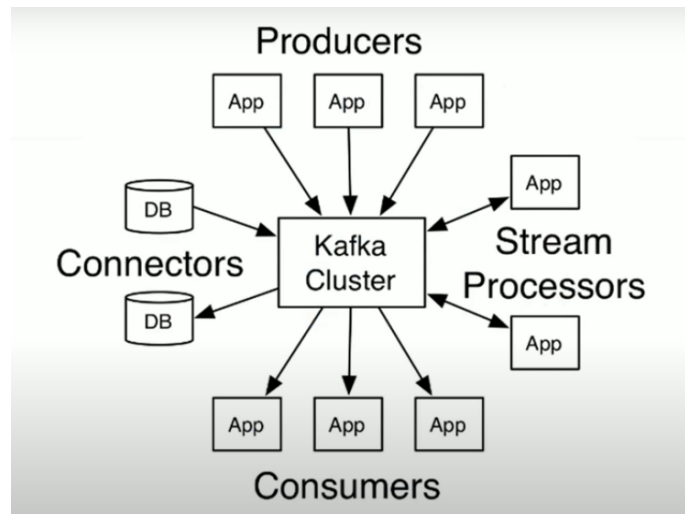


# Kafka

Kafka was introduced by LinkedIn. Open-source distributed event streaming platform. It is written in Scala and Java.



## Benefits of Kafka

- Reliability
- Scalability
- Durability
- Performance

Kafka is very fast and guarantees zero downtime and zero data loss.

## Use Cases

Kafka can be used in many Use Cases. Some of them are listed below :

- Metrics – Kafka is often used for operational monitoring data. This involves aggregating statistics from distributed applications to produce centralized feeds of operational data.
- Log Aggregation Solution – Kafka can be used across an organization to collect logs from multiple services and make them available in a standard format to multiple consumers.
- Stream Processing – Popular frameworks such as Storm and Spark Streaming read data from a topic, process it, and write processed data to a new topic where it becomes available for users and applications. Kafka's strong durability is also very useful in the context of stream processing.

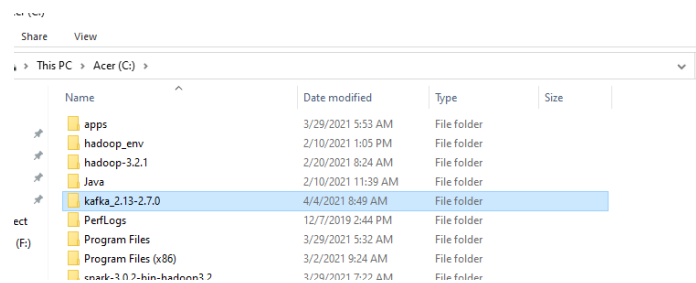
To work with Kafka, must have a good understanding of Java, Scala, Dis-tributed messaging system, and Linux environment.

## Main components in Kafka

- Producers - Produce messages and send them to the kafka cluster.
- Consumers - Read messages from kafka clusters and proceed with their requirements.
- Kafka cluster - Bunch of brokers that are stored in multiple machines. Take messages from producers and store them in kafka logs.
- Connectors - Use to import and export data to kafka.
- Stream Processors - Help in real time processing applications in Kafka.

## Installing Kafka on Windows

- 1.Go to <https://kafka.apache.org/downloads> and download Apache Kafka on the system.
- 2.Extract folder and move the Kafka folder to the local disk C.



- 3.Open Command Prompt then move to the Kafka directory. Check the java version by **java -version**.(If the installation of java is correct, the java version will show as below).

```
C:\kafka_2.13-2.7.0>java -version
java version "1.8.0_281"
Java(TM) SE Runtime Environment (build 1.8.0_281-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.281-b09, mixed mode)
```

