

Kafka + Spark Streaming + PySpark on Windows 11

Step 1: Installing Spark

1.1 Download Spark

<http://spark.apache.org/downloads.html>

```
wget http://www-us.apache.org/dist/spark/spark-3.3.1/spark-3.3.1-bin-hadoop3.tgz
```

```
PS C:\Users\Sharo> wget https://www.apache.org/dyn/closer.lua/spark/spark-3.3.1/spark-3.3.1-bin-hadoop3.tgz

StatusCode      : 200
StatusDescription : OK
Content         : <!DOCTYPE html>

                  <html lang="en">
                  <head>
                    <meta charset="utf-8">
                    <meta http-equiv="X-UA-Compatible" content="IE=edge">
                    <meta name="viewport" content="width=device-width, initial-scale=1">
                    <meta...
RawContent      : HTTP/1.1 200 OK
                  Connection: keep-alive
                  X-Served-By: cache-hel1410033-HEL, cache-pao17466-PA0
                  X-Cache: MISS, MISS
                  X-Cache-Hits: 0, 0
                  X-Timer: S1669964368.357487,VS0,VE166
                  Vary: Accept-Encoding
```

1.2 Unpack

```
tar -xvf spark-3.3.1-bin-hadoop3.tgz
```

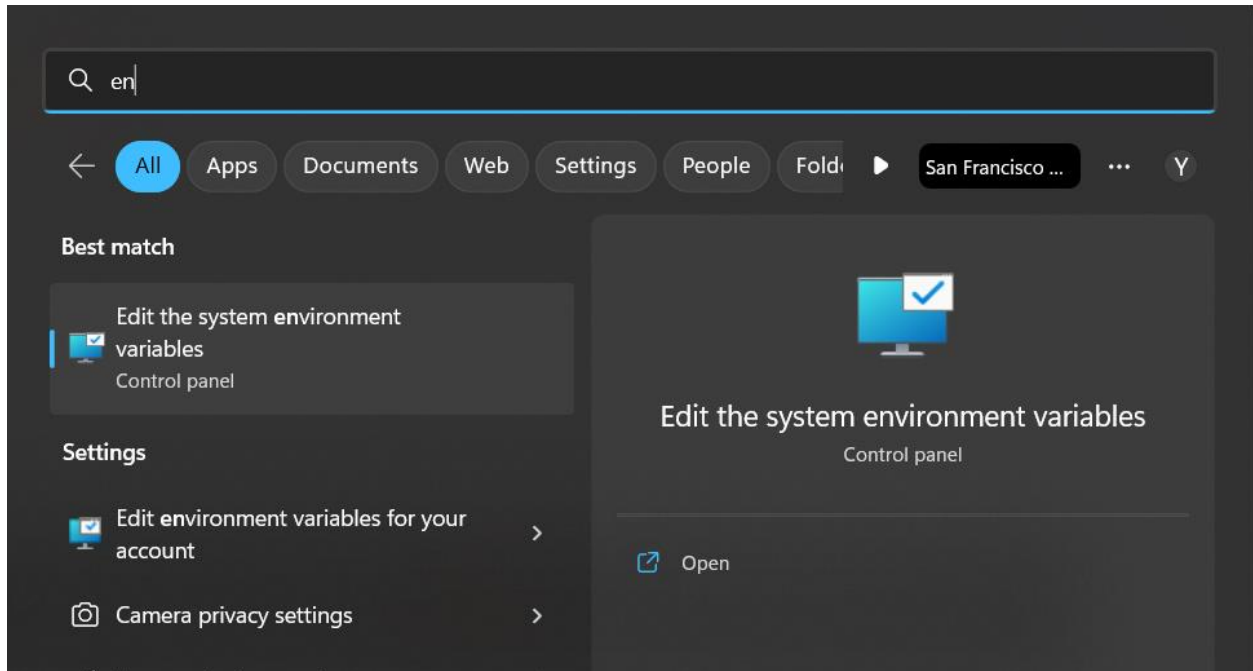
```
PS C:\Users\Sharo\Downloads> tar -xvf spark-3.3.1-bin-hadoop3.tgz
x spark-3.3.1-bin-hadoop3/
x spark-3.3.1-bin-hadoop3/LICENSE
x spark-3.3.1-bin-hadoop3/NOTICE
x spark-3.3.1-bin-hadoop3/R/
x spark-3.3.1-bin-hadoop3/R/lib/
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/DESCRIPTION
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/INDEX
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/Meta/
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/Meta/Rd.rds
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/Meta/features.rds
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/Meta/hsearch.rds
x spark-3.3.1-bin-hadoop3/R/lib/SparkR/Meta/links.rds
```

1.3 Move folder to desired location (You can simply copy-paste the folder in Windows)

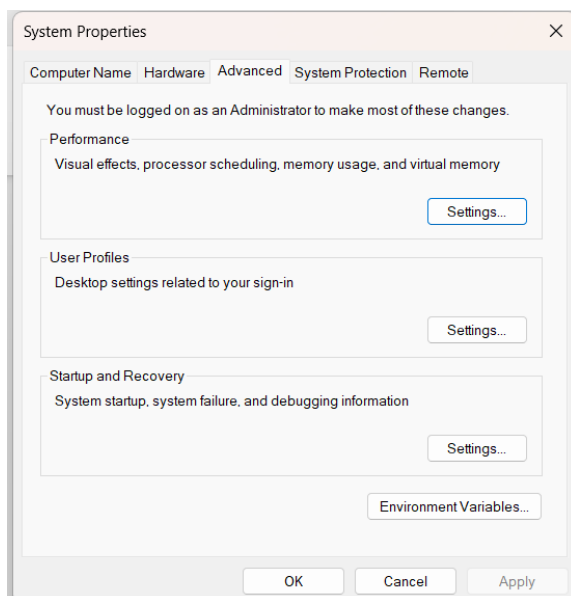
```
PS C:\Users\Sharo\Downloads> mv spark-3.3.1-bin-hadoop3 ..  
PS C:\Users\Sharo\Downloads> ls ..
```

1.4 Set Spark Environment Variables

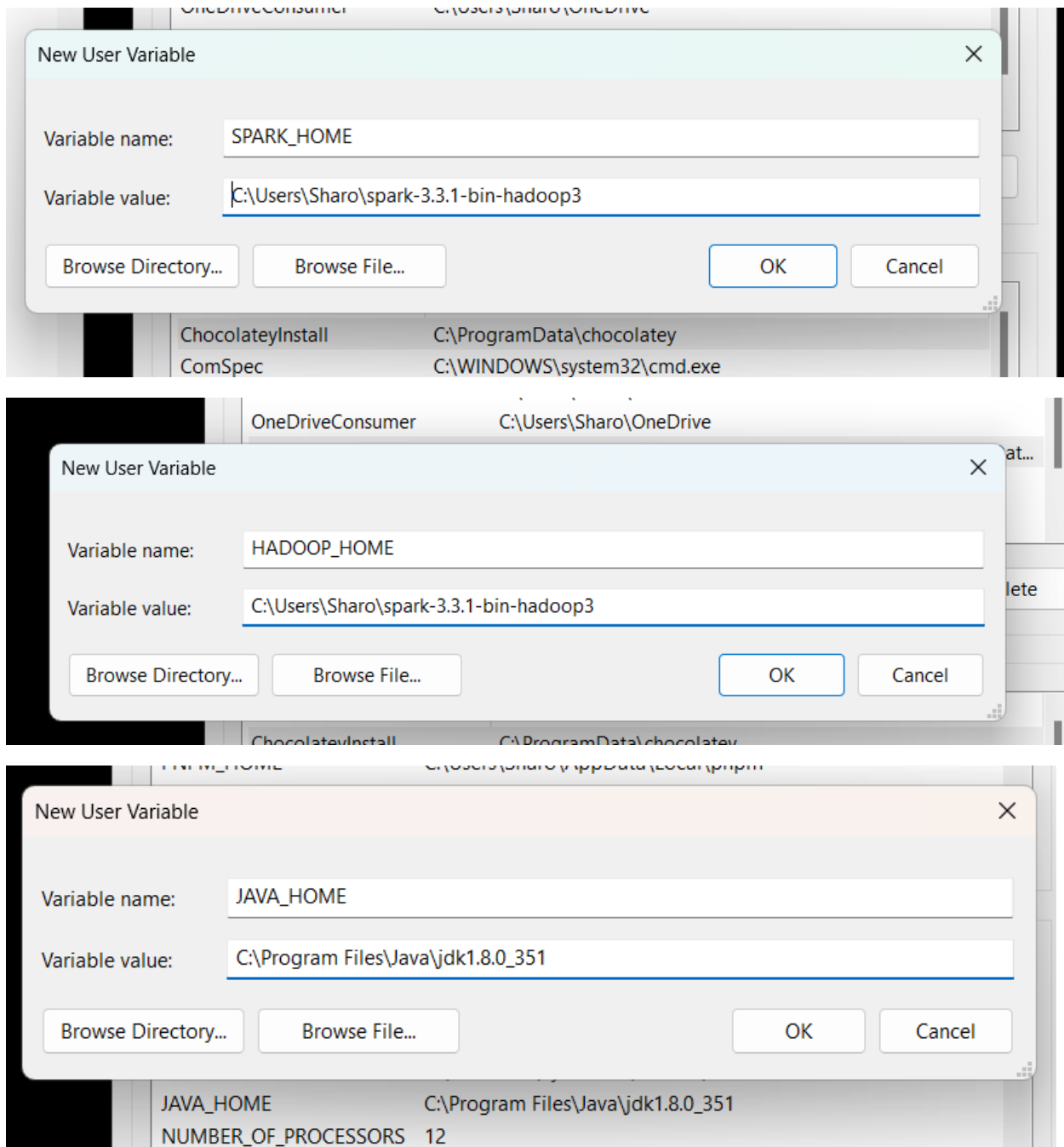
1.41 Search → Type “Environment Variables” → clicks



