Lab # : 2; Lab Name : An Introduction to HDFS ; Subject Name : Information Storage and Retrieval; Week #: 2; Lab Duration : 20 to 30 mins

# 

# **Intro**

In this lab, we will go over a few features of data loading and handling data or files in HDFS. We will go over the following features:

* How to see what files are present in HDFS
* How to mkdir in HDFS
* How to load/remove a file
* How to find size of a file
* How to see all commands available
* How to see help for a specific command

# 

Here are a few points to get to know HDFS :

* HDFS stands for Hadoop Distributed File System.
* Data is stored in HDFS in a Hadoop System.
* HDFS can store data in various file formats like RC, ORC, Parquet, etc.
* Virtually all interactions with Hadoop involve HDFS directly or indirectly.

# **Let’s go!**

# **Step-1. How to see what files are present in HDFS**

The following command is used on a daily basis by a Hadoop developer to check on the files, permissions, etc.

hadoop dfs -ls /yourhdfsdirectory

# **Step-2. How to mkdir**

# To get organized like we do with local file systems, we can create directories in HDFS folders as well. We can change permissions on those using ‘chmod’ unix command as well.

# 

# 

# 

# hadoop dfs -mkdir /hdfslocation/yourdirectory

# hadoop dfs -ls /hdfslocation/yourdirectory

# **Step-3. How to load/remove a file**

This is to load a data file/s manually or for automated processes.

# hadoop dfs -put /localdir/yourfile /hdfsdirectory/

# hadoop dfs -rm /hdfsdirectory/yourfile

# 

You can use unix ‘rmr’ option as well to remove recursively content from a hdfs directory.

Here are a few useful commands to move data from HDFS to Local filesystem.

# hadoop dfs -get /hdfsdirectory/yourfile /localdir/

# hadoop dfs -getmerge /hdfsdirectory/ /localdir/Mygetmergedfiles.txt

# 

# **Step-4. How to find size of a file/directory**

# hadoop dfs -du /hdfsdirectory/yourfile

# hadoop dfs -dus /hdfsdirectory/yourfile

# 

# **Step-5. How to see all hadoop commands available**

# hadoop dfs -help

# **Step-6. How to see help for a specific command**

hadoop dfs -help dus

Here is very useful link : http://hadoop.apache.org/docs/r1.2.1/file\_system\_shell.html.

**Questions:**

Q1 : What is the basic difference between RDBMS and Hadoop?

Q2 : List out the components of Hadoop?

Q3 : What are the processes in Hadoop FW?

Q4 : What kind of replication is present for hdfs?

Q5 : Is Hadoop better for datafiles with small size?