**********25/09/2023**********

Agenda:

- Azure
- Cloud concepts
- Types
- Lass, paas, Saas
- Diff between on premises & cloud based'
- Job responsibities as azure data engineer
- Types of data

Azure data engineer also help ensure that data pipeline and data stores are high performing, efficient, organized, and reliable, given a set of business.

Server

If we are going to use cloud, In cloud no need to maintain any server, with internet we can access locally.

No need to take care of insfrastructure, cloud is cheaper.

To access azure login: https://portal.azure.com/#home

GitHub - deeksharm/DP203

https://github.com/deeksharm/DP203

DP203/Azure_Sep25.xlsx at main · deeksharm/DP203 · GitHub

Big data, hadoop: Big Data Tutorial – Learn Big Data from Scratch - DataFlair (data-flair.training)

AZURE DOCUMENTATION: Azure documentation | Microsoft Learn

PAYMENT MODEL

1.Pay-as-you-go
2.1 year reservation -->
up to 30% discount
3. 3 years reservation -->
up to 50-70%

Learning link:

Data redundancy - Azure Storage | Microsoft Learn

Introduction to Azure Storage - Cloud storage on Azure | Microsoft Learn

Azure cross-region replication | Microsoft Learn

Dynamic data masking - SQL Server | Microsoft Learn

What is a Public Cloud - Definition | Microsoft Azure

Storage account overview - Azure Storage | Microsoft Learn

Azure Storage Explorer – cloud storage management | Microsoft Azure

Adding Users to Your SQL Azure Database | Azure Blog | Microsoft Azure

Azure portal: Dynamic data masking - Azure SQL Database | Microsoft Learn

Dynamic data masking - Azure SQL Database | Microsoft Learn

Pricing Pricing Calculator | Microsoft Azure

Cosmos db:

Databases, containers, and items - Azure Cosmos DB | Microsoft Learn

Azure resource manager:

Create template - Visual Studio Code - Azure Resource Manager | Microsoft Learn

Hadoop Ecosystem and Their Components - A Complete Tutorial - DataFlair (data-flair.training)

Refer:

Python, Big data → data flair.com(Free Online Certification Courses – Learn Today. Lead Tomorrow. - DataFlair (data-flair.training))

The Azure Storage platform includes the following data services:

- Azure Blobs: A massively scalable object store for text and binary data. Also includes support for big data analytics through Data Lake Storage Gen2.
- Azure Files: Managed file shares for cloud or on-premises deployments.
  Existing configuration file we can use no need to do congifiguartion for all 50 virtual machines . in files we can store configuration files.
  SMB protocol, NFS.
- Azure Elastic SAN (preview): A fully integrated solution that simplifies deploying, scaling, managing, and configuring a SAN in Azure.
- Azure Queues: data stored in message format.
- Azure Tables: A NoSQL store for schemaless storage of structured data.
- Azure managed Disks: Block-level storage volumes for Azure V

To check pricing of service we are going to create:
Pricing Calculator | Microsoft Azure

STORAGE ACCOUNT:

CREATE RESOURCE GROUP:

1) Home ➔ Subscriptions ➔ Pay-as-you go ➔ Resource Group ➔ + create ➔ Resource group: RG_UST_Training ➔ Region : EAST(US)
2) Tags ➔ Project : UST_Training, Name: Monisha (it can be any name)
3) Click Review : create
4) Create

DATA LAKE(HIERARCHICAL NAMESPACE):

1) Go inside the resource group created(RG_UST_Training) ➔ +create ➔ search: storage account ➔ expand create ➔ click storage account ➔ Account Name(should be in small letters, globally unique, alpha numeric value because the url will use this name so it should be unique) : adlsusttraining(azure datalake storage)
2) 
3) Above tick the check box for Standard(if you premium expensive you can choose that also)
4) Redundancy : Choose LRS(Locally Redundant Storage)➔ less cost
5) Click Next: Advanced
6) Go with default option: Tick 1st : https secure and 3rd : authentication method key access
7) Click Next
8) Tick Enable Storage
9) Hierarchical Namespace Tick it (DATA LAKE)
10) Access Tier by default Hot go with that option
11) Click Next: Networking
12) Networking Connectivity: Tick 1st: Enable public access
13) Network Routing : Tick Microsoft Network Routing
14) Click Next

15) Tick Enable soft delete or blobs
16) Encryption go with default
17) Tags:

Project: UST_Training
Owner : Monisha

18)Review and create 19)Click Create

CREATE VIRTUAL MACHINE:((WINDOWS))

1) Go to the Resource Group(RG_UST_Training) ➔ +Create ➔ search : Virtual Machine ➔Click create ➔ click virtual machine ➔ VM name: vmmoni, Region : EAST US
2) Availablity Option : No infrastructure redundancy required (As of now because we r not creating for production)
3) Security type : Trusted launch virtual machine
4) Image ➔ configuration file Windows server 2022 datacenter: Azure Edition - * 64 Gen 2
5) Spot instance : B1'S
6) Administrator Account :
   USER NAME : MonishaB
   PASSWORD : ********
7) Public Inbound ports: Allow selected ports
8) Select inbound ports: RDP: 3389
9) Os disk type : Premium SSD
10) Key ➔ platform managed key
11) Default
12) Next
13) Tags

Project: UST_Training
Owner : Monisha

14) Review and create
15) Click Create

16) How to connect vm(windows):

17) Start
18) Connect
19) Click select
20) Download RDP file ➔ click open file
21) Connect
22) Enter username and password
23) trust

CREATE VIRTUAL MACHINE:((LINUX))

1) Go to the Resource Group(RG_UST_Training) ➔ +Create ➔ search : Virtual Machine ➔Click create ➔ click virtual machine ➔ VM name: ustvm, Region : South central

2) Availablity Option : No infrastructure redundancy required (As of now because we r not creating for production)

3) Security type : Trusted launch virtual machine
Image ➔ configuration file UBUNTU server 22.04 LTS

4) Spot instance : B1'S to increase use DS1_V2

5) SSH:
USER NAME : MonishaB
Key value : monisha(anyname we can give)

6) Public Inbound ports: Allow selected ports

7) Select inbound ports: SSH:22

8) Os disk type : Premium SSD

9) Key ➔ platform managed key

10) Default

11) Next

12) Tags
Project: UST_Training
Owner : Monisha

13) Review and create

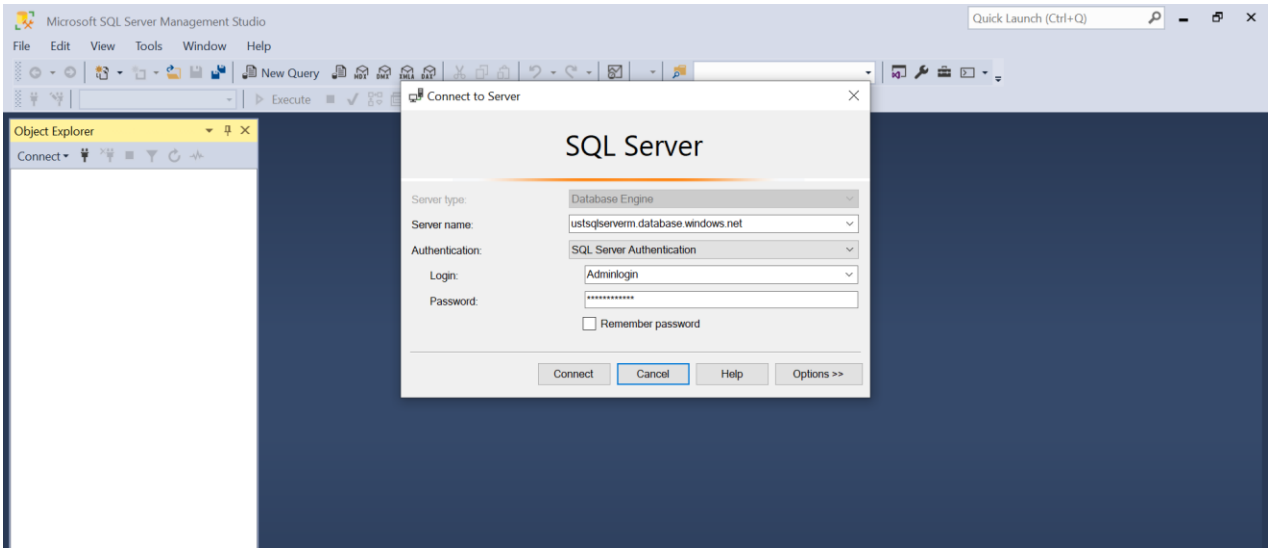
14) Click Create

15) Download pem file store in a local

How to connect :LINUX

1)Start
2)connect
3)Search puttygen:
4)Load ➔ choose .pem file ➔ open ➔ save private key
5)Search putty :
6)Ssh ➔ auth ➔ credential ➔ browse ➔ ppk file ➔ open
7)Go to session ➔ port number 22 ➔ IPAddress: username@ipaddress
8)Open
9)Connect once
10Type hostname and check hostname

11)Stop(After using linux vm stop it)



AZURE SQL:

TYPES:

1) SQL with vm ➔ (SQL Server + vm) IAAS
2) Managed Instance ➔ (Fully managed, SQL Server) PAAS
3) SQL Database ➔ (Complete SQL DB) PAAS
    1)Singleton DB
    2)Elastic Pool

LAB AZURE SQL:

1) Search Azure SQL
2) Click create ➔ AZURE SQL
3) Choose SQL Databases ➔ single database ➔ create
4) Choose Resource group ➔ RG_UST_Training
5) Database name : USTSqlDB
6) Server : Create New : ➔ Server Name (lowercase should be unique): ustsqlserverm
7) Location : (US) East US
8) Authentication method : 1)SQL Authentication
    2) Azure active directory auth ➔ Microsoft entra authentication
    ➔ use both sql & Microsoft entra auth
9) Choose sql authentication Server Admin Name : Adminlogin , Password : Monisha@ishu
10) Ok
11) Want to use sql Elastic Pool : No
12) Work Environment : Tick Development,
    Production.
13) Compute + Storage : Service & compute tier : choose basic (DTU'S 5 DATA SIZE 2GB)➔ Apply
    Standard
    Premium
    VCore

14) Backup Storage Redundancy : Tick LRS,
GRS,
ZRS.
15) Connectivity Method : Tick Public Endpoint ➔ allow azure service : yes , hybrid benefit: yes
No access,
Private Endpoint
16) Connection Policy: Tick Default (uses ReDirect Policy)
Proxy
Redirect
17) Microsoft Defender for SQL: Enable defender ➔ Tick Not now,
Start Free Trial
18) Next
19) Additional settings: Data source ➔ Use Existing Data ➔ Choose Sample ➔ ok
20) Tags
Project : ust_training
Owner : monisha
21) Review + create
22) Create

AZURE SQL QUERING:

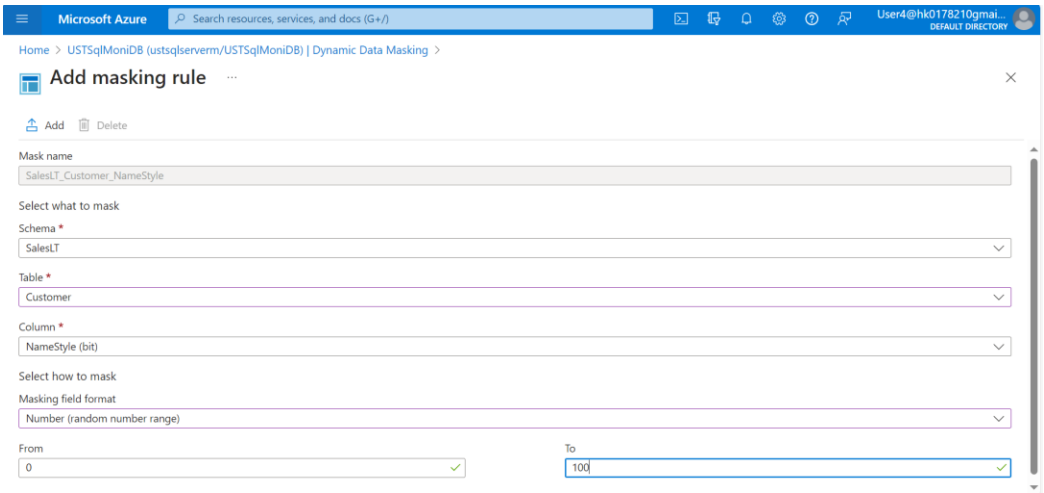
1) Overview ➔ setserverfirewall ➔ +create ➔ 0.0.0.0 254.254.254.254 ➔ save
2) Query editor ➔ username : , password : ➔ ok
3) +new query➔ Select * from [SalesLT].[Address]
4) Run

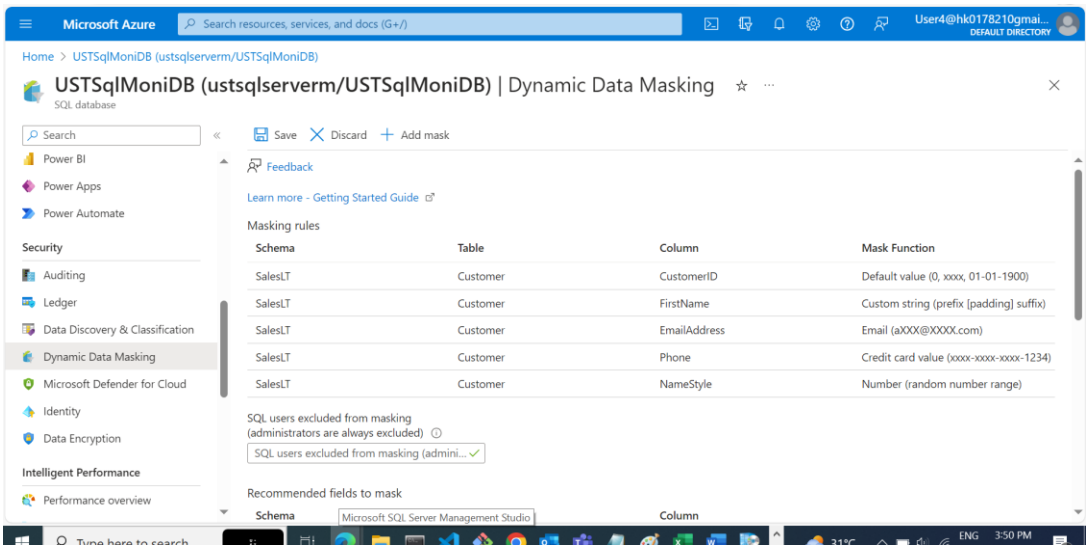
HOW TO CONNECT AZURE SQL TO SSMS :

1) In azure sql ➔ overview ➔ copy server name
2) Open SSMS Software ➔ + connect ➔ Database Engine ➔ server : paste server name ➔ SQL Server Authentication ➔ server_name: password: from azure portal adminlogin and password
3) Connect

DIFFERENT MASKING RULE:

HOW TO ADD MASKING RULE:





**********28/9/2023**********

USE CASE 1:

ADF

BLOB =====➔ DATALAKE

E T load

CREATE DATA FACTORY:

1) Go to resource group ➔ + create ➔ search : data factory ➔ create ➔ click data factory
2) Name(unique) : Adfmonisha
3) As of now Repository type :Tick Git later. Tick (Repository type : tick github or azure
4) Connect via : Tick public endpoint,
   Private endpoint
5) Tags :
   Project ➔ ust_training
   Owner ➔ Monisha
6) Review + Create
7) Create

We need blob storage and datalake service to copy data via data factory: blob to data lake

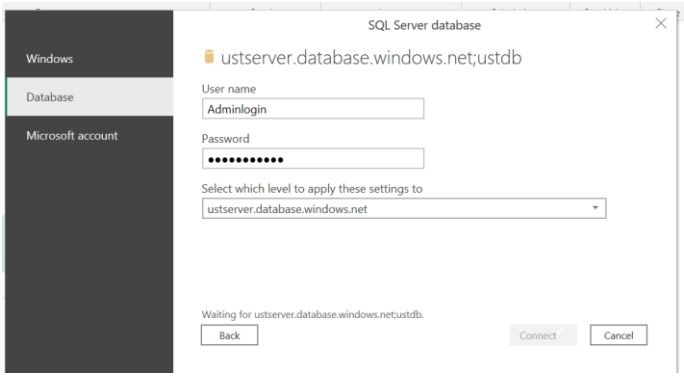
1) Click container in blob ➔ +container ➔Create container named raw in blob storage. Name:raw➔ create
2) Click container in datalake ➔ +container ➔Create container named refined in default storage.
3) Name:refined➔ create
4) In blob storage ➔ click upload ➔ browse for files ➔ choose person.csv file ➔ open➔upload
5) Go to data factory(adfmonisha) ➔ Overview ➔ click launch studio

6) Click Manage ➔ Linked service ➔ + New ➔ search:blob ➔ choose Azure Blob Storage ➔ continue
7) Name (LS is best practice): LS_AzureBlobStorage
8) **Connect via integration runtime :** Default (AutoResolveIntegrationRuntime)
9) **Authentication type:** Account Type ➔Connection String ➔
10) **Tick** From Azure subscription((((Enter manually to connect with other account))))
11) Azure Subscription : Choose pay as you go
12) Storage account name : Blob Storage name(absmonishaust)
13) Test connection:To linked Service
14) Click create
15) Click Manage ➔ Linked service ➔ + New ➔ search:data ➔ choose Azure Data Lake Storage Gen 2➔ continue
16) Name (LS is best practice): LS_AzureDataLakeStorage
17) **Connect via integration runtime :** Default (AutoResolveIntegrationRuntime)
18) **Authentication type:** Account Type ➔Connection String ➔
19) **Tick** From Azure subscription((((Enter manually to connect with other account))))
20) Azure Subscription : Choose pay as you go
21) Storage account name : Data Lake Storage name(adlsmonishaust)
22) Test connection:To linked Service
23) Click create
24) Create DATASETS ➔ Author ➔ dataset ➔ … ➔ new dataset➔ search:blob ➔ choose Azure Blob Storage ➔ continue
25) Select format : choose csv ➔ continue
26) Name(DS is a best practice) : DS_Source_raw
27) LinkedService: LS_AzureBlobStorage
28) Browse ➔ select raw ➔ person.csv ➔ ok ➔ ok
29) … ➔ new dataset➔ search:data ➔ Azure Data Lake Storage Gen 2➔ continue
30) Choose json ➔ Name : DS_Target_refined ➔Linkedservice: LS_AzureDataLakeStorage
31) Browse ➔ select refined ➔ ok ➔ ok
32) Pipelines ➔ … ➔Name: PL_Raw_to_Refined ➔Activities➔ search:copy(Becoz here we are going to copy person.csv from blob to person.json to data lake)➔ drag and drop copy data ➔ give Name : Raw_to_refined ➔
33) Source ➔ Source datset : DS_Source_raw
34) Sink ➔ Sink dataset : DS_Target_refined
35) Click Validate → it will validate and resolve if any error is there means
36) Click debug ➔ in debug we can do validation without saving (publish) files once debug is succeeded
37) Click publish all ➔ then click Add trigger ➔ trigger now ➔ to trigger pipeline
38) Add Trigger ➔ new/edit ➔ schedule , tumbling window, storage events, custom events

HOW TO CONNECT POWERBI WITH AZURE SQL DATABASE:

1) Go to PowerBI → Get Data → more → Azure → choose Azure SQL database → connect

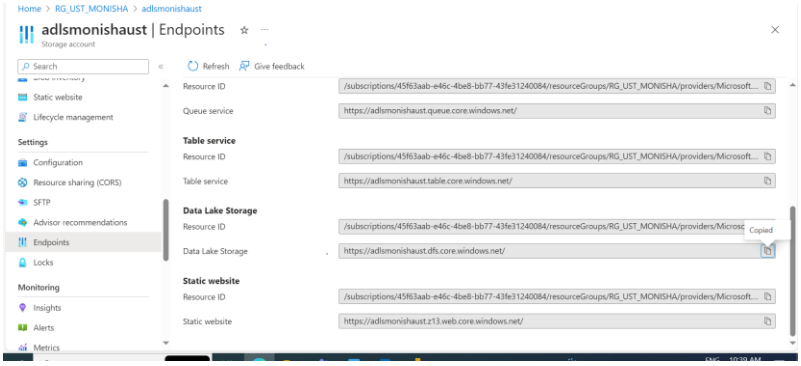
2) In portal➔ go to azure sql (resource_group: RG_UST_Training, ustdb:azure sql)➔ overview ➔ copy server name and paste in power bi server ➔ database name also copy from portal and paste in powerBi ➔ ok
3) Choose database ➔ username: Adminlogin, password: Welcome@123 ➔ connect



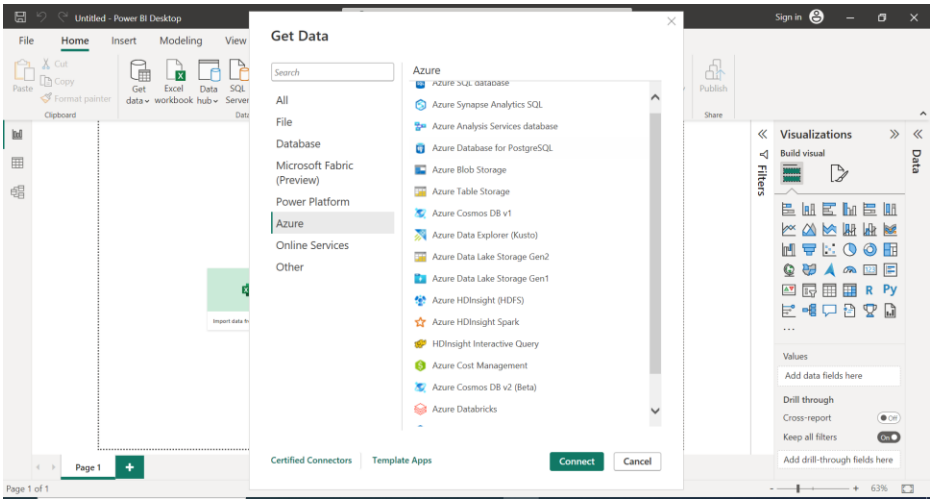
4) Connect mobile data(UST_SECURE is not

HOW TO CONNECT DATALAKE WITH POWERBI:

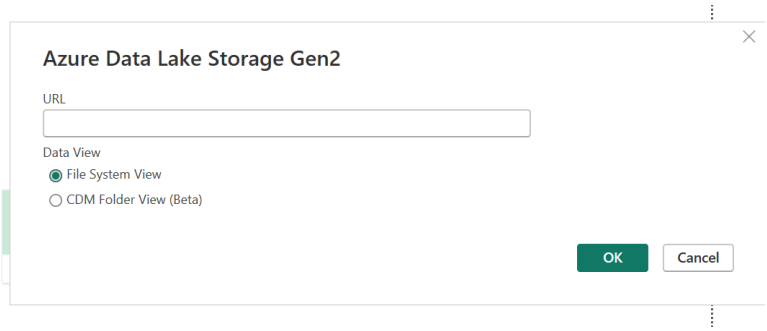
1) Go to data lake storage ➔ endpoint ➔ copy datalake storage account 2nd url ➔ ok



2) In Power Bi → Get Data ➔ Get More ➔ Azure ➔ Azure Data Lake Storage Gen 2 ➔ connect

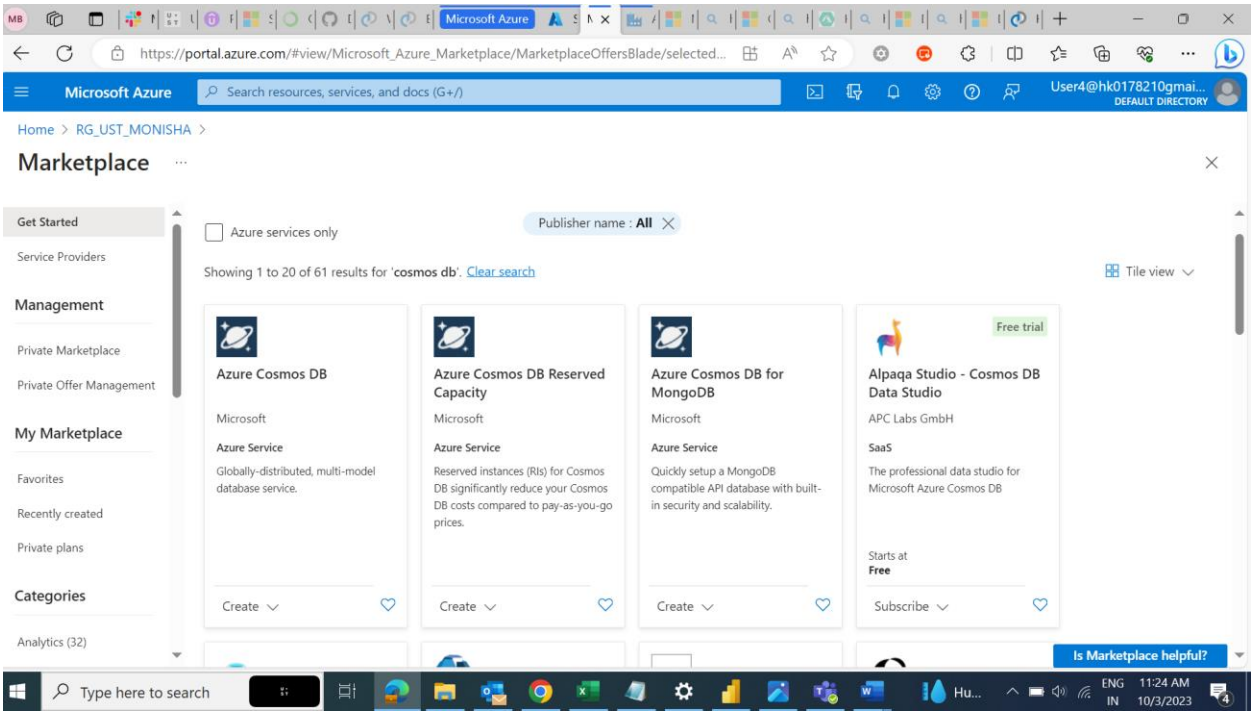


3) Paste the datalake url ➔ click ok



4) Authentication : Organisational Account,
Account key,
SAS ➔ Access choose ➔ generate SAS ➔ Token
5) Choose Account key ➔ copy and paste account key ➔ connect
6) Combine or load or transform data

AZURE COSMOS DB:



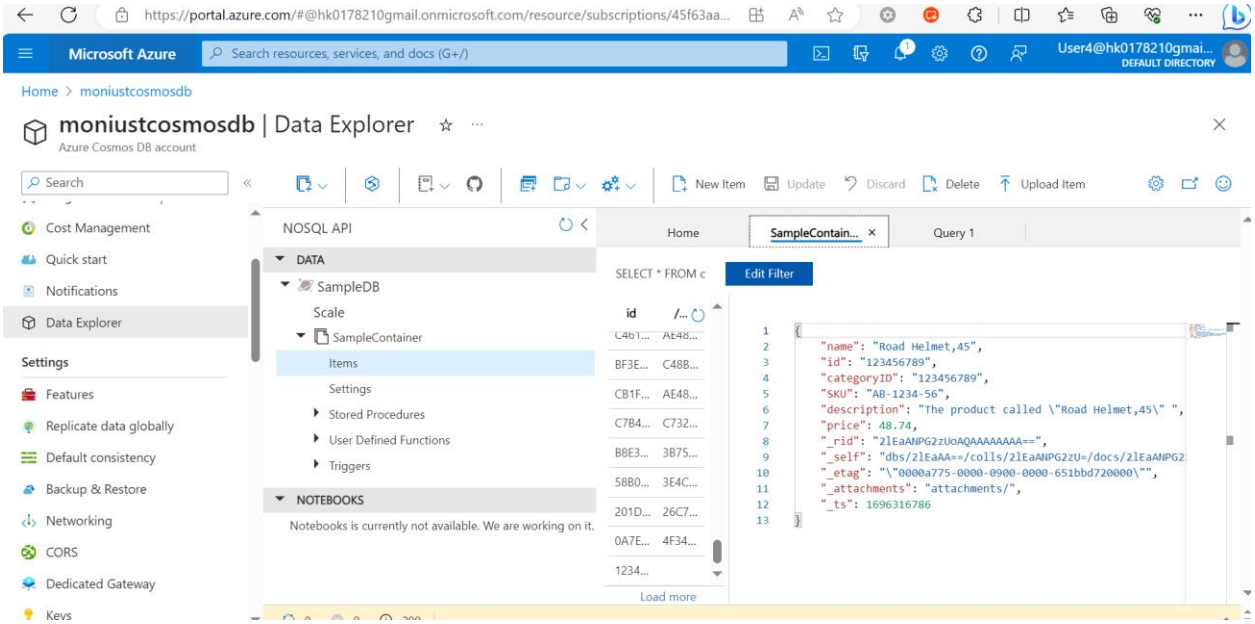
- Go to resource group ➔ search azure cosmos db ➔ create ➔ azure cosmos db
- Azure cosmos db for nosql ➔create
- Account Name : moniustcosmosdb

- Location: Canada East(Account level region)
- Capacity mode : tick provisioned throughput,
  Serverless
  Limit total account tck it
- Next
- GRS Disable
- Next
- Connectivity network : tick all network(default)
- next
- Backup policy: periodic(default)
- Next'
- Data encryption: tick service managed key
- Tags
- Review +create
- Create

How to create container and item, database:

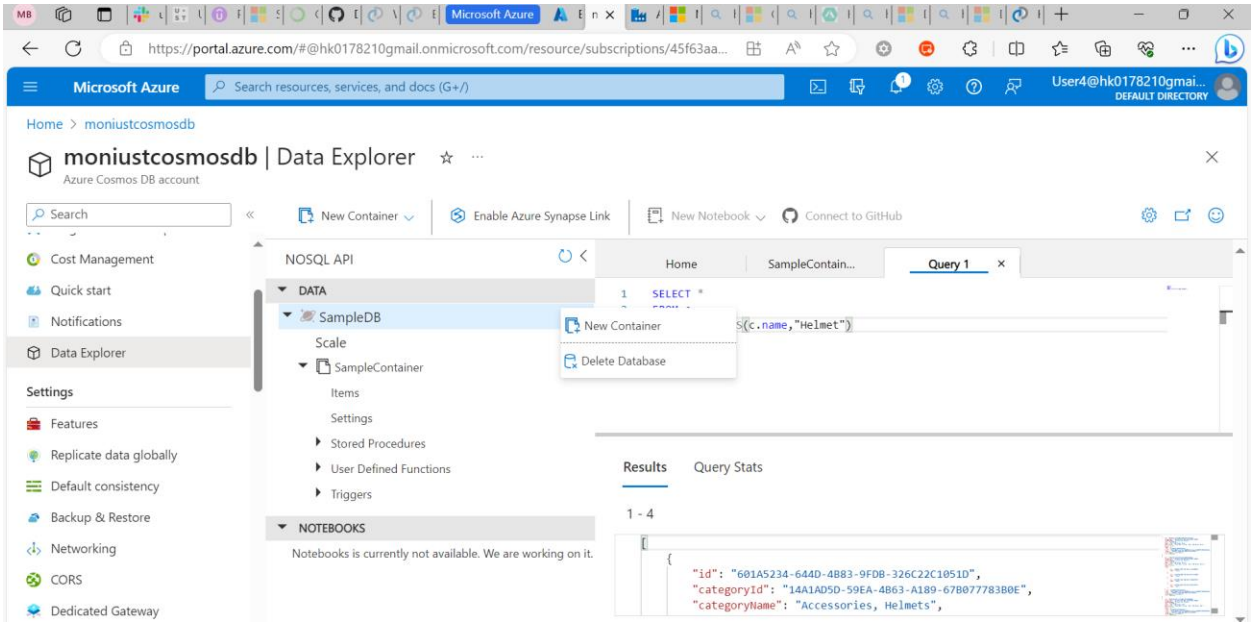
- Click data explorer ➔ Launch quick start ➔ create new ➔ SampleDB (Default)
- Tick share throughput across container
- Database throughput: tick Autoscale ➔ 1000
- Container id : SampleContainer
- Partition key: /categoryid
- Click ok
- Click items ➔ new item ➔ paste item json from cosmosdnjsonformat ➔ new sql query:

```
SELECT *
FROM c
WHERE CONTAINS(c.name,"Helmet")
```

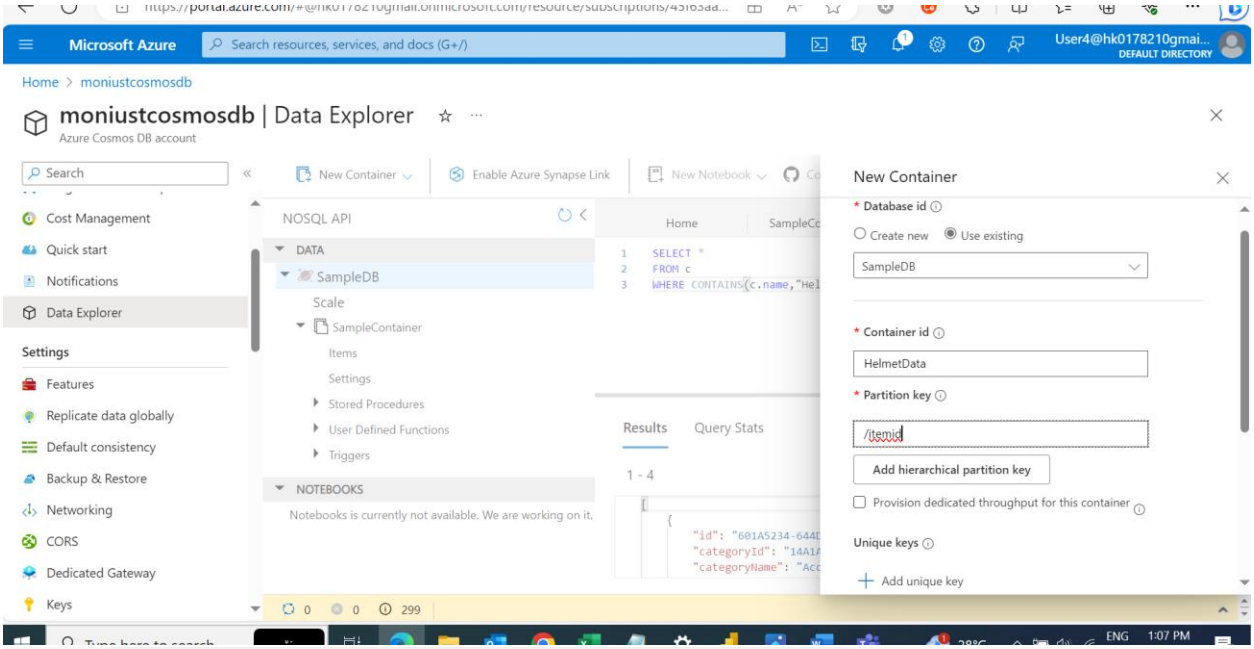


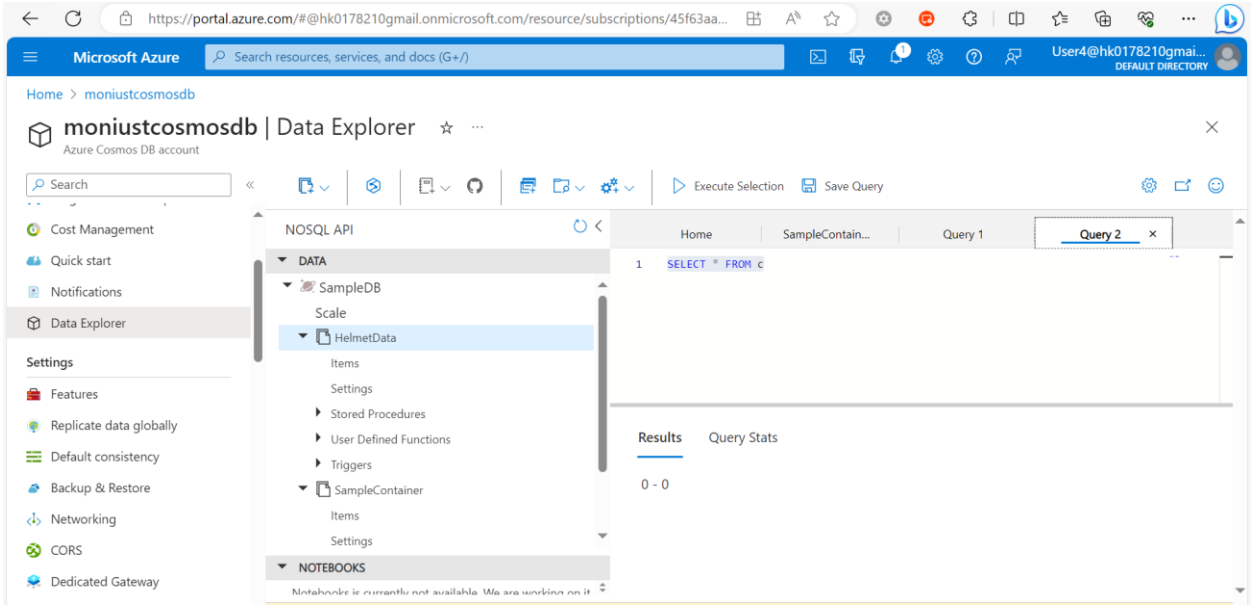
MOVING DATA FROM ONE CONTAINER TO ANOTHER CONTAINER USING DATA FACTORY:

- … ➔ New container ➔

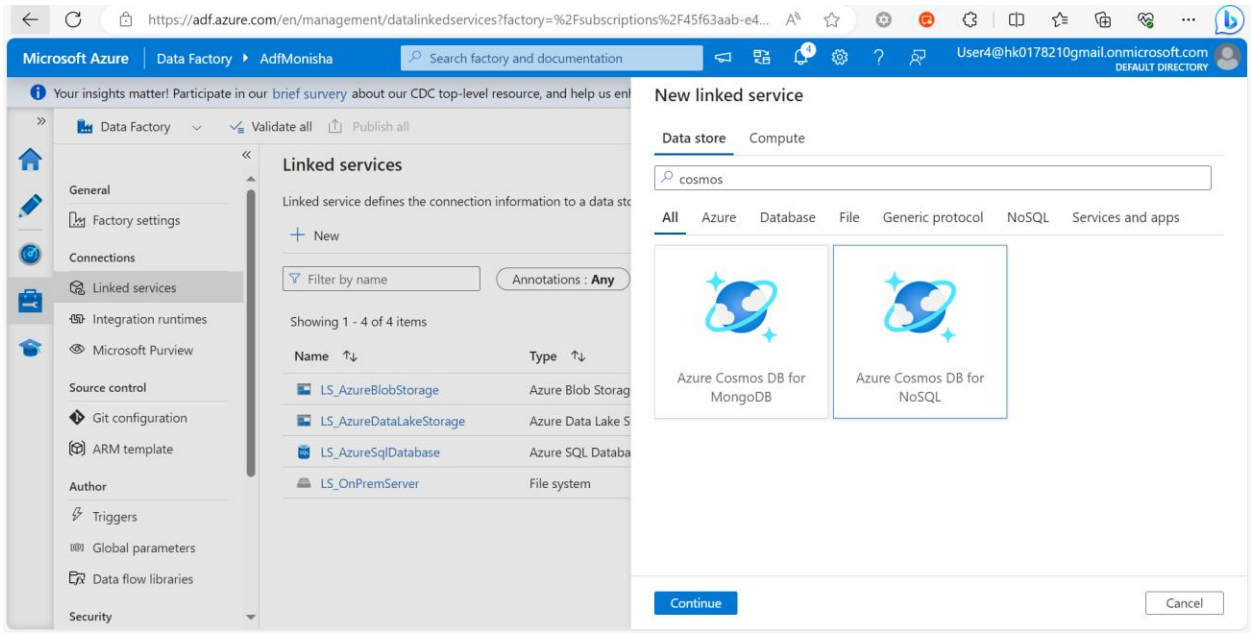


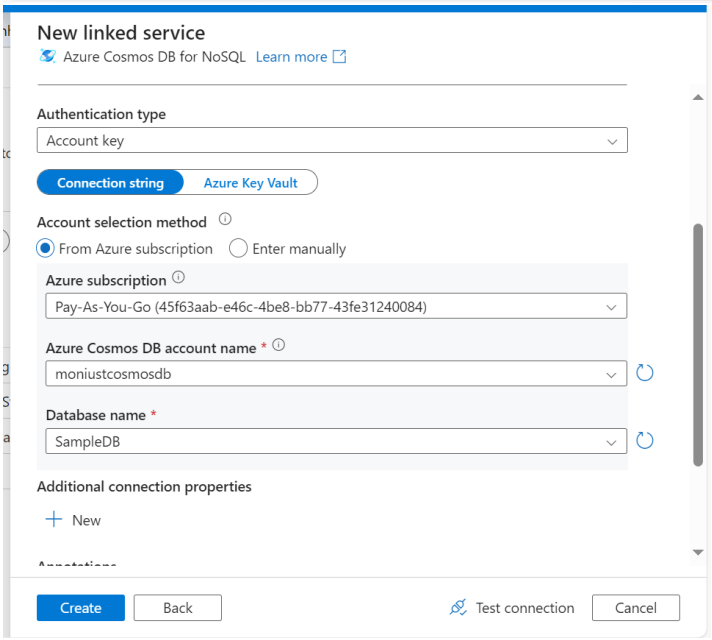
- Container id : HelmetData, partition key: /itemdi➔ click ok
-

- 

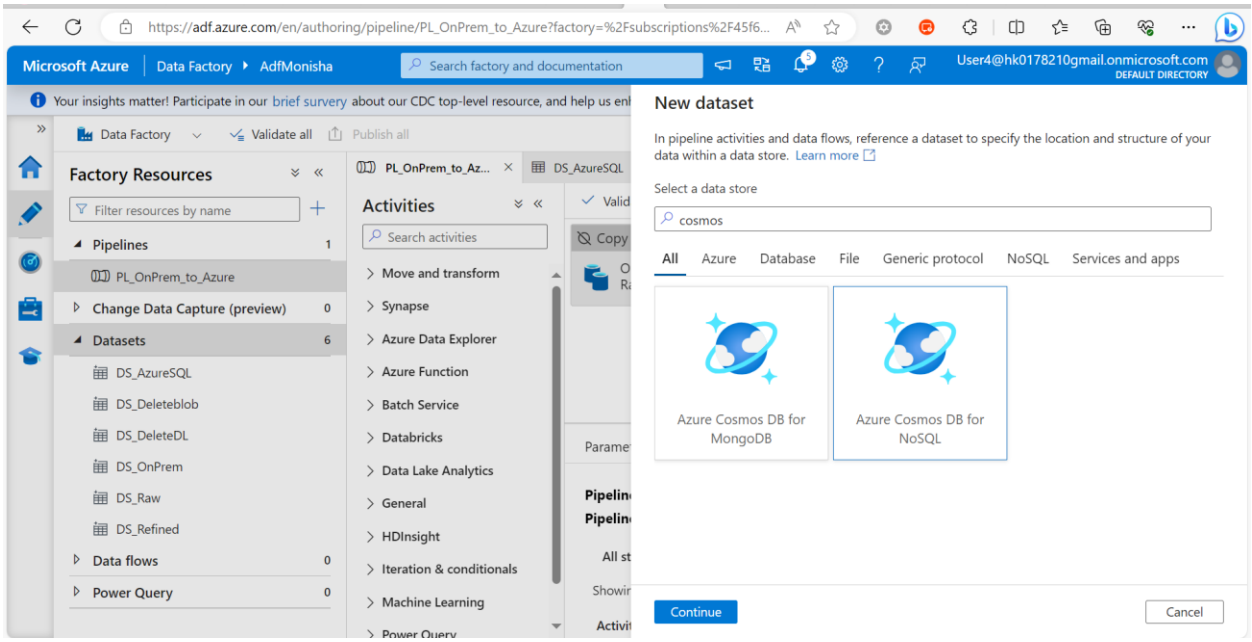
- 

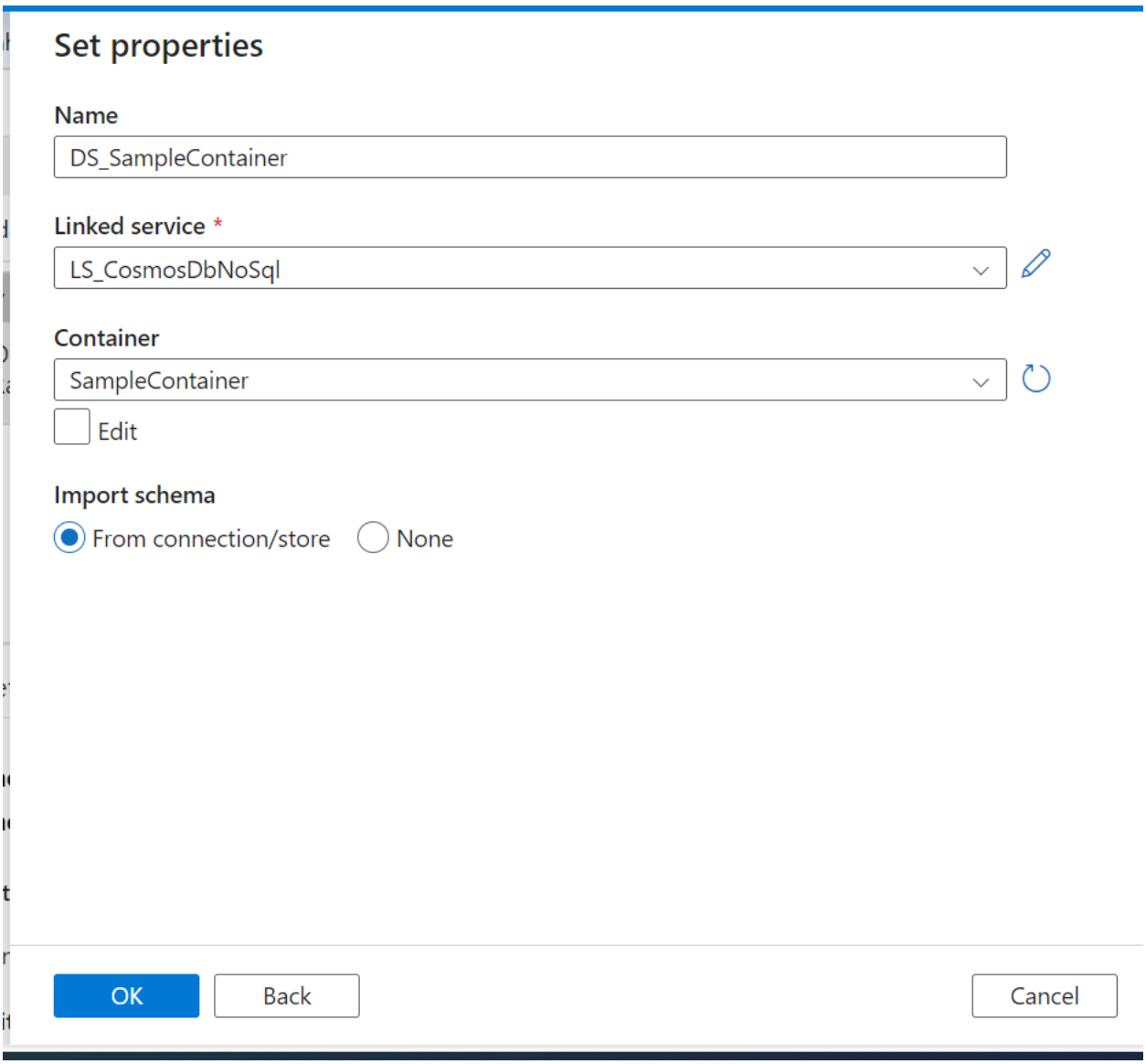
- Go to data factory ➔ Linked service ➔cosmosdb nosql ➔ continue

