



# Data Warehouse

---

Sirojul Munir S.SI, M.KOM – Semester Genap TA 20182

# Pengantar Data Warehouse

---

# Data, Informasi & Knowledge

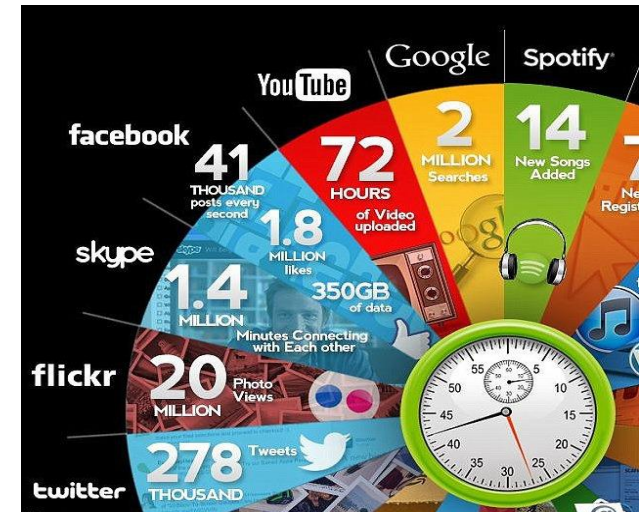
## Definisi Data:

- ✿ **Navathe dan Elmasri, 2000:** Data yaitu fakta yang dapat disimpan dan memiliki arti
- ✿ **Hoffer, Prescott, dan McFadden, 2005:** Data yaitu sesuatu yang mewakilkan objek dan peristiwa yang memiliki arti dan sangat penting bagi pemakai atau user
- ✿ **Data:** fakta, teks, hasil pengukuran, gambar, suara, dan video yang bernilai informasi.

# Data, Informasi & Knowledge

## Jenis Data:

- **Structured data**
  - Tipe data yang dapat disimpan di database atau spreadsheet, diperlukan untuk dikelola sesuai dengan format penyimpanan standar dan ontologi, seperti : nama, alamat, telpon,
  - Contoh : Aplikasi sistem informasi akademik, aplikasi work flow, aplikasi SDM dll
  - Solusi Kelola: **DBMS**
- **Unstructured data**
  - text, audio, imagery, video
  - Contoh : data sistem email siswa, chat rooms, hasil questioner, video / audio di sistem e-elearning , RFID , barcode
  - Solusi Kelola: **BIGDATA** (4V: Volume, Velocity, Varieaty, Veracity)



# Data, Informasi & Knowledge

## Contoh data terstruktur:

Tabel Data Calon Ketua Kelas

No.	Nama	Turus	Banyak Pemilih
1.	Andi	NI NI	10
2.	Ika	NI NI NI I	16
3.	Santi	NI III	8
4.	Rudi	NI I	6
Jumlah			40

Nama Toko Anda  
Jl. Abcdefghijkl No. 1234567890  
Telp. 123456789 Kota Anda  
Sedia berbagai kebutuhan sehari-hari

Pelanggan : PLG0003  
: Ida Ayu Setia  
: Jl. Soekarno Hatta No 18  
: 08000000000

No. Transaksi : PJL20130827-0003-MAS-192944  
Kasir : Master  
Tanggal : 27-08-2013 19:29:44

P. Nama	Qty	Harga	Total
T Eskulin Gel 100ml Drea	5	6.468	32.340
Y Eskulin Gel 100ml Mond	6	6.600	39.600

Total--- 2 ---item(s). Rp. 71.940,-  
Pembulatan Rp. 60,-  
GrandTotal Rp. 72.000,-  
Tunai Rp. 80.000,-  
Kembali Rp. 8.000,-

Belanja produk trntu mndptkan 1 poin tiap klptn 100rb  
Dapatkan poin sebanyak-banyaknya  
Tukarkan poin dengan hadiah yang keren!  
Barang yg sudah dibeli tdk dapat ditukar/dikembalikan

# Data, Informasi & Knowledge

## Contoh data tidak terstruktur:

**12+ TBs**  
of tweet data  
every day

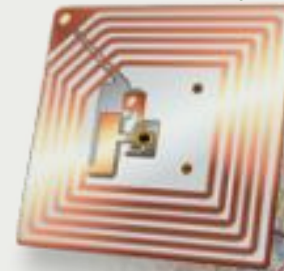


? TBs of  
data every  
day



**25+ TBs** of  
log data  
every day

**30 billion** RFID tags  
today  
(1.3B in 2005)



**76 million** smart meters in 2009...  
200M by 2014

**4.6 billion**  
camera  
phones  
world  
wide



**100s of millions of**  
**GPS**  
**enabled**  
devices  
sold  
annually



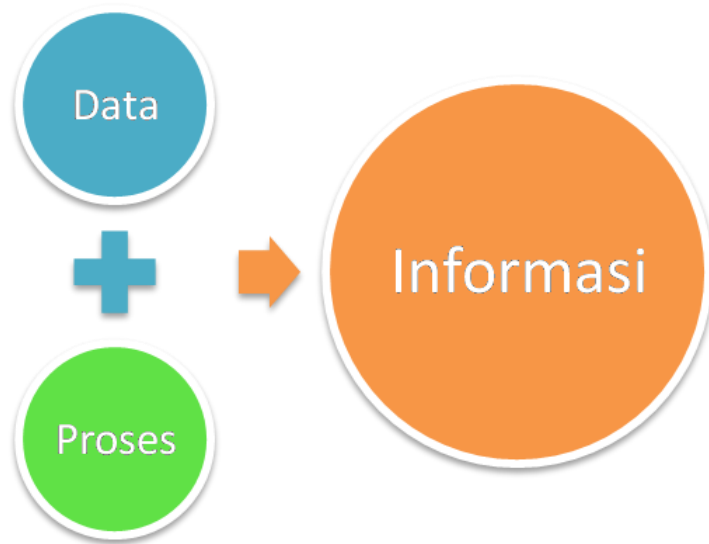
**2+ billion**  
people on  
the Web  
by end  
2011





