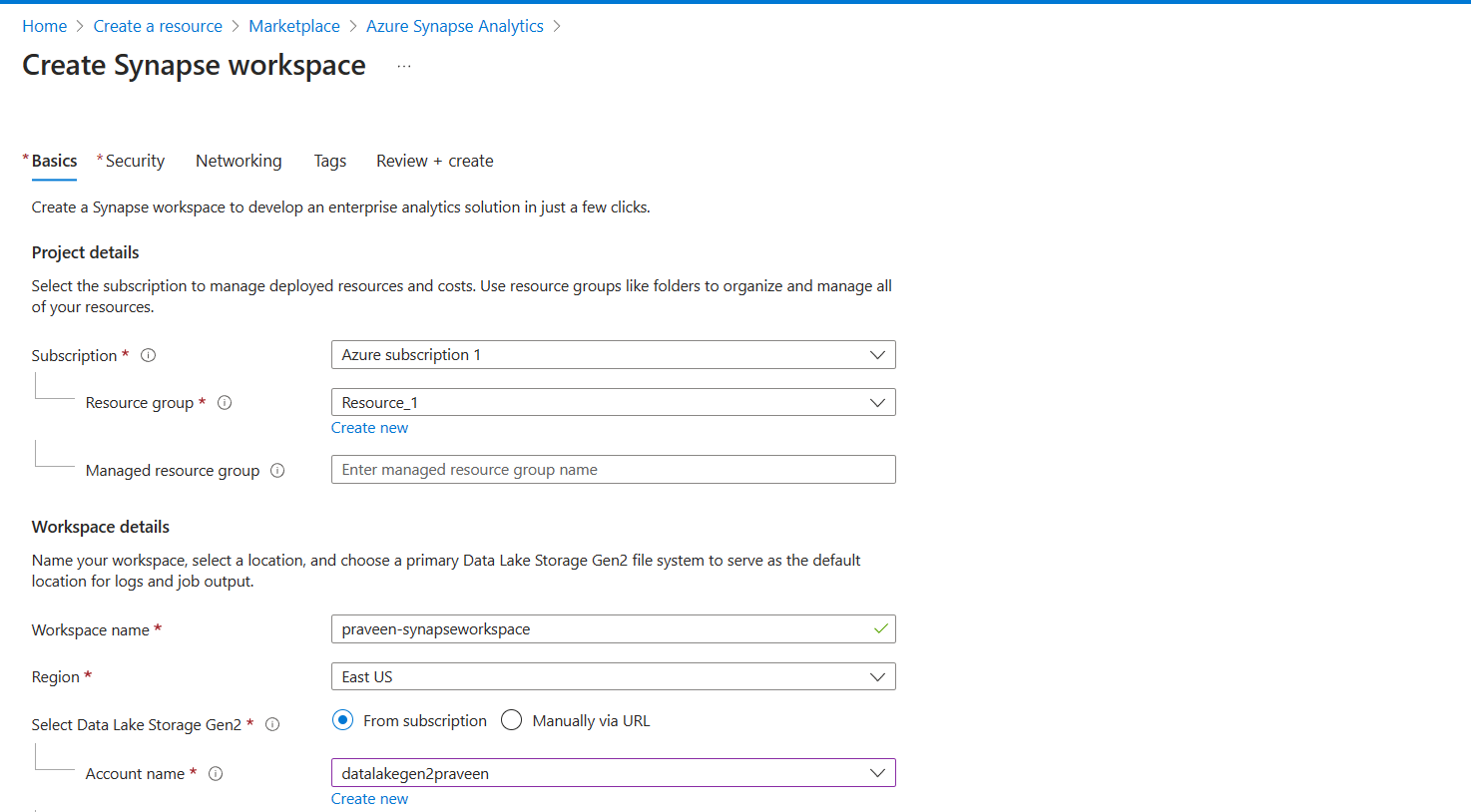
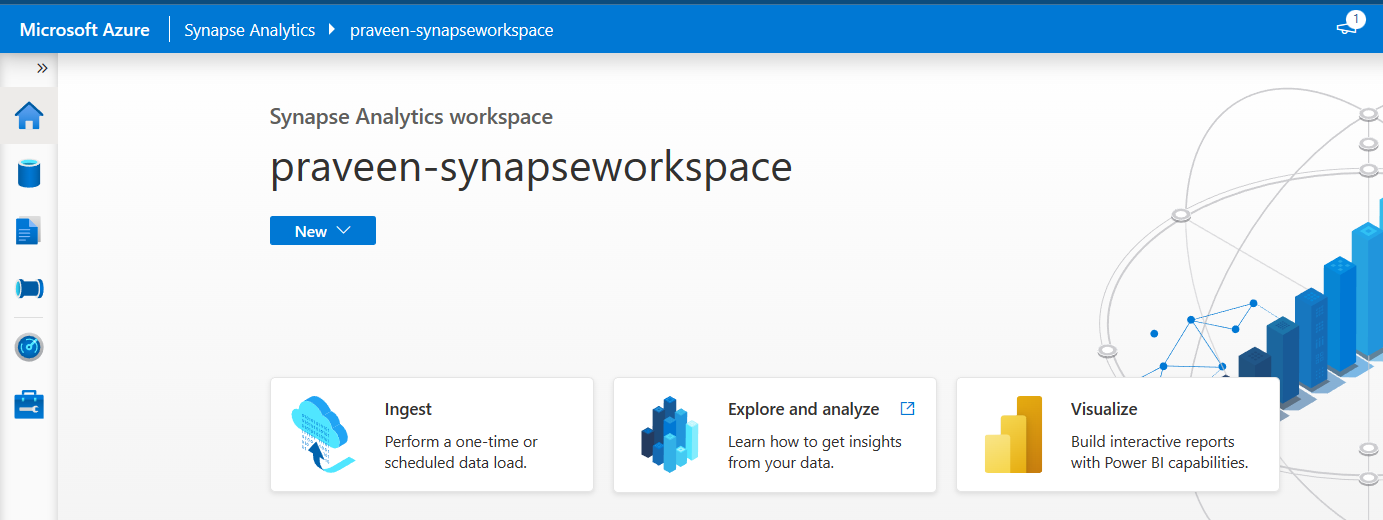
Task 1 – Creating Azure Synapse Workspace and launch a Serverless SQL pool

