

Session 22:
KAFKA INTRODUCTION
Assignment 1

- Start zookeeper using command `"bin/zookeeper-server-start.sh config/zookeeper.properties"`

```

acadgild@localhost:~/install/kafka/kafka_2.12-0.10.1.1
File Edit View Search Terminal Help
kafka-reassign-partitions.sh      zookeeper-server-stop.sh
kafka-replay-log-producer.sh     zookeeper-shell.sh
[acadgild@localhost bin]$ cd ..
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost kafka_2.12-0.10.1.1]$ bin/zookeeper-server-start.sh config/z
ookeeper.properties
[2019-02-05 19:54:44,560] INFO Reading configuration from: config/zookeeper.prop
erties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2019-02-05 19:54:44,597] INFO autopurge.snapRetainCount set to 3 (org.apache.zo
ookeeper.server.DataDirCleanupManager)
[2019-02-05 19:54:44,598] INFO autopurge.purgeInterval set to 0 (org.apache.zook
eeper.server.DataDirCleanupManager)
[2019-02-05 19:54:44,598] INFO Purge task is not scheduled. (org.apache.zookeep
er.server.DataDirCleanupManager)
[2019-02-05 19:54:44,598] WARN Either no config or no quorum defined in config,
running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2019-02-05 19:54:44,774] INFO Reading configuration from: config/zookeeper.prop
erties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2019-02-05 19:54:44,781] INFO Starting server (org.apache.zookeeper.server.ZooK
eeperServerMain)
[2019-02-05 19:54:44,864] INFO Server environment:zookeeper.version=3.4.8--1, bu
ilt on 02/06/2016 03:18 GMT (org.apache.zookeeper.server.ZooKeeperServer)
[2019-02-05 19:54:44,864] INFO Server environment:host.name=localhost (org.apach
e.zookeeper.server.ZooKeeperServer)

```

- Start kafka server using command `"bin/kafka-server-start.sh config/kafka.properties"`

```

acadgild@localhost:~/install/kafka/kafka_2.12-0.10.1.1
File Edit View Search Terminal Help
rty.ratio -> 0.5, index.interval.bytes -> 4096, unclean.leader.election.enable -
> true, retention.bytes -> -1, delete.retention.ms -> 86400000, cleanup.policy -
> compact, flush.ms -> 9223372036854775807, segment.ms -> 604800000, segment.by
tes -> 104857600, retention.ms -> 604800000, message.timestamp.difference.max.ms
-> 9223372036854775807, segment.index.bytes -> 10485760, flush.messages -> 92233
72036854775807}. (kafka.log.LogManager)
[2019-02-05 20:15:23,925] INFO Partition [__consumer_offsets,19] on broker 0: No
checkpointed highwatermark is found for partition [__consumer_offsets,19] (kafk
a.cluster.Partition)
[2019-02-05 20:15:23,932] INFO Completed load of log __consumer_offsets-28 with
1 log segments and log end offset 0 in 1 ms (kafka.log.Log)
[2019-02-05 20:15:23,935] INFO Created log for partition [__consumer_offsets,28]
in /tmp/kafka-logs with properties {compression.type -> producer, message.forma
t.version -> 0.10.1-IV2, file.delete.delay.ms -> 60000, max.message.bytes -> 100
0012, min.compaction.lag.ms -> 0, message.timestamp.type -> CreateTime, min.insy
nc.replicas -> 1, segment.jitter.ms -> 0, preallocate -> false, min.cleanable.di
rty.ratio -> 0.5, index.interval.bytes -> 4096, unclean.leader.election.enable -
> true, retention.bytes -> -1, delete.retention.ms -> 86400000, cleanup.policy -
> compact, flush.ms -> 9223372036854775807, segment.ms -> 604800000, segment.by
tes -> 104857600, retention.ms -> 604800000, message.timestamp.difference.max.ms
-> 9223372036854775807, segment.index.bytes -> 10485760, flush.messages -> 92233
72036854775807}. (kafka.log.LogManager)
[2019-02-05 20:15:23,936] INFO Partition [__consumer_offsets,28] on broker 0: No
checkpointed highwatermark is found for partition [__consumer_offsets,28] (kafk

```

Task – 1:

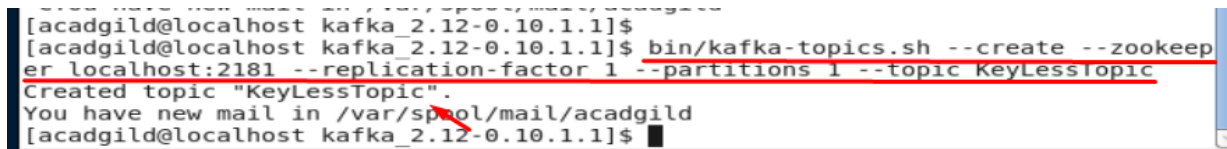
Create a kafka topic named KeyLessTopic.

