Họ và tên: Lê Thái Giao Tiên TÓM TẮT LÝ THUYẾT – HK1, 2022-2023

MSSV: B1906586 Môn: QUẢN TRỊ CƠ SỞ DỮ LIỆU

Nhóm: CT205 nhóm 1 Ngày nộp 10 tháng 10 năm 2022

TÓM TẮT LÝ THUYẾT

# HQTCSDL là gì? Các HQTCSDL hiện nay

* ***HQTCSDL*** là hệ thống phần mềm cho phép người dùng định nghĩa, tạo và duy trì CSDL, đồng thời quản lý việc cung cấp dịch vụ truy cập đến CSDL này.

*(Chương 1 – Trang 2)*

* HQTCSDL phổ biến hiện nay: Oracle, DB2, MS SQL Server, MS Access và các phần mềm nguồn mở MySQL, PostGreSQL,…

(Chương 1 – Trang 2)

# Khác nhau giữa HQTCSDL và bảng tính (vd, Excel) là gì?

* ***HQTCSDL:*** Là một tập hợp các chương trình để truy cập dữ liệu cung cấp thông tin về một tổ chức cụ thể. Thực hiện các hoạt động như: tạo, chèn, xóa, cập nhật và truy xuất thông tin. Nó là một hệ thống phần mềm mục đích chung tạo điều kiện thuận lợi cho quá trình xác định, xây dựng và thao tác cơ sở dữ liệu cho các ứng dụng khác nhau và liên quan đến dữ liệu.
* ***Bảng tính:*** Là một tệp có thể giúp tính toán và sắp xếp dữ liệu. Bảng tính có thể có một giá trị số, văn bản, công thức và hàm. Nó có các cột và hàng để giữ cho thông tin được chèn dễ đọc và dễ hiểu và là một bảng đồ thị điện tử.

|  |  |
| --- | --- |
| **HQTCSDL** | **Bảng tính** |
| Một tập hợp dữ liệu liên quan đến nhau | Một bảng biểu đồ điện tử với các hàng và cột |
| Lưu trữ và truy xuất dữ liệu | Lưu trữ và phân tích dữ liệu |
| Hỗ trợ cho việc tổ chức và lựa chọn dữ liệu | Hỗ trợ cho các tính toán phức tạp |
| Công suất chỉ bị giới hạn bởi máy tính | Chứa lượng thông tin hạn chế |
| Hướng tới việc liên kết và tổ chức dữ liệu nhiều hơn | Hướng tới tính toán nhiều hơn |
| Có thể thực hiện các tính toán đơn giản | Có thể thực hiện tổ chức và lựa chọn dữ liệu đơn giản |
| Xử lý dữ liệu hiệu quả | Xử lý dữ liệu không hiệu quả |
| Ví dụ: Oracle, MySQL, MS Access | Ví dụ: Excel, Lotus 1-2-3 |

*(https://www.geeksforgeeks.org/difference-between-dbms-and-spreadsheet/)*

# DBA là ai? Nhiệm vụ?

DBA là ***nhà quản trị CSDL*** *(Database Administrator)*

Nhiệm vụ:

* Đảm bảo cho sự hoạt động thông suốt của các chức năng, tính hiệu quả và ứng dụng truy cập vào CSDL của tổ chức
* Khi cần phát triển một ứng dụng mới:
  + DBA tham gia vào tất cả giai đoạn của đề án nhằm đảm bảo ứng dụng có sự truy cập chính xác, hiệu quả đến dữ liệu của tổ chức.
* Khi ứng dụng chuẩn bị đưa vào sử dụng:
  + DBA đảm bảo HQTCSDL đã được chuẩn bị sẵn sàng cho tải mới và cũng có trách nhiệm đưa CSDL mới từ môi trường thử nghiệm vào hoạt động.
* Khi ứng dụng đã được đưa vào sử dụng:
  + DBA đảm bảo tính sẵn sàng, an ninh, toàn vẹn của hệ thống (cài đặt các ràng buộc trên CSDL); giám sát hiệu suất, điều chỉnh, sao lưu và phục hồi hệ thống.
* Khi ứng dụng không còn được sử dụng:
  + DBA xác định tình trạng cuối cùng của dữ liệu sử dụng bởi ứng dụng để có thể cần cho ứng dụng nào khác, hay cần lưu trữ theo quy định nào đó,…

