WORKSHEET 1 SQL

Q1 and Q2 have one or more correct answer. Choose all the correct option to answer your question.

1. Which of the following is/are DDL commands in SQL?
   A) Create B) Update
   C) Delete D) ALTER
2. Which of the following is/are DML commands in SQL?
   A) Update B) Delete
   C) Select D) Drop

Q3 to Q10 have only one correct answer. Choose the correct option to answer your question.

3. Full form of SQL is:
   A) Strut querying language B) Structured Query Language
   C) Simple Query Language D) None of them
4. Full form of DDL is:
   A) Descriptive Designed Language B) Data Definition Language
   C) Data Descriptive Language D) None of the above.
5. DML is:
   A) Data Manipulation Language B) Data Management Language
   C) Data Modeling Language D) None of these
6. Which of the following statements can be used to create a table with column B int type and C float type?
   A) Table A (B int, C float) B) Create A (b int, C float)
   C) Create Table A (B int,C float) D) All of them
7. Which of the following statements can be used to add a column D (float type) to the table A created above?
   A) Table A (D float) B) Alter Table A ADD COLUMN D float
   C) Table A(B int, C float, D float) D) None of them
8. Which of the following statements can be used to drop the column added in the above question?
   A) Table A Drop D B) Alter Table A Drop Column D
   C) Delete D from A D) None of them
9. Which of the following statements can be used to change the data type (from float to int) of the column Dof table A created in above questions?
   A) Table A (D float int) B) Alter Table A Alter Column D int
   C) Alter Table A D float int D) Alter table A Column D float to int
10. 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 B) Alter table (B primary key)
    C) Alter Table A Add Primary key B D) None of them

Q11 to Q15 are subjective answer type questions, Answer them briefly.

11. What is data-warehouse?
12. What is the difference between OLTP VS OLAP?
13. What are the various characteristics of data-warehouse?
14. What is Star-Schema??
15. What do you mean by SETL?

11. What is data-warehouse?

A Data Warehouse (DW) is a relational database that is designed for query and analysis rather processing. It includes historical data derived from transaction data from single and multiple sources.

A data warehouse provides integrated, enterprise-wide, historical data and forces on providing support for decision makers for data modeling and analysis.

A Data Warehouse is a group of data specific to the entire organization, not only to a particular group of users.

A Data Warehouse can be viewed as a data system with the following attributes:

- It is a database designed for investigative tasks, using data from various applications.
- It supports a relatively small no of clients with relatively long interactions.
- It includes current and historical data to provide a historical perspective of information.
- Its usage is read-intensive
- It contains a few large tables.

"Data Warehouse" is a subject-oriented, integrated, and time-variant store of information in support of management's decisions."

Subject-Oriented:

A data warehouse target on the modeling and analysis of data for decision-makers. Therefore, data warehouses typically provide a concise and straightforward view around a particular subject, such as customer, product, or sales, instead of the global organization's ongoing operations. This is done by excluding data that are not useful concerning the subject and including all data needed by the users to understand the subject.

Integrated:

A data warehouse integrates various heterogeneous data sources like RDBMS, flat files, and online transaction records. It requires performing data cleaning and integration during data warehousing to ensure consistency in naming conventions, attributes types etc. among different data sources.

Time-Variant:

Historical information is kept in a data warehouse. For example, one can retrieve files from 3 months, 6 months, 12 months, or even previous data from a data warehouse. These variations with a transactions system, where often only the most current file is kept.

Non-Volatile:

The data warehouse is a physically separate data storage, which is transformed from the sources operational RDBMS. The operational updates of data do not occur in the data warehouse, update, insert, and delete operations are not performed. It usually requires only two procedures in data accessing: Initial loading of data and access to data. Therefore, the DW does not require transaction processing, recovery, and concurrency capabilities, which allows for substantial speedup of data retrieval. Non-Volatile defines that once entered into the warehouse, and data should not change.

