

Data Warehouse

Arna Fariza

Data Warehouse

- **Rainardi (2008)** : Data warehouse adalah sebuah sistem yang mengambil dan menggabungkan data secara periodik dari sistem sumber data ke penyimpanan data bentuk dimensional atau normal.
- **Han (2006)**: Data warehouse merupakan penyimpanan data yang berorientasi objek, terintegrasi, mempunyai variant waktu, dan menyimpan data dalam bentuk nonvolatile sebagai pendukung manajemen dalam proses pengambilan keputusan.

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

- Data warehouse menyatukan dan menggabungkan data dalam bentuk multidimensi.
- Pembangunan data warehouse meliputi pembersihan data, penyatuan data dan transformasi data dan dapat dilihat sebagai praproses yang penting untuk digunakan dalam data mining.

Data Warehouse

- Meliputi: extraction, transportation, transformation, loading solution, online analytical processing (OLAP), client analysis tool, dan aplikasi lain yang mengatur proses pengumpulan data dan mengirimkan
- Contoh: data penjualan pada suatu perusahaan. Dengan data warehouse maka dapat menjawab pertanyaan “siapakah customer terbesar pada akhir tahun?”

Data Warehouse vs OLTP (Online Transaction Processing)

OLTP		Data Warehouse
Complex data structures (3NF databases)		Multidimensional data structures
Few	Indexes	Many
Many	Joins	Some
Normalized DBMS	Duplicated Data	Denormalized DBMS
Rare	Derived Data and Aggregates	Common

Salah satu perbedaan utama data warehouse dengan OLTP adalah data warehouse tidak selalu dalam bentuk normal ketiga (3NF), sedangkan OLTP biasanya dalam bentuk normal ketiga (3NF)

Data Warehouse vs OLTP

1. Workload

- Data warehouse didesain untuk menampung query dalam jumlah yang besar
- OLTP hanya mendukung operasi tertentu

2. Data modification

- Data warehouse diupdate secara regular (setiap minggu atau setiap hari) menggunakan teknik modifikasi data sehingga user tidak secara langsung mengupdate data warehouse
- Pada OLTP, user melakukan proses update data secara rutin dan langsung

3. Schema Design

- Data warehouse biasanya ternormalisasi secara sebagian bahkan dalam keadaan tidak ternormalisasi
- OLTP ternormalisasi penuh untuk meningkatkan proses update/insert/delete dan meningkatkan konsistensi data

Data Warehouse vs OLTP

4. Typical operation

- Data warehouse menjalankan query yang memproses banyak baris (ratusan atau milyaran), contoh : total penjualan semua customer pada akhir tahun
- OLTP hanya mengkases record tertentu, contoh : mencari data order untuk customer tertentu

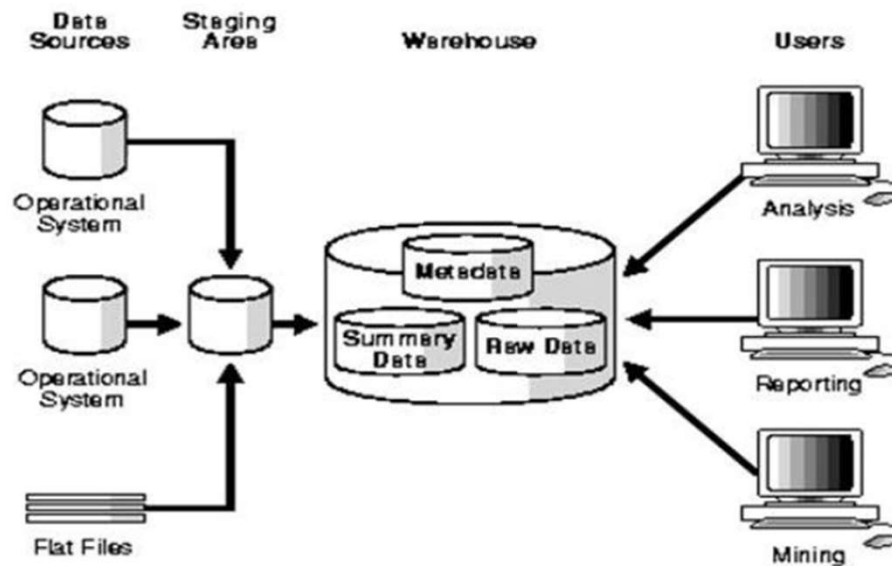
5. Historical data

- Data warehouse menyimpan data selama beberapa bulan atau tahun. Hal ini mendukung proses historical analysis
- OLTP menyimpan data hanya beberapa minggu atau bulan

