

BISNIS CERDAS



Pertemuan 5 **Data Warehouse**

Sri Herawati

PRODI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO

Bisnis Cerdas



- **Tujuan Instruksional Umum**
Mahasiswa mampu memahami dan mengimplementasikan Bisnis cerdas
- **Tujuan Instruksional Khusus**
Mahasiswa dapat menjelaskan konsep datawarehouse

Topik Bahasan



- Pengertian Datawarehouse
- Karakteristik Datawarehouse
- Proses Datawarehouse
- Arsitektur Datawarehouse
- ***OLAP*** (On-line analytical processing)

Data Warehouse

Definisi :

- Data Warehouse adalah Pusat repositori informasi yang mampu memberikan database berorientasi subyek untuk informasi yang bersifat historis yang mendukung DSS (Decision Suport System) dan EIS (Executive Information System).
- Salinan dari transaksi data yang terstruktur secara spesifik pada query dan analisa.
- Salinan dari transaksi data yang terstruktur spesifik untuk query dan laporan.

Tujuan :

Meningkatkan kualitas dan akurasi informasi bisnis dan mengirimkan informasi ke pemakai dalam bentuk yang dimengerti dan dapat diakses dengan mudah.

Karakteristik data warehouse

- Subject oriented
- Integrated
- Time variant
- Non-volatile

Karakteristik data warehouse

- **Subject oriented**

- Data yang disusun menurut subyek berisi hanya informasi yang penting bagi pemrosesan decision support.
- Database yang semua informasi yang tersimpan di kelompokkan berdasarkan subyek tertentu misalnya: pelanggan, gudang, pasar, dsb.
- Semua Informasi tersebut disimpan dalam suatu sistem *data warehouse*.
- Data-data di setiap subyek dirangkum ke dalam dimensi, **misalnya** : periode waktu, produk, wilayah, dsb, sehingga dapat memberikan nilai sejarah untuk bahan analisa.

Karakteristik data warehouse

- **Integrated**

- Jika data terletak pada berbagai aplikasi yang terpisah dalam suatu lingkungan operasional, encoding data sering tidak seragam sehingga bila data dipindahkan ke data warehouse maka coding akan diasumsikan sama seperti lazimnya.

Karakteristik data warehouse

- **Time variant**

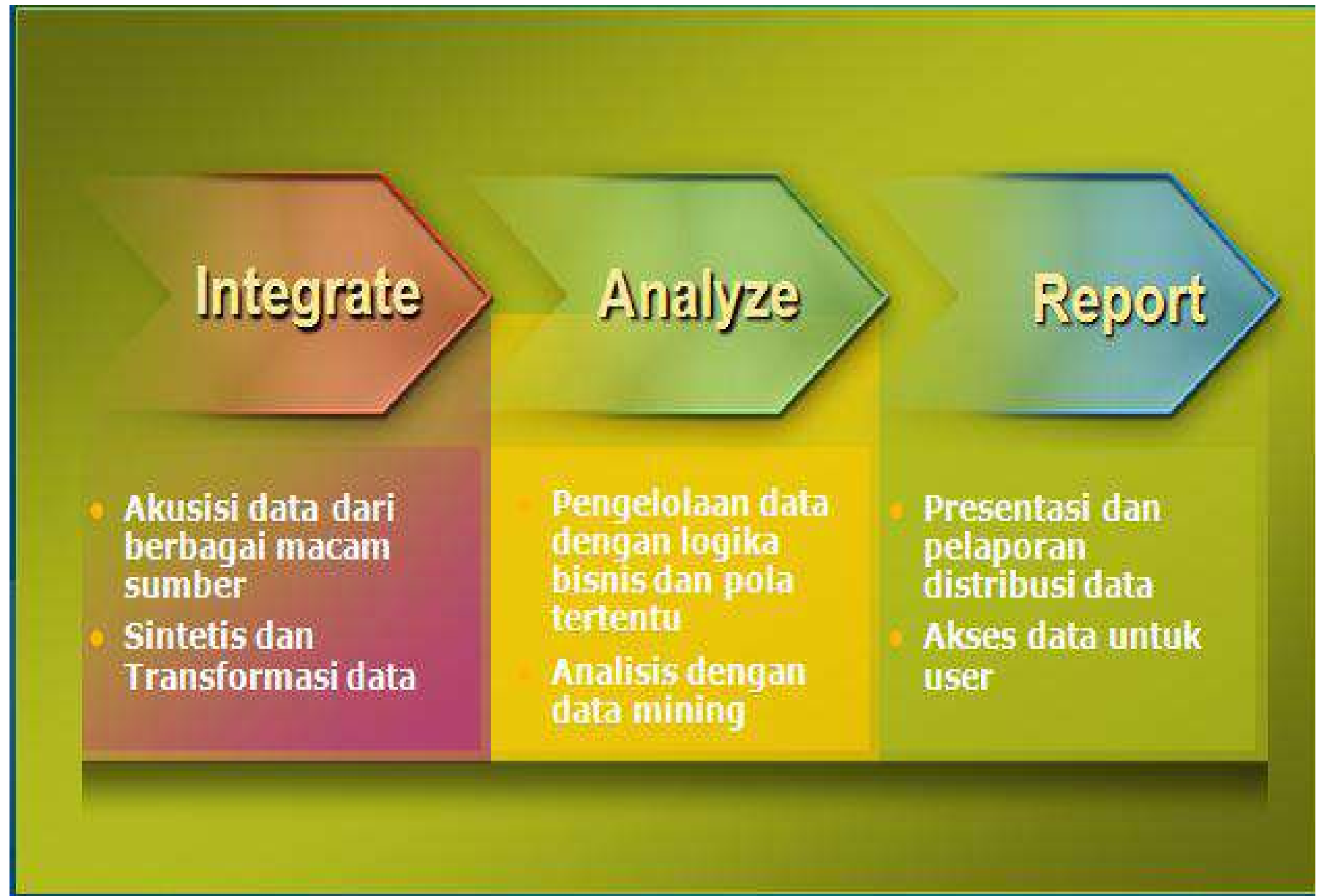
Data warehouse adalah tempat untuk storing data selama 5 sampai 10 tahun atau lebih, data digunakan untuk perbandingan atau perkiraan dan data ini tidak dapat diperbaharui.

