



# Hệ quản trị CSDL

---

- Trình bày: Đỗ Tuấn Anh
- Bộ môn KHMT
- Viện CNTT và Truyền Thông



# Mục đích môn học

---

- Tìm hiểu nguyên lý và chức năng của Hệ QT CSDL
- Xây dựng ứng dụng trên nền hệ quản trị csdl



# Yêu cầu

---

- Xây dựng được hệ thống thông tin quản lý dựa trên thiết kế
- Đảm bảo toàn vẹn và an toàn dữ liệu
- Nắm bắt được xu hướng phát triển các ứng dụng CSDL
- Hiểu về phân tích dữ liệu, khai phá tri thức trên CSDL



# Tài liệu

---

- Ramakrishnan.R., Gehrke J. *Database Management Systems*. McGraw Hill Professional, 2003
- Beynon-Davies P. *Database Systems*. Palgrave MacMillan, London, 2003
- Dejan Šunderi, Microsoft SQL Server 2005 Stored Procedure Programming in T-SQL & .NET, McGraw-Hill/Osborne, 2006
- Jay Greenspan, My SQL /PHP Database Applications, M&T Books, 2001
- Jason Price, *Oracle Database 11g SQL*, Oracle Press, McGrawHill, 2008
- David J. Gallardo, *Java Oracle Database Development*, Prentice Hall, 2003



# Nhiệm vụ đối với sinh viên

---

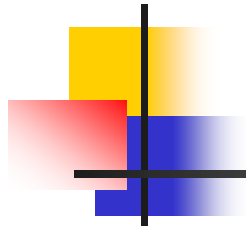
- Dự lớp: đầy đủ
- Bài tập: Hoàn thành bài tập môn học
- Phương pháp học
  - Hiểu rõ lý thuyết
  - Thực hành bài tập cụ thể
- Hình thức đánh giá
  - Điểm quá trình: Điểm dự lớp + Kiểm tra
  - Điểm cuối kỳ



# Nội dung môn học

---

- Chương 1: Khái niệm cơ bản
- Chương 2: Ngôn ngữ SQL
- Chương 3: Xây dựng ứng dụng CSDL truyền thống
- Chương 4: Xu hướng phát triển Hệ thống CSLD
- Chương 5: Một số ứng dụng nâng cao



# Chương 1: Khái niệm cơ bản

---

- Mục đích:
  - Các khái niệm về CSLD, HQT CSLD
  - Các yêu cầu khi xây dựng HQTCSDL
- Yêu cầu
  - Nắm bắt các khái niệm
  - Hiểu các vấn đề khi XD một HQTCSDL: thiết kế, tính chất nhất quán.
  - Cấu trúc của hệ thống HQTCSDL
  - Vai trò của người dùng trong hệ thống



# 1. Khái niệm

---

- CSDL (Database) là một tập hợp các thông tin có liên quan với nhau, được thiết kế nhằm làm giảm thiểu sự lặp lại dữ liệu
- HQT CSDL (DBMS) là một hệ thống gồm một CSDL và các thao tác trên CSDL, được thiết kế trên một nền tảng phần cứng, phần mềm và với một kiến trúc nhất định.
- Hệ CSDL (DBS) là một phần mềm cho phép xây dựng một HQT CSDL.

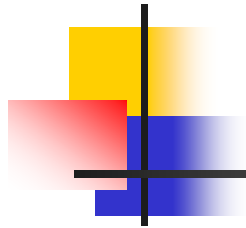




## 2. Hệ Cơ sở dữ liệu

---

- Tại sao lại xuất hiện Hệ QTCSDL?
- Hệ thống xử lý lưu trữ thông thường
  - Dư thừa dữ liệu và tính không nhất quán (Data redundancy and inconsistency)
  - Khó khăn trong việc truy xuất dữ liệu
  - Sự nhất quán dữ liệu (Data isolation)
  - Các vấn đề về tính nguyên tử (Atomicity problems)
  - Tính bất thường trong truy xuất đồng thời
  - Vấn đề an toàn (Security problems)
- ➔ Mở ra sự phát triển Hệ QTCSDL

