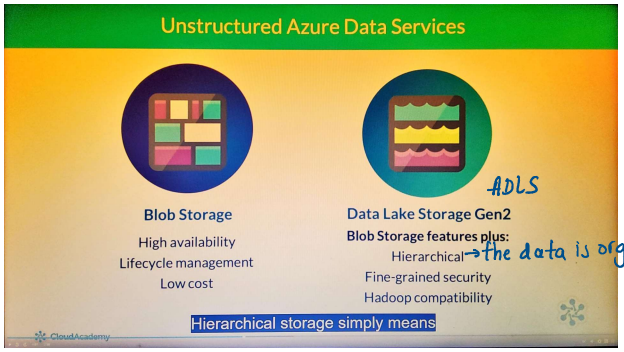
Using Azure Data Lake Storage Gen2

Friday, January 6, 2023 7:29 PM

Overview

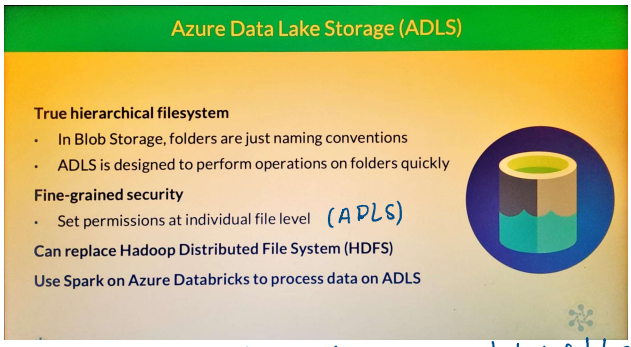
In fact, one common set up is to process data in the data lake. and then export it to the data warehouse.



ADLS is built on top of Blob storage, to get the best of both worlds.

ADLS

→ the data is organized into a tree of folders and files.



for Blob storage to operate on a simulated folder, it has to perform a separate operation on each file.

→ BS can only restrict access at the container level, rather than at the individual blob level.

