# FINAL PROJECT

## Processing live streaming data via event hub and Databricks.

**EMPLOYEE ID:** 2320213

**NAME:** MAJJI VIJAY VAMSI

CSDAIA24AZ005

# INTRODUCTION

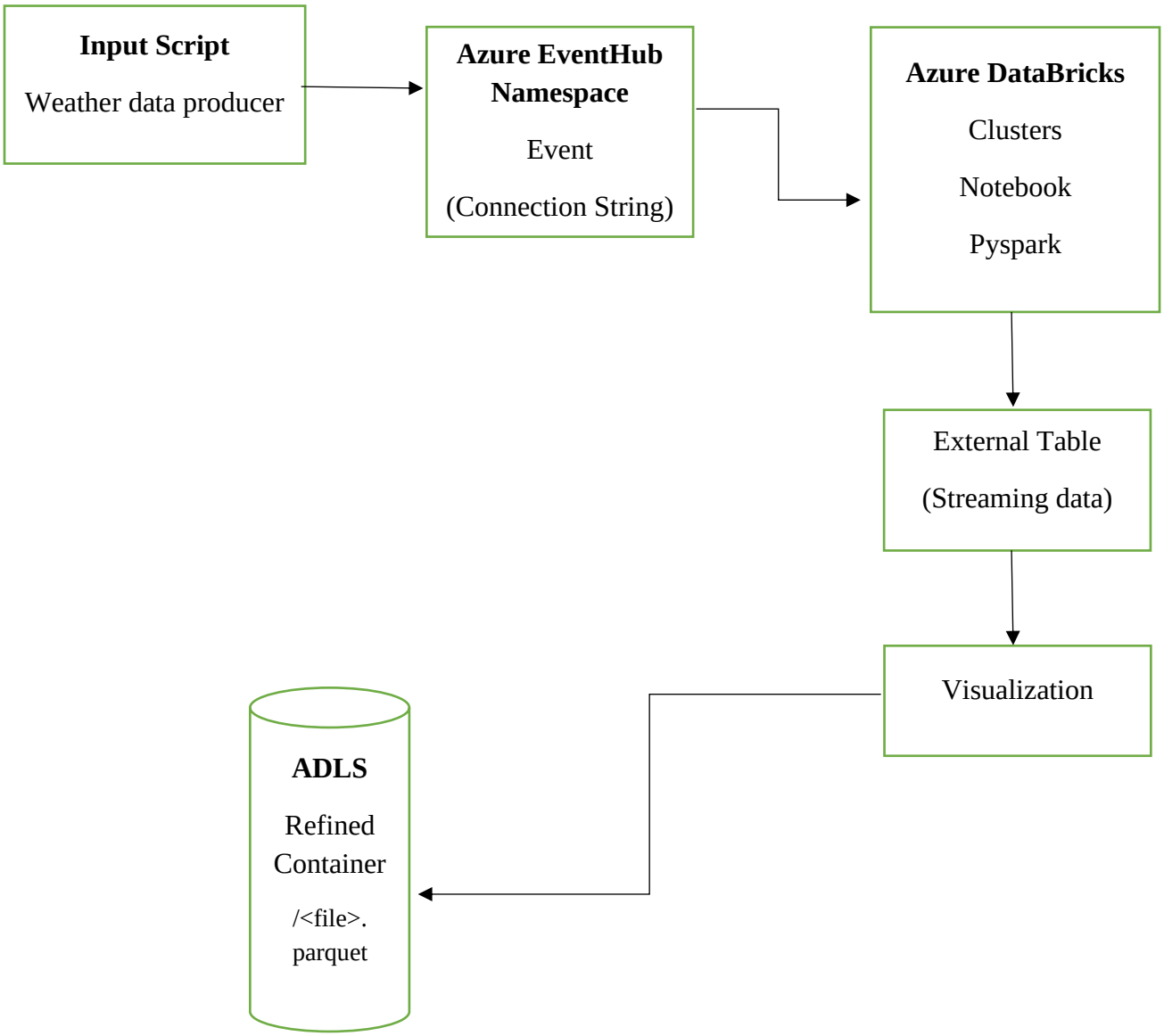
Azure EventHub and Azure DataBricks are the tools were used in the project. With this use of Azure components I learnt how to fetch the live streaming data of weather report from the Azure Event hub and convert that data into to the table in the Azure DataBricks by using the commands and the functions present in the pyspark and I loaded that streaming data in the container which present in the Azure Data Lake Storage(ADLS). By performing this operations I learnt how to get the live streaming data and how to use it based on the business requirement.

# OBJECTIVE

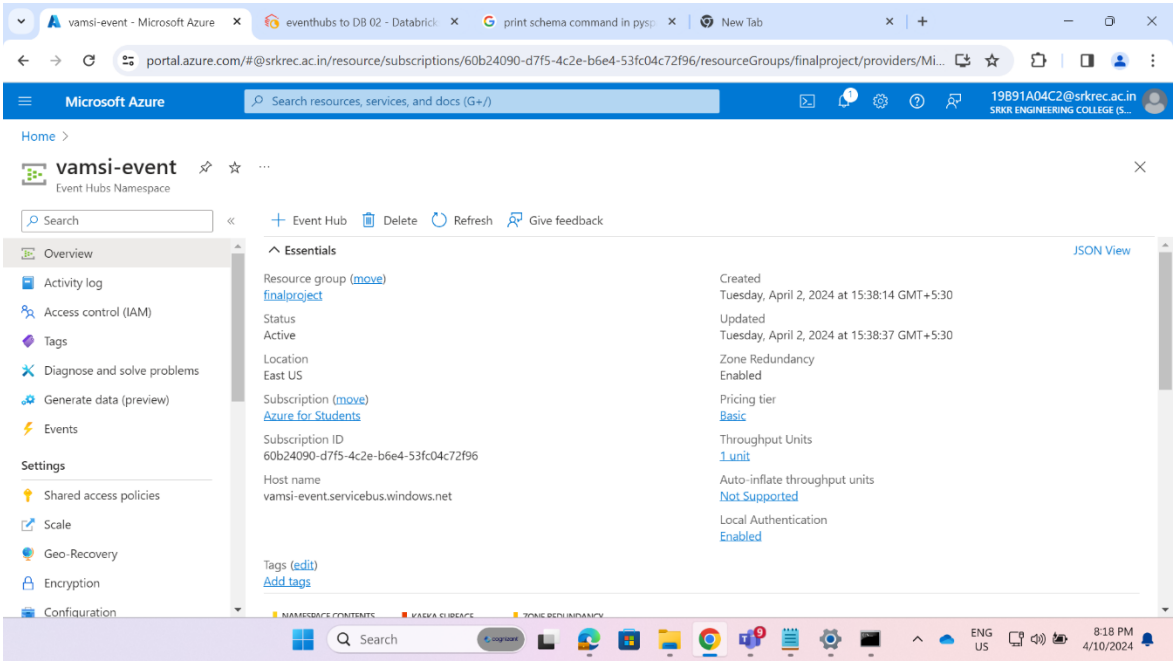
The objective of the project is to fetch the live streaming data by using the python script and loading that live streaming data into the Azure DataBricks by using the connection string in the EventHub. After fetching the data in EventHub, giving the output of the EventHub to the Azure Databricks and by making use of pyspark commands in the notebook connecting the both EventHub and ADB. Finally fetching the data from the EventHub and converting the data received into the table format and loading that data into the Adls(refined container) .

