

Bab 9

Data Warehousing dan Decision Support

POKOK BAHASAN:

- ✓ Hubungan antara Data Warehouse dan Decision Support
- ✓ Model Data Multidimensi
- ✓ *Online Analytical Processing (OLAP)*
- ✓ Arsitektur Data Warehouse
- ✓ Implementasi Data Warehouse
- ✓ View dan Decision Support

TUJUAN BELAJAR:

Setelah mempelajari materi dalam bab ini, mahasiswa diharapkan mampu:

- ✓ Memahami hubungan antara Data Warehouse dan Decision Support
- ✓ Dapat mendesain model data multidimensi
- ✓ Memahami *Online Analytical Processing (OLAP)*
- ✓ Memahami Arsitektur dan Implementasi Data Warehouse
- ✓ Memahami penggunaan view pada aplikasi Decision Support

9.1. PENDAHULUAN

Data Warehouse adalah suatu database penunjang keputusan yang dikelola secara terpisah dari database operasional perusahaan. Dan merupakan penunjang pemrosesan informasi dengan menyediakan suatu platform yang kokoh untuk analisa data yang mengandung histori dan yang terkonsolidasi.

Berikut beberapa definisi tentang data warehouse :

“Suatu DW adalah suatu koleksi data yang bisa digunakan untuk menunjang pengambilan keputusan manajemen, yang berorientasi subjek (topik), terpadu, time variant, dan tidak mudah berubah” —W. H. Inmon (Bp. Data Warehousing)

“Suatu data warehouse sederhananya adalah suatu penyimpanan data tunggal, lengkap dan konsisten, yang diperoleh dari berbagai sumber dan dibuat tersedia bagi end user dalam suatu cara yang bisa mereka pahami dan bisa mereka gunakan dalam suatu konteks bisnis.” - - Barry Devlin, *IBM Consultant*.

9.2. DARI DATA WAREHOUSE KE DECISION SUPPORT

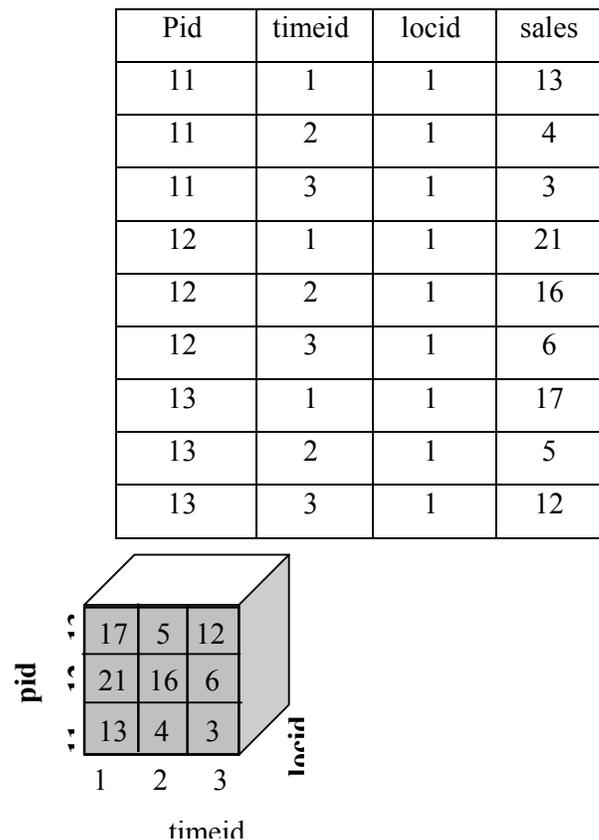
Pembuatan keputusan organisasi memerlukan view menyeluruh pada segala aspek perusahaan, sehingga organisasi membuat data warehouse gabungan yang berisi data yang berasal dari berbagai sumber. DBMS juga didesain untuk mendukung query OLAP secara efisien dan dioptimalisasi untuk mendukung aplikasi decision support.

