Day 5 notes

Task:

It represents a unit of logic to be run as a step in a job.They can range in complexity and can include

- Notebook
- JAR
- SQL queries
- A DLT pipeline
- Another job
- Control of flow tasks

Control flowoptions:

- Triggers
- Retries
- Run if cond tasks
- If/else
- For each tasks
- Duration thresholds
- Concurencey

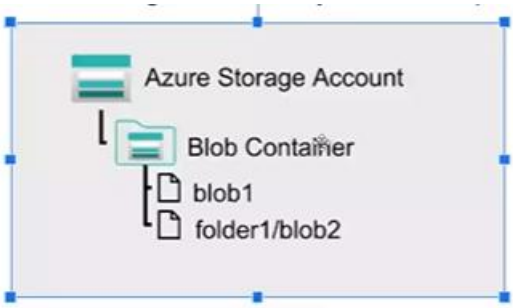
Trigger types

- Scheduling
- File arrival
- Gg

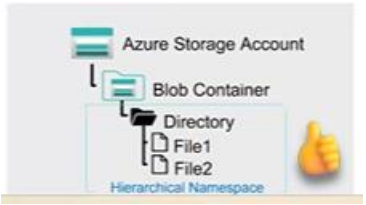
ADLS

Azure Data Lake storage

- In azure blob storage we can store large amount of unstructured data
- Storage in azure-→Blob
- “In blob manageability blob are always in single level hierarchal in a **FLAT** namespace”
- Azure storage--→(adls)



- Azure datalake storage zen2---→(adlszen2)
- “Hierarchical namespace”
- Azure data can be accessed via web by https, http
- This structure allows operations like directory renaming, deleting

- 

- Store data without performing analysis disable hierarchical namespace
- Adlszen2 integrates with data factory & synapse & azdevops for analytics

Understanding the stages of BIG DATA processing:

Data lakes have fundamental role in architectures:

Four stages/phases:

1. **Ingest**----------→ acquiring source of data
   Eg: Apache Kafka for streaming applns-→process financial transactions in real time, track and monitor vehicles
2. **Store**----------→ where the ingested data is placed
   Eg: adlsgen2
3. **Prep and train**----------→ identifies technologies that are used to perform data preparation and model training.
   Eg: Azure Synapse Analytics, Azure Databricks, Azure HDInsight
4. **Model and serve**----------→this phase involves the technologies that will present the data to users.

   Eg:These technologies provide visualization tools like Microsoft Power BI, Analytical data stores like Azure synapse analytics

---

Subscription

MML Learners

Resource group

rg-azuser2361_mml.local-PhZfL

Location

East US

Storage account name

datastoresinadb

Primary service

Azure Blob Storage or Azure Data Lake Storage Gen 2

Performance

Standard

Replication

Locally-redundant storage (LRS)

https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace