DATA WAREHOUSE

http://www.brigidaarie.com

MASIH INGAT DENGAN DATABASE???

- **Apa itu database?**
  - tempat penyimpanan data yang saling berhubungan secara logika
- **Untuk apa database itu??**
  - untuk mendapatkan suatu informasi yang diperlukan oleh suatu organisasi atau perusahaan
- **Dapat darimana seh data-data itu??**
  - didapat dari kegiatan operasional sehari-hari atau hasil dari transaksi

Perkembangan model database

DATA WAREHOUSE

APA ITU DATA WAREHOUSE??

- Menurut W.H. Inmon dan Richard D.H.
  - koleksi data yang mempunyai sifat berorientasi subjek,terintegrasi,time-variant, dan bersifat tetap dari koleksi data dalam **mendukung proses pengambilan keputusan management**
- Menurut Vidette Poe
  - database yang bersifat analisis dan *read only* yang digunakan sebagai **fondasi dari sistem penunjang keputusan**

DATA WAREHOUSE??

- Menyimpan data sekarang dan data masa lalu yang berasal dari berbagai sistem operasional dan sumber yang lain (sumber eksternal)
- Menjadi perhatian penting bagi manajemen dalam organisasi
- Ditujukan untuk keperluan analisis dan pelaporan manajemen
- Dalam rangka pengambilan keputusan

DATA WAREHOUSE??

- Digunakan untuk mendukung pengambilan keputusan, bukan untuk melaksanakan pemrosesan transaksi
- Hanya berisi informasi-informasi yang relevan bagi kebutuhan pemakai yang dipakai untuk pengambilan keputusan

DATA WAREHOUSE

DATABASE

TAPI.....

PERANCANGAN DATA WAREHOUSE

$\neq$

PERANCANGAN DATABASE

JADI DATA WAREHOUSE =

- Database yang **saling berelasi** yang dapat **digunakan untuk *query* dan analisisis**, bersifat orientasi subjek, terintegrasi, time-variant,tidak berubah yang digunakan untuk **membantu para pengambil keputusan**

BERORIENTASI SUBYEK

- Didesain untuk menganalisa data berdasarkan subject-subject tertentu dalam organisasi,bukan pada proses atau fungsi aplikasi tertentu

TERINTEGRASI

- Dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah kedalam suatu format yang konsisten dan saling terintegrasi satu dengan lainnya.
- Dengan demikian data tidak bisa dipecah-pecah.

TIME VARIANT

- Seluruh data pada data warehouse dapat dikatakan akurat atau valid pada rentang waktu tertentu.