Karakteristik data warehouse

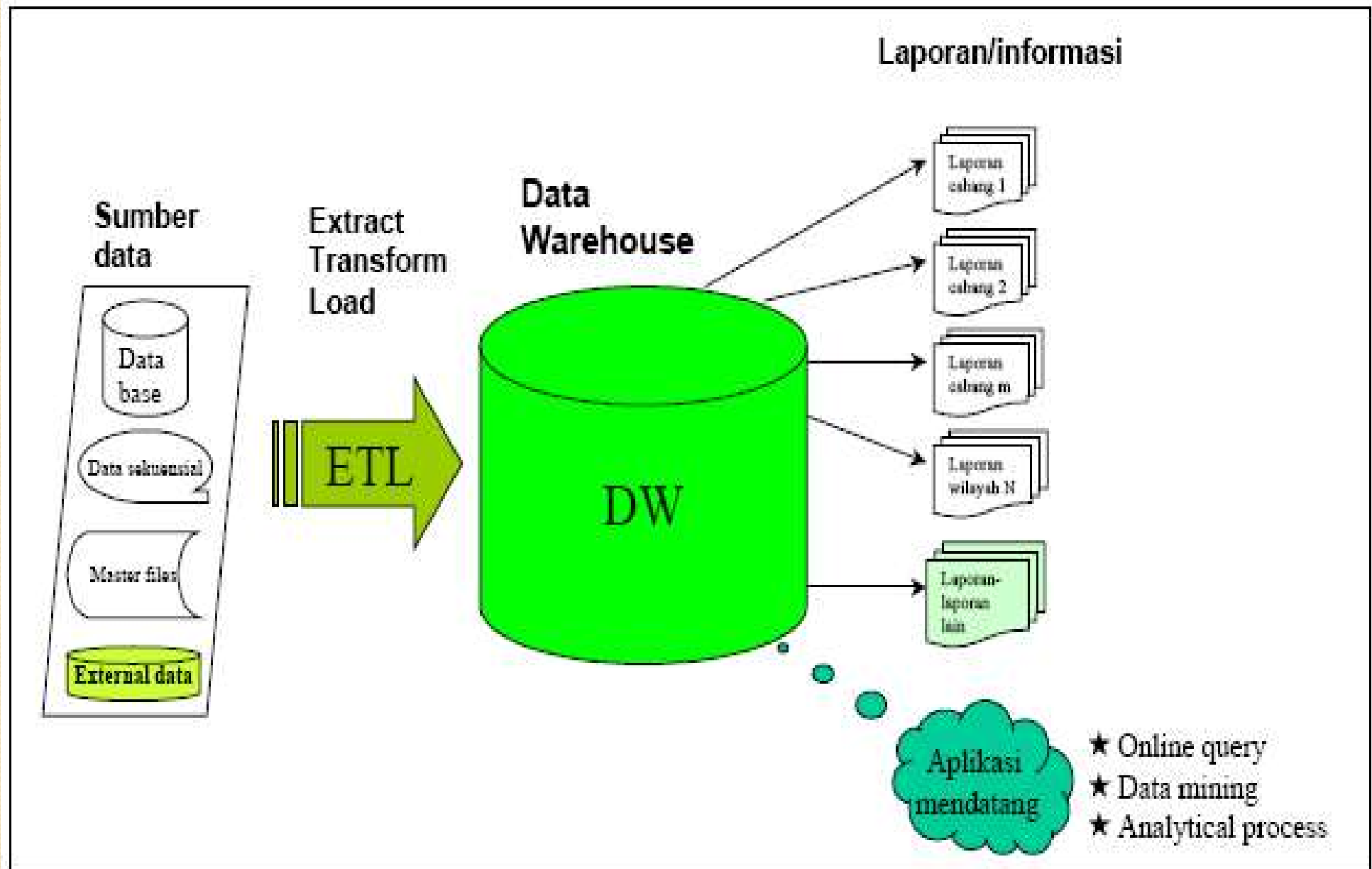
- **Non-volatile**

Data tidak dapat diperbaharui atau dirubah tetapi hanya dapat ditambah dan dilihat.

Konsep data warehouse



Proses Data warehouse



ARSITEKTUR DATA WAREHOUSE

Pilihan berikut harus dibuat didalam perancangan data warehouse

- **Process model**

Tipe apa yang akan dimodelkan?

- **Grain**

Apa dasar data dan level data yang akan disajikan?

