



# Billing and Secure Collaboration

Managing cloud costs and collaborating securely



# Session Goals

- ▶ What cloud computing vendors work with Terra?
- ▶ What costs money in Terra and how do you pay for it?
- ▶ What is a Terra billing project and how is it connected with a workspace?
- ▶ Where can you see how much you've spent?
- ▶ Any recommendations for getting started with running analyses and securely sharing workspaces?

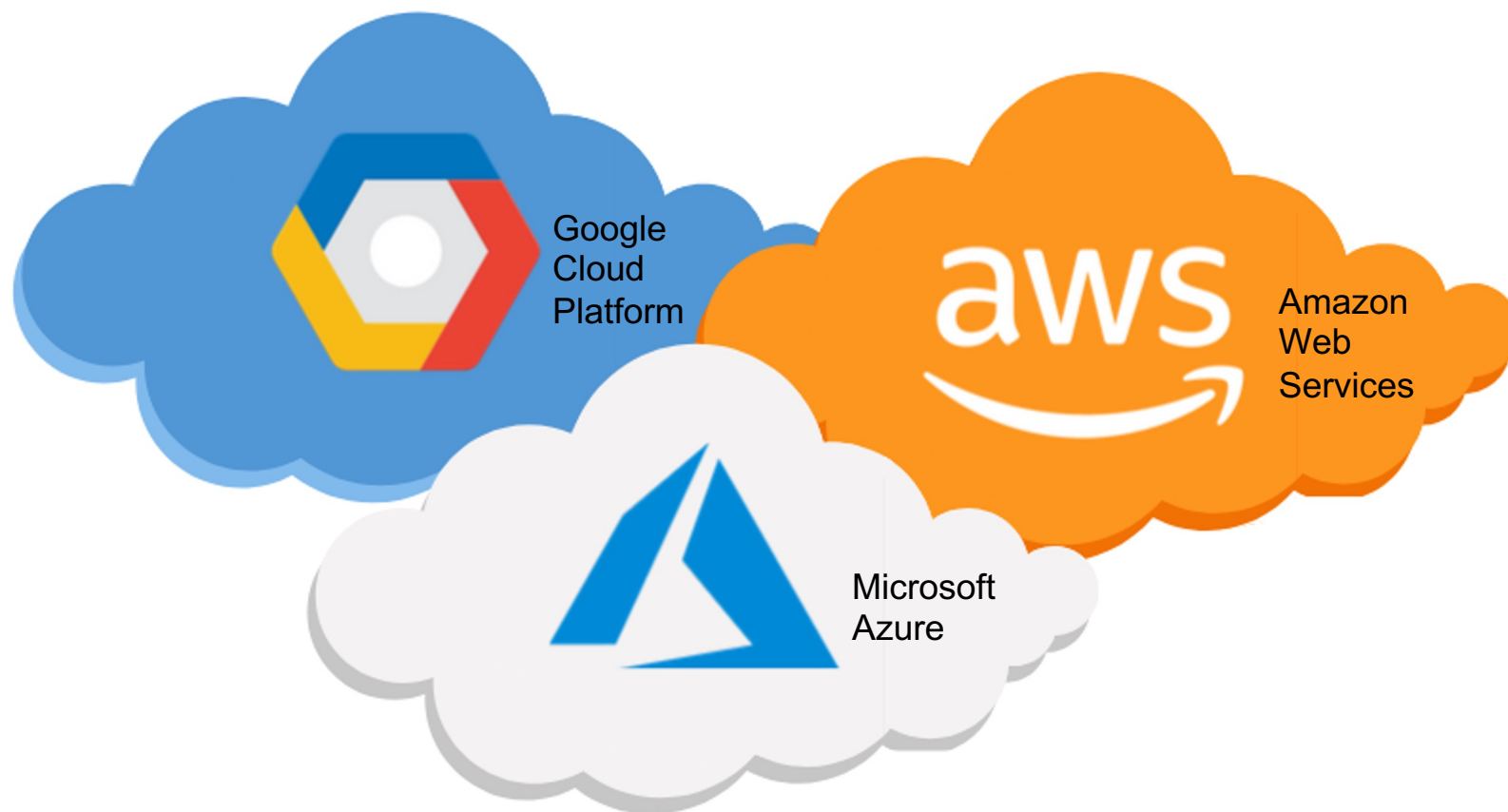


# Cloud Vendors



# Pay for what you use in “the cloud”

- ▶ Terra is free to sign up for and browse
- ▶ Computing services cost money
- ▶ Cloud providers charge for usage





