# PRODUCT SALES ANALYSIS

EMP ID:2320097
NAME: ADINA JOSHI SATYA VARDAN
Cohort Id: CSDAIA24AZ003

# Problem Statement:

- Fetch data from source URL & load in raw layer as zip file.
- Fetch files that has been arrived in last 24 hrs from source URL.
- Unzip files into refined container which is staging layer and
- Processed or clean data & load into ADLS processed container in parquet file as delta table.

# Requirements 1 - Data Preparation (using ADF):

**1. Fetching and Loading Raw Data:**

- The source data will be fetched from the provided URL.
- The fetched data will be loaded into the raw container in ADLS in zip file format.

**2. Unzipping and Loading Data into Staging:**

- Unzip the files that have arrived in the last 24 hours from the source in raw container.
- Load the unzipped files into the ADLS refine container in the path: refine/<participant_name>/Unzippeddata/file.

**3. Processing and Loading Clean Data:**

- Process or clean the data from the unzipped files using Azure Databricks (ADB).
- Load the processed or clean data under the processed container using Azure Databricks in the form of parquet files.

# Requirements 2 - Data Transformation (using ADB):

**1. Accessing Data Lake using Secret Keys:**

- Access the data lake using secret keys by creating secrets in Key Vault.

**2. Adding Header to Raw File:**

- Add headers to the raw file "sourceorderheader" with the provided schema:

  SalesOrderID,RevisionNumber,OrderDate,DueDate,ShipDate,Status,OnlineOrderFlag,SalesOrderNumber,PurchaseOrderNumber,AccountNumber,CustomerID,ShipToAddressID,BillToAddressID,ShipMethod,CreditCardApprovalCode,SubTotal,TaxAmt,Freight,TotalDue,Comment,rowguid,ModifiedDate.

**3. Joining and Cleaning Data:**

- Join the "product" file and "salesorderdetails" file.
- Perform necessary cleaning:
  - Drop all entries where color is null.

- Drop all entries where size is null.
- Check the difference between StandardCost and ListPrice for each sales order.
- Replace product sizes: S -> 30, M -> 32, L -> 34, XL -> 36, XXL -> 38 to display available product groups by size.

**4. Creating Managed Delta Tables:**

- Create managed delta tables in Azure Databricks for performing analysis using SQL queries.

**5. Creating External Delta Tables:**

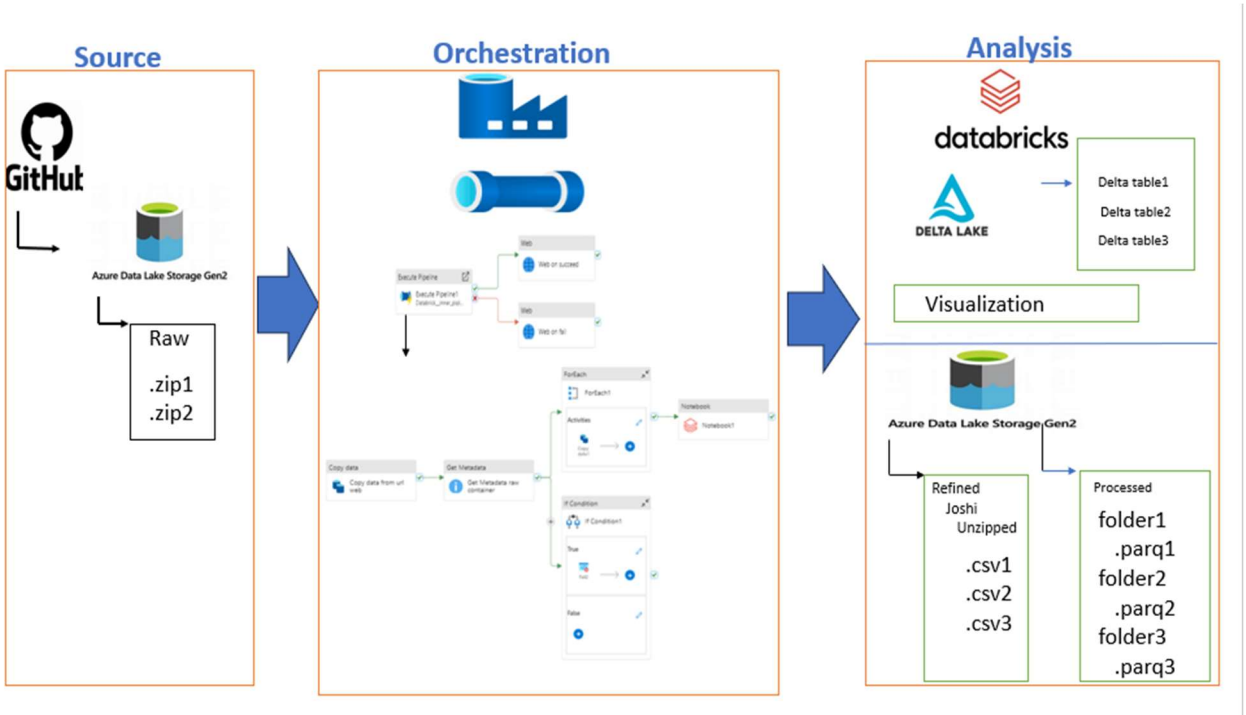
- Create external delta tables in Azure Databricks and load data in the form of parquet files into the processed data lake container.

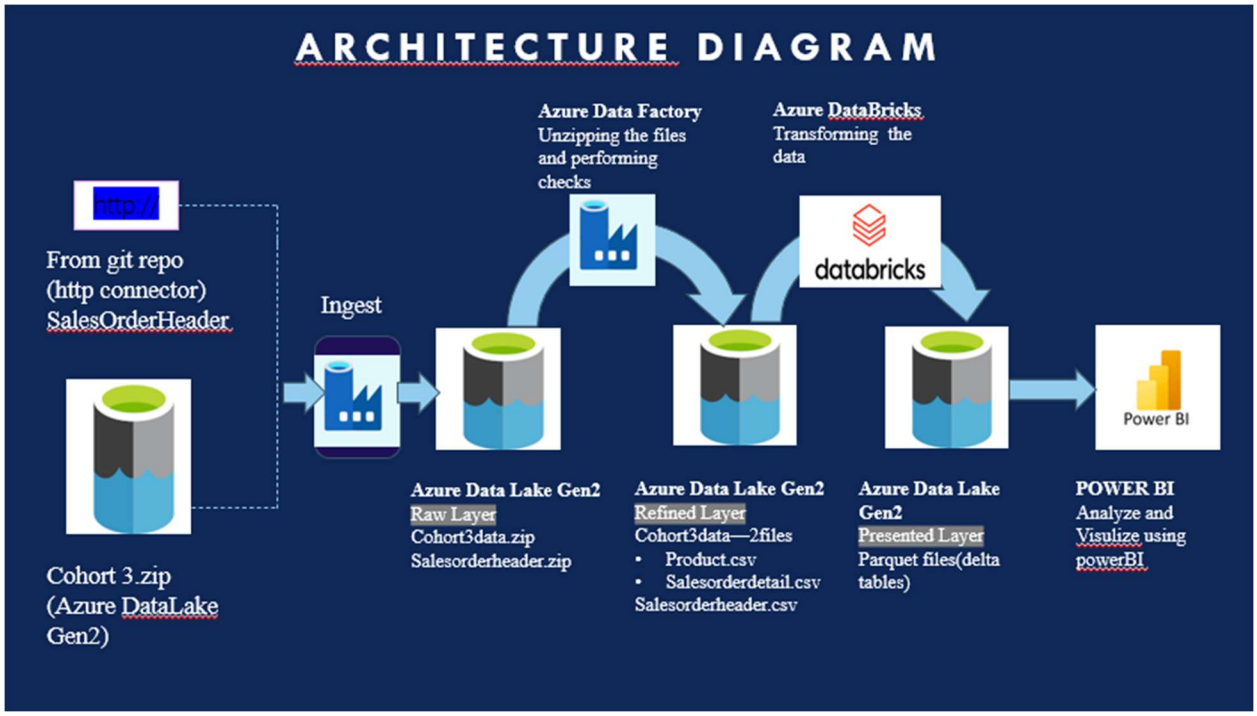
**6. Visualization and Analysis:**

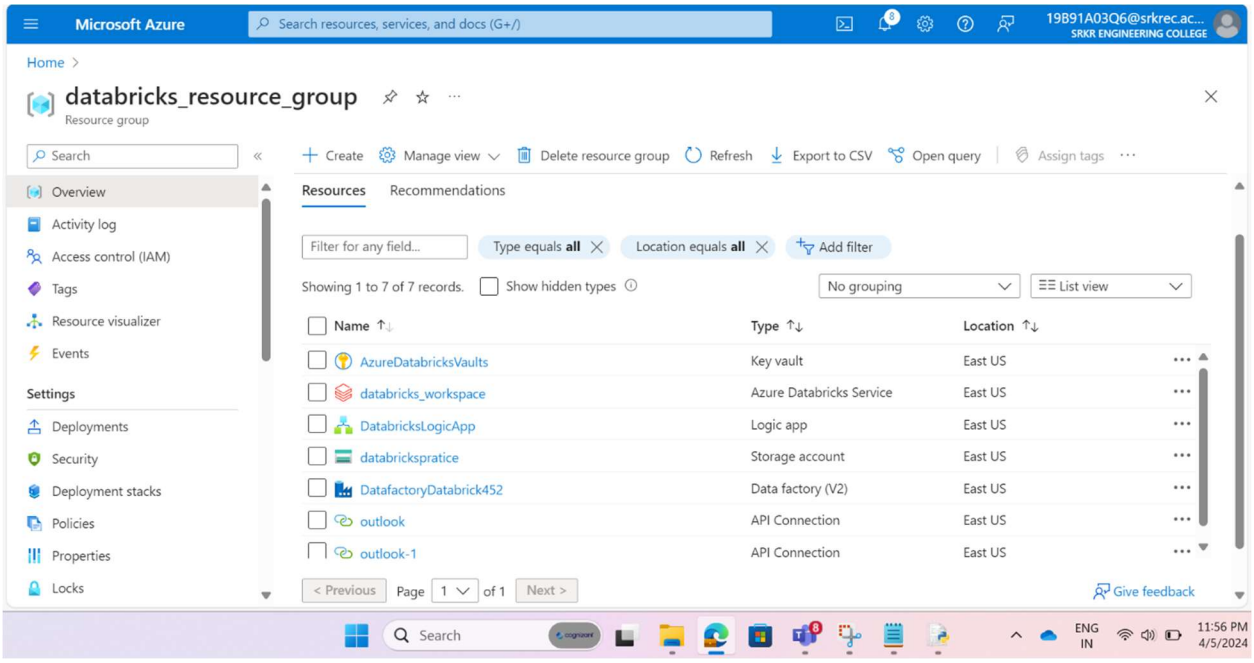
- Create visualizations on the cleaned data and generate graphs to understand product order sales details easily.

# Setup Email Configuration:

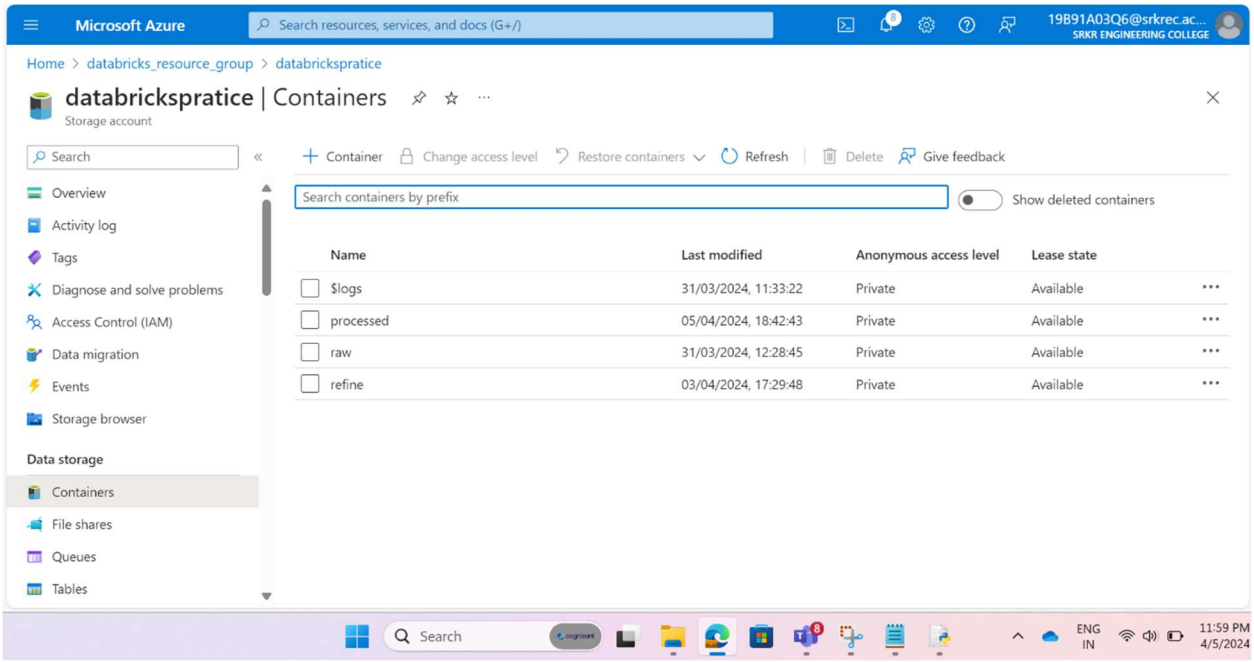
- Configure email notifications for pipeline failure and success.
- The email should contain pipeline name, run ID, and failure/success message.



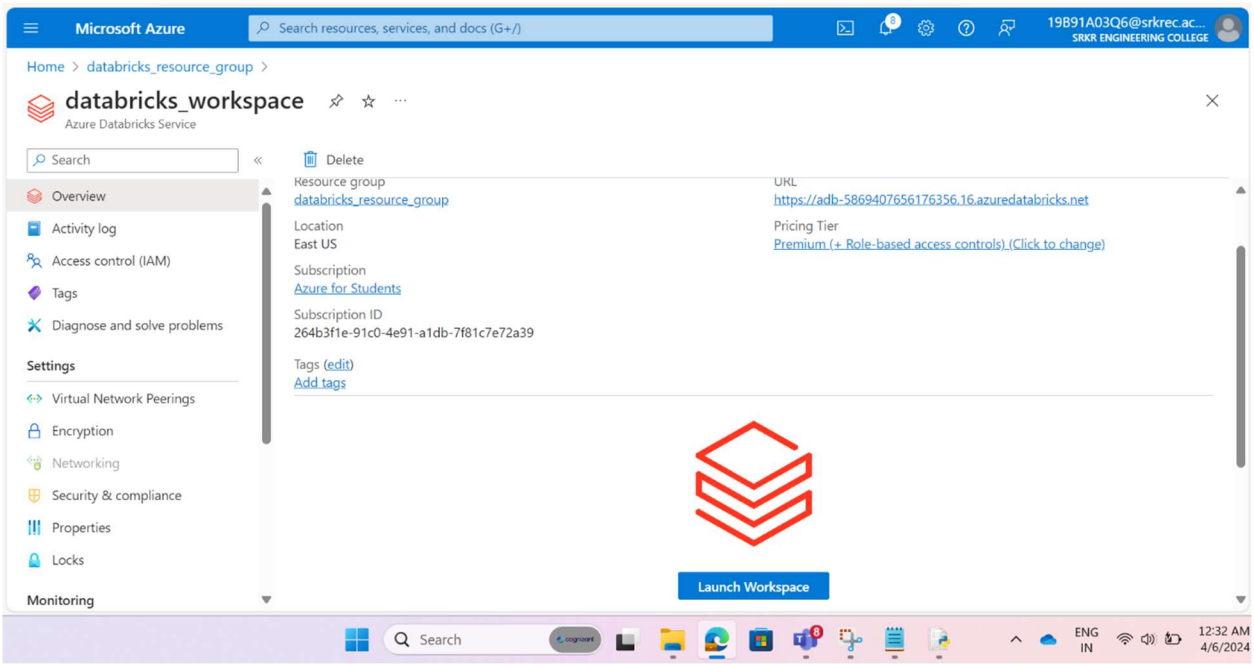


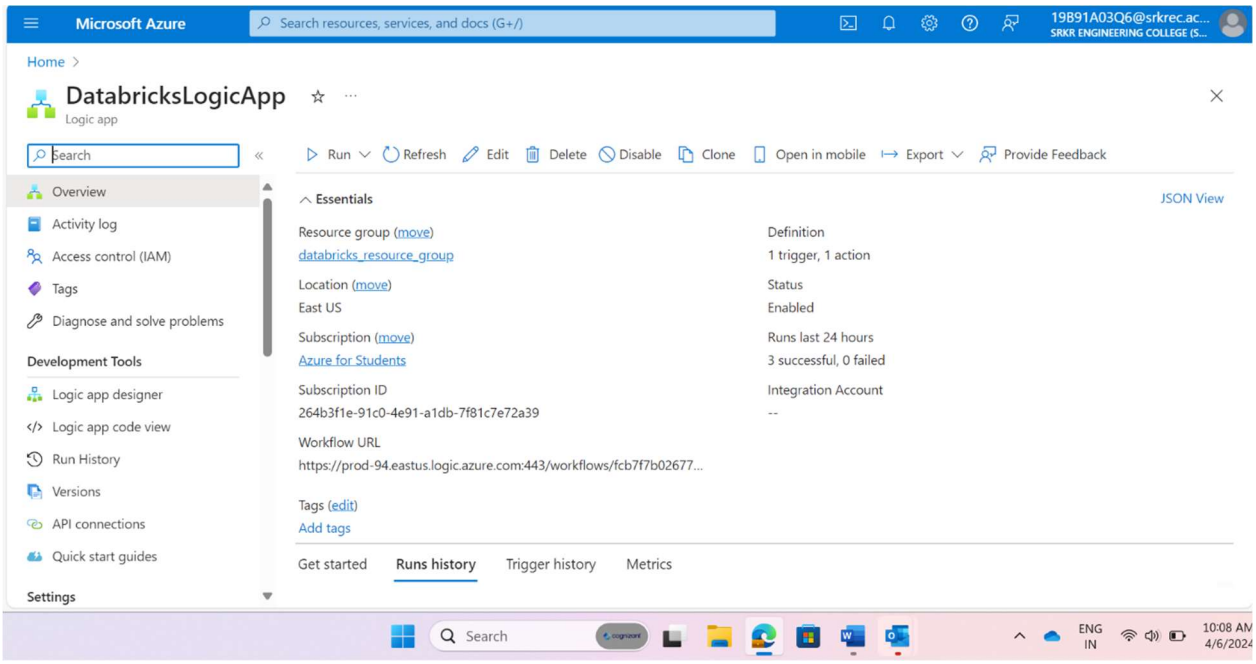

Step 1: create resource group “databricks_resource_group” which is used to hold resources.



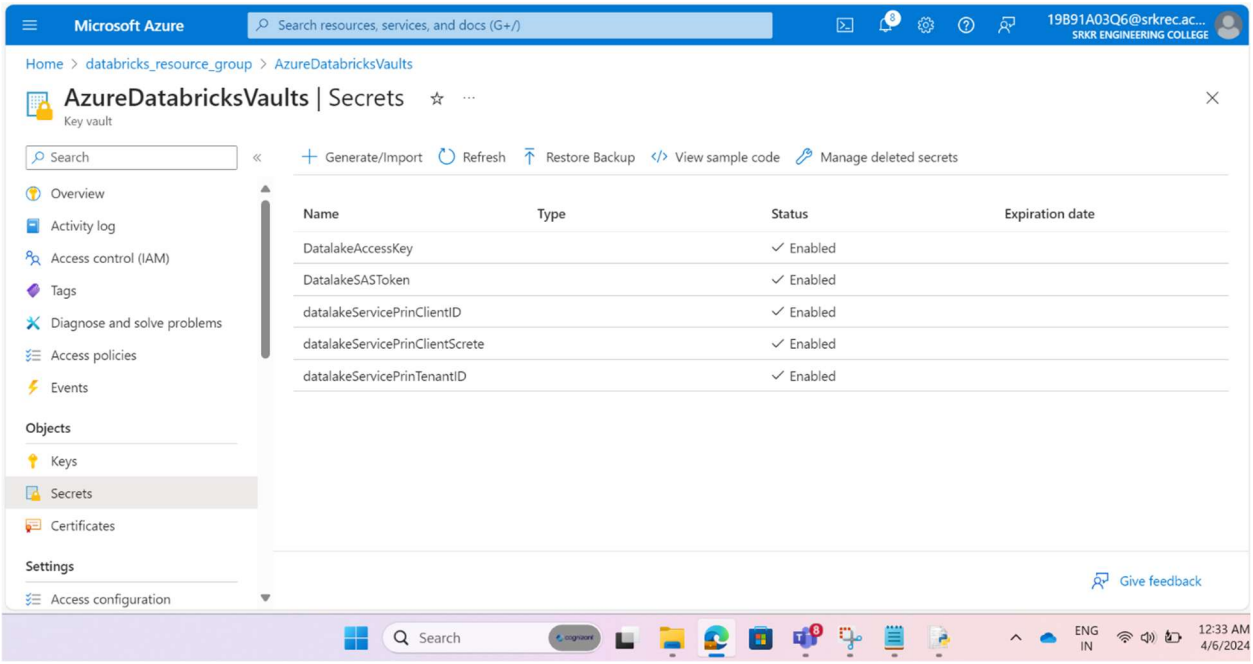

Step 2: Create storage account (Azure datalake ) with name “databrickspratice” for hierarchical namespace for storing raw data ,refined data, processed data in different folders inside containers.



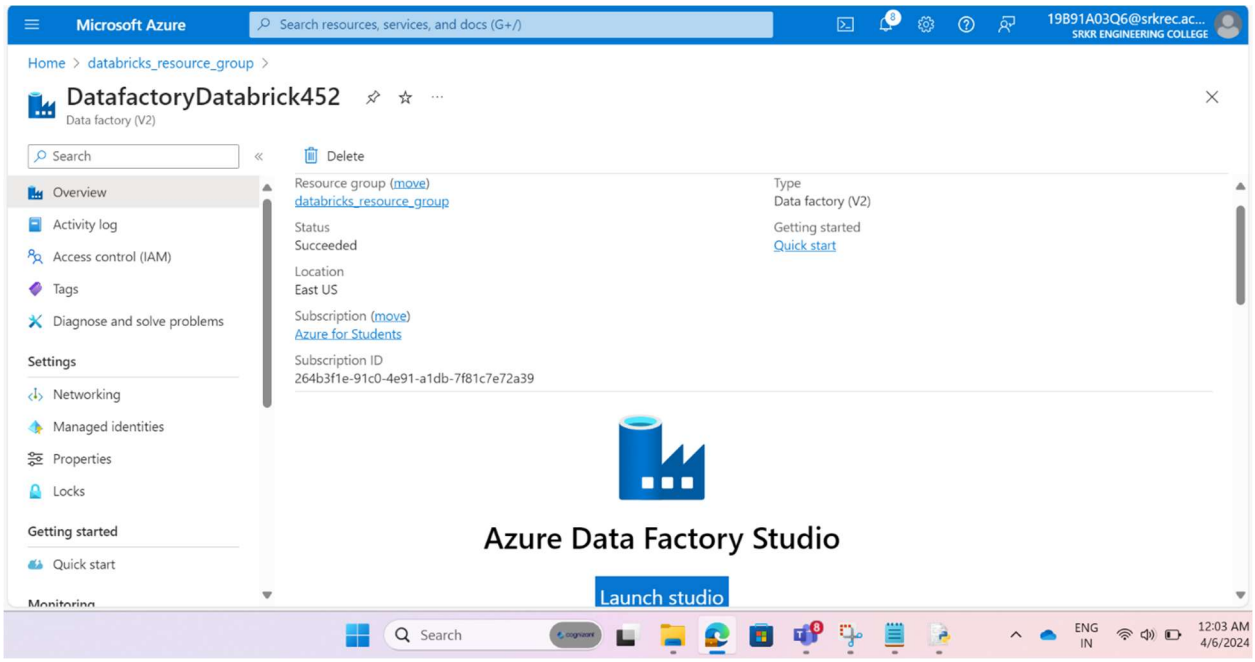

Step 3:Create data bricks workspace with name “databricks_workspace”.




Step 4: Create logic app with name “DatabricksLogicApp”.

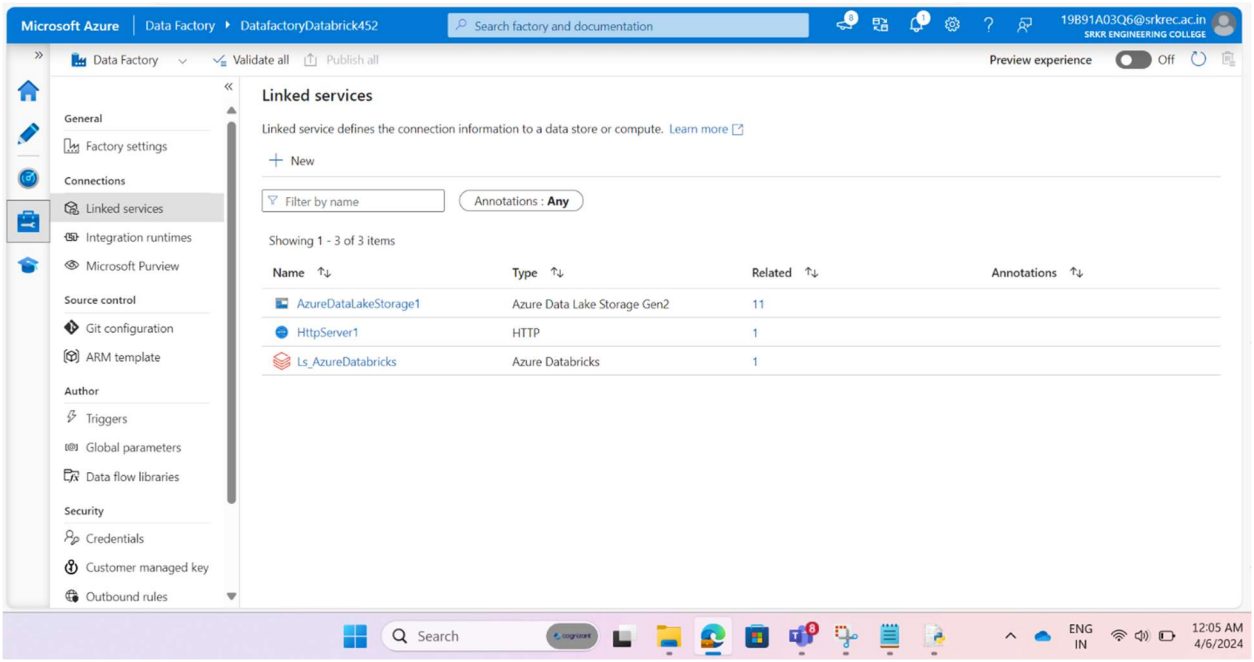



Step 5: Create Azure key vault  with name “AzureDatabricksVaults” to store secrets in key valuts for project .

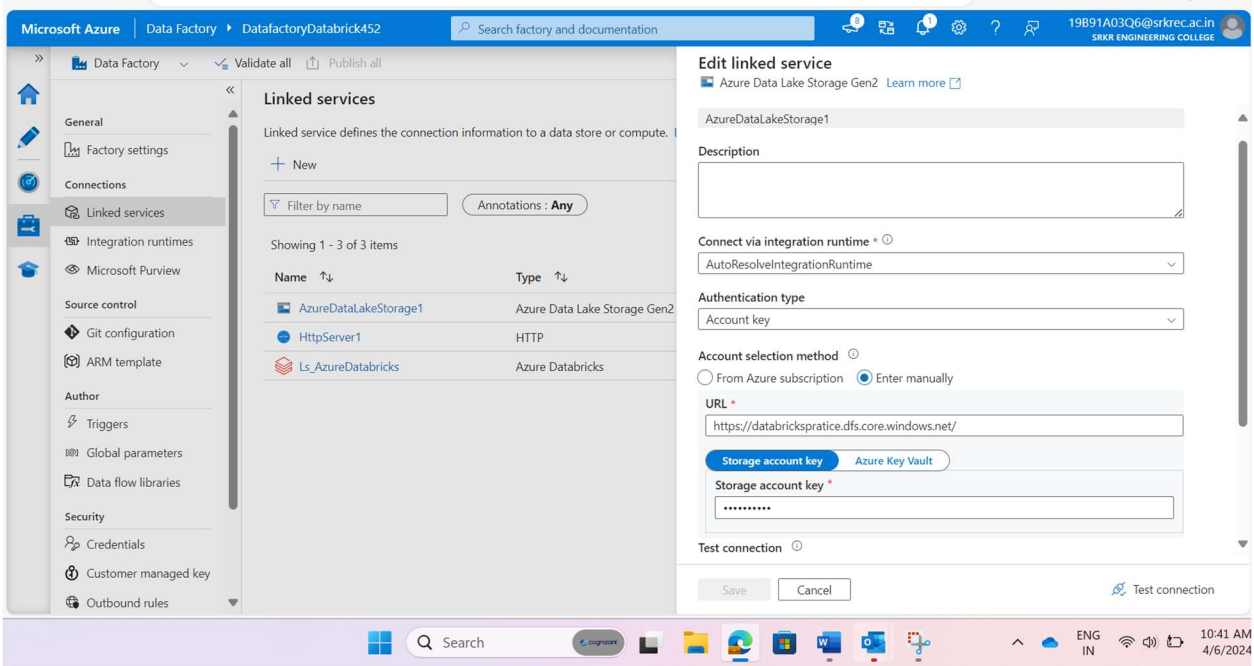



Step 6: Create data factory with name" DatafactoryDatabricks452" from creating pipelines and scheduling pipelines.

