MỤC LỤC

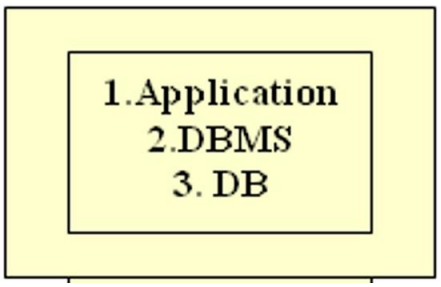
CHƯƠNG 1: TỔNG QUAN VỀ AN TOÀN THÔNG TIN TRONG CSDL

Câu 1: Các mối đe dọa, các tấn công có thể đến với CSDL là gì?

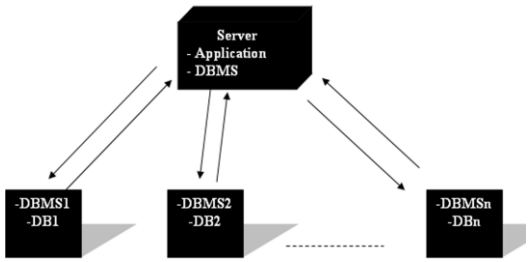
- Các mối đe dọa có thể đến với CSDL là những hiểm họa có thể được xác định khi đối phương sử dụng các kỹ thuật đặc biệt để tiếp cận nhằm khám phá, sửa đổi trái phép thông tin quan trọng do hệ thống quản lý.
  - Những mối đe dọa có thể xuất phát từ nhiều nguyên nhân: ngẫu nhiên hay có chủ ý.
  - Xâm phạm: đọc, sửa, xóa dữ liệu trái phép
    - Khai thác trái phép thông qua suy diễn thông tin được phép
    - Sửa đổi dữ liệu trái phép
    - Từ chối dịch vụ hợp phép (DoS)
  - Hiểm họa ngẫu nhiên:
    - Các thảm họa trong thiên nhiên.
    - Các lỗi phần cứng hay phần mềm có thể dẫn đến việc áp dụng các chính sách an toàn thông tin không đúng.
    - Các sai phạm vô ý do con người gây ra.
  - Hiểm họa có chủ ý:
    - Người dùng hợp pháp: lạm quyền, sử dụng vượt mức quyền hạn cho phép.
    - Người dùng truy nhập thông tin trái phép, có thể là người ngoài tổ chức hoặc bên trong tổ chức: tấn công, phá hoại, leo thang đặc quyền.

Câu 2: Tìm hiểu các cấu hình xử lý CSDL (CSDL tập trung, phân tán, Client/Server). Các cấu hình này được áp dụng như thế nào trong thực tế. (Chú ý: nêu rõ đặc điểm – bản chất và vẽ hình minh họa).

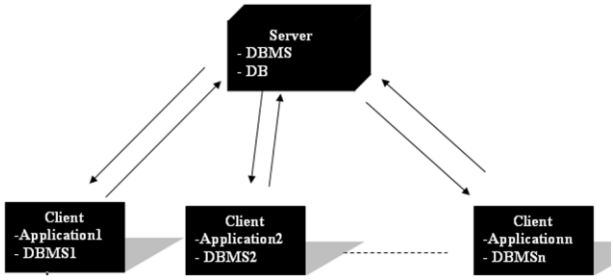
- Có 3 thành phần trong mô hình xử lý CSDL: ứng dụng, hệ quản trị CSDL (DBMS), cơ sở dữ liệu.
- Có 3 mô hình:
  - Mô hình CSDL tập trung
  - Mô hình CSDL phân tán
  - Mô hình CSDL Client/Server
- Mô hình CSDL tập trung: cả 3 thành phần là ứng dụng, DBMS, DB cùng nằm chung trên 1 máy.
  - Ưu điểm: xử lý nhanh, dễ dàng, không tốn chi phí.
  - Nhược điểm: không lưu trữ nhiều; nếu một máy hỏng thì mất toàn bộ dữ liệu; không lưu trữ khi có nhiều dự án.
  - Áp dụng cho các cá nhân, tổ chức nhỏ, ít dữ liệu.
  - Mô hình:



- Mô hình CSDL phân tán: Bao gồm 1 server được kế nối tới nhiều máy tính. Mỗi máy tính chứa DB riêng và có một DBMS. Khi máy chủ cần truy xuất dữ liệu, nó sẽ gọi đến các máy tính con, yêu cầu cần quá trình đồng bộ dữ liệu.
    - Ưu điểm: lưu trữ nhiều; 1 máy hỏng thì dữ liệu không bị mất; lưu trữ khi đang ở nhiều địa điểm khác nhau.
    - Nhược điểm: chi phí cao, khó quản lý.
    - Mô hình được áp dụng phù hợp cho các công ty có nhiều chi nhánh khác nhau.
    - Mô hình:



- Mô hình CSDL client/server: bao gồm một máy chủ server và các máy trạm. Dữ liệu và hệ quản trị cơ sở dữ liệu nằm trên máy chủ, các thành phần ứng dụng làm trên các client.
    - Ưu điểm: xử lý nhanh, có thể truy xuất dữ liệu từ xa.
    - Nhược điểm: nếu server sập thì toàn bộ dữ liệu sẽ bị mất.
    - Mô hình:



Câu 3: Trình bày về SQL, các câu lệnh SQL cơ bản.

- Mỗi DBMS đều phải có ngôn ngữ giao tiếp giữa người sử dụng với cơ sở dữ liệu. SQL (Structure Query Language) là ngôn ngữ hỏi đáp có cấu trúc cho phép người sử dụng khai thác CSDL để truy vấn các thông tin cần thiết trong CSDL.
- SQL cho phép truy cập CSDL và thực hiện các thao tác như: Lấy dữ liệu, chèn, xóa, sửa,…
    - Có thể thực thi các câu lệnh SQL trên CSDL như: Select, Insert, Update, Delete…
    - SQL hoạt động hầu hết với các chương trình CSDL như MS Acsess, Oracle, SQL Server, My SQL
    - SQL có các ngôn ngữ con sau:
        - DDL (Data Definition Language) là ngôn ngữ máy tính để định nghĩa lược đồ CSDL logic (creat, alter, drop table (index- chỉ mục)).
          Ví dụ lệnh Create sau được viết trong Oracle sẽ tạo ra 1 bảng có tên SINHVIEN:
          CREATE TABLE SINHVIEN(MaSV INT primary key, HoTen varchar(100), GioiTinh varchar(10));

- DML (Data Manipulation Language) là họ các ngôn ngữ máy tính được sử dụng để tìm kiếm, chèn, xóa và cập nhập dữ liệu trong một CSDL (SELECT, INSERT, UPDATE, DELETE).

  Ví dụ một số câu lệnh DML được viết trong Oracle:

  ```
  SELECT MaSV FROM SINHVIEN WHERE (MaSV = 1);
  INSERT INTO SINHVIEN VALUES(1,'Son','Nam');
  UPDATE SINHVIEN SET HoTen = 'THAI' WHERE MaSV = 1;
  DELETE FROM SINHVIEN WHERE MaSV = 1;
  ```

- DCL (Data Control Language) là ngôn ngữ điều khiển dữ liệu, sử dụng hai từ khóa là Grant, Revoke thực hiện việc gán và thu hồi quyền.
- QL (Query Language) là ngôn ngữ khai báo hỗ trợ cho những người dùng cuối, giúp người dùng tìm kiếm dữ liệu trong cơ sở dữ liệu. Ví dụ về QL như là câu lệnh lựa chọn trong SQL là Select.

- Các câu lệnh trong SQL:

Create table	Create index	Insert	Alter table	Delete
Drop table	Drop index	Update	Select	

Câu 4: Các yêu cầu bảo vệ CSDL

Có 10 yêu cầu bảo vệ CSDL:

- Bảo vệ chống truy nhập trái phép:
  - Chỉ trao quyền cho những người dùng hợp pháp.
  - Việc kiểm soát truy nhập cần tiến hành trên các đối tượng dữ liệu ở mức thấp hơn file: bản ghi, thuộc tính, giá trị.
- Bảo vệ chống suy diễn: Suy diễn là khả năng có được các thông tin bí mật từ các thông tin không bí mật (công khai).
- Bảo vệ toàn vẹn CSDL:
  - Bảo vệ cơ sở dữ liệu khỏ những người dùng không hợp pháp, tránh sửa đổi nội dụng dữ liệu trái phép.
  - Hệ quản trị cơ sở dữ liệu kiểm soát bằng các ràng buộc dữ liệu, thủ tục sao lưu, phục hồi, các thủ tục an toàn đặc biệt, file nhật ký.
  - Một số phương pháp bảo vệ toàn vẹn dữ liệu như: kiểu dữ liệu, không cho phép định nghĩa null, định nghĩa mặc định, các thuộc tính định danh, các ràng buộc, các quy tắc, triggers, các chỉ mục.
- Toàn vẹn dữ liệu thao tác:
  - Yêu cầu đảm bảo tính tương thích logic của dữ liệu khi có nhiều giao tác thực hiện đồng thời.
  - Một giao dịch (transaction): là một loạt các hoạt động xảy ra được xem như 1 đơn vị công việc nghĩa là hoặc thành công toàn bộ hoặc không làm gì cả.
- Toàn vẹn ngữ nghĩa của dữ liệu:
  - Yêu cầu này đảm bảo tính tương thích logic của các dữ liệu thay đổi, bằng cách kiểm tra các giá trị dữ liệu có nằm trong khoảng cho phép hay không (đó là các ràng buộc toàn vẹn).
  - Ràng buộc là những thuộc tính mà ta áp đặt lên một bảng hay một cột để tránh việc lưu dữ liệu không chính xác vào CSDL.

    - Các ràng buộc bao gồm: ràng buộc khóa chính, ràng buộc khóa ngoại, ràng buộc kiểm tra.
- Khả năng lưu vết và kiểm tra:
    - Bao gồm khả năng ghi lại mọi truy nhập tới dữ liệu (các phép toàn read, write).
    - Đảo bảo tính toàn vẹn dữ liệu vật lý, trợ giúp cho việc phân tích dãy truy nhập vào CSDL.
- Xác thực người dùng:
    - Xác định tính duy nhất của người dùng. Dịnh danh người dùng làm cơ sở cho việc trao quyền. Người dùng được phép truy nhập dữu liệu khi được hệ thống xác thực là hợp pháp.
- Bảo vệ dữ liệu nhạy cảm:
    - Dữ liệu nhạy cảm là những dữ liệu không nên đưa ra công bố công khai.
    - Kiểm soát truy nhập vào các CSDL bao hàm: bảo vệ tính tin cậy của dữ liệu nhạy cảm, chỉ cho phép người dùng hợp pháp truy nhập. Người dùng được trao quyền trên các dữ liệu này không được phép lan truyền chúng.
- Bảo vệ nhiều mức:
    - Dữ liệu được phân loại thành nhiều mức nhạy cảm.
    - Mục đích: phân loại các mục thống tin khác nhau, đồng thời phân quyền cho các mức truy nhập khác nhau vào các mục riêng biệt.
- Sự hạn chế:
    - Tránh chuyển các thông tin không mong muốn giữa các chương trình trong hệ thống.
    - Kênh được cho phép cung cấp thông tin qua các hoạt động được phép: soạn thảo, biên dịch 1 file.

Câu 5: Trình bày một số phương pháp (được tích hợp sẵn trong các DBMS) để đảm bảo tính toàn vẹn dữ liệu.

Một số phương pháp đảm bảo tính toàn vẹn dữ liệu tích hợp trong các DBMS là:

- Kiểu dữ liệu.
- Không cho phép định nghĩa NULL.
- Định nghĩa mặc định.
- Các thuộc tính định danh.
- Các ràng buộc: ràng buộc khóa chính, ràng buộc khóa ngoại, ràng buộc kiểm tra.
- Các quy tắc.
- Triggers.
- Các chỉ mục.

Câu 6: Tìm hiểu về transaction. Quá trình thực hiện 1 transaction.

- Transaction:
    - Transaction trong SQL là một đơn vị công việc hoặc 1 dãy công việc được thực hiện theo 1 thứ tự logic và hợp lý, có thể được thao tác bởi người dùng hoặc 1 DB program.
    - Transaction là một nhóm các câu lệnh SQL, xử lý có tuần tự các thao tác trên CSDL nhưng được xem như 1 đơn vị. Vì vậy một transaction sẽ không được coi như thành

công nếu như trong quá trình xử lý có 1 thao tác trong nó không hoàn thành => giao dịch thất bại.

- Transaction có 1 chuẩn gọi là ACID bao gồm 4 thuộc tính chuẩn:
  - Atomicity (tính tự trị): đảm bảo tata cả các thao tác, hành động trong phạm vị một đơn vị giao dịch là thành công hoàn toàn. Ngược lại transaction được coi là thất bại.
  - Consistency (tính nhất quán): đảm bảo tất cả các thao tác trên CSDL được thay đổi sau khi giao dịch thành công và không xảy ra lỗi.
  - Isolation (tính cô lập): đảm bảo transaction này hoạt động độc lập với transaction khác.
  - Durability (tính bền vững): đảm bảo kết quả hoặc tác động của transaction vẫn luôn tồn tại, kể cả khi hệ thống xảy ra lỗi.

