jarang beberapa tabel dimensi memiliki lebih dari 50 atribut. Oleh karena itu, kita katakan bahwa tabel dimensinya lebar. Jika Anda menatannya sebagai tabel dengan kolom dan baris, tabel tersebut tersebar secara horizontal.
- ❖ Atribut Tekstual; Dalam tabel dimensi Anda jarang menemukan nilai numerik yang digunakan untuk perhitungan. Atribut dalam tabel dimensi berformat tekstual. Atribut-atribut ini mewakili deskripsi tekstual dari komponen-komponen dalam dimensi bisnis. Pengguna akan menyusun pertanyaan mereka menggunakan deskriptor ini.
- ❖ Atribut yang Tidak Berhubungan Langsung; Seringkali Anda akan menemukan bahwa beberapa atribut dalam tabel dimensi tidak berhubungan langsung dengan atribut lain dalam tabel. Misalnya, ukuran paket tidak berhubungan langsung dengan merek produk; namun demikian, ukuran paket dan merek produk keduanya dapat menjadi atribut tabel dimensi produk.
- ❖ Tidak Dinormalisasi; Atribut dalam tabel dimensi digunakan berulang kali dalam kueri. Atribut diambil sebagai batasan dalam kueri dan diterapkan langsung ke metrik dalam tabel fakta. Untuk kinerja kueri yang efisien, yang terbaik adalah jika kueri mengambil atribut dari tabel dimensi dan langsung menuju ke tabel fakta dan tidak melalui tabel perantara lainnya. Jika Anda menormalkan tabel dimensi, Anda akan membuat tabel perantara dan itu tidak akan efisien. Oleh karena itu, tabel dimensi diratakan, bukan dinormalisasi.
- ❖ Mengebor; Menggulung; Atribut dalam tabel dimensi memberikan kemampuan untuk mendapatkan rincian dari tingkat agregasi yang lebih tinggi ke tingkat rincian yang lebih rendah. Misalnya, tiga atribut kode pos, kota, dan negara bagian membentuk hierarki. Anda bisa mendapatkan total penjualan berdasarkan negara bagian, lalu menelusuri total penjualan berdasarkan kota, lalu berdasarkan kode pos. Dengan cara lain, pertama-tama Anda dapat memperoleh total berdasarkan kode pos, lalu menggabungkannya menjadi total berdasarkan kota dan negara bagian.
- ❖ Berbagai Hirarki; Dalam contoh dimensi pelanggan, terdapat hierarki tunggal yang naik dari pelanggan individu ke wilayah, kota, dan negara bagian. Namun tabel dimensi sering kali menyediakan beberapa hierarki, sehingga penelusuran dapat dilakukan di sepanjang beberapa hierarki tersebut. Ambil contoh tabel dimensi produk untuk department store. Dalam bisnis ini, departemen pemasaran mungkin mempunyai caranya sendiri dalam mengklasifikasikan produk ke dalam kategori produk dan

departemen produk. Di sisi lain, departemen akuntansi dapat mengelompokkan produk secara berbeda ke dalam kategori dan departemen produk. Jadi dalam hal ini, tabel dimensi produk akan memiliki atribut pemasaran– produk– kategori, pemasaran– produk– departemen, keuangan– produk– kategori, dan keuangan– produk– departemen.

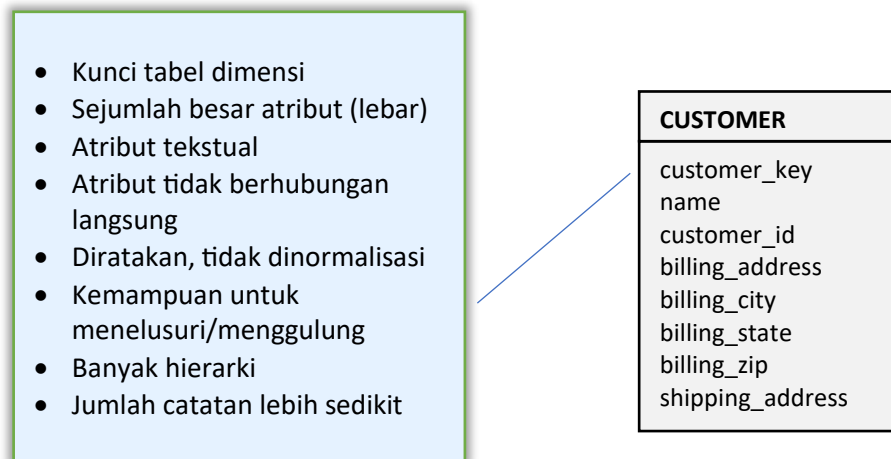
- ❖ Lebih Sedikit Catatan; Tabel dimensi biasanya memiliki lebih sedikit catatan atau baris dibandingkan tabel fakta. Tabel dimensi produk untuk produsen mobil mungkin hanya memiliki 500 baris. Di sisi lain, tabel fakta mungkin berisi jutaan baris.



Step Downdrill



Gambar 10.9 Memahami analisis penelusuran dari skema STAR.

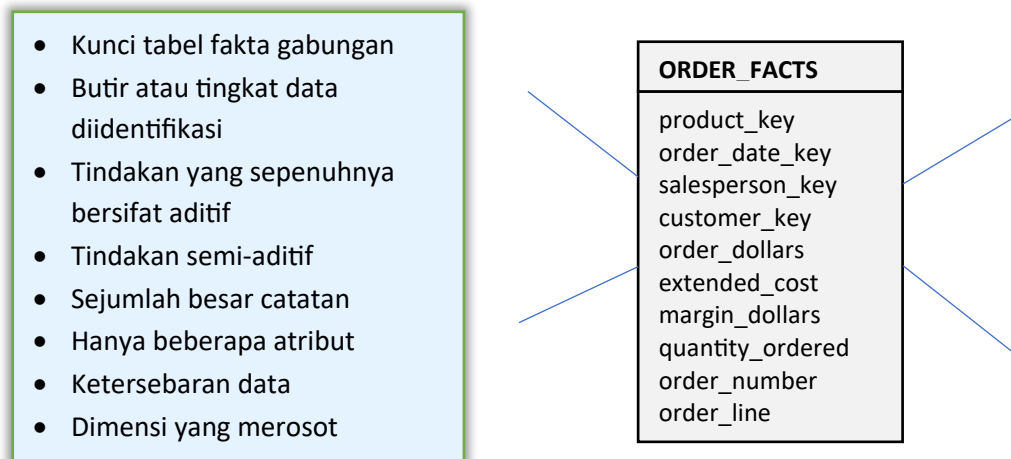


Gambar 10.10 Di dalam tabel dimensi.

Di dalam Tabel Fakta

Sekarang mari kita masuk ke tabel fakta dan memeriksa komponen-komponennya. Ingat di sinilah kami melakukan pengukuran. Kami mungkin menjaga detailnya serendah mungkin. Dalam tabel fakta department store untuk analisis penjualan, kami dapat menyimpan unit yang terjual berdasarkan transaksi individual di kasir. Beberapa tabel fakta mungkin hanya berisi data ringkasan. Ini disebut tabel fakta agregat. Gambar 10.11 mencantumkan karakteristik tabel fakta. Mari kita tinjau ciri-ciri ini:

- ❖ **Kunci Gabungan;** Sebuah baris dalam tabel fakta berkaitan dengan kombinasi baris dari semua tabel dimensi. Dalam contoh tabel fakta ini, Anda menemukan kuantitas yang dipesan sebagai atribut. Katakanlah tabel dimensinya adalah produk, waktu, pelanggan, dan perwakilan penjualan. Untuk tabel dimensi ini, asumsikan bahwa tingkat terendah dalam hierarki dimensi adalah produk individual, tanggal kalender, pelanggan tertentu, dan perwakilan penjualan tunggal. Kemudian satu baris dalam tabel fakta harus berhubungan dengan produk tertentu, tanggal kalender tertentu, pelanggan tertentu, dan perwakilan penjualan individu. Artinya baris pada tabel fakta harus diidentifikasi dengan kunci utama tabel empat dimensi tersebut. Jadi, kunci utama tabel fakta harus merupakan rangkaian kunci utama semua tabel dimensi.
- ❖ **Butir Data;** Ini merupakan karakteristik penting dari tabel fakta. Seperti yang kita ketahui, data grain adalah tingkat detail suatu pengukuran atau metrik. Dalam contoh ini, metrik berada pada tingkat detail. Kuantitas yang dipesan berkaitan dengan kuantitas produk tertentu dalam satu pesanan, pada tanggal tertentu, untuk pelanggan tertentu, dan dibeli oleh perwakilan penjualan tertentu. Jika kita menyimpan kuantitas yang dipesan sebagai kuantitas produk tertentu untuk setiap bulan, maka butir datanya berbeda dan berada pada level yang lebih tinggi.



Gambar 10.11 Di dalam tabel fakta.

- ❖ Tindakan Aditif Sepenuhnya; Mari kita lihat atribut `order_dollars`, `extended_cost`, dan `quantity_ordered`. Masing-masing hal ini berkaitan dengan produk tertentu pada tanggal tertentu untuk pelanggan tertentu yang diperoleh oleh perwakilan penjualan individu. Dalam kueri tertentu, katakanlah pengguna menginginkan total produk tertentu pada tanggal tertentu, bukan untuk pelanggan tertentu, namun untuk pelanggan di negara bagian tertentu. Kemudian kita perlu mencari semua baris dalam tabel fakta yang berkaitan dengan semua pelanggan di negara bagian tersebut dan menambahkan `order_dollars`, `extended_cost`, dan `quantity_ordered` untuk mendapatkan totalnya. Nilai-nilai atribut ini dapat diringkas dengan penjumlahan sederhana. Tindakan seperti ini dikenal sebagai tindakan yang bersifat aditif penuh. Agregasi langkah-langkah aditif penuh dilakukan dengan penambahan sederhana. Saat kita menjalankan kueri untuk menggabungkan pengukuran dalam tabel fakta, kita harus memastikan bahwa pengukuran tersebut sepenuhnya bersifat aditif. Jika tidak, angka agregat mungkin tidak menunjukkan jumlah total yang benar.
- ❖ Tindakan Semiaditif; Pertimbangkan atribut `margin_dollars` di tabel fakta. Misalnya, jika `order_dollars` adalah 120 dan `extended_cost` adalah 100, `margin_percentage` adalah 20. Ini adalah metrik terhitung yang berasal dari `order_dollars` dan `extended_cost`. Jika Anda menggabungkan angka-angka dari baris-baris dalam tabel fakta yang berkaitan dengan semua pelanggan di negara bagian tertentu, Anda tidak dapat menjumlahkan `margin_percentages` dari semua baris ini dan menghasilkan angka gabungan. Atribut turunan seperti `margin_percentage` tidak bersifat aditif. Tindakan tersebut dikenal sebagai tindakan semiaditif. Bedakan tindakan semiaditif dari tindakan aditif penuh saat Anda melakukan agregasi dalam kueri.
- ❖ Meja Dalam, Tidak Lebar; Biasanya tabel fakta berisi lebih sedikit atribut dibandingkan tabel dimensi. Biasanya, ada sekitar 10 atribut atau kurang. Namun jumlah record dalam tabel fakta sangat besar jika dibandingkan. Ambil contoh yang sangat sederhana dari 3 produk, 5 pelanggan, 30 hari, dan 10 perwakilan penjualan yang direpresentasikan sebagai baris dalam tabel dimensi. Bahkan dalam contoh ini, jumlah

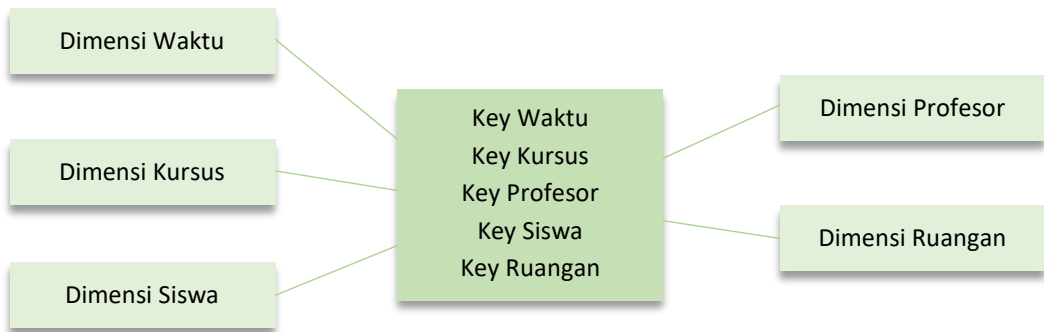
baris tabel fakta akan menjadi 4500, sangat besar dibandingkan dengan baris tabel dimensi. Jika Anda menata tabel fakta sebagai tabel dua dimensi, Anda akan melihat bahwa tabel fakta itu sempit, dengan jumlah kolom yang sedikit, namun sangat dalam, dengan jumlah baris yang banyak.

- ❖ Data Jarang; Kita telah mengatakan bahwa satu baris dalam tabel fakta berkaitan dengan produk tertentu, tanggal kalender tertentu, pelanggan tertentu, dan perwakilan penjualan individu. Dengan kata lain, untuk produk tertentu, tanggal kalender tertentu, pelanggan tertentu, dan perwakilan penjualan individu, terdapat baris yang sesuai di tabel fakta. Apa yang terjadi jika tanggal tersebut mewakili hari libur tertutup dan tidak ada pesanan yang diterima dan diproses? Baris tabel fakta untuk tanggal tersebut tidak akan memiliki nilai untuk pengukurannya. Selain itu, mungkin terdapat kombinasi atribut tabel dimensi lainnya, yang nilainya pada baris tabel fakta akan memiliki ukuran nol. Apakah kita perlu menyimpan baris-baris tersebut dengan ukuran nol di tabel fakta? Hal ini tidak perlu dilakukan. Oleh karena itu, penting untuk menyadari jenis data renggang ini dan memahami bahwa tabel fakta mungkin memiliki kesenjangan.
- ❖ Dimensi yang Merosot; Perhatikan baik-baik contoh tabel fakta. Anda menemukan atribut `order_number` dan `order_line`. Ini bukanlah ukuran, metrik, atau fakta. Lalu mengapa atribut-atribut tersebut ada di tabel fakta? Ketika Anda mengambil atribut untuk tabel dimensi dan tabel fakta dari sistem operasional, Anda akan mendapatkan beberapa elemen data dalam sistem operasional yang bukan merupakan fakta atau atribut dimensi yang ketat. Contoh atribut tersebut adalah nomor referensi seperti nomor pesanan, nomor faktur, nomor baris pesanan, dan sebagainya. Atribut-atribut ini berguna dalam beberapa jenis analisis. Misalnya, Anda mungkin mencari jumlah rata-rata produk per pesanan. Kemudian Anda harus menghubungkan produk dengan nomor pesanan untuk menghitung rata-ratanya. Atribut seperti `order_number` dan `order_line` dalam contoh disebut dimensi degenerasi dan disimpan sebagai atribut tabel fakta.

Tabel Fakta Tanpa Fakta

Selain kunci utama yang digabungkan, tabel fakta berisi fakta atau ukuran. Katakanlah kita sedang membuat tabel fakta untuk melacak kehadiran siswa. Untuk menganalisis kehadiran mahasiswa, dimensi yang mungkin adalah mahasiswa, mata kuliah, tanggal, ruangan, dan dosen. Kehadiran dapat dipengaruhi oleh salah satu dimensi ini. Saat Anda ingin menandai kehadiran yang berkaitan dengan mata kuliah, tanggal, ruangan, dan dosen tertentu, berapa ukuran yang Anda gunakan untuk mencatat acara tersebut? Pada baris tabel fakta, kehadiran akan ditandai dengan angka satu. Setiap baris tabel fakta akan memuat angka satu sebagai kehadiran. Kalau iya, buat apa repot-repot mencatat nomor satu di setiap baris tabel fakta? Tidak perlu melakukan ini. Kehadiran baris tabel fakta yang sesuai dapat mengindikasikan kehadiran. Situasi seperti ini muncul ketika tabel fakta merepresentasikan peristiwa. Tabel fakta seperti itu sebenarnya tidak perlu memuat fakta. Itu adalah tabel fakta tanpa fakta. Gambar 10.12 menunjukkan tabel fakta tanpa fakta.

Ukuran atau fakta direpresentasikan dalam tabel fakta. Namun, terdapat peristiwa atau liputan bisnis yang dapat disajikan dalam tabel fakta, meskipun tidak ada ukuran atau fakta yang dikaitkan dengan hal tersebut.



Gambar 10.12 Tabel fakta tanpa fakta.

Granularitas Data

Saat ini, kita tahu bahwa granularitas mewakili tingkat detail dalam tabel fakta. Jika tabel fakta berada pada tingkat yang paling rendah, maka fakta atau metrik tersebut berada pada tingkat yang paling rendah yang dapat diambil dari sistem operasional. Apa keuntungan menjaga tabel fakta tetap pada tingkat yang paling rendah? Apa trade-offnya?

Ketika Anda menjaga tabel fakta pada tingkat terendah, pengguna dapat menelusuri hingga tingkat detail terendah dari gudang data tanpa perlu pergi ke sistem operasional itu sendiri. Tabel fakta tingkat dasar harus berada pada tingkat terendah alami dari semua dimensi terkait. Dengan melakukan ini, kueri untuk menelusuri dan menggulung dapat dilakukan secara efisien.

Lalu berapa tingkat terendah alami dari dimensi yang bersangkutan? Dalam contoh dimensi produk, tanggal, pelanggan, dan perwakilan penjualan, tingkat terendah alaminya masing-masing adalah produk individual, tanggal individual tertentu, pelanggan individual, dan perwakilan penjualan individual. Jadi, dalam hal ini, satu baris dalam tabel fakta harus berisi pengukuran pada tingkat terendah untuk masing-masing produk, dipesan pada tanggal tertentu, berkaitan dengan pelanggan individual, dan dibeli oleh perwakilan penjualan individual.

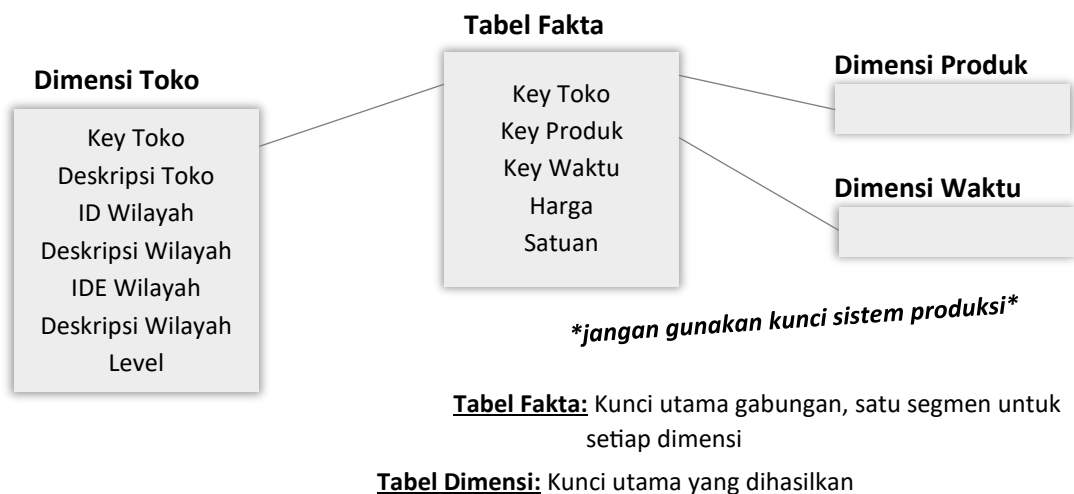
Misalkan kita ingin menambahkan atribut baru distrik pada dimensi perwakilan penjualan. Perubahan ini tidak menjamin perubahan apa pun pada baris tabel fakta karena baris tersebut sudah berada pada tingkat terendah dari masing-masing perwakilan penjualan. Ini adalah perubahan yang “baik” karena semua kueri lama akan terus berjalan tanpa perubahan apa pun. Demikian pula, mari kita asumsikan kita ingin menambahkan dimensi promosi yang baru. Sekarang Anda harus menyusun ulang baris tabel fakta untuk memasukkan dimensi promosi. Namun, butir tabel fakta akan berada pada level terendah.

Bahkan di sini, kueri lama akan tetap berjalan tanpa perubahan apa pun. Ini juga merupakan perubahan yang “anggun”. Tabel fakta di bagian paling bawah memfasilitasi perluasan yang “anggun”. Namun kita harus membayar harga dalam hal penyimpanan dan pemeliharaan tabel fakta hingga tingkat terendah. Butir terendah berarti sejumlah besar baris tabel fakta. Namun dalam praktiknya, kami membuat tabel fakta agregat untuk mendukung

kueri yang mencari nomor ringkasan. Ada dua keuntungan lagi dari tabel fakta granular. Tabel fakta granular berfungsi sebagai tujuan alami untuk data operasional saat ini yang mungkin sering diambil dari sistem operasional. Lebih jauh lagi, aplikasi penambangan data yang lebih baru memerlukan detail yang paling detail. Gudang data memasukkan data ke dalam aplikasi penambangan data.

10.3 KUNCI SKEMA BINTANG

Gambar 10.13 mengilustrasikan bagaimana kunci dibentuk untuk tabel dimensi dan fakta.



Gambar 10.13 Kunci skema STAR.

Kunci Utama

Setiap baris dalam tabel dimensi diidentifikasi oleh nilai unik dari suatu atribut yang ditunjuk sebagai kunci utama dimensi tersebut. Dalam tabel dimensi produk, kunci utama mengidentifikasi setiap produk secara unik. Dalam tabel dimensi pelanggan, nomor pelanggan mengidentifikasi setiap pelanggan secara unik. Demikian pula, dalam tabel dimensi perwakilan penjualan, nomor jaminan sosial perwakilan penjualan mengidentifikasi setiap perwakilan penjualan.

Kami telah memilih ini sebagai kunci kandidat untuk tabel dimensi. Sekarang mari kita pertimbangkan beberapa implikasi dari kunci kandidat ini. Misalkan kode produk dalam sistem operasional adalah kode 8 posisi, dua posisi di antaranya menunjukkan kode gudang tempat produk biasanya disimpan, dan dua posisi lainnya menunjukkan kategori produk. Mari kita lihat apa yang terjadi jika kita menggunakan kode produk sistem operasional sebagai kunci utama untuk tabel dimensi produk.

Gudang data berisi data historis. Asumsikan kode produk diubah pada pertengahan tahun, karena produk tersebut sekarang disimpan di gudang perusahaan yang berbeda. Jadi kita harus mengubah kode produk di data warehouse. Jika kode produk adalah kunci utama tabel dimensi produk, maka data baru untuk produk yang sama akan berada di gudang data

dengan nilai kunci berbeda. Hal ini dapat menimbulkan masalah jika kita perlu menggabungkan data sebelum perubahan dengan data setelah perubahan kode produk. Apa sebenarnya yang menyebabkan masalah ini? Masalahnya adalah hasil dari keputusan kami untuk menggunakan kunci sistem operasional sebagai kunci tabel dimensi.

Kunci Pengganti

Bagaimana kita mengatasi masalah yang dihadapi pada bagian sebelumnya? Bisakah kita menggunakan kunci sistem produksi sebagai kunci utama untuk tabel dimensi? Jika tidak, apa kunci kandidat lainnya?

Ada dua prinsip umum yang diterapkan ketika memilih kunci utama untuk tabel dimensi. Prinsip pertama berasal dari masalah yang timbul ketika produk mulai disimpan di gudang yang berbeda. Dengan kata lain, kunci produk dalam sistem operasional memiliki arti bawaan. Beberapa posisi dalam kunci produk sistem operasional menunjukkan gudang dan beberapa posisi lain dalam kunci menunjukkan kategori produk. Ini adalah makna bawaan dalam kuncinya. Prinsip pertama yang harus diikuti adalah: hindari makna bawaan pada kunci utama tabel dimensi.

Di beberapa perusahaan, beberapa pelanggan tidak lagi terdaftar di perusahaan tersebut. Mereka bisa saja meninggalkan perusahaannya masing-masing bertahun-tahun yang lalu. Ada kemungkinan bahwa nomor pelanggan dari pelanggan yang dihentikan tersebut dialihkan ke pelanggan baru. Sekarang, katakanlah kita telah menggunakan kunci pelanggan sistem operasional sebagai kunci utama untuk tabel dimensi pelanggan. Kita akan mendapat masalah karena nomor pelanggan yang sama bisa saja berhubungan dengan data pelanggan baru dan juga data pelanggan pensiunan. Data pelanggan yang sudah pensiun masih dapat digunakan untuk agregasi dan perbandingan menurut kota dan negara bagian. Oleh karena itu, prinsip kedua adalah: jangan gunakan kunci sistem produksi sebagai kunci utama untuk tabel dimensi. Lalu apa yang harus kita gunakan sebagai kunci utama untuk tabel dimensi? Jawabannya adalah dengan menggunakan kunci pengganti.

Kunci pengganti hanyalah nomor urut yang dihasilkan sistem. Mereka tidak memiliki makna bawaan apa pun. Tentu saja, kunci pengganti akan dipetakan ke kunci sistem produksi. Meski begitu, keduanya berbeda. Praktik umumnya adalah menjaga kunci sistem operasional sebagai atribut tambahan dalam tabel dimensi. Lihat kembali Gambar 10-13. STORE KEY adalah kunci utama pengganti untuk tabel dimensi penyimpanan. Kunci utama sistem operasional untuk tabel referensi penyimpanan dapat disimpan hanya sebagai atribut non-kunci dalam tabel dimensi penyimpanan.

Kunci Asing

Setiap tabel dimensi berada dalam hubungan satu-ke-banyak dengan tabel fakta pusat. Jadi kunci utama setiap tabel dimensi haruslah merupakan kunci asing pada tabel fakta. Jika terdapat empat tabel dimensi produk, tanggal, pelanggan, dan perwakilan penjualan, maka kunci utama dari keempat tabel tersebut harus ada dalam tabel fakta pesanan sebagai kunci asing.

Mari kita periksa kembali kunci utama tabel fakta. Ada tiga opsi:

1. Satu kunci primer majemuk yang panjangnya merupakan panjang total kunci dari tabel dimensi individual. Dalam opsi ini, selain kunci utama gabungan, kunci asing juga harus disimpan dalam tabel fakta sebagai atribut tambahan. Opsi ini meningkatkan ukuran tabel fakta.
2. Kunci utama gabungan yang merupakan gabungan semua kunci utama tabel dimensi. Di sini Anda tidak perlu menyimpan kunci utama tabel dimensi sebagai atribut tambahan untuk dijadikan kunci asing. Masing-masing bagian dari kunci primer itu sendiri akan berfungsi sebagai kunci asing.
3. Kunci utama yang dihasilkan tidak bergantung pada kunci tabel dimensi. Selain kunci utama yang dihasilkan, kunci asing juga harus disimpan dalam tabel fakta sebagai atribut tambahan. Opsi ini juga meningkatkan ukuran tabel fakta.

Dalam praktiknya, opsi (2) digunakan di sebagian besar tabel fakta. Opsi ini memungkinkan Anda dengan mudah menghubungkan baris tabel fakta dengan baris tabel dimensi.

10.4 KEUNGGULAN SKEMA BINTANG

Ketika Anda melihat skema STAR, Anda menemukan bahwa skema ini hanyalah sebuah model relasional dengan hubungan satu ke banyak antara setiap tabel dimensi dan tabel fakta. Apa istimewanya susunan skema STAR? Mengapa dinyatakan sangat cocok untuk data warehouse? Apa alasan penggunaannya secara luas dan keberhasilannya dalam menyediakan pengoptimalan untuk pemrosesan kueri?

Meskipun skema STAR merupakan model relasional, namun skema ini bukanlah model yang dinormalisasi. Tabel dimensi sengaja didnormalisasi. Ini adalah perbedaan mendasar antara skema STAR dan skema relasional untuk sistem OLTP. Sebelum kita membahas beberapa keuntungan signifikan skema STAR, kita perlu menyadari bahwa kepatuhan ketat terhadap pengaturan ini tidak selalu merupakan pilihan terbaik. Misalnya, jika pelanggan adalah salah satu dimensi dan jika perusahaan mempunyai jumlah pelanggan yang sangat besar, tabel dimensi pelanggan yang didnormalisasi tidak diinginkan. Tabel dimensi yang besar dapat meningkatkan ukuran tabel fakta. Namun, kelebihanannya jauh lebih besar dibandingkan kekurangannya. Jadi, mari kita bahas keuntungan skema STAR.

Mudah Dipahami Pengguna

Pengguna sistem OLTP berinteraksi dengan aplikasi melalui layar GUI yang telah ditentukan sebelumnya atau templat kueri yang telah ditentukan sebelumnya. Praktisnya, pengguna tidak perlu memahami struktur data di balik layar. Struktur data dan skema database tetap berada di ranah profesional TI.

Pengguna sistem pendukung keputusan seperti gudang data berbeda-beda. Di sini pengguna sendiri yang akan merumuskan pertanyaan. Saat mereka berinteraksi dengan gudang data melalui alat kueri pihak ketiga, pengguna harus mengetahui apa yang harus diminta. Mereka harus memahami data apa yang tersedia bagi mereka di gudang data. Mereka harus memiliki pemahaman tentang struktur data dan bagaimana berbagai bagian dihubungkan satu sama lain dalam skema keseluruhan. Mereka harus memahami hubungannya tanpa kesulitan.

Skema STAR mencerminkan dengan tepat bagaimana pengguna berpikir dan membutuhkan data untuk kueri dan analisis. Mereka berpikir dalam kaitannya dengan metrik bisnis yang signifikan. Tabel fakta berisi metrik. Pengguna memikirkan dimensi bisnis untuk menganalisis metrik. Tabel dimensi berisi atribut yang biasanya digunakan pengguna untuk melakukan kueri dan analisis. Saat Anda menjelaskan kepada pengguna bahwa unit produk A disimpan dalam tabel fakta dan menunjukkan hubungan bagian data ini dengan setiap tabel dimensi, pengguna dengan mudah memahami hubungannya. Hal ini karena skema STAR mendefinisikan jalur gabungan dengan cara yang persis sama dengan cara pengguna biasanya memvisualisasikan hubungan. Skema STAR dipahami secara intuitif oleh pengguna.

Cobalah untuk memandu pengguna melalui skema relasional sistem OLTP. Agar mereka dapat memahami hubungannya, Anda harus membawanya melalui labirin tabel yang dinormalisasi, terkadang melewati beberapa tabel, satu per satu, untuk mendapatkan kumpulan hasil terkecil sekalipun. Skema STAR jelas merupakan pemenang karena kesederhanaannya. Pengguna memahami struktur dan koneksinya dengan sangat mudah.

Skema STAR memiliki keuntungan yang pasti setelah implementasi. Namun, keuntungan yang ada bahkan pada tahap pengembangan tidak dapat diabaikan. Karena pengguna memahami skema STAR dengan sangat baik, maka mudah untuk menggunakannya sebagai sarana komunikasi dengan pengguna selama pengembangan gudang data.

Mengoptimalkan Navigasi

Dalam skema database, apa tujuan dari hubungan atau koneksi antar entitas data? Hubungan tersebut digunakan untuk berpindah dari satu tabel ke tabel lainnya untuk mendapatkan informasi yang Anda cari. Hubungan tersebut memberikan kemampuan untuk menavigasi database. Anda melompat dari meja ke meja menggunakan jalur gabungan. Jika jalur gabungan banyak dan berbelit-belit, navigasi Anda melalui database menjadi sulit dan lambat. Di sisi lain, jika jalur bergabungnya sederhana dan mudah, navigasi Anda akan dioptimalkan dan menjadi lebih cepat.

Keuntungan utama skema STAR adalah mengoptimalkan navigasi melalui database. Bahkan ketika Anda mencari hasil kueri yang tampak rumit, navigasinya tetap sederhana dan mudah. Mari kita lihat contohnya dan pahami cara kerjanya. Lihatlah Gambar 10-14 yang menunjukkan skema STAR untuk menganalisis kerusakan pada mobil. Asumsikan Anda adalah manajer layanan di dealer mobil yang menjual mobil GM. Anda melihat tingginya insiden cat putih terkelupas pada Corvette pada bulan Januari 2009. Anda memerlukan alat untuk menganalisis cacat tersebut, menentukan penyebab utamanya, dan menyelesaikan masalahnya.

Dalam skema STAR, jumlah cacat disimpan sebagai metrik di tengah-tengah tabel fakta cacat. Dimensi waktu berisi tahun model. Dimensi komponen memiliki informasi bagian; misalnya cat putih mutiara. Dimensi masalah membawa jenis-jenis masalah; misalnya cat terkelupas. Dimensi produk berisi merek, model, dan paket trim mobil. Dimensi pemasok berisi data tentang pemasok suku cadang.



Gambar 10.14 Skema STAR mengoptimalkan navigasi.

Sekarang lihat betapa mudahnya menentukan pemasok yang menyebabkan cat terkelupas pada Corvette putih mutiara. Lihatlah empat anak panah yang menunjuk ke tabel fakta dari tabel empat dimensi. Panah ini menunjukkan bagaimana Anda akan menavigasi ke baris dalam tabel fakta dengan mengisolasi Corvette dari dimensi produk, cat terkelupas dari dimensi masalah, cat putih mutiara dari dimensi komponen, dan Januari 2009 dari dimensi waktu. Dari tabel fakta, navigasi langsung menuju ke dimensi pemasok untuk mengisolasi pemasok penyebab masalah.

Paling Cocok untuk Pemrosesan Kueri

Kami telah menyebutkan beberapa kali bahwa skema STAR adalah struktur yang berpusat pada kueri. Artinya skema STAR paling cocok untuk pemrosesan kueri. Mari kita lihat kebenarannya. Mari kita bentuk query sederhana pada skema STAR untuk analisis pesanan yang ditunjukkan pada Gambar 10.7. Berapa total biaya tambahan produk A yang dijual kepada pelanggan di San Francisco selama bulan Januari 2009? Ini adalah kueri tiga dimensi. Apa karakteristik struktur data atau skema yang paling cocok untuk memproses kueri ini? Hasil akhir, yaitu total biaya tambahan, akan diperoleh dari baris-baris dalam tabel fakta. Tapi dari baris mana? Jawabannya adalah baris-baris yang berkaitan dengan produk A, berkaitan dengan pelanggan di San Francisco, dan berkaitan dengan Januari 2009.

Mari kita lihat bagaimana query akan diproses. Pertama, pilih baris dari tabel dimensi pelanggan dengan kota San Francisco. Kemudian, dari tabel fakta, pilih hanya baris yang terkait dengan baris dimensi pelanggan tersebut. Ini adalah kumpulan baris hasil pertama dari tabel fakta. Selanjutnya, pilih baris dalam tabel Dimensi waktu dengan bulan Januari 2009. Pilih dari kumpulan hasil pertama hanya baris yang terkait dengan baris dimensi waktu tersebut. Ini sekarang merupakan kumpulan hasil kedua dari baris tabel fakta. Pindah ke

dimensi produk berikutnya. Pilih baris dalam tabel dimensi produk yang produknya adalah produk A. Pilih dari hasil kedua hanya baris yang terkait dengan baris dimensi produk yang dipilih. Anda sekarang memiliki hasil akhir dari baris tabel fakta. Tambahkan biaya tambahan untuk mendapatkan totalnya.

Terlepas dari jumlah dimensi yang berpartisipasi dalam kueri dan kompleksitas kueri, setiap kueri dieksekusi terlebih dahulu dengan memilih baris dari tabel dimensi menggunakan filter berdasarkan parameter kueri dan kemudian menemukan tabel fakta yang sesuai. baris. Hal ini dimungkinkan karena jalur gabungan yang sederhana dan mudah dipahami serta pengaturan skema STAR. Tidak ada labirin perantara yang harus dinavigasi untuk mencapai tabel fakta dari tabel dimensi.

Aspek penting lainnya dari kueri gudang data adalah kemampuan untuk menelusuri atau menggulung. Mari kita segera menjalankan skenario penelusuran. Katakanlah kita telah menanyakan dan mendapatkan total biaya tambahan untuk semua pelanggan di negara bagian California. Hasilnya berasal dari kumpulan baris tabel fakta. Kemudian kami ingin menelusuri dan melihat hasilnya berdasarkan rentang kode pos. Hal ini diperoleh dengan membuat pilihan lebih lanjut dari baris tabel fakta yang dipilih terkait dengan rentang kode pos yang dipilih. Drill down merupakan proses seleksi lebih lanjut terhadap baris-baris tabel fakta. Sebaliknya, rolling up adalah proses memperluas pilihan baris tabel fakta.

STARjoin dan STARindex

Skema STAR memungkinkan perangkat lunak pemroses kueri menggunakan rencana eksekusi yang lebih baik. Ini memungkinkan skema kinerja tertentu untuk diterapkan pada kueri. Susunan skema STAR sangat cocok untuk teknik kinerja khusus seperti STARjoin dan STARindex. STARjoin adalah gabungan multitabel berkecepatan tinggi, single-pass, dapat diparalelkan. Itu dapat menggabungkan lebih dari dua tabel dalam satu operasi. Skema khusus ini meningkatkan kinerja kueri.

STARindex adalah indeks khusus untuk mempercepat kinerja gabungan. Ini adalah indeks yang dibuat pada satu atau lebih kunci asing dari tabel fakta. Indeks ini mempercepat penggabungan antara tabel dimensi dan tabel fakta. Bab 18 membahas desain fisik dan metode pengindeksan untuk gudang data.

