Synopsis: Azure Banking Data Platform – End-to-End Project (Expanded Edition)

1. Introduction

This document presents a comprehensive overview of a modern, cloud-native **Banking Data Platform built on Microsoft Azure**, designed to process financial transactions in real time, store data across optimized analytical layers, and deliver business intelligence dashboards. The platform is structured to be beginner-friendly yet robust enough to reflect industry standards used in digital banking ecosystems.

The system simulates real-world banking events such as ATM withdrawals, UPI payments, and customer login activities. It demonstrates how these events move across Azure Event Grid, Azure Functions, Cosmos DB, Databricks Lakehouse layers, and ultimately into Power BI for analytics. This expanded synopsis provides enough depth for academic submission, project demonstrations, and professional portfolios.

---

2. Problem Statement

Banks produce large volumes of structured and semi-structured data every second. Handling such data presents several challenges:

- Multiple disparate sources (ATM, UPI, mobile banking)
- Lack of real-time ingestion in legacy systems
- Manual, batch-based ETL jobs causing delays
- Difficulty in creating unified customer views
- Scalability challenges during peak transaction loads
- Complexities in fraud detection and regulatory reporting

An integrated, real-time data platform is required to ingest, validate, store, enrich, model, and analyze banking transactions at scale.

3. Objectives

The project aims to build an end-to-end, learning-oriented data engineering solution capable of:

1. Real-time ingestion of banking events using Azure serverless components.
2. Storing operational data in Cosmos DB with high throughput.
3. Implementing a multi-layer Delta Lake architecture—Bronze, Silver, Gold.
4. Performing data quality improvements and transformations using Databricks.
5. Creating a business-ready dimensional data model.
6. Enabling financial analytics, customer insights, and fraud monitoring through Power BI.
7. Automating deployments using GitHub Actions.
8. Ensuring extensibility, scalability, and ease of understanding for beginners.

---

4. High-Level Architecture Overview

The platform follows a modular, layered architecture:

1. **Event Sources** – ATM systems, UPI processors, login systems.
2. **Event Grid** – Listens for new events and triggers downstream logic.
3. **Azure Functions** – Performs validation and routing.
4. **Cosmos DB** – Stores operational event data.
5. **Databricks Lakehouse**:
    - Bronze → Raw, unmodified data
    - Silver → Cleaned, standardized data
    - Gold → Business model (Star Schema)

6. **Azure SQL (optional)** – Secondary reporting layer.
7. **Power BI** – Dashboards and insights.
8. **CI/CD Pipelines** – GitHub Actions deploying Functions and notebooks.

This layered approach ensures auditability, performance, data governance, and analytical readiness.

---

5. Detailed System Flow (End-to-End Logic)

This section describes each operational stage in detail.

5.1 Event Creation Layer (Simulated Banking Systems)

The system generates real-time financial events including:

- ATM cash withdrawals, deposits, mini-statements
- UPI transactions (P2P, P2M, QR code, refunds)
- Login and device identity events from mobile banking

Each event contains:

- Customer ID
- Timestamp
- Transaction type
- Amount
- Location (ATM ID or mobile geolocation)
- Device information

These events mimic real-world core banking systems.

5.2 Event Grid – Event Routing and Notification

Event Grid acts as the event broker. Its responsibilities include:

- Monitoring ADLS raw container or upstream systems
- Triggering Azure Functions instantly on new event arrival
- Passing metadata such as file path, event size, creation time

It provides low latency, reliability, and scalability up to millions of events.

5.3 Azure Functions – Real-Time Ingestion Engine

Functions perform critical preprocessing operations:

- Read event payload from Event Grid
- Parse JSON or CSV structures
- Validate schema and ensure required fields are present
- Convert timestamps to uniform format
- Identify type (ATM, UPI, Login)
- Enrich data with ingestion metadata
- Handle malformed records
- Insert cleaned data into Cosmos DB collections

Functions create the first reliable, structured checkpoint for all incoming events.