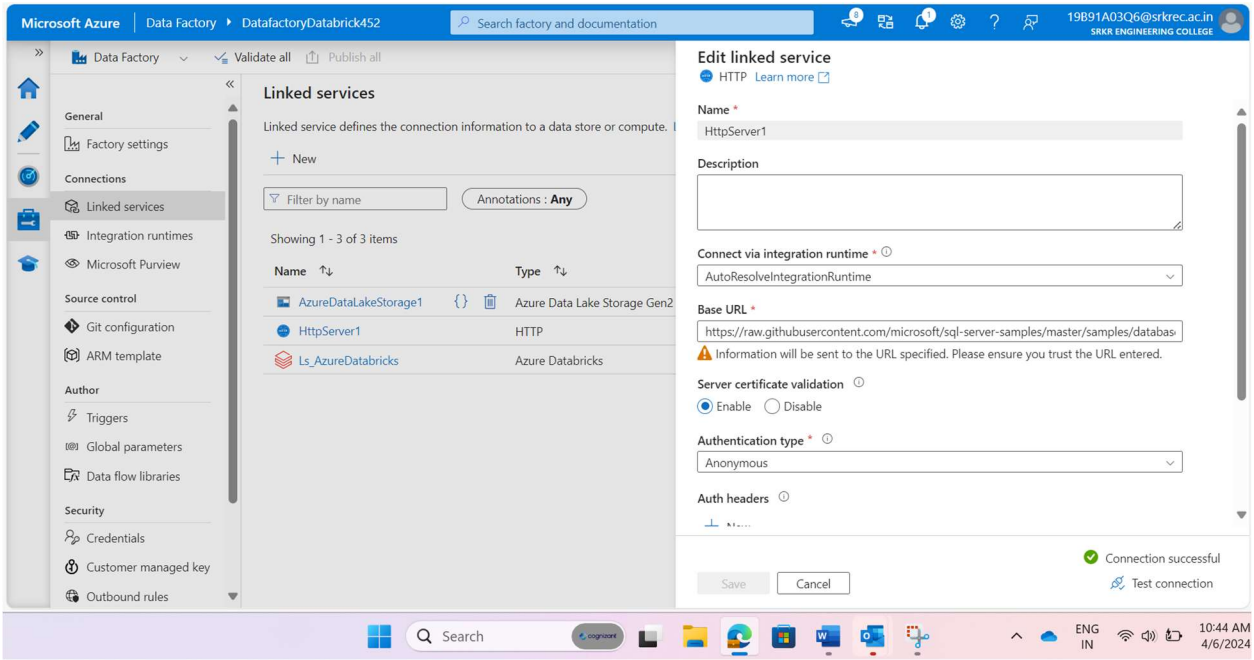
## Related to Azure DataFactory:



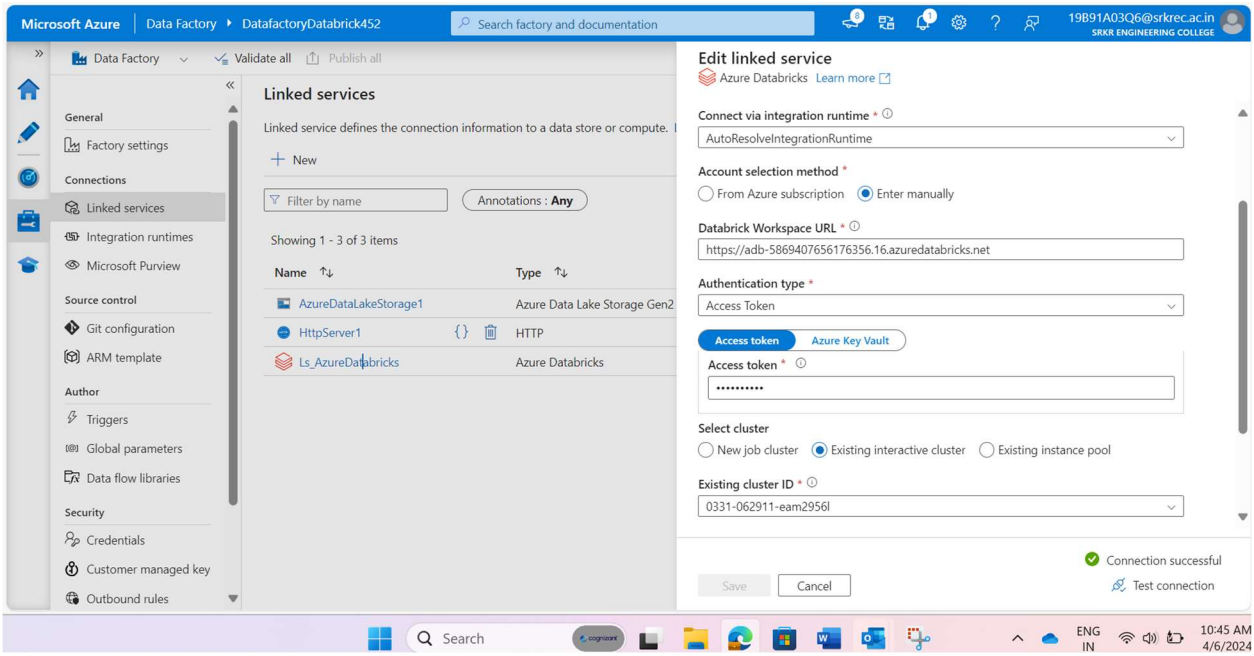
Step 7: Create Linked Services for data factory to external services provided in azure like data lake **(AzureDataLakeStorage**), data bricks **(LS_AzureDatabricks)** and http server **(HTTPSServer1).**



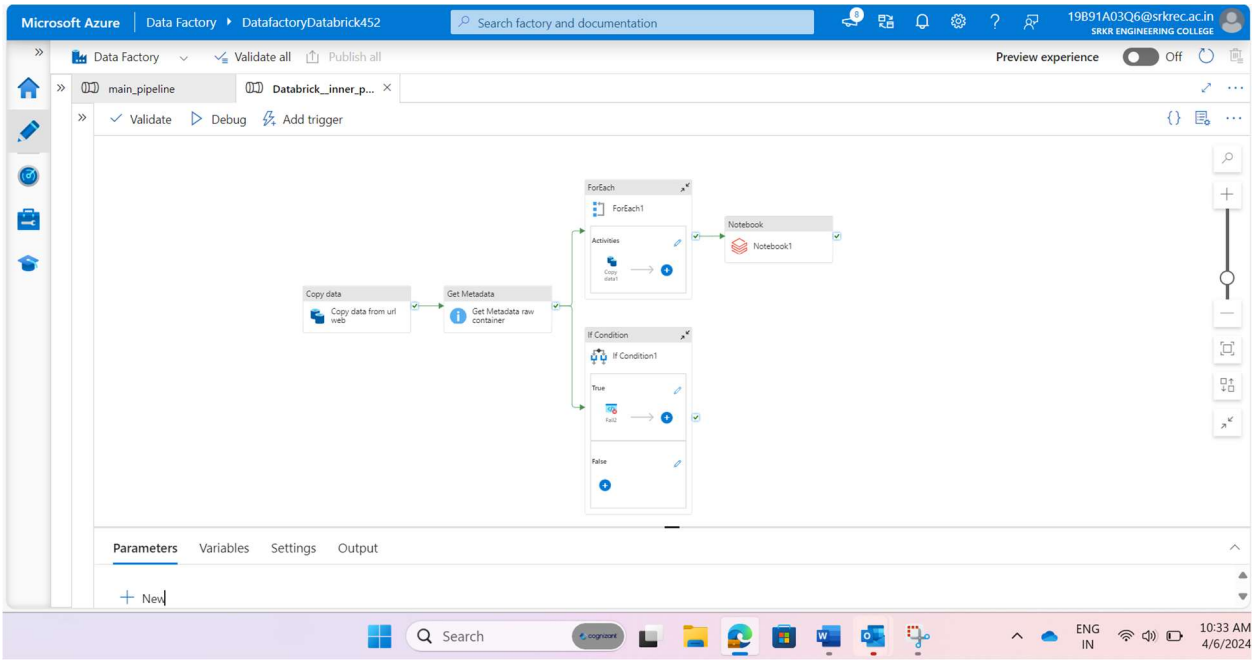

Step 8: Create Linked service between data lake and data factory with name "**AzureDataLakeStorage**".



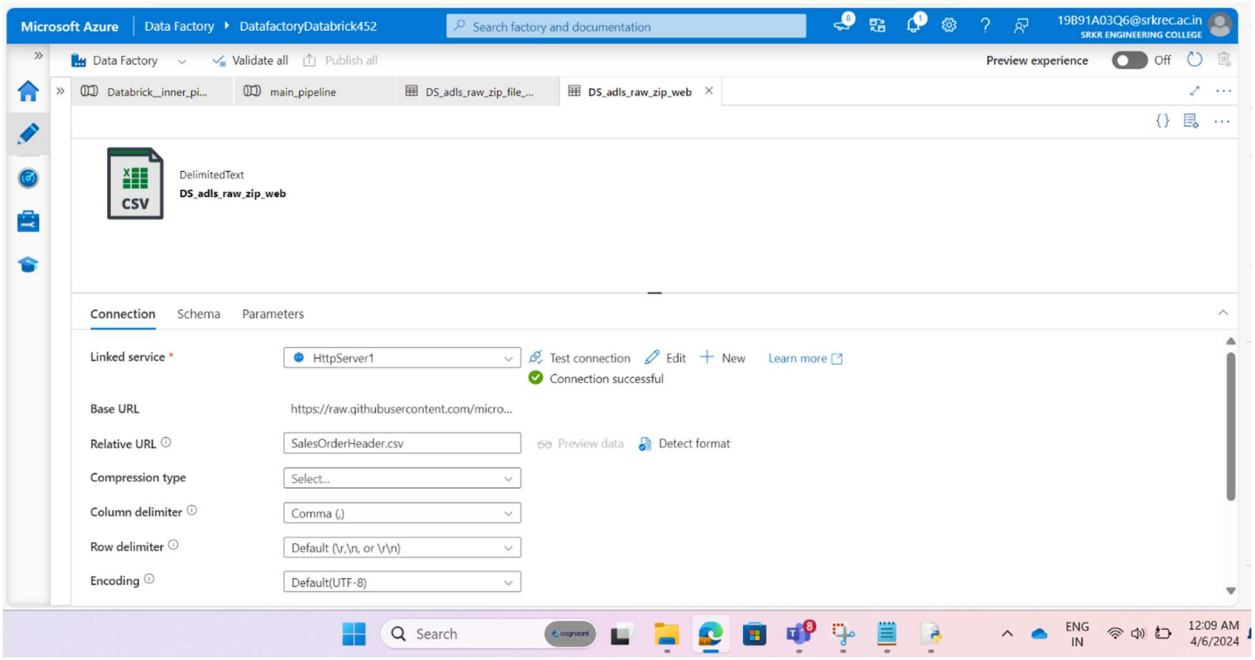
Step 9: Creating Linked Service between external data source to data factory with name "**HTTPSServer1**".



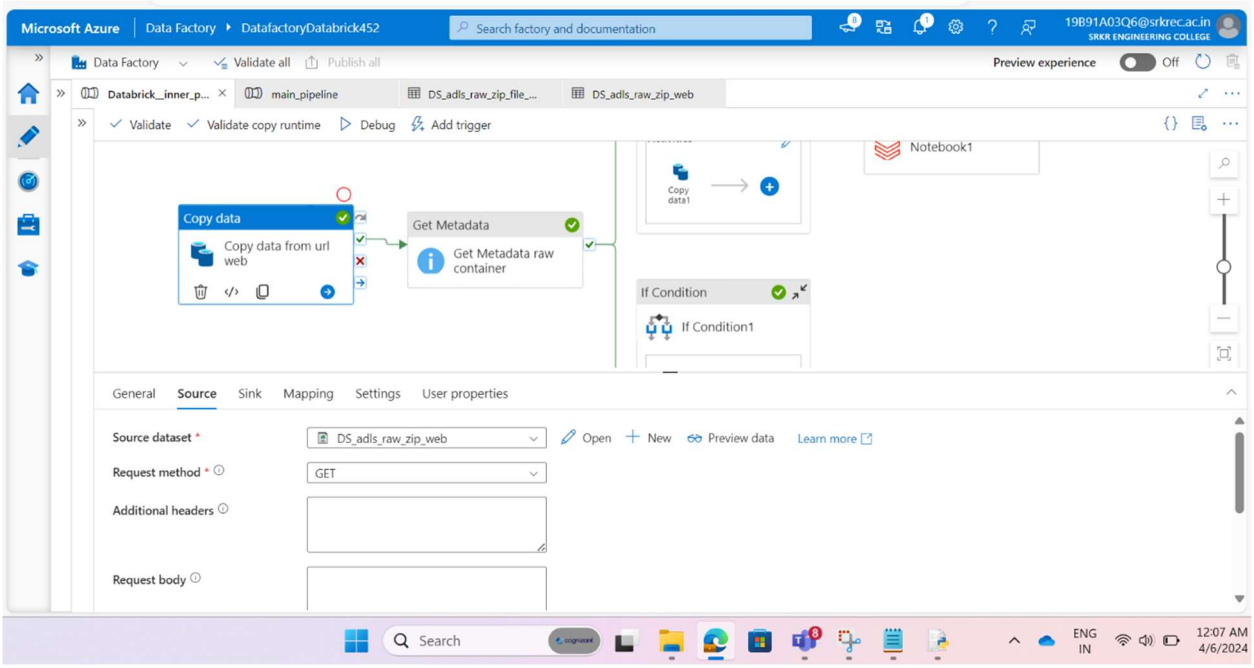

Step 10: Create Linked Service between data bricks and data factory with name "**LS_AzureDatabricks**".



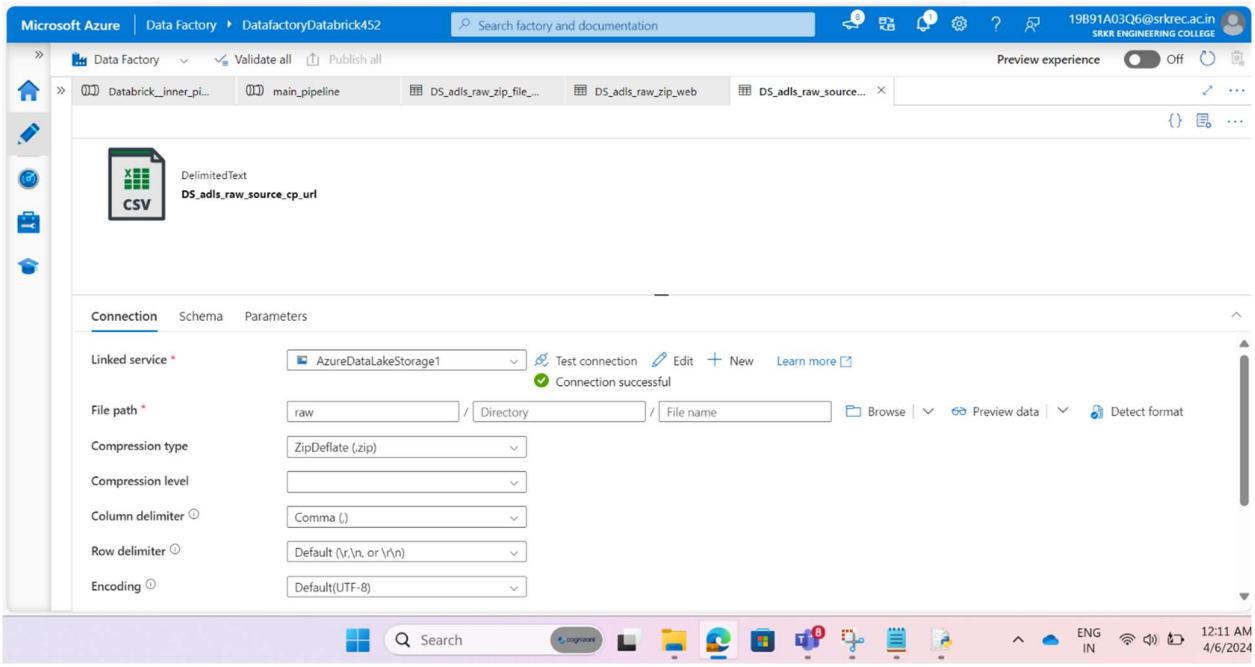

Above picture will show inner pipeline designed for data ingestion ,data transformation and data loading.



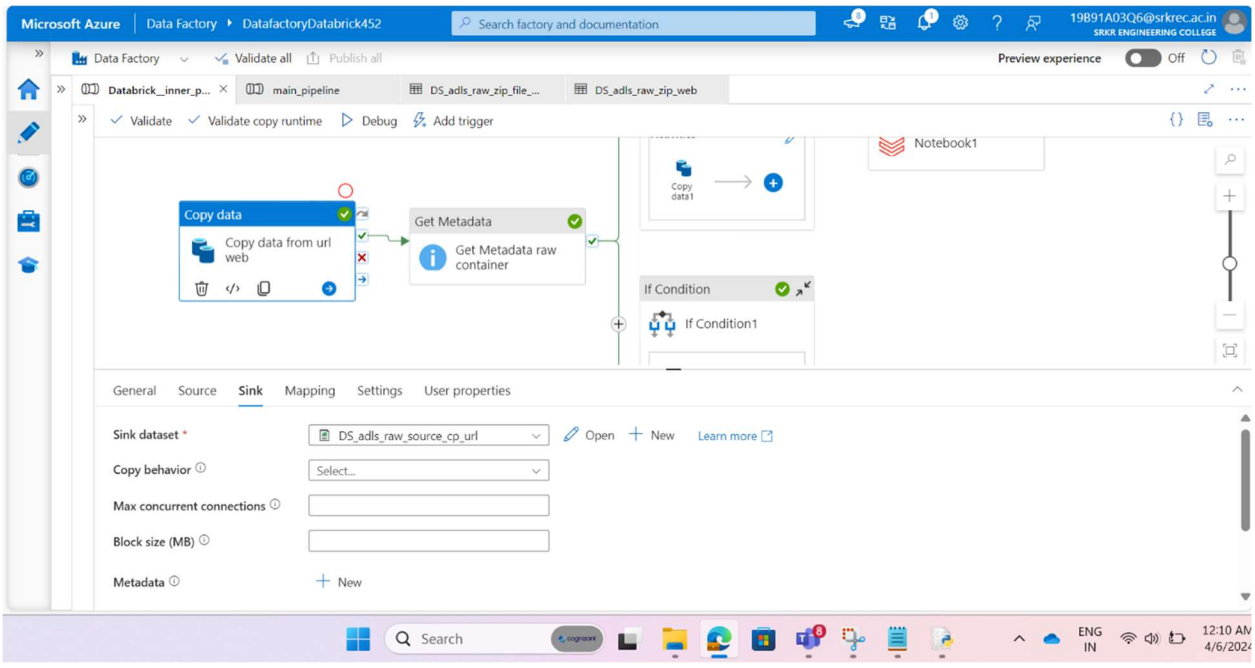

Step 11: Create source dataset for fetching data from https URL .



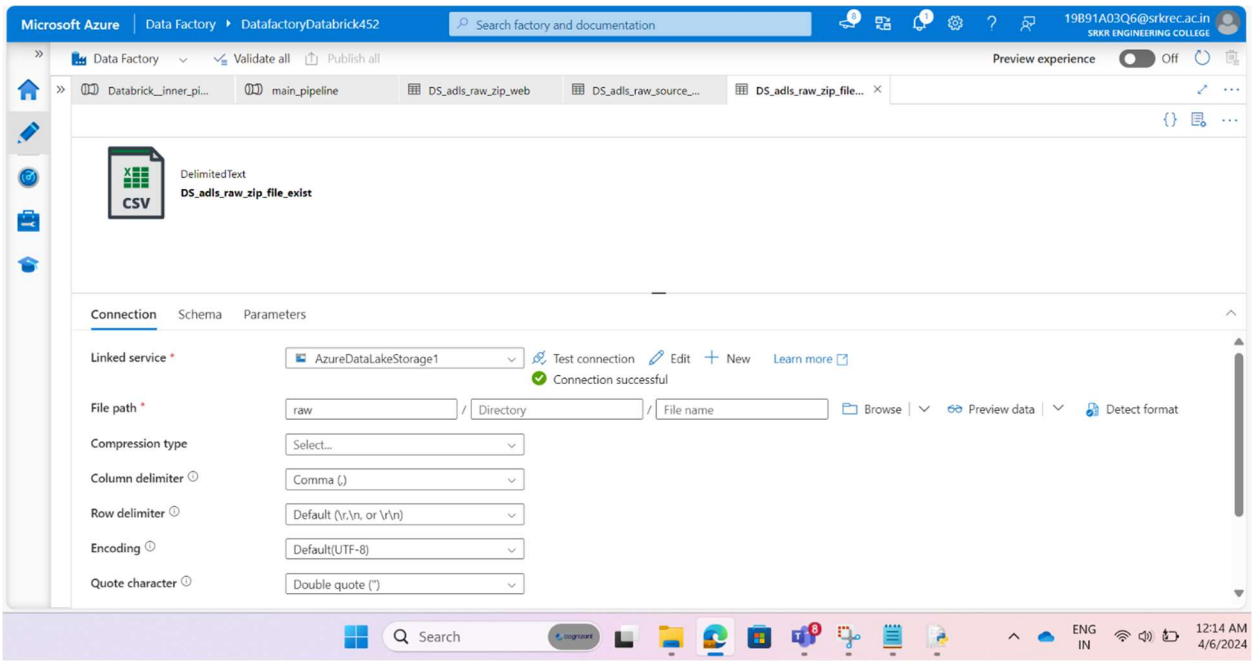

Step 12: given that **“DS_adls_raw_zip_web”** dataset for source side of copy data activity with name as **“Copy data from URL web”.**



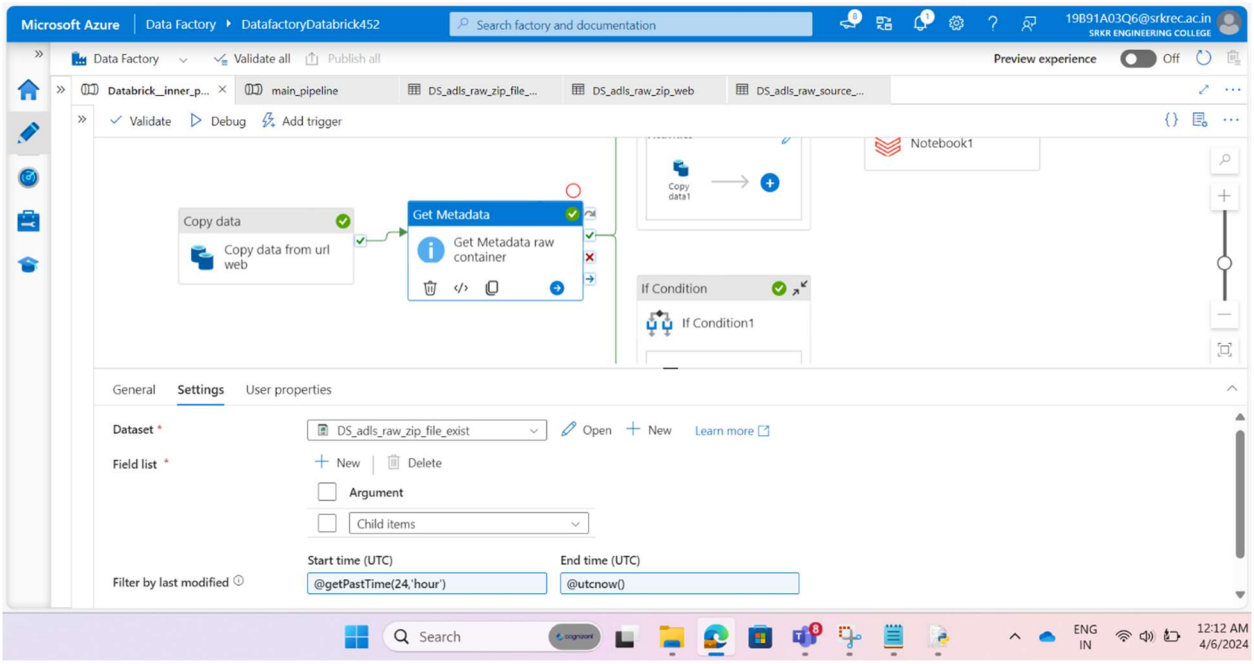

Step 13: Create sink data set for storing data set into raw container in data lake in zip format by giving compression type as (ZIP).



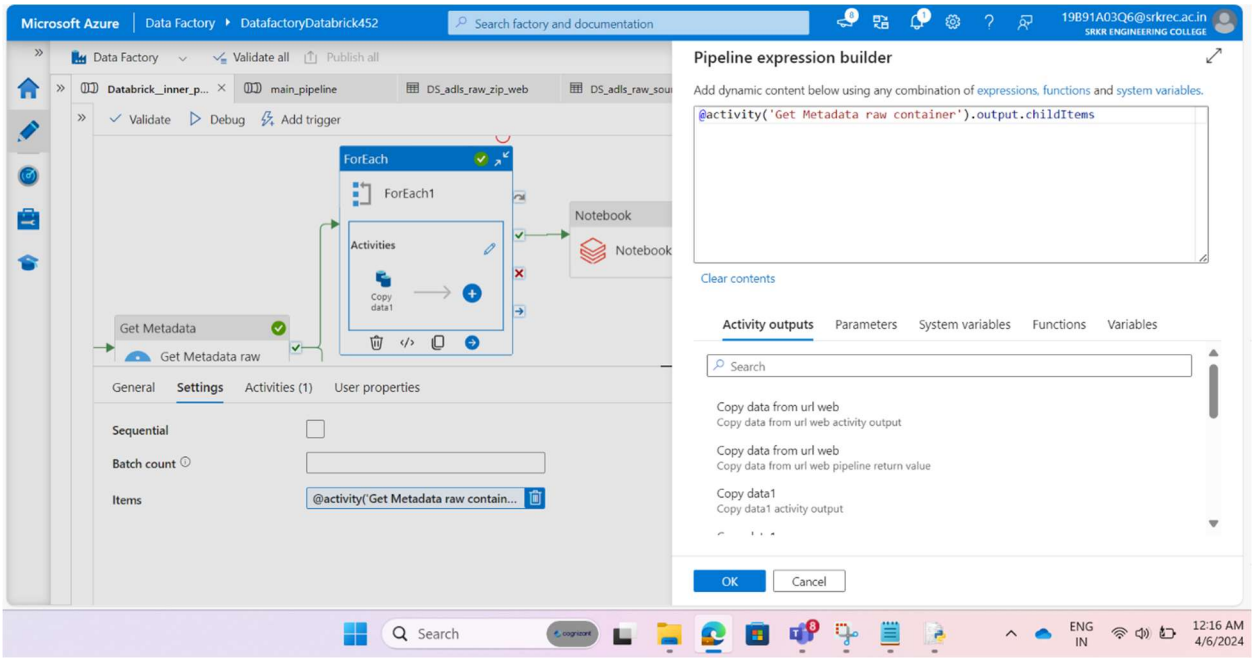

Step 14: given that **"DS_adls_raw_source_cp_url"** dataset for sink side of copy data activity **"Copy data from URL web".**



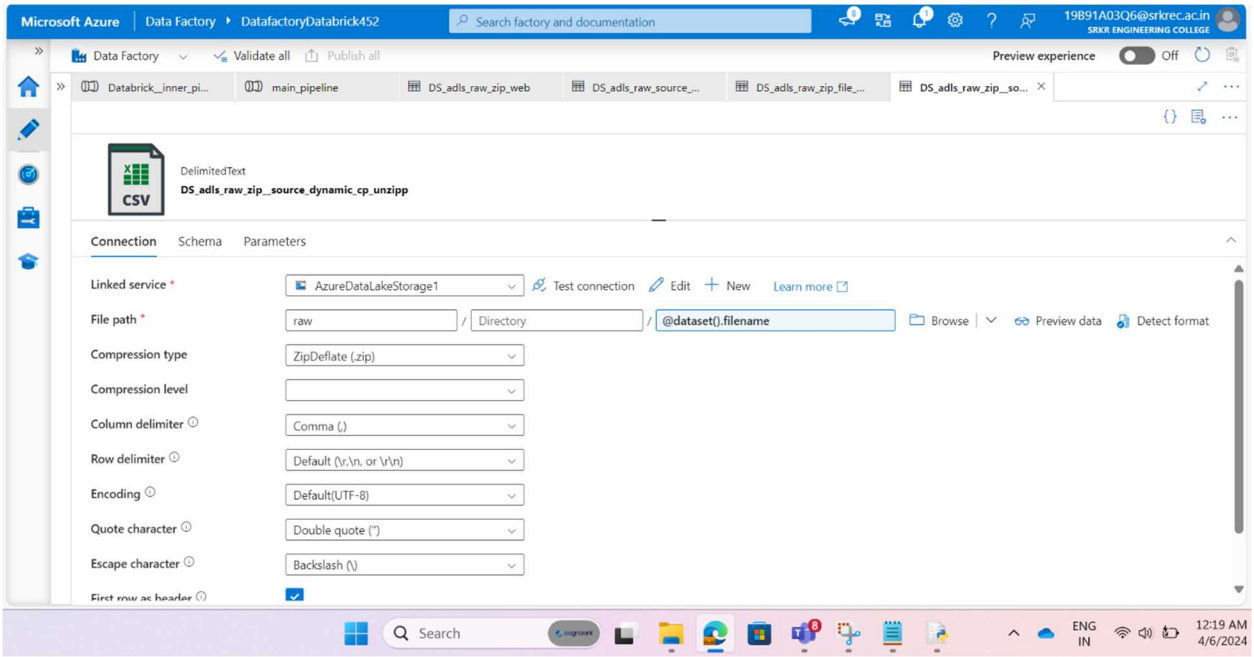

Step 15: Create dataset for get metadata activity to point out to raw container for feaching metadata about files and folders in that container.



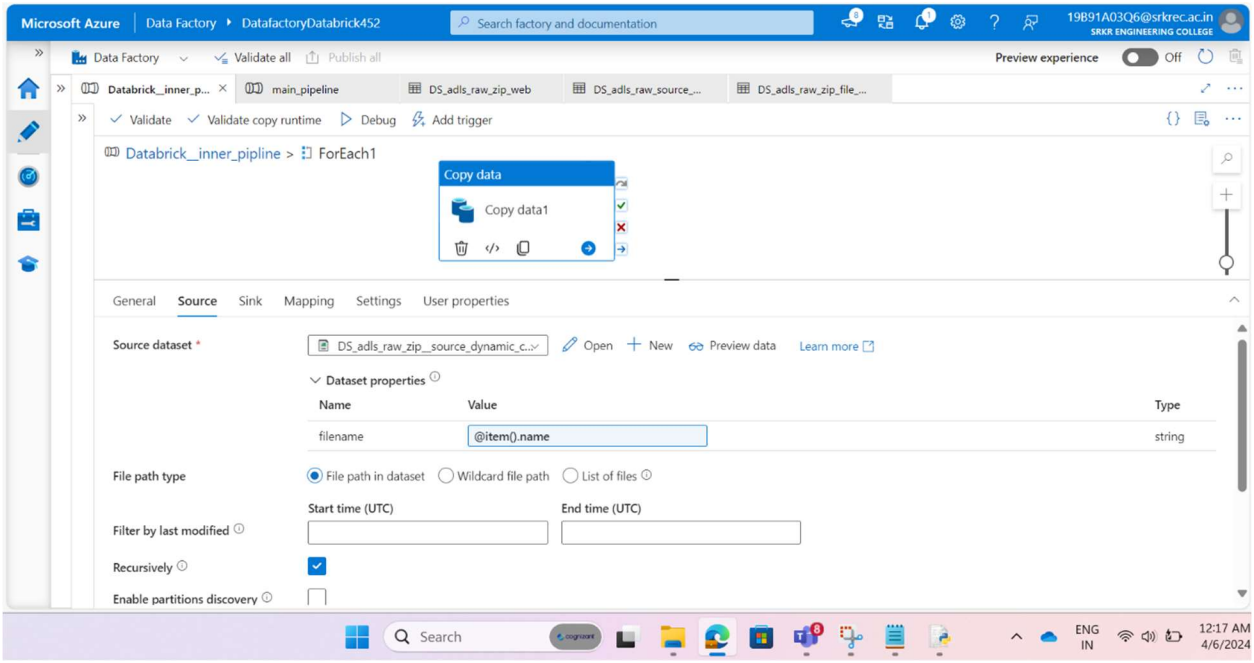

Step 16 : Given that “DS_adls_raw_zip_file_exist” dataset for picking childitems of files in raw containers within 24 hours from present time when pipeline start .



