

Data Warehouse, Data Mart, OLAP, dan Data Mining

Data Warehouse

- *Data warehouse* adalah basis data yang menyimpan data sekarang dan data masa lalu yang berasal dari berbagai sistem operasional dan sumber yang lain (sumber eksternal) yang menjadi perhatian penting bagi manajemen dalam organisasi dan ditujukan untuk keperluan analisis dan pelaporan manajemen dalam rangka pengambilan keputusan
- *Data warehouse* digunakan untuk mendukung pengambilan keputusan, bukan untuk melaksanakan pemrosesan transaksi
- *Data warehouse* hanya berisi informasi-informasi yang relevan bagi kebutuhan pemakai yang dipakai untuk pengambilan keputusan

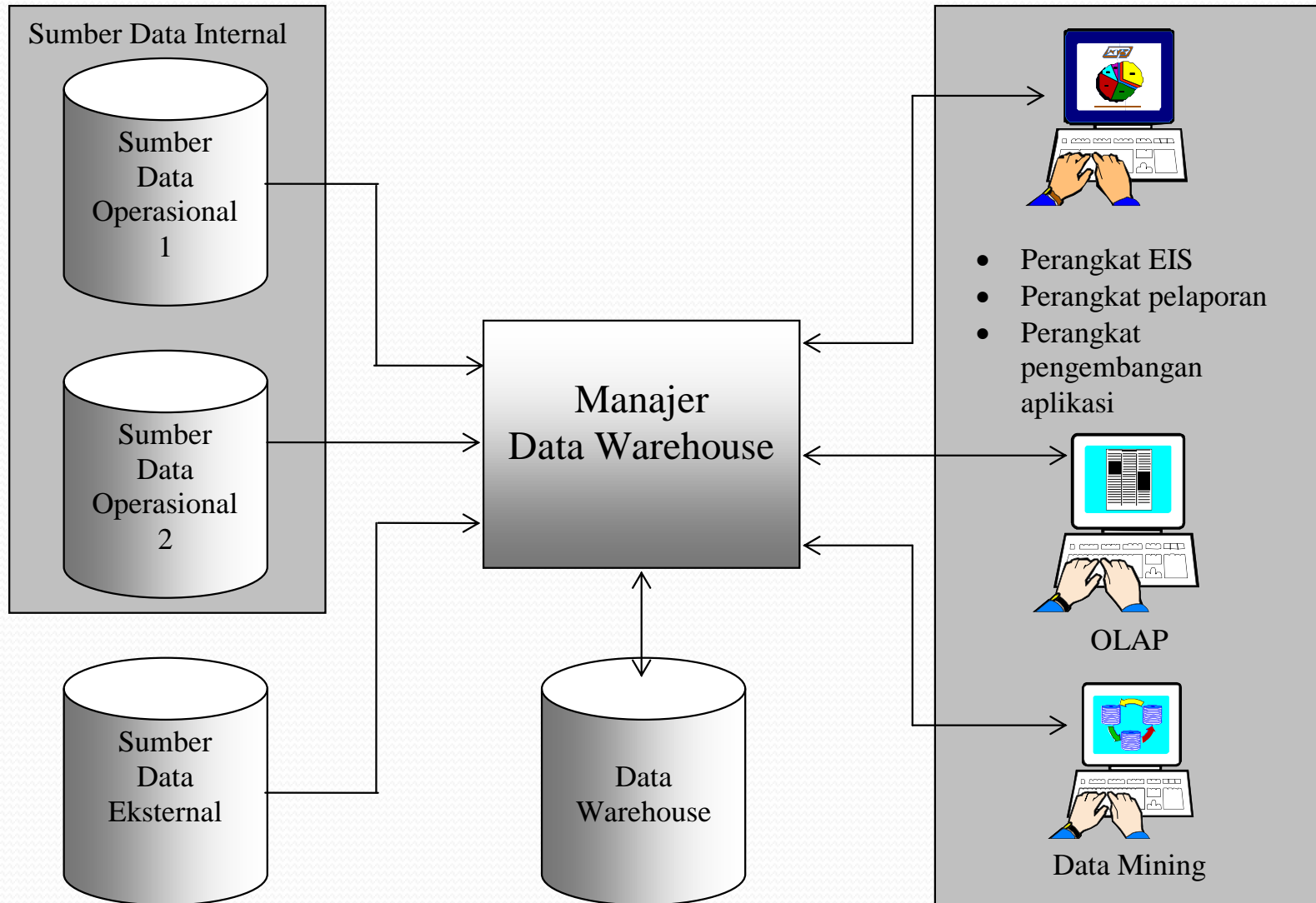
Perbedaan DW dan OLTP

OLTP	Data Warehouse
<ul style="list-style-type: none">• Menangani data saat ini• Data bisa saja disimpan pada beberapa platform• Data diorganisasikan berdasarkan fungsi atau operasi seperti penjualan, produksi, dan pemrosesan pesanan• Pemrosesan bersifat berulang• Untuk mendukung keputusan harian (operasional)• Melayani banyak pemakai operasional• Berorientasi pada transaksi	<ul style="list-style-type: none">• Lebih cenderung menangani data masa lalu• Data disimpan dalam satu