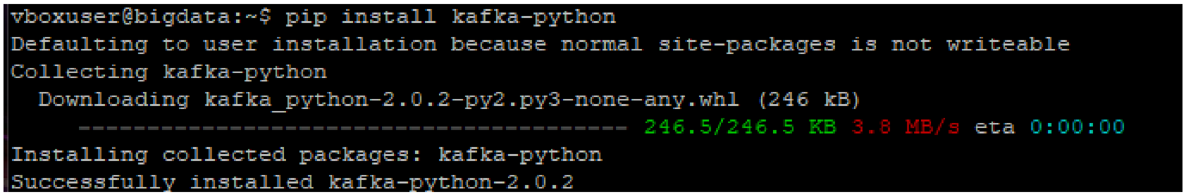
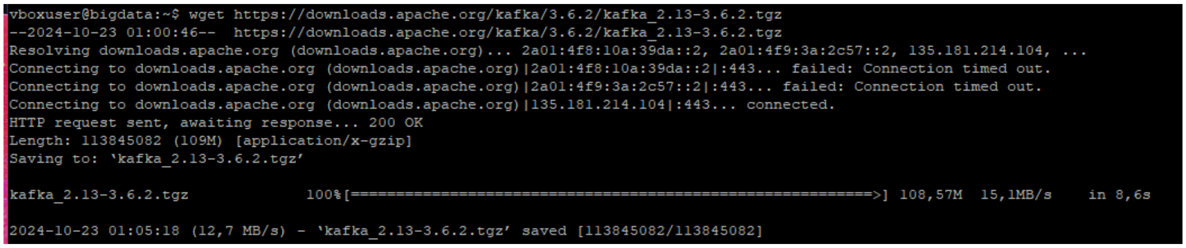
pip install kafka-python



```
vboxuser@bigdata:~$ pip install kafka-python
Defaulting to user installation because normal site-packages is not writeable
Collecting kafka-python
  Downloading kafka_python-2.0.2-py2.py3-none-any.whl (246 kB)
     ---------------------------------------- 246.5/246.5 KB 3.8 MB/s eta 0:00:00
Installing collected packages: kafka-python
Successfully installed kafka-python-2.0.2
```

Descargamos KAFKA

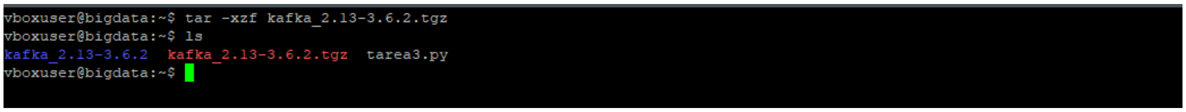
wget https://downloads.apache.org/kafka/3.6.2/kafka_2.13-3.6.2.tgz



```
vboxuser@bigdata:~$ wget https://downloads.apache.org/kafka/3.6.2/kafka_2.13-3.6.2.tgz
--2024-10-23 01:00:46--  https://downloads.apache.org/kafka/3.6.2/kafka_2.13-3.6.2.tgz
Resolving downloads.apache.org (downloads.apache.org)... 2a01:4f8:10a:39da::2, 2a01:4f9:3a:2c57::2, 135.181.214.104, ...
Connecting to downloads.apache.org (downloads.apache.org)|2a01:4f8:10a:39da::2|:443... failed: Connection timed out.
Connecting to downloads.apache.org (downloads.apache.org)|2a01:4f9:3a:2c57::2|:443... failed: Connection timed out.
Connecting to downloads.apache.org (downloads.apache.org)|135.181.214.104|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 113845082 (109M) [application/x-gzip]
Saving to: 'kafka_2.13-3.6.2.tgz'

kafka_2.13-3.6.2.tgz            100%[=========================================================>] 108,57M  15,1MB/s    in 8,6s

2024-10-23 01:05:18 (12,7 MB/s) - 'kafka_2.13-3.6.2.tgz' saved [113845082/113845082]
```

Descomprimimos KAFKA

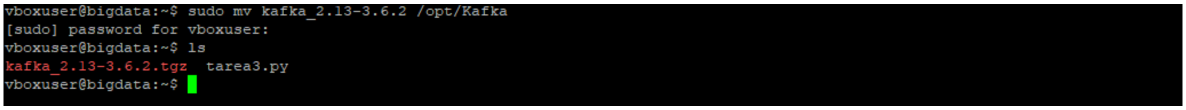
tar -xzf kafka_2.13-3.6.2.tgz



```
vboxuser@bigdata:~$ tar -xzf kafka_2.13-3.6.2.tgz
vboxuser@bigdata:~$ ls
kafka_2.13-3.6.2  kafka_2.13-3.6.2.tgz  tarea3.py
vboxuser@bigdata:~$
```

Movemos lo descargado a una carpeta llamada KAFKA en el directorio opt

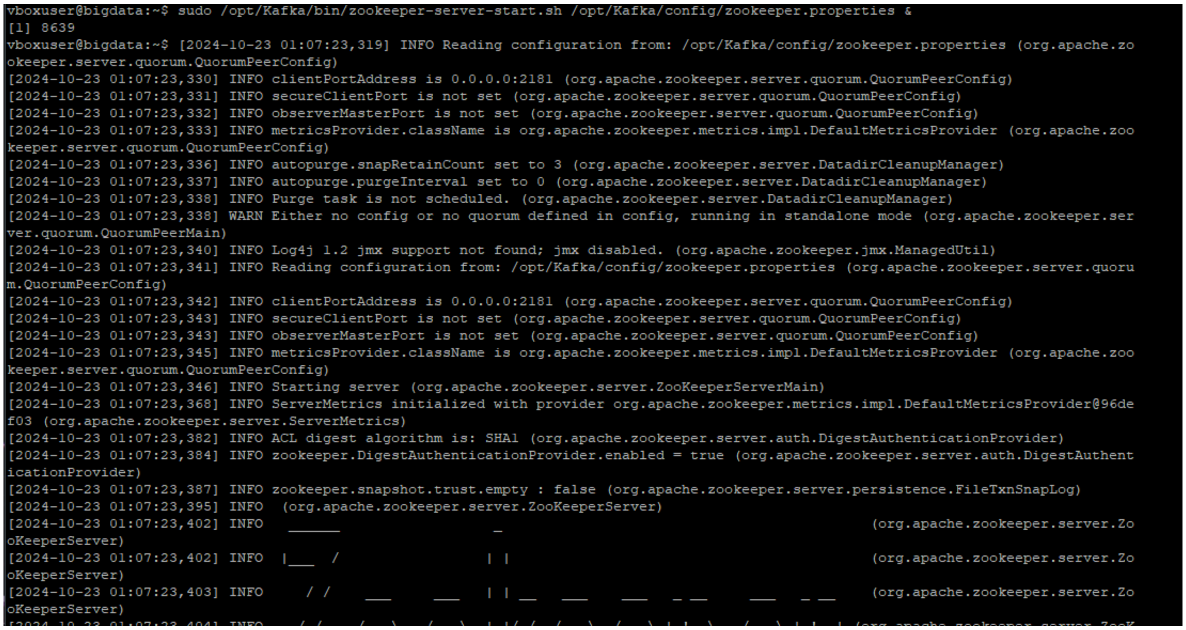
sudo mv kafka_2.13-3.6.2 /opt/Kafka



```
vboxuser@bigdata:~$ sudo mv kafka_2.13-3.6.2 /opt/Kafka
[sudo] password for vboxuser:
vboxuser@bigdata:~$ ls
kafka_2.13-3.6.2.tgz  tarea3.py
vboxuser@bigdata:~$
```

