

**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:**

**Input Script**

Weather data producer

**Azure EventHub Namespace**

Event

(Connection String)

**Azure DataBricks**

Clusters

Notebook

Pyspark

External Table

(Streaming data)

**STEP 1:** Create an event-hub namespace and event-hub (namespace-vamsi-event) (eventhub-**A screenshot of a computer

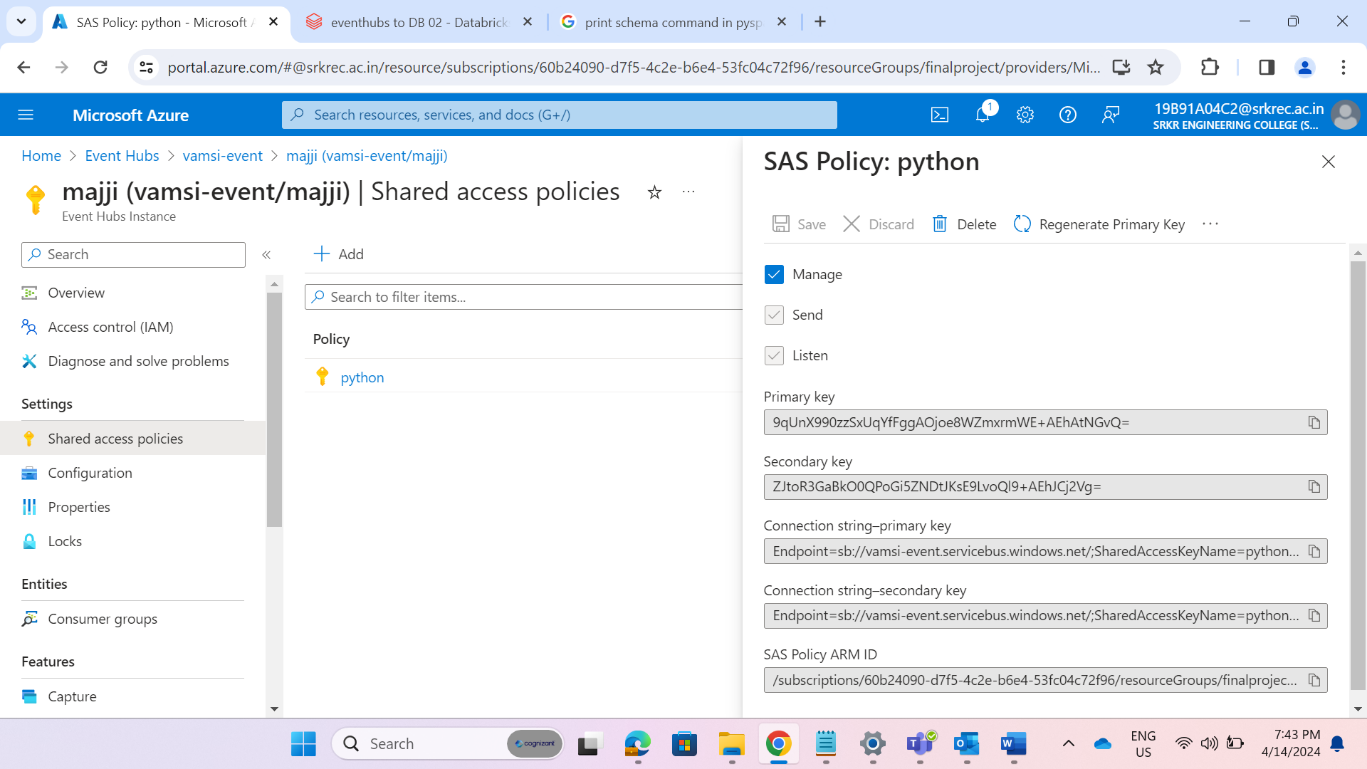
Description automatically generated**majji).

