

# coreservlets.com – Hadoop Course

## HDFS Installation and Shell

In this exercise, you will have a chance to get antiquated with Pseudo-Distributed installation as well as practice using HDFS shell commands. You will get a chance to create, explore, delete and copy files to/from HDFS.

**Approx Time: 60 minutes**

### Set Up

Virtual Machine has been set up for you with an installation of HDFS in the Pseudo-Distributed Mode. We used Cloudera Distribution for Hadoop version 4 (CDH4) which can be found at

[~/Training/CDH4](#)

This directory can also be referenced with `$CDH_HOME` environment variable. HDFS installation can be found here:

[~/Training/CDH4/hadoop-2.0.0-cdh4.0.0](#)

This directory can also be referenced with `$HADOOP_HOME` environment variable. In addition `$HADOOP_HOME/bin` is added to the `$PATH` therefore you can use `hdfs` command anywhere:

The sample data for this exercise can be found at

[/home/hadoop/Training/exercises/sample\\_data](#)

### Answer

1. What is the Namenode's URI and which file is it configured in?
2. Where on a local file system will Namenode store its image and which file is it configured in?
3. Where on a local file system will Datanode store its blocks and which file is it configured in?
4. What is the block replication and which file is it configured in?

### Perform

1. Start HDFS and verify that it's running
2. Create a new directory [/exercise1](#) on HDFS
3. Upload `$PLAY_AREA/exercises/filesystem/hamlet.txt` to HDFS under [/exercise1](#) directory
4. View the content of the [/exercise1](#) directory
5. Determine the size of the `hamlet.txt` file in KB that resides on HDFS (not local directory)
6. Print the first 25 lines to the screen from `hamlet.txt` on HDFS
7. Copy `hamlet.txt` to `hamlet_hdfsCopy.txt`
8. Copy `hamlet.txt` back to local file system and name it `hamlet_copy.txt`
9. Check the entire filesystem for inconsistencies/problems
10. Delete `hamlet.txt` from HDFS
11. Delete the [/exercise1](#) directory from HDFS
12. Take a second to look at other available shell options

## Answer Solution

1. Namenode's URI is **hdfs://localhost:8020**, it's configured with *fs.default.name* property that's specified in `$HADOOP_CONF_DIR/core-site.xml`
2. Namenode will store its image under `/home/hadoop/Training/hadoop_work/data/name`, it's configured with *dfs.namenode.name.dir* property that's specified in `$HADOOP_CONF_DIR/hdfs-site.xml`
3. Datanode will store data blocks under `/home/hadoop/Training/hadoop_work/data/data`, it's configured with *dfs.datanode.data.dir* property that's specified in `$HADOOP_CONF_DIR/hdfs-site.xml`
4. Replication is set to 1, it's configured with *dfs.replication* property that's specified in `$HADOOP_CONF_DIR/hdfs-site.xml`

## Perform Solution

1. Perform the following steps:
  - a. `$ cd $HADOOP_HOME/sbin`
  - b. `$ ./start-dfs.sh` This will start the Namenode, Secondary Namenode all the configured Datanodes, which in this case is just one (localhost)
  - c. You can verify with the browser or via command line:
    - i. Open a browser and navigate to `http://localhost:50070`, make sure there are no warnings under 'Cluster Summary' section and there is 1 live node. Make sure there are no 'Dead Nodes' and has 0 under replicated blocks Click on 'Live Nodes' links and verify that there are no failed volumes and 'Admin State' is listed as 'In Service'
    - ii. Secondary Namenode can be confirmed via `http://localhost:50090`
    - iii. Execute on the command line `$ hadoop dfsadmin -report`, you will get a report about the status of the cluster. Make sure there is 1 live node, 0 dead nodes and 0 under-replicated blocks.
2. `$ hdfs dfs -mkdir /exercise1`
3. Perform the following steps:
  - a. `$ cd $PLAY_AREA/exercises/filesystem`
  - b. `$ hdfs dfs -put hamlet.txt /exercise1/`
4. `$ hdfs dfs -ls /exercise1/`
5. Perform the following steps:
  - a. `$ hdfs dfs -du -h /exercise1/hamlet.txt`  
**206.3k** /exercise1/hamlet.txt
6. `$ hdfs dfs -cat /exercise1/hamlet.txt | head -n 25`
7. `$ hdfs dfs -cp /exercise1/hamlet.txt /exercise1/hamlet_hdfsCopy.txt`
8. `$ hdfs dfs -get /exercise1/hamlet.txt hamlet_copy.txt`
9. `$ hdfs fsck /`
10. `$ hdfs dfs -rm /exercise1/hamlet.txt`
11. `$ hdfs dfs -rm -r /exercise1`
12. `$ hdfs dfs -help`