Assignment 10

1. Download **https://dlcdn.apache.org/spark/spark-3.5.5/spark-3.5.5-bin-hadoop3.tgz**

2. Extract it.

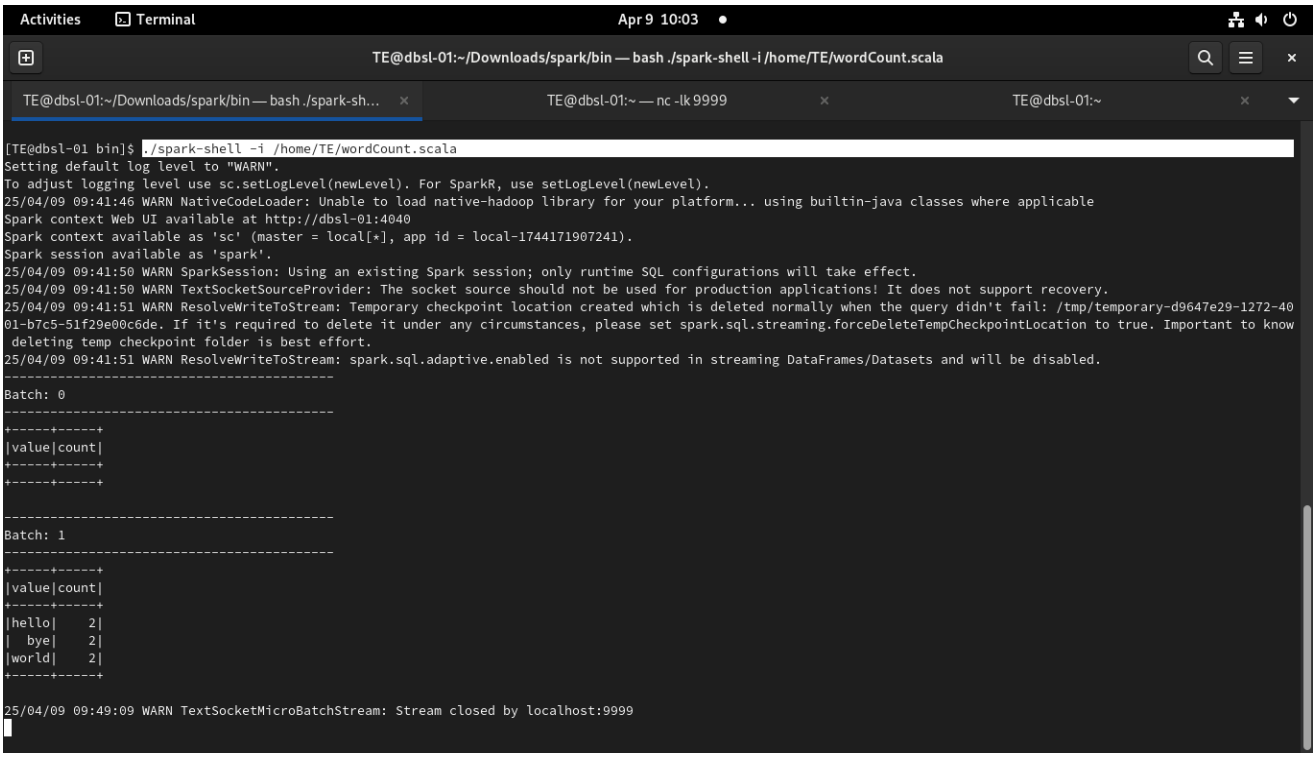
3.Open terminal and type the command: nc -lk 9999

4.Go to the extracted folder/bin and open terminal.

5.In the terminal type the command: ./spark-shell -i ***pathToYourScalaFile***

wordCount.scala =>

```
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._


val spark = SparkSession
  .builder
  .appName("StructuredNetworkWordCount")
  .master("local[*]")
  .getOrCreate()


import spark.implicits._


val lines = spark.readStream
  .format("socket")
  .option("host", "localhost")
  .option("port", 9999)
  .load()


val words = lines.as[String].flatMap(_.split(" "))


val wordCounts = words.groupBy("value").count()

val query = wordCounts.writeStream
  .outputMode("complete")
  .format("console")
  .start()

query.awaitTermination()
```

