### **AWS Assignment 1**

### **Q1) What is OLAP and OLTP? Ans) a) Online Analytical Processing(OLAP):** OLAP stands for Online Analytical Processing. OLAP systems have the capability to analyze database information of multiple systems at the current time. The primary goal of OLAP Service is data analysis and not data processing.  **Benefits of OLAP Services**

* OLAP services help in keeping consistency and calculation.
* We can store planning, analysis, and budgeting for business analytics within one platform.
* OLAP services help in handling large volumes of data, which helps in enterprise-level business applications.
* OLAP services help in applying security restrictions for data protection.
* OLAP services provide a multidimensional view of data, which helps in applying operations on data in various ways.

**Drawbacks of OLAP Services**

* OLAP Services requires professionals to handle the data because of its complex modeling procedure.
* OLAP services are expensive to implement and maintain in cases when datasets are large.
* We can perform an analysis of data only after extraction and transformation of data in the case of OLAP which delays the system.
* OLAP services are not efficient for decision-making, as it is updated on a periodic basis.

### **b) Online Transaction Processing(OLTP):**  OLTP has the work to administer day-to-day transactions in any organization. The main goal of OLTP is data processing not data analysis. **Benefits of OLTP Services**

* OLTP services allow users to read, write and delete data operations quickly.
* OLTP services help in increasing users and transactions which helps in real-time access to data.
* OLTP services help to provide better security by applying multiple security features.
* OLTP services help in making better decision making by providing accurate data or current data.
* OLTP Services provide Data Integrity, Consistency, and High Availability to the data.

### **Drawbacks of OLTP Services**

* OLTP has limited analysis capability as they are not capable of intending complex analysis or reporting.
* OLTP has high maintenance costs because of frequent maintenance, backups, and recovery.
* OLTP Services get hampered in the case whenever there is a hardware failure which leads to the failure of online transactions.
* OLTP Services many times experience issues such as duplicate or inconsistent data.

**Q2) What are the differences between OLTP and OLAP?**

| **Category** | **OLAP (Online Analytical Processing)** | **OLTP (Online Transaction Processing)** |
| --- | --- | --- |
| Definition | It is well-known as an online database query management system. | It is well-known as an online database modifying system. |
| Data source | Consists of historical data from various Databases. | Consists of only operational current data. |
| Method used | It makes use of a data warehouse. | It makes use of a standard Database Management System (DBMS). |
| Application | It is subject-oriented. Used for Data Mining, Analytics, Decisions making, etc. | It is application oriented. Used for business tasks. |
| Normalized | In an OLAP database, tables are not normalized. | In an OLTP database, tables are normalized(3NF). |
| Usage of data | The data is used in planning, problem-solving, and decision-making. | The data is used to perform day-to-day fundamental operations. |
| Task | It provides a multi-dimensional view of different business tasks. | It reveals a snapshot of present business tasks. |
| Purpose | It serves the purpose to extract information for analysis and decision-making. | It serves the purpose to Insert, Update, and Delete information from the database. |
| Volume of data | A large amount of data is stored typically in TB, PB | The size of the data is relatively small as the historical data is archived in MB, and GB. |
| Queries | Relatively slow as the amount of data involved is large. Queries may take hours. | Very Fast as the queries operate on 5% of the data. |
| Update | The OLAP database is not often updated. As a result, data integrity is unaffected. | The data integrity constraint must be maintained in an OLTP database. |
| Backup and Recovery | It only needs backup from time to time as compared to OLTP. | The backup and recovery process is maintained rigorously |
| Processing time | The processing of complex queries can take a lengthy time. | It is comparatively fast in processing because of simple and straightforward queries. |
| Types of users | This data is generally managed by CEO, MD, and GM. | This data is managed by clerksForex and managers. |
| Operations | Only read and rarely write operations. | Both read and write operations. |
| Updates | With lengthy, scheduled batch operations, data is refreshed on a regular basis. | The user initiates data updates, which are brief and quick. |
| Nature of audience | The process is focused on the customer. | The process is focused on the market. |
| Database Design | Design with a focus on the subject. | Design that is focused on the application. |
| Productivity | Improves the efficiency of business analysts. | Enhances the user’s productivity. |

### **Q3) What are all the DB Normal Forms? Ans) Normalization** is the process of minimizing **redundancy** from a relation or set of relations. **Normal forms** are used to eliminate or reduce redundancy in database tables. Important Points Regarding Normal Forms in DBMS

* **First Normal Form (1NF):** This is the most basic level of normalization. In 1NF, each table cell should contain only a single value, and each column should have a unique name. The first normal form helps to eliminate duplicate data and simplify queries.
* **Second Normal Form (2NF):** 2NF eliminates redundant data by requiring that each non-key attribute be dependent on the primary key. This means that each column should be directly related to the primary key, and not to other columns.
* **Third Normal Form (3NF):** 3NF builds on 2NF by requiring that all non-key attributes are independent of each other. This means that each column should be directly related to the primary key, and not to any other columns in the same table.
* **Boyce-Codd Normal Form (BCNF):** BCNF is a stricter form of 3NF that ensures that each determinant in a table is a candidate key. In other words, BCNF ensures that each non-key attribute is dependent only on the candidate key.
* **Fourth Normal Form (4NF):** 4NF is a further refinement of BCNF that ensures that a table does not contain any multi-valued dependencies.
* **Fifth Normal Form (5NF):** 5NF is the highest level of normalization and involves decomposing a table into smaller tables to remove data redundancy and improve data integrity.