Se inicia el servidor ZooKeeper

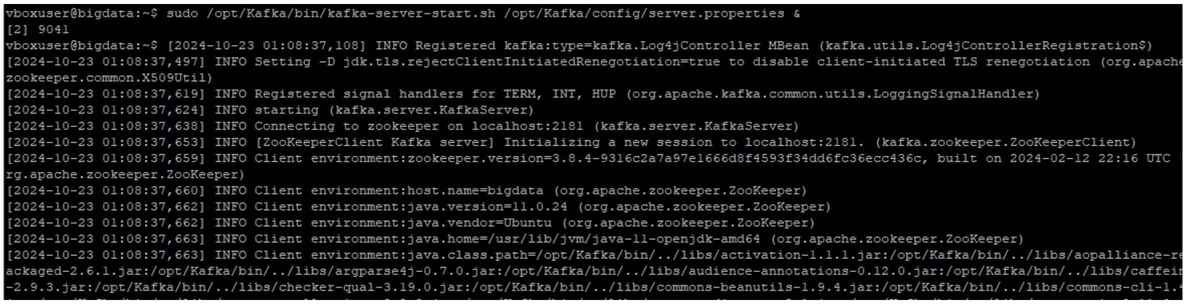
sudo /opt/Kafka/bin/zookeeper-server-start.sh /opt/Kafka/config/zookeeper.properties &



```
vboxuser@bigdata:~$ sudo /opt/Kafka/bin/zookeeper-server-start.sh /opt/Kafka/config/zookeeper.properties &
[1] 8639
vboxuser@bigdata:~$ [2024-10-23 01:07:23,319] INFO Reading configuration from: /opt/Kafka/config/zookeeper.properties (org.apache.zo
okeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,330] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,331] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,332] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,333] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zoo
keeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,336] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-10-23 01:07:23,337] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-10-23 01:07:23,338] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DatadirCleanupManager)
[2024-10-23 01:07:23,338] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.ser
ver.quorum.QuorumPeerMain)
[2024-10-23 01:07:23,340] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2024-10-23 01:07:23,341] INFO Reading configuration from: /opt/Kafka/config/zookeeper.properties (org.apache.zookeeper.server.quoru
m.QuorumPeerConfig)
[2024-10-23 01:07:23,342] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,343] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,343] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,345] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zoo
keeper.server.quorum.QuorumPeerConfig)
[2024-10-23 01:07:23,346] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2024-10-23 01:07:23,368] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@96de
f03 (org.apache.zookeeper.server.ServerMetrics)
[2024-10-23 01:07:23,382] INFO ACL digest algorithm is: SHA1 (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
[2024-10-23 01:07:23,384] INFO zookeeper.DigestAuthenticationProvider.enabled = true (org.apache.zookeeper.server.auth.DigestAuthent
icationProvider)
[2024-10-23 01:07:23,387] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2024-10-23 01:07:23,395] INFO  (org.apache.zookeeper.server.ZooKeeperServer)
[2024-10-23 01:07:23,402] INFO   ______                  _                                          (org.apache.zookeeper.server.Zo
oKeeperServer)
[2024-10-23 01:07:23,402] INFO  |___  /                 | |                                         (org.apache.zookeeper.server.Zo
oKeeperServer)
[2024-10-23 01:07:23,403] INFO     / /    ___     ___   | | __   ___    ___   _ __     ___   _ __    (org.apache.zookeeper.server.Zo
oKeeperServer)
```

Iniciamos el servidor Kafka:

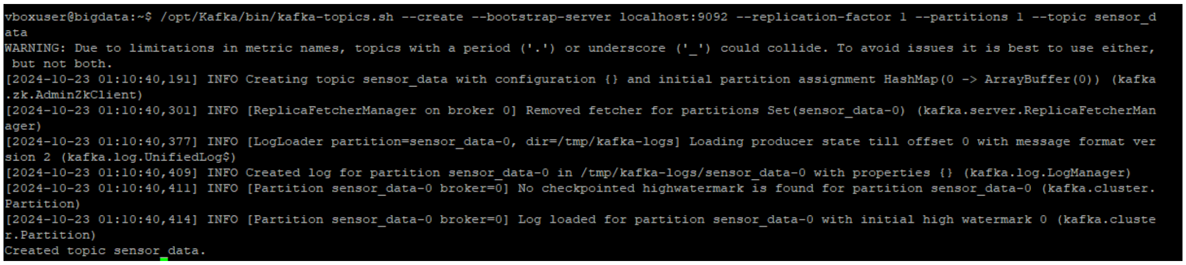
sudo /opt/Kafka/bin/kafka-server-start.sh /opt/Kafka/config/server.properties &



```
vboxuser@bigdata:~$ sudo /opt/Kafka/bin/kafka-server-start.sh /opt/Kafka/config/server.properties &
[2] 9041
vboxuser@bigdata:~$ [2024-10-23 01:08:37,108] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2024-10-23 01:08:37,497] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache
zookeeper.common.X509Util)
[2024-10-23 01:08:37,619] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2024-10-23 01:08:37,624] INFO starting (kafka.server.KafkaServer)
[2024-10-23 01:08:37,638] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2024-10-23 01:08:37,653] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2024-10-23 01:08:37,659] INFO Client environment:zookeeper.version=3.8.4-9316c2a7a97e1666d8f4593f34dd6fc36ecc436c, built on 2024-02-12 22:16 UTC
rg.apache.zookeeper.ZooKeeper)
[2024-10-23 01:08:37,660] INFO Client environment:host.name=bigdata (org.apache.zookeeper.ZooKeeper)
[2024-10-23 01:08:37,662] INFO Client environment:java.version=11.0.24 (org.apache.zookeeper.ZooKeeper)
[2024-10-23 01:08:37,662] INFO Client environment:java.vendor=Ubuntu (org.apache.zookeeper.ZooKeeper)
[2024-10-23 01:08:37,663] INFO Client environment:java.home=/usr/lib/jvm/java-11-openjdk-amd64 (org.apache.zookeeper.ZooKeeper)
[2024-10-23 01:08:37,663] INFO Client environment:java.class.path=/opt/Kafka/bin/../libs/activation-1.1.1.jar:/opt/Kafka/bin/../libs/aopalliance-re
ackaged-2.6.1.jar:/opt/Kafka/bin/../libs/argparse4j-0.7.0.jar:/opt/Kafka/bin/../libs/audience-annotations-0.12.0.jar:/opt/Kafka/bin/../libs/caffein
-2.9.3.jar:/opt/Kafka/bin/../libs/checker-qual-3.19.0.jar:/opt/Kafka/bin/../libs/commons-beanutils-1.9.4.jar:/opt/Kafka/bin/../libs/commons-cli-1.
```

