

1. Start HDFS and verify that it's running

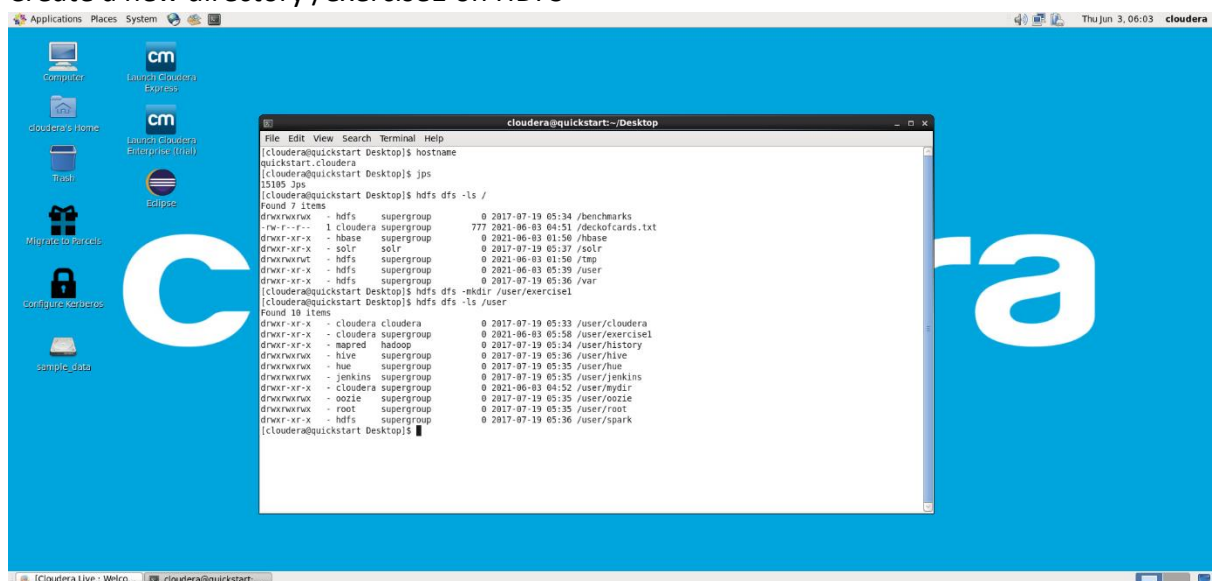
## Overview 'quickstart.cloudera:8020' (active)

<b>Started:</b>	Thu Jun 03 14:16:09 +0530 2021
<b>Version:</b>	2.6.0-cdh5.12.0, rdba647c5a8bc5e09b572d76a8d29481c78d1a0dd
<b>Compiled:</b>	Thu Jun 29 17:02:00 +0530 2017 by jenkins from Unknown
<b>Cluster ID:</b>	CID-70a4b6a0-c9f5-4ddf-bd57-1a338fca4f7a
<b>Block Pool ID:</b>	BP-1914853243-127.0.0.1-1500467607052

