

# FPT POLYTECHNIC



Bài 1:

T NG QUAN V C S D LI U

---

- Giới thiệu khái niệm dữ liệu và cơ sở dữ liệu (CSDL)
- Các phương pháp quản lý dữ liệu và các cấu trúc
- Giới thiệu các mô hình dữ liệu khác nhau
- Hệ quản trị CSDL (DBMS) và hệ quản trị CSDL quan hệ (RDBMS)

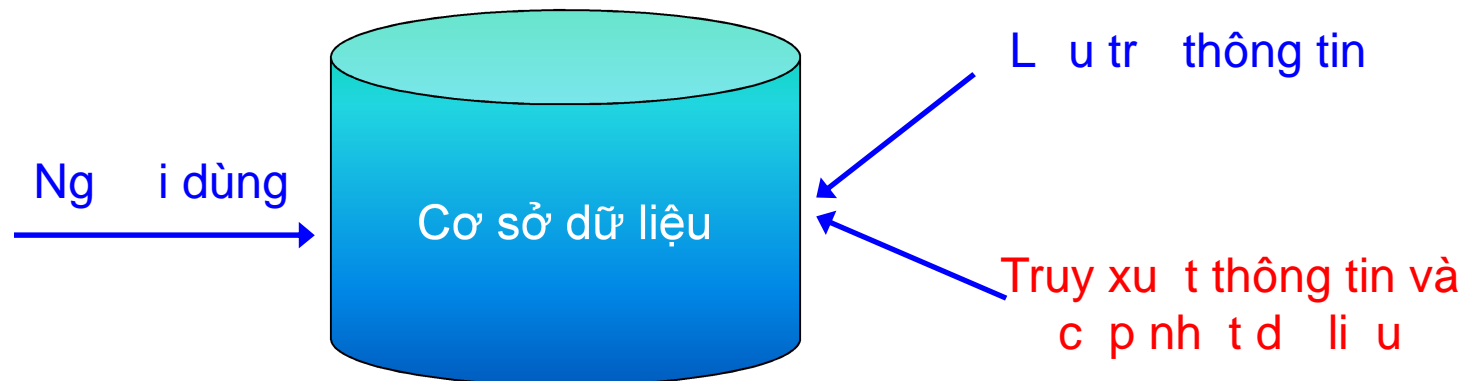
## ■ D ị u (data)

- Là các thông tin c a i t ã ng (ng ã i, v t, m t khái ni m, s vi c...) c l u tr ã trên máy tính.
- Có th truy nh p vào d ị u tr ích xu t thông tin.

## ■ D ị u c mô t d ị nh i u d ã ng khác nhau (các ký t , ký s , hình ã nh, ký hi u, âm thanh...). M i cách mô t g ã n v i m t ã ng ã ngh a nào ó.

- D ị u v ị t ̣ ng có th ̣ khác nhau, tùy thu ̣ vào ng ̣ c nh.
- Ví d ̣ : d ị u v ị t ̣ ng sinh viên có th ̣ khác nhau tùy vào m ̣ c ̣ ích qu ̣ n lý:
  - Qu ̣ n lý ̣ i m: Tên, mã sinh viên, ̣ i m môn 1, ̣ i m môn 2, ̣ i m môn 3
  - Qu ̣ n lý nhân thân: Tên, ̣ a ch , ngày sinh, quê quán, l ̣ p

- C s d li u (CSDL) = T p h p d li u c t ch c có c u trúc liên quan v i nhau và c l u tr trong máy tính.
- CSDL c thi t k , xây d ng cho phép ng i dùng l u tr d li u, truy xu t thông tin ho c c p nh t d li u



