

HDFS Task-1

Answer

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

Ans: Name node is the master and management node of Hadoop system.in core-site.xml file

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

Ans: FSImage is where Namenode stores its Image and it is stored in XML.config

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

Ans:In Hadoop home etc folder the hdfs-site.xml file is there in that file data node property is configured

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

Ans: Its a copy of the data of node of rack, each data should be placed in different nodes of different rack.It is configured in hdfs-site.xml

Perform:

1. Start HDFS and verify that it's running ?

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs supergroup 0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup 0 2021-06-02 11:36 /hbase
-rw-r--r-- 1 cloudera supergroup 726663168 2021-06-02 11:39 /largedeck.txt
drwxr-xr-x - cloudera supergroup 0 2021-06-02 11:47 /result1
drwxr-xr-x - solr solr 0 2017-07-19 05:37 /solr
drwxrwxrwt - hdfs supergroup 0 2021-05-25 07:17 /tmp
drwxr-xr-x - hdfs supergroup 0 2021-05-27 09:18 /user
drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /var
[cloudera@quickstart ~]$
```

2. Create a new directory /exercise1 on HDFS ?

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/exercise1;
[cloudera@quickstart ~]$ hdfs dfs -ls /user
Found 11 items
drwxr-xr-x - cloudera cloudera          0 2017-07-19 05:33 /user/cloudera
-rw-r--r-- 1 cloudera supergroup        693 2021-05-27 09:18 /user/deckofcards.txt
drwxr-xr-x - cloudera supergroup          0 2021-06-03 08:20 /user/exercise
drwxr-xr-x - cloudera supergroup          0 2021-06-03 08:20 /user/exercise1
drwxr-xr-x - mapred  hadoop              0 2017-07-19 05:34 /user/history
drwxrwxrwx - hive    supergroup          0 2017-07-19 05:36 /user/hive
drwxrwxrwx - hue     supergroup          0 2017-07-19 05:35 /user/hue
drwxrwxrwx - jenkins supergroup          0 2017-07-19 05:35 /user/jenkins
drwxrwxrwx - oozie   supergroup          0 2017-07-19 05:35 /user/oozie
drwxrwxrwx - root    supergroup          0 2017-07-19 05:35 /user/root
drwxr-xr-x - hdfs    supergroup          0 2017-07-19 05:36 /user/spark
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Downloads/deckofcards.txt /user/exercise1
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise 1
ls: `1': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise1
Found 1 items
-rw-r--r-- 1 cloudera supergroup        695 2021-06-03 08:26 /user/exercise1/deckofcards.txt
[cloudera@quickstart ~]$
```

4. View the content of the /exercise1 directory

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Downloads/deckofcards.txt /user/exercise1
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise 1
ls: `1': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise1
Found 1 items
-rw-r--r-- 1 cloudera supergroup        695 2021-06-03 08:26 /user/exercise1/deckofcards.txt
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercise1/deckofcards.txt
-rw-r--r-- 1 cloudera supergroup        695 2021-06-03 08:26 /user/exercise1/deckofcards.txt
[cloudera@quickstart ~]$
```

695-Bytes

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

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/exercisel/deckofcards.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
```

7. Copy deckofcards.txt to deckofcardsCopy.txt

```
[cloudera@quickstart ~]$ hdfs dfs -cp /user/exercisel/deckofcards.txt /user/exercisel/deckofcardsCopy.txt
21/06/03 09:13:43 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:952)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:690)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:879)
[cloudera@quickstart ~]$ hdfs dfs ls /user/exercisel
ls: Unknown command
Did you mean -ls? This command begins with a dash.
[cloudera@quickstart ~]$ hdfs dfs -ls /user/exercisel
Found 2 items
-rw-r--r-- 1 cloudera supergroup 695 2021-06-03 08:26 /user/exercisel/deckofcards.txt
-rw-r--r-- 1 cloudera supergroup 695 2021-06-03 09:13 /user/exercisel/deckofcardsCopy.txt
[cloudera@quickstart ~]$
```