# Data, **Informasi** & Knowledge

## Definisi Informasi:

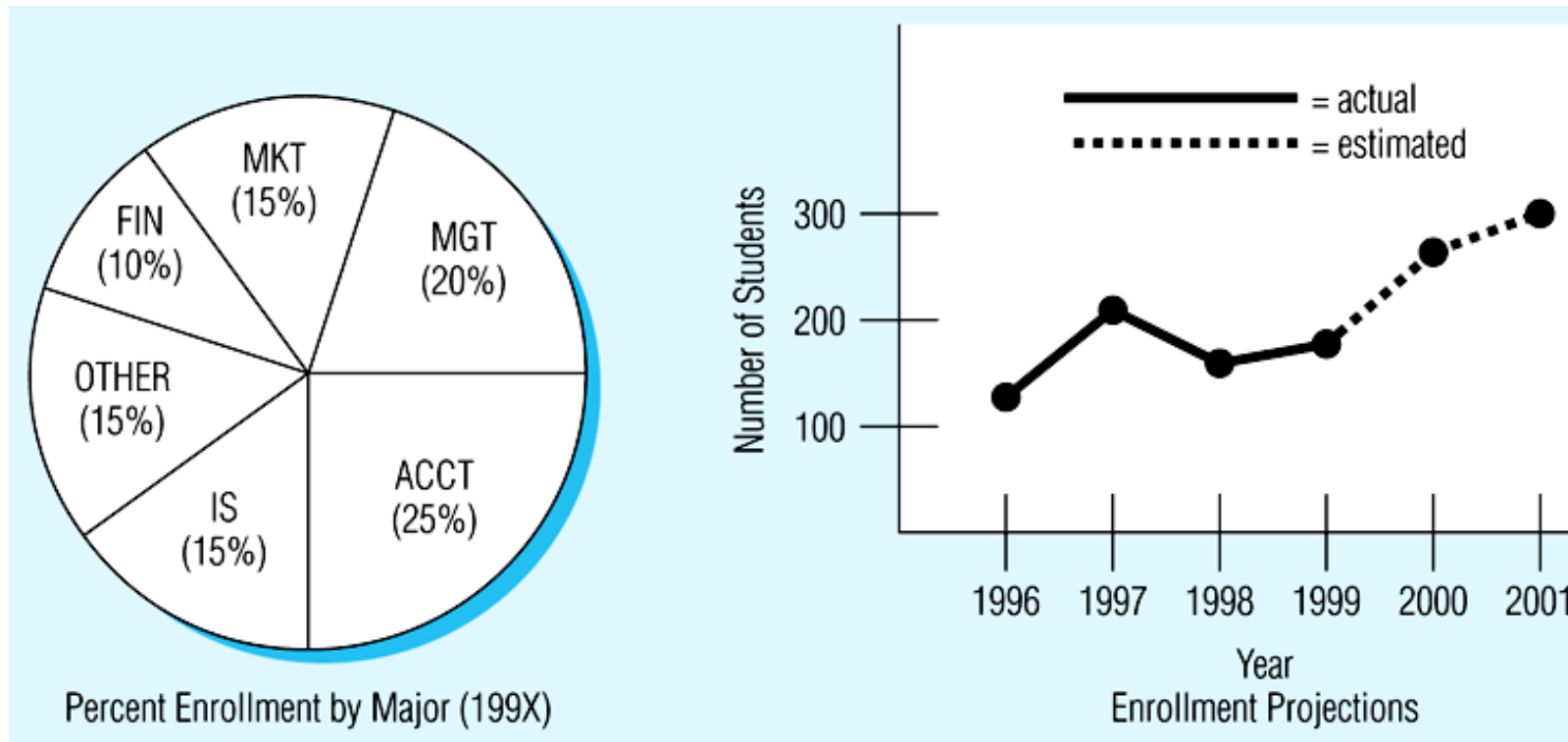


Gambar Sistem Pengolahan Data

- ❁ **Informasi:** data yang telah diproses sebagai bahan dalam proses pengambilan keputusan.

# Data, **Informasi** & Knowledge

Informasi - dapat dimanfaatkan sebagai dasar untuk pengambilan keputusan dan memahami permasalahan/situasi



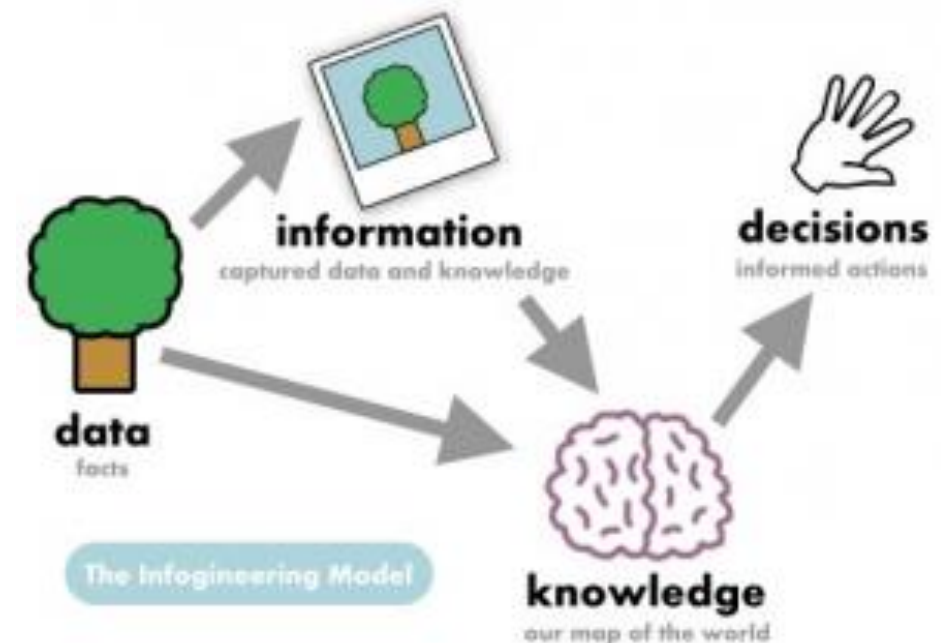


# Data, Informasi & Knowledge

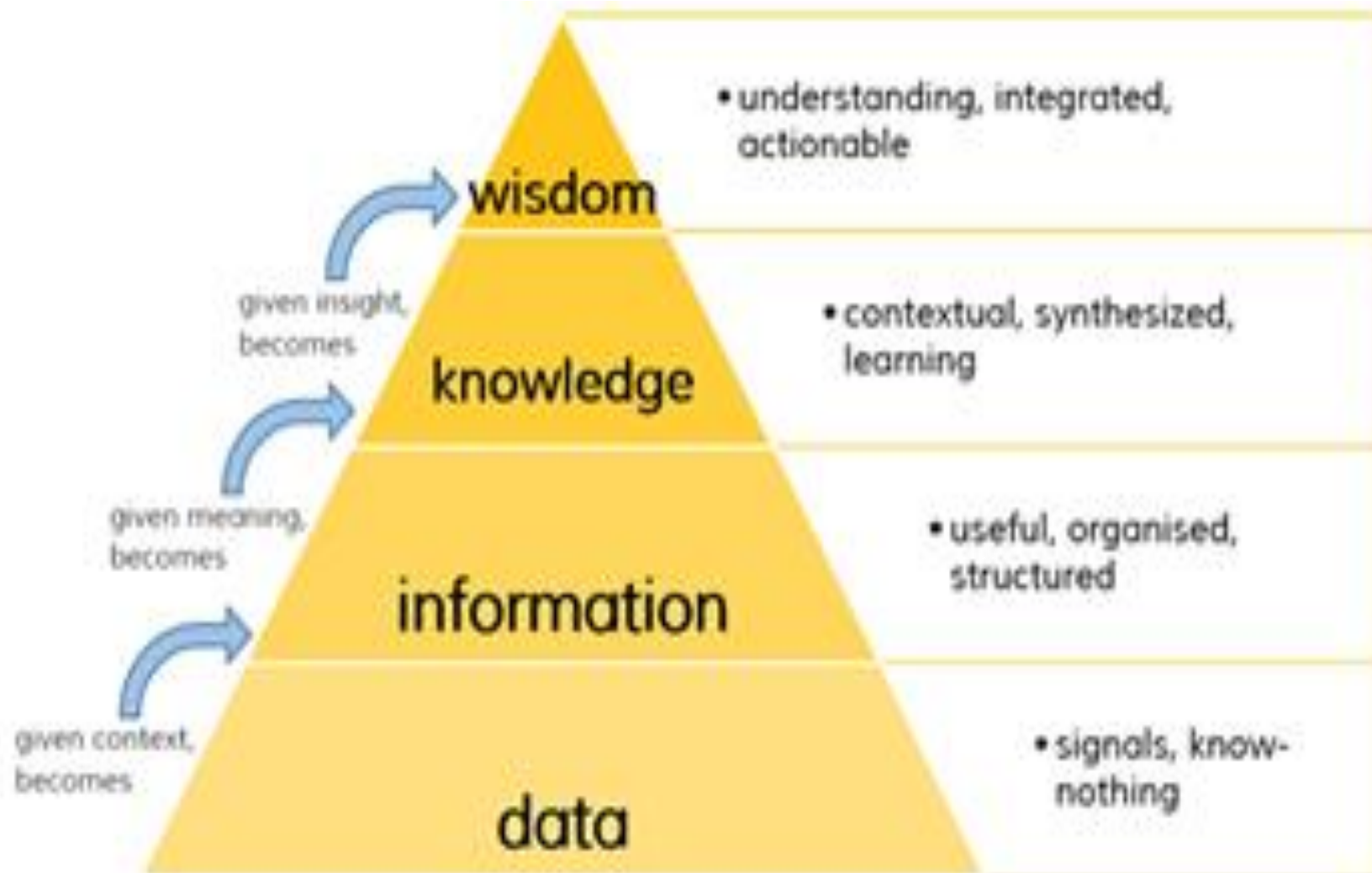
- Knowledge: adalah informasi yang dilengkapi dengan pemahaman pola hubungan dari informasi disertai pengalaman, baik individu maupun kelompok dalam organisasi.

