

19th September, 2018

Weekly Report

D Swami

Project Title: On studying the performance of Hadoop Map Reduce vs MPI for Aggregation Operations: A Big Data Challenge

The following project aimed at benchmarking various parameters of Map Reduce & MPI for parallel I/O. In the first week of the work, I have accomplished following tasks:

- 1) Build the Hadoop from source using required libraries (zlib and snappy, to be used for compression).
- 2) Check the build and ran a small word count application to confirm their working.
- 3) Developed the strategy to find min and max values in same Map Reduce program and implemented the same using Java.
- 4) A small test of the working of the Map Reduce Jar was carried out using Eclipse.
- 5) The complete dataset has been downloaded and stored in an external drive; ready for ingestion.

Issues tackled in the current week:

- 1) Apache Hadoop build errors resulting from using Oracle Java 9. The problem was identified to be Activation package (javax.activation) was discontinued to be packed from the Java 9 bundle. Currently used the JAVA_OPTS parameter to get activation package but plan to downgrade to oracle Java 9 in next week.
- 2) Hadoop jars class path errors resolved for Map Reduce program builds when using Eclipse.

Tasks for the upcoming week:

- 1) Identify the best way to ingest the data. Also, includes a literature review of available tools. (Wednesday)
- 2) Ingest the data and record the ingestion rate. (Wednesday)
- 3) Debug the Map Reduce code for a small number of input files and record the time stamp and memory usage. This would serve as a benchmark for carrying out complete analysis. (Thursday, Friday, Saturday)
- 4) Literature review of MPI for group by aggregate queries. (Monday & Tuesday).

Expected Issues in the coming week:

- 1) Ingestion issues.
- 2) Out-of-Memory Heap issues with Map Reduce using `mapred.child.java.opts=-Xmx1024M` should solve the problem. The following property would allocate a max of 1 GB per mapper task and should suffice.