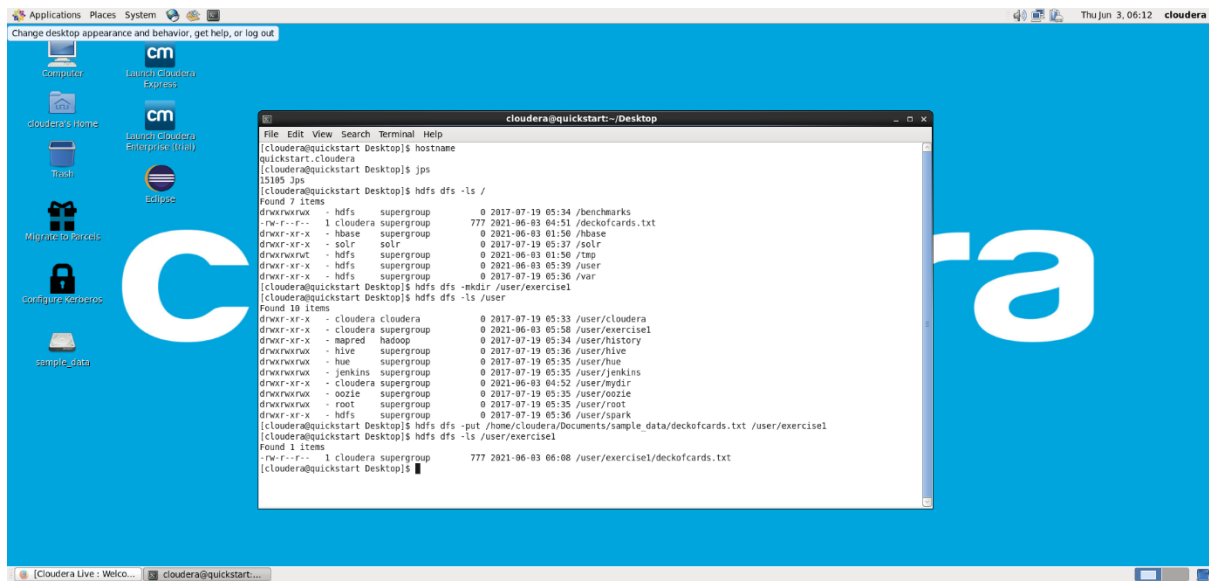
Non Heap Memory used 39.18 MB of 39.5 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

<b>Configured Capacity:</b>	54.51 GB
<b>DFS Used:</b>	830.43 MB (1.49%)
<b>Non DFS Used:</b>	32.39 GB
<b>DFS Remaining:</b>	42.82 GB (78.56%)
<b>Block Pool Used:</b>	830.43 MB (1.49%)
<b>DataNodes usages% (Min/Median/Max/stdDev):</b>	1.49% / 1.49% / 1.49% / 0.00%
<b>Live Nodes</b>	1 (Decommissioned: 0, In Maintenance: 0)
<b>Dead Nodes</b>	0 (Decommissioned: 0, In Maintenance: 0)
<b>Decommissioning Nodes</b>	0
<b>Entering Maintenance Nodes</b>	0
<b>Total Datanode Volume Failures</b>	0 (0 B)
<b>Number of Under-Replicated Blocks</b>	0
<b>Number of Blocks Pending Deletion</b>	0

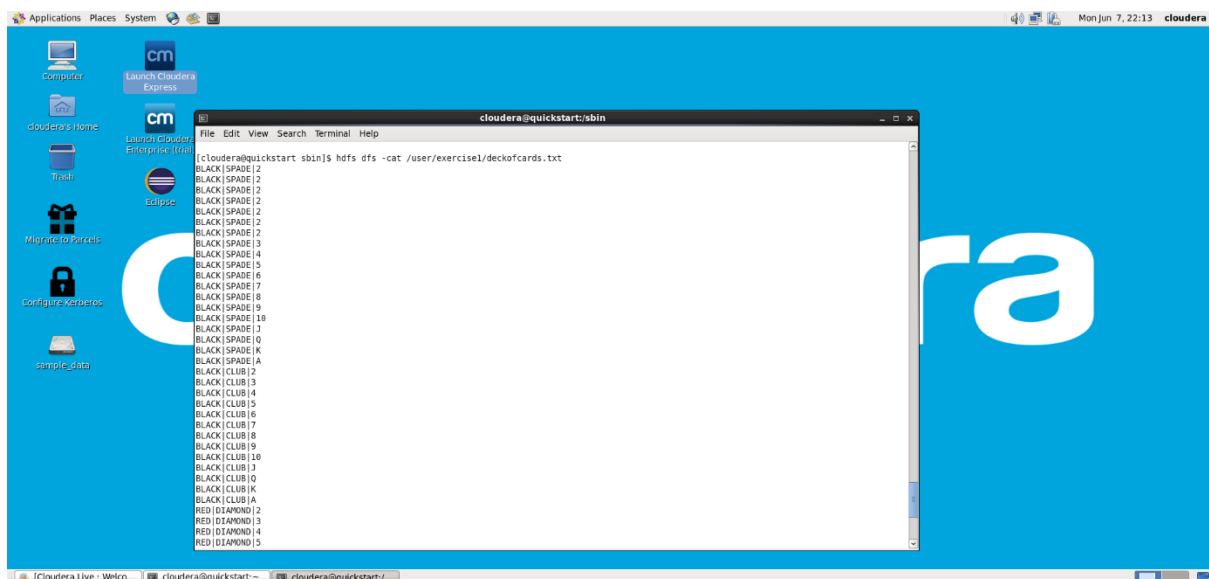
2. Create a new directory /exercise1 on HDFS