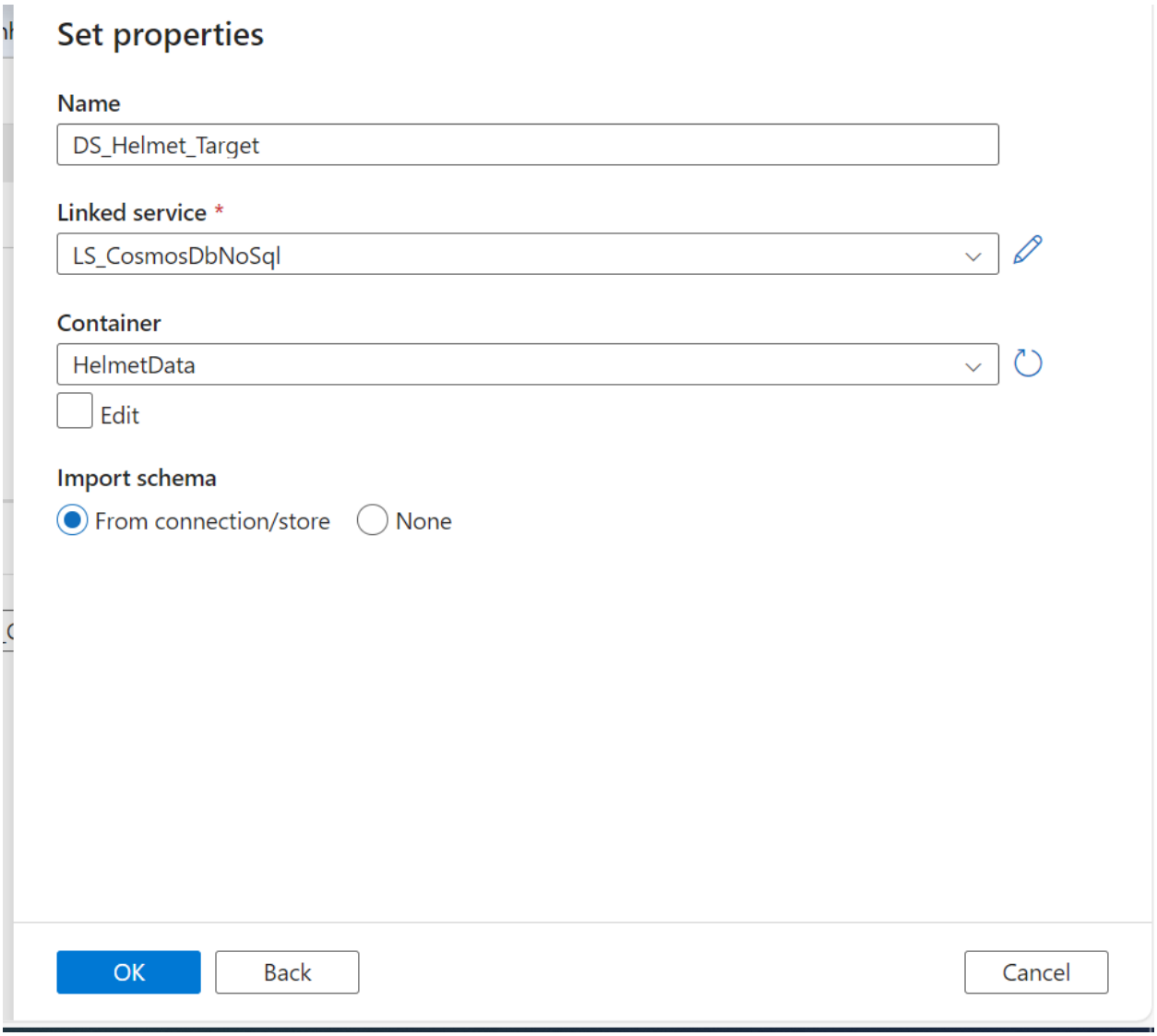
- 

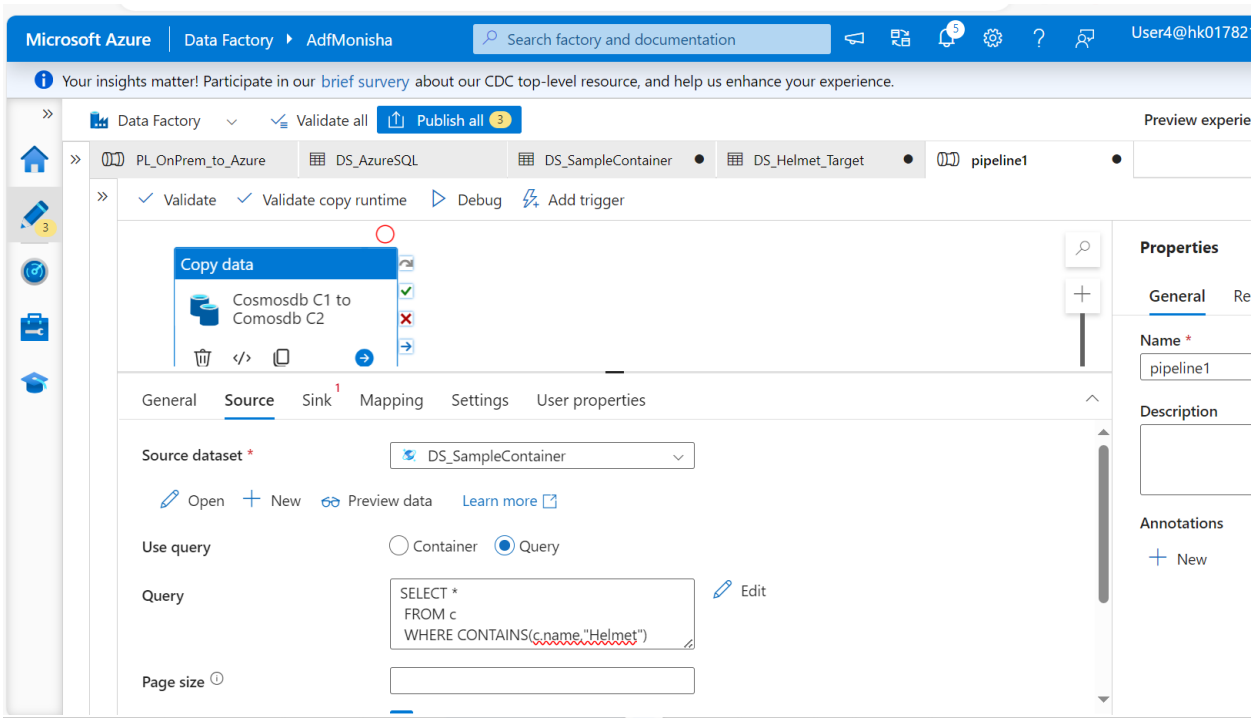
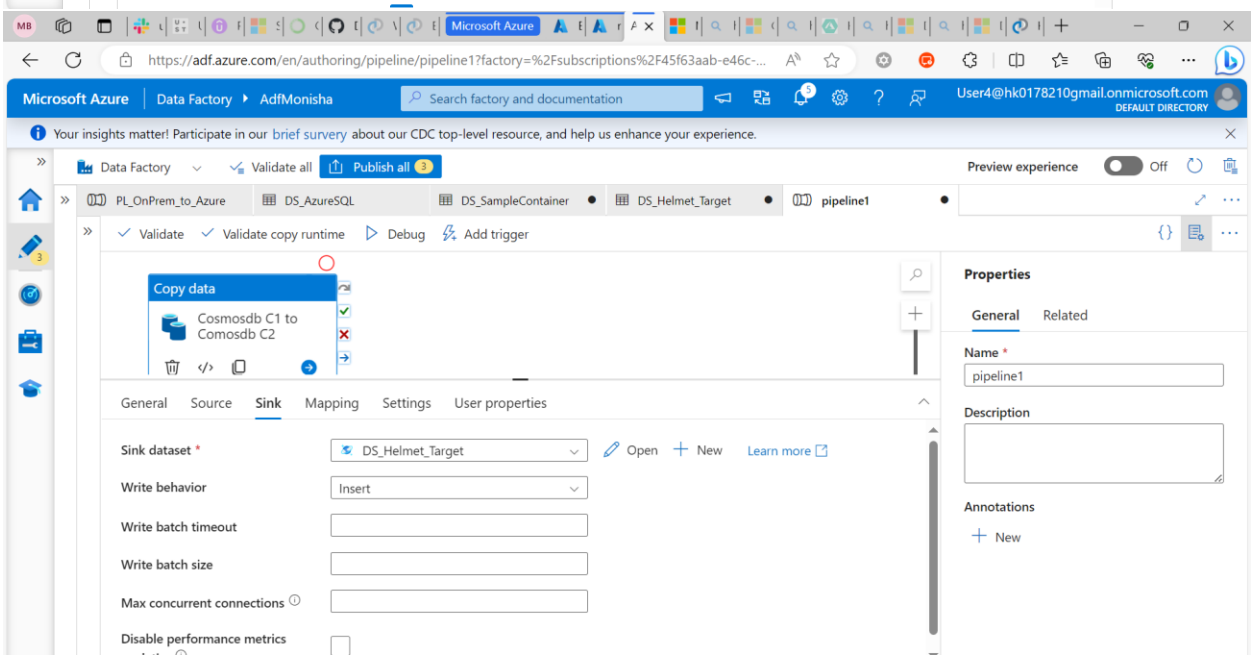
- 

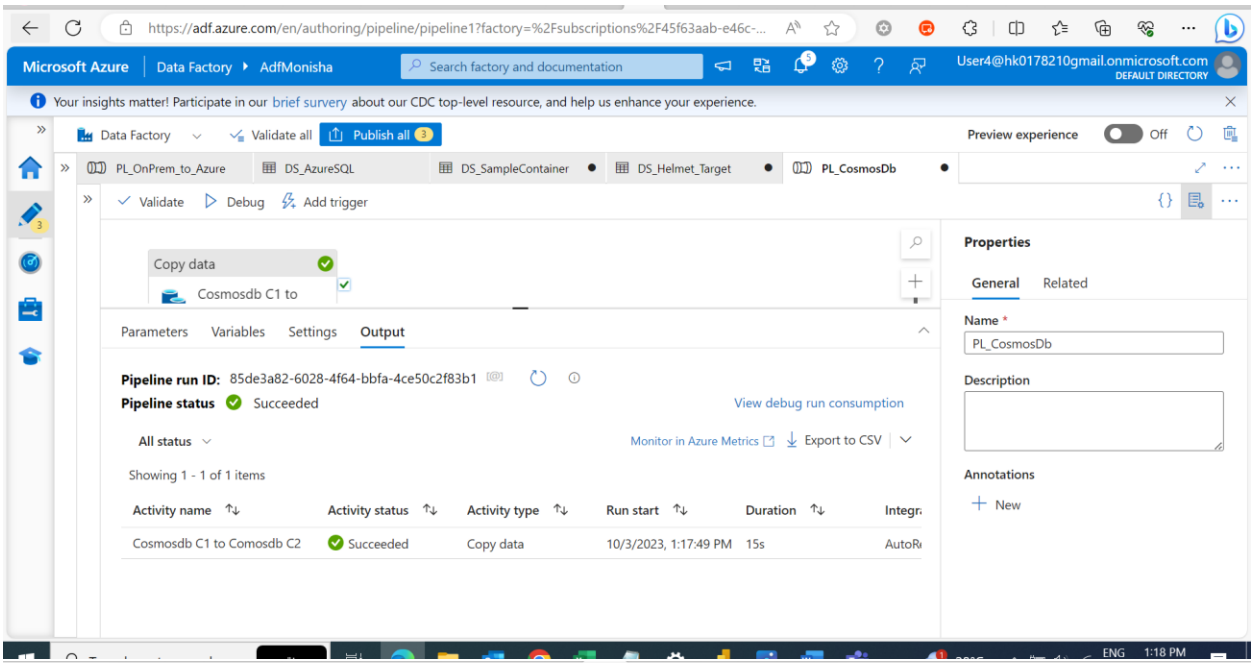
- Create dataset

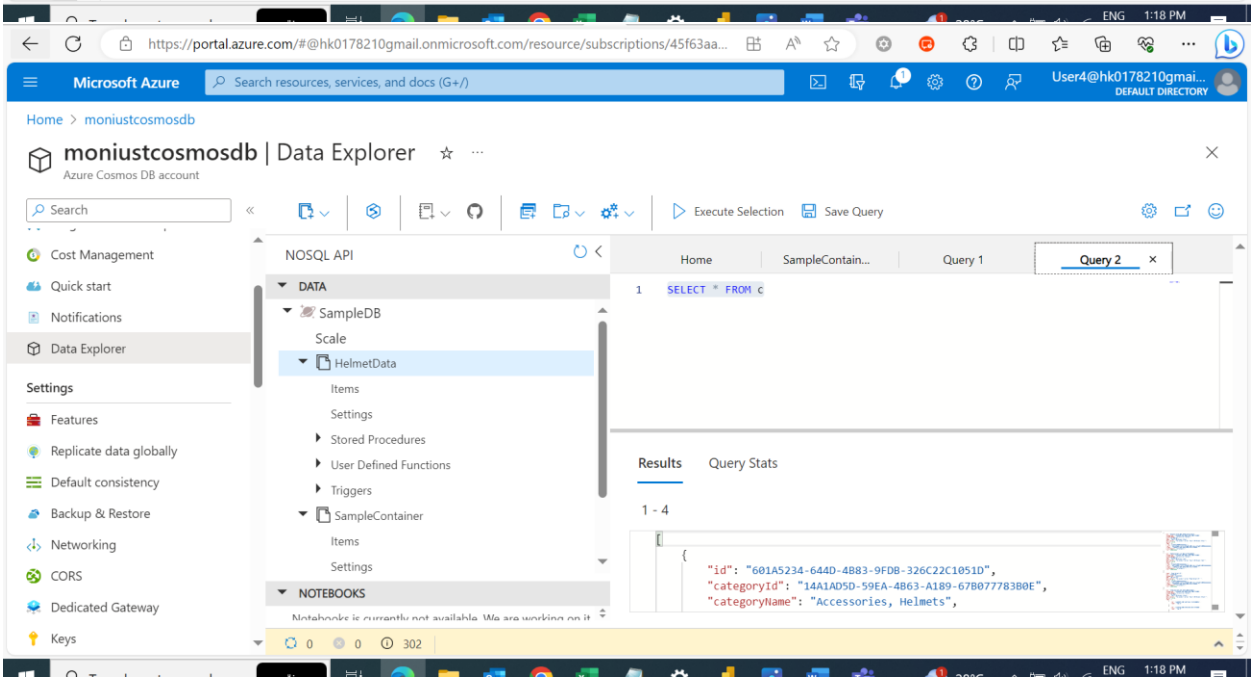
- 
https://adf.azure.com/en/authoring/pipeline/PL_OnPrem_to_Azure?factory=%2Fsubscriptions%2F45f6...
Microsoft Azure
Data Factory
AdfMonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enh
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines 1
PL_OnPrem_to_Azure
Change Data Capture (preview) 0
Datasets 6
DS_AzureSQL
DS_Deleteblob
DS_DeleteDL
DS_OnPrem
DS_Raw
DS_Refined
Data flows 0
Power Query 0
PL_OnPrem_to_Az...
DS_AzureSQL
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Power Query
Valid
Copy
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
cosmos
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Azure Cosmos DB for MongoDB
Azure Cosmos DB for NoSQL
Continue
Cancel

- 
Set properties
Name
DS_SampleContainer
Linked service *
LS_CosmosDbNoSql
Container
SampleContainer
Edit
Import schema
From connection/store
None
OK
Back
Cancel

- 
Set properties
Name
DS_Helmet_Target
Linked service *
LS_CosmosDbNoSql
Container
HelmetData
Edit
Import schema
From connection/store
None
OK
Back
Cancel

- 
Microsoft Azure
Data Factory
AdfMonisha
Search factory and documentation
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all 3
PL_OnPrem_to_Azure
DS_AzureSQL
DS_SampleContainer
DS_Helmet_Target
pipeline1
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Cosmosdb C1 to Comosdb C2
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
DS_SampleContainer
Open
New
Preview data
Learn more
Use query
Container
Query
Query
SELECT *
FROM c
WHERE CONTAINS(c.name,"Helmet")
Edit
Page size
Properties
General
Name *
pipeline1
Description
Annotations
New
- 
https://adf.azure.com/en/authoring/pipeline/pipeline1?factory=%2Fsubscriptions%2F45f63aab-e46c-...
Microsoft Azure
Data Factory
AdfMonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all 3
Preview experience
Off
PL_OnPrem_to_Azure
DS_AzureSQL
DS_SampleContainer
DS_Helmet_Target
pipeline1
Validate
Validate copy runtime
Debug
Add trigger
Copy data
Cosmosdb C1 to Comosdb C2
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
DS_Helmet_Target
Open
New
Learn more
Write behavior
Insert
Write batch timeout
Write batch size
Max concurrent connections
Disable performance metrics
Properties
General
Related
Name *
pipeline1
Description
Annotations
New

- 

- 

HOW TO CONNECT COSMOSDB TO LOCAL:

Inside vm

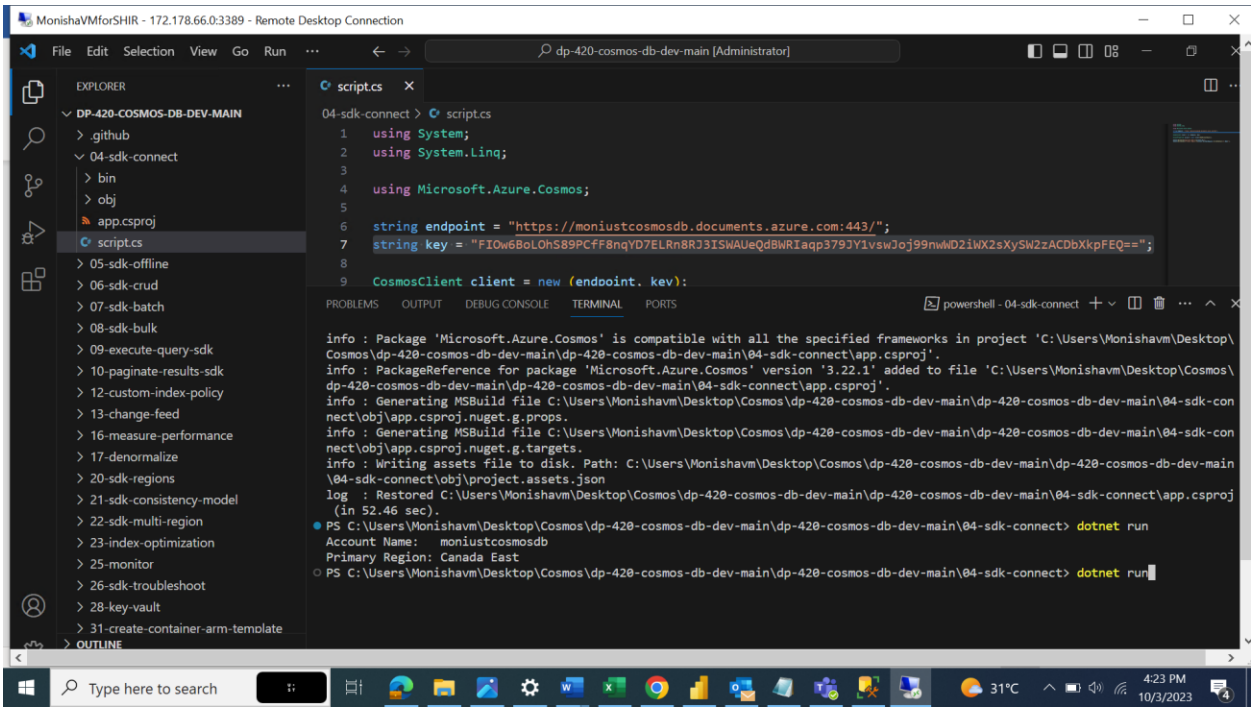
- Download and install vs code: [Download Visual Studio Code - Mac, Linux, Windows]
- Download .net runtime *64
- https://dotnet.microsoft.com/en-us/download/dotnet/7.0
- Download .net SDK
- In desktop ➔ create folder named : cosmos

- https://github.com/microsoftlearning/dp-420-cosmos-db-dev from this link if u have git clone this file if not download zip and extract that file inside Cosmos folder
- Open that extracted file with VS Code
- Go to 04-sdk-Connect ➔ open script.cs ➔

put below code in script.cs

```
using System;

using System.Linq;


using Microsoft.Azure.Cosmos;


string endpoint = "<cosmos-endpoint>";

string key = "<cosmos-key>";


CosmosClient client = new (endpoint, key);


AccountProperties account = await client.ReadAccountAsync();


Console.WriteLine($"Account Name:\t{account.Id}");

Console.WriteLine($"Primary Region:\t{account.WritableRegions.FirstOrDefault()?.Name}");
```

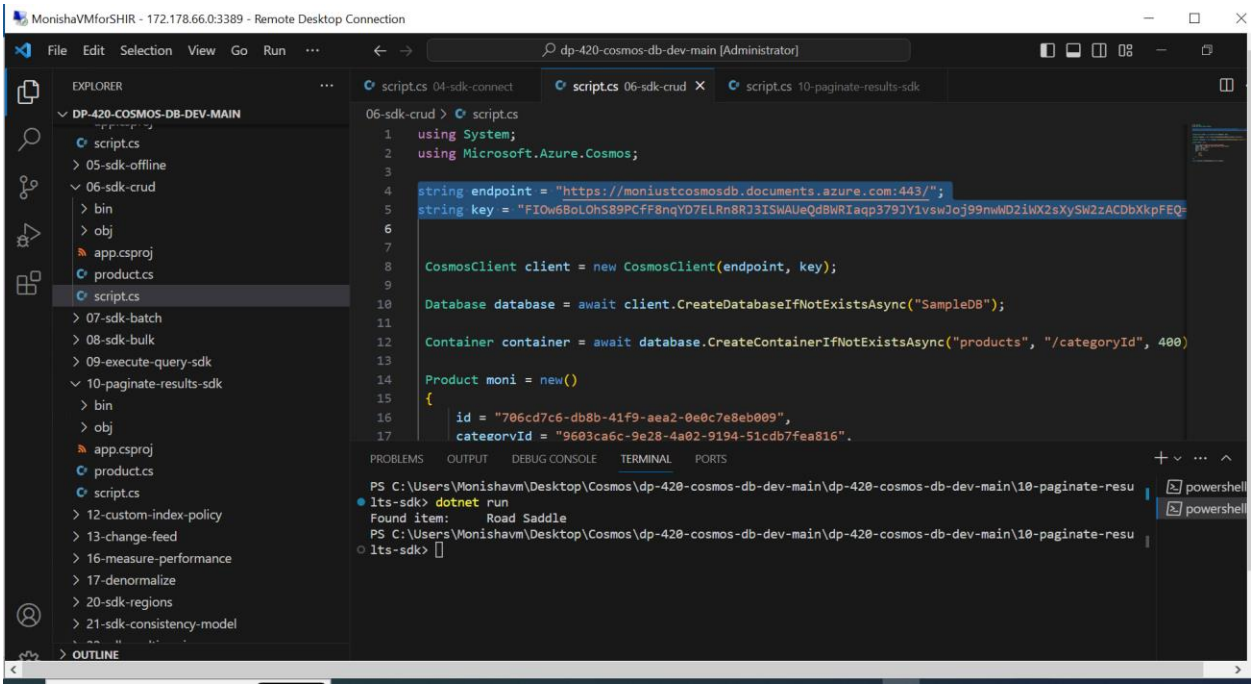
- Go to cosmosdb ➔ key ➔ copy url
- ```
  string endpoint = "https://moniustcosmosdb.documents.azure.com:443/";
  ```
- copy primary key ➔ paste
- ```
  string key = "FIOw6BoLOhS89PCfF8nqYD7ELRn8RJ3ISWAUeQdBWRIaqp379JY1vswJoj99nwWD2iWX2sXySW2zACDbXkpFEQ==";
  ```
- save
- right click on script.cs ➔ open in integrated terminal ➔ paste the below command and run dotnet add package Microsoft.Azure.Cosmos --version 3.22.1
- dotnet run



-
- HOW TO DO CRUD, READ/WRITE OPERATION:
- Open script.cs from 06-sdk-offline ➔ paste script ➔ change url and key ➔ type dotnet run➔check in cosmosdb to see new product container

```
using System;
 using Microsoft.Azure.Cosmos;

 string endpoint = "https://madhangicosmosdb.documents.azure.com:443/";

 string key =
"QCh8V9HVHkDiePiYVRrgezRlK04yFb2N6eX3yLmEAO0jr9m3S23yZi98ProAmwLykbxXxLRzAnbLA
CDbfS3KFA==";


 CosmosClient client = new CosmosClient(endpoint, key);

 Database database = await client.CreateDatabaseIfNotExistsAsync("SampleDB");

 Container container = await database.CreateContainerIfNotExistsAsync("products",
"/categoryId", 400);

 Product saddle = new()
 {
    id = "706cd7c6-db8b-41f9-aea2-0e0c7e8eb009",
    categoryId = "9603ca6c-9e28-4a02-9194-51cdb7fea816",
    name = "Road Saddle",
    price = 45.99d,
```

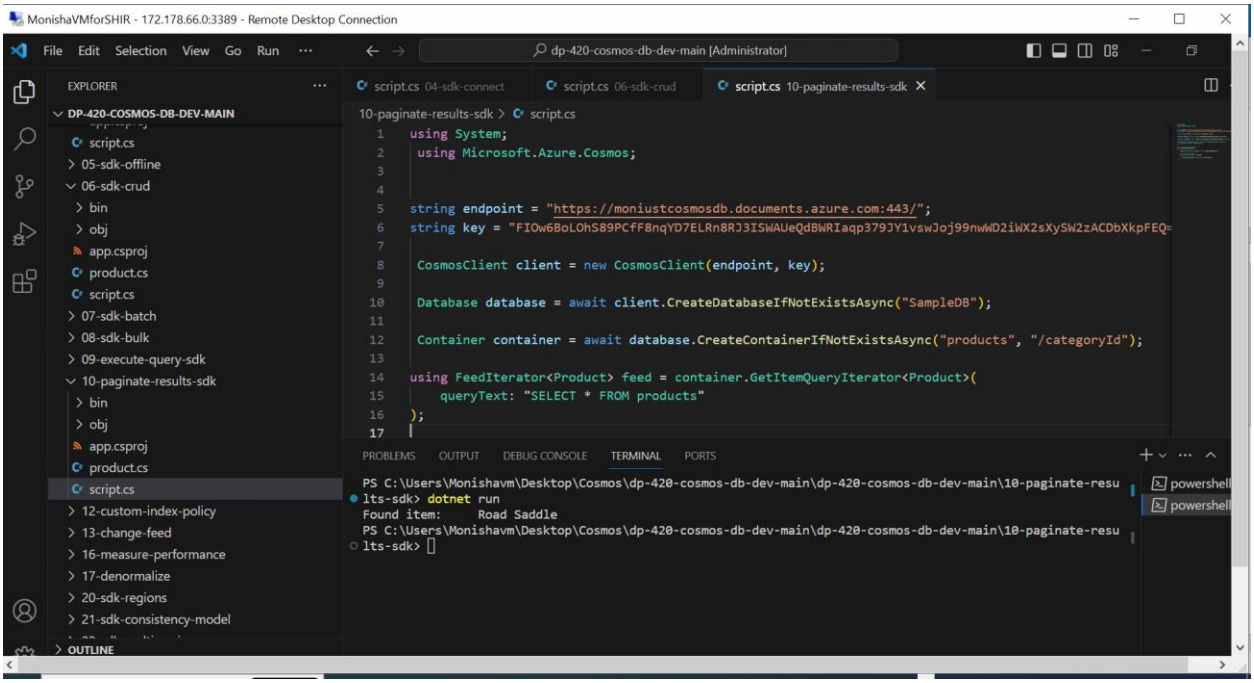
```
    tags = new string[]
    {
        "tan",
        "new",
        "crisp"
    }
};

await container.CreateItemAsync<Product>(saddle);
```

- 

- In 10-paginate-results-sdk, script.cs paste script ➔ change url, key ➔ dotnet run

```
using System;
 using Microsoft.Azure.Cosmos;

 string endpoint = "https://madhangicosmosdb.documents.azure.com:443/";

 string key =
"QCh8V9HVHkDiePiYVRrgezRlK04yFb2N6eX3yLmEAO0jr9m3S23yZi98ProAmwLykbxXxLRzAnbLA
CDbfS3KFA==";

 CosmosClient client = new CosmosClient(endpoint, key);
```

```
Database database = await client.CreateDatabaseIfNotExistsAsync("SampleDB");

Container container = await database.CreateContainerIfNotExistsAsync("products",
"/categoryId");

using FeedIterator<Product> feed = container.GetItemQueryIterator<Product>(
   queryText: "SELECT * FROM products"
);

// Iterate query result pages
while (feed.HasMoreResults)
{
   FeedResponse<Product> response = await feed.ReadNextAsync();

   // Iterate query results
   foreach (Product item in response)
   {
      Console.WriteLine($"Found item:\t{item.name}");
   }
}
```

- 

To add one more container and new database:

In 06-sdk, change database name to FourOctober and Container name to address.

```
using System;
using Microsoft.Azure.Cosmos;
```

```
string endpoint = "https://moniustcosmosdb.documents.azure.com:443/";
string key =
"FIOw6BoLOhS89PCfF8nqYD7ELRn8RJ3ISWAUeQdBWRIaqp379JY1vswJoj99nwWD2iWX2sXySW2zACDb
XkpFEQ==";


 CosmosClient client = new CosmosClient(endpoint, key);

 Database database = await client.CreateDatabaseIfNotExistsAsync("FourOctober");

 Container container = await database.CreateContainerIfNotExistsAsync("address",
"/categoryId", 400);

 Product moni = new()
 {
     id = "406cd7c6-db8b-41f9-aea2-0e0c7e8eb012",
     categoryId = "2745ca6c-9e28-4a02-9194-51cdb7fea815",
     name = "Address",
     price = 45.99d,
     tags = new string[]
     {
         "tan",
         "new",
         "crisp"
     }
 };

 await container.CreateItemAsync<Product>(moni);
```

To Query go to 10-sdk,

```
using System;
 using Microsoft.Azure.Cosmos;


string endpoint = "https://moniustcosmosdb.documents.azure.com:443/";
string key =
"FIOw6BoLOhS89PCfF8nqYD7ELRn8RJ3ISWAUeQdBWRIaqp379JY1vswJoj99nwWD2iWX2sXySW2zACDb
XkpFEQ==";

 CosmosClient client = new CosmosClient(endpoint, key);

 Database database = await client.CreateDatabaseIfNotExistsAsync("FourOctober");
```

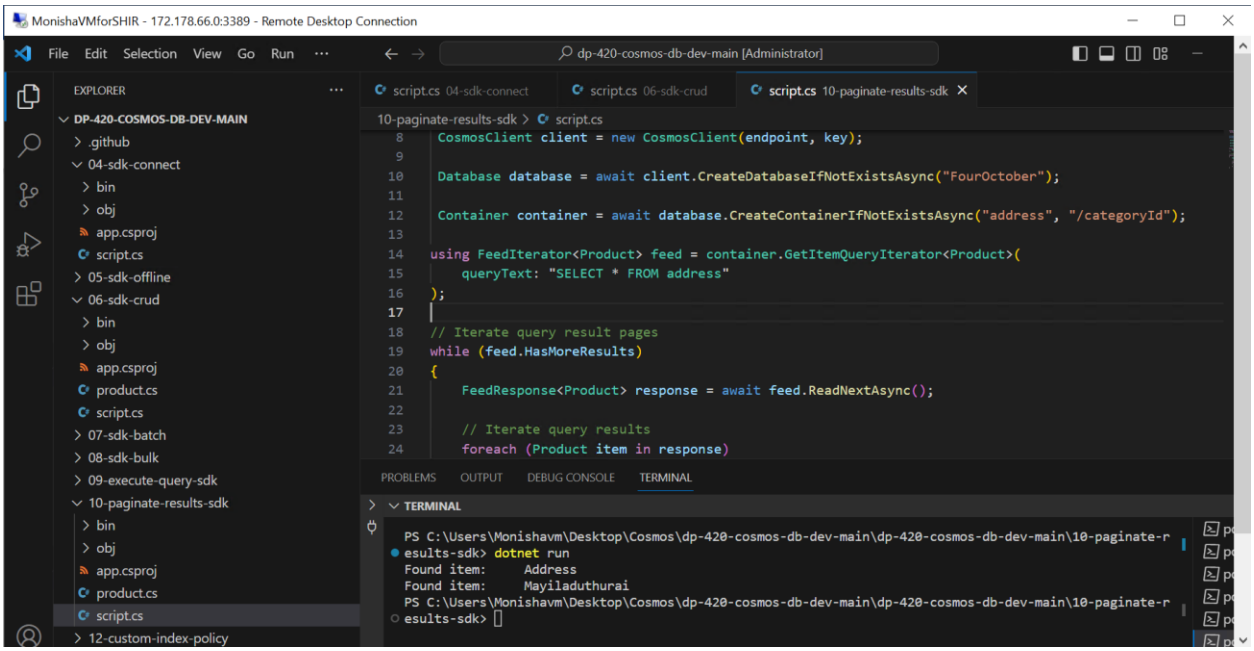
```
 Container container = await database.CreateContainerIfNotExistsAsync("address",
"/categoryId");

using FeedIterator<Product> feed = container.GetItemQueryIterator<Product>(
    queryText: "SELECT * FROM address"
);

// Iterate query result pages
while (feed.HasMoreResults)
{
    FeedResponse<Product> response = await feed.ReadNextAsync();

    // Iterate query results
    foreach (Product item in response)
    {
        Console.WriteLine($"Found item:\t{item.name}");
    }
}
```

To add one more item, in 06-sdk-crud, change item name and name:

```
using System;
using Microsoft.Azure.Cosmos;

string endpoint = "https://moniustcosmosdb.documents.azure.com:443/";
string key =
"FIOw6BoLOhS89PCfF8nqYD7ELRn8RJ3ISWAUeQdBWRIaqp379JY1vswJoj99nwWD2iWX2sXySW2zACDb
XkpFEQ==";


 CosmosClient client = new CosmosClient(endpoint, key);

 Database database = await client.CreateDatabaseIfNotExistsAsync("FourOctober");

 Container container = await database.CreateContainerIfNotExistsAsync("address",
"/categoryId", 400);

 Product monisha = new()
 {
     id = "506cd7c6-db8b-41f9-aea2-0e0c7e8eb012",
     categoryId = "1745ca6c-9e28-4a02-9194-51cdb7fea815",
     name = "Mayiladuthurai",
     price = 45.99d,
     tags = new string[]
     {
```

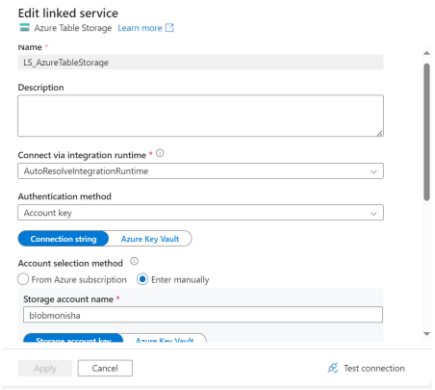
```
        "tan",
        "new",
        "crisp"
    }
};

await container.CreateItemAsync<Product>(monisha);
```

then if u run query: we can see two items

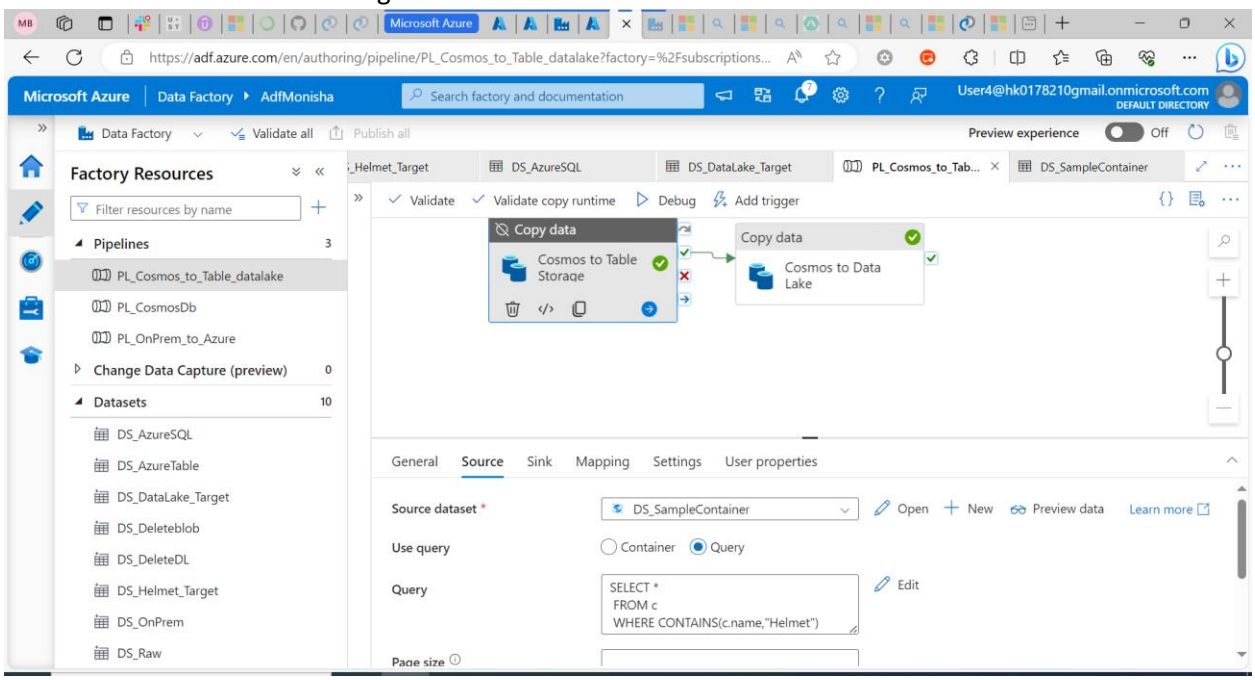


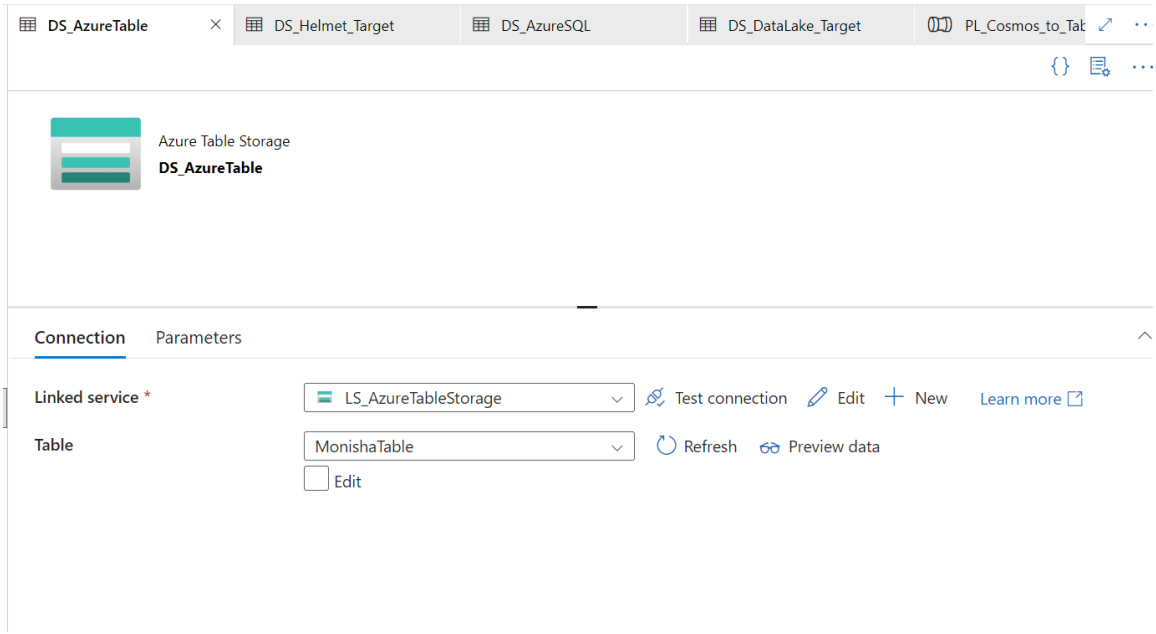
TO LOAD CONTAINER DATA TO TABLE STORAGE:

1. Go to blob storage ➔ Click Table
2. +Table ➔ Table Name: Monisha Table
3. Go to data factory ➔ Create Linked service ➔ choose Table Storage ➔ Table Name ➔Create (use datalake or blob to create table)

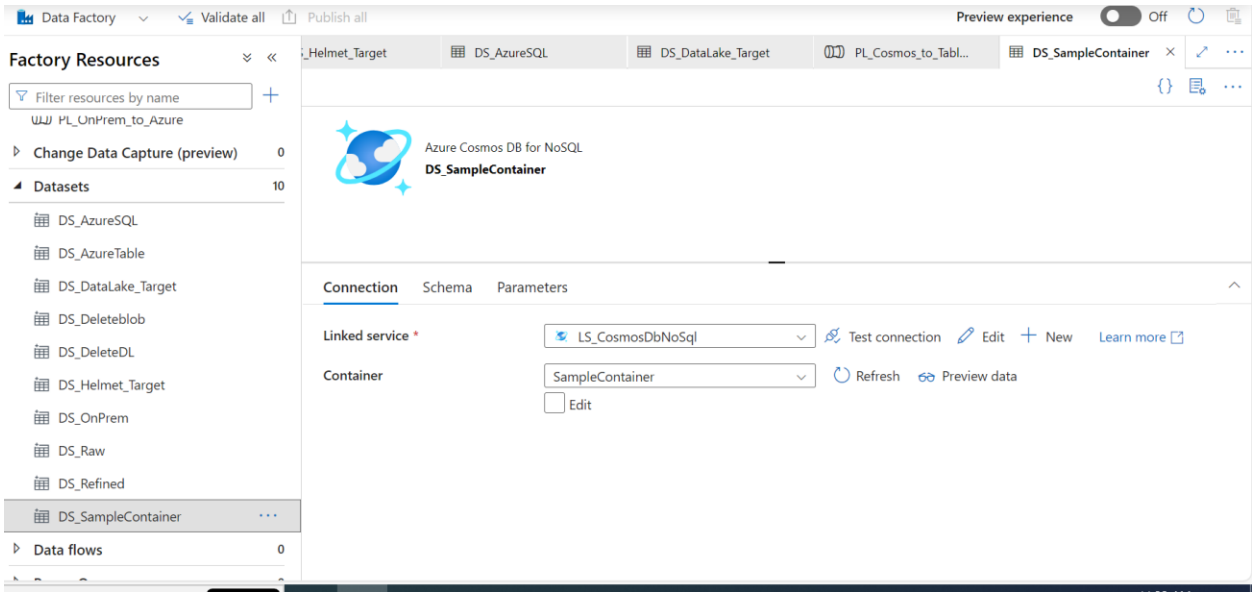


5) Create a dataset ➔ Table storage ➔ create





6) Another dataset ➔ cosmos nosql ➔



7)

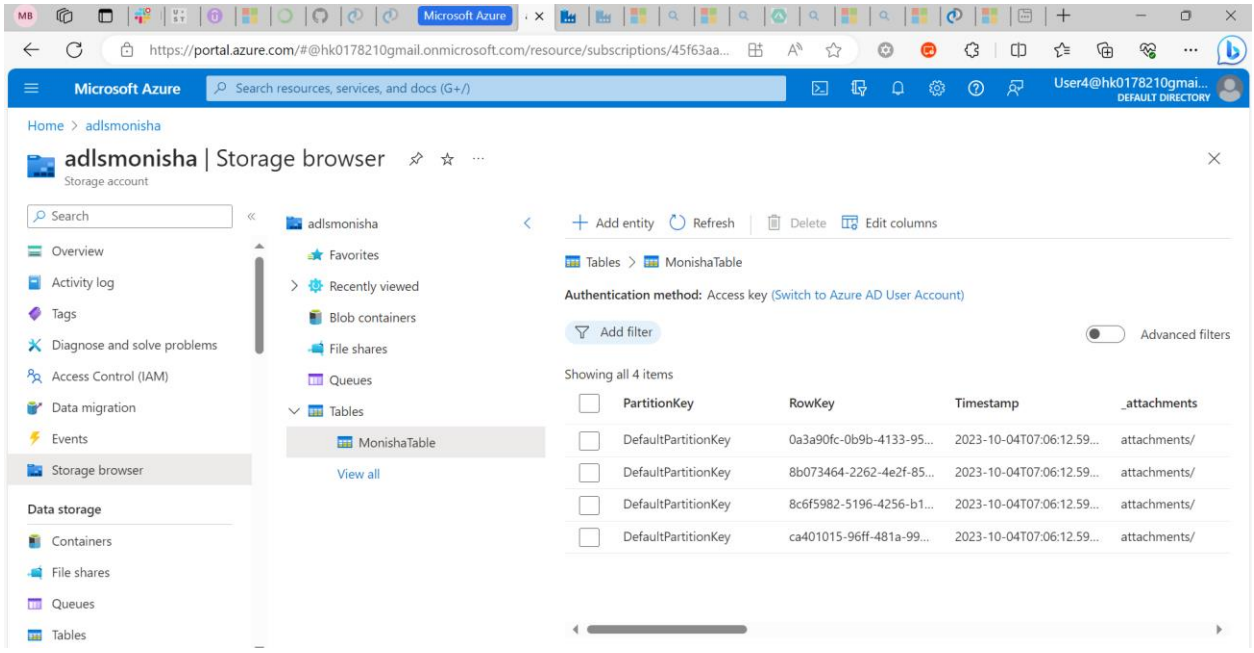
8) Query ➔

```
SELECT *
 FROM c
 WHERE CONTAINS(c.name,"Helmet")
```

9)sink ➔ Ds_AzureTable

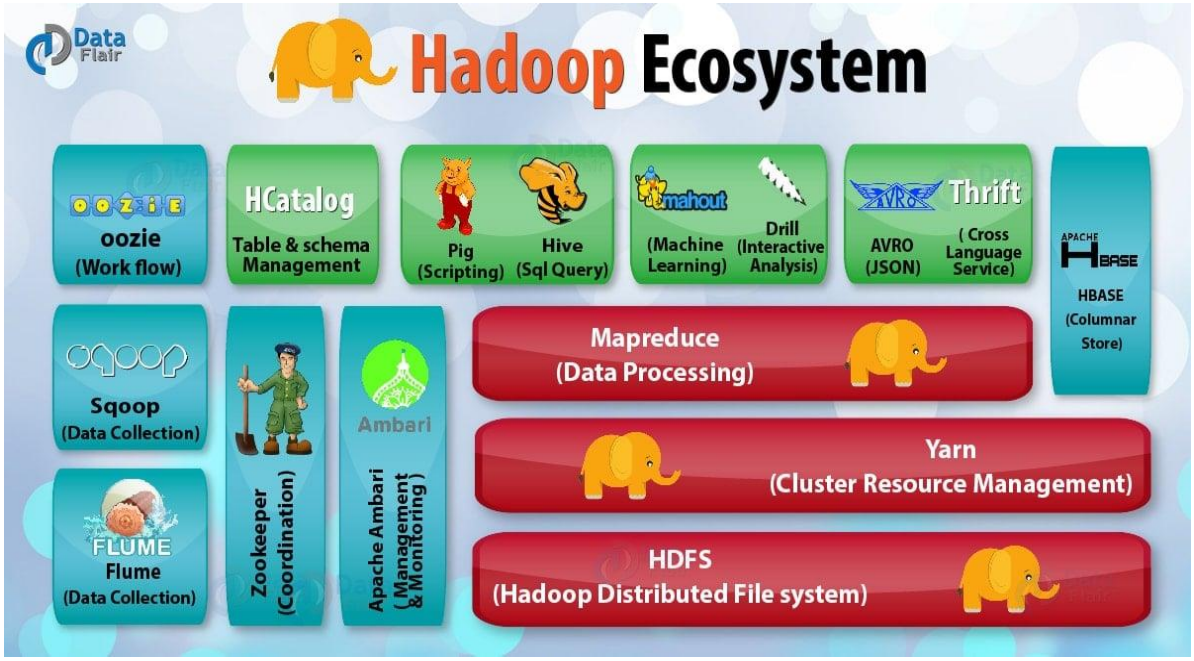
10)debug

11) to check in table data stored or not : go to datalake ➔ click Storage browser ➔ tabes ➔ click on table name ➔ validate data by checking data with cosmos

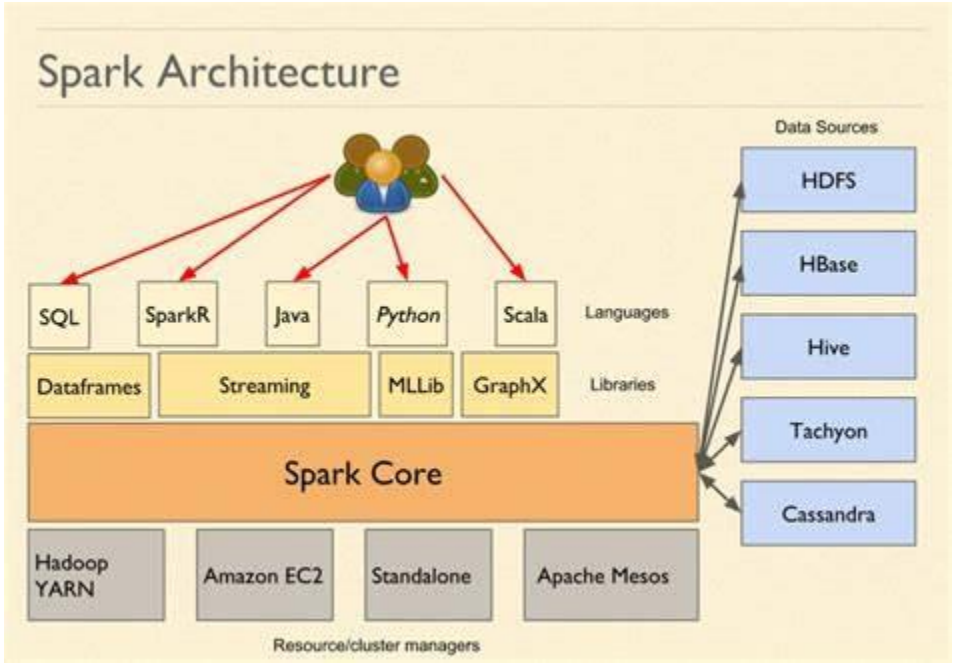


BIG DATA:

Hadoop Ecosystem:



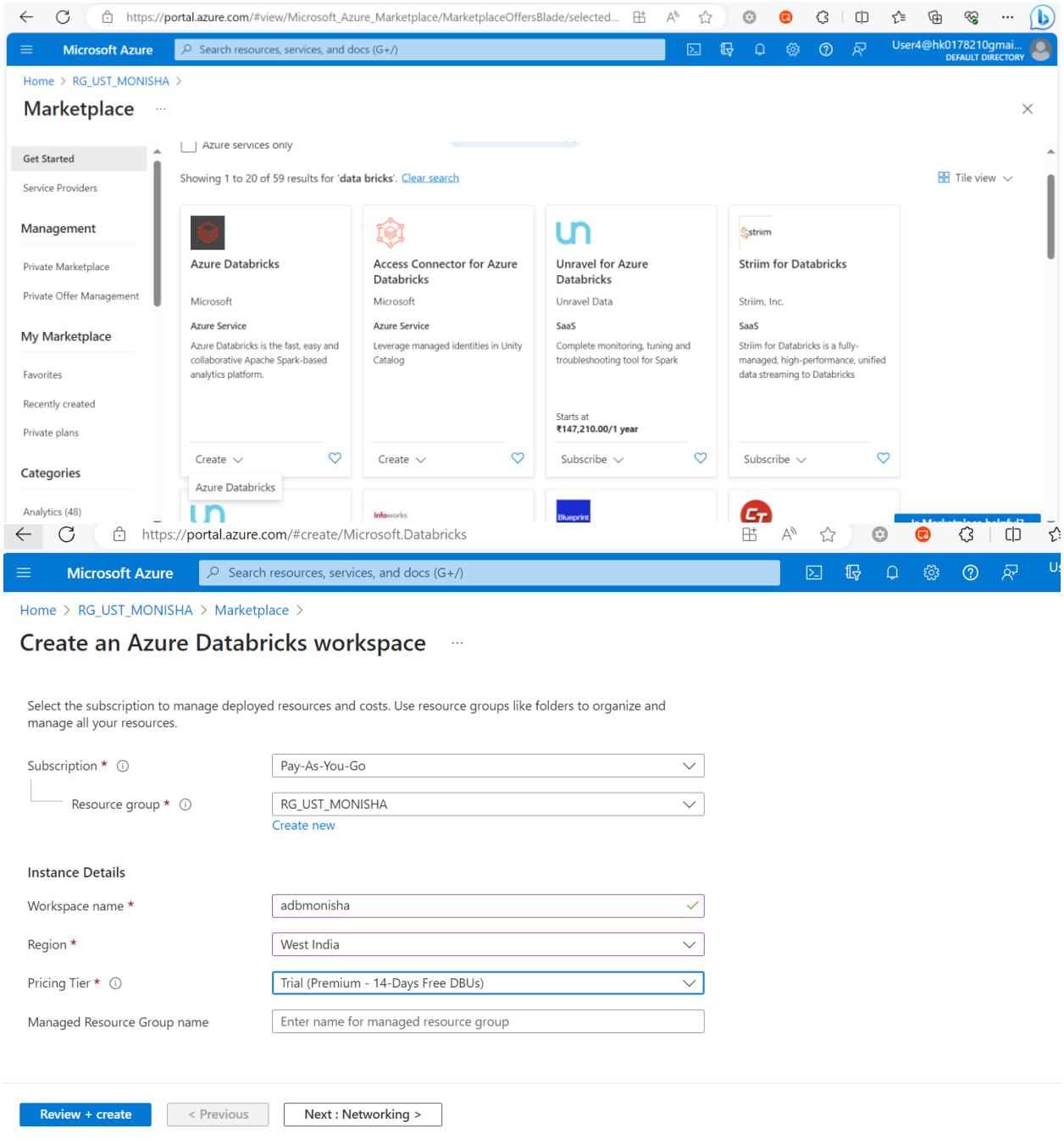
Spark Architecture:

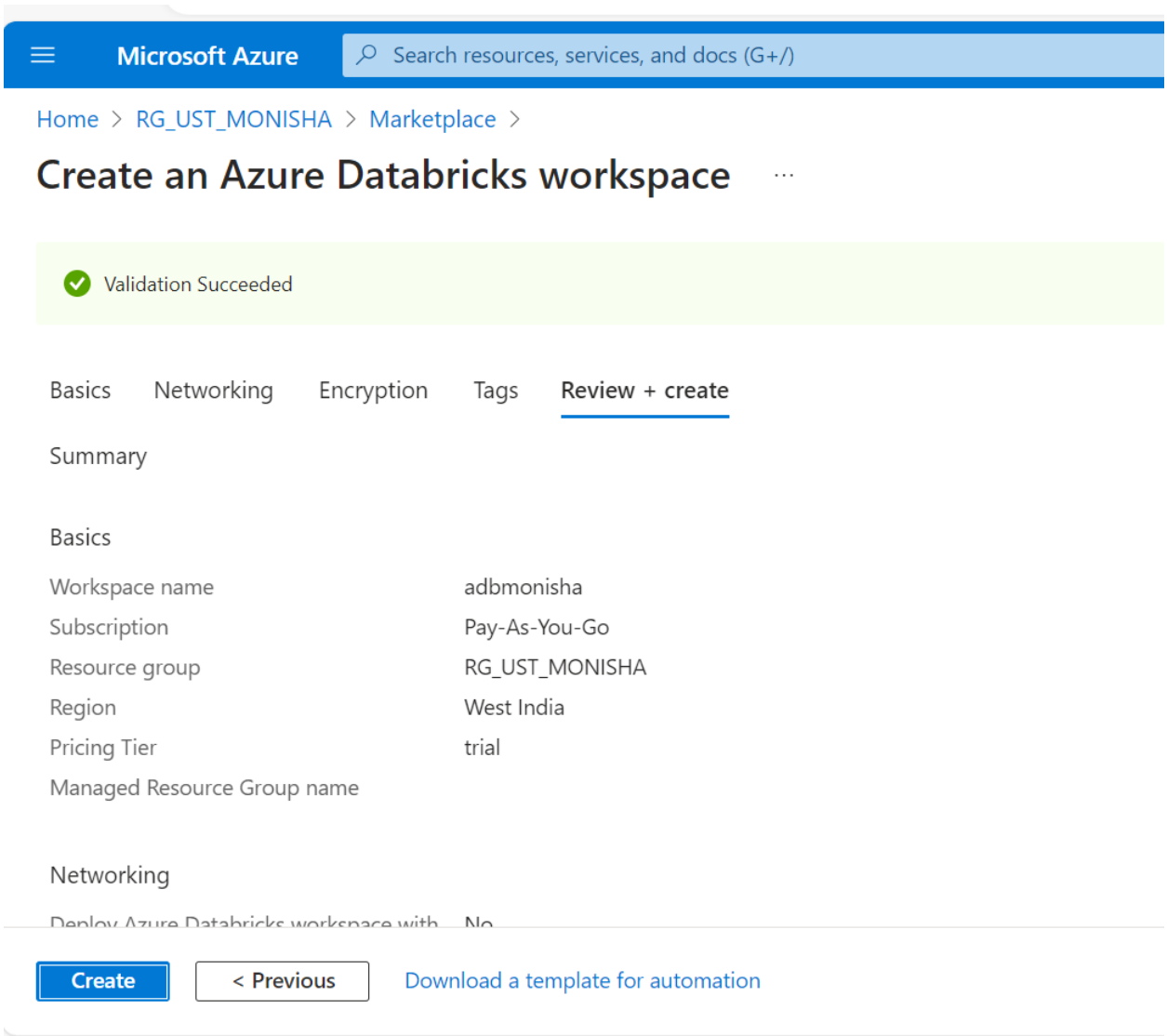


HOW TO CRETE DATA BRICKS:

- Go inside Resource Group ➔ click +create
- Workspace Name : adbmonisha
- Region: west India

- Pricing Tier : Trial (as we are testing)
- Rest everything go with default
- Tags: Project: ust_training , Onwer: Monisha
- Review+ create
- create

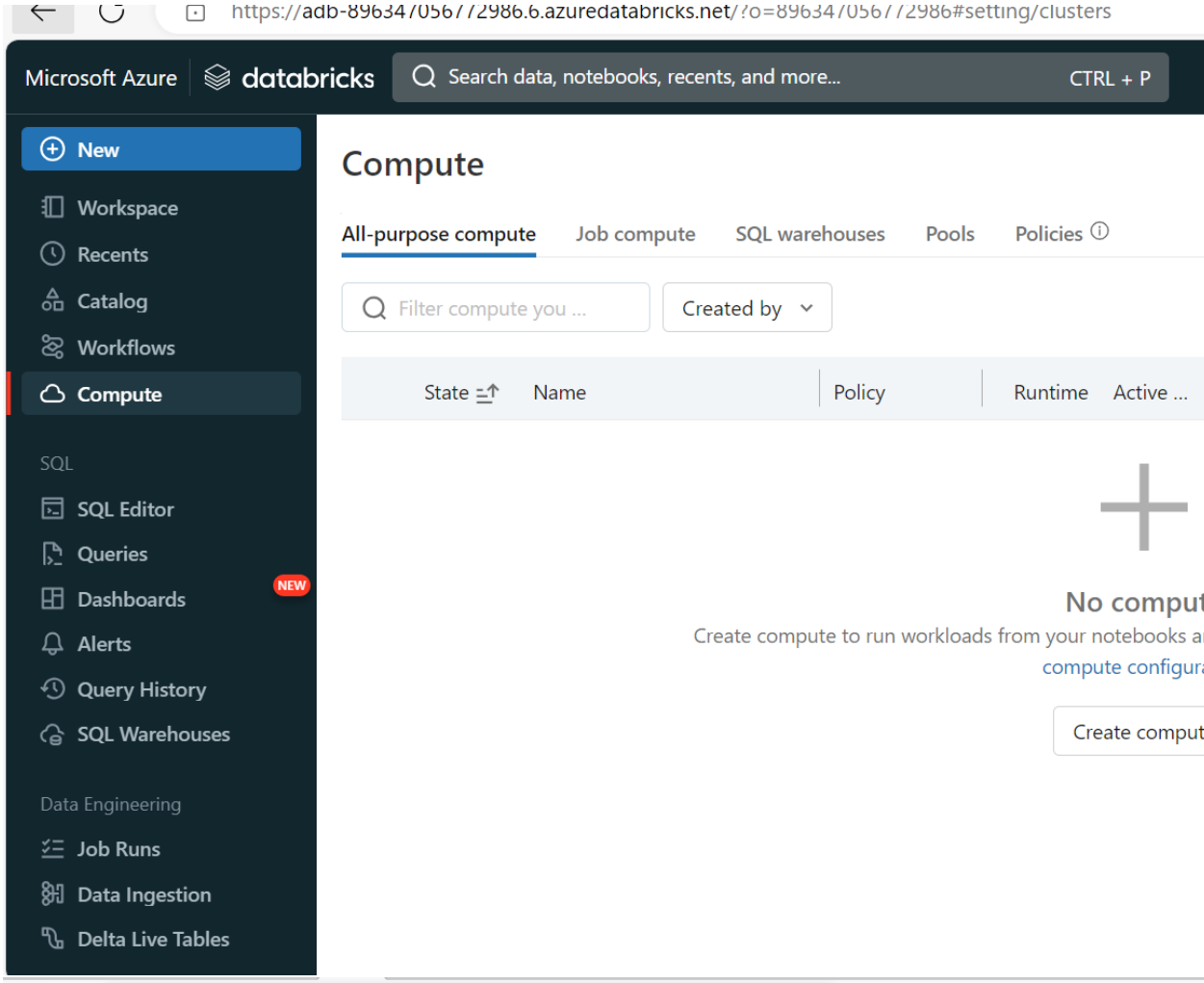




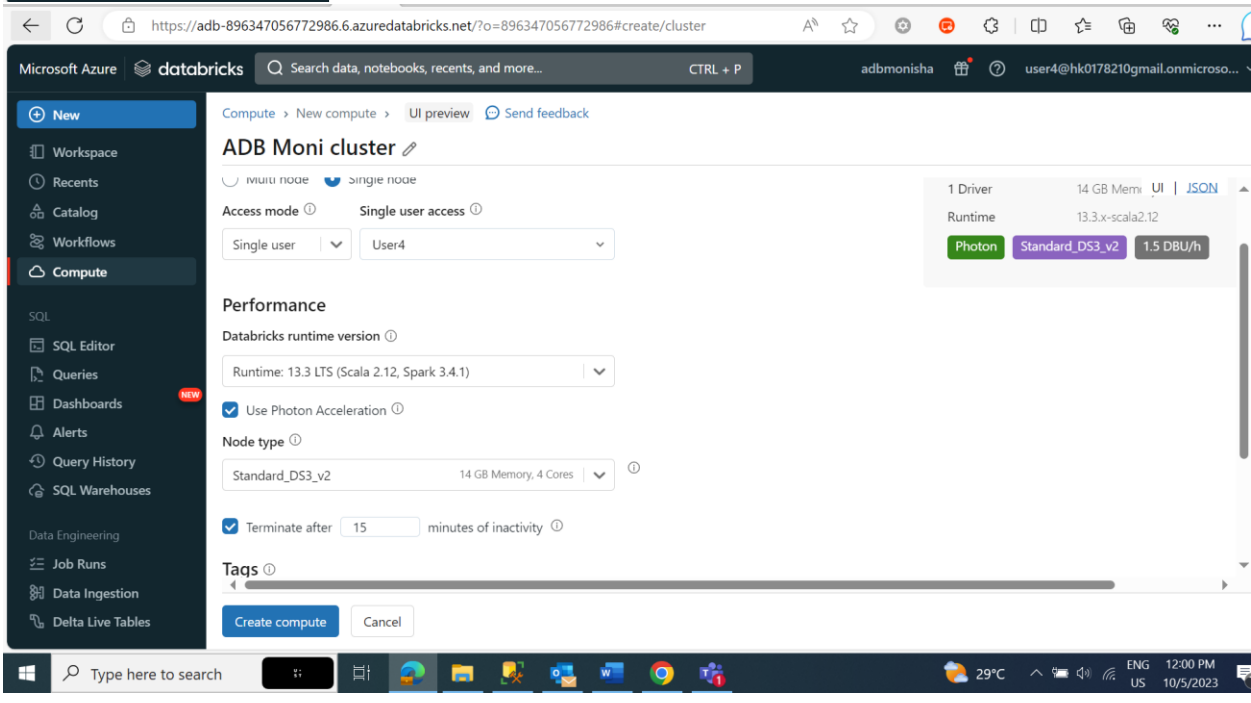
HOW TO CREATE COMPUTE:

1. LAUNCH workspace ➔ COMPUTE ➔

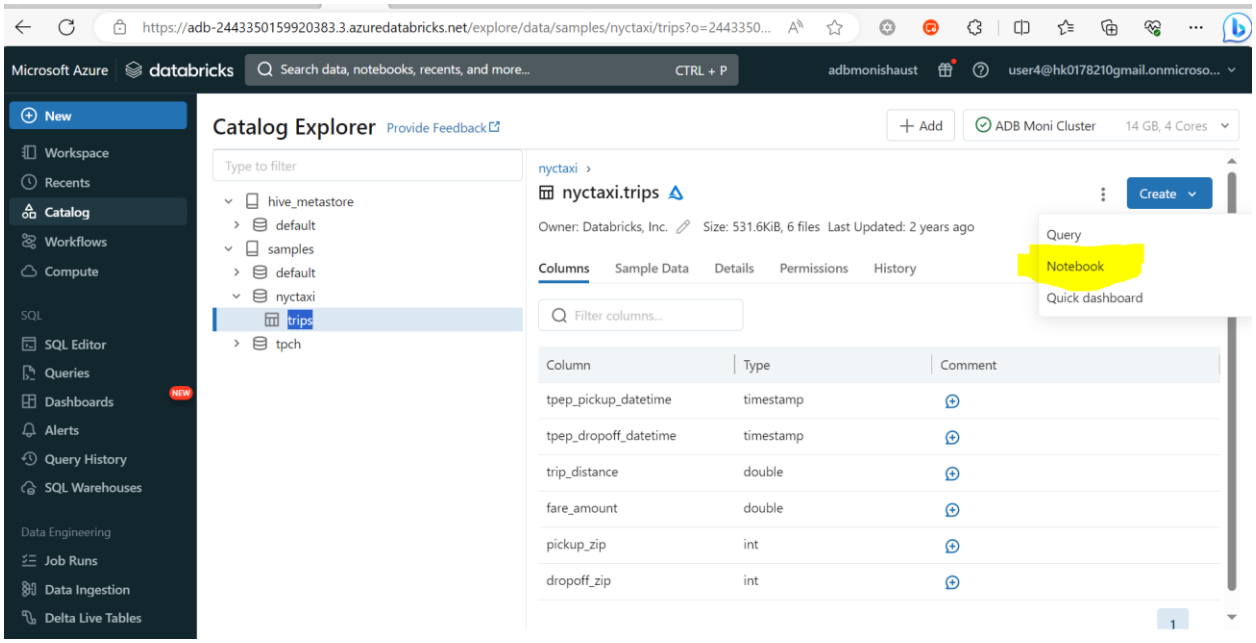
2.


https://adb-89634705677 2986.6.azuredatabricks.net/?o=89634705677 2986#setting/clusters
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
NEW
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Compute
All-purpose compute
Job compute
SQL warehouses
Pools
Policies
Filter compute you ...
Created by
State
Name
Policy
Runtime
Active ...
No comput
Create compute to run workloads from your notebooks a
compute configura
Create comput

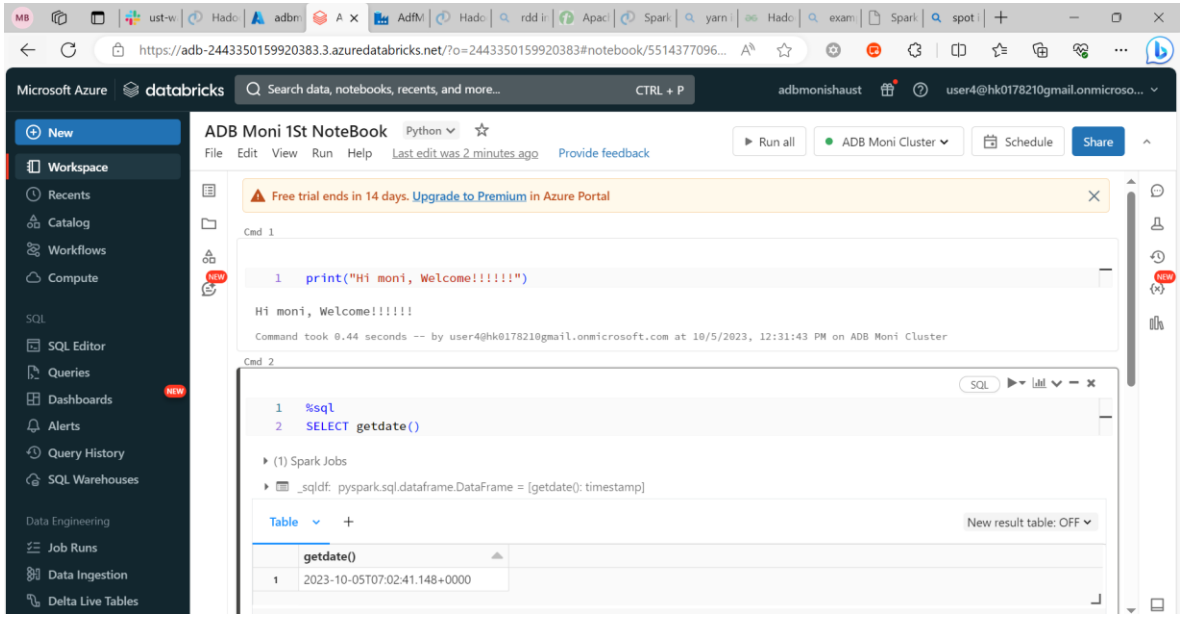
3.


https://adb-896347056772986.6.azuredatabricks.net/?o=896347056772986#create/cluster
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
adbmonisha
user4@hk0178210gmail.onmicroso...
Compute › New compute ›
UI preview
Send feedback
ADB Moni cluster
Access mode
Single user access
Single user
User4
1 Driver
14 GB Mem
UI | JSON
Runtime
13.3.x-scala2.12
Photon
Standard_DS3_v2
1.5 DBU/h
Performance
Databricks runtime version
Runtime: 13.3 LTS (Scala 2.12, Spark 3.4.1)
Use Photon Acceleration
Node type
Standard_DS3_v2
14 GB Memory, 4 Cores
Terminate after
15
minutes of inactivity
Tags
Create compute
Cancel
Type here to search
29°C
ENG
US
12:00 PM
10/5/2023

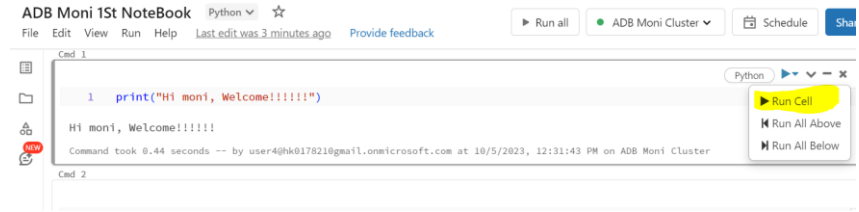
Another way of querying:



To create notebook ➔ new ➔ notebook

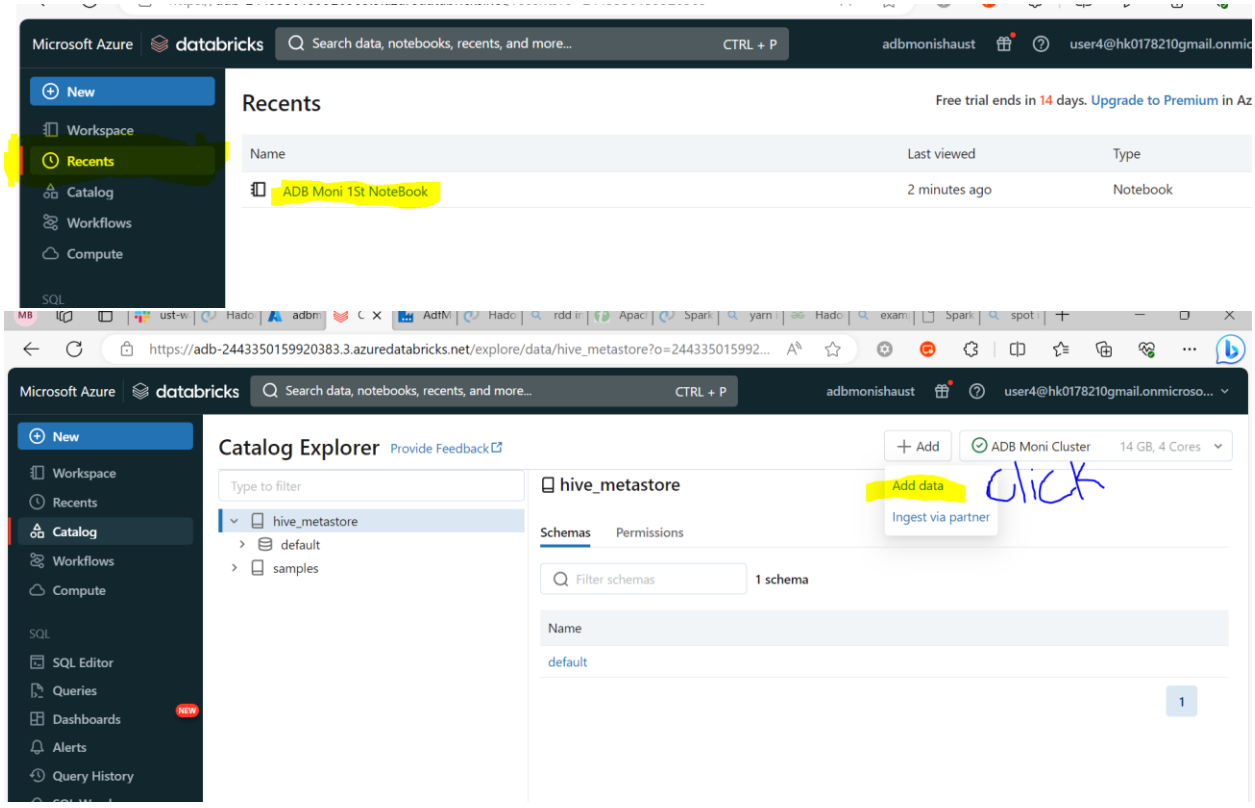


To run cell : run cell Or shift + enter

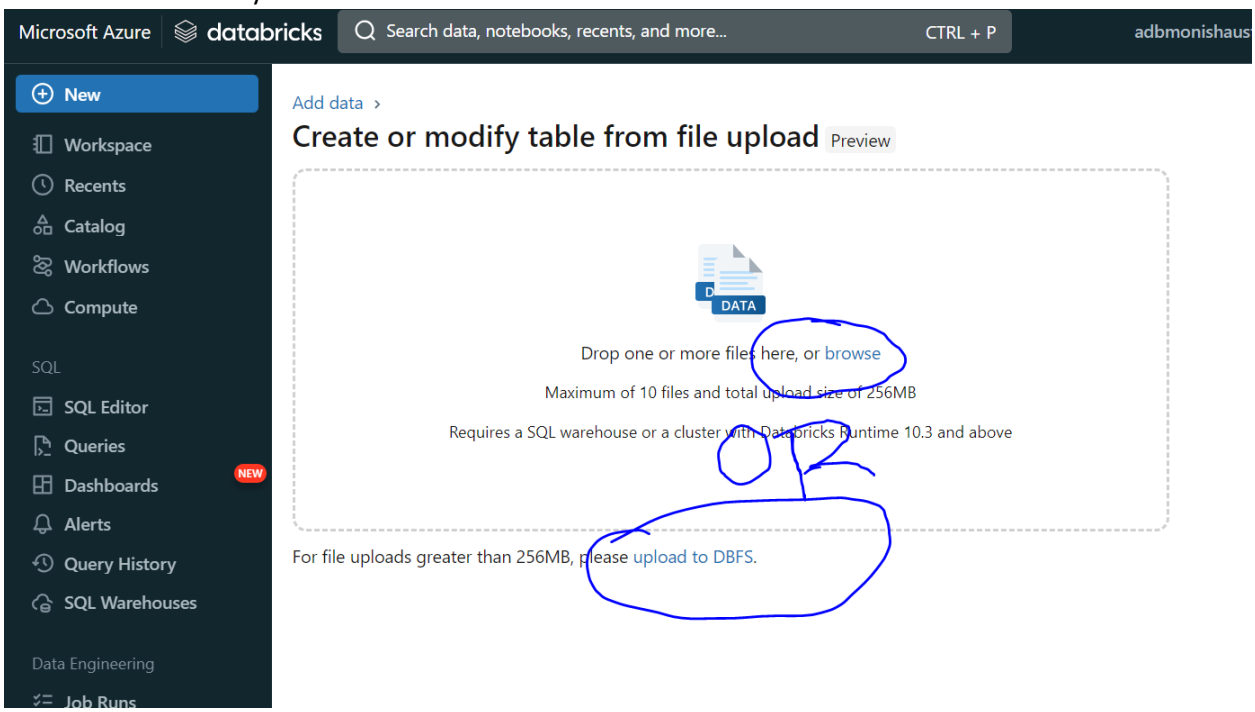


Shift+Ctrl+Enter to run selected text

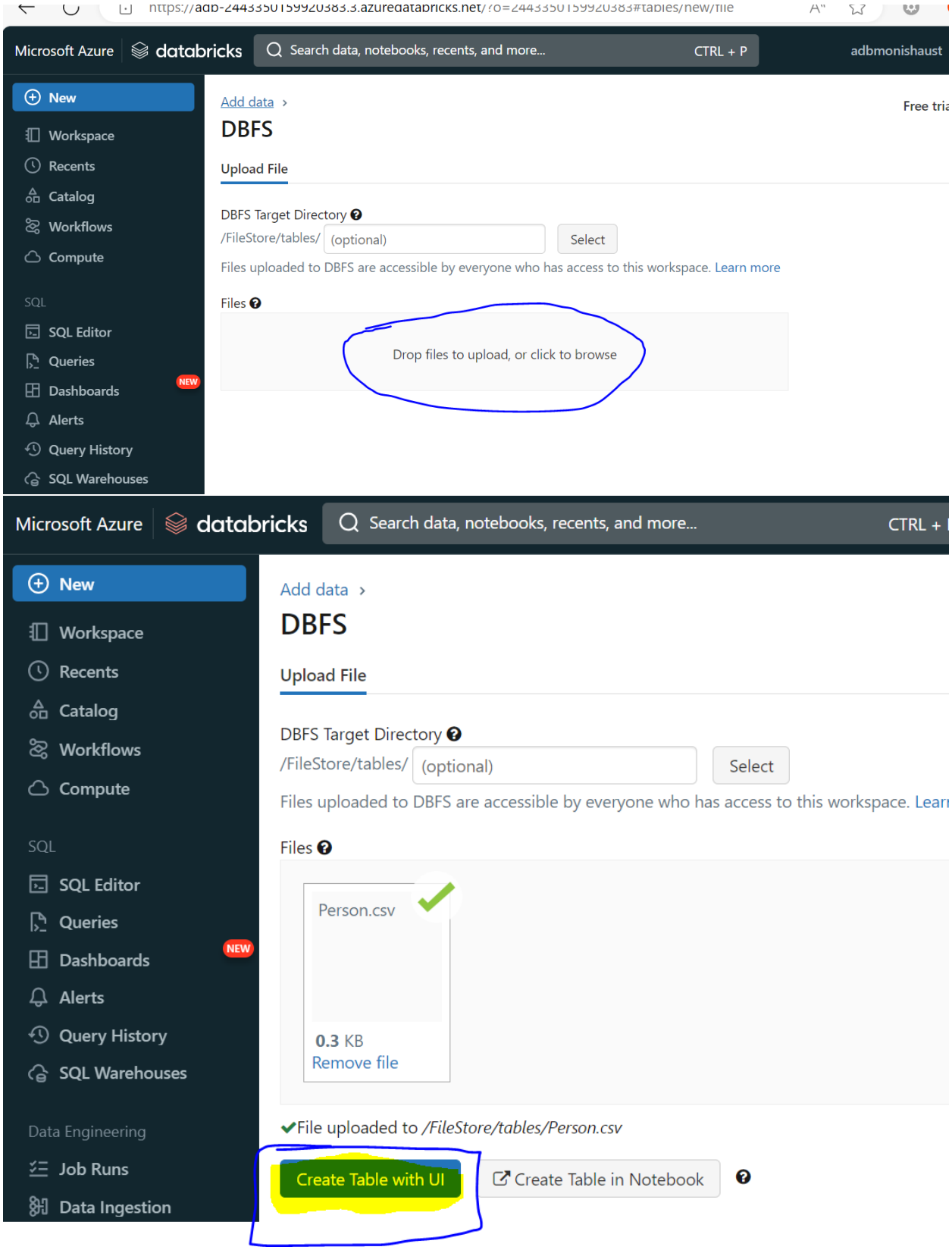
To go to recent notebook:



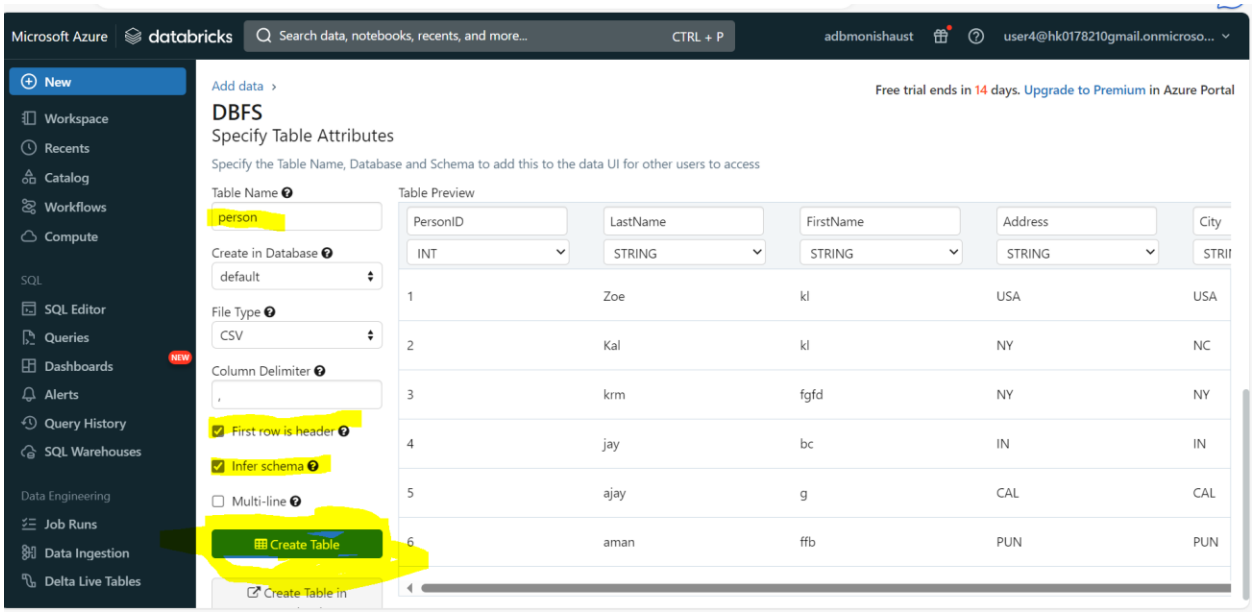
Click create or modify table ➔



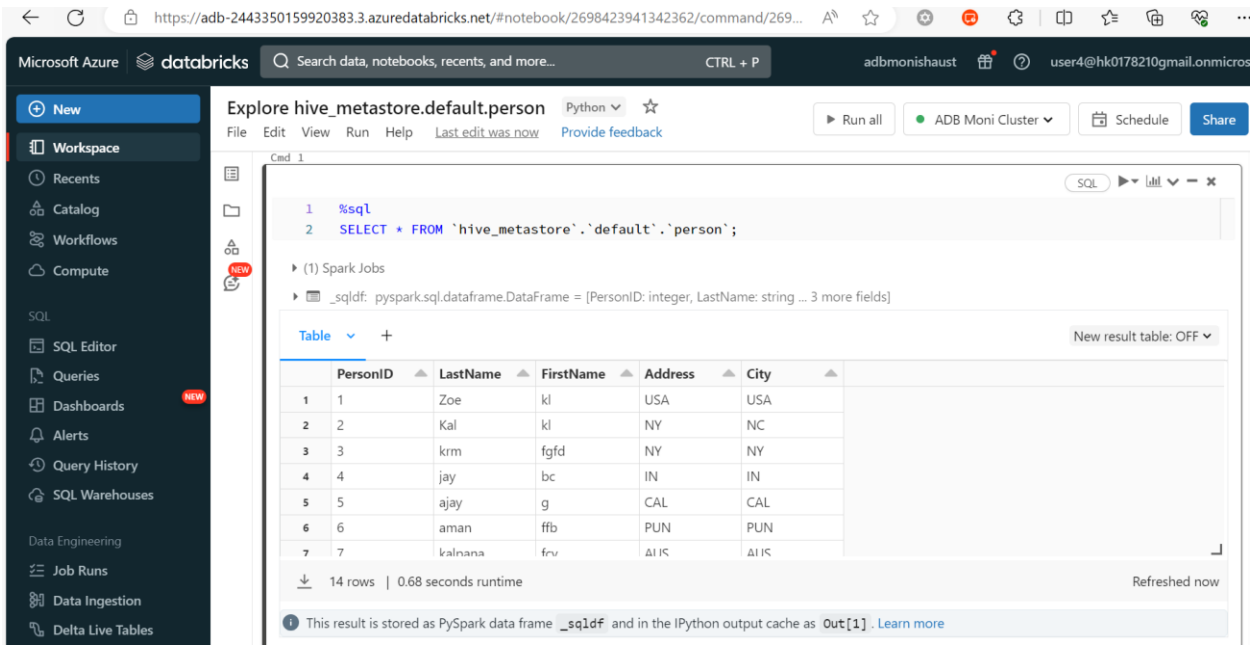
Click upload to dbfs ➔



Cluster : choose cluster(ADB Moni Cluster) ➔ PREVIEW table ➔table name: person, tick first row, tick infer schema



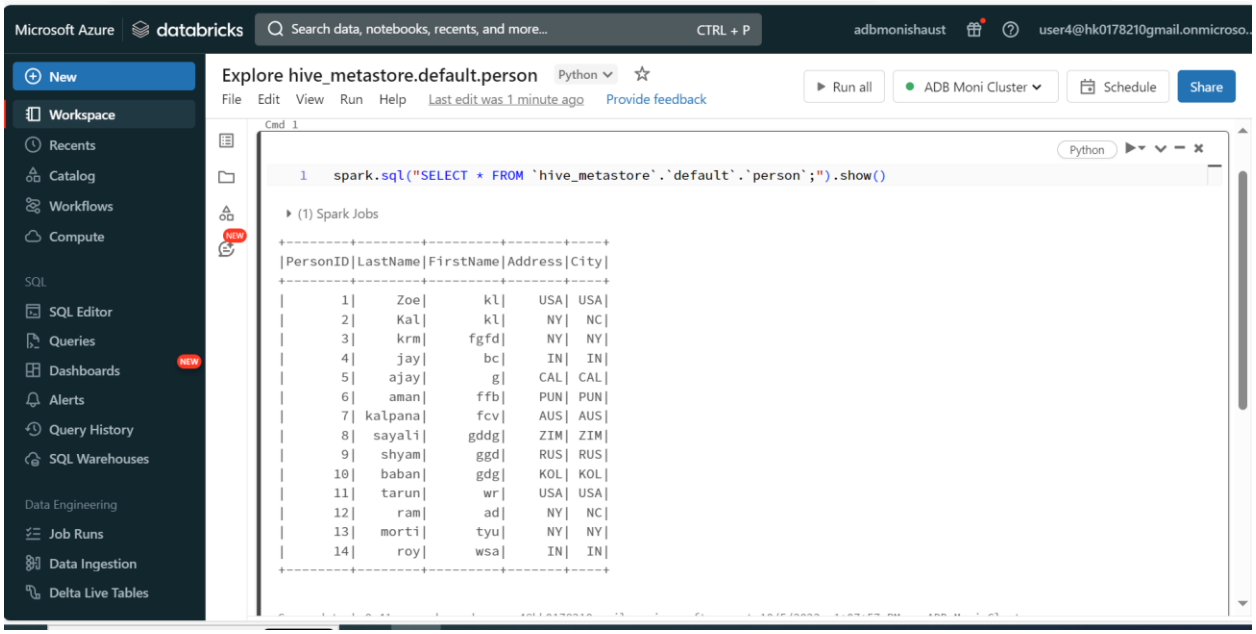

```
%sql
SELECT * FROM `hive_metastore`.`default`.`person`;
```



OR

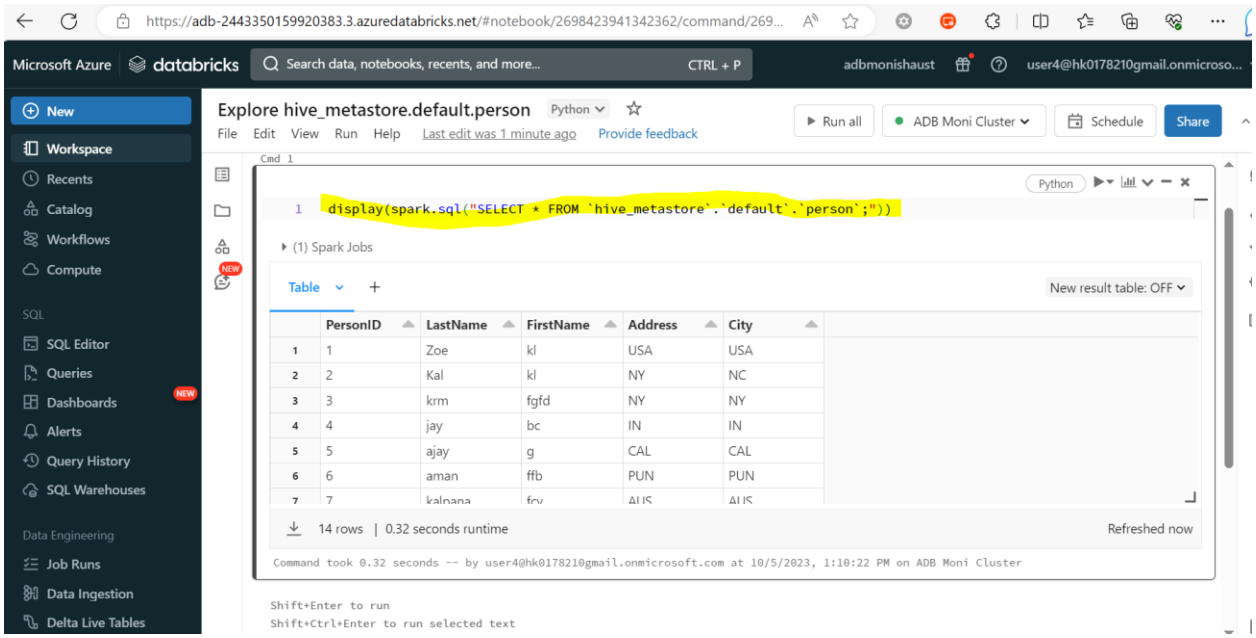
Spark.sql (“query”).Show():

```
spark.sql("SELECT * FROM `hive_metastore`.`default`.`person`;").show()
```



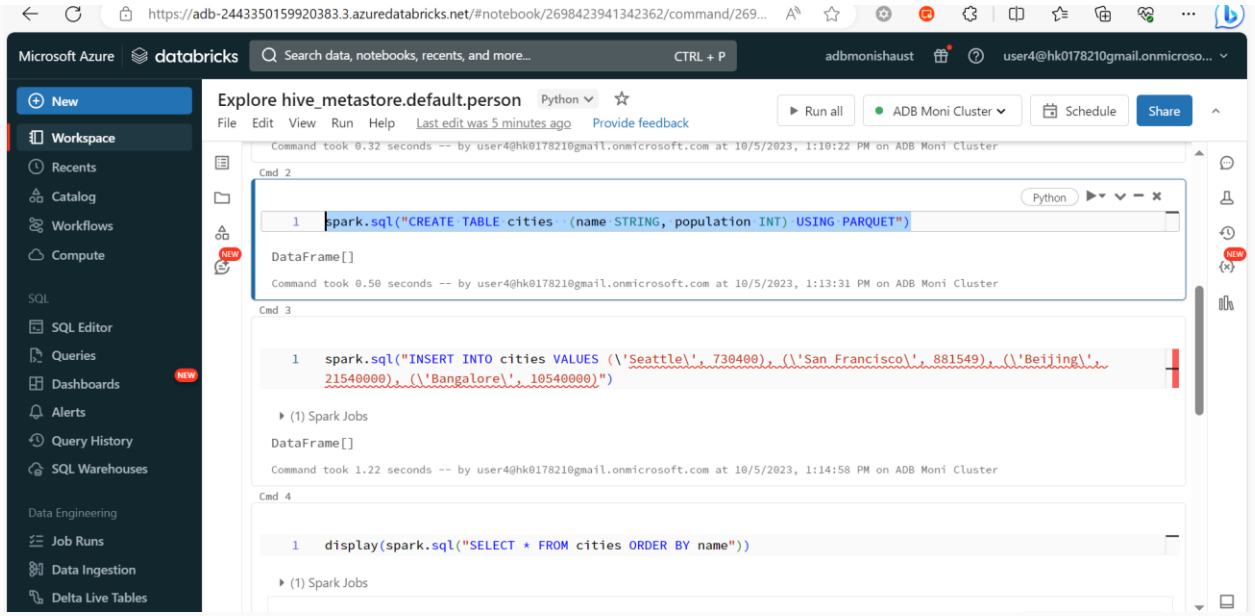
OR

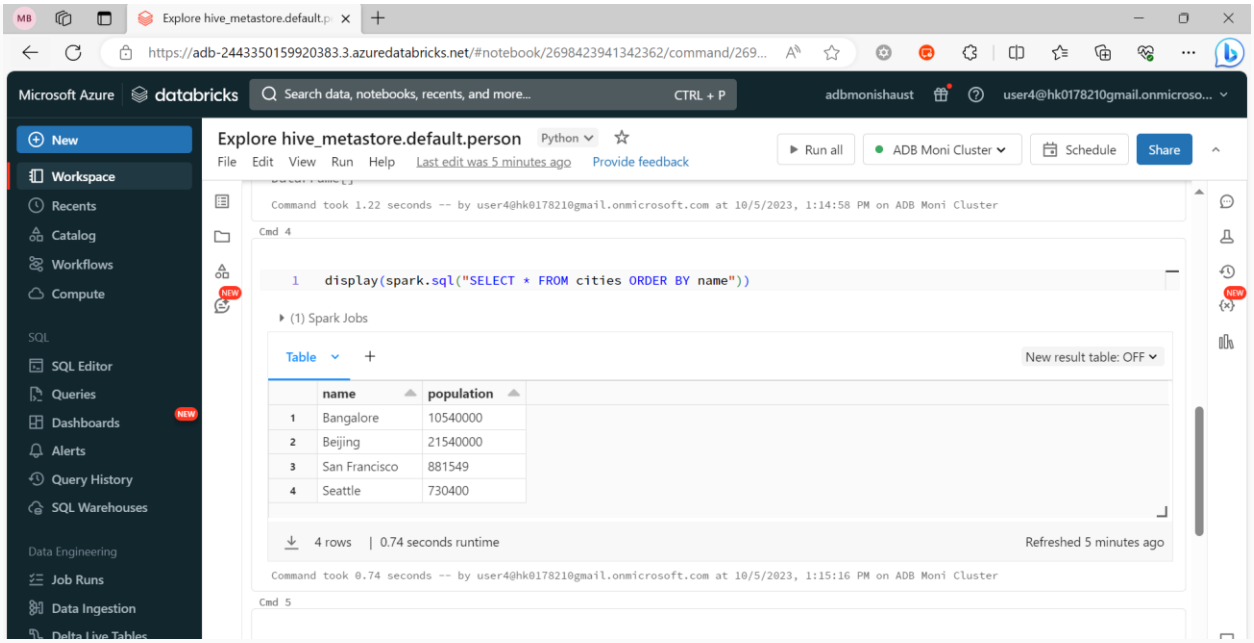
```
display(spark.sql("SELECT * FROM `hive_metastore`.`default`.`person`;"))
```



HOW TO CREATE TABLE WITH NOTEBOOK:

```
1. spark.sql("CREATE TABLE cities  (name STRING, population INT) USING PARQUET")
2. spark.sql("INSERT INTO cities VALUES (\'Seattle\', 730400), (\'San Francisco\', 881549), (\'Beijing\', 21540000), (\'Bangalore\', 10540000)")
3. display(spark.sql("SELECT * FROM cities ORDER BY name"))
```




Explore hive_metastore.default.p
https://adb-2443350159920383.3.azuredatabricks.net/#notebook/2698423941342362/command/269...
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
adbmonishaust
user4@hk0178210gmail.onmicroso...
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Explore hive_metastore.default.person
Python
File Edit View Run Help
Last edit was 5 minutes ago
Provide feedback
Run all
ADB Moni Cluster
Schedule
Share
Command took 1.22 seconds -- by user4@hk0178210gmail.onmicrosoft.com at 10/5/2023, 1:14:58 PM on ADB Moni Cluster
Cmd 4
1 display(spark.sql("SELECT * FROM cities ORDER BY name"))
(1) Spark Jobs
Table
New result table: OFF
name
population
1 Bangalore 10540000
2 Beijing 21540000
3 San Francisco 881549
4 Seattle 730400
4 rows | 0.74 seconds runtime
Refreshed 5 minutes ago
Command took 0.74 seconds -- by user4@hk0178210gmail.onmicrosoft.com at 10/5/2023, 1:15:16 PM on ADB Moni Cluster
Cmd 5

MOUNT STORAGE TO DATA BRICKS:

1. Create data lake storage
2. In cluster notebook ➔ New ➔ Notebook ➔ Name : ADB to ADLS Connectivity
3. In cell, To add markdown:
   %md
   Mount DataLake with DataBricks → bold
4. To configure data lake or connect:

```
spark.conf.set(
      'fs.azure.account.key.adlsmonisha.dfs.core.windows.net'
      ,
'bs+6/GW6Ps+g8kKPGEeITgCZO0eHu1Fe2WAyeizdR9s0090o/TlQCGs90tWD8OfoRMi1YedWKkQl+AStavXOMg=='
)
```

   Syntax : spark.conf.set(‘fs.azure.account.key(or)SAS.endpoint of storage’,’account key or SAS’)
   Fs ➔ file system
   Authentication method ➔ account key or SAS

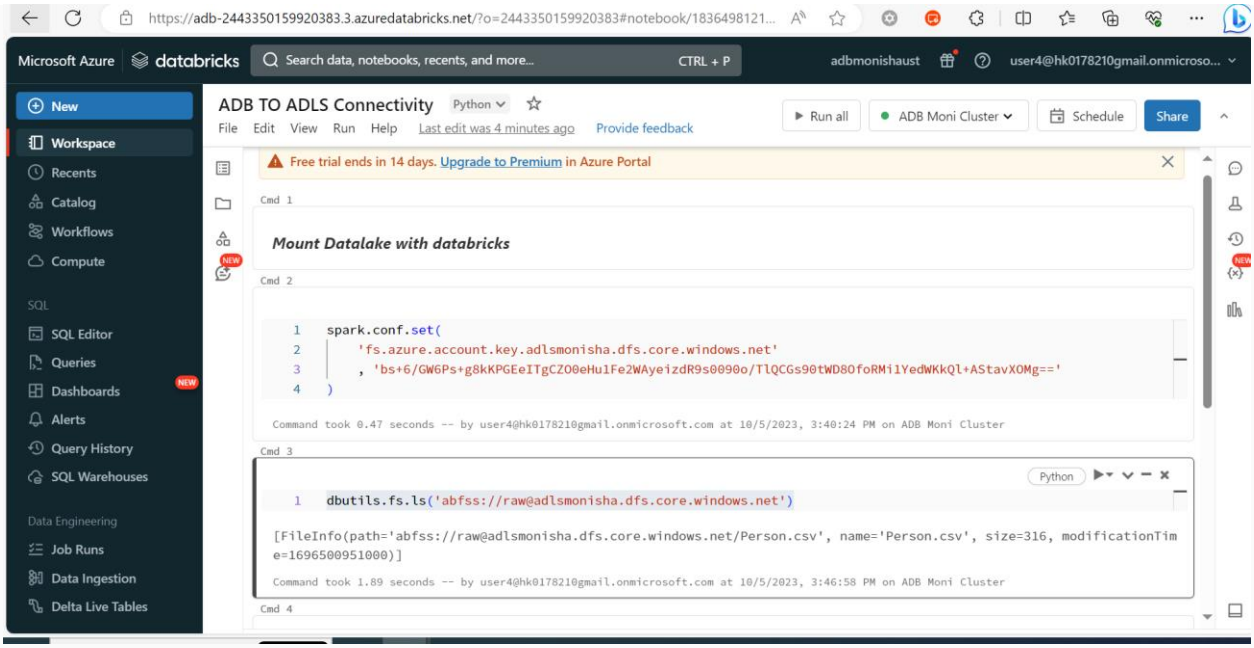
5. To copy endpoint : Go to storage account ➔ end points ➔ DataLake Storage url without https:// ➔ copy
6. To copy account key: Go to storage account ➔ Access key ➔ key ➔ click show ➔ copy key
7. Shift enter and run the cell
8. To list the files in the storage account :

   ```
   dbutils.fs.ls('abfss://raw@adlsmonisha.dfs.core.windows.net')
   ```

   syntax:

   ```
   dbutils.fs.ls('abfss://containername@storagename.endpoints/folder/subfolder/file.exe’)
   ```

   abfss ➔ azure blob file system



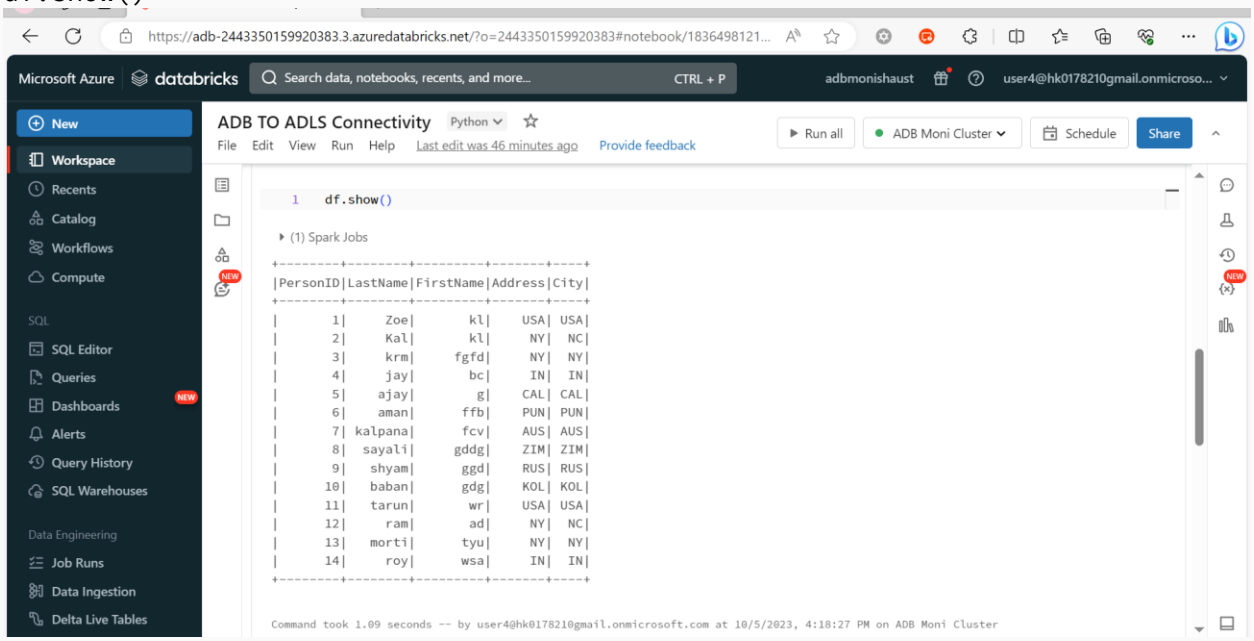
HOW TO READ THE FILE:

df ➔ data frame
read ➔ to read data
inferSchema ➔ by default csv data will be string but if u correct data type. inferSchema will auto deduct data type
header ➔ header of table is true
delimiter ➔ it can be , : . this file is comma separated.

