



TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

Nội dung

1. **Khái niệm thông tin, dữ liệu**
2. Khái niệm về CSDL.
3. Hệ quản trị CSDL.
4. Các mô hình CSDL

Thông tin vs dữ liệu

- Dữ liệu là thông tin của đối tượng (ví dụ: người, vật, một khái niệm, sự việc...) được lưu trữ trên máy tính.

http://www.diffen.com/difference/Data_vs_Information#Examples_of_Data_and_Information

- Dữ liệu được lưu trữ dưới nhiều dạng khác nhau (các ký tự, ký số, hình ảnh, ký hiệu, âm thanh...)
- Dữ liệu về đối tượng có thể khác nhau, tùy thuộc vào ngữ cảnh.
 - Ví dụ: dữ liệu về đối tượng sinh viên có thể khác nhau tùy vào mục đích quản lý:
 - Quản lý điểm: Tên, mã sinh viên, điểm môn 1, điểm môn 2, điểm môn 3
 - Quản lý nhân thân: Tên, địa chỉ, ngày sinh, quê quán

Thông tin vs dữ liệu

- Thông tin là dữ liệu đã được xử lý như phân loại, tính toán, chọn lọc Sao cho phù hợp với ngữ cảnh.
- Ví dụ: điểm trung bình của lớp học, của khoa... là thông tin có được từ dữ liệu điểm sinh viên



Knowledge

Know-how, understanding,
experience, insight, intuition, and
contextualized information

Information

Contextualized, categorized, calculated and
condensed data

Data

Facts and figures which relay something specific, but
which are not organized in any way

Thông tin vs dữ liệu

- Dữ liệu hợp thành thông tin
- Thông tin có được từ việc xử lý dữ liệu
- Thông tin lột tả ý nghĩa của dữ liệu
- Thông tin đúng đắn, kịp thời và đầy đủ là mấu chốt để thực hiện những quyết định
- Quyết định đúng đắn là then chốt quyết định sự sống còn của tổ chức

http://www.jvn.edu.vn/uploads/news/2017_12/jvn-datascience.pdf

Nội dung

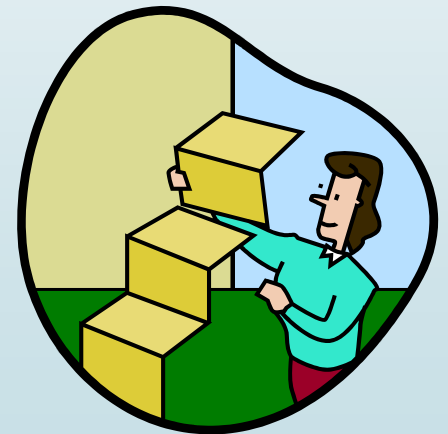
1. Khái niệm thông tin, dữ liệu
2. **Khái niệm về CSDL.**
3. Hệ quản trị CSDL.
4. Các mô hình CSDL

Khái niệm về CSDL

- Hệ thống tập tin cố định
- Định nghĩa cơ sở dữ liệu

Hệ thống tập tin cổ điển

- Dữ liệu được tổ chức riêng rẽ, phục vụ cho một mục đích của một đơn vị cụ thể
- Ưu điểm:
 - Triển khai ứng dụng nhanh.
 - Có khả năng đáp ứng nhanh chóng, kịp thời



Hệ thống tập tin cổ điển – nhược điểm

- Dư thừa dữ liệu – data redundancy
 - Các phiên bản khác nhau và mâu thuẫn nhau của cùng một dữ liệu
 - Hậu quả của việc dư thừa dữ liệu không thể kiểm soát được
 - Lãng phí công sức nhập liệu và không gian lưu trữ.
 - Những dị thường về dữ liệu - data anomalies
 - Khi sửa đổi
 - Khi thêm
 - Khi xóa
 - Dữ liệu không nhất quán - Data inconsistency
 - Thiếu các ràng buộc về dữ liệu - data integrity
- Thiếu sự chia sẻ thông tin giữa các nơi.