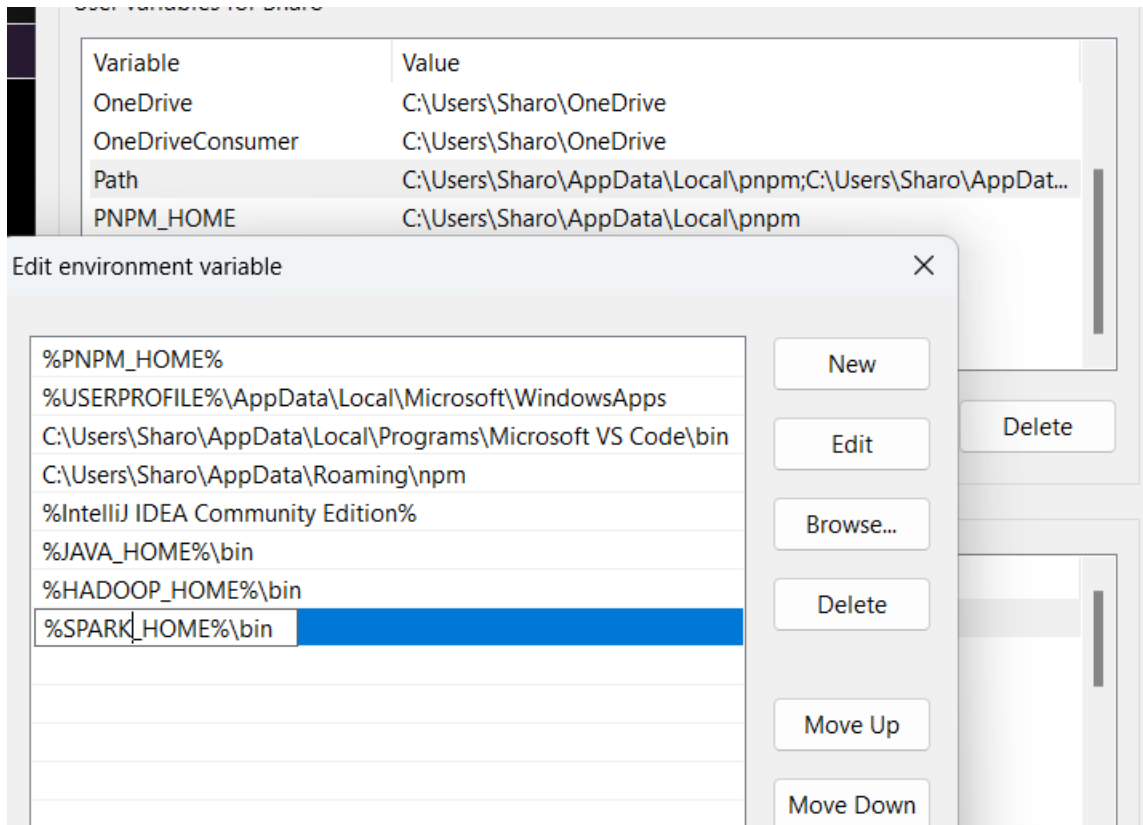
1.42 Select Environment Variables



1.43 Set JAVA_HOME, HADOOP_HOME, SPARK_HOME in user variables

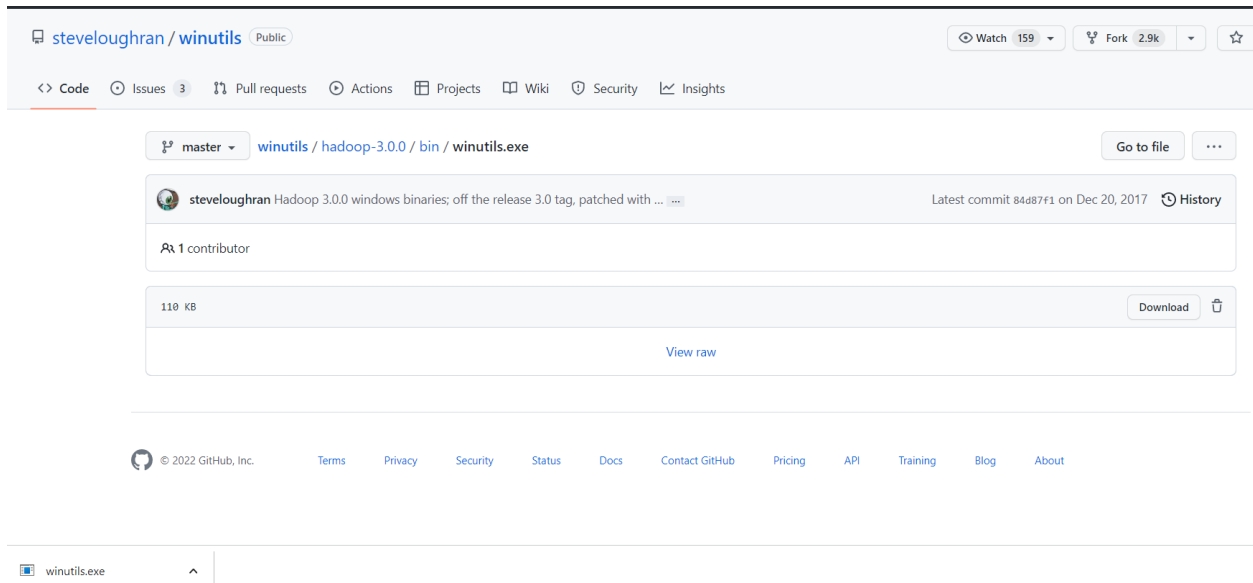


1.44 Set Path



1.5 Download Spark with winutils.exe on Windows

➔ Download winutils.exe according to your spark version



➔ Move downloaded winutils.exe to your spark-3.3.1-bin-hadoop3 folder

File Explorer window showing the contents of the 'bin' directory in the 'spark-3.3.1-bin-hadoop3' folder. The left sidebar shows the navigation pane with 'Users' selected. The main pane displays a list of files and folders. The file 'winutils' is highlighted in blue, indicating it is selected. The status bar at the bottom shows '28 items' and '1 item selected 107 KB'.

Name	Date modified	Type	Size
sparkR	10/15/2022 3:32 AM	Windows Comma...	2 KB
sparkR2	10/15/2022 3:32 AM	Windows Comma...	2 KB
spark-shell	10/15/2022 3:32 AM	File	4 KB
spark-shell	10/15/2022 3:32 AM	Windows Comma...	2 KB
spark-shell2	10/15/2022 3:32 AM	Windows Comma...	2 KB
spark-sql	10/15/2022 3:32 AM	File	2 KB
spark-sql	10/15/2022 3:32 AM	Windows Comma...	2 KB
spark-sql2	10/15/2022 3:32 AM	Windows Comma...	2 KB
spark-submit	10/15/2022 3:32 AM	File	2 KB
spark-submit	10/15/2022 3:32 AM	Windows Command Script	2 KB
spark-submit2	10/15/2022 3:32 AM	Windows Comma...	2 KB
winutils	12/3/2022 1:15 PM	Application	107 KB