5.4 Cosmos DB – Operational Data Store (ODS)

Cosmos DB acts as the real-time landing store. It maintains separate collections:

- ATMTransactions
- UPIEvents
- AccountProfiles
- FraudAlerts

Key benefits:

- Low-latency writes (single-digit ms)
- Global distribution support
- Flexible schema for evolving banking events
- Partitioning ensures scalability under heavy load

This layer supports real-time analytics or APIs if needed.

5.5 Databricks Bronze Layer – Raw Delta Store

The Bronze layer stores the **exact** raw event records from Cosmos DB. It:

- Appends ingestion metadata (load timestamp, source system)
- Preserves all original values for auditing
- Enables replay and recovery scenarios

Bronze is the immutable foundation of the Lakehouse.

5.6 Databricks Silver Layer – Cleansed & Conformed Data

Silver performs heavy data refinement:

- Deduplication rules remove repeated events
- Data type corrections
- Null and missing field handling
- Standardized datetime formats
- Normalized structure across ATM + UPI datasets
- Joins with reference data such as branch codes, account profiles
- Creation of derived attributes (DateKey, ChannelType, GeoRegion)

The Silver layer is analytics-ready and significantly cleaner.

5.7 Databricks Gold Layer – Dimensional Data Model

Gold represents the curated business model using a **Star Schema**.

Dimension Tables:

- **DimCustomer**: Personal data, demographics, risk score, KYC validity, SCD2 history
- **DimAccount**: Account type, branch, currency, status, open date
- **DimDate**: Calendar table with daily, monthly, fiscal attributes

Fact Table:

- **FactTransactions**: Unified transactional dataset combining ATM and UPI with surrogate keys

Benefits:

- Faster analytical queries

- Business-friendly structure
- Optimized for BI tools

5.8 Azure SQL Layer (Optional)

Gold tables may be further loaded into Azure SQL Database when BI teams require relational structures or when integration with legacy reporting systems is needed.

5.9 Power BI Analytics Layer

Power BI connects to Gold Delta tables or Azure SQL to create interactive dashboards such as:

- Customer 360 overview
- Transaction patterns by channel
- ATM vs UPI comparison trends
- Fraud detection insights
- Daily, weekly, and monthly KPIs

This layer enables business users, analysts, and auditors to explore the data.

5.10 CI/CD Pipeline – GitHub Actions

The pipeline automates:

- Azure Function deployment via ZIP publish
- Uploading Databricks notebooks using API
- Maintaining version control and consistent environments

This ensures seamless updates and minimizes manual effort.

6. Data Modeling Strategy

The solution adopts a Lakehouse approach:

- **Bronze**: Raw, append-only
- **Silver**: Clean, structured, enriched
- **Gold**: Business dimensional model

Dimensional modeling improves performance, usability, and reporting consistency.

---

7. Use Cases Enabled

- Real-time fraud alerting
- Customer behavioral analysis
- ATM and UPI performance dashboards
- Financial reporting and ledger summaries
- Compliance and regulatory audits (RBI guidelines)
- Customer segmentation and marketing analytics

---

8. Technologies Used

- Azure Event Grid
- Azure Functions (Python)
- Cosmos DB
- ADLS Gen2

- Databricks (PySpark, Delta Lake)
- Azure SQL
- Power BI
- GitHub Actions

---

9. Repository Structure

```
project-root/
├── functions/
├── databricks/
│   └── notebooks/
├── powerbi/
├── adls/
└── .github/workflows/
```

---

10. Implementation Steps

1. Deploy Azure resources manually.
2. Push code to GitHub main branch.
3. Allow CI/CD to deploy Functions and notebooks.
4. In Databricks, run Bronze → Silver → Gold pipelines.
5. Connect Power BI to Gold tables.
6. Publish dashboards.

