

Eureka User Guide

UPDATED 09.2024

CHANGES TO NOTE IN EUREKA V4

1. When logging in with your hdcuser credentials, ensure that your account name, "firstname.lastname", is all lowercase.
2. The new login page (the teal xrdp screen) does not allow copy-pasting, so password managers will not work.
3. The Limited Internet App is now located under Applications -> Internet
4. As of V4, an unlimited number of users can login to a machine at once, as opposed to the limit of 4 imposed by NoMachine in Eureka V3.
5. After 5 minutes of inactivity, your Eureka screen will sleep. This does not log you out, but you will be prompted for a password upon returning to your screen.
6. A separate R Studio link is no longer available in V4. Users will be required to use the R Studio desktop application via the remote desktop.

Eureka User Guide Contents

[Connecting to Eureka App VM](#)

[Disconnecting from Eureka App VM](#)

[Accessing Compass Data Marts in Eureka](#)

[Moving Files In & Out of Eureka App VM](#)

[Preinstalled Software](#)

[Limited Internet Access from Eureka App VM](#)

[Idle Shutdown of Eureka App VM](#)

[Using Eureka HPC](#)

[Storage Options](#)

[Google Cloud Source Repository](#)

Connecting to Eureka App VM

Each user has been provided a unique URL to connect to your Eureka App VM. Here is how to connect to Eureka App VM:

7. Using the current version of Chrome or Firefox web browser, open a Private or Incognito window, and go to your Eureka App VM URL.
8. Use your Compass User Account to authenticate with Google (including 2-factor authentication).
9. If you get a "Server Error" message, it is because Google is booting the VM (it may take up to 3 minutes for the VM to boot). Refresh the page, and if the machine is booted the error message will go away. You may need to close and re-open the tab if refreshing the page does not work.
10. A turquoise page should appear. Use your Compass User Account name or email (in all lower case) and password to log into Guacamole (i.e. `firstname.lastname`, firstname.lastname@hdcuser.org).
11. You're in your App VM once you see the Eureka background image.

If you abandon a session (this can happen by closing the web browser instead of disconnecting from Eureka) your session will remain active for 30 minutes before automatically disconnecting. To connect to your abandoned session, simply log back in to the custom App VM URL.

If you have forgotten your password, please [contact Compass](#) to reset your account credentials. Any links will be sent to the e-mail that you used to request your Compass User Account.

Disconnecting from Eureka App VM

When you are done working in Eureka, it is a good security practice to end your login session. This can be done by manually logging out of Guacamole to end your session, or by shutting down the VM if no other users are logged in.

Manually logging out of Guacamole (without shutting down the VM):

12. In the upper right hand corner of your VM desktop, click on your username and a drop down window will appear.
13. Select 'Log Out' and a window will appear in the middle of the screen.
14. Select 'Log Out' again, and you will be logged out of your Guacamole session.

Manually shutting down your VM:

15. Open a terminal window.
16. Type "sudo poweroff", then press Enter.
17. If prompted, enter your password.

Accessing Compass Data Marts in Eureka

If you have been authorized access to a Health Data Compass data mart in BigQuery, you can safely view it from your Eureka App VM. You can also download it to your Eureka App VM for further analysis on your Eureka App VM.

Via the Web User Interface

You can interact with BigQuery through the BigQuery user interface via your Eureka App VM. Simply open a web browser to <https://console.cloud.google.com/bigquery>. The user interface should be fairly self-explanatory. Some helpful documentation here: <https://cloud.google.com/bigquery/docs/bigquery-web-ui>

Important: The BigQuery Web UI at the above links will allow you to access Compass data marts in BigQuery from your local workstation. This access is not approved and will fire an alert with our security monitoring team. Only access Compass data marts in BigQuery from your Eureka App VM.

Via the Command Line

You can access BigQuery datasets using the “bq” command line utility. This is a powerful utility, and full documentation can be found here: <https://cloud.google.com/bigquery/docs/bq-command-line-tool>.

