

HDFS Task-1

Answer

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

Name node URI is default web UI which is 9000,50070 or 8020 .

hdfs://localhost:9000.

It is configured in core-site.xml.

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

Namenode uses file in its local host to store the EditLog.FsImage is stored as a file in the namenode local file system. It is configured in hdfs-site.xml

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

Datanode stores the actual data in HDFS and it is configured with disk space. It is configured in hdfs-site.xml

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

The blocks of file are replicated for fault tolerance. Default replication factor is 3 and block size is 128 MB. The replication factor is property set in the HDFS configuration file. Replication factor is configured in hdfs-site.xml

Perform

1. Start HDFS and verify that it's running

To start:

Start namenode -format

Start-dfs.sh

To verify:

https://localhost:9870

2. Create a new directory /exercise1 on HDFS

hadoop fs -mkdir/user/exercise1

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

hadoop fs -put/home/deckofcards.text/user/exercise1/

4. View the content of the /exercise1 directory

```
hadoop fs -ls/user/exercise1
```

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

```
hadoop fs -du[-s][-h] /user/exercise1/hamlet.txt
```

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

```
hadoop dfs -cat /user/exercise1/deckofcards.txt/head-25
```

7. Copy deckofcards.txt to deckofcardsCopy.txt

```
Hadoop fs -mv  
/user/exercise1/deckofcards.txt/user/exercise1/deckofcardsCopy.txt
```

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

```
hadoop fs -cp  
/user/exercise1/deckofcards.txt/user/exercise1/deckofcardsCopy.txt
```

9. Check the entire filesystem for inconsistencies/problems

```
Fsck command to check file inconsistencies  
hdfs fsck/
```

10. Delete deckofcards.txt from HDFS

```
hadoop fs -rm/user/exercise1/deckofcards.txt
```

11. Delete the /exercise1 directory from HDFS

```
Hadoop fs -rm -r/user/exercise1
```

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