

MAP REDUCE IN SCALA USING SPARK SHELL

Step 1 :- Create Input file word_count.txt

The image shows a text editor window titled "word_count.txt". The text inside the editor is as follows:

During the 1980s, as the computer revolution was sweeping the world, empowering ambitious professionals to bring their dreams to reality, Pune Institute of Computer Technology (PICT) was established in 1983 under the aegis of Society for Computer Technology and Research (SCTR) by noble personalities namely, Shri. Gopal Krishna, Shri. N. Ramakrishna, and Shri. R. S. Kothavale. The goal of the great visionaries was to make PICT the leader in higher technical education and research and thereby contribute towards the growth of the organization and society at large by attracting and sustaining gifted individuals who have enormous potential.

Taking into account, the ever-increasing demand for Computer Engineers, globally, PICT started the Bachelor of Engineering (B.E.) programme in Computer Engineering with affiliation to University of Pune (now Savitribai Phule Pune University (SPPU)). PICT is the first self-financed institute to start a Computer Engineering (B.E.) programme in Maharashtra, with an intake capacity of 60 which was subsequently increased to 120 in the Academic Year (A.Y.) 1984-85, and at present, is 240. In the year 1995, considering the rapid growth in the electronic industry and the huge demand for competent engineers, the institute started Electronics & Telecommunication Engineering (B.E.), programme with an intake of 60, and at present, it is 240. In A.Y. 2001-02, the Information Technology (B.E.) programme was started to meet the requirements of the booming IT industry, with an intake of 60, which is now 180.

The institute has been a pioneer in Post Graduation of Computer Engineering (M.E.) among the un-aided Engineering Colleges affiliated to SPPU since the year 2000. Subsequently, the Post-Graduate (PG) programmes in Computer Engineering, Data Science, Electronics & Communication (Wireless Communication Technology), and Information Technology were added. PICT is also a Savitribai Phule Pune University (SPPU) recognized Research Center in Computer Engineering and Electronics & Telecommunication Engineering.

To inculcate research culture and pave the way to a career in industries/organizations that are research and innovation oriented, PICT started Ph.D. programme research center approved by SPPU, in Computer Engineering and Electronics & Telecommunication Engineering in the A.Y. 2014-15.

The philosophy of the founding members' which emphasizes "Quality in Engineering Education and Research" is continuously nurtured and carried forward by the Management, Director, Principal, staff and students; to make PICT a leader in technical education and research of international repute by becoming most sought after destination for students from India and abroad.

Step2 :- Open Spark Shell in Terminal

```
[TE@fedora ~]$ spark-shell
```

Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).

24/03/14 12:00:30 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Spark context Web UI available at http://fedora:4040
Spark context available as 'sc' (master = local[*], app id = local-1710397831531).
Spark session available as 'spark'.
Welcome to

```
 _   _   _____      _         _____    _____ 
| | | | \_ / ____ \|_| |_|       |_____)___/_\_/_____||__| | | |
| |_| | ) / |__) || | | |        |____)___/_\_/_____||__|
|  __/ / ___ \|_| |_|       |_____)___/_\_/_____||__|
|_|_____\_/\_\\_||_|_|_     |_____)___/_\_/_____||__|

version 3.3.2
```

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 17.0.7)
Type in expressions to have them evaluated.
Type :help for more information.

```
scala> 
```

Step3 :- Write the Scala program word_count.scala using Map-Reduce

```
val data=sc.textFile("word_count.txt")
data.collect;
val splitdata = data.flatMap(line => line.split(" "));
splitdata.collect;
val mapdata = splitdata.map(word => (word,1));
mapdata.collect;
val reducedata = mapdata.reduceByKey(_+_);
reducedata.collect;
```