platform• Data diorganisasikan menurut subjek seperti pelanggan atau produk• Pemrosesan sewaktu-waktu, tak terstruktur, dan bersifat heuristik• Untuk mendukung keputusan yang strategis• Untuk mendukung pemakai manajerial yang berjumlah relatif sedikit• Berorientasi pada analisis

Sumber Data untuk DW

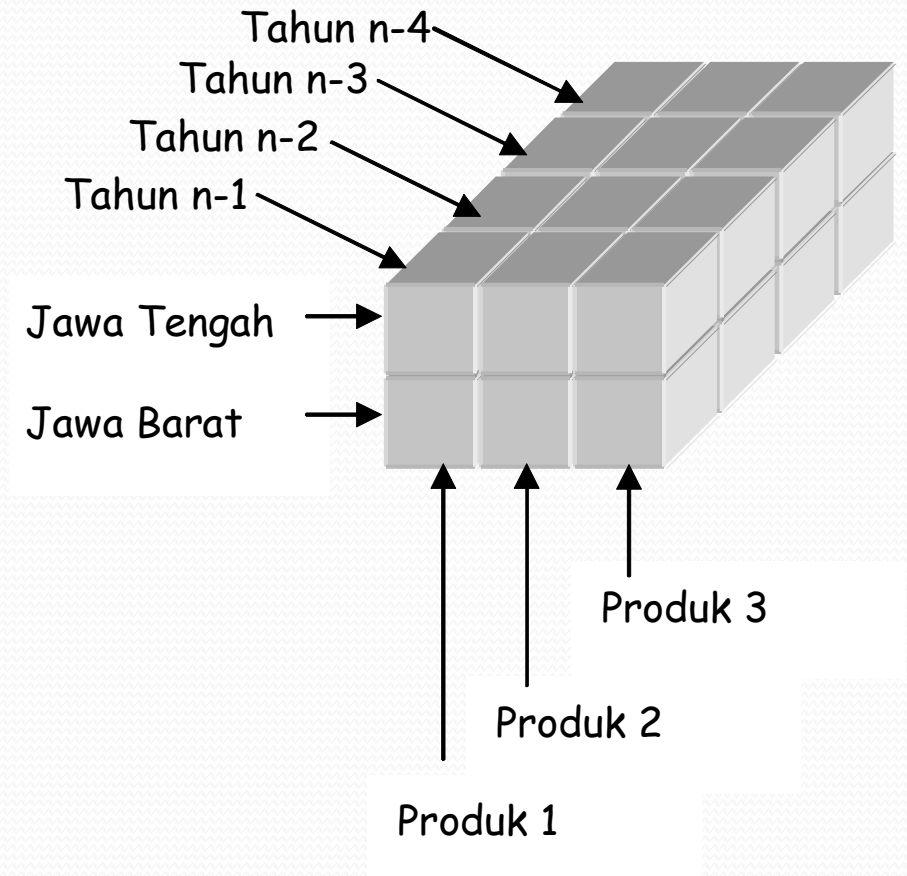
1. Data operasional dalam organisasi, misalnya basis data pelanggan dan produk, dan
 2. Sumber eksternal yang diperoleh misalnya melalui Internet, basis data komersial, basis data pemasok atau pelanggan
- Berbagai data yang berasal dari sumber digabungkan dan diproses lebih lanjut oleh manajer *data warehouse* dan disimpan dalam basis data tersendiri.
 - Selanjutnya, perangkat lunak seperti OLAP dan *data mining* dapat digunakan oleh pemakai untuk mengakses *data warehouse*

Prinsip Data Warehouse



Sifat Data Warehouse

- Multidimensional yang berarti bahwa terdapat banyak lapisan kolom dan baris (Ini berbeda dengan tabel pada model relasional yang hanya berdimensi dua)
- Berdasarkan susunan data seperti itu, amatlah mudah untuk memperoleh jawaban atas pertanyaan seperti: “Berapakah jumlah produk 1 terjual di Jawa Tengah pada tahun n-3?”



