

Week 4

Felipe Rodriguez

Bellevue University

DSC 650 Big Data

Professor Nasheb Ismaily

December 24, 2023

SparkPi Output

```
55183 [main] INFO org.apache.spark.scheduler.DAGScheduler - Job 0 finished: reduce at SparkPi.scala:38, took 12.963425 s
P1 is roughly 3.139143139143139
55232 [main] INFO org.sparkproject.jetty.server.AbstractConnector - Stopped Spark@46b695ec(HTTP/1.1,[http/1.1]){172.28.1.1:4040}
55236 [main] INFO org.apache.spark.ui.SparkUI - Stopped Spark web UI at http://localhost:4040
55253 [YARN application state monitor] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend - Interrupting monitor thread
55311 [main] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend - Shutting down all executors
55313 [dispatcher-coarseGrainedScheduler] INFO org.apache.spark.scheduler.cluster.YarnSchedulerBackend$YarnDriverEndpoint - Asking each executor to shut down
55325 [main] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend - YARN client scheduler backend Stopped
55781 [dispatcher-event-loop-3] INFO org.apache.spark.MapOutputTrackerMasterEndpoint - MapOutputTrackerMasterEndpoint stopped!
55798 [main] INFO org.apache.spark.storage.memory.MemoryStore - MemoryStore cleared
55800 [main] INFO org.apache.spark.storage.BlockManager - BlockManager stopped
55821 [main] INFO org.apache.spark.storage.BlockManagerMaster - BlockManagerMaster stopped
55825 [dispatcher-event-loop-3] INFO org.apache.spark.scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint - OutputCommitCoordinator stopped!
55891 [main] INFO org.apache.spark.SparkContext - Successfully stopped SparkContext
55919 [shutdown-hook-0] INFO org.apache.spark.util.ShutdownHookManager - Shutdown hook called
55925 [shutdown-hook-0] INFO org.apache.spark.util.ShutdownHookManager - Deleting directory /tmp/spark-1504a2bd-a3ea-4770-a063-5e4144a52258
55937 [shutdown-hook-0] INFO org.apache.spark.util.ShutdownHookManager - Deleting directory /tmp/spark-f7dc6085-70a2-4e42-9425-4e592c629eb5
bash-5.0# █
```

First 100 Numbers Generated

```
[scala> numbersRDD.take(100).foreach(println)
278
272
294
144
531
369
964
137
56
805
634
374
123
843
589
334
780
160
345
180
121
588
476
752
308
33
733
152
895
184
59
846
216
922
209
703
167
904
919
208
804
799
524
284
589
363
839
694
389
527
933
536
729
79
765
302
932
298
104
970
506
775
812
309
865
```

```
466
415
766
195
595
472
722
688
380
425
752
465
5
280
511
583
902
157
151
986
72
23
1
821
440
575
580
895
473
110
93
492
133
282
602
```

```
scala> █
```

Transformation – substring()

```
[scala> val transformedSentences = sentencesRDD.map(sentence => sentence.substring(0,4))
transformedSentences: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at map at <console>:25

[scala> sentences.take(25).foreach(println)
cherry.
date cherry.
date fig cherry apple.
apple grape.
cherry honeydew elderberry grape banana fig.
banana fig elderberry date apple cherry.
fig date banana.
grape apple banana.
grape apple date cherry fig.
honeydew.
honeydew cherry banana.
grape apple fig banana honeydew cherry.
grape.
honeydew grape apple fig date elderberry.
grape date honeydew cherry banana apple.
elderberry date cherry grape banana fig.
fig apple.
elderberry banana fig grape date cherry.
apple cherry honeydew fig date banana.
elderberry.
banana date grape honeydew.
grape cherry honeydew banana.
cherry date banana elderberry honeydew grape.
honeydew fig.
grape cherry elderberry.

[scala> transformedSentences.take(25).foreach(println)
cher
date
date
appl
cher
bana
fig
grap
grap
hone
hone
grap
grap
hone
grap
elde
fig
elde
appl
elde
bana
grap
cher
hone
grap

scala> █
```

The transformation chosen is the substring function. This transformation allows to select a set number of characters specified in the function. Since this is set to 0,4, the first four characters of the sentence are selected. The first part of the screen shot contains the command followed by the original output. The last line is the transformed data. As you can see, the data is now limited to the first four positions.