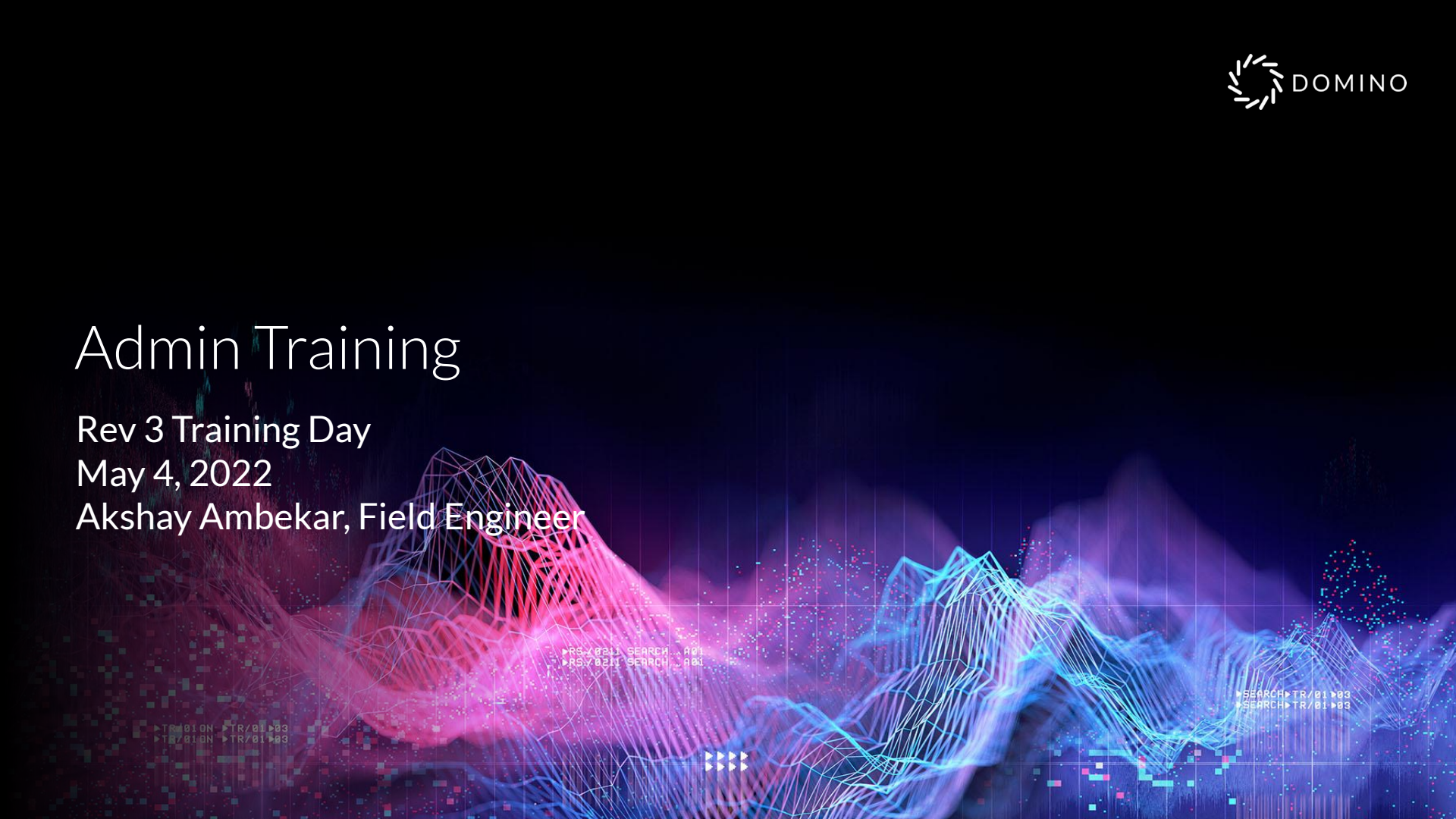
Admin Training

Rev 3 Training Day
May 4, 2022
Akshay Ambekar, Field Engineer



Admin Course Objectives

DOMINO

Upon Completion of this Course, Learners will be able to:

Describe the Domino Architecture, Components, and Configuration.

Explain Managing the Domino Environment & Infrastructure.

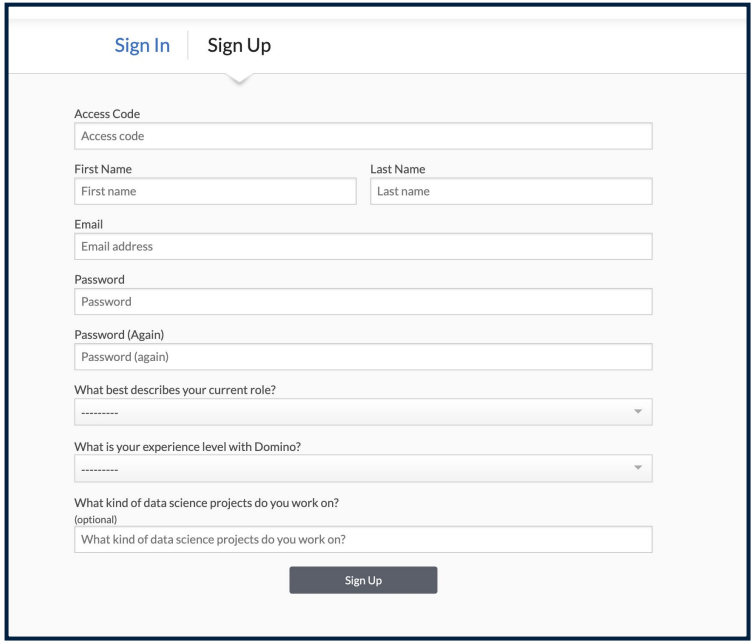
Discuss Domino Storage and how to Manage it.

Define Optimizing and Troubleshooting Domino.

Describe Domino Monitoring, Data Management, and Security.

DOMINO ACADEMY

1. Go to learn.dominodatalab.com
2. Click on “Sign Up”
3. Input your access code 2sypw-iehkb44g
4. Select the Admin Training course



PREREQUISITES

You are familiar with Kubernetes.

The level of required familiarity will depend on how deep you want to go.

Recommendations for learning:

- A Cloud Guru online course for surface level Kubernetes (4 hours)
- Online training from Linux Foundation to train for Kubernetes Admin Certification (35 hours + homework)

PREREQUISITES

Skillsets for Domino Administration

The following table lists some of the skillsets needed to administer Domino, and a few options for how to learn about various technologies. You may spread administration among numerous people, in which case these tasks would be shared, or you may have a Domino managed instance where you do not need to do more than application level administration work.

Technology	Recommendations
Kubernetes	Kubernetes Basics, Cloud Guru Online Course (4 hours), Online Training from the Linux Foundation (35 hours)
Docker	Intro to Containers, VMs, and Docker, Docker for Beginners
Git	Learn Git, Git on the Command Line
Networking	Networking for Developers
Linux	Intro to Linux
Python Package Management	Overview of Python management tools, Intro to Python Modules and Packages
AWS (if using)	Understanding of the AWS services used for Domino. Consider taking the AWS Certified Cloud Practitioner exam.
Google Cloud (if using)	Understanding of the GKE services used for Domino.
Azure (if using)	Understanding of the AKS services used for Domino.

PREREQUISITES II

Admin access to Domino

During the presentation, we will take short pauses so that you can click around areas of the product to get familiar.

However, please don't modify anything during the admin training unless asked to do so.

RESPONSIBILITIES OF DOMINO ADMINISTRATORS

As ***infrastructure admins***

- Set up network connectivity to data sources and other external resources
- Maintain and support infrastructure
- Manage the compute grid (hardware tiers and resource quotas)

As ***application admins***

- Manage user access
- Administer global environments
- Manage application-level configurations (built-in features to Domino that require customization or enablement)
- Certify best practices and spreading those to the user community

