

# Logs archiving to GCP

- **File location**

The file is named *archive.py*, in the home directory of the Dataiku user on design prod VM.

```
[dataiku@ce-bda-dss-design-ew1-prod ~]$ ls
archive.py
```

- **Scheduling**

To set up the cron job, we run :

```
crontab -e
```

The job is set up to run every day at 9AM French time / 3PM Singapour Time

```
0 7 * * * python3 ~/archive.py
```

- **Logs location**

Logs are located in the */dataiku/dss\_data/run* directory.

```
[dataiku@ce-bda-dss-design-ew1-prod ~]$ cd /dataiku/dss_data/run
[dataiku@ce-bda-dss-design-ew1-prod run]$ ls
archivelogs.zip  backend.log.2  backend.log.7  frontend.log.0.lck  install-impersonation.log
audit            backend.log.3  backend.log.8  frontend.log.1      install.log
backend.log      backend.log.4  backend.log.9  frontend.log.2      ipython.log
backend.log.1    backend.log.5  dbservers.json  frontend.log.3      ipython.log.1
backend.log.10   backend.log.6  frontend.log.0  frontend.log.4      ipython.log.2
```

- **Execution**

The program lists every file in the run directory, then zip all the backend.log.\* into a archivelogs.zip file. We use google cloud storage python package. We can install the package with :

```
pip install google-cloud-storage
```

```
from google.cloud import storage

storage_client = storage.Client()
bucket = storage_client.get_bucket('cs-bda-dss-ew-prod-design')
blob = bucket.blob('dataiku/Logs/'+ date + 'archivelogs.zip')
blob.upload_from_filename('archivelogs.zip')
```

- **Cloud Storage**

The zip files are located in the *cs-bda-dss-ew-prod-design* bucket, in the *dataiku/logs* folder. Files are structured with this subfolder structure : Year => Month => Daily zip