

# CHƯƠNG I

## GIỚI THIỆU

### (Introduction)

#### MỤC ĐÍCH

Chương này trình bày một cái nhìn bao quát về cơ sở dữ liệu (CSDL/DB), về hệ quản trị cơ sở dữ liệu (HQTCSDL/DBMS) và về hệ cơ sở dữ liệu (HCSDL/DBS). Các đòi hỏi khi xây dựng một HQTCSDL đó cũng chính là những chức năng mà một HCSDL cần phải có. Một khái niệm quan trọng là khái niệm giao dịch (Transaction). Các tính chất một giao dịch phải có để đảm bảo một HQTCSDL, được xây dựng trên HCSDL tương ứng, trong suốt quá trình hoạt động sẽ luôn cho một CSDL tin cậy (dữ liệu luôn nhất quán). Quản trị giao dịch nhằm đảm bảo mọi giao dịch trong hệ thống có các tính chất mà một giao dịch phải có. Một điều cần chú ý là trong các tính chất của một giao dịch, *tính chất nhất quán* trước hết phải được đảm bảo bởi người lập trình-người viết ra giao dịch.

#### YÊU CẦU

Hiểu các khái niệm.

Hiểu các vấn đề đặt ra khi xây dựng một HQTCSDL: thiết kế CSDL, đảm bảo tính nhất quán của CSDL trong suốt cuộc sống của nó, nền tảng phần cứng trên đó một HQTCSDL được xây dựng.

Hiểu cấu trúc hệ thống tổng thể

Hiểu vai trò của các người sử dụng hệ thống.

#### MỘT SỐ KHÁI NIỆM

- Một cơ sở dữ liệu (CSDL/ DB: DataBase) là một tập hợp các tập tin có liên quan với nhau, được thiết kế nhằm làm giảm thiểu sự lặp lại dữ liệu.
- Một hệ quản trị cơ sở dữ liệu (HQTCSDL/ DBMS: DataBase Management System) là một hệ thống gồm một 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.
- Một hệ cơ sở dữ liệu (HCSDL/ DBS: DataBase System) là một phần mềm cho phép xây dựng một HQTCSDL.

## HỆ CƠ SỞ DỮ LIỆU

Một số điểm bất lợi chính của việc lưu giữ *thông tin có tổ chức* trong hệ thống xử lý file thông thường:

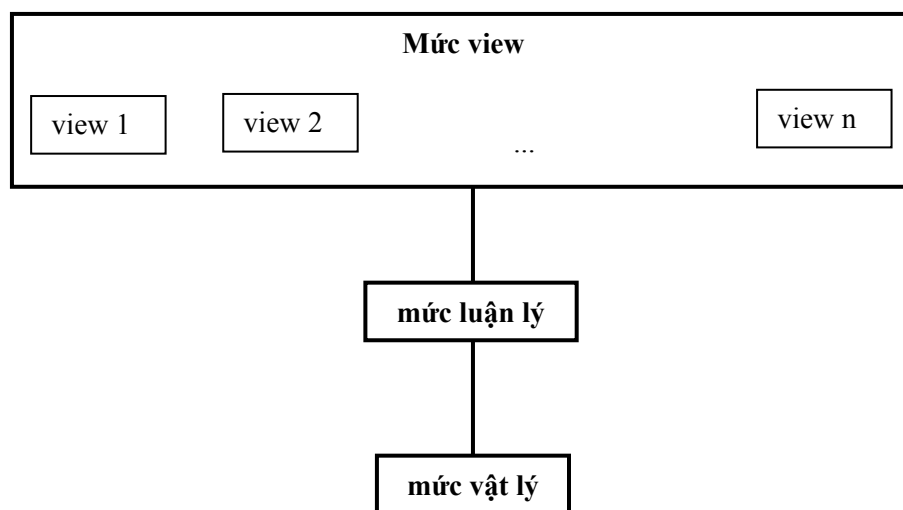
- **Dư thừa dữ liệu và tính không nhất quán** (Data redundancy and inconsistency): Do các file và các trình ứng dụng được tạo ra bởi các người lập trình khác nhau, nên các file có định dạng khác nhau, các chương trình được viết trong các ngôn ngữ lập trình khác nhau, cùng một thông tin có thể được lưu giữ trong các file khác nhau. Tính không thống nhất và dư thừa này sẽ làm *tăng chi phí truy xuất và lưu trữ*, hơn nữa, nó sẽ dẫn đến tính không nhất quán của dữ liệu: *các bản sao của cùng một dữ liệu có thể không nhất quán*.
- **Khó khăn trong việc truy xuất dữ liệu**: Môi trường của hệ thống xử lý file thông thường không cung cấp các công cụ cho phép truy xuất thông tin một cách hiệu quả và thuận lợi.
- **Sự cô lập dữ liệu** (Data isolation): Các giá trị dữ liệu được lưu trữ trong cơ sở dữ liệu phải thỏa mãn một số các *ràng buộc về tính nhất quán của dữ liệu ( ràng buộc nhất quán/consistency constraints )*. Trong hệ thống xử lý file thông thường, rất khó khăn trong việc thay đổi các chương trình để thỏa mãn các yêu cầu thay đổi ràng buộc. Vấn đề trở nên khó khăn hơn khi các ràng buộc liên quan đến các hạng mục dữ liệu nằm trong các file khác nhau.
- **Các vấn đề về tính nguyên tử** (Atomicity problems): Tính nguyên tử của một hoạt động (giao dịch) là: *hoặc nó được hoàn tất trọn vẹn hoặc không có gì cả*. Điều này có nghĩa là một hoạt động (giao dịch) *chỉ làm thay đổi* các dữ liệu bền vững khi nó đã hoàn tất (kết thúc thành công) nếu không, giao dịch không để lại một dấu vết nào trên CSDL. Trong hệ thống xử lý file thông thường khó đảm bảo được tính chất này.
- **Tính bất thường trong truy xuất cạnh tranh**: Một hệ thống cho phép nhiều người sử dụng cập nhật dữ liệu đồng thời, có thể dẫn đến kết quả là dữ liệu không nhất quán. Điều này đòi hỏi một sự giám sát. Hệ thống xử lý file thông thường không cung cấp chức năng này.
- **Vấn đề an toàn** (Security problems): một người sử dụng hệ cơ sở dữ liệu không cần thiết và cũng không có quyền truy xuất tất cả các dữ liệu. Vấn đề này đòi hỏi hệ thống phải đảm bảo được tính phân quyền, chống truy xuất trái phép ...