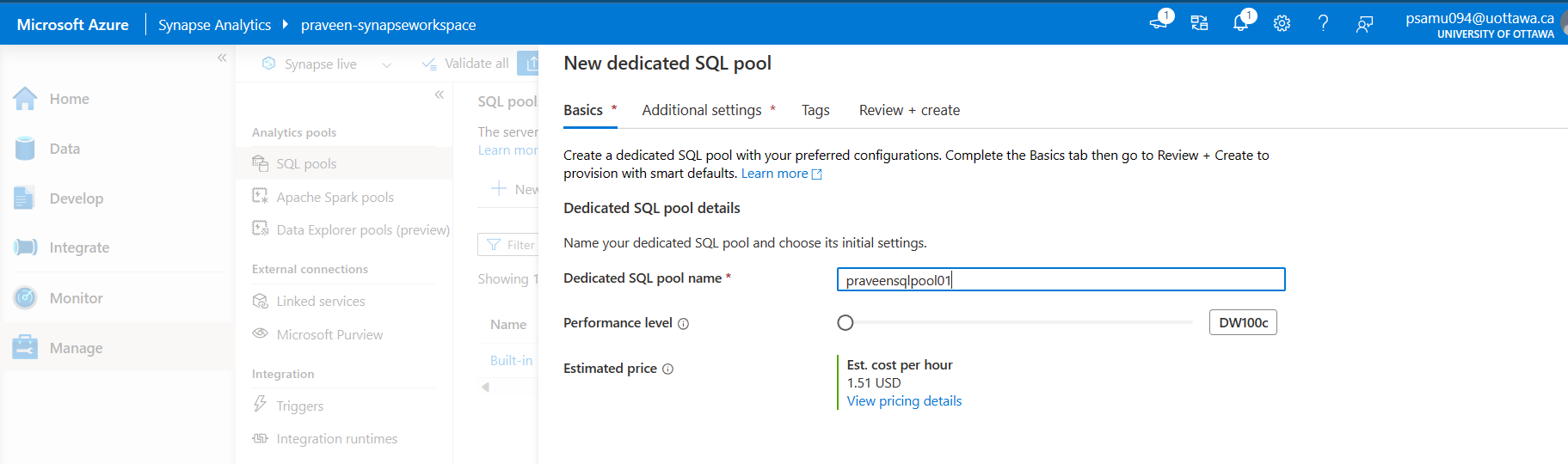


Sqladminuser

Praveen1234

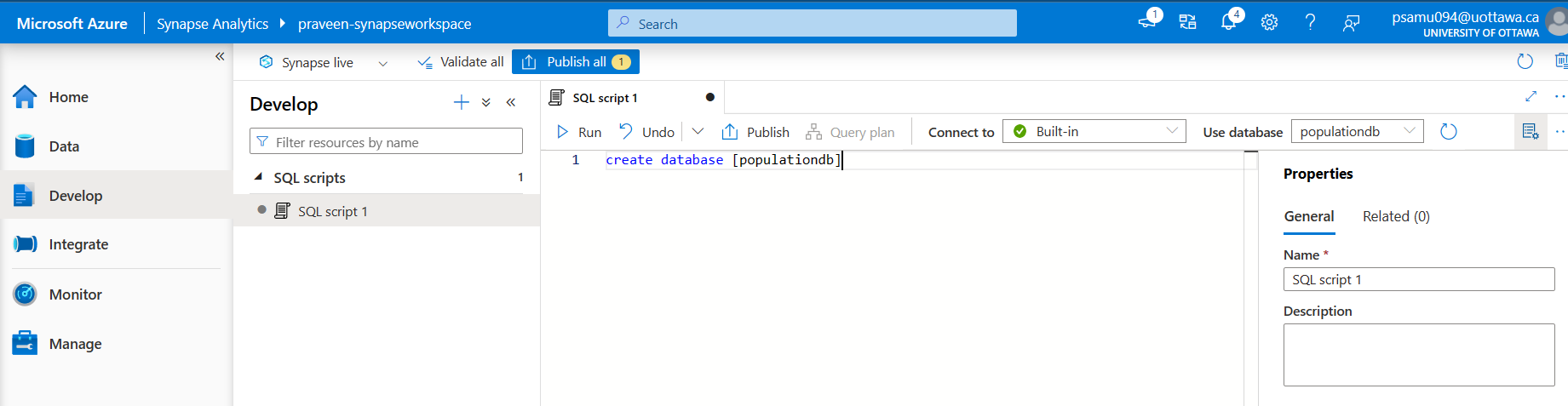


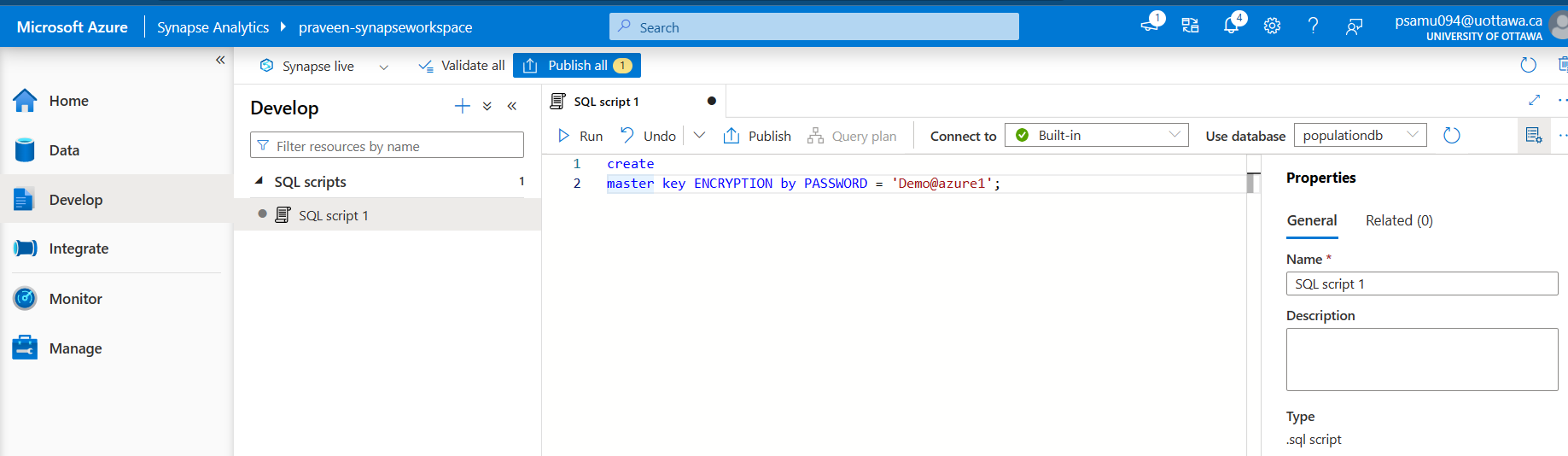
To create a new Dedicated SQL Pool, Manage -> SQL pools -> New



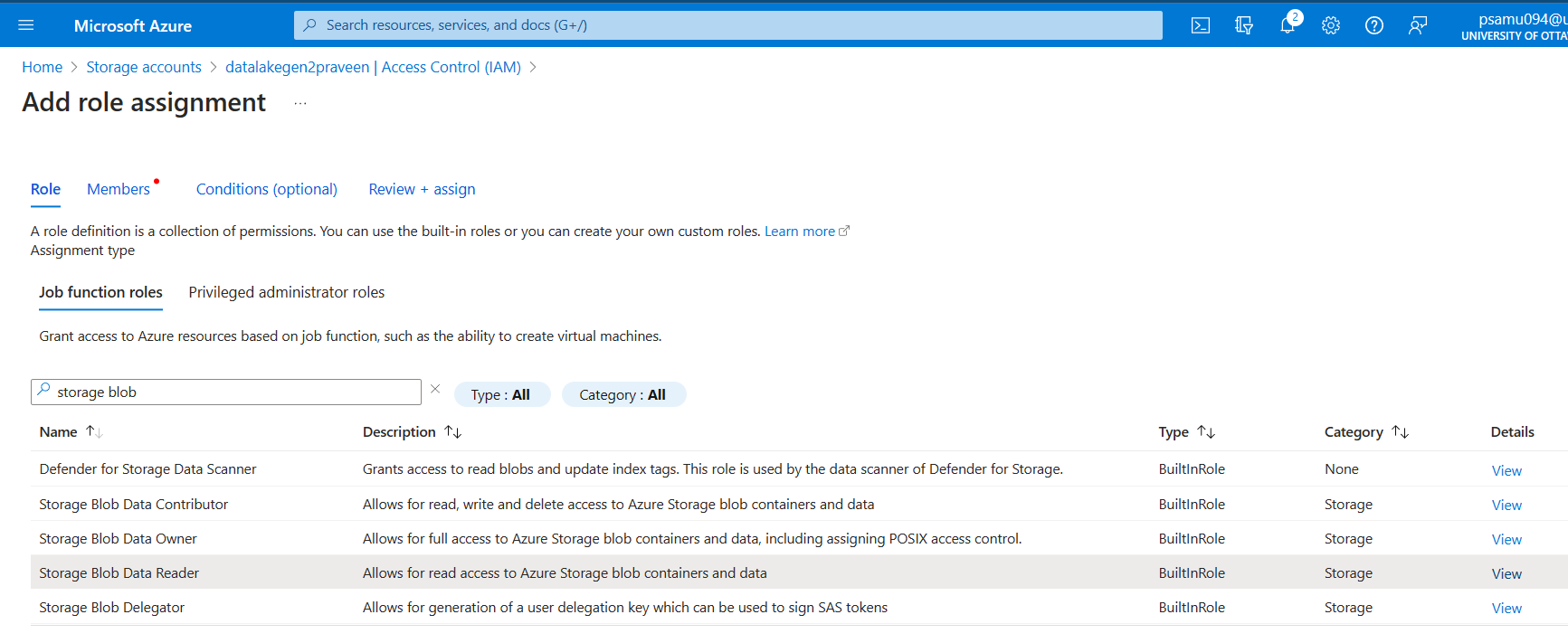
Task 2 – Loading data into external table

