

1. Looking at the below results , with Spark there is an improvement of 30% - 50% in each and every scenario - lookup of a single row from the entire data set , Filtering multiple rows and Group by and then order by and then Sort on an input dataset of ~ 7GB.
2. Hence, in comparison of Spark and Pig which internally implements Map Reduce : Spark is the clear winner in our POC.
- 3.The time taken to load data , process data and retrieve results for certain logic can be considerably reduced with Spark in Big Data Problems
4. With Spark, one can definitely look for enhanced performance and hence better decision making in real time.

Job Run Time	Single Record Lookup	Filter	Group by Accompanied with Order by
Pig	00:04:39	00:04:57	00:05:48
Spark RDD	00:02:58	00:03:07	00:03:13

Total Count of Records in Output	Group by Accompanied with Order by	Group by Accompanied with Order by	Group by Accompanied with Order by
Pig	1	31029	4
Spark RDD	1	31029	4