

HDFS Task-1

Answer:

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

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

Ans :- Namenode will store its image under `/var/lib/hadoop-hdfs/cache/${user.name}/dfs/name/`, it's configured with `dfs.namenode.name.dir` property that's specified in `$HADOOP_CONF_DIR/hdfs-site.xml`

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

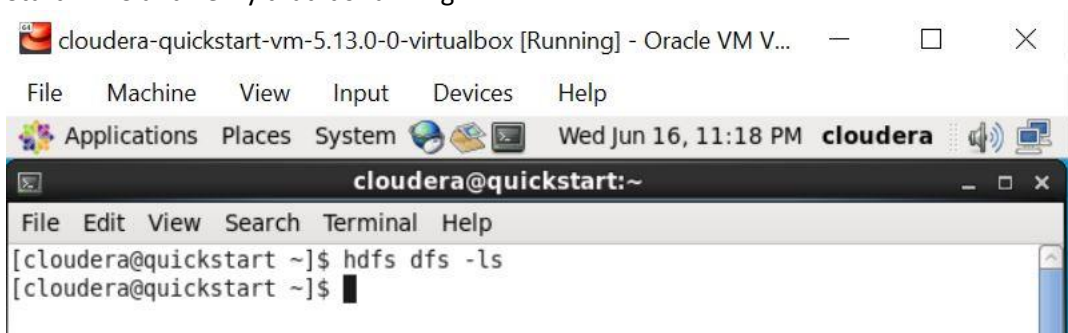
Ans : - Datanode will store data blocks under `/var/lib/hadoop-hdfs/cache/${user.name}/dfs/data/`, it's configured with `dfs.datanode.data.dir` property that's specified in `$HADOOP_CONF_DIR/hdfs-site.xml`

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

Ans : - Replication is set to 1, it's configured with `dfs.replication` property that's specified in `$HADOOP_CONF_DIR/hdfs-site.xml`

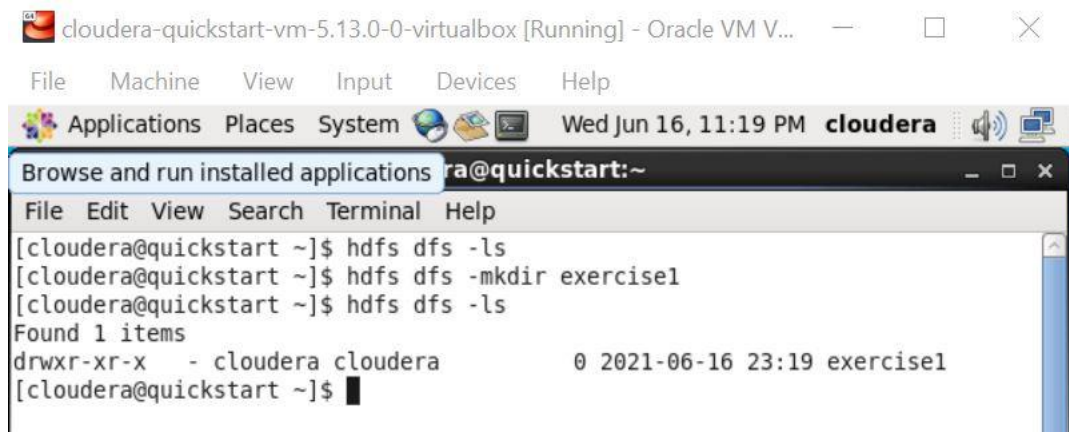
Perform:

1. Start HDFS and verify that it's running



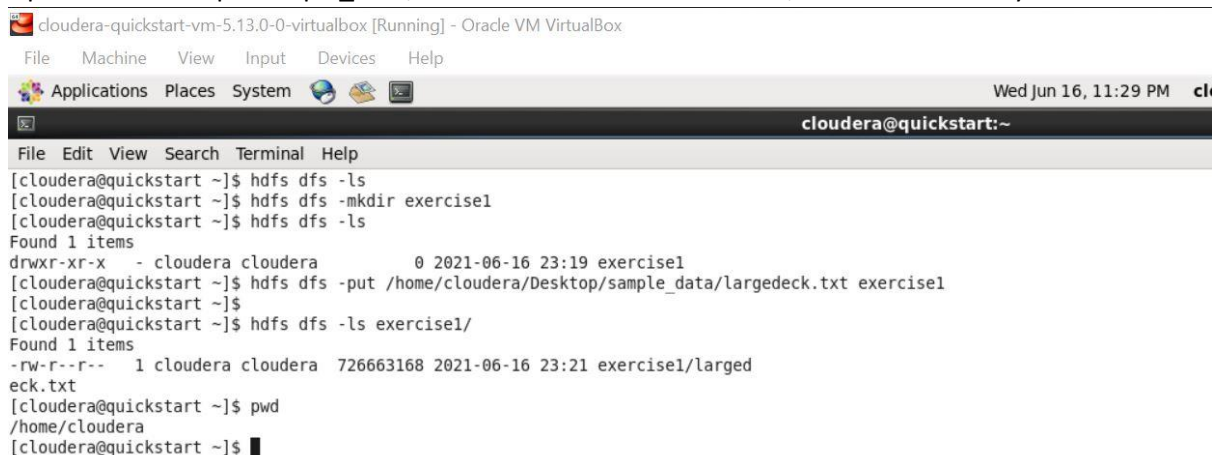
The screenshot shows a virtual machine window titled "cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM V...". Inside the VM, there is a terminal window titled "cloudera@quickstart:~". The terminal has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The command prompt shows the user "cloudera" at the host "quickstart" in the directory "~". The command `hdfs dfs -ls` has been entered and executed, resulting in a blank line, indicating that the HDFS filesystem is accessible.

2. Create a new directory /exercise1 on HDFS



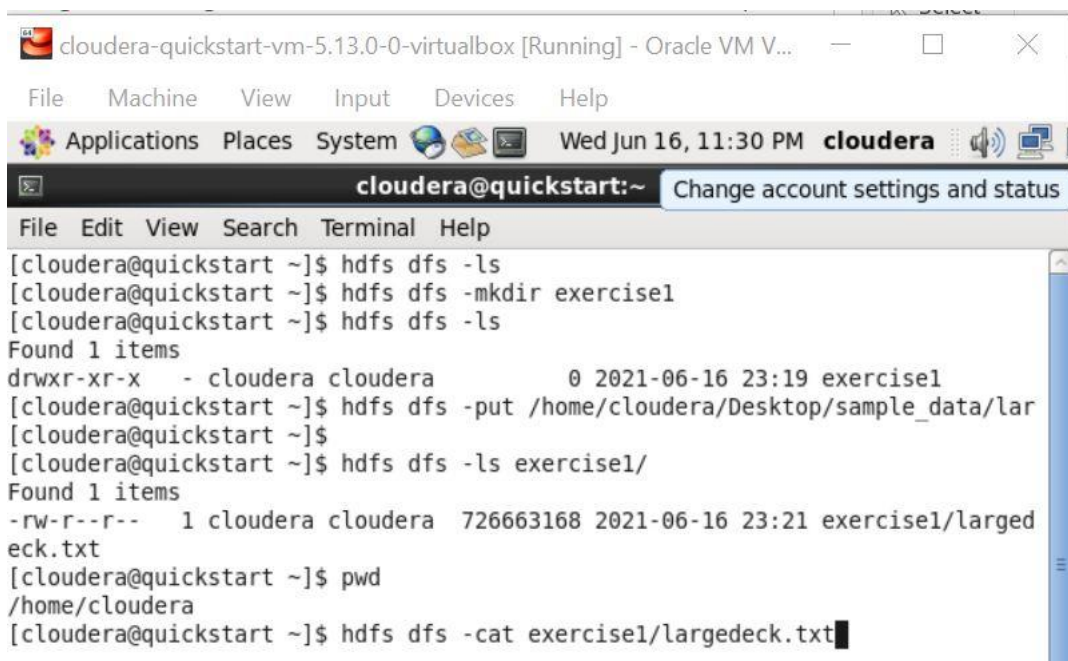
```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM V...
File Machine View Input Devices Help
Applications Places System Wed Jun 16, 11:19 PM cloudera
Browse and run installed applications cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -mkdir exercise1
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - cloudera cloudera 0 2021-06-16 23:19 exercise1
[cloudera@quickstart ~]$
```

3. Upload GitHub repo sample_data/deckofcards.txt to HDFS under /exercise1 directory



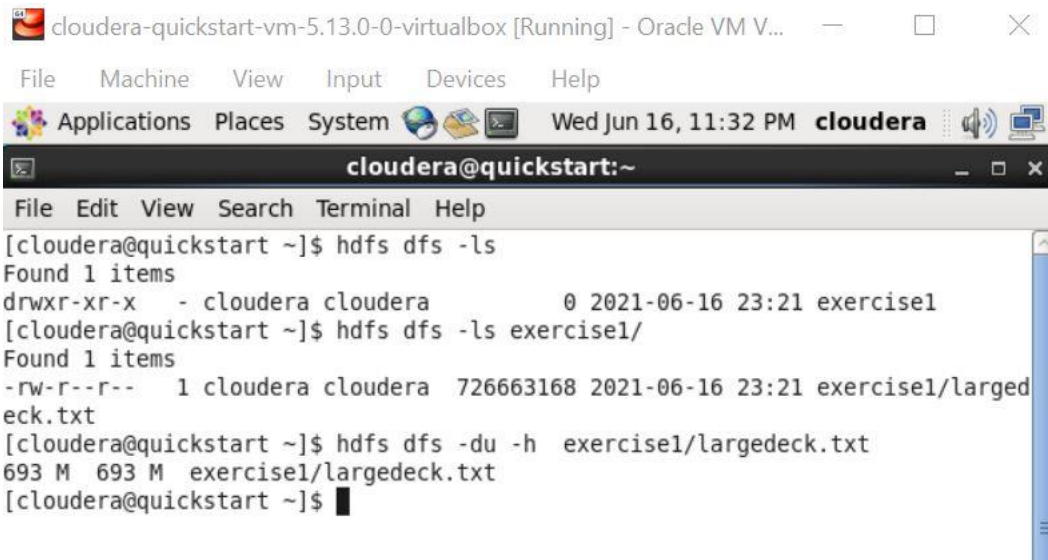
```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Wed Jun 16, 11:29 PM cl
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -mkdir exercise1
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - cloudera cloudera 0 2021-06-16 23:19 exercise1
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/sample_data/largedecck.txt exercise1
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hdfs dfs -ls exercise1/
Found 1 items
-rw-r--r-- 1 cloudera cloudera 726663168 2021-06-16 23:21 exercise1/largedecck.txt
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$
```

4. View the content of the /exercise1 directory



```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM V...
File Machine View Input Devices Help
Applications Places System Wed Jun 16, 11:30 PM cloudera
Change account settings and status
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls
[cloudera@quickstart ~]$ hdfs dfs -mkdir exercise1
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - cloudera cloudera 0 2021-06-16 23:19 exercise1
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/sample_data/largedecck.txt
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hdfs dfs -ls exercise1/
Found 1 items
-rw-r--r-- 1 cloudera cloudera 726663168 2021-06-16 23:21 exercise1/largedecck.txt
[cloudera@quickstart ~]$ pwd
/home/cloudera
[cloudera@quickstart ~]$ hdfs dfs -cat exercise1/largedecck.txt
```

5. Determine the size of the hamlet.txt file in KB that resides on HDFS (not local directory)



```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM V...
File Machine View Input Devices Help
Applications Places System Wed Jun 16, 11:32 PM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - cloudera cloudera 0 2021-06-16 23:21 exercisel
[cloudera@quickstart ~]$ hdfs dfs -ls exercisel/
Found 1 items
-rw-r--r-- 1 cloudera cloudera 726663168 2021-06-16 23:21 exercisel/larged
eck.txt
[cloudera@quickstart ~]$ hdfs dfs -du -h exercisel/largedeck.txt
693 M 693 M exercisel/largedeck.txt
[cloudera@quickstart ~]$
```

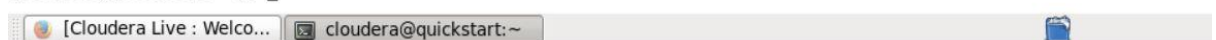
6. Print the first 25 lines to the screen from deckofcards.txt on HDFS



```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System Wed Jun 16, 11:33 PM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - cloudera cloudera 0 2021-06-16 23:21 exercisel
[cloudera@quickstart ~]$ hdfs dfs -ls exercisel/
Found 1 items
-rw-r--r-- 1 cloudera cloudera 726663168 2021-06-16 23:21 exercisel/larged
eck.txt
[cloudera@quickstart ~]$ hdfs dfs -du -h exercisel/largedeck.txt
693 M 693 M exercisel/largedeck.txt
[cloudera@quickstart ~]$ hdfs dfs -cat exercisel/largedeck.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
cat: Unable to write to output stream.
[cloudera@quickstart ~]$
```

