

```

hadoop-yarn@ip-10-0-0-96:/opt/hadoop$ hls=$(cat /dev/urandom | fold -n 100 | tr -dc 'a-z0-9' | fold -w 64 | xargs sha512sum)
hadoop-yarn@ip-10-0-0-96:/opt/hadoop$ sudo docker-compose exec master bash
bash-5.0# spark-submit --class org.apache.spark.examples.SparkPi \
--master yarn \
--deploy-mode client \
--driver-memory 2g \
--executor-memory 1g \
--executor-cores 1 \
$SPARK_HOME/examples/jars/spark-examples*.jar \
10
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.log4j.LoggerFactory]
[main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1484 [main] INFO org.apache.spark.SparkContext - Running Spark version 3.0.0
1486 [main] INFO org.apache.spark.ResourceUtils - =====
1478 [main] INFO org.apache.spark.ResourceUtils - Resources for Spark driver:
1472 [main] INFO org.apache.spark.ResourceUtils -
1474 [main] INFO org.apache.spark.SparkContext - Submitted application: Spark Pi
1784 [main] INFO org.apache.spark.SecurityManager - Changing view acls to: root
1797 [main] INFO org.apache.spark.SecurityManager - Changing modify acls to: root
1790 [main] INFO org.apache.spark.SecurityManager - Changing view acls groups to:
1795 [main] INFO org.apache.spark.SecurityManager - Changing modify acls groups to:
1799 [main] INFO org.apache.spark.SecurityManager - SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); gr
groups with view permissions: Set(); users with modify permissions: Set(root); groups with modify permissions: Set()
1868 [main] INFO org.apache.spark.SparkEnv - Successfully started service "SparkDriver" on port 7001.
1898 [main] INFO org.apache.spark.SparkEnv - Registering MapOutputTracker
1354 [main] INFO org.apache.spark.storage - Registering BlockManagerMaster
1444 [main] INFO org.apache.spark.storage.BlockManagerMasterEndpoint - Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
1446 [main] INFO org.apache.spark.storage.BlockManagerMasterEndpoint - BlockManagerMasterEndpoint up
1446 [main] INFO org.apache.spark.SparkEnv - Registering BlockManagerMasterHeartbeat
1446 [main] INFO org.apache.spark.storage.DiskBlockManager - Created local directory at /tmp/blockmgr-d3ba625-749f-443b-999b-546d0d2420f1
1815 [main] INFO org.apache.spark.storage.memory.MemoryStore - MemoryStore started with capacity 912.3 MiB
4007 [main] INFO org.apache.spark.SparkEnv - Registering OutputCommitCoordinator
3762 [main] INFO org.sparkproject.jetty.server.Server - Logging initialized @1097ms to org.sparkproject.jetty.servlet.LogFileLog
6702 [main] INFO org.sparkproject.jetty.server.Server - jetty.v4.2-SNAPSHOT; built: 2019-04-29T20:42:08.995Z; git: elbc3512da6be17e3405d2294a43f3a2ce5079
7; jvm 1.8.0_275-b01
7782 [main] INFO org.sparkproject.jetty.server.Server - Started @10379ms
6883 [main] INFO org.sparkproject.jetty.server.AbstractConnector - Started ServerConnector@846c996{HTTP/1.1,[http/1.1]}{172.28.1.1:14040}
6883 [main] INFO org.apache.spark.ui.Utils - Successfully started service "SparkUI" on port 4040.
```

```

11936 [main] INFO org.apache.spark.deploy.yarn.client - Uploading resource file:/tmp/spark-44fbfbbd-7007-4347-43bf-52d3c7c0a70e/_spark_conf_165043686011
11937 [main] INFO dfps://master:9000/user/root/_spark$tagging/application/1704515461368_0001/_spark_conf_*.zip
11938 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - No unit for dfs.client.datanode-restart.timeout(30) assuming SECONDS
11939 [main] INFO org.apache.spark.security.SecurityManager - Changing view acls to: root
11940 [main] INFO org.apache.spark.SecurityManager - Changing modify acls to: root
11941 [main] INFO org.apache.spark.SecurityManager - Changing view acls groups to:
11942 [main] INFO org.apache.spark.SecurityManager - Changing modify acls groups to:
11943 [main] INFO org.apache.spark.SecurityManager - SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root);
11944 [main] INFO org.apache.spark.SecurityManager - SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); users with modify permissions: Set()
11945 [main] INFO org.apache.spark.deploy.yarn.client - Submitting application application_1704515461368_0001 to ResourceManager
11946 [main] INFO org.apache.spark.deploy.yarn.client - Submitted application application_1704515461368_0001
11947 [main] INFO org.apache.spark.deploy.yarn.client - Application report for application_1704515461368_0001 (state: ACCEPTED)
11948 [main] INFO org.apache.spark.deploy.yarn.client - client token: N/A
11949 [main] INFO org.apache.spark.deploy.yarn.client - diagnostics: N/A
11950 [main] INFO org.apache.spark.deploy.yarn.client - ApplicationMaster host: N/A
11951 [main] INFO org.apache.spark.deploy.yarn.client - ApplicationMaster RPC port: -1
11952 [main] INFO org.apache.spark.deploy.yarn.client - queue: root:root
11953 [main] INFO org.apache.spark.deploy.yarn.client - start time: 170451544823
11954 [main] INFO org.apache.spark.deploy.yarn.client - final status: UNKOWNED
11955 [main] INFO org.apache.spark.deploy.yarn.client - tracking URL: http://master:8088/proxy/application_1704515461368_0001/
11956 [main] INFO org.apache.spark.deploy.yarn.client - user: root
11957 [main] INFO org.apache.spark.deploy.yarn.client - Application report for application_1704515461368_0001 (state: ACCEPTED)