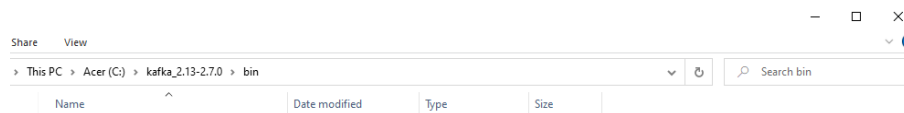
4. Execute **bin\windows\kafka-topics.bat** command.

```
C:\Windows\System32\cmd.exe - bin\windows\zookeeper-server-start.bat config\zookeeper.properties

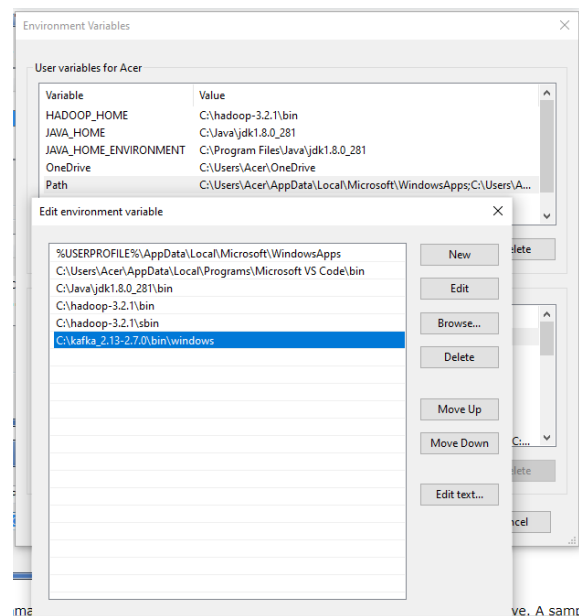
C:\kafka_2.13-2.7.0\bin\windows>kafka-topics.bat
Create, delete, describe, or change a topic.
Option                                Description
-----                                -
--alter                                Alter the number of partitions,
                                        replica assignment, and/or
                                        configuration for the topic.
--at-min-isr-partitions                if set when describing topics, only
                                        show partitions whose isr count is
                                        equal to the configured minimum. Not
                                        supported with the --zookeeper
                                        option.
--bootstrap-server <String: server to REQUIRED: The Kafka server to connect
                                        connect to> to. In case of providing this, a
                                        direct Zookeeper connection won't be
                                        required.
--command-config <String: command Property file containing configs to be
                                        config property file> passed to Admin Client. This is used
                                        only with --bootstrap-server option
```

## Setting Kafka directory path

5. Open **Kafka folder>bin>windows** and copy its address.



6. Edit Environment Variables as below. (Edit path by pasting above copied address)



7. Open Command Prompt again, apply the command '**kafka-topics.bat**' from any drive.

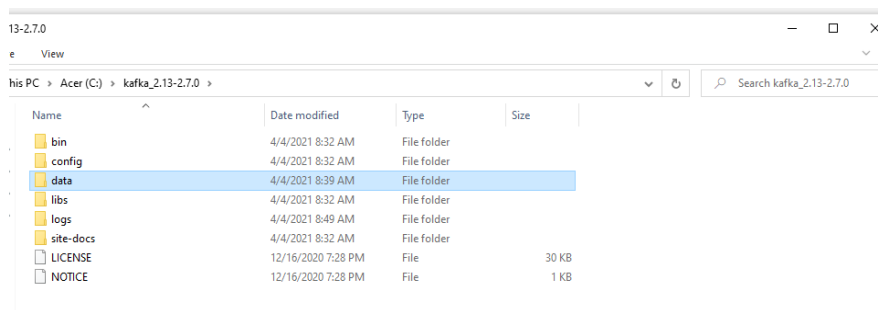
```
Command Prompt
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\Users\Acer>kafka-topics.bat
Create, delete, describe, or change a topic.
Option                                Description
-----                                -
--alter                                Alter the number of partitions,
                                        replica assignment, and/or
                                        configuration for the topic.
--at-min-isr-partitions                if set when describing topics, only
                                        show partitions whose isr count is
                                        equal to the configured minimum. Not
                                        supported with the --zookeeper
```

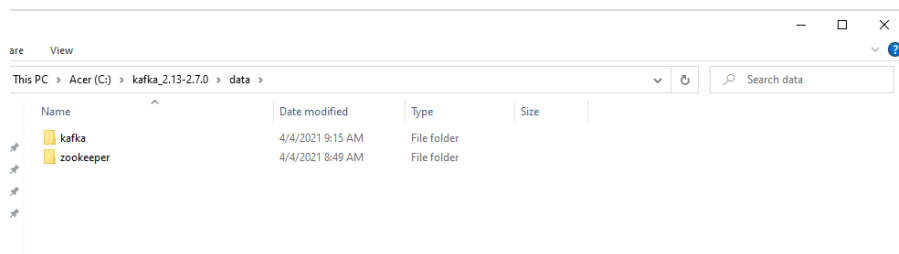
If the path is successfully set, it will display the above output. Now, the kafka commands can be executed from anywhere.