1.6 Verify installation

➔ Type pyspark in terminal to test

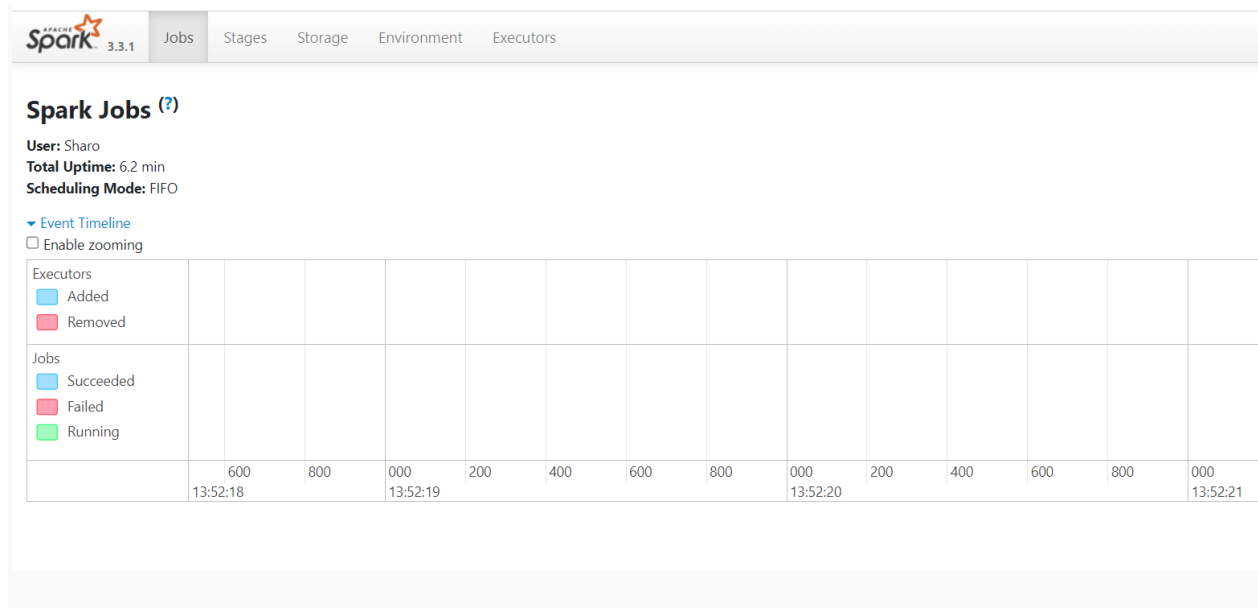
```
C:\Users\Sharo\spark-3.3.1-bin-hadoop3\bin>pyspark
Python 3.11.0 (main, Oct 24 2022, 18:26:48) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
22/12/03 13:52:17 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: java.io.FileNotFoundException: HADOOP_HOME and hadoop.home.dir are unset. -see https://wiki.apache.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/12/03 13:52:18 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Welcome to

      /\_/\
     /__\/   version 3.3.1
    /  _ \
   /___/\

Using Python version 3.11.0 (main, Oct 24 2022 18:26:48)
Spark context Web UI available at http://Sharon:4040
Spark context available as 'sc' (master = local[*], app id = local-1670104341732).
SparkSession available as 'spark'.
>>>
```

1.7 GUI on <http://localhost:4040>

<https://sparkbyexamples.com/spark/spark-web-ui-understanding/>




Done!

Step 2: Install Kafka

2.1 Download Kafka

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



GET STARTED DOCS POWERED BY COMMUNITY APACHE **DOWNLOAD KAFKA**

DOWNLOAD

3.3.1 is the latest release. The current stable version is 3.3.1.

You can verify your download by following these [procedures](#) and using these [KEYS](#).

3.3.1

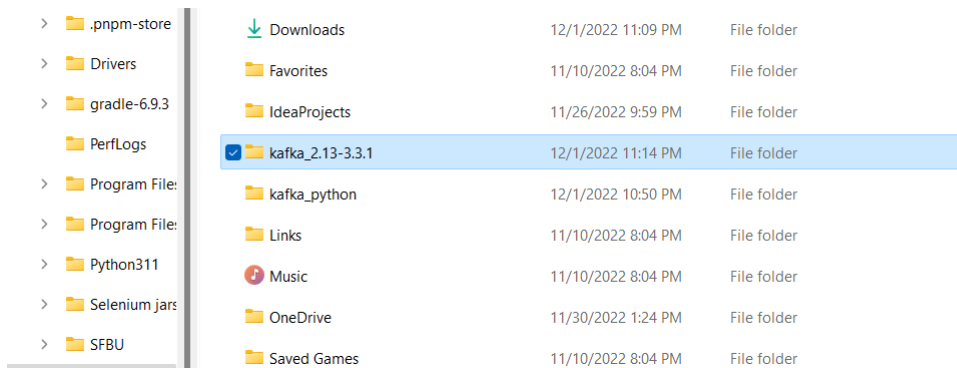
- Released October 3, 2022
- [3.3.1](#) and [3.3.0](#) Release Notes
- Source download: [kafka-3.3.1-src.tgz](#) (asc, sha512)
- Binary downloads:
 - Scala 2.12 - [kafka_2.12-3.3.1.tgz](#) (asc, sha512)
 - Scala 2.13 - [kafka_2.13-3.3.1.tgz](#) (asc, sha512)

2.2 Unpack and move to desire location

```
tar -xvf kafka_2.13-3.3.1.tgz
```

```
PS C:\Users\Sharo\Downloads> tar -xvf kafka_2.13-3.3.1.tgz
x kafka_2.13-3.3.1/
x kafka_2.13-3.3.1/LICENSE
x kafka_2.13-3.3.1/NOTICE
x kafka_2.13-3.3.1/bin/
x kafka_2.13-3.3.1/bin/kafka-console-consumer.sh
x kafka_2.13-3.3.1/bin/kafka-log-dirs.sh
x kafka_2.13-3.3.1/bin/kafka-producer-perf-test.sh
x kafka_2.13-3.3.1/bin/kafka-console-producer.sh
x kafka_2.13-3.3.1/bin/kafka-streams-application-reset.sh
x kafka_2.13-3.3.1/bin/kafka-configs.sh
x kafka_2.13-3.3.1/bin/kafka-get-offsets.sh
```

2.3 Create data files and update files



→ cd to kafka_2.13-3.3.1 directory

```
PS C:\Users\Sharo\Downloads> cd kafka_2.13-3.3.1
PS C:\Users\Sharo\Downloads\kafka_2.13-3.3.1> ls

Directory: C:\Users\Sharo\Downloads\kafka_2.13-3.3.1

Mode                LastWriteTime         Length Name
----                -
d-----          9/29/2022 12:06 PM             bin
d-----          9/29/2022 12:06 PM            config
d-----          9/29/2022 12:06 PM             libs
d-----          9/29/2022 12:06 PM          licenses
d-----          9/29/2022 12:06 PM          site-docs
-a-----          9/29/2022 12:03 PM        14842 LICENSE
-a-----          9/29/2022 12:03 PM        28184 NOTICE
```

→ Make a directory called data and within the directory create another two directories broker and zookeeper

```
PS C:\Users\Sharo\Downloads\kafka_2.13-3.3.1> mkdir data
```

```
Directory: C:\Users\Sharo\Downloads\kafka_2.13-3.3.1
```

Mode	LastWriteTime	Length	Name
d----	12/1/2022 11:10 PM		data

```
PS C:\Users\Sharo\Downloads\kafka_2.13-3.3.1> cd data
```

```
PS C:\Users\Sharo\Downloads\kafka_2.13-3.3.1\data> mkdir broker
```

```
Directory: C:\Users\Sharo\Downloads\kafka_2.13-3.3.1\data
```

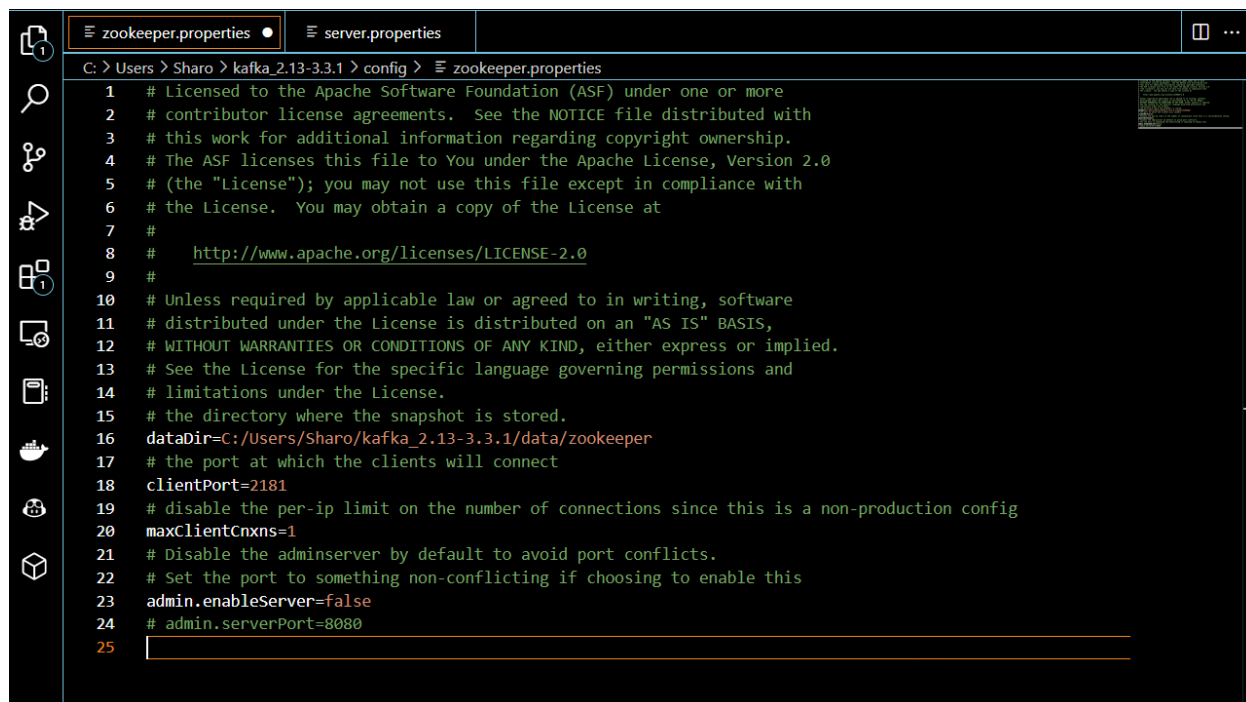
Mode	LastWriteTime	Length	Name
d----	12/1/2022 11:10 PM		broker

```
PS C:\Users\Sharo\Downloads\kafka_2.13-3.3.1\data> mkdir zookeeper
```

```
Directory: C:\Users\Sharo\Downloads\kafka_2.13-3.3.1\data
```

