



BÀI 9

CƠ SỞ DỮ LIỆU VÀ HỆ THỐNG THÔNG TIN

Viện Công nghệ thông tin và Truyền thông
2017

Nội dung

1. Dữ liệu
2. Tổ chức dữ liệu
3. Cơ sở dữ liệu
4. Các mô hình cơ sở dữ liệu
5. Các kiểu cơ sở dữ liệu
6. Sử dụng CSDL

1. Dữ liệu

- Dữ liệu (data) bao gồm các yếu tố hay các quan sát về con người, địa điểm, vạn vật và các sự kiện.
- Các kiểu dữ liệu:
 - Số
 - Ký tự
 - Âm thanh
 - Âm nhạc
 - Hình ảnh
 - Video

Cách nhìn về dữ liệu

- Cách nhìn về vật lý: Tập trung vào dạng thức thực tế và nơi đặt dữ liệu.
- Cách nhìn về logic: Tập trung vào ý nghĩa, nội dung và ngữ cảnh của dữ liệu.

2. Tổ chức dữ liệu

- Ký tự (Character): là phần tử dữ liệu cơ bản nhất.
- Trường (Field) : là nhóm các ký tự có quan hệ với nhau để mô tả một thuộc tính của thực thể (entity) (một người, vị trí, sự vật hay một đối tượng).
- Bản ghi (Record): tập hợp các trường có quan hệ với nhau để mô tả các thuộc tính của thực thể.
- Bảng (Table): tập hợp các bản ghi có quan hệ với nhau, được lưu trữ dưới dạng file.
- Cơ sở dữ liệu (Database): Tập hợp các bảng có quan hệ logic với nhau

Trường khóa (Key Field)

- Còn gọi là Khóa chính (Primary Key)
- Định danh duy nhất cho bản ghi
- Ví dụ
 - Mã số công dân
 - Mã số sinh viên
 - Mã số cán bộ
 - Mã số phụ tùng (Part Number)
 - Số seri (Serial number)

Ví dụ

	Họ đệm	Tên	Mã số SV	Ngày sinh	Nơi sinh
Bản ghi 1	Nguyễn Hoàng	Long	20172168	02/06/1999	Hà Nội
Bản ghi 2	Trần Hoài	Nam	20172912	21/10/1999	Hải Dương
Bản ghi 3	Lê Thu	Trang	20173265	30/05/1999	Hải Phòng

- Bảng dữ liệu sinh viên KT22 K62
- Cơ sở dữ liệu sinh viên K62-ĐHBKHN

Xử lý theo lô và xử lý thời gian thực

- Xử lý theo lô (Batch processing): Dữ liệu được thu thập và tập hợp sau một khoảng thời gian xác định, khi đó mới xử lý tất cả cùng một lần.
- Xử lý thời gian thực (Real-time processing): dữ liệu được xử lý cùng lúc với các giao tác dữ liệu.

3. Cơ sở dữ liệu (CSDL)

- CSDL là tập các bảng dữ liệu có quan hệ logic được tích hợp với nhau
- CSDL cần có các đặc điểm sau:
 - Chia sẻ thông tin: thông tin từ bộ phận này có thể sẵn sàng được chia sẻ với bộ phận khác.
 - An toàn: người dùng chỉ có thể truy cập được thông tin cần thiết cho họ.
 - Ít dư thừa dữ liệu (less data redundancy): Nếu không có CSDL chung, khi đó các bộ phận riêng lẻ cần phải tạo ra và duy trì dữ liệu của mình và dẫn đến dư thừa dữ liệu.
 - Tính toàn vẹn dữ liệu (data integrity): khi có nhiều nguồn dữ liệu, thông tin trong các nguồn có thể khác nhau.

