



Cơ sở dữ liệu

BÀI 1: GIỚI THIỆU MÔN HỌC, TỔNG QUAN VỀ HỆ CSDL

Nguyễn Thị Hậu

Đại học Công nghệ - Đại học Quốc Gia Hà Nội

nguyenhau@vnu.edu.vn

GIỚI THIỆU MÔN HỌC

Nội dung

- **Lý thuyết:**

1. Các khái niệm cơ bản về CSDL
2. Mô hình thực thể liên kết
3. Mô hình quan hệ, các ràng buộc quan hệ, đại số quan hệ
4. Phụ thuộc hàm và chuẩn hóa CSDL quan hệ, các thuật toán thiết kế lược đồ CSDL quan hệ

- **Thực hành:**

1. Thực hành sử dụng Mysql
2. Tuần 2 - 15

Chuẩn đầu ra học phần

Kiến Thức

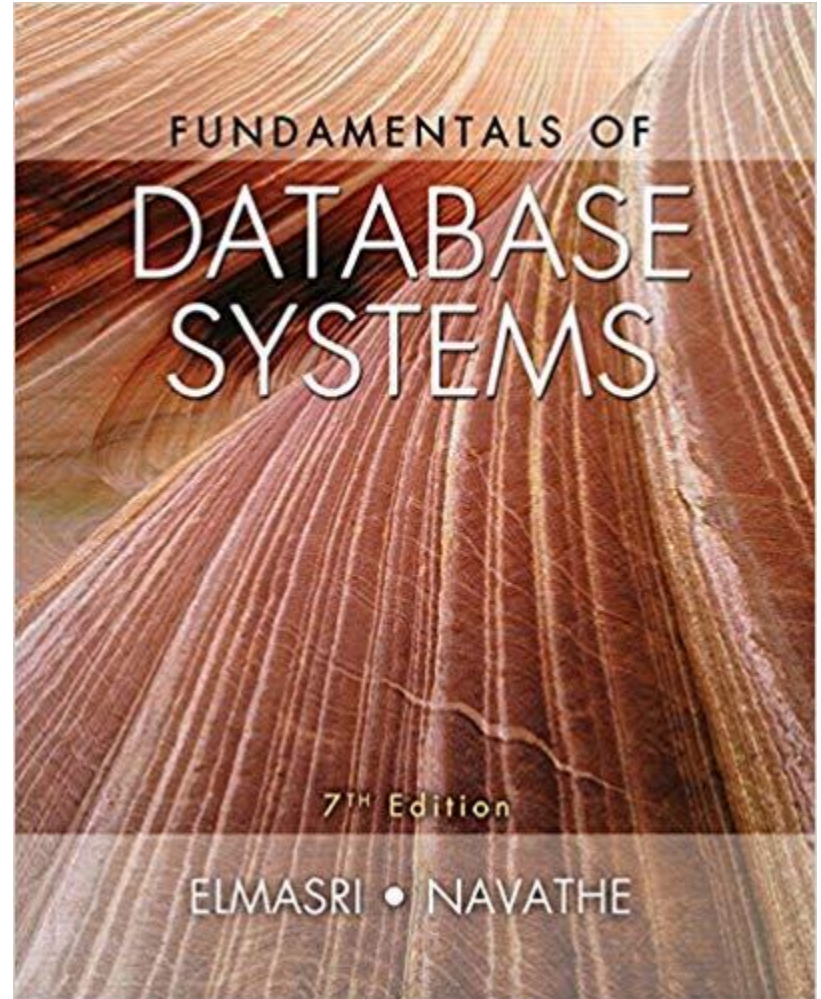
- Vận dụng được các kiến thức về mô hình thực thể - liên kết
- Vận dụng được các kiến thức về mô hình quan hệ và đại số quan hệ
- Vận dụng được các kiến thức về phụ thuộc hàm và các dạng chuẩn
- Vận dụng được các thuật toán thiết kế CSDL quan hệ

Kỹ năng

- Cài đặt được một hệ quản trị CSDL, chẳng hạn MySQL, PostgreSQL, MS SQL
- Sử dụng thành thạo ngôn ngữ truy vấn SQL để thực hiện các yêu cầu truy xuất dữ liệu

Tài liệu tham khảo

- Ramez Elmasri & Shamkant B. Navathe, Fundamentals of Database Systems, 7th Edition, Pearson 2016.
- CORONEL, C., and MORRIS, S. 2017, Database Systems: Design, Implementation, and Management. 13th Ed., Thomson Course Technology



Kiểm tra, đánh giá

- **Nhiệm vụ của sinh viên:**

- Có tài khoản truy cập vào website môn học:

- <https://portal.uet.vnu.edu.vn/>

- Tham gia các bài giảng, thực hành trên lớp

- Tự học và nghiên cứu

- **Kiểm tra, đánh giá:**

- **Thực hành:** 30% (*Bài tập hàng tuần, kiểm tra giữa kỳ, cuối kỳ*)

- **Lý thuyết:** 70% (*Điểm danh: 10%, Bài tập hàng tuần: 10%, kiểm tra lần 1 tuần 6-7: 15%, kiểm tra lần 2 theo lịch của nhà trường: 35%*)

- **Điểm cộng:** chữa bài trên lớp (tối đa 15%)

- *Sinh viên **vắng quá 20%** số buổi học sẽ không có điểm tổng kết*

- *Điểm bài tập bằng 0 nếu thiếu/gian lận 3 bài tập bất kỳ*

Database Career Opportunities

JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	SQL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc.

Top 10 Hệ quản trị CSDL 2024

423 systems in ranking, September 2024

Rank	Score			DBMS	Database Model	Score		
	Sep 2024	Aug 2024	Sep 2023			Sep 2024	Aug 2024	Sep 2023
1.	1.	1.		Oracle +	Relational, Multi-model ⓘ	1286.59	+28.11	+45.72
2.	2.	2.		MySQL +	Relational, Multi-model ⓘ	1029.49	+2.63	-82.00
3.	3.	3.		Microsoft SQL Server +	Relational, Multi-model ⓘ	807.76	-7.41	-94.45
4.	4.	4.		PostgreSQL +	Relational, Multi-model ⓘ	644.36	+6.97	+23.61
5.	5.	5.		MongoDB +	Document, Multi-model ⓘ	410.24	-10.74	-29.18
6.	6.	6.		Redis +	Key-value, Multi-model ⓘ	149.43	-3.28	-14.26
7.	7.	↑ 11.		Snowflake +	Relational	133.72	-2.25	+12.83
8.	8.	↓ 7.		Elasticsearch	Search engine, Multi-model ⓘ	128.79	-1.04	-10.20
9.	9.	↓ 8.		IBM Db2	Relational, Multi-model ⓘ	123.05	+0.04	-13.67
10.	10.	↓ 9.		SQLite +	Relational	103.35	-1.44	-25.85
11.	11.	↑ 12.		Apache Cassandra +	Wide column, Multi-model ⓘ	98.94	+1.94	-11.11
12.	12.	↓ 10.		Microsoft Access	Relational	93.76	-2.61	-34.81
13.	13.	↑ 14.		Splunk	Search engine	93.02	-3.08	+1.63
14.	↑ 15.	↑ 17.		Databricks +	Multi-model ⓘ	84.24	-0.22	+9.06
15.	↓ 14.	↓ 13.		MariaDB +	Relational, Multi-model ⓘ	83.44	-3.09	-17.01
16.	16.	↓ 15.		Microsoft Azure SQL Database	Relational, Multi-model ⓘ	72.95	-2.08	-9.78

Các HQT CSDL dùng bởi Google, Facebook, Twitter...

