

Khushboo Tekchandani

CSE587 Homework 3

Solution:

The problem requires us to calculate the volatilities of NASDAQ stocks and find the stocks with top 10 maximum and the top 10 minimum volatilities.

There are two different solutions to this problem based on the technologies used. The two technologies employed to solve this problem are Pig Latin and Hive QL.

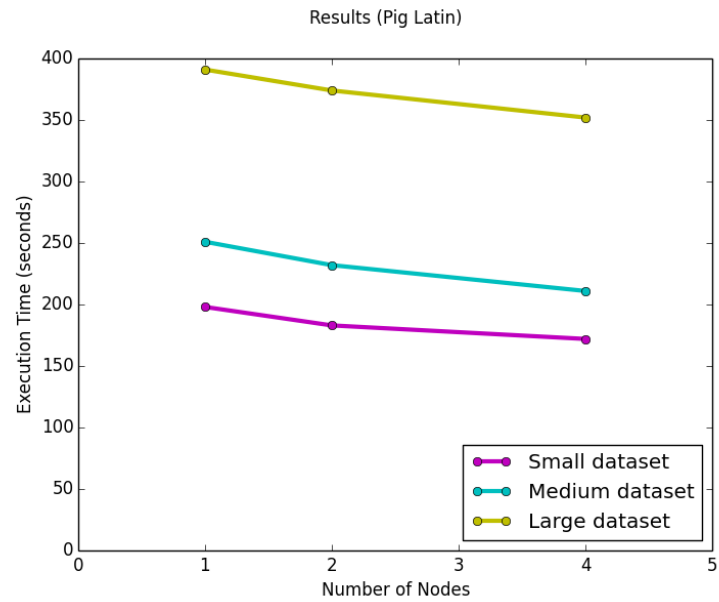
I have used the Pig Latin script and the Hive QL without the use of any UDFs as both these technologies have a wide variety of inbuilt query to find a solution to our problem.

Evaluation:

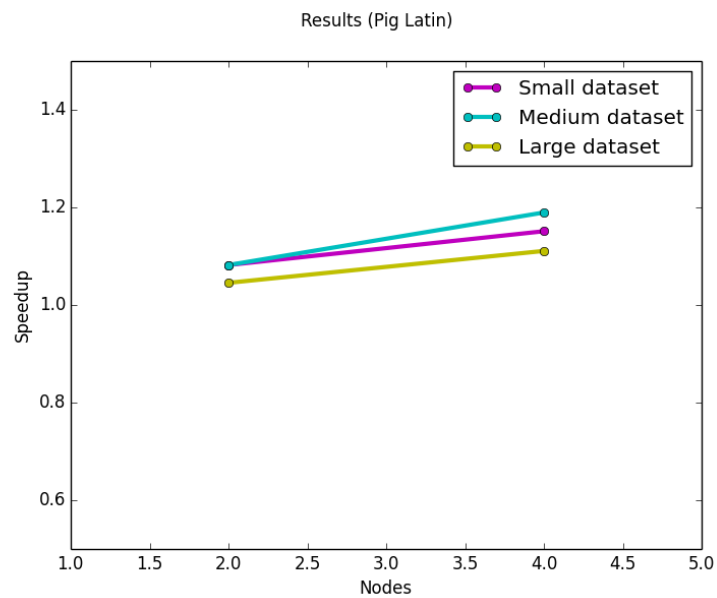
Using these implementations, I executed my code on CCR to understand the scalability and performance of my solution using the two technologies. The experiment was done using different data sets and different number of nodes. Following data was obtained by carrying out these experiments:

Problem Size	Execution Time in Seconds		
	1 node (12 cores)	2 nodes (24 cores)	4 nodes (48 cores)
Small	198	183	172
Medium	251	232	211
Large	391	374	352

Table 1: Execution time of the program on different number of nodes, using multiple data sets.
(Pig Latin)



