Student Name	Imaad Imran Hajwane
SRN No	202101132
Roll No	23
Program	Computer Engineering
Year	Last Year
Division	A
Subject	BDA
Assignment No	2

1. mkdir:

This is no different from the UNIX **mkdir** command and is used to create a directory on an HDFS environment.

2. ls:

This is no different from the UNIX ls command and it is used for listing the directories present under a specific directory in an HDFS system.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -mkdir /test1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-08-01 16:31 /test1
```

3. put:

This command is used to copy files from the local file system to the HDFS filesystem. This command is similar to **–copyFromLocal** command.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -put ls_help.txt /test1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test1
Found 1 items
-rw-r--r--   1 hadoop supergroup       2945 2024-08-01 16:34 /test1/ls_help.txt
```

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -mkdir /test2
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - hadoop supergroup          0 2024-08-01 16:34 /test1
drwxr-xr-x   - hadoop supergroup          0 2024-08-01 17:58 /test2
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -copyFromLocal ls_help.txt /test1
copyFromLocal: `/test1/ls_help.txt': File exists
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -copyFromLocal ls_help.txt /test2
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test2
Found 1 items
Show Apps      1 hadoop supergroup       2945 2024-08-01 17:59 /test2/ls_help.txt
hadoop@iamincognito-VirtualBox:~$
```

```
hadoop@iamincognito-VirtualBox:~$ ls
Desktop    Downloads  hadoop-3.3.6.tar.gz  hadoopdata  Pictures  snap       Videos
Documents  hadoop     hadoop-3.4.0.tar.gz  Music       Public    Templates
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -copyToLocal /test2/ls_help.txt
hadoop@iamincognito-VirtualBox:~$ ls
Desktop    Downloads  hadoop-3.3.6.tar.gz  hadoopdata   Music     Public  Templates
Documents  hadoop     hadoop-3.4.0.tar.gz  ls_help.txt  Pictures  snap    Videos
hadoop@iamincognito-VirtualBox:~$
```

4. get:

This command is used to copy files from HDFS file system to the local file system, just the opposite to put command.

```
hadoop@iamincognito-VirtualBox:~$ ls
Desktop     Downloads  hadoop-3.3.6.tar.gz  hadoopdata  Pictures  snap       Videos
Documents   hadoop     hadoop-3.4.0.tar.gz  Music       Public    Templates
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -get /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$ ls
Desktop     Downloads  hadoop-3.3.6.tar.gz  hadoopdata   Music     Public  Templates
Documents   hadoop     hadoop-3.4.0.tar.gz  ls_help.txt  Pictures  snap    Videos
hadoop@iamincognito-VirtualBox:~$
```

5. cat:

This command is similar to the **UNIX cat command** and is used for displaying the contents of a file on the console.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -cat /test1/ls_help.txt
ls/: Unknown command
Usage: hadoop fs [generic options]
        [-appendToFile [-n] <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum [-v] <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-concat <target path> <src path> <src path> ...]
        [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
        [-copyToLocal [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <locald
st>]
        [-count [-q] [-h] [-v] [-t [<storage type>]] [-u] [-x] [-e] [-s] <path> ...]
        [-cp [-f] [-p | -p[topax]] [-d] [-t <thread count>] [-q <thread pool queue size>] <src> ... <dst>]
        [-createSnapshot <snapshotDir> [<snapshotName>]]
        [-deleteSnapshot <snapshotDir> <snapshotName>]
        [-df [-h] [<path> ...]]
        [-du [-s] [-h] [-v] [-x] <path> ...]
        [-expunge [-immediate] [-fs <path>]]
        [-find <path> ... <expression> ...]
        [-get [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
        [-getfacl [-R] <path>]
        [-getfattr [-R] {-n name | -d} [-e en] <path>]
        [-getmerge [-nl] [-skip-empty-file] <src> <localdst>]
        [-head <file>]
        [-help [cmd ...]]
        [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]
        [-mkdir [-p] <path> ...]
        [-moveFromLocal [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
        [-moveToLocal <src> <localdst>]
        [-mv <src> ... <dst>]
        [-put [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
        [-renameSnapshot <snapshotDir> <oldName> <newName>]
        [-rm [-f] [-r|-R] [-skipTrash] [-safely] <src> ...]
        [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
        [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]|[--set <acl_spec> <path>]]
        [-setfattr {-n name [-v value] | -x name} <path>]
        [-setrep [-R] [-w] <rep> <path> ...]
        [-stat [format] <path> ...]
        [-tail [-f] [-s <sleep interval>] <file>]
        [-test -[defswrz] <path>]
        [-text [-ignoreCrc] <src> ...]
        [-touch [-a] [-m] [-t TIMESTAMP (yyyyMMdd:HHmmss) ] [-c] <path> ...]
        [-touchz <path> ...]
        [-truncate [-w] <length> <path> ...]
        [-usage [cmd ...]]

Generic options supported are:
-conf <configuration file>        specify an application configuration file
-D <property=value>               define a value for a given property
-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from confi
gurations.
-jt <local|resourcemanager:port>  specify a ResourceManager
-files <file1,...>                specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>               specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>          specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]
```