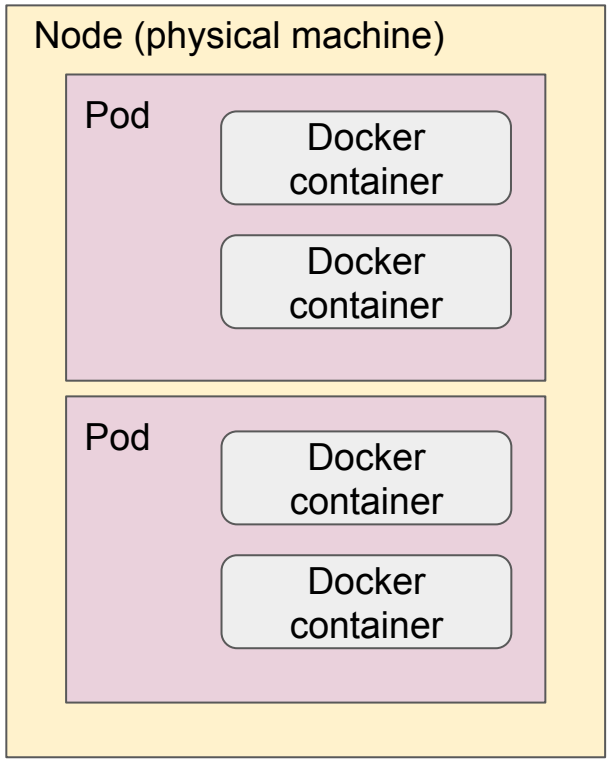
Both are responsible for

- Helping debug issues with Domino support

Introduction to Domino

KUBERNETES TERMINOLOGY REVIEW

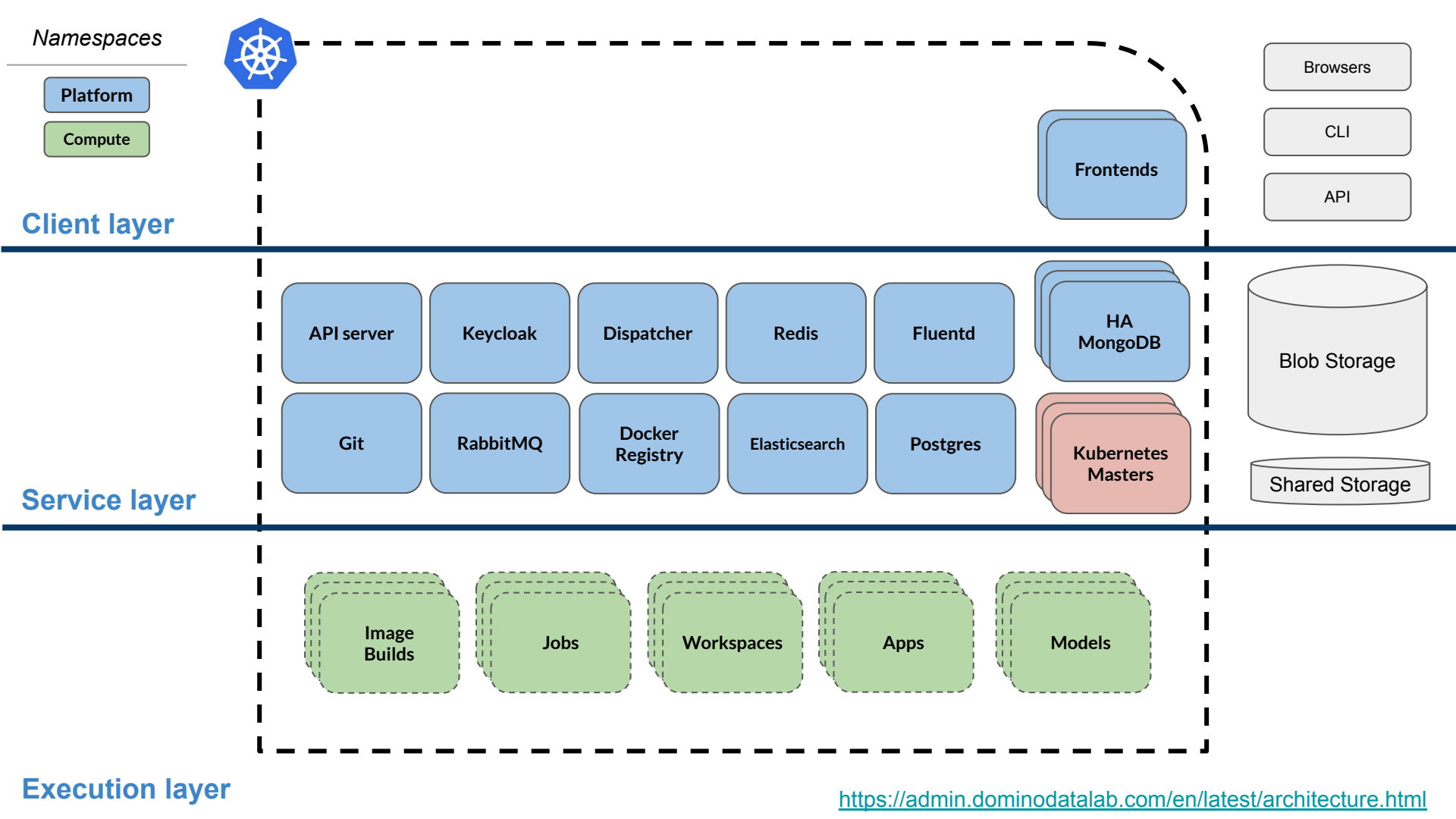
Docker Container	A standard unit of software that packages up code and all its dependencies
Pod	A group of containers, co-located and co-scheduled, to run some application-specific logic
Node	A worker machine, which can host pods
Namespace	A virtual cluster within the larger Kubernetes cluster



Overview of Critical Product Areas

1	Workspaces
2	Environments
3	Model APIs

DOMINO



https://admin.dominodatalab.com/en/latest/architecture.html

GETTING STARTED

Lesson on navigating the Admin Panel - https://learn.dominodatalab.com/admin-training/855680

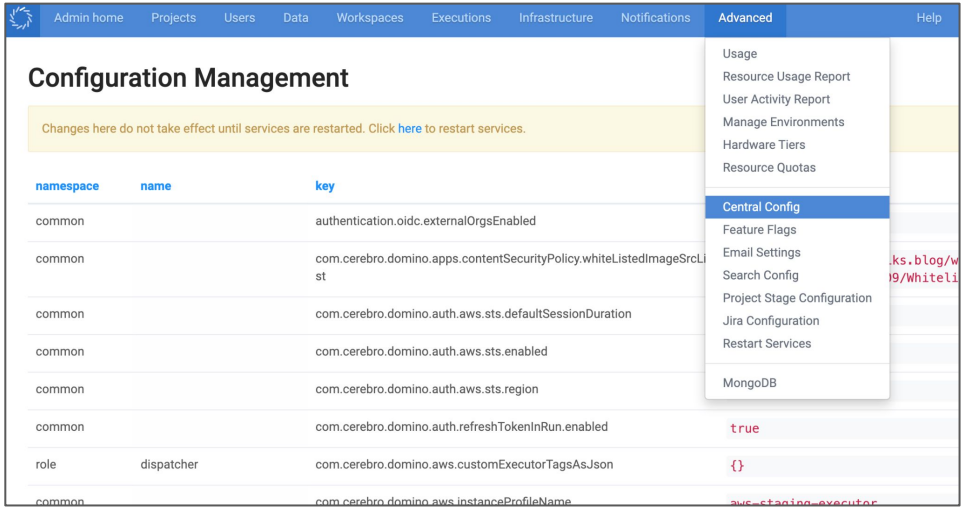
Set-up Checklist - https://learn.dominodatalab.com/admin-training/894574

Configuration

CENTRAL CONFIGS

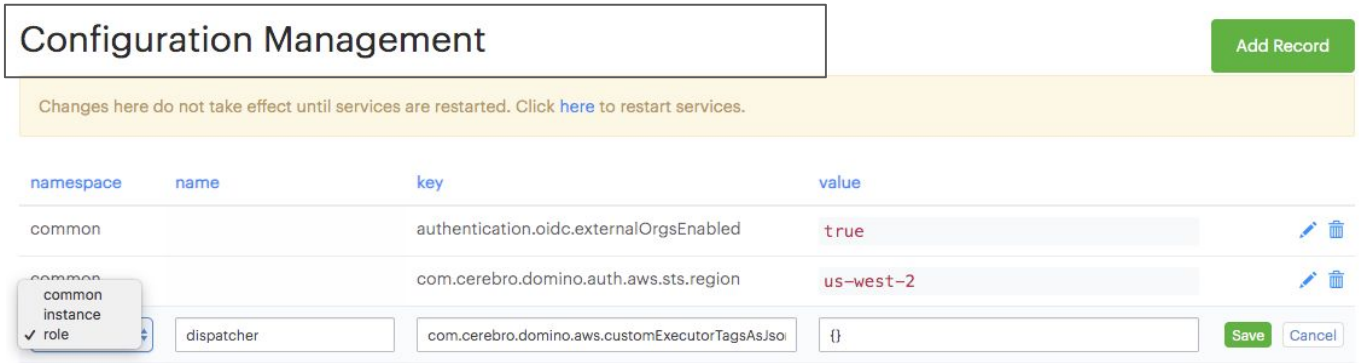
- Various configuration options of your Domino installation.
- In order for changes to take effect, you'll need to restart the Domino frontend process.

- ***BE CAREFUL!!!***
  - Typos can cause downtime!
  - We recommend consulting with Domino support before altering any values.



CENTRAL CONFIG COLUMNS

Column	Description
`namespace`	[`common`] NOT K8s namespace. Directs configuration at specific components of Domino.
`name`	[`<empty>`] If namespace `role` or `instance` are specified, `name` applies to the specific role or instance.
`key`	The key of the configuration
`value`	The value for the configuration



CENTRAL CONFIG DEFAULTS

If there is no record explicitly set, then the default value is used. See our docs for keys and default values:

https://admin.dominodatalab.com/en/latest/configuration/Central_configuration.html

Admin Lesson - https://learn.dominodatalab.com/admin-training/855692

ACTIVITY: LOCATE CENTRAL CONFIG VALUES

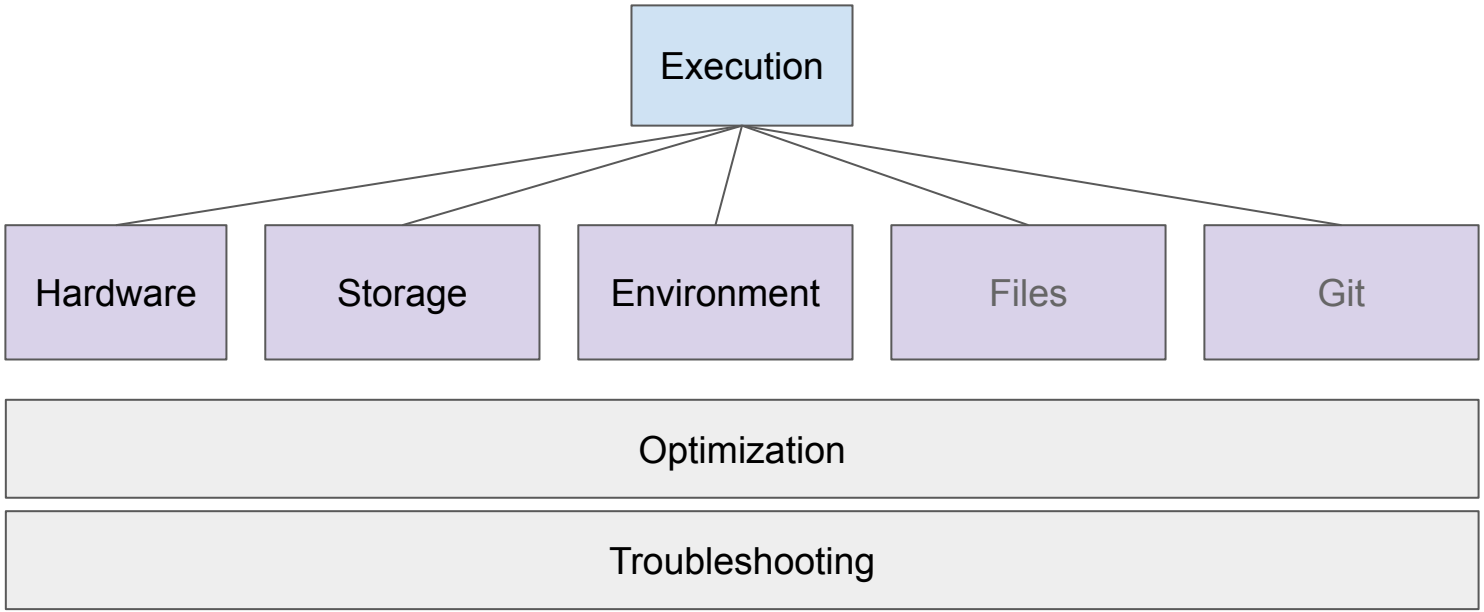
1. Sign into the temporary training Domino instance as an Admin:

   https://admin-rev.workshop.domino.tech/
2. Determine the central config values for the following:
   a. Maximum executions per user
   b. Default volume size for jobs and workspaces