**ADLS**

Refined Container

/<file>. parquet

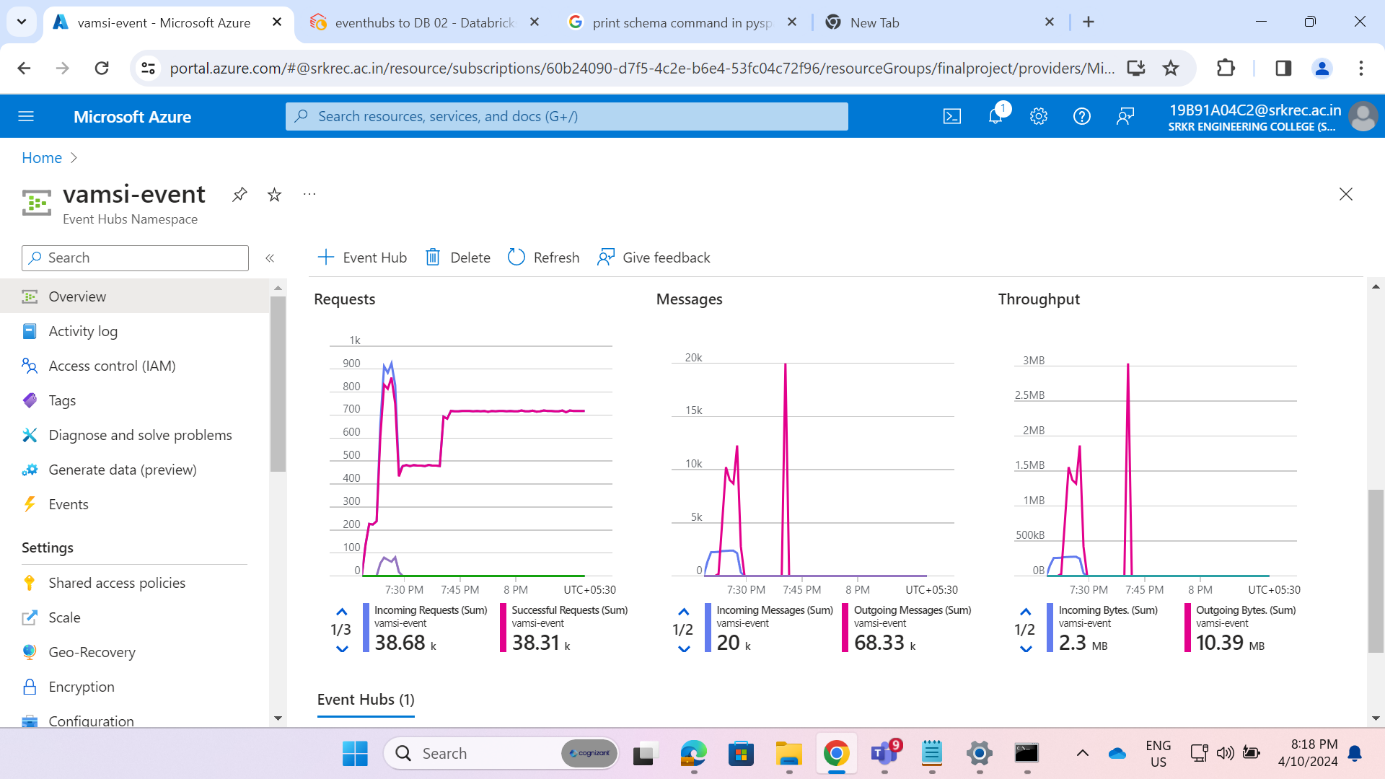
Visualization

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

A screenshot of a computer screen

Description automatically generated**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.



A screenshot of a computer

Description automatically generated**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).

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

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

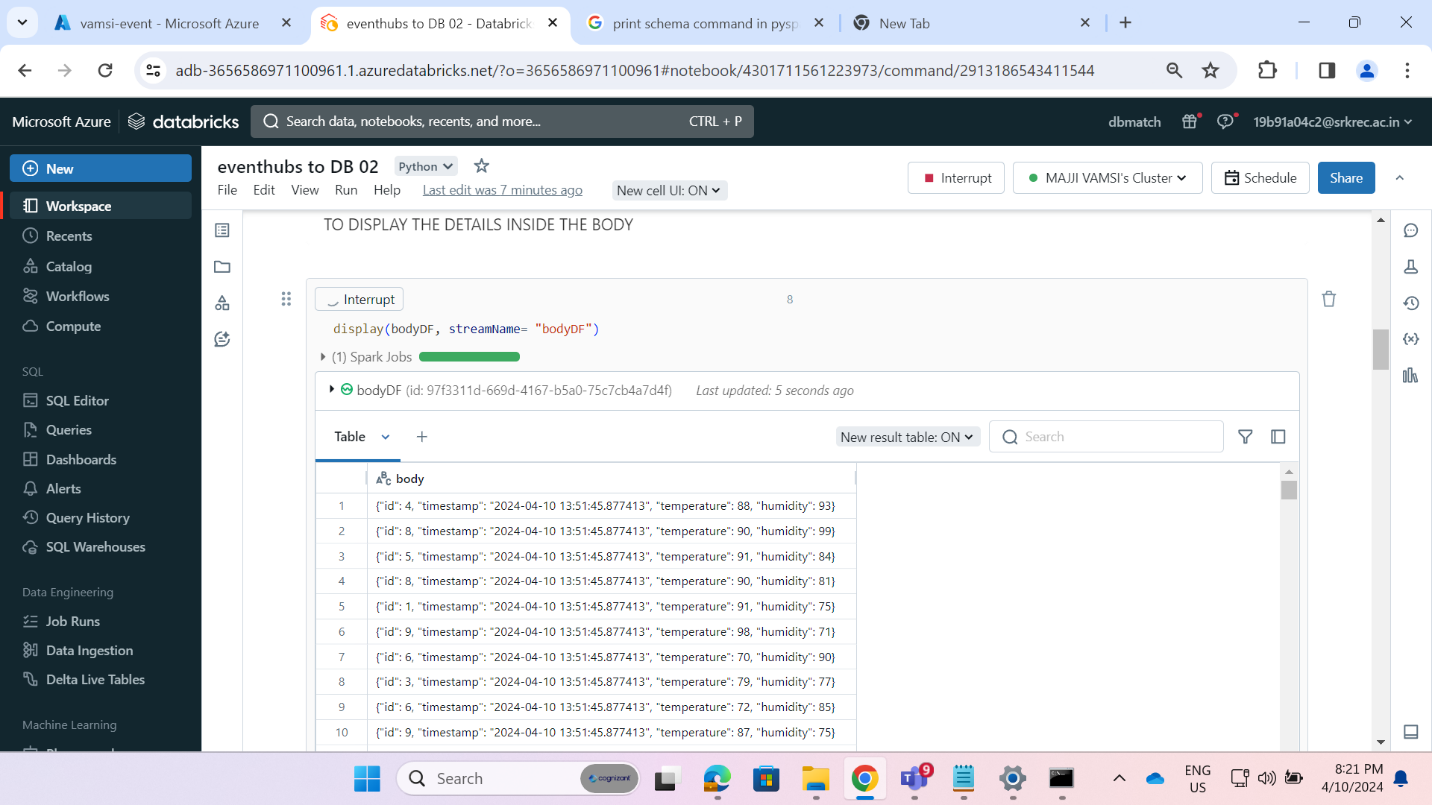
**A screenshot of a computer

Description automatically generatedSTEP 8:** Reading the streaming data form the EventHub.