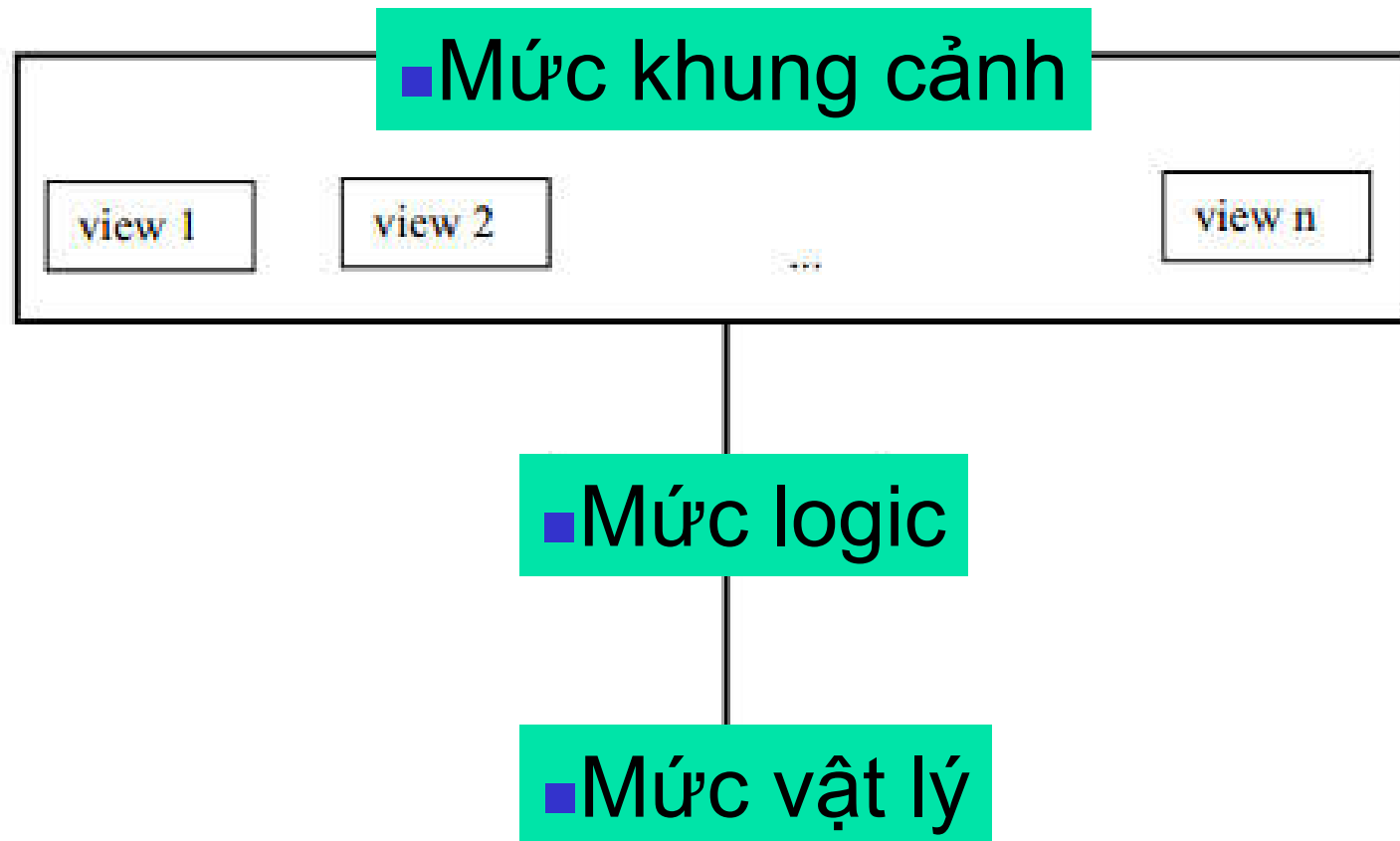


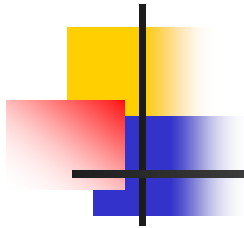
### 3. Góc nhìn dữ liệu

---

- Tính hiệu quả của hệ thống đòi hỏi phải thiết kế cấu trúc dữ liệu phức tạp để biểu diễn dữ liệu trong CSDL
- Các mức độ
  - Mức Vật lý (Physical level)
  - Mức Logic (Logical level)
  - Mức khung cảnh (View level)

### 3. Góc nhìn dữ liệu





# 3. Góc nhìn dữ liệu

- Thể hiện và sơ đồ (instances and schemas)
  - Tập hợp các thông tin được lưu trữ trong CSDL tại một thời điểm được gọi là một thể hiện
  - Thiết kế tổng thể của CSDL được gọi là sơ đồ
- Sự độc lập dữ liệu (data independence) thể hiện ở khả năng sửa đổi một định nghĩa không ảnh hưởng tới một định nghĩa sơ đồ ở mức cao hơn
  - Độc lập dữ liệu vật lý
  - Độc lập dữ liệu lôgics



## 4. Mô hình dữ liệu

---

- Nằm dưới cấu trúc của một cơ sở dữ liệu là mô hình dữ liệu bao gồm công cụ
  - Mô tả dữ liệu
  - Mô tả quan hệ dữ liệu
  - Mô tả ngữ nghĩa dữ liệu
  - Mô tả ràng buộc dữ liệu
- Các nhóm mô hình
  - Mô hình logic dựa trên đối tượng
  - Mô hình logic dựa trên các bản ghi
  - Mô hình vật lý



## 5. Ngôn ngữ CSDL

---

- Ngôn ngữ định nghĩa dữ liệu (Data Definition Language: DDL)
  - cho phép định nghĩa sơ đồ CSDL.
  - kết quả biên dịch các lệnh của DDL là được lưu trữ trong một file đặc biệt được gọi là từ điển dữ liệu (data dictionary) hay thư mục dữ liệu (data directory)



## 5. Ngôn ngữ CSDL

---

- Ngôn ngữ thao tác dữ liệu (Data manipulation language: DML)
  - Cho phép người sử dụng truy xuất hoặc thao tác dữ liệu.
  - Có hai kiểu ngôn ngữ thao tác dữ liệu
    - DML thủ tục (procedural DML): Người dùng đặc tả dữ liệu và nhận dữ liệu
    - DML không thủ tục (Nonprocedural DML): Người dùng chỉ cần đặt tả dữ liệu vd: query



## 6. Giao dịch và quản lý giao dịch

---

- Xét bài toán chuyển khoản
  - Chuyển số tiền  $x$  từ tài khoản A vào tài khoản B
  - Các công việc cần thực hiện
    - $A = A - x$
    - $B = B + x$
  - Thao tác chuyển khoản  $\rightarrow$  giao dịch
  - Kết quả giao dịch: Thành công hoặc Không





## 6. Giao dịch và quản lý giao dịch

---

- Tính nguyên tử: giao dịch phải xảy ra toàn thể hoặc không
- Tính nhất quán: một số yếu tố cần đảm bảo nguyên vẹn trước và sau giao dịch.  
VD:  $A+B$
- Tính bền vững: các giá trị phải bền vững cho dù có thể có sự cố hệ thống



## 6. Giao dịch và quản lý giao dịch

---

- → Giao dịch: một tập các hoạt động thực hiện chỉ một chức năng logic trong một ứng dụng cơ sở dữ liệu
- Mỗi giao dịch là một đơn vị mang tính nguyên tử và nhất quán
- → CSDL là nhất quán khi một giao dịch khởi động thì nó cũng phải là nhất quán khi giao dịch kết thúc



## 6. Giao dịch và quản lý giao dịch

---

- Hệ CSDL cần có thành phần quản lý giao dịch (transaction-management component)
  - Nếu giao dịch thành công → tính nguyên tử
  - Nếu giao dịch 0 thành công và tính nguyên tử được đảm bảo → trạng thái CSDL trước và sau là giống nhau
  - Khi một số giao dịch cạnh tranh cập nhật dữ liệu → tính nhất quán → TP quản lý điều khiển cạnh tranh



## 7. Thành phần Quản lý lưu trữ

---

- Thông tin CSDL lưu trữ → Bộ nhớ ngoài
- Vận chuyển dữ liệu từ bộ ngoài vào bộ nhớ xử lý sẽ chậm → Tối ưu cấu trúc dữ liệu → Đáp ứng truy xuất dữ liệu
- Bộ quản trị lưu trữ (storage manager) là một module chương trình cung cấp giao diện giữa dữ liệu mức thấp được lưu trữ trong CSDL với các chương trình ứng dụng thông qua truy vấn.



## 8. NHÀ QUẢN TRỊ CƠ SỞ DỮ LIỆU

---

- Người điều khiển trung tâm trên toàn hệ thống gọi là nhà quản trị CSDL (DataBase Administrator - DBA)
- Chức năng
  - Định nghĩa sơ đồ
  - Định nghĩa cấu trúc lưu trữ và phương pháp truy xuất
  - Sửa đổi sơ đồ và tổ chức vật lý
  - Cấp quyền truy xuất dữ liệu
  - Đặc tả ràng buộc toàn vẹn



## 9. Đối tượng sử dụng CSDL

---

- Mục đích của Hệ CSDL là cung cấp môi trường để tìm, lưu trữ thông tin. Vì vậy có thể phân thành 4 nhóm theo cách thức trao đổi với hệ thống
  - Nhóm lập trình ứng dụng: DML
  - Nhóm sử dụng thành thạo: Truy vấn trực tiếp
  - Nhóm cấp cao: Sử dụng tính năng nâng cao được hỗ trợ để khai thác trong các hệ thống Thiết kế, tri thức, hệ chuyên gia, mô phỏng
  - Nhóm cấp thấp: không thành thạo, trao đổi với hệ thống qua cầu dẫn



