

CHƯƠNG 1: TỔNG QUAN VỀ QUẢN LÝ THÔNG TIN

Khoa Khoa học và kỹ thuật thông tin
Bộ môn Thiết bị di động và Công nghệ Web

NỘI DUNG

1. Quản lý thông tin là gì.
2. Quy trình quản lý thông tin.
3. Các khó khăn khi quản lý thông tin.
4. Dữ liệu và thông tin.
5. Cơ sở dữ liệu.

KHÁI NIỆM VỀ QUẢN LÝ THÔNG TIN

QUẢN LÝ THÔNG TIN LÀ GÌ

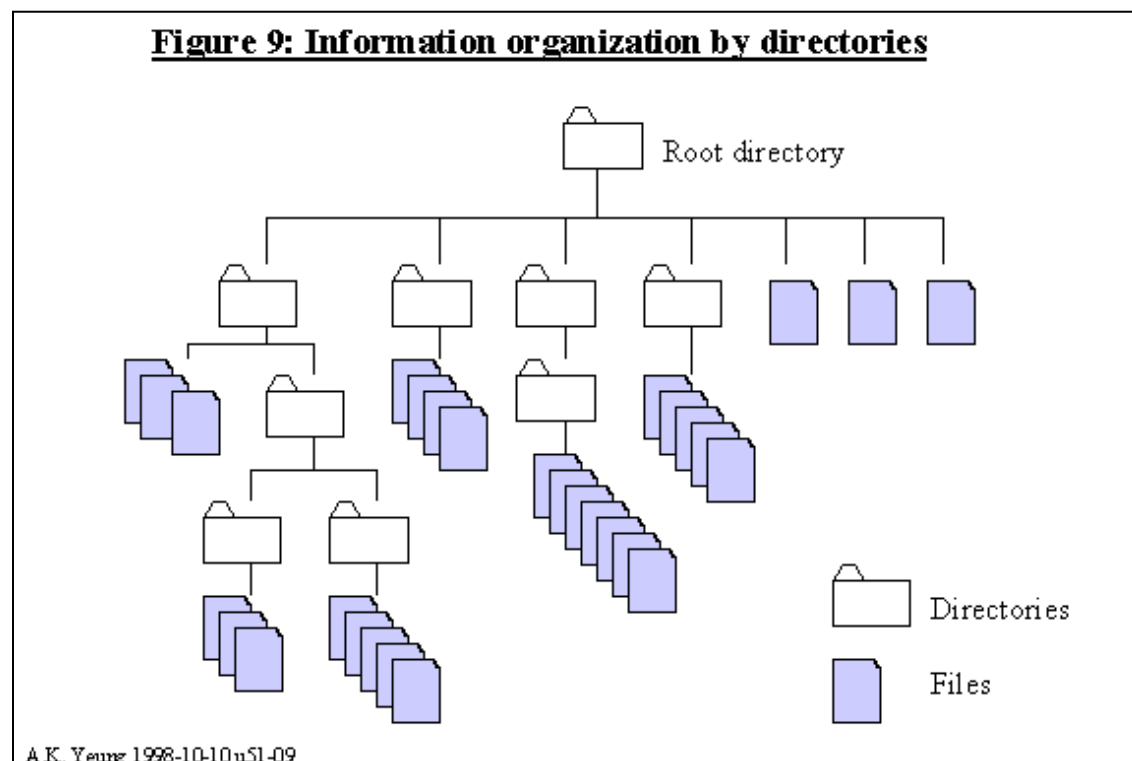
- Quản lý thông tin gồm 2 khái niệm lớn như sau:
 - + Tổ chức thông tin.
 - + Xử lý thông tin.
- Mục tiêu:
 - + Giảm dư thừa dữ liệu.
 - + Tăng khả năng tổ chức và xử lý thông tin.

Tổ chức thông tin

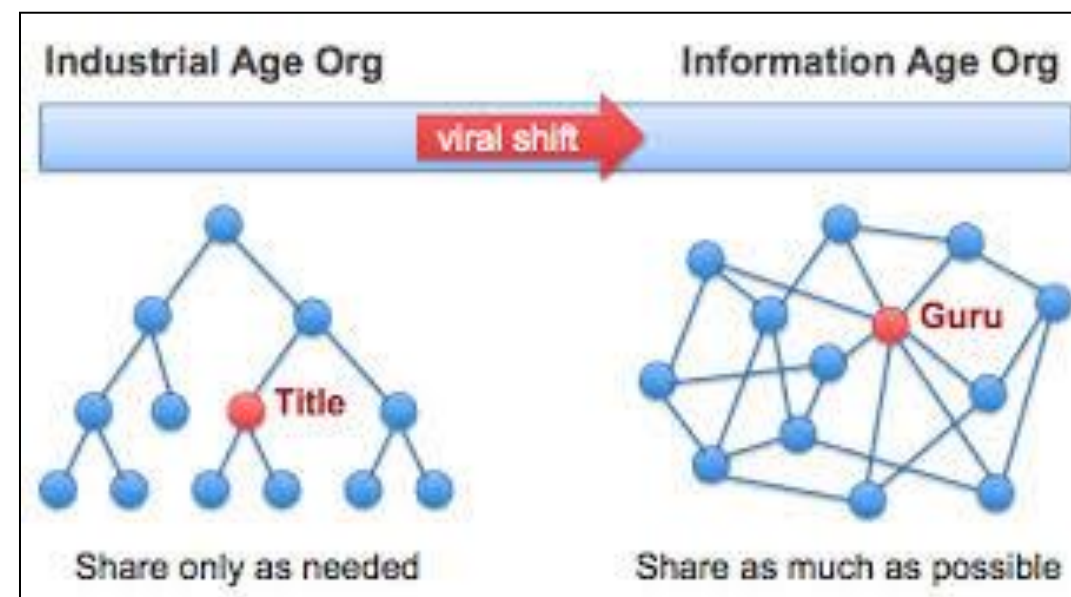
- Tổ chức thông tin gồm các hoạt động sau:
 - + Mô tả tài liệu, lập chỉ mục và phân loại.
 - + Được thực hiện trong các thư viện, cơ sở dữ liệu, ...
 - + Những hoạt động này được thực hiện bởi cán bộ thư viện, các chuyên gia, cũng như bởi các thuật toán máy tính.