9.3. MODEL DATA MULTIDIMENSI

Model data multidimensional dirancang untuk memfasilitasi analisis dan bukan transaksi. Model ini umum digunakan dalam data warehouse. Memiliki konsep intuitif dari banyak dimensi atau perspektif pengukuran bisnis atau fakta-fakta. Contohnya : untuk melihat penjualan dari perspektif customer, product dan time.

Model data multi dimensi adalah himpunan pengukuran numerik yang tergantung pada himpunan dimensi. Misalnya untuk mengetahui Penjualan/Sales, dimensinya adalah Produk (pid), Lokasi (locid), dan Waktu (timeid).

Data multi dimensi dapat disimpan secara fisik dalam sebuah array yang disebut sistem MOLAP. Alternatif lainnya, data dapat disimpan sebagai relasi yang disebut sistem ROLAP. Relasi utama yang berhubungan dengan dimensi yang diukur dinamakan tabel fakta (fact table). Tiap dimensi dapat diberi tambahan atribut dan berasosiasi dengan suatu tabel dimensi (dimension table). Tabel fakta mempunyai ukuran yang lebih besar dibandingkan dengan tabel dimensi.



Gambar 10.1. Model Data Multidimensi

9.4. **ONLINE ANALYTICAL PROCESSING (OLAP)**

Query OLAP dipengaruhi oleh dua hal, yaitu : SQL dan spreadsheet. Operasi yang umum di dalam query OLAP adalah melakukan agregasi pada satu atau lebih dimensi. Misalnya, cari total penjualan (sales), cari total penjualan tiap propinsi, cari 5 ranking produk teratas berdasarkan total penjualan. Jenis-jenis query OLAP adalah :

1. Roll up

Yaitu dengan melakukan agregasi pada level yang berbeda dari hirarki dimensi. Misalnya untuk setiap kota diberikan total penjualan, maka untuk total penjualan tiap propinsi bisa didapatkan dengan menambahkan total penjualan pada semua kota dalam satu propinsi.

2. Drill down

Adalah kebalikan dari roll up. Misalnya untuk setiap propinsi dapat diberikan total penjualan, maka total penjualan tiap kota dapat di-drill down.

3. Pivoting

Yaitu melakukan agregasi pada dimensi terpilih. Misalnya jika dilakukan pivoting pada Location dan Time didapatkan cross-tabulation sebagai berikut :

	LA	NY	tota
1998	25	41	66
1999	64	53	11
tota	89	94	18

Cross-tabulation tersebut merupakan kumpulan dari query SQL berikut :

```
SELECT SUM (S.Sales)
FROM Sales S, Times T
WHERE S.timeid = T.timeid
GROUP BY T.year
```

dan

```
SELECT SUM (S.Sales)
FROM Sales S, Location L
WHERE S.timeid = L.timeid
GROUP BY L.state
```

Sehingga menjadi query baru sebagai berikut :

```
SELECT SUM (S.Sales)
FROM Sales S, Times T, Location L
WHERE S.timeid = T.timeid AND S.timeid = L.timeid
GROUP BY T.year, L.state
```

4. Slicing dan Dicing

Yaitu mencari kesamaan dan jangkauan seleksi pada satu atau lebih dimensi.

9.5. ARSITEKTUR DATA WAREHOUSE

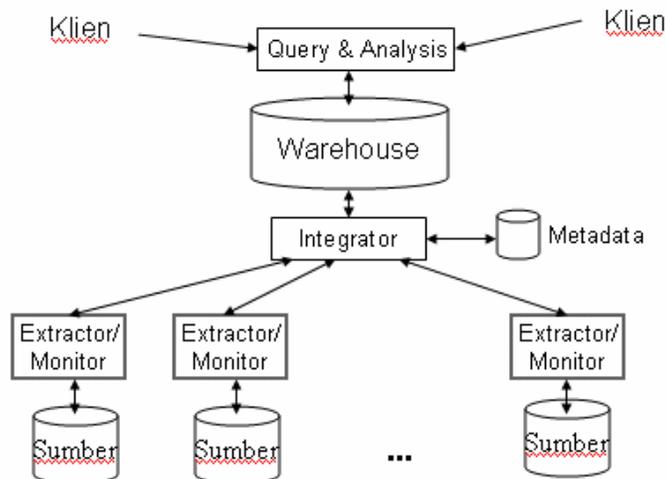
Pilihan berikut harus dibuat didalam perancangan data warehouse