**Fungsi Informasi:**  $I = i(D, S, T)$

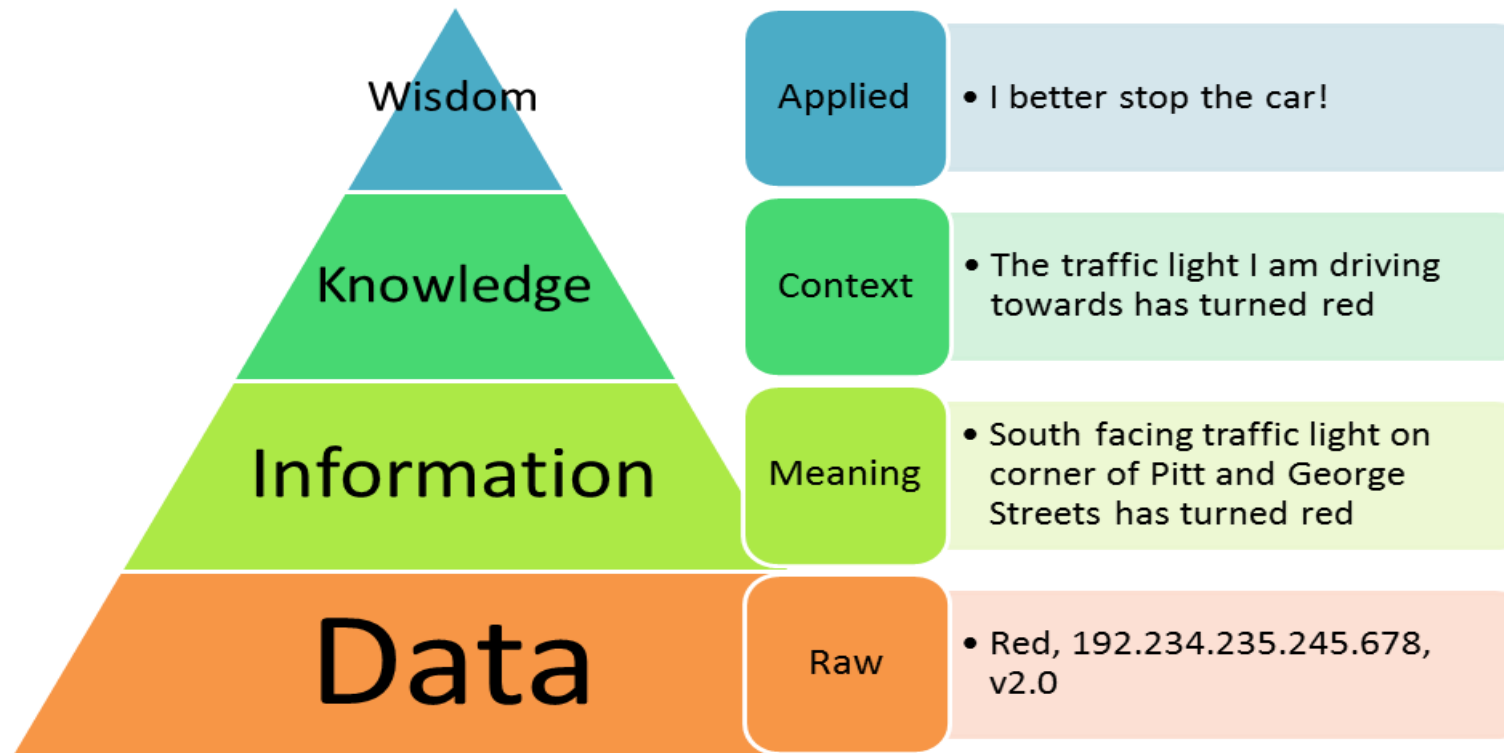
- I** : Informasi
- D**: Data
- S** : Pengetahuan awal
- T** : Waktu



# Piramida : Data, Informasi, Knowledge, Wisdom



# Piramida : Data, Informasi, Knowledge, Wisdom

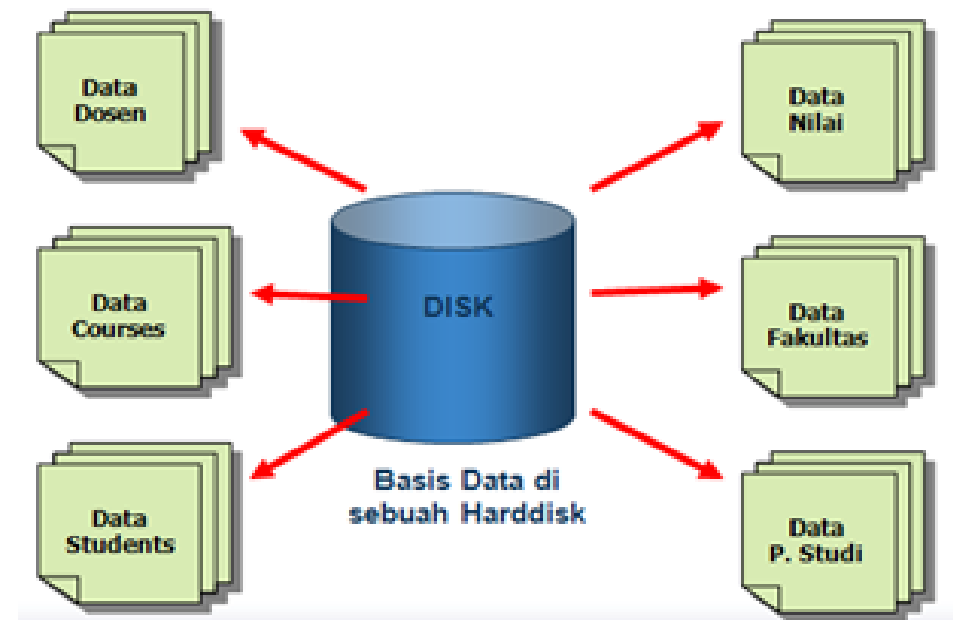
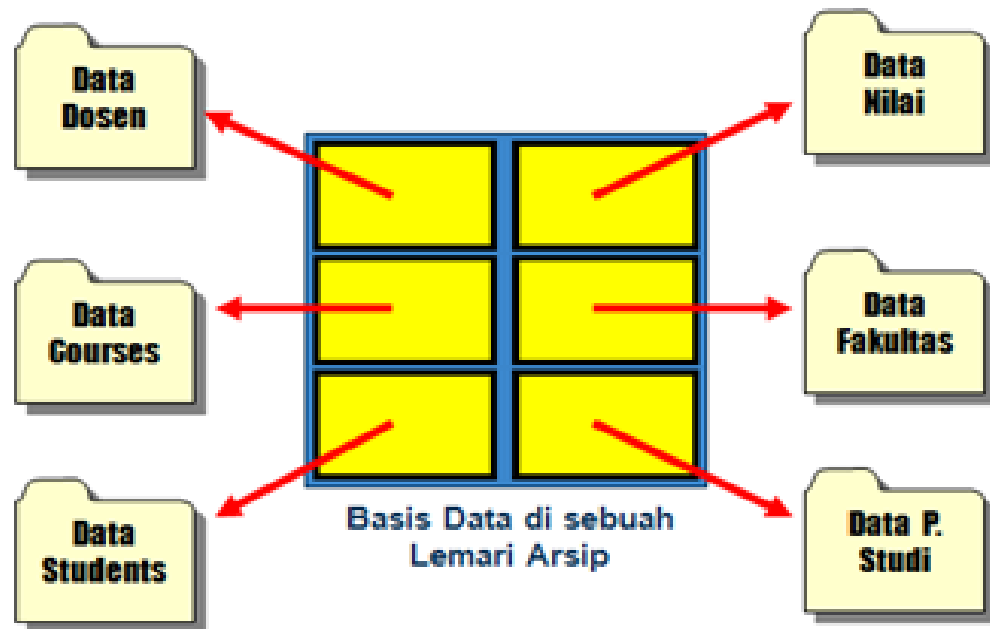


