

```
Activities Terminal Apr 17 15:51 root@student: /home/student/DSBDAL
root@student: /home/student/DSBDAL
root@student: /home/student/DSBDAL/wc_out
su
(base) student@student: $ su
Password:
root@student:/home/student# /home/student/DSBDAL
bash: /home/student/DSBDAL: Is a directory
root@student:/home/student# cd /home/student/DSBDAL
root@student:/home/student/DSBDAL# nano wc.txt
root@student:/home/student/DSBDAL# sprak-shell
sprak-shell: command not found
root@student:/home/student/DSBDAL# spark-shell
24/04/17 15:28:25 WARN Utils: Your hostname, student resolves to a loopback address: 127.0.1.1; using 10.11.5.21 instead (on interface enp3s0)
24/04/17 15:28:25 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/04/17 15:28:31 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context Web UI available at http://10.11.5.21:4040
Spark context available as 'sc' (master = local[*], app id = local-1713347911994).
Spark session available as 'spark'.
Welcome to

  _____
 /_ _ _ _ _ \
|  _ _ _ _ |
| | _ _ _ |
| | _ _ _ |
| | _ _ _ |
| | _ _ _ |
|_|_ _ _ _|

version 3.3.1

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 11.0.22)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val inputfile = sc.textfile("wc.txt")
<console>:23: error: value textfile is not a member of org.apache.spark.SparkContext
    val inputfile = sc.textfile("wc.txt")
                        ^
scala> val inputfile = sc.textFile("wc.txt")
inputfile: org.apache.spark.rdd.RDD[String] = wc.txt MapPartitionsRDD[1] at textFile at <console>:23
```

```
Activities Terminal Apr 17 15:51 root@student: /home/student/DSBDAL
root@student: /home/student/DSBDAL
root@student: /home/student/DSBDAL/wc_out
/_/_/

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 11.0.22)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val inputfile = sc.textfile("wc.txt")
<console>:23: error: value textfile is not a member of org.apache.spark.SparkContext
    val inputfile = sc.textfile("wc.txt")
                        ^
scala> val inputfile = sc.textFile("wc.txt")
inputfile: org.apache.spark.rdd.RDD[String] = wc.txt MapPartitionsRDD[1] at textFile at <console>:23

scala> val counts = inputfile.flatMap(line => line.split(" ")).map(word, 1)).reduceByKey(_+_);
<console>:1: error: ';' expected but ')' found.
    val counts = inputfile.flatMap(line => line.split(" ")).map(word, 1)).reduceByKey(_+_);
                                   ^
scala> val counts = inputfile.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(_+_);
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:23

scala> counts.toDebugString
res0: String =
(2) ShuffledRDD[4] at reduceByKey at <console>:23 []
+- (2) MapPartitionsRDD[3] at map at <console>:23 []
|   MapPartitionsRDD[2] at flatMap at <console>:23 []
|   wc.txt MapPartitionsRDD[1] at textFile at <console>:23 []
|   wc.txt HadoopRDD[0] at textFile at <console>:23 []

scala> counts.cache()
res1: counts.type = ShuffledRDD[4] at reduceByKey at <console>:23

scala> counts.saveAsTextFile("wc_out")

scala>
```

```
Activities Terminal Apr 17 15:51
root@student: /home/student/DSBDAL/wc_out

root@student: /home/student/DSBDAL
root@student: /home/student/DSBDAL/wc_out

su
(base) student@student: $ su
Password:
root@student: /home/student# cd /home/student/DSBDAL
root@student: /home/student/DSBDAL# wc_out# ls
wc_out#: command not found
root@student: /home/student/DSBDAL# wc_out# ls
wc_out#: command not found
root@student: /home/student/DSBDAL# cd wc_out
root@student: /home/student/DSBDAL/wc_out# ls
part-00000 part-00001 _SUCCESS
root@student: /home/student/DSBDAL/wc_out# cat part-00000
(are,1)
(review,1)
(ls,1)
(am,3)
(girl,1)
(our,1)
(we,1)
(,1)
(there,1)
(hello,1)
(class,2)
(successful,1)
root@student: /home/student/DSBDAL/wc_out# ls
part-00000 part-00001 _SUCCESS
root@student: /home/student/DSBDAL/wc_out# cat part-00000
(are,1)
(review,1)
(ls,1)
(am,3)
(girl,1)
(our,1)
(we,1)
(,1)
(there,1)
```

```
Activities Terminal Apr 17 15:51
root@student: /home/student/DSBDAL/wc_out

root@student: /home/student/DSBDAL
root@student: /home/student/DSBDAL/wc_out

root@student: /home/student/DSBDAL/wc_out# ls
wc_out#: command not found
root@student: /home/student/DSBDAL# wc_out# ls
wc_out#: command not found
root@student: /home/student/DSBDAL# cd wc_out
root@student: /home/student/DSBDAL/wc_out# ls
part-00000 part-00001 _SUCCESS
root@student: /home/student/DSBDAL/wc_out# cat part-00000
(are,1)
(review,1)
(ls,1)
(am,3)
(girl,1)
(our,1)
(we,1)
(,1)
(there,1)
(hello,1)
(class,2)
(successful,1)
root@student: /home/student/DSBDAL/wc_out# ls
part-00000 part-00001 _SUCCESS
root@student: /home/student/DSBDAL/wc_out# cat part-00000
(are,1)
(review,1)
(ls,1)
(am,3)
(girl,1)
(our,1)
(we,1)
(,1)
(there,1)
(hello,1)
(class,2)
(successful,1)
root@student: /home/student/DSBDAL/wc_out#
```

The screenshot displays the Apache Spark Jobs application interface in a web browser. The browser's address bar shows the URL `localhost:4040/jobs/`. The application header includes the Apache Spark logo and version `3.3.1`, along with navigation tabs for `Jobs`, `Stages`, `Storage`, `Environment`, and `Executors`. The `Jobs` tab is active, showing the title `Spark Jobs (?)` and the following summary information:

- User: root
- Total Uptime: 22 min
- Scheduling Mode: FIFO
- Completed Jobs: 1

Below the summary, there is a section for the `Event Timeline` with a checkbox for `Enable zooming`. The main area features a timeline chart with two categories: `Executors` and `Jobs`. The `Executors` category includes `Added` (blue) and `Removed` (pink) events. The `Jobs` category includes `Succeeded` (blue), `Failed` (pink), and `Running` (green) events. A single event, `Executor driver added`, is shown as a blue dot on the timeline at approximately 15:28:30 on Wednesday, April 17. The timeline axis is labeled with times from 15:28 to 15:33. At the bottom, a section titled `Completed Jobs (1)` shows a pagination control for page 1 of 1, with options to show 100 items per page. A table header is visible at the very bottom, listing columns: `Job Id`, `Description`, `Submitted`, `Duration`, `Stages: Succeeded/Total`, and `Tasks (for all stages): Succeeded/Total`.