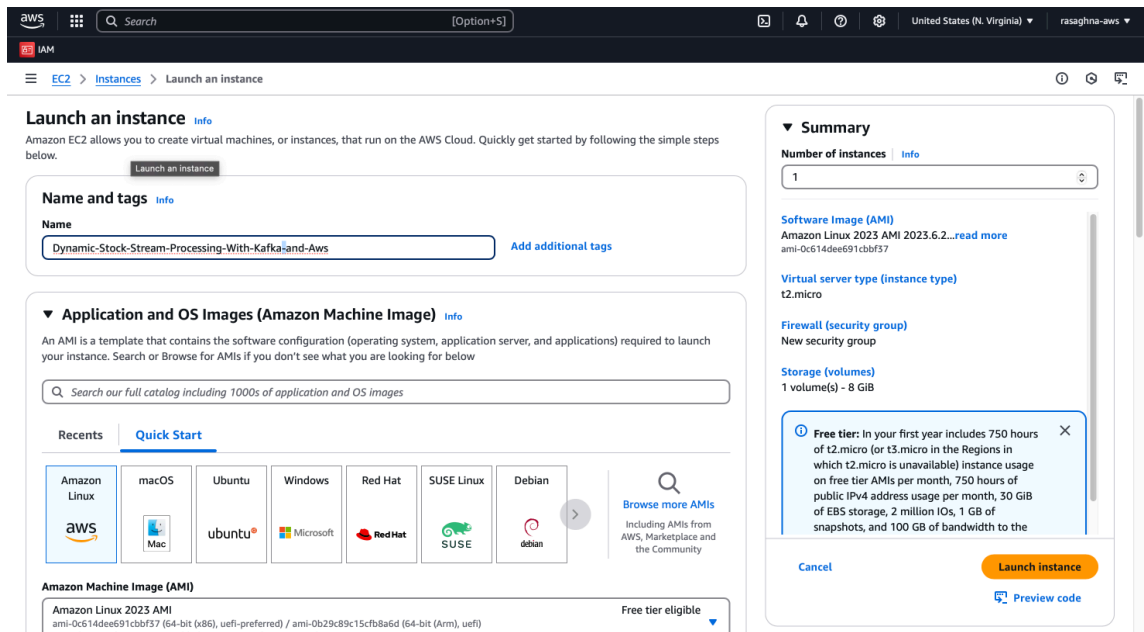1.Create a EC2 instance in your AWS console

2. Refer to this document and connect to the EC2 instance through the Terminal "https://docs.aws.amazon.com/msk/latest/developerguide/mkc-create-topic.html"

3.Download apache kafka compressed version "`wget` `https://archive.apache.org/dist/kafka/2.2.1/kafka_2.12-2.2.1.tgz`"

4. Kafka runs on JAVA so install java "`sudo yum install java-1.8.0-openjdk`" on your local machine

5. Start zookeeper in your Kafka "c"

6. Connect to EC2 instance via new terminal

7. Run this command "`cd kafka_2.12-2.2.1/`" check the kafka version while running the command

8. Edit the inbound rules to take all the traffic from Ipv4

9. Open a separate terminal to create topics for consumer and producer to consume.

10. bin/kafka-topics.sh --create --topic stock_market_demo --bootstrap-server 3.83.127.4:9092 --replication-factor 1 --partitions 1

11. As I am using older version of kafka here is the command to run the producers "bin/kafka-console-producer.sh --topic stock_market_demo --broker-list 3.83.127.4:9092"

12. Here is the code to run the consumers "bin/kafka-console-consumer.sh --topic stock_market_demo--bootstrap-server 3.83.127.4:9092"

bin/kafka-console-consumer.sh --topic stock_market_demo --bootstrap-server 3.83.127.4:9092

Possible error: Missing the EC2 port in the producer and consumer code

Python version error python >3.10 might cause some errors, so downgrade the python version
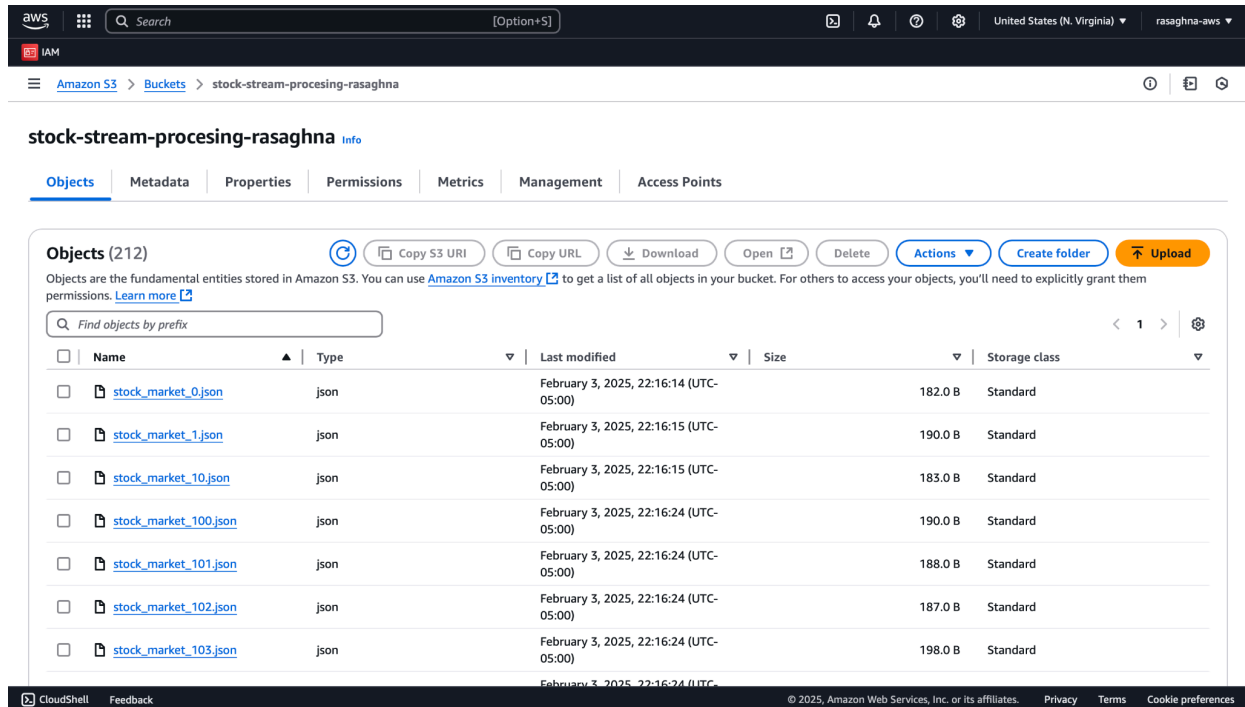
Producer not actually sending messages to Kafka
Consumer is listening but unable to find messages
Wrong Kafka topic setup
Offset reset settings in the consumer
Kafka logs for potential errors

Data in S3 bucket :

Setting up the Glue crawler :

AWS Glue > Crawlers > stock_market_kafka_rk

AWS Glue

Getting started
ETL jobs
   Visual ETL
   Notebooks
   Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)
Zero-ETL integrations New
Data Catalog
   Databases
   Tables
Stream schema registries
   Schemas
Connections
Crawlers
   Classifiers
Catalog settings
Data Integration and ETL
Legacy pages

What's New
Documentation
AWS Marketplace

Enable compact mode

stock_market_kafka_rk

Last updated (UTC) February 4, 2025 at 03:54:19

Run crawler | Edit | Delete

Crawler properties

Name	IAM role	Database	State
stock_market_kafka_rk	aws-stock-market-kafka	stock-market-kafka	READY

Description	Security configuration	Lake Formation configuration	Table prefix
-	-	-	-

Maximum table threshold
-

▶ Advanced settings

Crawler runs | Schedule | Data sources | Classifiers | Tags

Crawler runs (3)
The list of crawler runs for this crawler.

Stop run | View CloudWatch logs | View run details

Filter data

Filter by a date and time range

< 1 >

	Start time (UTC) ▲	End time (UTC) ▼	Current/last duration ▼	Status ▼	DPU hours ▼	Table changes ▼
○	February 4, 2025 at 03:18:50	February 4, 2025 at 03:20:13	01 min 22 s	⊘ Completed	0.042	1 table change, 0 partition changes
○	February 4, 2025 at 03:13:56	February 4, 2025 at 03:15:09	01 min 12 s	⊘ Completed	0.047	2 table changes, 0 partition changes
○	February 4, 2025 at 03:02:04	February 4, 2025 at 03:04:14	02 min 10 s	⊘ Completed	0.042	1 table change, 1 partition change

Querying in AWS Athena:

Amazon Athena > Query editor tabs

```
1  SELECT * FROM "stock-market-kafka"."stock_stream_procesing_rasaghna" limit 10;
```

Data source
AwsDataCatalog ▼

catalogue
None ▼

Database
stock-market-kafka ▼

Tables and views Create ▼

Filter tables and views

▼ Tables (1) < 1 >
   stock_stream_procesing_rasaghna

▼ Views (0) < 1 >

SQL   Ln 1, Col 79

Run again | Explain | Cancel | Clear | Create ▼

Reuse query results
up to 60 minutes ago

Query results | Query status

⊘ Completed | Time in queue: 118 ms | Run time: 532 ms | Data scanned: 2.40 KB

Results (10)

Copy | Download results

Search rows

< 1 >

# ▽	index ▽	date ▽	open ▽	high ▽	low ▽	close ▽	adj close ▽	volume ▽
1	HSI	1997-03-13	13056.29981	13056.29981	12877.09961	12917.09961	12917.09961	0.0
2	IXIC	1980-04-21	133.080002	133.080002	133.080002	133.080002	133.080002	0.0