[stackshare.io]



mongoDB



redis

Chương 1. Tổng quan về Hệ CSDL

- Một số khái niệm
- Quá trình phát triển
- Một số đặc tính của CSDL
- Người sử dụng CSDL
- Kiến trúc của HQT CSDL
- Các tính năng của HQT CSDL
- Một số khái niệm: Mô hình, Lược đồ, Thể hiện
- Ngôn ngữ CSDL

1.1 MỘT SỐ KHÁI NIỆM

Thông tin và dữ liệu

Dữ liệu	Thông tin
<ul style="list-style-type: none">- Là các số liệu, sự kiện thô- Là ký tự, chuỗi ký tự, số, hình ảnh, âm thanh, video...- Không có ngữ cảnh- VD: 50117	<ul style="list-style-type: none">- Dữ liệu có ngữ cảnh- Dữ liệu được xử lý- Là kết quả của việc tóm tắt, tổ chức, phân tích dữ liệu- VD<ul style="list-style-type: none">- Ngày kiểm tra cuối kỳ môn CSDL:05/01/2017- 50 117 VNĐ giá của một cuốn lịch để bàn năm 2018- 50117 Mã một sản phẩm trong siêu thị

Cơ sở dữ liệu

❖ **Cơ sở dữ liệu (Database):** Một tập hợp những dữ liệu có liên quan với nhau được lưu trữ trong máy tính đáp ứng nhu cầu khai thác của nhóm người sử dụng với các mục đích xác định nào đó.

❖ **Các tính chất:**

- Một CSDL biểu diễn một phần của thế giới thực (thế giới nhỏ - miniworld), được cập nhật phản ánh sự thay đổi của thế giới nó biểu diễn.
- Một cơ sở dữ liệu là một tập hợp dữ liệu liên kết với nhau một cách logic và mang một nghĩa nào đó.
- Một cơ sở dữ liệu được thiết kế và được phổ biến cho một mục đích riêng, có nhóm người sử dụng.

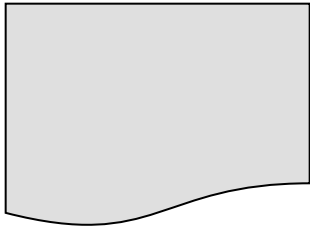
Ví dụ Một ứng dụng sử dụng CSDL truyền thống

- Giả sử cần xây dựng một hệ thống quản lý các thông tin sau:
 - Sinh viên
 - Môn học
 - Giáo viên
 - Mỗi sinh viên đang học một số môn học, mỗi giảng viên đang dạy một số lớp môn học

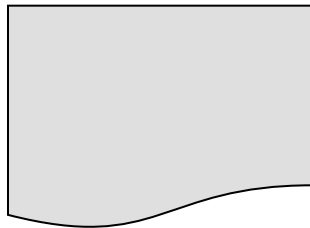
Không sử dụng HQT CSDL ?

Lưu trữ dữ liệu ở các tập tin:

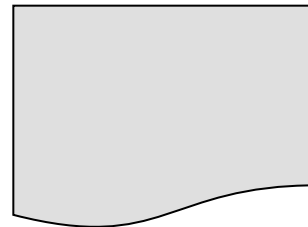
sinhvien.txt



Monhoc.txt



Giaovien.txt



Viết chương trình (C++, Java...) thực hiện các nhiệm vụ

Không sử dụng HQT CSDL

- Thêm sinh viên “Nguyễn Văn Minh” vào lớp môn học “INT2207-1”:

Viết chương trình C++ thực hiện các nhiệm vụ:

Đọc ‘sinhvien.txt’

Đọc ‘Monhoc.txt’

Tìm và cập nhật bản ghi “Nguyễn Văn Minh”

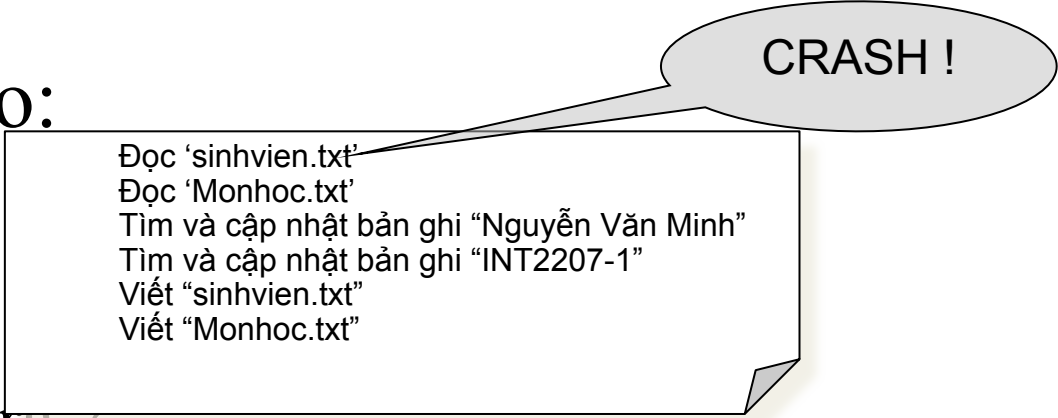
Tìm và cập nhật bản ghi “INT2207-1”

Viết “sinhvien.txt”

Viết “Monhoc.txt”

Không sử dụng HQT CSDL

- Hệ thống bị treo:

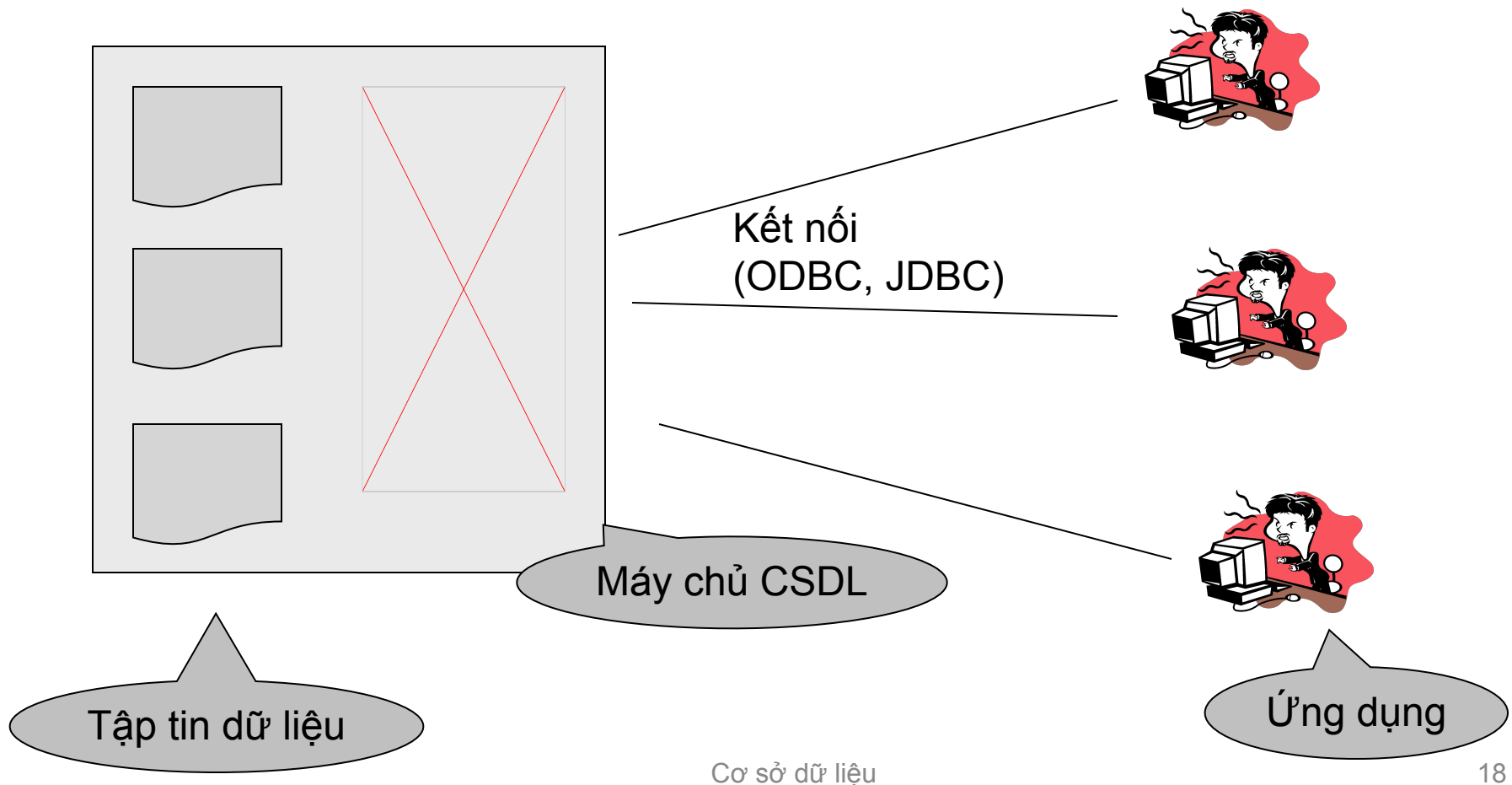


Đọc 'sinhvien.txt'
Đọc 'Monhoc.txt'
Tìm và cập nhật bản ghi "Nguyễn Văn Minh"
Tìm và cập nhật bản ghi "INT2207-1"
Viết "sinhvien.txt"
Viết "Monhoc.txt"

CRASH !

- Vấn đề gì xảy ra ?
- Bộ dữ liệu lớn (giả sử 50GB)
 - Vấn đề gì xảy ra ?
- Nhiều người dùng truy cập đồng thời

Giải quyết vấn đề với HQT CSDL



Hệ quản trị CSDL

□ ***Hệ quản trị CSDL (Database Management System)***

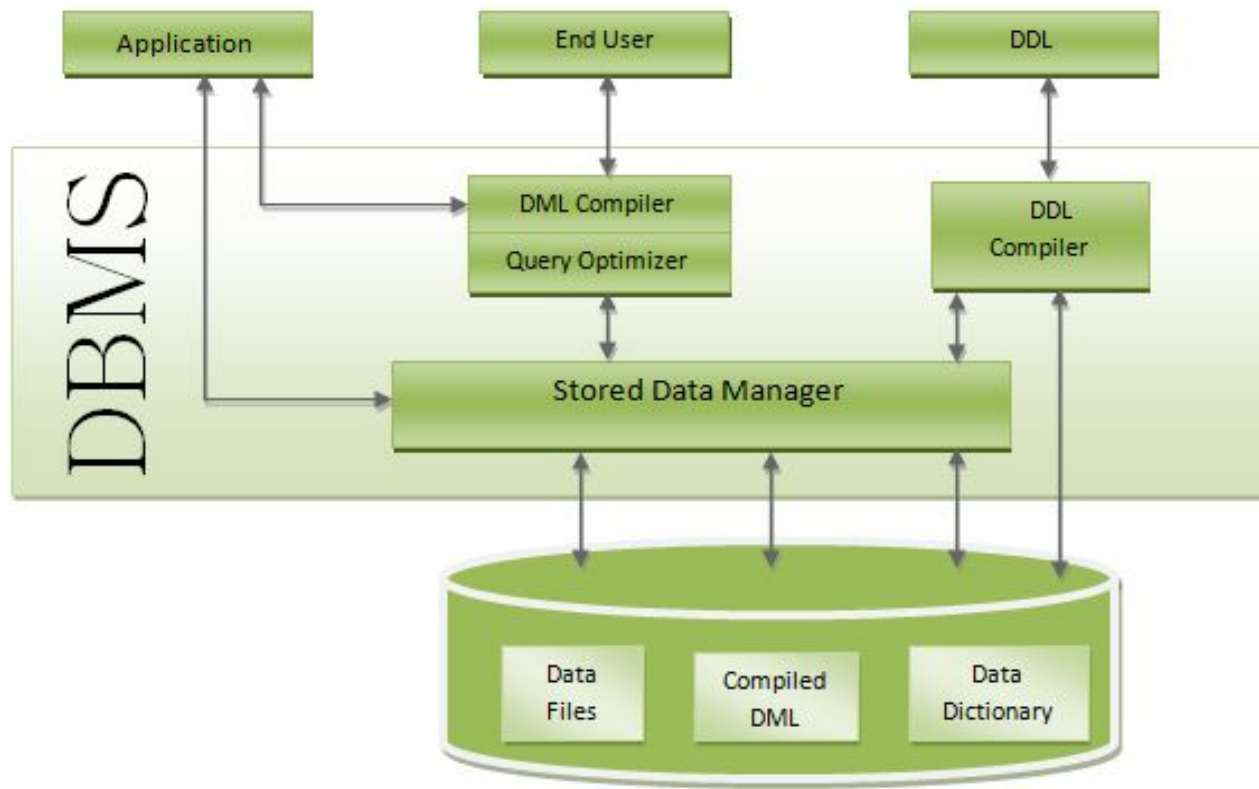
– Tập hợp các chương trình cho phép người sử dụng tạo, bảo trì và khai thác CSDL; tức là phần mềm cho phép ***định nghĩa, xây dựng*** và ***thao tác*** với dữ liệu.

