FEC Bulk Download – Make Local Copy

November 30, 2019

```
[1]: import datetime
     startTime = datetime.datetime.now()
     print(startTime.strftime("%Y-%m-%d %H-%M-%S"))
```

```
2019-11-29 22-48-22
```

1 FEC Bulk Downloads

1.1 AWS setup info

```
[2]: REGION  = "us-gov-west-1"
     FECBULK = "s3://cg-519a459a-0ea3-42c2-b7bc-fa1143481f74/bulk-downloads"
```

1.2 Save list of FEC Bulk Download folders and files in a file

This list of nearly 30,000 files is quite useful for future `diffs` to determine what has changed on the site and when.

```
[3]: !aws2 --region $REGION s3 ls $FECBULK/ --recursive --no-sign-request  >␣
     ↪FEC-Bulk-Download-List.txt
```

Read and display partial results

```
[4]: import pandas as pd
     import numpy as np

     columnBreaks = [(0, 19), (20, 30), (31, 120)]
     columnNames  = ["DateTime", "Bytes", "Filename"]
     columnTypes  = {"DateTime": np.object, "Bytes": np.int64, "Filename": np.object}

     data = pd.read_fwf('FEC-Bulk-Download-List.txt',
                        header   = None,
                        names    = columnNames,
                        colspecs = columnBreaks,
                        dtype    = columnTypes,
                        parse_dates = [0],
                        infer_datetime_format = True)
```

```
[5]: data.head(10)
```

```
[5]:             DateTime    Bytes                            Filename
     0 2017-08-31 13:19:38   239793              bulk-downloads/1978/can78.txt
     1 2017-08-31 13:19:38  4394124              bulk-downloads/1978/pac78.txt
     2 2017-08-31 13:20:47     1540  bulk-downloads/1980/Old Format/cm80.txt
     3 2017-08-31 13:20:47   429965  bulk-downloads/1980/Old Format/cm80.zip
     4 2017-08-31 13:20:48     1590  bulk-downloads/1980/Old Format/cn80.txt
     5 2017-08-31 13:20:48   155689  bulk-downloads/1980/Old Format/cn80.zip
     6 2017-08-31 13:20:48  1066208             bulk-downloads/1980/cansum80.dat
     7 2017-08-31 13:20:49     3165             bulk-downloads/1980/cansum80.txt
     8 2017-08-31 13:20:49   159883             bulk-downloads/1980/cansum80.zip
     9 2017-08-31 13:20:49     1540                bulk-downloads/1980/cm80.txt
```

```
[6]: data.dtypes
```

```
[6]: DateTime    datetime64[ns]
     Bytes                int64
     Filename            object
     dtype: object
```

1.3 Save local copy in folder K:/FEC

This will download over 29,600 files that will require about 62 GB of disk space.

```
[7]: !aws2 --region $REGION s3 cp --recursive $FECBULK/ --no-sign-request K:/FEC/  >␣
     ↪FEC-Bulk-Download.log
```

```
[8]: stopTime = datetime.datetime.now()
     print(stopTime.strftime("%Y-%m-%d %H-%M-%S"))
     print(stopTime - startTime)
```

```
2019-11-30 00-24-40
1:36:18.152291
```

Earl F Glynn, Olathe, KS