**TASK NO: 2**

**What is a Data Mart?**

A data mart is a subset of a data warehouse focused on a particular line of business, department, or subject area.

Data marts make specific data available to a defined group of users, which allows those users to quickly access critical insights without wasting time searching through an entire data warehouse.

For example, many companies may have a data mart that aligns with a specific department in the business, such as finance, sales, or marketing.

Data marts are designed to meet the needs of specific groups by having a comparatively narrow subject of data. And while a data mart can still contain millions of records, its objective is to provide business users with the most relevant data in the shortest amount of time.

**Types of Data Marts:**

**Dependent data marts** are partitioned segments within an enterprise data warehouse. This top-down approach begins with the storage of all business data in one central location. The newly created data marts extract a defined subset of the primary data whenever required for analysis.

**Independent data marts** act as a standalone system that doesn't rely on a data warehouse. Analysts can extract data on a particular subject or business process from internal or external data sources, process it, and then store it in a data mart repository until the team needs it.

**Hybrid data marts** combine data from existing data warehouses and other operational sources. This unified approach leverages the speed and user-friendly interface of a top-down approach and also offers the enterprise-level integration of the independent method.

# **What is a Data Lakehouse?**

A Data lake house can be defined as a modern data platform built from a combination of a data lake and a data warehouse.

More specifically, a data lake house takes the flexible storage of unstructured data from a data lake and the management features and tools from data warehouses, then strategically implements them together as a larger system.

This integration of two unique tools brings the best of both worlds to users.

## **Advantages of a data lake house: A modern data platform**

By building a data lake house, organizations can streamline their overall data management process with a unified modern data platform.

**Less administration:** By using a data lake house, any sources connected to it can have their data accessible and consolidated for usage, as opposed to extracting it from raw data and preparing to work within a data warehouse.

**Better data governance:** Data lake houses simplify and improve governance by consolidating resources and data sources, and are built with standardized open scheme, which allows for greater control over security, metrics, role-based access, and other critical management elements.

**Simplified standards:** Data warehouses originated in the 1980s, when connectivity was extremely limited, meaning localized schema standards were often created within organizations, even departments. Today, open standards for schema exist for many types of data, and data lake houses take advantage of that by ingesting multiple data sources with overlapping standardized schema to simplify processes.

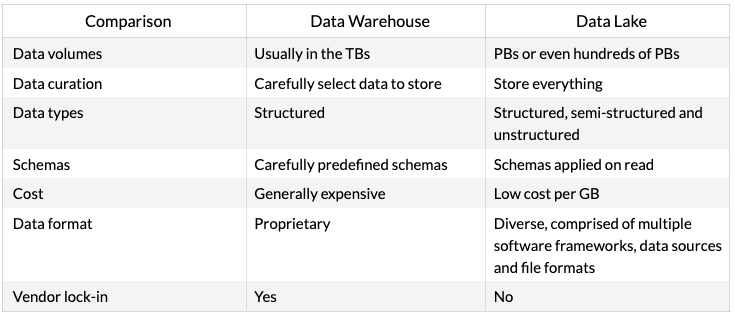
**Increased cost-effectiveness:** Data lake houses are built with modern infrastructure that separates compute and storage, which allows for easy addition of storage without the need to augment compute power. This creates cost-effective scaling with the simple use of low-cost data storage.

## **What is a Data Mesh?**

A data mesh is an architectural framework that solves advanced data security challenges through distributed, decentralized ownership. Organizations have multiple data sources from different lines of business that must be integrated for analytics. A data mesh architecture effectively unites the disparate data sources and links them together through centrally managed data sharing and governance guidelines. Business functions can maintain control over how shared data is accessed, who accesses it, and in what formats it’s accessed. A data mesh adds complexities to architecture but also brings efficiency by improving data access, security, and scalability.

## **Data Lake vs. Data Warehouse**

**A data lake is a massive repository of structured and unstructured data, and the purpose for this data has not been defined. A data warehouse is a repository of highly structured historical data which has been processed for a defined purpose.**



## **What is OLAP?**

[**Online analytical processing (OLAP)**](https://www.ibm.com/cloud/learn/olap) is a system for performing multi-dimensional analysis at high speeds on large volumes of data. Typically, this data is from a data warehouse, data mart or some other centralized data store. OLAP is ideal for data mining, business intelligence and complex analytical calculations, as well as business reporting functions like financial analysis, budgeting and sales forecasting.

OLAP is optimized for conducting complex data analysis for smarter decision-making. OLAP systems are designed for use by data scientists, business analysts and knowledge workers, and they support business intelligence (BI), data mining and other decision support applications.

## **What is OLTP?**

[**Online transactional processing (OLTP)**](https://www.ibm.com/cloud/learn/oltp) enables the real-time execution of large numbers of database transactions by large numbers of people, typically over the Internet. OLTP systems are behind many of our everyday transactions, from ATMs to in-store purchases to hotel reservations. OLTP can also drive non-financial transactions, including password changes and text messages.

OLTP, on the other hand, is optimized for processing a massive number of transactions. OLTP systems are designed for use by frontline workers (e.g., cashiers, bank tellers, hotel desk clerks) or for customer self-service applications (e.g., online banking, e-commerce, travel reservations).