Job Run Time (Approx In Minutes)

	Single Record Lookup	Filter	Group By Accompained With Order By
Pig	2	7	12
Spark RDD	2	4	5

Total Count Of Records

	Single Record Lookup	Filter	Group By Accompained With Order By
Pig	1	31035	(1,36492406) (2,17648209) (3,292090) (4,85287)
Spark RDD	1	31035	(1,36492406) (2,17648209) (3,292090) (4,85287)

Analysis And Conclusion

Execution Environment:

1. Single node Cloudera Hadoop Cluster
2. On an Oracle VM VirtualBox
   - 12 GB RAM
   - i5 4 core processor

Based the analysis of the above statistics in reference to the execution environment, it is evident that there is a significant throughput when processing data in Spark as compared to Pig (or Any Other MapReduce program).

Spark is an in-memory computation engine which provides us the benefit of higher throughput in comparision to Pig scripts which are converted to a MapReduce program(s) before the MapReduce program(s) is executed. A MapReduce program is essentially used for batch-processing where disk I/O is involved resulting in lower throughput.