Ví dụ về tổ chức thông tin

Dạng cây thư mục

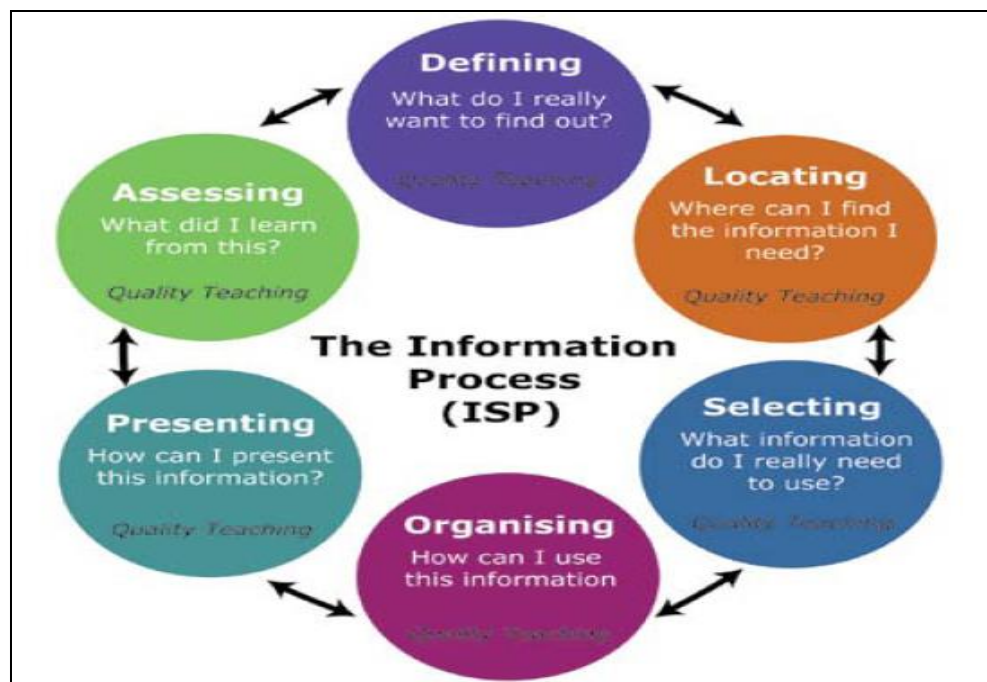


Dạng đồ thị



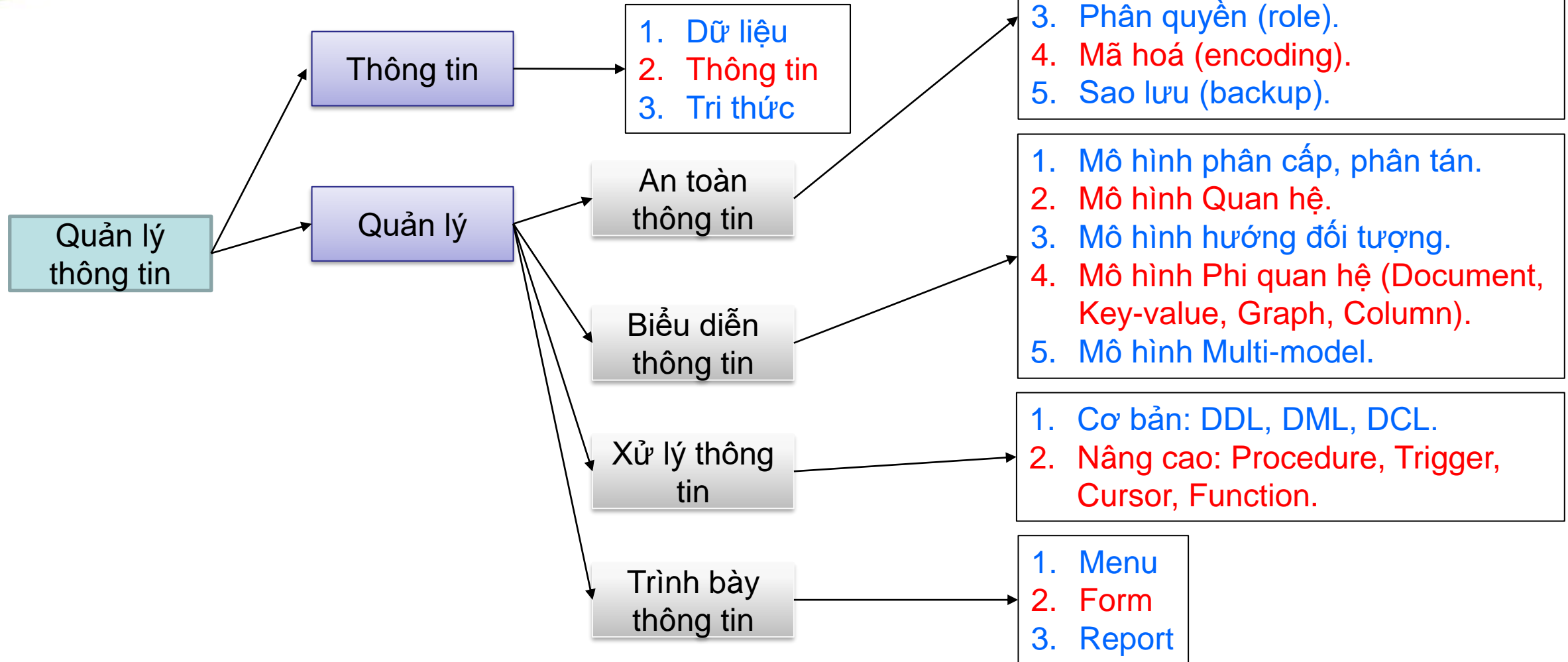
Xử lý thông tin

- Xử lý thông tin là các ngành khoa học liên quan đến việc thu thập, thao tác, lưu trữ, tìm kiếm, và phân loại thông tin đã ghi nhận.



QUY TRÌNH QUẢN LÝ THÔNG TIN

QUY TRÌNH QUẢN LÝ THÔNG TIN



KHÓ KHĂN KHI QUẢN LÝ THÔNG TIN

SỰ PHỨC TẠP CỦA QUẢN LÝ THÔNG TIN (1)

- Thiếu các nguồn lực để tổ chức, quản lý và cải thiện hệ thống thông tin.
- Thiếu định nghĩa với nhiều thông tin.
- Ít hoặc không có sự đánh giá và hỗ trợ quản lý thông tin của các quản lý cấp cao.
- Sự kém chất lượng của thông tin, vd: sự thiếu nhất quán, sự lặp lại.
- Hệ thống đã lỗi thời cần thay thế hoặc nâng cấp.
- Số lượng lớn các hệ thống quản lý thông tin phân tán ít liên kết nhau.

