Azure Data lake Storage

ADLS

- In Azure blob Storage, you can store large amounts of unstructured data in a flat namespace within a blob container.

- Azure DL build on blob Storage and optimizes I/O of high-volume data by using hierarchical namespace that organizes blob data into directories, and stores metadata about each directory and files within it.

⇒ If you want to store data without performing analysis on the data, set the ~~Hierar~~ Hierarchical namespace option to disabled to set up the Storage account as an Azure blob Storage account.

- You can also use blob Storage to archive rarely used data or to shared website assets such as images and media.

- If you are performing analytics on the data, set up the Storage account as an Azure DL Storage Gen 2 account by setting the Hierarchical Namespace option to Enabled.
- Because Azure Data lake Storage Gen 2 is integrated into the Azure Storage Platform, applications can use either the blob APIs or the Azure Data lake Storage Gen 2 file system APIs to access data.

(2)

classmate
Date
Page

Understanding the Stages for processing ~~dat~~ big data.

→ DL have a fundamental role in a wide range of big data architecture. These architecture can involve creation of :-

(i) An enterprise data warehouse
(ii) Advanced analytics against big data.
(iii) A real time solution.

There a 4 stages of processing data sol^n that are common to all architectures :-

1) Ingest :- Identifies the technology and processes that are used to acquire the source data.

2) Store :- Identifies where the ingested data should be placed.

3) Prep and Train :- Identifies the technology that are used to perform data preparation and model training.

4) Model and Serve :- Finally, the model and serve phase includes the technologies that will present the data to users. These technologies can include visualization tools such as Microsoft Power BI, or analytical data stores such as Azure Synapse Analytics.