## Installing Zookeeper

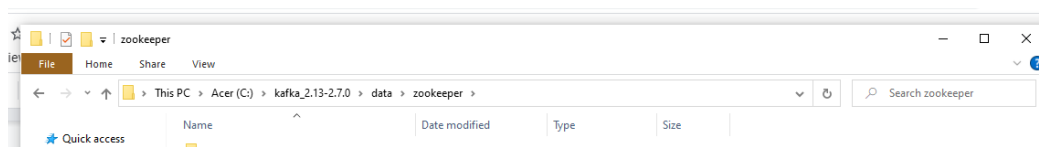
8. Go to the Kafka directory and create a new folder as '**data**'.



9. Open the data folder and create two more folders under it as '**zookeeper**' and '**kafka**'.



10. Copy the address of the zookeeper folder. After copying, go back to the Kafka directory.



11. Move to the **config** folder under the Kafka directory. Open the config folder and go to the **zookeeper.properties** file. Edit by pasting the copied address here.

```
# this work for additional information regarding copyright ownership.
# The ASF licenses this file to You under the Apache License, Version 2.0
# (the "License"); you may not use this file except in compliance with
# the License. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
# the directory where the snapshot is stored.
dataDir=C:/kafka_2.13-2.7.0/data/zookeeper
# the port at which the clients will connect
clientPort=2181
# disable the per-ip limit on the number of connections since this is a non-production config
maxClientCnxns=0
# Disable the adminserver by default to avoid port conflicts.
# Set the port to something non-conflicting if choosing to enable this
```

12. The Zookeeper server is ready to start. Open the command prompt and go to the Kafka directory. Then, type the command: **'zookeeper-server-start.bat config\zookeeper.properties'**.

```
C:\kafka_2.13-2.7.0>zookeeper-server-start.bat config\zookeeper.properties
'zookeeper-server-start.bat' is not recognized as an internal or external command,
operable program or batch file.

C:\kafka_2.13-2.7.0>bin\windows\zookeeper-server-start.bat config\zookeeper.properties
```

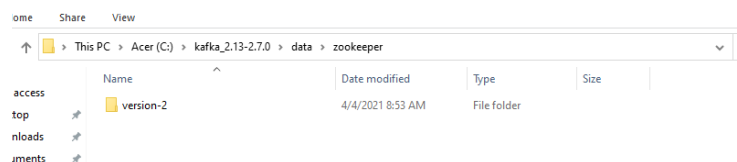
13. Here I fixed this issue by, applying **bin\windows\zookeeper-server-start.bat config\zookeeper.properties**.

```
C:\Windows\System32\cmd.exe - bin\windows\zookeeper-server-start.bat config\zookeeper.properties

C:\kafka_2.13-2.7.0>zookeeper-server-start.bat config\zookeeper.properties
'zookeeper-server-start.bat' is not recognized as an internal or external command,
operable program or batch file.

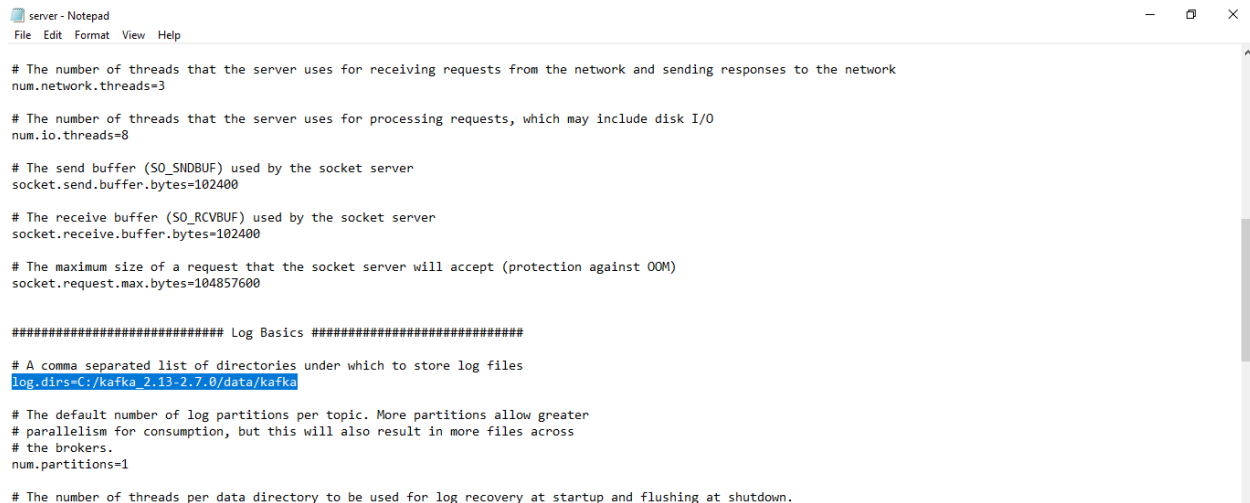
C:\kafka_2.13-2.7.0>bin\windows\zookeeper-server-start.bat config\zookeeper.properties
[2021-04-04 08:49:29,531] INFO Reading configuration from: config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 08:49:29,559] WARN config\zookeeper.properties is relative. Prepend .\ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 08:49:29,588] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 08:49:29,588] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 08:49:29,605] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2021-04-04 08:49:29,605] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2021-04-04 08:49:29,606] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
```

