LAB - 02

Interfacing to Hadoop Distributed File System (HDFS)

Aim : Interfacing to Hadoop Distributed File System (HDFS) using command line interface (CLI) ang graphical user interface (GUI) on a stand alone Apache Hadoop instance running locally. Storage as a Service or STaaS is cloud storage which can be architecture using hadoop-like frameworks which works upon horizontal scaling.

EXERCISE :-

EXERCISE-1 : hdfs-site.xml

Here, we can see that the settings of "staticuser" are not there.

```
hadoop@celab3:/opt/hadoop$ cat /opt/hadoop/etc/hadoop/hdfs-site.xml
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/opt/hadoop/dfs/name</value>
  </property>

  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/opt/hadoop/dfs/data</value>
  </property>
  <property>
    <name>dfs.namenode.http-address</name>
    <value>localhost:50070</value>
  </property>

  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>localhost:50090</value>
  </property>

  <property>
  <name>dfs.webhdfs.enabled</name>
  <value>true</value>
  </property>

</configuration>
```

Hence, we can't create a file at frontend and we get the following error :

Modifying the configuration file and adding the property for "staticuser", as below :
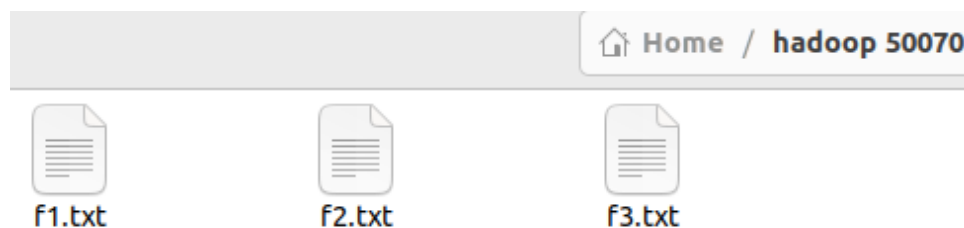
Now I'm able to create a directory at the frontend as below :

EXERCISE - 2: Create a text file in the local file system and put the same file in HDFS using appropriate command.

I created these files using terminal command.

Uploaded these 3 files on the HDFS localhost:50070 by specifying the entire path.

```
hadoop@celab3:~$ hadoop dfs -cat hdfs://localhost:50070/user/hadoop/f1.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2025-01-29 16:40:11,156 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
cat: RPC response exceeds maximum data length
```

```
hadoop@celab3:~$ hadoop dfs -put f2.txt /.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. U
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" inst

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. U
2025-01-29 16:58:58,325 WARN util.NativeCodeLoader: Unable
hadoop@celab3:~$ hadoop dfs -put f1.txt /.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. U
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" inst

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. U
2025-01-29 16:59:48,350 WARN util.NativeCodeLoader: Unable
hadoop@celab3:~$ hadoop dfs -put f3.txt /.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. U
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" inst

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. U
2025-01-29 16:59:52,826 WARN util.NativeCodeLoader: Unable
```

These three files are thus created on the frontend as shown :

Browse Directory

| / | | | | | | | | Go! |

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	drwxr-xr-x	hadoop	supergroup	0 B	Jan 29 16:59	0	0 B	demo	🗑
☐	-rw-r--r--	hadoop	supergroup	11 B	Jan 29 16:59	3	128 MB	f1.txt	🗑
☐	-rw-r--r--	hadoop	supergroup	12 B	Jan 29 16:58	3	128 MB	f2.txt	🗑
☐	-rw-r--r--	hadoop	supergroup	11 B	Jan 29 16:59	3	128 MB	f3.txt	🗑
☐	drwxr-xr-x	hadoop	supergroup	0 B	Jan 29 16:59	0	0 B	student	🗑

Showing 1 to 5 of 5 entries   Previous 1 Next

Hadoop, 2023.

Replication	Block Size	Name	
0	0 B	demo	🗑
3	128 MB	f1.txt	🗑
3	128 MB	f2.txt	🗑
3	128 MB	f3.txt	🗑
0	0 B	student	🗑

Previous **1** Next

EXERCISE-3: Display the contents of this file of HDFS on terminal:

Here, I'm displaying the output of the files using cat command on terminal.

f1.txt

```
hadoop@celab3:~$ hadoop dfs -cat /f1.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2025-01-29 17:01:42,375 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
first data
```

f2.txt

```
hadoop@celab3:~$ hadoop dfs -cat /f2.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2025-01-29 17:01:49,734 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
second data
```

f3.txt

```
hadoop@celab3:~$ hadoop dfs -cat /f3.txt
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2025-01-29 17:01:55,186 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
third data
```
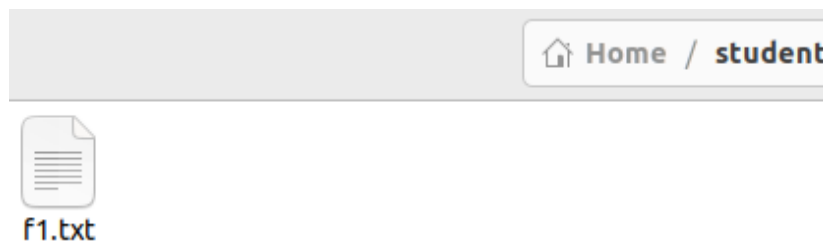
EXERCISE-4: Download this file from HDFS on a local machine and change its content. Upload this file again on HDFS and verify its content. Also, make sure that this file is deleted automatically from the local machine.

Creating an empty folder "student", and copying the f1.txt file here using the following command.

Now, displaying it's content on the terminal.

This file is deleted automatically from the local machine.

The command 'hadoop fs -help' displays a list of available commands for the HDFS shell. It provides a brief description and usage about them.

This command is useful for learning about the different operations one can perform on HDFS, such as creating directories, uploading and downloading files, and managing permissions.

```
hadoop@celab3:~$ hadoop fs -help
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
2025-01-29 17:16:50,709 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
Usage: hadoop fs [generic options]
        [-appendToFile <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
        [-copyToLocal [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread po
        [-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 siz
        [-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
        [-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>] <l
        [-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 ...]]
```

Summarised learning:

I learnt HADOOP and its components. Commands to perform
read and write directory and file structures and data content also with properties of the
same. Also understanding the abstraction imposed by HDFS being a guest file system
on top of the host file system provided by the underlying operating system itself.