SỰ PHỨC TẠP CỦA QUẢN LÝ THÔNG TIN (2)

- Thiếu đồng bộ giữa các hệ thống thông tin.
- Cạnh tranh giữa các hệ thống quản lý thông tin.
- Thiếu con đường chiến lược về quản lý hệ thống thông tin.
- Thiếu sự rõ ràng của các chiến lược và mục tiêu của tổ chức.
- Có khác nhau về các vấn đề đa dạng và nhu cầu chưa được giải quyết.
- Nhân viên hạn chế cập nhật hệ thống thông tin hiện có.
- Cản trở trong việc chuyển đổi tập quán làm việc cũ và các quá trình của nhân viên.

Dữ liệu và thông tin

Phân biệt giữa dữ liệu và thông tin

Dữ liệu

- Sự kiện.
- Không có ngữ cảnh.
- Con số, hình ảnh, âm thanh, văn bản.

Thông tin

- Dữ liệu có ngữ cảnh.
- Dữ liệu đã xử lý.
- Thêm giá trị vào dữ liệu nhờ
 - + Phân tích.
 - + Tổng hợp.
 - + Tổ chức lại.

Phân biệt giữa dữ liệu và thông tin

Dữ liệu

— VD: 51014

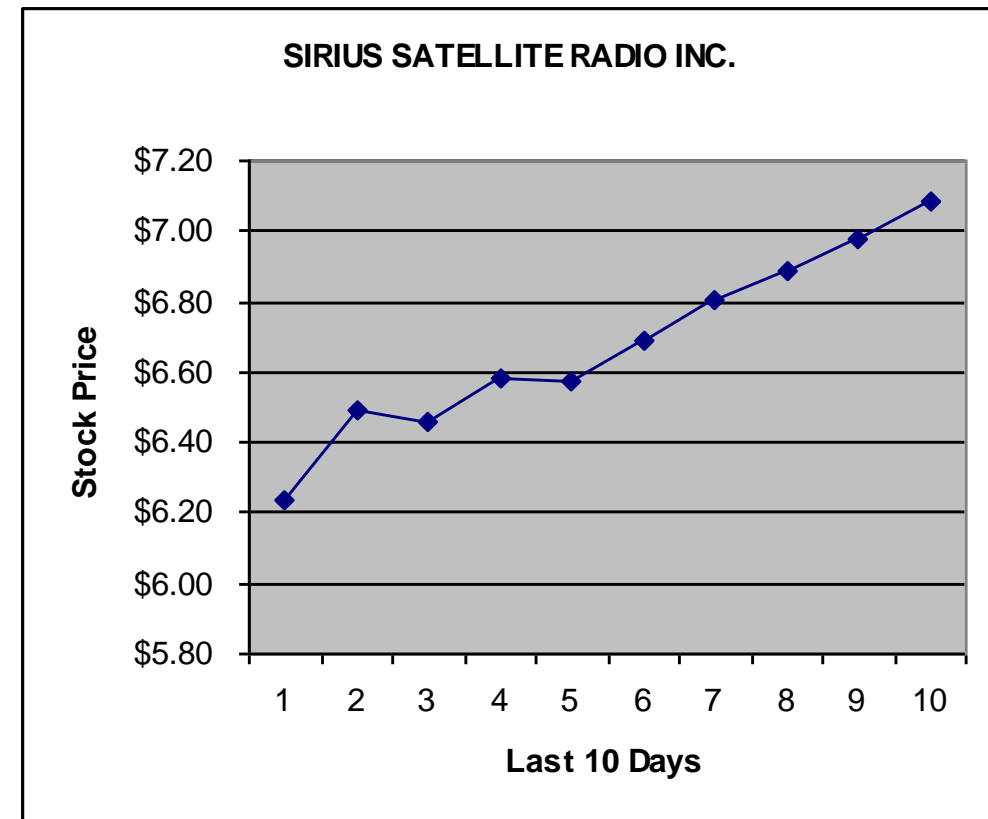
Thông tin

- 5/10/14: Sinh nhật của SV N.T.A.
- \$51,014: Thu nhập bình quân của một người lao động tại quốc gia AB.

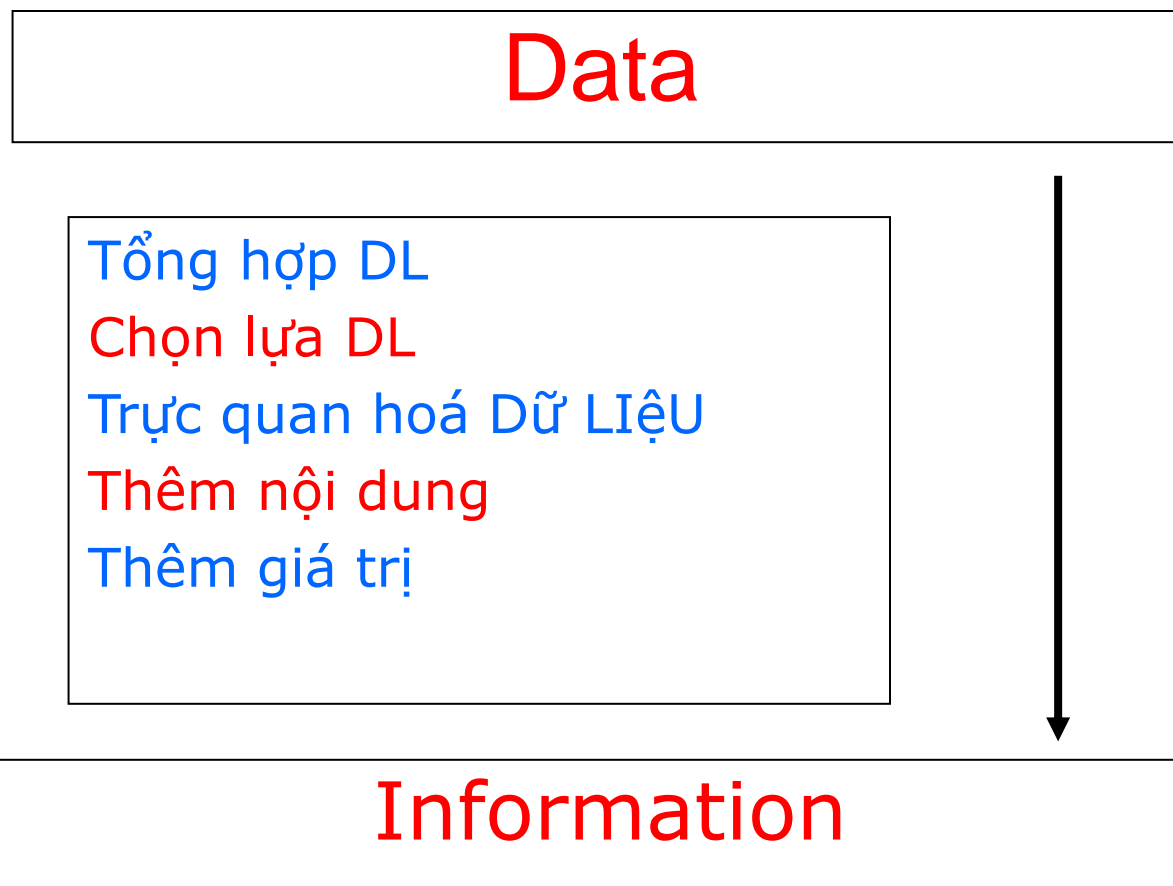
Ví dụ 2

— Data

+ 6.34
 + 6.45
 + 6.39
 + 6.62
 + 6.57
 + 6.64
 + 6.71
 + 6.82
 + 7.12
 + 7.06

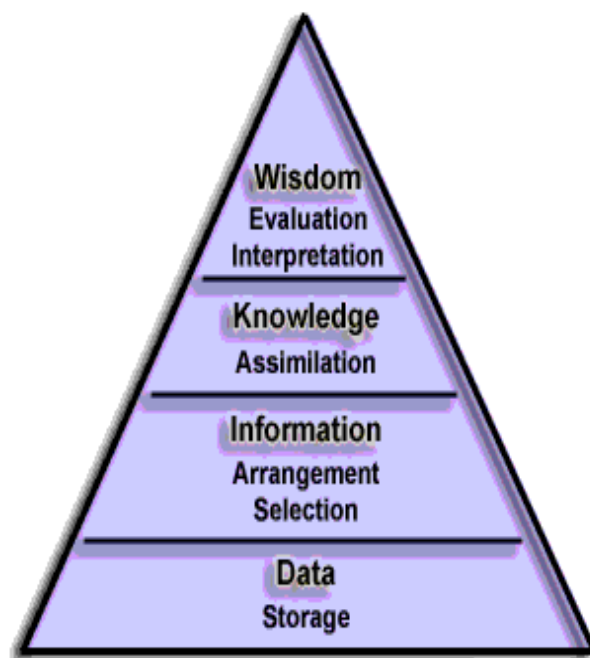


Dữ liệu và thông tin



Thông tin và tri thức

Information



Data Hierarchy Model
developed at 3M by G. Stephen Tuthill, 1990

- ☐ Làm thế nào là thông tin gắn với kết quả?
- ☐ Có bất kỳ mô hình trong thông tin?
- ☐ Thông tin gì có liên quan đến vấn đề này?
- ☐ Làm thế nào để thông tin này ảnh hưởng đến hệ thống?
- ☐ Cách tốt nhất để sử dụng các thông tin là gì?
- ☐ Làm thế nào chúng ta có thể tăng giá trị cho các thông tin?

Knowledge

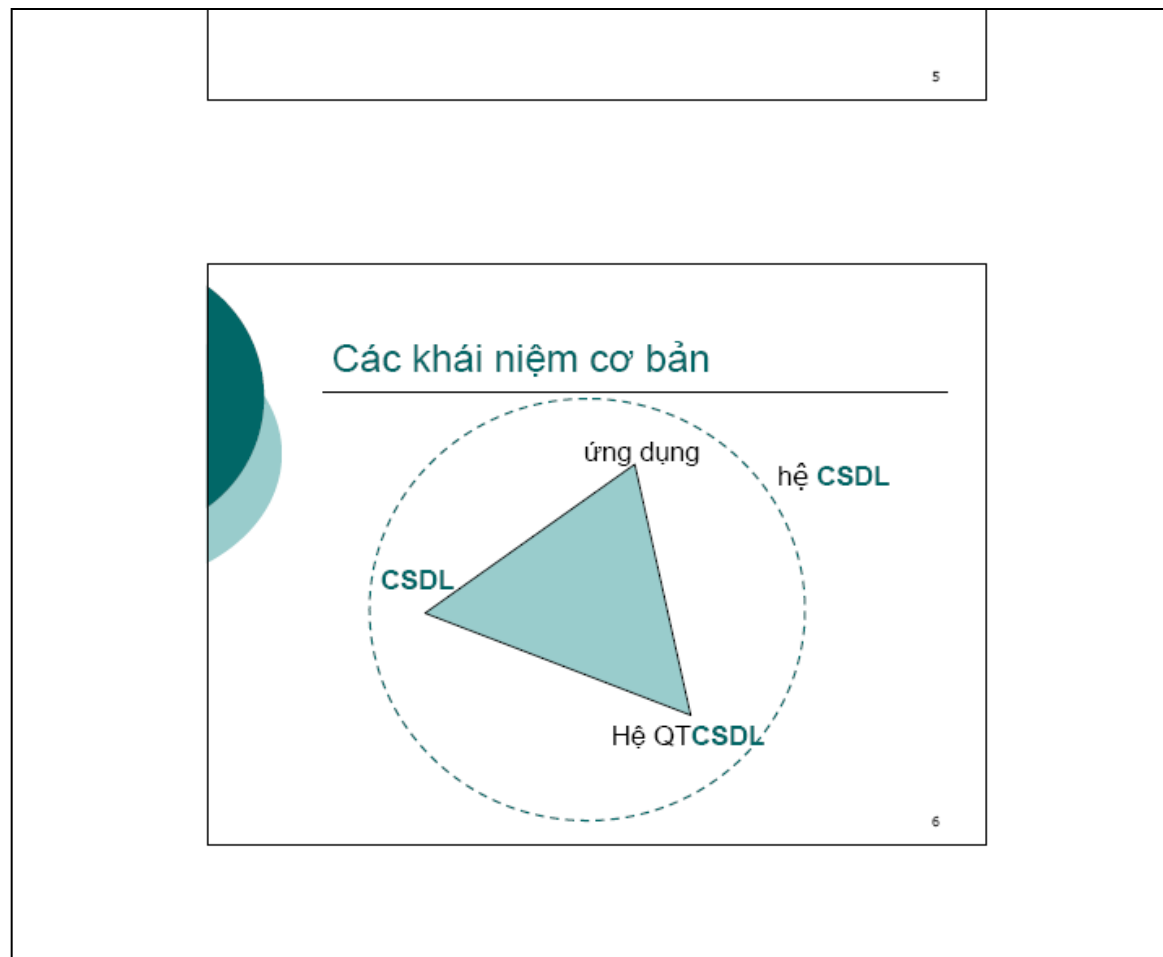
Dữ liệu và thông tin

- Mục tiêu của hệ thống thông tin:
 - + Chuyển dữ liệu thành thông tin
 - + Lõi của hệ thống thông tin là Cơ sở dữ liệu.
- Cơ sở dữ liệu (CSDL) dùng để quản trị thông tin và dữ liệu trên máy tính.