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

A screenshot of a computer

Description automatically generated

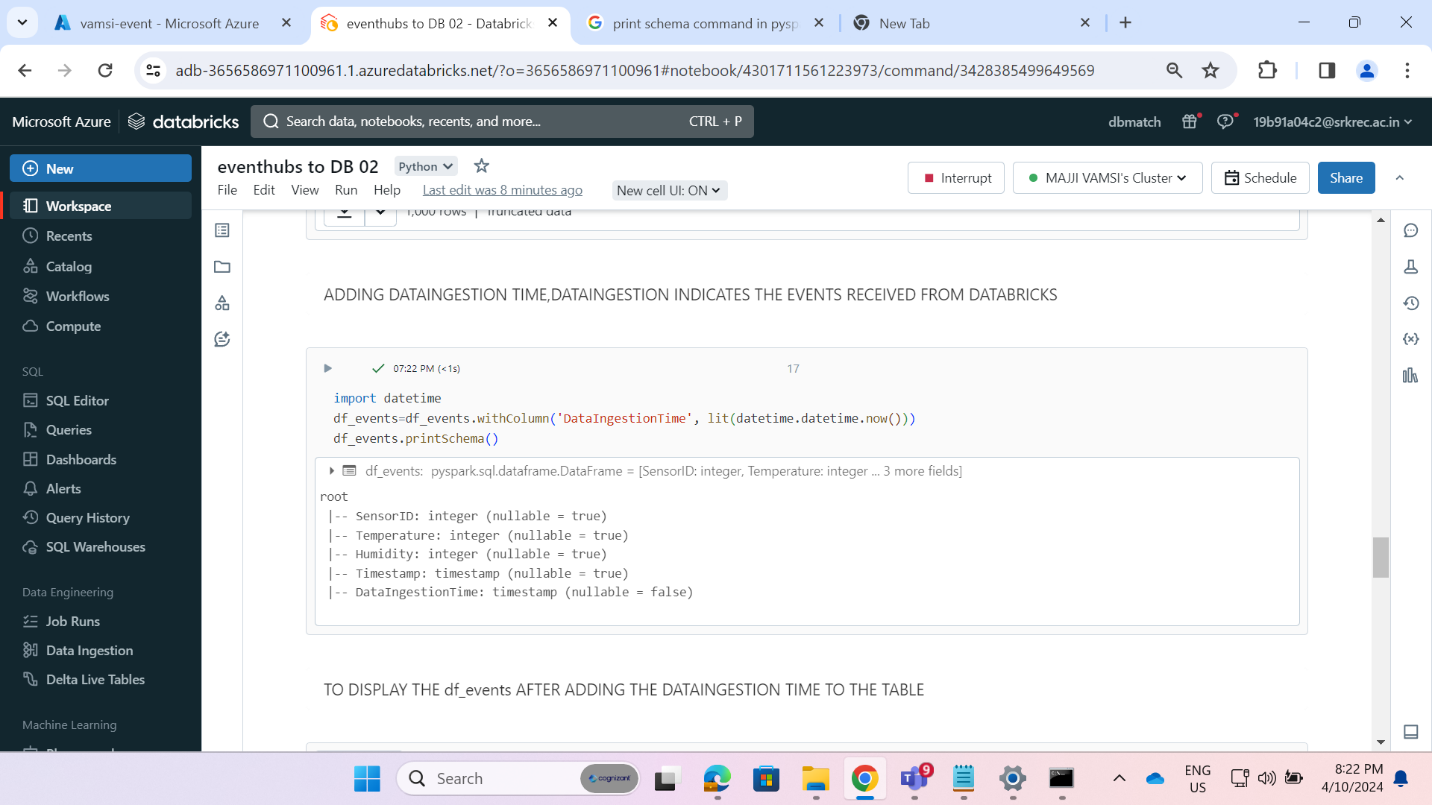


**STEP 10**: creating schema for the dataframe.

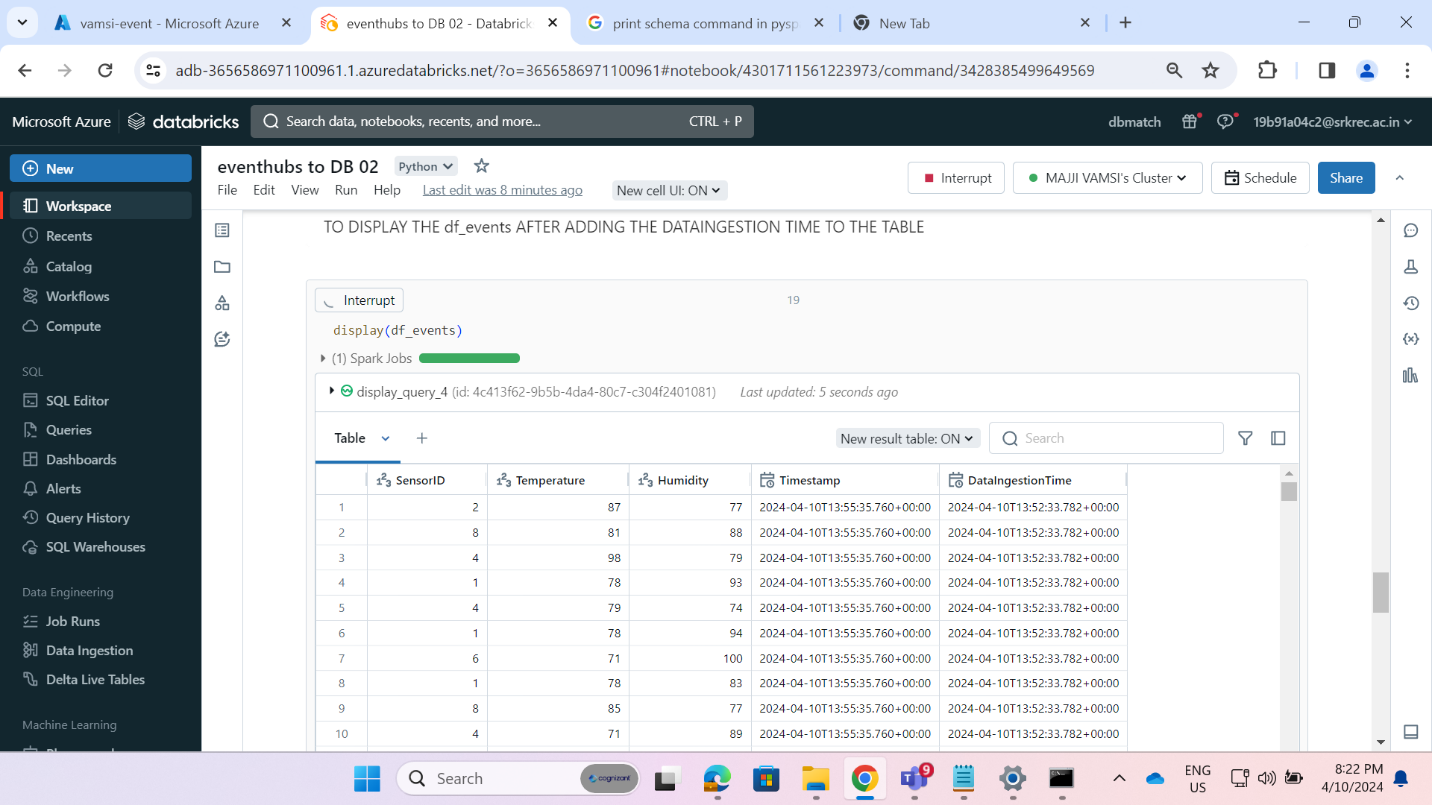
