**29/08/2023**

**Data - Factual or piece of information**

* Semi structured over structured -> not fixed schema…flexible, space is also wasted in structured.
* Data mart -> single database -> localization advantage
* Database: structured, semi structured, unstructured

**Normalization:**

* Primary key, foreign key, candidate key, super key, unique key
* 1NF – no multivalued attribute
* 2NF – no partial dependency between non key and key attributes.
* 3NF – no transitive dependency between key and non-key attributes
* BCNF, 4NF, 5NFd
* Normalization removes update, insertion and delete anomalies.

**Denormalization:**

* Process of reversing the normalized tables.
* Denormalization is required when we need to query all the results easily.

**ER (Entity Relationship) model:**

* one to one
* one to many
* many to one
* many to many

**Case study on E-R diagram:**

* Created a E-R diagram for the bookstore database.
* Book, Author, Customer, Order, Publisher.
* The relationship is established between the tables using the primary keys and foreign keys.

**Case study on Normalization:**

* The order table from the above database has redundancy.
* It is first converted to 1NF by removing the multivalued attributes.
* Then to 2NF by removing the partial dependency.

**Dimensional Models:**

* Dimensional tables – will have all the business values columns.
* Fact tables – quantitative measures or important columns.
* 2 types – Star schema and snowflake schema
* Star schema – dimensions are surrounded by facts.
* Snowflake – centralized fact table is connected to multiple dimensions.

**Slowly Changing Dimensions**

* 3 types SCD-1, SCD-2, SCD-3
* **SCD-1** historic data is overridden, and it will be lost forever.
* Add the update data as a separate row in the table.
* In **SCD-2** we will add 3 columns to the existing table, effective start date, effective end date, and current flag.
* All the other data will have the end data infinite and the old data will have the data when it is updates and flag will be N.
* Mostly scd-3 is used to track the historic data.
* While writing the query add where condition flag =Y.
* In SCD-3 we add one extra column for the updated value. For example, if location is updated then column named new location is created which has the updated location.