

## **WORKSHEET 1 SQL**

**Q1. Which of the following is/are DDL commands in SQL?**

A. Create and D. Alter

**Q2. Which of the following is/are DML commands in SQL?**

A) Update B) Delete C) Select

**Q3. Full form of SQL is:**

B) Structured Query Language

**Q4. Full form of DDL is:**

B) Data Definition Language

**Q5. DML is:**

A) Data Manipulation Language

**Q6. Which of the following statements can be used to create a table with column B int type and C float type?**

C) Create Table A (B int,C float)

**Q7. Which of the following statements can be used to add a column D (float type) to the table A created above?**

B) Alter Table A ADD COLUMN D float

**Q8. Which of the following statements can be used to drop the column added in the above question?**

B) Alter Table A Drop Column D

**Q9. Which of the following statements can be used to change the data type (from float to int ) of the column D of table A created in above questions?**

B) Alter Table A Alter Column D int

**Q10. Suppose we want to make Column B of Table A as primary key of the table. By which of the following statements we can do it?**

A) Alter Table A Add Constraint Primary Key B

**Q 11. What is data-warehouse?**

- A data warehouse is a centralized storage system which allows for the storing, analyzing, and interpreting of data which gathers raw data from multiple sources into a central repository, structured using predefined schemas designed for data analytics.

- It is a relational database that is designed for query and analysis of data which further enable and support business intelligence decisions.
- A data warehouse may contain multiple databases. Within each database, data is organized into tables and columns. Within each column, the description of the data can be defined, such as integer, data field, or string. Tables can be organized inside of schemas, such as folders. When data is ingested, it is stored in various tables described by the schema. Query tools use the schema to determine which data tables to access and analyze.
- A data warehouse architecture is made up of tiers. The top tier is the front-end client that presents results through reporting, analysis, and data mining tools. The middle tier consists of the analytics engine that is used to access and analyze the data. The bottom tier of the architecture is the database server, where data is loaded and stored.

#### **Q12. What is the difference between OLTP VS OLAP?**

- **OLTP stands for Online Transaction Processing.** It deals with transaction-related operations such as query processing, data integrity, efficiency, and effectiveness. it enables the real-time execution of large numbers of database transactions
- Relational databases serve as the organized and detailed source for Online Transaction Processing. The atomicity of the OLTP system ensures that either the transaction succeeds completely, or the transaction is completely canceled, thus maintaining transaction integrity.
- OLTP queries are simple and straightforward (insert, update, delete), however, they require little time in processing, and they also take up less space. In order to maintain data integrity, OLTP needs frequent incremental backups.
- It is a market-driven system that includes applications such as online banking, ticket booking, storing information, etc. Furthermore, it assists in managing the day-to-day transaction-oriented operations of an organization. Due to the timely modification of all transactions, OLTP provides a solid foundation for a stable business.
- ❖ **OLAP stands for Online Analytical Processing** and is used to handle data analysis and retrieval. It is a customer-driven data analysis system that helps organizations make more informed decisions by analyzing business data.
- ❖ OLAP is based on a data warehouse method and uses complicated queries to find insights and solve specific business problems. In OLAP, multidimensional data can be analyzed logically and effectively by using OLAP cubes which store and analyze multidimensional data in an orderly fashion. It is a powerful technology for discovering data. It ensures data is secure by implementing appropriate security measures. It is commonly used for financial reporting tools, and forecasting as well as for business intelligence tools and data mining.

- ❖ OLAP tools are designed for multidimensional analysis of data in a data warehouse, which contains both historical and transactional data. Common uses of OLAP include data mining and other business intelligence applications, complex analytical calculations, and predictive scenarios, as well as business reporting functions like financial analysis, budgeting, and forecast planning.

### Q13. What are the various characteristics of data-warehouse?

#### Characteristics of Data Warehouse:

- **Subject-oriented**

A data warehouse is always a subject oriented as it delivers information about a theme instead of organization's current operations. It can be achieved on specific theme. That means the data warehousing process is proposed to handle with a specific theme which is more defined. These themes can be sales, distributions, marketing etc.

A data warehouse never put emphasis only current operations. Instead, it focuses on demonstrating and analysis of data to make various decision. It also delivers an easy and precise demonstration around particular theme by eliminating data which is not required to make the decisions.

- **Integrated**

It is somewhere same as subject orientation which is made in a reliable format. Integration means founding a shared entity to scale the all similar data from the different databases. The data also required to be resided into various data warehouse in shared and generally granted manner.

A data warehouse is built by integrating data from various sources of data such that a mainframe and a relational database. In addition, it must have reliable naming conventions, format and codes. Integration of data warehouse benefits in effective analysis of data. Reliability in naming conventions, column scaling, encoding structure etc. should be confirmed. Integration of data warehouse handles various subject related warehouse.