Arsitektur Data Warehouse

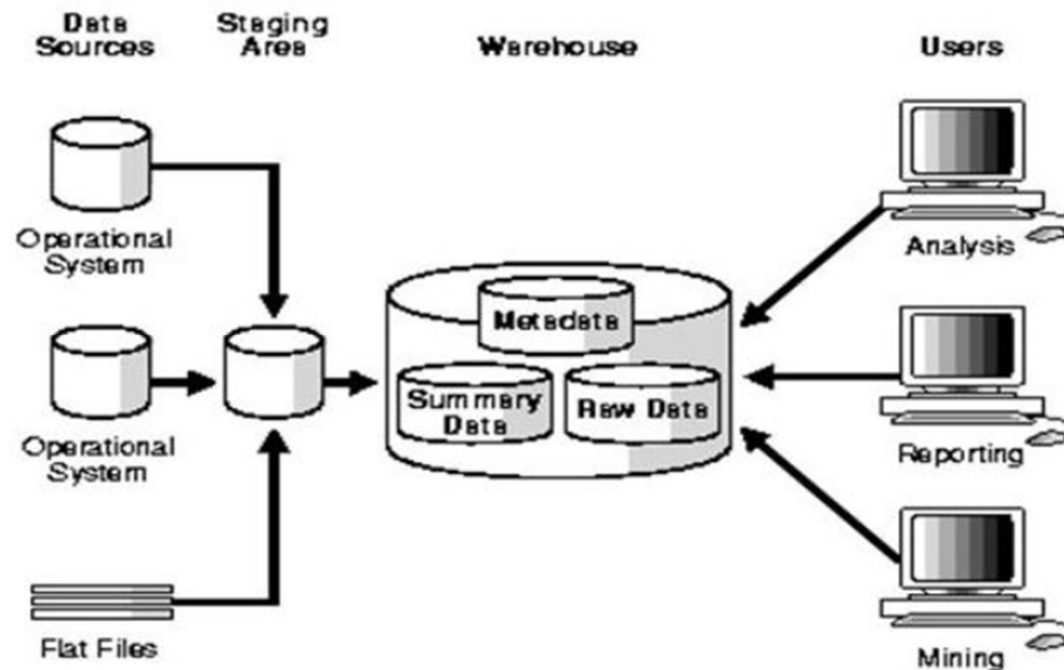
1. Arsitektur Dasar Data Warehouse

Gambar berikut menampilkan arsitektur sederhana dari suatu Data Warehouse. User dapat secara langsung mengakses data yang diambil dari beberapa source melalui data warehouse



Arsitektur Data Warehouse

2. Arsitektur Data Warehouse dengan menggunakan Staging Area

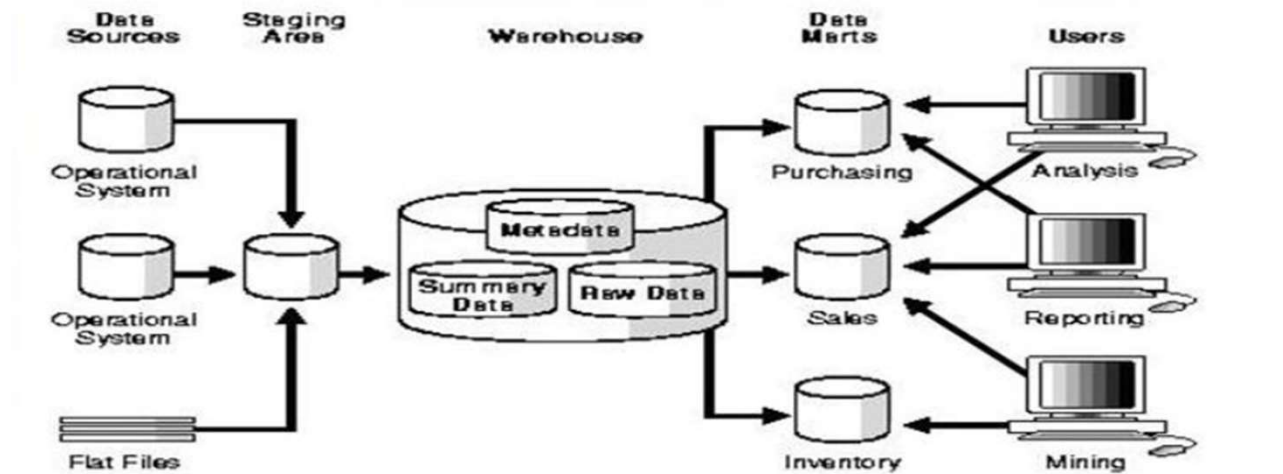


Staging area menyederhanakan proses pembuatan summary dan management warehouse secara umum

