



RETAIL X

*This document presents a comprehensive overview of the **RetailX project**, an end-to-end real-time retail analytics pipeline built using Microsoft Azure and Python. Developed as part of a data engineering learning initiative, this project demonstrates the integration of cloud-native tools to simulate, ingest, process, store, and visualize streaming retail transaction data.*

The pipeline leverages a combination of Azure services and open-source tools to simulate real-time transaction events, stream them through the Azure ecosystem, and generate actionable business insights through Power BI dashboards. The project emphasizes the practical application of cloud architecture and real-time data engineering principles in the retail domain.

Krishna Cha Muttevi

Project Description:

RetailX is a cloud-native, real-time retail data pipeline built using Microsoft Azure, Python, and Power BI. It simulates live transaction data, processes it on the fly using Azure Stream Analytics, stores it in Azure Data Lake Gen2, and visualizes the insights using Power BI dashboards. This project demonstrates scalable event-driven architecture and live business intelligence.

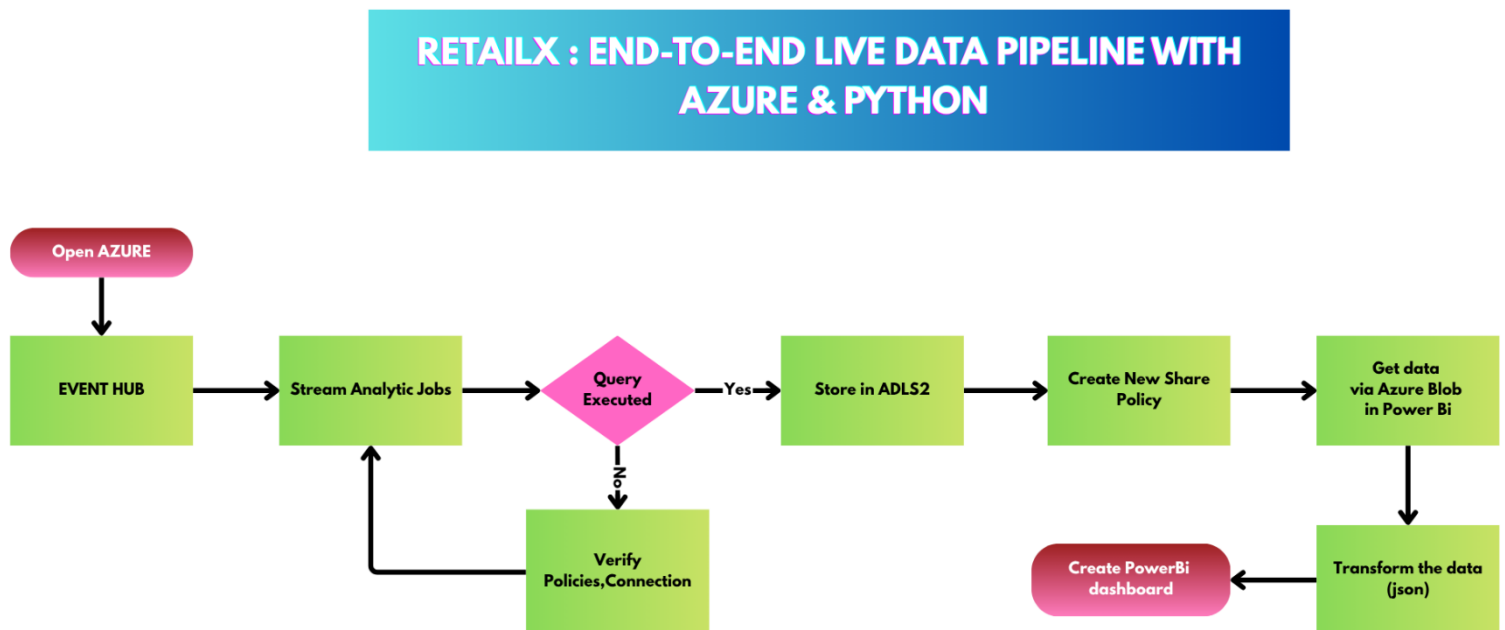
Tech Stack:

Tool	Purpose
Azure Event Hub	Real-time event ingestion
Azure Stream Analytics	SQL-based stream processing
Azure Data Lake Gen2	Storage for raw and transformed data
Python (VS Code)	Real-time data simulator
Power BI	Live data visualization
JSON	Stream data format

Key Features:

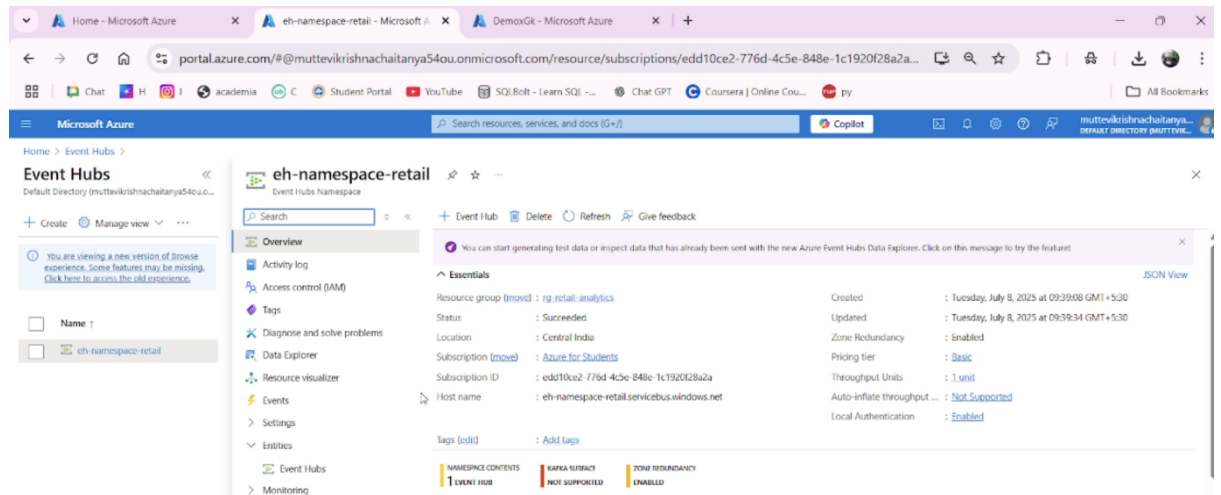
- Real-time data ingestion using **Azure Event Hub**
- SQL-based stream processing with **Azure Stream Analytics**
- Scalable cloud storage via **ADLS Gen2**
- Integration with **Power BI** for live dashboards
- Python-based data simulator for real-time publishing

Architecture Diagram:



EVENT-HUB:

1.Namespace:

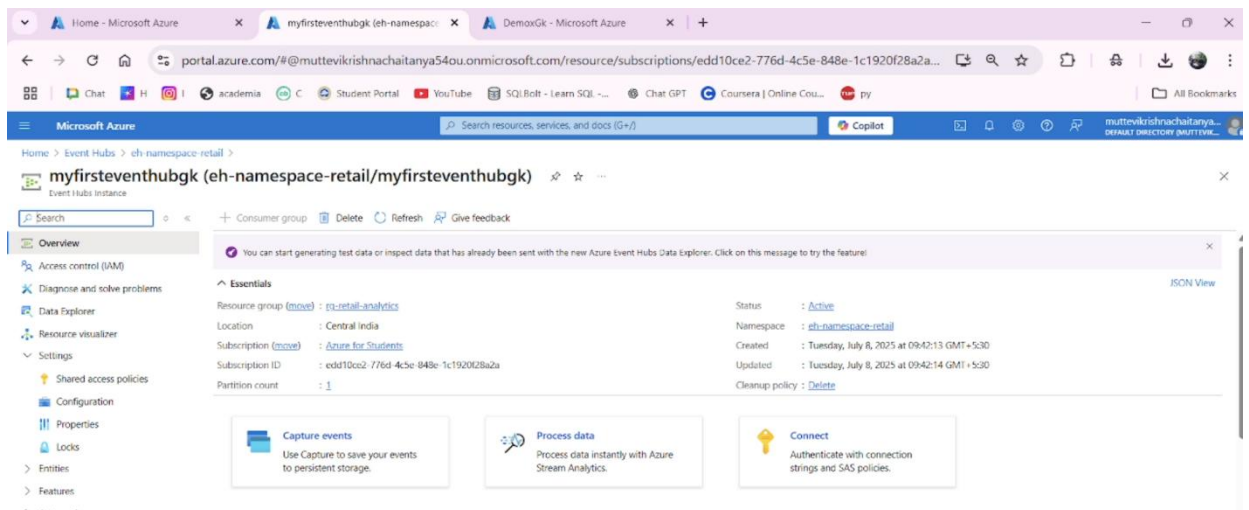


The screenshot displays the Microsoft Azure portal interface for the 'eh-namespace-retail' Event Hubs Namespace. The left sidebar shows the navigation menu with options like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Data Explorer', 'Resource visualizer', 'Settings', 'Entities', 'Event Hubs', and 'Monitoring'. The main content area shows the 'Overview' tab for the namespace. Key details include:

- Resource group (moved):** rg-retail-analytics
- Status:** Succeeded
- Location:** Central India
- Subscription (moved):** Azure for Students
- Subscription ID:** edd10ce2-776d-4c5e-848e-1c1920f28a2a
- Host name:** eh-namespace-retail.servicebus.windows.net
- Created:** Tuesday, July 8, 2025 at 09:39:08 GMT+5:30
- Updated:** Tuesday, July 8, 2025 at 09:39:34 GMT+5:30
- Zone Redundancy:** Enabled
- Pricing tier:** Basic
- Throughput Units:** 1 unit
- Auto-inflate throughput ...:** Not Supported
- Local Authentication:** Enabled

At the bottom, there are three status indicators: 'NAMESPACE CONTENTS' (1 EVENT HUB), 'KAFKA SUPPORT' (NOT SUPPORTED), and 'ZONE REDUNDANCY' (ENABLED).

