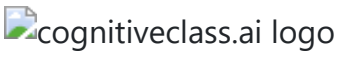
Hands-on Lab: Working with streaming data using Kafka



Estimated time needed: **20** minutes

Objectives

After completing this lab you will be able to:

- Download and install Kafka
- Start the Zookeeper server for Kafka metadata management
- Start the Kafka message broker service
- Create a topic
- Start a producer
- Start a consumer

About Skills Network Cloud IDE

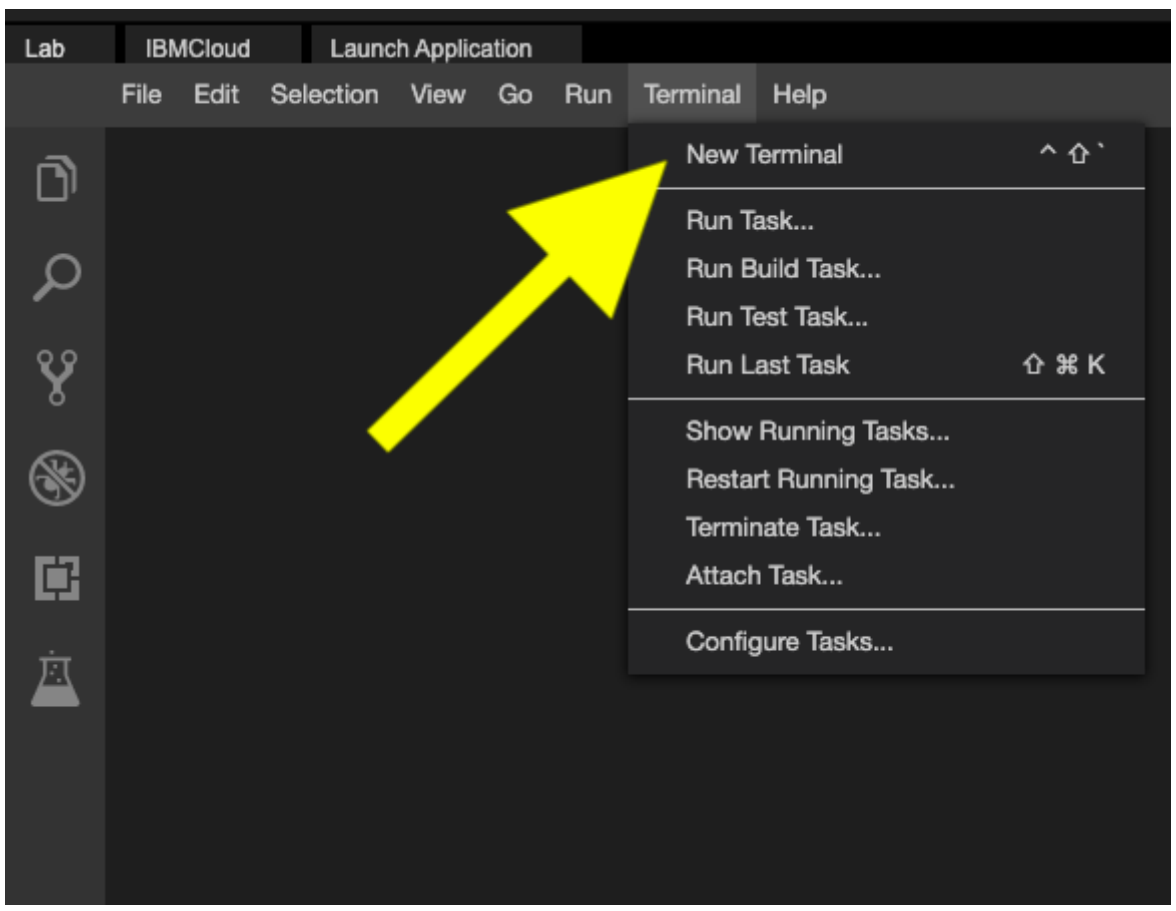
Skills Network Cloud IDE (based on Theia and Docker) provides an environment for hands on labs for course and project related labs. Theia is an open source IDE (Integrated Development Environment), that can be run on desktop or on the cloud. to complete this lab, we will be using the Cloud IDE based on Theia running in a Docker container.

