- `gcloud components update`
- `gcloud init` -- if needed
  - pick project `hedera-regression`
  - pick zone 'us-west1-c` -- arbitrarily (?)
- `gcloud auth application-default login` -- and sign in w/ account

Python setup

- have `venv` available, install `3.12.0` (or something else recent)
- ▶ `pip install -r requirements.txt`
  - ▶ `pip install google-cloud-storage`
- `open /Applications/Python\ 3.12/Install\ Certificates.command`

Get hostnames from `swirlds/infrastructure` repo

- see `mainnet/ansible/host_vars_mainnet/node*.yml`
- `grep source_hostname * | sed 's/[.]yml:source_hostname://g'`
- use file `mainnet_hostnames-2024-02-04.txt`

To run:

1. First create the list of blobs (record/event files) you want to download from gcp:

```
python3 main.py get_blob_names -root <dir-for-files> \
                      -b <bloblist-filename>         \
                      -s <start-time> -e <end-time>  \
                      -node <node#>
```

   where the start and end of the interval are specified like `2024-02-01T00:00:00` , and the node number is the node you want to pull files for.
   This will tell you how many files it found in that interval.
   But it may happen the node you picked was down for some time during that interval. So run the script again using the command `reget_blob_names` (instead of `get_blob_names`) and specify a different node number. It will *merge* additional files found into the bloblist you already have. Repeat until it finds no new files.
2. Download the blobs with the command

```
python3 main.py download_blobs -root <dir-for-files> \
                    -b <bloblist-filename>
```

It will fetch files in batches - and give you a progress report on how many batches it is doing.
Files can fail to download. Keep repeating this command until you see, by the metrics reported, that all the files are downloaded.
You can "tune" the performance by changing the batch size with the `-batch nnn` argument, and by changing the level of concurrency with the `-concurrency nnn` argument.