2. Event Hub:

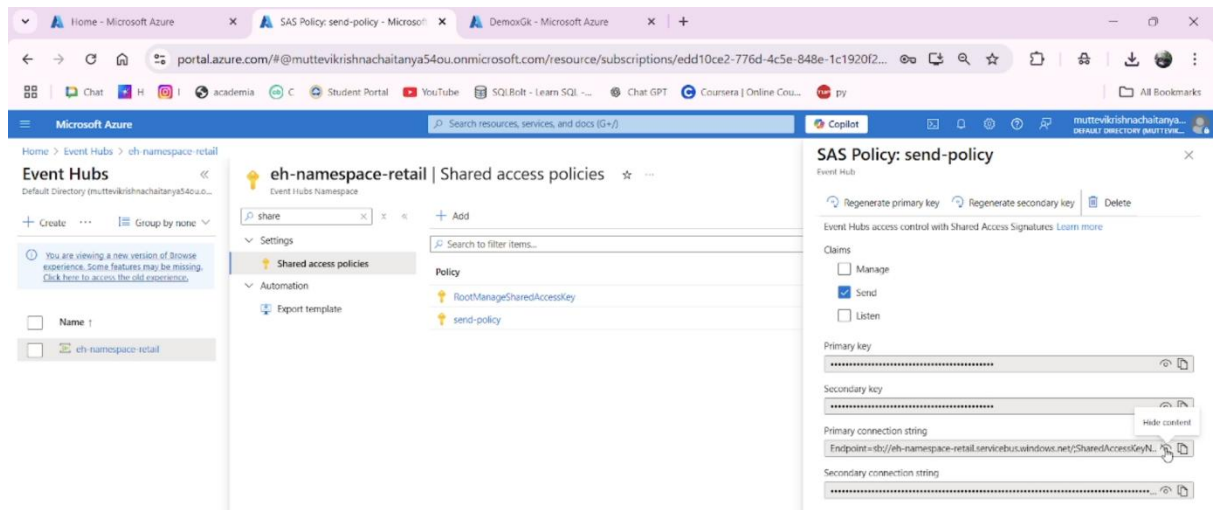


The screenshot displays the Microsoft Azure portal interface for the 'myfirsteventhubgk' Event Hub within the 'eh-namespace-retail' namespace. The left sidebar shows the navigation menu with options like 'Overview', 'Access control (IAM)', 'Diagnose and solve problems', 'Data Explorer', 'Resource visualizer', 'Settings', 'Shared access policies', 'Configuration', 'Properties', 'Locks', 'Entities', 'Features', and 'Automation'. The main content area shows the 'Overview' tab for the event hub. Key details include:

- Resource group (moved):** rg-retail-analytics
- Location:** Central India
- Subscription (moved):** Azure for Students
- Subscription ID:** edd10ce2-776d-4c5e-848e-1c1920f28a2a
- Partition count:** 1
- Status:** Active
- Namespace:** eh-namespace-retail
- Created:** Tuesday, July 8, 2025 at 09:42:13 GMT+5:30
- Updated:** Tuesday, July 8, 2025 at 09:42:14 GMT+5:30
- Cleanup policy:** Delete

At the bottom, there are three action buttons: 'Capture events' (Use Capture to save your events to persistent storage), 'Process data' (Process data instantly with Azure Stream Analytics), and 'Connect' (Authenticate with connection strings and SAS policies).