*(Chương 1 – Trang 7)*

# Các phương pháp bảo vệ dữ liệu

* Phương pháp không sử dụng máy tính:
  + Bố trí thiết bị một cách an toàn
  + Quản trị nhân sự
  + Chính sách an ninh và kế hoạch đối phó với những bất ngờ
* Phương pháp có sử dụng máy tính:
  + Cấp quyền (authorization)
  + Khung nhìn (Views)
  + Sao lưu và phục hồi (Backup and restore)
  + Toàn vẹn dữ liệu (Integrity)
  + Mật hóa (Encryption)

*(Chương 2 – Trang 9)*

# Giao dịch là gì? Tại sao cần phải giao dịch?

* ***Giao dịch*** là một hành động hay một chuỗi các hành động được thực hiện bởi 1 người dùng hoặc 1 chương trình ứng dụng, trong đó có truy cập hoặc thay đổi nội dung của một CSDL. Nó có thể là toàn bộ chương trình, một phần của chương trình, hoặc một lệnh đơn lẻ như INSERT hay UPDATE, và nó có thể bao gồm nhiều thao tác trên CSDL. *(Chương 3 – Trang 18)*
* ***Cần phải giao dịch vì:*** Lợi ích chính của việc sử dụng các giao dịch là tính toàn vẹn dữ liệu. Nhiều cách sử dụng cơ sở dữ liệu yêu cầu lưu trữ dữ liệu cho nhiều bảng hoặc nhiều hàng trong cùng một bảng để duy trì một tập dữ liệu nhất quán. Sử dụng giao dịch đảm bảo rằng các kết nối khác với cùng một cơ sở dữ liệu sẽ thấy tất cả các bản cập nhật hoặc không có bản cập nhật nào trong số chúng.

*(https://www.justsoftwaresolutions.co.uk/database/database\_tip\_use\_transactions.html)*

# Các trạng thái của giao dịch

* ***Hoạt động* (Active):** Trạng thái khởi đầu; giao dịch giữ trong trạng thái này trong khi nó đang thực hiện.
* ***Hoàn tất một phần*** *(Partially Committed):* Sau khi lệnh cuối cùng được thực hiện.
* ***Thất bại*** *(Failed):* Sau khi phát hiện rằng sự thực hiện không thể tiếp tục được nữa.
* ***Hủy bỏ*** *(Aborted):* Sau khi giao dịch bị cuộn lại (rolled back) và CSDL đã phục hồi lại trạng thái của nó trước khi khởi động giao dịch.
* ***Hoàn tất*** *(Committed):* Sau khi hoàn thành thành công giao dịch CSDL đạt tới trạng thái nhất quán mới.

*(Chương 3 – Trang 19)*

# Thuộc tính ACID là gì?

Thuộc tính ACID là viết tắt 4 thuộc tính của các giao dịch đều phải có, bao gồm:

* ***Tính nguyên tử*** *(Atomicity):* hay tính ‘Tất cả hoặc không có gì’: Một giao dịch hoặc được thực hiện toàn bộ hoặc không thực hiện gì.
* ***Tính nhất quán*** *(Consistency):* Một giao dịch phải chuyển CSDL từ một trạng thái nhất quán này sang một trạng thái nhất quán khác. Tính nhất quán của CSDL yêu cầu chỉ những dữ liệu không vi phạm ràng buộc dữ liệu mới được ghi vào CSDL.
* ***Tính cô lập*** *(Isolation):* Các kết quả giữa chừng của một giao dịch chưa hoàn tất thì không để các giao dịch khác nhìn thấy.
* ***Tính bền vững*** *(Duration):* Mỗi một giao dịch hoàn tất, tất cả các cập nhật đã thực hiện trên CSDL vẫn còn đó, ngay cả khi sự cố hệ thống sau khi giao dịch đã hoàn tất.

*(Chương 3 – Trang 20)*

# Các vấn đề cạnh tranh trong môi trường đa người dùng là gì? Giải pháp khắc phục?

* ***Các vấn đề cạnh tranh*** trong môi trường đa người dùng:
  + Mất dữ liệu đã cập nhật (lost update)
  + Sự phụ thuộc vào các giao dịch không hoàn tất (uncommitted dependency)
  + Phân tích không nhất quán (inconsistent analysis)
* ***Khắc phục*** bằng điều khiển cạnh tranh (Concurrency control): Quản lý cạnh tranh là quá trình quản lý các thao tác đang diễn ra đồng thời trên CSDL mà không cho phép chúng can thiệp lẫn nhau.

