Assignment 1 - HDFS

1. Start HDFS and verify that it's running.

```
hadoop-daemon.sh start namenode
hadoop-daemon.sh start datanode
hdfs dfsadmin -report
```

   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*. Execute on the command line `hdfs 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. Create a new directory `/sics` on HDFS.

```
hdfs dfs -mkdir /sics
```

3. Create a file, name it `big`, on your local filesystem and upload it to HDFS under `/sics`.

```
hdfs dfs -put big /sics
```

4. View the content of `/sics` directory.

```
hdfs dfs -ls big /sics
```

5. Determine the size of `big` on HDFS.

```
hdfs dfs -du -h /sics/big
```

6. Print the first 5 lines to screen from `big` on HDFS.

```
hdfs dfs -cat /sics/big | head -n 5
```

7. Copy `big` to `/big_hdfscopy` on HDFS.

```
hdfs dfs -cp /sics/big /sics/big_hdfscopy
```

8. Copy `big` back to local filesystem and name it `big_localcopy`.

```
hdfs dfs -get /sics/big big_localcopy
```

9. Check the entire HDFS filesystem for inconsistencies/problems.

```
hdfs fsck /
```

10. Delete `big` from HDFS.

```
hdfs dfs -rm /sics/big
```

11. Delete `/sics` directory from HDFS.

```
hdfs dfs -rm -r /sics
```