Q?:

Berikan contoh keterkaitan data, informasi, knowledge dan wisdom. Dalam kehidupan sehari2 disekitar anda !

# Basis Data

- ✦ **Basis Data (Database):** kumpulan data yang terorganisir berdasarkan suatu struktur hubungan (konsep entitas).



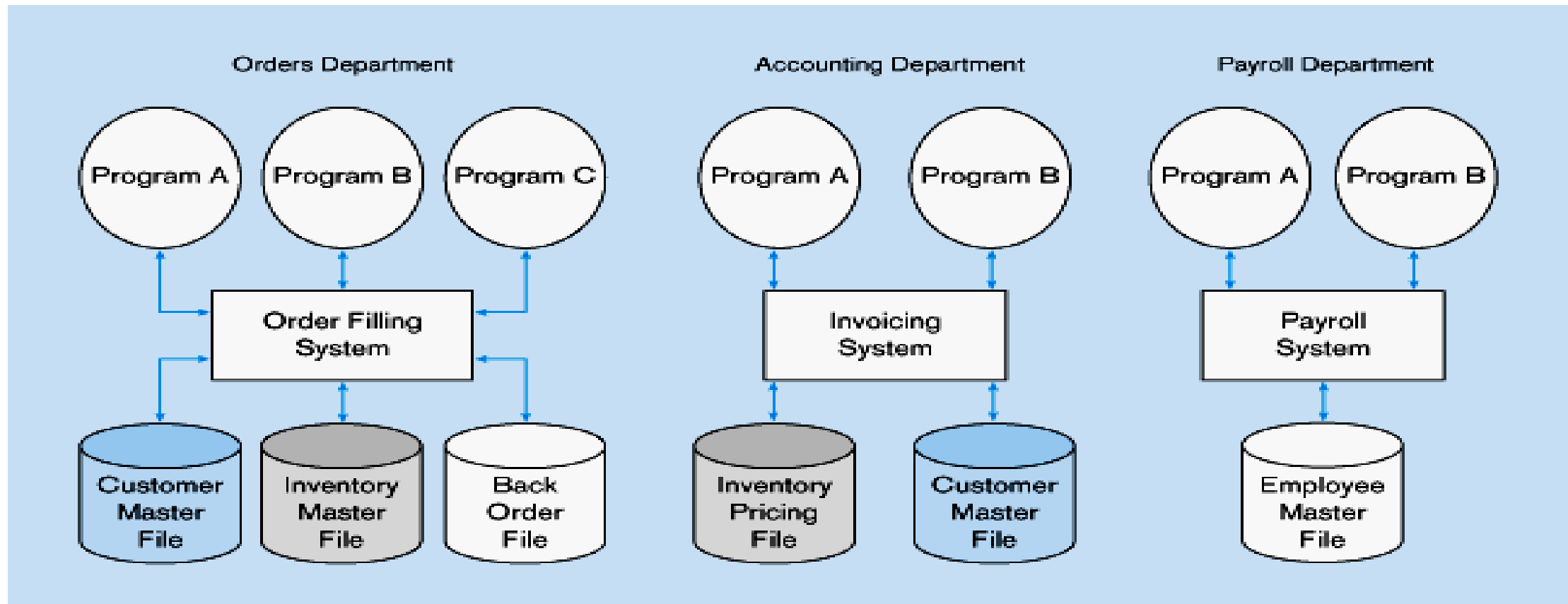
# Metadata

Deskripsi tentang format dan karakteristik data, termasuk tipenya, ukurannya, nilai-nilai yang absah, dan dokumentasi lainnya.

<i>Data Item</i>			<i>Value</i>		
<b>Name</b>	<b>Type</b>	<b>Length</b>	<b>Min</b>	<b>Max</b>	<b>Description</b>
Course	Alphanumeric	30			Course ID and name
Section	Integer	1	1	9	Section number
Semester	Alphanumeric	10			Semester and year
Name	Alphanumeric	30			Student name
ID	Integer	9			Student ID (SSN)
Major	Alphanumeric	4			Student major
GPA	Decimal	3	0.0	4.0	Student grade point average

# Sistem Berbasis File

Program-program aplikasi menyimpan data masing-masing dalam file



# Kelemahan Sistem Berbasis File

- **Dependensi antara program & data**
  - Fungsi penyimpanan dan akses data merupakan bagian tak terpisahkan dari setiap program aplikasi.
- **Redundansi (duplikasi) data**
  - Masing-masing aplikasi/program memiliki kopi sendiri untuk data yang sama.
- **Inkonsistensi data**
  - Tidak ada pengawasan data secara terpusat.
- **Sulit berbagi (*sharing*) data**
  - Format file antar program aplikasi dapat berbeda-beda.
- **Biaya pemeliharaan tinggi**



# Permasalahan Dependensi Data

- ❑ Setiap program aplikasi harus **memiliki data** sendiri – memungkinkan duplikasi data.
- ❑ Setiap program aplikasi harus berurusan dengan **metadata** (format data) untuk setiap file yang digunakan.
- ❑ Setiap program aplikasi harus memiliki **fungsi** untuk membaca, menulis, mengubah dan menghapus data.
- ❑ Tidak ada **koordinasi pengelolaan data** yang sama antar aplikasi.
- ❑ Sulit **membakukan format-format** file.

