

### Experiment 3

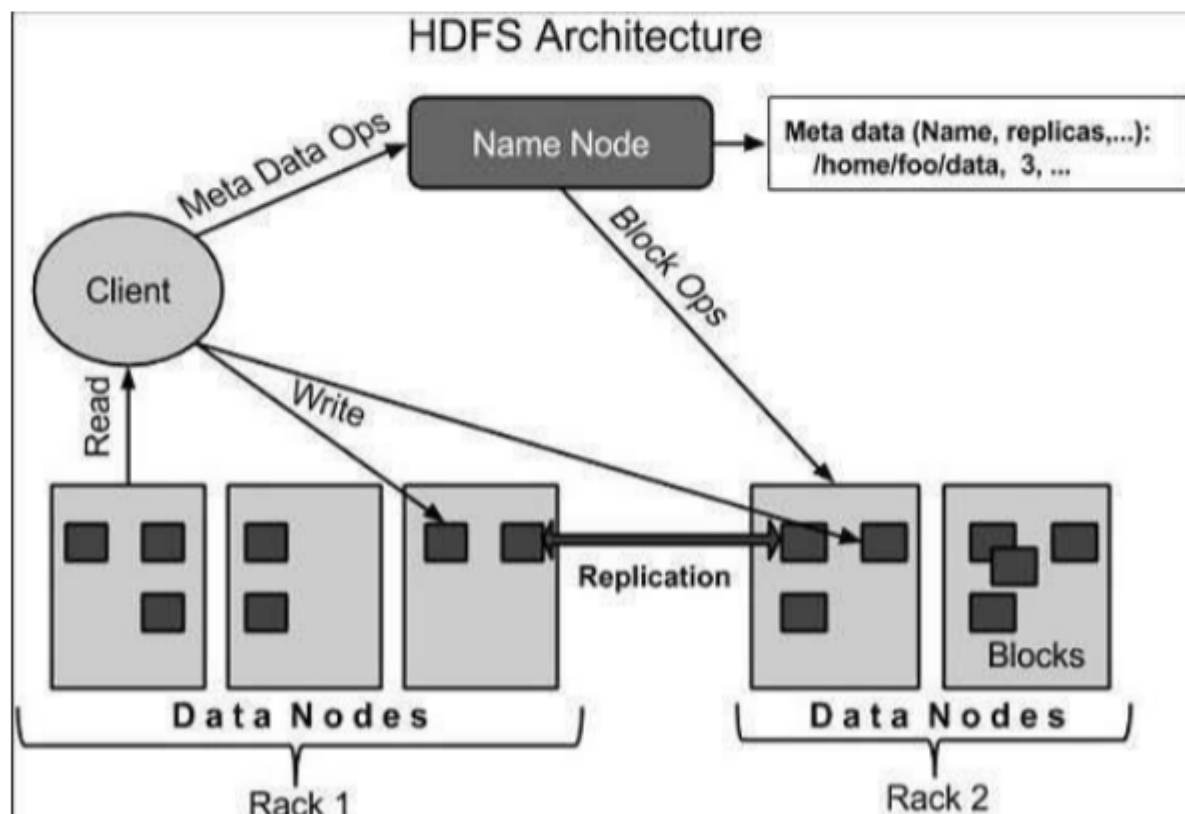
Aim : Execute different HDFS commands.

Theory :

Hadoop is a software framework that enables distributed storage and processing of large data sets. It consists of several open source projects, including HDFS, MapReduce, and Yarn. While Hadoop can be used for different purposes, the two most common are Big Data analytics and NoSQL database management. HDFS stands for "Hadoop Distributed File System" and is a decentralized file system that stores data across multiple computers in a cluster. This makes it ideal for large-scale storage as it distributes the load across multiple machines so there's less pressure on each individual machine. MapReduce is a programming model that allows users to write code once and execute it across many servers. When combined with HDFS, MapReduce can be used to process massive data sets in parallel by dividing work up into smaller chunks and executing them simultaneously.

