

Part2gsutilHIVE

March 23, 2024

```
[1]: %pip install gsutil
```

```
Collecting gsutil
  Downloading gsutil-5.27.tar.gz (3.0 MB)
      3.0/3.0 MB
13.9 MB/s eta 0:00:0000:0100:01
  Preparing metadata (setup.py) ... done
Collecting argcomplete>=1.9.4 (from gsutil)
  Downloading argcomplete-3.2.3-py3-none-any.whl.metadata (16 kB)
Collecting crcmod>=1.7 (from gsutil)
  Downloading crcmod-1.7.tar.gz (89 kB)
      89.7/89.7 kB
12.8 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting fasteners>=0.14.1 (from gsutil)
  Downloading fasteners-0.19-py3-none-any.whl.metadata (4.9 kB)
Collecting gcs-oauth2-boto-plugin>=3.0 (from gsutil)
  Downloading gcs-oauth2-boto-plugin-3.0.tar.gz (20 kB)
  Preparing metadata (setup.py) ... done
Collecting google-apitools>=0.5.32 (from gsutil)
  Downloading google_apitools-0.5.32-py3-none-any.whl.metadata (2.3 kB)
Requirement already satisfied: httpLib2==0.20.4 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from gsutil) (0.20.4)
Collecting google-reauth>=0.1.0 (from gsutil)
  Downloading google_reauth-0.1.1-py2.py3-none-any.whl.metadata (2.6 kB)
Collecting monotonic>=1.4 (from gsutil)
  Downloading monotonic-1.6-py2.py3-none-any.whl.metadata (1.5 kB)
Requirement already satisfied: pyOpenSSL>=0.13 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from gsutil) (23.2.0)
Collecting retry_decorator>=1.0.0 (from gsutil)
  Downloading retry_decorator-1.1.1.tar.gz (3.9 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: six>=1.16.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from gsutil) (1.16.0)
Requirement already satisfied: google-auth>=2.5.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from google-
auth[aiohttp]>=2.5.0->gsutil) (2.28.1)
Requirement already satisfied:
```

```

pyparsing!=3.0.0,!3.0.1,!3.0.2,!3.0.3,<4,>=2.4.2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
httplib2==0.20.4->gsutil) (3.1.1)
Collecting rsa==4.7.2 (from gcs-oauth2-boto-plugin>=3.0->gsutil)
  Downloading rsa-4.7.2-py3-none-any.whl.metadata (3.6 kB)
Collecting boto>=2.29.1 (from gcs-oauth2-boto-plugin>=3.0->gsutil)
  Downloading boto-2.49.0-py2.py3-none-any.whl.metadata (7.3 kB)
Requirement already satisfied: oauth2client>=2.2.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from gcs-oauth2-boto-
plugin>=3.0->gsutil) (4.1.3)
Requirement already satisfied: pyasn1>=0.1.3 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from rsa==4.7.2->gcs-
oauth2-boto-plugin>=3.0->gsutil) (0.5.1)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from google-
auth>=2.5.0->google-auth[aiohttp]>=2.5.0->gsutil) (5.3.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from google-
auth>=2.5.0->google-auth[aiohttp]>=2.5.0->gsutil) (0.3.0)
Requirement already satisfied: aiohttp<4.0.0.dev0,>=3.6.2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from google-
auth[aiohttp]>=2.5.0->gsutil) (3.9.3)
Requirement already satisfied: requests<3.0.0.dev0,>=2.20.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from google-
auth[aiohttp]>=2.5.0->gsutil) (2.28.2)
Requirement already satisfied: pyu2f in
/opt/conda/miniconda3/lib/python3.10/site-packages (from google-
reauth>=0.1.0->gsutil) (0.1.5)
Requirement already satisfied: cryptography!=40.0.0,!40.0.1,<42,>=38.0.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
pyOpenSSL>=0.13->gsutil) (39.0.0)
Requirement already satisfied: aiosignal>=1.1.2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
aiohttp<4.0.0.dev0,>=3.6.2->google-auth[aiohttp]>=2.5.0->gsutil) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
aiohttp<4.0.0.dev0,>=3.6.2->google-auth[aiohttp]>=2.5.0->gsutil) (23.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
aiohttp<4.0.0.dev0,>=3.6.2->google-auth[aiohttp]>=2.5.0->gsutil) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
aiohttp<4.0.0.dev0,>=3.6.2->google-auth[aiohttp]>=2.5.0->gsutil) (6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
aiohttp<4.0.0.dev0,>=3.6.2->google-auth[aiohttp]>=2.5.0->gsutil) (1.9.4)
Requirement already satisfied: async-timeout<5.0,>=4.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from

