HDFS Task-1

Answer

1. What is the Namenode's URI and which file is it configured in?
     URI : Uniform resource identifier (URI) in HDFS to locate a file.
   it can be configured in core-site.xml file in "fs.default.name" name property

2. Where on a local file system will Namenode store its image and which file is it configured in?
  /opt/hadoop-2.6.0/Hadoop-data/dfs/data/name
  it is configured in "hdfs-site.xml" file
  <property>
     <name>dfs.namenode.name.dir</name>
     <value>file:///opt/hadoop-2.6.0/hadoop_data/dfs/name</value>
  </property>

3. Where on a local file system will Datanode store its blocks and which file is it configured in?
  /opt/hadoop-2.6.0/Hadoop-data/dfs/data/name
  it is configured in "hdfs-site.xml" file
  <property>
     <name>dfs.datanode.data.dir</name>
     <value>file:///opt/hadoop-2.6.0/hadoop_data/dfs/data</value>
  </property>

4. What is the block replication and which file is it configured in?
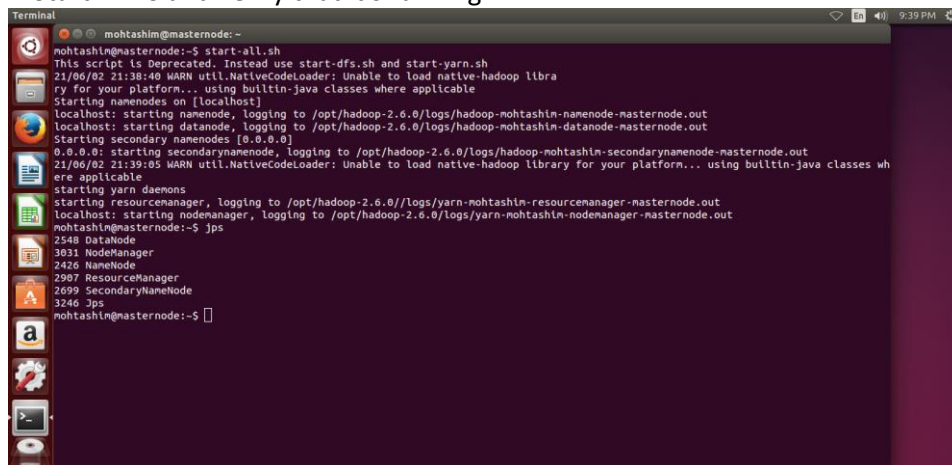 Block replication means no of copies of a data for fault tolerance
  it is configured in "hdfs-site.xml" file
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>

Perform

1. Start HDFS and verify that it's running

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)

```
21/06/02 21:40:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
Found 2 items
drwxr-xr-x   - mohtashim supergroup          0 2021-05-28 16:42 /MyDir
drwxr-xr-x   - mohtashim supergroup          0 2021-06-02 21:40 /exercise1
mohtashim@masternode:~$ sudo ifconfig eth0 up
[sudo] password for mohtashim:
mohtashim@masternode:~$ sudo dhclient eth0
mohtashim@masternode:~$ gedit

(gedit:3674): GLib-GObject-CRITICAL **: g_value_set_object: assertion 'G_VALUE_HOLDS_OBJECT (value)' failed

(gedit:3674): Gtk-CRITICAL **: _gtk_widget_captured_event: assertion 'WIDGET_REALIZED_FOR_EVENT (widget, event)' failed
mohtashim@masternode:~$ hdfs dfs -put sample_data/deckofcards.txt /exercise1
21/06/02 21:50:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
mohtashim@masternode:~$ hdfs dfs -ls /exercise1
21/06/02 21:51:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
Found 1 items
-rw-r--r--   1 mohtashim supergroup        693 2021-06-02 21:50 /exercise1/deckofcards.txt
mohtashim@masternode:~$ hdfs dfs -ls /
21/06/02 21:53:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
Found 2 items
drwxr-xr-x   - mohtashim supergroup          0 2021-05-28 16:42 /MyDir
drwxr-xr-x   - mohtashim supergroup          0 2021-06-02 21:50 /exercise1
mohtashim@masternode:~$ hdfs dfs -ls /exercise1
21/06/02 21:54:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
Found 1 items
-rw-r--r--   1 mohtashim_supergroup        693 2021-06-02 21:50 /exercise1/deckofcards.txt
mohtashim@masternode:~$
```

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


```
BLACK|SPADE|10
BLACK|SPADE|J
mohtashim@masternode:~$ hdfs dfs -cat /exercise1/deckofcards.txt | head -n 25
21/06/02 22:29:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
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
mohtashim@masternode:~$ ^C
mohtashim@masternode:~$
```

7. Copy deckofcards.txt to deckofcardsCopy.txt


```
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
mohtashim@masternode:~$ ^C
mohtashim@masternode:~$ hdfs dfs -cp /exercise1/deckofcards.txt   /exercise1/deckofcardsCopy.txt
21/06/02 22:30:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
mohtashim@masternode:~$ hdfs dfs -ls /exercise1
21/06/02 22:30:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Found 2 items
-rw-r--r--   1 mohtashim supergroup        693 2021-06-02 21:50 /exercise1/deckofcards.txt
-rw-r--r--   1 mohtashim_supergroup        693 2021-06-02 22:30 /exercise1/deckofcardsCopy.txt
mohtashim@masternode:~$
```

