

Chương 1: TỔNG QUAN

Khuyết điểm của Hệ Thống Tập Tin

- ❑ Sự dư thừa dữ liệu và không nhất quán trong dữ liệu
 - Thông tin trùng lặp trong những tập tin khác nhau => Dữ liệu tăng nhanh.
- ❑ Sự phụ thuộc giữa dữ liệu và chương trình
- ❑ Chương trình ứng dụng với các chức năng cố định
 - Cần liên hệ với nhà lập trình của ứng dụng để viết cho mỗi công việc mới
- ❑ Sự cô lập dữ liệu — nhiều tập tin và dạng thức.
 - Khó khăn trong việc truy xuất dữ liệu

Nội Dung

- ❑ Giới thiệu CSDL
 - Lịch sử hình thành CSDL
 - Các khái niệm cơ bản
 - Kiến trúc ba-mức ANSI_SPARC
 - Sự độc lập dữ liệu
 - Các mô hình dữ liệu
 - Các vai trò trong môi trường CSDL
- ❑ Hệ quản trị CSDL
 - Các chức năng của HQTCSL
 - Các thành phần của HQTCSL
 - Kiến trúc cho HQTCSL đa người dùng
 - Ưu, khuyết điểm của HQTCSL

Sự hình thành CSDL

- ❑ Sự giới hạn của tiếp cận hệ thống tập tin chủ yếu là do:
 - Định nghĩa dữ liệu được nhúng trong các CTUĐ, chứ không được lưu trữ một cách riêng biệt và độc lập
 - Việc truy cập và thao tác trên dữ liệu không nằm ngoài những gì các CTUĐ đã cài đặt.
- ❑ Một tiếp cận mới ra đời có thể giải quyết các vấn đề này: Cơ sở dữ liệu (CSDL) và Hệ quản trị CSDL

Hệ Thống Tập Tin (file-based system)

- ❑ Hệ thống tập tin là một tập hợp các chương trình ứng dụng (CTUD) phục vụ cho người dùng cuối (end-users); chẳng hạn như, tạo báo cáo. *Mỗi chương trình định nghĩa và quản lý dữ liệu của chính nó.*
- ❑ Ví dụ: Xét hệ thống tin học của một trường Đại học:
 - HT tập tin của Phòng Đào Tạo tổ chức lưu trữ: Hồ sơ tất cả giảng viên, nhân viên của Trường, Hồ sơ sinh viên, quá trình học tập và kết quả.
 - HT tập tin của Phòng Tài Vụ cũng muốn lưu trữ: Hồ sơ của giảng viên, nhân viên và sinh viên.
 - HT tập tin tại mỗi Khoa cũng cần có: Hồ sơ giảng viên của Khoa, sinh viên của Khoa, kết quả học tập của các sinh viên này.
- ❑ Một tập tin thực chất là một tập hợp các mẫu tin, trong đó mỗi mẫu tin chứa các dữ liệu có liên quan luận lý với nhau.

Cơ Sở Dữ Liệu (Database - DB)

- ❑ Cơ Sở Dữ Liệu:
 - Là một tập hợp chia sẻ các dữ liệu có liên quan với nhau về các đối tượng cần quản lý, được thiết kế để đáp ứng nhu cầu thông tin của một tổ chức.
 - Được lưu trữ trên các thiết bị lưu trữ của máy tính
 - Được quản lý theo cơ chế thống nhất của hệ quản trị CSDL nhằm thực hiện 3 chức năng: Mô tả dữ liệu, Cập nhật dữ liệu, Tìm kiếm dữ liệu

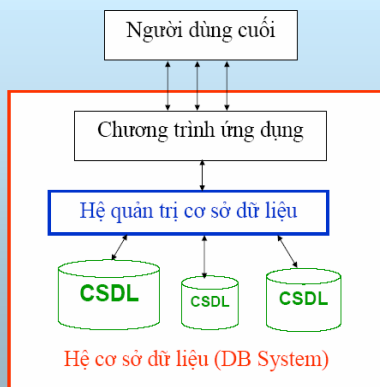
Hệ Quản Trị Cơ Sở Dữ Liệu (Database Management System - DBMS)