Skema Bintang: Contoh

Pada bagian ini kami memberikan empat contoh skema STAR yang berkaitan dengan empat jenis bisnis berbeda. Dalam setiap kasus, perhatikan hal berikut dengan cermat dan pahami skema STAR:

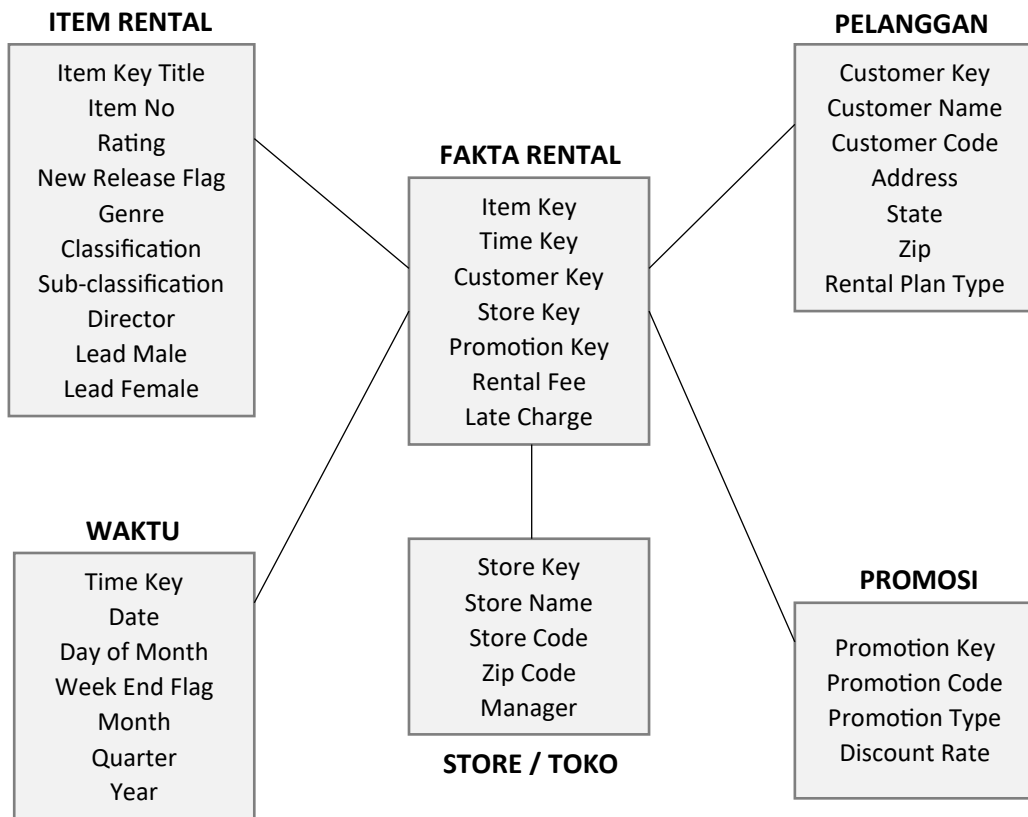
- Metrik atau fakta yang dianalisis
- Dimensi bisnis yang digunakan untuk analisis
- Hierarki tersedia dalam setiap dimensi untuk menelusuri dan menggulung
- Kunci utama dan kunci asing

☞ Penyewaan Video Lihat Gambar 10-15.

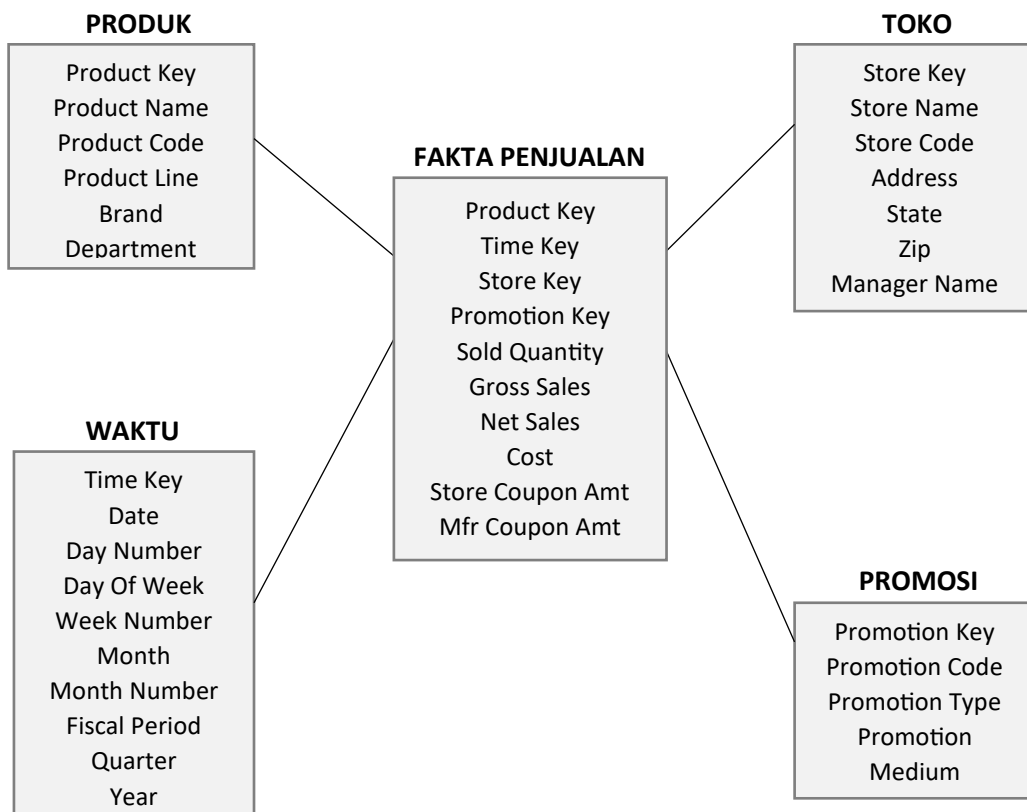
☞ Supermarket Lihat Gambar 10-16.

☞ Layanan Telepon Nirkabel Lihat Gambar 10-17.

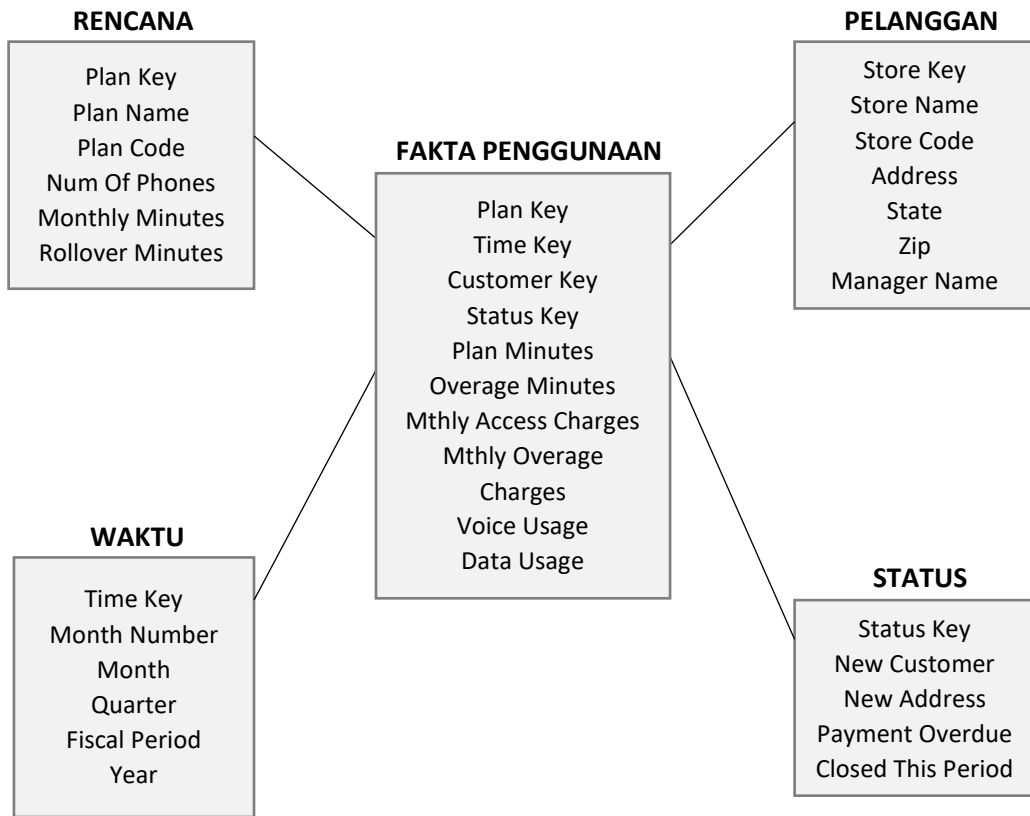
☞ Perusahaan Lelang Lihat Gambar 10-18.



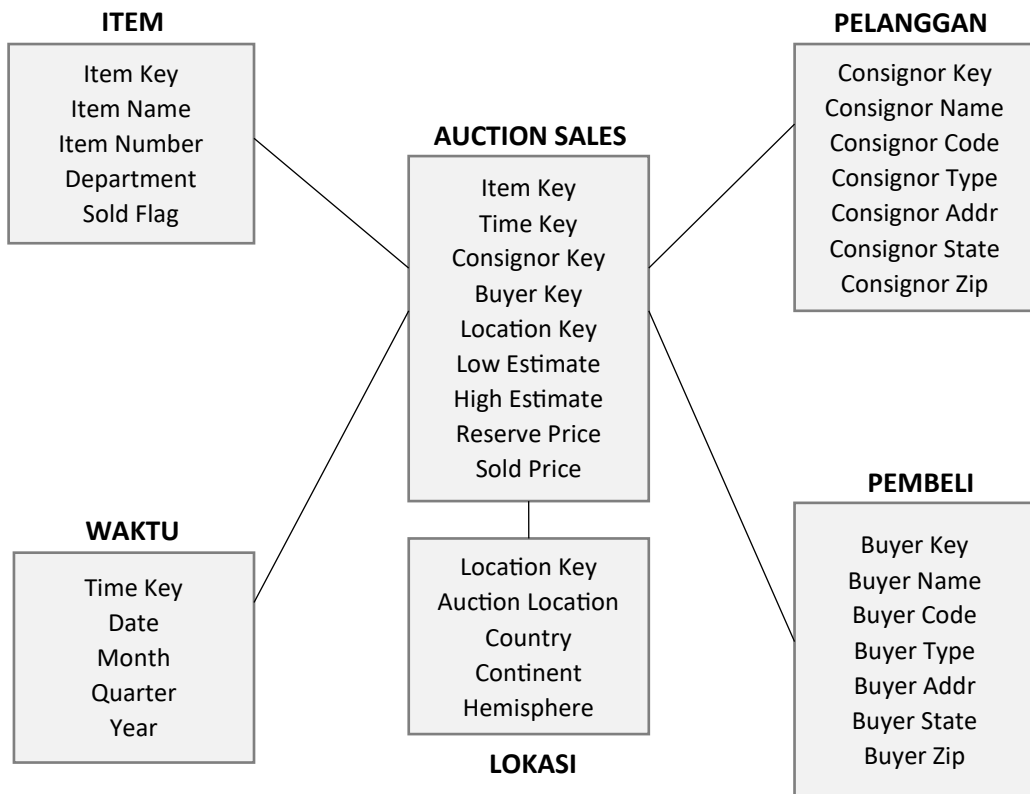
Gambar 10.15 Contoh skema STAR: rental video.



Gambar 10.16 Contoh skema STAR: supermarket.



Gambar 10.17 Contoh skema STAR: layanan telepon nirkabel.



Gambar 10-18 Contoh skema STAR: perusahaan lelang.

RINGKASAN BAB

- ❖ Komponen model dimensi berasal dari paket informasi dalam definisi persyaratan.
- ❖ Teknik pemodelan hubungan entitas tidak cocok untuk gudang data; teknik pemodelan dimensi sudah tepat.
- ❖ Skema STAR yang digunakan untuk desain data adalah model relasional yang terdiri dari tabel fakta dan dimensi.
- ❖ Tabel fakta berisi metrik atau pengukuran bisnis; tabel dimensi berisi dimensi bisnis. Hierarki dalam setiap tabel dimensi digunakan untuk menelusuri tingkat data yang lebih rendah.
- ❖ Keuntungan skema STAR adalah mudah dipahami oleh pengguna, mengoptimalkan navigasi, paling cocok untuk pemrosesan kueri, dan memungkinkan skema kinerja tertentu.

LATIHAN SOAL

1. Diskusikan masalah desain utama yang perlu diatasi sebelum melanjutkan desain data.
2. Mengapa teknik pemodelan hubungan entitas tidak cocok untuk data warehouse? Apa perbedaan pemodelan dimensi?
3. Apa yang dimaksud dengan skema STAR? Apa saja tabel komponennya?
4. Tabel dimensi lebar; tabel faktanya mendalam. Menjelaskan.
5. Hierarki dan kategori apa yang dapat diterapkan pada tabel dimensi?
6. Bedakan antara tindakan yang sepenuhnya aditif dan semi-aditif.
7. Jelaskan sifat renggang data pada tabel fakta.
8. Jelaskan komposisi kunci utama untuk tabel dimensi dan fakta.
9. Diskusikan granularitas data di gudang data.
10. Sebutkan tiga keuntungan skema STAR. Dapatkah Anda memikirkan kelemahan skema STAR?

BAB 11

PEMODELAN DIMENSI

TUJUAN BAB

- Diskusikan dan pahami dengan baik dimensi yang berubah secara perlahan
- Memahami dimensi besar dan cara menghadapinya
- Periksa skema kepingan salju secara detail
- Pelajari tentang tabel agregat dan tentukan kapan menggunakannya
- Survei lengkap keluarga STARS dan aplikasinya

Dari bab sebelumnya, Anda telah mempelajari dasar-dasar pemodelan dimensi. Anda tahu bahwa skema STAR terdiri dari tabel fakta di tengah yang dikelilingi oleh tabel dimensi. Meskipun ini merupakan representasi visual yang baik, ini masih merupakan model relasional di mana setiap tabel dimensi berada dalam hubungan induk-anak dengan tabel fakta. Oleh karena itu, kunci utama setiap tabel dimensi adalah kunci asing dalam tabel fakta.

Anda juga telah memahami sifat atribut dalam tabel fakta dan tabel dimensi. Anda telah memahami keuntungan skema STAR dalam sistem pendukung keputusan. Skema STAR mudah dipahami pengguna; ini mengoptimalkan navigasi melalui konten gudang data dan paling cocok untuk lingkungan yang berpusat pada kueri.

Studi kita tentang pemodelan dimensi tidak akan lengkap sampai kita mempertimbangkan beberapa topik lagi. Dalam skema STAR, tabel dimensi memungkinkan analisis dalam berbagai cara. Kita perlu menjelajahi tabel dimensi lebih detail. Bagaimana dengan meringkas metrik dan menyimpan angka agregat dalam tabel fakta tambahan? Berapa banyak agregasi yang telah dihitung sebelumnya yang diperlukan? Skema STAR adalah desain yang didenormalisasi. Apakah hal ini mengakibatkan terlalu banyak redundansi dan inefisiensi? Jika ya, apakah ada pendekatan alternatif?

Sekarang mari kita melampaui dasar-dasar pemodelan dimensi dan mempertimbangkan fitur dan masalah tambahan. Mari kita bahas topik lanjutan terkait dan memperluas studi kita lebih jauh.

11.1 PEMBARUAN TABEL DIMENSI

Kembali ke Gambar 10.4 pada bab sebelumnya, Anda melihat skema STAR untuk penjualan produsen mobil. Tabel fakta Penjualan Otomatis berisi pengukuran atau metrik seperti Harga Jual Aktual, Harga Opsi, dan sebagainya. Seiring berjalannya waktu, apa yang terjadi pada tabel fakta? Setiap hari seiring dengan semakin banyaknya penjualan yang terjadi, semakin banyak baris yang ditambahkan ke tabel fakta. Seiring berjalannya waktu jumlah baris pada tabel fakta terus bertambah. Sangat jarang baris dalam tabel fakta diperbarui dengan perubahan. Bahkan ketika ada penyesuaian pada angka-angka sebelumnya, angka-angka tersebut juga diproses sebagai baris penyesuaian tambahan dan ditambahkan ke tabel fakta.

Sekarang perhatikan tabel dimensi. Dibandingkan dengan tabel fakta, tabel dimensi lebih stabil dan tidak terlalu fluktuatif. Namun, berbeda dengan tabel fakta yang berubah melalui penambahan jumlah baris, tabel dimensi tidak hanya berubah melalui penambahan jumlah baris, namun juga melalui perubahan pada atribut itu sendiri. Lihatlah tabel dimensi produk. Setiap tahun, baris ditambahkan seiring tersedianya model baru. Namun bagaimana dengan atribut dalam tabel dimensi produk? Jika produk tertentu dipindahkan ke kategori produk lain, maka nilai terkait dalam tabel dimensi produk harus diubah. Mari kita periksa jenis perubahan yang mempengaruhi tabel dimensi dan diskusikan cara menangani jenis ini.

Perlahan Mengubah Dimensi

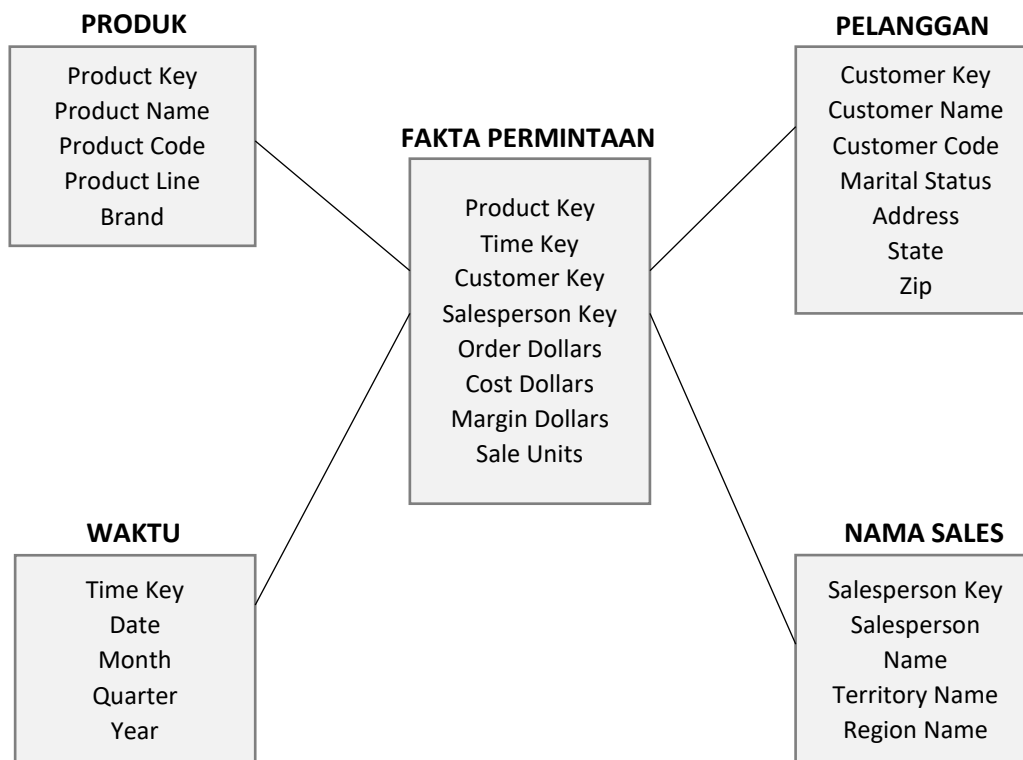
Pada contoh di atas, kami menyebutkan perubahan tabel dimensi produk karena kategori produk suatu produk diubah. Pertimbangkan tabel dimensi demografi pelanggan. Apa jadinya jika status pelanggan berubah dari rumah sewa menjadi rumah milik sendiri? Baris terkait dalam tabel dimensi tersebut harus diubah. Selanjutnya lihat tabel dimensi metode pembayaran. Jika jenis pembiayaan salah satu metode pembayaran berubah, perubahan ini harus tercermin dalam tabel dimensi metode pembayaran.

Dari pertimbangan perubahan tabel dimensi, kita dapat memperoleh prinsip-prinsip berikut:

- ✘ Sebagian besar dimensi umumnya konstan sepanjang waktu.
- ✘ Banyak dimensi, meskipun tidak konstan seiring berjalannya waktu, berubah secara perlahan.
- ✘ Kunci produk dari rekaman sumber tidak berubah.
- ✘ Deskripsi dan atribut lainnya berubah secara perlahan seiring berjalannya waktu.
- ✘ Dalam sistem OLTP sumber, nilai-nilai baru menimpa nilai-nilai lama.
- ✘ Menimpa atribut tabel dimensi tidak selalu merupakan pilihan yang tepat di gudang data.
- ✘ Cara perubahan dilakukan pada tabel dimensi bergantung pada jenis perubahan dan informasi apa yang harus disimpan dalam gudang data.

Teknik untuk membuat perubahan terbagi dalam tiga jenis berbeda. Praktisi data warehousing telah menemukan teknik berbeda untuk menerapkan perubahan. Mereka juga memberi nama pada ketiga jenis perubahan tabel dimensi ini. Ya, tebakanmu benar. Nama yang diberikan adalah perubahan tipe 1, perubahan tipe 2, dan perubahan tipe 3.

Kita akan mempelajari ketiga jenis ini dengan menggunakan skema STAR sederhana untuk melacak pesanan distributor produk industri, seperti yang ditunjukkan pada Gambar 11.1. Skema STAR ini terdiri dari tabel fakta dan tabel empat dimensi. Mari kita asumsikan beberapa perubahan pada dimensi ini dan meninjau teknik untuk menerapkan perubahan pada tabel dimensi.



Gambar 11.1 Skema STAR untuk pelacakan pesanan.

Perubahan Tipe 1: Koreksi Kesalahan

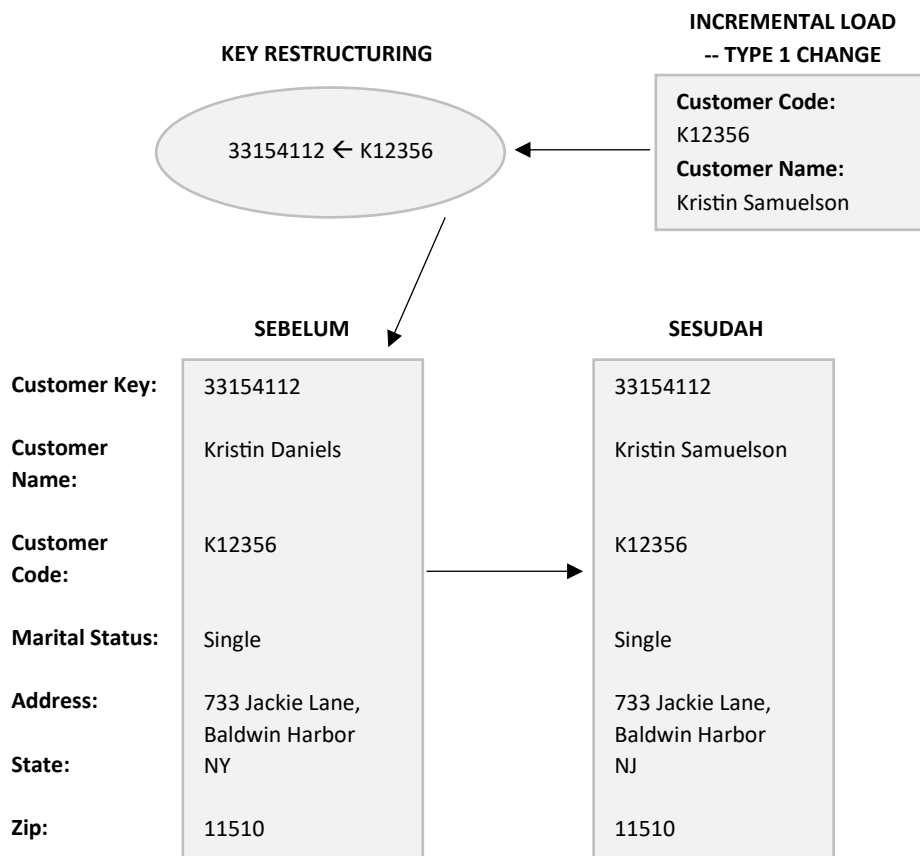
Sifat Perubahan Tipe 1 Perubahan ini biasanya berhubungan dengan koreksi kesalahan pada sistem sumber. Misalnya, kesalahan ejaan pada nama pelanggan dikoreksi menjadi Michael Romano dari entri Michel Romano yang salah. Juga, misalkan nama pelanggan untuk pelanggan lain diubah dari Kristin Daniels menjadi Kristin Samuelson, dan status perkawinan diubah dari lajang untuk menikah.

Pertimbangkan perubahan nama pelanggan dalam kedua kasus. Tidak perlu melestarikan nilai-nilai lama. Dalam kasus Michael Romano, nama lama tersebut salah dan perlu dibuang. Ketika pengguna perlu menemukan semua pesanan dari Michael Romano, dia akan menggunakan nama yang benar. Prinsip yang sama berlaku untuk perubahan nama pelanggan Kristin Samuelson.

Namun perubahan status perkawinannya sedikit berbeda. Perubahan ini dapat ditangani dengan cara yang sama seperti perubahan nama pelanggan hanya jika perubahan tersebut merupakan koreksi kesalahan. Jika tidak, Anda akan menimbulkan masalah saat pengguna ingin menganalisis pesanan berdasarkan status perkawinan.

Berikut adalah prinsip umum untuk perubahan tipe 1:

- ✿ Biasanya, perubahan berhubungan dengan koreksi kesalahan dalam sistem sumber.
- ✿ Kadang-kadang perubahan pada sistem sumber tidak mempunyai arti penting.
- ✿ Nilai lama dalam sistem sumber perlu dibuang.
- ✿ Perubahan pada sistem sumber tidak perlu disimpan di gudang data.



Gambar 11.2 Metode penerapan perubahan tipe 1.

Menerapkan Perubahan Tipe 1 pada Gudang Data Gambar 11.2 menunjukkan penerapan perubahan tipe 1 pada tabel dimensi pelanggan. Cara menerapkan perubahan tipe I adalah:

- ⊗ Timpa nilai atribut pada baris tabel dimensi dengan nilai baru.
- ⊗ Nilai atribut yang lama tidak dipertahankan.
- ⊗ Tidak ada perubahan lain yang dilakukan pada baris tabel dimensi.
- ⊗ Kunci tabel dimensi ini atau nilai kunci lainnya tidak terpengaruh.
- ⊗ Jenis ini paling mudah diterapkan.

Perubahan Tipe 2: Pelestarian Sejarah

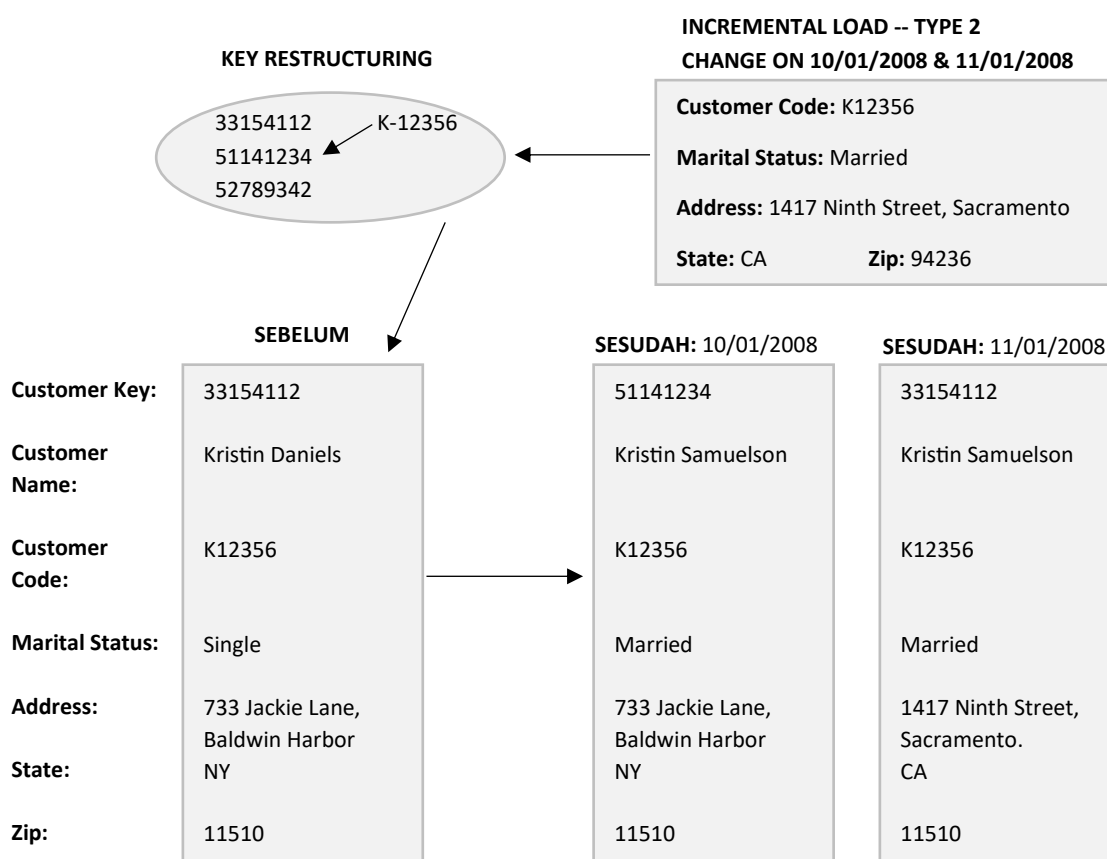
Sifat Perubahan Tipe 2 Kembali ke perubahan status perkawinan Kristin Samuelson. Asumsikan bahwa di gudang data Anda, salah satu persyaratan penting adalah melacak pesanan berdasarkan status perkawinan selain melacak berdasarkan atribut lainnya. Jika perubahan status perkawinan terjadi pada tanggal 1 Oktober 2008, maka seluruh pesanan dari Kristin Samuelson sebelum tanggal tersebut harus dimasukkan dalam status perkawinan: lajang, dan semua pesanan pada atau setelah tanggal 1 Oktober 2008 harus dimasukkan dalam status perkawinan: menikah.

Apa sebenarnya yang dibutuhkan dalam kasus ini? Pada data warehouse, Anda harus mempunyai cara untuk memisahkan pesanan pelanggan agar pesanan sebelum dan sesudah tanggal tersebut dapat dijumlahkan secara terpisah.

Sekarang mari kita tambahkan perubahan lain pada informasi tentang Kristin Samuelson. Asumsikan dia pindah ke alamat baru di California dari alamat lamanya di New York pada tanggal 1 November 2008. Jika gudang data Anda mengharuskan Anda untuk dapat melacak pesanan berdasarkan negara bagian, maka perubahan ini juga harus ditangani seperti perubahan status perkawinan. Setiap pesanan sebelum tanggal 1 November 2008 akan berada di bawah negara bagian: NY.

Jenis perubahan yang telah kita bahas untuk status perkawinan dan alamat pelanggan adalah perubahan tipe 2. Berikut adalah prinsip umum untuk jenis perubahan ini:

- ✿ Biasanya berhubungan dengan perubahan nyata dalam sistem sumber.
- ✿ Ada kebutuhan untuk melestarikan sejarah di gudang data.
- ✿ Jenis perubahan ini mempartisi riwayat di gudang data.
- ✿ Setiap perubahan untuk atribut yang sama harus dipertahankan.



Gambar 11.3 Metode penerapan perubahan tipe 2.

Menerapkan Perubahan Tipe 2 pada Data Warehouse Gambar 11.3 menunjukkan penerapan perubahan tipe 2 pada tabel dimensi pelanggan. Cara menerapkan perubahan tipe 2 adalah:

- ✿ Tambahkan baris tabel dimensi baru dengan nilai baru dari atribut yang diubah.
- ✿ Bidang tanggal efektif dapat dimasukkan dalam tabel dimensi.

- ❖ Tidak ada perubahan pada baris asli pada tabel dimensi.
- ❖ Kunci baris asli tidak terpengaruh.
- ❖ Baris baru disisipkan dengan kunci pengganti baru.

Perubahan Tipe 3: Revisi Lembut Tentatif

Sifat Perubahan Tipe 3 Hampir semua perubahan yang biasa terjadi pada nilai dimensi adalah perubahan tipe 1 atau tipe 2. Dari keduanya, perubahan tipe 1 lebih umum terjadi. Perubahan tipe 2 melestarikan sejarah. Ketika Anda menerapkan perubahan tipe 2 pada tanggal tertentu, tanggal tersebut adalah titik batas. Dalam hal perubahan status perkawinan di atas pada tanggal 1 Oktober 2008, maka tanggal tersebut merupakan tanggal batasnya. Setiap pesanan dari pelanggan sebelum tanggal tersebut termasuk dalam kelompok pesanan yang lebih lama; pesanan pada atau setelah tanggal tersebut termasuk dalam grup pesanan baru. Pesanan untuk pelanggan ini harus termasuk dalam satu kelompok atau lainnya; itu tidak dapat dihitung dalam kedua kelompok untuk jangka waktu tertentu.

Bagaimana jika Anda perlu menghitung pesanan pada atau setelah tanggal batas waktu di kedua grup selama periode tertentu setelah tanggal batas waktu? Anda tidak dapat menangani perubahan ini sebagai perubahan tipe 2. Kadang-kadang, meskipun jarang, ada kebutuhan untuk melacak nilai-nilai lama dan baru dari atribut yang diubah untuk jangka waktu tertentu, baik dalam arah maju maupun mundur. Jenis perubahan ini adalah perubahan tipe 3.

Perubahan tipe 3 bersifat tentatif atau perubahan lunak. Sebuah contoh akan memperjelas hal ini. Asumsikan departemen pemasaran Anda sedang mempertimbangkan penataan kembali penugasan teritorial untuk tenaga penjualan. Sebelum melakukan penataan kembali secara permanen, mereka ingin menghitung perintah dengan dua cara: menurut penataan wilayah saat ini dan juga menurut usulan penataan kembali. Jenis perubahan sementara atau tentatif ini merupakan perubahan tipe 3.

Sebagai contoh, katakanlah Anda ingin memindahkan penjual Robert Smith dari wilayah New England ke wilayah Chicago dengan kemampuan melacak pesannya di kedua wilayah. Anda perlu melacak semua pesanan melalui Robert Smith di kedua wilayah.

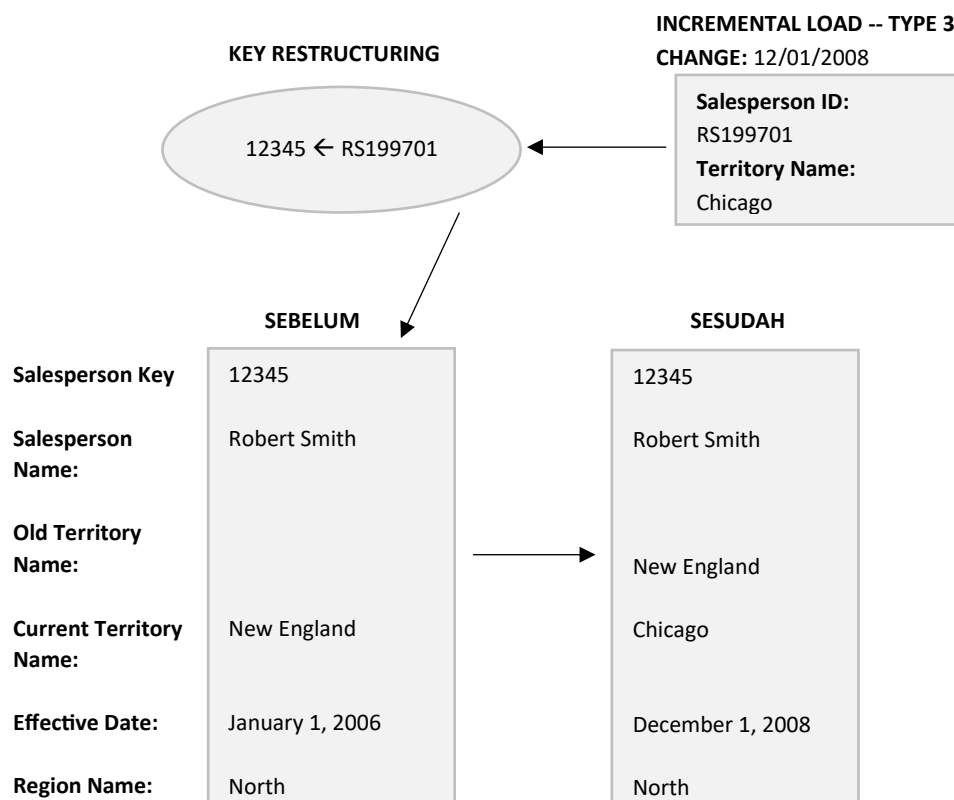
Berikut adalah prinsip umum untuk perubahan tipe 3:

- ❖ Hal ini biasanya berkaitan dengan perubahan “lunak” atau tentatif dalam sistem sumber.
- ❖ Ada kebutuhan untuk melacak sejarah dengan nilai-nilai lama dan baru dari atribut yang diubah.
- ❖ Indikator ini digunakan untuk membandingkan kinerja selama masa transisi.
- ❖ Mereka memberikan kemampuan untuk melacak maju dan mundur.

Menerapkan Perubahan Tipe 3 pada Data Warehouse Gambar 11-4 menunjukkan penerapan perubahan tipe 3 pada tabel dimensi pelanggan. Metode untuk menerapkan perubahan tipe 3 adalah:

- ❖ Tambahkan kolom “lama” di tabel dimensi untuk atribut yang terpengaruh.
- ❖ Tekan nilai atribut yang ada dari kolom “saat ini” ke kolom “lama”.
- ❖ Simpan nilai atribut baru di kolom “saat ini”.

- ❖ Selain itu, Anda juga dapat menambahkan kolom tanggal efektif “saat ini” untuk atribut tersebut.
- ❖ Kunci baris tidak terpengaruh.
- ❖ Tidak diperlukan baris dimensi baru.
- ❖ Kueri yang ada akan dengan mudah beralih ke nilai “saat ini”.
- ❖ Setiap pertanyaan yang perlu menggunakan nilai “lama” harus direvisi.
- ❖ Teknik ini bekerja paling baik untuk satu perubahan “lunak” dalam satu waktu.
- ❖ Jika terjadi perubahan secara berurutan, teknik yang lebih canggih harus dirancang.



Gambar 11.4 Menerapkan perubahan tipe 3.

11.2 DIMENSI LAIN-LAIN

Setelah mempertimbangkan jenis perubahan atribut dimensi dan cara menangani perubahan dimensi dalam data warehouse, sekarang mari kita mengalihkan perhatian kita ke beberapa isu penting lainnya tentang dimensi. Salah satu permasalahan berkaitan dengan tabel dimensi yang sangat luas dan sangat dalam.

Dalam pembahasan kita sebelumnya, kita berasumsi bahwa atribut dimensi tidak berubah terlalu cepat. Jika perubahannya adalah perubahan tipe 2, Anda tahu bahwa Anda harus membuat baris lain dengan nilai atribut yang baru. Jika nilai atribut berubah lagi, maka Anda membuat baris lain dengan nilai yang lebih baru. Bagaimana jika nilainya berubah terlalu sering atau terlalu cepat? Dimensi seperti itu bukan lagi dimensi yang berubah secara perlahan. Apa yang harus Anda lakukan terhadap dimensi yang berubah tidak terlalu lambat?