# What are the costs of working on Terra?

## 1. Compute costs

- ▶ Instance type (high performance costs more; preemptibles cost less)
- ▶ CPU hours
- ▶ Memory & Disk

## 2. Data storage


- ▶ Anything in the Workspace bucket
- ▶ Workflow submission outputs (whether the workflow succeeds or fails)

## 3. Data egress (i.e. moving data) costs

- ▶ Downloading data to a non GCP location
- ▶ Moving hosted data to a GCS bucket in a different region



# Costs are organized by workspace

 BETA WORKSPACES

WORKSPACES +

Tags  Project Owner  Billing project  Submission status

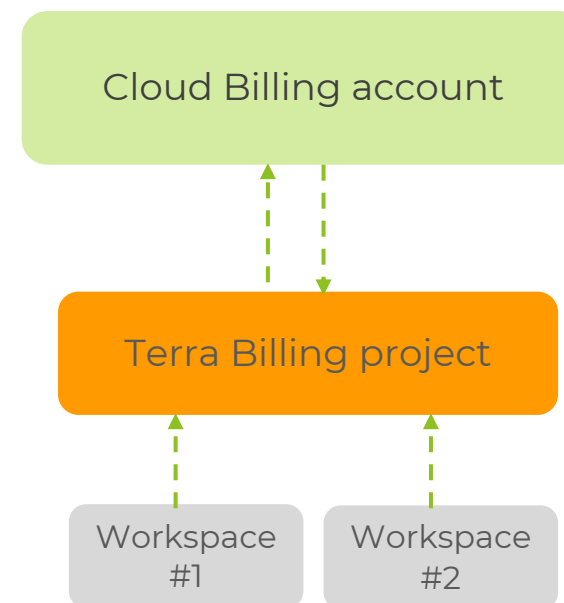
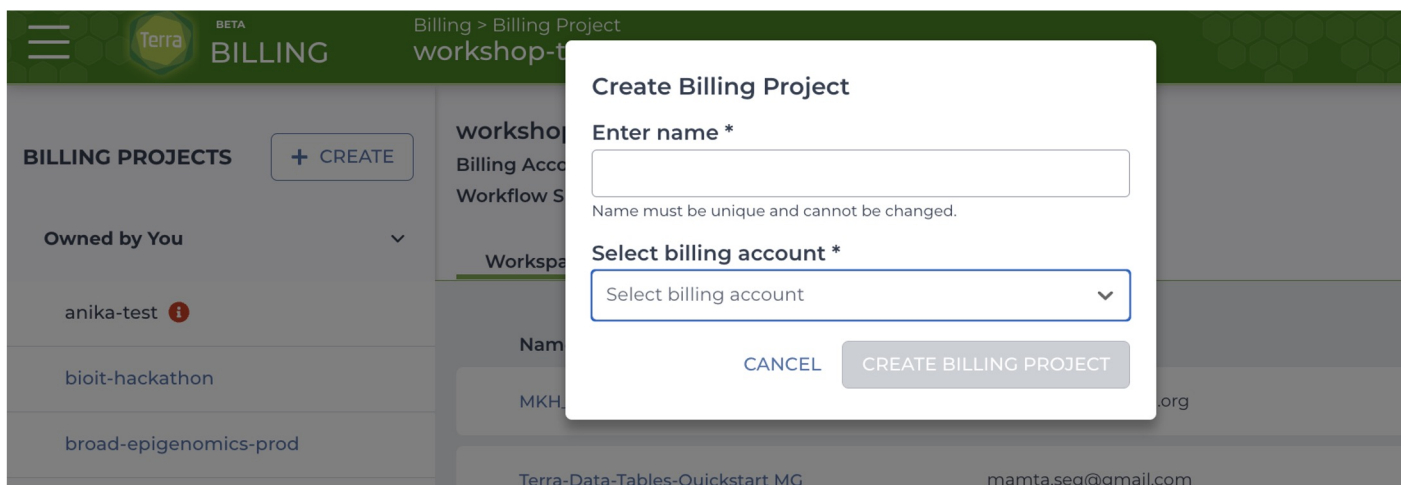
MY WORKSPACES (1228) NEW AND INTERESTING (1) **FEATURED (22)** PUBLIC (76)