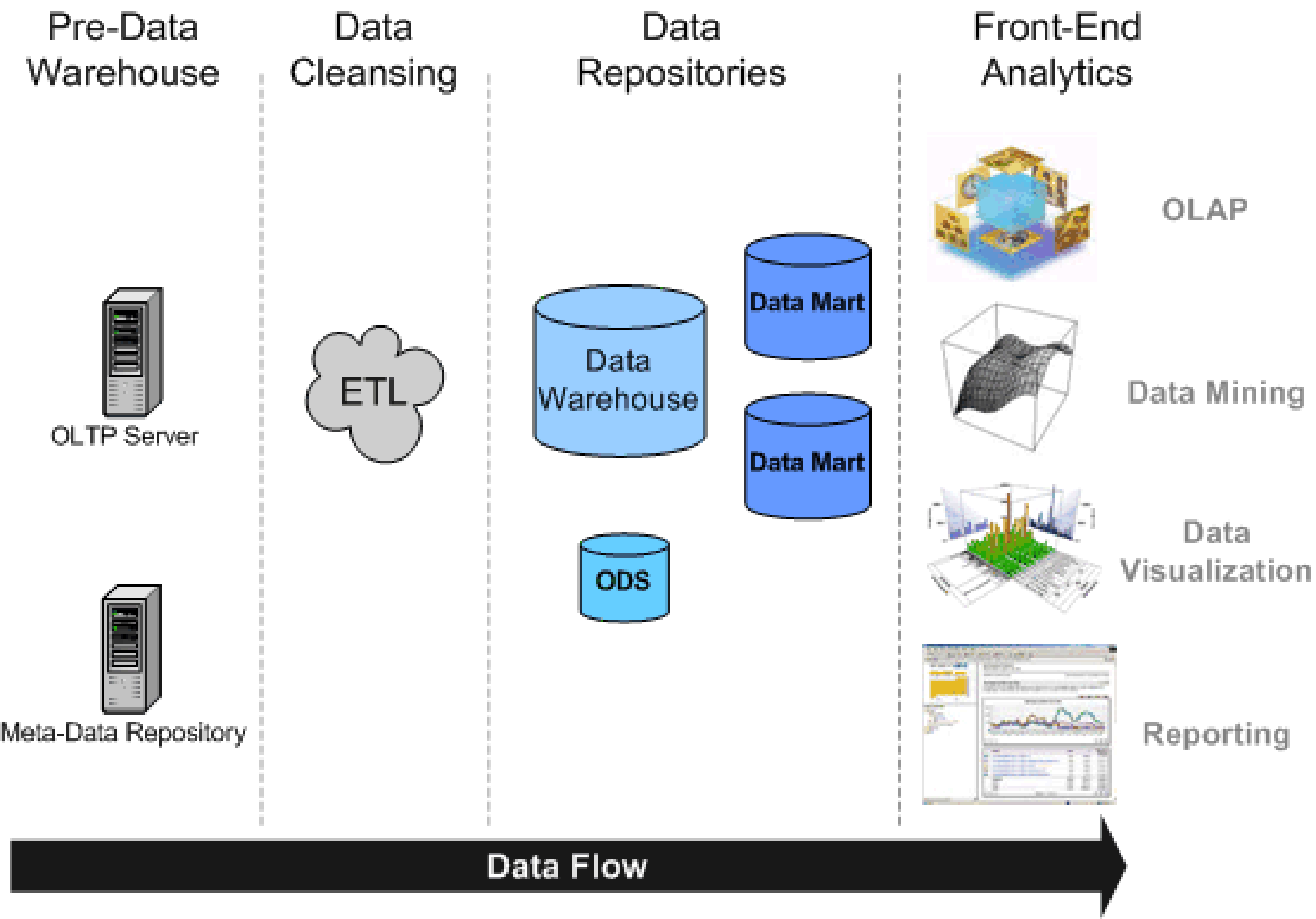
NON-VOLATILE

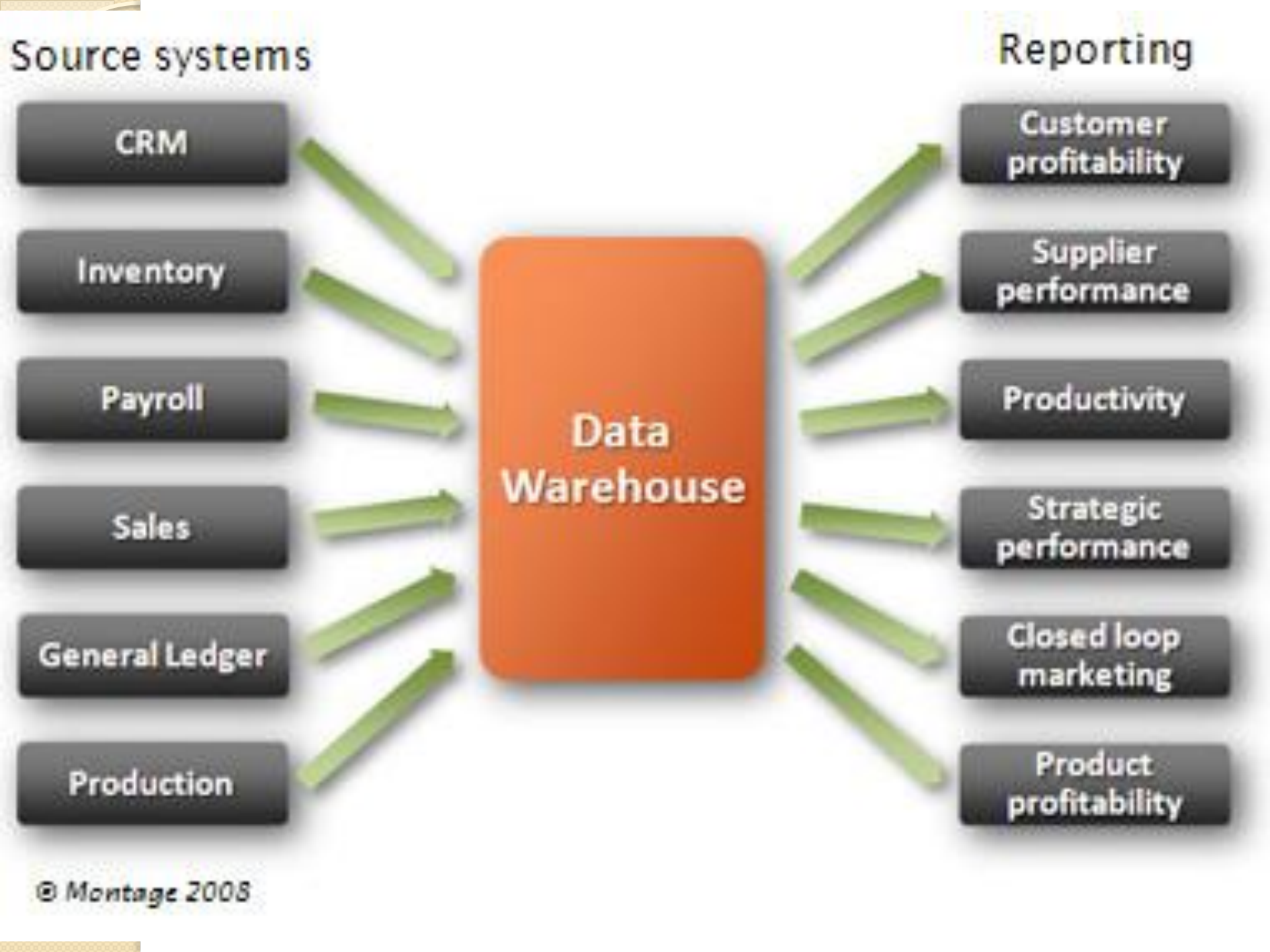
- Data pada data warehouse tidak di-update secara real time tetapi di refresh dari sistem operasional secara reguler.
- Data yang baru selalu ditambahkan sebagai suplemen bagi database itu sendiri dari pada sebagai sebuah perubahan

Fungsinya buat apa seh???

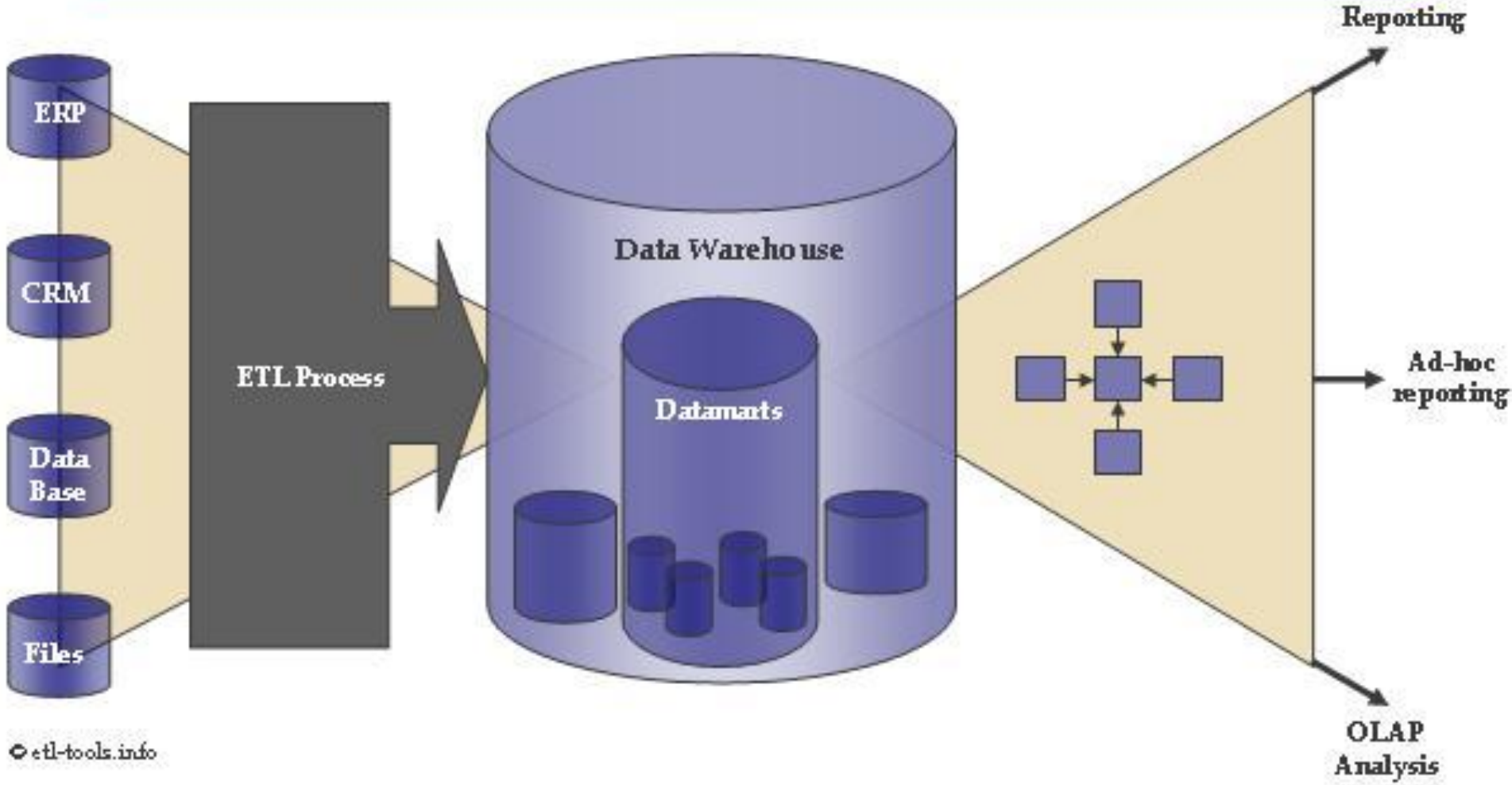
- Pembuatan laporan
- On-Line Analytical Processing (OLAP)
- Data Mining
- Proses informasi executive