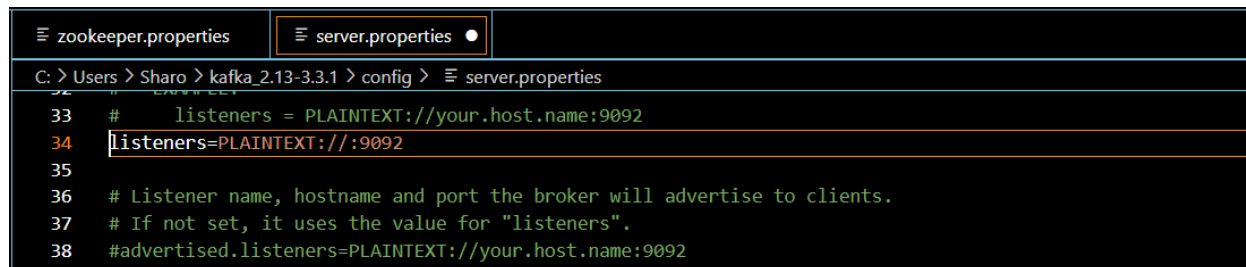
Mode	LastWriteTime	Length	Name
d----	12/1/2022 11:11 PM		zookeeper

→ Edit config/Zookeeper.properties

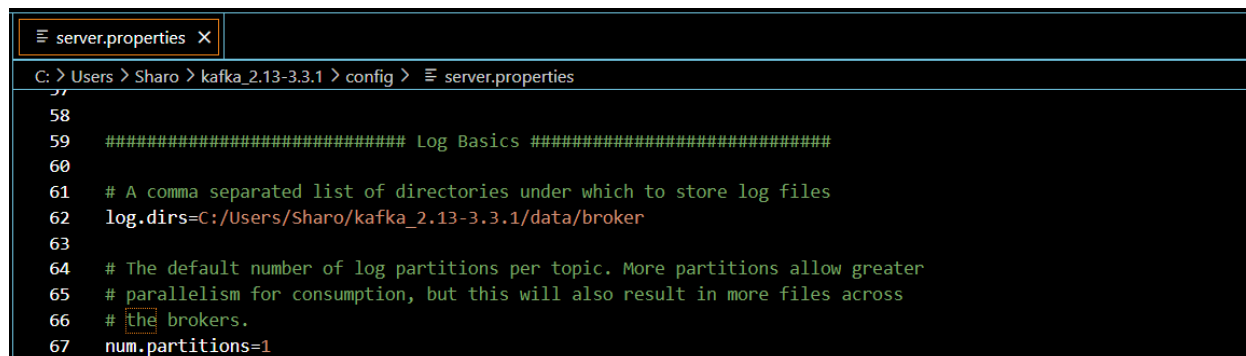


```
1 # Licensed to the Apache Software Foundation (ASF) under one or more
2 # contributor license agreements. See the NOTICE file distributed with
3 # this work for additional information regarding copyright ownership.
4 # The ASF licenses this file to You under the Apache License, Version 2.0
5 # (the "License"); you may not use this file except in compliance with
6 # the License. You may obtain a copy of the License at
7 #
8 # http://www.apache.org/licenses/LICENSE-2.0
9 #
10 # Unless required by applicable law or agreed to in writing, software
11 # distributed under the License is distributed on an "AS IS" BASIS,
12 # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 # See the License for the specific language governing permissions and
14 # limitations under the License.
15 # the directory where the snapshot is stored.
16 dataDir=C:/Users/Sharo/kafka_2.13-3.3.1/data/zookeeper
17 # the port at which the clients will connect
18 clientPort=2181
19 # disable the per-ip limit on the number of connections since this is a non-production config
20 maxClientCnxns=1
21 # Disable the adminserver by default to avoid port conflicts.
22 # Set the port to something non-conflicting if choosing to enable this
23 admin.enableServer=false
24 # admin.serverPort=8080
25
```

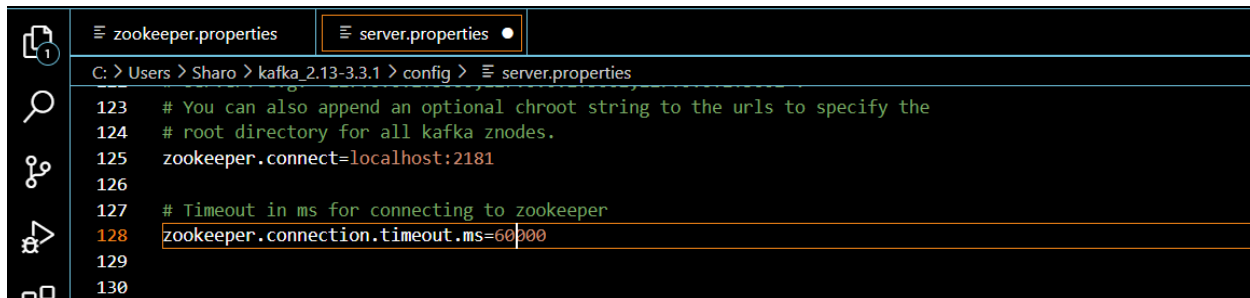
→ Edit config/Server.properties



```
33 # listeners = PLAINTEXT://your.host.name:9092
34 listeners=PLAINTEXT://:9092
35
36 # Listener name, hostname and port the broker will advertise to clients.
37 # If not set, it uses the value for "listeners".
38 #advertised.listeners=PLAINTEXT://your.host.name:9092
```



```
58
59 ##### Log Basics #####
60
61 # A comma separated list of directories under which to store log files
62 log.dirs=C:/Users/Sharo/kafka_2.13-3.3.1/data/broker
63
64 # The default number of log partitions per topic. More partitions allow greater
65 # parallelism for consumption, but this will also result in more files across
66 # the brokers.
67 num.partitions=1
```