Important Notice about this lab environment

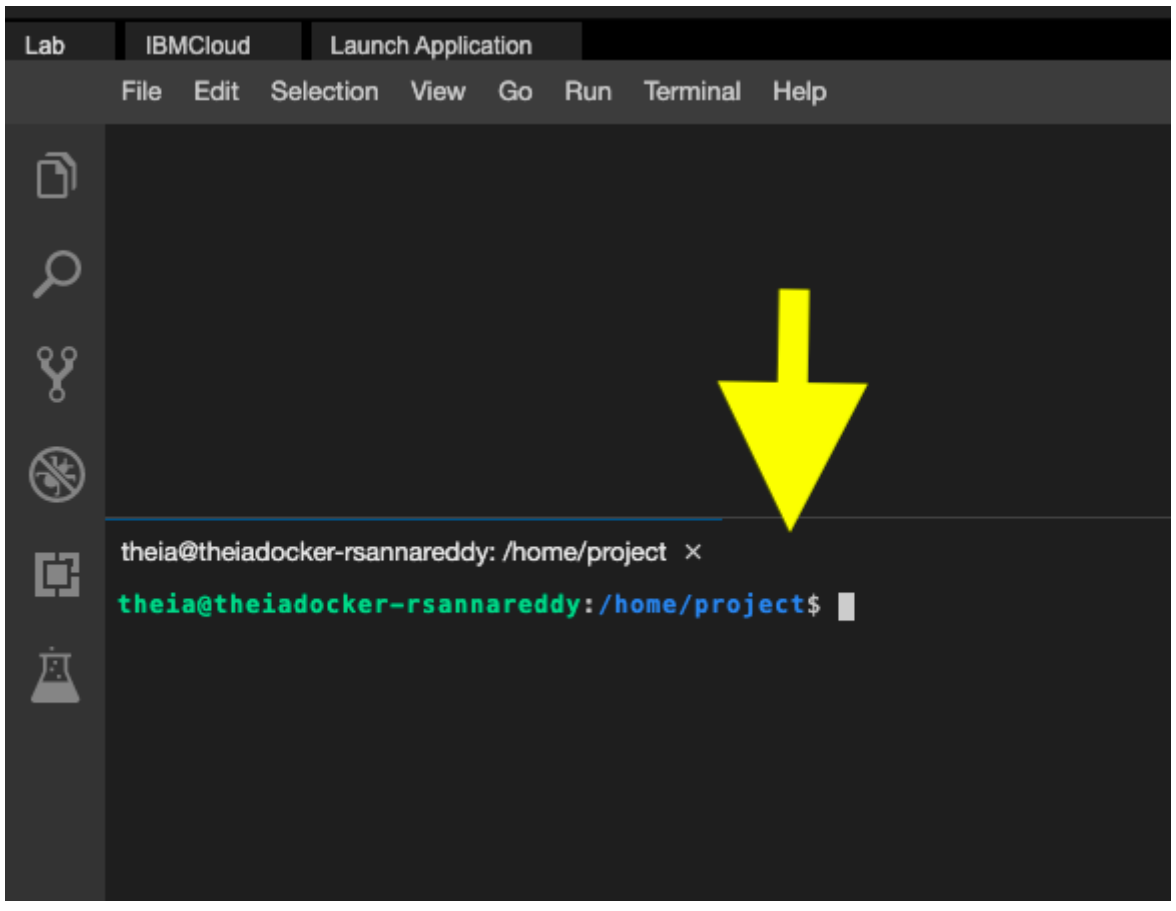
Please be aware that sessions for this lab environment are not persistent. A new environment is created for you every time you connect to this lab. Any data you may have saved in an earlier session will get lost. To avoid losing your data, please plan to complete these labs in a single session.

Exercise 1 - Download and extract Kafka

Open a new terminal, by clicking on the menu bar and selecting **Terminal**->**New Terminal**, as shown in the image below.



This will open a new terminal at the bottom of the screen.



Run the commands below on the newly opened terminal. (You can copy the code by clicking on the little copy button on the bottom right of the codeblock below and then paste it, wherever you wish.)

