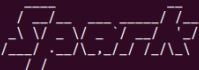


Apache Sparks and SCALA Assignment

```
Activities Terminal Apr 12 09:26 root@student:/home/student/sid438/DSBDAL

scala> (base) student@student:~$ su
Password:
root@student:/home/student# cd sid438
root@student:/home/student/sid438# cd DSBDAL
root@student:/home/student/sid438/DSBDAL# nano input.txt
root@student:/home/student/sid438/DSBDAL# spark-shell
24/04/12 09:13:31 WARN Utils: Your hostname, student resolves to a loopback address: 127.0.1.1; using 10.11.5.76 instead (on interface enp3s0)
24/04/12 09:13:31 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).
24/04/12 09:13:39 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Spark context Web UI available at http://10.11.5.76:4040
Spark context available as 'sc' (master = local[*], app id = local-1712893421056).
Spark session available as 'spark'.
Welcome to

 version 3.3.1

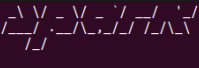
Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 11.0.22)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val inputFile = sc.textFile("input.txt")
InputFile: org.apache.spark.rdd.RDD[String] = input.txt MapPartitionsRDD[1] at textFile at <console>:23

scala> val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
<console>:22: error: not found: value inputFile
      val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
                        ^

scala> val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
<console>:23: error: value reduceByKey is not a member of (String, Int)
      val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
                                          ^
```

```
Activities Terminal Apr 12 09:26 root@student:/home/student/sid438/DSBDAL

 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("input.txt")
InputFile: org.apache.spark.rdd.RDD[String] = input.txt MapPartitionsRDD[1] at textFile at <console>:23

scala> val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
<console>:22: error: not found: value inputFile
      val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
                        ^

scala> val counts = inputFile.flatMap(line => line.split(" ")).map(word => (word,1).reduceByKey(_+_))
<console>:23: error: value reduceByKey is not a member of (String, Int)
      val counts = inputFile.flatMap(line => line.split(" ")).map(word => (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 []
|   input.txt MapPartitionsRDD[1] at textFile at <console>:23 []
|   input.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_output")

scala>
```

Activities Google Chrome Apr 12 09:26

Spark shell - Spark Jobs x how to take screenshot x +

localhost:4040/jobs/

Spark 3.3.1 Jobs Stages Storage Environment Executors Spark shell application UI

Spark Jobs (?)

User: root
Total Uptime: 11 min
Scheduling Mode: FIFO
Completed Jobs: 1

Event Timeline

Completed Jobs (1)

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	runJob at SparkHadoopWriter.scala:83 runJob at SparkHadoopWriter.scala:83	2024/04/12 09:22:23	0.8 s	2/2	4/4

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

Activities Google Chrome Apr 12 09:26

Spark shell - Executors x how to take screenshot x +

localhost:4040/executors/

Executors

Show Additional Metrics

Summary

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write	Excluded
Active(1)	2	33.6 KiB / 434.4 MiB	0.0 B	4	0	0	4	4	13 min (0.3 s)	122 B	290 B	290 B	0
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0.0 ms (0.0 ms)	0.0 B	0.0 B	0.0 B	0
Total(1)	2	33.6 KiB / 434.4 MiB	0.0 B	4	0	0	4	4	13 min (0.3 s)	122 B	290 B	290 B	0

Executors

Show 20 entries Search:

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read	Shuffle Write
driver	10.11.5.76:39651	Active	2	33.6 KiB / 434.4 MiB	0.0 B	4	0	0	4	4	13 min (0.3 s)	122 B	290 B	290 B

