**Class 1: Assignment**

1. **What is OLTP and OLAP**

Ans) **OLTP (Online Transaction Processing):** Applications that are focused on transactions are managed by OLTP systems.  
They are designed to handle a high volume of brief online transactions, like the real-time insertion, updating, or deletion of small bits of data.  
These systems are commonly employed in settings like retail sales, financial transactions, or airline reservations where concurrency and responsiveness are critical.  
Example: An internet banking system, a retail point-of-sale system, or an ATM that handles withdrawal and deposit transactions

**OLAP (Online Analytical Processing):** OLAP systems are made to analyse and query massive amounts of data rapidly.  
They work best with intricate queries that need to be aggregated and calculated across huge datasets.  
With the help of multidimensional views offered by OLAP systems, users can dig down, slice, and pivot data to obtain new insights.  
Data mining, decision support systems, and business intelligence frequently use these systems.  
A sales analysis system, for instance, would enable managers to assess sales results based on a range of factors, including time, product, geography, or client segment. A system for financial reporting that analyses firm financial data across time periods and business units could serve as another illustration.

**2)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 |

**3)Database Normal forms?**

Ans) Normalization is a crucial database design technique aimed at reducing data redundancy and eliminating undesirable characteristics like insertion, update, and deletion anomalies. Here's a brief overview of the types of normal forms in database management systems (DBMS):

\*\*1NF (First Normal Form): Ensures that each column contains atomic (indivisible) values, and each record is unique, eliminating repeating groups.

\*\*2NF (Second Normal Form): Requires removal of redundant data by placing them in separate tables. It ensures all non-key attributes are fully functional on the primary key.

\*\*3NF (Third Normal Form): Extends 2NF by ensuring 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 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 focuses on reconstructing 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 stricter adherence to normalization rules.

**4)What is fact and dimension table and difference between them?**

A) \*\*Fact Table\*\*: Contains quantitative data or metrics for a specific business process or event, typically at a detailed level. It represents the "what, where, when, and how many" aspects of the data. Fact tables are often large and are at the center of a star or snowflake schema.

- \*\*Dimension Table\*\*: Contains descriptive attributes that provide context to the data in the fact table. It represents the "who, what, when, where, and why" aspects of the data. Dimension tables are usually smaller and are joined to fact tables through foreign key relationships. They are denormalized to some extent to facilitate easier querying and analysis.

**Differences**

1.     **Content**:

* + Fact tables contain quantitative data or measures.
  + Dimension tables contain descriptive attributes.

2.     **Size and Granularity**:

* + Fact tables are typically larger and store detailed, granular data.
  + Dimension tables are usually smaller and contain aggregated, descriptive data.

3.     **Relationship**:

* + Fact tables are related to dimension tables through foreign key relationships.
  + Dimension tables provide context to the measures stored in the fact table.

4.     **Purpose**:

* + Fact tables primarily store numeric data that can be aggregated for analysis.
  + Dimension tables provide descriptive context for the measures and facilitate slicing and dicing of data for analysis.

**Types Of Dimensions:**

1. Slowly Changing Dimensions
2. Junk Dimension
3. Degenerate Dimension
4. Shrunken Dimension
5. Role Play Dimension
6. Static Dimension
7. Rapid Or Fast Changing Dimension

**5)What is difference between snowflake and star schema?**

A) **Star Schema:**

- Simple, star-like structure with one central fact table and multiple directly connected dimension tables.

- Less normalized, suitable for simpler data warehousing needs.

- Easier to understand and query, offering better performance for ad-hoc queries.

**Snowflake Schema:**

- More normalized structure with dimension tables broken down into multiple related tables.

- Suitable for complex data models with many-to-many relationships.

- More complex to understand and query due to increased normalization, may require more optimization techniques for performance.

In essence, star schemas are simpler and more denormalized, whereas snowflake schemas are more normalized and suitable for complex data models.