Arsitektur Data Warehouse

3. Arsitektur Data Warehouse dengan menggunakan Staging Area dan Data Mart

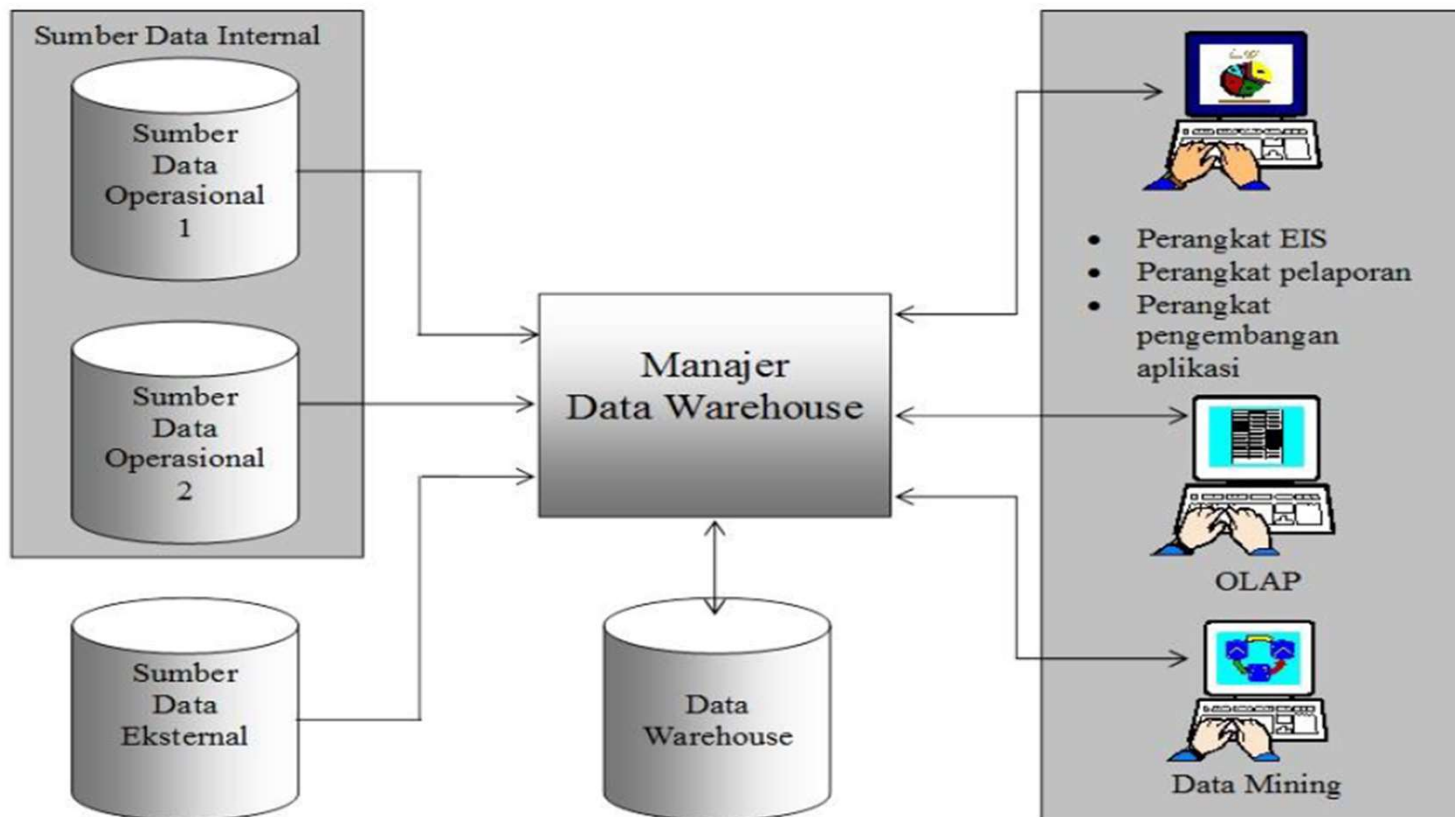
- Data mart merupakan subset dari data resource, biasanya berorientasi untuk suatu tujuan yang spesifik atau subjek data yang didistribusikan untuk mendukung kebutuhan bisnis.