## 10. Cấu trúc hệ thống

---

- hệ CSDL được phân thành các module, mỗi một thực hiện một nhiệm vụ
- Một số chức năng của hệ CSDL có thể được cung cấp bởi hệ điều hành
- Hầu hết, các thành phần của hệ CSDL sử dụng các dịch vụ cơ bản của HDH
- →thiết kế hệ CSDL phải xem xét đến giao diện giữa hệ CSDL và hệ điều hành



# 10. Cấu trúc hệ thống

---

- Các thành phần chức năng của hệ CSDL
  - xử lý truy vấn (query processor components)
  - quản lý lưu trữ(storage manager components )





# 10. Cấu trúc hệ thống

---

- Xử lý truy vấn
  - Trình biên dịch DML (DML compiler)
  - Trình tiền biên dịch DML nhúng (Embedded DML Precompiler)
  - Bộ thông dịch DDL (DDL interpreter)
  - Engine đánh giá truy vấn (Query evaluation engine)



# 10. Cấu trúc hệ thống

---

- quản lý lưu trữ
  - Bộ quản trị quyền và tính toàn vẹn (Authorization and integrity manager)
  - Bộ quản trị giao dịch (Transaction manager)
  - Bộ quản trị file (File manager)
  - Bộ quản trị bộ đệm (Buffer manager)

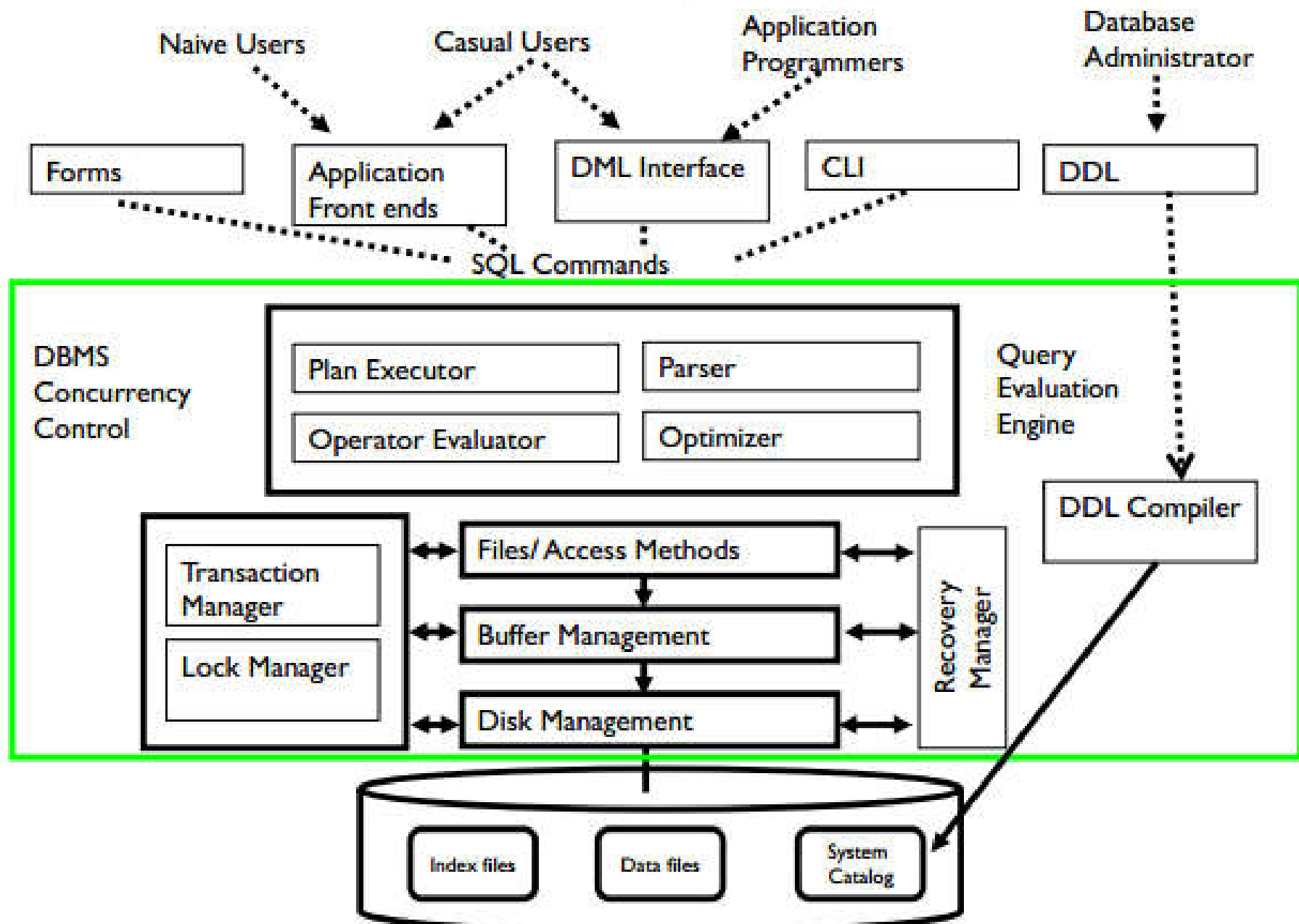


# 10. Cấu trúc hệ thống

---

- Mối số thành phần cấu trúc hỗ trợ cho hệ thống vật lý
  - Các file dữ liệu
  - Từ điển dữ liệu (Data Dictionary)
  - Chỉ mục ( Indices )
  - Dữ liệu thống kê ( Statistical data )

# Kiến trúc của một hệ QTCSDL





# 11. Kiến trúc hệ cơ sở dữ liệu

---

- Kiến trúc hệ CSDL liên quan đến hệ thống máy tính
- Các kiến trúc hệ thống máy
  - Kiến trúc mạng
  - Kiến trúc song song
  - Kiến trúc phân tán



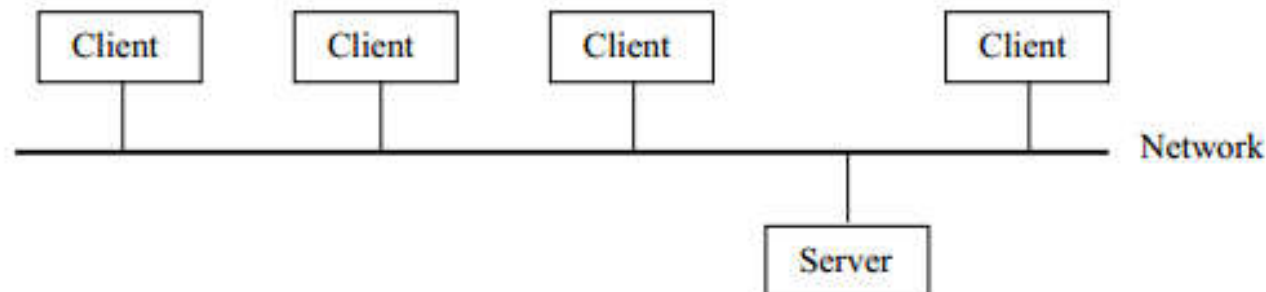
## 12. Hệ thống tập trung

---

- Hệ CSDL tập trung chạy trên máy đơn và không trao đổi với máy khác
- Hệ CSDL chạy trên một máy đơn hay Mainframe
- Đặc điểm
  - Truy xuất và vùng nhớ chung
  - Tạo các bản sao giống nhau → tăng tốc độ