Kami akan menyelesaikan diskusi kami tentang dimensi dengan mempertimbangkan isu-isu relevan tersebut.

Dimensi Besar

Anda dapat mempertimbangkan suatu dimensi besar berdasarkan dua faktor. Dimensi yang besar sangatlah dalam; yaitu dimensi yang mempunyai jumlah baris yang sangat banyak. Dimensi yang besar mungkin juga sangat lebar; artinya, dimensi tersebut mungkin memiliki sejumlah besar atribut atau kolom. Dalam kedua kasus tersebut, Anda dapat menyatakan dimensinya besar. Ada pertimbangan khusus untuk dimensi besar. Anda mungkin harus mengisi tabel berdimensi besar dengan cara khusus. Anda mungkin ingin memisahkan beberapa dimensi mini dari dimensi besar. Kami akan mengambil skema STAR sederhana yang dirancang untuk analisis pesanan. Anggaplah ini adalah skema untuk perusahaan manufaktur dan departemen pemasaran tertarik untuk menentukan bagaimana mereka membuat kemajuan dengan pesanan yang diterima oleh perusahaan.

Dalam gudang data, biasanya dimensi pelanggan dan produk cenderung besar. Setiap kali suatu perusahaan berhubungan dengan masyarakat umum, dimensi pelanggan diharapkan menjadi sangat besar. Dimensi pelanggan dari jaringan ritel nasional dapat mendekati ukuran jumlah rumah tangga di AS. Tabel dimensi pelanggan tersebut mungkin memiliki sebanyak 100 juta baris. Pada skala berikutnya, jumlah baris tabel dimensi perusahaan di industri telekomunikasi dan perjalanan juga dapat mencapai jutaan. Memiliki 10 atau 20 juta baris pelanggan bukanlah hal yang aneh. Dimensi produk pengecer besar juga cukup besar.

Berikut adalah beberapa ciri khas dari dimensi pelanggan dan produk yang besar:

A. Pelanggan

- ☞ Sangat besar dalam kisaran 20 juta baris
- ☞ Mudah hingga 150 atribut dimensi
- ☞ Dapat memiliki banyak hierarki

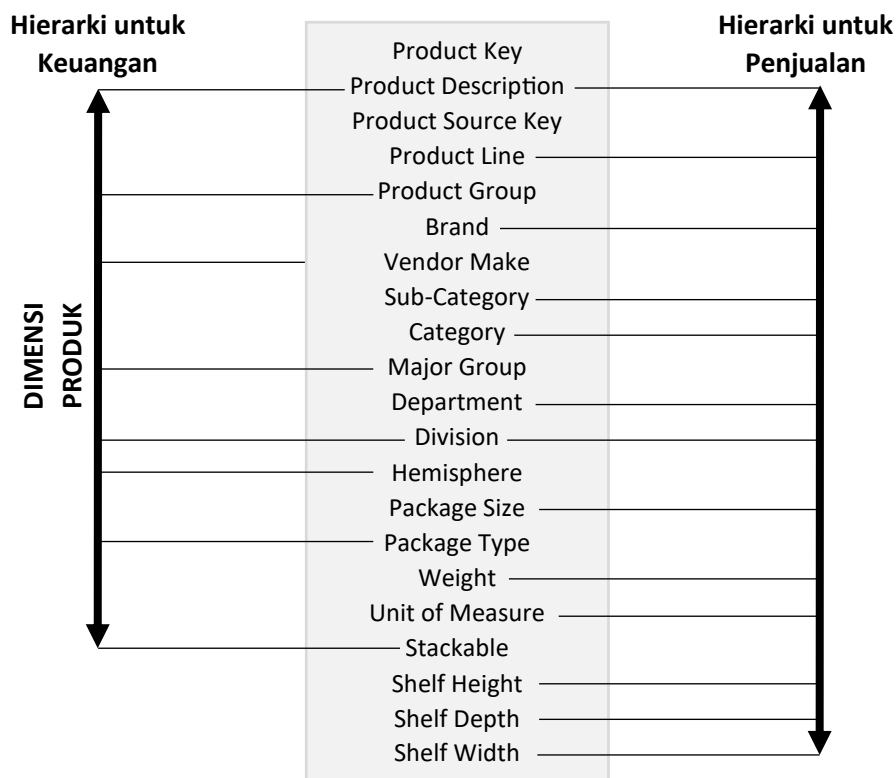
B. Produk

- ☞ Terkadang sebanyak 100.000 variasi produk
- ☞ Dapat memiliki lebih dari 100 atribut dimensi
- ☞ Dapat memiliki banyak hierarki

Dimensi yang besar memerlukan pertimbangan khusus. Karena ukurannya yang besar, banyak fungsi gudang data yang melibatkan dimensi besar mungkin menjadi lambat dan tidak efisien. Anda perlu mengatasi masalah berikut dengan menggunakan metode desain yang efektif, dengan memilih indeks yang tepat, dan dengan menerapkan teknik pengoptimalan lainnya:

- Mengisi tabel dimensi yang sangat besar
- Telusuri kinerja dimensi yang tidak dibatasi, terutama ketika kardinalitas atributnya rendah
- Waktu penelusuran untuk nilai atribut dimensi yang dibatasi silang
- Inefisiensi dalam query tabel fakta ketika dimensi besar perlu digunakan

➤ Baris tambahan dibuat untuk menangani perubahan dimensi tipe 2 secara perlahan. Hierarki Berganda Dimensi besar biasanya memiliki karakteristik lain yang berbeda. Mereka cenderung memiliki banyak hierarki. Ambil contoh dimensi produk untuk pengecer besar. Satu set atribut dapat membentuk hierarki departemen pemasaran. Pengguna dari departemen tersebut menggunakan atribut ini untuk menelusuri atau meningkatkan. Dengan cara yang sama, departemen keuangan mungkin perlu menggunakan seperangkat atributnya sendiri dari dimensi produk yang sama untuk menelusuri atau meningkatkannya. Gambar 11.5 menunjukkan beberapa hierarki dalam dimensi produk yang besar.



Gambar 11.5 Beberapa hierarki dalam dimensi produk yang besar.

Dimensi yang Berubah dengan Cepat

Seperti yang Anda ketahui, saat Anda menangani perubahan tipe 2, Anda membuat baris tabel dimensi tambahan dengan nilai baru dari atribut yang diubah. Dengan demikian, Anda bisa melestarikan sejarah. Jika atribut yang sama berubah untuk kedua kalinya, Anda membuat satu baris tabel dimensi lagi dengan nilai terbaru.

Kebanyakan dimensi produk sangat jarang berubah, mungkin sekali atau dua kali setahun. Jika jumlah baris dalam dimensi produk tersebut sekitar 100.000 atau lebih, pendekatan pembuatan baris tambahan dengan nilai atribut baru dapat dikelola dengan mudah. Sekalipun jumlah barisnya berkisar beberapa ribu, pendekatan penerapan perubahan sebagai perubahan tipe 2 masih cukup layak.

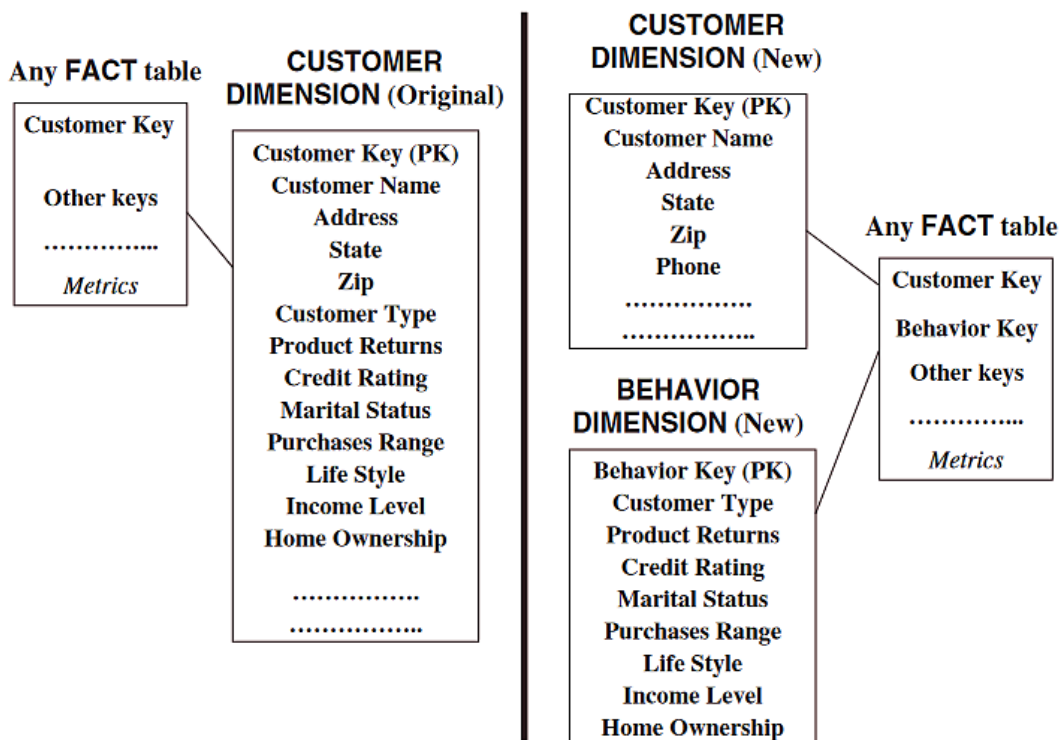
Namun, pertimbangkan dimensi lain, seperti dimensi pelanggan. Di sini jumlah barisnya cenderung banyak, bahkan terkadang berkisar satu juta baris atau lebih. Jika atribut

dari sejumlah besar baris berubah, namun jarang berubah, pendekatan tipe 2 tidak terlalu sulit. Namun atribut signifikan dalam dimensi pelanggan dapat berubah berkali-kali dalam setahun. Mengubah dimensi besar dengan cepat bisa menjadi masalah bagi pendekatan tipe 2. Tabel dimensi dapat dipenuhi dengan sejumlah besar baris tambahan yang dibuat setiap kali ada penambahan beban.

Sebelum terburu-buru mengeksplorasi pilihan lain untuk menangani dimensi besar yang berubah dengan cepat, tangani setiap dimensi besar satu per satu. Pendekatan tipe 2 masih bagus dalam desain skema STAR. Berikut beberapa alasan mengapa pendekatan tipe 2 dapat berhasil dalam banyak kasus untuk dimensi yang berubah dengan cepat:

- a) Jika tabel dimensi dibuat datar, hal ini memungkinkan penelusuran silang simetris di antara berbagai atribut dimensi.
- b) Bahkan ketika baris tabel dimensi tambahan dibuat, struktur dimensi dasar tetap dipertahankan. Tabel fakta dihubungkan ke semua tabel dimensi dengan kunci asing. Keunggulan skema STAR masih tersedia.
- c) Hanya ketika kueri pengguna akhir didasarkan pada atribut yang diubah barulah keberadaan beberapa baris untuk pelanggan yang sama menjadi jelas. Untuk kueri lain, keberadaan beberapa baris praktis tersembunyi.

Bagaimana jika tabel dimensi terlalu besar dan berubah terlalu cepat? Kemudian carilah alternatif untuk penerapan langsung pendekatan tipe 2. Salah satu pendekatan yang efektif adalah dengan memecah tabel dimensi besar menjadi satu atau lebih tabel dimensi sederhana. Bagaimana Anda dapat mencapai hal ini?



Gambar 11.6 Membagi tabel dimensi yang besar dan berubah dengan cepat.

Jelasnya, Anda perlu memecah atribut yang berubah dengan cepat ke dalam tabel dimensi lain, meninggalkan atribut yang berubah secara perlahan di tabel aslinya. Gambar 11.6 menunjukkan bagaimana tabel dimensi pelanggan dapat dipisahkan menjadi tabel dua dimensi. Gambar tersebut mengilustrasikan teknik umum untuk memisahkan atribut-atribut yang berubah dengan cepat. Gunakan ini sebagai panduan ketika berhadapan dengan dimensi yang besar dan berubah dengan cepat di lingkungan gudang data Anda.

Dimensi Sampah

Periksa sistem warisan sumber Anda dan tinjau masing-masing bidang dalam struktur data sumber untuk pelanggan, produk, pesanan, wilayah penjualan, kampanye promosi, dan sebagainya. Sebagian besar bidang ini dimasukkan ke dalam tabel dimensi. Anda akan melihat bahwa beberapa bidang seperti bendera lain-lain dan bidang tekstual tertinggal di struktur data sumber. Ini termasuk tanda ya/tidak, kode tekstual, dan teks bentuk bebas.

Beberapa dari tanda dan data tekstual ini mungkin terlalu kabur untuk dapat memberikan nilai sebenarnya. Ini mungkin merupakan sisa dari konversi sebelumnya dari catatan manual yang dibuat sejak lama. Namun, banyak tanda dan teks yang dapat berguna sesekali dalam kueri. Ini mungkin tidak dimasukkan sebagai bidang penting dalam dimensi utama. Pada saat yang sama, bendera dan teks ini juga tidak dapat dibuang. Jadi, apa pilihan Anda? Berikut pilihan utamanya:

- Kecualikan dan buang semua bendera dan teks. Jelas, ini bukan pilihan yang baik, karena alasan sederhana yaitu Anda cenderung membuang beberapa informasi berguna.
- Tempatkan bendera dan teks tanpa perubahan di tabel fakta. Opsi ini kemungkinan besar akan menambah tabel fakta sehingga tidak memberikan keuntungan yang spesifik.
- Jadikan setiap bendera dan teks sebagai tabel dimensi tersendiri. Dengan menggunakan opsi ini, jumlah tabel dimensi akan bertambah banyak.
- Simpan hanya bendera dan teks yang bermakna; kelompokkan semua tanda yang berguna ke dalam satu dimensi "sampah". Atribut dimensi "Sampah" berguna untuk membatasi kueri berdasarkan nilai tanda/teks.

SKEMA Kepingan Salju

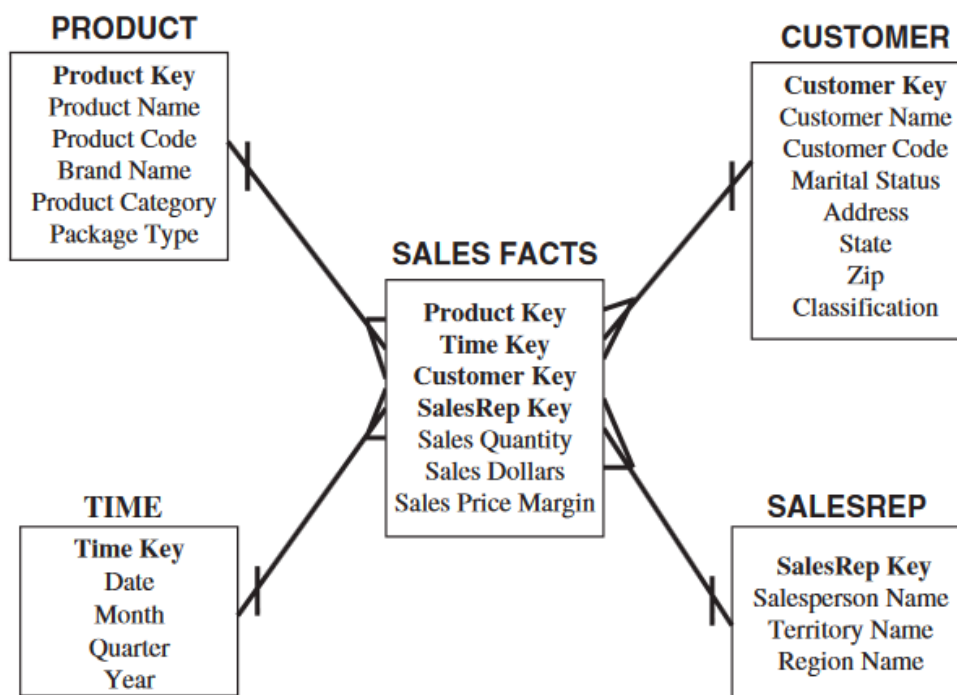
"Snowflaking" adalah metode normalisasi tabel dimensi dalam skema STAR. Ketika Anda sepenuhnya menormalkan semua tabel dimensi, struktur yang dihasilkan menyerupai kepingan salju dengan tabel fakta di tengahnya. Pertama, mari kita mulai dengan Gambar 11.7, yang menunjukkan skema STAR sederhana untuk penjualan di perusahaan manufaktur.

Tabel fakta penjualan berisi kuantitas, harga, dan metrik relevan lainnya. Tenaga penjualan, pelanggan, produk, dan waktu adalah tabel dimensi. Ini adalah skema STAR klasik, yang didenormalisasi untuk akses kueri optimal yang melibatkan semua atau sebagian besar dimensi. Model tersebut tidak berada dalam bentuk normal ketiga.

Pilihan untuk Menormalkan

Asumsikan ada 500.000 baris dimensi produk. Produk-produk ini termasuk dalam 500 merek produk dan merek-merek produk ini termasuk dalam 10 kategori produk. Sekarang

anggaplah salah satu pengguna Anda menjalankan kueri yang membatasi hanya pada kategori produk. Jika tabel dimensi produk tidak diindeks pada kategori produk, kueri harus menelusuri 500.000 baris. Di sisi lain, bahkan jika dimensi produk dinormalisasi sebagian dengan memisahkan merek produk dan kategori produk ke dalam tabel terpisah, pencarian awal untuk kueri harus melalui hanya 10 baris dalam tabel kategori produk. Gambar 11.8 mengilustrasikan pengurangan proses pencarian ini.

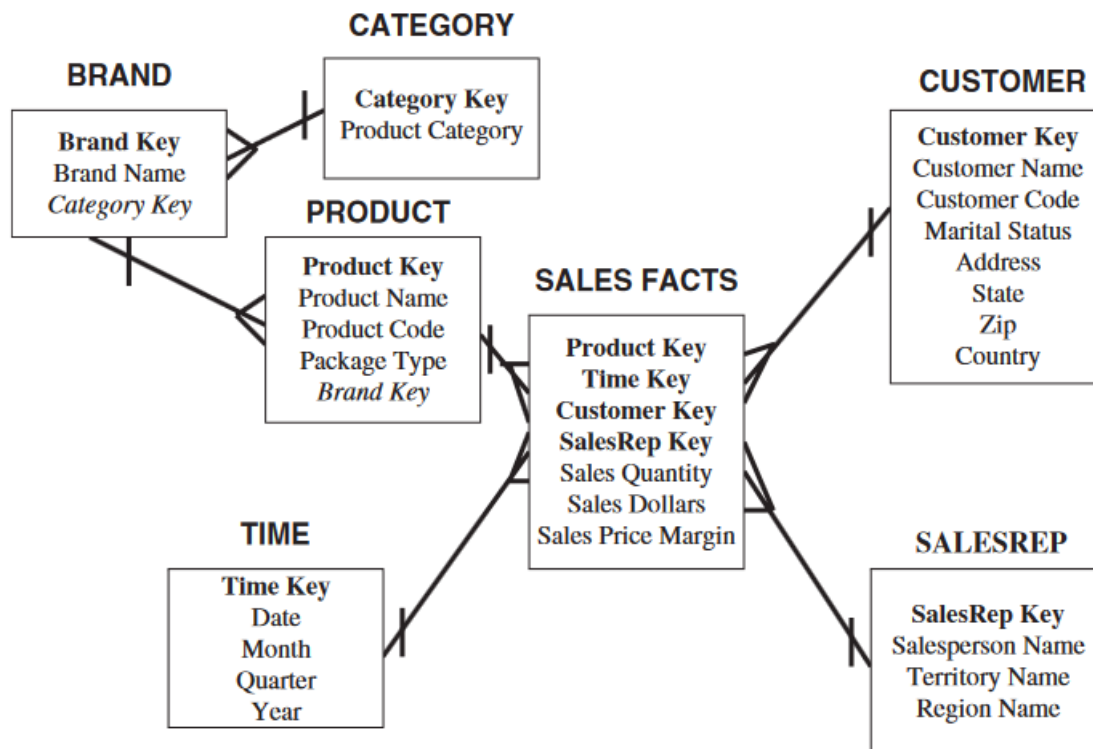


Gambar 11.7 Penjualan: skema STAR sederhana.

Pada Gambar 11.8, kita belum sepenuhnya menormalisasi dimensi produk. Kita juga dapat memindahkan atribut lain dari tabel dimensi produk dan membentuk struktur yang dinormalisasi. “Snowflaking” atau normalisasi tabel dimensi dapat dicapai dengan beberapa cara berbeda. Saat Anda ingin “kepingan salju”, periksa konten dan penggunaan normal setiap tabel dimensi.

Opsi berikut menunjukkan berbagai metode yang mungkin ingin Anda pertimbangkan untuk normalisasi tabel dimensi:

- ✘ Normalisasikan sebagian hanya beberapa tabel dimensi, biarkan tabel lainnya tetap utuh.
- ✘ Normalisasikan sebagian atau seluruhnya hanya beberapa tabel dimensi, biarkan sisanya tetap utuh.
- ✘ Menormalkan sebagian setiap tabel dimensi.
- ✘ Menormalkan sepenuhnya setiap tabel dimensi.



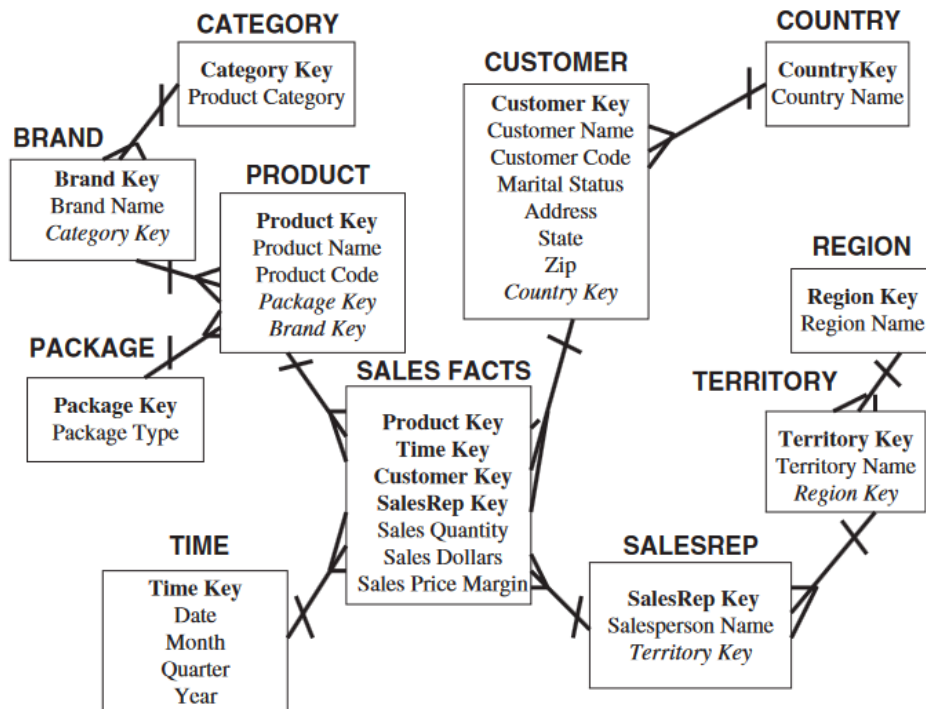
Gambar 11.8 Dimensi produk: dinormalisasi sebagian.

Gambar 11.9 menunjukkan versi skema kepingan salju untuk penjualan di mana setiap tabel dimensi dinormalisasi sebagian atau seluruhnya. Skema STAR asli untuk penjualan seperti yang ditunjukkan pada Gambar 11-7 hanya berisi 5 tabel, sedangkan versi yang dinormalisasi sekarang bertambah menjadi 11 tabel. Anda akan melihat bahwa dalam skema kepingan salju, atribut dengan kardinalitas rendah di setiap tabel dimensi asli dihilangkan untuk membentuk tabel terpisah. Tabel baru ini dihubungkan kembali ke tabel dimensi asli melalui kunci buatan.

Keuntungan dan kerugian

Anda mungkin ingin membuat kepingan salju karena satu alasan yang jelas. Dengan menghilangkan semua bidang teks panjang dari tabel dimensi, Anda berharap dapat menghemat ruang penyimpanan. Misalnya, jika Anda memiliki “perabotan pria” sebagai salah satu nama kategori, teks tersebut akan diulangi di setiap baris produk dalam kategori tersebut. Sekilas, menghilangkan redundansi seperti itu mungkin terlihat menghemat ruang penyimpanan secara signifikan ketika dimensinya besar.

Misalkan tabel dimensi produk Anda memiliki 500.000 baris. Dengan melakukan snowflaking Anda dapat menghapus 500.000 nama kategori 20-byte. Pada saat yang sama, Anda harus menambahkan kunci kategori buatan 4-byte ke tabel dimensi. Penghematan bersihnya kira-kira 500.000 kali 16, yaitu sekitar 8 MB. Tabel dimensi produk rata-rata 500.000 baris menempati sekitar 200 MB ruang penyimpanan dan tabel fakta terkait membutuhkan 20 GB lagi. Penghematannya hanya 4%. Anda akan menemukan bahwa penghematan ruang yang kecil tidak mengimbangi kerugian lain dari kepingan salju.



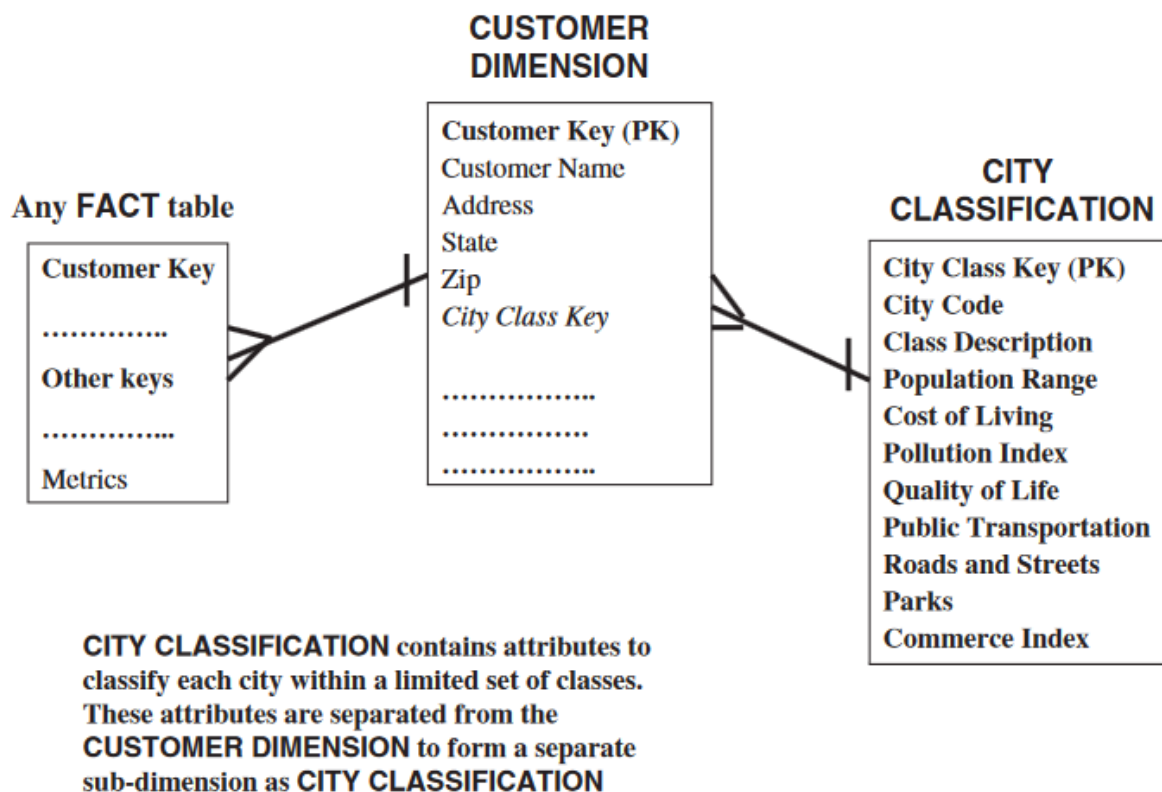
Gambar 11.9 Penjualan: skema “kepingan salju”.

Berikut ringkasan singkat kelebihan dan keterbatasan kepingan salju:

❖ **Keuntungan**

- ✓ Penghematan kecil dalam ruang penyimpanan
- ✓ Struktur yang dinormalisasi lebih mudah untuk diperbarui dan dipelihara. Kerugian
- ✓ Skema kurang intuitif dan pengguna akhir tidak tertarik dengan kompleksitasnya
- ✓ Kemampuan untuk menelusuri isi sulit
- ✓ Performa kueri menurun karena adanya penggabungan tambahan

Kepingan salju umumnya tidak direkomendasikan di lingkungan gudang data. Kinerja kueri memiliki arti paling penting dalam gudang data dan kepingan salju menghambat kinerja.



Gambar 11.10 Membentuk subdimensi.

Kapan Kepingan Salju

Sebagai seorang profesional TI, Anda memiliki ketertarikan terhadap struktur bentuk normal ketiga. Kami mengetahui betul semua masalah yang dapat ditimbulkan oleh struktur yang tidak dinormalisasi. Selain itu, ruang yang terbuang bisa menjadi pertimbangan lain untuk pembuatan kepingan salju.

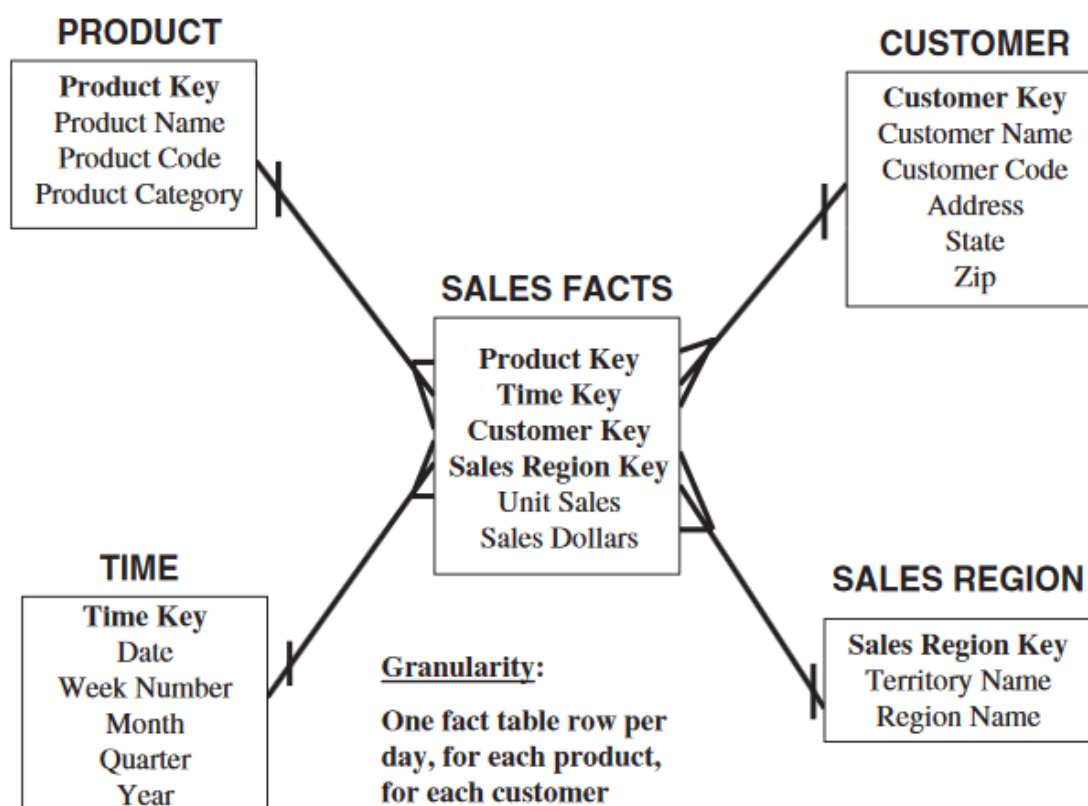
Terlepas dari kerugian yang tampak, adakah kondisi yang memungkinkan pengelupasan salju diperbolehkan? Prinsip di balik kepingan salju adalah normalisasi tabel dimensi dengan menghilangkan atribut berkardinalitas rendah dan membentuk tabel terpisah. Dengan cara yang sama, beberapa situasi memberikan peluang untuk memisahkan sekumpulan atribut dan membentuk subdimensi. Proses ini sangat mirip dengan teknik kepingan salju. Gambar 11.10 menunjukkan bagaimana subdimensi klasifikasi kota dibentuk dari dimensi pelanggan.

Meskipun pembentukan subdimensi dapat dianggap sebagai kepingan salju, sangat masuk akal untuk memisahkan atribut demografi ke dalam tabel lain. Anda biasanya memuat data demografi pada waktu yang berbeda dengan waktu ketika Anda memuat atribut dimensi lainnya. Kedua kumpulan atribut tersebut berbeda dalam rinciannya. Jika dimensi pelanggan sangat besar, mencapai jutaan baris, penghematan ruang penyimpanan bisa sangat besar. Alasan valid lainnya untuk memisahkan atribut demografis berkaitan dengan penelusuran atribut. Pengguna dapat menjelajahi atribut demografis lebih banyak dibandingkan atribut lainnya dalam tabel dimensi pelanggan.

11.3 TABEL FAKTA AGREGAT

Agregat adalah ringkasan yang telah dihitung sebelumnya dan diperoleh dari tabel fakta paling terperinci. Ringkasan ini membentuk satu set tabel fakta agregat yang terpisah. Anda dapat membuat setiap tabel fakta agregat sebagai ringkasan spesifik pada sejumlah dimensi. Mari kita mulai dengan memeriksa contoh skema STAR. Pilih skema STAR sederhana dengan tabel fakta pada tingkat perincian serendah mungkin. Asumsikan ada tabel empat dimensi yang mengelilingi tabel fakta paling terperinci ini. Gambar 11.11 menunjukkan contoh yang ingin kita periksa.

Apa perbedaan mendasar antara kueri yang dijalankan di sistem operasional dan kueri yang dijalankan di lingkungan gudang data? Saat Anda menjalankan kueri dalam sistem operasional, ini menghasilkan kumpulan hasil tentang satu pelanggan, satu pesanan, satu faktur, satu produk, dan seterusnya. Namun, seperti yang Anda ketahui, kueri di lingkungan gudang data menghasilkan kumpulan hasil yang besar. Kueri di gudang data mengambil ratusan dan ribuan baris tabel, memanipulasi metrik dalam tabel fakta, dan kemudian menghasilkan kumpulan hasil. Manipulasi metrik tabel fakta dapat berupa penambahan sederhana, penambahan dengan beberapa penyesuaian, perhitungan rata-rata, atau bahkan penerapan algoritma aritmatika yang kompleks.



Gambar 11.11 Skema STAR dengan tabel fakta paling terperinci.

Mari kita tinjau beberapa pertanyaan umum terhadap contoh skema STAR yang ditunjukkan pada Gambar 11.11.

- **Kueri 1:** Total penjualan untuk nomor pelanggan 12345678 selama minggu pertama bulan Desember 2008 untuk produk Widget-1.
- **Kueri 2:** Total penjualan untuk nomor pelanggan 12345678 selama tiga bulan pertama tahun 2009 untuk produk Widget-1.
- **Kueri 3:** Total penjualan untuk seluruh pelanggan di wilayah selatan-tengah selama dua kuartal pertama tahun 2009 untuk kategori produk Bigtools.

Cermati pertanyaan-pertanyaan ini dan tentukan bagaimana totalnya akan dihitung dalam setiap kasus. Totalnya akan dihitung dengan menjumlahkan jumlah penjualan dan jumlah penjualan dari baris kualifikasi tabel fakta. Dalam setiap kasus, mari kita tinjau baris kualifikasi yang berkontribusi terhadap total kumpulan hasil.

- a) **Kueri 1:** Semua baris tabel fakta di mana kunci pelanggan berkaitan dengan nomor pelanggan 12345678, kunci produk berkaitan dengan Widget-1 produk, dan kunci waktu berkaitan dengan tujuh hari pada minggu pertama bulan Desember 2008. Dengan asumsi bahwa pelanggan dapat membuat paling banyak satu pembelian satu produk dalam satu hari, hanya maksimal tujuh baris tabel fakta yang berpartisipasi dalam penjumlahan.
- b) **Kueri 2:** Semua baris tabel fakta di mana kunci pelanggan berkaitan dengan nomor pelanggan 12345678, kunci produk berkaitan dengan Widget-1 produk, dan kunci waktu berkaitan dengan sekitar 90 hari pada kuartal pertama tahun 2009. Dengan asumsi bahwa pelanggan dapat menghasilkan di sebagian besar pembelian satu produk dalam satu hari, hanya sekitar 90 baris tabel fakta atau kurang yang berpartisipasi dalam penjumlahan.
- c) **Kueri 3:** Semua baris tabel fakta dengan kunci pelanggan berkaitan dengan semua pelanggan di wilayah selatan-tengah, kunci produk berkaitan dengan semua produk dalam kategori produk Bigtools, dan kunci waktu berkaitan dengan sekitar 180 hari dalam dua kuartal pertama 2009. Dalam hal ini, jelas sejumlah besar baris tabel fakta berpartisipasi dalam penjumlahan.

Jelasnya, kueri 3 akan berjalan lama karena banyaknya baris tabel fakta yang harus diambil. Apa yang bisa dilakukan untuk mengurangi waktu kueri? Di sinilah tabel agregat dapat membantu. Sebelum kita membahas tabel fakta agregat secara mendetail, mari kita tinjau ukuran beberapa tabel fakta umum di gudang data dunia nyata.