# Solusi BASIS DATA

## ***Database***

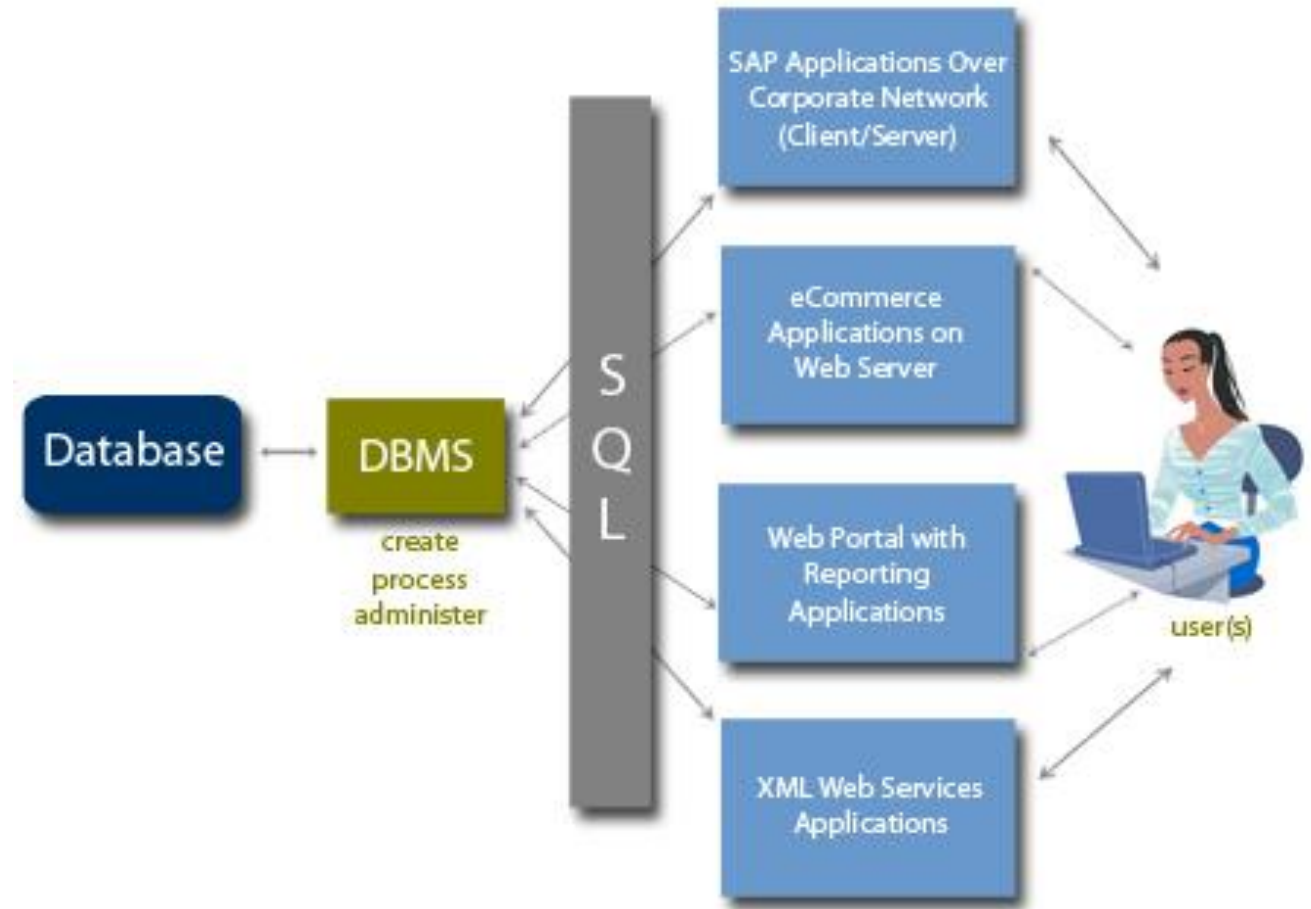
- Pusat repositori data bagi seluruh organisasi.
- Data dikelola oleh suatu sistem pengendali.
- Data disimpan dalam format yang baku dan mudah dimengerti.



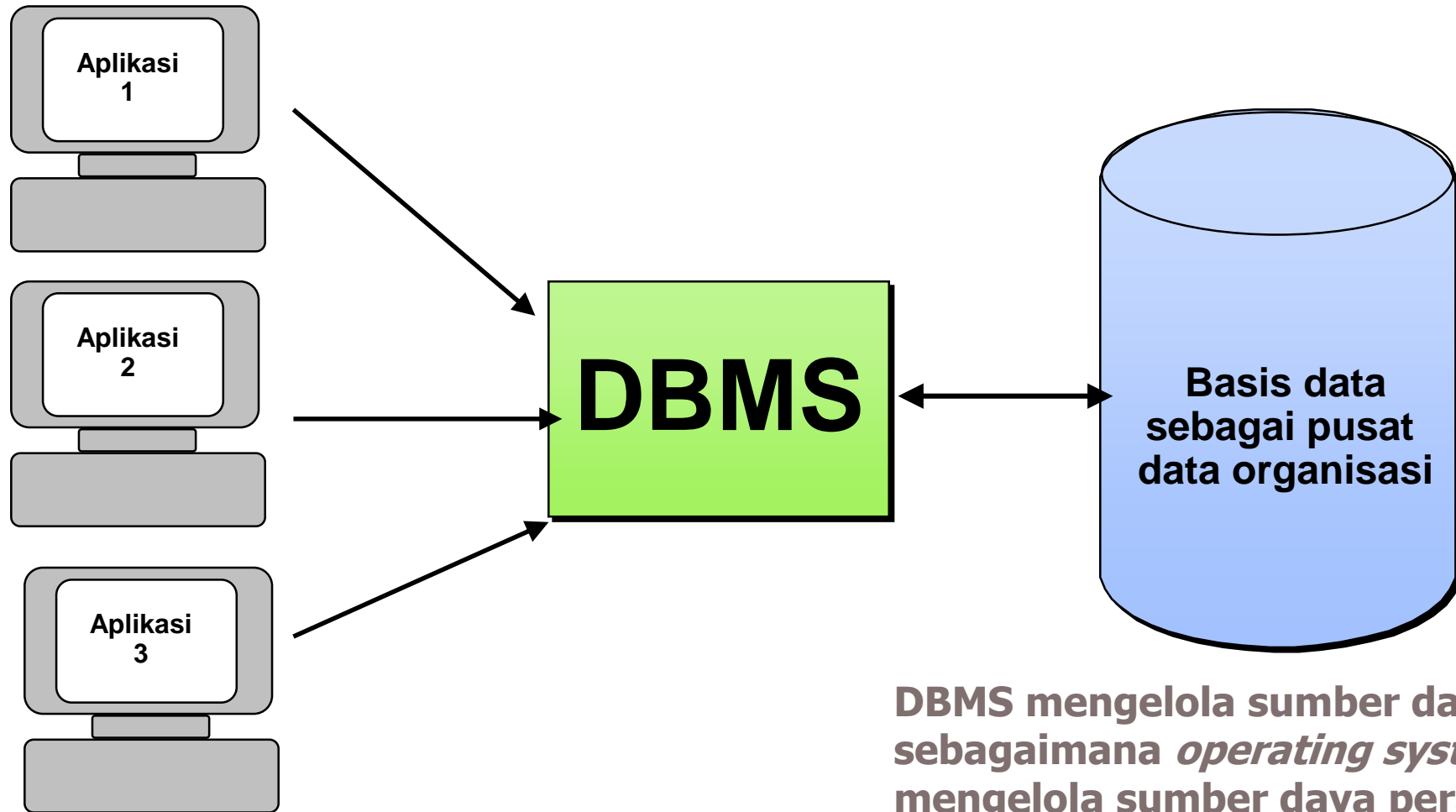
Membutuhan suatu *Database Management System* (DBMS).

# Database ?

*Database Management System (DBMS)* adalah sistem penyimpanan dan pengambilan data yang memungkinkan data untuk disimpan secara **tidak terduplikasi** dan mengorganisasikan data dengan **struktur** yang dikehendaki pengguna.

