

## 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()
```