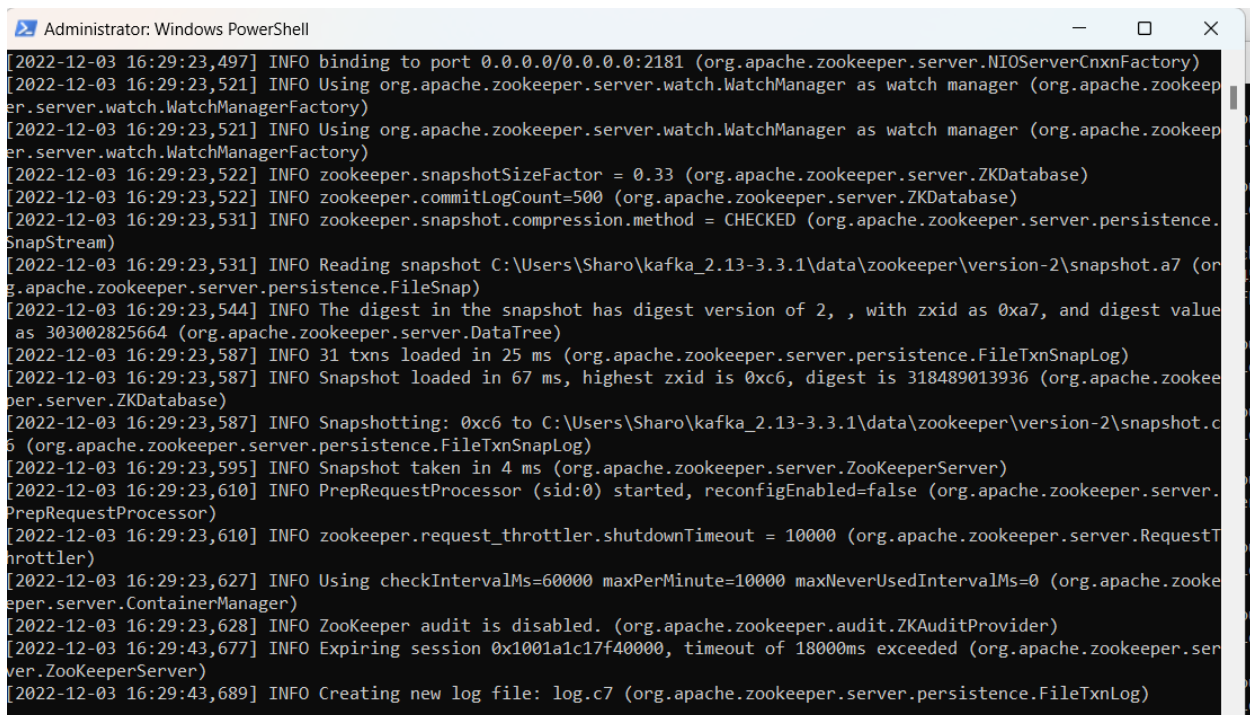


Step 3: Starting Kafka and Test with Example

3.1 Start Kafka Zookeeper

→ Open PowerShell terminal as Administrator

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/zookeeper-server-start.bat
config/zookeeper.properties
```



```
Select Administrator: Windows PowerShell

[2022-12-03 16:29:23,366] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-03 16:29:23,366] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-03 16:29:23,366] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-03 16:29:23,366] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2022-12-03 16:29:23,366] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2022-12-03 16:29:23,390] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@7a5d012c (org.apache.zookeeper.server.ServerMetrics)
[2022-12-03 16:29:23,405] INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapshotLog)
[2022-12-03 16:29:23,426] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,426] INFO _____ (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,426] INFO |___ / | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,427] INFO // _ _ | | _ _ _ _ _ (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,428] INFO // / _ \ / _ \ | / / / _ \ / _ \ | ' \ / _ \ | ' _ | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,428] INFO / / _ | ( ) | | ( ) | | < | _ / | _ / | | _ / | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,428] INFO / _ _ | \ _ / \ _ / | _ \ \ _ / \ _ / | | _ / \ _ / | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,428] INFO | | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,428] INFO | _ | (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,428] INFO (org.apache.zookeeper.server.ZooKeeperServer)
[2022-12-03 16:29:23,439] INFO Server environment:zookeeper.version=3.6.3--6401e4ad2087061bc6b9f80dec2d69f2e3c8660a, built on 04/08/2021 16:35 GMT (org.apache.zookeeper.server.ZooKeeperServer)
```

3.2 Start Kafka Brokers

→ Open another PowerShell terminal as Administrator

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-server-start.bat
config/server.properties
```

```
Administrator: Windows PowerShell

group.GroupMetadataManager)
[2022-12-03 16:31:10,629] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-42 in 50 milliseconds for epoch 0, of which 50 milliseconds was spent in the scheduler. (kafka.coordinator.group.GroupMetadataManager)
[2022-12-03 16:31:10,630] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-12 in 50 milliseconds for epoch 0, of which 50 milliseconds was spent in the scheduler. (kafka.coordinator.group.GroupMetadataManager)
[2022-12-03 16:31:10,634] INFO Loaded member MemberMetadata(memberId=console-consumer-f259df26-fcbf-4447-b69b-7726577a71fd, groupId=None, clientId=console-consumer, clientHost=/192.168.1.11, sessionTimeoutMs=45000, rebalanceTimeoutMs=300000, supportedProtocols=List(range)) in group console-consumer-67719 with generation 1. (kafka.coordinator.group.GroupMetadataManager)
[2022-12-03 16:31:10,635] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-21 in 54 milliseconds for epoch 0, of which 49 milliseconds was spent in the scheduler. (kafka.coordinator.group.GroupMetadataManager)
[2022-12-03 16:31:10,635] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-36 in 49 milliseconds for epoch 0, of which 49 milliseconds was spent in the scheduler. (kafka.coordinator.group.GroupMetadataManager)
[2022-12-03 16:31:10,636] INFO [GroupMetadataManager brokerId=0] Finished loading offsets and group metadata from __consumer_offsets-6 in 50 milliseconds for epoch 0, of which 50 milliseconds was spent in the scheduler. (kafka.coordinator.group.GroupMetadataManager)
```

3.2 Create Kafka Topics

→ Open another PowerShell terminal as Administrator

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --create --  
topic test_step_1 --bootstrap-server localhost:9092 --partitions 1 --  
replication-factor 1
```

```
PS C:\WINDOWS\system32> C:\Users\Sharo\kafka_2.13-3.3.1\bin\windows\kafka-topics.bat --create --topic test_step_1 --boot  
strap-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 issu  
es it is best to use either, but not both.  
Created topic test_step_1.  
PS C:\WINDOWS\system32> █
```

→ To check current topics created

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --list --  
bootstrap-server localhost:9092
```

3.3 Using Example in package

→ Write events to topic

Open a new terminal and type

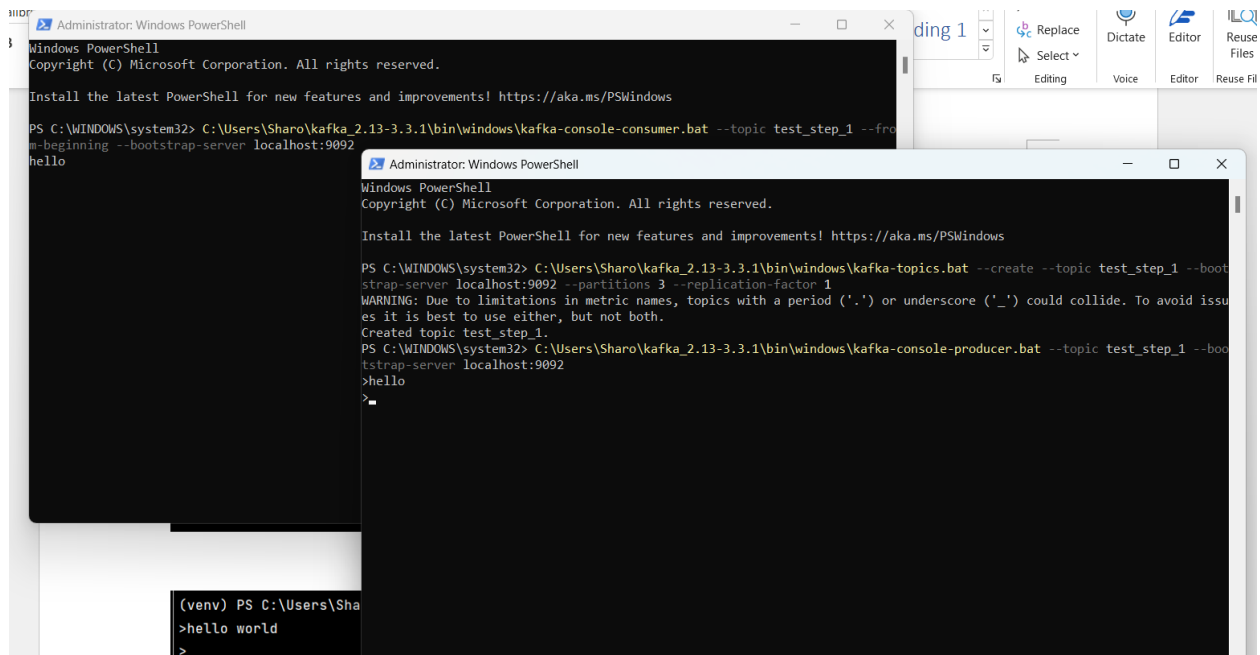
```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-console-producer.bat --  
topic test_step_1 --bootstrap-server localhost:9092
```

→ Read events from topic

Open a new terminal and type

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-console-consumer.bat --  
topic test_step_1 --bootstrap-server localhost:9092
```

Notice: When you type some events in producer, it will send to consumer when “enter” tabbed



3.4 Using Python for Kafka

→ Stop producer by tab "Ctrl+C" then install kafka-python

```
pip install kafka-python
```

Notice: kafka-python will not work with co-existence of kafka package

```
PS C:\WINDOWS\system32> pip install kafka-python
Collecting kafka-python
  Using cached kafka_python-2.0.2-py2.py3-none-any.whl (246 kB)
Installing collected packages: kafka-python
Successfully installed kafka-python-2.0.2

[notice] A new release of pip available: 22.3 -> 22.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

3.4 Test Python for Kafka with Current Consumer.bat

→ Create a main.py file and test with current running consumer

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

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

```

for e in range(1000):
    data = {'number' : e}
    print(data)
    producer.send(topic_name, value=data)
    sleep(5)

