# Big Data Hadoop Training

Session 13 Assignment 3 Solution

**Write the code to Turn a collection into a RDD and perform map operation on it to cube every number and filter the number which are divided by two and three**.

1. Run all Spark Daemons and start “spark scala” shell:

Create a simple RDD :

**Scala> val input = sc.parallelize(List(2,12,3,6,1))**

Here sc is SparkContext.

We can see in the screenshot, a collection has been turned into a RDD[Int].

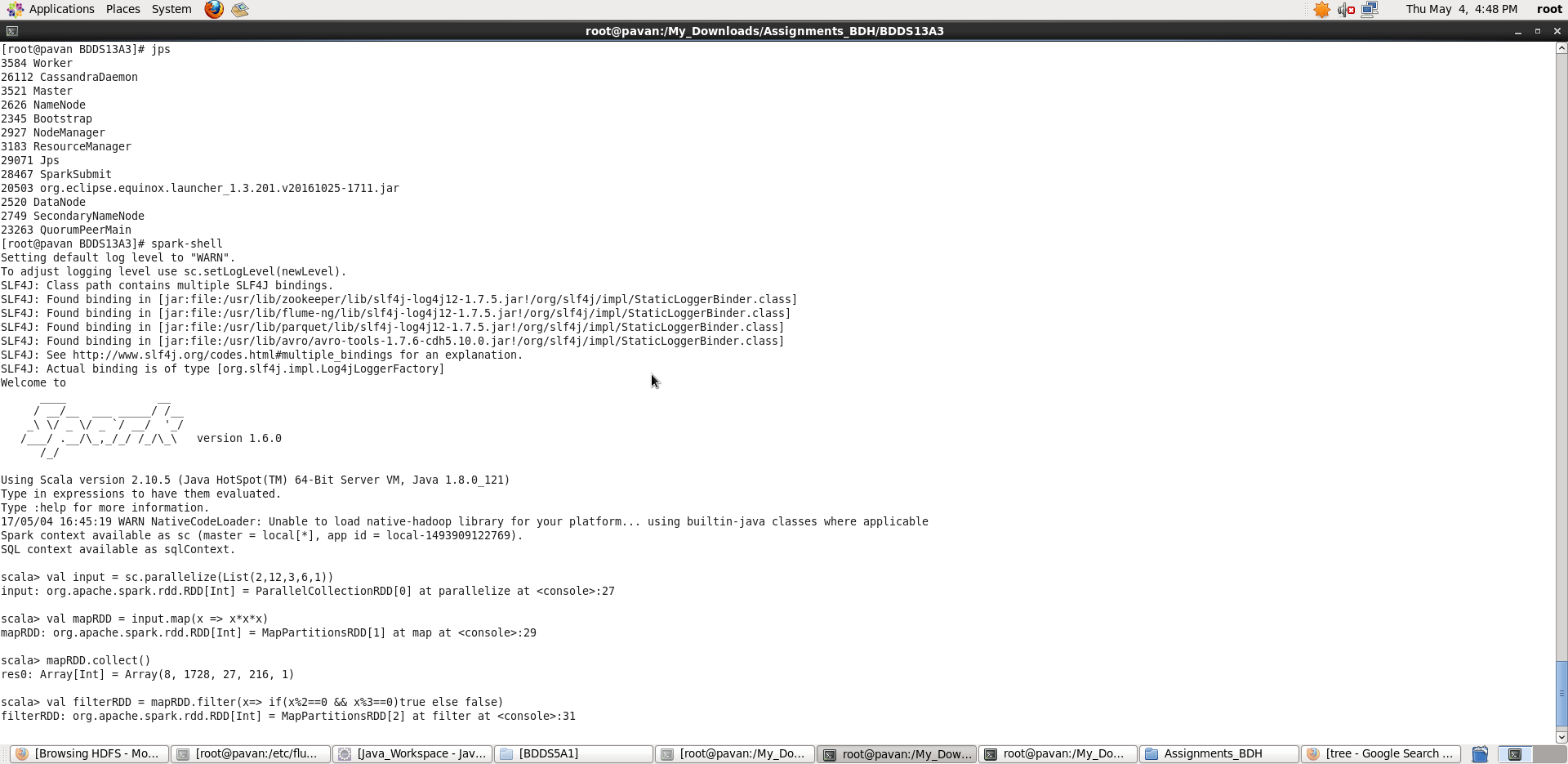
**Scala> val mapRDD = input.map(x => x\*x\*x)**

Then, I have used the map(func) function to find the cube of every element. We can see the output of map(func) function using collect() function.

