Lab 18.1: Using HDFS Snapshots

|  |  |
| --- | --- |
| **Objective:** | Understand how snapshots work in Hadoop. |
| **Successful Outcome:** | The data folder has a snapshot taken of it. |
| **Before You Begin:** | SSH into node1. |

1. Load a File into HDFS
   1. Start by creating a new directory in HDFS:

# hadoop fs -mkdir data

* 1. Put a file in the new directory:

# hadoop fs -put ~/labs/constitution.txt data/

* 1. Run the fsck command on the file:

# hdfs fsck /user/root/data/constitution.txt -files -blocks -locations

Select-and-copy the block ID.

* 1. Use the find command to identify the location of the block on your local file system. The command will look like the following, and you may need to run it on node2 or node3:

# find / -name "blk\_ 1073743186"

* 1. Which node and folder is the block stored in? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Enable Snapshots
   1. Now let’s enable the /user/root/data directory for taking snapshots:

# su -l hdfs -c "hdfs dfsadmin -allowSnapshot /user/root/data"

You should see the following confirmation:

Allowing snaphot on /user/root/data succeeded

1. Create a Snapshot
   1. Now create a snapshot of /user/root/data:

# hdfs dfs -createSnapshot /user/root/data ss01

You should see the following confirmation:

Created snapshot /user/root/data/.snapshot/ss01

* 1. Verify the snapshot was created by viewing the contents of the data/.snapshot folder:

# hadoop fs -ls -R data/.snapshot

drwxr-xr-x - root hadoop 0 data/.snapshot/ss01

-rw-r--r-- 3 root hadoop 44841 data/.snapshot/ ss01/constitution.txt

* 1. Try to delete the data folder:

# hadoop fs -rm -R data

You cannot delete the folder. Why not? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Delete the File
   1. Delete the constitution.txt file in data:

# hadoop fs -rm data/constitution.txt

* 1. Use the ls command to verify the file is no longer in the data folder in HDFS.
  2. Check whether the file still exists in /user/root/data/.snapshot/ss01. It should still be there.
  3. Run the same find command again that you ran in the earlier step. Does the block file still exist on your local file system? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Recover the File
   1. Let’s copy this file from data/.snapshot/ss01 to the data directory.

# hadoop fs -cp data/.snapshot/ss01/constitution.txt data/

* 1. Run the fsck command again on data/constitution.txt. Notice that the block and location information have changed for this file.
  2. Run the find command for the new blocks. Notice the blocks for the constitution.txt file appear in two locations on your local file system (before deleting the file and after copying the file).

**RESULT**: This lab demonstrates how the snapshot process locks down the blocks from deleting and editing, and the blocks are always available in case you need to recover your file in future.

Answers:

Step 1.5: In a subfolder of: /hadoop/hdfs/data/current/

Step 3.3: The command should read:

# hadoop fs -rm –R –skipTrash data

Step 4.4: Yes