| Name ↓  | Last Modified | Created By                  | Access Level  |
|---|---------------|-----------------------------|---------------|
| <b>Bioconductor</b><br>Explore common Bioconductor packages that can be used to perform bulk RNA differential expressio...    | Jul 15, 2021  | ekiernan@broadinstitute.org | Project Owner |
| <b>GATK-Structural-Variants-Single-Sample</b><br>### GATK Best Practices for Structural Variation Discovery on Single Samples | Mar 28, 2022  | cwhelan@broadinstitute.org  | Project Owner |
| <b>GATK4-Germline-Preprocessing-VariantCalling-JointCalling</b><br>### GATK Best Practices for Germline SNPs & Indels         | Feb 22, 2022  | tiffanym@broadinstitute.org | Project Owner |
| <b>GATK4-RNA-Germline-VariantCalling</b><br>### GATK Best Practices for Germline Variant Calling in RNAseq                    | Nov 15, 2021  | bshifaw@broadinstitute.org  | Project Owner |
| <b>GATKTutorials-Germline</b><br># What's in this workspace?  | Sep 28, 2021  | knoblett@broadinstitute.org | Project Owner |
| <b>GATKTutorials-Pipelining</b><br># What's in this workspace?  | Sep 16, 2021  | knoblett@broadinstitute.org | Project Owner |
| <b>GATKTutorials-Somatic</b><br># What's in this workspace?   | Sep 30, 2021  | knoblett@broadinstitute.org | Project Owner |
| <b>Germline-CNVs-GATK4</b><br>### GATK Best Practices for Germline Copy Number Variation                                      | Jan 13, 2022  | bshifaw@broadinstitute.org  | Project Owner |
| <b>HCA_Optimus_Pipeline</b><br># Optimus Pipeline for Analysis of 3' Single-cell and Single-nucleus Transcriptomic Data       | Oct 28, 2021  | ekiernan@broadinstitute.org | Project Owner |



# Terra Billing Project

- Terra billing project – a feature that links a cloud billing account with a workspace





# Compute Costs



# Estimating workflow spend

- Find cost estimates in featured workspace dashboard descriptions

CRAM\_To\_BAM

| Sample Name   | Sample Total Size | Time     | Cost \$ |
|---------------|-------------------|----------|---------|
| NA12878_small | 6.84 MB           | 00:07:00 | 0.02    |

- Use **built-in cost reporting** in the Job History page (see how to set it up [here](#))

The screenshot shows the Terra Job History page for a submission. The 'Total Run Cost' is highlighted as \$0.02. Below, a table lists the workflow details, with the 'Run Cost' column also highlighted.

| Data Entity        | Last Changed          | Status      | Run Cost | Messages | Workflow ID                       | Links |
|--------------------|-----------------------|-------------|----------|----------|-----------------------------------|-------|
| my_sample (sample) | May 17, 2021, 8:57 AM | ✓ Succeeded | \$0.02   |          | 128542af-5345-46f8-8f73-b3017b... |       |

- 1) *Run on a small sample*
- 2) *Multiply to scale*
- 3) *Add buffer for failed jobs*



# Estimating interactive analysis costs

**Cloud Environment** ✕

A cloud environment consists of application configuration, cloud compute and persistent disk(s).

|                            |                           |                         |
|----------------------------|---------------------------|-------------------------|
| Running cloud compute cost | Paused cloud compute cost | Persistent disk cost    |
| <b>\$0.06 per hr</b>       | <b>&lt; \$0.01 per hr</b> | <b>\$2.00 per month</b> |

---

**Application configuration** ⓘ

Default: (GATK 4.2.0.0, Python 3.7.10, R 4.1.1) ▼

[What's installed on this environment?](#) Updated: Sep 23, 2021  
Version: 2.0.3

---

**Cloud compute profile**

CPU: 1 ▼ Memory (GB): 3.75 ▼

☐ **Enable GPUs** BETA [Learn more about GPU cost and restrictions.](#)