Creamos un tema (topic) de Kafka, el tema se llamará sensor_data

/opt/Kafka/bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic sensor_data



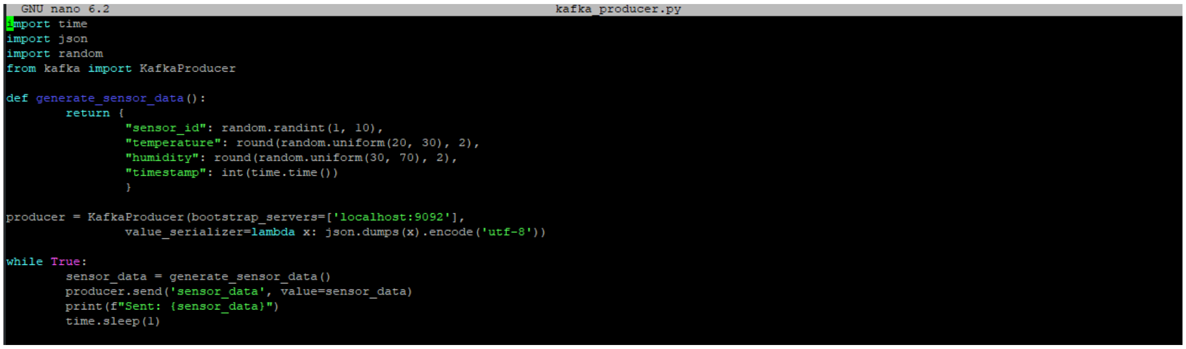
```
vboxuser@bigdata:~$ /opt/Kafka/bin/kafka-topics.sh --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic sensor_d
ata
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either,
 but not both.
[2024-10-23 01:10:40,191] INFO Creating topic sensor_data with configuration {} and initial partition assignment HashMap(0 -> ArrayBuffer(0)) (kafka
.zk.AdminZkClient)
[2024-10-23 01:10:40,301] INFO [ReplicaFetcherManager on broker 0] Removed fetcher for partitions Set(sensor_data-0) (kafka.server.ReplicaFetcherMan
ager)
[2024-10-23 01:10:40,377] INFO [LogLoader partition=sensor_data-0, dir=/tmp/kafka-logs] Loading producer state till offset 0 with message format ver
sion 2 (kafka.log.UnifiedLog$)
[2024-10-23 01:10:40,409] INFO Created log for partition sensor_data-0 in /tmp/kafka-logs/sensor_data-0 with properties {} (kafka.log.LogManager)
[2024-10-23 01:10:40,411] INFO [Partition sensor_data-0 broker=0] No checkpointed highwatermark is found for partition sensor_data-0 (kafka.cluster.
Partition)
[2024-10-23 01:10:40,414] INFO [Partition sensor_data-0 broker=0] Log loaded for partition sensor_data-0 with initial high watermark 0 (kafka.cluste
r.Partition)
Created topic sensor_data.
```

Implementación del productor(producer) de Kafka

Creamos un archivo llamado kafka_producer.py

```
nano kafka_producer.py
```

```
import time
import json
import random
from kafka import KafkaProducer

def generate_sensor_data():
    return {
        "sensor_id": random.randint(1, 10),
        "temperature": round(random.uniform(20, 30), 2),
        "humidity": round(random.uniform(30, 70), 2),
        "timestamp": int(time.time())
        }

producer = KafkaProducer(bootstrap_servers=['localhost:9092'],
        value_serializer=lambda x: json.dumps(x).encode('utf-8'))

while True:
    sensor_data = generate_sensor_data()
    producer.send('sensor_data', value=sensor_data)
    print(f"Sent: {sensor_data}")
    time.sleep(1)
```



```
import time
import json
import random
from kafka import KafkaProducer

def generate_sensor_data():
        return {
                "sensor_id": random.randint(1, 10),
                "temperature": round(random.uniform(20, 30), 2),
                "humidity": round(random.uniform(30, 70), 2),
                "timestamp": int(time.time())
                }

producer = KafkaProducer(bootstrap_servers=['localhost:9092'],
                value_serializer=lambda x: json.dumps(x).encode('utf-8'))

while True:
        sensor_data = generate_sensor_data()
        producer.send('sensor_data', value=sensor_data)
        print(f"Sent: {sensor_data}")
        time.sleep(1)
```

Este script genera datos simulados de sensores y los envía al tema (topic) de Kafka que creamos anteriormente (sensor_data).

### Implementación del consumidor con Spark Streaming

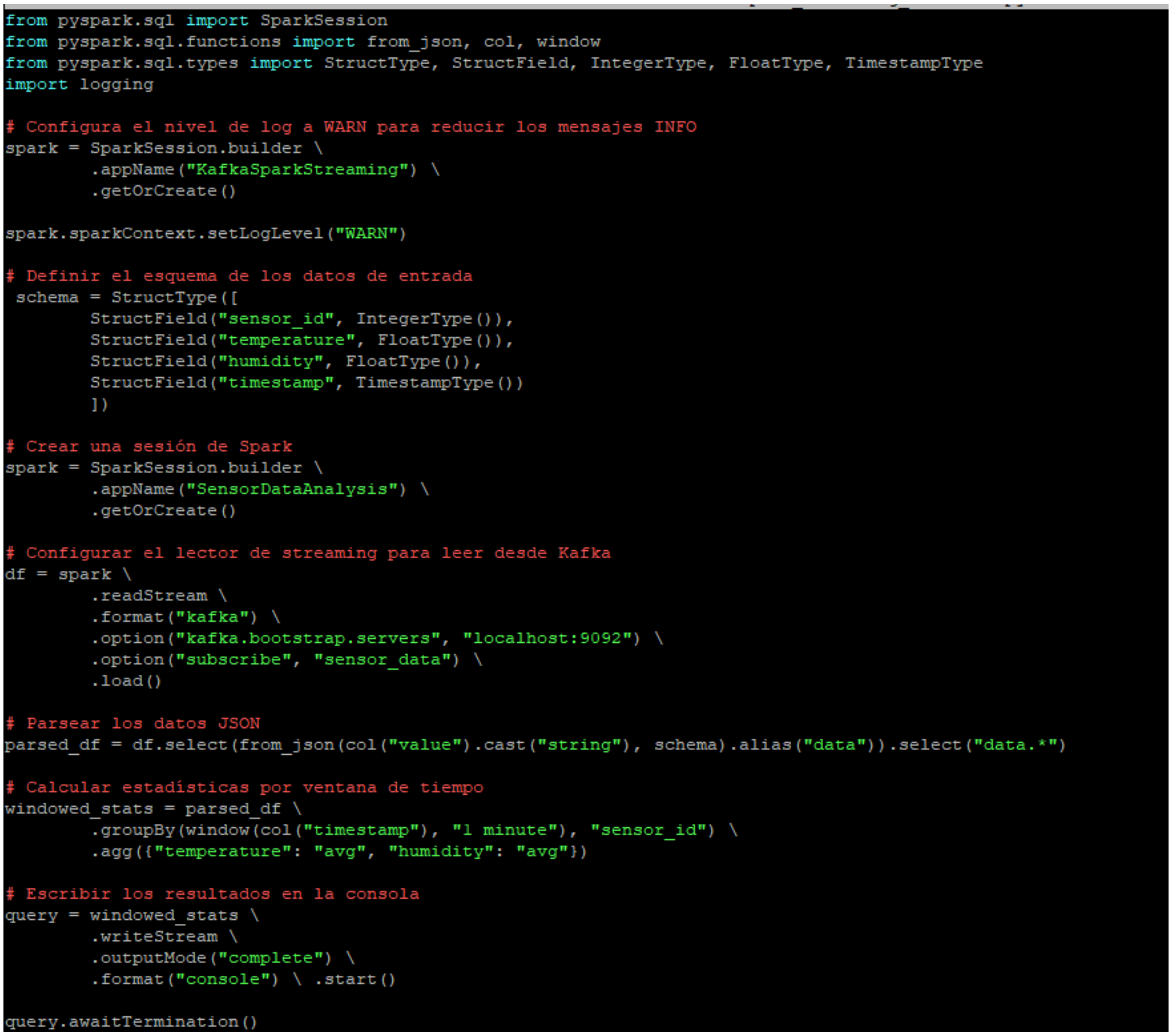
Ahora, crearemos un consumidor(consumer) utilizando Spark Streaming para procesar los datos en tiempo real. Crea un archivo llamado spark_streaming_consumer.py

```
nano spark_streaming_consumer.py
```

```
from pyspark.sql import SparkSession

from pyspark.sql.functions import from_json, col, window

from pyspark.sql.types import StructType, StructField, IntegerType, FloatType, TimestampType

import logging


# Configura el nivel de log a WARN para reducir los mensajes INFO

spark = SparkSession.builder \

    .appName("KafkaSparkStreaming") \

    .getOrCreate()


spark.sparkContext.setLogLevel("WARN")


# Definir el esquema de los datos de entrada
```

```
schema = StructType([
        StructField("sensor_id", IntegerType()),
        StructField("temperature", FloatType()),
        StructField("humidity", FloatType()),
        StructField("timestamp", TimestampType())
        ])

# Crear una sesión de Spark
spark = SparkSession.builder \
        .appName("SensorDataAnalysis") \
        .getOrCreate()

# Configurar el lector de streaming para leer desde Kafka
df = spark \
        .readStream \
        .format("kafka") \
        .option("kafka.bootstrap.servers", "localhost:9092") \
        .option("subscribe", "sensor_data") \
        .load()

# Parsear los datos JSON
parsed_df = df.select(from_json(col("value").cast("string"), schema).alias("data")).select("data.*")

# Calcular estadísticas por ventana de tiempo
windowed_stats = parsed_df \
        .groupBy(window(col("timestamp"), "1 minute"), "sensor_id") \
        .agg({"temperature": "avg", "humidity": "avg"})
```

```
# Escribir los resultados en la consola
query = windowed_stats \
        .writeStream \
        .outputMode("complete") \
        .format("console") \
        .start()

query.awaitTermination()
```