- ***Định nghĩa*** CSDL: đặc tả các kiểu dữ liệu, cấu trúc mô tả chi tiết về dữ liệu, các ràng buộc của dữ liệu trong CSDL.

- ***Xây dựng*** CSDL: lưu trữ dữ liệu trên thiết bị bộ nhớ

- ***Thao tác***: truy vấn, cập nhật và sinh báo cáo.

Cấu trúc Hệ CSDL, Hệ quản trị CSDL



Hệ CSDL

Ví dụ CSDL Quản lý Đào tạo

- *Các đối tượng:* Sinh viên, Môn học, Học phần, Ngành..
- *Mối quan hệ:*
 - Mỗi sinh viên có thể học một số học phần trong mỗi kỳ
 - Mỗi môn học được dạy trong một hoặc nhiều học phần xác định
 - Mỗi môn học có thể có các môn điều kiện
 - Mỗi học phần do một giáo viên giảng
 - Mỗi môn học do một khoa quản lý
 - Mỗi sinh viên thuộc một ngành nào đó
 - Mỗi ngành do một khoa quản lý.

Ví dụ CSDL Quản lý Đào tạo

SINHVIEN			
MaSv	Hoten	Lop	MaNganh

MONHOC			
MaMH	TenMH	SoTC	MaKhoa

HOCPHAN			
MaHP	MaMH	Hocky	Giaovien

KHOA		
MaKhoa	TenKhoa	MaNganh

DIEM		
MaSv	MaHP	Diem

MONDK	
MaMH	MaMonDK
NGANH	
MaNg	TenNg

Ví dụ CSDL Quản lý Đào tạo

- **Định nghĩa CSDL:** Xác định cấu trúc các bảng, Các kiểu dữ liệu của các trường,...
- **Xây dựng CSDL:** Dữ liệu được ghi trong các tệp như thế nào, các quan hệ giữa các bản ghi,
- **Thao tác CSDL:** Cập nhật và khai thác dữ liệu.

Meta data

employee_id	first_name	last_name	nin	department_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Comelsen	NE 22 63 82	2
47	Andrew	Petculescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Bary	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Bemdt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1
54	Bonnie	Hall	WW 53 77 68 A	15
55	Taylor	Li	ZE 55 22 80 B	1

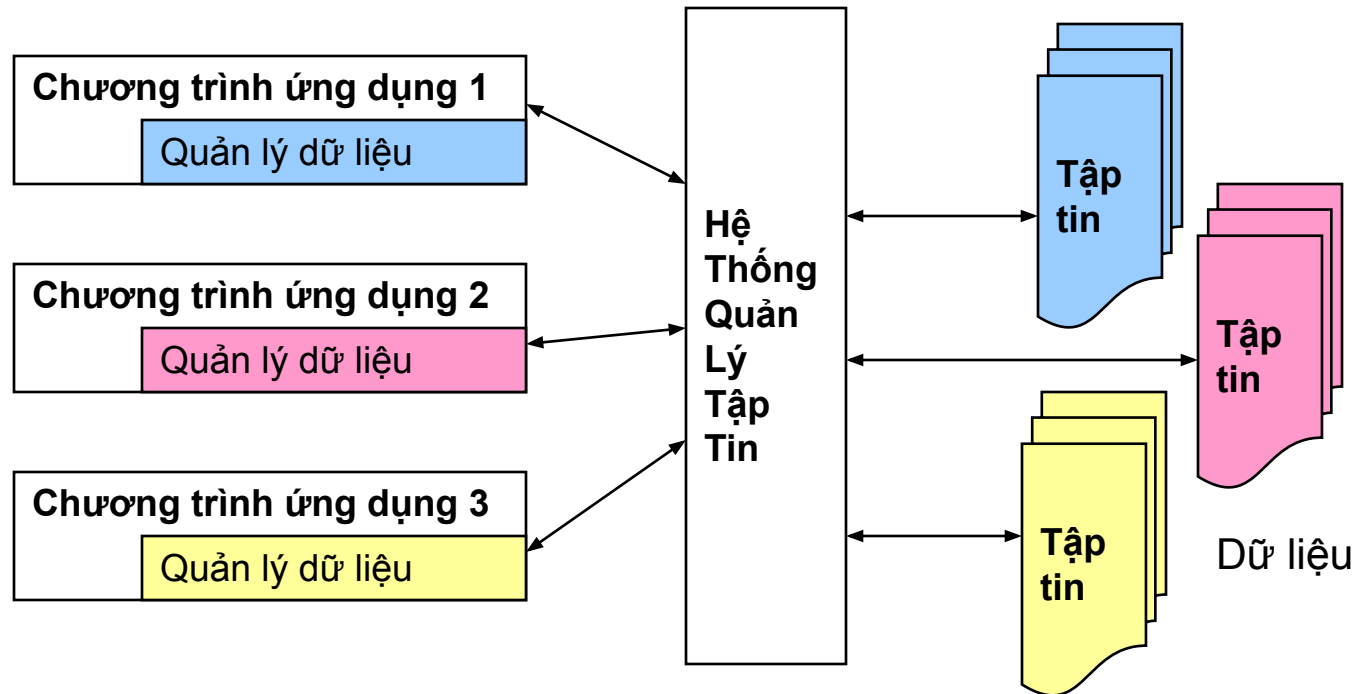
Data

Metadata

Column	Data Type	Description
employee_id	int	Primary key of a table
first_name	nvarchar(50)	Employee first name
last_name	nvarchar(50)	Employee last name
nin	nvarchar(15)	National Identification Number
position	nvarchar(50)	Current position title, e.g. Secretary
department_id	int	Employee department. Ref: Departments
gender	char(1)	M = Male, F = Female, Null = unknown
employment_start_date	date	Start date of employment in organization.
employment_end_date	date	Employment end date. Null if employee still employed.