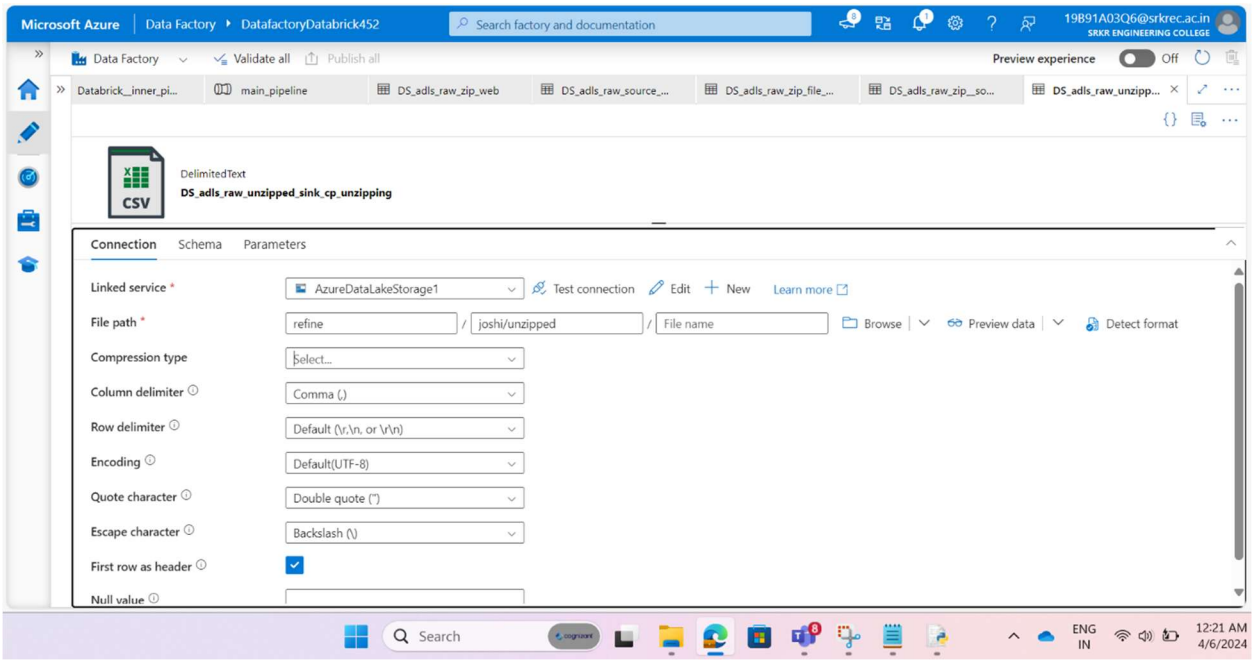

Step 17: Drag and drop for each activity for iterate over those child items pick by get metadata. Connect get metadata output with for each activity.



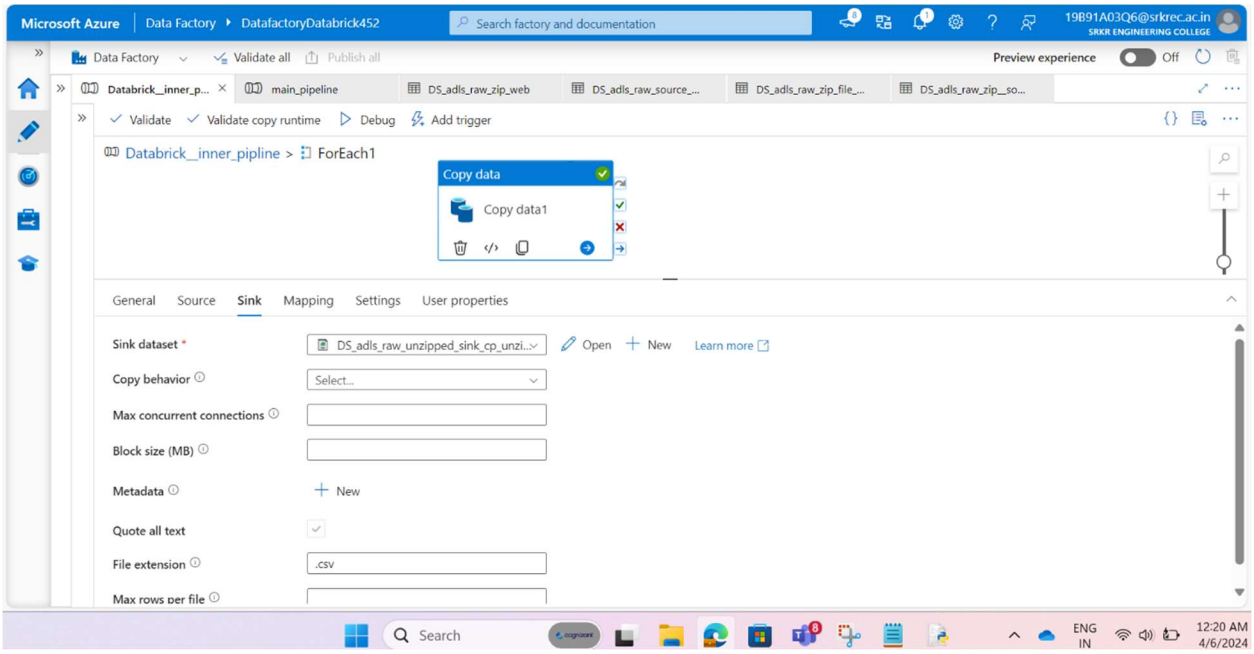

Step 18: Create source dataset for copy activity inside for each activity by passing file name dynamically by creating parameter with **"filename"** and passing value **"@dataset().filename"** at dataset level dynamically and giving compression type as .Zip .



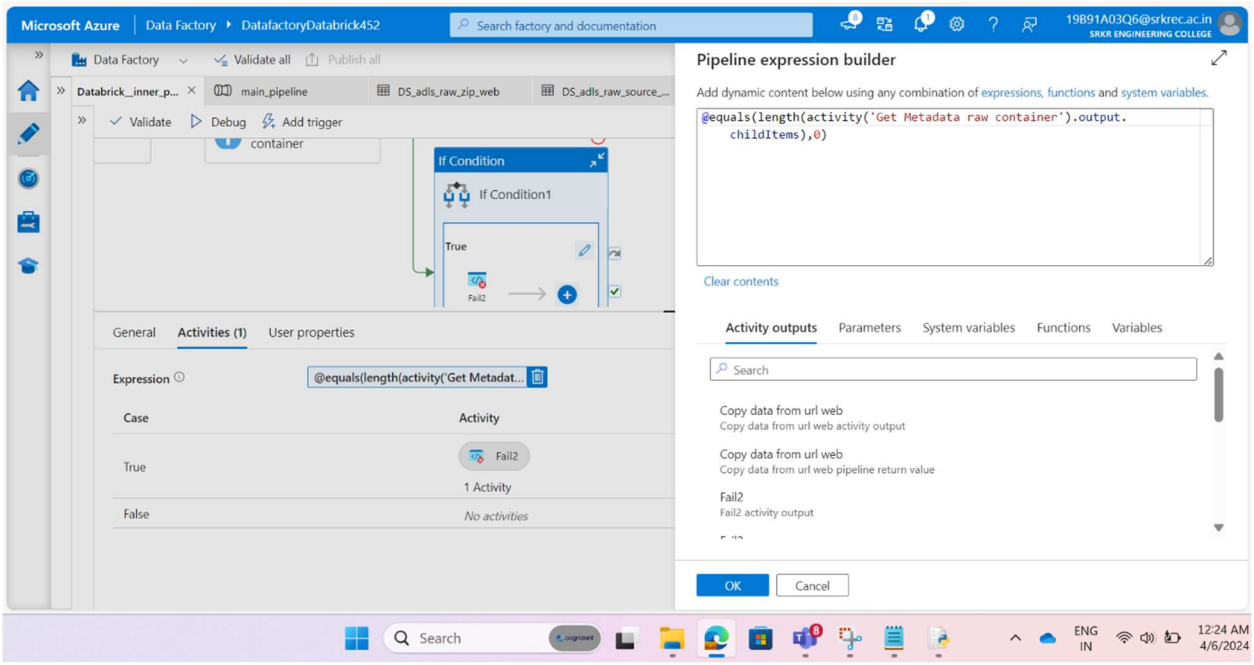

Step 19: Given that "**DS_adls_raw_zip_source_dynamcially_cp"** dataset at source side to copy data activity inside for each activity.



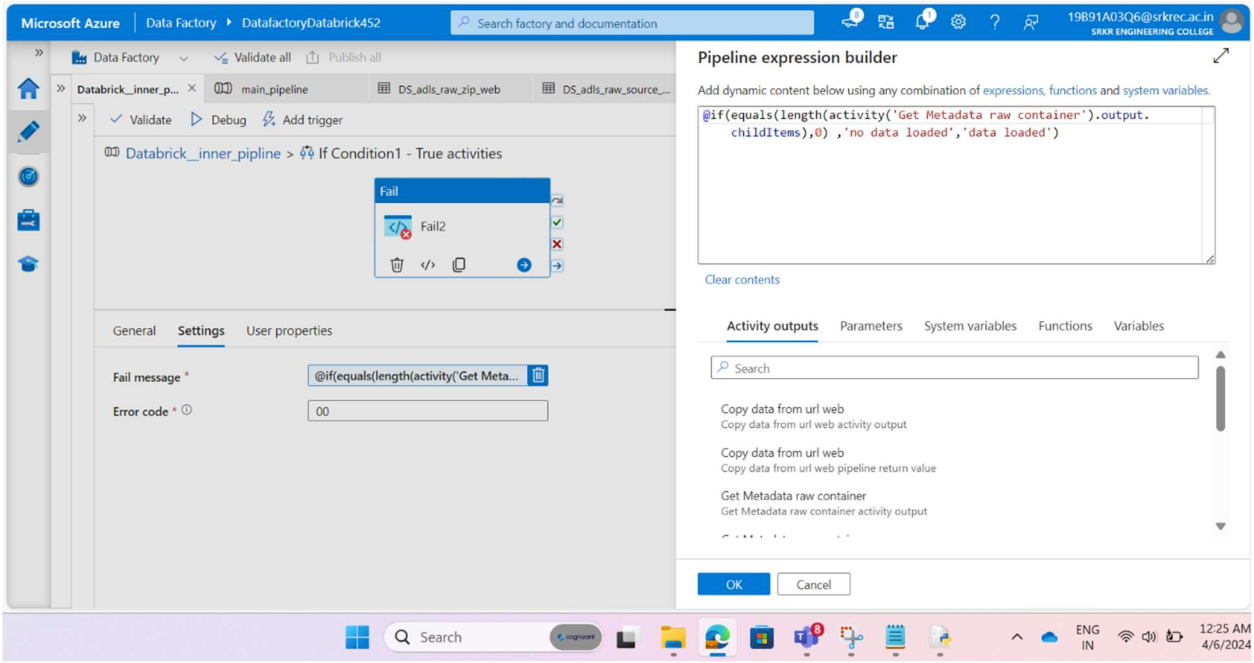

Step 20: Create dataset for sink side of copy data activity inside for each activity for copying zip file data from raw container to refine container inside folder name **"joshi/unzipped"** by unzipping files and keeping only **.csv files** inside refined container.



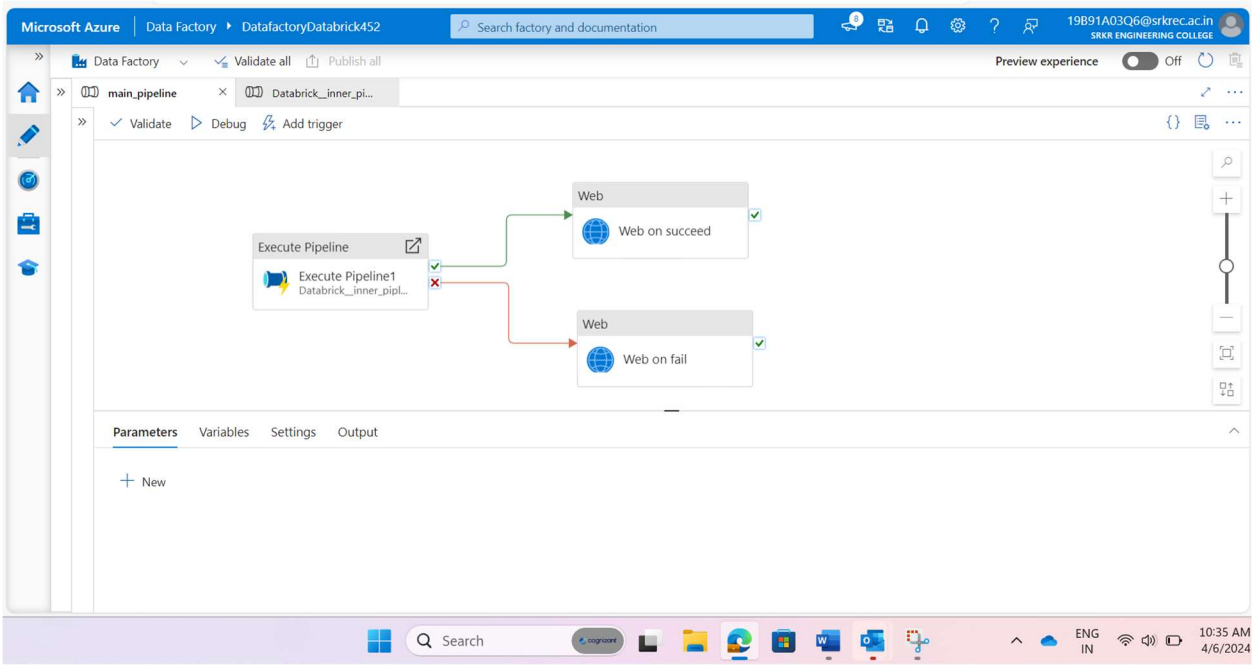

Step 21: given that **“DS_adls_raw_unzipped_sink_cp”** dataset to sink side of copy data activity inside for each activity.



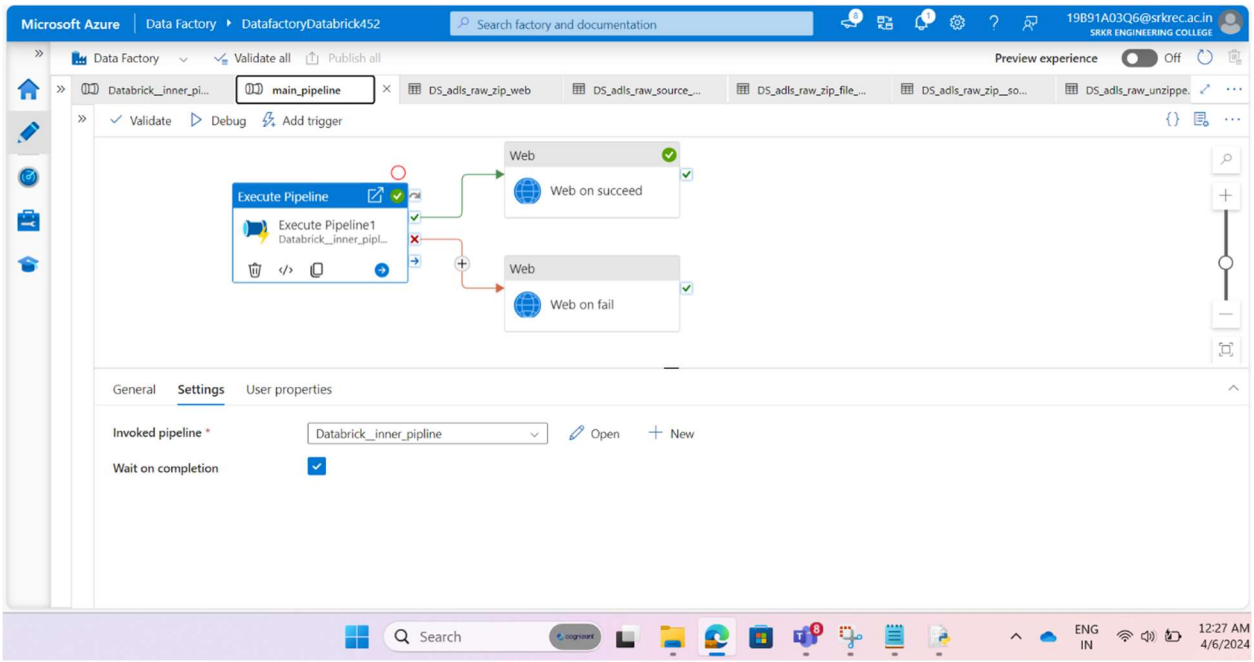

Step 22: Drag and drop if activity for checking child items length return by get metadata activity by giving condition like above it will check whether child item length is equals to zero or not.



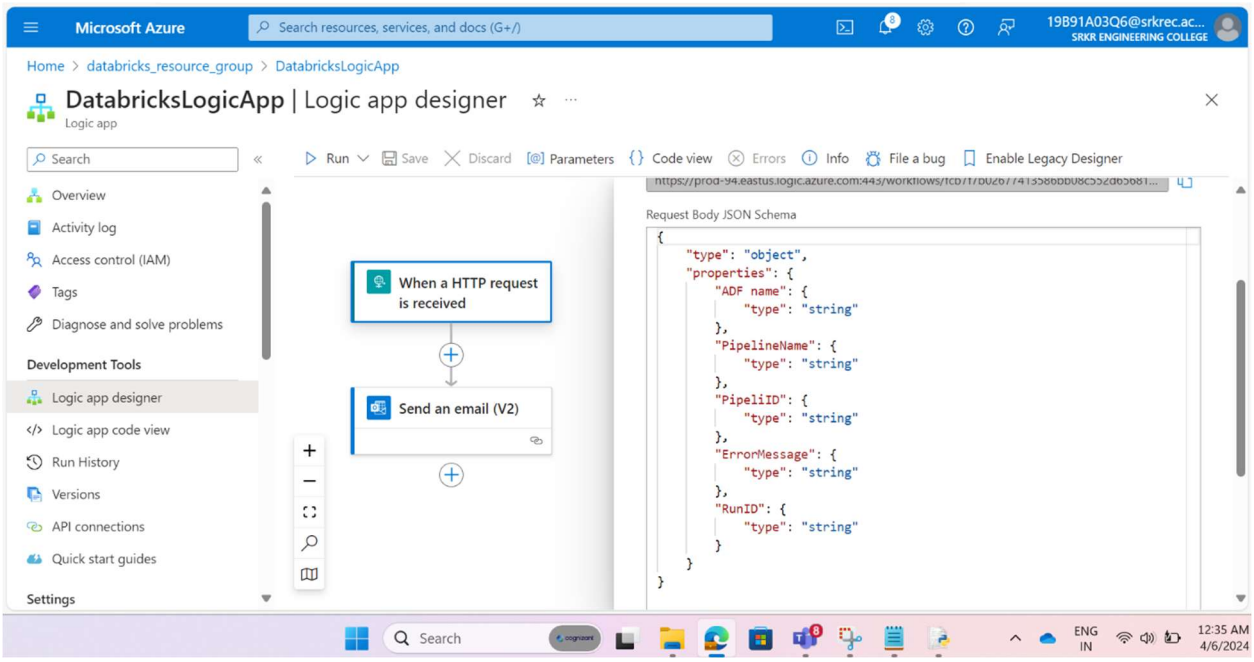

Step 23: Drag and drop Fail activity inside if condition while condition is true fail activity will work. Check for child item length whether zero or not depending upon that fail activity will fail pipeline when condition is true.



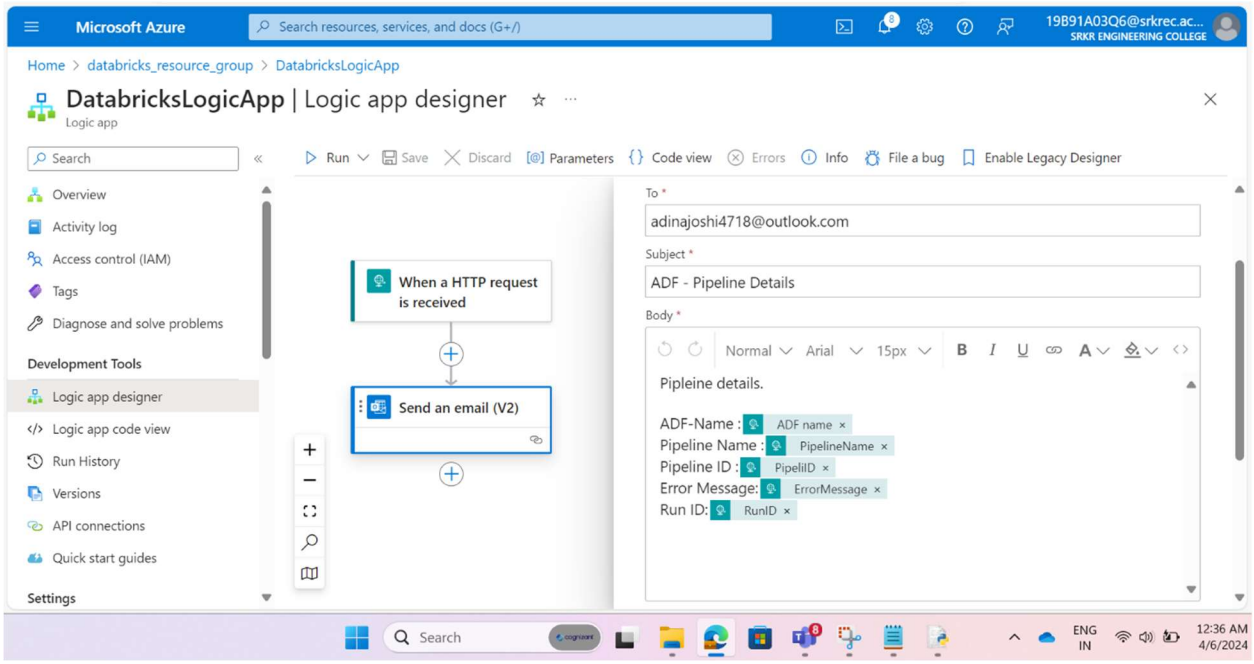

Step 24: Create another main pipeline with Execute pipeline activity, web activity.



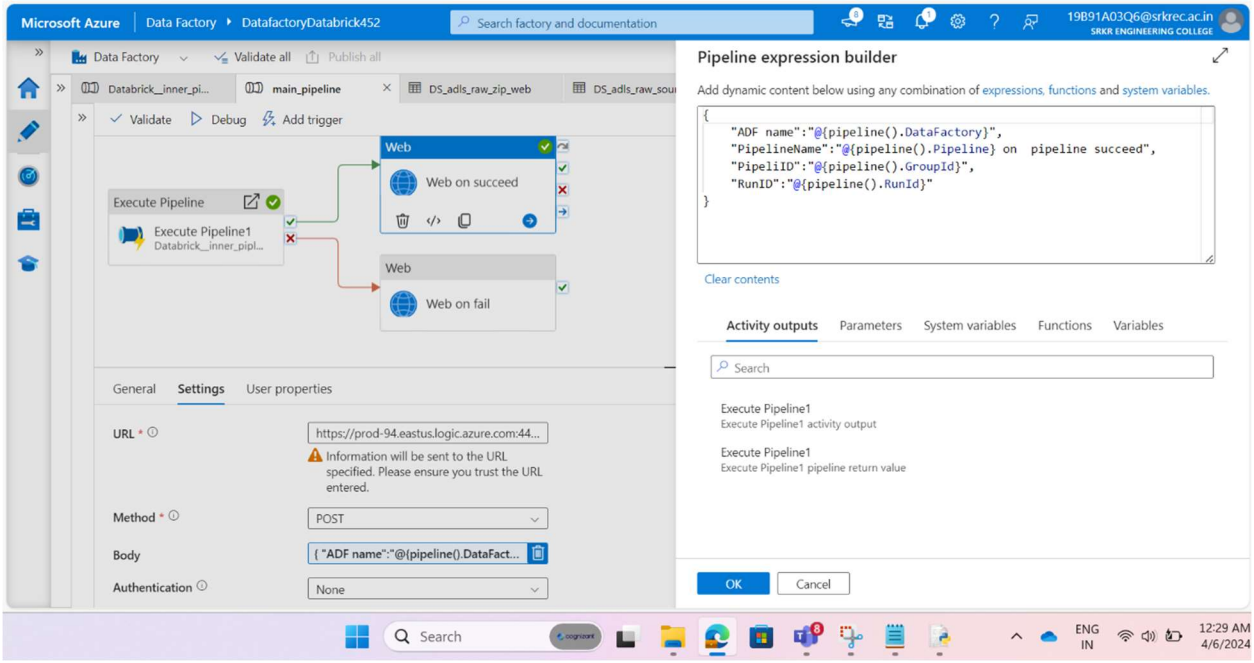

Step 25: given another pipeline to this activity which is nothing previous pipeline with name **“Databrick_inner_pipelin”.**



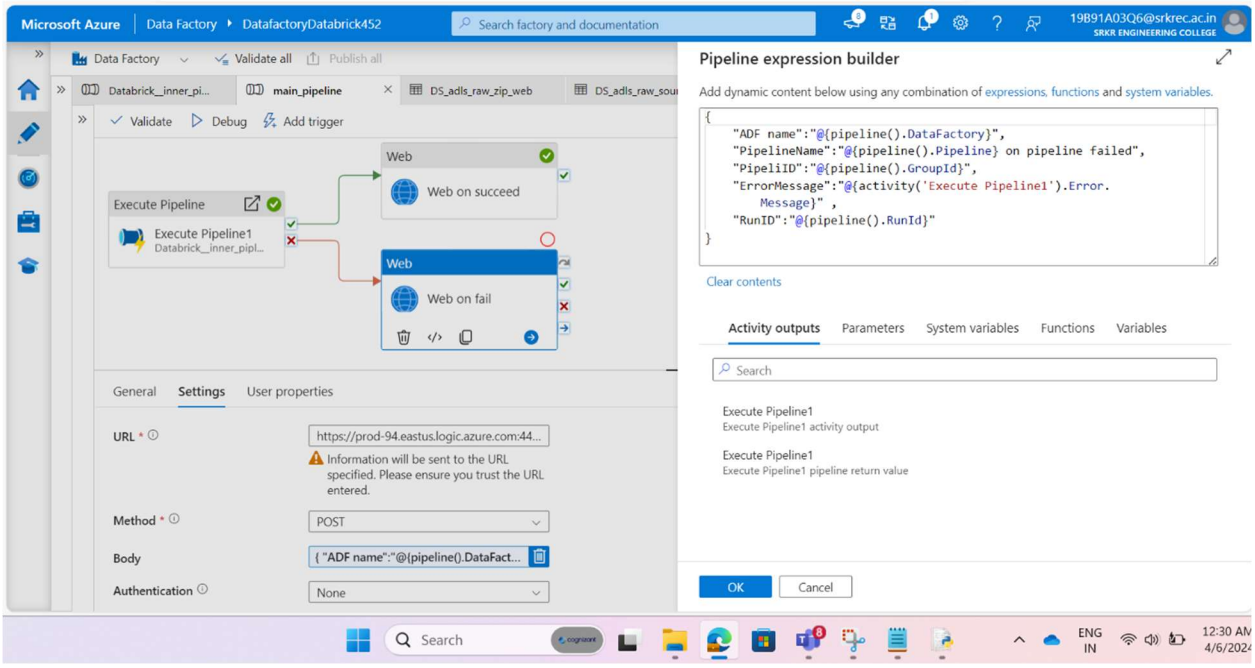

Step 26: Design log app for triggering mail when pipeline fail or on pipeline succeed by passing Json format which are to be appear in mail like pipeline id, ADF name, pipeline name, error message.




Step 27: Design custom mail by giving personal mail id and craeting body with requried details reagarding pipeline fails details.

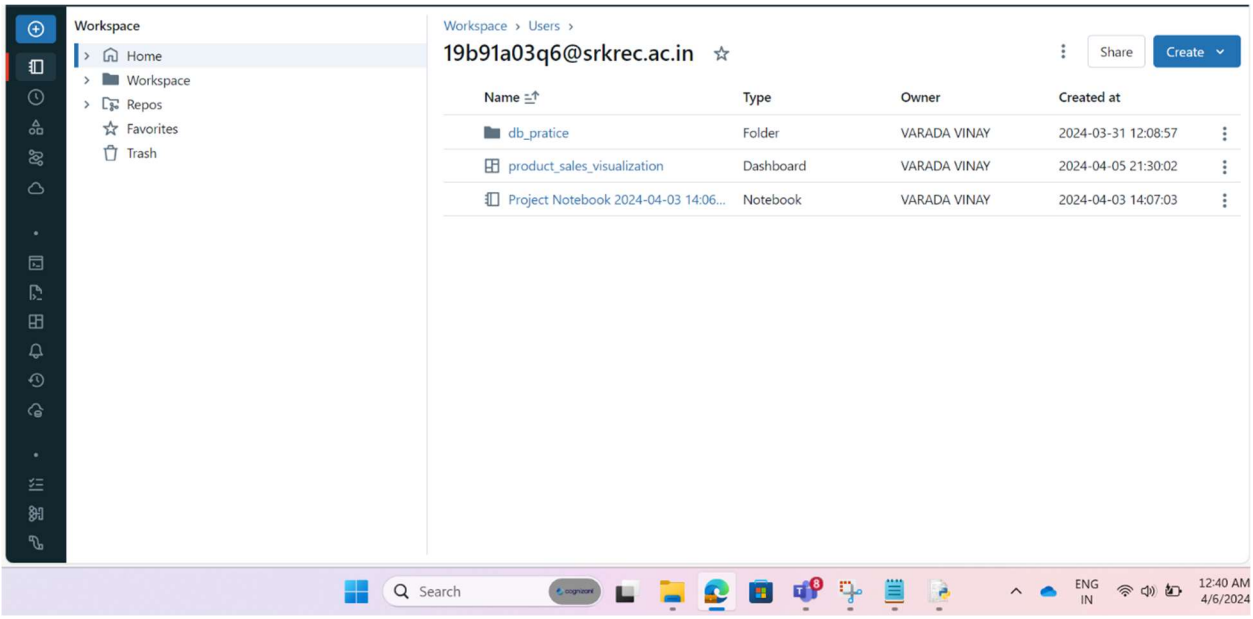


Step 28: given ADF name, pipeline name, pipeline id, Run id dynamically to https request to trigger  mail in logic app on pipeline succeed.

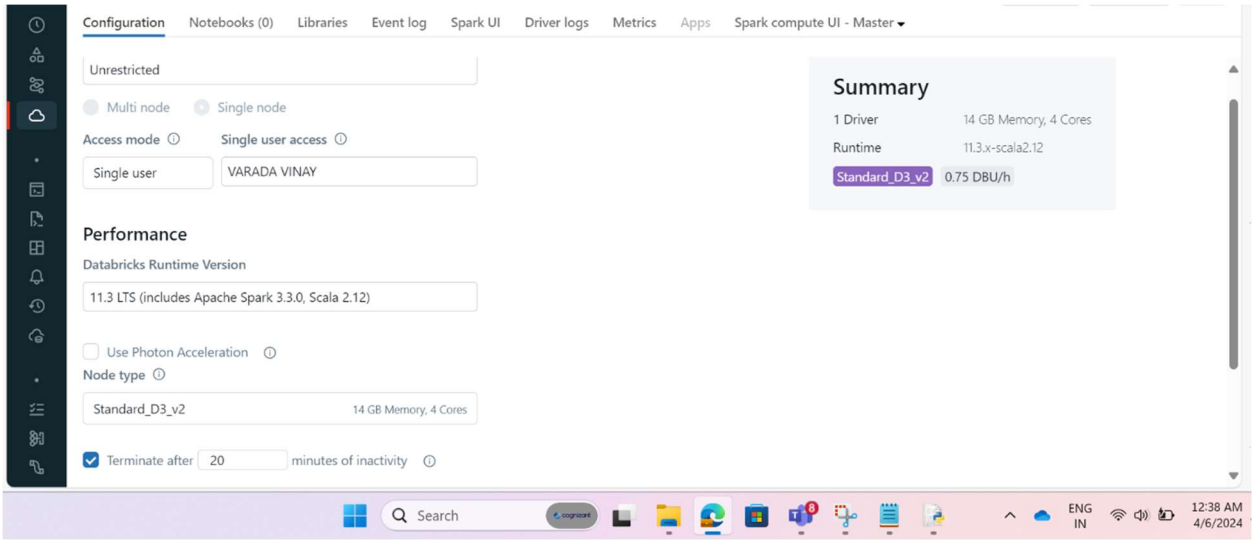



Step 29: given ADF name, pipeline name, pipeline id, Run id, error message dynamically to https request to trigger  mail in logic app on pipeline fail.