Download Kafka, by running the command below:

```
wget https://archive.apache.org/dist/kafka/2.8.0/kafka_2.12-2.8.0.tgz
```

Extract kafka from the zip file by running the command below.

```
tar -xzf kafka_2.12-2.8.0.tgz
```

This creates a new directory 'kafka_2.12-2.8.0' in the current directory.

Exercise 2 - start ZooKeeper

ZooKeeper is required for Kafka to work. Start the ZooKeeper server.

```
cd kafka_2.12-2.8.0
bin/zookeeper-server-start.sh config/zookeeper.properties
```

When ZooKeeper starts you should see an output like this:

```
theia@theiadocker-rsannareddy:/home/project/kafka_2.12-2.8.0$ bin/zookeeper-server-start.sh config/zookeeper.properties
JVMJ9VM007W Command-line option unrecognised: -Xlog:gc*:file=/home/project/kafka_2.12-2.8.0/bin/../logs/zookeeper-gc.log:time
,tags:filecount=10,filesize=100M
[2021-08-24 13:19:41,160] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.Qu
orumPeerConfig)
[2021-08-24 13:19:41,163] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.
zookeeper.server.quorum.QuorumPeerConfig)
[2021-08-24 13:19:41,178] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-08-24 13:19:41,178] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-08-24 13:19:41,183] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DatadirCleanupManager)
[2021-08-24 13:19:41,183] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DatadirCleanupManager)
[2021-08-24 13:19:41,183] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DatadirCleanupManager)
[2021-08-24 13:19:41,183] WARN Either no config or no quorum defined in config, running  in standalone mode (org.apache.zooke
eper.server.quorum.QuorumPeerMain)
[2021-08-24 13:19:41,190] INFO Log4j 1.2 jmx support found and enabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2021-08-24 13:19:41,217] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.Qu
orumPeerConfig)
[2021-08-24 13:19:41,217] WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.
zookeeper.server.quorum.QuorumPeerConfig)
```

You can sure it has started when you see an output like this:

```
[2021-08-24 13:19:41,253] INFO Server environment:os.memory.max=512MB (org.apache.zookeeper.server.ZooKeeperServer)
[2021-08-24 13:19:41,253] INFO Server environment:os.memory.total=512MB (org.apache.zookeeper.server.ZooKeeperServer)
[2021-08-24 13:19:41,256] INFO minSessionTimeout set to 6000 (org.apache.zookeeper.server.ZooKeeperServer)
[2021-08-24 13:19:41,256] INFO maxSessionTimeout set to 60000 (org.apache.zookeeper.server.ZooKeeperServer)
[2021-08-24 13:19:41,257] INFO Created server with tickTime 3000 minSessionTimeout 6000 maxSessionTimeout 60000 datadir /tmp/
zookeeper/version-2 snapdir /tmp/zookeeper/version-2 (org.apache.zookeeper.server.ZooKeeperServer)
[2021-08-24 13:19:41,275] INFO Using org.apache.zookeeper.server.NIOServerCnxnFactory as server connection factory (org.apach
e.zookeeper.server.ServerCnxnFactory)
[2021-08-24 13:19:41,281] INFO Configuring NIO connection handler with 10s sessionless connection timeout, 1 selector thread(
s), 4 worker threads, and 64 kB direct buffers. (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2021-08-24 13:19:41,290] INFO binding to port 0.0.0.0/0.0.0.0:2181 (org.apache.zookeeper.server.NIOServerCnxnFactory)
[2021-08-24 13:19:41,314] INFO zookeeper.snapshotSizeFactor = 0.33 (org.apache.zookeeper.server.ZKDatabase)
[2021-08-24 13:19:41,321] INFO Snapshotting: 0x0 to /tmp/zookeeper/version-2/snapshot.0 (org.apache.zookeeper.server.persiste
nce.FileTxnSnapLog)
[2021-08-24 13:19:41,324] INFO Snapshotting: 0x0 to /tmp/zookeeper/version-2/snapshot.0 (org.apache.zookeeper.server.persiste
nce.FileTxnSnapLog)
[2021-08-24 13:19:41,345] INFO PrepRequestProcessor (sid:0) started, reconfigEnabled=false (org.apache.zookeeper.server.PrepR
equestProcessor)
[2021-08-24 13:19:41,355] INFO Using checkIntervalMs=60000 maxPerMinute=10000 (org.apache.zookeeper.server.ContainerManager)
```

ZooKeeper, as of this version, is required for Kafka to work. ZooKeeper is responsible for the overall management of Kafka cluster. It monitors the Kafka brokers and notifies Kafka if any broker or partition goes down, or if a new broker or partition goes up.

