SCALA

1.Download apache spark:

* <https://dlcdn.apache.org/spark/spark-3.5.5/spark-3.5.5-bin-hadoop3.tgz>

2.On home terminal type the command(Run Netcat server on port 9999 - for streaming data) :

* nc -lk 9999

3.Navigate to the extracted Apache Spark folder and create a Scala file:

* Go to the bin directory
* Create wordCount.scala file with your Spark code

4.Open bin in terminal and type the command:

* ./spark-shell -i wordCount.scala

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() |
| --- |