Goals of Data Warehousing:

- To help reporting as well as analysis
- Maintain the organization's historical information
- Be the foundation for decision making

Need for Data Warehouse:

1. **Business User:** Business users require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.
2. **Store historical Data:** Data warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.
3. **Make Strategic Decisions:** Some strategies may be depending upon the data in the data warehouse. So data warehouse contributes to making strategic decisions.
4. **For Data Consistency and quality:** Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.
5. **High Response Time:** Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time.

Benefits of Data Warehouse:

1. Understand business trends and make better forecasting decisions.
2. Data Warehouses are designed to perform well enormous amounts of data.
3. The structure of data warehouses is more accessible for end-users to navigate, understand, and query.
4. Queries that would be complex in many normalized databases could be easier to build and maintain in data warehouses.
5. Data warehousing is an efficient method to manage demand for lots of information from lots of users.
6. Data warehousing provide the capabilities to analyze a large amount of historical data.

12. What is the difference between OLTP VS OLAP?

OLTP and OLAP: The two terms look similar but refer to different kinds of systems. Online transaction processing (OLTP) captures, stores, and processes data from transactions in real time. Online analytics process (OLAP) uses complex queries to analyze aggregated historical data from OLTP systems.

What is OLTP:

An OLTP system captures and maintains transaction data in a database. Each transaction involves individual database records made up of multiple fields or columns. Examples include banking and credit card activity or retail checkout scanning.

In OLTP, the emphasis is on fast processing, because OLTP databases are read, written, and updated frequently. If a transaction fails, built-in system logic ensures data integrity.

What is OLAP?

OLAP applies complex queries to large amounts of historical data, aggregated from OLTP databases and other sources, for data mining, analytics, and business intelligence projects. In OLAP, the emphasis is on response time to these complex queries. Each query involves one or more columns of data aggregated from many rows. Examples include year-over-year financial performance or marketing lead generation trends. OLAP databases and data warehouses give analysts and decision-makers the ability to use custom reporting tools to turn data into information.

Query failure in OLAP does not interrupt or delay transaction processing for customers, but it can delay or impact the accuracy of business intelligence insights.

	OLTP	OLAP
Characteristics	Handles a large number of small transactions	Handles large volumes of data with complex queries
Query types	Simple standardized queries	Complex queries
Operations	Based on INSERT, UPDATE, DELETE commands	Based on SELECT commands to aggregate data for reporting
Response time	Milliseconds	Seconds, minutes, or hours depending on the amount of data to process
Design	Industry-specific, such as retail, manufacturing, or banking	Subject-specific, such as sales, inventory, or marketing
Source	Transactions	Aggregated data from transactions
Purpose	Control and run essential business operations in real time	Plan, solve problems, support decisions, discover hidden insights
Data updates	Short, fast updates initiated by user	Data periodically refreshed with scheduled, long-running batch jobs
Space requirements	Generally small if historical data is archived	Generally large due to aggregating large datasets
Database design	Normalized databases for efficiency	Denormalized databases for analysis

13. What are the various characteristics of data-warehouse?

Data-Warehouse can be controlled when the user has a shared way of explaining the trends that are introduced as specific subject. Below are major characteristics of Data Warehouse.

Characteristics of Data-Warehouse:

1. Subject Oriented
2. Integrated
3. Time-Variant
4. Non-Volatile

1. **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.

2. **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.

3. **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.

4. **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

14. **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:

The star schema is intensely suitable for data warehouse database design because of the following features:

- 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.

Advantages of Star Schema:

Star Schemas are easy for end-users and application to understand and navigate. With a well-designed schema, the customer can instantly analyze large, multidimensional data sets.

The main advantage of star schemas in a decision-support environment are:

- Query Performance
- Load performance and administration
- Built-in referential integrity
- Easily Understood

Disadvantage of Star Schema:

There is some condition which cannot be meet by star schemas like the relationship between the user, and bank account cannot describe as star schema as the relationship between them is many to many.