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

```
[cloudera@quickstart ~]$ hdfs dfs -get /user/exercisel/deckofcards.txt /home/cloudera/Downloads/deckofcardsCopy.txt
[cloudera@quickstart ~]$ ls /home/cloudera/Downloads
l_spark_overview.pdf  deckofcardsCopy.txt  deckofcards.txt~  Hive Command  largedeck.txt (2)  META-INF
cardstest-1.0-SNAPSHOT.jar  deckofcards.txt  Hadoop.docx  largedeck.txt  largedeck.txt.gz  org
[cloudera@quickstart ~]$
```

9. Check the entire filesystem for inconsistencies/problems

```
[cloudera@quickstart ~]$ hdfs fsck /user/exercise1/deckofcards.txt
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&path=%2Fuser%2Fexercise1%2Fdeckofcards.txt
FSCK started by cloudera (auth:SIMPLE) from /10.0.2.15 for path /user/exercise1/deckofcards.txt at Thu Jun 03 09:24:50 PDT 2021
1
Status: HEALTHY
Total size:      695 B
Total dirs:      0
Total files:     1
Total symlinks:  0
Total blocks (validated): 1 (avg. block size 695 B)
Minimally replicated blocks: 1 (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 Thu Jun 03 09:24:50 PDT 2021 in 10 milliseconds

The filesystem under path '/user/exercise1/deckofcards.txt' is HEALTHY
[cloudera@quickstart ~]$ █
```

10. Delete deckofcards.txt from HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -rm /user/exercise1/deckofcards.txt
Deleted /user/exercise1/deckofcards.txt
[cloudera@quickstart ~]$ █
```

11. Delete the /exercise1 directory from HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -rm -r /user/exercise1
Deleted /user/exercise1
[cloudera@quickstart ~]$ █
```

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

Task 2 :

1. What is the default replication factor of Hadoop cluster?

- a. 3
- b. 2
- c. 4
- d. 1

Ans:- a

2. Which component in Hadoop Cluster is responsible for serving read and write requests from the file system's clients?

- a. Name Node
- b. Data Node

- c. Both a & b
- d. None of the above

Ans:- b

3. Which component of Hadoop Cluster manages the file system namespace and regulates access to files by clients?

- a. Name Node
- b. Data Node
- c. Both a & b
- d. None of the above

Ans:- a

4. If a file size of size 100 MB is stored on HDFS, what would be the split size?

- a. 64 MB & 64 MB
- b. 64 MB & 36 MB
- c. 100 MB
- d. None of the above

Ans:- c

5. State true or false: ?

- a. FALSE
- b. True

Ans:- a

6. Which comand of HDFS helps copy files from HDFS to Local file system?

- a. copyFromLocal
- b. copyToLocal
- c. put
- d. Mv

Ans:- b or get

7. Which Eco system component of Hadoop is good for non sql programmers?

- a. Hive
- b. Hbase
- c. Flume
- d. Pig

Ans:- b

8. Block size of a Hadoop cluster is configurable by Administrator?

- a. TRUE
- b. FALSE

Ans:- a

9. The functions performed by DataNodes in Hadoop Cluster is/are?

- a. Data Block Creation
- b. Data Block Deletion
- c. Data Block Replication
- d. All above

Ans:- d

10. Find error in below command:

hdfs dfs -put /home/user1/abc.txt

- a. Target name missing
- b. Source name should include hdfs://
- c. No error

Ans:- a

11. Hadoop block size should be multiple of which unit?

- a. 32 MB
- b. 50 MB
- c. 64 MB
- d. 70 MB

Ans:- c in old new version has 128mb

12. Which component of the hadoop cluster manages data on slave nodes?

- a. Name node
- b. Data Node
- c. Task Tracker
- d. Job Tracker

Ans:- b

13. MR1 and MR2 are two modes of processing in Hadoop?

- a. TRUE
- b. FALSE

Ans:- a

14. What is Hadoop?

- a. Open source software for reliable, scalable, distributed computing.
- b. A framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models.
- c. Both a & b
- d. None of the above

Ans:- c

15. Hadoop provides

- a. A reliable distributed storage and processing system
- b. Only distributed storage
- c. Only processing system
- d. None of the above

Ans:- a