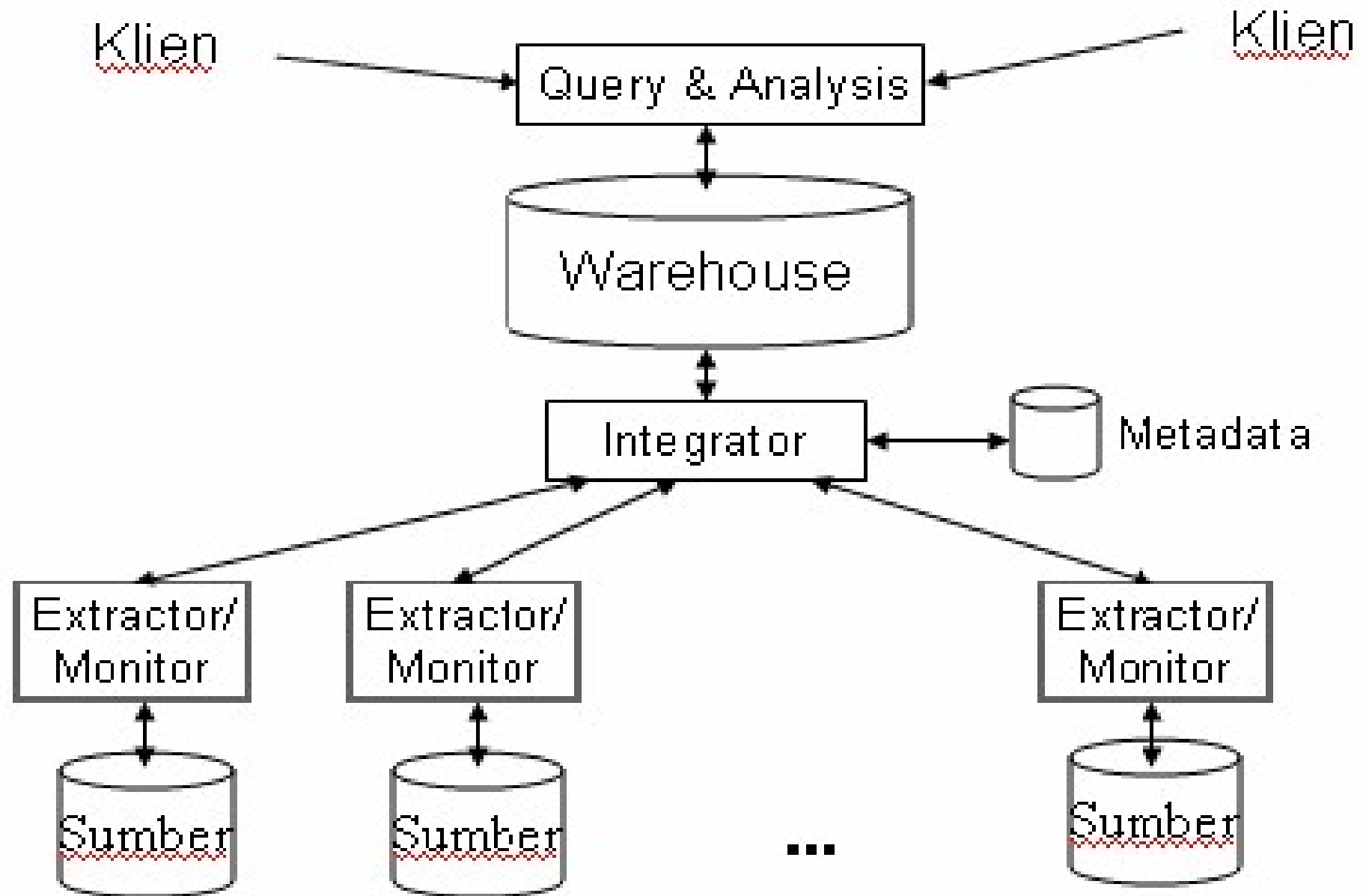
- **Dimensi**

Dimensi apa yang dipakai untuk masing-masing record tabel fakta?

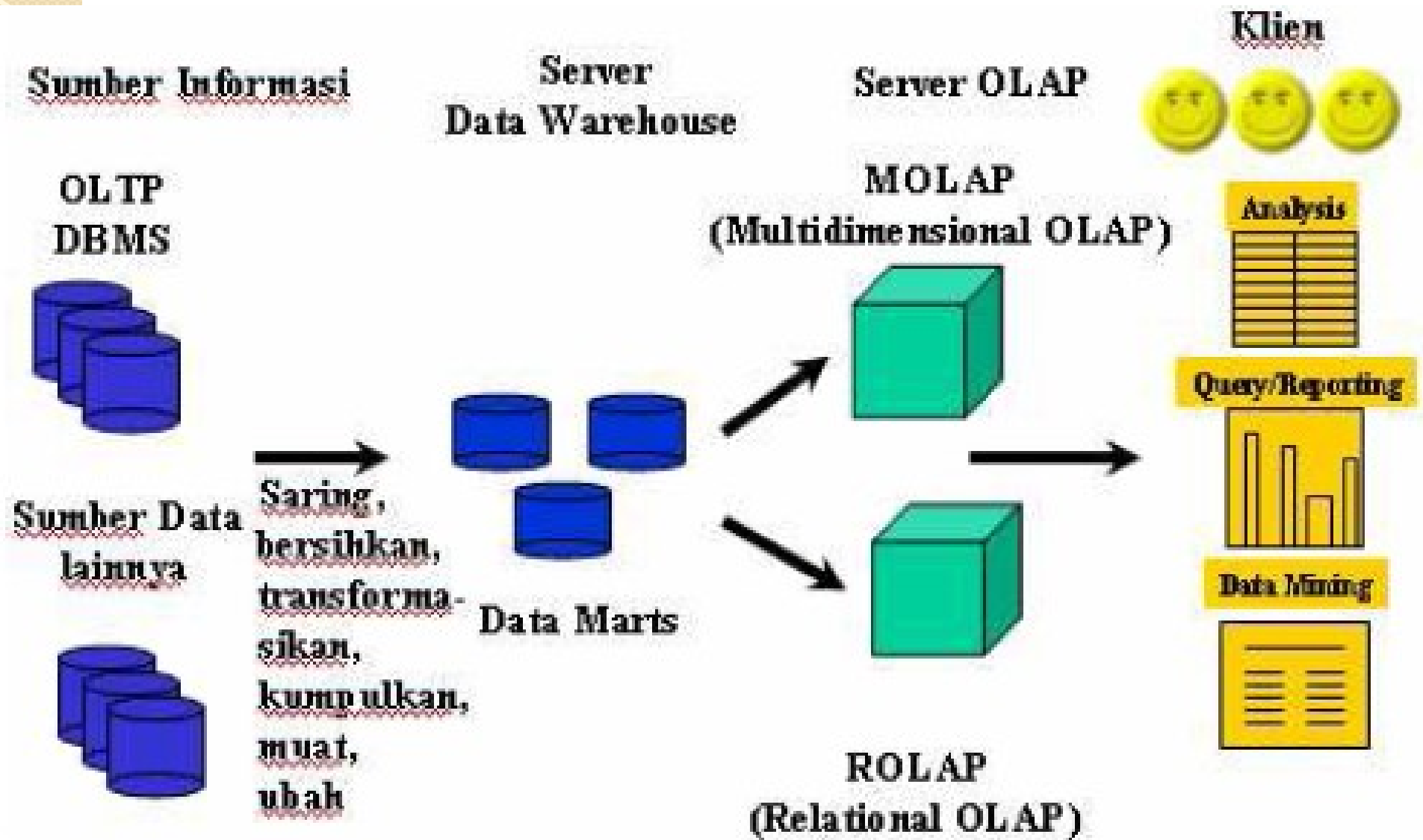
- **Ukuran**

Ukuran apa yang akan mengumpulkan masing-masing record tabel fakta?

