Roll No :31418
Batch: K4
Assignment 13

Step 1: Create a Folder to Work In

Open your terminal and run:

```
mkdir ~/spark-wordcount
cd ~/spark-wordcount
```

This is your working directory.

Step 2: Create `passage.txt`

```
nano passage.txt
```

Paste this inside:

```
hello world
hello spark
hello scala
I am Farkhanda
spark is fast
```

Then:

- Press `Ctrl + O`, then `Enter` to save.
- Press `Ctrl + X` to exit.

You now have a file named `passage.txt`.

Step 3: Start the Spark Shell

From the **same folder** (`~/spark-wordcount`), run:

```
spark-shell
```

This opens a Scala REPL with Spark already set up (`sc` is available).

Step 4: Run the Word Count Program

In the `spark-shell`:

```
val input = sc.textFile("passage.txt")

val words = input.flatMap(line => line.split(" "))

val counts = words.map(word => (word, 1))

val reducedCounts = counts.reduceByKey(_ + _)

reducedCounts.foreach(println)
```

OUTPUT:

```
(hello,3)
(world,1)
(spark,2)
(is,1)
(fast,1)
(scala,1)
```