```

→ Run python code

```
python C:/Users/Sharo/kafka_python/main.py
```

```

PS C:\Users\Sharo> python C:\Users\Sharo\kafka_python\main.py
b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)'
b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)'
b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)'
b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)'
b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)'
Traceback (most recent call last):
  File "C:\Users\Sharo\kafka_python\main.py", line 11, in <module>
    sleep(5)
KeyboardInterrupt
PS C:\Users\Sharo> |

```

→ Go to the current running kafka-console-consumer.bat

```

PS C:\Users\Sharo> C:\Users\Sharo\kafka_2.13-3.3.1\bin\windows\kafka-console-consumer.bat --topic test_step_1 --from-beg
inning --bootstrap-server localhost:9092
hello
hello world
123456
987
(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)
(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)
(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)
(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)
(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)
|

```

3.5 Test Python for Kafka with Self-defined Python Code

→ Open a cmd terminal and start with consumer.py

```

from kafka import KafkaConsumer

consumer = KafkaConsumer('test_step_1', bootstrap_servers=['localhost:9092'])
for msg in consumer:
    print(msg.value)

```

```
Command Prompt - python  X  Windows PowerShell  X  +  v

Microsoft Windows [Version 10.0.22621.819]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Sharo>python
Python 3.11.0 (main, Oct 24 2022, 18:26:48) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from kafka import KafkaConsumer
>>> consumer = KafkaConsumer('test_step_1', bootstrap_servers=['localhost:9092'])
>>> for msg in consumer:
...     print(msg.value)
...

```

→Open another cmd terminal and start with producer.py

```
from kafka import KafkaProducer

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

producer.send('test_step_1', b'(1, Main Menu), (2, Phone) , (3, Smart Phone),
(4, iPhone)')
```

```
Command Prompt - python  X  Windows PowerShell  X  +  v

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Sharo> python
Python 3.11.0 (main, Oct 24 2022, 18:26:48) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from kafka import KafkaProducer
>>> producer = KafkaProducer(bootstrap_servers='localhost:9092')
>>> producer.send('test_step_1', b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)')
<kafka.producer.future.FutureRecordMetadata object at 0x00000267214EC7D0>
>>> |

```

→Go back to consumer.py and see the result

```
Command Prompt - python  X  Windows PowerShell  X  +  v

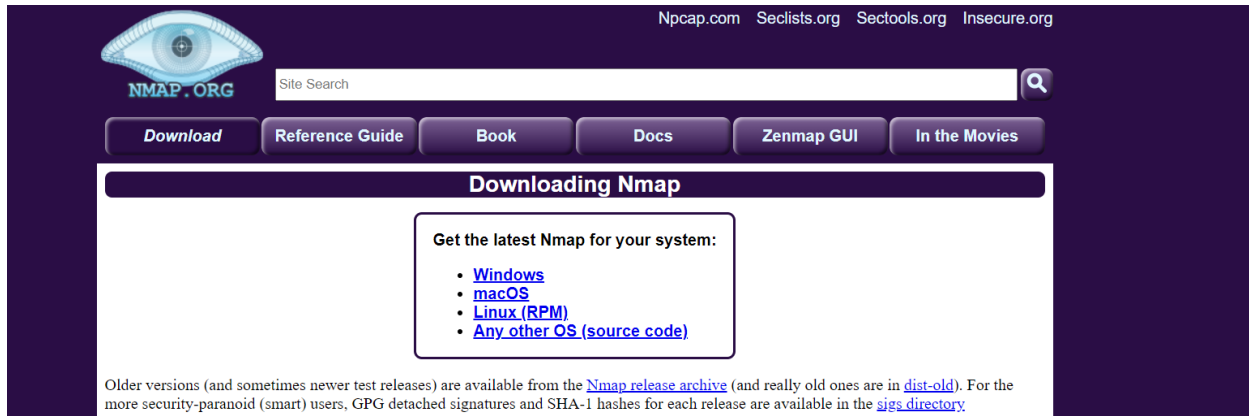
Microsoft Windows [Version 10.0.22621.819]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Sharo>python
Python 3.11.0 (main, Oct 24 2022, 18:26:48) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from kafka import KafkaConsumer
>>> consumer = KafkaConsumer('test_step_1', bootstrap_servers=['localhost:9092'])
>>> for msg in consumer:
...     print(msg.value)
...
b'(1, Main Menu), (2, Phone) , (3, Smart Phone), (4, iPhone)'
```

Step 4: Spark Streaming

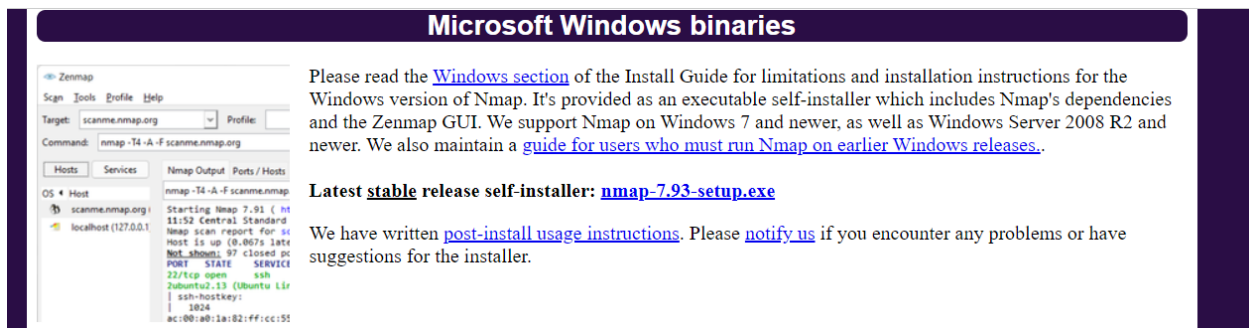
4.1 Install Netcat

<https://nmap.org/ncat/>



The screenshot shows the Nmap.org website with a dark blue header. The main content area is titled 'Downloading Nmap' and contains a box with the text 'Get the latest Nmap for your system:' followed by a bulleted list of links: Windows, macOS, Linux (RPM), and Any other OS (source code). Below this list, there is a paragraph of text about older versions and release archives.

Older versions (and sometimes newer test releases) are available from the [Nmap release archive](#) (and really old ones are in [dist-old](#)). For the more security-paranoid (smart) users, GPG detached signatures and SHA-1 hashes for each release are available in the [sigs directory](#) ([verification instructions](#)). Before downloading, be sure to read the relevant section for your platform from the [Nmap User's Guide](#). The next

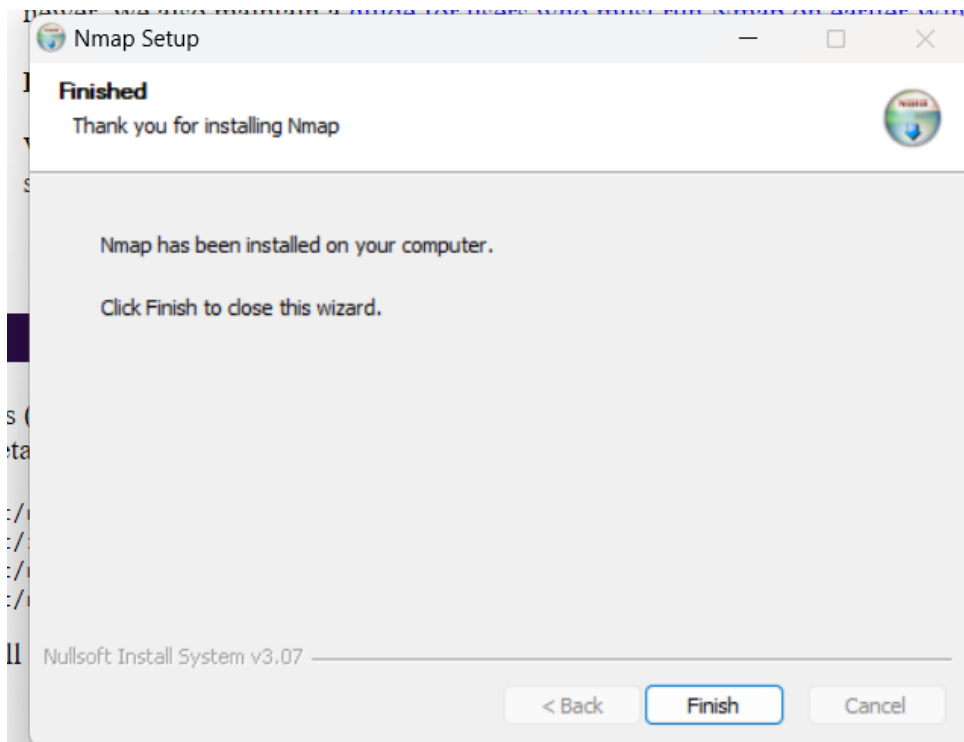


The screenshot shows the Zenmap GUI on the left, displaying a scan of localhost (127.0.0.1) with open ports 22/tcp (ssh) and 22/tcp (ssh). On the right, there is a section titled 'Microsoft Windows binaries' with text about the Windows version of Nmap and a link to the 'Latest stable release self-installer: nmap-7.93-setup.exe'. Below this, there is a paragraph about post-install usage instructions.

