



## **Module 9**

# **Introduction to Kafka**

Thanachart Numnonda, Executive Director, IMC Institute

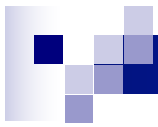
Thanisa Numnonda, Faculty of Information Technology,  
King Mongkut's Institute of Technology Ladkrabang

# Introduction

Open-source message broker project

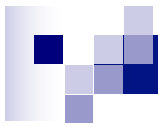


**An open-source message broker project developed by the Apache Software Foundation written in Scala. The project aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds. It is, in its essence, a "massively scalable pub/sub message queue architected as a distributed transaction log", making it highly valuable for enterprise infrastructures.**



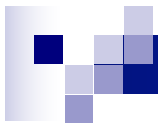
# What is Kafka?

- An apache project initially developed at LinkedIn
- Distributed publish-subscribe messaging system
- Designed for processing of real time activity stream data  
e.g. logs, metrics collections
- Written in Scala
- Does not follow JMS Standards, neither uses JMS APIs



# Kafka: Features

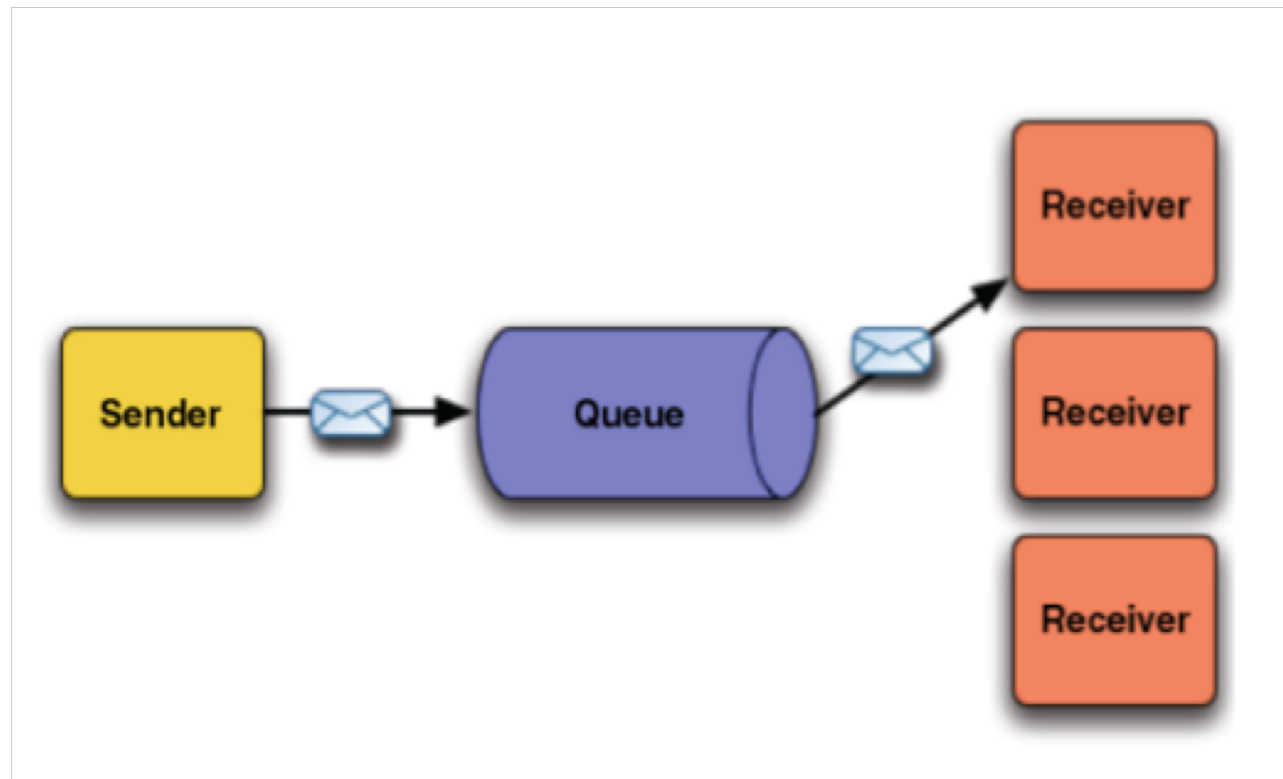
- Persistent messaging
- High-throughput
- Supports both queue and topic semantics
- Uses Zookeeper for forming a cluster of nodes (producer/consumer/broker) and many more...