**Scala> mapRDD.collect()**

**Map(func) –** returns a new distributed dataset formed by passing each element of the source through a function **func.**

**Collect() –** returns all the elements of the dataset as an array at the driver program



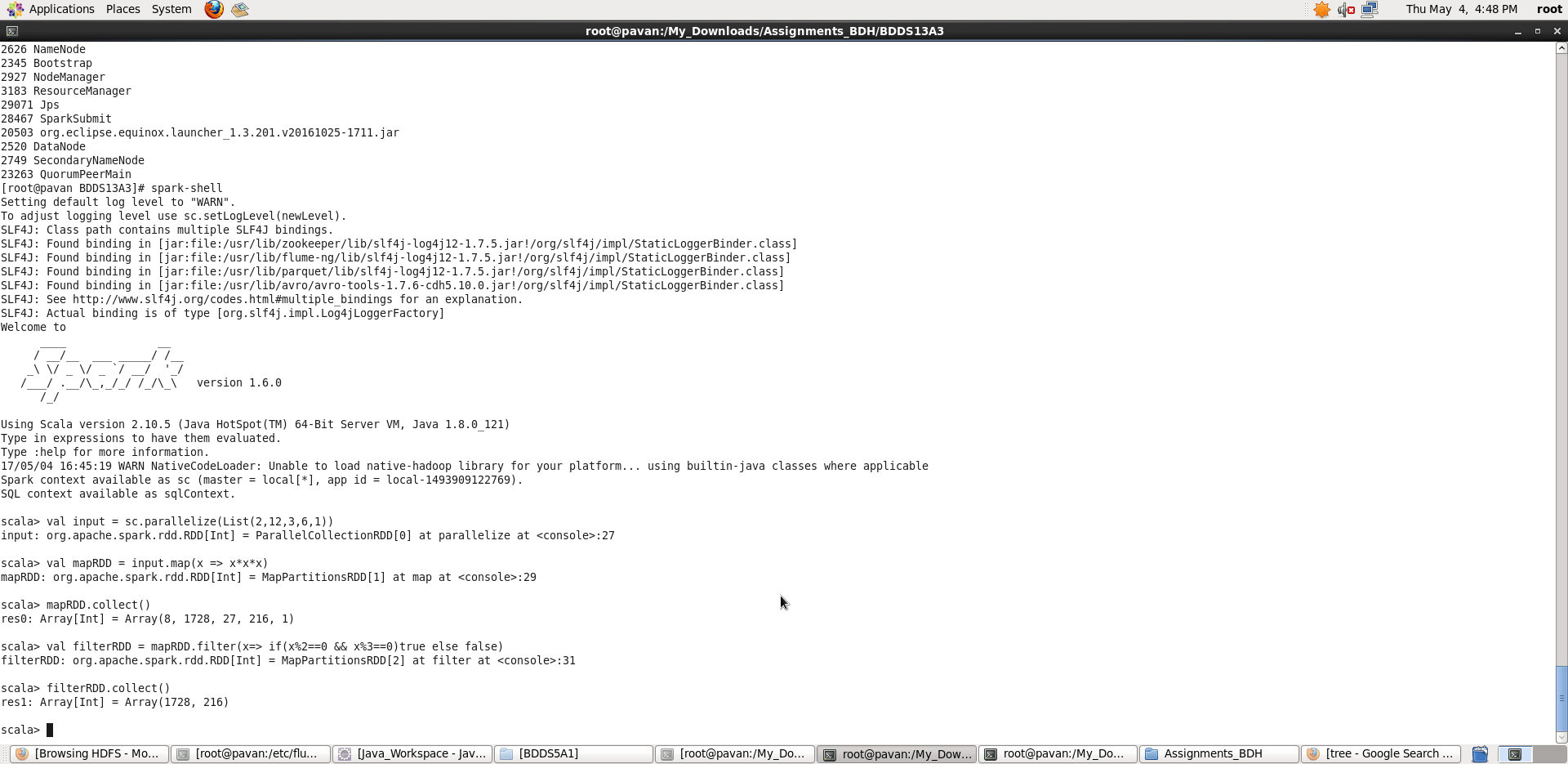
**Scala> val filterRDD = mapRDD.filter(x => if(x%2==0 && x%3==0)true else false)**

Finally, I have applied filter(func) on the mapRDD to get the desired result.

**filter(func) –** returns a new distributed dataset formed by selecting those elements of the source on which function **func** returns **true.**

**Scala> filterRDD.collect()**

**Final output is shown in below screenshot:**



Thus, in Scala prompt, we have performed the required task.