Please read the [Windows section](#) of the Install Guide for limitations and installation instructions for the Windows version of Nmap. It's provided as an executable self-installer which includes Nmap's dependencies and the Zenmap GUI. We support Nmap on Windows 7 and newer, as well as Windows Server 2008 R2 and newer. We also maintain a [guide for users who must run Nmap on earlier Windows releases](#).

Latest stable release self-installer: [nmap-7.93-setup.exe](#)

We have written [post-install usage instructions](#). Please [notify us](#) if you encounter any problems or have suggestions for the installer.



4.2 Test Example

➔ Open Powershell to run:

ncat -lk 9999

```
PS C:\Users\Sharo> ncat -lk 9999
```

➔ Open another Powershell to run code below in spark-3.3.1-bin-hadoop3 folder:

./bin/spark-submit examples/src/main/python/streaming/network_wordcount.py localhost 9999

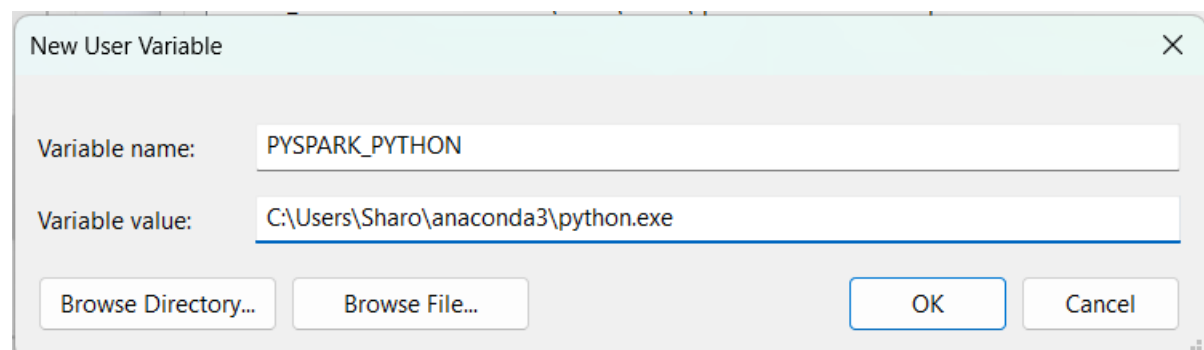
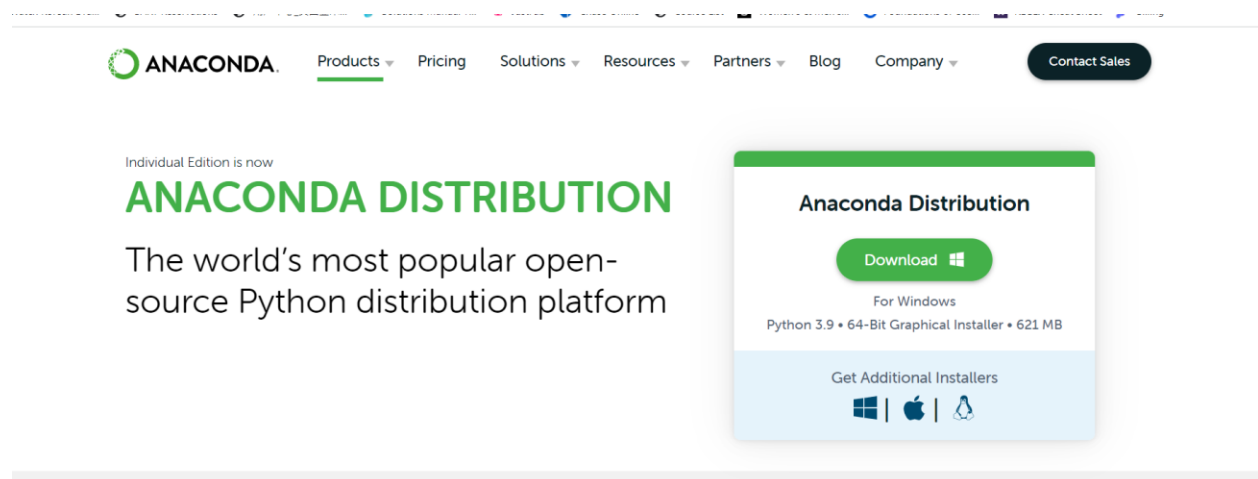
Error

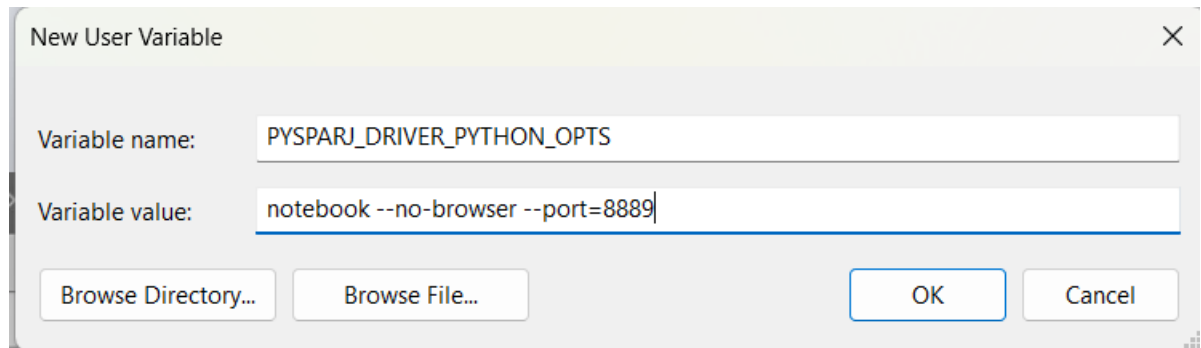
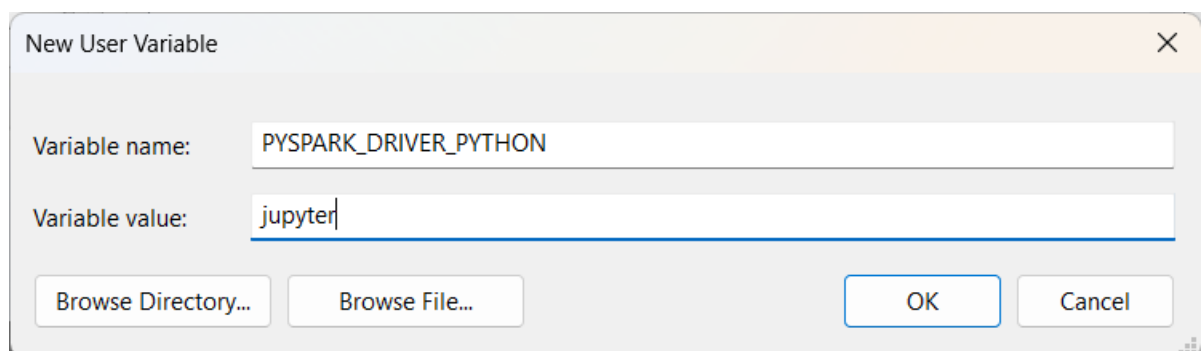
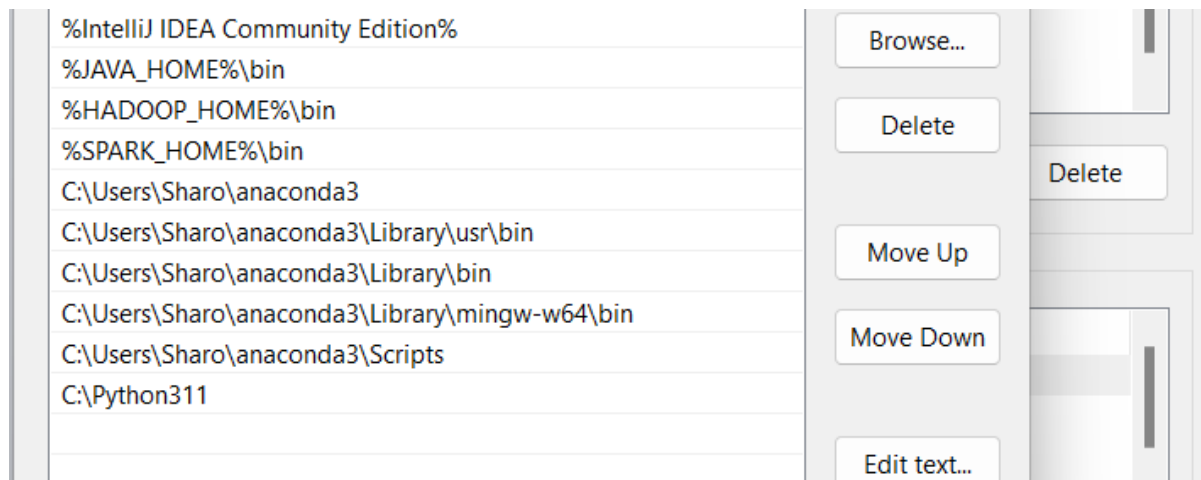
```
PS C:\Users\Sharo\spark-3.3.1-bin-hadoop3> ./bin/spark-submit examples/src/main/python/streaming/network_wordcount.py localhost 9999
Python was not found; run without arguments to install from the Microsoft Store, or disable this shortcut from Settings
> Manage App Execution Aliases.
22/12/05 21:03:51 INFO ShutdownHookManager: Shutdown hook called
22/12/05 21:03:51 INFO ShutdownHookManager: Deleting directory C:\Users\Sharo\AppData\Local\Temp\spark-fbe490fa-5801-4970-a73c-bad1551d4263
```