Data Warehouse

- Mengingat sistem *data warehouse* memerlukan pemrosesan data dengan volume yang besar, sistem ini biasa diterapkan dengan menggunakan teknologi pemrosesan SMP dan MPP
- *Data warehouse* dapat dibangun sendiri dengan menggunakan perangkat pengembangan aplikasi ataupun dengan menggunakan perangkat lunak khusus yang ditujukan untuk menangani hal ini
- Beberapa contoh perangkat lunak yang digunakan untuk administrasi dan manajemen *data warehouse*:
 - HP Intelligent Warehouse (Hewlett Packard)
 - FlowMark (IBM)
 - SourcePoint (Software AG)

Petunjuk Membangun DW

- Menentukan misi dan sasaran bisnis bagi pembentukan *data warehouse*
- Mengidentifikasi data dari basis data operasional dan sumber lain yang diperlukan bagi *data warehouse*
- Menentukan item-item data dalam perusahaan dengan melakukan standarisasi penamaan data dan maknanya
- Merancang basis data untuk *data warehouse*
- Membangun kebijakan dalam mengarsipkan data lama sehingga ruang penyimpanan tak menjadi terlalu besar dan agar pengambilan keputusan tidak menjadi terlalu lamban.
- Menarik data produksi (operasional) dan meletakkan ke basis data milik *data warehouse*

Data Mart

- Bagian dari *data warehouse* yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu dalam perusahaan. Karakteristik yang membedakan data mart dan *data warehouse* adalah sebagai berikut (Connolly, Begg, Strachan 1999).
 - Data mart memfokuskan hanya pada kebutuhan-kebutuhan pemakai yang terkait dalam sebuah departemen atau fungsi bisnis.
 - Data mart biasanya tidak mengandung data operasional yang rinci seperti pada *data warehouse*.
 - Data mart hanya mengandung sedikit informasi dibandingkan dengan *data warehouse*. Data mart lebih mudah dipahami dan dinavigasi.

Contoh Software Data Mart

- SmartMart (IBM)
- Visual Warehouse (IBM)
- PowerMart (Informatica)

OLAP

- *OnLine Analytical Processing*
- Suatu jenis pemrosesan yang memanipulasi dan menganalisa data bervolume besar dari berbagai perspektif (multidimensi). OLAP seringkali disebut **analisis data multidimensi**.

OLAP (Lanjutan...)

- Data multidimensi adalah data yang dapat dimodelkan sebagai atribut dimensi dan atribut ukuran
- Contoh atribut dimensi adalah nama barang dan warna barang, sedangkan contoh atribut ukuran adalah jumlah barang

OLAP : Contoh Data 2 Dimensi

Kota →	Kudus	Magelang	Semarang	...
Triwulan ↓				
1	6.000.000	8.500.000	12.500.000	...
2	4.500.000	3.500.000	14.000.000	...
3	7.600.000	5.500.000	13.700.000	...
4	5.400.000	7.200.000	12.800.00	...

Kemampuan OLAP

- **Konsolidasi** melibatkan pengelompokan data. Sebagai contoh kantor-kantor cabang dapat dikelompokkan menurut kota atau bahkan propinsi. Transaksi penjualan dapat ditinjau menurut tahun, triwulan, bulan, dan sebagainya. Kadangkala istilah *rollup* digunakan untuk menyatakan konsolidasi
- **Drill-down** adalah suatu bentuk yang merupakan kebalikan dari konsolidasi, yang memungkinkan data yang ringkas dijabarkan menjadi data yang lebih detail
- **Slicing and dicing** (atau dikenal dengan istilah pivoting) menjabarkan pada kemampuan untuk melihat data dari berbagai sudut pandang