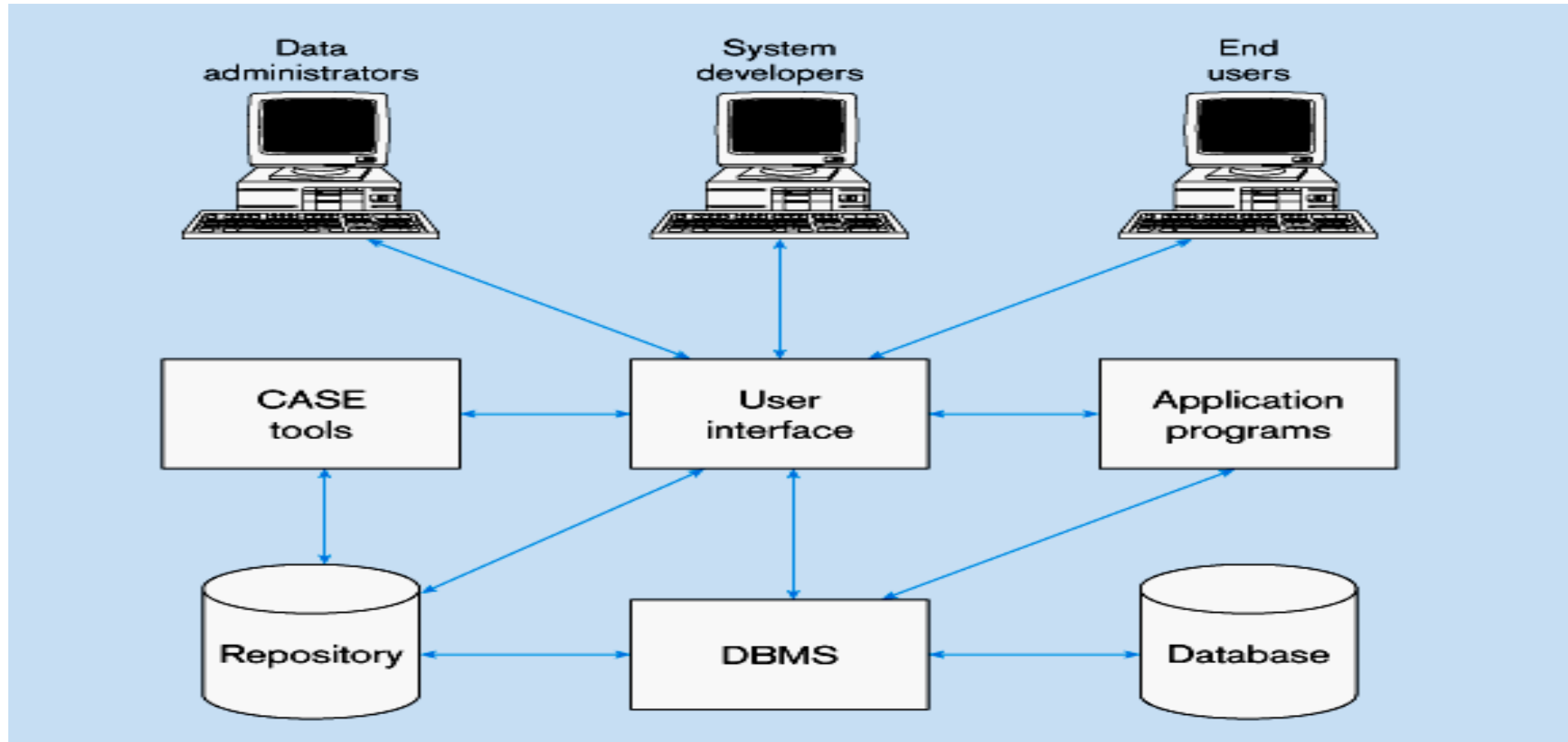


# DBMS



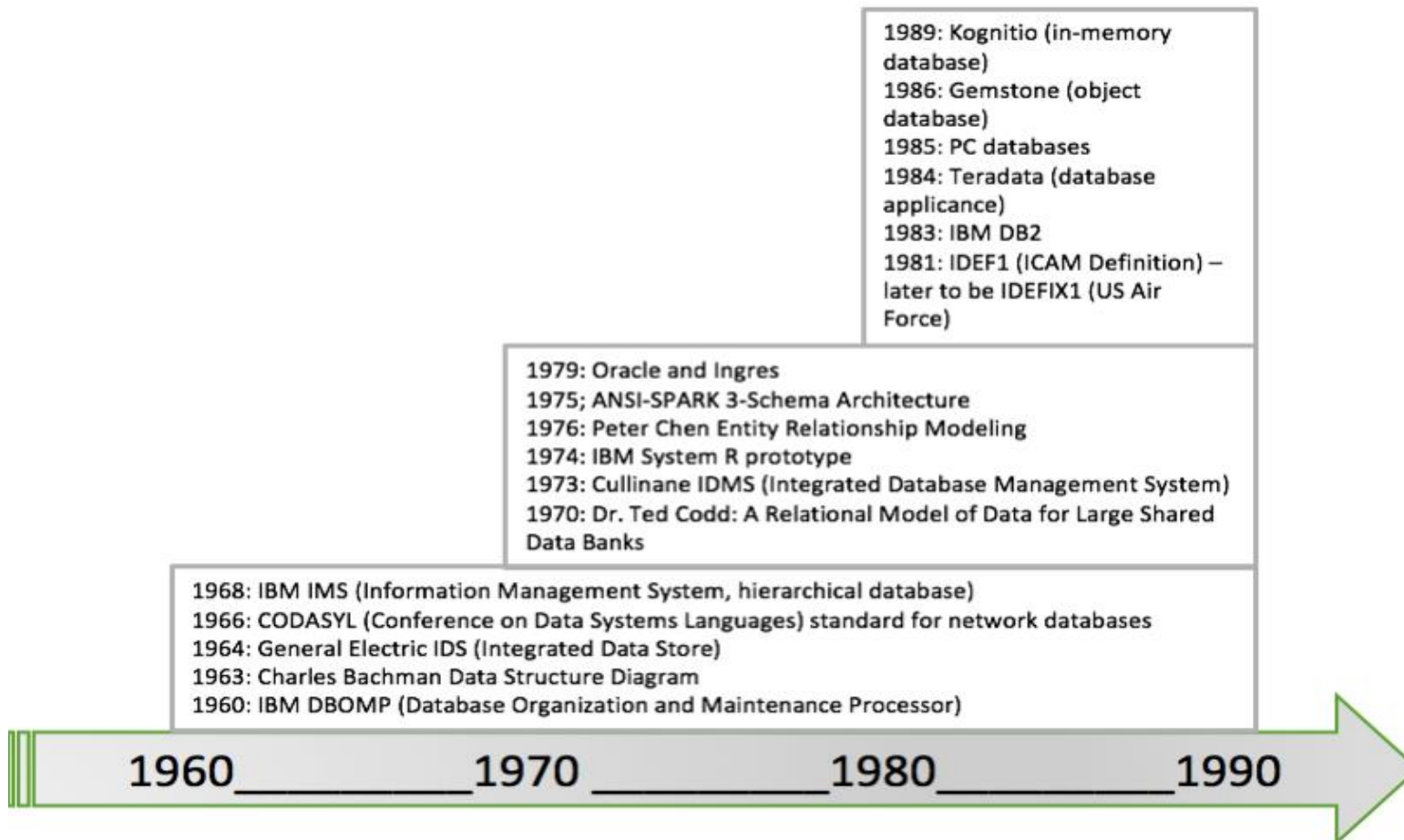
DBMS mengelola sumber daya data  
sebagaimana *operating system* (OS)  
mengelola sumber daya perangkat keras.

