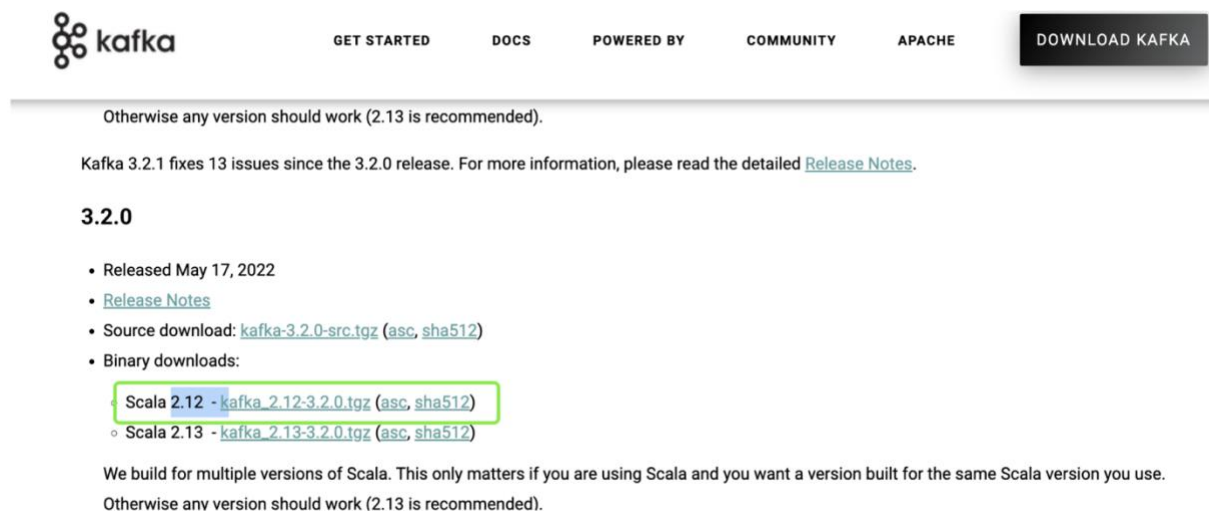


Week 12

Project: Kafka + Spark Streaming + PySpark

Kafka QuickStart — Apache Kafka + Kafka-Python

1. The latest version of Kafka binary distribution is available at <https://kafka.apache.org/downloads>.



2. Starting Zookeeper. Unzip it, get into the folders cd into it

```
Last login: Tue Dec 6 20:20:11 on tty000
karthikchatla@Karthiks-MacBook-Pro ~ % cd Downloads
karthikchatla@Karthiks-MacBook-Pro Downloads % cd kafka_2.12-3.3.1
karthikchatla@Karthiks-MacBook-Pro kafka_2.12-3.3.1 % ls
LICENSE      bin          libs         logs
NOTICE       config       licenses     site-docs
karthikchatla@Karthiks-MacBook-Pro kafka_2.12-3.3.1 % bin/zookeeper-server-start.sh config/zookeeper.properties
[2022-12-06 20:22:12,774] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,775] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,780] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-06 20:22:12,781] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-06 20:22:12,781] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-06 20:22:12,781] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,782] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
```