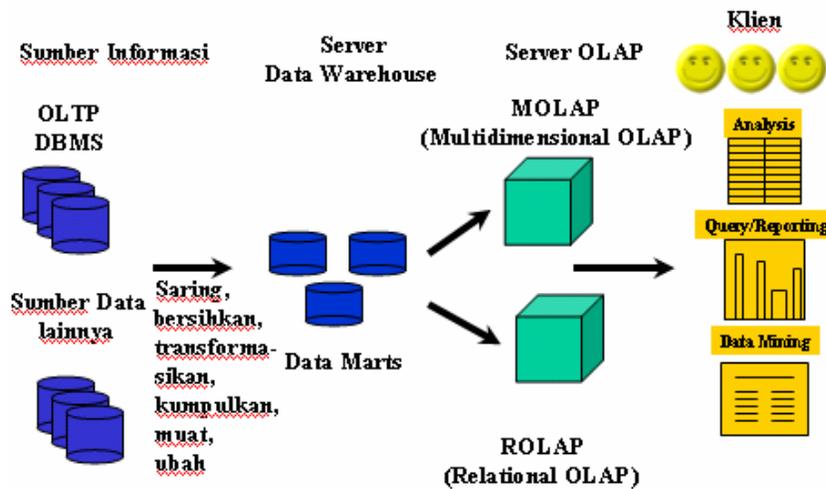
- process model
Tipe apa yang akan dimodelkan?
- grain
Apa dasar data dan level atom data yang akan disajikan?
- dimensi
Dimensi apa yang dipakai untuk masing-masing record tabel fakta?
- ukuran
Ukuran apa yang akan mengumpulkan masing-masing record tabel fakta?

Suatu data warehouse didasarkan kepada suatu model data multidimensi yang melihat data dalam bentuk suatu kubus data. Suatu kubus data seperti Sales memungkinkan data untuk dimodelkan dan dilihat dari banyak dimensi

- Dimensi tabel, seperti item (item_name, brand, type), atau time(day, week, month, quarter, year)
- Tabel fakta memuat ukuran (seperti dollars_sold) dan kunci untuk setiap dimensi tabel terkait

Berikut, arsitektur dari Data Warehouse :





Gambar 10.2. Arsitektur Data Warehouse

9.6. IMPLEMENTASI DATA WAREHOUSE

Data Warehouse diorganisasikan berdasarkan kegunaan disekitar subjek bukan aplikasi, misal: customer, product, sales. Perhatian dipusatkan pada pemodelan dan analisa data untuk pembuat keputusan, bukan untuk operasi harian atau pemrosesan transaksi. View sederhana dan ringkas disekitar subjek pembicaraan disediakan dengan cara memisahkan data-data yang tidak berkaitan dengan proses penunjang keputusan.

Data Warehouse dibangun dengan memadukan banyak sumber data yang heterogen, misal : Database relasional, flat file, catatan transaksi on-line. Teknik pembersihan dan integrasi data juga diterapkan dalam datawarehouse.

Dikarenakan data berasal dari sumber yang berbeda-beda, maka harus dapat dijamin konsistensi penamaan, penyandian struktur, ukuran atribut, dsb., dari antara sumber-sumber data yang berbeda, misal, tarif hotel: mata uang, pajak, breakfast covered, dsb. Ketika data dipindahkan ke warehouse, data ini telah terkonversi

9.7. VIEW DAN DECISION SUPPORT

View digunakan secara luas dalam aplikasi decision support. Kelompok analis yang berbeda dalam suatu organisasi biasanya berhubungan dengan aspek bisnis yang berbeda, dan akan lebih memudahkan untuk mendefinisikan view yang memberi tiap grup wawasan detail bisnis yang sesuai. Setelah view didefinisikan, barulah dapat ditulis query atau definisi view baru yang menggunakannya. Mengevaluasi query yang diajukan untuk view sangat penting untuk aplikasi decision support. Berikut ini akan dibahas bagaimana query dapat dievaluasi secara efisien setelah menempatkan view ke dalam konteks aplikasi decision support.

