

Data Pipeline on AWS (ETL)

Description:

1. User: Starts the process by uploading a CSV file to the S3 bucket.

2. AWS S3 Bucket:

- Triggers an event when a new file is uploaded.
- Monitoring: CloudWatch logs S3 events for successful or failed uploads.

3. Lambda Function 1 (TransformData):

- Processes the uploaded CSV file.
- Sends a message to the Amazon SQS queue.
- Monitoring: CloudWatch tracks Lambda execution metrics such as invocation count, duration, errors, and throttles.

4. Amazon SQS Queue:

- Acts as a message buffer, ensuring asynchronous communication.
- Monitoring: CloudWatch tracks queue depth, message age, and any dead-letter queue (DLQ) activity.

5. Lambda Function 2 (SaveToMongo):

- Consumes messages from SQS.
- Writes the processed data to MongoDB.
- Monitoring: CloudWatch monitors invocation metrics and logs errors or exceptions.

MongoDB Atlas:

- The final destination where the data is stored.
- Monitoring: MongoDB Atlas provides its own monitoring tools for database performance and query metrics.

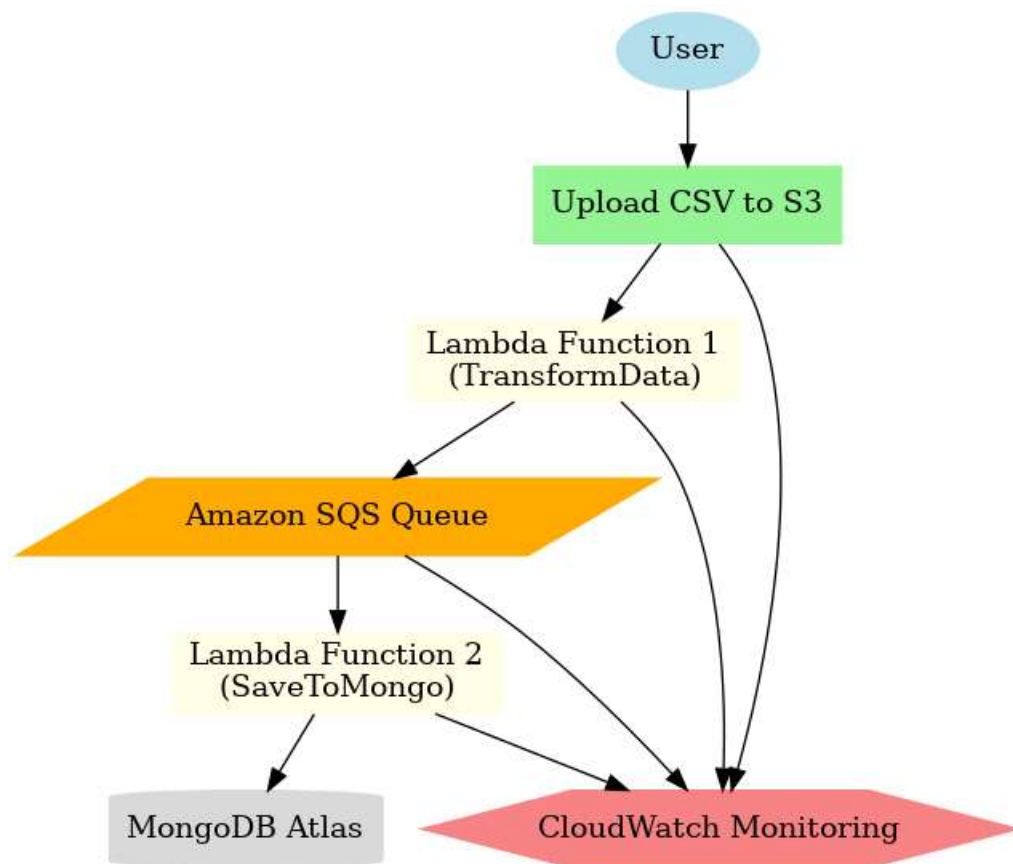
Dataset:

Financial Transactions

Dataset: Cards transactions (anonymized).

Source: Kaggle

Flow Chart:



Appendix:

Appendix A: Code

Lambda Function 1 (TransformData): Function name is TriggerSparkJob.

Code is attached as a separate file.

File Name : LambdaFunction1-TransformData.py

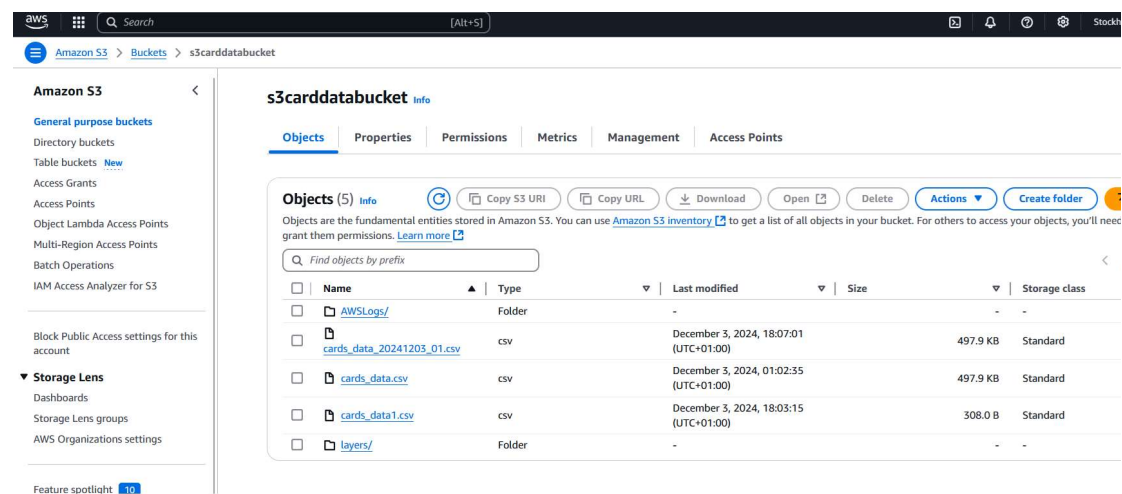
Lambda Function 2(MongoDBInsertion):

Code is attached as a separate file.

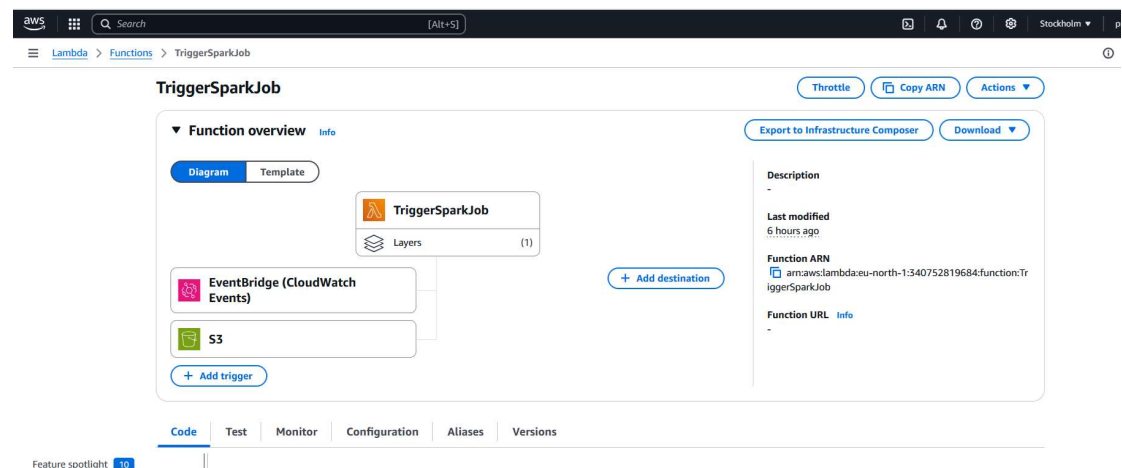
File Name : LambdaFunction2-MongoDBInsertion.py

Appendix B: Screenshots

AWS S3 Bucket:



Lambda Function 1 (TransformData): Function name is TriggerSparkJob.



Amazon SQS Queue:

TransformedDataQueue

[Edit](#) [Delete](#) [Purge](#) [Send and receive messages](#) [Start DLQ redrive](#)

Details [Info](#)

Name TransformedDataQueue	Type Standard	ARN arn:aws:sqs:eu-north-1:340752819684:TransformedDataQueue
Encryption Amazon SQS key (SSE-SQS)	URL https://sqs.eu-north-1.amazonaws.com/340752819684/TransformedDataQueue	Dead-letter queue -

► More

◀ [SNS subscriptions](#) [Lambda triggers](#) [EventBridge Pipes](#) [Dead-letter queue](#) [Monitoring](#) [Tagging](#) [Queue policies](#) [Encryption](#) ▶

Subscription region
eu-north-1

SNS subscriptions (0) [Info](#)

[View in SNS](#) [Delete](#) [Subscribe to Amazon SNS topic](#)

◀ 1 ▶


Lambda Function 2 (MongoDBInsertionLambda):


MongoDBInsertionLambda


[Throttle](#) [Copy ARN](#) [Actions](#)

▼ **Function overview** [Info](#)

[Diagram](#) [Template](#)

 **MongoDBInsertionLambda**

 **Layers** (1)

 **SQS**

[+ Add trigger](#)

[+ Add destination](#)

[Export to Infrastructure Composer](#) [Download](#)

Description
-

Last modified
6 hours ago

Function ARN
arn:aws:lambda:eu-north-1:340752819684:function:MongoDBInsertionLambda

Function URL [Info](#)
-

[Code](#) [Test](#) [Monitor](#) [Configuration](#) [Aliases](#) [Versions](#)

Code source [Info](#)

[Upload from](#)

◀ 1 ▶

Lambda Layers:

Lambda > Layers

Layers (1)

Last fetched 0 seconds ago [Create layer](#)

◀ 1 ▶

Name	Version	Description	Compatible runtimes	Compatible architectures	Created
pymongo-layer	4	Layer with pymongo package	python3.12	-	7 hours ago

Lambda

- Dashboard
- Applications
- Functions

▼ **Additional resources**

- Code signing configurations
- Event source mappings
- [Layers](#)
- Replicas

▼ **Related AWS resources**

- Step Functions state machines

MongoDB :

Atlas

Nackademini

Access Manager

Billing

All ClustersGet HelpPriyanka

Project 0

Data Services

Charts

Database

Clusters

Services

Atlas Search

Stream Processing

Triggers

Migration

Data Federation

Security

Quickstart

Backup

Database Access

Network Access

Advanced

NACKADEMIN > PROJECT 0

Network Access

IP Access List

Peering

Private Endpoint

+ADD IP ADDRESS

You will only be able to connect to your cluster from the following list of IP Addresses:

IP Address	Comment	Status	Actions
89.255.238.193/32 (includes your current IP address)		Active	EDITDELETE
0.0.0.0/0 (includes your current IP address)		Active	EDITDELETE

System Status: All Good

©2024 MongoDB, Inc. Status Terms Privacy Atlas Blog Contact Sales

MongoDB Compass - cluster0.ssdor.mongodb.net/test1.cards_data1

ConnectionsEditViewCollectionHelp

Compass

My Queries

Connections (1)

cluster0.ssdor.mongodb.net

admin

config

local

test

test1

cards_data1

cluster0.ssdor.mongodb.net > test1 > cards_data1

Documents2AggregationsSchemaIndexes1Validation

Type a query: { field: 'value' } or Generate query

ExplainResetFindOptions

ADD DATAEXPORT DATAUPDATEDELETE

1001 - 100 of 9274

	cards_data1	_id	ObjectId	id	Int32	client_id	Int32	card_brand	String	card_type	String	card_n
1	ObjectId('674f3ab854b7cea...')	4524		825		"Visa"		"Debit"		43446761		
2	ObjectId('674f3ab83d2848a...')	3781		825		"Visa"		"Debit"		45823134		
3	ObjectId('674f3ab8cbe04b2...')	2731		825		"Visa"		"Debit"		49569655		
4	ObjectId('674f3ab828d50a3...')	42		825		"Visa"		"Credit"		48794941		
5	ObjectId('674f3ab954b7cea...')	1278		1746		"Visa"		"Debit"		48814821		
6	ObjectId('674f3aba54b7cea...')	4537		1746		"Visa"		"Credit"		44048988		
7	ObjectId('674f3aba28d50a3...')	3687		1746		"Mastercard"		"Debit"		56272208		
8	ObjectId('674f3aba28d50a3...')	2732		1718		"Visa"		"Debit"		42420151		
9	ObjectId('674f3aba28d50a3...')	5435		68		"Visa"		"Credit"		44114561		
10	ObjectId('674f3aba28d50a3...')	4985		1116		"Mastercard"		"Debit"		56688311		
11	ObjectId('674f3aba28d50a3...')	3755		1894		"Mastercard"		"Debit"		55157821		
12	ObjectId('674f3abacbe04b2...')	2379		1718		"Mastercard"		"Debit"		57663521		
13	ObjectId('674f3aba28d50a3...')	4455		1660		"Mastercard"		"Debit"		52314621		
14	ObjectId('674f3abacbe04b2...')	5559		1116		"Mastercard"		"Debit"		53410491		
15	ObjectId('674f3abacbe04b2...')	1768		1894		"Visa"		"Credit"		45883881		
16	ObjectId('674f3abacbe04b2...')	4198		1747		"Mastercard"		"Debit"		51791571		

Appendix C: Dashboards

CloudWatch Dashboards

Custom Dashboards (3) Info

Share dashboard Delete Create dashboard

Filter dashboards

Name	Sharing	Favorite	Last update (Local)
CloudWatch-Metrics-For-LambdaFunction1		★	2024-12-04 00:17
CloudWatch-Metrics-For-AmazonSQSQueue		☆	2024-12-04 00:22
CloudWatch-Metrics-For-LambdaFunction2		☆	2024-12-04 00:18

CloudWatch

Alarms

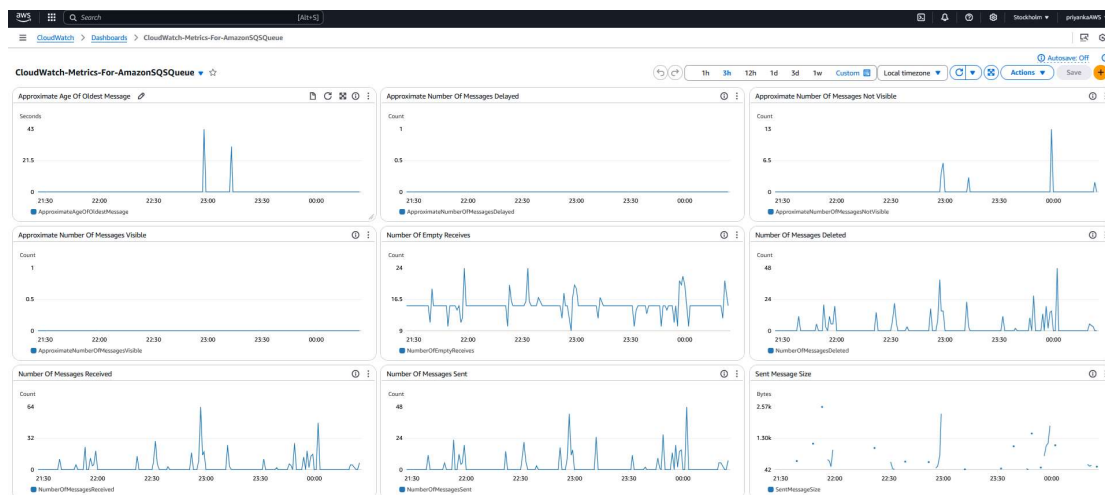
Logs

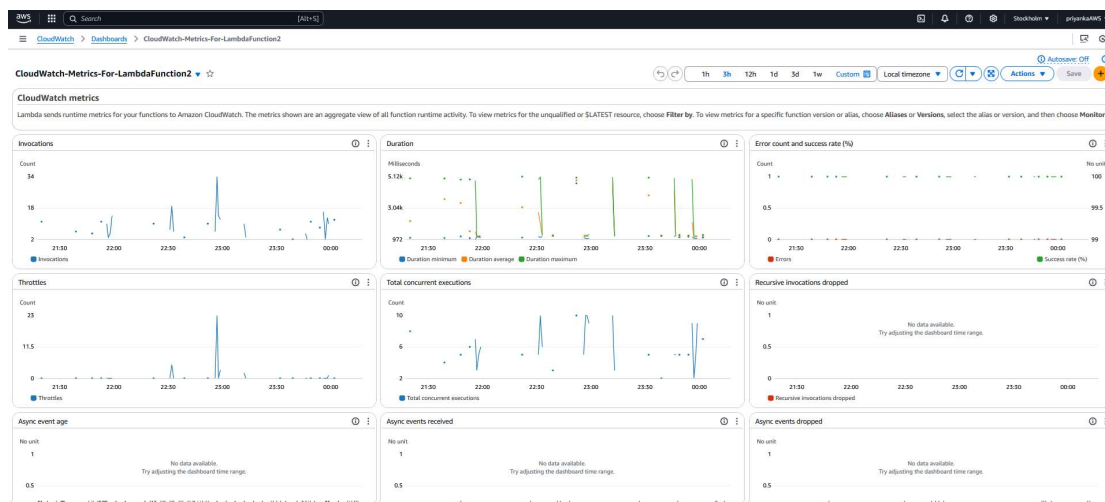
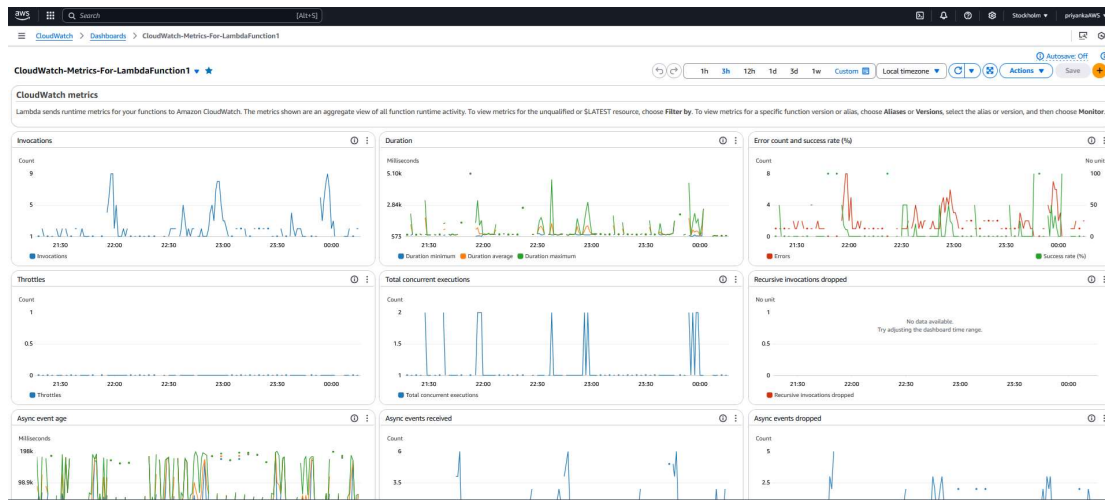
- Log groups
- Log Anomalies
- Live Tail
- Logs Insights
- Contributor Insights

Metrics

- All metrics
- Explorer
- Streams

X-Ray traces





NOTE: Public Links for CloudWatch Dashboards are given in CloudWatch-Dashboard-Public-Links.rtf document.