Hệ thống tập tin cổ điển – nhược điểm

- Truy cập đồng thời bất thường
- Vấn đề bảo mật và toàn vẹn dữ liệu



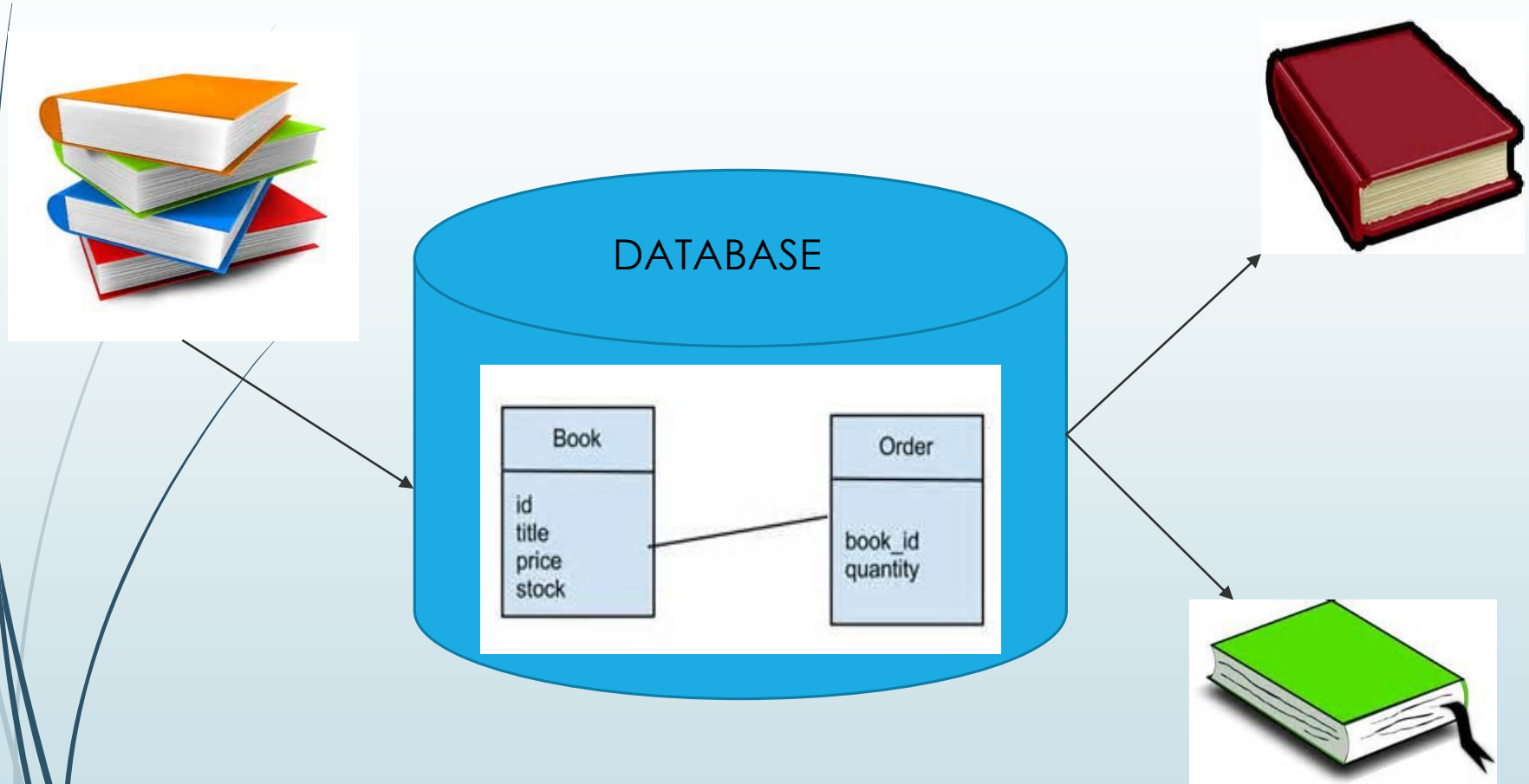
Khái niệm về CSDL

- Hệ thống tập tin cố định
- Định nghĩa cơ sở dữ liệu

Khái niệm về cơ sở dữ liệu

- Cơ sở dữ liệu là một hệ thống các **thông tin có cấu trúc** được lưu trữ trên các thiết bị lưu trữ thông tin thứ cấp (như băng từ, đĩa từ...) để có thể thỏa mãn yêu **cầu khai thác thông tin đồng thời** của nhiều người sử dụng hay nhiều chương trình ứng dụng với nhiều mục đích khác nhau.