Hint: The docs can be found here

Managing the Compute Grid Infrastructure

COMPONENTS REQUIRED FOR AN EXECUTION

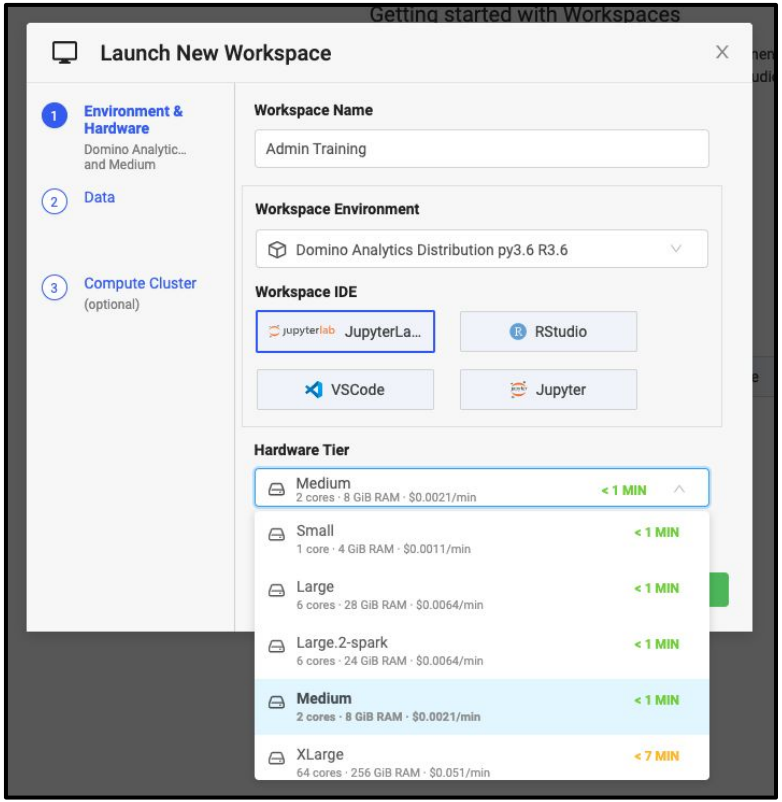


UNDERSTANDING THE STATE OF THE CLUSTER WHEN EXECUTING CODE

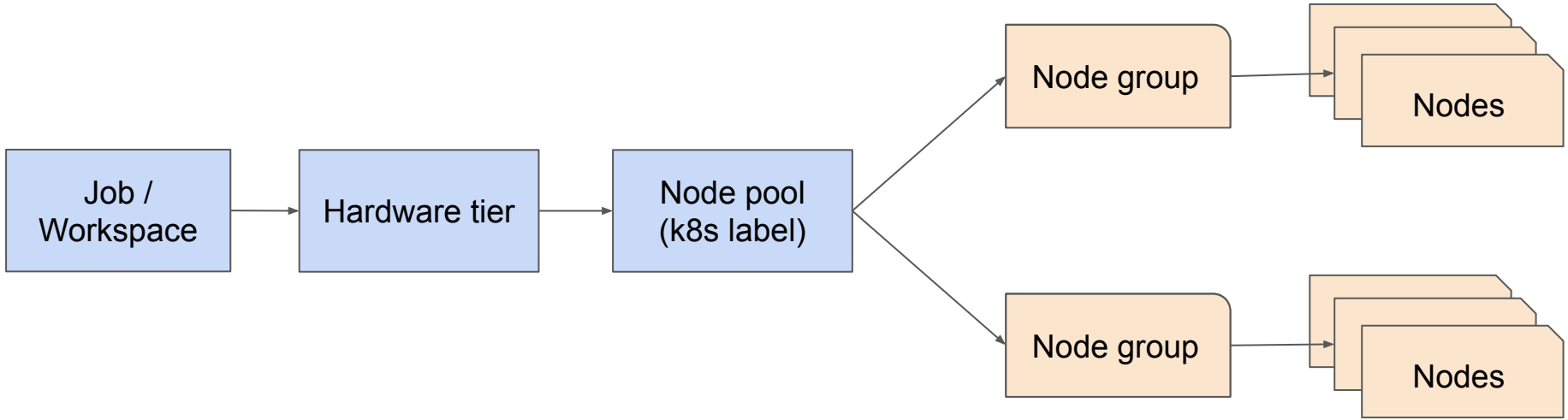
When running a Job or a Workspace, the workload gets scheduled on a hardware tier.

The hardware tier defines

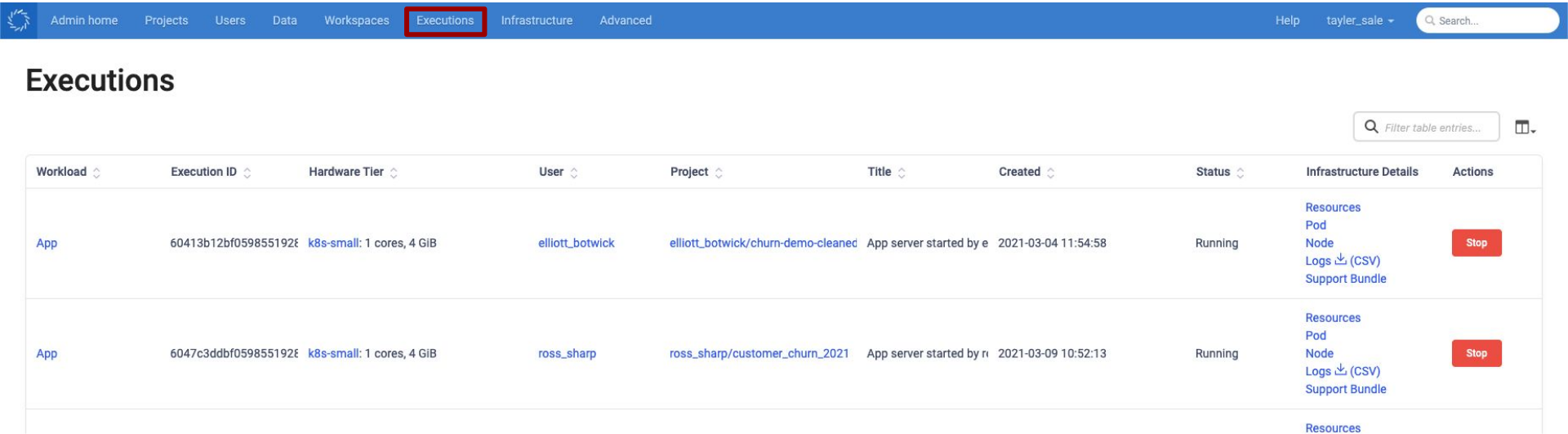
- The type of machine that it will run on
- The resource limits for the pod that the run will execute in



EXECUTION LIFECYCLE

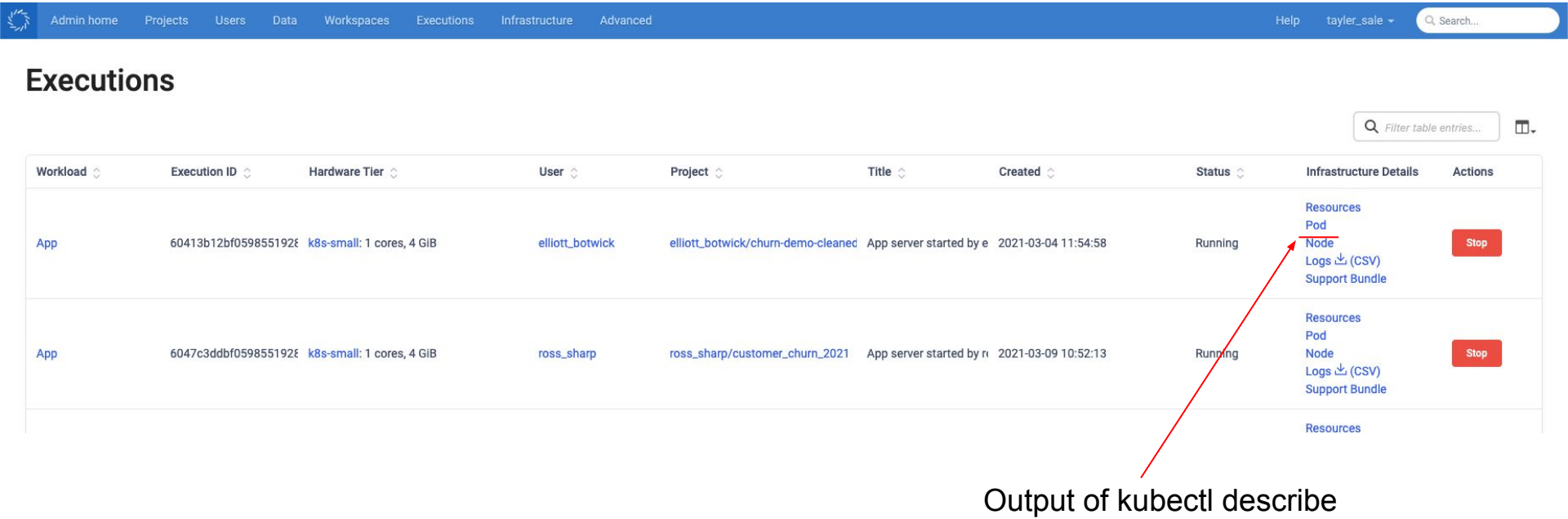


ADMIN PANEL - EXECUTIONS PAGE



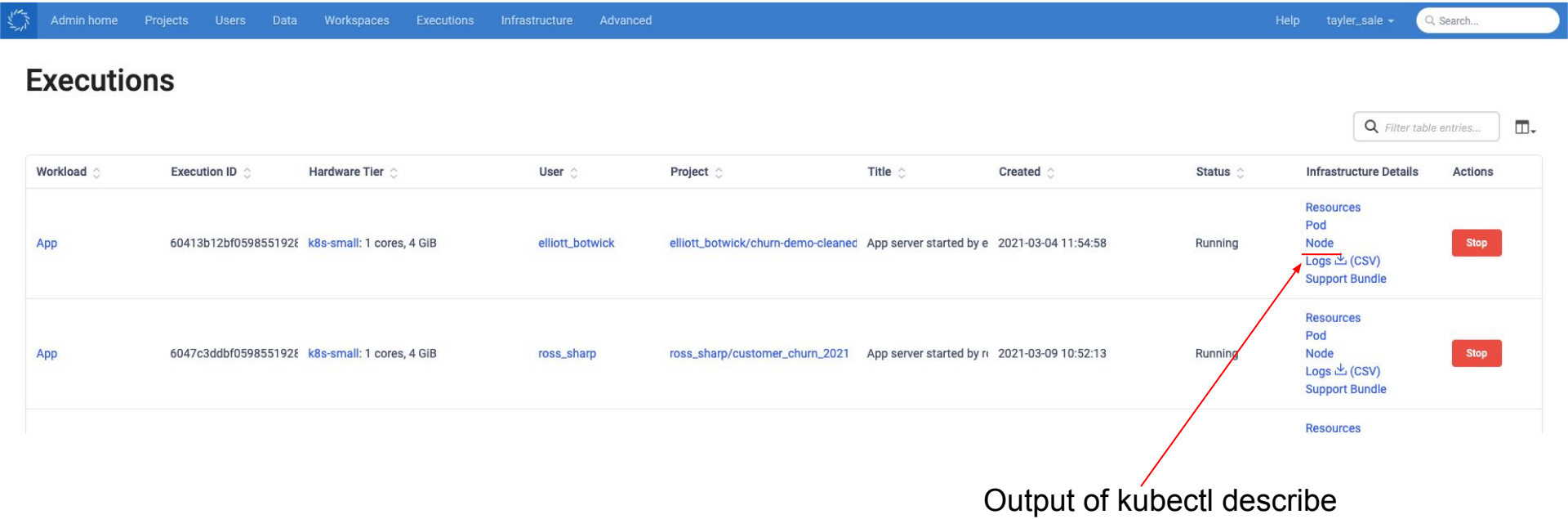
Admin lesson - https://learn.dominodatalab.com/admin-training/879952

