Computer Science & Information Systems

**Big Data Systems – Hadoop Lab Sheet 2**

**HDFS Operations**

1. Objective:

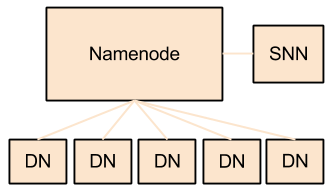
Students should be able to

1. Get familiarity with the Hadoop file system
2. Get hands-on experience with HDFS commands

HDFS is structured similarly to a regular UNIX filesystem except that data storage is distributed across several machines. It is not intended as a replacement to a regular filesystem, but rather as a filesystem-like layer for large distributed systems to use. It has in built mechanisms to handle machine outages, and is optimized for throughput rather than latency.

There are two types of machine in a HDFS cluster:

* Datanode - where HDFS actually stores the data, there are usually quite a few of these.
* Namenode - the ‘master’ machine. It controls all the Meta data for the cluster. E.g. - what blocks make up a file, and what data nodes those blocks are stored on.
* Secondary Namenode - this is NOT a backup name node, but is a separate service that keeps a copy of both the edit logs, and filesystem image, merging them periodically to keep the size reasonable.



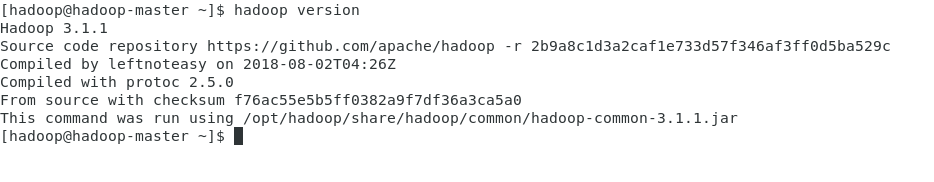
Data can be accessed using either the Java API, or the Hadoop command line client. Many operations are similar to their UNIX counterparts.

1. Steps to be performed:
2. Checking version of Hadoop with “version”

“version” command can be used to check the version of Hadoop installation.

Hadoop HDFS version Command Example

* hadoop version



1. Checking the documentation for HDFS commands with “help”

Returns usage information for the commands listed.

Hadoop HDFS help Command Example

* hadoop help

1. Creating the directory with “mkdir” command

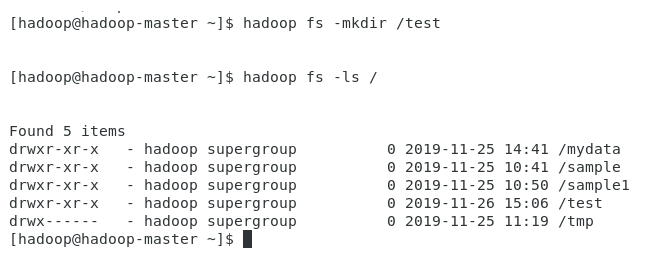
This HDFS command takes path URI’s as an argument and creates directories. Creates any parent directories in path that are missing.

Hadoop HDFS mkdir Command Usage

* mkdir <path>

Hadoop HDFS mkdir Command Example

* hadoop fs -mkdir /test



1. Listing the contents of directory with “ls”

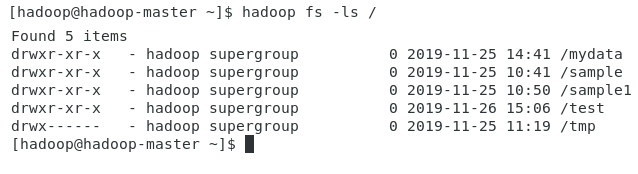
This command is used to list all the files. Use “lsr” for recursive approach. It is useful when we want a hierarchy of a folder.

Hadoop HDFS mkdir Command Usage

* hadoop fs -ls <path>

Hadoop HDFS mkdir Command Example

* hadoop fs -ls /



1. Copying file from local file system to HDFS with “put” or “copyFromLocal”

To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.

Hadoop HDFS ls Command Usage

hadoop fs -copyFromLocal <local file path> <destination(present on hdfs)>

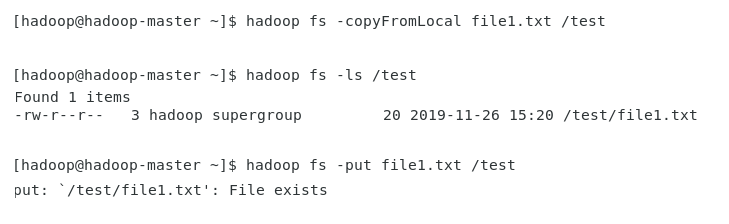
Hadoop HDFS ls Command Example

Let’s suppose we have a file file1.txt on local file system which we want to copy to folder "test" present on hdfs.

hadoop fs -copyFromLocal file1.txt /test

(OR)

hadoop fs -put file1.txt /test



1. Copying file from HDFS to local file system with “get” or “copyToLocal”

To copy files/folders from hdfs store to local file system.

