Load Testing and Benchmarking for BigData

Aayush Agrawal, Sunil Raiyani, Jayam Modi

July 4, 2014

Aim of the Project

The aim of the project is Load Testing and Benchmarking for BigData.

The major task is to setup a distributed file system on a cluster, test the data and query processing capacity of the system using BigBench and predict the performance of the system for larger data sets.
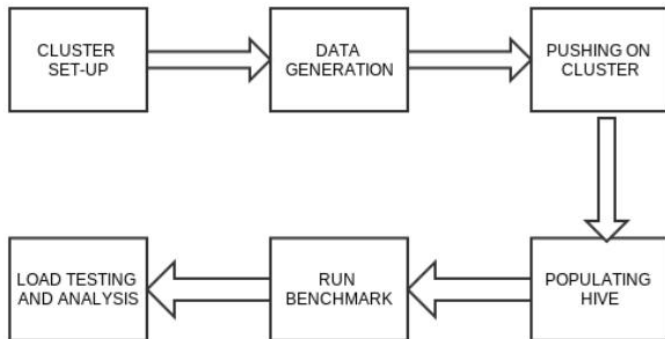
Block Diagram

Figure : Average Query Response Times for different data sizes

Achievements

1. Customized version of BigBench.[1]
2. Evaluating factors affecting the average query response time on the cluster.

Future Scope

The following work is considered for future:

- Run the experiment for clusters of larger size and try to determine the optimum size of cluster of the given configuration of nodes to manage a particular data-set size.

- Analyze other parameters affecting the average query response time of the system.

- Open Source Release of our customized version of BigBench.[2]

References I

[1] "BigBench Installation
   https://github.com/intel-hadoop/Big-Bench/blob/
   master/README.md. Accessed on July 4, 2014."

[2] T. Rabl, A. Ghazal, M. Hu, A. Crolotte, F. Raab, M. Poess, and
   H.-A. Jacobsen, "Bigbench specification v0. 1," in *Specifying Big
   Data Benchmarks*, pp. 164–201, Springer, 2014.