ABSTRACT

Hadoop is a framework that supports data-intensive distributed applications. And is used to process data on thousands of computers Two scalable distributed clustering methods are proposed: the Single Clustered File method and the Multiple Clustered Files method Both methods were executed on one to eight machines in Hadoop.

This Project is dealing with implementation of bigdata using hadoop framework .we have provided the support of data hadoop version 1.2.1 and version 2.6.4. The project includes implementation of Docker Technology. My project provide two way to set up hadoop cluster

- **AUTOMATIC CONFIGURATION**
- **ON DEMAND CONFIGURATION**

AUTOMATIC CONFIGURATION

In this phase we configure hadoop cluster in such a way that it uses computer resources in optimized way . So it selects automatically a node which has more storage and ram used as a Datanode and TaskTracker and which has less ram and storage used as a Namenode and JobTracker among all the available node and we are also providing a simple and well-structured web interface to the user so that he can use the cluster in efficient way.

ON DEMAND CONFIGURATION

In this phase we ask the user to enter number of datanode and task tracker and it creates instantly. To implement this we use a tool called Docker. Docker containers wrap a piece of software in a complete filesystem that contains everything needed to run: code, runtime, system tools, system libraries – anything that can be installed on a server. This guarantees that the software will always run the same, regardless of its environment.

LIST OF CONTENT