View berhubungan erat dengan OLAP dan data warehousing. Query OLAP biasanya merupakan query agregasi. Data warehouse adalah kumpulan table yang direplikasi secara asynchronous dan view yang disinkronisasi secara periodik.

View berikut ini menghitung penjualan produk berdasarkan kategori dan Negara bagian:

```
CREATE VIEW RegionalSales(category, sales, state)
AS SELECT P.category, S.sales, L.state
FROM Products P, Sales S, Locations L
WHERE P.pid=S.pid AND S.locid=L.locid
```

Query berikut menghitung penjualan total untuk tiap kategori berdasarkan Negara bagian :

```
SELECT R.category, R.state, SUM(R.sales)
FROM RegionalSales R
GROUP BY R.category, R.state
```

Query berikut ini adalah hasil modifikasi dari query sebelumnya dengan menggantikan *RegionalSales* dengan sebuah view yang ditempatkan pada klausa FROM yaitu :

```
SELECT R.category, R.state, SUM(R.sales)
FROM (SELECT P.category, S.sales, L.state
      FROM Products P, Sales S, Locations L
      WHERE P.pid = S.pid AND S.locid=L.locid) AS R
GROUP BY R.category, R.state
```

RINGKASAN:

- Data Warehouse adalah suatu database penunjang keputusan yang dikelola secara terpisah dari database operasional perusahaan, yang menyediakan suatu platform untuk analisa data yang mengandung histori dan yang terkonsolidasi.
- Aplikasi Decision Support memerlukan view yang menyeluruh pada segala aspek perusahaan, dan didapatkan dari data yang berasal dari berbagai sumber.
- Model data multidimensional dirancang untuk melakukan analisa data dan umum digunakan dalam data warehouse.
- Model data multidimensi memiliki konsep intuitif dari banyak dimensi atau perspektif pengukuran bisnis atau fakta-fakta. Contohnya : untuk melihat penjualan dari perspektif customer, product dan time.
- Jenis-jenis dari query OLAP yaitu : Roll up, Drill down, Pivoting, Slicing dan Dicing.
- Pilihan-pilihan yang harus dibuat didalam perancangan data warehouse : process model, level atom data yang akan disajikan, dimensi dan ukuran
- View digunakan secara luas dalam aplikasi decision support dan berhubungan erat dengan OLAP dan data warehousing dan perlu disinkronisasi secara periodik.

LATIHAN SOAL :

1. Apa yang dimaksud dengan aplikasi decision support ?
2. Apa pertimbangan yang digunakan dalam mendesain Data Warehouse ?
3. Deskripsikan dimensi dan pengukuran dalam model data multidimensi, dan jelaskan perbedaan antara table fakta dan table dimensi.
4. Apa yang dimaksud dengan table fakta, dan mengapa sangat penting dari sudut pandang performa ?
5. Perhatikan contoh relasi sales pada gambar 10.1, tunjukkan hasil pivoting relasi pada *pid* dan *timeid*.
6. Pada gambar 10.1, tunjukkan hasil rool-up pada *locid* (misalnya, negara bagian).

-
7. Tuliskan query dalam SQL, sesuai dengan data yang ada pada gambar 10.1
 - a. Carilah perubahan persentase dalam penjualan bulanan total untuk tiap lokasi
 - b. Carilah perubahan persentase dalam penjualan kuartalan total untuk tiap produk
 - c. Carilah tiga besar lokasi yang diurutkan berdasarkan penjualan total.
 8. Mengapa view begitu penting dalam lingkungan decision support ? Bagaimana view dihubungkan dalam lingkungan data warehousing dan OLAP ?