

Contoh Membuat Data Model untuk Data Warehouse

Djoni Darmawikarta

djoni_darmawikarta@yahoo.ca

Lisensi Dokumen:

Copyright © 2003 IlmuKomputer.Com

Seluruh dokumen di IlmuKomputer.Com dapat digunakan, dimodifikasi dan disebarkan secara bebas untuk tujuan bukan komersial (nonprofit), dengan syarat tidak menghapus atau merubah atribut penulis dan pernyataan copyright yang disertakan dalam setiap dokumen. Tidak diperbolehkan melakukan penulisan ulang, kecuali mendapatkan ijin terlebih dahulu dari IlmuKomputer.Com.

Sebelum mulai membuat data model untuk data warehouse, sebaiknya spesifikasi kebutuhan informasi dan data yang tersedia (sumber data) sudah harus diketahui.

Untuk contoh dalam tulisan ini, kita sederhanakan spesifikasi kebutuhan informasi dan sumber data yaitu hanya spesifikasi kebutuhan fungsional. Tentu saja dalam kenyataan sesungguhnya akan lebih lengkap dan detil. Kita tidak membahas informasi teknis, misalnya spesifikasi kecepatan respons dan penggunaan teknologi internet (web access). Juga tentang sumber data, kita tidak membahas spesifikasi teknis, misalnya format dan ukuran data (field type & length), volume, dan kualitasnya.

Spesifikasi kebutuhan fungsional informasi secara teknis menyatakan data apa yang harus disimpan didalam data warehouse untuk memenuhi analisa yang akan dilakukan oleh pemakai. Ini dapat dinyatakan dalam bentuk pertanyaan yang harus dapat dijawab oleh data didalam data warehouse, misalnya untuk contoh kita sebagai berikut:

Berapa besar penjualan dan laba per bulan, per produk, per tipe pelanggan/pembeli (grosir, pabrikan, industri rumah, atau exportir), jenis industri (dari pelanggan/pembeli), dan kemana barang pesanan dikirim (propinsi atau nama negara)

Pertanyaan pemakai semacam diatas biasa disebut analisa multi-dimensi (*multidimensional analysis*). Besaran (*measure*) yang ditanyakan (dianalisa) adalah 'penjualan' dan 'laba'. Batasan (*dimension*) dari besaran tersebut adalah bulan (*period*), produk, tipe pelanggan, jenis industri, dan tujuan pengiriman (*location*).

Selain harus memenuhi spesifikasi fungsional dari pemakai, data model untuk data warehouse yang kita buat (rancang) nantinya harus menampung data dari sumber, yang biasanya adalah data didalam sistem operasional, misalnya dalam contoh kita adalah sistem pesanan (sales order system).

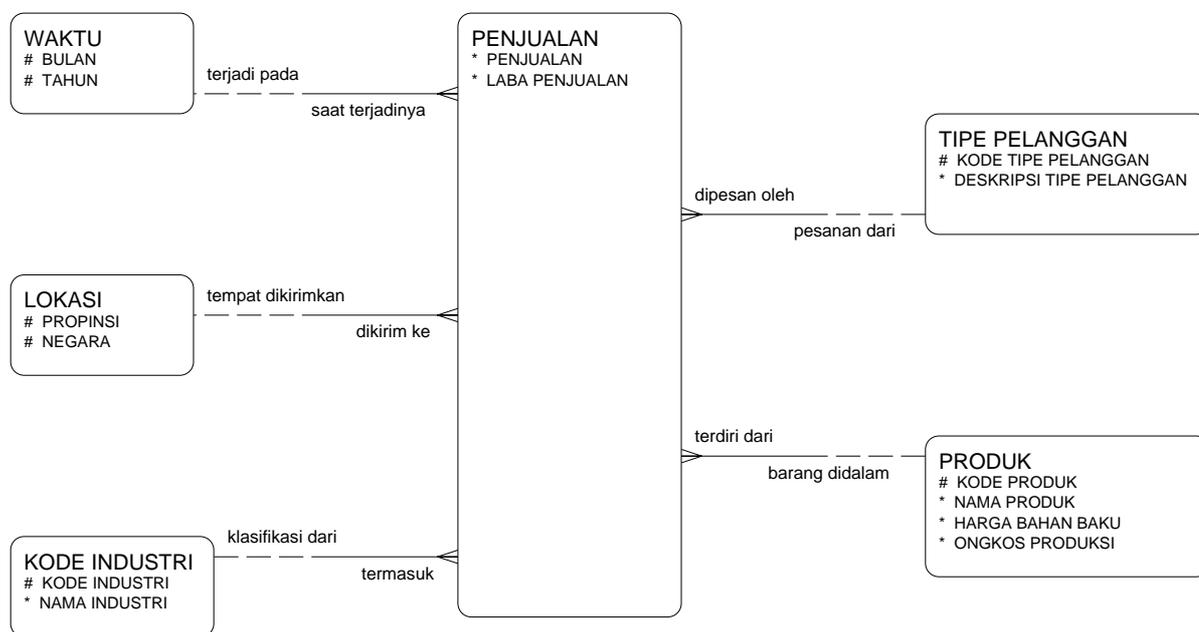
Data Model

Data model (*logical*) dari sumber data ini tergambar sebagai berikut. Diagram data model ini dibuat dengan menggunakan Oracle Designer (salah satu modeling software yang populer) Catatan:

- Attribute yang bertanda # didepan namanya adalah Primary key.
- Foreign key tidak ditunjukkan dalam diagram.



