SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

COURSE NAME:BDI **CLASS:** Third Year Beach

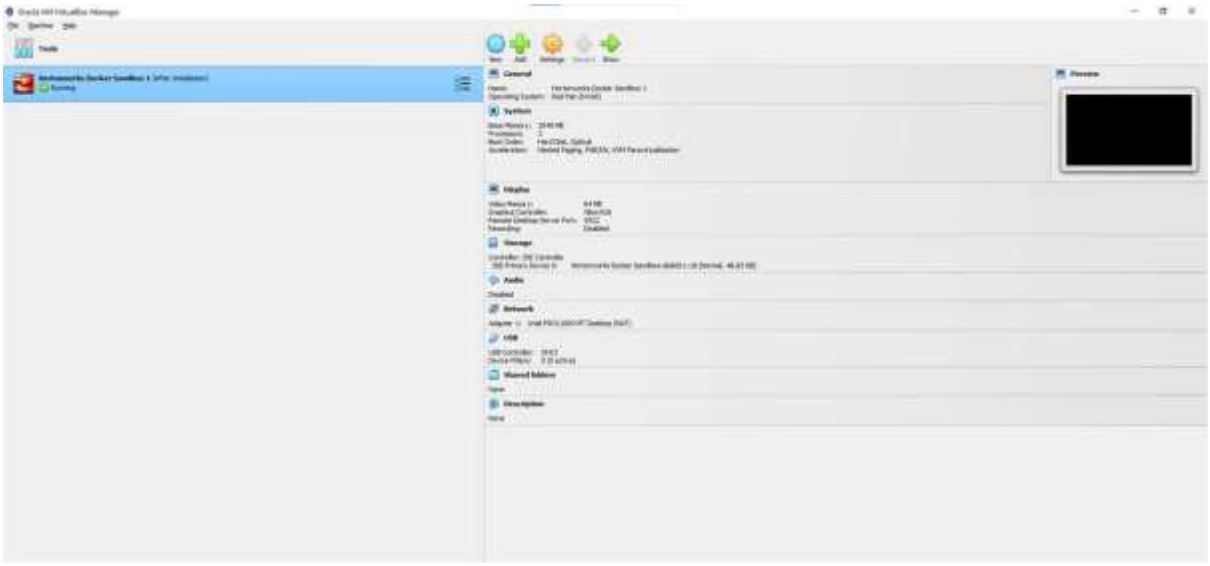
NAME: Falguni Parmar **BATCH:** C22

Experiment - 2

Aim: Installation of Hadoop on a Single Node Cluster.

Theory:

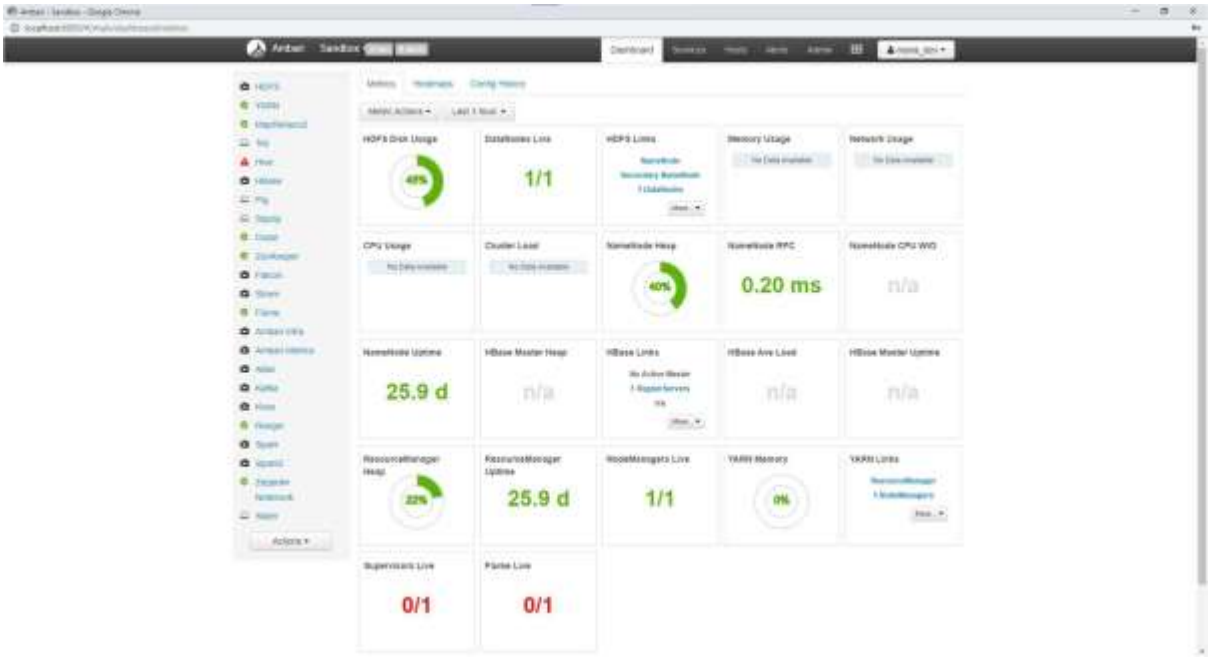
To install Hadoop on a Single Node Cluster, we can use the Hortonworks Docker Sandbox 1. This can be done on a virtual machine running on the Red Hat (64-bit) operating system. First, we need to download and install the Docker software on the virtual machine. Once the Docker is installed, we can pull the Hortonworks Sandbox image from the Docker Hub repository. Then, we can start the container using the "docker run" command, and map the container's ports to the virtual machine's ports. After that, we can access the Hadoop web interface on the virtual machine's browser, which will allow us to run and manage Hadoop jobs. Finally, we can use the Hadoop command line interface to submit jobs and monitor their progress. This installation method provides an easy and convenient way to experiment with Hadoop in a Single Node Cluster environment without the need for a dedicated physical machine.



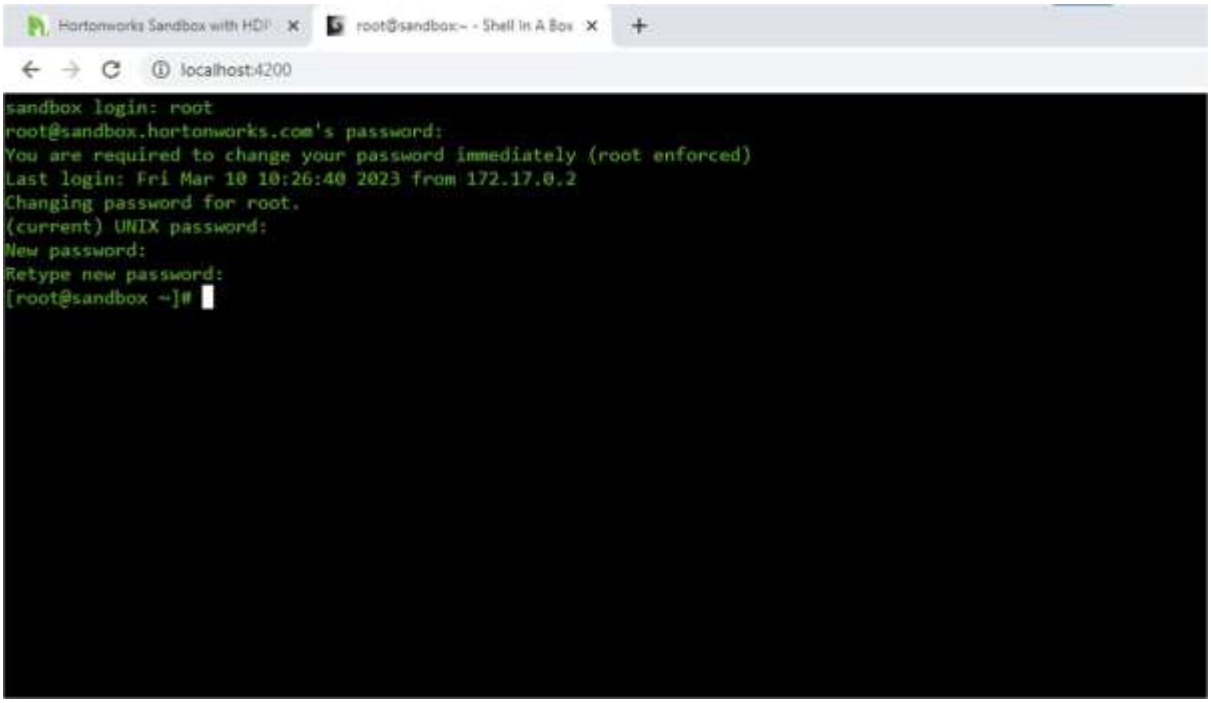
After the successful installation of Hadoop on the virtual machine, a localhost port is generated on port 8080 which loads the Ambari dashboard homepage. The Ambari dashboard is a user-friendly web interface that allows users to manage and monitor their Hadoop cluster with ease. From the dashboard, users can manage and monitor services, set up alerts, view performance metrics, and monitor system health. This dashboard provides an intuitive graphical interface that simplifies the management of a Hadoop cluster, making it accessible to a wider range of users, including those who may not be familiar with commandline interfaces.

SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

Where credentials are 'maria_dev'



After logging in to HDFS using the provided login credentials, users can start performing various HDFS commands to manage their files and directories in the distributed file system. Some common HDFS commands include creating and deleting directories, uploading and downloading files, and setting permissions on files and directories. HDFS also supports other advanced features, such as data replication and block size configuration, which can be useful for optimizing the performance and reliability of the distributed file system. Overall, HDFS provides a robust and scalable platform for storing and managing large amounts of data across a Hadoop cluster, making it an essential component of any big data processing infrastructure.



SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

Conclusion:

Installing Hadoop on a Single Node Cluster using the Hortonworks Docker Sandbox 1 on a virtual machine running Red Hat (64-bit) OS is a straightforward process. The Ambari dashboard can be accessed on localhost port 8080, and users can manage their Hadoop cluster and access HDFS to manage files and directories using HDFS commands. This installation method provides an easy and efficient way to experiment with Hadoop in a Single Node Cluster environment.