* Các bước chính trong quá trình thiết kế một cơ sở dữ liệu.

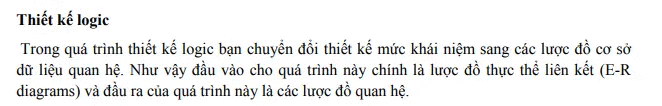
1. Phân tích yêu cầu
   * + Dữ liệu nào đc lưu trữ?
     + UD nào đc sd?
     + Thao tác nào sd thường xuyên
2. Thiết kế CSDL mức k/n: ttin 🡪 phát triển thành 1 mô tả tổng quát dữ liệu có ràng buộc cần thiết với nhau
3. Thiết kế CSDL mức logic: 1 hệ QT CSDL 🡪 chọn 🡪 cài đặt (mức k/n 🡪 lược đồ CSDL)
4. Cải tiến lược đồ: Các lược đồ đc phát triển => phân tích vấn đề tiềm ẩn => giải quyết => chuẩn hóa CSDL
5. Thiết kế CSDL mức vật lý:
   * + Khối lượng cvc tiêm ẩn
     + Phương pháp truy nhập đc mô phỏng

* Giải quyết => tiềm ẩn dlieu

1. Thiết kế an toàn bảo mật: ? giải quyết hacker

* CSDL:
  + Là 1 bộ sưu tập rất lớn (tài nguyên)
  + Bao gồm các loại dữ liệu: âm thanh, hình ảnh, văn bản, tiếng nói, … 🡪gắn kết với nhau 🡪 bảng
  + Được mã hóa thành các chuỗi bit và được lưu trữ dưới dạng file
  + Có 2 loại dữ liệu:
    - Dữ liệu vật lý: thô, thực được nhập vào CSDL
    - Dữ liệu dẫn xuất: dữ liệu được tính toán từ dữ liệu thuộc CSDL.
    - VD: CSDL lưu thông tin về SV, trong đó lưu các điểm thi của SV. Giá trị tb điểm thi của SV sẽ đc tính toán khi cần
* Hệ QTCSDL:
  + Hệ thống phần mềm
  + Cho phép tạo lập và điều khiển mọi truy nhập đến CSDL đó
  + Có 4 đặc tính:
    - Tạo mới CSDL = ngôn ngữ định nghĩa dữ liệu (DDLs)
    - Truy vấn = ngôn ngữ thao tác dữ liệu (DMLs)
    - Hỗ trỡ lưu trữ số lượng lớn trong time dài
    - Kiểm soát truy nhập dữ liệu từ nhiều người tại cùng 1 thời điểm
* Hệ CSDL:
  + là 1 CSDL được quản trị bởi 1 hệ QTCSDL
  + có 4 tp:
    - phần mềm hệ QTCSDL
    - phần cứng
    - CSDL hợp nhất:
      * Tối ưu hóa dư thừa
      * Được chia sẻ
    - Người dùng

- VD về hệ QTCSDL: MS Access, My SQL, SQL server, …



Thiết kế KN:

Đầu vào: thông tin, thuộc tính của các thực thể và quan hệ ( kết quả từ bước phân tích yêu cầu bài toán )

Đầu ra: sơ đồ quan hệ thực thể Er

**phép trừ** yêu cầu 2 quan hệ toán hạng phải là khả hợp

***Dạng chưa chuẩn hoá:***

Các quan hệ ở dạng chưa chuẩn hóa đồng nghĩa với việc chúng chưa ở dạng chuẩn 1.  
- ***Các quan hệ chưa ở dạng chuẩn 1*** chứa một hoặc một số thuộc tính không nguyên tố, các thuộc tính lặp, và các thuộc tính dẫn xuất.  
- ***Thuộc tính chứa giá trị nguyên tố:*** là những thuộc tính chứa giá trị đơn và không thể phân rã được nữa.

Một số khái niệm về khoá:

- ***Siêu khóa:*** Là một tập các thuộc tính xác định duy nhất thực thể trong quan hệ.  
- ***Khóa:*** là một siêu khóa mà khi loại bỏ bất kỳ thuộc tính nào từ khóa này thì nó không còn là một siêu khóa nữa. Nghĩa là, khóa có số thuộc tính là nhỏ nhất  
- ***Khóa dự bị:*** là một tập các thuộc tính khóa nhỏ nhất của lược đồ quan hệ.  
- ***Khóa chính:*** là một khóa dự bị được chọn ra. Tất cả các khóa dự bị còn lại trở thành khóa phụ hay khóa thứ cấp.

- ***Thuộc tính khóa:*** là thuộc tính của quan hệ R và là thành viên của một khóa dự bị nào đó.  
- ***Thuộc tính không khóa:*** là thuộc tính của quan hệ R mà không phải là thành viên của một khóa dự bị nào.

***\*Một quan hệ ở dạng chuẩn 1:***- Mọi giá trị thuộc tính của quan hệ đều ở dạng nguyên tố.  
- Không có thuộc tính đa trị.  
- Không có thuộc tính dẫn xuất

***\* Định nghĩa dạng chuẩn 2:*** Một lược đồ quan hệ R với tập phụ thuộc hàm F ở dạng chuẩn 2NF khi thỏa mãn:  
- Là dạng chuẩn 1NF  
- Mọi thuộc tính không khóa đều phụ thuộc hàm đầy đủ vào mọi khóa dự bị của R.