Ukuran Tabel Fakta

Gambar 11.12 mewakili skema STAR untuk penjualan jaringan supermarket besar. Terdapat sekitar dua miliar baris tabel fakta dasar dengan tingkat perincian terendah. Pelajari perhitungan yang ditunjukkan di bawah ini:

- **Dimensi waktu:** 5 tahun × 365 hari $\frac{1}{4}$ 1825 Dimensi toko: 300 toko melaporkan penjualan harian
- **Dimensi produk:** 40.000 produk di setiap toko (sekitar 4000 terjual di setiap toko setiap hari)

- **Dimensi promosi:** item yang terjual hanya boleh ada dalam satu promosi di toko pada hari tertentu Jumlah maksimum catatan tabel fakta dasar: $1825 \times 300 \times 4000 \times 1 \frac{1}{2}$ miliar

Berikut adalah beberapa perkiraan ukuran tabel fakta dalam kasus-kasus umum lainnya:
Pemantauan panggilan telepon

- **Dimensi waktu:** 5 tahun $\frac{1}{4}$ 1825 hari Jumlah panggilan yang dilacak setiap hari: 150 juta
- **Jumlah maksimum catatan tabel fakta dasar:** 274 miliar Pelacakan transaksi kartu kredit
- **Dimensi waktu:** 5 tahun $\frac{1}{4}$ 60 bulan Jumlah rekening kartu kredit: 150 juta
- **Jumlah rata-rata transaksi bulanan per akun:** 20 Jumlah maksimum catatan tabel fakta dasar: 180 miliar

Dari contoh di atas, Anda dapat melihat besarnya tabel fakta yang berada pada tingkat perincian paling rendah. Meskipun tidak ada kueri dari pengguna yang memerlukan data hanya dari satu baris dalam tabel fakta ini, diperlukan data dengan tingkat detail terendah. Hal ini karena ketika pengguna melakukan berbagai bentuk analisis, dia harus mendapatkan kumpulan hasil yang terdiri dari berbagai kombinasi baris tabel fakta individual. Jika Anda tidak menyimpan rincian di masing-masing toko, Anda tidak dapat mengambil kumpulan hasil untuk produk di masing-masing toko. Di sisi lain, jika Anda tidak menyimpan detail berdasarkan produk individual, Anda tidak dapat mengambil kumpulan hasil untuk toko berdasarkan produk individual.

Jadi, inilah pertanyaannya. Jika Anda memerlukan data terperinci pada tingkat perincian terendah dalam tabel fakta dasar, bagaimana Anda menangani penjumlahan baris tabel fakta dalam jumlah besar untuk menghasilkan hasil kueri? Pertimbangkan pertanyaan berikut yang terkait dengan gudang data rantai grosir:

1. Bagaimana kinerja ketiga toko baru di Wisconsin selama tiga bulan terakhir dibandingkan dengan rata-rata nasional?
2. Apa dampak kampanye penjualan liburan terbaru terhadap daging dan unggas?
3. Bagaimana penjualan liburan tanggal 4 Juli berdasarkan kategori produk dibandingkan dengan tahun lalu?

Masing-masing dari ketiga kueri ini memerlukan pilihan dan penjumlahan dari baris tabel fakta. Untuk jenis penjumlahan ini, Anda memerlukan data terperinci berdasarkan satu atau beberapa dimensi, namun hanya ringkasan total berdasarkan dimensi lainnya. Misalnya, untuk kueri terakhir, Anda memerlukan data harian terperinci berdasarkan dimensi waktu, namun ringkasan total berdasarkan kategori produk. Dalam kasus apa pun, jika Anda memiliki ringkasan total atau agregat yang telah dihitung sebelumnya, kueri akan berjalan lebih cepat. Dengan ringkasan yang dikumpulkan dengan benar, kinerja setiap kueri ini dapat ditingkatkan secara signifikan.

Kebutuhan Agregat

Lihat Gambar 11.12 yang menunjukkan skema STAR untuk jaringan toko kelontong. Di 300 toko tersebut, asumsikan ada 500 produk per merek. Dari 40.000 produk, asumsikan

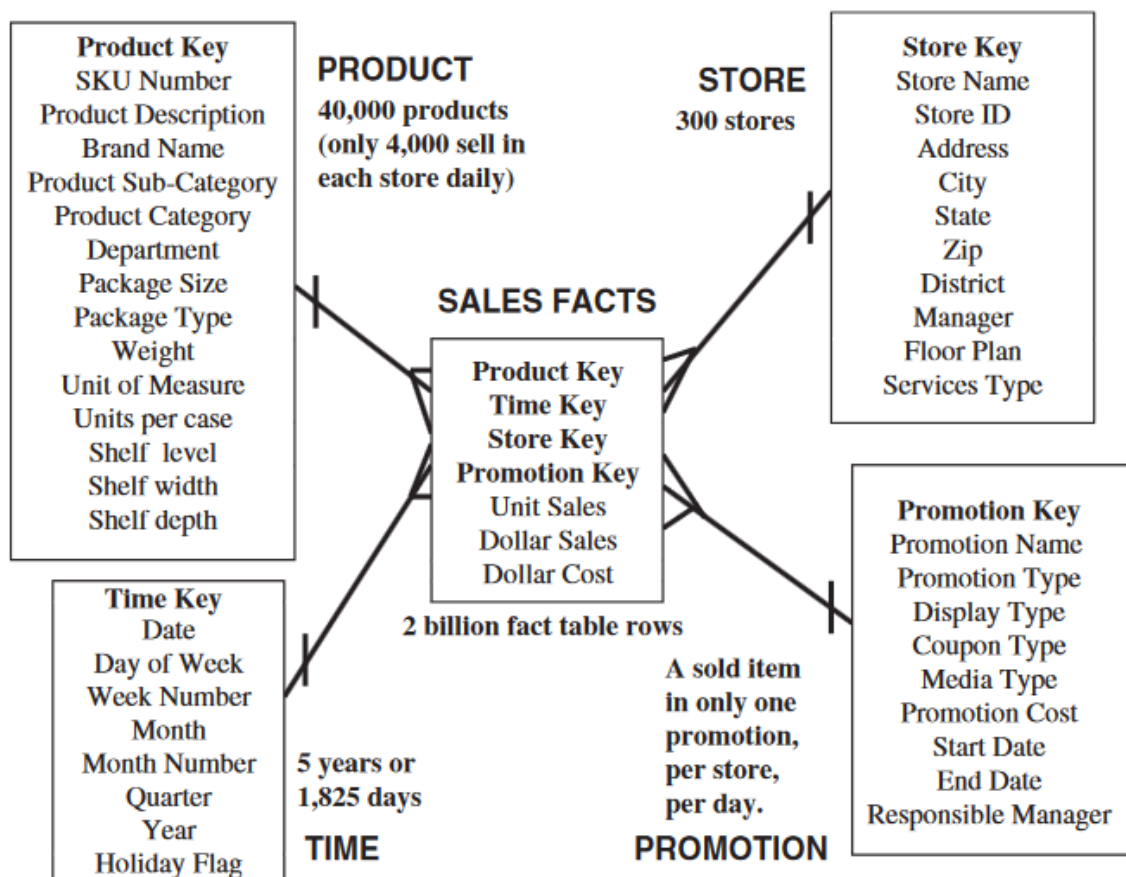
setidaknya ada satu penjualan per produk per toko per minggu. Mari kita perkirakan jumlah baris tabel fakta yang akan diambil dan diringkas untuk jenis kueri berikut:

Kueri melibatkan 1 produk, 1 toko, 1 minggu—ambil/ringkas hanya 1 baris tabel fakta
 Kueri melibatkan 1 produk, semua toko, 1 minggu—ambil/ringkas 300 baris tabel fakta
 Kueri melibatkan 1 merek, 1 toko, 1 minggu—ambil/merangkum 500 baris tabel fakta
 Kueri melibatkan 1 merek, semua toko, 1 tahun—mengambil/meringkas 7.800.000 baris tabel fakta

Misalkan Anda telah melakukan perhitungan sebelumnya dan membuat tabel fakta agregat yang setiap barisnya merangkum total total suatu merek, per toko, per minggu. Kemudian kueri ketiga harus mengambil hanya satu baris dari tabel fakta agregat ini. Demikian pula, kueri terakhir harus mengambil hanya 15.600 baris dari tabel fakta agregat ini, jauh lebih sedikit dari 7 juta baris.

Selanjutnya, jika Anda melakukan perhitungan awal dan membuat tabel fakta agregat lain yang setiap barisnya merangkum total merek, per toko, per tahun, kueri terakhir harus mengambil hanya 300 baris.

Agregat memiliki baris lebih sedikit dibandingkan tabel dasar. Oleh karena itu, ketika sebagian besar kueri dijalankan terhadap tabel fakta agregat dan bukan tabel fakta dasar, Anda akan melihat peningkatan kinerja yang luar biasa di gudang data. Pembentukan tabel fakta agregat tentunya merupakan metode yang sangat efektif untuk meningkatkan kinerja query.

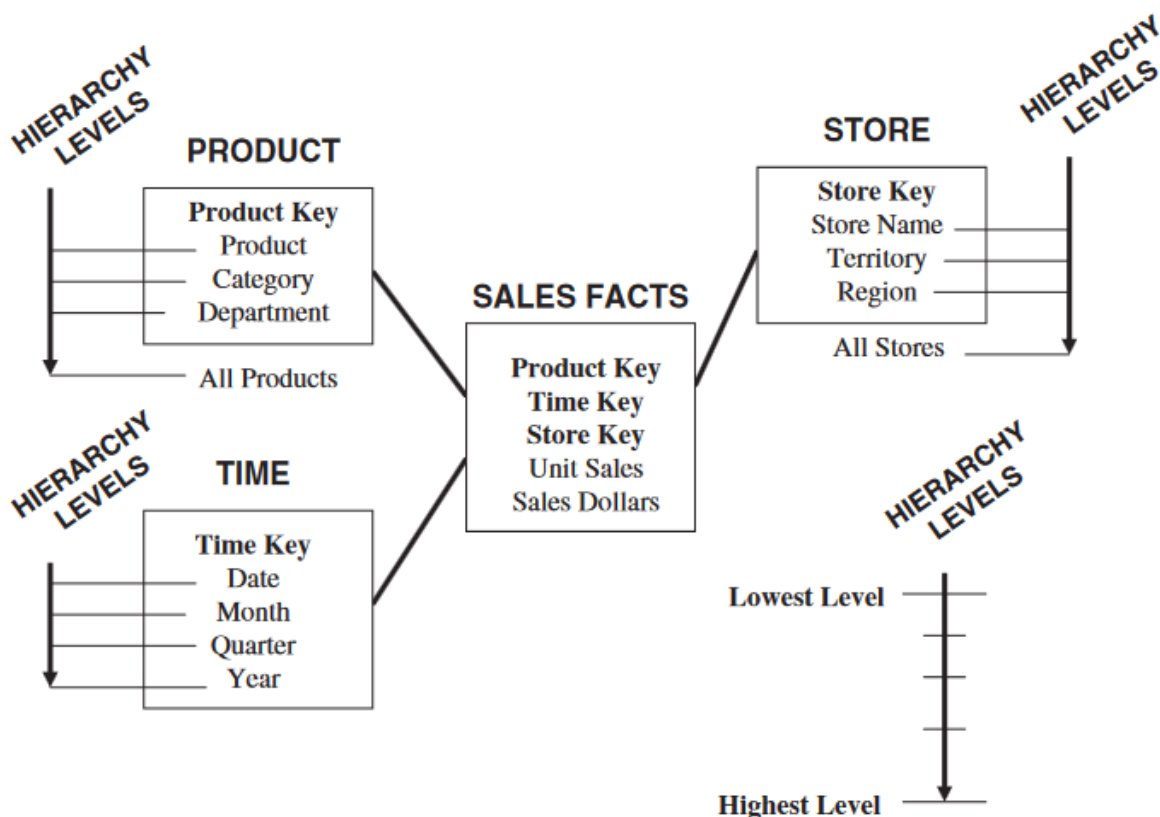


Gambar 11.12 Skema STAR: rantai grosir.

Mengumpulkan Tabel Fakta

Seperti yang telah kita lihat, tabel fakta agregat hanyalah ringkasan dari data paling terperinci pada tingkat yang lebih tinggi sepanjang hierarki dimensi. Gambar 11.13 mengilustrasikan hierarki dalam tiga dimensi. Periksa hierarki dalam tiga dimensi. Tingkat hierarki dalam dimensi waktu bergerak naik dari hari pada tingkat terendah hingga tahun pada tingkat tertinggi. Kota berada pada level terendah dalam dimensi toko dan produk berada pada level terendah dalam dimensi produk.

Dalam tabel fakta dasar, baris mencerminkan angka pada tingkat terendah hierarki dimensi. Misalnya, setiap baris dalam tabel fakta dasar menunjukkan unit penjualan dan dolar penjualan yang berkaitan dengan satu tanggal, satu toko, dan satu produk. Dengan menaikkan satu tingkat atau lebih sepanjang hierarki di setiap dimensi, Anda dapat membuat berbagai tabel fakta agregat. Mari kita jelajahi kemungkinannya.

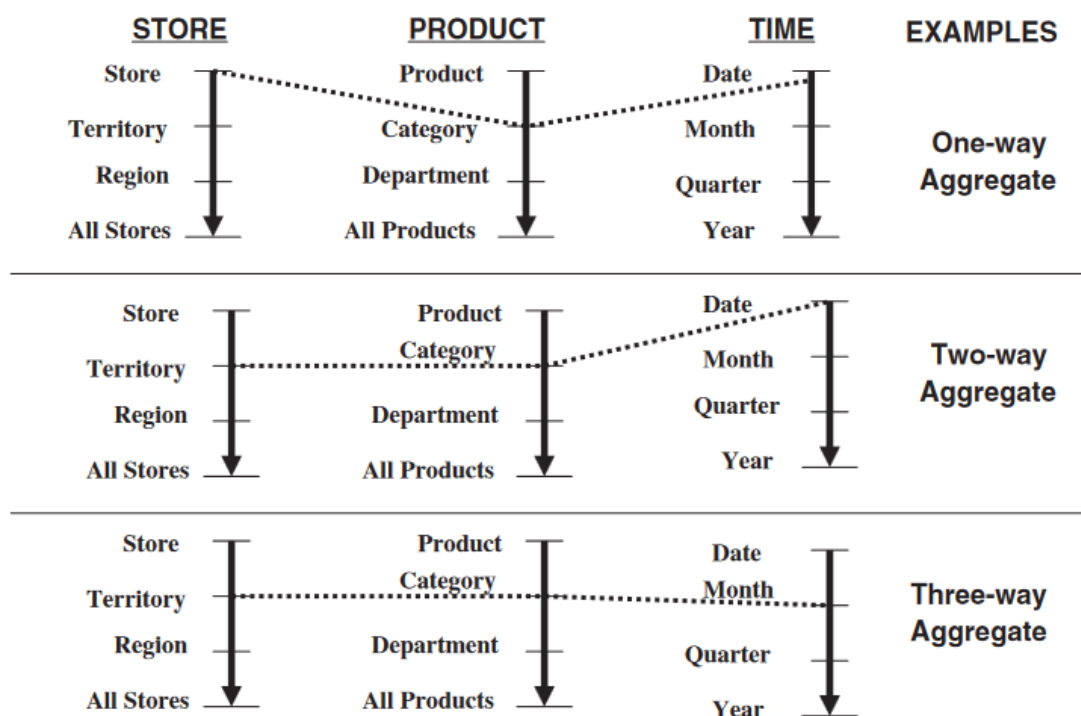


Gambar 11.13 Hierarki dimensi.

Tabel Fakta Agregat Multi-Arah Gambar 11.14 mengilustrasikan berbagai cara pembentukan tabel fakta agregat. Paragraf berikut menjelaskan kemungkinan agregat. Agregat Satu Arah Ketika Anda naik ke level yang lebih tinggi dalam hierarki satu dimensi dan mempertahankan level terendah di dimensi lain, Anda membuat tabel agregat satu arah. Tinjau contoh berikut:

- Kategori produk berdasarkan toko berdasarkan tanggal
- Departemen produk berdasarkan toko berdasarkan tanggal
- Semua produk berdasarkan toko berdasarkan tanggal

- Wilayah berdasarkan produk berdasarkan tanggal
- Wilayah berdasarkan produk berdasarkan tanggal
- Semua toko berdasarkan produk berdasarkan tanggal
- Bulan menurut toko menurut produk
- Seperempat berdasarkan toko berdasarkan produk
- Tahun demi toko berdasarkan produk



Gambar 11.14 Membentuk tabel fakta agregat.

Agregat Dua Arah Ketika Anda naik ke level yang lebih tinggi dalam hierarki dua dimensi dan mempertahankan level terendah di dimensi lain, Anda membuat tabel agregat dua arah. Tinjau contoh berikut:

- ▶ Kategori produk berdasarkan wilayah berdasarkan tanggal
- ▶ Kategori produk berdasarkan wilayah berdasarkan tanggal
- ▶ Kategori produk menurut semua toko berdasarkan tanggal
- ▶ Kategori produk berdasarkan bulan berdasarkan toko
- ▶ Kategori produk per kuartal per toko
- ▶ Kategori produk menurut tahun menurut toko
- ▶ Departemen produk berdasarkan wilayah berdasarkan tanggal
- ▶ Departemen produk berdasarkan wilayah berdasarkan tanggal
- ▶ Departemen produk menurut semua toko berdasarkan tanggal
- ▶ Departemen produk berdasarkan bulan berdasarkan toko
- ▶ Departemen produk per kuartal per toko
- ▶ Departemen produk menurut tahun menurut toko

- ▶ Semua produk berdasarkan wilayah berdasarkan tanggal
- ▶ Semua produk berdasarkan wilayah berdasarkan tanggal
- ▶ Semua produk menurut semua toko berdasarkan tanggal
- ▶ Semua produk berdasarkan bulan berdasarkan toko
- ▶ Semua produk per kuartal per toko
- ▶ Semua produk berdasarkan tahun berdasarkan toko
- ▶ Distrik berdasarkan bulan berdasarkan produk
- ▶ Distrik per kuartal berdasarkan produk
- ▶ Distrik berdasarkan tahun berdasarkan produk
- ▶ Wilayah berdasarkan bulan berdasarkan produk
- ▶ Wilayah per kuartal berdasarkan produk
- ▶ Wilayah berdasarkan tahun berdasarkan produk
- ▶ Wilayah berdasarkan bulan berdasarkan produk
- ▶ Wilayah per kuartal berdasarkan produk
- ▶ Wilayah berdasarkan tahun berdasarkan produk
- ▶ Semua toko berdasarkan bulan berdasarkan produk
- ▶ Semua toko per kuartal berdasarkan produk
- ▶ Semua toko berdasarkan tahun berdasarkan produk

Agregat Tiga Arah Ketika Anda naik ke tingkat yang lebih tinggi dalam hierarki ketiga dimensi, Anda membuat tabel agregat tiga arah. Tinjau contoh berikut:

- Kategori produk berdasarkan wilayah berdasarkan bulan
- Departemen produk berdasarkan wilayah berdasarkan bulan
- Semua produk berdasarkan wilayah berdasarkan bulan
- Kategori produk berdasarkan wilayah berdasarkan bulan
- Departemen produk berdasarkan wilayah berdasarkan bulan
- Semua produk berdasarkan wilayah berdasarkan bulan
- Kategori produk menurut semua toko berdasarkan bulan
- Departemen produk menurut semua toko berdasarkan bulan
- Kategori produk berdasarkan wilayah per kuartal
- Departemen produk berdasarkan wilayah per kuartal
- Semua produk berdasarkan wilayah per kuartal
- Kategori produk berdasarkan wilayah per kuartal
- Departemen produk berdasarkan wilayah per kuartal
- Semua produk berdasarkan wilayah per kuartal
- Kategori produk menurut semua toko per kuartal
- Departemen produk menurut semua toko per kuartal
- Kategori produk berdasarkan wilayah berdasarkan tahun
- Departemen produk berdasarkan wilayah berdasarkan tahun
- Semua produk berdasarkan wilayah berdasarkan tahun
- Kategori produk berdasarkan wilayah berdasarkan tahun

- Departemen produk menurut wilayah menurut tahun
- Semua produk berdasarkan wilayah berdasarkan tahun
- Kategori produk menurut semua toko menurut tahun
- Departemen produk menurut semua toko berdasarkan tahun
- Semua produk menurut semua toko berdasarkan tahun

Masing-masing tabel fakta agregat ini berasal dari tabel fakta dasar tunggal. Tabel fakta agregat turunan digabungkan ke satu atau lebih tabel dimensi turunan. Gambar 11.15 menunjukkan tabel fakta agregat turunan yang dihubungkan ke tabel dimensi turunan.

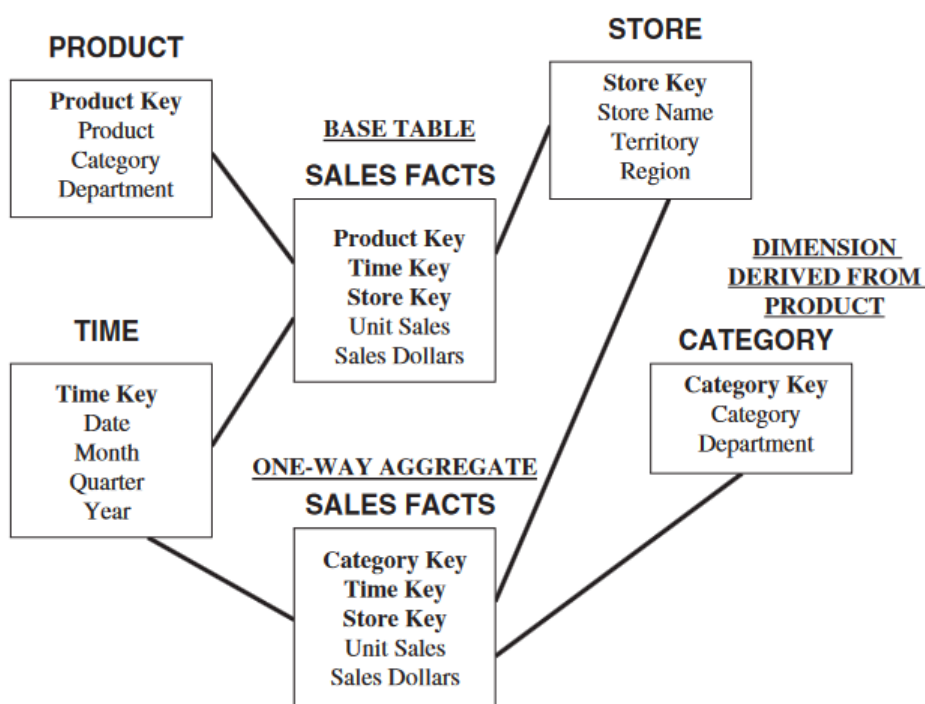
Pengaruh Ketersebaran pada Agregasi Pertimbangkan kasus jaringan toko kelontong dengan 300 toko, 40.000 produk di setiap toko, namun hanya 4.000 produk yang terjual di setiap toko dalam sehari. Seperti yang telah dibahas sebelumnya, dengan asumsi Anda menyimpan catatan selama 5 tahun atau 1825 hari, jumlah maksimum baris tabel fakta dasar dihitung sebagai berikut:

Produk $\frac{1}{4}$ 40.000

Simpan $\frac{1}{4}$ 300

Waktu $\frac{1}{4}$ 1825

Jumlah maksimum baris tabel fakta dasar $\frac{1}{4}$ 22 miliar



Gambar 11.15 Tabel fakta agregat dan tabel dimensi turunan.

Karena hanya 4000 produk yang terjual di setiap toko dalam sehari, tidak semua dari 22 miliar baris ini terisi. Karena ketersebaran ini, hanya 10% baris yang terisi. Oleh karena itu, perkiraan sebenarnya jumlah baris tabel dasar adalah 2 miliar. Sekarang mari kita lihat apa yang terjadi ketika Anda membentuk agregat. Teliti agregat satu arah: total merek berdasarkan toko per hari. Hitung jumlah maksimum baris dalam agregat satu arah ini.

Merek $\frac{1}{4}$ 80

Simpan $\frac{1}{4}$ 300

Waktu $\frac{1}{4}$ 1825

Jumlah maksimum baris tabel agregat $\frac{1}{4}$ 43.800.000

Saat membuat agregat satu arah, Anda akan melihat bahwa ketersebaran agregat ini tidak 10% seperti pada tabel dasar. Hal ini karena bila Anda menggabungkan berdasarkan merek, lebih banyak kode merek yang akan digabungkan dengan kode toko dan waktu. Ketersebaran agregat satu arah akan menjadi sekitar 50%, sehingga menghasilkan perkiraan riil sebesar 21.900.000. Jika ketersebaran tetap sebesar 10% yang berlaku pada tabel dasar, perkiraan sebenarnya jumlah baris dalam tabel agregat akan jauh lebih sedikit.

Jika Anda memilih tingkat agregat yang lebih tinggi, persentase ketersebaran naik dan bahkan mencapai 100%. Karena kegagalan ketersebaran untuk tetap rendah, Anda dihadapkan pada pertanyaan apakah agregat benar-benar meningkatkan kinerja sebanyak itu. Apakah mereka mengurangi jumlah baris secara drastis?

Praktisi data warehousing yang berpengalaman punya saran. Saat Anda membentuk agregat, pastikan bahwa setiap baris tabel agregat merangkum setidaknya 10 baris di tabel tingkat bawah. Jika Anda menambahnya menjadi 20 baris atau lebih, itu akan sangat luar biasa.

Opsi Agregasi

Kembali ke diskusi kita tentang agregat satu arah, dua arah, dan tiga arah untuk skema STAR dasar hanya dengan tiga dimensi, Anda dapat menghitung lebih dari 50 cara berbeda untuk membuat agregat. Di dunia nyata, jumlah dimensinya tidak hanya tiga, tapi masih banyak lagi. Oleh karena itu, jumlah tabel agregat yang mungkin meningkat menjadi ratusan. Lebih lanjut, dari referensi kegagalan ketersebaran pada tabel agregat, Anda mengetahui bahwa proses agregasi tidak mengurangi jumlah baris secara proporsional. Dengan kata lain, jika ketersebaran tabel fakta dasar adalah 10%, ketersebaran tabel agregat tingkat yang lebih tinggi tidak akan tetap sebesar 10%. Persentase ketersebaran semakin meningkat seiring dengan tabel agregat Anda naik lebih tinggi dan lebih tinggi lagi dalam tingkat peringkasan.

Apakah agregasi memang begitu efektif? Apa sajakah pilihannya? Bagaimana Anda memutuskan apa yang akan dikumpulkan? Pertama, tetapkan beberapa sasaran agregasi untuk lingkungan gudang data Anda. Sasaran Strategi Agregasi Terlepas dari tujuan umum keseluruhan untuk meningkatkan kinerja gudang data, berikut adalah beberapa sasaran praktis dan spesifik:

- ❖ Jangan terjebak dengan terlalu banyak agregat. Ingat, Anda juga harus membuat dimensi turunan tambahan untuk mendukung agregat.
- ❖ Cobalah untuk melayani berbagai kelompok pengguna. Bagaimanapun, sediakan untuk pengguna kuat Anda.

- ❖ Gunakan agregat yang tidak terlalu meningkatkan penggunaan penyimpanan secara keseluruhan. Perhatikan baik-baik agregat yang lebih besar dengan persentase ketersebaran yang rendah.
- ❖ Menyembunyikan agregat dari pengguna akhir namun tetap transparan terhadap permintaan pengguna akhir. Artinya, alat kuerilah yang perlu mengetahui agregat untuk mengarahkan kueri agar mendapatkan akses yang tepat.
- ❖ Berusaha untuk menjaga dampak pada proses pementasan data seintensif mungkin.

Saran Praktis Sebelum melakukan penghitungan apa pun untuk menentukan jenis agregat yang diperlukan untuk lingkungan gudang data Anda, luangkan banyak waktu untuk menentukan sifat dari kueri umum. Bagaimana biasanya pengguna Anda melaporkan hasil? Apa tingkat pelaporannya? Berdasarkan toko? Berdasarkan bulan? Berdasarkan kategori produk? Telusuri dimensinya, satu per satu, dan tinjau tingkat hierarkinya. Periksa apakah ada beberapa hierarki dalam dimensi yang sama. Jika ya, cari tahu hierarki mana yang lebih penting. Di setiap dimensi, pastikan atribut mana yang digunakan untuk mengelompokkan metrik tabel fakta. Langkah selanjutnya adalah menentukan atribut mana yang digunakan dalam kombinasi dan kombinasi apa yang paling umum.

Setelah Anda menentukan atribut dan kemungkinan kombinasinya, lihat jumlah nilai untuk setiap atribut. Misalnya, dalam skema jaringan hotel, asumsikan bahwa hotel berada pada level terendah dan kota berada pada level berikutnya yang lebih tinggi dalam dimensi hotel. Katakanlah ada 25.000 nilai untuk hotel dan 15.000 nilai untuk kota. Jelas, tidak ada keuntungan besar jika digabungkan berdasarkan kota. Di sisi lain, jika kota hanya memiliki 500 nilai, maka kota adalah tingkat di mana Anda dapat mempertimbangkan agregasi. Periksa setiap atribut dalam hierarki dalam suatu dimensi. Periksa nilai untuk setiap atribut. Bandingkan nilai atribut pada berbagai tingkat hierarki yang sama dan tentukan mana yang merupakan kandidat kuat untuk berpartisipasi dalam agregasi.

Kembangkan daftar atribut yang merupakan kandidat berguna untuk agregasi, lalu hitung kombinasi atribut ini untuk menghasilkan kumpulan tabel fakta agregat multi-arah yang pertama. Tentukan tabel dimensi turunan yang Anda perlukan untuk mendukung tabel fakta agregat ini. Silakan terapkan tabel fakta agregat ini sebagai kumpulan awal.

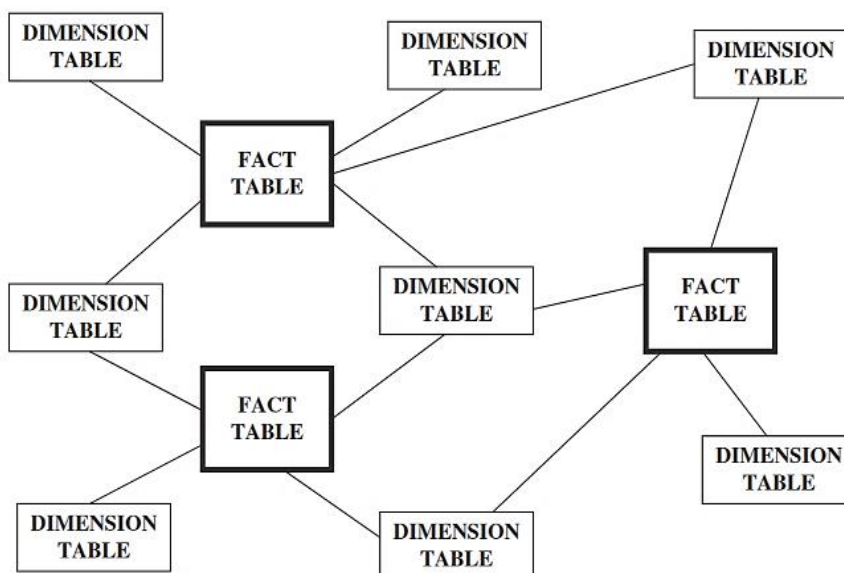
Ingatlah bahwa agregasi adalah mekanisme penyesuaian kinerja. Peningkatan kinerja kueri mendorong kebutuhan untuk meringkas, jadi jangan terlalu khawatir jika rangkaian tabel agregat pertama Anda tidak bekerja dengan sempurna. Agregat Anda dimaksudkan untuk dipantau dan direvisi seperlunya. Sifat sebagian besar permintaan kueri kemungkinan akan berubah. Ketika pengguna Anda menjadi lebih mahir dalam menggunakan gudang data, mereka akan menemukan cara baru untuk mengelompokkan dan menganalisis data. Jadi apa saran praktisnya? Lakukan pekerjaan persiapan Anda. Mulailah dengan serangkaian tabel agregat yang masuk akal, dan terus lakukan penyesuaian seperlunya.

11.4 SKEMA BINTANG

Saat Anda melihat skema STAR tunggal dengan tabel faktanya dan tabel dimensi di sekitarnya, Anda tahu bahwa itu bukanlah sebatas gudang data. Hampir semua gudang data

berisi beberapa struktur skema STAR. Setiap STAR memiliki tujuan tertentu untuk melacak tindakan yang disimpan dalam tabel fakta. Bila Anda memiliki kumpulan skema STAR terkait, Anda dapat menyebut koleksi tersebut sebagai keluarga STARS. Keluarga STARS terbentuk karena berbagai alasan. Anda dapat membentuk sebuah kelompok hanya dengan menambahkan tabel fakta agregat dan tabel dimensi turunan untuk mendukung agregat. Terkadang, Anda dapat membuat tabel fakta inti yang berisi fakta yang menarik bagi sebagian besar pengguna dan tabel fakta yang disesuaikan untuk kelompok pengguna tertentu. Banyak faktor yang menyebabkan keberadaan keluarga STARS. Pertama, lihat contoh yang diberikan pada Gambar 11.16.

Tabel fakta STARS dalam tabel dimensi berbagi keluarga. Biasanya, dimensi waktu digunakan bersama oleh sebagian besar tabel fakta dalam grup. Pada contoh di atas, ketiga tabel fakta cenderung berbagi dimensi waktu. Sebaliknya, tabel dimensi dari beberapa STARS dapat berbagi tabel fakta dari satu STAR.



Gambar 11.16 Keluarga BINTANG.

Jika Anda berkecimpung dalam bisnis seperti perbankan atau layanan telepon, masuk akal untuk menangkap transaksi individu serta snapshot pada interval tertentu. Anda kemudian dapat menggunakan kelompok STARS yang terdiri dari skema transaksi dan snapshot. Jika Anda berada di perusahaan manufaktur atau perusahaan dengan jenis produksi serupa, perusahaan Anda perlu memantau metrik di sepanjang rantai nilai. Beberapa institusi lain seperti pusat kesehatan, di mana nilai tambah tidak diberikan secara berantai tetapi di berbagai stasiun dalam perusahaan. Bagi perusahaan-perusahaan ini, keluarga STARS mendukung rantai nilai atau lingkaran nilai. Kami akan membahas detailnya di beberapa bagian berikutnya.

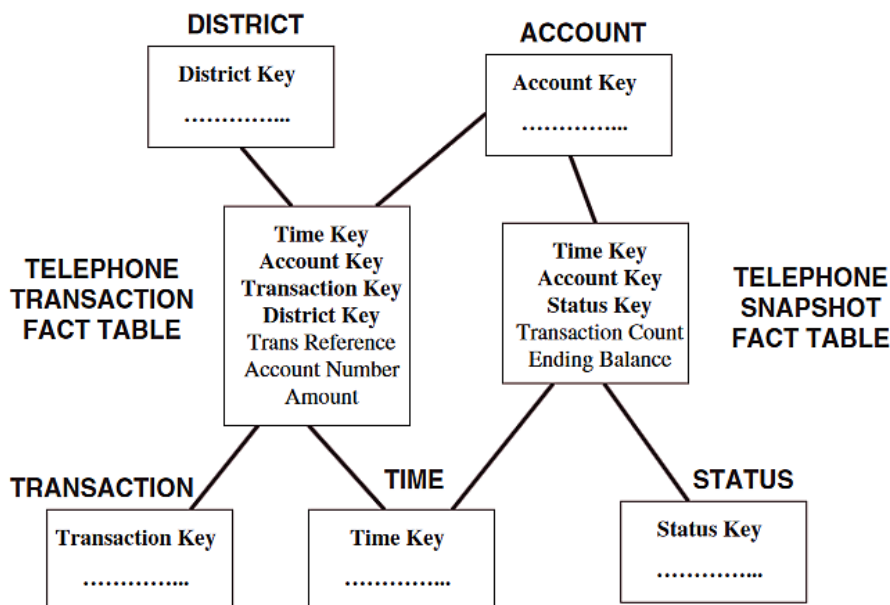
Tabel Snapshot dan Transaksi

Mari kita tinjau beberapa persyaratan dasar perusahaan telepon. Sejumlah transaksi individu membentuk rekening pelanggan telepon. Banyak transaksi terjadi pada jam 06.00 hingga 22.00. hari pelanggan. Transaksi lebih banyak terjadi pada hari libur dan akhir pekan untuk pelanggan residensial. Pelanggan institusi menggunakan telepon pada hari kerja, bukan pada akhir pekan. Sebuah perusahaan telepon mengumpulkan kumpulan data transaksi kaya yang sangat besar yang dapat digunakan untuk berbagai jenis analisis berharga. Perusahaan telepon memerlukan skema yang menangkap data transaksi yang mendukung pengambilan keputusan strategis untuk perluasan, peningkatan layanan baru, dan sebagainya. Skema transaksi ini menjawab pertanyaan seperti bagaimana pendapatan jam sibuk pada akhir pekan dan hari libur dibandingkan dengan jam sibuk pada hari kerja.

Selain itu, perusahaan telepon perlu menjawab pertanyaan pelanggan mengenai saldo rekening. Departemen layanan pelanggan terus-menerus dibombardir dengan pertanyaan tentang status masing-masing akun pelanggan. Secara berkala, departemen akuntansi mungkin tertarik dengan jumlah yang diperkirakan akan diterima pada pertengahan bulan depan. Berapa saldo terutang yang tagihannya akan dikirimkan akhir bulan ini? Untuk tujuan ini, perusahaan telepon memerlukan skema untuk menangkap snapshot secara berkala. Gambar 11.17 memperlihatkan snapshot dan tabel fakta transaksi untuk perusahaan telepon. Catatlah atribut-atribut pada kedua tabel fakta tersebut. Satu tabel melacak transaksi telepon individual. Tabel lainnya menyimpan cuplikan akun individual pada interval tertentu. Perhatikan juga bagaimana tabel dimensi dibagi antara dua tabel fakta.

Snapshot dan tabel transaksi juga umum terjadi di bank. Misalnya, tabel transaksi ATM menyimpan transaksi ATM individual. Tabel fakta ini melacak jumlah transaksi individual untuk akun pelanggan. Tabel snapshot menyimpan saldo untuk setiap akun pada akhir setiap hari. Kedua tabel tersebut mempunyai dua fungsi yang berbeda. Dari tabel transaksi, Anda dapat melakukan berbagai jenis analisis transaksi ATM. Tabel snapshot menyediakan jumlah total yang disimpan pada interval periodik yang menunjukkan pergeseran dan pergerakan saldo.

Gudang data keuangan juga memerlukan snapshot dan tabel transaksi karena sifat analisis dalam kasus ini. Rangkaian pertanyaan pertama untuk gudang ini berkaitan dengan transaksi yang memengaruhi akun tertentu selama periode waktu tertentu. Rangkaian pertanyaan lainnya berpusat pada saldo masing-masing akun pada interval tertentu atau total kelompok akun pada akhir periode tertentu. Tabel transaksi menjawab pertanyaan set pertama; tabel snapshot menangani pertanyaan set kedua.



Gambar 11.17 Snapshot dan tabel transaksi.