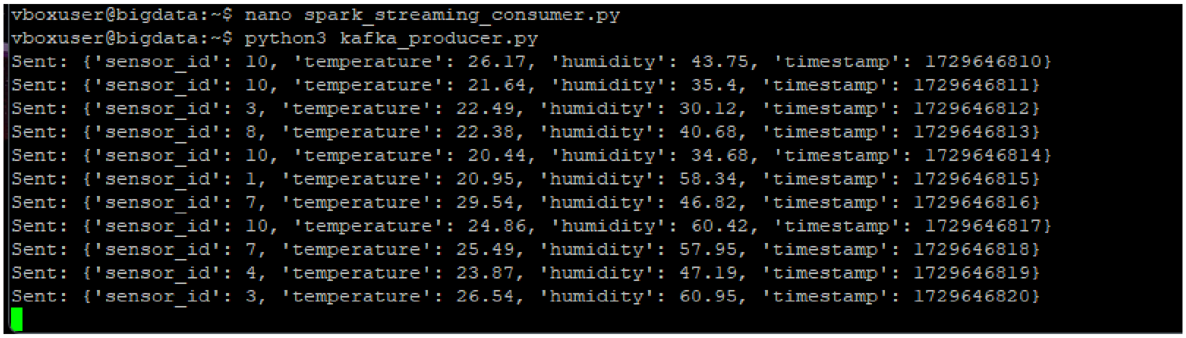


```
from pyspark.sql import SparkSession
from pyspark.sql.functions import from_json, col, window
from pyspark.sql.types import StructType, StructField, IntegerType, FloatType, TimestampType
import logging

# Configura el nivel de log a WARN para reducir los mensajes INFO
spark = SparkSession.builder \
        .appName("KafkaSparkStreaming") \
        .getOrCreate()

spark.sparkContext.setLogLevel("WARN")

# Definir el esquema de los datos de entrada
 schema = StructType([
        StructField("sensor_id", IntegerType()),
        StructField("temperature", FloatType()),
        StructField("humidity", FloatType()),
        StructField("timestamp", TimestampType())
        ])

# Crear una sesión de Spark
spark = SparkSession.builder \
        .appName("SensorDataAnalysis") \
        .getOrCreate()

# Configurar el lector de streaming para leer desde Kafka
df = spark \
        .readStream \
        .format("kafka") \
        .option("kafka.bootstrap.servers", "localhost:9092") \
        .option("subscribe", "sensor_data") \
        .load()

# Parsear los datos JSON
parsed_df = df.select(from_json(col("value").cast("string"), schema).alias("data")).select("data.*")

# Calcular estadisticas por ventana de tiempo
windowed_stats = parsed_df \
        .groupBy(window(col("timestamp"), "1 minute"), "sensor_id") \
        .agg({"temperature": "avg", "humidity": "avg"})

# Escribir los resultados en la consola
query = windowed_stats \
        .writeStream \
        .outputMode("complete") \
        .format("console") \ .start()

query.awaitTermination()
```