Open **data>zookeeper**. A new folder, **'version-2'** will be created automatically. If there is no such folder, something went wrong.



14. As the zookeeper server is started, go to the **kafka directory>data>kafka**. Copy the address of the kafka folder.

15. Move back to the **Kafka directory>config>server.properties**. Paste the copied address of the kafka folder as a value of **log.dirs**.



```
server - Notepad
File Edit Format View Help

# The number of threads that the server uses for receiving requests from the network and sending responses to the network
num.network.threads=3

# The number of threads that the server uses for processing requests, which may include disk I/O
num.io.threads=8

# The send buffer (SO_SNDBUF) used by the socket server
socket.send.buffer.bytes=102400

# The receive buffer (SO_RCVBUF) used by the socket server
socket.receive.buffer.bytes=102400

# The maximum size of a request that the socket server will accept (protection against OOM)
socket.request.max.bytes=104857600

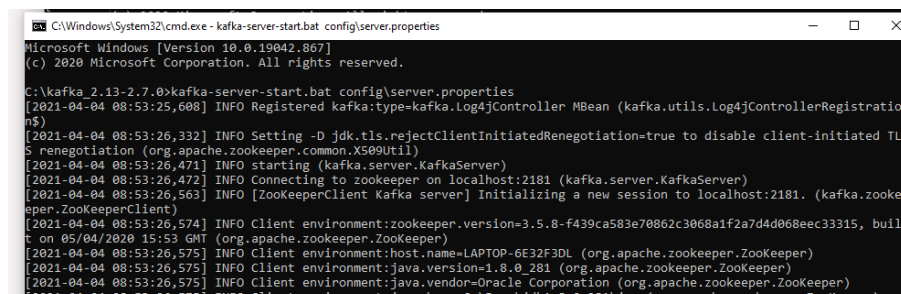
##### Log Basics #####

# A comma separated list of directories under which to store log files
log.dirs=C:\kafka_2.13-2.7.0\data\kafka

# The default number of log partitions per topic. More partitions allow greater
# parallelism for consumption, but this will also result in more files across
# the brokers.
num.partitions=1

# The number of threads per data directory to be used for log recovery at startup and flushing at shutdown.
```

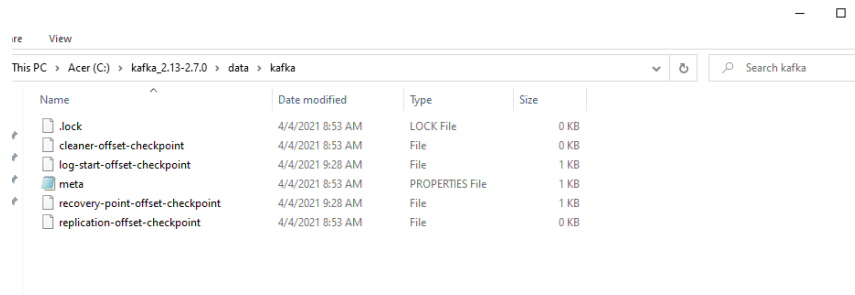
16. Open the command prompt and move to the Kafka directory. Type the command: **'kafka-server-start.bat config\server.properties'**.



```
C:\Windows\System32\cmd.exe - kafka-server-start.bat config\server.properties
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>kafka-server-start.bat config\server.properties
[2021-04-04 08:53:25,608] INFO Registered kafka:type-kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2021-04-04 08:53:26,332] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2021-04-04 08:53:26,471] INFO starting (kafka.server.KafkaServer)
[2021-04-04 08:53:26,472] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2021-04-04 08:53:26,563] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2021-04-04 08:53:26,574] INFO Client environment:zookeeper.version=3.5.8-f439ca583e70862c3068a1f2a7d4d068eec33315, built on 05/04/2020 15:53 GMT (org.apache.zookeeper.ZooKeeper)
[2021-04-04 08:53:26,575] INFO Client environment:host.name=LAPTOP-6E32F3DL (org.apache.zookeeper.ZooKeeper)
[2021-04-04 08:53:26,575] INFO Client environment:java.version=1.8.0_281 (org.apache.zookeeper.ZooKeeper)
[2021-04-04 08:53:26,575] INFO Client environment:java.vendor=Oracle Corporation (org.apache.zookeeper.ZooKeeper)
[2021-04-04 08:53:26,575] INFO Client environment:java.home=C:\Java\jdk1.8.0_281\jre (org.apache.zookeeper.ZooKeeper)
```