ADMIN PANEL - EXECUTIONS PAGE



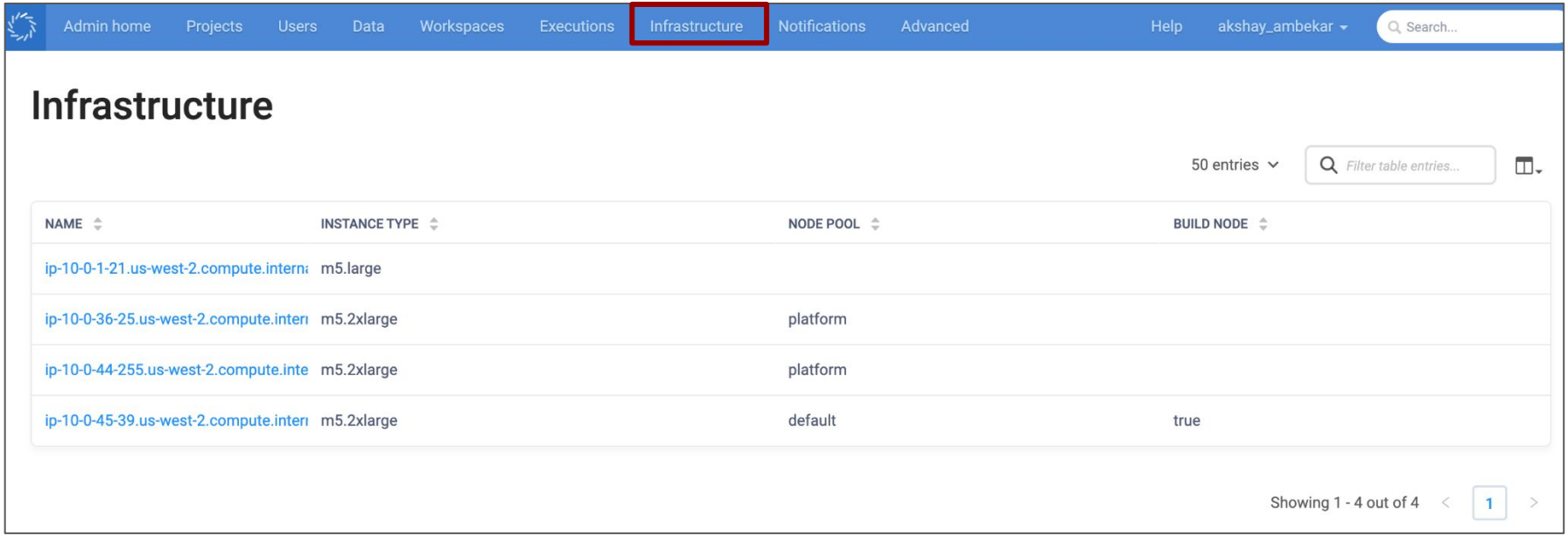
Output of kubectl describe

ADMIN PANEL - EXECUTIONS PAGE



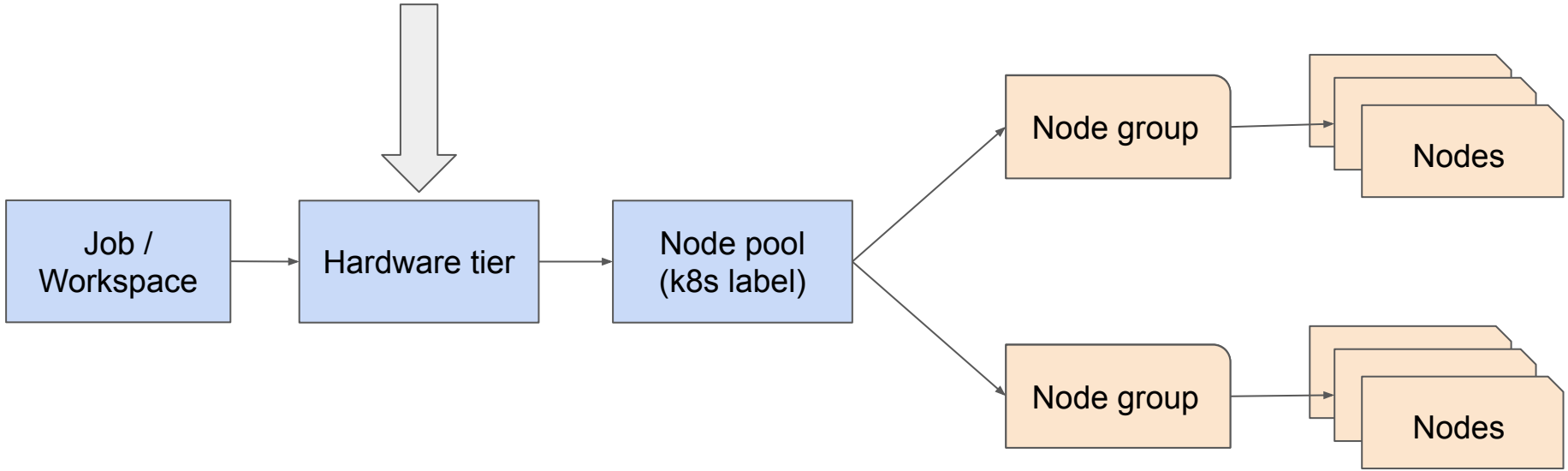
Output of kubectl describe

ADMIN PANEL - INFRASTRUCTURE PAGE



https://admin.dominodatalab.com/en/latest/compute/compute-grid.html#how-do-i-view-the-current-nodes-in-my-compute-grid

HOW DO YOU CREATE A NEW HARDWARE TIER?

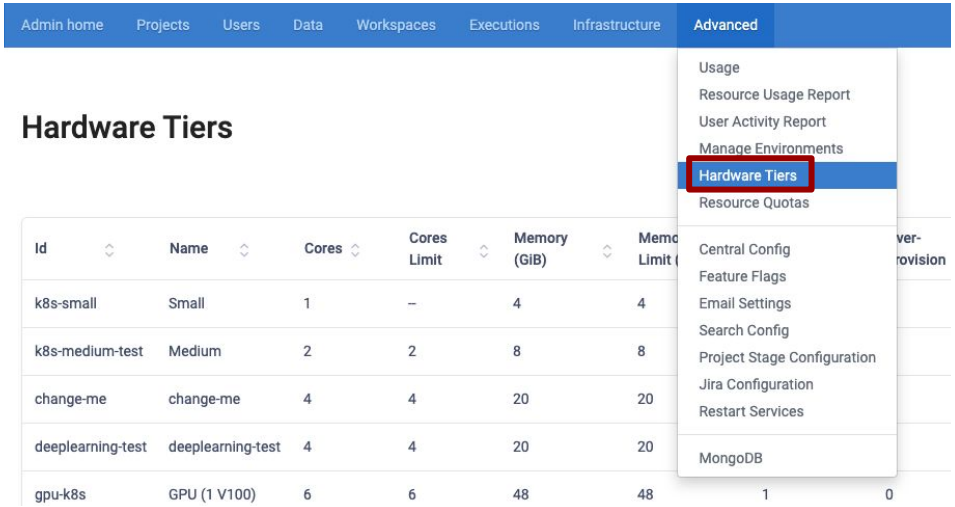


CREATING A HARDWARE TIER

1. Create a Node Pool in Kubernetes
2. Define the Hardware Tier using the Admin Panel

Node pools lesson - https://learn.dominodatalab.com/admin-training/894982

Hardware tiers lesson - https://learn.dominodatalab.com/admin-training/879953



HARDWARE TIER OVERHEAD

When defining hardware tiers, leave room for overhead.

As a rule of thumb, **node overhead is 1.5 cores and 2 GiB of RAM.**

Additional **overhead per execution is about 1 core and 1.5 GiB.**

Taking account of overhead is especially important when you have only 1 concurrent execution per node.

https://admin.dominodatalab.com/en/latest/compute/hardware-tiers.html#accounting-for-overhead

ACTIVITY: DIAGNOSE ERROR WITH HARDWARE TIER

1. Sign into the temporary training Domino instance as an Admin:

   https://admin-rev.workshop.domino.tech/

2. Find the logs for this failed workspace:

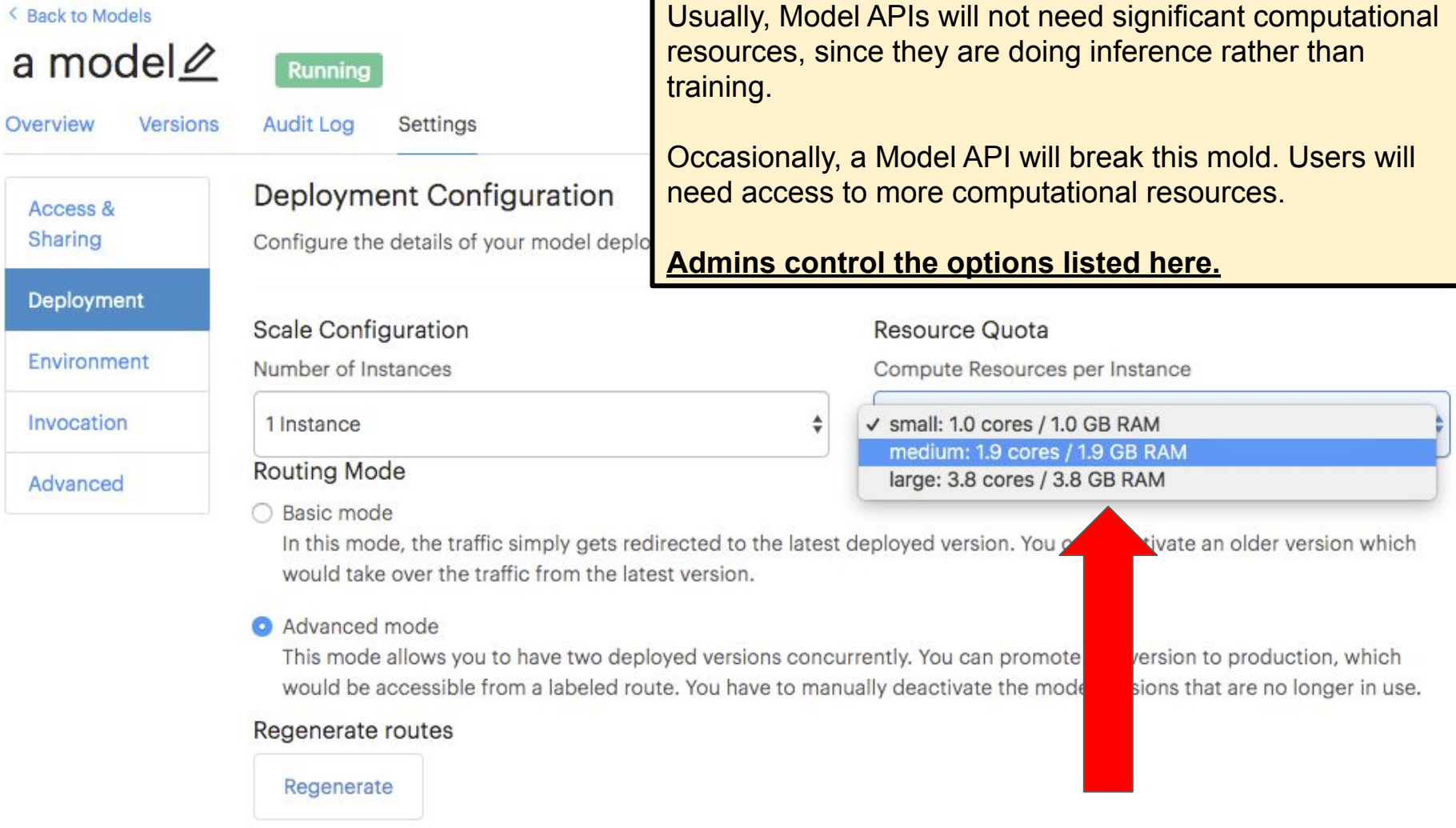
3. Determine why the hardware tier didn't work

Managing Model API Resources

Usually, Model APIs will not need significant computational resources, since they are doing inference rather than training.

