

Session 22: KAFKA INTRODUCTION

Assignment 1

Task 1

Create a Kafka Topic named KeyLessTopic

Inside KeyLessTopic insert following data:

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

Solution:

First we need to start the zookeeper

Starting the zookeeper by following command

\$KAFKA_HOME/bin/zookeeper-server-start.sh \$KAFKA_HOME/config/zookeeper.properties

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/zookeeper-server-start.sh $KAFKA_HOME/config/zookeeper.properties
[2018-10-02 14:51:49,301] INFO Reading configuration from: /home/acadgild/install/kafka/kafka_2.12-0.10.1.1/config/zookeeper.properties (
org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2018-10-02 14:51:49,348] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2018-10-02 14:51:49,349] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2018-10-02 14:51:49,350] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2018-10-02 14:51:49,350] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.
quorum.QuorumPeerMain)
[2018-10-02 14:51:49,679] INFO Reading configuration from: /home/acadgild/install/kafka/kafka_2.12-0.10.1.1/config/zookeeper.properties (
org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2018-10-02 14:51:49,689] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2018-10-02 14:51:49,768] INFO Server environment:zookeeper.version=3.4.8--1, built on 02/06/2016 03:18 GMT (org.apache.zookeeper.server.
ZooKeeperServer)
[2018-10-02 14:51:49,768] INFO Server environment:host.name=localhost (org.apache.zookeeper.server.ZooKeeperServer)
[2018-10-02 14:51:49,768] INFO Server environment:java.version=1.8.0_151 (org.apache.zookeeper.server.ZooKeeperServer)
[2018-10-02 14:51:49,768] INFO Server environment:java.vendor=Oracle Corporation (org.apache.zookeeper.server.ZooKeeperServer)
[2018-10-02 14:51:49,769] INFO Server environment:java.home=/usr/java/jdk1.8.0_151/jre (org.apache.zookeeper.server.ZooKeeperServer)
[2018-10-02 14:51:49,769] INFO Server environment:java.class.path=:/home/acadgild/install/kafka/kafka_2.12-0.10.1.1/bin/../libs/aopallian
ce-repackaged-2.4.0-b34.jar:/home/acadgild/install/kafka/kafka_2.12-0.10.1.1/bin/../libs/argparse4j-0.5.0.jar:/home/acadgild/install/kafk
```

Now, we have to start the Kafka broker with default properties by using following command:

\$KAFKA_HOME/bin/kafka-server-start.sh \$KAFKA_HOME/config/server.properties

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-server-start.sh $KAFKA_HOME/config/server.properties
[2018-10-02 14:52:29,115] INFO KafkaConfig values:
  advertised.host.name = null
  advertised.listeners = null
  advertised.port = null
  authorizer.class.name =
  auto.create.topics.enable = true
  auto.leader.rebalance.enable = true
  background.threads = 10
  broker.id = 0
  broker.id.generation.enable = true
  broker.rack = null
  compression.type = producer
  connections.max.idle.ms = 600000
  controlled.shutdown.enable = true
  controlled.shutdown.max.retries = 3
  controlled.shutdown.retry.backoff.ms = 5000
  controller.socket.timeout.ms = 30000
  default.replication.factor = 1
  delete.topic.enable = true
  fetch.purgatory.purge.interval.requests = 1000
  group.max.session.timeout.ms = 300000
  group.min.session.timeout.ms = 6000
```

We have copied all messages into 'kafka_data_for_topics' text file. This will help us to write or read all lines at once from this file.