---

11. Benefits of the Platform

- Real-time event processing
- Scalable and cloud-native design
- Clean separation of data layers
- Business-ready analytics
- Ideal for beginners learning Azure data engineering
- Extensible for production-grade enhancements

---

12. Conclusion

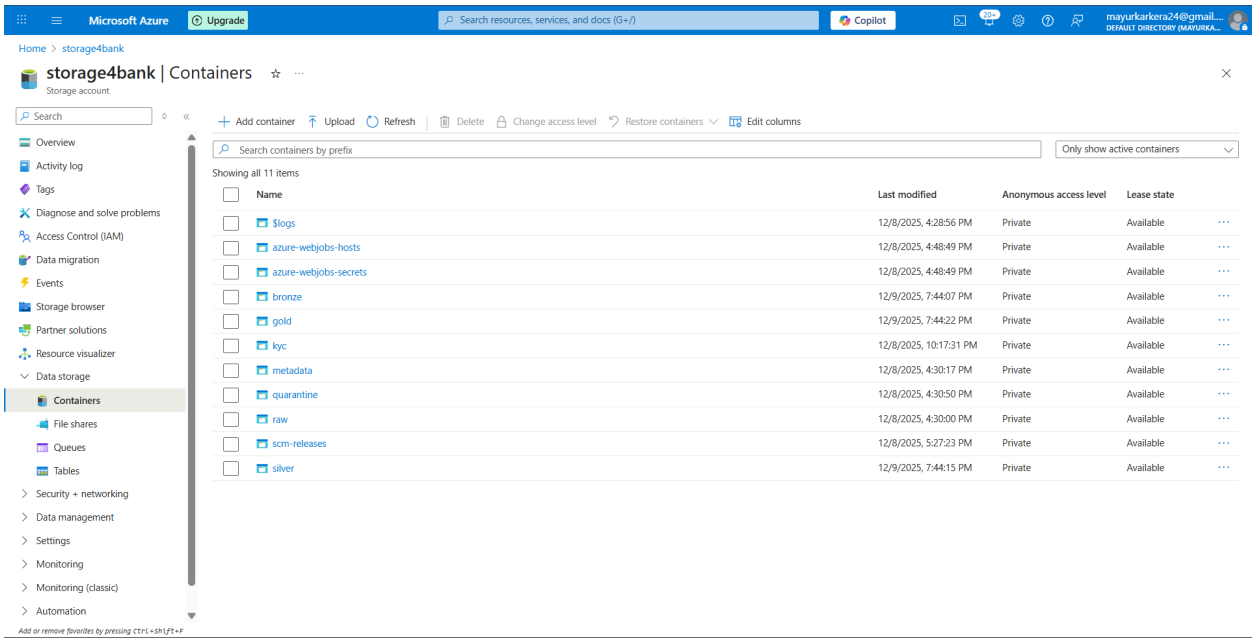
This expanded synopsis outlines the architecture, workflow, transformations, modeling, and analytics involved in building a complete Azure Banking Data Platform. The solution represents an end-to-end, industry-aligned data engineering pipeline suitable for academic projects, organizational prototypes, and hands-on learning.