**A screenshot of a computer

Description automatically generatedSTEP 11:** Changing the datatypes and storing the streaming data into columns.

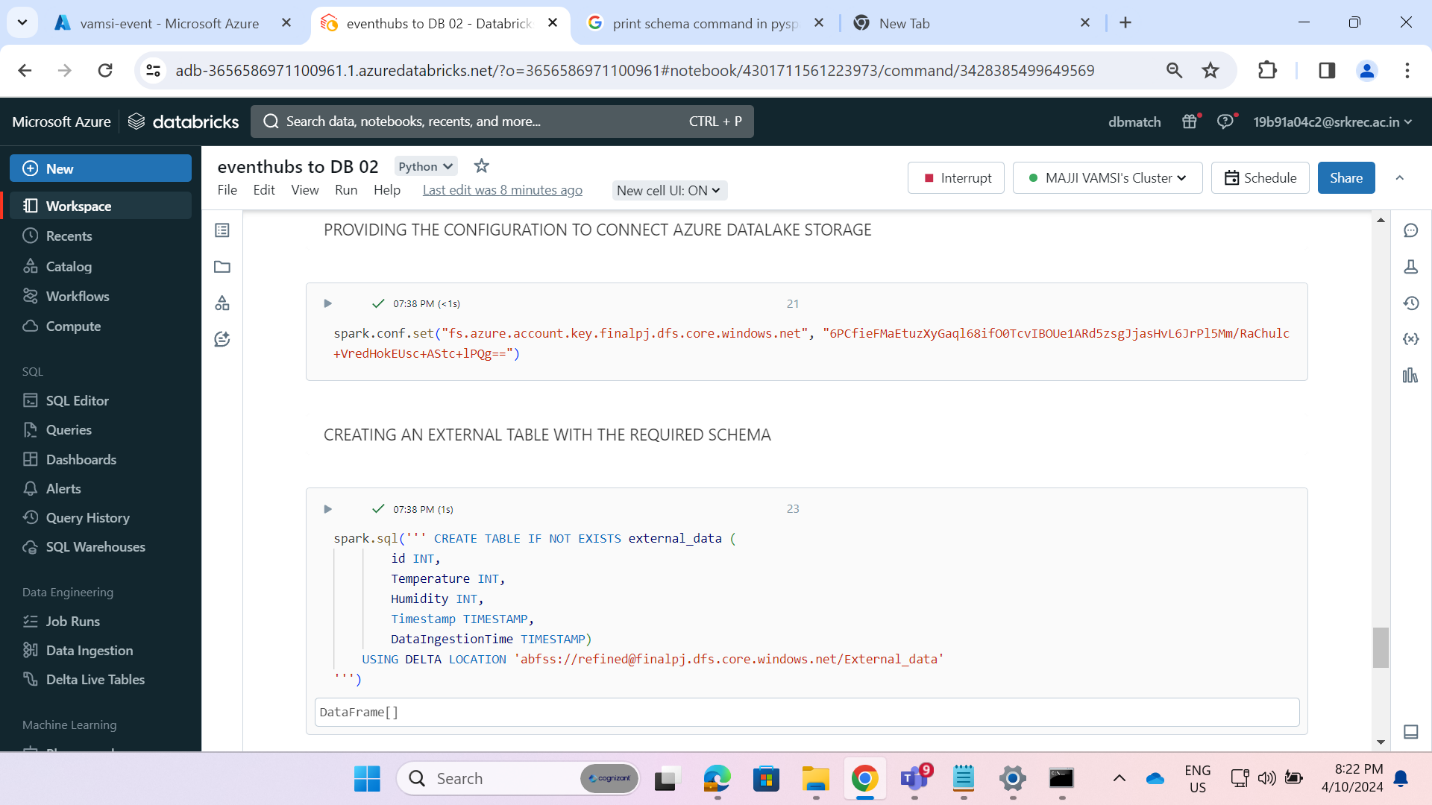
**STEP 12:** Ingesting a new column for data ingestion timing.

