



Name : Manav Pahilwani	Class/Roll No. : D16AD/37	Grade :
-------------------------------	----------------------------------	----------------

Title of Experiment : To study Hadoop Ecosystem and to demonstrate Basic Hadoop Commands.

Objective of Experiment : Acquire a foundational understanding of the Hadoop ecosystem and its components, focusing on basic Hadoop commands for effective data management and processing.

Outcome of Experiment : We successfully installed Hadoop Eco-system and executed basic Hadoop commands on it.

Problem Statement : Learn how to use Hadoop for managing and processing big data by mastering essential commands and concepts.

Description / Theory : Hadoop is an open-source software framework that is used for storing and processing large amounts of data in a distributed computing environment. It is designed to handle big data and is based on the MapReduce programming model, which allows for the parallel processing of large datasets. It is used for storing a large amount of data and performing the computation. Its framework is based on Java programming with some native code in C and shell scripts. Hadoop is an open-source software framework that is used for storing and processing large amounts of data in a distributed computing environment. It is designed to handle big data and is based on the MapReduce programming model, which allows for the parallel processing of large datasets.

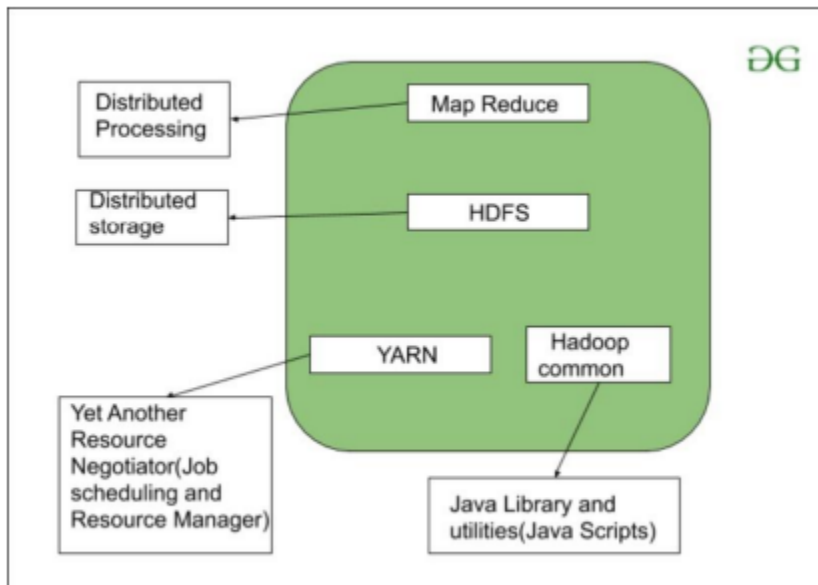
Hadoop has two main components: HDFS (Hadoop Distributed File System): This is the storage component of Hadoop, which allows for the storage of large amounts of data across multiple machines. It is designed to work with commodity hardware, which makes it cost-effective. YARN (Yet Another Resource Negotiator): This is



Subject/Odd Sem 2023-23/Experiment 1

the resource management component of Hadoop, which manages the allocation of resources (such as CPU and memory) for processing the data stored in HDFS. Hadoop also includes several additional modules that provide additional functionality, such as Hive (a SQL-like query language), Pig (a high-level platform for creating MapReduce programs), and HBase (a non-relational, distributed database). Hadoop is commonly used in big data scenarios such as data warehousing, business intelligence, and machine learning. It's also used for data processing, data analysis, and data mining. It enables the distributed processing of large data sets across clusters of computers using a simple programming model.

Hadoop Architecture:





Output:

```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ sudo jps
5739 JobHistoryServer
7467 HistoryServer
6905 RunJar
6077 ResourceManager
5562 SecondaryNameNode
7431 Bootstrap
8387
6355 HMaster
5430 NameNode
8117 Bootstrap
5841 NodeManager
6753 RunJar
8310 Bootstrap
5690 Bootstrap
7574 HRegionServer
6481 RESTServer
6616 ThriftServer
5241 DataNode
5171 QuorumPeerMain
5343 JournalNode
8341
9509 Jps
[cloudera@quickstart ~]$ pwd
```



Vivekanand Education Society's Institute of Technology

Approved by AICTE & Affiliated to University of Mumbai

Artificial Intelligence and Data Science Department

Subject/Odd Sem 2023-23/Experiment 1

```
Eclipse
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
8341
9509 Jps
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$ hadoop fs -mkdir /user/cloudera/RTDIR1
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop fs -mkdir RTDIR2

[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$ ls
cloudera-manager Downloads kerberos Pictures workspace
cm_api.py eclipse lib Public
Desktop enterprise-deployment.json Music Templates
Documents express-deployment.json parcels Videos
[cloudera@quickstart ~]$ hdfs dfs -mkdir ManavP1
[cloudera@quickstart ~]$ hadoop fs -ls
Found 3 items
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:19 ManavP1
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:17 RTDIR1
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:18 RTDIR2
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera
```



Vivekanand Education Society's Institute of Technology

Approved by AICTE & Affiliated to University of Mumbai

Artificial Intelligence and Data Science Department

Subject/Odd Sem 2023-23/Experiment 1

```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:18 RTDIR2
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera
Found 3 items
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:19 /user/cloudera/ManavP1
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:17 /user/cloudera/RTDIR1
drwxr-xr-x - cloudera cloudera 0 2023-10-15 05:18 /user/cloudera/RTDIR2
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2023-10-15 05:08 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2023-10-15 05:08 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir -p Manav1/Manav2/Manav3
[cloudera@quickstart ~]$ hdfs dfs -mkdir Data1 Data2 Data3
[cloudera@quickstart ~]$ hdfs dfs -touchz /user/cloudera/Manav1/answer.txt
[cloudera@quickstart ~]$ ls
cloudera-manager Downloads kerberos Pictures workspace
cm_api.py eclipse lib Public
Desktop enterprise-deployment.json Music Templates
```



Vivekanand Education Society's Institute of Technology

Approved by AICTE & Affiliated to University of Mumbai

Artificial Intelligence and Data Science Department

Subject/Odd Sem 2023-23/Experiment 1

Cloudera [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Sun Oct 15, 05:37 cloudera

Browsing HDFS - Mozilla Firefox

Browsing HDFS

quickstart.cloudera:50070/explorer.html#/user/cloudera

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	cloudera	0 B	Sun Oct 15 05:19:01 -0700 2023	0	0 B	ManavP1
drwxr-xr-x	cloudera	cloudera	0 B	Sun Oct 15 05:17:57 -0700 2023	0	0 B	RTDIR1
drwxr-xr-x	cloudera	cloudera	0 B	Sun Oct 15 05:18:19 -0700 2023	0	0 B	RTDIR2

Hadoop, 2017.

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Right Ctrl



Vivekanand Education Society's Institute of Technology

Approved by AICTE & Affiliated to University of Mumbai

Artificial Intelligence and Data Science Department

Subject/Odd Sem 2023-23/Experiment 1

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -ls /  
Found 6 items  
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks  
drwxr-xr-x - hbase supergroup 0 2023-10-15 05:08 /hbase  
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr  
drwxrwxrwt - hdfs supergroup 0 2023-10-15 05:08 /tmp  
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user  
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var  
[cloudera@quickstart ~]$ hdfs dfs -mkdir -p Manav1/Manav2/Manav3  
[cloudera@quickstart ~]$ hdfs dfs -mkdir Data1 Data2 Data3  
[cloudera@quickstart ~]$ hdfs dfs -touchz /user/cloudera/Manav1/answer.txt  
[cloudera@quickstart ~]$ ls  
cloudera-manager Downloads kerberos Pictures workspace  
cm_api.py eclipse lib Public  
Desktop enterprise-deployment.json Music Templates  
Documents express-deployment.json parcels Videos  
[cloudera@quickstart ~]$  
[cloudera@quickstart ~]$ cat >file1.txt  
  
^Z  
[1]+ Stopped cat > file1.txt  
[cloudera@quickstart ~]$ cd Desktop
```



Subject/Odd Sem 2023-23/Experiment 1

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
  
^Z  
[1]+  Stopped                  cat > file1.txt  
[cloudera@quickstart ~]$ cd Desktop  
[cloudera@quickstart Desktop]$ cat >file1.txt  
^Z  
[2]+  Stopped                  cat > file1.txt  
[cloudera@quickstart Desktop]$ cat file1.txt  
[cloudera@quickstart Desktop]$ ls  
Eclipse.desktop  Express.desktop  Kerberos.desktop  
Enterprise.desktop  file1.txt        Parcels.desktop  
[cloudera@quickstart Desktop]$ cd  
[cloudera@quickstart ~]$ hdfs dfs -du /user/cloudera/Manav1  
0  0  /user/cloudera/Manav1/Manav2  
0  0  /user/cloudera/Manav1/answer.txt  
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/file1.txt /user/cl  
oudera/Manav1  
[cloudera@quickstart ~]$ hdfs dfs -du /user/cloudera/Manav1  
0  0  /user/cloudera/Manav1/Manav2  
0  0  /user/cloudera/Manav1/answer.txt  
23 23 /user/cloudera/Manav1/file1.txt  
[cloudera@quickstart ~]$
```

Results and Discussions : We learnt about Hadoop's important features and structure. We practiced basic Hadoop commands, like making folders, moving files, and checking data. These actions showed us how Hadoop can handle big data tasks effectively.