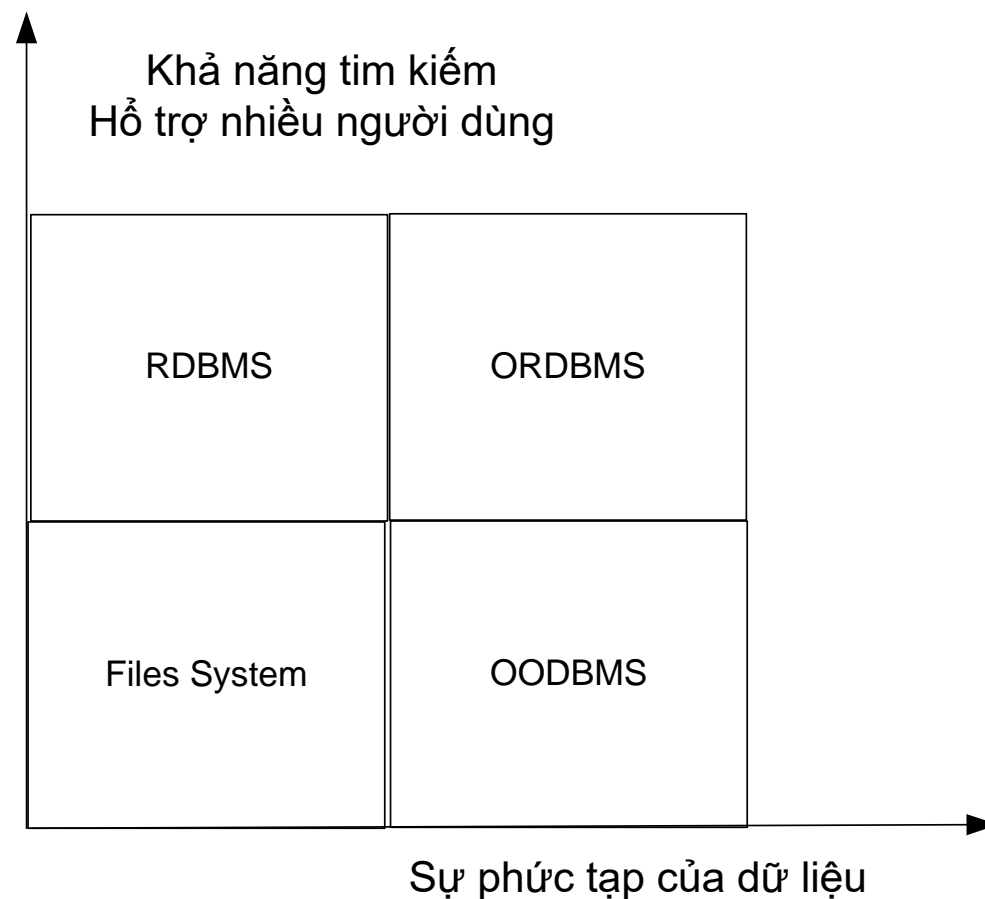
Hệ quản trị CSDL

- Một cơ sở dữ liệu (Database) là một tập hợp các tập tin có liên quan với nhau, được thiết kế theo một chuẩn chung được các nhà khoa học đề xuất.
- Một hệ quản trị cơ sở dữ liệu (Database Management System) là một hệ thống gồm các CSDL và các thao tác trên CSDL đó, được thiết kế trên một nền tảng phần cứng, phần mềm và với một kiến trúc nhất định.
- Các hệ QTCSDL thông dụng: *SQL Server, MySQL, Oracle, DB2, ...*