Hadoop HDFS ls Command Usage

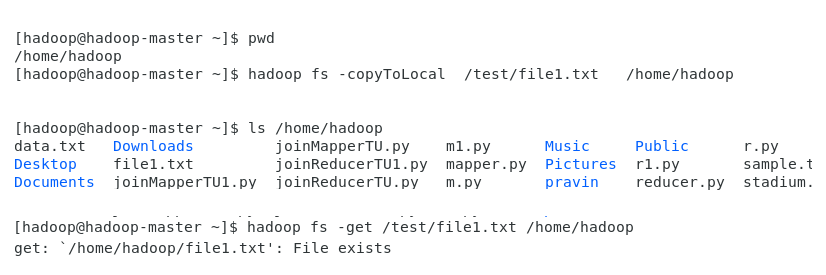
hadoop fs -copyToLocal <<source file(on hdfs)> <local file destination>

Hadoop HDFS ls Command Example

hadoop fs -copyToLocal /test/file1.txt /home/hadoop

(OR)

hadoop fs fs -get /test/file1.txt /home/hadoop



1. Listing the contents of HDFS file with “cat”

To print file contents.

Hadoop HDFS ls Command Usage

hadoop fs -cat <path>

Hadoop HDFS ls Command Example

hadoop fs -cat /test/file1.txt



1. Copying files within HDFS with “cp”

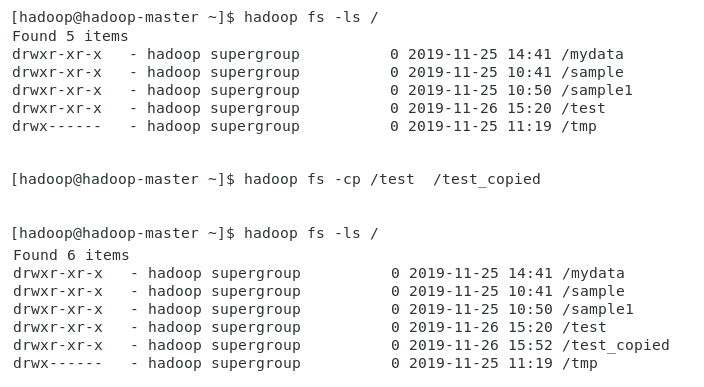
This command is used to copy files within hdfs. Lets copy folder test to test\_copied.

Hadoop HDFS ls Command Usage

hadoop fs -cp <soure directory(on hdfs)> <destination directory(on hdfs)>

Hadoop HDFS ls Command Example

hadoop fs -cp /test /test\_copied



1. Removing the directory with “rm”

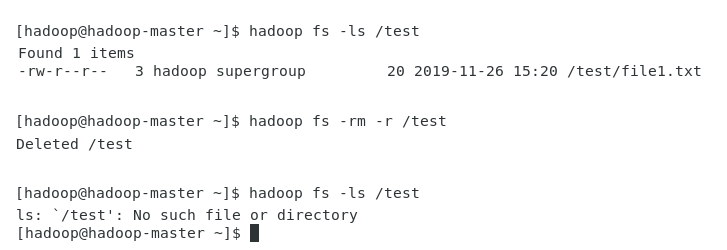
This command deletes a file from HDFS recursively. It is very useful command when you want to delete a non-empty directory.

Hadoop HDFS ls Command Usage

hadoop fs -rmr <filename/directoryName>

Hadoop HDFS ls Command Example

hadoop fs -rmr /test



1. Outputs/Results:

Students should be able to use HDFS commands for

* Listing the directories
* Creating the directory
* Copying file from local file system to HDFS
* Copying file from HDFS to local file system
* Listing the contents of HDFS file
* Copying files within HDFS
* Removing the directory

1. Observations:

Students carefully needs to observe the syntax the HDFS commands and use it properly.

1. References:
2. [HDFS Commands Guide](https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html)
3. [TutorialPoint Hadoop - Command Reference](https://www.tutorialspoint.com/hadoop/hadoop_command_reference.htm)
4. [HDFS Command Examples](https://blog.matthewrathbone.com/2013/04/17/what-is-hadoop.html)