- Quá trình thực hiện 1 transaction:
  - Bắt đầu transaction với câu lệnh:
    START TRANSACTION;
    BEGIN;
    Lưu ý: nên sử dụng câu lệnh SET autocomit = 0 trước khi bắt đầu transaction vì mặc định autocomit = 1, transaction sẽ tự động hoàn thành mà không phải sử dụng comit hay rollback.
  - Thông báo 1 hay nhiều lênh như SELECT, INSERT, UPDATE, DELETE sau khi bắt đầu transaction.
  - Kiểm tra xem có lỗi nào hay không và mọi thứ có theo như yêu cầu của bạn hay không?
  - Khi 1 transaction hoàn thành thì cần đưa ra câu lệnh COMIT để mọi hành động tác động đến table được thực sự thay đổi.
  - Khi 1 transaction thất bại cần đưa ra câu lệnh ROLLBACK để hủy toàn bộ hành động, phục hồi dữ liệu về trạng thái trước khi bắt đầu transaction.

CHƯƠNG 2. CÁC MÔ HÌNH VÀ CHÍNH SÁCH AN TOÀN

Câu 1: Nêu rõ đặc điểm của kiểm soát truy nhập MAC và DAC trong CSDL, nêu sự khác nhau giữa chúng. Ứng dụng 2 chính sách này trong thực tế các hệ quản trị như thế nào?

- Kiểm soát truy nhập tùy ý DAC:
  - Kiểm soát truy nhập dựa trên định danh của chủ thể hoặc định danh nhóm.

  Chỉ rõ đặc quyền mà mỗi chủ thể có thể có được trên các đối tượng và trên hệ thống (object privilege, system prilvilege).
  - Các yêu cầu truy nhập được kiểm tra, thông qua 1 cơ chế kiểm soát tùy ý, truy nhập chỉ được trao cho các chủ thể thỏa mãn các quy tắc cấp quyền của hệ thống.
  - Được định nghĩa trên 1 tập:
    - Các đối tượng an toàn (security objects)
    - Các chủ thể an toàn (security objects)
    - Các đặc quyền truy nhập (access prilvilege)

    (Quyền truy nhập gồm object privilege, system prilvilege)

- Đặc điểm:
  - Người dùng có thể bảo vệ dữ liệu mà họ sở hữu. Người chủ sở hữu có quyền cao nhất đối với cơ sở dữ liệu mà họ sở hữu.
  - Người chủ sở hữu (owner) có thể gán quyền truy nhập (read, write, execute…) tới các user khác.
  - Việc gán và thu hồi quyền là tùy ý do những người dùng này.
  - Được biểu diễn bởi ma trận truy nhập (ACM).
- Ưu điểm:
  - là kỹ thuật phổ biến, chỉ có một vài vấn đề nghiên cứu mở. Hầu hết các hệ quản trị thương mại đều hỗ trợ nó như Access, Oracle…
  - dễ thực hiện, hệ thống linh hoạt
- Nhược điểm:
  - Khó quản lí việc gán/thu hồi quyền
  - Dễ lộ thông tin
  - Kiểm soát an toàn không tốt

- Kiểm soát truy nhập bắt buộc MAC:
  - Được áp dụng cho các thông tin có yêu cầu bảo vệ nghiêm ngặt.
  - Hoạt động trong hệ thống mà dữ liệu hệ thống và ng dung đc phân loại rõ ràng
  - Hạn chế truy nhập của các chủ thể vào các đối tượng bằng cách sử dụng các nhãn an toàn (label).
  - Cơ chế kiểm soát: dữ liệu được phân loại theo độ mật. Các chủ thể được cấp nhãn truy nhập. Chủ thể chỉ được phép truy nhập đến những dữ liệu có độ mật tương đương với nhãn truy nhập hoặc thấp hơn.
  - Lớp an toàn = (Mức nhạy cảm, vùng ứng dụng)
    - Mức nhạy cảm: thành phần phân cáp
    - Vùng ứng dụng: tp không phân cấp
    - trong quân sự: có 4 mức nhạy cảm: U, C, S, TS
    - trong thương mại: có 3 mức: U, S, HS
    - trong Oracle: mỗi lớp an toàn xác định = 1 nhãn.

      Label = (Level, Compartment, Group)

      Level: tp bắt buộc: phân cấp, thể hiện mức nhạy cảm

      Compartment (tùy chọn): không phân cấp, dùng phân loại dl

      Group (tùy chọn): phân cấp, phân loại ng dùng
  - Ưu điểm: độ an toàn cao vì sử dụng các nhãn an toàn, phù hợp với các môi trường đòi hỏi độ an toàn cao như quốc phòng, quân sự. Khác phục được hạn chế của DAC trong vấn đề trao quyền.
  - Nhược điểm:

- phức tạp.
- Thiếu kỹ thuật gắn nhãn an toàn tự động.
- Không giải quyết được hoàn toàn tấn công trojan horse.
- Giảm tính linh hoạt hệ thống
- Ng dung k đc phép thay đổi quyền

- Sự khác nhau giữa MAC và DAC:

MAC	DAC
Kiểm soát quyền dựa vào các nhãn an toàn gắn với chủ thể và đối tượng.	Kiểm soát quyền dựa trên quyền sở hữu đối tượng.
Việc gán/thu hồi quyền chỉ do 1 nhân viên an toàn	Việc gán/thu hồi quyền là tùy ý với những chủ thể có đặc quyền.
User không thể thay đổi nhãn hay quyền	User có thể thay đổi quyền tùy vào đặc quyền của người dùng.
Dùng cho các hệ thống yêu cầu bảo vệ nghiêm ngặt (quân sự)	Dùng được cho mọi hệ thống.
Độ an toàn cao nhưng phức tạp.	Linh hoạt, độ an toàn không cao.

- Ứng dụng 2 chính sách trong thực tế:
  - DAC: Oracle, DB2, Sybase
  - MAC: Access, SQL, SQL Server, …

Câu 1.1: Trình bày về hệ thống Multilevel security: đ/n, các đặc điểm chính, ví dụ

- Hệ thống Multi-level Security (MLS) là hệ thống an toàn nhiều mức, mỗi chủ thể và đối tượng trong đó đều được gắn nhãn an toàn thể hiện mức độ nhạy cảm của các chủ thể và các đối tượng đó.
- Mục đích: là phân loại các mục thông tin khác nhau, đồng thời phân quyền cho các mức truy nhập khác nha và các mục riêng biệt. Đảm bảo tính bí mật. Thường được áp dụng cho lĩnh vực quân sự.
- CSDL đa mức: Là CSDL mà người dùng và dữ liệu được phân thành các mức an toàn khác nhau (chẳng hạn như không phân lớp – U, mật – C, tuyệt mật – S, tối mật – TS).

Bảo vệ nhiều mức: Ví dụ

User	C_{user}	Dept	C_{dept}	Salary	C_{salary}	TC
Bob	S	Math	S	10K	S	S
Ann	S	CIS	S	20K	TS	TS
Sam	TS	CIS	TS	30K	TS	TS

Ví dụ trên là gán nhãn theo ô

- Chủ thể khi truy nhập bị giới hạn bởi những điều khiển truy nhập bắt buộc là "not read up, not write down", theo mô hình của Bell – Lapadula.

- Đa thể hiện (polyinstantiation): Là một kỹ thuật trong CSDL cho phép CSDL có thể chứa nhiều thể hiện của cùng một dữ liệu với các mức nhạy cảm khác nhau. Trong các DBMS quan hệ, có thể có nhiều bản ghi khác nhau nhưng có cùng một khóa chính với các mức nhạy cảm khác nhau. Các bản ghi đa thể hiện là các bản ghi với cùng khóa chính nhưng có các lớp user truy nhập khác nhau gắn với các khóa chính đó. Vấn đề đa thể hiện xuất hiện nhằm tránh kênh ngầm (convert channel) – Lampson (1973) đã định nghĩa kênh ngầm như một cách để đi vào luồng thông tin.
- Ưu điểm: Độ an toàn cao vì sử dụng các nhãn an toàn phù hợp với các môi trường đòi hỏi độ an toàn nghiêm ngặt như quân sự, quốc phòng. Khắc phục hạn chế của DAC trong vấn đề trao đổi quyền.
- Nhược điểm: Phức tạp do việc gán nhãn không tốt có thể dẫn đến việc gán nhãn không đầy đủ hoặc không nhất quán. Thiếu kỹ thuật gán nhãn an toàn tự động nên tốn công sức. Vấn đề kênh ngầm – không giải quyết được hoàn toàn tấn công Trojan Horse.
- Chính sách MAC dùng cho Multi-level Security còn DAC không sử dụng

Câu 2: Thế nào là mô hình an toàn? Sự khác nhau giữa mô hình an toàn và chính sách an toàn. Tìm hiểu mô hình an toàn là Bell-Lapadula.

- **Mô hình an toàn:** là một mô hình khái niệm mức cao, độc lập phần mềm và xuất phát từ các đặc tả yêu cầu của tổ chức để mô tả nhu cầu bảo vệ của một hệ thống.
- **Chính sách an toàn:** là những phát biểu mức tổng quát và an toàn thông tin từ phía nhà quản lí
- Sự khác nhau giữa mô hình an toàn và chính sách an toàn:

Mô hình an toàn	Chính sách an toàn
Là mô hình khái niệm mức cao, độc lập phần mềm, xuất phát từ các đặc tả yêu cầu của tổ chức, mô tả nhu cầu bảo vệ của một hệ thống.	Là các quy tắc, hướng dẫn ở mức cao, liên quan đến việc thiết kế và quản lý hệ thống trao quyền, là phát biểu mức tổng quát về ATTT của nhà quản lý.
Quan tâm đến 2 vấn đề: chủ thể và đối tượng.	Quan tâm đến 3 vấn đề: tính bí mật, tính toàn vẹn, tính sẵn sàng của dữ liệu.
Mô hình an toàn được xây dựng đầu tiên trong quá trình thiết kế hệ thống.	Chính sách an toàn được xây dựng từ mô hình an toàn.

- Mô hình Bell Paladula:
  - Xuất hiện năm 1975, do quân đội Mỹ.
  - Phù hợp sử dụng trong các hệ thống của quân đội, chính phủ.
  - Mục đích: đảm bảo tính bí mật.
  - Đây là mô hình chính tắc đầu tiên về điều khiển luồng thông tin.
  - Là mô hình tĩnh: mức an toàn (nhãn an toàn) không thay đổi.
  - Người dùng được phân mức độ an toàn, đối tượng được phân mức độ nhạy cảm.
  - Có hai quy tắc:

    Not Read up: Một chủ thể S được phép truy cập đọc đến đối tượng O khi và chỉ khi Clear(S) >= Class(O).

    => Các chủ thể chỉ được đọc thông tin có mức nhạy cảm ngang hoặc thấp hơn mức an toàn mà nó được gán.

    => Không bị lộ thông tin cho những không được quyền truy xuất đến đối tượng đó.

Not write down: Một chủ thể S được phép ghi lên một đối tượng O khi Clear(S) <= Class(O).
=> Các chủ thể chỉ được ghi dữ liệu lên mức nhạy cảm ngang hoặc cao hơn mức an toàn mà nó được gán
=> tránh người dùng vô tình ghi dữ liệu mức cao xuống mức thấp => làm lộ thông tin.

- Ưu điểm: các nhãn an toàn của các chủ thể và các đối tượng không bao giờ được thay đổi trong suốt thời gian hệ thống hoạt động.
- Nhược điểm:
  - mới chỉ quan tâm đến tính bí mật
  - chưa chỉ ra cách thay đổi quyền truy nhập

Câu 3: Trình bày cơ bản về một số phương pháp có thể bảo vệ CSDL trong hệ quản trị Oracle (chẳng hạn: VPD, mã hóa CSDL, OLS, Kiểm toán…), các phương pháp đảm bảo tính toàn vẹn CSDL trong DBMS.

- Cơ sở dữ liệu riêng ảo (VPD - Virtual Private Database)
  - Là kiểm soát truy nhập mức mịn hay cơ chế an toàn mức hàng, cung cấp tính năng bảo mật mức hàng cho cơ sở dữ liệu.
  - Cung cấp giải pháp bảo mật tới mức mịn trực tiếp trên các table, view, synonym; gán trực tiếp các chính sách bảo mật lên các đối tượng cơ sở dữ liệu, các chính sách tự động được thực hiện mỗi khi người dùng truy nhập dữ liệu đến đối tượng đó.
  - Ưu điểm: chi phí thấp, trong suốt với người dùng, tăng cao cơ hội kinh doanh
- An toàn dựa trên nhãn trong Oracle (OLS)
  - Cho phép bảo vệ dữ liệu của các bảng đến mức hàng, mức bản ghi.
  - Cho phép định nghĩa 1 chính sách an toàn được thực thi bằng cách gắn cho các bản ghi trong bảng bởi các nhãn an toàn, thể hiện quyền mà người dùng có thể đọc hay ghi dữ liệu lên các bản ghi.
  - Các tính năng của OLS:
    - Nhãn người dùng cung cấp thông tin về quyền hạn của người dùng.
    - Nhãn dữ liệu cho thấy độ nhạy cảm của thông tin trong hàng đó.
    - Chính sách đặc quyền của người dùng có thể cho phép bỏ qua 1 số khía cạnh của kiểm soát truy nhập dựa vào nhãn.
    - Tùy chọn thực thi chính sách của 1 bảng xác định các khía cạnh khác nhau về cách điều khiển truy nhập thực thi để đọc, ghi lên bảng đó.
- Cơ chế kiểm toán mịn
  - Cho phép giám sát và ghi lại việc truy nhập dữ liệu dựa trên nội dung của dữ liệu.
  - Cho phép định nghĩa một chính sách kiểm toán trên một bảng và các cột tùy chọn.
  - Cung cấp cơ chế điều khiển tốt hơn và mức chi tiết nhỏ hơn so với phương pháp kiểm toán thông thường như: kiểm toán câu lệnh, kiểm toán đặc quyền, kiểm toán đối tượng lược đồ.
- Oracle Advanced Security
  - Là cơ chế an toàn nâng cao trong Oracle, cho phép kiểm soát phòng ngừa, giúp giải quyết nhiều yêu cầu đặt ra, ngăn chặn các hành vi vi phạm dữ liệu, bảo vệ thông tin.
  - Cung cấp tính riêng tư, tính toàn vẹn, xác thực, cấp quyền truy nhập với nhiều cách thức khác nhau.

