student@student-OptiPlex-9020:~$ gedit word\_count.scala

student@student-OptiPlex-9020:~$ gedit 31181.txt

student@student-OptiPlex-9020:~$ java -version

openjdk version "11.0.22" 2024-01-16

OpenJDK Runtime Environment (build 11.0.22+7-post-Ubuntu-0ubuntu220.04.1)

OpenJDK 64-Bit Server VM (build 11.0.22+7-post-Ubuntu-0ubuntu220.04.1, mixed mode, sharing)

student@student-OptiPlex-9020:~$ scala -version

Scala code runner version 2.11.12 -- Copyright 2002-2017, LAMP/EPFL

student@student-OptiPlex-9020:~$ spark-shell

24/04/24 10:22:47 WARN Utils: Your hostname, student-OptiPlex-9020 resolves to a loopback address: 127.0.1.1; using 10.10.15.34 instead (on interface eno1)

24/04/24 10:22:47 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).

24/04/24 10:22:51 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Spark context Web UI available at http://10.10.15.34:4040

Spark context available as 'sc' (master = local[\*], app id = local-1713934372677).

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> load word\_count.scala

<console>:1: error: ';' expected but '.' found.

load word\_count.scala

^

scala> :load word\_count.scala

Loading word\_count.scala...

data: org.apache.spark.rdd.RDD[String] = 31181.txt MapPartitionsRDD[1] at textFile at word\_count.scala:23

res0: Array[String] = Array(My name is Vivek, Rohil and Aniket are my friends., "")

splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at word\_count.scala:23

res1: Array[String] = Array(My, name, is, Vivek,, Rohil, and, Aniket, are, my, friends., "")

mapdata: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at word\_count.scala:23

res2: Array[(String, Int)] = Array((My,1), (name,1), (is,1), (Vivek,,1), (Rohil,1), (and,1), (Aniket,1), (are,1), (my,1), (friends.,1), ("",1))

reducedata: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at word\_count.scala:23

res3: Array[(String, Int)] = Array((are,1), (Aniket,1), (is,1), (Rohil,1), ("",1), (my,1), (My,1), (friends.,1), (name,1), (Vivek,,1), (and,1))

scala> :load word\_count.scala

Loading word\_count.scala...

data: org.apache.spark.rdd.RDD[String] = 31181.txt MapPartitionsRDD[6] at textFile at word\_count.scala:23

res4: Array[String] = Array(My name is Vivek, Rohil and Aniket are my friends., is is is is Rohil IS, "")

splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[7] at flatMap at word\_count.scala:23

res5: Array[String] = Array(My, name, is, Vivek,, Rohil, and, Aniket, are, my, friends., is, is, is, is, Rohil, IS, "")

mapdata: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[8] at map at word\_count.scala:23

res6: Array[(String, Int)] = Array((My,1), (name,1), (is,1), (Vivek,,1), (Rohil,1), (and,1), (Aniket,1), (are,1), (my,1), (friends.,1), (is,1), (is,1), (is,1), (is,1), (Rohil,1), (IS,1), ("",1))

reducedata: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[9] at reduceByKey at word\_count.scala:23

res7: Array[(String, Int)] = Array((are,1), (Aniket,1), (is,5), (IS,1), (Rohil,2), ("",1), (my,1), (My,1), (friends.,1), (name,1), (Vivek,,1), (and,1))

scala>