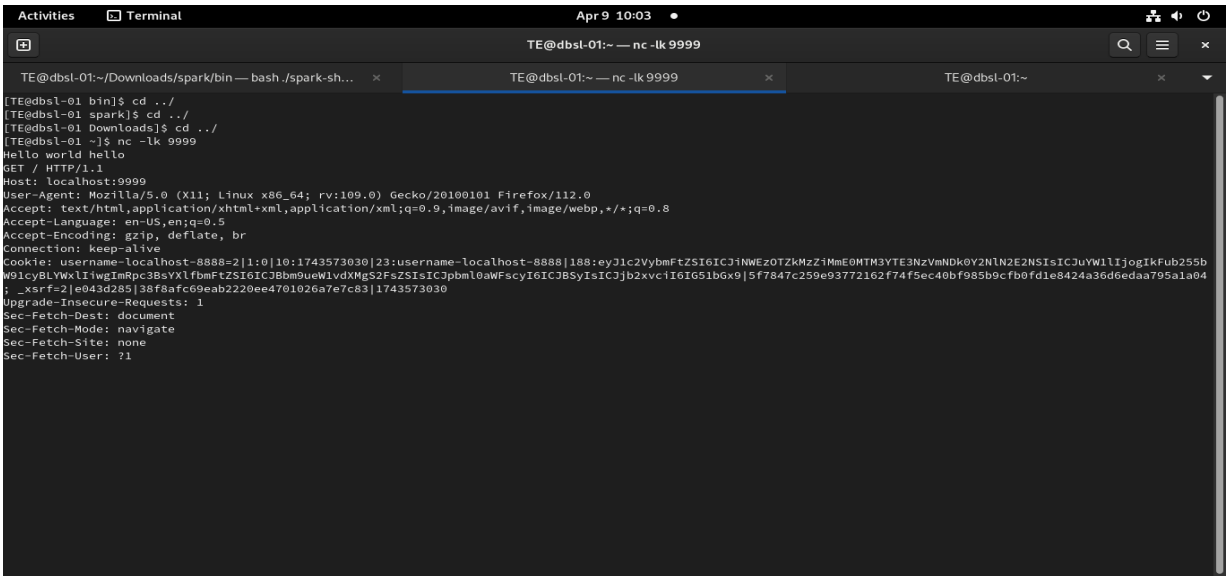

Activities
Terminal
Apr 9 10:03
TE@dbsl-01:~/Downloads/spark/bin — bash ./spark-shell -i /home/TE/wordCount.scala
TE@dbsl-01:~/Downloads/spark/bin — bash ./spark-sh...
TE@dbsl-01:~ — nc -lk 9999
TE@dbsl-01:~
[TE@dbsl-01 bin]$ ./spark-shell -i /home/TE/wordCount.scala
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
25/04/09 09:41:46 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context Web UI available at http://dbsl-01:4040
Spark context available as 'sc' (master = local[*], app id = local-1744171907241).
Spark session available as 'spark'.
25/04/09 09:41:50 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
25/04/09 09:41:50 WARN TextSocketSourceProvider: The socket source should not be used for production applications! It does not support recovery.
25/04/09 09:41:51 WARN ResolveWriteToStream: Temporary checkpoint location created which is deleted normally when the query didn't fail: /tmp/temporary-d9647e29-1272-4001-b7c5-51f29e00c6de. If it's required to delete it under any circumstances, please set spark.sql.streaming.forceDeleteTempCheckpointLocation to true. Important to know deleting temp checkpoint folder is best effort.
25/04/09 09:41:51 WARN ResolveWriteToStream: spark.sql.adaptive.enabled is not supported in streaming DataFrames/Datasets and will be disabled.

Batch: 0

+-----+-----+
|value|count|
+-----+-----+
+-----+-----+

Batch: 1

+-----+-----+
|value|count|
+-----+-----+
|hello| 2|
| bye| 2|
|world| 2|
+-----+-----+
25/04/09 09:49:09 WARN TextSocketMicroBatchStream: Stream closed by localhost:9999


Activities
Terminal
Apr 9 10:03
TE@dbsl-01:~ — nc -lk 9999
TE@dbsl-01:~/Downloads/spark/bin — bash ./spark-sh...
TE@dbsl-01:~ — nc -lk 9999
TE@dbsl-01:~
[TE@dbsl-01 bin]$ cd ../
[TE@dbsl-01 spark]$ cd ../
[TE@dbsl-01 Downloads]$ cd ../
[TE@dbsl-01 ~]$ nc -lk 9999
Hello world hello
GET / HTTP/1.1
Host: localhost:9999
User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:109.0) Gecko/20100101 Firefox/112.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate, br
Connection: keep-alive
Cookie: username-localhost-8888=2|1:0|10:1743573030|23:username-localhost-8888|188:eyJ1c2VybmFtZSI6ICJiNWEzOTZkMzZiMmE0MTM3YTE3NzVmNDk0Y2NlN2E2NSIsICJuYW1lIjogIkFub255bW91cyBLYWxlIiwgImRpc3BsYXlfbmFtZSI6ICJBbm9ueW1vdXMgS2FsZSIsICJpbml0aWFscyI6ICJBSyIsICJjb2xvciI6IG51bGx9|5f7847c259e93772162f74f5ec40bf985b9cfb0fd1e8424a36d6edaa795a1a04; _xsrf=2|e043d285|38f8afc69eab2220ee4701026a7e7c83|1743573030
Upgrade-Insecure-Requests: 1
Sec-Fetch-Dest: document
Sec-Fetch-Mode: navigate
Sec-Fetch-Site: none
Sec-Fetch-User: ?1