

Second Deliverable

The second deliverable consists of 5 major parts which will be used for the final integration given in the 3rd deliverable. About 60% of the project has been completed. The language dependent code has to be written. After that if time permits, additional functionalities will be added. The 5 areas which have been worked on are:

1. **Installation of framework:** We chose 4 architectures Hadoop, RHadoop, Mahout and Giraph. All these 4 run on distributed mode. All these 4 have been installed manually. This part contained the maximum amount of error and hence the maximum amount of time and effort.
2. **Installation scripts:** Basic installation scripts for the installation of all the frameworks on any new architecture has been written. *install_all.sh* is the file which will install all the frameworks in */usr/local*. (Assumption: user has configured the host and slave files according to his needs and the changes which are required apart from the core install have also been committed.)
3. **API Specifications:** 2 major tasks execution and addition have been considered and 2 subtasks in each of these have also been identified. RESTles API and the corresponding JSON output format has been given in the file *API Specifications.pdf*.
4. **Algorithm argument and parameters:** Top 5-10 needed/accessed/searched algorithms from each of the frameworks has been studied to identify how are they run on any specific framework. Argument structure and parameter value are checked for all the these algorithms and mentioned in the file *Algorithm argument and parameters.pdf*.
5. **Algorithms:** *Algorithms.pdf* contains the algorithm for running the analytics engine. The algorithms are subdivided into 3 major types and explained in details with diagram flow thereon.