

HDFS Task-1

1. What is the Namenode's URI and which file is it configured in?

Answer : name node is the master and management node of Hadoop system
in core-site.xml file the Uri is mentioned.

2. Where on a local file system will Namenode store its image and which file is it configured in?

Answer: In Hadoop home etc folder the hdfs-site.xml file is there in that file
name node property is configured

3. Where on a local file system will Datanode store its blocks and which file is it configured in?

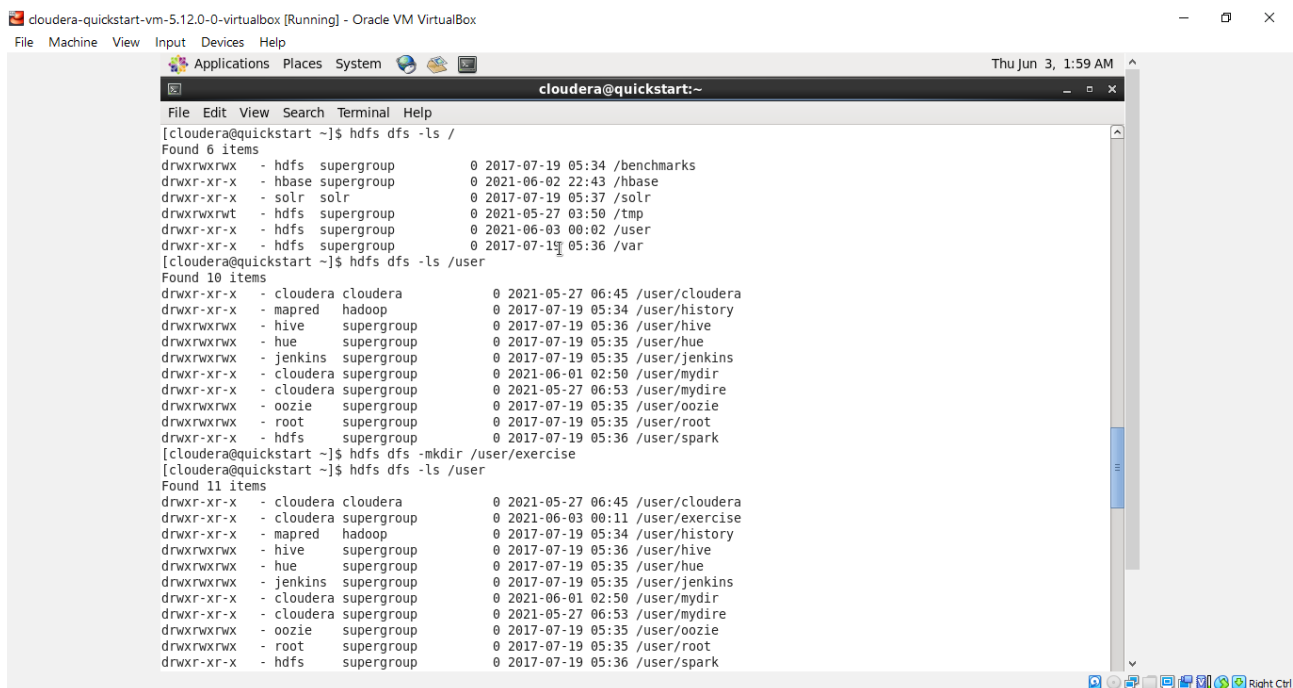
Answer: In Hadoop home etc folder the hdfs-site.xml file is there in that file
data node property is configured.

4. What is the block replication and which file is it configured in?

Answer: Its a copy of the data of node of rack, each data should be placed in
different nodes of different rack.It is configured in hdfs-site.xml