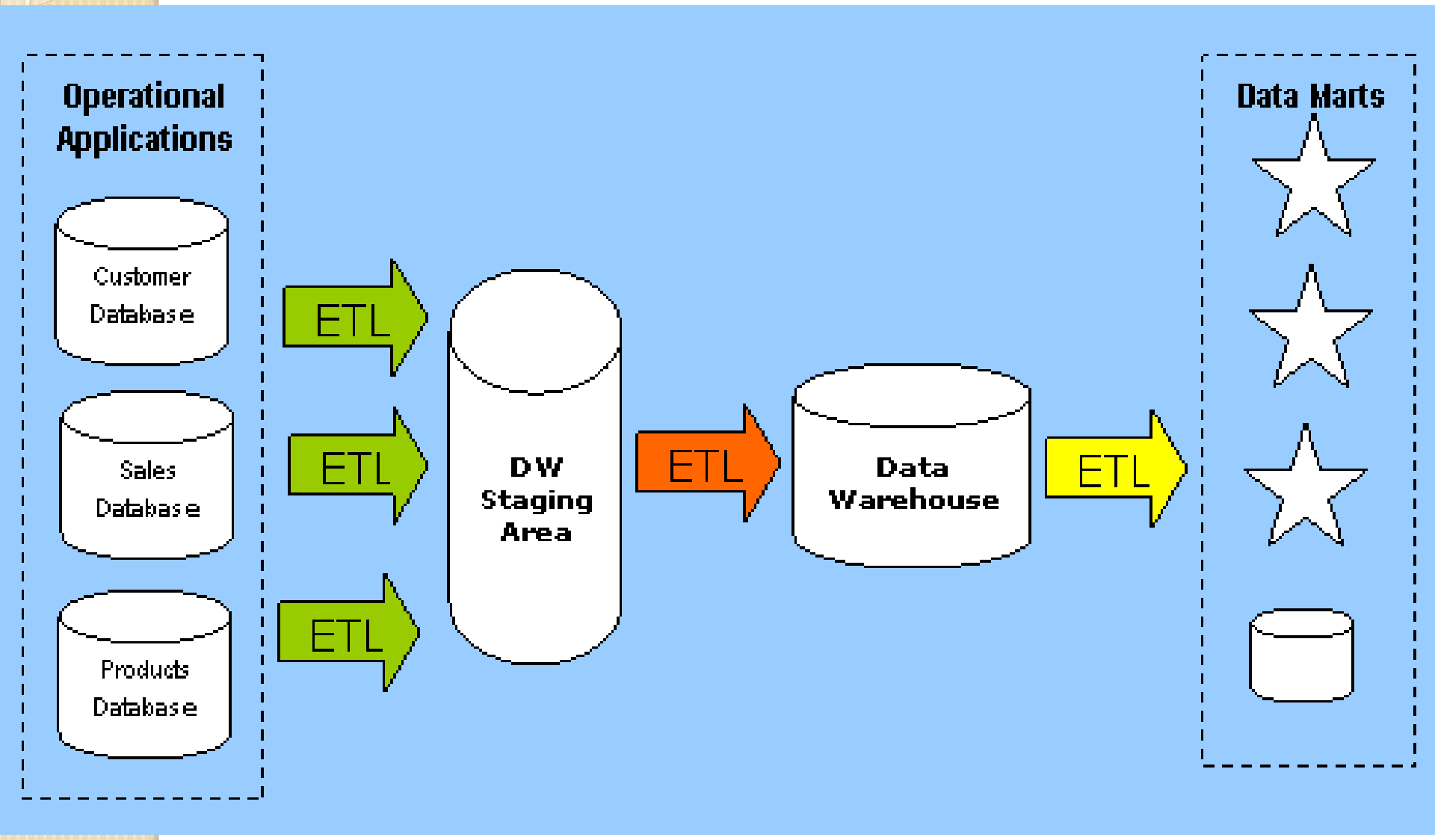
- Menyediakan data organisasi yang mudah diakses oleh manager.
- Data yang berada di datawarehouse bersifat konsisten, dan merupakan kebenaran.
- Datawarehouse merupakan tempat, dimana data yang telah digunakan di publikasikan.
- Kualitas data di datawarehouse dapat diandalkan.

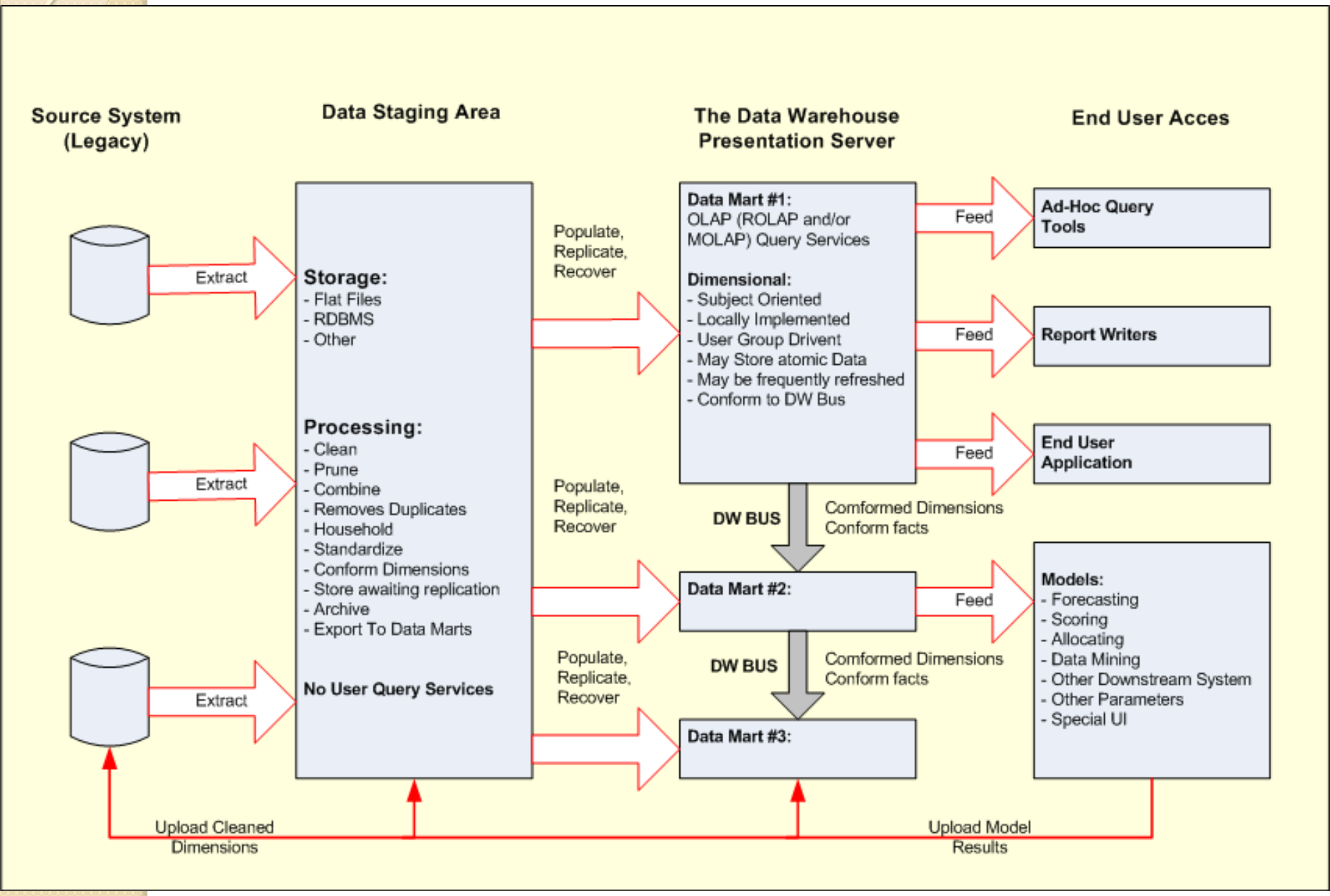

Pre-Data Warehouse
Data Cleansing
Data Repositories
Front-End Analytics
OLTP Server
Meta-Data Repository
ETL
Data Warehouse
Data Mart
Data Mart
ODS
OLAP
Data Mining
Data Visualization
Reporting
Data Flow


