WORKSHEET 1 SQL

1. a) Create
   d) ALTER
2. a) Update
   b) Delete
3. b) Structured Query Language
4. b) Data definition Language
5. a) Data Manipulation Language
6. c) Create A(B int, C float)
7. b) Alter Table A ADD COLUMN D float
8. b)Alter Table A Drop Column D
9. b)Alter Table A Alter Column D int
10. c)Alter Table A Add Primary key B

11. A data warehouse is a centralized storage system that allows for the storing, analyzing, and interpreting of data in order to facilitate better decision-making. Transactional systems, relational databases, and other sources provide data into data warehouses on a regular basis.

A data warehouse is a type of data management system that facilitates and supports business intelligence (BI) activities

Data warehouse applications can be categorized as:

- **Query and reporting tools**
- **Application Development tools**
- **Data mining tools**
- **OLAP tools**

Some popular data warehouse tools are Xplenty, Amazon Redshift, Teradata, Oracle 12c, Informatica, IBM Infosphere, Cloudera, and Panoply.

Several benefits of data warehouse for end users.

- Improved data consistency
- Better business decisions
- Easier access to enterprise data for end-users
- Better documentation of data

- Reduced computer costs and higher productivity
- Enabling end-users to ask ad-hoc queries or reports without deterring the performance of operational systems
- Collection of related data from various sources into a place

Data Warehouse Architecture

Usually, data warehouse architecture comprises a three-tier structure.

Bottom Tier -The bottom tier or data warehouse server usually represents a relational database system. Back-end tools are used to cleanse, transform and feed data into this layer.

Middle Tier - The middle tier represents an OLAP server that can be implemented in two ways.

- The ROLAP or Relational OLAP model is an extended relational database management system that maps multidimensional data process to standard relational process.
- The MOLAP or multidimensional OLAP directly acts on multidimensional data and operations.

Top Tier - This is the front-end client interface that gets data out from the data warehouse. It holds various tools like query tools, analysis tools, reporting tools, and data mining tools.

12. **Online Analytical Processing (OLAP):** Online Analytical Processing consists of a type of software tools that are used for data analysis for business decisions. OLAP provides an environment to get insights from the database retrieved from multiple database systems at one time. Examples – Any type of Data warehouse system is an OLAP system. The uses of OLAP are as follows:

Spotify analyzed songs by users to come up with a personalized homepage of their songs and playlist.Netflix movie recommendation system.

Online transaction processing (OLTP): Online transaction processing provides transaction-oriented applications in a 3-tier architecture. OLTP administers the day-to-day transactions of an organization.

Examples: Uses of OLTP are as follows:

- ATM center is an OLTP application.
- OLTP handles the ACID properties during data transactions via the application.
- It’s also used for Online banking, Online airline ticket booking, sending a text message, add a book to the shopping cart.

OLAP vs OLTP

1 OLAP stands for Online analytical processing. OLTP stands for online transaction processing.

2 It includes software tools that help in analyzing data mainly for business decisions. It helps in managing online database modification.

3 It utilizes the data warehouse. It utilizes traditional approaches of DBMS.

4 It is popular as an online database query management system. It is popular as an online database modifying system.

5 OLAP employs the data warehouse. OLTP employs traditional DBMS.

6 It holds old data from various Databases. It holds current operational data.

7 Here the tables are not normalized. Here, the tables are normalized.

8 It allows only read and hardly write operations. It allows both read and write operations.

9 Here, the complex queries are involved. Here, the queries are si

13. Key Characteristics of Data Warehouse

The main characteristics of a data warehouse are as follows:

- **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:**

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.

- **Non-Volatile:**

Data warehouse is a physically separate data storage, which is transformed from the source operational RDBMS. The operational updates of data do not occur in the data warehouse, i.e., 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.

- **Time-Variant:**

The data stored in a data warehouse is documented with an element of time, either explicitly or implicitly. An example of time variance in Data Warehouse is exhibited in the Primary Key, which must have an element of time like the day, week, or month.

14. Star Schema in data warehouse, is a schema in which the center of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure resembles a star. The Star Schema data model is the simplest type of Data Warehouse schema. It is also known as Star Join Schema and is optimized for querying large data sets.

Fact Tables

A Fact table in a star schema contains facts and is connected to dimensions. A fact table has two types of columns: • A column that includes Facts • Foreign Key to Dimensions Table Generally, the primary key of a fact table is a composite key that is made up of all the foreign keys that make up the table. Fact tables can contain detail-level facts or aggregated facts. Fact tables that include aggregated facts are often called summary tables. Fact tables usually contain facts that have been aggregated to some level.

Dimension Tables

A dimension is an architecture that categorizes data in a hierarchy. A dimension without hierarchies and levels is called a flat dimension or list. Each dimension table's primary key is part of the composite primary key of the fact table. A dimension attribute is a descriptive, textual attribute that helps describe a dimensional value. Fact tables are usually larger than dimension tables. Characteristics of Star Schema Every dimension in a star schema is represented with the only one-dimension table. The dimension table should contain the set of attributes. The dimension table is joined to the fact table using a

foreign key The dimension table are not joined to each other Fact table would contain key and measure The Star schema is easy to understand and provides optimal disk usage. The dimension tables are not normalized. For instance, in the above figure, Country_ID does not have Country lookup table as an OLTP design would have. The schema is widely supported by BI Tools Advantages of Star Schema Star schemas have a more straightforward join logic compared to other schemas for fetching data from highly normalized transactional schemas. As opposed to highly normalized transactional schemas, the star schema simplifies common business reporting logic, such as reporting and period-over-period. Star schemas are widely used by OLAP systems to design cubes efficiently. A star schema can be used as a source without designing a cube structure in most major OLAP systems. By enabling specific performance schemes that can be applied to queries, the query processor software in Star Schema can offer better execution plans. Disadvantage of Star Schema Since the schema is highly de-normalized, data integrity is not enforced well. Not flexible in terms of analytical needs. Star schemas do not reinforce many-to-many relationships within business entities.

15. SQL set operators are used to combine the results obtained from two or more queries into a single result. The queries which contain two or more subqueries are known as compounded queries.

There are four major types of SQL operators, namely:

- **Union:** Combines distinct results of two or more SELECT statements.
- **Union all:** Combines all results of two or more SELECT statements, including duplicates
- **Intersect:** Returns only the common records obtained from two or more SELECT statements.
- **Minus:** Returns only those records which are exclusive to the first table.
- **Syntax and Parameters of SQL Set Operators**
- The generic syntax for working with SQL set operators is as follows:

- **Syntax:**
- SELECT column_name

  FROM table_name_1

```
SET OPERATOR

SELECT column_name

FROM table_name_2

SET OPERATOR

SELECT column_name

FROM table_name_3

.

.

.
```

Parameters:

The different parameters used in the syntax are :

- **SET OPERATOR:** Mention the type of set operation you want to perform from { Union, Union all, Intersect, Minus}
- **column_name:** Mention the column name on which you want to perform the set operation and want in the result set
- **FROM table_name_1:** Mention the first table name from which the column has to be fetched
- **FROM table_name_2:** Mention the second table name from which the column has to be fetched

From the above-mentioned parameters, all the parameters are mandatory. You may use WHERE GROUP BY and HAVING clauses based on your requirements.