**AWS EXERCISE – 1**

**1)OLTP:** Online Transaction Processing is referred to as OLTP. It is a kind of system that controls transaction-oriented applications, usually for real-time online business and financial transactions. OLTP systems are made to manage a high volume of quick transactions, like adding, changing, or removing entries from databases.   
 **Key features of OLTP systems include:**

* Transactional Integrity: OLTP systems maintain transactional integrity by abiding by the ACID (Atomicity, Consistency, Isolation, Durability) properties, which ensure error-free and dependable transaction processing.
* Concurrent Access: Multiple users can access OLTP systems concurrently, enabling the processing of numerous transactions at once without interfering with one another.
* Quick Response Time: OLTP systems are designed to respond quickly, usually in milliseconds, enabling users to complete transactions effectively.
* Online Interaction: OLTP systems are made to be accessed online, enabling users to communicate with the system in real time via a network, like the

**OLAP:** Online Analytical Processing is known as OLAP. This kind of technology is employed in the analysis and querying of multidimensional data from various angles. OLAP systems give users flexible and interactive tools to explore vast amounts of data, making complex data analysis, reporting, and decision-making tasks easier.  
 **OLAP's salient characteristics include:**

* Multidimensional Analysis: OLAP systems enable users to analyze data along multiple dimensions, including time, geography, product, or customer. They do this by organizing data into multidimensional structures, such as cubes or hypercubes.
* Aggregation: By allowing users to view summarized data or drill down to detailed information as needed, OLAP systems support the aggregation of data at various levels of granularity.
* Interactive Querying: Ad hoc queries, data slicing and dicing, and sophisticated analytical operations can be carried out in real-time by users of OLAP systems thanks to their interactive query capabilities.

**Difference between OLTP & OLAP:** The following are the differences between OLTP and OLAP  
 **OLTP:**

* It is an online transactional system. It manages database modification.
* It is characterized by large numbers of short online transactions.
* OLTP is an online database modifying system.

**OLAP:**

* OLAP is an online analysis and data retrieving process.
* It is characterized by a large volume of data.
* OLAP is an online database query management system.

**2)DATA BASE Normal forms:** The process of arranging a relational database's attributes and tables to reduce dependencies and redundancies is known as database normalization. Using a set of guidelines known as normal forms, the normalization process makes sure that the database schema is efficiently organized and devoid of anomalies. Normal forms that are most frequently used are:

1. First Normal Form (1NF)
2. Second Normal Form (2NF)
3. Third Normal Form (3NF)
4. Boyce-Codd Normal Form (BCNF)
5. Fourth Normal Form (4NF)
6. Fifth Normal Form (5NF)
7. Domain-Key Normal Form (DKNF)
8. Sixth Normal Form (6NF)

**First Normal Form, or 1NF:**  
verifies that a table's attributes only contain atomic, or indivisible, values and that no repeating groups of attributes exist. Divide a column containing multiple values into separate columns as an example.

**Second Normal Form, or Form 2NF:**  
requires the database to be in 1NF and all non-key attributes to be fully functionally dependent on the entire primary key. eliminates partial dependencies, which occur when a feature needs only a portion of the main key to work.   
For example, think about creating separate tables for attributes whose functionality depends on part of the primary key.

**Three-Normal Form (3NF):**  
requires the database to be in 2NF and that all non-key attributes be transitively dependent on the primary key. eliminates transitive functional dependencies between attributes that rely on non-key attributes.  
For example, think about creating separate tables for attributes that are dependent on other non-key attributes.

Higher normal forms, such as Boyce-Codd Normal Form (BCNF), Fourth Normal Form (4NF), and Fifth Normal Form (5NF), go beyond these fundamental normal forms and further refine the normalization process to handle more intricate dependencies and anomalies.

**3) what is Dimension and Fact table:** Fact tables and dimensions are two essential tools used to store and organize data in data warehousing and business intelligence.  
**Dimension:**   
 A dimension table contains descriptive features or attributes that provide classification and context for the data in a data warehouse.  
Among the common data kept in dimension tables are entities, goods, locations, timings, and other reference information.  
Each row in a dimension table represents a unique instance of the entity being described, and each column indicates an attribute or characteristic of that entity.  
Analytical queries employ dimension tables to perform data filtering, grouping, and aggregation. Oftentimes, they are smaller than fact tables.

