Azure Synapse

Need to have a Data Warehouse – It helps us do analytics on the data that we have. The data is stored in a way where it is made to process high volumes of read requests.

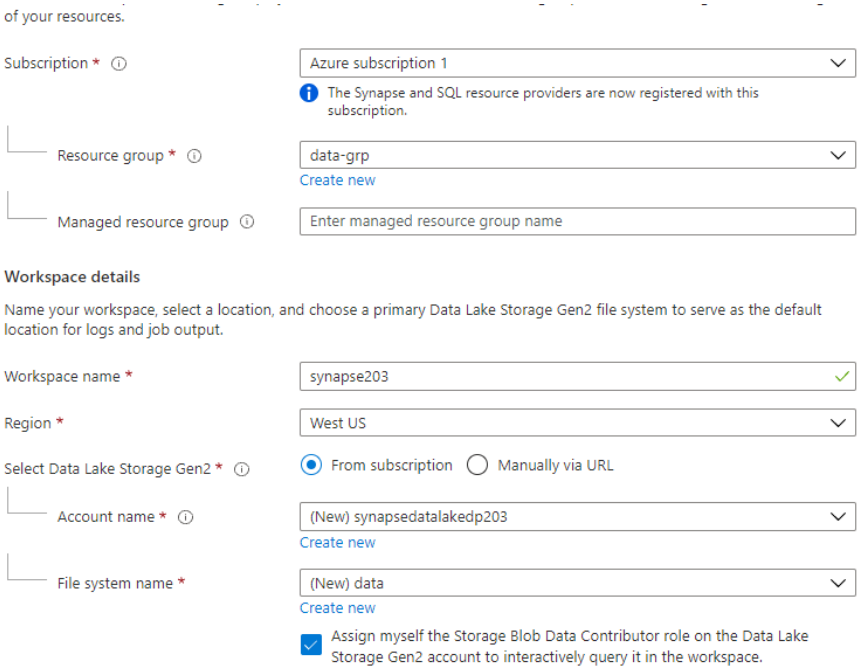
Synapse initially was just a data warehouse but now it is known as Azure Synapse Analytics. Now we can create warehouses with the help of SQL, integrate the data using pipelines and also use data from data lakes.

We can also use Spark for processing and the data and services like Azure monitor and Azure Active Directory with synapse.

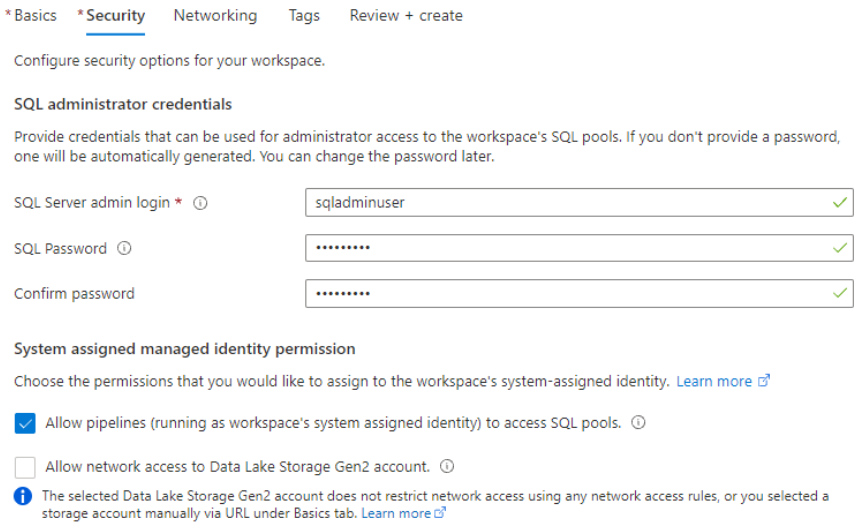
Creating an Azure Synapse Workspace

Go to your home screen > Create a Resource > Search for Azure Synapse Analytics > Create.

Enter : Your subscription, your resource group, *unique* workspace name, region, data lake gen2 details (new or old)



On the next screen give the password for your SQL Administrator Credentials, make sure “allow pipelines” checkbox is ticked.



Leave everything else as is and create.

Synapse Compute Options

There are different compute options – Serverless SQL pool and SQL pool.

Serverless SQL Pool

- You can user this option to perform quick adhoc analysis of data
- Can use T-SQL
- Can only create external tables but cannot persist the data
- Charged based on how much you use the service and how much data your process

SQL Pool

- User to build your warehouse
- Can use T-SQL
- Used if you want to persist the data
- Charged based on the data warehousing units (which includes things like compute, memory, etc.)

External tables - Can be defined in the Serverless pool and the dedicated SQL pool. We use external tables when the table data is lying in an external source, but the table definition is lying in Azure synapse. This is useful when you don't want to load the table on to the server itself.

For example if there are tables that exist on an external source and there's data on the sql server, then to perform a join operation between the two, an external table can be used.

There are a few important checks that need to be done in order to access the external data:

- We first need to have authorization to use the external source of data
- We then need to define the format of the external file that we want to use as an external table
- Finally, create the external table

Using External Tables

Open up your synapse dashboard and click on Open Synapse Studio. In the synapse studio you can use SQL commands against your Serverless SQL pool as well as dedicated SQL pool. Can create pipelines to integrate your data and just view your data as well.

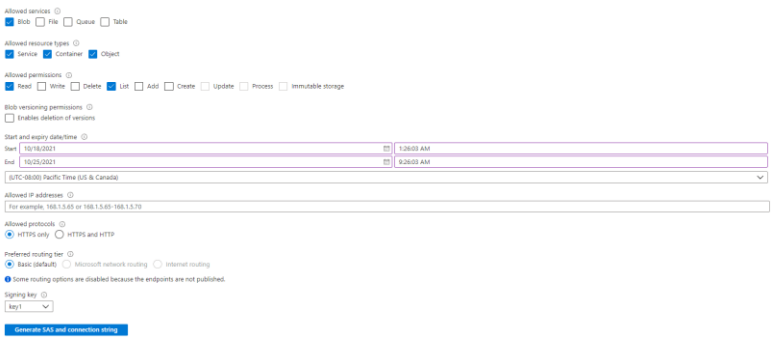
Executing a script on Azure Synapse Studio

In the left-hand menu, click on develop, click on the plus icon in the develop screen and select SQL Script out of the options given.

Name the script on the left and copy and paste this SQL script on to the editor.

Now, we would be running a series of commands –

- Firstly, we would run the create data base command to create a database in the serverless pool
- Change the database from the top right of the editor where master is selected (refresh if the newly created database is not showing)
- Next, we create a master key that would be used to encrypt the database scope credentials which will allow ourselves to use the file that we would be using in our Data Lake Gen2 account.
- Now to create the scope credentials we need to get the shared access signature like follows and copy the SAS token –



- Paste it in the SECRET variable and remove the ‘?’ from the front. Run the command for creating scoped credential.
-