```

```

97429 [task-result-getter-3] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 3.0 in stage 0.0 (TID 3) in 110 ms on worker2 (executor 2) (4/6)
97537 [dispatcher-CoarseGrainedScheduler] INFO org.apache.spark.scheduler.TaskSetManager - Starting task 5.0 in stage 0.0 (TID 5, worker2, executor 2, partition 5, PROCESS_LOCAL, 7404 bytes)
97538 [task-result-getter-0] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 4.0 in stage 0.0 (TID 4) in 120 ms on worker2 (executor 2) (5/6)
97600 [dispatcher-CoarseGrainedScheduler] INFO org.apache.spark.scheduler.TaskSetManager - Starting task 6.0 in stage 0.0 (TID 6, worker2, executor 2, partition 6, PROCESS_LOCAL, 7404 bytes)
97664 [task-result-getter-1] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 5.0 in stage 0.0 (TID 5) in 130 ms on worker2 (executor 2) (6/6)
97759 [dispatcher-CoarseGrainedScheduler] INFO org.apache.spark.scheduler.TaskSetManager - Starting task 7.0 in stage 0.0 (TID 7, worker2, executor 2, partition 7, PROCESS_LOCAL, 7404 bytes)
97761 [task-result-getter-2] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 6.0 in stage 0.0 (TID 6) in 101 ms on worker2 (executor 2) (7/6)
97908 [dispatcher-CoarseGrainedScheduler] INFO org.apache.spark.scheduler.TaskSetManager - Starting task 8.0 in stage 0.0 (TID 8, worker2, executor 2, partition 8, PROCESS_LOCAL, 7404 bytes)
97913 [task-result-getter-3] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 7.0 in stage 0.0 (TID 7) in 155 ms on worker2 (executor 2) (8/6)
97967 [dispatcher-CoarseGrainedScheduler] INFO org.apache.spark.scheduler.TaskSetManager - Starting task 9.0 in stage 0.0 (TID 9, worker2, executor 2, partition 9, PROCESS_LOCAL, 7404 bytes)
97969 [task-result-getter-0] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 8.0 in stage 0.0 (TID 8) in 61 ms on worker2 (executor 2) (9/10)
98065 [task-result-getter-1] INFO org.apache.spark.scheduler.TaskSetManager - Finished task 9.0 in stage 0.0 (TID 9) in 103 ms on worker2 (executor 2) (10/10)
98070 [task-result-getter-1] INFO org.apache.spark.scheduler.cluster.YarnScheduler - Removed TaskSet 0.0, whose tasks have all completed, from pool
98074 [dag-scheduler-event-loop] INFO org.apache.spark.scheduler.DAGScheduler - ResultStage 0 (reduce at SparkPi.scala:38) finished in 14.984 s
98122 [dag-scheduler-event-loop] INFO org.apache.spark.scheduler.DAGScheduler - Job 0 finished. Cancelling potential speculative or zombie tasks for this job
98124 [dag-scheduler-event-loop] INFO org.apache.spark.scheduler.cluster.YarnScheduler - Killing all running tasks in stage 0: Stage finished
98152 [main] INFO org.apache.spark.scheduler.dag.DAGScheduler - Job 0 finished: reduce at SparkPi.scala:38, took 15.236094 s
98154 PI is roughly 3.141191141191441
98199 [main] INFO org.sparkproject.jetty.server.AbstractConnector - Stopped Spark44b6d59cse[HTTP/1.1]{http://172.28.1.14040}
98201 [main] INFO org.apache.spark.util.SparkUI - stopped Spark web UI at http://localhost:14040
98238 [YARN application state monitor] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend - Interrupting monitor thread
98327 [main] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend - Shutting down all executors
98329 [dispatcher-CoarseGrainedScheduler] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend.YarnDriverEndpoint - Asking each executor to shut down
98349 [main] INFO org.apache.spark.scheduler.cluster.YarnClientSchedulerBackend - YARN client scheduler backend stopped
98373 [dispatcher-event-loop-3] INFO org.apache.spark.MapOutputTrackerMasterEndpoint - MapOutputTrackerMasterEndpoint stopped!
98395 [main] INFO org.apache.spark.storage.MemoryStore - MemoryStore cleared
98911 [main] INFO org.apache.spark.storage.BlockManager - BlockManager stopped
98929 [main] INFO org.apache.spark.storage.BlockManagerMaster - BlockManagerMaster stopped
98933 [dispatcher-event-loop-3] INFO org.apache.spark.scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint - OutputCommitCoordinator stopped!
98934 [main] INFO org.apache.spark.scheduler.SuccessfullyStoppedSparkContext - Successfully stopped SparkContext
99010 [shutdown-hook-0] INFO org.apache.spark.util.ShutdownHookManager - Shutdown hook called
99012 [shutdown-hook-0] INFO org.apache.spark.util.ShutdownHookManager - Deleting directory /tmp/spark-44bfb7bd-6707-43ab-ab3f-52d2cf70a7de
99032 [shutdown-hook-0] INFO org.apache.spark.util.ShutdownHookManager - Deleting directory /tmp/spark-5a6a1af1-5f70-4755-8327-ea3b233bae2f