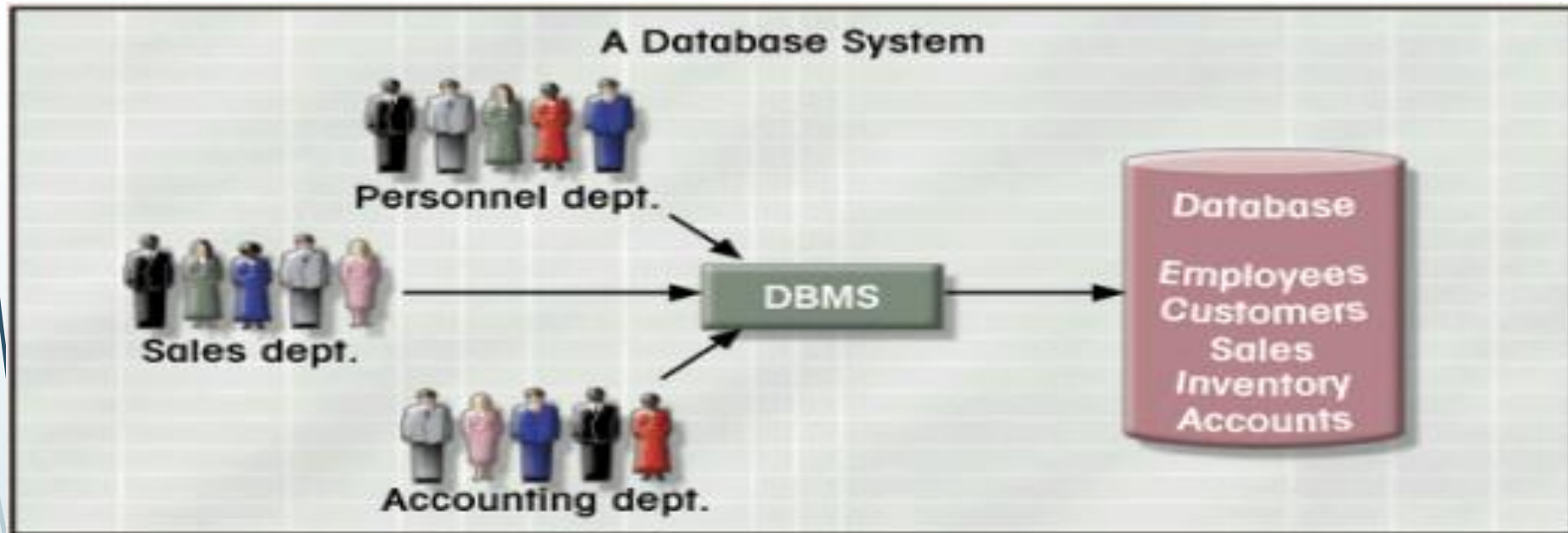
Cơ sở dữ liệu



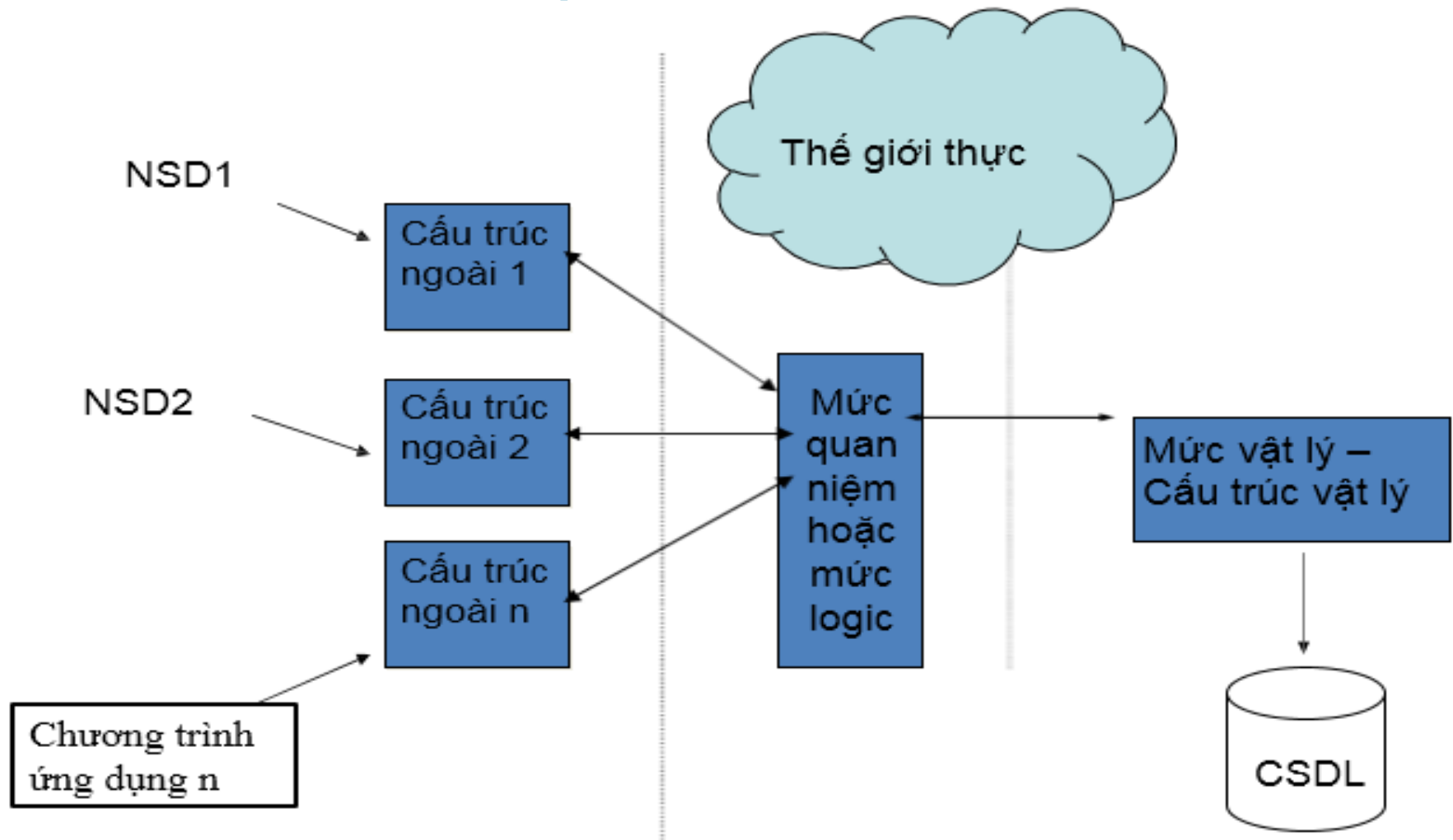
Ưu điểm của CSDL

- Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó bảo đảm được tính nhất quán và toàn vẹn dữ liệu.
- Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác nhau.
- Khả năng chia sẻ thông tin cho nhiều người sử dụng và nhiều ứng dụng khác nhau.

A database System vs a File System



Kiến trúc ba lớp của CSDL



Kiến trúc ba lớp của CSDL

- Mức trong: (mức vật lý – Physical) là mức lưu trữ CSDL
 - Lưu trữ như thế nào? ở đâu?
 - Dành cho người quản trị CSDL (Database Administrator).
- Mức quan niệm: (Conception hay Logical)
 - Cần phải lưu trữ bao nhiêu loại dữ liệu? là dữ liệu gì? mối quan hệ
- Mức ngoài: của người sử dụng không chuyên và các chương trình ứng dụng

Các đối tượng sử dụng CSDL

- Người sử dụng CSDL không chuyên về lĩnh vực tin học và CSDL → cần công cụ để họ có thể khai thác CSDL khi cần. (**End User**)
- Chuyên viên tin học → xây dựng các ứng dụng để phục vụ cho các mục đích quản lý. (**Application User**)
