

**ID - 202218061**  
**NAME - JATAN SAHU**

**IT 609 - Big Data Processing Lab 6**  
**Teaching Assistant - Saran Pandian P**

**March 23, 2023**

**1 Aim**

- learn how to use apache Kafka for processing streaming Data.

**2 Problem description**

Kafka provides APIs for processing streaming Data. In this lab assignment, you build a producer and consumer function in python using Kafka which produces and consumes streaming data.

**3 Implementation**

**3.1 Exercise**

- Install Apache python-Kafka (zookeeper).
- start the zookeeper service.
- write the code for producer and consumer services.
- stream the current day, the timestamp from producer to consumer shell.
- Take a screenshot of the results from the producer and consumer terminal.

**4 Submission guidelines**

You have to submit three files

- Instructions for installing Kafka in Linux/Windows and final results in a pdf file.
- Consumer.py
- Producer.py

**REFERENCE**

- [https://github.com/nancyyanyu/kafka\\_stock/tree/master/pipeline](https://github.com/nancyyanyu/kafka_stock/tree/master/pipeline)
- <https://www.geeksforgeeks.org/how-to-install-and-run-apache-kafka-on-windows/>
- <https://betterdatascience.com/apache-kafka-in-python-how-to-stream-data-with-producers-and-consumers/>
- <https://www.youtube.com/watch?v=Hlz0pUXhM3U>

**Instructions for installing Kafka in Linux/Windows and final results.**

**Step 1: Go to the Downloads folder and select the downloaded Binary file.**

Link - <https://kafka.apache.org/downloads>

**Step 2: Extract the file and move the extracted folder to the directory where you wish to keep the files.**

**Path KAFKA FOLDER - C:\Hadoop\kafka\_2.12-3.4.0**

**Note - kafka-logs and zookeeper-data folder was added after configuration**

Name	Date modified	Type
bin	07-04-2023 02:57	File folder
config	07-04-2023 02:57	File folder
kafka-logs	07-04-2023 21:39	File folder
libs	07-04-2023 02:57	File folder
licenses	07-04-2023 02:57	File folder
logs	07-04-2023 19:03	File folder
site-docs	07-04-2023 02:57	File folder
zookeeper-data	07-04-2023 03:01	File folder
LICENSE	31-01-2023 23:32	File
NOTICE	31-01-2023 23:32	File

**Step 3: Copy the path of the Kafka folder. Now go to *config* inside kafka folder and open *zookeeper.properties* file. Copy the path against the field *dataDir* and add */zookeeper-data* to the path.**

```
# the directory where the snapshot is stored.  
dataDir=/C:/Hadoop/kafka_2.12-3.4.0/zookeeper-data  
# the port at which the clients will connect  
clientPort=2181  
# disable the per-ip limit on the number of connections s
```

**Step 4: Now in the same folder *config* open *server.properties* and scroll down to *log.dirs* and paste the path. To the path add */kafka-logs***

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

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

# A comma separated list of directories under which to store log files
log.dirs=/C:/Hadoop/kafka_2.12-3.4.0/kafka-logs

# 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.
```

**Step 5: This completes the configuration of zookeeper and kafka server. Now open command prompt and change the directory to the kafka folder. First start zookeeper using the command given below:**

`.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties`

```
Microsoft Windows [Version 10.0.19045.2728]
(c) Microsoft Corporation. All rights reserved.

C:\Users\M.K. COMPUTERS>cd C:\Hadoop\kafka_2.12-3.4.0

C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
[2023-04-07 19:04:14,631] INFO Reading configuration from: .\config\zookeeper.properties (org.apache
[2023-04-07 19:04:14,651] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum
[2023-04-07 19:04:14,659] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.Quoru
[2023-04-07 19:04:14,659] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.Quo
[2023-04-07 19:04:14,662] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.Default
```

**Step 6: Now open another command prompt and change the directory to the kafka folder. Run kafka server using the command:**

`.\bin\windows\kafka-server-start.bat .\config\server.properties`

