Lab : Access HDFS in Terminal

|  |  |
| --- | --- |
| **Exercise Path** | ~/materials/data/html\_site |

**In this exercise, we will simply explore the HDFS file system install on our virtual machines, using terminal.**

**Uploading to HDFS**

Start by creating a new top-level directory for exercises. You will use this directory throughout the rest of the course.

$ hdfs dfs -mkdir lanier

Change directories to the local filesystem directory containing the sample data we will be using in the course.

$ cd ~/materials/data

$ ls

1. Insert this directory into HDFS:

$ hdfs dfs -put html\_site lanier

This copies the local html\_site directory and its contents into a remote HDFS directory named lanier/html\_site. Notice we do not “cd” into HDFS.

2. List the contents of the new HDFS directory now:

$ hdfs dfs -ls lanier/html\_site

You should see the html\_site articles that were in the local directory.

NOTE: In HDFS, relative (non-absolute) paths are considered

relative to your home directory. There is no concept of a “current” or

“working” directory as there is in Linux and similar file systems.

3. Practice uploading a directory, confirm the upload, and then remove it, as it is not actually needed for the exercises.

$ hdfs dfs -put calllogs/ lanier/

$ hdfs dfs -ls lanier/calllogs

$ hdfs dfs -rm -r lanier/calllogs

**Viewing HDFS Files**

Now view some of the data you just copied into HDFS.

1. Enter:

$ hdfs dfs -cat lanier/html\_site/KBDOC-00289.html | head

-n 20

This prints the first 20 lines of the article to your terminal.

2. To download a file to work with on the local filesystem use the

hdfs dfs -get command. This command takes two arguments: an HDFS path and a local path. It copies the HDFS contents into the local filesystem:

$ hdfs dfs -get lanier/html\_site/ KBDOC-00289.html ~/article.html

$ less ~/article.html

Enter the letter [q] to quit the less command after reviewing the downloaded

file.

There are several other operations available with the hdfs dfs command to perform most common filesystem manipulations such as mv, cp, and mkdir. In the terminal window, enter:

$ hdfs dfs

You see a help message describing all the file system commands provided by

HDFS. Try playing around with a few of these commands if you like.

**END**