- Quản trị CSDL → tổ chức CSDL, bảo mật, cấp quyền, sao lưu, phục hồi dữ liệu, giải quyết các tranh chấp dữ liệu ...(**Database Administrator**)

Nội dung

1. Khái niệm thông tin, dữ liệu
2. Khái niệm về CSDL.
3. **Hệ quản trị CSDL.**
4. Các mô hình CSDL

Hệ quản trị cơ sở dữ liệu (DataBase Management System - DBMS)

- Phần mềm dùng quản lý cơ sở dữ liệu
- Một DBMS phải có:
 - Ngôn ngữ giao tiếp giữa người sử dụng và CSDL
 - Cơ chế bảo mật
 - Cơ chế giải quyết tranh chấp dữ liệu
 - Có cơ chế sao lưu (backup), phục hồi (restore)



Hệ quản trị cơ sở dữ liệu (DataBase Management System - DBMS)

► Ngôn ngữ giao tiếp:

- Ngôn ngữ mô tả dữ liệu (DDL – Data Definition Language): cho phép khai báo cấu trúc CSDL, các mối liên hệ của dữ liệu, các quy định, ràng buộc dữ liệu.
- Ngôn ngữ thao tác dữ liệu (DML – Data Manipulation Language): cho phép thực hiện thao tác thêm, xóa, sửa dữ liệu.
- Ngôn ngữ truy vấn có cấu trúc (SQL – Structured Query Language): cho phép người khai thác sử dụng để truy vấn thông tin cần thiết.
- Ngôn ngữ quản lý dữ liệu (DCL – Data Control Language) cho phép thay đổi cấu trúc bảng, khai báo bảo mật, cấp quyền cho người sử dụng.