```
C:\Users\M.K. COMPUTERS>cd C:\Hadoop\kafka_2.12-3.4.0

C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\kafka-server-start.bat .\config\server.properties
[2023-04-07 19:05:02,825] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Lo
[2023-04-07 19:05:03,812] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to dis
9Util)
[2023-04-07 19:05:04,034] INFO starting (kafka.server.KafkaServer)
[2023-04-07 19:05:04,036] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServ
[2023-04-07 19:05:04,097] INFO [ZooKeeperClient Kafka server] Initializing a new session to loca
[2023-04-07 19:05:04,118] INFO Client environment:zookeeper.version=3.6.3--6401e4ad2087061bc6b9f
.ZooKeeper)
[2023-04-07 19:05:04,118] INFO Client environment:host.name=10.200.45.208 (org.apache.zookeeper.
[2023-04-07 19:05:04,119] INFO Client environment:java.version=1.8.0_202 (org.apache.zookeeper.Z
[2023-04-07 19:05:04,119] INFO Client environment:java.vendor=Oracle Corporation (org.apache.zoo
```

Now kafka is running and ready to stream data.

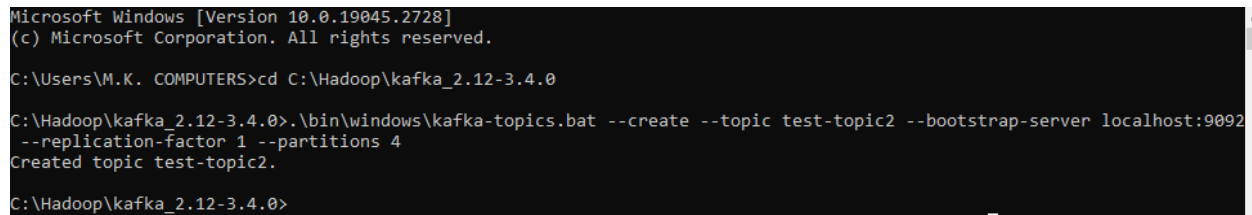
**3. Writing the code for producer and consumer services.**

**AND**

**4. Stream the current day, the timestamp from producer to consumer shell.**

**Creating a new topic by opening a new command prompt, navigating to .../kafka/bin/windows and execute:**

**.\bin\windows\kafka-topics.bat --create --topic test-topic  
--bootstrap-server localhost:9092 --replication-factor 1 --partitions 4**



```
Microsoft Windows [Version 10.0.19045.2728]
(c) Microsoft Corporation. All rights reserved.

C:\Users\M.K. COMPUTERS>cd C:\Hadoop\kafka_2.12-3.4.0

C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\kafka-topics.bat --create --topic test-topic2 --bootstrap-server localhost:9092
--replication-factor 1 --partitions 4
Created topic test-topic2.

C:\Hadoop\kafka_2.12-3.4.0>
```

**CODE FOR PRODUCER:- (producer\_202218061\_lab06\_bdp)**

```
: import json
  from time import sleep
  from json import dumps
  from kafka import KafkaProducer
  from datetime import datetime
```

```
: from kafka import KafkaProducer
import time

producer = KafkaProducer(bootstrap_servers='localhost:9092')

for i in range(10):
    producer.send('test-topic2', b'Hello, World!')
    time.sleep(1)
```

```
: now = datetime.now()
now

: datetime.datetime(2023, 4, 8, 0, 13, 43, 767624)
```

```
: current_time = now.strftime("%d/%m/%Y %H:%M:%S")
current_time
```

```
: '08/04/2023 00:13:43'
```

```
: for i in range(10):
    message = "Message {}".format(str(datetime.now().time()))
    producer.send('test-topic2', json.dumps(message).encode('utf-8'))
    sleep(2)
    print("message sent ",i)
```

## **OUTPUT:-**

```
message sent 0
message sent 1
message sent 2
message sent 3
message sent 4
message sent 5
message sent 6
message sent 7
message sent 8
message sent 9
```

## **CODE FOR CONSUMER:-(consumer\_202218061\_lab06\_bdp)**

```
[n [1]: from kafka import KafkaConsumer
        from json import loads
        import json
```

```
[n [*]: from kafka import KafkaConsumer

        consumer = KafkaConsumer('test-topic2', bootstrap_servers=['localhost:9092'])

        for message in consumer:
            print(message.value.decode())
```

**5.Take a screenshot of the results from the producer and consumer terminal.**

## **#SERVER NAME LIST**

```
.\bin\windows\kafka-topics.bat --list --bootstrap-server
localhost:9092
```

```
C:\Users\M.K. COMPUTERS>cd C:\Hadoop\kafka_2.12-3.4.0
C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\kafka-topics.bat --list --bootstrap-server localhost:9092
__consumer_offsets
humtest
test
test-topic
test-topic1
test-topic2
C:\Hadoop\kafka_2.12-3.4.0>_
```

## **1.Command for producer:-**

```
.\bin\windows\kafka-console-producer.bat --broker-list
localhost:9092 --topic test-topic2
```

```
C:\Windows\system32>cd C:\Hadoop\kafka_2.12-3.4.0
C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\kafka-console-producer.bat --broker-list localhost:9092 --topic test-topic2
_
```

## **2.Command for consumer:-**

```
.\bin\windows\kafka-console-consumer.bat --bootstrap-server
localhost:9092 --topic test-topic2 --from-beginning
```

```
Microsoft Windows [Version 10.0.19045.2728]  
(c) Microsoft Corporation. All rights reserved.  
  
C:\Windows\system32>cd C:\Hadoop\kafka_2.12-3.4.0  
  
C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic test-topic2 --from-beginning  
Hello, World!  
Hello, World!  
"Message 00:13:43.835154"
```

```
C:\Hadoop\kafka_2.12-3.4.0>.\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic test-topic2 --from-beginning  
Hello, World!  
Hello, World!  
"Message 00:13:43.835154"  
"Message 00:13:45.852097"  
"Message 00:13:47.864536"  
"Message 00:13:49.870160"  
"Message 00:13:57.913717"  
"Message 00:13:59.928531"  
Hello, World!  
Hello, World!  
Hello, World!  
Hello, World!  
Hello, World!  
Hello, World!  
"Message 00:13:53.897834"  
"Message 00:13:55.905443"  
Hello, World!  
Hello, World!  
"Message 00:13:51.885655"  
"Message 00:14:01.940137"
```