

Name : Parth Nandedkar

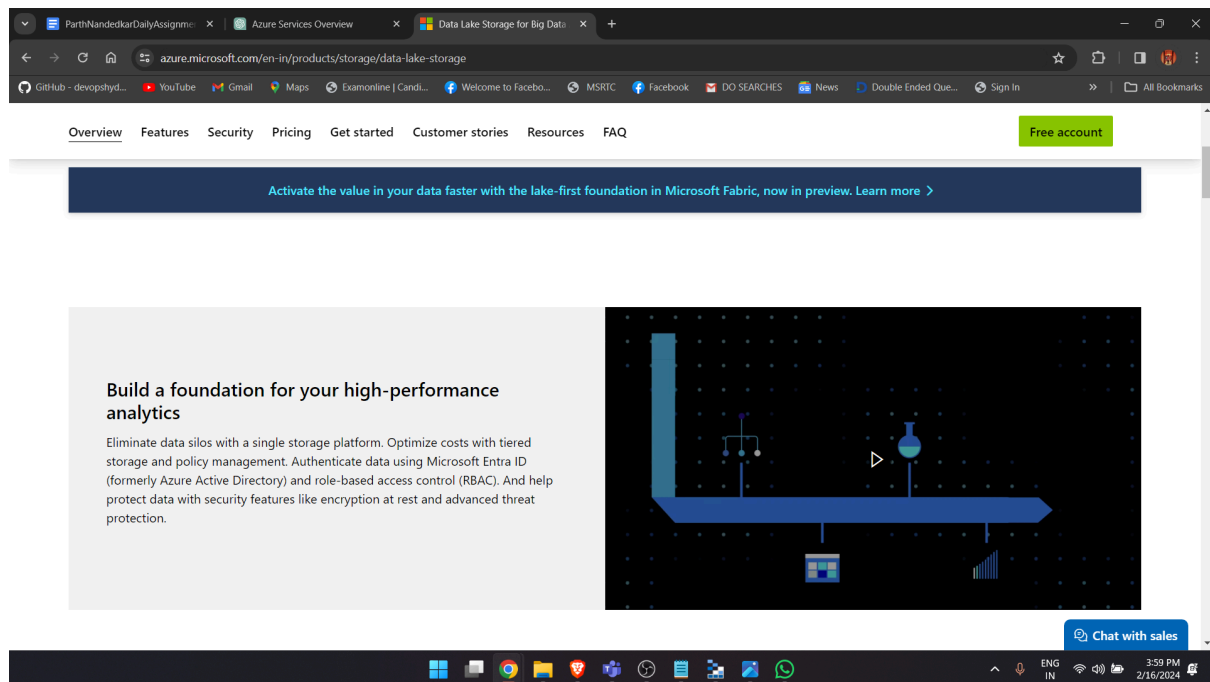
Date : 16 Feb 2024

Topics : Azure Data Lake Storage

Batch : Data Engineering Batch-1

Azure Data Lake Storage (ADLS):

Azure Data Lake Storage (ADLS) is a highly scalable and secure cloud-based storage solution optimised for big data analytics. It's part of the Azure Data Lake suite, designed to handle large volumes of structured and unstructured data for analytics, machine learning, and other data-intensive workloads.



Handwritten Notes :

16/02/2024

Data Lake storage Gen1

ADLS → Azure Data Lake Storage →

Service Name → Azure data lake storage gen-1

If you want to store data without performing analysis on the data, set the hierarchical Namespace option to Disabled to set up storage account as an azure blob storage account. You can use blob storage to archive rarely used data or to store website assets such as images & media.

If you are performing analytics on data ; set hierarchical Namespace to Enabled. Because Azure Data Lake Storage Gen2 is integrated into Azure data storage platform.

Stages for Processing

- An enterprise data warehouse
- Advanced structural data
- A real-time analytical solution.

Creating data lake on Azure Storage :

The image shows two screenshots from the Azure portal. The top screenshot is the 'Create a storage account' wizard, and the bottom screenshot is the 'workspace1064' storage account overview page.

Create a storage account wizard:

- Advanced** tab is selected.
- Hierarchical Namespace:** Enable hierarchical namespace is checked.
- Access protocols:** Enable SFTP and Enable network file system v3 are unchecked.
- Blob storage:** No options are visible.
- Buttons: Review, Previous, Next: Networking.

workspace1064 Storage account Overview:

- Essentials:**
 - Resource group: Test1064
 - Location: eastus
 - Primary/Secondary Location: Primary: East US, Secondary: West US
 - Subscription: Azure subscription 1
 - Subscription ID: 984f097c-963c-4eb6-a20d-839457ae9f08
 - Disk state: Primary: Available, Secondary: Available
 - Performance: Standard
 - Replication: Read-access geo-redundant storage (RA-GRS)
 - Account kind: StorageV2 (general purpose v2)
 - Provisioning state: Succeeded
 - Created: 2/16/2024, 4:24:21 PM
- Tags:** Add tags
- Properties:**
 - Data Lake Storage:**
 - Hierarchical namespace: Enabled
 - Default access tier: Hot
 - Blob anonymous access: Disabled
 - Blob soft delete: Enabled (7 days)
 - Container soft delete: Enabled (7 days)
 - Versioning: Disabled
 - Security:**
 - Require secure transfer for REST API operations: Enabled
 - Storage account key access: Enabled
 - Minimum TLS version: Version 1.2
 - Infrastructure encryption: Disabled
- Networking:**

Connecting account to local :