Tried to install python and set variables, not working

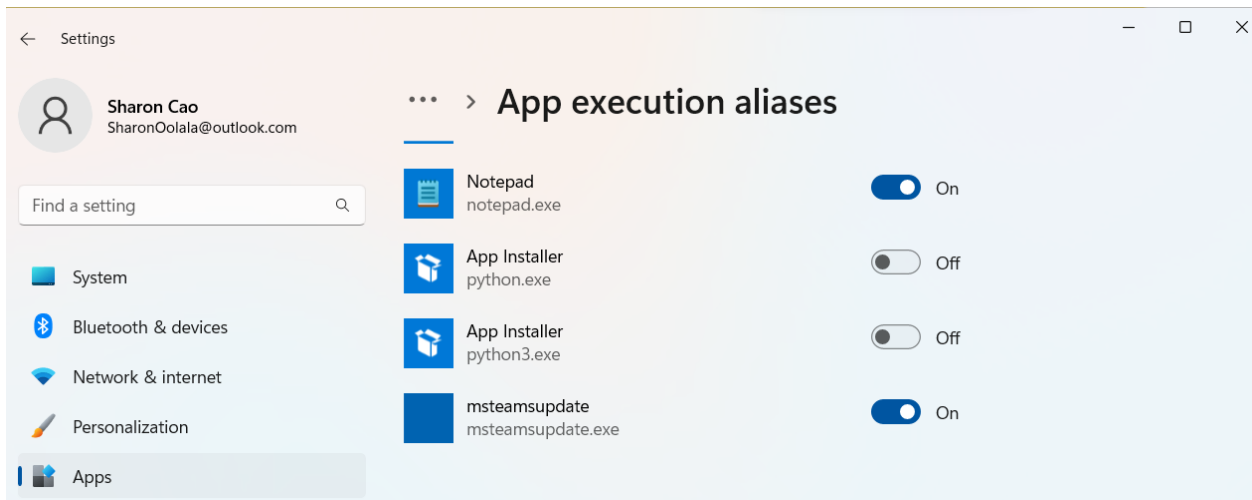
Python

<https://www.anaconda.com/products/distribution>





Fixed by turning of App Execution aliases



Run again, still error:

```
PS C:\Users\Sharo\spark-3.3.1-bin-hadoop3> ./bin/spark-submit examples/src/main/python/streaming/network_wordcount.py lo
calhost 9999
22/12/07 19:36:58 INFO SparkContext: Running Spark version 3.3.1
22/12/07 19:36:58 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
22/12/07 19:36:59 INFO ResourceUtils: =====
22/12/07 19:36:59 INFO ResourceUtils: No custom resources configured for spark.driver.
22/12/07 19:36:59 INFO ResourceUtils: =====
22/12/07 19:36:59 INFO SparkContext: Submitted application: PythonStreamingNetworkWordCount
22/12/07 19:36:59 INFO ResourceProfile: Default ResourceProfile created, executor resources: Map(cores -> name: cores, a
mount: 1, script: , vendor: , memory -> name: memory, amount: 1024, script: , vendor: , offHeap -> name: offHeap, amount
: 0, script: , vendor: ), task resources: Map(cpus -> name: cpus, amount: 1.0)
22/12/07 19:36:59 INFO ResourceProfile: Limiting resource is cpu
22/12/07 19:36:59 INFO ResourceProfileManager: Added ResourceProfile id: 0
22/12/07 19:36:59 INFO SecurityManager: Changing view acls to: Sharo
22/12/07 19:36:59 INFO SecurityManager: Changing modify acls to: Sharo
22/12/07 19:36:59 INFO SecurityManager: Changing view acls groups to:
22/12/07 19:36:59 INFO SecurityManager: Changing modify acls groups to:
22/12/07 19:36:59 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view per
missions: Set(Sharo); groups with view permissions: Set(); users with modify permissions: Set(Sharo); groups with modif
y permissions: Set()
22/12/07 19:36:59 INFO Utils: Successfully started service 'sparkDriver' on port 25734.
22/12/07 19:36:59 INFO SparkEnv: Registering MapOutputTracker
22/12/07 19:36:59 INFO SparkEnv: Registering BlockManagerMaster
```

```
Traceback (most recent call last):
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\streaming\util.py", line 71, in call
    r = self.func(t, *rdds)
    ^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\streaming\dstream.py", line 918, in <lambda
a>
    ] = lambda t, rdd: func(t, prev_func(t, rdd))
    ^^^^^^^^^^^^^^^^^
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\streaming\dstream.py", line 410, in func
    return oldfunc(rdd) # type: ignore[arg-type, call-arg]
    ^^^^^^^^^^^^^^^^^
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\streaming\dstream.py", line 202, in func
    return rdd.combineByKey(createCombiner, mergeValue, mergeCombiners, numPartitions)
    ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\rdd.py", line 2558, in combineByKey
    shuffled = locally_combined.partitionBy(numPartitions, partitionFunc)
    ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\rdd.py", line 2486, in partitionBy
    pairRDD = self.ctx._jvm.PairwiseRDD(keyed._jrd.rdd()).asJavaPairRDD()
    ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Sharo\spark-3.3.1-bin-hadoop3\python\lib\pyspark.zip\pyspark\rdd.py", line 3505, in _jrd
    wrapped_func = _wrap_function(
```

It seems problems occurred with build-in functions of spark packages.

Step 5: Try Kafka-python streaming

5.1 Open PowerShell and run code to create topic:

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --create --topic input_event --bootstrap-server localhost:9092 --partitions 3 --replication-factor 1
```

```
C:\Users\Sharo>C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --create --topic input_event --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.
Created topic input_event.
```

```
C:\Users\Sharo>
```

5.2 Create another topic:

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --create --topic output_event --bootstrap-server localhost:9092 --partitions 3 --replication-factor 1
```

```
C:\Users\Sharo>C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --create --topic output_event --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.
Created topic output_event.
```

```
C:\Users\Sharo>
```

5.3 List all topics:

```
C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --list --bootstrap-server localhost:9092
```

```
C:\Users\Sharo>C:/Users/Sharo/kafka_2.13-3.3.1/bin/windows/kafka-topics.bat --list --bootstrap-server localhost:9092
__consumer_offsets
helloworld
input_event
input_recommend_product
input_recomment_product
output_event
test_step_1
testing
```

```
C:\Users\Sharo>
```

5.4 Open another PowerShell and run:

```
spark-submit --jars C:\Users\Sharo\spark-3.3.1-bin-hadoop3\spark_script\spark-streaming-kafka_2.11-1.6.3.jar --master spark://abc.def.ghi.jkl:7077 --deploy-mode client C:\Users\Sharo\spark-3.3.1-bin-hadoop3\spark_script\spark_processor.py
```

```
C:\Users\Sharo>spark-submit --jars C:\Users\Sharo\spark-3.3.1-bin-hadoop3\spark_script\spark-streaming-kaf
ka_2.11-1.6.3.jar --master spark://abc.def.ghi.jkl:7077 --deploy-mode client C:\Users\Sharo\spark-3.3.1-b
in-hadoop3\spark_script\spark_processor.py
22/12/03 19:17:45 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using b
uiltin-java classes where applicable
Python was not found; run without arguments to install from the Microsoft Store, or disable this shortcut
from Settings > Manage App Execution Aliases.
22/12/03 19:17:46 INFO ShutdownHookManager: Shutdown hook called
22/12/03 19:17:46 INFO ShutdownHookManager: Deleting directory C:\Users\Sharo\AppData\Local\Temp\spark-dbe
5a09d-cc01-4807-abc2-bd8a5fb6b34b
C:\Users\Sharo>
```

Still get error...