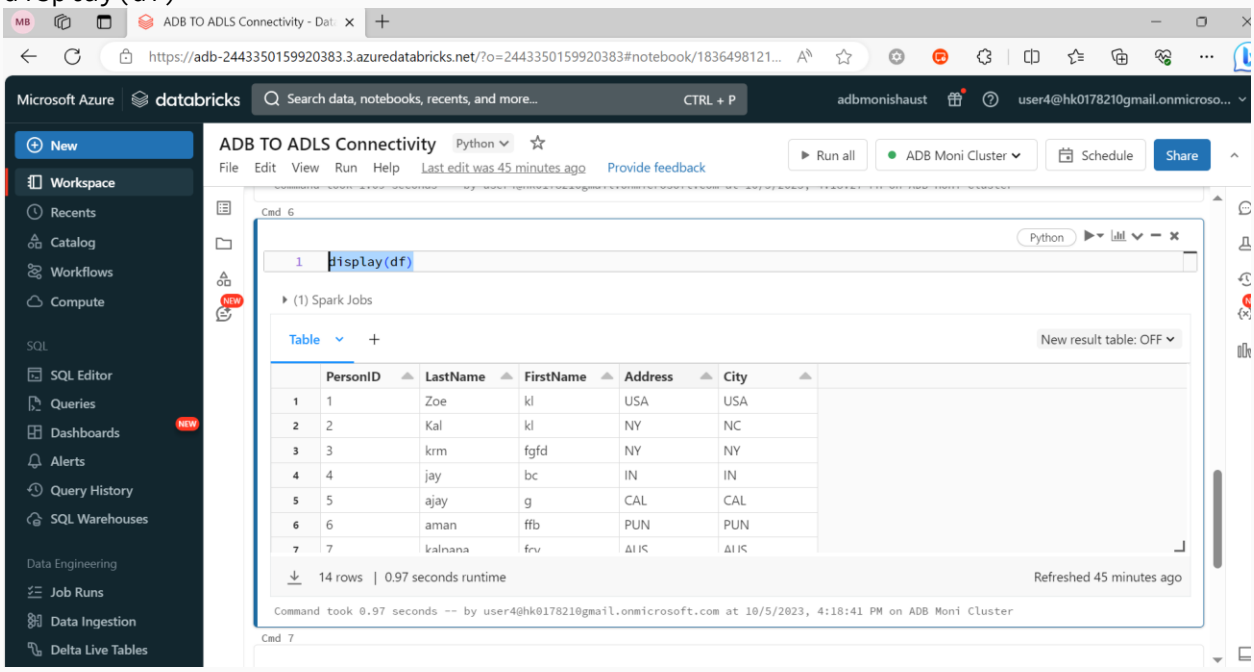
```
# Read File from ADLS
source_file = 'abfss://raw@adlsmonisha.dfs.core.windows.net/Person.csv'

df = spark.read\
    .format('csv')\
    .option('inferSchema', True)\
    .option('header', True)\
    .option('delimiter', ',')\
    .load(source_file)
```

To view data:

df.show()



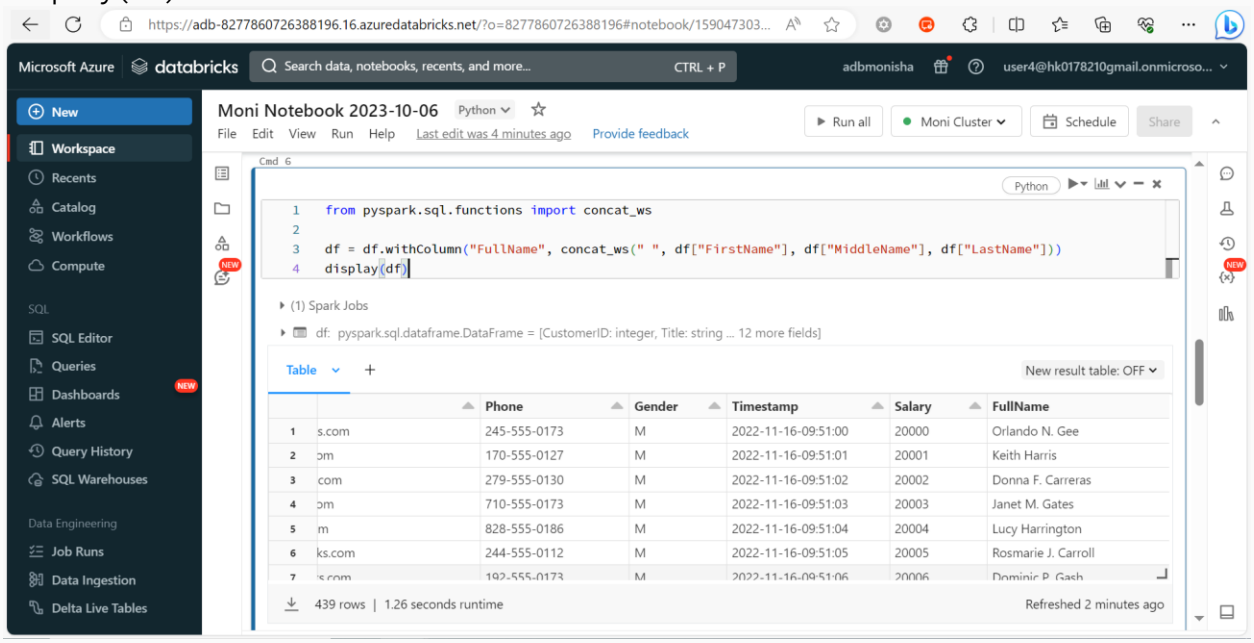
display(df)



ASSIGNMENT:

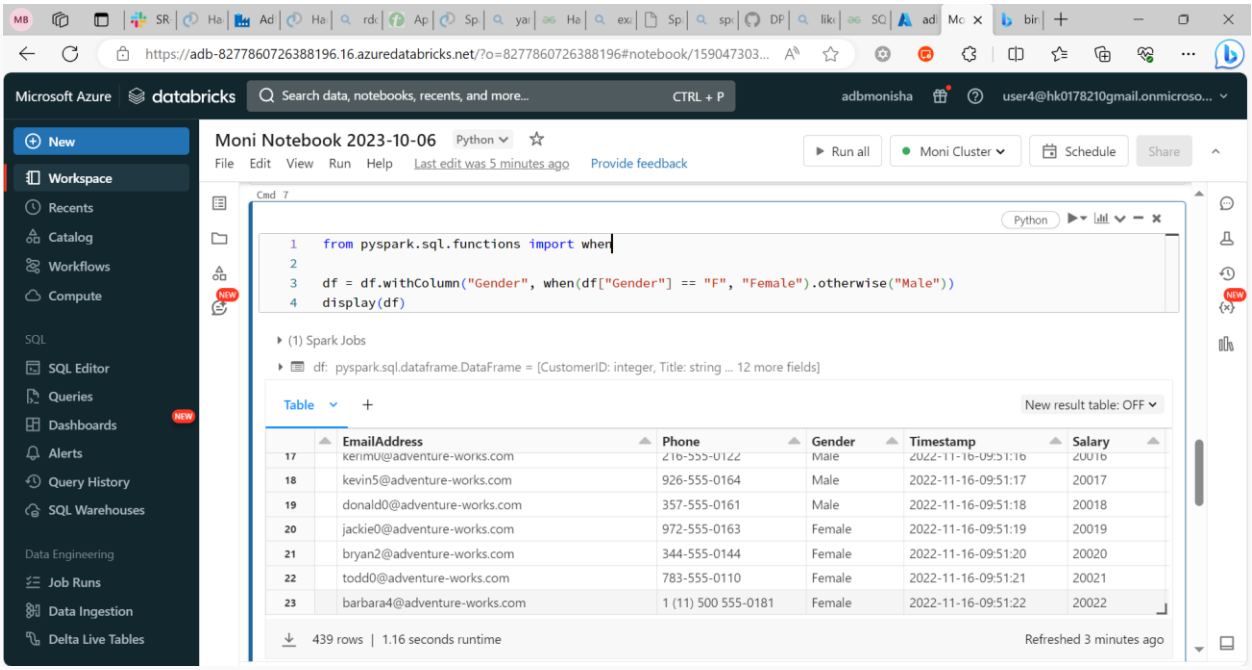
1)

```
from pyspark.sql.functions import concat_ws

df = df.withColumn("FullName", concat_ws(" ", df["FirstName"],
df["MiddleName"], df["LastName"]))
display(df)
```



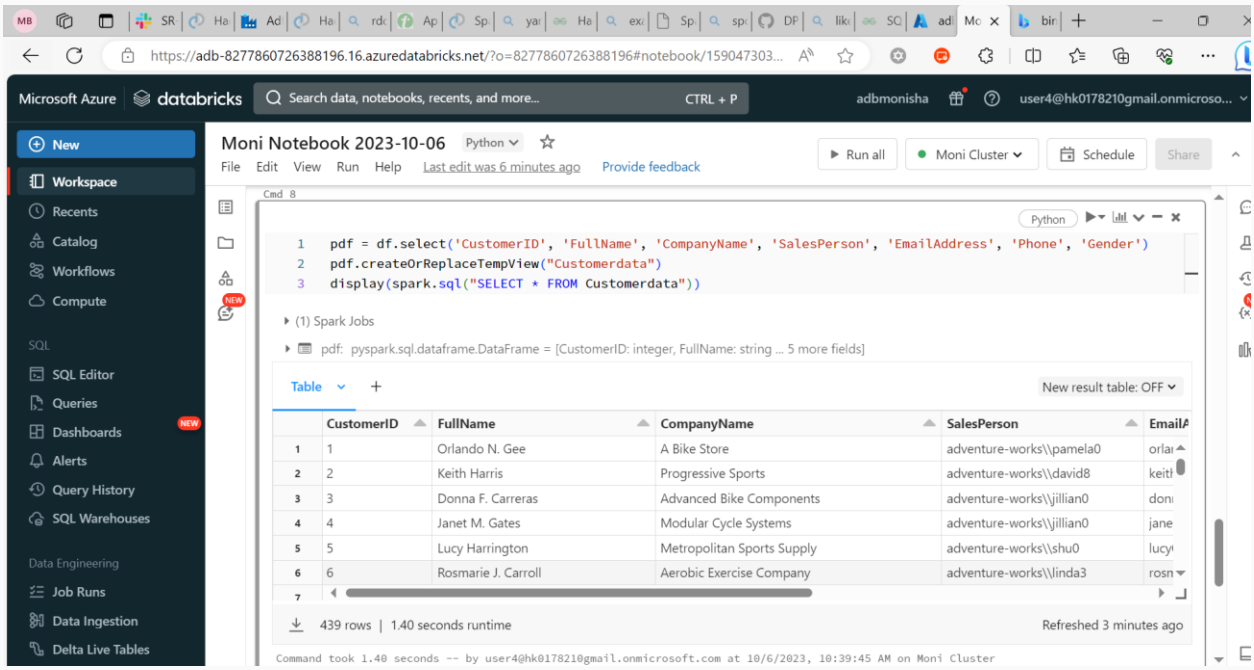
2)

```
from pyspark.sql.functions import when

df = df.withColumn("Gender", when(df["Gender"] == "F",
"Female").otherwise("Male"))
display(df)
```



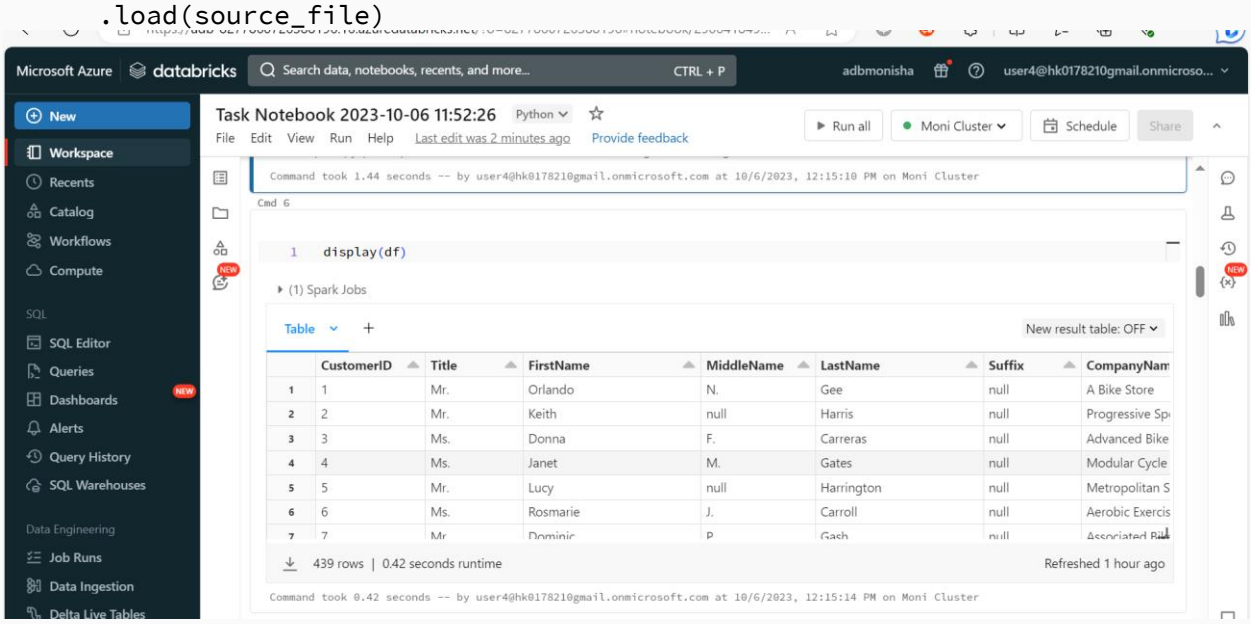
3)

```
pdf = df.select('CustomerID', 'FullName', 'CompanyName', 'SalesPerson', 'EmailAddress', 'Phone', 'Gender')
pdf.createOrReplaceTempView("Customerdata")
display(spark.sql("SELECT * FROM Customerdata"))
```



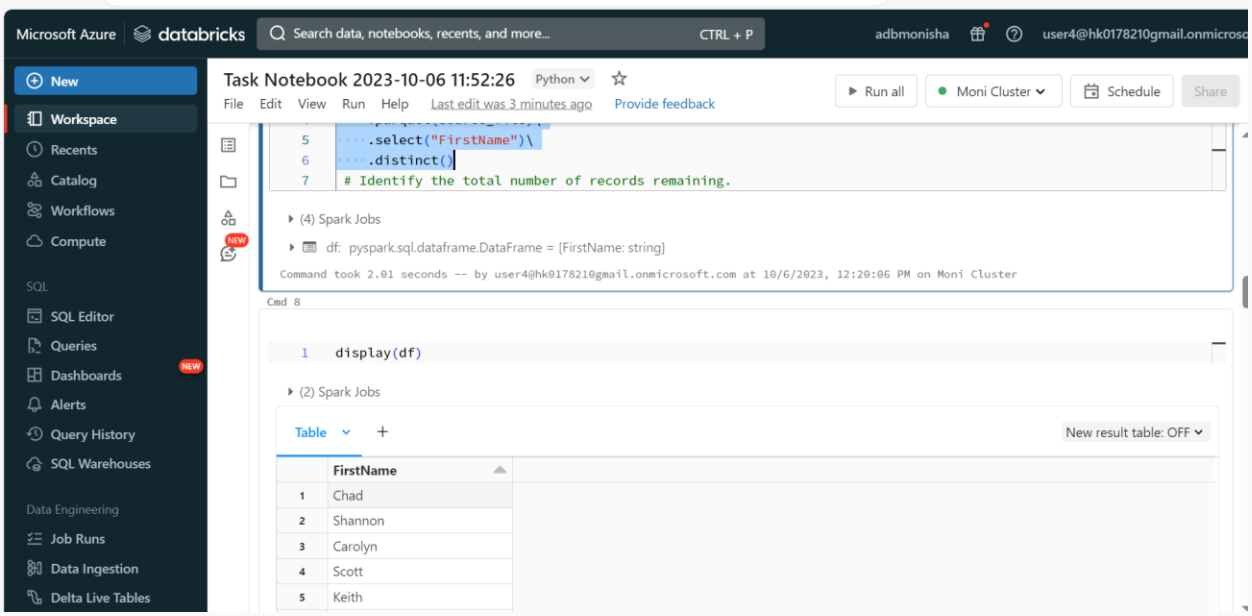
Task:
1)

```
# TODO
# Replace <<FILL_IN>> with your code.
source_file =
'abfss://raw@adlsmonisha.dfs.core.windows.net/Source/customerdemo.parquet'

pdf = spark.read\
    .parquet(source_file)
    .load(source_file)
```



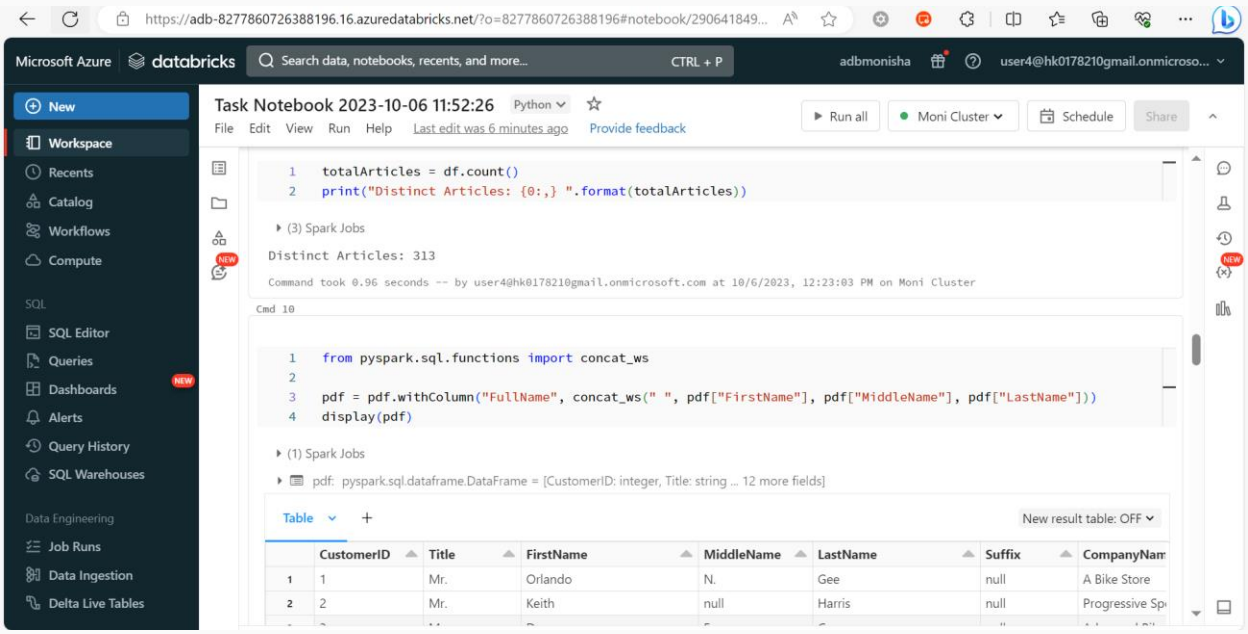
2)

```
source_file =
'abfss://raw@adlsmonisha.dfs.core.windows.net/Source/customerdemo.parquet'

df = spark.read\
    .parquet(source_file)\
    .select("FirstName")\
    .distinct()
```



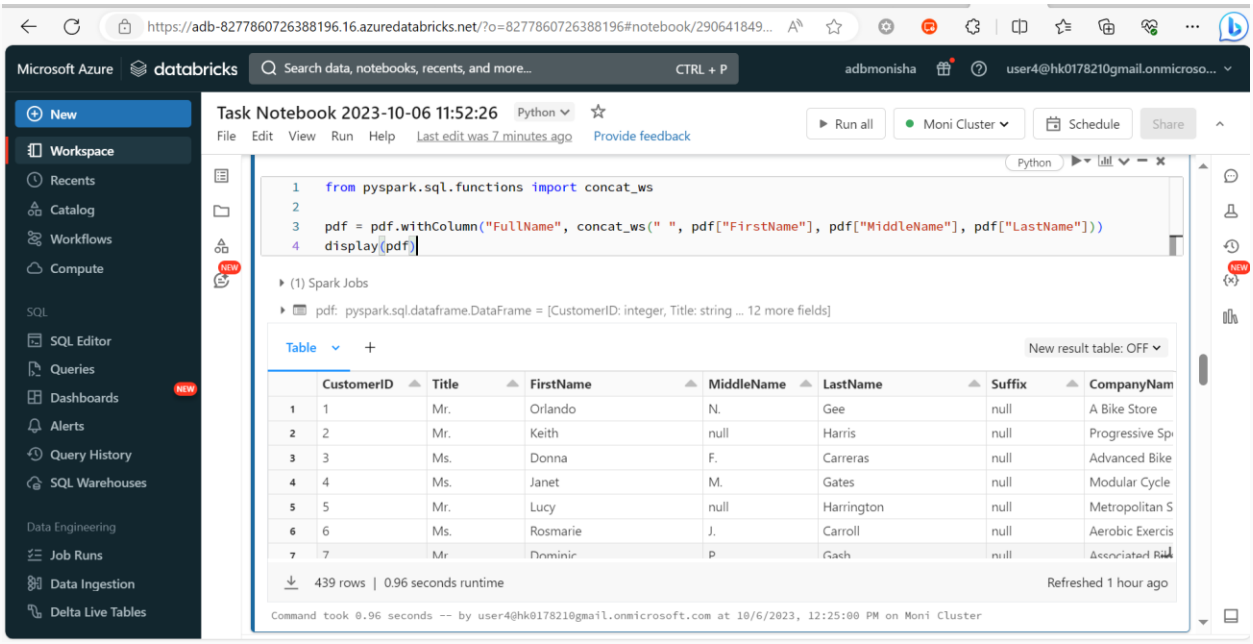
3)

```
totalArticles = df.count()
print("Distinct Articles: {0:,} ".format(totalArticles))
```



4)

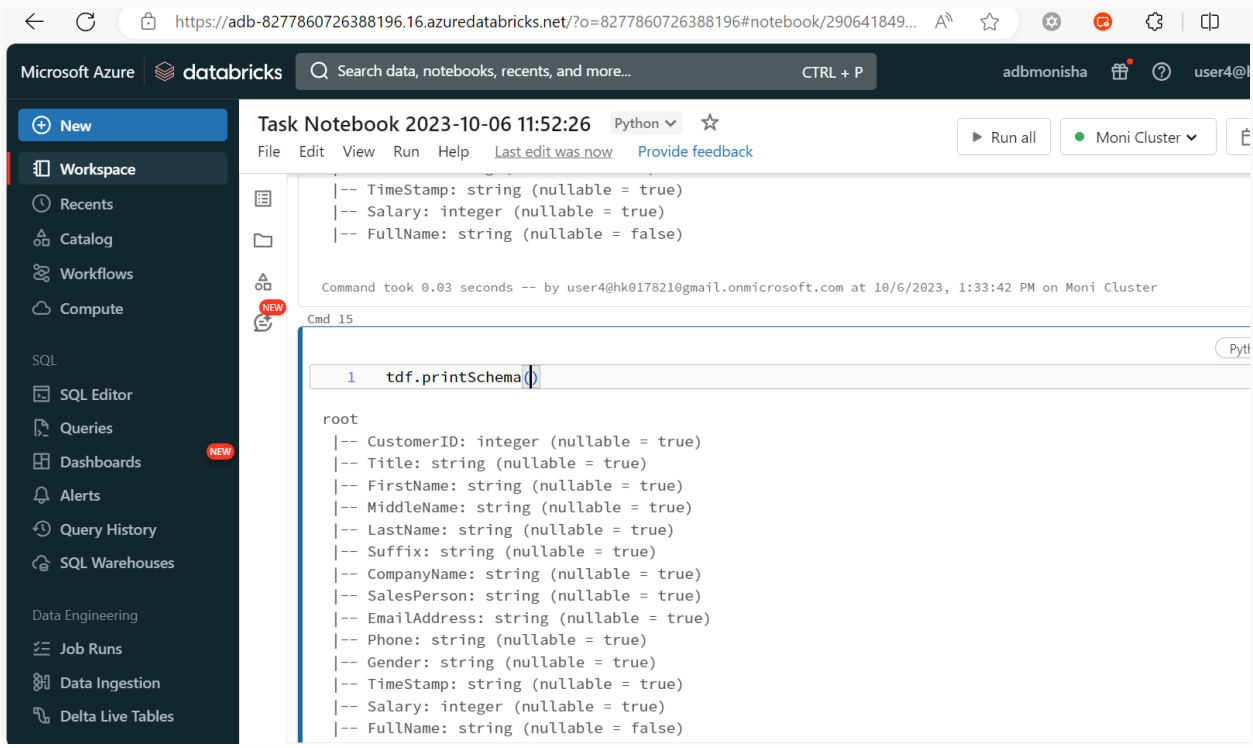
```
from pyspark.sql.functions import concat_ws

pdf = pdf.withColumn("FullName", concat_ws(" ", pdf["FirstName"],
pdf["MiddleName"], pdf["LastName"]))
display(pdf)
```



5)

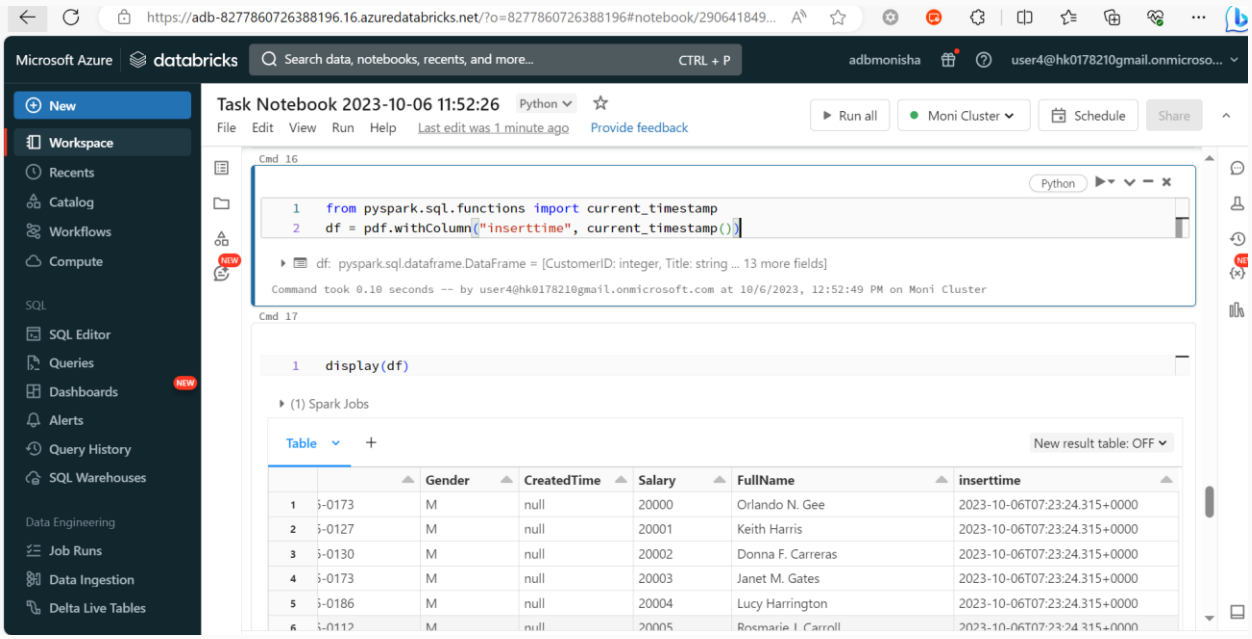
To rename column and change data type:

```
from pyspark.sql.functions import *
from pyspark.sql.types import StringType
tdf = pdf.withColumnRenamed('TimeStamp',
'CreatedTime').withColumn('TimeStamp', unix_timestamp(col('TimeStamp'), "yyyy-MM-dd" 't' "HH:mm:ss").cast(StringType()))
```



6)

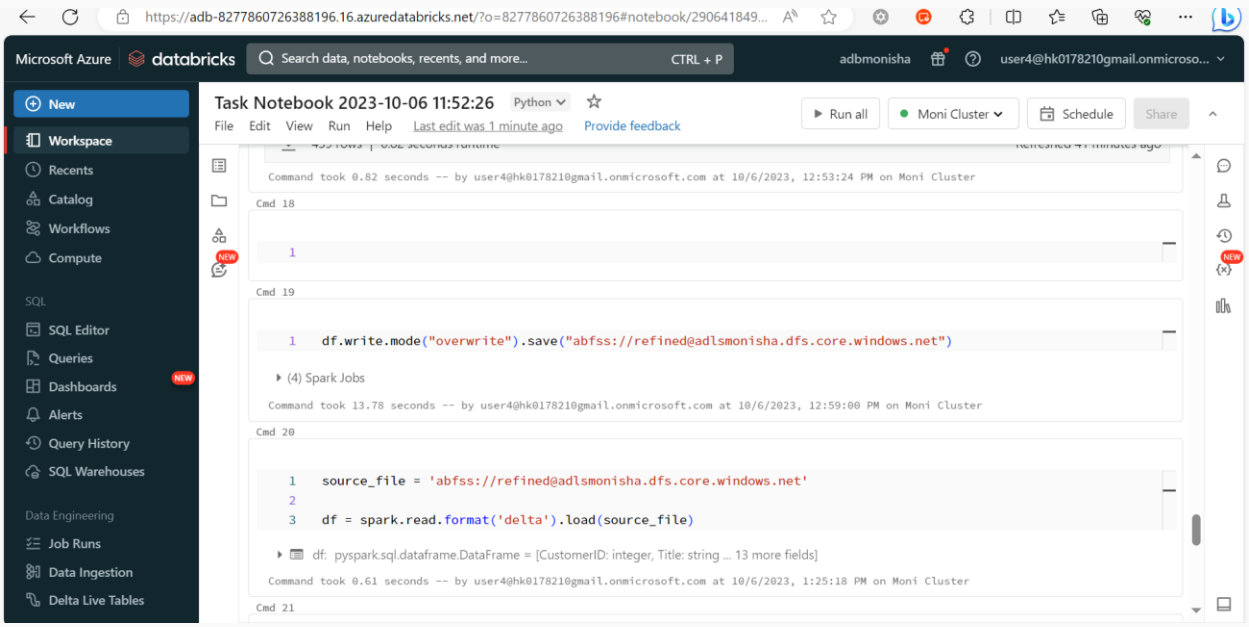
```
from pyspark.sql.functions import current_timestamp
df = pdf.withColumn("inserttime", current_timestamp())
```



7)

To write the file and store it in datalake another container:

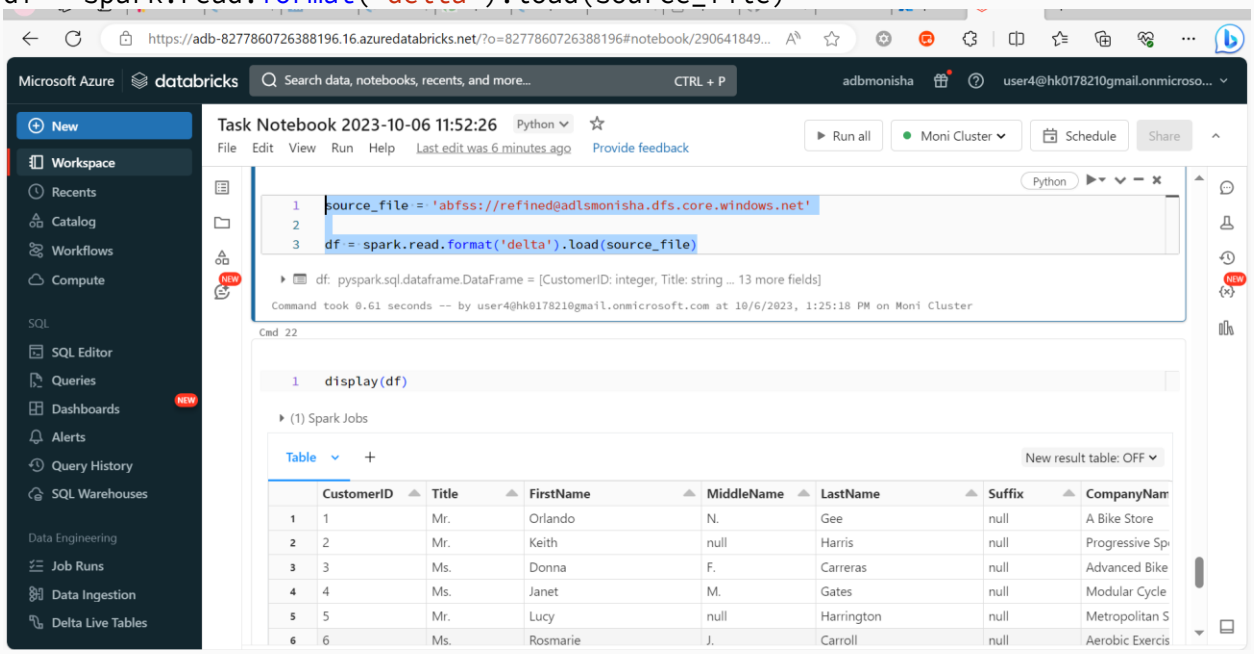
```
df.write.mode("overwrite").save("abfss://refined@adlsmonisha.dfs.core.windows.net")
```



8)

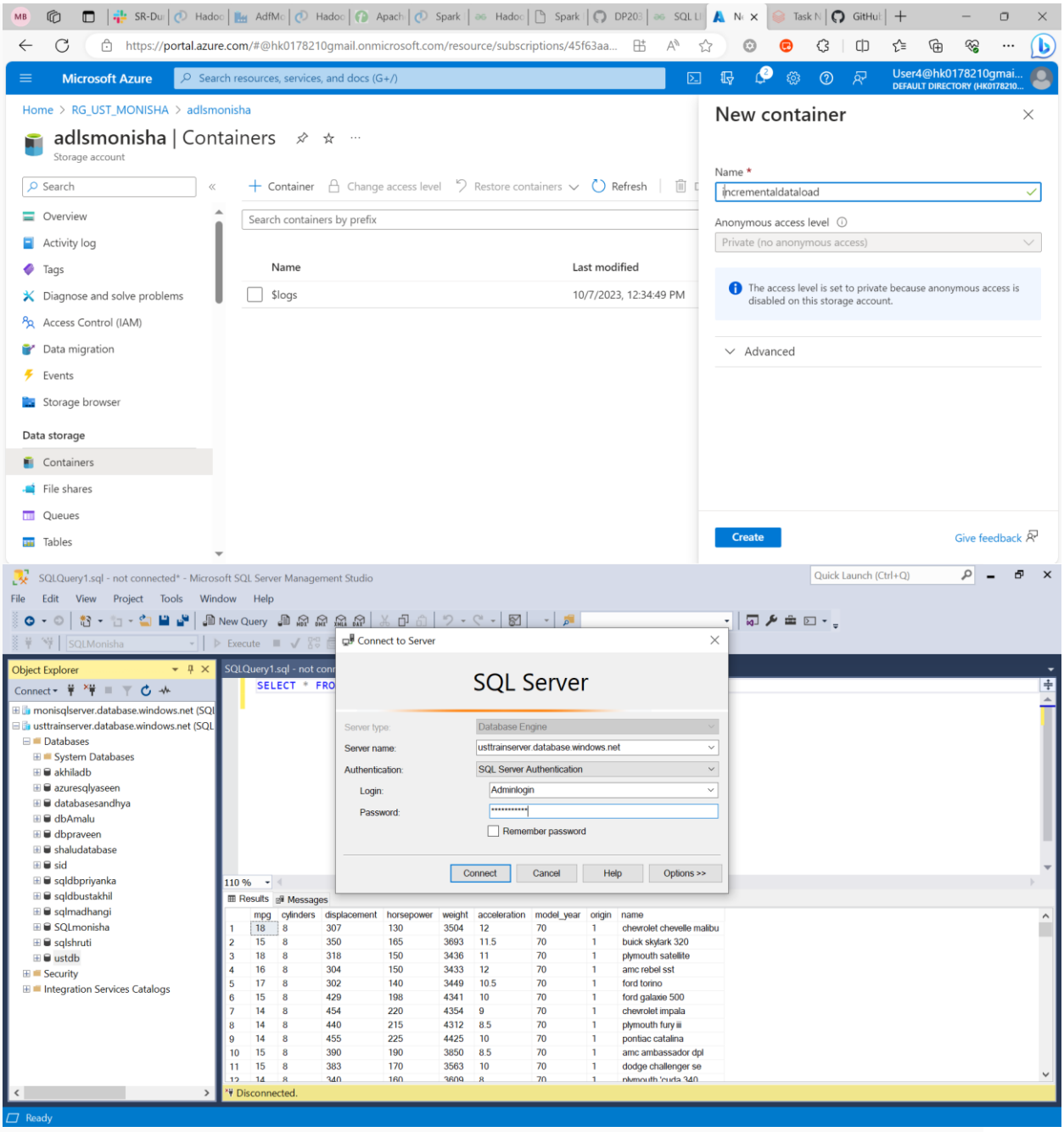
To read the delta file:

```
source_file = 'abfss://refined@adlsmonisha.dfs.core.windows.net'

df = spark.read.format('delta').load(source_file)
```



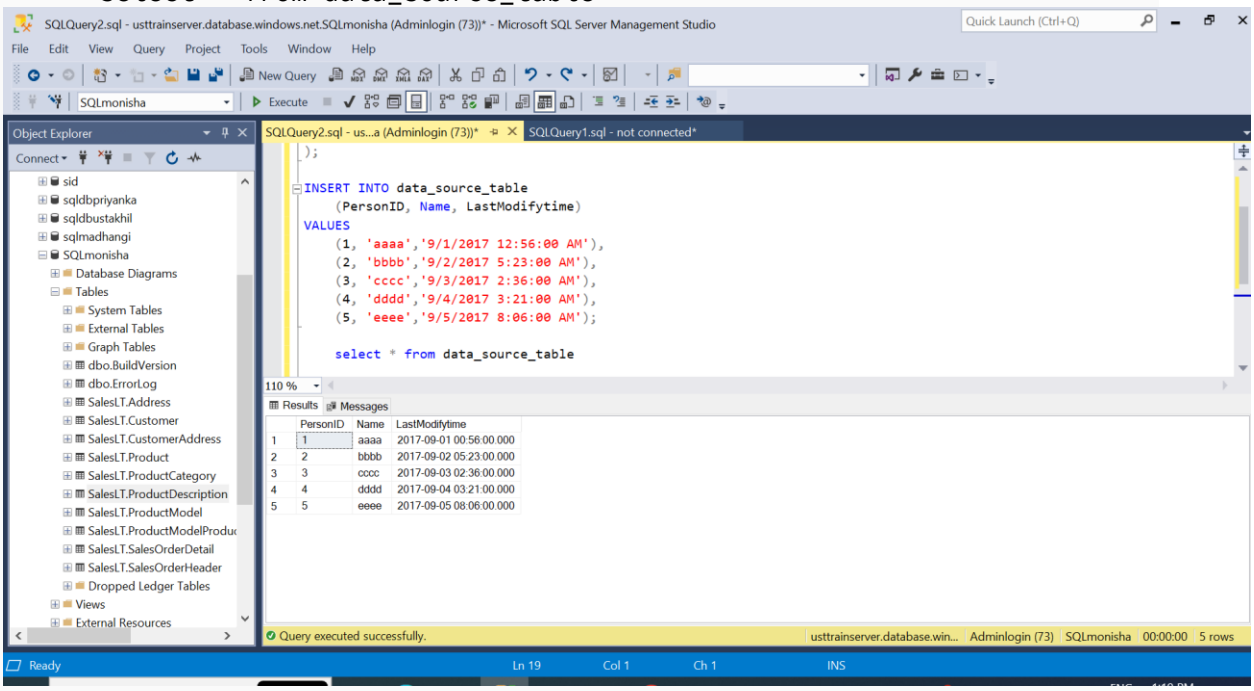
7/10/2023

1) Create data lake storage(adlsmonisha)
2) Create sql database use server existing(SQLMonisha)
3) Data factory(adfmonisha)
4) Create a container named IncrementalDataLoad in DataLake Storage

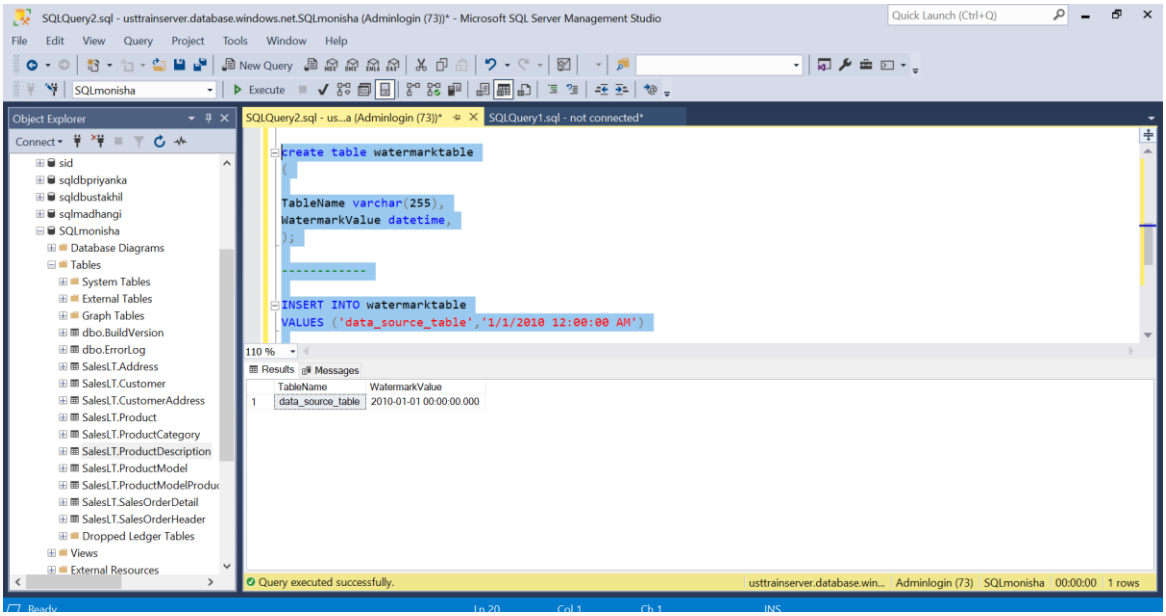



```
create table data_source_table
(
    PersonID int,
    Name varchar(255),
    LastModifytime datetime
);

INSERT INTO data_source_table
    (PersonID, Name, LastModifytime)
VALUES
```

```
    (1, 'aaaa','9/1/2017 12:56:00 AM'),
    (2, 'bbbb','9/2/2017 5:23:00 AM'),
    (3, 'cccc','9/3/2017 2:36:00 AM'),
    (4, 'dddd','9/4/2017 3:21:00 AM'),
    (5, 'eeee','9/5/2017 8:06:00 AM');

      select * from data_source_table
```




```
5)
6) create table watermarktable
7) (
8)
9) TableName varchar(255),
10)WatermarkValue datetime,
11));
12)
13)------------
14)
15)INSERT INTO watermarktable
16)VALUES ('data_source_table','1/1/2010 12:00:00 AM')
17)
18)---------------
19)
20)Select * from watermarktable
21)
```

22) 

23)CREATE PROCEDURE usp_write_watermark @LastModifiedtime datetime,
@TableName varchar(50)

24)AS

25)

26)BEGIN

27)