Tabel Inti dan Tabel Kustom

Perhatikan dua jenis bisnis yang tampaknya berbeda. Pertama ambil kasus bank. Bank menawarkan berbagai macam layanan yang semuanya berkaitan dengan keuangan dalam satu atau lain bentuk. Sebagian besar layanan berbeda satu sama lain. Layanan rekening giro dan layanan rekening tabungan serupa dalam banyak hal. Namun layanan rekening tabungan sama sekali tidak menyerupai layanan kartu kredit. Bagaimana Anda melacak layanan yang berbeda ini?

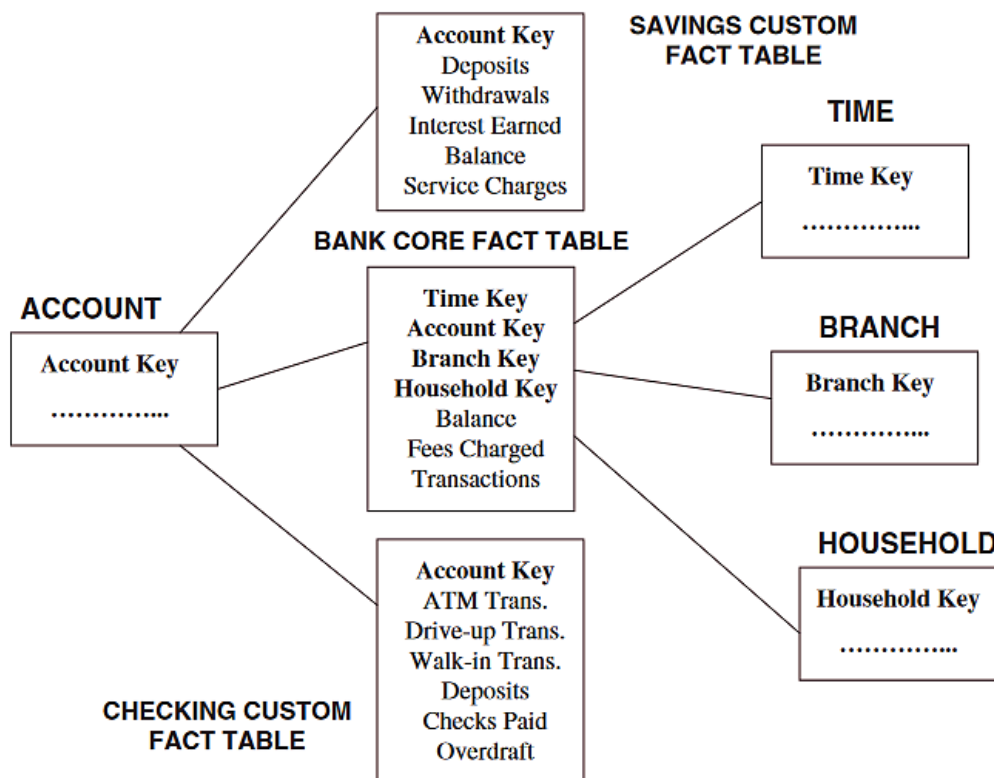
Selanjutnya, pertimbangkan sebuah perusahaan manufaktur yang memproduksi sejumlah produk heterogen. Meskipun beberapa faktor mungkin sama pada berbagai produk, pada umumnya faktor-faktor tersebut berbeda. Apa yang harus Anda lakukan untuk mendapatkan informasi tentang produk yang heterogen?

Jenis keluarga STARS yang berbeda memenuhi persyaratan perusahaan-perusahaan ini. Dalam jenis kelompok ini, semua produk dan layanan terhubung ke tabel fakta inti dan setiap produk atau layanan terkait dengan tabel kustom individual. Pada Gambar 11.18, Anda akan melihat tabel inti dan tabel kustom untuk sebuah bank. Perhatikan bagaimana tabel fakta inti menyimpan metrik yang umum untuk semua jenis akun. Setiap tabel fakta khusus berisi metrik khusus untuk lini layanan tersebut. Perhatikan juga dimensi bersama dan perhatikan bagaimana tabel membentuk keluarga STARS.

Mendukung Rantai Nilai Perusahaan atau Value Circle

Dalam bisnis manufaktur, suatu produk melewati berbagai tahapan, dimulai dari bahan mentah dan berakhir sebagai barang jadi di inventaris gudang. Biasanya langkah-langkahnya meliputi penambahan bahan, perakitan bahan, pengendalian proses, pengemasan, dan pengiriman ke gudang. Dari persediaan barang jadi, suatu produk berpindah ke pengiriman ke distributor, persediaan distributor, pengiriman distributor, persediaan eceran, dan penjualan eceran. Pada setiap langkah, nilai tambah pada produk.

Beberapa sistem operasional mendukung alur melalui langkah-langkah ini. Keseluruhan aliran membentuk rantai pasokan atau rantai nilai. Demikian pula, dalam perusahaan asuransi, rantai nilai dapat mencakup sejumlah langkah mulai dari penjualan asuransi hingga penerbitan polis dan akhirnya pemrosesan klaim. Dalam hal ini, rantai nilai berkaitan dengan layanan.



Gambar 11.18 Tabel inti dan tabel kustom.

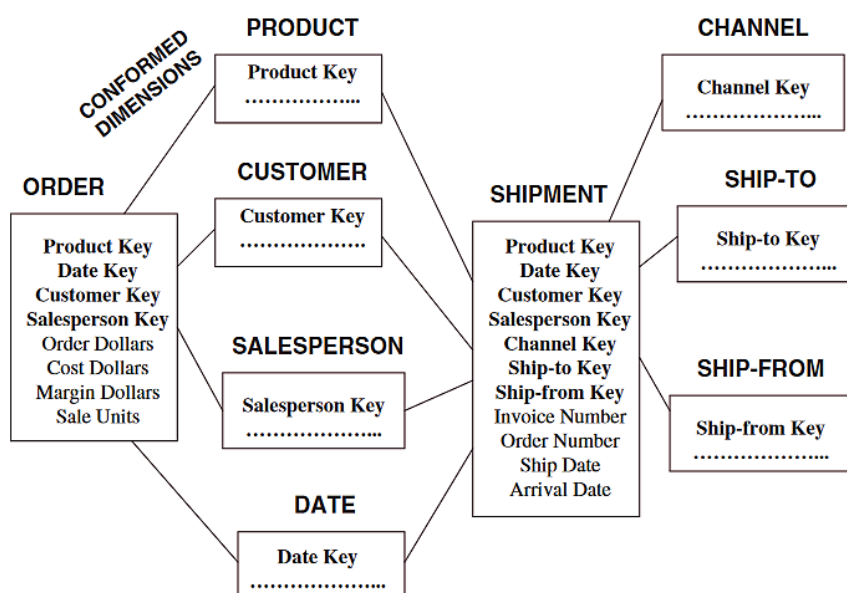
Jika Anda menjalankan salah satu bisnis ini, Anda perlu melacak metrik penting di berbagai langkah di sepanjang rantai nilai. Anda membuat skema STAR untuk langkah-langkah penting dan rangkaian lengkap skema terkait membentuk keluarga STARS. Anda menentukan tabel fakta dan serangkaian dimensi yang sesuai untuk setiap langkah penting dalam rantai. Jika perusahaan Anda memiliki beberapa rantai nilai, maka Anda harus mendukung setiap rantai dengan rangkaian STARS yang terpisah.

Rantai pasok atau rantai nilai berjalan secara linier, dimulai dengan langkah tertentu dan berakhir pada langkah berikutnya dengan banyak langkah di antaranya. Sekali lagi, pada setiap langkah, nilai tambah. Di beberapa jenis bisnis lain yang memberikan nilai tambah pada jasa, pergerakan linier serupa tidak terjadi. Misalnya saja sebuah institusi layanan kesehatan yang memberikan nilai tambah pada layanan pasien dari unit-unit yang berbeda seolah-olah mereka membentuk lingkaran di sekitar layanan tersebut. Kami merasakan lingkaran nilai dalam organisasi semacam itu. Lingkaran nilai dari organisasi pemeliharaan kesehatan yang besar dapat mencakup rumah sakit, klinik, kantor dokter, apotek, laboratorium, lembaga pemerintah, dan perusahaan asuransi. Masing-masing unit ini menyediakan perawatan

pasien atau mengukur perawatan pasien. Perawatan pasien oleh setiap unit dapat diukur dalam metrik yang berbeda. Namun sebagian besar unit akan menganalisis metrik menggunakan serangkaian dimensi yang sama seperti waktu, pasien, penyedia layanan kesehatan, pengobatan, diagnosis, dan pembayar. Untuk lingkaran nilai, keluarga STARS terdiri dari beberapa tabel fakta dan serangkaian dimensi yang disesuaikan.

Dimensi yang Sesuai

Saat menjelajahi keluarga STARS, Anda akan memperhatikan bahwa dimensi dibagi di antara tabel fakta. Dimensi membentuk hubungan umum antar BINTANG. Agar dimensi dapat disesuaikan, Anda harus dengan sengaja memastikan bahwa dimensi yang sama dapat digunakan antara dua BINTANG yang lebih banyak. Jika dimensi produk dibagi antara dua tabel fakta penjualan dan inventaris, maka atribut dimensi produk harus mempunyai arti yang sama dalam kaitannya dengan masing-masing dua tabel fakta tersebut. Gambar 11.19 menunjukkan sekumpulan dimensi yang disesuaikan.

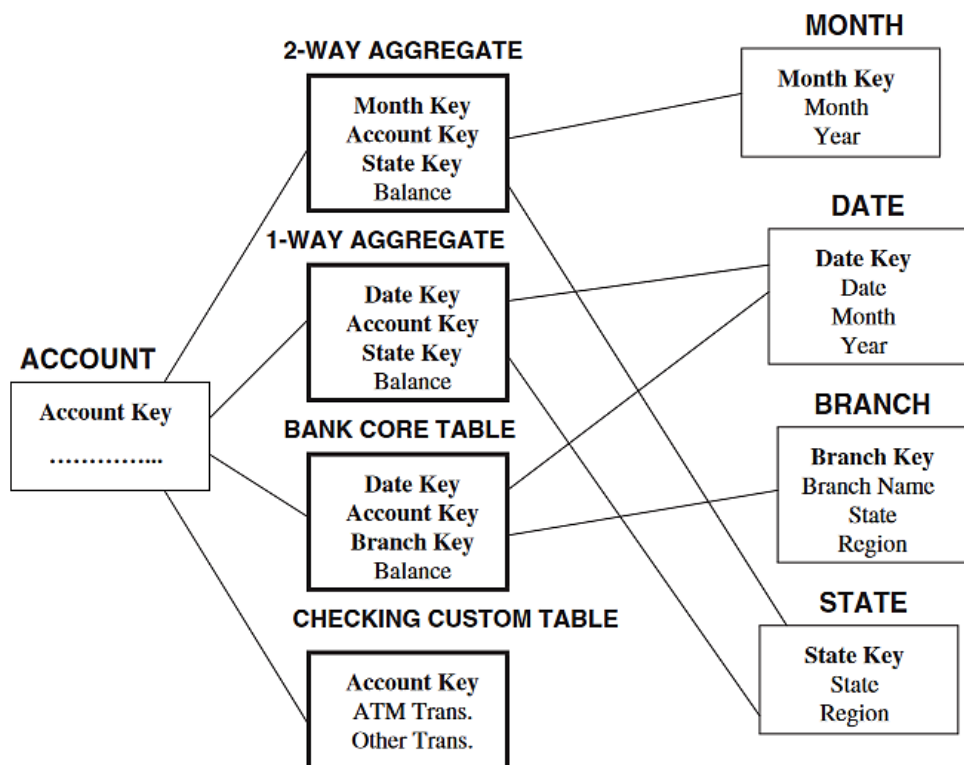


Gambar 11.19 Dimensi yang sesuai.

Tabel fakta pesanan dan pengiriman berbagi dimensi produk, tanggal, pelanggan, dan tenaga penjualan yang sesuai. Dimensi yang disesuaikan adalah kombinasi atribut yang komprehensif dari sistem sumber setelah menyelesaikan semua perbedaan dan konflik. Misalnya, dimensi produk yang disesuaikan harus benar-benar mencerminkan daftar produk utama perusahaan dan harus mencakup semua hierarki yang mungkin. Setiap atribut harus memiliki tipe data yang benar dan harus memiliki panjang serta batasan yang sesuai.

Menyesuaikan dimensi adalah persyaratan dasar dalam gudang data. Berikan perhatian khusus dan ambil langkah-langkah yang diperlukan untuk menyesuaikan semua dimensi Anda. Ini adalah tanggung jawab utama tim proyek. Dimensi yang disesuaikan memungkinkan penggabungan di seluruh data mart. Antarmuka pengguna akan konsisten

apa pun jenis kuerinya. Kumpulan hasil kueri akan konsisten di seluruh data mart. Tentu saja, satu dimensi yang disesuaikan dapat digunakan terhadap beberapa tabel fakta.



Gambar 11.20 Rangkaian lengkap STARS.

Standarisasi Fakta

Selain tugas menyesuaikan dimensi adalah persyaratan untuk membakukan fakta. Kita telah melihat bahwa tabel fakta berfungsi di seluruh skema STAR. Tinjau isu-isu berikut yang berkaitan dengan standarisasi atribut tabel fakta:

- ✓ Pastikan definisi dan terminologi yang sama di seluruh data mart.
- ✓ Menyelesaikan homonim dan sinonim.
- ✓ Jenis fakta yang akan distandarisasi meliputi pendapatan, harga, biaya, dan margin keuntungan.
- ✓ Menjamin bahwa algoritma yang sama digunakan untuk setiap unit turunan di setiap tabel fakta.
- ✓ Pastikan setiap fakta menggunakan satuan pengukuran yang tepat.

Ringkasan Keluarga STARS

Mari kita akhiri diskusi kita tentang keluarga STARS dengan diagram komprehensif yang menunjukkan serangkaian tabel fakta standar dan tabel dimensi yang disesuaikan. Pelajari Gambar 11.20 dengan cermat. Perhatikan tabel fakta agregat dan tabel dimensi turunan yang terkait. Jenis agregat apakah ini? Manakah tabel fakta dasarnya? Perhatikan dimensi bersama. Apakah dimensi ini sesuai? Lihat bagaimana berbagai tabel fakta dan tabel dimensi saling terkait.

RINGKASAN BAB

- Dimensi yang berubah secara perlahan dapat diklasifikasikan menjadi tiga jenis berdasarkan sifat perubahannya. Tipe 1 berkaitan dengan koreksi, tipe 2 berkaitan dengan pelestarian sejarah, dan tipe 3 berkaitan dengan revisi lunak. Penerapan setiap jenis revisi pada data warehouse berbeda-beda.
- Tabel berdimensi besar seperti pelanggan atau produk memerlukan pertimbangan khusus untuk menerapkan teknik pengoptimalan.
- “Snowflaking” atau membuat skema kepingan salju adalah metode normalisasi skema STAR. Meskipun beberapa kondisi membenarkan skema kepingan salju, hal ini umumnya tidak disarankan.
- Berbagai macam flag dan data tekstual disatukan dalam satu tabel yang disebut tabel dimensi sampah.
- Tabel agregat atau ringkasan meningkatkan kinerja. Merumuskan strategi untuk membangun tabel agregat.
- Seperangkat skema STAR yang terkait membentuk keluarga STARS. Contohnya adalah tabel snapshot dan transaksi, tabel inti dan tabel kustom, serta tabel yang mendukung rantai nilai atau lingkaran nilai. Keluarga STARS bergantung pada tabel dimensi yang disesuaikan dan tabel fakta yang terstandarisasi.

LATIHAN SOAL

1. Jelaskan perubahan dimensi secara perlahan. Apa sajakah ketiga jenis tersebut? Jelaskan setiap jenis dengan sangat singkat.
2. Bandingkan dan kontraskan tipe 2 dan tipe 3 yang perlahan-lahan berubah dimensinya.
3. Bisakah Anda memperlakukan dimensi yang berubah dengan cepat dengan cara yang sama seperti tipe 2 yang mengubah dimensi secara perlahan? Membahas.
4. Apa yang dimaksud dengan dimensi sampah? Apakah mereka diperlukan di gudang data?
5. Apa perbedaan skema kepingan salju dengan skema STAR? Sebutkan dua kelebihan dan dua kelemahan skema kepingan salju.
6. Membedakan dimensi yang berubah secara perlahan dan cepat.
7. Apa yang dimaksud dengan tabel fakta agregat? Mengapa mereka dibutuhkan? Berikan contoh.
8. Jelaskan dengan contoh snapshot dan tabel fakta transaksi. Bagaimana hubungan mereka?
9. Berikan contoh lingkaran nilai. Jelaskan bagaimana keluarga STARS dapat mendukung lingkaran nilai.
10. Apa yang dimaksud dengan penyesuaian dimensi? Mengapa hal ini penting dalam gudang data?

LATIHAN TAMBAHAN

1. Tunjukkan apakah benar atau salah:
 - A. Perubahan tipe 1 untuk perubahan dimensi secara perlahan berhubungan dengan koreksi kesalahan.
 - B. Untuk menerapkan perubahan tipe 3 dimensi yang berubah secara perlahan, timpa nilai atribut pada baris tabel dimensi dengan nilai baru.
 - C. Dimensi besar biasanya memiliki banyak hierarki.
 - D. Skema STAR adalah versi normal dari skema kepingan salju.
 - E. Agregat adalah ringkasan yang telah dihitung sebelumnya.
 - F. Persentase ketersebaran tabel dasar cenderung lebih tinggi dibandingkan tabel agregat.
 - G. Tabel fakta STARS dalam tabel dimensi berbagi keluarga.
 - H. Tabel fakta inti dan khusus berguna bagi perusahaan dengan beberapa lini layanan.
 - I. Menyesuaikan dimensi tidak mutlak diperlukan dalam gudang data.
 - J. Lingkaran nilai biasanya membutuhkan keluarga STARS untuk mendukung bisnisnya.
2. Asumsikan Anda berkecimpung dalam bisnis asuransi. Temukan dua contoh tipe 2 yang secara perlahan mengubah dimensi dalam bisnis tersebut. Sebagai seorang analis proyek, tuliskan spesifikasi untuk menerapkan perubahan tipe 2 pada gudang data dengan memperhatikan dua contoh.
3. Anda adalah spesialis desain data di tim proyek gudang data untuk perusahaan ritel. Rancang skema STAR untuk melacak unit penjualan dan dolar penjualan dengan tabel tiga dimensi. Jelaskan bagaimana Anda akan memutuskan untuk memilih dan membangun empat agregat dua arah.
4. Sebagai perancang data untuk bank internasional, pertimbangkan kemungkinan jenis snapshot dan tabel transaksi. Selesaikan desain satu snapshot dan satu tabel transaksi.
5. Untuk perusahaan manufaktur, rancang rangkaian tiga STARS untuk mendukung rantai nilai.

BAB 12

EKSTRAKSI, TRANSFORMASI, DAN PEMUATAN DATA

TUJUAN BAB

- Melakukan survei secara luas terhadap berbagai aspek fungsi ekstraksi, transformasi, dan pemuatan data (ETL).
- Menguji fungsi ekstraksi data, tantangannya, tekniknya, dan mempelajari cara mengevaluasi dan menerapkan teknik tersebut
- Membahas berbagai tugas dan jenis fungsi transformasi data
- Memahami arti integrasi dan konsolidasi data
- Memahami pentingnya fungsi pemuatan data dan menyelidiki metode utama untuk menerapkan data ke gudang
- Dapatkan pemahaman sebenarnya mengapa ETL penting, memakan waktu, dan sulit

Anda mungkin yakin bahwa data dalam sistem operasional organisasi Anda sama sekali tidak memadai untuk menyediakan informasi bagi pengambilan keputusan strategis. Sebagai profesional di bidang teknologi informasi, kami sepenuhnya menyadari upaya sia-sia dalam beberapa dekade terakhir dalam menyediakan informasi strategis dari sistem operasional. Upaya ini tidak berhasil. Pergudangan data telah mulai memenuhi kebutuhan mendesak akan informasi strategis.

Sebagian besar, informasi yang terkandung dalam gudang mengalir dari sistem operasional yang sama yang tidak dapat langsung digunakan untuk menyediakan informasi strategis. Apa perbedaan antara data dalam sistem operasional sumber dan informasi dalam gudang data? Ini adalah serangkaian fungsi yang termasuk dalam kelompok besar ekstraksi data, transformasi, dan pemuatan (ETL).

Fungsi ETL membentuk kembali data yang relevan dari sistem sumber menjadi informasi berguna untuk disimpan di gudang data. Tanpa fungsi-fungsi ini, tidak akan ada informasi strategis di data warehouse. Jika sumber data tidak diekstraksi dengan benar, dibersihkan, dan diintegrasikan dalam format yang tepat, pemrosesan kueri dan penyampaian intelijen bisnis, tulang punggung gudang data, tidak dapat terjadi.

Pada Bab 2, ketika kita membahas blok bangunan data warehouse, kita secara singkat melihat fungsi ETL sebagai bagian dari data staging area. Pada Bab 6 kita meninjau kembali fungsi ETL dan memeriksa bagaimana kebutuhan bisnis mendorong fungsi-fungsi ini juga. Lebih lanjut, di Bab 8, kita mengeksplorasi opsi infrastruktur perangkat keras dan perangkat lunak untuk mendukung fungsi pergerakan data. Lalu, mengapa peninjauan tambahan terhadap ETL diperlukan?

Fungsi ETL membentuk prasyarat untuk konten informasi gudang data. Fungsi ETL patut mendapat lebih banyak pertimbangan dan diskusi. Dalam bab ini, kita akan mempelajari lebih dalam isu-isu yang berkaitan dengan fungsi ETL. Kami akan meninjau banyak aktivitas penting dalam ETL. Pada bab berikutnya, kita perlu melanjutkan diskusi dengan mempelajari

fungsi penting lainnya yang termasuk dalam lingkup ETL secara keseluruhan kualitas data. Sekarang, mari kita mulai dengan gambaran umum tentang ETL.

12.1 GAMBARAN UMUM ETL

Jika Anda mengingat pembahasan kita tentang fungsi dan layanan arsitektur teknis gudang data, Anda akan melihat bahwa kita membagi lingkungan menjadi tiga area fungsional. Area-area ini adalah akuisisi data, penyimpanan data, dan pengiriman informasi. Ekstraksi data, transformasi data, dan pemuatan data mencakup bidang akuisisi data dan penyimpanan data. Ini adalah proses back-end yang mencakup ekstraksi data dari sistem sumber. Selanjutnya, mereka mencakup semua fungsi dan prosedur untuk mengubah data sumber ke dalam format dan struktur yang sesuai untuk penyimpanan dalam database gudang data. Setelah transformasi data, proses ini terdiri dari semua fungsi untuk memindahkan data secara fisik ke dalam repositori data warehouse.

Ekstraksi data, tentu saja, mendahului semua fungsi lainnya. Namun sejauh mana cakupan dan cakupan data yang akan Anda ekstrak dari sistem sumber? Tidakkah menurut Anda pengguna gudang data Anda tertarik pada semua data operasional untuk beberapa jenis kueri atau analisis? Jadi, mengapa tidak mengekstrak semua data operasional dan membuangnya ke gudang data? Tampaknya ini merupakan pendekatan yang mudah. Namun, pendekatan ekstraksi data ditentukan oleh kebutuhan pengguna. Definisi kebutuhan Anda harus memandu Anda mengenai data apa yang perlu Anda ekstrak dan dari sistem sumber mana. Hindari membuat data junkhouse dengan membuang semua data yang tersedia dari sistem sumber dan menunggu untuk melihat apa yang akan dilakukan pengguna terhadap data tersebut. Ekstraksi data melibatkan proses seleksi. Pilih data yang dibutuhkan berdasarkan kebutuhan pengguna.

Luas dan kompleksitas proses back-end berbeda dari satu gudang data ke gudang data lainnya. Jika perusahaan Anda didukung oleh sejumlah besar sistem operasional yang berjalan pada beberapa platform komputasi, proses back-end dalam kasus Anda akan sangat luas dan mungkin juga rumit. Jadi, dalam situasi Anda, ekstraksi data menjadi cukup menantang. Fungsi transformasi data dan pemuatan data mungkin juga sama sulitnya. Terlebih lagi, jika kualitas data sumber berada di bawah standar, kondisi ini akan semakin memperparah proses back-end. Selain tantangan-tantangan ini, jika hanya sedikit metode pemuatan yang sesuai untuk situasi Anda, maka pemuatan data juga bisa menjadi sulit. Mari kita bahas secara spesifik tentang sifat fungsi ETL.

Paling Penting dan Paling Menantang

Masing-masing fungsi ETL memenuhi tujuan penting. Saat Anda ingin mengubah data dari sistem sumber menjadi informasi bisnis untuk disimpan di gudang data, masing-masing fungsi ini penting. Untuk mengubah data menjadi informasi, Anda perlu menangkap datanya terlebih dahulu. Setelah Anda mengambil data, Anda tidak bisa begitu saja membuang data tersebut ke gudang data dan menyebutnya sebagai informasi strategis. Anda harus melakukan segala macam transformasi pada data yang diekstraksi sehingga data tersebut cocok untuk diubah menjadi informasi strategis. Setelah Anda mengubah data, data tersebut masih belum

berguna bagi pengguna akhir sampai data tersebut dipindahkan ke repositori gudang data. Pemuatan data adalah fungsi penting. Anda harus menjalankan ketiga fungsi ETL agar berhasil mengubah data menjadi informasi strategis atau intelijen bisnis. Ambil contoh analisis yang ingin dilakukan pengguna Anda.

Pengguna ingin membandingkan dan menganalisis penjualan berdasarkan toko, produk, dan bulan. Angka penjualan tersedia di beberapa aplikasi penjualan di perusahaan Anda. Selain itu, Anda memiliki file master produk. Selanjutnya, setiap transaksi penjualan mengacu pada toko tertentu. Semua ini adalah potongan data dalam sistem operasional sumber. Untuk melakukan analisis, Anda harus memberikan informasi tentang penjualan di database gudang data. Anda harus menyediakan unit penjualan dan dolar dalam tabel fakta, produk dalam tabel dimensi produk, jumlah toko dalam tabel dimensi toko, dan bulan dalam tabel dimensi waktu. Bagaimana kamu melakukan ini? Ekstrak data dari masing-masing sistem operasional, rekonsiliasi variasi representasi data di antara sistem sumber, dan transformasikan semua penjualan semua produk. Kemudian memuat penjualan ke dalam tabel fakta dan dimensi. Kini, setelah ketiga fungsi ini selesai, data yang diekstraksi disimpan di gudang data, diubah menjadi informasi strategis, siap dikirimkan sebagai intelijen bisnis untuk dianalisis. Perhatikan bahwa penting bagi setiap fungsi untuk dilakukan, dan dilakukan secara berurutan.

Fungsi ETL merupakan tantangan terutama karena sifat dari sistem sumber. Sebagian besar tantangan dalam ETL muncul dari kesenjangan antar sistem operasional sumber. Tinjau daftar berikut alasan jenis kesulitan dalam fungsi ETL. Pertimbangkan masing-masing dengan hati-hati dan kaitkan dengan lingkungan Anda sehingga Anda dapat menemukan resolusi yang tepat.

- ❖ Sistem sumber sangat beragam dan berbeda.
- ❖ Biasanya terdapat kebutuhan untuk menangani sistem sumber pada berbagai platform dan sistem operasi yang berbeda.
- ❖ Banyak sistem sumber merupakan aplikasi lama yang berjalan pada teknologi database yang sudah ketinggalan zaman.
- ❖ Secara umum, data historis mengenai perubahan nilai tidak disimpan dalam sistem operasional sumber. Informasi historis sangat penting dalam gudang data.
- ❖ Kualitas data diragukan di banyak sistem sumber lama yang telah berkembang seiring berjalannya waktu.
- ❖ Struktur sistem sumber terus berubah seiring waktu karena kondisi bisnis baru. Fungsi ETL juga harus dimodifikasi.
- ❖ Kurangnya konsistensi antar sistem sumber sering terjadi. Data yang sama kemungkinan besar direpresentasikan secara berbeda di berbagai sistem sumber. Misalnya, data gaji dapat direpresentasikan sebagai gaji bulanan, gaji mingguan, dan gaji dua bulanan dalam sistem penggajian sumber yang berbeda.
- ❖ Bahkan ketika data yang tidak konsisten terdeteksi di antara sistem sumber yang berbeda, kurangnya cara untuk menyelesaikan ketidaksesuaian akan semakin memperparah masalah ketidakkonsistenan.

- ❖ Kebanyakan sistem sumber tidak merepresentasikan data dalam tipe atau format yang berguna bagi pengguna. Banyak representasi yang samar dan ambigu.

Memakan Waktu dan Sulit

Ketika tim proyek merancang fungsi ETL, menguji berbagai proses, dan menerapkannya, Anda akan menemukan bahwa fungsi ini menghabiskan persentase yang sangat tinggi dari total upaya proyek. Bukan hal yang aneh jika tim proyek menghabiskan 50% hingga 70% upaya proyek untuk fungsi ETL. Anda telah mencatat beberapa faktor yang menambah kompleksitas fungsi ETL.

Ekstraksi data itu sendiri bisa sangat rumit tergantung pada sifat dan kompleksitas sistem sumber. Metadata pada sistem sumber harus berisi informasi pada setiap database dan setiap struktur data yang diperlukan dari sistem sumber. Anda memerlukan informasi yang sangat detail, termasuk ukuran database dan volatilitas data. Anda harus mengetahui jangka waktu setiap hari ketika Anda dapat mengekstrak data tanpa memengaruhi penggunaan sistem operasional. Anda juga perlu menentukan mekanisme untuk menangkap perubahan data di setiap sistem sumber yang relevan. Ini adalah aktivitas yang berat dan memakan waktu.

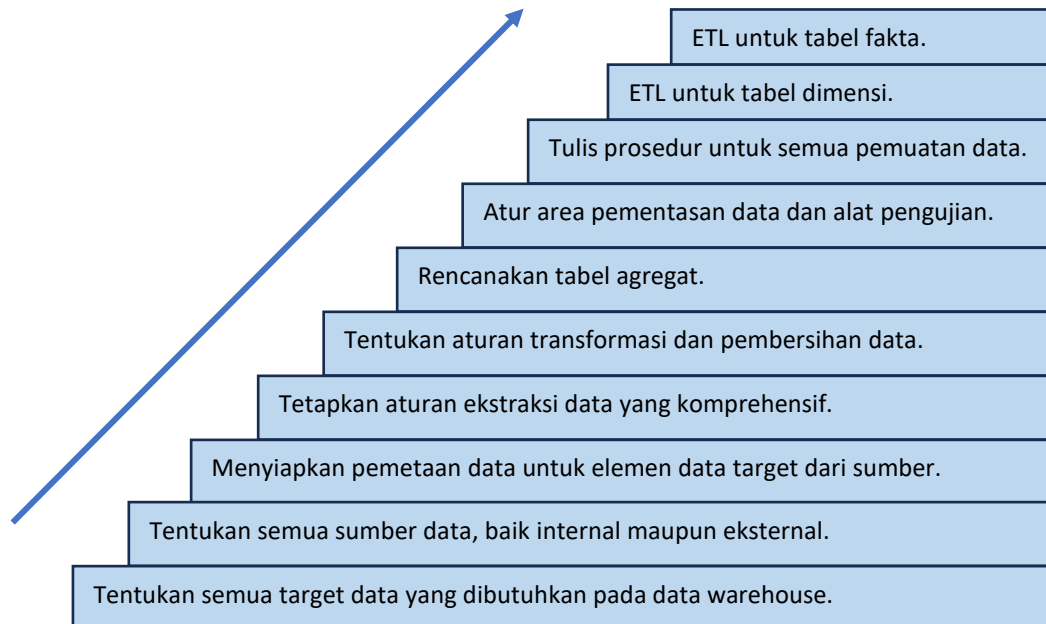
Aktivitas dalam fungsi transformasi data dapat menjalankan keseluruhan metode transformasi. Anda harus memformat ulang struktur data internal, mengurutkan ulang data, menerapkan berbagai bentuk teknik konversi, memberikan nilai default jika ada nilai yang hilang, dan Anda harus merancang seluruh rangkaian agregat yang diperlukan untuk peningkatan kinerja. Dalam banyak kasus, Anda perlu mengkonversi dari format EBCDIC ke ASCII.

Sekarang alihkan perhatian Anda ke fungsi pemuatan data. Pemuatan awal yang sangat besar dapat mengisi jutaan baris dalam database gudang data. Membuat dan mengelola load image untuk jumlah yang besar bukanlah tugas yang mudah. Yang lebih sulit lagi adalah tugas menguji dan menerapkan gambar beban untuk benar-benar mengisi file fisik di gudang data. Terkadang, diperlukan waktu dua minggu atau lebih untuk menyelesaikan pemuatan fisik awal. Sehubungan dengan mengekstraksi dan menerapkan perubahan bertahap yang sedang berlangsung, terdapat beberapa kesulitan. Menemukan metode ekstraksi yang tepat untuk kumpulan data sumber individual bisa jadi sulit. Setelah Anda menentukan metode ekstraksi, temukan jangka waktu untuk menerapkan perubahan gudang data bisa menjadi rumit jika gudang data Anda tidak dapat mengalami downtime yang lama.

12.2 PERSYARATAN DAN LANGKAH-LANGKAH ETL

Sebelum kita menyoroti beberapa masalah utama yang berkaitan dengan ETL, mari kita meninjau langkah-langkah fungsionalnya. Untuk penyegaran massal awal serta untuk pemuatan data tambahan, urutannya seperti yang disebutkan di sini: memicu perubahan tambahan, memfilter untuk penyegaran dan pemuatan tambahan, ekstraksi data, transformasi, integrasi, pembersihan, dan penerapan ke database gudang data.

Apa langkah utama dalam proses ETL? Lihatlah daftar yang ditunjukkan pada Gambar 12.1. Masing-masing langkah besar ini dipecah menjadi serangkaian kegiatan dan tugas. Gunakan gambar ini sebagai panduan untuk membuat daftar langkah-langkah untuk proses ETL gudang data Anda.



Gambar 12-1 Langkah-langkah utama dalam proses ETL.

Daftar berikut menyebutkan jenis aktivitas dan tugas yang membentuk proses ETL. Daftar ini tidak lengkap untuk setiap gudang data, namun memberikan wawasan yang baik tentang apa yang terlibat untuk menyelesaikan proses ETL.

- 1) Rencanakan tabel fakta agregat.
- 2) Tentukan aturan transformasi dan pembersihan data. Tetapkan aturan ekstraksi data yang komprehensif.
- 3) Menyiapkan pemetaan data untuk elemen data target dari sumber. Integrasikan semua sumber data, baik internal maupun eksternal.
- 4) Tentukan semua target data yang dibutuhkan dalam data warehouse.
- 5) Gabungkan beberapa struktur data sumber menjadi satu baris dalam database target gudang data.
- 6) Pisahkan satu struktur data sumber menjadi beberapa struktur untuk masuk ke beberapa baris database target.
- 7) Membaca data dari kamus data dan katalog sistem sumber.
- 8) Membaca data dari berbagai struktur file termasuk file datar, file terindeks (VSAM), dan database sistem lama (hierarki/jaringan).
- 9) Muat detail untuk mengisi tabel fakta atom.
- 10) Agregat untuk mengisi tabel fakta agregat atau ringkasan.
- 11) Ubah data dari satu format di platform sumber ke format lain di platform target.
- 12) Turunkan nilai target untuk kolom masukan (contoh: usia sejak tanggal lahir).

Ubah nilai samar menjadi nilai yang berarti bagi pengguna (contoh: 1 dan 2 untuk pria dan wanita).

Faktor faktor kunci

Sebelum kita melanjutkan, mari kita tunjukkan beberapa faktor kunci. Yang pertama berkaitan dengan kompleksitas fungsi ekstraksi dan transformasi data. Yang kedua adalah tentang fungsi pemuatan data.

Ingatlah bahwa alasan utama kompleksitas fungsi ekstraksi dan transformasi data adalah keragaman sistem sumber yang luar biasa. Di perusahaan besar, kita mungkin memiliki kombinasi yang membingungkan antara platform komputasi, sistem operasi, sistem manajemen basis data, protokol jaringan, dan sistem warisan sumber. Anda perlu memberikan perhatian khusus pada berbagai sumber dan memulai dengan inventarisasi lengkap sistem sumber. Dengan inventarisasi ini sebagai titik awal, kerjakan semua detail ekstraksi data. Kesulitan yang dihadapi dalam fungsi transformasi data juga berhubungan dengan heterogenitas sistem sumber.

Sekarang, dengan mengalihkan perhatian Anda ke fungsi pemuatan data, Anda memiliki sejumlah masalah yang harus diperhatikan. Biasanya, penyegaran massal, baik untuk pemuatan awal atau penyegaran berkala, menimbulkan kesulitan, bukan karena kerumitannya, namun karena pekerjaan pemuatan ini berjalan terlalu lama. Anda harus menemukan waktu yang tepat untuk menjadwalkan penyegaran penuh ini. Beban tambahan mempunyai beberapa jenis kesulitan lain. Pertama, Anda harus menentukan metode terbaik untuk menangkap perubahan yang sedang berlangsung dari setiap sistem sumber. Selanjutnya, Anda harus menjalankan penangkapan tanpa berdampak pada sistem sumber. Setelah itu, di sisi lain, Anda harus menjadwalkan pemuatan tambahan tanpa memengaruhi penggunaan gudang data oleh pengguna. Berikan perhatian khusus pada isu-isu utama ini saat merancang fungsi ETL untuk gudang data Anda. Sekarang mari kita ambil masing-masing dari ketiga fungsi ETL, satu per satu, dan pelajari detailnya.

12.3 EKSTRAKSI DATA

Sebagai seorang profesional TI, Anda pasti pernah berpartisipasi dalam ekstraksi dan konversi data saat menerapkan sistem operasional. Saat Anda beralih dari sistem entri pesanan berorientasi file VSAM ke sistem pemrosesan pesanan baru menggunakan teknologi database relasional, Anda mungkin telah menulis program ekstraksi data untuk mengambil data dari file VSAM guna menyiapkan data untuk mengisi database relasional.

Dua faktor utama membedakan ekstraksi data untuk sistem operasional baru dari ekstraksi data untuk gudang data. Pertama, untuk gudang data, Anda harus mengekstrak data dari banyak sumber berbeda. Selanjutnya, untuk gudang data, Anda harus mengekstrak data tentang perubahan untuk muatan tambahan yang sedang berlangsung serta untuk muatan penuh awal satu kali. Untuk sistem operasional, yang Anda butuhkan hanyalah ekstraksi satu kali dan konversi data.