Hệ quản trị cơ sở dữ liệu

- Hệ quản trị CSDL (DBMS - Database Management System): là phần mềm để tạo, biến đổi và truy cập cơ sở dữ liệu.
- Ví dụ:
 - Microsoft Access
 - SQL Server
 - My SQL
 - Oracle

Các thành phần của DBMS

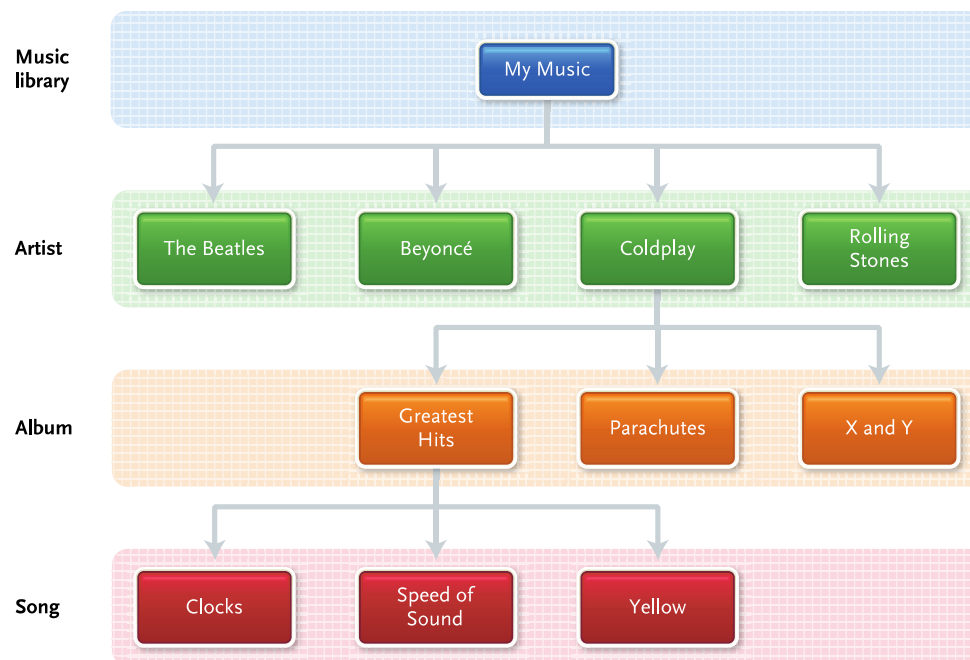
- DBMS engine: cung cấp cầu nối giữa dữ liệu logic và dữ liệu vật lý.
- Hệ thống con định nghĩa dữ liệu: định nghĩa cấu trúc logic của CSDL bằng cách sử dụng từ điển dữ liệu hay lược đồ dữ liệu.
- Hệ thống con thao tác dữ liệu: cung cấp các công cụ để duy trì và phân tích dữ liệu (thêm, xóa, sửa đổi, truy vấn).
- Hệ thống con tạo ứng dụng: cung cấp các công cụ để tạo các dạng vào dữ liệu và các ngôn ngữ lập trình chuyên biệt.
- Hệ thống con quản trị dữ liệu: để quản lý CSDL, nó bao gồm: duy trì an toàn, cung cấp sự hỗ trợ khôi phục thảm họa và giám sát hiệu năng tổng thể của các thao tác CSDL.

4. Các mô hình cơ sở dữ liệu

- Cơ sở dữ liệu phân cấp (Hierarchical Database)
- Cơ sở dữ liệu mạng (Network Database)
- Cơ sở dữ liệu quan hệ (Relational Database)
- Cơ sở dữ liệu đa chiều (Multidimensional Database)
- Cơ sở dữ liệu hướng đối tượng (Object-Oriented Database)