Các bất lợi nêu trên đã gợi mở sự phát triển các DBMS. Phần sau của giáo trình sẽ đề cập đến các quan niệm và các thuật toán được sử dụng để phát triển một hệ cơ sở dữ liệu nhằm *giải quyết các vấn đề nêu trên*. Một số khái niệm

## GÓC NHÌN DỮ LIỆU

Tính hiệu quả của hệ thống đòi hỏi phải thiết kế các cấu trúc dữ liệu phức tạp để biểu diễn dữ liệu trong cơ sở dữ liệu. Các nhà phát triển che dấu sự phức tạp này thông qua các *mức trừu tượng* nhằm đơn giản hóa sự trao đổi của người sử dụng với hệ thống:

- **Mức vật lý ( Physical level )**: Mức thấp nhất của sự trừu tượng, mô tả dữ liệu hiện được lưu trữ thế nào. Ở mức này, cấu trúc dữ liệu mức thấp, phức tạp được mô tả chi tiết.
- **Mức luận lý ( Logical level )**: Mức kế cao hơn về sự trừu tượng, mô tả dữ liệu gì được lưu trữ trong cơ sở dữ liệu và các mối quan hệ gì giữa các dữ liệu này. Mức logic của sự trừu tượng được dùng bởi các người quản trị cơ sở dữ liệu.
- **Mức view ( view level )**: Mức cao nhất của sự trừu tượng, mô tả chỉ một phần của cơ sở dữ liệu toàn thể. Một người sử dụng cơ sở dữ liệu liên quan đến chỉ một bộ phận của cơ sở dữ liệu. Như vậy sự trao đổi của họ với hệ thống được làm đơn giản bởi việc định nghĩa view. Hệ thống có thể cung cấp nhiều mức view đối với cùng một cơ sở dữ liệu.



*Figure 1*

- **Thể hiện và sơ đồ (instances and schemas)**: Tập hợp các thông tin được lưu trữ trong cơ sở dữ liệu tại một thời điểm được gọi là một thể hiện (instance) của cơ sở dữ liệu. Thiết kế tổng thể của cơ sở dữ liệu được gọi là sơ đồ (schema).

Một hệ cơ sở dữ liệu có một vài sơ đồ, được phân tương ứng với các mức trừu tượng. ở mức thấp nhất là **sơ đồ vật lý** (physical schema), ở mức trung gian là **sơ đồ luận lý** (logical schema), ở mức cao nhất là **sơ đồ con** (subschema). Nói chung một hệ cơ sở dữ liệu hỗ trợ một sơ đồ vật lý, một sơ đồ luận lý và một vài sơ đồ con.

- Khả năng sửa đổi một định nghĩa ở một mức không ảnh hưởng một định nghĩa sơ đồ ở mức cao hơn được gọi là **sự độc lập dữ liệu** (data independence). Có hai mức độc lập dữ liệu:
  - **Độc lập dữ liệu vật lý** (Physical data independence) là khả năng sửa đổi sơ đồ vật lý không làm cho các chương trình ứng dụng phải viết lại. Các sửa đổi ở mức vật lý là cần thiết để cải thiện hiệu năng.

- **Độc lập dữ liệu luận lý** (Logical data independence) là khả năng sửa đổi sơ đồ luận lý không làm cho các chương trình ứng dụng phải viết lại. Các sửa đổi ở mức luận lý là cần thiết khi cấu trúc luận lý của cơ sở dữ liệu bị thay thế.