### Ejecución y análisis

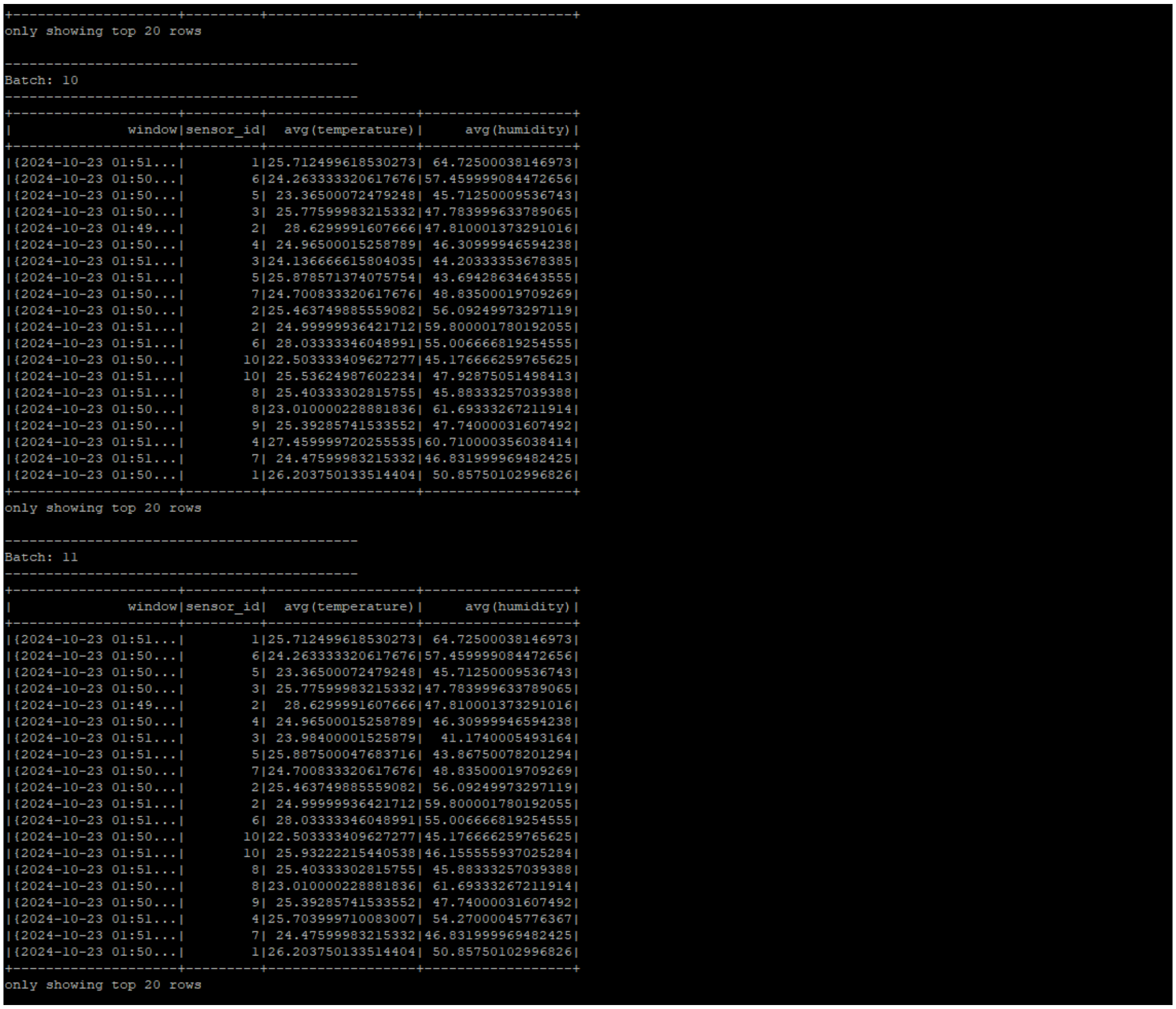
En una terminal, ejecutamos el productor(producer) de Kafka:

python3 kafka_producer.py

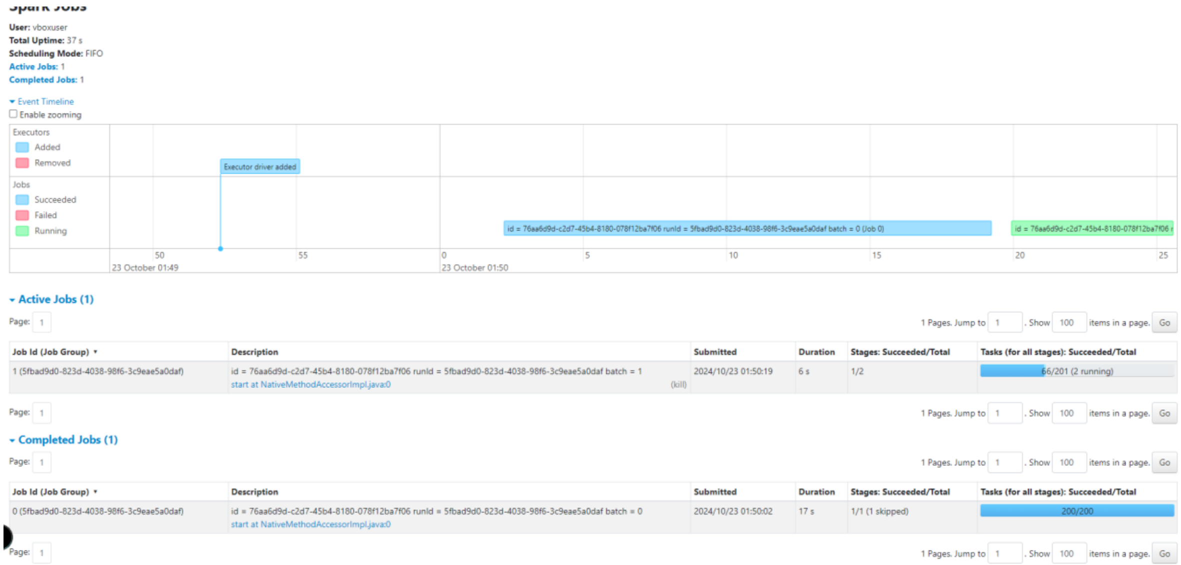


```
vboxuser@bigdata:~$ nano spark_streaming_consumer.py
vboxuser@bigdata:~$ python3 kafka_producer.py
Sent: {'sensor_id': 10, 'temperature': 26.17, 'humidity': 43.75, 'timestamp': 1729646810}
Sent: {'sensor_id': 10, 'temperature': 21.64, 'humidity': 35.4, 'timestamp': 1729646811}
Sent: {'sensor_id': 3, 'temperature': 22.49, 'humidity': 30.12, 'timestamp': 1729646812}
Sent: {'sensor_id': 8, 'temperature': 22.38, 'humidity': 40.68, 'timestamp': 1729646813}
Sent: {'sensor_id': 10, 'temperature': 20.44, 'humidity': 34.68, 'timestamp': 1729646814}
Sent: {'sensor_id': 1, 'temperature': 20.95, 'humidity': 58.34, 'timestamp': 1729646815}
Sent: {'sensor_id': 7, 'temperature': 29.54, 'humidity': 46.82, 'timestamp': 1729646816}
Sent: {'sensor_id': 10, 'temperature': 24.86, 'humidity': 60.42, 'timestamp': 1729646817}
Sent: {'sensor_id': 7, 'temperature': 25.49, 'humidity': 57.95, 'timestamp': 1729646818}
Sent: {'sensor_id': 4, 'temperature': 23.87, 'humidity': 47.19, 'timestamp': 1729646819}
Sent: {'sensor_id': 3, 'temperature': 26.54, 'humidity': 60.95, 'timestamp': 1729646820}
```

spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.5.3 spark_streaming_consumer.py



```
+--------------------+---------+------------------+------------------+
only showing top 20 rows

-------------------------------------------
Batch: 10
-------------------------------------------
+--------------------+---------+------------------+------------------+
|              window|sensor_id|  avg(temperature)|     avg(humidity)|
+--------------------+---------+------------------+------------------+
|{2024-10-23 01:51...|        1|25.712499618530273| 64.72500038146973|
|{2024-10-23 01:50...|        6|24.263333320617676|57.459999084472656|
|{2024-10-23 01:50...|        5| 23.36500072479248| 45.71250009536743|
|{2024-10-23 01:50...|        3| 25.77599983215332|47.783999633789065|
|{2024-10-23 01:49...|        2|  28.6299991607666|47.810001373291016|
|{2024-10-23 01:50...|        4| 24.96500015258789| 46.30999946594238|
|{2024-10-23 01:51...|        3|24.136666615804035| 44.20333353678385|
|{2024-10-23 01:51...|        5|25.878571374075754| 43.69428634643555|
|{2024-10-23 01:50...|        7|24.700833320617676| 48.83500019709269|
|{2024-10-23 01:50...|        2|25.463749885559082| 56.09249973297119|
|{2024-10-23 01:51...|        2| 24.99999936421712|59.800001780192055|
|{2024-10-23 01:51...|        6| 28.03333346048991|55.006666819254555|
|{2024-10-23 01:50...|       10|22.503333409627277|45.176666259765625|
|{2024-10-23 01:51...|       10| 25.53624987602234| 47.92875051498413|
|{2024-10-23 01:51...|        8| 25.40333302815755| 45.88333257039388|
|{2024-10-23 01:50...|        8|23.010000228881836| 61.69333267211914|
|{2024-10-23 01:50...|        9| 25.39285741533552| 47.74000031607492|
|{2024-10-23 01:51...|        4|27.459999720255535|60.710000356038414|
|{2024-10-23 01:51...|        7| 24.47599983215332|46.831999969482425|
|{2024-10-23 01:50...|        1|26.203750133514404| 50.85750102996826|
+--------------------+---------+------------------+------------------+
only showing top 20 rows

-------------------------------------------
Batch: 11
-------------------------------------------
+--------------------+---------+------------------+------------------+
|              window|sensor_id|  avg(temperature)|     avg(humidity)|
+--------------------+---------+------------------+------------------+
|{2024-10-23 01:51...|        1|25.712499618530273| 64.72500038146973|
|{2024-10-23 01:50...|        6|24.263333320617676|57.459999084472656|
|{2024-10-23 01:50...|        5| 23.36500072479248| 45.71250009536743|
|{2024-10-23 01:50...|        3| 25.77599983215332|47.783999633789065|
|{2024-10-23 01:49...|        2|  28.6299991607666|47.810001373291016|
|{2024-10-23 01:50...|        4| 24.96500015258789| 46.30999946594238|
|{2024-10-23 01:51...|        3| 23.98400001525879|  41.1740005493164|
|{2024-10-23 01:51...|        5|25.887500047683716| 43.86750078201294|
|{2024-10-23 01:50...|        7|24.700833320617676| 48.83500019709269|
|{2024-10-23 01:50...|        2|25.463749885559082| 56.09249973297119|
|{2024-10-23 01:51...|        2| 24.99999936421712|59.800001780192055|
|{2024-10-23 01:51...|        6| 28.03333346048991|55.006666819254555|
|{2024-10-23 01:50...|       10|22.503333409627277|45.176666259765625|
|{2024-10-23 01:51...|       10| 25.93222215440538|46.155555937025284|
|{2024-10-23 01:51...|        8| 25.40333302815755| 45.88333257039388|
|{2024-10-23 01:50...|        8|23.010000228881836| 61.69333267211914|
|{2024-10-23 01:50...|        9| 25.39285741533552| 47.74000031607492|
|{2024-10-23 01:51...|        4|25.703999710083007| 54.27000045776367|
|{2024-10-23 01:51...|        7| 24.47599983215332|46.831999969482425|
|{2024-10-23 01:50...|        1|26.203750133514404| 50.85750102996826|
+--------------------+---------+------------------+------------------+
only showing top 20 rows
```



## Spark Jobs

**User:** vboxuser
**Total Uptime:** 37 s
**Scheduling Mode:** FIFO
**Active Jobs:** 1
**Completed Jobs:** 1

▾ Event Timeline
□ Enable zooming

Executors
🟦 Added
🟥 Removed
Executor driver added
Jobs
🟦 Succeeded
🟥 Failed
🟩 Running
id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 0 (Job 0)
id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 r
50
55
0
5
10
15
20
25
23 October 01:49
23 October 01:50

### ▾ Active Jobs (1)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

| Job Id (Job Group) ▾ | Description | Submitted | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total |
|---|---|---|---|---|---|
| 1 (5fbad9d0-823d-4038-98f6-3c9eae5a0daf) | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 1<br>start at NativeMethodAccessorImpl.java:0 (kill) | 2024/10/23 01:50:19 | 6 s | 1/2 | 66/201 (2 running) |

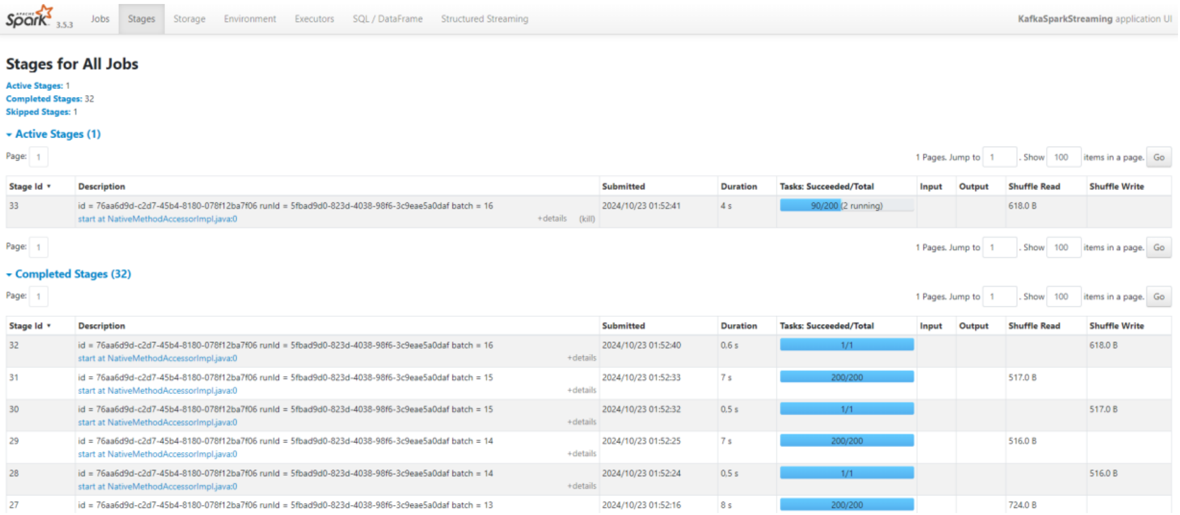
Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

### ▾ Completed Jobs (1)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

| Job Id (Job Group) ▾ | Description | Submitted | Duration | Stages: Succeeded/Total | Tasks (for all stages): Succeeded/Total |
|---|---|---|---|---|---|
| 0 (5fbad9d0-823d-4038-98f6-3c9eae5a0daf) | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 0<br>start at NativeMethodAccessorImpl.java:0 | 2024/10/23 01:50:02 | 17 s | 1/1 (1 skipped) | 200/200 |

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go



## Stages for All Jobs

**Active Stages:** 1
**Completed Stages:** 32
**Skipped Stages:** 1

### ▾ Active Stages (1)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

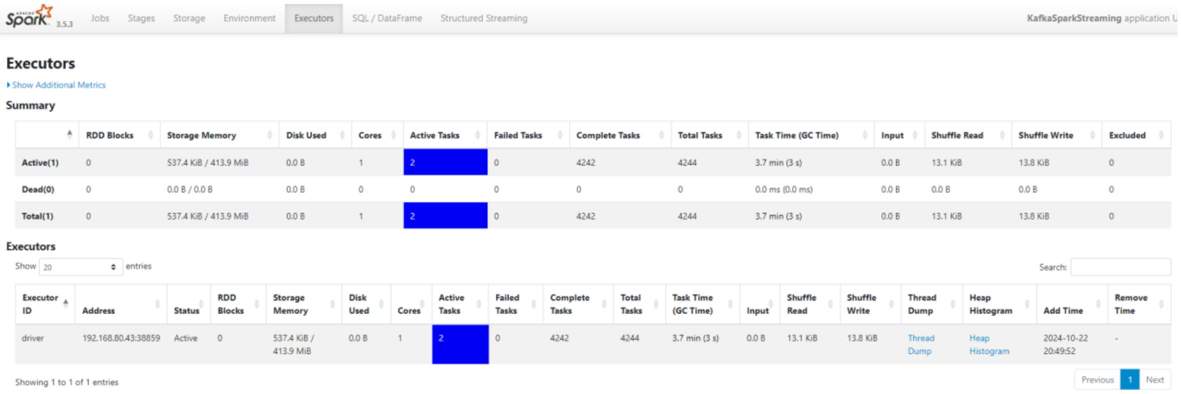
| Stage Id ▾ | Description | Submitted | Duration | Tasks: Succeeded/Total | Input | Output | Shuffle Read | Shuffle Write |
|---|---|---|---|---|---|---|---|---|
| 33 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 16<br>start at NativeMethodAccessorImpl.java:0 +details (kill) | 2024/10/23 01:52:41 | 4 s | 90/200 (2 running) | | | 618.0 B | |