- Oracle secure backup (OSB)
  - Cho phép bảo vệ dữ liệu đáng tin cậy thông quâ hệ thống tập tin sao lưu.
  - Hỗ trợ băng từ, ổ địa, môi trường SAN…
  - Là một phần của giải pháp lưu trữ Oracle, giảm tính phức tạp và giảm chi phí cho việc mua phần mềm bổ sung.
  - Cung cấp khả năng mở rộng phân phối và khôi phục sao lưu dự phòng.
- Một số phương pháp đảm bảo toàn vẹn dữ liệu trong DBMS:
  - Kiểu dữ liệu (Data Type)
  - Không cho phép định nghĩa Null (Not Null Definitions)
  - Định nghĩa mặc định (Default Definitions)
  - Các thuộc tính định danh (Identity Properties)
  - Các ràng buộc (Constraints)
  - Các quy tắc (Rules)
  - Triggers
  - Các chỉ mục (Indexes)

Câu 4: Trình bày những lớp người dùng chính của một hệ thống an toàn CSDL và vai trò của họ.

- Người dùng cơ sở dữ liệu có thể chia thành các lớp như sau:
  - Lớp thứ nhất: Lập trình viên cơ sở dữ liệu là người viết chương trình ứng dụng sử dụng cơ sở dữ liệu thông qua một ngôn ngữ nào đó, như: C++, C#, ASP, PHP,… Các chương trình này sử dụng các phép toán CSDL thông thường như: thêm, xóa, sửa… chủ yếu sử dụng các câu lệnh SQL. Các chương tình có thể được viết theo lô các lệnh hoặc cũng có thể hoạt động trực tuyến – nghĩa là giao tiếp trực tiếp với DBMS. Chức năng hoạt động trực tuyến thường được sử dụng để quản trị CSDL.
  - Lớp thứ hai: Người dùng cuối, là người sử dụng các chương trình đã lập sẵn để giao tiếp với cơ sở dữ liệu. Các chương trình đã lập sẵn gồm các chương trình được lập bởi lập trình viên hoặc là 1 phần của DBMS. Phần lớn các DBMS đều cung cấp nhiều tiện ích lập sẵn. Một trong các tiện lợi cơ bản đó là giao diện truy vấn. Trong giao diện này, người dùng có thể đưa ra các câu lệnh SQL và phần mềm sẽ cho kết quả của câu lệnh đó.
  - Lớp thứ ba: Những người quản trị cơ sở dữ liệu (DBA – Database Administrator), là người làm công tác quản trị cơ sở dữ liệu.
- Đối với một hệ thống quản lý an toàn cơ sở dữ liệu, chúng ta cần có một lớp các người dùng sau:
  - Người quản lý ứng dụng: Có trách nhiệm đối với việc phát triển và duy trì, hoặc các chương trình thư viện.

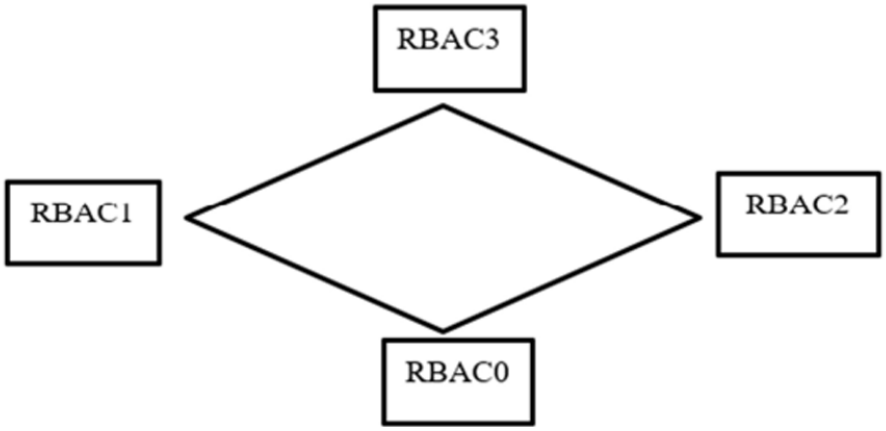
  DBA: Quản lý các lược đồ khái niệm và lược đồ bên trong của cơ sở dữ liệu.
  - Nhân viên an toàn: xác định các quyền truy nhập, các tiên đề, thông qua các quy tắc trong một ngôn ngữ thích hợp (có thể là DDL, hoặc DML).
  - Kiểm toán viên: chịu trách nhiệm kiểm tra các yêu cầu kết nối và các câu hỏi truy nhập, nhằm phát hiện ra các xâm phạm quyền.
  - Nhân viên sao lưu phục hồi: Thực hiện việc sao lưu dữ liệu và phục hồi dữ liệu cho cơ sở dữ liệu.

Câu 5: Trình bày việc gán/thu hồi quyền trong MAC và DAC.

- DAC
  - Trao truyền: việc trao quyền do người sở hữu đối tượng. Trong DAC có thể lan truyền quyền. Cần các cơ chế trao quyền phức tạp hơn, nhằm tránh mất quyền kiểm soát khi lan truyền quyền từ người trao quyền, hoặc những người có trách nhiệm khác. Ví dụ trong Oracle có grant option, admin option.
  - Thu hồi quyền: người dùng muốn thu hồi quyền phải có đặc quyền để thu hồi quyền. Trong Oracle nếu một người dùng có "grant option" thì người dùng đó có thể thu hồi quyền đã trao cho người khác.
- MAC
  - Hệ thống được quản lý bởi 1 người quản trị viên trung tâm, gọi là người trao quyền. Người trao quyền sẽ kiểm soát toàn bộ hệ thống trao quyền.
  - Các chủ thể và đối tượng được gán các nhãn an toàn nhất định. Từ đó quy định chủ thể có những quyền gì đối với đối tượng.
  - Người dùng không thể trao hay thu hồi quyền được gán. Mọi sự thay đổi chỉ được phép khi có sự đồng ý của người trao quyền.

Câu 6: Trình bày đặc điểm cơ bản về mô hình RBAC.

- Hầu hết các hệ quản trị cơ sở dữ liệu đều hỗ trợ RBAC. RBAC có thể dùng kết hợp với mô hình DAC hoặc MAC hoặc dùng độc lập.
- RBAC được áp dụng vào đầu những năm 1970. Khái niệm chính của RBAC là những quyền hạn được liên kết với những vai trò (role).
- Mục đích chính của RBAC là giúp cho việc quản trị an toàn một cách dễ dàng hơn.
- RBAC có thể giới han trước các mối quan hệ vai trò – quyền hạn, làm cho việc gán người dùng đến các vai trò được xác định trước 1 cách dễ dàng hơn.
- Vai trò (role) là một tập các quyền. RBAC thực hiện gán cho chủ thể một vai trò, khi đó chủ thể đó có mọi quyền thuộc vai trò đó.



- Mô hình RBAC gồm 4 mô hình con là: RBAC0, RBAC1, RBAC2, RBAC3.
  - Mô hình nền tàng RBAC0 ở dưới cùng, là yêu cầu tối thiểu cho bất kỳ hệ thống nào có hỗ trợ RBAC.
  - Mô hình RBAC1, RBAC2 được phát triển từ mô hình RBAC0 nhưn có thêm các điểm đặc trưng cho từng mô hình. RBAC1 thêm vào khái niệm của hệ thống phân cấp vai trò. RBAC2 thêm vào các ràng buộc. RBAC1 và RBAC2 không liên quan đến nhau.
  - Mô hình RBAC3 là mô hình tổng hợp của 3 mô hình RBAC0, RBAC1, RBAC2.

Câu 7: Tìm hiểu mô hình cấp quyền System R, bộ quyền trong System R. Đặc biệt lưu ý vấn đề thu hồi quyền đệ quy và không đệ quy

- System R là một hệ quản trị CSDL quan hệ đầu tiên của IBM, dựa trên nguyên tắc cấp quyền quản trị cho người sở hữu.
- Việc bảo vệ được thực hiện tại mức table:
  - Chủ thể: người dùng.
  - Đối tượng: các bảng và khung nhìn.
- Các chế độ truy nhập vào một table:
  - Read: đọc các bộ của 1 bảng. 1 user truy nhập read có thể định nghĩa các "views" trên table đó
  - Insert: thêm các bộ vào bảng
  - Delete: xóa các bộ trong bảng
  - Update: sửa đổi các bộ có trong bảng
  - Drop: xóa toàn bộ bảng
- System R hỗ trợ quản trị quyền phi tập trung:
  - Người tạo ra bảng có mọi đặc quyền trên bảng đó và có thể trao/thu hồi (grant/revoke) quyền cho các user khác.
  - Điều này có thể không đúng với các khung nhìn.
- Việc trao và thu hồi quyền của System R được thực hiện bằng các lệnh SQL.
- Người tạo ra bảng có mọi đặc quyền trên bảng đó và có thể trao/thu hồi quyền cho các user khác, mỗi quyền là 1 bộ gồm:

  <s, p, t, ts, g, go>
  - s: chủ thể được gán quyền.
  - p: đặc quyền được quyền.
  - t: tên bảng, trên đó truy nhập được gán.
  - ts: thời điểm quyền được gán.
  - g: người gán quyền.
  - go (yes/no): grant option.

  ⇨ Tham số go = yes, s có GRANT OPTION, nên có thể gán đặc quyền p cho các user khác.

  ⇨ Tham số: g và ts, để có thể thực hiện các hoạt động revoke quyền sau này 1 cách chính xác, bằng cách kiểm tra 1 loạt các quyền đã đc gán.
- Gán quyền: nếu một user được gán quyền trên một table với Grant option, người dùng có thể gán và thu hồi quyền cho các user khác các quyền anh ta có.
- Thu hồi quyền:
  - Sử dụng cơ chế thu hồi đệ quy.
  - Nếu x thu hồi quyền của y, trong khi đó x không gán quyền gì cho y trước đó, thì việc thu hồi quyền này bị loại bỏ.
  - Người dùng (người trao đặc quyền trên 1 bảng) cũng có thể ghi rõ từ khóa Public, thay cho users. Khi đó, tất cả những người dùng của CSDL đều được trao đặc quyền trên bảng.
- Thu hồi quyền đệ quy: khi người dùng A thu hồi quyền truy cập của người B thì tất cả các quyền mà B đã gán cho người khác đều được thu hồi.
  - Thu hồi quyền đệ quy trong system R dựa vào nhãn thời gian mỗi lần cấp quyền truy nhập cho người dùng.

- Thu hồi quyền không đệ quy: khi người A thu hồi quyền truy nhập trên B thì tất cả quyền truy nhập B cấp cho chủ thể khác được thay bằng A đã cấp cho những chủ thể này.
    - Thực tế khi một người dùng A thay đổi công việc hay vị trí thì đôi khi tổ chức chỉ muốn lấy lại quyền truy nhập của A mà không muốn lấy lại các quyền truy nhập mà A đã cấp => áp dụng thu hồi không đệ quy.
    - Vẫn dựa vào nhãn thời gian.

Câu 8: Nêu ví dụ về đặc quyền hệ thống (system prilvilege) và đặc quyền đối tượng (object privilega). Viết câu lệnh SQL cho các ví dụ đó. Nêu sự khác nhau giữa admin option và grant option. Ví dụ các câu lệnh SQL.