17. Open the **Kafka directory>data>kafka**. Now can find some files, which will be generated on the successful startup of the Kafka server.



## Start work with Kafka

1. Open a command prompt and start the Zookeeper as done before.

```
C:\Windows\System32\cmd.exe - .\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>.bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
[2021-04-04 12:03:38,769] INFO Reading configuration from: .\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 12:03:38,813] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 12:03:38,819] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 12:03:38,845] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2021-04-04 12:03:38,857] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2021-04-04 12:03:38,858] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2021-04-04 12:03:38,858] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2021-04-04 12:03:38,907] INFO Log4j 1.2 jmx support found and enabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2021-04-04 12:03:38,964] INFO Reading configuration from: .\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2021-04-04 12:03:38,966] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
```

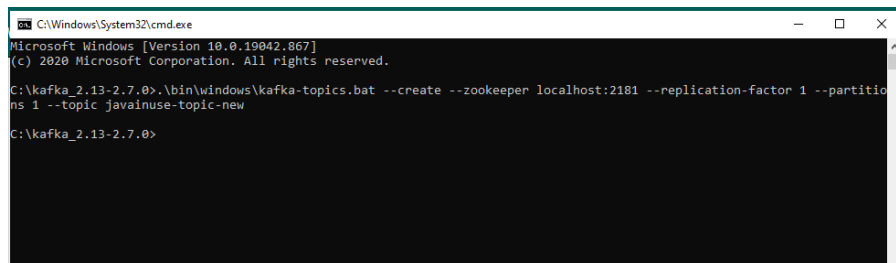
2. Open a new command prompt and start the Apache Kafka as done before.

```
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-server-start.bat .\config\server.properties
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>.bin\windows\kafka-server-start.bat .\config\server.properties
[2021-04-04 12:05:08,724] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2021-04-04 12:05:10,106] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2021-04-04 12:05:10,845] INFO starting (kafka.server.KafkaServer)
[2021-04-04 12:05:10,848] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2021-04-04 12:05:10,977] INFO [ZookeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZookeeperClient)
[2021-04-04 12:05:11,017] INFO Client environment:zookeeper.version=3.5.8-f439ca583e70862c3b68a1f2a7d4d068ec33315, built on 05/04/2020 15:53 GMT (org.apache.zookeeper.ZooKeeper)
[2021-04-04 12:05:11,018] INFO Client environment:host.name=LAPTOP-6E32F3DL (org.apache.zookeeper.ZooKeeper)
[2021-04-04 12:05:11,018] INFO Client environment:java.version=1.8.0_281 (org.apache.zookeeper.ZooKeeper)
```

3. Open a new command prompt and create a topic with the name javainuse-topic-new, that has only one partition & one replica.

(.\bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic javainuse-topic-new)



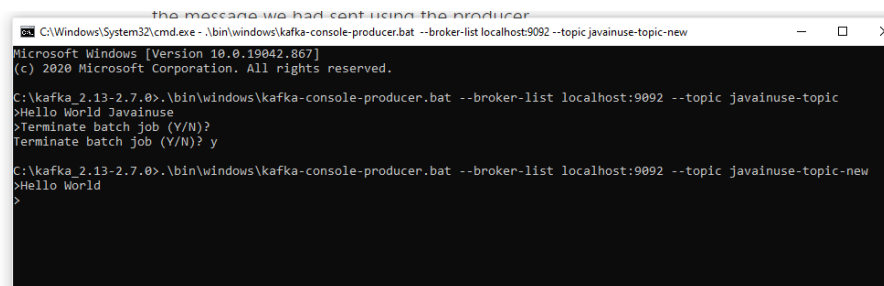
```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-topics.bat --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic javainuse-topic-new

C:\kafka_2.13-2.7.0>
```

4. Next Open a new command prompt and create a producer to send a message to the above created javainuse-topic and send a message - Hello World Javainuse to it.