# 13. Hệ thống Client – Server

- Hệ thống server giao dịch (transaction-server systems) bao gồm hai thành phần
  - phần trước (front-end): Công cụ cho người dùng
  - Phần sau (back-end): quản trị cấu trúc truy vấn
- Kết nối Client-Server
  - ODBC (Open DataBase Connectivity): giao diện trình ứng dụng cho phép các clients sinh ra các lệnh SQL và gửi đến một server
  - Giao diện kết nối khác (transactional remote procedure calls)









# Chương 2: SQL

---





# Kiến thức nền tảng về Cơ sở dữ liệu

---

## **Bài 1**



# Mục tiêu của bài học (1)

---

- Giải thích vì sao chúng ta cần có cơ sở dữ liệu
- Giải thích sự phát triển của các hệ quản trị cơ sở dữ liệu
- Liệt kê các mô hình cơ sở dữ liệu khác nhau
- Mô tả các khái niệm cơ bản liên quan đến cơ sở dữ liệu
  - Concepts of RDBMS and



## Mục tiêu của bài học (2)

---

- Mô tả các khái niệm cơ bản liên quan đến cơ sở dữ liệu quan hệ
  - Toàn vẹn thực thể (entity integrity)
  - Khóa chính (primary key)
  - Khóa ngoại (foreign key)
  - Mỗi quan hệ và bậc của nó (relation and its degree)
- Mô tả tóm tắt về SQL

■ Concepts of RDBMS and

# Cơ sở dữ liệu là gì?

Một **cơ sở dữ liệu (CSDL - Database)** là dạng được sắp xếp của các dữ liệu, được xây dựng từ một hoặc nhiều mục thông tin có liên quan đến nhau gọi là các **bản ghi (record)**.



Một CSDL giống như một tập hợp dữ liệu và cho phép ta có thể đặt nhiều câu hỏi khác nhau với nó.

Ví dụ, “Số điện thoại và địa chỉ của 5 bưu điện gần trường nhất là gì?”

■ Concepts of RDBMS and



# Vì sao lại phải dùng CSDL?

Các CSDL được dùng để lưu trữ dữ liệu theo 1 cách thức có tổ chức và hiệu quả. CSDL cho phép chúng ta quản lý dữ liệu nhanh chóng và dễ dàng.

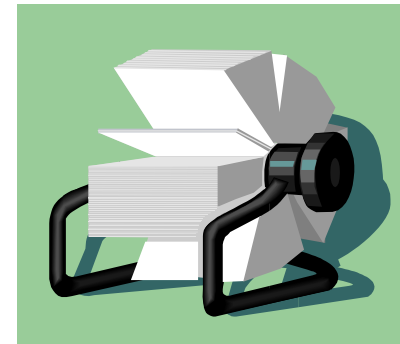
Ví dụ, 1 công ty có thể lưu các thông tin chi tiết về nhân viên của mình trong 1 CSDL.

FirstName	LastName	Hiredate
Karla	Jablonski	1994-03-11
Martine	Rance	1992-02-05
Margaret	Smith	1988-09-29
Gary	Thomas	1988-08-09

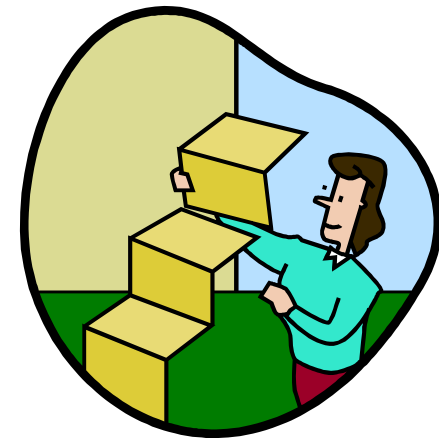
■ Concepts of RDBMS and

# Nhu cầu về tính bền vững dữ liệu

Ta thậm chí có thể dùng các tệp thủ công để lưu trữ dữ liệu.



Tuy nhiên, dữ liệu được lưu trữ theo dạng này là không bền vững.



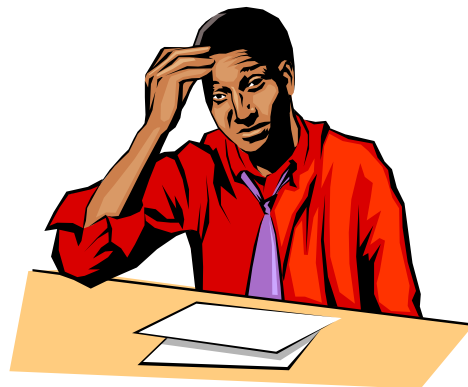
■ Concepts of RDBMS and



# Vấn đề về cách lưu trữ theo kiểu tệp

Quy trình lưu trữ dữ liệu bằng các tệp hồ sơ:

- Tẻ nhạt
- Tốn thời gian
- Nguy cơ lỗi cao



■ Concepts of RDBMS and

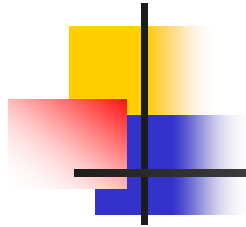


## Ưu điểm của CSDL được tin học hóa

---

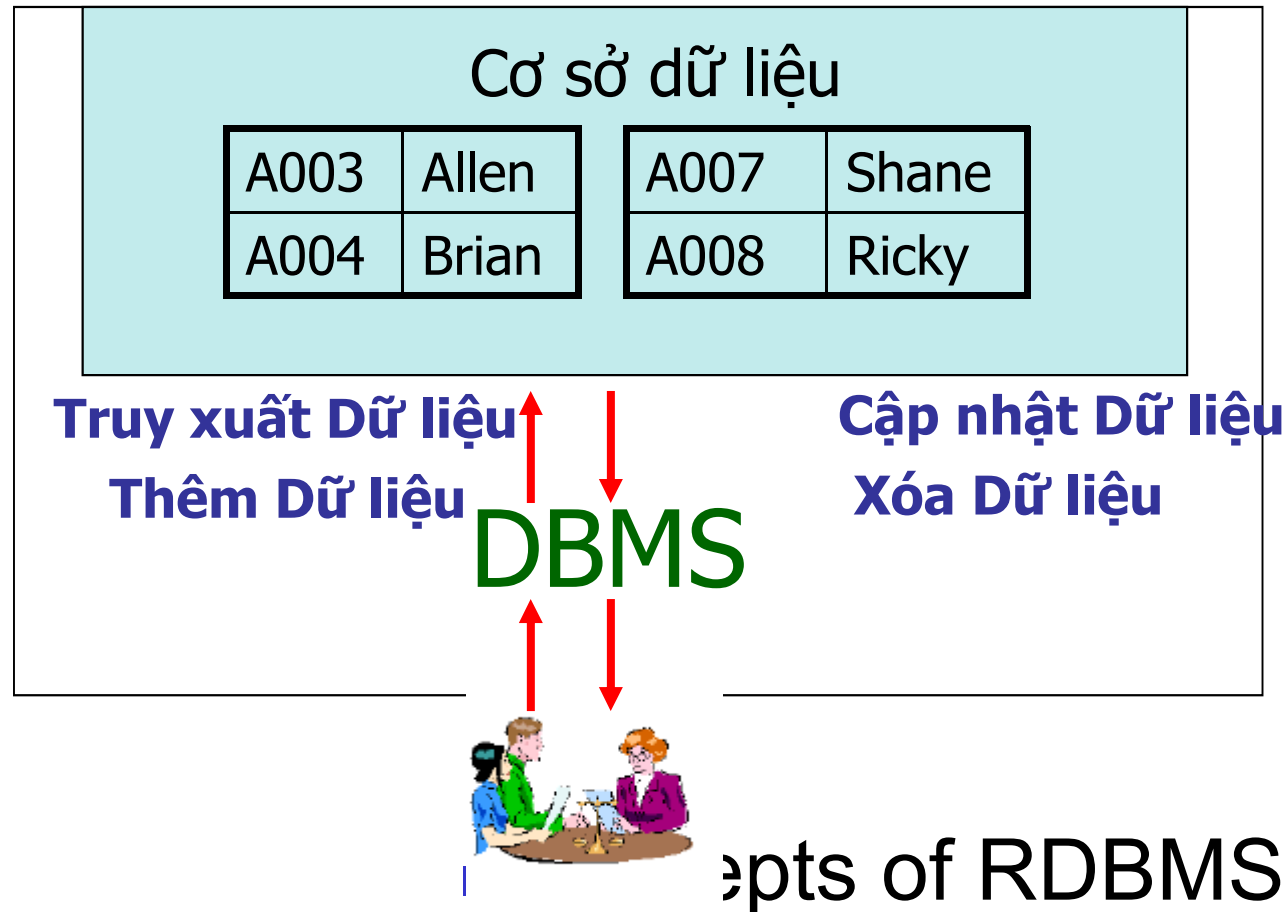
- Giảm thiểu được sự dư thừa trong các dữ liệu được lưu trữ
- Không còn sự thiếu nhất quán trong dữ liệu
- Có thể chia xẻ dữ liệu đã lưu trữ
- Có thể đặt ra các tiêu chuẩn và phải tuân thủ các tiêu chuẩn đặt ra khi thao tác với dữ liệu
- Có thể duy trì được sự nhất quán dữ liệu
- Có thể thực hiện được việc bảo mật dữ liệu

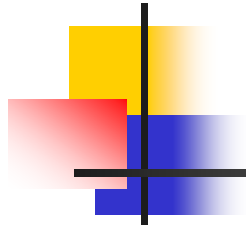
■ Concepts of RDBMS and



# Sự phát triển của DBMS-1

DBMS là gì?

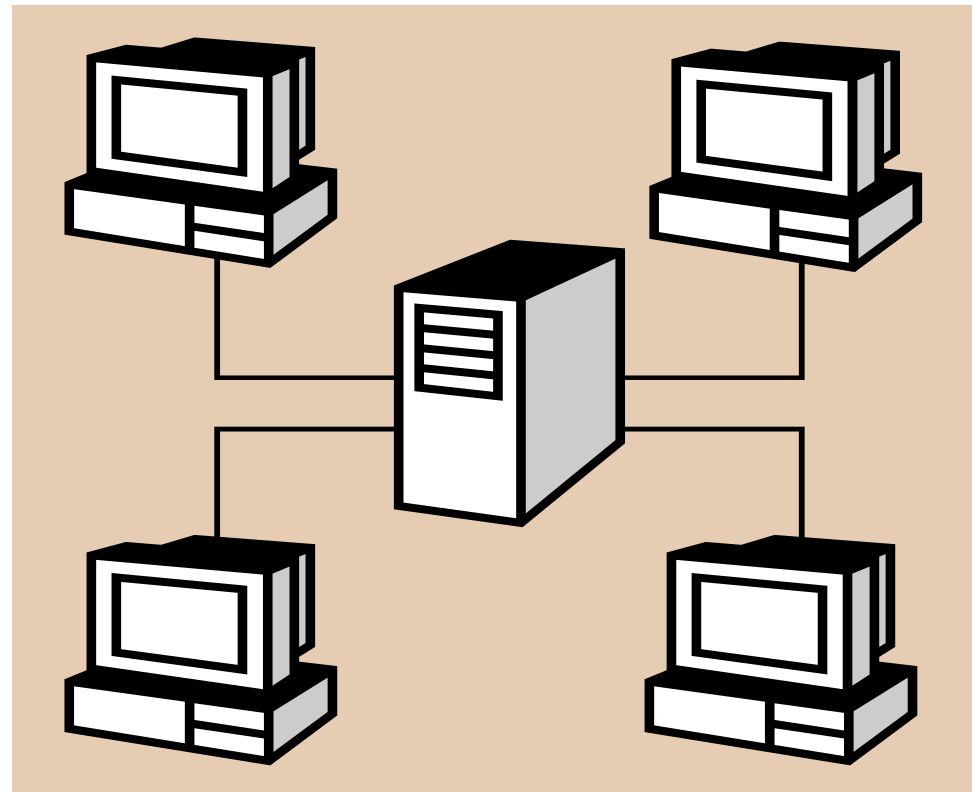




# Sự phát triển của DBMS-2

## Hệ CSDL tập trung

- Dữ liệu được lưu tại 1 vị trí trung tâm chung.
- Mọi người sử dụng được phép truy cập vào CSDL chung