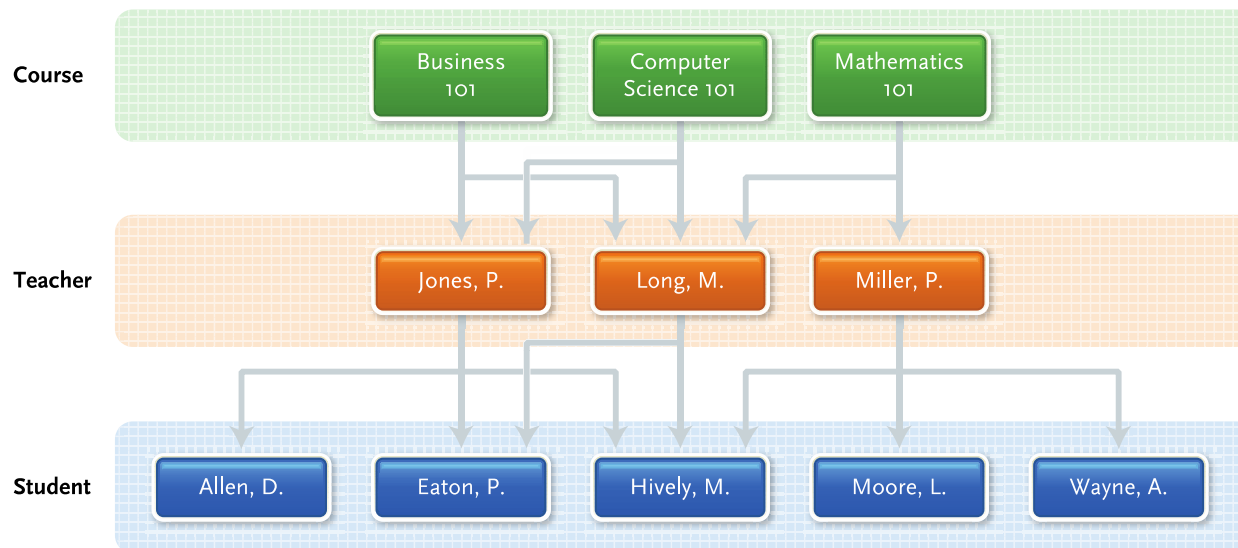
Cơ sở dữ liệu phân cấp

- Các trường hoặc các bản ghi được cấu trúc theo node
- Mỗi điểm vào có thể có một node cha với một vài node con theo quan hệ một nhiều
- Mỗi node con chỉ có một node cha
