CSP 554 Big Data Technologies

Assignment – #9

Shiva Sankar Modala(A20517528)

Exercise 1 (5 points)

Read the article "Real-time stream processing for Big Data" available on the blackboard in the 'Articles' section and then answer the following questions:

a) (1.25 points) What is the Kappa architecture and how does it differ from the lambda architecture?

Ans: In the Kappa Architecture, "everything is a stream – we only need a stream processing engine". Kappa architecture involves doing all data computations in a stream processing system. The Streaming Layer is where the computation takes place. The Kappa Architecture does not have a batch layer unlike the Lambda architecture and it continuously processes the data as it arrives, where it was batch processing.

b) (1.25 points) What are the advantages and drawbacks of pure streaming versus micro-batch real-time processing systems?

Ans: While all stream processors have some common ground in terms of basic concepts and working principles, there is a significant difference between the particular systems that directly affects the processing speed. Handling data items as they arrive reduces latency at the expense of high per-item costs (e.g., through messaging), whereas buffering and processing them in batches improves efficiency while increasing the time each item spends in the data pipeline. Storm and Samza are pure stream-oriented systems with very low latency and somewhat high per-item costs, whereas batch-oriented systems offer unmatched resource efficiency at the tradeoff of unreasonably high latency for real-time applications. Because the gap between these two extremes is so large, several systems, such as Storm Trident and Spark Streaming, use micro-batching algorithms to exchange latency for throughput: Trident groups tuples into batches in order to relax the one-at-a-time processing approach in favour of higher throughput, whereas Spark Streaming limits batch size in a native batch processor in order to reduce latency.

c) (1.25 points) In few sentences describe the data processing pipeline in Storm.

Ans: A data pipeline or application in Storm is called a topology and as illustrated is a directed graph that represents data flow as directed edges between nodes which again represent the individual processing steps: The nodes that ingest data and thus initiate the data flow in the topology are called spouts and emit tuples to the nodes downstream which are called bolts and do processing, write data to external storage and may send tuples further downstream themselves. Storm comes with several groupings that control data flow between nodes, e.g. for shuffling or hash-partitioning a stream of tuples by some attribute value, but also allows arbitrary custom groupings. By default, Storm distributes spouts and bolts across the nodes in the cluster in a round-robin fashion, though the scheduler is pluggable to account for scenarios in which a certain processing step has to be executed on a particular node, for example because of hardware dependencies.

d) (1.25 points) How does Spark streaming shift the Spark batch processing approach to work on real-time data streams?

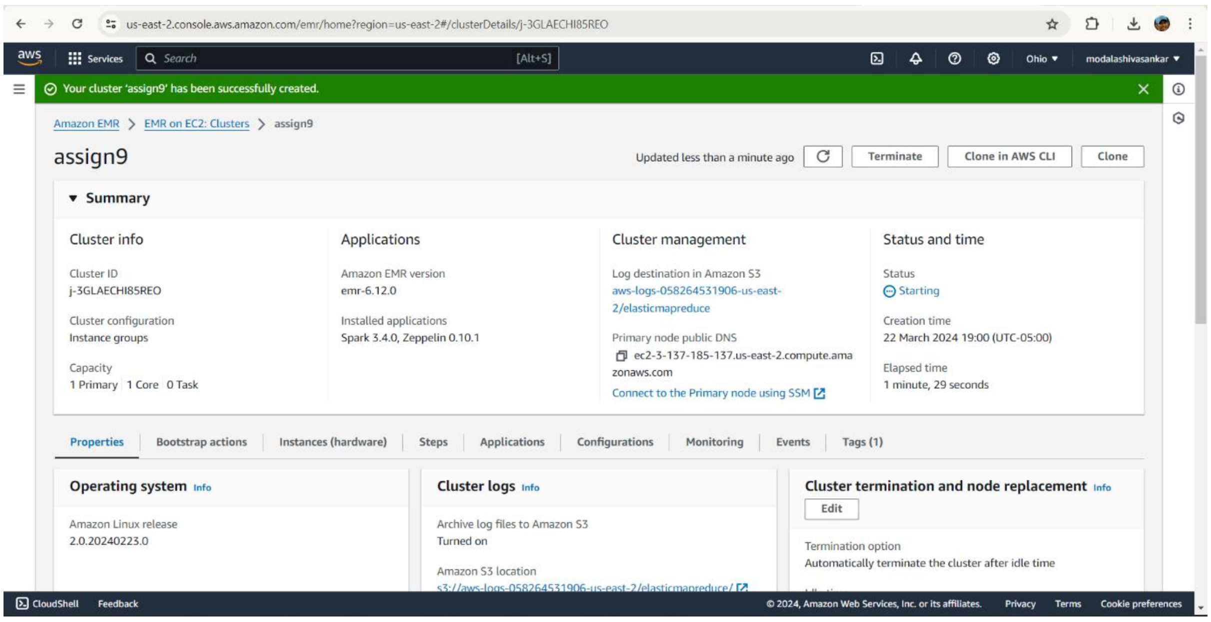
Ans: Spark Streaming shifts Spark's batch-processing approach towards real-time requirements by chunking the stream of incoming data items into small batches, transforming them into Resilient Distributed Datasets (RDDs) and processing them as usual. It further takes care of data flow and distribution automatically. Data is ingested and transformed into a sequence of RDDs which is called Discretized Stream (DStream) before processing through workers. All RDDs in a DStream are processed in order, whereas data items inside an RDD are processed in parallel without any ordering guarantees.

Exercise 2(5 points)

Step A – Start an EMR cluster

Start up an EMR/Hadoop cluster as previously, but instead of choosing the "Core Hadoop" configuration chose the "Spark" configuration (see below), otherwise proceed as before.

Created an EMR cluster names Spark cluster



us-east-2.console.aws.amazon.com/emr/home?region=us-east-2#/clusterDetails/j-3GLAECHI85REO

Your cluster 'assign9' has been successfully created.

Amazon EMR > EMR on EC2: Clusters > assign9

assign9

Updated less than a minute ago

Summary

Cluster info

Cluster ID
j-3GLAECHI85REO

Cluster configuration
Instance groups

Capacity
1 Primary | 1 Core | 0 Task

Applications

Amazon EMR version
emr-6.12.0

Installed applications
Spark 3.4.0, Zeppelin 0.10.1

Cluster management

Log destination in Amazon S3
aws-logs-058264531906-us-east-2/elasticmapreduce

Primary node public DNS
ec2-3-137-185-137.us-east-2.compute.amazonaws.com

Connect to the Primary node using SSM

Status and time

Status
Starting

Creation time
22 March 2024 19:00 (UTC-05:00)

Elapsed time
1 minute, 29 seconds

Operating system

Amazon Linux release
2.0.20240223.0

Cluster logs

Archive log files to Amazon S3
Turned on

Amazon S3 location
s3://aws-logs-058264531906-us-east-2/elasticmapreduce/

Cluster termination and node replacement

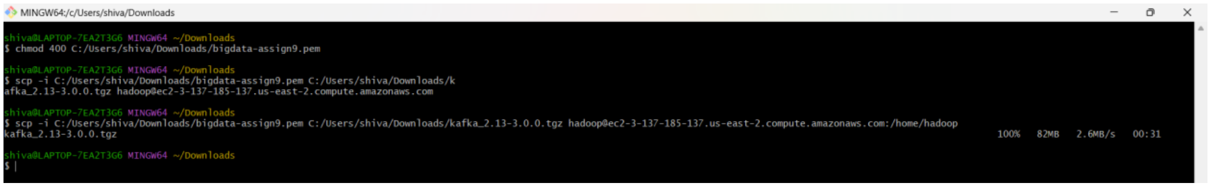
Termination option
Automatically terminate the cluster after idle time

Step B – Copy the Kafka software to the EMR master node

Download the kafka_2.13-3.0.0.tgz file from the blackboard to you PC/MAC. Use the secure copy (scp) program to move this file to the /home/hadoop directory of the master node. Here is an example of how your command line might look (yours will be somewhat different because your master node DNS name, key-pair and kafka file locations will vary)

Downloaded the file kafka_2.13-3.0.0.tgz.

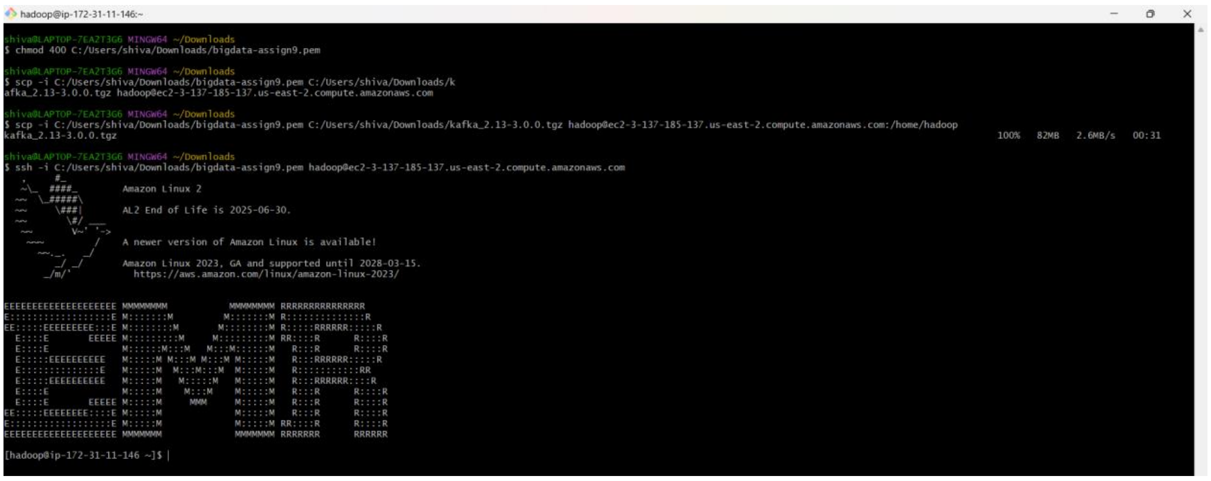
Used the secure copy (scp) program to move the file to the /home/hadoop directory of the master node.



MINGW64:/c/Users/shiva/Downloads

```
shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ chmod 400 C:/Users/shiva/Downloads/bigdata-assign9.pem

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ scp -i C:/Users/shiva/Downloads/bigdata-assign9.pem C:/Users/shiva/Downloads/k
afka_2.13-3.0.0.tgz hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ scp -i C:/Users/shiva/Downloads/bigdata-assign9.pem C:/Users/shiva/Downloads/kafka_2.13-3.0.0.tgz hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com:/home/hadoop
kafka_2.13-3.0.0.tgz                                                                                                            100%   82MB   2.6MB/s   00:31

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ |
```

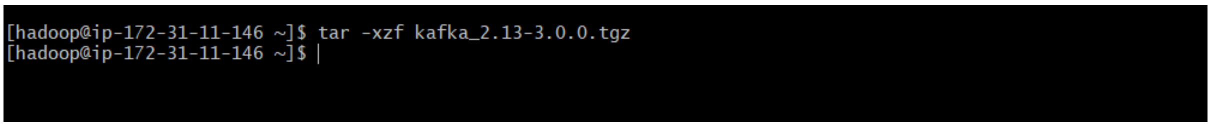
Step C – Install the Kafka software and start it



```
shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ chmod 400 C:/Users/shiva/Downloads/bigdata-assign9.pem

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ scp -i C:/Users/shiva/Downloads/bigdata-assign9.pem C:/Users/shiva/Downloads/k
afka_2.13-3.0.0.tgz hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ scp -i C:/Users/shiva/Downloads/bigdata-assign9.pem C:/Users/shiva/Downloads/kafka_2.13-3.0.0.tgz hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com:/home/hadoop
kafka_2.13-3.0.0.tgz                                                                                                                  100%   82MB   2.6MB/s   00:31

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ ssh -i C:/Users/shiva/Downloads/bigdata-assign9.pem hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com
   ,     #_
   ~\_  ####_        Amazon Linux 2
  ~~  \_#####\
  ~~     \###|       AL2 End of Life is 2025-06-30.
  ~~       \#/ ___
   ~~       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/

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::E M::::::::M       M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M     M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M   M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE   M:::::M M:::M M:::M M:::::M   R:::RRRRRR:::::R
  E::::::::::::::E   M:::::M  M:::M:::M  M:::::M   R:::::::::::RR
  E:::::EEEEEEEEEE   M:::::M   M:::::M   M:::::M   R:::RRRRRR::::R
  E::::E             M:::::M    M:::M    M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M   R:::R      R::::R
EE:::::EEEEEEEE::::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-11-146 ~]$ |
```

Enter the following command:

tar -xzf kafka_2.13-3.0.0.tgz

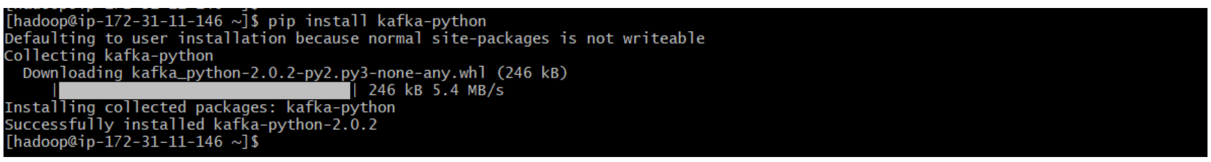


```
[hadoop@ip-172-31-11-146 ~]$ tar -xzf kafka_2.13-3.0.0.tgz
[hadoop@ip-172-31-11-146 ~]$ |
```

Note, this will create a new directory (kafka_2.13-3.0.0) holding the kakfa software release.

Then enter this command:

pip install kafka-python



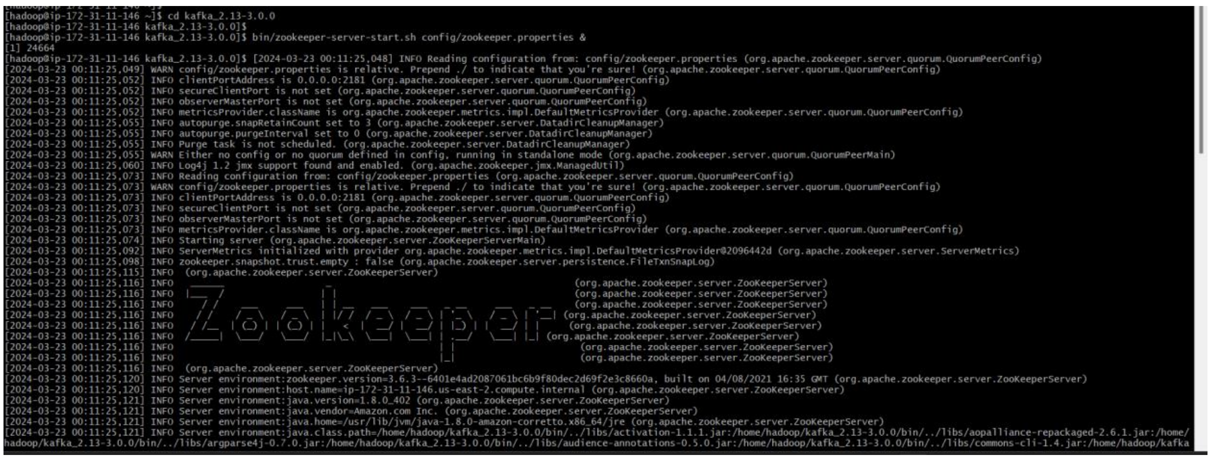
```
[hadoop@ip-172-31-11-146 ~]$ pip install kafka-python
Defaulting to user installation because normal site-packages is not writeable
Collecting kafka-python
  Downloading kafka_python-2.0.2-py2.py3-none-any.whl (246 kB)
     |████████████████████████████████| 246 kB 5.4 MB/s
Installing collected packages: kafka-python
Successfully installed kafka-python-2.0.2
[hadoop@ip-172-31-11-146 ~]$
```

This installs the kafka-python package.

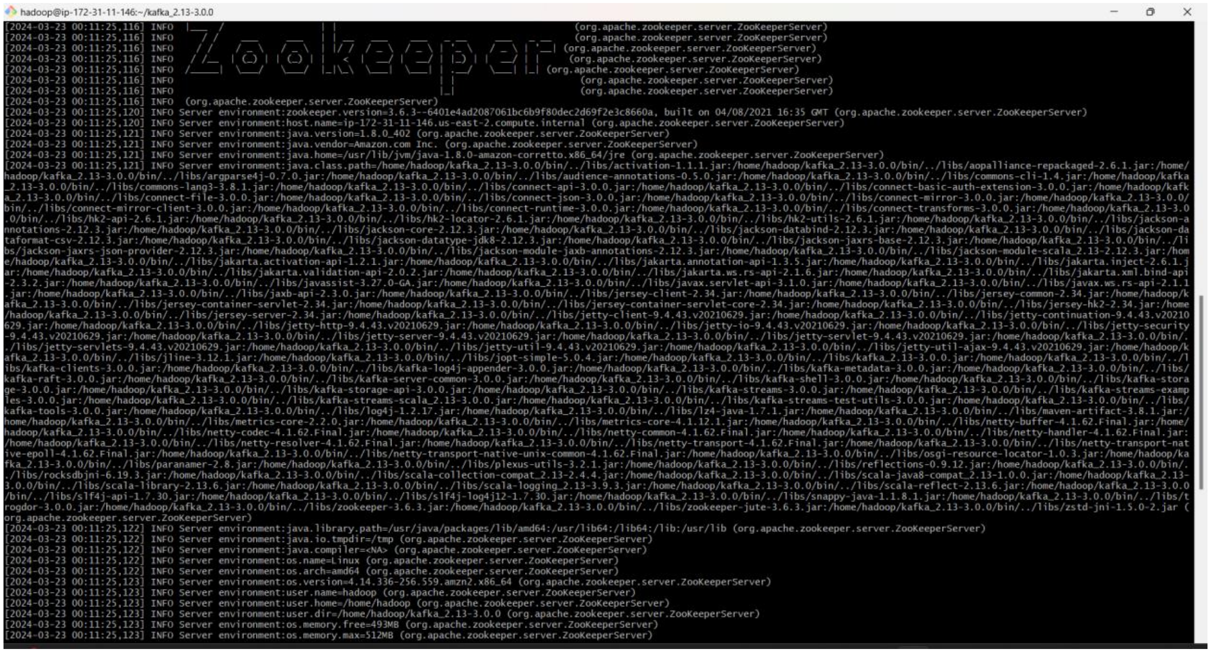
Now enter the following commands into the terminal:

cd kafka_2.13-3.0.0

bin/zookeeper-server-start.sh config/zookeeper.properties &

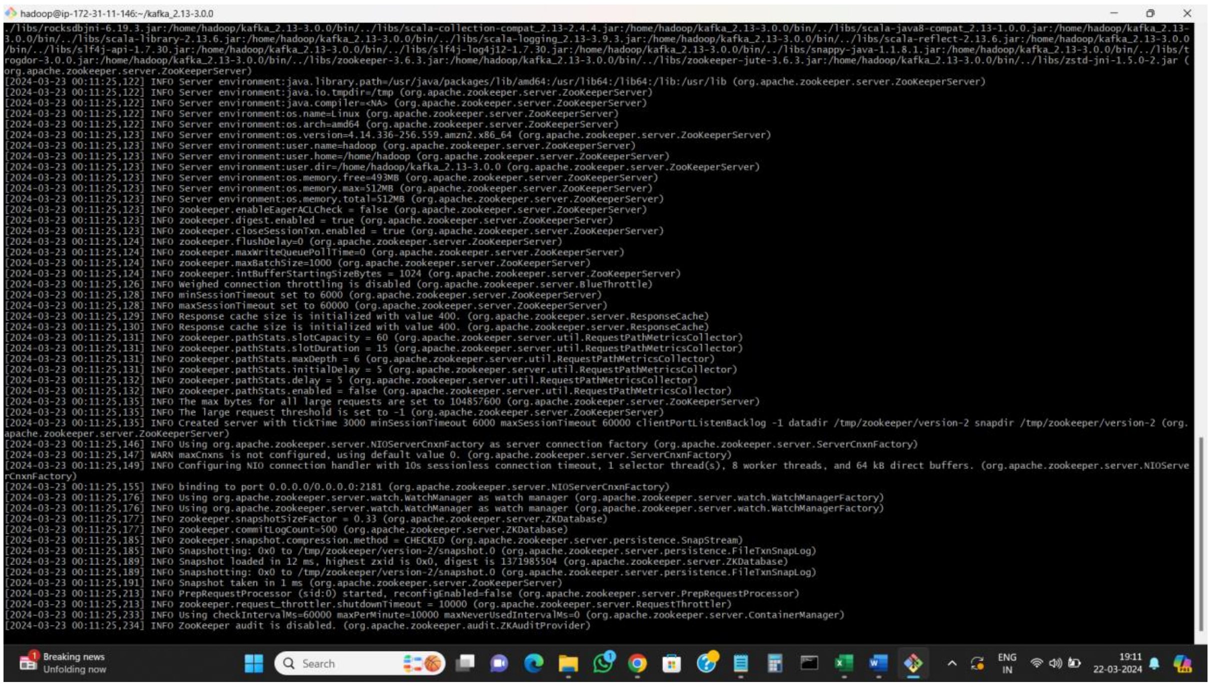


```
[hadoop@ip-172-31-11-146 ~]$
[hadoop@ip-172-31-11-146 ~]$ cd kafka_2.13-3.0.0
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ bin/zookeeper-server-start.sh config/zookeeper.properties &
[1] 24664
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ [2024-03-23 00:11:25,048] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,049] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,052] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,052] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,052] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,052] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,055] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-03-23 00:11:25,055] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-03-23 00:11:25,055] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-03-23 00:11:25,055] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2024-03-23 00:11:25,060] INFO Log4j 1.2 jmx support found and enabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2024-03-23 00:11:25,073] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,073] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,073] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,073] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,073] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,073] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-03-23 00:11:25,074] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2024-03-23 00:11:25,092] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@2096442d (org.apache.zookeeper.server.ServerMetrics)
[2024-03-23 00:11:25,098] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-03-23 00:11:25,115] INFO  (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO   ______                  _                                          (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO  |___  /                 | |                                         (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO     / /    ___     ___   | | __   ___    ___   _ __     ___   _ __   (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO    / /    / _ \   / _ \  | |/ /  / _ \  / _ \ | '_ \   / _ \ | '__| (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO   / /__  | (_) | | (_) | |   <  |  __/ |  __/ | |_) | |  __/ | |     (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO  /_____|  \___/   \___/  |_|\_\  \___|  \___| | .__/   \___| |_| (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO                                               | |                     (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO                                               |_|                     (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO  (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,120] INFO Server environment:zookeeper.version=3.6.3--6401e4ad2087061bc6b9f80dec2d69f2e3c8660a, built on 04/08/2021 16:35 GMT (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,120] INFO Server environment:host.name=ip-172-31-11-146.us-east-2.compute.internal (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.version=1.8.0_402 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.vendor=Amazon.com Inc. (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.home=/usr/lib/jvm/java-1.8.0-amazon-corretto.x86_64/jre (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.class.path=/home/hadoop/kafka_2.13-3.0.0/bin/../libs/activation-1.1.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/aopalliance-repackaged-2.6.1.jar:/home/
hadoop/kafka_2.13-3.0.0/bin/../libs/argparse4j-0.7.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/audience-annotations-0.5.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/commons-cli-1.4.jar:/home/hadoop/kafka
```



hadoop@ip-172-31-11-146:~/kafka_2.13-3.0.0

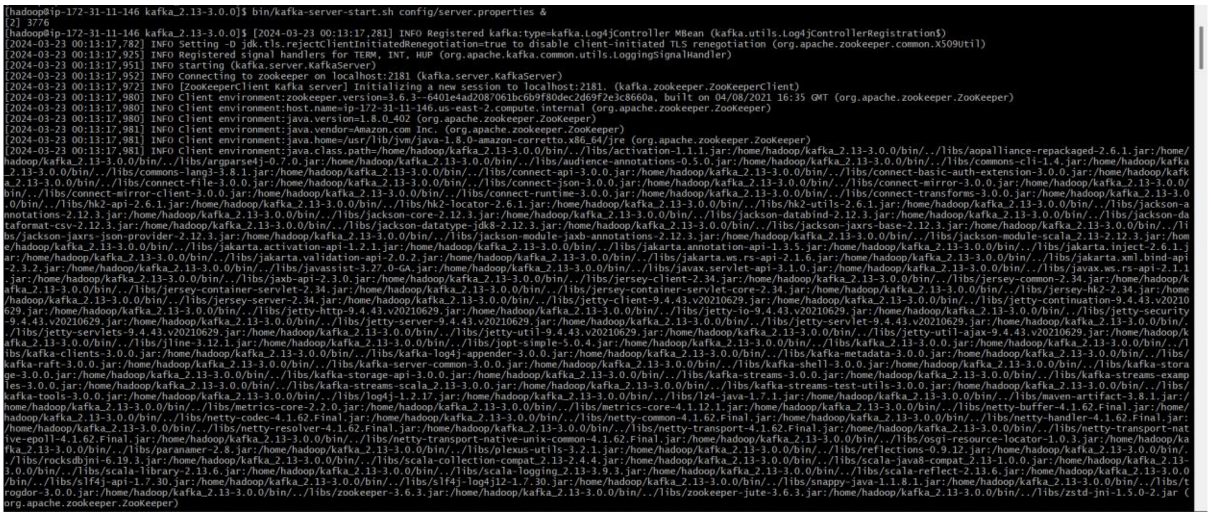
```
[2024-03-23 00:11:25,116] INFO  |___  /                 | |                                         (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO     / /    ___     ___   | | __   ___    ___   _ __     ___   _ __   (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO    / /    / _ \   / _ \  | |/ /  / _ \  / _ \ | '_ \   / _ \ | '__| (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO   / /__  | (_) | | (_) | |   <  |  __/ |  __/ | |_) | |  __/ | |     (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO  /_____|  \___/   \___/  |_|\_\  \___|  \___| | .__/   \___| |_| (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO                                               | |                     (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO                                               |_|                     (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,116] INFO  (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,120] INFO Server environment:zookeeper.version=3.6.3--6401e4ad2087061bc6b9f80dec2d69f2e3c8660a, built on 04/08/2021 16:35 GMT (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,120] INFO Server environment:host.name=ip-172-31-11-146.us-east-2.compute.internal (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.version=1.8.0_402 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.vendor=Amazon.com Inc. (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.home=/usr/lib/jvm/java-1.8.0-amazon-corretto.x86_64/jre (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,121] INFO Server environment:java.class.path=/home/hadoop/kafka_2.13-3.0.0/bin/../libs/activation-1.1.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/aopalliance-repackaged-2.6.1.jar:/home/
hadoop/kafka_2.13-3.0.0/bin/../libs/argparse4j-0.7.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/audience-annotations-0.5.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/commons-cli-1.4.jar:/home/hadoop/kafka
_2.13-3.0.0/bin/../libs/commons-lang3-3.8.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-api-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-basic-auth-extension-3.0.0.jar:/home/hadoop/kafk
a_2.13-3.0.0/bin/../libs/connect-file-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-json-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-mirror-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/
bin/../libs/connect-mirror-client-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-runtime-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-transforms-3.0.0.jar:/home/hadoop/kafka_2.13-3.0
.0/bin/../libs/hk2-api-2.6.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/hk2-locator-2.6.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/hk2-utils-2.6.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-a
nnotations-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-core-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-databind-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-da
taformat-csv-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-datatype-jdk8-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-jaxrs-base-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../li
bs/jackson-jaxrs-json-provider-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-module-jaxb-annotations-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-module-scala_2.13-2.12.3.jar:/hom
e/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.activation-api-1.2.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.annotation-api-1.3.5.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.inject-2.6.1.j
ar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.validation-api-2.0.2.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.ws.rs-api-2.1.6.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.xml.bind-api
-2.3.2.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/javassist-3.27.0-GA.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/javax.servlet-api-3.1.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/javax.ws.rs-api-2.1.1
.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jaxb-api-2.3.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-client-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-common-2.34.jar:/home/hadoop/k
afka_2.13-3.0.0/bin/../libs/jersey-container-servlet-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-container-servlet-core-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-hk2-2.34.jar:/home
/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-server-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-client-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-continuation-9.4.43.v20210
629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-http-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-io-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-security
-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-server-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-servlet-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/.
./libs/jetty-servlets-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-util-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-util-ajax-9.4.43.v20210629.jar:/home/hadoop/k
afka_2.13-3.0.0/bin/../libs/jline-3.12.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jopt-simple-5.0.4.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka_2.13-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../l
ibs/kafka-clients-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-log4j-appender-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-metadata-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/
kafka-raft-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-server-common-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-shell-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-stora
ge-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-storage-api-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-examp
les-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-scala_2.13-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-test-utils-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/
kafka-tools-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/log4j-1.2.17.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/lz4-java-1.7.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/maven-artifact-3.8.1.jar:/
home/hadoop/kafka_2.13-3.0.0/bin/../libs/metrics-core-2.2.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/metrics-core-4.1.12.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-buffer-4.1.62.Final.jar:/home/
hadoop/kafka_2.13-3.0.0/bin/../libs/netty-codec-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-common-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-handler-4.1.62.Final.jar:
/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-resolver-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-nat
ive-epoll-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-native-unix-common-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/osgi-resource-locator-1.0.3.jar:/home/hadoop/ka
fka_2.13-3.0.0/bin/../libs/paranamer-2.8.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/plexus-utils-3.2.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/reflections-0.9.12.jar:/home/hadoop/kafka_2.13-3.0.0/bin/.
./libs/rocksdbjni-6.19.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-collection-compat_2.13-2.4.4.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-java8-compat_2.13-1.0.0.jar:/home/hadoop/kafka_2.13-
3.0.0/bin/../libs/scala-library-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-logging_2.13-3.9.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-reflect-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0
/bin/../libs/slf4j-api-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/slf4j-log4j12-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/snappy-java-1.1.8.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/t
rogdor-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-jute-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zstd-jni-1.5.0-2.jar (
org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:java.io.tmpdir=/tmp (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:java.compiler=<NA> (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:os.name=Linux (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:os.arch=amd64 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.version=4.14.336-256.559.amzn2.x86_64 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:user.name=hadoop (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:user.home=/home/hadoop (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:user.dir=/home/hadoop/kafka_2.13-3.0.0 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.memory.free=493MB (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.memory.max=512MB (org.apache.zookeeper.server.ZooKeeperServer)
```



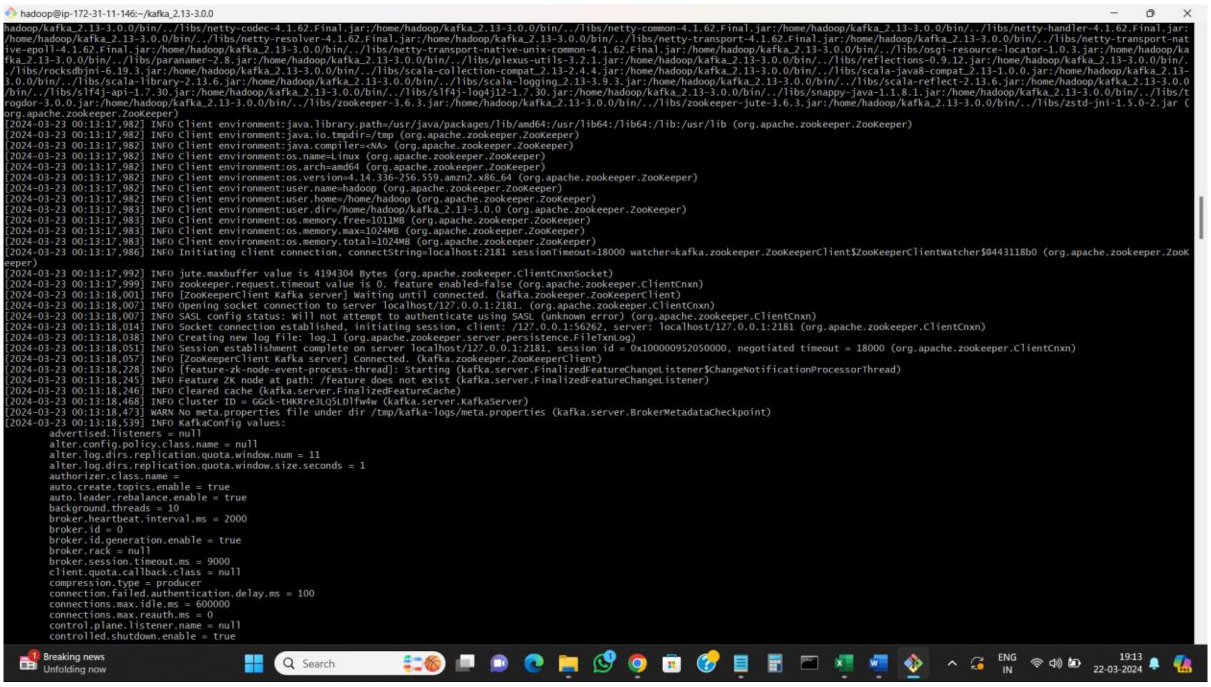
hadoop@ip-172-31-11-146:~/kafka_2.13-3.0.0

```
./libs/rocksdbjni-6.19.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-collection-compat_2.13-2.4.4.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-java8-compat_2.13-1.0.0.jar:/home/hadoop/kafka_2.13-
3.0.0/bin/../libs/scala-library-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-logging_2.13-3.9.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-reflect-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0
/bin/../libs/slf4j-api-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/slf4j-log4j12-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/snappy-java-1.1.8.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/t
rogdor-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-jute-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zstd-jni-1.5.0-2.jar (
org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:java.io.tmpdir=/tmp (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:java.compiler=<NA> (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:os.name=Linux (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,122] INFO Server environment:os.arch=amd64 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.version=4.14.336-256.559.amzn2.x86_64 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:user.name=hadoop (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:user.home=/home/hadoop (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:user.dir=/home/hadoop/kafka_2.13-3.0.0 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.memory.free=493MB (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.memory.max=512MB (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO Server environment:os.memory.total=512MB (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO zookeeper.enableEagerACLCheck = false (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO zookeeper.digest.enabled = true (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,123] INFO zookeeper.closeSessionTxn.enabled = true (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,124] INFO zookeeper.flushDelay=0 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,124] INFO zookeeper.maxWriteQueuePollTime=0 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,124] INFO zookeeper.maxBatchSize=1000 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,124] INFO zookeeper.intBufferStartingSizeBytes = 1024 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,126] INFO Weighed connection throttling is disabled (org.apache.zookeeper.server.BlueThrottle)
[2024-03-23 00:11:25,128] INFO minSessionTimeout set to 6000 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,128] INFO maxSessionTimeout set to 60000 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,129] INFO Response cache size is initialized with value 400. (org.apache.zookeeper.server.ResponseCache)
[2024-03-23 00:11:25,130] INFO Response cache size is initialized with value 400. (org.apache.zookeeper.server.ResponseCache)
[2024-03-23 00:11:25,131] INFO zookeeper.pathStats.slotCapacity = 60 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-03-23 00:11:25,131] INFO zookeeper.pathStats.slotDuration = 15 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-03-23 00:11:25,131] INFO zookeeper.pathStats.maxDepth = 6 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-03-23 00:11:25,131] INFO zookeeper.pathStats.initialDelay = 5 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-03-23 00:11:25,132] INFO zookeeper.pathStats.delay = 5 (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-03-23 00:11:25,132] INFO zookeeper.pathStats.enabled = false (org.apache.zookeeper.server.util.RequestPathMetricsCollector)
[2024-03-23 00:11:25,135] INFO The max bytes for all large requests are set to 104857600 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,135] INFO The large request threshold is set to -1 (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,135] INFO Created server with tickTime 3000 minSessionTimeout 6000 maxSessionTimeout 60000 clientPortListenBacklog -1 datadir /tmp/zookeeper/version-2 snapdir /tmp/zookeeper/version-2 (org.
apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,146] INFO Using org.apache.zookeeper.server.NIOServerCnxnFactory as server connection factory (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-03-23 00:11:25,147] WARN maxCnxns is not configured, using default value 0. (org.apache.zookeeper.server.ServerCnxnFactory)
[2024-03-23 00:11:25,149] INFO Configuring NIO connection handler with 10s sessionless connection timeout, 1 selector thread(s), 8 worker threads, and 64 kB direct buffers. (org.apache.zookeeper.server.NIOServe
rCnxnFactory)
[2024-03-23 00:11:25,155] INFO binding to port 0.0.0.0/0.0.0.0:2181 (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2024-03-23 00:11:25,176] INFO Using org.apache.zookeeper.server.watch.WatchManager as watch manager (org.apache.zookeeper.server.watch.WatchManagerFactory)
[2024-03-23 00:11:25,176] INFO Using org.apache.zookeeper.server.watch.WatchManager as watch manager (org.apache.zookeeper.server.watch.WatchManagerFactory)
[2024-03-23 00:11:25,177] INFO zookeeper.snapshotSizeFactor = 0.33 (org.apache.zookeeper.server.ZKDatabase)
[2024-03-23 00:11:25,177] INFO zookeeper.commitLogCount=500 (org.apache.zookeeper.server.ZKDatabase)
[2024-03-23 00:11:25,185] INFO zookeeper.snapshot.compression.method = CHECKED (org.apache.zookeeper.server.persistence.SnapStream)
[2024-03-23 00:11:25,185] INFO Snapshotting: 0x0 to /tmp/zookeeper/version-2/snapshot.0 (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-03-23 00:11:25,189] INFO Snapshot loaded in 12 ms, highest zxid is 0x0, digest is 1371985504 (org.apache.zookeeper.server.ZKDatabase)
[2024-03-23 00:11:25,189] INFO Snapshotting: 0x0 to /tmp/zookeeper/version-2/snapshot.0 (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-03-23 00:11:25,191] INFO Snapshot taken in 1 ms (org.apache.zookeeper.server.ZooKeeperServer)
[2024-03-23 00:11:25,213] INFO PrepRequestProcessor (sid:0) started, reconfigEnabled=false (org.apache.zookeeper.server.PrepRequestProcessor)
[2024-03-23 00:11:25,213] INFO zookeeper.request_throttler.shutdownTimeout = 10000 (org.apache.zookeeper.server.RequestThrottler)
[2024-03-23 00:11:25,233] INFO Using checkIntervalMs=60000 maxPerMinute=10000 maxNeverUsedIntervalMs=0 (org.apache.zookeeper.server.ContainerManager)
[2024-03-23 00:11:25,234] INFO ZooKeeper audit is disabled. (org.apache.zookeeper.audit.ZKAuditProvider)
```

bin/kafka-server-start.sh config/server.properties &



```
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ bin/kafka-server-start.sh config/server.properties &
[2] 3776
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ [2024-03-23 00:13:17,281] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2024-03-23 00:13:17,782] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2024-03-23 00:13:17,925] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2024-03-23 00:13:17,951] INFO starting (kafka.server.KafkaServer)
[2024-03-23 00:13:17,952] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2024-03-23 00:13:17,972] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2024-03-23 00:13:17,980] INFO Client environment:zookeeper.version=3.6.3--6401e4ad2087061bc6b9f80dec2d69f2e3c8660a, built on 04/08/2021 16:35 GMT (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,980] INFO Client environment:host.name=ip-172-31-11-146.us-east-2.compute.internal (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,980] INFO Client environment:java.version=1.8.0_402 (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,981] INFO Client environment:java.vendor=Amazon.com Inc. (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,981] INFO Client environment:java.home=/usr/lib/jvm/java-1.8.0-amazon-corretto.x86_64/jre (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,981] INFO Client environment:java.class.path=/home/hadoop/kafka_2.13-3.0.0/bin/../libs/activation-1.1.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/aopalliance-repackaged-2.6.1.jar:/home/
hadoop/kafka_2.13-3.0.0/bin/../libs/argparse4j-0.7.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/audience-annotations-0.5.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/commons-cli-1.4.jar:/home/hadoop/kafka
_2.13-3.0.0/bin/../libs/commons-lang3-3.8.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-api-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-basic-auth-extension-3.0.0.jar:/home/hadoop/kafk
a_2.13-3.0.0/bin/../libs/connect-file-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-json-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-mirror-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/
bin/../libs/connect-mirror-client-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-runtime-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/connect-transforms-3.0.0.jar:/home/hadoop/kafka_2.13-3.0
.0/bin/../libs/hk2-api-2.6.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/hk2-locator-2.6.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/hk2-utils-2.6.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-a
nnotations-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-core-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-databind-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-da
taformat-csv-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-datatype-jdk8-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-jaxrs-base-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../li
bs/jackson-jaxrs-json-provider-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-module-jaxb-annotations-2.12.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jackson-module-scala_2.13-2.12.3.jar:/hom
e/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.activation-api-1.2.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.annotation-api-1.3.5.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.inject-2.6.1.j
ar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.validation-api-2.0.2.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.ws.rs-api-2.1.6.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jakarta.xml.bind-api
-2.3.2.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/javassist-3.27.0-GA.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/javax.servlet-api-3.1.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/javax.ws.rs-api-2.1.1
.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jaxb-api-2.3.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-client-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-common-2.34.jar:/home/hadoop/k
afka_2.13-3.0.0/bin/../libs/jersey-container-servlet-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-container-servlet-core-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-hk2-2.34.jar:/home
/hadoop/kafka_2.13-3.0.0/bin/../libs/jersey-server-2.34.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-client-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-continuation-9.4.43.v20210
629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-http-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-io-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-security
-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-server-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-servlet-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/.
./libs/jetty-servlets-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-util-9.4.43.v20210629.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jetty-util-ajax-9.4.43.v20210629.jar:/home/hadoop/k
afka_2.13-3.0.0/bin/../libs/jline-3.12.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/jopt-simple-5.0.4.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka_2.13-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../l
ibs/kafka-clients-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-log4j-appender-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-metadata-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/
kafka-raft-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-server-common-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-shell-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-stora
ge-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-storage-api-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-examp
les-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-scala_2.13-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/kafka-streams-test-utils-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/
kafka-tools-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/log4j-1.2.17.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/lz4-java-1.7.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/maven-artifact-3.8.1.jar:/
home/hadoop/kafka_2.13-3.0.0/bin/../libs/metrics-core-2.2.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/metrics-core-4.1.12.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-buffer-4.1.62.Final.jar:/home/
hadoop/kafka_2.13-3.0.0/bin/../libs/netty-codec-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-common-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-handler-4.1.62.Final.jar:
/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-resolver-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-nat
ive-epoll-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-native-unix-common-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/osgi-resource-locator-1.0.3.jar:/home/hadoop/ka
fka_2.13-3.0.0/bin/../libs/paranamer-2.8.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/plexus-utils-3.2.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/reflections-0.9.12.jar:/home/hadoop/kafka_2.13-3.0.0/bin/.
./libs/rocksdbjni-6.19.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-collection-compat_2.13-2.4.4.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-java8-compat_2.13-1.0.0.jar:/home/hadoop/kafka_2.13-
3.0.0/bin/../libs/scala-library-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-logging_2.13-3.9.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-reflect-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0
/bin/../libs/slf4j-api-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/slf4j-log4j12-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/snappy-java-1.1.8.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/t
rogdor-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-jute-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zstd-jni-1.5.0-2.jar (
org.apache.zookeeper.ZooKeeper)
```

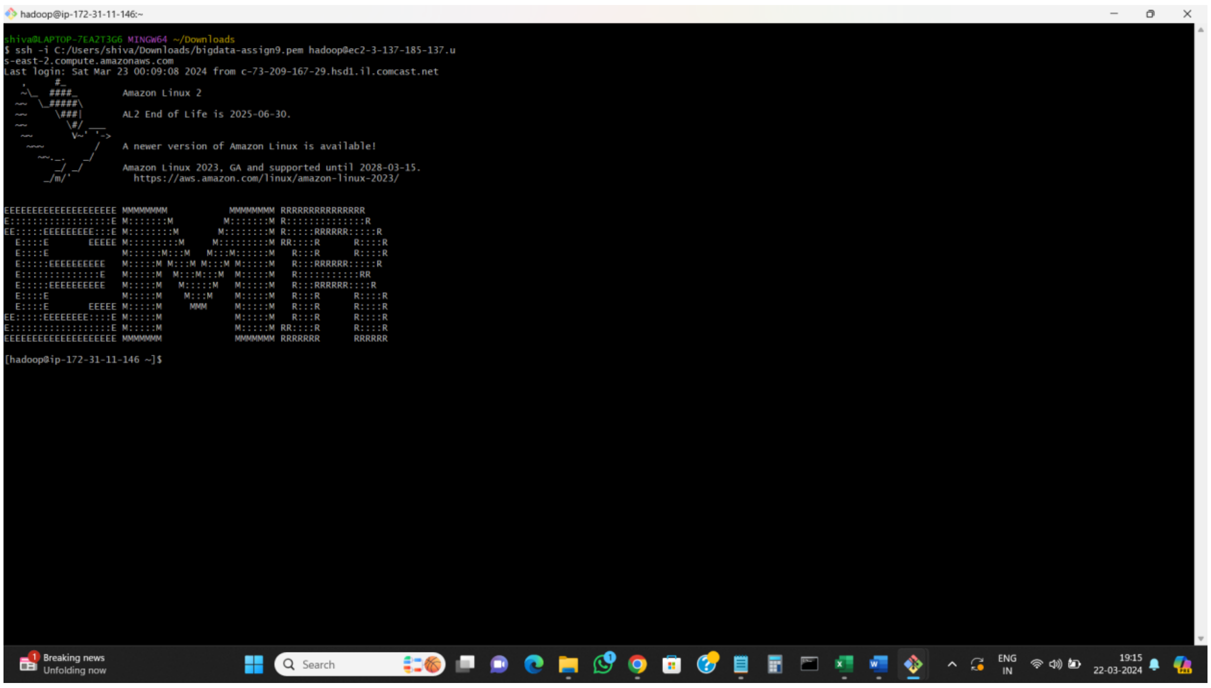


```
hadoop/kafka_2.13-3.0.0/bin/../libs/netty-codec-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-common-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-handler-4.1.62.Final.jar:
/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-resolver-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-nat
ive-epoll-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/netty-transport-native-unix-common-4.1.62.Final.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/osgi-resource-locator-1.0.3.jar:/home/hadoop/ka
fka_2.13-3.0.0/bin/../libs/paranamer-2.8.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/plexus-utils-3.2.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/reflections-0.9.12.jar:/home/hadoop/kafka_2.13-3.0.0/bin/.
./libs/rocksdbjni-6.19.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-collection-compat_2.13-2.4.4.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-java8-compat_2.13-1.0.0.jar:/home/hadoop/kafka_2.13-
3.0.0/bin/../libs/scala-library-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-logging_2.13-3.9.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/scala-reflect-2.13.6.jar:/home/hadoop/kafka_2.13-3.0.0
/bin/../libs/slf4j-api-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/slf4j-log4j12-1.7.30.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/snappy-java-1.1.8.1.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/t
rogdor-3.0.0.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zookeeper-jute-3.6.3.jar:/home/hadoop/kafka_2.13-3.0.0/bin/../libs/zstd-jni-1.5.0-2.jar (
org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:java.io.tmpdir=/tmp (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:java.compiler=<NA> (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:os.name=Linux (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:os.arch=amd64 (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:os.version=4.14.336-256.559.amzn2.x86_64 (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:user.name=hadoop (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,982] INFO Client environment:user.home=/home/hadoop (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,983] INFO Client environment:user.dir=/home/hadoop/kafka_2.13-3.0.0 (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,983] INFO Client environment:os.memory.free=1011MB (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,983] INFO Client environment:os.memory.max=1024MB (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,983] INFO Client environment:os.memory.total=1024MB (org.apache.zookeeper.ZooKeeper)
[2024-03-23 00:13:17,986] INFO Initiating client connection, connectString=localhost:2181 sessionTimeout=18000 watcher=kafka.zookeeper.ZooKeeperClient$ZooKeeperClientWatcher$@443118b0 (org.apache.zookeeper.ZooK
eeper)
[2024-03-23 00:13:17,992] INFO jute.maxbuffer value is 4194304 Bytes (org.apache.zookeeper.ClientCnxnSocket)
[2024-03-23 00:13:17,999] INFO zookeeper.request.timeout value is 0. feature enabled=false (org.apache.zookeeper.ClientCnxn)
[2024-03-23 00:13:18,001] INFO [ZooKeeperClient Kafka server] Waiting until connected. (kafka.zookeeper.ZooKeeperClient)
[2024-03-23 00:13:18,007] INFO Opening socket connection to server localhost/127.0.0.1:2181. (org.apache.zookeeper.ClientCnxn)
[2024-03-23 00:13:18,007] INFO SASL config status: Will not attempt to authenticate using SASL (unknown error) (org.apache.zookeeper.ClientCnxn)
[2024-03-23 00:13:18,014] INFO Socket connection established, initiating session, client: /127.0.0.1:56262, server: localhost/127.0.0.1:2181 (org.apache.zookeeper.ClientCnxn)
[2024-03-23 00:13:18,038] INFO Creating new log file: log.1 (org.apache.zookeeper.server.persistence.FileTxnLog)
[2024-03-23 00:13:18,051] INFO Session establishment complete on server localhost/127.0.0.1:2181, session id = 0x100000952050000, negotiated timeout = 18000 (org.apache.zookeeper.ClientCnxn)
[2024-03-23 00:13:18,057] INFO [ZooKeeperClient Kafka server] Connected. (kafka.zookeeper.ZooKeeperClient)
[2024-03-23 00:13:18,228] INFO [feature-zk-node-event-process-thread]: Starting (kafka.server.FinalizedFeatureChangeListener$ChangeNotificationProcessorThread)
[2024-03-23 00:13:18,245] INFO Feature ZK node at path: /feature does not exist (kafka.server.FinalizedFeatureChangeListener)
[2024-03-23 00:13:18,246] INFO Cleared cache (kafka.server.FinalizedFeatureCache)
[2024-03-23 00:13:18,468] INFO Cluster ID = GGck-tHKRreJLQ5LDlfw4w (kafka.server.KafkaServer)
[2024-03-23 00:13:18,473] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties (kafka.server.BrokerMetadataCheckpoint)
[2024-03-23 00:13:18,539] INFO KafkaConfig values:
        advertised.listeners = null
        alter.config.policy.class.name = null
        alter.log.dirs.replication.quota.window.num = 11
        alter.log.dirs.replication.quota.window.size.seconds = 1
        authorizer.class.name =
        auto.create.topics.enable = true
        auto.leader.rebalance.enable = true
        background.threads = 10
        broker.heartbeat.interval.ms = 2000
        broker.id = 0
        broker.id.generation.enable = true
        broker.rack = null
        broker.session.timeout.ms = 9000
        client.quota.callback.class = null
        compression.type = producer
        connection.failed.authentication.delay.ms = 100
        connections.max.idle.ms = 600000
        connections.max.reauth.ms = 0
        control.plane.listener.name = null
        controlled.shutdown.enable = true
```

Step D – Prepare to run Kafka producers and consumers

Open a second terminal connection to the EMR master node. Going forward we will call this terminal connection: Producer-Term.

Open a third terminal connection to the EMR master node. Going forward we will call this terminal connection: Consumer-Term.



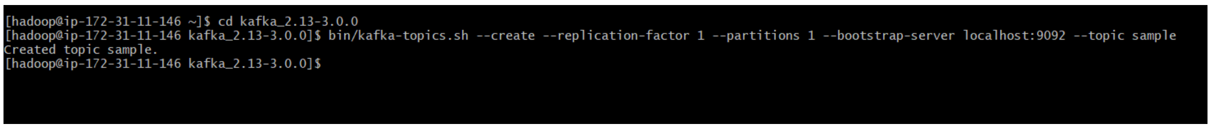
```
shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ ssh -i C:/Users/shiva/Downloads/bigdata-assign9.pem hadoop@ec2-3-137-185-137.u
s-east-2.compute.amazonaws.com
Last login: Sat Mar 23 00:09:08 2024 from c-73-209-167-29.hsd1.il.comcast.net
   ,     #_
   ~\_  ####_        Amazon Linux 2
  ~~  \_#####\
  ~~     \###|       AL2 End of Life is 2025-06-30.
  ~~       \#/ ___
   ~~       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/

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::E M::::::::M       M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M     M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M   M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE   M:::::M M:::M M:::M M:::::M   R:::RRRRRR:::::R
  E::::::::::::::E   M:::::M  M:::M:::M  M:::::M   R:::::::::::RR
  E:::::EEEEEEEEEE   M:::::M   M:::::M   M:::::M   R:::RRRRRR::::R
  E::::E             M:::::M    M:::M    M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M   R:::R      R::::R
EE:::::EEEEEEEE::::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-11-146 ~]$
```

Step E – Create a Kafka topic

In the Producer-Term, enter the following command:

cd kafka_2.13-3.0.0

bin/kafka-topics.sh --create --replication-factor 1 --partitions 1 --bootstrap-server localhost:9092 --topic sample

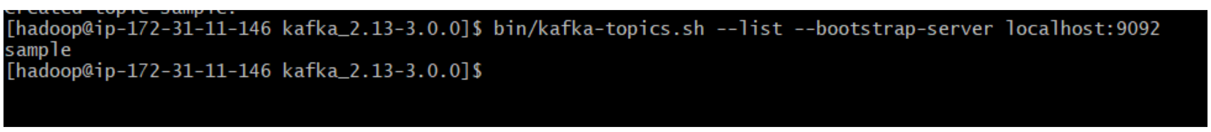


```
[hadoop@ip-172-31-11-146 ~]$ cd kafka_2.13-3.0.0
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ bin/kafka-topics.sh --create --replication-factor 1 --partitions 1 --bootstrap-server localhost:9092 --topic sample
Created topic sample.
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$
```

Here we create a new kafka topic called ‘sample’. You can use this command to create a topic with any name you like. Try creating a few more topics.

To list the topics that you created you can enter the following into the Producer-Term (note some default topics already exist):

bin/kafka-topics.sh --list --bootstrap-server localhost:9092



```
Created topic sample.
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ bin/kafka-topics.sh --list --bootstrap-server localhost:9092
sample
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$
```

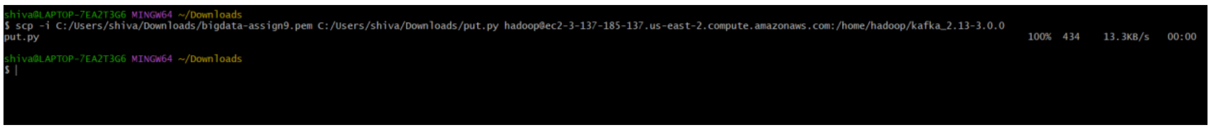
a)

In the Producer-Term (or some other way) write a small program, call it ‘put.py’, using the vi text or some other way of putting a python program onto the EMR master node. If you like you could use a text editor on your PC/MAC to write the program and then scp it over to your EMR master name.

This program should implement a kafka producer that writes three messages to the topic ‘sample’. Recall that you need to convert values and keys to type bytes. The three messages should have keys and values as follows:

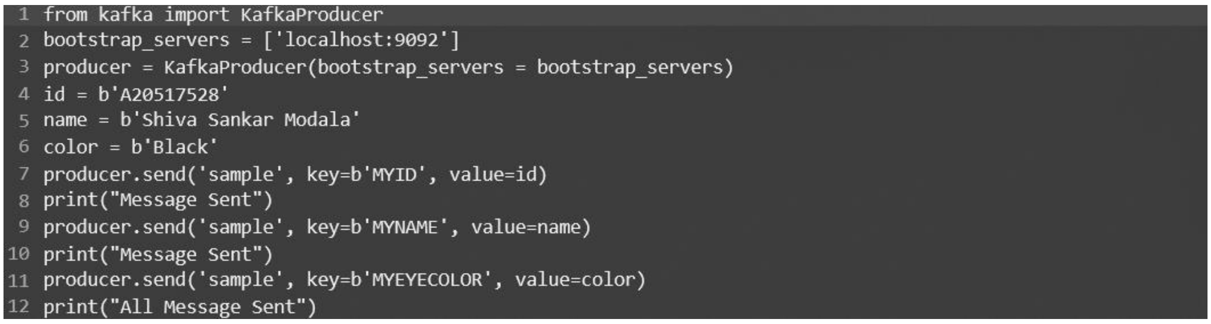
Key	Value
‘MYID’	Your student id
‘MYNAME’	Your name
‘MYEYECOLOR’	Your eye color (make it up if you can’t remember)

Execute this program in the Producer-Term, use the command line (you might need to provide a full pathname depending on where your python program is such as /home/hadoop/someplace/put.py):

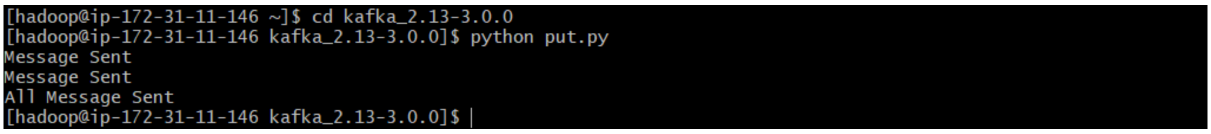


```
shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ scp -i C:/Users/shiva/Downloads/bigdata-assign9.pem C:/Users/shiva/Downloads/put.py hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com:/home/hadoop/kafka_2.13-3.0.0
put.py                                                                                          100%  434    13.3KB/s   00:00

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ |
```

python put.py



```
from kafka import KafkaProducer
bootstrap_servers = ['localhost:9092']
producer = KafkaProducer(bootstrap_servers = bootstrap_servers)
id = b'A20517528'
name = b'Shiva Sankar Modala'
color = b'Black'
producer.send('sample', key=b'MYID', value=id)
print("Message Sent")
producer.send('sample', key=b'MYNAME', value=name)
print("Message Sent")
producer.send('sample', key=b'MYEYECOLOR', value=color)
print("All Message Sent")
```

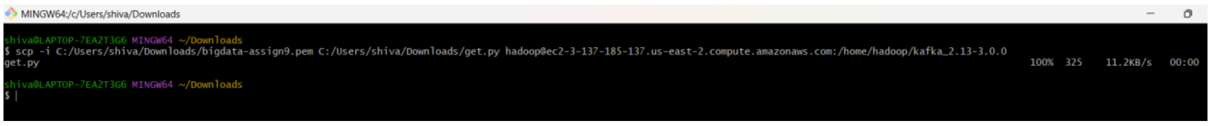


```
[hadoop@ip-172-31-11-146 ~]$ cd kafka_2.13-3.0.0
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ python put.py
Message Sent
Message Sent
All Message Sent
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ |
```

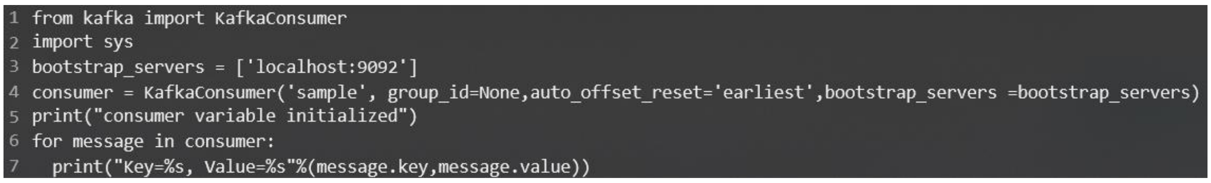
b)

In the Consumer-Term, write another small program, call it ‘get.py’, using the vi text or some other way of putting a python program onto the EMR master node.

This program should implement a kafka consumer that reads the messages you wrote previously from the topic ‘sample’ and writes them to the terminal.



```
MINGW64:/c/Users/shiva/Downloads

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ scp -i C:/Users/shiva/Downloads/bigdata-assign9.pem C:/Users/shiva/Downloads/get.py hadoop@ec2-3-137-185-137.us-east-2.compute.amazonaws.com:/home/hadoop/kafka_2.13-3.0.0
get.py                                                                                          100%  325    11.2KB/s   00:00

shiva@LAPTOP-7EA2T3G6 MINGW64 ~/Downloads
$ |
```



```
from kafka import KafkaConsumer
import sys
bootstrap_servers = ['localhost:9092']
consumer = KafkaConsumer('sample', group_id=None,auto_offset_reset='earliest',bootstrap_servers =bootstrap_servers)
print("consumer variable initialized")
for message in consumer:
  print("Key=%s, Value=%s"%(message.key,message.value))
```

The output should look something like this:

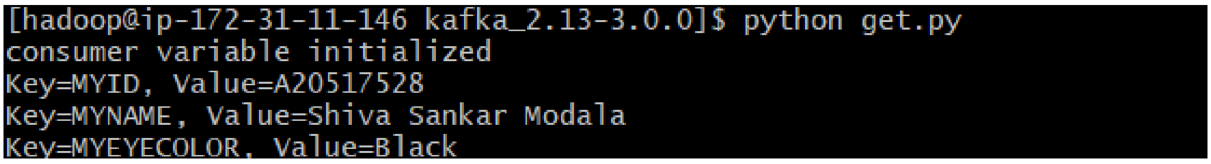
Key=MYID, Value=’your id number’

Key=MYNAME, Value=’your name’

Key=MYEYECOLOR, Value=’your eye color’

Execute this program in the Consumer-Term. Use the command line:

python get.py



```
[hadoop@ip-172-31-11-146 kafka_2.13-3.0.0]$ python get.py
consumer variable initialized
Key=MYID, Value=A20517528
Key=MYNAME, Value=Shiva Sankar Modala
Key=MYEYECOLOR, Value=Black
```