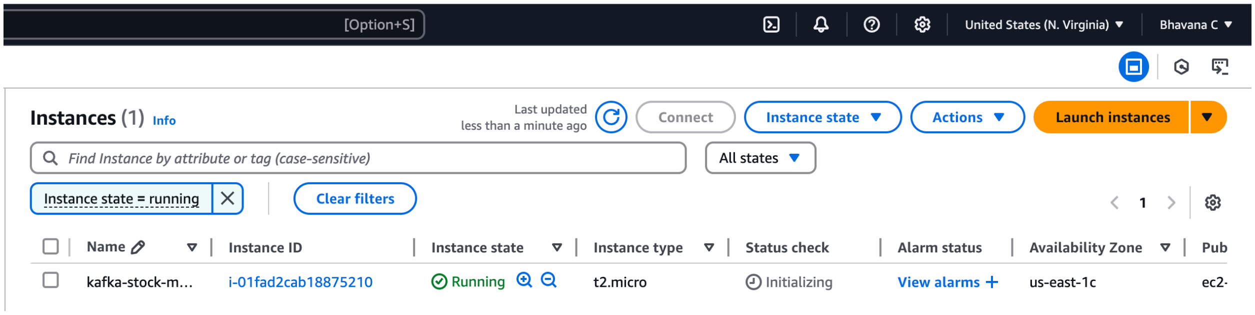
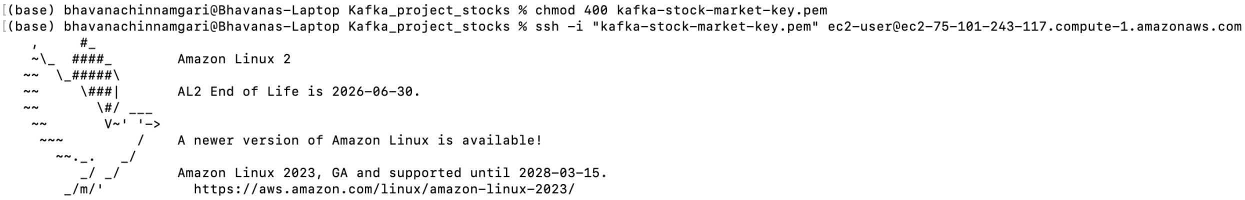
Stock Market Project – Kafka and AWS

Create an EC2 instance in AWS:



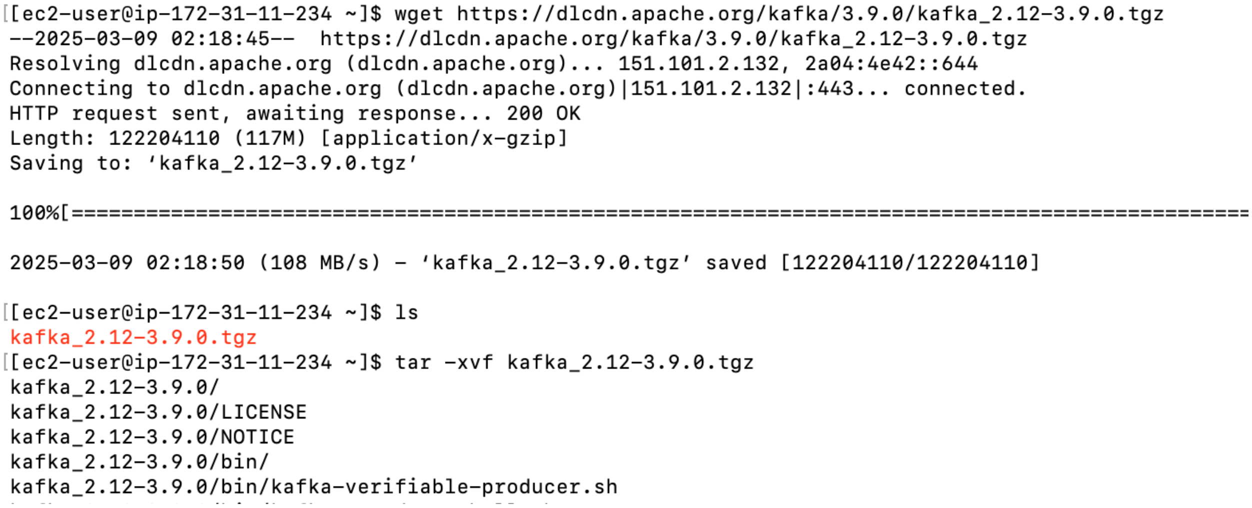
□	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Pub
□	kafka-stock-m...	i-01fad2cab18875210	Running	t2.micro	Initializing	View alarms +	us-east-1c	ec2-

Start EC2:
cd "/Users/bhavanachinnamgari/Projects/Kafka_project_stocks"
ssh -i "kafka-stock-market-key.pem" ec2-user@ec2-75-101-243-117.compute-1.amazonaws.com



```
[(base) bhavanachinnamgari@Bhavanas-Laptop Kafka_project_stocks % chmod 400 kafka-stock-market-key.pem
[(base) bhavanachinnamgari@Bhavanas-Laptop Kafka_project_stocks % ssh -i "kafka-stock-market-key.pem" ec2-user@ec2-75-101-243-117.compute-1.amazonaws.com
   ,     #_
   ~\_  ####_        Amazon Linux 2
  ~~  \_#####\
  ~~     \###|       AL2 End of Life is 2026-06-30.
  ~~       \#/ ___
   ~~       V~' '->
    ~~~         /    A newer version of Amazon Linux is available!
      ~~._.   _/
         _/ _/       Amazon Linux 2023, GA and supported until 2028-03-15.
       _/m/'           https://aws.amazon.com/linux/amazon-linux-2023/
```

Download Kafka and unzip the folder:
wget https://dlcdn.apache.org/kafka/3.9.0/kafka-3.9.0-src.tgz
tar -xvf kafka_2.12-3.9.0.tgz

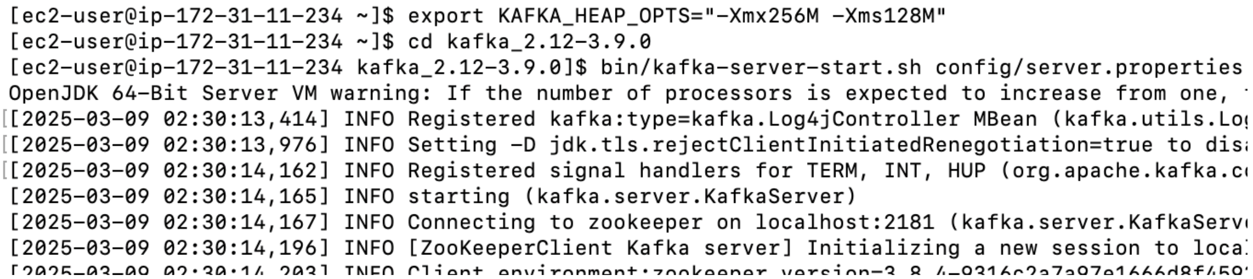


```
[[ec2-user@ip-172-31-11-234 ~]$ wget https://dlcdn.apache.org/kafka/3.9.0/kafka_2.12-3.9.0.tgz
 --2025-03-09 02:18:45--  https://dlcdn.apache.org/kafka/3.9.0/kafka_2.12-3.9.0.tgz
 Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::644
 Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.
 HTTP request sent, awaiting response... 200 OK
 Length: 122204110 (117M) [application/x-gzip]
 Saving to: ‘kafka_2.12-3.9.0.tgz’

 100%[=============================================================================================

 2025-03-09 02:18:50 (108 MB/s) - ‘kafka_2.12-3.9.0.tgz’ saved [122204110/122204110]

[[ec2-user@ip-172-31-11-234 ~]$ ls
 kafka_2.12-3.9.0.tgz
[[ec2-user@ip-172-31-11-234 ~]$ tar -xvf kafka_2.12-3.9.0.tgz
 kafka_2.12-3.9.0/
 kafka_2.12-3.9.0/LICENSE
 kafka_2.12-3.9.0/NOTICE
 kafka_2.12-3.9.0/bin/
 kafka_2.12-3.9.0/bin/kafka-verifiable-producer.sh
```

Check for Java JDK, install if not present: sudo yum install java-1.8.0-openjdk
Change path into folder: cd kafka_2.12-3.9.0
Run zookeeper in Terminal-1: bin/zookeeper-server-start.sh config/zookeeper.properties

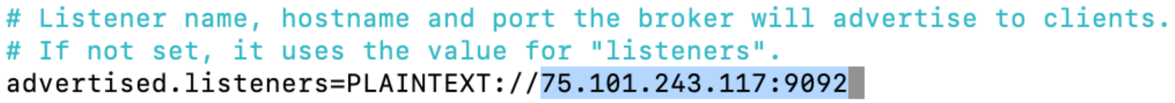
---

Start EC2 in Terminal-2 and run Kafka server:
export KAFKA_HEAP_OPTS="-Xmx256M -Xms128M"
cd kafka_2.12-3.9.0
bin/kafka-server-start.sh config/server.properties



```
[ec2-user@ip-172-31-11-234 ~]$ export KAFKA_HEAP_OPTS="-Xmx256M -Xms128M"
[ec2-user@ip-172-31-11-234 ~]$ cd kafka_2.12-3.9.0
[ec2-user@ip-172-31-11-234 kafka_2.12-3.9.0]$ bin/kafka-server-start.sh config/server.properties
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to increase from one,
[2025-03-09 02:30:13,414] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log
[2025-03-09 02:30:13,976] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disa
[2025-03-09 02:30:14,162] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.co
[2025-03-09 02:30:14,165] INFO starting (kafka.server.KafkaServer)
[2025-03-09 02:30:14,167] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServe
[2025-03-09 02:30:14,196] INFO [ZooKeeperClient Kafka server] Initializing a new session to local
[2025-03-09 02:30:14,203] INFO Client environment:zookeeper.version=3.8.4-9316c2a7a97e1666d8f459
```

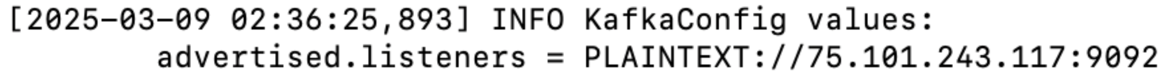
Change properties file with EC2 port number: sudo nano config/server.properties



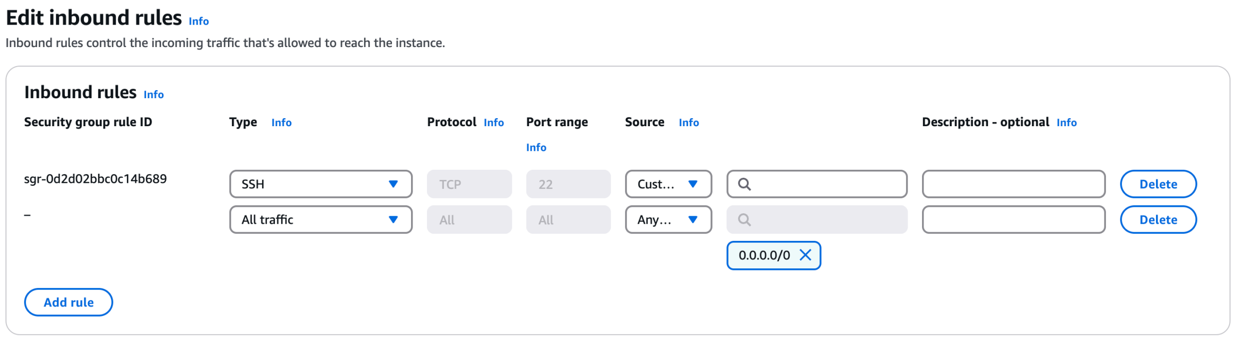
```
# Listener name, hostname and port the broker will advertise to clients.
# If not set, it uses the value for "listeners".
advertised.listeners=PLAINTEXT://75.101.243.117:9092
```

CTRL + X, Y, RETURN

Run both Zookeeper and Kafka again and check below in Kafka server:



```
[2025-03-09 02:36:25,893] INFO KafkaConfig values:
        advertised.listeners = PLAINTEXT://75.101.243.117:9092
```



Edit inbound rules Info

Inbound rules control the incoming traffic that's allowed to reach the instance.

Inbound rules Info

Security group rule ID	Type Info	Protocol Info	Port range Info	Source Info		Description - optional Info	
sgr-0d2d02bbc0c14b689	SSH	TCP	22	Cust...			Delete
–	All traffic	All	All	Any...	0.0.0.0/0		Delete

Now, we have Zookeeper and Kafka setup done!

Let’s create Topic, Producer and Consumer. Start EC2 instances in new terminals and Goto Kafka folder

Create the topic with replication factor and partitions: Terminal-3

bin/kafka-topics.sh --create --topic stocks-test6 --bootstrap-server 75.101.243.117:9092 --partitions 1 --replication-factor 1

General command: bin/kafka-topics.sh --create --topic {topic name} --bootstrap-server {EC2 instance/localhost}:9092 --partitions 1 --replication-factor 1

# “stocks-test6” Topic gets created.

Start Producer: Terminal-3

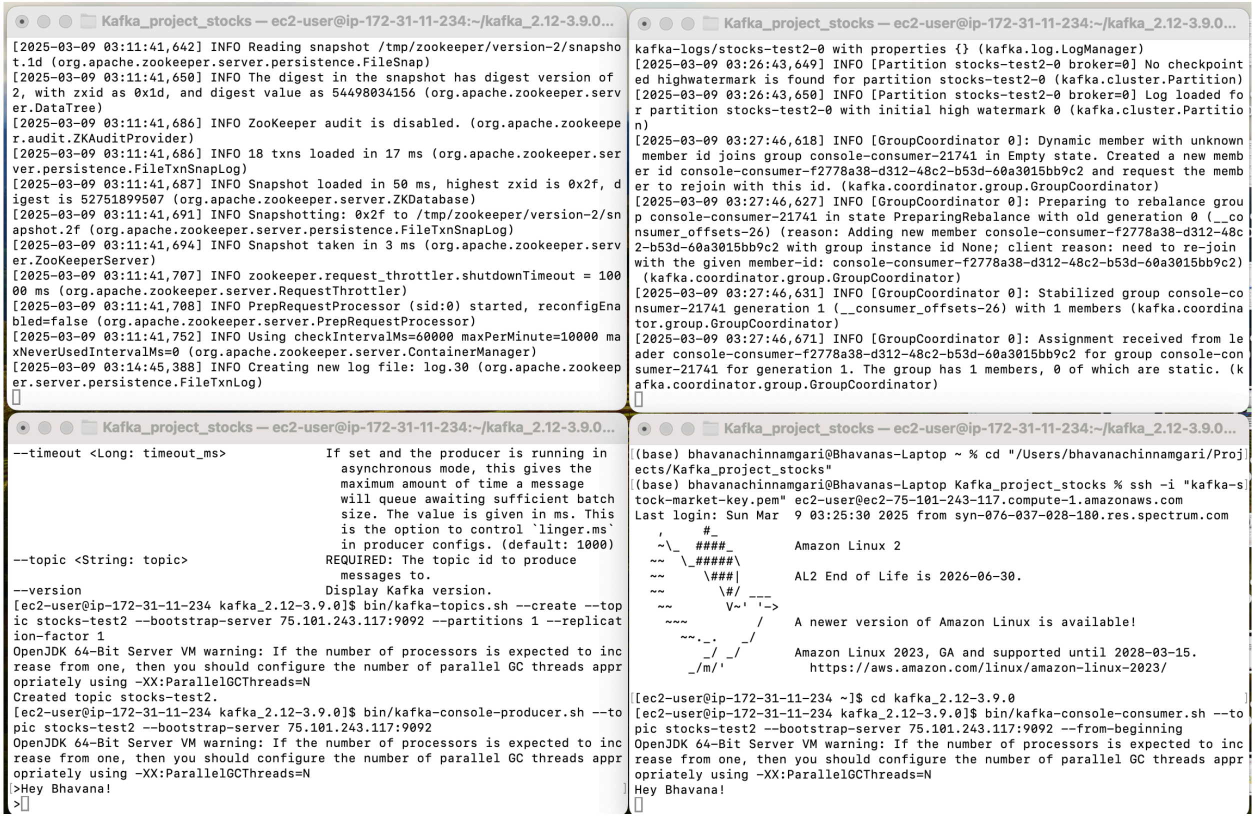
bin/kafka-console-producer.sh --topic stocks-test6 --bootstrap-server 75.101.243.117:9092

Start Consumer in a new Terminal-4

cd kafka_2.12-3.9.0

bin/kafka-console-consumer.sh --topic stocks-test6 --bootstrap-server 75.101.243.117:9092 --from-beginning

---


Kafka_project_stocks — ec2-user@ip-172-31-11-234:~/kafka_2.12-3.9.0...
Kafka_project_stocks — ec2-user@ip-172-31-11-234:~/kafka_2.12-3.9.0...
Kafka_project_stocks — ec2-user@ip-172-31-11-234:~/kafka_2.12-3.9.0...
Kafka_project_stocks — ec2-user@ip-172-31-11-234:~/kafka_2.12-3.9.0...

```
[2025-03-09 03:11:41,642] INFO Reading snapshot /tmp/zookeeper/version-2/snapsho
t.1d (org.apache.zookeeper.server.persistence.FileSnap)
[2025-03-09 03:11:41,650] INFO The digest in the snapshot has digest version of
2, with zxid as 0x1d, and digest value as 54498034156 (org.apache.zookeeper.serv
er.DataTree)
[2025-03-09 03:11:41,686] INFO ZooKeeper audit is disabled. (org.apache.zookeepe
r.audit.ZKAuditProvider)
[2025-03-09 03:11:41,686] INFO 18 txns loaded in 17 ms (org.apache.zookeeper.ser
ver.persistence.FileTxnSnapLog)
[2025-03-09 03:11:41,687] INFO Snapshot loaded in 50 ms, highest zxid is 0x2f, d
igest is 52751899507 (org.apache.zookeeper.server.ZKDatabase)
[2025-03-09 03:11:41,691] INFO Snapshotting: 0x2f to /tmp/zookeeper/version-2/sn
apshot.2f (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2025-03-09 03:11:41,694] INFO Snapshot taken in 3 ms (org.apache.zookeeper.serv
er.ZooKeeperServer)
[2025-03-09 03:11:41,707] INFO zookeeper.request_throttler.shutdownTimeout = 100
00 ms (org.apache.zookeeper.server.RequestThrottler)
[2025-03-09 03:11:41,708] INFO PrepRequestProcessor (sid:0) started, reconfigEna
bled=false (org.apache.zookeeper.server.PrepRequestProcessor)
[2025-03-09 03:11:41,752] INFO Using checkIntervalMs=60000 maxPerMinute=10000 ma
xNeverUsedIntervalMs=0 (org.apache.zookeeper.server.ContainerManager)
[2025-03-09 03:14:45,388] INFO Creating new log file: log.30 (org.apache.zookeep
er.server.persistence.FileTxnLog)
```

```
kafka-logs/stocks-test2-0 with properties {} (kafka.log.LogManager)
[2025-03-09 03:26:43,649] INFO [Partition stocks-test2-0 broker=0] No checkpoint
ed highwatermark is found for partition stocks-test2-0 (kafka.cluster.Partition)
[2025-03-09 03:26:43,650] INFO [Partition stocks-test2-0 broker=0] Log loaded fo
r partition stocks-test2-0 with initial high watermark 0 (kafka.cluster.Partitio
n)
[2025-03-09 03:27:46,618] INFO [GroupCoordinator 0]: Dynamic member with unknown
 member id joins group console-consumer-21741 in Empty state. Created a new memb
er id console-consumer-f2778a38-d312-48c2-b53d-60a3015bb9c2 and request the memb
er to rejoin with this id. (kafka.coordinator.group.GroupCoordinator)
[2025-03-09 03:27:46,627] INFO [GroupCoordinator 0]: Preparing to rebalance grou
p console-consumer-21741 in state PreparingRebalance with old generation 0 (__co
nsumer_offsets-26) (reason: Adding new member console-consumer-f2778a38-d312-48c
2-b53d-60a3015bb9c2 with group instance id None; client reason: need to re-join
with the given member-id: console-consumer-f2778a38-d312-48c2-b53d-60a3015bb9c2)
 (kafka.coordinator.group.GroupCoordinator)
[2025-03-09 03:27:46,631] INFO [GroupCoordinator 0]: Stabilized group console-co
nsumer-21741 generation 1 (__consumer_offsets-26) with 1 members (kafka.coordina
tor.group.GroupCoordinator)
[2025-03-09 03:27:46,671] INFO [GroupCoordinator 0]: Assignment received from le
ader console-consumer-f2778a38-d312-48c2-b53d-60a3015bb9c2 for group console-con
sumer-21741 for generation 1. The group has 1 members, 0 of which are static. (k
afka.coordinator.group.GroupCoordinator)
```

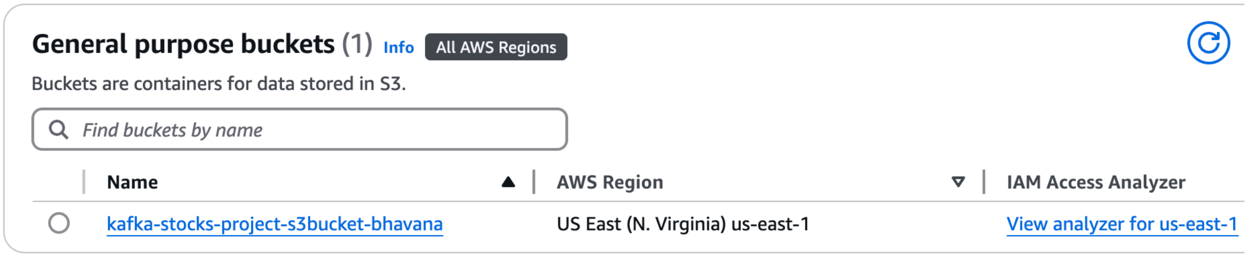
```
--timeout <Long: timeout_ms>              If set and the producer is running in
                                            asynchronous mode, this gives the
                                            maximum amount of time a message
                                            will queue awaiting sufficient batch
                                            size. The value is given in ms. This
                                            is the option to control `linger.ms`
                                            in producer configs. (default: 1000)
--topic <String: topic>                   REQUIRED: The topic id to produce
                                            messages to.
--version                                 Display Kafka version.
[ec2-user@ip-172-31-11-234 kafka_2.12-3.9.0]$ bin/kafka-topics.sh --create --top
ic stocks-test2 --bootstrap-server 75.101.243.117:9092 --partitions 1 --replicat
ion-factor 1
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to inc
rease from one, then you should configure the number of parallel GC threads appr
opriately using -XX:ParallelGCThreads=N
Created topic stocks-test2.
[ec2-user@ip-172-31-11-234 kafka_2.12-3.9.0]$ bin/kafka-console-producer.sh --to
pic stocks-test2 --bootstrap-server 75.101.243.117:9092
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to inc
rease from one, then you should configure the number of parallel GC threads appr
opriately using -XX:ParallelGCThreads=N
>Hey Bhavana!
>
```

```
(base) bhavanachinnamgari@Bhavanas-Laptop ~ % cd "/Users/bhavanachinnamgari/Proj
ects/Kafka_project_stocks"
(base) bhavanachinnamgari@Bhavanas-Laptop Kafka_project_stocks % ssh -i "kafka-s
tock-market-key.pem" ec2-user@ec2-75-101-243-117.compute-1.amazonaws.com
Last login: Sun Mar  9 03:25:30 2025 from syn-076-037-028-180.res.spectrum.com
   ,     #_
   ~\_  ####_        Amazon Linux 2
  ~~  \_#####\
  ~~     \###|       AL2 End of Life is 2026-06-30.
  ~~       \#/ ___
   ~~       V~' '->
    ~~~         /    A newer version of Amazon Linux is available!
      ~~._.   _/
         _/ _/       Amazon Linux 2023, GA and supported until 2028-03-15.
       _/m/'           https://aws.amazon.com/linux/amazon-linux-2023/

[ec2-user@ip-172-31-11-234 ~]$ cd kafka_2.12-3.9.0
[ec2-user@ip-172-31-11-234 kafka_2.12-3.9.0]$ bin/kafka-console-consumer.sh --to
pic stocks-test2 --bootstrap-server 75.101.243.117:9092 --from-beginning
OpenJDK 64-Bit Server VM warning: If the number of processors is expected to inc
rease from one, then you should configure the number of parallel GC threads appr
opriately using -XX:ParallelGCThreads=N
Hey Bhavana!
```

Producer and Consumer are working well!

Login to AWS using CLI with Access ID and Secret ID. Write the python code for the kafka project to display output in Jupyter/ VS code for testing. Then use stocks_data.csv to populate them in S3 bucket.

Create S3 bucket to save the output:

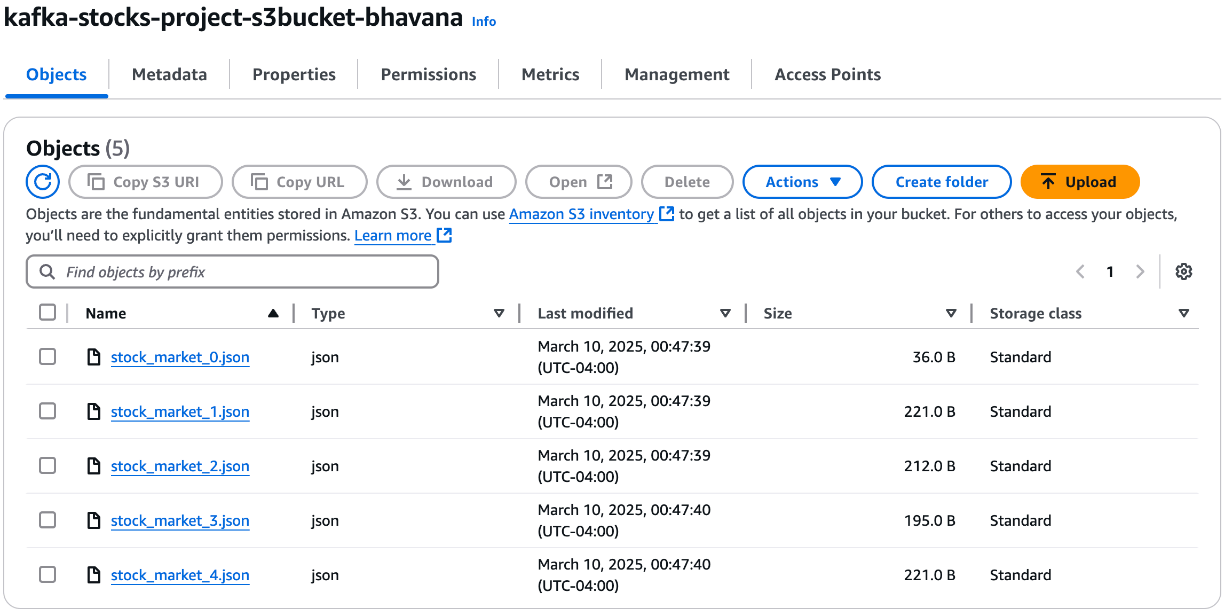


General purpose buckets (1)

Buckets are containers for data stored in S3.

	Name	AWS Region	IAM Access Analyzer
○	kafka-stocks-project-s3bucket-bhavana	US East (N. Virginia) us-east-1	View analyzer for us-east-1

Run for 5 sec and check the console if the json data is being populated.



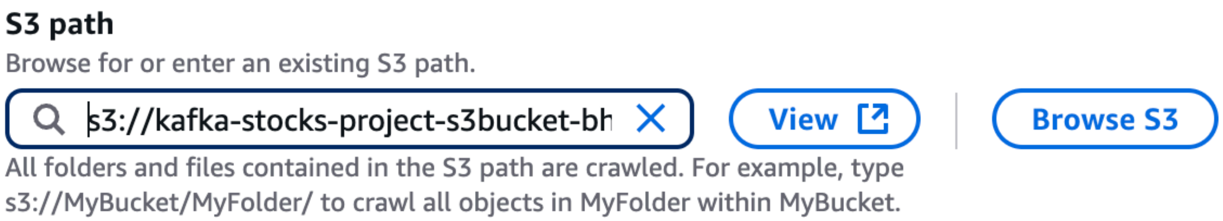
kafka-stocks-project-s3bucket-bhavana

Objects (5)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

□	Name	Type	Last modified	Size	Storage class
□	stock_market_0.json	json	March 10, 2025, 00:47:39 (UTC-04:00)	36.0 B	Standard
□	stock_market_1.json	json	March 10, 2025, 00:47:39 (UTC-04:00)	221.0 B	Standard
□	stock_market_2.json	json	March 10, 2025, 00:47:39 (UTC-04:00)	212.0 B	Standard
□	stock_market_3.json	json	March 10, 2025, 00:47:40 (UTC-04:00)	195.0 B	Standard
□	stock_market_4.json	json	March 10, 2025, 00:47:40 (UTC-04:00)	221.0 B	Standard

Create a AWS Crawler using our S3 bucket



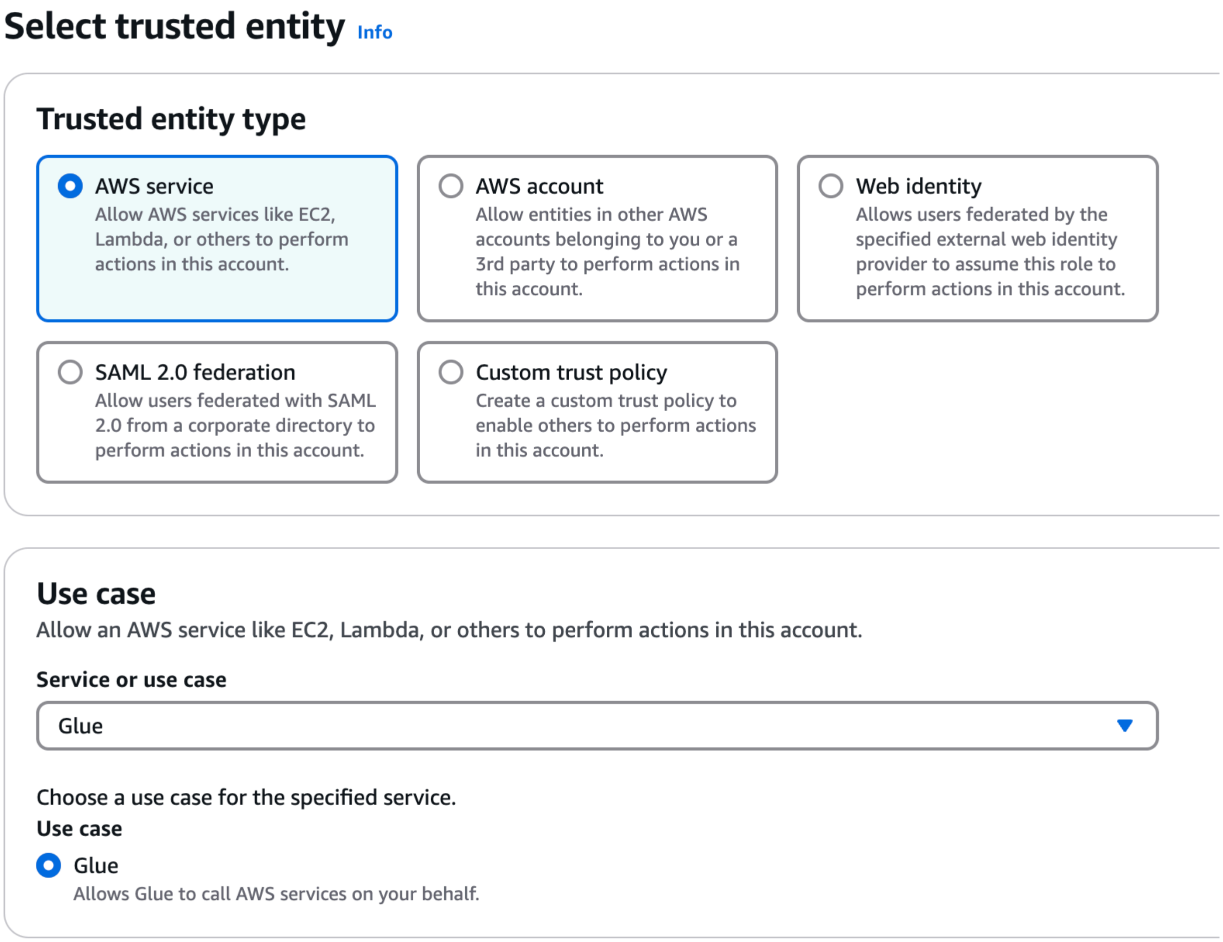
S3 path

Browse for or enter an existing S3 path.

s3://kafka-stocks-project-s3bucket-bh

All folders and files contained in the S3 path are crawled. For example, type s3://MyBucket/MyFolder/ to crawl all objects in MyFolder within MyBucket.

Create a Role to give access to AWS Glue:



Select trusted entity

Trusted entity type

- ● AWS service: Allow AWS services like EC2, Lambda, or others to perform actions in this account.
- ○ AWS account: Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.
- ○ Web identity: Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.
- ○ SAML 2.0 federation: Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.
- ○ Custom trust policy: Create a custom trust policy to enable others to perform actions in this account.

Use case

Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

Service or use case

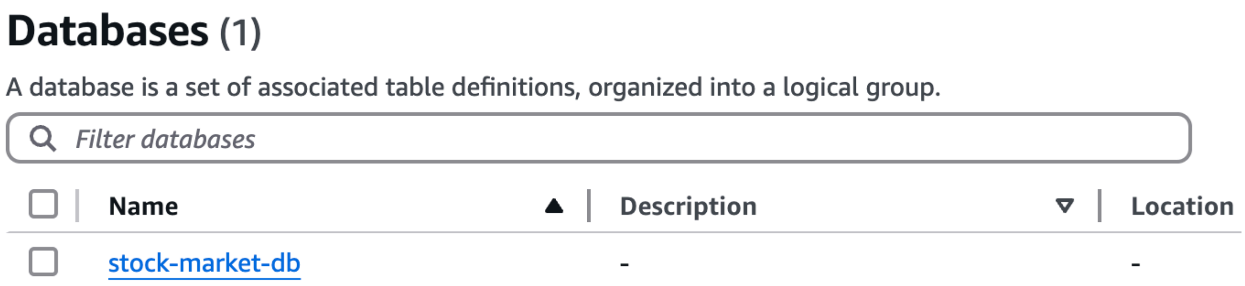
Glue

Choose a use case for the specified service.

Use case

● Glue
Allows Glue to call AWS services on your behalf.

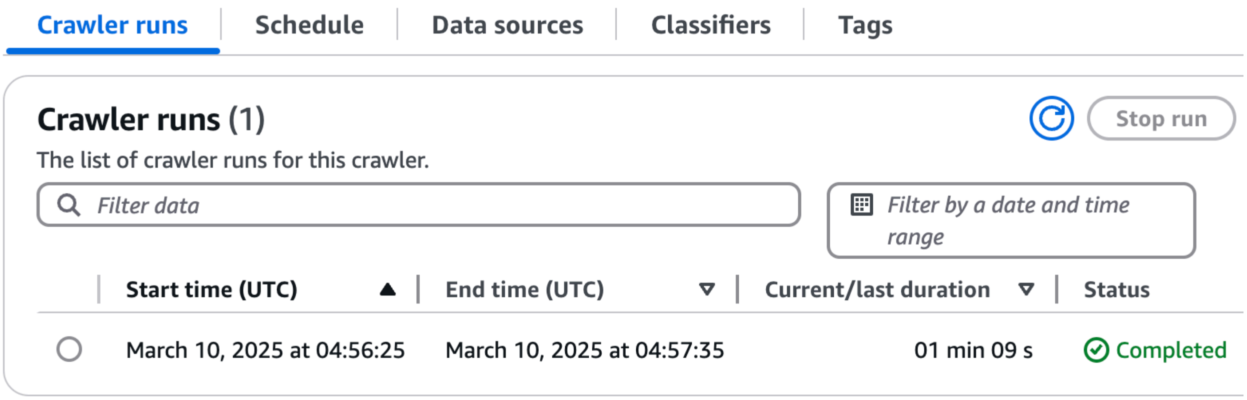
Create a Database “stock-market-db” to store the table created by Crawler:



Databases (1)

A database is a set of associated table definitions, organized into a logical group.

□	Name	Description	Location
□	stock-market-db	-	-

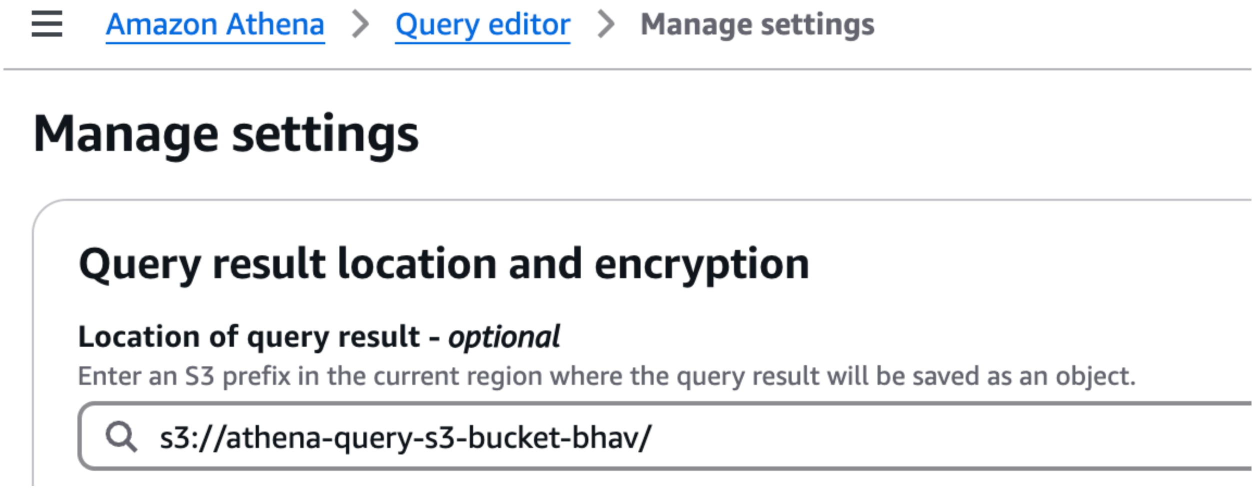


Crawler runs (1)

The list of crawler runs for this crawler.

	Start time (UTC)	End time (UTC)	Current/last duration	Status
○	March 10, 2025 at 04:56:25	March 10, 2025 at 04:57:35	01 min 09 s	✓ Completed

A Table is created under the database “stock-market-db”. Now we can execute queries using AWS Athena. Create another S3 bucket to save the query information.



Manage settings

Query result location and encryption

Location of query result - *optional*

Enter an S3 prefix in the current region where the query result will be saved as an object.

s3://athena-query-s3-bucket-bhav/