Nội dung

1. Khái niệm thông tin, dữ liệu
2. Khái niệm về CSDL.
3. Hệ quản trị CSDL.
4. **Các mô hình CSDL**

Các mô hình dữ liệu

- Mô hình CSDL là kiến trúc DBMS sử dụng để lưu trữ các đối tượng trong CSDL và các mối liên hệ giữa chúng

Các mô hình dữ liệu

Mô hình dữ liệu
file phẳng
(Flat file)

Mô hình dữ liệu
phân cấp
(Hierarchical
model)

Mô hình dữ liệu
mạng (Network
model)

Mô hình dữ liệu
quan hệ
(Relational
model)

Mô hình dữ liệu
hướng đối
tượng
(ObjectOriented
model)

Mô hình dữ liệu file phẳng

- Các file dữ liệu được lưu rời rạc, không có mối liên hệ với nhau.

Ví dụ
mô
hình
file
phẳng

File Customer

Customer ID	Company Name	Contact First Name	Contact Last Name	Job Title	City	State
6	Company F	Francisco	Pérez-Olaeta	Purchasing Manager	Milwaukee	WI
26	Company Z	Run	Liu	Accounting Assistant	Miami	FL

File Employee

Employee ID	First Name	Last Name	Title
2	Andrew	Cencini	Vice President, Sales
5	Steven	Thrope	Sales Manager
9	Anne	Hellung-Larsen	Sales Representative

File Product

Product ID	Product Code	Product Name	Category	Quantity Per Unit	List Price
5	NWTO-5	Northwind Traders Olive Oil	Oil	36 boxes	\$21.35
7	NWTFN-7	Northwind Traders Dried Pears	Dried Fruit & Nuts	12 - 1 lb pkgs	\$30.00
40	NWTCM-40	Northwind Traders Crab Meat	Canned Meat	24 - 4 oz tins	\$18.40
41	NWTSO-41	Northwind Traders Clam Chowder	Soups	12 - 12 oz cans	\$9.65
48	NWTCA-48	Northwind Traders Chocolate	Candy	10 pkgs	\$12.75

File Order

Order ID	Customer ID	Employee ID	Order Date	Shipped Date	Shipping Fee
51	26	9	4/5/2006	4/5/2006	\$60.00
56	6	2	4/3/2006	4/3/2006	\$0.00
79	6	2	6/23/2006	6/23/2006	\$0.00

File Order Detail

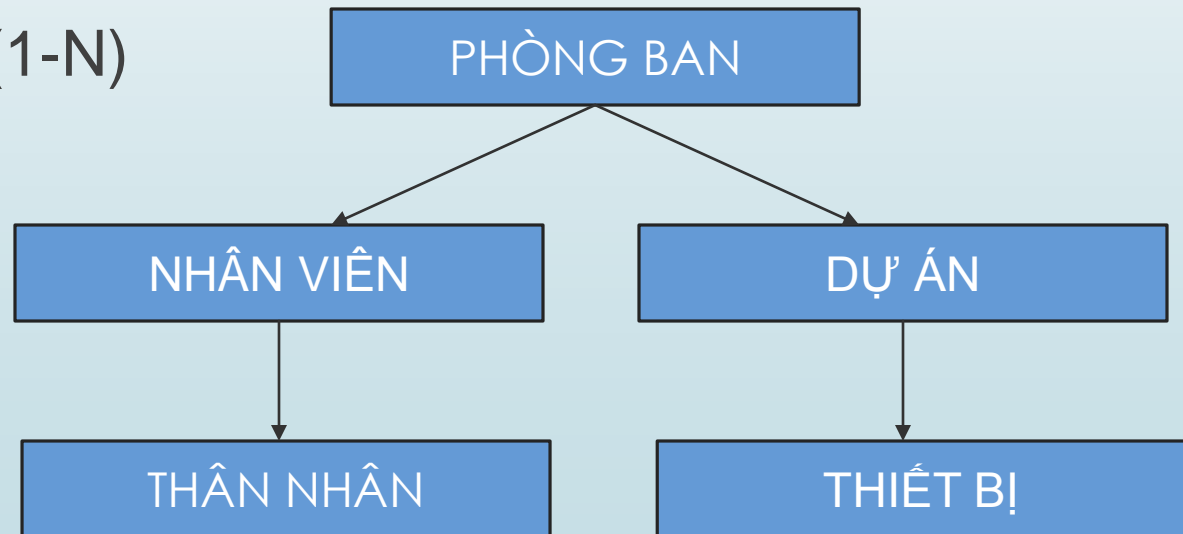
Order ID	Product ID	Unit Price	Quantity
51	5	\$21.35	15
51	41	\$9.65	21
51	40	\$18.40	2
56	48	\$12.75	20
79	7	\$30.00	14
79	51	\$53.00	8