Kedua faktor ini meningkatkan kompleksitas ekstraksi data untuk gudang data dan, oleh karena itu, memerlukan penggunaan alat ekstraksi data pihak ketiga selain program atau

skrip internal. Alat pihak ketiga umumnya lebih mahal dibandingkan program internal, namun alat tersebut mencatat metadatanya sendiri. Di sisi lain, program internal meningkatkan biaya pemeliharaan dan sulit dipertahankan seiring dengan perubahan sistem sumber. Jika perusahaan Anda berada dalam industri yang sering mengalami perubahan kondisi bisnis, maka Anda mungkin ingin meminimalkan penggunaan program internal. Alat pihak ketiga biasanya memberikan fleksibilitas bawaan. Yang harus Anda lakukan adalah mengubah parameter masukan untuk alat pihak ketiga yang Anda gunakan.

Ekstraksi data yang efektif adalah kunci keberhasilan gudang data Anda. Oleh karena itu, Anda perlu memberikan perhatian khusus terhadap permasalahan tersebut dan merumuskan strategi ekstraksi data untuk gudang data Anda. Berikut adalah daftar masalah ekstraksi data:

- ✓ Identifikasi sumber: mengidentifikasi aplikasi sumber dan struktur sumber.
- ✓ Metode ekstraksi: untuk setiap sumber data, tentukan apakah proses ekstraksi dilakukan secara manual atau berbasis alat.
- ✓ Frekuensi ekstraksi: untuk setiap sumber data, tetapkan seberapa sering ekstraksi data harus dilakukan; harian, mingguan, triwulanan, dan seterusnya.
- ✓ Jangka waktu: untuk setiap sumber data, menunjukkan jangka waktu untuk proses ekstraksi.
- ✓ Urutan pekerjaan: menentukan apakah permulaan suatu pekerjaan dalam aliran pekerjaan ekstraksi harus menunggu hingga pekerjaan sebelumnya berhasil diselesaikan.
- ✓ Penanganan pengecualian: menentukan cara menangani rekaman masukan yang tidak dapat diekstraksi.

Identifikasi Sumber

Mari kita bahas masalah pertama di atas, yaitu identifikasi sumber. Kita akan membahas permasalahan lainnya nanti saat kita melanjutkan sisa bab ini. Identifikasi sumber tentu saja mencakup identifikasi semua sumber data yang tepat. Hal ini tidak berhenti pada identifikasi sumber data saja. Hal ini lebih dari itu mencakup pemeriksaan dan verifikasi bahwa sumber yang diidentifikasi akan memberikan nilai yang diperlukan untuk gudang data. Mari kita telusuri proses identifikasi sumber secara mendetail.

Asumsikan bahwa bagian dari database Anda, mungkin salah satu data mart Anda, dirancang untuk memberikan informasi strategis mengenai pemenuhan pesanan. Untuk tujuan ini, Anda perlu menyimpan informasi historis tentang pesanan yang telah dipenuhi dan tertunda. Jika Anda mengirimkan pesanan melalui beberapa saluran pengiriman, Anda perlu mengambil data tentang saluran tersebut. Jika pengguna Anda tertarik untuk menganalisis pesanan berdasarkan status pesanan saat pesanan melalui proses pemenuhan, maka Anda perlu mengekstrak data tentang status pesanan.

Dalam tabel fakta untuk pemenuhan pesanan, Anda memerlukan atribut tentang jumlah total pesanan, diskon, komisi, perkiraan waktu pengiriman, waktu pengiriman aktual, dan tanggal pada berbagai tahap proses. Anda memerlukan tabel dimensi untuk produk, disposisi pesanan, saluran pengiriman, dan pelanggan. Pertama, Anda harus menentukan

apakah Anda memiliki sistem sumber untuk menyediakan data yang diperlukan untuk data mart ini. Kemudian, dari sistem sumber, Anda harus menetapkan sumber data yang benar untuk setiap elemen data di data mart. Selanjutnya, Anda harus melalui proses verifikasi untuk memastikan bahwa sumber yang diidentifikasi benar-benar benar. Gambar 12.2 menjelaskan pendekatan bertahap untuk identifikasi sumber untuk pemenuhan pesanan.

Identifikasi sumber bukanlah proses yang sederhana. Ini adalah proses pertama yang penting dalam fungsi ekstraksi data. Anda harus melalui proses identifikasi sumber untuk setiap informasi yang harus Anda simpan di gudang data. Seperti yang mungkin sudah Anda ketahui, identifikasi sumber memerlukan ketelitian, banyak waktu, dan analisis mendalam.

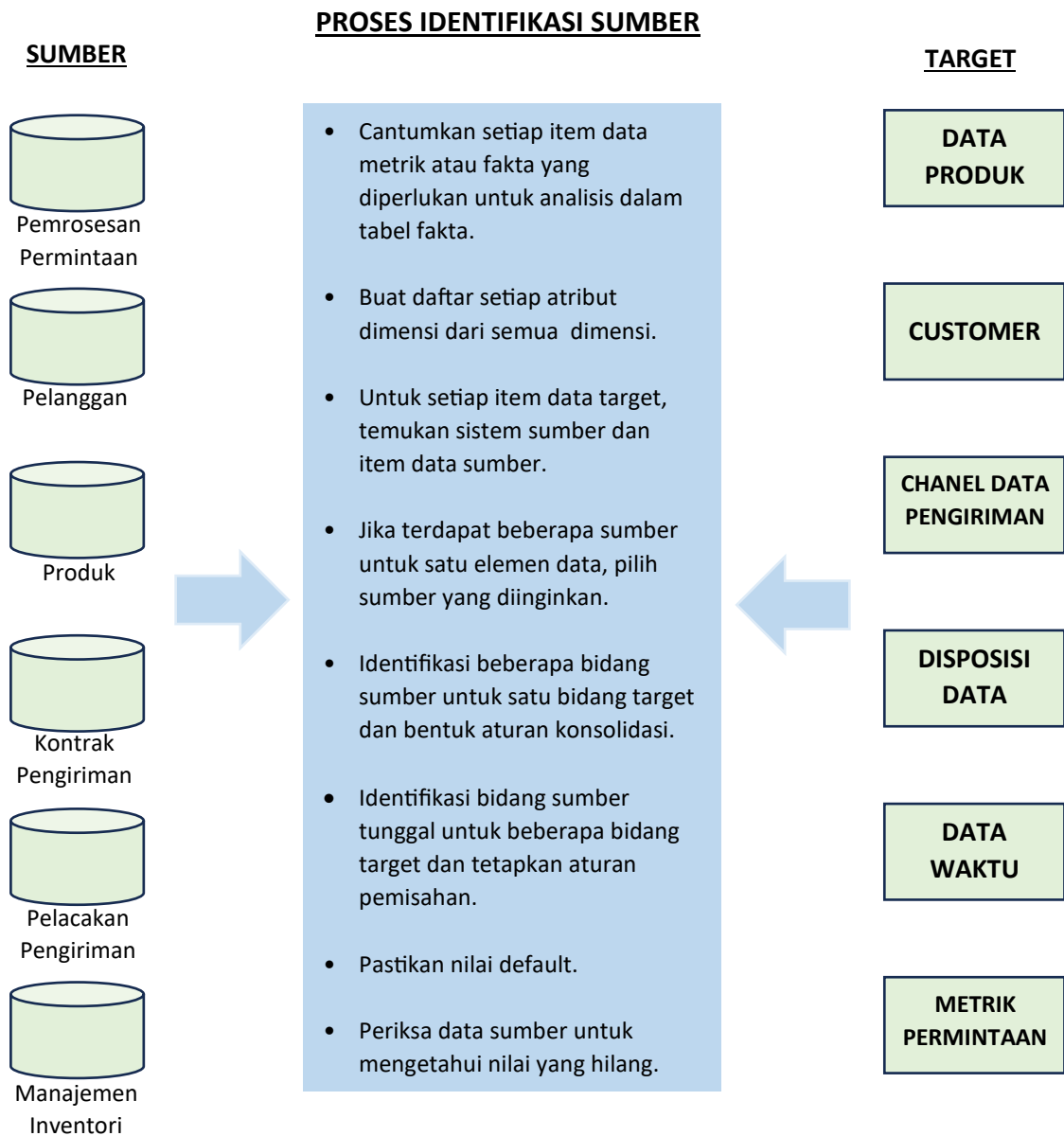
Teknik Ekstraksi Data

Sebelum mempelajari berbagai teknik ekstraksi data, Anda harus memahami dengan jelas sifat sumber data yang Anda ekstrak atau tangkap. Selain itu, Anda perlu mendapatkan wawasan tentang bagaimana data yang diekstraksi akan digunakan. Sumber data berada dalam kondisi yang terus berubah.

Transaksi bisnis terus mengubah data dalam sistem sumber. Dalam kebanyakan kasus, nilai suatu atribut dalam sistem sumber adalah nilai atribut tersebut pada saat itu. Jika Anda melihat setiap struktur data dalam sistem operasional sumber, transaksi bisnis sehari-hari secara konstan mengubah nilai atribut dalam struktur ini. Saat pelanggan berpindah ke negara bagian lain, data tentang pelanggan tersebut berubah di tabel pelanggan di sistem sumber. Ketika dua jenis paket tambahan ditambahkan ke cara suatu produk dijual, data produk berubah di sistem sumber. Ketika koreksi diterapkan pada kuantitas yang dipesan, data tentang pesanan tersebut akan berubah di sistem sumber.

Data dalam sistem sumber dikatakan bergantung pada waktu atau temporal. Ini karena data sumber berubah seiring waktu. Nilai suatu variabel berubah seiring waktu. Sekali lagi, ambil contoh perubahan alamat seorang pelanggan yang pindah dari New York ke California. Dalam sistem operasional, yang penting alamat pelanggan saat ini memiliki CA sebagai kode negara. Transaksi perubahan aktual itu sendiri, yang menyatakan bahwa kode negara bagian sebelumnya adalah NY dan kode negara bagian yang direvisi adalah CA, tidak perlu dipertahankan. Namun pikirkan bagaimana perubahan ini memengaruhi informasi di gudang data.

Jika kode negara bagian digunakan untuk menganalisis beberapa pengukuran seperti penjualan, penjualan kepada pelanggan sebelum perubahan harus dihitung di New York dan penjualan setelah perpindahan harus dihitung di California. Dengan kata lain, riwayat di gudang data tidak bisa diabaikan. Hal ini membawa kita pada pertanyaan: bagaimana Anda menangkap sejarah dari sistem sumber? Jawabannya bergantung pada bagaimana sebenarnya data disimpan dalam sistem sumber. Jadi mari kita periksa dan pahami bagaimana data disimpan dalam sistem operasional sumber.



Gambar 12.2 Identifikasi sumber: pendekatan bertahap.

Data dalam Sistem Operasional Sistem sumber ini umumnya menyimpan data dalam dua cara. Data operasional dalam sistem sumber dapat dianggap terbagi dalam dua kategori besar. Jenis teknik ekstraksi data yang harus Anda gunakan bergantung pada sifat dari kedua kategori tersebut.

Nilai Saat Ini Sebagian besar atribut dalam sistem sumber termasuk dalam kategori ini. Di sini nilai atribut yang disimpan mewakili nilai atribut pada saat itu. Nilai-nilai tersebut bersifat sementara atau sementara. Ketika transaksi bisnis terjadi, nilainya berubah. Tidak ada cara untuk memprediksi berapa lama nilai saat ini akan bertahan atau kapan akan berubah selanjutnya. Nama dan alamat pelanggan, saldo rekening bank, dan jumlah terutang pada pesanan individu adalah beberapa contoh dari kategori ini.

Apa implikasi kategori ini terhadap ekstraksi data? Nilai suatu atribut tetap konstan hanya sampai transaksi bisnis mengubahnya. Tidak ada yang tahu kapan itu akan diubah.

Ekstraksi data untuk melestarikan riwayat perubahan dalam gudang data cukup terlibat dalam kategori data ini. Status Berkala Kategori ini tidak seumum kategori sebelumnya. Dalam kategori ini, nilai atribut dipertahankan sebagai status setiap kali terjadi perubahan. Pada setiap titik waktu ini, nilai status disimpan dengan mengacu pada waktu ketika nilai baru mulai berlaku.

Kategori ini juga mencakup peristiwa yang disimpan dengan mengacu pada waktu terjadinya setiap peristiwa. Lihatlah cara data tentang polis asuransi biasanya dicatat dalam sistem operasional perusahaan asuransi. Basis data operasional menyimpan data status kebijakan pada setiap titik waktu ketika sesuatu dalam kebijakan berubah. Demikian pula, untuk klaim asuransi, setiap peristiwa, seperti permulaan klaim, verifikasi, penilaian, dan penyelesaian, dicatat dengan mengacu pada titik waktu.

Untuk data operasional dalam kategori ini, riwayat perubahan disimpan dalam sistem sumber itu sendiri. Oleh karena itu, ekstraksi data untuk tujuan menyimpan riwayat di data warehouse relatif lebih mudah. Baik itu data status atau data tentang suatu peristiwa, sistem sumber berisi data pada setiap titik waktu ketika terjadi perubahan. Pelajari Gambar 12-3 untuk memahami dua kategori data yang disimpan dalam sistem operasional. Berikan perhatian khusus pada contoh-contohnya.

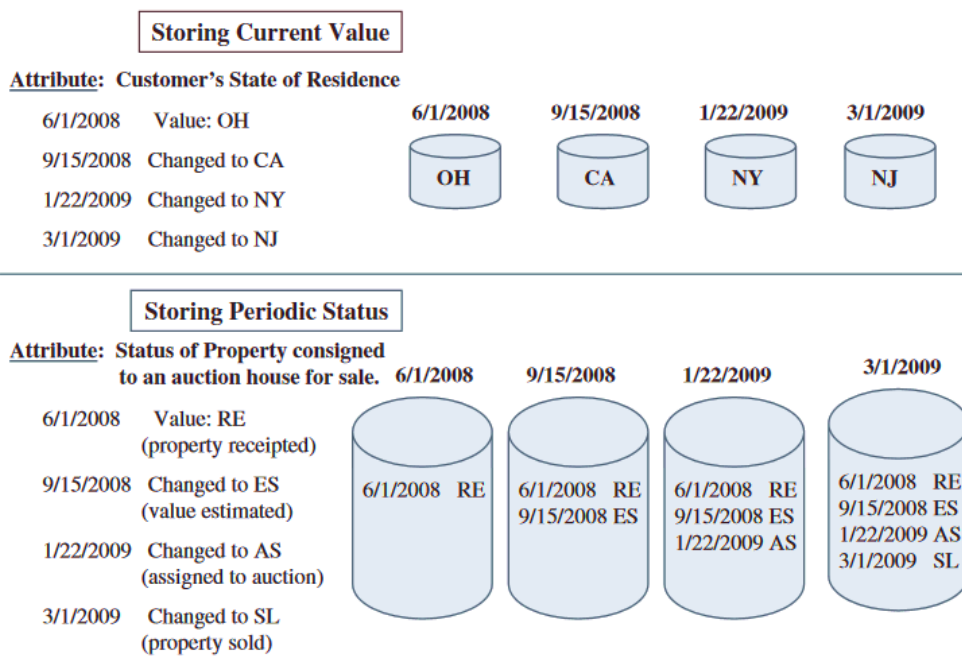
Setelah meninjau kategori yang menunjukkan bagaimana data disimpan dalam sistem operasional, kini kita dapat mendiskusikan teknik umum untuk ekstraksi data. Saat Anda menyebarkan gudang data, data awal pada waktu tertentu harus dipindahkan ke gudang data untuk memulainya. Ini adalah beban awal. Setelah pemuatan awal, gudang data Anda harus terus diperbarui sehingga riwayat perubahan dan status tercermin dalam gudang data. Secara umum, ada dua jenis ekstraksi data utama dari sistem operasional sumber: data “sebagaimana adanya” (statis) dan data revisi.

Data “sebagaimana adanya” atau statis adalah pengambilan data pada suatu titik waktu tertentu. Hal ini seperti mengambil snapshot dari sumber data yang relevan pada titik waktu tertentu. Untuk data saat ini atau data sementara, tangkapan ini akan mencakup semua data sementara yang diidentifikasi untuk diekstraksi. Selain itu, untuk data yang dikategorikan periodik, pengambilan data ini akan mencakup setiap status atau peristiwa pada setiap titik waktu sebagaimana tersedia dalam sistem operasional sumber.

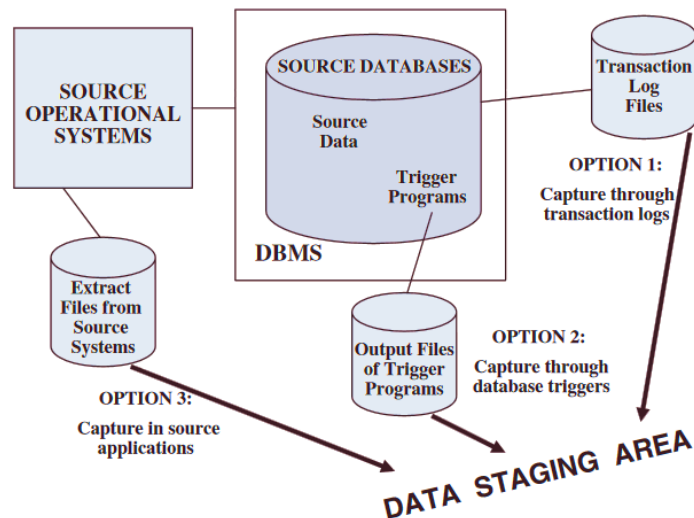
Anda akan menggunakan pengambilan data statis terutama untuk pemuatan awal gudang data. Terkadang, Anda mungkin ingin menyegarkan tabel dimensi secara menyeluruh. Misalnya, asumsikan bahwa master produk aplikasi sumber Anda telah diubah sepenuhnya. Dalam hal ini, Anda mungkin akan lebih mudah melakukan penyegaran penuh tabel dimensi produk gudang data target. Jadi, untuk tujuan ini, Anda akan melakukan pengambilan data statis terhadap data produk. Data revisi juga dikenal sebagai pengambilan data tambahan. Sebenarnya, ini bukan data tambahan melainkan revisi sejak terakhir kali data diambil. Jika sumber datanya bersifat sementara, perolehan revisinya tidaklah mudah. Untuk data status periodik atau data peristiwa periodik, pengambilan data tambahan mencakup nilai atribut pada waktu tertentu. Ekstrak status dan peristiwa yang telah dicatat sejak ekstraksi data terakhir.

Pengambilan data tambahan mungkin dilakukan segera atau ditunda. Dalam kelompok pengambilan data langsung ada tiga opsi berbeda. Tersedia dua opsi terpisah untuk pengambilan data yang ditangguhkan. Ekstraksi Data Segera Dalam opsi ini, ekstraksi data dilakukan secara real-time. Ini terjadi ketika transaksi terjadi pada database dan file sumber. Gambar 12.4 menunjukkan opsi ekstraksi data langsung. Sekarang mari kita masuk ke beberapa detail tentang tiga opsi untuk ekstraksi data langsung.

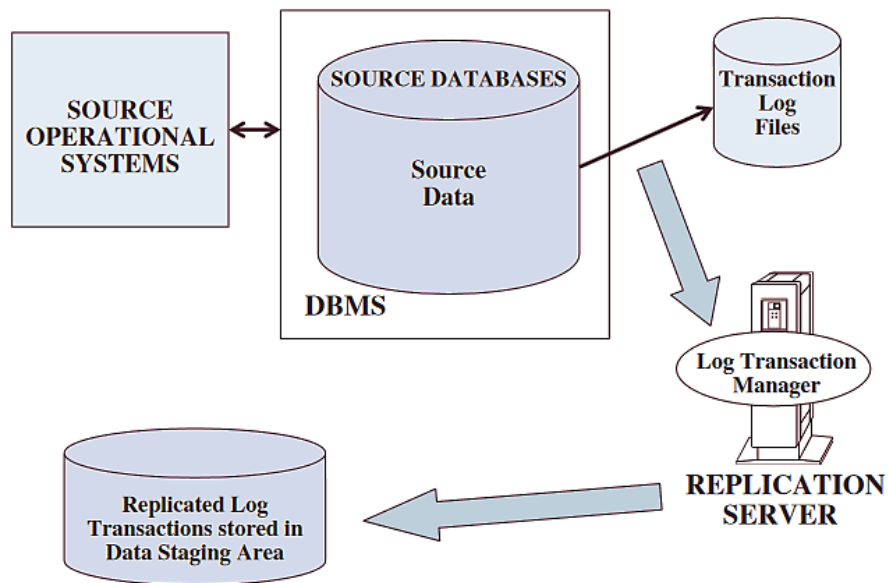
**VALUES OF ATTRIBUTES AS STORED IN
EXAMPLES OF ATTRIBUTES OPERATIONAL SYSTEMS AT DIFFERENT DATES**



Gambar 12.3 Data dalam sistem operasional.



Gambar 12.4 Opsi untuk ekstraksi data segera.



Gambar 12.5 Ekstraksi data menggunakan teknologi replikasi.

Menangkap melalui Log Transaksi Opsi ini menggunakan log transaksi DBMS yang dikelola untuk pemulihan dari kemungkinan kegagalan. Ketika setiap transaksi menambah, memperbarui, atau menghapus baris dari tabel database, DBMS segera menulis entri pada file log. Teknik ekstraksi data ini membaca log transaksi dan memilih semua transaksi yang dilakukan. Tidak ada biaya tambahan dalam sistem operasional karena logging sudah menjadi bagian dari pemrosesan transaksi.

Anda harus memastikan bahwa semua transaksi diekstraksi sebelum file log disegarkan. Saat file log pada penyimpanan disk terisi, kontennya dicadangkan di media lain dan file log disk digunakan kembali. Pastikan semua transaksi log diekstraksi untuk pembaruan gudang data. Jika semua sistem sumber Anda adalah aplikasi database, tidak ada masalah dengan teknik ini. Namun jika beberapa data sistem sumber Anda ada pada file terindeks dan file datar lainnya, opsi ini tidak akan berfungsi untuk kasus ini. Tidak ada file log untuk aplikasi non-database ini. Anda harus merancang beberapa teknik ekstraksi data lain untuk kasus ini.

Sementara kita membahas topik pengambilan data melalui log transaksi, mari kita mengambil penjelajahan sampingan dan melihat penggunaan replikasi. Replikasi data hanyalah sebuah metode untuk membuat salinan data dalam lingkungan terdistribusi. Gambar 12.5 mengilustrasikan bagaimana teknologi replikasi dapat digunakan untuk menangkap perubahan pada sumber data.

Log transaksi yang sesuai berisi semua perubahan pada berbagai tabel database sumber. Berikut adalah langkah-langkah umum dalam menggunakan replikasi untuk menangkap perubahan pada data sumber:

- Identifikasi tabel database sistem sumber
- Identifikasi dan tentukan file target di area pementasan
- Membuat pemetaan antara tabel sumber dan file target
- Tentukan mode replikasi
- Jadwalkan proses replikasi

- Catat perubahan dari log transaksi
- Mentransfer data yang diambil dari log ke file target
- Verifikasi transfer perubahan data
- Konfirmasikan keberhasilan atau kegagalan replikasi
- Dalam metadata, dokumentasikan hasil replikasi
- Mempertahankan definisi sumber, target, dan pemetaan

Menangkap melalui Pemicu Basis Data Sekali lagi, opsi ini berlaku untuk sistem sumber Anda yang merupakan aplikasi basis data. Seperti yang Anda ketahui, trigger adalah prosedur tersimpan khusus (program) yang disimpan di database dan dijalankan ketika kejadian tertentu yang telah ditentukan terjadi. Anda dapat membuat program pemicu untuk semua peristiwa yang datanya perlu diambil. Output dari program pemicu ditulis ke file terpisah yang akan digunakan untuk mengekstrak data untuk gudang data. Misalnya, jika Anda perlu mencatat semua perubahan pada catatan di tabel pelanggan, tulis program pemicu untuk menangkap semua pembaruan dan penghapusan di tabel tersebut.

Pengambilan data melalui pemicu basis data terjadi langsung dari sumbernya dan oleh karena itu cukup dapat diandalkan. Anda dapat mengambil gambar sebelum dan sesudah. Namun, membangun dan memelihara program pemicu memberikan beban tambahan pada upaya pembangunan. Selain itu, pelaksanaan prosedur pemicu selama pemrosesan transaksi sistem sumber memberikan overhead tambahan pada sistem sumber. Selanjutnya, opsi ini hanya berlaku untuk data sumber dalam database.

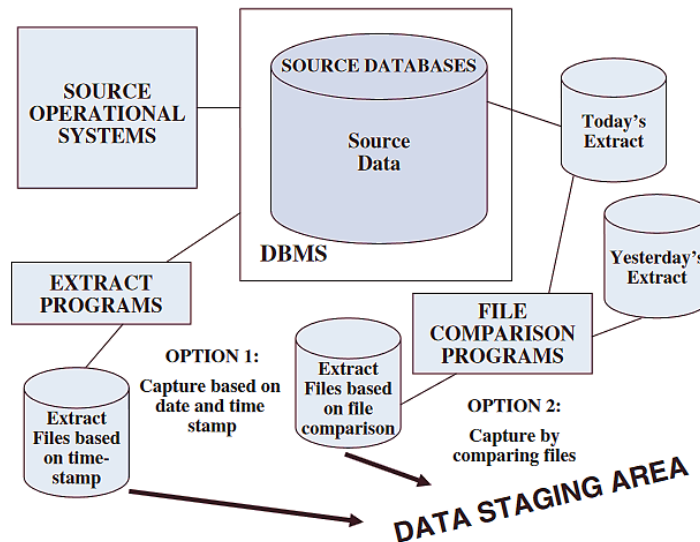
Pengambilan dalam Aplikasi Sumber Teknik ini juga disebut sebagai pengambilan data yang dibantu aplikasi. Dengan kata lain, aplikasi sumber dibuat untuk membantu dalam pengambilan data untuk data warehouse. Anda harus memodifikasi program aplikasi relevan yang menulis ke file sumber dan database. Anda merevisi program untuk menulis semua penambahan, pembaruan, dan penghapusan ke file sumber dan tabel database. Kemudian program ekstrak lainnya dapat menggunakan file terpisah yang berisi perubahan pada data sumber.

Berbeda dengan dua kasus sebelumnya, teknik ini dapat digunakan untuk semua jenis data sumber terlepas dari apakah itu dalam database, file yang diindeks, atau file datar lainnya. Namun Anda harus merevisi program dalam sistem operasional sumber dan menjaganya tetap terpelihara. Ini bisa menjadi tugas berat jika jumlah program sistem sumber banyak. Selain itu, teknik ini dapat menurunkan kinerja aplikasi sumber karena diperlukan pemrosesan tambahan untuk menangkap perubahan pada file terpisah.

Ekstraksi Data yang Ditangguhkan Dalam kasus yang dibahas di atas, pengambilan data terjadi saat transaksi terjadi dalam sistem operasional sumber. Pengambilan data bersifat langsung atau real-time. Sebaliknya, teknik ekstraksi data yang ditangguhkan tidak menangkap perubahan secara real-time. Penangkapan terjadi kemudian. Gambar 12.6 menunjukkan opsi ekstraksi data yang ditangguhkan. Sekarang mari kita bahas dua opsi untuk ekstraksi data yang ditangguhkan.

Pengambilan Berdasarkan Stempel Tanggal dan Waktu Setiap kali rekaman sumber dibuat atau diperbarui, rekaman tersebut mungkin ditandai dengan stempel yang

menunjukkan tanggal dan waktu. Stempel waktu memberikan dasar untuk memilih rekaman untuk ekstraksi data. Di sini pengambilan data terjadi di lain waktu, bukan saat setiap rekaman sumber dibuat atau diperbarui. Jika Anda menjalankan program ekstraksi data pada tengah malam setiap hari, setiap hari Anda hanya akan mengekstrak data yang memiliki cap tanggal dan waktu lebih lambat dari tengah malam pada hari sebelumnya. Teknik ini bekerja dengan baik jika jumlah catatan yang direvisi sedikit.



Gambar 12.6 Opsi untuk ekstraksi data yang ditangguhkan.

Tentu saja, teknik ini mengasumsikan bahwa semua catatan sumber yang relevan berisi cap tanggal dan waktu. Asalkan hal ini benar, pengambilan data berdasarkan cap tanggal dan waktu dapat berfungsi untuk semua jenis file sumber. Teknik ini menangkap keadaan terkini dari sumber data. Status perantara antara dua proses ekstraksi data akan hilang.

Penghapusan catatan sumber menimbulkan masalah khusus. Jika rekaman sumber terhapus di antara dua proses ekstrak, informasi tentang penghapusan tersebut tidak terdeteksi. Anda dapat menyiasatinya dengan menandai rekaman sumber untuk dihapus terlebih dahulu, melakukan proses ekstraksi, lalu melanjutkan dan menghapus rekaman secara fisik. Ini berarti Anda harus menambahkan lebih banyak logika ke aplikasi sumber. Ambil dengan Membandingkan File Jika tidak ada teknik di atas yang dapat dilakukan untuk file sumber tertentu di lingkungan Anda, maka pertimbangkan teknik ini sebagai pilihan terakhir. Teknik ini juga disebut teknik snapshot differential karena membandingkan dua snapshot dari data sumber. Mari kita lihat cara kerja teknik ini.

Misalkan Anda ingin menerapkan teknik ini untuk menangkap perubahan pada data produk Anda. Saat melakukan ekstraksi data hari ini untuk perubahan pada data produk, Anda melakukan perbandingan file lengkap antara salinan data produk hari ini dan salinan data produk kemarin. Anda juga membandingkan kunci rekaman untuk menemukan sisipan dan penghapusan. Kemudian Anda menangkap perubahan apa pun di antara kedua salinan tersebut.

Teknik ini mengharuskan penyimpanan salinan semua sumber data yang relevan sebelumnya. Meskipun sederhana dan lugas, perbandingan baris penuh dalam file besar bisa sangat tidak efisien. Namun, ini mungkin satu-satunya opsi yang layak untuk beberapa sumber data lama yang tidak memiliki log transaksi atau stempel waktu pada rekaman sumber.

Evaluasi Teknik

Ringkasnya, opsi berikut tersedia untuk ekstraksi data:

- Pengambilan data statis
- Tangkap melalui log transaksi
- Tangkap melalui pemicu basis data
- Tangkap dalam aplikasi sumber
- Tangkap berdasarkan cap tanggal dan waktu
- Tangkap dengan membandingkan file

Anda dihadapkan pada beberapa pertanyaan besar. Manakah yang dapat diterapkan di lingkungan Anda? Teknik apa yang harus Anda gunakan? Anda akan menggunakan teknik pengambilan data statis setidaknya dalam satu situasi ketika Anda mengisi gudang data pada awalnya pada saat penerapan. Setelah itu, Anda biasanya akan menyadari bahwa Anda memerlukan kombinasi beberapa teknik ini untuk lingkungan Anda. Jika Anda memiliki sistem lama, Anda mungkin memerlukan metode perbandingan file.

Gambar 12.7 menyoroti kelebihan dan kekurangan berbagai teknik. Pelajarilah dengan cermat dan gunakan untuk menentukan teknik yang perlu Anda gunakan di lingkungan Anda. Mari kita membuat beberapa komentar umum. Teknik manakah yang mudah dan murah untuk diterapkan? Pertimbangkan teknik penggunaan log transaksi dan pemicu database. Kedua teknik ini sudah tersedia melalui produk database. Keduanya relatif murah dan mudah diterapkan. Teknik berdasarkan log transaksi mungkin yang paling murah. Tidak ada overhead tambahan pada sistem operasional sumber. Dalam kasus pemicu basis data, terdapat kebutuhan untuk membuat dan memelihara program pemicu. Bahkan di sini, upaya pemeliharaan dan biaya tambahan pada sistem operasional sumber tidak terlalu banyak dibandingkan dengan teknik lainnya.

Pengambilan data dalam sistem sumber bisa menjadi hal yang paling mahal dalam hal pengembangan dan pemeliharaan. Teknik ini memerlukan revisi besar terhadap sistem sumber yang ada. Untuk banyak aplikasi sumber lama, menemukan kode sumber dan memodifikasinya mungkin tidak dapat dilakukan sama sekali. Namun, jika data sumber tidak berada di file database dan stempel tanggal dan waktu tidak ada di rekaman sumber, ini adalah salah satu dari sedikit opsi yang tersedia. Apa dampaknya terhadap kinerja sistem operasional sumber? Tentu saja, metode ekstraksi data yang ditanggihkan memiliki dampak paling kecil terhadap sistem operasional.

Ekstraksi data berdasarkan stempel waktu dan ekstraksi data berdasarkan perbandingan file dilakukan di luar operasi normal sistem sumber. Oleh karena itu, kedua hal ini merupakan pilihan yang lebih disukai ketika meminimalkan dampak terhadap sistem operasional merupakan prioritas. Namun, opsi penangkapan yang ditanggihkan ini memiliki beberapa kekurangan. Mereka melacak perubahan status data sumber pada saat ekstraksi

saat ini dibandingkan dengan statusnya pada saat ekstraksi sebelumnya. Perubahan sementara apa pun tidak dicatat. Oleh karena itu, di mana pun Anda berurusan dengan data sumber sementara, Anda hanya dapat memberikan perkiraan sejarahnya.



Gambar 12.7 Teknik pengambilan data: kelebihan dan kekurangan.

Jadi, apa intinya? Gunakan teknik pengambilan data dalam sistem sumber dengan hemat karena melibatkan terlalu banyak pekerjaan pengembangan dan pemeliharaan. Untuk data sumber Anda di database, pengambilan melalui log transaksi dan pengambilan melalui pemicu database adalah pilihan pertama yang jelas. Di antara keduanya, penangkapan melalui log transaksi adalah pilihan yang lebih baik karena kinerjanya lebih baik. Selain itu, teknik ini juga berlaku untuk database nonrelasional. Metode perbandingan file adalah yang paling memakan waktu untuk ekstraksi data. Gunakan hanya jika yang lainnya tidak dapat diterapkan.

12.4 TRANSFORMASI DATA

Dengan memanfaatkan beberapa teknik yang dibahas di bagian sebelumnya, Anda merancang fungsi ekstraksi data. Sekarang data yang diekstraksi adalah data mentah dan tidak bisa langsung diterapkan ke data warehouse. Pertama, semua data yang diekstraksi harus dapat digunakan di gudang data. Memiliki informasi yang dapat digunakan untuk pengambilan keputusan strategis adalah prinsip yang mendasari gudang data. Anda tahu bahwa data dalam sistem operasional tidak dapat digunakan untuk tujuan ini. Selanjutnya, karena data operasional diambil dari banyak sistem lama, kualitas data dalam sistem tersebut cenderung kurang baik untuk gudang data. Anda harus memperkaya dan meningkatkan kualitas data sebelum dapat digunakan di gudang data.

Sebelum memindahkan data yang diekstraksi dari sistem sumber ke dalam data warehouse, mau tidak mau Anda harus melakukan berbagai macam transformasi data. Anda harus mengubah data sesuai standar karena data tersebut berasal dari banyak sistem sumber yang berbeda. Anda harus memastikan bahwa setelah semua data digabungkan, data gabungan tersebut tidak melanggar aturan bisnis apa pun.

Pertimbangkan struktur data dan elemen data yang Anda perlukan di gudang data Anda. Sekarang pikirkan tentang semua data relevan yang akan diambil dari sistem sumber. Dari beragam format data sumber, nilai data, dan kondisi kualitas data, Anda tahu bahwa Anda harus melakukan beberapa jenis transformasi agar data sumber sesuai untuk gudang data Anda. Transformasi data sumber mencakup berbagai macam manipulasi untuk mengubah semua data sumber yang diekstraksi menjadi informasi yang dapat digunakan untuk disimpan di gudang data

Banyak organisasi meremehkan tingkat dan kompleksitas fungsi transformasi data. Mereka memulai dengan data mart departemen sederhana sebagai proyek percontohan. Hampir semua data untuk uji coba ini berasal dari satu sumber aplikasi. Transformasi data hanya memerlukan konversi lapangan dan beberapa pemformatan ulang struktur data. Jangan membuat kesalahan dengan menganggap remeh fungsi transformasi data. Bersiaplah untuk mempertimbangkan semua permasalahan yang berbeda dan alokasikan waktu dan upaya yang cukup untuk tugas merancang transformasi.

Praktisi data warehouse telah mencoba mengklasifikasikan transformasi data dalam beberapa cara, dimulai dengan klasifikasi yang sangat umum dan luas yaitu transformasi sederhana dan transformasi kompleks. Ada juga beberapa kebingungan tentang semantik. Seorang praktisi mungkin menyebut integrasi data sebagai proses dalam fungsi transformasi data yang merupakan semacam pra-pemrosesan data sumber. Bagi praktisi lain, integrasi data dapat berarti pemetaan bidang sumber ke bidang target di gudang data. Menolak godaan untuk menggeneralisasi dan mengklasifikasikan, kami akan menyoroti dan mendiskusikan jenis umum dari fungsi transformasi utama. Anda dapat meninjau setiap tipe dan memutuskan sendiri apakah tipe tersebut akan sederhana atau kompleks di lingkungan gudang data Anda sendiri.

Salah satu upaya besar dalam transformasi data adalah peningkatan kualitas data. Secara sederhana, ini termasuk mengisi nilai atribut yang hilang dalam data yang diekstraksi.

Kualitas data sangat penting dalam gudang data karena dampak keputusan strategis berdasarkan informasi yang salah dapat sangat merugikan. Oleh karena itu, kami akan membahas masalah kualitas data secara mendalam pada bab berikutnya.

