

Utilisation de Hadoop

Manipulation de HDFS

Pr. Ghazouani mohamed

Manipulation du contenu HDFS

9 novembre 2024

Il existe différentes façons d'interagir avec un système de fichiers HDFS telles que des **API Java**, des **protocoles réseaux classiques** (HTTP, FTP, WebHDFS, HttpFS...), des **protocoles propriétaires** (Amazon S3) ou tout simplement via la **ligne de commandes**. **C'est ce dernier moyen d'interaction que nous allons détailler dans la suite de cette section.**

Toutes les commandes supportées par HDFS sont disponibles via la commande suivante :

```
$ hadoop fs
```

```
$ hdfs dfs -fs
```

L'appel d'une commande HDFS doit se faire comme une sous-commande de la commande Hadoop.

Commands Option

9 novembre 2024

- **Usage:** **-ls**
 - **hdfs dfs -ls uri or path**
 - For a file ls returns stat on the file with the following format:
permissions number_of_replicas userid groupid filesize modification_date
modification_time filename
 - For a directory it returns list of its direct children as in Unix. A directory is listed as:
permissions userid groupid modification_date modification_time dirname
 - Files within a directory are order by filename by default
- **Example:**
Nous donnons ci-dessous un exemple qui liste le contenu du répertoire racine en utilisant un chemin absolu via une **URI** /.
 - hdfs dfs -ls /
 Ci-dessous nous obtenons le même résultat en utilisant cette fois un chemin absolu via une **URL** hdfs://localhost:9000/
 - hdfs dfs -ls **hdfs://localhost:9000/**

Pr. M. Ghazouani

3

Commands Option

9 novembre 2024

-mkdir

- **Usage:**
 - **hdfs dfs -mkdir [-p] <paths>**
 - Takes path uri's as argument and creates directories
- **Options:**
 - The -p option behaviour is much like Unix mkdir -p, creating parent directories along the path
- **Example:**
 - hdfs dfs -mkdir /ghazouani/
 - hdfs dfs -mkdir -p /ghazouani/mohamed/dir1
 - hdfs dfs -mkdir -p hdfs://localhost:9000/ghazouani/hadoop/dir1

Pr. M. Ghazouani

4

Commands Option

-touchz

9 novembre 2024

- **Usage:**
 - `hdfs dfs -touchz URI [URI ...]`
 - Create a file of zero length
- **Example:**
 - `hdfs dfs -touchz /ghazouani/fichier1.txt`

Pr. M. Ghazouani

5

Commands Option

-put

9 novembre 2024

- **Usage:**
 - `hdfs dfs -put <localsrc> <dst>`
 - Copy single src, or multiple srcs from local file system to the destination file system
- **Examples:**
 - `hdfs dfs -put localfile5.txt /user/root/localfile6`
 - `hdfs dfs -put localfile1.txt localfile2.txt /user/root/hadoopfile`

-copyFromLocal

- **Usage:**
 - `hdfs dfs -copyFromLocal <localsrc> URI`
 - Similar to put command.
- **Options:**
 - The `-f` option will overwrite the destination if it already exists
- **Examples:**
 - `hdfs dfs -copyFromLocal /localfile3`

Pr. M. Ghazouani

6

Commands Option

9 novembre 2024

- **Usage:** **-get**
 - **hdfs dfs -get <src> <localdst>**
 - Copy files to the local file system
- **Examples:**
 - *hdfs dfs -get /user/root/localfile3*
 - ***hdfs dfs -get /ghazouani/fichier.txt***
 - *hdfs dfs -get hdfs://localhost/user/root/hadoopfile localfiles*
 - ***hdfs dfs -get /ghazouani/fichier.txt Bureau***
- **Usage:** **-copyToLocal**
 - **hdfs dfs -copyToLocal URI <localdst>**
 - Similar to get command.
- **Examples:**
 - ***hdfs dfs -copyToLocal /ghazouani/fichier.txt***
 - *hdfs dfs -copyToLocal /user/root/localfile3 Bureau*

Pr. M. Ghazouani

7

Commands Option

9 novembre 2024

-moveFromLocal, -moveToLocal

- **Command: moveFromLocal**
 - **Usage:**
 - **hdfs dfs -moveFromLocal <localsrc> <dst>**
 - Similar to put command, except that the source local src is deleted after it's copied
 - **Example:**
 - *hdfs dfs -moveFromLocal localfile.txt /user/root/*
 - ***hdfs dfs -moveFromLocal fichier1.txt /ghazouani***
- **Command: moveToLocal**
 - **Usage:**
 - **hdfs dfs -moveToLocal <src> <dst>**
 - Displays a "Not implemented yet" message

Pr. M. Ghazouani

8

Commands Option

9 novembre 2024

-cp

- **Usage:**
 - **hdfs dfs -cp [-f] URI or path <dest>**
 - Copy files from source (HDFS) to destination (HDFS).
- **Examples:**
 - **hdfs dfs -cp /user/root/localfile1 /user/root/localfile4**
 - **hdfs dfs -cp /user/root/localfile1 /user/root/localfile2 /user/root/dir1**
 - **hdfs dfs -cp hdfs://localhost/user/root/localfile1 hdfs://localhost/user/root/hadoopfile4**
 - **hdfs dfs -cp /ghazouani/fichier.txt /result**