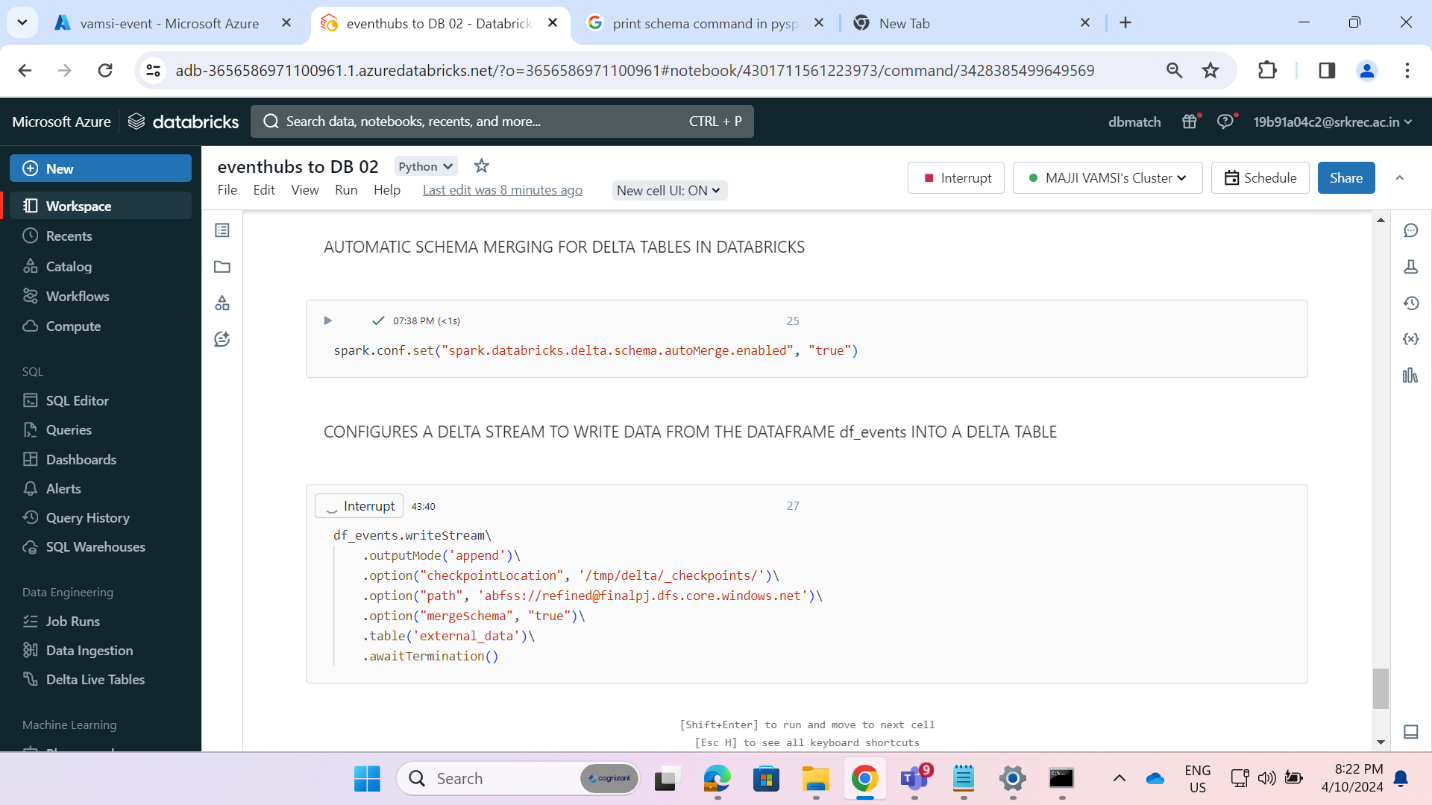
****

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

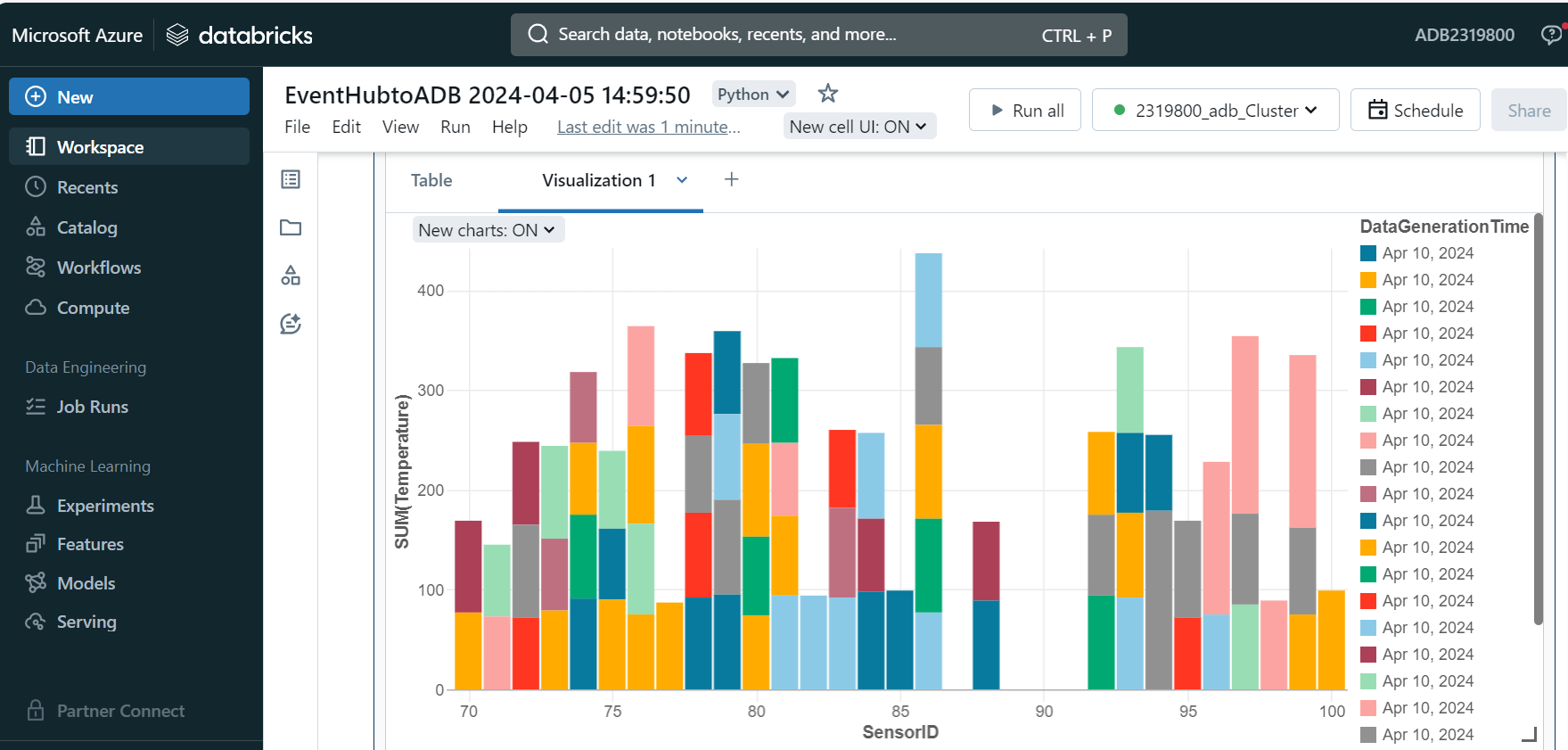