## ■ CSDL có thể có cấu trúc:

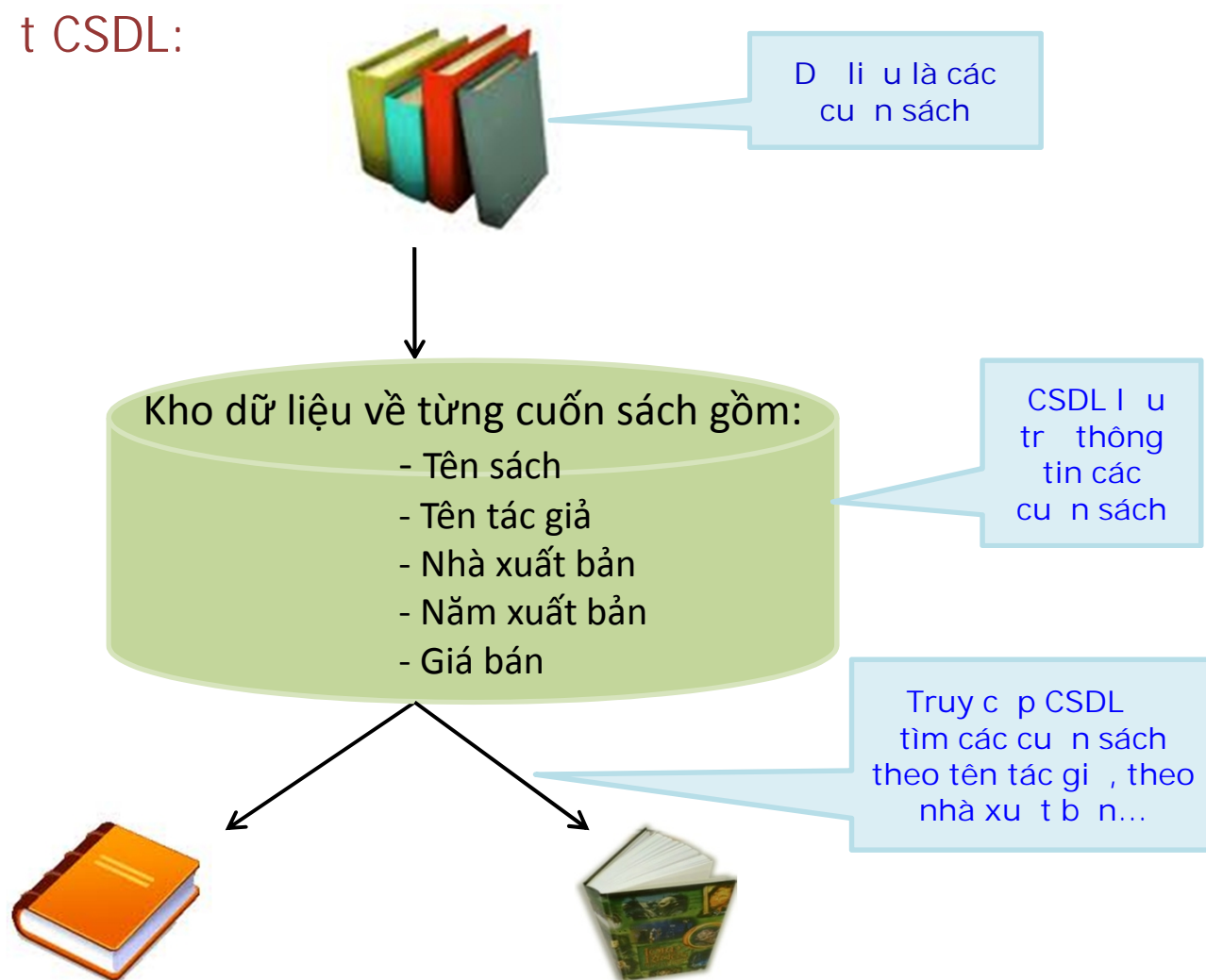
- Các dữ liệu có cấu trúc thành các bản ghi (record), các trường dữ liệu (field).
- Các dữ liệu có mối quan hệ (**relation**) với nhau.

## ■ Khả năng truy xuất thông tin từ CSDL:

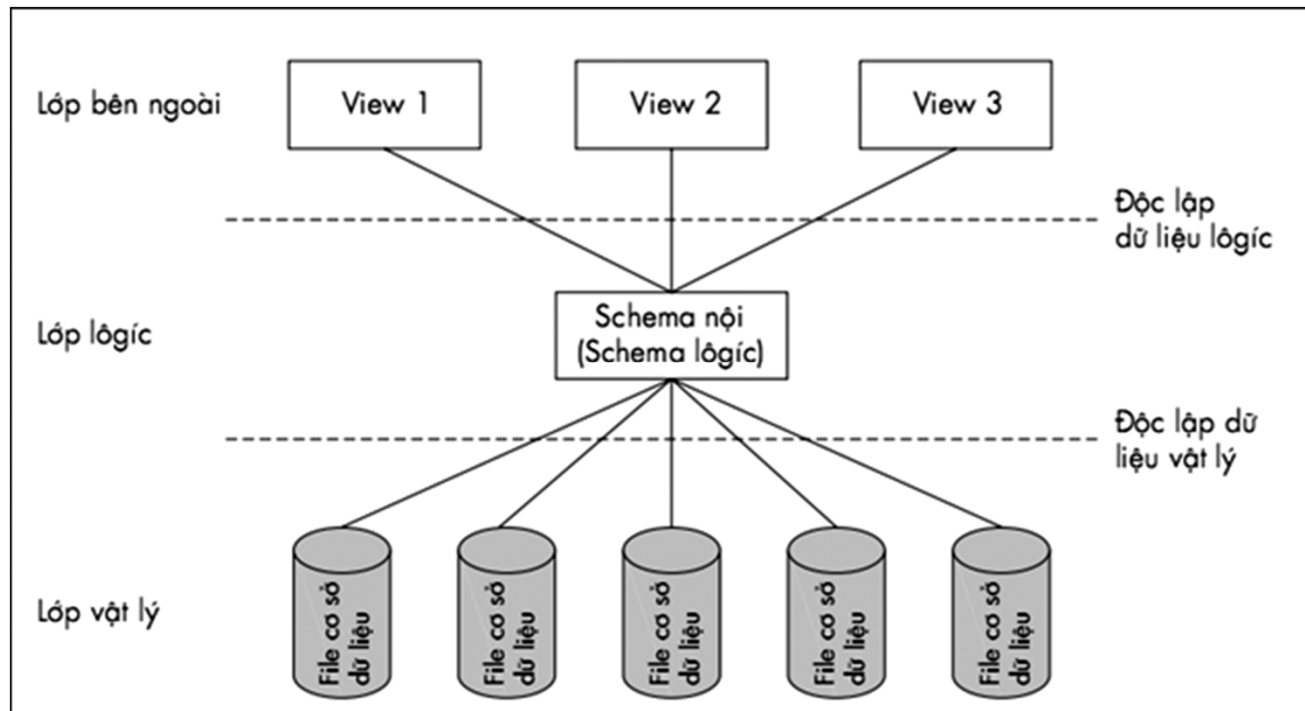
- CSDL có cấu trúc dễ dàng truy cập, quản lý và cập nhật

