Niroop777 / Bank_Project

Type / to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main Bank_Project / README.md

Go to file t

Niroop777 Update README.md 8579a29 · now History

Preview Code Blame Raw

Azure Banking Data Platform

A production-grade, real-time data engineering pipeline built using **Azure Event-Driven Architecture**, **Cosmos DB**, and the **Databricks Lakehouse Platform**.

---

1. Project Overview

This solution delivers an end-to-end banking data platform capable of handling **real-time ATM and UPI transactions**, **customer profile updates**, **KYC records**, and **branch performance analytics**.

The platform integrates:

- **Event-driven ingestion** using Event Grid, Azure Functions, and Service Bus
- **Operational data store** using Azure Cosmos DB

- **Data Lakehouse ETL** using Databricks (Bronze → Silver → Gold)
- **SCD-2 dimension management**
- **SQL Data Warehouse serving layer**
- **Power BI dashboards for Customer 360, Fraud Metrics, Branch Performance**
- **CI/CD workflow using GitHub Actions / Azure DevOps**

---

2. High-Level Architecture

```
+------------------+          +-----------------+         +----------------+
|   ADLS Raw       |  Event   |  Azure Function |  Msg    |   Service Bus  |
| (ATM/UPI/KYC)    +--------->+      (A)        +-------->+     Queue      |
+------------------+          +-----------------+         +----------------+
                                           |  Trigger
                                           v
                                    +---------------+
                                    | Azure Function|
                                    |      (B)      |
                                    +-------+-------+
                                            |
                   +------------------------+-----------------------------+
                   |                                                      |
          Valid Data to Cosmos DB                               Invalid → ADLS/quarantine
                   |                                              Metadata → ADLS/metadata
                   v
        +----------------------+
        |   Cosmos DB (ODS)    |
        +----------+-----------+
                   |
                   v
    +----------------------------------+
    |     Databricks Lakehouse        |
    |  Bronze → Silver → Gold Layers  |
    +------------------+----------------+
                       |
                       v
              +-----------------+
              | SQL Data Ware-  |
              | house (Serving) |
              +-------+---------+
                      |
                      v
              +----------------+
              |   Power BI     |
              | Dashboards     |
              +----------------+
```

3. Azure Resources Used

Service	Purpose
ADLS Gen2	Raw, Quarantine, Metadata, Bronze, Silver, Gold zones
Event Grid	Detects new file uploads in ADLS
Azure Function A	Receives EventGrid events and pushes file metadata into Service Bus
Service Bus Queue	Reliable buffer for downstream processing
Azure Function B (Queue Trigger)	Validates payload → writes valid records to Cosmos DB / invalid to ADLS
Cosmos DB	Operational store for ATM, UPI, and Customer data

Service	Purpose
Databricks	Orchestrates Bronze → Silver → Gold ETL
Azure SQL DB / SQL DWH	Serving layer for BI tools
Power BI	Analytics dashboards
CI/CD (GitHub Actions)	Notebook sync + Function App deployment

---

4. ADLS Folder Structure

```
/raw
    /atm
    /upi
    /customers
    /kyc

/quarantine
/metadata

/etl
    /bronze
    /silver
    /gold
```

---

5. Ingestion Workflow

Step 1 — File lands in ADLS (raw)

ATM, UPI, Customer, KYC files are uploaded into `/raw`.

Step 2 — Event Grid triggers Function A

Detects `BlobCreated` events and extracts:

- file_url
- file_name
- source_type (ATM/UPI/CUSTOMER/KYC)

Step 3 — Function A → Event Grid Trigger

Minimal enrichment and routing.

Step 4 — Function B (Queue-trigger)

Validates schema, datatype, required fields.

- **If valid:** → Insert into Cosmos DB → Store metadata in `/metadata`
- **If invalid:** → Store full payload in `/quarantine`

Step 5 — Databricks Ingestion

- Reads **Cosmos DB** for transactions & customer profiles
- Reads **KYC directly from ADLS**

---

6. Databricks ETL (Lakehouse)

6.1 Bronze Layer

- Raw ingestion from Cosmos DB
- KYC raw ingestion from ADLS
- Light parsing
- No transformations
- Stored as Delta tables

6.2 Silver Layer

- Standardized schema
- Data quality cleanup
- Normalized structures
- Reference cleanup
- Joins across sources
- Error-handled and type-safe

6.3 Gold Layer (Final Analytics Layer)

Dimensions

- `dim_customer`
- `dim_account`
- `dim_kyc`

- `dim_date`

SCD-2 Dimensions

- `dim_customer_scd2`
- `dim_account_scd2`

Facts

- `fact_transactions`
- `fact_customer_profile`

Aggregates

- Daily transaction summary
- Customer monthly spending
- Branch performance
- Channel-wise distribution

---

7. Serving Layer (SQL Data Warehouse)

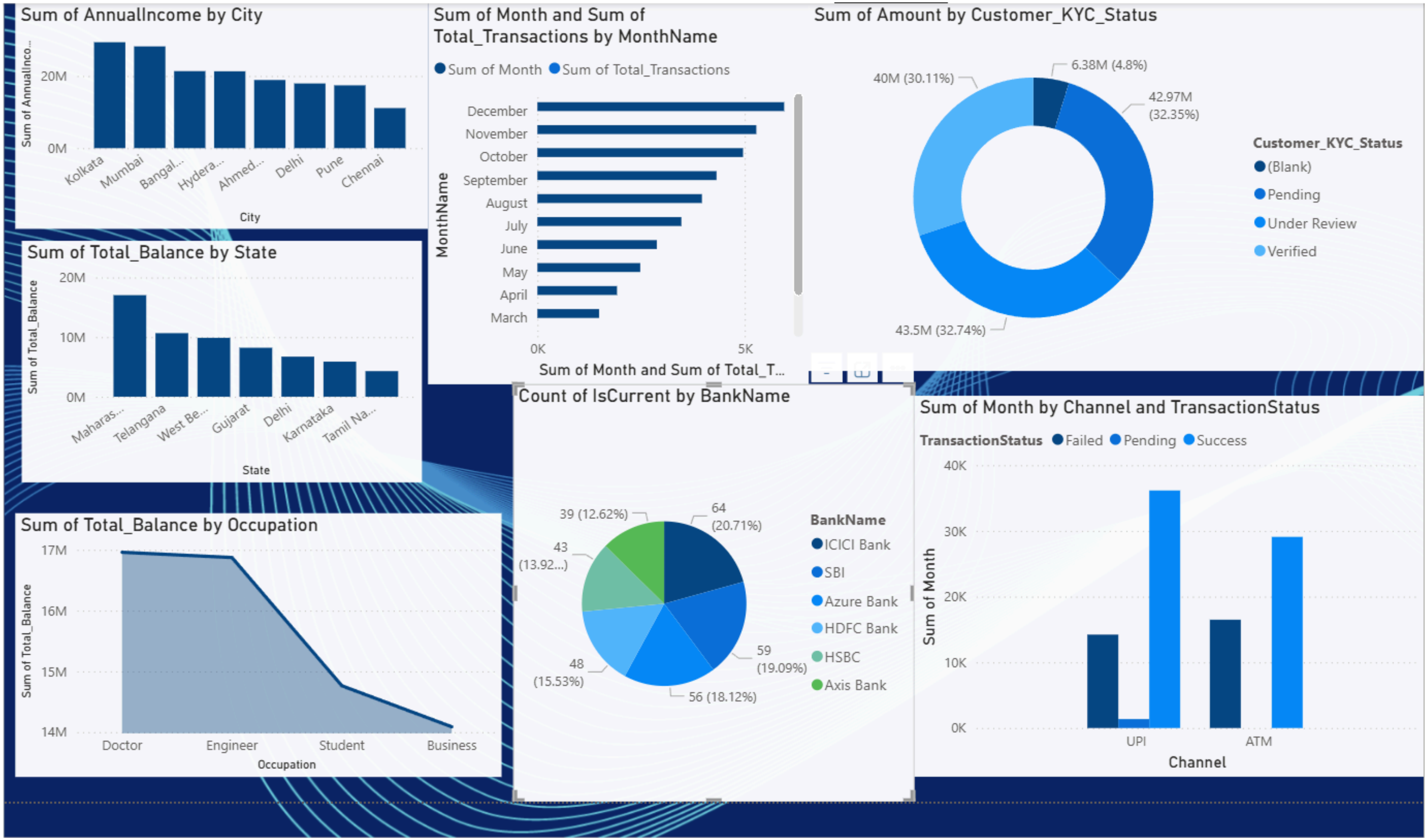
Gold tables and aggregates are written to **Azure SQL DB / SQL Warehouse**, enabling optimized BI reporting.

---

8. Power BI Dashboards

Dashboards supported:

Example placeholders:



9. CI/CD

CI Pipeline

- Linting & quality checks
- Validate Python Function code
- Validate Databricks notebooks
- Package Function App

CD Pipeline

- Deploy Azure Function App
- Sync Databricks notebooks (via Databricks CLI)
- Optionally trigger Databricks job
- Validate Cosmos DB + ADLS connectivity

Pipeline can be implemented in:

- GitHub Actions
- Azure DevOps Pipelines

---

10. Repository Structure

```
/src
    /function_app
    /notebooks
         /bronze
         /silver
         /gold
    /powerbi
    /sql
/docs
    /images
    architecture.md
/readme.md
```

---

11. How to Run the Project

Prerequisites

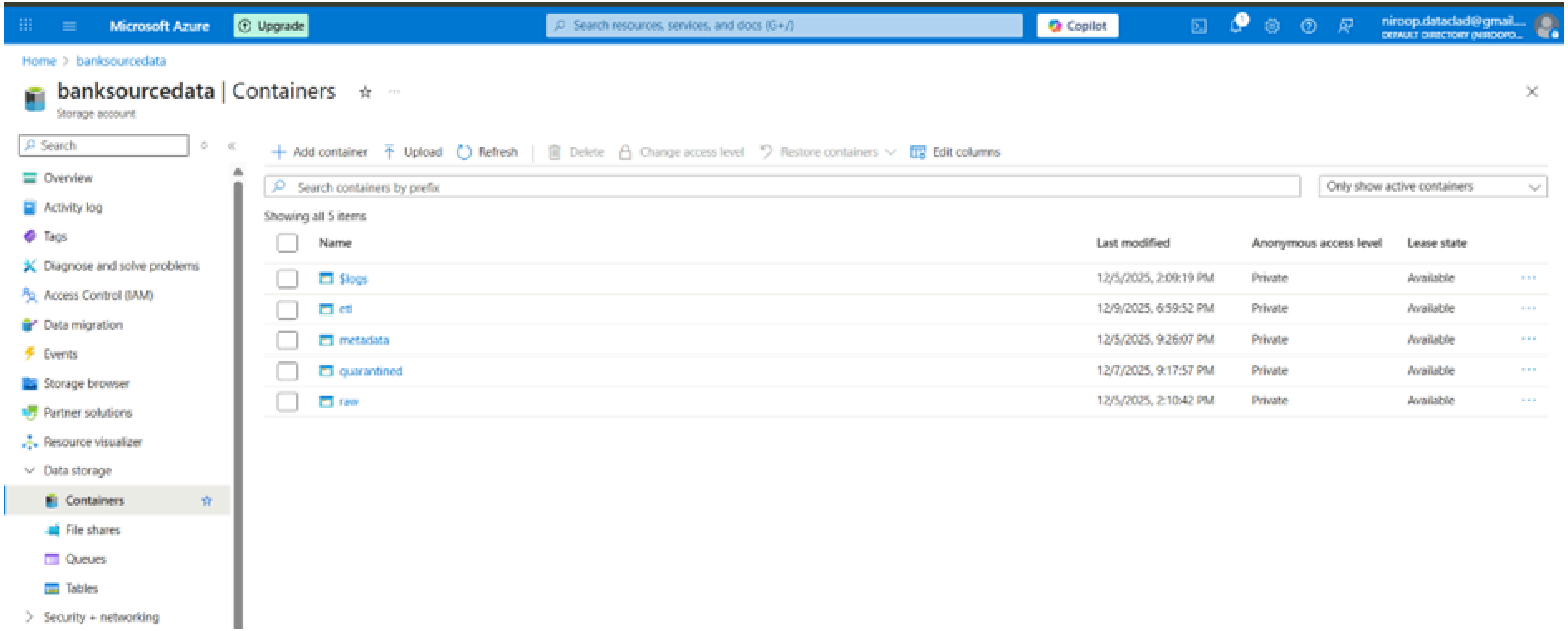
- Azure subscription
- Cosmos DB + ADLS + Function App + Service Bus
- Databricks Workspace
- SQL Database

Execution Steps

1. Upload sample data into `/raw` in ADLS
2. Event Grid → Functions → Service Bus → Cosmos DB automatically ingests
3. Databricks Jobs run Bronze → Silver → Gold pipeline
4. Gold tables are published to SQL DWH
5. Power BI connects to SQL DWH for reporting
6. CI/CD handles notebook + function deployments

---

12. Screenshot Placeholders

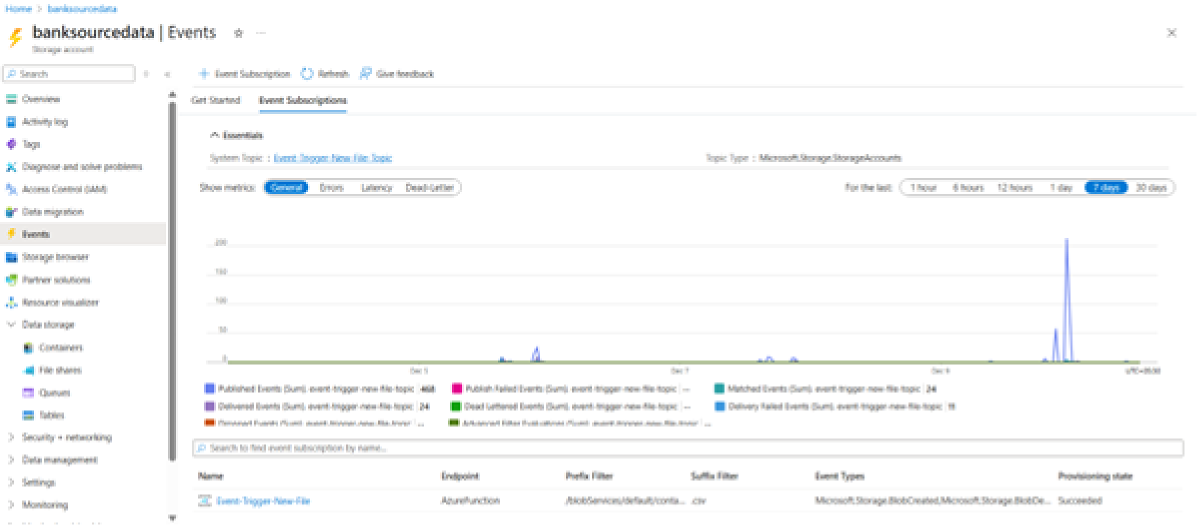

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
niroop.datacIad@gmail....
DEFAULT DIRECTORY (NIROOPD...
Home > banksourcedata

banksourcedata | Containers

Storage account

Showing all 5 items

Name	Last modified	Anonymous access level	Lease state
$logs	12/5/2025, 2:09:19 PM	Private	Available
etl	12/9/2025, 6:59:52 PM	Private	Available
metadata	12/5/2025, 9:26:07 PM	Private	Available
quarantined	12/7/2025, 9:17:57 PM	Private	Available
raw	12/5/2025, 2:10:42 PM	Private	Available



banksourcedata | Events

Storage account

Search
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Partner solutions
Resource visualizer
Data storage
Containers
File shares
Queues
Tables
Security + networking
Data management
Settings
Monitoring
Event Subscription
Refresh
Give feedback
Get Started
Event Subscriptions
Essentials

System Topic : Event-Trigger-New-File-Topic

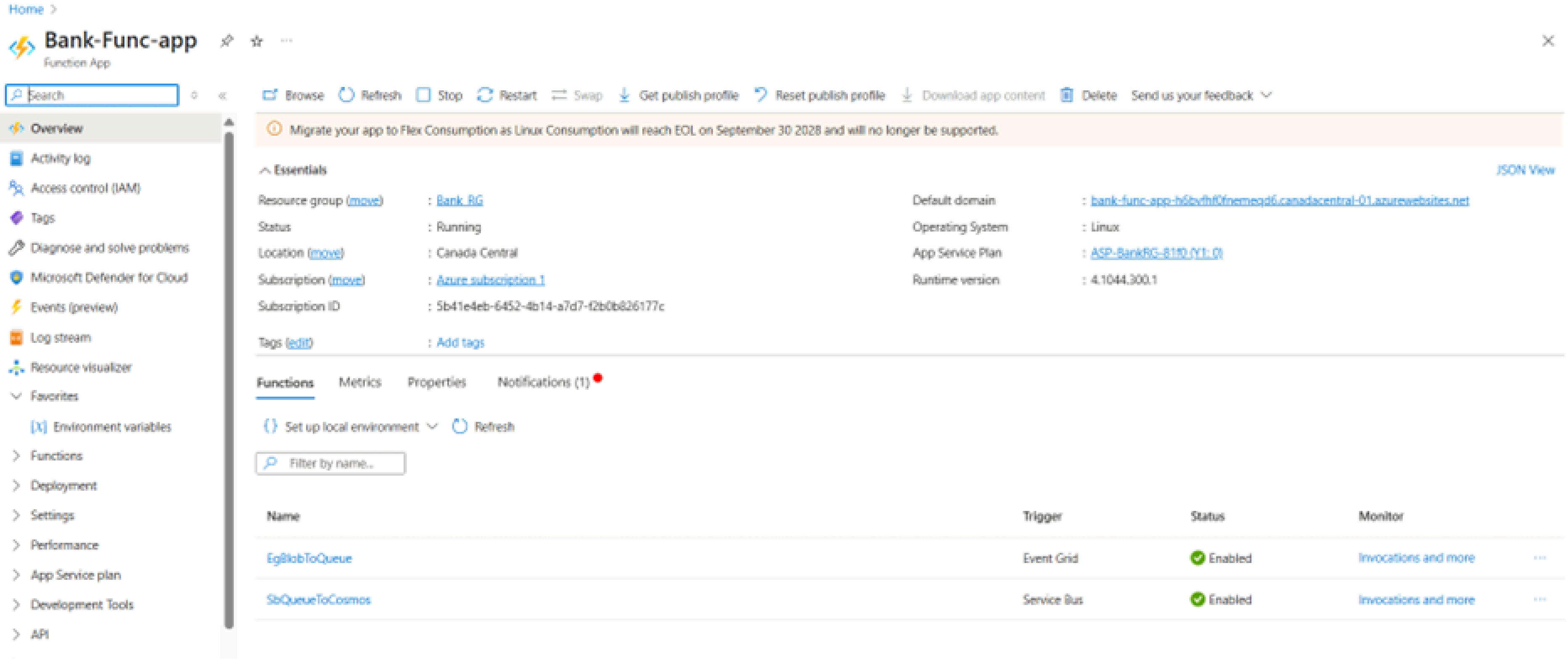
Topic Type : Microsoft.Storage.StorageAccounts

Show metrics: General Errors Latency Dead-Letter

For the last: 1 hour 6 hours 12 hours 1 day 7 days 30 days

200
150
100
50
0
Dec 5
Dec 7
Dec 9
UTC+05:30
Published Events (Sum), event-trigger-new-file-topic 468
Publish Failed Events (Sum), event-trigger-new-file-topic --
Matched Events (Sum), event-trigger-new-file-topic 24
Delivered Events (Sum), event-trigger-new-file-topic 24
Dead Lettered Events (Sum), event-trigger-new-file-topic --
Delivery Failed Events (Sum), event-trigger-new-file-topic 11
Dropped Events (Sum), event-trigger-new-file-topic --
Advanced Filter Evaluations (Sum), event-trigger-new-file-topic --
Search to find event subscription by name...

Name	Endpoint	Prefix Filter	Suffix Filter	Event Types	Provisioning state
Event-Trigger-New-File	AzureFunction	/blobServices/default/conta...	.csv	Microsoft.Storage.BlobCreated,Microsoft.Storage.BlobDe...	Succeeded



Bank-Func-app

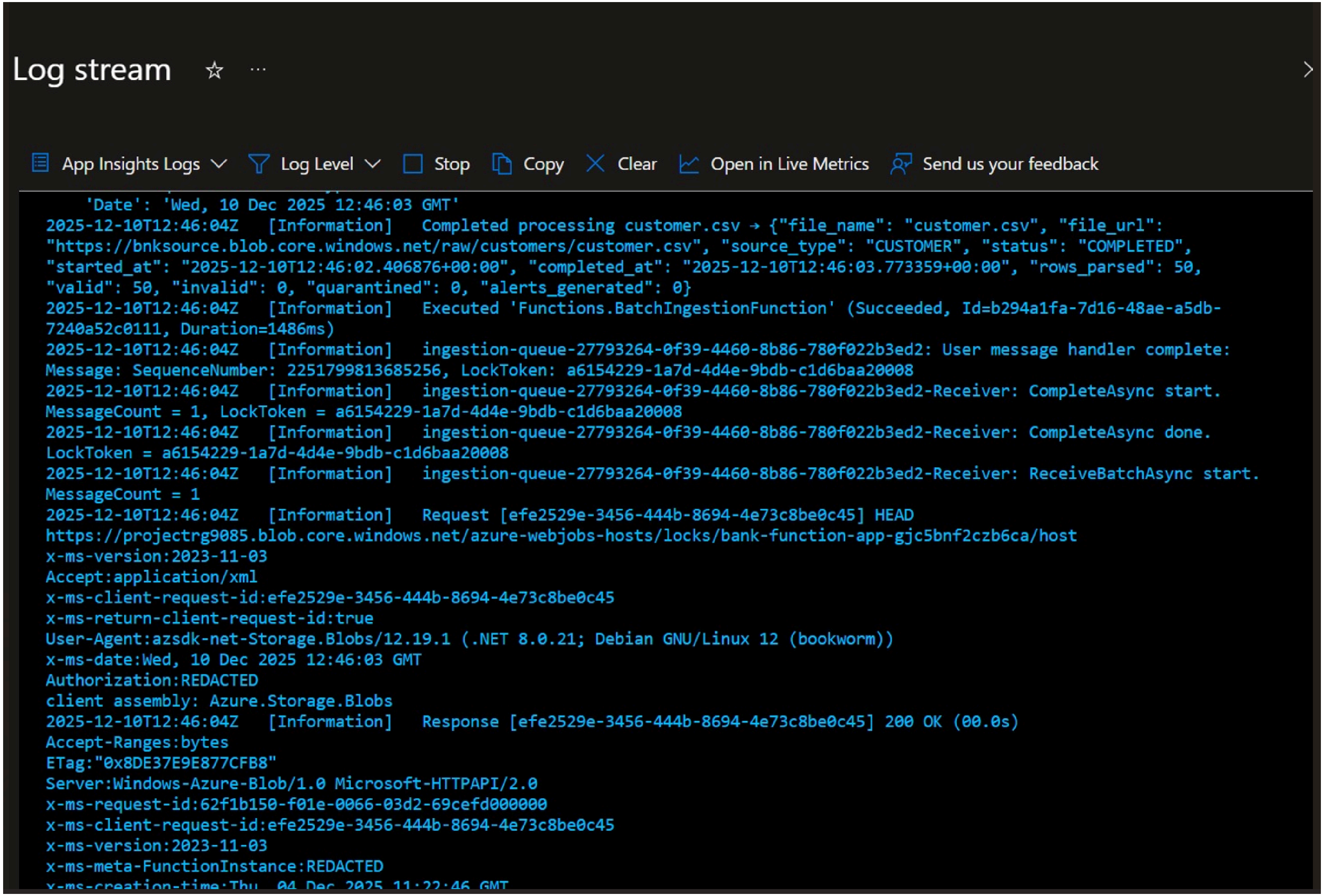
Function App

Migrate your app to Flex Consumption as Linux Consumption will reach EOL on September 30 2028 and will no longer be supported.

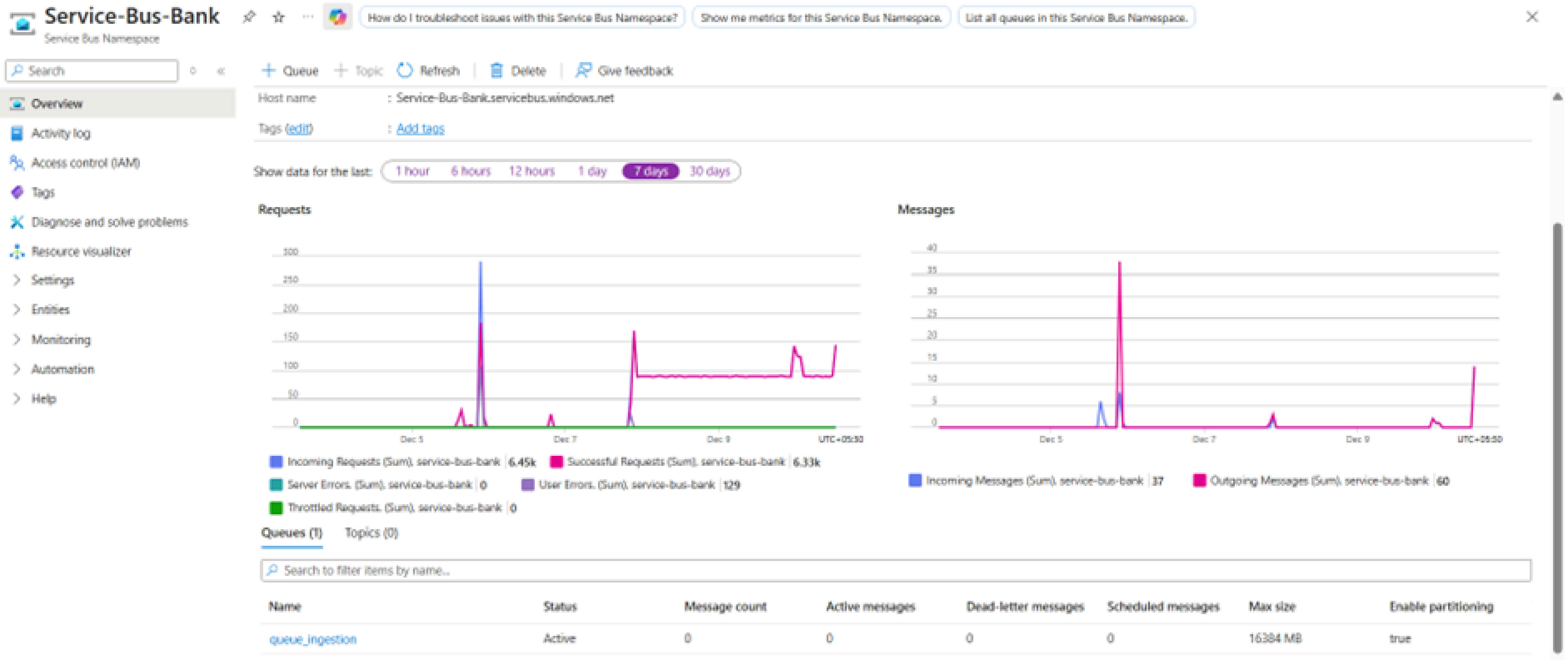
Essentials

Resource group (move)	: Bank_RG	Default domain	: bank-func-app-h6bvfhf0fnemeqd6.canadacentral-01.azurewebsites.net
Status	: Running	Operating System	: Linux
Location (move)	: Canada Central	App Service Plan	: ASP-BankRG-81f0 (Y1: 0)
Subscription (move)	: Azure subscription 1	Runtime version	: 4.1044.300.1
Subscription ID	: 5b41e4eb-6452-4b14-a7d7-f2b0b826177c		
Tags (edit)	: Add tags		

Name	Trigger	Status	Monitor
EgBlobToQueue	Event Grid	✓ Enabled	Invocations and more
SbQueueToCosmos	Service Bus	✓ Enabled	Invocations and more



```
    'Date': 'Wed, 10 Dec 2025 12:46:03 GMT'
2025-12-10T12:46:04Z   [Information]   Completed processing customer.csv → {"file_name": "customer.csv", "file_url":
"https://bnksource.blob.core.windows.net/raw/customers/customer.csv", "source_type": "CUSTOMER", "status": "COMPLETED",
"started_at": "2025-12-10T12:46:02.406876+00:00", "completed_at": "2025-12-10T12:46:03.773359+00:00", "rows_parsed": 50,
"valid": 50, "invalid": 0, "quarantined": 0, "alerts_generated": 0}
2025-12-10T12:46:04Z   [Information]   Executed 'Functions.BatchIngestionFunction' (Succeeded, Id=b294a1fa-7d16-48ae-a5db-
7240a52c0111, Duration=1486ms)
2025-12-10T12:46:04Z   [Information]   ingestion-queue-27793264-0f39-4460-8b86-780f022b3ed2: User message handler complete:
Message: SequenceNumber: 2251799813685256, LockToken: a6154229-1a7d-4d4e-9bdb-c1d6baa20008
2025-12-10T12:46:04Z   [Information]   ingestion-queue-27793264-0f39-4460-8b86-780f022b3ed2-Receiver: CompleteAsync start.
MessageCount = 1, LockToken = a6154229-1a7d-4d4e-9bdb-c1d6baa20008
2025-12-10T12:46:04Z   [Information]   ingestion-queue-27793264-0f39-4460-8b86-780f022b3ed2-Receiver: CompleteAsync done.
LockToken = a6154229-1a7d-4d4e-9bdb-c1d6baa20008
2025-12-10T12:46:04Z   [Information]   ingestion-queue-27793264-0f39-4460-8b86-780f022b3ed2-Receiver: ReceiveBatchAsync start.
MessageCount = 1
2025-12-10T12:46:04Z   [Information]   Request [efe2529e-3456-444b-8694-4e73c8be0c45] HEAD
https://projectrg9085.blob.core.windows.net/azure-webjobs-hosts/locks/bank-function-app-gjc5bnf2czb6ca/host
x-ms-version:2023-11-03
Accept:application/xml
x-ms-client-request-id:efe2529e-3456-444b-8694-4e73c8be0c45
x-ms-return-client-request-id:true
User-Agent:azsdk-net-Storage.Blobs/12.19.1 (.NET 8.0.21; Debian GNU/Linux 12 (bookworm))
x-ms-date:Wed, 10 Dec 2025 12:46:03 GMT
Authorization:REDACTED
client assembly: Azure.Storage.Blobs
2025-12-10T12:46:04Z   [Information]   Response [efe2529e-3456-444b-8694-4e73c8be0c45] 200 OK (00.0s)
Accept-Ranges:bytes
ETag:"0x8DE37E9E877CFB8"
Server:Windows-Azure-Blob/1.0 Microsoft-HTTPAPI/2.0
x-ms-request-id:62f1b150-f01e-0066-03d2-69cefd000000
x-ms-client-request-id:efe2529e-3456-444b-8694-4e73c8be0c45
x-ms-version:2023-11-03
x-ms-meta-FunctionInstance:REDACTED
x-ms-creation-time:Thu, 04 Dec 2025 11:22:46 GMT
```



Host name : Service-Bus-Bank.servicebus.windows.net

Tags (edit) : Add tags

Show data for the last: 1 hour 6 hours 12 hours 1 day [✓ 7 days] 30 days

Incoming Requests (Sum), service-bus-bank | 6.45k

Successful Requests (Sum), service-bus-bank | 6.33k

Server Errors. (Sum), service-bus-bank | 0

User Errors. (Sum), service-bus-bank | 129

Throttled Requests. (Sum), service-bus-bank | 0

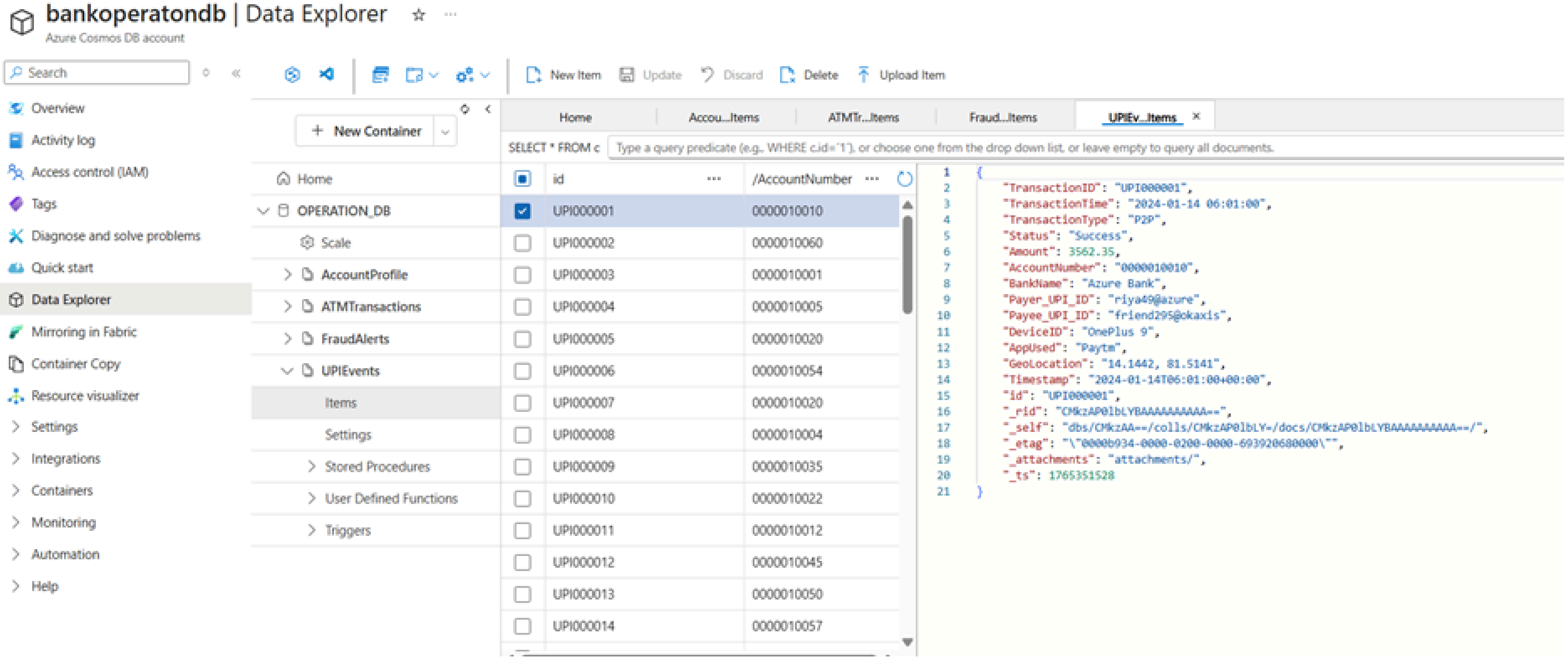
Incoming Messages (Sum), service-bus-bank | 37

Outgoing Messages (Sum), service-bus-bank | 60

Queues (1) Topics (0)

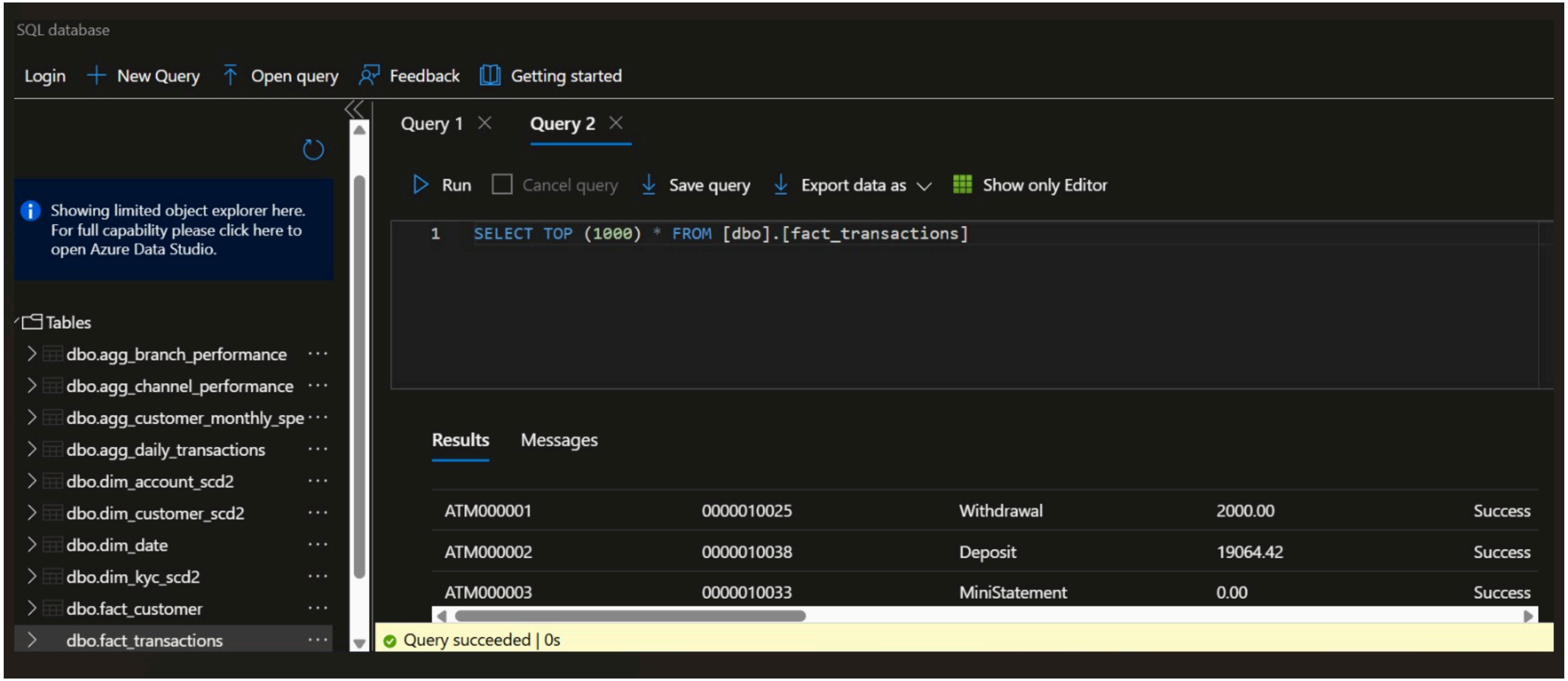
Search to filter items by name...

Name	Status	Message count	Active messages	Dead-letter messages	Scheduled messages	Max size	Enable partitioning
queue_ingestion	Active	0	0	0	0	16384 MB	true



☑	id	/AccountNumber
☑	UPI000001	0000010010
□	UPI000002	0000010060
□	UPI000003	0000010001
□	UPI000004	0000010005
□	UPI000005	0000010020
□	UPI000006	0000010054
□	UPI000007	0000010020
□	UPI000008	0000010004
□	UPI000009	0000010035
□	UPI000010	0000010022
□	UPI000011	0000010012
□	UPI000012	0000010045
□	UPI000013	0000010050
□	UPI000014	0000010057

```json
{
    "TransactionID": "UPI000001",
    "TransactionTime": "2024-01-14 06:01:00",
    "TransactionType": "P2P",
    "Status": "Success",
    "Amount": 3562.35,
    "AccountNumber": "0000010010",
    "BankName": "Azure Bank",
    "Payer_UPI_ID": "riya49@azure",
    "Payee_UPI_ID": "friend295@okaxis",
    "DeviceID": "OnePlus 9",
    "AppUsed": "Paytm",
    "GeoLocation": "14.1442, 81.5141",
    "Timestamp": "2024-01-14T06:01:00+00:00",
    "id": "UPI000001",
    "_rid": "CMkzAP0lbLYBAAAAAAAAAA==",
    "_self": "dbs/CMkzAA==/colls/CMkzAP0lbLY=/docs/CMkzAP0lbLYBAAAAAAAAAA==/",
    "_etag": "\"0000b934-0000-0200-0000-693920680000\"",
    "_attachments": "attachments/",
    "_ts": 1765351528
}
```



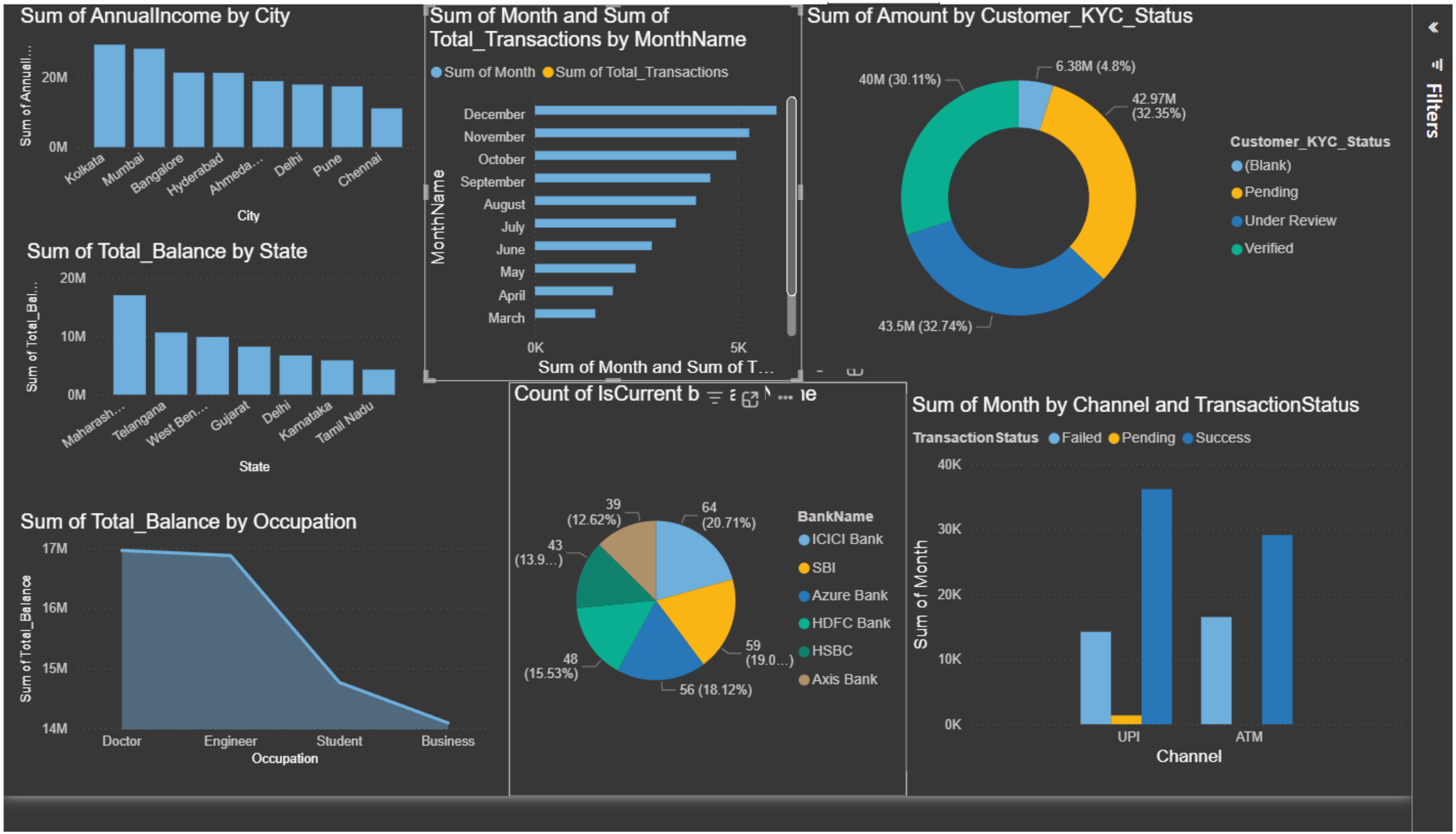
Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables
- dbo.agg_branch_performance
- dbo.agg_channel_performance
- dbo.agg_customer_monthly_spe
- dbo.agg_daily_transactions
- dbo.dim_account_scd2
- dbo.dim_customer_scd2
- dbo.dim_date
- dbo.dim_kyc_scd2
- dbo.fact_customer
- dbo.fact_transactions

```sql
SELECT TOP (1000) * FROM [dbo].[fact_transactions]
```

ATM000001	0000010025	Withdrawal	2000.00	Success
ATM000002	0000010038	Deposit	19064.42	Success
ATM000003	0000010033	MiniStatement	0.00	Success

Query succeeded | 0s


Sum of AnnualIncome by City
Sum of Annuall...
20M
0M
Kolkata
Mumbai
Bangalore
Hyderabad
Ahmeda...
Delhi
Pune
Chennai
City
Sum of Month and Sum of Total_Transactions by MonthName
● Sum of Month ● Sum of Total_Transactions
MonthName
December
November
October
September
August
July
June
May
April
March
0K
5K
Sum of Month and Sum of T...
Sum of Amount by Customer_KYC_Status
40M (30.11%)
6.38M (4.8%)
42.97M (32.35%)
43.5M (32.74%)
Customer_KYC_Status
● (Blank)
● Pending
● Under Review
● Verified
Filters
Sum of Total_Balance by State
Sum of Total_Bal...
20M
10M
0M
Maharash...
Telangana
West Ben...
Gujarat
Delhi
Karnataka
Tamil Nadu
State
Count of IsCurrent b ... ie
39 (12.62%)
64 (20.71%)
43 (13.9...)
59 (19.0...)
48 (15.53%)
56 (18.12%)
BankName
● ICICI Bank
● SBI
● Azure Bank
● HDFC Bank
● HSBC
● Axis Bank
Sum of Month by Channel and TransactionStatus
TransactionStatus ● Failed ● Pending ● Success
Sum of Month
40K
30K
20K
10K
0K
UPI
ATM
Channel
Sum of Total_Balance by Occupation
Sum of Total_Balance
17M
16M
15M
14M
Doctor
Engineer
Student
Business
Occupation