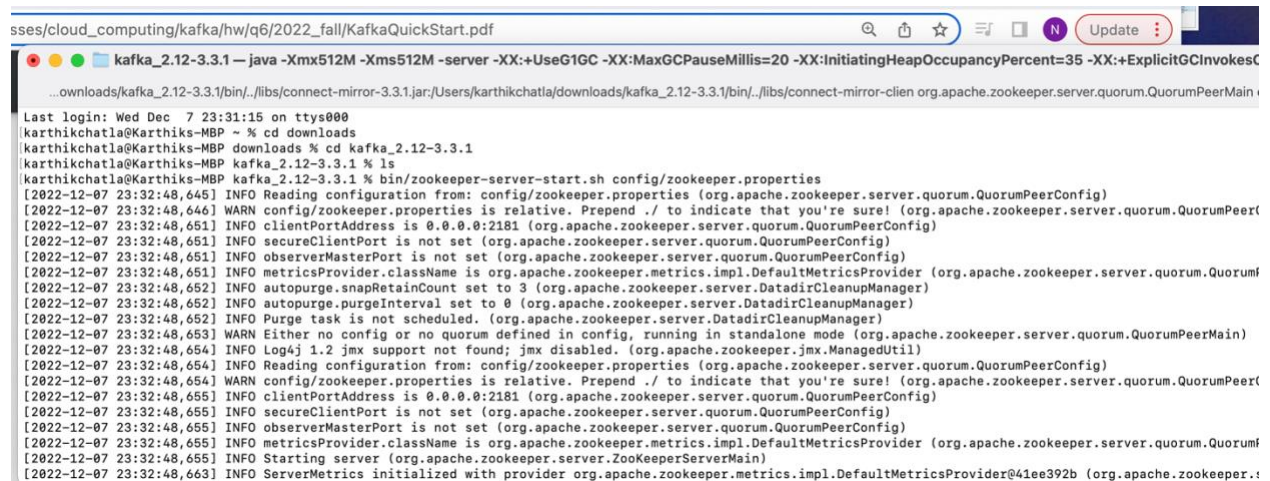
```

Last login: Tue Dec  6 20:20:11 on ttys000
karthikchatla@Karthiks-MacBook-Pro ~ % cd Downloads
karthikchatla@Karthiks-MacBook-Pro Downloads % cd kafka_2.12-3.3.1
karthikchatla@Karthiks-MacBook-Pro kafka_2.12-3.3.1 % ls
LICENSE      bin          libs        logs
NOTICE       config      licenses    site-docs
karthikchatla@Karthiks-MacBook-Pro kafka_2.12-3.3.1 % bin/zookeeper-server-start.sh config/zookeeper.properties
[2022-12-06 20:22:12,774] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,775] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.Q
[2022-12-06 20:22:12,779] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,779] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quor
[2022-12-06 20:22:12,780] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-06 20:22:12,781] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-06 20:22:12,781] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-06 20:22:12,781] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumP
[2022-12-06 20:22:12,782] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2022-12-06 20:22:12,782] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,782] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.Q
[2022-12-06 20:22:12,783] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,783] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,783] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-06 20:22:12,783] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quor
[2022-12-06 20:22:12,783] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2022-12-06 20:22:12,792] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@41ee392b (org.apache.z
[2022-12-06 20:22:12,795] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistance.FileTxnSnapLog)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,802] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,803] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,803] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,803] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-06 20:22:12,816] INFO Server environment: zookeeper.version=3.6.3--4d41e4ad7087061hc6b9f080der2d60f2e3e3e66a built on 06/08/2021 16:35 GMT (or

```

3. Starting Kafka Brokers

Create another terminal, do not close zookeeper



```

sses/cloud_computing/kafka/hw/q6/2022_fall/KafkaQuickStart.pdf
kafka_2.12-3.3.1 -- java -Xmx512M -Xms512M -server -XX:+UseG1GC -XX:MaxGCPauseMillis=20 -XX:InitiatingHeapOccupancyPercent=35 -XX:+ExplicitGCInvoke...
...ownloads/kafka_2.12-3.3.1/bin/./libs/connect-mirror-3.3.1.jar;Users[karthikchatla]/downloads/kafka_2.12-3.3.1/bin/./libs/connect-mirror-clien org.apache.zookeeper.server.quorum.QuorumPeerMain
Last login: Wed Dec  7 23:31:15 on ttys000
karthikchatla@Karthiks-MBP ~ % cd downloads
karthikchatla@Karthiks-MBP downloads % cd kafka_2.12-3.3.1
karthikchatla@Karthiks-MBP kafka_2.12-3.3.1 % ls
karthikchatla@Karthiks-MBP kafka_2.12-3.3.1 % bin/zookeeper-server-start.sh config/zookeeper.properties
[2022-12-07 23:32:48,646] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,646] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerC
[2022-12-07 23:32:48,651] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,651] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,651] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,651] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.Quorumf
[2022-12-07 23:32:48,652] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-07 23:32:48,652] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-07 23:32:48,652] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-12-07 23:32:48,653] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2022-12-07 23:32:48,654] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2022-12-07 23:32:48,654] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,654] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerC
[2022-12-07 23:32:48,655] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,655] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,655] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-07 23:32:48,655] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.Quorumf
[2022-12-07 23:32:48,655] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2022-12-07 23:32:48,663] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@41ee392b (org.apache.zookeeper.