## PROJECT REQUIREMENT:

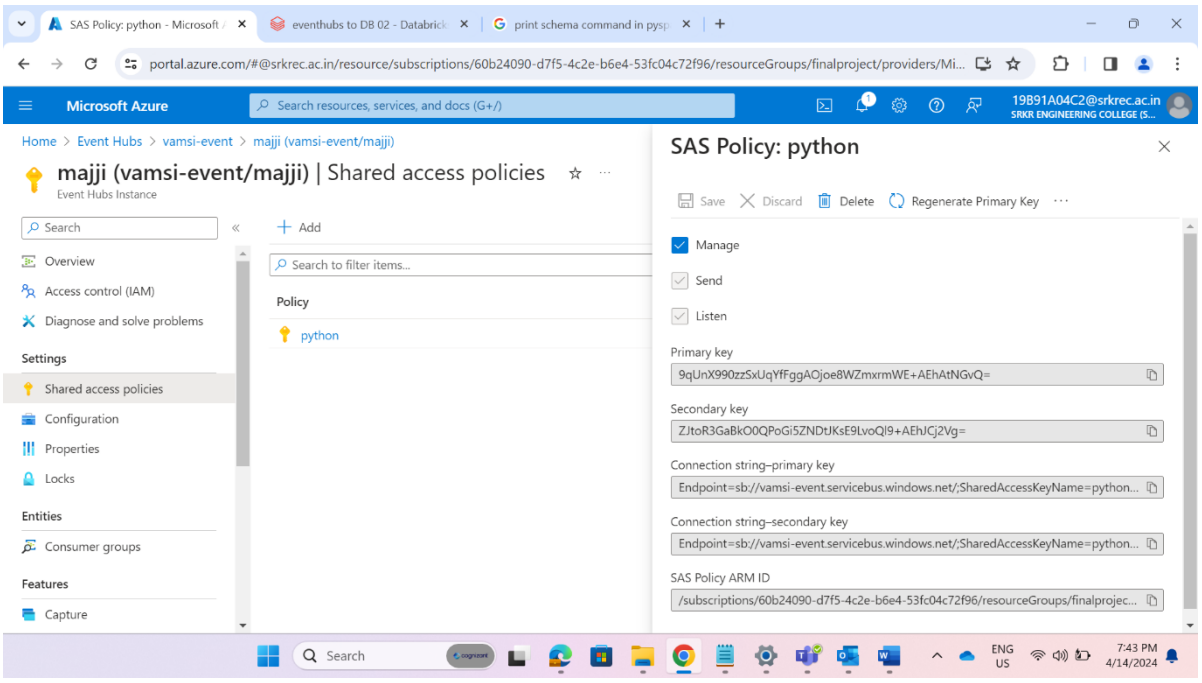
## BLOCK DIAGRAM:



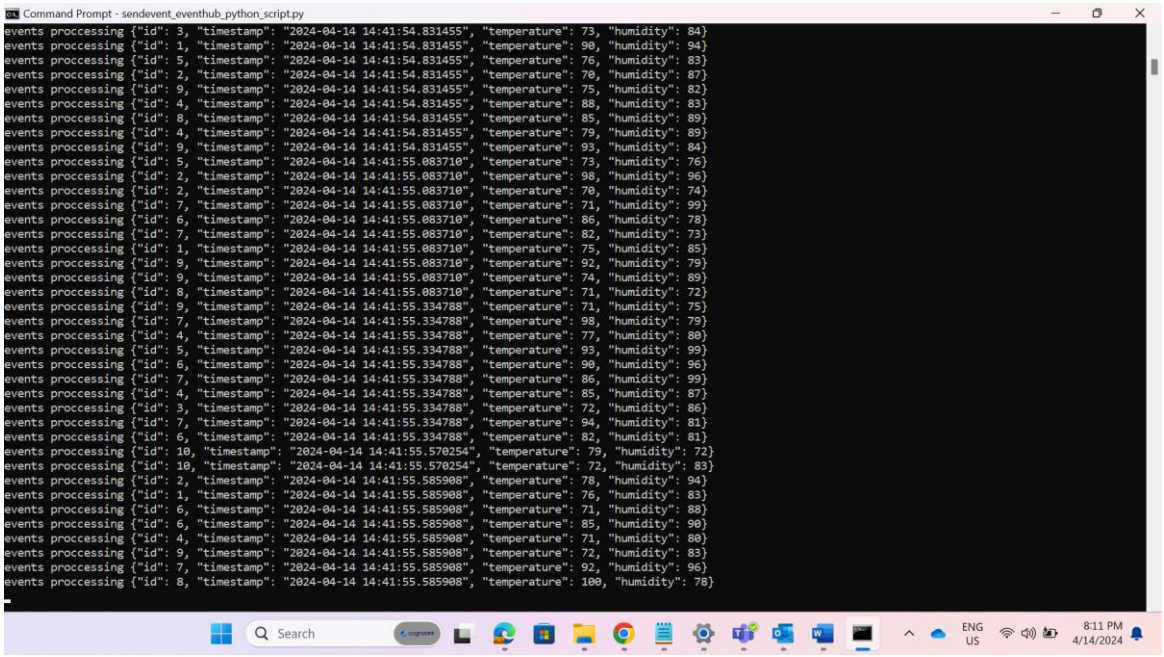
**STEP 1:** Create an event-hub namespace and event-hub (namespace-vamsi-event) (eventhub-majji).



