Worksheet 1 sql

1.a and d

2.a and b

3.b

4.b

5.a

6.c

7.b

8.b

9.d

10.a

11.DATA WAREHOUSE:

Data warehouses serve as a central repository for storing and analyzing information to make better informed decisions. An organization'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 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. It is a type of data management system that facilitates and supports business intelligence (BI) activities, specifically analysis.

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.

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.

	OLTP System (Operational Data)	OLAP System (Data Warehouse)
Data source	From Traditional or original source	From diverse databases
Aim of data	Organizational basic activities running	For planning and problem solving known Business support system
Data type	images of ongoing business transaction	Business activities from various sections
Insertion and updating	Insert and update are by end users which is fast	Periodic refreshing inform of Batch jobs
Queries	Simple queries that return small records	Complex aggregate queries
Efficiency	Very high speed to answer queries	Takes some time depending on the requirement but can be improved using index
Space hold	holds very small history	Comprehensive history due to aggregation
Database design	Greatly normalized due to many tables involvement	Highly denormalised with few tables using Star and snowflake schema.

13. Characteristics of data warehouse

- [ ] 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.

- [ ] Integrated –

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.

☐   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.

☐   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

14. 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.

Fact Tables
A table in a star schema which contains facts and connected to dimensions. A fact table has two types of columns: those that include fact and those that are foreign keys to the dimension table. The primary key of the fact tables is generally a composite key that is made up of all of its foreign keys.A fact table might involve either detail level fact or fact that have been aggregated (fact tables that include aggregated fact are often instead called summary tables). A fact table generally contains facts with the same level of aggregation.

Dimension Tables
A dimension is an architecture usually composed of one or more hierarchies that categorize data. If a dimension has not got hierarchies and levels, it is called a flat dimension or list. The primary keys of each of the dimensions table are part of the composite primary keys of the fact table. They are generally descriptive, textual values. Dimensional tables are usually small in size than fact table.

15.
- Current ETL tools neither support processing semantic data nor create a semantic Data Warehouse , a repository of semantically integrated data.
- SETl stands for semantic extract tansform load.it is a framework for semantic Data Warehouse.
- SETL builds on Semantic Web standards and tools and supports developers by offering a number of powerful modules, classes, and methods for (dimensional and semantic) Data Warehouse constructs and tasks. Thus it supports semantic data sources in addition to traditional data sources, semantic integration, and creating or publishing a semantic (multidimensional) Data Warehouse in terms of a knowledge base.
- SETL provides better programmer productivity, knowledge base quality, and performance than traditional tools