Big Data- Hadoop – Assignment 8- Kafka1

For this assignment we have to make sure that ZooKeeper and Kafka is running

1. Start the ZooKeeper with below command

Commands

- cd $KAFKA_HOME ($KAFKA_HOME = /home/acadgild/install/kafka/kafka_2.120.10.1.1/)
- ./bin/zookeeper-server-start.sh ./config/zookeeper.properties

```
[acadgild@localhost ~]$ cd $KAFKA_HOME
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/zookeeper-server-start.sh ./conf
ig/zookeeper.properties
[2018-08-26 04:53:38,606] INFO Reading configuration from: ./config/zookeeper.pr
operties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2018-08-26 04:53:38,627] INFO autopurge.snapRetainCount set to 3 (org.apache.zo
okeeper.server.DatadirCleanupManager)
[2018-08-26 04:53:38,627] INFO autopurge.purgeInterval set to 0 (org.apache.zook
eeper.server.DatadirCleanupManager)
[2018-08-26 04:53:38,627] INFO Purge task is not scheduled. (org.apache.zookeepe
r.server.DatadirCleanupManager)
[2018-08-26 04:53:38,627] WARN Either no config or no quorum defined in config,
running  in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2018-08-26 04:53:38,695] INFO Reading configuration from: ./config/zookeeper.pr
operties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2018-08-26 04:53:38,698] INFO Starting server (org.apache.zookeeper.server.ZooK
eeperServerMain)
[2018-08-26 04:53:38,730] INFO Server environment:zookeeper.version=3.4.8--1, bu
ilt on 02/06/2016 03:18 GMT (org.apache.zookeeper.server.ZooKeeperServer)
[2018-08-26 04:53:38,730] INFO Server environment:host.name=localhost (org.apach
e.zookeeper.server.ZooKeeperServer)
[2018-08-26 04:53:38,730] INFO Server environment:java.version=1.8.0_151 (org.ap
ache.zookeeper.server.ZooKeeperServer)
[2018-08-26 04:53:38,730] INFO Server environment:java.vendor=Oracle Corporation
 (org.apache.zookeeper.server.ZooKeeperServer)
```

2. Start the Kafka Broker

Commands

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

```
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/kafka-server-start.sh ./config/s
erver.properties
[2018-08-26 05:04:51,756] 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 = false
        fetch.purgatory.purge.interval.requests = 1000
        group.max.session.timeout.ms = 300000
        group.min.session.timeout.ms = 6000
```

Task 1:

Create a kafka topic named KeyLessTopic.

Solution Approach –

- Use the below command to create the topic

" bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic **topic-name** "

Code Execution as below

```
[acadgild@localhost bin]$ kafka-topics.sh --create --zookeeper localhost:2181 --
replication-factor 1 --partitions 1 --topic KeylessTopic1
Created topic "KeylessTopic1".
You have new mail in /var/spool/mail/acadgild
```

- Verify if topic is created with the list command as below

"bin/kafka-topics.sh --list --zookeeper localhost:2181 "

List of topics created

```
[acadgild@localhost bin]$ kafka-topics.sh --list --zookeeper localhost:2181
KeylessTopic
KeylessTopic1
```

Task 1.2

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}

- To Insert the data we must start the producer

Start Producer to Send Messages

"bin/kafka-console-producer.sh --broker-list localhost:9092 --topic topic-name"

Broker-list – The list of brokers that we want to send the messages to. In this case we only have one broker. The Config/server.properties file contains broker port id, since we know our broker is listening on port 9092, so you can specify it directly.

Topic name − Here is an example for the topic name.