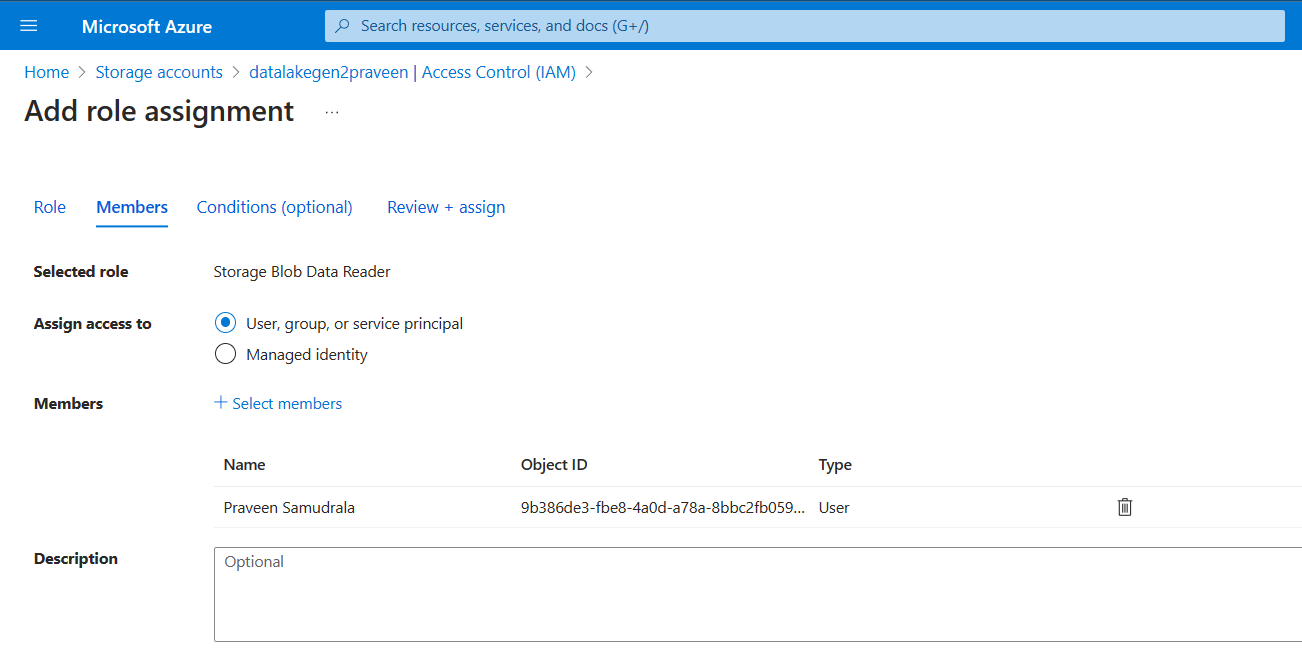
Create a database and connect to it



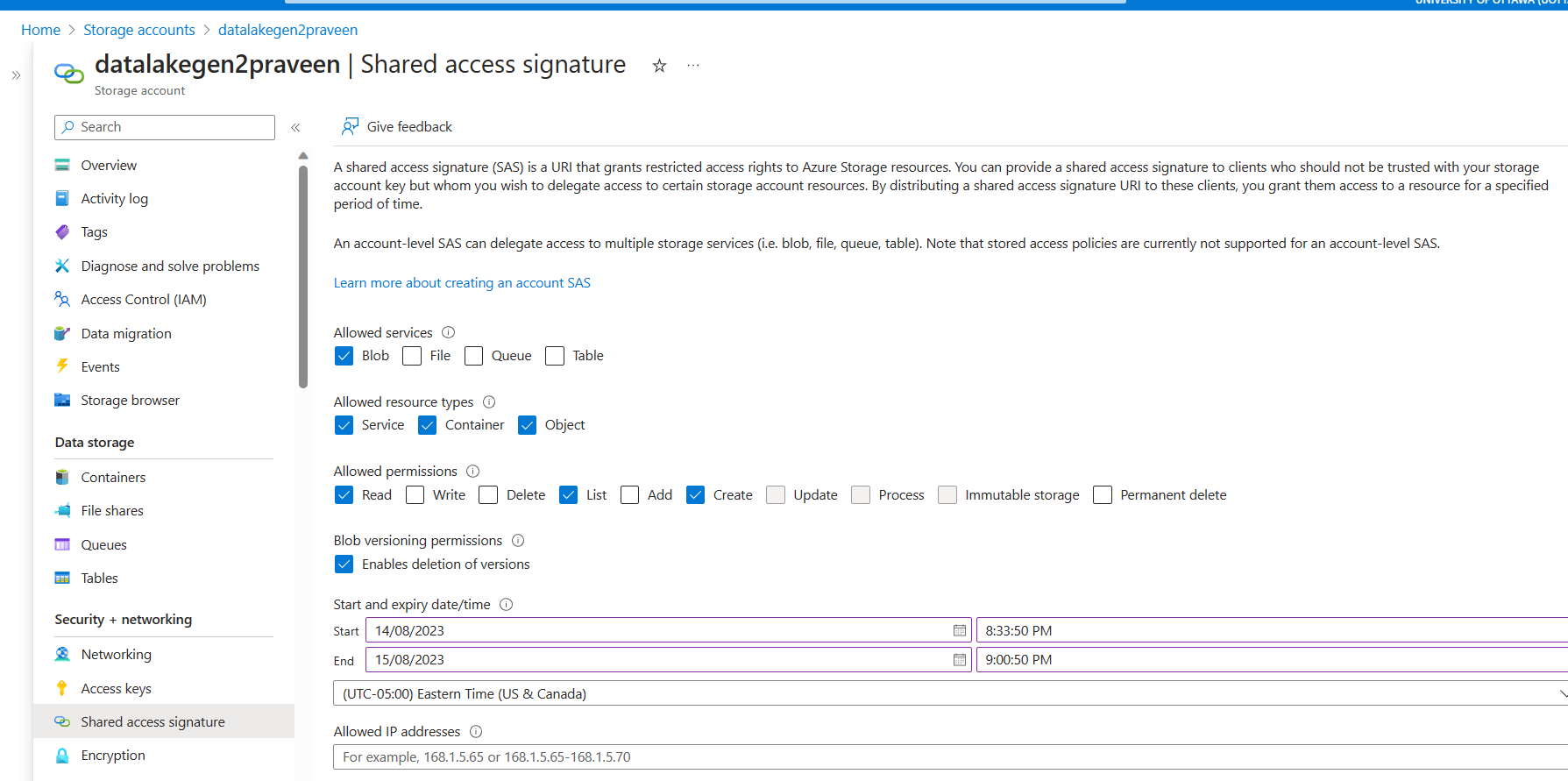


Adding an access role to the account to access datalake storage





Get Sas token credentials from DataLake container



Creating a SQL script creating database scoped credential telling the SAS token to use, and create a new fileformat telling it’s made of delimited text and comma separator.