**Startup script**

URI

**Compute type**

Standard VM ▼

---

**Persistent disk size (GB)**

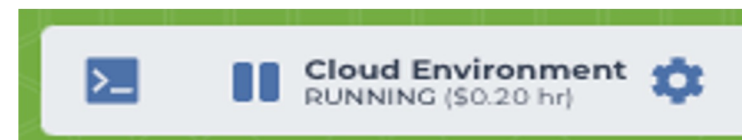
Persistent disks store analysis data. [Learn more about persistent disks and where your disk is mounted.](#)

50

**CREATE**

**Adjust settings to optimize cost**  
(Compute Type, CPU, and disk)

**Estimate costs using real-time cost/hour in Cloud Environment widget**



- Updates based on the machine configuration you choose
- Total cost (estimate) = (cost per hour) x (hours the instance will be active) + cost of the Persistent Disk
- Autopause function saves money! Delete when you are done!



# Storage Costs



# Estimating cloud storage costs

The screenshot shows the Terra workspace interface for a workspace named "amp-t2d-op/2019\_ASHG\_Reproduc...". The top navigation bar includes "WORKSPACES", "DASHBOARD", "DATA", "NOTEBOOKS", "WORKFLOWS", and "JOB HISTORY". The "ABOUT THE WORKSPACE" section describes the workspace's purpose and structure. The "Workspace Information" panel on the right displays key metrics: Creation Date (3/19/2020), Last Updated (9/27/2021), Submissions (2), Access Level (Writer), Estimated Monthly Cost (\$0.07), and Google Project ID (amp-t2d-op). An orange arrow points to the "EST. \$/MONTH \$0.07" value. The "OWNERS" section lists two email addresses, and the "TAGS" section shows a list of tags including "1000 Genomes", "GWAS", "Jupyter Notebooks", and "WDLs". The "Google Bucket" section provides details about the storage location.

**ABOUT THE WORKSPACE**

This workspace reproduces the fundamental steps in a genome wide association study (GWAS), using 1,000 Genomes Project<sup>1</sup> (phase 3) genotypes and simulated phenotypes.

The analysis is structured in two parts:

1. Explore phenotypes and population structure (Jupyter Notebook - Hail/Python)
2. Test for genetic associations using mixed-models and generate summary visualizations (WDL workflow)

The output of the notebook (part 1) serves as the input to the workflow (part 2).

Instructions for applying the analyses presented in this workspace on your own data are provided in the penultimate section of this documentation.

**Notes on data in this workspace**

To demonstrate an analysis that could be run on typical whole genome sequence data, this workspace provides mock phenotype data generated from publicly available 1000 Genomes phase 3 genotypes. Phenotypes have been simulated based on individual genotypes and known associated loci for multiple complex traits. The [GCTA software](#)<sup>6</sup> was used with lists of causal variants and an estimate of narrow sense heritability<sup>5</sup> for each phenotype.

**Traits and sources for causal variants**

a. BMI: Giant-UKBB meta-analysis<sup>2</sup>

**WORKSPACE INFORMATION**

|                            |                                 |
|----------------------------|---------------------------------|
| CREATION DATE<br>3/19/2020 | LAST UPDATED<br>9/27/2021       |
| SUBMISSIONS<br>2           | ACCESS LEVEL<br>Writer          |
| EST. \$/MONTH<br>\$0.07    | GOOGLE PROJECT ID<br>amp-t2d-op |

**OWNERS**

tmajaria@broadinstitute.org  
bshifaw@broadinstitute.org

**TAGS**

Add a tag

1000 Genomes × GWAS ×  
Jupyter Notebooks × WDLs ×

**Google Bucket**

Name: fc-d5eb5311-1cba-4c8f-84c5-...  
Location: multi-region: US  
[Open in browser](#)