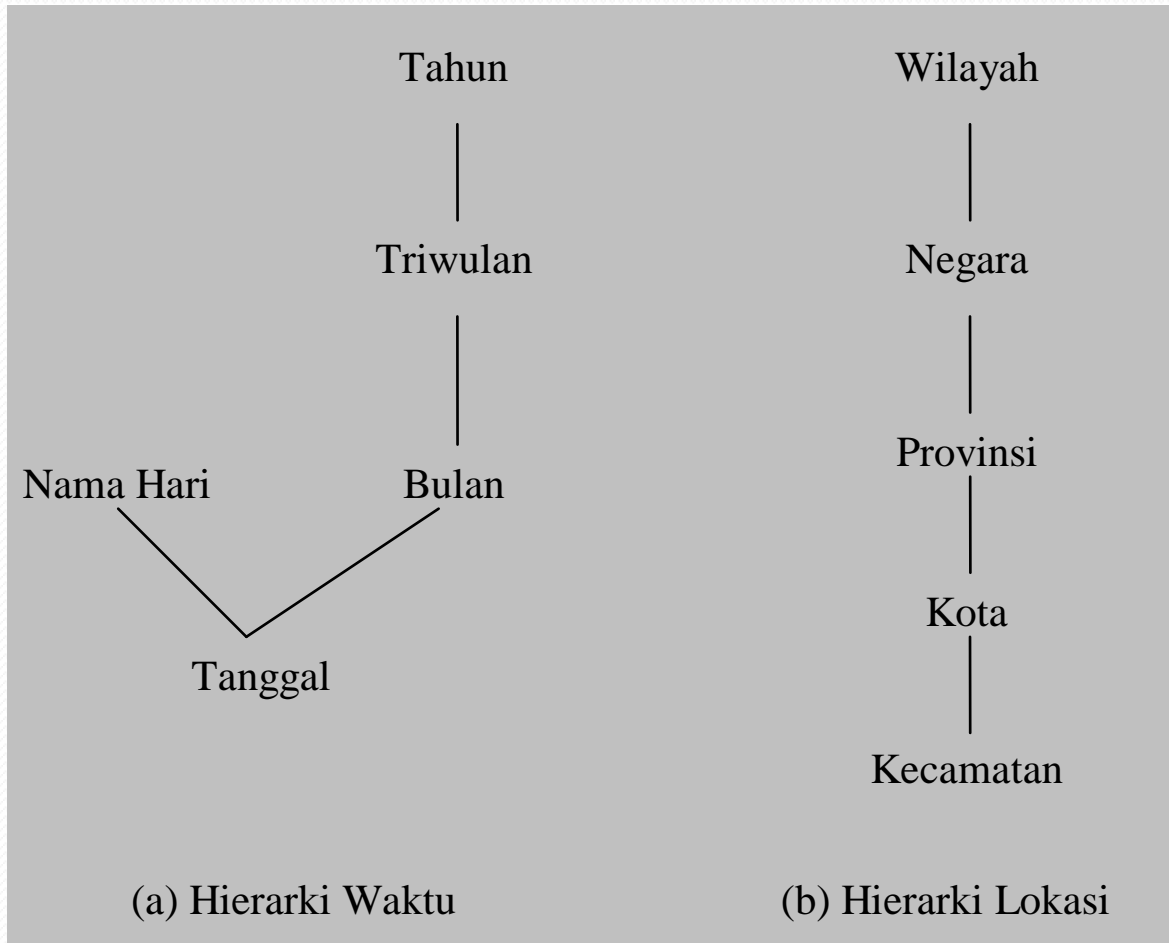
Contoh Tabel Pivoting

Rasa	Strawberry	Mangga	Nanas	Total
Sirup				
Biasa	3.500.000	1.750.000	500.000	5.750.000
Rendah Kalori				
	2.300.000	1.500.000	250.000	4.050.000
Total	5.800.000	3.250.000	750.000	9.800.000



Sirup	Rasa	Pendapatan
Biasa	Strawberry	3.500.000
Biasa	Mangga	1.750.000
Biasa	Nanas	500.000
Rendah Kalori	Strawberry	2.300.000
Rendah Kalori	Mangga	1.500.000
Rendah Kalori	Nanas	250.000

