MAPREDUCE - HADOOP ADMINISTRATION

This chapter explains Hadoop administration which includes both HDFS and MapReduce administration.
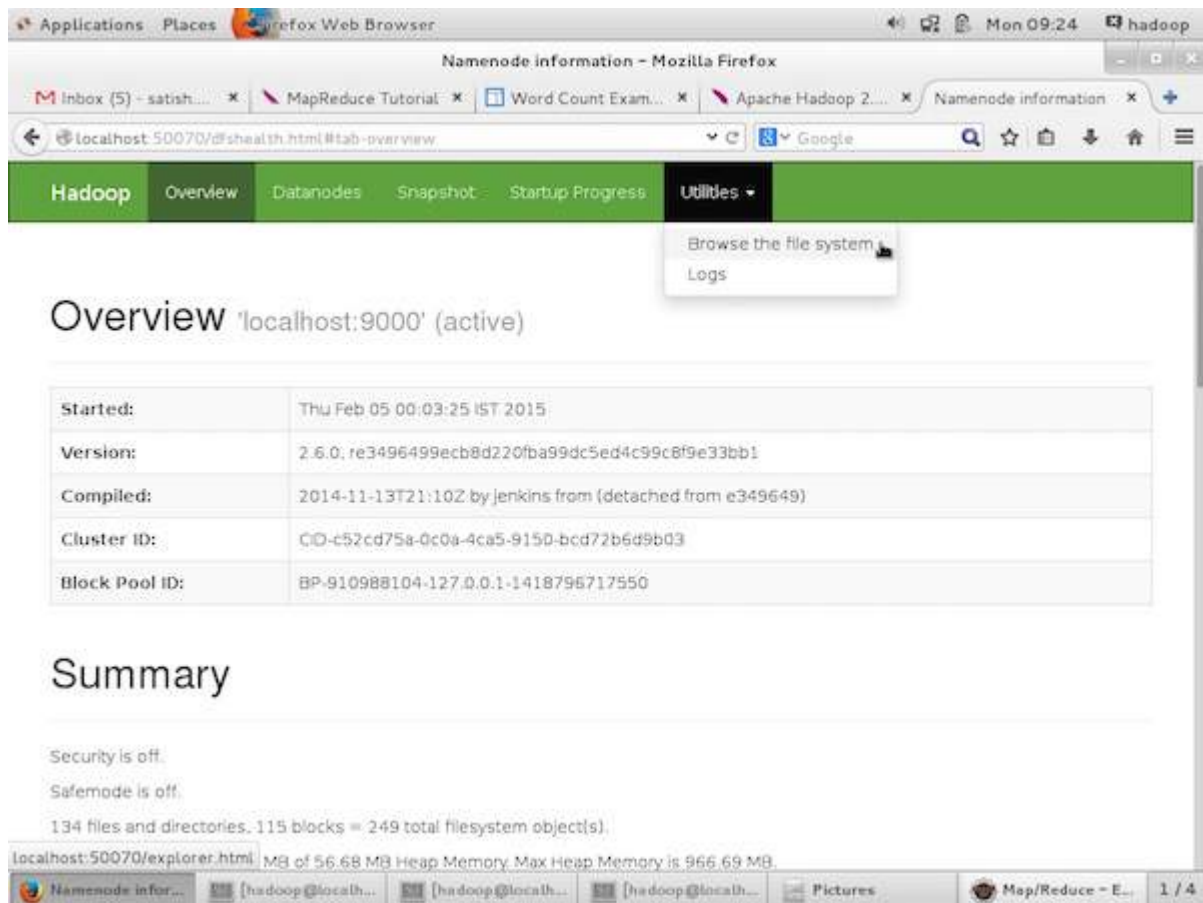
- HDFS administration includes monitoring the HDFS file structure, locations, and the updated files.

- MapReduce administration includes monitoring the list of applications, configuration of nodes, application status, etc.
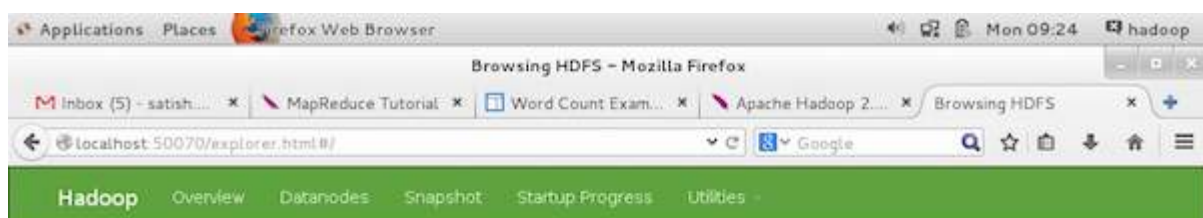
HDFS Monitoring

HDFS *HadoopDistributedFileSystem* contains the user directories, input files, and output files. Use the MapReduce commands, **put** and **get,** for storing and retrieving.

After starting the Hadoop framework *daemons* by passing the command "start-all.sh" on "/$HADOOP_HOME/sbin", pass the following URL to the browser "http://localhost:50070". You should see the following screen on your browser.

The following screenshot shows how to browse the browse HDFS.

The following screenshot show the file structure of HDFS. It shows the files in the "/user/hadoop" directory.

Browse Directory

/	Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	0	0 B	queryresult
drwx------	hadoop	supergroup	0 B	0	0 B	tmp
drwxr-xr-x	hadoop	supergroup	0 B	0	0 B	user

Hadoop, 2014.

The following screenshot shows the Datanode information in a cluster. Here you can find one node with its configurations and capacities.

Datanode Information

In operation

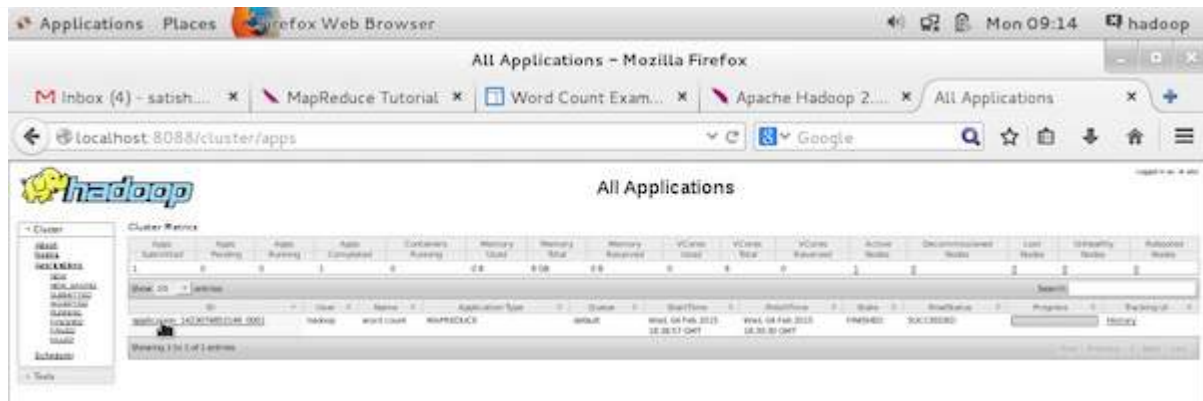| Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
localhost (127.0.0.1:50010)	0	In Service	38.16 GB	5.64 MB	1.31 GB	36.84 GB	115	5.64 MB (0.01%)	0	2.6.0

Decomissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction

MapReduce Job Monitoring

A MapReduce application is a collection of jobs *Mapjob, Combiner, Partitioner, andReducejob*. It is mandatory to monitor and maintain the following −

- Configuration of datanode where the application is suitable.
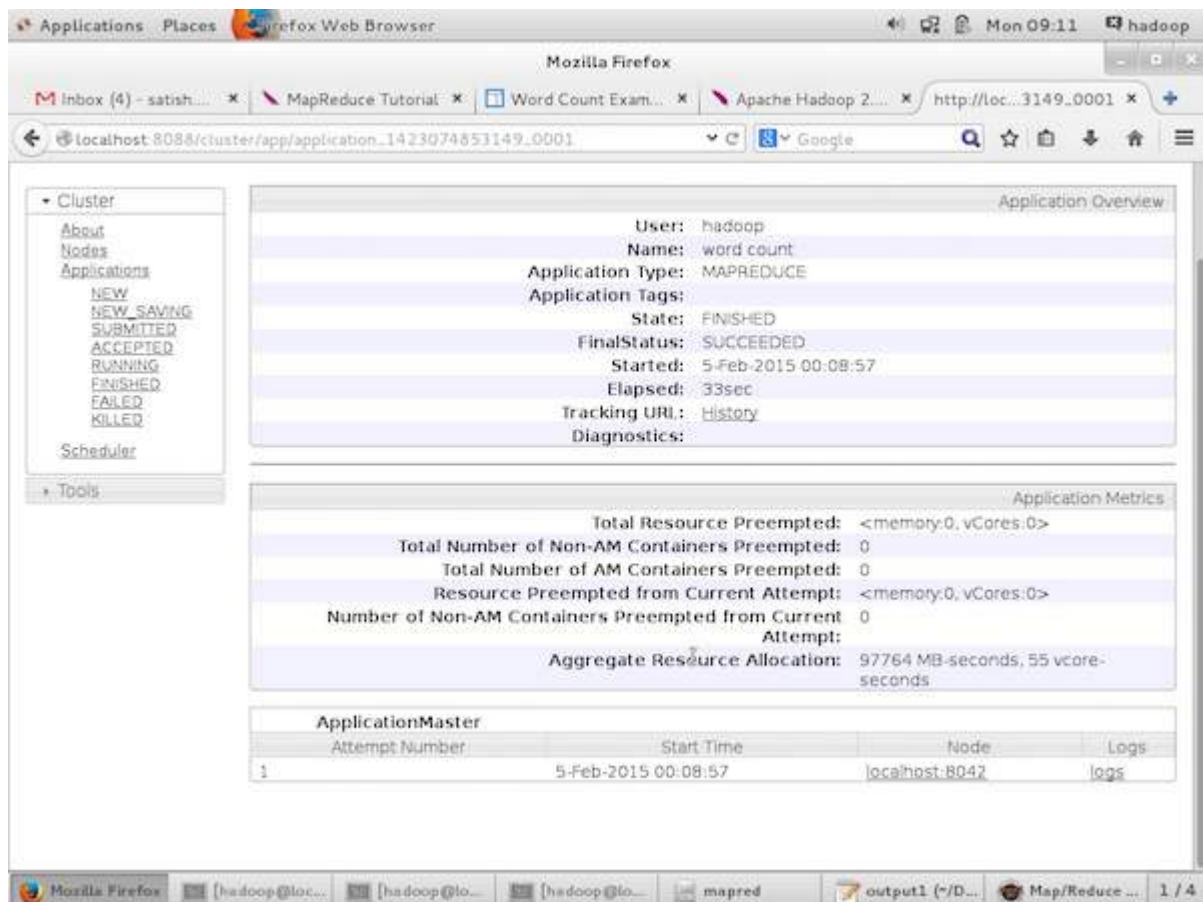- The number of datanodes and resources used per application.

To monitor all these things, it is imperative that we should have a user interface. After starting the Hadoop framework by passing the command "start-all.sh" on "/$HADOOP_HOME/sbin", pass the following URL to the browser "http://localhost:8080". You should see the following screen on your browser.

In the above screenshot, the hand pointer is on the application ID. Just click on it to find the following screen on your browser. It describes the following −

- On which user the current application is running

- The application name

- Type of that application

- Current status, Final status

- Application started time, elapsed $completedtime$, if it is complete at the time of monitoring

- The history of this application, i.e., log information

- And finally, the node information, i.e., the nodes that participated in running the application.

The following screenshot shows the details of a particular application −

The following screenshot describes the currently running nodes information. Here, the screenshot contains only one node. A hand pointer shows the localhost address of the running node.

Loading [MathJax]/jax/output/HTML-CSS/jax.js