```

```

aiohttp<4.0.0.dev0,>=3.6.2->google-auth[aiohttp]>=2.5.0->gsutil) (4.0.3)
Requirement already satisfied: cffi>=1.12 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
cryptography!=40.0.0,!40.0.1,<42,>=38.0.0->pyOpenSSL>=0.13->gsutil) (1.16.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
requests<3.0.0.dev0,>=2.20.0->google-auth[aiohttp]>=2.5.0->gsutil) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
requests<3.0.0.dev0,>=2.20.0->google-auth[aiohttp]>=2.5.0->gsutil) (3.6)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
requests<3.0.0.dev0,>=2.20.0->google-auth[aiohttp]>=2.5.0->gsutil) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
requests<3.0.0.dev0,>=2.20.0->google-auth[aiohttp]>=2.5.0->gsutil) (2024.2.2)
Requirement already satisfied: pycparser in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
cffi>=1.12->cryptography!=40.0.0,!40.0.1,<42,>=38.0.0->pyOpenSSL>=0.13->gsutil)
(2.21)
Downloading argcomplete-3.2.3-py3-none-any.whl (42 kB)
42.6/42.6 kB
5.7 MB/s eta 0:00:00
Downloading fasteners-0.19-py3-none-any.whl (18 kB)
Downloading rsa-4.7.2-py3-none-any.whl (34 kB)
Downloading google_apitools-0.5.32-py3-none-any.whl (135 kB)
135.7/135.7 kB
20.0 MB/s eta 0:00:00
Downloading google_reauth-0.1.1-py2.py3-none-any.whl (17 kB)
Downloading monotonic-1.6-py2.py3-none-any.whl (8.2 kB)
Downloading boto-2.49.0-py2.py3-none-any.whl (1.4 MB)
1.4/1.4 MB
77.6 MB/s eta 0:00:00
Building wheels for collected packages: gsutil, crcmod, gcs-oauth2-boto-
plugin, retry_decorator
  Building wheel for gsutil (setup.py) ... done
  Created wheel for gsutil: filename=gsutil-5.27-py3-none-any.whl
size=3785291
sha256=21c20b1280f6f383e196cba8500eaed94cbef6c1ef722e7f14b8dfa2db8bdadc
  Stored in directory: /root/.cache/pip/wheels/03/0c/cd/24471307835689188d498272
32a1430bdd36587f913d8f707f
  Building wheel for crcmod (setup.py) ... done
  Created wheel for crcmod:
filename=crcmod-1.7-cp310-cp310-linux_x86_64.whl size=23544
sha256=731553a2bbff685c46010755b17aaad3d0a17fc72f3eded7ffcd482edcceb8c
  Stored in directory: /root/.cache/pip/wheels/85/4c/07/72215c529bd59d67e3dac297
11d7aba1b692f543c808ba9e86
  Building wheel for gcs-oauth2-boto-plugin (setup.py) ... done