Mentioning an external data source from where data is fetched, a file format on how to read data from that source. Then creating a table in db, reading the data from the external source, in the specified format and loading it into the external table.

create master key ENCRYPTION by PASSWORD = 'Welcome123'

CREATE DATABASE SCOPED CREDENTIAL SasToken

WITH IDENTITY = 'SHARED ACCESS SIGNATURE',

SECRET = SAS Token;

CREATE EXTERNAL FILE FORMAT TextFileFormat WITH (

    FORMAT\_TYPE = DELIMITEDTEXT,

    FORMAT\_OPTIONS(

        FIELD\_TERMINATOR = ',',

        FIRST\_ROW = 2

    )

)

CREATE EXTERNAL DATA SOURCE population\_data\_source

WITH (

    LOCATION = 'https://datalakegen2praveen.blob.core.windows.net',

    CREDENTIAL = SasToken

)

CREATE EXTERNAL TABLE [populationdatatable1] (

    [Country] [varchar](1000),

    [Population] [varchar](1000),

    [YearlyChange] [varchar](1000),

    [NetChange] [varchar](1000),

    [Density] [varchar](1000),

    [LandArea] [varchar](1000),

    [Migrants] [varchar](1000),

    [Fert.Rate] [varchar](1000),

    [Med.Age] [varchar](1000),

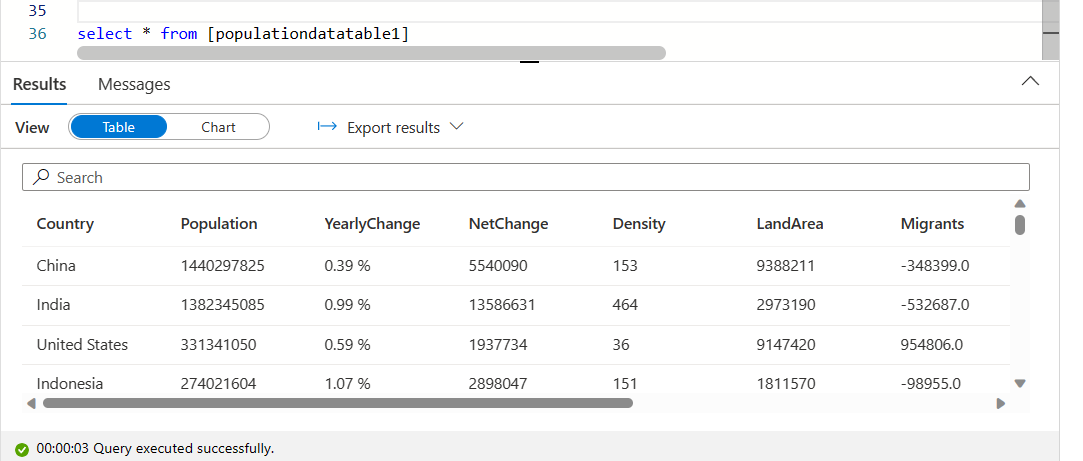
    [UrbanPop%] [varchar](1000),

    [WorldShare] [varchar](1000))

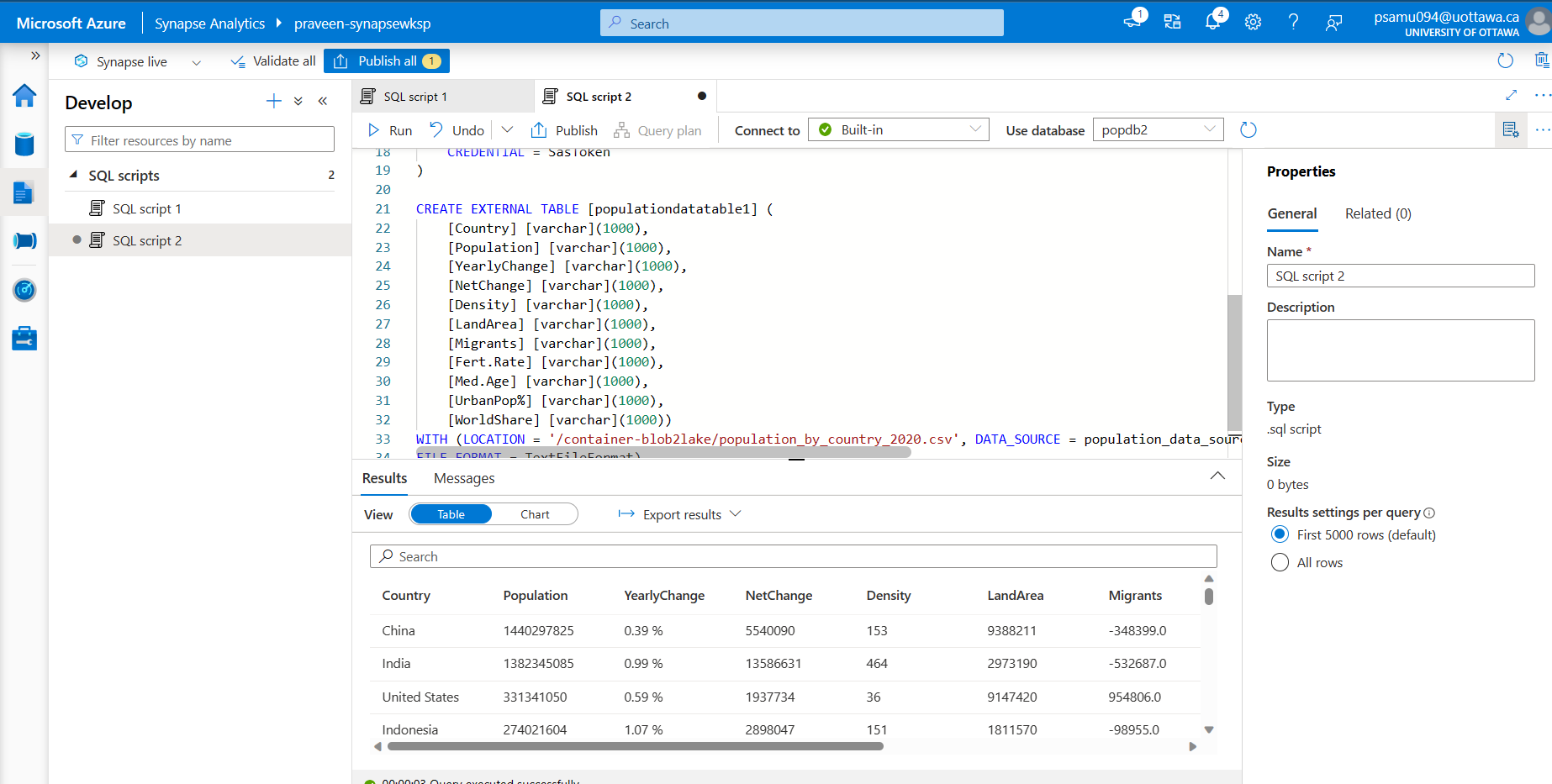
WITH (LOCATION = '/container-blob2lake/population\_by\_country\_2020.csv', DATA\_SOURCE = population\_data\_source,