(.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic)



```
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic-new
the message we had sent using the producer
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

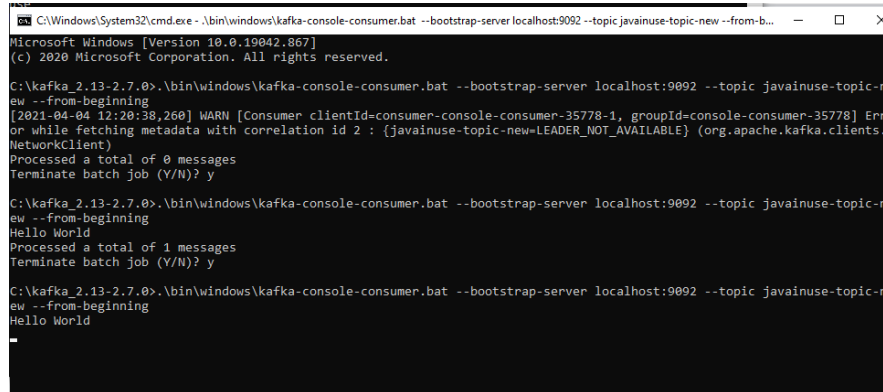
C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic
>Hello World Javainuse
>Terminate batch job (Y/N)?
Terminate batch job (Y/N)? y

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic-new
>Hello World
>
```



5. Finally Open a new command prompt and start the consumer which listens to the topic javainuse-topic we just created above. We will get the message we had sent using the producer.

(`.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic --from-beginning`)

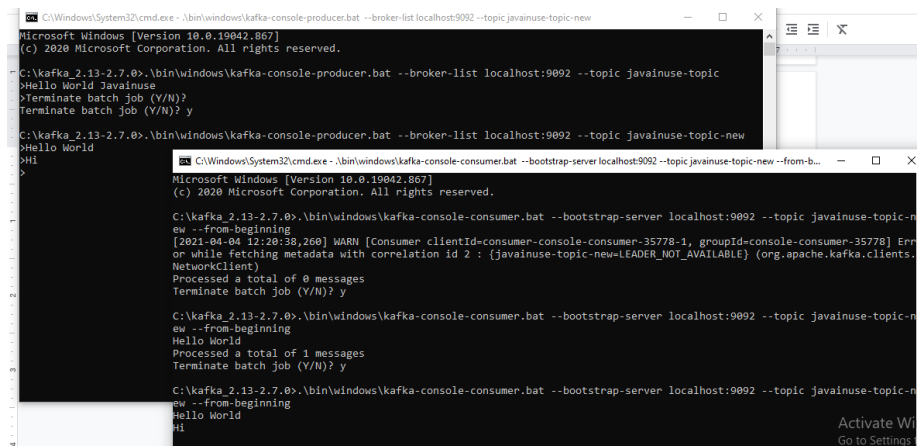


```
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-new --from-b...
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-n
ew --from-beginning
[2021-04-04 12:20:38,260] WARN [Consumer clientId=consumer-console-consumer-35778-1, groupId=console-consumer-35778] Err
or while fetching metadata with correlation id 2 : {javainuse-topic-new=LEADER_NOT_AVAILABLE} (org.apache.kafka.clients.
NetworkClient)
Processed a total of 0 messages
Terminate batch job (Y/N)? y

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-n
ew --from-beginning
Hello World
Processed a total of 1 messages
Terminate batch job (Y/N)? y

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-n
ew --from-beginning
Hello World
```



```
C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic-new
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic
>Hello World Javainuse
>Terminate batch job (Y/N)? y
>Terminate batch job (Y/N)? y

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic javainuse-topic-new
>Hello World
>Hi
>

C:\Windows\System32\cmd.exe - .\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-new --from-b...
Microsoft Windows [Version 10.0.19042.867]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-n
ew --from-beginning
[2021-04-04 12:20:38,260] WARN [Consumer clientId=consumer-console-consumer-35778-1, groupId=console-consumer-35778] Err
or while fetching metadata with correlation id 2 : {javainuse-topic-new=LEADER_NOT_AVAILABLE} (org.apache.kafka.clients.
NetworkClient)
Processed a total of 0 messages
Terminate batch job (Y/N)? y

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-n
ew --from-beginning
Hello World
Processed a total of 1 messages
Terminate batch job (Y/N)? y

C:\kafka_2.13-2.7.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic javainuse-topic-n
ew --from-beginning
Hello World
Hi
```