- Không có quan hệ giữa các node con



Cơ sở dữ liệu mạng

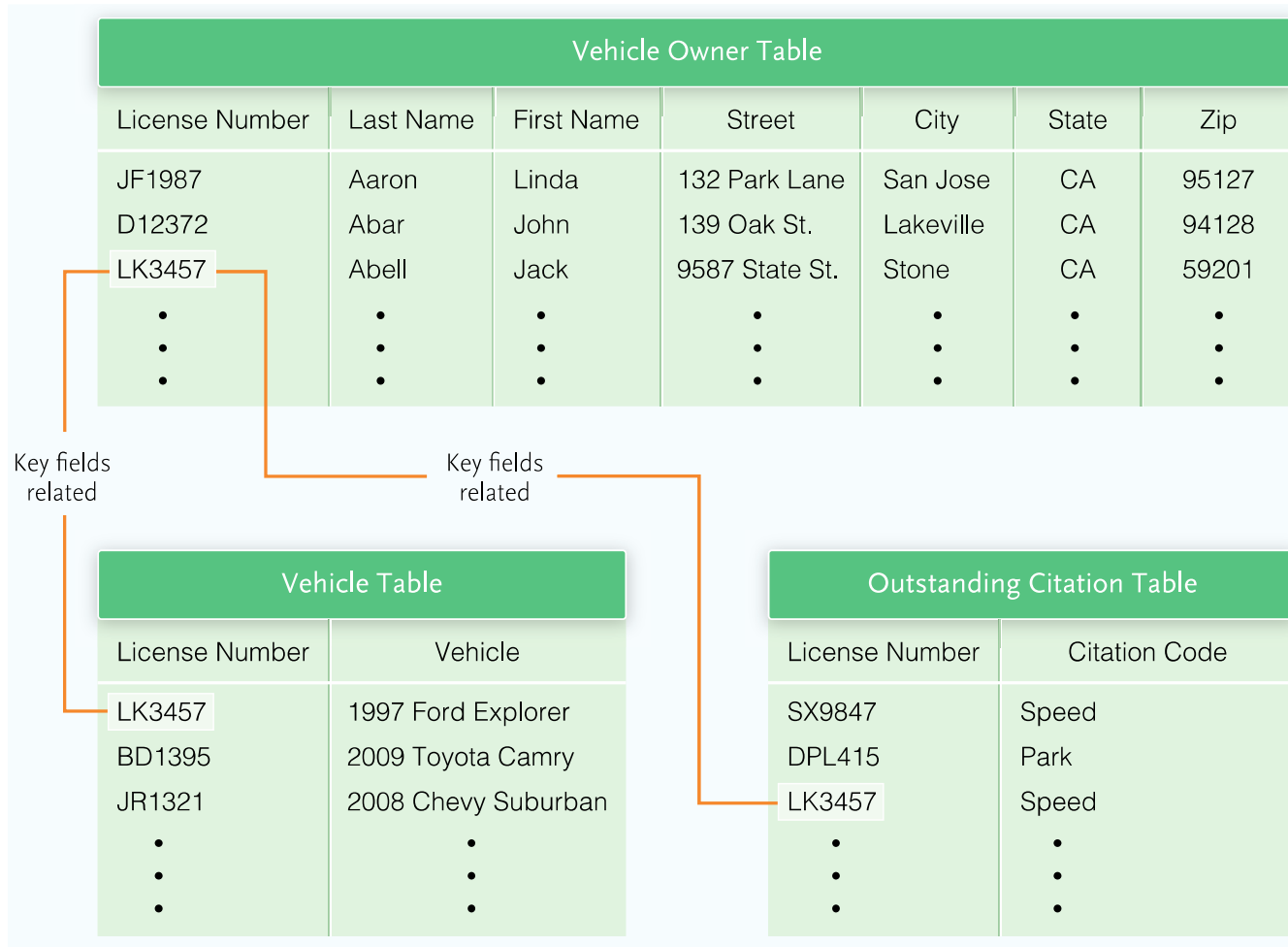
- Tổ chức phân cấp cho các node
- Mỗi node con có thể có nhiều hơn một node cha
- Có thể có quan hệ nhiều-nhiều
- Để truy cập vào một node có thể có các đường khác nhau.