\* Một lược đồ quan hệ R với tập phụ thuộc hàm F ở dạng chuẩn 2NF khi thỏa mãn một trong các điều kiện sau:  
- Tất cả các khóa dự bị đều có một thuộc tính  
- Không có thuộc tính không khóa nào.  
- Không tồn tại thuộc tính không khóa mà phụ thuộc hàm một phần vào một khóa dự bị của R.

***\*Định nghĩa dạng chuẩn 3NF:*** Một lược đồ quan hệ R ở dạng 3NF với một tập phụ thuộc hàm F nếu thỏa mãn:  
¬ Với bất kỳ phụ thuộc hàm X → A trong F thì hoặc X là một siêu khóa của R, hoặc A là một thuộc tính khóa.  
\* ***Một cách định nghĩa khác về dạng chuẩn 3NF:*** Một lược đồ quan hệ R ở dạng 3NF với một tập phụ thuộc hàm F nếu thỏa mãn:  
- R ở dạng 2NF  
- Không có thuộc tính không khóa nào phụ thuộc hàm bắc cầu vào khóa của R

***Định nghĩa:*** Một lược đồ quan hệ R được coi là ở **dạng chuẩn Boyce-Codd** với tập phụ thuộc hàm F nếu với bất kỳ phụ thuộc hàm X → A nào và A ⊈ X, thì X là một siêu khóa của R.  
ϖ ***Ví dụ:*** Cho lược đồ quan hệ R = (A, B, C), tập phụ thuộc hàm F = {AB → C, C → A} và khóa K = {AB} R không ở dạng chuẩn BCNF vì có C → A và C không phải là một siêu khóa của R.

***Sự khác nhau giữa dạng chuẩn 3NF và BCNF*** là BCNF cho phép bỏ đi luật với phụ thuộc hàm X → A thì X phải là thuộc tính khóa. 3nf vẫn có thể: thuộc tính khóa phụ thuộc vào thuộc tính k khóa còn bcnf thì k cho phép

* *Trong thực tế, hầu hết các lược đồ quan hệ ở dạng chuẩn 3NF thì cũng ở dạng chuẩn BCNF. Chỉ khi trong lược đồ có X* → *A mà X không phải là một siêu khóa hoặc A là một thuộc tính khóa thì lược đồ này ở dạng  
  chuẩn 3NF mà không ở dạng chuẩn BCNF.\*

**Chuẩn hóa từ 1NF sang 2NF:**

* Loại bỏ các thuộc tính không khóa phụ thuộc 1 phần vào khóa chính để tách thành bảng riêng. Bảng này sẽ có khóa chính là bộ phận khóa mà chúng phụ thuộc
* Các thuộc tính còn lại thành một bảng với khóa là khóa chính ban đầu

**Chuẩn hóa từ 2NF sang 3NF:**

Graphical user interface, text, application, email

Description automatically generated

**- Kiểm tra đạt chuẩn 3NF -> BCNF**

- đã đạt 1NF

- các thuộc tính không khoá trong các bảng đã phụ thuộc hàm đầy đủ vào khoá => đạt 2NF

- các thuộc tính không khoá đã được suy ra trực tiếp từ thuộc tính khoá trong các bảng => đạt 3NF

- không có thuộc tính khoá nào phụ thuộc vào thuộc tính không khoá => đạt BCNF

*Mục đích chuẩn hoá:* **Mục đích** của **chuẩn hoá** là loại bỏ các dư thừa dữ liệu và các lỗi khi thao tác dư thừa và các lỗi khi thao tác dữ liệu (Insert, Delete, Update)

Khi xoá lưu ý: Hãy cẩn thận khi xóa các bản ghi trong bảng! Lưu ý mệnh đề WHERE trong câu lệnh DELETE. Mệnh đề WHERE chỉ định (các) bản ghi nào nên được xóa. Nếu bạn bỏ qua mệnh đề WHERE, tất cả các bản ghi trong bảng sẽ bị xóa! Quan tâm đến khoá ngoài

* Khóa ngoại của một table được xem như con trỏ trỏ tới khóa chính của table khác.

Khi INSERT cần lưu ý: quan tâm đến khoá chính

**Bao đóng của phụ thuộc hàm**

* Tập tất cả các phụ thuộc hàm được suy diễn từ một tập phụ thuộc hàm F được gọi là bao đóng của F và được ký hiệu là F+
* X+ là bao đóng của tập thuộc tính X trên tập phụ thuộc hàm F nếu mọi thuộc tính trong X+ đều được sinh ra từ X nhờ F

**Khoá chính:** được sử dụng để định danh duy nhất mỗi record trong table của cơ sở dữ liệu.

**Khoá ngoại:** Khóa ngoại của một table được xem như con trỏ trỏ tới khóa chính của table khác.

DDL: ngôn ngữ định nghĩa DL

DML: ngôn ngữ thao tác DL