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

### ▾ Completed Stages (32)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

| Stage Id ▾ | Description | Submitted | Duration | Tasks: Succeeded/Total | Input | Output | Shuffle Read | Shuffle Write |
|---|---|---|---|---|---|---|---|---|
| 32 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 16<br>start at NativeMethodAccessorImpl.java:0 +details | 2024/10/23 01:52:40 | 0.6 s | 1/1 | | | | 618.0 B |
| 31 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 15<br>start at NativeMethodAccessorImpl.java:0 +details | 2024/10/23 01:52:33 | 7 s | 200/200 | | | 517.0 B | |
| 30 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 15<br>start at NativeMethodAccessorImpl.java:0 +details | 2024/10/23 01:52:32 | 0.5 s | 1/1 | | | | 517.0 B |
| 29 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 14<br>start at NativeMethodAccessorImpl.java:0 +details | 2024/10/23 01:52:25 | 7 s | 200/200 | | | 516.0 B | |
| 28 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 14<br>start at NativeMethodAccessorImpl.java:0 +details | 2024/10/23 01:52:24 | 0.5 s | 1/1 | | | | 516.0 B |
| 27 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 13 | 2024/10/23 01:52:16 | 8 s | 200/200 | | | 724.0 B | |



## Executors

▸ Show Additional Metrics

### Summary

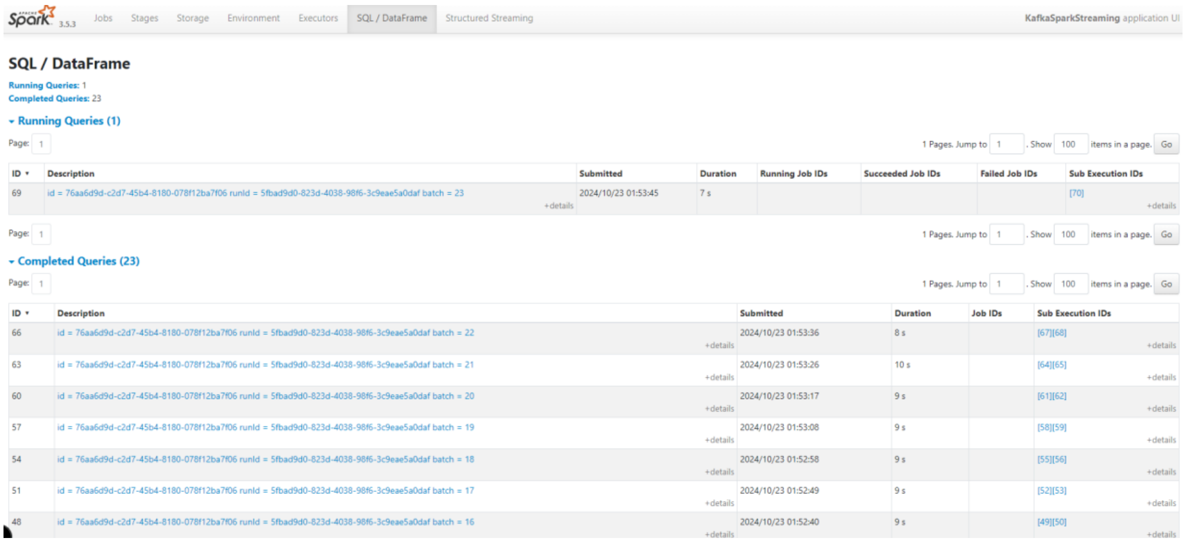
| | RDD Blocks | Storage Memory | Disk Used | Cores | Active Tasks | Failed Tasks | Complete Tasks | Total Tasks | Task Time (GC Time) | Input | Shuffle Read | Shuffle Write | Excluded |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Active(1)** | 0 | 537.4 KiB / 413.9 MiB | 0.0 B | 1 | 2 | 0 | 4242 | 4244 | 3.7 min (3 s) | 0.0 B | 13.1 KiB | 13.8 KiB | 0 |
| **Dead(0)** | 0 | 0.0 B / 0.0 B | 0.0 B | 0 | 0 | 0 | 0 | 0 | 0.0 ms (0.0 ms) | 0.0 B | 0.0 B | 0.0 B | 0 |
| **Total(1)** | 0 | 537.4 KiB / 413.9 MiB | 0.0 B | 1 | 2 | 0 | 4242 | 4244 | 3.7 min (3 s) | 0.0 B | 13.1 KiB | 13.8 KiB | 0 |

### Executors

Show 20 entries — Search:

| Executor ID | Address | Status | RDD Blocks | Storage Memory | Disk Used | Cores | Active Tasks | Failed Tasks | Complete Tasks | Total Tasks | Task Time (GC Time) | Input | Shuffle Read | Shuffle Write | Thread Dump | Heap Histogram | Add Time | Remove Time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| driver | 192.168.80.43:38859 | Active | 0 | 537.4 KiB / 413.9 MiB | 0.0 B | 1 | 2 | 0 | 4242 | 4244 | 3.7 min (3 s) | 0.0 B | 13.1 KiB | 13.8 KiB | Thread Dump | Heap Histogram | 2024-10-22 20:49:52 | - |