- Đặc quyền hệ thống: cho phép người dùng tạo những cơ sở dữ liệu mới, tạo các đối tượng mới bên trong cơ sở dữ liệu có sẵn, hay sao lưu cơ sở dữ liệu hoặc nhật ký giao tác.
    - Một số đặc quyền hệ thống:
      CREATE DATABASE
      CREATE TABLE
      CREATE PROCEDURE
      CREATE DEFAULT
      CREATE RULE
      CREATE VIEW
      BACKUP DATABASE
      BACKUP LOG
- Đặc quyền đối tượng: các quyền dùng đối tượng cho phép người sử dụng, role thực hiện những hành động trên một đối tượng cụ thể trong cơ sở dữ liệu.
    - Một số đặc quyền đối tượng:
      Select
      Insert
      Update
      Delete
      Execute
      Reference
- Lệnh gán quyền và thu hồi quyền
    - Lệnh gán quyền có dạng như sau:
      GRANT {ALL RIGHT (privileges) ALL BUT (privileges)} ON (table) TO (user-list) [WITH GRANT OPTION]

      Người sử dụng (người trao đặc quyền trên một bảng) cũng có thể ghi rõ từ khoá PUBLIC, thay cho (user-list). Khi đó, tất cả những người sử dụng của cơ sở dữ liệu đều được trao đặc quyền trên bảng.
    - Những người sử dụng (người có đặc quyền trên một bảng với tuỳ chọn trao) cũng có thể thu hồi đặc quyền trên bảng. Tuy nhiên, anh ta chỉ có thể thu hồi các quyền mà anh ta đã trao. Lệnh thu hồi của SQL có dạng như sau:
      REVOKE {ALL RIGHTS (privileges)} ON (table) FROM (user-list)
    - Một số câu lệnh SQL để gán và thu hồi các quyền này:
      Ví dụ: Hoa: SV (MaSV, Hoten, Điachi, SĐT, GioiTinh)
        - Gán quyền:

      Hoa GRANT Select, Insert ON SV TO Cuong WITH Grant Option;

GRANT Create Database, Alter Table TO HONG WITH Admin Option;

- Thu hồi quyền:

REVOKE Select, Insert ON SV FROM Cuong CASCADE;

- Sự khác nhau giữa Grant option và Admin option:

Admin option	Grant option
- Tùy chọn trong câu lệnh gán quyền hệ thống. - Cho phép chủ thể lan truyền quyền đó cho chủ thể khác.	- Tùy chọn trong câu lệnh gán quyền đối tượng. - Cho phép chủ thể lan truyền quyền đó sang chủ thể khác.

Câu 9: Đặc điểm cơ bản của hai phương pháp kiểm soát phụ thuộc dữ liệu gồm: Kiểm soát truy cập dựa trên khung nhìn và Sửa đổi truy vấn

a) Kiểm soát truy cập dựa trên khung nhìn

- Kiểm soát truy vấn trên khung nhìn:
  - Là cơ chế bảo vệ bên dưới của các DBMS dựa trên System R
  - Thay vì truy nhập trực tiếp vào bảng quan hệ cơ sở, người dùng chỉ được truy nhập vào các bảng khung nhìn ảo.
- Một bảng cơ sở (table) một bảng "thực" trong cơ sở dữ liệu
- Một khung nhìn (View) là một bảng "ảo" được đưa ra từ các bảng cơ sở và các khung nhìn khác
- Views:
  - Một user muốn tạo các view trên các table cơ sở, anh ta phải được:
  - Admin trao quyền create View.
  - Anh ta ít nhất phải có quyền read (select) trên các bảng cơ sở này, mới có quyền tạo các view.
- Người dùng (người định nghĩa một khung nhìn) là chủ sở hữu của khung nhìn. Tuy nhiên, chưa chắc anh ta đã được phép thực hiện tất cả các đặc quyền trên khung nhìn.
- Phụ thuộc vào các quyền mà người dùng có được trên các bảng có khung nhìn tham chiếu trực tiếp vào các bảng này.
  - Nếu user (người định nghĩa khung nhìn) được phép thực hiện một đặc quyền (ví dụ: SELECT) trên tất cả các bảng cơ sở với GRANT OPTION, thì anh ta cũng có đặc quyền đó với GRANT OPTION trên khung nhìn.
- Sau khi có một khung nhìn đã được định nghĩa, nếu người sở hữu khung nhìn nhận thêm hoặc bị thu hồi các đặc quyền trên các bảng cơ sở, thì các đặc quyền này sẽ được áp dụng trên khung nhìn, có nghĩa là người dùng sẽ được thêm hoặc bị thu hồi chúng trên khung nhìn.
  - Một view có thể được tạo từ một hoặc nhiều table cơ sở (Join).
  - Cú pháp tạo View:

```
create view view_name
as query_definition;
```

VD: Create View View_LopSV as
Select SV .* , Lop.TenLop From SV, Lop
Where SV.MaLop = Lop.MaLop;

- Kiểm soát truy cập cơ sở dữ liệu thường phụ thuộc vào dữ liệu.

Ví dụ: một số người dùng có thể bị giới hạn chỉ nhìn thấy các giá trị lương nhỏ hơn $30,000. một người trưởng phòng chỉ nhìn thấy các giá trị lương của các nhân viên trong phòng mình quản lý.

b) Sửa đổi truy vấn:

- Là một kỹ thuật khác phục vụ cho việc thực thi kiểm soát truy cập phụ thuộc dữ liệu.
- Đối mỗi một truy van ma ngươi dùng thuc hiện se đươc sửa đoi lại để hạn chế thêm người dùng sao cho phù hợp với quyền của người đó.
- Kỹ thuật này đã được ứng dụng trong Oracle và gọi là cơ chế VPD (Virtual Private Database).
- VD: có bảng EMPLOYEE

NAME	DEPT	SALARY	MANAGER
Smith	Toy	10,000	Jones
Jones	Toy	15,000	Baker
Baker	Admin	40,000	Harding
Adams	Candy	20,000	Harding
Harding	Admin	50,000	NULL

Thomas được cấp quyền sau:

GRANT SELECT ON EMPLOYEE TO Thomas WHERE DEPT = 'Toy'

Bây giờ g/s Thomas thực hiện:

SELECT NAME, DEPT, SALARY,MANAGER FROM EMPLOYEE

DBMS sẽ tự động sửa đổi truy vấn này:

```
SELECT NAME, DEPT, SALARY,MANAGER FROM EMPLOYEE
WHERE DEPT = 'Toy'
```

Câu 10: Khái niệm về chủ thể an toàn và đối tượng an toàn trong mô hình an toàn CSDL. Kể tên các mô hình an toàn CSDL và lấy ví dụ.

- Chủ thể an toàn (Security Subject): Là thực thể chủ động, có khả năng truy cập, thao tác, hoặc ảnh hưởng đến đối tượng an toàn.
  Các chủ thể có thể bao gồm: Là user hay các tiến trình (process)
- Đối tượng an toàn (Security Object): Là thực thể thụ động, được các chủ thể truy cập hoặc thao tác.
  Các đối tượng bao gồm: Là file, CSDL, bảng, khung nhìn, cột, hàng, ô (entry)
- Hai loại mô hình an toàn là:
  - Mô hình an toàn tùy ý (Discretionary security models - DAC) Mô hình ma trận truy nhập, Take- Grant
    Ví dụ: Một người dùng có thể cấp quyền đọc/ghi một tệp cho người khác.
  - Mô hình an toàn bắt buộc (Mandatory security models - MAC). mô hình Bell - Lapadula, BiBa
    Ví dụ: Một hệ thống phân cấp bảo mật (Top Secret, Secret, Confidential, Unclassified) chỉ cho phép người dùng ở cấp độ tương ứng truy cập.

CHƯƠNG 3. AN TOÀN TRONG DBMS

Câu 1: Tìm hiểu đặc điểm cơ bản của kiến trúc chủ thể tin cậy (Trusted Subject) và kiến trúc Integrity Lock. Đặc biệt chú ý kiến trúc Intergrity Lock

- Kiến trúc chủ thể tin cậy Trusted subject
  - Giả sử hệ quản trị cơ sở dữ liệu và một OS tin cậy.
  - DBMS hoạt động như là một chủ thể tin cậy của OS.
  - DBMS có trách nhiệm trong việc bảo vệ đa mức các đối tượng của CSDL.
  - Được sử dụng trong nhiều DBMS thương mại: sybase, informix, oracle…
  - Người dùng kết nới tơi DBMS qua các phần mềm untrusted front end (vì họ kết nối qua Internet).
  - Người dùng được phân loại các mức nhạy cảm khác nhau: high (cao), low (thấp), và một mức DBMS khác với hai mức trên.
  - Các chủ thể và đối tượng được gán một nhãn DBMS không giống với mức high và low.
  - Chỉ có các chủ thể được gán nhãn DBMS mới được phép thực hiện mã lệnh và truy nhập vào dữ liệu.
  - Các chủ thể có nhãn DBMS được coi là chủ thể tin cậy và được miễn kiểm soát bắt buộc của OS.
  - Các đối tượng CSDL được gán nhãn nhạy cảm. Ví dụ: các bộ, các giá trị.
  - Hệ quản trị Sybase tuân theo giải pháp này, với kiến trúc máy khách/máy chủ, sybase thực hiện gán nhãn mức bản ghi (mức hàng).
- Kiến trúc Intergrity Lock
  - Khóa toàn vẹn được đề xuất lần đầu tiên tại Viện nghiên cứu của lực lương không quân về an toàn cơ sở dữ liệu, được dùng để kiểm soát tính toàn vẹn và sự truy nhập cho cơ sở dữ liệu.

  - Kiến trúc Intergrity Lock đã có trong hệ quản trị thương mại TRUDATA.
  - TFE (bộ lọc tin cậy) thực thi bảo vệ nhiều mức bằng cách gắn các nhãn an toàn vào các đối tượng CSDL dưới dạng các tem – Stamp.
  - Một tem là một trường đặc biệt của một đối tượng, lưu thông tin về nhãn an toàn và các dữ liệu điều khiển liên quan khác.

- Tem là dang mã hóa của các thông tin trên, sử dụng một kỹ thuật niêm phong mật mã gọi là Intergrity Lock.
- TFE có nhiệm vụ tạo và kiểm tra các tem.
  - TFE sử dụng mật mã khóa bí mật để tạo tem và giải mã các tem. Các tem này có thể tạo ra dựa vào tổng kiểm tra (checksum).
  - Khóa bí mật chỉ có TFE biết.
- Insert dữ liệu: khi người dùng muốn insert một mục dữ liệu, TFE sẽ tính:
  - Tổng kiểm tra = mức nhạy cảm dữ liệu + dữ liệu.
  - Mã hóa tổng kiểm tra này bằng một khóa bí mật K, tạo ra tem, và lưu vào trong CSDL cùng với mục dữ liệu đó (gắn với mục dữ liệu).
- Đưa ra dữ liệu: khi đưa ra dữ liệu trả cho người dùng, TFE nhận được dữ liệu từ DBMS không tin cậy, nó sẽ kiểm tra tem gắn với mục dữ liệu xem có chính xác không:
  - Giải mã tem gắn với dữ liệu.
  - So sánh dữ liệu nhận được với dữ liệu sau khi giải mã tem. Nếu không khớp chứng tỏ dữ liệu đã bị sửa đổi.
  - Lưu ý: nếu dùng hàm băm để tạo tem, thì sau khi DBMS nhận được dữ liệu và tem tương ứng, nó sẽ băm dữ liệu này ra và so sánh với tem nhận được xem có trùng nhau không?
- Bộ lọc giao hoán: là quá trình tương tác với cả người sử dụng và hệ quản trị cơ sở dữ liệu.
- Bộ lọc back end: có trách nhiệm trong việc định nghĩa khung nhìn được phép tối đa bằng cách phát hiện tất cả các bản ghi/ thuộc tính không được phép, thay thế các yếu tố không được phép bằng giá trị 0.
- Bộ lọc front end.

Câu 2: Tìm hiểu kỹ về khái niệm và ý tưởng của tấn công SQL Injection

- SQL Injection một kỹ thuật tấn công cho phép kẻ tấn công lời dụng lỗ hổng của việc kiểm tra dữ liệu đầu vào trong các ứng dụng web và các thông báo lỗi của hệ quản trị cơ sở dữ liệu trả về để tiêm vào và thi hành các câu lệnh SQL một cách trái phép.
- SQL Injection có thể cho phép kẻ tấn công thực hiện các thao tác delete, insert, update… trên cơ sở dữ liệu của ứng dụng, thậm chí là máy chủ của ứng dụng đang chạy.
- Là kỹ thuật được sử dụng dựa trên các lỗi và lỗ hổng bảo mật các lớp ứng dụng để thực hiện một cuộc tấn công vào cơ sở dữ liệu hoặc dữ liệu.
- Hậu quả để lại rất nghiêm trọng vì nó cho phép kẻ tấn công có thể thực hiện các thao tác xóa, hiệu chỉnh, hoặc có toàn quyền trên cơ sở dữ liệu của ứng dụng.

- Thường xảy ra trên các ứng dụng web có dữ liệu được quản lý bằng các hệ quản trị cơ sở dữ liệu như SQL server, My SQL, Oracle…
- Cách phòng tránh SQL Injection:
  - Giới hạn quyền người dùng.
  - Loại bỏ các dấu, ký tự đặc biệt như: ‘:’, ‘- -‘, ‘/’,…
  - Giới hạn những Textox và Input.
  - Mã hóa cơ sở dữ liệu ở các trường hoặc các bản ghi quan trọng.
  - **…**

Câu 3: Các đặc điểm khác nhau giữa DBMS và OS

- Độ chi tiết của đối tượng (Object granularity):
  - OS: độ chi tiết ở mức tệp (file), thư mục, thiết bị.
  - DBMS: chi tiết hơn tới table, rows, fields, entry.
- Các tương quan ngữ nghĩa trong dữ liệu (Semantic correlations):
  - OS: không có.
  - DBMS: dữ liệu có ngữ nghĩa và liên quan với nhau thông qua các quan hệ ngữ nghĩa như:
    - Data
    - Time
    - Context
- Histo Siêu dữ liệu (Metadata):
  - OS: không có
  - DBMS: siêu dữ liệu cung cấp thông tin về cấu trúc của dữ liệu như: table, view, rows, fields,
- Các đối tượng logic và vật lý:
  - OS: chứa các đối tượng vật lý như: file, memory, process, devices ....
  - DBMS: chứa các đối tượng logic như: table, view, index, column, rows, entry ... và chúng độc lập với các đối tượng của OS.
- Multi-datatypes:
  - OS: có các truy nhập vật lý: như Read, write, execute ...
  - DBMS: có rất nhiều kiểu dữ liệu, do đó các CSDL cũng yêu cầu nhiều chế độ truy nhập như: chế độ thống kê, chế độ quản trị, select, insert, update, delete ...
- Các đối tượng động và tĩnh:
  - OS: quản lý các đối tượng tĩnh và tương ứng với các đối tượng thực.
  - DBMS: quản lý cả các đối tượng có thể được tạo ra động như: views hay SQL query và không có các đối tượng thực tương ứng.