Exercise 3 - Start the Kafka broker service

Start a new terminal.
Run the commands below. This will start the Kafka message broker service.

```
cd kafka_2.12-2.8.0
bin/kafka-server-start.sh config/server.properties
```

When Kafka starts, you should see an output like this:

```
theia@theiadocker-rsannareddy:/home/project$ cd kafka_2.12-2.8.0
theia@theiadocker-rsannareddy:/home/project/kafka_2.12-2.8.0$ bin/kafka-server-start.sh config/server.properties
JVMJ9VM007W Command-line option unrecognised: -Xlog:gc*:file=/home/project/kafka_2.12-2.8.0/bin/../logs/kafkaServer-gc.log:t
ime,tags:filecount=10,filesize=100M
[2021-08-24 13:23:34,367] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2021-08-24 13:23:34,856] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS re
negotiation (org.apache.zookeeper.common.X509Util)
[2021-08-24 13:23:34,975] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHan
dler)
[2021-08-24 13:23:34,984] INFO starting (kafka.server.KafkaServer)
[2021-08-24 13:23:34,985] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2021-08-24 13:23:35,026] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper
.ZooKeeperClient)
[2021-08-24 13:23:35,037] INFO Client environment:zookeeper.version=3.5.9-83df9301aa5c2a5d284a9940177808c01bc35cef, built on
 01/06/2021 20:03 GMT (org.apache.zookeeper.ZooKeeper)
[2021-08-24 13:23:35,037] INFO Client environment:host.name=theiadocker-rsannareddy (org.apache.zookeeper.ZooKeeper)
[2021-08-24 13:23:35,037] INFO Client environment:java.version=11.0.11 (org.apache.zookeeper.ZooKeeper)
[2021-08-24 13:23:35,037] INFO Client environment:java.vendor=AdoptOpenJDK (org.apache.zookeeper.ZooKeeper)
[2021-08-24 13:23:35,037] INFO Client environment:java.home=/usr/lib/jvm/java-jdk-11.0.11+9 (org.apache.zookeeper.ZooKeeper)
[2021-08-24 13:23:35,037] INFO Client environment:java.class.path=/usr/lib/jvm/java-jdk-11.0.11+9/lib:/home/project:/home/pr
oject/kafka_2.12-2.8.0/bin/../libs/activation-1.1.1.jar:/home/project/kafka_2.12-2.8.0/bin/../libs/aopalliance-repackaged-2.
6.1.jar:/home/project/kafka_2.12-2.8.0/bin/../libs/argparse4j-0.7.0.jar:/home/project/kafka_2.12-2.8.0/bin/../libs/audience-
annotations-0.5.0.jar:/home/project/kafka_2.12-2.8.0/bin/../libs/commons-cli-1.4.jar:/home/project/kafka_2.12-2.8.0/bin/../l
```

