ATA Data Google Cloud Storage

Jon Richards

SETI Institute

June 25, 2019

**Introduction**

This document explains details of the Google Cloud Storage used for ATA data. This includes the SQL database instance.

If you have any access issues, need access to the data buckets, etc., contact Jon Richards ([jrichards@seti.or](mailto:jrichards@seti.or)g) and he will give you any access you want. You may wish to contact him first to make sure you should have access, before you get confused in the installation steps below.

For data storage the cloud is accessed using the gsutil utility provided by Google. There is a list of available commands:

* Go to: <https://cloud.google.com/storage/docs/gsutil_install>
* In the menu on the left igrate to gsutil tool -> gsutil Commands

**Data Storage Bucket**

The ATA data storage is located under the SETI-ATA Cloud project. The project name is “probable-cove-796”. It is in a bucket called ata\_data. You access the files with the path “gs://ata\_test\_data”, once you have successfully completed “gcloud config” described below.

To view all the data directories in the probable-cove-796 project go to:

* <https://console.cloud.google.com/storage/browser/ata_test_data?project=probable-cove-796>

The ata\_test\_data directory is under there; you can click to view all the files.

Currently, the following ATA computers have access to this data configured and working, under the jrseti@gmail account:

* NSG-head – this is the head node where the on/off data is stored in the /data directory.
* sonata1 – this is the computer that contains the database and various utilities. The database on this computer is eventually going to be migrated to the cloud.
* chan1x – this computer is used to read data from the beamformer, usually beam 1, x pol, and store it on a large disk mounted as “/data”. Relating to the cloud, the usual use case for this computer is to read and store beamformer data as it streams in (using hashpipe), convert the data to filterbank format, and push this data in the cloud. It is possible that some raw beamformer data may be moved to the cloud, but it is so large and oppressive that hopefully storing raw beamformer data in the cloud will be a rare case.
* chan2x – Same as “chan1x”, but for another beam or pol

**SQL Database Instance**

The SQL database instance stores all the feed sensor data. Grafana uses this instance as a data source. In theory anyone should be able to run his or her own Grafana server and use this SQL database instance as a data source.

The SQL database instance is at: 34.83.6.3, the database name is “ants”

To access read only, should work for anyone from any computer:

* mysql --host=34.83.6.3--user=ata-sensors ants

Read/Write access is available if you connect from any computer originating from the ip address of the ATA, 70.100.31.22. Also, the user must be “ata-sensors-rw”.

* mysql --host=34.83.6.3 --user=ata-sensors-rw ants

**Utilities Installation**

On your computer you need to do the following to access the data. Note that you will be using your own

Install gsutil, follow the instructions at:

* <https://cloud.google.com/storage/docs/gsutil_install>

Once gsutil is installed, init your access to the gloud:

* gcloud auth login --no-launch-browser (then follow directions)

Tell your computer which data bucket to use:

* gcloud config set project probable-cove-796

You should now be able to access the data. Test that you can read a file from the gs://ata\_data directory. There is a test file just for this purpose:

* gsutil cp gs://ata\_test\_data/access\_test\_file.txt
* Verify that access\_test\_file.txt transferred to your computer and you can view the simple text contents of the file.

Test your ability to write to the data directory, if you have write permissions. It may be the case you do not have write permissions if your use case is just to read data files. It is assumed that most users will not need or want write access, but if that is the case for you, test:

* On your computer create a dummy file: “touch my\_test.dat”
* gsutil cp my\_test.dat gs://ata\_test\_data
* Go to <https://console.cloud.google.com/storage/browser/ata_test_data?project=probable-cove-796>
* View the list of files. You should see “my\_test.dat” in the list. You may need to perform a refresh in your browser to see the data.
* Remove the file: gsutil r mgs://ata\_test\_data/my\_test.dat. Refresh the browser and confirm the file is removed.

Examples:

Rsync 6GB of files from /data/20190625 on NSG-head to gs://ata\_test\_data/rfi/:

* Do this in a screen!
* cd /data
* gsutil rsync -r 20190625 gs://ata\_test\_data/rfi/20190625