*(Chương 3 – Trang 21)*

# Lịch trình là gì? Lịch trình tuần tự và không tuần tự

* ***Lịch trình*** một chuỗi các thao tác thực hiện bởi một tập hợp các GD cạnh tranh mà vẫn đảm bảo thứ tự của các thao tác trong từng GD đơn lẻ.
* ***Lịch trình tuần tự*** *(serial schedule)* là một lịch trình trong đó các thao tác của một GD được thực hiện liên tiếp nhau, không có bất kỳ thao tác nào của các GD khác xen vào giữa.
* ***Lịch trình không tuần tự*** *(nonserial schedule)* là một lịch trình trong đó các thao tác từ một tập hợp các giao dịch cạnh tranh đan xen lẫn nhau. Khi đó, các lịch trình này có thể đưa CSDL về tình trạng không nhất quán hay các kết quả không chính xác như phân tích trong các phần kế tiếp.

*(Chương 3 – Trang 23)*

# Các kỹ thuật quản lý cạnh tranh

## Bi quan (lock)

Cho phép các GD thực hiện song song một cách an toàn dựa trên một số ràng buộc nào đó.

* 2 phương pháp: Khóa chốt và nhãn thời gian làm cho các GD bị ngưng trệ khi chúng xung đột với các giao dịch khác tại một thời điểm nào đó trong tương lai.
* ***Khóa chốt (locking):*** là một thủ tục sử dụng để điều khiển truy cập cạnh tranh đến dữ liệu. Khi một GD đang truy cập vào CSDL, một khóa chốt có thể từ chối truy cập đến từ các CD khác để tránh kết quả không đúng.
  + *Khóa đọc (read lock):* Nếu một GD có một khóa đọc trên một mục dữ liệu, nó có thể đọc nhưng không thể cập nhật mục dữ liệu đó.
  + *Khóa ghi (write lock):* Nếu một GD có một khóa ghi trên một mục dữ liệu, nó có thể đọc và cập nhật mục dữ liệu đó.
* Bởi vì các thao tác đọc không thể xung đột, nên có thể có nhiều GD giữ khóa đọc một cách đồng thời trên cùng một mục dữ liệu. Ngược lại, một khóa ghi cho phép một GD độc quyền truy cập trên mục dữ liệu đó. Vì vậy, ngay khi một GD giữ một khóa ghi trên mục dữ liệu, thì không một GD nào khác có thể đọc hoặc cập nhật mục dữ liệu đó.

*(Chương 3 – Trang 27)*

## Lạc quan: 3 kỳ (đọc, kiểm tra, ghi)

Các ***kỹ thuật lạc quan (optimistic techniques)*** dựa trên giả thiết rằng xung đột thì hiếm, và sẽ hiệu quả hơn nếu cho phép các GD được tiến hành mà không cần bắt chúng phải trì hoãn để đảm bảo tính khả tuần tự (Kung & Robinson, 1981).

Khi một giao dịch muốn hoàn tất, hệ thống mới kiểm tra xem liệu có xung đột đã xảy ra hay không. Nếu có xung đột, thì GD phải bị hủy và khởi động lại. Vì giả thiết là xung đột hiếm xảy ra, nên việc cuộn lại cũng sẽ hiếm.

Giao thức điều khiển cạnh tranh lạc quan: Một GD sẽ trải qua ba kỳ nếu là GD cập nhật, và trải qua 2 kỳ nếu là GD chỉ đọc:

* ***Kỳ đọc:*** kỳ này kéo dài từ lúc bắt đầu GD cho đến ngay trước hành động commit. GD đọc các giá trị của tất cả các mục dữ liệu nó cần từ CSDL và lưu chúng vào các biến cục bộ. Các cập nhật sẽ được áp trên bản sao chép cục bộ của dữ liệu, không phải trên CSDL.
* ***Kỳ kiểm tra:*** Kỳ này theo sau kỳ đọc. Các kiểm tra sẽ được thực hiện để đảm bảo tính khả tuần tự không bị vi phạm nếu các cập nhật của GD được đưa vào CSDL.