- Hệ quản trị CSDL (HQTCSDL): Là một hệ thống phần mềm cho phép người dùng định nghĩa, tạo và duy trì CSDL đồng thời cung cấp dịch vụ truy cập đến CSDL này một cách có quản lý.
- Một số HQTCSDL thông dụng:
 - Microsofts: MS Access, Visual Foxpro, MS SQL Server
 - Oracle: Oracle
 - IBM: DB2
 - Phần mềm nguồn mở: MySQL, PostgreSQL, ...

Lược đồ & thể hiện của CSDL (Database Schemas & Instances)

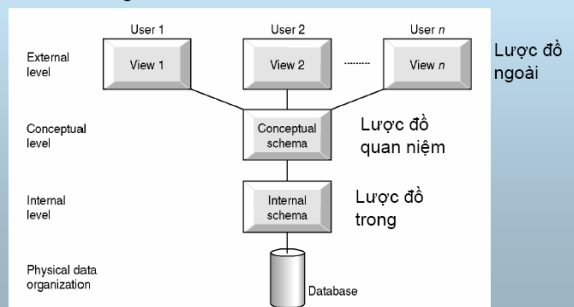
- Lược đồ CSDL (DB Schema) là cấu trúc logic của CSDL
 - Ví dụ: 1 CSDL bao gồm những thông tin về một tập khách hàng, tài khoản và mối liên hệ giữa chúng:
 - KH_HANG gồm có HOTEN_KH, DIACHI_KH, TEL
 - TAI_KHOAN gồm có MASO_TK, NGAY_MO, SO_TIEN
 - KH_HANG có mối quan hệ "sở hữu" đối với TAI_KHOAN
 - Tương tự với thông tin về kiểu của một biến trong 1 chương trình
- Thể hiện của CSDL (DB Instance) là dữ liệu thật sự, nội dung thật sự của CSDL tại một thời điểm nào đó.
 - Ví dụ: 1 CSDL bao gồm những thông tin về một tập khách hàng, tài khoản và mối liên hệ giữa chúng:
 - KH_HANG_1 ("Nguyễn Văn A", "1 Ly Tu Trong", "831301")
 - TAI_KHOAN_1 ("0010221025", "1/1/2007", 50000)
 - KH_HANG_1 "sở hữu" TAI_KHOAN_1
 - Tương tự với giá trị của một biến.

Hệ CSDL (Database System)



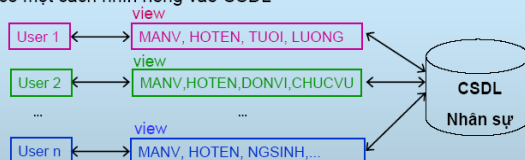
Kiến Trúc Ba - Mức ANSI_SPARC (ANSI_SPARC three – level architecture)

- Mục đích: Tách biệt mỗi khung nhìn của người dùng khỏi cách thức lưu trữ vật lý của cơ sở dữ liệu => tách chương trình khỏi các biểu diễn bên dưới



Khung nhìn dữ liệu (View)

- Khung nhìn: là một tập con của CSDL cho phép mỗi người dùng có một cách nhìn riêng vào CSDL

