JUNAID GIRKAR

60004190057

TE COMPS A4

**EXPERIMENT - 2**

**AIM**: Execution of HDFS commands

**THEORY**:

### **What is DFS?**

DFS stands for the distributed file system, it is a concept of storing the file in multiple nodes in a distributed manner. DFS actually provides the Abstraction for a single large system whose storage is equal to the sum of storage of other nodes in a cluster.

HDFS(Hadoop Distributed File System) is utilised for storage permission as a Hadoop cluster. It is mainly designed for working on commodity Hardware devices(devices that are inexpensive), working on a distributed file system design. HDFS is designed in such a way that it believes more in storing the data in a large chunk of blocks rather than storing small data blocks. HDFS in Hadoop provides Fault-tolerance and High availability to the storage layer and the other devices present in that Hadoop cluster.

HDFS is capable of handling larger size data with high volume velocity and variety makes Hadoop work more efficient and reliable with easy access to all its components. HDFS stores the data in the form of the block where the size of each data block is 128MB in size which is configurable means you can change it according to your requirement in the hdfs-site.xml file in your Hadoop directory.

### Some Important Features of HDFS(Hadoop Distributed File System)

* It’s easy to access the files stored in HDFS.
* HDFS also provides high availability and fault tolerance.
* Provides scalability to scaleup or scaledown nodes as per our requirement.
* Data is stored in a distributed manner i.e. various Datanodes are responsible for storing the data.
* HDFS provides Replication because of which there is no fear of Data Loss.
* HDFS Provides High Reliability as it can store data in a large range of Petabytes.
* HDFS has in-built servers in Name node and Data Node that helps them to easily retrieve the cluster information.
* Provides high throughput.

### HDFS Storage Daemon’s

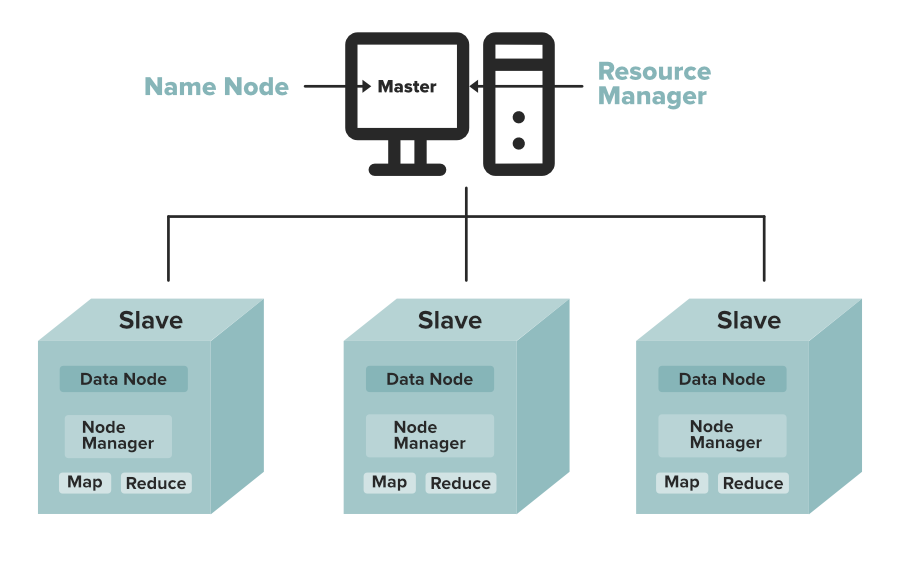
As we all know Hadoop works on the MapReduce algorithm which is a master-slave architecture, HDFS has NameNode and DataNode that works in the similar pattern.

**1. NameNode**: NameNode works as a Master in a Hadoop cluster that Guides the Datanode(Slaves). Namenode is mainly used for storing the Metadata i.e. nothing but the data about the data. Meta Data can be the transaction logs that keep track of the user’s activity in a Hadoop cluster.

Meta Data can also be the name of the file, size, and the information about the location(Block number, Block ids) of Datanode that Namenode stores to find the closest DataNode for Faster Communication. Namenode instructs the DataNodes with the operation like delete, create, Replicate, etc.

As our NameNode is working as a Master it should have a high RAM or Processing power in order to Maintain or Guide all the slaves in a Hadoop cluster. Namenode receives heartbeat signals and block reports from all the slaves i.e. DataNodes.

**2.** **DataNode**: DataNodes works as a Slave DataNodes are mainly utilized for storing the data in a Hadoop cluster, the number of DataNodes can be from 1 to 500 or even more than that, the more number of DataNode your Hadoop cluster has More Data can be stored. so it is advised that the DataNode should have High storing capacity to store a large number of file blocks. Datanode performs operations like creation, deletion, etc. according to the instruction provided by the NameNode.