1.2 QUÁ TRÌNH PHÁT TRIỂN

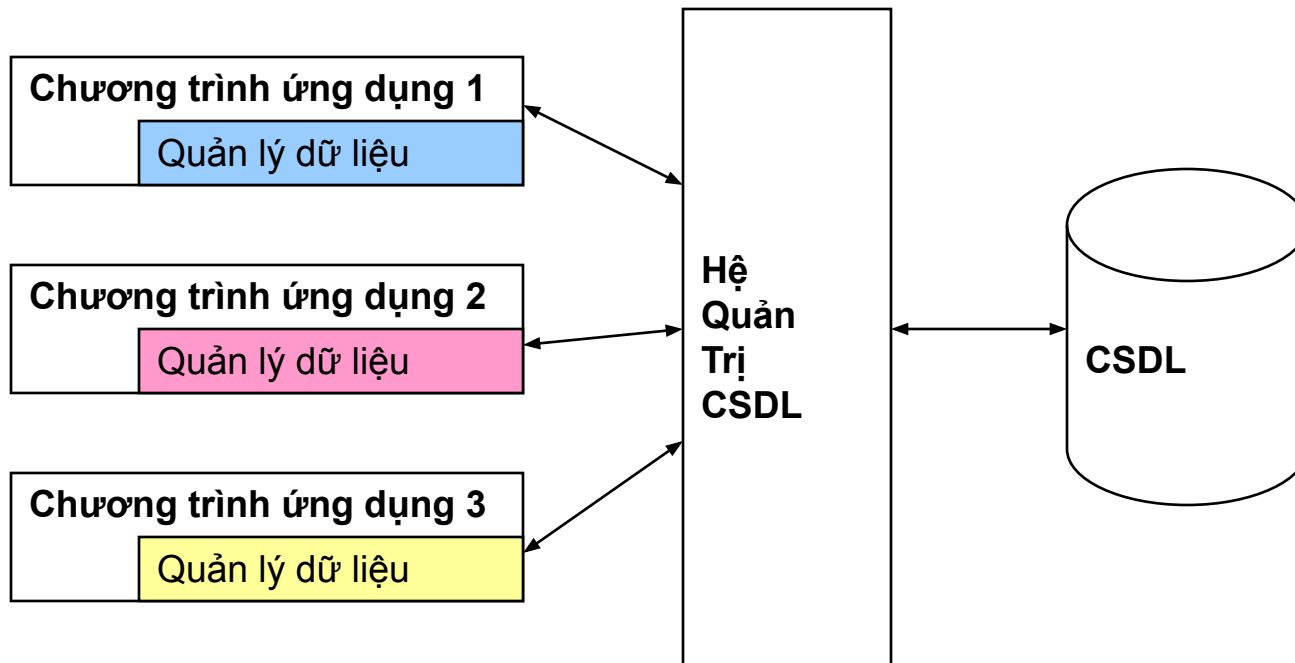
Tập tin (trước 1960)



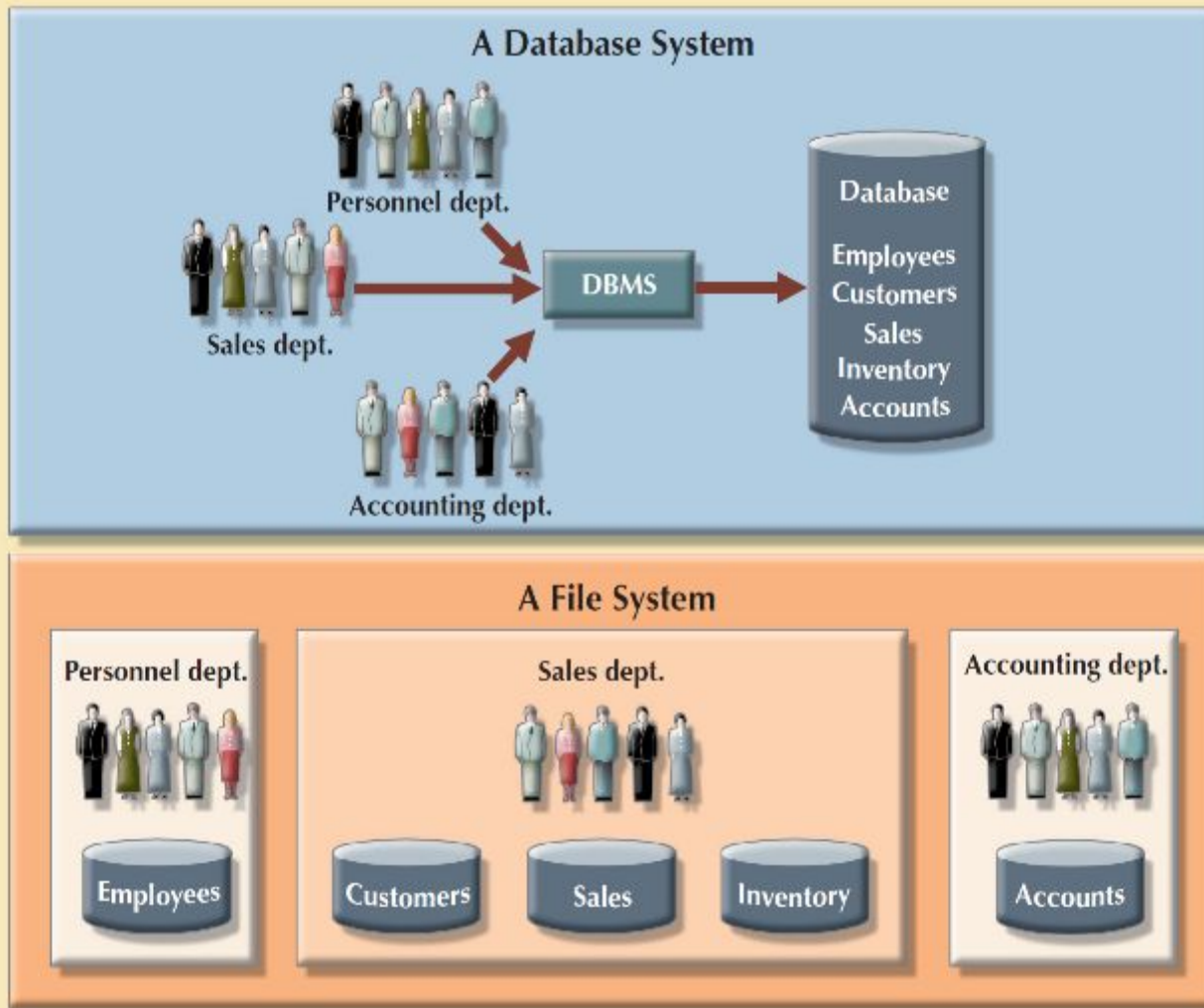
Tập tin (trước 1960)

- *Hạn chế*
 - Dữ liệu bị trùng lặp và dư thừa
 - Thiếu tính nhất quán giữa các dữ liệu
 - Khó khăn trong việc truy xuất
 - Việc chia sẻ dữ liệu bị hạn chế
 - Khó khôi phục