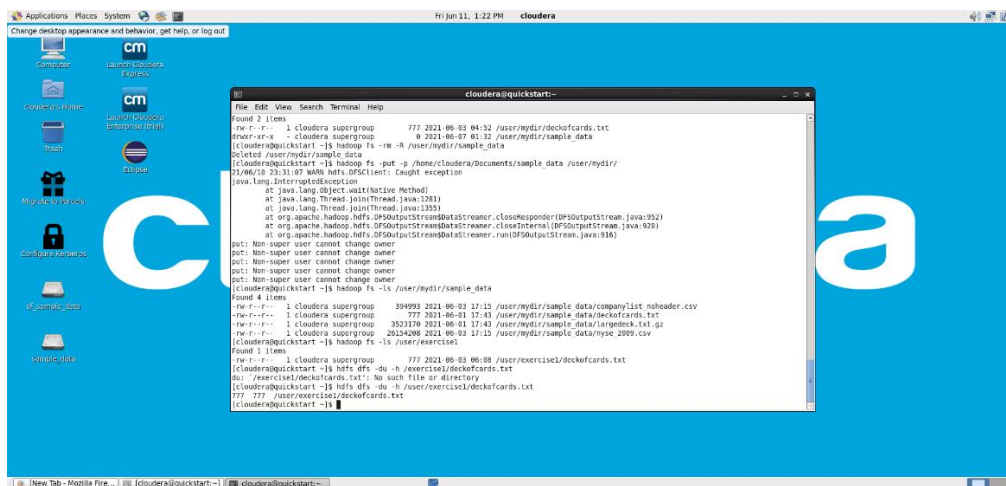
3. Upload GitHub repo sample\_data/deckofcards.txt to HDFS under /exercise1 directory



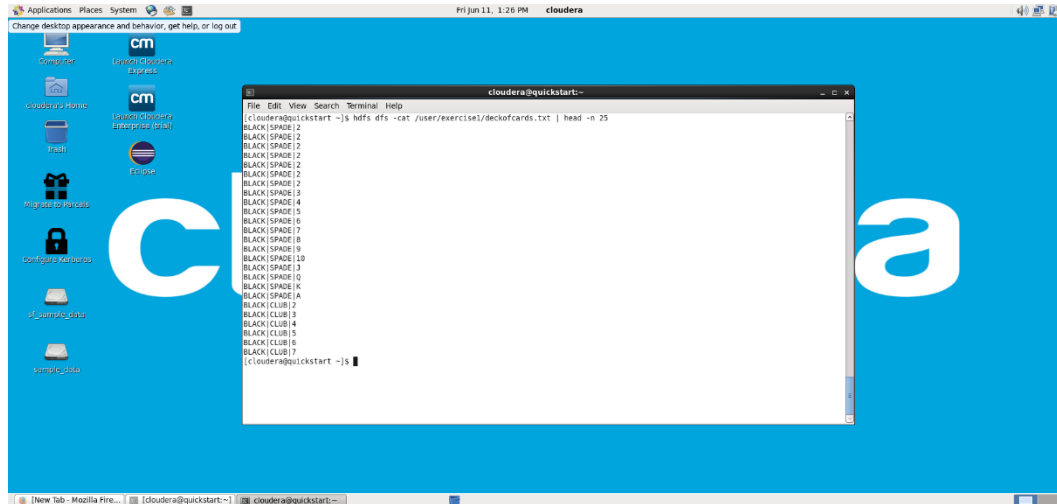
4. View the content of the /exercise1 directory



