Job Run Time

	Single Record Lookup	Filter	Group by Accompanied with Order by
Pig	Start time: 04:21:36 End time: 04:24:28 Total time taken: 00:02:52 (2 min 52 sec)	Start time: 04:29:17 End time: 04:32:17 Total time taken: 00:03:00 (3 min)	Start time: 04:36:49 End time: 04:42:26 Total time taken: 00:05:37 (5 min 37 sec)
Spark	Start time: 03:55:17 End time: 03:56:54 Total time taken: 00:01:37 (1 min 37 sec)	Start time: 04:03:37 End time: 04:05:07 Total time taken: 00:01:30 (1 min 30 sec)	Start time: 04:09:56 End time: 04:11:41 Total time taken: 00:01:45 (1 min 45 sec)

Total Count of Records in Output

	Single Record Lookup	Filter	Group by Accompanied with Order by
Pig	1	31029	4
Spark	1	31029	4

Logical explanation

1. Output is same for both pig and spark. Count of records is same and records are also same. It is just that output of pig is tab separated and output of spark is comma separated.
2. Time taken by spark is quite less in comparison to pig.
3. **In single record lookup, spark is 43.6% faster than pig.**
   Time taken by pig in seconds = 172
   Time taken by spark in seconds = 97
   % faster = (172-97)/172 = 75/172 * 100 = 43.6%
4. **In filter by operation, spark is 50% faster than pig.**
   Time taken by pig in seconds = 180
   Time taken by spark in seconds = 90
   % faster = (180-90)/180 = 90/180 * 100 = 50%
5. **In group by accompanied with order by operation, spark is 68.8% faster than pig.**
   Time taken by pig in seconds = 337
   Time taken by spark in seconds = 105
   % faster = (337-105)/337 = 232/337 * 100 = 68.8%

Conclusion

Spark which is based on in-memory processing is much faster than pig (map reduce) which is based on disk-based processing.