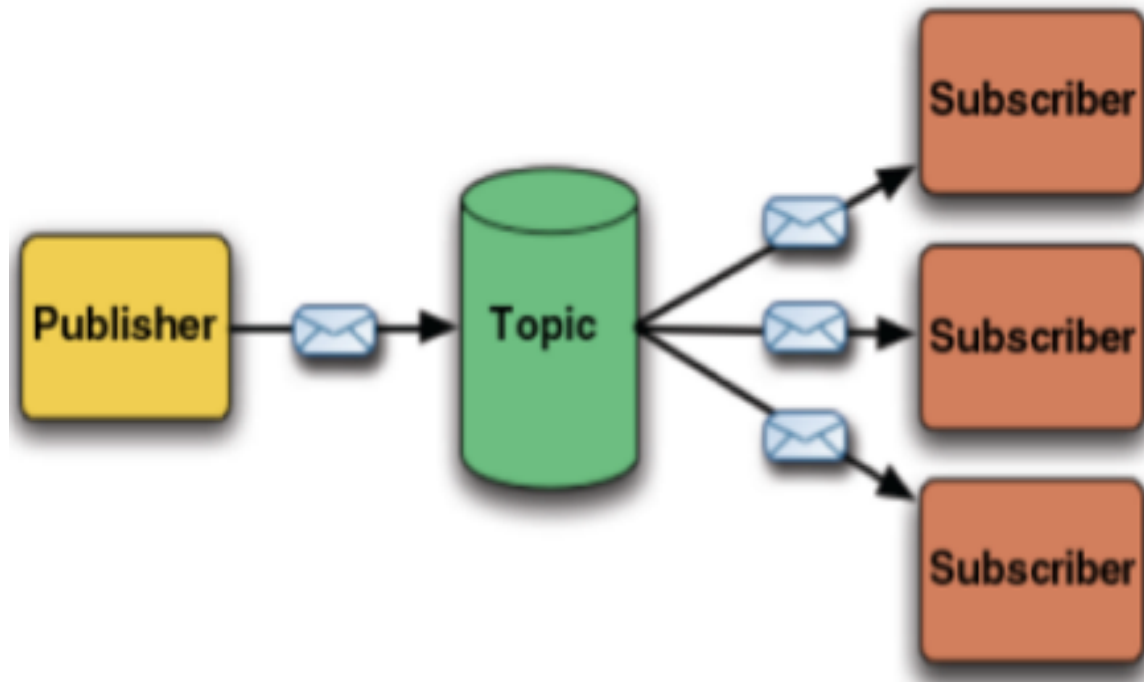
# Why Kafka?

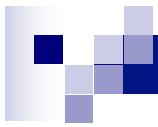
- Built with speed and scalability in mind.
- Enabled near real-time access to any data source
- Empowered hadoop jobs
- Allowed us to build real-time analytics
- Vastly improved our site monitoring and alerting capability
- Enabled us to visualize and track our call graphs.

