**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 thats 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

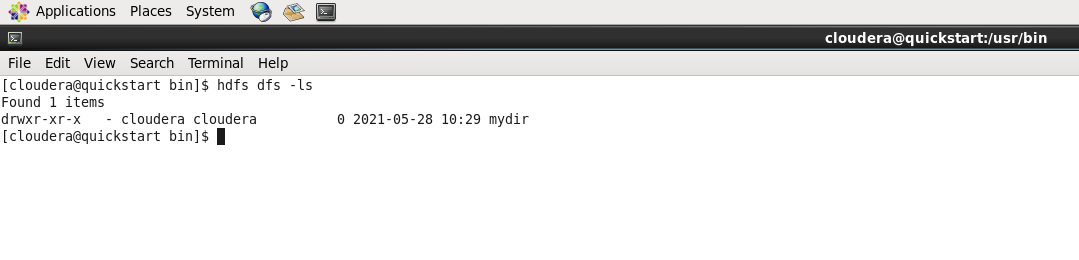
4. What is the block replication and which file is it configured in?

Ans : - Replication is set to 1, it's configured with dfs.replication property that’s specified in

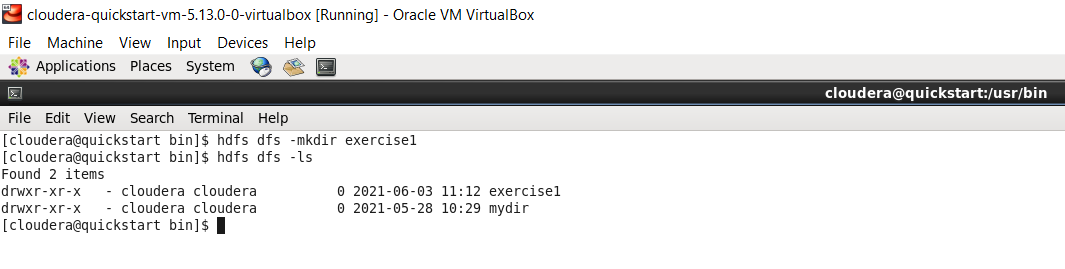
$HADOOP\_CONF\_DIR/hdfs-site.xml

**Perform**

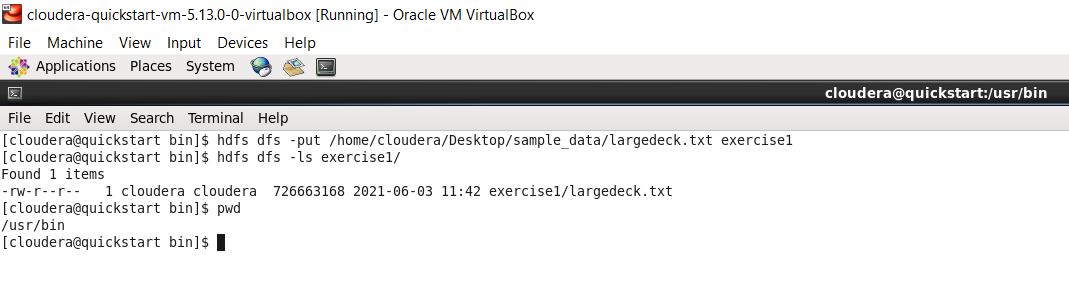
1. Start HDFS and verify that it's running



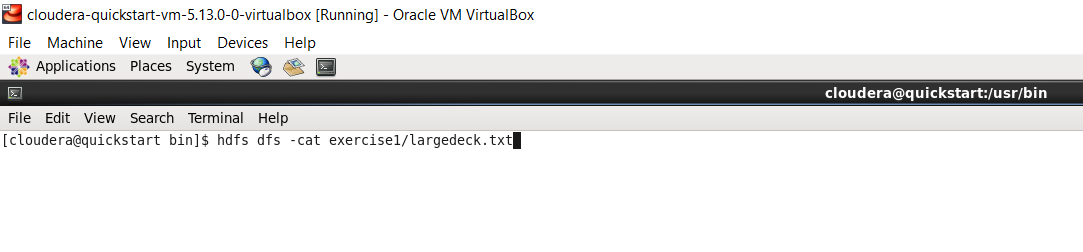
2. Create a new directory /exercise1 on HDFS



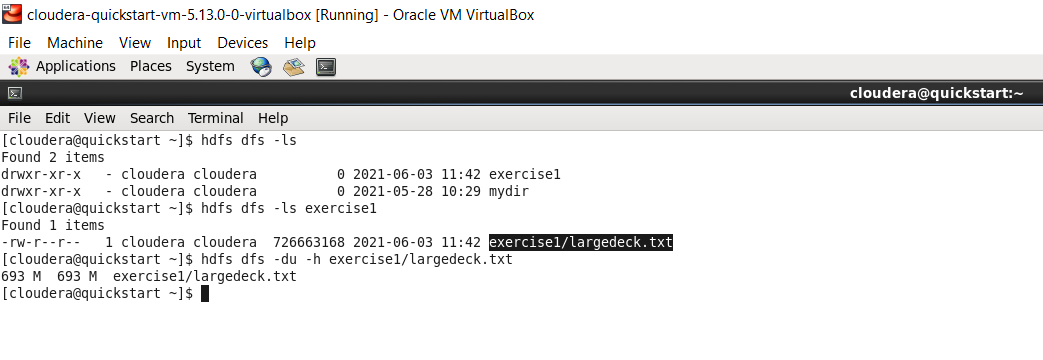
3. Upload GitHub repo sample\_data/largedeck.txt to HDFS under /exercise1 directory