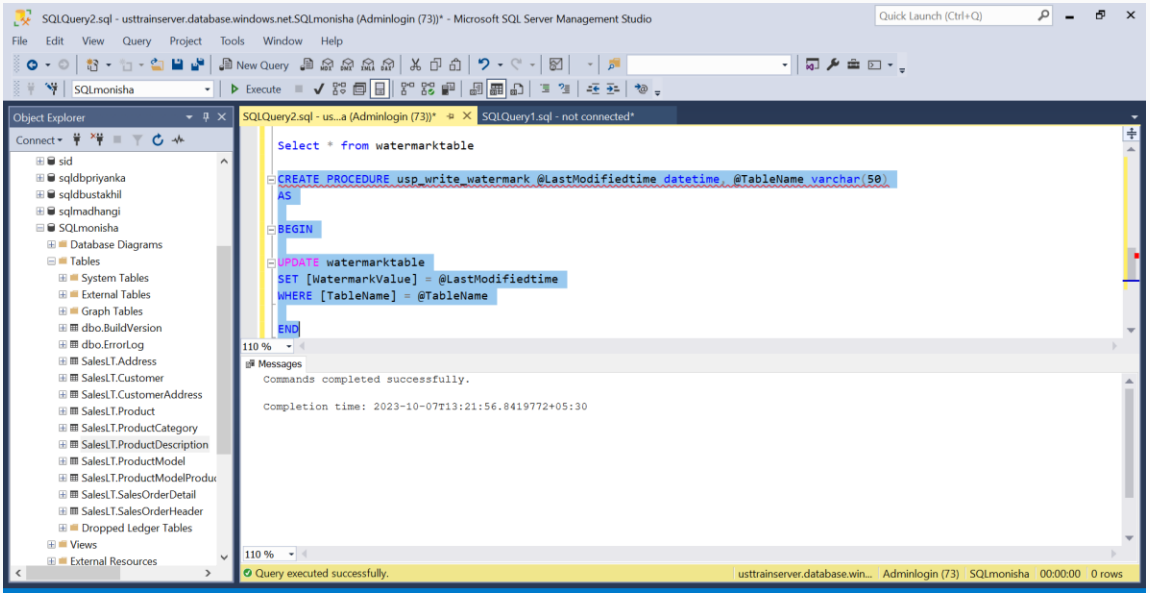
28)UPDATE watermarktable

29)SET [WatermarkValue] = @LastModifiedtime

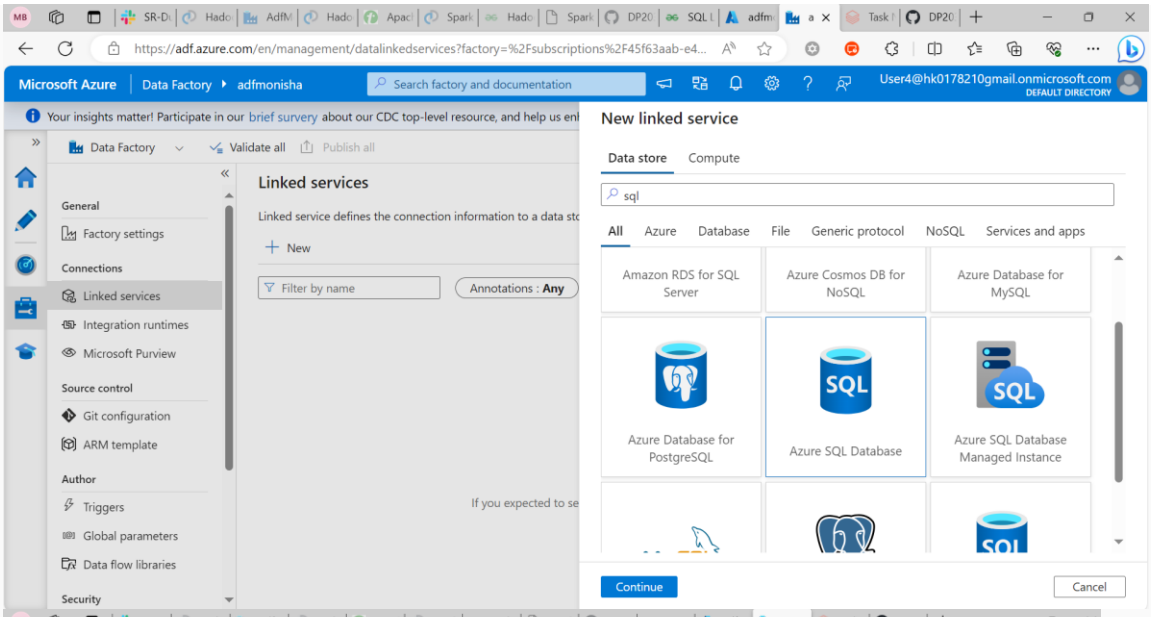
30)WHERE [TableName] = @TableName

31)

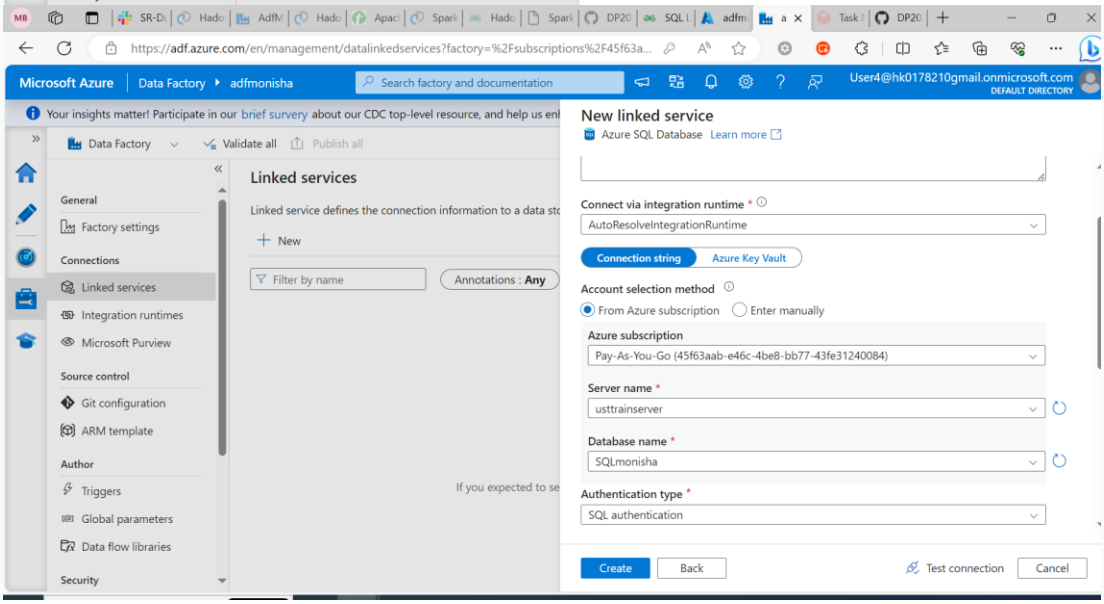
32)END

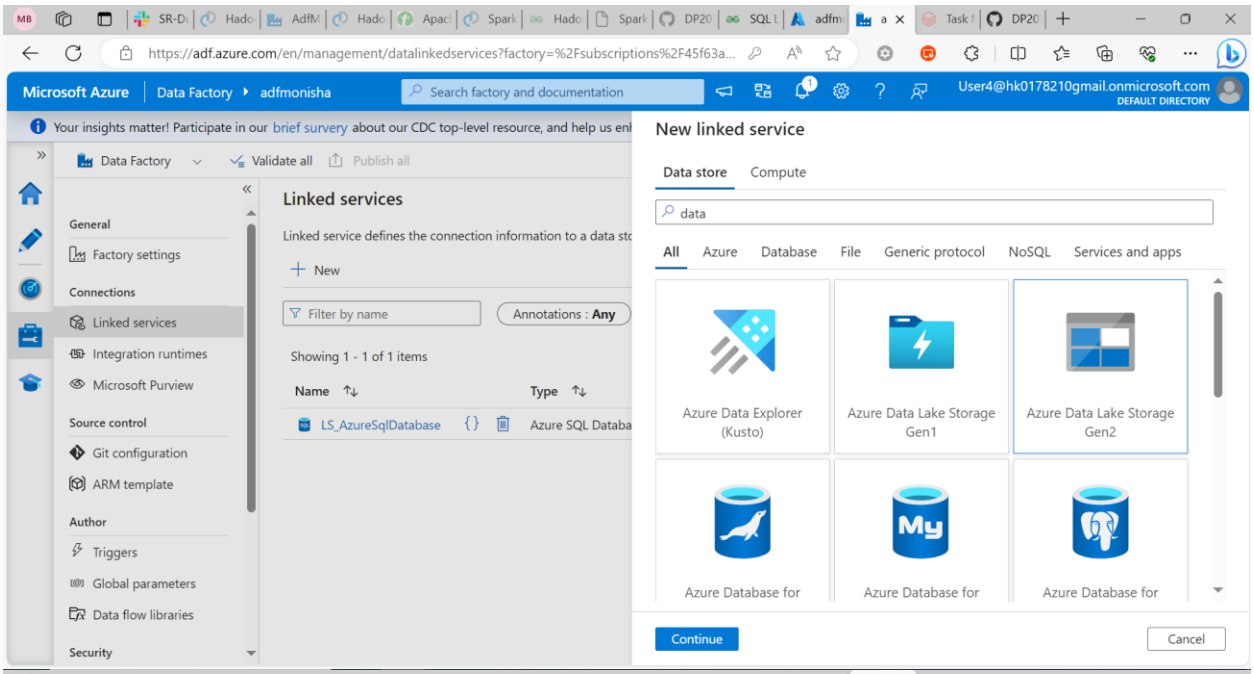
33) 

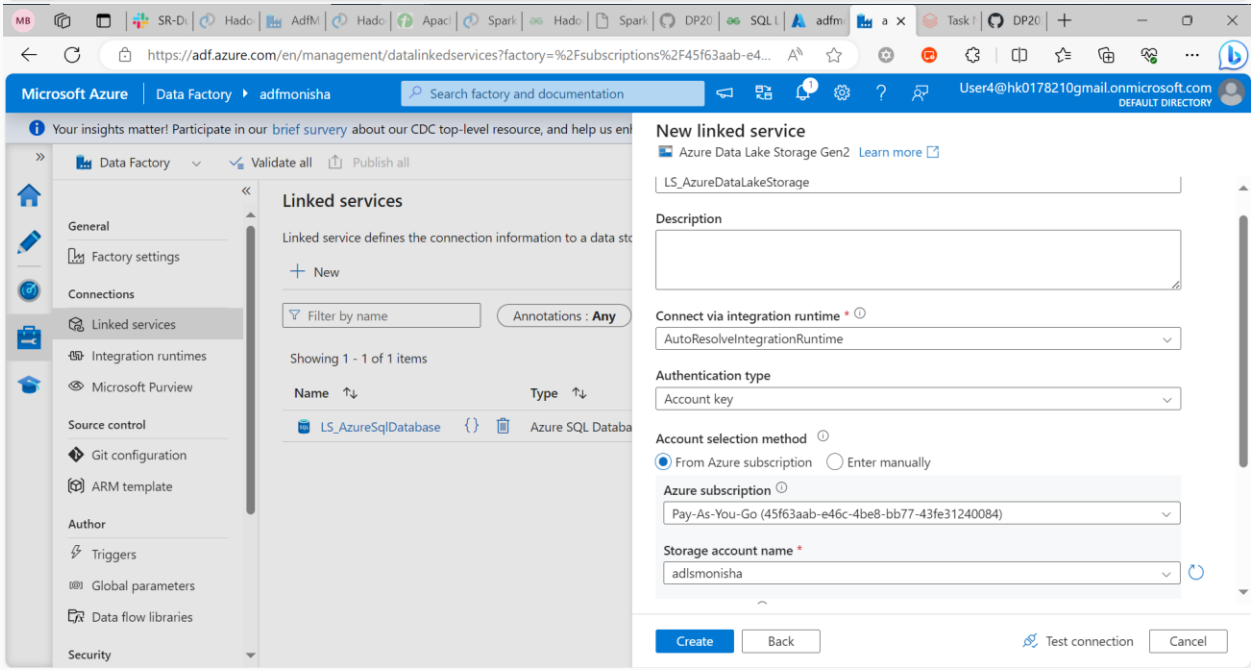
34)

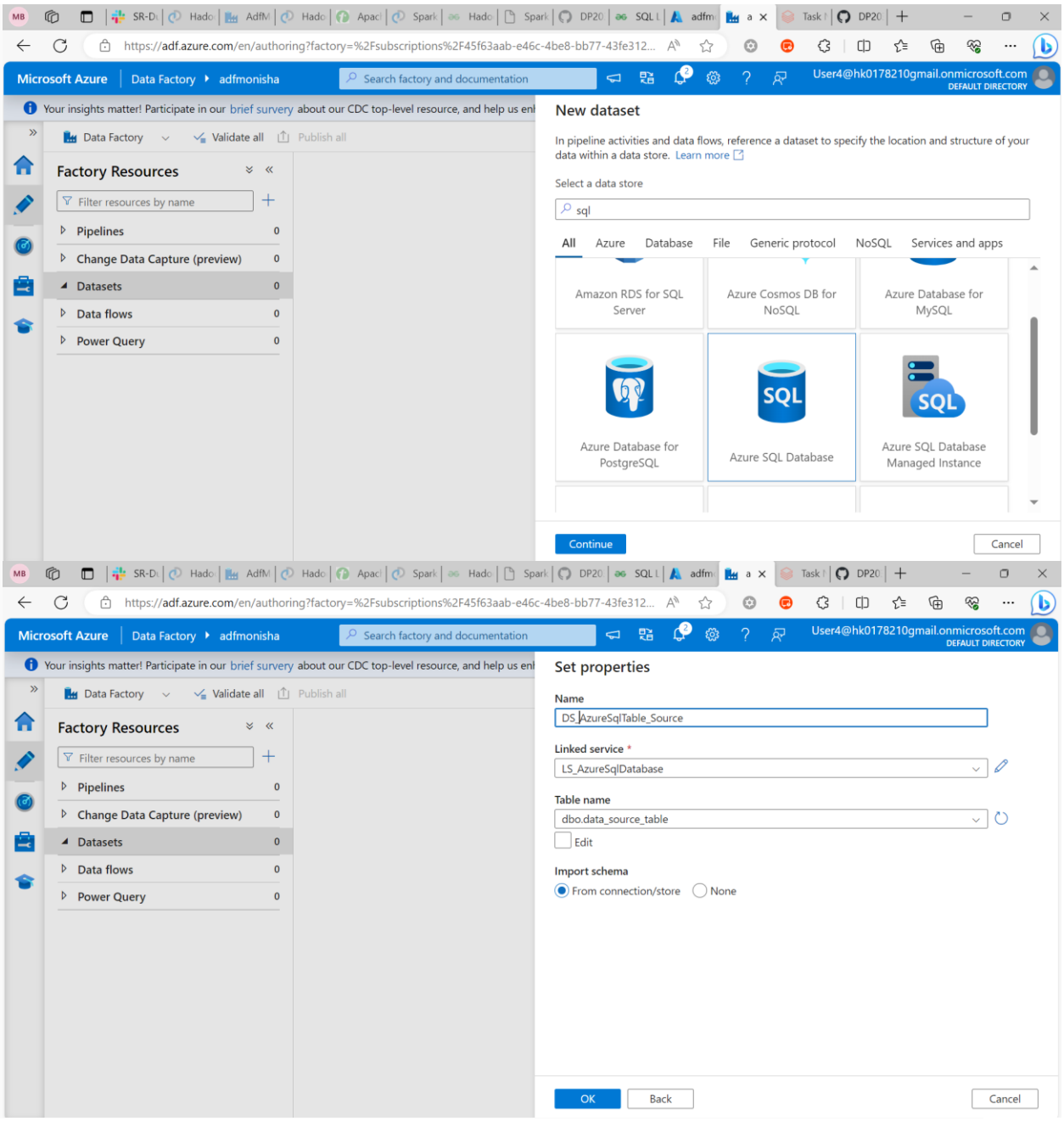

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enl
Data Factory
Validate all
Publish all
Linked services
Linked service defines the connection information to a data sto
New
Filter by name
Annotations : Any
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
If you expected to se
New linked service
Data store
Compute
sql
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Amazon RDS for SQL Server
Azure Cosmos DB for NoSQL
Azure Database for MySQL
Azure Database for PostgreSQL
Azure SQL Database
Azure SQL Database Managed Instance
Continue
Cancel

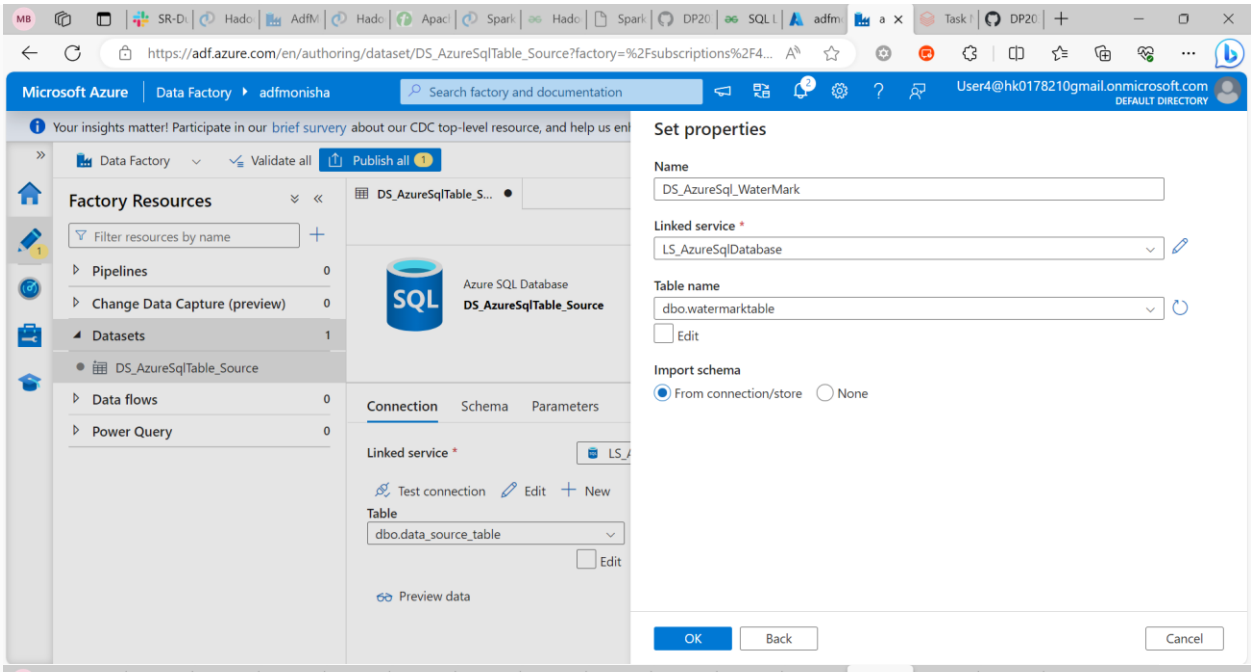
35)

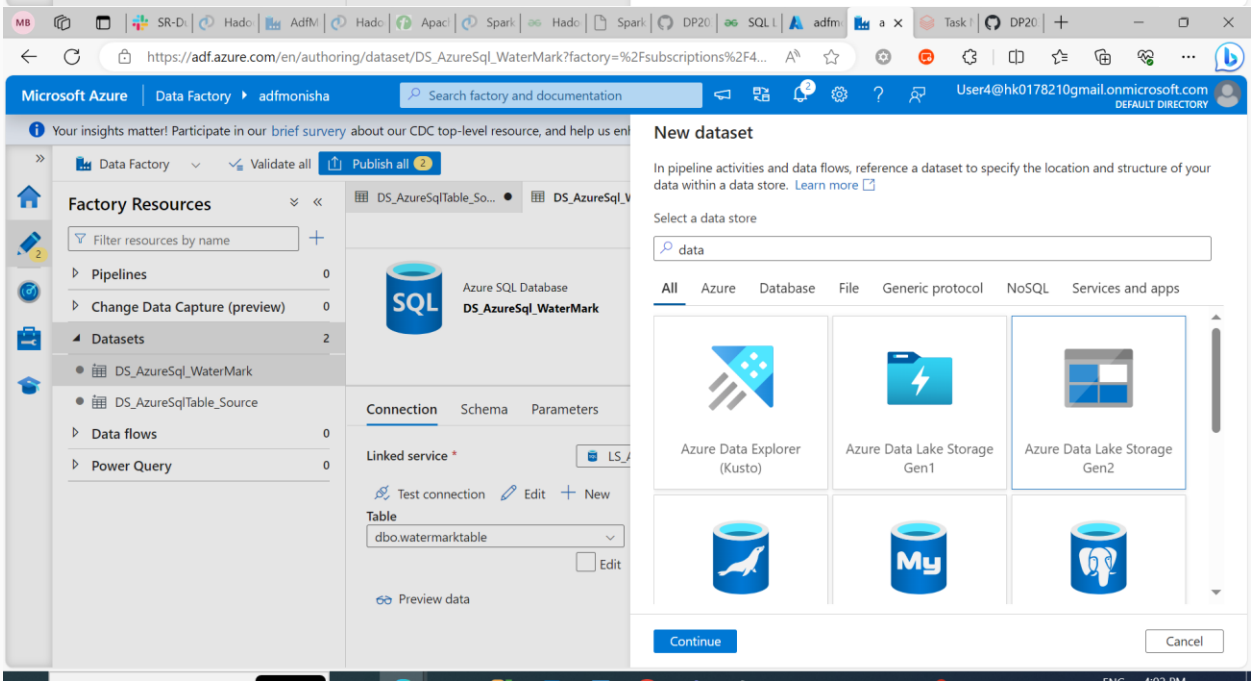

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enl
Data Factory
Validate all
Publish all
Linked services
Linked service defines the connection information to a data sto
New
Filter by name
Annotations : Any
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
If you expected to se
New linked service
Azure SQL Database
Learn more
Connect via integration runtime *
AutoResolveIntegrationRuntime
Connection string
Azure Key Vault
Account selection method
From Azure subscription
Enter manually
Azure subscription
Pay-As-You-Go (45f63aab-e46c-4be8-bb77-43fe31240084)
Server name *
usttrainserver
Database name *
SQLmonisha
Authentication type *
SQL authentication
Create
Back
Test connection
Cancel

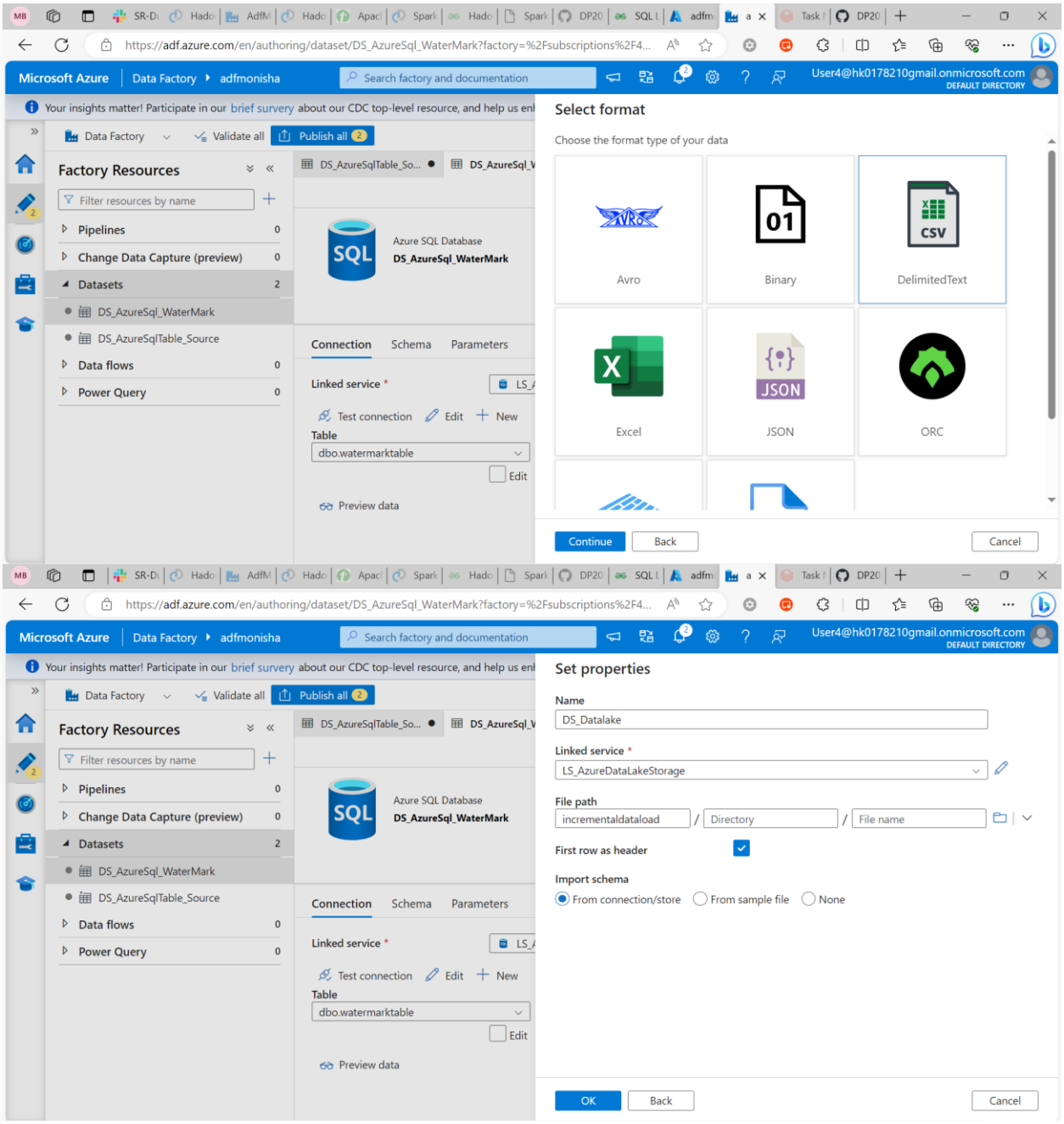

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enh
Data Factory
Validate all
Publish all
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Linked services
Linked service defines the connection information to a data sto
New
Filter by name
Annotations : Any
Showing 1 - 1 of 1 items
Name
Type
LS_AzureSqlDatabase
Azure SQL Databa
New linked service
Data store
Compute
data
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2
Azure Database for
Azure Database for
Azure Database for
Continue
Cancel


Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enh
Data Factory
Validate all
Publish all
General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Linked services
Linked service defines the connection information to a data sto
New
Filter by name
Annotations : Any
Showing 1 - 1 of 1 items
Name
Type
LS_AzureSqlDatabase
Azure SQL Databa
New linked service
Azure Data Lake Storage Gen2
Learn more
LS_AzureDataLakeStorage
Description
Connect via integration runtime *
AutoResolveIntegrationRuntime
Authentication type
Account key
Account selection method
From Azure subscription
Enter manually
Azure subscription
Pay-As-You-Go (45f63aab-e46c-4be8-bb77-43fe31240084)
Storage account name *
adlsmonisha
Create
Back
Test connection
Cancel

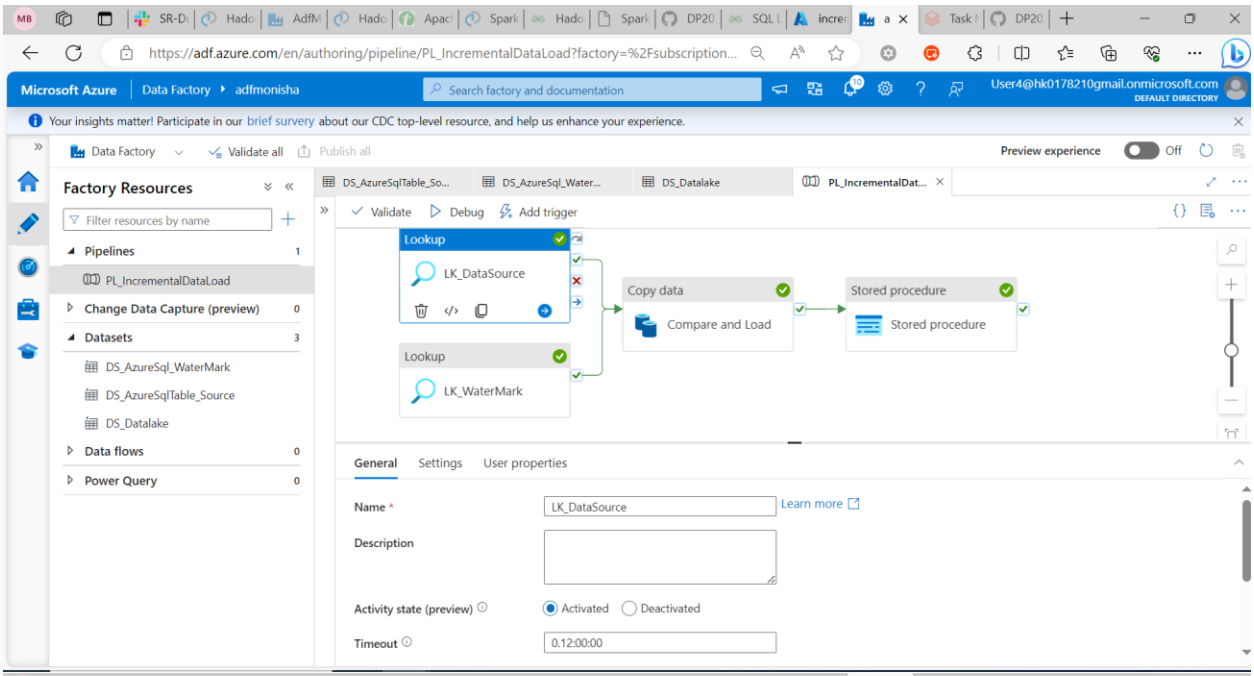

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enh
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 0
Data flows 0
Power Query 0
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
sql
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Amazon RDS for SQL Server
Azure Cosmos DB for NoSQL
Azure Database for MySQL
Azure Database for PostgreSQL
Azure SQL Database
Azure SQL Database Managed Instance
Continue
Cancel
Set properties
Name
DS_AzureSqlTable_Source
Linked service *
LS_AzureSqlDatabase
Table name
dbo.data_source_table
Edit
Import schema
From connection/store
None
OK
Back
Cancel

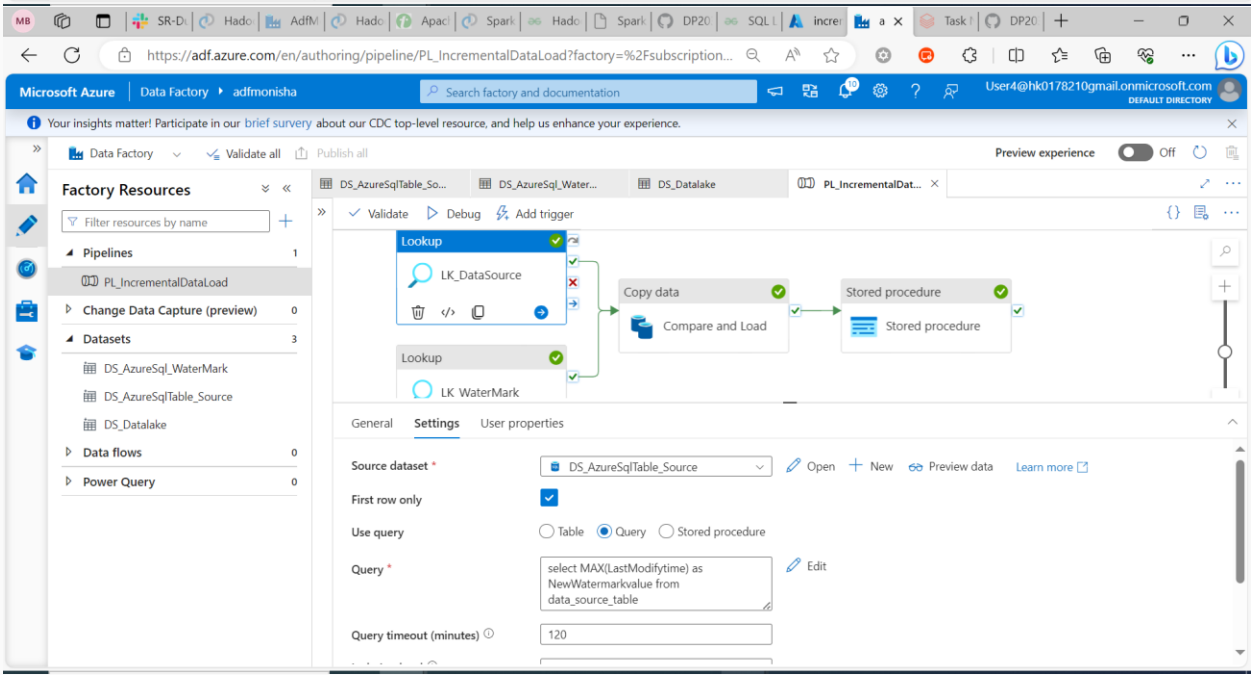

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enl
Data Factory
Validate all
Publish all 1
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 1
DS_AzureSqlTable_Source
Data flows 0
Power Query 0
DS_AzureSqlTable_S...
Azure SQL Database
DS_AzureSqlTable_Source
Connection
Schema
Parameters
Linked service *
Test connection
Edit
New
Table
dbo.data_source_table
Edit
Preview data
Set properties
Name
DS_AzureSql_WaterMark
Linked service *
LS_AzureSqlDatabase
Table name
dbo.watermarktable
Edit
Import schema
From connection/store
None
OK
Back
Cancel

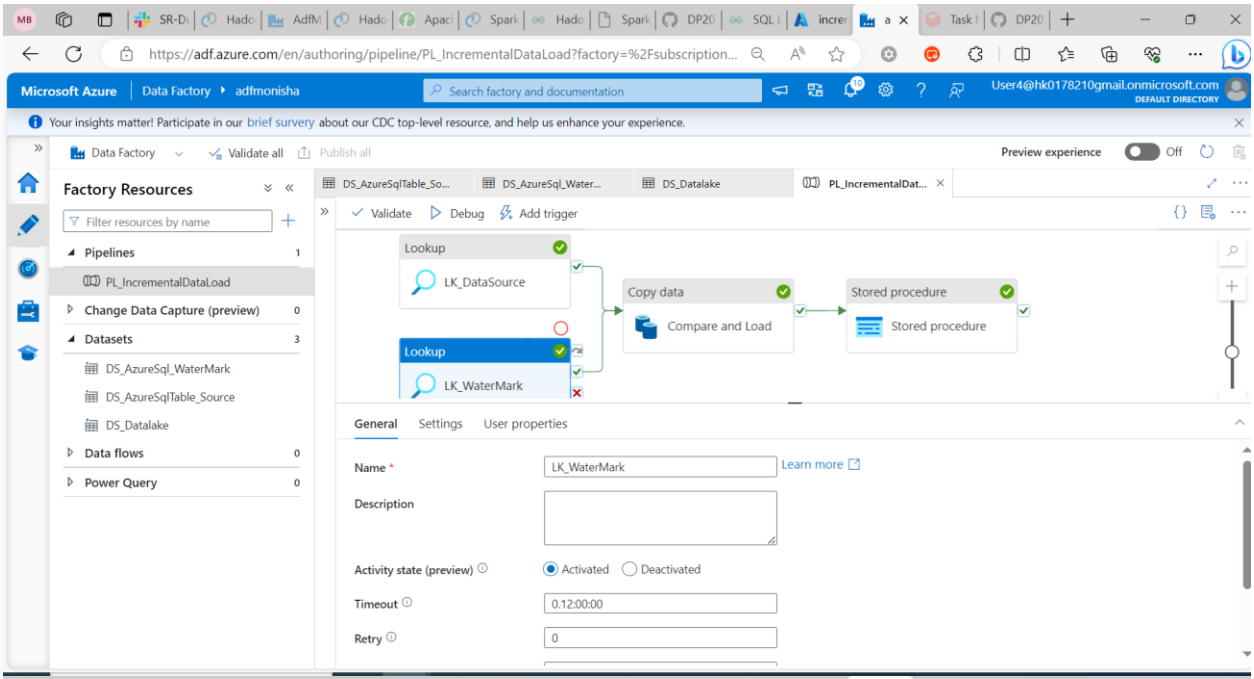

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enl
Data Factory
Validate all
Publish all 2
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 2
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_W
Azure SQL Database
DS_AzureSql_WaterMark
Connection
Schema
Parameters
Linked service *
Test connection
Edit
New
Table
dbo.watermarktable
Edit
Preview data
New dataset
In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more
Select a data store
data
All
Azure
Database
File
Generic protocol
NoSQL
Services and apps
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2
Continue
Cancel

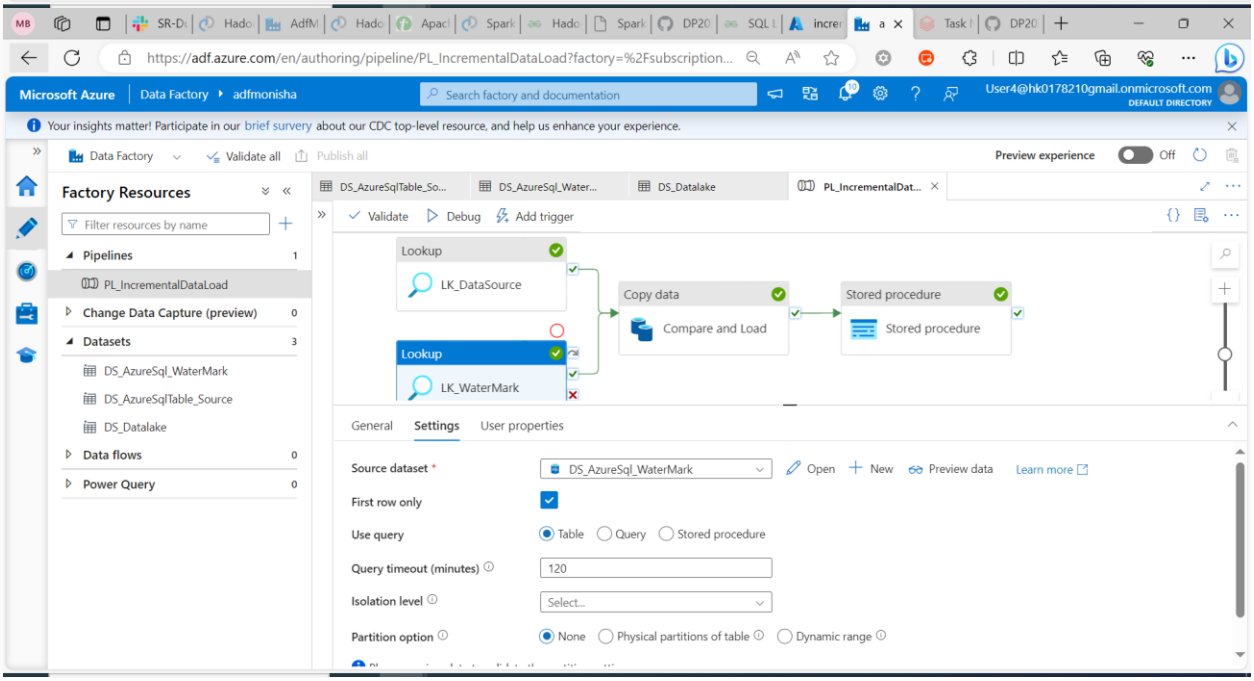

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enl
Data Factory
Validate all
Publish all 2
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 2
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_W
Azure SQL Database
DS_AzureSql_WaterMark
Connection
Schema
Parameters
Linked service *
Test connection
Edit
New
Table
dbo.watermarktable
Edit
Preview data
Select format
Choose the format type of your data
Avro
Binary
DelimitedText
Excel
JSON
ORC
Continue
Back
Cancel

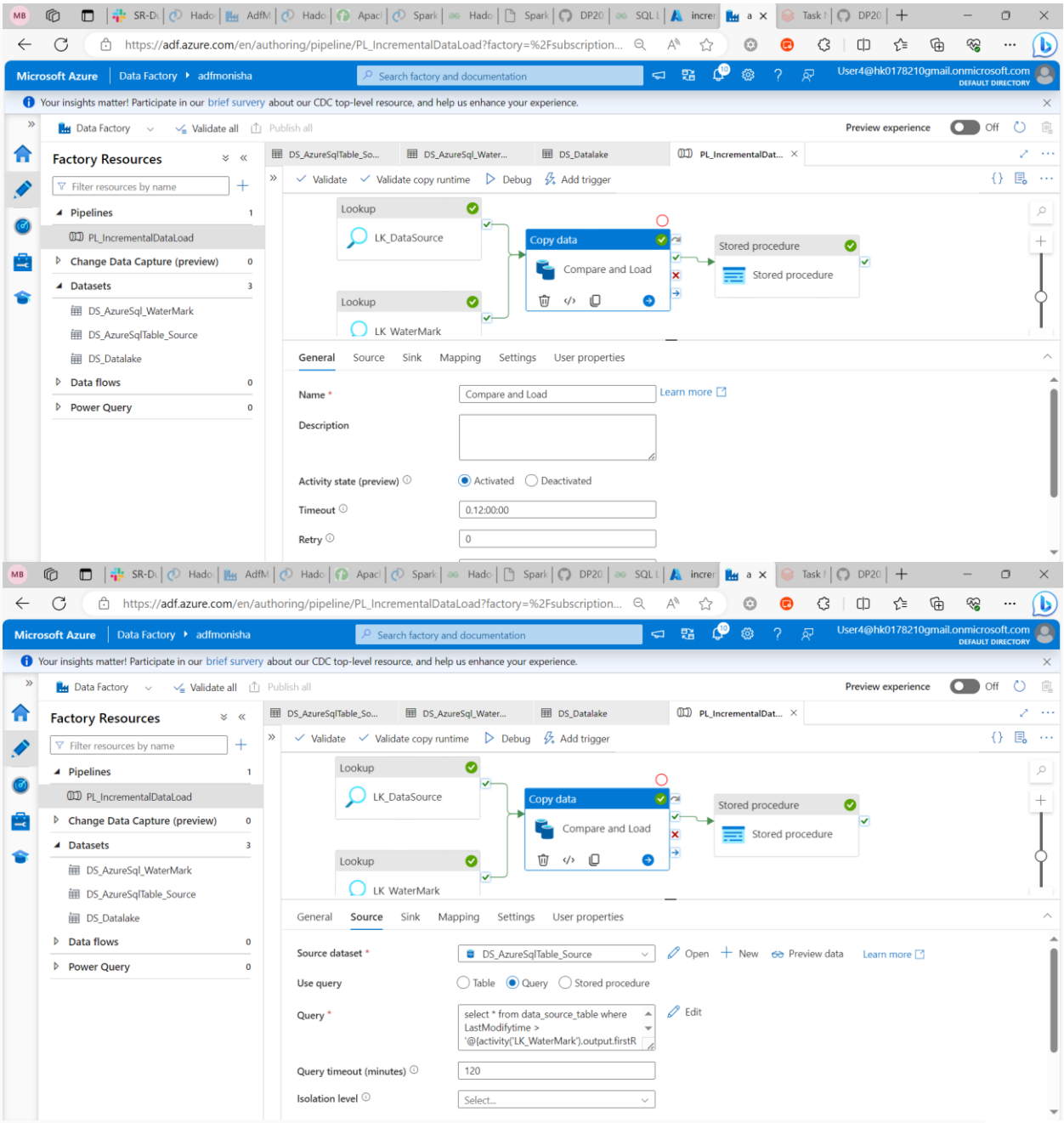
Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enl
Data Factory
Validate all
Publish all 2
Factory Resources
Filter resources by name
Pipelines 0
Change Data Capture (preview) 0
Datasets 2
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_W
Azure SQL Database
DS_AzureSql_WaterMark
Connection
Schema
Parameters
Linked service *
Test connection
Edit
New
Table
dbo.watermarktable
Edit
Preview data
Set properties
Name
DS_Datalake
Linked service *
LS_AzureDataLakeStorage
File path
incrementaldataload
Directory
File name
First row as header
Import schema
From connection/store
From sample file
None
OK
Back
Cancel


https://adf.azure.com/en/authoring/pipeline/PL_IncrementalDataLoad?factory=%2Fsubscription...
Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
PL_IncrementalDataLoad
Change Data Capture (preview) 0
Datasets 3
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
DS_Datalake
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_Water...
DS_Datalake
PL_IncrementalDat...
Validate
Debug
Add trigger
Lookup
LK_DataSource
Copy data
Compare and Load
Stored procedure
Stored procedure
Lookup
LK_WaterMark
General
Settings
User properties
Name *
LK_DataSource
Learn more
Description
Activity state (preview)
Activated
Deactivated
Timeout
0.12:00:00

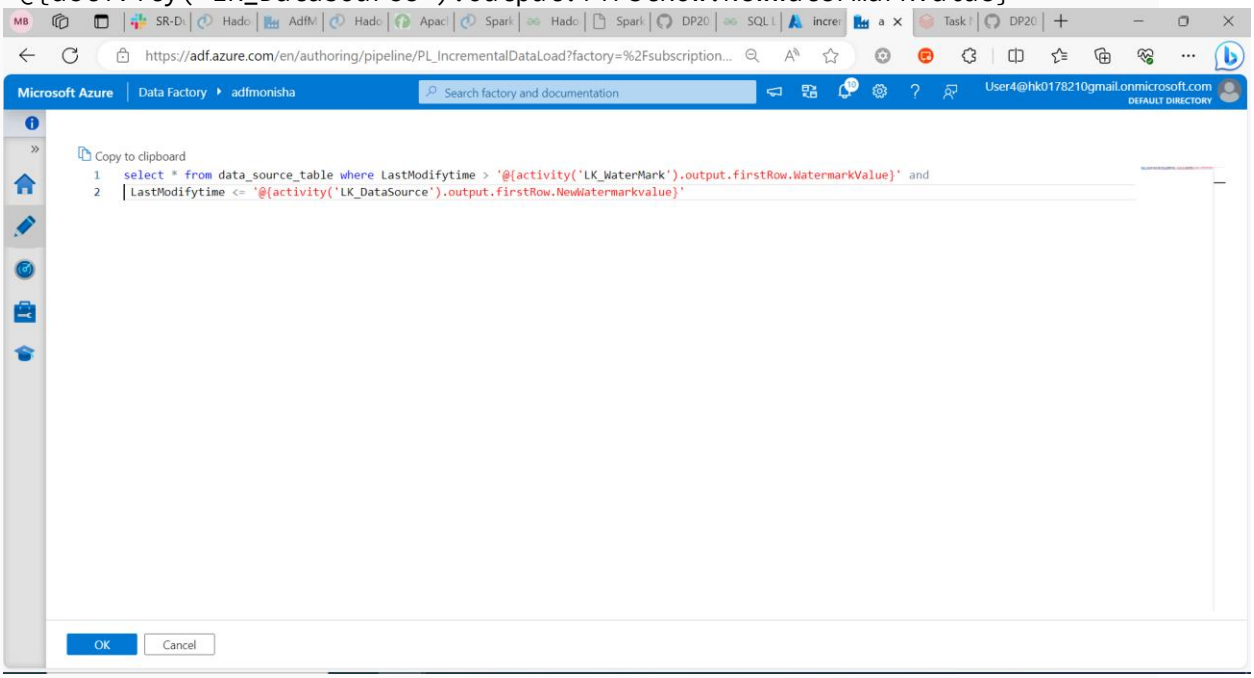

https://adf.azure.com/en/authoring/pipeline/PL_IncrementalDataLoad?factory=%2Fsubscription...
Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
PL_IncrementalDataLoad
Change Data Capture (preview) 0
Datasets 3
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
DS_Datalake
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_Water...
DS_Datalake
PL_IncrementalDat...
Validate
Debug
Add trigger
Lookup
LK_DataSource
Copy data
Compare and Load
Stored procedure
Stored procedure
Lookup
LK_WaterMark
General
Settings
User properties
Source dataset *
DS_AzureSqlTable_Source
Open
New
Preview data
Learn more
First row only
Use query
Table
Query
Stored procedure
Query *
select MAX(LastModifytime) as NewWatermarkvalue from data_source_table
Edit
Query timeout (minutes)
120

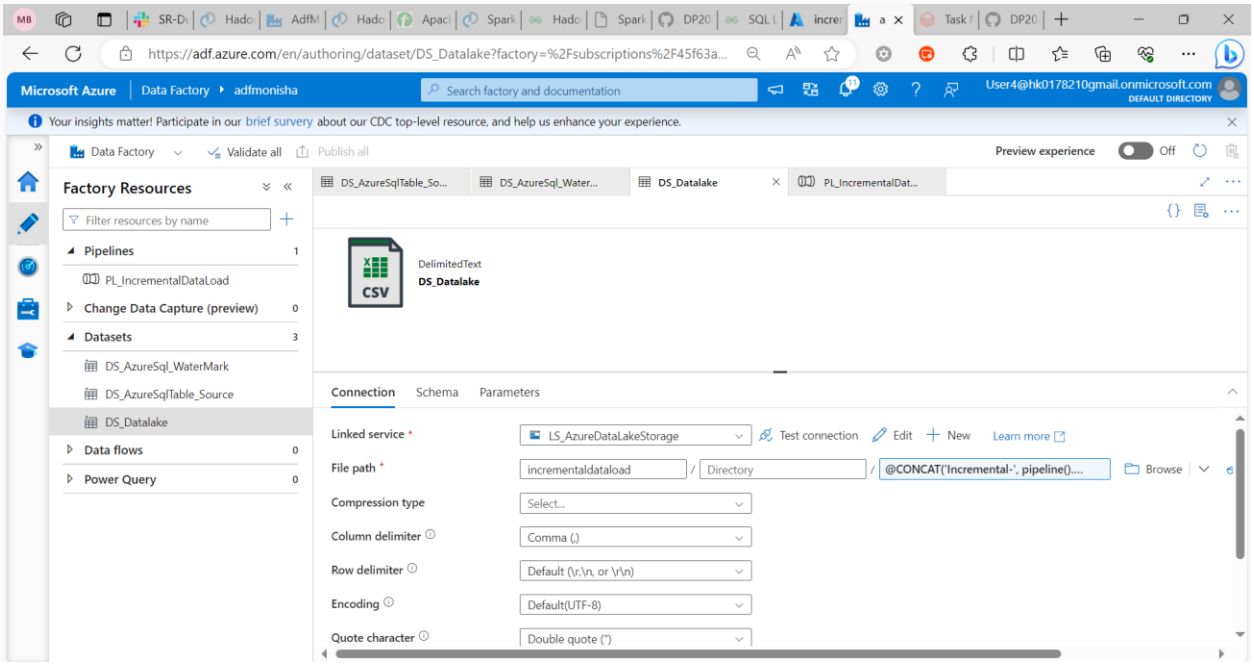

https://adf.azure.com/en/authoring/pipeline/PL_IncrementalDataLoad?factory=%2Fsubscription...
Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
PL_IncrementalDataLoad
Change Data Capture (preview) 0
Datasets 3
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
DS_Datalake
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_Water...
DS_Datalake
PL_IncrementalDat...
Validate
Debug
Add trigger
Lookup
LK_DataSource
Copy data
Compare and Load
Stored procedure
Stored procedure
Lookup
LK_WaterMark
General
Settings
User properties
Name *
LK_WaterMark
Learn more
Description
Activity state (preview)
Activated
Deactivated
Timeout
0.12:00:00
Retry
0


https://adf.azure.com/en/authoring/pipeline/PL_IncrementalDataLoad?factory=%2Fsubscription...
Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
PL_IncrementalDataLoad
Change Data Capture (preview) 0
Datasets 3
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
DS_Datalake
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_Water...
DS_Datalake
PL_IncrementalDat...
Validate
Debug
Add trigger
Lookup
LK_DataSource
Copy data
Compare and Load
Stored procedure
Stored procedure
Lookup
LK_WaterMark
General
Settings
User properties
Source dataset *
DS_AzureSql_WaterMark
Open
New
Preview data
Learn more
First row only
Use query
Table
Query
Stored procedure
Query timeout (minutes)
120
Isolation level
Select...
Partition option
None
Physical partitions of table
Dynamic range

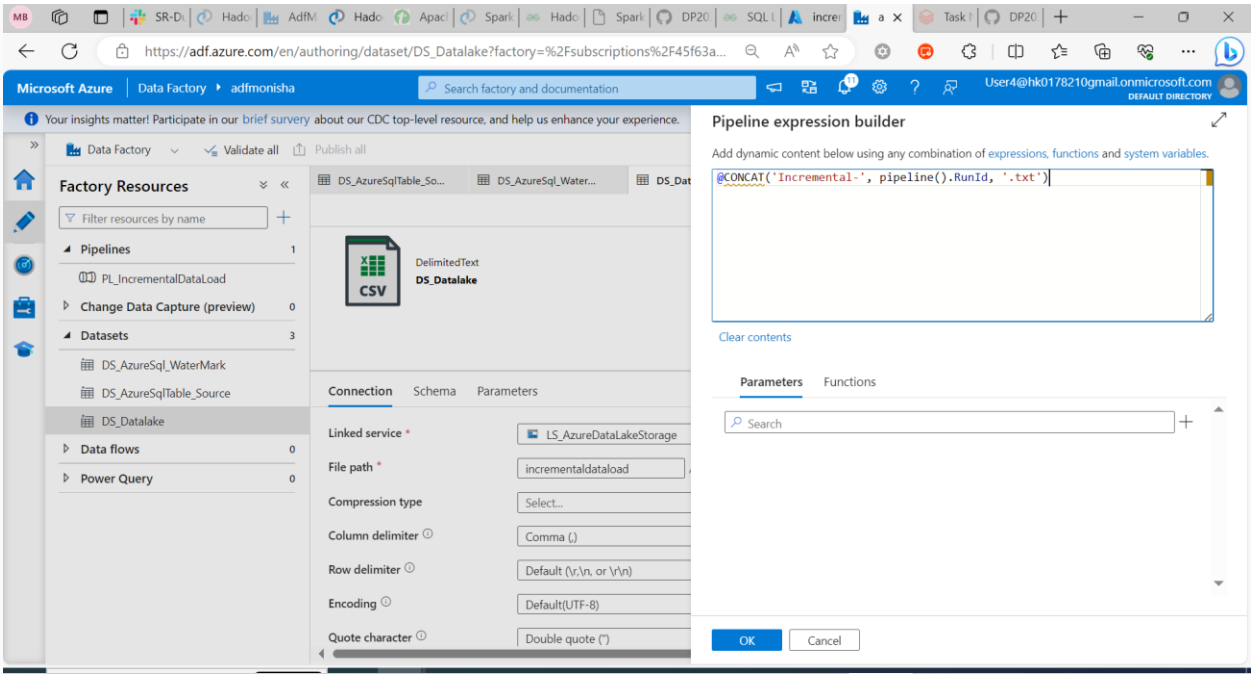

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines 1
PL_IncrementalDataLoad
Change Data Capture (preview) 0
Datasets 3
DS_AzureSql_WaterMark
DS_AzureSqlTable_Source
DS_Datalake
Data flows 0
Power Query 0
DS_AzureSqlTable_So...
DS_AzureSql_Water...
DS_Datalake
PL_IncrementalDat...
Validate
Validate copy runtime
Debug
Add trigger
Lookup
LK_DataSource
Copy data
Compare and Load
Stored procedure
Stored procedure
Lookup
LK WaterMark
General
Source
Sink
Mapping
Settings
User properties
Name *
Compare and Load
Learn more
Description
Activity state (preview)
Activated
Deactivated
Timeout
0.12:00:00
Retry
0
Source dataset *
DS_AzureSqlTable_Source
Open
New
Preview data
Learn more
Use query
Table
Query
Stored procedure
Query *
select * from data_source_table where LastModifytime > '@{activity('LK_WaterMark').output.firstR
Edit
Query timeout (minutes)
120
Isolation level
Select...