Source systems
CRM
Inventory
Payroll
Sales
General Ledger
Production
Data Warehouse
Reporting
Customer profitability
Supplier performance
Productivity
Strategic performance
Closed loop marketing
Product profitability
© Montage 2008

Business Intelligence




Operational Applications
Customer Database
Sales Database
Products Database
ETL
ETL
ETL
DW Staging Area
ETL
Data Warehouse
ETL
Data Marts


Source System (Legacy)
Data Staging Area
The Data Warehouse Presentation Server
End User Acces
Extract
Extract
Extract
Storage:
- Flat Files
- RDBMS
- Other
Processing:
- Clean
- Prune
- Combine
- Removes Duplicates
- Household
- Standardize
- Conform Dimensions
- Store awaiting replication
- Archive
- Export To Data Marts
No User Query Services
Populate, Replicate, Recover
Populate, Replicate, Recover
Populate, Replicate, Recover
Data Mart #1:
OLAP (ROLAP and/or MOLAP) Query Services
Dimensional:
- Subject Oriented
- Locally Implemented
- User Group Drivent
- May Store atomic Data
- May be frequently refreshed
- Conform to DW Bus
DW BUS
Comformed Dimensions
Conform facts
Data Mart #2:
DW BUS
Comformed Dimensions
Conform facts
Data Mart #3:
Feed
Feed
Feed
Feed
Ad-Hoc Query Tools
Report Writers
End User Application
Models:
- Forecasting
- Scoring
- Allocating
- Data Mining
- Other Downstream System
- Other Parameters
- Special UI
Upload Cleaned Dimensions
Upload Model Results

Source System

- Operasional sistem yang berfungsi mencatat transaksi dari suatu bisnis.

Data Staging Area

- Dalam tahap ini, data diolah dari sumbernya untuk siap menjawab query.
- Prosesnya terdiri dari *extract, transform, load (ETL)*.

DATA

Operational Vs Informational

Operational Data

- Data yang digunakan untuk menjalankan bisnis.
  - Cirinya : disimpan, diperoleh dan diupdate oleh system Online Transactional Processing (OLTP).
  - Contohnya : system pemesanan, aplikasi perhitungan atau an order entry application

Informational Data

- Biasanya disimpan dalam relational database,
- Tetapi mungkin disimpan dalam legacy hierarchical atau flat formats as well

OLTP VS Data Warehouse

OLTP	Data Warehouse
• 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	• 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

Datanya dari mana aja tu??

Sumber Data Warehouse

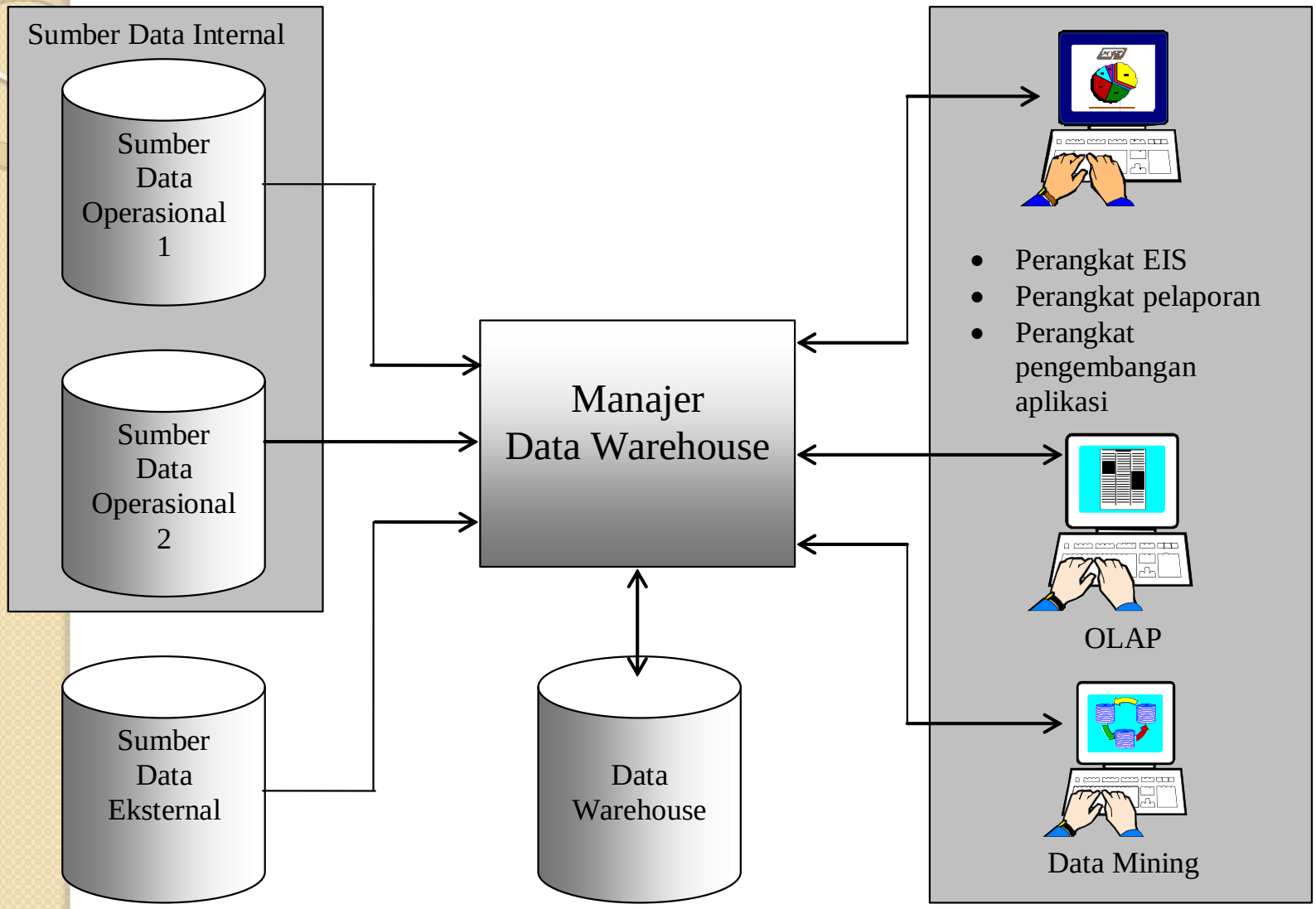
1. Data operasional dalam organisasi
   - misalnya basis data pelanggan dan produk
2. Sumber eksternal yang diperoleh
   - misalnya melalui Internet, basis data komersial, basis data pemasok atau pelanggan

Trus mau diapain data-data itu?

Setelah itu….maka….