Commands Option

9 novembre 2024

-mv

- **Usage:**
 - **hdfs dfs -mv URI [URI ...] <dest>**
 - Moves files from source (HDFS) to destination (HDFS)
- **Example:**
 - **hdfs dfs -mv /user/root/localfile1 /user/root/localfile5**
 - **hdfs dfs -mv /ghazouani/fichier1.txt /result**

Commands Option

9 novembre 2024

-rm

- **Usage:**
 - **hdfs dfs -rm [-f] [-r |-R] URI or path [URI ...]**
 - Delete files specified as args
- **Options:**
 - The -f option will not display a diagnostic message or modify the exit status to reflect an error if the file does not exist
 - The -R option deletes the directory and any content under it recursively.
 - The -r option is equivalent to -R
- **Example:**
 - **hdfs dfs -mkdir /repasup**
 - **hdfs dfs -touchz /repasup/fichasup.txt**
 - **hdfs dfs -touchz /repasup/fichasup1.txt**
 - **hdfs dfs -rm /repasup/fichasup.txt**
 - **hdfs dfs -rm -r /repasup/**

Pr. M. Ghazouani

11

Commands Option

9 novembre 2024

-cat

- **Usage:**
 - **hdfs dfs -cat URI or path**
 - Copies file content to stdout
- **Examples:**
 - **hdfs dfs -cat hdfs://localhost/user/root/localfile3**
 - **hdfs dfs -cat /ghazouani/testl.txt**

Pr. M. Ghazouani

12

Commands Option

9 novembre 2024

-tail

- **Usage:**
 - `hdfs dfs -tail [-f] URI or path`
 - Displays last kilobyte of the file to stdout
- **Options:**
 - The -f option will output appended data as the file grows, as in Unix
- **Example:**
 - `hdfs dfs -tail pathname`
 - `hdfs dfs -tail /test1/test1.txt`

Commands Option

9 novembre 2024

-text

- **Usage:**
 - `hdfs dfs -text <src>`
 - Takes a source file and outputs the file in text format
- **Example:**
 - `hdfs dfs -text /user/root/filename`

Commands Option

9 novembre 2024

-appendToFile

- **Usage:**
 - **hdfs dfs -appendToFile <localsrc> <dst>**
 - Append single src, or multiple srcs from local file system to the destination file system
 - Also reads input from stdin and appends to destination file system
- **Examples:**
 - **hdfs dfs -appendToFile localfile /user/hadoop/localfile2**
 - **hdfs dfs -appendToFile localfile1 localfile2 /user/hadoop/hadoopfile**
 - **hdfs dfs -appendToFile localfile hdfs://localhost/hadoop/localfile2**
 - **hdfs dfs -appendToFile fichier.txt /poeme.txt**

Pr. M. Ghazouani

15

Commands Option

9 novembre 2024

-setrep

- **Usage:**
 - **hdfs dfs -setrep [-w] <numReplicas> <path>**
 - Changes the replication factor of a file
 - If path is a directory then the command recursively changes the replication factor of all files under the directory tree rooted at path
- **Options:**
 - The -w flag requests that the command wait for the replication to complete. This can potentially take a very long time.
- **Example:**
 - **hdfs dfs -setrep -w 3 /user/root/dir1**
 - **hdfs dfs -setrep -w 2 /ghazouani/poeme.txt**

Pr. M. Ghazouani

16

Commands Option

9 novembre 2024

-getmerge

- **Usage:**

- **hdfs dfs -getmerge <src> <localdst> [addnl]**
- Takes a source directory and a destination file as input and concatenates files in src into the destination local file
- Optionally -nl can be set to enable adding a newline character at the end of each file

- **Example:**

- **hdfs dfs -getmerge /user/root/temp localfile**
- **hdfs dfs -getmerge /poeme.txt /ghazouani/fichier1.txt getmerge.txt**
- **hdfs dfs -getmerge -nl /poeme.txt /ghazouani/fichier1.txt getmerge.txt**

Pr. M. Ghazouani

17

Commands Option

9 novembre 2024

-stat

- **Usage:**

- **hdfs dfs -stat [format] <path> ...**
- Print statistics about the file/directory at <path> in the specified format.
- Format accepts filesize in blocks (%b), type (%F), group name of owner (%g), name (%n), block size (%o), replication (%r), user name of owner(%u), and modification date (%y, %Y). %y shows UTC date as "yyyy-MM-dd HH:mm:ss" and %Y shows milliseconds since January 1, 1970 UTC.
- If the format is not specified, %y is used by default

- **Example:**

- **hdfs dfs -stat "%F %u:%g %b %y %r %n" /file**

Pr. M. Ghazouani

18

Commands Option

9 novembre 2024

-du

- **Usage:**
 - **hdfs dfs -du [-s] [-h] URI or path [URI ...]**
 - Displays sizes of files and directories contained in the given directory or the length of a file in case its just a file
- **Options:**
 - The -s option will result in an aggregate summary of file lengths being displayed, rather than the individual files.
 - The -h option will format file sizes in a "human-readable" fashion (e.g 64.0m instead of 67108864)
- **Example:**
 - **hdfs dfs -du /user/root/**
 - **hdfs dfs -du -s hdfs://localhost/user/root/**

Pr. M. Ghazouani

19

Commands Option

9 novembre 2024

-help

- **Usage**
 - **hdfs dfs -help <command>**
 - Returns usage output
- **Example**
 - **hdfs dfs -help**
 - **hdfs dfs -help ls**
 - **hdfs dfs -help getmerge**

Pr. M. Ghazouani

20

9 novembre 2024



Travail pratique HDFS

Pr. M. Ghazouani

21