Submitted By: Arjun B
Batch Number: DS0082

Worksheet 1- SQL

1. a) Create

   d) Alter

2. a) Update

   b) Delete

   c) Select

3. b) Structured Query Language

4. b) Data Definition Language

5. a) Data Manipulation Language

6. c) Create Table 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 a) Alter Table A Add Primary Key B

11. Data Warehousing is a method for gathering, transforming, and storing data from several heterogeneous sources in order to offer valuable business insights. Typically, corporate data from various sources is connected and analyzed in a data warehouse. Other names for it include Business Intelligence Solutions, Management Information Systems, and Decision Support Systems (DSS). Enterprise data warehouses (EDW), operational data stores (ODS), and data marts are the three main categories of data warehouses.
    Eg: Health care industry
    A DWH is the backbone of healthcare systems because the latest, up-to-date treatment information is crucial for saving lives. EDWs forecast outcomes, generate treatment reports and share data with insurance providers, and research labs.

Submitted By: Arjun B
Batch Number: DS0082

12 comparison between OLTP and OLAP

OLTP	**OLAP**
It is a well-known online database modifying system	It is a well-known online database query management system.
Consists of current operational data	Consists of historical data from various Databases.
In an OLTP database, tables are normalized (3NF).	In an OLAP database, tables are not normalized.
The data is used to perform day-to-day fundamental operations.	The data is used in planning, problem-solving, and decision-making.
The data integrity constraint must be maintained in an OLTP database.	The OLAP database is not often updated. As a result, data integrity is unaffected.
It is comparatively fast in processing because of simple and straightforward queries.	The processing of complex queries can take a lengthy time.
It is application-oriented. Used for business tasks.	It is subject-oriented. Used for Data Mining, Analytics, Decisions making, etc.

13.. **Characteristics of Data Warehouse**

Subject Oriented - The data warehouse is subject-oriented as it delivers information about a theme instead of the organization's current operations. It can be achieved on a specific theme that is more defined, such as sales, distributions, marketing, etc.

· **Integrated** - Data warehouse is a collection of integrated data from various sources of data such as transactional databases and operational reports. The mainframe and relational database are integrated to form a single unit, creating accessibility for analysis. Reliability in naming conventions, format, and codes is also required when building a data warehouse.

·**Time Variant** - Time variance is the feature of a data warehouse that keeps data in chronological order and it defines the time element explicitly or implicitly. 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.

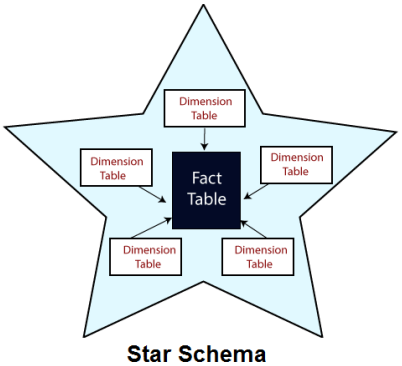
· **Non-Volatile** - The data warehouse is the primary source of information for decision-making. The data in a data warehouse is permanent. It means that it will not be erased or deleted when new data is inserted into modifications between the selected quantity on logical business

14 Star Schema

A star schema is a relational database design that establishes a central fact table and surrounds it with dimension tables to represent the underlying multidimensional data model. The star schema can be used as an analytical tool to access data from several sources and create aggregations across those sources.

Characteristics:

- 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

Submitted By: Arjun B
Batch Number: DS0082

15. SETL (SET Language) is a very high-level programming language based on the mathematical theory of sets. SETL provides two basic aggregate data types: unordered sets, and sequences (the latter also called tuples). The elements of sets and tuples can be of any arbitrary type, including sets and tuples themselves. Maps are provided as sets of pairs (i.e., tuples of length 2) and can have arbitrary domain and range types. Primitive operations in SETL include set membership, union, intersection, and power set construction, among others.

ELT-Extract, Transform, and Load (ETL) is a methodology used to transform unstructured or semi-structured data from its source format into another format that can be loaded into a database or data warehouse. It is most commonly used to extract data from various sources, transform this data by business rules, and then load it into a destination database. The need for ETL arises from the fact that many organizations have multiple systems containing data in different formats.

The ETL process has 3 main steps, which are Extract, Transform, and Load.

Extract – The first step in the ETL process is extracting the data from various sources. Each of the source systems may store its data in a completely different format from the rest. The sources are usually flat files or RDBMS, but almost any data storage can be used as a source for an ETL process.

Transform – Once the data has been extracted and converted in the expected format, it's time for the next step in the ETL process, which is transforming the data according to a set of business rules. The data transformation may include various operations including but not limited to filtering, sorting, aggregating, joining data, cleaning data, generating calculated data based on existing values, validating data, etc.

Load – The final ETL step involves loading the transformed data into the destination target, which might be a database or data warehouse.