DBMS, Big Data Fundamentals

Tuesday, August 29, 2023 9:12 AM

What is Data?

- Piece of Info
- Data needed to retain customers, to know capacity, targeted ads
- Types: Structured (excel, schema), Semi-structured (json, no schema), Unstructured (media)

What is Database?

- Tables, Views, etc.
- Primary Key: Unique and Not Null, need it to identify which id is being referred to
- Unique Key: Unique values, not repeated, can be null
- Foreign Key: Should be primary key in another table
- Candidate Key: Combination of two keys
- Super key: superset of all keys

Data Engineer- need to know how to develop an application, will deploy for testing or production.

Normalization -> Break down into multiple tables according to 1NF, 2NF, etc.
Renormalization -> Redoing the process to join the tables again

2Nf: should be 1NF and no partial dependency
3Nf: No transitive dependency
BCNf: non key attr should be individual candidate key, no dependency

Left join Vs Right join, when to use?
ER Diagram

1NF -> Each table has only 1 value, and column names are uni
2NF -> 1NF and no partial dependency (partially dependent o
non key is directly dependent on primary.
3NF -> all non-key attributes are independent to each other
BCNF -> all non-key attributes dependent on candidate keys

Dimensional Modeling

- Dimension table vs Fact table
- Only important features from dimension table will be a
- Dimension table has all data, but only imp qualitative c
- Star and Snowflake schema available in dimensional m
- Star -> dimension table in center
- Snowflake -> Fact table in center
- In snowflake there will be multiple links, whereas in sta
  table connecting all.

SCD: Slowly Changing Dimension

- SCD1, SCD2, SCD3
- SCD1 - Update table when changing data. Old data lost.
- SCD2 - Keeps old data. To keep old data and avoid data
  columns (from date, to date, flag) to the table. It is use
- SD3 - Keeps old data. To update data, add new column
  to be updated, with the value to be changed. SD3 takes
  almost doubles the amount of columns needed.