Hệ thống file
quản lý đơn
hàng

TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

Mô hình dữ liệu phân cấp (Hierarchical model)

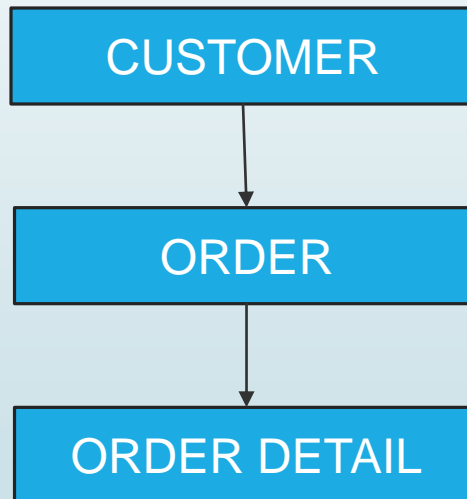
- Mỗi file trong flat file trở thành một nút (node)
- Các nút nối với nhau theo mối quan hệ *cha – con*, nghĩa là một cha có thể có nhiều con, nhưng mỗi con chỉ có một cha
- Thể hiện mối quan hệ
 - một – một (1 - 1)
 - một – nhiều (1-N)



Mô hình dữ liệu phân cấp (Hierarchical model)

HẠN CHẾ

Không thể hiện được mối quan hệ phức tạp

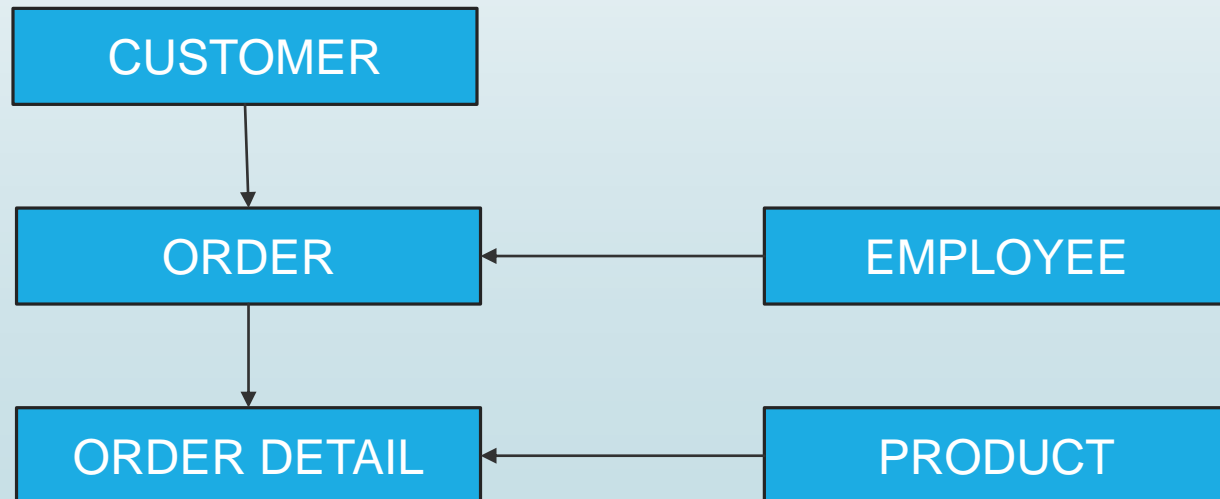


EMPLOYEE

PRODUCT

Mô hình dữ liệu mạng (Network model)