Sumber Data untuk DW

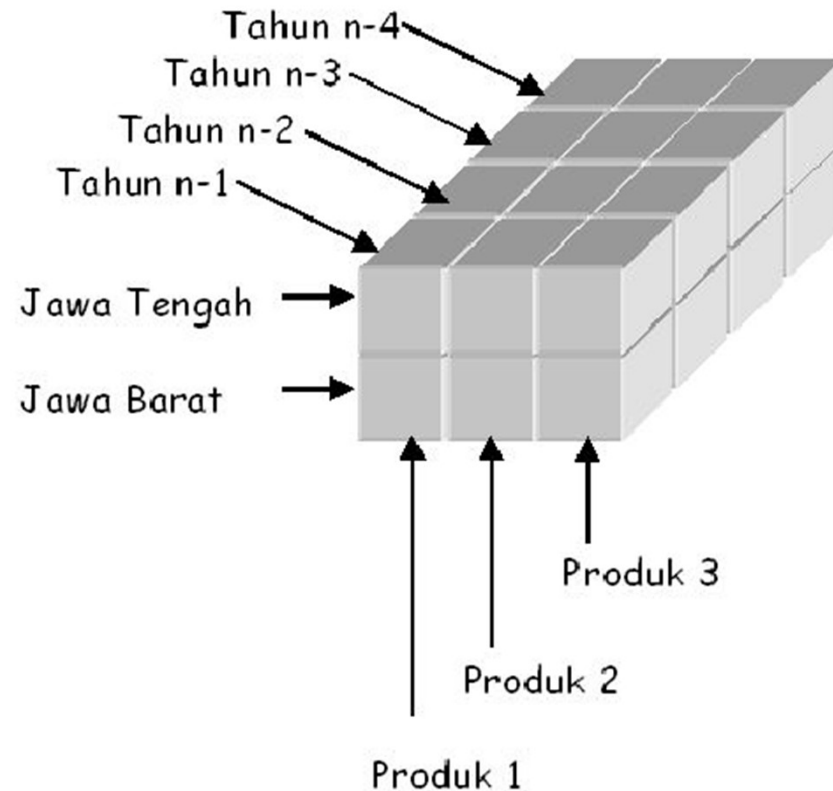
1. Data operasional dalam organisasi, misalnya basis data pelanggan dan produk, dan
 2. Sumber eksternal yang diperoleh misalnya melalui Internet, basis data komersial, basis data pemasok atau pelanggan
- Berbagai data yang berasal dari sumber digabungkan dan diproses lebih lanjut oleh manajer data warehouse 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?”



Multidimensional Database Structure

		Denver		Los Angeles		San Francisco		West	
Sales	Camera	February		March					
		Actual	Budget	Actual	Budget				
Sales	TV								
	VCR								
	Audio								
Margin	Camera								
	TV								
	VCR								
	Audio								

		Profit		Total Expenses		Margin		COGS	
Sales	TV	January		February		March		Qtr 1	
		Actual	Budget	Actual	Budget	Actual	Budget	Actual	Budget
Sales	VCR								
	Qtr 1								
	Qtr 1								

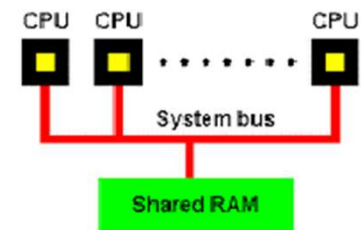
		April		Qtr 1		March		February	
TV	East	January		Actual		Budget			
		Sales	Margin	Sales	Margin				
TV	West								
	South								
	Total								
	Total								
VCR	East								
	West								
	South								
	Total								

		April		Qtr 1		March		February	
East	Actual	January		Sales		Margin			
		TV	VCR	TV	VCR				
East	Budget								
	Forecast								
	Variance								
	Variance								
West	Actual								
	Budget								
	Forecast								
	Variance								