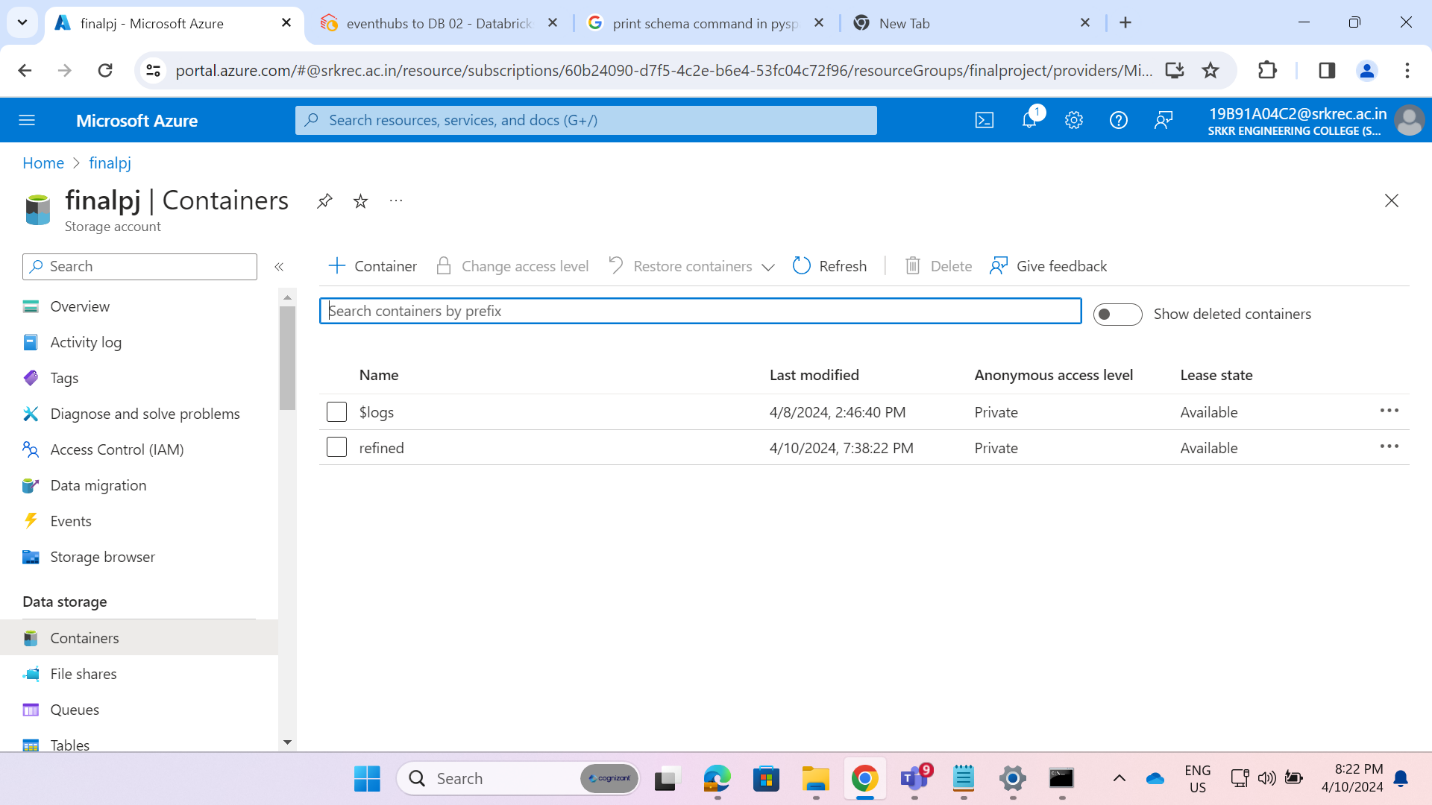
**STEP 14:** Connecting with storage account Databricks and creating a required schema for external table.