**Fact table:**Fact tables are collections of numerical and quantitative data that depict events, activities, or business transactions. They are also sometimes referred to as facts or measures.  
Fact tables are frequently employed to store data in an all-inclusive fashion, covering the who, what, where, when, and how much of business operations.  
Fact tables often contain foreign keys that link to dimension tables, providing the background information required to comprehend the facts.  
Fact tables may also include summarized or aggregated data, also known as aggregated facts, to aid in higher-level analysis and reporting.  
Fact tables are usually much larger than dimension tables and are made for query performance and analytical processing.

**conclusion:** fact tables hold the quantitative data that reflects business activities, whereas dimension tables offer a descriptive context for the data in a data warehouse. A dimensional model, a popular schema used in data warehousing for organizing and structuring data for analytical purposes, is primarily composed of dimension and fact tables.  
  
**4) Types of Dimensions:** Data warehousing dimensions are commonly classified into different types according to the type of data they represent and the analytical viewpoints they offer. Among the widely acknowledged categories of dimensions are:

* **conformed dimension:**

A conformed dimension is one that is shared by multiple fact tables within a single data warehouse or data mart Conformed dimensions ensure uniformity and consistency in reporting and analysis by providing a single, consistent view of a particular business aspect, such as customers, products, or time.

* **Junk dimension**:

A dimension that does not fit well into an existing dimension and has low cardinality is considered junk. It is sometimes called a utility dimension or a garbage dimension.  
Junk dimensions are typically made up of indicators, flags, or other categorical features that are useful for filtering or grouping data but don't have any intrinsic meaning.  
By merging these features into a single junk dimension, the data warehouse's number of dimension tables can be reduced, streamlining the schema, and improving query performance.

* **role-playing dimension:**

A role-playing dimension is one that is used in multiple roles or scenarios within a single data model.  
Using role-playing dimensions, the same dimension table can be joined to a fact table more than once, each time indicating a different perspective or relationship.  
Typical examples of role-playing dimensions are time dimensions (such as order date, ship date, and delivery date) and geographic dimensions (such as billing location, shipping location).

* **Slowly Changing Dimension:**

A dimension that changes slowly is one whose properties change with time and require special handling to maintain historical accuracy.  
Slowly changing dimensions are divided into three types: SCD Type 1 (overwrite), SCD Type 2 (add new row), and SCD Type 3 (add new attribute), depending on how they handle changes to attribute values.  
By capturing previous changes to dimension attributes, gradually changing the dimensions enables trend analysis, historical reporting, and the analysis of data at different points in time.

* **degenerate dimension:**  
  A degenerate dimension is one that, instead of being a separate dimension table, contains one or more attributes that are stored in the fact table itself.  
  Degenerate dimensions are typically used to represent operational or transactional data that does not fit well into traditional dimension tables.  
  Typical examples of degenerate dimensions are transaction numbers, order numbers, invoice numbers, and other unique identifiers that function as keys to identify particular transactions.

These dimensions offer scalability and flexibility in the design of data marts and warehouses to accommodate various business needs and analytical requirements.

5) **Snowflake vs Star Schema:** There are two popular methods for creating data warehouse schemas: snowflake and star schema. Although they are appropriate for different scenarios and have different characteristics, both schemas have the same function of efficiently organizing and structuring data for querying and analysis.

**Snowflake Schema:** A snowflake schema is a type of database schema used in data warehousing and business intelligence environments. It organizes data into a structure resembling a snowflake, with a central fact table surrounded by multiple dimension tables. In data warehousing environments, where data integrity, flexibility, and storage efficiency are crucial factors, snowflake schemas are frequently utilized. They work especially well with intricate data models that have several dimensions and hierarchies. Nevertheless, several variables, including query patterns, data type, performance needs, and maintenance considerations, influence the decision between a snowflake schema and alternative schema designs.

**Star Schema:** One kind of database schema that's used in business intelligence and data warehousing settings is the star schema. It arranges data into a star-shaped structure, with several dimension tables encircling a central fact table. A star schema's primary attributes are as follows:  
Star schemas are commonly employed in data warehousing settings where scalability, query performance, and ease of use are crucial factors. They work especially well in situations where analytical reporting and querying are prioritized over transactional processing. The selection between a star schema and alternative schema designs is contingent upon various factors, including but not limited to the data's characteristics.