■ Concepts of RDBMS and



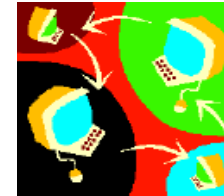
# Các mô hình CSDL

---

- Mô hình phân cấp



- Mô hình mạng lưới



- Mô hình quan hệ

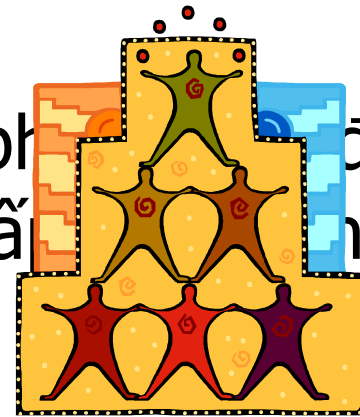


■ Concepts of RDBMS and



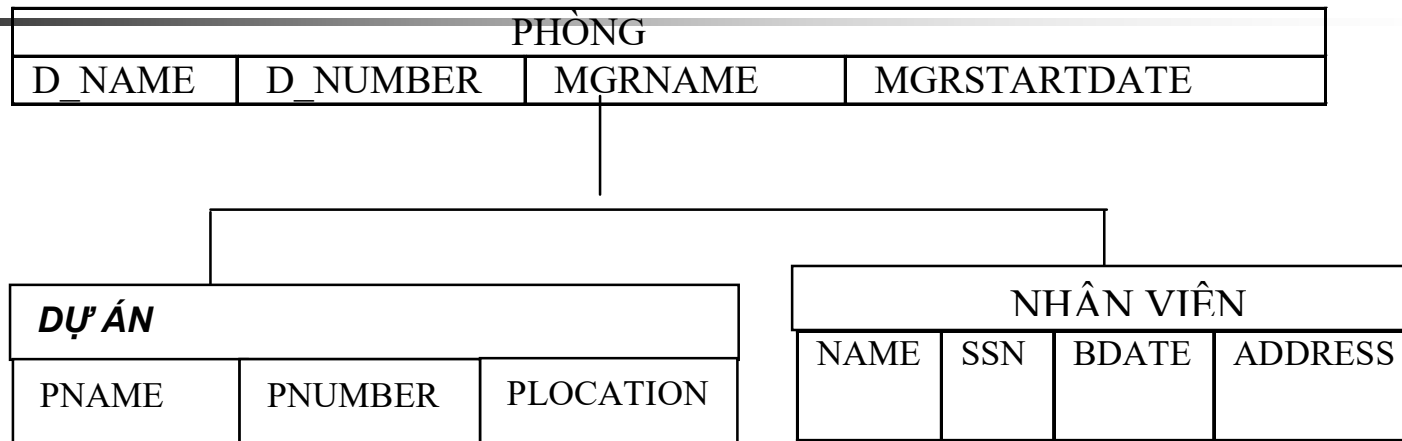
# Mô hình phân cấp-1

Mô hình dữ liệu phân cấp được phát triển để mô hình hóa các loại cơ quan phân cấp nhằm giới thiệu thực



- Concepts of RDBMS and

# Mô hình phân cấp-2



a. **PHÒNG:**

Nghiên cứu

Hành chính

**NHÂN VIÊN:**

Smith

Max John

Grace

Elite

James

Frank

b. **PHÒNG :**

Nghiên cứu

Hành Chính

**DỰ ÁN**

Sản phẩm A

Sản phẩm B

Tin học hóa

Những lợi ích mới

■ Concepts of RDBMS and



# Mô hình mạng - 1

---

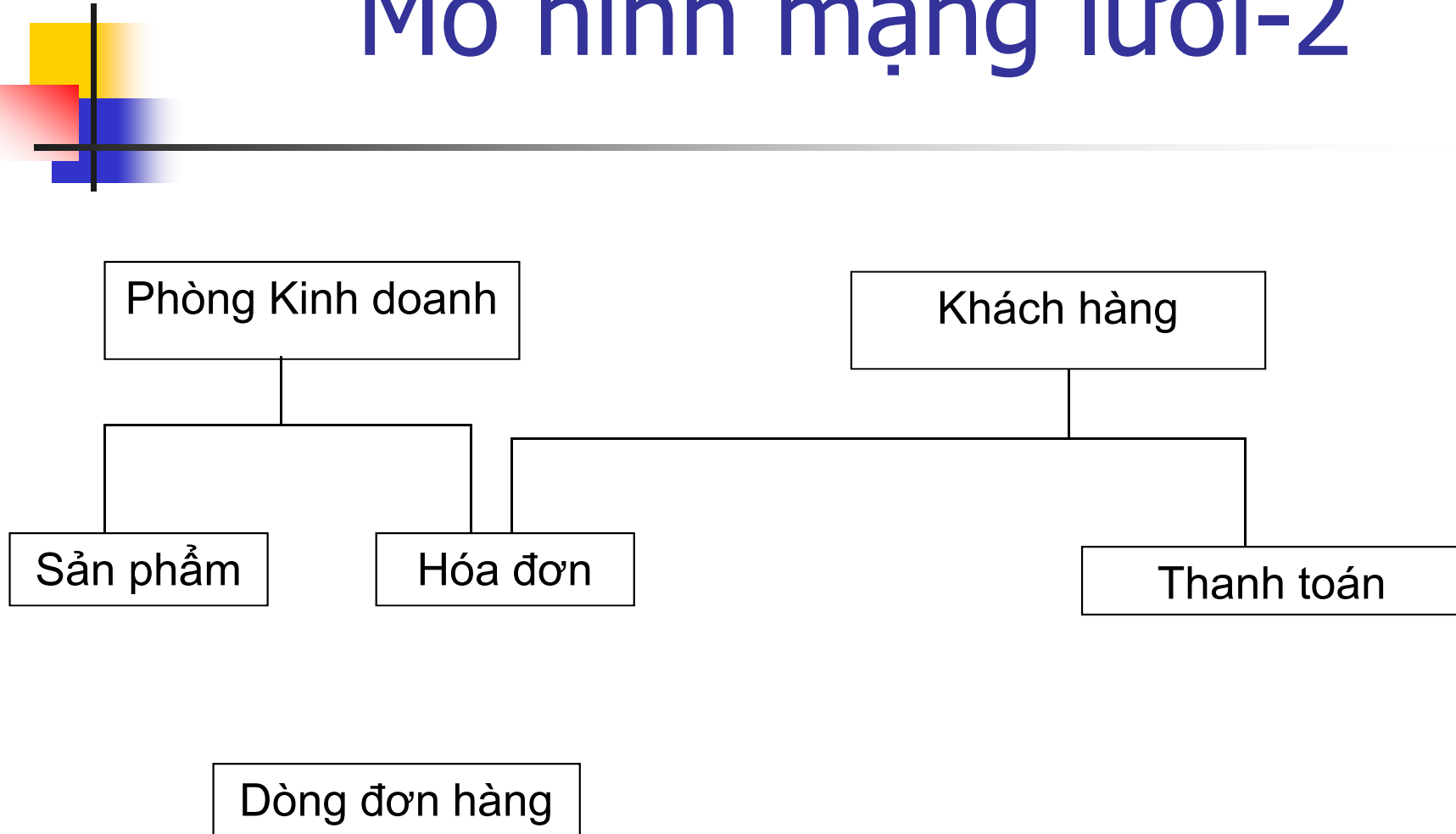
Cùng với mô hình phân cấp, mô hình CSDL mạng (network model) là các mô hình dữ liệu chính để hiện thực rất nhiều hệ quản trị CSDL thương mại.

Các mô hình dữ liệu này được sử dụng từ rất sớm, nhưng không được phổ biến cho lắm vì chúng quá phức tạp.

- Concepts of RDBMS and



# Mô hình mạng lưới-2

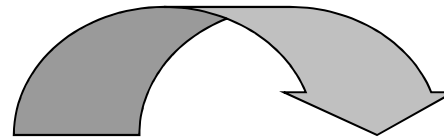


■ Concepts of RDBMS and

# Khái niệm cơ bản về một CSDL-1

**Dữ liệu**

<b>Tay vợt</b>	<b>Số trận thắng trong các loạt đấu</b>
Lleyton Hewitt	20
Carlos Moya	15
Marat Safin	10
Paradorn Srichaphan	8



**Thông tin**

**Tay vợt xuất sắc nhất qua các loạt đấu**

Lleyton Hewitt

■ Concepts of RDBMS and

# Khái niệm cơ bản về một CSDL-2

Các bảng là những đối tượng CSDL trong đó chứa tất cả các dữ liệu trong 1 CSDL.