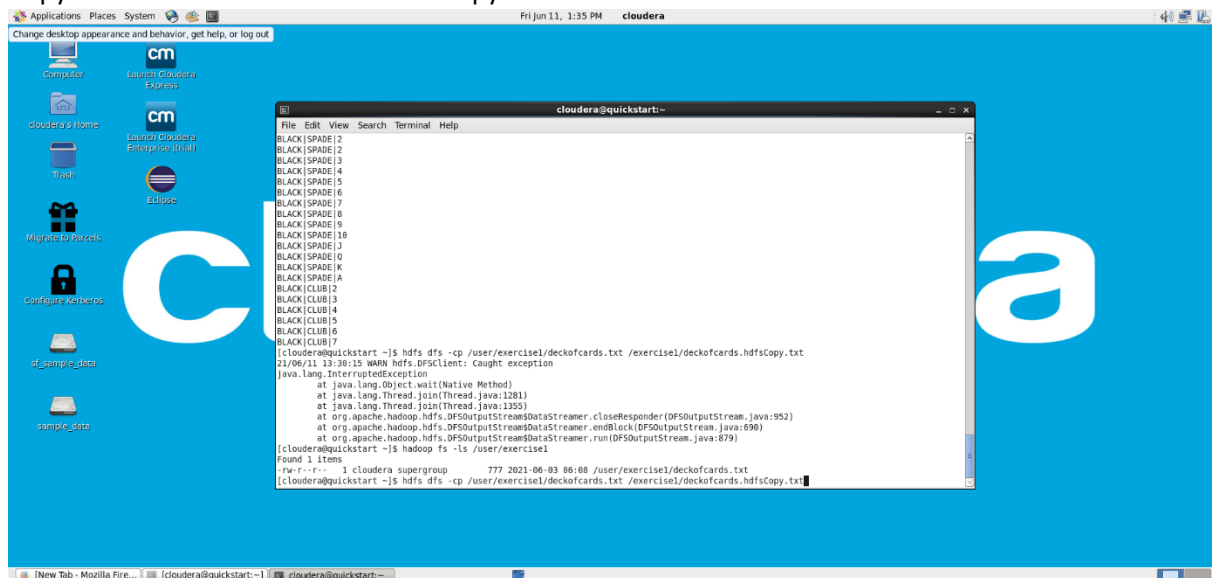
5. Determine the size of the hamlet.txt file in KB that resides on HDFS (not local directory)



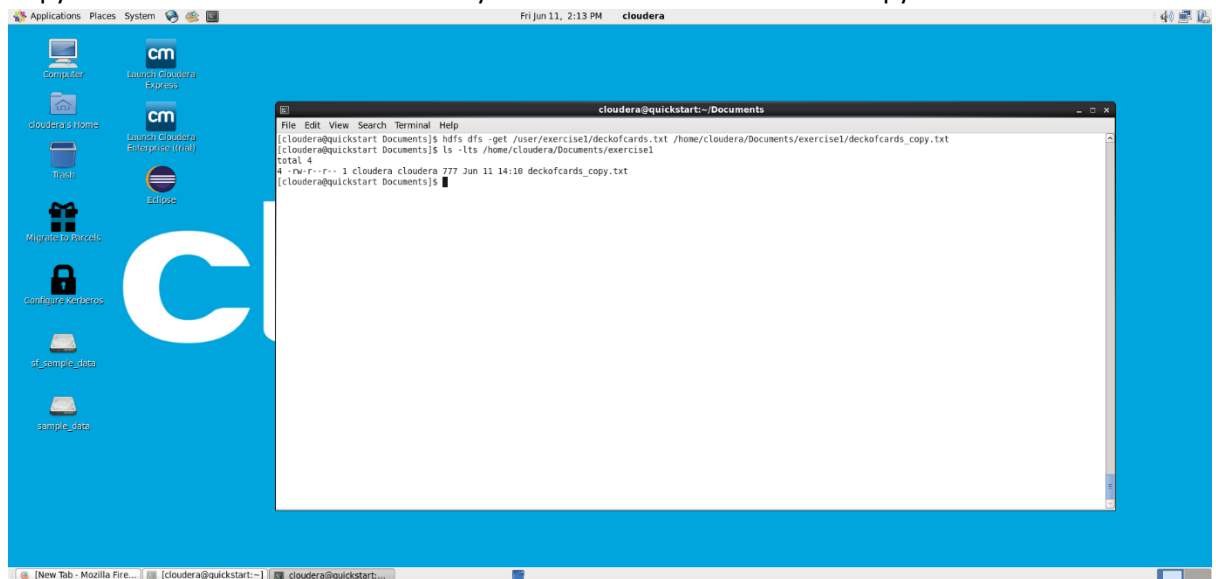
6. Print the first 25 lines to the screen from deckofcards.txt on HDFS