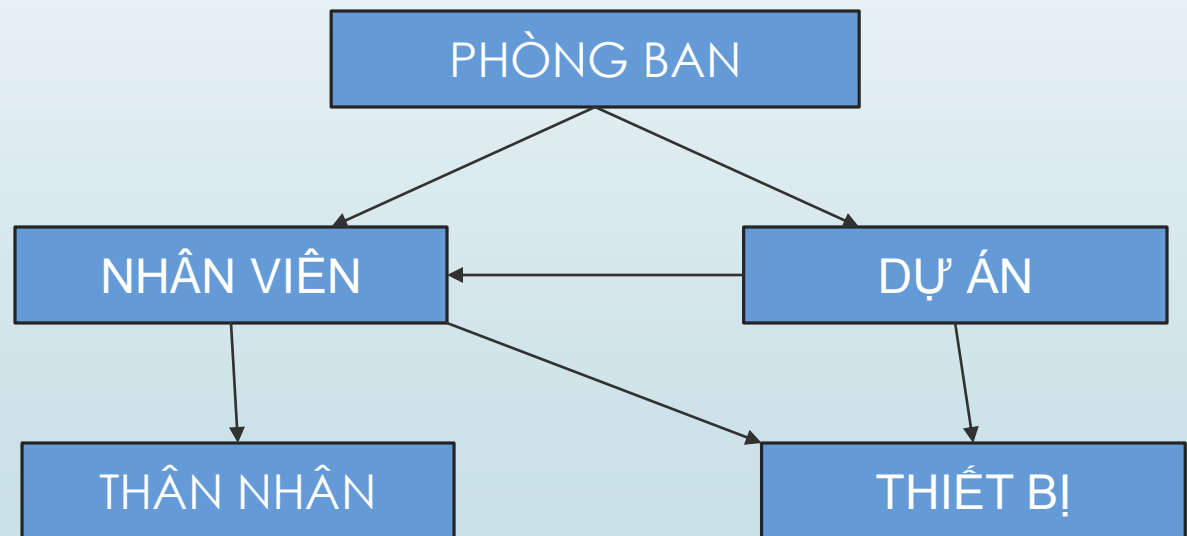
- Là mô hình phân cấp được cải tiến, một nút **con** có thể có nhiều nút **cha**.
- Cũng chỉ có mối quan hệ
 - một – một (1 - 1)
 - một – nhiều (1-N)



Mô hình dữ liệu mạng (Network model)

HẠN CHẾ

Phức tạp trong lập trình xử lý số liệu



Mô hình dữ liệu quan hệ (Relational model)

- Mô hình dữ liệu quan hệ dựa trên lý thuyết tập hợp và đại số quan hệ.
- Dữ liệu được biểu diễn dưới dạng bảng hai chiều, còn được gọi là quan hệ :
 - Mỗi hàng là một bản ghi (record), hay bộ (tuple)
 - Mỗi cột là một thuộc tính (attribute), hay trường (field)
- Trong mô hình dữ liệu quan hệ *không có các liên kết vật lý*. Dữ liệu trong hai bảng liên hệ với nhau thông qua *các cột chung*.

Mô hình dữ liệu quan hệ (Relational model)

Bảng customer

Customer ID	Company Name	Contact First Name	Contact Last Name	Job Title	City	State
6	Company F	Francisco	Pérez-Olaeta	Purchasing Manager	Milwaukee	WI
26	Company Z	Rui	Liu	Accounting Assistant	Miami	FL

Bảng Order

Order ID	Customer ID	Employee ID	Order Date	Shipped Date	Shipping Fee
51	26	9	4/5/2006	4/5/2006	\$60.00
56	6	2	4/3/2006	4/3/2006	\$ 0.00
79	6	2	6/23/2006	6/23/2006	\$ 0.00