*Estimated cloud storage costs for your workspace*



# Egress Costs



# Estimating egress costs

WORKSPACES

DASHBOARD DATA NOTEBOOKS

TABLES

|                               |                          |                 |
|-------------------------------|--------------------------|-----------------|
| aligned_reads_index (2504)    | <input type="checkbox"/> | align           |
| aliquot (2504)                | <input type="checkbox"/> | 000             |
| germline_variation_... (2504) | <input type="checkbox"/> | 001             |
| pheno-data (2504)             | <input type="checkbox"/> | 003             |
| program (1)                   | <input type="checkbox"/> | 005             |
| project (1)                   | <input type="checkbox"/> | 007             |
| read_group (2504)             | <input type="checkbox"/> | 0098ed18-93...  |
| reference_file (44)           | <input type="checkbox"/> | 00acfc66-78b... |
| sample (2504)                 | <input type="checkbox"/> | 00b72a8c-dc7... |
| simple_germline_va... (2504)  | <input type="checkbox"/> | 00c6b6fe-9d6... |
| study (1)                     | <input type="checkbox"/> | 00dad5ac-683... |

### File Details

Filename  
HG03642.final.cram.crai

DOS uri's can't be previewed

File size  
1.31 MB

[View this file in the Google Cloud Storage Browser](#)

**DOWNLOAD FOR < \$0.01\***

Terminal download command  
`gsutil cp gs://nih-nhlbi-biodata-catalys`

[More Information](#)

\* Estimated. Download cost may be higher in China or Australia.

**DONE**

***Find price to egress***

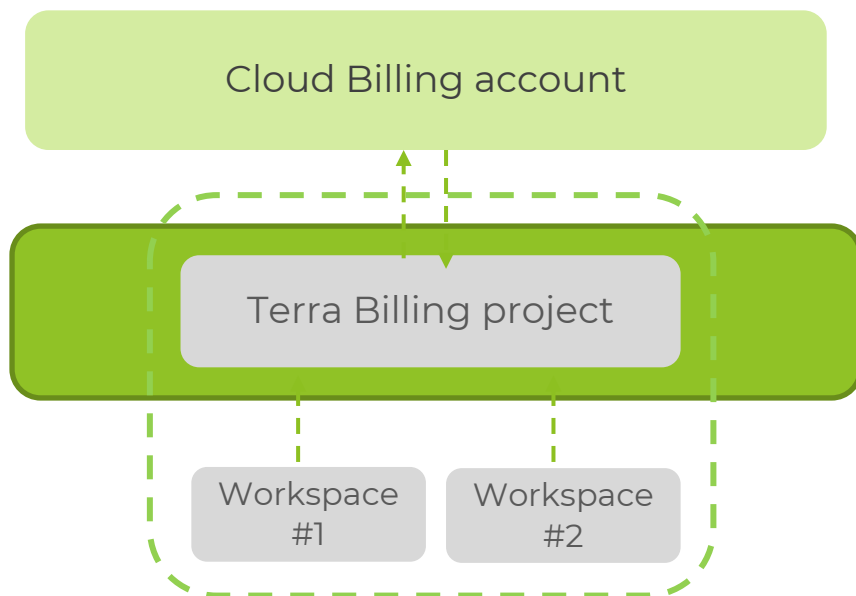
Will you download  
derived data to save  
locally or elsewhere?



# Roles & permissions



# Terra Billing project roles



Billing roles determine who can create/manage resources

- ▶ Owners & Users can create Terra workspaces
- ▶ Owners can grant Terra Billing project permissions to other collaborators



# Terra Billing project roles

The screenshot shows the Terra Billing Project interface. On the left, a sidebar lists billing projects: 'ac-billing-08-19-2021', 'ac-test-09-21-2021' (selected), 'ashg-2018-02543846-f8da', and 'billing-project-1'. The main content area shows details for 'ac-test-09-21-2021', including the Billing Account 'NIH.NHLBI.BDC.Cohort3.Fellow.008' and Workflow Spend Report Configuration. The 'Users' tab is active, displaying a table with columns 'Email' and 'Roles'. A blue '+ Add User' button is located above the table. An orange arrow points from this button to a modal dialog titled 'Add user to Billing Project'.

