Data Factory stores pipeline-run data for only 45 days.

The term total cost of ownership (TCO) describes the final cost of owning a given technology.

For SQL (Core) API accounts, when analytical store is enabled, the default schema representation in the analytical store is well-defined. Whereas for Azure Cosmos DB API for MongoDB accounts, the default schema representation in the analytical store is full fidelity schema representation.

df.orderBy(col("request").desc())

Azure Synapse Studio notebook is purely Spark based.

copy data command work for public end point.

All IR support private networks.

Materialized views are prewritten queries with joins and filters whose definition is saved and the results persisted to a dedicated SQL pool. They are not supported by serverless SQL pools.

cosmos.olap is the method that connects to the analytical store in Azure Cosmos DB.

Within an Apache Spark Pool it is possible to configure a fixed size when you disable autoscaling. When you enable autoscale, you can set a minimum and maximum number of nodes in order to control the scale that you'd like. Once you have enabled autoscale, Synapse Analytics will monitor for you the requirement of resources of the load. Accordingly, it will scale the number of nodes up or down. There will be continuously monitoring depending on CPU usage, pending memory, free CPU, free memory, and the used memory per node when it comes to the metrics involved to make a decision to scale up or down. It checks these metrics every 30 seconds and makes scaling decisions based on the values. There's no additional charge for this feature.

Setting Global parameters in an Azure Data Factory pipeline, allows you to use constants for consumption in pipeline expressions.

Invoke spark.streams.active which is the correct syntax to view the list of active streams.

While you can deploy more than one Workspace in a VNet by keeping the associated subnet pairs separate from other workspaces, MS recommends that you should only deploy one workspace in any VNet.

SAS:

You can give access to the specific option which is available in storage account.

You can give selective access on to the services which is available in your storage account (like blob, file, queue).

why :

if you access storage account with access key, you will have access of all the service like blob, file, queue for that storage account but if you want to give access to the user only for blob then you can use SAS here instead of access key.

you can give permissions like read, write , delete ,list.

you can set time limit for SAS with the help of start time and end time.

Can set IP address which can access SAS.

One of the access key required to generate the SAS.

ADLS RBAC :

If you create a user in Azure AAD, then this new user will not be able to acess resources for that subscription initially. you have to provide access to that user with the help of RBAC. F

Example : if you want to provide access of data lake to new user , below steps are needed.

1. Go to resource and click on access control.

2. click add for "add role assignment" for new user by providing the user id( email id).

3. give the specific role like reader , contributor or owner.

Now user will have the access for this resource.

Still we have not given permission to container or object placed in that container.user will have access of that data lake resource but will not be able to see the data available in containers.

For this we need ACl.

ACL :

With the help of ACL, you can grant permission to the files or directories in gen2.

Example : If I want to give the access to a user for 1 container only , follow below steps:

1. go to storage account and go to access control.

2. Add role assignment (Storage blob data reader specificlly) for that user.

3. again go to storage account.

4. Select the container and click on manage access control list.

5. add user and provide access permissions for that container.

Now we have given permission on container level , now we have to provide the access on folder level for that we have to follow below steps.

1. go to container and select the folder.

2. click on manage access control list and add user.

This will give us permission on folder level but not on object level.

3. once permission is given to that folder, now you click on propagate access control list for giving permission for all that file which is available in that folder.

Virtual Network Service End Point :

If there is a virtual machine in azure and one app is running on it and the requirement is like this app wants to fetch some data from adls gen2 account, then it can only be through internet.

This app will fetch adls gen2 data using internet only even this virtual machine resides in virtual network.

One way to secure this data trasfer is using service endpoints.

we can limit the connectivity of this storage account to virtual network only.only resources in that virtual network can access the storage account

Manage Identity :

Azure Active Directory :

Used for Authentication

Key Vault :

it is uded to manage certificate, secrets and encryption keys.

example : you can store database password as secret in key vault.

ADF already encrypt data at rest with help of microsoft managed keys.you can also define your own key using key vault service. for the key vault, soft delete and setting of do not purge should enabled.also grant access to adf for get, wrap key and unwrep key.

Customer Managed Keys:

You can create your own key for encryption on data on azure services.

Example. if you have to use customer managed key for azure synpase work space.

step 1: create a key in key vault

step 2: create a new synapse resouce and in "work space encryption " use customer managed keys.

it can only be possible with new instances