Câu 4: Những đặc điểm cơ bản của Kiến trúc Trusted Subject và Kiến trúc Woods Hole.

- Các kiến trúc Woods Hole sử dụng DBMS không tin cậy cùng với một bộ lọc tin cậy và không quan tâm đến OS có tin cậy hay không.
  - Được phát triển năm 1982 bởi National Research Council
- Nhận xét:
  - Phần mềm front ends va DBMS đều không tin cậy (Không quan tâm OS có tin cậy hay không)
  - Phần mềm untrusted front-end thuc hiện các công việc xử lý trước và sau các câu truy vấn (phân tích, tối ưu hóa, phép chiếu).
  - Phần mềm trusted front end (TFE) ở giữa thực thi các chức năng an toàn và bảo vệ nhiều mức, vì vậy hoạt động như một TCB (Trusted Computing Base).
    - Kiến trúc Integrity Lock
    - Kiến trúc Kernelized
    - Kiến trúc Replicated (còn được gọi là kiến trúc Distributed)

Câu 5: Các đặc điểm cơ bản của ba cơ chế: Xác thực, ủy quyền và kiểm toán trong các DBMS.

Các đặc điểm cơ bản của ba cơ chế: Xác thực, ủy quyền và kiểm toán trong các DBMS.

- Xác thực (Authentication):
  - Là quá trình xác nhận định danh của các cá nhân hay ứng dụng có yêu cầu truy nhập tới một môi trường an toàn.
  - Có ba mức xác thực thường xuyên trong môi trường cơ sở dữ liệu, đó là:
    - Mức hệ điều hành.
    - Mức cơ sở dữ liệu.
    - Hỗ trợ của bên thứ ba.
- Ủy quyền (Authorization): Là quá trình đảm bảo rằng những cá nhân hoặc các ứng dụng yêu cầu truy nhập vào một môi trường hoặc một đối tượng trong môi trường có sự cho phép hay không.
- Kiểm toán (Auditing): Là giám sát việc sử dụng tài nguyên hệ thống của người dùng. Các cơ chế này bao gồm hai giai đoạn:
  - Giai đoạn ghi vào nhật ký: tất cả các câu hỏi truy nhập và câu trả lời liên quan đều được ghi lại (dù được trả lời hay bị từ chối).
  - Giai đoạn báo cáo: các báo cáo của giai đoạn trước được kiểm tra, nhằm phát hiện các xâm phạm hoặc tấn công có thể xảy ra.

CHƯƠNG 4. ỨNG DỤNG MẬT MÃ TRONG AT CSDL

Câu 1: Nêu các lợi ích, nguyên tắc và tác động của việc mã hóa dữ liệu đối với một tổ chức, doanh nghiệp

Lợi ích, nguyên tắc và tác động của mã hóa

- Mã hóa co sở dữ liệu mang lại những lợi ích sau:
  - Bảo đảm tính bí mật cho cá nhân và tổ chức
  - Phương pháp đơn giản, hiệu quả nhất đáp ứng các yêu cầu của tổ chức.
  - Bảo đảm an toàn cho dữ liệu có giá trị nhất của tổ chức.
  - Nâng cao bảo vệ và giảm rủi ro an toàn cho dữ liệu.
  - Góp phần bảo đảm an toàn cho hoạt động của tổ chức.
  - Duy trì tính cạnh tranh.
  - Bảo đảm an toàn cho dữ liệu outsource.
  - Đáp ứng các yêu cầu và quy định quản trị.
- Nguyên tắc mã hóa:
  - Đề mã hóa dữ liệu, các khóa ngẫu nhiên được sử dụng. Các khóa tương ứng được sử dụng để giải mã dữ liệu.
  - Tính bảo mật của dữ liệu được mã hóa phụ thuộc vào thuật toán mã hóa được sử dụng, khóa, kích thước khóa và việc thực hiện thuật toán mã hóa.
- Tác động của mã hóa dữ liệu
  - Mã hóa dữ liệu thường có thể ảnh hưởng nghiêm trọng đến hiệu suất nếu không có kế hoạch khôn ngoan về chiến lược mã hóa.
  - Mã hóa và giải mã dữ liệu chắc chắn sẽ gây ra một số suy giảm hiệu suất do bản chất của tính toán mã hóa, tùy thuộc vào lượng dữ liệu được mã hóa, thuật toán mã hóa và kích thước khóa.

Câu 2: Các chiến lược mã hóa CSDL (mã hóa bên trong DBMS và bên ngoài DBMS)

Dữ liệu sẽ được mã hóa ngay sau khi chúng được lưu trữ

Lợi thế của chiến lược mã hóa này là nó trong suốt đối với các ứng dụng, do đó không cần thay đổi đối với các ứng dụng.

- Mã hóa bên trong DBMS Có hai dạng sau:
  - Mã hóa mức lưu trữ
  - Thực hiện mã hóa dữ liệu là mã hóa dữ liệu trong hệ thống lưu trữ phụ (mã hóa toàn bộ tệp và thư mục và bảo vệ dữ liệu lưu trữ).
  - Chiến lược mã hóa này không thể liên quan đến đặc quyền của người dùng.
  - Mã hóa mức cơ sở dữ liệu: Mã hóa này có thể liên quan đến lược đồ cơ sở dữ liệu (có thể được mã hóa ở cấp độ bảng, hàng hoặc cột.)
  - Mã hóa ở muc co sở du liệu có thể làm giảm hiệu suat vì nó phức tạp ở việc lập chỉ mục dữ liệu được mã hóa.

- Mã hóa bên ngoài DBMS
    - Nếu cần mã hóa dữ liệu trong quá trình truyền thì một giải pháp phù hợp hơn là mã hóa dữ liệu bên ngoài DBMS ở mức ứng dụng. Do đó, dữ liệu được truyền dưới dạng văn bản mã và được lưu trữ, truy xuất từ DBMS ở dạng mã hóa.
    - Giải pháp này mang đến sự linh hoạt về thuật toán mã hóa có thể làm giảm chi phí hoạt động và tăng tính bảo mật.
    - Ngoài ra, giải pháp này có khả năng mở rộng cao liên quan đến số lượng người dùng và cơ sở dữ liệu được mã hóa (nghĩa là ngươi ta có thể thêm nhiều cơ sở dữ liệu mà không cần sửa đổi máy chủ mã hóa).

Câu 3: Tìm hiểu những vấn đề cần thiết khi mã hóa CSDL (vấn đề mã hóa ở đâu, bảo vệ khóa, phân phối khóa như thế nào…)

- Mã hóa ở: mã hóa khi dữ liệu lưu trữ trên các hệ thống máy tính và truyền qua internet hay các mạng máy tính khác, mã hóa ổ đĩa, file, email, điện mật
- Bảo vệ khóa: tăng cường độ khó các thuật toán, đảm bảo tính bí mật toàn vẹn và xác thực, làm khó viec phân tích thành thừa số nguyên tố.
- Phân phối khóa được định nghĩa là cơ chế một nhóm chọn khoa mật và sau đó truyền nó đến các nhóm khác. Còn thỏa thuận khoa là giao thức để hai nhóm (hoặc nhiều hơn) liên kết với nhau cùng thiết lập một khoa mật bằng cách liên lạc trên kênh công khai.

CHƯƠNG 5 CƠ SỞ DỮ LIỆU THỐNG KÊ

Câu 1: Cơ sở dữ liệu thống kê (statistical database) là gì? (Viết được các câu lệnh SQL cho các thống kê). Ứng dụng trong thực tế? Các dạng biểu diễn.

- Cơ sở dữ liệu thống kê (SDB):
    - Là một cơ sở dữ liệu được sử dụng cho mục đích phân tích thống kê.
    - Là một cơ sở dữ liệu chứa các bản ghi nhạy cảm mô tả về các cá nhân nhưng chỉ các câu truy vấn thống kê như: COUNT, SUM, AVERAGE, MAX, MIN… mới được trả lời, ngoài các câu truy vấn này thì những truy vấn vào các mục dữ liệu riêng sẽ không được đáp lại.
    - Sự khác biệt chính với CSDL quan hệ thông thường đó là với một SDB những câu truy vấn thống kê mới đc phép truy vấn, những câu truy vấn vào từng trường hợp dl riêng lẻ đều coi không hợp lệ
- Các câu lệnh SQL thống kê
    - Select Count (*) from NV.
    - Select SUM(luong) as sum_luong from NV.
    - Select AVG(luong) as avg_luong from NV.
    - Select MIN(luong) from NV.
    - Select MAX(luong) from NV.

- Ứng dụng của SDB trong thực tế:
  - Điều tra dân số.
  - Thống kê số người tử vong.
  - Về kế hoạch kinh tế.
  - Thống kê khám chữa bệnh.
  - Thống kê các vụ tai nạn ô tô.
  - Thống kê về các lĩnh vực kinh tế, giáo dục, tài chính, thương mại…
  - Phân tích và đưa ra chiến lược.
- Các dạng biểu diễn của SDB:
  - Dạng cơ sở dữ liệu quan hệ: SDB được biểu diễn ở dạng bảng hai chiều bình thường như các cơ sở dữ liệu quan hệ khác.
  - Dạng cơ sở dữ liệu vĩ mô: biểu diễn bằng các bảng chứa các thống kê vĩ mô. Các thống kê thường là count, sum, min, max, avg…

Câu 2: Tìm hiểu những khái niệm cơ bản trong CSDL thống kê

- SDB tĩnh: là SDB không thay đổi trong suốt thời gian tồn tại của chúng. VD: CSDL thống kê dân số
- SDB động: thay đổi liên tục theo sự thay đổi của dữ liệu thực, cho phép sửa đổi để phản ánh các thay đổi động của thế giới thực. VD: CSDL nghiên cứu trực tuyến, lớp học trực tuyến khi bổ sung thành viên…
- SDB trực tuyến (online): người sử dụng nhận được các phản hồi thời gian thực cho các câu truy vấn thống kê của mình.
- SDB ngoại tuyến (offline): người sử dụng không biết khi nào các thống kê của họ được xử lý, việc SDB bị lộ sẽ khó khăn.
- Kiến thức làm việc (working knowledge): là tập các mực thông tin (field) và giá trị thuộc tính trong SDB và các kiểu thống kê có sẵn trong SDB mà người dùng có thể biết một cách hợp lệ.
- Kiến thức bổ sung của người sử dụng (sumplementary knowledge): người sử dụng có thể có kiến thức bên ngoài về các cá nhân được biểu diễn trong SDB. Người dùng hoàn toàn có thể lợi dụng kiến thức này cho các mục đích xấu để suy diễn.
- Công thức đặc trưng: là một công thức logic, được ký hiệu bởi 1 chữ cái viết hoa, trong đó các giá trị thuộc tính được kết hợp với nhau thông qua các toán tử boolean như or, and, not.
- Tập truy vấn (query set) của một công thức đặc trưng C là tập tất cả các bản ghi thỏa mãn C. Kí hiệu: X(C).

  COUNT(C) = |X(C)| đây được gọi là lực lượng của X(C).
- Thống kê trên C: là các câu truy vấn thống kê trên C. Kí hiệu q (C). Ví dụ: COUNT (C).

- Khái niệm bậc: một thống kê gồm m thuộc tính khác nhau được gọi là thống kê bậc m. VD: Count ((GioiTinh = F) ∧ (MaPhong = Phong1)) là một thống kê bậc 2.
  Count(*) là một thống kê bậc 0.
- Thống kê nhạy cảm: là thống kê được tính toán trên một thuộc tính bí mật trong tập truy vấn có kích cỡ bằng 1.
  VD: Count(Tuoi < 30) = 1 => SUM(Luong, Tuoi < 30) là thống kê nhạy cảm.

Câu 3: Thế nào là thống kê nhạy cảm, cho ví dụ? Working knowledge và Supplementary knowledge?

- Thống kê nhạy cảm:
  - Là thống kê có thể được sử dụng để nhận dạng thông tin bí mật về 1 cá nhân được biểu diễn trong SDB.
  - Là thống kê được tính toán trên một thuộc tính bí mật trong tập truy vấn có kích cỡ bằng 1.
- Ví dụ: COUNT(AGE > 50) =1

  ⇨ SUM (Salary, AGE > 50) là thống kê nhạy cảm.
- Kiến thức làm việc (working knowledge): là tập các mục thông tin (field) và các giá trị thuộc tính trong SDB và các kiểu thống kê có sẵn trong SDB mà người dùng có thể biết một cách hợp lệ.
- Kiến thức bổ sung (sumplementary knowledge): người sử dụng có thể có kiến thức bên ngoài về các cá nhân được biểu diễn trong SDB. Người dùng hoàn toàn có thể lợi dụng những kiến thức này cho các mục đích xấu để suy diễn.