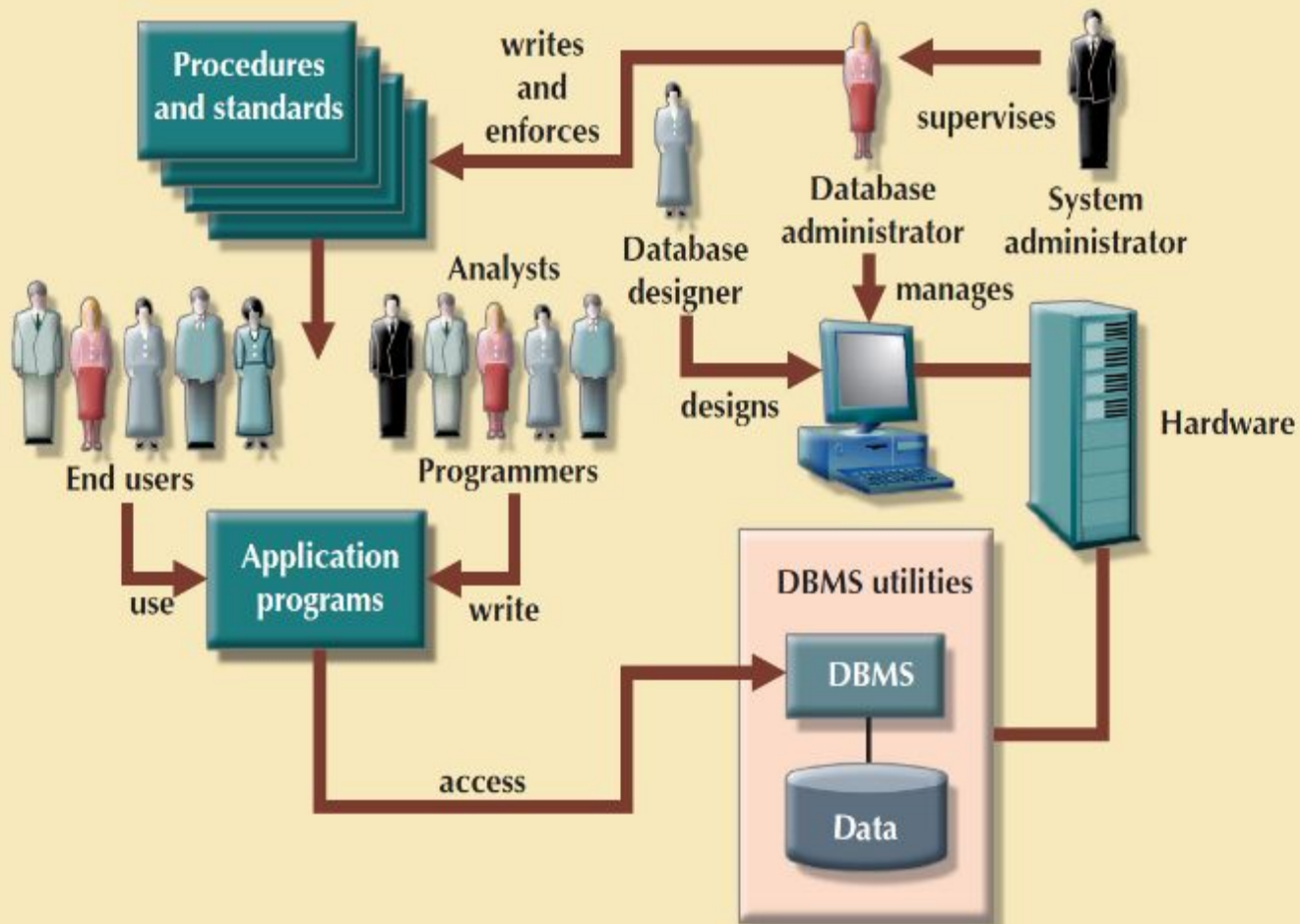
Cơ sở dữ liệu (Database)



