

# BIG DATA HADOOP & SPARK TRAINING

Assignment on Kafka Introduction-1

Rashmi Krishna

---

## Few definitions:

**Topic:** A topic is a category or feed name to which records are published. Topics in Kafka are always multi-subscriber; that is, a topic can have zero, one, or many consumers that subscribe to the data written to it

**Producers:** Producers publish data to the topics of their choice. The producer is responsible for choosing which record to assign to which partition within the topic. This can be done in a round-robin fashion simply to balance load or it can be done according to some semantic partition function

**Consumers:** Consumers label themselves with a *consumer group* name, and each record published to a topic is delivered to one consumer instance within each subscribing consumer group.

If all the consumer instances have the same consumer group, then the records will effectively be load balanced over the consumer instances.

If all the consumer instances have different consumer groups, then each record will be broadcast to all the consumer processes.

**Zookeeper:** ZooKeeper is a service to which clients can connect to. It provides access to clients to a tree like structure or a hierarchical namespace.

## Task 1:

### Create a kafka topic named KeyLessTopic.

Command:-

```
kafka-topics.sh --create --topic KeyLessTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
```

### Inside KeyLessTopic insert following data:

```
kafka-console-producer.sh --broker-list localhost:9092 --topic KeyLessTopic
```

Insert the below data:-

```
{"name":"John", "exp":16}
```

```
{"name":"Finn", "exp":20}
```

```
{"name":"Cylin", "exp":18}
```

```
{"name":"Mark", "exp":2}
```

```
{"name":"Akshay", "exp":14}
```

## Task 2:

**Create a console consumer that reads KeyLessTopic from beginning**

Command:-

```
kafka-console-consumer.sh --topic KeyLessTopic --from-beginning \  
--zookeeper localhost:2181 \  
--property print.key=true
```

```
[acadgild@localhost bin]$ kafka-topics.sh --create --topic KeyLessTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1  
Created topic "KeyLessTopic".  
[acadgild@localhost bin]$ kafka-console-producer.sh --broker-list localhost:9092 --topic KeyLessTopic  
{ "name": "John", "exp": 16 }  
{ "name": "Finn", "exp": 20 }  
{ "name": "Cylin", "exp": 18 }  
{ "name": "Mark", "exp": 2 }  
{ "name": "Akshay", "exp": 14 }  
^CYou have new mail in /var/spool/mail/acadgild  
[acadgild@localhost bin]$ kafka-console-consumer.sh --topic KeyLessTopic --from-beginning \  
--zookeeper localhost:2181 \  
--property print.key=true  
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a future major release. Consider using the new consumer by passing [bootstrap-server] instead of [zookeeper].  
null { "name": "John", "exp": 16 }  
null { "name": "Finn", "exp": 20 }  
null { "name": "Cylin", "exp": 18 }  
null { "name": "Mark", "exp": 2 }  
null { "name": "Akshay", "exp": 14 }
```

Inserting values

output of values stored in KeyLessTopic

## Task 3:

**Create a kafka topic named KeyedTopic. Inside KeyedTopic insert following data:**

The part before comma(,) should be treated as key and after comma(,) should be treated as value

```
{"name": "John"}, {"exp": 16}  
{"name": "Finn"}, {"exp": 20}  
{"name": "Cylin"}, {"exp": 18}  
{"name": "Mark"}, {"exp": 2}  
{"name": "Akshay"}, {"exp": 14}
```

Command:-

```
kafka-topics.sh --create --topic KeyedTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
```

```
kafka-console-producer.sh --broker-list localhost:9092 --topic KeyedTopic --property parse.key=true --property key.separator=","
```


The above command, produces data with key value separator, where key and values are separated with “,”. Now, Insert the above mentioned data

The below command is used to read the values of the topic:

```
kafka-console-consumer.sh --topic KeyedTopic --from-beginning \
--zookeeper localhost:2181 \
--property print.key=true
```

```
[acadgild@localhost bin]$ kafka-topics.sh --create --topic KeyedTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
Created topic "KeyedTopic".
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost bin]$ kafka-console-producer.sh --broker-list localhost:9092 --topic KeyedTopic --property parse.key=true --property key.separator=","
{"name":"John"}, {"exp":16}
{"name":"Finn"}, {"exp":20}
{"name":"Cylin"}, {"exp":18}
{"name":"Mark"}, {"exp":2}
{"name":"Akshay"}, {"exp":14}
^C[acadgild@localhost bin]$ kafka-console-consumer.sh --topic KeyedTopic --from-beginning \
> --zookeeper localhost:2181 \
> --property print.key=true
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a future major release. Consider using the new consumer by passing [bootstrap-server] instead of [zookeeper].
{"name":"John"} {"exp":16}
{"name":"Finn"} {"exp":20}
{"name":"Cylin"} {"exp":18}
{"name":"Mark"} {"exp":2}
{"name":"Akshay"} {"exp":14}
```

The part before comma(,) is treated as key and after comma(,) is treated as value



## Task 4:

**Create a console consumer that reads KeyedTopic from beginning. The key and value should be separated by '-'**

Creating topic:

```
kafka-topics.sh --create --topic KeydTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
```

creating Producer with key and value separator "--"

```
kafka-console-producer.sh --broker-list localhost:9092 --topic KeydTopic --property  
parse.key=true --property key.separator="-"
```

```
{"name":"John"}-{"exp":16}
```

```
{"name":"Finn"}-{"exp":20}
```

```
{"name":"Cylin"}-{"exp":18}
```

```
{"name":"Mark"}-{"exp":2}
```

```
{"name":"Akshay"}-{"exp":14}
```

Creating consumer:

```
kafka-console-consumer.sh --topic KeydTopic --from-beginning \  
--zookeeper localhost:2181 \  
--property print.key=true
```

```
[acadgild@localhost bin]$ kafka-topics.sh --create --topic KeydTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1  
Created topic "KeydTopic".  
[acadgild@localhost bin]$ kafka-console-producer.sh --broker-list localhost:9092 --topic KeydTopic --property parse.key=true --propert  
y key.separator="-"  
{"name":"John"}-{"exp":16}  
{"name":"Finn"}-{"exp":20}  
{"name":"Cylin"}-{"exp":18}  
{"name":"Mark"}-{"exp":2}  
{"name":"Akshay"}-{"exp":14}  
^C[acadgild@localhost bin]$ kafka-console-consumer.sh --topic KeydTopic --from-beginning \  
> --zookeeper localhost:2181 \  
> --property print.key=true  
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a future major release. Consider using the new consume  
r by passing [bootstrap-server] instead of [zookeeper].  
{"name":"John"} {"exp":16}  
{"name":"Finn"} {"exp":20}  
{"name":"Cylin"} {"exp":18}  
{"name":"Mark"} {"exp":2}  
{"name":"Akshay"} {"exp":14}
```