→ Cần phải quản trị CSDL

## ■ Ví dụ minh họa CSDL:



- CSDL cung cấp khả năng trừu tượng hóa dữ liệu thông qua các lớp. Bao gồm 3 lớp: Lớp vật lý, Lớp logic, Lớp bên ngoài.
- Sự phân biệt giữa các lớp tạo nên 2 tính năng: tính độc lập dữ liệu vật lý và tính độc lập dữ liệu logic.





## ■ Lập v t lý:

- Lập v t lý ch a toàn b các file d li u
- Ng i dùng CSDL không nh t thi t ph i n m c c u trúc t ch c c a các file d li u v t lý khi s d ng c s d li u.

## ■ Logic (còn gọi là schema logic):

- Cấu trúc dữ liệu trừu tượng được tổ thành thành lập vật lý.
- Logic có thể chia thành hai phần các bảng hai chiều, một cấu trúc phân cấp tổ chức như sơ đồ tổ chức của một công ty hay một sơ cấu trúc khác

## ■ Lập bên ngoài:

- Bao gồm các khung nhìn (view) và các g i chung là subschema
- Khung nhìn: là khái niệm cho phép người dùng quan sát dữ liệu theo những cách khác nhau **v cùng một dữ liệu để trình bày và tổ chức.**

## ■ clip dữ liệu vật lý:

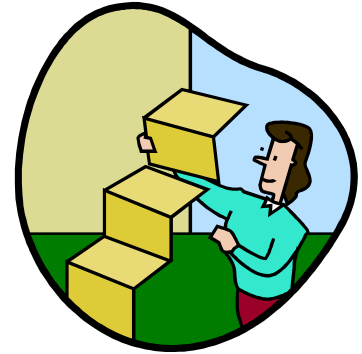
- Khả năng thay đổi cấu trúc file vật lý của một cơ sở dữ liệu mà không làm gián đoạn **ng** **i dùng hi n t i** và **các quá trình ang** **di n ra** **c g i là** clip dữ liệu vật lý
- Việc phân biệt clip vật lý và clip logic tạo ra sự clip dữ liệu vật lý

## ■ Các lập luận logic:

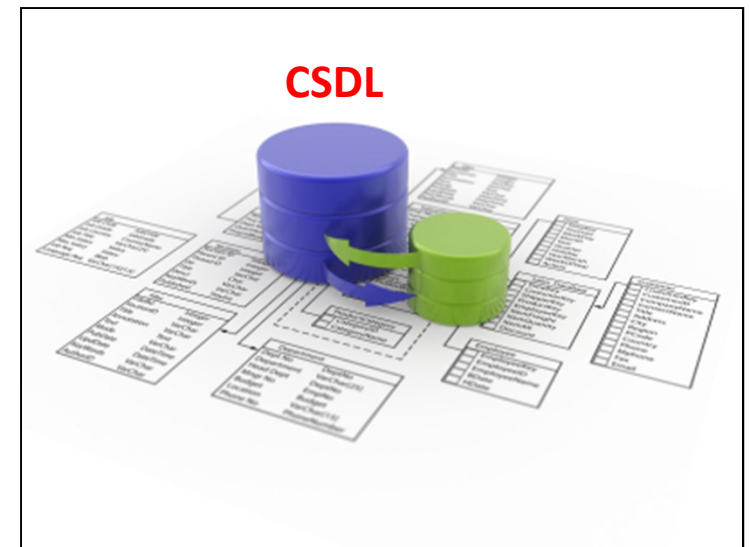
- Khi cần ng t o ra các thay i t i l p logic mà không làm gián o n ng i dùng hi n t i và các quá trình ang di n ra c g i là c l p d li u logic.
- Vì c phân bi t gi a l p logic và l p bên ngoài t o ra tính c l p d li u logic

- **Quản lý dữ liệu**: là quản lý m t s l ng l n d li u, bao g m c vi c l u tr và cung c p c ch cho phép **Thao tác** (thêm, s a, xóa d li u) và **Truy v n d li u**.
- 2 ph ng pháp quản lý dữ liệu:
  - H th ng quản lý b ng file
  - H th ng quản lý b ng CSDL

- Dữ liệu được lưu trữ trong các file riêng biệt
- Ví dụ: các chương trình lưu trữ thông tin bằng hình thức các file dạng text
- Những điểm cần lưu ý về quản lý lưu trữ file:
  - Đồng nhất và mâu thuẫn dữ liệu
  - Kiểm tra hiệu suất trong truy xuất ngẫu nhiên hoặc xử lý hàng loạt
  - Dữ liệu lưu trữ rỗng
  - Giảm thiểu rủi ro an toàn và bảo mật