FILE\_FORMAT = TextFileFormat)

select \* from [populationdatatable1]



Task 3 – Working on SQL pool external tables - csv



Task 4 – Working with parquet files

--Parquet File

create external FILE FORMAT parquetfile with (

    FORMAT\_TYPE = PARQUET,

    DATA\_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'

);

create EXTERNAL table [population\_parquet] (

    [Id] [int],

    [Correlationid] [varchar](200),

    [Operationname] [varchar](200),

    [Status] [varchar](100),

    [Eventcategory] [varchar](100),

    [Level] [varchar](100),

    [Time] [datetime],

    [Subscription] [varchar](200),

    [Eventinitiatedby] [varchar](1000),

    [Resourcetype] [varchar](1000),

    [Resourcegroup] [varchar](1000))

WITH (

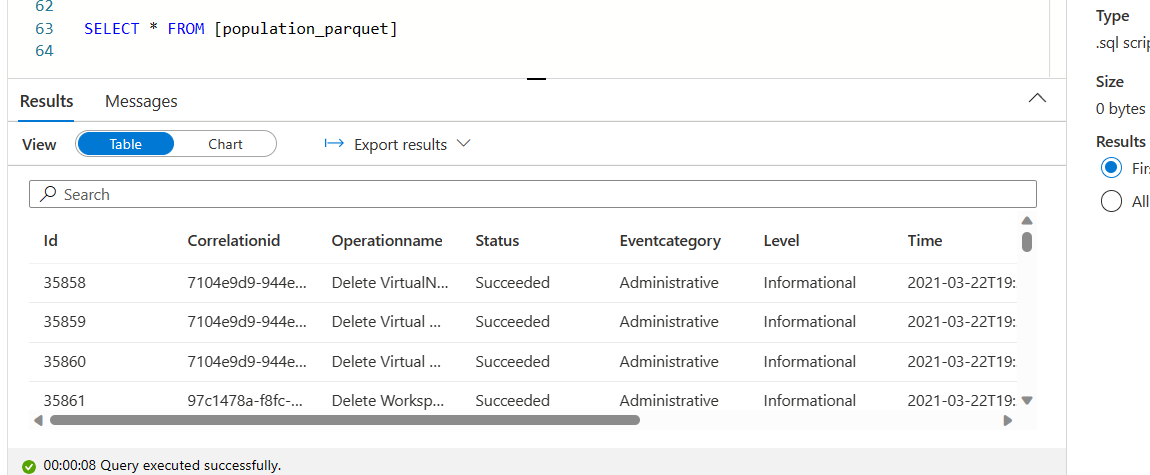
    LOCATION = '/container-blob2lake/parquet2.parquet',

    DATA\_SOURCE = population\_data\_source,

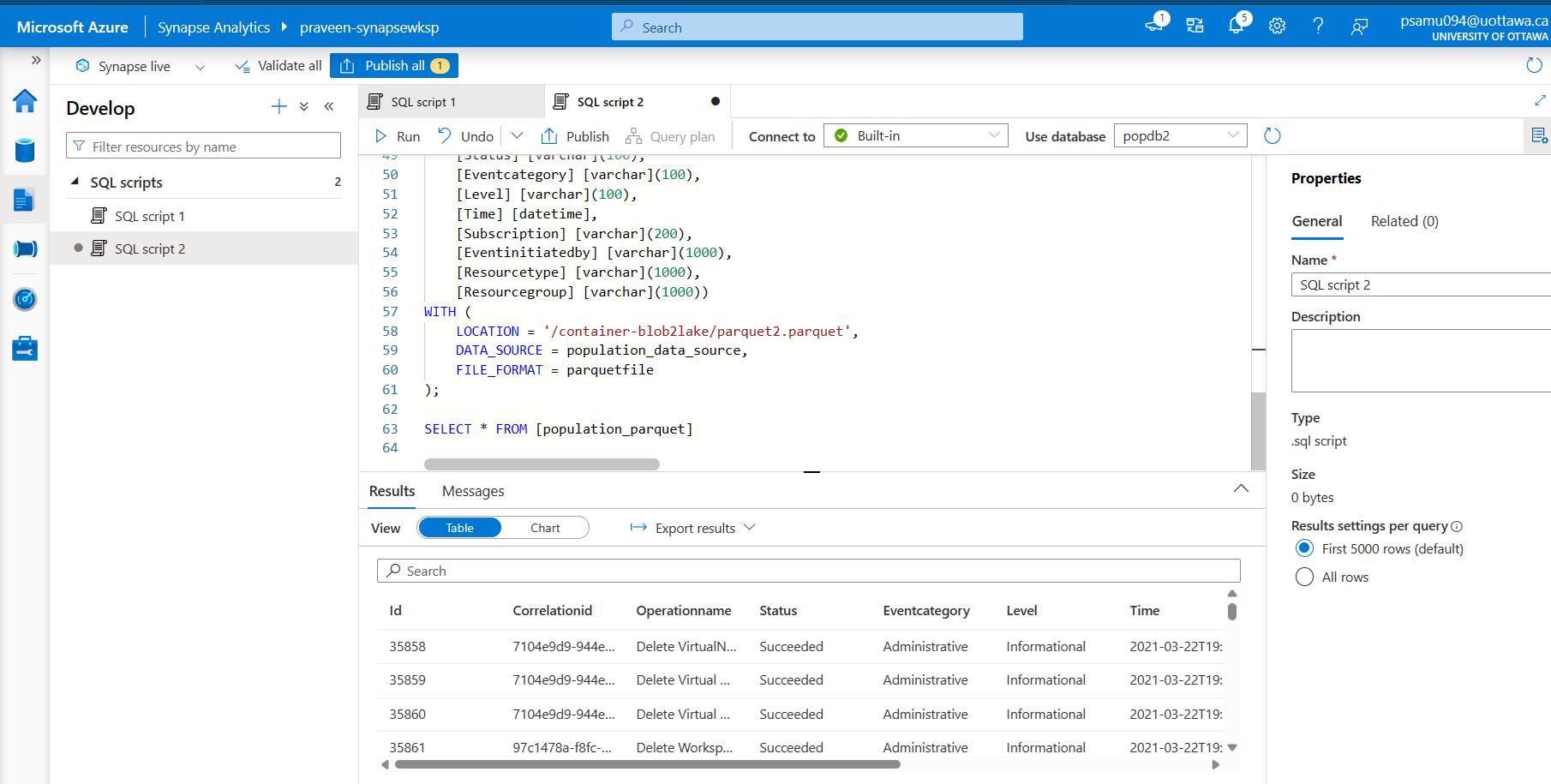
    FILE\_FORMAT = parquetfile

);

SELECT \* FROM [population\_parquet]

