SQL FOR DB2

Chapter 4
Introduction to Data Modeling and Database Design

Chapter Objectives

- Explain why logical data modeling and physical database design are necessary
- Explain the purpose of logical data modeling
- Explain the purpose of physical database design
- Explain the database models
- Describe the basics of a table
- Discuss the importance of integrity constraints

The Importance of Modeling and Design

- Before creating database tables, one should know exactly what kinds of tables need to be created
- Determining business requirements and documenting then, are tasks often referred to as “modeling”
- Logical data modeling or data modeling concentrates on the data requirements
- A data model says what the business needs, not how a system will be implemented
- Physical database design or database design concentrates on how to implement the data requirements discovered in data modeling

Logical Data Modeling

- A logical data model represents the data used in an organization and the relationships between the data.
- A logical data model specifies the following items of interest to the organization:
  - Entities
  - Events
  - Relationships
  - Attributes
  - Business rules
- The logical data model specifies the requirements for the database design

Physical Database Design

- Physical database design is the process of producing a description of the implementation of the database
- The physical design is concerned with:
  - Tables
  - File organizations
  - Indexes
  - Integrity constraints
  - Security
- Physical database design lays out how tables and other elements of the system will be implemented

Data Modeling vs Physical Design

Logical Data Modeling	Physical Database Design
Includes entities, attributes, and relationships	Includes tables, columns, keys, data types, validation rules, database triggers, stored procedures, and access constraints
Business names are used for attributes	Abbreviated column names are used that are limited by the database management system (DBMS) being used
Independent of the technology that will be used to implement the database.	Primary keys and indices are defined for fast data access.
Normalized to at least 3rd normal form	May be de-normalized to meet performance requirements.

Database Models

- Four major database models:
  - Hierarchical
  - Network
  - Relational
  - Object oriented
- Hierarchical model permits a dependent to have only one parent, just as in from of hierarchy, such as an organization chart
- Network model lets a row type be the dependent of any number of parent row types, as in which a course row type could be dependent of both an instructor row type and a student row type
- To relate information between different tables, the relational model uses duplicate data fields

The Basics of Tables

- Table, sometimes referred to as a file or entity is a collection of related rows
- Rows, sometimes referred to as a record or tuple is a collection of related columns
- Columns, sometimes referred to as a field or attribute is a group of storage positions reserved for a specific data element
- A database is an organized collection of related tables that can be joined to provide information to users
- Each row in a table contains one or more columns that uniquely identify that row in the table

Integrity Constraints

- Data Integrity
  - Data type and length
  - Null value acceptance
  - Allowable values (domain)
  - Default values
- Entity Integrity (Key Integrity)
  - Every table should have a primary key
  - The primary key should never be allowed to be null
- Referential Integrity
  - A foreign key value in a table must have a matching primary key value in a related table

Chapter Summary

- Before implementing database tables, the developer needs to know what the tables should contain and how they are related
- This information is developed by logical data modeling and physical database design
- A logical data model says what the business needs, not how a system will be implemented
- Physical database design concentrates on how to implement the data requirements discovered in data modeling
- A DBMS provides:
  - A data definition language
  - A data manipulation language

Chapter Summary

- In the relational model:
  - Tables represent entities
  - Each row in a table represents one instance of the table's entity
  - A column represents some entity attribute
  - Each row is identified by a primary key value
  - Rows in one table can be associated with rows in another table by foreign key values
- Integrity constraints:
  - Data integrity defines the values for a column
  - Entity defines a unique primary key
  - Referential integrity state that foreign key values must match some existing row's primary key value