```

4. Creating Kafka Topics.

Create another terminal, do not close zookeeper and kafka brokers

bin/kafka-topics.sh --create --topic input_recommend_product --zookeeper localhost:2181 -- partitions 3 --replication-factor 1

```
[2022-12-09 00:27:51,172] INFO App info kafka.server for 0 unregistered (org.apache.kafka.common.utils.AppInfoParser)
[2022-12-09 00:27:51,173] INFO shut down completed (kafka.server.KafkaServer)
[2022-12-09 00:27:51,173] ERROR Exiting Kafka due to fatal exception during startup. (kafka.Kafka$)
kafka.zookeeper.ZooKeeperClientTimeoutException: Timed out waiting for connection while in state: CONNECTING
    at kafka.zookeeper.ZooKeeperClient.$anonfun$waitUntilConnected$3(ZooKeeperClient.scala:254)
    at kafka.zookeeper.ZooKeeperClient.waitUntilConnected(ZooKeeperClient.scala:250)
    at kafka.zookeeper.ZooKeeperClient.<init>(ZooKeeperClient.scala:108)
    at kafka.zk.KafkaZkClient$.apply(KafkaZkClient.scala:1980)
    at kafka.server.KafkaServer.initZkClient(KafkaServer.scala:503)
    at kafka.server.KafkaServer.startup(KafkaServer.scala:203)
    at kafka.Kafka$.main(Kafka.scala:109)
    at kafka.Kafka.main(Kafka.scala)
[2022-12-09 00:27:51,173] INFO shutting down (kafka.server.KafkaServer)
karthikchatla@Karthiks-MBP kafka_2.12-3.3.1 % bin/kafka-topics.sh --create --topic input_recommend_product --zookeeper localhost:2181 --partition
--replication-factor 1
Exception in thread "main" joptsimple.UnrecognizedOptionException: zookeeper is not a recognized option
    at joptsimple.OptionException.unrecognizedOption(OptionException.java:108)
    at joptsimple.OptionParser.handleLongOptionToken(OptionParser.java:510)
    at joptsimple.OptionParserState$2.handleArgument(OptionParserState.java:56)
    at joptsimple.OptionParser.parse(OptionParser.java:396)
    at kafka.admin.TopicCommand$TopicCommandOptions.<init>(TopicCommand.scala:567)
    at kafka.admin.TopicCommand$.main(TopicCommand.scala:47)
    at kafka.admin.TopicCommand.main(TopicCommand.scala)
karthikchatla@Karthiks-MBP kafka_2.12-3.3.1 %
```

Meet with problem. Solve it : <https://stackoverflow.com/questions/69297020/exception-inthread-main-joptsimple-unrecognizedoptionexception-zookeeper-is>

▲ Newer versions(2.2+) of Kafka no longer requires ZooKeeper connection string

109 --zookeeper localhost:2181

▼ It throws the following exception while creating a topic

📌 Exception in thread "main" joptsimple.UnrecognizedOptionException: zookeeper is not a recognized option

🔄 Instead, add Kafka Broker --bootstrap-server localhost:9092 connection string.

```
./kafka-topics.sh --create --topic test-topic --bootstrap-server localhost
```

bin/kafka-topics.sh --create --topic input_recommend_product --bootstrap-server localhost:9092 --partitions 3 --replication-factor 1

```
karthikchatla@Karthiks-MBP kafka_2.12-3.3.1 % bin/kafka-topics.sh --create --topic input_recommend_product --bootstrap-server localhost:9092
--partitions 3 --replication-factor 1
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either
but not both.
[2022-12-09 00:36:17,429] WARN [AdminClient clientId=adminclient-1] Connection to node -1 (localhost/127.0.0.1:9092) could not be established. Bro
may not be available. (org.apache.kafka.clients.NetworkClient)
[2022-12-09 00:36:17,553] WARN [AdminClient clientId=adminclient-1] Connection to node -1 (localhost/127.0.0.1:9092) could not be established. Bro
may not be available. (org.apache.kafka.clients.NetworkClient)
[2022-12-09 00:36:17,659] WARN [AdminClient clientId=adminclient-1] Connection to node -1 (localhost/127.0.0.1:9092) could not be established. Bro
may not be available. (org.apache.kafka.clients.NetworkClient)
[2022-12-09 00:36:17,867] WARN [AdminClient clientId=adminclient-1] Connection to node -1 (localhost/127.0.0.1:9092) could not be established. Bro
may not be available. (org.apache.kafka.clients.NetworkClient)
[2022-12-09 00:36:18,388] WARN [AdminClient clientId=adminclient-1] Connection to node -1 (localhost/127.0.0.1:9092) could not be established. Bro
may not be available. (org.apache.kafka.clients.NetworkClient)
[2022-12-09 00:36:19,217] WARN [AdminClient clientId=adminclient-1] Connection to node -1 (localhost/127.0.0.1:9092) could not be established. Bro
may not be available. (org.apache.kafka.clients.NetworkClient)
```

To see the topics: bin/kafka-topics.sh --bootstrap-server localhost:9092 --list

5 Creating Producer and Consumer using Kafka-python

5.1 Create producer.py

```
from kafka import KafkaProducer
producer = KafkaProducer(bootstrap_servers='localhost:9092')
producer.send('input_recommend_product', b'(1, Main Menu), (2, Phone) ,(3, Smart Phone), (4, iPhone)')
producer.close()
```

5.2 Create Consumerr.py

Code

```
from kafka import KafkaConsumer
consumer = KafkaConsumer('input_recommend_product',
bootstrap_servers=['localhost:9092'])
for msg in consumer:
    print(msg)
```

5.3 Run comsumer.py first (you can run it in your IDE)

5.4 Create another terminal, run the producer.py

5.5 Go to the consumer terminal, you can see the result