HDFS Commands

DYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTE

See List of Commands

```
hadoop fs
hdfs dfs
```

To get help

```
hadoop fs -help ls
```

TRENDYTECH - TRENDYTECH - TRENDYTECH

Listing Files & Directories

ls /home/cloudera (list all the files in local system)

hadoop fs -ls /user/cloudera (list the files in hdfs)

Note:

The home directory in local is /home/cloudera

The home directory in hdfs is /user/cloudera

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

How does HDFS commands work?

All the hdfs commands will interact with namenode and show the results.

Example: ls

however some commands will also go to the datanodes to read the data from files.

Ex: cat, tail

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

ls command (various sorting techniques)

Sort by Name

```
hadoop fs -ls /
```

sort by time in reverse order

```
hadoop fs -ls -t -r /
```

sort by size, by default largest is at the top

```
hadoop fs -ls -S /
```

sort by size, size will be displayed in human readable form

```
hadoop fs -ls -S -h /
```

TRENDYTECH - UPLIFT YOUR CAREER!

A few more ls commands

list recursively

hadoop fs -ls -R /

search in the list using grep

hadoop fs -ls /user | grep cloudera

home directory for a user in hdfs is /user/cloudera

hadoop fs -ls (or) hadoop fs - ls /user/cloudera

TRENDYTECH - UPLIFT YOUR CAREER!

Creating a new directory

Creating a new directory inside you home

hadoop fs -mkdir /user/cloudera/testing

Creating a hierarchy of directories

hadoop fs -mkdir -p /user/cloudera/folder1/folder2

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Removing a file or directory

rm can be used to remove files (but not directories)

hadoop fs -rm /user/cloudera/file1.txt

To remove a directory we have to give -R

hadoop fs -rm -R /user/cloudera/testing

rmdir can only remove empty directories

hadoop fs -rmdir /user/cloudera/folder1 (this will give error as folder1 directory is not empty)

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Copying files or folder from local to hdfs

2 commands used for this purpose are:

copyFromLocal (or) put

Step 1: hadoop fs -mkdir /data (creating hdfs dir)

Step 2: create file1.txt on local Desktop

Step 3: Run the below command

hadoop fs -copyFromLocal Desktop/file1.txt /data

TRENDYTECH - TRENDYTECH - TRENDYTECH

Copying files or folder from local to hdfs

copy folder from local to hdfs

Scenario 1: destination folder does not have the same name

```
hadoop  fs  -copyFromLocal  Desktop/folder1  /data
```

Scenario 2: destination folder has the same name

```
hadoop fs -copyFromLocal Desktop/folder1  /data/folder1
```

TRENDYTECH - TRENDYTECH - TRENDYTECH

Copying files or folder from hdfs to local

2 commands used for this purpose are:

copyToLocal (or) get

hadoop fs -copyToLocal <hdfs path> <local path>

Or

hadoop fs -get <hdfs path> <local path>

TRENDYTECH - UPLIFT YOUR CAREER!

To view first few lines or last few lines

tail to view last 10 lines

hadoop fs -tail <hdfs filepath>

head command does not work to see first 10 lines

We can use more command

hadoop fs -cat <hdfs filepath> | more

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Copy files/folders from one hdfs location to other

We can use cp command for this

hadoop fs -cp <hdfs file path> <hdfs location2>

So basically hdfs file path is the path where the file exists in hdfs.

hdfs location2 is the new path in hdfs where we want to move the file.

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Cut paste the files/folders from one location in hdfs to other

This is actually nothing but renaming a file.

So just the metadata needs to be updated and this is very quick.

hadoop fs -mv <hdfs file path> <hdfs location2>

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Check the free disk space

df command is used to check the free disk space.

hadoop fs -df -h /user/cloudera

Here -h stands for human readable. If we do not give this, then it will show in bytes.

TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Check the disk usage

du command is used to check the disk used.

```
hadoop fs -du -h /user/cloudera
```

Here -h stands for human readable. If we do not give this, then it will show in bytes.

To summarize the results use -s

```
hadoop fs -du -s -h /user/cloudera
```

TRENDYTECH - UPLIFT YOUR CAREER!

Dynamically change the replication factor

We need to set the below property:

dfs.replication=5 (we can give any number here)

This below command will make more sense in cloudxlab:

hadoop fs -Ddfs.replication=5 -put loan_stats_3c.csv /user/crazycodingteam6223

TRENDYTECH - TRENDYTECH - TRENDYTECH

Use fsck command to see metadata in hdfs

fsck stands for filesystem check

hdfs fsck /user/crazycodingteam6223/loan_stats_3c.csv -files -blocks -locations

The above command will give the block information and replication information along with the ip address of datanodes where the block are kept.

TECH - TRENDYTECH - TRENDYTECH - TRENDYTECH - TRENDYTECH

Thank you!

Google REVIEWS

5 Star Google Rated
Big Data Course

LEARN FROM THE EXPERT

9108179578

Call for more details

Follow US

Trainer	Mr. Sumit Mittal
Phone	9108179578
Email	trendytech.sumit@gmail.com
Website	https://trendytech.in/courses/big-data-online-training/
LinkedIn	https://www.linkedin.com/in/bigdatabysumit/
Twitter	@BigdataBySumit
Instagram	bigdatabysumit
Facebook	https://www.facebook.com/trendytech.in/
Youtube	TrendyTech