ARSITEKTUR DATA WAREHOUSE



ARSITEKTUR DATA WAREHOUSE



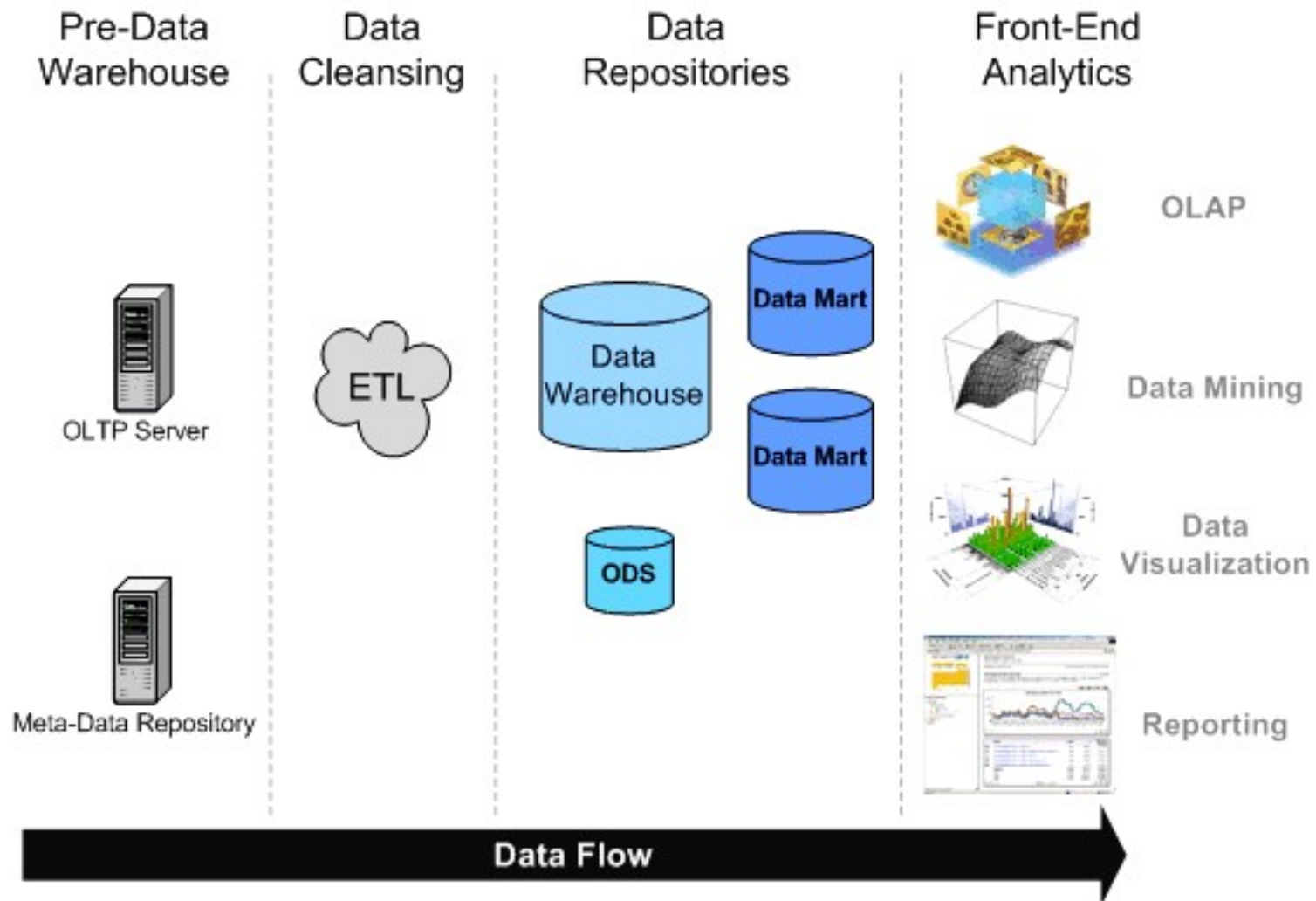
OLAP (On-line Analytical Processing)

- OLAP adalah suatu sistem atau teknologi yang dirancang untuk mendukung proses analisis kompleks dalam rangka mengungkapkan kecenderungan pasar dan faktor-faktor penting dalam bisnis
- OLAP ditandai dengan kemampuannya menaikkan atau menurunkan dimensi data sehingga kita dapat menggali data sampai pada level yang sangat detail dan memperoleh pandangan yang lebih luas mengenai objek yang sedang kita analisis.

OLAP (On-line Analytical Processing)

- OLAP secara khusus memfokuskan pada pembuatan data agar dapat diakses pada saat pendefinisian kembali dimensi.
- OLAP dapat digunakan membuat rangkuman dari multidimensi data yang berbeda, rangkuman baru dan mendapatkan respon secara online, dan memberikan view dua dimensi pada data menggunakan cube multidimensi secara interaktif.

Analisis Multidimensional



OLAP vs OLTP

- ⦿ OLTP adalah singkatan dari On Line Transaction Processing.
- ⦿ OLTP sering kita jumpai di sekitar kita seperti toko atau swalayan contohnya database pada sistem informasi penjualan.
- ⦿ Berbeda dengan OLAP, OLAP adalah On Line Analytical Processing yang maksudnya adalah database yang menyimpan data sekarang dan masa lalu yang berasal dari OLTP.
- ⦿ OLAP bertujuan untuk keperluan analisis dan pelaporan manajemen dalam rangka pengambilan keputusan.

OLAP vs OLTP

- OLAP merupakan kunci dari BI, yang digunakan untuk menganalisis data dan informasi yang pada akhirnya akan menjadi dasar basis *Decision Support System* (DSS) dan *Expert Information System* (EIS).
- Beberapa aktivitas yang dapat dilakukan melalui OLAP antara lain seperti : melakukan *query*, meminta laporan yang *ad hoc* dan mendukung analisis statistik.

OLAP vs OLTP

	OLTP	OLAP
users	clerk, IT professional	knowledge worker
function	day to day operations	decision support
DB design	application-oriented (E-R based)	subject-oriented (star, snowflake)
data	current, up-to-date detailed, flat relational isolated	historical, summarized, multidimensional integrated, consolidated
usage	repetitive	ad-hoc
access	read/write index/hash on prim. key	lots of scans/read mostly
unit of work	short, simple transaction	complex query
# records accessed	tens	millions
#users	thousands	hundreds
DB size	100MB-GB	100GB-TB
metric	transaction throughput	query throughput, response

OLAP vs OLTP

- **User**

Dalam OLTP, penggunaanya adalah IT Professional sedangkan OLAP penggunaanya adalah Knowledge worker maksudnya penggunaanya adalah seorang yang bertindak dalam subyek tertentu, atau petinggi dalam suatu perusahaan.

- **Function**

OLTP digunakan sehari-hari untuk proses bisnis seperti toko atau swalayan, sedangkan OLAP digunakan untuk pengambilan keputusan.

- **Design DB**

Desain dalam OLTP bersifat Entity Relational atau databasenya dinormalisasi dulu sebelum digunakan. Untuk OLAP desain databasenya di de-normalisasi.

OLAP vs OLTP

- **Data**

Dalam OLTP datanya adalah hari ini, update setiap saat sedangkan OLAP datanya adalah sekarang dan hari ini yang berguna untuk melakukan analisis ke depan.

- **Penggunaan**

OLTP digunakan setiap saat, sedangkan OLAP digunakan seperlunya saja.

- **Access**

OLTP aksesnya bisa write, read dan lain-lain. Sedangkan OLAP sering dibaca karena digunakan untuk analisa.