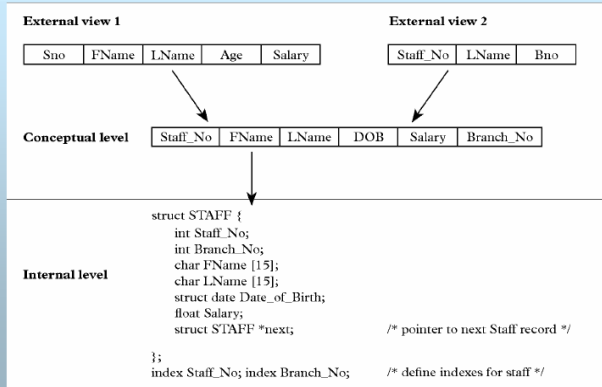


- Lợi ích của Khung nhìn:
 - Cung cấp một mức độ an toàn cho dữ liệu
 - Cung cấp một cơ chế để điều chỉnh diện mạo của cơ sở dữ liệu theo từng loại người dùng.
 - Một khung nhìn thể hiện cho một hình ảnh nhất quán, không thay đổi của cấu trúc của CSDL

Kiến Trúc Ba - Mức ANSI_SPARC (tt.)

- Mức ngoài (External Level): các khung nhìn CSDL của các người dùng khác nhau, mô tả dữ liệu cho mỗi khung nhìn.
- Mức quan niệm (Conceptual Level): mô tả dữ liệu gì được lưu trữ và mối quan hệ giữa chúng. Bao gồm:
 - Cấu trúc logic của toàn CSDL, các ràng buộc trên dữ liệu: thể hiện trên lược đồ quan niệm.
 - Thông tin về ngữ nghĩa của dữ liệu,
 - Các thông tin về an ninh, toàn vẹn,...
- Mức trong (Internal Level): mô tả cách biểu diễn vật lý của CSDL trên máy, thể hiện trên lược đồ trong.

Kiến Trúc Ba - Mức ANSI_SPARC (tt.)



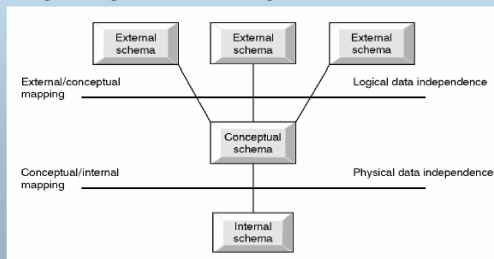
Mô hình quan hệ thực thể (Entity Relationship Model - ERM)

- ERM được sử dụng rộng rãi để thiết kế CSDL
- ERM gồm các khái niệm chính:
 - Thực thể (Entities): là một đối tượng, khái niệm tồn tại độc lập
 - Ví dụ: khách hàng, tài khoản...
 - Mối quan hệ giữa các thực thể (Relationship)
 - Ví dụ: KH Nguyễn Văn A "gửi tiền" vào tài khoản A-101



Sự Độc Lập Dữ Liệu (Data Independence)

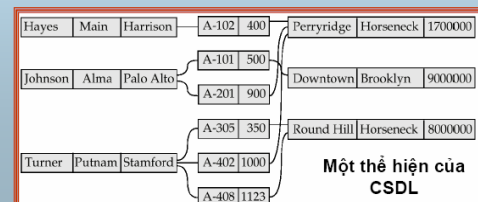
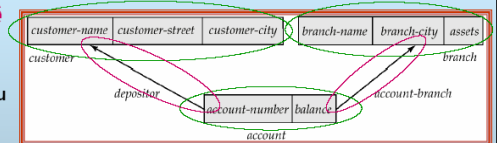
- Độc lập DL ở mức luận lý:** Các thay đổi của lược đồ quan niệm không làm ảnh hưởng đến các lược đồ ngoài.
- Độc lập DL ở mức vật lý:** Các thay đổi của lược đồ trong không làm ảnh hưởng đến lược đồ khái niệm.