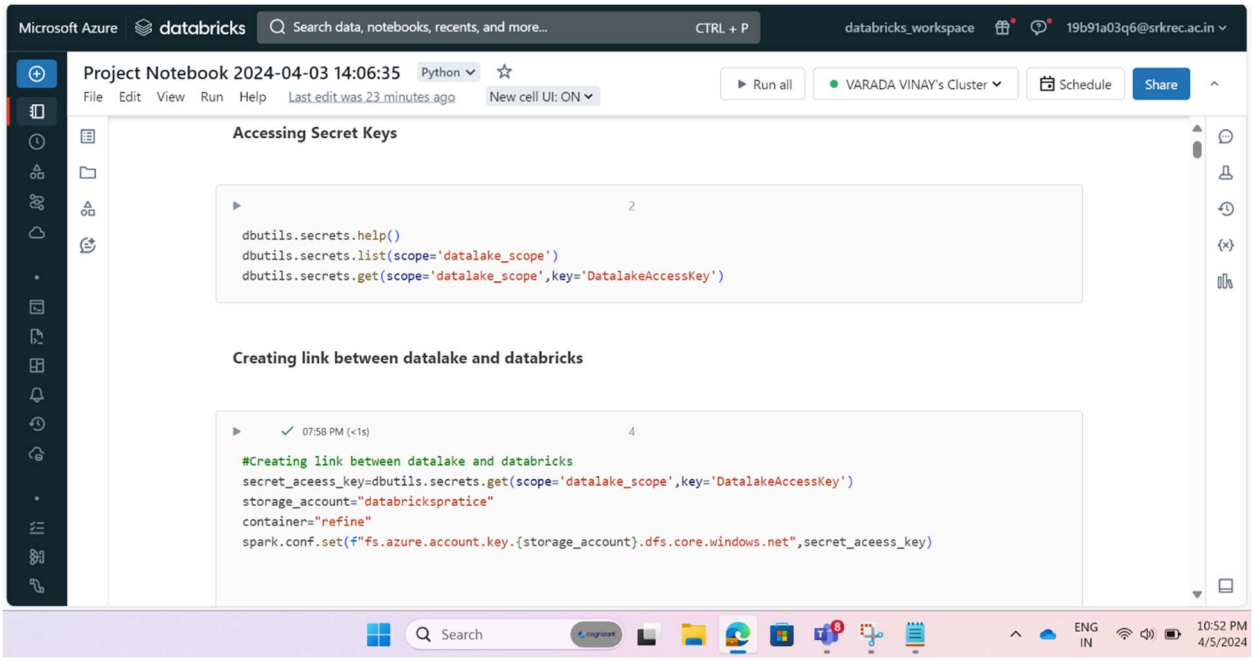
## Related to Data bricks:



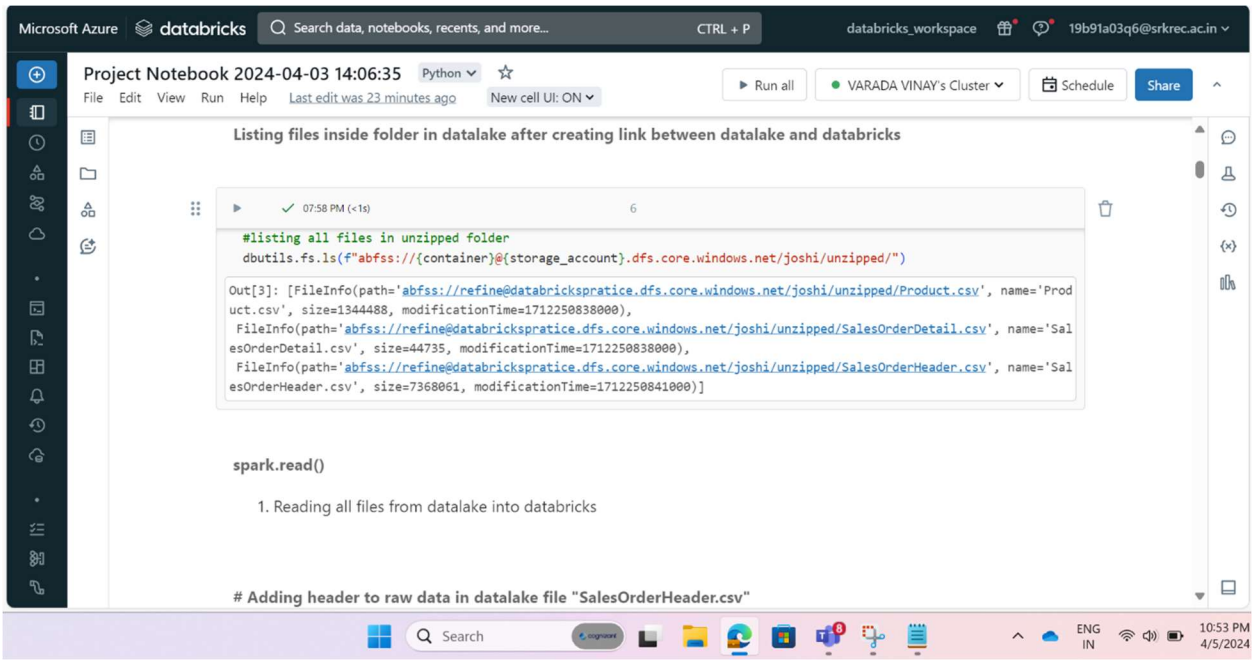

Step 30: Create workspace for doing data cleaning and data transformation on raw data.



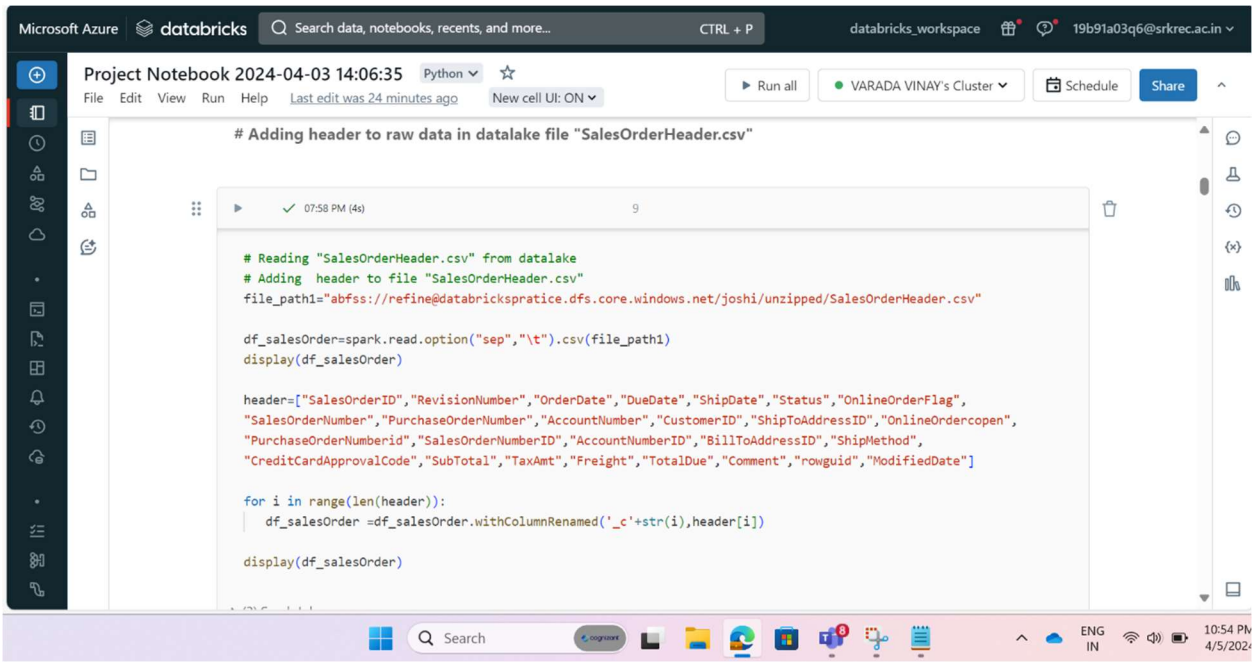

Step 31: Create cluster with single node for running queries written in notebook and to give output.



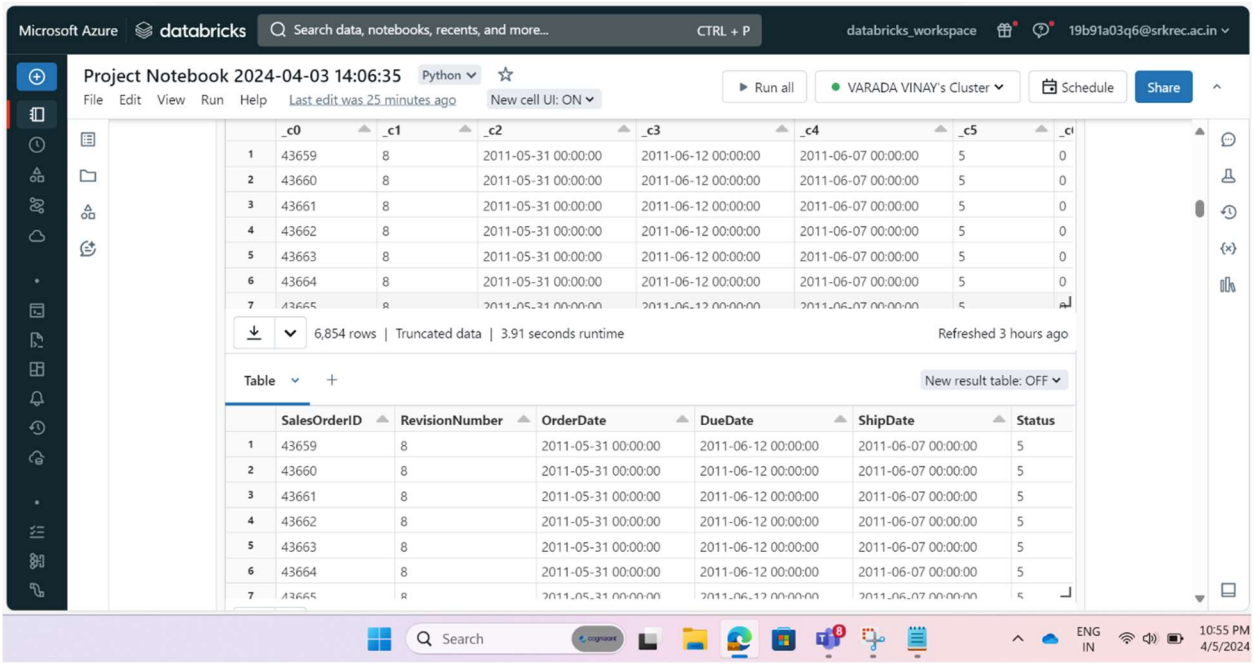
Step 32: By using above code we can fetch secrets creating in data bricks and we can create link and authentication for data lake to data bricks.



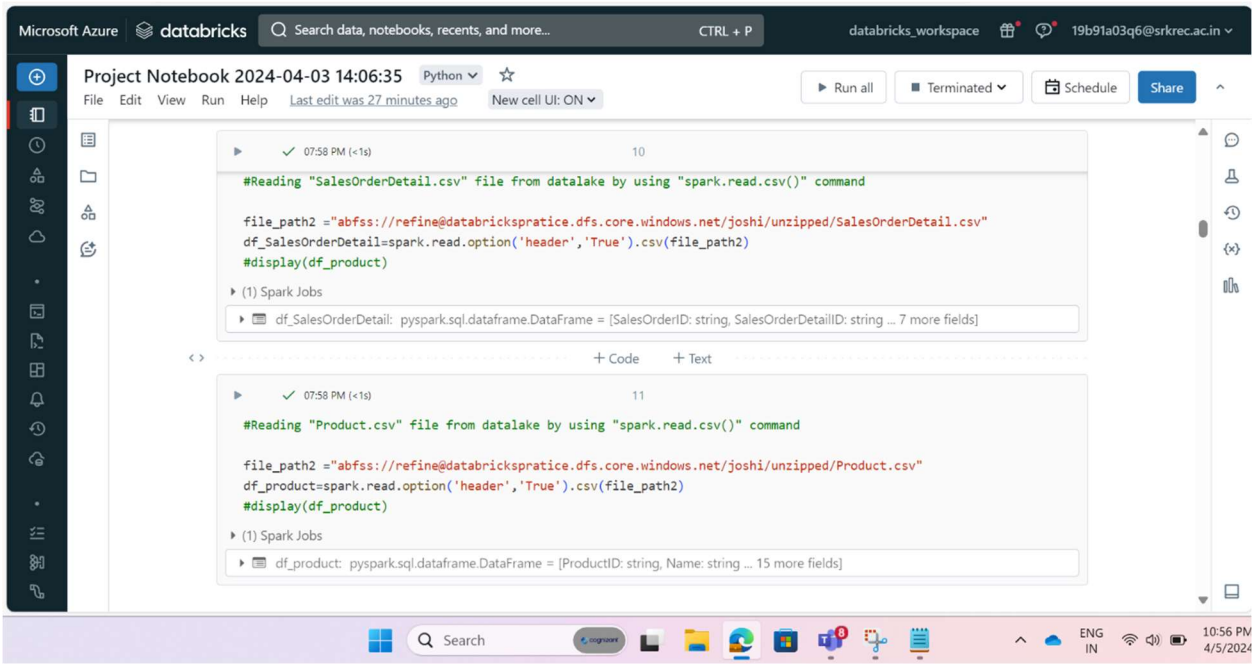

Step 33: Above code is used to fetch files from data lake into data bricks to picks file for cleaning and transformation.



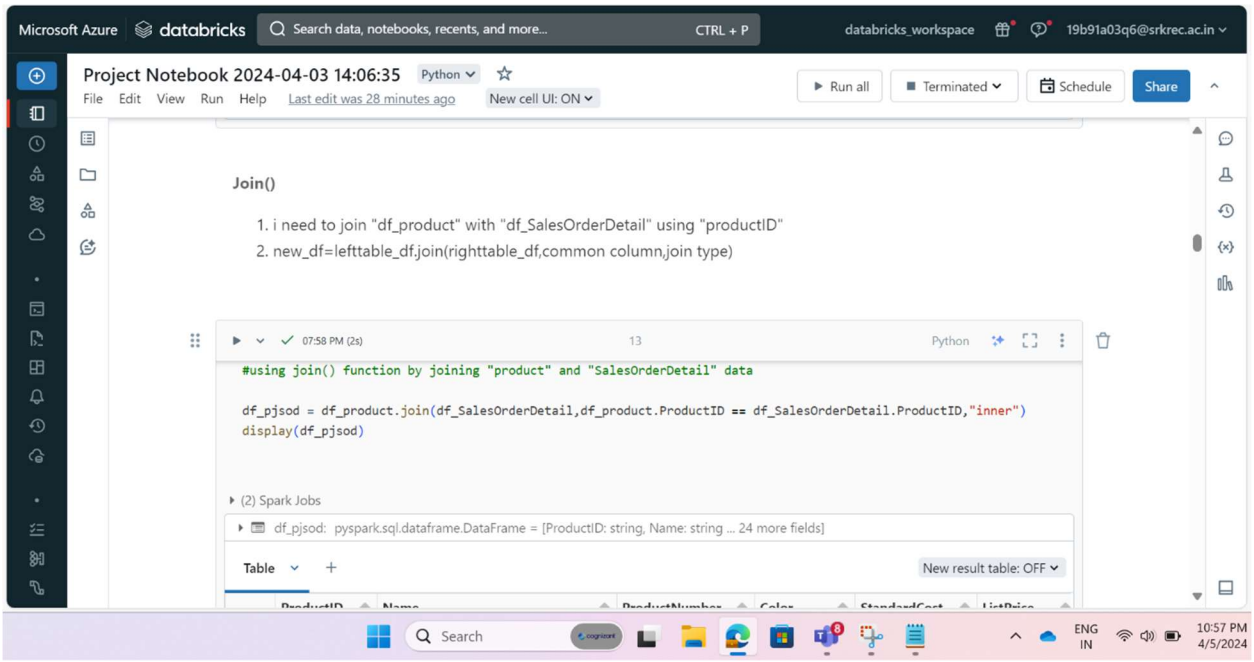
Step 34: Above code in used to read raw file from data lake and add header to that raw file.



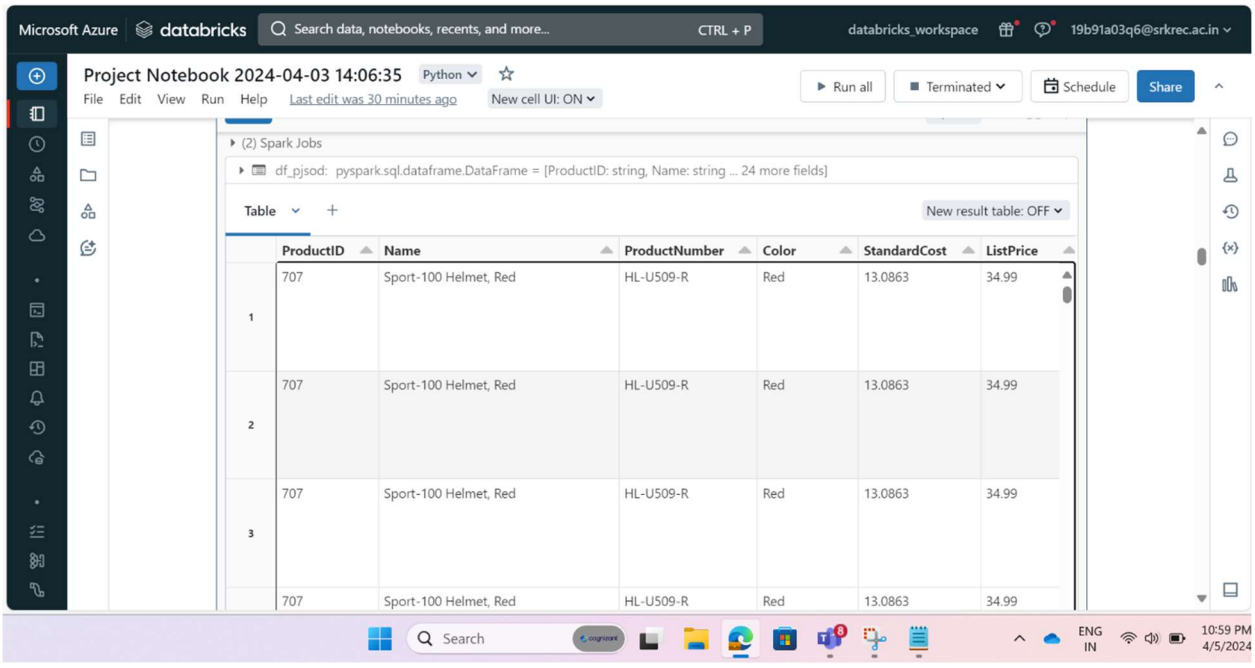

Step 35: Above picture will show the output for above code to add header to raw file.



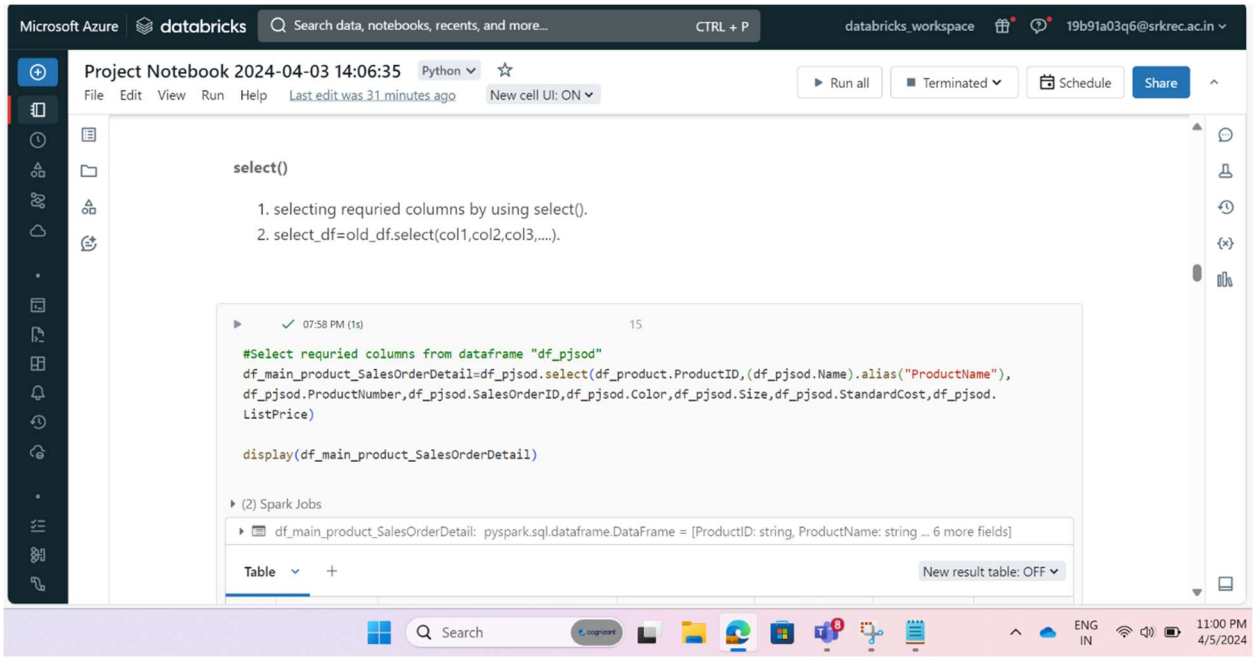

Step 36: Above code is used for reading csv files from datalake into data bricks.



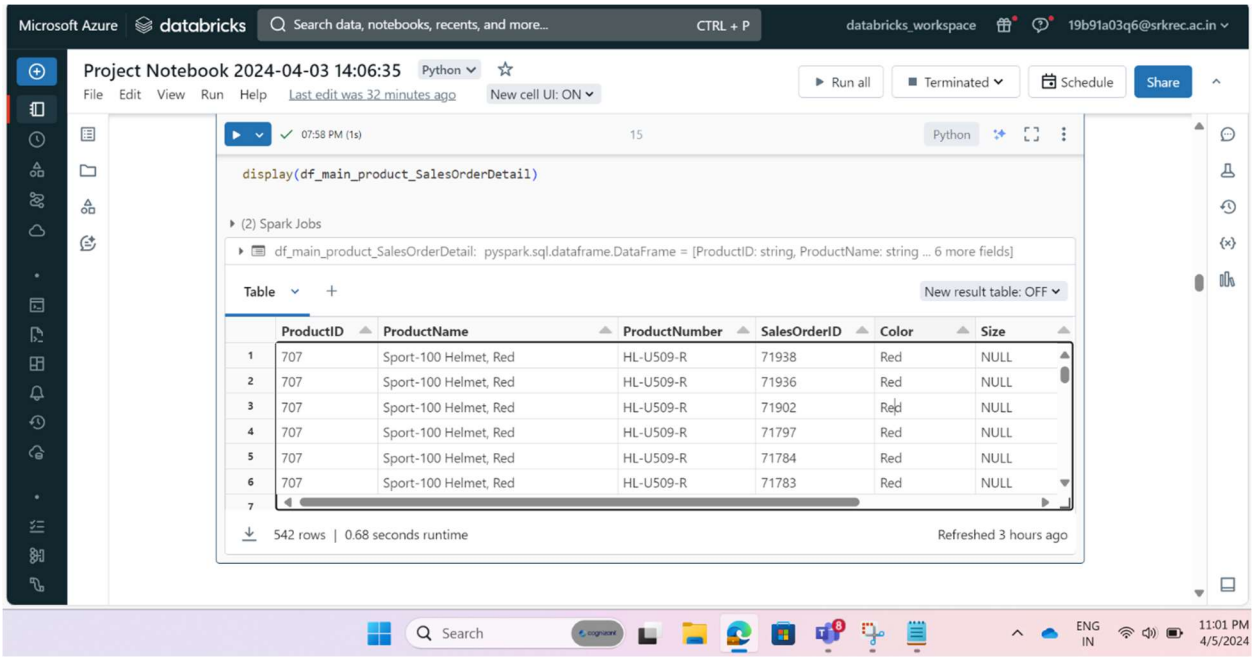

Step 37: Above code is used for apply join between to datasets “product” and “salesorder details” files.



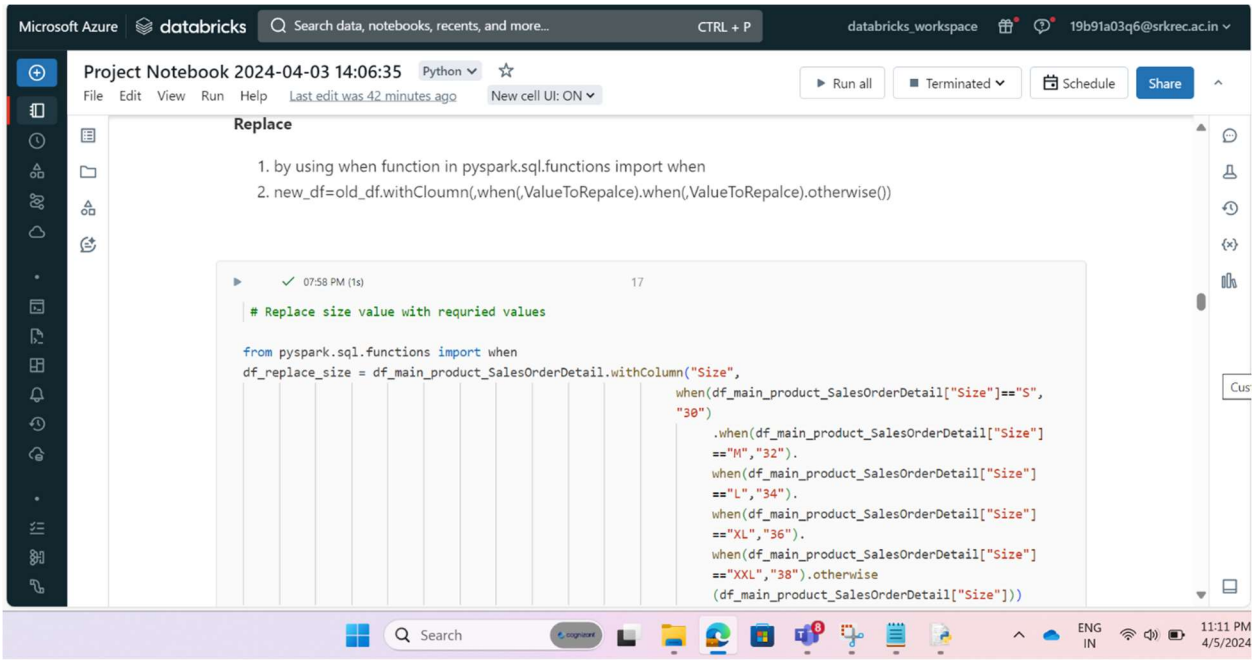

Step 38: After applying join by above code the output appear is like this.



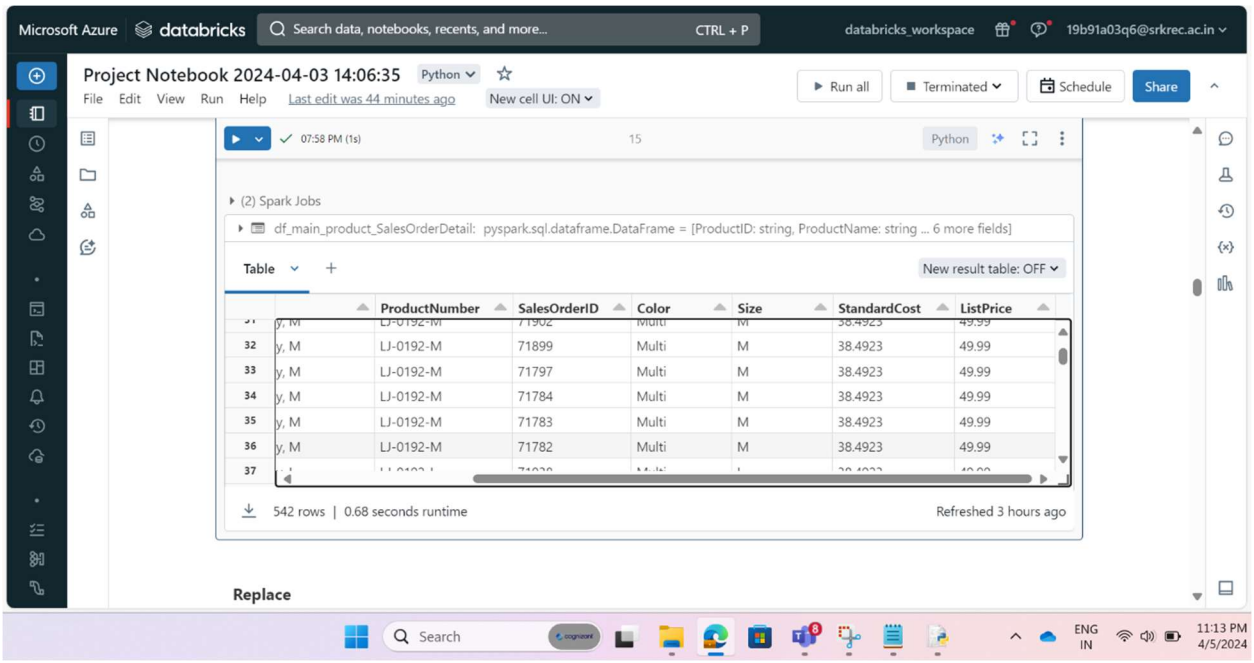

Step 39: Above code is to select requried columns from that join table for our data cleaning and data transformation .



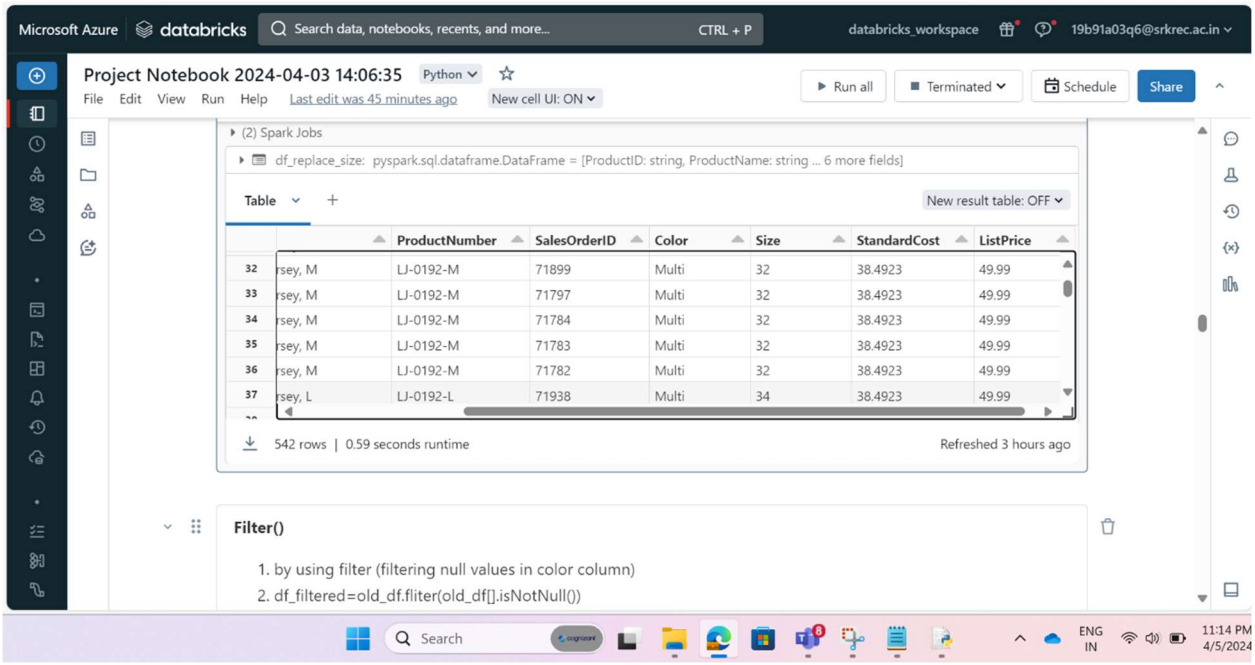

Step 40: After selecting required columns by using above code we got this output.



