DATA INTENSIVE COMPUTING
CSE587

PROGRAMMING ASSIGNMENT-4
REPORT

By
Name: Vishwas Shanbhog
UBIT No: 50135707

Introduction

In this assignment, calculation of the volatility for stocks was used to evaluate the performance of HBase with Hadoop Map reduce. There were three sets of stocks given. The goal of this assignment was to calculate the top 10 stocks with maximum volatility and bottom 10 stocks with lowest volatility. The platform used was HBASE. All the datasets (small, medium and large) were run on with different node configurations which were 1 nodes (12 cores), 2nodes (24 cores) and 4 nodes (48 cores) on HBASE. The results of the runtime are described below.

The HBASE execution times are given below:

HBASE:-

Problem Size	Execution time(1 node) 12 cores in secs	Execution time(2 node) 24 cores in secs	Execution time(3 nodes) 36 cores In secs	Execution time(4 node) 48 cores In secs
small	3568	1675	956	723
medium	4553	3635	2803	1934
large	12700	10700	8537	6326

HADOOP MAPREDUCE:

The runtime of calculating in pig for small, medium and large datasets for 1,2 and 4 nodes are as follows. The results:

Problem size	Execution time(1 node- 12cores)	Execution time(2 nodes- 24 cores)	Execution time(4 nodes- 48 cores)
small	2623 sec	1461 sec	875 sec
medium	8313 sec	5032 sec	2231 sec
Large	26568 sec	12967 sec	7199 sec

PIG:

The runtime of calculating in pig for small, medium and large datasets for 1, 2 and 4 nodes are as follows:

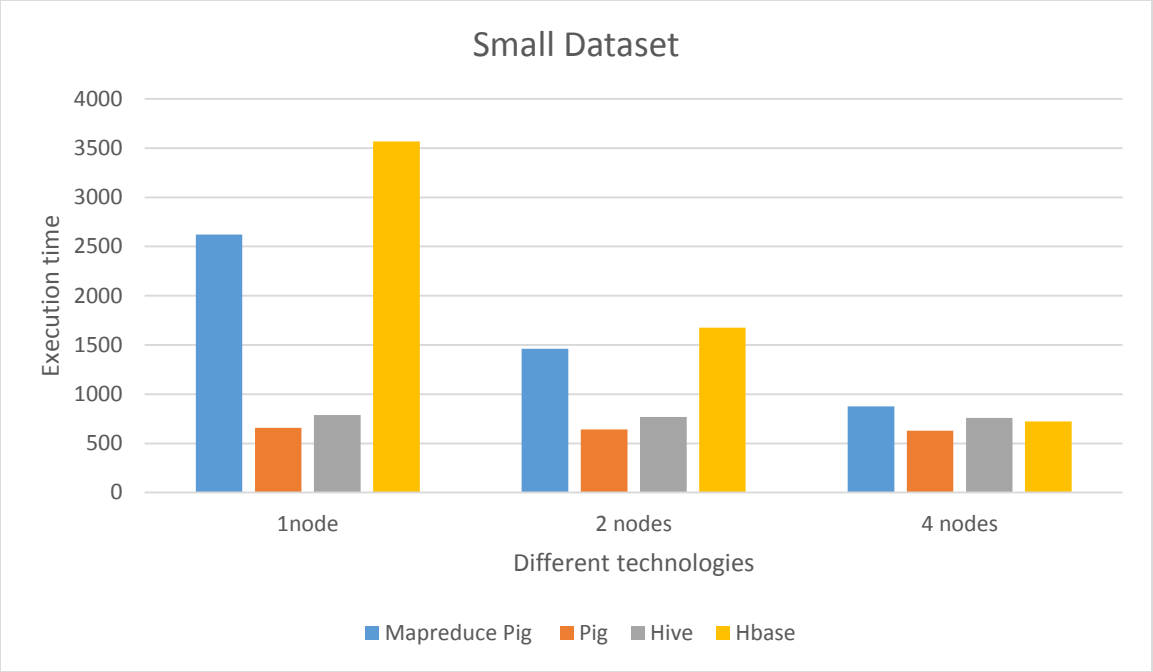
Problem size	Execution time(1 node- 12cores)	Execution time(2 nodes- 24 cores)	Execution time(4 nodes- 48 cores)
Small	657 sec	643 sec	630 sec
medium	1369 sec	1350 sec	1347 sec
Large	2704 sec	2678 sec	2669 sec