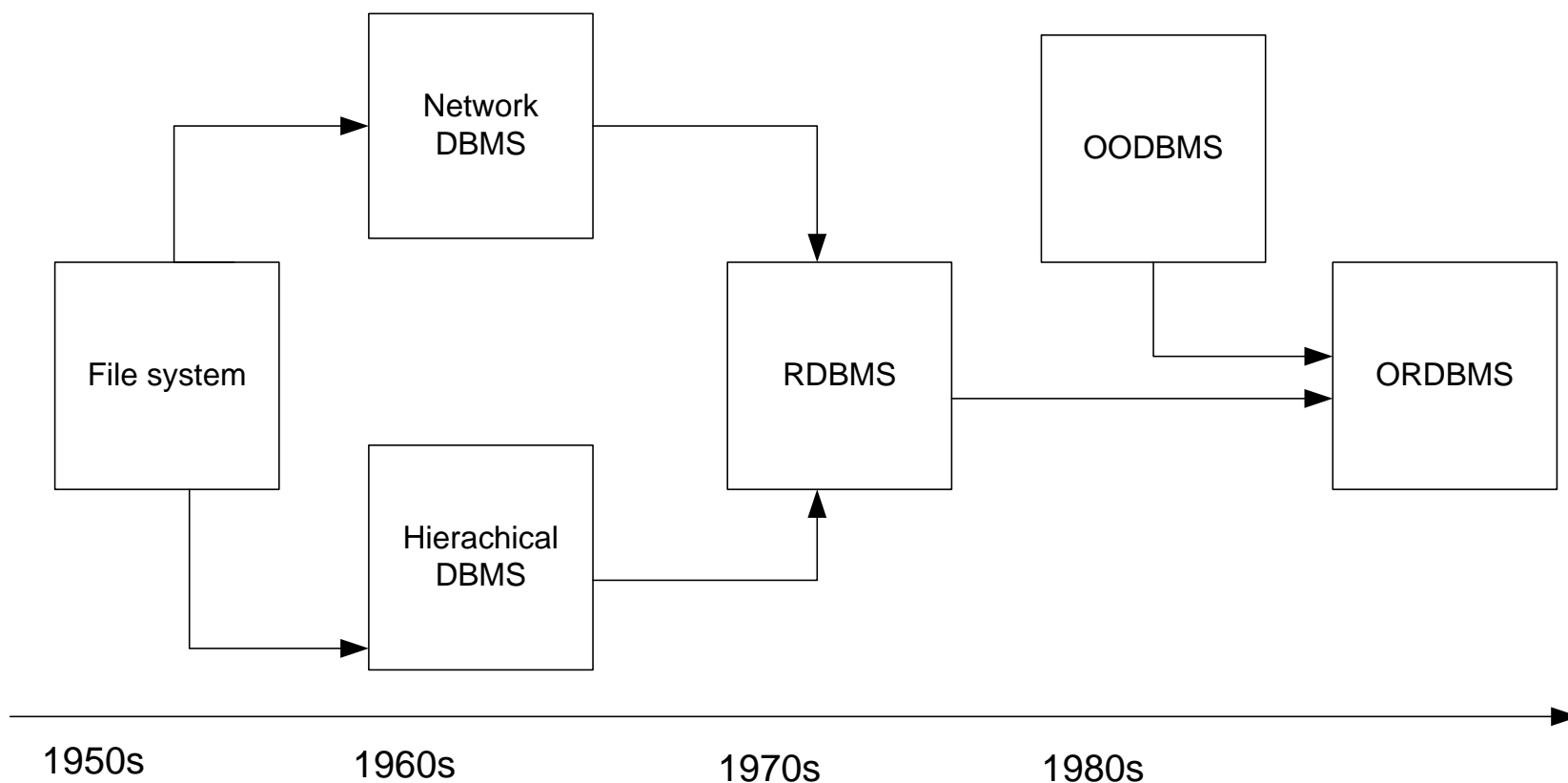
Hệ quản trị CSDL



Các mô hình quản trị CSDL



Lịch sử phát triển các mô hình CSDL



Các tiêu chí phân loại CSDL

- Dựa theo mô hình: Quan hệ; Mạng; Cây; Đối tượng.
- Dựa vào số người dùng: Single user; Multi-user.
- Dựa vào sự phân bố dữ liệu: Tập trung (centralized DB); Phân tán (distributed DB).
- Dựa vào đồng nhất của các nguồn dữ liệu: Đồng nhất; Không đồng nhất.
- Dựa vào chi phí: Thấp: [100, 10000]; Trung bình: [10000, 100000]; Lớn: >100000USD.
- Dựa vào mục đích: Tổng quát; Chuyên biệt.

Các thao tác chính trên CSDL

Các thao tác trên 1 hệ quản trị CSDL

- Thu thập dữ liệu
- Làm sạch dữ liệu
- Phân tích dữ liệu
- Xử lý dữ liệu
- Khôi phục dữ liệu
- Sao lưu dữ liệu

1. Thu thập dữ liệu (Data collection)

- Dữ liệu được thu thập từ nhiều nguồn khác nhau.
- Các yêu cầu được đặt ra bởi các nhà phân tích, khách hàng.
- Các dữ liệu cũng có thể được thu thập từ các cảm biến trong môi trường (camera giao thông, vệ tinh, thiết bị ghi âm, vv)
- Có thể thu được thông qua các cuộc phỏng vấn, tải từ các nguồn trực tuyến, hoặc tài liệu đọc.

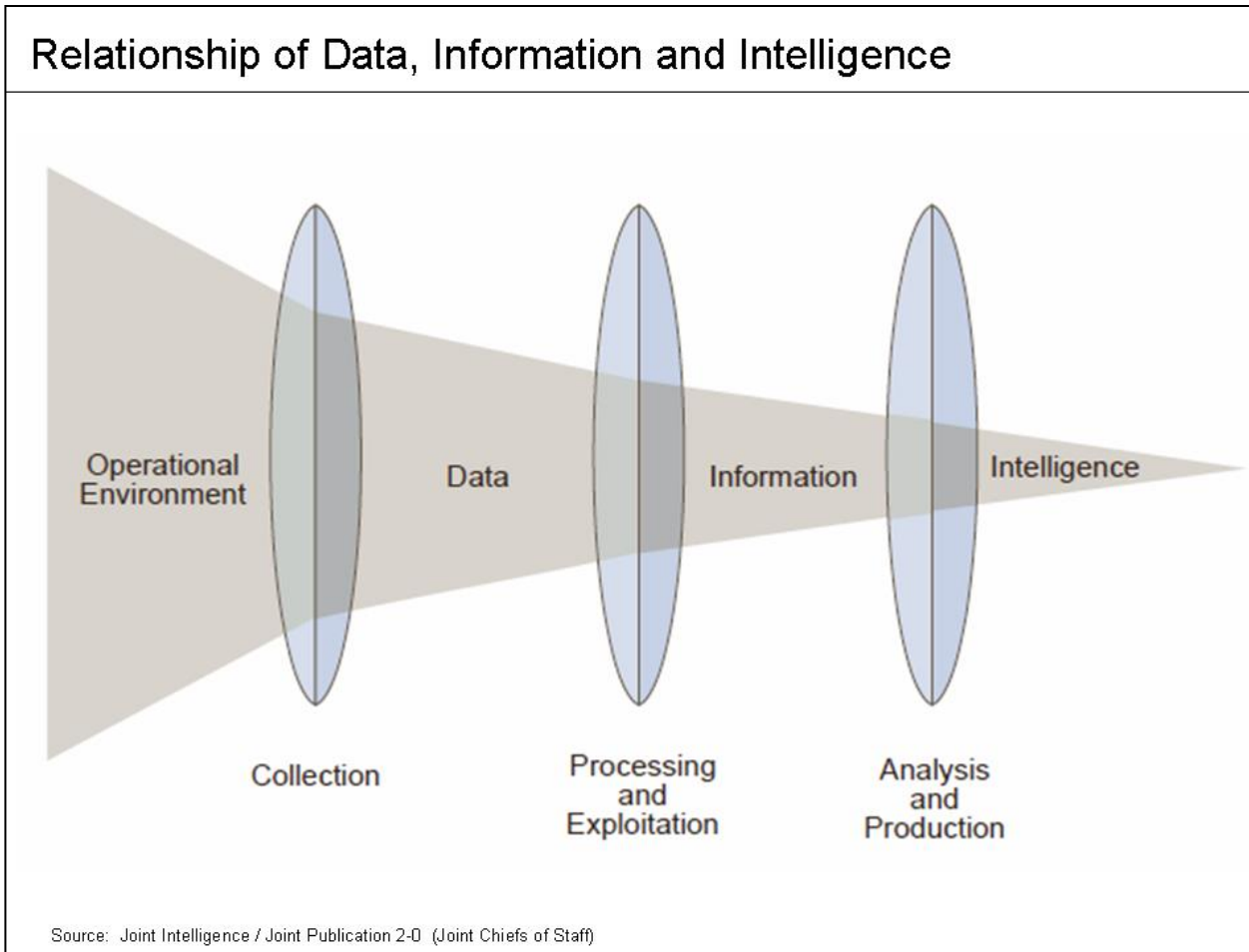
2 - Làm sạch dữ liệu (Data cleaning)

- Sau khi xử lý và tổ chức, các dữ liệu có thể:
 - + Không đầy đủ.
 - + Bị trùng lặp.
 - + Có sai sót.
- Sự cần thiết phải làm sạch dữ liệu sẽ phát sinh từ phương cách mà dữ liệu đã được nhập và lưu trữ.
- Làm sạch dữ liệu là quá trình ngăn ngừa và sửa chữa các lỗi này.