## Roles include

- Owner (i.e. Project Owner)
- User

### Add user to Billing Project

User email \*

#### Role

☐ Can manage users (Owner)

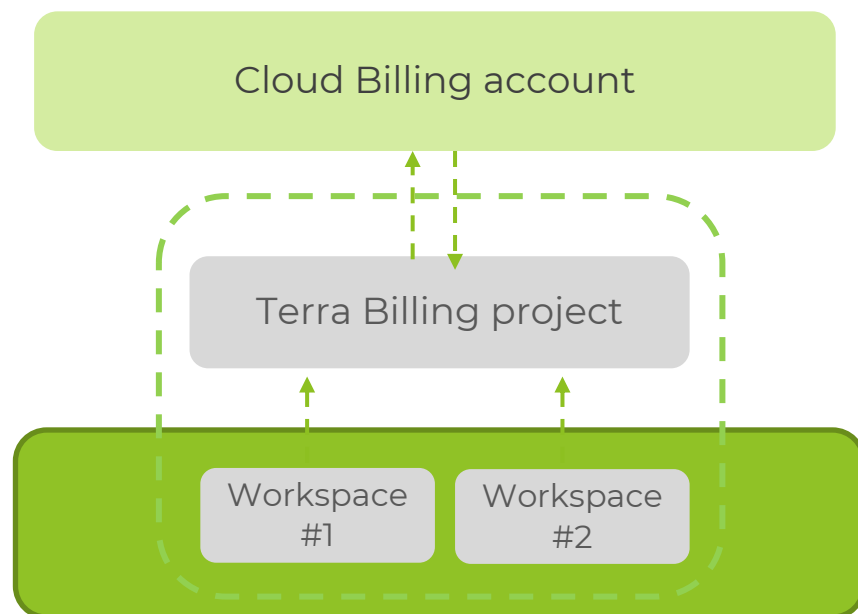
Warning: Adding any user to this project will mean they can incur costs to the billing associated with this project.

CANCEL

ADD USER



# Workspace roles



Workspace roles determine who can perform actions with a cost (spend money)

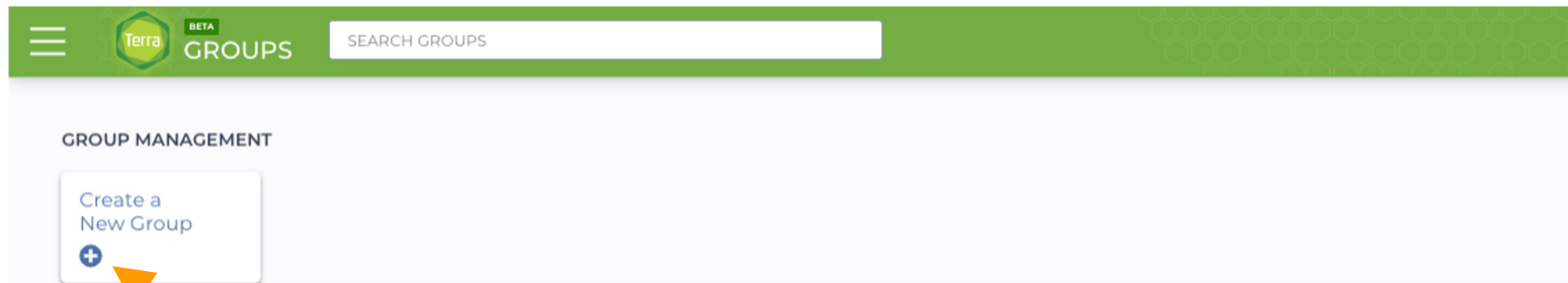
► **Roles** include

- Owner
- Writer
- Reader
- Can share
- Can compute

***Costs are charged to the workspace Billing project – not the user!***



# Streamline shared access with Groups



- ▶ Add groups at **Main Menu navigation > groups**
- ▶ Assign **roles** en masse for a given **resource**
- ▶ Can be updated as needed
  - ▶ Member - Has access to any resources granted group permission
  - ▶ Admin - May add or remove members or other admins to the group - also members of the group
- ▶ Permissions across all resources updates when group is updated



# Authorization Domains can contain one or more groups

## ► Authorization Domain-

A group or multiple groups that are allowed to access a workspace.

## ► Example ->

If you are in both ASHG18-dev and the Broad\_Product\_Team groups, you can access this workspace if you are given Owner, Writer, or Reader role.