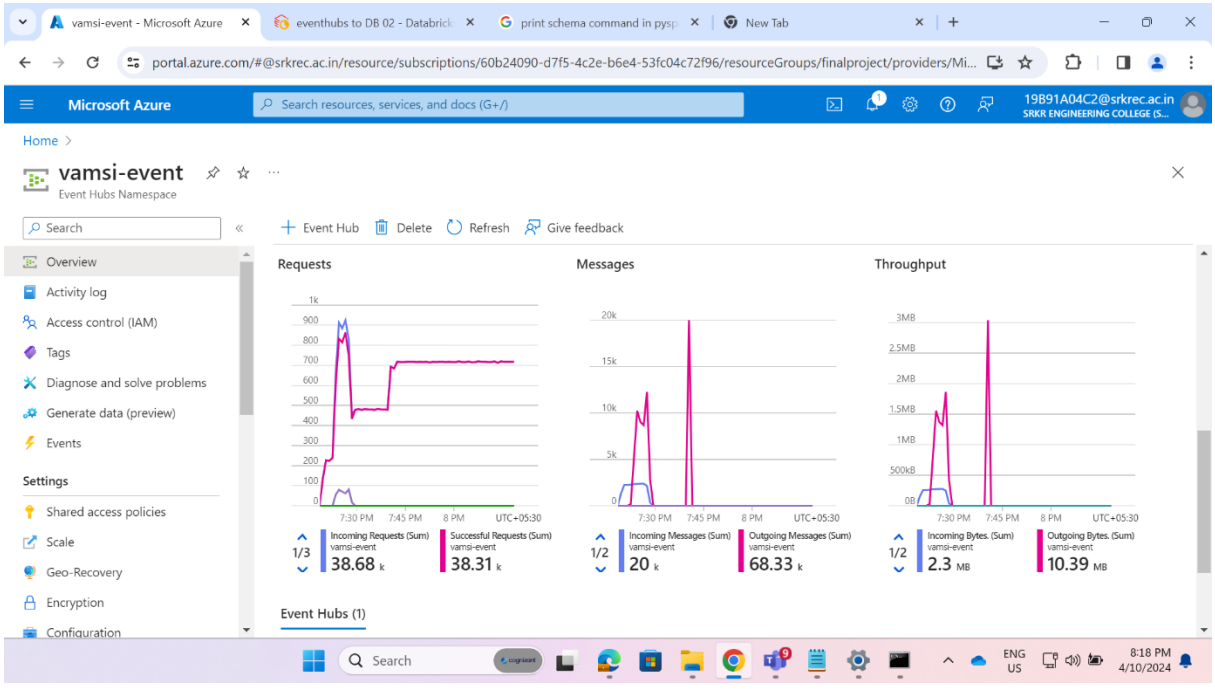

**STEP 2:** Create a SAS policy to give a connection string to script to send the events to event-hub.



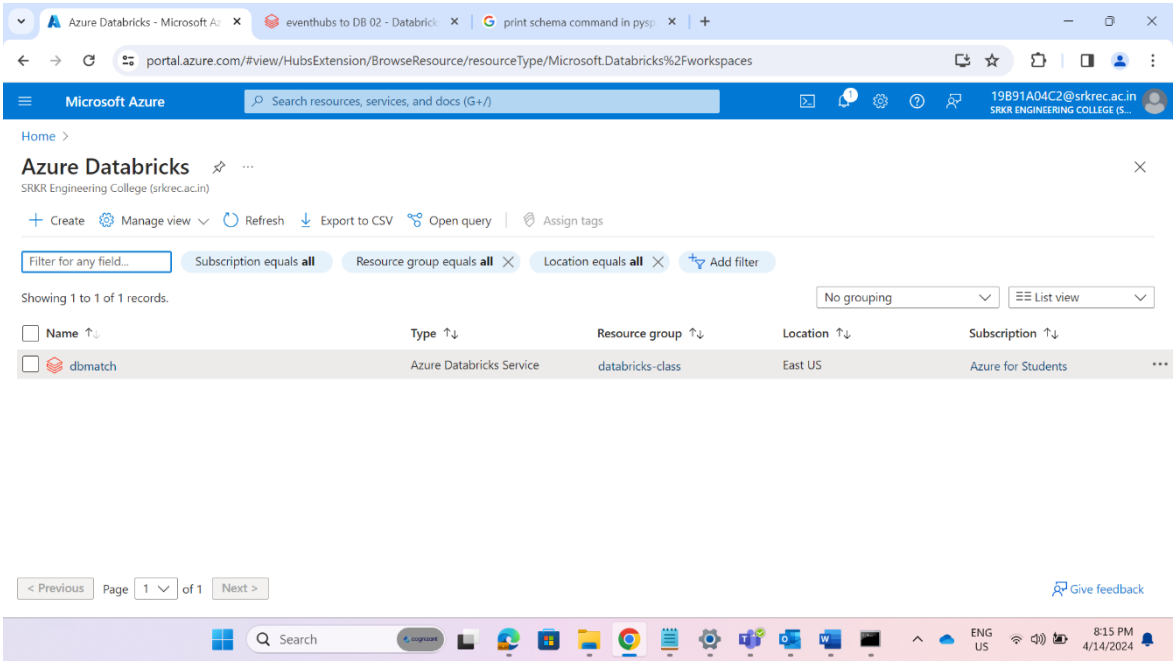
**STEP 3:** Install event hub libraries (pip install-eventhub) in cmd and run the script in command prompt (sending the events to event-hub).



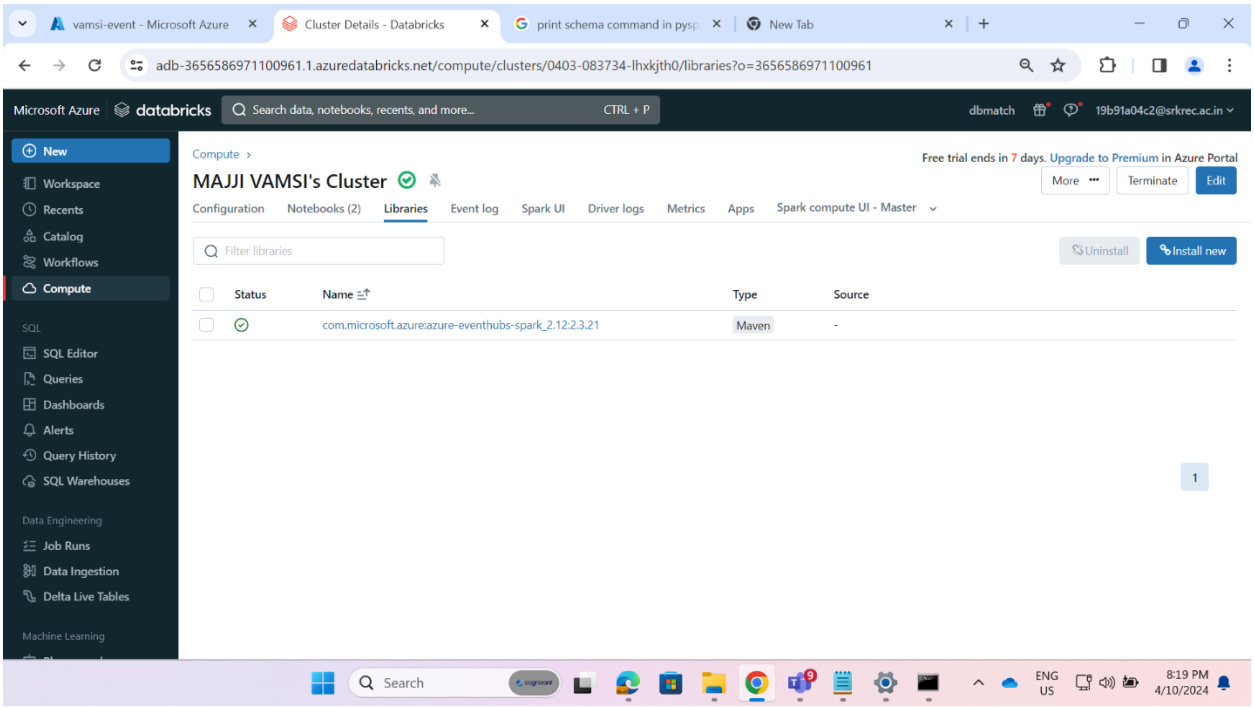

**STEP 4:** Receiving events in event-hub in graphical presentation.



**STEP 5:** Create a data Bricks resource (dbmatch) and Launch Workspace.

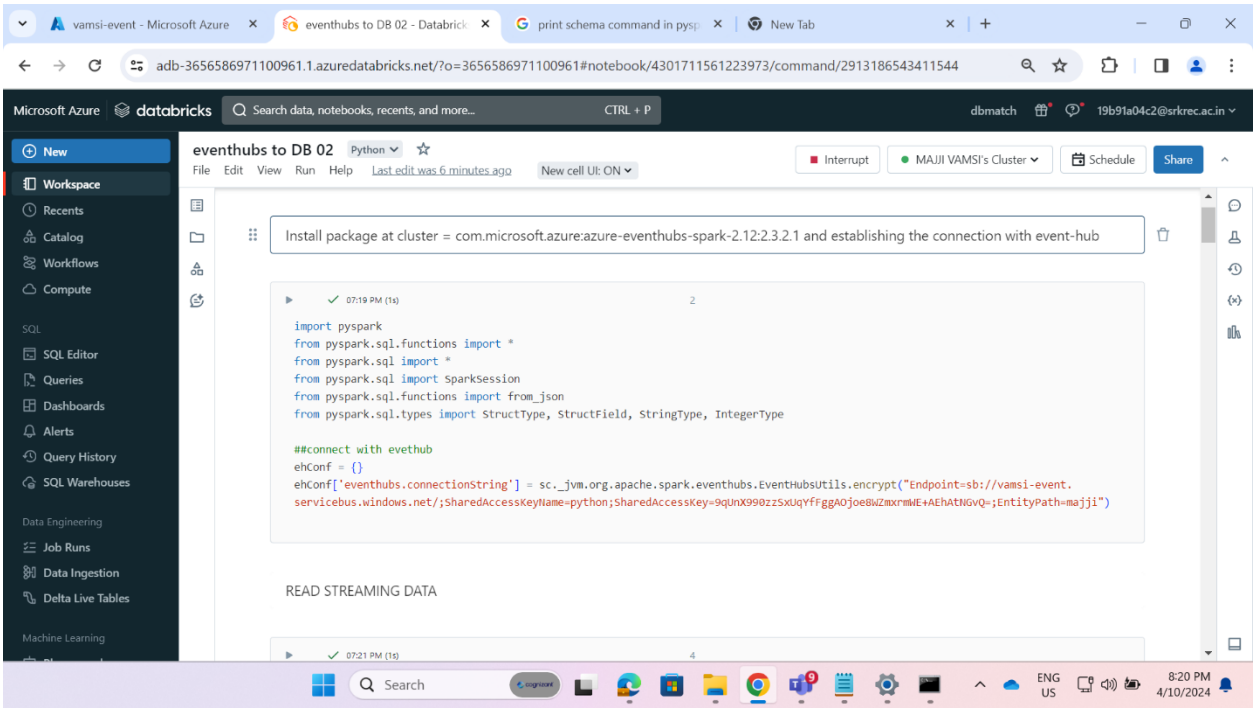



**STEP 6**: In data bricks create a cluster (MAJJI VAMSI). Install libraries to connect azure event hub to spark (com.microsoft.azure:azure-eventhubs-spark_2.12:2.3.22).