Mô Hình Dữ Liệu Mạng (Network Data Model)

- Gồm 2 khái niệm: các **tập dữ liệu** và các **tập quan hệ**

Lược đồ dữ liệu

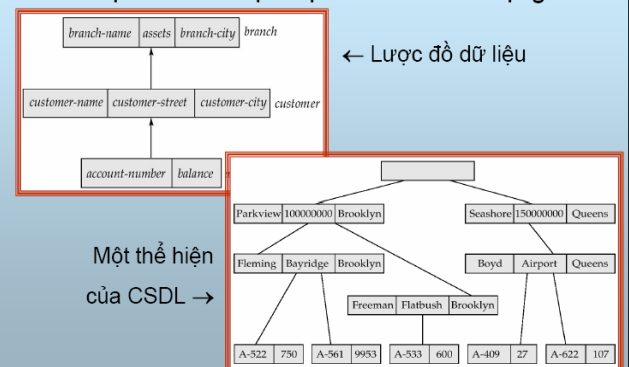


Các mô hình dữ liệu (Data Models)

- Mô hình dữ liệu: tập hợp các khái niệm dùng để mô tả và thao tác dữ liệu, các mối quan hệ giữa dữ liệu, và ràng buộc trên dữ liệu trong một tổ chức
- Có nhiều mô hình dữ liệu (DL), được chia làm 3 loại:
 - Mô hình DL dựa trên đối tượng (object-based data models):
 - Mô hình thực thể - quan hệ
 - Mô hình hướng đối tượng
 - Mô hình DL dựa trên mẫu tin (record-based data models):
 - Mô hình DL mạng
 - Mô hình DL phân cấp
 - Mô hình DL quan hệ
 - Mô hình DL vật lý (physical data models)
 - Mô hình Unifying
 - Mô hình Frame Memory

Mô Hình Dữ Liệu Phân Cấp (Hierarchical Data Models)

- Là một mô hình đặc biệt của mô hình mạng



Chi tiết trong chương 2

Mô hình quan hệ (Relational Data Models)

- Mô hình quan hệ gồm nhiều *quan hệ (relation)*, hay *bảng (table)*
- Mỗi quan hệ là một tập hợp các *bộ (tuple)*, hay *mẫu tin (record)*.
- Mỗi mẫu tin thể hiện trên một *dòng (row)* gồm nhiều giá trị thể hiện trên nhiều *cột (column)* hay còn gọi là *thuộc tính (attribute)*.
- Ví dụ:

Lược đồ quan hệ:

account(account-number, balance)

Một thể hiện của quan hệ →

account-number	balance
A-101	500
A-215	700
A-102	400
A-305	350
A-201	900
A-217	750
A-222	700

(b) The account table

So Sánh OLTP và OLAP/DS

	OLTP	OLAP/DS
Người dùng	Giao dịch viên, thư ký, kỹ sư CNTT,...	Kỹ sư tri thức
Chức năng	Công việc hàng ngày	Hỗ trợ quyết định (Decision Support)
Tkế CSDL	Hướng ứng dụng	Hướng chủ đề
Dữ liệu	Hiện tại, cập nhật, chi tiết, ...	Lịch sử, tổng hợp, tích hợp đa chiều,...
Công việc	Các giao dịch nhỏ, đơn giản	Các truy vấn rất phức tạp
Kích cỡ	100 MB - GB	100GB-TB
Tiêu chuẩn đánh giá	Hiệu suất của GD	Hiệu suất, đáp ứng của truy vấn

Ví dụ Mô hình quan hệ

□ Lược đồ CSDL:

- *account(account-number, balance)*
- *customer(customer-id, customer-name, customer-street, customer-city)*
- *depositor(account-number, customer-id)*

account-number	balance
A-101	500
A-215	700
A-102	400
A-305	350
A-201	900
A-217	750
A-222	700

(b) The account table

customer-id	customer-name	customer-street	customer-city
192-83-7465	Johnson	12 Alma St.	Palo Alto
019-28-3746	Smith	4 North St.	Rye
677-89-9011	Hayes	3 Main St.	Harrison
182-73-6091	Turner	123 Putnam Ave.	Stamford
		100 Main St.	Harrison
		175 Park Ave.	Pittsfield
		72 North St.	Rye