# Komponen Basis Data



# Perkembangan Database

## The Pioneers of the DBMS Trail

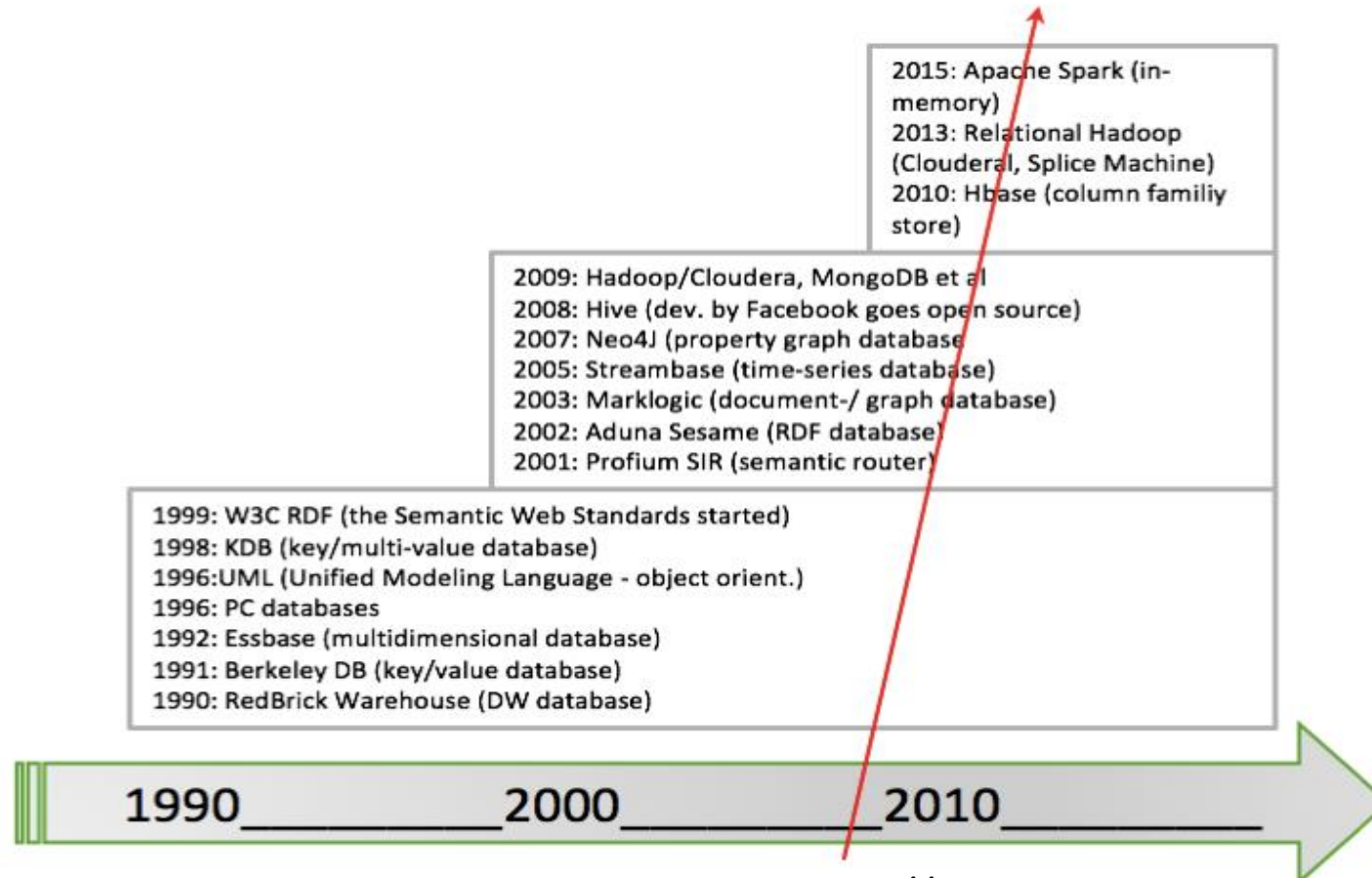


<http://graphdatamodeling.com/GraphDataModeling/History.html>

# Perkembangan Database

## The Relational Empire

## NoSQL



<http://graphdatamodeling.com/GraphDataModeling/History.html>



# Database Relational



## ✦ **EF Codd (1970):**

*Memperkenalkan **model relasional** data, pada saat itu sebagian besar sistem database berdasarkan dua model data: **Model hirarki** (hierarchical model) dan **model jaringan** (network model).  
Prototipe sistem database model relasional dikembangkan di IBM dan di UC-Berkeley pada pertengahan tahun 1974*

*Model Relasional telah digunakan sebagaian besar sistem database saat ini*

# Relational Model

## ✚ Simple & Elegan

- ✚ Database adalah kumpulan dari satu atau lebih dari relasi, dimana setiap relasi adalah berupa tabel, kolom dan baris

## ✚ Keuntungan

- ✚ Tampilan data berbentuk tabular mudah dimengerti
- ✚ Kemudahan tampilan data walaupun dengan query yang rumit

# Tabel

column / field

row / record

No ▲	NIM ◆	Nama ◆	Prodi ◆	Thn Angkatan ◆	IPK ◆	Predikat ◆
1	02011	Faiz Fikri	2012	TI	3.8	Cum Laude
2	02012	Alissa Khairunnisa	2012	TI	3.9	Cum Laude
3	01011	Rosalie Naurah	2010	SI	3.46	Memuaskan
4	01012	Defghi Muhammad	2010	SI	3.2	Memuaskan

- **Field/Column** – Satu jenis informasi/data yang Mempunyai Tipe Data Sama
- **Record/Row** – Satu kesatuan informasi yang terdiri atas satu Field atau lebih
- **Character** – Satuan terkecil dari data

# SQL - RDBMS

- **SQL (Structure Query Language)** - adalah suatu bahasa query yang dikembangkan oleh IBM pada proyek DBMS relasional system -R.
- Selama beberapa tahun SQL menjadi bahasa query yang digunakan untuk melakukan pendefinisian skema, manipulasi data dan query pada DBMS relasional.

# Standarisasi SQL #1

- ✚ Banyak digunakan vendor pada produk RDBMS
- ✚ Dibutuhkan standar penulisan SQL.
- ✚ Produk RDBMS diharapkan mengacu pada suatu standar yang ditetapkan.
- ✚ Diharapkan tidak ada perbedaan perintah SQL pada produk yang dikeluarkan.