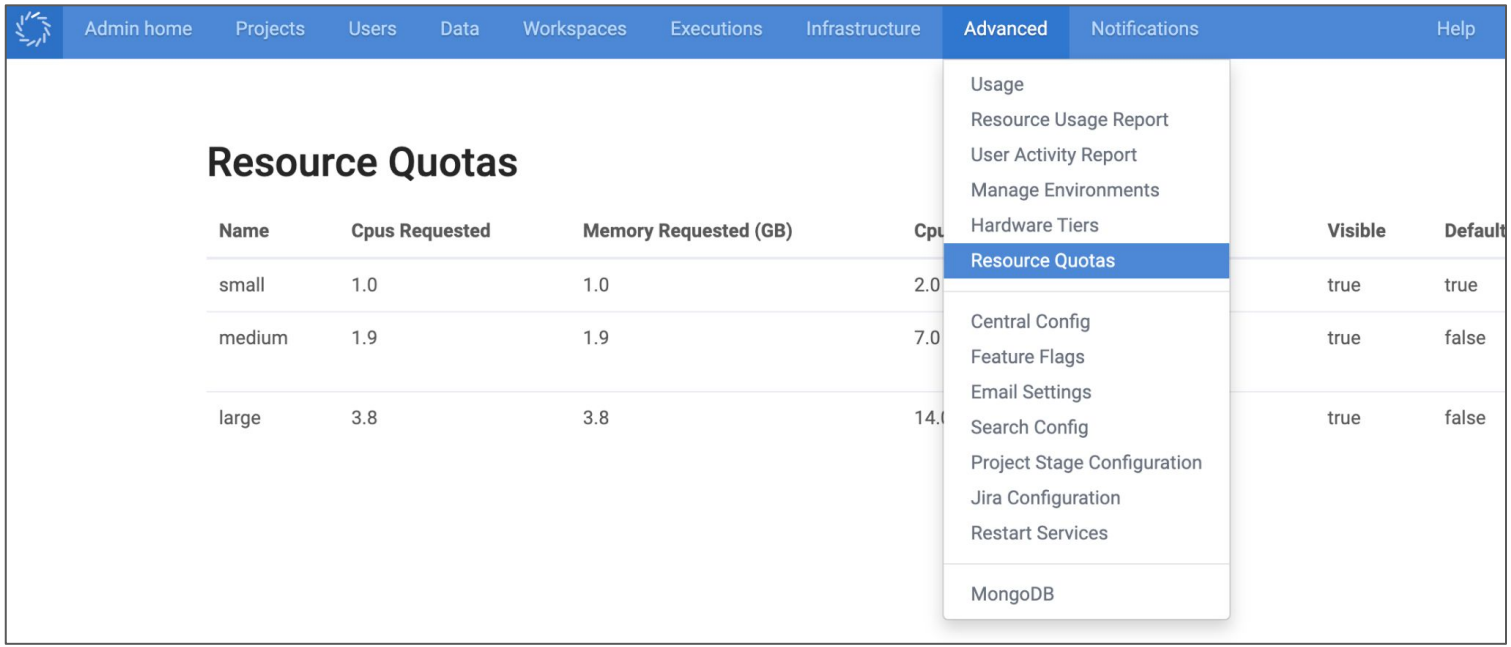
Occasionally, a Model API will break this mold. Users will need access to more computational resources.

<u>Admins control the options listed here.</u>

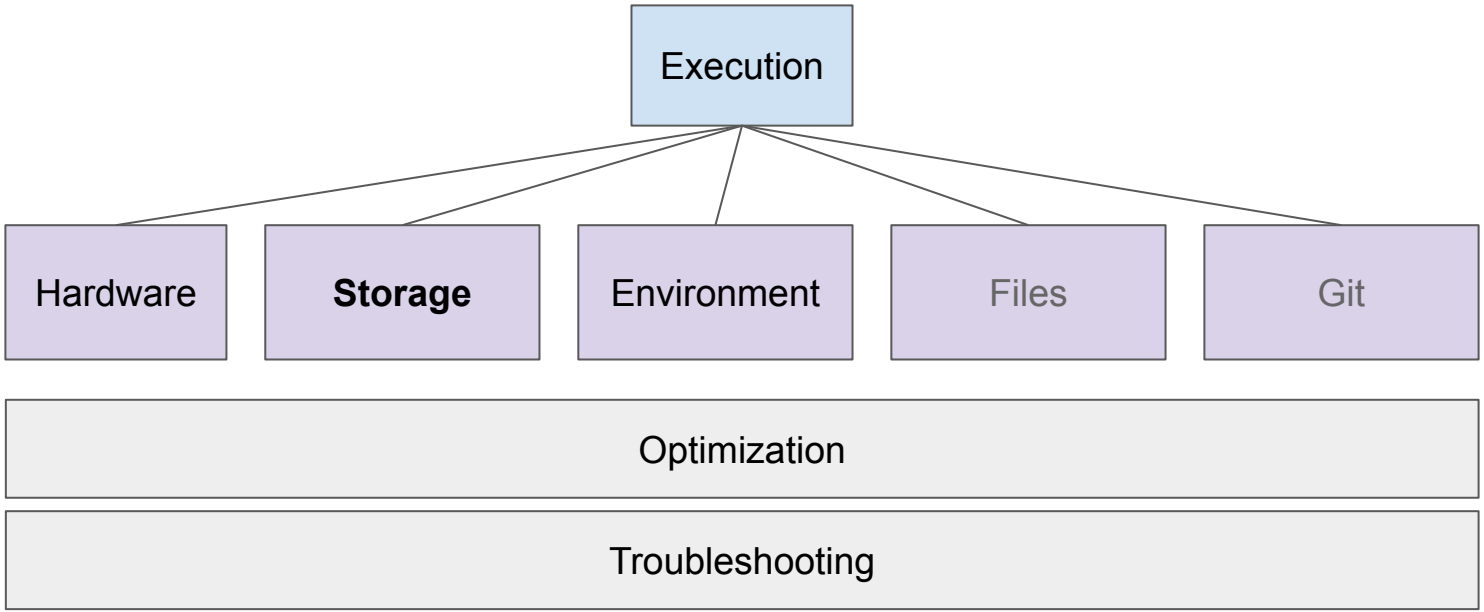


RESOURCE QUOTAS

Lessons on creating Resource Quotas and other configs for Model APIs:
https://learn.dominodatalab.com/admin-training/894358



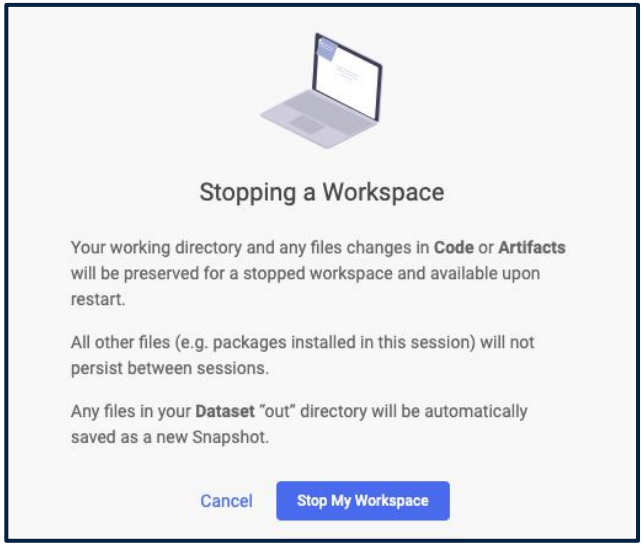
COMPONENTS REQUIRED FOR AN EXECUTION



Persistent Volume Management - Workspaces

WORKSPACES

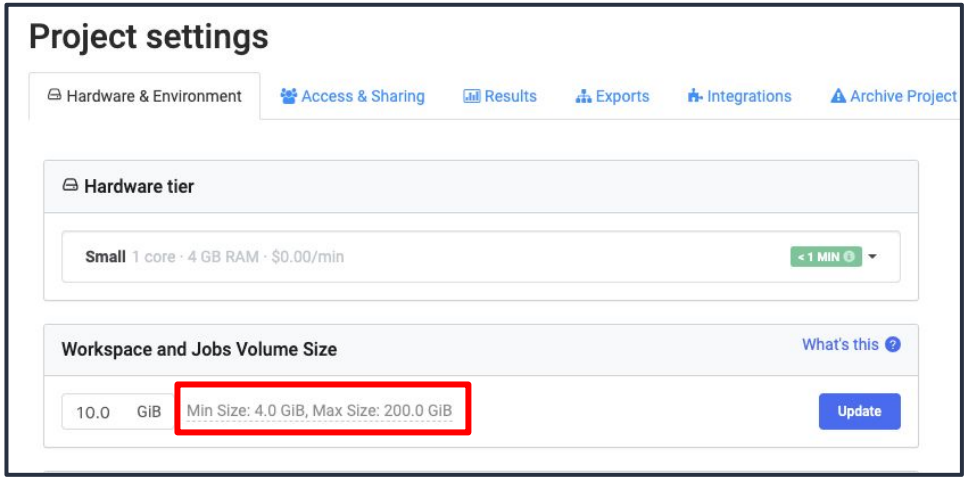
- You can stop and start workspaces without committing files
- Everything in the directory mapped to the *Files (*or *Code* and *Artifacts)* folders will persist to next session
- Objects in memory and files outside of the */mnt* directory (including pip installed packages) will not persist
- See docs here for user info



DOMINO

VOLUME SIZE

- Users can control the amount of storage available to workspaces and jobs in their projects
- This is updated in project settings
- Default limits are shown - this is configurable as an Admin



DOMINO

WORKSPACE LIMITS

Even *Stopped* Workspaces use storage space

Default limits are in place for cost management:

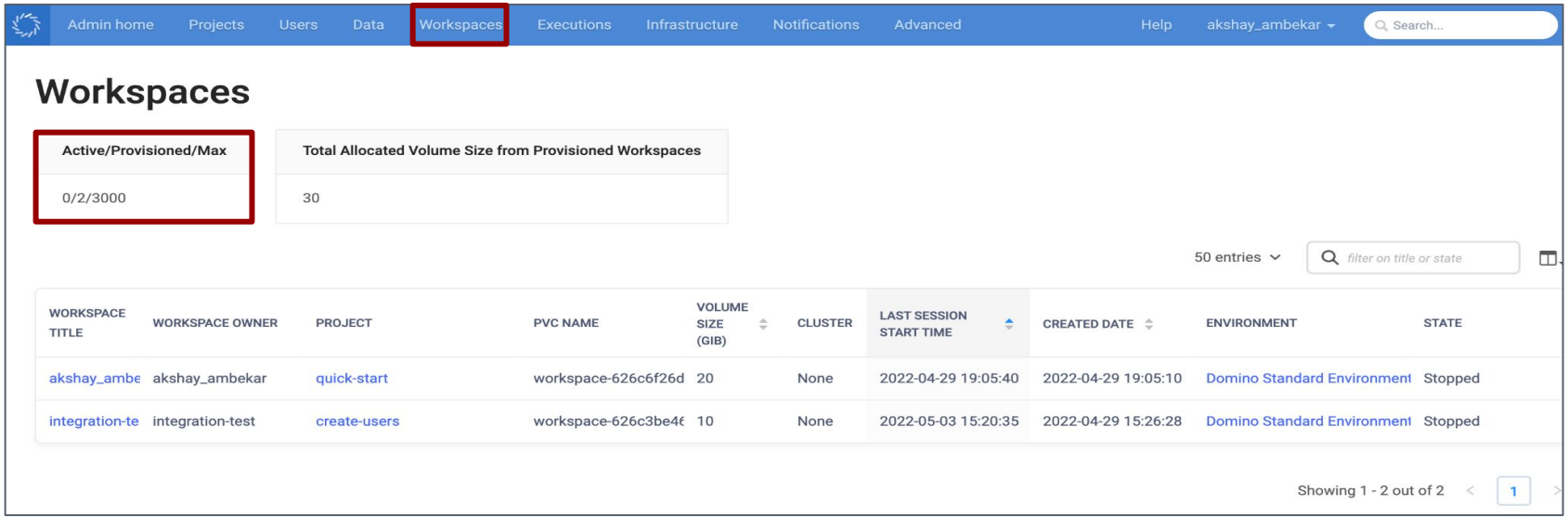
- Four workspaces per *user* per *project*
- Sixteen workspaces *total* per user
- 3000 workspaces in *deployment*