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}
```

Execution:

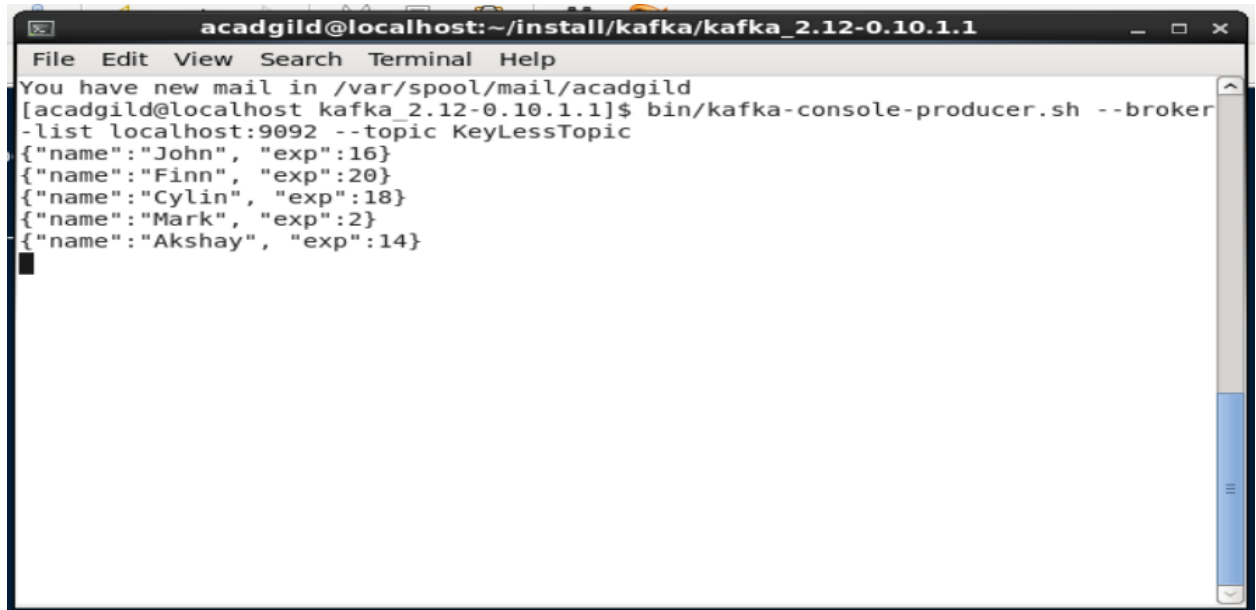
- Create new topic "KeyLessTopic"



```
[acadgild@localhost kafka_2.12-0.10.1.1]$  
[acadgild@localhost kafka_2.12-0.10.1.1]$ bin/kafka-topics.sh --create --zookeep  
er localhost:2181 --replication-factor 1 --partitions 1 --topic KeyLessTopic  
Created topic "KeyLessTopic".  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost kafka_2.12-0.10.1.1]$
```

- Create messages and publish them to KeyLessTopic

Command : bin/kafka-console-producer.sh --broker-list localhost:9092 --topic
KeyLessTopic