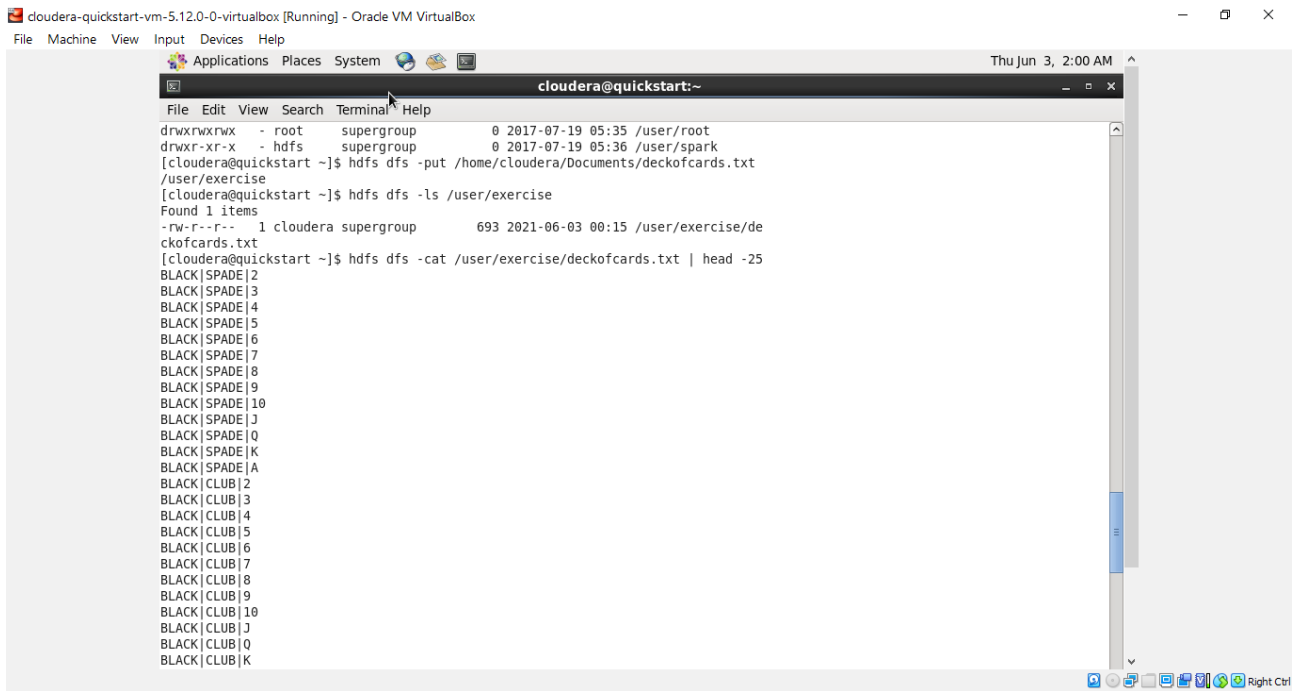
Perform

1. Start HDFS and verify that it's running 2.Create a new directory /exercise1 on HDFS



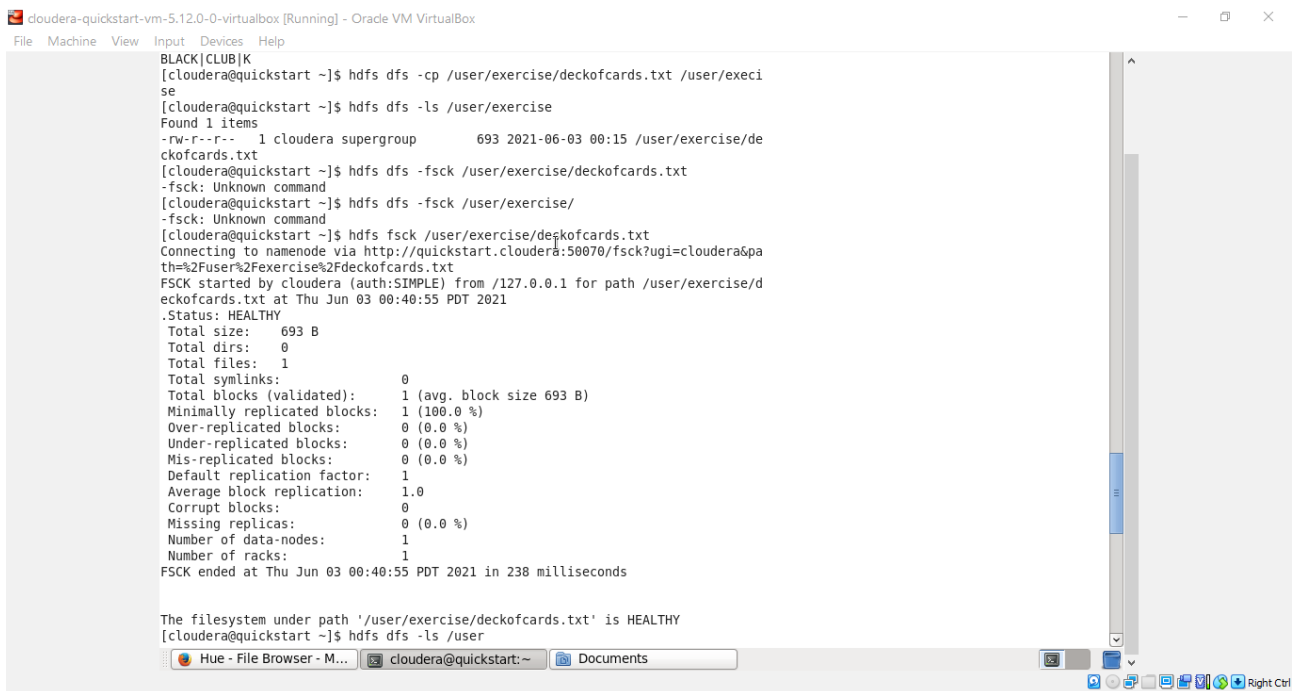
```
cloudera-quickstart-vm-5.12.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Thu Jun 3, 1:59 AM
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup 0 2021-06-02 22:43 /hbase
drwxr-xr-x - solr solr 0 2017-07-19 05:37 /solr
drwxrwxrwt - hdfs supergroup 0 2021-05-27 03:50 /tmp
drwxr-xr-x - hdfs supergroup 0 2021-06-03 00:02 /user
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /var
[cloudera@quickstart ~]$ hdfs dfs -ls /user
Found 10 items
drwxr-xr-x - cloudera cloudera 0 2021-05-27 06:45 /user/cloudera
drwxr-xr-x - mapred hadoop 0 2017-07-19 05:34 /user/history
drwxrwxrwx - hive supergroup 0 2017-07-19 05:36 /user/hive
drwxrwxrwx - hue supergroup 0 2017-07-19 05:35 /user/hue
drwxrwxrwx - jenkins supergroup 0 2017-07-19 05:35 /user/jenkins
drwxr-xr-x - cloudera supergroup 0 2021-06-01 02:50 /user/mydir
drwxr-xr-x - cloudera supergroup 0 2021-05-27 06:53 /user/mydire
drwxrwxrwx - oozie supergroup 0 2017-07-19 05:35 /user/oozie
drwxrwxrwx - root supergroup 0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/exercise
[cloudera@quickstart ~]$ hdfs dfs -ls /user
Found 11 items
drwxr-xr-x - cloudera cloudera 0 2021-05-27 06:45 /user/cloudera
drwxr-xr-x - cloudera supergroup 0 2021-06-03 00:11 /user/exercise
drwxr-xr-x - mapred hadoop 0 2017-07-19 05:34 /user/history
drwxrwxrwx - hive supergroup 0 2017-07-19 05:36 /user/hive
drwxrwxrwx - hue supergroup 0 2017-07-19 05:35 /user/hue
drwxrwxrwx - jenkins supergroup 0 2017-07-19 05:35 /user/jenkins
drwxr-xr-x - cloudera supergroup 0 2021-06-01 02:50 /user/mydir
drwxr-xr-x - cloudera supergroup 0 2021-05-27 06:53 /user/mydire
drwxrwxrwx - oozie supergroup 0 2017-07-19 05:35 /user/oozie
drwxrwxrwx - root supergroup 0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark
```

3. Upload GitHub repo sample_data/deckofcards.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 deckofcards.txt on HDFS



```
cloudera@quickstart:~  
drwxrwxrwx - root supergroup 0 2017-07-19 05:35 /user/root  
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark  
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Documents/deckofcards.txt  
/user/exercise  
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise  
Found 1 items  
-rw-r--r-- 1 cloudera supergroup 693 2021-06-03 00:15 /user/exercise/de  
ckofcards.txt  
[cloudera@quickstart ~]$ hdfs dfs -cat /user/exercise/deckofcards.txt | head -25  
BLACK|SPADE|2  
BLACK|SPADE|3  
BLACK|SPADE|4  
BLACK|SPADE|5  
BLACK|SPADE|6  
BLACK|SPADE|7  
BLACK|SPADE|8  
BLACK|SPADE|9  
BLACK|SPADE|10  
BLACK|SPADE|J  
BLACK|SPADE|Q  
BLACK|SPADE|K  
BLACK|SPADE|A  
BLACK|CLUB|2  
BLACK|CLUB|3  
BLACK|CLUB|4  
BLACK|CLUB|5  
BLACK|CLUB|6  
BLACK|CLUB|7  
BLACK|CLUB|8  
BLACK|CLUB|9  
BLACK|CLUB|10  
BLACK|CLUB|J  
BLACK|CLUB|Q  
BLACK|CLUB|K
```

