**Class 1 (03/21/2024) Assignment**

1. What is OLTP and OLAP

Ans) **OLTP (Online Transaction Processing):** It is an operational system which supports transaction- oriented applications in a 3-tier architecture. It is focused on maintaining data integrity, query processing in multiple environments. The architecture of OLTP contains Business/enterprise Strategy, Business process, Customers, orders, products, ETL processes, Data Mart, Data mining, Analytics and decision making.

Important characteristics of OLTP is:

* It uses small amount of data for transactions.
* Databases are directly accessible to end-users.
* It supports complex data models and tables.

Examples of the OLTP system is the ATM centre, Online banking, adding book to shopping cart.

**OLAP (Online Analytical Processing):** It allows users to analyse information from multiple databases at same time. These OLAPoperations in Data miningare resource intensive. The OLAP cube is a data structure optimized for very quick data analysis. OLAP cube is also called the hypercube.

A data warehouse would information from multiple data sources and formats like text files, excel sheet, multimedia files etc… The extracted data is cleaned and transformed. Then data is loaded into an OLAP server where information is pre-calculated in advanced analysis.

Basic analytical Operations are: **Roll-up, drill down, Slice and Dice, Pivot (Rotate).**

Types of OLAP are: **ROLAP, MOLAP, HOLAP**….

1. What is difference between OLTP and OLAP

Ans)

|  |  |
| --- | --- |
| **OLTP**   * OLTP is an online transactional system. * It is characterized by large numbers of short online transactions. * OLTP uses traditional DBMS. * It is a market-orientated process. * It is easy to create and maintain. * It allows Read/ Write Operations. * Transaction throughput is the performance metric. * Queries in this process are standardized and simple. * OLTP is designed to have fast response time, low data redundancy, and is normalized. | **OLAP**   * OLAP is an online analysis and data retrieving process. * It is characterized by a large volume of data * OLAP uses the data warehouse. * It is a customer orientated process. * The user creates a view with the help of a spreadsheet. * Only read and sometimes Write. * Query throughput is the performance metric. * Complex queries involving aggregations. * A data warehouse is created uniquely, it can integrate different data sources for building a consolidated database |

1. Database Normal forms

Ans) **Normalization** is a database design technique that reduces data redundancy and eliminates undesirable characteristics like Insertion, Update and Deletion Anomalies. Normalization rules divides larger tables into smaller tables and links them using relationships. Normalization rules divides larger tables into smaller tables and links them using relationships.

Types of Normal Forms in DBMS

* **1NF (First Normal Form):** Ensures that the database table is organized such that each column contains atomic (indivisible) values, and each record is unique. This eliminates repeating groups, thereby structuring data into tables and columns.
* **2NF (Second Normal Form):** Builds on 1NF by We need to remove redundant data from a table that is being applied to multiple rows. and placing them in separate tables. It requires all non-key attributes to be fully functional on the primary key.
* **3NF (Third Normal Form):** Extends 2NF by ensuring that all non-key attributes are not only fully functional on the primary key but also independent of each other. This eliminates transitive dependency.
* **4NF (Fourth Normal Form):** Addresses multi-valued dependencies. It ensures 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), It pertains to the reconstruction of information from smaller, differently arranged data pieces.
* **BCNF (Boyce-Codd Normal Form):** A refinement of 3NF that addresses anomalies not handled by 3NF. It requires every determinant to be a candidate key, ensuring even stricter adherence to normalization rules.

1. What is fact and dimension table and difference between them?

**Dimension Table:**

Think of a dimension as something you want to analyse or categorize your data by.

Dimension tables contain descriptive attributes that provide context to your data.

They are typically narrower but can have many rows.

Each row in a dimension table represents a unique value of that dimension.

**Fact Table:**

Fact tables contain quantitative data, often referred to as measures or metrics.

They usually have many rows but fewer columns compared to dimension tables.

Fact tables contain keys that reference the dimension tables, creating relationships between them.

These keys act as foreign keys that connect the fact table to the dimension tables.

**Types of Dimensions:**

**Conformed Dimension:** A dimension that has the same meaning and content when referred from different fact tables. For example, a "Time" dimension could be used in both sales and inventory fact tables.

**Junk Dimension:** A dimension that contains attributes that don't fit well in the existing dimensions, so they are grouped together. For example, flags or indicators.

**Degenerate Dimension:** A dimension that is derived from the fact table itself rather than a separate dimension table. For example, an invoice number or transaction ID.

**Slowly Changing Dimension (SCD): A** dimension that changes slowly over time. There are different types of SCDs **(Type 1, Type 2, Type 3)** which determine how historical data is managed when changes occur in the dimension.

1. What is difference between snowflake and star schema?

**Star Schema:**

Star schema is a type of data warehouse schema where there is one central "fact" table surrounded by multiple "dimension" tables. The fact table contains the primary metrics or measures of interest, such as sales revenue or quantity sold.

Dimension tables contain descriptive attributes that provide context to the measures in the fact table, such as product, customer, time, etc. The fact table is at the centre, with direct relationships to each dimension table. This creates a star-like structure, hence the name "star schema."

Star schema is simple and easy to understand, making it efficient for querying and reporting.

**Snowflake Schema:**

Snowflake schema is a more normalized form of a data warehouse schema compared to the star schema. In a snowflake schema, dimension tables are normalized into multiple related tables, rather than being denormalized into a single table as in the star schema.

This normalization reduces data redundancy and can save storage space, but it can make querying more complex due to the need for additional joins. Snowflake schema is useful when dealing with large and complex datasets where space optimization is critical, and when updates to dimension data are frequent.