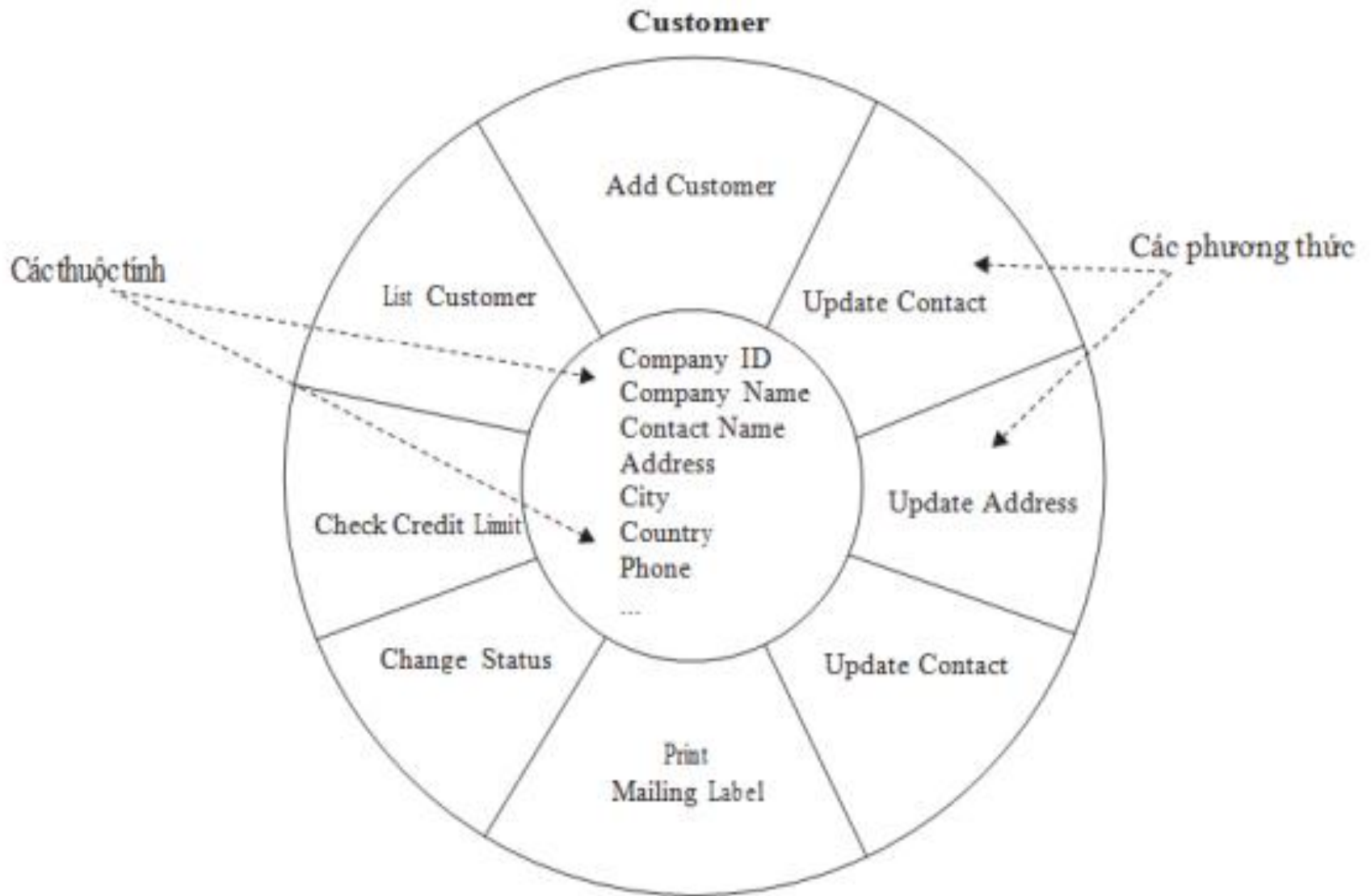
Bảng Employee

Employee ID	First Name	Last Name	Title
2	Andrew	Cencini	Vice President, Sales
5	Steven	Thrope	Sales Manager
9	Anne	Hellung-Larsen	Sales Representative

Mô hình dữ liệu hướng đối tượng (ObjectOriented model)

- Dựa trên cách tiếp cận của phương pháp lập trình hướng đối tượng.
- CSDL bao gồm các đối tượng:
 - Mỗi đối tượng bao gồm các thuộc tính, phương thức (hành vi) của đối tượng.
 - Các đối tượng trao đổi với nhau thông qua các phương thức.
 - Một đối tượng có thể được sinh ra từ việc thừa kế từ đối tượng khác, nạp chồng (hay định nghĩa lại) phương thức của đối tượng khác...

Ví dụ mô hình dữ liệu đối tượng Customer



Tóm tắt lịch sử cơ sở dữ liệu

1960 cơ sở dữ liệu đầu tiên

Integrated Data Store - IDS dựa trên *mô hình mạng*,

1970, Edgar Codd giới thiệu *mô hình dữ liệu quan hệ*

1990, hệ thống quan *hệ hướng đối tượng* xuất hiện như Illustra, Cohera, StreamBase Systems ...

1964 – 1966, Information Management System - IMS ra đời dựa trên *mô hình phân cấp*

1980 – 1990, ngôn ngữ *SQL* ra đời và trở thành ngôn ngữ truy vấn dữ liệu chuẩn sử dụng trong cơ sở dữ liệu quan hệ (SQL Server, Oracle, DB2....)

Câu hỏi

1. Phân biệt *thông tin* và *dữ liệu*?
2. Phân biệt *cơ sở dữ liệu* và *hệ quản trị cơ sở dữ liệu*?
3. Liệt kê các mô hình dữ liệu? Mô hình nào đang được sử dụng nhiều nhất?

Tham khảo

➡ http://www.jvn.edu.vn/uploads/news/2017_12/jvn-datascience.pdf