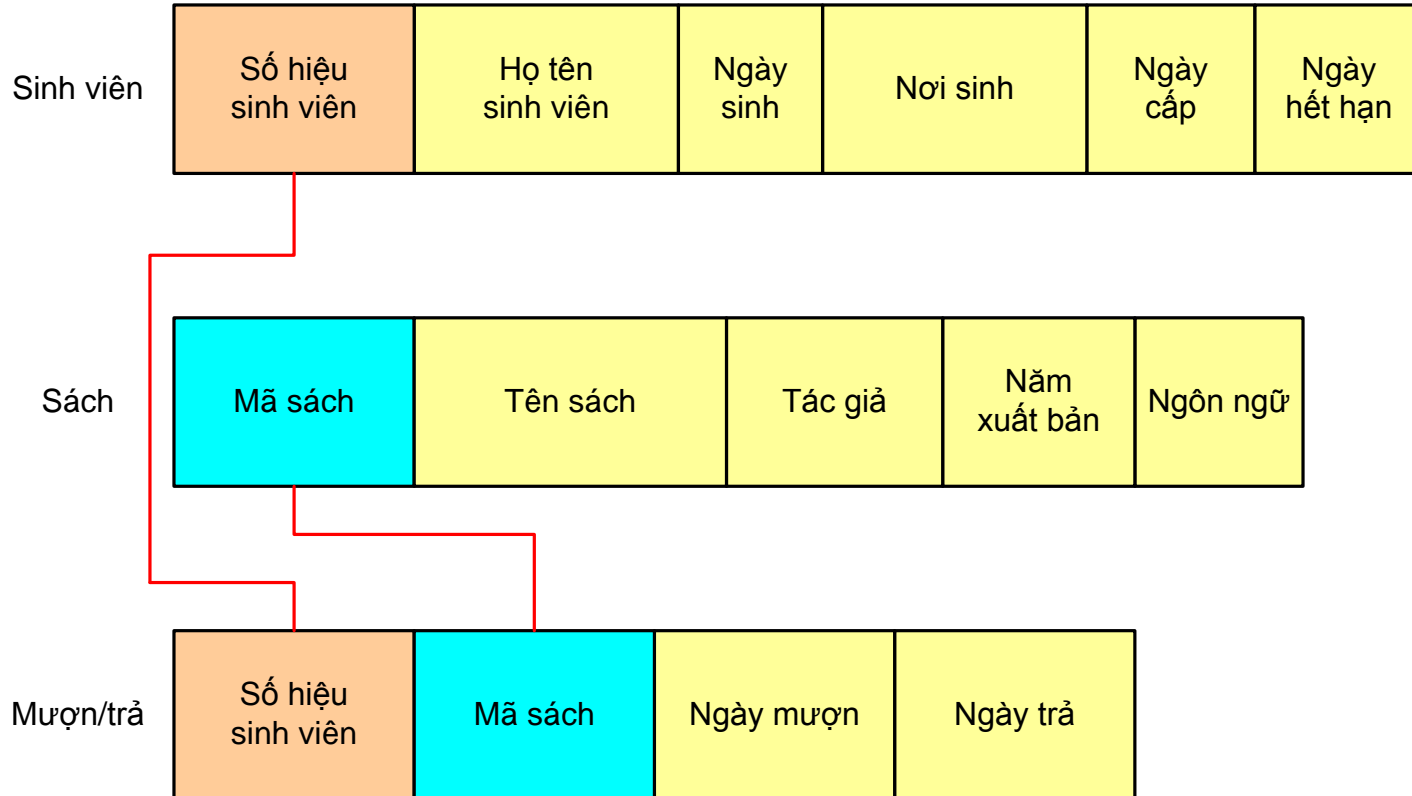
Cơ sở dữ liệu quan hệ

- Các phần tử dữ liệu được lưu trữ trong các bảng khác nhau
- Các bảng bao gồm các hàng và các cột
- Các bảng cần có mục dữ liệu chung (trường khóa) cho phép thông tin được lưu trong bảng này có thể được liên kết với thông tin trong bảng khác
- Tổ chức đơn giản và dễ hiểu
- Các thực thể dễ dàng thêm, xóa và thay đổi
- Được sử dụng phổ biến nhất

Ví dụ cơ sở dữ liệu quan hệ



Ví dụ cơ sở dữ liệu quan hệ



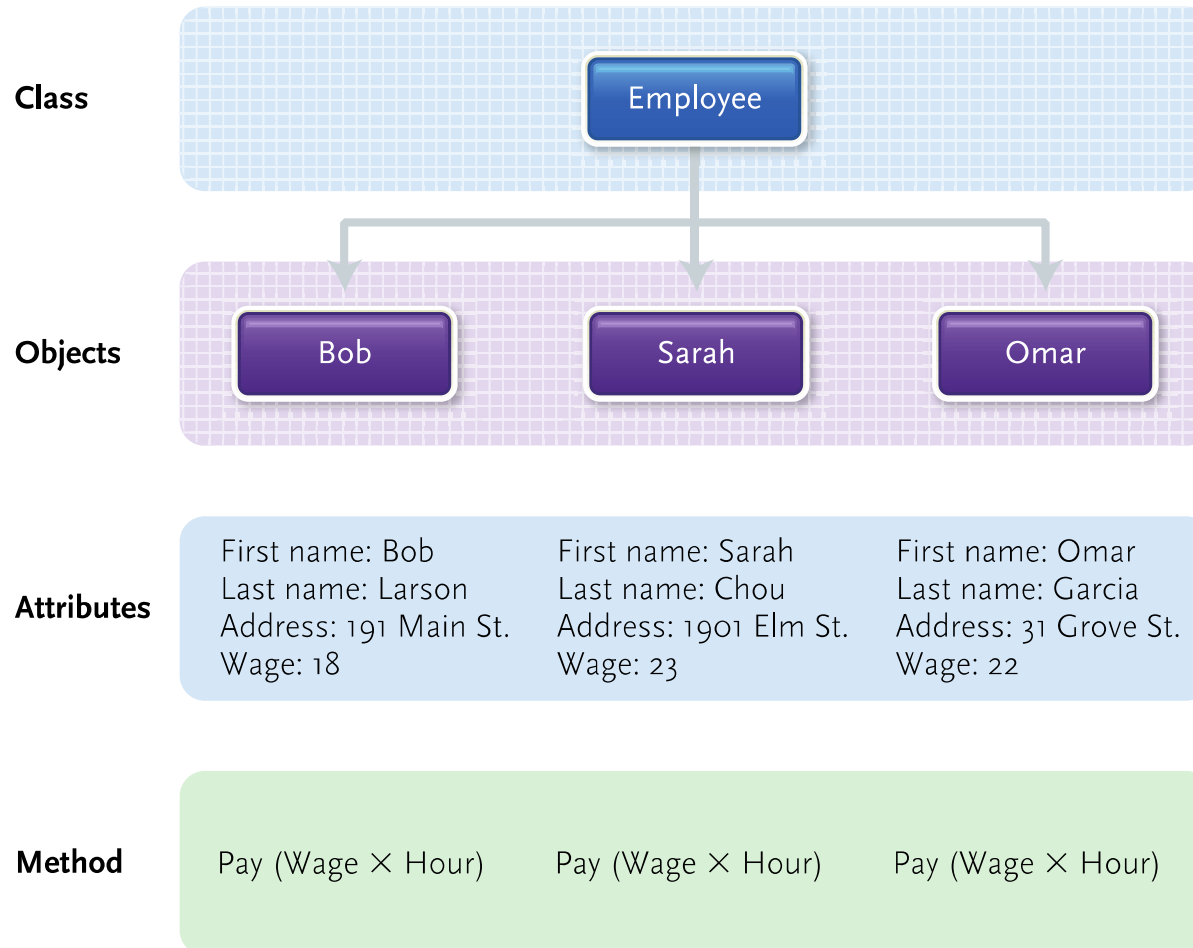
Cơ sở dữ liệu đa chiều

- Mở rộng của cơ sở dữ liệu quan hệ từ hai chiều thành nhiều chiều hơn
- Sử dụng cho các quan hệ dữ liệu phức tạp
- Tăng tốc độ xử lý

Cơ sở dữ liệu hướng đối tượng

- Cơ sở dữ liệu hướng đối tượng lưu giữ dữ liệu và cả lệnh để thao tác với dữ liệu đó.
- Tổ chức của CSDL hướng đối tượng bao gồm:
 - Lớp (Classes): Định nghĩa khuôn mẫu chung cho các đối tượng
 - Đối tượng (Objects): là các trường hợp cụ thể của lớp chứa cả dữ liệu và lệnh để thao tác dữ liệu
 - Thuộc tính (Attributes): là các trường dữ liệu của đối tượng sở hữu
 - Phương thức (Methods): là các lệnh để tìm kiếm hoặc thao tác các giá trị của thuộc tính.

Cơ sở dữ liệu hướng đối tượng



5. Các kiểu cơ sở dữ liệu

- CSDL cá nhân: dùng cho một người
- CSDL công ty: được dùng chung chia sẻ bởi nhiều người trong công ty
- CSDL phân tán: phân bố ở các vị trí địa lý khác nhau, được truy nhập thông qua các đường truyền thông
- CSDL thương mại: sẵn dùng với các thuê bao (ngân hàng dữ liệu)

