

Data Warehouse

Rengga Asmara, S.Kom. OCA
2010

Pokok Bahasan

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

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

Pendahuluan (1)

- 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.

Pendahuluan (2)

- 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*.

Hubungan Data Warehouse dan 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.

Model Data Multidimensi

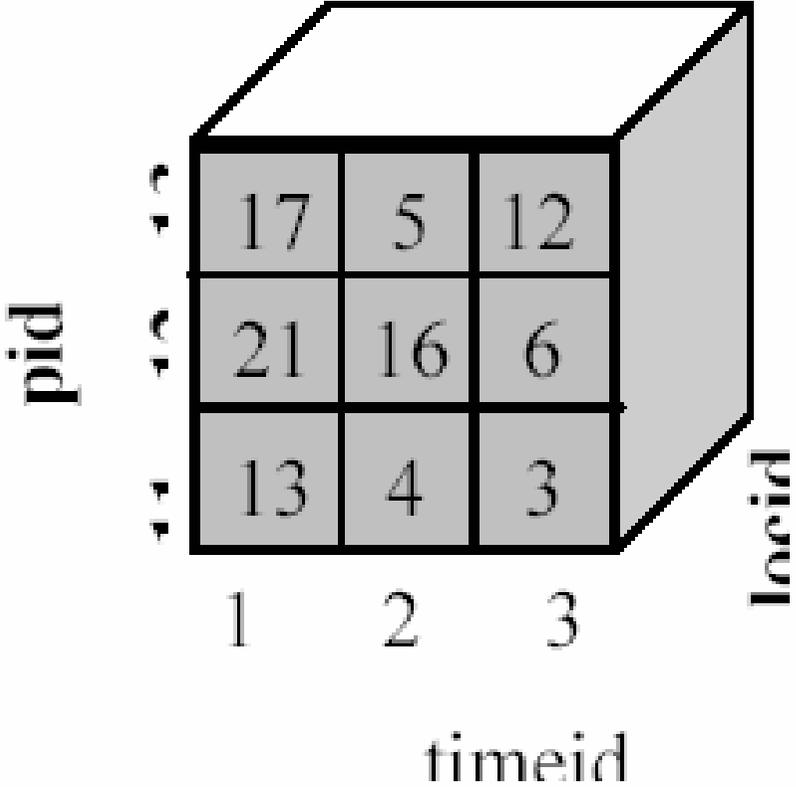
- Model data multidimensional dirancang untuk memfasilitasi analisis dan bukan transaksi.
- Model ini umum digunakan dalam data warehouse.
- Memiliki konsep intuitif dari banyak dimensi atau perspektif pengukuran bisnis atau fakta-fakta.
- Contohnya :
 - untuk melihat penjualan dari perspektif customer,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

Model Data Multidimensi – cont'd

- 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.

Model Data Multidimensi – cont'd

Pid	timeid	locid	sales
11	1	1	13
11	2	1	4
11	3	1	3
12	1	1	21
12	2	1	16
12	3	1	6
13	1	1	17
13	2	1	5
13	3	1	12



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.

OLAP – cont'd

- Jenis-jenis query OLAP adalah :
 - 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.

OLAP – cont'd

– Drill down

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

– Pivoting

- Yaitu melakukan agregasi pada dimensi terpilih.