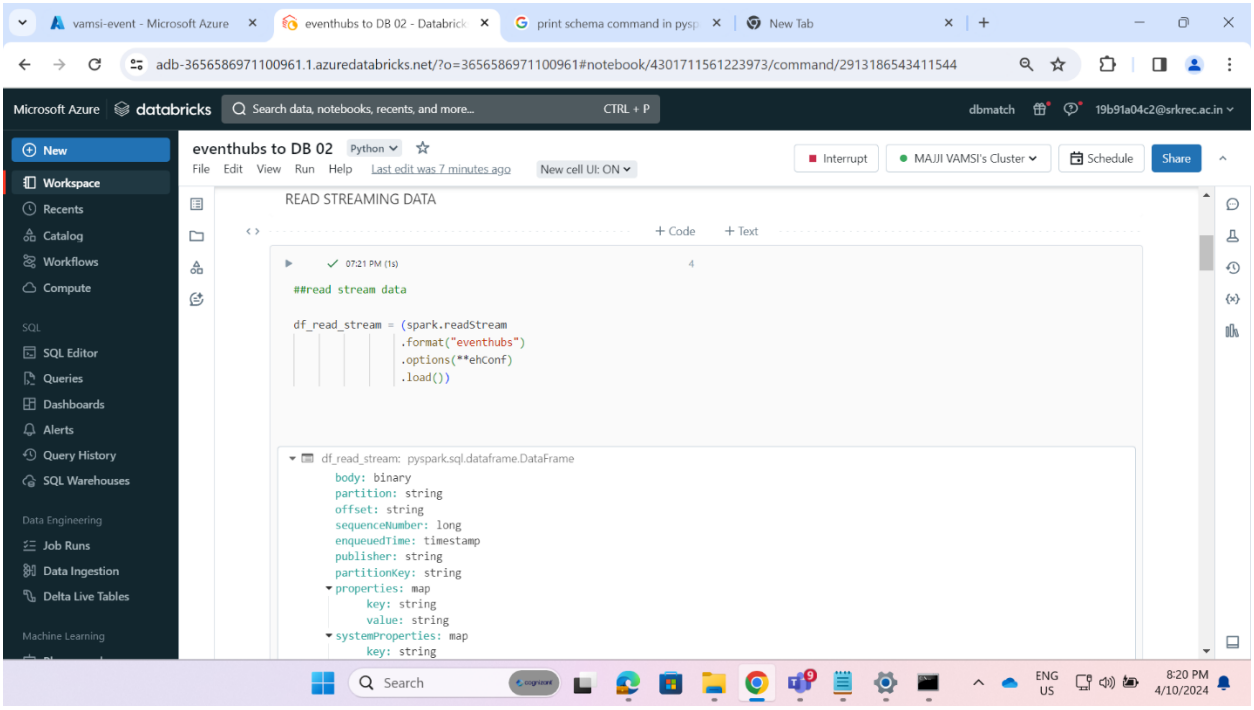


**STEP 7:** Creating a Notebook as eventhubs to DB 02.

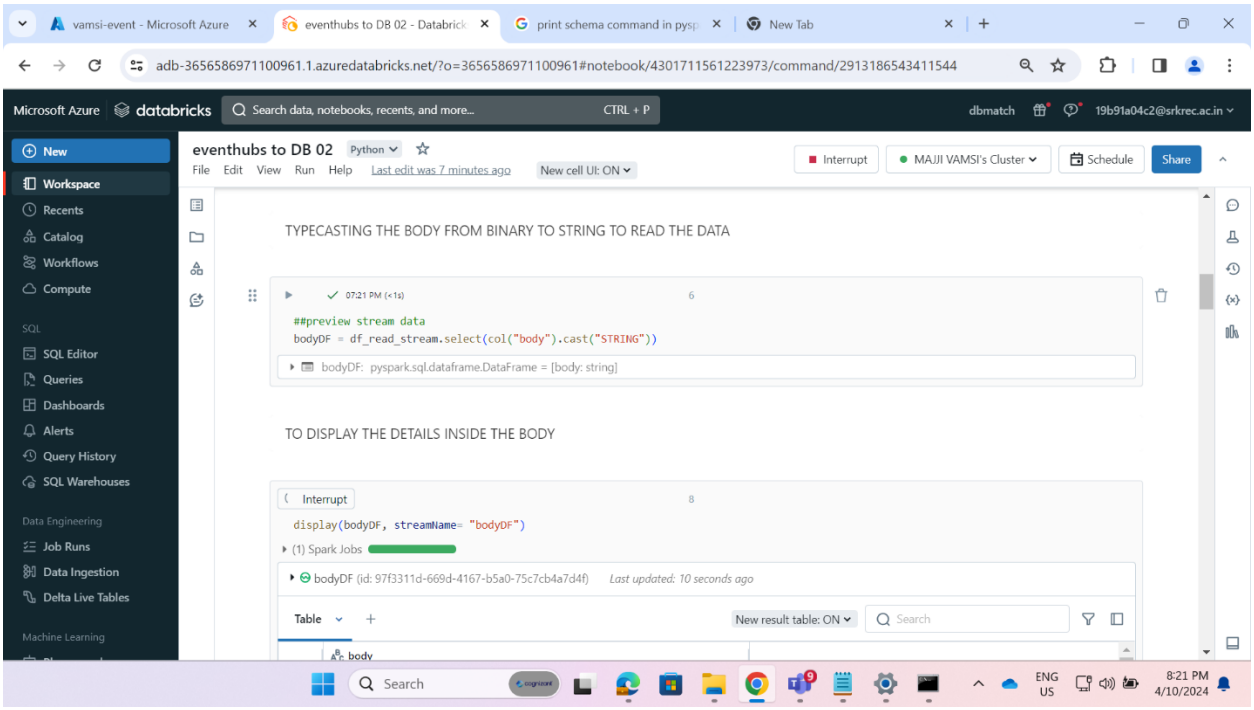
Create a cell in the notebook Import packages to work with pyspark and sql.

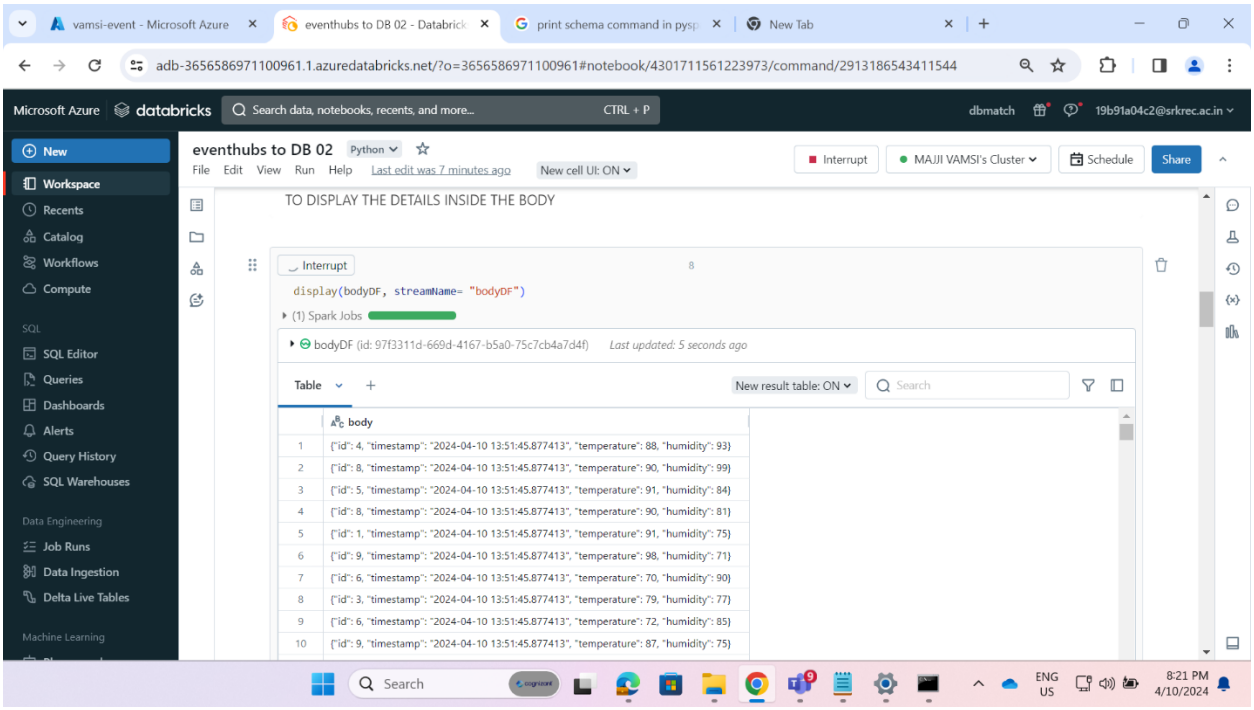



**STEP 8:** Reading the streaming data form the EventHub.

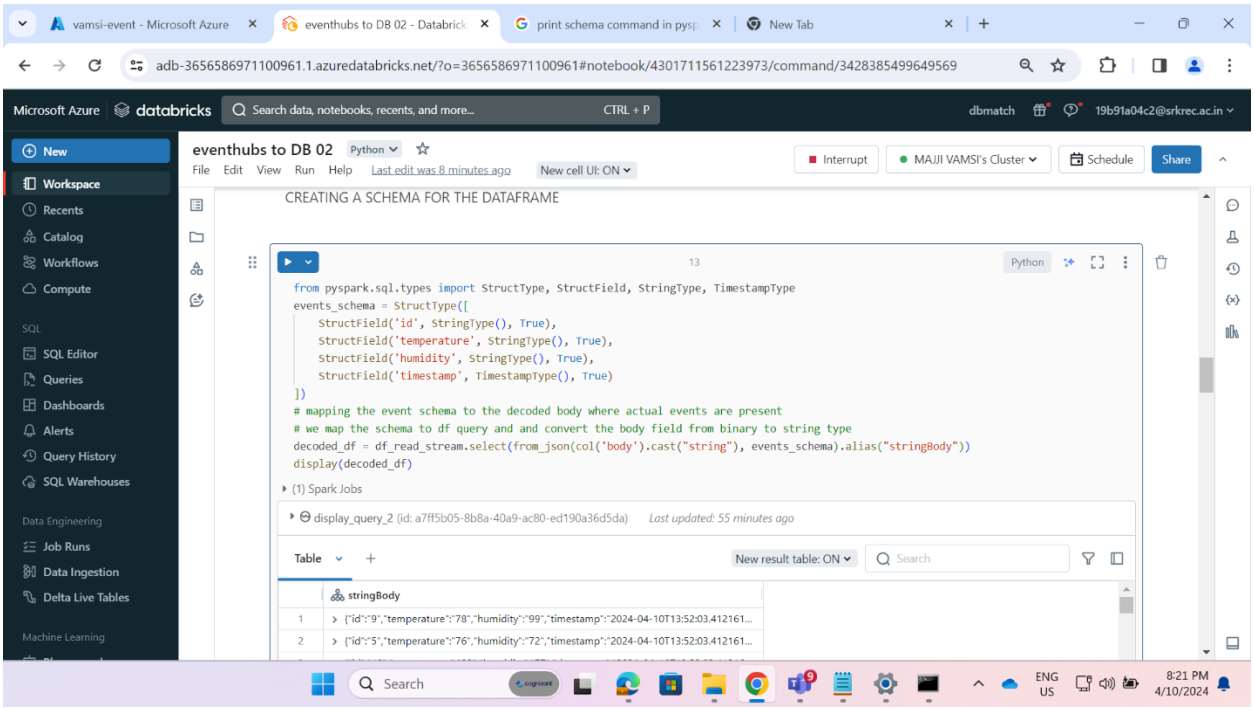


**STEP 9**: Preview the streaming data.

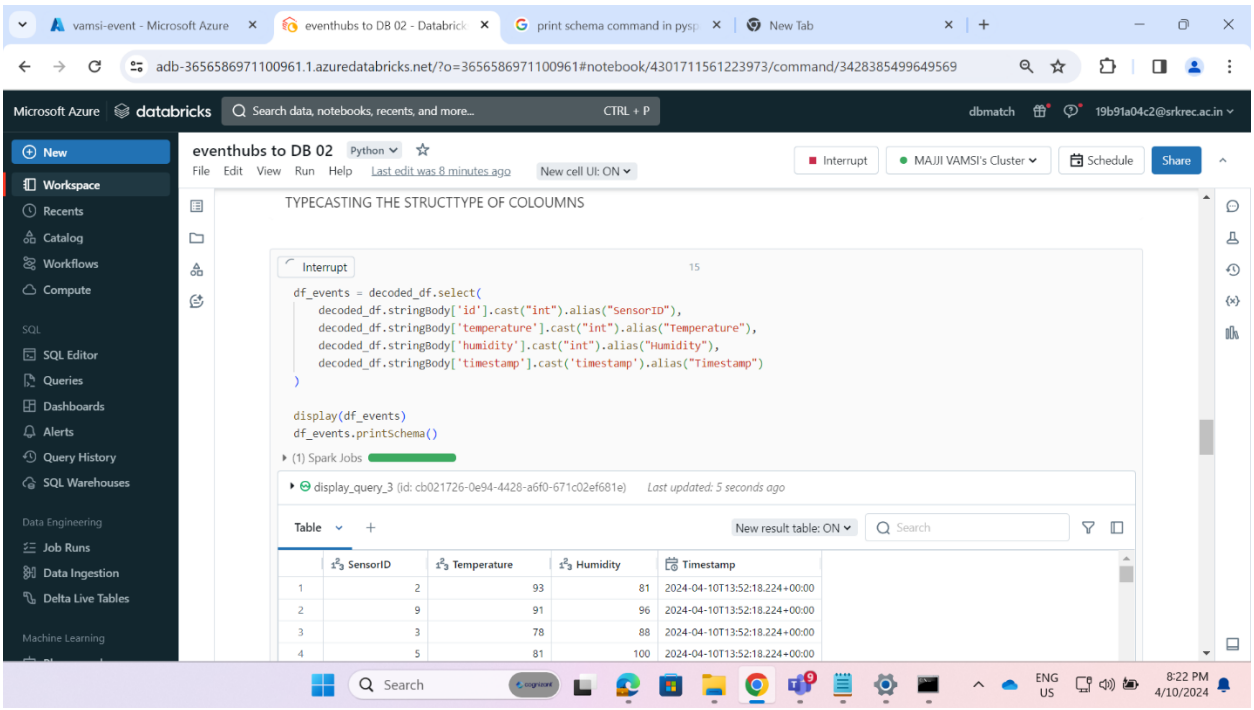


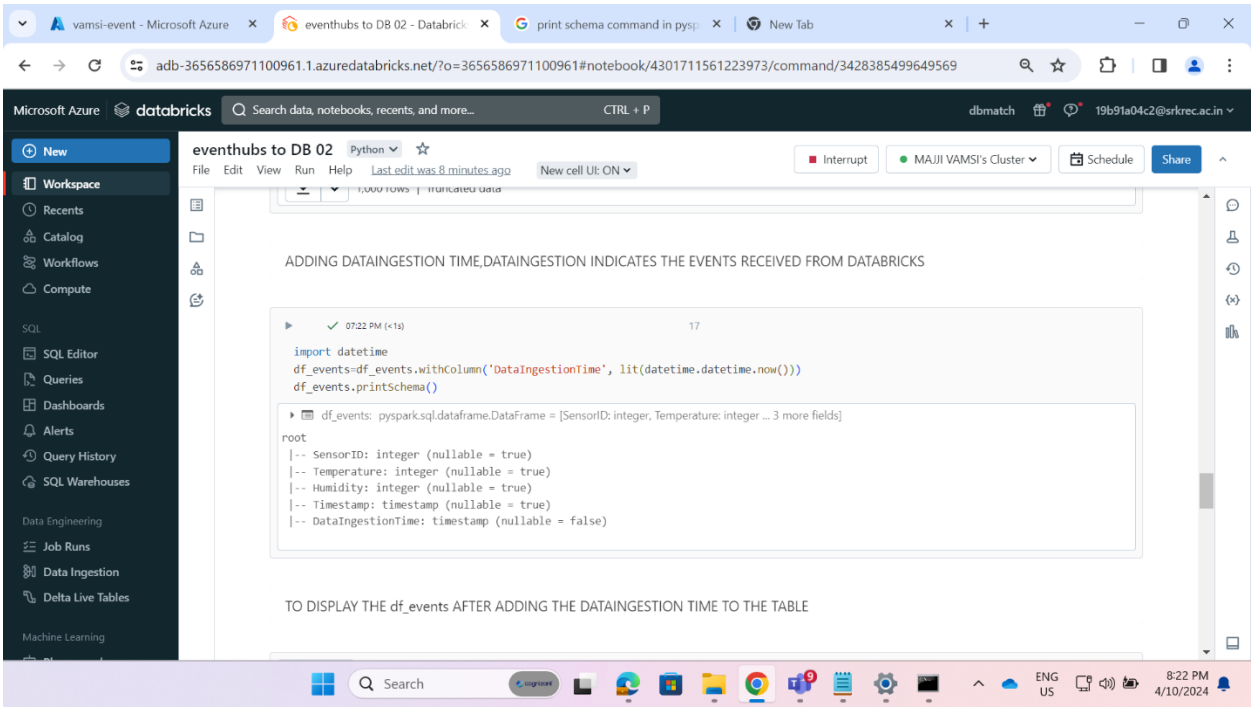
**STEP 10**: creating schema for the dataframe.