**Create a New Workspace**

Workspace name \*  
Example\_of\_Auth\_domain

Billing project \*  
help-terra

Bucket location ⓘ  
US multi-regional (default)

Description  
Enter a description

Authorization domain ⓘ  
ASHG18-dev x Broad\_Product\_Team x

CANCEL CREATE WORKSPACE



# Setting up billing

Options for working in Terra

How to set up a Cloud Billing account



# How to pay for costs when working in Terra

- ▶ A collaborator can grant you “can compute” permission in a [shared workspace](#)
  - ▶ **Your** costs are covered by **their** Billing project



# How to pay for costs when working in Terra

- ▶ A collaborator can grant you “can compute” permission in a [shared workspace](#)
  - ▶ **Your** costs are covered by **their** Billing project
- ▶ The owner of an existing [Terra Billing project](#) can authorize you to use their Billing project
  - ▶ You can create/clone a workspace to work in
  - ▶ **Your** workspace costs are covered by **their** Billing project



# How to pay for costs when working in Terra

- ▶ A collaborator can grant you “can compute” permission in a [shared workspace](#)
  - ▶ **Your** costs are covered by **their** Billing project).
- ▶ The owner of an existing [Terra Billing project](#) can authorize you to use their Billing project
  - ▶ You can create/clone a workspace to work in
  - ▶ **Your** workspace costs are covered by **their** Billing project
- ▶ Use your own Cloud Billing
  - ▶ Set up a Cloud Billing account
  - ▶ [Link to Terra](#)
  - ▶ Create a new Terra Billing Project
  - ▶ Create or clone a workspace



# How to set up a Google Cloud Billing account

- ▶ Option 1: Create your own Cloud Billing account on GCP console
  - ▶ \$300 Google [Getting Started credits](#)
  - ▶ Institution or personal [credit card](#)
- ▶ Option 2: Use a [third-party reseller](#) (such as [STRIDES](#) for NIH or NIH funded researchers or Onix Networking)

To learn more, see: [How to set up billing in Terra](#)

For step-by-step instructions to set up a Google Cloud Billing account and link it to Terra, see: [Set up billing with \\$300 Google credits to explore Terra](#)



# Session Goals



# Revisit Session Goals

- ▶ What cloud vendors work with Terra?
- ▶ What costs money in Terra and how do you pay for it?
- ▶ What is a Terra billing project and how is it connected with a workspace?
- ▶ Where can you see how much you've spent in Terra?
- ▶ Any recommendations for getting started with running workflows, cloud environments, and securely sharing workspaces?



# Revisit Session Goals

- ▶ What cloud vendors work with Terra?



# Revisit Session Goals

- ▶ What cloud vendors work with Terra?

Google Cloud Platform and Azure (soon)



# Revisit Session Goals

- ▶ What cloud vendors work with Terra?

Google Cloud Platform and Azure (soon)

- ▶ What costs money in Terra and how do you pay for it?



# Revisit Session Goals

- What cloud vendors work with Terra?

Google Cloud Platform and Azure (soon)

- What costs money in Terra and how do you pay for it?

Compute, data storage, and egress cost money.

Pay with Terra Billing project linked to Cloud billing account.



# Revisit Session Goals

- ▶ What is a Terra billing project and how is it connected with a workspace?



# Revisit Session Goals

- ▶ What is a Terra billing project and how is it connected with a workspace?

A feature in Terra that connects a Cloud billing account and a workspace so that costs accrued in the workspace get paid for.



# Revisit Session Goals

- ▶ What is a Terra billing project and how is it connected with a workspace?

A feature in Terra that connects a Cloud billing account and a workspace so that costs accrued in the workspace get paid for.

- ▶ Where can you see how much you've spent in Terra?



# Revisit Session Goals

- ▶ What is a Terra billing project and how is it connected with a workspace?

A feature in Terra that connects a Cloud billing account and a workspace so that costs accrued in the workspace get paid for.

- ▶ Where can you see how much you've spent in Terra?

Combination of the Job History tab & Workspace dashboard. The source of truth is in the Google Cloud console.



# Revisit Session Goals

- ▶ Any recommendations for getting started with running workflows, cloud environments, and securely sharing workspaces?



# Revisit Session Goals

- ▶ Any recommendations for getting started with running workflows, cloud environments, and securely sharing workspaces?

Start small. Run workflows you are interested in on one sample first.

See how much it cost. Then scale up.

When you are done using a cloud environment, delete it.

Use groups to share and only share with folks you are comfortable with having the ability to spend \$.



Questions?