**Data Engineering Assignment-1**

Question: Explain the data warehousing concept?

Answer:

Data warehouse is a collection of data designed to support management decision making by presenting a coherent picture of business conditions at a single point of time. In simple terms we can say that, a data warehouse is a centralized repository that is used for storing, integrating, and managing large volumes of data from various sources within an organization.

It is designed to support decision-making processes by providing a unified, historical, and consistent view of data across the organization. The main purpose of data warehouse is to provide a platform for efficient data analysis and reporting.

Features of data warehouse:

* Subject-oriented: Instead of organizing the data according to application, the data is organized based on subject. Its main focus is to model and analyze the data for decision makers.
* Integrated: Data in data warehouse is collected from various heterogenous sources like relational databases, flat files, online transaction records etc.
* Time-variant: The time variant feature of a data warehouse refers to its ability to manage and store historical data over time. Unlike operational databases that often focus on the current state of data, a data warehouse is designed to capture, organize data over different points in time.
* Non-volatile: Non-volatile feature states that no updates are allowed. Once the data entered into the warehouse, it cannot be changed.

Data Warehouse Architectures:

* OLTP: Online Transaction Processing
* OTAP: Online Analytical Processing

OLTP:

OLTP is a methodology where end users are provided access to large amounts

of data. It refers to a class of systems and processes that handle the real-time, day-to-day transactional operations of an organization. These transactions typically involve interactions with customers, such as sales, order processing, and other routine business operations. OLTP systems are designed to ensure the consistency and accuracy of data in the operational databases that support an organization's core business functions.

Example of OLTP is Atm.

OLAP:

Unlike OLTP (Online Transaction Processing), which focuses on transactional data and day-to-day operations, OLAP is designed for complex queries and data analysis, providing a multidimensional view of historical and aggregated data. Here are key characteristics and aspects of OLAP:

Operational data store:

An Operational Database is a type of database that is optimized for the day-to-day processing and management of operational data within an organization. Unlike data warehouses, which are designed for analytical processing and historical analysis, operational databases focus on supporting transactional activities and the immediate needs of an organization's daily operations. Operational database uses OLTP architecture which is optimized for faster transaction processing.

Data Marts:

The data in the data warehouse is stored in the form of Data marts. It allows the user to access the data in terms of a specific business line or team. In simple terms we can define it as, Data mart is a subset of a data warehouse that is designed to serve the specific data and analytical needs of a particular business unit, department, or group within an organization. Data marts are smaller, more focused data repositories that provide a targeted view of data, tailored to meet the requirements of a specific user community.

And there are other topics that are discussed. They are:

Types of data:

* Raw data: Unprocessed data and no schema applied
* Processed data: Raw data with schema applied
* Cooked data: Processed data that has been summarized

Big data Properties:

* Volume: It is the amount of data that you have
* Velocity: It is stated as how fast the data is getting to you
* Variety: It refers to different types of data
* Veracity: It refers to how accurate your data is

Data Processing techniques:

* Batch processing
* Stream processing

Decision Support Systems (DSS):

Decision Support Systems helps us to assess and resolve everyday business queries. It works by filtering useful information from a combination of raw data, documents, personal knowledge or business models. Its main objective is to assist decision-makers in analyzing information and making decisions that impact the organization's goals and objectives.

And we also talked about structures and unstructured components of DSS.

* A Structured component is one which helps us to directly proceed with decisions.
* An Unstructured component is one which is to be still processed and requires human interaction.