3. Creation of New Shared Access Policy:

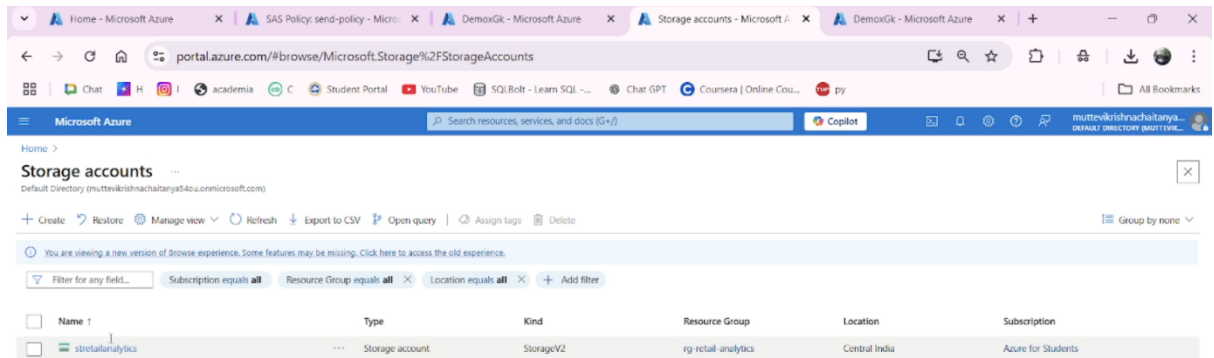


4. Python file:

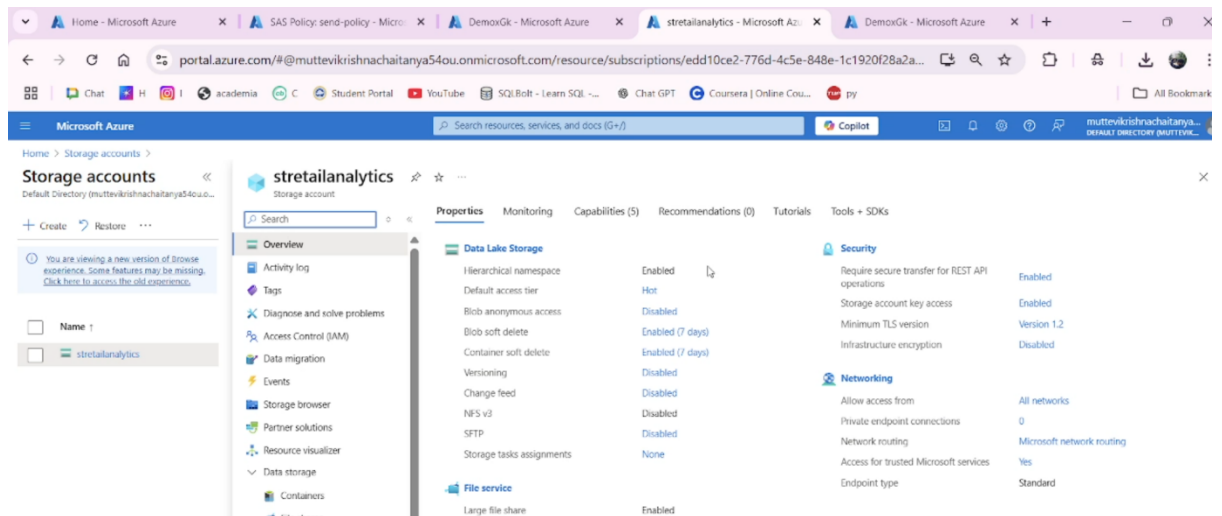
```
financial animation.ipynb 1.mod_importer.ipynb simulate trans.py password.py stats.py time.py samp.txt
1 import json, time, random
2 from azure.eventhub import EventHubProducerClient, EventData
3 from faker import Faker
4
5 fake = Faker()
6 conn_str = "Endpoint=sb://eh-namespace-retail.servicebus.windows.net/;SharedAccessKeyName=send-policy;SharedAccessKey=ek4zvE7q"
7 eventhub_name = "myfirsteventhubgk"
8
9 producer = EventHubProducerClient.from_connection_string(conn_str, eventhub_name=eventhub_name)
10
11 def generate_transaction():
12     return {
13         "transaction_id": f"TX-{random.randint(1000,9999)}",
14         "product_id": f"P{random.randint(1, 10):03d}",
15         "store_id": f"S{random.randint(1, 5):02d}",
16         "channel": random.choice(["online", "store"]),
17         "timestamp": fake.date_time_this_year().isoformat(),
18         "quantity": random.randint(1, 5),
19         "price": round(random.uniform(10.0, 100.0), 2)
20     }
21
22 while True:
23     event_data = EventData(json.dumps(generate_transaction()))
24     producer.send_batch([event_data])
25     print("Event sent")
26     time.sleep(1)
27
```

ADLS 2:

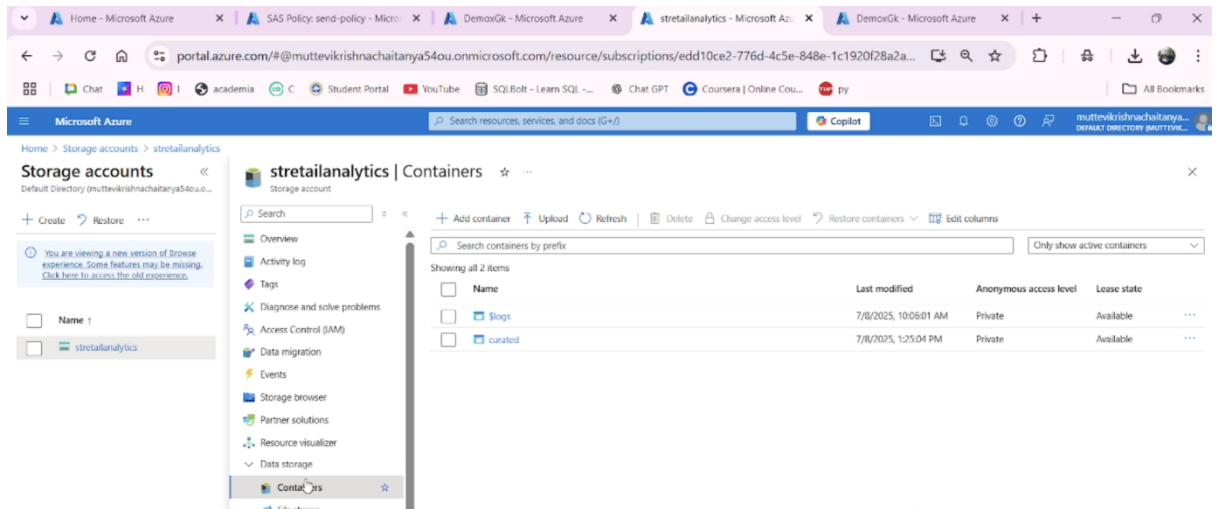
1. New Storage Account:



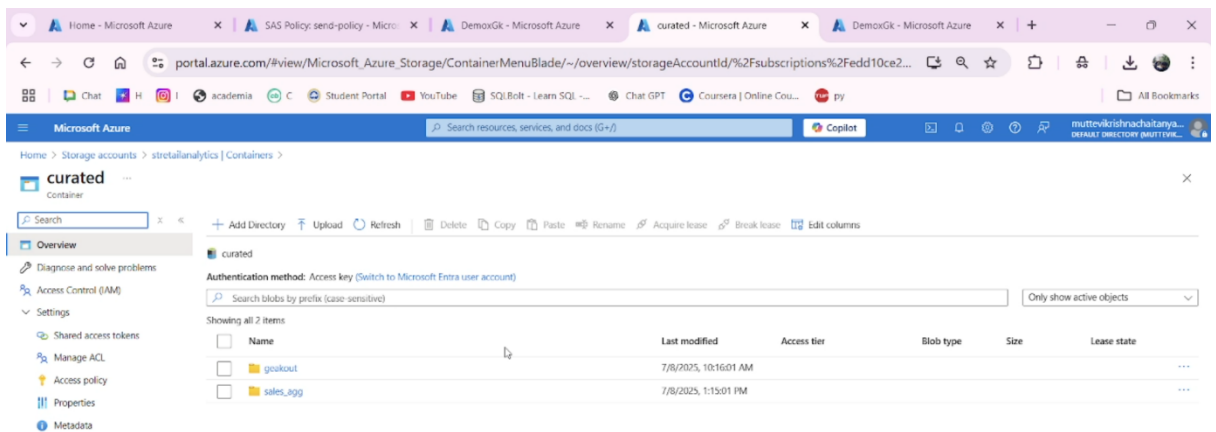
2. Enable Hierarchical Namespace:



3. Create a Container:

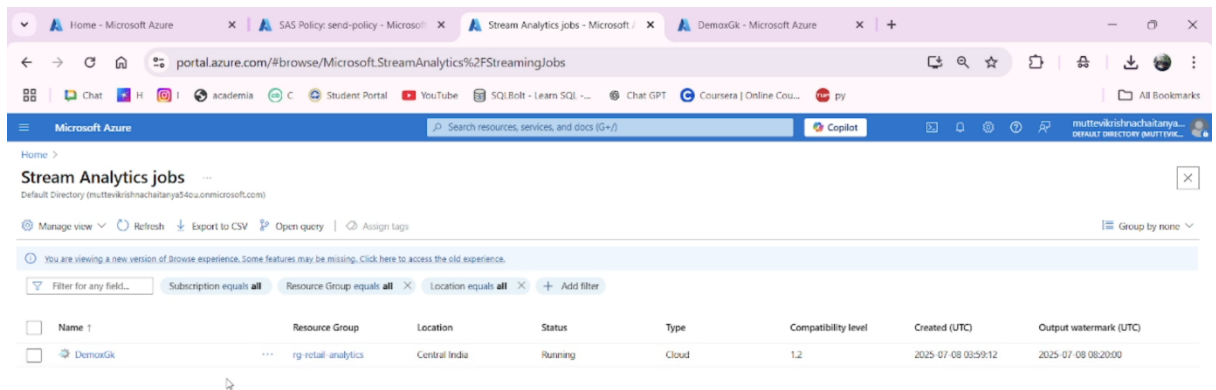


4. Inside Container Create a new Directory:



Stream Analytics:

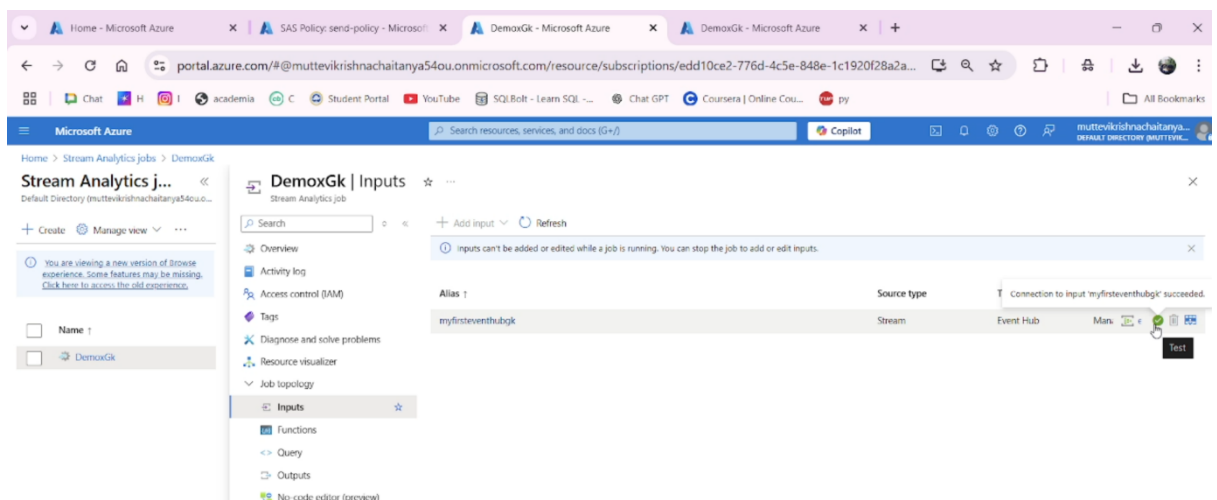
1. Job Creation:



The screenshot shows the 'Stream Analytics jobs' page in the Microsoft Azure portal. The page title is 'Stream Analytics jobs' and the URL is 'portal.azure.com/#browse/Microsoft.StreamAnalytics%2FStreamingJobs'. The page includes a search bar, a 'Manage view' dropdown, and a 'Refresh' button. A message states: 'You are viewing a new version of Browse experience. Some features may be missing. Click here to access the old experience.' Below this, there are filters for 'Subscription equals all', 'Resource Group equals all', and 'Location equals all'. A table lists the jobs:

Name	Resource Group	Location	Status	Type	Compatibility level	Created (UTC)	Output watermark (UTC)
DemoxGk	rg-retail-analytics	Central India	Running	Cloud	1.2	2025-07-08 03:59:12	2025-07-08 08:20:00

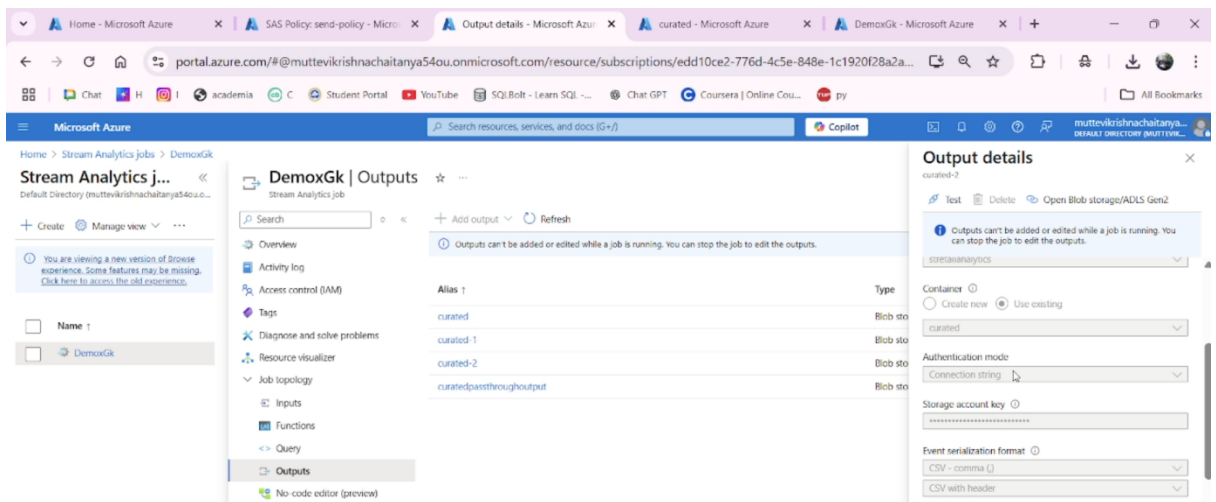
2. Input Definition:



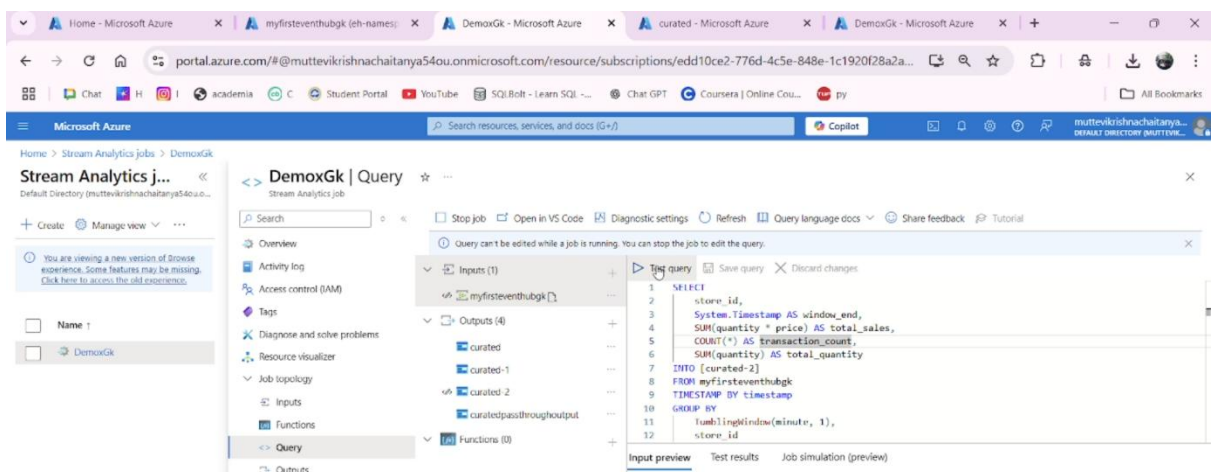
The screenshot shows the 'DemoxGk | Inputs' page in the Microsoft Azure portal. The page title is 'DemoxGk | Inputs' and the URL is 'portal.azure.com/#/@muttevikrishnachaitanya54ou.onmicrosoft.com/resource/subscriptions/edd10ce2-776d-4c5e-848e-1c1920f28a2a...'. The page includes a search bar, a 'Refresh' button, and a message: 'Inputs can't be added or edited while a job is running. You can stop the job to add or edit inputs.' Below this, there are tabs for 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Resource visualizer', 'Job topology', 'Inputs', 'Functions', 'Query', 'Outputs', and 'No-code editor (preview)'. A table lists the inputs:

Alias	Source type	Connection to input 'myfirsteventhubjk' succeeded.
myfirsteventhubjk	Stream	Event Hub

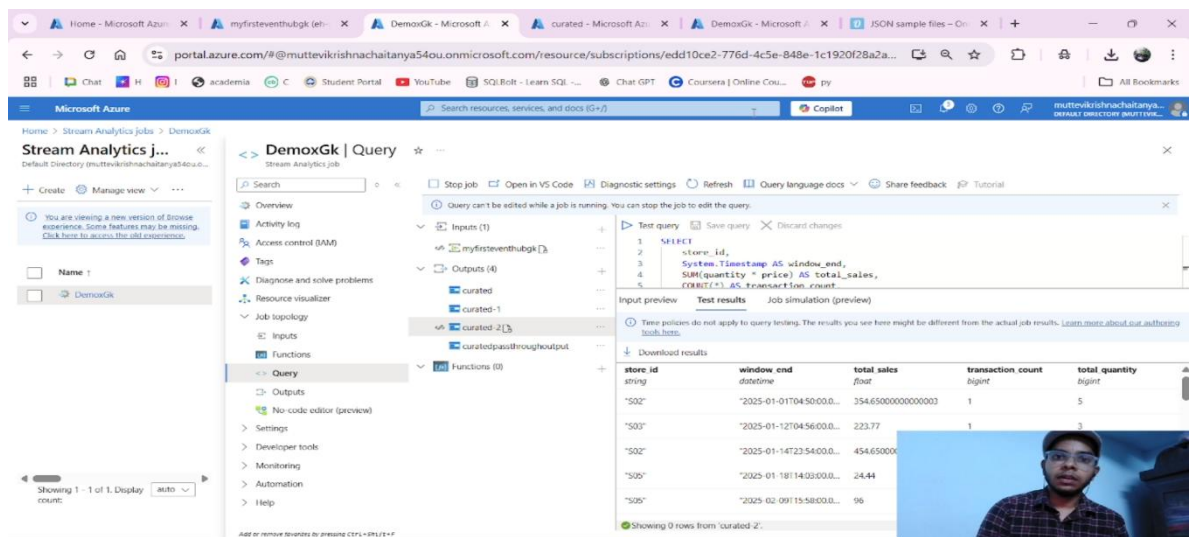
3. Output Definition:



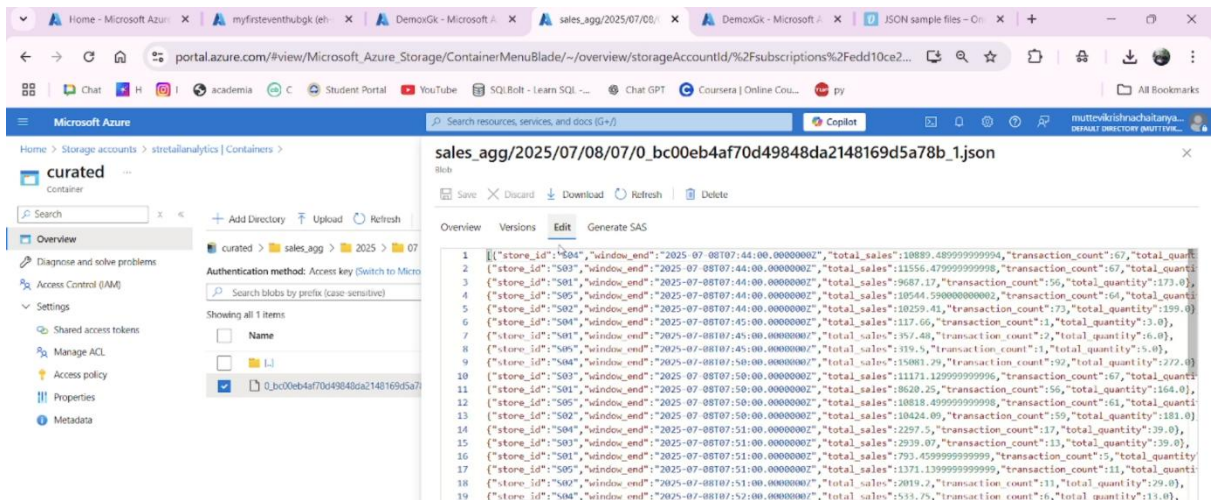
4. Start Job + Run Query:



5. Verify Query Output:

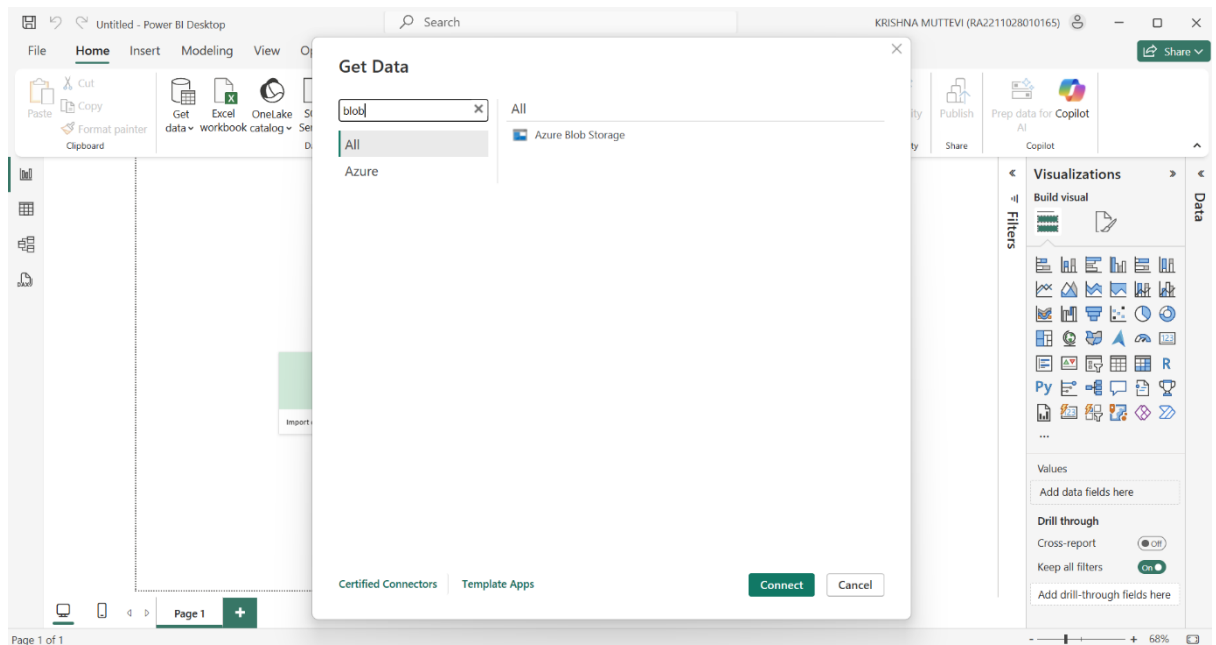


6. Check Output in ADLS2 :



Power Bi Connection:

1. Get Blob Storage link and Use in Get Data of Power Bi:



2. Enter Storage Account Details:

Azure Blob Storage

Account name or URL

OK

Cancel

3. Enter Secure Blob Storage Url + Transform Data:

Navigator

Display Options ▾

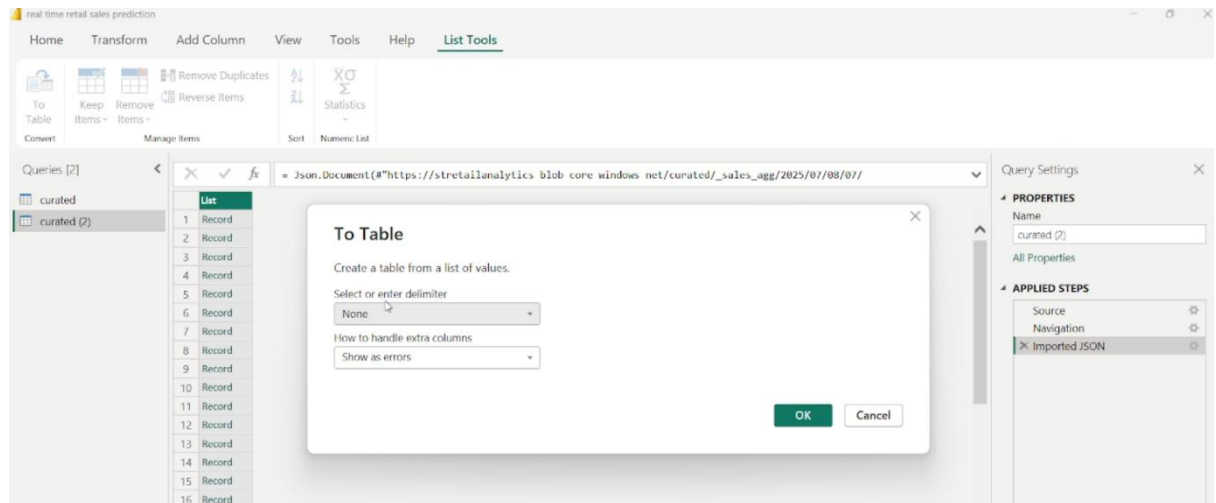
stretailanalytics [1]

☐ curated

curated

Content	Name	Extension
Binary	geakout	
Binary	sales_agg	
Binary	sales_agg/2025	
Binary	sales_agg/2025/07	
Binary	sales_agg/2025/07/08	
Binary	sales_agg/2025/07/08/07	
Binary	sales_agg/2025/07/08/07/0_bc00eb4af70d49848da2148169d5a78b_1.jsc	.json
Binary	sales_agg/2025/07/08/08	
Binary	sales_agg/2025/07/08/08/0_13bd1ab5805a413baaa04c495f983de7_1.jsc	.json

4. Select Json and Convert to Table:



5. Create a Dashboard:

