

File Edit View Search Terminal Help

```
prerna@prerna-Inspiron-5559:~$ cd spark-2.0.2-bin-hadoop2.6/
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.ProductSales ../jars
jars/      jars_file/
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.ProductSales ../jars_file/retail_productsca
ala.jar
bin/      licenses/      ret-out-01/      sbin/
conf/     NOTICE        ret-out-02/      spark-warehouse/
data/     python/        ret-out-03/      wc-out-01/
examples/ R/             ret-out-prod-01/ wc-scala-out-001/
jars/     README.md      ret-out-store-01/ yarn/
LICENSE   RELEASE        ret-out-total-01/

prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.ProductSales ../jars_file/retail_productsca
ala.jar ../data/Retail_Sample_Data_Set.txt ret-out-prodSalesScala-01
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
20/06/28 22:12:09 INFO SparkContext: Running Spark version 2.0.2
20/06/28 22:12:10 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/06/28 22:12:10 WARN Utils: Your hostname, prerna-Inspiron-5559 resolves to a loopback address: 127.0.1.1; using 192.168.1.10 instead (on in
terface wlp1s0)
20/06/28 22:12:10 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
20/06/28 22:12:10 INFO SecurityManager: Changing view acls to: prerna
20/06/28 22:12:10 INFO SecurityManager: Changing modify acls to: prerna
20/06/28 22:12:10 INFO SecurityManager: Changing view acls groups to:
20/06/28 22:12:10 INFO SecurityManager: Changing modify acls groups to:
20/06/28 22:12:10 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(prerna);
groups with view permissions: Set(); users with modify permissions: Set(prerna); groups with modify permissions: Set()
20/06/28 22:12:10 INFO Utils: Successfully started service 'sparkDriver' on port 38231.
20/06/28 22:12:10 INFO SparkEnv: Registering MapOutputTracker
20/06/28 22:12:10 INFO SparkEnv: Registering BlockManagerMaster
20/06/28 22:12:10 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-5c6b137d-a057-4994-85bc-416d70cac008
20/06/28 22:12:10 INFO MemoryStore: MemoryStore started with capacity 366.3 MB
20/06/28 22:12:10 INFO SparkEnv: Registering OutputCommitCoordinator
20/06/28 22:12:11 INFO Utils: Successfully started service 'SparkUI' on port 4040.
20/06/28 22:12:11 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.1.10:4040
20/06/28 22:12:11 INFO SparkContext: Added JAR file:/home/prerna/spark-2.0.2-bin-hadoop2.6/../jars_file/retail_productscala.jar at spark://192
.168.1.10:38231/jars/retail_productscala.jar with timestamp 1593362531137
20/06/28 22:12:11 INFO Executor: Starting executor ID driver on host localhost
20/06/28 22:12:11 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 36867.
20/06/28 22:12:11 INFO NettyBlockTransferService: Server created on 192.168.1.10:36867
```

File Edit View Search Terminal Help

```
20/06/28 22:12:11 WARN SparkContext: Use an existing SparkContext, some configuration may not take effect.
20/06/28 22:12:11 INFO SharedState: Warehouse path is 'file:/home/prerna/spark-2.0.2-bin-hadoop2.6/spark-warehouse'.
20/06/28 22:12:14 INFO FileSourceStrategy: Pruning directories with:
20/06/28 22:12:14 INFO FileSourceStrategy: Post-Scan Filters:
20/06/28 22:12:14 INFO FileSourceStrategy: Pruned Data Schema: struct<value: string>
20/06/28 22:12:14 INFO FileSourceStrategy: Pushed Filters:
20/06/28 22:12:14 INFO MemoryStore: Block broadcast_0 stored as values in memory (estimated size 214.3 KB, free 366.1 MB)
20/06/28 22:12:14 INFO MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 20.5 KB, free 366.1 MB)
20/06/28 22:12:14 INFO BlockManagerInfo: Added broadcast_0_piece0 in memory on 192.168.1.10:36867 (size: 20.5 KB, free: 366.3 MB)
20/06/28 22:12:14 INFO SparkContext: Created broadcast 0 from rdd at ProductSales.scala:20
20/06/28 22:12:14 INFO FileSourceStrategy: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
20/06/28 22:12:14 INFO CodeGenerator: Code generated in 233.935013 ms
20/06/28 22:12:15 INFO deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id
20/06/28 22:12:15 INFO deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id
20/06/28 22:12:15 INFO deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap
20/06/28 22:12:15 INFO deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition
20/06/28 22:12:15 INFO deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id
20/06/28 22:12:15 INFO SparkContext: Starting job: saveAsTextFile at ProductSales.scala:28
20/06/28 22:12:15 INFO DAGScheduler: Registering RDD 4 (map at ProductSales.scala:22)
20/06/28 22:12:15 INFO DAGScheduler: Got job 0 (saveAsTextFile at ProductSales.scala:28) with 1 output partitions
20/06/28 22:12:15 INFO DAGScheduler: Final stage: ResultStage 1 (saveAsTextFile at ProductSales.scala:28)
20/06/28 22:12:15 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 0)
20/06/28 22:12:15 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 0)
20/06/28 22:12:15 INFO DAGScheduler: Submitting ShuffleMapStage 0 (MapPartitionsRDD[4] at map at ProductSales.scala:22), which has no missing parents
20/06/28 22:12:15 INFO MemoryStore: Block broadcast_1 stored as values in memory (estimated size 9.9 KB, free 366.1 MB)
20/06/28 22:12:15 INFO MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 5.3 KB, free 366.1 MB)
20/06/28 22:12:15 INFO BlockManagerInfo: Added broadcast_1_piece0 in memory on 192.168.1.10:36867 (size: 5.3 KB, free: 366.3 MB)
20/06/28 22:12:15 INFO SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1012
20/06/28 22:12:15 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 0 (MapPartitionsRDD[4] at map at ProductSales.scala:22)
20/06/28 22:12:15 INFO TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
20/06/28 22:12:15 INFO TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, partition 0, PROCESS_LOCAL, 5983 bytes)
20/06/28 22:12:15 INFO Executor: Running task 0.0 in stage 0.0 (TID 0)
20/06/28 22:12:15 INFO Executor: Fetching spark://192.168.1.10:38231/jars/retail_productscala.jar with timestamp 1593362531137
20/06/28 22:12:15 INFO TransportClientFactory: Successfully created connection to /192.168.1.10:38231 after 48 ms (0 ms spent in bootstraps)
20/06/28 22:12:15 INFO Utils: Fetching spark://192.168.1.10:38231/jars/retail_productscala.jar to /tmp/spark-7e68a22-de2e-41a9-9004-3dfff5dbf713b/userFiles-1ead47aa-a26c-45d3-a2db-855ecd73e75c/fetchFileTemp4650897513303964301.tmp
```


Activities Terminal ▾

Sun 22:21 ●

📶 🔊 🔌

prerna@prerna-Inspiron-5559: ~/spark-2.0.2-bin-hadoop2.6

File Edit View Search Terminal Help

```
20/06/28 22:12:15 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
20/06/28 22:12:15 INFO DAGScheduler: ShuffleMapStage 0 (map at ProductSales.scala:22) finished in 0.640 s
20/06/28 22:12:15 INFO DAGScheduler: looking for newly runnable stages
20/06/28 22:12:15 INFO DAGScheduler: running: Set()
20/06/28 22:12:15 INFO DAGScheduler: waiting: Set(ResultStage 1)
20/06/28 22:12:15 INFO DAGScheduler: failed: Set()
20/06/28 22:12:15 INFO DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[6] at saveAsTextFile at ProductSales.scala:28), which has no missing parents
20/06/28 22:12:16 INFO MemoryStore: Block broadcast_2 stored as values in memory (estimated size 65.8 KB, free 366.0 MB)
20/06/28 22:12:16 INFO MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 23.7 KB, free 366.0 MB)
20/06/28 22:12:16 INFO BlockManagerInfo: Added broadcast_2_piece0 in memory on 192.168.1.10:36867 (size: 23.7 KB, free: 366.3 MB)
20/06/28 22:12:16 INFO SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1012
20/06/28 22:12:16 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[6] at saveAsTextFile at ProductSales.scala:28)
20/06/28 22:12:16 INFO TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
20/06/28 22:12:16 INFO TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, partition 0, ANY, 5241 bytes)
20/06/28 22:12:16 INFO Executor: Running task 0.0 in stage 1.0 (TID 1)
20/06/28 22:12:16 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
20/06/28 22:12:16 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 7 ms
20/06/28 22:12:16 INFO FileOutputCommitter: Saved output of task 'attempt_202006282212_0001_m_000000_1' to file:/home/prerna/spark-2.0.2-bin-hadoop2.6/ret-out-prodSalesScala-01/_temporary/0/task_202006282212_0001_m_000000
20/06/28 22:12:16 INFO SparkHadoopMapRedUtil: attempt_202006282212_0001_m_000000_1: Committed
20/06/28 22:12:16 INFO Executor: Finished task 0.0 in stage 1.0 (TID 1). 1721 bytes result sent to driver
20/06/28 22:12:16 INFO TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 134 ms on localhost (1/1)
20/06/28 22:12:16 INFO TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
20/06/28 22:12:16 INFO DAGScheduler: ResultStage 1 (saveAsTextFile at ProductSales.scala:28) finished in 0.134 s
20/06/28 22:12:16 INFO DAGScheduler: Job 0 finished: saveAsTextFile at ProductSales.scala:28, took 1.037269 s
20/06/28 22:12:16 INFO SparkUI: Stopped Spark web UI at http://192.168.1.10:4040
20/06/28 22:12:16 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/06/28 22:12:16 INFO MemoryStore: MemoryStore cleared
20/06/28 22:12:16 INFO BlockManager: BlockManager stopped
20/06/28 22:12:16 INFO BlockManagerMaster: BlockManagerMaster stopped
20/06/28 22:12:16 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/06/28 22:12:16 INFO SparkContext: Successfully stopped SparkContext
20/06/28 22:12:16 INFO ShutdownHookManager: Shutdown hook called
20/06/28 22:12:16 INFO ShutdownHookManager: Deleting directory /tmp/spark-7e868a22-de2e-41a9-9004-3dff5dbf713b
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ ls ret-out-prodSalesScala-01/
part-000000 _SUCCESS
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ cat ret-out-prodSalesScala-01/part-000000
```

```
Sun 22:21
prema@prerna-Inspiron-5559: ~/spark-2.0.2-bin-hadoop2.6

File Edit View Search Terminal Help

20/06/28 22:12:16 INFO SparkHadoopMapRedUtil: attempt_202006282212_0001_m_000000_1: Committed
20/06/28 22:12:16 INFO Executor: Finished task 0.0 in stage 1.0 (TID 1). 1721 bytes result sent to driver
20/06/28 22:12:16 INFO TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 134 ms on localhost (1/1)
20/06/28 22:12:16 INFO TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
20/06/28 22:12:16 INFO DAGScheduler: ResultStage 1 (saveAsTextFile at ProductSales.scala:28) finished in 0.134 s
20/06/28 22:12:16 INFO DAGScheduler: Job 0 finished: saveAsTextFile at ProductSales.scala:28, took 1.037269 s
20/06/28 22:12:16 INFO SparkUI: Stopped Spark web UI at http://192.168.1.10:4040
20/06/28 22:12:16 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/06/28 22:12:16 INFO MemoryStore: MemoryStore cleared
20/06/28 22:12:16 INFO BlockManager: BlockManager stopped
20/06/28 22:12:16 INFO BlockManagerMaster: BlockManagerMaster stopped
20/06/28 22:12:16 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/06/28 22:12:16 INFO SparkContext: Successfully stopped SparkContext
20/06/28 22:12:16 INFO ShutdownHookManager: Shutdown hook called
20/06/28 22:12:16 INFO ShutdownHookManager: Deleting directory /tmp/spark-7e868a22-de2e-41a9-9004-3dff5dbf713b
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ ls ret-out-prodSalesScala-01/
part-000000 SUCCESS
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ cat ret-out-prodSalesScala-01/part-000000
(Men's Clothing,4030.89)
(Video Games,2573.38)
(Computers,2102.66)
(Women's Clothing,3736.8699)
(Sporting Goods,1952.89)
(Music,2396.4)
(CDs,2644.51)
(Garden,1882.25)
(Pet Supplies,2660.83)
(Cameras,2591.27)
(Books,3492.7998)
(Health and Beauty,2467.3198)
(Baby,2034.2297)
(Children's Clothing,2778.2102)
(Toys,3188.18)
(Consumer Electronics,2963.59)
(Crafts,3258.09)
(DVDs,2831.0002)
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$
```



```
Sun 22:26 ●
prema@prerna-Inspiron-5559: ~/spark-2.0.2-bin-hadoop2.6

File Edit View Search Terminal Help

prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.StoreSales ../jars_file/retail_../data/Retail_Sample_Data_Set.txt ret-out-prodSalesScala-01 retail_productscala.jar retail_storescala.jar
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.StoreSales ../jars_file/retail_storescala.jar ../data/Retail_Sample_Data_Set.txt ret-out-StoreSalesScala-01
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
20/06/28 22:25:53 INFO SparkContext: Running Spark version 2.0.2
20/06/28 22:25:54 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/06/28 22:25:54 WARN Utils: Your hostname, prerna-Inspiron-5559 resolves to a loopback address: 127.0.1.1; using 192.168.1.10 instead (on interface wlp1s0)
20/06/28 22:25:54 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
20/06/28 22:25:54 INFO SecurityManager: Changing view acls to: prerna
20/06/28 22:25:54 INFO SecurityManager: Changing modify acls to: prerna
20/06/28 22:25:54 INFO SecurityManager: Changing view acls groups to:
20/06/28 22:25:54 INFO SecurityManager: Changing modify acls groups to:
20/06/28 22:25:54 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(prerna); groups with view permissions: Set(); users with modify permissions: Set(prerna); groups with modify permissions: Set()
20/06/28 22:25:54 INFO Utils: Successfully started service 'sparkDriver' on port 34181.
20/06/28 22:25:54 INFO SparkEnv: Registering MapOutputTracker
20/06/28 22:25:54 INFO SparkEnv: Registering BlockManagerMaster
20/06/28 22:25:54 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-22cd6fa7-daeb-4dc0-8d2e-0710da4a84f1
20/06/28 22:25:54 INFO MemoryStore: MemoryStore started with capacity 366.3 MB
20/06/28 22:25:54 INFO SparkEnv: Registering OutputCommitCoordinator
20/06/28 22:25:55 INFO Utils: Successfully started service 'SparkUI' on port 4040.
20/06/28 22:25:55 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.1.10:4040
20/06/28 22:25:55 INFO SparkContext: Added JAR file:/home/prerna/spark-2.0.2-bin-hadoop2.6/./jars_file/retail_storescala.jar at spark://192.168.1.10:34181/jars/retail_storescala.jar with timestamp 1593363355075
20/06/28 22:25:55 INFO Executor: Starting executor ID driver on host localhost
20/06/28 22:25:55 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 34491.
20/06/28 22:25:55 INFO NettyBlockTransferService: Server created on 192.168.1.10:34491
20/06/28 22:25:55 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.1.10, 34491)
20/06/28 22:25:55 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.1.10:34491 with 366.3 MB RAM, BlockManagerId(driver, 192.168.1.10, 34491)
20/06/28 22:25:55 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.1.10, 34491)
20/06/28 22:25:55 WARN SparkContext: Use an existing SparkContext, some configuration may not take effect.
20/06/28 22:25:55 INFO SharedState: Warehouse path is 'file:/home/prerna/spark-2.0.2-bin-hadoop2.6/spark-warehouse'.
20/06/28 22:25:57 INFO FileSourceStrategy: Pruning directories with:
20/06/28 22:25:57 INFO FileSourceStrategy: Post-Scan Filters:
```

File Edit View Search Terminal Help

```
20/06/28 22:25:59 INFO DAGScheduler: running: Set()
20/06/28 22:25:59 INFO DAGScheduler: waiting: Set(ResultStage 1)
20/06/28 22:25:59 INFO DAGScheduler: failed: Set()
20/06/28 22:25:59 INFO DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[6] at saveAsTextFile at StoreSales.scala:29), which has no missing parents
20/06/28 22:25:59 INFO MemoryStore: Block broadcast_2 stored as values in memory (estimated size 65.8 KB, free 366.0 MB)
20/06/28 22:25:59 INFO MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 23.7 KB, free 366.0 MB)
20/06/28 22:25:59 INFO BlockManagerInfo: Added broadcast_2_piece0 in memory on 192.168.1.10:34491 (size: 23.7 KB, free: 366.3 MB)
20/06/28 22:25:59 INFO SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1012
20/06/28 22:25:59 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[6] at saveAsTextFile at StoreSales.scala:29)
20/06/28 22:25:59 INFO TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
20/06/28 22:25:59 INFO TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, partition 0, ANY, 5239 bytes)
20/06/28 22:25:59 INFO Executor: Running task 0.0 in stage 1.0 (TID 1)
20/06/28 22:25:59 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
20/06/28 22:25:59 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 8 ms
20/06/28 22:25:59 INFO FileOutputCommitter: Saved output of task 'attempt_202006282225_0001_m_000000_1' to file:/home/prerna/spark-2.0.2-bin-hadoop2.6/ret-out-StoreSalesScala-01/_temporary/0/task_202006282225_0001_m_000000
20/06/28 22:25:59 INFO SparkHadoopMapRedUtil: attempt_202006282225_0001_m_000000_1: Committed
20/06/28 22:26:00 INFO Executor: Finished task 0.0 in stage 1.0 (TID 1). 1721 bytes result sent to driver
20/06/28 22:26:00 INFO TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 152 ms on localhost (1/1)
20/06/28 22:26:00 INFO TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
20/06/28 22:26:00 INFO DAGScheduler: ResultStage 1 (saveAsTextFile at StoreSales.scala:29) finished in 0.153 s
20/06/28 22:26:00 INFO DAGScheduler: Job 0 finished: saveAsTextFile at StoreSales.scala:29, took 1.020878 s
20/06/28 22:26:00 INFO SparkUI: Stopped Spark web UI at http://192.168.1.10:4040
20/06/28 22:26:00 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/06/28 22:26:00 INFO MemoryStore: MemoryStore cleared
20/06/28 22:26:00 INFO BlockManager: BlockManager stopped
20/06/28 22:26:00 INFO BlockManagerMaster: BlockManagerMaster stopped
20/06/28 22:26:00 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/06/28 22:26:00 INFO SparkContext: Successfully stopped SparkContext
20/06/28 22:26:00 INFO ShutdownHookManager: Shutdown hook called
20/06/28 22:26:00 INFO ShutdownHookManager: Deleting directory /tmp/spark-563ba68f-4106-48d4-bd16-9951ecbd09ea
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ ls ret-out-StoreSalesScala-01/
part-000000 SUCCESS
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ cat ret-out-StoreSalesScala-01/part-000000
(Albuquerque,1074.88)
(Pittsburgh,1271.35)
(Monahie,565.21)
```



```
File Edit View Search Terminal Help
20/06/28 22:26:00 INFO MemoryStore: MemoryStore cleared
20/06/28 22:26:00 INFO BlockManager: BlockManager stopped
20/06/28 22:26:00 INFO BlockManagerMaster: BlockManagerMaster stopped
20/06/28 22:26:00 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/06/28 22:26:00 INFO SparkContext: Successfully stopped SparkContext
20/06/28 22:26:00 INFO ShutdownHookManager: Shutdown hook called
20/06/28 22:26:00 INFO ShutdownHookManager: Deleting directory /tmp/spark-563ba68f-4106-48d4-bd16-9951ecbd09ea
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ ls ret-out-StoreSalesScala-01/
part-000000 _SUCCESS
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ cat ret-out-StoreSalesScala-01/part-000000
(Albuquerque,1074.88)
(Pittsburgh,1271.35)
(Memphis,565.21)
(Garland,134.33)
(Seattle,934.39)
(Orlando,160.07)
(Dallas,273.49)
(Newark,39.75)
(SCottsdale,880.26)
(Anaheim,114.41)
(Fort Worth,1128.14)
(Irvine,208.16)
(Fremont,465.66)
(Aurora,117.81)
(Reno,230.84999)
(Stockton,247.18)
(San Antonio,125.35)
(Rochester,75.73)
(Lincoln,712.76996)
(Corpus Christi,25.38)
(Jersey City,465.96)
(Omaha,1811.8899)
(Miami,331.75)
(Portland,108.69)
(San Jose,429.87)
(Raleigh,314.1)
(Austin,1787.88)
(Charlotte,440.11)
(Silbert,274.00000)
```

File Edit View Search Terminal Help

(Chandler,1648.77)
(Buffalo,483.82)
(WinstonSalem,361.66)
(Detroit,134.89)
(San Francisco,260.65)
(Cincinnati,323.37997)
(Long Beach,98.23)
(Los Angeles,37.77)
(Bakersfield,217.79)
(Tulsa,431.95)
(Indianapolis,135.96)
(Atlanta,254.62)
(Wichita,400.38)
(Tucson,367.76)
(Denver,413.21002)
(Columbus,392.5)
(Washington,650.48)
(Chesapeake,676.35)
(Saint Paul,172.06)
(El Paso,103.01)
(Lubbock,1517.0801)
(Anchorage,1086.22)
(Phoenix,955.31)
(Spokane,448.97998)
(Santa Ana,922.30005)
(Baltimore,7.98)
(Greensboro,749.73)
(Cleveland,427.43)
(Milwaukee,418.91)
(Las Vegas,146.65)
(Lexington,1080.47)
(Norfolk,669.25)
(Arlington,400.08)
(Philadelphia,1166.76)
(Minneapolis,504.41998)
(Houston,1101.95)
(Toledo,605.23)
(North Las Vegas,409.26)


```
Activities Terminal ▾ Sun 22:27 ● prema@prerna-Inspiron-5559: ~/spark-2.0.2-bin-hadoop2.6
File Edit View Search Terminal Help
(Lubbock,1517.0801)
(Anchorage,1086.22)
(Phoenix,955.31)
(Spokane,448.97998)
(Santa Ana,922.30005)
(Baltimore,7.98)
(Greensboro,749.73)
(Cleveland,427.43)
(Milwaukee,418.91)
(Las Vegas,146.05)
(Lexington,1080.47)
(Norfolk,669.25)
(Arlington,400.08)
(Philadelphia,1166.76)
(Minneapolis,504.41998)
(Houston,1101.95)
(Toledo,605.23)
(North Las Vegas,409.26)
(New Orleans,1118.0801)
(Honolulu,717.57996)
(Fresno,1185.8701)
(Oakland,501.99)
(Boise,481.08997)
(Nashville,765.05)
(Mesa,384.8)
(Tampa,437.19)
(Glendale,220.63)
(Madison,515.48)
(Durham,980.32)
(San Diego,448.92)
(Boston,1114.54)
(Jacksonville,1330.1799)
(St. Petersburg,187.44)
(Fort Wayne,370.55)
(Riverside,1106.01)
(San Bernardino,270.58)
(Virginia Beach,647.07)
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$
```

File Edit View Search Terminal Help

```
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.TotalSales ../jars_file/retail_../data/Re
tail_Sample_Data_Set.txt ret-out-StoreSalesScala-01
retail_products.scala.jar retail_stores.scala.jar retail_totals.scala.jar
prerna@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ bin/spark-submit --class com.prerna.retail.TotalSales ../jars_file/retail_totals.scala.
jar ../data/Retail_Sample_Data_Set.txt ret-out-TotalSalesScala-01
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
20/06/28 22:32:20 INFO SparkContext: Running Spark version 2.0.2
20/06/28 22:32:20 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/06/28 22:32:21 WARN Utils: Your hostname, prerna-Inspiron-5559 resolves to a loopback address: 127.0.1.1; using 192.168.1.10 instead (on in
terface wlp1s0)
20/06/28 22:32:21 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
20/06/28 22:32:21 INFO SecurityManager: Changing view acls to: prerna
20/06/28 22:32:21 INFO SecurityManager: Changing modify acls to: prerna
20/06/28 22:32:21 INFO SecurityManager: Changing view acls groups to:
20/06/28 22:32:21 INFO SecurityManager: Changing modify acls groups to:
20/06/28 22:32:21 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(prerna);
groups with view permissions: Set(); users with modify permissions: Set(prerna); groups with modify permissions: Set()
20/06/28 22:32:21 INFO Utils: Successfully started service 'sparkDriver' on port 33239.
20/06/28 22:32:21 INFO SparkEnv: Registering MapOutputTracker
20/06/28 22:32:21 INFO SparkEnv: Registering BlockManagerMaster
20/06/28 22:32:21 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-30806fd1-b89e-415c-98ef-e95a800bf78e
20/06/28 22:32:21 INFO MemoryStore: MemoryStore started with capacity 366.3 MB
20/06/28 22:32:21 INFO SparkEnv: Registering OutputCommitCoordinator
20/06/28 22:32:21 INFO Utils: Successfully started service 'SparkUI' on port 4040.
