Module 4 Database Design

Tuesday, 1 July 2025   7:05 PM

Module by Dhaanesh S

Objectives

➢ Learn what is database design and how to design a database schema
➢ Differentiate between different types of database schema's
➢ Relational database design
➢ How are tables related and what are types of keys used to maintain relationships and data integrity

Schema : Defines how data is related between tables within the database.

Database schema consists of

➢ Data relationships
➢ Unique object keys
➢ Name and data type of each column

Advantages

➢ Management
➢ Accessibility
➢ Security
➢ Ownership

Database schema design : The database schema is like a blueprint of a database. Before anyone can use a database to store and manipulate data, one should design a database schema first. This process of database schema design in called **data modelling**.

Database schema can be broadly divided into 3 categories:

➢ Conceptual or logical schema with entity relationship diagrams.
➢ Physical schema (That defines how data is stored in secondary storage)
➢ External (defines different user views)

Logical Schema - Its just a conceptual way of showing how a database design is, and physical storage and retrieval of data is hidden.

Here's an example of a logical schema.
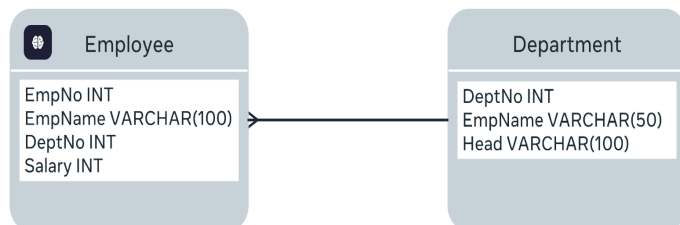
This depicts the employee and department entities in the database along with their attributes and how these two entities are related to each other. This is just a simple example and there'll be more entities in a real database.

Internal or physical schema - This shows how data is stored and what data is stored in database in the form of tables, columns and records

External schema - This schema limits the viewer to see only relevant database to its user. If a sales employee is working, he is only allowed to view sales database.

Conceptual/Logical Schema	Conceptual Level
Internal/Physical Schema	Internal Level
	Database
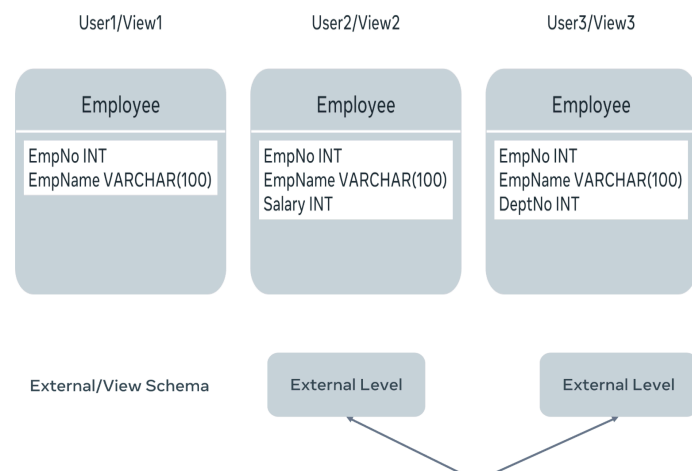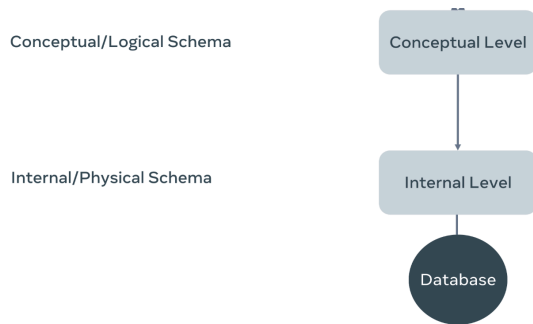
Database schema's advantages:

➤ They help to give clear view of how data is to be stored into tables with proper data types and manipulated.
➤ Prevents reverse engineering
➤ Write efficient queries to retrieve data for reporting purposes, analytics and so on.

Hands on: Build a schema for shopping cart database

CREATE TABLE products
(
    Product_id INT PRIMARY KEY,
    ..
)

RELATIONAL DATABASE DESIGN

Types of relationships between relational database model

1. One to one
2. One to many
3. Many to many

A relational model is built around 3 main concepts which are

➤ Data
➤ Relationships
➤ Constraints

Cardinality refers to how many records of information are present in the given table. Degree refers to no of columns.

PRIMARY KEY

What if we come across duplicates in a table. To solve those issues we define keys.

Candidate key

ID	Name	Date of Birth	Email	Grade
1	Jack	2000-03-02	jack.r@college.coursera	
2	Mary	2001-02-03	mary.g@college.coursera	
3	Dan	2001-02-03	dan.e@college.coursera	
4				

In this we can either define candidate key (primary key) as ID or email which are both unique. If we select ID as PK, then the other rejected column becomes secondary key (alternate key).

In the case if u cannot find a unique column to assign primary key because all values are duplicated, we can form a composite primary key (combination of two or more attributes).

FOREIGN KEY

	ID	Name	Date of Birth	Age	Email	Grade
	1	Jack Ray	2000-03-02	22	jack.r@college.coursera / jray@email.com	A
	2	Mary Gordon	2001-02-03	21	mary.g@college.coursera / marygo@email.com	B
	3	Dan Earls	2001-02-12	21	dan.e@college.coursera / dannye@email.com	A
	4	Mary Carter	1998-09-18	23	mary.c@college.coursera / maryar@email.com	C

Student (table title)

contained in the student ID column.