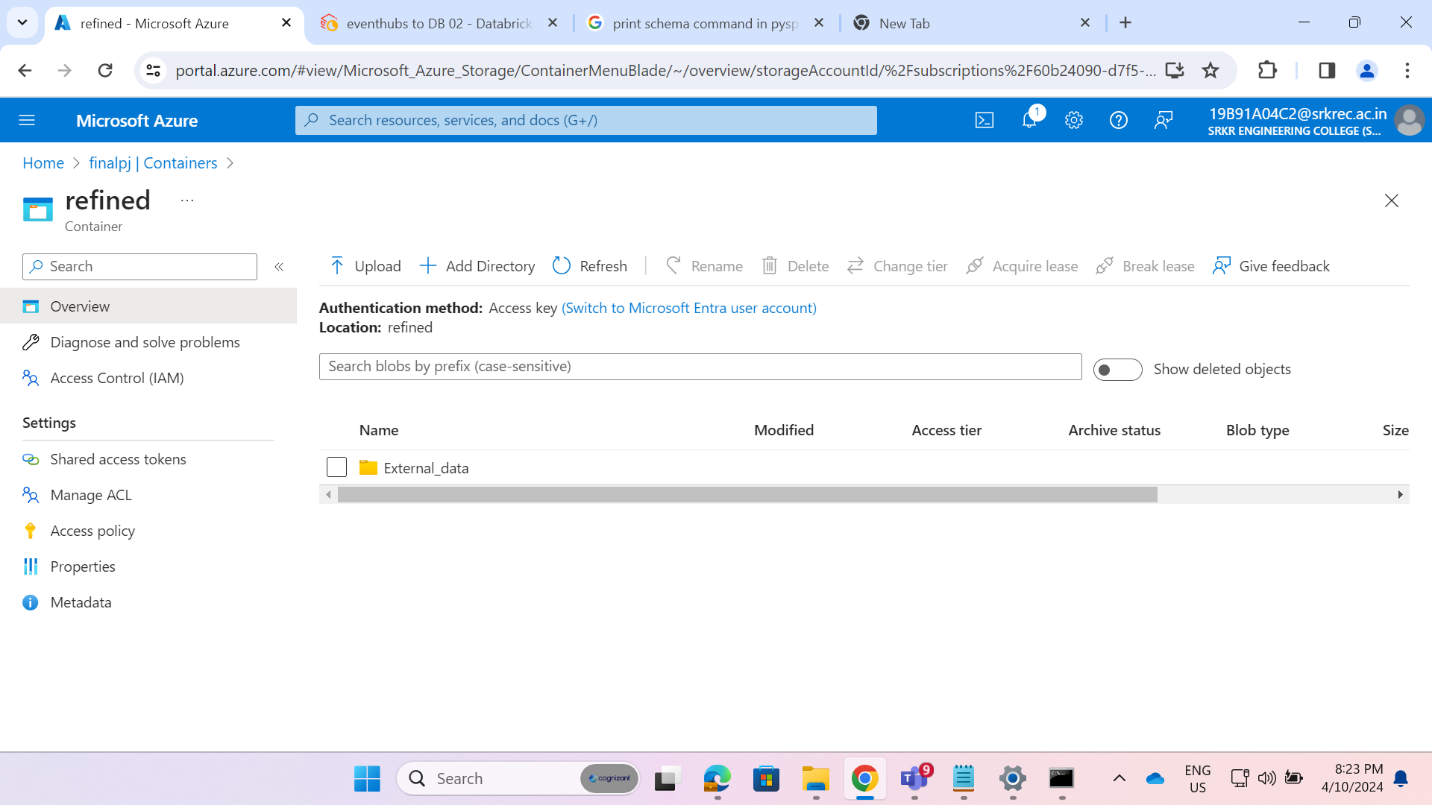
****

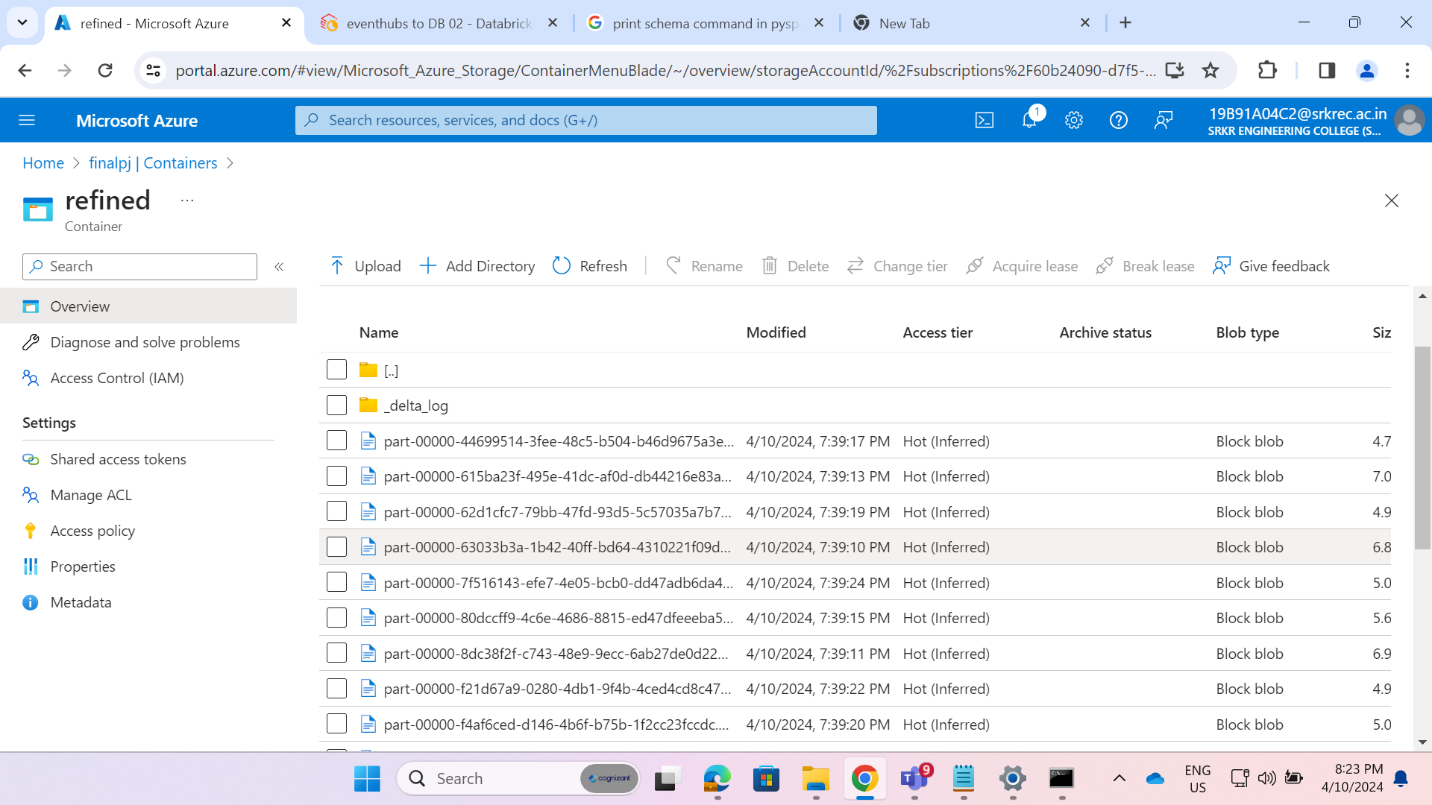
**STEP 15:** Ingesting Streaming data into the external table.