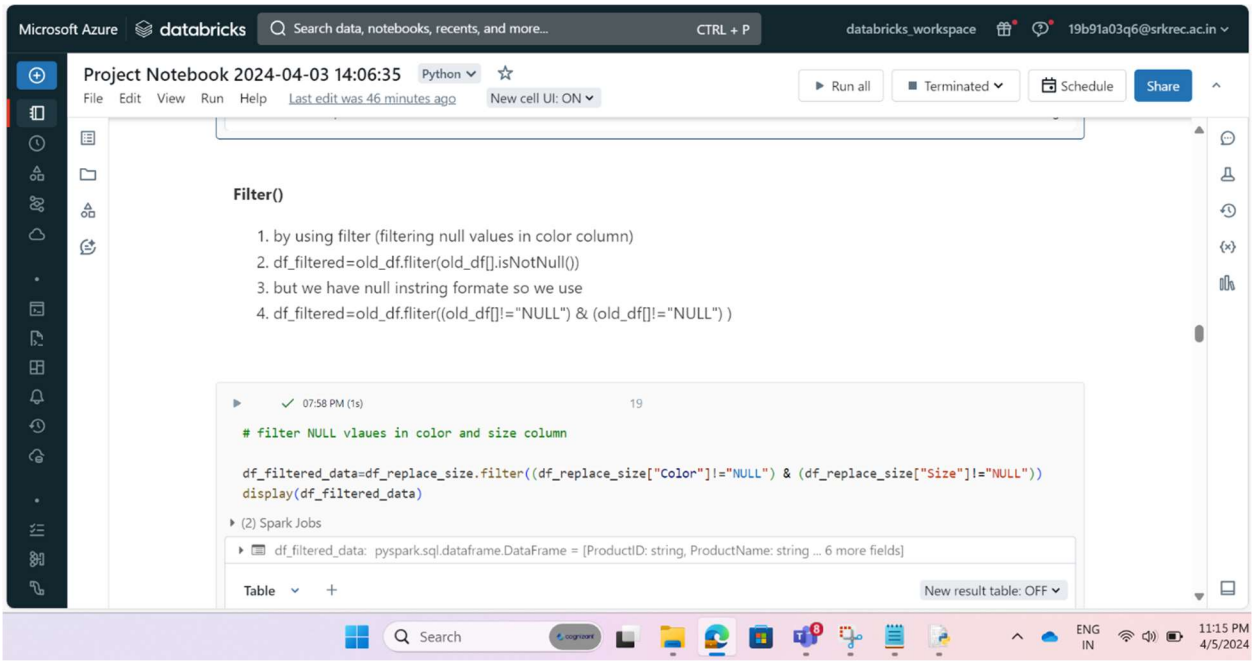

Step 41: Above code is used for cleaning data by replace size values which are letter with a number.



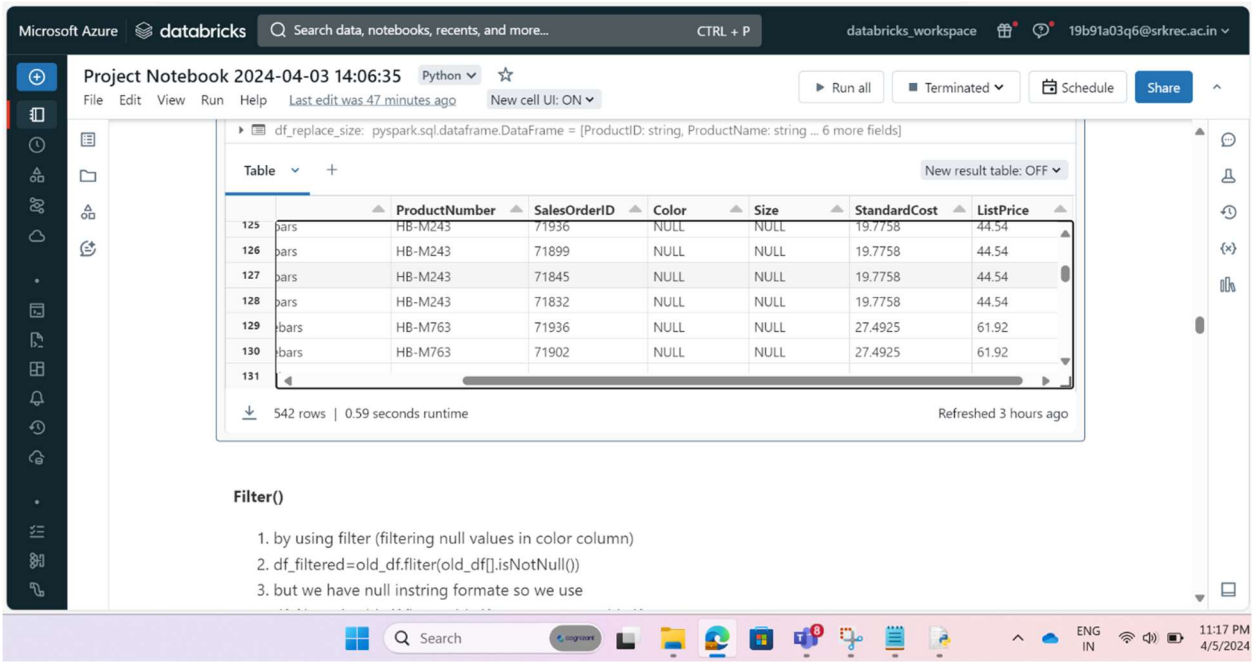

Step 42: picture is showing size column data before applying transformation.



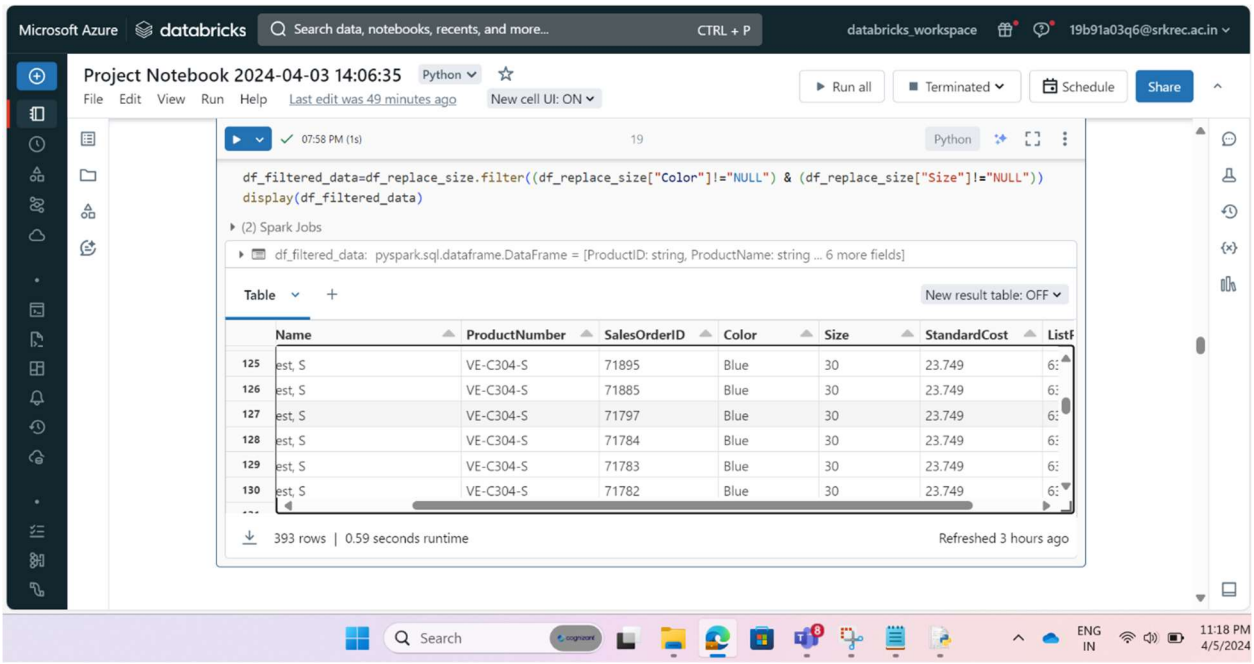

Step 43: picture is showing size column data after applying transformation.



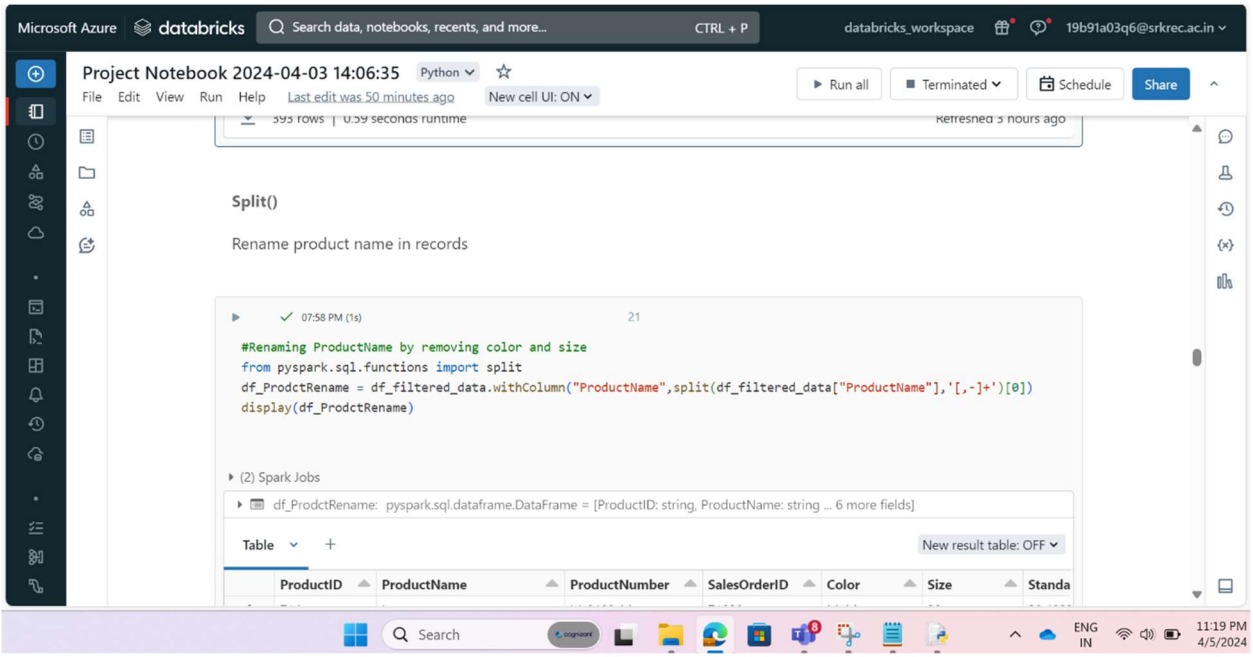

Step 44: Above code is used for filtering color,size column recorders whivh are NULL.



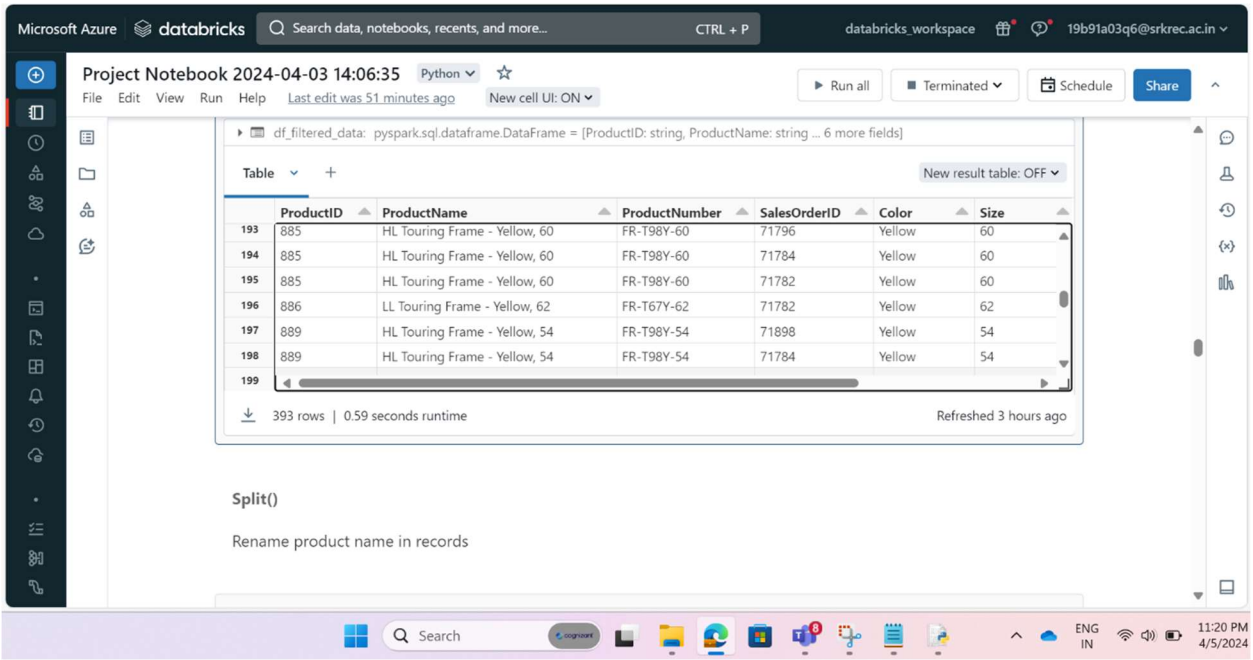

Step 45 : Above picture shows output before filtering null values.



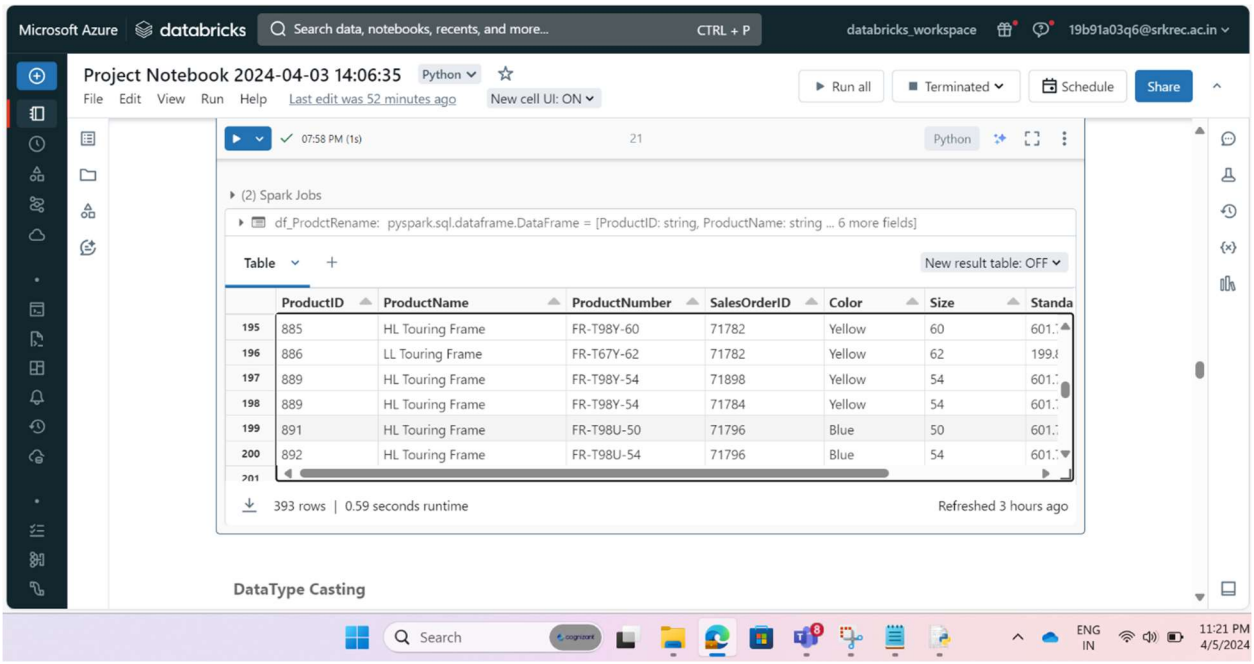

Step 46: Above picture shows output after filtering null values.



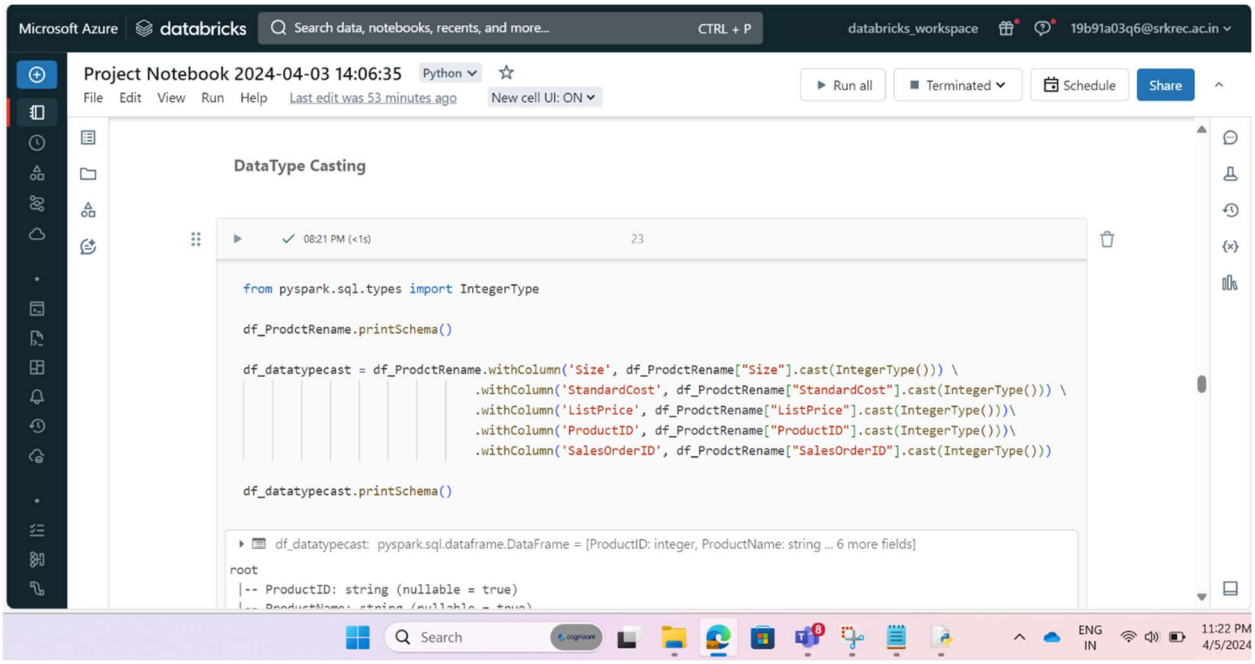

Step 47: Above code is to clean product name is useful manner by using split () function in pyspark.



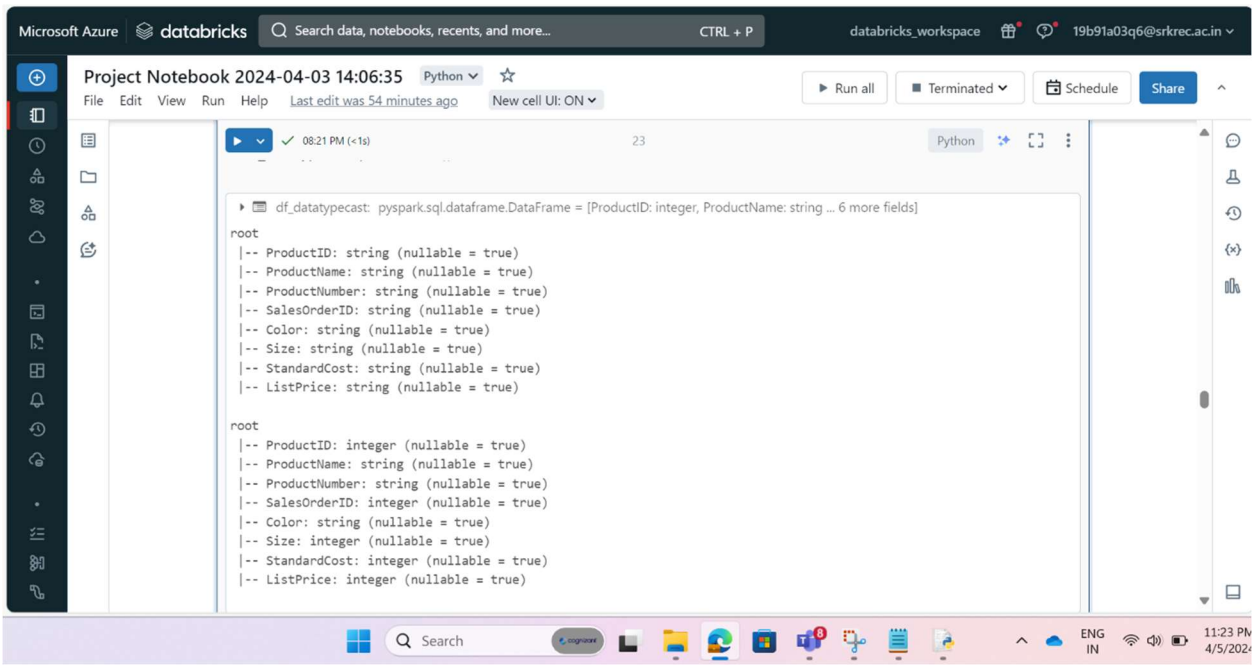

Step 48: Above picture shows output before applying split() function.



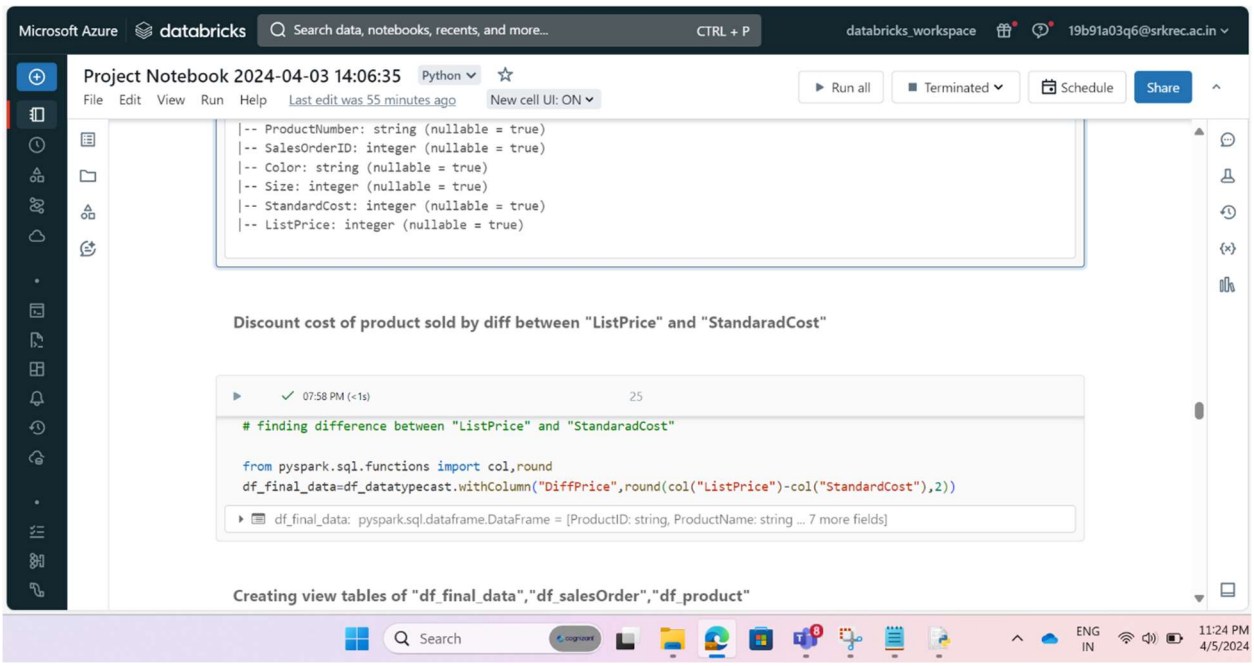

Step 49: Above picture shows output after applying split() function.



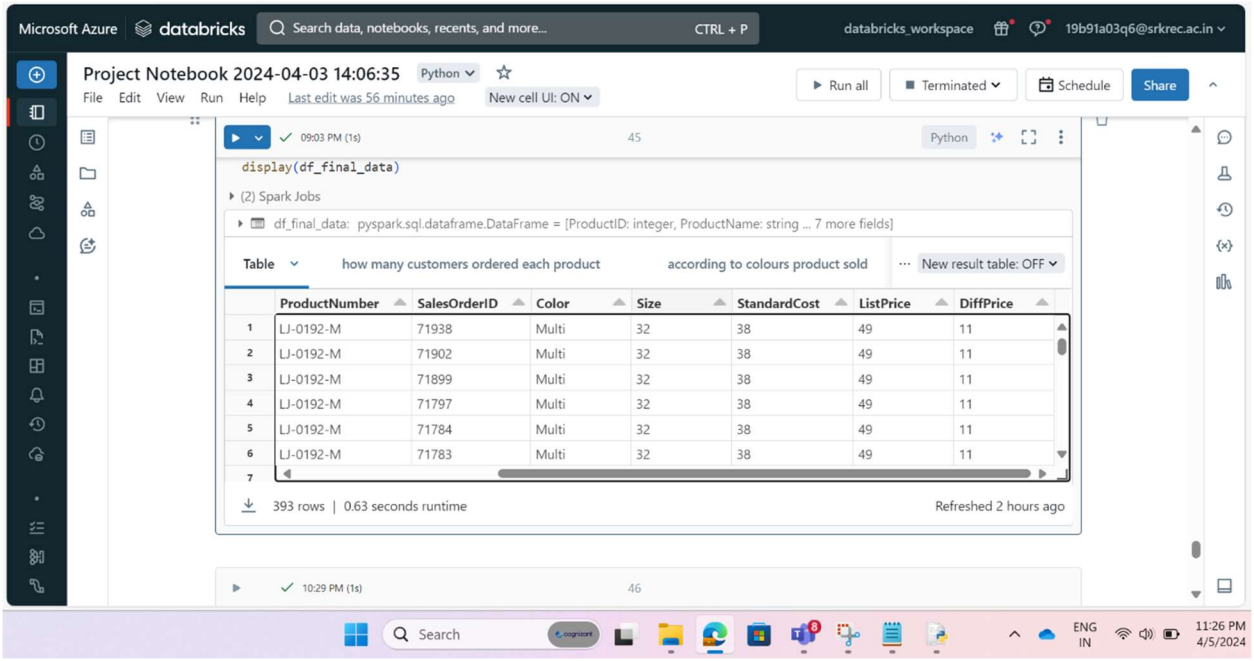

Step 50: above code is to type cat datatype form string to integers.




Step 51: result of above code after and before type casting schema.

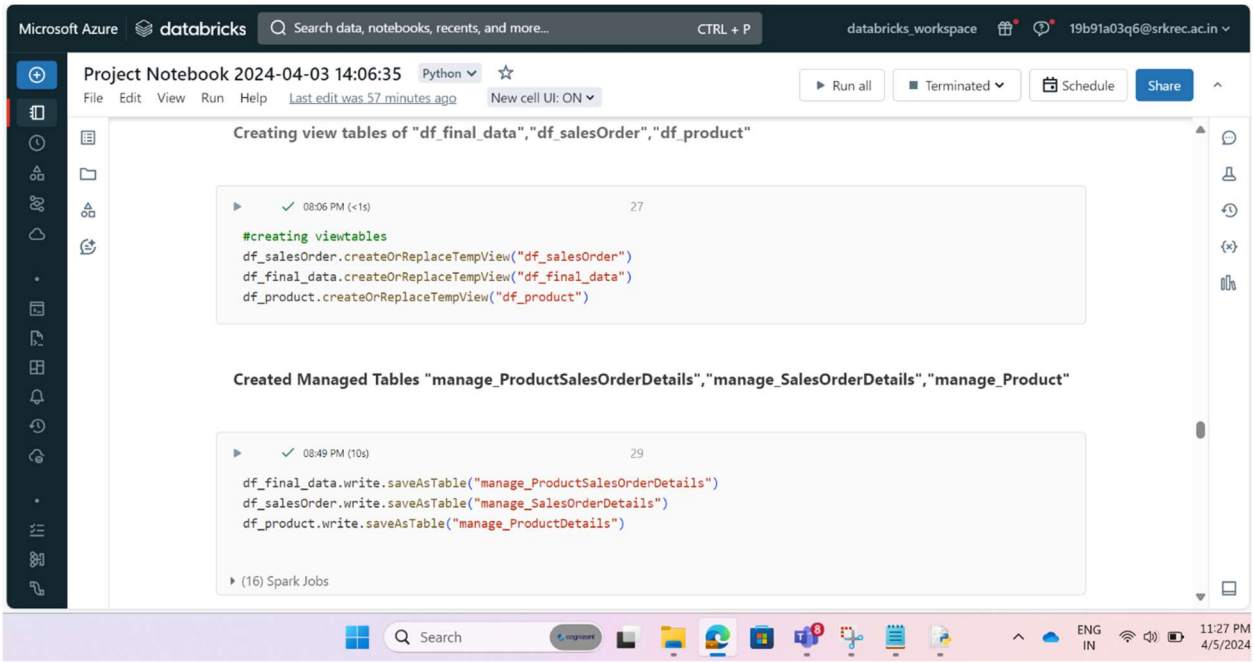


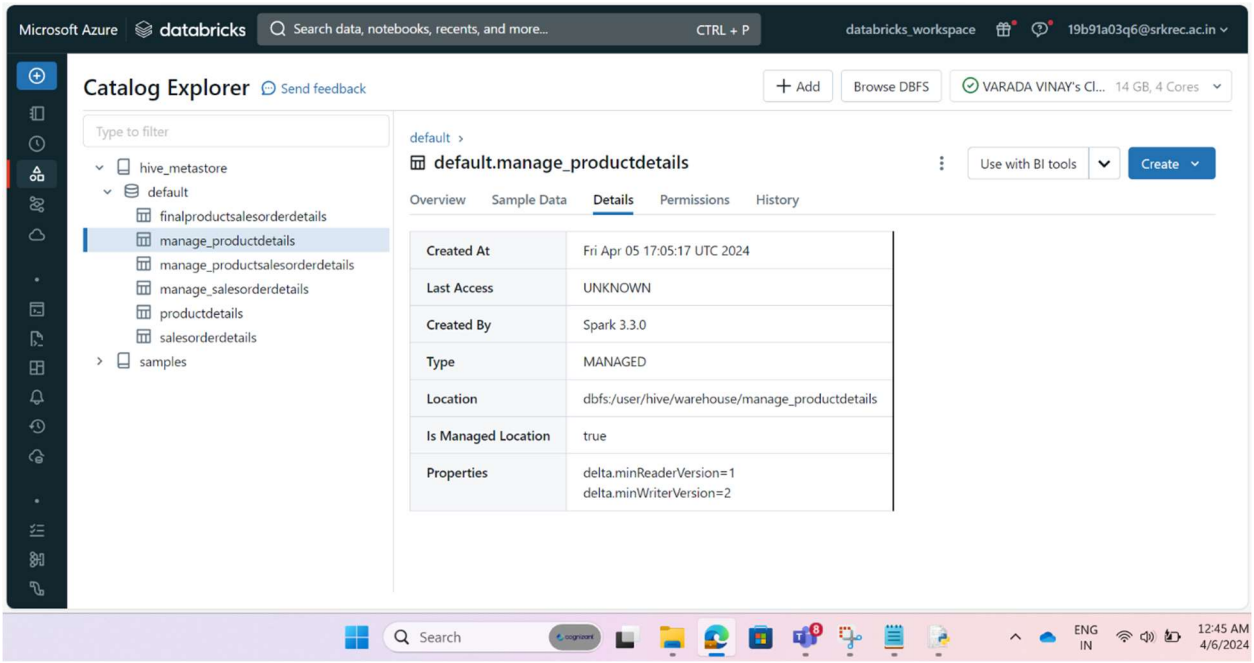

Step 52: Find different between listprice and standardprice .

