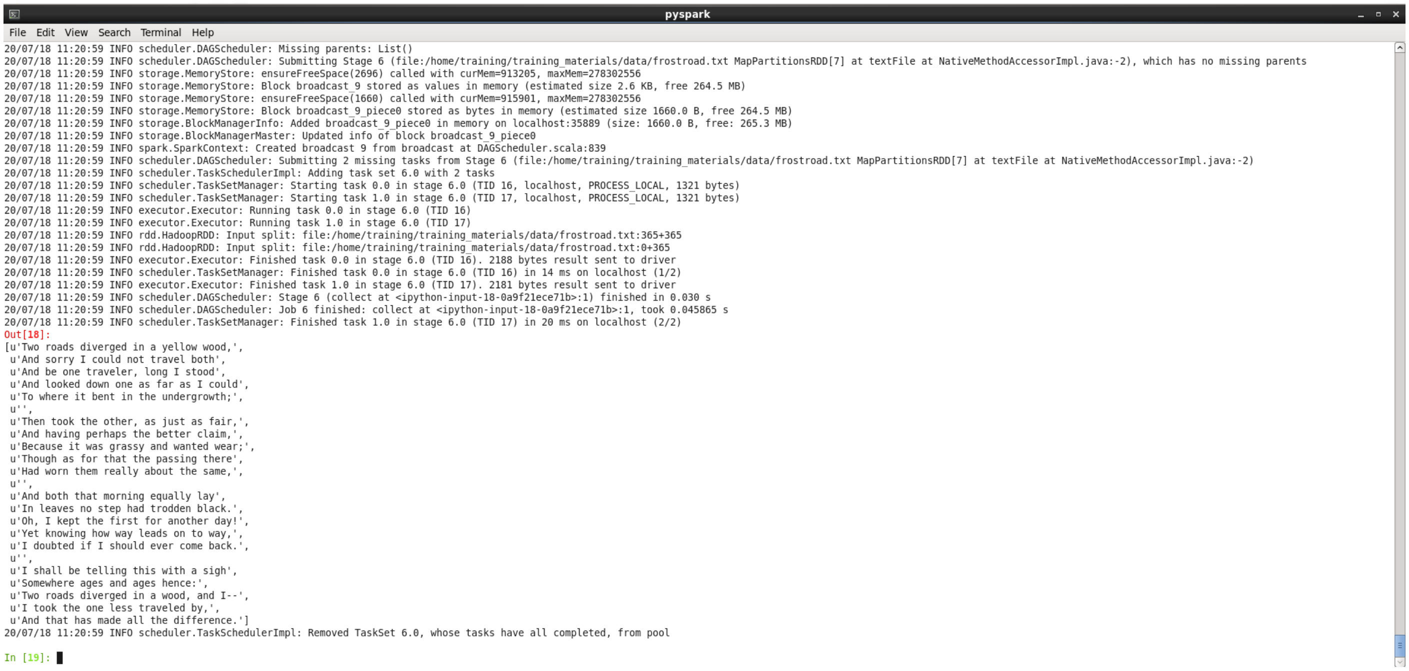
BUAN 6346 Big Data

Homework 10b

Explore RDDs using the Spark Shell

Step11: mydata.collect()



```
20/07/18 11:20:59 INFO scheduler.DAGScheduler: Missing parents: List()
20/07/18 11:20:59 INFO scheduler.DAGScheduler: Submitting Stage 6 (file:/home/training/training_materials/data/frostroad.txt MapPartitionsRDD[7] at textFile at NativeMethodAccessorImpl.java:-2), which has no missing parents
20/07/18 11:20:59 INFO storage.MemoryStore: ensureFreeSpace(2696) called with curMem=913205, maxMem=278302556
20/07/18 11:20:59 INFO storage.MemoryStore: Block broadcast_9 stored as values in memory (estimated size 2.6 KB, free 264.5 MB)
20/07/18 11:20:59 INFO storage.MemoryStore: ensureFreeSpace(1660) called with curMem=915901, maxMem=278302556
20/07/18 11:20:59 INFO storage.MemoryStore: Block broadcast_9_piece0 stored as bytes in memory (estimated size 1660.0 B, free 264.5 MB)
20/07/18 11:20:59 INFO storage.BlockManagerInfo: Added broadcast_9_piece0 in memory on localhost:35889 (size: 1660.0 B, free: 265.3 MB)
20/07/18 11:20:59 INFO storage.BlockManagerMaster: Updated info of block broadcast_9_piece0
20/07/18 11:20:59 INFO spark.SparkContext: Created broadcast 9 from broadcast at DAGScheduler.scala:839
20/07/18 11:20:59 INFO scheduler.DAGScheduler: Submitting 2 missing tasks from Stage 6 (file:/home/training/training_materials/data/frostroad.txt MapPartitionsRDD[7] at textFile at NativeMethodAccessorImpl.java:-2)
20/07/18 11:20:59 INFO scheduler.TaskSchedulerImpl: Adding task set 6.0 with 2 tasks
20/07/18 11:20:59 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 6.0 (TID 16, localhost, PROCESS_LOCAL, 1321 bytes)
20/07/18 11:20:59 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 6.0 (TID 17, localhost, PROCESS_LOCAL, 1321 bytes)
20/07/18 11:20:59 INFO executor.Executor: Running task 0.0 in stage 6.0 (TID 16)
20/07/18 11:20:59 INFO executor.Executor: Running task 1.0 in stage 6.0 (TID 17)
20/07/18 11:20:59 INFO rdd.HadoopRDD: Input split: file:/home/training/training_materials/data/frostroad.txt:365+365
20/07/18 11:20:59 INFO rdd.HadoopRDD: Input split: file:/home/training/training_materials/data/frostroad.txt:0+365
20/07/18 11:20:59 INFO executor.Executor: Finished task 0.0 in stage 6.0 (TID 16). 2188 bytes result sent to driver
20/07/18 11:20:59 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 6.0 (TID 16) in 14 ms on localhost (1/2)
20/07/18 11:20:59 INFO executor.Executor: Finished task 1.0 in stage 6.0 (TID 17). 2181 bytes result sent to driver
20/07/18 11:20:59 INFO scheduler.DAGScheduler: Stage 6 (collect at <ipython-input-18-0a9f21ece71b>:1) finished in 0.030 s
20/07/18 11:20:59 INFO scheduler.DAGScheduler: Job 6 finished: collect at <ipython-input-18-0a9f21ece71b>:1, took 0.045865 s
20/07/18 11:20:59 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 6.0 (TID 17) in 20 ms on localhost (2/2)
Out[18]:
[u'Two roads diverged in a yellow wood,',
 u'And sorry I could not travel both',
 u'And be one traveler, long I stood',
 u'And looked down one as far as I could',
 u'To where it bent in the undergrowth;',
 u'',
 u'Then took the other, as just as fair,',
 u'And having perhaps the better claim,',
 u'Because it was grassy and wanted wear;',
 u'Though as for that the passing there',
 u'Had worn them really about the same,',
 u'',
 u'And both that morning equally lay',
 u'In leaves no step had trodden black.',
 u'Oh, I kept the first for another day!',
 u'Yet knowing how way leads on to way,',
 u'I doubted if I should ever come back.',
 u'',
 u'I shall be telling this with a sigh',
 u'Somewhere ages and ages hence:',
 u'Two roads diverged in a wood, and I--',
 u'I took the one less traveled by,',
 u'And that has made all the difference.']
20/07/18 11:20:59 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool

In [19]: 
```