16/02/2024

ADLS Extraction

→ If you want to store data without performing analysis on the data, set the hierarchical namespace given option to Disabled to set up the story account as an Azure Blob storage account. You can also use blob storage to achieve rarely used data (or) to store website assests such as images & media

→ If you are performing analytical on the data, set up the storage account as Azure data lake storage Gen2 account by setting the Hierarchical Namespace option to Enabled. Because Azure Data lake storage platform applications can use either the blob API's (or) Azure Data lake storage Gen2 file system API's to access data

Understand the stages for processing big data:-

Data lakes have a fundamental role in wide range of big data architectures. These architectures can involve the creation of:-

- An enterprise data warehouse
- Advanced analytical against big data
- A real-time analytical solution

They are 4 ~~types~~ stages:-

① Ingest:-

→ This phase identifies the technology and process that are used to acquire the source data

→ Data can come from files, logs & other data types of unstructed data that must be put in data lake

→ Azure data factory (or) Azure synapse Analytics are the most appropriate technologies to use.

② store:-

→ It identifies where the ingest data should be placed

→ Azure data lake storage Gen2 provides a secure & scalable storage solution that is compatible with commonly used big data processing technologies

③ Prep and train:

→ This identifies the technologies that are used to perform data preparation and model training and scaling for machine learning solutions. Eg:- Azure Synapse Analytics, Azure Databricks.

④ Model & serve:-

→ Finally the model & serve phase involves 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

→ often a combine of multiple technologies will be used depending on the business requirements