## **MÔ HÌNH DỮ LIỆU**

Nằm dưới cấu trúc của một cơ sở dữ liệu là mô hình dữ liệu: một bộ các công cụ quan niệm để mô tả dữ liệu, quan hệ dữ liệu, ngữ nghĩa dữ liệu và các ràng buộc nhất quán. Có ba nhóm mô hình: Các mô hình luận lý dựa trên đối tượng (Object-based logical models), các mô hình luận lý dựa trên mẫu tin (record-based logical models), các mô hình vật lý (physical models).

- **Các mô hình luận lý dựa trên đối tượng** được dùng mô tả dữ liệu ở mức luận lý và mức view. Chúng được đặc trưng bởi việc chúng cung cấp khả năng cấu trúc linh hoạt và cho phép các ràng buộc dữ liệu được xác định một cách tường minh. Dưới đây là một vài mô hình được biết rộng rãi: Mô hình **thực thể - quan hệ** (entity-relationship model), mô hình **hướng đối tượng** (object-oriented model), mô hình **dữ liệu ngữ nghĩa** (semantic data model), mô hình **dữ liệu hàm** (function data model).
- **Các mô hình luận lý dựa trên mẫu tin** được dùng để miêu tả dữ liệu ở mức luận lý hay mức view. Chúng được dùng để xác định cấu trúc luận lý toàn thể của cơ sở dữ liệu và cung cấp sự mô tả mức cao hơn việc thực hiện. Cơ sở dữ liệu được cấu trúc ở dạng mẫu tin định dạng cố định (fixed format record): mỗi mẫu tin xác định một số cố định các trường, mỗi trường thường có độ dài cố định. Một vài mô hình được biết rộng rãi là: **Mô hình quan hệ, mô hình mạng, mô hình phân cấp**.
- **Mô hình dữ liệu vật lý** được dùng để mô tả dữ liệu ở mức thấp nhất. Hai mô hình dữ liệu vật lý được biết rộng rãi nhất là **mô hình hợp nhất** (unifying model) và **mô hình khung-bộ nhớ** (frame-memory model).

## **NGÔN NGỮ CƠ SỞ DỮ LIỆU**

Một hệ cơ sở dữ liệu cung cấp hai kiểu ngôn ngữ khác nhau: một để xác định sơ đồ cơ sở dữ liệu, một để biểu diễn các vấn tin cơ sở dữ liệu và cập nhật.

- **Ngôn ngữ định nghĩa dữ liệu (Data Definition Language: DDL)** cho phép định nghĩa sơ đồ cơ sở dữ liệu. Kết quả biên dịch các lệnh của DDL là tập hợp các bảng được lưu trữ trong một *file đặc biệt được gọi là tự điển dữ liệu (data dictionary)* hay thư mục dữ liệu (*data directory*). Tự điển dữ liệu là một file chứa *metadata*. File này được tra cứu trước khi dữ liệu hiện hành được đọc hay sửa đổi. Cấu trúc lưu trữ và phương pháp truy cập được sử dụng bởi hệ cơ sở dữ liệu được xác định bởi một tập hợp các định nghĩa trong một kiểu đặc biệt của DDL được gọi là **ngôn ngữ định nghĩa và lưu trữ dữ liệu (data storage and definition language)**. Kết quả biên dịch của các định nghĩa này là một tập hợp các chỉ thị xác định sự thực hiện chi tiết của các sơ đồ cơ sở dữ liệu (thường được che dấu).
- **Ngôn ngữ thao tác dữ liệu (Data manipulation language: DML)** là ngôn ngữ 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)** yêu cầu người sử dụng đặc tả dữ liệu nào cần và làm thế nào để nhận được nó. **DML không thủ tục (Nonprocedural DML)** yêu cầu người sử dụng đặc tả dữ liệu nào cần nhưng không cần đặc tả làm thế nào để nhận được nó. Một vấn tin (*query*) là một lệnh yêu cầu tìm lại dữ liệu (*information*).

*retrieval*). Phần ngôn ngữ DML liên quan đến sự tìm lại thông tin được gọi là ngôn ngữ vấn tin (*query language*).

## QUẢN TRỊ GIAO DỊCH

Thông thường, một số thao tác trên cơ sở dữ liệu tạo thành một đơn vị logic công việc. Ta hãy xét ví dụ chuyển khoản, trong đó một số tiền  $x$  được chuyển từ tài khoản A ( $A := A - x$ ) sang một tài khoản B ( $B := B + x$ ). Một yếu tố cần thiết là cả hai thao tác này hoặc cùng xảy ra hoặc không hoạt động nào xảy ra cả. Việc chuyển khoản phải xảy ra trong tính toàn thể của nó hoặc không. Đòi hỏi **toàn thể-hoặc-không** này được gọi là tính nguyên tử (atomicity). Một yếu tố cần thiết khác là sự thực hiện việc chuyển khoản bảo tồn tính nhất quán của cơ sở dữ liệu: *giá trị của tổng  $A + B$  phải được bảo tồn*. Đòi hỏi về tính chính xác này được gọi là tính nhất quán (consistency). Cuối cùng, sau khi thực hiện thành công hoạt động chuyển khoản, các giá trị của các tài khoản A và B phải bền vững cho dù có thể có sự cố hệ thống. Đòi hỏi về tính bền vững này được gọi là tính lâu bền (durability).

