ASSIGNMENT

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 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. C) Alter Table A Add Primary key B

11. What Is a Data Warehouse

Data warehouses serve as a central repository for storing and analysing information to make better informed decisions. An organisation's data warehouse receives data from a variety of sources, typically on a regular basis, including transactional systems, relational databases, and other sources.

A data warehouse is a centralised storage system that allows for the storing, analysing, 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, specifically analysis. Data warehouses are primarily designed to facilitate searches and analyses and usually contain large amounts of historical data.

A data warehouse can be defined as a collection of organisational data and information extracted from operational sources and external data sources. The data is periodically pulled from various internal applications like sales, marketing, and finance; customer-interface

applications; as well as external partner systems. This data is then made available for decision-makers to access and analyse.

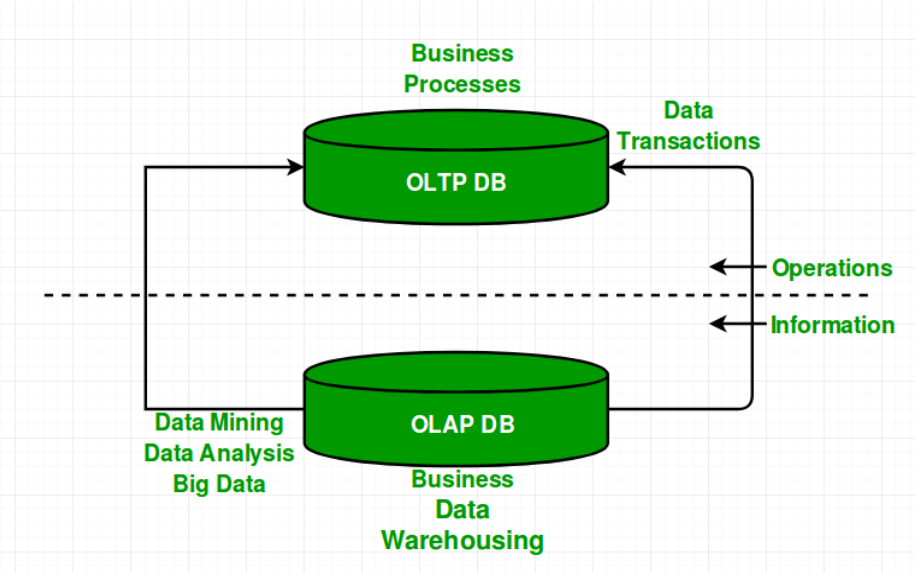
12. Online Analytical Processing (OLAP): Online Analytical Processing consists of a type of software tools that are used for data analysis for business decisions. OLAP provides an environment to get insights from the database retrieved from multiple database systems at one time. **Examples –** Any type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

- Spotify analysed songs by users to come up with a personalised 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 organisation.

Examples: Uses of OLTP are as follows:

- ATM centre 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.



Comparisons of OLAP vs OLTP :

Sr. No.	Category	OLAP (Online analytical processing)	OLTP (Online transaction processing)
1.	Definition	It is well-known as an online database query management system.	It is well-known as an online database modifying system.
2.	Data source	Consists of historical data from various Databases.	Consists of only operational current data.

3.	Method used	It makes use of a data warehouse.	It makes use of a standard database management system (DBMS).
4.	Application	It is subject-oriented. Used for Data Mining, Analytics, Decisions making, etc.	It is application-oriented. Used for business tasks.
5.	Normalised	In an OLAP database, tables are not normalised.	In an OLTP database, tables are normalised (3NF).
6.	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.
7.	Task	It provides a multi-dimensional view of different business tasks.	It reveals a snapshot of present business tasks.
8.	Purpose	It serves the purpose to extract information for	It serves the purpose to Insert, Update, and Delete

		analysis and decision-making.	information from the database.
9.	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. For ex MB, GB
10.	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.
11.	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.
12.	Backup and Recovery	It only need backup from time to time as compared to OLTP.	Backup and recovery process is maintained rigorously
13.	Processing time	The processing of complex queries can take a lengthy time.	It is comparatively fast in processing because of simple and straightforward queries.

14.	Types of users	This data is generally managed by CEO, MD, GM.	This data is managed by clerks, managers.
15.	Operations	Only read and rarely write operation.	Both read and write operations.
16.	Updates	With lengthy, scheduled batch operations, data is refreshed on a regular basis.	The user initiates data updates, which are brief and quick.
17.	Nature of audience	Process that is focused on the customer.	Process that is focused on the market.
18.	Database Design	Design with a focus on the subject.	Design that is focused on the application.
19.	Productivity	Improves the efficiency of business analysts.	Enhances the user's productivity.

13. Key Characteristics of Data Warehouse

The main characteristics of a data warehouse are as follows:

- Subject-Oriented

A data warehouse is subject-oriented since it provides topic-wise information rather than the overall processes of a business. Such subjects may be sales, promotion, inventory, etc. For example, if you want to analyse your company's sales data, you need to build a data warehouse that concentrates on sales. Such a warehouse would provide valuable information like 'who was your best customer last year?' or 'who is likely to be your best customer in the coming year?'

- Integrated

A data warehouse is developed by integrating data from varied sources into a consistent format. The data must be stored in the warehouse in a consistent and universally acceptable manner in terms of naming, format, and coding. This facilitates effective data analysis.

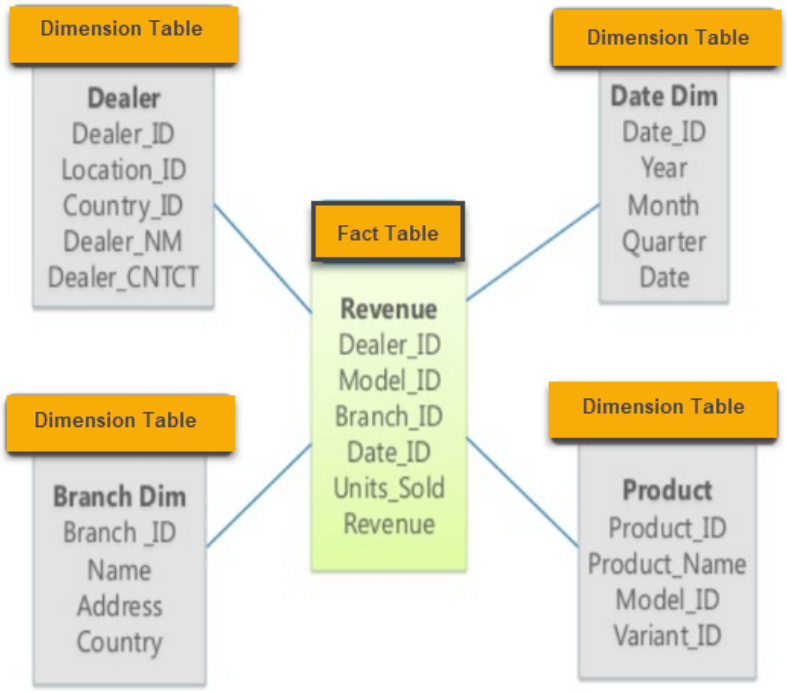
- Non-Volatile

Data once entered into a data warehouse must remain unchanged. All data is read-only. Previous data is not erased when current data is entered. This helps you to analyse what has happened and when.

- 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. A star schema is a database organisational structure optimised for use in a data warehouse or business intelligence that uses a single large fact table to store transactional or measured data, and one or more smaller dimensional tables that store attributes about the data. It is called a star schema because the fact table sits at the centre of the logical diagram, and the small dimensional tables branch off to form the points of the star.



15.

SETL: A programmable semantic extract-transform-load framework for semantic data warehouses

In order to create better decisions for business analytics, organizations increasingly use external structured, semi-structured, and unstructured data in addition to the (mostly structured) internal data. Current Extract-Transform-Load (ETL) tools are not suitable for this “open world scenario” because they do not consider semantic issues in the integration processing. Current ETL tools neither support processing semantic data nor create a semantic Data Warehouse (DW), a repository of semantically integrated data. This paper describes our programmable 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. Thus 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.