Dữ liệu trong bảng được sắp thành các dòng và các cột

Các cột

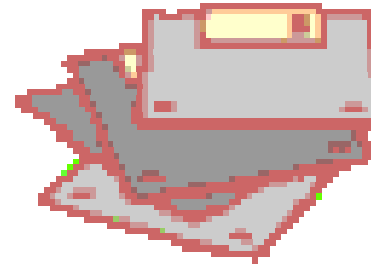
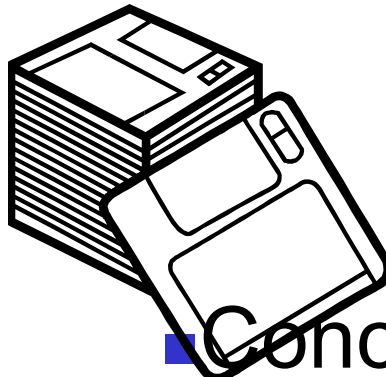
Aircode	Airline_name
9W	Jet Airways
AI	Air India
BA	British Airways
IA	Indian Airlines

Các dòng

■ Concepts of RDBMS and

# Khái niệm cơ bản về một CSDL-3

- Một Hệ thống CSDL là một hệ thống sử dụng máy tính để quản lý các bản ghi mà mục đích tổng quát là nhằm ghi lại và bảo quản thông tin.
- Một Hệ quản trị CSDL (DBMS) là một tập hợp các bản ghi có liên quan với nhau và một bộ các chương trình để truy xuất và thao tác các bản ghi này.

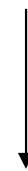


■ Concepts of RDBMS and

# Khái niệm cơ bản về một CSDL-4

- Có thể làm giảm sự lặp lại dữ liệu hoặc 'sự trùng lặp dữ liệu' bằng cách sử dụng 1 Hệ quản trị CSDL.
- Việc đảm bảo tính chính xác của dữ liệu trong CSDL được gọi là duy trì tính toàn vẹn của dữ liệu.

Khóa duy nhất

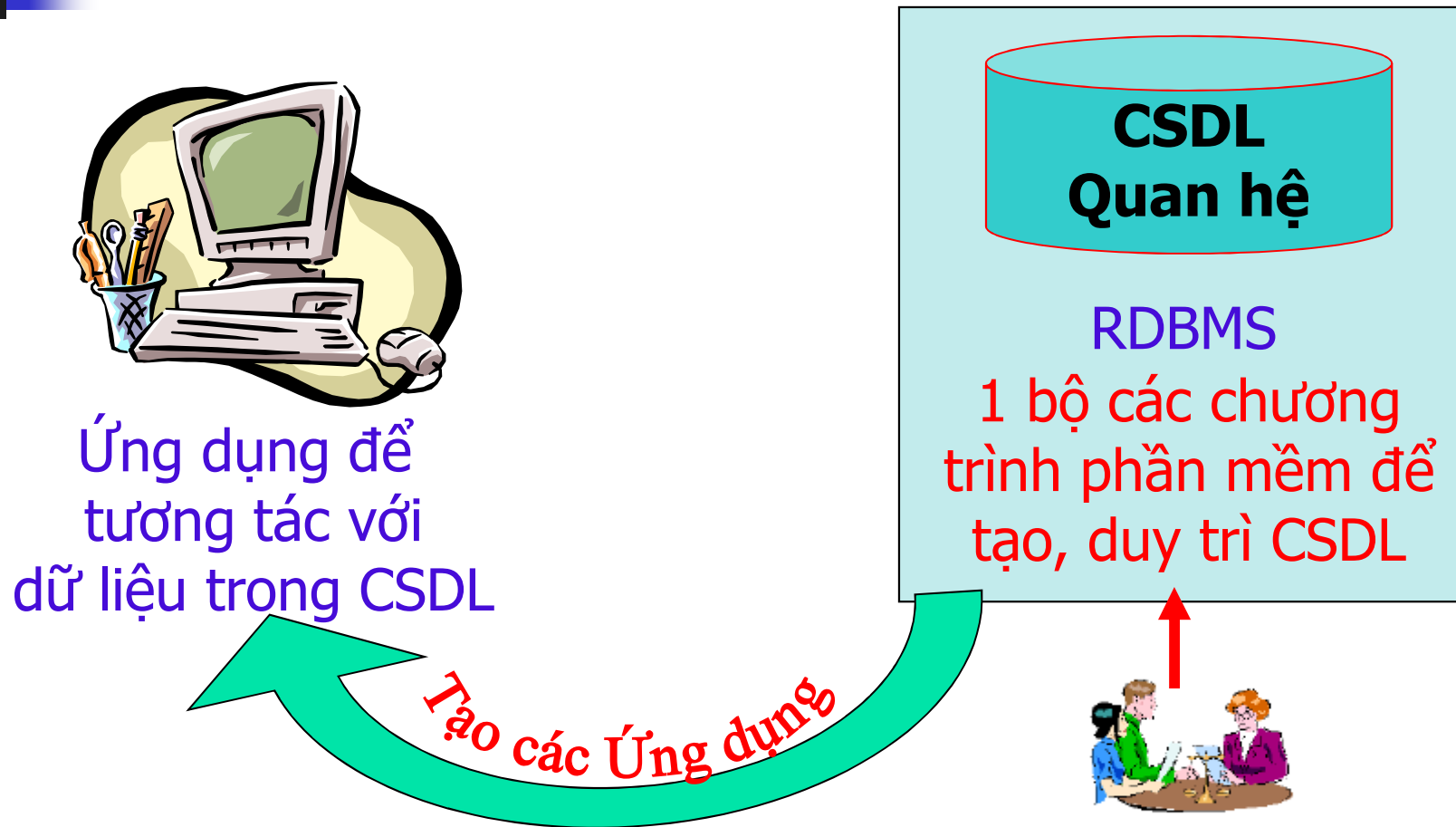


Mã Sinh viên	Tên
1	Sam
2	John
3	Jenny
3	Lisa

Sai

- Concepts of RDBMS and

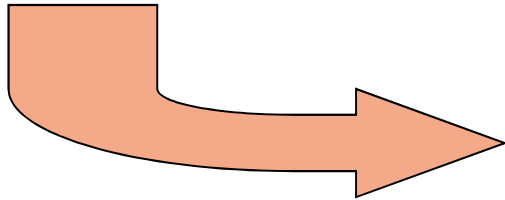
# Khái niệm cơ bản về một CSDL Quan hệ-1



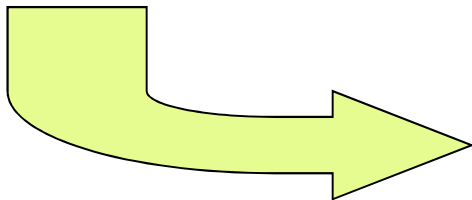
■ Concepts of RDBMS and

# Khái niệm cơ bản về một CSDL Quan hệ-2

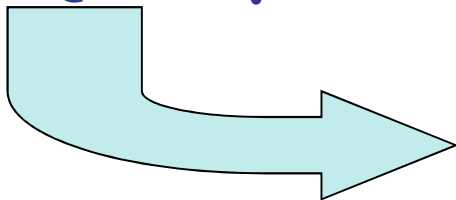
Các Thực thể



Các Thuộc tính



Các Quan hệ

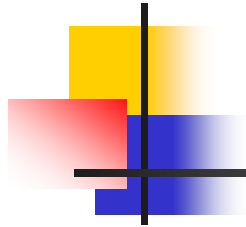


Ảnh xạ thành

Quan hệ

■ Concepts of RDBMS and

# Khái niệm cơ bản về một CSDL Quan hệ-3



1 CSDL là 1 tập hợp các bảng. Mỗi bảng chứa thông tin về 1 thực thể đơn

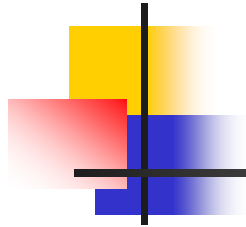
Khóa chính là 1 hoặc tổ hợp 1 số cột mà qua đó có khả năng xác định duy nhất 1 bản ghi trong một bảng.