- Quản lý dữ liệu bằng CSDL giúp dữ liệu có lưu trữ một cách hiệu quả và có tổ chức, cho phép quản lý dữ liệu nhanh chóng và hiệu quả
- Lợi ích của hệ thống quản lý bằng CSDL:
  - Tránh dữ liệu thừa, trùng lặp dữ liệu
  - Giảm bộ nhớ lưu trữ trong CSDL
  - Các dữ liệu lưu trữ có thể chia sẻ
  - Có thể thiết lập các chuẩn trên dữ liệu
  - Duy trì tính toàn vẹn dữ liệu
  - Giảm bộ nhớ lưu trữ



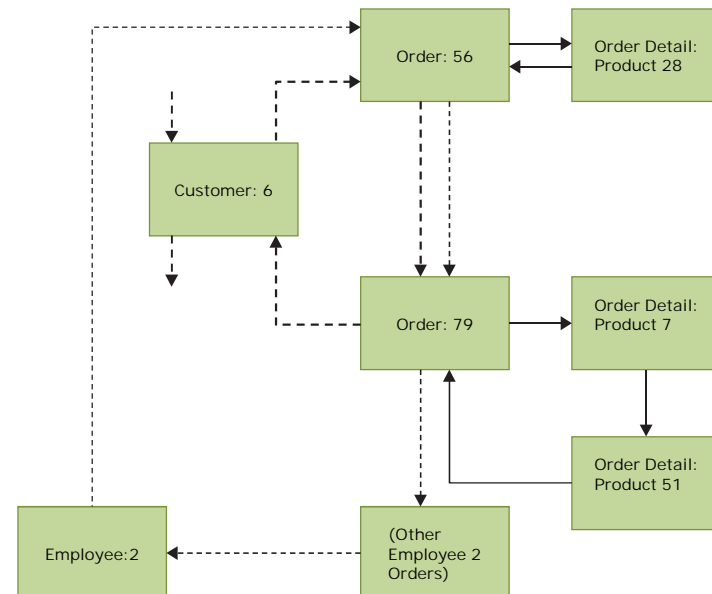


- Các CSDL có thể khác nhau về chức năng và mô hình dữ liệu (data model).
- Mô hình dữ liệu sẽ quy định cách thức lưu trữ và truy cập dữ liệu.
- Tùy từng ngôn ngữ quan hệ giữa các thành phần dữ liệu trong CSDL, mô hình phân cấp áp dụng vì cách lưu trữ và truy xuất dữ liệu tối ưu cao nhất.

## ■ Các mô hình:

- Mô hình dữ liệu file phẳng (Flat file)
- Mô hình dữ liệu mạng (Network model)
- Mô hình dữ liệu phân cấp (Hierarchical model)
- Mô hình dữ liệu quan hệ (Relational model)
- Mô hình dữ liệu hướng đối tượng (Object-Oriented model)

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



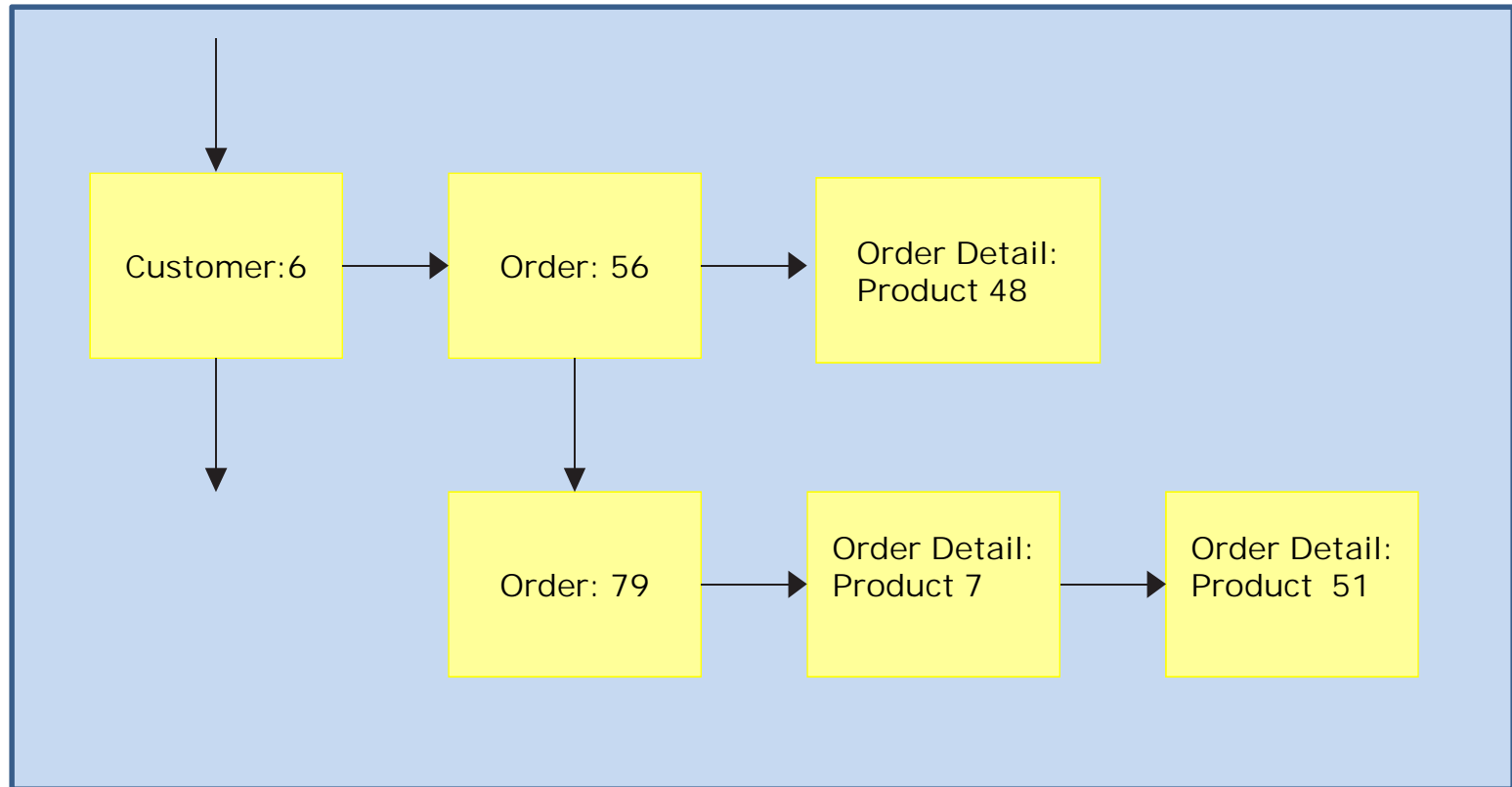
- Mô hình này chỉ dùng cho các CSDL ngôn ngữ
- CSDL dạng file phông thường là file kiểu văn bản chứa dữ liệu dạng bảng