Contoh solusi data model (*logical*) untuk data warehouse kita dengan menggunakan skema bintang (*star schema*), ditunjukkan pada gambar berikut.



Dalam skema bintang ini, tabel Penjualan adalah fakta (*fact table*) yang lain adalah dimensi (*dimension table*) Tabel fakta berisi besaran; tabel dimensi berisi dimensi yang diperlukan untuk menjawab spesifikasi fungsional dari pemakai (*analisa multi-dimensi*)

Pemikiran (pada waktu menganalisa dan merancang) dalam pembuatan data model ini adalah:

- Hanya mencakup data yang dibutuhkan, yang lain tidak akan dimasukkan kedalam data warehouse kita (misalnya pajak penjualan, ongkos kirim, cara kirim, alamat pelanggan) untuk menghemat tempat simpan (disk storage space) dan mempercepat query response. Data-data yang tidak dimasukkan ini, bisa ikut diextract dari sumbernya, disimpan ditempat terpisah (seperti misalnya offline storage) untuk memudahkan penambahannya pada saatnya harus ditambahkan kedalam data warehouse kita.
- Tiga dimensi dibuat umum (common dimension), dengan demikian dimensi-dimensi ini dapat dimanfaatkan untuk analisa lain (mendatang)
 1. Waktu - untuk memenuhi spesifikasi “per bulan”, yang akan dihubungkan dengan data “tanggal pesan”. Table ini akan diisi semua bulan dan sekian tahun (harus diputuskan berapa tahun, karena tabel ini kecil volumenya, misalnya diisi untuk 50 tahun pada awal implementasi data warehouse)
 2. Jenis industri – untuk memenuhi spesifikasi “jenis industri”, yang akan dihubungkan dengan data “industri pelanggan”; diisi lengkap (lihat butir berikut tentang NAICS)
 3. Lokasi – untuk memenuhi spesifikasi “propinsi/negara pemakai”, yang akan dihubungkan dengan bagian data “alamat kirim”, dengan kata lain didalam proses memasukkan data dari sumber kedalam warehouse harus mencari dan menentukan propinsi atau negara didalam/dari data alamat kirim.
- Ada data yang ditambahkan, tidak berasal dari sumber data kita, yaitu “kode industri” dan “nama industri”. Data ini berasal dari NAICS (North America Industry Code Standards), misalnya untuk industri manufaktur ditunjukkan dalam tabel dibawah ini (kolom pertama adalah kode industri bersangkutan). Ini merupakan contoh penggunaan “external data” yang kadang muncul dalam membangun data warehouse. Dalam contoh kita ini, penggunaan standard ini merupakan keinginan pimpinan perusahaan (agar sejalan dengan strategi perusahaan untuk “go international”)

Contoh data NAICS untuk industri manufaktur:

Manufacturing

311	Food Manufacturing
312	Beverage and Tobacco Product Manufacturing
313	Textile Mills
314	Textile Product Mills
315	Apparel Manufacturing
316	Leather and Allied Product Manufacturing
321	Wood Product Manufacturing
322	Paper Manufacturing
323	Printing and Related Support Activities
...	...

- Penjualan diakumulasikan per bulan (monthly aggregated), juga per dimensi yang lain sesuai spesifikasi: produk, tipe pelanggan, jenis industry, dan lokasi kirim. Pajak penjualan dan Ongkos kirim tidak diperhitungkan.
- Data hasil perhitungan “laba” ikut disimpan (pre calculated), agar query cepat. Formula yang diberikan oleh user adalah: Laba penjualan = (Harga satuan – (Harga bahan baku + Ongkos produksi)) x Jumlah pesanan

Isi data

Agar lebih memahami data model diatas, berikut diberikan contoh isi tabel-tabelnya.

Tabel-tabel sumber data

PELANGGAN

NAMA PELANGGAN	ALAMAT PELANGGAN	INDUSTRI PELANGGAN	TIPE PELANGGAN
P.T. Tekstil Indonesia Raya	Jln. Bromo 100 Surabaya Jatim Indonesia	Pabrik tekstil	Pabrik
Japanese International Pulps Trading Company	Tokyo Japan	Pabrik bahan kertas	Importir
P.T. Tekstil Kalimantan	Jln. Kayu Kalimantan no. 007 Jakarta Indonesia	Pabrik tekstil	Pabrik

