

Database Design

Database 2 - Lecture 3

1. Introduction

Database design process integrates relevant data in such a manner that it can be processed through a mechanism for recording the facts. A database of an organization is an information repository that represents facts about the organization. It is manipulated by some software to incorporate the changes that take place in the organization. The database design is a complex process. The complexity arises mainly because of the identification of relationships among individual components and their representation for maintaining correct functionality are highly involved. The degree of complexity increases if there are many-to-many relationships among individual components. The process of database design usually requires a number of steps which are in Fig. 1.

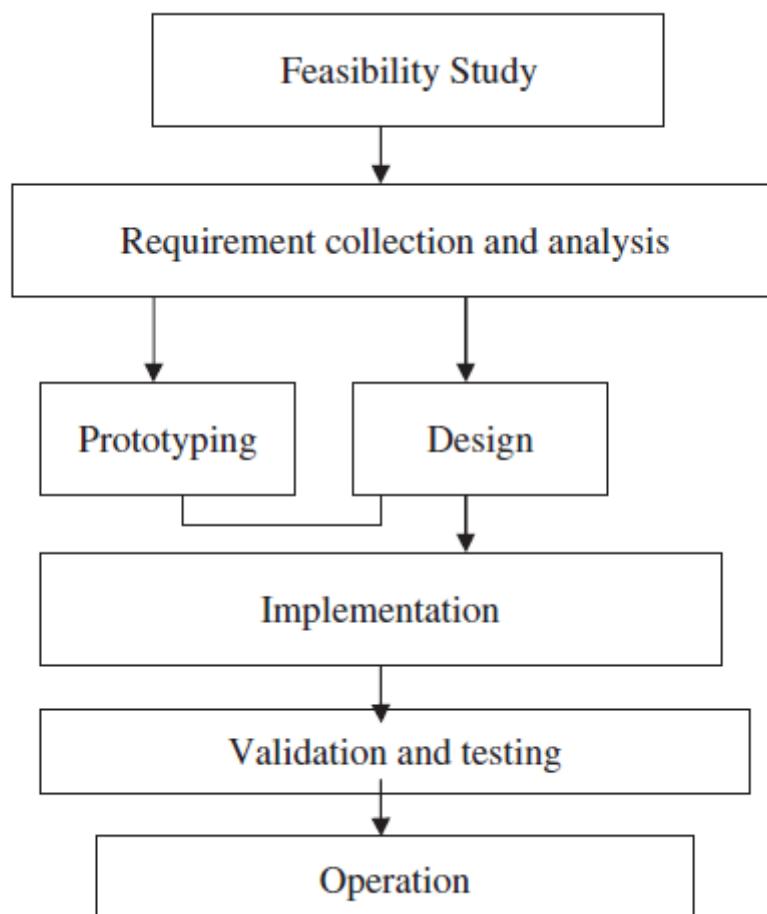


Fig. 1. Steps in database design

Feasibility Study

When designing a database, the purpose for which the database is being designed must be clearly defined. In other words, the objective of creating the database must be crystal clear.

Requirement Collection and Analysis

In requirement collection, one has to decide what data are to be stored, and to some extent, how that data will be used. The people who are going to use the database must be interviewed repeatedly. Assumptions about the stated relationships between various parts of the data must be questioned again and again. For example, in designing the database about medical records of a patient, the following queries must be clearly defined. Does a patient have more than one doctor? Is there a separate billing number for each drug ordered by a patient?

Prototyping and Design

Design implies a procedure for analyzing and organizing data into a form suitable to support business requirements and makes use of strategic technology. The three phases in relational database design are conceptual design, logical design, and physical design.

Implementation

Database implementation involves development of code for database processing, and also the installation of new database contents, usually from existing data sources.

2. Objectives of Database Design

The objectives of database design vary from implementation to implementation. Some of the important factors like efficiency, integrity, privacy, security, implementability, flexibility have to be considered in the design of the database.

Efficiency

Efficiency is generally considered to be the most important. Given a piece of hardware on which the database will run and a piece of software (DBMS) to run it, the design should make full and efficient use of the facilities provided. If the database is made online, then the users should interact with the database without any time delay.

Integrity

The term integrity means that the database should be as accurate as possible. The problem of preserving the integrity of data in a database can be viewed at a number of levels. At a low level it concerns ensuring that the data are not corrupted by hardware or software errors. At a higher level, the problem of preserving database integrity concerns maintaining an accurate representation of the real world.

Privacy

The database should not allow unauthorized access to files. This is very important in the case of financial data. For example, the bank balance of one customer should not be revealed to other customers.