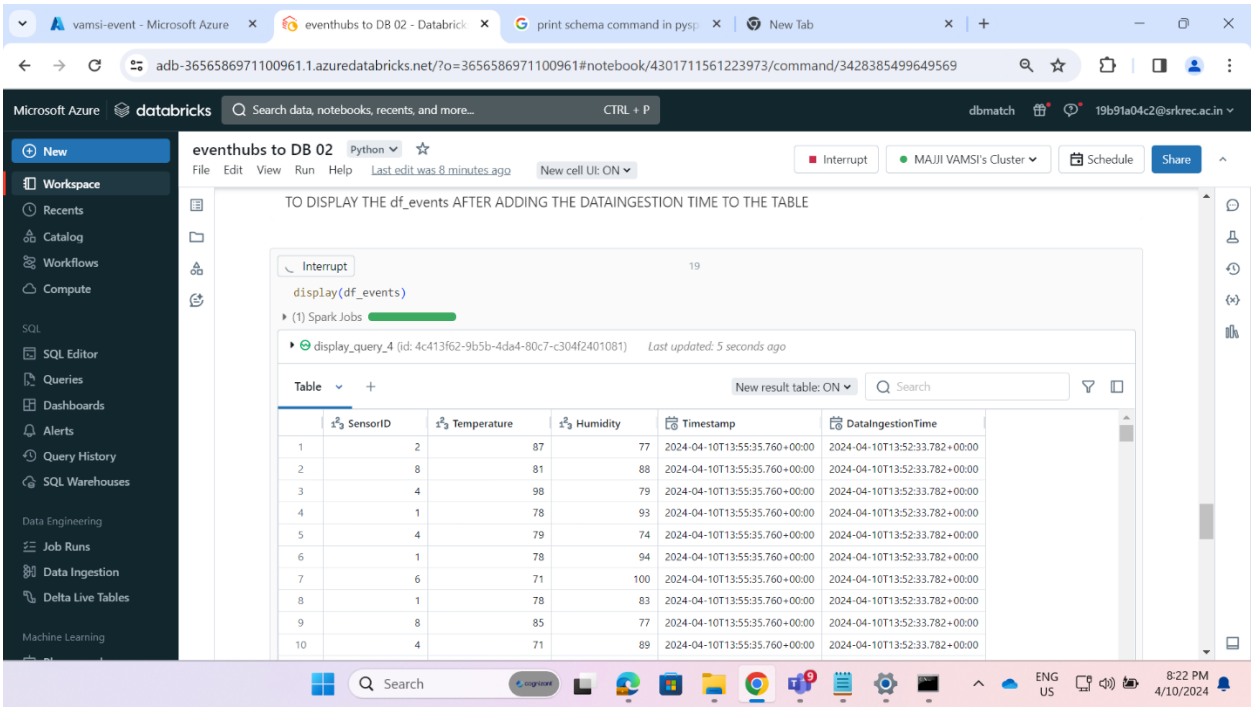

**STEP 11:** Changing the datatypes and storing the streaming data into columns.



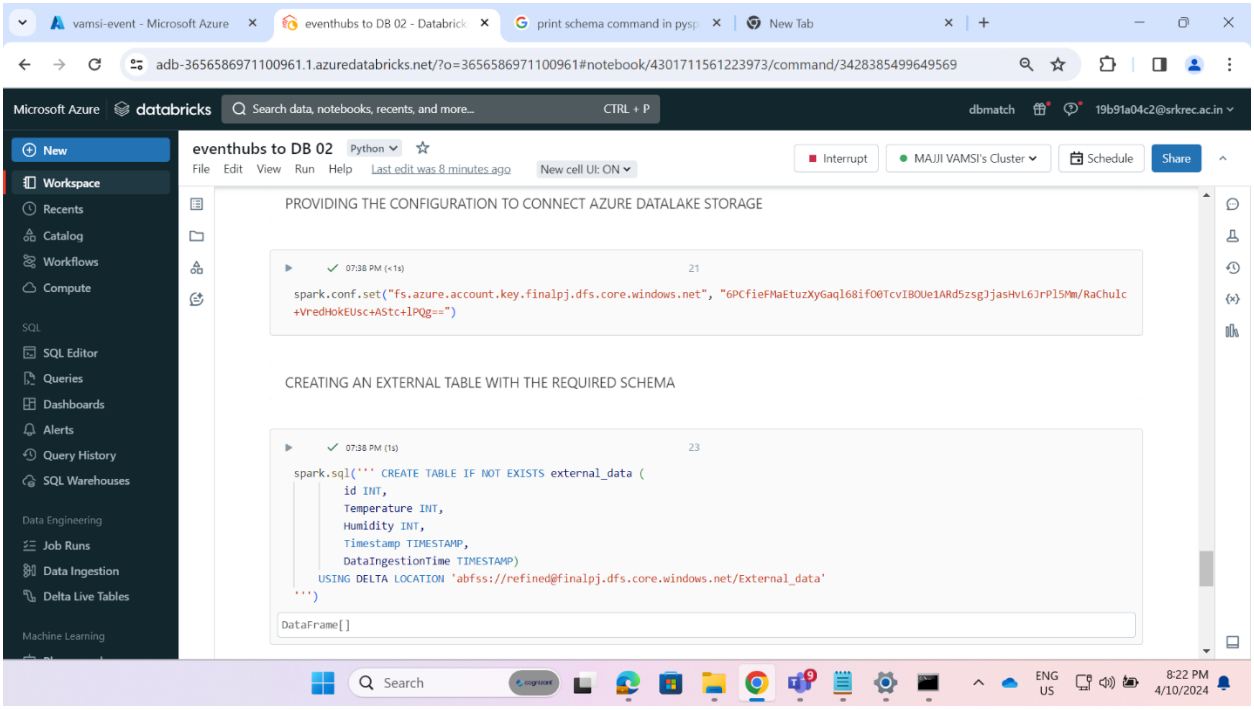
**STEP 12:** Ingesting a new column for data ingestion timing.