Deleting a workspace stops it permanently and removes it from limit count

Admin lesson - https://learn.dominodatalab.com/admin-training/880015

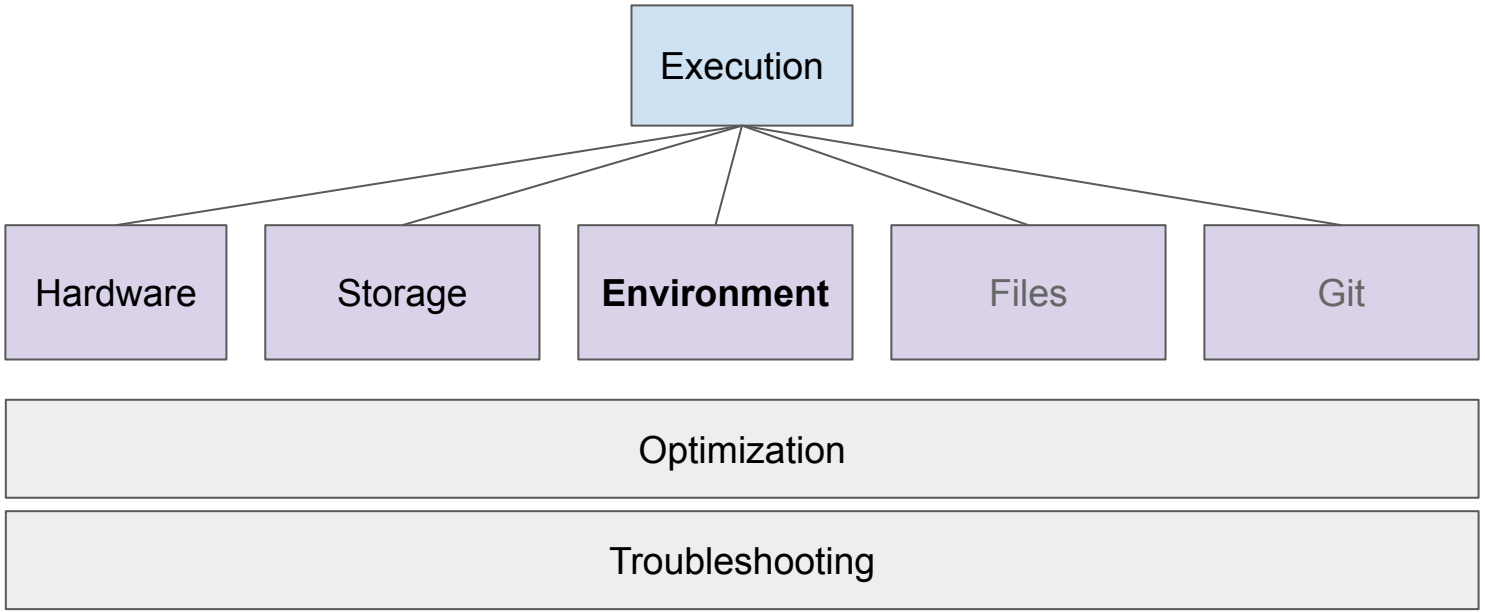
WORKSPACES ADMIN PAGE

View all stopped and running workspaces and see total volume size in the Admin Workspaces page



DOMINO

COMPONENTS REQUIRED FOR AN EXECUTION



Managing Compute Environments

Environments are Dockerfile, scripts, and configurations ***together***.

Users have root access inside the Docker container.

Data scientists will probably use them a lot.

https://docs.dominodatalab.com/en/5.1/reference/environments/Environment_management.html

MAINTAINING ENVIRONMENTS

Admins are responsible for maintaining global environments

Note that in Domino 5.1, most custom images can be automatically adapted for use in Workspaces and Jobs

Lessons on customizing environments and best practices can be found here:

https://learn.dominodatalab.com/admin-training/879956

On Demand Distributed Computing

ENVIRONMENT SETUP

On-demand Distributed Computing in Domino requires two separate environments:

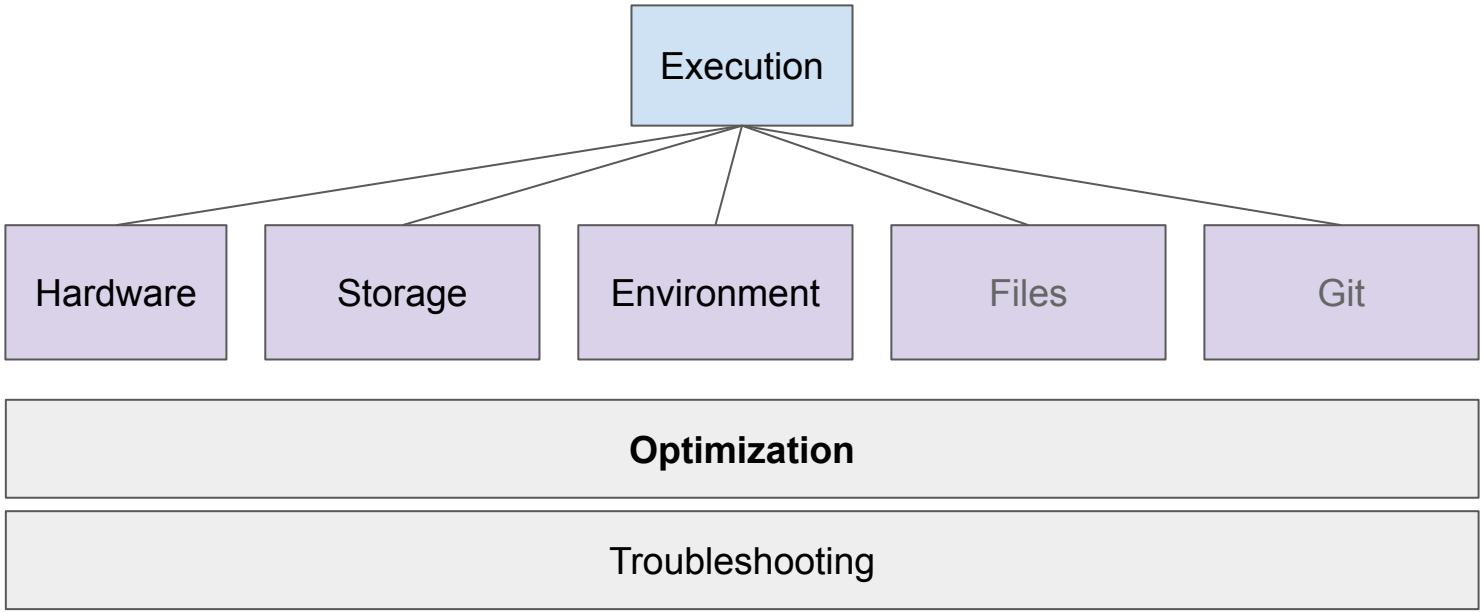
1. Base environment for the workspace/job
2. Cluster environment

Admin lesson here - https://learn.dominodatalab.com/admin-training/880011

User training is found in the 201 class -

https://learn.dominodatalab.com/domino-201-advanced-features-46/936218

COMPONENTS REQUIRED FOR AN EXECUTION



Optimizing Start Times for Jobs and Workspaces

SUMMARY OF STEPS TO OPTIMIZE START TIMES

1. Start a new node if there isn't one with requested resources
2. Download the Domino images (executor, NGINX, Prometheus, fluentd, etc.)
3. Download the Environment image
4. Copy the user's project files into a persistent volume
5. Start the executor process
6. Execute any environment setup scripts