Một giao dịch là 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 cả tính nguyên tử lẫn tính nhất quán. Như vậy, các giao dịch phải không được vi phạm bất kỳ ràng buộc nhất quán nào: ***Nếu cơ sở dữ liệu là nhất quán khi một giao dịch khởi động thì nó cũng phải là nhất quán khi giao dịch kết thúc thành công***. Tuy nhiên, trong khi đang thực hiện giao dịch, phải cho phép sự không nhất quán tạm thời. Sự không nhất quán tạm thời này tuy là cần thiết nhưng lại có thể dẫn đến các khó khăn nếu xảy ra sự cố.

Trách nhiệm của người lập trình là xác định đúng đắn các giao dịch sao cho mỗi một bảo tồn tính nhất quán của cơ sở dữ liệu.

Đảm bảo tính nguyên tử và tính lâu bền là trách nhiệm của hệ cơ sở dữ liệu nói chung và của **thành phần quản trị giao dịch (transaction-management component)** nói riêng. Nếu không có sự cố, tất cả giao dịch hoàn tất thành công và tính nguyên tử được hoàn thành dễ dàng. Tuy nhiên, do sự hiện diện của các sự cố, một giao dịch có thể không hoàn tất thành công sự thực hiện của nó. Nếu tính nguyên tử được đảm bảo, một giao dịch thất bại không gây hiệu quả đến trạng thái của cơ sở dữ liệu. Như vậy, cơ sở dữ liệu phải được hoàn lại trạng thái của nó trước khi giao dịch bắt đầu. Hệ cơ sở dữ liệu phải có trách nhiệm phát hiện sự cố hệ thống và trả lại cơ sở dữ liệu về trạng thái trước khi xảy ra sự cố.

Khi một số giao dịch cạnh tranh cập nhật cơ sở dữ liệu, tính nhất quán của dữ liệu có thể không được bảo tồn, ngay cả khi mỗi giao dịch là chính xác. **Bộ quản trị điều khiển cạnh tranh (concurrency-control manager)** có trách nhiệm điều khiển các trao đổi giữa các giao dịch cạnh tranh để đảm bảo tính thống nhất của CSDL.

## QUẢN TRỊ LƯU TRỮ

Các CSDL đòi hỏi một khối lượng lớn không gian lưu trữ, có thể lên đến nhiều terabytes (1 terabyte =  $10^3$  Gigabytes =  $10^6$  Megabytes). Các thông tin phải được lưu trữ trên lưu trữ ngoài (đĩa). Dữ liệu được di chuyển giữa lưu trữ đĩa và bộ nhớ chính khi cần thiết. Do việc di chuyển dữ liệu từ và lên đĩa tương đối chậm so với tốc độ của đơn vị xử lý trung tâm, điều này ép buộc hệ CSDL phải cấu trúc dữ liệu sao cho tối ưu hóa nhu cầu di chuyển dữ liệu giữa đĩa và bộ nhớ chính.

Mục đích của một hệ CSDL là làm đơn giản và dễ dàng việc truy xuất dữ liệu. Người sử dụng hệ thống có thể không cần quan tâm đến chi tiết vật lý của sự thực thi hệ thống. Phần lớn họ chỉ quan tâm đến hiệu năng của hệ thống (thời gian trả lời một câu vấn tin ...).

**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 và các câu vấn tin được đệ trình cho hệ thống. Bộ quản trị lưu trữ có trách nhiệm trao đổi với bộ quản trị file (file manager). Dữ liệu thô được lưu trữ trên đĩa sử dụng hệ thống file (file system), hệ thống này thường được cung cấp bởi hệ điều hành. Bộ quản trị lưu trữ dịch các câu lệnh DML thành các lệnh của hệ thống file mức thấp. Như vậy, *bộ quản trị lưu trữ có nhiệm vụ lưu trữ, tìm lại và cập nhật dữ liệu trong CSDL.*

## NHÀ QUẢN TRỊ CƠ SỞ DỮ LIỆU

Một trong các lý do chính đối với việc sử dụng DBMS là có sự điều khiển trung tâm cho cả dữ liệu lẫn các chương trình truy cập các dữ liệu này. Người điều khiển trung tâm trên toàn hệ thống như vậy gọi là nhà quản trị cơ sở dữ liệu (DataBase Administrator - DBA). Các chức năng của DBA như sau:

- **Định nghĩa sơ đồ:** DBA tạo ra sơ đồ CSDL gốc bằng cách viết một tập các định nghĩa mà nó sẽ được dịch bởi trình biên dịch DDL thành một tập các bảng được lưu trữ thường trực trong tự điển dữ liệu.
- **Định nghĩa cấu trúc lưu trữ và phương pháp truy xuất:** DBA tạo ra một cấu trúc lưu trữ thích hợp và các phương pháp truy xuất bằng cách viết một tập hợp các định nghĩa mà nó sẽ được dịch bởi trình biên dịch lưu trữ dữ liệu và ngôn ngữ định nghĩa dữ liệu.
- **Sửa đổi sơ đồ và tổ chức vật lý**
- **Cấp quyền truy xuất dữ liệu:** Việc cấp các dạng quyền truy cập khác nhau cho phép DBA điều hoà những phần của CSDL mà nhiều người có thể truy xuất. Thông tin về quyền được lưu giữ trong một cấu trúc hệ thống đặc biệt, nó được tham khảo bởi hệ CSDL mỗi khi có sự truy xuất dữ liệu của hệ thống.
- **Đặc tả ràng buộc toàn vẹn ( integrity-constraint ):** Các giá trị dữ liệu được lưu trữ trong CSDL phải thoả mãn một số các ràng buộc nhất quán nhất định. Ví dụ số giờ làm việc của một nhân viên trong một tuần không thể vượt quá một giới hạn 80 giờ chẳng hạn. Một ràng buộc như vậy phải được đặc tả một cách tường minh bởi DBA. Các ràng buộc toàn vẹn được lưu giữ trong một cấu trúc hệ thống đặc biệt được tham khảo bởi hệ CSDL mỗi khi có sự cập nhật dữ liệu.

## NGƯỜI SỬ DỤNG CSDL

Mục đích đầu tiên của hệ CSDL là cung cấp một môi trường để tìm lại thông tin và lưu thông tin trong CSDL. Các người sử dụng cơ sở dữ liệu được phân thành bốn nhóm tùy theo cách thức họ trao đổi với hệ thống.

- **Các người lập trình ứng dụng:** Là nhà chuyên môn máy tính người trao đổi với hệ thống thông qua các lời gọi DML được nhúng trong một chương trình được viết trong một ngôn ngữ chủ - *host language* (Pascal, C, Cobol ...). Các chương trình này thường được tham khảo như các chương trình ứng dụng. Vì cú pháp DML thường rất khác với cú pháp của ngôn ngữ chủ, các lời gọi DML thường được bắt đầu bởi một ký tự đặc biệt như vậy mã thích hợp mới có thể được sinh. Một bộ tiền xử lý đặc biệt, được gọi là tiền

biên dịch (precompiler) DML, chuyển các lệnh DML thành các lời gọi thủ tục chuẩn trong ngôn ngữ chủ. Bộ biên dịch ngôn ngữ chủ sẽ sinh mã đối tượng thích hợp. Có những ngôn ngữ lập trình phối hợp cấu trúc điều khiển của các ngôn ngữ giống như Pascal với cấu trúc điều khiển để thao tác đối tượng CSDL. Các ngôn ngữ này (đôi khi được gọi là ngôn ngữ thể hệ thứ tư) thường bao gồm các đặc điểm đặc biệt để làm dễ dàng việc sinh các dạng và hiển thị dữ liệu trên màn hình.

- **Các người sử dụng thành thạo ( *Sophisticated users* )**: Trao đổi với hệ thống không qua viết trình. Thay vào đó họ đặt ra các yêu cầu của họ trong ngôn ngữ truy vấn CSDL ( Database query language ). Mỗi câu văn tin như vậy được đệ trình cho bộ xử lý văn tin, chức năng của bộ xử lý văn tin là "dịch" các lệnh DML thành các chỉ thị mà bộ quản trị lưu trữ hiểu. Các nhà phân tích đệ trình các câu văn tin thăm dò dữ liệu trong cơ sở dữ liệu thuộc vào phạm trù này.
- **Các người sử dụng chuyên biệt ( *Specialized users* )**: Là các người sử dụng thành thạo, họ viết các ứng dụng CSDL chuyên biệt không nằm trong khung xử lý dữ liệu truyền thống. Trong đó, phải kể đến các hệ thống thiết kế được trợ giúp bởi máy tính (computer-aided design systems), Cơ sở tri thức (knowledge-base) và hệ chuyên gia (expert systems), các hệ thống lưu trữ dữ liệu với kiểu dữ liệu phức tạp (dữ liệu đồ họa, hình ảnh, âm thanh) và các hệ thống mô hình môi trường (environment-modeling systems)
- **Các người sử dụng ngây thơ ( *Naive users* )**: là các người sử dụng không thành thạo, họ trao đổi với hệ thống bởi câu dẫn một trong các chương trình ứng dụng thường trực đã được viết sẵn.

## CẤU TRÚC HỆ THỐNG TỔNG THỂ

Một hệ CSDL được phân thành các module, mỗi một thực hiện một trách nhiệm trong hệ thống tổng thể. Một số chức năng của hệ CSDL có thể được cung cấp bởi hệ điều hành. Trong hầu hết các trường hợp, hệ điều hành chỉ cung cấp các dịch vụ cơ sở nhất, hệ CSDL phải xây dựng trên cơ sở đó. Như vậy, thiết kế hệ CSDL phải xem xét đến giao diện giữa hệ CSDL và hệ điều hành.