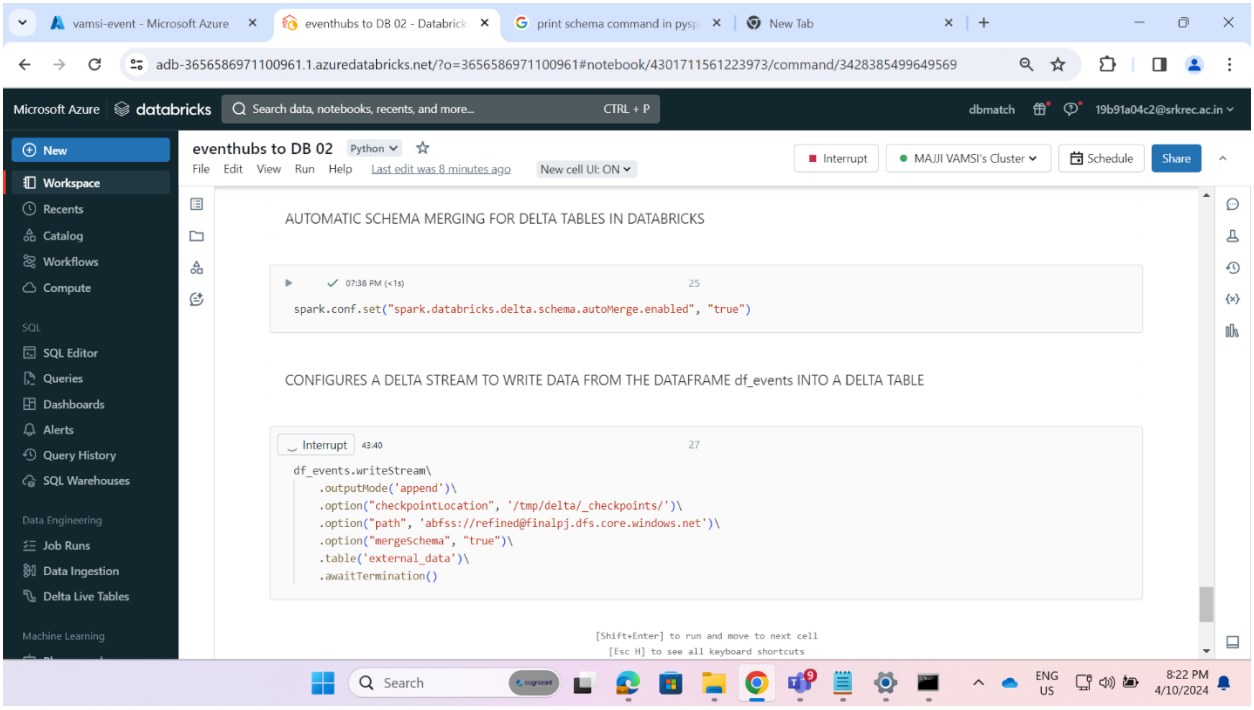

**STEP 13:** Display the newly added column to table.



**STEP 14:** Connecting with storage account Databricks and creating a required schema for external table.

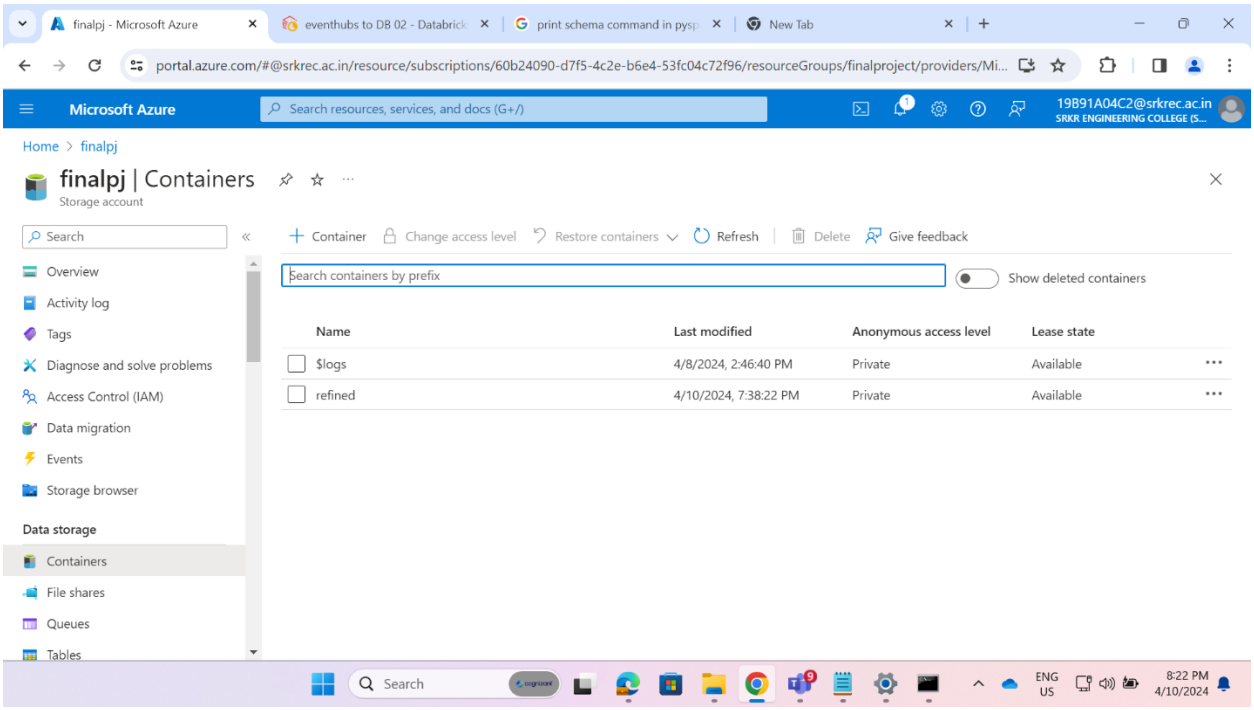


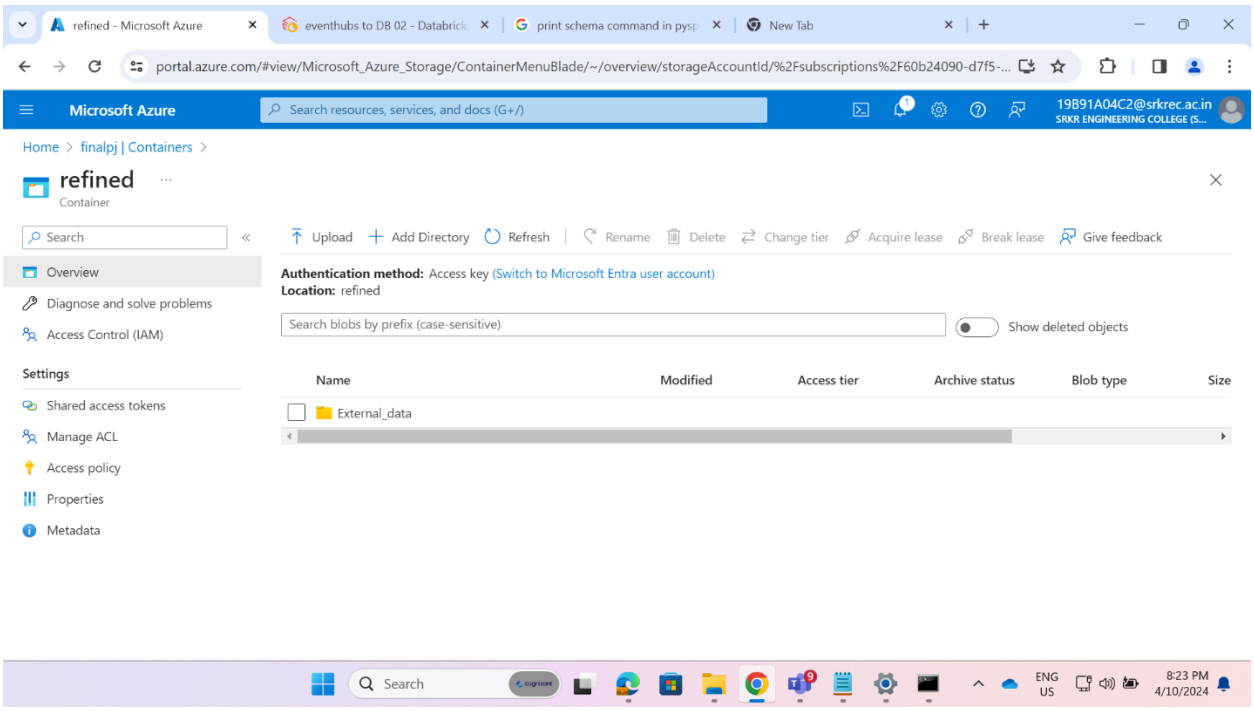

**STEP 15:** Ingesting Streaming data into the external table.

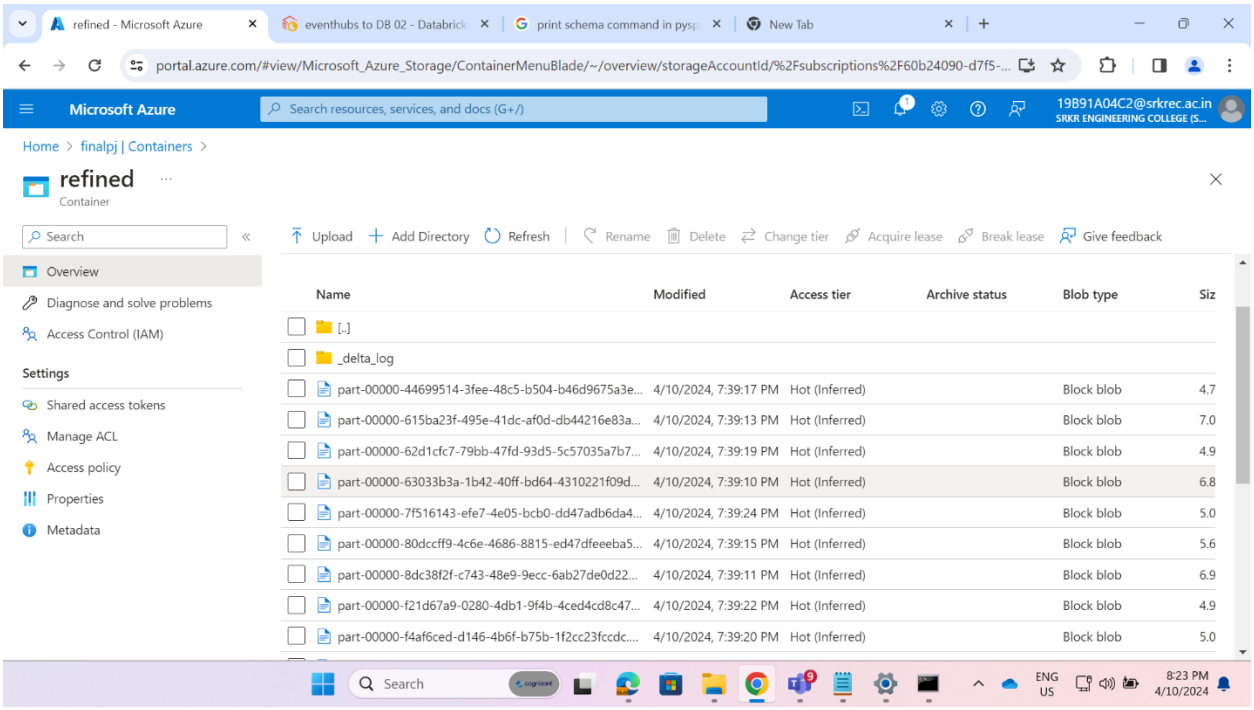


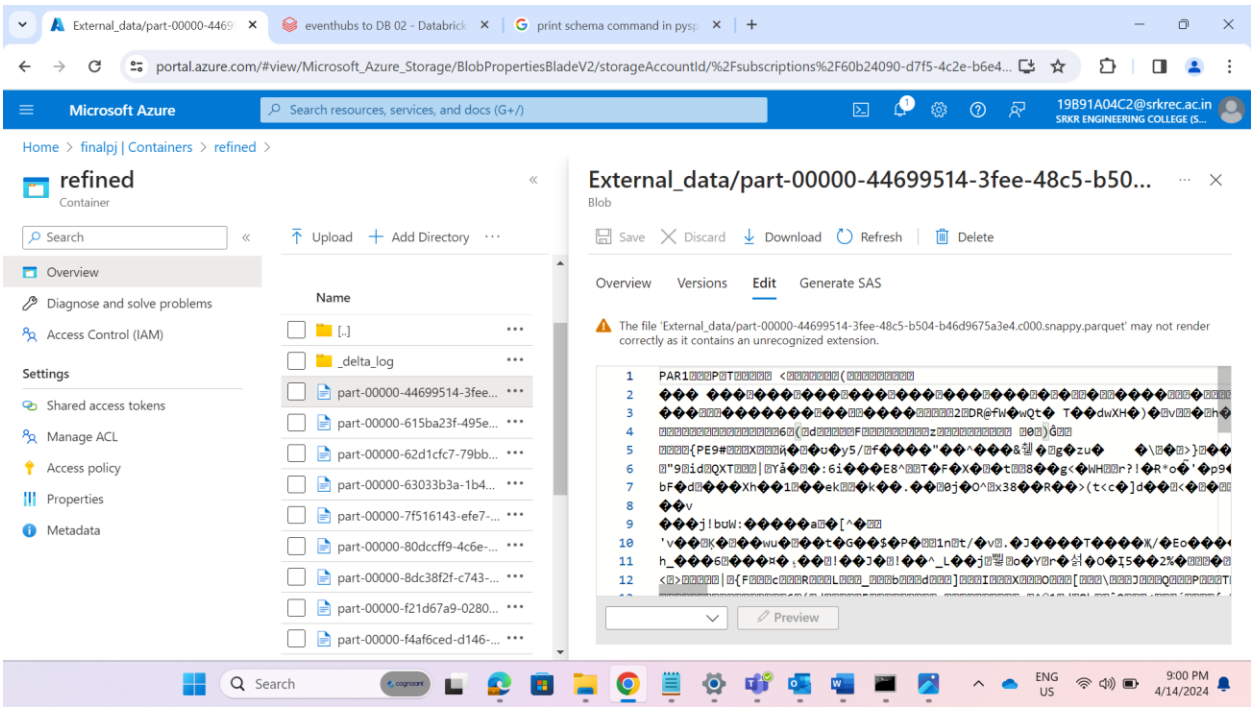
# VISUALIZATION OF DATA:


# Result:




refined - Microsoft Azure
eventhubs to DB 02 - Databrick
print schema command in pysp
New Tab
portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F60b24090-d7f5-...
Microsoft Azure
Search resources, services, and docs (G+/)
19B91A04C2@srkrec.ac.in
SRKR ENGINEERING COLLEGE (S...
Home > finalpj | Containers >
refined
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: refined
Search blobs by prefix (case-sensitive)
Show deleted objects
Name
Modified
Access tier
Archive status
Blob type
Size
External_data
Search
ENG
US
8:23 PM
4/10/2024


refined - Microsoft Azure
eventhubs to DB 02 - Databrick
print schema command in pysp
New Tab
portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F60b24090-d7f5-...
Microsoft Azure
Search resources, services, and docs (G+/)
19B91A04C2@srkrec.ac.in
SRKR ENGINEERING COLLEGE (S...
Home > finalpj | Containers >
refined
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
Name
Modified
Access tier
Archive status
Blob type
Siz
[..]
_delta_log
part-00000-44699514-3fee-48c5-b504-b46d9675a3e... 4/10/2024, 7:39:17 PM Hot (Inferred) Block blob 4.7
part-00000-615ba23f-495e-41dc-af0d-db44216e83a... 4/10/2024, 7:39:13 PM Hot (Inferred) Block blob 7.0
part-00000-62d1cfc7-79bb-47fd-93d5-5c57035a7b7... 4/10/2024, 7:39:19 PM Hot (Inferred) Block blob 4.9