- Ví dụ: một file phông thể hiện thông tin về Customer (Khách hàng) đã đi đăng ký công ty Northwind Traders

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

- Tổ chức theo hình cây, mỗi nút biểu diễn một thực thể dữ liệu.
- Liên hệ dữ liệu thể hiện trên liên hệ giữa nút cha và nút con. Mỗi nút cha có thể có một hoặc nhiều nút con, nhưng mỗi nút con chỉ có thể có một nút cha.
- Do đó mô hình phân cấp thể hiện các kiểu quan hệ :
  - 1-1
  - 1-N

- Ví dụ : mô hình dữ liệu phân cấp của CSDL Northwind



## ■ Hình thức :

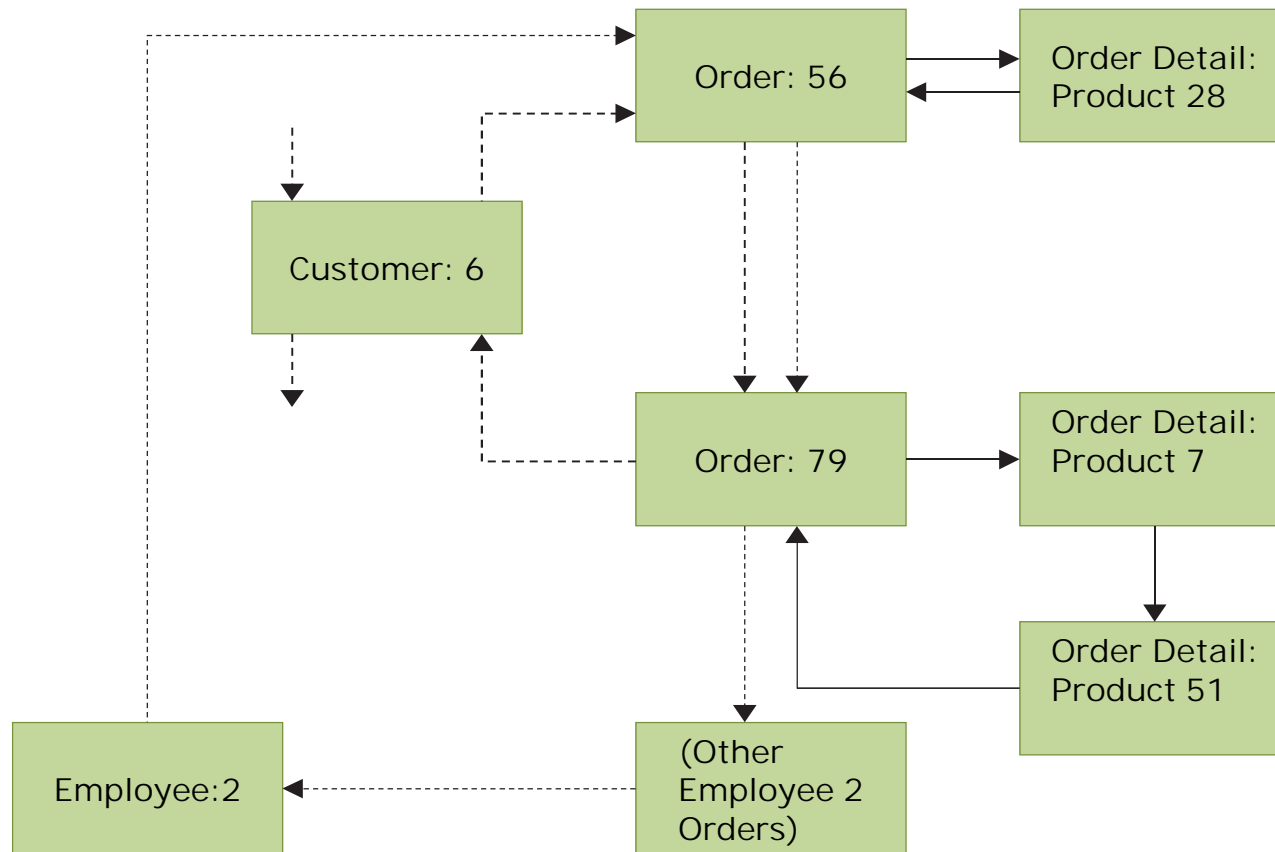
- Một nút con không thể có quá một nút cha -> Không biểu diễn được các quan hệ dữ liệu phức tạp