Các thành phần chức năng của hệ CSDL có thể được chia thành các thành phần xử lý văn tin (query processor components) và các thành phần quản trị lưu trữ (storage manager components ).

Các thành phần xử lý văn tin gồm:

- **Trình biên dịch DML ( *DML compiler* )**: dịch các lệnh DML trong một ngôn ngữ văn tin thành các chỉ thị mức thấp mà engine định giá văn tin ( query evaluation engine ) có thể hiểu. Hơn nữa, Trình biên dịch DML phải biến đổi một yêu cầu của người sử dụng thành một đích tương đương nhưng ở dạng hiệu quả hơn có nghĩa là tìm một chiến lược tốt để thực hiện câu văn tin.
- **Trình tiền biên dịch DML nhúng ( *Embedded DML Precompiler* )**: biến đổi các lệnh DML được nhúng trong một chương trình ứng dụng thành các lời gọi thủ tục chuẩn trong ngôn ngữ chủ. Trình tiền biên dịch phải trao đổi với trình biên dịch DML để sinh mã thích hợp.
- **Bộ thông dịch DDL ( *DDL interpreter* )**: thông dịch các lệnh DDL và ghi chúng vào một tập hợp các bảng chứa metadata.

- **Engine định giá vấn tin ( Query evaluation engine )**: Thực hiện các chỉ thị mức thấp được sinh ra bởi trình biên dịch DML.

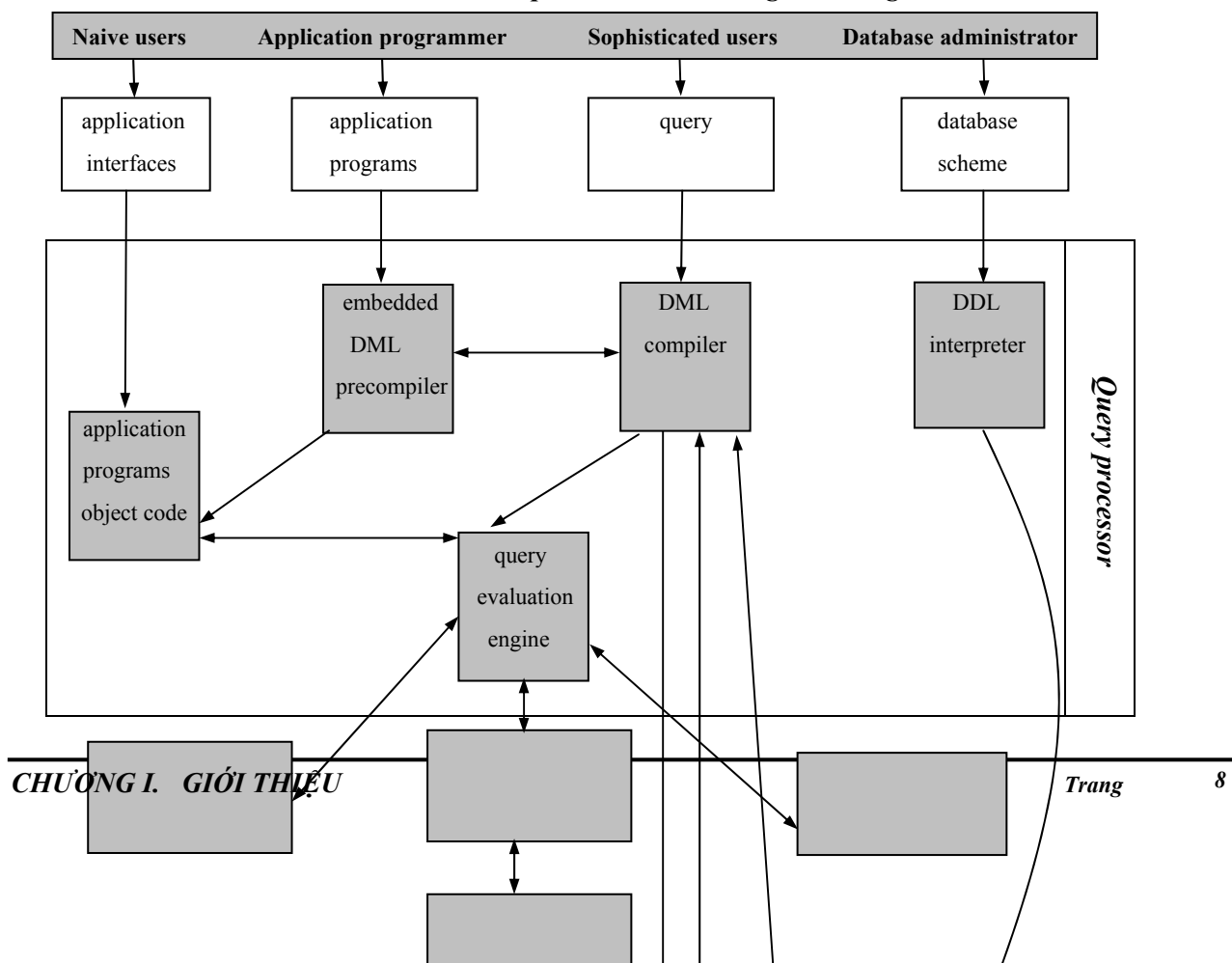
Các thành phần quản trị lưu trữ cung cấp các giao diện giữa dữ liệu mức thấp được lưu trữ trong CSDL và các chương trình ứng dụng, các vấn tin được đệ trình cho hệ thống. Các thành phần quản trị lưu trữ gồm:

- **Bộ quản trị quyền và tính toàn vẹn ( Authorization and integrity manager )**: kiểm tra sự thoả mãn các ràng buộc toàn vẹn và kiểm tra quyền truy xuất dữ liệu của người sử dụng.
- **Bộ quản trị giao dịch ( Transaction manager )**: Đảm bảo rằng CSDL được duy trì trong trạng thái nhất quán cho dù hệ thống có sự cố và đảm bảo rằng các thực hiện giao dịch cạnh tranh tiến triển không xung đột.
- **Bộ quản trị file ( File manager )**: Quản trị cấp phát không gian trên lưu trữ đĩa và các cấu trúc dữ liệu được dùng để biểu diễn thông tin được lưu trữ trên đĩa.
- **Bộ quản trị bộ đệm ( Buffer manager )**: có trách nhiệm đem dữ liệu từ lưu trữ đĩa vào bộ nhớ chính và quyết định dữ liệu nào trữ trong bộ nhớ.

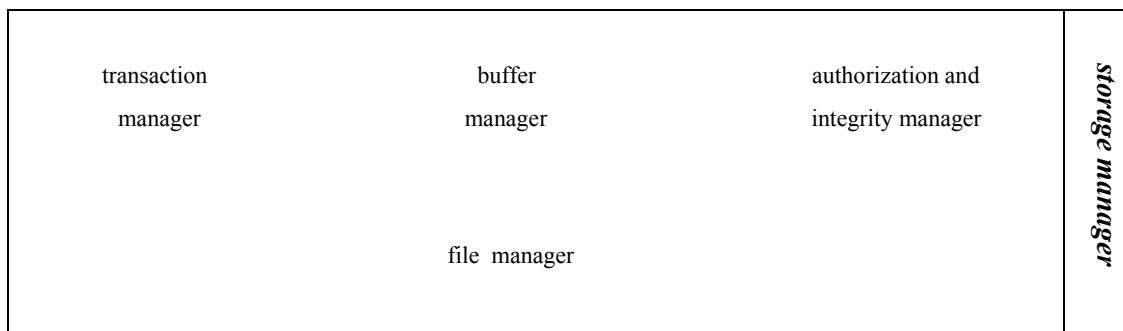
Hơn nữa, một số cấu trúc dữ liệu được cần đến như bộ phận của sự thực thi hệ thống vật lý:

- **Các file dữ liệu:** Lưu trữ CSDL
- **Tự điển dữ liệu ( Data Dictionary ):** lưu metadata về cấu trúc CSDL.
- **Chỉ mục ( Indices ):** cung cấp truy xuất nhanh đến các hạng mục dữ liệu chứa các giá trị tìm kiếm.
- **Dữ liệu thống kê ( Statistical data ):** lưu trữ thông tin thống kê về dữ liệu trong cơ sở dữ liệu. Thông tin này được dùng bởi bộ xử lý văn tin để chọn những phương pháp hiệu quả thực hiện câu văn tin.

### Sơ đồ các thành phần và các nối kết giữa chúng







**Figure 2**

## **KIẾN TRÚC HỆ CƠ SỞ DỮ LIỆU**

Kiến trúc hệ CSDL bị ảnh hưởng nhiều bởi hệ thống máy nền. Các sắc thái của kiến trúc máy như mạng, song song và phân tán được phản ánh trong kiến trúc của hệ CSDL.

- Mạng máy tính cho phép thực hiện một số công việc trên một hệ thống các server, một số công việc trên các hệ thống client. Việc phân chia công việc này dẫn đến sự phát triển hệ CSDL *client-server*.
- Xử lý song song trong một hệ thống máy tính làm tăng tốc độ các hoạt động của hệ CSDL, trả lời các giao dịch nhanh hơn. Các vấn tin được xử lý theo cách khai thác tính song song. Sự cần thiết xử lý vấn tin song song này dẫn tới sự phát triển của hệ CSDL *song song*.
- Dữ liệu phân tán trên các site hoặc trên các bộ phận trong một cơ quan cho phép các dữ liệu thường trú tại nơi chúng được sinh ra nhưng vẫn có thể truy xuất chúng từ các site khác hay các bộ phận khác. Việc lưu nhiều bản sao của CSDL trên các site khác nhau cho phép các tổ chức lớn vẫn có thể tiếp tục hoạt động khi một hay một vài site bị sự cố. Hệ CSDL phân tán được phát triển để quản lý dữ liệu phân tán, trên phương diện địa lý hay quản trị, trải rộng trên nhiều hệ CSDL .