Step4 :- Load and Execute the word_count.scala file

```
[TE@fedora ~]$ spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/03/14 12:00:30 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
Spark context Web UI available at http://fedora:4040
Spark context available as 'sc' (master = local[*], app id = local-1710397831531).
Spark session available as 'spark'.
Welcome to

      ____ _ 
     / ___ \_/_  __\__  /\_/\___\//_____\_____
    /  _ \| | | ||| |\_| |_||_\_\\_____/_____/_____\
   /  ___ \| |_| ||| |\_| |_||_\_\\_____/_____/_____\
  /_____\|_____|_____|_____|_____|_____|_____|_____|
                                     version 3.3.2

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 17.0.7)
Type in expressions to have them evaluated.
Type :help for more information.

scala> :load word_count.scala
Loading word_count.scala...
data: org.apache.spark.rdd.RDD[String] = word_count.txt MapPartitionsRDD[1] at textFile at word_count.scala:23
res0: Array[String] = Array(During the 1980s, as the computer revolution was sweeping the world, empowering ambitious prof
essionals to bring their dreams to reality, Pune Institute of Computer Technology (PICT) was established in 1983 under the
aegis of Society for Computer Technology and Research (SCTR) by noble personalities namely, Shri. Gopal Krishna, Shri. N.
Ramakrishna, and Shri. R. S. Kothavale. The goal of the great visionaries was to make PICT the leader in higher technical
education and research and thereby contribute towards the growth of the organization and society at large by attracting a
nd sustaining gifted individuals who have enormous potential., "", Taking into account, the ever-increasing demand for Com
puter Engineers, globally, PICT started...)
splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at word_count.scala:23

Activities Terminal Mar 14 12:01 TE@fedora:~ — bash /usr/local/spark/bin/spark-shell

aegis of Society for Computer Technology and Research (SCTR) by noble personalities namely, Shri. Gopal Krishna, Shri. N.
Ramakrishna, and Shri. R. S. Kothavale. The goal of the great visionaries was to make PICT the leader in higher technical
education and research and thereby contribute towards the growth of the organization and society at large by attracting a
nd sustaining gifted individuals who have enormous potential., "", Taking into account, the ever-increasing demand for Com
puter Engineers, globally, PICT started...
splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at word_count.scala:23
res1: Array[String] = Array(During, the, 1980s,, as, the, computer, revolution, was, sweeping, the, world,, empowering, am
bitious, professionals, to, bring, their, dreams, to, reality,, Pune, Institute, of, Computer, Technology, (PICT), was, es
tablished, in, 1983, under, the, aegis, of, Society, for, Computer, Technology, and, Research, (SCTR), by, noble, personal
ities, namely,, Shri., Gopal, Krishna, Shri., N., Ramakrishna,, and, Shri., R., S., Kothavale., The, goal, of, the, great
visionaries, was, to, make, PICT, the, leader, in, higher, technical, education, and, research, and, thereby, contribute
towards, the, growth, of, the, organization, and, society, at, large, by, attracting, and, sustaining, gifted, individua
ls, who, have, enormous, potential., ""...)
mapdata: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at word_count.scala:23
res2: Array[(String, Int)] = Array((During,1),(the,1),(1980s,,1),(as,1),(the,1),(computer,1),(revolution,1),(was,1)
,(sweeping,1),(the,1),(world,,1),(empowering,1),(ambitious,1),(professionals,1),(to,1),(bring,1),(their,1),(drea
ms,1),(to,1),(reality,,1),(Pune,1),(Institute,1),(of,1),(Computer,1),(Technology,1),((PICT),1),(was,1),(establis
hed,1),(in,1),(1983,1),(under,1),(the,1),(aegis,1),(of,1),(Society,1),(for,1),(Computer,1),(Technology,1),(and,
1),(Research,1),((SCTR),1),(by,1),(noble,1),(personalities,1),(namely,,1),(Shri.,1),(Gopal,1),(Krishna,1),(Shri
.,1),(N.,1),(Ramakrishna,1),(and,1),(Shri.,1),(R.,1),(S.,1),(Kothavale.,1),(The,1),(goal,1),(of,1),(the,1),(g
reat,1),(visionaries,1),(was,1),(to,...))
reduceData: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at word_count.scala:23
res3: Array[(String, Int)] = Array(((1984-85,,1),(Communication,2),(under,1),(Technology,4),(Graduation,1),(A.Y.,2),(
2001-02,,1),(intake,3),(have,1),(60,,2),(Telecommunication,3),(Bachelor,1),(carried,1),(inculcate,1),(Research",1)
,(established,1),(globally,,1),(first,1),(been,1),(who,1),(namely,,1),(2014-15,,1),(Science,1),(Pune,4),(indust
ry,,1),(Savitribai,2),(continuously,1),(SPPU,,1),(make,2),(Phule,2),(8&,4),(Data,1),(PICT,6),(programme,5),(be
coming,1),(Subsequently,,1),(capacity,1),(rapid,1),(personalities,1),(are,1),(60,1),(enormous,1),((B.E.),3),(Ph.D
.,1),(SPPU,1),(Colleges,1),(sought,1),(recognized,1),(society,1),(into,1),(as,1),(1995,,1),(international,1),(s
ince,1),(affiliation,1),(destination,...))

scala>
```