OLAP vs OLTP

- **Unit Pekerjaan**

Kalau OLTP pekerjaannya hanya sederhana misalnya transaksi dalam swalayan. Untuk OLAP query untuk menampilkan data sangat kompleks

- **Jumlah rekaman yang di akses**

Kalau OLTP sekitar ratusan sampai ribuan, tapi jika OLAP data yang diakses bisa sampai jutaan bahkan milyaran.

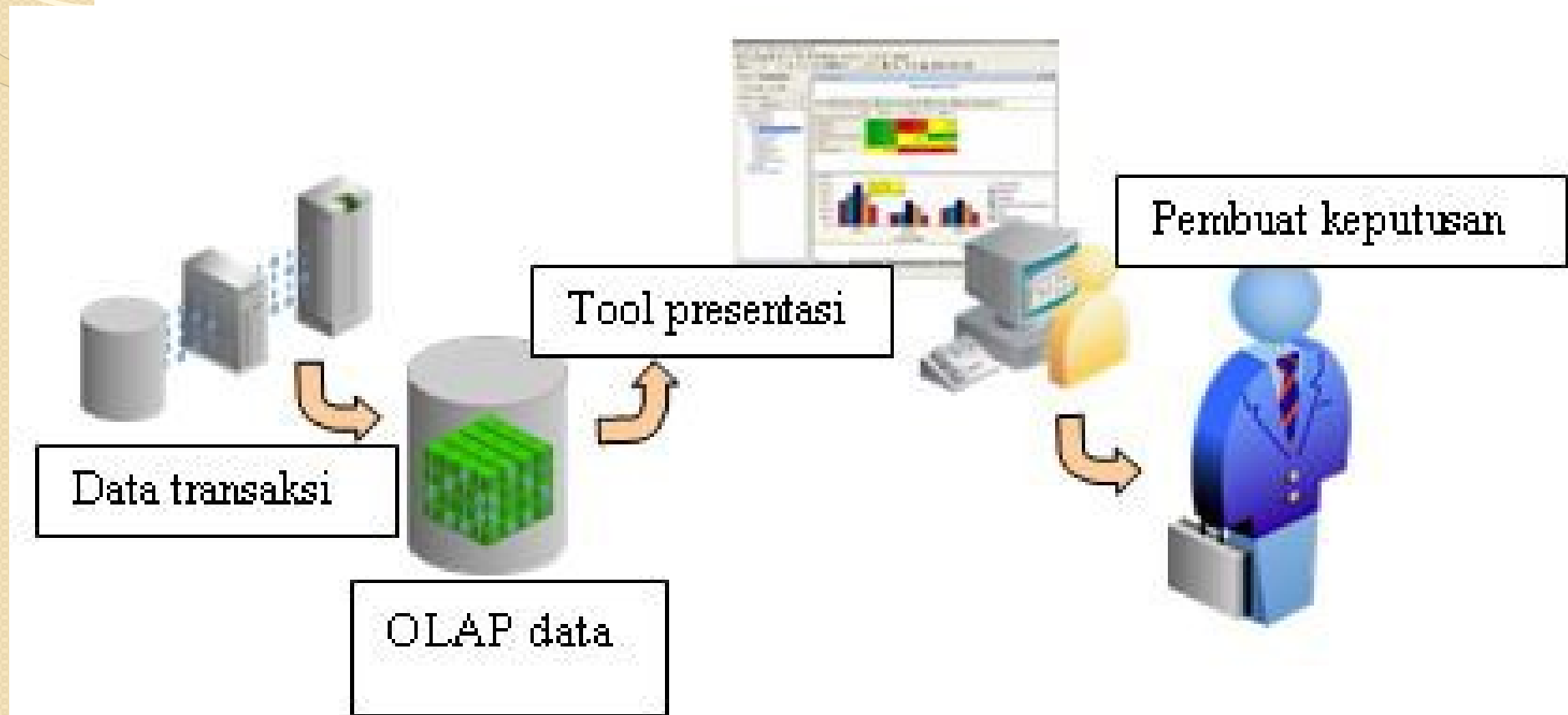
- **Jumlah Pengguna**

Untuk OLTP penggunanya adalah puluhan, tapi kalau OLAP penggunanya bisa sampai ratusan bahkan ribuan