**IMPLEMENTATION**:

| [training@localhost ~]$ pwd |
| --- |

| /home/training |
| --- |

| [training@localhost ~]$ ls |
| --- |

| 201212daily.txt KMeans$Map.class sample5.txt aaa.txt KMeans$Reduce.class sample6.txt abcde km.jar sample-bloomfilter a.txt Kpack1 .java sample-filter backgrounds Kpack1.java sorrt.jar bb.txt Kpack1.java~ sortData BloomFilter.java kriti sortData~ bloom.jar lib SortData book ma SortData.class bookinfo ma.jar SortData.java bookinfo~ mat sorting Books mat1 sorting1.jar b.txt mat1.jar sort.jar build mat2.jar SortMapper Centroid mat3.jar SortMapper.class Centroid~ mat4.jar SortMapper.java centroids.txt mat5.jar SortReducer centroid.txt mat6.jar SortReducer.class Clustering1.java mat.jar SortReducer.java Clustering1.java~ mat.jar-C stored cust\_file\_class\_2009 matrcies1.jar stored~ Data matrices student1.java data1\_2009 matrices1.jar student.java data1\_2009~ matrix t data2\_2009 Matrix.class table1.txt data2\_2009~ matrix.jar table1.txt~ data3\_2009~ matrix.java table2.txt data4\_2009~ #Matrix.java# table2.txt~ data.txt Matrix.java temp data.txt~ Matrix.java~ Templates deptinfo Matrix$Map.class TempStatsStore derby.log matrixmuln test Desktop Matrixmulti test1.txt dir6 Matrixmulti.java test1.txt~ dirbook Matrixmulti.java~ test.jar djs Matrix$Reduce.class textfile2.txt djs.jar Matrix.txt textfile3 djss matrixx.jar textfile4 eclipse matt textfile.txt emp mattt training\_materials emp~ MRDemo.jar tushar emp2 new file tushar1 emp2~ new file~ tushar1.jar emp2.java new file 1 tushar.jar emp3.java new file 2 untitled folder emp.java new file 2~ w emp\_name order\_custid WeatherData example~ pig\_1443861414011.log w.jar file1.txt pig\_1443947154908.log word file1.txt~ pig\_1443949486150.log Word1.java filetext4 pig\_1444470259655.log word2.jar FilterJob.java pig\_1444547817901.log WordCount12.java FilterMapper.java pig\_1448259294240.log wordcount1.jar gender pig\_1448278570362.log WordCount.class hadoop pig\_1448684754725.log wordcount\_classes HadoopPoem1.txt pig\_1448685352126.log wordcount.java HadoopPoem2.txt poem912 WordCount.java inner2.txt posts.txt WordCount.java~ inner.txt product\_file\_class\_2009 WordCount.java~~ input proto\_file1.txt WordCount$MapClass.class java proto\_file2.txt WordCount$Reduce.class kmeans proto\_file3.txt word.jar kmeans1.jar proto\_file4.txt wordtext KMeans.class qqq workspace kmeans.jar QueryResult.java w.txt "KMeans.java" sample w.txt~ "KMeans.java"~ sample~ www.java KMeans.java sample1 wwww.jar KMeans.java~ sample1~ |
| --- |