```

```
Using Scala version 2.12.10 (OpenJDK 64-Bit Server VM, Java 1.8.0_275)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val numNumbers = 10000
numNumbers: Int = 10000

scala> val numbers = (1 to numNumbers).map(_ => scala.util.Random.nextInt(1000))
numbers: scala.collection.immutable.IndexedSeq[Int] = Vector(299, 963, 546, 212, 805, 592, 951, 575, 773, 703, 212, 378, 805, 282, 702, 229, 34, 920, 975, 13
3, 160, 240, 201, 962, 449, 330, 245, 811, 332, 634, 682, 126, 498, 273, 1, 848, 396, 914, 516, 173, 977, 166, 791, 789, 764, 701, 934, 243, 297, 642, 870, 1
79, 672, 937, 881, 384, 530, 145, 417, 583, 993, 406, 909, 259, 28, 229, 945, 478, 159, 423, 447, 944, 425, 583, 940, 727, 463, 288, 759, 492, 238, 682, 521,
846, 574, 911, 577, 408, 177, 312, 73, 892, 363, 172, 106, 772, 876, 54, 87, 786, 288, 610, 765, 708, 160, 140, 672, 757, 706, 806, 809, 626, 603, 716, 620,
640, 700, 420, 333, 588, 42, 743, 258, 551, 535, 671, 148, 260, 671, 125, 991, 464, 397, 416, 601, 892, 723, 987, 927, 521, 968, 251, 891, 58,...

scala> val numbersRDD = sc.parallelize(numbers)
numbersRDD: org.apache.spark.rdd.RDD[Int] = ParallelCollectionRDD[0] at parallelize at <console>:26

scala> numbersRDD.take(100).foreach(println)
299
963
546
212
805
592
951
575
773
703
212
378
805
282
702
229
34
920
975
133
160
240
201
962
449
330
245
811
332
634
```

```

937
881
384
530
143
417
583
993
406
909
259
28
228
945
478
159
423
447
944
425
583
940
727
463
288
758
492
238
682
521
846
574
911
577
405
177
312
73
692
363
172
106
772
876
54
87
786
scala>