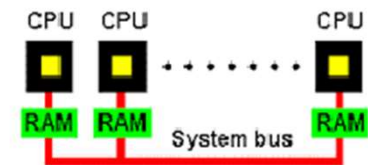
Data Warehouse

- Mengingat sistem data warehouse memerlukan pemrosesan data dengan volume yang besar, sistem ini biasa diterapkan dengan menggunakan teknologi pemrosesan SMP (Symmetric Multiprocessing) dan MPP (Multiple Parallel Processing).
- 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)

- SMP (Symmetric Multiprocessing) adalah pengolahan program dengan beberapa prosesor yang berbagi sistem operasi umum dan memori.
- MPP (Massively Parallel Processing) adalah pengolahan terkoordinasi dari program oleh beberapa prosesor yang bekerja pada bagian yang berbeda dari program ini, dengan masing-masing prosesor menggunakan sistem operasi sendiri dan memori.



SMP: Multiple processors share RAM and system bus



MPP: message-passing moves data through system

Data Marts

- **Connolly, Begg, Strachan (1999):** Bagian dari data warehouse yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu dalam perusahaan.
- Karakteristik yang membedakan data mart dan data warehouse :
 - 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:
 - 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

1. 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
2. **Drill-down** adalah suatu bentuk yang merupakan kebalikan dari konsolidasi, yang memungkinkan data yang ringkas dijabarkan menjadi data yang lebih detail.
3. **Slicing and dicing** (atau dikenal dengan istilah **pivoting**) menjabarkan pada kemampuan untuk melihat data dari berbagai sudut pandang.

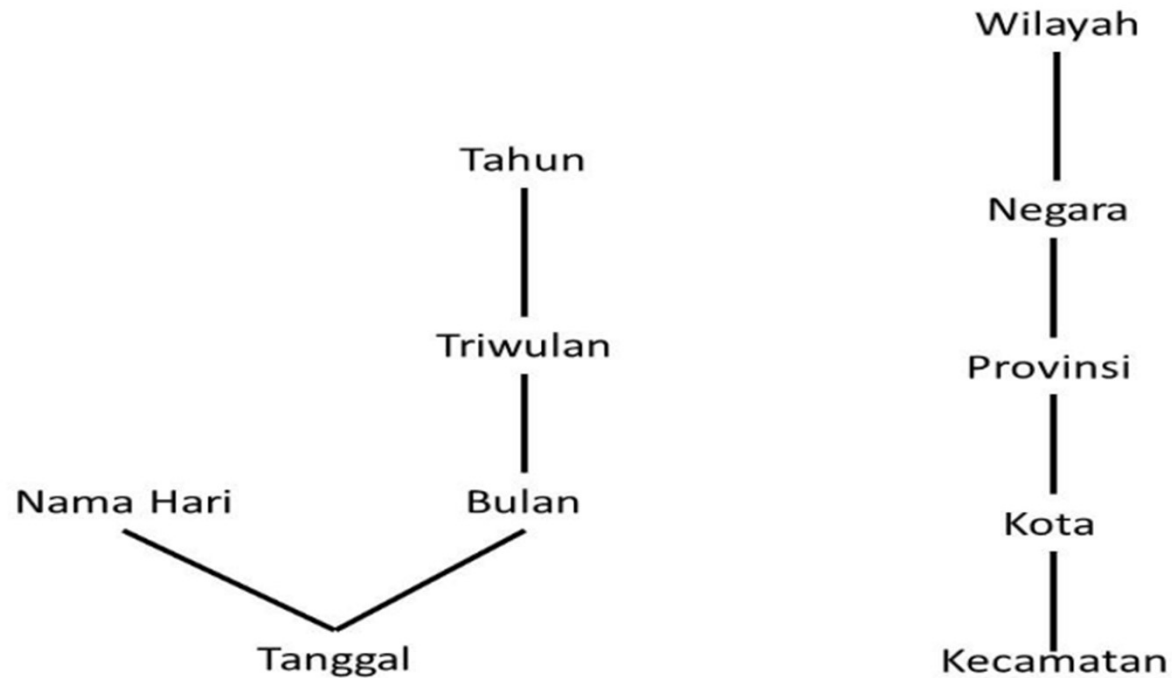
Contoh Tabel Pivoting

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



(a) Hierarki Waktu

(b) Hierarki Lokasi

Software OLAP

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

Data mining

- Data mining merupakan proses untuk menggali pengetahuan dan informasi baru dari data yang berjumlah banyak pada data warehouse, dengan menggunakan kecerdasan buatan (Artificial Intelligence), statistik dan matematika.
- Data mining merupakan teknologi yang diharapkan dapat menjembatani komunikasi antara data dan pemakainya.