6. cp:

This command is similar to the **UNIX cp command**, and it is used for copying files from one directory to another directory within the HDFS file system.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -copyFromLocal hadooptest.txt /test2
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -cp test2/hadooptest.txt test1/
cp: `test1/': No such file or directory: `hdfs://localhost:9000/user/hadoop/test1'
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -cp test2/hadooptest.txt /test1
cp: `test2/hadooptest.txt': No such file or directory
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -cp /test2/hadooptest.txt /test1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test1
Found 2 items
-rw-r--r--   1 hadoop supergroup       1168 2024-08-01 19:53 /test1/hadooptest.txt
-rw-r--r--   1 hadoop supergroup       2945 2024-08-01 16:34 /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$
```

7. mv:

This command is similar to the **UNIX mv command**, and it is used for moving a file from one directory to another directory within the HDFS file system.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test1
Found 1 items
-rw-r--r--   1 hadoop supergroup       2945 2024-08-01 16:34 /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -mv /test2/hadooptest.txt /test1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test1
Found 2 items
-rw-r--r--   1 hadoop supergroup       1168 2024-08-01 19:50 /test1/hadooptest.txt
-rw-r--r--   1 hadoop supergroup       2945 2024-08-01 16:34 /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test2
Found 1 items
-rw-r--r--   1 hadoop supergroup       2945 2024-08-01 17:59 /test2/ls_help.txt
hadoop@iamincognito-VirtualBox:~$
```

8. rm:

This command is similar to the **UNIX rm command**, and it is used for removing a file from the HDFS file system. The command **–rmr** can be used to delete files recursively.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -ls /test1
Found 2 items
-rw-r--r--   1 hadoop supergroup       1168 2024-08-01 19:53 /test1/hadooptest.txt
-rw-r--r--   1 hadoop supergroup       2945 2024-08-01 16:34 /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -rm /test1/hadooptest.txt
Deleted /test1/hadooptest.txt
hadoop@iamincognito-VirtualBox:~$
```

9. getmerge:

This is the most important and the most useful command on the HDFS filesystem when trying to read the contents of a MapReduce job or PIG job's output files. This is used for merging a list of files in a directory on the HDFS filesystem into a single local file on the local filesystem.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -getmerge /test2 /test1
getmerge: /test1 (Permission denied)
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -getmerge /test1 /test2
getmerge: /test2 (Permission denied)
```

10. setrep:

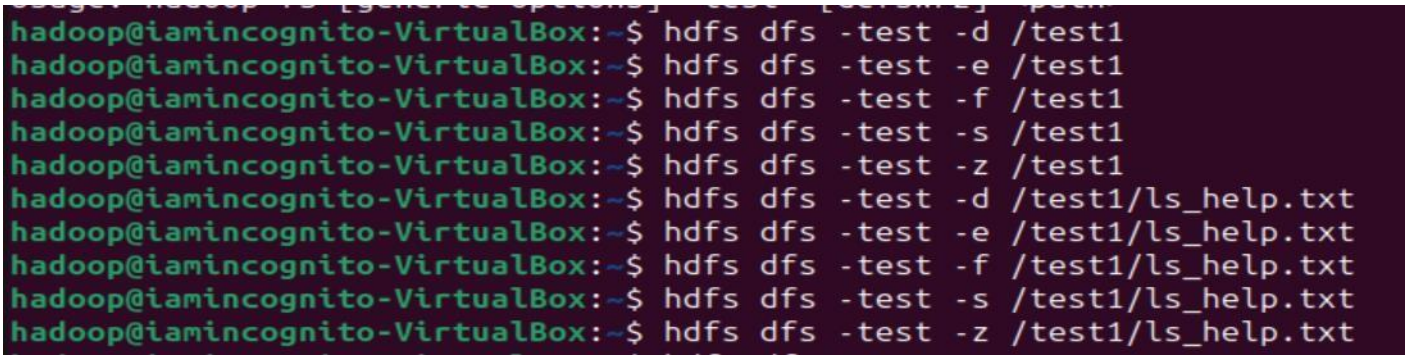
This command is used to change the replication factor of a file to a specific count instead of the default replication factor for the remaining in the HDFS file system. It is a directory then the command will recursively change the replication factor of all the residing files in the directory tree as per the input provided.

11. test:

This command is used to test an HDFS file's existence of zero length of the file or whether if it is a directory or not.

options:

–d	used to check whether if it is a directory or not, returns 0 if it is a directory
–e	used to check whether they exist or not, returns 0 if the exists
–f	used to check whether there is a file or not, returns 0 if the file exists
–s	used to check whether the file size is greater than 0 bytes or not, returns 0 if the size is greater than 0 bytes
–z	used to check whether the file size is zero bytes or not. If the file size is zero bytes, then returns 0 or else returns 1.



12. expunge:

This command is used to empty the trash available in an HDFS system.

13. appendToFile:

This command appends the contents of all the given local files to the provided destination file on the HDFS filesystem. The destination file will be created if it is not existing earlier.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -appendToFile ls_help.txt hadooptest.txt /test3
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -cat /test3
ls/: Unknown command
Usage: hadoop fs [generic options]
        [-appendToFile [-n] <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum [-v] <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-concat <target path> <src path> <src path> ...]
        [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
        [-copyToLocal [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <locald
st>]
        [-count [-q] [-h] [-v] [-t [<storage type>]] [-u] [-x] [-e] [-s] <path> ...]
        [-cp [-f] [-p | -p[topax]] [-d] [-t <thread count>] [-q <thread pool queue size>] <src> ... <dst>]
        [-createSnapshot <snapshotDir> [<snapshotName>]]
        [-deleteSnapshot <snapshotDir> <snapshotName>]
        [-df [-h] [<path> ...]]
        [-du [-s] [-h] [-v] [-x] <path> ...]
        [-expunge [-immediate] [-fs <path>]]
        [-find <path> ... <expression> ...]
        [-get [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
        [-getfacl [-R] <path>]
        [-getfattr [-R] {-n name | -d} [-e en] <path>]
        [-getmerge [-nl] [-skip-empty-file] <src> <localdst>]
        [-head <file>]
        [-help [cmd ...]]
        [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]
        [-mkdir [-p] <path> ...]
        [-moveFromLocal [-f] [-p] [-l] [-d] <localsrc> ... <dst>]
        [-moveToLocal <src> <localdst>]
        [-mv <src> ... <dst>]
        [-put [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
        [-renameSnapshot <snapshotDir> <oldName> <newName>]
        [-rm [-f] [-r|-R] [-skipTrash] [-safely] <src> ...]
        [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
        [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]|[--set <acl_spec> <path>]]
        [-setfattr {-n name [-v value] | -x name} <path>]
        [-setrep [-R] [-w] <rep> <path> ...]
        [-stat [format] <path> ...]
        [-tail [-f] [-s <sleep interval>] <file>]
        [-test -[defswrz] <path>]
        [-text [-ignoreCrc] <src> ...]
        [-touch [-a] [-m] [-t TIMESTAMP (yyyyMMdd:HHmmss) ] [-c] <path> ...]
        [-touchz <path> ...]
        [-truncate [-w] <length> <path> ...]
        [-usage [cmd ...]]