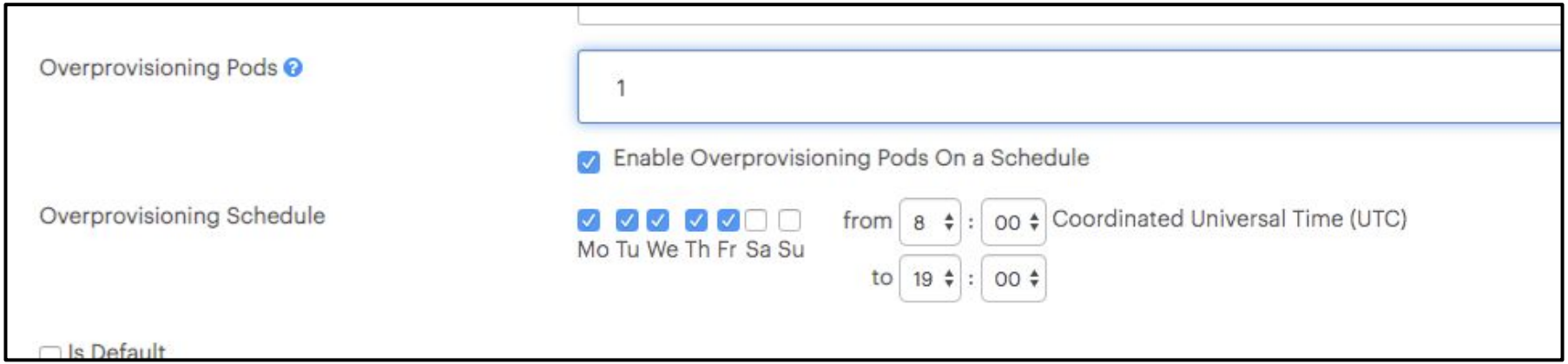
Overprovisioning pods

Cache Docker images on AMI

Domino Datasets

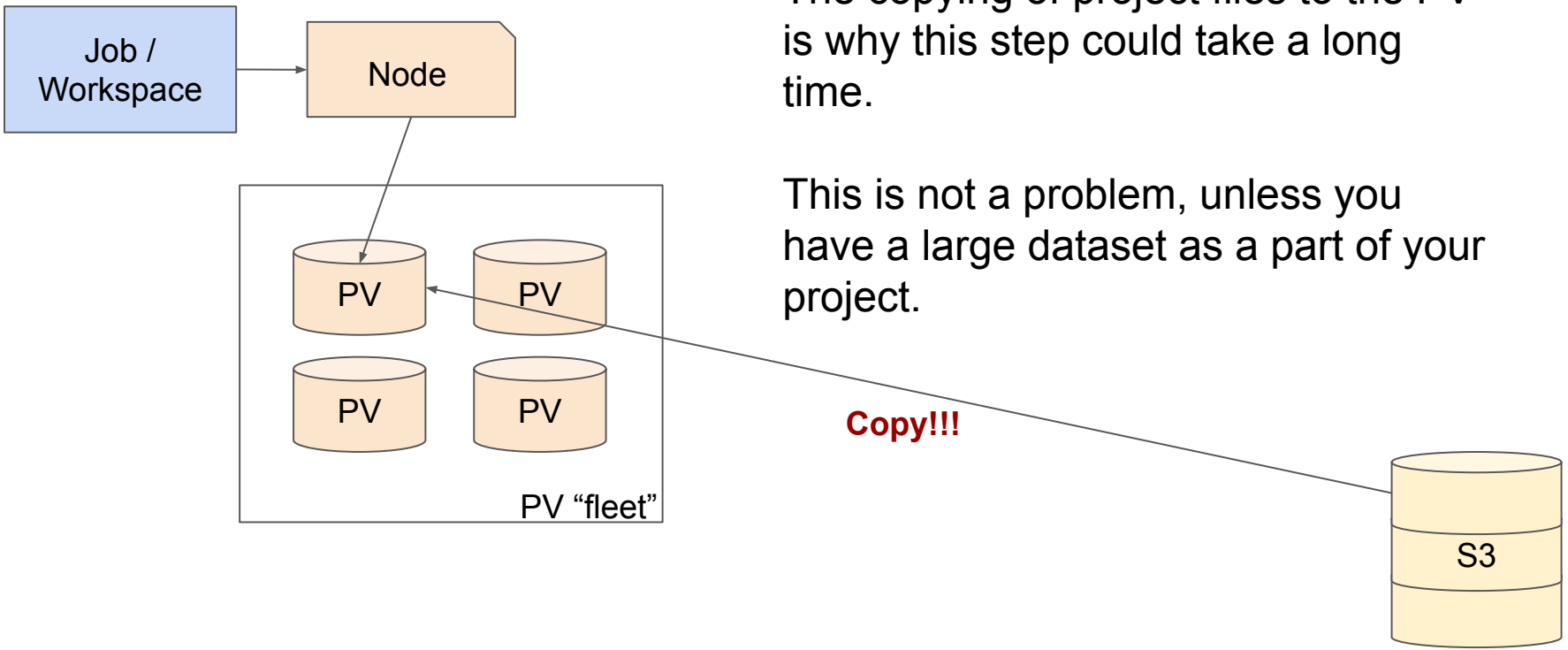
Admin lesson -
https://learn.dominodatalab.com/admin-training/880096

OVERPROVISIONING IS SET IN THE HARDWARE TIER DEFINITION



Overprovision pods on a schedule to prepare for known times when there is a sharp increase of usage.

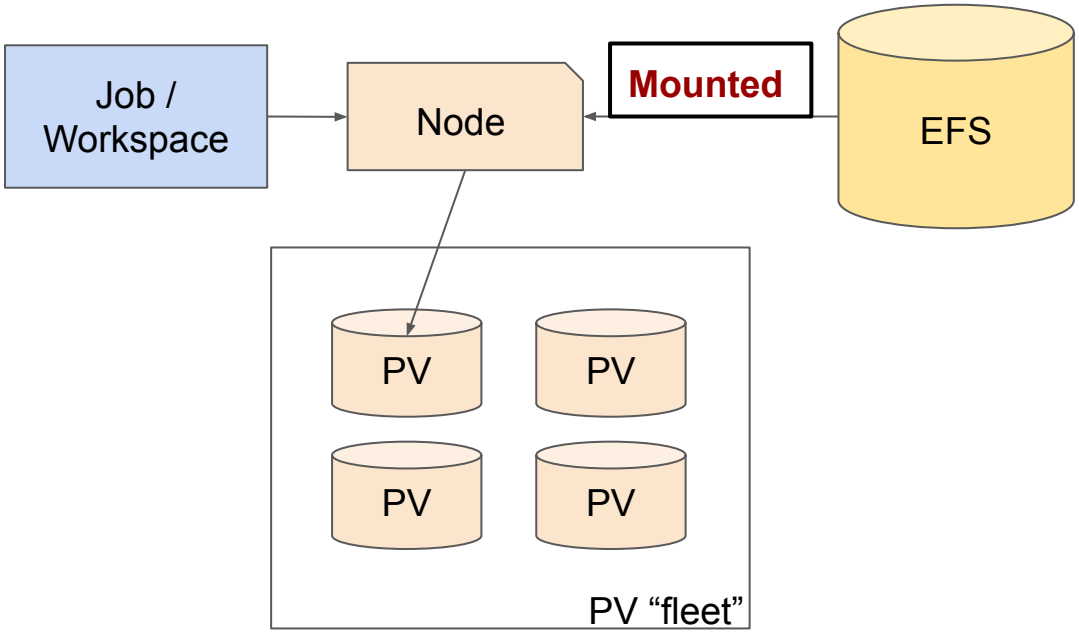
STORAGE WORKFLOW



The copying of project files to the PV is why this step could take a long time.

This is not a problem, unless you have a large dataset as a part of your project.

STORAGE WORKFLOW WITH DOMINO DATASETS



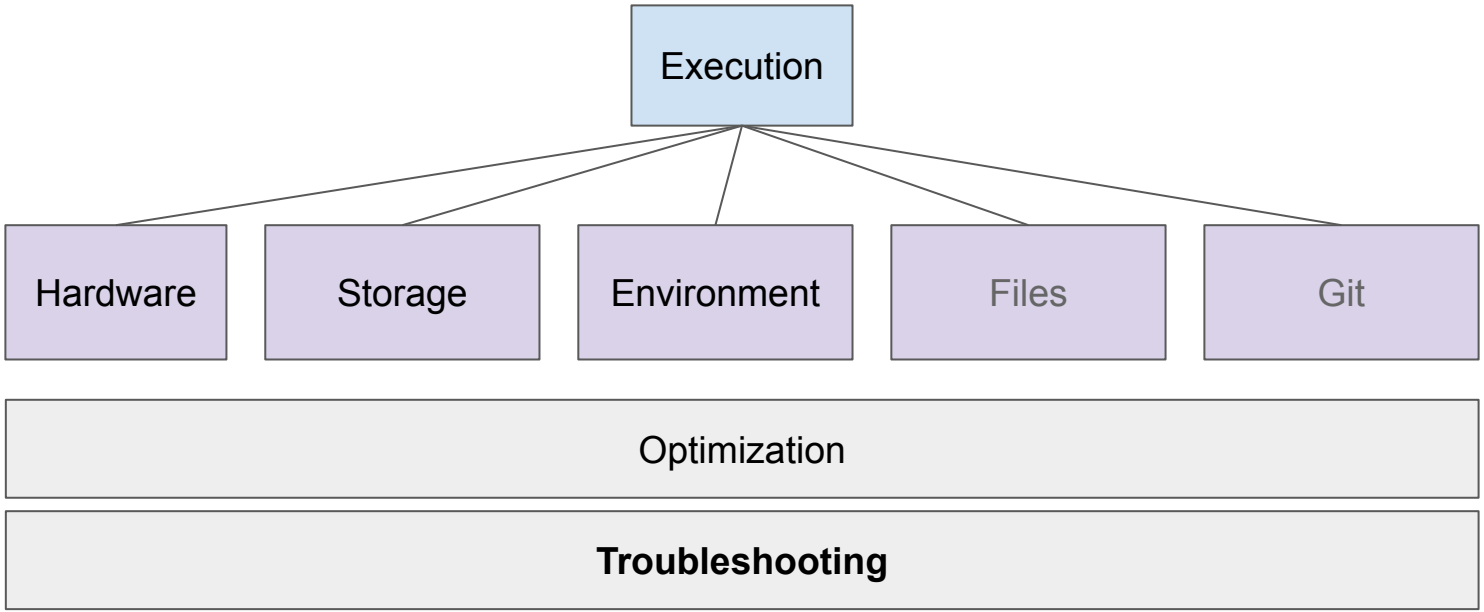
Domino Datasets use EFS/NFS and are mounted directly into execution pods.

Datasets become available to the Domino execution immediately.

https://admin.dominodatalab.com/en/latest/data-management/Data_in_Domino.html#about-domino-datasets

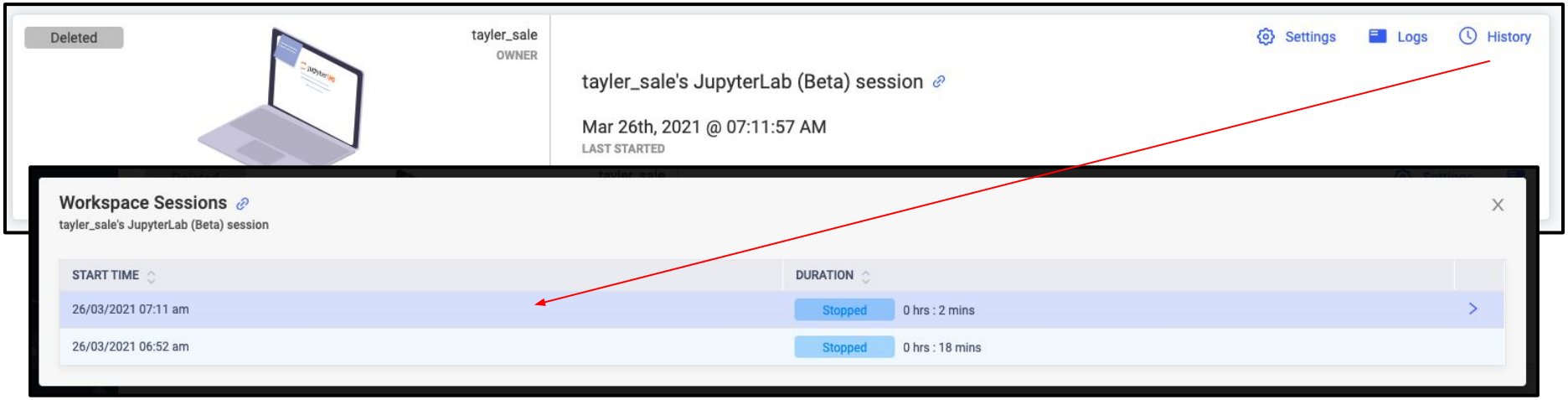
https://docs.dominodatalab.com/en/5.1/reference/data/data_in_domino/datasets.html

COMPONENTS REQUIRED FOR AN EXECUTION



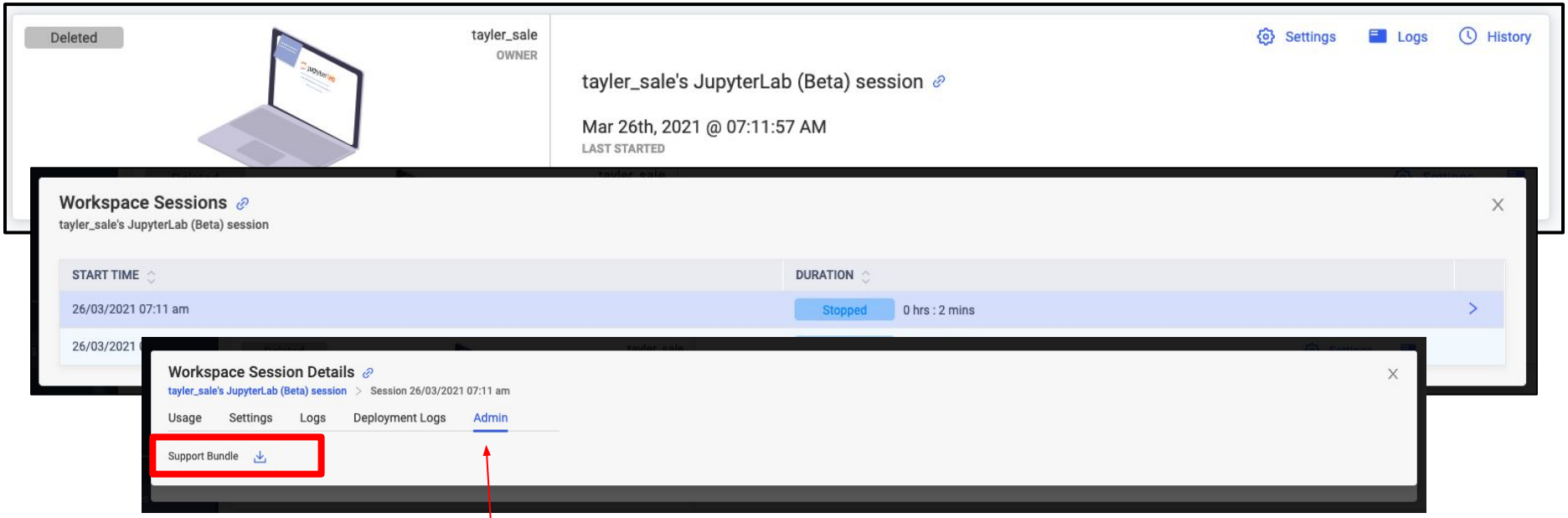
Troubleshooting the Compute Grid: Getting Information