OLAP – cont'd

- 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

OLAP – cont'd

- 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
```

OLAP – cont'd

– Slicing dan Dicing

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

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?

Arsitektur Data Warehouse – cont'd

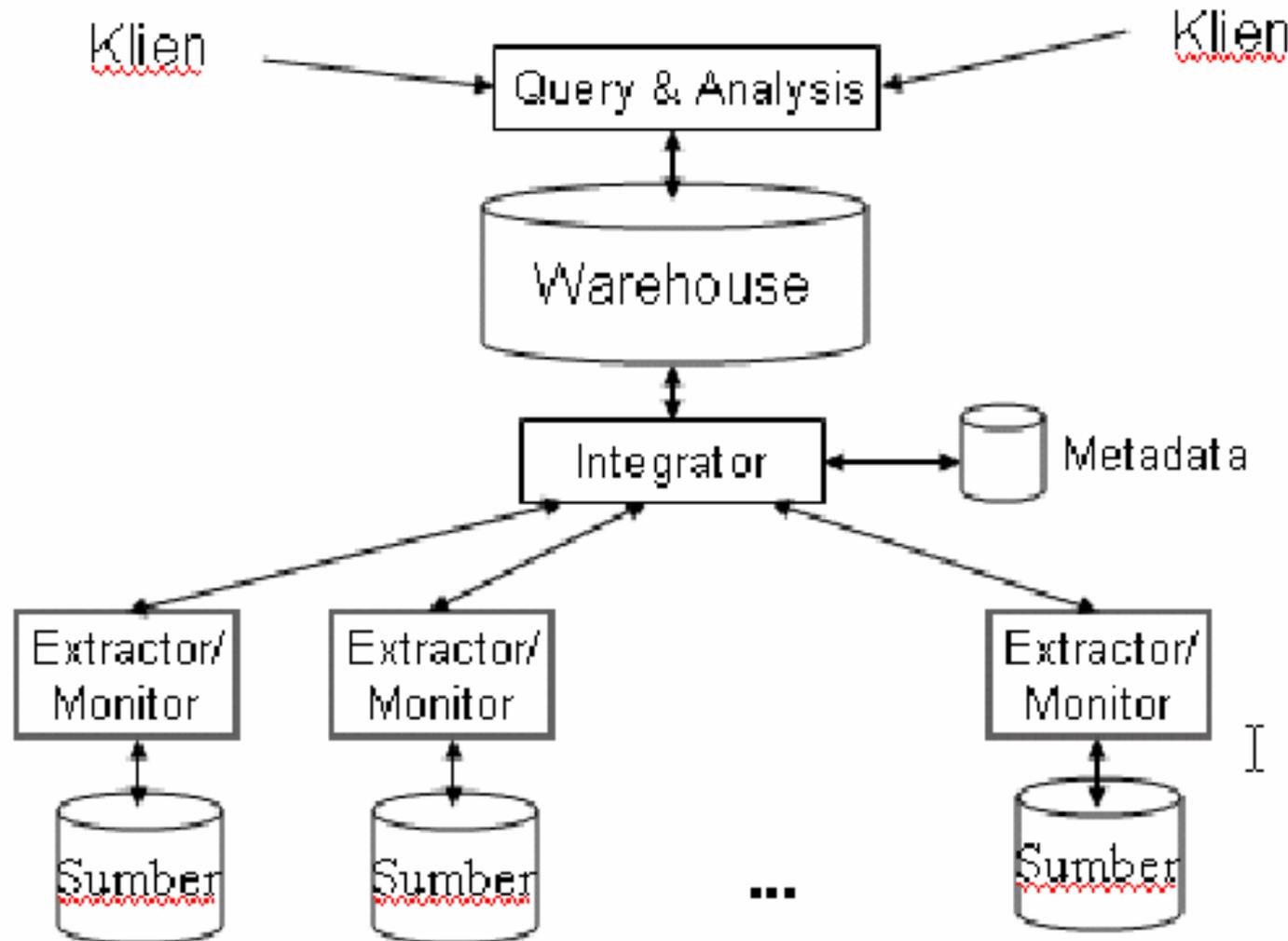
- Suatu data warehouse didasarkan kepada suatu model data multidimensi yang melihat data dalam bentuk suatu kubus data.
- Suatu kubus data seperti Sales memungkinkan data untuk dimodelkan dan dilihat dari banyak dimensi.

Arsitektur Data Warehouse – cont'd

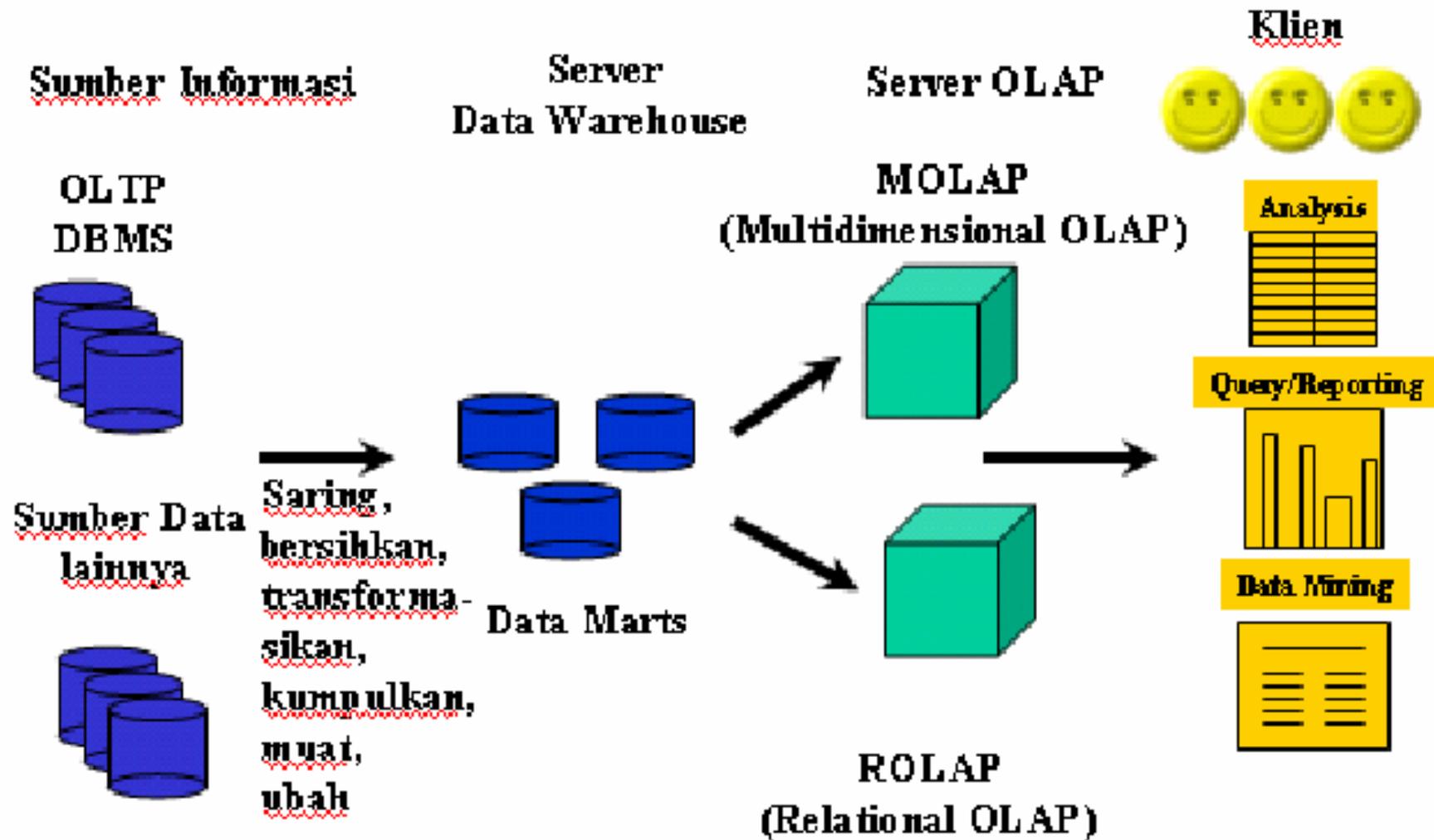
- Dimensi tabel, seperti item (item_name, brand, type), atau time(day, week, month, quarter, year).
- Tabel fakta memuat ukuran (seperti dollars_sold) dan kunci untuk setiap dimensi tabel terkait.

Arsitektur Data Warehouse – cont'd

- Berikut, arsitektur dari Data Warehouse :



Arsitektur Data Warehouse – cont'd



Gambar 10.2. Arsitektur Data Warehouse

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.

Implementasi Data Warehouse – cont'd

- 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 data warehouse.

Implementasi Data Warehouse – cont'd

- 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 harus terkonversi.

Question?

End of Session