20/06/28 22:32:21 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.1.10:4040
20/06/28 22:32:21 INFO SparkContext: Added JAR file:/home/prerna/spark-2.0.2-bin-hadoop2.6/../../jars_file/retail_totals.scala.jar at spark://192.1
68.1.10:33239/jars/retail_totals.scala.jar with timestamp 1593363741882
20/06/28 22:32:21 INFO Executor: Starting executor ID driver on host localhost
20/06/28 22:32:21 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 35793.
20/06/28 22:32:21 INFO NettyBlockTransferService: Server created on 192.168.1.10:35793
20/06/28 22:32:21 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.1.10, 35793)
20/06/28 22:32:21 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.1.10:35793 with 366.3 MB RAM, BlockManagerId(driver, 192.
168.1.10, 35793)
20/06/28 22:32:22 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.1.10, 35793)
20/06/28 22:32:22 WARN SparkContext: Use an existing SparkContext, some configuration may not take effect.
20/06/28 22:32:22 INFO SharedState: Warehouse path is 'file:/home/prerna/spark-2.0.2-bin-hadoop2.6/spark-warehouse'.
20/06/28 22:32:24 INFO FileSourceStrategy: Pruning directories with:
20/06/28 22:32:24 INFO FileSourceStrategy: Post-Scan Filters:
```


File Edit View Search Terminal Help

```
20/06/28 22:32:21 INFO SparkContext: Added JAR file:/home/prerna/spark-2.0.2-bin-hadoop2.6/./jars_file/retail_totalscala.jar at spark://192.168.1.10:33239/jars/retail_totalscala.jar with timestamp 1593363741882
20/06/28 22:32:21 INFO Executor: Starting executor ID driver on host localhost
20/06/28 22:32:21 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 35793.
20/06/28 22:32:21 INFO NettyBlockTransferService: Server created on 192.168.1.10:35793
20/06/28 22:32:21 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.1.10, 35793)
20/06/28 22:32:21 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.1.10:35793 with 366.3 MB RAM, BlockManagerId(driver, 192.168.1.10, 35793)
20/06/28 22:32:22 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.1.10, 35793)
20/06/28 22:32:22 WARN SparkContext: Use an existing SparkContext, some configuration may not take effect.
20/06/28 22:32:22 INFO SharedState: Warehouse path is 'file:/home/prerna/spark-2.0.2-bin-hadoop2.6/spark-warehouse'.
20/06/28 22:32:24 INFO FileSourceStrategy: Pruning directories with:
20/06/28 22:32:24 INFO FileSourceStrategy: Post-Scan Filters:
20/06/28 22:32:24 INFO FileSourceStrategy: Pruned Data Schema: struct<value: string>
20/06/28 22:32:24 INFO FileSourceStrategy: Pushed Filters:
20/06/28 22:32:25 INFO MemoryStore: Block broadcast_0 stored as values in memory (estimated size 214.3 KB, free 366.1 MB)
20/06/28 22:32:25 INFO MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 20.5 KB, free 366.1 MB)
20/06/28 22:32:25 INFO BlockManagerInfo: Added broadcast_0_piece0 in memory on 192.168.1.10:35793 (size: 20.5 KB, free: 366.3 MB)
20/06/28 22:32:25 INFO SparkContext: Created broadcast 0 from rdd at TotalsSales.scala:20
20/06/28 22:32:25 INFO FileSourceStrategy: Planning scan with bin packing, max size: 4194304 bytes, open cost is considered as scanning 4194304 bytes.