Every line entered on the console is treated as a value with NULL as key as this topic is key less

Big Data- Hadoop – Assignment 8- Kafka1

Code execution

```
[acadgild@localhost bin]$ cd $KAFKA_HOME
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/kafka-console-producer.sh --brok
er-list localhost:9092 --topic KeylessTopic1
```

Inserting the data

```
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/kafka-console-producer.sh --bro
er-list localhost:9092 --topic KeylessTopic1
{"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 To read we have to Start Consumer to Receive Messages

Similar to producer, the default consumer properties are specified in config/consumer.properties file. Open a new terminal and type the below syntax for consuming messages.

“bin/kafka-console-consumer.sh --zookeeper localhost:2181 —topic topic-name --from-beginning “

Code Execution with output

```
[acadgild@localhost ~]$ cd $KAFKA_HOME
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/kafka-console-consumer.sh --topi
c KeylessTopic1 --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-s
erver] 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}
```

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

Data:

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

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

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

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

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

Create a topic which accepts key-value pair of data

Code Execution and Output

```
[acadgild@localhost ~]$ cd $KAFKA_HOME
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./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 kafka_2.12-0.10.1.1]$ ./bin/kafka-topics.sh --list --zookeeper local
host:2181
KeyedTopic
KeylessTopic
KeylessTopic1
You have new mail in /var/spool/mail/acadgild
```

Create a producer to insert the data with key value pair . To achive the same we have to metion two properties while creating producer

- Parse-key as true
- Key separator as ‘,’ (comma) in this case.

Code Execution and Output

```
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./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}
^CYou have new mail in /var/spool/mail/acadgild
```

Task 4:

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

The topic has data inserted in key value format by the producer created in above task (3). However the producer gave the key separator as ‘,’. However we can overwrite the key separator value in console consumer and mandate to print key-value pair. To achive the same we must write console consumer with the two properties

- Print.key = true
- Key-separator = ‘-‘

Command is

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

Code Execution and Output

```
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/kafka-console-consumer.sh --topic KeyedT
opic --property print.key=true --property key.separator="-" --from-beginning \
> --zookeeper localhost:2181 \
> --property print.key=true
Using the ConsoleConsumer with old consumer is deprecated and will be removed in a futur
e major release. Consider using the new consumer by passing [bootstrap-server] instead o
f [zookeeper].
{"name":"John"}-{"exp":16}
{"name":"Finn"}-{"exp":20}
{"name":"Cylin"}-{"exp":18}
{"name":"Mark"}-{"exp":2}
{"name":"Akshay"}-{"exp":14}
```

Task 5:

Create a java program MyKafkaProducer.java that takes a file name and delimiter as input arguments.

It should read the content of file line by line.

Fields in the file are in following order

1. Kafka Topic Name

2. Key

3. value

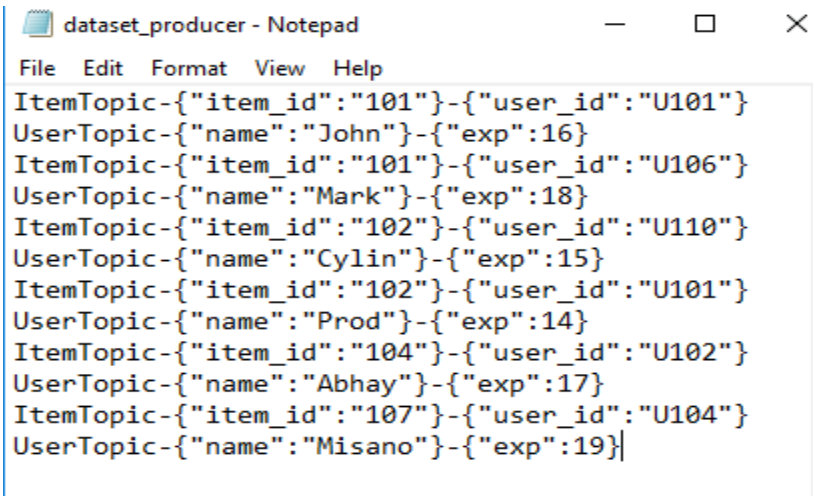
For every line, insert the key and value to the respective Kafka broker in a fire and forget mode.

After record is sent, it should print appropriate message on screen.

Pass dataset_producer.txt as the input file and -as delimiter.

LINK: https://drive.google.com/file/d/0B_Qjau8wv1KoSnR5eHpKOF9rTFU/view?usp=sharing

Dataset Used for this assignment is as below



Big Data- Hadoop – Assignment 8- Kafka1

Topics Creation from Command Line

It has two topics. These topics needs to be created from command line as below

- Created topic named as **ItemTopic**

“bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 2 --topic ItemTopic “

- Created topic named as **UserTopic**

“bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 2 --topic UserTopic “