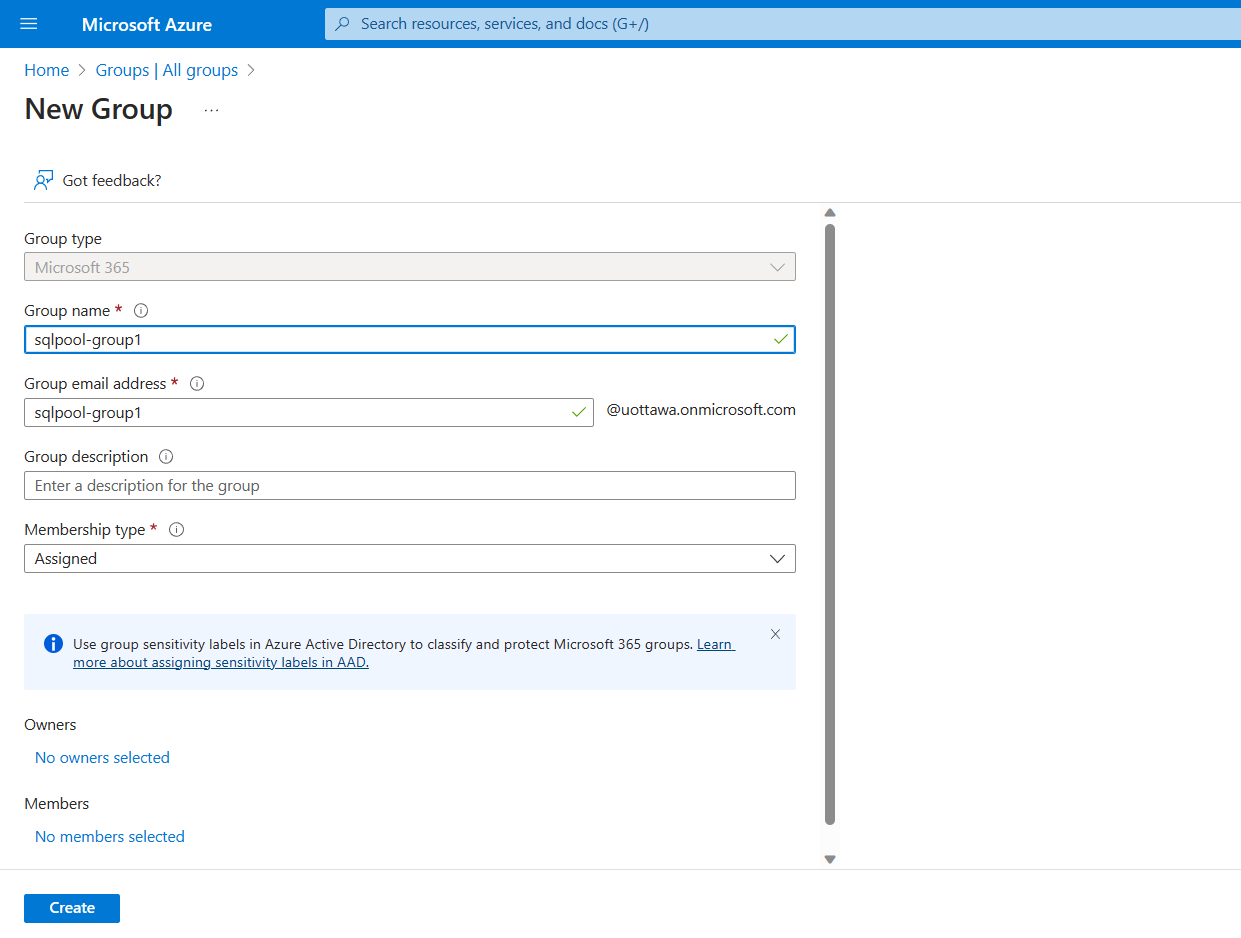


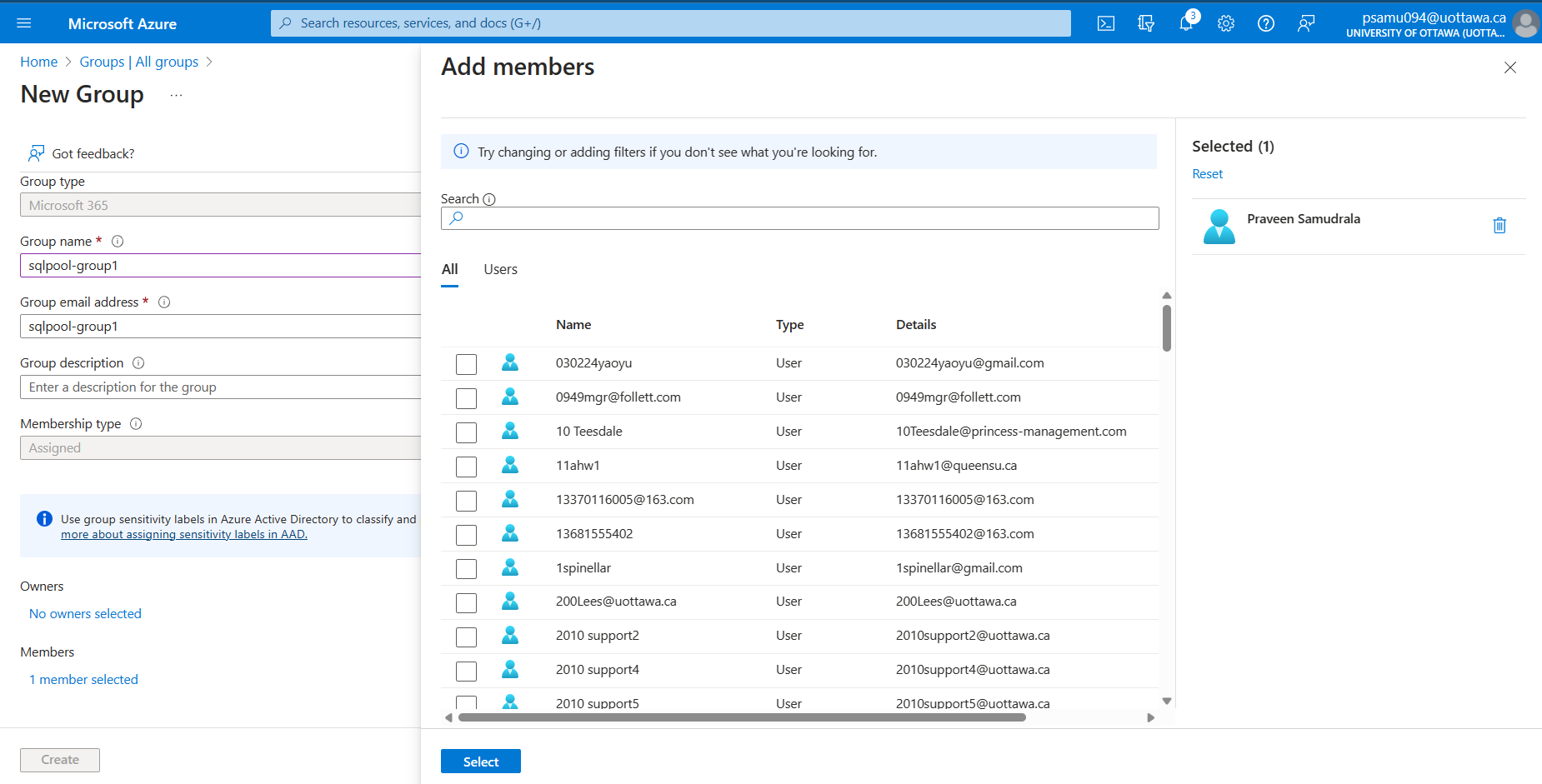
Task 5 – Loading Parquet file onto External table in serverless SQL pool



Task 6 – Securing Secure access to data in Data Lake from Serverless SQL pool

Step 1 – Create Azure active directory security group – Groups -> Create New Group