Contrasting Database and File Systems

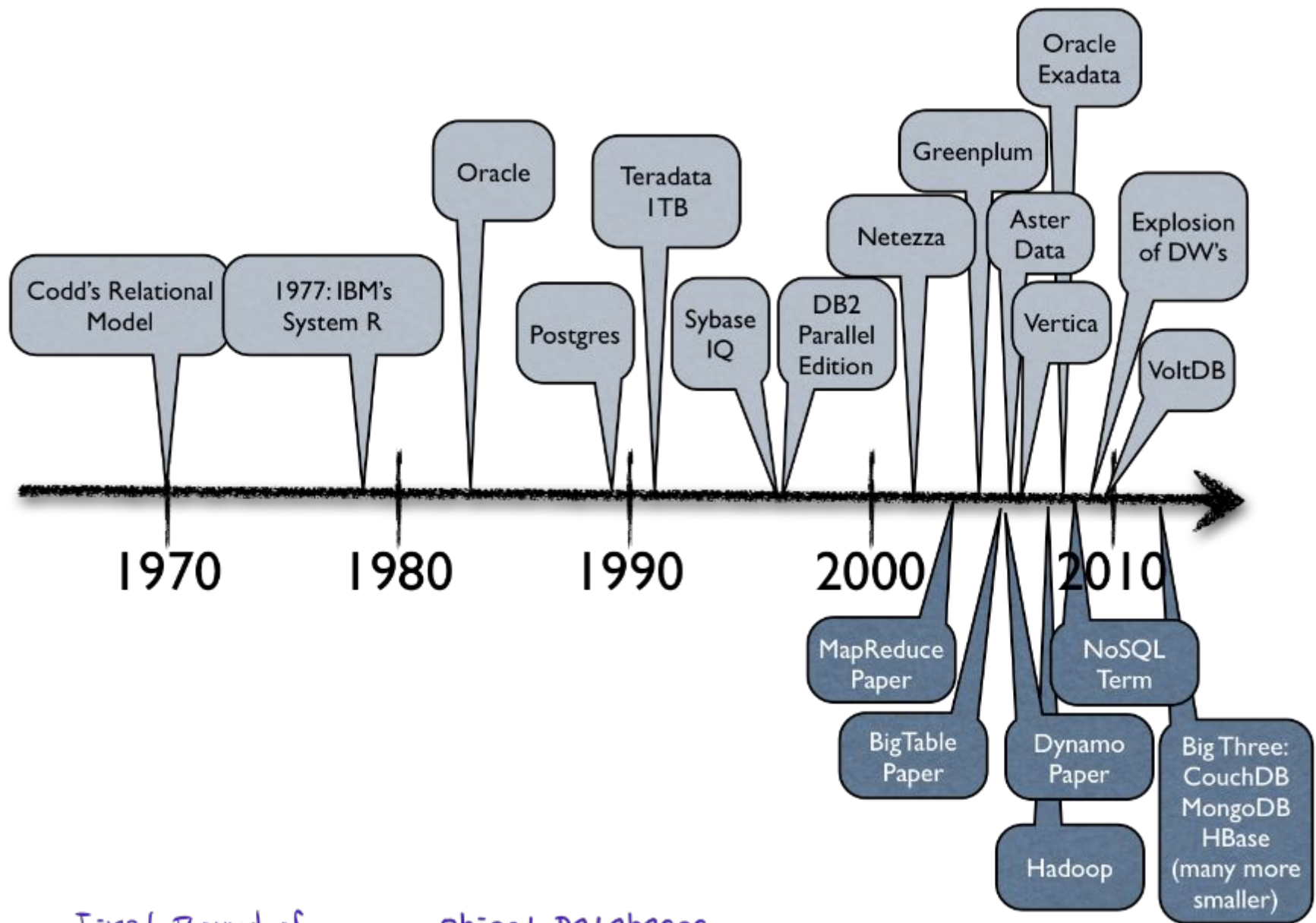


The Database System Environment



Quá trình phát triển

- **Những năm 1960s:** HQT CSDL đầu tiên ra đời
 - IMS (Information Management System) của IBM dựa trên mô hình phân cấp,
 - IDS (Integrated Data Store) do Charles Bachmann phát triển dựa trên mô hình mạng, sau này dùng cho CODASYL
- **1970-1972:** Edgar Codd giới thiệu mô hình quan hệ của dữ liệu
- **1974-1977:** HQT CSDL theo mô hình quan hệ đầu tiên xuất hiện
 - Ingres của UBC
 - System R của IBM
- **1976:** P. Chen giới thiệu mô hình thực thể liên kết (Entity Relationship model)
- **Từ 1980:**
 - SQL trở thành ngôn ngữ truy vấn chuẩn.
 - Nhiều HQT CSDL mới ra đời: Oracle, DB2, Access, Foxpro, Paradox
- **Từ 1990** Xuất hiện HQT CSDL hướng đối tượng
- **1998:** Hệ quản trị CSDL theo mô hình NoSQL (phổ biến 2009)
- **~2000:** Semi structured, XML (XQuery)



First Round of
Database Wars

Object Databases
challenge

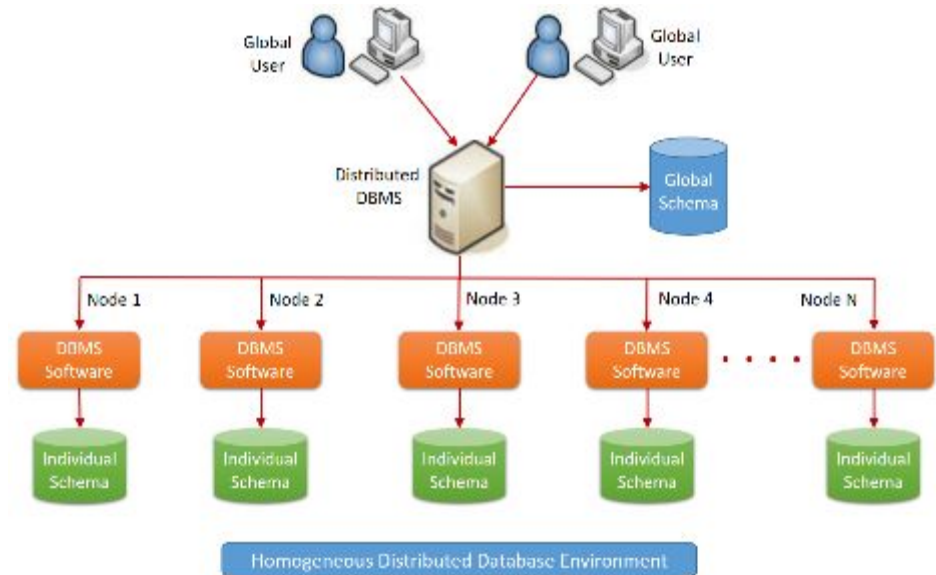
Semi-structured

Types of Databases

- **Single-user database:** Supports one user at a time
 - **Desktop database:** Runs on PC
- **Multi-user database:** Supports multiple users at the same time
 - **Workgroup databases:** Supports a small number of users or a specific department
 - **Enterprise database:** Supports many users across many departments

Types of Databases

- **Centralized database:**
Data is located at a single site
- **Distributed database:**
Data is distributed across different sites
- **Cloud database:**
Created and maintained using cloud data services that provide defined performance measures for the database