Security

The database, once loaded, should be safe from physical corruption whether from hardware or software failure or from unauthorized access. This is a general requirement of most databases.

Implementation

The conceptual model should be simple and effective so that mapping from conceptual model to logical model is easy. Moreover, while designing the database, care has to be taken such that application programs should interact effectively with the database.

Flexibility

The database should not be implemented in a rigid way that assumes the business will remain constant forever. Changes will occur and the database must be capable of responding readily to such change. Other than the factors which were mentioned above, the design of the database should ensure that data redundancy is not there.

3. Database Design Tools

Once the objectives of the database design and the various steps in database design is known, it is essential to know the database design tools which are used to automate the task of designing a business system. Using automated design tools is the process of using a GUI tool to assist in the design of a database or database application. Many database design tools are available with a variety of features. The design tools are vendor-specific. CASE tools are software that provides automated support for some portion of the systems development process. Database drawing tools are used in enterprise modeling, conceptual data modeling, logical database design, and physical data modeling.

3.1 Need for Database Design Tool

The database design tools increase the overall productivity because the manual tasks are automated and less time is spent in performing tedious tasks and more time is spent in thinking about the actual design of the database. The quality of the end product is improved in using database design tools; because the design tool automates much of the design process as a result the time taken to design a database is reduced. As a result, more time is available to interview the customer, conduct user feedback sessions, and naturally the quality of the product is improved.

3.2 Desired Features of Database Design Tools

The database design tools should help the developer to complete the database model of database application in a timely fashion. Some of the features of the database design tools are given below:

- The database design tools should capture the user needs.
- The capability to model the flow of data in an organization.
- The database design tool should have the capability to model entities and their relationships.
- The database design tool should have the capability to generate Data Definition Language (DDL) to create database object.

- The database design tool should support full life cycle database support.
- Database and application version control.
- The database design tool should generate reports for documentation and user-feedback sessions.

3.3 Advantages of Database Design Tools

Some of the advantages of using database design tools for system design or application development are given as:

- The amount of code to be written is reduced as a result the database design time is reduced.
- Chances of errors because of manual work are reduced.
- Easy to convert the business model to working database model.
- Easy to ensure that all business requirements are met with.
- A higher quality, more accurate product is produced.

3.4 Disadvantages of Database Design Tools

Some of the disadvantages of database design tools are given below:

- More expenses involved for the tool itself.
- Developers might require special training to use the tool.

4. Redundancy and Data Anomaly

Redundant data means storing the same information more than once, i.e., redundant data could be removed without the loss of information. Redundancy can lead to anomalies. The different anomalies are insertion, deletion, and update anomalies.

4.1 Problems of Redundancy

Redundancy can cause problems during normal database operations. For example, when data are inserted into the database, the data must be duplicated wherever redundant versions of that data exist. Also, when the data are updated, all redundant data must be simultaneously updated to reflect that change.

4.2 Insertion, Deletion, and Update Anomaly

A table anomaly is a structure for which a normal database operation cannot be executed without information loss or full search of the data table. The table anomaly can be broadly classified into (1) Insertion Anomaly, (2) Deletion Anomaly, and (3) Update or Modification Anomaly.

Example 1

Staff no.	Job	Dept. no.	Dept. name	City
100	sales man	10	sales	Trichy
101	manager	20	accounts	Coimbatore
102	clerk	30	accounts	Chennai
103	clerk	30	operations	Chennai

Insertion Anomaly

We cannot insert a department without inserting a member of staff that works in that department.

Update Anomaly

Employee 519 is shown as having different addresses on different records.

Employees' Skills

Employee ID	Employee Address	Skill
426	87 Sycamore Grove	Typing
426	87 Sycamore Grove	Shorthand
519	94 Chestnut Street	Public Speaking
519	96 Walnut Avenue	Carpentry

Deletion Anomaly

By removing, employee 100, we have removed all information pertaining to the sales department.

Repeating Group

A repeating group is an attribute (or set of attributes) that can have more than one value for a primary key value. To understand the concept of repeating group, consider the example of the table STAFF. A staff can have more than one contact number. For each contact number, we have to store the data of the STAFF which leads to more storage space (more memory).

STAFF

Staff no.	Job	Dept. name	DeptID	City	Contact number
100	sales man	sales	01	Coimbatore	5434, 54221, 54241
101	manager	accounts	02	Chennai	56332, -----
102	clerk	accounts	03	Chennai	-----, -----, -----
103	clerk	operations	04	Chennai	-----, -----, -----

Repeating groups are not allowed in a relational design, since all attributes have to be atomic, i.e., there can only be one value per cell in a table.