- **Ukuran Database**

Ukuran database untuk OLTP sekitar MB-GB, sedangkan OLAP bisa sampai GB-TB

OLAP vs OLTP



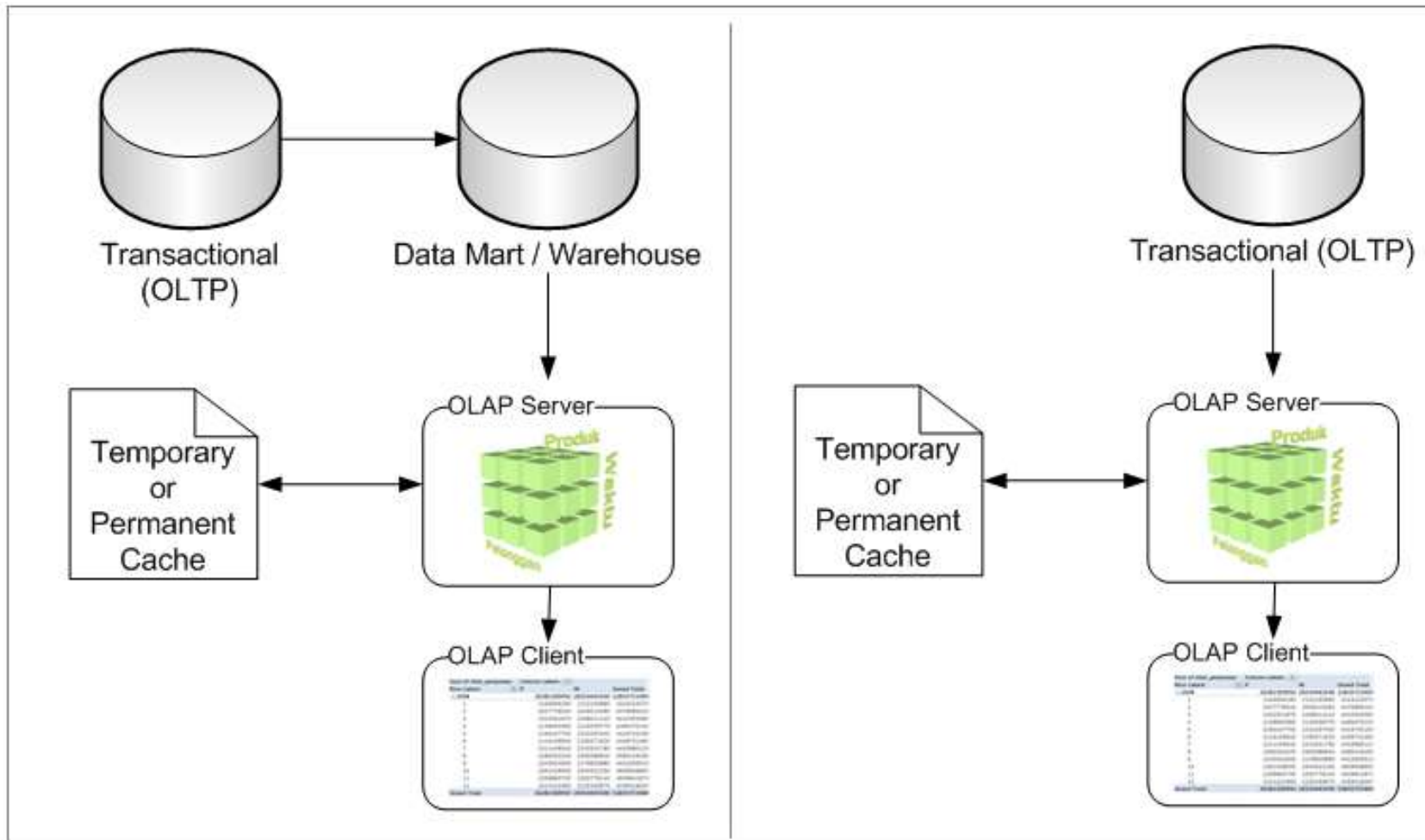
OLAP

- OLAP adalah aplikasi analytical dengan kemampuan pivot menyerupai spreadsheet - seperti Microsoft Excel, OpenOffice Calc, dll. OLAP merupakan komponen penting dari aplikasi BI (Business Intelligence).
- Perbedaan dengan spreadsheet adalah OLAP dirancang khusus untuk mampu menangani jumlah data besar dan memiliki ekspresi bahasa analisis yang lebih baik. Dan aplikasi OLAP ini biasanya memiliki arsitektur client / server.
- OLAP menyediakan cara untuk menampilkan data Multi Dimensional yang ada dalam Data Mart atau Data Warehouse, dengan OLAP dapat dibuat Cube yang mengorganisasikan data dan membuat summary data untuk query yang efisien.

OLAP

- Database OLAP memiliki struktur skema tersendiri dan biasanya berupa suatu data warehouse. Namun tidak tertutup kemungkinan OLAP mengambil dari database operasional (transaksional) - ini dengan catatan database ini telah memiliki struktur rancangan yang "OLAP friendly" (Lihat Gambar).

OLAP



OLAP

- Pada gambar di atas, terlihat OLAP Server menggunakan buffer / cache yang bersifat temporer dan permanen. Ini akan erat kaitannya dengan jenis OLAP yang akan kita bahas. Selain itu, walaupun skema infrastruktur di bagian kanan gambar memungkinkan, ini sangat tidak disarankan dengan alasan performa dan kendala perancangan *cube (pivotal data)* yang lebih kompleks.

Jenis OLAP : MOLAP & ROLAP

- MOLAP (Multidimensional Analytical Processing)
- ROLAP (Relational Analytical Processing)
- Perbedaan keduanya terdapat pada "**Temporary Or Permanent Cache**", dimana :
 - ROLAP menggunakan Temporary Cache (*SQL Result Cache*)
 - MOLAP menggunakan Permanent Cache (*Precomputed Storage*)

ROLAP

- ⦿ Merupakan cara lain yang digunakan untuk melakukan partisi menggunakan tabel relasional dalam data warehouse. Banyak orang berpendapat bahwa basis data yang dirancang secara khusus untuk sebuah keperluan analisis tidak dibutuhkan karena sebuah basis data relasional sudah cukup mampu untuk menampilkan data OLAP. Hal ini hanya berlaku pada tingkat tertentu saja, pada sebuah basis data yang terdiri dari ribuan atau ratusan ribu records maka menampilkan data OLAP akan menjadi sebuah masalah, karena banyak data yang harus diquery. Dan hal ini lah yang menjadi keterbatasan partisi ROLAP.

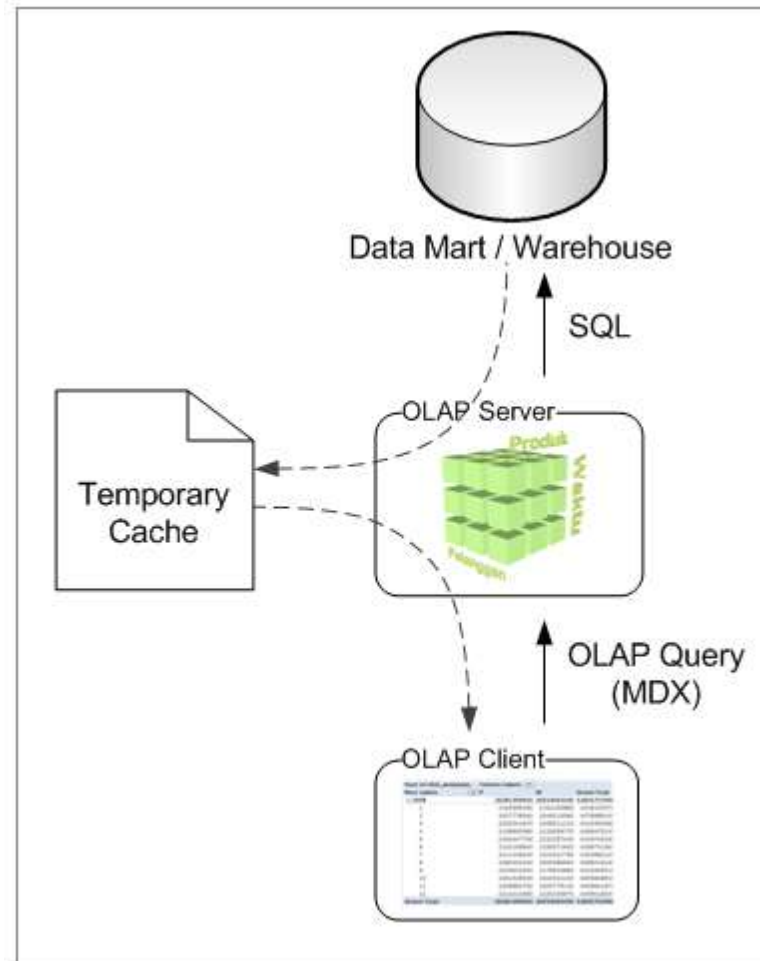
ROLAP

- **Plus (+) :**
 - **Dapat menangani jumlah volume data yang sangat besar**, batasan ukuran volume data yang ditangani pada teknologi ROLAP adalah batas dari volume dari Relational Database yang dipakai.
 - Dengan kata lain pada ROLAP sendiri **tidak ada batasan volume** data.
 - **Dapat memanfaatkan fungsi-fungsi yang ada pada Relational Database** yang dipakai.

