**Session 17 Assignment 1**

* **TASK 1**

scala> val inputFile = sc.textFile("/user/session17\_assignment1.txt")

inputFile: org.apache.spark.rdd.RDD[String] = /user/session17\_assignment1.txt MapPartitionsRDD[1] at textFile at <console>:27

scala> inputFile.count()

res0: Long = 4

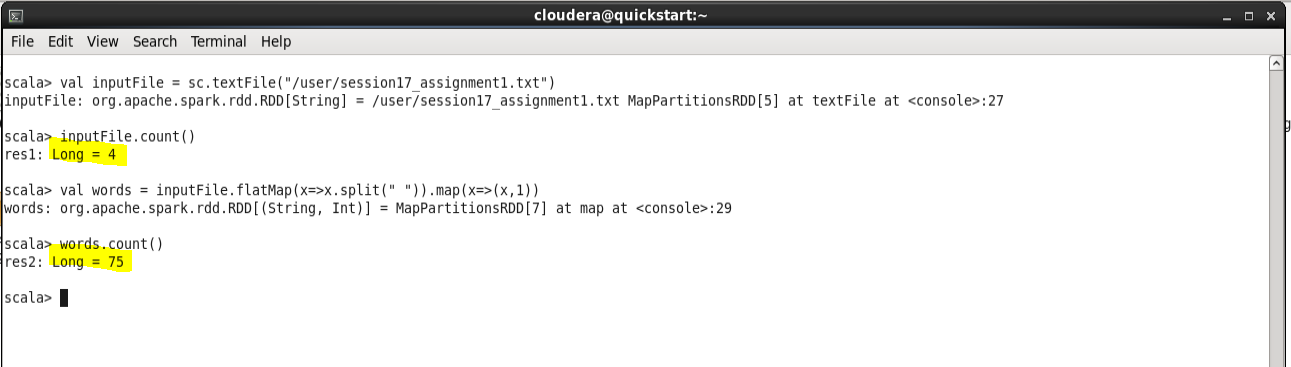
* **TASK 2**

scala> val words = inputFile.flatMap(x=>x.split(" ")).map(x=>(x,1))

words: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at <console>:29

scala> words.count()

res0: Long = 75



* **TASK 3**

scala> val inputSample = sc.textFile("/user/sample.txt")

inputSample: org.apache.spark.rdd.RDD[String] = /user/sample.txt MapPartitionsRDD[1] at textFile at <console>:27

scala> val flatmap = inputSample.flatMap(x=>x.split("-"))

flatmap: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <console>:29

scala> val words = flatmap.map(x=>(x,1))

words: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at <console>:31

scala> words.count()

res0: Long = 22