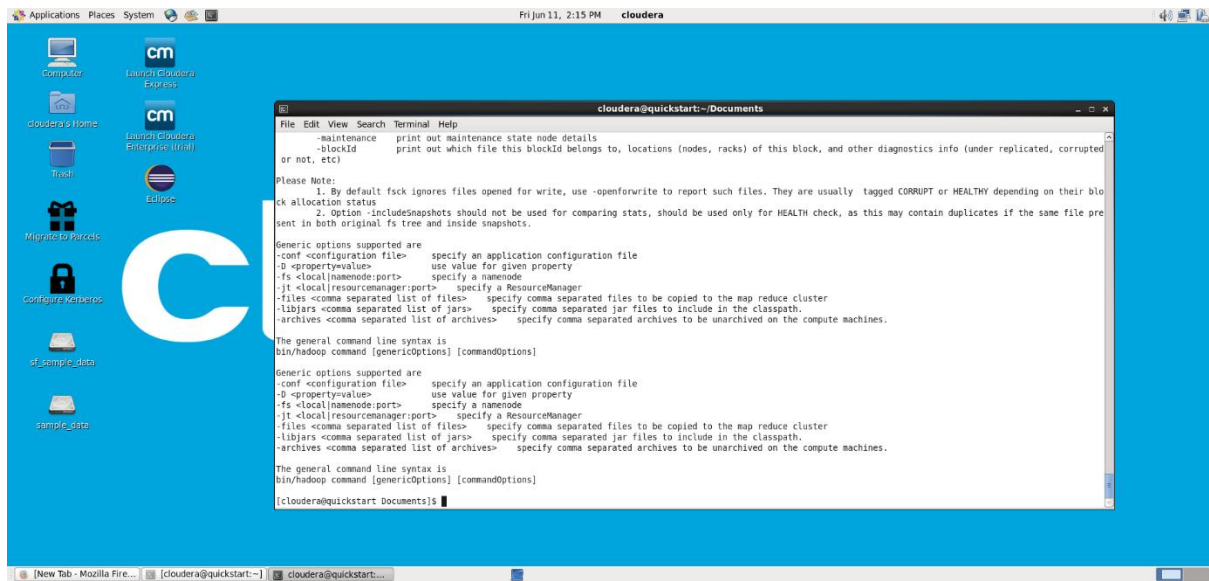
7. Copy deckofcards.txt to deckofcardsCopy.txt



