


```
Administrator: Command Prompt - spark-shell
Microsoft Windows [Version 10.0.19045.4046]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\system32>spark-shell
24/02/20 15:20:25 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: Hadoop bin directory does not exist: C:\mydrive\hadoop\bin\bin -see https://wiki.apa
che.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://DESKTOP-086K1H4:4040
Spark context available as 'sc' (master = local[*], app id = local-1708422636250).
Spark session available as 'spark'.
Welcome to

 version 3.5.0

Using Scala version 2.12.18 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.
```

```
scala> val text = sc.textFile("C:/wordcount_sandeeep.txt")
text: org.apache.spark.rdd.RDD[String] = C:/wordcount_sandeeep.txt MapPartitionsRDD[7] at textFile at <console>:23

scala> text.collect
res5: Array[String] = Array("In the realm of algorithms and code, a software engineer named Sandeep Kumar. ", "His days were a symphony of lines, debugging melodies, and in
novation harmonies. ", "One fateful night, a bug eluded his grasp, teasing his with its complexity. ", "With unwavering determination, Sandeep delved deep into its labyrinth
n, emerging victorious at dawn, his code pristine. ", A testament to his passion and resilience in the digital domain.)

scala> val counts = text.flatMap(line => line.split(" "))
counts: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[8] at flatMap at <console>:23

scala> counts.collect
res6: Array[String] = Array(In, the, realm, of, algorithms, and, code,, a, software, engineer, named, Sandeep, Kumar., His, days, were, a, symphony, of, lines,, debugging,
melodies,, and, innovation, harmonies., One, fateful, night,, a, bug, eluded, his, grasp,, teasing, his, with, its, complexity., With, unwavering, determination,, Sandeep,
delved, deep, into, its, labyrinth, emerging, victorious, at, dawn,, his, code, pristine., A, testament, to, his, passion, and, resilience, in, the, digital, domain.)

scala> val mapf = counts.map(word => (word,1))
mapf: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[9] at map at <console>:23

scala> mapf.collect
res7: Array[(String, Int)] = Array((In,1), (the,1), (realm,1), (of,1), (algorithms,1), (and,1), (code,,1), (a,1), (software,1), (engineer,1), (named,1), (Sandeep,1), (Kumar
.,1), (His,1), (days,1), (were,1), (a,1), (symphony,1), (of,1), (lines,,1), (debugging,1), (melodies,,1), (and,1), (innovation,1), (harmonies.,1), (One,1), (fateful,1), (ni
ght,1), (a,1), (bug,1), (eluded,1), (his,1), (grasp,,1), (teasing,1), (his,1), (with,1), (its,1), (complexity.,1), (With,1), (unwavering,1), (determination,1), (Sandeep,1
), (delved,1), (deep,1), (into,1), (its,1), (labyrinth,1), (emerging,1), (victorious,1), (at,1), (dawn,,1), (his,1), (code,1), (pristine.,1), (A,1), (testament,1), (to,1),
(his,1), (passion,1), (and,1), (resilience,1), (in,1), (the,1), (digital,1), (domain,,1))

scala> val reducef = mapf.reduceByKey(_+_ )
reducef: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[10] at reduceByKey at <console>:23

scala> reducef.collect
res8: Array[(String, Int)] = Array((his,4), (dawn,,1), (night,,1), (its,2), (domain.,1), (deep,1), (One,1), (melodies,,1), (debugging,1), (with,1), (Sandeep,2), (into,1), (
unwavering,1), (delved,1), (complexity.,1), (Kumar.,1), (pristine.,1), (emerging,1), (digital,1), (his,1), (with,1), (harmonies.,1), (algorithms,1), (bug,1), (fateful,1), (
grasp,,1), (resilience,1), (passion,1), (determination,1), (a,3), (teasing,1), (days,1), (symphony,1), (testament,1), (A,1), (code,,1), (code,1), (labyrinth,1), (at,1), (
in,1), (to,1), (of,2), (In,1), (software,1), (were,1), (victorious,1), (and,3), (named,1), (innovation,1), (engineer,1), (eluded,1), (realm,1), (the,2), (lines,,1))

scala> reducef.saveAsTextFile("C:/spark_output")

scala>
```

PC > Local Disk (C:) > spark_output

Name	Date modified	Type	Size
._SUCCESS.crc	2/20/2024 3:46 PM	CRC File	1 KB
.part-00000.crc	2/20/2024 3:46 PM	CRC File	1 KB
.part-00001.crc	2/20/2024 3:46 PM	CRC File	1 KB
._SUCCESS	2/20/2024 3:46 PM	File	0 KB
part-00000	2/20/2024 3:46 PM	File	1 KB
part-00001	2/20/2024 3:46 PM	File	1 KB

```
part-00000 - Notepad
File Edit Format View Help
(his,4)
(dawn,,1)
(night,,1)
(its,2)
(domain.,1)
(deep,1)
(One,1)
(melodies,,1)
(debugging,1)
(with,1)
(Sandeep,2)
(into,1)
(unwavering,1)
(delved,1)
(complexity.,1)
(Kumar.,1)
(pristine.,1)
(emerging,1)
(digital,1)
(His,1)
(With,1)
(harmonies.,1)
(algorithms,1)
(bug,1)
```