3 - Phân tích dữ liệu (Data analysis)

- Một khi dữ liệu được làm sạch, nó có thể được phân tích.
- Các nhà phân tích có thể áp dụng một loạt các kỹ thuật phân tích dữ liệu được gọi là thăm dò để bắt đầu tìm hiểu các thông điệp chứa trong dữ liệu.
- Quá trình thăm dò có thể dẫn đến làm sạch dữ liệu bổ sung hoặc yêu cầu bổ sung cho dữ liệu, do đó, những hoạt động có thể được lặp đi lặp lại trong tự nhiên.
- Phương pháp thống kê có thể được tạo ra để giúp hiểu được dữ liệu (sum, max, min, avg, count. . .).
- Trực quan dữ liệu cũng có thể được sử dụng để kiểm tra dữ liệu ở định dạng đồ họa, để có được cái nhìn sâu sắc thêm về các thông điệp trong dữ liệu (insight).

3.1 – Quy trình phân tích dữ liệu (Data analysis)



4 - Xử lý dữ liệu (Data Processing)

- Gồm các giai đoạn khôn ngoan được sử dụng để chuyển đổi các dữ liệu thô thành dữ liệu tinh luyện hay kiến thức.
- Dữ liệu ban đầu thu được phải được xử lý hoặc tổ chức để phân tích.

Ví dụ, đưa dữ liệu thành các hàng và cột trong một bảng để phân tích thêm, nhờ các công cụ có sẵn.

5 - Khôi phục dữ liệu

- Khôi phục dữ liệu gồm 2 thao tác chính:
 - + Khôi phục dữ liệu file.
 - + Khôi phục CSDL.

Các nguyên nhân mất mát dữ liệu

— Lỗi HĐH (lỗi logic):

- + Sao chép các dữ liệu quan trọng trên một thiết bị lưu trữ khác trước khi có sự cố
- + Nếu xảy ra sự cố, có thể khôi phục dữ liệu như phần mềm chuyên biệt: Testdisk. . .

— Lỗi đĩa (lỗi physic):

- + Do nhiều nguyên nhân: trầy trên bề mặt, hỏng động cơ.
- + Dữ liệu bị mất ít hoặc nhiều.
- + Phụ thuộc tình hình mà dùng các phần mềm khôi phục dữ liệu
- + Nên vận hành máy tính đúng qui tắc để tránh thiệt hại ở dạng này.

— Xóa nhầm (lỗi chủ quan):

- + Thông thường các files chỉ bị đánh dấu xóa chứ không xóa thực sự trong thời gian đầu. Một số công cụ có thể hỗ trợ khôi phục trong tình huống này.
- + Nếu file xóa đã bị ghi chồng lên (overwrite) thì việc khôi phục là không thể hoặc rất khó khăn.

Khôi phục dữ liệu file

- Phục hồi dữ liệu là quá trình **cứu vớt** (salvage) và xử lý các dữ liệu **từ các dữ liệu bị hư hỏng**, hoặc các phương tiện lưu trữ thứ cấp không thể tiếp cận bình thường.
- Một số tình huống đặc biệt, khôi phục dữ liệu được hiểu là:
 - + Chuyển dữ liệu từ mã hóa đến không mã hoá. (Ví dụ: giải mã file nén winrar bị đặt mật khẩu.
 - + Từ ẩn (hidden) đến không ẩn (unhidden).
- Các phần mềm hỗ trợ khôi phục: *Ultimate Boot CD for Windows*, *Boot Repair Disk*, *Recova*, *Recover my files*.

Khôi phục CSDL

- Khôi phục CSDL là đưa CSDL về trạng thái nhất quán gần nhất, nghĩa là khôi phục về trạng thái đúng gần nhất ở quá khứ.
- Để có thể khôi phục được cần giữ các thông tin làm thay đổi dữ liệu trong thời gian thực hiện giao tác (transaction), các thông tin này nằm ngoài CSDL, nó được ghi trong Systemlog.
- Systemlog: file nhật kí hệ thống sử dụng cho việc khôi phục.
- Chiến thuật
 - + CSDL hổng nặng: sử dụng file backup gần nhất để khôi phục.
 - + CSDL hổng nhẹ: dùng cơ chế UNDO và REDO.