We can see the content by using following command

cat kafka_data_for_topics.txt

```
[acadgild@localhost ~]$ cat kafka_data_for_topics.txt
{"name":"John", "exp":16}
{"name":"Finn", "exp":20}
{"name":"Cylin", "exp":18}
{"name":"Mark", "exp":2}
{"name":"Akshay", "exp":14}[acadgild@localhost ~]$
```

Now, creating topic with name "KeyLessTopic" by following command

\$KAFKA_HOME/bin/kafka-topics.sh --create --topic KeyLessTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-topics.sh --create --topic KeyLessTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
Created topic "KeyLessTopic".
You have new mail in /var/spool/mail/acadgild
```

Now, verifying whether "KeyLessTopic" is created or not by using following command.

\$KAFKA_HOME/bin/kafka-topics.sh --list --zookeeper localhost:2181

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-topics.sh --list --zookeeper localhost:2181
KeyLessTopic
TestTopic
__consumer_offsets
my-replicated-topic
my-topic
```

Now, we have created Kafka producer (publisher) to publish the messages on the “KeyLessTopic” and messages are provided by “kafka_data_for_topics.txt” file.

```
$KAFKA_HOME/bin/kafka-console-producer.sh --broker-list localhost:9092 --topic KeyLessTopic < kafka_data_for_topics.txt
```

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-console-producer.sh --broker-list localhost:9092 --topic KeyLessTopic < kafka_data_for_topics.txt
```

Task 2

Create a console consumer that reads KeyLessTopic from beginning

Solution:

We have created console consumer for the KeyLessTopic topic to receive the messages from the beginning. Here we have set print.key option to false as we are not using keys in this case.

```
$KAFKA_HOME/bin/kafka-console-consumer.sh --topic KeyLessTopic --from-beginning --zookeeper localhost:2181 --property print.key=false
```

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-console-consumer.sh --topic KeyLessTopic --from-beginning --zookeeper localhost:2181 --property print.key=false
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}
```

We could see that all messages from producer are delivered to consumer.

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}
```

Solution:

Here we have created Topic named as ‘KeyedTopic’ by using below command:

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

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-topics.sh --create --topic KeyedTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
Created topic "KeyedTopic".
```

Then we have verified '**KeyedTopic**' topic is created by using below List topic command which shows list of all topics created:

\$KAFKA_HOME/bin/kafka-topics.sh --list --zookeeper localhost:2181

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-topics.sh --list --zookeeper localhost:2181
KeyLessTopic
KeyedTopic
TestTopic
_consumer_offsets
my-replicated-topic
my-topic
```

We have copied all messages into a new text file : 'kafka_data_for_topics_with_key'. This will help us to write or read all lines at once from this file.

We could see content of this file by using below command:

cat kafka_data_for_topics_with_key.txt

```
[acadgild@localhost ~]$ cat kafka_data_for_topics_with_key.txt
{"name": "John"}, {"exp": 16}
{"name": "Finn"}, {"exp": 20}
{"name": "Cylin"}, {"exp": 18}
{"name": "Mark"}, {"exp": 2}
{"name": "Akshay"}, {"exp": 14}
```

We have created Kafka producer (publisher) to publish the messages on the KeyedTopic and messages are provided by kafka_data_for_topics_with_key.txt file.

We also provide extra options to the command to send messages as key/value pair and to use "," as the separator between key and value.

\$KAFKA_HOME/bin/kafka-console-producer.sh --broker-list localhost:9092 --topic KeyedTopic --property "parse.key"=true --property "key.separator=" < kafka_data_for_topics_with_key.txt

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-console-producer.sh --broker-list localhost:9092 --topic KeyedTopic --property "parse.key"=true --property "key.separator=" < kafka_data_for_topics_with_key.txt
```

Task 4

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

Solution:

We have created console consumer for the KeyedTopic topic to receive the messages from the beginning. Here we have set print.key option to true as we are expecting both keys and values.

Since we are expecting to receive messages in the form of Key/Value pair, we provide extra options to the command to print the key and the value and use key.separator as '- '.


```
$KAFKA_HOME/bin/kafka-console-consumer.sh --topic KeyedTopic --from-beginning --zookeeper localhost:2181 --property "print.key=true" --property "parse.key=true" --property "key.separator=-"
```

```
[acadgild@localhost ~]$ $KAFKA_HOME/bin/kafka-console-consumer.sh --topic KeyedTopic --from-beginning --zookeeper localhost:2181 --property "print.key=true" --property "parse.key=true" --property "key.separator=-"
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}
```

We could see that keys and values are separated and printed using “-” as the separator.
e.g. For the key : {"name":"John"}, it's value is {"exp":16}.