```

```

Created wheel for gcs-oauth2-boto-plugin:
filename=gcs_oauth2_boto_plugin-3.0-py3-none-any.whl size=23213
sha256=4c2dbafdadfaa44dd330e8fd3857990d0174cf278a7fbd0ff725c78a30c48e03
Stored in directory: /root/.cache/pip/wheels/77/2a/cd/c7b51b51563ebefe01488e3e
cb19df401ea6725771937d3bc6
Building wheel for retry_decorator (setup.py) ... done
Created wheel for retry_decorator:
filename=retry_decorator-1.1.1-py2.py3-none-any.whl size=3639
sha256=5cdb64822fe7590b498240692ce6222c7d6063b32aaf83f238019df7c89c1b8f
Stored in directory: /root/.cache/pip/wheels/dd/ac/77/8c54eac0d373d9eacfbe4259
9710c9bf91b4c5985297f6922a
Successfully built gsutil crcmod gcs-oauth2-boto-plugin retry_decorator
Installing collected packages: retry_decorator, monotonic, crcmod, boto, rsa,
fasteners, argcomplete, google-reauth, google-apitools, gcs-oauth2-boto-plugin,
gsutil
Attempting uninstall: rsa
Found existing installation: rsa 4.9
Uninstalling rsa-4.9:
Successfully uninstalled rsa-4.9
Successfully installed argcomplete-3.2.3 boto-2.49.0 crcmod-1.7 fasteners-0.19
gcs-oauth2-boto-plugin-3.0 google-apitools-0.5.32 google-reauth-0.1.1
gsutil-5.27 monotonic-1.6 retry_decorator-1.1.1 rsa-4.7.2
WARNING: Running pip as the 'root' user can result in broken permissions
and conflicting behaviour with the system package manager. It is recommended to
use a virtual environment instead: https://pip.pypa.io/warnings/venv
Note: you may need to restart the kernel to use updated packages.

```

```

[2]: from google.colab import auth
auth.authenticate_user()

import gspread
from google.auth import default
creds, _ = default()
#Deben iniciar sesion con la cuenta de la up para que les deje ver mi proyecto
gc = gspread.authorize(creds)

```

```

-----
ModuleNotFoundError                                Traceback (most recent call last)
/tmp/ipykernel_3793/122713332.py in <cell line: 1>()
----> 1 from google.colab import auth
      2 auth.authenticate_user()
      3
      4 import gspread
      5 from google.auth import default

```

```
ModuleNotFoundError: No module named 'google.colab'
```

```
[3]: !gsutil config
```

This command will create a boto config file at /root/.boto containing your credentials, based on your responses to the following questions.
CommandException: The user account authentication flow no longer works as of February 1, 2023. Tokens generated before this date will continue to work. To authenticate with your user account, install gsutil via Cloud SDK and run "gcloud auth login"

```
[4]: #Muestra cuanto pesa todos los CSV
!gsutil du -hs "gs://bucketteam3/*.csv"
```