# Messaging System Concept: Queue



# Messaging System Concept: Topic

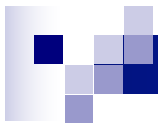




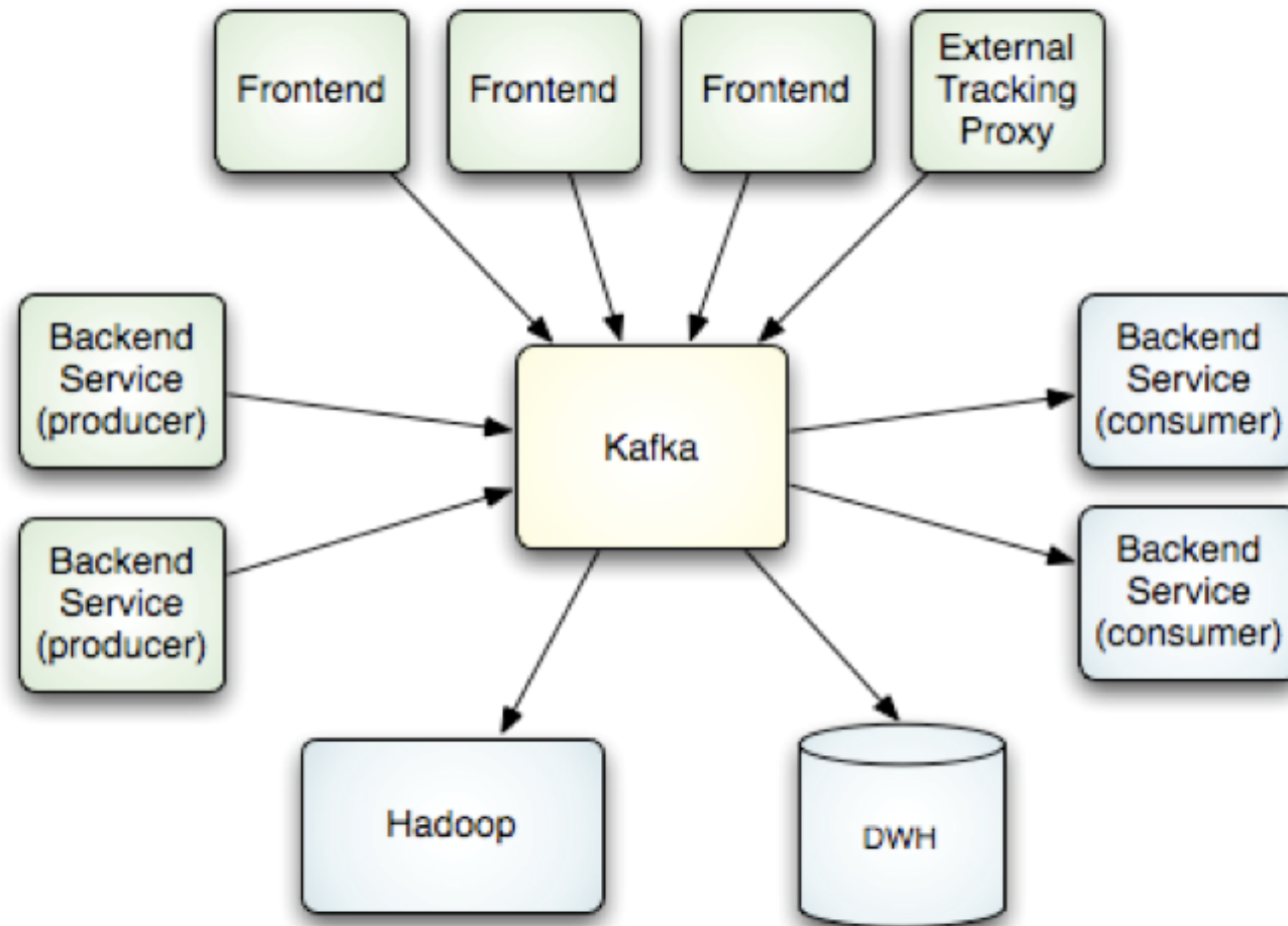
# Terminology

- Kafka maintains feeds of messages in categories called topics.
- Processes that publish messages to a Kafka topic are called producers.
- Processes that subscribe to topics and process the feed of published messages are called consumers.
- Kafka is run as a cluster comprised of one or more servers, each of which is called a broker.