Attributes

➢ Single valued attribute
➢ Derived attribute
➢ Composite attribute
➢ Multivalued attribute
➢ Key attribute

DATABASE NORMALIZATION

What is database normalization ?

Normalization is an important process used in database systems. Its structures tables in the way that minimize challenges, by reducing data duplication avoiding data modification implications and helping to simplify data queries from database.

Database normalization challenges

1. **Insert Anomaly** : Insert anomaly occurs when new data is inserted into a table, which then requires addition of several other data.

   For example

Insert anomaly

Student ID	Student Name	Course name	Course Credit	Department Name	Department Director	Department Tel no.
01	Anita	Computer Science	160	Computing	Dr. Jones	+1 464-371
02	Pedro	Computer Science	160	Computing	Dr. Jones	+1 464-371
03	Nick	Ancient Civilizations	80	History	Dr. Shalvey	+1 464-371
04						+1 464-381

And I can't enroll new students without assigning each student and ID.

In this if we insert a new course under course name, we will have to enter a corresponding student name to tackle the empty column of our database, this way it causes insert anomaly problem.

2. **Update Anomaly** : Updating a record in a table requires further update in the columns of the database.

3. **Deletion Anomaly** : deletion of a particular record poses a problem on other columns, for example if a student rose is deleted, her record will be deleted but it still is depended by the other records in the table.

Different forms of Normalization

1. **1NF - First Normal Form** : This reduces the data duplication within the table. To fix multiple instances of data in the same column. Patient name and total cost columns.

   The first normal form uses data atomicity rule and eliminates the repeating data groups. It implies that you can only have one single instance value of column attribute in any table cell.

Patient ID	Patient name	Slot ID	Total Cost
P1	Rami	A1	1500
P2	Kim	A2	1200
P3	Nora	A3	1600
P4	Kamel	A1	2500
P5	Sami	A2	1000

However the patient table needs to be divided into patient table and slot table

P6	Norma	A5	2000
P7	Rose	A6	1000

However the patient table needs to be divided into patient table and slot table

Doctor ID	Doctor name	Region	Surgery Number	Surgery council	Postcode
D1	Karl	West London	3	Harrow	HA9SDE
D1	Karl	East London	4	Hackney	E1 6AW
D2	Mark	West London	4	Hackney	E1 6AW
D2	Mark	East London	5	Harrow	HA862E

The doctor table contains repeated no of instances under same column as in reign and council and name. Lets separate the above table into doctor table and surgery table.

Doctor table

Doctor ID	Doctor name
D1	Karl
D2	Mark

Surgery table

Surgery Number	Region	Surgery council	Postcode
3	West London	Harrow	HA9SDE
4	East London	Hackney	E1 6AW
5	West London	Harrow	HA862E

2. **2NF - Second Normal Form** : This form helps you to avoid partial dependency. Means the table with composite primary key. A non attribute key depends only on the on part of composite primary key. Here total cost depends only on the slot id. Ann patient name depends on patient id.

Patient ID	Patient name	Slot ID	Total Cost
P1	Rami	A1	1500
P2	Kim	A2	1200
P3	Nora	A3	1600
P4	Kamel	A1	2500
P5	Sami	A2	1000
P5	Sami	A3	1000
P6	Sami	A4	1500
P7	Norma	A5	2000
P8	Rose	A6	1000
P1	Rami	A7	1500

This can be solved by splitting above table into patient table and appointment table.

Patient table	
Patient ID	Patient name

Patient ID	Patient name
P1	Rami
P2	Kim
P3	Nora
P4	Kamel
P5	Sami
P7	Norma
P8	Rose

Appointment table before adding a new unique key

Appointment table

Slot ID	Total Cost
A1	1500
A2	1200
A3	1600
A1	2500
A2	1000
A3	1000
A4	1500
A5	2000
A6	1000
A7	1500

Appointment table after adding a new unique key

Appointment table

Appointment ID	Slot ID	Total Cost
1	A1	1500
2	A2	1200
3	A3	1600
4	A1	2500
5	A2	1000
6	A3	1000
7	A4	1500
8	A5	2000
9	A6	1000
10	A7	1500

In appointment table you need to add a unique key, like appointment id to make it as a primary key to identify unique records in the table.

3. **3NF - Third Normal Form** : For a table to be in 3NF form it should already be in 2NF form. It defines that no non-key attribute in a table should be dependent of any other non-key attribute. In the below table council is dependent on postcode. Explains np transitive dependency.

Surgery number	Region	Surgery council	Postcode
3	West London	Harrow	HA9SDE
4	East London	Hackney	E1 6AW
5	West London	Harrow	HA862E

Split the table into two, location table and council table.

Location table

Surgery number	Postcode
3	HA9SDE
4	E1 6AW
5	HA862E

Council table

Surgery council	Region
Harrow	West London
Hackney	East London

Hackney	East London

Unnormalized Form

Surgery: Doctor ID | Doctor name | Patient ID | Patient Name | Region | Council | Postcode | Surgery Number | Slot ID | Total Cost

First Normal Form

Patients: Patient ID | Patient Name | Slot ID | Total Cost

Surgery: Doctor ID | Doctor name | Region | Council | Postcode | Surgery Number

Doctors: Doctor ID | Doctor name

Surgery: Region | Council | Postcode | Surgery Number

Second Normal Form

Patients: Patient ID | Patient Name

Appointments: Appointment ID | Slot ID | Total Cost

Third Normal Form

Council: Council | Region

Surgery: Postcode | Surgery Number