```

Screenshot of a segment of the generated transformed sentences and an explanation of your unique transformation :

```

scala> val numberOfSentences = 1000
numberOfSentences: Int = 1000

scala> val words = List("apple", "banana", "cherry", "date", "elderberry", "fig", "grape", "honeydew")
words: List[String] = List(apple, banana, cherry, date, elderberry, fig, grape, honeydew)

scala> val sentences = (1 to numberOfSentences).map(_ => scala.util.Random.shuffle(words).take(scala.util.Random.nextInt(6) + 1).mkString(" ") + ".")
sentences: scala.collection.immutable.IndexedSeq[String] = Vector(banana elderberry cherry., apple grape cherry date fig., fig date honeydew banana., honeydew fig cherry., fig honeydew grape apple date cherry., date cherry honeydew elderberry., grape fig., grape apple banana honeydew fig., apple cherry fig banana., cherry date grape honeydew., date honeydew banana fig., date fig., apple date., grape., honeydew banana apple date., honeydew apple cherry., banana fig cherry elderberry date., apple date cherry grape., fig elderberry., cherry elderberry date., fig cherry honeydew elderberry., banana cherry., cherry., grape apple., elderberry apple date., cherry banana elderberry., honeydew cherry apple fig., apple fig., grape., honeydew grape banana apple., honeydew...)

scala> val sentencesRDD = sc.parallelize(sentences)
sentencesRDD: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[1] at parallelize at <console>:26

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

scala> transformedSentences.take(100).foreach(println)
BANANA ELDERBERRY CHERRY.
APPLE GRAPE CHERRY DATE FIG.
FIG DATE HONEYDEW BANANA.
HONEYDEW FIG CHERRY.
FIG HONEYDEW GRAPE APPLE DATE CHERRY.
DATE CHERRY HONEYDEW ELDERBERRY.
GRAPE FIG.
GRAPE APPLE BANANA HONEYDEW FIG.
APPLE CHERRY FIG BANANA.
CHERRY DATE GRAPE HONEYDEW.
DATE HONEYDEW BANANA FIG.
DATE FIG.
APPLE DATE.
GRAPE.
HONEYDEW BANANA APPLE DATE.
HONEYDEW APPLE CHERRY.
BANANA FIG CHERRY ELDERBERRY DATE.
APPLE DATE CHERRY GRAPE.
FIG ELDERBERRY.
CHERRY ELDERBERRY DATE.
FIG CHERRY HONEYDEW ELDERBERRY.
BANANA CHERRY.
CHERRY.
GRAPE APPLE.
ELDERBERRY APPLE DATE.
CHERRY BANANA ELDERBERRY.
HONEYDEW CHERRY APPLE FIG.
APPLE FIG.

```

```

DATE.
DATE HONEYDEW.
FIG HONEYDEW.
HONEYDEW.
GRAPE ELDERBERRY CHERRY BANANA.
FIG HONEYDEW.
DATE FIG.
CHERRY BANANA DATE GRAPE APPLE HONEYDEW.
APPLE FIG BANANA GRAPE CHERRY DATE.
DATE.
GRAPE CHERRY DATE APPLE HONEYDEW.
CHERRY ELDERBERRY.
FIG ELDERBERRY BANANA GRAPE DATE CHERRY.
CHERRY ELDERBERRY BANANA HONEYDEW.
FIG.
ELDERBERRY FIG APPLE DATE GRAPE.
GRAPE ELDERBERRY.
APPLE ELDERBERRY GRAPE.
BANANA.
APPLE BANANA FIG DATE HONEYDEW.
HONEYDEW BANANA ELDERBERRY FIG DATE GRAPE.
DATE GRAPE CHERRY.
DATE HONEYDEW BANANA CHERRY APPLE ELDERBERRY.
APPLE FIG CHERRY.
GRAPE DATE BANANA.
FIG CHERRY.
CHERRY HONEYDEW APPLE GRAPE DATE BANANA.
FIG HONEYDEW APPLE GRAPE BANANA.
DATE BANANA HONEYDEW.
HONEYDEW DATE.
CHERRY HONEYDEW BANANA.
BANANA FIG GRAPE CHERRY DATE HONEYDEW.
GRAPE.
DATE CHERRY GRAPE FIG.
GRAPE HONEYDEW FIG DATE ELDERBERRY BANANA.
HONEYDEW.
CHERRY.
ELDERBERRY FIG APPLE GRAPE CHERRY.
DATE HONEYDEW.
BANANA ELDERBERRY CHERRY DATE FIG APPLE.
HONEYDEW ELDERBERRY FIG DATE APPLE.
APPLE BANANA HONEYDEW CHERRY ELDERBERRY.
HONEYDEW BANANA GRAPE.
DATE ELDERBERRY HONEYDEW GRAPE FIG.
ELDERBERRY FIG CHERRY HONEYDEW GRAPE APPLE.
FIG APPLE.
CHERRY HONEYDEW GRAPE.
scala> []

```

Used toUpperCase transformation code to convert the all the sentences with Uppercase letters. The program uses numberOfSentences which basically specifies the total number of sentences to be generated. Words defines a list of words that can be used to build the sentences. Then it defines how the sentences will be formed using random words from the list and sentences will be generated from one to numberOfSentences. sc.parallelize(sentences): Converts the list of sentences into an RDD. sc is assumed to be the SparkContext.