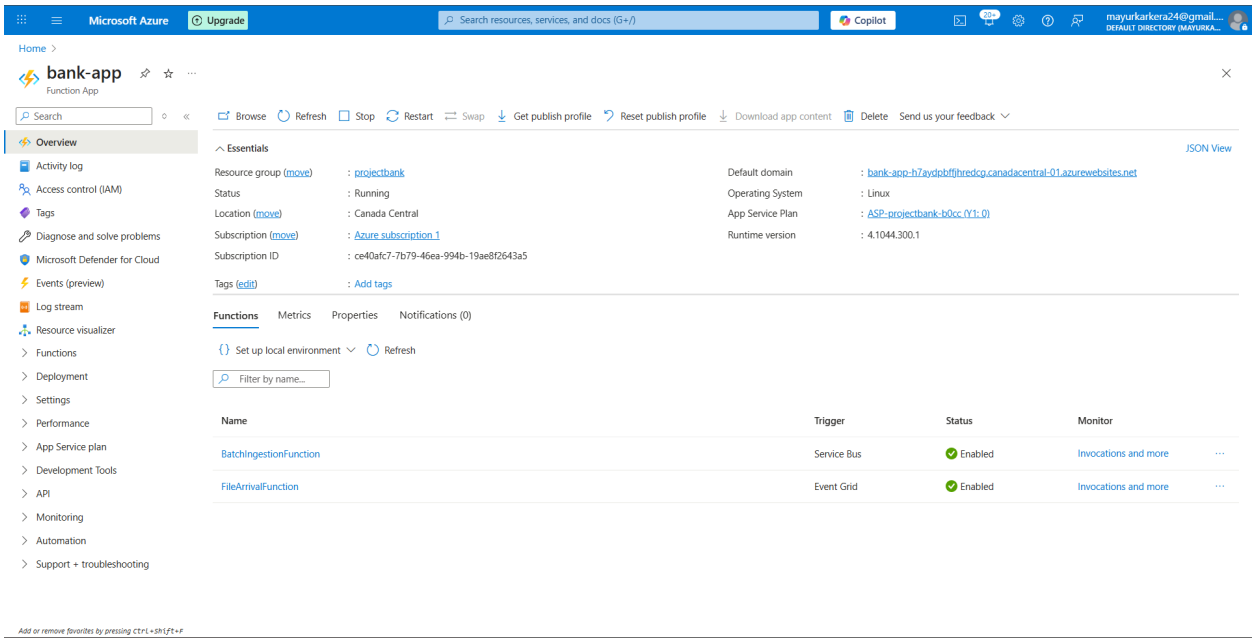
---

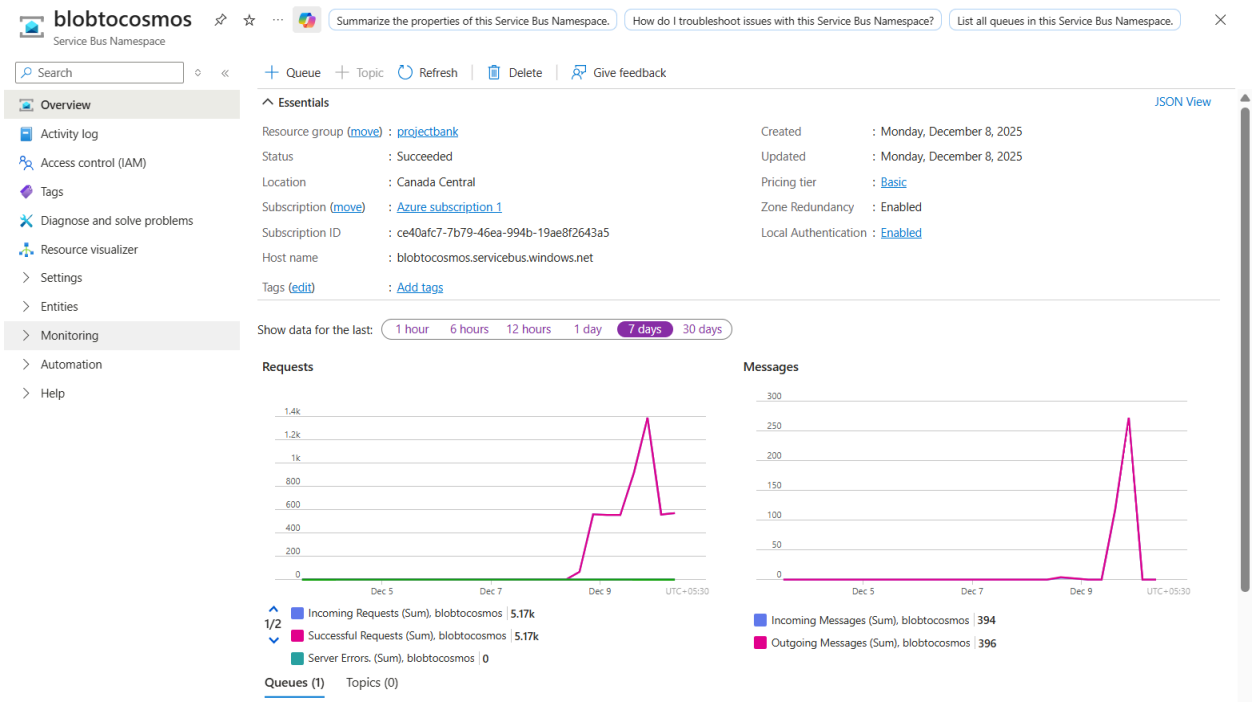
13. Keywords

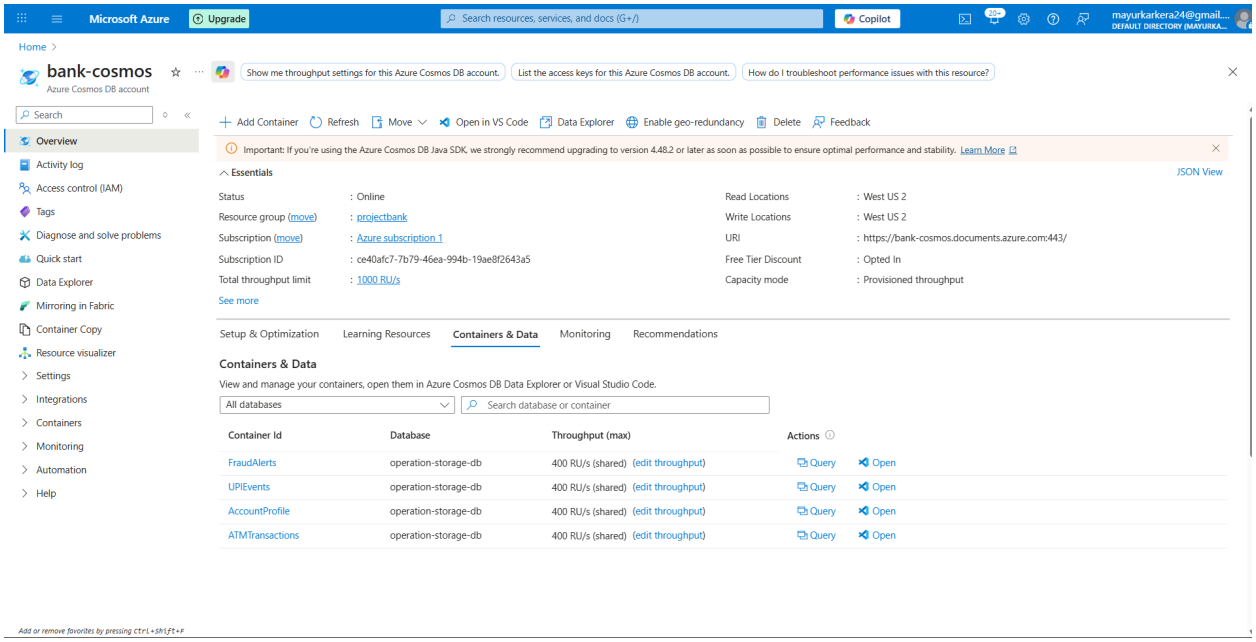
Azure Databricks, Cosmos DB, Event Grid, Azure Functions, Delta Lake, Star Schema, Banking Analytics, ETL Pipeline, Data Platform