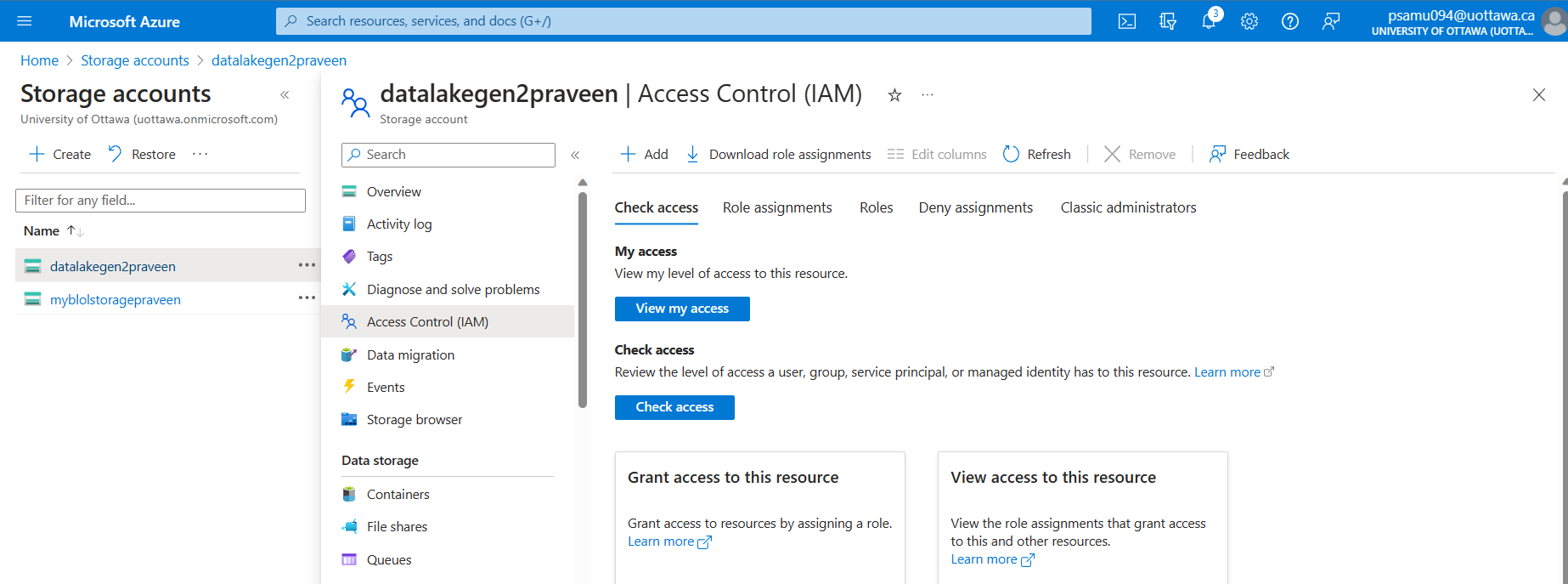




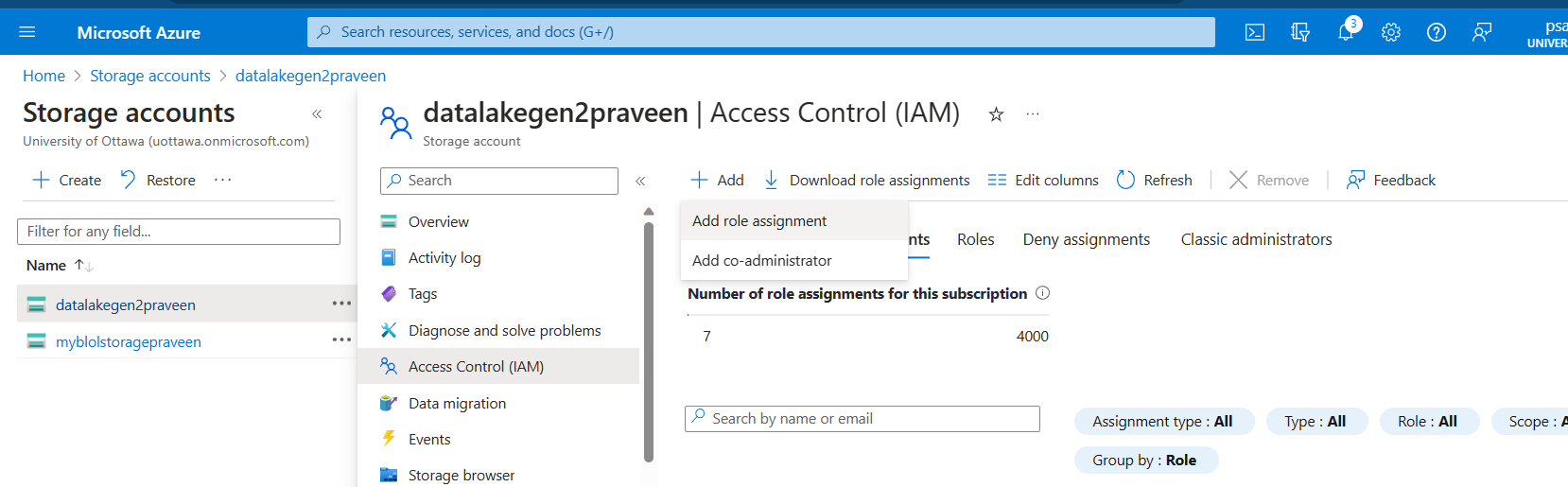
My student account is not permitting to create a resource group.

Task 8 – Configure Data Lake security using Role-based access control list (RBACL) and ACL.

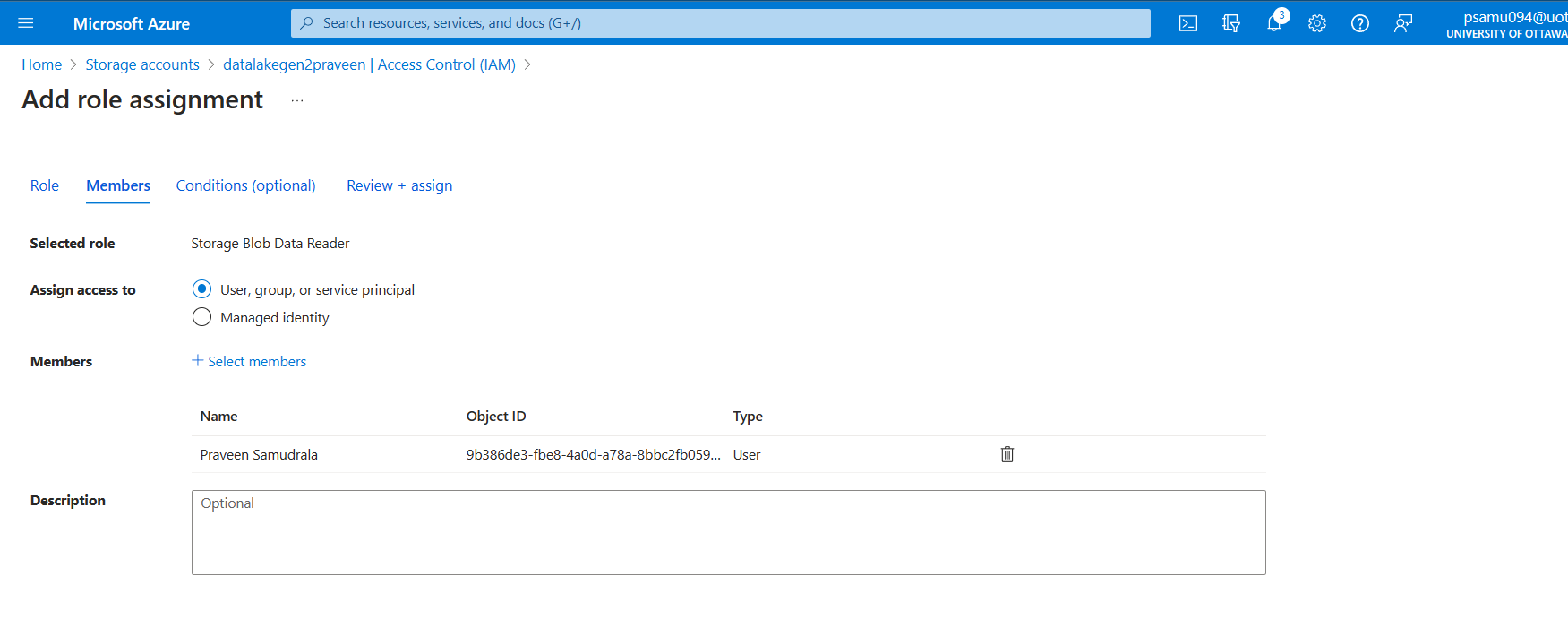
Step 1 – Storage Account -> Choose the Storage -> Access Control IAM