Generic options supported are:
-conf <configuration file>        specify an application configuration file
-D <property=value>               define a value for a given property
-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from confi
gurations.
-jt <local|resourcemanager:port>  specify a ResourceManager
-files <file1,...>                specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>               specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>          specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

In the ever-expanding world of big data, managing and processing vast amounts of information efficiently has become par
amount for businesses and organizations. At the forefront of this data revolution stands Hadoop, a powerful open-source
 framework designed to tackle the challenges of distributed data storage and processing. If you're eager to harness the
 potential of Hadoop for your data projects but find the installation process daunting, fear not! This comprehensive gu
ide will walk you through the essential steps to install Hadoop on your Ubuntu system, demystifying the process and set
ting you on the path to unlocking the limitless possibilities of big data analytics.

Whether you are a seasoned data professional looking to bolster your skills or someone just beginning your data journey
, this guide aims to make Hadoop more accessible than ever before. By the end of this journey, you'll be well-equipped
to explore the world of distributed data processing, leverage the capabilities of Hadoop, and harness its full potentia
l for your data-driven endeavors. So, let's dive in and embark on an exciting journey into the realm of big data analyt
ics.
hadoop@iamincognito-VirtualBox:~$
```

14. tail:

This command is used to show the last 1KB of the file.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -tail /test3
 organizations. At the forefront of this data revolution stands Hadoop, a powerful open-source framework designed to ta
ckle the challenges of distributed data storage and processing. If you're eager to harness the potential of Hadoop for
your data projects but find the installation process daunting, fear not! This comprehensive guide will walk you through
 the essential steps to install Hadoop on your Ubuntu system, demystifying the process and setting you on the path to u
nlocking the limitless possibilities of big data analytics.

Whether you are a seasoned data professional looking to bolster your skills or someone just beginning your data journey
, this guide aims to make Hadoop more accessible than ever before. By the end of this journey, you'll be well-equipped
to explore the world of distributed data processing, leverage the capabilities of Hadoop, and harness its full potentia
l for your data-driven endeavors. So, let's dive in and embark on an exciting journey into the realm of big data analyt
ics.
```

15. Stat:

This command is used to print the statistics about the file/directory in the specified format. Format accepts file size in blocks (%b), the group name of the owner (%g) and the file name (%n), block size (%o), replication (%r), the username of the owner (%u), modification date (%y, %Y)

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -stat /test3
2024-08-01 14:48:35
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -stat %y /test3
2024-08-01 14:48:35
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -stat %Y /test3
1722523715528
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -stat %b /test3
4113
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -stat %r /test3
1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -stat %o /test3
134217728
hadoop@iamincognito-VirtualBox:~$
```

16. df:

This command is used to show the capacity, free and used space available on the **HDFS filesystem**. If the filesystem has multiple partitions and if there is no path is mentioned to any specific partition, then the status of the root partition will be displayed for us to know.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -df -h
Filesystem                Size  Used  Available  Use%
hdfs://localhost:9000  115.5 G  72 K     90.7 G    0%
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -df
Filesystem                     Size   Used    Available  Use%
hdfs://localhost:9000  124024193024  73728  97347387392    0%
hadoop@iamincognito-VirtualBox:~$
```

17. du:

This command is used to show the amount of space in bytes that have been used by the files that match the specified file pattern. Even without the –s option, this only shows the size summaries one level deep in the directory.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -du /test1
1168  1168  /test1/hadooptest.txt
2945  2945  /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -du -s /test1
4113  4113  /test1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -du -h /test1
1.1 K  1.1 K  /test1/hadooptest.txt
2.9 K  2.9 K  /test1/ls_help.txt
hadoop@iamincognito-VirtualBox:~$
```

20. count:

This command is used to count the number of directories, files, and bytes under the path that matches the provided file pattern.

```
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -count /test1
           1            2               4113 /test1
hadoop@iamincognito-VirtualBox:~$ hdfs dfs -count -q /test1
        none             inf            none             inf            1            2               4113 /test1
hadoop@iamincognito-VirtualBox:~$
```