#### HDFS Architecture

HDFS is an Open source component of the Apache Software Foundation that manages data. HDFS has scalability, availability, and replication as key features. Name nodes, secondary name nodes, data nodes, checkpoint nodes, backup nodes, and blocks all make up the architecture of HDFS. HDFS is faulttolerant and is replicated. Files are distributed across the cluster systems using the Name node and Data Nodes.



The primary difference between Hadoop and Apache HBase is that Apache HBase is a nonrelational database and Apache Hadoop is a non-relational data store





```

root@hadoop104:~# hdfs dfs -rm /user/arnold
rm: DEPRECATED: Please use 'rm -r' instead.
2016-02-14 04:22:42 INFO fs.TrashView$Info: Moved: 'hdfs://sandbox.hortonworks.com:8020/krone_copyed' to trash at: hdfs://sandbox.hortonworks.com:8020/user/trash/.Trash/Current/krone_copyed
root@hadoop104:~# hdfs dfs -rm -r /user/arnold
rm: /user/arnold: No such file or directory
root@hadoop104:~# hdfs dfs -ls /
Found 14 items
drwxr-xr-x   3 yarn          hadoop    0 2016-10-25 08:18 /app-logs
drwxr-xr-x   3 hdfs          hadoop    0 2016-10-25 07:54 /apps
drwxr-xr-x   3 yarn          hadoop    0 2016-10-25 07:00 /ats
drwxr-xr-x   3 hdfs          hadoop    0 2016-10-25 08:01 /dms
drwxr-xr-x   3 hdfs          hadoop    0 2016-10-25 07:00 /dms
drwxr-xr-x   3 root          hadoop    0 2014-02-14 04:21 /krone
drwxr-xr-x   3 mapred         hadoop    0 2016-10-25 07:00 /mapred
drwxr-xr-x   3 mapred         hadoop    0 2016-10-25 07:00 /mr-history
drwxr-xr-x   3 hdfs          hadoop    0 2016-10-25 07:47 /ranger
drwxr-xr-x   3 spark          hadoop    0 2014-02-14 04:25 /spark-history
drwxr-xr-x   3 spark          hadoop    0 2016-10-25 08:14 /spark2-history
drwxr-xr-x   3 hdfs          hadoop    0 2016-10-25 08:11 /sp
drwxr-xr-x   3 root          hadoop    0 2014-02-14 04:00 /tr
drwxr-xr-x   3 hdfs          hadoop    0 2014-02-14 04:22 /user
root@hadoop104:~# hdfs dfs -rm -r /
rm: Cannot move 'hdfs://sandbox.hortonworks.com:8020/' to the trash, as it contains the trash. Consider using -skiptrash option
root@hadoop104:~# hdfs dfs -rm -r / -skiptrash
rm: Cannot move 'hdfs://sandbox.hortonworks.com:8020/' to the trash, as it contains the trash. Consider using -skiptrash option
rm: '-skiptrash': No such file or directory
root@hadoop104:~# hdfs dfs -ls /
Found 1 items
drwxr-xr-x   3 rag_aps hdfs          5 2016-02-14 04:24 /a/test2.txt
root@hadoop104:~# hdfs dfs -rm -r /
rm: Cannot move 'hdfs://sandbox.hortonworks.com:8020/' to the trash, as it contains the trash. Consider using -skiptrash option
root@hadoop104:~# hdfs dfs -du /krone
0 /krone/283
0 /krone/hdfs_commands.txt
0 /krone/localDoc_283.txt
0 /krone/test2.txt
root@hadoop104:~# hdfs dfs -du /krone
rm: DEPRECATED: Please use 'du -s' instead.
0 /krone
root@hadoop104:~# hdfs dfs stat /krone
2014-02-14 04:22:50
root@hadoop104:~# hdfs dfs -setrep 3 /krone
Replication 4 set: /krone/hdfs_commands.txt
Replication 4 set: /krone/localDoc_283.txt
Replication 4 set: /krone/test2.txt

```

Conclusion : We have successfully execute different HDFS Commands