BIG DATA
EXPERIENCE

Python for Big Data

HDFS Demo

Asst. Prof. Dr. Santitham Prom-on

Department of Computer Engineering, Faculty of Engineering
King Mongkut's University of Technology Thonburi

BIG DATA
EXPERIENCE
CENTER

FIRST
LOGIC

KMUTT g-able

HDFS: Put file from Local into HDFS

```
[mea-de@debootcamp ~]$ hadoop fs -put myfile.txt
[mea-de@debootcamp ~]$
```

Command

hadoop fs –put <filename> <destination>

Options

-f : Overwrite if already exists

HDFS: List files/folders

```
[mea-de@debootcamp ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 myfile.txt
[mea-de@debootcamp ~]$
```

Command

hadoop fs –ls <path>

Options

-R : Recursive (also visit all subdirectories)
-h : Human-readable file size formatting
-S : Sort result by file size (-r to reverse)
-t : Sort result by modification time

HDFS: Create new directory

```
[mea-de@debootcamp ~]$ hadoop fs -mkdir my_dir
[mea-de@debootcamp ~]$ hadoop fs -ls
Found 3 items
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:41 my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 myfile.txt
[mea-de@debootcamp ~]$
```

Command

hadoop fs –mkdir <dir_name>

Options

-p : Creating parent directories along the path

HDFS: Move file/directory

```
[mea-de@debootcamp ~]$ hadoop fs -ls -R
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:41 my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 myfile.txt
[mea-de@debootcamp ~]$ hadoop fs -mv myfile.txt my_dir/
[mea-de@debootcamp ~]$ hadoop fs -ls -R
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:45 my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 my_dir/myfile.txt
[mea-de@debootcamp ~]$
```

Command

Options

`hadoop fs -mv <source> <destination>`

HDFS: Copy file/directory

```
[mea-de@debootcamp ~]$ hadoop fs -ls -R
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:45 my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 my_dir/myfile.txt
[mea-de@debootcamp ~]$ hadoop fs -cp my_dir/myfile.txt myfile_copy.txt
[mea-de@debootcamp ~]$ hadoop fs -ls -R
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:45 my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 my_dir/myfile.txt
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:47 myfile_copy.txt
[mea-de@debootcamp ~]$
```

Command

Options

`hadoop fs -cp <source> <destination>`

HDFS: Delete file/directory

```
[mea-de@debootcamp ~]$ hadoop fs -ls -R
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:45 my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 my_dir/myfile.txt
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:47 myfile_copy.txt
[mea-de@debootcamp ~]$ hadoop fs -rm -r my_dir
21/04/18 16:52:48 INFO fs.TrashPolicyDefault: Moved: 'hdfs://debootcamp.meavm:8020/user/mea-de/my_dir' to trash
s://debootcamp.meavm:8020/user/mea-de/.Trash/Current/user/mea-de/my_dir
[mea-de@debootcamp ~]$ hadoop fs -ls -R
drwx------   - mea-de supergroup          0 2021-04-18 16:52 .Trash
drwx------   - mea-de supergroup          0 2021-04-18 16:52 .Trash/Current
drwx------   - mea-de supergroup          0 2021-04-18 16:52 .Trash/Current/user
drwx------   - mea-de supergroup          0 2021-04-18 16:52 .Trash/Current/user/mea-de
drwxr-xr-x   - mea-de supergroup          0 2021-04-18 16:45 .Trash/Current/user/mea-de/my_dir
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:39 .Trash/Current/user/mea-de/my_dir/myfile.txt
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:47 myfile_copy.txt
[mea-de@debootcamp ~]$
```

Command

hadoop fs –rm <path>

Options

-r : Recursive (also delete files/subfolders)
-skipTrash : Force permanent deletion

HDFS: Read content of the file

```
[mea-de@debootcamp ~]$ hadoop fs -cat myfile_copy.txt
Hello World!
[mea-de@debootcamp ~]$ hadoop fs -tail myfile_copy.txt
Hello World!
[mea-de@debootcamp ~]$
```

Command

```
hadoop fs –cat <path>
hadoop fs –tail <path>
```

Options

HDFS: Count number of file/directory

```
[mea-de@debootcamp ~]$ hadoop fs -count -v -h /user/mea-de
   DIR_COUNT   FILE_COUNT       CONTENT_SIZE PATHNAME
           7             2                 26 /user/mea-de
[mea-de@debootcamp ~]$
```

Command

`hadoop fs -count <path>`

Options

-v : Displays a header line
-h : Shows sizes in human readable format

HDFS: Download a file from HDFS to Local

```
[mea-de@debootcamp ~]$ hadoop fs -ls
Found 3 items
drwx------   - mea-de supergroup          0 2021-04-18 17:00 .Trash
drwxr-xr-x   - mea-de supergroup          0 2020-12-31 15:27 .sparkStaging
-rw-r--r--   3 mea-de supergroup         13 2021-04-18 16:47 myfile_copy.txt
[mea-de@debootcamp ~]$ hadoop fs -get myfile_copy.txt
[mea-de@debootcamp ~]$ ls
cme-cm-export.json  Documents  Music           myfile.txt  Public     Untitled.ipynb
Desktop             Downloads  myfile_copy.txt  Pictures    Templates  Videos
[mea-de@debootcamp ~]$
```

Command

hadoop fs –get <path>

Options

-f : Overwrite if already exists

BIG DATA
EXPERIENCE

More on Hadoop File System Shell

http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html

KMUTT g-able