WORKSHEET-1

***Q.1: ans**. – a,d.*

***Q.2: ans**. – a,b.*

***Q.3: ans**. – b.*

***Q.4: ans**. – b.*

***Q.5: ans**. – a.*

***Q.6: ans**. – d.*

***Q.7: ans**. – a.*

***Q.8: ans**. – a.*

***Q.9: ans**. – b.*

***Q.10: ans**. – b.*

***Q.11: ans**. - A **Data Warehousing** (DW) is process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.*

It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

***Q.12: ans**. - **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.*

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 of operational current data.
3.	Method used	It makes use of a data warehouse.	It makes use of a standard database

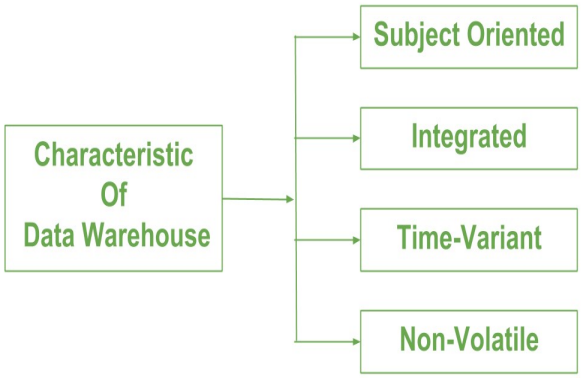
WORKSHEET-1

Sr.No.	Category	OLAP (Online analytical processing)	OLTP (Online transaction processing)
			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.	Normalized	In an OLAP database, tables are not normalized.	In an OLTP database, tables are normalized (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 analysis and decision-making.	It serves the purpose to Insert, Update, and Delete 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.

WORKSHEET-1

Sr.No.	Category	OLAP (Online analytical processing)	OLTP (Online transaction processing)
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.

Q.13: ans. - 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:



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.

WORKSHEET-1

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*

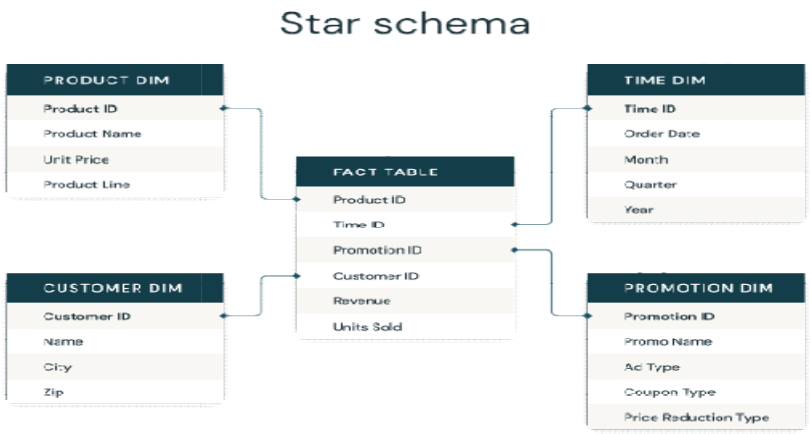
Functions of Data warehouse:
It works as a collection of data and here is organized by various communities that endures the features to recover the data functions. It has stocked facts about the tables which have high transaction levels which are observed so as to define the data warehousing techniques and major functions which are involved in this are mentioned below:

1. *Data consolidation*
2. *Data Cleaning*
3. *Data Integration*

Q.14: ans*. - A* ***star schema*** *is a multi-dimensional data model used to organize data in a database so that it is easy to understand and analyze. Star schemas can be applied to data warehouses, databases, data marts, and other tools. The star schema design is optimized for querying large data sets.*

Introduced by Ralph Kimball in the 1990s, star schemas are efficient at storing data, maintaining history, and updating data by reducing the duplication of repetitive business definitions, making it fast to aggregate and filter data in the data warehouse.

WORKSHEET-1



Fact tables and dimension tables

A star schema is used to denormalize business data into dimensions (like time and product) and facts (like transactions in amounts and quantities).

A star schema has a single fact table in the center, containing business "facts" (like transaction amounts and quantities). The fact table connects to multiple other dimension tables along "dimensions" like time, or product. Star schemas enable users to slice and dice the data however they see fit, typically by joining two or more fact tables and dimension tables together.

Denormalized data

*Star schemas **denormalize** the data, which means adding redundant columns to some dimension tables to make querying and working with the data faster and easier. The purpose is to trade some redundancy (duplication of data) in the data model for increased query speed, by avoiding computationally expensive join operations.*

In this model, the fact table is normalized but the dimensions tables are not. That is, data from the fact table exists only on the fact table, but dimensional tables may hold redundant data.

Benefits of star schemas

- *Fact/dimensional models like star schemas are **simple** to understand and implement, and make it easy for end users to find the data they need. They can be applied to data marts and other data resources.*
- ***Great for simple queries** because of their reduced dependency on joins when accessing the data, as compared to normalized models like snowflake schemas.*
- ***Adapt well to fit OLAP models.***
- ***Improved query performance** as compared to normalized data, because star schemas attempt to avoid computationally expensive joins.*

WORKSHEET-1

How does a star schema differ from 3NF (Third Normal Form)?

3NF, or Third Normal Form, *is a method of reducing data-redundancy through normalization. It is a common standard for databases that are considered fully normalized. It typically has more tables than a star schema due to data normalization. On the flip-side, queries tend to be more complex due to the increased number of joins between large tables.*

Q.15: ans. SETL *(SET Language) is a very high-level programming language based on the mathematical theory of sets. It was originally developed by (Jack) Jacob T. Schwartz at the New York University (NYU) Courant Institute of Mathematical Sciences in the late 1960s.*

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.

SETL provides quantified boolean expressions constructed using the universal and existential quantifiers of first-order predicate logic.

SETL provides several iterators to produce a variety of loops over aggregate data structures.

Implementations of SETL were available on the DEC VAX, IBM/370, SUN workstation and APOLLO. In the 1970s, SETL was ported to the BESM-6, ES EVM and other Russian computer systems.

SETL was used for an early implementation of the programming language Ada, named the NYU Ada/ED translator. This later became the first validated Ada implementation, certified on April 11, 1983.

According to Guido van Rossum, "Python's predecessor, ABC, was inspired by SETL -- Lambert Meertens spent a year with the SETL group at NYU before coming up with the final ABC design!" .