## ■ Cách tổ chức:

- Các file riêng biệt trong hệ thống file phẳng có nghĩa là các bản ghi. Tập hợp bản ghi cùng kiểu tạo thành một kiểu thực thể dữ liệu.
- Các kiểu thực thể kết nối với nhau thông qua mối quan hệ cha-con.
- Mô hình dữ liệu mạng biểu diễn bất kỳ một thực thể nào, và các mối tên chính thức thực thể cha sang kiểu thực thể con.



## ■ Ví dụ : Mô hình dữ liệu của CSDL Northwind



## ■ Tính chất:

- Ưu điểm:

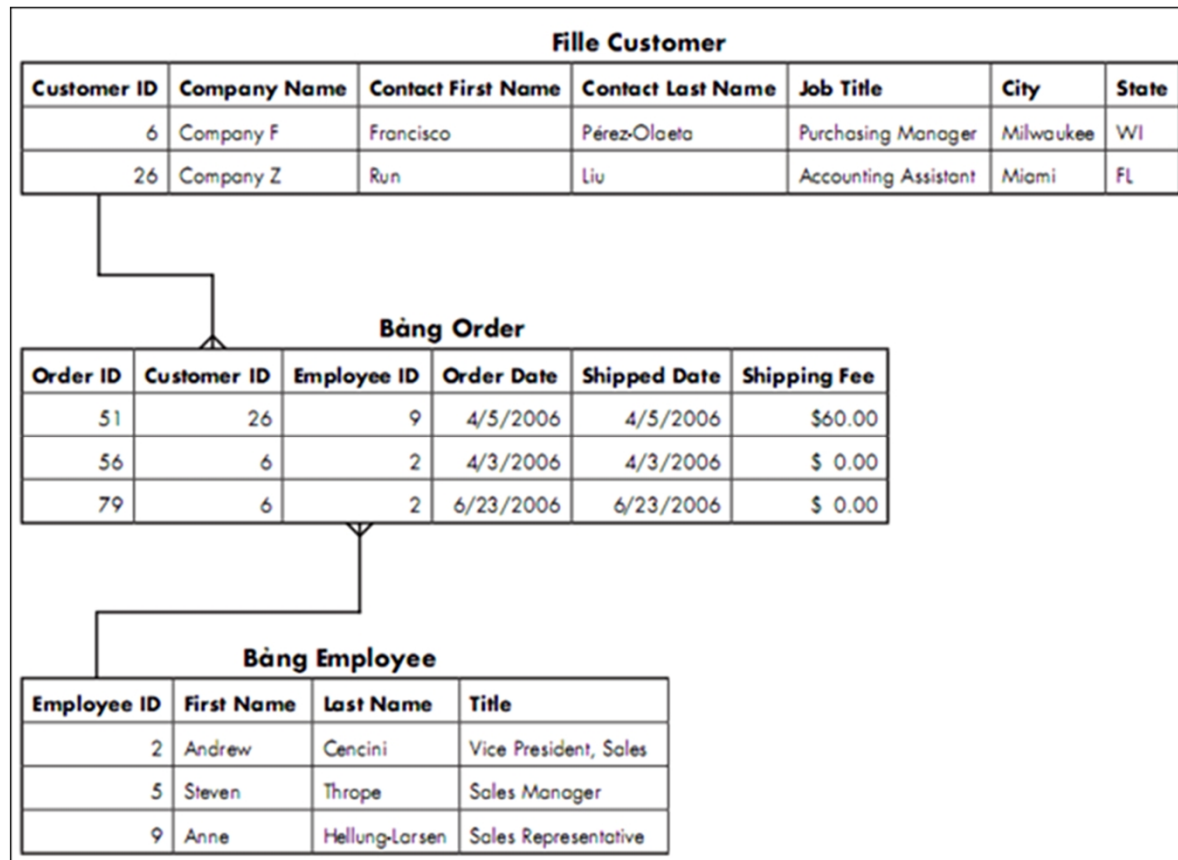
- Dễ biểu diễn mô hình
- Dễ nhận thấy các liên hệ dữ liệu phức tạp

- Nhược điểm:

- Truy xuất chậm
- Không thích hợp với các CSDL có quy mô lớn

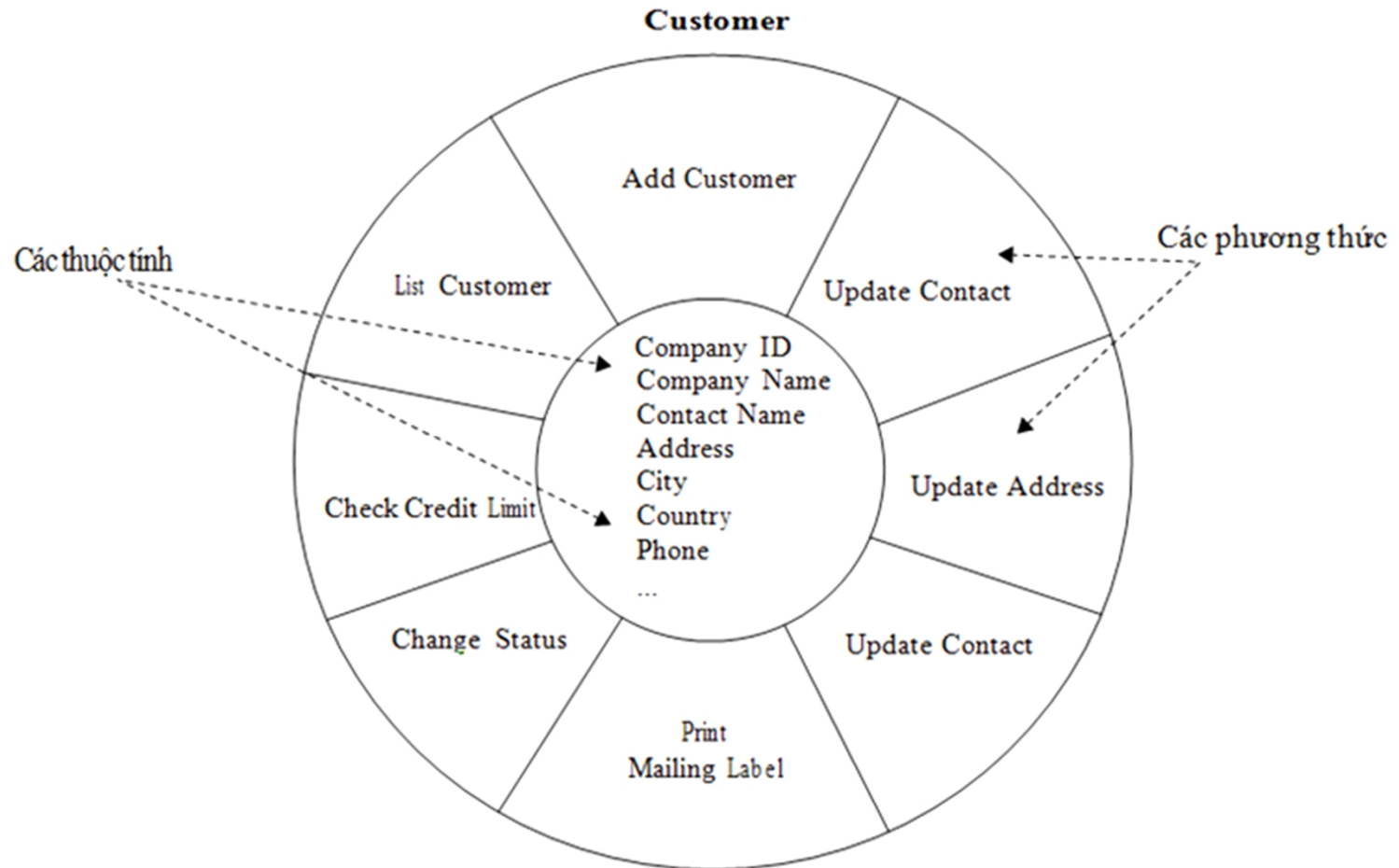
- Trong mô hình dữ liệu quan hệ, không có các liên kết vật lý. Dữ liệu được biểu diễn dưới dạng bảng với các hàng và các cột:
  - CSDL là tập hợp các bảng (còn gọi là quan hệ)
  - Mỗi hàng là một bản ghi (record), còn mỗi cột là bộ (tuple)
  - Mỗi cột là một thuộc tính, còn mỗi giá trị là trường (field)
- Dữ liệu trong hai bảng liên hệ với nhau thông qua các cột chung.
- Có các toán tử thao tác trên các hàng của bảng.

- Ví dụ: mô hình dữ liệu quan hệ trong CSDL Northwind gồm 3 bảng: Customer, Order, Employee



- Ra đi vào kho ngữ luận m 90, 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 sinh ra một vài cá thể khác biệt đối tượng khác, n p ch ng (hay nh ngh a l i) phương thức của đối tượng khác...

## ■ Ví dụ mô hình dữ liệu nghiệp vụ "Customer"



- Các mô hình CSDL xác định các hình thức tổ chức lưu trữ và truy cập dữ liệu.
- Hệ quản trị CSDL (DataBase Management System – DBMS) là các phần mềm giúp tạo các CSDL và cung cấp cách lưu trữ, truy cập theo các mô hình CSDL.
- Ví dụ :
  - SQL Server, Microsoft Access, Oracle là các hệ quản trị CSDL hiện hành cho mô hình quan hệ.
  - IMS của IBM là hệ quản trị CSDL cho mô hình phân cấp
  - IDMS là hệ quản trị CSDL cho mô hình mạng