Đối với các GD chỉ đọc, điều này bao gồm việc kiểm tra xem các giá trị dữ liệu đã đọc vào các biến vẫn còn là các giá trị hiện hành trong các mục dữ liệu tương ứng. Nếu không có sự can thiệp nào đã diễn ra, thì GD được hoàn tất. Nếu không, GD phải bị hủy và khởi động lại.

Đối với một GD có cập nhật, việc kiểm tra bao gồm việc xác định liệu GD đó có đưa CSDL về một tình trạng nhất quán với tính khả tuần tự được duy trì hay không. Nếu không thì GD phải bị hủy.

* ***Kỳ ghi:*** Kỳ này theo sau kỳ kiểm tra thành công đối với các giao dịch cập nhật. Trong kỳ này, các cập nhật đã thực hiện trên biến cục bộ sẽ được chép vào CSDL.

*(Chương 3 – Trang 35)*

## Độ mịn của mục dữ liệu

***Độ mịn*** là kích cỡ của mục dữ liệu được chọn như là một đơn vị bảo vệ bởi giao thức điều khiển cạnh tranh. Một mục dữ liệu điển hình được chọn như là một trong những thứ có kích cỡ từ lớn đến nhỏ sau đây:

* Toàn bộ CSDL
* Một tập tin (file)
* Một trang tin (page), đôi khi còn gọi là một vùng hay một không gian CSDL – một đoạn đĩa vật l ý nơi lưu trữ các quan hệ
* Một mẫu tin (record)
* Một giá trị của một trường trong một mẫu tin.

*(Chương 3 – Trang 37)*

# Phục hồi dữ liệu

## Tại sao cần phục hồi dữ liệu

Vì có rất nhiều nguyên nhân gây lỗi như hệ thống bị ngưng lại, lỗi thiết bị lưu trữ, lỗi phần mềm ứng dụng, thiên tai, bất cẩn hay phá hoại dữ liệu có chủ ý nên cần giảm thiểu các ảnh hưởng và phục hồi CSDL về trạng thái đúng sau sự cố có lỗi xảy ra.

*(Chương 3 – Trang 39)*

## Điểm kiểm tra là gì (check point)

***Điểm kiểm tra (checkpoint)*** là một điểm mà tại đó sự đồng bộ giữa CSDL và tập tin nhật ký GD được ghi nhận. Khi đó, tất cả các vùng đệm phải được ghi-ép-buộc ra bộ lưu trữ thứ cấp.

*(Chương 3 – Trang 42)*

## Các kỹ thuật phục hồi dữ liệu

* *Nếu CSDL bị hư,* không đọc được nữa, thì ta cần sử dụng bản sao lưu cuối cùng và chạy lại các thao tác của các GD hoàn tất đã ghi nhận trong tập tin nhật ký.
* *Nếu CSDL không bị hư vật lý* nhưng trở nên không nhất quán, chẳng hạn như khi hệ thống bị treo khi đang thực hiện giao dịch, thì ta cần undo các thay đổi gây ra tình trạng không nhất quán. Cũng có thể phải redo một số giao dịch để đảm bảo các cập nhật của nó được lưu ra bộ lưu trữ thứ cấp.
* Ba kỹ thuật để phục hồi cho trường hợp này là cập nhật trì hoãn (deferred update), cập nhật tức thì (immediate update) và tạo trang bóng (shadow paging).
  + ***Cập nhật trì hoãn:*** Ý tưởng chủ yếu là trì hoãn các cập nhật thực sự lên CSDL cho đến khi GD kết thúc thành công và đạt đến điểm hoàn tất.
  + ***Cập nhật tức thì:*** CSDL có thể được cập nhật ngay bởi các thao tác của một GD, trước khi GD tiến đến điểm hoàn tất.
  + ***Tạo trang bóng:*** Cơ chế này duy trì các bảng hai-trang trong suốt quá trình thực hiện của GD: một bảng trang hiện tại và một bảng trang bóng. Khi một GD bắt đầu, các bảng này giống nhau. Bảng trang bóng không bao giờ thay đổi, vì vậy được dùng để phục hồi CSDL khi có sự cố xảy ra. Trong suốt GD, bảng trang hiện tại được dùng để ghi nhận mọi thay đổi trên dữ liệu. Khi GD hoàn tất, bảng trang hiện tại trở thành bảng trang bóng.

*(Chương 3 – Trang 42)*