| [training@localhost ~]$ hadoop dfs -ls |
| --- |

| Found 109 items -rw-r--r-- 1 training supergroup 1390 2015-10-02 20:20 /user/training/Books drwxr-xr-x - training supergroup 0 2014-08-17 03:50 /user/training/WeatherData -rw-r--r-- 1 training supergroup 3103 2016-02-16 21:54 /user/training/WordCount.java drwxr-xr-x - training supergroup 0 2015-10-17 02:34 /user/training/\_sqoop -rw-r--r-- 1 training supergroup 0 2018-02-01 23:46 /user/training/aaa.txt drwxr-xr-x - training supergroup 0 2018-02-02 00:45 /user/training/abcdef -rw-r--r-- 1 training supergroup 12 2020-09-25 00:58 /user/training/asd -rw-r--r-- 1 training supergroup 18 2018-02-01 23:52 /user/training/bb.txt drwxr-xr-x - training supergroup 0 2016-03-09 22:55 /user/training/bbb11451728037137 drwxr-xr-x - training supergroup 0 2016-03-10 01:05 /user/training/bbbb17422138347695 drwxr-xr-x - training supergroup 0 2016-03-10 01:06 /user/training/bbbb17569614160750 drwxr-xr-x - training supergroup 0 2016-03-10 01:15 /user/training/bbbb18073723471225 -rw-r--r-- 1 training supergroup 2944 2015-09-26 02:37 /user/training/bookinfo drwxr-xr-x - training supergroup 0 2015-09-20 01:38 /user/training/class2009\_dir1 -rw-r--r-- 1 training supergroup 81 2015-09-18 21:05 /user/training/deptinfo drwxr-xr-x - training supergroup 0 2014-08-17 01:59 /user/training/dir1 drwxr-xr-x - training supergroup 0 2015-09-12 02:42 /user/training/dir10 drwxr-xr-x - training supergroup 0 2015-09-19 02:00 /user/training/dir100 drwxr-xr-x - training supergroup 0 2015-09-19 02:42 /user/training/dir1000 drwxr-xr-x - training supergroup 0 2014-08-22 02:30 /user/training/dir11 drwxr-xr-x - training supergroup 0 2014-08-22 02:35 /user/training/dir12 drwxr-xr-x - training supergroup 0 2014-08-17 01:59 /user/training/dir2 drwxr-xr-x - training supergroup 0 2014-08-17 01:59 /user/training/dir3 drwxr-xr-x - training supergroup 0 2014-08-17 01:59 /user/training/dir4 drwxr-xr-x - training supergroup 0 2014-08-17 01:59 /user/training/dir5 drwxr-xr-x - training supergroup 0 2014-08-17 03:03 /user/training/dir6 drwxr-xr-x - training supergroup 0 2014-08-17 03:04 /user/training/dir7 drwxr-xr-x - training supergroup 0 2014-08-17 04:22 /user/training/dir9 -rw-r--r-- 1 training supergroup 1390 2015-10-02 20:22 /user/training/dirbook drwxr-xr-x - training supergroup 0 2021-10-05 01:06 /user/training/djsin drwxr-xr-x - training supergroup 0 2021-10-05 01:06 /user/training/djsout drwxr-xr-x - training supergroup 0 2021-10-05 01:16 /user/training/djsout1 drwxr-xr-x - training supergroup 0 2021-10-05 01:12 /user/training/djss drwxr-xr-x - training supergroup 0 2015-10-17 02:14 /user/training/emp drwxr-xr-x - training supergroup 0 2015-10-17 02:34 /user/training/emp2 drwxr-xr-x - training supergroup 0 2015-10-10 23:18 /user/training/emp\_dir drwxr-xr-x - training supergroup 0 2015-10-17 02:26 /user/training/employee drwxr-xr-x - training supergroup 0 2015-10-17 23:39 /user/training/employee2 drwxr-xr-x - training supergroup 0 2015-10-17 23:42 /user/training/employee3 drwxr-xr-x - training supergroup 0 2015-11-27 20:53 /user/training/hadoop drwxr-xr-x - training supergroup 0 2021-07-30 02:17 /user/training/hadoopinword drwxr-xr-x - training supergroup 0 2021-07-30 02:12 /user/training/hadoopword drwxr-xr-x - training supergroup 0 2021-07-30 02:21 /user/training/hadoopword1 drwxr-xr-x - training supergroup 0 2016-02-16 02:30 /user/training/imput1234 -rw-r--r-- 1 root supergroup 13 2017-03-10 23:22 /user/training/inword -rw-r--r-- 1 training supergroup 69 2015-10-10 02:35 /user/training/join -rw-r--r-- 1 training supergroup 105 2015-10-10 02:38 /user/training/join2 drwxr-xr-x - training supergroup 0 2016-03-09 22:47 /user/training/kmeansinput drwxr-xr-x - training supergroup 0 2016-03-09 22:42 /user/training/kmeansoutput drwxr-xr-x - training supergroup 0 2016-03-09 22:48 /user/training/kmeansoutput10970575692823 drwxr-xr-x - training supergroup 0 2016-03-08 22:12 /user/training/kminput drwxr-xr-x - training supergroup 0 2016-03-08 22:13 /user/training/kmoutput -rw-r--r-- 1 root supergroup 235 2017-03-10 23:44 /user/training/ma drwxr-xr-x - training supergroup 0 2021-10-12 00:14 /user/training/mat1in drwxr-xr-x - training supergroup 0 2021-10-12 00:12 /user/training/mat1out drwxr-xr-x - training supergroup 0 2021-10-12 00:41 /user/training/mat1out1 drwxr-xr-x - training supergroup 0 2021-10-13 02:35 /user/training/matricesin drwxr-xr-x - training supergroup 0 2021-10-13 02:33 /user/training/matricesout drwxr-xr-x - training supergroup 0 2021-10-13 02:37 /user/training/matricesout1 drwxr-xr-x - training supergroup 0 2016-02-23 00:16 /user/training/matrixinput drwxr-xr-x - training supergroup 0 2016-02-23 00:17 /user/training/matrixoutput drwxr-xr-x - training supergroup 0 2021-10-12 01:09 /user/training/mattin drwxr-xr-x - training supergroup 0 2021-10-12 01:05 /user/training/mattout drwxr-xr-x - training supergroup 0 2021-10-12 01:11 /user/training/mattout1 drwxr-xr-x - training supergroup 0 2021-10-12 01:21 /user/training/matttin drwxr-xr-x - training supergroup 0 2021-10-12 01:20 /user/training/matttout drwxr-xr-x - training supergroup 0 2021-10-12 01:22 /user/training/matttout1 -rw-r--r-- 1 training supergroup 235 2017-03-17 01:44 /user/training/mm drwxr-xr-x - training supergroup 0 2020-09-25 02:00 /user/training/out1 drwxr-xr-x - training supergroup 0 2021-10-13 03:18 /user/training/out25 drwxr-xr-x - training supergroup 0 2016-02-16 02:32 /user/training/output1234 drwxr-xr-x - training supergroup 0 2017-03-10 23:27 /user/training/oword2 -rw-r--r-- 1 training supergroup 16 2015-10-04 01:23 /user/training/pig -rw-r--r-- 1 training supergroup 32 2015-10-04 02:04 /user/training/pig1 -rw-r--r-- 1 training supergroup 90 2015-11-23 03:12 /user/training/poem -rw-r--r-- 1 training supergroup 90 2015-09-12 02:41 /user/training/poem912 drwxr-xr-x - training supergroup 0 2021-10-13 03:02 /user/training/sortingin drwxr-xr-x - training supergroup 0 2021-10-13 03:00 /user/training/sortingout drwxr-xr-x - training supergroup 0 2021-10-13 03:05 /user/training/sortingout1 drwxr-xr-x - training supergroup 0 2021-10-13 03:08 /user/training/sortingout2 drwxr-xr-x - training supergroup 0 2021-10-13 03:12 /user/training/sortingout5 drwxr-xr-x - training supergroup 0 2016-02-22 22:34 /user/training/sortinput drwxr-xr-x - training supergroup 0 2021-10-13 03:17 /user/training/sortinputf drwxr-xr-x - training supergroup 0 2021-08-28 00:36 /user/training/sortout drwxr-xr-x - training supergroup 0 2016-02-22 22:34 /user/training/sortoutput drwxr-xr-x - training supergroup 0 2016-02-22 22:35 /user/training/sortoutput2 drwxr-xr-x - training supergroup 0 2017-03-03 21:50 /user/training/sortoutputtt drwxr-xr-x - training supergroup 0 2021-08-28 01:04 /user/training/ss1 drwxr-xr-x - training supergroup 0 2021-08-28 01:03 /user/training/sss drwxr-xr-x - training supergroup 0 2015-10-10 03:53 /user/training/str drwxr-xr-x - training supergroup 0 2015-10-17 01:26 /user/training/stud drwxr-xr-x - training supergroup 0 2015-10-17 01:42 /user/training/stud1 drwxr-xr-x - training supergroup 0 2015-10-17 01:43 /user/training/stud2 drwxr-xr-x - training supergroup 0 2015-10-17 02:01 /user/training/student -rw-r--r-- 1 training supergroup 86 2015-10-10 03:40 /user/training/table2 drwxr-xr-x - training supergroup 0 2016-02-05 19:51 /user/training/test drwxr-xr-x - training supergroup 0 2021-10-05 01:48 /user/training/test1 -rw-r--r-- 1 training supergroup 29 2021-10-05 01:54 /user/training/test2 drwxr-xr-x - training supergroup 0 2021-10-05 01:51 /user/training/testout -rw-r--r-- 1 training supergroup 24 2016-02-12 00:40 /user/training/textfile.txt drwxr-xr-x - training supergroup 0 2020-11-07 00:08 /user/training/tushar1input drwxr-xr-x - training supergroup 0 2020-11-07 00:13 /user/training/tushar1output drwxr-xr-x - training supergroup 0 2020-11-07 00:24 /user/training/tushar1output9 drwxr-xr-x - training supergroup 0 2020-11-06 22:17 /user/training/tusharinput drwxr-xr-x - training supergroup 0 2020-11-06 22:12 /user/training/tusharoutput drwxr-xr-x - training supergroup 0 2020-11-06 22:14 /user/training/tusharoutput1 drwxr-xr-x - training supergroup 0 2020-11-06 22:17 /user/training/tusharoutput2 drwxr-xr-x - training supergroup 0 2015-09-12 02:33 /user/training/user -rw-r--r-- 1 training supergroup 59 2015-10-10 00:56 /user/training/wordcount |
| --- |