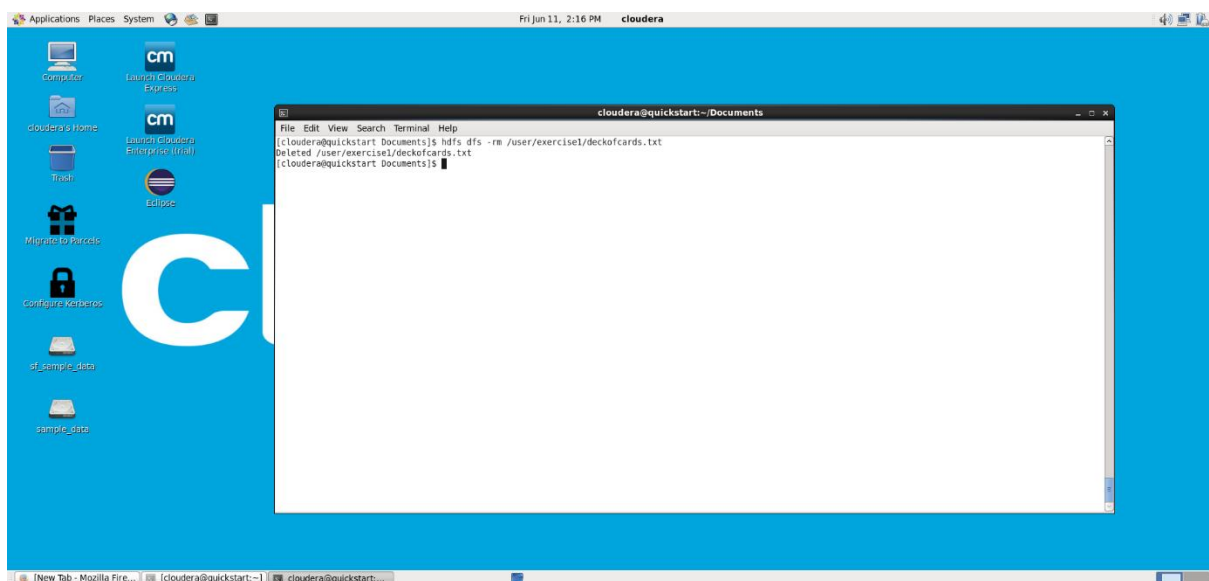
8. Copy deckofcards.txt back to local file system and name it deckofcards\_copy.txt



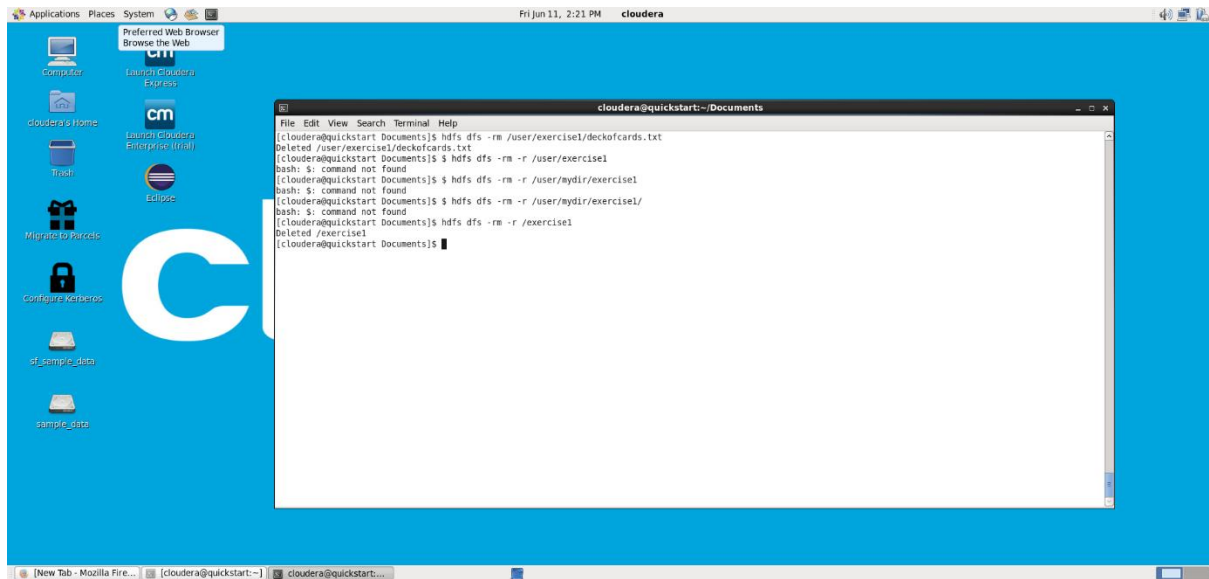
## 9. Check the entire filesystem for inconsistencies/problems



## 10. Delete deckofcards.txt from HDFS



## 11. Delete the /exercise1 directory from HDFS



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