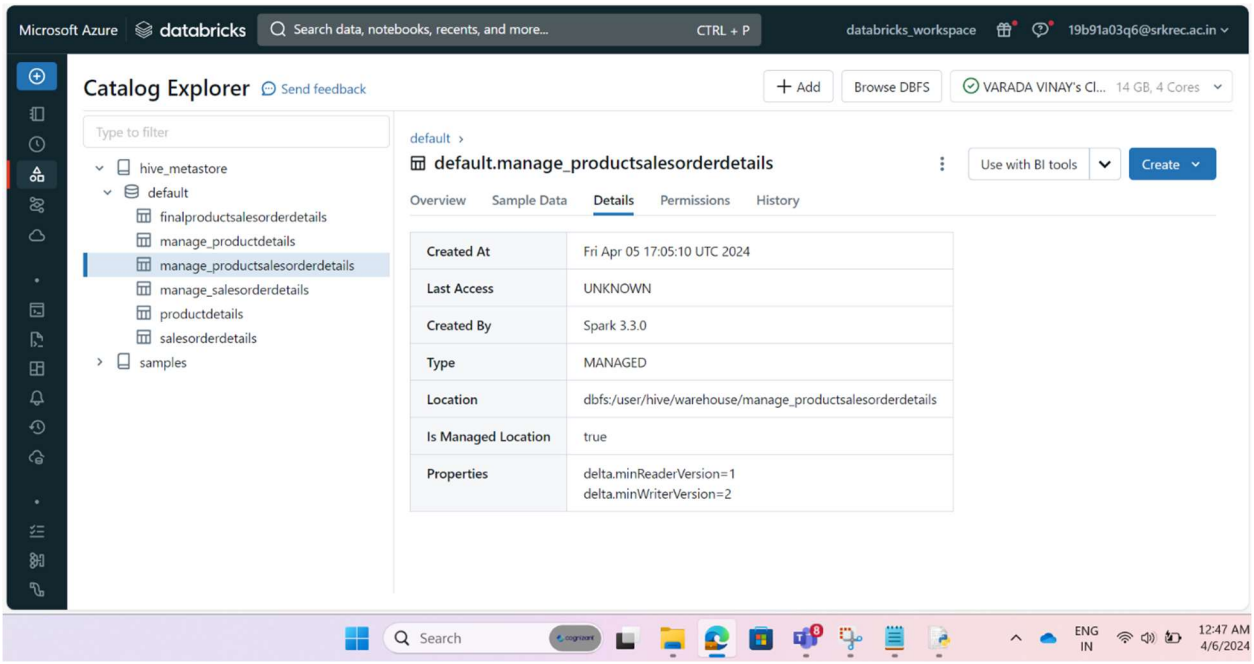


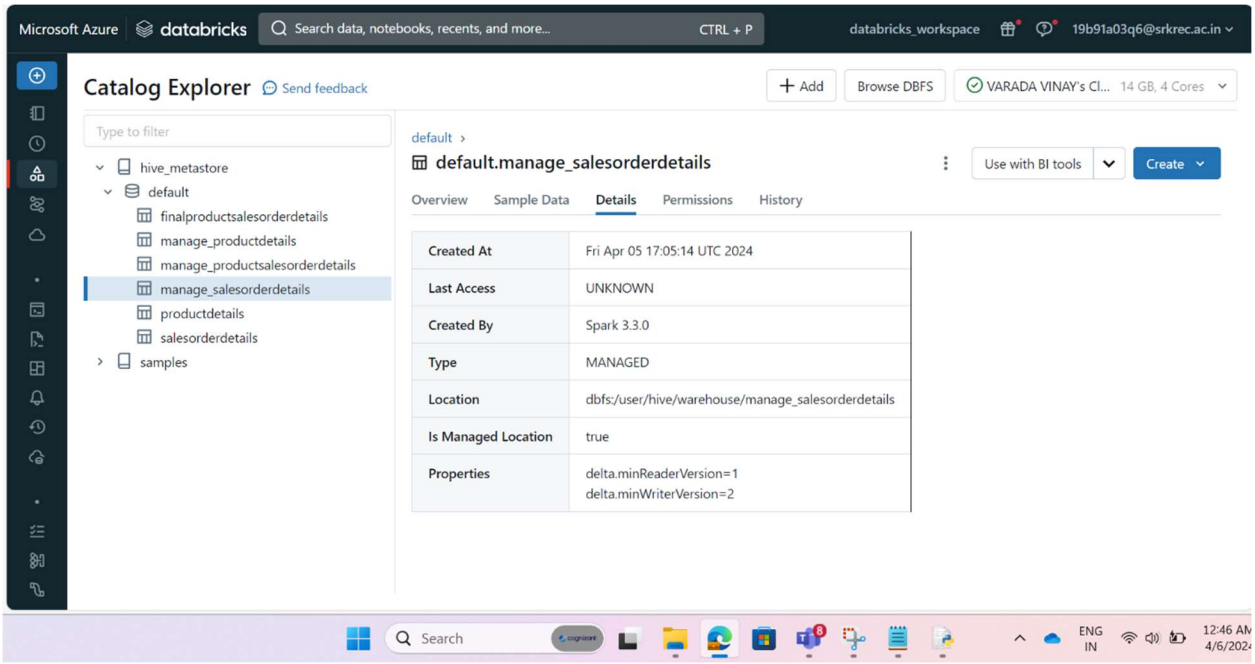

Step 53: Output for above code using to find difference between listprice and standardprice.

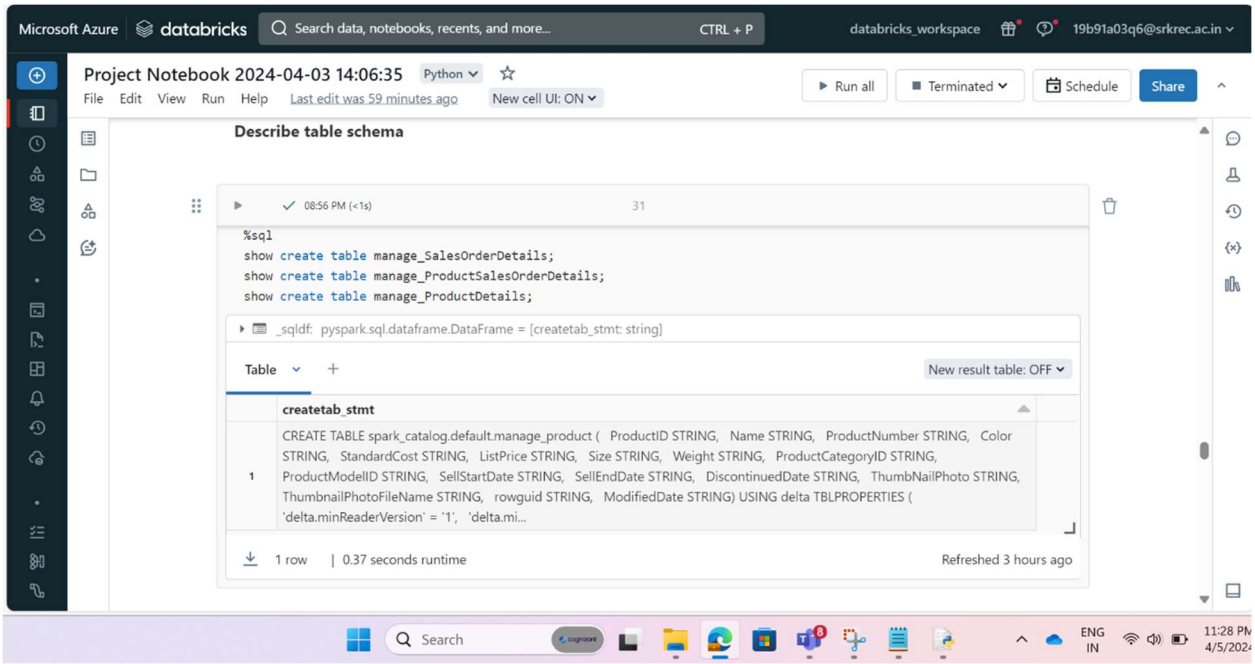
**Creating delta tables (manage tables) in data bricks in hive meta store which are stored in DBFS.**



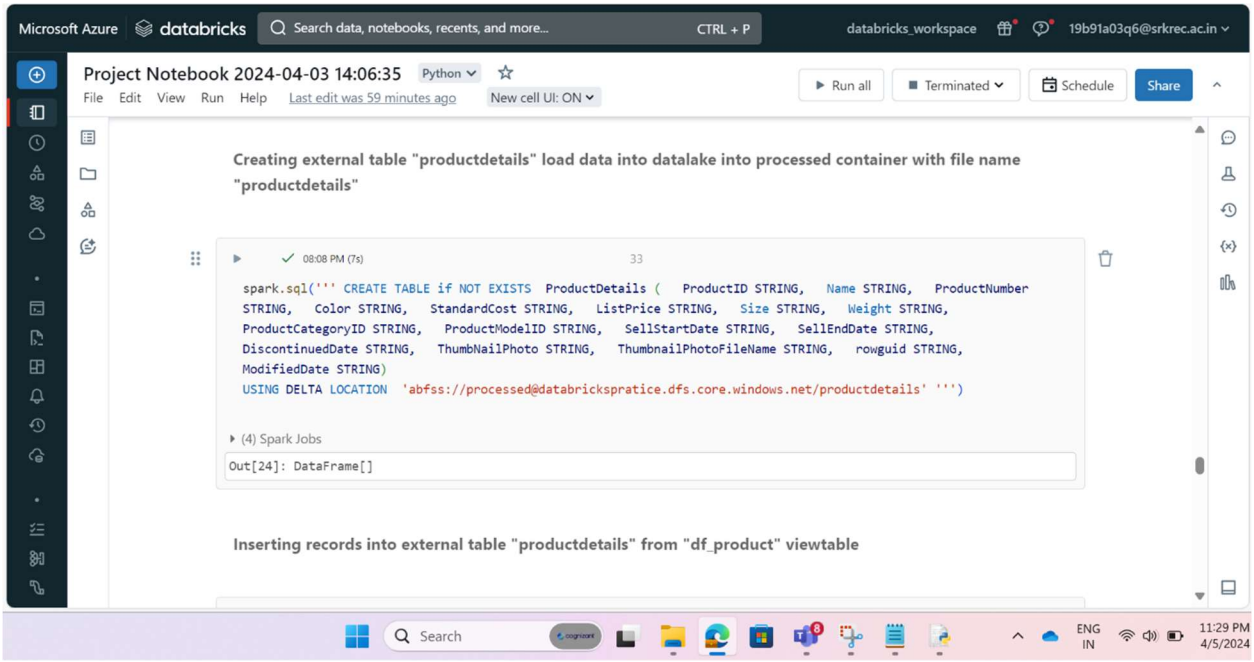

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.manage_productdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Created At
Fri Apr 05 17:05:17 UTC 2024
Last Access
UNKNOWN
Created By
Spark 3.3.0
Type
MANAGED
Location
dbfs:/user/hive/warehouse/manage_productdetails
Is Managed Location
true
Properties
delta.minReaderVersion=1
delta.minWriterVersion=2
Search
ENG
IN
12:45 AM
4/6/2024

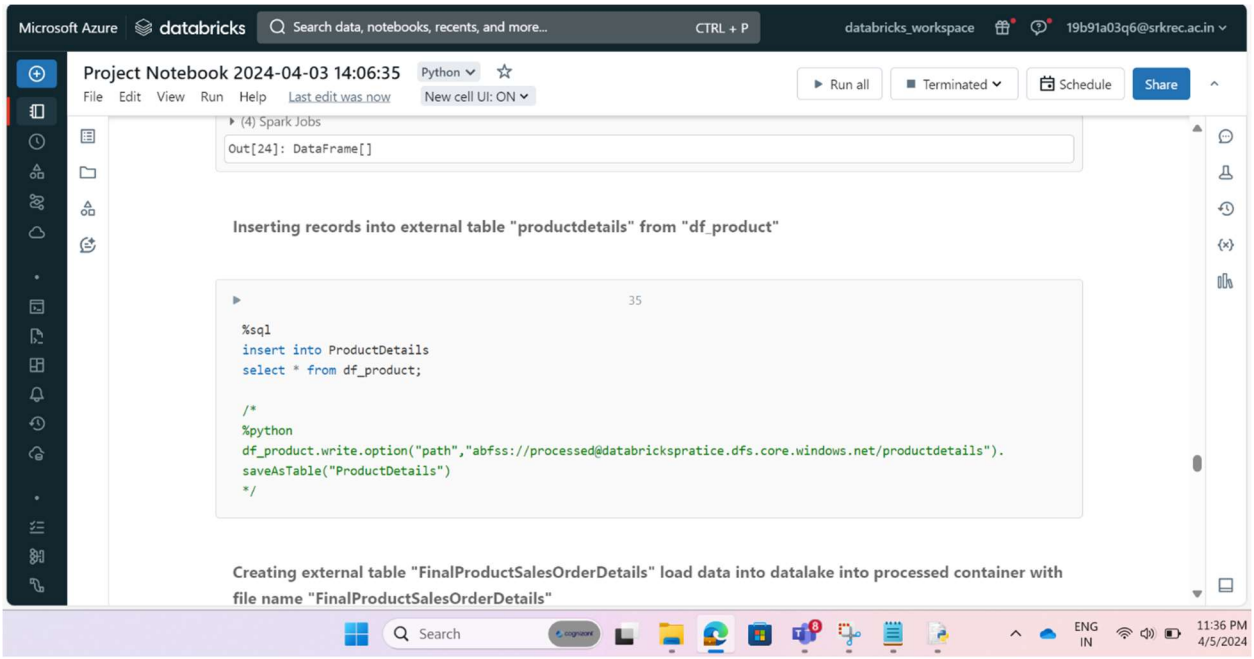

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.manage_productsalesorderdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Created At
Fri Apr 05 17:05:10 UTC 2024
Last Access
UNKNOWN
Created By
Spark 3.3.0
Type
MANAGED
Location
dbfs:/user/hive/warehouse/manage_productsalesorderdetails
Is Managed Location
true
Properties
delta.minReaderVersion=1
delta.minWriterVersion=2
Search
ENG
IN
12:47 AM
4/6/2024

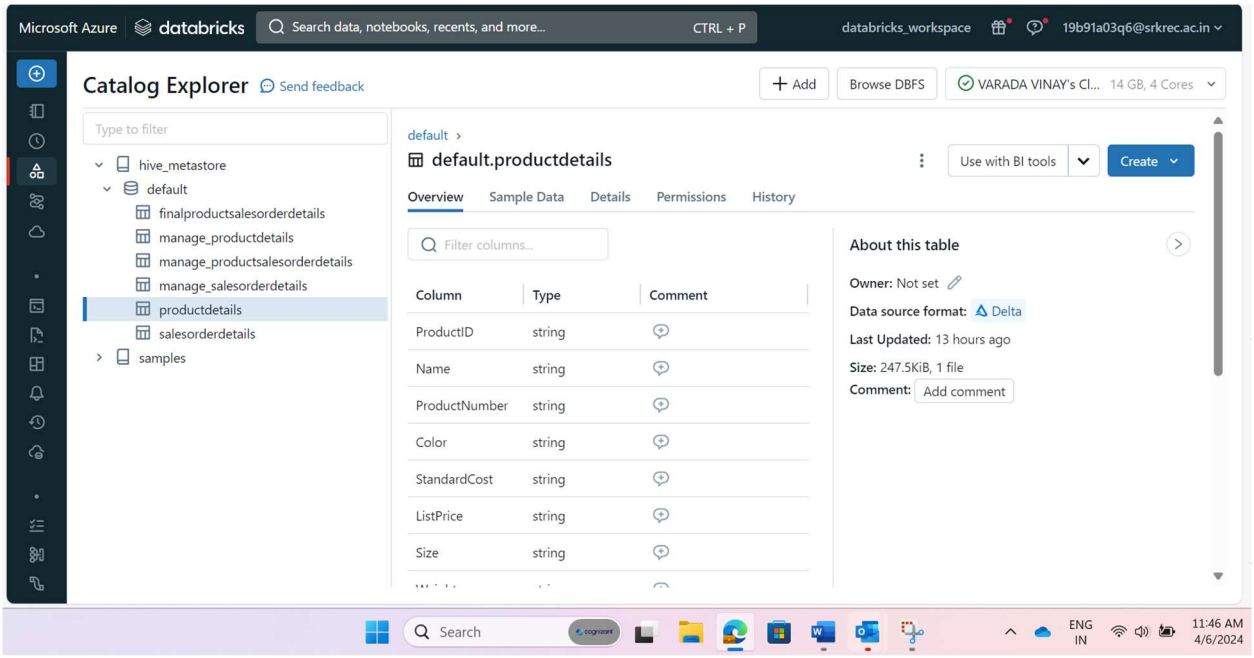

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.manage_salesorderdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Created At
Fri Apr 05 17:05:14 UTC 2024
Last Access
UNKNOWN
Created By
Spark 3.3.0
Type
MANAGED
Location
dbfs:/user/hive/warehouse/manage_salesorderdetails
Is Managed Location
true
Properties
delta.minReaderVersion=1
delta.minWriterVersion=2
Search
ENG
IN
12:46 AM
4/6/2024

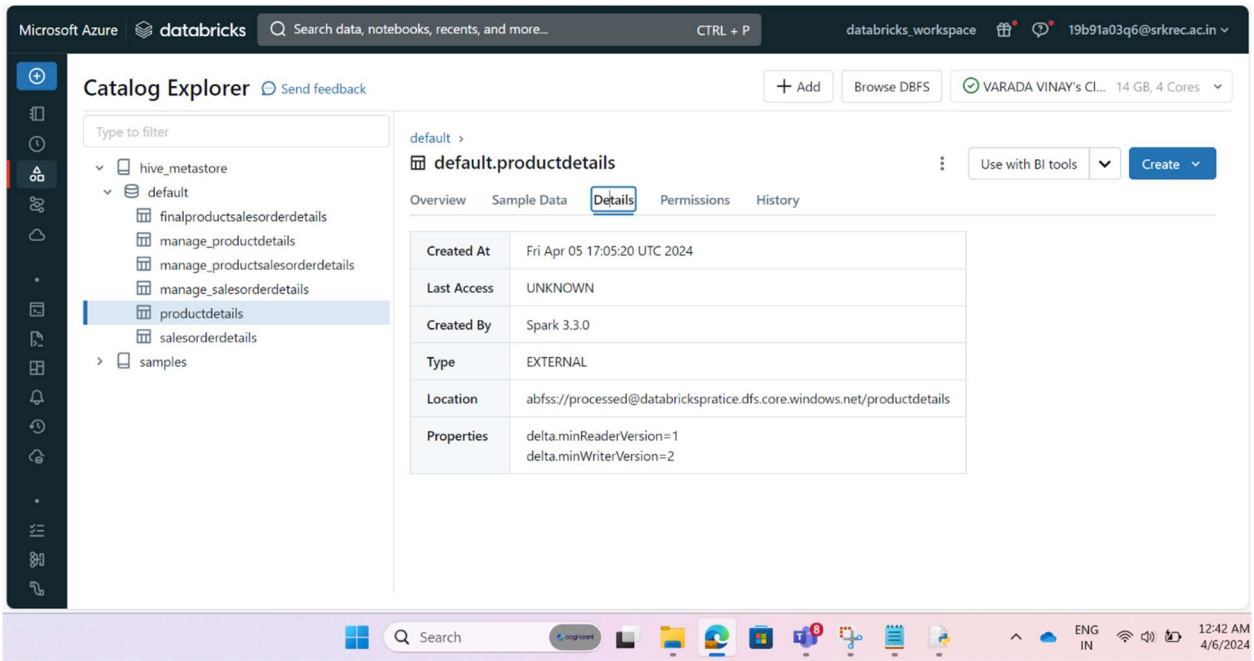

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File
Edit
View
Run
Help
Last edit was 59 minutes ago
New cell UI: ON
Run all
Terminated
Schedule
Share
Describe table schema
08:56 PM (<1s)
31
%sql
show create table manage_SalesOrderDetails;
show create table manage_ProductSalesOrderDetails;
show create table manage_ProductDetails;
_sqldf: pyspark.sql.dataframe.DataFrame = [createtab_stmt: string]
Table
New result table: OFF
createtab_stmt
1
CREATE TABLE spark_catalog.default.manage_product ( ProductID STRING, Name STRING, ProductNumber STRING, Color STRING, StandardCost STRING, ListPrice STRING, Size STRING, Weight STRING, ProductCategoryID STRING, ProductModelID STRING, SellStartDate STRING, SellEndDate STRING, DiscontinuedDate STRING, ThumbNailPhoto STRING, ThumbnailPhotoFileName STRING, rowguid STRING, ModifiedDate STRING) USING delta TBLPROPERTIES ( 'delta.minReaderVersion' = '1', 'delta.mi...
1 row | 0.37 seconds runtime
Refreshed 3 hours ago
Search
ENG
IN
11:28 PM
4/5/2024

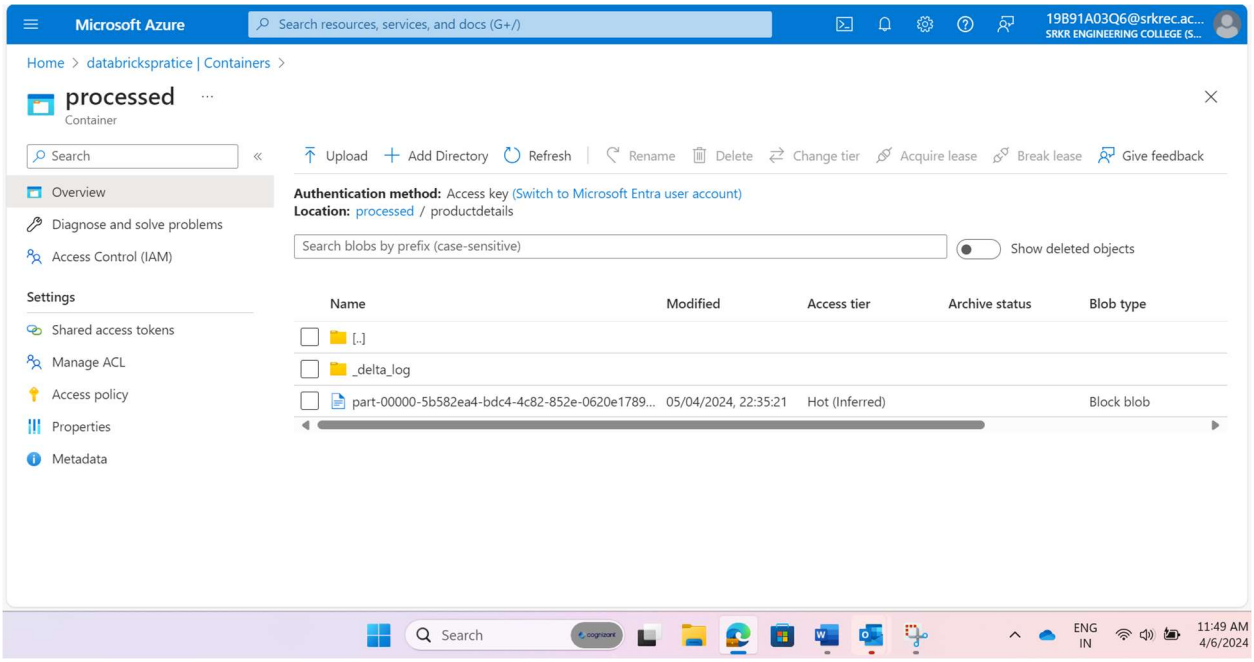
**Creating delta tables (external tables) in data bricks and there results of parquet files in datalake.**

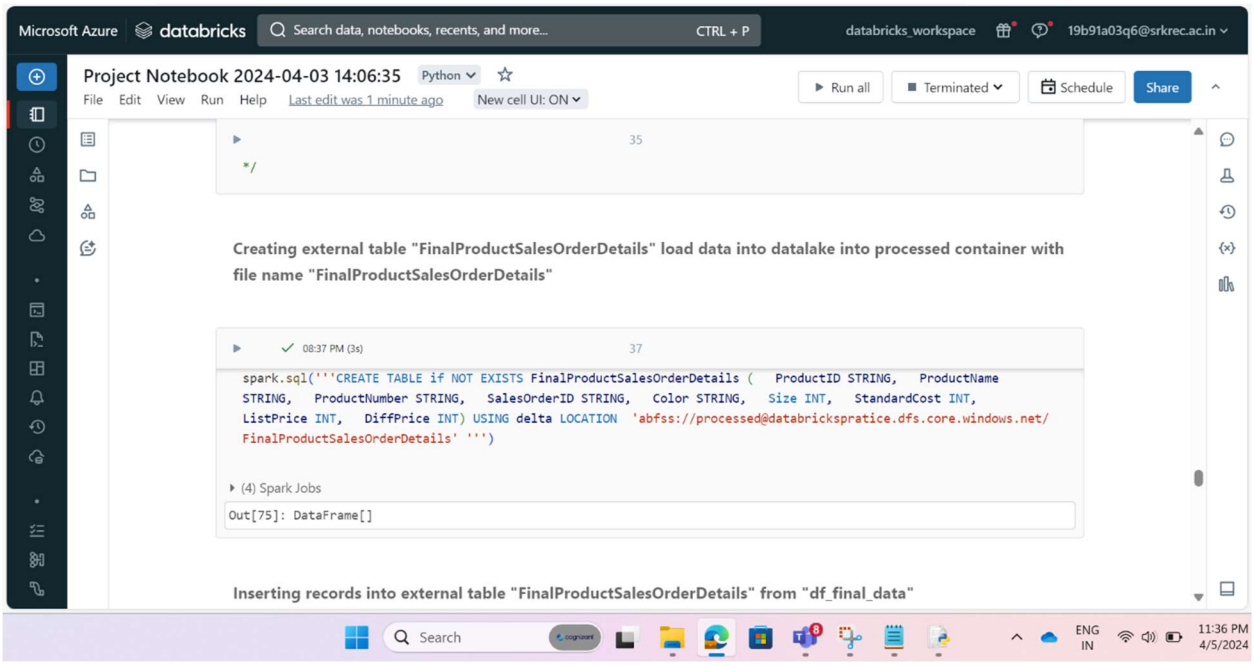


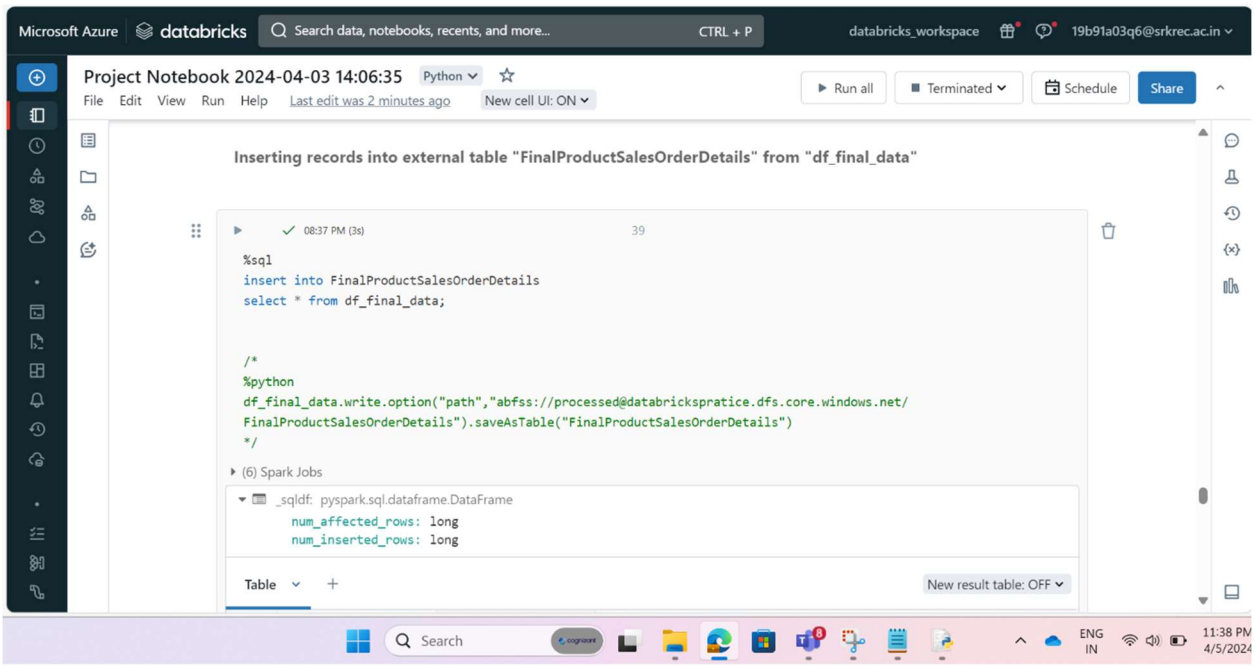



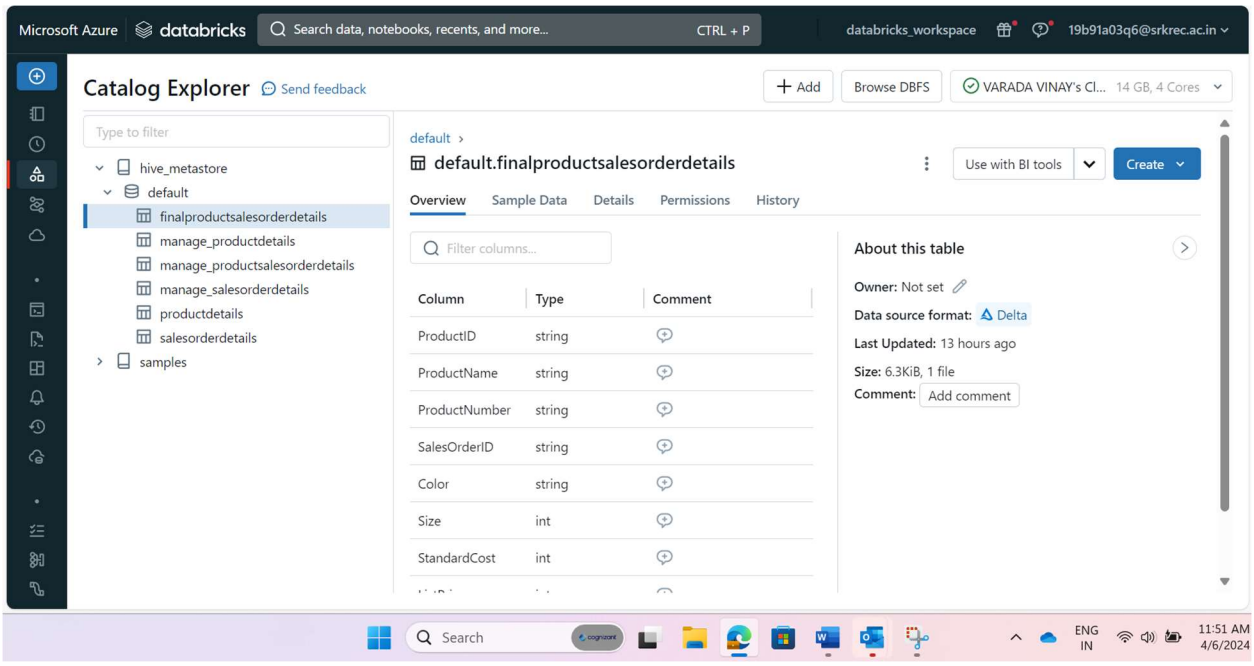

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.productdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Filter columns...
Column
Type
Comment
ProductID string
Name string
ProductNumber string
Color string
StandardCost string
ListPrice string
Size string
About this table
Owner: Not set
Data source format: Delta
Last Updated: 13 hours ago
Size: 247.5KiB, 1 file
Comment: Add comment
Search
ENG
IN
11:46 AM
4/6/2024