Transformasi Data: Tugas Dasar

Terlepas dari keragaman dan kompleksitas sistem operasional sumber, dan terlepas dari luas gudang data Anda, Anda akan menemukan bahwa sebagian besar fungsi transformasi data Anda dipecah menjadi beberapa tugas dasar. Mari kita bahas tugas-tugas dasar ini sehingga Anda dapat melihat transformasi data dari perspektif mendasar. Berikut adalah serangkaian tugas dasar:

- 1) **Pilihan:** Ini terjadi pada awal seluruh proses transformasi data. Anda memilih seluruh catatan atau bagian dari beberapa catatan dari sistem sumber. Tugas seleksi biasanya merupakan bagian dari fungsi ekstraksi itu sendiri. Namun, dalam beberapa kasus, komposisi struktur sumber mungkin tidak dapat disesuaikan dengan pemilihan bagian yang diperlukan selama ekstraksi data. Dalam kasus ini, adalah bijaksana untuk mengekstrak seluruh rekaman dan kemudian melakukan seleksi sebagai bagian dari fungsi transformasi.
- 2) **Pemisahan/Penggabungan:** Tugas ini mencakup tipe manipulasi data yang perlu Anda lakukan pada bagian rekaman sumber yang dipilih. Terkadang (jarang), Anda akan memisahkan bagian-bagian yang dipilih lebih jauh lagi selama transformasi data. Penggabungan bagian-bagian yang dipilih dari banyak sistem sumber lebih luas di lingkungan gudang data.
- 3) **Konversi:** Ini adalah tugas yang mencakup semua. Ini mencakup berbagai macam konversi dasar dari satu bidang karena dua alasan utama satu untuk menstandarisasi ekstraksi data dari sistem sumber yang berbeda, dan yang lainnya untuk membuat bidang tersebut dapat digunakan dan dipahami oleh pengguna.
- 4) **Peringkasan:** Terkadang Anda mungkin merasa tidak mungkin menyimpan data pada tingkat detail terendah di gudang data Anda. Mungkin tidak ada pengguna Anda yang memerlukan data dengan rincian terendah untuk analisis atau pembuatan kueri. Misalnya, untuk jaringan toko kelontong, data penjualan pada tingkat detail terendah untuk setiap transaksi di kasir mungkin tidak diperlukan. Menyimpan penjualan per produk per toko per hari di gudang data mungkin cukup memadai. Jadi, dalam hal ini, fungsi transformasi data mencakup ringkasan penjualan harian berdasarkan produk dan toko.
- 5) **Penyuburan:** Tugas ini adalah penataan ulang dan penyederhanaan masing-masing bidang agar lebih berguna bagi lingkungan gudang data. Anda dapat menggunakan satu atau lebih bidang dari catatan masukan yang sama untuk membuat tampilan data yang lebih baik untuk gudang data. Prinsip ini diperluas ketika satu atau lebih bidang berasal dari beberapa catatan, sehingga menghasilkan satu bidang untuk gudang data.

Jenis Transformasi Utama

Anda telah melihat serangkaian tugas transformasi dasar. Saat Anda mempertimbangkan kumpulan struktur data tertentu yang diekstraksi, Anda akan

menemukan bahwa fungsi transformasi yang perlu Anda lakukan pada kumpulan ini dapat dilakukan dengan kombinasi tugas dasar yang dibahas.

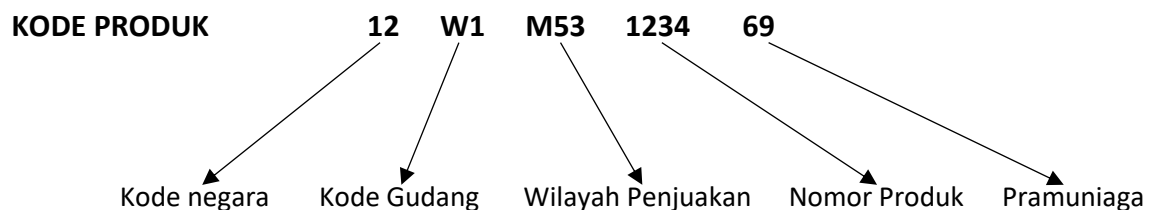
Sekarang mari kita pertimbangkan jenis fungsi transformasi tertentu. Ini adalah jenis transformasi yang paling umum:

- a) **Revisi Format:** Anda akan sering menjumpai hal ini. Revisi ini mencakup perubahan pada tipe data dan panjang masing-masing bidang. Dalam sistem sumber Anda, tipe paket produk dapat ditunjukkan dengan kode dan nama yang bidangnya berupa tipe data numerik dan teks. Sekali lagi, panjang jenis paket mungkin berbeda-beda di setiap sistem sumber. Sebaiknya standarisasi dan ubah tipe data menjadi teks untuk memberikan nilai yang berarti bagi pengguna.
- b) **Decoding Bidang:** Ini juga merupakan jenis transformasi data yang umum. Saat Anda berurusan dengan beberapa sistem sumber, Anda pasti memiliki item data yang sama yang dijelaskan oleh sejumlah besar nilai bidang. Contoh klasiknya adalah pengkodean gender, dengan satu sistem sumber menggunakan 1 dan 2 untuk pria dan wanita dan sistem lainnya menggunakan M dan F. Selain itu, banyak sistem lama yang terkenal menggunakan kode rahasia untuk mewakili nilai bisnis. Apa arti kode AC, IN, RE, dan SU di file pelanggan? Anda perlu memecahkan kode semua kode rahasia tersebut dan mengubahnya menjadi nilai yang masuk akal bagi pengguna. Ubah kodenya menjadi Aktif, Tidak Aktif, Reguler, dan Ditangguhkan.
- c) **Nilai yang Dihitung dan Diperoleh:** Bagaimana jika Anda ingin menyimpan margin keuntungan beserta jumlah penjualan dan biaya di tabel gudang data Anda? Data yang diambil dari sistem penjualan berisi jumlah penjualan, unit penjualan, dan perkiraan biaya operasional berdasarkan produk. Anda harus menghitung total biaya dan margin keuntungan sebelum data dapat disimpan di gudang data. Saldo harian rata-rata dan rasio operasi adalah contoh bidang turunan.
- d) **Pemisahan Bidang Tunggal:** Sistem warisan sebelumnya menyimpan nama dan alamat pelanggan dan karyawan dalam kolom teks besar. Nama depan, inisial tengah, dan nama belakang disimpan sebagai teks besar dalam satu kolom. Demikian pula, beberapa sistem sebelumnya menyimpan data kota, negara bagian, dan kode pos secara bersamaan dalam satu bidang. Anda perlu menyimpan masing-masing komponen nama dan alamat di bidang terpisah di gudang data Anda karena dua alasan. Pertama, Anda dapat meningkatkan kinerja pengoperasian dengan mengindeks masing-masing komponen. Kedua, pengguna Anda mungkin perlu melakukan analisis menggunakan masing-masing komponen seperti kota, negara bagian, dan kode pos.
- e) **Penggabungan Informasi:** Hal ini tidak sepenuhnya merupakan kebalikan dari pemisahan bidang tunggal. Jenis transformasi data ini tidak berarti menggabungkan beberapa bidang untuk membuat satu bidang data. Misalnya, informasi tentang suatu produk mungkin berasal dari sumber data yang berbeda. Kode dan deskripsi produk mungkin berasal dari satu sumber data. Jenis paket yang relevan mungkin ditemukan di sumber data lain. Data biaya mungkin berasal dari sumber lain. Dalam hal ini,

penggabungan informasi menunjukkan kombinasi kode produk, deskripsi, jenis paket, dan biaya menjadi satu kesatuan.

- f) **Konversi kumpulan karakter:** Jenis transformasi data ini berkaitan dengan konversi kumpulan karakter menjadi kumpulan karakter standar yang disepakati untuk data tekstual di gudang data. Jika Anda memiliki sistem warisan mainframe sebagai sistem sumber, sumber data dari sistem ini akan menggunakan karakter EBCDIC. Jika arsitektur berbasis PC adalah pilihan untuk gudang data Anda, maka Anda harus mengubah format mainframe EBCDIC ke format ASCII. Ketika data sumber Anda berada pada jenis perangkat keras dan sistem operasi lain, Anda dihadapkan pada konversi rangkaian karakter yang serupa.
- g) **Konversi Satuan Pengukuran:** Banyak perusahaan saat ini memiliki cabang global. Pengukuran di banyak negara Eropa menggunakan satuan metrik. Jika perusahaan Anda beroperasi di luar negeri, Anda mungkin harus mengonversi metriknya sehingga semuanya berada dalam satu unit pengukuran standar.
- h) **Konversi Tanggal/Waktu:** Tipe ini berkaitan dengan representasi tanggal dan waktu dalam format standar. Misalnya, format tanggal Amerika dan Inggris mungkin distandarisasi ke format internasional. Tanggal 11 Oktober 2008 ditulis 11/10/2008 dalam format AS dan 10/11/2008 dalam format Inggris. Tanggal ini mungkin dibakukan untuk ditulis sebagai 11 Oktober 2008.
- i) **Peringkasan:** Jenis transformasi ini adalah pembuatan ringkasan untuk dimuat di gudang data alih-alih memuat data pada tingkat paling granular. Misalnya, bagi perusahaan kartu kredit untuk menganalisis pola penjualan, mungkin tidak perlu menyimpan di gudang data setiap transaksi pada setiap kartu kredit. Sebaliknya, Anda mungkin ingin meringkas transaksi harian untuk setiap kartu kredit dan menyimpan data ringkasan daripada menyimpan data paling terperinci berdasarkan transaksi individual.

KUNCI SISTEM PRODUKSI



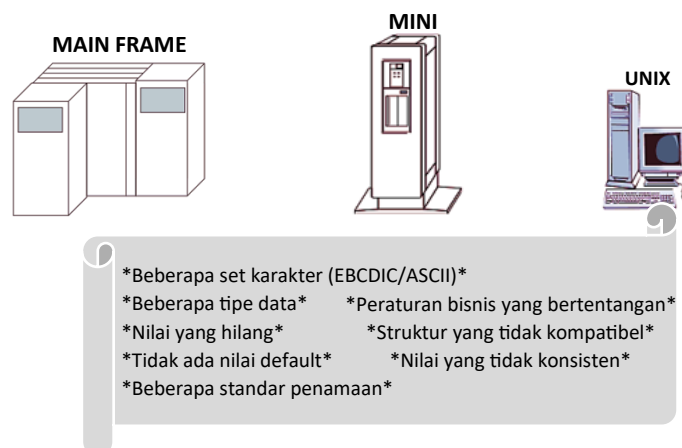
**GUDANG DATA – KUNCI PRODUK
12345678**

Gambar 12.8 Transformasi data: restrukturisasi kunci.

- j) **Restrukturisasi Kunci:** Saat mengekstrak data dari sumber input Anda, lihat kunci utama dari rekaman yang diekstraksi. Anda harus membuat kunci untuk tabel fakta dan dimensi berdasarkan kunci dalam rekaman yang diekstraksi. Lihat Gambar 12-8. Dalam contoh yang ditunjukkan pada gambar, kode produk di organisasi ini disusun agar memiliki makna yang melekat. Jika Anda menggunakan kode produk ini sebagai kunci utama, akan ada masalah. Jika produk dipindahkan ke gudang lain, bagian gudang dari kunci produk harus diubah. Ini adalah masalah umum pada sistem lama. Saat memilih kunci untuk tabel database gudang data Anda, hindari kunci tersebut dengan makna bawaan. Ubah kunci tersebut menjadi kunci generik yang dihasilkan oleh sistem itu sendiri. Ini disebut restrukturisasi kunci.
- k) **Deduplikasi:** Di banyak perusahaan, file pelanggan memiliki beberapa catatan untuk pelanggan yang sama. Kebanyakan, duplikat tersebut merupakan hasil dari pembuatan catatan tambahan secara tidak sengaja. Di gudang data, Anda ingin menyimpan satu catatan untuk satu pelanggan dan menautkan semua duplikat di sistem sumber ke catatan tunggal ini. Proses ini disebut deduplikasi file pelanggan. File karyawan dan, terkadang, file master produk memiliki masalah duplikasi seperti ini.

Integrasi dan Konsolidasi Data

Tantangan sebenarnya dari fungsi ETL adalah menyatukan semua sumber data dari banyak sistem sumber yang berbeda dan berbeda. Bahkan saat ini, banyak gudang data mendapatkan data yang diekstraksi dari kombinasi sistem mainframe lama dan aplikasi komputer mini lama, selain sistem klien/server yang lebih baru. Sebagian besar sistem sumber ini tidak mematuhi seperangkat aturan bisnis yang sama. Seringkali mereka mengikuti konvensi penamaan yang berbeda dan standar representasi data yang bervariasi. Gambar 12-9 menunjukkan lingkungan sumber data yang khas. Perhatikan permasalahan menantang yang ditunjukkan pada gambar.



Gambar 12.9 Lingkungan sumber data pada umumnya.

Mengintegrasikan data melibatkan penggabungan semua data operasional yang relevan ke dalam struktur data yang koheren agar siap untuk dimuat ke dalam gudang data. Anda mungkin ingin menganggap integrasi dan konsolidasi data sebagai jenis pra-proses

sebelum rutinitas transformasi besar lainnya diterapkan. Anda harus membakukan nama dan representasi data serta menyelesaikan perbedaan dalam cara data yang sama direpresentasikan dalam sistem sumber yang berbeda. Meskipun memakan waktu, banyak tugas integrasi data yang dapat dikelola. Namun, mari kita bahas beberapa tantangan yang lebih sulit.

Masalah Identifikasi Entitas Jika Anda memiliki tiga aplikasi warisan berbeda yang dikembangkan di organisasi Anda pada waktu yang berbeda di masa lalu, kemungkinan besar Anda memiliki tiga file pelanggan berbeda yang mendukung sistem tersebut. Satu sistem mungkin adalah sistem entri pesanan lama, sistem lainnya adalah sistem dukungan layanan pelanggan, dan yang ketiga adalah sistem pemasaran. Sejumlah besar pelanggan akan umum untuk ketiga file tersebut. Pelanggan yang sama pada setiap file mungkin memiliki nomor identifikasi unik. Nomor identifikasi unik untuk pelanggan yang sama mungkin tidak sama di ketiga sistem.

Ini adalah masalah identifikasi di mana Anda tidak mengetahui catatan pelanggan mana yang berhubungan dengan pelanggan yang sama. Namun di gudang data Anda perlu menyimpan satu catatan untuk setiap pelanggan. Anda harus bisa mendapatkan aktivitas satu pelanggan dari berbagai sistem sumber dan kemudian mencocokkannya dengan catatan tunggal untuk dimuat ke gudang data. Ini adalah masalah yang umum namun sangat sulit di banyak perusahaan di mana aplikasi telah berkembang seiring berjalannya waktu. Jenis masalah ini lazim terjadi ketika ada banyak sumber untuk entitas yang sama. Vendor, pemasok, karyawan, dan terkadang produk adalah entitas yang rentan terhadap masalah seperti ini.

Dalam contoh tiga file pelanggan di atas, Anda harus merancang algoritme kompleks untuk mencocokkan catatan dari ketiga file dan membentuk kelompok catatan yang cocok. Tidak ada algoritma pencocokan yang dapat menentukan kelompok secara lengkap. Jika kriteria pencocokan terlalu ketat, beberapa rekaman akan lolos dari grup. Di sisi lain, jika kriteria pencocokan terlalu longgar, kelompok tertentu dapat mencakup catatan lebih dari satu pelanggan. Anda perlu melibatkan pengguna dalam meninjau pengecualian pada prosedur otomatis. Anda harus mempertimbangkan masalah yang berkaitan dengan sistem sumber Anda dan memutuskan bagaimana menangani masalah identifikasi entitas.

Setiap kali fungsi ekstrak data dilakukan untuk gudang data Anda, yang mungkin dilakukan setiap hari, apakah Anda berhenti sejenak untuk menyelesaikan masalah identifikasi entitas sebelum memuat gudang data? Bagaimana pengaruhnya terhadap ketersediaan gudang data bagi pengguna Anda? Beberapa perusahaan, tergantung pada situasi masing-masing, mengambil opsi untuk memecahkan masalah identifikasi entitas dalam dua tahap. Pada tahap pertama, semua catatan, terlepas dari apakah itu duplikat atau tidak, diberi pengidentifikasi unik. Tahap kedua terdiri dari rekonsiliasi duplikat secara berkala melalui algoritma otomatis dan verifikasi manual.

Masalah Berbagai Sumber Ini adalah jenis masalah lain yang mempengaruhi integrasi data, meskipun kurang umum dan tidak sekompleks masalah identifikasi entitas. Masalah ini diakibatkan oleh satu elemen data yang memiliki lebih dari satu sumber. Misalnya, biaya

satuan produk tersedia dari dua sistem. Dalam aplikasi penetapan biaya standar, nilai biaya dihitung dan diperbarui pada interval tertentu. Sistem pemrosesan pesanan Anda juga memuat biaya per unit untuk semua produk. Mungkin terdapat sedikit perbedaan dalam angka biaya dari kedua sistem ini. Dari sistem manakah Anda harus mendapatkan biaya penyimpanan di gudang data?

Solusi langsungnya adalah dengan menetapkan prioritas yang lebih tinggi pada salah satu dari dua sumber tersebut dan mengambil biaya unit produk dari sumber tersebut. Terkadang, solusi sederhana seperti ini mungkin tidak sesuai dengan kebutuhan pengguna gudang data. Anda mungkin harus memilih salah satu file berdasarkan tanggal pembaruan terakhir. Atau, dalam beberapa kasus lain, penentuan Anda terhadap sumber yang sesuai bergantung pada bidang terkait lainnya.

Transformasi Atribut Dimensi

Pada Bab 11, kita membahas perubahan atribut tabel dimensi. Kami meninjau jenis perubahan pada atribut ini. Selain itu, kami menyarankan cara untuk menangani ketiga jenis dimensi yang berubah secara perlahan. Perubahan tipe 1 adalah koreksi kesalahan. Perubahan ini diterapkan pada gudang data tanpa perlu menyimpan riwayat. Perubahan tipe 2 mempertahankan riwayat di gudang data. Perubahan tipe 3 adalah perubahan tentatif di mana pengguna Anda memerlukan kemampuan menganalisis metrik dengan dua cara dengan perubahan dan tanpa perubahan.

Untuk menerapkan perubahan dengan benar, Anda perlu mengubah perubahan yang masuk dan menyiapkan perubahan pada data untuk dimuat ke dalam gudang data. Gambar 12.10 mengilustrasikan bagaimana perubahan data yang diekstraksi dari sistem sumber ditransformasikan dan disiapkan untuk pemuatan data. Gambar ini menunjukkan penanganan setiap jenis perubahan tabel dimensi. Tipe 1, 2, dan 3 ditampilkan dengan jelas. Tinjau gambar tersebut dengan cermat untuk mendapatkan pemahaman yang baik tentang solusinya. Anda akan dihadapkan pada perubahan tabel dimensi setiap saat.

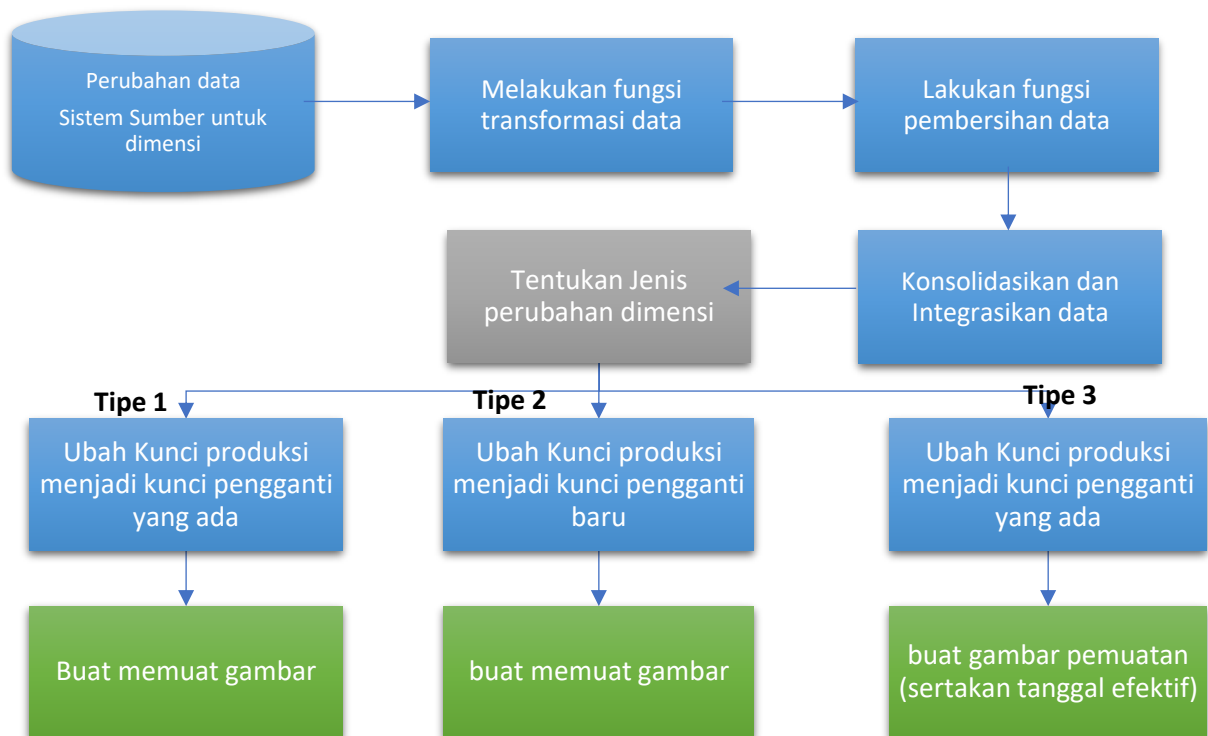
Bagaimana Menerapkan Transformasi

Kompleksitas dan luasnya transformasi data menunjukkan bahwa metode manual saja tidak akan cukup. Anda harus melampaui metode penulisan program konversi yang biasa ketika Anda menerapkan sistem operasional. Jenis transformasi data jauh lebih sulit dan menantang.

Metode yang mungkin ingin Anda terapkan bergantung pada beberapa faktor penting. Jika Anda mempertimbangkan untuk mengotomatisasi sebagian besar fungsi transformasi data, pertama-tama pertimbangkan apakah Anda memiliki waktu untuk memilih alat, mengkonfigurasi dan menginstalnya, melatih tim proyek tentang alat tersebut, dan mengintegrasikan alat tersebut ke dalam lingkungan gudang data. Alat transformasi data bisa mahal. Jika cakupan gudang data Anda sederhana, anggaran proyek mungkin tidak memiliki ruang untuk alat transformasi.

Mari kita lihat isu-isu yang berkaitan dengan penggunaan teknik manual dan penggunaan alat transformasi data. Dalam banyak kasus, kombinasi kedua metode yang

sesuai akan terbukti efektif. Temukan keseimbangan yang tepat berdasarkan kerangka waktu yang tersedia dan uang dalam anggaran.



Gambar 12.10 Ditransformasikan untuk perubahan dimensi.

Menggunakan Alat Transformasi Dalam beberapa tahun terakhir, alat transformasi telah meningkat pesat dalam hal fungsionalitas dan fleksibilitas. Meskipun tujuan yang diinginkan dari penggunaan alat transformasi adalah untuk menghilangkan metode manual sama sekali, dalam praktiknya hal ini tidak sepenuhnya memungkinkan. Bahkan jika Anda memiliki seperangkat alat transformasi yang paling canggih dan komprehensif, bersiaplah untuk menggunakan program internal di sana-sini.

Penggunaan alat otomatis tentu saja meningkatkan efisiensi dan akurasi. Sebagai spesialis transformasi data, Anda hanya perlu menentukan parameter, definisi data, dan aturan pada alat transformasi. Jika masukan Anda ke alat tersebut akurat, maka pekerjaan selanjutnya akan dilakukan secara efisien oleh alat tersebut.

Anda mendapatkan keuntungan besar dari penggunaan alat transformasi karena pencatatan metadata oleh alat tersebut. Saat Anda menentukan parameter dan aturan transformasi, parameter dan aturan ini disimpan sebagai metadata oleh alat tersebut. Metadata ini kemudian menjadi bagian dari keseluruhan komponen metadata gudang data. Ini mungkin digunakan bersama oleh komponen lain. Ketika terjadi perubahan pada fungsi transformasi karena perubahan aturan bisnis atau definisi data, Anda hanya perlu memasukkan perubahan tersebut ke dalam alat. Metadata untuk transformasi disesuaikan secara otomatis oleh alat ini.

Menggunakan Teknik Manual Ini adalah metode utama pada masa-masa awal pergudangan data ketika alat transformasi mulai muncul di pasar. Teknik manual mungkin masih memadai untuk gudang data yang lebih kecil. Dalam kasus seperti itu, program dan skrip yang dikodekan secara manual melakukan setiap transformasi data. Sebagian besar, program-program ini dijalankan di area pementasan data. Analis dan programmer yang telah memiliki pengetahuan dan keahlian mampu menghasilkan program dan skrip.

Tentu saja, metode ini melibatkan pengkodean dan pengujian yang rumit. Meskipun biaya awal mungkin masuk akal, pemeliharaan berkelanjutan dapat meningkatkan biayanya. Berbeda dengan alat otomatis, cara manual lebih rentan terhadap kesalahan. Mungkin juga beberapa program individual diperlukan di lingkungan Anda. Kerugian utama berkaitan dengan metadata. Alat otomatis merekam metadatanya sendiri, namun program internal harus dirancang berbeda jika Anda perlu menyimpan dan menggunakan metadata. Meskipun program internal mencatat metadata transformasi data pada awalnya, setiap kali terjadi perubahan pada aturan transformasi, metadata tersebut harus dipertahankan. Hal ini memberikan beban tambahan pada pemeliharaan program transformasi yang dikodekan secara manual.

12.5 PEMUATAN DATA

Secara umum disepakati bahwa fungsi transformasi berakhir segera setelah memuat gambar dibuat. Rangkaian fungsi utama berikutnya terdiri dari fungsi mengambil data yang telah disiapkan, menerapkannya ke gudang data, dan menyimpannya dalam database di sana. Anda membuat gambar pemuatan agar sesuai dengan file target yang akan dimuat di database gudang data.

Seluruh proses pemindahan data ke dalam repositori data warehouse dirujuk dalam beberapa cara. Anda pasti pernah mendengar ungkapan menerapkan data, memuat data, dan menyegarkan data. Agar lebih jelas, kami akan menggunakan frasa seperti yang ditunjukkan di bawah ini:

- ❖ Pemuatan awal—mengisi semua tabel gudang data untuk pertama kalinya. Beban tambahan—menerapkan perubahan berkelanjutan seperlunya secara berkala.
- ❖ Penyegaran penuh—menghapus seluruh isi satu atau beberapa tabel dan memuat ulang dengan data baru (pemuatan awal adalah penyegaran semua tabel).

Karena memuat gudang data mungkin memerlukan banyak waktu, pemuatan biasanya menjadi perhatian besar. Selama pemuatan, gudang data harus offline. Anda perlu menemukan jangka waktu kapan pemuatan dapat dijadwalkan tanpa memengaruhi pengguna gudang data Anda. Oleh karena itu, pertimbangkan untuk membagi seluruh proses pemuatan menjadi bagian-bagian yang lebih kecil dan mengisi beberapa file sekaligus. Ini akan memberi Anda dua manfaat. Anda mungkin dapat menjalankan beban yang lebih kecil secara paralel. Selain itu, Anda mungkin dapat menjaga beberapa bagian gudang data tetap aktif dan berjalan saat memuat bagian lainnya. Sulit untuk memperkirakan waktu berjalannya muatan, terutama muatan awal atau penyegaran total. Lakukan uji beban untuk memverifikasi kebenaran dan memperkirakan waktu pengoperasian.

Saat Anda menjalankan suatu beban, jangan berharap setiap rekaman dalam file gambar beban sumber berhasil diterapkan ke gudang data. Untuk rekaman yang Anda coba muat ke tabel fakta, kunci gabungan mungkin salah dan tidak sesuai dengan tabel dimensi. Menyediakan prosedur untuk menangani pemuatan gambar yang tidak dimuat. Juga, miliki rencana untuk jaminan kualitas dari catatan yang dimuat.

Jika area pementasan data dan database gudang data berada di server yang sama, hal itu akan menghemat upaya Anda untuk memindahkan gambar beban ke server gudang data. Namun jika Anda harus memindahkan gambar beban ke server gudang data, pertimbangkan opsi dengan hati-hati dan pilih yang paling sesuai untuk lingkungan Anda. Tautan Web, FTP, dan basis data adalah beberapa opsi. Anda harus mempertimbangkan kebutuhan bandwidth dan juga dampak transmisi pada jaringan. Pikirkan kompresi data dan buat rencana darurat. Apa metode umum untuk menerapkan data? Metode yang paling mudah adalah menulis program beban khusus.

Tergantung pada ukuran gudang data Anda, jumlah program beban bisa jadi besar. Mengelola beban sejumlah besar program dapat menjadi suatu tantangan. Selain itu, memelihara sejumlah besar program beban khusus akan menghabiskan banyak waktu dan tenaga. Utilitas beban yang disertakan dengan DBMS menyediakan metode pemuatan yang cepat. Anggaplah metode ini sebagai pilihan utama. Ketika file area pementasan dan repositori gudang data berada di server yang berbeda, link database berguna. Anda sudah mengetahui beberapa kekhawatiran dan kesulitan dalam memuat data. Tim proyek harus sangat memahami tantangan-tantangan umum sehingga dapat menghasilkan resolusi yang tepat. Sekarang mari kita beralih ke detail teknik dan proses pemuatan data.

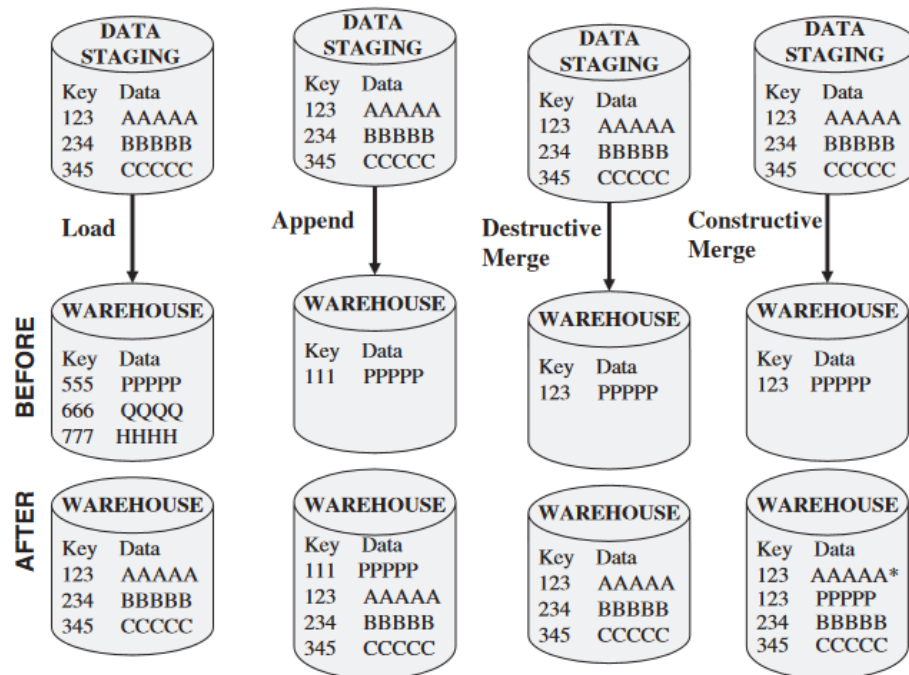
Menerapkan Data: Teknik dan Proses

Sebelumnya di bagian ini, kita mendefinisikan tiga jenis penerapan data ke gudang data: pemuatan awal, pemuatan tambahan, dan penyegaran penuh. Pertimbangkan bagaimana data diterapkan di masing-masing jenis ini. Mari kita ambil contoh data produk. Untuk pemuatan awal, Anda mengekstrak data untuk semua produk dari berbagai sistem sumber, mengintegrasikan dan mengubah data, lalu membuat gambar pemuatan untuk memuat data ke dalam tabel dimensi produk. Untuk beban tambahan, Anda mengumpulkan perubahan pada data produk untuk rekaman produk yang telah berubah di sistem sumber sejak ekstrak sebelumnya, menjalankan perubahan melalui proses integrasi dan transformasi, dan membuat rekaman keluaran untuk diterapkan ke dimensi produk meja. Penyegaran penuh mirip dengan pemuatan awal.

Dalam setiap kasus, Anda membuat file data untuk diterapkan ke tabel dimensi produk di gudang data. Bagaimana Anda bisa menerapkan data ke gudang? Apa saja modusnya? Data dapat diterapkan dalam empat mode berbeda berikut: memuat, menambahkan, menggabungkan destruktif, dan menggabungkan konstruktif. Pelajari Gambar 12.11 untuk memahami pengaruh penerapan data pada masing-masing empat mode ini. Mari kami jelaskan cara kerja setiap mode.

Load Jika tabel target yang akan dimuat sudah ada dan data sudah ada di dalam tabel, proses load akan menghapus data yang ada dan menerapkan data dari file yang masuk. Jika tabel sudah kosong sebelum dimuat, proses pemuatan cukup menerapkan data dari file yang masuk. Append Anda mungkin menganggap append sebagai perpanjangan dari beban. Jika data sudah ada dalam tabel, proses penambahan menambahkan data masuk tanpa syarat, mempertahankan data yang ada dalam tabel target. Jika rekaman masuk merupakan duplikat dari rekaman yang sudah ada, Anda dapat menentukan cara menangani duplikat masuk. Catatan yang masuk mungkin diperbolehkan untuk ditambahkan sebagai duplikat. Dalam opsi lain, rekaman duplikat yang masuk mungkin ditolak selama proses penambahan.

Penggabungan Destruktif Dalam mode ini, Anda menerapkan data yang masuk ke data target. Jika kunci utama rekaman masuk cocok dengan kunci rekaman yang sudah ada, perbarui rekaman target yang cocok. Jika record yang masuk adalah record baru tanpa kecocokan dengan record yang sudah ada, tambahkan record yang masuk ke tabel target.



Gambar 12.11 Cara penerapan data.

Penggabungan Konstruktif Mode ini sedikit berbeda dengan penggabungan destruktif. Jika kunci utama rekaman masuk cocok dengan kunci rekaman yang sudah ada, tinggalkan rekaman yang ada, tambahkan rekaman masuk, dan tandai rekaman yang ditambahkan sebagai pengganti rekaman lama. Sekarang mari kita pertimbangkan bagaimana mode penerapan data ke gudang data ini sesuai dengan tiga jenis beban. Kita akan membahasnya satu per satu.

Pemuatan Awal Katakanlah Anda dapat memuat seluruh gudang data dalam sekali proses. Sebagai variasi dari proses tunggal ini, katakanlah Anda dapat membagi beban menjadi subbeban terpisah dan menjalankan masing-masing subbeban ini sebagai beban tunggal.

Dengan kata lain, setiap proses pemuatan akan membuat tabel database dari awal. Dalam kasus ini, Anda akan menggunakan mode muat yang dibahas di atas.

Jika Anda memerlukan lebih dari satu proses untuk membuat satu tabel, dan proses pemuatan Anda untuk satu tabel harus dijadwalkan untuk dijalankan dalam beberapa hari, maka pendekatannya berbeda. Untuk menjalankan pemuatan awal tabel tertentu untuk pertama kalinya, gunakan mode pemuatan. Semua proses selanjutnya akan menerapkan data yang masuk menggunakan mode penambahan.

Pembuatan indeks pada pemuatan awal atau penyegaran penuh memerlukan pertimbangan khusus. Pembuatan indeks pada muatan massal bisa memakan waktu terlalu lama. Jadi jatuhkan indeks sebelum pemuatan agar pemuatan berjalan lebih cepat. Anda dapat membangun kembali atau membuat ulang indeks ketika pemuatan selesai. Beban Tambahan Ini adalah penerapan perubahan yang sedang berlangsung dari sistem sumber. Perubahan pada sistem sumber selalu terikat pada waktu tertentu, terlepas dari apakah perubahan tersebut didasarkan pada stempel waktu eksplisit dalam sistem sumber atau tidak. Oleh karena itu, Anda memerlukan metode untuk menjaga sifat periodik dari perubahan di gudang data.

Mari kita tinjau mode penggabungan konstruktif. Dalam mode ini, jika kunci utama dari rekaman yang masuk cocok dengan kunci dari rekaman yang sudah ada, rekaman yang ada akan dibiarkan di tabel target sebagaimana adanya dan rekaman yang masuk akan ditambahkan dan ditandai sebagai menggantikan rekaman yang lama. Jika stempel waktu juga merupakan bagian dari kunci utama atau jika stempel waktu disertakan dalam perbandingan antara rekaman yang masuk dan yang ada, maka penggabungan konstruktif dapat digunakan untuk menjaga sifat periodik dari perubahan tersebut. Ini merupakan penyederhanaan yang berlebihan terhadap rincian yang tepat tentang bagaimana penggabungan konstruktif dapat digunakan. Namun demikian, intinya adalah mode penggabungan konstruktif adalah metode yang tepat untuk beban tambahan. Rinciannya harus dikerjakan berdasarkan sifat dari masing-masing tabel target.

Apakah ada kasus di mana modus penggabungan destruktif dapat diterapkan? Bagaimana dengan tipe 1 yang perlahan-lahan mengubah dimensi? Dalam hal ini, perubahan pada record tabel dimensi dimaksudkan untuk memperbaiki kesalahan pada record yang ada. Catatan yang ada harus diganti dengan catatan masuk yang telah diperbaiki, sehingga Anda dapat menggunakan mode penggabungan destruktif. Mode ini juga berlaku untuk tabel target mana pun yang perspektif sejarahnya tidak penting.

Penyegaran Penuh Jenis penerapan data ini melibatkan penulisan ulang seluruh gudang data secara berkala. Terkadang, Anda juga dapat melakukan penyegaran sebagian untuk menulis ulang tabel tertentu saja. Penyegaran sebagian jarang terjadi karena setiap tabel dimensi terikat erat dengan tabel fakta. Sejauh menyangkut mode aplikasi data, penyegaran penuh mirip dengan pemuatan awal. Namun, dalam kasus penyegaran penuh, data sudah ada di tabel target sebelum data masuk diterapkan. Data yang ada harus dihapus sebelum menerapkan data yang masuk. Sama seperti pemuatan awal, mode pemuatan dan penambahan berlaku untuk penyegaran penuh.

Penyegaran Data versus Pembaruan

Setelah pemuatan awal, Anda dapat memelihara gudang data dan selalu memperbaruinya dengan menggunakan dua metode:

- ❖ Pembaruan—penerapan perubahan tambahan pada sumber data. Refresh—menyelesaikan pengisian ulang pada interval tertentu.

