**Assignment – 1**

**What are OLTP and OLAP?**

**OLAP** (Online Analytical Processing) databases are typically structured as data warehouses, storing large amounts of both current and historical data. Due to the potential for structural changes in the data over time, OLAP databases commonly adopt a multi-dimensional schema. This multi-dimensional array of data abstraction is commonly known as the OLAP cube in computing terminology.

The critical aspect of a successful OLAP database implementation is its response time to complex queries. While slow queries may not directly impact transaction processing within applications, any delays in query results can significantly affect the accuracy of insights for data analysts and decision-makers relying on these intelligence insights.

The processing speed of OLAP databases is typically slower due to the nature of read-intensive, complex queries performed on extensive datasets. As OLAP databases are primarily used for analysis rather than transactional processing, the need for frequent data backups is minimized, as modifications to current data are infrequent.

**OLTP** (Online Transaction Processing) is a database system designed for fast, real-time tasks. It is commonly used in applications where multiple users access data simultaneously and quick response times are crucial. OLTP databases prioritize reliability and speed for operations like password resets and in-app purchases. They excel at handling insertions, updates, and deletions rather than storing massive datasets like OLAP.

In contrast to OLAP, OLTP queries are short-lived, typically lasting only milliseconds. They are less complex and involve processing a small number of records quickly. Because applications relying on OLTP are vital to organizations, any downtime affecting transaction processing can harm reputation.

To minimize downtime risks, OLTP databases require an aggressive backup strategy. Since OLTP systems constantly modify data, frequent backup snapshots or concurrent backups are necessary to maintain data integrity in case of disruptions.

**Differences between OLTP and OLAP**

Here are the key differences between OLTP and OLAP:

1. Process:

- OLTP: Manages database modifications through online transactional systems.

- OLAP: Performs online analysis and retrieves data.

2. Characteristic:

- OLTP: Characterized by many short online transactions.

- OLAP: Characterized by a large volume of data.

3. Functionality:

- OLTP: Modifies data in online databases.

- OLAP: Queries and manages data in a data warehouse.

4. Method:

- OLTP: Uses traditional DBMS.

- OLAP: Utilizes the data warehouse.

5. Query:

- OLTP: Performs Insert, Update, and Delete operations on the database.

- OLAP: Mostly involves select operations.

6. Table:

- OLTP: Tables are normalized.

- OLAP: Tables are not normalized.

7. Source:

- OLTP: Transactions in OLTP systems are the sources of data.

- OLAP: Various OLTP databases serve as the sources of data for OLAP.

**Database Normal Forms (5 Normal forms).**

Below is an overview of the various Normal Forms:

**1NF (First Normal Form):** Ensures that each column in a database table contains atomic values, eliminating repeating groups and ensuring each record is unique.

**2NF (Second Normal Form):** Eliminates redundant data by separating attributes into separate tables and requiring non-key attributes to be fully functional on the primary key.

**3NF (Third Normal Form):** Ensures that non-key attributes are not only fully functional on the primary key but also independent of each other, eliminating transitive dependency.

**4NF (Fourth Normal Form):** Addresses multi-valued dependencies, ensuring that there are no multiple independent multi-valued facts about an entity in a record.

**5NF (Fifth Normal Form):** Also known as "Projection-Join Normal Form" (PJNF), refers to the reconstruction of information from smaller, differently arranged data pieces.

**Dimension vs Fact Table and Types of Dimensions**

Dimension vs Fact Table:

1. Definition:

- Fact Table: Contains facts or measurements about a business process.

- Dimension Table: Stores descriptive characteristics used as query constraints.

2. Position:

- Fact Table: Positioned centrally within a snowflake or star schema.

- Dimension Table: Located at the edges of the snowflake or star schema, connected to the fact table.

3. Design:

- Fact Table: Defined by the grain or atomic level of data.

- Dimension Table: Designed to be extensive, detailed, and of high quality.

4. Task:

- Fact Table: Quantifies events for reporting and analysis, using data from dimension tables.

- Dimension Table: Provides background data about a company or entity.

5. Data Type:

- Fact Table: Contains data related to specific events or measures, such as sales by product and date.

- Dimension Table: Includes attributes that describe specific dimensions, like product ID or category.

**Types of Dimensions:**

1. Conformed Dimension

2. Degenerate Dimension

3. Junk Dimension

4. Role-Playing Dimension

5. Slowly Changing Dimension (SCD)

6. Rapidly Changing Dimension (RCD)

7. Fixed Dimension

**Snowflake Vs Star Schema**

**Star Schema:**

- Hierarchies for dimensions are stored in the dimensional table.

- Contains a fact table surrounded by dimension tables.

- Only a single join creates the relationship between the fact table and any dimension tables.

- Simple database design.

- Denormalized data structure.

- High level of data redundancy.

- Single dimension table contains aggregated data.

- Cube processing is faster.

- Offers higher performing queries using Star Join Query Optimization.

- Tables may relate to multiple dimensions.

**Snowflake Schema:**

- Hierarchies are divided into separate tables.

- One fact table surrounded by dimension tables, which are in turn surrounded by dimension tables.

- Requires many joins to fetch the data.

- Very complex database design.

- Normalized data structure.

- Very low-level data redundancy.

- Data split into different dimension tables.

- Cube processing might be slow because of the complex join.

- The snowflake schema is represented by a centralized fact table, which is unlikely connected with multiple dimensions.