- 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



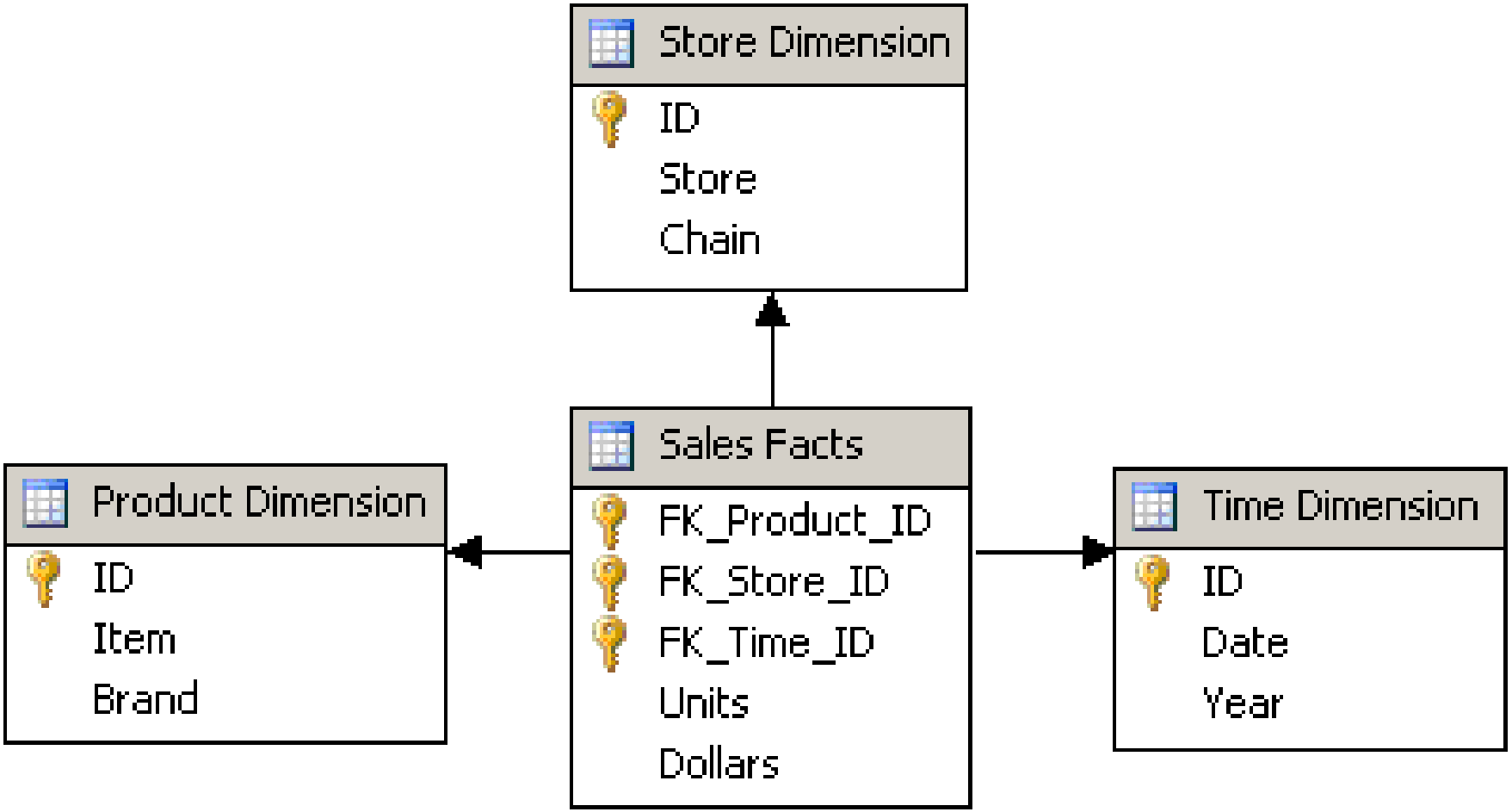
Sifatnya gimana data warehouse itu?

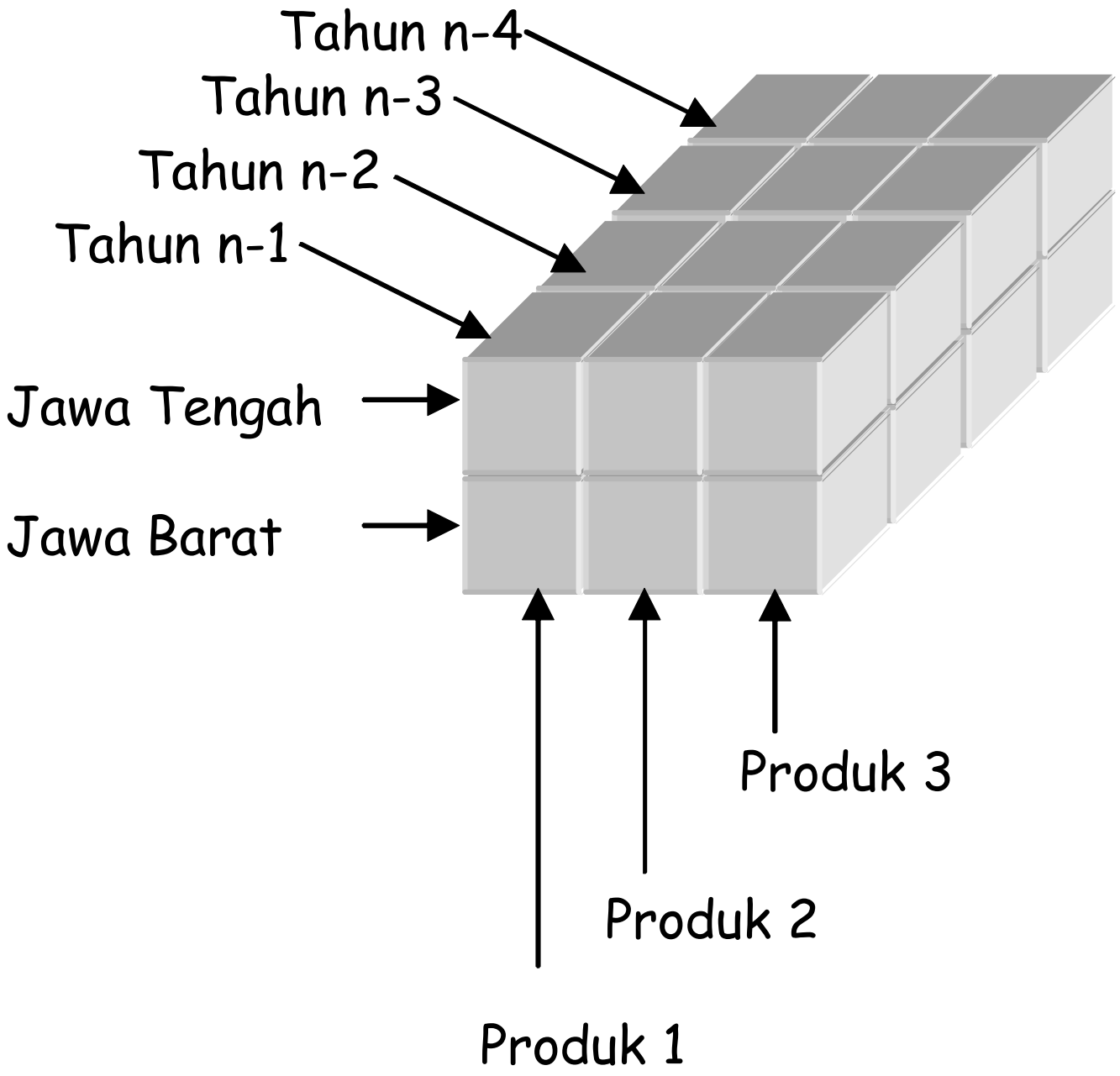
Sifat Data Warehouse

- **Multidimensional** yang berarti bahwa terdapat banyak lapisan kolom dan baris
- Berbeda dengan tabel pada model relasional yang hanya berdimensi dua

Kayak gimana Multidimensional??