# Standarisasi SQL #2

## ❖ ***American National Standards Institute (ANSI) mengeluarkan Standard SQL***

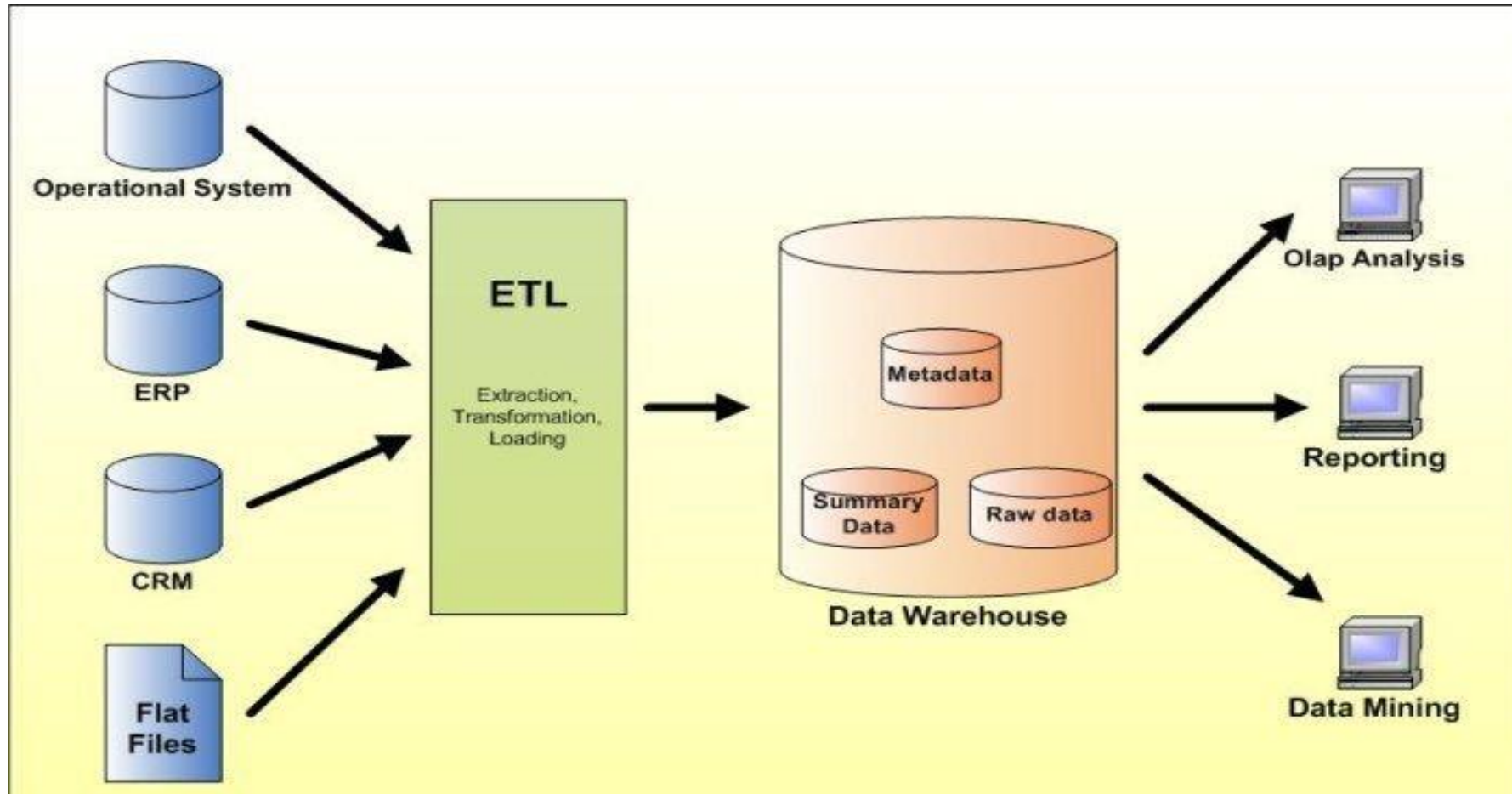
- ❖ 1986: SQL-86
- ❖ 1989: SQL-89 ; perubahan minor
- ❖ 1992: SQL-92 ; ANSI+ISO – perubahan besar
- ❖ 1999: SQL-99
- ❖ 2003: SQL-2003

# Perintah SQL

- ❖ **Data Definition Language (DDL)** – Perintah SQL untuk mendefinisikan skema database atau table : CREATE, DROP, ALTER
- ❖ **Data Manipulation Language (DML)** – Perintah SQL untuk manipulasi data dan pencarian data : INSERT, UPDATE, DELETE, SELECT
- ❖ **Data Control Language (DCL)** – Perintah SQL untuk mengontrol database atau table : GRANT, REVOKE

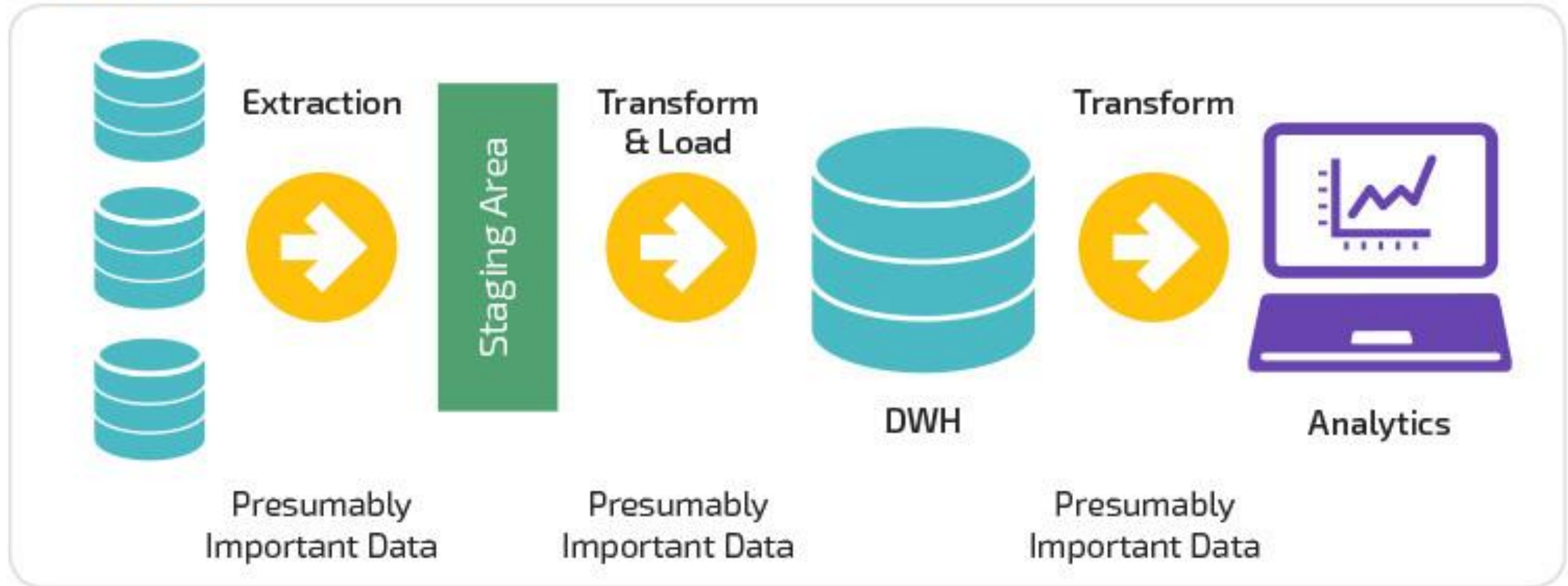


# Database → Data Warehouse



# ETL :: Data Warehouse

## ETL



# Definisi Data Warehouse

Data Warehouse adalah sekumpulan data yang memiliki enam buah sifat atau karakteristik : subject oriented, integrated, process oriented, time variant, accessible, dan non volatile. ( **W.H. Bill Inmon : 1991** )

Data Warehouse merupakan sebuah system untuk pengumpulan data transaksional dari berbagai sumber data, yang mengutamakan Query dan Analisa Data. ( **Ralph Kimball : 1996** )

Data Warehouse merupakan sebuah system untuk pengumpulan dan konsolidasi data secara periodik dari sebuah system sumber data menjadi sumber data yang bersifat dimensional atau dalam bentuk normalisasi data. ( **Vincent Rainardi : 2008** )

# Quiz!!

---

✚ Apa manfaat dari data warehouse bagi organisasi bisnis (perusahaan) ?

# Manfaat Data Warehouse

- ❖ Mempertahankan sejarah data, bahkan jika sistem transaksi sumber tidak beroperasi.
- ❖ Mengintegrasikan data dari beberapa sistem sumber, memungkinkan pandangan sentral di seluruh perusahaan.
- ❖ Meningkatkan kualitas data, dengan menyediakan kode konsisten dan deskripsi, bahkan memperbaiki data yang buruk.
- ❖ Menampilkan seluruh informasi organisasi secara konsisten.
- ❖ Memperbaiki data sehingga memberikan kinerja query yang sangat baik, bahkan untuk pertanyaan analitik kompleks, tanpa mempengaruhi sistem operasional.
- ❖ Peningkatan produktivitas perusahaan dalam pengambilan keputusan