Moving Files In & Out of Eureka App VM

Eureka is designed first and foremost to protect sensitive data files. One important aspect of this is that you cannot access the Internet directly from your app server in order to upload or download files via the usual mechanisms such as FTP, email, or web sites. Instead, you will use a specially configured location on Google Cloud Storage, called your Eureka Staging Bucket. Your Eureka Staging Bucket can be used to transfer files between your local workstation and your Eureka App VM.

This is a two-step process:

18. Upload files from your app server or local workstation to your staging bucket
19. Download the files from your staging bucket to your app server or local workstation

There are three options for doing this: [Google Cloud Console](#), [gsutil](#), and [GCSFuse](#).

Important:

- 20. You have the ability to use your Eureka Staging Bucket to download data to your local workstation.**
- 21. Just because you can do this does not mean that you should.**
- 22. Sensitive data such as PHI may only be downloaded to workstations or servers that comply with your institutional HIPAA policies.**

Please [contact us](#) if you have any questions.

Preinstalled Software

Each Eureka App-VM comes with the following software preinstalled:

Google Cloud SDK - utilities to access & manage Google Cloud Platform resources

LibreOffice - office productivity suite

PyCharm - Python IDE

Python - high-level, general purpose programming language

R - statistical analysis toolkit

R Studio Desktop - desktop-based IDE for R

Visual Studio Code - code editor

Limited Internet Access from Eureka App VM

Eureka App VM V4 has the ability to connect to the following URLs from within Eureka via the Eureka Limited Internet App. Google Chrome is the only optimized browser to use in Eureka App VM with the limited internet access functionality.

The first time you use the Eureka Limited Internet App you will need to run the following command from your Eureka App VM and follow the prompts. It may provide you with a long URL which you should paste into a web browser in Eureka App VM and authenticate using your Eureka credentials.

```
gcloud auth login
```

There are two ways to interact with the Eureka Limited Internet App.

Option #1: Locate the Eureka Limited Internet App in the applications directory and select the website to which you wish to connect.

Option #2: Open a terminal window and use one of the eureka-internet commands listed below.

After you select the site from the Eureka Limited Internet App, connection to the chosen site will be allowed after a short delay. This is usually around 5 seconds, but can take up to 15. Access to the site is limited to 30 minutes, if you need the connection open for longer, re-select the site from the Eureka Limited Internet App and that will add another 30 minutes of connection. Below are the options for internet connectivity:

Limited Internet App Button: **CRAN & Bioconductor**

Console Command: **eureka-internet-CRAN-Bioconductor**

Sites:

<https://cloud.r-project.org>

<https://www.bioconductor.org>

Limited Internet App Button: **Github.com**

Console Command: **eureka-internet-GitHub.com**

Sites:

<https://github.com>

<https://raw.githubusercontent.com>

Limited Internet App Button: **PyPi.org & Python.org**

Console Command: **eureka-internet-Python.org**

Sites:

<https://pypi.org>

<https://www.python.org>

Limited Internet App Button: **REDCap**

Console Command: **eureka-internet-RedCap**

Sites:

<https://redcap.ucdenver.edu>

NOTE: Some R Packages require access to GitHub at the same time to CRAN so make sure you select both sites from the Eureka Limited Internet App to ensure complete installation of those packages.

When you are done with your session and no longer need to use the Eureka Limited Internet App, you can logout of GCloud by running the following command from your Eureka App VM:

```
gcloud auth revoke
```

Internet Security & Eureka App VM

Security is a group effort between you and Compass. We cannot do it without you. Please be sure to follow all rules in the Eureka User Agreement.

Some common problems with software downloaded from the internet include:

- 23. Outdated software with known security vulnerabilities
- 24. Software that includes poor programming or security practices
- 25. Malicious software such as viruses

You must ensure that you have carefully reviewed software from any source for these problems, but be particularly careful with container hubs (such as Docker Hub) and software from GitHub that is not widely used. Due to the difficulty of determining the trustworthiness of software on container hubs, we discourage their use. You are responsible for vetting software you upload to Eureka.