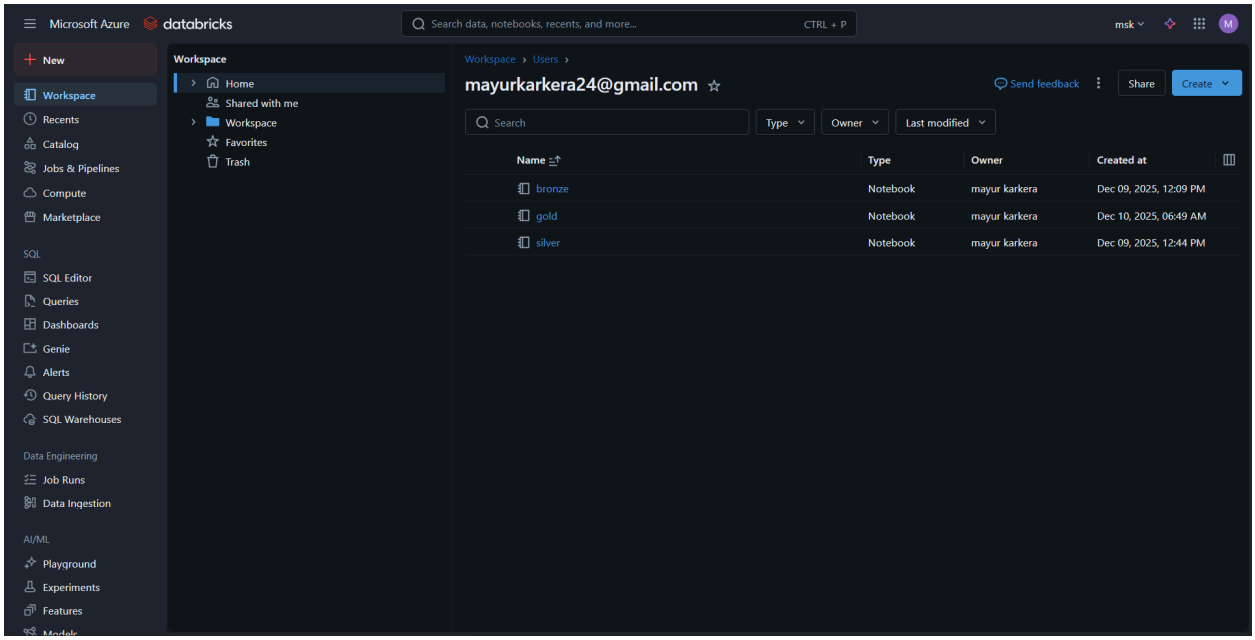
Please find all the screenshot of the resource below