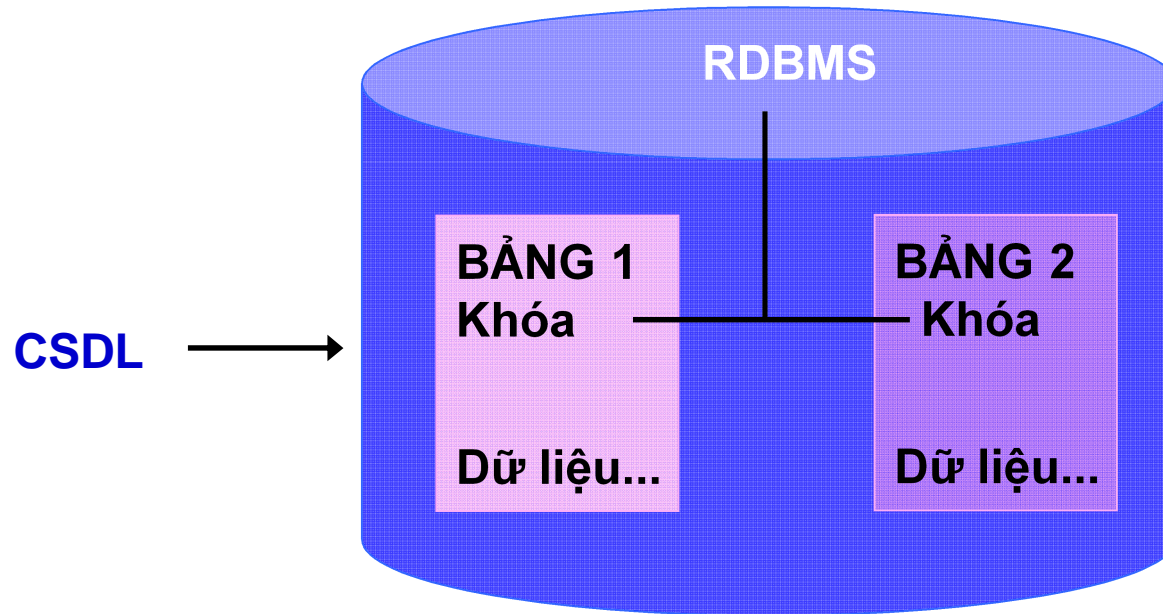
## ■ Những lợi ích DBMS mang lại:

- Quản trị các CSDL
- Cung cấp giao diện truy cập để đưa các tính phức tạp về mặt cấu trúc tổ chức dữ liệu về mặt lý
- Hỗ trợ các ngôn ngữ giao tiếp. Ví dụ:
  - Ngôn ngữ mô tả, định nghĩa dữ liệu – DDL
  - Ngôn ngữ thao tác dữ liệu – DML
  - Ngôn ngữ truy vấn dữ liệu có cấu trúc – SQL
- Có các chức năng bảo mật, bảo toàn, bảo mật cao



- Hệ quản trị CSDL quan hệ (Relational DataBase Management System = RDBMS)
- RDBMS là một dạng DBMS sử dụng phương tiện nhất, trong đó tổ chức dữ liệu có thể thực hiện được các **biến đổi** dữ liệu.
- Thực hiện các thao tác trên CSDL chủ yếu diễn ra trên các **biến đổi**.

- Minh họa hệ quản trị CSDL quan hệ



- **Người dùng tham gia vào hệ thống RDBMS:**
  - Người quản trị CSDL (**DataBase Administrator**)
  - Người thiết kế CSDL (**DataBase Designer**)
  - Người phân tích hệ thống (**System Analyst**)
  - Người lập trình ứng dụng (**Application Programmer**)
  - Người thiết kế và triển khai CSDL (**DBMS Designer and Implementer**)
  - Người dùng cuối (**End User**)



- D ường nh ngh a, duy trì và thao tác d li u l u tr
- D ường trích xu t d li u
- D li u c chu n hóa và c b o v t t
- Nhi u nhà cung c p cung c p ph n m m
- D ường chuy n i gi a nhà cung c p và nhà tri n khai
- RDBMS là các s n ph m tr ng thành và n nh

- **CSDL** là tập hợp dữ liệu liên quan với nhau được lưu trữ có cấu trúc. Người dùng có thể dễ dàng cập nhật dữ liệu hoặc trích xuất thông tin từ CSDL.
- Tập hợp CSDL tạo ra các **lớp trừu tượng CSDL**: lớp vật lý, lớp logic và lớp bên ngoài.
- Ban đầu dữ liệu lưu trữ rời rạc dưới dạng các file, gọi là mô hình dữ liệu file phôi.
- Sau đó, các **mô hình dữ liệu** khác nhau cho phép mô tả cách thức lưu trữ dữ liệu và cách thức truy cập dữ liệu dễ dàng.

- Hệ quản trị CSDL (DBMS) là tập hợp các chương trình cho phép người dùng lưu trữ, cập nhật và trích xuất thông tin từ CSDL.
- Hệ quản trị CSDL quan hệ (RDBMS) là tập hợp các chương trình cho phép tạo và thao tác với CSDL quan hệ.
- Có nhiều loại ngôn ngữ dùng RDBMS như: hệ quản trị CSDL, thiết kế CSDL, phân tích và thiết kế ngôn ngữ, cài đặt CSDL, ngôn ngữ truy vấn.

XIN C M N!