CommandException: One or more URLs matched no objects.
0 B gs://bucketteam3/*.csv

```
[5]: !gsutil ls gs://bucketteam3
```

```
[6]: %pip install hive
```

```
Collecting hive
  Downloading hive-0.1.3dev.tar.gz (39 kB)
  Preparing metadata (setup.py) ... done
Collecting Pylons>=1.0 (from hive)
  Downloading Pylons-1.0.3-py2.py3-none-any.whl.metadata (3.7 kB)
Requirement already satisfied: Jinja2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from hive) (3.1.3)
Collecting Routes>=1.12.3 (from Pylons>=1.0->hive)
  Downloading Routes-2.5.1-py2.py3-none-any.whl.metadata (25 kB)
Collecting WebHelpers>=0.6.4 (from Pylons>=1.0->hive)
  Downloading WebHelpers-1.3.tar.gz (729 kB)
    730.0/730.0
kB 5.9 MB/s eta 0:00:00a 0:00:01
  Preparing metadata (setup.py) ... done
Collecting Beaker>=1.5.4 (from Pylons>=1.0->hive)
  Downloading Beaker-1.12.1.tar.gz (67 kB)
    67.3/67.3 kB
12.2 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Collecting Paste>=1.7.5.1 (from Pylons>=1.0->hive)
  Using cached Paste-3.8.0-py3-none-any.whl.metadata (4.9 kB)
Collecting PasteDeploy>=1.5.0 (from Pylons>=1.0->hive)
  Using cached PasteDeploy-3.1.0-py3-none-any.whl.metadata (2.7 kB)
Collecting PasteScript>=1.7.4.2 (from Pylons>=1.0->hive)
  Using cached PasteScript-3.5.1-py2.py3-none-any.whl.metadata (2.4 kB)
Collecting FormEncode>=1.2.4 (from Pylons>=1.0->hive)
```

```

    Downloading FormEncode-2.1.0-py3-none-any.whl.metadata (2.6 kB)
Requirement already satisfied: simplejson>=2.2.1 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from Pylons>=1.0->hive)
(3.19.2)
Requirement already satisfied: decorator>=3.3.2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from Pylons>=1.0->hive)
(5.1.1)
Collecting nose>=1.1.2 (from Pylons>=1.0->hive)
    Downloading nose-1.3.7-py3-none-any.whl.metadata (1.7 kB)
Requirement already satisfied: Mako>=0.5.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from Pylons>=1.0->hive)
(1.3.2)
Collecting WebError>=0.10.3 (from Pylons>=1.0->hive)
    Downloading WebError-0.13.1.tar.gz (85 kB)
                                85.9/85.9 kB
13.9 MB/s eta 0:00:00
    Preparing metadata (setup.py) ... done
Collecting WebTest>=1.3.1 (from Pylons>=1.0->hive)
    Downloading WebTest-3.0.0-py3-none-any.whl.metadata (1.8 kB)
Collecting Tempita>=0.5.1 (from Pylons>=1.0->hive)
    Downloading Tempita-0.5.2-py3-none-any.whl.metadata (1.0 kB)
Requirement already satisfied: MarkupSafe>=0.15 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from Pylons>=1.0->hive)
(2.1.1)
Collecting WebOb>=1.1.1 (from Pylons>=1.0->hive)
    Downloading WebOb-1.8.7-py2.py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: setuptools in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
Paste>=1.7.5.1->Pylons>=1.0->hive) (69.1.0)
Requirement already satisfied: six>=1.4.0 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
Paste>=1.7.5.1->Pylons>=1.0->hive) (1.16.0)
Collecting repoze.lru>=0.3 (from Routes>=1.12.3->Pylons>=1.0->hive)
    Downloading repoze.lru-0.7-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: Pygments in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
WebError>=0.10.3->Pylons>=1.0->hive) (2.17.2)
Collecting waitress>=0.8.5 (from WebTest>=1.3.1->Pylons>=1.0->hive)
    Downloading waitress-3.0.0-py3-none-any.whl.metadata (4.2 kB)
Requirement already satisfied: beautifulsoup4 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
WebTest>=1.3.1->Pylons>=1.0->hive) (4.12.3)
Requirement already satisfied: soupsieve>1.2 in
/opt/conda/miniconda3/lib/python3.10/site-packages (from
beautifulsoup4->WebTest>=1.3.1->Pylons>=1.0->hive) (2.5)
Downloading Pylons-1.0.3-py2.py3-none-any.whl (147 kB)
                                147.7/147.7 kB
15.7 MB/s eta 0:00:00