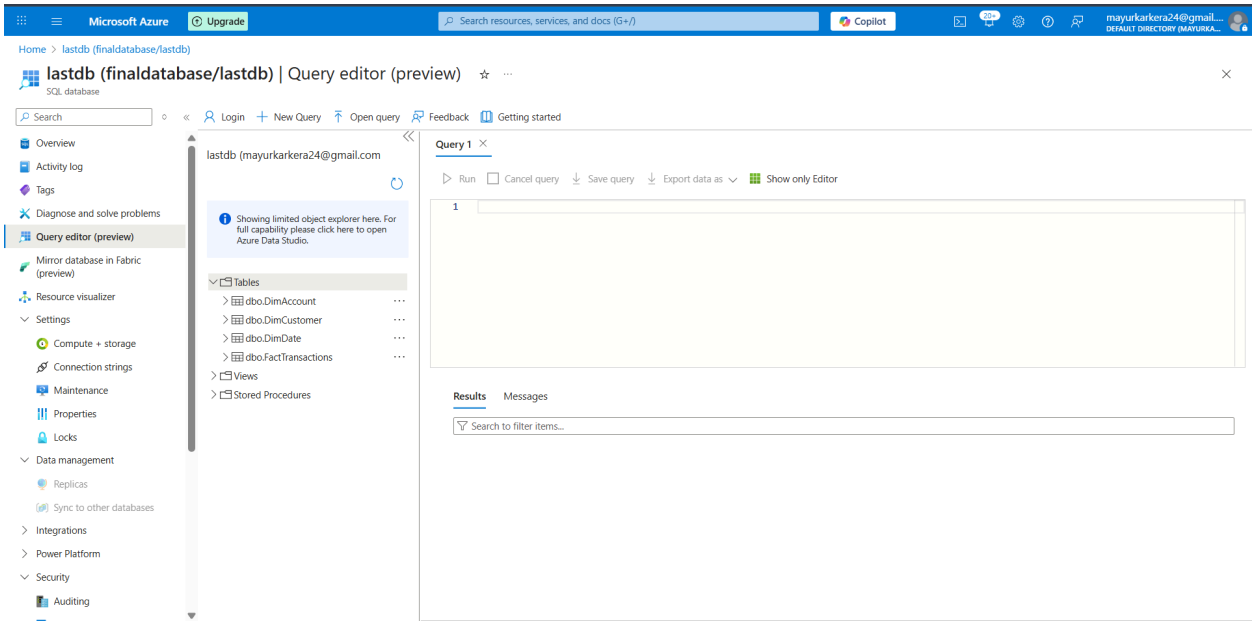

Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
mayurkarkera24@gmail....
DEFAULT DIRECTORY (MAYURKA...
Home > storage4bank
storage4bank | Containers
Storage account
Search
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage browser
Partner solutions
Resource visualizer
Data storage
Containers
File shares
Queues
Tables
Security + networking
Data management
Settings
Monitoring
Monitoring (classic)
Automation
Add or remove favorites by pressing Ctrl+Shift+F
Add container
Upload
Refresh
Delete
Change access level
Restore containers
Edit columns
Search containers by prefix
Only show active containers
Showing all 11 items
Name
Last modified
Anonymous access level
Lease state
$logs 12/8/2025, 4:28:56 PM Private Available
azure-webjobs-hosts 12/8/2025, 4:48:49 PM Private Available
azure-webjobs-secrets 12/8/2025, 4:48:49 PM Private Available
bronze 12/9/2025, 7:44:07 PM Private Available
gold 12/9/2025, 7:44:22 PM Private Available
kyc 12/8/2025, 10:17:31 PM Private Available
metadata 12/8/2025, 4:30:17 PM Private Available
quarantine 12/8/2025, 4:30:50 PM Private Available
raw 12/8/2025, 4:30:00 PM Private Available
scm-releases 12/8/2025, 5:27:23 PM Private Available
silver 12/9/2025, 7:44:15 PM Private Available


Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
mayurkarkera24@gmail....
DEFAULT DIRECTORY (MAYURKA...
Home
bank-app
Function App
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Microsoft Defender for Cloud
Events (preview)
Log stream
Resource visualizer
Functions
Deployment
Settings
Performance
App Service plan
Development Tools
API
Monitoring
Automation
Support + troubleshooting
Add or remove favorites by pressing Ctrl+Shift+F
Browse
Refresh
Stop
Restart
Swap
Get publish profile
Reset publish profile
Download app content
Delete
Send us your feedback
Essentials
JSON View
Resource group (move) : projectbank
Status : Running
Location (move) : Canada Central
Subscription (move) : Azure subscription 1
Subscription ID : ce40afc7-7b79-46ea-994b-19ae8f2643a5
Tags (edit) : Add tags
Default domain : bank-app-h7aydpbffjhredcg.canadacentral-01.azurewebsites.net
Operating System : Linux
App Service Plan : ASP-projectbank-b0cc (Y1: 0)
Runtime version : 4.1044.300.1
Functions
Metrics
Properties
Notifications (0)
Set up local environment
Refresh
Filter by name...
Name
Trigger
Status
Monitor
BatchIngestionFunction
Service Bus
Enabled
Invocations and more
FileArrivalFunction
Event Grid
Enabled
Invocations and more


blobtocosmos
Service Bus Namespace
Summarize the properties of this Service Bus Namespace.
How do I troubleshoot issues with this Service Bus Namespace?
List all queues in this Service Bus Namespace.
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Resource visualizer
Settings
Entities
Monitoring
Automation
Help
Queue
Topic
Refresh
Delete
Give feedback
Essentials
JSON View
Resource group (move) : projectbank
Status : Succeeded
Location : Canada Central
Subscription (move) : Azure subscription 1
Subscription ID : ce40afc7-7b79-46ea-994b-19ae8f2643a5
Host name : blobtocosmos.servicebus.windows.net
Tags (edit) : Add tags
Created : Monday, December 8, 2025
Updated : Monday, December 8, 2025
Pricing tier : Basic
Zone Redundancy : Enabled
Local Authentication : Enabled
Show data for the last:
1 hour
6 hours
12 hours
1 day
7 days
30 days
Requests
1.4k
1.2k
1k
800
600
400
200
0
Dec 5
Dec 7
Dec 9
UTC+05:30
1/2
Incoming Requests (Sum), blobtocosmos 5.17k
Successful Requests (Sum), blobtocosmos 5.17k
Server Errors. (Sum), blobtocosmos 0
Messages
300
250
200
150
100
50
0
Dec 5
Dec 7
Dec 9
UTC+05:30
Incoming Messages (Sum), blobtocosmos 394
Outgoing Messages (Sum), blobtocosmos 396
Queues (1)
Topics (0)


Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
mayurkarkera24@gmail....
DEFAULT DIRECTORY (MAYURKA...
Home
bank-cosmos
Azure Cosmos DB account
Show me throughput settings for this Azure Cosmos DB account.
List the access keys for this Azure Cosmos DB account.
How do I troubleshoot performance issues with this resource?
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Quick start
Data Explorer
Mirroring in Fabric
Container Copy
Resource visualizer
Settings
Integrations
Containers
Monitoring
Automation
Help
Add Container
Refresh
Move
Open in VS Code
Data Explorer
Enable geo-redundancy
Delete
Feedback
Important: If you're using the Azure Cosmos DB Java SDK, we strongly recommend upgrading to version 4.48.2 or later as soon as possible to ensure optimal performance and stability. Learn More
Essentials
JSON View
Status : Online
Resource group (move) : projectbank
Subscription (move) : Azure subscription 1
Subscription ID : ce40afc7-7b79-46ea-994b-19ae8f2643a5
Total throughput limit : 1000 RU/s
Read Locations : West US 2
Write Locations : West US 2
URI : https://bank-cosmos.documents.azure.com:443/
Free Tier Discount : Opted In
Capacity mode : Provisioned throughput
See more
Setup & Optimization
Learning Resources
Containers & Data
Monitoring
Recommendations
Containers & Data
View and manage your containers, open them in Azure Cosmos DB Data Explorer or Visual Studio Code.
All databases
Search database or container
Container Id
Database
Throughput (max)
Actions
FraudAlerts
operation-storage-db
400 RU/s (shared) (edit throughput)
Query
Open
UPIEvents
operation-storage-db
400 RU/s (shared) (edit throughput)
Query
Open
AccountProfile
operation-storage-db
400 RU/s (shared) (edit throughput)
Query
Open
ATMTransactions
operation-storage-db
400 RU/s (shared) (edit throughput)
Query
Open
Add or remove favorites by pressing Ctrl+Shift+F


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
msk
New
Workspace
Recents
Catalog
Jobs & Pipelines
Compute
Marketplace
SQL
SQL Editor
Queries
Dashboards
Genie
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
AI/ML
Playground
Experiments
Features
Workspace
Home
Shared with me
Workspace
Favorites
Trash
Workspace › Users ›
mayurkarkera24@gmail.com
Send feedback
Share
Create
Search
Type
Owner
Last modified
Name
Type
Owner
Created at
bronze
Notebook
mayur karkera
Dec 09, 2025, 12:09 PM
gold
Notebook
mayur karkera
Dec 10, 2025, 06:49 AM
silver
Notebook
mayur karkera
Dec 09, 2025, 12:44 PM


Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
mayurkarkera24@gmail....
DEFAULT DIRECTORY (MAYURKA...
Home › lastdb (finaldatabase/lastdb)
lastdb (finaldatabase/lastdb) | Query editor (preview)
SQL database
Search
Overview
Activity log
Tags
Diagnose and solve problems
Query editor (preview)
Mirror database in Fabric (preview)
Resource visualizer
Settings
Compute + storage
Connection strings
Maintenance
Properties
Locks
Data management
Replicas
Sync to other databases
Integrations
Power Platform
Security
Auditing
Login
New Query
Open query
Feedback
Getting started
lastdb (mayurkarkera24@gmail.com
Showing limited object explorer here. For full capability please click here to open Azure Data Studio.
Tables
dbo.DimAccount
dbo.DimCustomer
dbo.DimDate
dbo.FactTransactions
Views
Stored Procedures
Query 1
Run
Cancel query
Save query
Export data as
Show only Editor
1
Results
Messages
Search to filter items...