9. Check the entire filesystem for inconsistencies/problems



```
cloudera@quickstart:~  
[cloudera@quickstart ~]$ hdfs dfs -cp /user/exercise/deckofcards.txt /user/exercise  
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise  
Found 1 items  
-rw-r--r-- 1 cloudera supergroup 693 2021-06-03 00:15 /user/exercise/de  
ckofcards.txt  
[cloudera@quickstart ~]$ hdfs dfs -fsck /user/exercise/deckofcards.txt  
-fsck: Unknown command  
[cloudera@quickstart ~]$ hdfs dfs -fsck /user/exercise/  
-fsck: Unknown command  
[cloudera@quickstart ~]$ hdfs fsck /user/exercise/deckofcards.txt  
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&pa  
th=%2Fuser%2Fexercise%2Fdeckofcards.txt  
FSCK started by cloudera (auth:SIMPLE) from /127.0.0.1 for path /user/exercise/d  
eckofcards.txt at Thu Jun 03 00:40:55 PDT 2021  
.Status: HEALTHY  
Total size: 693 B  
Total dirs: 0  
Total files: 1  
Total symlinks: 0  
Total blocks (validated): 1 (avg. block size 693 B)  
Minimally replicated blocks: 1 (100.0 %)  
Over-replicated blocks: 0 (0.0 %)  
Under-replicated blocks: 0 (0.0 %)  
Mis-replicated blocks: 0 (0.0 %)  
Default replication factor: 1  
Average block replication: 1.0  
Corrupt blocks: 0  
Missing replicas: 0 (0.0 %)  
Number of data-nodes: 1  
Number of racks: 1  
FSCK ended at Thu Jun 03 00:40:55 PDT 2021 in 238 milliseconds  
  
The filesystem under path '/user/exercise/deckofcards.txt' is HEALTHY  
[cloudera@quickstart ~]$ hdfs dfs -ls /user
```

7. Copy deckofcards.txt to deckofcardsCopy.txt
8. Copy deckofcards.txt back to local file system and name it deckofcards.copy.txt
10. Delete deckofcards.txt from HDFS
11. Delete the /exercise1 directory from HDFS
12. Take a second to look at other available shell options.

```
File Machine View Input Devices Help
drwxrwxrwx - oozie supergroup 0 2017-07-19 05:35 /user/oozie
drwxrwxrwx - root supergroup 0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark
[cloudera@quickstart ~]$ hdfs dfs -rm -r /user/exercise
Deleted /user/exercise
[cloudera@quickstart ~]$ hdfs dfs -cp /user/exercise/deckofcards.txt /user/exercise
cp: '/user/exercise/deckofcards.txt': File exists
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal /home/cloudera/Documents/deckofcards.txt /user/exercise
copyFromLocal: '/home/cloudera/Documents/deckofcards.txt': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal /home/cloudera/Documents/deckofcards.txt /user/exercise
copyFromLocal: '/user/exercise/deckofcards.txt': File exists
[cloudera@quickstart ~]$ hdfs dfs -get /user/exercise/deckofcards.txt /home/cloudera/Downloads
[cloudera@quickstart ~]$ hdfs dfs -get /user/exercise/deckofcards.txt /home/cloudera/Documents
get: '/home/cloudera/Documents/deckofcards.txt': File exists
[cloudera@quickstart ~]$ hdfs dfs -rm -r /user/exercise/deckofcards.txt
Deleted /user/exercise/deckofcards.txt
[cloudera@quickstart ~]$ hdfs dfs -rm -r /user/exercise/
Deleted /user/exercise
[cloudera@quickstart ~]$ hdfs dfs -ls /user
Found 10 items
drwxr-xr-x - cloudera cloudera 0 2021-05-27 06:45 /user/cloudera
drwxr-xr-x - mapred hadoop 0 2017-07-19 05:34 /user/history
drwxrwxrwx - hive supergroup 0 2017-07-19 05:36 /user/hive
drwxrwxrwx - hue supergroup 0 2017-07-19 05:35 /user/hue
drwxrwxrwx - jenkins supergroup 0 2017-07-19 05:35 /user/jenkins
drwxr-xr-x - cloudera supergroup 0 2021-06-01 02:50 /user/mydir
drwxr-xr-x - cloudera supergroup 0 2021-05-27 06:53 /user/mydire
drwxrwxrwx - oozie supergroup 0 2017-07-19 05:35 /user/oozie
drwxrwxrwx - root supergroup 0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /user/spark
[cloudera@quickstart ~]$ hdfs dfs -ls /user
```

8 & 10 -12