Star-Schema


Store Dimension
ID
Store
Chain
Sales Facts
FK_Product_ID
FK_Store_ID
FK_Time_ID
Units
Dollars
Product Dimension
ID
Item
Brand
Time Dimension
ID
Date
Year


Tahun n-4
Tahun n-3
Tahun n-2
Tahun n-1
Jawa Tengah
Jawa Barat
Produk 3
Produk 2
Produk 1

Sifat Data Warehouse

- 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?"

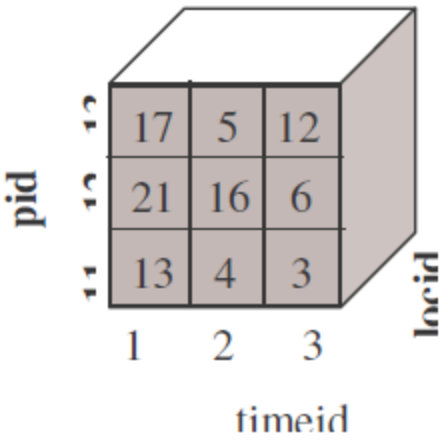
Multidimensional??

- Himpunan pengukuran numerik yang tergantung pada himpunan dimensi.
- Misalnya untuk mengetahui Penjualan/Sales, dimensinya
  - Produk (pid)
  - Lokasi (locid)
  - Waktu (timeid).

Multidimensional??

- Dapat disimpan secara fisik dalam sebuah array yang disebut sistem **MOLAP**. Alternatif lainnya, data dapat disimpan sebagai relasi yang disebut sistem **ROLAP**.

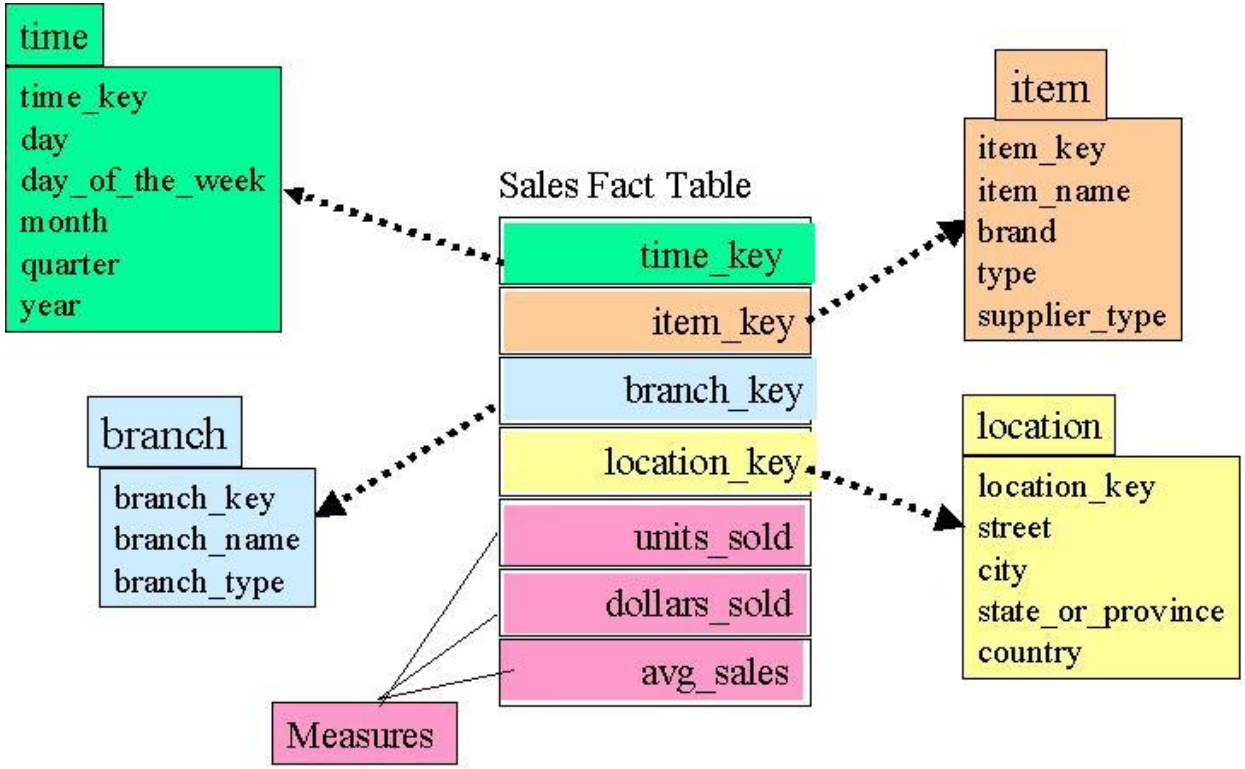
Pid	timeid	locid	sales
11	1	1	13
11	2	1	4
11	3	1	3
12	1	1	21
12	2	1	16
12	3	1	6
13	1	1	17
13	2	1	5
13	3	1	12



Tabel Fakta dan Tabel Dimensi

- 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.

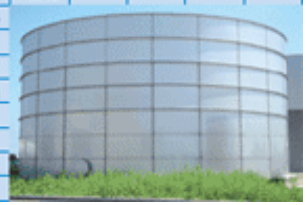
Tabel Fakta



Merupakan tabel yang umumnya mengandung angka dan data history dimana *key* (kunci) yang dihasilkan sangat unik

Tabel Dimensi

	Height / Höhe	1,25	2,45	3,65	4,85	6,05	7,25	8,45	9,65	10,85	12,05	13,25	14,45	15,65
Ref.	Diameter	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net	Cap Net
Diam	m	m^3	m^3	m^3	m^3	m^3	m^3	m^3	m^3	m^3	m^3	m^3	m^3	m^3
04	3,07	6	15	24	33	42	51	60	68	77	86	95	104	113
05	3,89	10	24	39	53	67	81	96	110	124	138	153	167	181
06	4,60	14	34	54	74	94	114	134	154	174	194	214	233	253
07	5,42	20	47	75	103	130	158	186	213	241	269	296	324	352
08	6,14	25	61	96	132	167	203	238	274	309	345	380	416	452
09	6,96	32	78	124	169	215	261	306	352	398	443	489	535	580
10	7,67	39	95	150	206	261	316	372	427	483	538	594	649	705
11	8,49	48	116	184	252	320	388	456	524	592	660	727	795	863
12	9,21	57	137	217	296	376	456	536	616	696	776	856	936	1.016
13	10,02	67	162	256	351	446	540	635	729	824	919	1.013	1.108	1.202
14	10,74	77	186	294	403	512	621	729	838	947	1.055	1.164	1.273	1.382
15	11,56	89	215	341	467	593	719	845	971	1.097	1.223	1.349	1.475	1.601
16	12,27	101	242	384	526	668	810	952	1.094	1.236	1.378	1.519	1.661	1.803
17	13,09	114	276	437	599	760	922	1.083	1.245	1.406	1.568	1.729	1.891	2.052
18	13,81	127	307	487	667	846	1.026	1.206	1.385	1.565	1.745	1.925	2.104	2.284
19	14,63	143	345	546	748	950	1.151	1.353	1.555	1.757	1.958	2.160	2.362	2.564
20	15,34	157	379	601	822	1.044	1.266	1.488	1.710	1.931	2.153	2.375	2.597	2.818
21	16,16	174	420	667	913	1.159	1.405	1.651	1.897	2.143	2.389	2.635	2.882	3.128
22	16,88	190	459	727	996	1.264	1.533	1.801	2.070	2.339	2.607	2.876	3.144	3.413
23	17,70	209	504	800	1.095	1.390	1.685	1.981	2.276	2.571	2.866	3.162	3.457	3.752
24	18,41	226	546	865	1.185	1.504	1.823	2.143	2.462	2.782	3.101	3.420	3.740	4.059
25	19,23	247	595	944	1.292	1.641	1.989	2.338	2.686	3.035	3.383	3.732	4.080	4.429
26	19,95	266	641	1.016	1.391	1.766	2.141	2.516	2.891	3.266	3.642	4.017	4.392	4.767
27	20,76	288	694	1.100	1.506	1.912	2.319	2.725	3.131	3.537	3.943	4.349	4.756	5.162
28	21,48	308	743	1.178	1.613	2.047	2.482	2.917	3.352	3.787	4.222	4.656	5.091	5.526
29	22,30	332	801	1.269	1.738	2.207	2.675	3.144	3.613	4.081	4.550	5.019	5.487	5.956
30	23,01	353	852	1.351	1.850	2.349	2.848	3.347	3.846	4.345	4.844	5.343		
31	23,83	379	914	1.449	1.985	2.520	3.055	3.590	4.125	4.661	5.196			
32	24,55	402	970	1.538	2.106	2.674	3.242	3.810	4.378	4.946				
33	25,37	430	1.036	1.643	2.249	2.856	3.463	4.069	4.676					
34	26,08	454	1.095	1.736	2.377	3.018	3.659	4.300	4.941					
35	26,90	483	1.165	1.847	2.529	3.211	3.893	4.575	5.257					
36	27,62	509	1.228	1.947	2.666	3.385	4.104	4.823	5.542					
37	28,44	540	1.302	2.065	2.827	3.589	4.351	5.114						
38	29,15	567	1.368	2.169	2.970	3.771	4.571	5.372						
39	29,97	600	1.446	2.293	3.139	3.986	4.832							
40	30,69	629	1.516	2.404	3.292	4.179	5.067							