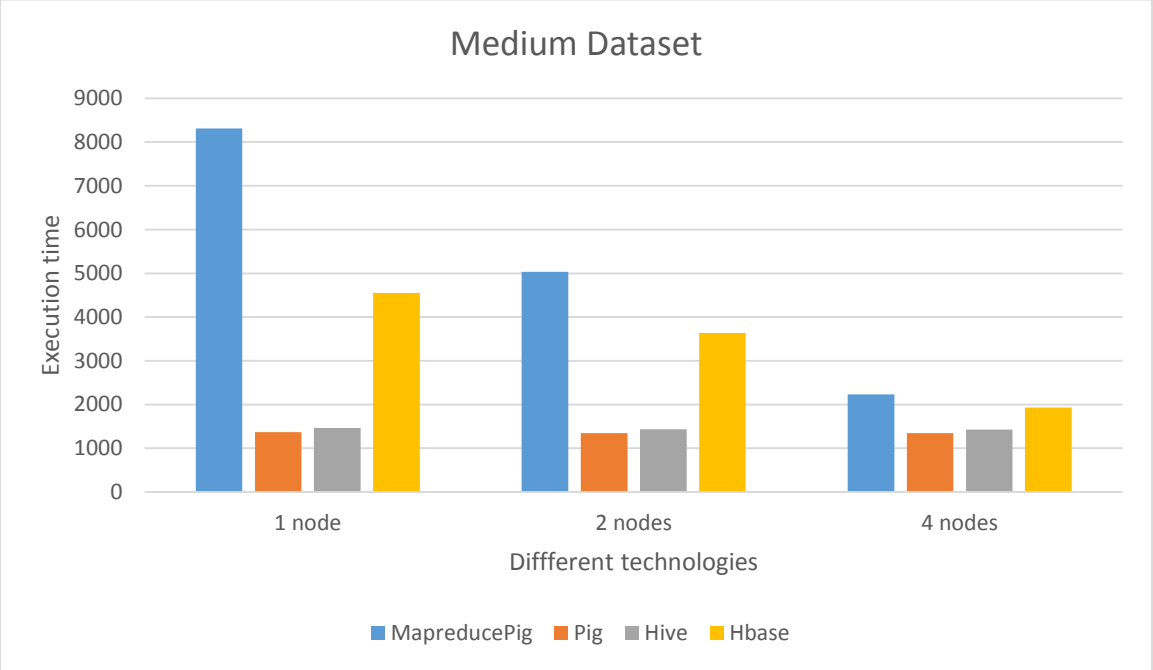
HIVE:

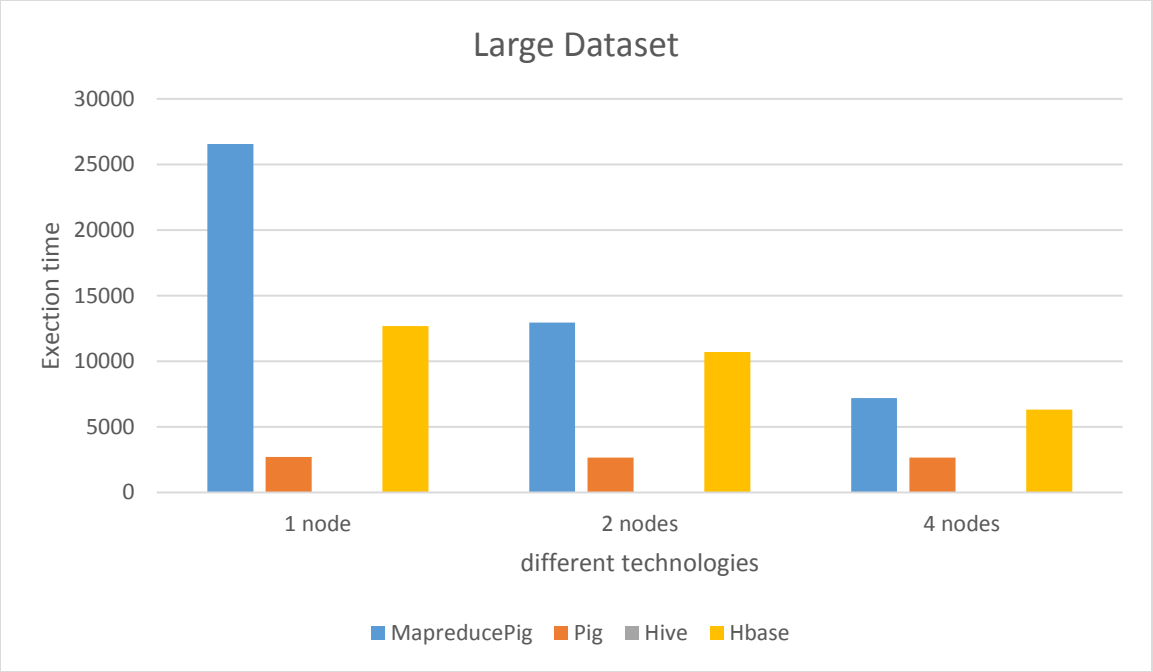
The runtime of calculating in pig for small, medium and large datasets for 1,2 and 4 nodes are as follows. :

Problem size	Execution time(1 node- 12cores)	Execution time(2 nodes- 24 cores)	Execution time(4 nodes- 48 cores)
small	787 sec	768 sec	760 sec
medium	1465 sec	1432 sec	1430 sec
Large	General compute error	General compute error	General compute error

Comparison of values:-


Small Dataset
4000
3500
3000
2500
2000
1500
1000
500
0
Execution time
1node
2 nodes
4 nodes
Different technologies
Mapreduce Pig
Pig
Hive
Hbase


Medium Dataset
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Execution time
1 node
2 nodes
4 nodes
Diffferent technologies
MapreducePig
Pig
Hive
Hbase


Large Dataset
30000
25000
20000
15000
10000
5000
0
Exection time
1 node
2 nodes
4 nodes
different technologies
MapreducePig
Pig
Hive
Hbase