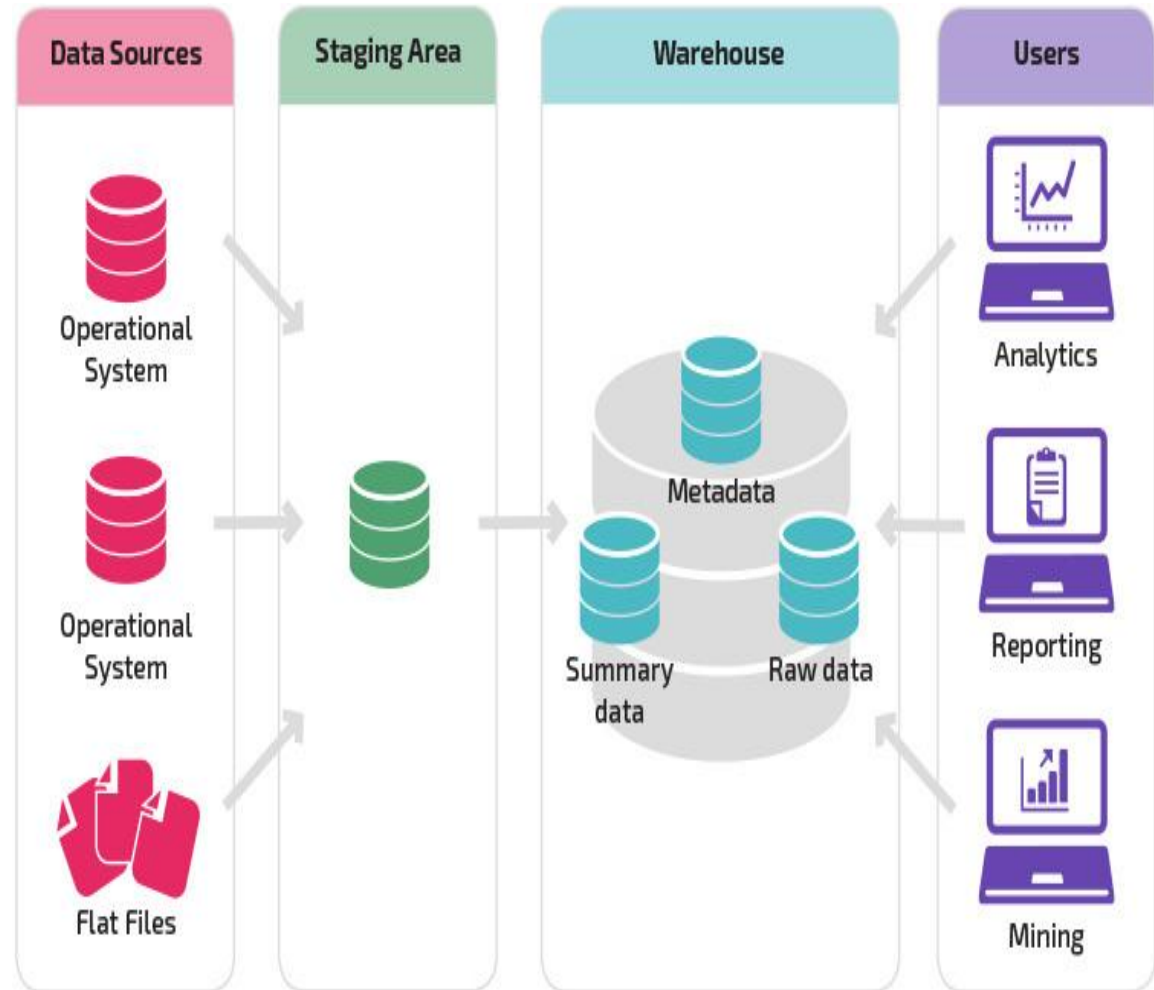


Types of Databases

- **General-purpose databases:** Contains a wide variety of data used in multiple disciplines
- **Discipline-specific databases:** Contains data focused on specific subject areas

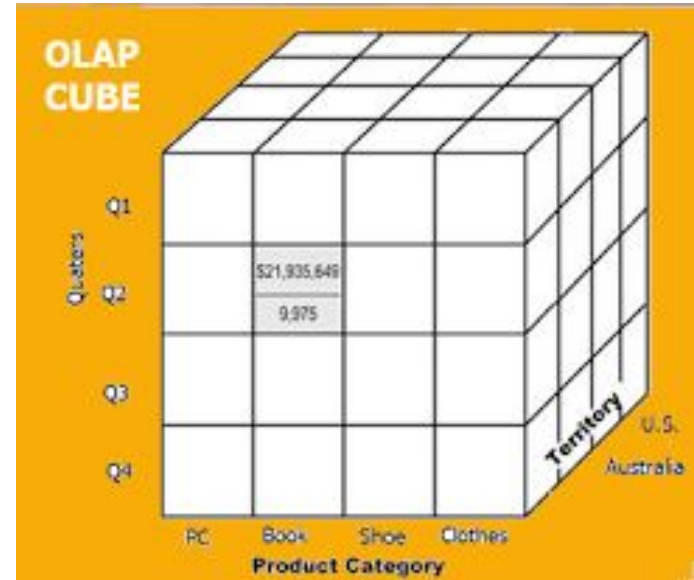
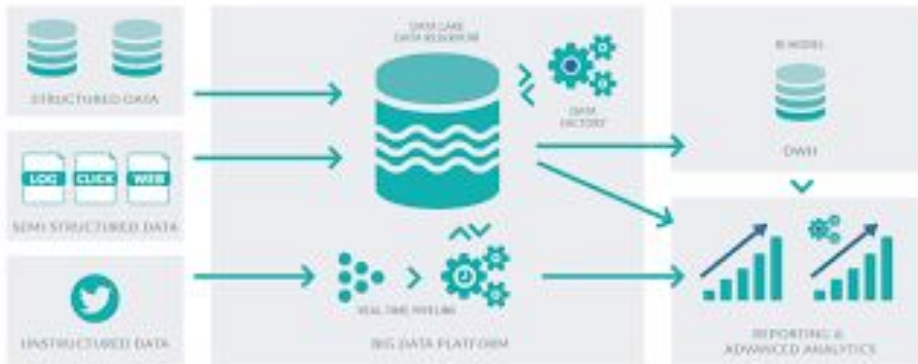
Types of Databases

- **Operational database:**
Designed to support a company's day-to-day operations
- **Analytical database:**
Stores historical data and business metrics used exclusively for tactical or strategic decision making
 - **Data warehouse:**
Stores data in a format optimized for decision support



Types of Databases

- **Business intelligence:** Captures and processes business data to generate information that support decision making

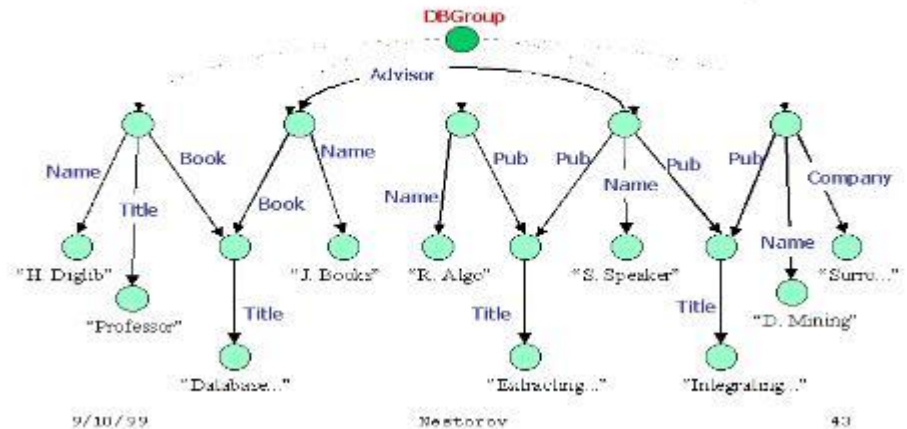


- **Online analytical processing (OLAP)**
 - Enable retrieving, processing, and modeling data from the data warehouse

Types of Databases

- **Unstructured data:** It exists in their original state
- **Structured data:** It results from formatting
 - Structure is applied based on type of processing to be performed
- **Semi-structured data:** Processed to some extent
- **Extensible Markup Language (XML)**
 - Represents data elements in textual format

Semistructured Data: Example



```
<Persons>
<!--List of Demo Persons-->
<Person id="1">
  <FirstName>Christy</FirstName>
  <LastName>Rudolph</LastName>
  <Company>Howard Miller Co</Company>
  <Address>2920 Dry Pine Bay Rd</Address>
  <City>Killarney</City>
  <State>ON</State>
  <ZIP>P0M 2A0</ZIP>
  <Phone>705-287-8488</Phone>
  <Fax>705-287-0745</Fax>
  <Email>christy@alternativepoker.com</Email>
  <Web>http://www.alternativepoker.com</Web>
</Person id="1">
<Person id="2">
</Person id="3">
```