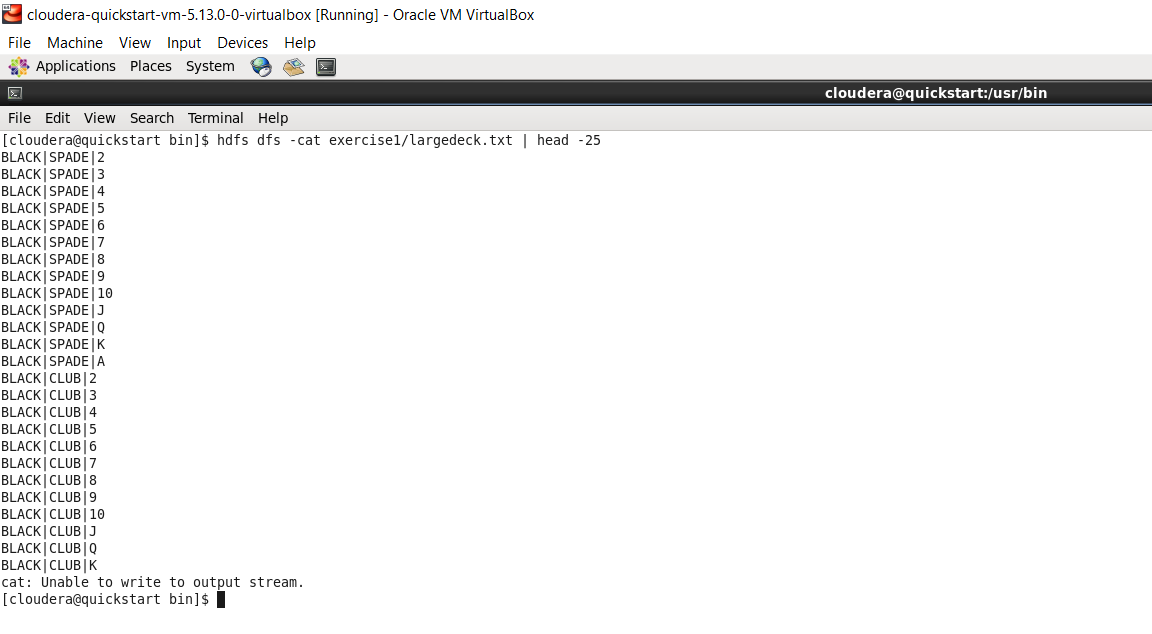
4. View the content of the /exercise1 directory



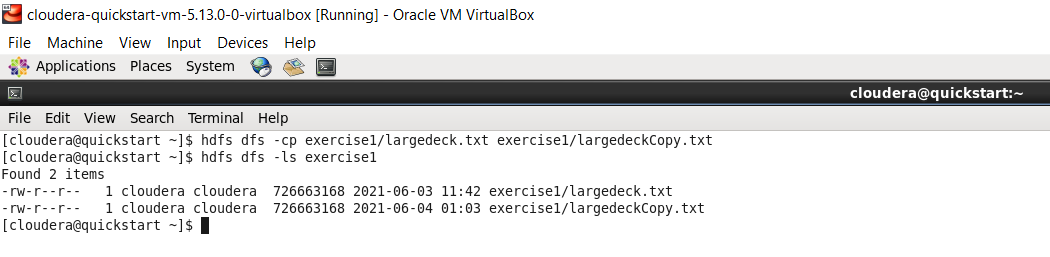
5. Determine the size of the largedeck.txt file in KB that resides on HDFS (not local directory)



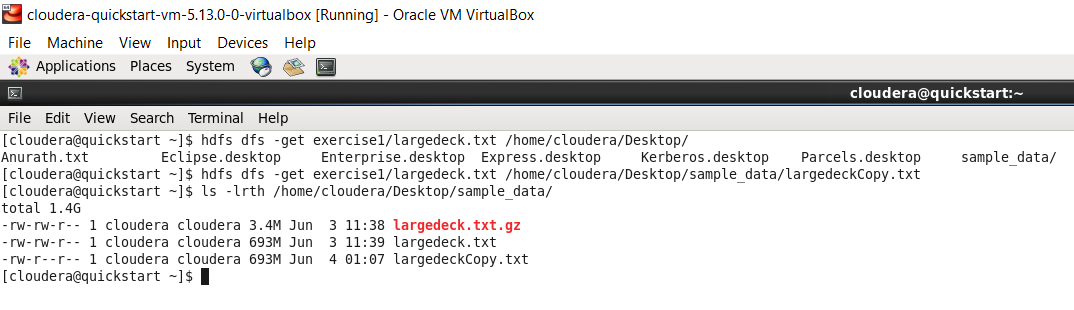
6. Print the first 25 lines to the screen from deckofcards.txt on HDFS