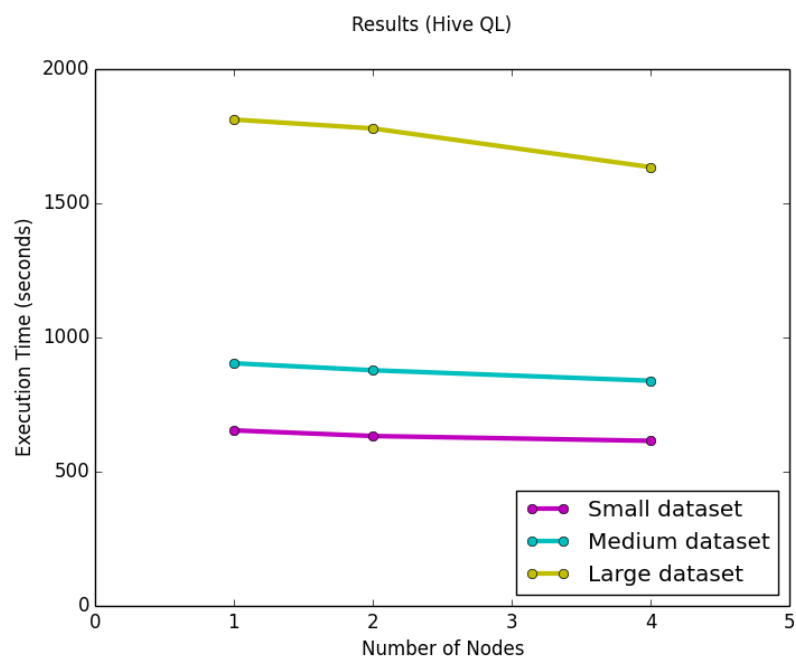
Graph1: Execution time for different problem sizes on different number of compute nodes



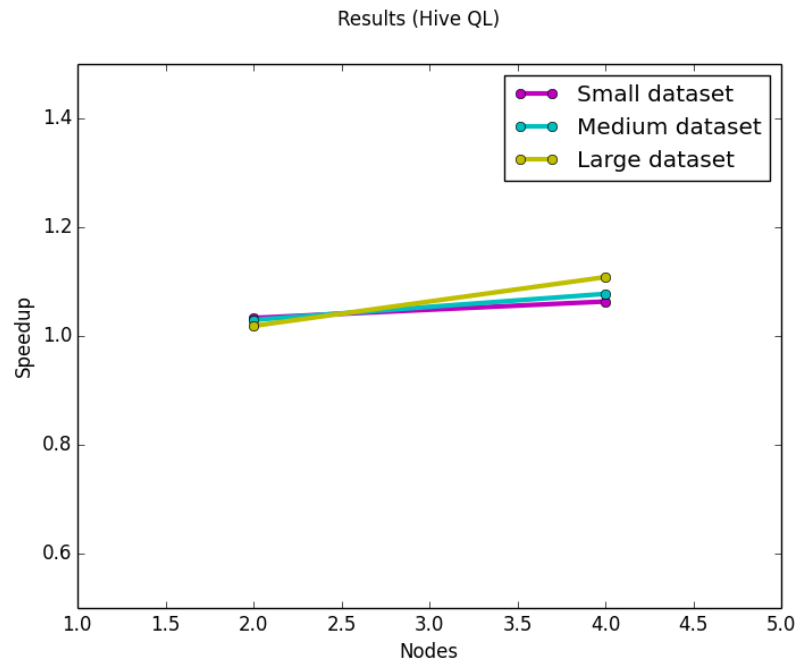
Graph2: Speedup of the result of increasing the compute nodes at different data sets

	Execution Time in Seconds		
Problem Size	1 node (12 cores)	2 nodes (24 cores)	4 nodes (48 cores)
Small	654	633	615
Medium	904	878	839
Large	1812	1779	1635

Table 2: Execution time of the program on different number of nodes, using multiple data sets.
(Hive QL)



Graph3: Execution time for different problem sizes on different number of compute nodes



Graph2: Speedup of the result of increasing the compute nodes at different data sets

Observations:

For both Pig and Hive, execution times for various number of nodes follow a similar graph over different data sets. Same holds true for Speed up.

Both Pig and Hive give a better speedup for a larger data set.

Pig performs much better than Hive QL.