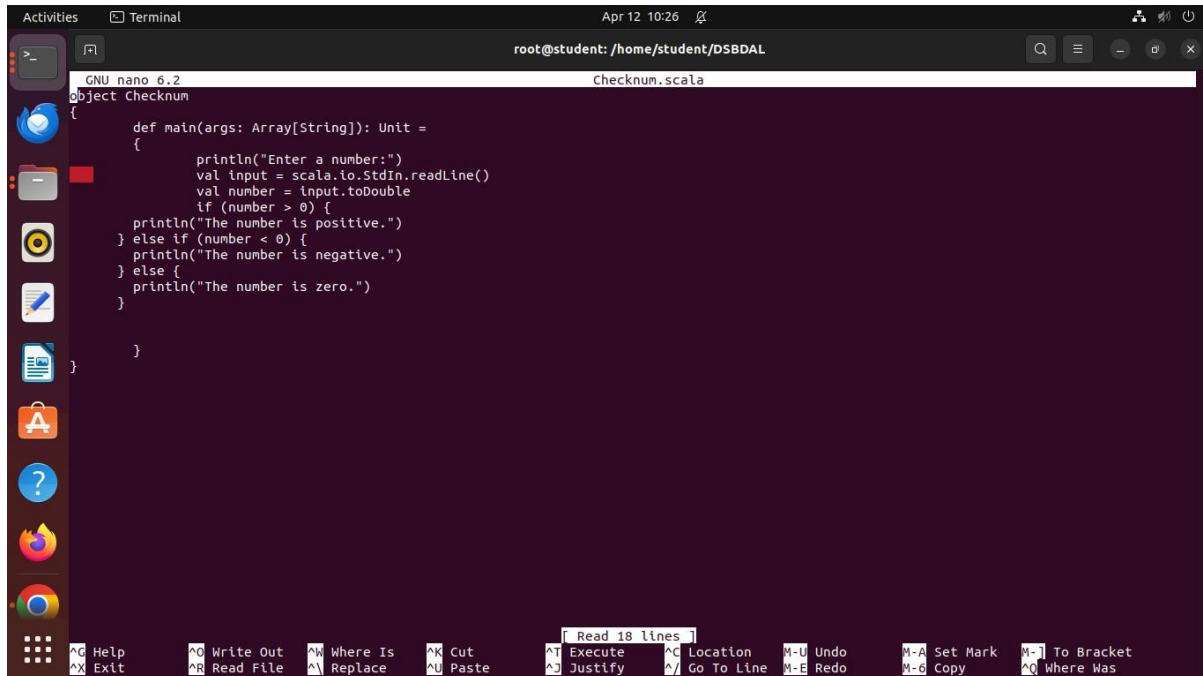
Showing 1 to 1 of 1 entries Previous 1 Next

Activities Terminal Apr 12 09:29

root@student: /home/student/sid43B/DSBDAL/wc_output

```
(base) student@student:~$ su
Password:
root@student:/home/student# cd sid43B
root@student:/home/student/sid43B# cd DSBDAL
root@student:/home/student/sid43B/DSBDAL# cd wc_output
root@student:/home/student/sid43B/DSBDAL/wc_output# ls
part-000000 part-000001 SUCCESS
root@student:/home/student/sid43B/DSBDAL/wc_output# cat part-000000
(learning,1)
(is,3)
(an,1)
(very,1)
(Hil,1)
(My,1)
root@student:/home/student/sid43B/DSBDAL/wc_output# cat part-000001
(intersting,1)
(it,1)
(friday,1)
(palash,1)
(name,1)
(i,1)
(sapr,1)
(today,1)
root@student:/home/student/sid43B/DSBDAL/wc_output#
```