Secara teknis, penyegaran adalah opsi yang jauh lebih sederhana daripada pembaruan. Untuk menggunakan opsi pembaruan, Anda harus merancang strategi yang tepat untuk mengekstrak perubahan dari setiap sumber data. Kemudian Anda harus menentukan strategi terbaik untuk menerapkan perubahan pada data warehouse. Opsi penyegaran hanya melibatkan penggantian berkala tabel gudang data lengkap. Namun pekerjaan penyegaran membutuhkan waktu lama untuk dijalankan. Jika Anda harus menjalankan pekerjaan penyegaran setiap hari, Anda mungkin harus membiarkan gudang data tidak aktif dalam waktu yang sangat lama. Kasusnya menjadi lebih buruk jika database Anda memiliki tabel besar.

Apakah ada pedoman kapan refresh lebih baik daripada update atau sebaliknya? Gambar 12.12 menunjukkan grafik yang membandingkan refresh dengan update. Biaya penyegaran tetap konstan terlepas dari jumlah perubahan dalam sistem sumber. Jika jumlah perubahan bertambah, waktu dan upaya untuk melakukan penyegaran penuh tetap sama. Di sisi lain, biaya pembaruan bervariasi sesuai dengan jumlah catatan yang akan diperbarui.

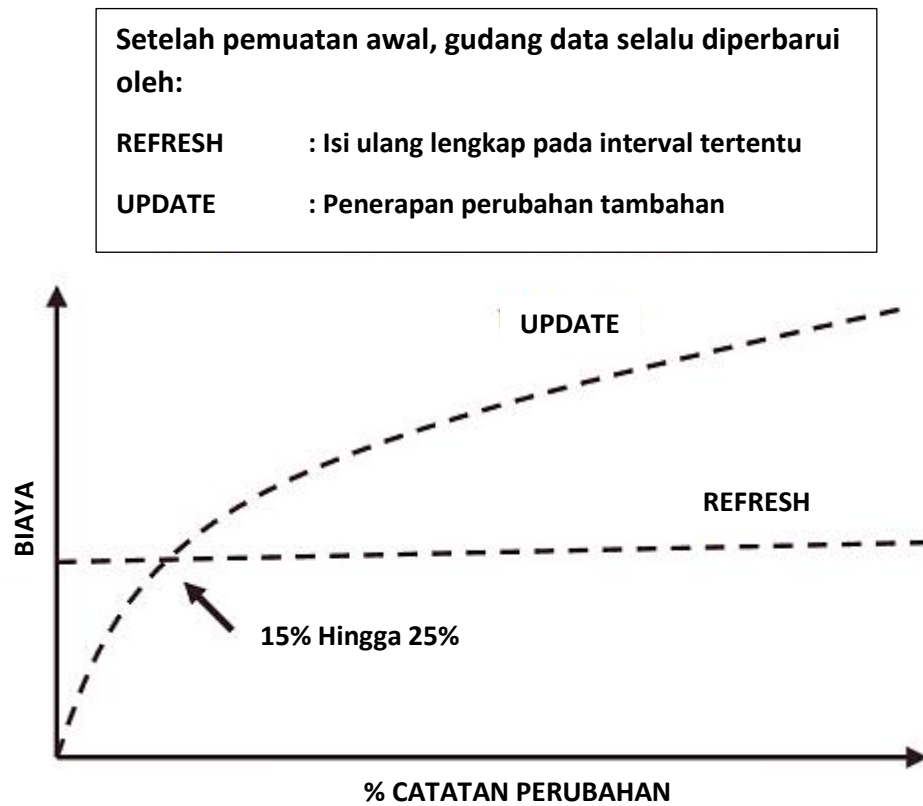
Jika jumlah rekaman yang akan diperbarui turun antara 15% dan 25% dari total jumlah rekaman, biaya pemuatan per rekaman cenderung sama, baik Anda memilih penyegaran penuh seluruh gudang data atau melakukan pembaruan. . Kisaran ini hanyalah panduan umum. Jika lebih dari 25% catatan sumber berubah setiap hari, pertimbangkan penyegaran penuh secara serius. Umumnya, administrator gudang data menggunakan proses pembaruan. Kadang-kadang, Anda mungkin ingin mengulang gudang data dengan penyegaran penuh ketika beberapa restrukturisasi besar atau perubahan massal serupa terjadi.

Prosedur Tabel Dimensi

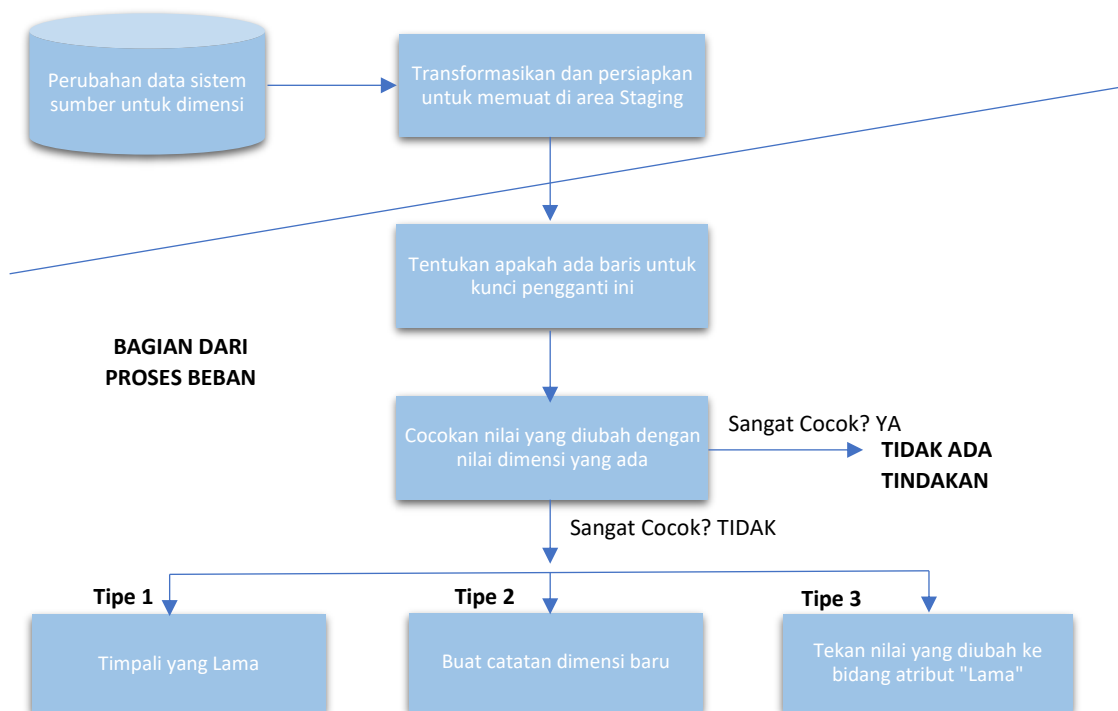
Di gudang data, tabel dimensi berisi atribut yang digunakan untuk menganalisis pengukuran dasar seperti penjualan dan biaya. Seperti yang Anda ketahui dengan baik, pelanggan, produk, waktu, dan wilayah penjualan adalah contoh tabel dimensi. Prosedur pemeliharaan tabel dimensi mencakup dua fungsi: pertama, pemuatan awal tabel; setelah itu, menerapkan perubahan secara berkelanjutan. Mari kita pertimbangkan dua persoalan.

Yang pertama adalah tentang kunci catatan di sistem sumber dan kunci catatan di gudang data. Untuk alasan yang dibahas sebelumnya, kami tidak menggunakan kunci sistem produksi untuk pencatatan di gudang data. Di gudang data, Anda menggunakan kunci yang dihasilkan sistem. Catatan dalam sistem sumber memiliki kuncinya sendiri. Oleh karena itu, sebelum data sumber dapat diterapkan ke tabel dimensi, baik untuk pemuatan awal atau untuk perubahan yang sedang berlangsung, kunci produksi harus dikonversi ke kunci yang dihasilkan sistem di gudang data. Anda dapat melakukan konversi kunci sebagai bagian dari fungsi transformasi atau Anda dapat melakukannya secara terpisah sebelum fungsi pemuatan sebenarnya. Sistem terjemahan kunci terpisah lebih disukai. Permasalahan berikutnya

berkaitan dengan penerapan perubahan dimensi tipe 1, tipe 2, dan tipe 3 pada data warehouse. Gambar 12.13 menunjukkan bagaimana tipe-tipe yang berbeda ini ditangani.



Gambar 12.12 Penyegaran versus pembaruan.



Gambar 12.13 Memuat perubahan pada tabel dimensi.

Tabel Fakta: Riwayat dan Beban Tambahan

Kunci dari tabel fakta adalah rangkaian kunci dari tabel dimensi. Oleh karena itu, karena alasan ini, rekaman dimensi dimuat terlebih dahulu. Kemudian, sebelum memuat setiap rekaman tabel fakta, Anda harus membuat kunci gabungan untuk rekaman tabel fakta dari kunci rekaman dimensi terkait.

Berikut beberapa tip untuk memuat riwayat tabel fakta:

- ❖ Mengidentifikasi data historis yang berguna dan menarik untuk data warehouse.
- ❖ Mendefinisikan dan menyempurnakan aturan bisnis ekstrak.
- ❖ Mengambil statistik audit untuk dikaitkan dengan sistem operasional.
- ❖ Melakukan pencarian kunci pengganti tabel fakta.
- ❖ Meningkatkan isi tabel fakta.
- ❖ Merestrukturisasi data.
- ❖ Siapkan file beban.

Di bawah ini adalah beberapa komentar berguna tentang penambahan beban untuk tabel fakta:

- Ekstrak tambahan untuk tabel fakta
- Terdiri dari transaksi baru
- Terdiri dari update transaksi
- Gunakan log transaksi database untuk pengambilan data
- Pemuatan tambahan untuk tabel fakta
- Muat sesering mungkin
- Gunakan file dan indeks yang dipartisi
- Menerapkan teknik pemrosesan paralel

12.6 RINGKASAN ETL

Saat ini Anda seharusnya sudah yakin sepenuhnya bahwa fungsi ekstraksi, transformasi, dan pemuatan data untuk gudang data mencakup bidang yang sangat luas. Fungsi konversi yang biasanya dikaitkan dengan pengembangan sistem operasional apa pun tidak sebanding dengan tingkat dan kompleksitas fungsi ETL dalam lingkungan data warehouse. Fungsi ekstraksi data di gudang data mencakup beberapa sistem sumber yang bervariasi. Sebagai pengembang gudang data, Anda perlu memeriksa dengan cermat tantangan yang ditimbulkan oleh beragam sistem sumber Anda dan menemukan metode ekstraksi data yang tepat. Kami telah membahas sebagian besar metode umum. Ekstraksi data untuk gudang data bukanlah peristiwa yang terjadi satu kali saja; ini adalah fungsi berkelanjutan yang dilakukan pada interval yang sangat sering.

Ada banyak jenis transformasi data di gudang data dengan banyak tugas berbeda. Ini bukan sekedar konversi lapangan ke lapangan. Dalam diskusi kita, kita membahas banyak jenis transformasi data yang umum. Daftar tipe yang dapat kami pertimbangkan sama sekali tidak

lengkap atau lengkap. Di lingkungan gudang data, Anda akan menemukan jenis transformasi data tambahan.

Bagaimana dengan beban data di gudang data dibandingkan dengan beban sistem operasional baru? Untuk penerapan sistem operasional baru, Anda mengonversi dan memuat data satu kali untuk memulai sistem baru. Pemuatan data di gudang data tidak berhenti pada implementasi awal. Sama seperti ekstraksi dan transformasi, pemuatan data bukan sekadar aktivitas awal untuk memulai data warehouse. Selain pemuatan data awal, Anda memiliki pemuatan data tambahan yang berkelanjutan dan penyegaran penuh secara berkala. Untungnya, banyak vendor telah mengembangkan alat canggih untuk ekstraksi data, transformasi data, dan pemuatan data. Anda tidak lagi dibiarkan sendiri untuk menangani tantangan ini dengan metode manual yang tidak canggih. Anda memiliki solusi vendor yang fleksibel dan sesuai. Alat vendor mencakup berbagai pilihan fungsional. Anda memiliki alat yang efektif untuk menjalankan fungsi di setiap bagian proses ETL. Alat dapat mengekstrak data dari berbagai sumber, melakukan sejumlah fungsi transformasi, dan melakukan muatan massal serta muatan tambahan. Mari kita tinjau beberapa opsi yang Anda miliki sehubungan dengan alat ETL.

Opsi Alat ETL

Vendor telah mendekati tantangan ETL dan mengatasinya dengan menyediakan alat yang termasuk dalam tiga kategori fungsional berikut:

1. Mesin transformasi data; ini terdiri dari algoritma manipulasi data yang dinamis dan canggih. Rangkaian alat ini menangkap data dari sekumpulan sistem sumber yang ditentukan pada interval yang ditentukan pengguna, melakukan transformasi data yang rumit, mengirimkan hasilnya ke lingkungan target, dan menerapkan data ke file target. Alat-alat ini memberi Anda fleksibilitas maksimum untuk menunjuk ke berbagai sistem sumber, memilih metode transformasi data yang sesuai, dan menerapkan beban penuh dan beban tambahan. Fungsionalitas alat ini mencakup keseluruhan proses ETL.
2. Pengambilan data melalui replikasi; sebagian besar alat ini menggunakan log pemulihan transaksi yang dikelola oleh DBMS. Perubahan pada sistem sumber yang dicatat dalam log transaksi direplikasi hampir secara real-time ke area pementasan data untuk diproses lebih lanjut. Beberapa alat menyediakan kemampuan untuk mereplikasi data melalui penggunaan pemicu basis data. Prosedur tersimpan khusus dalam database ini memberi sinyal kepada agen replikasi untuk menangkap dan memindahkan perubahan.
3. Pembuat kode; Ini adalah alat yang secara langsung berhubungan dengan ekstraksi, transformasi, dan pemuatan data. Alat ini mengaktifkan proses dengan menghasilkan kode program untuk menjalankan fungsi-fungsi ini. Generator kode membuat program ekstraksi dan transformasi data 3GL/4GL. Anda memberikan parameter sumber data dan tata letak target beserta aturan bisnis. Alat tersebut menghasilkan sebagian besar kode program dalam beberapa bahasa pemrograman umum. Bila Anda ingin menambahkan lebih banyak kode untuk menangani jenis transformasi yang tidak

tercakup dalam alat ini, Anda dapat melakukannya dengan kode program Anda sendiri. Kode yang dihasilkan secara otomatis oleh alat ini memiliki titik keluar di mana Anda dapat menambahkan kode untuk menangani kondisi khusus.

Saat ini, beberapa vendor DBMS menyertakan kemampuan ETL dalam produk database mereka dengan sedikit atau tanpa biaya tambahan. Pilihan ini menarik bagi banyak organisasi karena implikasi biayanya. Selain itu, jika fungsionalitas ETL dan database berasal dari vendor yang sama, masalah dukungan dapat diminimalkan.

Lebih khusus lagi, apa yang dapat dilakukan alat ETL? Tinjau daftar berikut dan saat Anda membaca setiap item, pertimbangkan apakah Anda memerlukan fitur tersebut di alat ETL untuk lingkungan Anda:

- ◆ Ekstraksi data dari berbagai database relasional vendor terkemuka
- ◆ Ekstraksi data dari database lama, file yang diindeks, dan file datar
- ◆ Transformasi data dari satu format ke format lainnya dengan variasi kolom sumber dan target
- ◆ Melakukan konversi standar, pemformatan ulang kunci, dan perubahan struktural
- ◆ Penyediaan jejak audit dari sumber ke sasaran
- ◆ Penerapan aturan bisnis untuk ekstraksi dan transformasi
- ◆ Menggabungkan beberapa catatan dari sistem sumber menjadi satu catatan target yang terintegrasi
- ◆ Pencatatan dan pengelolaan metadata

Menekankan kembali Metadata ETL

Bab 9 membahas metadata gudang data dengan sangat rinci. Kami membahas peran dan pentingnya metadata dalam tiga area fungsional utama data warehouse. Kami meninjau penangkapan dan penggunaan metadata di tiga bidang akuisisi data, penyimpanan data, dan penyampaian informasi. Metadata dalam akuisisi data dan penyimpanan data berhubungan dengan fungsi ETL.

Saat Anda menggunakan alat vendor untuk menjalankan sebagian atau seluruh fungsi ETL, sebagian besar alat ini mencatat dan mengelola metadatanya sendiri. Meskipun metadatanya berada dalam format kepemilikan alat tersebut, metadata tersebut dapat digunakan dan tersedia. Metadata ETL berisi informasi tentang sistem sumber, pemetaan antara data sumber dan struktur data target gudang data, transformasi data, dan pemuatan data.

Namun seperti yang Anda ketahui, alat yang Anda pilih mungkin tidak secara eksklusif menjalankan semua fungsi ETL. Anda harus meningkatkan upaya ETL Anda dengan program internal. Dalam setiap fungsi ekstraksi data, transformasi data, dan pemuatan data, Anda dapat menggunakan program yang ditulis oleh tim proyek. Tergantung pada situasi di lingkungan Anda, jumlah program internal ini mungkin sangat bervariasi. Meskipun program internal memberi Anda lebih banyak kendali dan fleksibilitas, ada satu kelemahan. Berbeda dengan alat ETL, program internal tidak mencatat atau mengelola metadata. Anda harus melakukan upaya khusus untuk menangani metadata. Meskipun kami telah meninjau metadata secara ekstensif di bab sebelumnya, kami ingin menegaskan kembali bahwa Anda

perlu memberikan perhatian khusus dan memastikan bahwa metadata tidak diabaikan saat Anda menggunakan program internal untuk fungsi ETL. Semua aturan bisnis, informasi data sumber, pemetaan sumber-ke-target, transformasi, dan informasi pemuatan harus dicatat secara manual di direktori metadata. Ini sangat penting untuk membuat komponen metadata Anda lengkap dan akurat.

Ringkasan dan Pendekatan ETL

Mari kita rangkum fungsi-fungsi yang dibahas dalam bab ini dengan Gambar 12.14. Lihatlah gambar dan lakukan tinjauan singkat terhadap fungsi-fungsi utama. Apa pendapat Anda tentang ukuran bab ini dan topik yang dibahas? Panjang bab ini saja sudah menyoroti pentingnya dan kompleksitas fungsi ekstraksi, transformasi, dan pemuatan data. Kenapa begitu? Berkali-kali, keragaman dan sifat heterogen dari sistem sumber menjadi alasan utama untuk memberikan perhatian khusus pada ETL. Di satu sisi, keragaman dan heterogenitas menambah tantangan ekstraksi data. Namun jika Anda mempertimbangkan jumlah sistem sumber yang berbeda, semakin banyak jumlahnya, semakin intens dan kompleks fungsinya. Kemungkinan besar akan terjadi lebih banyak inkonsistensi dan diperkirakan akan ada lebih banyak variasi dari standar.

Namun yang diperlukan hanyalah pendekatan sistematis dan praktis. Kapan pun Anda dapat membagi suatu tugas menjadi dua, lakukanlah tanpa ragu-ragu. Misalnya, mencari cara untuk memecah muatan awal menjadi beberapa submuatan. Selain itu, analisis terperinci juga penting. Anda tidak dapat menganggap enteng sistem sumber apa pun. Setiap sistem sumber mungkin mempunyai tantangan tersendiri. Pelajari detailnya. Luangkan waktu yang cukup dalam pemetaan sumber-ke-target. Buatlah daftar awal transformasi data dan biarkan daftar ini berkembang. Lakukan lebih banyak analisis dan tambahkan ke daftar.

Anda harus hidup dengan banyak data setiap hari. Pemuatan tambahan yang sering sangat penting untuk menjaga gudang data Anda tetap mutakhir. Cobalah untuk mengotomatiskan pemuatan tambahan sebanyak mungkin. Jaga agar program internal tetap pada tingkat yang wajar. Pemeliharaan metadata secara manual dapat memberikan beban yang besar. Kami menyadari fungsi ETL memakan waktu, rumit, dan sulit; namun demikian, hal-hal tersebut sangat penting. Setiap kekurangan dalam proses ETL muncul di gudang data. Pengguna Anda pada akhirnya akan menggunakan informasi yang salah. Menurut Anda, keputusan seperti apa yang akan mereka ambil dengan informasi yang salah dan tidak lengkap?

12.7 PENDEKATAN INTEGRASI LAINNYA

Setelah mempertimbangkan semua yang dibahas dalam bab ini tentang ekstraksi, transformasi, dan pemuatan data, Anda dapat memahami bahwa ETL adalah metodologi terdepan untuk integrasi data perusahaan. Inti dari ETL adalah konsolidasi data perusahaan, mungkin dari banyak sumber berbeda. Namun demikian, seiring kemajuan perusahaan dalam integrasi informasi untuk memberikan satu pandangan perusahaan yang terpadu, ada pendekatan integrasi lain yang perlu dipertimbangkan.

Kami sekarang akan meninjau dua pendekatan utama untuk integrasi informasi tanpa terlalu banyak membahas rincian teknis. Kami juga akan membandingkan masing-masing pendekatan ETL ini untuk membantu Anda membuat pilihan di lingkungan khusus Anda. Dua pendekatan lainnya umumnya dikenal sebagai Enterprise Information Integration (EII) dan Enterprise Application Integration (EAI).

Integrasi Informasi Perusahaan (EII)

Pendekatan integrasi ini didasarkan pada konfigurasi dan penyediaan pandangan terpadu atas data perusahaan yang tersebar. Namun integrasinya bukan secara fisik melainkan virtual. Produk EII berupaya memberikan pandangan bisnis virtual atas data yang tersebar. Berbagai kumpulan sumber data yang berbeda dianggap sebagai federasi informasi. Koleksinya mungkin mencakup data transaksi bisnis, data mart, dan bahkan beberapa kumpulan konten data tidak terstruktur. EII mendukung pendekatan federasi informasi terhadap integrasi.

Aplikasi yang perlu mengakses data secara virtual melihat kumpulan data yang tersebar seolah-olah semuanya berada dalam satu database. Kueri diarahkan terhadap tampilan virtual dari data yang diperlukan yang mungkin berada di berbagai sumber. Produk EII memecah kueri menjadi subkueri. Setiap sub-kueri diarahkan ke kumpulan data yang sesuai. Ketika semua subkueri selesai dieksekusi, hasilnya digabungkan menjadi hasil terpadu dan disajikan kepada pengguna. Teknologi DBMS dan XML relasional mendorong penerapan alat EII. Produk EII ini mendukung antarmuka data SQL dan XML.

EII dibandingkan dengan ETL

- ✘ EII mungkin berlaku untuk gudang data gabungan; namun arsitektur gabungan hampir selalu tidak direkomendasikan untuk gudang data baru. Oleh karena itu, EII umumnya bukan pengganti total ETL.
- ✘ Saat menggunakan ETL untuk menangkap dan mengkonsolidasikan data, mungkin terjadi latensi. Namun, jika data operasional sangat fluktuatif dan kueri serta analisis real-time sangat penting, maka EII dapat dipertimbangkan sebagai pilihan.
- ✘ Kueri melalui EII menyiratkan akses data ke sistem operasional saat ini serta data historis. Jika intelijen bisnis layanan mandiri dibuka untuk pengguna yang tidak dikenal atau kurang terlatih, pertanyaan yang tidak menentu melalui EII dapat menyebabkan degradasi sistem yang parah dalam sistem operasional.
- ✘ Dalam kasus dimana volume data terlalu besar untuk konsolidasi dan hanya sebagian kecil dari data konsolidasi yang digunakan secara teratur, pengambilan dan konsolidasi data menggunakan ETL mungkin tidak dapat dibenarkan. EII mungkin terbukti menjadi alternatif yang layak.
- ✘ Transformasi data sebagai bagian dari ETL terkadang menjadi sangat kompleks. Dalam situasi ini, transformasi data EII yang kompleks dengan cepat dapat mengakibatkan overhead sistem yang sangat tinggi.
- ✘ ETL memungkinkan pembangunan data historis dari waktu ke waktu dengan cara yang sistematis. Dimanapun tren dan pola historis menjadi fokus intelijen bisnis, EII mungkin bukan pendekatan pilihan.

Integrasi Aplikasi Perusahaan (EAI)

Dalam pendekatan ini, integrasi mengambil bentuk menghubungkan sistem aplikasi dengan memungkinkan mereka untuk bertukar data, pesan, dan transaksi. Aplikasi tertentu dapat mengakses data yang diperlukan secara transparan tanpa mengetahui format bagaimana data disimpan atau di mana data disimpan. Pada dasarnya, EAI umumnya diterapkan pada pemrosesan transaksi real-time dalam suatu sistem operasional.

Jadi, bagaimana kinerja EAI sebagai pendekatan integrasi? EAI dapat digunakan untuk memindahkan data dari peristiwa real-time ke pendekatan integrasi data lainnya seperti ETL. EAI dirancang untuk mengangkut sejumlah kecil data secara real-time. Transformasi data apa pun di EAI masih belum sempurna. Pendekatan ini tidak dimaksudkan untuk pengambilan, transformasi, dan pengiriman data dalam jumlah besar.

EAI dibandingkan dengan ETL

- 1) Pertama, EAI tidak dapat dianggap sebagai pendekatan integrasi yang bersaing dengan ETL. Kedua pendekatan ini dapat digabungkan untuk saling melengkapi. EAI dapat digunakan sebagai mekanisme penghasil masukan untuk ETL; ETL dapat dianggap sebagai mekanisme layanan untuk EAI.
- 2) Dimanapun aplikasi ETL memerlukan data berlatensi rendah, EAI dapat digunakan untuk menyediakan antarmuka EAI-ke-ETL untuk ekstraksi dan konsolidasi data secara real-time.
- 3) Meskipun antarmuka EAI-ke-ETL yang dinamis diinginkan, sebagian besar organisasi menggunakan EAI untuk membuat kumpulan data yang selanjutnya dimasukkan ke dalam aplikasi ETL.

RINGKASAN BAB

- ◆ Fungsi ETL dalam gudang data adalah hal yang paling penting, menantang, memakan waktu, dan padat karya.
- ◆ Ekstraksi data rumit karena sistem sumbernya berbeda; transformasi data sulit dilakukan karena banyaknya tugas; pemuatan data merupakan tantangan karena volume data.
- ◆ Terdapat beberapa teknik ekstraksi data, masing-masing dengan kelebihan dan kekurangannya. Pilih teknik yang tepat berdasarkan kondisi di lingkungan Anda.
- ◆ Fungsi transformasi data mencakup konversi, pembersihan, konsolidasi, dan integrasi data. Menerapkan fungsi transformasi menggunakan kombinasi alat khusus dan perangkat lunak yang dikembangkan sendiri.
- ◆ Fungsi pemuatan data berkaitan dengan pemuatan awal, pemuatan tambahan berkala secara teratur, dan penyegaran penuh dari waktu ke waktu. Empat metode untuk menerapkan data adalah: memuat, menambahkan, penggabungan destruktif, dan penggabungan konstruktif.
- ◆ Alat untuk fungsi ETL terbagi dalam tiga kategori fungsional yang luas: mesin transformasi data, pengambilan data melalui replikasi, dan pembuat kode.

- ◆ Integrasi Informasi Perusahaan (EII) dan Integrasi Aplikasi Perusahaan (EAI) adalah beberapa pendekatan lain terhadap integrasi data.

LATIHAN SOAL

1. Berikan tiga alasan mengapa menurut Anda fungsi ETL paling menantang dalam lingkungan data warehouse.
2. Sebutkan lima jenis aktivitas yang merupakan bagian dari proses ETL. Manakah dari berikut ini yang memakan waktu?
3. Keberagaman sistem sumber yang luar biasa merupakan alasan utama kompleksitasnya. Apa kamu setuju? Jika ya, jelaskan alasannya secara singkat.
4. Apa dua kategori umum data yang disimpan dalam sistem operasional sumber? Berikan dua contoh untuk masing-masingnya.
5. Sebutkan lima jenis tugas transformasi utama. Berikan masing-masing contohnya.
6. Jelaskan secara singkat masalah identifikasi entitas dalam integrasi dan konsolidasi data. Bagaimana Anda mengatasi masalah ini?
7. Apa yang dimaksud dengan restrukturisasi utama? Jelaskan mengapa itu diperlukan.
8. Tentukan muatan awal, muatan tambahan, dan penyegaran penuh.
9. Jelaskan perbedaan antara penggabungan destruktif dan penggabungan konstruktif untuk menerapkan data ke repositori data warehouse. Kapan Anda menggunakan mode ini?
10. Kapan penyegaran data penuh lebih baik daripada pemuatan bertahap? Dapatkah Anda memikirkan sebuah contoh?

DAFTAR PUSTAKA

- Abai, N. H. Z., Yahaya, J. H., & Deraman, A. (2013). User requirement analysis in data warehouse design: a review. *Procedia Technology*, *11*, 801-806.
- Adamson, Christopher, dan Joseph A. Peppard. (2014). "Master Data Management and Data Governance." New York: McGraw-Hill Education.
- Agarwal, Anol, dan Sid Adelman. (2016). "Data Warehouse Design: Modern Principles and Methodologies." New York: McGraw-Hill Education.
- Ariyachandra, T., & Watson, H. (2010). Key organizational factors in data warehouse architecture selection. *Decision support systems*, *49*(2), 200-212.
- Ariyachandra, T., & Watson, H. J. (2006). Which data warehouse architecture is most successful?. *Business intelligence journal*, *11*(1), 4.
- Ballou, D. P., & Tayi, G. K. (1999). Enhancing data quality in data warehouse environments. *Communications of the ACM*, *42*(1), 73-78.
- Chau, K. W., Cao, Y., Anson, M., & Zhang, J. (2003). Application of data warehouse and decision support system in construction management. *Automation in construction*, *12*(2), 213-224.
- Date, C. J. (2019). "An Introduction to Database Systems." Boston: Addison-Wesley Professional.
- Elmasri, Ramez, dan Shamkant B. Navathe. (2015). "Fundamentals of Database Systems." Boston: Pearson Education.
- Gardner, S. R. (1998). Building the data warehouse. *Communications of the ACM*, *41*(9), 52-60.
- Golfarelli, M., & Rizzi, S. (1998, November). A methodological framework for data warehouse design. In *Proceedings of the 1st ACM international workshop on Data warehousing and OLAP* (pp. 3-9).
- Golfarelli, M., & Rizzi, S. (1999). Designing the data warehouse: Key steps and crucial issues. *Journal of computer science and Information Management*, *2*(3), 88-100.
- Golfarelli, Matteo, Stefano Rizzi, dan Dario Mantovani. (2009). "Data Warehouse Design: Modern Principles and Methodologies." New York: McGraw-Hill Education.
- Gupta, Ashish. (2019). "Data Warehousing: Concepts, Techniques, Products and Applications." New Delhi: PHI Learning Pvt. Ltd.
- Gupta, H., & Mumick, I. S. (2005). Selection of views to materialize in a data warehouse. *IEEE Transactions on Knowledge and Data Engineering*, *17*(1), 24-43.
- Hackathorn, Richard D. (2000). "Data Warehouse Design Solutions." New York: John Wiley & Sons.
- Hsu, John. (2019). "Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance." New York: Packt Publishing.
- Hughes, Larry. (2019). "Data Warehousing For Dummies." Hoboken: For Dummies.
- Hüsemann, B., Lechtenbörger, J., & Vossen, G. (2000). *Conceptual data warehouse design* (Vol. 168). Münster, Germany: Universität Münster. Angewandte Mathematik und Informatik.
- Imhoff, C., Gallemmo, N., & Geiger, J. G. (2003). *Mastering data warehouse design: relational and dimensional techniques*. John Wiley & Sons.

- Imhoff, Claudia, Nicholas Gallemmo, dan Jonathan G. Geiger. (2017). "Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance." New York: John Wiley & Sons.
- Imhoff, Claudia. (2018). "Mastering Data Warehousing: Theory and Practice." Boca Raton: CRC Press.
- Inmon, W. H. (1995). What is a data warehouse. *Prism Tech Topic*, 1(1), 1-5.
- Inmon, William H. (2005). "Building the Data Warehouse." Indianapolis: John Wiley & Sons.
- Inmon, William H., dan Daniel Linstedt. (2011). "Building the Data Warehouse." Indianapolis: John Wiley & Sons.
- Jarke, M., Lenzerini, M., Vassiliou, Y., & Vassiliadis, P. (2002). *Fundamentals of data warehouses*. Springer Science & Business Media.
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Kimball, Ralph, dan Margy Ross. (2002). "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling." New York: John Wiley & Sons.
- Kimball, Ralph, dan Margy Ross. (2017). "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling." San Francisco: Wiley.
- Kimball, Ralph, Margy Ross, dan Bob Becker. (2018). "The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data." New York: John Wiley & Sons.
- Kimball, Ralph, Margy Ross, dan Warren Thornthwaite. (2008). "The Data Warehouse Lifecycle Toolkit." New York: John Wiley & Sons.
- Kimball, Ralph, Margy Ross, dan Warren Thornthwaite. (2016). "The Data Warehouse Lifecycle Toolkit." San Francisco: Wiley.
- Lahdenmaki, Jani, dan Henry Lipponen. (2018). "Agile Data Warehousing Project Management: Business Intelligence Systems Using Scrum." Boston: Pearson Education.
- List, B., Bruckner, R. M., Machaczek, K., & Schiefer, J. (2002). A comparison of data warehouse development methodologies case study of the process warehouse. In *Database and Expert Systems Applications: 13th International Conference, DEXA 2002 Aix-en-Provence, France, September 2–6, 2002 Proceedings 13* (pp. 203-215). Springer Berlin Heidelberg.
- Loshin, David. (2011). "The Practitioner's Guide to Data Quality Improvement." Burlington: Morgan Kaufmann.
- Maier, Ronny. (2020). "Building a Data Warehouse: With Examples in SQL Server." New York: Apress.
- McPherson, Mark. (2017). "Data Warehousing and Business Intelligence: Concepts, Methodologies, Tools, and Applications." Hershey: IGI Global.
- Motenko, H., Neuhauser, S. B., O'keefe, M., & Richardson, J. E. (2015). MouseMine: a new data warehouse for MGI. *Mammalian Genome*, 26, 325-330.
- Mundy, Joy, dan Warren Thornthwaite. (2016). "The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset." New York: John Wiley & Sons.
- O'Neil, Patrick E., dan Elizabeth J. O'Neil. (2015). "Database: Principles, Programming, and Performance." San Francisco: Morgan Kaufmann.

- Paulraj, Ponniah. (2018). "Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals." Indianapolis: John Wiley & Sons.
- Rahm, Erhard, dan Hong Hai Do. (2000). "Data Cleaning: Problems and Current Approaches." *IEEE Data Engineering Bulletin*, 23(4), 3-13.
- Rahm, Erhard, dan Hong Hai Do. (2021). "Data Cleaning: Problems and Current Approaches." New York: Springer.
- Rai, Ramesh. (2016). "Data Warehouse and Business Intelligence: Concepts and Tools." New Delhi: Pearson India.
- Redmond, Eric, dan Jim Rymarczyk. (2004). "Data Warehousing: Architecture and Implementation." Upper Saddle River: Prentice Hall.
- Rosenthal, A., & Sciore, E. (2000, June). View security as the basis for data warehouse security. In *DMDW* (p. 8).
- Shroff, Gautam. (2014). "Data Warehousing: Using the Wal-Mart Model." New Delhi: Tata McGraw-Hill Education.
- Sinclair, Adam. (2018). "Data Warehouse Design: Understanding Star Schema." Boston: Addison-Wesley Professional.
- Sisodia, Anil K., dan Parimal Vyas. (2017). "Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals." New Delhi: Wiley India Pvt. Ltd.
- Siva, R. S. (2016). "Data Warehouse Design Solutions: Implementing the Kimball Dimensional Modeling Techniques." New Delhi: BPB Publications.
- Smith, Paulraj. (2017). "Data Warehousing Fundamentals: Comprehensive Guide for IT Professionals." New Delhi: Wiley India Pvt. Ltd.
- Theodoratos, D., & Sellis, T. (1997, August). Data warehouse configuration. In *VLDB* (Vol. 97, pp. 126-135).
- Tryfona, N., Busborg, F., & Borch Christiansen, J. G. (1999, November). starER: A conceptual model for data warehouse design. In *Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP* (pp. 3-8).
- Vaisman, A., & Zimányi, E. (2014). Data warehouse systems. *Data-Centric Systems and Applications*.
- Watson, Hugh J. (2013). "Data Management: Databases & Organizations." Hoboken: John Wiley & Sons.
- Widom, J. (1995, December). Research problems in data warehousing. In *Proceedings of the fourth international conference on Information and knowledge management* (pp. 25-30).
- Williams, B., dan C. H. Kip. (2016). "Data Warehouse Design Solutions." New York: John Wiley & Sons.
- Wrembel, Robert, dan Christian Koncilia. (2019). "ETL and Data Warehousing: Issues and Challenges in a Modern Context." Berlin: Springer.

PERGUDANGAN DATA (Data Warehousing)

JILID 1

Dr. Budi Raharjo, S.Kom.,M.Kom.,MM.

BIODATA PENULIS



Dr. Budi Raharjo, S.Kom, M.Kom, MM lahir di Semarang, tanggal 22 Februari 1985. Beliau adalah Alumni dari Universitas Bina Nusantara (BINUS University) Jakarta dan juga alumni Universitas Kristen Satya wacana (UKSW) Salatiga. Dr. Budi Raharjo telah menjadi Dosen pada Universitas STEKOM pada mata kuliah Kepemimpinan (Leadership), mata kuliah Pengantar Akuntansi, Manajemen Proses, Manajemen Akuntansi dan Manajemen Resiko Bisnis. Selain sebagai dosen Universitas STEKOM, Dr. Budi Raharjo, M.Kom, MM juga mempunyai bisnis sendiri dalam bidang perhotelan dan juga sebagai wirausaha dalam bidang pemasok unggas (ayam) beku, ke berbagai kota besar, khususnya Jakarta dan sekitarnya.

Pengalaman beliau berwirausaha menjadi bekal utama dalam penulisan buku ajar yang diterbitkan oleh Yayasan Prima Agus Teknik (YPAT) Semarang. Oleh sebab itu bukunya berisi langkah langkah praktis yang mudah diikuti oleh para mahasiswa, saat mahasiswa mengikuti proses perkuliahan pada Universitas Sains dan Teknologi Komputer (Universitas STEKOM). Jabatan struktural yang di embannya saat ini adalah Wakil Rektor 1 (Akademik) Universitas STEKOM Semarang.



YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

ISBN 978-623-8642-04-5 (jil.1 PDF)



PERGUDANGAN DATA (Data Warehousing)

JILID 1

Dr. Budi Raharjo, S.Kom.,M.Kom.,MM.



YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id