20/06/28 22:32:25 INFO CodeGenerator: Code generated in 255.75533 ms
20/06/28 22:32:25 INFO deprecation: mapred.ttip.id is deprecated. Instead, use mapreduce.task.id
20/06/28 22:32:25 INFO deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id
20/06/28 22:32:25 INFO deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap
20/06/28 22:32:25 INFO deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition
20/06/28 22:32:25 INFO deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id
20/06/28 22:32:25 INFO SparkContext: Starting job: saveAsTextFile at TotalsSales.scala:29
20/06/28 22:32:26 INFO DAGScheduler: Registering RDD 4 (map at TotalsSales.scala:22)
20/06/28 22:32:26 INFO DAGScheduler: Got job 0 (saveAsTextFile at TotalsSales.scala:29) with 1 output partitions
20/06/28 22:32:26 INFO DAGScheduler: Final stage: ResultStage 1 (saveAsTextFile at TotalsSales.scala:29)
20/06/28 22:32:26 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 0)
20/06/28 22:32:26 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 0)
20/06/28 22:32:26 INFO DAGScheduler: Submitting ShuffleMapStage 0 (MapPartitionsRDD[4] at map at TotalsSales.scala:22), which has no missing parents
20/06/28 22:32:26 INFO MemoryStore: Block broadcast_1 stored as values in memory (estimated size 9.9 KB, free 366.1 MB)
20/06/28 22:32:26 INFO MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 5.3 KB, free 366.1 MB)
20/06/28 22:32:26 INFO BlockManagerInfo: Added broadcast_1_piece0 in memory on 192.168.1.10:35793 (size: 5.3 KB, free: 366.3 MB)
```

```
Sun 22:32
prema@prerna-Inspiron-5559: ~/spark-2.0.2-bin-hadoop2.6

File Edit View Search Terminal Help

20/06/28 22:32:26 INFO DAGScheduler: looking for newly runnable stages
20/06/28 22:32:26 INFO DAGScheduler: running: Set()
20/06/28 22:32:26 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
20/06/28 22:32:26 INFO DAGScheduler: waiting: Set(ResultStage 1)
20/06/28 22:32:26 INFO DAGScheduler: failed: Set()
20/06/28 22:32:26 INFO DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[6] at saveAsTextFile at TotalSales.scala:29), which has no missing parents
20/06/28 22:32:26 INFO MemoryStore: Block broadcast_2 stored as values in memory (estimated size 65.8 KB, free 366.0 MB)
20/06/28 22:32:26 INFO MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 23.7 KB, free 366.0 MB)
20/06/28 22:32:26 INFO BlockManagerInfo: Added broadcast_2_piece0 in memory on 192.168.1.10:35793 (size: 23.7 KB, free: 366.3 MB)
20/06/28 22:32:26 INFO SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1012
20/06/28 22:32:26 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[6] at saveAsTextFile at TotalSales.scala:29)
20/06/28 22:32:26 INFO TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
20/06/28 22:32:26 INFO TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, partition 0, ANY, 5239 bytes)
20/06/28 22:32:26 INFO Executor: Running task 0.0 in stage 1.0 (TID 1)
20/06/28 22:32:26 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
20/06/28 22:32:26 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 7 ms
20/06/28 22:32:26 INFO FileOutputCommitter: Saved output of task 'attempt_202006282232_0001_m_000000_1' to file:/home/prerna/spark-2.0.2-bin-hadoop2.6/ret-out-TotalSalesScala-01/temporary/0/task_202006282232_0001_m_000000
20/06/28 22:32:26 INFO SparkHadoopMapRedUtil: attempt_202006282232_0001_m_000000_1: Committed
20/06/28 22:32:26 INFO Executor: Finished task 0.0 in stage 1.0 (TID 1). 1808 bytes result sent to driver
20/06/28 22:32:26 INFO TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 121 ms on localhost (1/1)
20/06/28 22:32:26 INFO TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
20/06/28 22:32:26 INFO DAGScheduler: ResultStage 1 (saveAsTextFile at TotalSales.scala:29) finished in 0.122 s
20/06/28 22:32:26 INFO DAGScheduler: Job 0 finished: saveAsTextFile at TotalSales.scala:29, took 0.943037 s
20/06/28 22:32:26 INFO SparkUI: Stopped Spark web UI at http://192.168.1.10:4040
20/06/28 22:32:26 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
20/06/28 22:32:26 INFO MemoryStore: MemoryStore cleared
20/06/28 22:32:26 INFO BlockManager: BlockManager stopped
20/06/28 22:32:26 INFO BlockManagerMaster: BlockManagerMaster stopped
20/06/28 22:32:26 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
20/06/28 22:32:26 INFO SparkContext: Successfully stopped SparkContext
20/06/28 22:32:26 INFO ShutdownHookManager: Shutdown hook called
20/06/28 22:32:26 INFO ShutdownHookManager: Deleting directory /tmp/spark-c770d2fc-7e4b-45d3-b213-1f31804e637c
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$ cat ret-out-TotalSalesScala-01/part-00000
(Total Sales Done: ,49585.37)
prema@prerna-Inspiron-5559:~/spark-2.0.2-bin-hadoop2.6$
```