```
acadgild@localhost:~/install/kafka/kafka_2.12-0.10.1.1  
File Edit View Search Terminal Help  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost kafka_2.12-0.10.1.1]$ 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}  
█
```

Task – 2:

Create a console consumer that reads KeyLessTopic from beginning

- Create a console consumer and read messages from KeyLessTopic

Command : `bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic KeyLessTopic --from-beginning`

A terminal window titled 'acadgild@localhost:~/install/kafka/kafka_2.12-0.10.1.1' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic KeyLessTopic --from-beginning' is entered. A red arrow points to the end of the command. The output shows five JSON messages: {'name': 'John', 'exp': 16}, {'name': 'Finn', 'exp': 20}, {'name': 'Cylin', 'exp': 18}, {'name': 'Mark', 'exp': 2}, and {'name': 'Akshay', 'exp': 14}.

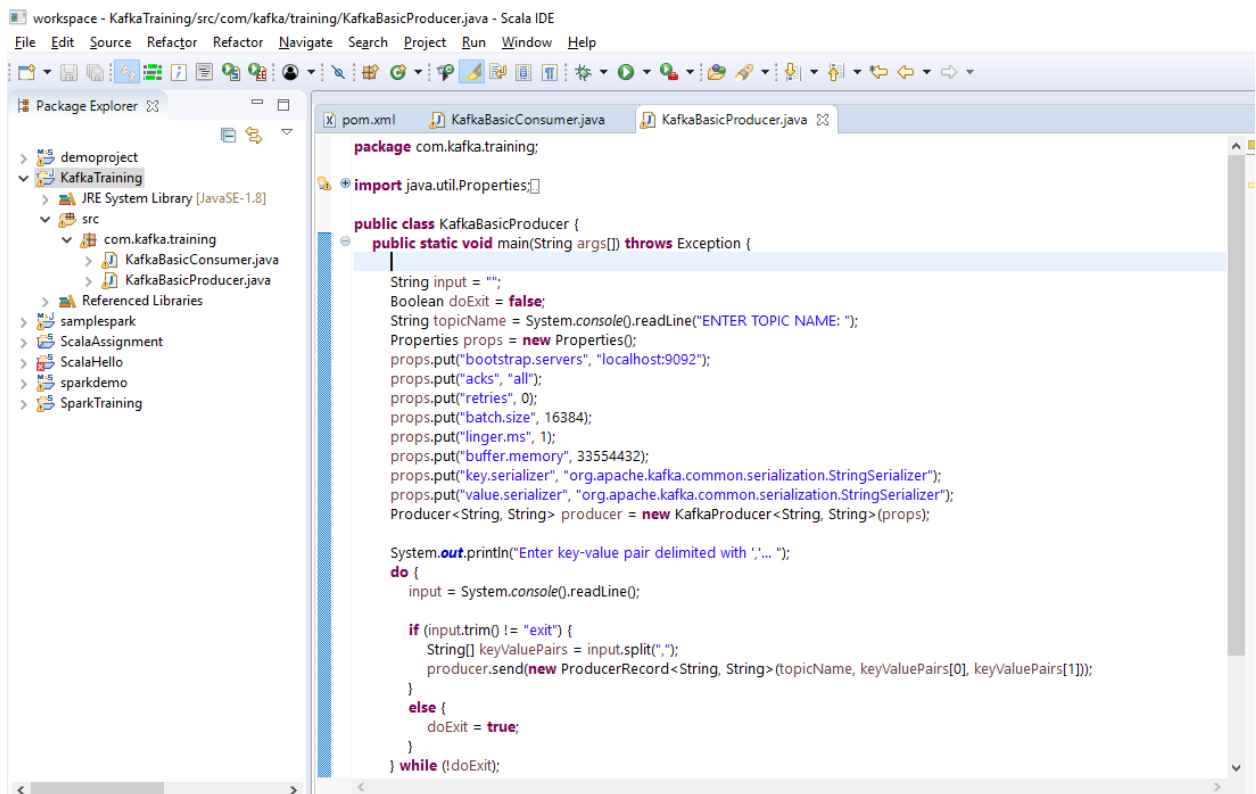
```
acadgild@localhost:~/install/kafka/kafka_2.12-0.10.1.1
File Edit View Search Terminal Help
[acadgild@localhost kafka_2.12-0.10.1.1]$ bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic KeyLessTopic --from-beginning
{"name": "John", "exp": 16}
{"name": "Finn", "exp": 20}
{"name": "Cylin", "exp": 18}
{"name": "Mark", "exp": 2}
{"name": "Akshay", "exp": 14}
```

Task – 3:

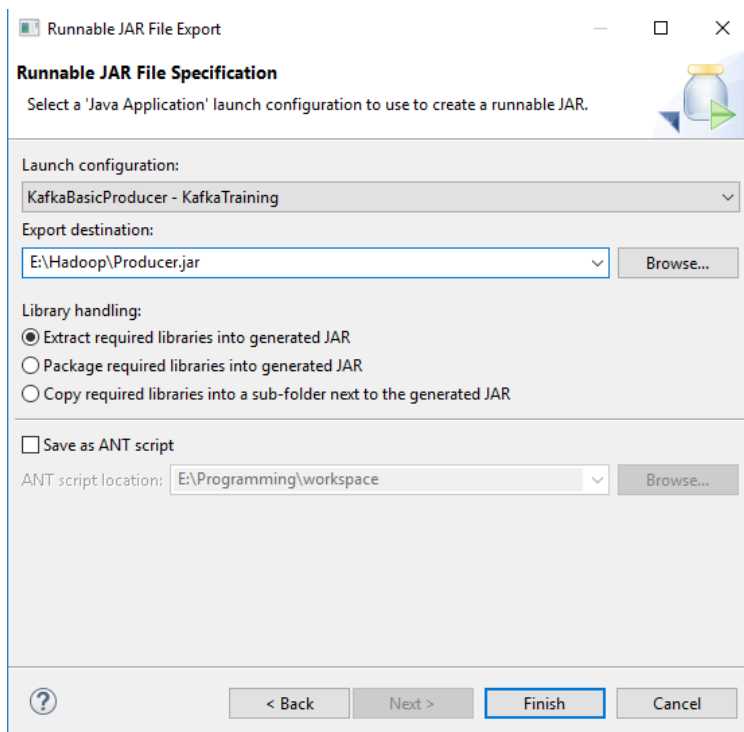
Create a kafka topic named KeyedTopic. Inside KeyedTopic insert following data:

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

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

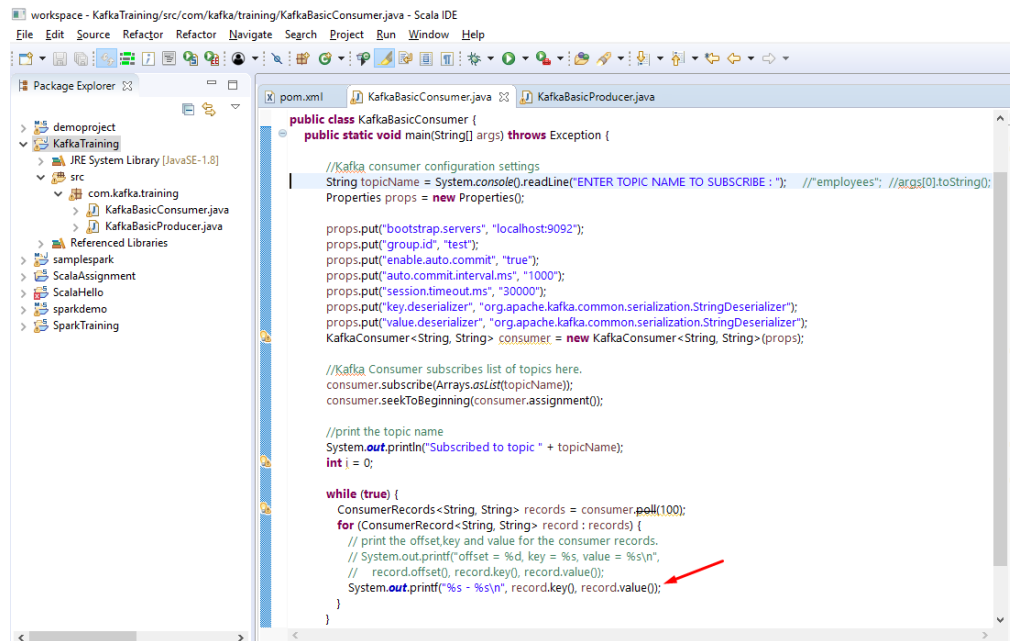


Export jar file of producer program:

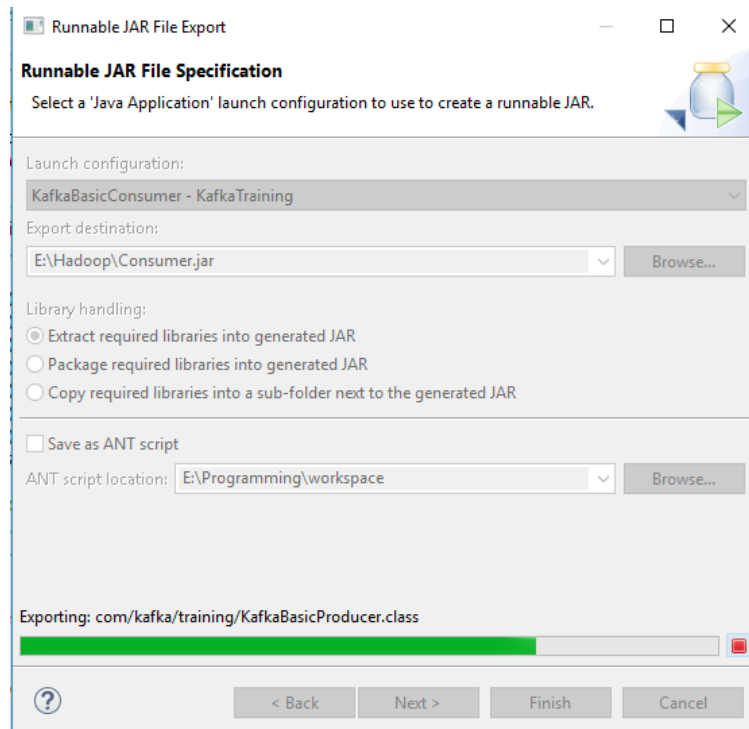


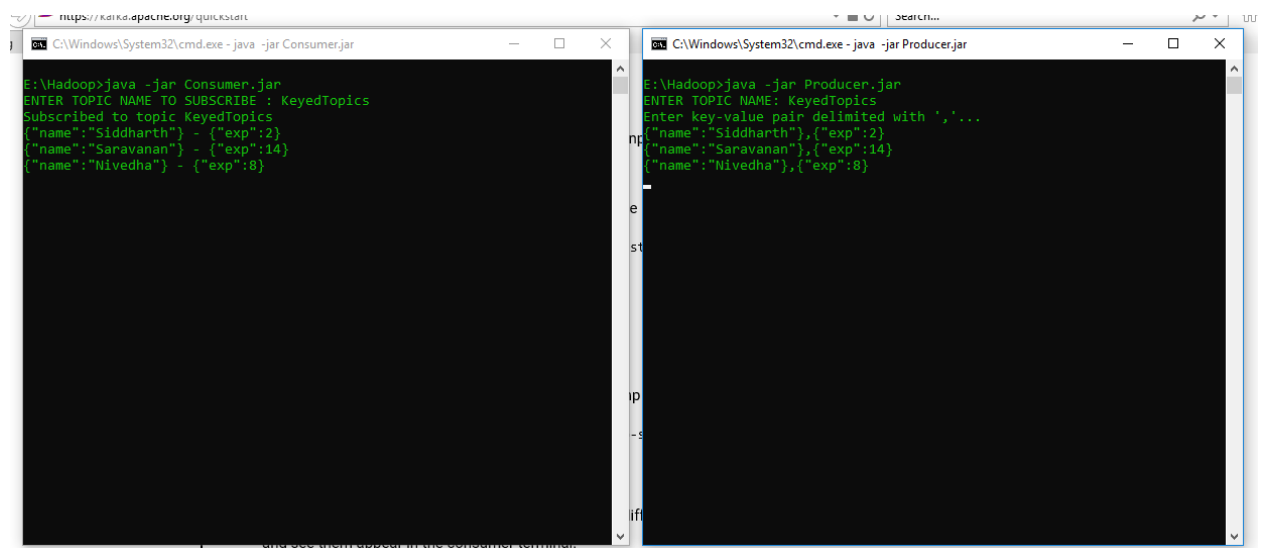
Task – 4:

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



Export jar file of consumer program





The image shows two terminal windows side-by-side. The left window is titled 'C:\Windows\System32\cmd.exe - java -jar Consumer.jar' and the right window is titled 'C:\Windows\System32\cmd.exe - java -jar Producer.jar'. Both windows show the execution of Java programs for Kafka. The left window shows a consumer subscribing to a topic and receiving three JSON messages. The right window shows a producer sending three JSON messages to the same topic.

```
E:\Hadoop>java -jar Consumer.jar
ENTER TOPIC NAME TO SUBSCRIBE : KeyedTopics
Subscribed to topic KeyedTopics
{"name":"Siddharth"} - {"exp":2}
{"name":"Saravanan"} - {"exp":14}
{"name":"Nivedha"} - {"exp":8}
```

```
E:\Hadoop>java -jar Producer.jar
ENTER TOPIC NAME: KeyedTopics
Enter key-value pair delimited with ','
{"name":"Siddharth"}, {"exp":2}
{"name":"Saravanan"}, {"exp":14}
{"name":"Nivedha"}, {"exp":8}
```