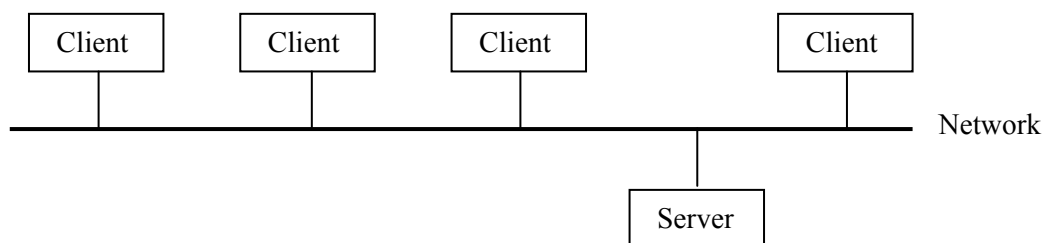
## **HỆ THỐNG TẬP TRUNG**

Các hệ CSDL tập trung chạy trên máy đơn và không trao đổi với các máy khác. Các hệ thống như vậy trải từ các hệ CSDL một người sử dụng chạy trên các máy cá nhân (PC) đến các hệ CSDL hiệu năng cao chạy trên các hệ mainframe. Một hệ máy tính mục đích chung hiện đại gồm một hoặc một vài CPU và một số bộ điều khiển thiết bị được nối với nhau thông qua một bus

chung, cho phép truy xuất đến bộ nhớ chia sẻ. CPU có bộ nhớ cache cục bộ lưu các bản sao của một số phần của bộ nhớ chính nhằm tăng tốc độ truy xuất dữ liệu. Mỗi bộ điều khiển thiết bị phụ trách một kiểu thiết bị xác định. Các CPU và các bộ điều khiển thiết bị có thể thực hiện đồng thời, cạnh tranh truy cập bộ nhớ. Bộ nhớ cache giúp làm giảm sự tranh chấp truy xuất bộ nhớ. Ta phân biệt hai cách các máy tính được sử dụng: Hệ thống một người dùng và hệ thống nhiều người dùng. Hệ CSDL được thiết kế cho hệ thống một người dùng không hỗ trợ điều khiển cạnh tranh, chức năng phục hồi hoặc là thiếu hoặc chỉ là một sự chập dờn đơn giản.

## **HỆ THỐNG CLIENT-SERVER**

Các máy tính cá nhân ( PC ) ngày càng trở nên mạnh hơn, nhanh hơn, và rẻ hơn. Có sự chuyển dịch trong hệ thống tập trung. Các đầu cuối (terminal) được nối với hệ thống tập trung bây giờ được thế chỗ bởi các máy tính cá nhân. Chức năng giao diện người dùng (user interface) thường được quản lý trực tiếp bởi các hệ thống tập trung nay được quản lý bởi các máy tính cá nhân. Như vậy, các hệ thống tập trung ngày nay hoạt động như các hệ thống server nó làm thỏa mãn các đòi hỏi của các client. Chức năng CSDL có thể được chia thành hai phần: phần trước (front-end) và phần sau (back-end). Phần sau quản trị truy xuất cấu trúc, định giá câu vấn tin và tối ưu hoá, điều khiển sự xảy ra đồng thời và phục hồi. Phần trước của hệ CSDL gồm các công cụ như: tạo mẫu (form), các bộ soạn báo cáo (report writer), giao diện đồ họa người dùng (graphical user interface). Giao diện giữa phần trước và phần sau thông qua SQL hoặc một chương trình ứng dụng. Các hệ thống server có thể được phân thành các phạm trù : server giao dịch (transaction server), server dữ liệu (data server).



**Figure 3**

- **Hệ thống server giao dịch (transaction-server systems):** còn được gọi là hệ thống server vấn tin (query-server system), cung cấp một giao diện mà các client có thể gửi đến nó các yêu cầu thực hiện một hành động. Để đáp ứng các yêu cầu, hệ thống thực hiện các hành động và gửi lại client các kết quả. Các người sử dụng có thể đặc tả các yêu cầu trong SQL hoặc trong một giao diện trình ứng dụng sử dụng một cơ chế gọi thủ tục xa ( remote-procedure-call ).
  - **Các servers giao dịch ( Transaction servers ):** Trong các hệ thống tập trung, phần trước (front-end) và phần sau (back-end) được thực hiện trong một hệ thống. Kiến trúc server giao dịch cho phép chia chức năng giữa phần trước và phần sau. Chức năng phần trước được hỗ trợ trên các máy tính cá nhân (PC). Các PC hành động như những khách hàng của các hệ thống server nơi lưu trữ một khối lượng lớn dữ liệu và hỗ trợ các chức năng phần sau. Các clients gửi các giao dịch đến các hệ thống server tại đó các giao dịch được thực hiện và