

***DATA ENGINEERING***

***TASK-02***

*Submitted to*

***Sir Muhammad Ahtisham***

*Submitted by*

***Aabra Malik***

***Data Mart***

A smaller portion of a bigger data warehouse called a "data mart" is created to cater to the demands of a particular department or group inside an organization.

Data marts are often focused on a single business function, such as sales, marketing, finance, or human resources, and contain a portion of the data that is housed in the organization's data warehouse. The organization and optimization of the data in a data mart are tailored to the unique requirements of the department or group that uses it, making it simpler and more effective to access and analyze.

Data marts can be configured to be either autonomous or dependent on the bigger data warehouse. Independent data marts are freestanding entities that are built and maintained separately from the data warehouse, while dependent data marts are created from subsets of data that are extracted from the larger data warehouse.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

***Data Lakehouse***

A data lakehouse is a modern data management architecture that combines the benefits of data lakes with data warehouses. It is a unified platform that allows users to store and analyze structured, semi-structured, and unstructured data in a single spot.

In a data lakehouse, data is absorbed into a data lake, which is a huge, centralized repository of raw, unstructured data. This data can then be changed and arranged using a combination of schema-on-read and schema-on-write techniques to make it more accessible and usable for analysis. The data warehouse provides users with fast and efficient access to the data, making it easier to run complex queries and generate insights.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

***Data Mesh***

An architectural paradigm called "data mesh" aims to address the problems associated with scaling data management in big businesses.

In typical centralized data architectures, data is managed and controlled by a centralized team, which can cause bottlenecks and impede down the supply of data-driven insights. Since each domain or business unit is in charge of controlling its own data goods and services, data mesh advocates for decentralized data ownership.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

***DWH vs Data Lake***

In order to assist business intelligence (BI) and decision-making processes, data warehouses are centrally located, structured repositories of data. It typically has data that has been cleaned, transformed, and organized to fit a particular schema, and is designed with querying and reporting in mind. Data warehouses are frequently used to preserve historical data and offer the company a single source of truth.

In contrast, a data lake is a distributed, unstructured repository of data that is meant to hold enormous volumes of raw, unprocessed data in its natural format. It is capable of storing organized, semi-structured, and unstructured data, among others. Data lakes allow data scientists and analysts to deal with a variety of data sources, making them a common tool for data exploration and discovery, as they enable data scientists and analysts to work with a wide range of data sources and types, without the need for extensive data preparation.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

***OTLP vs OLAP***

OTLP (Online Transaction Processing) and OLAP (Online Analytical Processing) are two different types of data processing systems that are used for different purposes.

OTLP is a system that is used to process real-time transactions or data as they occur. OTLP systems are designed to efficiently manage and process large volumes of transactions, such as credit card purchases, bank deposits and withdrawals, and online orders. The primary focus of an OTLP system is to ensure that transactions are processed accurately, quickly, and securely, without any loss of data.

OLAP, on the other hand, is a system that is used to analyze large amounts of data and provide insights into business operations. OLAP systems are designed to handle complex queries and perform data analysis, often using advanced algorithms and statistical methods. The primary focus of an OLAP system is to provide business intelligence and support decision-making.

***\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****