Data Engineer FellowShip with Bytewise Limited.

Meer Danish 15/March/2023

horizontal line

Task # 2

# What is a data mart?

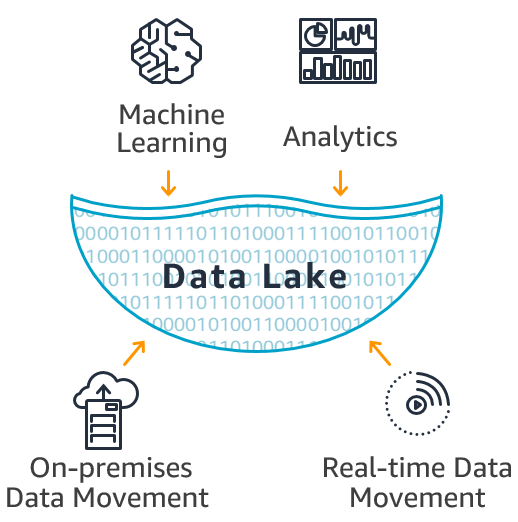
## Data Mart:

A data mart is a data storage system that contains information specific to an organization's business unit. It contains a small and selected part of the data that the company stores in a larger storage system. Companies use a data mart to analyze department-specific information more efficiently.

# What is a data lake?

## Data Lake:

A data lake is a centralized repository that allows you to store all your structured and unstructured data at any scale. You can store your data as-is, without having to first structure the data, and run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and machine learning to guide better decisions.



# 

# What is a data mesh?

## Data Mesh:

Data mesh is a new approach to thinking about data based on a distributed architecture for data management. The idea is to make data more accessible and available to business users by directly connecting data owners, data producers, and data consumers. Data mesh aims to improve business outcomes of data-centric solutions as well as drive adoption of modern data architectures.

# DWH vs Data Lake?

## Data Warehouse:

* A data warehouse contains structured data that has been cleaned and processed, ready for strategic analysis based on predefined business needs.
* Data from a data warehouse is typically accessed by managers and business-end users looking to gain insights from business KPIs, as the data has already been structured to provide answers to pre-determined questions for analysis.
* Data visualization, BI, data analytics.
* In a data warehouse, the schema is defined before the data is stored. This lengthens the time it takes to process the data, but once complete, the data is at the ready for consistent, confident use across the organization.
* ETL (Extract, Transform, Load). In this process, data is extracted from its source(s), scrubbed, then structured so it's ready for business-end analysis.
* Data warehouses cost more than data lakes, and also require more time to manage, resulting in additional operational costs.

## Data Lake:

* A data lake contains all an organization's data in a raw, unstructured form, and can store the data indefinitely — for immediate or future use.
* Data from a data lake — with its large volume of unstructured data — is typically used by data scientists and engineers who prefer to study data in its raw form to gain new, unique business insights.
* Predictive analytics, machine learning, data visualization, BI, [big data analytics](https://www.qlik.com/us/data-analytics/big-data-analytics).
* Schema is defined after the data is stored in a data lake vs data warehouse, making the process of capturing and storing the data faster.
* ELT (Extract, Load, Transform). In this process, the data is extracted from its source for storage in the data lake, and structured only when needed.
* Storage costs are fairly inexpensive in a data lake vs data warehouse. Data lakes are also less time-consuming to manage, which reduces operational costs.

# OLTP vs OLAP?

## OLAP:

Online Analytical Processing (OLAP): Online Analytical Processing consists of a type of software tools that are used for data analysis for business decisions. [OLAP](https://www.geeksforgeeks.org/olap-servers/) provides an environment to get insights from the database retrieved from multiple database systems at one time. **Examples –** Any type of Data warehouse system is an OLAP system.

**The uses of OLAP are as follows:**

* Spotify analyzed songs by users to come up with a personalized homepage of their songs and playlist.
* Netflix movie recommendation system.

## OLTP:

[Online transaction processing](https://www.geeksforgeeks.org/on-line-transaction-processing-oltp-system-in-dbms/) provides transaction-oriented applications in a [3-tier architecture](https://www.geeksforgeeks.org/introduction-of-3-tier-architecture-in-dbms-set-2/). OLTP administers the day-to-day transactions of an organization.

**Uses of OLTP are as follows:**

* ATM center is an OLTP application.
* OLTP handles the ACID properties during data transactions via the application.
* It’s also used for Online banking, Online airline ticket booking, sending a text message, and adding a book to the shopping cart.