7. Copy deckofcards.txt to deckofcardsCopy.txt

```
[cloudera@quickstart ~]$ hdfs dfs -cp exercisel/largedeck.txt exercisel/largedeckcopy.txt
[cloudera@quickstart ~]$ hdfs dfs -du -h exercisel
693 M 693 M exercisel/largedeck.txt
693 M 693 M exercisel/largedeckcopy.txt
[cloudera@quickstart ~]$
```



8. Copy `deckofcards.txt` back to local file system and name it `deckofcards.copy.txt`

```

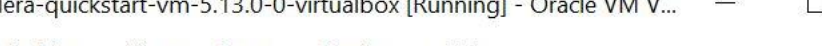
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -get exercisel/largedeck.txt /home/cloudera/Desktop/
[cloudera@quickstart ~]$ hdfs dfs -get exercisel/largedeck.txt /home/cloudera/Desktop/sample_data/largedeckcopy.txt
[cloudera@quickstart ~]$ ls -lrth /home/cloudera/Desktop/sample_data/
total 1.4G
-rwxrwxrwx 1 cloudera cloudera 745 Jun 15 12:10 deckofcards.txt
-rw-rw-r-- 1 cloudera cloudera 3.4M Jun 15 23:43 largedeck.txt.gz
-rw-rw-r-- 1 cloudera cloudera 693M Jun 16 21:27 largedeck.txt
-rw-r--r-- 1 cloudera cloudera 693M Jun 16 23:41 largedeckcopy.txt
[cloudera@quickstart ~]$

```

9. Check the entire filesystem for inconsistencies/problems

```
cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox  
File Machine View Input Devices Help  
Applications Places System  
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs fsck /  
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&path=/2F  
FSCK started by cloudera (auth:SIMPLE) from /10.0.2.15 for path / at Wed Jun 16 23:43:45 PDT 2021  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....Status: HEALTHY  
Total size: 2314016470 B (Total open files size: 166 B)  
Total dirs: 112  
Total files: 933  
Total symlinks: 0 (Files currently being written: 3)  
Total blocks (validated): 941 (avg. block size 2459741 B) (Total open f  
ile blocks (not validated): 2)  
Minimally replicated blocks: 941 (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 Wed Jun 16 23:43:45 PDT 2021 in 170 milliseconds  
  
The filesystem under path '/' is HEALTHY  
[cloudera@quickstart ~]$
```

- ## 10. Delete deckofcards.txt from HDFS



cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM V...

File Machine View Input Devices Help

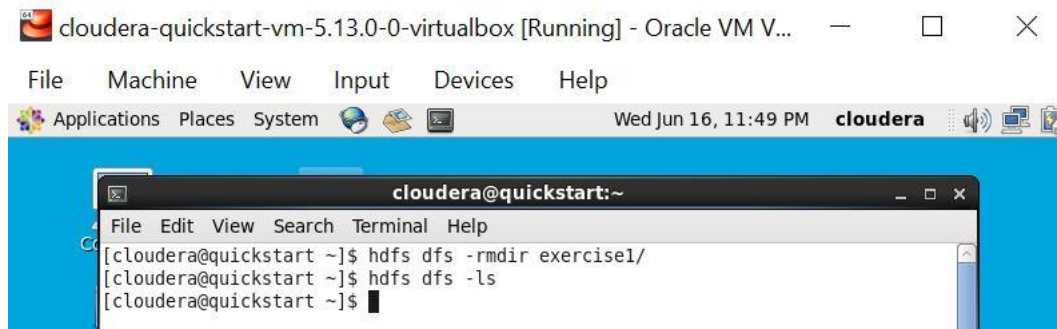
Applications Places System Wed Jun 16, 11:48 PM cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hdfs dfs -rm exercisel/largedeck.txt
Deleted exercisel/largedeck.txt
[cloudera@quickstart ~]$ hdfs dfs -rm exercisel/largedeckcopy.txt
Deleted exercisel/largedeckcopy.txt
```

11. Delete the /exercise1 directory from HDFS



12. Take a second to look at other available shell options
hdfs dfs -help