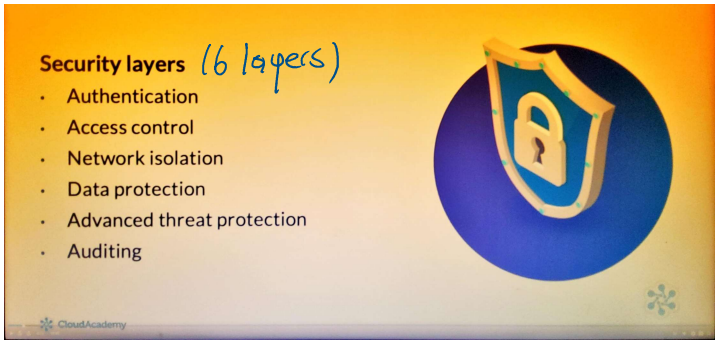
Directory ↔ Folder

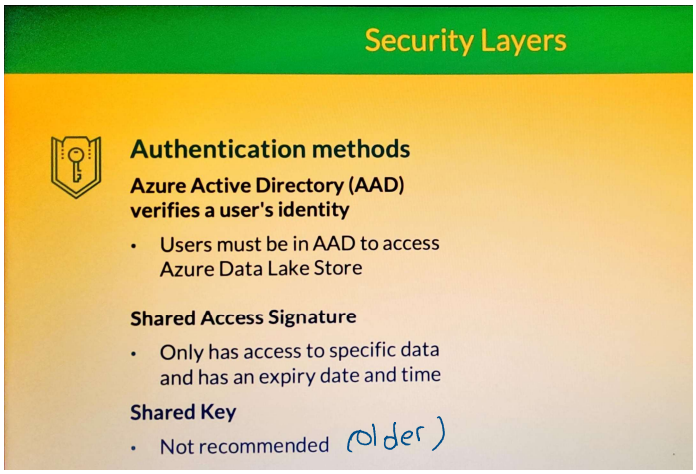
Container ↔ filesystem

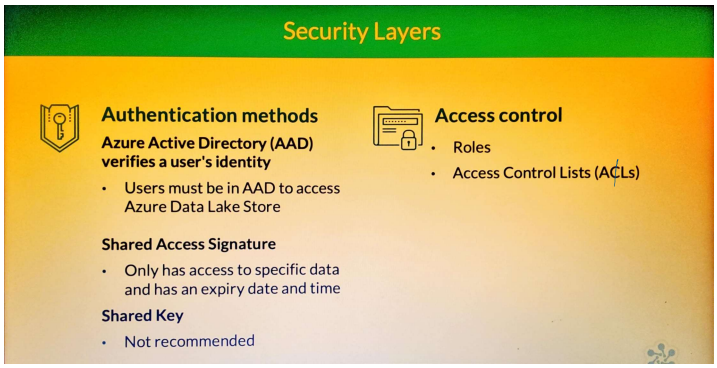
BS ADLS

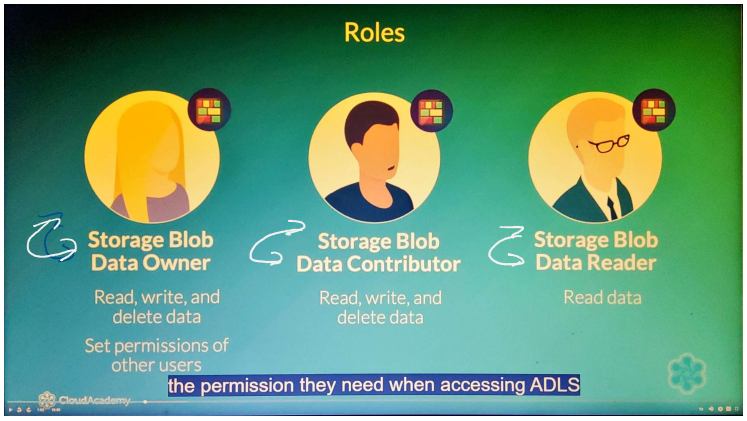
→ ADLS can seamlessly (perfectly) integrate. with the huge ecosystem of Hadoop SW.

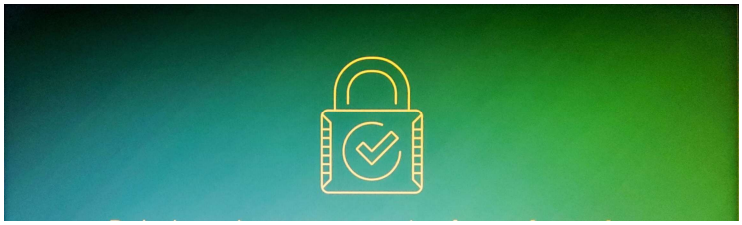
Security

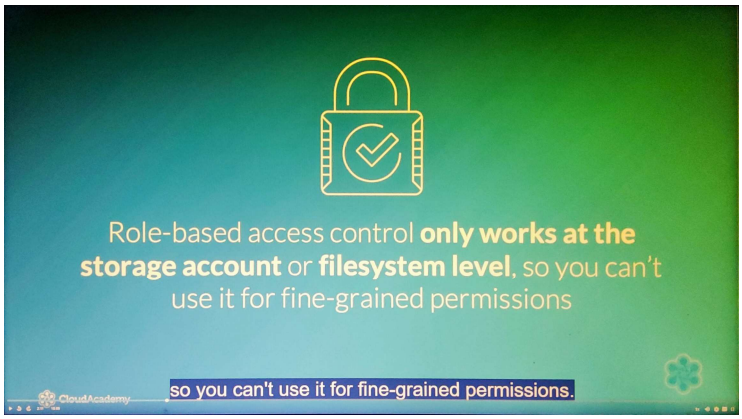

Security layers (6 layers)
Authentication
Access control
Network isolation
Data protection
Advanced threat protection
Auditing
CloudAcademy

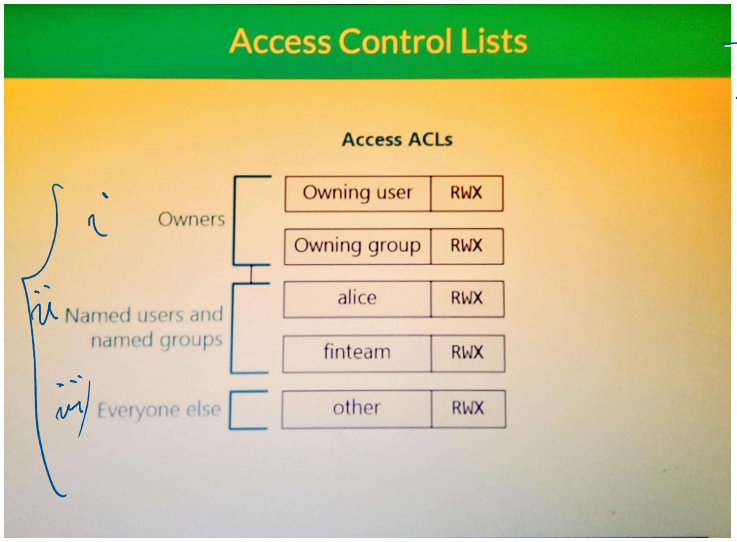

Security Layers
Authentication methods
Azure Active Directory (AAD) verifies a user's identity
Users must be in AAD to access Azure Data Lake Store
Shared Access Signature
Only has access to specific data and has an expiry date and time
Shared Key
Not recommended (older)