The customer table

customer-id	account-number
192-83-7465	A-101
192-83-7465	A-201
019-28-3746	A-215
677-89-9011	A-102
182-73-6091	A-305
321-12-3123	A-217
336-66-9999	A-222
019-28-3746	A-201

(c) The depositor table

Một thể hiện của CSDL

Các vai trò trong môi trường CSDL

- Nhà quản trị dữ liệu (DA) và nhà quản trị CSDL (DBAs)
- Người thiết kế CSDL (DB designers)
- Người lập trình ứng dụng (application developers)
- Người dùng cuối (end-users):
 - Người sử dụng sơ cấp (Naive users) – không biết gì về HQTCSDL, giao tiếp với hệ thống bằng cách kích hoạt một chương trình ứng dụng
 - Người sử dụng thành thạo (Sophisticated users) – quen thuộc với cấu trúc dữ liệu và các tiện ích của HQTCSDL.

OLTP & OLAP/DS

- Các CSDL hiện nay được sử dụng trong hai ngữ cảnh chính:
 - OLTP (OnLine Transaction Processing): xử lý giao dịch trực tuyến, hỗ trợ các hoạt động thường nhật của hệ thống
 - VD: Hệ thống tính tiền tại các quầy trong siêu thị
 - OLAP/DS (OnLine Analytical Processing/Decision Support): xử lý phân tích trực tuyến, hỗ trợ hoạt động ra quyết định của ban lãnh đạo
 - VD: Hệ thống hỗ trợ quyết định trong vấn đề nhập hàng cho siêu thị

Hệ Quản Trị Cơ Sở Dữ Liệu



Các chức năng của HQTCSDL (DBMS Functions)

Một HQTCSDL thường cung cấp cho người dùng các chức năng sau:

- ❑ Lưu trữ, truy xuất và cập nhật dữ liệu
 - Ngôn ngữ định nghĩa, thao tác và truy vấn dữ liệu
- ❑ Khả năng truy cập dữ liệu có quản lý với các tính năng:
 - **An toàn (Security):** ngăn chặn người dùng không được phép
 - **Toàn vẹn (Integrity):** duy trì tính nhất quán của dữ liệu
 - Phương tiện bảo đảm ràng buộc trên dữ liệu (data constraints)
 - Hỗ trợ cơ chế quản lý giao dịch (transaction management)
 - **Cạnh tranh (concurrency):** cho phép nhiều truy cập cùng lúc.
 - **Quản lý phục hồi (recovery management):** cho phép phục hồi hệ thống sau sự cố
- ❑ Quản lý danh mục hệ thống (system catalog) hay còn gọi là từ điển dữ liệu (data dictionary) hay meta-data
- ❑ Các tiện ích khác như xuất khẩu, nhập khẩu dữ liệu từ HQTCSDL này sang HQTCSDL khác,...

Ngôn ngữ truy vấn SQL (Structured Query Language)

- ❑ SQL là ngôn ngữ phi thủ tục được sử dụng rộng rãi nhất

➢ Ví dụ: Tìm tên của khách hàng với customer-id 192-83-7465

```
select customer.customer-name
from customer
where customer.customer-id = '192-83-7465'
```

- ❑ Ngôn ngữ SQL thao tác trên một tập hợp các mẫu tin, thay vì trên từng mẫu tin một như các ngôn ngữ thủ tục.

Ngôn ngữ định nghĩa dữ liệu (Data Definition Language – DDL)

- ❑ Là ngôn ngữ cho phép khai báo tên đối tượng, cấu trúc của đối tượng và các quan hệ giữa các đối tượng (định nghĩa lược đồ quan hệ)
 - Ví dụ: CREATE TABLE account (

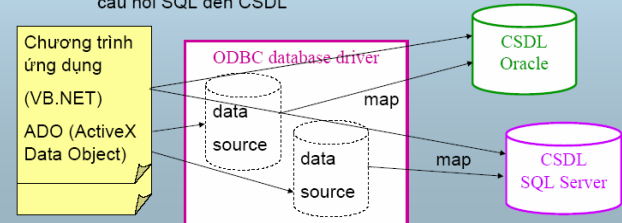
account-number char(10),
balance integer