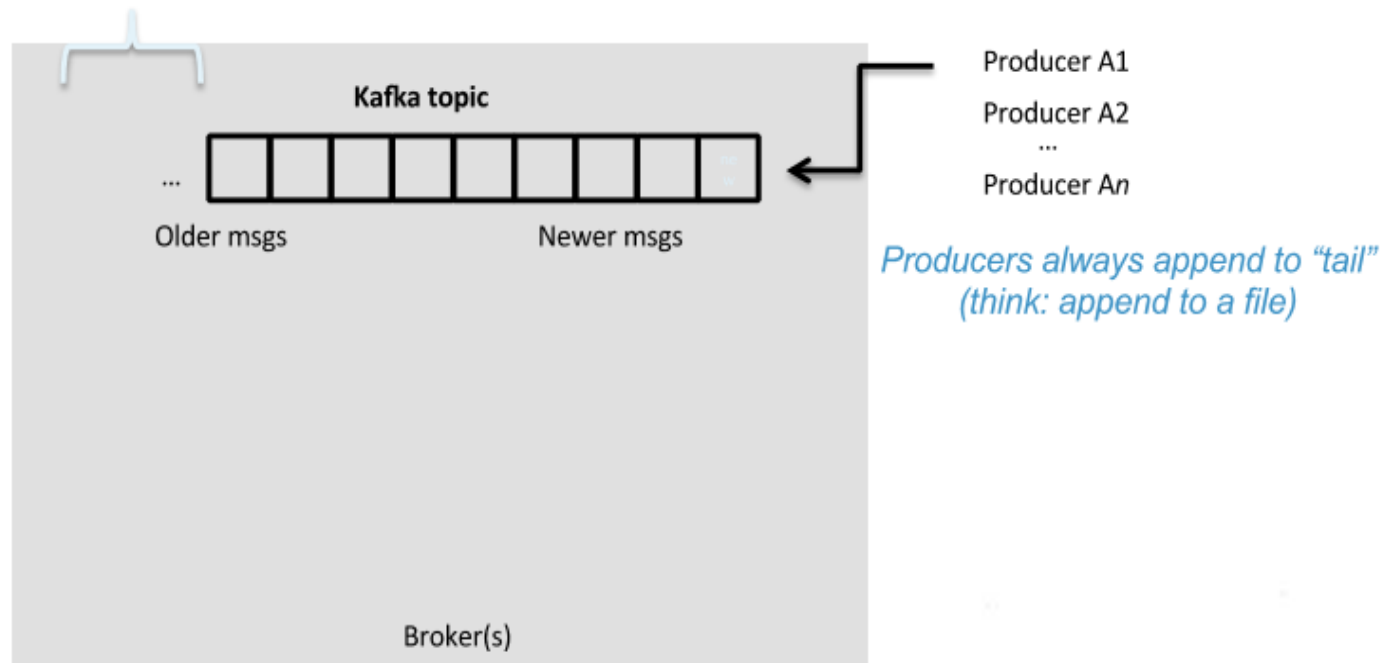
# Kafka



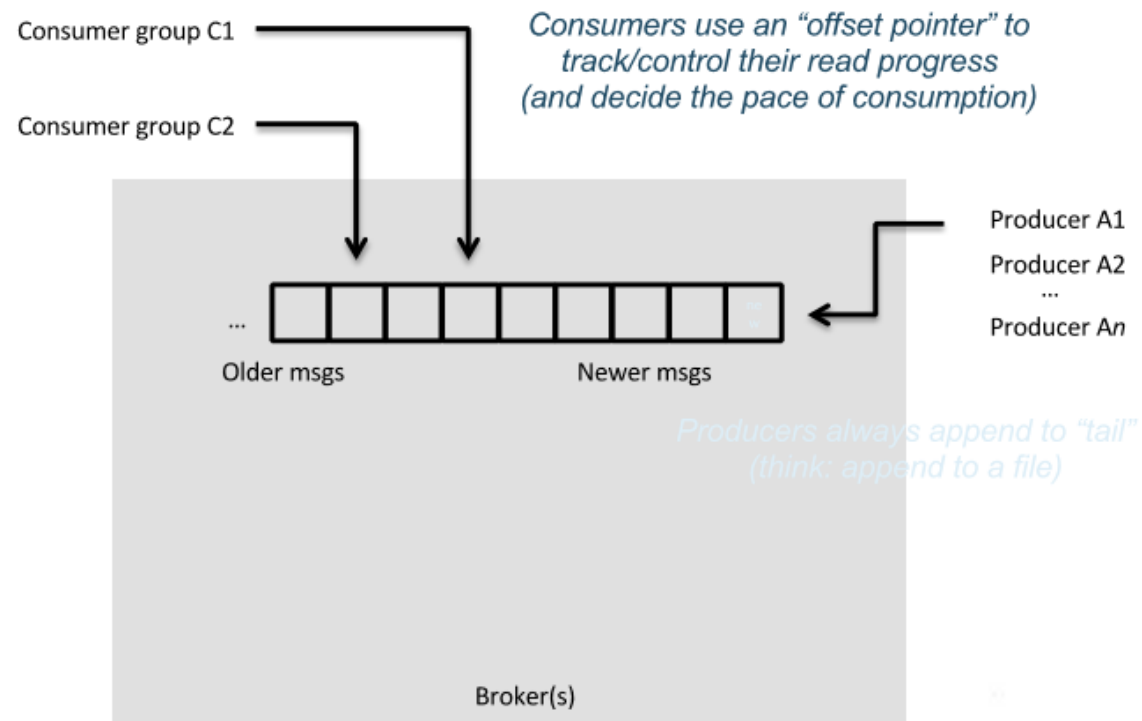
# Topics

- Topic: feed name to which messages are published

*Kafka prunes "head" based on age or max size or "key"*

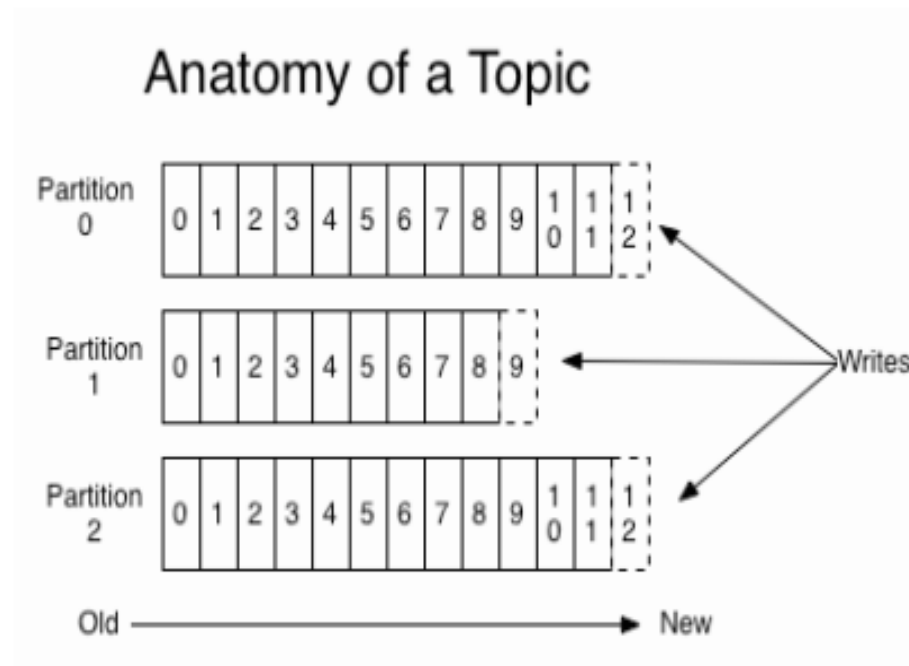


# Topics

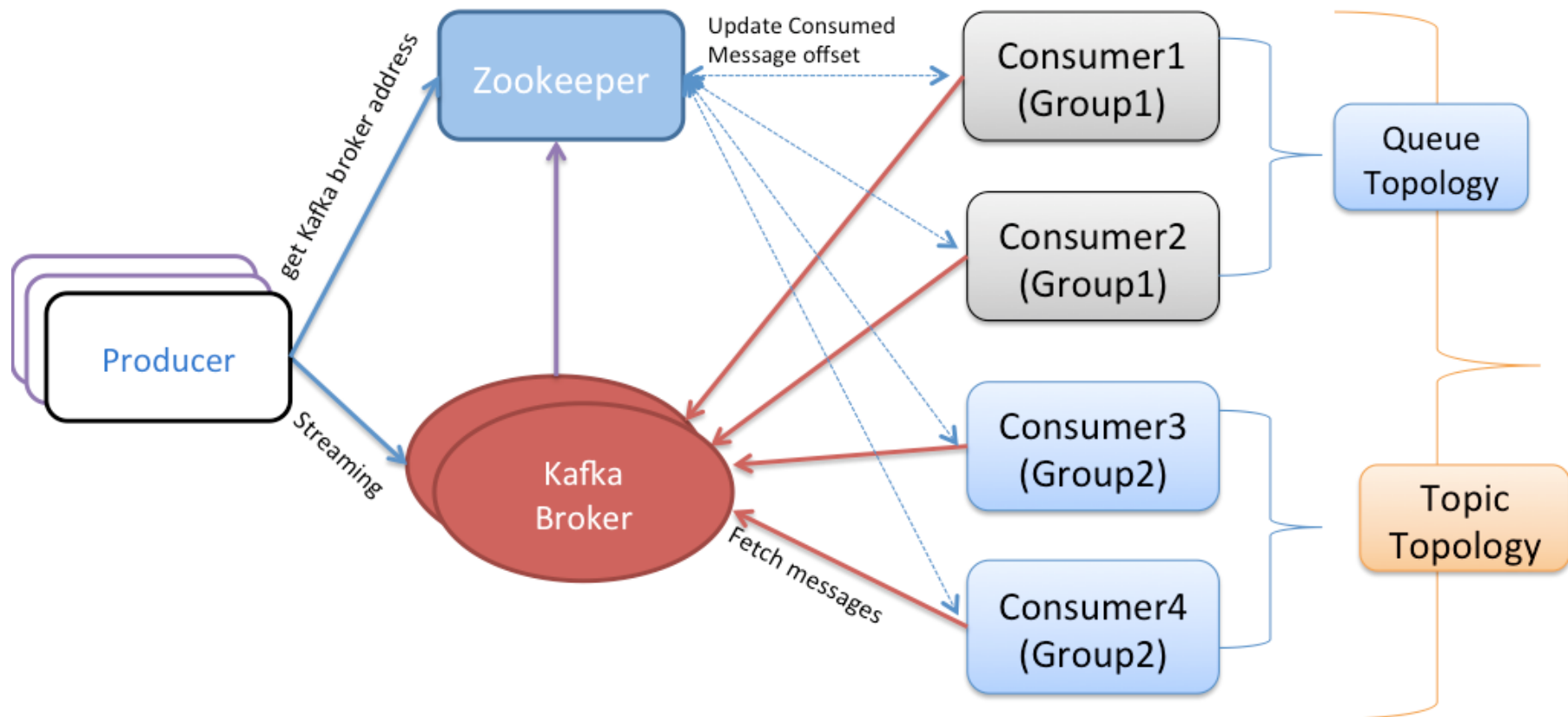


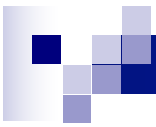
# Topics

- A topic consists of partitions.
- Partition: ordered + immutable sequence of messages that is continually appended



# Kafka Architecture





# **Hands-on**

---

# **SparkStreaming with Kafka**



# Install & Start Kafka Server

```
# wget http://www-us.apache.org/dist/kafka/0.11.0.2/  
kafka_2.11-0.11.0.2.tgz  
  