Câu 4: Nêu đặc điểm cơ bản về Tấn công dựa vào Trình theo dõi (trình bày được ý tưởng của 2 kiểu tấn công này) và cho ví dụ.

a) Tấn công dựa vào trình theo dõi (Tracker)

- Trình theo dõi là một tập các công thức đặc trưng, có thể được sử dụng để đưa thêm bản ghi vào các tập truy vấn kích cỡ nhỏ, làm cho kích cỡ của chúng nằm trong khoảng [k, N-k].
- Thông qua các trình theo dõi có thể tính toán được các thống kê bị hạn chế.
- Giả sử C là công thức đặc trưng người dùng yêu cầu. T là một trình theo dõi. T thỏa mãn điều kiện: K <= |X(T)| <= N - K.
  - Tấn công kiểu 1: K = 2 (chỉ thực hiện được khi C = A ∧ B)
    - Giả sử:

      User cần tính count (C), sum (C, luong).

      Công thức C (A ∧ B), count (C) = 1. Câu truy vấn này bị cấm.
    - Tấn công:

      Tính T = A ∧ $\overline{B}$ thỏa mãn k ≤ |X (T)| ≤ N – k.

Tính gián tiếp count (C): Q (C) = Q (A ∧ B) = Q (A) – Q (A ∧ $\overline{B}$)

=> Q (C) = Q (A) – Q (T).

- Tấn công kiểu 2:
  - Giả sử: cần tính count (C), count (C) < k. Đây là thống kê bị cấm.
  - Tấn công:

    Chọn T thỏa mãn: k ≤ |X (T)|, |X($\overline{T}$)| ≤ N – k.

    Q (D) = Q(all) = Q(T) + Q($\overline{T}$) (Q(all) bị cấm)

    Tính gián tiếp Q (C): Q(C) = Q(C ∨ T) + Q(C ∨ $\overline{T}$) – Q (D)

b) Tấn công dựa vào hệ tuyến tính

- Là loại tấn công bằng cách giải một hệ phương trình có dạng HX = Q với mỗi phương trình tương ứng một câu truy vấn.

$$\lambda_{1,1}x_1 + \lambda_{1,2}x_2 + \cdots + \lambda_{1,n}x_N = q_1$$
$$\lambda_{2,1}x_1 + \lambda_{2,2}x_2 + \cdots + \lambda_{2,n}x_N = q_2$$
$$\cdots$$
$$\lambda_{k,1}x_1 + \lambda_{k,2}x_2 + \cdots + \lambda_{k,n}x_N = q_k$$

  Mỗi phương trình tương ứng một câu truy vấn.
- H là ma trận truy vấn
  - H[i, j] = 1 nếu bản ghi $x_j \in X(C_j)$ (tương ứng với q_j)
  - H[i, j] = 0 nếu ngược lại

$$H = \begin{vmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,n} \\ \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{k,1} & \lambda_{k,2} & \cdots & \lambda_{k,n} \end{vmatrix}$$

- $x_1, x_2 \ldots, x_N$ là giá trị của N bản ghi.
- $Q = (q_1, q_2 \ldots., q_N)$ là vector của các thông kê đưa ra.
- Ví dụ:

NhanVien

ID	Ten	ChucVu	Phong	Tuoi	GioiTinh	Luong
01	Nam	Nhân viên	Maketing	29	F	3500
02	Lan	Trưởng phòng	Kế hoạch	33	M	6200
03	Huệ	Nhân viên	Kế hoạch	27	F	4000
04	Minh	Giám sát viên	Maketing	24	F	3600
05	Quỳnh	Nhân viên	Tài vụ	24	F	2900

  - Giả thiết:

    C = (Phong = "Kế hoạch") ∧ (GioiTinh = F)

    Cần tính q = Count(C) = 1 => bị chặn!

- Thực hiện:

  Tính q_1 = Count(Phong = "Kế hoạch")

  Tính q_2 = Count(Phong = "Kế hoạch", GioiTinh = M)

$$\begin{cases} q_1 = 0x_1 + 1x_2 + 1x_3 + 0x_4 + 1x_5 = 3 \\ q_2 = 0x_1 + 1x_2 + 1x_3 + 0x_4 + 0x_5 = 2 \end{cases}$$

  ⇨ q_3 = Count(Phong = "Kế hoạch", GioiTinh = F)

  = $q_1 - q_2 = 3 - 2 = 1$

  ⇨ $q = q_3 = 1$

  Ví dụ trên: C = (Phong = "Kế hoạch") ∧ (GioiTinh = F)

  Cần tính: q = Sum(Luong, C)

  Tính $q_1 = X(C_1)$ = Count(Phong = "Kế hoạch") = 3

  Tính $q_2 = X(C_2)$ = Count(Phong = "Kế hoạch", GioiTinh = M) = 2

  Sum(Luong, C) = Sum(Luong, C_1) – Sum(Luong, C_2)

  = (6200 + 4000 + 2900) – (6200 + 4000)

  = 2900

  ⇨ như vậy kẻ tấn công đã tìm ra lương của người thỏa mãn C

Câu 5: Kể tên các tấn công suy diễn vào SDB và lấy ví dụ về tấn công dựa vào đếm.

- Các kỹ thuật chống suy diễn bao gồm:
  - Kỹ thuật khái niệm.
  - Kỹ thuật hạn chế.
  - Kỹ thuật gây nhiễu.
  - Kỹ thuật mẫu ngẫu nhiên.
- Ví dụ tấn công dựa vào đếm:
  - Đây là loại tấn công bằng cách kết hợp giá trị đếm với giá trị tổng (hoặc giá trị Min, Max, AVG) để thu được thông tin bí mật.
  - Ví dụ, xét một SDB về Công nhân:

    CONGNHAN(MACN, HoTen, ChucVu, DiaChi, Phong, Luong)

    Khi thực hiện: COUNT (Chuc Vu = "Phó phòng", Phong= "Tài vụ") = 1.

    Kẻ tấn công thực hiện tiếp:

    Min (Luong, (Chuc Vu= "Phó phòng", Phong= "Tài vụ"))

    Chẳng hạn kết quả thu được là 8000, như vậy kẻ tấn công có thể khám phá ra lương chính xác của người phó phòng, phòng Tài vụ.

Câu 6: Giải thích tấn công suy diễn "Interference attack", lấy ví dụ về tấn công này.

- Tấn công suy diễn là dạng tấn công kết hợp các câu truy vấn thống kê như: Count, Sum, Min, Max, AVG để thu được thông tin bí mật về một cá nhân trong SDB.
  - Ví dụ: Xét một SDB về Đảng viên: DANGVIEN(MaDV, HoTen, ChucVu, Phong, DV,HeSoLuong)

Khi thực hiện: COUNT (ChucVu = "Phó bí thư", Phong = "ATTT") = 1.

Kẻ tấn công thực hiện tiếp:

AVG(HeSoLuong, (ChucVu = "Phó bí thư", Phong= "ATTT"))

Chẳng hạn kết quả thu được là 10000, như vậy kẻ tấn công có thể khám phá ra lương chính xác của người phó phó bí thư đó.

Câu 7: Có những dạng biểu diễn nào của một SDB, giải thích từng dạng biểu diễn này và cho ví dụ đơn giản.

Có hai dạng biểu diễn của một SDB gồm:

- Dạng CSDL quan hệ: là dạng biểu diễn CSDL SDB dưới dạng các bảng CSDL dạng quan hệ gồm các bản ghi và các trường.
- Dạng CSDL vĩ mô: là dạng biểu diễn SDB gắn với một thống kê cụ thể như Count, Sum, Min, Max, AVG.

  Ví dụ:

  CSDL quan hệ về Sinh viên:

Mã sinh viên	Họ và tên	Giới tính	Địa chỉ	Lớp
MS01	Đinh Văn Tùng	M	Hà Nội	Toán
MS02	Bạch Hải Phương	F	Hà Nội	Lý

  CSDL vĩ mô về Sinh viên: Tổng(Sum) học bổng theo giới tính và theo lớp

	ATA	ATB	ATC	ATD
M	12	34	56	6
F	12	42	34	95
Tổng	24	76	90	101

Câu 8: Trình theo dõi (Tracker)? Trình bày sự khác nhau giữa hai kiểu tấn công trình theo dõi (kiểu 1 và kiểu 2).

- Trình theo dõi (Tracker):
  - Là một tập các công thức đặc trưng, có thể được sử dụng để đưa thêm bản ghi vào các các tập truy vấn kích cỡ nhỏ, làm cho kích cỡ của chúng nằm trong khoảng [k, N – k].
  - Thông qua các trình theo dõi có thể tính toán được các thống kê bị hạn chế.
- Hai kiểu tấn công trình theo dõi khác nhau ở chỗ:
  - Kiểu 1: chỉ thực hiện tấn công được với giả thiết công thức đặc trưng C có dạng: C = (A ^ B). Khi đó T có dạng: T = A ^ $\overline{B}$ phụ thuộc vào C.
  - Kiểu 2: Có thể tấn công được với mọi công thức đặc trưng C, và cần phải chọn công thức T, T không phụ thuộc vào C.

Cấu 9: Ưu, nhược điểm của Kiểm soát kích cỡ tập truy vấn với SDB.

- Ưu điểm:
  - Đưa ra kết quả chính xác
  - Chỉ chống được các tấn công đơn giản
- Nhược điểm:
  - Hạn chế khả năng hữu ích của SDB
  - Chỉ ngăn chặn được các tấn công đơn giản, khó có thể ngăn chặn được các tấn công phức tạp, như: Trình theo dõi, Tấn công hệ tuyến tính

Câu 10: Đặc điểm cơ bản của Kỹ thuật giấu ô.

- Kỹ thuật này được thiết kế cho các SDB vĩ mô (đưa ra các thống kê trong bảng 2 - chiều, như các thống kê dân số).
- Giấu ô trong các bảng như sau:
  - Giấu đi tất cả các ô tương ứng với các thống kê nhạy cảm
  - Giấu thêm các ô tương ứng với các thống kê có thế gián tiếp khám phá ra các thống kê nhạy cảm (Giấu bổ sung).

Câu 11: Ưu, nhược điểm của kỹ thuật giấu ô và kỹ thuật gây nhiễu dữ liệu.

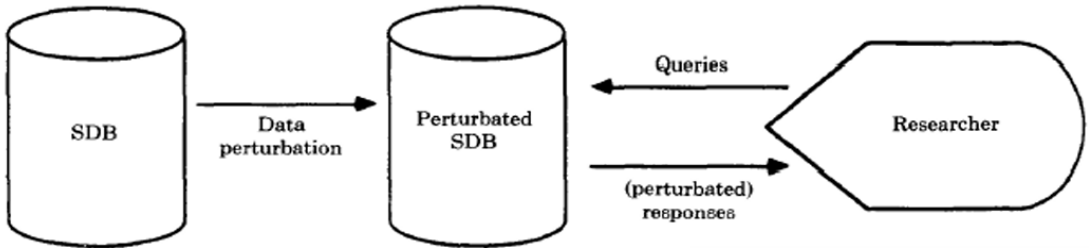
- Kỹ thuật giấu ô:
  - Ưu điểm: Chống đuoc các tấn công kết hợp dựa vào Count và Sum
  - Nhược điểm: Hạn chế khả năng hữu ích của SDB, vì phải che giấu một số ô trong CSDL.
- Kỹ thuật gây nhiễu
  - Gây nhiễu cố định
    - Ưu điểm: Chống được nhiều tấn công, kể cả tấn công tính trung bình (lặp nhiều lần)
    - Nhược điểm: - Chỉ áp dụng cho thuộc tính số
      - Kết quả trả về không chính xác
  - Gây nhiễu dựa vào truy vấn
    - Ưu điểm: Gây nhiễu dữ liệu nên chống được nhiều tấn công
    - Nhược điểm: - Với mỗi thống kê, lại phải áp dụng một hàm gây nhiễu f, với giá trị nhiễu=> tốn công, giảm hiệu năng hệ thống.
      - Kết quả đưa ra không chính xác.

Câu 12: Công thức đặc trưng là gì. Hãy viết các câu truy vấn ví dụ về Count, Sum, Min trên C.

