



Department of Computer Science and Engineering (Data Science)

Rishabh Singhvi

60009210206

CSE DS D2

Subject: Big Data Engineering (DJ19DSL604)

AY: 2022-23

Experiment 4

(Messaging Service)

Aim: Implement messaging system using Kafka.

Theory:

Kafka Overview

Apache Kafka is a distributed publish-subscribe messaging system and a robust queue that can handle a high volume of data and enables you to pass messages from one end-point to another. Kafka is suitable for both offline and online message consumption. Kafka messages are persisted on the disk and replicated within the cluster to prevent data loss. Kafka is built on top of the ZooKeeper synchronization service. It integrates very well with Apache Storm and Spark for real-time streaming data analysis.

Need for Kafka

Kafka is a unified platform for handling all the real-time data feeds. Kafka supports low latency message delivery and gives guarantee for fault tolerance in the presence of machine failures. It has the ability to handle a large number of diverse consumers. Kafka is very fast, performs 2 million writes/sec. Kafka persists all data to the disk, which essentially means that all the writes go to the page cache of the OS (RAM). This makes it very efficient to transfer data from page cache to a network socket.

Kafka Cluster Architecture

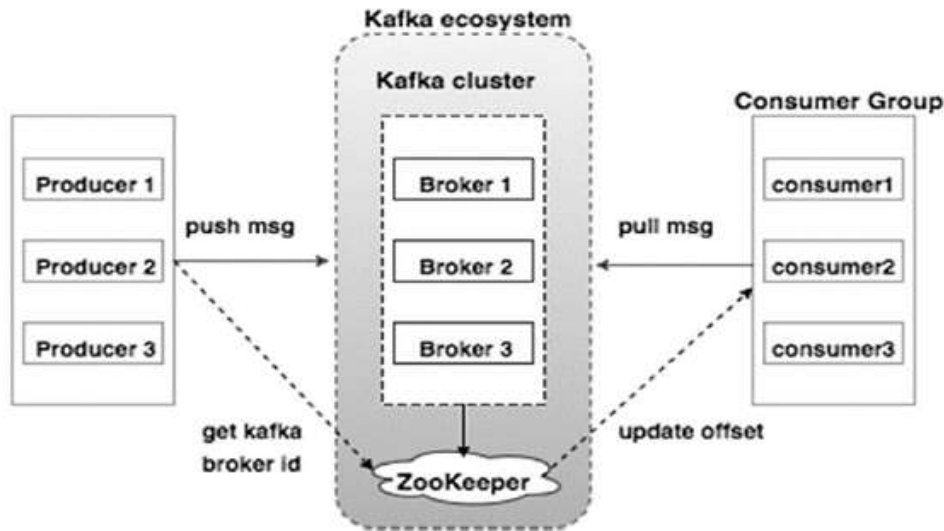


Department of Computer Science and Engineering (Data Science)

Rishabh Singhvi

60009210206

CSE DS D2



Broker

Kafka cluster typically consists of multiple brokers to maintain load balance. Kafka brokers are stateless, so they use ZooKeeper for maintaining their cluster state. One Kafka broker instance can handle hundreds of thousands of reads and writes per second and each broker can handle TB of messages without performance impact. Kafka broker leader election can be done by ZooKeeper.

ZooKeeper

ZooKeeper is used for managing and coordinating Kafka broker. ZooKeeper service is mainly used to notify producer and consumer about the presence of any new broker in the Kafka system or failure of the broker in the Kafka system. As per the notification received by the Zookeeper regarding presence or failure of the broker then producer and consumer takes decision and starts coordinating their task with some other broker.

Producers

Producers push data to brokers. When the new broker is started, all the producers search it and automatically sends a message to that new broker. Kafka producer doesn't wait for acknowledgements from the broker and sends messages as fast as the broker can handle.



Department of Computer Science and Engineering (Data Science)

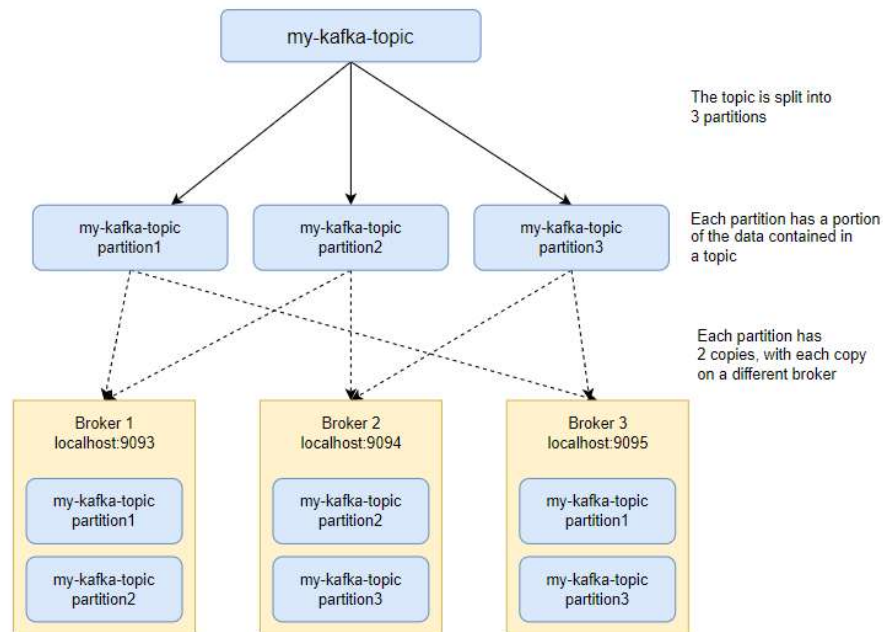
Rishabh Singhvi

60009210206

CSE DS D2

Consumers

Since Kafka brokers are stateless, which means that the consumer has to maintain how many messages have been consumed by using partition offset. If the consumer acknowledges a particular message offset, it implies that the consumer has consumed all prior messages. The consumer issues an asynchronous pull request to the broker to have a buffer of bytes ready to consume. The consumers can rewind or skip to any point in a partition simply by supplying an offset value. Consumer offset value is notified by ZooKeeper.



Lab Assignment:

1. Installation of Kafka 2.13-3.0.0 locally.
2. Create a Kafka local cluster with 3 brokers and create a topic to which the data should belong.
3. Send message to a topic using producer and read the data from the cluster using a consumer.
4. Testing Replication after a failed broker.



Department of Computer Science and Engineering (Data Science)

Rishabh Singhvi

60009210206

CSE DS D2

```
*C:\kafka\config\server.properties - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

46 # The number of threads that the server uses for processing requests, which may include disk I/O
47 num.io.threads=8
48
49 # The send buffer (SO_SNDBUF) used by the socket server
50 socket.send.buffer.bytes=102400
51
52 # The receive buffer (SO_RCVBUF) used by the socket server
53 socket.receive.buffer.bytes=102400
54
55 # The maximum size of a request that the socket server will accept (protection against OOM)
56 socket.request.max.bytes=104857600
57
58 ##### Log Basics #####
59
60 # A comma separated list of directories under which to store log files
61 log.dirs=c:/kafka/kafka-logs
62
63 # The default number of partitions per topic. More partitions allow greater
64 # parallelism for consumption, but this will also result in more files across
65 # the brokers.
66 num.partitions=1
67
68 # The number of threads per data directory to be used for log recovery at startup and flushing at shutdown.
69 # This value is recommended to be increased for installations with data dirs located in RAID array.
70 num.recovery.threads.per.data.dir=1
71
72 ##### Internal Topic Settings #####
73 # The replication factor for the group metadata internal topics "__consumer_offsets" and "__transaction_state"
74 # For anything other than development testing, a value greater than 1 is recommended to ensure availability such as 3.
75 offsets.topic.replication.factor=1
76 transaction.state.log.replication.factor=1
77 transaction.state.log.min.isr=1
78
79 ##### Log Flush Policy #####
80
81 # Licensed to the Apache Software Foundation (ASF) under one or more
82 # contributor license agreements. See the NOTICE file distributed with
83 # this work for additional information regarding copyright ownership.
84 # The ASF licenses this file to You under the Apache License, Version 2.0
85 # (the "License"); you may not use this file except in compliance with
86 # the License. You may obtain a copy of the License at
87 #
88 # http://www.apache.org/licenses/LICENSE-2.0
89 #
90 # Unless required by applicable law or agreed to in writing, software
91 # distributed under the License is distributed on an "AS IS" BASIS,
92 # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
93 # See the License for the specific language governing permissions and
94 # limitations under the License.
95 # the directory where the snapshot is stored.
96 dataDir=c:/kafka/zookeeper-data
97 # the port at which the clients will connect
98 clientPort=2181
99 # disable the per-ip limit on the number of connections since this is a non-production config
100 maxClientCnxns=0
101 # Disable the adminserver by default to avoid port conflicts.
102 # Set the port to something non-conflicting if choosing to enable this
103 admin.enableServer=false
104 # admin.serverPort=8080
105
```




Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

Rishabh Singhvi

60009210206

CSE DS D2

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.1889]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
```

```
C:\Windows\System32\cmd.exe - .\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
[2022-09-09 00:57:07,627] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-09-09 00:57:07,627] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-09-09 00:57:07,628] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-09-09 00:57:07,628] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-09-09 00:57:07,628] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2022-09-09 00:57:07,643] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@6bf256fa (org.apache.zookeeper.server.ServerMetrics)
[2022-09-09 00:57:07,647] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnLog)
[2022-09-09 00:57:07,670] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,670] INFO _____ (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,671] INFO |__ / | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,674] INFO / / _ _ | | _ _ _ _ _ _ (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,675] INFO / / / _ \ / _ \ | | / / _ \ / _ \ | ' _ | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,675] INFO / / _ | ( ) | | ( ) | | < | _ / | _ / | | _ / | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,676] INFO / _ _ | \ _ \ | | \ _ \ | \ _ \ | _ \ | _ | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,676] INFO | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-09-09 00:57:07,677] INFO | _ | (org.apache.zookeeper.server.ZooKeeperServer)
```



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

Rishabh Singhvi

60009210206

CSE DS D2

```
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-server-start.bat .\config\server.properties
Microsoft Windows [Version 10.0.19044.1889]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>. \bin\windows\kafka-server-start.bat .\config\server.properties
[2022-09-09 00:59:32,682] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2022-09-09 00:59:33,070] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.XS9Util)
[2022-09-09 00:59:33,201] INFO starting (kafka.server.KafkaServer)
[2022-09-09 00:59:33,202] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2022-09-09 00:59:33,210] INFO [ZooKeeperClient Kafka server-] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
.
```

```
[2022-09-09 00:59:43,930] INFO [GroupCoordinator 0]: Starting up. (kafka.coordinator.group.GroupCoordinator)
[2022-09-09 00:59:43,936] INFO [GroupCoordinator 0]: Startup complete. (kafka.coordinator.group.GroupCoordinator)
[2022-09-09 00:59:43,939] INFO Feature ZK node created at path: /feature (kafka.server.FinalizedFeatureChangeListener)
[2022-09-09 00:59:43,961] INFO [TransactionCoordinator id=0] Starting up. (kafka.coordinator.transaction.TransactionCoordinator)
[2022-09-09 00:59:43,977] INFO [Transaction Marker Channel Manager 0]: Starting (kafka.coordinator.transaction.TransactionMarkerChannelManager)
[2022-09-09 00:59:43,978] INFO [TransactionCoordinator id=0] Startup complete. (kafka.coordinator.transaction.TransactionCoordinator)
[2022-09-09 00:59:44,006] INFO Updated cache from existing <empty> to latest FinalizedFeaturesAndEpoch(features=Features(), epoch=0). (kafka.server.FinalizedFeatureCache)
[2022-09-09 00:59:44,030] INFO [ExpirationReaper-0-All-Acls]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOperationReaper)
[2022-09-09 00:59:44,062] INFO [/config/changes-event-process-thread]: Starting (kafka.common.ZkNodeChangeNotificationListener$ChangeEventProcessThread)
[2022-09-09 00:59:44,075] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Starting socket server acceptors and processors (kafka.network.SocketServer)
[2022-09-09 00:59:44,081] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Started data-plane acceptor and processor(s) for endpoint : ListenerName(PLAINTEXT) (kafka.network.SocketServer)
[2022-09-09 00:59:44,083] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Started socket server acceptors and processors (kafka.network.SocketServer)
[2022-09-09 00:59:44,091] INFO Kafka version: 3.2.1 (org.apache.kafka.common.utils.AppInfoParser)
[2022-09-09 00:59:44,093] INFO Kafka commitId: b172a0a94f4eb9f (org.apache.kafka.common.utils.AppInfoParser)
[2022-09-09 00:59:44,094] INFO Kafka startTimeMs: 1662665384033 (org.apache.kafka.common.utils.AppInfoParser)
[2022-09-09 00:59:44,097] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)
[2022-09-09 00:59:44,202] INFO [BrokerToControllerChannelManager broker=0 name=forwarding]: Recorded new controller, from now on will use broker Aashay:9092 (id: 0 rack: null) (kafka.server.BrokerToControllerRequestThread)
[2022-09-09 00:59:44,235] INFO [BrokerToControllerChannelManager broker=0 name=alter-Partition]: Recorded new controller, from now on will use broker Aashay:9092 (id: 0 rack: null) (kafka.server.BrokerToControllerRequestThread)
```



Shri Vile Parle Kelavani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

Rishabh Singhvi

60009210206

CSE DS D2

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.1889]
(c) Microsoft Corporation. All rights reserved.

C:\kafka\bin\windows>kafka-topics.bat --create --bootstrap-server localhost:9092
```

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.1889]
(c) Microsoft Corporation. All rights reserved.

C:\kafka\bin\windows>kafka-topics.bat --create --bootstrap-server localhost:9092 --topic test
Created topic test.
```

```
C:\kafka\bin\windows>kafka-console-producer.bat --broker-list localhost:9092 --topic test
>
```

```
C:\kafka\bin\windows>kafka-console-consumer.bat --topic test --bootstrap-server localhost:9092 --from-beginning
```

```
C:\kafka\bin\windows>kafka-topics.bat --create --bootstrap-server localhost:9092 --topic test
Created topic test.

C:\kafka\bin\windows>kafka-console-producer.bat --broker-list localhost:9092 --topic test
>
```

Producer Consumer Test:

```
C:\Windows\System32\cmd.exe - kafka-console-producer.bat --broker-list localhost:9092 --topic test
Microsoft Windows [Version 10.0.19044.1889]
(c) Microsoft Corporation. All rights reserved.

C:\kafka\bin\windows>kafka-topics.bat --create --bootstrap-server localhost:9092 --topic test
Created topic test.

C:\kafka\bin\windows>kafka-console-producer.bat --broker-list localhost:9092 --topic test
>{"Name": "John", "Age": "31", "Gender": "Male"}

C:\Windows\System32\cmd.exe - kafka-console-consumer.bat --topic test --bootstrap-server localhost:9092 --from-beginning
Microsoft Windows [Version 10.0.19044.1889]
(c) Microsoft Corporation. All rights reserved.

C:\kafka\bin\windows>kafka-console-consumer.bat --topic test --bootstrap-server localhost:9092 --from-beginning
{"Name": "John", "Age": "31", "Gender": "Male"}
```