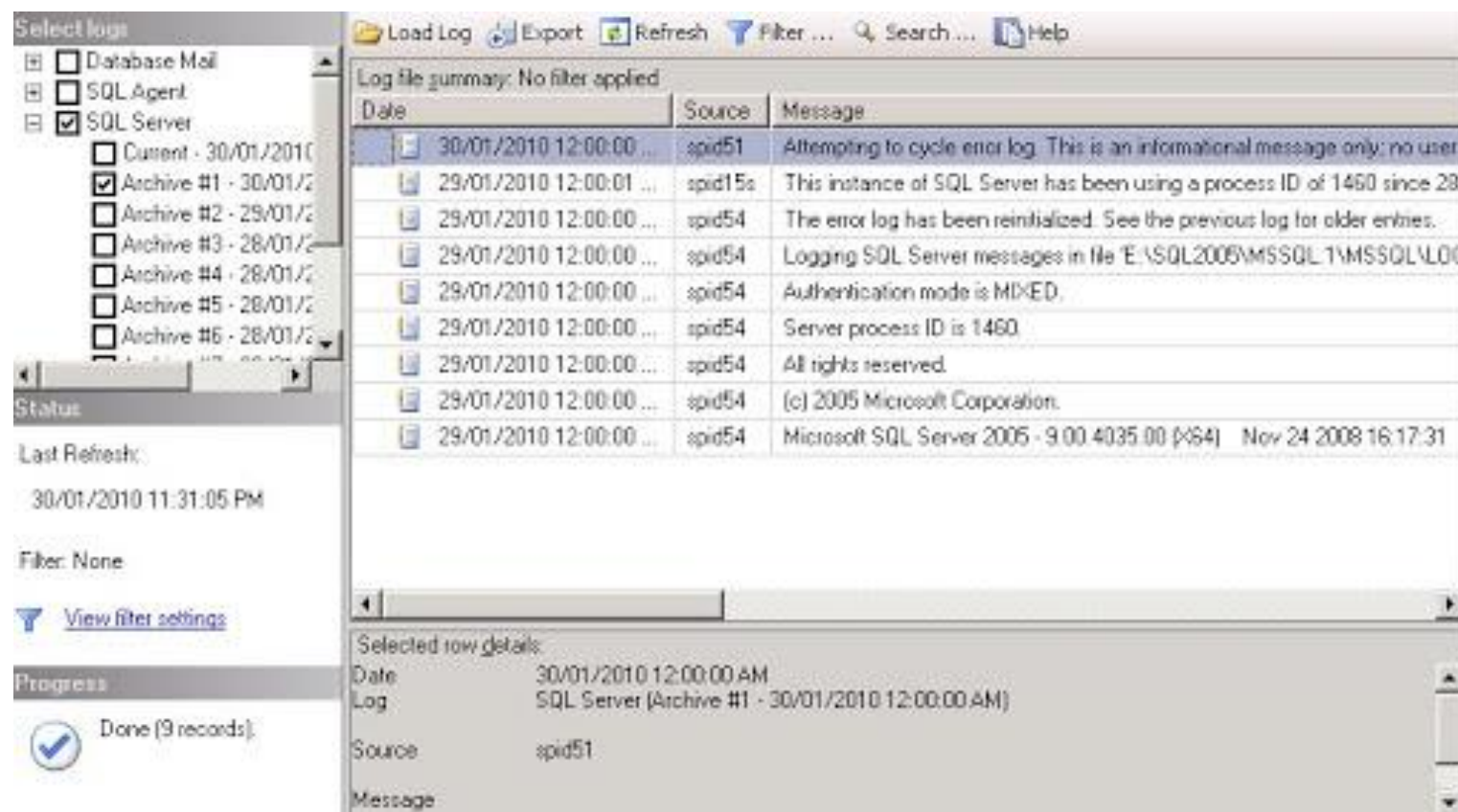
Transaction

- Giao tác là **tập các hành vi** được định nghĩa như một đơn vị nhỏ nhất **khi thao tác với CSDL** (ví dụ: insert vào CSDL là 1 giao tác).
- Một transaction chỉ có thể tồn tại:
 - + Thành công
 - + Không làm gì (All or nothing).
- Một giao tác thất bại thì cập nhật CSDL buộc phải **rollback**, tất cả dữ liệu bị thay đổi bởi giao tác cần khôi phục về giá trị trước đó (before value).
- Một giao tác T1 bị rollback có thể kéo theo T2, T3, ... bị rollback, do T2, T3, ... đọc các dữ liệu cần bị undo của T1. Các trans T2, T3 ... rollback kéo theo T1', T2' ... rollback theo cũng do nguyên nhân giống trên và chúng tạo thành hiện tượng cascade rollback.

Systemlog

- Systemlog là công cụ giống như là Nhật ký, dùng để ghi lại mọi sự thay đổi ảnh hưởng đến hệ thống tính từ thời điểm hiện tại.
- System log là công cụ cực kỳ quan trọng, giúp cho người quản trị có thể nắm bắt được những gì đang xảy ra trong hệ thống, để từ đó có hướng xử lý phù hợp.

Ví dụ System log



Có 2 thông tin quan trọng cần biết khi xem log:

1. Thời gian thực hiện giao tác.

2. Nội dung giao tác.

6 – Sao lưu dữ liệu

- Trong CNTT, một bản sao lưu, hoặc quá trình sao lưu, đề cập đến việc sao chép và lưu trữ dữ liệu máy tính và sử dụng nó để khôi phục lại dữ liệu ban đầu sau khi dữ liệu hư hay mất.
- Các thiết bị dùng để sao lưu:
 - + Hard disk trên máy tính.
 - + Thiết bị ngoại vi: USB, CD, Ổ cứng rời.
 - + Sao lưu qua mạng.
 - + Clouding server (AWS, Google cloud platform, ...).

TỔNG KẾT

1. Quản lý thông tin là một tập các hành vi lớn, liên quan đến việc tổ chức và xử lý thông tin trên máy tính.
2. Đối tượng chính của quản lý thông tin là: dữ liệu và thông tin.
3. Hệ quản trị CSDL là công cụ được tạo ra để quản lý thông tin trên máy tính.
4. Các thao tác chính trên 1 hệ Quản trị CSDL: thu thập dữ liệu, làm sạch dữ liệu, phân tích dữ liệu, xử lý dữ liệu, khôi phục dữ liệu, sao lưu dữ liệu.

TÀI LIỆU THAM KHẢO

1. Nguyễn Gia Tuấn Anh, Trương Châu Long, *Bài tập và bài giải SQL Server*, NXB Thanh niên (2005).
2. Đỗ Phúc, Nguyễn Đăng Ty, *Cơ sở dữ liệu*, NXB Đại học quốc gia TP HCM (2010).
3. Nguyễn Gia Tuấn Anh, Mai Văn Cường, Bùi Danh Hương, *Cơ sở dữ liệu nâng cao*, NXB Đại học quốc gia TP HCM (2019).
4. Itzik Ben-Gan, *Microsoft SQL Server 2012- TSQL Fundamentals*.



BÀI TẬP (1)

1. Hãy tổ chức các folder trên máy tính thể hiện tính khoa học với các hoạt động phổ biến của một sinh viên.
2. Hãy tổ chức các thư nhận và gửi trên gmail đảm bảo tính khoa học với các hoạt động phổ biến của một sinh viên.
3. Cho biết các file quan trọng phổ biến nào mà một sinh viên cần backup.

BÀI TẬP (2)

4. Sự khác nhau giữa việc lưu trữ dữ liệu trên CD và USB?
5. Khi tạo 1 CSDL trong SQL Server, cho biết hệ thống tối thiểu tạo bao nhiêu file, ý nghĩa mỗi file là gì?
6. Hãy thực hiện việc backup và restore 1 CSDL bằng SQL Server bằng 2 cách:

Full Backup

Different Backup