Data Mining

- Mengotomatisasi penemuan pola-pola yang tidak diketahui sebelumnya.
- Contoh dari penemuan pola ini adalah analisis pada data penjualan ritel untuk mengidentifikasi produk-produk, yang kelihatannya tidak berkaitan, yang seringkali dibeli secara bersamaan oleh kustomer.
- Contoh lain adalah pendeteksian transaksi palsu dengan kartu kredit

Cara Kerja Data Mining

1. Data MINING untuk menentukan Prospek Model ini dapat diterapkan untuk menarik kesimpulan dari informasi khusus (sebagai data prospek), dimana saat ini informasi khusus tersebut tidak dimiliki oleh perusahaan. Dengan model ini, calon-calon Customer baru dapat ditarget secara selektif.

Tabel 1. Data Mining untuk menentukan Prospek

	Customer	Prospek
Informasi Umum (Contoh data Demografis)	diketahui	diketahui
Informasi Khusus (Contoh Transaksi Customer)	diketahui	target

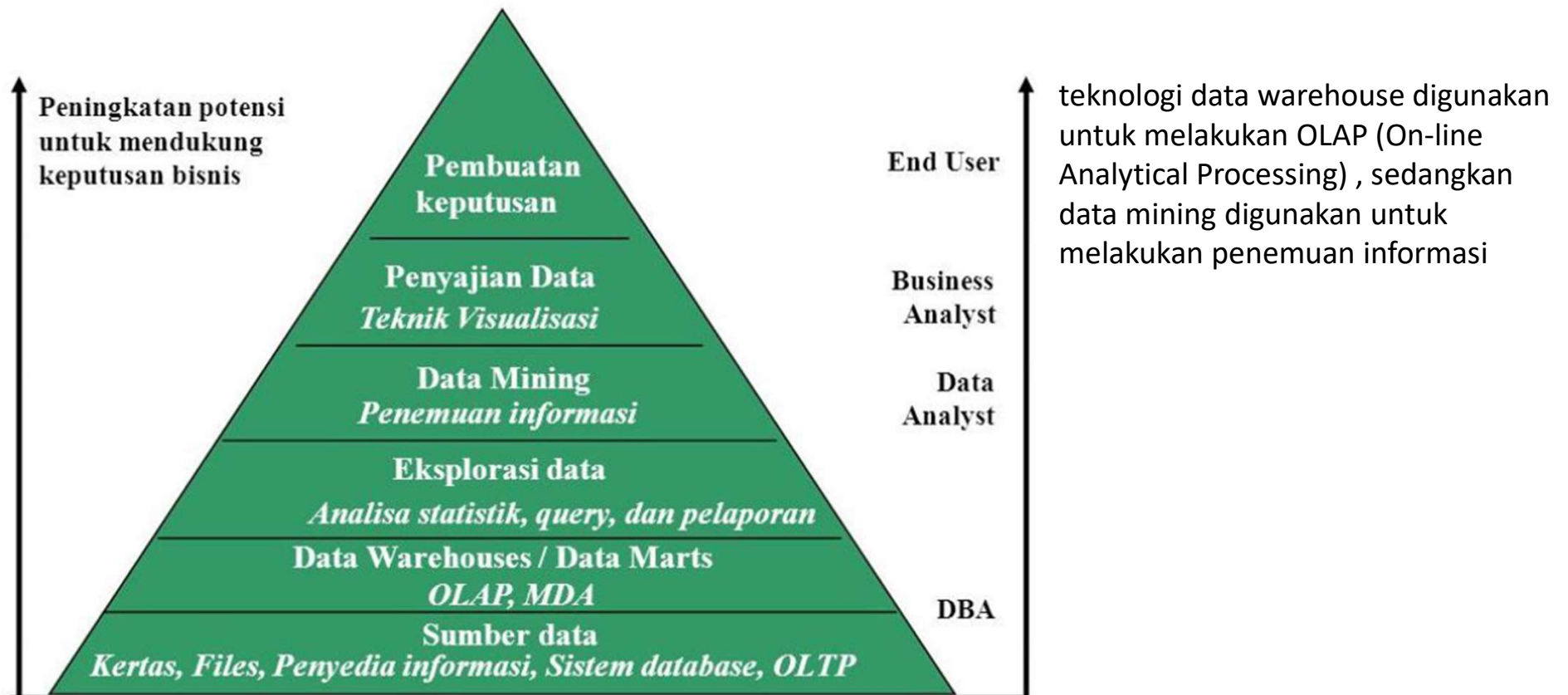
Cara Kerja Data Mining

2. Data MINING untuk prediksi memprediksi apa yang akan terjadi di masa mendatang