ROLAP

- **Minus (-) :**
 - **Performance dapat lambat**, karena setiap ROLAP report pada dasarnya adalah SQL Query pada Relational Database, waktu query dapat lebih lama jika volume data semakin besar.
 - **Fungsi SQL yang terbatas**, karena teknologi ROLAP terutama tergantung pada pembentukan statement Query pada Relational Database, dan tidak semua kebutuhan dapat terpenuhi dengan SQL Statement. ROLAP vendor telah mengantisipasi resiko ini dengan cara membuat Tool out-of-the-box untuk fungsi-fungsi yang kompleks bahkan memungkinkan user untuk mendefinisikan fungsi-fungsi yang dibutuhkannya sendiri.

ROLAP

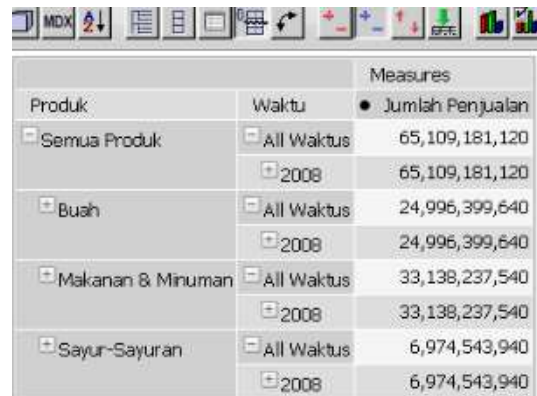


Gambaran Mekanisme kerja ROLAP

Cara Kerja ROLAP

1. OLAP client mengirimkan query analisis ke OLAP Server.
2. OLAP server akan melakukan pemeriksaan di cache apakah sudah bisa melayani permintaan query dari client tersebut, jika sudah akan dikirimkan.
3. Jika pada cache belum terdapat data diminta, akan dilakukan query SQL ke data mart dan hasil eksekusinya disimpan di cache dan dikirimkan kepada client.
4. Demikian seterusnya.
5. Cache akan disimpan selama periode waktu tertentu dan akan dibersihkan total jika server dimatikan.

Contoh ROLAP



The screenshot shows the Mondrian client interface. At the top is a toolbar with icons for MDX, Pivot, and other functions. Below is a table with columns 'Produk' and 'Waktu'. The 'Produk' column has a tree view with categories: Semua Produk, Buah, Makanan & Minuman, and Sayur-Sayuran. The 'Waktu' column has a tree view with 'All Waktus' and '2008'. The 'Measures' column shows 'Jumlah Penjualan' with values ranging from 6,974,543,940 to 65,109,181,120. An arrow points from the text 'Mondrian's ROLAP Pivot View' to this table.

Produk	Waktu	Measures
[-] Semua Produk	All Waktus	65,109,181,120
	2008	65,109,181,120
[-] Buah	All Waktus	24,996,399,640
	2008	24,996,399,640
[-] Makanan & Minuman	All Waktus	33,138,237,540
	2008	33,138,237,540
[-] Sayur-Sayuran	All Waktus	6,974,543,940
	2008	6,974,543,940

Mondrian's
ROLAP
Pivot View

SQL
Generated

```
2010-11-06 10:54:15,437 DEBUG [mondrian.sql] 5:
SqlTupleReader.readTuples [[Produk].[Kategori Produk],
[Waktu].[Tahun]]: executing sql [select
'dim_produk`.`kode_kategori` as `c0`,
'dim_produk`.`nama_kategori` as `c1`, 'dim_waktu`.`year`
as `c2` from `dim_produk` as `dim_produk`,
`fact_penjualan` as `fact_penjualan`, `dim waktu` as
`dim_waktu` where `fact_penjualan`.`sk_produk` =
`dim_produk`.`sk_produk` and `fact_penjualan`.`date_tk` =
`dim_waktu`.`date_tk` and (`dim_produk`.`kode_kategori` =
'KTG-002') group by `dim_produk`.`kode_kategori`,
`dim_produk`.`nama_kategori`, `dim_waktu`.`year` order by
ISNULL(`dim_produk`.`kode_kategori`),
`dim_produk`.`kode_kategori` ASC,
ISNULL(`dim_waktu`.`year`), `dim_waktu`.`year` ASC]
2010-11-06 10:54:16,609 DEBUG [mondrian.sql] 5: , exec
1172 ms
2010-11-06 10:54:16,609 DEBUG [mondrian.sql] 5: ,
exec+fetch 1172 ms, 1 rows
2010-11-06 10:54:16,609 DEBUG [mondrian.sql] 6:
```

Contoh ROLAP

- Gambar di atas adalah contoh tampilan web dari Mondrian / Pentaho Analysis yang merupakan ROLAP (bagian atas gambar). Tiap level data yang dianalisis akan dikonstruksi menjadi SQL yang terlihat pada bagian bawah gambar.

MOLAP

- ⦿ Ini adalah cara tradisional dalam analisis OLAP. Dalam MOLAP data disimpan dalam bentuk Multidimensional Database. Tidak seperti basis data relasional yang menyimpan data dalam jumlah baris dalam tabel, sebuah data multidimensional menyimpan data dalam sejumlah array multidimensional.
- ⦿ Karena dimensi dapat diakses secara mudah, maka user dapat melakukan query pada basis data MOLAP dengan sangat cepat. Selain mengandung data mentah, basis data MOLAP juga mengandung agregasi data sehingga dapat memberikan respon yang cepat terhadap query.