Security Layers
Authentication methods
Azure Active Directory (AAD) verifies a user's identity
Users must be in AAD to access Azure Data Lake Store
Shared Access Signature
Only has access to specific data and has an expiry date and time
Shared Key
Not recommended
Access control
Roles
Access Control Lists (ACLs)


Roles
Storage Blob Data Owner
Read, write, and delete data
Set permissions of other users
Storage Blob Data Contributor
Read, write, and delete data
Storage Blob Data Reader
Read data
the permission they need when accessing ADLS
CloudAcademy

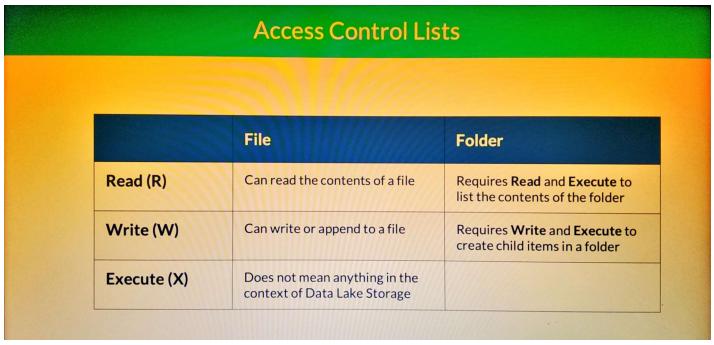




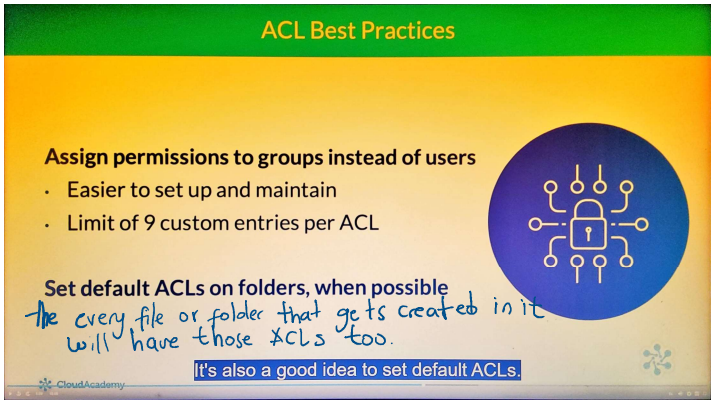


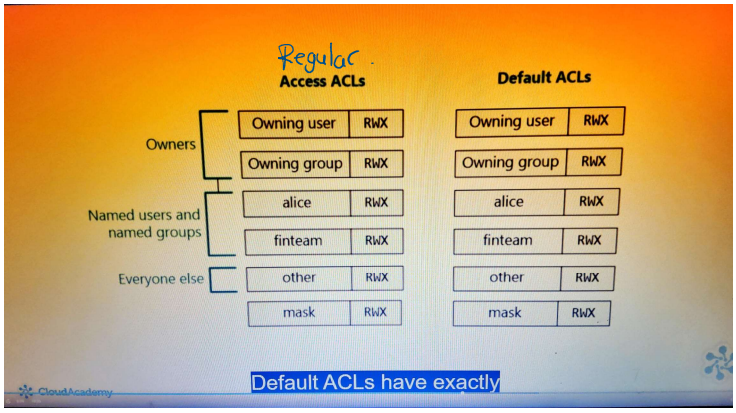
→ to handle permissions for files and folders

→ Each entry in an ACL specifies the read, write and execute permissions for a specific user or group.



	File	Folder
Read (R)	Can read the contents of a file	Requires **Read** and **Execute** to list the contents of the folder
Write (W)	Can write or append to a file	Requires **Write** and **Execute** to create child items in a folder
Execute (X)	Does not mean anything in the context of Data Lake Storage	





iii) Network Isolation: you can actually set up a firewall just for your datalake.

iv) Encryption: data protected (in transint - or not in transit) with Azure Storage Encryption or your own.

v) Defender: potential malicious

vi) Auditory: Activity log.

Ingesting