# tar xzf kafka_2.11-0.11.0.2.tgz  
  
# cd kafka_2.11-0.11.0.2  
  
# bin/kafka-server-start.sh config/server.properties&
```

```
[2018-04-06 21:32:56,084] INFO Creating /brokers/ids/0 (is it secure? false) (kafka.utils.ZKCheckedEphemeral)  
[2018-04-06 21:32:56,090] INFO Result of znode creation is: OK (kafka.utils.ZKCheckedEphemeral)  
[2018-04-06 21:32:56,092] INFO Registered broker 0 at path /brokers/ids/0 with addresses: EndPoint(quickstart.cloudera,9092,ListenerName(PLAINTEXT),PLAINTEXT) (kafka.utils.ZkUtils)  
[2018-04-06 21:32:56,092] WARN No meta.properties file under dir /tmp/kafka-logs/meta.properties (kafka.server.BrokerMetadataCheckpoint)  
[2018-04-06 21:32:56,106] INFO Kafka version : 0.11.0.2 (org.apache.kafka.common.utils.AppInfoParser)  
[2018-04-06 21:32:56,106] INFO Kafka commitId : 73bele1168f91ee2 (org.apache.kafka.common.utils.AppInfoParser)  
[2018-04-06 21:32:56,107] INFO [Kafka Server 0], started (kafka.server.KafkaServer)
```



# Running Kafka Producer

*Suggestion: Please open new terminal before run any following commands*

```
# bin/kafka-console-producer.sh --topic test --broker-list  
localhost:9092
```

type some random messages

```
> This is a test message from IT KMITL.  
> Test  
> Test again !
```

(press Ctrl+c to exit)





## Running Kafka Consumer

```
# bin/kafka-console-consumer.sh --topic test --zookeeper  
localhost:2181 --from-beginning
```

```
This is a test message from IT KMITL.  
Test  
Test again !
```

(press Ctrl+c to exit)