Name - Aparna Bharti

Batch - Data Engineering batch 1

**Topic - Data Warehouse**

Data warehouse is a type ofdata management system that is designed to support business intelligence (BI) activities Data warehouses are used to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications,spreadsheets.

**Data Warehouse features**

**Subject-oriented**

This means that the data is organised around specific subjects, such as customers, products, or sales. This allows for easy access to the data relevant to a specific subject.

**Integrated**

This means that the data is collected from various sources, such as transactional systems, and then cleaned, transformed, and into a single view. This allows for easy access and analysis of the data

**Time-variant**

which means that the data is stored with a time dimension. This allows for easy access to data for specific time periods, such as last quarter or last year

**Nonvolatile**

This means that the data in the warehouse is never updated or deleted, only added

**DSS(decision support system)**

A decision support system (DSS) is a computerised system that gathers and analyses data to produce information reports.

**Components of DSS**

**Structured component**

A Structured component is one which directly helps us to proceed towards a decision.

**Unstructured component**

An Unstructured component is that which is to be processed and requires human interaction with the DSS.

**Architectural Style of DSS**

**OLTP**

OLTP stands for online transaction processing.It is responsible for collecting, storing, and processing data from transactions.

**Example** - ATM( Automated Teller Machine ),online ticketing

**Advantage of OLTP**

OLTP systems provide fast query processing

**Disadvantage of OLTP**

OLTP requires instant updates.

The data that we get from OLTP is not suitable for data analysis.

**OLAP**

OLAP stands for online analytical processing.It is responsible for analysing data from OLTP systems.

**Why do we need Data Engineering?**

Companies have access to so much information multiple data sources can also lead to information overload. This results in scattered data, which restricts the organisation from drawing relevant insights. That is where data engineering plays an integral part. It simplifies data and makes it more reliable

**4 V’s of Data Engineering**

**Volume**

The volume of data refers to the size of the data sets that need to be analyzed and processed.

**Example -** facebook

**Velocity**

Velocity refers to the speed with which data is generated.

**Example -** facebook generates every second millions of data

**Variety**

Data generated could be of great varieties and generally is one out of three types: structured, semi structured and unstructured data.

**Veracity**

Veracity refers to the quality of the data that is being analysed

**Types of Processing of Data**

**Batch processing**

In batch processing data in collected and analysed after that insight’s are generated

**Stream processing**

Stream processing refers to processing of a continuous stream of data immediately as it is produced. It analyses streaming data in real time

**Types of Data storage**

**Relational Database(SQl)**

A relational database organise data in rows and columns.

**Example -** Mysql,PostgreSQl,Oracle

**Document Store(NoSQL)**

Non-relational databases organise data in documents like JSOn , BLOB,XML

**Example -** MongoDB