```

```

Downloading FormEncode-2.1.0-py3-none-any.whl (179 kB)
179.4/179.4 kB
26.4 MB/s eta 0:00:00
Downloading nose-1.3.7-py3-none-any.whl (154 kB)
154.7/154.7 kB
25.8 MB/s eta 0:00:00
Using cached Paste-3.8.0-py3-none-any.whl (278 kB)
Using cached PasteDeploy-3.1.0-py3-none-any.whl (16 kB)
Using cached PasteScript-3.5.1-py2.py3-none-any.whl (73 kB)
Downloading Routes-2.5.1-py2.py3-none-any.whl (40 kB)
40.1/40.1 kB
6.2 MB/s eta 0:00:00
Downloading Tempita-0.5.2-py3-none-any.whl (12 kB)
Downloading WebOb-1.8.7-py2.py3-none-any.whl (114 kB)
115.0/115.0 kB
16.0 MB/s eta 0:00:00
Downloading WebTest-3.0.0-py3-none-any.whl (31 kB)
Downloading repoze.lru-0.7-py3-none-any.whl (10 kB)
Downloading waitress-3.0.0-py3-none-any.whl (56 kB)
56.7/56.7 kB
9.2 MB/s eta 0:00:00
Building wheels for collected packages: hive, Beaker, WebError, WebHelpers
  Building wheel for hive (setup.py) ... done
  Created wheel for hive: filename=hive-0.1.3.dev0-py3-none-any.whl
size=51201
sha256=ceb362768eb35305b2689d21c43b578af30a5c00f57d48a0f9f2621b1891942a
  Stored in directory: /root/.cache/pip/wheels/de/82/29/b79d6d56a73a355372dc7252
f581b61fe7a79c4196064e5a5e
  Building wheel for Beaker (setup.py) ... done
  Created wheel for Beaker: filename=Beaker-1.12.1-py3-none-any.whl
size=51625
sha256=9dd9916a085a69fbc17381b31156df1d7bd1416e8255d537f3f893662d37b73c
  Stored in directory: /root/.cache/pip/wheels/76/14/05/c8d0d173ee1a71b0aa4589bb
6a69dec189b9e0ffc3facb6e7d
  Building wheel for WebError (setup.py) ... done
  Created wheel for WebError: filename=WebError-0.13.1-py3-none-any.whl
size=89411
sha256=6b0c3005a687e4c601fd1f07ba3518d36feb3ea6ede1850d819f1e762a02ed80
  Stored in directory: /root/.cache/pip/wheels/39/eb/54/21f8f808adf42ecd9c314cc7
91f59ddca11d90134721a2c33f
  Building wheel for WebHelpers (setup.py) ... done
  Created wheel for WebHelpers: filename=WebHelpers-1.3-py3-none-any.whl
size=149026
sha256=db7601df00fc1fed2d0ebba3ec7b12e02b2a5b899e409d15b693db4b3f5bf0a9
  Stored in directory: /root/.cache/pip/wheels/4a/44/46/932d613b343bbd3f7d769d29
447b8a00dc251c69b6b91625d9
Successfully built hive Beaker WebError WebHelpers
Installing collected packages: Tempita, repoze.lru, nose, Beaker, WebOb,

```

```
WebHelpers, waitress, Routes, PasteDeploy, Paste, FormEncode, WebTest, WebError,
PasteScript, Pylons, hive
Successfully installed Beaker-1.12.1 FormEncode-2.1.0 Paste-3.8.0
PasteDeploy-3.1.0 PasteScript-3.5.1 Pylons-1.0.3 Routes-2.5.1 Tempita-0.5.2
WebError-0.13.1 WebHelpers-1.3 WebOb-1.8.7 WebTest-3.0.0 hive-0.1.3.dev0
nose-1.3.7 repoze.lru-0.7 waitress-3.0.0
```

WARNING: Running pip as the 'root' user can result in broken permissions
and conflicting behaviour with the system package manager. It is recommended to
use a virtual environment instead: <https://pip.pypa.io/warnings/venv>

Note: you may need to restart the kernel to use updated packages.

```
[7]: !ls
```

```
Movies  copyright  hadoop      lib    libx32    mnt    root    snap    tmp
bin     dev        home        lib32  lost+found  opt    run     srv     usr
boot    etc        kaggle.json lib64   media     proc   sbin    sys     var
```

```
[8]: %cd Movies
```

```
/Movies
```

```
[9]: !ls
```

```
Movies      countries.csv  genres.csv    letterboxd.zip  posters
actors.csv  crew.csv      languages.csv  movies.csv
```

```
[10]: %cd Movies
```

```
/Movies/Movies
```

```
[1]: #Muestra cuanto pesa todos los CSV
!gsutil du -hs "gs://bucketteam3/*.csv"
```

```
CommandException: One or more URLs matched no objects.
0 B              gs://bucketteam3/*.csv
```

```
[ ]:
```