)
- ❑ Kết quả sau khi thực thi các câu lệnh DDL là một tập các bảng (table) được lưu trữ trong một từ điển dữ liệu (data dictionary)
- ❑ Từ điển dữ liệu bao gồm metadata:
 - Lược đồ CSDL
 - Tên của người dùng có quyền sử dụng CSDL
 - Các mục dữ liệu mà mỗi người dùng có thể truy cập và kiểu truy cập (cho phép thêm, cập nhật, xóa hay chỉ đọc).

Ngôn ngữ truy vấn SQL (Structured Query Language)

- ❑ Những chương trình ứng dụng truy xuất CSDL thông qua việc sử dụng:

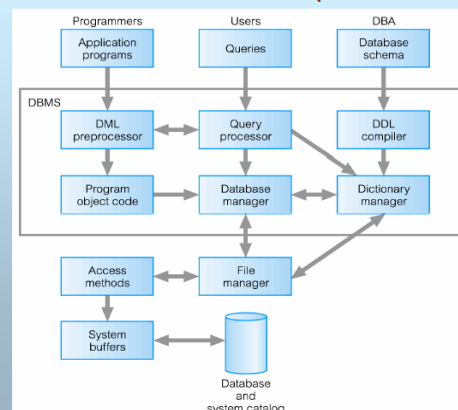
- Một ngôn ngữ mở rộng cho phép nhúng SQL
 - Foxpro for Dos, Visual Foxpro,...
- Một giao diện chương trình ứng dụng như ODBC (Open DB Connectivity) hay JDBC (Java DB Connectivity), để gửi các câu hỏi SQL đến CSDL