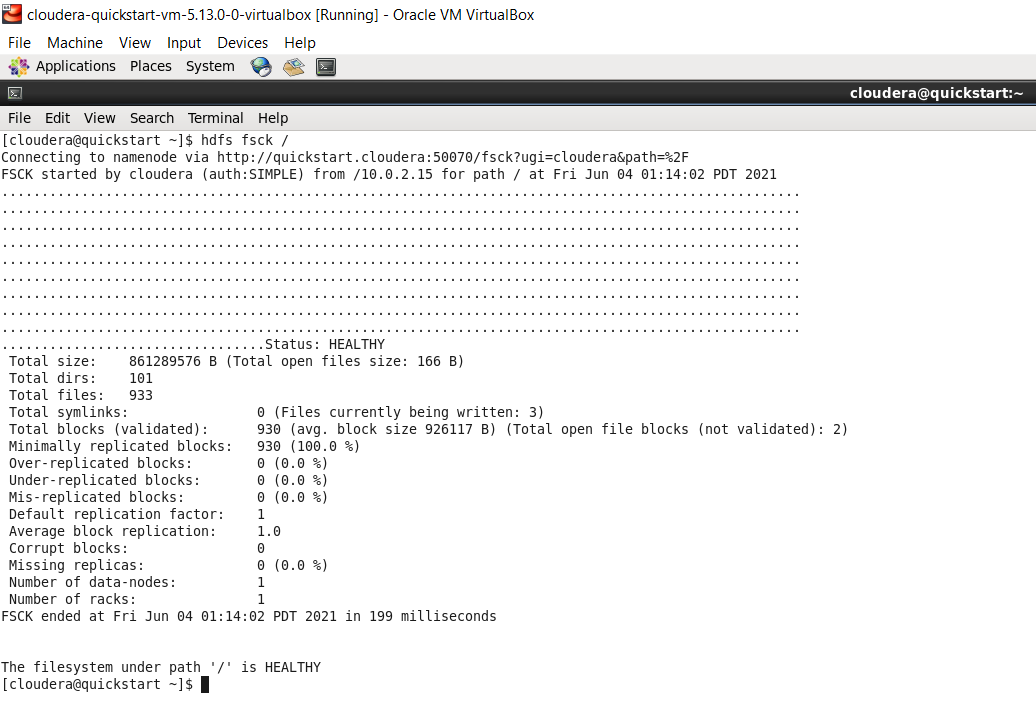
7. Copy largedeck.txt to largedeckCopy.txt



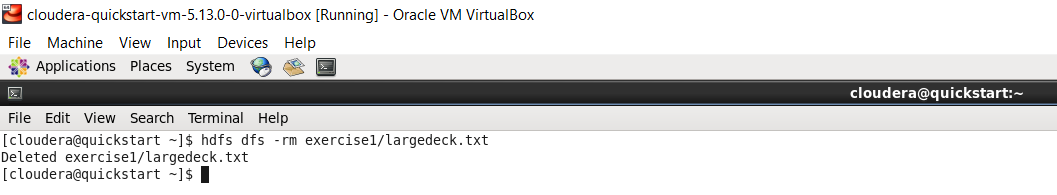
8. Copy largedeck.txt back to local file system and name it largedeck.copy.txt



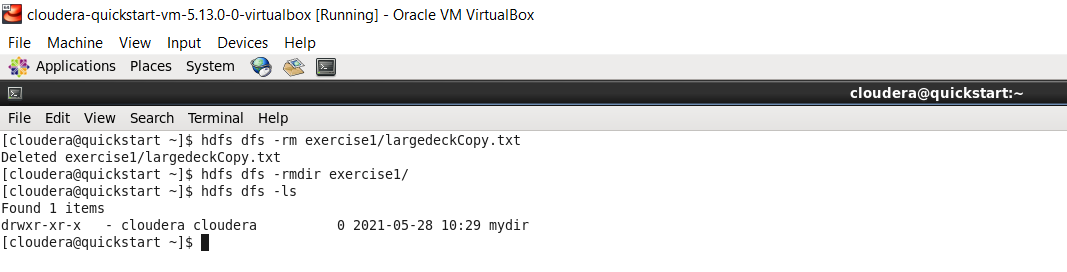
9. Check the entire filesystem for inconsistencies/problems



10. Delete largedeck.txt from HDFS



11. Delete the /exercise1 directory from HDFS



12. Take a second to look at other available shell options.

hdfs dfs -help