

Proposal Document

- Project Overview

Evaluation of Big Data frameworks using different Machine Learning algorithms

- Problem Definition

The processes involved in Machine Learning (training/testing) are computationally very intensive and take a lot of time for completion. Various Big Data frameworks like Hadoop, Spark, Tensorflow, CUDA exists for this purpose. Our goal is to evaluate the performance of these big data technologies using ML algorithms.

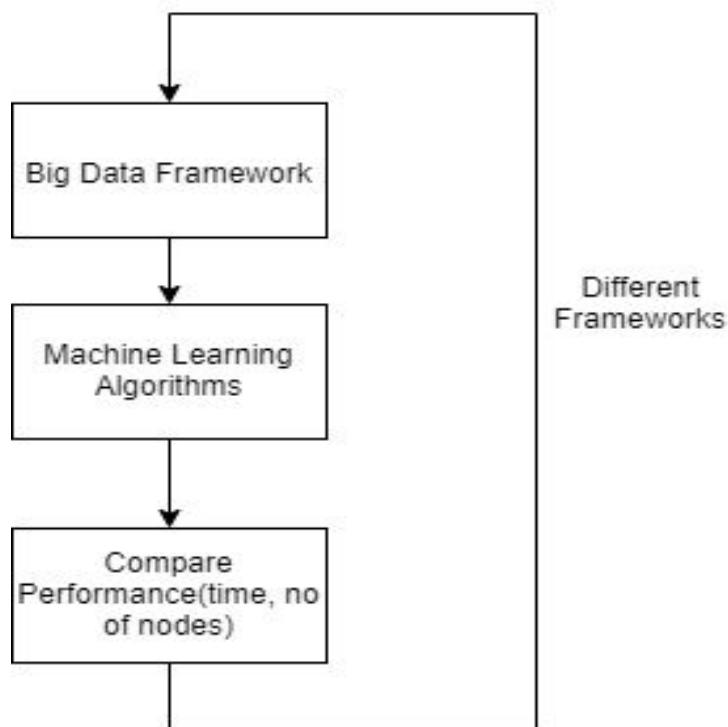
- Scope

This project will be useful for people to identify a Big Data framework on basis of comparison study.

- Objectives

- To identify the popular Big Data frameworks in use today.
- To identify the ML algorithm and dataset for the benchmarking of the Big Data frameworks.
- To implement and use all the framework for a particular ML algorithm and vary the number of computational nodes.
- To compare the performance of the Big Data frameworks

- Generic project life cycle for the chosen technology



- Planned Technology to be used
 - Programming language: Python
 - Machine Learning framework: Tensorflow
 - Big Data framework: Hadoop - Map reduce, Spark, Hana, Tensorflow, CUDA