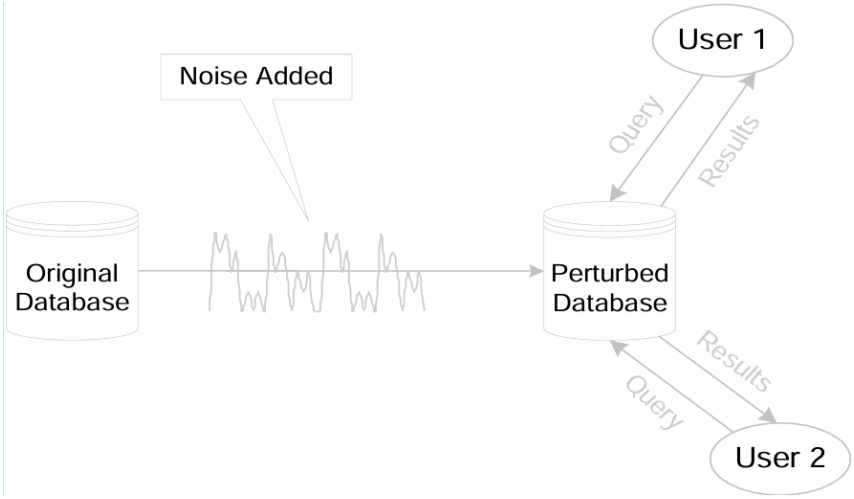
- Công thức đặc trưng: Là một công thức lôgíc, được ký hiệu bởi một chữ cái viết hoa (A,B,C,...), trong đó các giá trị thuộc tính được kết hợp với nhau thông qua các toán tử Boolean như OR, AND, NOT ($\vee$, $\wedge$, $\neg$).
- Các câu truy vấn ví dụ trên C như sau:
  Cho C = (GioiTinh = F) ^ [(MaPhong = "KếHoạch")(Maphong = "Tài vụ")] ^ (NamSinh < 1965)
  Select Count(*) from NHANVIEN where ((GioiTinh = F) and ((MaPhong = "Kếhoạch") or (MaPhong = "Tàivụ")) and (NamSinh < 1965)).
  Select Sum(Tuoi) from NHANVIEN where ((GioiTinh = F) and ((MaPhong = "Kếhoạch") or(MaPhong = "Tài vụ")) and (NamSinh < 1965)).
  Select Min(Luong) from NHANVIEN where ((GioiTinh = F) and ((MaPhong = "Kếhoạch") or(MaPhong = "Tài vụ")) and (NamSinh < 1965)).

Câu 13: So sánh, vẽ hình hai kỹ thuật gây nhiễu dữ liệu.

- Kỹ thuật gây nhiễu dữ liệu:



- Chia thành 2 loại:
  - Kỹ thuật gây nhiễu cố định (fixed perturbation);
  - Kỹ thuật gây nhiễu dựa vào truy vấn



- Kỹ thuật gây nhiễu cố định:
  - Ưu điểm: Chống được nhiều tấn công, kể cả tấn công tính trung bình (lặp nhiều lần)

- Nhược điểm: Chỉ áp dụng cho thuộc tính dân số, Kết quả trả về không chính xác

- Kỹ thuật gây nhiễu dựa vào truy vấn:
  - Ưu điểm: Gây nhiễu dữ liệu chống được nhiều tấn công
  - Nhược điểm: Với mỗi thống kê, lại phải áp dụng một hàm gây nhiễu f, với giá trị nhiễu => tốn công, giảm hiệu năng hệ thống; Kết quả đưa ra không chính xác

Câu 14: Tìm hiểu các kỹ thuật chống suy diễn trong CSDL thống kê, nêu ưu nhược điểm của từng phương pháp. (Chú ý tìm hiểu kỹ các kiểm soát này: Kiểm soát kích cỡ tập truy vấn, Kỹ thuật giấu ô, Kỹ thuật gây nhiễu)

- Các kỹ thuật chống suy diễn bao gồm:
  - Kỹ thuật khái niệm.
  - Kỹ thuật hạn chế.
  - Kỹ thuật gây nhiễu.
  - Kỹ thuật mẫu ngẫu nhiên.

a) Kỹ thuật khái niệm

- Bao gồm 2 kỹ thuật: mô hình lưới và phân hoạch khái niệm.
- ***Mô hình lưới:***
  - Là một mô hình khái niệm cung cấp nền tảng cho việc phát hiện những tấn công suy diễn có thể xảy ra với SDB.
  - Xuất phát từ thông tin thống kê được gộp ở nhiều mức khác nhau có thể gây dư thừa dữ liệu => người dùng có thể khám phá dữ liệu nhạy cảm.
  - Dựa trên cấu trúc lưới.
  - Gồm các bảng m – chiều ($0 \leq m \leq N$, N là số thuộc tính của bảng SDB): là các bảng được gộp dữ liệu từ 1 hay nhiều thuộc tính.
  - Tính trên một thống kê nào đó như: count, sum, avg…
  - Ưu điểm: là mô hình an toàn hiệu quả cho nghiên cứu các vấn đề suy diễn và các phương pháp kiểm soát suy diễn. Với nhiều bảng ở các mức gộp khác nhau, ta có thể phân tích:
    - Các kiểu tấn công suy diễn bằng câu truy vấn count, sum, average…
    - Các tấn công kiểu kết hợp các câu truy vấn khác nhau để suy diễn ra dữ liệu nhạy cảm…
    - So sánh các kiểm soát suy diễn: hạn chế tập truy vấn và gây nhiễu dữ liệu.
  - Nhược điểm:
    - Không thể cung cấp tính đầy đủ của cơ sở dữ liệu.
    - Không phù hợp với cơ sở dữ liệu động vì khi cập nhật SDB cần phải cập nhập tất cả các bảng trong mô hình lưới => tốn công.
- ***Phân hoạch khái niệm:***
  - Giải quyết các vấn đề chống suy diễn trong giai đoạn thiết kế khái niệm của SDB.

- Dựa vào việc định nghĩa tập các cá thể của SDB tại mức khái niệm, được gọi là các lực lượng.
- Dựa vào các điều kiện cần kiểm tra nhằm tránh suy diễn.
- Để hỗ trợ việc xác định các yêu cầu an toàn thống kê trong mô hình khái niệm này, người ta đề xuất hệ thống tiện ích quản lý an toàn thống kê (SSMF) gồm có 3 modul: PDC, UKC, CEC
  - PDC: xây dựng định nghĩa lực lượng – population definition construct.
  - UKC: xây dựng trình độ người dùng – user knowledge construct.
  - CEC: bộ thi hành và kiểm tra ràng buộc – constraint enforcer and checker.

b) Kỹ thuật hạn chế

- ***Kiểm soát kích cỡ tập truy vấn: (học kỹ)***
  - Một thống kê q(C) chỉ được phép nếu tập truy vấn của nó X(C) thỏa mãn quan hệ sau: $K \leq |X(C)| \leq N - K$ với $0 \leq K \leq N/2$.
  - Trong đó N là tổng số bản ghi trong SDB, k do DBA định nghĩa.
  - Ưu điểm:
    - Đưa ra được kết quả chính xác.
    - Kiểm soát này ngăn chặn các tấn công đơn giản, dựa vào các tập truy vấn rất nhỏ hoặc rất lớn.
  - Nhược điểm:
    - Hạn chế khả năng hữu ích của SDB
    - Chỉ ngăn chặn được các tấn công đơn giản, khó ngăn chặn được các tấn công phức tạp như trình theo dõi, tấn công hệ tuyến tính.
- ***Kiểm soát kích cỡ tập truy vấn mở rộng: (học kỹ)***
  - Tăng số lượng các tập truy vấn cần được kiểm soát.
  - Với một công thức đặc trưng C cho trước, chúng ta có thể định nghĩa các tập truy vấn ngầm định.
  - Các tập truy vấn ngầm sẽ được định dạng trực tiếp bằng cách kết hợp với các thuộc tính của C. Sau đó, các tập truy vấn này sẽ được kiểm soát để quyết định khi nào thì có thể trả lời câu truy vấn của người dùng.
  - Ưu điểm: chống được các kiểu tấn công: trình theo dõi, hệ tuyến tính.
  - Nhược điểm:
    - Phải kiểm tra 2^m tập truy vấn ngầm định (hàm mũ tăng rất lớn theo m) nên rất tốn công => giải pháp này khó thực hiện.
    - Ngoài tập truy vấn ngầm định, kẻ tấn công có thể sử dụng những công thức khác liên quan đến tập truy vấn này để tính ra truy vấn yêu cầu.
- ***Kiểm soát chồng lấp tập truy vấn:***

- Ý tưởng: các câu truy vấn liên tiếp được kiểm tra, dựa vào số lượng bản ghi của chúng, không cho phép chúng có số lượng bản ghi chung lớn hơn so với ngưỡng cho phép.
- Cho phép một thống kê yêu cầu q(C) chỉ khi thỏa mãn số lượng các bản ghi giữa tập truy vấn của q (C) và tập truy vấn của tất cả các thống kê q(D) đã được đưa ra phải nhỏ hơn hoặc ngang bằng với α (α là ngưỡng của hệ thống, chỉ ra số lượng các bản ghi chung tối đa được phép cho các tập truy vấn của 2 câu truy vấn liên tiếp).
- Ưu điểm: ngăn chặn được các tấn công như trình theo dõi, hệ tuyến tính.
- Nhược điểm:
  - Vẫn có khả năng bị phá vỡ bởi những tấn công được thiết kế theo dạng chuỗi các câu truy vấn liên tiếp.
  - Hạn chế khả năng hữu ích của SDB.
  - Kém hiệu quả do yêu cầu nhiều so sánh.
  - Tốn thời gian.

- ***Kỹ thuật gộp***
  - Các câu truy vấn thống kê được tính toán trên các nhóm gộp. dữ liệu riêng sẽ được nhóm lại thành một khối nhỏ trước khi đưa ra.
  - Giá trị trung bình của nhóm gộp sẽ thay thế cho mỗi giá trị riêng của sữ liệu được gộp.
  - Kỹ thuật này giúp ngăn chặn khám phá dữ liệu riêng.
  - Ưu điểm: tránh được việc để lộ thông tin nhạy cảm.
  - Nhược điểm: kết quả đưa ra không chính xác.
- ***Kỹ thuật giấu ô: (học kỹ)***
  - Được thiết kế cho các SDB vĩ mô (đưa ra các thống kê trong bản 2 chiều, như thống kê dân số).
  - Giấu ô: trong các bảng
    - Giấu đi tất cả các ô tương ứng với các thống kê nhạy cảm.
    - Giấu thêm các ô tương ứng với các thống kê có thể gián tiếp khám phá ra các thống kê nhạy cảm (giấu bổ sung).
  - Tiêu chuẩn giấu ô:
    - Thống kê count: kích cỡ tập truy vấn nhỏ hơn hoặc bằng 1, nghĩa là count (C) = 0, count (C) = 1.
    - Thống kê sum: tiêu chuẩn nhạy cảm được sử dụng là quy tắc <<đáp ứng n, trội k%>>.
    - Một thống kê là nhạy cảm nếu n giá trị thuộc tính của n hoặc ít hơn n bản ghi tạo thành k% hoặc lớn hơn k% trong toàn bộ thống kê sum cảu ô đó => ô này bị giấu. Các tham số n, k được giữ bí mật và do DBA xác định (n < N).

- Ưu điểm: chống được các tấn công kết hợp dựa vào count, sum.
- Nhược điểm: hạn chế khả năng hữu ích của SDB, vì phải che giấu một số ô trong cơ sở dữ liệu.

c) Kỹ thuật gây nhiễu

- ***Kỹ thuật gây nhiễu dữ liệu (học kỹ):***
  - Tiến hành trên các giá trị của bản ghi (được lưu trưc trong SDB) và sinh ra dữ liệu mới là các dữ liệu bị sửa đổi.
  - Chuyển đổi dữ liệu: còn gọi là chuyển đổi đa chiều. Tạo ra một SDB sửa đổi, trong đó các thống kê bậc t đầu tiên phải chính xác (t là một thống kê được tính toán trên t thuộc tính của các bản ghi SDB), các thống kê bậc cao hơn không nhất thiết phải chính xác.
  - Gây nhiễu cố định:
    - Tạo ra một CSDL được sửa đổi so với CSDL ban đầu. CSDL sửa đổi được tạo ra bởi việc thay đổi giá trị của mỗi trường bằng một giá trị nhiễu được sinh ra ngẫu nhiên.
    - Bao gồm: sửa đổi giá trị của các thuộc tính và tính toán các thống kê dựa vào các giá trị gây nhiễu. Với mỗi câu truy vấn, việc gây nhiễu không khác nhau.
    - Nhiễu được thực hiện chỉ 1 lần nên các truy vấn lặp đi lặp lại sẽ có kết quả là các giá trị tương thích. => lấy trung bình giá trị kết quả sẽ không thu được giá trị ban đầu của dữ liệu.
    - Ưu điểm: chống được nhiều tấn công, kể cả tấn công tính trung bình (lặp nhiều lần).
    - Nhược điểm: chỉ áp dụng cho thuộc tính số; kết quả trả về không chính xác.
  - Gây nhiễu dựa vào truy vấn:
    - Không yêu cầu tạo 1 SDB nhiễu.
    - Với mỗi truy vấn được tạo ra trong SDB, một hàm gây nhiễu sẽ được áp dụng với tất cả các thuộc tính của tập truy vấn đó.
    - Giả sử thống kê q(C), với mọi giá trị xij thuộc X(C): x'ij = fc(xij).
    - Giá trị $\varepsilon = x'_{ij} - x_{ij}$ là ngẫu nhiên.
    - Ưu điểm: gây nhiễu dữ liệu nên chống được nhiều tấn công.
    - Nhược điểm:
      - Với mỗi thống kê, lại phải áp dụng 1 hàm gây nhiễu f, với giá trị nhiễu => tốn công, giảm hiệu năng hệ thống.
      - Kết quả đưa ra không chính xác.
- ***Kỹ thuật gây nhiễu đầu ra:***
  - thực hiện sửa đổi trên các kết quả được tính toán chính xác của một câu truy vấn thống kê, trước khi chuyển nó cho người sử dụng.