**Q4)a) what are the types of dimension table ?  
 b) what are the differences between Dimension table and Fact table?  
Ans)  
a) Types of Dimension Table:  
Dimension table:** This is a table in a star schema of a data warehouse. Data warehouses are built using dimensional data models which consist of fact and dimension tables. The latter is used to describe dimensions. They contain dimension keys, values, and attributes. **\*) Slowly Changing Dimensions:** Dimension attributes that change slowly over a period rather than changing regularly is grouped as SCDs.   
Type 1 is to over write the old value, Type 2 is to add a new row and Type 3 is to create a new column.  
**Type 1:  
 Advantage:** It is very easy to follow and it results  in huge space savings and hence cost savings.   
 **Disadvantage:** It is that no history is maintained.  
**Type 2:**   
 **Advantage:** It is that the complete history is maintained.   
 **Disadvantage:** The only disadvantage lies in the huge space allocation because the entire history right from the start has to be maintained.  
Type 3:  
 **Advantage:** The best approach could be to add a new column where you add two new columns. In this case keeping a tracking of the history becomes very easy.  
  
**\*) Conformed Dimension:**This is used in multiple locations**.**It helps in creating consistency so that the same can be maintained across the fact tables. Different tables can use the table across the fact table and it can help in creating different reports.  
  
**\*) Degenerate Dimension**: A degenerate dimension is when the dimension attribute is stored as part of the fact table and not in a separate table.   
  
**\*) Junk Dimension**: It is a single table with a combination of different and unrelated attributes to avoid having many foreign keys in the fact table. They are often created to manage the foreign keys created by rapidly changing dimensions.  
  
**\*) Role play dimension**: It is a table that has multiple valid relationships with a fact table. For example, a fact table may include foreign keys for both ship date and delivery date. But the same attributes apply to each foreign key so the same tables can be joined to the foreign keys.  
  
**b)** The **Fact Table or Reality Table** helps the user to investigate the business dimensions that helps him in call taking to enhance his business.

On the opposite hand, **Dimension Tables** facilitate the reality table or fact table to gather dimensions on that the measures needs to be taken.

| **.NO** | **Fact Table** | **Dimension Table** |
| --- | --- | --- |
| 1. | Fact table contains the measuring of the attributes of a dimension table. | Dimension table contains the attributes on that truth table calculates the metric. |
| 2. | In fact table, There is less attributes than dimension table. | While in dimension table, There is more attributes than fact table. |
| 3. | In fact table, There is more records than dimension table. | While in dimension table, There is less records than fact table. |
| 4. | Fact table forms a vertical table. | While dimension table forms a horizontal table. |
| 5. | The attribute format of fact table is in numerical format and text format. | While the attribute format of dimension table is in text format. |
| 6. | It comes after dimension table. | While it comes before fact table. |
| 7. | The number of fact table is less than dimension table in a schema. | While the number of dimension is more than fact table in a schema. |
| 8. | It is used for analysis purpose and decision making. | While the main task of dimension table is to store the information about a business and its process. |

**Q5) What is star schema and snowflake schema? What are the differences between them?  
Ans)  
Star Schema:** Star schema is the type of multidimensional model which is used for data warehouse. In star schema, The fact tables and the dimension tables are contained. In this schema fewer foreign-key join is used. This schema forms a star with fact table and dimension tables.   
  
**Star Schema:** Star schema is the type of multidimensional model which is used for data warehouse. In star schema, The fact tables and the dimension tables are contained. In this schema fewer foreign-key join is used. This schema forms a star with fact table and dimension tables. 

| **S.NO** | **Star Schema** | **Snowflake Schema** |
| --- | --- | --- |
| 1. | In Star Schema, the fact tables and the dimension tables are contained. | While in snowflake Schema, the fact tables, dimension tables as well as sub dimension tables are contained. |
| 2. | Star schema is a top-down model. | While it is a bottom-up model. |
| 3. | Star schema uses more space. | While it uses less space. |
| 4. | It takes less time for the execution of queries. | While it takes more time than star schema for the execution of queries. |
| 5. | In star schema, Normalization is not used. | While in this, both normalization and denormalization are used. |
| 6. | It’s design is very simple. | While it’s design is complex. |
| 7. | The query complexity of star schema is low. | While the query complexity of snowflake schema is higher than star schema. |
| 8. | It’s understanding is very simple. | While it’s understanding is difficult. |
| 9. | It has less number of foreign keys. | While it has more number of foreign keys. |
| 10. | It has high data redundancy. | While it has low data redundancy. |