- ✓Berisikan kategori dengan ringkasan data detail yang dapat dilaporkan
- ✓Seperti laporan laba pada tabel fakta dapat dilaporkan sebagai dimensi waktu(yang berupa perbulan, perkwartal dan pertahun)

OLAP
(*OnLine Analytical Processing*)

Apaan tu???

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

- **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

Contoh 2 Dimensi

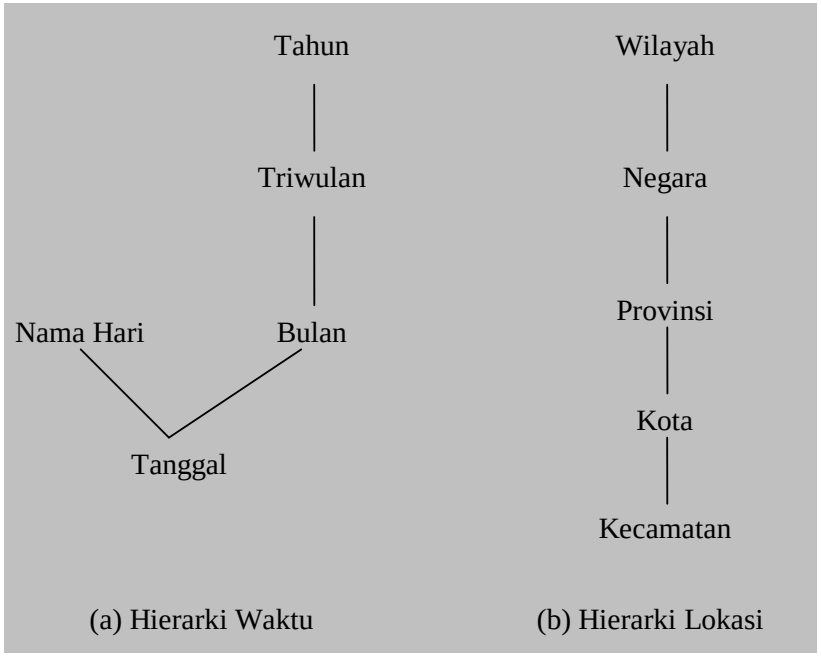
Kota → / Triwulan ↓	Kudus	Magelang	Semarang	…
1	6.000.000	8.500.000	12.500.000	…
2	4.500.000	3.500.000	14.000.000	…
3	7.600.000	5.500.000	13.700.000	…
4	5.400.000	7.200.000	12.800.00	…