part-00000-63033b3a-1b42-40ff-bd64-4310221f09d... 4/10/2024, 7:39:10 PM Hot (Inferred) Block blob 6.8
part-00000-7f516143-efe7-4e05-bcb0-dd47adb6da4... 4/10/2024, 7:39:24 PM Hot (Inferred) Block blob 5.0
part-00000-80dcff9-4c6e-4686-8815-ed47dfeeeba5... 4/10/2024, 7:39:15 PM Hot (Inferred) Block blob 5.6
part-00000-8dc38f2f-c743-48e9-9ecc-6ab27de0d22... 4/10/2024, 7:39:11 PM Hot (Inferred) Block blob 6.9
part-00000-f21d67a9-0280-4db1-9f4b-4ced4cd8c47... 4/10/2024, 7:39:22 PM Hot (Inferred) Block blob 4.9
part-00000-f4af6ced-d146-4b6f-b75b-1f2cc23fccdc.... 4/10/2024, 7:39:20 PM Hot (Inferred) Block blob 5.0
Search
ENG
US
8:23 PM
4/10/2024


External_data/part-00000-44699514-3fee-48c5-b50...
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Edit
Generate SAS
The file 'External_data/part-00000-44699514-3fee-48c5-b504-b46d9675a3e4.c000.snappy.parquet' may not render correctly as it contains an unrecognized extension.
Preview
refined
Container
Search
Upload
Add Directory
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Shared access tokens
Manage ACL
Access policy
Properties
Metadata
Name
[..]
_delta_log
part-00000-44699514-3fee...
part-00000-615ba23f-495e...
part-00000-62d1cfc7-79bb...
part-00000-63033b3a-1b4...
part-00000-7f516143-efe7-...
part-00000-80dccff9-4c6e-...
part-00000-8dc38f2f-c743-...
part-00000-f21d67a9-0280...
part-00000-f4af6ced-d146-...
Microsoft Azure
Search resources, services, and docs (G+/)
19B91A04C2@srkrec.ac.in
SRKR ENGINEERING COLLEGE (S...
Home > finalpj | Containers > refined >
portal.azure.com/#view/Microsoft_Azure_Storage/BlobPropertiesBladeV2/storageAccountId/%2Fsubscriptions%2F60b24090-d7f5-4c2e-b6e4...
eventhubs to DB 02 - Databrick
print schema command in pysp
Search
ENG
US
9:00 PM
4/14/2024