

---

# **webhdfspy Documentation**

***Release 0.3.1***

**Fabio Souto**

May 11, 2015



<b>1</b>	<b>Installation</b>	<b>3</b>
<b>2</b>	<b>Python versions</b>	<b>5</b>
<b>3</b>	<b>Usage</b>	<b>7</b>
<b>4</b>	<b>API Documentation</b>	<b>9</b>
<b>5</b>	<b>WebHDFS documentation</b>	<b>13</b>
<b>6</b>	<b>Indices and tables</b>	<b>15</b>



A Python 2/3 wrapper library to access [Hadoop WebHDFS REST API](#)



---

## Installation

---

To install webhdfspy from PyPI:

```
$ pip install webhdfspy
```





---

## Python versions

---

webhdfspy supports Python 2.7 and 3.4



---

## Usage

---

```
>>> import webhdfspy
>>> webHDFS = webhdfspy.WebHDFSClient("localhost", 50070, "username")
>>> print(webHDFS.listdir('/'))
[]
>>> webHDFS.mkdir('/foo')
True
>>> print(webHDFS.listdir('/'))
[{'group': u'supergroup', u'permission': u'755', u'blockSize': 0, u'accessTime': 0, u'pathSuffix': u'foo'}]
>>> print(webHDFS.create('/foo/foo.txt', "just put some text here", True))
True
>>> print(webHDFS.open('/pywebhdfs_test/foo.txt'))
just put some text here
>>> webHDFS.remove('/foo')
True
>>> print(webHDFS.listdir('/'))
[]
```



---

## API Documentation

---

**class** webhdfspy.webhdfspy.**WebHDFSClient** (*host, port, username=None, logger=None*)

**\_\_init\_\_** (*host, port, username=None, logger=None*)

Create a new WebHDFS client.

When security is on, we need to specify an username :param host: hostname of the HDFS namenode  
:param port: port of the namenode :param username: used for authentication

**append** (*path, file\_data, buffersize=None*)

Append file\_data to a file

### Parameters

- **path** – path of the file
- **file\_data** – data to append to the file
- **buffersize** – the size of the buffer used to transfer the data

**chmod** (*path, permission*)

Set the permissions of a file or directory

### Parameters

- **path** – path of the file/dir
- **permission** – dir permissions in octal (0-777)

**copyfromlocal** (*local\_path, hdfs\_path, overwrite=None*)

Copy a file from the local filesystem to HDFS

### Parameters

- **local\_path** – path of the file to move
- **hdfs\_path** – hdfs path to copy the file

**create** (*path, file\_data, overwrite=None*)

Create a new file in HDFS with the content of file\_data

<https://hadoop.apache.org/docs/r1.0.4/webhdfs.html#CREATE>

### Parameters

- **path** – the file path to create the file
- **data** – the data to write to the

**environ\_home** ()

**Returns** the home directory of the user

**get\_checksum** (*path*)

Returns the checksum of a file

**Parameters** **path** – path of the file

**Returns** a FileChecksum JSON object

**listdir** (*path*='/' )

List all the contents of a directory

**Parameters** **path** – path of the directory

**Returns** a list of fileStatusProperties:

<http://hadoop.apache.org/common/docs/r1.0.0/webhdfs.html#fileStatusProperties> False on error

**mkdir** (*path*, *permission*=None)

Create a directory hierarchy, like the unix command mkdir -p

**Parameters**

- **path** – the path of the directory
- **permission** – dir permissions in octal (0-777)

**open** (*path*, *offset*=None, *length*=None, *buffer*size=None)

Open a file to read

**Parameters**

- **path** – path of the file
- **offset** – starting bit position
- **length** – number of bits to read
- **buffer**size – the size of the buffer used to transfer the data

**Returns** the file data

**remove** (*path*, *recursive*=False)

Delete a file o directory

**Parameters**

- **path** – path of the file or dir to delete
- **recursive** – set to true to delete the content in subdirectories

**rename** (*src*, *dst*)

Rename a file or directory

**Parameters**

- **src** – path of the file or dir to rename
- **dst** – path of the final file/dir

**set\_replication** (*path*, *replication\_factor*)

Set the replication factor of a file

**Parameters**

- **path** – path of the file
- **replication\_factor** – number of replications, should be > 0

**status** (*path*)

Returns the status of a file/dir

**Parameters** **path** – path of the file/dir

**Returns** a FileStatus dictionary on success, false otherwise





---

## WebHDFS documentation

---

<https://hadoop.apache.org/docs/r1.0.4/webhdfs.html>



---

## Indices and tables

---

- `genindex`
- `modindex`
- `search`



## Symbols

`__init__()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 9

## A

`append()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 9

## C

`chmod()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 9

`copyfromlocal()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 9

`create()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 9

## E

`environ_home()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 9

## G

`get_checksum()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 10

## L

`listdir()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 10

## M

`mkdir()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 10

## O

`open()` (webhdfspy.webhdfspy.WebHDFSCient method),  
10

## R

`remove()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 10

`rename()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 10

## S

`set_replication()` (webhdf-  
spy.webhdfspy.WebHDFSCient method),  
10

`status()` (webhdfspy.webhdfspy.WebHDFSCient  
method), 10

## W

`WebHDFSCient` (class in webhdfspy.webhdfspy), 9