

---

# Volatility Calculation Using PIG and HIVE

**CSE 587**

---

Prakash Natarajan - 30 March 2015

Mail id : [pn33@buffalo.edu](mailto:pn33@buffalo.edu)

Id: 50134214

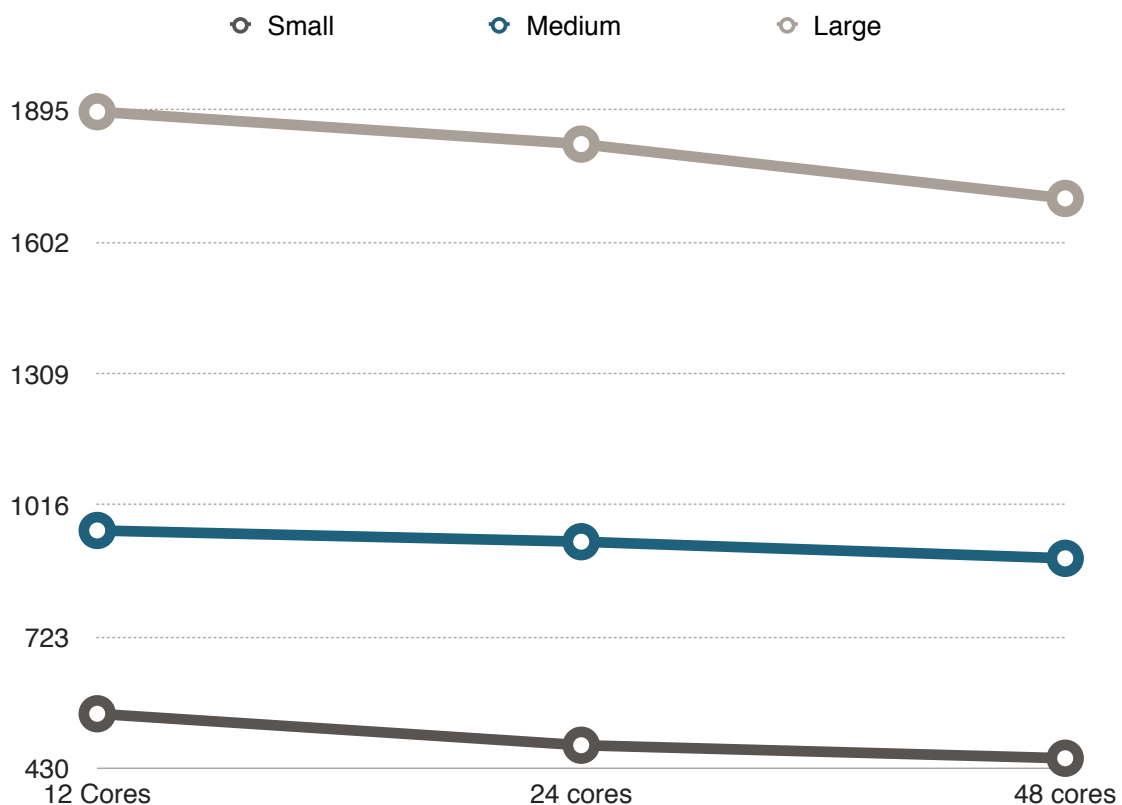


## PIG

Pig is a high-level platform for creating MapReduce programs used with Hadoop. The language for this platform is called Pig Latin. Pig Latin abstracts the programming from the Java MapReduce idiom into a notation which makes MapReduce programming high level, similar to that of SQL for RDBMS systems. Pig Latin can be extended using UDF (User Defined Functions) which the user can write in Java, Python, JavaScript, Ruby or Groovy and then call directly from the language.

In this project we are using java as a user defined function for calculating the values of xi and Volatility.

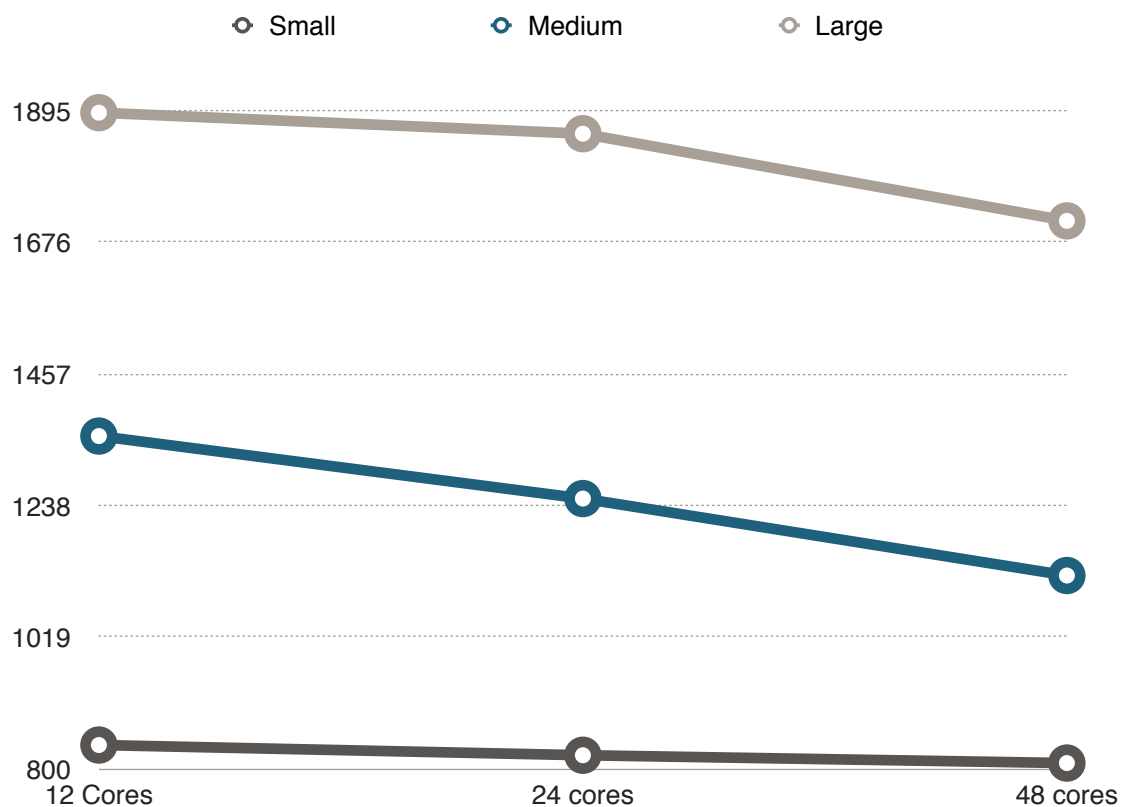
Problem Size	Execution Time : 1 Node(12 cores)	Execution Time: 2 Node(24 cores)	Execution Time: 3 Node(48 cores)
Small	552	482	453
Medium	960	935	898
Large	1892	1820	1699



## HIVE

Apache Hive is a data warehouse infrastructure built on top of Hadoop for providing data summarisation, query, and analysis. While initially developed by Facebook, Apache Hive is now used and developed by other companies such as Netflix.

Problem Size	Execution Time : 1 Node(12 cores)	Execution Time: 2 Node(24 cores)	Execution Time: 3 Node(48 cores)
Small	840	823	810
Medium	1354	1250	1122
Large	1892	1857	1712



---

## Conclusion

Pig and Hive seems to play a greater part in mapreduce programs. It had reduced the job of the mapreduce programmer to a greater extent. Coding entire mapreduce program in java looked cumbersome. Pig and hive both invariantly reduced the time as well as effort needed to write map reduce programs.