| Name: Subrato Tapaswi	Class/Roll No.: D16AD/60	Grade:

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

Objective of Experiment: Explore the Hadoop Ecosystem and its components, and demonstrate Hadoop commands for effective data management and processing.

Outcome of Experiment : Successfully install Hadoop Ecosystem and execute basic Hadoop commands on it.

Problem Statement: Install Cloudera CDH and learn how to use Hadoop commands for managing and processing big data.
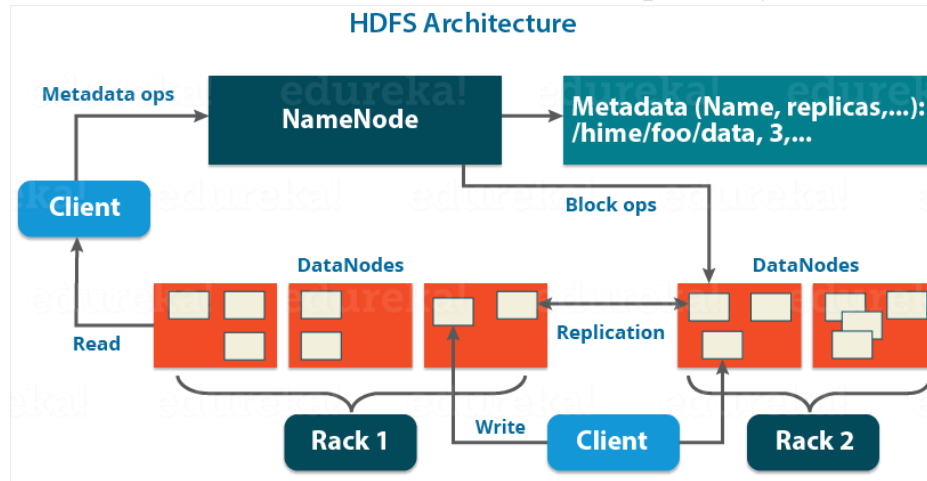
Theory:
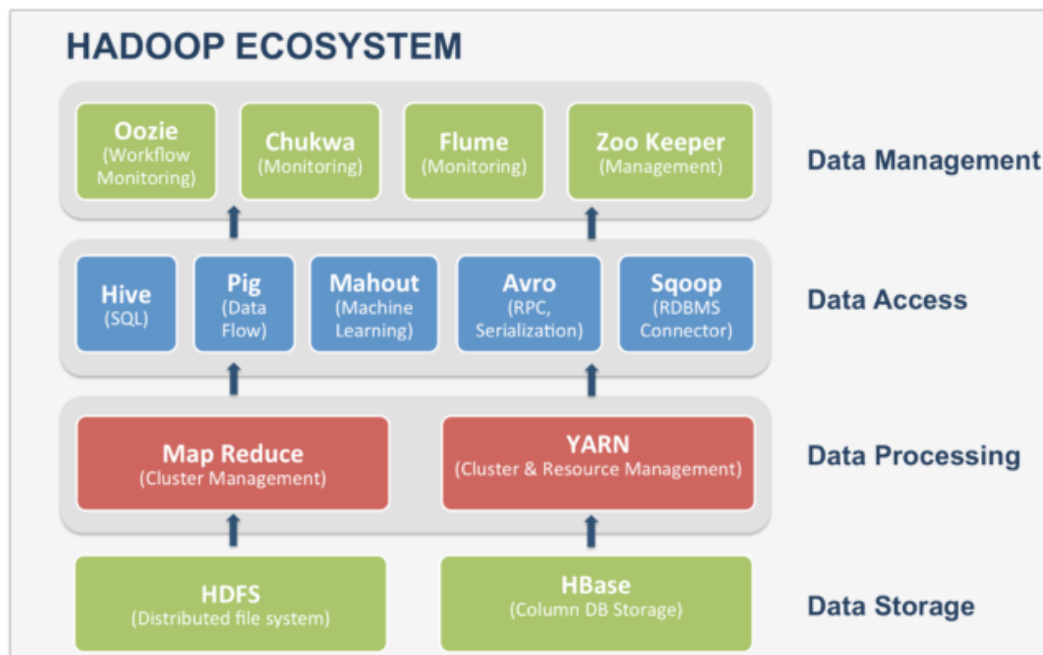
HDFS Architecture
Given below is the architecture of a Hadoop File System.

HDFS follows the master-slave architecture and it has the following elements.
The main components of HDFS are as described below:

1. NameNode is the master of the system. It maintains the name system i.e. the directories and files and manages the blocks which are present on the DataNodes.

2. DataNodes are the slaves which are deployed on each machine and provide the actual storage. They are responsible for serving read and write requests for the clients

3. Secondary NameNode is responsible for performing periodic checkpoints. In the event of NameNode failure, the NameNode can be restarted using the checkpoint

4. Block

Generally, the user data is stored in the files of HDFS. The file in a file system will be divided into one or more segments and/or stored in individual data nodes. These file segments are called blocks. In other words, the minimum amount of data that HDFS can read or write is called a Block. The default block size is 64MB, but it can be increased as per the need to change in HDFS configuration.

Goals of HDFS

● **Fault detection and recovery** : Since HDFS includes a large number of commodity hardware, failure of components is frequent. Therefore HDFS should have mechanisms for quick and automatic fault detection and recovery.

● **Huge datasets** : HDFS should have hundreds of nodes per cluster to manage the applications having huge datasets.

● **Hardware at data** : A requested task can be done efficiently, when the computation takes place near the data. Especially where huge datasets are involved, it reduces the network traffic and increases the throughput.

➢ **Output:**

1. Create a directory and list items in directory:

```
[cloudera@quickstart ~]$ hadoop fs -mkdir /Subrato
[cloudera@quickstart ~]$ hadoop fs -mkdir /Subrato/Exp1
[cloudera@quickstart ~]$ hadoop fs -ls /
Found 7 items
drwxr-xr-x   - cloudera supergroup          0 2023-10-11 20:07 /Subrato
drwxrwxrwx   - hdfs     supergroup          0 2017-07-19 05:34 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2023-10-02 08:26 /hbase
drwxr-xr-x   - solr     solr                0 2017-07-19 05:37 /solr
drwxrwxrwt   - hdfs     supergroup          0 2023-10-02 08:24 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-07-19 05:36 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-07-19 05:36 /var
```

2. Copy files to HDFS, Copy HDFS file to Local:

```
[cloudera@quickstart ~]$ hadoop fs -copyFromLocal /home/cloudera/Desktop/Subrat
.txt /Subrato/
copyFromLocal: `/home/cloudera/Desktop/Subrato.txt': No such file or directory
[cloudera@quickstart ~]$ hadoop fs -copyFromLocal /home/cloudera/Desktop/Subrat
.txt /Subrato/
[cloudera@quickstart ~]$ hadoop fs -copyToLocal /Subrato/Subrato.txt /home/clou
era/Desktop
copyToLocal: `/home/cloudera/Desktop/Subrato.txt': File exists
```

3. Move file from one location to another in HDFS:

```
[cloudera@quickstart ~]$ hadoop fs -mv /Subrato/Subrato.txt /Subrato/Exp1
[cloudera@quickstart ~]$ hadoop fs -ls /Subrato/
Found 1 items
drwxr-xr-x   - cloudera supergroup          0 2023-10-11 20:12 /Subrato/Exp1
[cloudera@quickstart ~]$ ▮
```

4. Remove file from specified location:

```
[cloudera@quickstart ~]$ hadoop fs -rm /Subrato/Exp1/Subrato.txt
Deleted /Subrato/Exp1/Subrato.txt
[cloudera@quickstart ~]$ ▮
```

5.   Show content of file stored in HDFS :

```
[cloudera@quickstart ~]$ hadoop fs -cat /Subrato/Subrato.txt
hello
how are vou
```

6.   Show last 10 lines of HDFS file :

hadoop fs -tail /Subrato/Subrato.txt

```
hello
how are you
this is 1st
experiment
bda is in
semester
seven
```

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