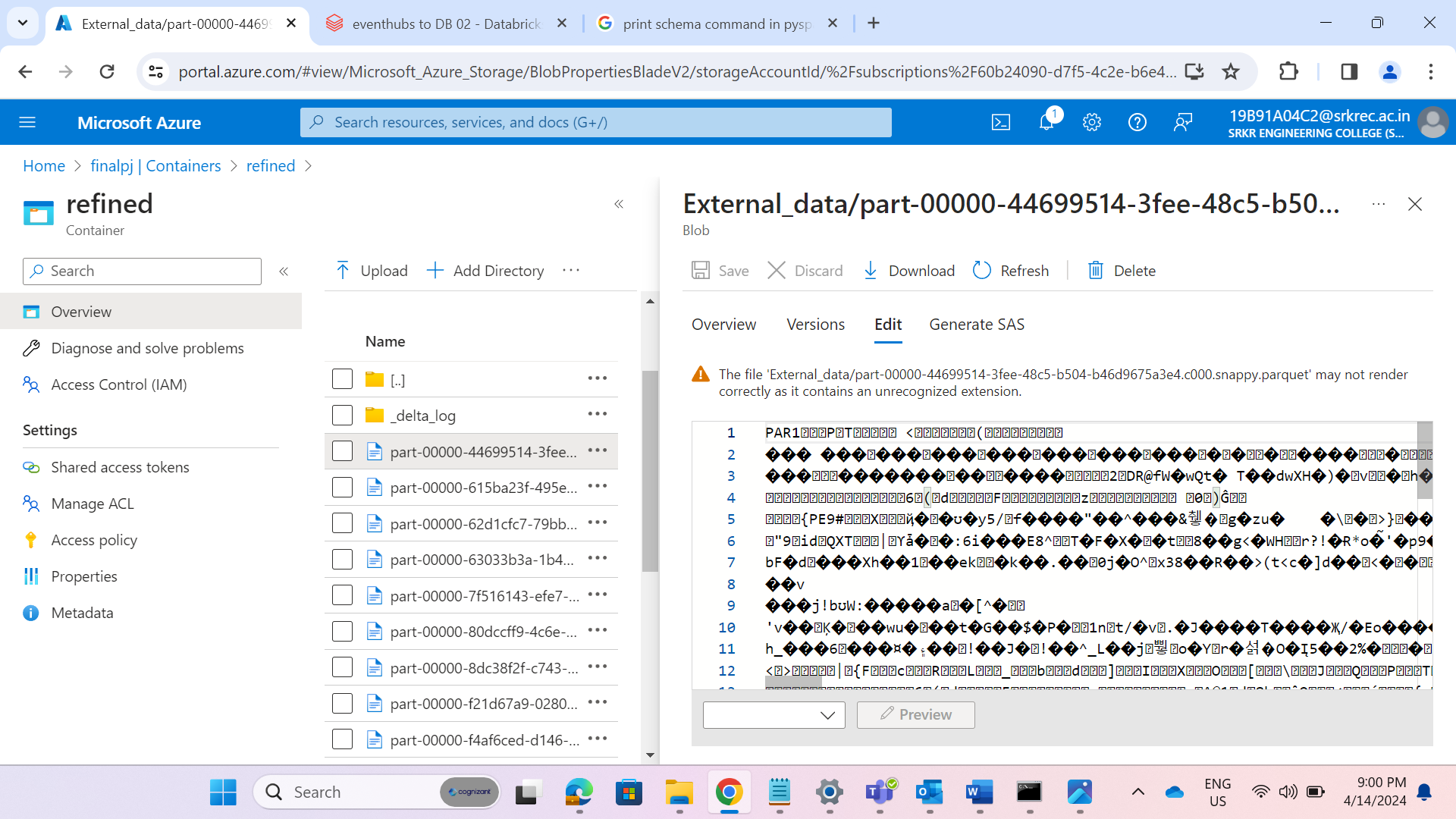
**VISUALIZATION OF DATA:**



**Result:**

****

****

****