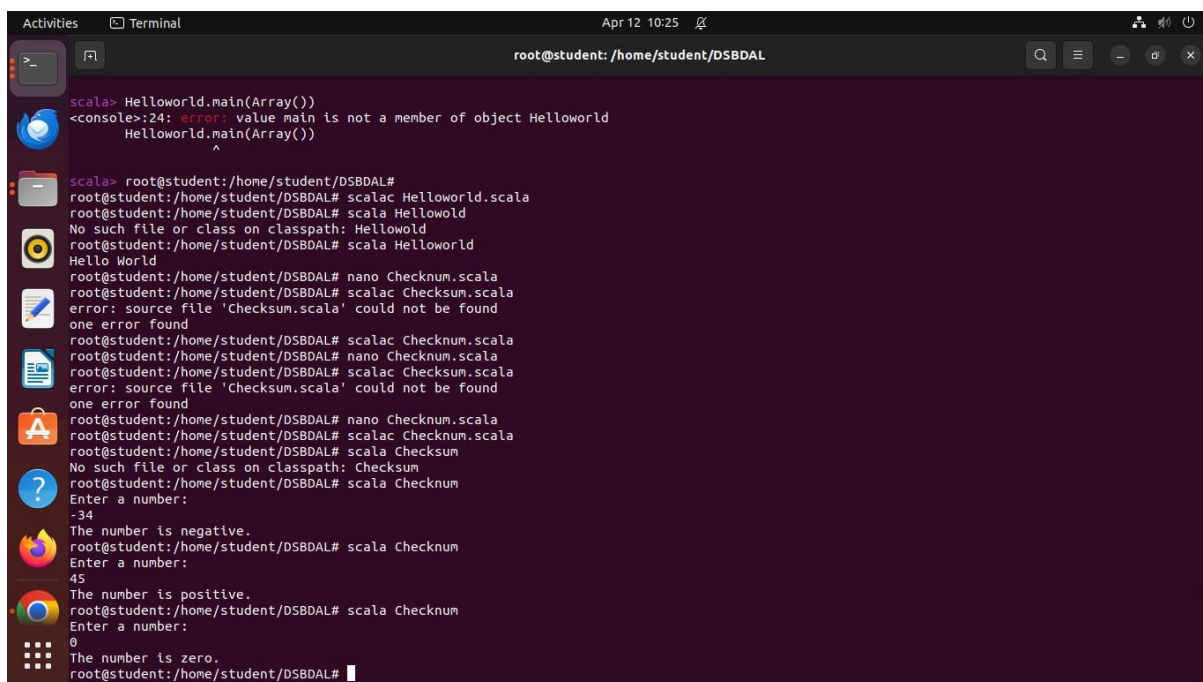
Check for a number is +ve , -ve or 0 using scala



The screenshot shows a terminal window with the nano 6.2 editor open. The file being edited is Checknum.scala. The code inside the editor is as follows:

```
object Checknum
{
    def main(args: Array[String]): Unit =
    {
        println("Enter a number:")
        val input = scala.io.StdIn.readLine()
        val number = input.toDouble
        if (number > 0) {
            println("The number is positive.")
        } else if (number < 0) {
            println("The number is negative.")
        } else {
            println("The number is zero.")
        }
    }
}
```

The terminal window title is "root@student: /home/student/DSBDAL". The bottom status bar of the nano editor shows various keyboard shortcuts like "Read 18 lines", "Execute", "Location", "Undo", "Set Mark", "To Bracket", "Where Was", "Write Out", "Read File", "Where Is", "Replace", "Cut", "Paste", "Justify", "Go To Line", and "Redo".



The screenshot shows a terminal window with the following commands and output:

```
scala> Helloworld.main(Array())
<console>:24: error: value main is not a member of object Helloworld
Helloworld.main(Array())
      ^
scala> root@student:/home/student/DSBDAL#
root@student:/home/student/DSBDAL# scalac Helloworld.scala
root@student:/home/student/DSBDAL# scala Hellowold
No such file or class on classpath: Hellowold
root@student:/home/student/DSBDAL# scala Helloworld
Hello World
root@student:/home/student/DSBDAL# nano Checknum.scala
root@student:/home/student/DSBDAL# scalac Checksum.scala
error: source file 'Checksum.scala' could not be found
one error found
root@student:/home/student/DSBDAL# scalac Checknum.scala
root@student:/home/student/DSBDAL# nano Checknum.scala
root@student:/home/student/DSBDAL# scalac Checksum.scala
error: source file 'Checksum.scala' could not be found
one error found
root@student:/home/student/DSBDAL# nano Checknum.scala
root@student:/home/student/DSBDAL# scalac Checknum.scala
root@student:/home/student/DSBDAL# scala Checksum
No such file or class on classpath: Checksum
root@student:/home/student/DSBDAL# scala Checknum
Enter a number:
-34
The number is negative.
root@student:/home/student/DSBDAL# scala Checknum
Enter a number:
45
The number is positive.
root@student:/home/student/DSBDAL# scala Checknum
Enter a number:
0
The number is zero.
root@student:/home/student/DSBDAL#
```