You can be sure it has started when you see an output like this:

```
[2021-08-24 13:23:37,628] INFO [Transaction Marker Channel Manager 0]: Starting (kafka.coordinator.transaction.TransactionMa
rkerChannelManager)
[2021-08-24 13:23:37,628] INFO [TransactionCoordinator id=0] Startup complete. (kafka.coordinator.transaction.TransactionCoo
rdinator)
[2021-08-24 13:23:37,645] INFO Updated cache from existing <empty> to latest FinalizedFeaturesAndEpoch(features=Features{},
epoch=0). (kafka.server.FinalizedFeatureCache)
[2021-08-24 13:23:37,701] INFO [ExpirationReaper-0-AlterAcls]: Starting (kafka.server.DelayedOperationPurgatory$ExpiredOpera
tionReaper)
[2021-08-24 13:23:37,764] INFO [/config/changes-event-process-thread]: Starting (kafka.common.ZkNodeChangeNotificationListen
er$ChangeEventProcessThread)
[2021-08-24 13:23:37,825] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Starting socket server acceptors and processo
rs (kafka.network.SocketServer)
[2021-08-24 13:23:37,845] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Started data-plane acceptor and processor(s)
for endpoint : ListenerName(PLAINTEXT) (kafka.network.SocketServer)
[2021-08-24 13:23:37,846] INFO [SocketServer listenerType=ZK_BROKER, nodeId=0] Started socket server acceptors and processor
s (kafka.network.SocketServer)
[2021-08-24 13:23:39,165] INFO Kafka version: 2.8.0 (org.apache.kafka.common.utils.AppInfoParser)
[2021-08-24 13:23:39,166] INFO Kafka commitId: ebb1d6e21cc92130 (org.apache.kafka.common.utils.AppInfoParser)
[2021-08-24 13:23:39,166] INFO Kafka startTimeMs: 1629811417846 (org.apache.kafka.common.utils.AppInfoParser)
[2021-08-24 13:23:39,192] INFO [KafkaServer id=0] started (kafka.server.KafkaServer)
[2021-08-24 13:23:39,384] INFO [broker-0-to-controller-send-thread]: Recorded new controller, from now on will use broker th
eiadocker-rsannareddy:9092 (id: 0 rack: null) (kafka.server.BrokerToControllerRequestThread)
```

Exercise 4 - Create a topic

You need to create a topic before you can start to post messages.

To create a topic named `news`, start a new terminal and run the command below.

```
cd kafka_2.12-2.8.0
bin/kafka-topics.sh --create --topic news --bootstrap-server localhost:9092
```

You will see the message: 'Created topic news.'

Exercise 5 - Start Producer

You need a producer to send messages to Kafka. Run the command below to start a producer.

```
bin/kafka-console-producer.sh --topic news --bootstrap-server localhost:9092
```

Once the producer starts, and you get the '>' prompt, type any text message and press enter. Or you can copy the text below and paste. The below text sends three messages to kafka.

```
Good morning
Good day
Enjoy the Kafka lab
```

Exercise 6 - Start Consumer

You need a consumer to read messages from kafka.

Open a new terminal.

Run the command below to listen to the messages in the topic `news`.

```
cd kafka_2.12-2.8.0
bin/kafka-console-consumer.sh --topic news --from-beginning --bootstrap-server localhost:9092
```

You should see all the messages you sent from the producer appear here.

You can go back to the producer terminal and type some more messages, one message per line, and you will see them appear here.

Exercise 7 - Explore Kafka directories.

Kafka uses the directory /tmp/kakfa-logs to store the messages.

Explore the folder news-0 inside /tmp/kakfa-logs.

This is where all the messages are stored.

Explore the folder /home/project/kafka_2.12-2.8.0

This folder has the below 3 sub directories.

Directory	Contents
bin	shell scripts to control kafka and zookeeper
config	configuration files
logs	log files forkafka and zookeeper

Exercise 8 - Clean up

Delete the kafka installation file.

```
rm kafka_2.12-2.8.0.tgz
```

Practice exercises

1. Problem:

> *Create a new topic named `weather`.*

▼ Click here for Hint

> Use `kafka-topics.sh` command with the `create` option.

▼ Click here for Solution

Make sure that you are in the 'kafka_2.12-2.8.0' directory. Run the following command:

```
bin/kafka-topics.sh --create --topic weather --bootstrap-server localhost:9092
```

2. Problem:

> *Post messages to the topic `weather`.*

▼ Click here for Hint

> Use `kafka-console-producer.sh` and point to topic `weather`.

▼ Click here for Solution

Make sure that you are in the 'kafka_2.12-2.8.0' directory. Run the following command:

```
bin/kafka-console-producer.sh --topic weather --bootstrap-server localhost:9092
```

Post some test messages.

3. Problem:

> *Read the messages from the topic `weather`.*

▼ Click here for Hint

> Use `kafka-console-consumer.sh` and read from the topic 'weather'

▼ Click here for Solution

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

Make sure that the messages you sent from the producer appear here.

Authors

Ramesh Sannareddy

Other Contributors

Rav Ahuja

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2021-08-24	0.2	Ramesh Sannareddy	Incorporated feedback from Yan
2021-06-22	0.1	Ramesh Sannareddy	Created initial version of the lab