1. **Write a program to read a text file and print the number of rows of data in the document.**

var sparkTextFile = sc.textFile("hdfs:///user/cloudera/SparkTextFile.txt")

sparkTextFile: org.apache.spark.rdd.RDD[String] = hdfs:///user/cloudera/SparkTextFile.txt MapPartitionsRDD[15] at textFile at <console>:27

sparkTextFile.count()

res27: Long = 20

1. **Write a program to read a text file and print the number of words in the document.**

val lines = sparkTextFile.flatMap(x => x.split(" "))

lines: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[16] at flatMap at <console>:29

lines.count()

res28: Long = 461

1. **We have a document where the word separator is -, instead of space. Write a spark**

**code, to obtain the count of the total number of words present in the document**.

var sparkDashTextFile = sc.textFile("hdfs:///user/cloudera/SparkDashTextFile.txt")

sparkDashTextFile: org.apache.spark.rdd.RDD[String] = hdfs:///user/cloudera/SparkDashTextFile.txt MapPartitionsRDD[18] at textFile at <console>:27

val lines = sparkDashTextFile.flatMap(x => x.split("-"))

lines: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[20] at flatMap at <console>:29

lines.count()

res30: Long = 477