- Dựa trên kỹ thuật làm tròn: kết quả mọi câu truy vấn sẽ được làm tròn: Q' = r (Q).
- Làm tròn có hệ thống: Q' là một kết quả sửa đổi, nó được tính toán cho thống kê yêu cầu q (C).
  - b' = ⌊(b+1)/2⌋
  - d = Q mod b.
  - r (Q) = Q nếu d=0
  - r (Q) = Q – d nếu d < b'
  - r (Q) = Q + b – d nếu d >= b'
- Làm tròn ngẫu nhiên: Q' là một kết quả sửa đổi, nó được tính toán cho thống kê yêu cầu q (C).
  - b'= ⌊(b+1)/2⌋
  - d = Q mod b.
  - r (Q) = Q nếu d=0
  - r (Q) = Q – d với xác suất 1 – p
  - r (Q) = Q +b – d với xác suất p
  - p = d/b
- Ưu điểm: bảo vệ được những tấn công đơn giản.
- Nhược điểm:
  - Không chống được những tấn công trung bình, tấn công trình theo dõi.
  - Kết quả đưa ra không chính xác.

d) Kỹ thuật mẫu ngẫu nhiên

- Sử dụng để ngăn chặn suy diễn trong các CSDL thống kê.
- Ý tưởng: sử dụng các mẫu bản ghi từ các tập truy vấn tương ứng với các truy vấn thống kê, thay vì lấy mẫu trong toàn bộ SDB.
- Cơ chế cơ bản là thay thế tập truy vấn (có liên quan đến một câu truy vấn thống kê) bằng một tập truy vấn được lấy mẫu gồm một tập con các bản ghi được chọn lựa chính xác trong tập truy vấn gốc.
- Sau đó tiến hành tính toán thống kê yêu cầu trên tập truy vấn mẫu này. Sử dụng hàm chọn f(C, i) để chọn lựa các bản ghi từ tập truy vấn gốc tương ứng với thống kê q(C) mà người dùng yêu cầu.
- Ưu điểm: chống lại các tấn công trình theo dõi, tấn công lấy trung bình kết quả được lấy mẫu.
- Nhược điểm: những câu truy vấn tương đương về mặt logic có thể đưa ra kết quả là các tập truy vấn khác nhau.

CHƯƠNG 6. KIỂM TOÁN + PHÁT HIỆN XÂM NHẬP CSDL

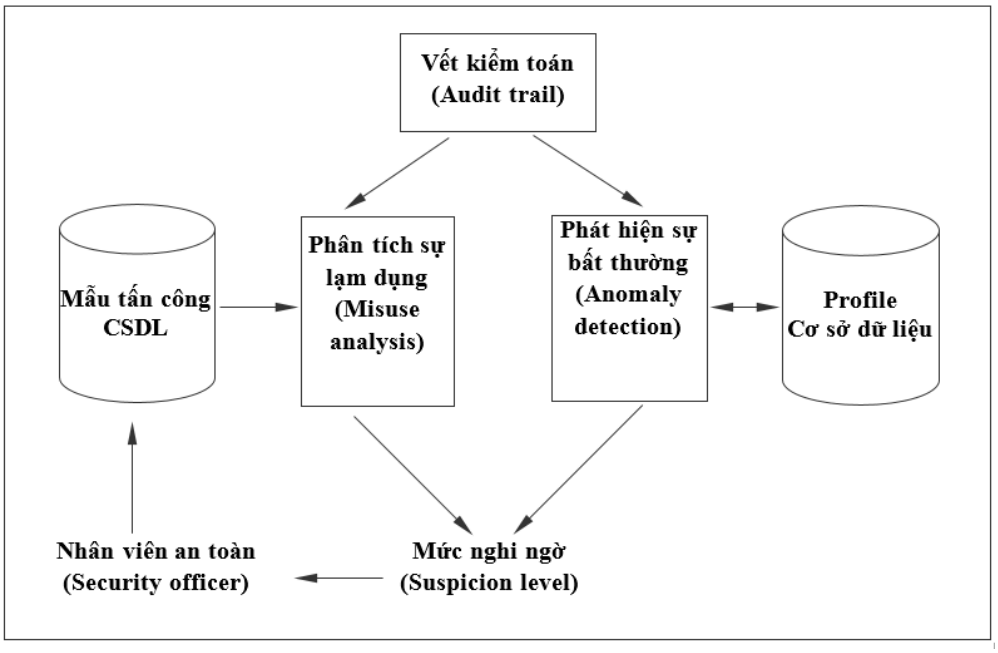
Câu 1: Vai trò và các cơ chế kiểm toán cơ bản cho CSDL

- Mục đích của kiểm toán là xem xét và đánh giá tính sẵn sàng, tính an toàn và tính chính xác thông qua việc trả lời những câu hỏi như:
  - Hệ thống máy tính có sẵn sàng cho hoạt động tại mọi thời điểm hay không?
  - Liệu môi trường CSDL có phải chỉ những người có thẩm quyền mới được sử dụng không?
  - Liệu môi trường CSDL đã cung cấp thông tin chính xác, trung thực và kịp thời hay chưa?
- Kiểm toán thường được sử dụng để:
  - Theo dõi các hành động hiện tại trong một lược đồ, bảng, hàng, cột hoặc một nội dung dữ liệu cụ thể.
  - Giúp người giám sát thấy được nếu có người sử dụng bất hợp pháp đang
  - ao tác với CSDL.
  - Điều tra hoạt động đáng ngờ.
  - Theo dõi và thu thập dữ liệu về các hoạt động CSDL cụ thể. Kiểm toán bắt buộc người dùng phải có trách nhiệm về hành động mà họ thực hiện, bằng cách theo dõi hành vi của họ.

Câu 2: Trình bày đặc điểm của mô hình phát hiện xâm nhập CSDL dựa trên bất thường và mô hình phát hiện xâm nhập CSDL dựa trên lạm dụng. Vẽ mô hình Kiến trúc của một IDS CSDL bao gồm cả hai mô hình phát hiện trên.

- Có hai mô hình mà hệ thống phát hiện xâm nhập áp dụng là: Các mô hình phát hiện bất thường và các mô hình phát hiện sự lạm dụng.
- Các mô hình phát hiện bất thường
  - Các mô hình này cho phép so sánh hồ sơ - profile (trong đó có lưu các hành vi bình thường của một người dùng) một cách có thống kê với các tham số trong phiên làm việc của người dùng hiện tại.
  - Các sai lệch "đáng kể" so với hành vi bình thường được báo cáo lại cho chuyên gia an toàn, trong đó sự "đáng kể" được xác định như là một ngưỡng, do mô hình xác định hoặc chuyên gia an toàn đặt ra.
  - Cho phép so sánh hồ sơ - profile (trong đó có lưu các hành vi bình thường của một người dùng) một cách có thống kê với các tham số trong phiên làm việc của người dùng hiện tại.
- Các mô hình phát hiện sự lạm dụng
  - Kiểu mô hình này trợ giúp việc so sánh các tham số trong phiên làm việc của người dùng và các dấu hiệu đã biết (được những kẻ tấn công sử dụng để xâm nhập vào một hệ thống), đây được gọi là các mẫu tấn công.

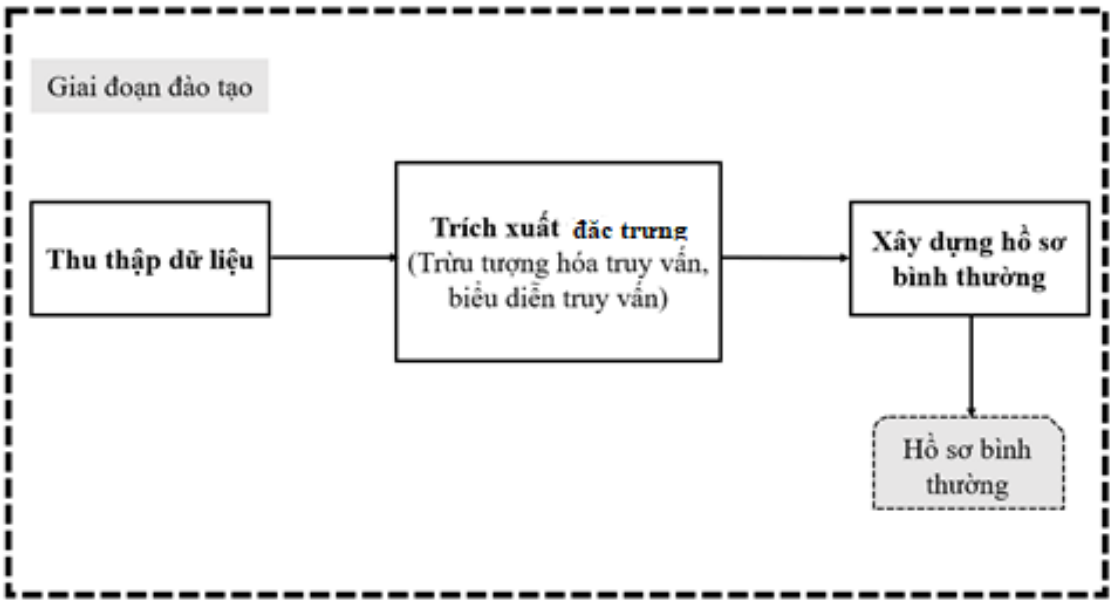
=> Nói chung, các kiểm soát đối với hành vi của người dùng trong hệ thống được giải quyết bằng cách theo dõi các yêu cầu mà người dùng thực hiện và ghi chúng vào một vết kiểm toán thích hợp.
=> Các kiểm soát kiểm toán trong các hệ thống kiểm toán truyền thống có điểm yếu là rất phức tạp và được tiến hành sau cùng.



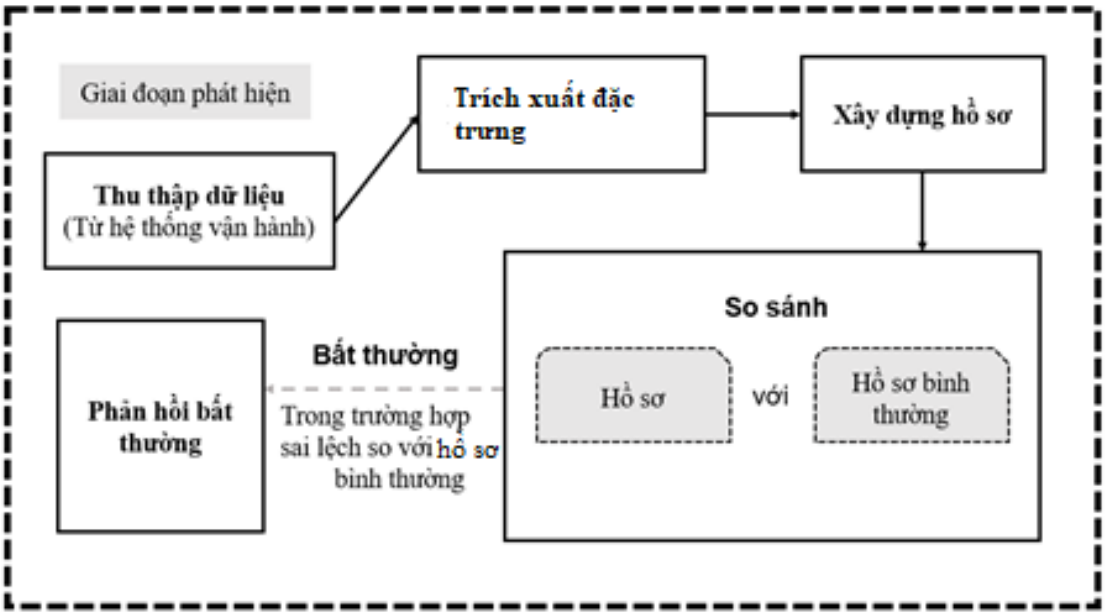
Kiến trúc của một IDS CSDL

Câu 3: Vẽ hình mô tả và giải thích hai giai đoạn: giai đoạn đào tạo và giai đoạn phát hiện của một hệ thống phát hiện xâm nhập cơ sở dữ liệu dựa trên bất thường.

- Giai đoạn đào tạo và giai đoạn phát hiện là hai giai đoạn trong kiến trúc này.
- Trong giai đoạn đào tạo, xây dựng hồ sơ hành vi chuẩn bằng cách so sánh với hồ sơ thời gian chạy trong giai đoạn phát hiện.
- Việc lập hồ sơ "hành vi chuẩn" đòi hỏi xem xét nhiều đặc điểm về truy vấn và xác định các đặc trưng quan trọng là một thách thức.
- Cả hai giai đoạn đều bao gồm hai thành phần cơ bản: *trích xuất đặc trưng* (trích rút) và *xây dựng hồ sơ*. Trong trích xuất đặc trưng, các đặc trưng cần thiết để xây dựng hồ sơ được trích xuất. Xây dựng hồ sơ là kỹ thuật để xây dựng hồ sơ bằng cách sử dụng các đặc trưng được chọn.
- Một hồ sơ về hành vi chuẩn được xây dựng trong giai đoạn đào tạo và được so sánh với hồ sơ thời gian chạy được xây dựng trong giai đoạn phát hiện. Những sai lệch của hồ sơ thời gian chạy so với hồ sơ chuẩn sẽ được gắn nhãn là bất thường.



Giai đoạn đào tạo của một hệ thống phát hiện bất thường.



Giai đoạn phát hiện của một hệ thống phát hiện bất thường.