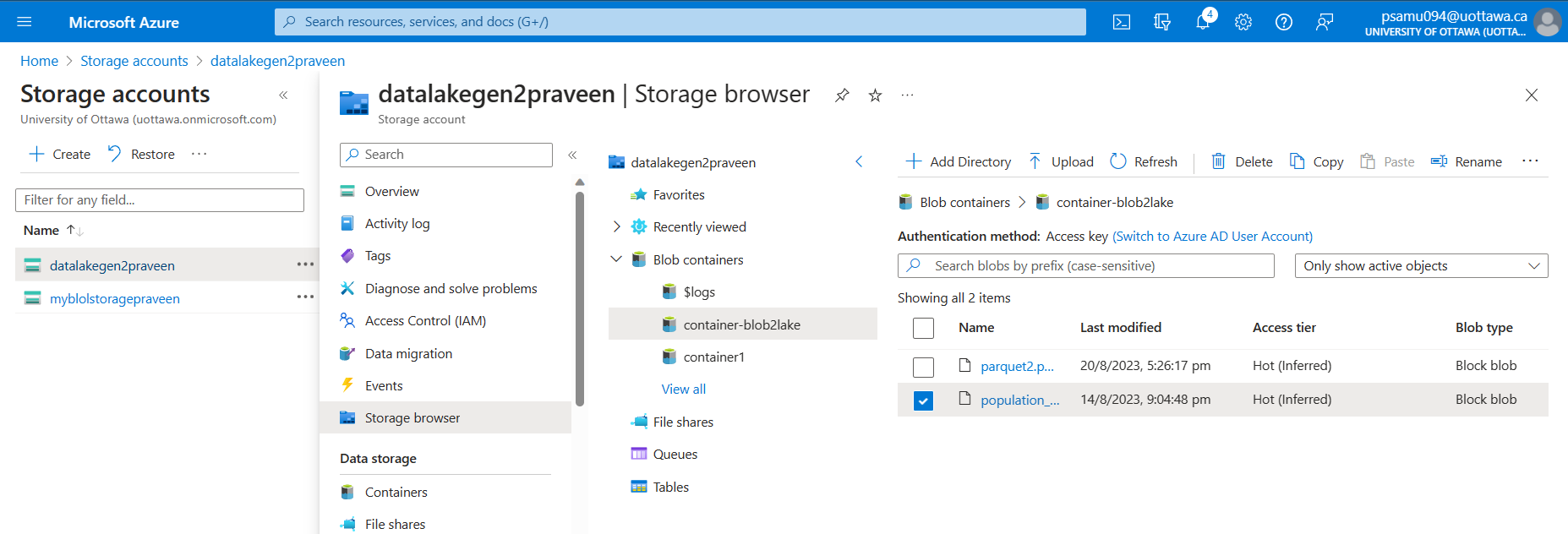
Step 2 – Add -> Add role assignment



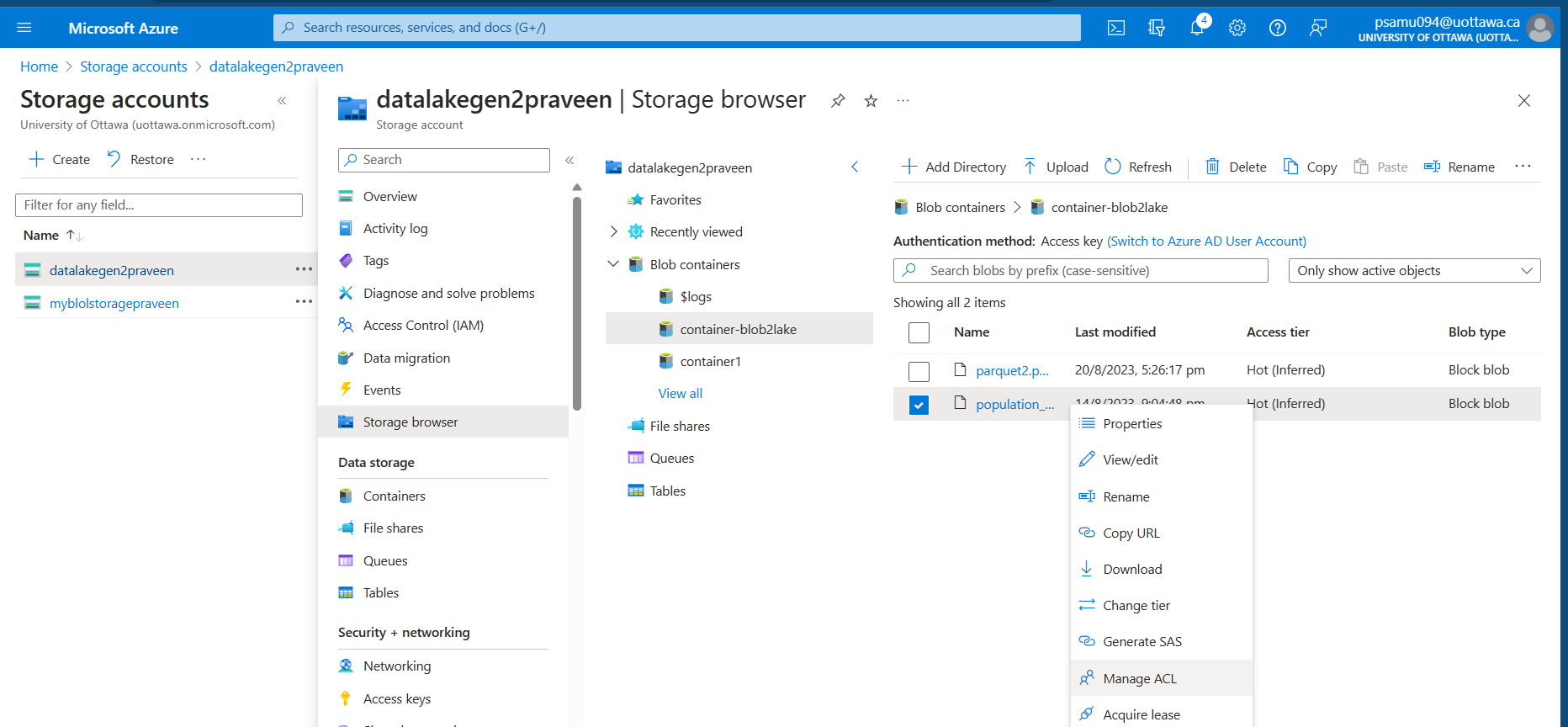
Step 3 – Choose the role and add a member



Step 4 – Configure Access Control Lists – Storage Account -> Storage Browser -> Choose the file in the container to which ACL are to be managed.



Step 5 – Right click on the file -> Manage ACL



Step 6 – Add Principal -> Choose the group created earlier and update the access by clicking on the check boxes

