

# upload\_donnees

March 19, 2019

## 1 Upload data

Material for the hackathon ENSAE / Red Cross / 2015. Upload data to a blob storage.

```
In [1]: from jyquickhelper import add_notebook_menu
        add_notebook_menu()
```

```
Out[1]: <IPython.core.display.HTML object>
```

### 1.1 Connexion to the blob storage

```
In [2]: import keyring
        hackathon = {}
        hackathon["blob_storage"] = keyring.get_password("CRCREDENTIALS", "blobstorage")
        hackathon["password"] = keyring.get_password("CRCREDENTIALS", "password")
```

```
In [3]: blobstorage = hackathon["blob_storage"]
        blobpassword = hackathon["password"]
```

```
In [4]: %load_ext pyensae
```

```
In [5]: cl, bs = %blob_open
        cl, bs
```

```
Out[5]: (<pyensae.remote.azure_connection.AzureClient at 0x925c2e8>,
        <azure.storage.blob.blobservice.BlobService at 0x925c320>)
```

```
In [6]: %blob_containers
```

```
Out[6]: ['croix-rouge']
```

### 1.2 Upload data

```
In [7]: %%file readme.txt
```

```
        This folder contains data for the Red Cross.
```

```
Writing readme.txt
```

```
In [8]: %blob_up readme.txt croix-rouge/readme.txt
```

```
Out[8]: 'readme.txt'
```

```
In [9]: %blob_ls croix-rouge
```

```

Out [9]:
          name                                     last_modified \
0      build/SINVOICE_M.csv Wed, 18 Nov 2015 18:56:27 GMT
1  data/ITMMASTER.schema.txt Mon, 16 Nov 2015 23:00:34 GMT
2      data/ITMMASTER.txt Mon, 09 Nov 2015 21:41:00 GMT
3  data/SINVOICE.schema.txt Mon, 16 Nov 2015 23:00:35 GMT
4      data/SINVOICE.txt Mon, 09 Nov 2015 21:42:32 GMT
5  data/SINVOICEV.schema.txt Mon, 16 Nov 2015 23:00:35 GMT
6      data/SINVOICEV.txt Mon, 09 Nov 2015 21:44:08 GMT
7  data/enseignes_france.csv Mon, 09 Nov 2015 21:40:54 GMT
8      data/stojou.csv Mon, 09 Nov 2015 21:55:23 GMT
9      readme.txt Sun, 22 Nov 2015 18:00:34 GMT

          content_type  content_length  blob_type
0  application/octet-stream      533771533  BlockBlob
1  application/octet-stream           5658  BlockBlob
2  application/octet-stream     103096479  BlockBlob
3  application/octet-stream       10252  BlockBlob
4  application/octet-stream     1362433753  BlockBlob
5  application/octet-stream        7999  BlockBlob
6  application/octet-stream     1252461865  BlockBlob
7  application/octet-stream      6303836  BlockBlob
8  application/octet-stream     8821375868  BlockBlob
9  application/octet-stream         45  BlockBlob

```

```

In [10]: import os
         files = ["ITMMASTER.schema.txt", "SINVOICE.schema.txt", "SINVOICEV.schema.txt"]
         for file in files:
             print("upload", file, os.stat(file).st_size)
             cl.upload(bs, "croix-rouge", "data/" + file, file)

```

```

upload ITMMASTER.schema.txt 5658
upload SINVOICE.schema.txt 10252
upload SINVOICEV.schema.txt 7999

```

```

In [11]: import os
         files = ["enseignes_france.csv", "ITMMASTER.txt", "SINVOICE.txt", "SINVOICEV.txt",
                 "skojou.csv"]
         for file in files:
             print("upload", file, os.stat(file).st_size)
             cl.upload(bs, "croix-rouge", "data/" + file, file)

```

```

upload enseignes_france.csv 6303836
upload ITMMASTER.txt 103096479
upload SINVOICE.txt 1362433753
upload SINVOICEV.txt 1252461865
upload stojou.csv 8821375868

```

```

In [12]: import os
         files = ["SINVOICE_.txt", "SINVOICEV_.txt"]
         for file in files:
             print("upload", file, os.stat(file).st_size)
             cl.upload(bs, "croix-rouge", "data/" + file, file)

```

```
upload SINVOICE_.txt 1362433753
upload SINVOICEV_.txt 1252461865
```

```
In [13]: %blob_ls croix-rouge/data
```

```
Out[13]:
```

	name	last_modified	\
0	data/ITMMASTER.schema.txt	Mon, 16 Nov 2015 23:00:34 GMT	
1	data/ITMMASTER.txt	Mon, 09 Nov 2015 21:41:00 GMT	
2	data/SINVOICE.schema.txt	Mon, 16 Nov 2015 23:00:35 GMT	
3	data/SINVOICE.txt	Mon, 09 Nov 2015 21:42:32 GMT	
4	data/SINVOICEV.schema.txt	Mon, 16 Nov 2015 23:00:35 GMT	
5	data/SINVOICEV.txt	Mon, 09 Nov 2015 21:44:08 GMT	
6	data/enseignes_france.csv	Mon, 09 Nov 2015 21:40:54 GMT	
7	data/stojou.csv	Mon, 09 Nov 2015 21:55:23 GMT	

  

	content_type	content_length	blob_type
0	application/octet-stream	5658	BlockBlob
1	application/octet-stream	103096479	BlockBlob
2	application/octet-stream	10252	BlockBlob
3	application/octet-stream	1362433753	BlockBlob
4	application/octet-stream	7999	BlockBlob
5	application/octet-stream	1252461865	BlockBlob
6	application/octet-stream	6303836	BlockBlob
7	application/octet-stream	8821375868	BlockBlob

### 1.3 Upload preprocessed data

```
In [14]: %blob_up SINVOICE_M.csv croix-rouge/build/SINVOICE_M.csv
```

```
Out[14]: 'build/SINVOICE_M.csv'
```

```
In [15]: %blob_up Test_CRFFOR.GACCTMPD.csv croix-rouge/build/Test_CRFFOR.GACCTMPD.csv
```

```
Out[15]: 'build/Test_CRFFOR.GACCTMPD.csv'
```

```
In [16]: %blob_ls croix-rouge/build
```

```
Out[16]:
```

	name	last_modified	\
0	build/SINVOICE_M.csv	Wed, 18 Nov 2015 18:56:27 GMT	
1	build/Test_CRFFOR.GACCTMPD.csv	Sun, 22 Nov 2015 21:53:38 GMT	

  

	content_type	content_length	blob_type
0	application/octet-stream	533771533	BlockBlob
1	application/octet-stream	822231942	BlockBlob

### 1.4 Close connection

```
In [17]: %blob_close
```

```
Out[17]: True
```

```
In [18]:
```