- **Time-Variant**

In this data is maintained via different intervals of time such as weekly, monthly, or annually etc. It founds various time limit which are structured between the large datasets and are held in online transaction process (OLTP). The time limits for data warehouse is wide-ranged than that of operational systems. The data resided in data warehouse is predictable with a specific interval of time and delivers information from the historical perspective. It comprises elements of time explicitly or implicitly. Another feature of time-variance is that once data is stored in the data warehouse then it cannot be modified, alter, or updated.

- **Non-Volatile**

As the name defines the data resided in data warehouse is permanent. It also means that data is not erased or deleted when new data is inserted. It includes the mammoth quantity of data that is inserted into modification between the selected quantity on logical business. It evaluates the analysis within the technologies of warehouse.

In this, data is read-only and refreshed at particular intervals. This is beneficial in analysing historical data and in comprehension the functionality. It does not need transaction process, recapture and concurrency control mechanism. Functionalities such as delete, update, and insert that are done in an operational application are lost in data warehouse environment. Two types of data operations done in the data warehouse are:

- Data Loading
- Data Access

#### **Q14. What is Star-Schema??**

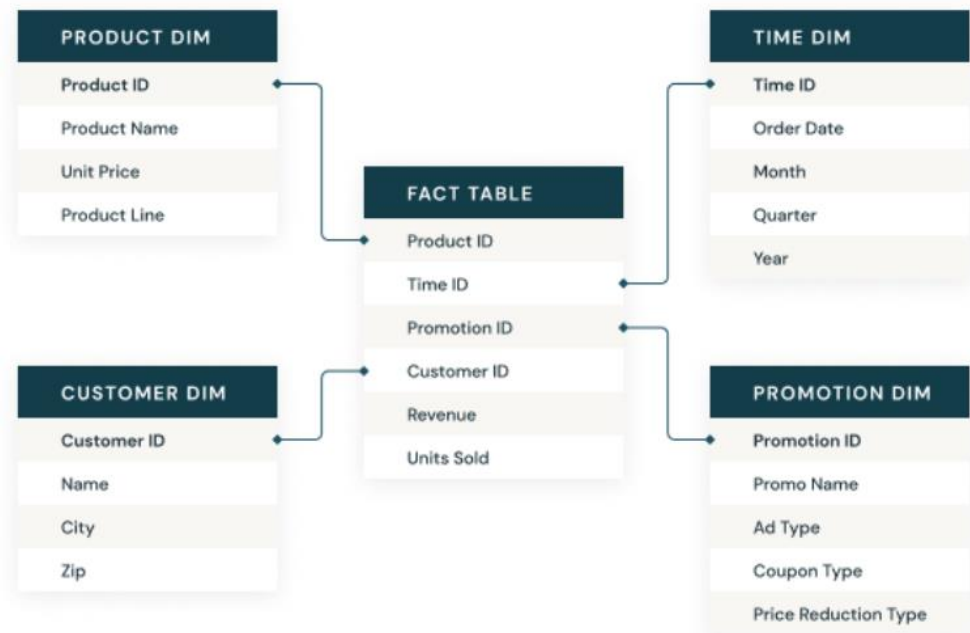
A star schema is the elementary form of a dimensional model, in which data are organized into **facts** and **dimensions**. A fact is an event that is counted or measured, such as a sale or log in. A dimension includes reference data about the fact, such as date, item, or customer.

A star schema is a relational schema where a relational schema whose design represents a multidimensional data model. The star schema is the explicit data warehouse schema. It is known as **star schema** because the entity-relationship diagram of this schemas simulates a star, with points, diverge from a central table. The center of the schema consists of a large fact table, and the points of the star are the dimension tables.

#### **Characteristics of Star Schema**

- It creates a DE-normalized database that can quickly provide query responses.
- It provides a flexible design that can be changed easily or added to throughout the development cycle, and as the database grows.
- It provides a parallel in design to how end-users typically think of and use the data.
- It reduces the complexity of metadata for both developers and end-users.

# Star schema



## Model of Star Schema:

In Star Schema, Business process data, that holds the quantitative data about a business is distributed in fact tables, and dimensions which are descriptive characteristics related to fact data. Sales price, sale quantity, distant, speed, weight, and weight measurements are few examples of fact data in star schema.

## **Q15. What do you mean by SETL?**

- Semantic ETL (SETL) framework. SETL builds on Semantic Web (SW) standards and tools and supports developers by offering a number of powerful modules, classes, and methods for (dimensional and semantic) DW constructs and tasks.
- It supports semantic data sources in addition to traditional data sources, semantic integration, and creating or publishing a semantic (multidimensional) DW in terms of a knowledge base. A comprehensive experimental evaluation comparing SETL to a solution made with traditional tools (requiring much more hand-coding) on a concrete use case, shows that SETL provides better programmer productivity, knowledge base quality, and performance.