25 Oct

Day 20

AZURE

(1) Create a Resource Group

- A container that holds resources for any Azure solution like Virtual machines, databases, accounts.
- It helps managing & organising resources as a SINGLE UNIT
  - ↳ allowing you to deploy, monitor, control access costs, lifecycle (create, manage, delete) together

Ex WebApp-RG
- ~~A VM~~
- storage account
- network

} If u delete RG → all 3 resources gets deleted automatically

:. Shared lifecycle

(2) Create a resource now, from inside RG

1. Storage Account
   - A container that holds all your storage data - blobs, files, tables

Azure Blobs → Cloud based object storage solution for unstructured data
- Can store vast amount of data like images, videos, audio
- highly scalable & cost effective

Azure Data Lake → Built on top of blob
- Optimized for big data analytics with a hierarchical namespace
  - ↳ organizes data like a file system not a blob

Normal Blob: → Virtual Hard drive in cloud.

- Data ↳ All files are stored as objects, just keys & paths. No true folders.

With HNS: Create folders within folders
↳ We can have real directories, subdirectories

∴ Data lake storage Gen 2

- file system + folder organization on top of that hard drive

→ Redundancy Options

LRS (Locally redundant storage)

- lowest cost
- Data is replicated 3 times within a single datacenter.

Why 3? → If 2: If 1 fails, only 1 remains so high risk of data loss
If 3+: Extra cost

∴ Many cloud providers use triple replication as a balance of cost, reliability, availability

Region: A geographic area, where Azure data centers are located Ex East US / Canada West
↳ choose region based on latency

Availability Zone: A collection of physically separated datacenters within a region designed to be isolated from failures of other zones.

Data Centre: Actual facility where servers & storage & networking exists

LRS

If a server or disk fails
↳ other copies in datacentre keeps us data safe
But if datacentre fails (fire / power outage)
↳ All 3 copies get affected

ZRS → Zone Redundant storage
Data is replicated across 3 zones in a region
Limitation → regional disaster

GRS → Geo-redundant storage
Data is replicated to another region

GZRS → Geo Zone redundant storey
- Highest durability & availability
- data remains safe if even if entire zone/region fails.

Redundancy → keeping extra copies of data or resources → so that if something fails, System continues to work

26 Oct

Day 21

(2) Data factory

↳ Cloud based ETL (Extract, Transform, Load) service used to move & transform data from various sources to destinations, for Analytics or storage

→ Used to build data pipelines → to collect/clean data automatically.

Pipeline : is a group of activities that together perform a specific data workflow, (ETL).

SHIR Self Hosted Integration Runtime

↳ Sometimes data is not in Azure, maybe its in your company's on premises SQL server, or private Azure Network

ADF can't reach the data as its not public

∴ So u install SHIR app (a small software agent) on the computer/server inside ur network. This agent acts as a bridge b/w your data & ADF

↳ Data needs to connect to ADF in cloud, somehow

① Public Endpoint
- SHIR connects to ADF over public internet (but securely encrypted)
- simple setup, safe (outbound traffic only (SHIR → Azure)

② Private Endpoint
- SHIR connects to ADF privately using Azure Private links
- Traffic stays 100% private
- Used when company don't allow internet communication
- High security/sensitive data

Ex : A company store its data in onprem SQL server
① behind a firewall
② ADF can't access it directly.
③ Install SHIR on the system
(which can access SQL server data locally & send data securely to ADF)

Ex Public Endpoint → How its secure/encrypted?
Ex : I login to bank website,
its public but encrypted (https)

Environments

- On prem : means to company's own servers, databases, & storage & are physically hosted in their own building/datacentre not in cloud.
  Ex: your official SQL server
- cloud : Azure SQL Database
- Hybrid : onprem DB + ADF pipelines

HTTPS →

- "S" stands for secure
- Uses encryption to keep data private & safe