| [training@localhost ~]$ hadoop dfs -mkdir hadoop [training@localhost ~]$ hadoop dfs -cp /user/training/WordCount.java /user/training/hadoop [training@localhost ~]$ hadoop dfs -touchz rem.txt [training@localhost ~]$ hadoop dfs -mv /user/training/rem.txt /user/training/hadoop [training@localhost ~]$ hadoop dfs -mkdir hadoop2 [training@localhost ~]$ hadoop dfs -cp /user/training/hadoop/rem.text /user/training/hadoop2 |
| --- |

| cp: File does not exist: /user/training/hadoop/rem.text |
| --- |

| [training@localhost ~]$ hadoop dfs -cp /user/training/hadoop/rem.txt /user/training/hadoop2 [training@localhost ~]$ hadoop dfs -rm  [training@localhost ~]$ hadoop dfs -rm /user/training/hadoop2/rem.txt |
| --- |

| Deleted hdfs://localhost/user/training/hadoop2/rem.txt |
| --- |

| [training@localhost ~]$ hadoop dfs -rm -rf /user/training/hadoop2 |
| --- |

| rm: cannot remove -rf: No such file or directory. rm: Cannot remove directory "hdfs://localhost/user/training/hadoop2", use -rmr instead |
| --- |

| [training@localhost ~]$ hadoop dfs -rmr /user/training/hadoop2 |
| --- |

| Deleted hdfs://localhost/user/training/hadoop2 |
| --- |

| [training@localhost ~]$ hadoop dfs -cat /user/training/bb.txt |
| --- |

| jhhhl kjhkjh laks |
| --- |

[training@localhost ~]$ hadoop dfs -cp /user/training/WordCount.java

/user/training/hadoop

[training@localhost ~]$ hadoop dfs -mv /user/training/WordCount.java

| /user/training/hadoop mv: Failed to rename hdfs://localhost/user/training/WordCount.java to /user/training/hadoop |
| --- |

[training@localhost ~]$ hadoop dfs -copyFromLocal /home/training/aaa.txt /user/training/hadoop

| [training@localhost ~]$ hadoop dfs -copyToLocal /user/training/hadoop/WordCount.java /home/training |
| --- |

| copyToLocal: Target /home/training/WordCount.java already exists |
| --- |

| [training@localhost ~]$ hadoop dfs -moveToLocal /user/training/hadoop/WordCount.java /home/training |
| --- |

| Option '-moveToLocal' is not implemented yet. |
| --- |

| [training@localhost ~]$ hadoop dfs -moveFromLocal /user/training/hadoop/WordCount.java /home/training |
| --- |

moveFromLocal: File /user/training/hadoop/WordCount.java does not exist.

| [training@localhost ~]$ hadoop dfs -moveFromLocal /home/training/bb.txt /user/training/hadoop |
| --- |

| [training@localhost ~]$ hadoop dfs -copyToLocal /user/training/hadoop/WordCount.java /home/training |
| --- |

| copyToLocal: Target /home/training/WordCount.java already exists |
| --- |

| [training@localhost ~]$ hadoop dfs -mkdir hadoop2 [training@localhost ~]$ hadoop dfs -copyFromLocal /home/training/aaa.txt /user/training/hadoop2 |
| --- |

| [training@localhost ~]$ hadoop dfs -rm r |
| --- |

| rm: cannot remove r: No such file or directory. |
| --- |

| [training@localhost ~]$ hadoop dfs -rmr hadoop2 |
| --- |

| Deleted hdfs://localhost/user/training/hadoop2 |
| --- |

| [training@localhost ~]$ hadoop dfs -touchz abc.txt [training@localhost ~]$ hadoop dfs -put /home/training/a.txt /user/training [training@localhost ~]$ hadoop dfs -rmr abc.txt |
| --- |

| Deleted hdfs://localhost/user/training/abc.txt |
| --- |

| [training@localhost ~]$ hadoop dfs -rm aaa.txt |
| --- |

| Deleted hdfs://localhost/user/training/aaa.txt |
| --- |

**CONCLUSION:** We learnt about DFS and HDFS. We then successfully implemented some HDFS commands.