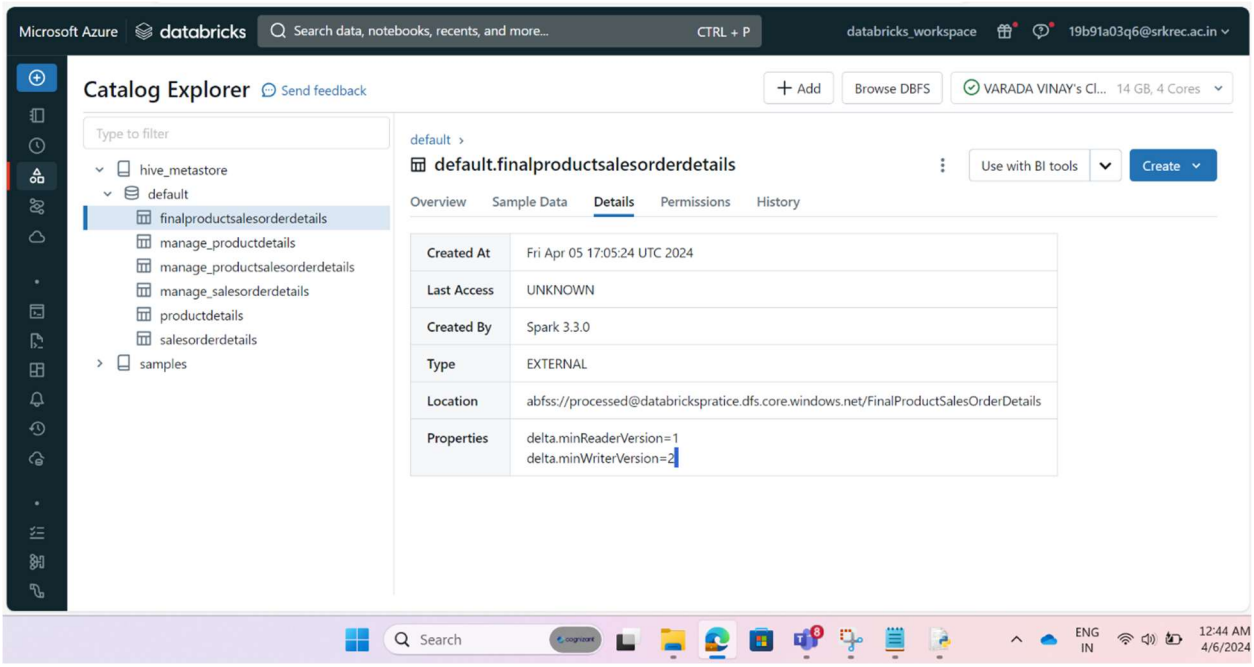

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.productdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Created At Fri Apr 05 17:05:20 UTC 2024
Last Access UNKNOWN
Created By Spark 3.3.0
Type EXTERNAL
Location abfss://processed@databrickspratice.dfs.core.windows.net/productdetails
Properties delta.minReaderVersion=1
delta.minWriterVersion=2
Search
ENG
IN
12:42 AM
4/6/2024

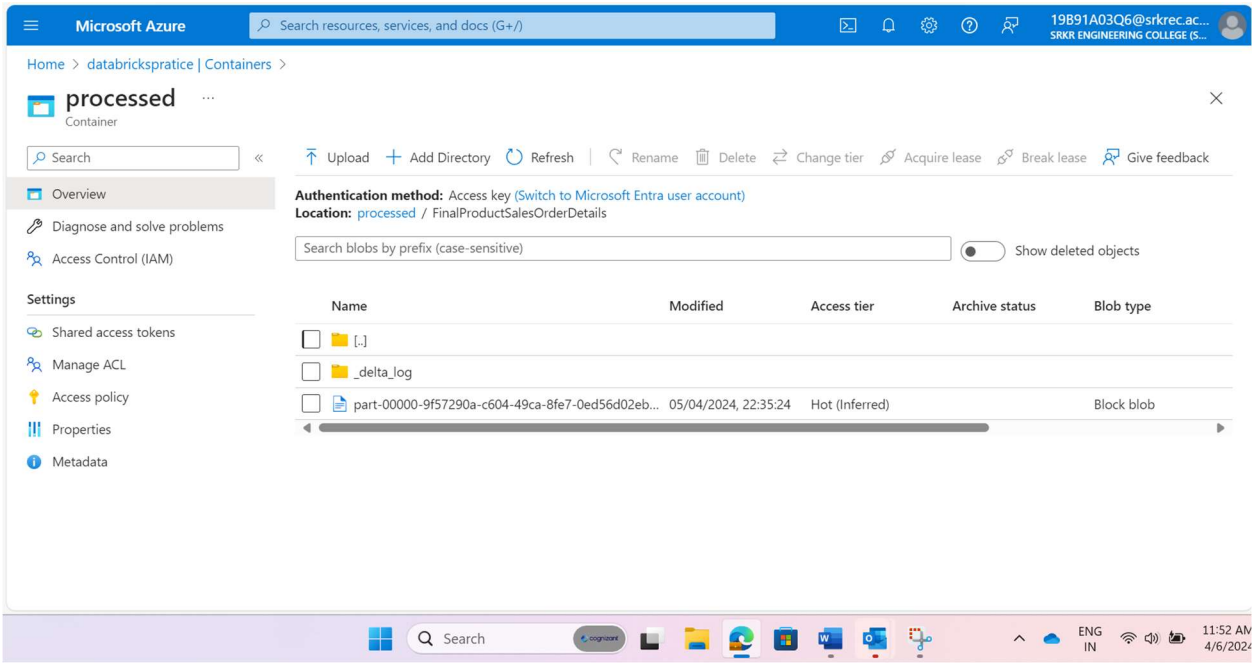

Microsoft Azure
Search resources, services, and docs (G+/)
19B91A03Q6@srkrec.ac...
SRKR ENGINEERING COLLEGE (S...
Home > databrickspratice | Containers >
processed
Container
Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: processed / productdetails
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
[..]
_delta_log
part-00000-5b582ea4-bdc4-4c82-852e-0620e1789...
05/04/2024, 22:35:21
Hot (Inferred)
Block blob
Search
ENG
IN
11:49 AM
4/6/2024

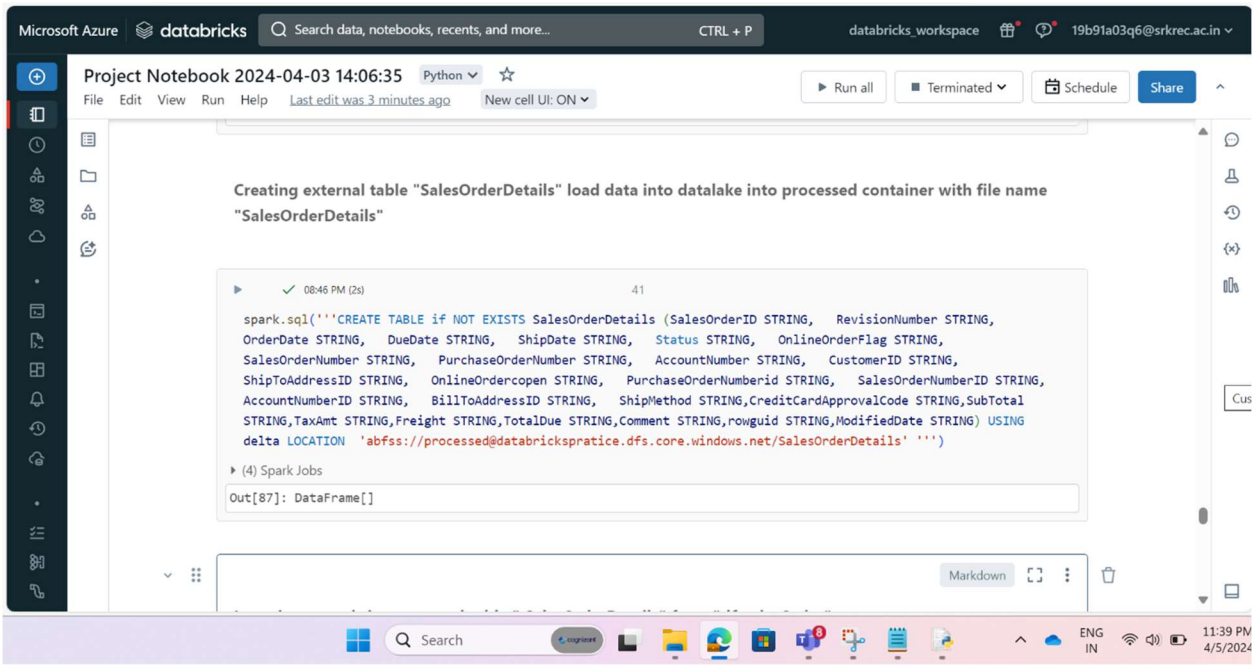

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File Edit View Run Help
Last edit was 1 minute ago
New cell UI: ON
Run all
Terminated
Schedule
Share
35
*/
Creating external table "FinalProductSalesOrderDetails" load data into datalake into processed container with file name "FinalProductSalesOrderDetails"
08:37 PM (3s)
37
spark.sql('''CREATE TABLE if NOT EXISTS FinalProductSalesOrderDetails ( ProductID STRING, ProductName STRING, ProductNumber STRING, SalesOrderID STRING, Color STRING, Size INT, StandardCost INT, ListPrice INT, DiffPrice INT) USING delta LOCATION 'abfss://processed@databrickspratice.dfs.core.windows.net/FinalProductSalesOrderDetails' ''')
(4) Spark Jobs
Out[75]: DataFrame[]
Inserting records into external table "FinalProductSalesOrderDetails" from "df_final_data"
Search
ENG
IN
11:36 PM
4/5/2024

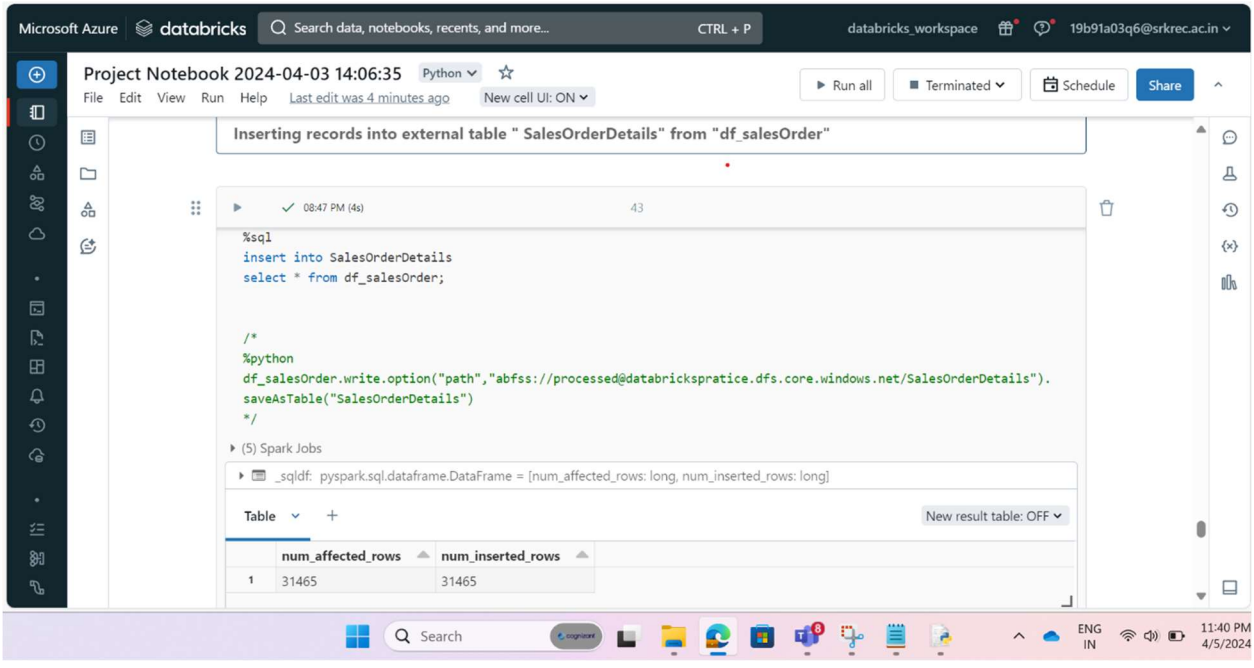

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File Edit View Run Help
Last edit was 2 minutes ago
New cell UI: ON
Run all
Terminated
Schedule
Share
Inserting records into external table "FinalProductSalesOrderDetails" from "df_final_data"
08:37 PM (3s)
39
%sql
insert into FinalProductSalesOrderDetails
select * from df_final_data;
/*
%python
df_final_data.write.option("path","abfss://processed@databrickspratice.dfs.core.windows.net/
FinalProductSalesOrderDetails").saveAsTable("FinalProductSalesOrderDetails")
*/
(6) Spark Jobs
_sqldf: pyspark.sql.dataframe.DataFrame
num_affected_rows: long
num_inserted_rows: long
Table
New result table: OFF
Search
ENG
IN
11:38 PM
4/5/2024

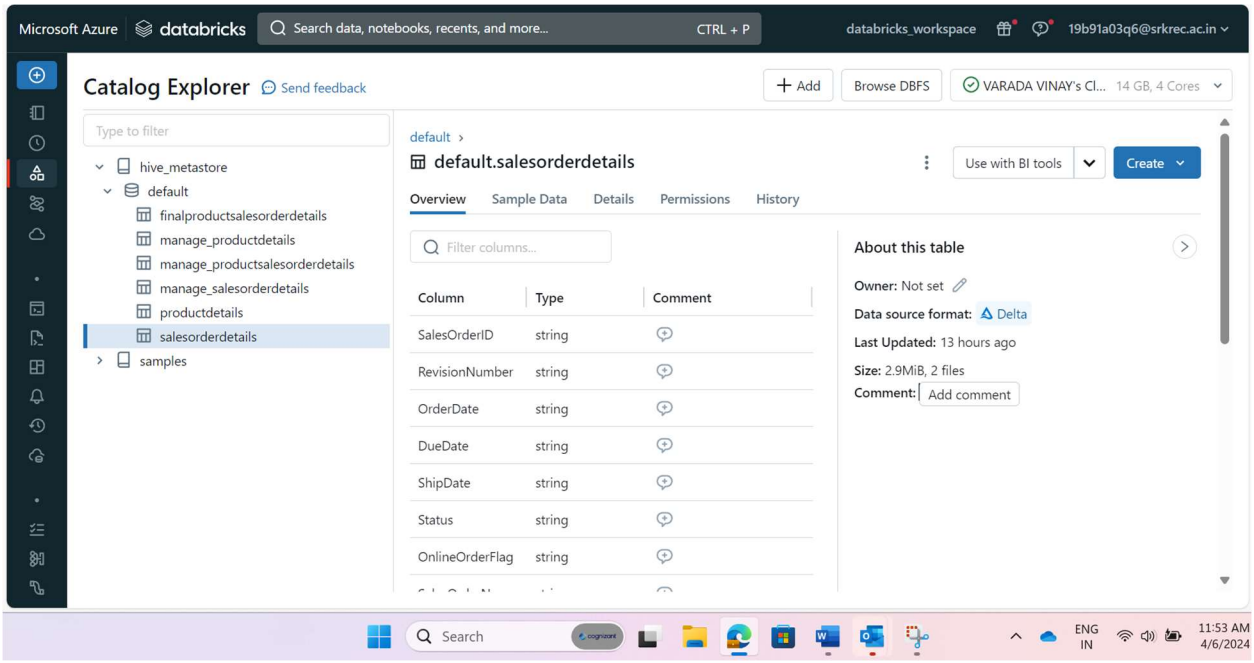

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default
default.finalproductsalesorderdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Filter columns...
Column
Type
Comment
ProductID string
ProductName string
ProductNumber string
SalesOrderID string
Color string
Size int
StandardCost int
About this table
Owner: Not set
Data source format: Delta
Last Updated: 13 hours ago
Size: 6.3KiB, 1 file
Comment: Add comment
Search
ENG
IN
11:51 AM
4/6/2024

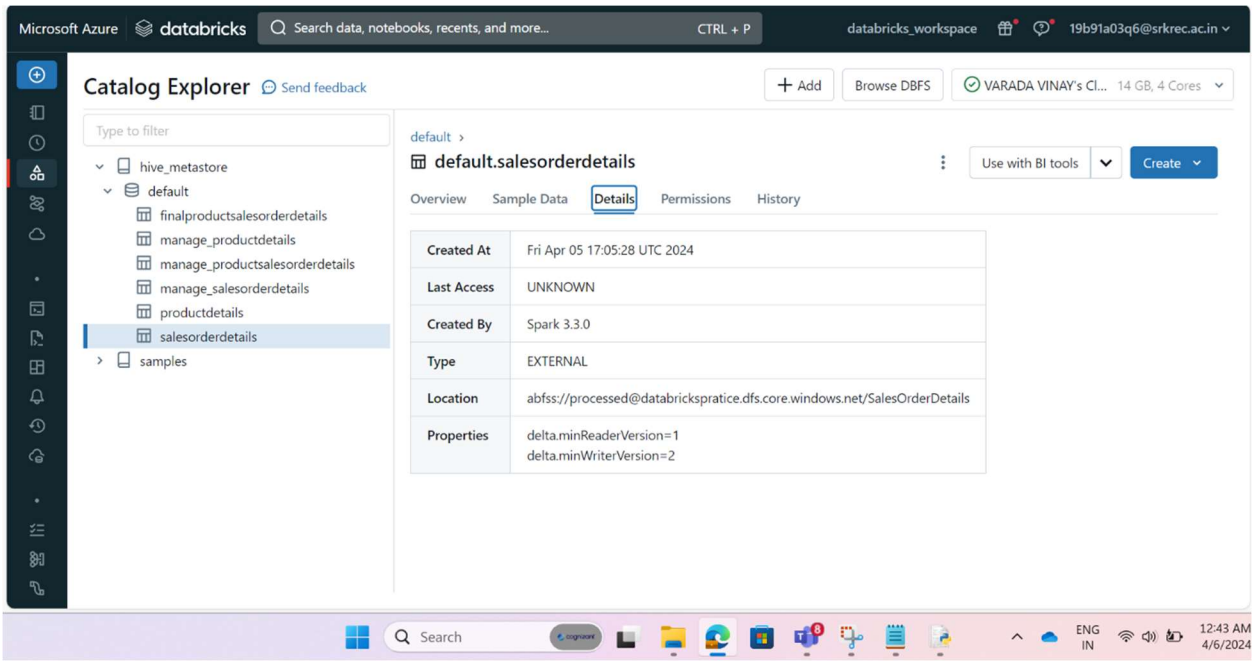

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl... 14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.finalproductsalesorderdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Created At
Fri Apr 05 17:05:24 UTC 2024
Last Access
UNKNOWN
Created By
Spark 3.3.0
Type
EXTERNAL
Location
abfss://processed@databrickspratice.dfs.core.windows.net/FinalProductSalesOrderDetails
Properties
delta.minReaderVersion=1
delta.minWriterVersion=2
Search
ENG
IN
12:44 AM
4/6/2024

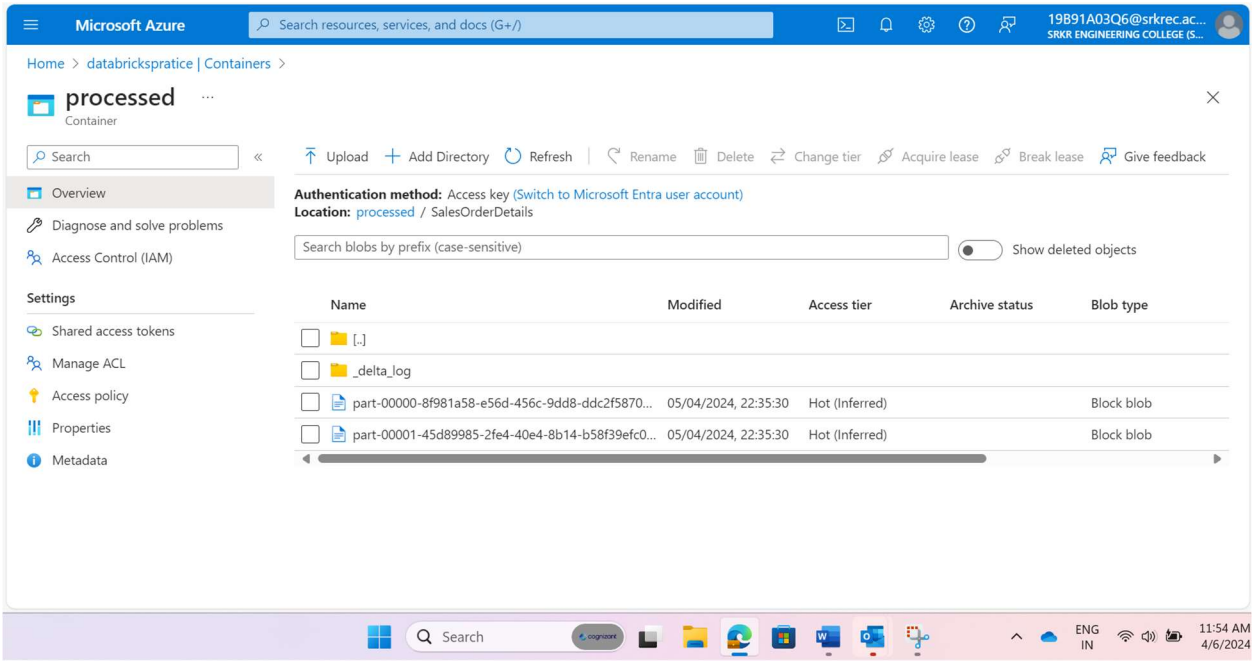

Microsoft Azure
Search resources, services, and docs (G+/)
19B91A03Q6@srkrec.ac...
SRKR ENGINEERING COLLEGE (S...
Home > databrickspratice | Containers >
processed
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: processed / FinalProductSalesOrderDetails
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
[..]
_delta_log
part-00000-9f57290a-c604-49ca-8fe7-0ed56d02eb...
05/04/2024, 22:35:24
Hot (Inferred)
Block blob
Search
ENG
IN
11:52 AM
4/6/2024


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File Edit View Run Help
Last edit was 3 minutes ago
New cell UI: ON
Run all
Terminated
Schedule
Share
Creating external table "SalesOrderDetails" load data into datalake into processed container with file name "SalesOrderDetails"
08:46 PM (2s)
41
spark.sql('''CREATE TABLE if NOT EXISTS SalesOrderDetails (SalesOrderID STRING, RevisionNumber STRING, OrderDate STRING, DueDate STRING, ShipDate STRING, Status STRING, OnlineOrderFlag STRING, SalesOrderNumber STRING, PurchaseOrderNumber STRING, AccountNumber STRING, CustomerID STRING, ShipToAddressID STRING, OnlineOrdercopen STRING, PurchaseOrderNumberid STRING, SalesOrderNumberID STRING, AccountNumberID STRING, BillToAddressID STRING, ShipMethod STRING,CreditCardApprovalCode STRING,SubTotal STRING,TaxAmt STRING,Freight STRING,TotalDue STRING,Comment STRING,rowguid STRING,ModifiedDate STRING) USING delta LOCATION 'abfss://processed@databrickspratice.dfs.core.windows.net/SalesOrderDetails' ''')
(4) Spark Jobs
Out[87]: DataFrame[]
Markdown
Search
ENG
IN
11:39 PM
4/5/2024

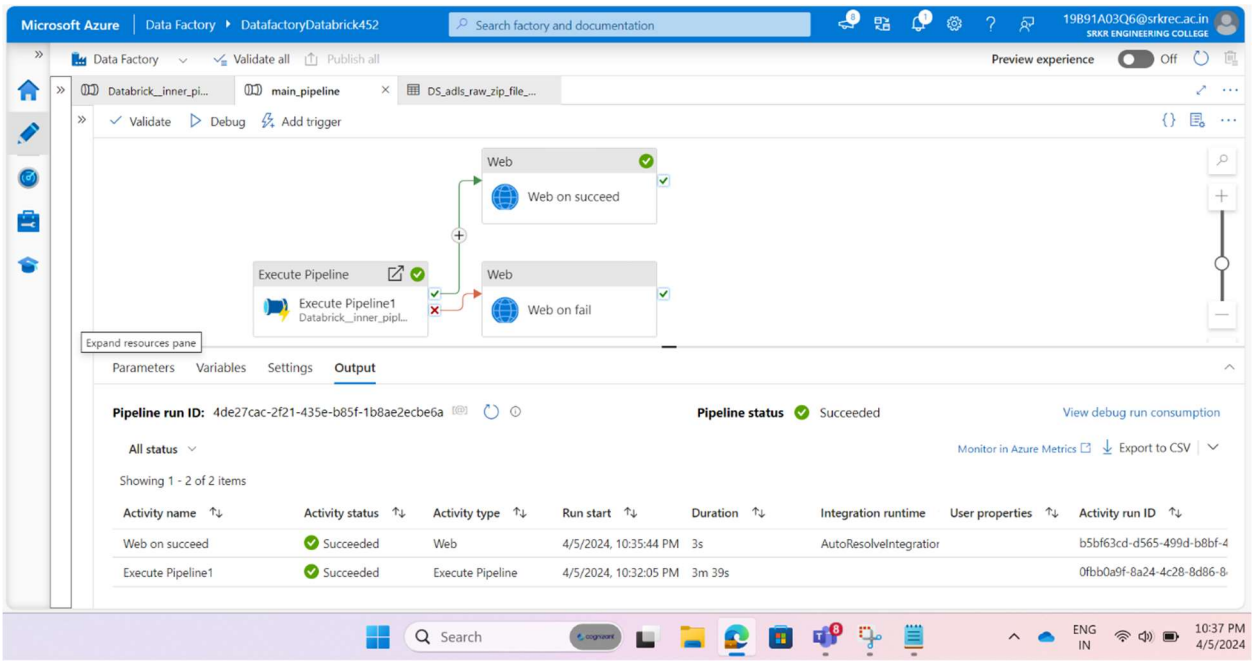

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File Edit View Run Help
Last edit was 4 minutes ago
New cell UI: ON
Run all
Terminated
Schedule
Share
Inserting records into external table " SalesOrderDetails" from "df_salesOrder"
08:47 PM (4s)
43
%sql
insert into SalesOrderDetails
select * from df_salesOrder;
/*
%python
df_salesOrder.write.option("path","abfss://processed@databrickspratice.dfs.core.windows.net/SalesOrderDetails").saveAsTable("SalesOrderDetails")
*/
(5) Spark Jobs
_sqldf: pyspark.sql.dataframe.DataFrame = [num_affected_rows: long, num_inserted_rows: long]
Table
New result table: OFF
num_affected_rows
num_inserted_rows
1
31465
31465
Search
ENG
IN
11:40 PM
4/5/2024

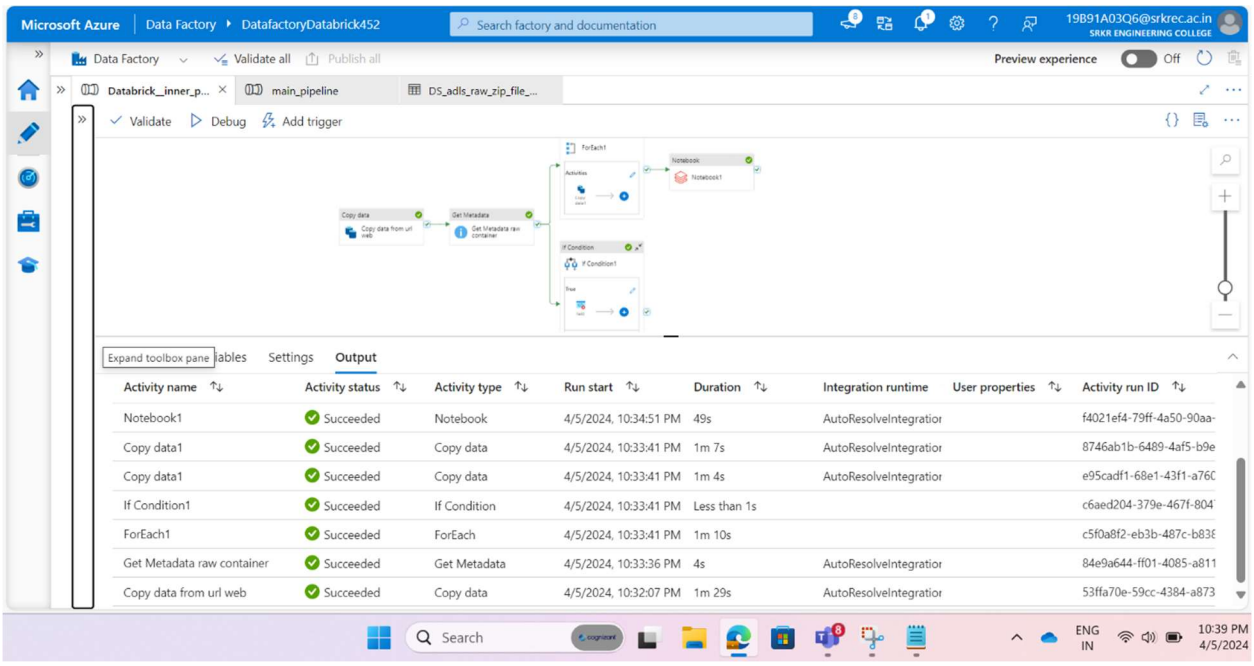

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl...
14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.salesorderdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Filter columns...
Column
Type
Comment
SalesOrderID
string
RevisionNumber
string
OrderDate
string
DueDate
string
ShipDate
string
Status
string
OnlineOrderFlag
string
About this table
Owner: Not set
Data source format: Delta
Last Updated: 13 hours ago
Size: 2.9MiB, 2 files
Comment: Add comment
Search
ENG
IN
11:53 AM
4/6/2024