PESANAN

NOMOR PESANAN	NAMA PELANGGAN	TANGGAL PESAN	TOTAL PEMBELIAN	PAJAK PENJUALAN	TOTAL TAGIHAN	ALAMAT TAGIH
2002-0001	P.T. Tekstil Indonesia Raya	01-JAN-2002	2900000000	2000000	2902000000	Jln. Bromo 100 Surabaya Jatim
2002-0002	Japanese International Pulps Trading Company	15-JAN-2002	9300000000	9000000	9309000000	Kantor pusat di Jakarta
2002-0003	P.T. Tekstil Kalimantan	20-JAN-2002	1000000000	2000000	1002000000	Kantor Pimpinan wilayah Kalimantan
2003-0001	P.T. Tekstil Indonesia Raya	01-APR-2003	15000000000	1000000	15001000000	Jln. Bromo 100 Surabaya Jatim

PRODUK

KODE PRODUK	NAMA PRODUK	HARGA BAHAN BAKU	ONGKOS PRODUKSI
WDM	Pewarna Dasar Merah	1000	100
WTP	Pewarna Tambahan Putih	500	100
PE1	Pengencer Tipe 1	500	10
PE2	Pengencer Tipe 2	1000	10

BARANG PESANAN

NOMOR PESANAN	KODE PRODUK	JUMLAH PESANAN	HARGA SATUAN
2002-0001	WDM	100000	5000
2002-0001	WTP	200000	3000
2002-0001	PE1	300000	2000
2002-0001	PE2	300000	4000
2002-0002	WDM	1000000	5100
2002-0002	PE2	1000000	4200
2002-0003	WDM	200000	5000
2003-0001	WTP	5000000	3000

PENGIRIMAN

NOMOR PESANAN	KODE PRODUK	ALAMAT KIRIM	CARA KIRIM	ONGKOS KIRIM
2002-0001	WDM	Balikpapan Kalimantan	Kapal pelanggan	
2002-0001	WTP	Balikpapan Kalimantan	Kapal pelanggan	
2002-0001	PE1	Balikpapan Kalimantan	Kapal pelanggan	
2002-0001	PE2	Balikpapan Kalimantan	Kapal pelanggan	
2002-0002	WDM	Tokyo Japan	Ocean liner	10000
2002-0002	PE2	Tokyo Japan	Ocean liner	10000
2002-0003	WDM	Balikpapan Kalimantan	Kapal laut	5000
2003-0001	WTP	Balikpapan Kalimantan	Kapal pelanggan	

Tabel-Tabel Data Warehouse

Berikut contoh isi tabel-tabel data warehouse dari model diatas. Semua tabel dimensi diisi data pada awal data warehouse di implementasikan (pre-loaded). Sedang untuk tabel fakta (Penjualan) datanya berasal dari sumber.

WAKTU

BULAN	TAHUN
...	...
JAN	2002
FEB	2002
MAR	2002
APR	2002
MEI	2002
JUN	2002
JUL	2002
AUG	2002
SEP	2002
OKT	2002
NOP	2002
DES	2002
JAN	2003
FEB	2003
MAR	2003
APR	2003
MEI	2003
...	...

LOKASI

PROPINSI	NEGARA
...	...
JATENG	INDONESIA
KALTENG	INDONESIA
	JAPAN

KODE INDUSTRI

KODE INDUSTRI	NAMA INDUSTRI
...	...
313	Textile Mills
314	Textile Product Mills
322	Paper Manufacturing
...	...

PRODUK

KODE PRODUK	NAMA PRODUK
WDM	Pewarna Dasar Merah
WTP	Pewarna Tambahan Putih
PE1	Pengencer Tipe 1
PE2	Pengencer Tipe 2

TIPE PELANGGAN

KODE TIPE PELANGGAN	DESKRIPSI TIPE PELANGGAN
...	...
GRSR	Grosir
PBRK	Pabrik
IMPR	Importir
...	...

Perhatikan aggregation dibulan Januari 2002 untuk produk WDM yang mencakup dua pesanan (no. pesananan 2002-0001 dan 2002-0003) dari dua pelanggan. Ini dapat diagregasikan karena semua dimensi dari keduanya sama percis (hilite biru). Sedang untuk pesanan 2002-0002 tidak dapat diagregasikan (dijumlahkan) karena dimensi Kode industri, Tipe pelanggan, Propinsi dan Negara, berbeda dengan yang dua itu, meskipun Bulan dan Tahunnya sama (hilite kuning)

PENJUALAN

BULAN	TAHUN	KODE INDUSTRI	TIPE PELANGGAN	KODE PRODUK	PROPINSI	NEGARA	PENJUALAN	LABA PEJUALAN
JAN	2002	313	PBRK	WDM	KALTENG	INDONESIA	1500000000	1170000000
JAN	2002	313	PBRK	WTP	KALTENG	INDONESIA	600000000	480000000
JAN	2002	313	PBRK	PE1	KALTENG	INDONESIA	600000000	447000000
JAN	2002	313	PBRK	PE2	KALTENG	INDONESIA	1200000000	897000000
JAN	2002	322	IMPR	WDM		JAPAN	1000000000	4000000000
JAN	2002	322	IMPR	PE2		JAPAN	4200000000	3190000000
APR	2003	313	PBRK	WTP	KALTENG	INDONESIA	1500000000	1245000000