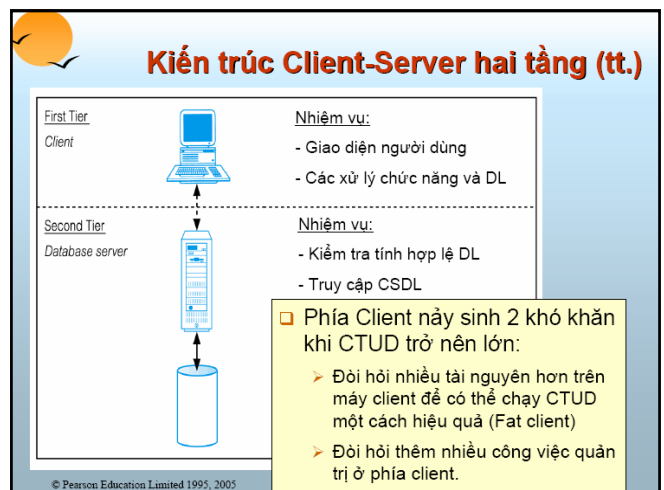
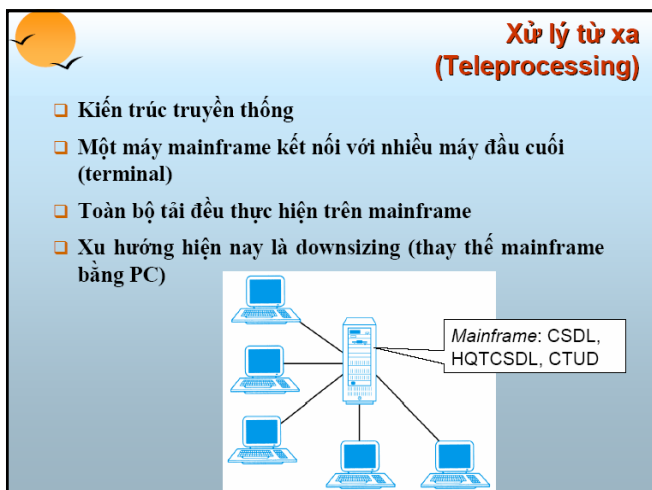
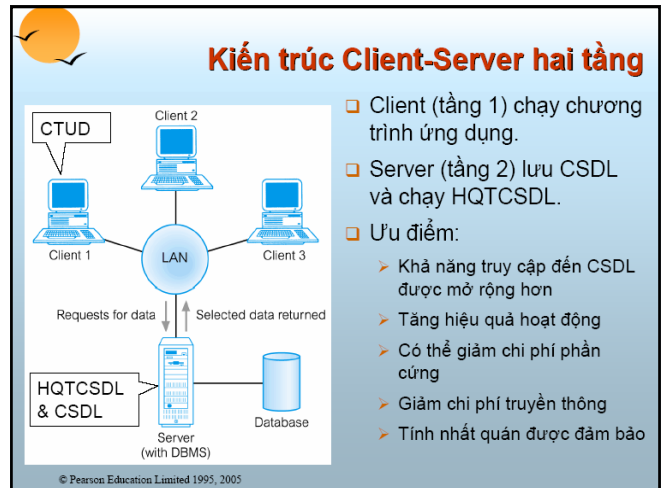
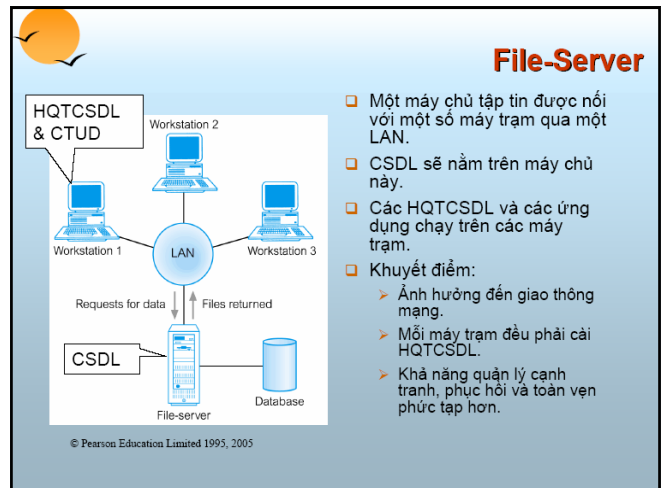
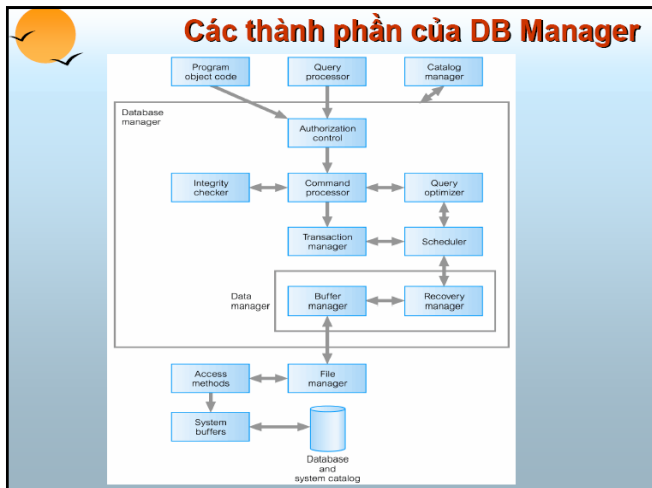


Ngôn ngữ thao tác dữ liệu (Data Manipulation Language – DML)

