**ASSESSMENT-27**

**Name: Vinutha**

**Date: 05/01/2024 (Friday)**

# **Azure Data Bricks**

## **What is Lakehouse**

* Lakehouse is a one stop destination for all the data, analytics and AI needs.
* It leverages the scalability and flexibility of data lakes and the reliability and performance of data warehouses, allowing organizations to store and process structured and unstructured data in a single system.
* The Lakehouse platform is built on top of Delta Lake, which is an open-source storage layer that adds reliability, scalability, and performance to data lakes.
* It provides ACID transactions, schema enforcement, and indexing to data lakes, making them more suitable for data warehousing workloads.

**features of Lakehouse:**

1. ACID transactions
2. Schema enforcement
3. Version control
4. Unified metadata management
5. Scalability & Performance
6. Flexibility

# **Architectural component of Lakehouse:**

## **Delta Lake**

Delta Lake provides a powerful set of features for managing data in a lakehouse architecture.

It is a optimized storage layer that allows users to store and manage data in various file formats, such as Parquet, Delta Lake, and ORC.

Along with provides support for indexing and partitioning to optimize data retrieval.

Delta Lake also provides a transaction log, which enables ACID transactions and allows for data versioning and rollbacks.

1. **Credits — Databricks**

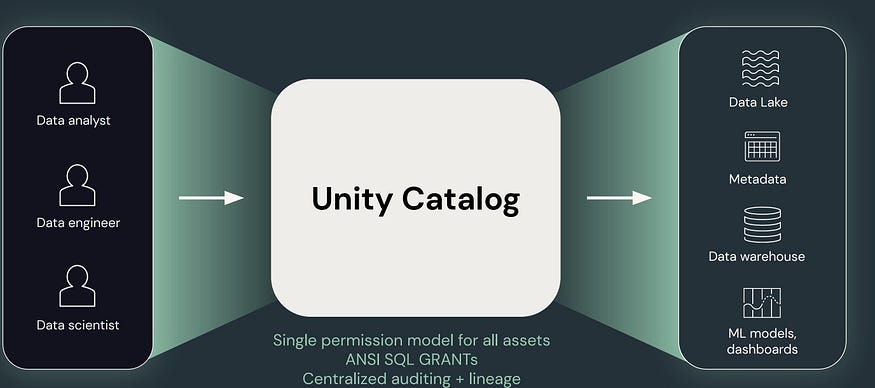
Delta lake enforce data governance policies and maintain data consistency across the organization by allowing users to apply schema enforcement on top of their data.

This ensures that data is of high quality and any changes made to the schema are tracked and managed in a centralized manner.

## **Unity Catalog**

Unity Catalog is a Unified Governance for All Data & AI Assets including files, tables, machine learning models and dashboards in your lakehouse on any cloud.

It is designed to help organizations manage and organize their data in a scalable and efficient manner.



**The three Data Ingestion Layers are:**

1. **Bronze Layer**
2. **Silver Layer**
3. **Gold Layer**

# **Azure Data Factory**

* Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.
* ADF does not store any data itself.
* It allows you to create data-driven workflows to orchestrate the movement of data between supported data stores and then process the data using compute services in other regions or in an on-premise environment.
* It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.
* The Data Factory service allows you to create data pipelines that move and transform data and then run the pipelines on a specified schedule (hourly, daily, weekly, etc.).
* This means the data that is consumed and produced by workflows is time-sliced data, and we can specify the pipeline mode as scheduled (once a day) or one time.
* Azure Data Factory pipelines (data-driven workflows) typically perform three steps.