Jenis Query OLAP

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

Jenis Query OLAP

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

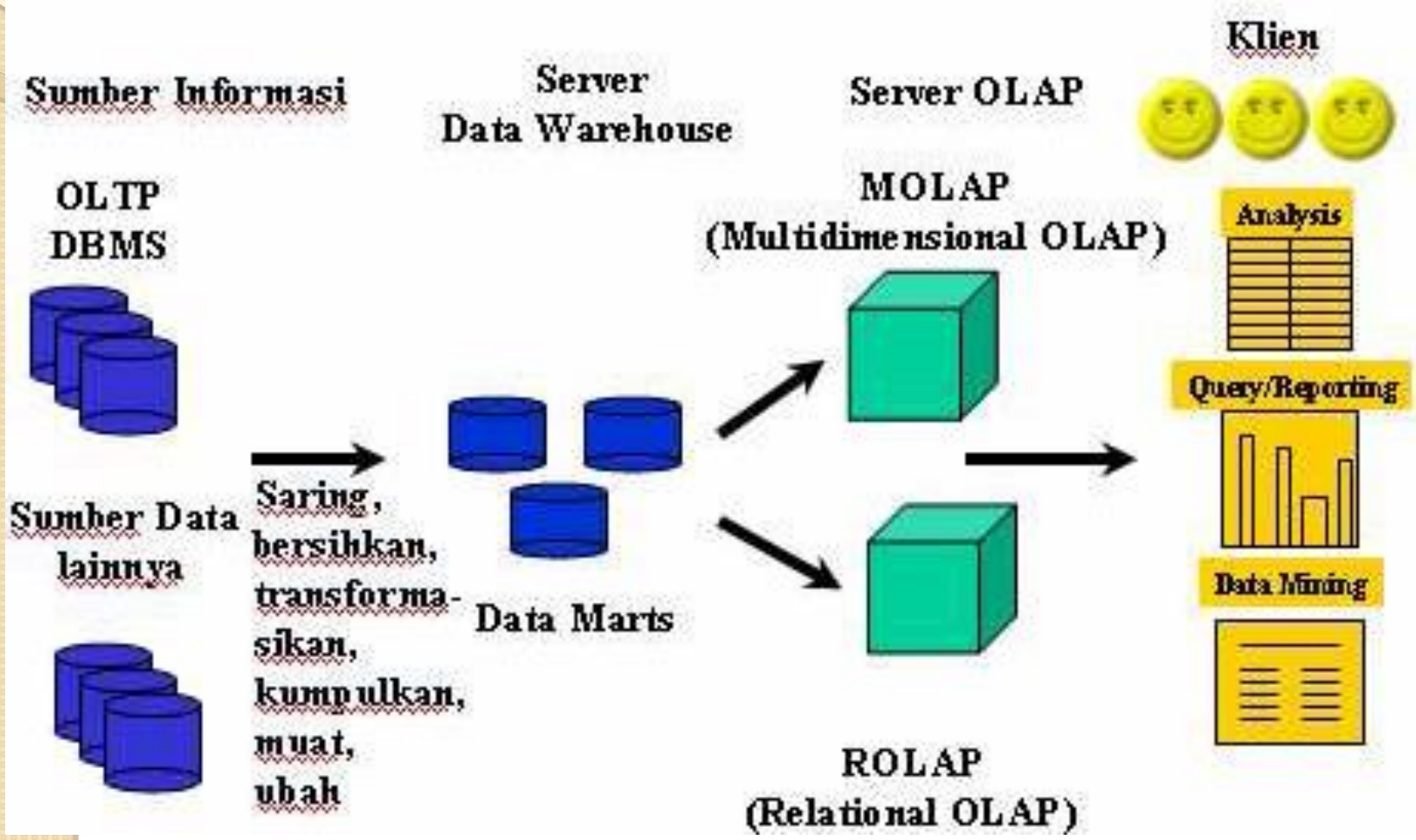


Jenis Query OLAP

- Pivoting
  - 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

Arsitektur Data Warehouse



Data Mart

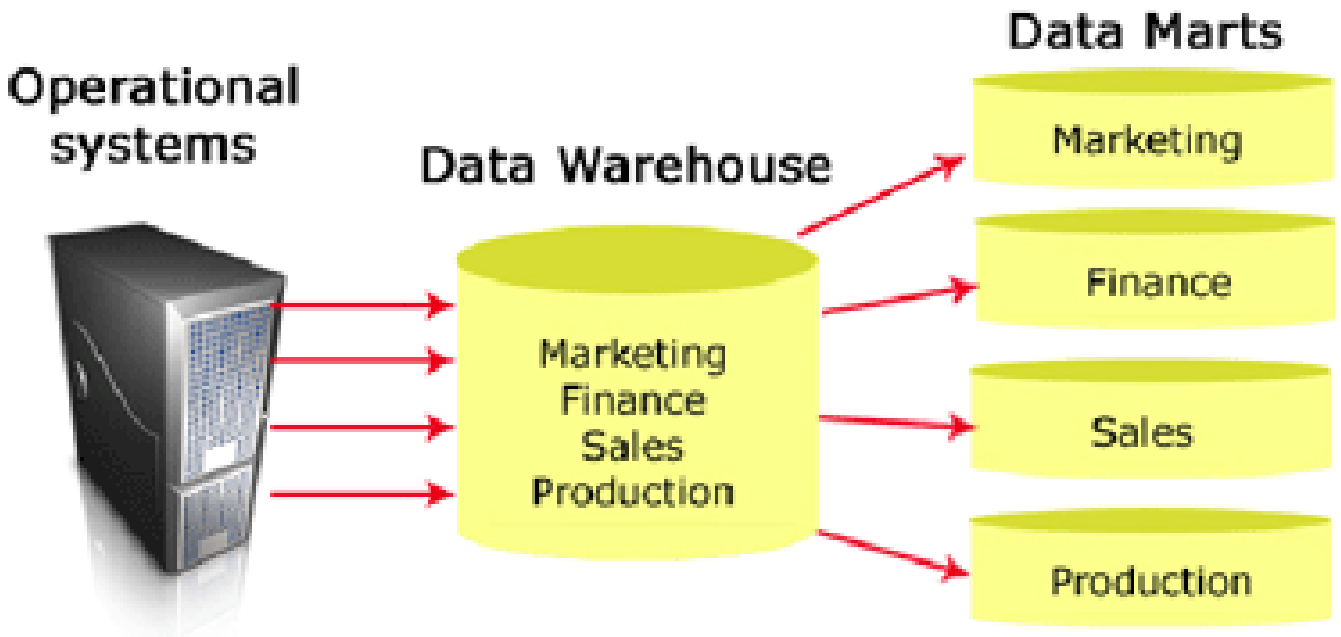
- Fasiltas penyimpan data yang berorentasi pada Subject tertentu atau berorentasi pada Departemen tertentu dari suatu organisasi
- Sehingga suatu Organisasi bisa mempunyai lebih dari satu Data Mart.

Dependent Data Mart
VS
Independent Data Mart

- Perbedaan dari kedua arsitektur tersebut hanya terletak pada ketergantungan sumber datanya terhadap data warehouse

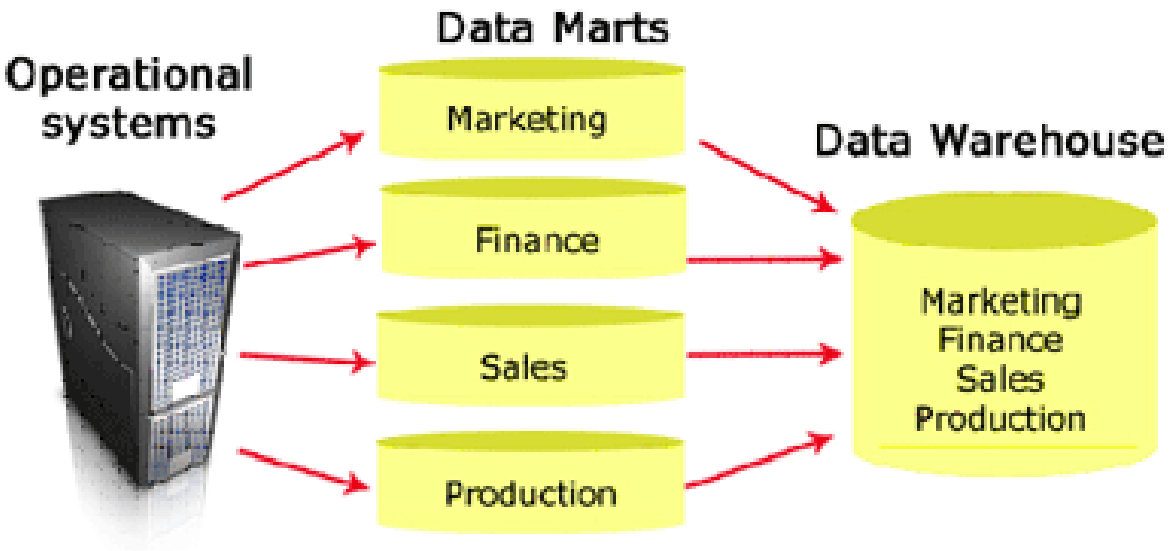
Dependent Data Mart (Inmon advocated)

- Berlaku sebagai komponen atau suatu bagian dari enterprise Data Warehouse, Data Mart dibangun dengan cara extract data dari Data Warehouse.



Independent Data Mart (Kimball advocated)

- Dibangun dengan cara extract langsung data dari berbagai Source System. Independent Data Mart tidak tergantung pada pusat penyimpan data seperti Data Warehouse.



Category	Data Mart	Data Warehouse
Scope	Departemental	Corporate
Granularity	High level	Lowest Level
Subject	Single	Multiple
Data Sources	Sedikit	Banyak
Index	Highly Indexed	Lightly Indexed
Size (typical)	Ratusan GB	Beberapa TB
Age	Beberapa Hari atau Bulan	Beberapa Tahun

Distributed Data Warehouse

- Dengan Local dan Global Data Warehouse
  - Local → merepresentasikan data dan proses-proses pada remote site.
  - Global → merepresentasikan bagian dari bisnis yang diintegrasikan.
- Technologically Distributed Data Warehouse
  - secara logic → single data warehouse
  - secara fisik → terdapat beberapa data warehouse yang diintegrasikan yang berasal dari beberapa prosesor/site yang berbeda.

- Independently Evolving Distributed Data Warehouse
  - Setiap bagian dari distributed data warehouse, memiliki otonomi untuk mengatur dan mengembangkan bisnisnya tanpa harus memperhatikan bagian yang lain