THE SUPPORT BUNDLE IS YOUR BEST FRIEND



Admin lesson here - https://learn.dominodatalab.com/admin-training/893101

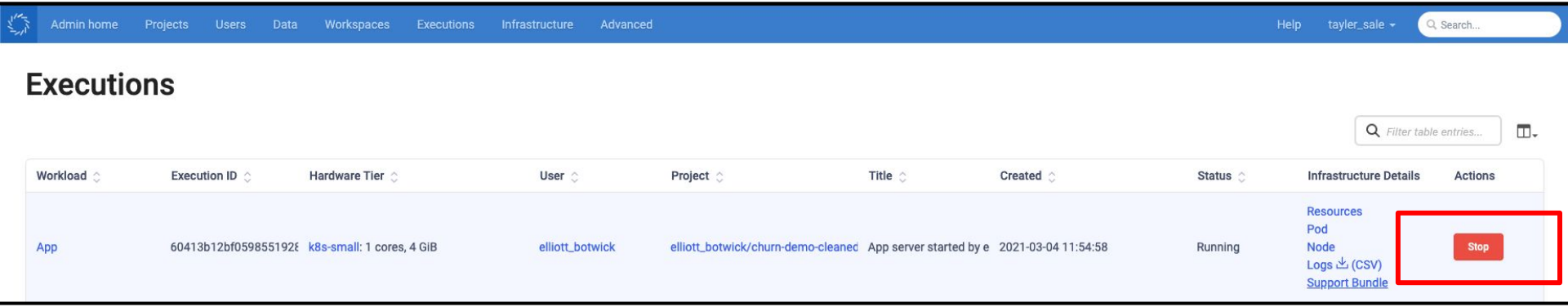
THE SUPPORT BUNDLE IS YOUR BEST FRIEND



CONTENTS OF THE SUPPORT BUNDLE

events.csv	Kubernetes event history in csv form
events.json	Kubernetes event history in json form
execution.log	Output of the execution logs that the user normally sees
executor.log	Logs for the executor container
kubernetes.yaml	Kubernetes YAML describing the execution pod that was created
manifest.csv	List of files in the support bundle with sha1
nginx.log	Logs for the nginx container
nucleus-dispatcher.logs	Dispatcher logs
nucleus-frontend.logs	Frontend logs

STOPPING EXECUTIONS AS A ADMIN



If an execution becomes stuck, an admin can use the Executions page to stop the execution. It may take some time, but this is the safest and quickest way of stopping.

Do not try to delete the execution pod.

Domino Model Monitoring

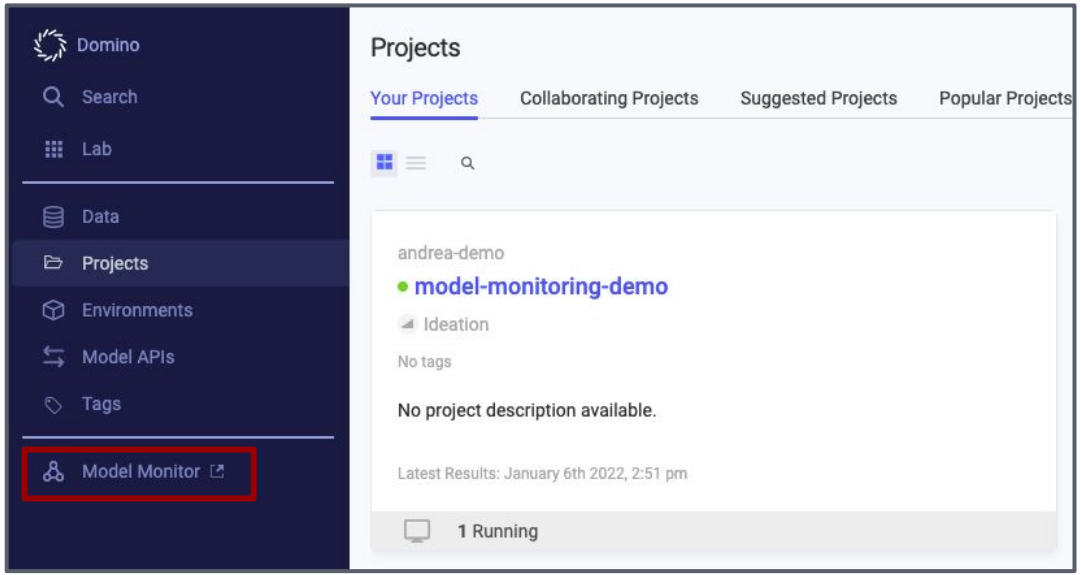
INTEGRATED MODEL MONITORING

Platform namespace

- 1 nodes (in addition to the 3 for Domino)
- Handles user interfaces, API server, scheduling, reporting, and other services

Compute namespace

- Model monitoring jobs that are scheduled on cluster



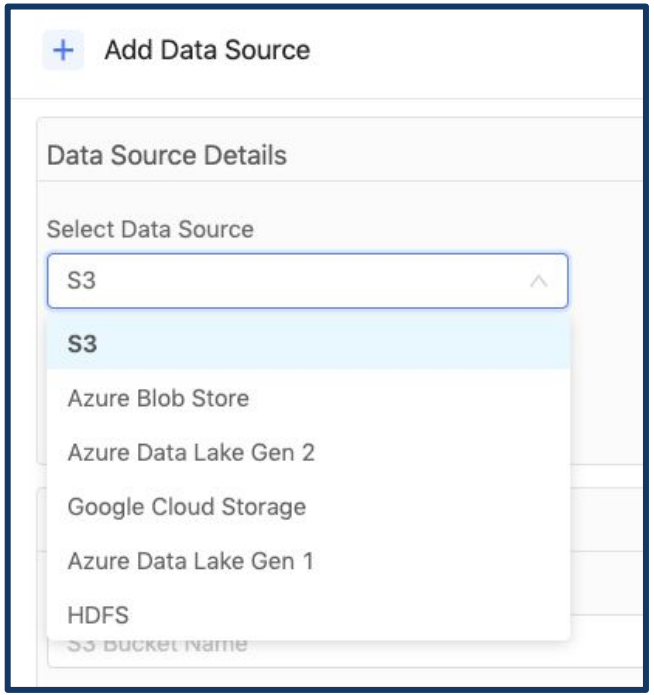
DOMINO

CONFIGURING DATA SOURCES

DMM can be used with the following data sources:

- Amazon S3
- Azure Blob
- Azure Data Lake Gen 1
- Azure Data Lake Gen 2
- Google Cloud Storage
- HDFS

See docs for specifics on using each data type



DOMINO

User Management and Authentication

KEYCLOAK IS AN OPEN SOURCE IDENTITY AND ACCESS MANAGEMENT SOLUTION

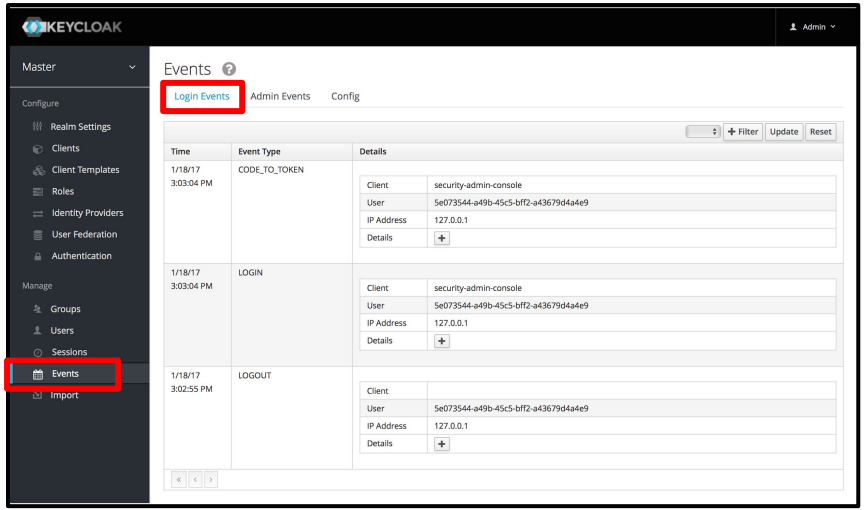
Additionally:

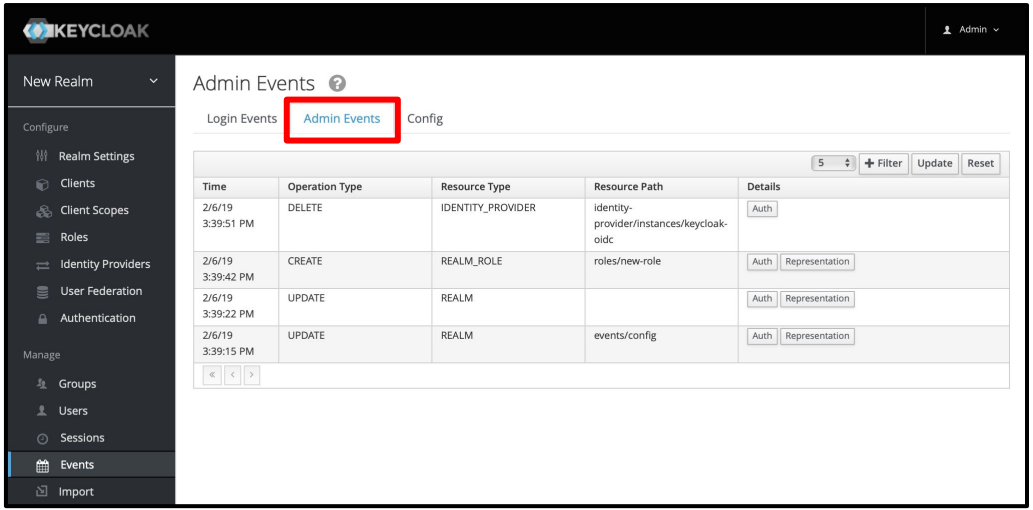
- Enables LDAP and SSO integration with Domino
- Enables propagation of AWS temporary credentials for accessing external systems into Domino run environments
- Supports SAML or SSO role synchronization with Domino
- