



**Department of Computer Science and Engineering (Data Science)**

**Subject: Big Data Engineering (DJ19DSL604)**

**AY: 2022-23**

**Experiment 4**

**(Messaging Service)**

*Bhuvi Ghosh*  
*60009210191*

**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.

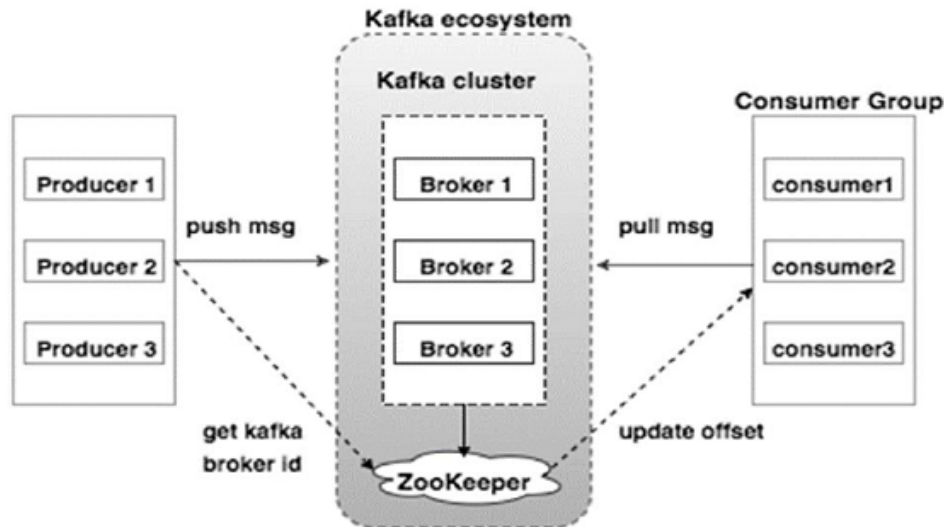
### **Need for Kafka**

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)**



## 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

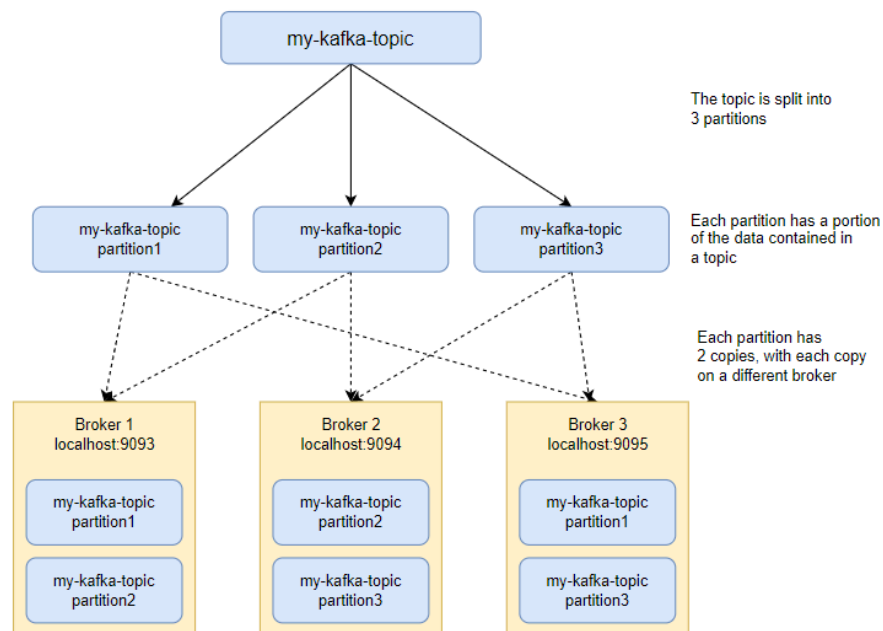


### Department of Computer Science and Engineering (Data Science)

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.

### 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.



Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

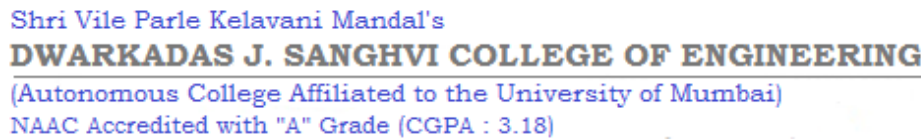
(Autonomous College Affiliated to the University of Mumbai)

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



**Department of Computer Science and Engineering (Data Science)**

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.

[illegible][illegible]





## Department of Computer Science and Engineering (Data Science)

### Creating cluster:

```
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ KAFKA_CLUSTER_ID="$(bin/kafka-storage.sh random-uuid)"
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-storage.sh format -t $KAFKA_CLUSTER_ID -c config/kraft/server.properties
Formatting /tmp/kraft-combined-logs with metadata.version 3.6-IV2.
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-server-start.sh config/kraft/server.properties
```

### Producer event creation:

```
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-topics.sh --create --topic quickstart-events --bootstrap-server localhost:9092
Created topic quickstart-events.
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-topics.sh --describe --topic quickstart-events --bootstrap-server localhost:9092
Topic: quickstart-events TopicId: NPMZHyhbR9y00wMgLMH2sg PartitionCount: 1 ReplicationFactor: 1 Configs:
  Topic: quickstart-events Partition: 0 Leader: 0 Replicas: 0 Isr: 0
Topic: quickstart-events TopicId: uDjcFFvqQVGAZTLha06bbw PartitionCount: 1 ReplicationFactor: 1 Configs:
  Topic: quickstart-events Partition: 0 Leader: 0 Replicas: 0 Isr: 0
Topic: quickstart-events Partition: 0 Leader: 0 Replicas: 0 Isr: 0
Topic: quickstart-events Partition: 0 Leader: 0 Replicas: 0 Isr: 0
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-topics.sh --describe --topic quickstart-events --bootstrap-server localhost:9092
Topic: quickstart-events TopicId: uDjcFFvqQVGAZTLha06bbw PartitionCount: 1 ReplicationFactor: 1 Configs:
  Topic: quickstart-events Partition: 0 Leader: 0 Replicas: 0 Isr: 0
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-console-producer.sh --topic quickstart-events --bootstrap-server localhost:9092
>This is my first event
>This is my second event
```

### Consumer event:

```
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ bin/kafka-console-consumer.sh --topic quickstart-events --from-beginning --bootstrap-server localhost:9092
This is my first event
This is my second event
```

### Editing test.txt :

```
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ more test.txt
foo
bar
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ echo Another line>> test.txt
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$
```

### New test.txt:

```
Text Editor Feb 21 12:00
test.txt
~/Downloads/kafka_2.13-3.6.1
1 foo
2 bar
3 Another line
```

### Cluster termination:

```
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$ rm -rf /tmp/kafka-logs /tmp/zookeeper /tmp/kraft-combined-logs
csds-student@mum0923cpu0962: ~/Downloads/kafka_2.13-3.6.1$
```