Khóa chính

Mã Sinh viên	Tên Sinh viên
1	Sam
2	John
3	Jenny
4	Lisa
5	Penny
6	Peter
7	Joe

■ Concepts of RDBMS and





# Toàn vẹn thực thể

- Luật toàn vẹn thực thể phát biểu rằng không thành phần nào trong khóa chính được phép chứa giá trị NULL.
- Một giá trị NULL là 1 giá trị bị bỏ qua.

Roll Number	Student Name
1	Sam
2	John
3	Jenny
4	Lisa
5	Penny
	Peter
7	Joe

Invalid

- Concepts of RDBMS and



# Khóa ngoại

Hai bảng được kết nối với nhau nhờ dùng một khóa ngoại.

**Khóa chính**

Mã Sinh viên	Tên Sinh viên
1	Sam
2	John
3	Jenny
4	Lisa
5	Penny
6	Peter
7	Joe

**Khóa ngoại**

Mã Sinh viên	Tên Sinh viên	Điểm đạt được
2	John	65
4	Lisa	90
6	Peter	87
7	Joe	89

■ Concepts of RDBMS and



# Toàn vẹn tham chiếu

Luật **Toàn vẹn tham chiếu** phát biểu rằng trong một CSDL không tồn tại bất kỳ một thuộc tính Khóa ngoại nào có giá trị không khớp với các giá trị tương ứng trong quan hệ mà thuộc tính đó làm khóa chính.

Hệ quản trị CSDL chịu trách nhiệm đảm bảo các Giá trị thuộc tính trong khóa ngoại là hợp lệ và có một giá trị tương ứng trong quan hệ mà thuộc tính đó là khóa chính. ■ **Concepts of RDBMS and**



# Quan hệ và bậc

Tùy vào mô hình quan hệ mà mỗi quan hệ hoặc bảng được tạo nên từ nhiều *bộ*. Chúng còn được gọi là các bản ghi.

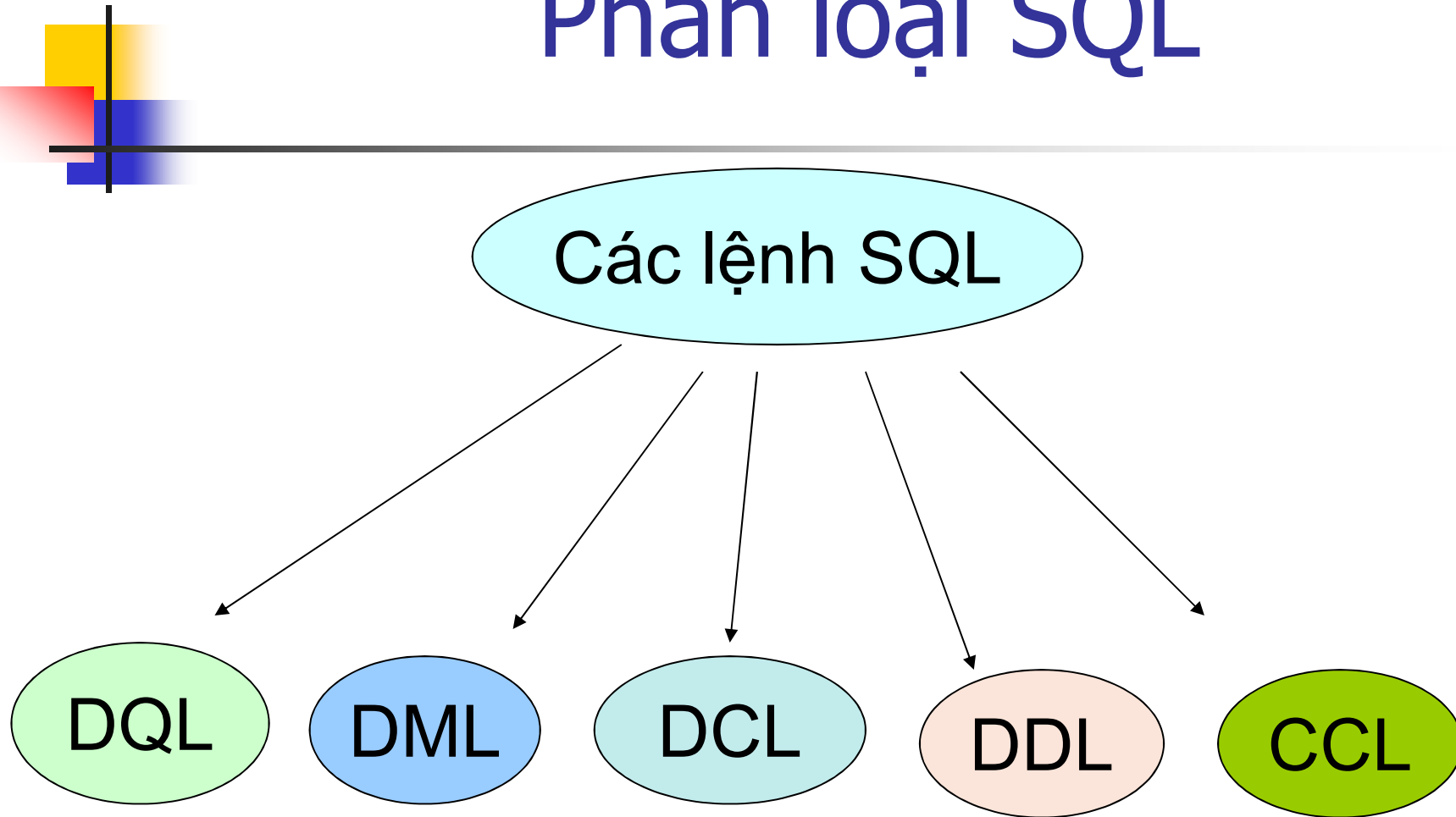
Ví dụ:

CHN	China	Renminbi (Yuan)
FRN	France	Francs

Số lượng các thuộc tính trong 1 bảng được gọi là bậc của bảng đó.

■ Concepts of RDBMS and

# Phân loại SQL



■ Concepts of RDBMS and