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

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
mohtashim@masternode:~$ ^C
mohtashim@masternode:~$ hdfs dfs -cp /exercise1/deckofcards.txt /exercise1/deckofcardsCopy.txt
21/06/02 22:30:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
mohtashim@masternode:~$ hdfs dfs -ls /exercise1
21/06/02 22:30:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Found 2 items
-rw-r--r--   1 mohtashim supergroup   693 2021-06-02 21:50 /exercise1/deckofcards.txt
-rw-r--r--   1 mohtashim supergroup   693 2021-06-02 22:30 /exercise1/deckofcardsCopy.txt
mohtashim@masternode:~$ hdfs dfs -get /exercise1/deckofcards.txt sample_data/deckofcards.copy.txt
21/06/02 22:32:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
mohtashim@masternode:~$ ls sample_data
deckofcards.copy.txt deckofcards.txt
mohtashim@masternode:~$

9. Check the entire filesystem for inconsistencies/problems

-libjars <comma separated list of jars>   specify comma separated jar files to include in the classpath.
-archives <comma separated list of archives>   specify comma separated archives to be unarchived on the compute machines.

The general command line syntax is
bin/hadoop command [genericOptions] [commandOptions]

mohtashim@masternode:~$ hdfs fsck /
21/06/02 22:36:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Connecting to namenode via http://localhost:50070
FSCK started by mohtashim (auth:SIMPLE) from /127.0.0.1 for path / at Wed Jun 02 22:36:39 IST 2021
...Status: HEALTHY
 Total size: 1415 B
 Total dirs: 3
 Total files: 3
 Total symlinks: 0
 Total blocks (validated):   3 (avg. block size 471 B)
 Minimally replicated blocks: 3 (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 02 22:36:39 IST 2021 in 27 milliseconds

The filesystem under path '/' is HEALTHY
mohtashim@masternode:~$

10. Delete deckofcards.txt from HDFS

21/06/02 22:37:42 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /exercise1/deckofcards.txt
mohtashim@masternode:~$ hdfs dfs -rm /exercise1/deckofcards.txt
21/06/02 22:38:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
rm: `/exercise1/deckofcards.txt': No such file or directory
mohtashim@masternode:~$ hdfs dfs -ls /
21/06/02 22:38:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Found 2 items
drwxr-xr-x   - mohtashim supergroup   0 2021-05-28 16:42 /MyDir
drwxr-xr-x   - mohtashim supergroup   0 2021-06-02 22:37 /exercise1
mohtashim@masternode:~$ hdfs dfs -ls /exercise1
21/06/02 22:38:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Found 1 items
-rw-r--r--   1 mohtashim supergroup   693 2021-06-02 22:30 /exercise1/deckofcardsCopy.txt
mohtashim@masternode:~$ hdfs dfs -cp /exercise1/deckofcardsCopy.txt /exercise1/deckofcards.txt
21/06/02 22:39:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
mohtashim@masternode:~$ hdfs dfs -ls /exercise1
21/06/02 22:39:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Found 2 items
-rw-r--r--   1 mohtashim supergroup   693 2021-06-02 22:39 /exercise1/deckofcards.txt
-rw-r--r--   1 mohtashim supergroup   693 2021-06-02 22:30 /exercise1/deckofcardsCopy.txt
mohtashim@masternode:~$ hdfs dfs -rm /exercise1/deckofcards.txt
21/06/02 22:39:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
21/06/02 22:39:42 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /exercise1/deckofcards.txt
mohtashim@masternode:~$

11. Delete the /exercise1 directory from HDFS

```
                        similar to 'rmdir a/b/c a/b a'
   -v, --verbose   output a diagnostic for every directory processed
        --help     display this help and exit
        --version  output version information and exit

Report rmdir bugs to bug-coreutils@gnu.org
GNU coreutils home page: <http://www.gnu.org/software/coreutils/>
General help using GNU software: <http://www.gnu.org/gethelp/>
For complete documentation, run: info coreutils 'rmdir invocation'
mohtashim@masternode:~$ hdfs dfs -rmdir /exercise1 | -v
-v: command not found
21/06/02 22:41:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
rmdir: `/exercise1': Directory is not empty
mohtashim@masternode:~$ hdfs dfs -rmr /exercise1
rmr: DEPRECATED: Please use 'rm -r' instead.
21/06/02 22:42:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
21/06/02 22:42:43 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /exercise1
mohtashim@masternode:~$ hdfs dfs -rmr /
rmr: DEPRECATED: Please use 'rm -r' instead.
21/06/02 22:42:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
21/06/02 22:42:59 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
rmr: `/': Input/output error
mohtashim@masternode:~$ hdfs dfs -ls /
21/06/02 22:43:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where a
pplicable
Found 1 items
drwxr-xr-x   - mohtashim_supergroup          0 2021-05-28 16:42 /MyDir
mohtashim@masternode:~$
```

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