

Table of contents

HDFS	1
Hadoop	3

HDFS

The content of a HDFS file can be accessed by means of

- Command line commands
- A basic web interface provided by Apache Hadoop. The HDFS content can only be browsed and its files downloaded from HDFS to the local file system, while uploading functionalities are not available.
- Vendor-specific web interfaces providing a full set of functionalities (upload, download, rename, delete, ...) (e.g., the HUE web application of Cloudera).

User folder

Each user of the Hadoop cluster has a personal folder in the HDFS file system. The default folder of a user is

`/user/username`

Command line

The `hdfs` command can be executed in a Linux shell to read/write/modify/delete the content of the distributed file system. The parameters/arguments of `hdfs` command are used to specify the operation to execute.

Content of a folder

To list the content of a folder of the HDFS file system, use

`hdfs dfs -ls folder`

i Example

The command `hdfs dfs -ls /user/garza` shows the content (list of files and folders) of the `/user/garza` folder.

The command `hdfs dfs -ls .` shows the content of the current folder (i.e., the content of `/user/current_username`).

Notice that the mapping between the local linux user and the user of the cluster is based on

- A Kerberos ticket if Kerberos is active
- Otherwise the local linux user is considered

Content of a file

To show the content of a file in the HDFS file system, use

```
hdfs dfs -cat file_name
```

i Example

The command `hdfs dfs -cat /user/garza/document.txt` shows the content of the `/user/garza/document.txt` file stored in HDFS.

Copy a file from local to HDFS

To copy a file from the local file system to the HDFS file system, use

```
hdfs dfs -put local_file HDFS_path
```

i Example

The command `hdfs dfs -put /data/document.txt /user/garza/` copies the local file `/data/document.txt` in the folder `/user/garza` in HDFS.

Copy a file from HDFS to local

To copy a file from the HDFS file system to the local file system, use

```
hdfs dfs -get HDFS_path local_file
```

i Example

The command `hdfs dfs -get /user/garza/document.txt /data/` copies the HDFS file `/user/garza/document.txt` in the local file system folder `/data/`.

Delete a file

To delete a file from the HDFS file system, use

```
hdfs dfs -rm HDFS_path
```

i Example

The command `hdfs dfs -rm /user/garza/document.txt` delete from HDFS the file `/user/garza/document.txt`

Other commands

There are many other linux-like commands, for example

- `rmdir`
- `du`
- `tail`

See the [HDFS commands guide](#) for a complete list.

Hadoop

The Hadoop programs are executed (submitted to the cluster) by using the `hadoop` command. It is a command line program, characterized by a set of parameters, such as

- the name of the jar file containing all the classes of the MapReduce application we want to execute
- the name of the Driver class
- the parameters/arguments of the MapReduce application

Example

The following command executes/submits a MapReduce application

```
hadoop jar MyApplication.jar \  
it.polito.bigdata.hadoop.DriverMyApplication \  
1 inputdatafolder/ outputdatafolder/
```

- It executes/submits the application contained in `MyApplication.jar`
- The Driver Class is `it.polito.bigdata.hadoop.DriverMyApplication`
- The application has three arguments
 - Number of reducers (1)
 - Input data folder (`inputdatafolder/`)
 - Output data folder (`outputdatafolder/`)