

Name: Rohan Vinayak Chaudhari

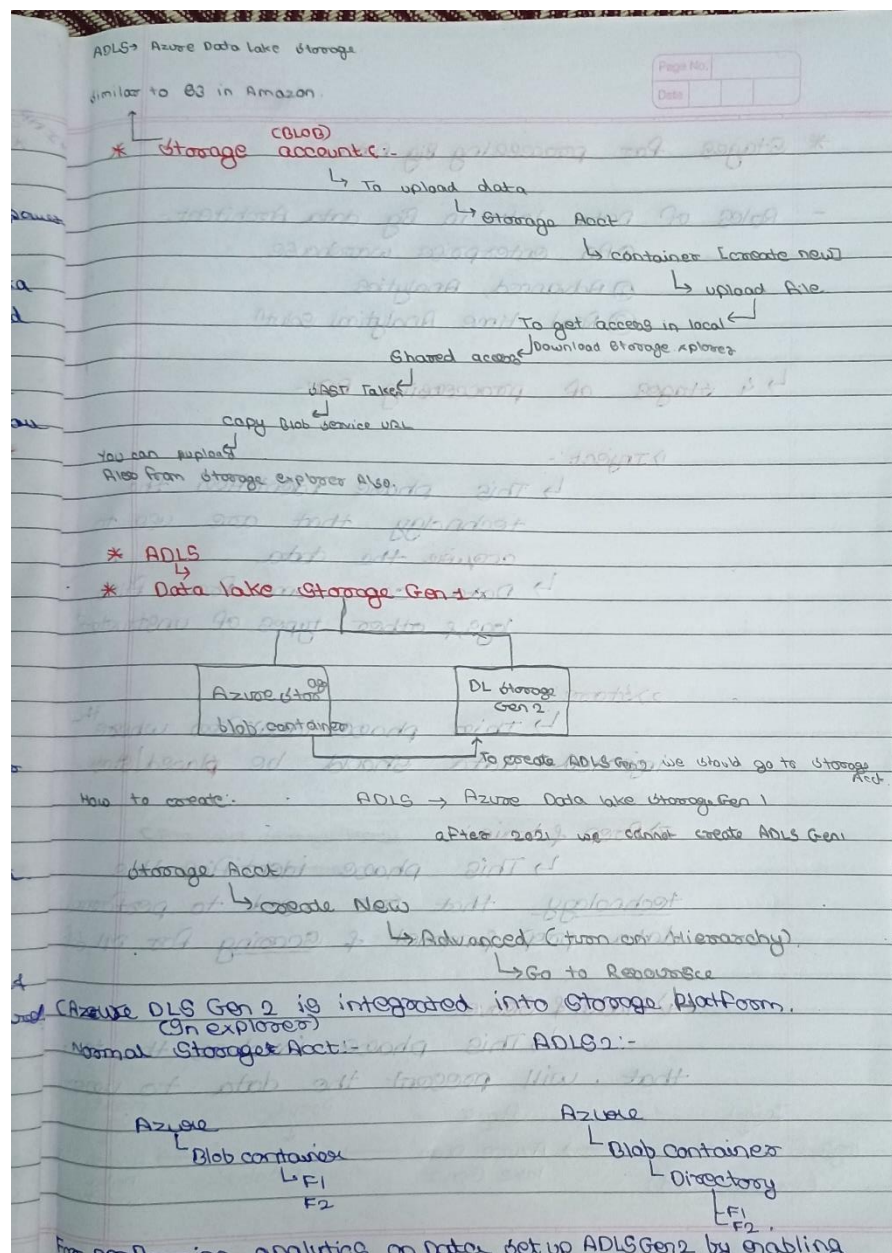
Batch: Data Engineering

Date: 16/02/2024

Topic: Azure DataBricks

Solution:

1. Azure Databricks (ADLS):



* Stages for processing Big Data

- Roles of Data lake in Big data Architect:-

- ① An enterprise warehouse
- ② Advanced Analytics
- ③ Real time Analytical solutiⁿ

↳ 4 stages of processing BD:-

1) Inject:-

↳ This phase identifies the technology that are use to acquire the data

↳ Data can come from Files, logs, & other types of unstructured

2) Store:-

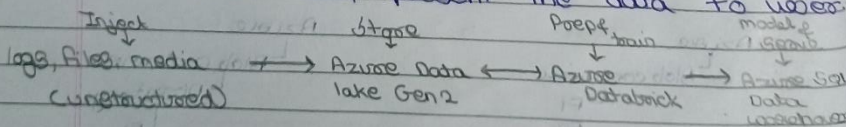
↳ This phase identifies where the injected data should be placed/store.

3) Prep & train:-

↳ This phase identifies the technology that are used to performed data prep & model & Scoring for ML solⁿ

4) Model & Serve:-

↳ This phase involves the tech that will present the data to users.



Deploying Storage account:

The screenshot shows the Microsoft Azure portal interface. At the top, the header includes the Microsoft Azure logo, a search bar, and the user's email address (azuser1076_mml.local@...). Below the header, the main content area displays the 'testing1076adlsgen2_1708066435867' deployment. The deployment status is 'Complete', indicated by a green checkmark. The deployment details show the name, subscription, resource group, start time, and correlation ID. A 'Go to resource' button is visible. On the right side, there are several informational cards: 'Cost Management' (Get notified to stay within your budget), 'Microsoft Defender for Cloud' (Secure your apps and infrastructure), 'Free Microsoft tutorials' (Start learning today), and 'Work with an expert' (Azure experts are service provider partners).

Data Lake is enabled:

The screenshot shows the Microsoft Azure portal interface for the 'testing1076adlsgen2' Storage account. The left sidebar contains navigation options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Data storage, Containers, File shares, Queues, Tables, Security + networking, and Networking. The main content area displays the 'Essentials' section, which includes the account's resource group, location, primary/secondary locations, subscription ID, disk state, and performance/replication details. Below the Essentials section, the 'Properties' tab is active, showing the 'Data Lake Storage' settings. The 'Data Lake Storage' section indicates that 'Hierarchical namespace' is enabled, 'Default access tier' is 'Hot', 'Blob anonymous access' is enabled, 'Blob soft delete' is enabled (7 days), and 'Container soft delete' is enabled (7 days). The 'Security' section shows that 'Require secure transfer for REST API operations' is enabled, 'Storage account key access' is enabled, 'Minimum TLS version' is 'Version 1.2', and 'Infrastructure encryption' is disabled.

Created container:

Microsoft Azure

Search resources, services, and docs (G+)

azuser1076_mm1local@...
BHT (BHTLONMICROSOFT.COM)

Home > testing1076adlsgen2

testing1076adlsgen2 | Containers

Storage account

Search

Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state	
<input type="checkbox"/> \$logs	16/02/2024, 12:24:29	Private	Available	...
<input type="checkbox"/> datalake1076	16/02/2024, 12:30:56	Private	Available	...

Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Data storage
Containers
File shares

Uploaded file in container:

Microsoft Azure

Search resources, services, and docs (G+)

azuser1076_mm1local@...
BHT (BHTLONMICROSOFT.COM)

Home > testing1076adlsgen2 | Containers >

datalake1076

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: datalake1076

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state	
<input type="checkbox"/> output_file.csv	16/02/2024, 17:17:41	Hot (Inferred)		Block blob	61.42 KiB	Available	...

Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata

Successfully uploaded blob(s)
Successfully uploaded 1 blob(s).