Hierarki Dimensi untuk Drill-down



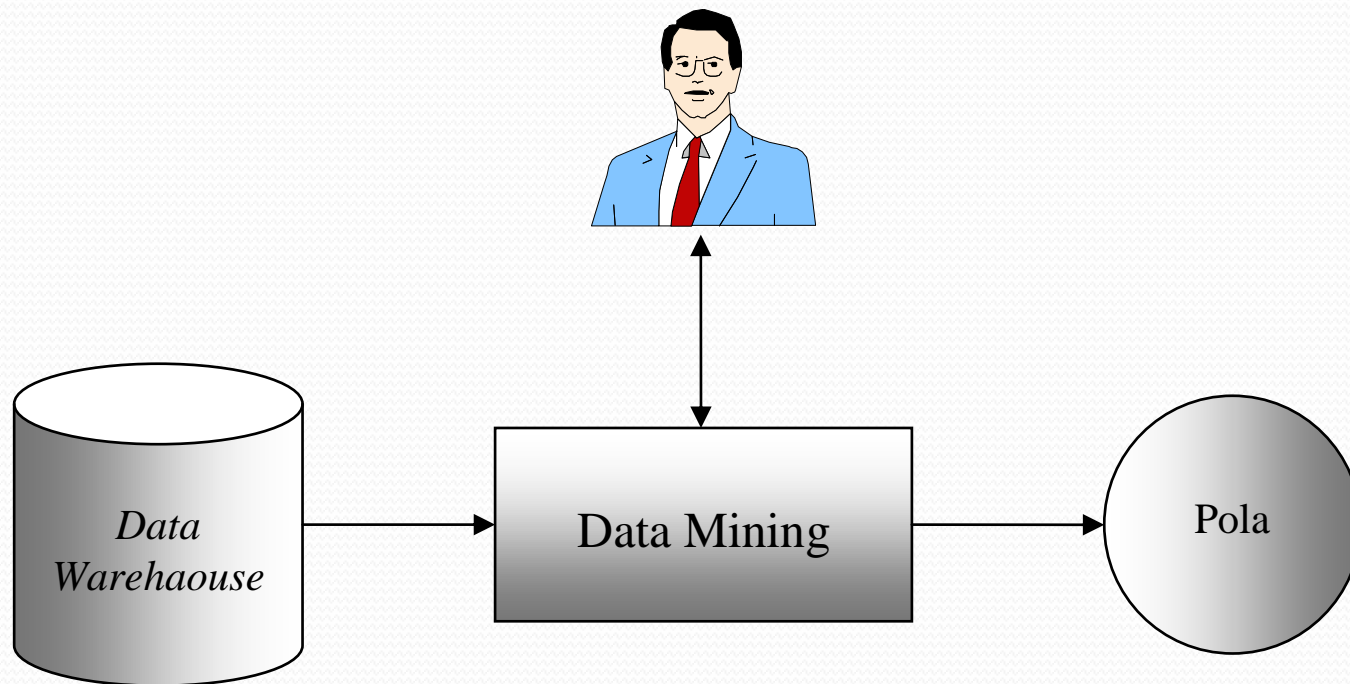
Software OLAP

- Express Server (Oracle)
- PowerPlay (Cognos Software)
- Metacube (Informix/Stanford Technology Group)
- HighGate Project (Sybase)

Data Mining

- Perangkat lunak yang digunakan untuk menemukan pola-pola tersembunyi maupun hubungan-hubungan yang terdapat dalam basis data yang besar dan menghasilkan aturan-aturan yang digunakan untuk memperkirakan perilaku di masa medatang
- *Data mining* sering dikatakan berurusan dengan “penemuan pengetahuan” dalam basis data. Suatu aturan yang dihasilkan oleh *data mining* misalnya seperti berikut : “Kebanyakan pembeli mobil Forsa adalah wanita berusia di atas 30 tahun”.

Prinsip Data Mining



Aplikasi Data Mining

Bidang	Contoh
Pemasaran	<ul style="list-style-type: none">• Mengidentifikasi pembelian yang dilakukan konsumen• Menemukan hubungan di antara karakteristik demografi pelanggan• Memperkirakan tanggapan penawaran melalui surat
Bank	<ul style="list-style-type: none">• Mendeteksi pola penyalahgunaan kartu kredit• Mengidentifikasi tingkat loyalitas pelanggan
Asuransi	<ul style="list-style-type: none">• Analisis klaim• Memperkirakan pelanggan yang akan membeli produk baru

Teknologi Untuk Data Mining

- Statistik
- Jaringan saraf (*neural network*)
- Logika kabur (*fuzzy logic*)
- Algoritma genetika
- dan berbagai teknologi kecerdasan buatan yang lain

Data Mining : Visualisasi Data

- Pendekatan *data mining* juga ada yang melalui visualisasi data
- Pada sistem seperti ini, pemakai akan dibantu untuk menemukan sendiri pola dari sejumlah data berukuran besar dengan didasarkan visualisasi oleh *data mining*

Data Mining : Visualisasi Data

