

DATABASE MANAGEMENT SYSTEM

DATA

Factual information is known as data.

Types of data

- i) Structured: tables
- ii) Semi-structured: json file
- iii) Unstructured: media

DATABASE

- Organized collection of structured information.
- It reduces data redundancy.

NORMALIZATION

- It is a process that allows us to decompose database tables in such a way that both data dependency and update anomalies are minimized.
- There are various forms of normalization: 1NF, 2NF, 3NF, BCNF

DENORMALIZATION

Database optimization technique in which we add redundant data to one or more tables. This can help us avoid costly joins in a relational database.

KEYS

There are various types of keys:

- **Candidate Key:** The minimal set of attributes that can uniquely identify a tuple.
- **Primary Key:** There can be more than one candidate key in relation out of which one can be chosen as the primary key.
- **Super Key:** The set of attributes that can uniquely identify a tuple.
- **Alternate Key:** The candidate key other than the primary key.
- **Foreign Key:** A field in one table that refers to the primary key in another table.

ANOMALIES

- **Insertion Anomalies:** These anomalies occur when it is not possible to insert data into a database because the required fields are missing or because the data is incomplete.
- **Deletion anomalies:** These anomalies occur when deleting a record from a database and can result in the unintentional loss of data.
- **Update anomalies:** These anomalies occur when modifying data in a database and can result in inconsistencies or errors.

E-R DIAGRAM

- It is known as Entity Relationship Diagram as it explains the relationship among the entities present in the database.
- **Activity-** To make an ER diagram for ordering a book from online store, perform normalization.

DATA MODELLING

- Data modeling is a process of creating a conceptual representation of data objects and their relationships to one another.
- **Fact Table:** consists of facts of the system as it's content data.
- **Dimension Table:** comprises of the content of the fact table, which in turn helps build a connection between the respective fact table and the dimension table.

SCHEMA

- **Star Schema:** data is organized into a central fact table that contains the measures of interest, surrounded by dimension tables that describe the attributes of the measures.
- **Snowflake Schema:** dimension tables are normalized into multiple related tables, creating a hierarchical or “snowflake” structure.

Activity- Implement the concepts of Star and Snowflake schema by drawing respective diagrams.

SLOWLY CHANGING DIMENSIONS

Dimensions that consist of relatively static data that can change slowly but unexpectedly.

- **SCD 1:** Update the columns in the dimension row without preserving any change history.
- **SCD 2:** Preserve the change history in the dimension table and create a new row when there are changes.
- **SCD 3:** Update new value for selective columns in the existing rows. It maintains current and previous (or initial) value for some identified columns.