Coding Challenge-4.2
Azure Databricks

Name-Ajay Chaudhary
Batch-Data Engineering(Batch-1)

Explain Overview of 3 level namespace and creating Unity Catalog objects.

Unity Catalog provides centralized access control, auditing, lineage, and data discovery capabilities across Azure Databricks workspaces.

The Unity Catalog object model in Azure Databricks provides a structured and hierarchical approach to organizing and managing metadata about data assets within the platform. Here's a brief overview of the Unity Catalog object model:

1. Metastore:
- The top-level container for metadata.
- Each metastore exposes a three-level namespace (catalog.schema.table) for organizing data.
- It serves as the overarching repository for all metadata within Azure Databricks.

2. Catalog:
- The first layer of the object hierarchy, used for organizing data assets.
- Within the catalog, you can group related data objects together for better management and organization.

3. Schema:
- Also known as databases, schemas are the second layer of the object hierarchy.
- Schemas contain tables and views, providing a logical separation of data assets based on their purpose or domain.

4. Tables, Views, and Volumes:
- These are the lowest-level data objects in the hierarchy.
- Tables represent structured data, views offer virtual representations of data, and volumes provide governance for non-tabular data.
- Tables and views contain the actual data or metadata definitions, while volumes provide governance capabilities for managing non-tabular data assets.

5. Models:

- Although not strictly data assets, registered models can also be managed within the Unity Catalog.
- Models reside at the lowest level in the object hierarchy alongside tables, views, and volumes.
- Models represent machine learning models or analytical artifacts that have been registered for reuse or management purposes.

The Unity Catalog object model provides a comprehensive framework for organizing, accessing, and managing metadata about various data assets, including structured tables, views, volumes, and even machine learning models. It facilitates centralized access control, auditing, lineage tracking, and data discovery capabilities across Azure Databricks workspaces, enhancing data governance and management within the platform.