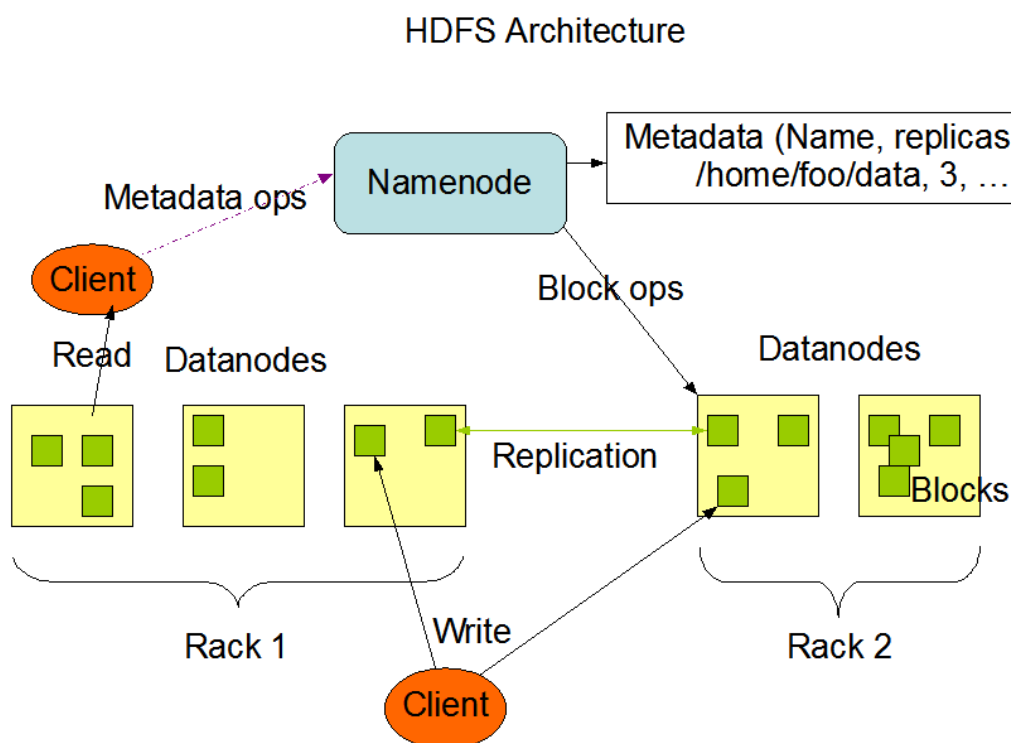


Program – 2

AIM: Implementing file management operations on HDFS 1) Adding files and Repositories, 2) Retrieving Files, 3) Deleting files.

About HDFS

Apache HDFS or Hadoop Distributed File System is a block-structured file system where each file is divided into blocks of a pre-determined size. These blocks are stored across a cluster of one or several machines. Apache Hadoop HDFS Architecture follows a Master/Slave Architecture, where a cluster comprises of a single NameNode (Master node) and all the other nodes are DataNodes (Slave nodes). HDFS can be deployed on a broad spectrum of machines that support Java. Though one can run several DataNodes on a single machine, but in the practical world, these DataNodes are spread across various machines.



NameNode is the master node in the Apache Hadoop HDFS Architecture that maintains and manages the blocks present on the DataNodes (slave nodes). NameNode is a very highly available server that manages the File System Namespace and controls access to files by clients. I will be discussing this High Availability feature of Apache Hadoop HDFS in my next blog. The HDFS architecture is built in such a way that the user data never resides on the NameNode. The data resides on DataNodes only.

DataNodes are the slave nodes in HDFS. Unlike NameNode, DataNode is a commodity hardware, that is, a non-expensive system which is not of high quality or high-availability. The DataNode is a block server that stores the data in the local file ext3 or ext4.

Program – 2

File Management in HDFS

1. Creating Directories

1	## Usage:
2	hadoop fs -mkdir <paths>
3	## Example:
4	hadoop fs -mkdir /user/saurzcode/dir1 /user/saurzcode/dir2

2. Adding files to HDFS

1	## Usage:
2	hadoop fs -put <localsrc> ... <HDFS_dest_Path>
3	## Example:
4	hadoop fs -put /home/saurzcode/Samplefile.txt /user/saurzcode/dir3/

3. Retrieving files

1	## Usage:
2	hadoop fs -get <hdfs_src> <localdst>
3	## Example:
4	hadoop fs -get /user/saurzcode/dir3/Samplefile.txt /home/

4. Removing files

1	## Usage :
2	hadoop fs -rm <arg>
3	## Example:
4	hadoop fs -rm /user/saurzcode/dir1/abc.txt

5. Removing directories

1	## Usage :
2	hadoop fs -rmr <arg>
3	## Example:
4	hadoop fs -rmr /user/saurzcode/

Findings and Learnings:

1. We studied file management and handling on the Hadoop DFS.