Tabel 2. Data Mining untuk Prediksi

	Kemarin	Sekarang	Besok
Informasi Statis dan Rencana terkini (Contoh data demografis, rencana pemasaran, dll)	diketahui	diketahui	diketahui
Informasi Dinamik (Contoh Transaksi Customer)	diketahui	diketahui	target

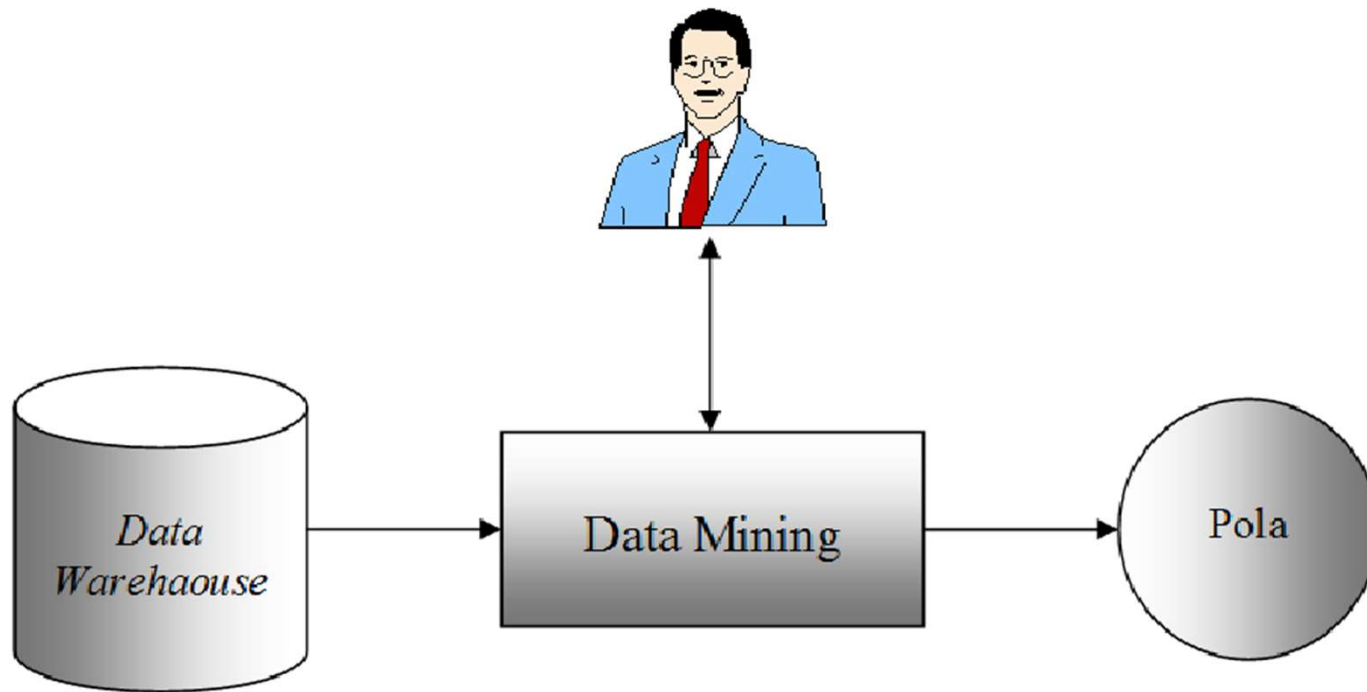
Data Mining dan Teknologi Database



Data Mining dan Teknologi Database

- Dari gambar di atas terlihat bahwa teknologi data warehouse digunakan untuk melakukan OLAP(On-line Analytical Processing)
- Data mining digunakan untuk melakukan information discovery yang informasinya lebih ditujukan untuk seorang Data Analyst dan Business Analyst.
- Dalam prakteknya, data mining juga mengambil data dari data warehouse. Hanya saja aplikasi dari data mining lebih khusus dan lebih spesifik dibandingkan OLAP mengingat database bukan satu-satunya bidang ilmu yang mempengaruhi data mining.

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