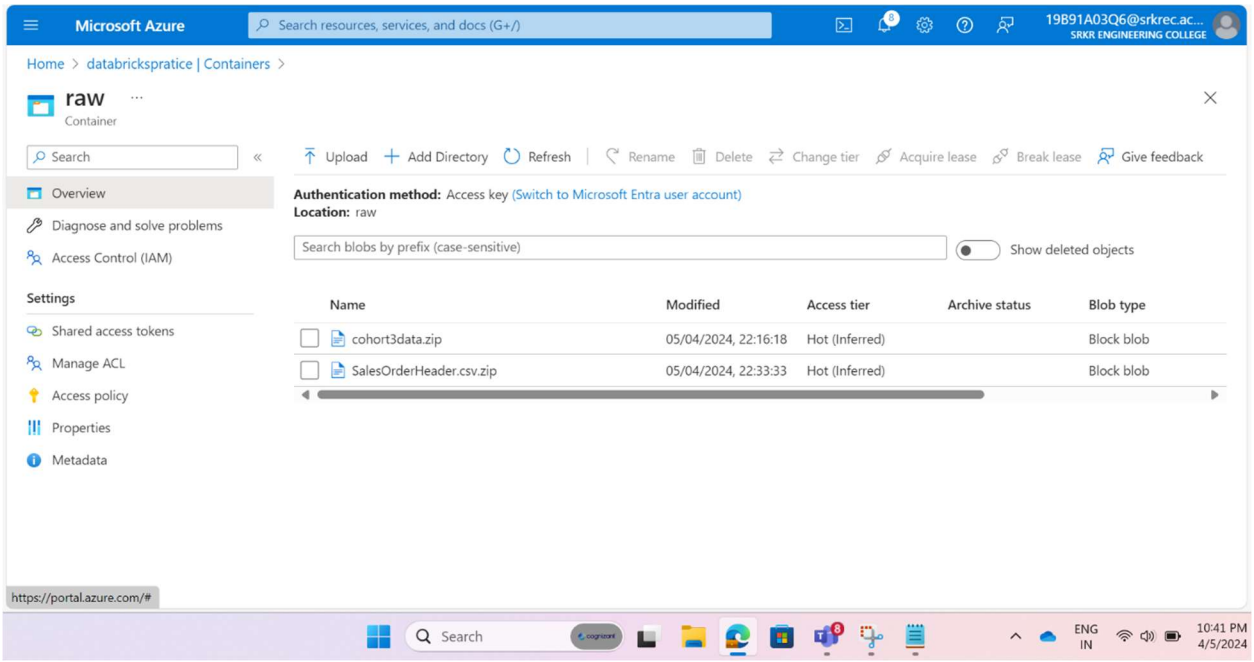

Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Catalog Explorer
Send feedback
Add
Browse DBFS
VARADA VINAY's Cl...
14 GB, 4 Cores
Type to filter
hive_metastore
default
finalproductsalesorderdetails
manage_productdetails
manage_productsalesorderdetails
manage_salesorderdetails
productdetails
salesorderdetails
samples
default >
default.salesorderdetails
Use with BI tools
Create
Overview
Sample Data
Details
Permissions
History
Created At
Fri Apr 05 17:05:28 UTC 2024
Last Access
UNKNOWN
Created By
Spark 3.3.0
Type
EXTERNAL
Location
abfss://processed@databrickspratice.dfs.core.windows.net/SalesOrderDetails
Properties
delta.minReaderVersion=1
delta.minWriterVersion=2
Search
ENG
IN
12:43 AM
4/6/2024

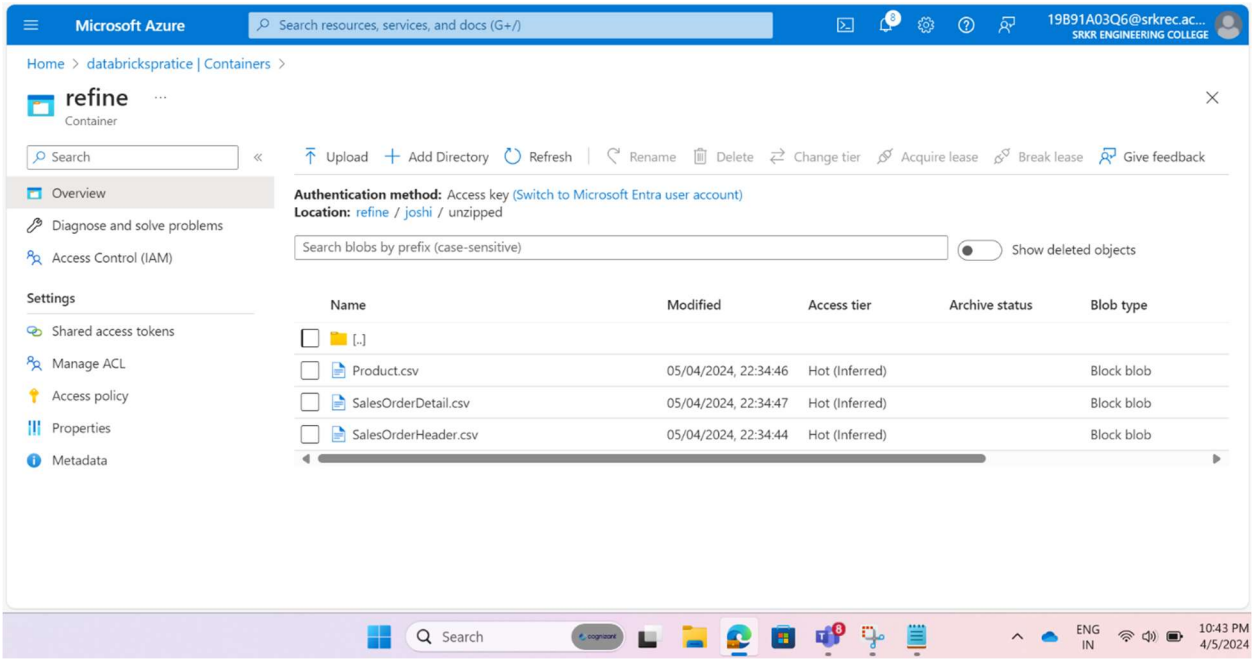


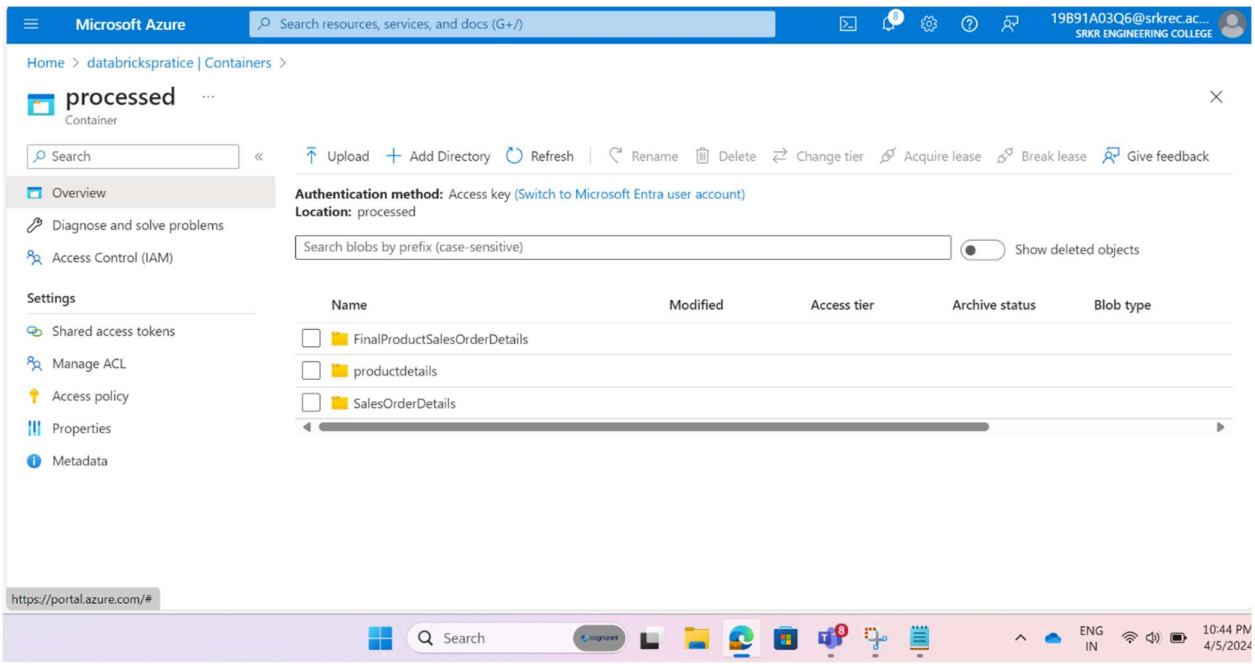

# Pipeline Result on success:

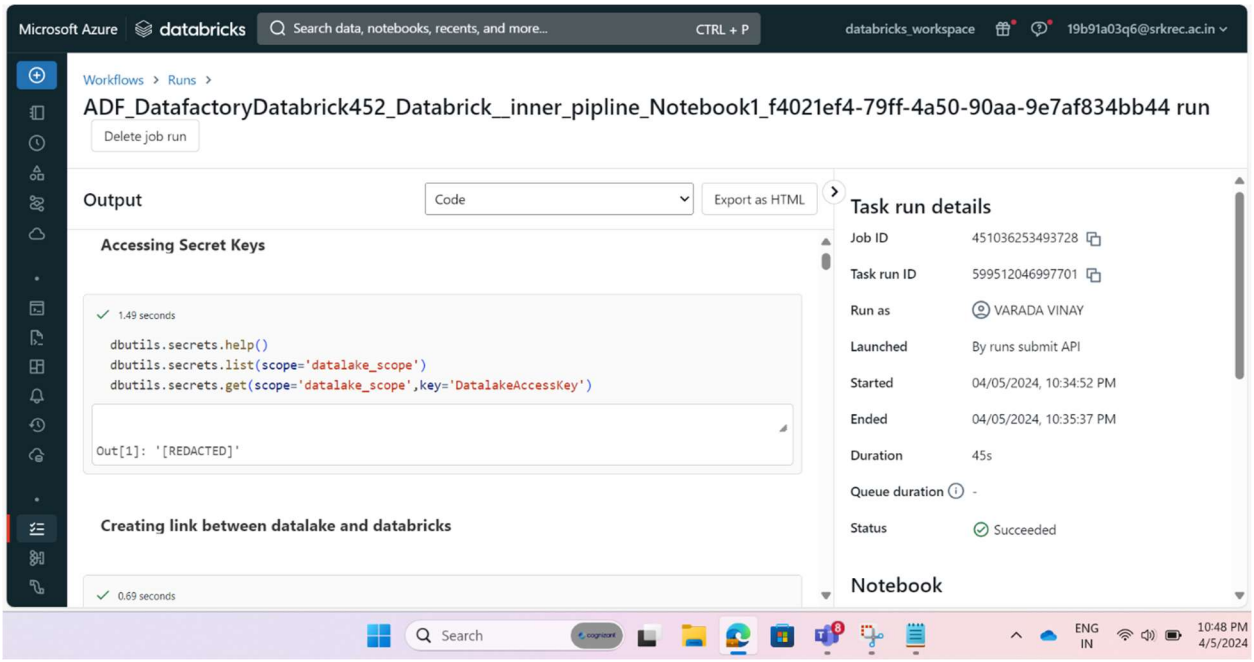


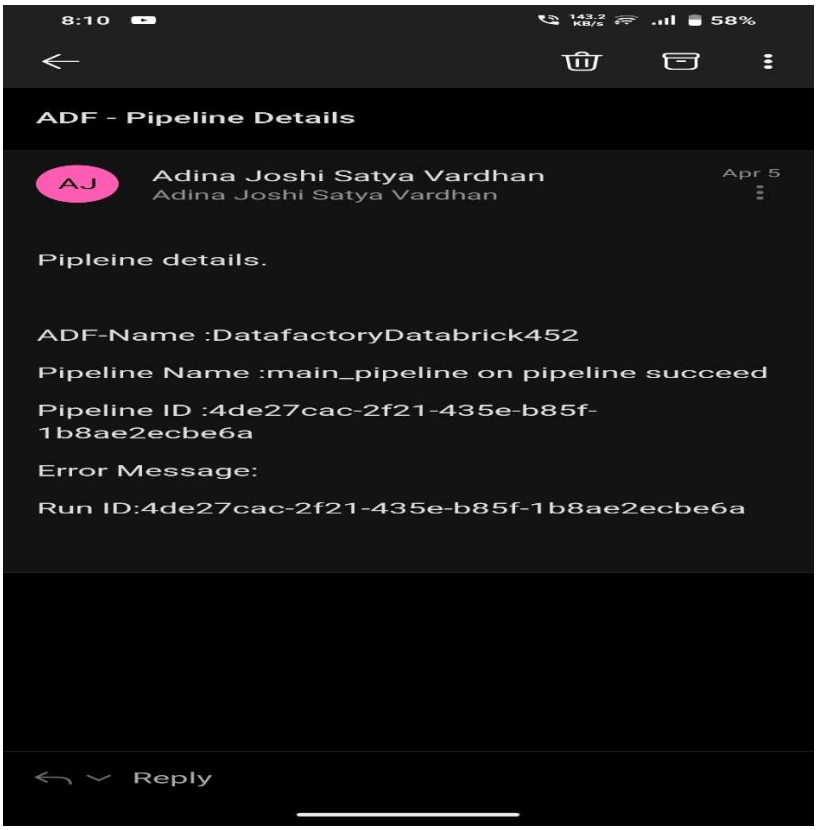

Microsoft Azure
Data Factory
DatafactoryDatabrick452
Search factory and documentation
19B91A03Q6@srkrec.ac.in
SRKR ENGINEERING COLLEGE
Data Factory
Validate all
Publish all
Preview experience
Off
Databrick_inner_p...
main_pipeline
DS_adls_raw_zip_file_...
Validate
Debug
Add trigger
Expand toolbox pane
Settings
Output
Activity name
Activity status
Activity type
Run start
Duration
Integration runtime
User properties
Activity run ID
Notebook1 Succeeded Notebook 4/5/2024, 10:34:51 PM 49s AutoResolveIntegratior f4021ef4-79ff-4a50-90aa-
Copy data1 Succeeded Copy data 4/5/2024, 10:33:41 PM 1m 7s AutoResolveIntegratior 8746ab1b-6489-4af5-b9e
Copy data1 Succeeded Copy data 4/5/2024, 10:33:41 PM 1m 4s AutoResolveIntegratior e95cadf1-68e1-43f1-a76(
If Condition1 Succeeded If Condition 4/5/2024, 10:33:41 PM Less than 1s c6aed204-379e-467f-804
ForEach1 Succeeded ForEach 4/5/2024, 10:33:41 PM 1m 10s c5f0a8f2-eb3b-487c-b83
Get Metadata raw container Succeeded Get Metadata 4/5/2024, 10:33:36 PM 4s AutoResolveIntegratior 84e9a644-ff01-4085-a811
Copy data from url web Succeeded Copy data 4/5/2024, 10:32:07 PM 1m 29s AutoResolveIntegratior 53ffa70e-59cc-4384-a873
Search
ENG
IN
10:39 PM
4/5/2024


Microsoft Azure
Search resources, services, and docs (G+/)
19B91A03Q6@srkrec.ac...
SRKR ENGINEERING COLLEGE
Home > databrickspratice | Containers >
raw
Container
Search
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: raw
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
cohort3data.zip 05/04/2024, 22:16:18 Hot (Inferred) Block blob
SalesOrderHeader.csv.zip 05/04/2024, 22:33:33 Hot (Inferred) Block blob
https://portal.azure.com/#
Search
ENG
IN
10:41 PM
4/5/2024


Microsoft Azure
Search resources, services, and docs (G+/)
19B91A03Q6@srkrec.ac...
SRKR ENGINEERING COLLEGE
Home > databrickspratice | Containers >
refine
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: refine / joshi / unzipped
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
[..]
Product.csv
05/04/2024, 22:34:46
Hot (Inferred)
Block blob
SalesOrderDetail.csv
05/04/2024, 22:34:47
Hot (Inferred)
Block blob
SalesOrderHeader.csv
05/04/2024, 22:34:44
Hot (Inferred)
Block blob
Search
ENG
IN
10:43 PM
4/5/2024

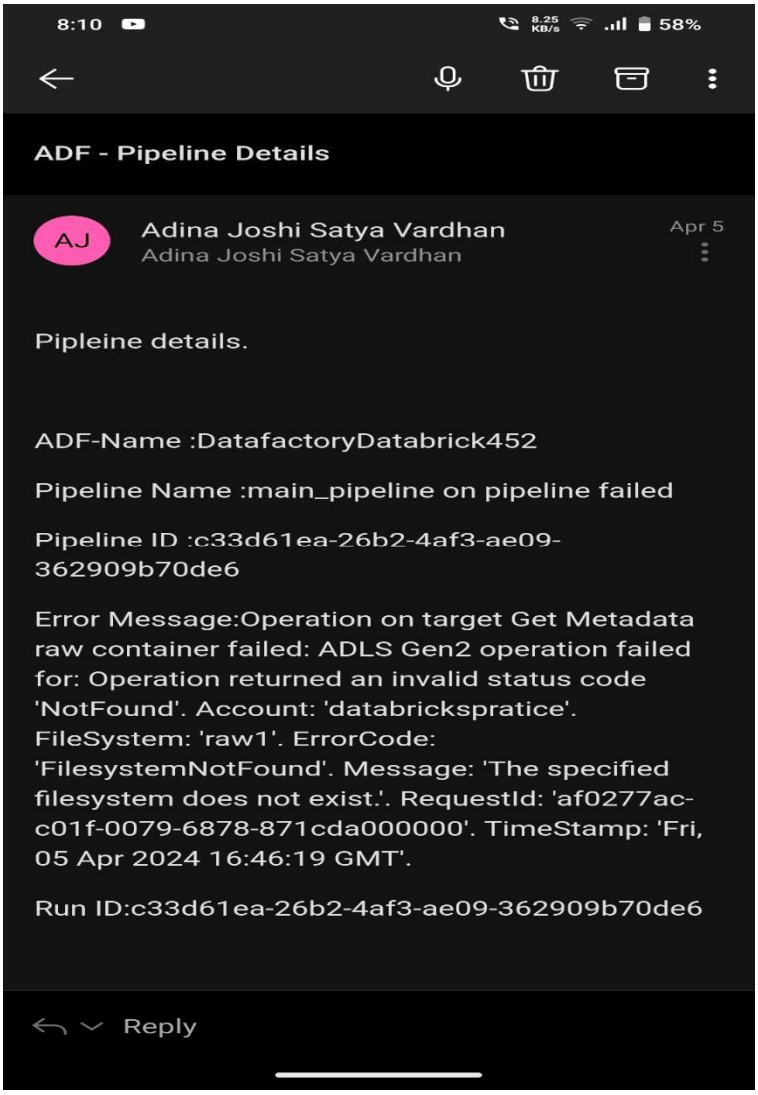

Microsoft Azure
Search resources, services, and docs (G+/)
19B91A03Q6@srkrec.ac...
SRKR ENGINEERING COLLEGE
Home > databrickspratice | Containers >
processed
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Upload
Add Directory
Refresh
Rename
Delete
Change tier
Acquire lease
Break lease
Give feedback
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: processed
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
FinalProductSalesOrderDetails
productdetails
SalesOrderDetails
https://portal.azure.com/#
Search
ENG
IN
10:44 PM
4/5/2024



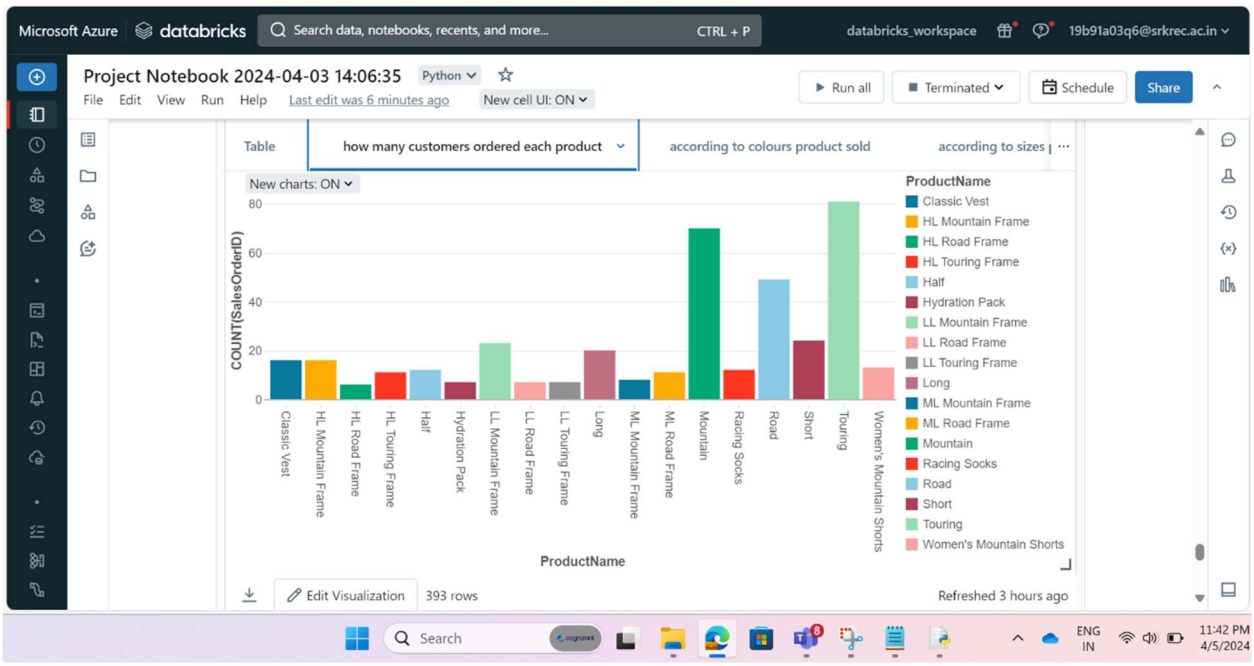



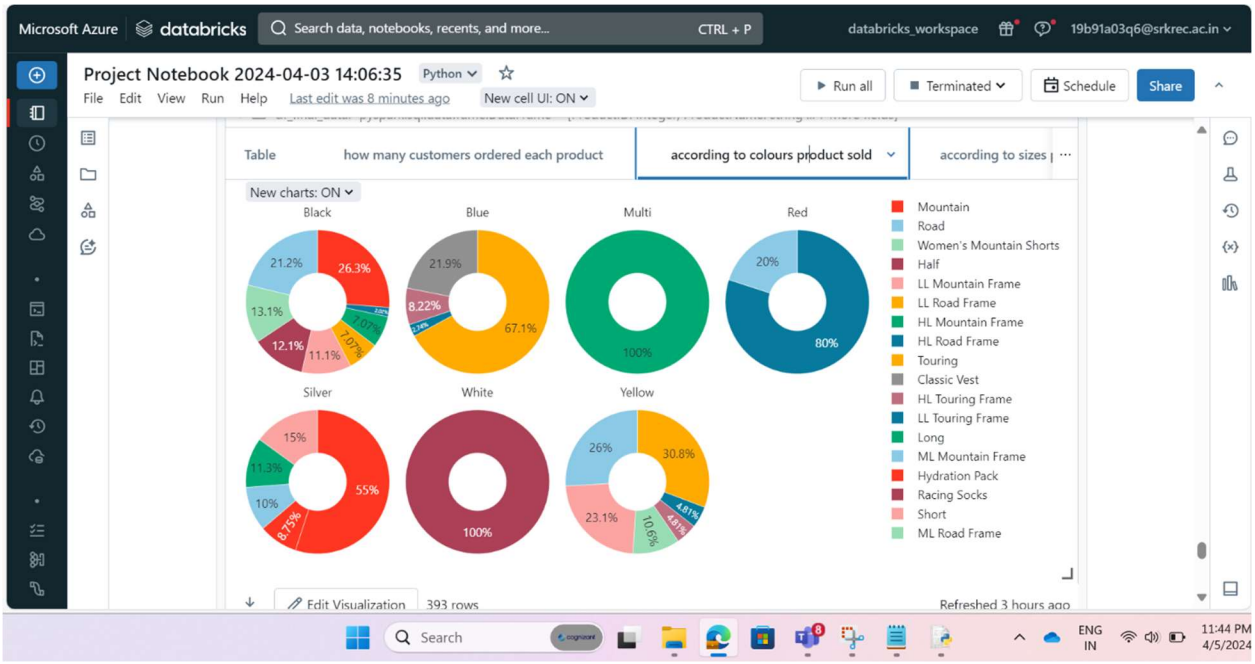

# Mail on pipeline succeed

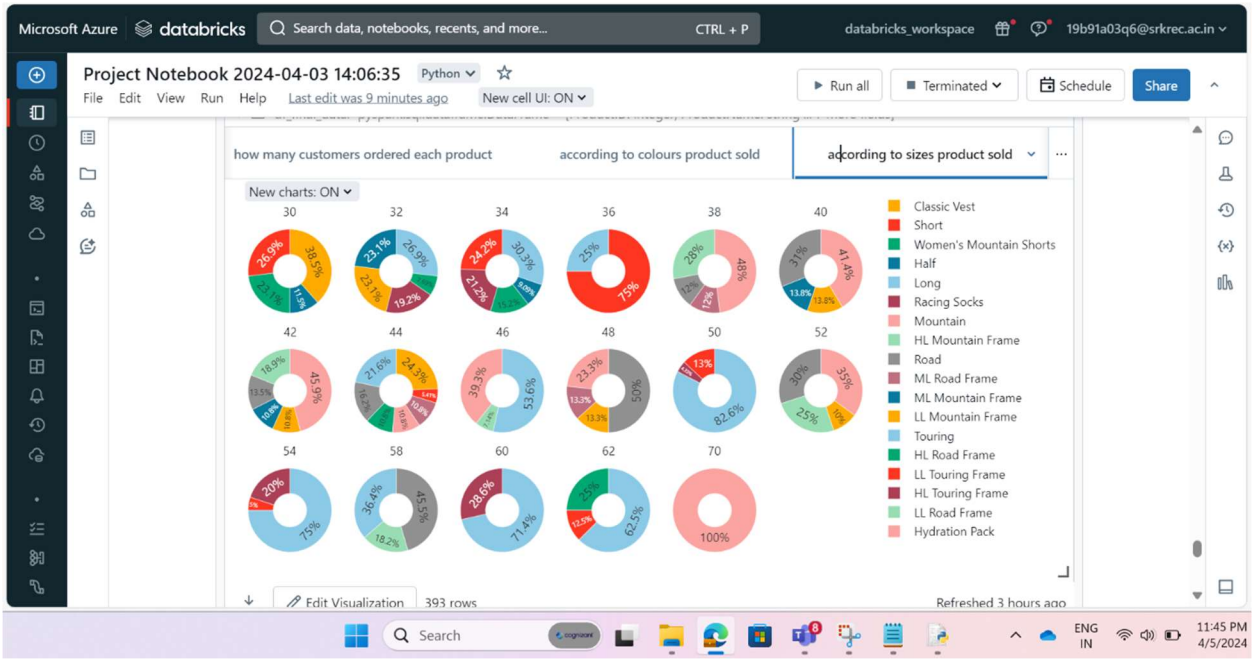
## Mail on pipeline failed:

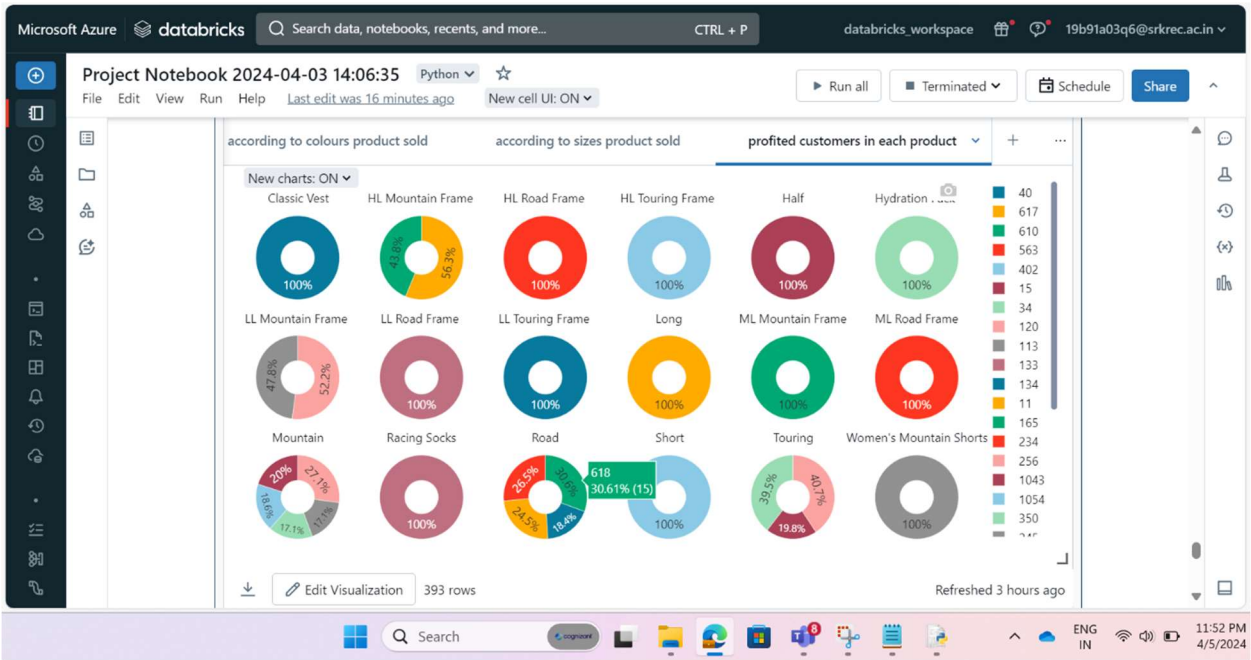


# Visualization:




Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File Edit View Run Help
Last edit was 9 minutes ago
New cell UI: ON
Run all
Terminated
Schedule
Share
how many customers ordered each product
according to colours product sold
according to sizes product sold
New charts: ON
30 32 34 36 38 40
42 44 46 48 50 52
54 58 60 62 70
Classic Vest
Short
Women's Mountain Shorts
Half
Long
Racing Socks
Mountain
HL Mountain Frame
Road
ML Road Frame
ML Mountain Frame
LL Mountain Frame
Touring
HL Road Frame
LL Touring Frame
HL Touring Frame
LL Road Frame
Hydration Pack
Edit Visualization
393 rows
Refreshed 3 hours ago
Search
ENG
IN
11:45 PM
4/5/2024


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
databricks_workspace
19b91a03q6@srkrec.ac.in
Project Notebook 2024-04-03 14:06:35
Python
File Edit View Run Help
Last edit was 16 minutes ago
New cell UI: ON
Run all
Terminated
Schedule
Share
according to colours product sold
according to sizes product sold
profited customers in each product
New charts: ON
Classic Vest
HL Mountain Frame
HL Road Frame
HL Touring Frame
Half
Hydration
LL Mountain Frame
LL Road Frame
LL Touring Frame
Long
ML Mountain Frame
ML Road Frame
Mountain
Racing Socks
Road
Short
Touring
Women's Mountain Shorts
618
30.61% (15)
40
617
610
563
402
15
34
120
113
133
134
11
165
234
256
1043
1054
350
Edit Visualization
393 rows
Refreshed 3 hours ago
Search
ENG
IN
11:52 PM
4/5/2024