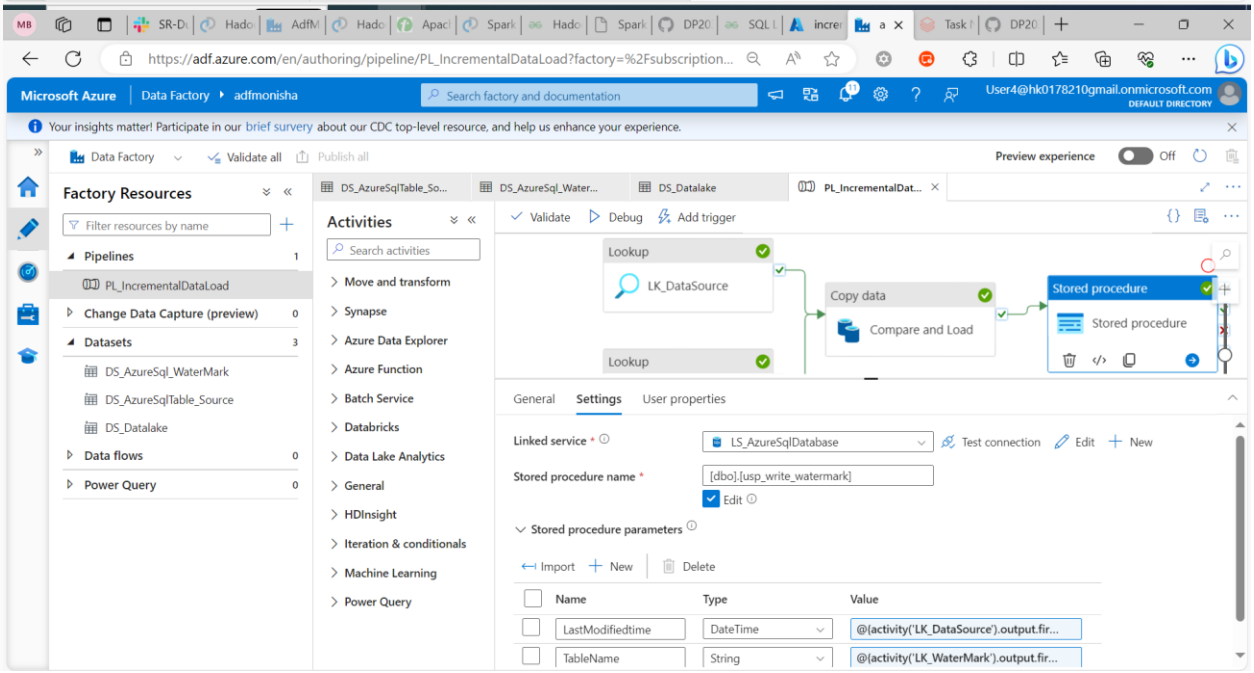
```
select * from data_source_table where LastModifytime >
'@{activity('LK_WaterMark').output.firstRow.WatermarkValue}' and
LastModifytime <=
'@{activity('LK_DataSource').output.firstRow.NewWatermarkvalue}'
```



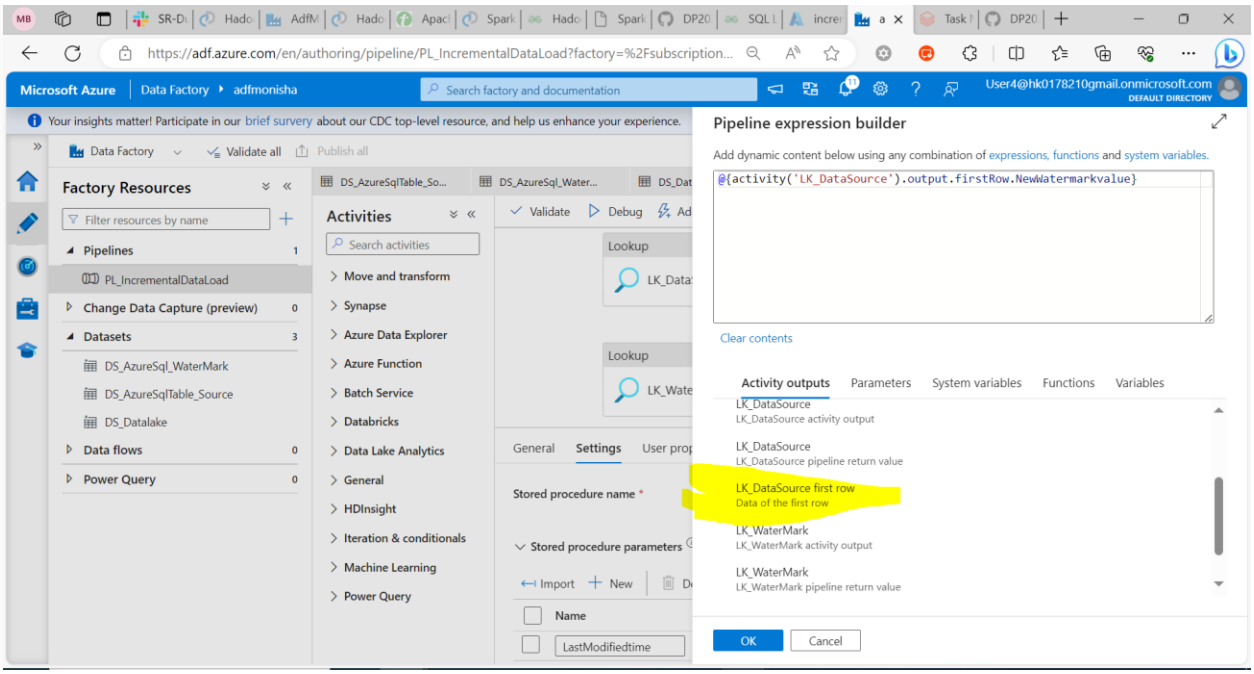



```
@CONCAT('Incremental-', pipeline().RunId, '.txt')
```

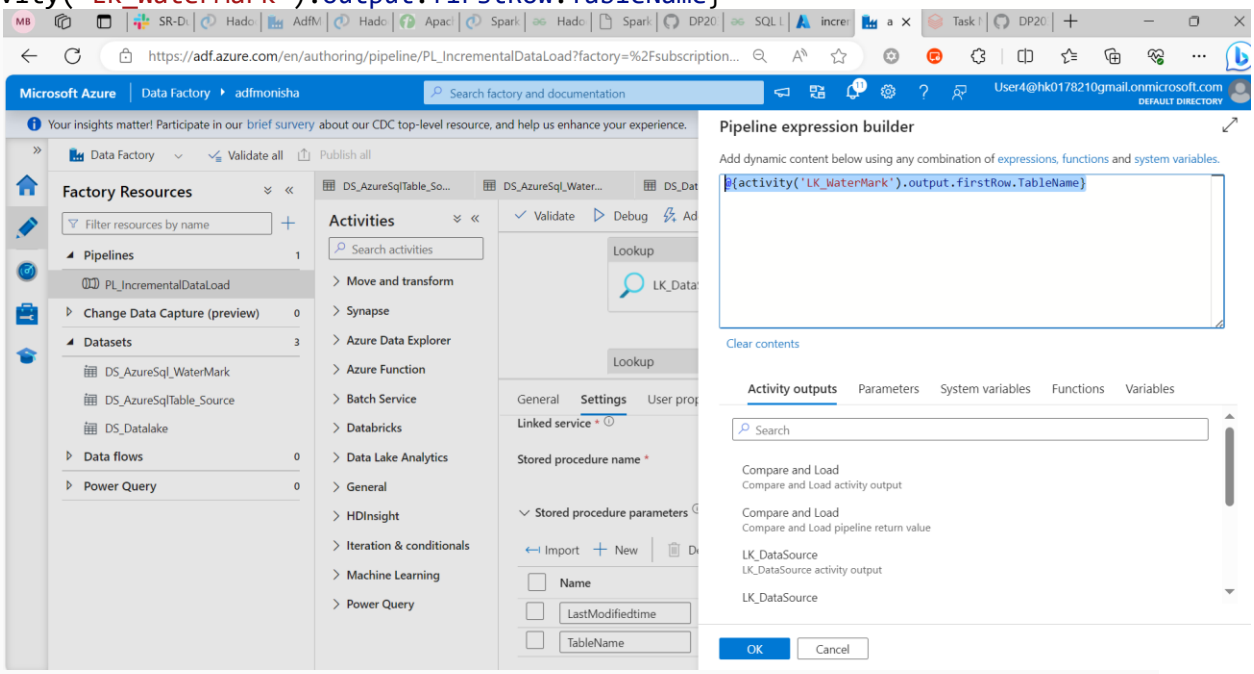


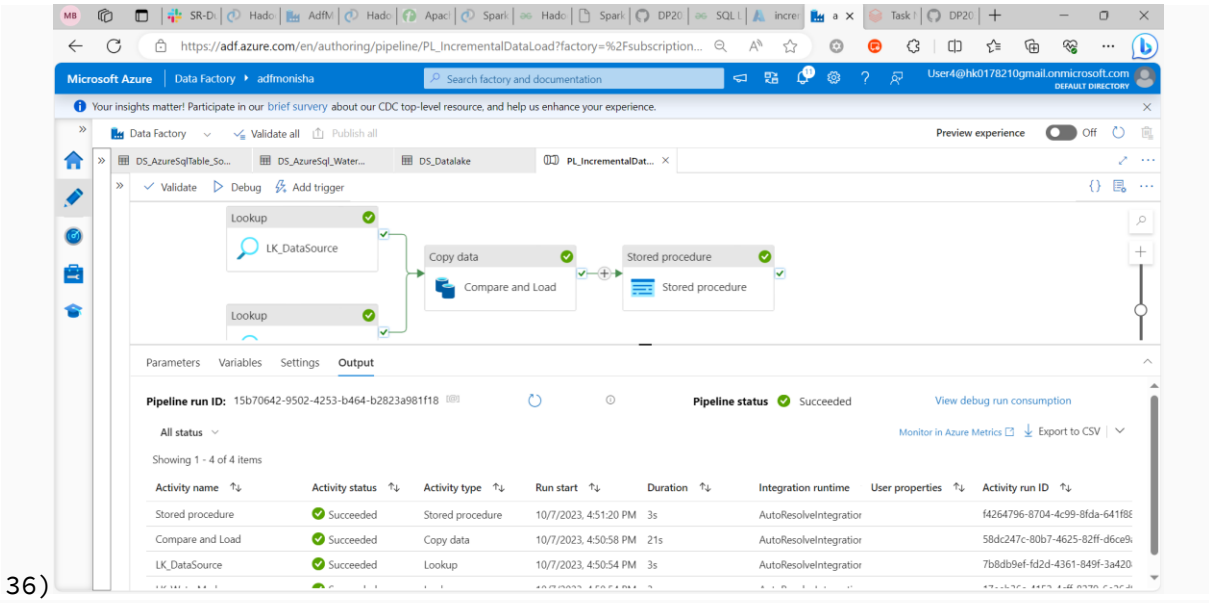



```
@{activity('LK_DataSource').output.firstRow.NewWatermarkvalue}
```



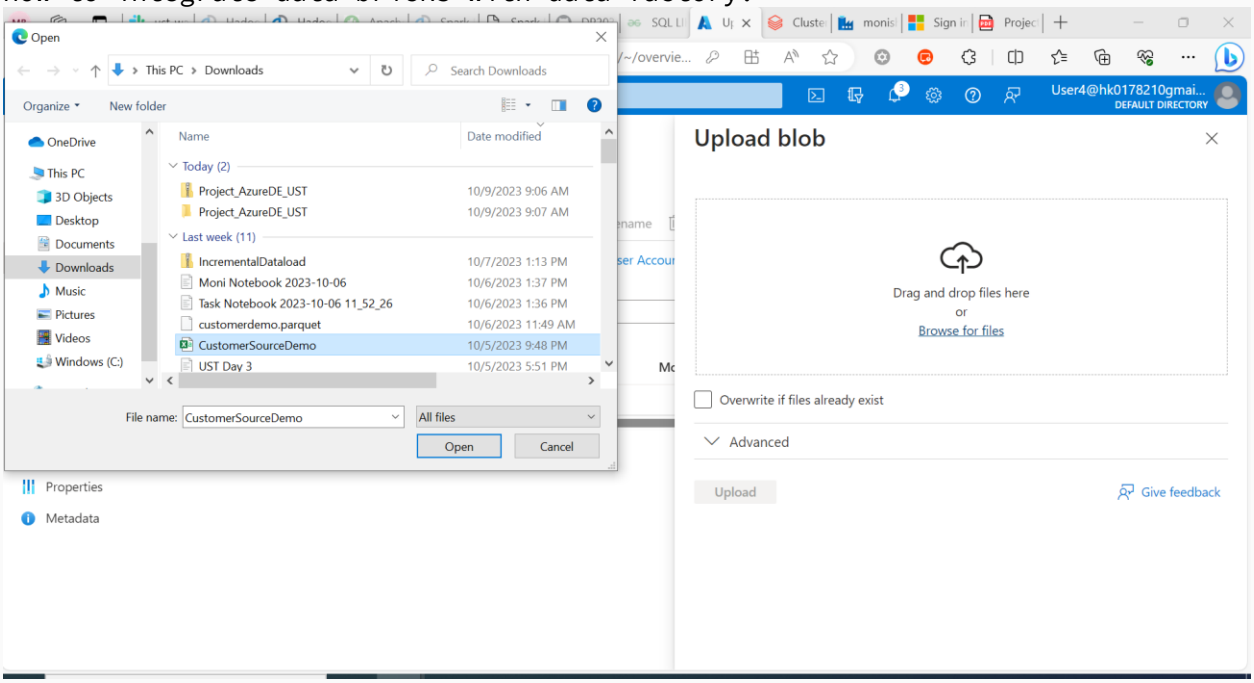

```
@{activity('LK_WaterMark').output.firstRow.TableName}
```



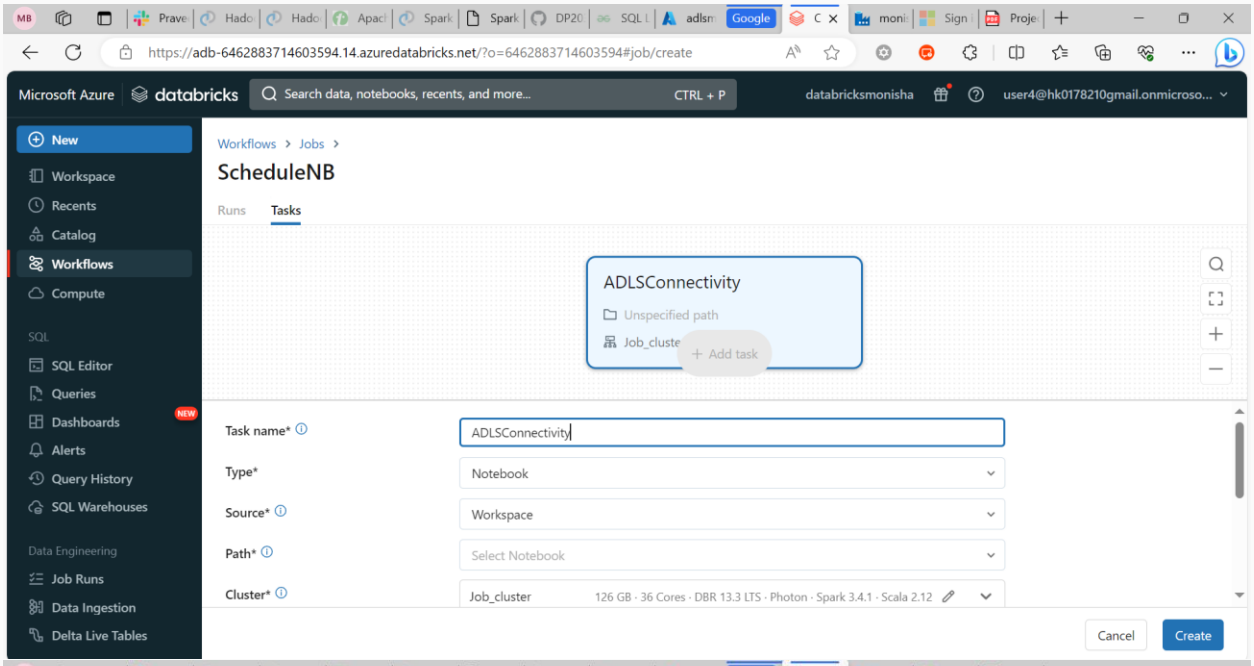


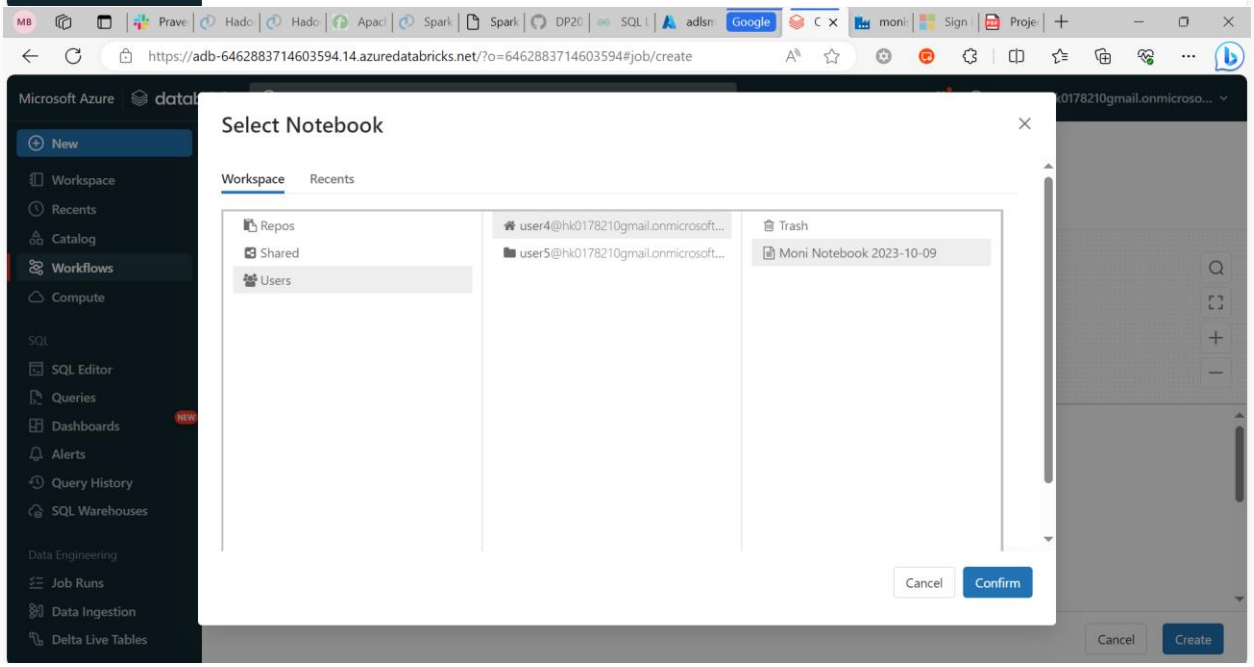
36)

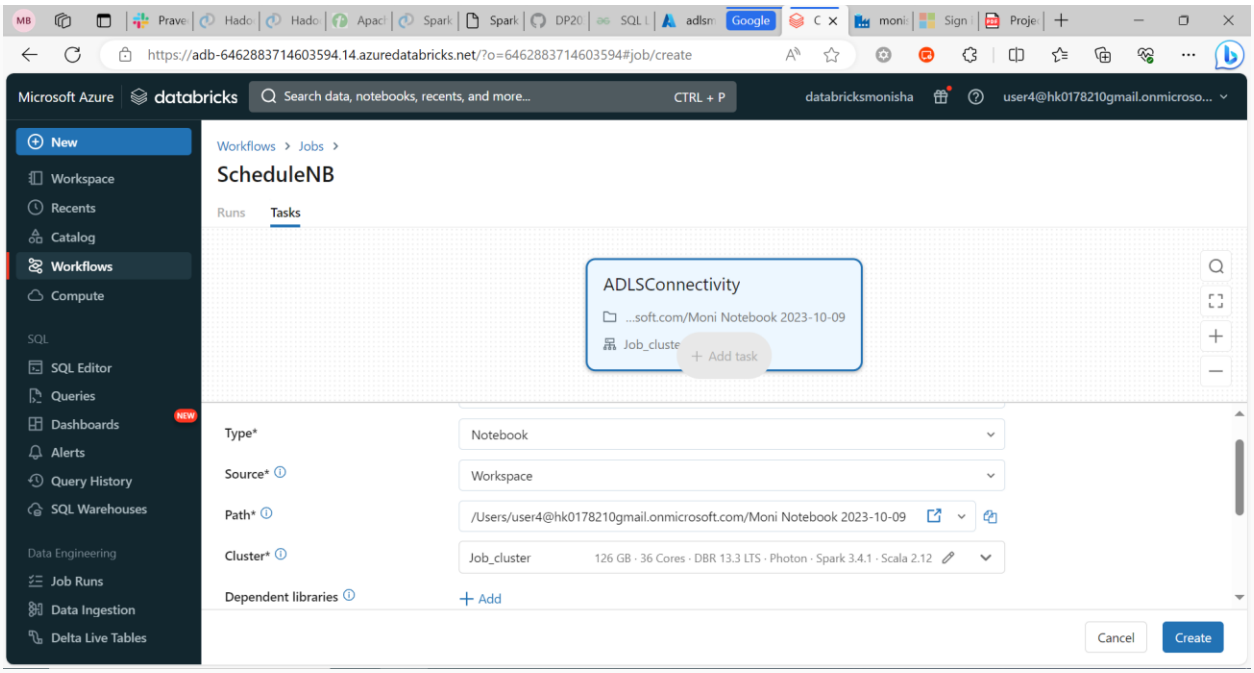
How to integrate data bricks with data factory:



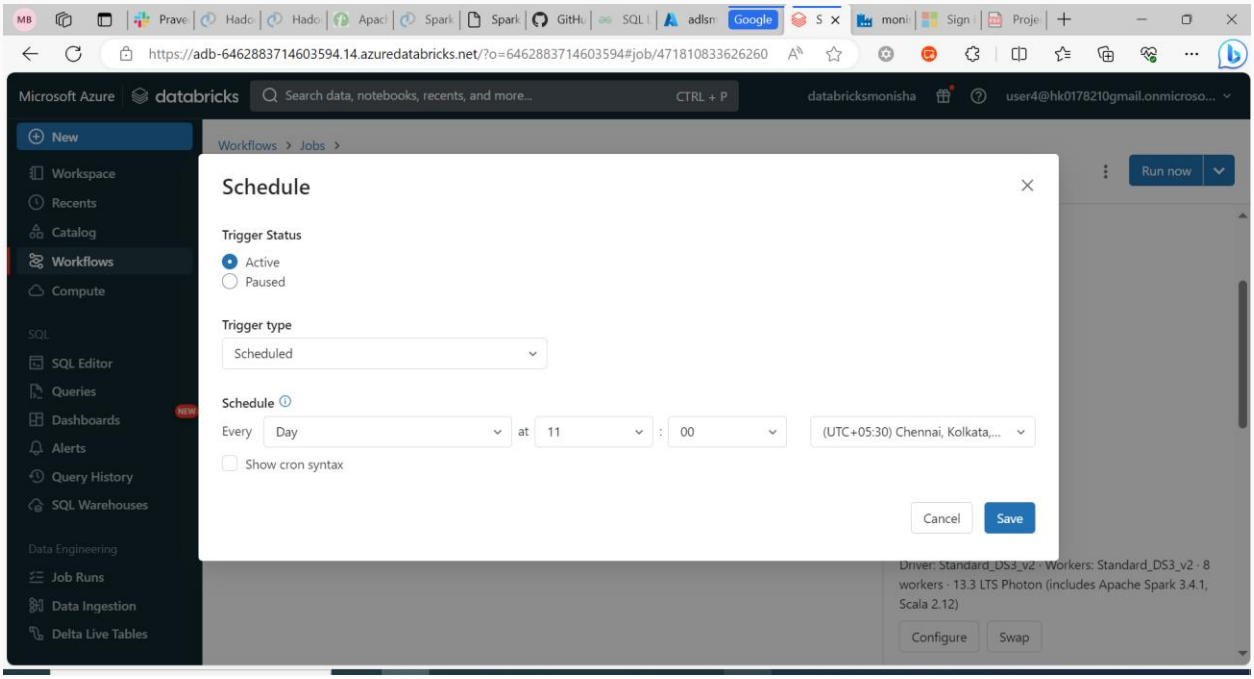
Create new notebook ➔

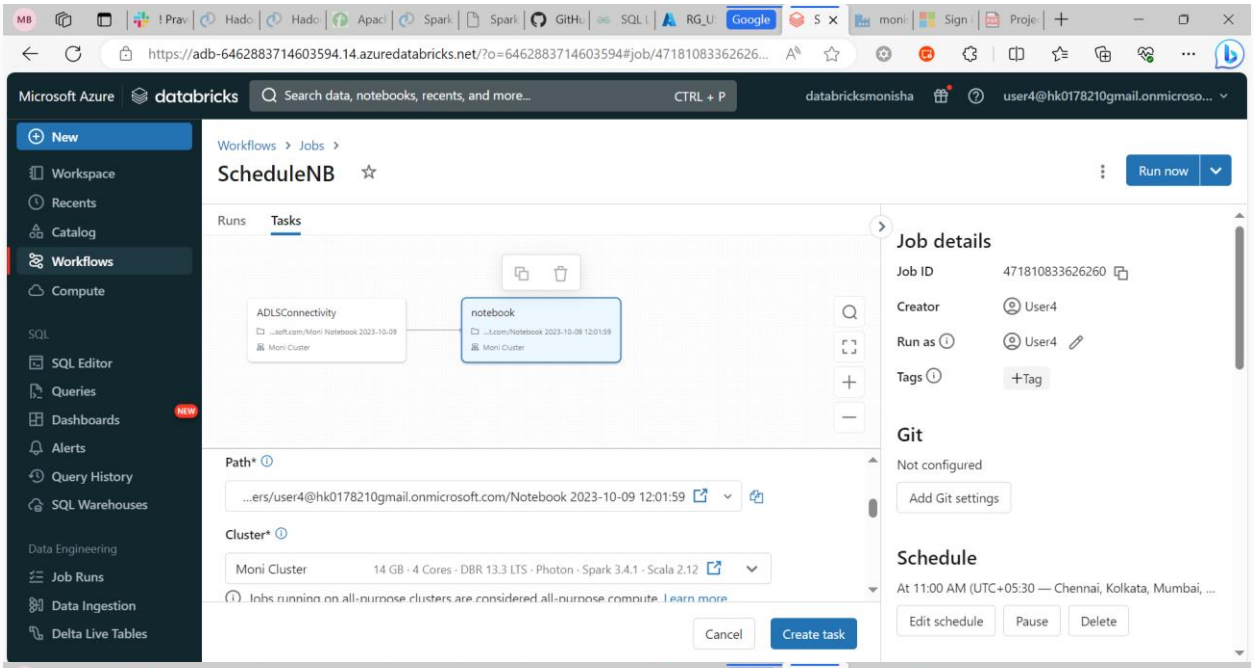

https://adb-6462883714603594.14.azuredatabricks.net/?o=6462883714603594#job/create
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricksmonisha
user4@hk0178210gmail.onmicroso...
New
Workspace
Recents
Catalog
Workflows
Compute
SQL
SQL Editor
Queries
Dashboards
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
Delta Live Tables
Workflows > Jobs >
ScheduleNB
Runs
Tasks
ADLSConnectivity
Unspecified path
Job_cluste
Add task
Task name*
ADLSConnectivity
Type*
Notebook
Source*
Workspace
Path*
Select Notebook
Cluster*
Job_cluster
126 GB · 36 Cores · DBR 13.3 LTS · Photon · Spark 3.4.1 · Scala 2.12
Cancel
Create

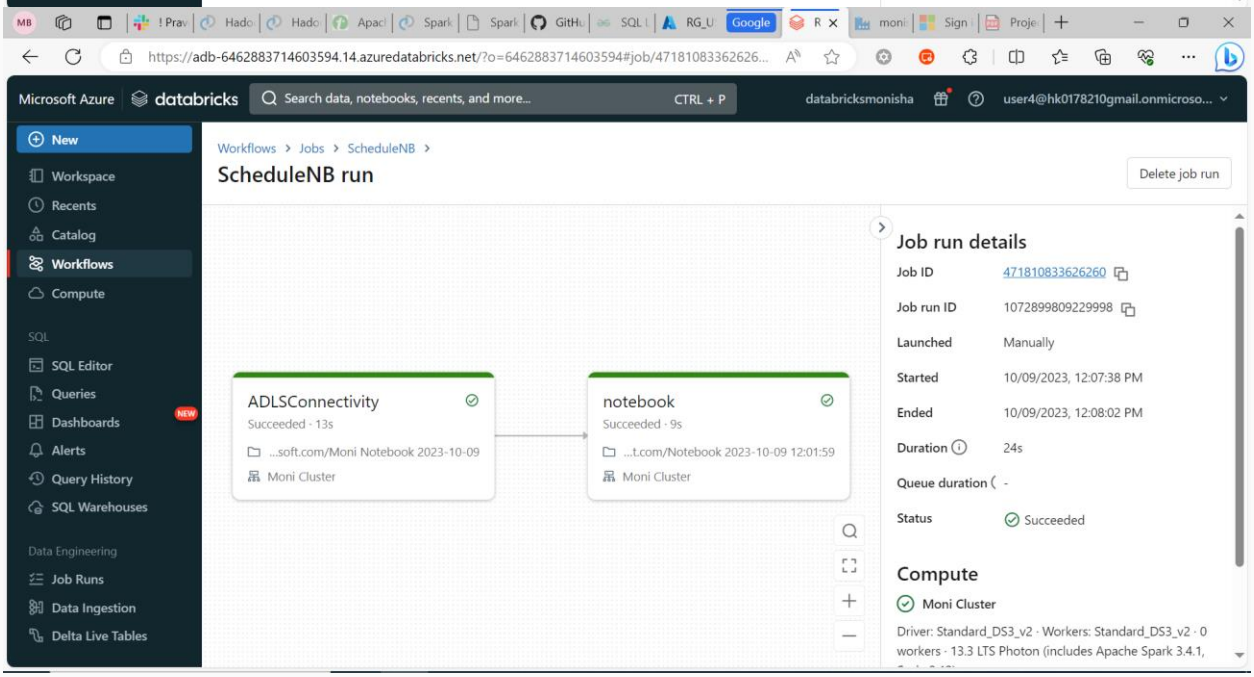

https://adb-6462883714603594.14.azuredatabricks.net/?o=6462883714603594#job/create
Select Notebook
Workspace
Recents
Repos
Shared
Users
user4@hk0178210gmail.onmicrosoft...
user5@hk0178210gmail.onmicrosoft...
Trash
Moni Notebook 2023-10-09
Cancel
Confirm
Cancel
Create



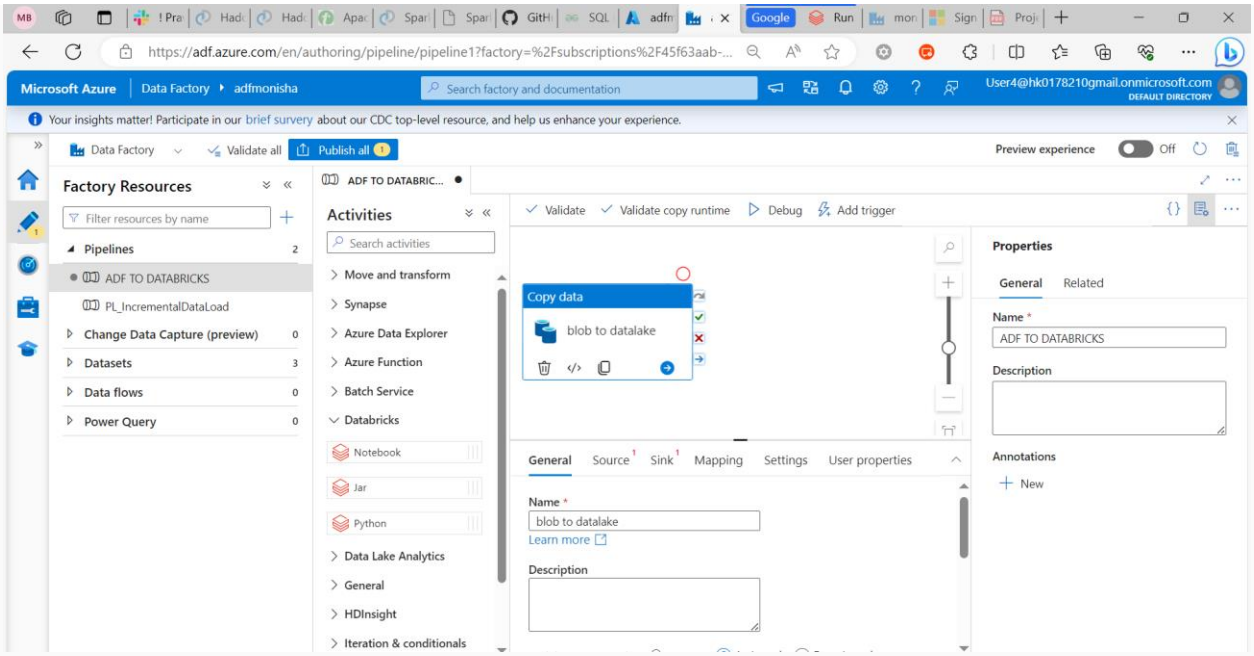
Schedule ➔ add schedule

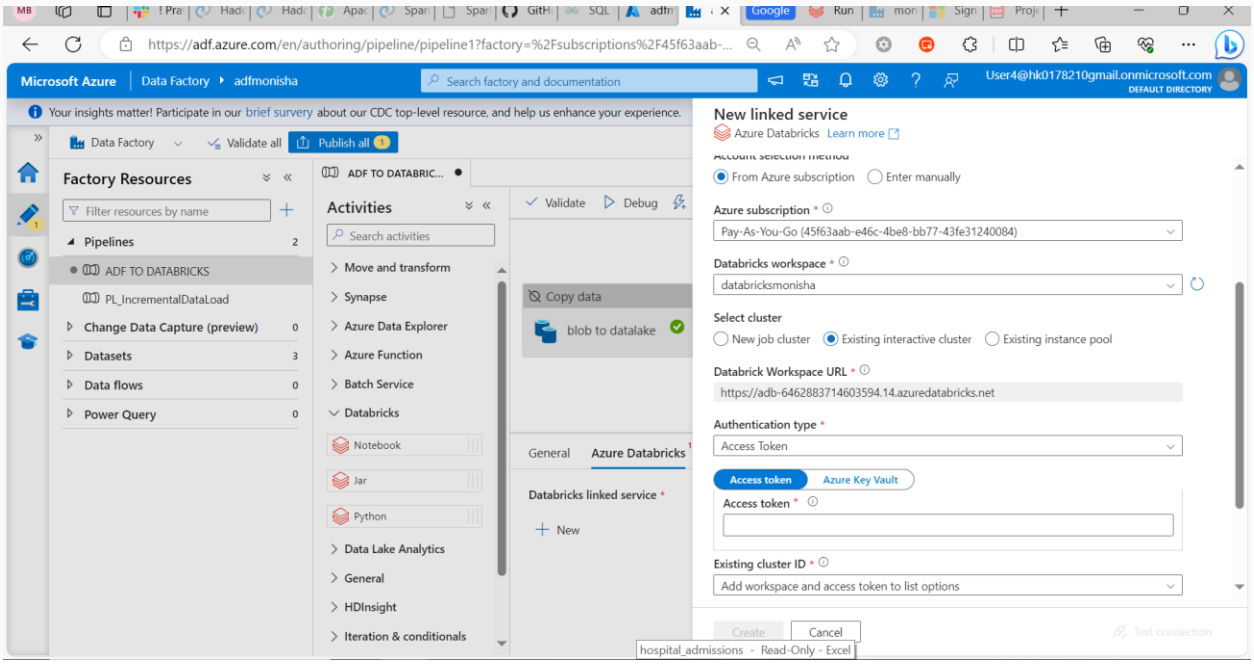


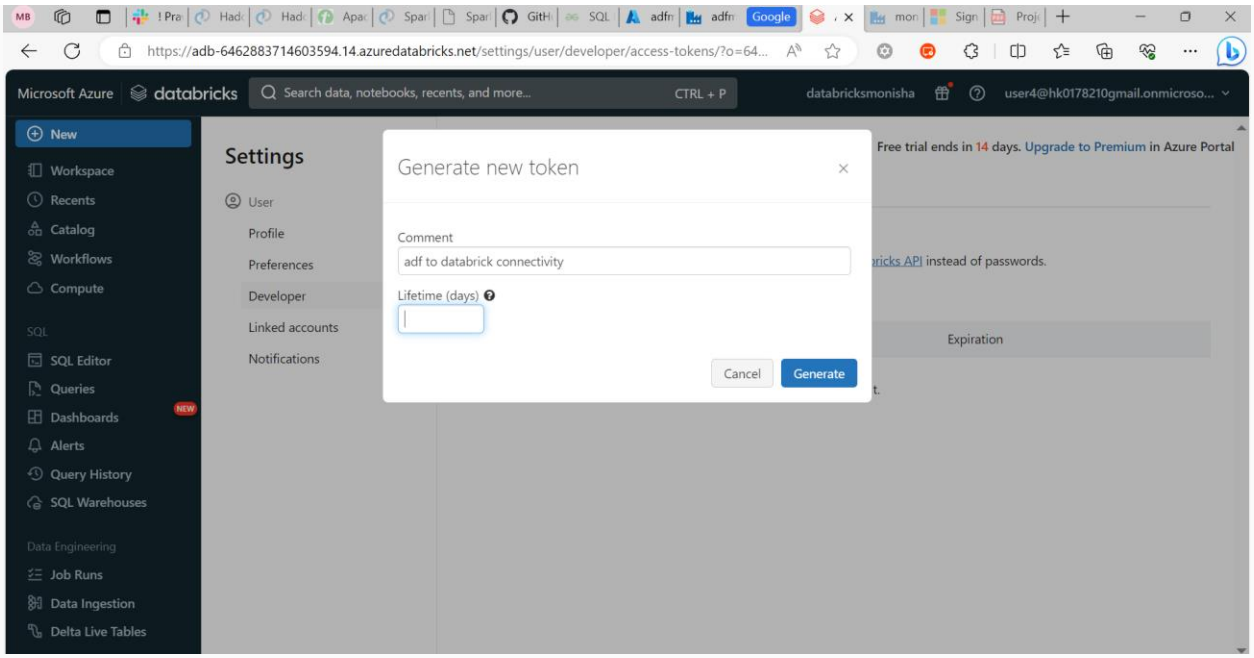




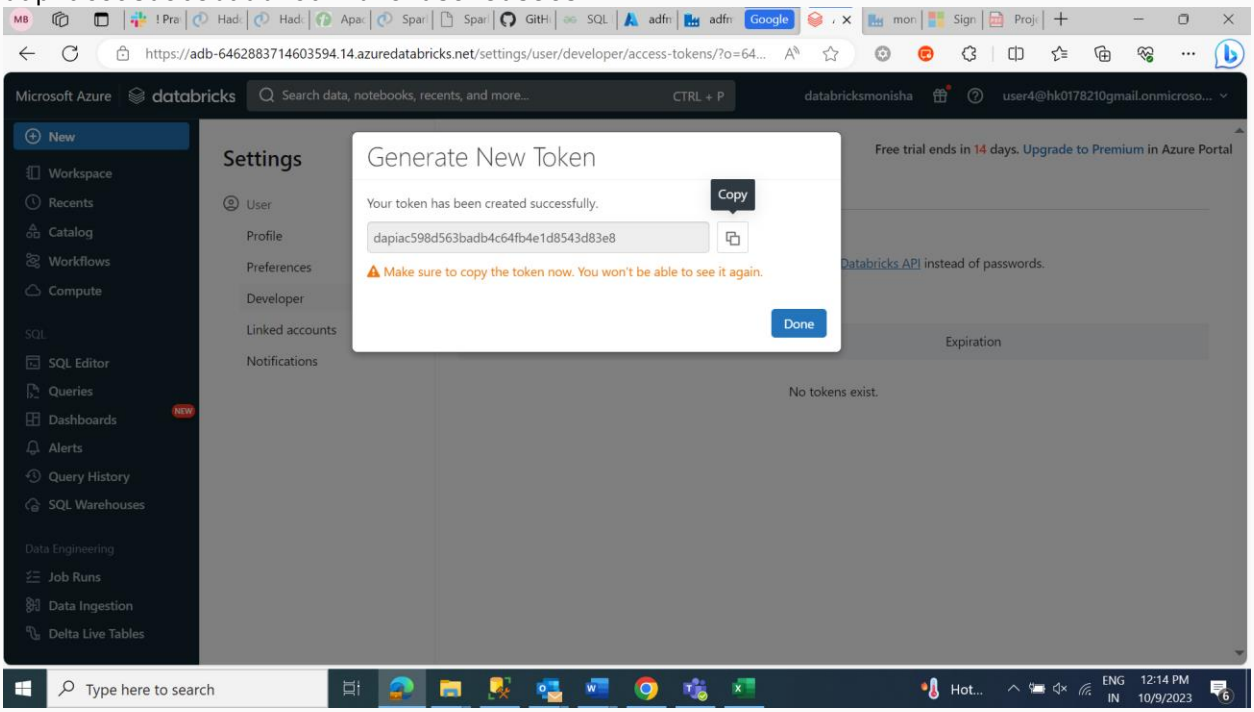
DATA FACTORY TO DATABRICKS:

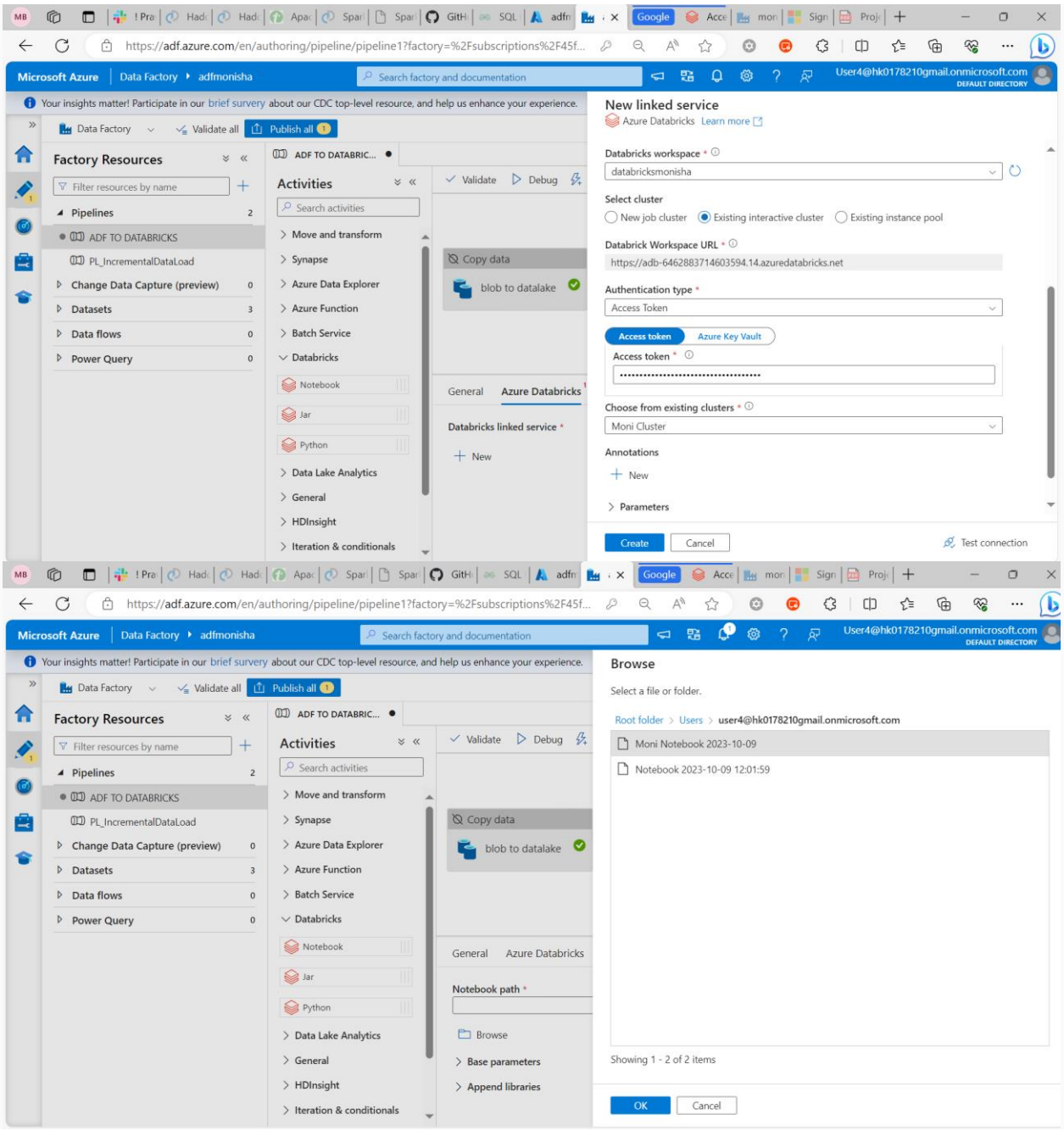

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Validate all
Publish all
Preview experience
Off
Factory Resources
Filter resources by name
Pipelines
ADF TO DATABRICKS
PL_IncrementalDataLoad
Change Data Capture (preview)
Datasets
Data flows
Power Query
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Notebook
Jar
Python
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Validate
Validate copy runtime
Debug
Add trigger
Copy data
blob to datalake
Properties
General
Related
Name *
ADF TO DATABRICKS
Description
Annotations
New
General
Source
Sink
Mapping
Settings
User properties
Name *
blob to datalake
Learn more
Description

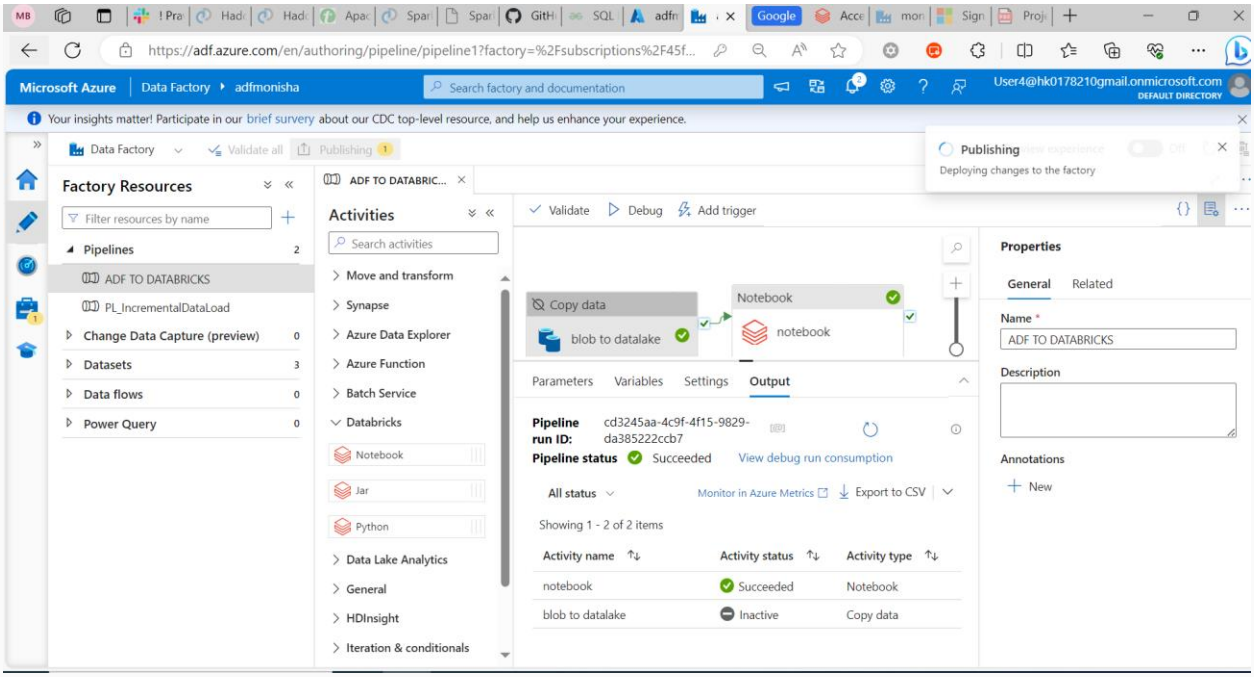

Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
ADF TO DATABRICKS
PL_IncrementalDataLoad
Change Data Capture (preview)
Datasets
Data flows
Power Query
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Notebook
Jar
Python
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Validate
Debug
Copy data
blob to datalake
General
Azure Databricks
Databricks linked service *
New
New linked service
Azure Databricks
Learn more
From Azure subscription
Enter manually
Azure subscription *
Pay-As-You-Go (45f63aab-e46c-4be8-bb77-43fe31240084)
Databricks workspace *
databricksmonisha
Select cluster
New job cluster
Existing interactive cluster
Existing instance pool
Databrick Workspace URL *
https://adb-6462883714603594.14.azuredatabricks.net
Authentication type *
Access Token
Access token
Azure Key Vault
Access token *
Existing cluster ID *
Add workspace and access token to list options
Create
Cancel
Test connection
hospital_admissions - Read-Only - Excel



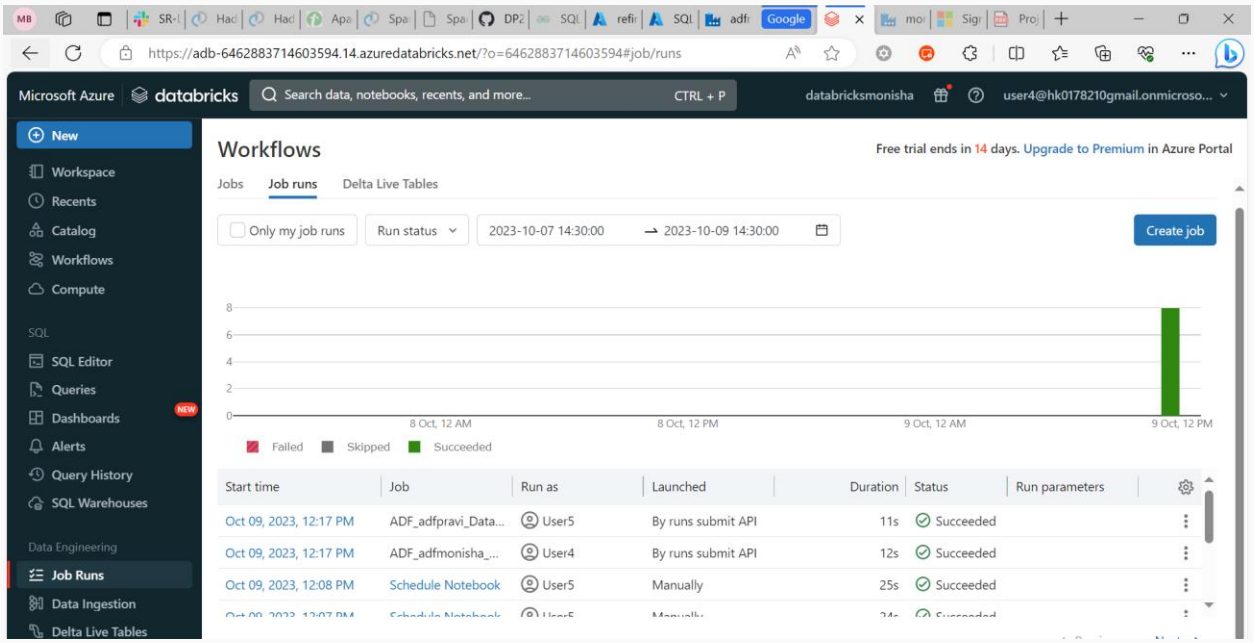
dapiac598d563badb4c64fb4e1d8543d83e8




Microsoft Azure
Data Factory
adfmonisha
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Factory Resources
Filter resources by name
Pipelines
ADF TO DATABRICKS
PL_IncrementalDataLoad
Change Data Capture (preview)
Datasets
Data flows
Power Query
ADF TO DATABRIC...
Activities
Search activities
Move and transform
Synapse
Azure Data Explorer
Azure Function
Batch Service
Databricks
Notebook
Jar
Python
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Validate
Debug
Copy data
blob to datalake
General
Azure Databricks
Databricks linked service *
New
New linked service
Azure Databricks
Learn more
Databricks workspace *
databricksmonisha
Select cluster
New job cluster
Existing interactive cluster
Existing instance pool
Databrick Workspace URL *
https://adb-6462883714603594.14.azuredatabricks.net
Authentication type *
Access Token
Access token
Azure Key Vault
Access token *
Choose from existing clusters *
Moni Cluster
Annotations
New
Parameters
Create
Cancel
Test connection
Browse
Select a file or folder.
Root folder
Users
user4@hk0178210gmail.onmicrosoft.com
Moni Notebook 2023-10-09
Notebook 2023-10-09 12:01:59
Showing 1 - 2 of 2 items
Notebook path *
Browse
Base parameters
Append libraries
OK
Cancel



To check:

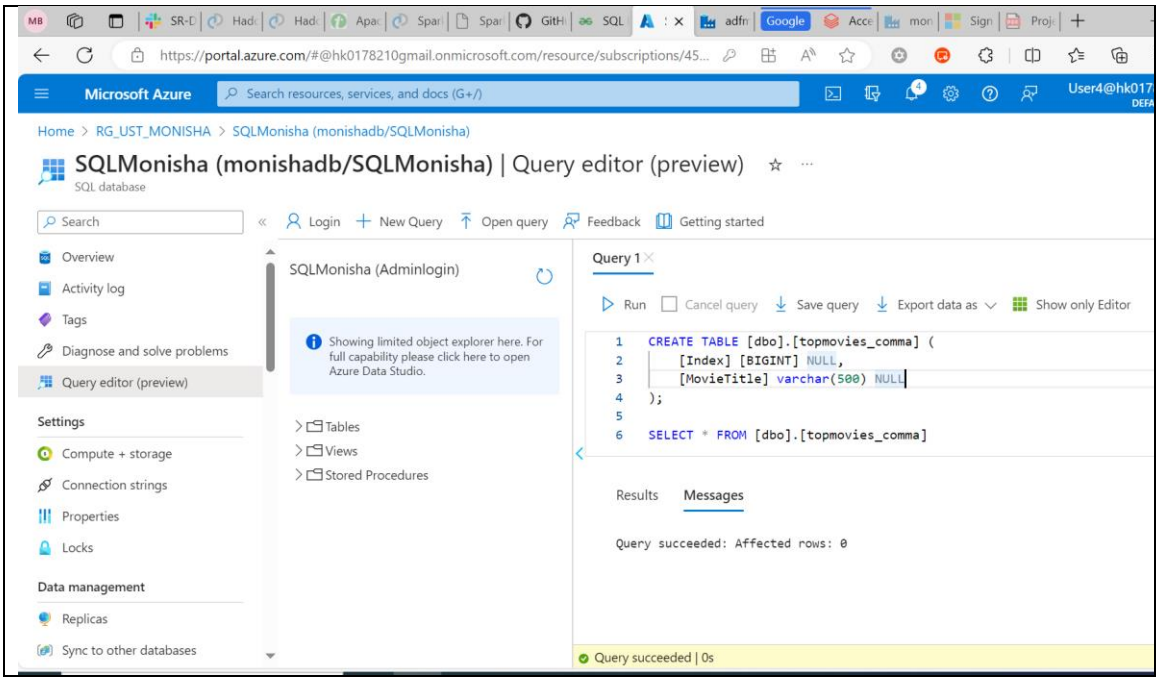


nono

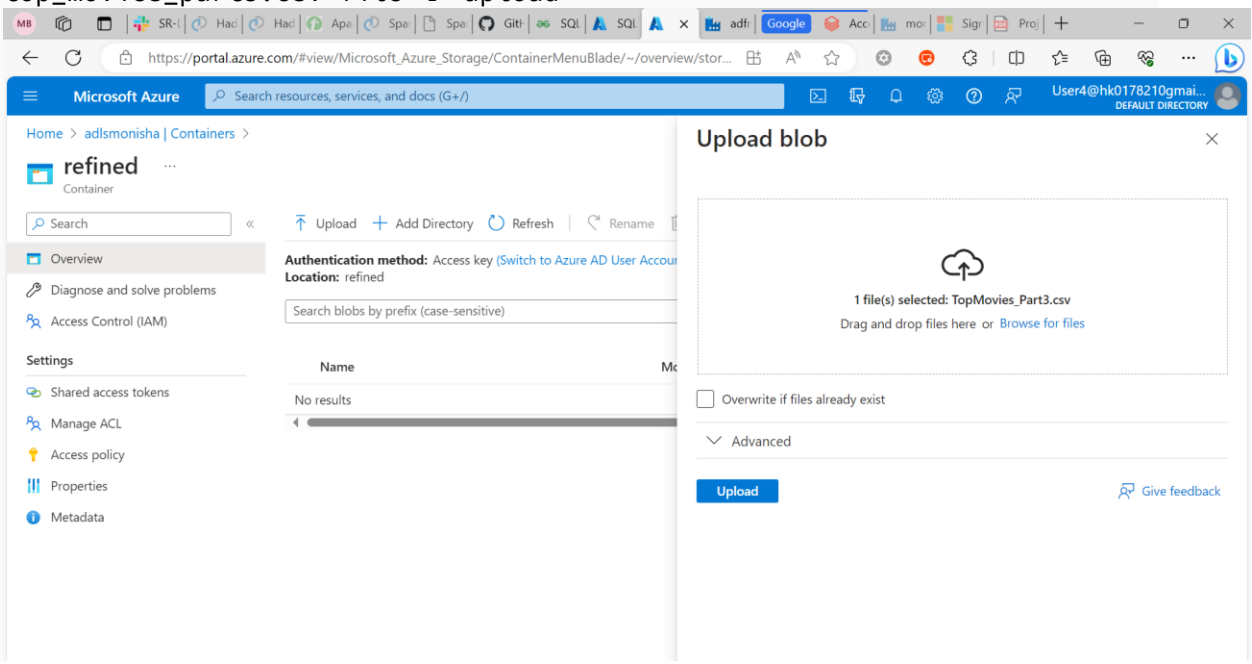
SQL TO DATABRICKS CONNECTION:

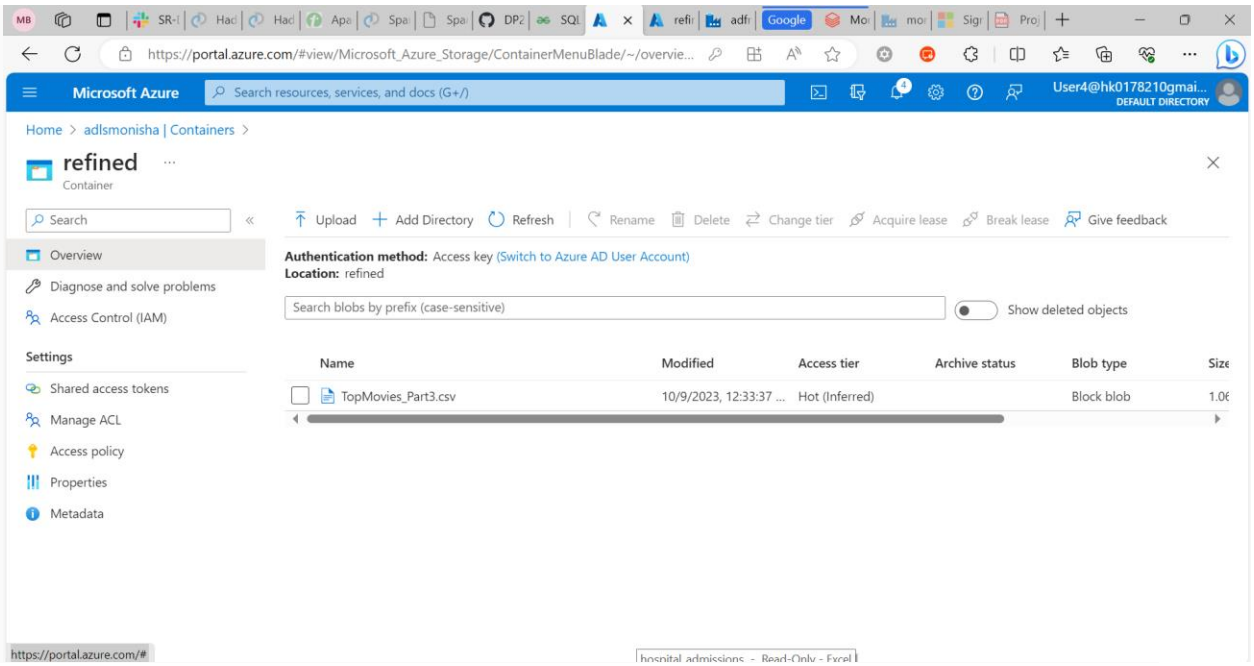
1) Go to sql database ➔ in query editor ➔ create table :

```
CREATE TABLE [dbo].[topmovies_comma] (
        [Index] [BIGINT] NULL,
 [MovieTitle] varchar(500) NULL
);
```



2) Go to datalake storage ➔ in refined container upload top_movies_part3.csv file ➔ upload





3) In sql data base ➔ connection strings ➔ jdbc

⇨ Give password we can connect

⇨ (OR)

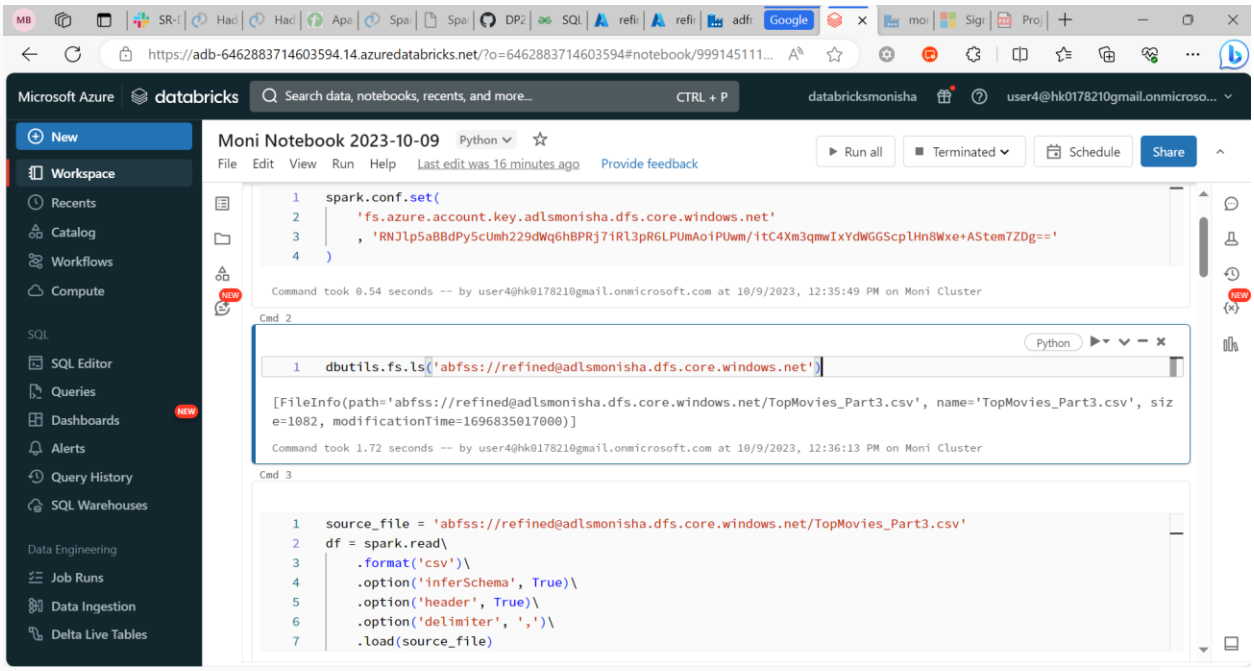
4) In data bricks notebook ➔ connect ➔ read file ➔

```
spark.conf.set(
    'fs.azure.account.key.adlsmonisha.dfs.core.windows.net'
    ,
'RNJlp5aBBdPy5cUmh229dWq6hBPRj7iRl3pR6LPUmAoiPUwm/itC4Xm3qmwIxYd
WGGScplHn8Wxe+AStem7ZDg=='
)
```

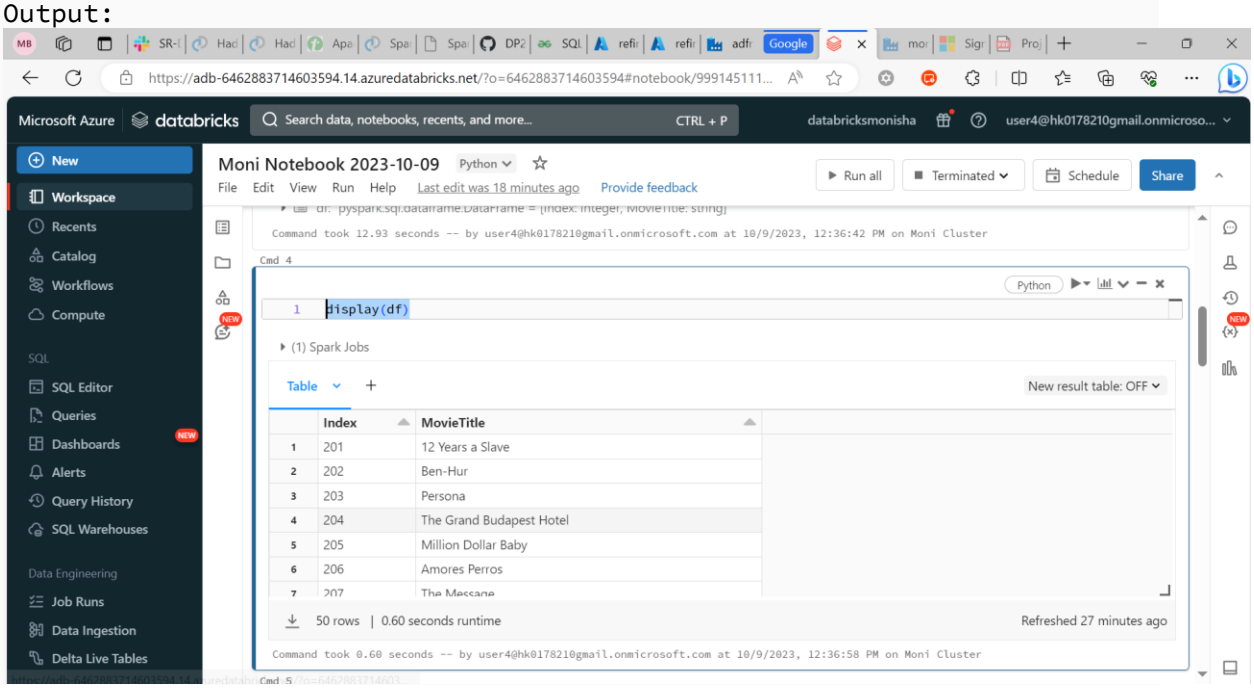
```
dbutils.fs.ls('abfss://refined@adlsmonisha.dfs.core.windows.net')
```

```
source_file =
'abfss://refined@adlsmonisha.dfs.core.windows.net/TopMovies_Part3.csv'
df = spark.read\
    .format('csv')\
    .option('inferSchema', True)\
    .option('header', True)\
    .option('delimiter', ',')\
    .load(source_file)
```

```
display(df)
```



Output:



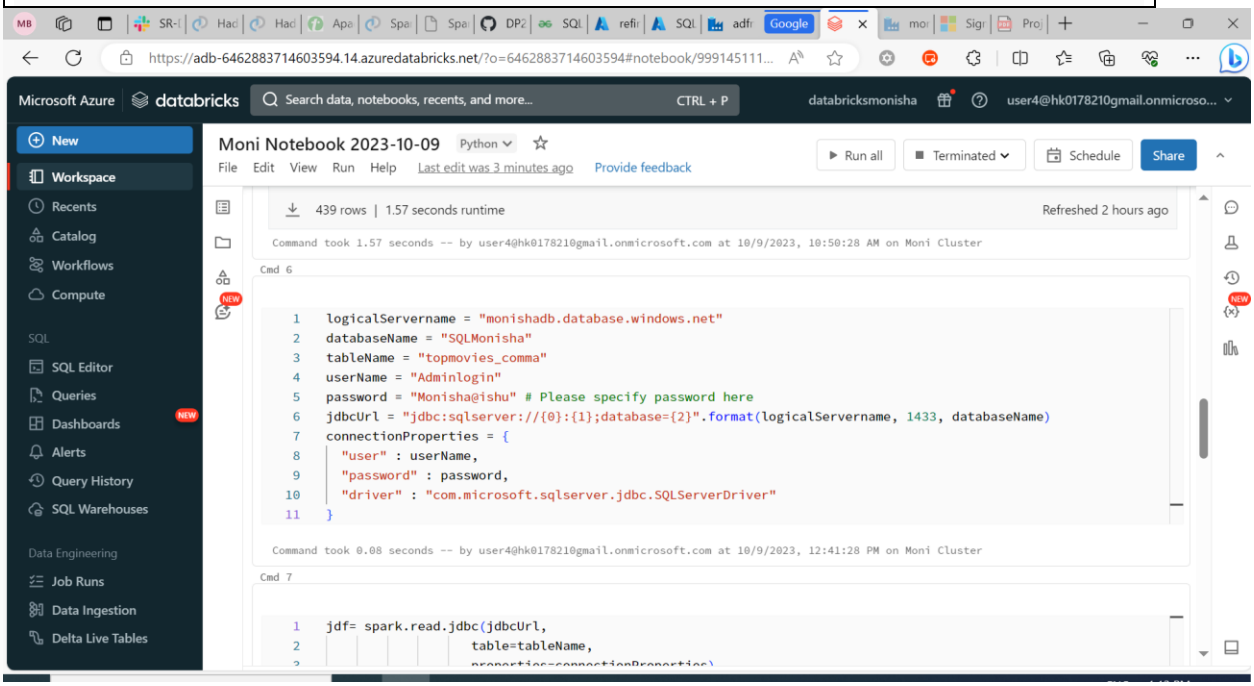
5) Go to data bricks note book ➔change logicalservername, databasename, tablename, username, password

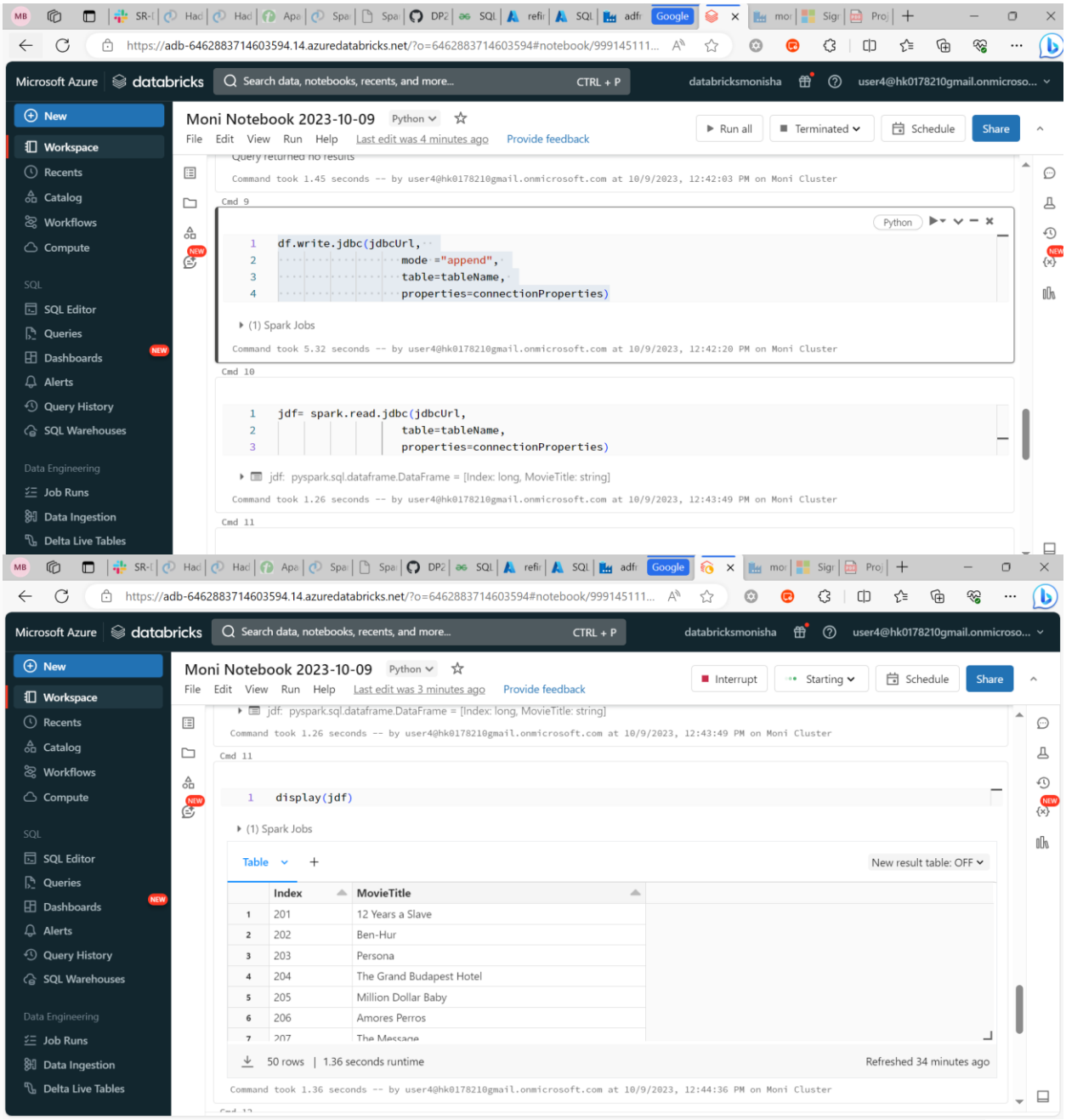
```
logicalServername = "monishadb.database.windows.net"
databaseName = "SQLMonisha"
tableName = "topmovies_comma"
userName = "Adminlogin"
password = "Monisha@ishu" # Please specify password here
```

```
jdbcUrl =
"jdbc:sqlserver://{0}:{1};database={2}".format(logicalServername
, 1433, databaseName)
connectionProperties = {
  "user" : userName,
  "password" : password,
  "driver" : "com.microsoft.sqlserver.jdbc.SQLServerDriver"
}
```

```
TO READ THE FILE:
jdf= spark.read.jdbc(jdbcUrl,
                     table=tableName,
                     properties=connectionProperties)
```

```
TO WRITE THE FILE:
df.write.jdbc(jdbcUrl,
                     mode ="append",
                     table=tableName,
                     properties=connectionProperties
```





6)

```
# Details about connection string
logicalServername = "sqlservera2023.database.windows.net"
databaseName = "sqldb"
tableName = "CustomerSource"
userName = "Adminlogin"
password = "Welcome@123" # Please specify password here
jdbcUrl =
"jdbc:sqlserver://{0}:{1};database={2}".format(logicalServername, 1433,
databaseName)
```

```
connectionProperties = {
  "user" : userName,
  "password" : password,
  "driver" : "com.microsoft.sqlserver.jdbc.SQLServerDriver"
}


df.write.jdbc(jdbcUrl,
                    mode ="append",
                    table=tableName,
                    properties=connectionProperties)


jdf= spark.read.jdbc(jdbcUrl,
                    table=tableName,
                    properties=connectionProperties)
```

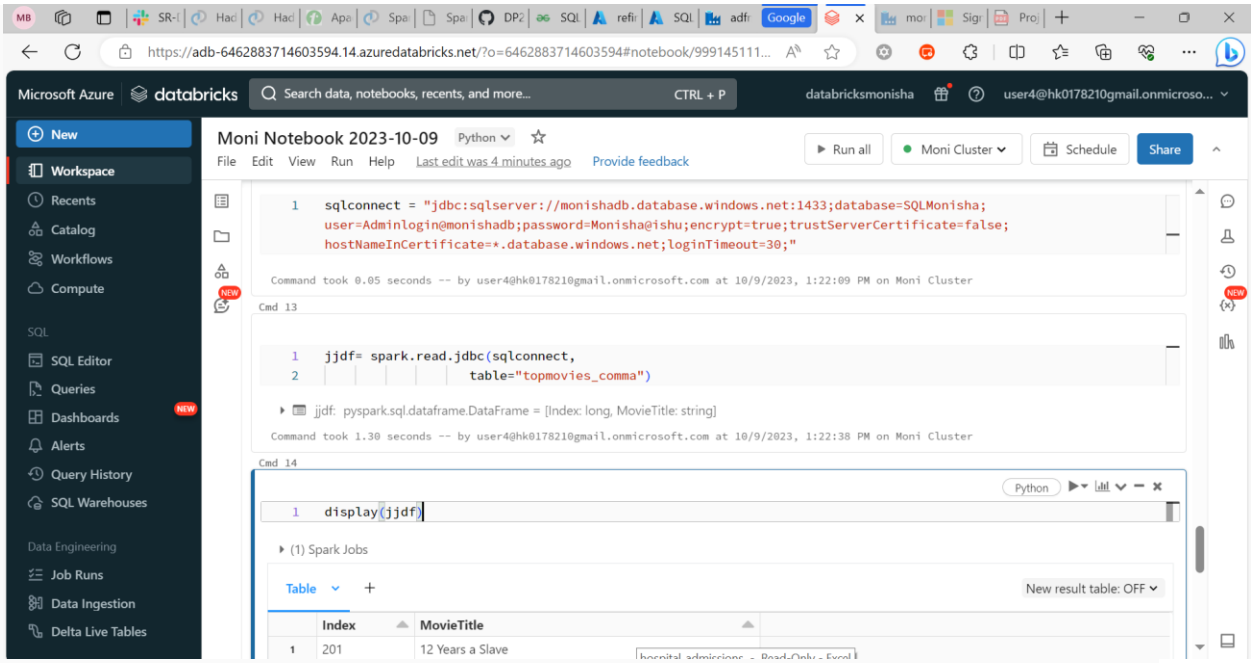
ANOTHER WAY TO CONNECT SQL:

1) Go to SQL Database ➔ Connection strings ➔ jdbc ➔ COPY THAT url
2) Go to databricks ➔ paste jdbc url and change password

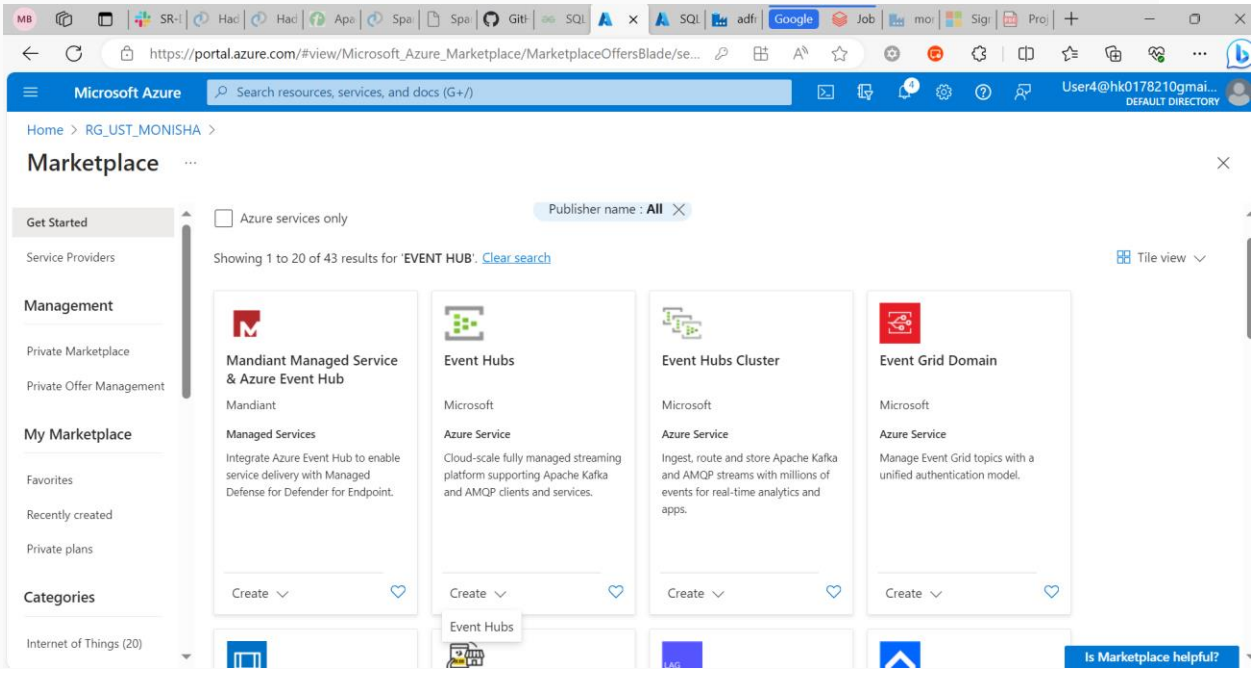
```
sqlconnect =
"jdbc:sqlserver://monishadb.database.windows.net:1433;data
base=SQLMonisha;user=Adminlogin@monishadb;password=Monisha
@ishu;encrypt=true;trustServerCertificate=false;hostNameIn
Certificate=*.database.windows.net;loginTimeout=30;"
```

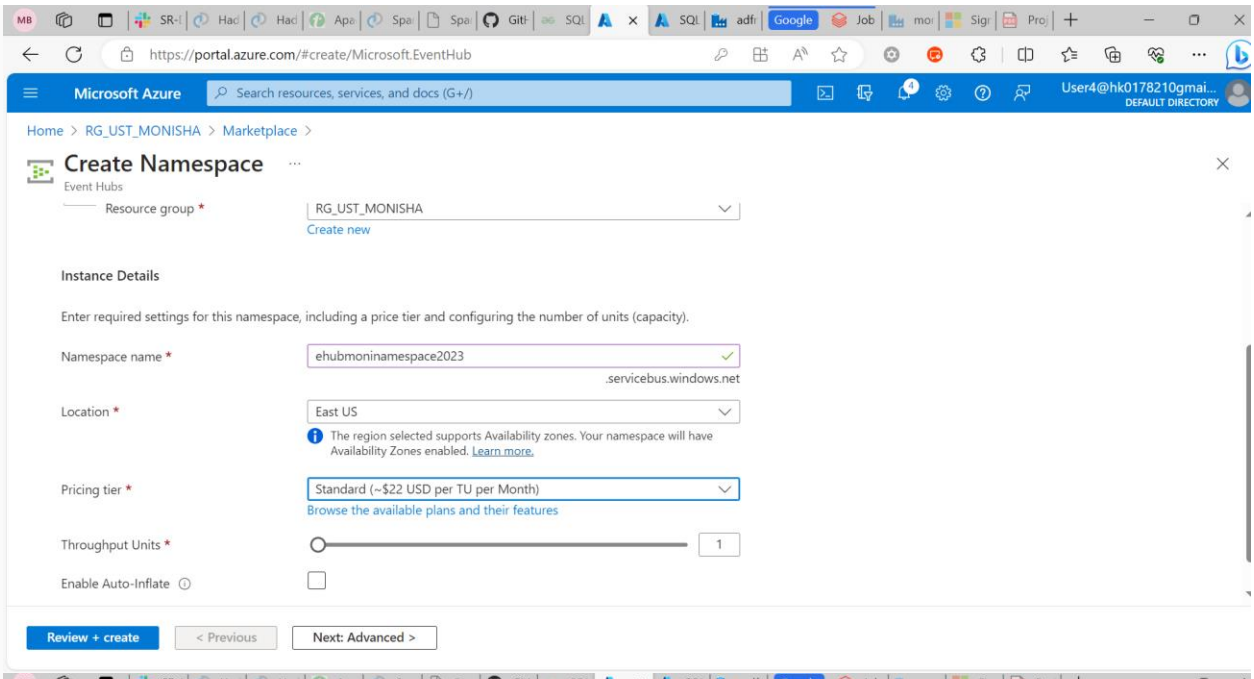
3) READ THE FILE : CHANGE variable name as sqlconnect and table name can be anything.