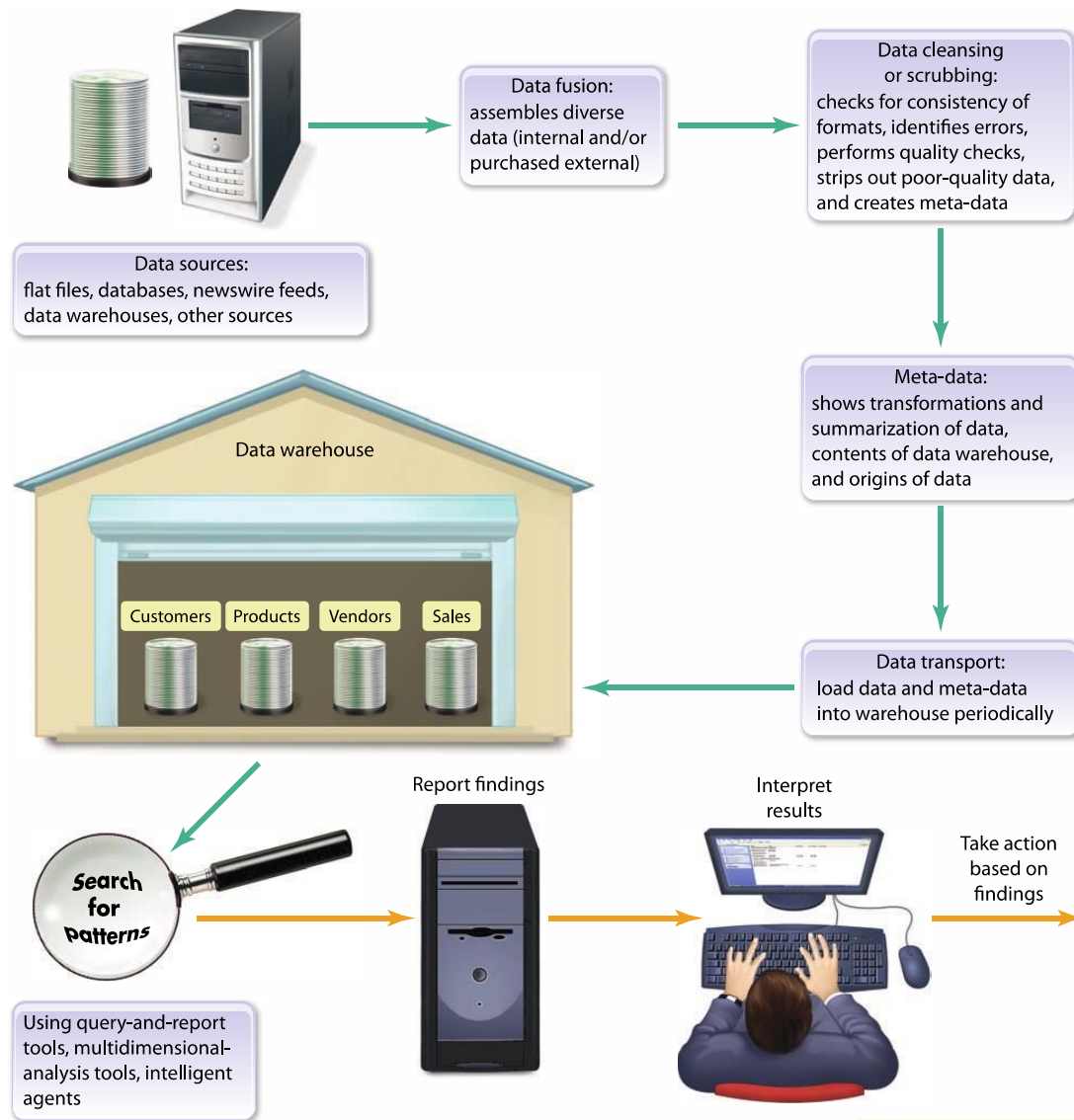
6. Sử dụng CSDL

- CSDL đưa ra cơ hội lớn để tăng năng suất lao động
- Việc duy trì CSDL cần cố gắng để không bị xáo trộn hoặc bị lạm dụng
- Chiến lược sử dụng:
 - Thường xuyên cập nhật từ các nguồn thông tin bên trong hoặc bên ngoài
 - Tổ chức thành kho dữ liệu (data warehouse)
 - Khai phá dữ liệu (data mining): là kỹ thuật dùng để tìm kiếm và khai phá dữ liệu để lấy ra các thông tin và các mẫu liên quan.
- An toàn cơ sở dữ liệu:
 - Chống sử dụng bất hợp pháp → sử dụng firewalls
 - Chống truy cập không đúng thẩm quyền → cần phân quyền truy cập cho người dùng

Khai phá dữ liệu và Kho dữ liệu

- Khai phá dữ liệu (Data Mining) là quá trình sàng lọc và phân tích lượng lớn dữ liệu để trích xuất ra các mẫu và các nghĩa ẩn và để phát hiện tri thức mới.
- Kho dữ liệu (Data Warehouse) là cơ sở dữ liệu đặc biệt của dữ liệu và siêu dữ liệu (metadata) sạch
- Dữ liệu được đưa vào một kho dữ liệu thông qua các bước sau:
 - Xác định và kết nối với các nguồn dữ liệu
 - Thực hiện tổng hợp dữ liệu và làm sạch dữ liệu
 - Nhận được cả dữ liệu và siêu dữ liệu (dữ liệu về dữ liệu)
 - Chuyển dữ liệu và siêu dữ liệu cho kho dữ liệu

Khai phá dữ liệu và Kho dữ liệu





Hết bài 9