You must not store confidential information on sites outside Eureka, unless you have received specific permission. You must never store confidential information on GitHub.

Idle Shutdown of Eureka App VM

Each Eureka instance is pre-configured to shut down the VM after 30 minutes of undetected usage of the VM. If you want to temporarily disable the idle shut down, run the following command from your VM terminal window:

```
sudo systemctl stop idleshutdown
```

If you disable the idle shut down, **you are responsible** for manually shutting down the VM if you are not longer using it.

The pre-configured idle shutdown will be re-enabled anytime the VM is rebooted, until then you will need to manually shut down the VM.

Using Eureka HPC

Jobs are submitted to the SLURM workload manager. Primarily, you will use the `sbatch` command to submit jobs, and the `squeue` command to monitor jobs. You can submit any valid shell script as a job using `sbatch`. Once it's submitted, Eureka HPC will create a temporary compute node just for this job. As soon as this node has no more work to do, it will be deleted. This is the core cost saving feature of Eureka HPC

If you submit a job with `sbatch` and provide no options to SLURM, your job will be submitted with the following defaults:

26. It will run on a 2 core node with 8 GB of RAM
27. It will have a maximum run time of 23 hours
28. It will be run on a Google preemptible VM for maximum cost savings

To exceed 23 hours of runtime, you must submit to a non-default Slurm partition. Each partition corresponds to a particular Google Cloud machine type. You can see what partitions exist by typing "`sinfo`" and get information about their hardware resources by typing "`man eureka-queues`". If your job can run in a short amount of time (under 23 hours) consider submitting it to a preemptible partition so that it will run at lower cost. Otherwise, submit to a partition with the `_nonpre` suffix to get a standard Google Cloud instance without the spot/preemptible cost-saving feature.

Running interactive SLURM jobs

You should not run any interactive processes that require more than minimal CPU on the login node. Things like text editors are fine, but commands like `sort` or `R` should be run as a SLURM interactive job so that they get run on powerful hardware. You can start an interactive job by typing `interact`. There may be a delay of approx. 1 minute after you type this, and then you will receive a shell prompt on a compute node.

SLURM Best Practices on Eureka

The ideal batch job is longer than a few minutes, but shorter than a day. Jobs in this range of lengths will be scheduled more efficiently by SLURM, and so you will get more work through the system and, on average experience less queue wait time if you keep your jobs within these limits.

Batch jobs should be able to be killed and restarted without losing too much progress. You can accomplish this by writing your code to checkpoint, or simply by breaking up a long running job into multiple shorter jobs.

Do not hard code the UNIX path of your home directory into your SLURM scripts. If you need to reference a file in your scripts, use the shell's \$HOME variable.

29. Try it: Type echo \$HOME
30. If you have a file called myfile in the top level of your home, you can reference it as \$HOME/myfile

Storage options on Eureka

To minimize dollars spent on computing and storage, you must be aware of the different types of Google storage available in Eureka HPC:

31. Google Cloud Storage
32. /home
33. /tmp on compute nodes
34. Optionally, shared storage mounted at /gpfs

Google Cloud Storage

This is Google's most cost effective storage method for long term storage. Ideally, you should store both input data here, and also results. You must move data in and out of Google Cloud Storage using the gsutil tool or the Google Cloud web console. Instructions on using Google Cloud storage can be found above.

/home

Your home directory exists only for holding files that are required for your Linux account to function. It is small and slow, but it's OK to temporarily allow SLURM logs to be written in your home, or other small files.

/tmp on compute nodes

Each compute node has temporary storage attached to it. This is faster and larger than /home, but exists only while the node is running and is destroyed when the node shuts down. It's only useful for working storage while your job is running, and so as the last step of your job, you must copy results that you want to save from /tmp to a permanent storage location, like Google Cloud Storage.

Google Cloud Source Repository

Each Eureka App VM instance has [Google Cloud Source Repository](#) set up and enabled for sharing code files between multiple users on a shared Eureka instance.

Note that sensitive data like PHI should never be included in code files. This includes those shared on other code sharing platforms like GitHub.