- ❑ Là ngôn ngữ dùng cho việc truy xuất và thao tác dữ liệu đã được tổ chức theo một mô hình dữ liệu thích hợp (DML cũng được xem như là ngôn ngữ hỏi – query language)
- ❑ Thường bao gồm các thao tác: Thêm dữ liệu (insert), cập nhật dữ liệu (update), xóa (delete) dữ liệu, truy xuất (retrieve) dữ liệu trong CSDL.
- ❑ Các kiểu ngôn ngữ:
 - Kiểu thủ tục: người sử dụng chỉ ra dữ liệu gì được yêu cầu và làm thế nào để thu được dữ liệu đó. (Ngôn ngữ Pascal, C,...)
 - Kiểu phi thủ tục (Còn gọi là ngôn ngữ khai báo): người sử dụng chỉ ra dữ liệu gì được yêu cầu và không chỉ ra làm thế nào để thu được dữ liệu đó. (Ngôn ngữ Prolog, SQL,...)

Các thành phần của HQTCSDL (DBMS Components)





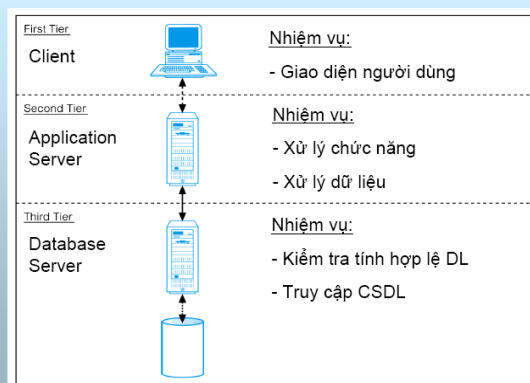
Kiến trúc Client-Server ba tầng

- ❑ Vào khoảng năm 1995, kiến trúc 3 tầng ra đời, mỗi tầng có thể chạy trên các nền khác nhau.
- ❑ Ưu điểm:
 - Client 'nhỏ', không cần phần cứng đắt tiền.
 - Việc bảo trì ứng dụng diễn ra tập trung.
 - Dễ thay đổi hoặc thay thế một tầng mà không làm ảnh hưởng đến các tầng khác
 - Dễ dàng cân bằng tải
 - Ảnh xạ khá tự nhiên vào môi trường web

Ưu điểm của HQTCSDL (2)

- ❑ Tiết kiệm chi phí
- ❑ Có thể hài hòa các yêu cầu mâu thuẫn
- ❑ Cải tiến khả năng truy cập và khả năng đáp ứng dữ liệu
- ❑ Tăng hiệu suất làm việc của các lập trình viên, từ đó giảm thời gian phát triển ứng dụng
- ❑ Cải tiến công tác bảo trì thông qua sự độc lập dữ liệu.
- ❑ Tạo khả năng truy cập dữ liệu cạnh tranh
- ❑ Cải tiến dịch vụ sao lưu và phục hồi

Kiến trúc Client-Server ba tầng (tt.)



Nhược điểm của HQTCSDL

- ❑ Các HQTCSDL tốt thì khá phức tạp
- ❑ Các HQTCSDL tốt có kích cỡ rất lớn, chiếm nhiều dung lượng nhớ.
- ❑ Giá cả của các HQTCSDL rất khác nhau tùy theo môi trường và chức năng.
- ❑ Chi phí thêm cho phần cứng
- ❑ Chi phí để chuyển đổi sang HTCSDL mới
- ❑ HQTCSDL được viết tổng quát cho mọi ứng dụng nên tốc độ thực thi có thể chậm.
- ❑ Lỗi của một bộ phận HQTCSDL sẽ ảnh hưởng đến nhiều người dùng và nhiều ứng dụng hơn.

Ưu điểm của HQTCSDL (1)

- ❑ Quản lý được dữ liệu dư thừa
- ❑ Đảm bảo tính nhất quán dữ liệu
- ❑ Tạo khả năng chia sẻ dữ liệu nhiều hơn
- ❑ Cải tiến tính toàn vẹn cho dữ liệu
- ❑ Cải tiến tính an toàn cho dữ liệu
- ❑ Có thể đặt ra các chuẩn về định dạng dữ liệu, về thủ tục cập nhật, qui tắc truy cập,...