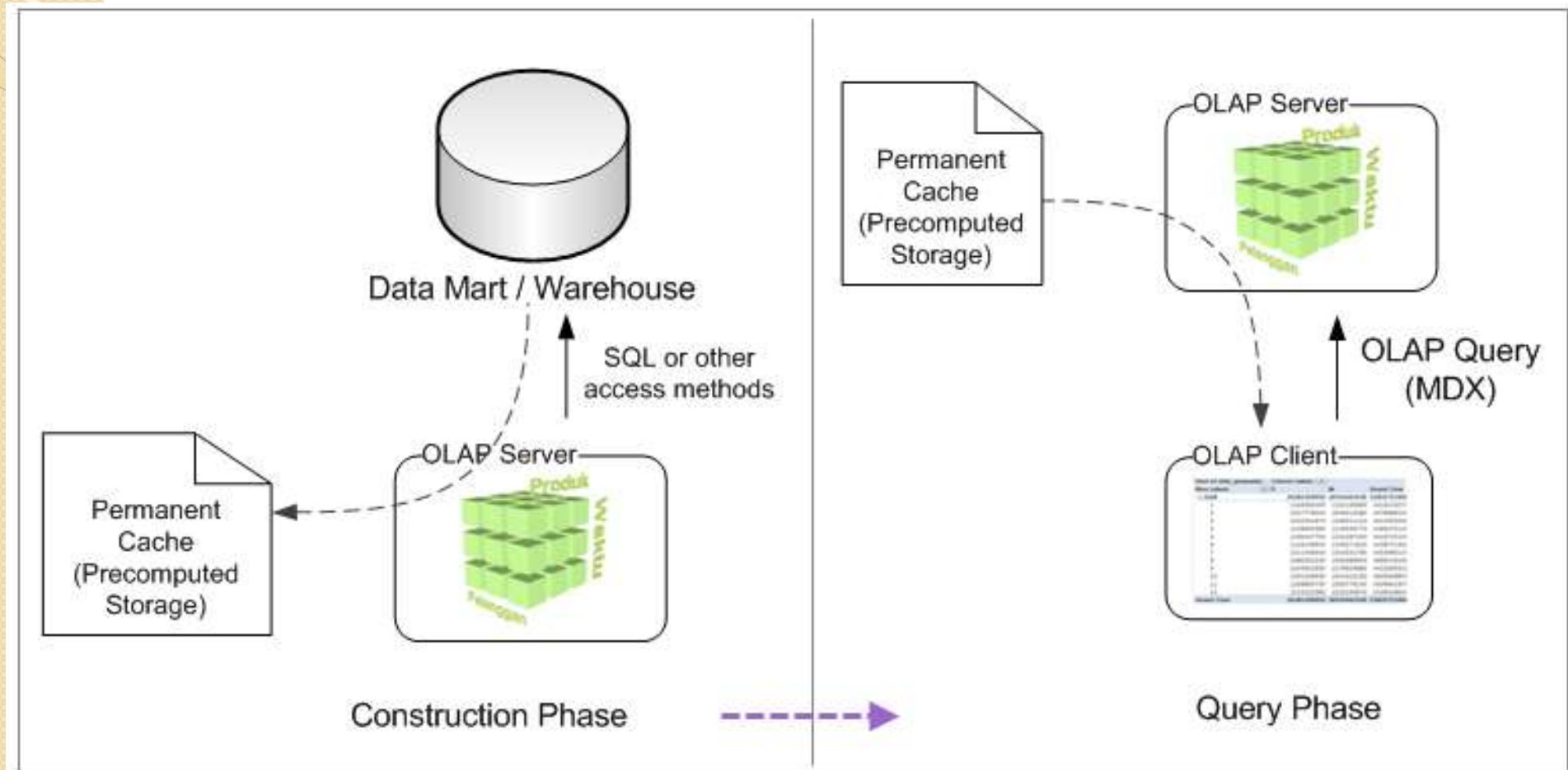
MOLAP

- Plus (+) :
 - Performance hebat, karena MOLAP memang dibangun untuk pengambilan data yang cepat dan optimal.
 - Dapat membentuk kalkulasi yang kompleks dan cepat.

Minus (-) :

- Jumlah volume data yang dapat ditangani terbatas.

MOLAP



Contoh MOLAP

Pivot View

year	(All)		
Sum of nilai_penjualan	Column Labels		
Row Labels	Buah	Makanan & Minuman	Sayur-Sayuran
PHI Mini Market - Jakarta Pusat 01	4,185,785,600.0	10,933,584,370.0	2,311,196,370.0
1	343,117,420.0	864,464,610.0	198,743,460.0
2	328,822,730.0	844,583,760.0	180,198,040.0
3	365,410,500.0	891,692,290.0	188,159,930.0
4	332,996,480.0	882,185,820.0	192,693,610.0
5	343,566,920.0	943,070,230.0	191,832,300.0
6	338,728,810.0	901,441,520.0	190,595,660.0
7	355,625,260.0	950,440,900.0	195,493,590.0
8	362,972,020.0	955,511,000.0	194,771,330.0
9	338,602,030.0	864,655,830.0	186,692,780.0
10	367,134,730.0	947,646,170.0	199,107,360.0
11	355,219,730.0	954,162,790.0	198,481,220.0
12	353,588,970.0	933,729,450.0	194,427,090.0
PHI Mini Market - Makassar 01	4,218,129,850.0	11,074,091,240.0	2,338,000,790.0
1	338,787,820.0	910,317,310.0	188,661,540.0
2	333,404,720.0	854,111,630.0	181,212,290.0
3	347,764,310.0	917,548,590.0	192,356,560.0
4	329,174,770.0	888,223,930.0	196,671,510.0

Precomputed Fragment Tables

Nama Kategori	Total Nilai Penjualan
Buah	12,722,972,010.0
Makanan & Minuman	33,138,237,540.0
Sayur-Sayuran	6,974,543,940.0

Total Nilai Penjualan
52,835,753,490

Cabang	Total Nilai Penjualan
Jakarta Pusat	17,430,566,340
Makassar	17,630,221,880
Surabaya	17,774,965,270

Tahun	Bulan	Cabang	Total Nilai Penjualan
2008	1	Jakarta Pusat	1,406,325,490
2008	1	Makassar	1,437,766,670
2008	1	Surabaya	1,472,060,910
2008	2	Jakarta Pusat	1,353,604,530
2008	2	Makassar	1,368,728,640
2008	2	Surabaya	1,354,655,930
2008	3	Jakarta Pusat	1,445,262,720
2008	3	Makassar	1,457,669,460
2008	3	Surabaya	1,507,420,900

...

Tugas

1. Jelaskan keterkaitan antara datawarehouse dan bisnis cerdas?
2. Apa yang dimaksud dengan Data Mart?
3. Jelaskan tentang Enterprise Data Warehouses (EDW)?
4. Cari dan buat resume jurnal yg berhubungan dengan penerapan datawarehouse!
5. Penerapan Data Warehouse bisa menggunakan Kettle (open source ETL). Lakukan proses instalasi dengan menggunakan Kettle tersebut!

Daftar Pustaka

- Djoni Darmawikarta, Mengenal Data Warehouse, 2003
- Yudho Giri Sucahyo, Data Mining, 2003
- Yudho Giri Sucahyo, Penerapan Data Mining, 2003
- Jeffrey A. Hoffer, Mary B. Prescott, Fred R. McFadden ; Modern Database Management
8th Edition; 2007