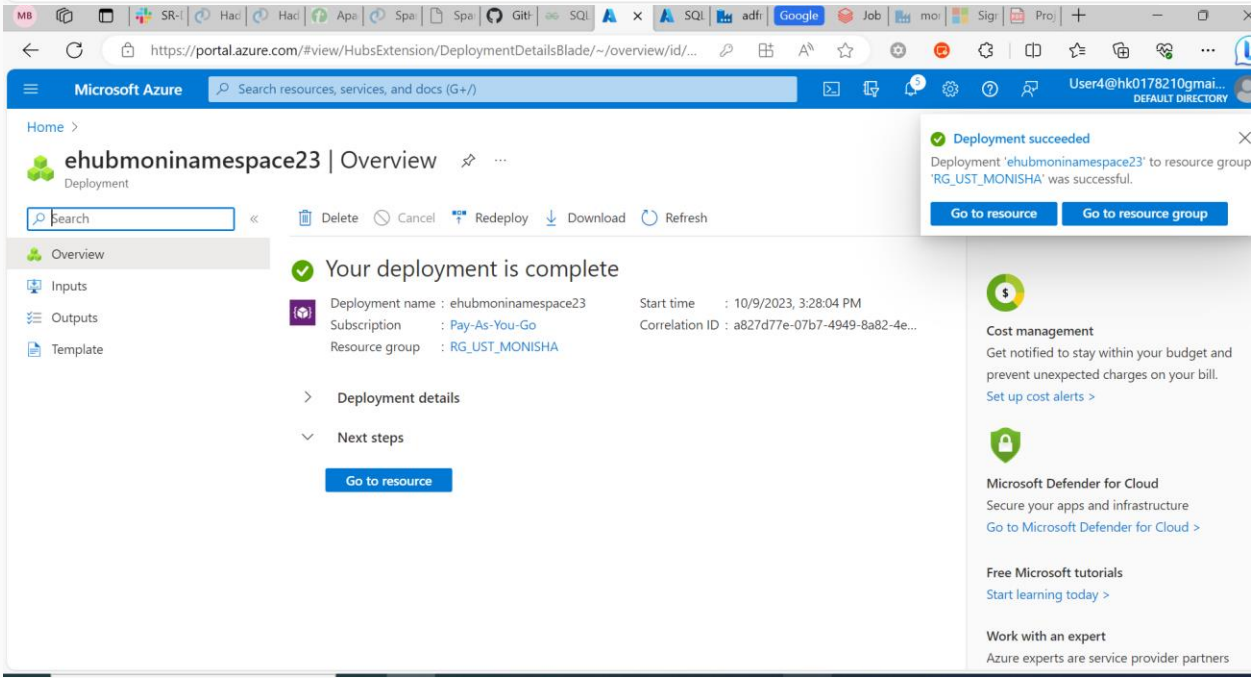
```
jjdf= spark.read.jdbc(sqlconnect,
                      table="topmovies_comma")
```

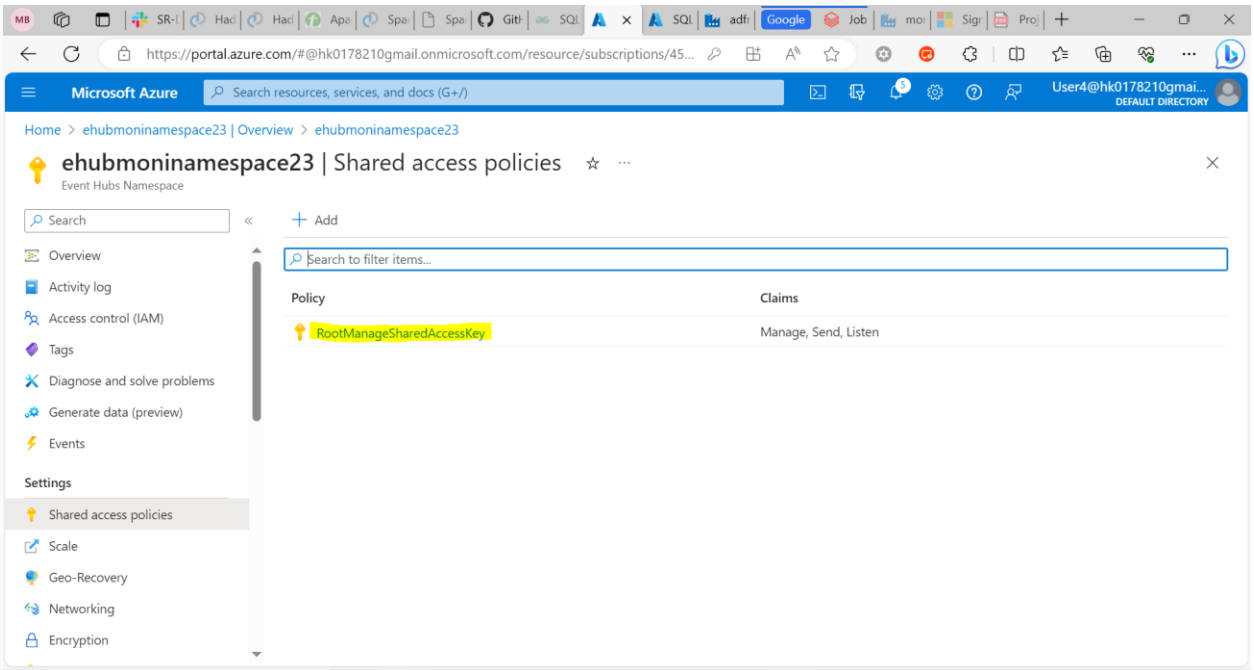


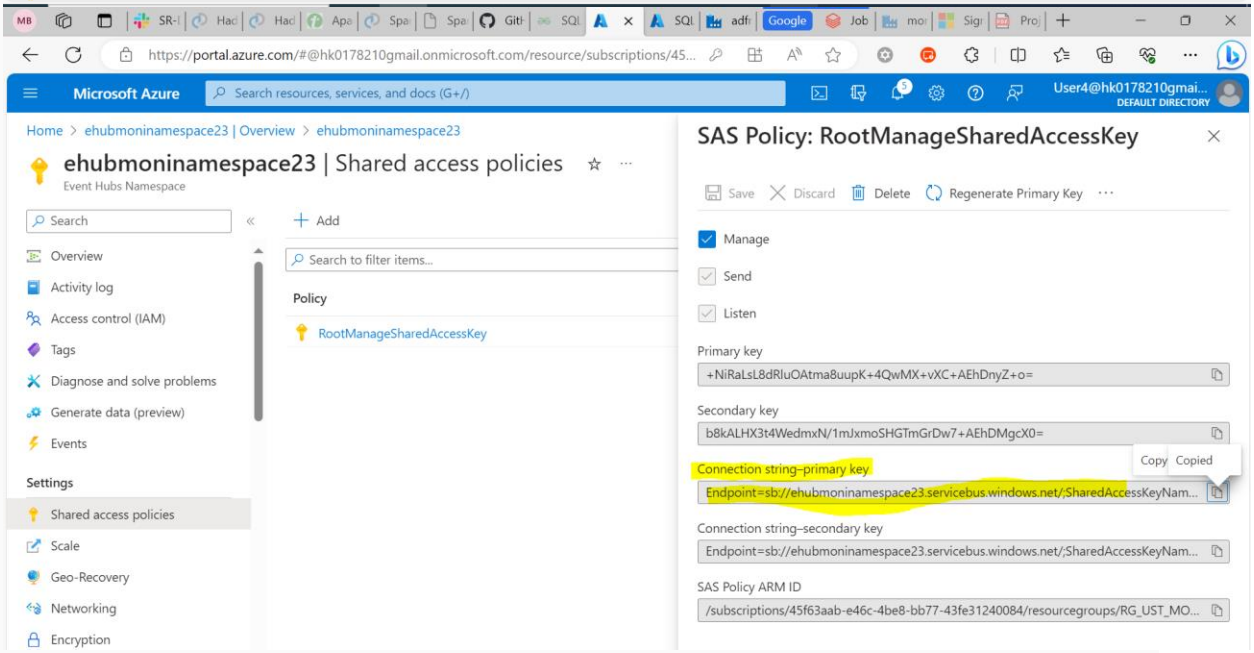
HOW TO CREATE EVENT HUB NAMESPACE:




https://portal.azure.com/#create/Microsoft.EventHub
Microsoft Azure
Search resources, services, and docs (G+/)
User4@hk0178210gmai...
DEFAULT DIRECTORY
Home > RG_UST_MONISHA > Marketplace >
Create Namespace
Event Hubs
Resource group *
RG_UST_MONISHA
Create new
Instance Details
Enter required settings for this namespace, including a price tier and configuring the number of units (capacity).
Namespace name *
ehubmoninamespace2023
.servicebus.windows.net
Location *
East US
The region selected supports Availability zones. Your namespace will have Availability Zones enabled. Learn more.
Pricing tier *
Standard (~$22 USD per TU per Month)
Browse the available plans and their features
Throughput Units *
1
Enable Auto-Inflate
Review + create
< Previous
Next: Advanced >


https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/...
Microsoft Azure
Search resources, services, and docs (G+/)
User4@hk0178210gmai...
DEFAULT DIRECTORY
Home >
ehubmoninamespace23 | Overview
Deployment
Search
Delete
Cancel
Redeploy
Download
Refresh
Overview
Inputs
Outputs
Template
Your deployment is complete
Deployment name : ehubmoninamespace23
Subscription : Pay-As-You-Go
Resource group : RG_UST_MONISHA
Start time : 10/9/2023, 3:28:04 PM
Correlation ID : a827d77e-07b7-4949-8a82-4e...
Deployment details
Next steps
Go to resource
Deployment succeeded
Deployment 'ehubmoninamespace23' to resource group 'RG_UST_MONISHA' was successful.
Go to resource
Go to resource group
Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >
Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >
Free Microsoft tutorials
Start learning today >
Work with an expert
Azure experts are service provider partners



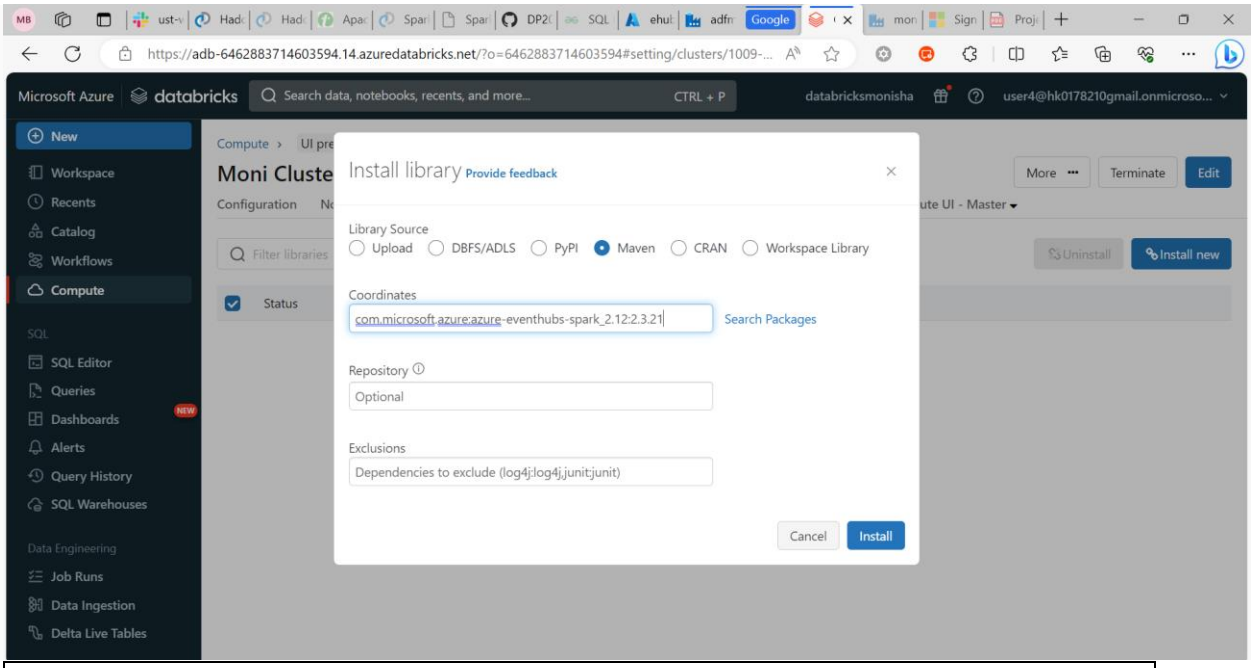


EVENT CONSUMER: CONNECT WITH EVENTHUB:

New ➔ notebook ➔ Name : ADB to Eventhub

Compute ➔ cluster ➔ libraries ➔ install new

com.microsoft.azure:azure-eventhubs-spark_2.12:2.3.21



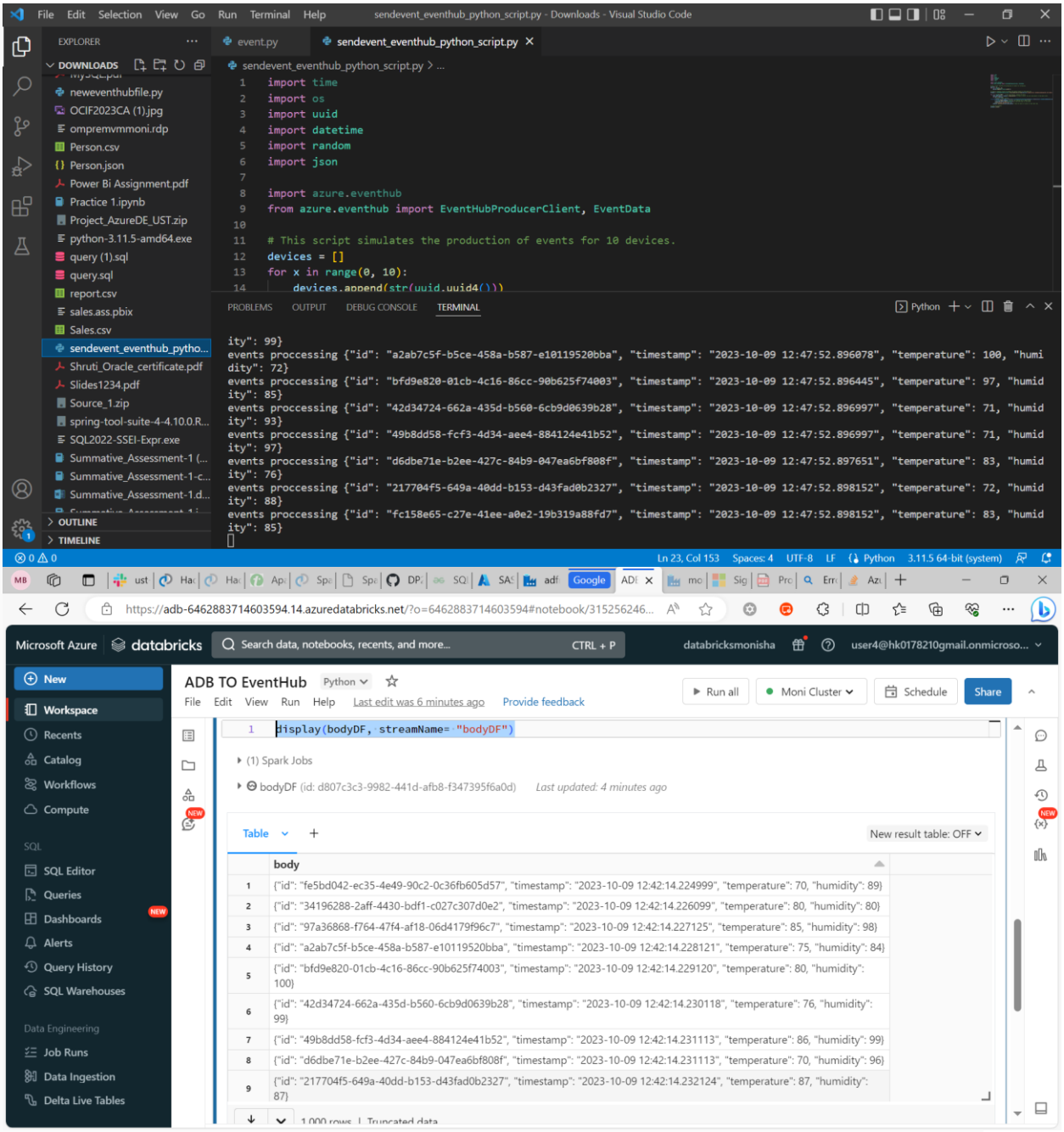

```
import pyspark
from pyspark.sql.functions import *
from pyspark.sql import *
from pyspark.sql import SparkSession
from pyspark.sql.functions import from_json
from pyspark.sql.types import StructType, StructField, StringType,
IntegerType
```

```
##connect with evethub
ehConf = {}
ehConf['eventhubs.connectionString'] =
sc._jvm.org.apache.spark.eventhubs.EventHubsUtils.encrypt("Endpoint=sb
://ehubmoninamespace23.servicebus.windows.net/;SharedAccessKeyName=Roo
tManageSharedAccessKey;SharedAccessKey=+NiRaLsL8dRluOAtma8uupK+4QwMX+v
XC+AEhDnyZ+o=;EntityPath=ehubmoni")
```

```
##read stream data

df_read_stream = (spark.readStream
                  .format("eventhubs")
                  .options(**ehConf)
                  .load())
```

```
bodyDF = df_read_stream.select(col("body").cast("STRING"))
```

```
display(bodyDF, streamName= "bodyDF")
```



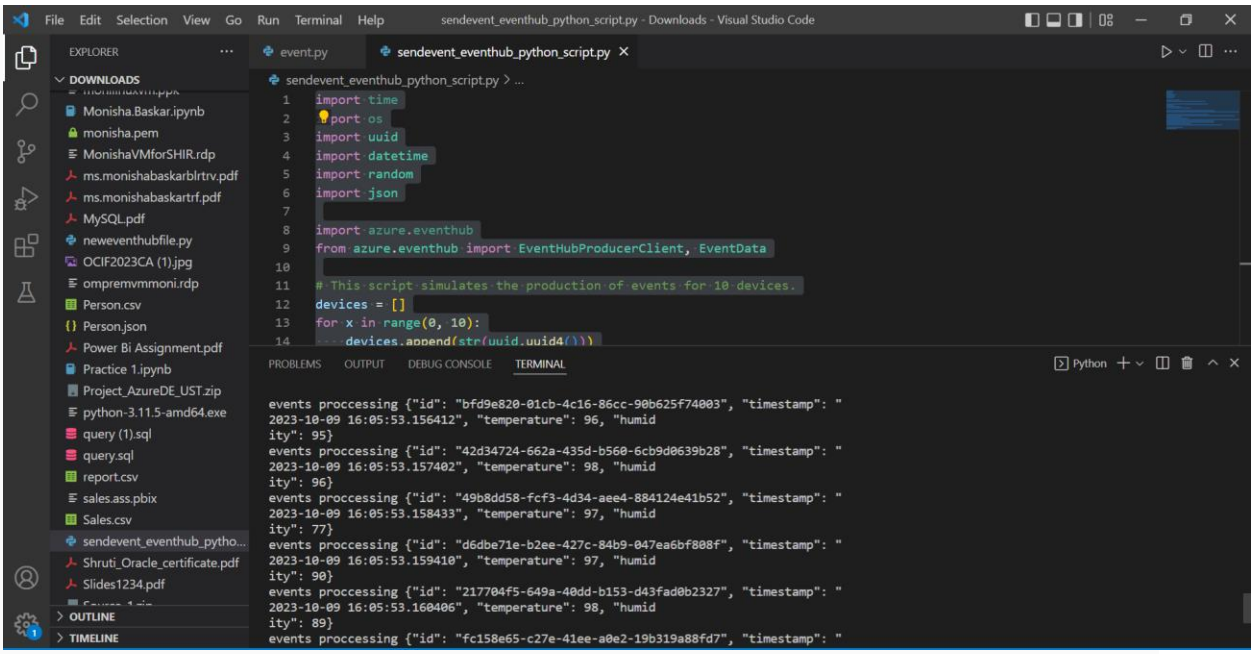
Cancel after getting 1000 rows

TO AUTOGENERATE SCRIPTS :

```
import time
import os
import uuid
import datetime
import random
import json

import azure.eventhub
from azure.eventhub import EventHubProducerClient, EventData

# This script simulates the production of events for 10 devices.
devices = []
for x in range(0, 10):
    devices.append(str(uuid.uuid4()))

# Create a producer client to produce and publish events to the event hub.
producer =
EventHubProducerClient.from_connection_string(conn_str="Endpoint=sb://ehubmoninam
espace23.servicebus.windows.net/;SharedAccessKeyName=RootManageSharedAccessKey;Sh
aredAccessKey=+NiRaLsL8dRluOAtma8uupK+4QwMX+vXC+AEhDnyZ+o=",
eventhub_name="ehubmoni")

for y in range(0,2000):    # For each device, produce 20 events.
    event_data_batch = producer.create_batch() # Create a batch. You will add
events to the batch later.
    for dev in devices:
        # Create a dummy reading.
        reading = {'id': dev, 'timestamp': str(datetime.datetime.utcnow()),
'temperature': random.randint(70, 100), 'humidity': random.randint(70, 100)}
        s = json.dumps(reading) # Convert the reading into a JSON string.
        event_data_batch.add(EventData(s)) # Add event data to the batch.
        print("events proccessing", s)
    producer.send_batch(event_data_batch) # Send the batch of events to the event
hub.

# Close the producer.
producer.close()
```

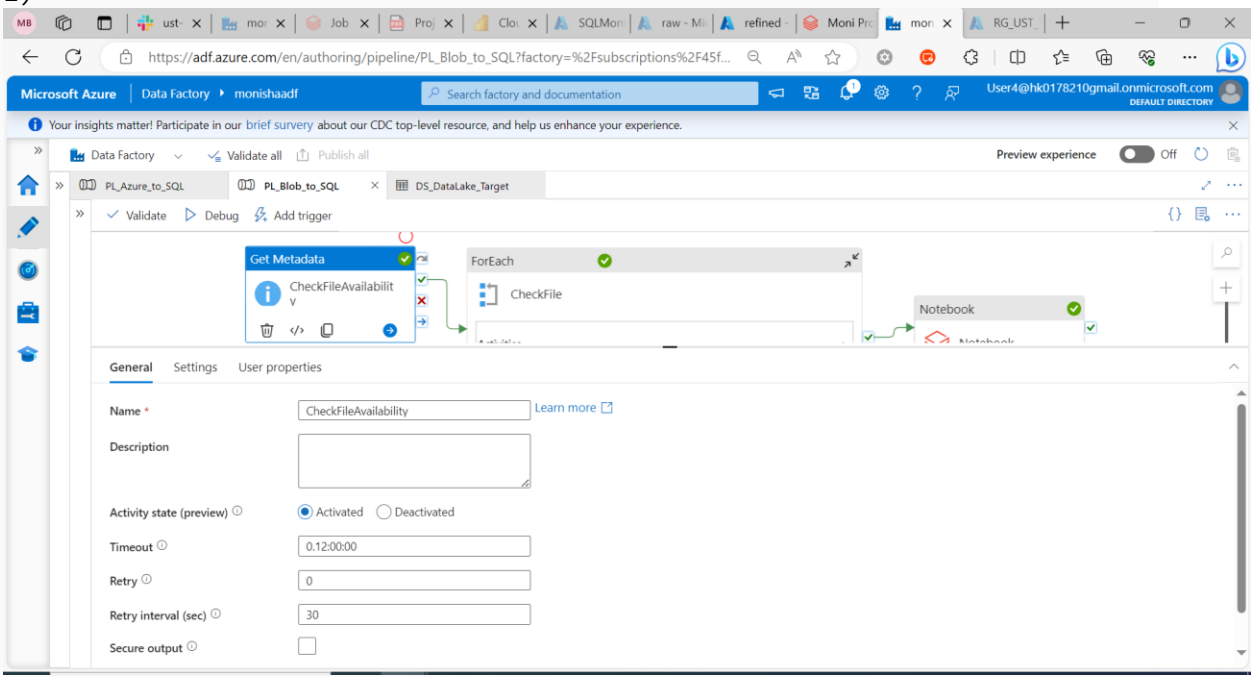
In visual studio code: right click on the file ➔ click run python file in terminal

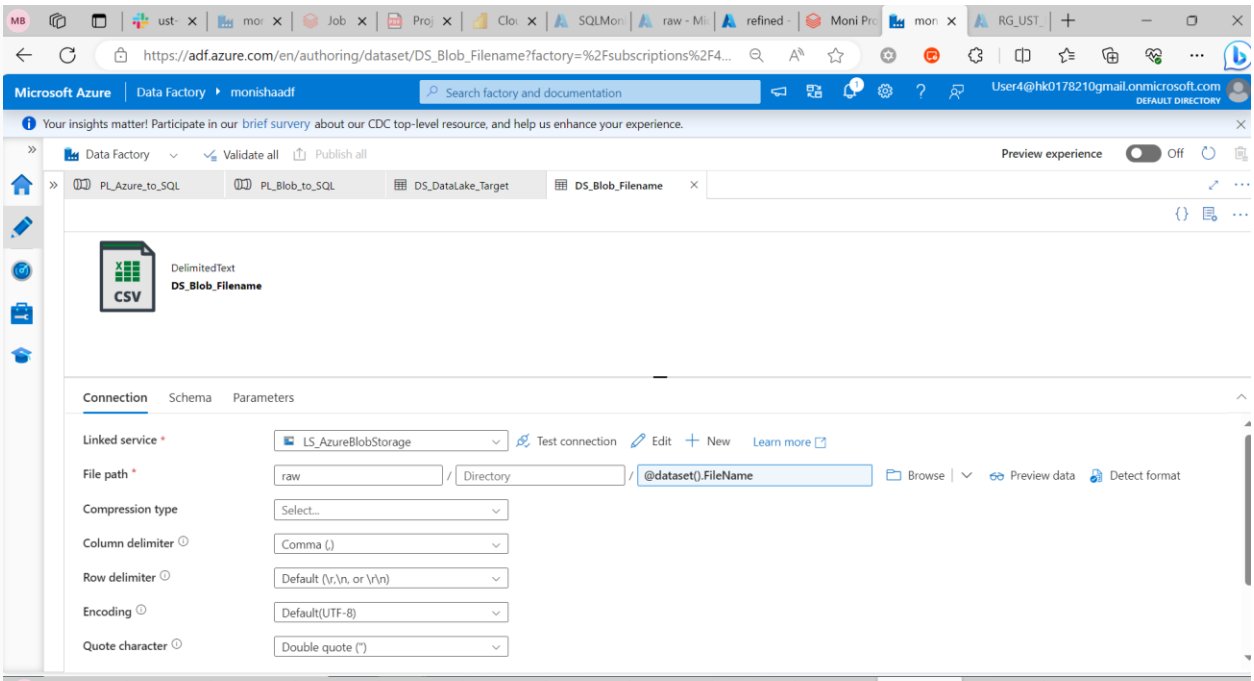
In terminal, cd to downloads ➔ cmd(command prompt) ➔ python file_name.py(python sendevent_eventhub_python_script.py)

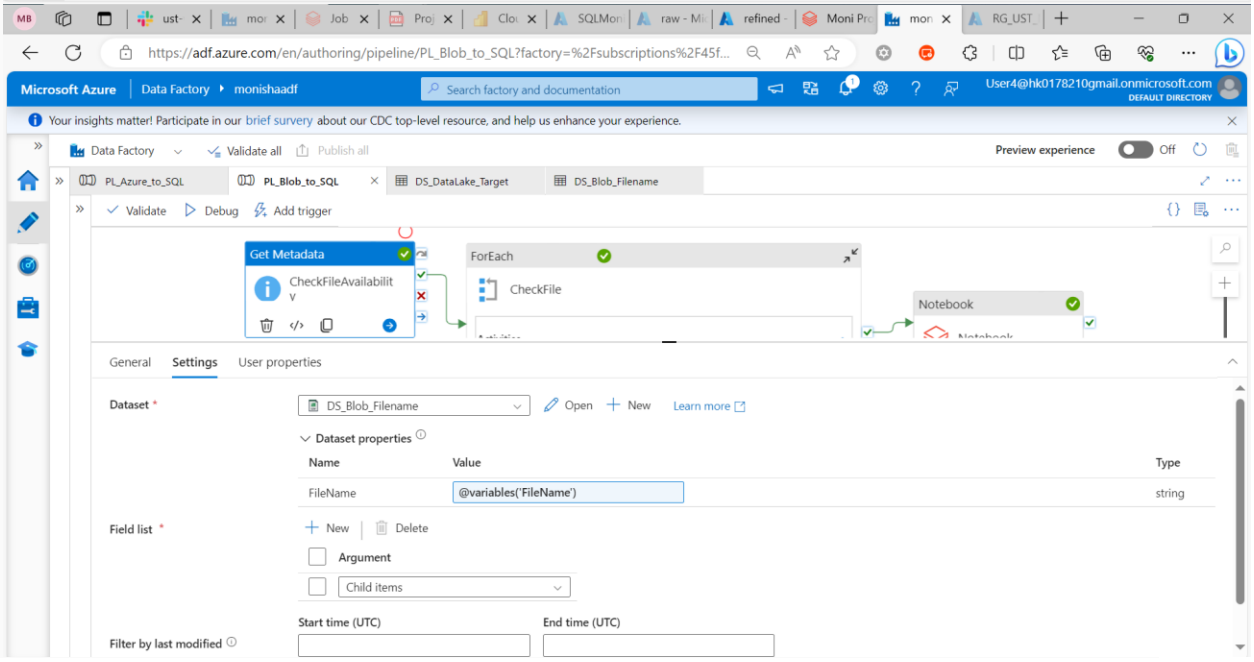


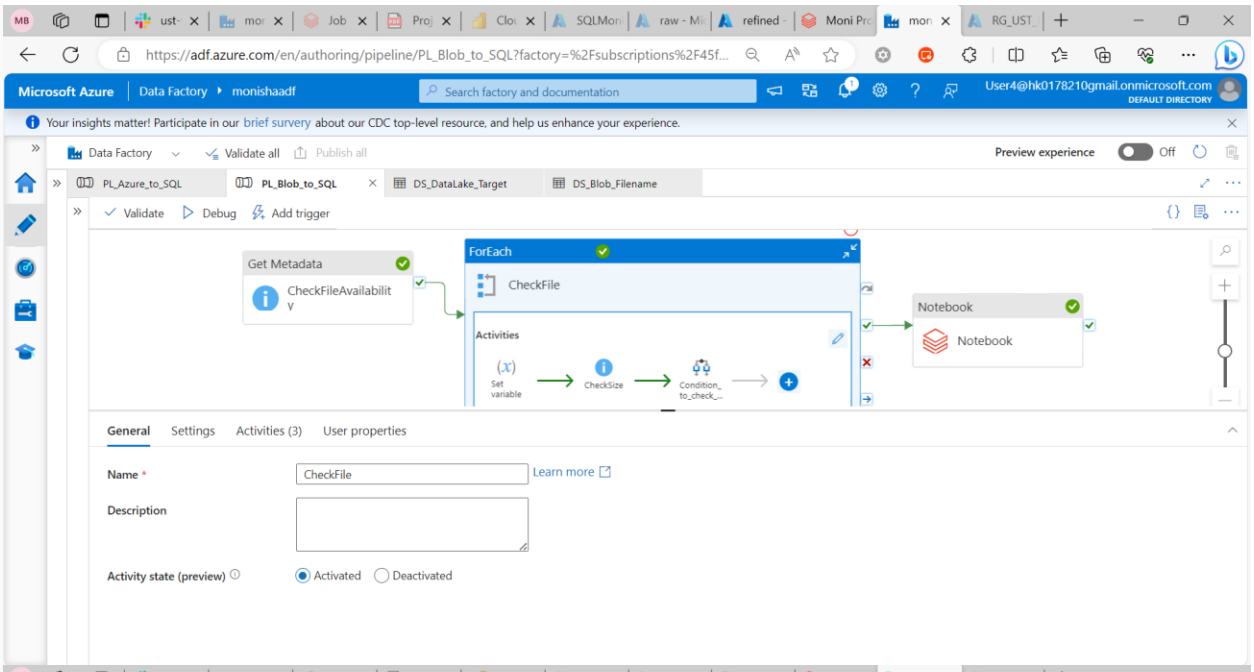
ASSIGNMENT:

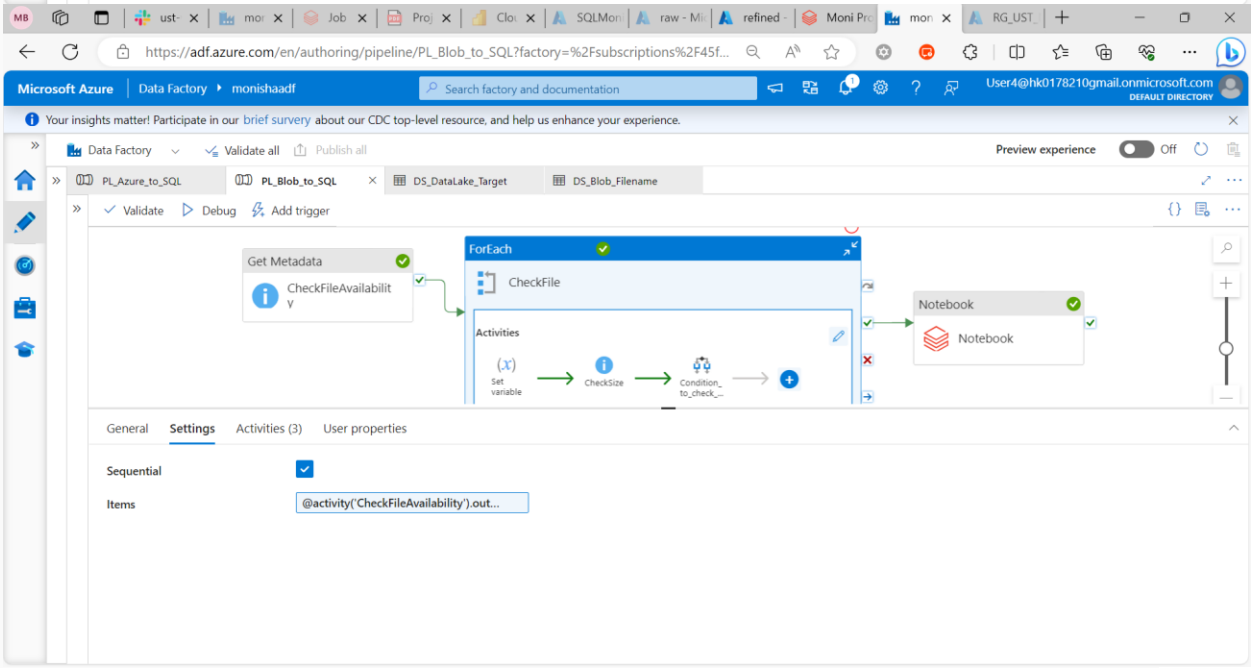
1)

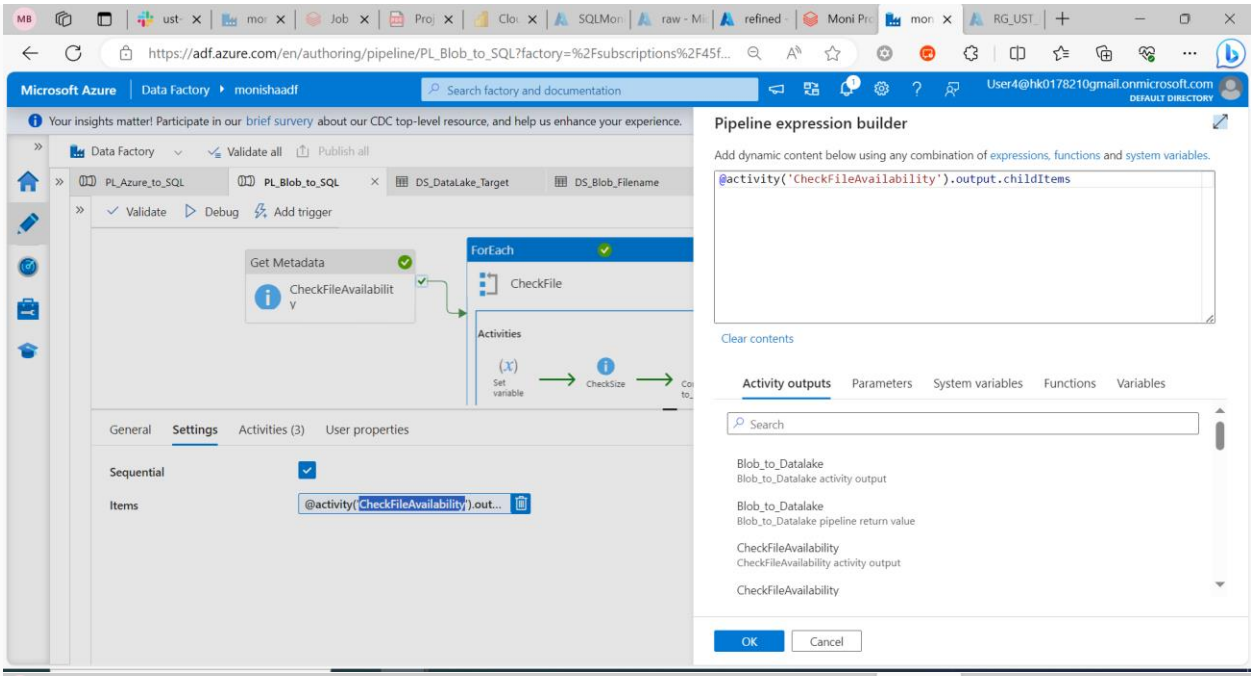


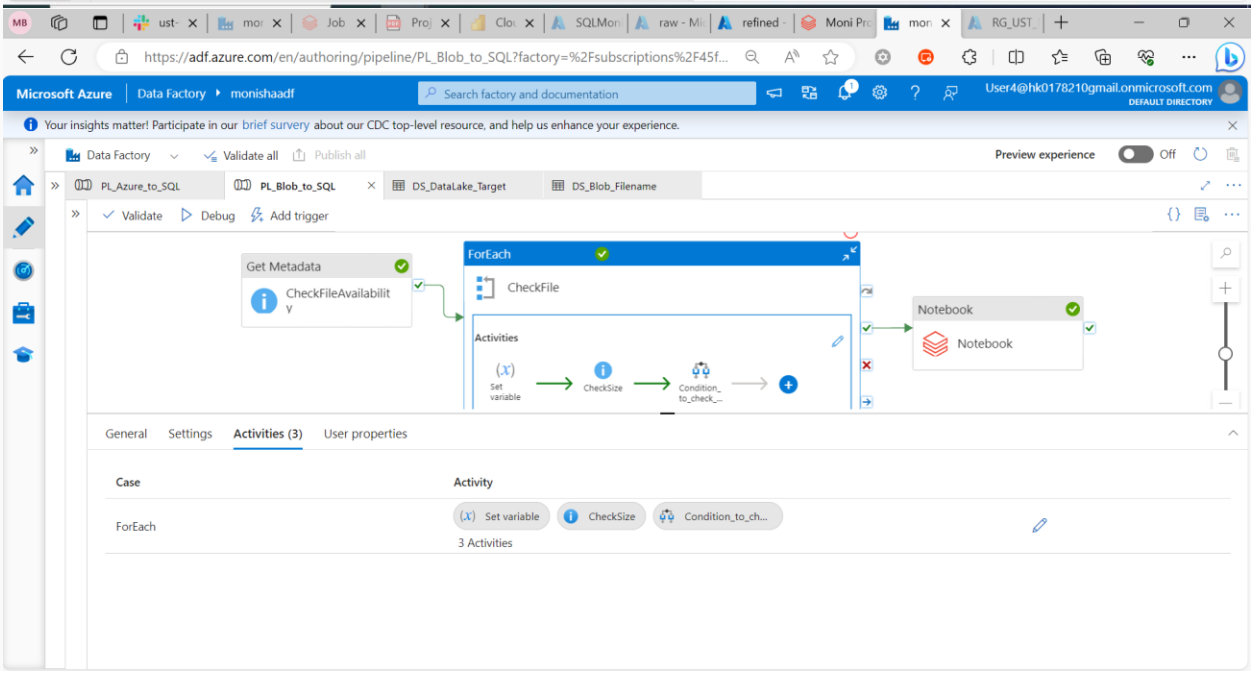

Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
DelimitedText
DS_Blob_Filename
Connection
Schema
Parameters
Linked service *
LS_AzureBlobStorage
Test connection
Edit
New
Learn more
File path *
raw
Directory
@dataset().FileName
Browse
Preview data
Detect format
Compression type
Select...
Column delimiter
Comma (,)
Row delimiter
Default (\r,\n, or \r\n)
Encoding
Default(UTF-8)
Quote character
Double quote (")

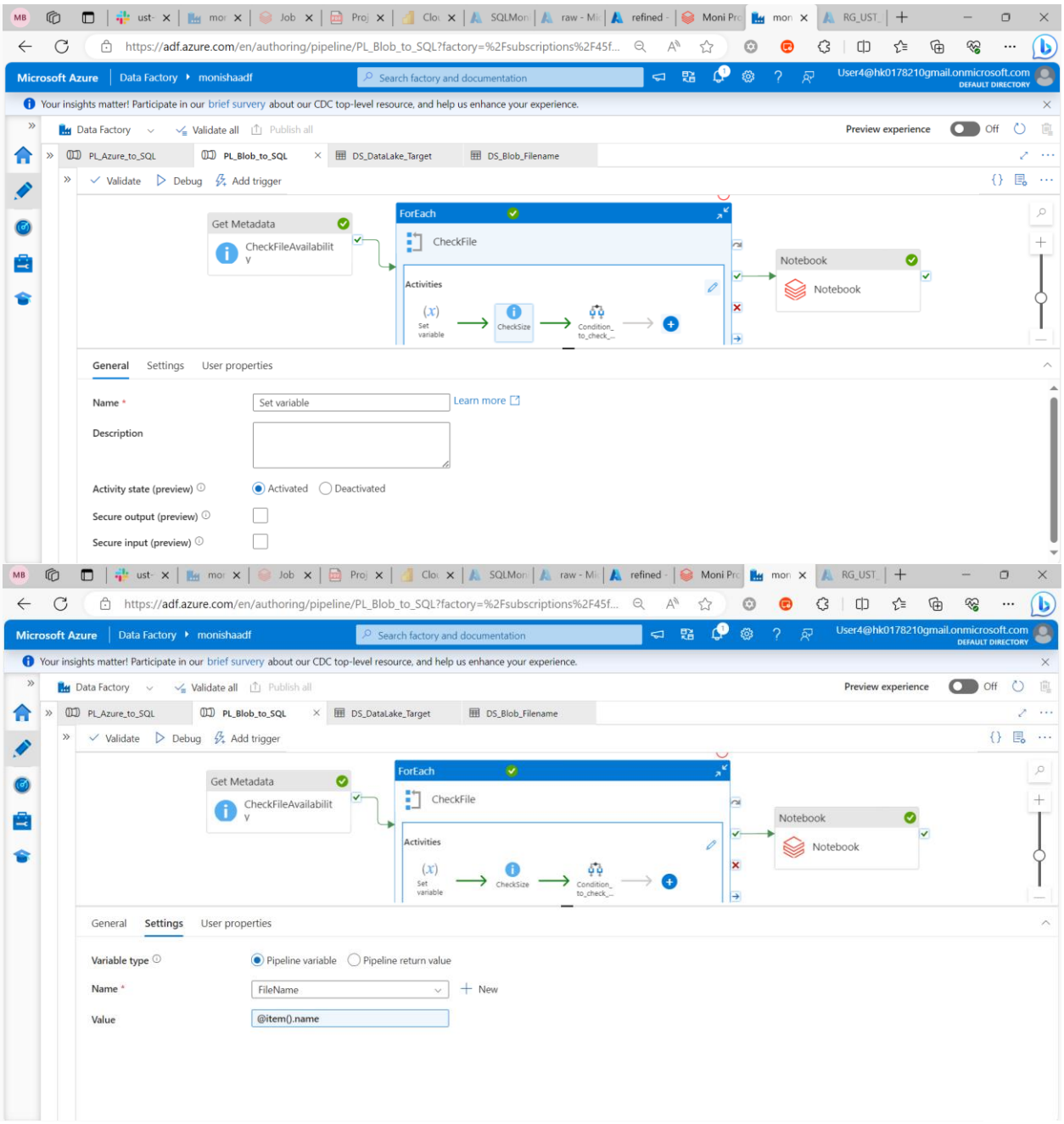

Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Notebook
General
Settings
User properties
Dataset *
DS_Blob_Filename
Open
New
Learn more
Dataset properties
Name
Value
Type
FileName
@variables('FileName')
string
Field list *
New
Delete
Argument
Child items
Start time (UTC)
End time (UTC)
Filter by last modified


https://adf.azure.com/en/authoring/pipeline/PL_Blob_to_SQL?factory=%2Fsubscriptions%2F45f...
Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survey about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Activities
Set
variable
CheckSize
Condition_
to_check_...
Notebook
Notebook
General
Settings
Activities (3)
User properties
Name *
CheckFile
Learn more
Description
Activity state (preview)
Activated
Deactivated

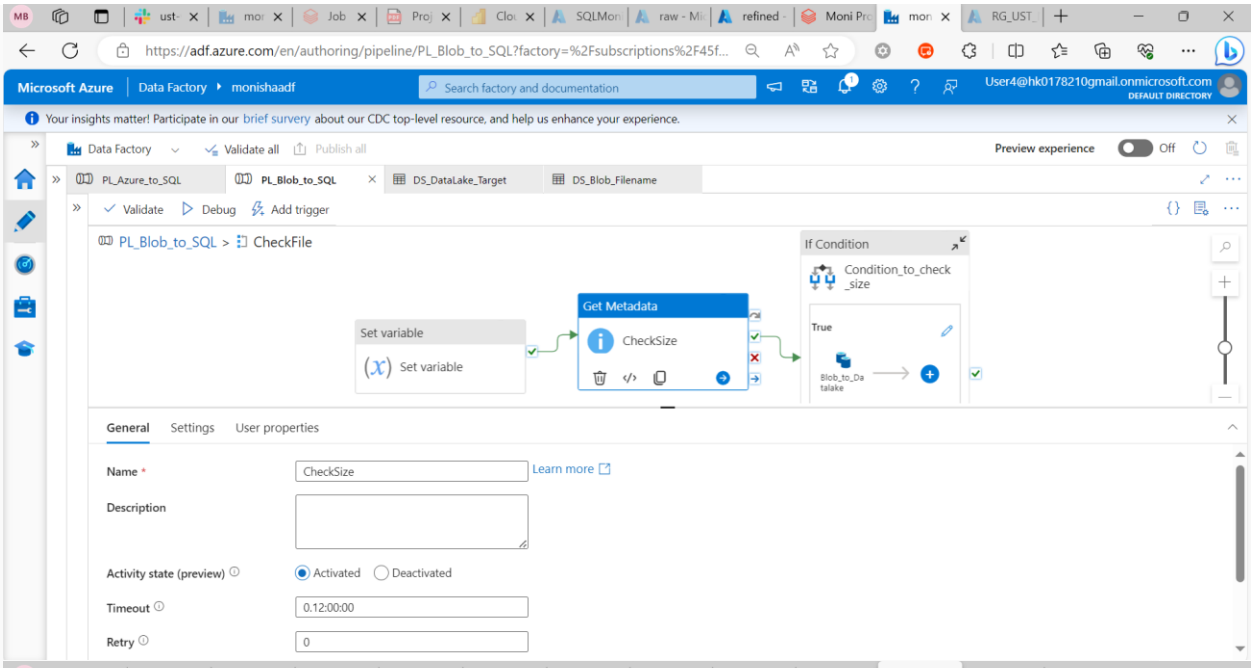

https://adf.azure.com/en/authoring/pipeline/PL_Blob_to_SQL?factory=%2Fsubscriptions%2F45f...
Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survey about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Activities
Set
variable
CheckSize
Condition_
to_check_...
Notebook
Notebook
General
Settings
Activities (3)
User properties
Sequential
Items
@activity('CheckFileAvailability').out...

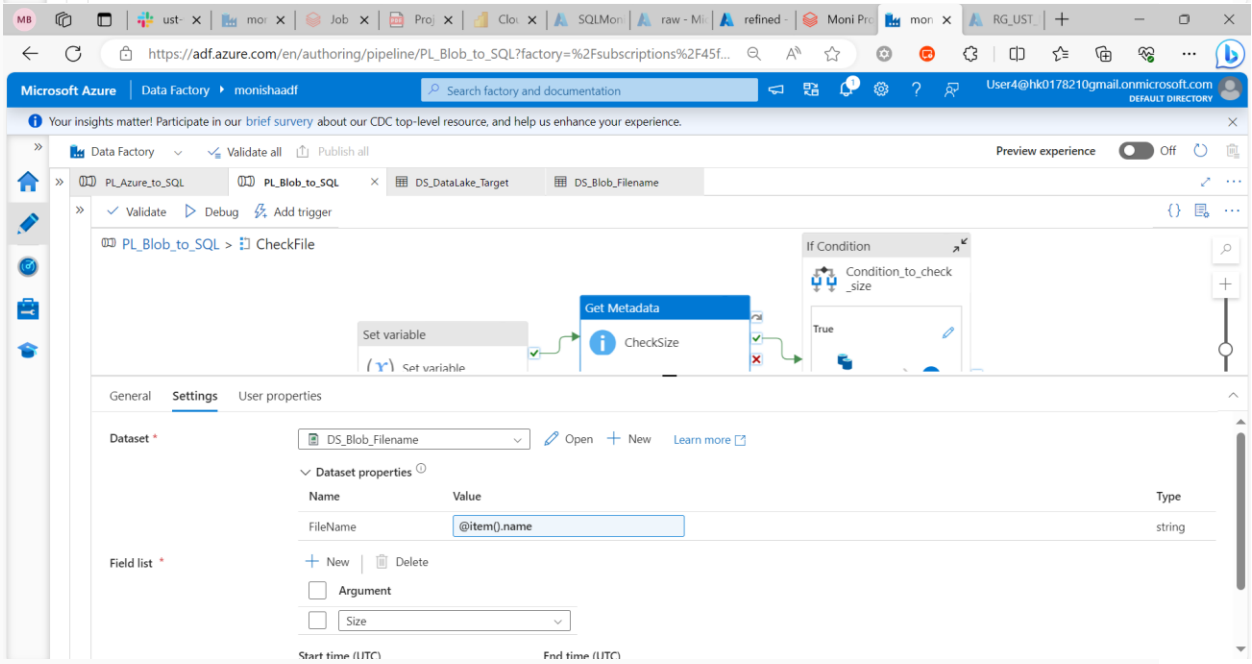

Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Activities
Set variable
CheckSize
General
Settings
Activities (3)
User properties
Sequential
Items
@activity('CheckFileAvailability').out...
Pipeline expression builder
Add dynamic content below using any combination of expressions, functions and system variables.
@activity('CheckFileAvailability').output.childItems
Clear contents
Activity outputs
Parameters
System variables
Functions
Variables
Search
Blob_to_Datalake
Blob_to_Datalake activity output
Blob_to_Datalake
Blob_to_Datalake pipeline return value
CheckFileAvailability
CheckFileAvailability activity output
CheckFileAvailability
OK
Cancel


Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Activities
Set variable
CheckSize
Condition_
to_check_...
Notebook
Notebook
General
Settings
Activities (3)
User properties
Case
Activity
ForEach
Set variable
CheckSize
Condition_to_ch...
3 Activities

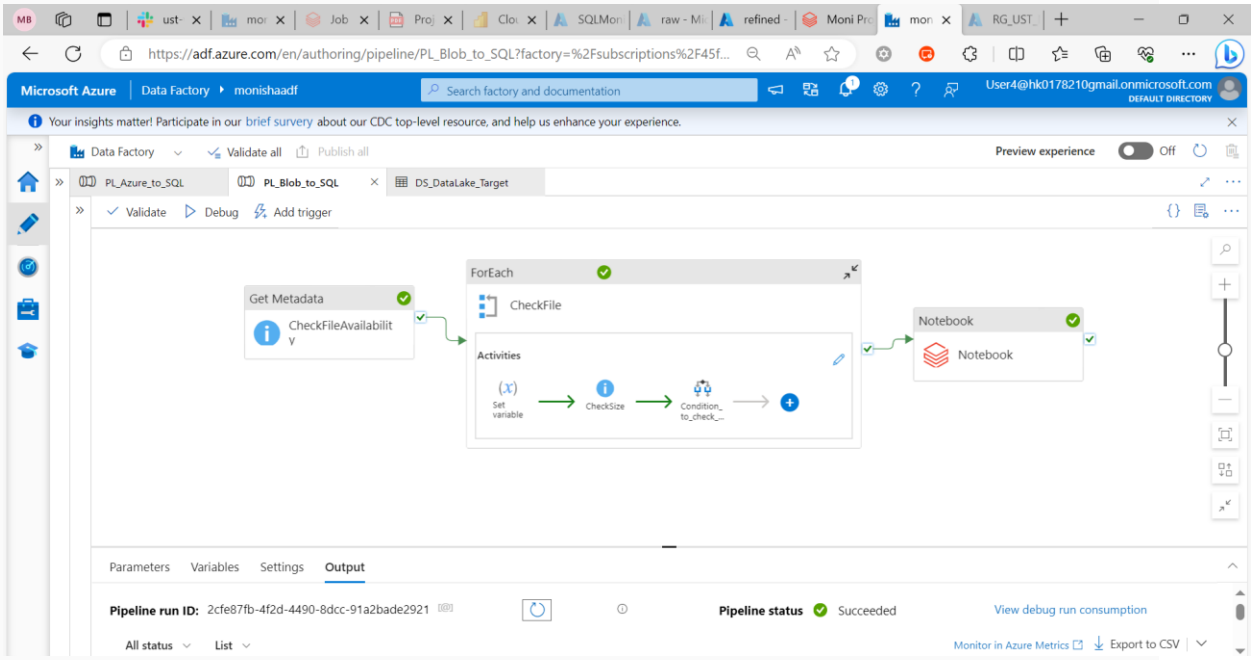

Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Activities
Set
variable
CheckSize
Condition_
to_check_...
Notebook
Notebook
General
Settings
User properties
Name *
Set variable
Learn more
Description
Activity state (preview)
Activated
Deactivated
Secure output (preview)
Secure input (preview)

Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
Get Metadata
CheckFileAvailabilit
y
ForEach
CheckFile
Activities
Set
variable
CheckSize
Condition_
to_check_...
Notebook
Notebook
General
Settings
User properties
Variable type
Pipeline variable
Pipeline return value
Name *
FileName
New
Value
@item().name


Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
PL_Blob_to_SQL > CheckFile
If Condition
Condition_to_check_size
Set variable
Get Metadata
CheckSize
True
General
Settings
User properties
Name *
CheckSize
Learn more
Description
Activity state (preview)
Activated
Deactivated
Timeout
0.12:00:00
Retry
0


Microsoft Azure
Data Factory
monishaadf
Search factory and documentation
User4@hk0178210gmail.onmicrosoft.com
DEFAULT DIRECTORY
Your insights matter! Participate in our brief survery about our CDC top-level resource, and help us enhance your experience.
Data Factory
Validate all
Publish all
Preview experience
Off
PL_Azure_to_SQL
PL_Blob_to_SQL
DS_DataLake_Target
DS_Blob_Filename
Validate
Debug
Add trigger
PL_Blob_to_SQL > CheckFile
If Condition
Condition_to_check_size
Set variable
Get Metadata
CheckSize
True
General
Settings
User properties
Dataset *
DS_Blob_Filename
Open
New
Learn more
Dataset properties
Name
Value
Type
FileName
@item().name
string
Field list *
New
Delete
Argument
Size
Start time (UTC)
End time (UTC)

M



IN DATAFACTORY ➔ MANAGE ➔ ARM TEMPLATE ➔ EXPORT

Resource group level also we can take backup

AFTER THAT EXTRACT ➔ BUILD UR TEMPLATE ➔ ARM ➔ LOAD FILE ➔ OPEN ➔ SAVE

Custom deployment:

(IT will not save password or credentials we need to provide)

MANAGE ➔ ARM TEMPLATE ➔ EXPORT ARM TEMPLATE