Showing 1 to 1 of 1 entries — Previous 1 Next



## SQL / DataFrame

**Running Queries:** 1
**Completed Queries:** 23

▾ **Running Queries (1)**

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

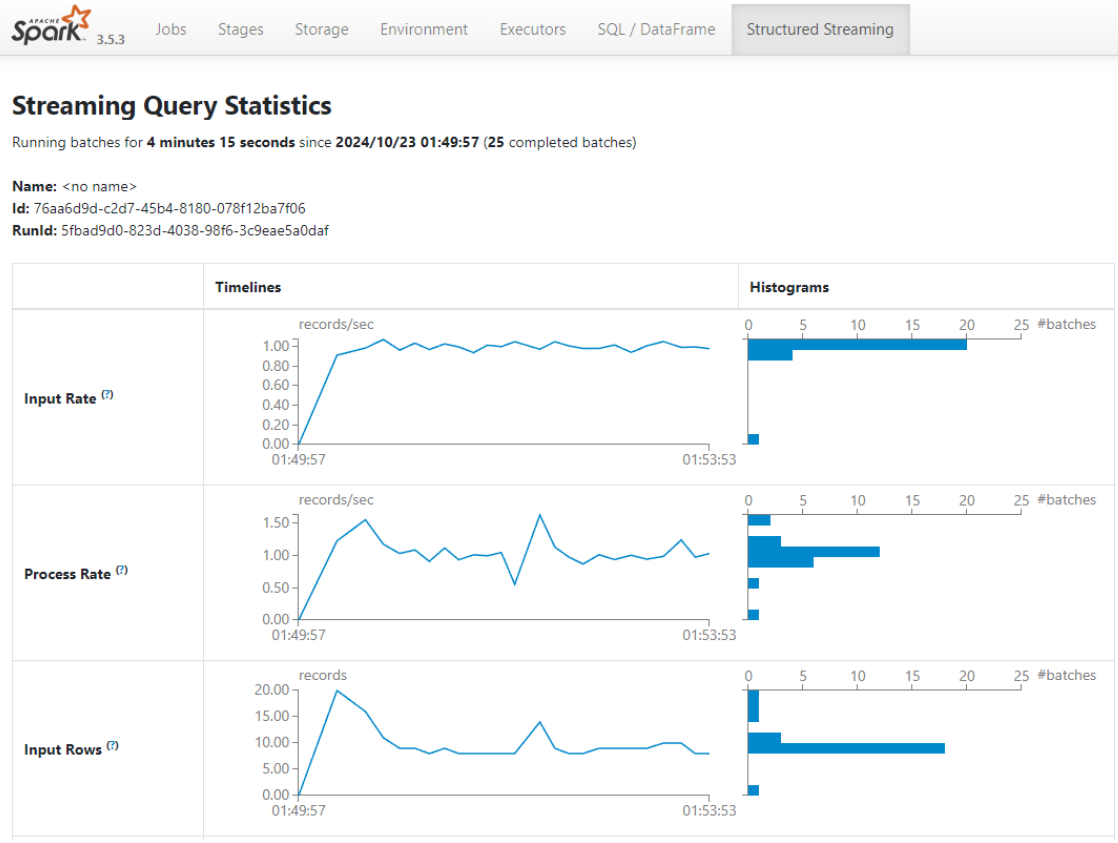
| ID ▾ | Description | Submitted | Duration | Running Job IDs | Succeeded Job IDs | Failed Job IDs | Sub Execution IDs |
|---|---|---|---|---|---|---|---|
| 69 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 23<br>+details | 2024/10/23 01:53:45 | 7 s | | | | [70]<br>+details |

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

▾ **Completed Queries (23)**

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

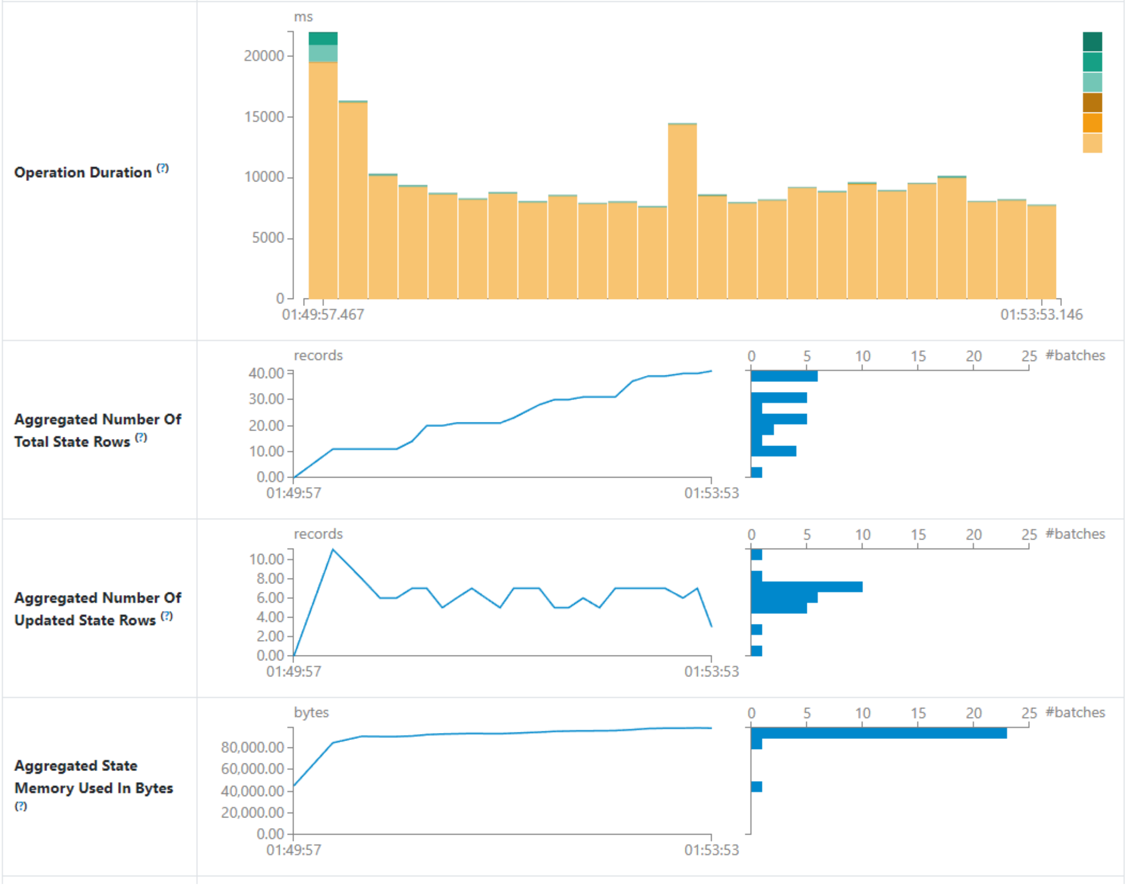
| ID ▾ | Description | Submitted | Duration | Job IDs | Sub Execution IDs |
|---|---|---|---|---|---|
| 66 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 22<br>+details | 2024/10/23 01:53:36 | 8 s | | [67][68]<br>+details |
| 63 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 21<br>+details | 2024/10/23 01:53:26 | 10 s | | [64][65]<br>+details |
| 60 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 20<br>+details | 2024/10/23 01:53:17 | 9 s | | [61][62]<br>+details |
| 57 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 19<br>+details | 2024/10/23 01:53:08 | 9 s | | [58][59]<br>+details |
| 54 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 18<br>+details | 2024/10/23 01:52:58 | 9 s | | [55][56]<br>+details |
| 51 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 17<br>+details | 2024/10/23 01:52:49 | 9 s | | [52][53]<br>+details |
| 48 | id = 76aa6d9d-c2d7-45b4-8180-078f12ba7f06 runId = 5fbad9d0-823d-4038-98f6-3c9eae5a0daf batch = 16<br>+details | 2024/10/23 01:52:40 | 9 s | | [49][50]<br>+details |



## Streaming Query Statistics

Running batches for **4 minutes 15 seconds** since **2024/10/23 01:49:57** (**25** completed batches)

**Name:** <no name>
**Id:** 76aa6d9d-c2d7-45b4-8180-078f12ba7f06
**RunId:** 5fbad9d0-823d-4038-98f6-3c9eae5a0daf



Operation Duration (?)

Aggregated Number Of Total State Rows (?)

Aggregated Number Of Updated State Rows (?)

Aggregated State Memory Used In Bytes (?)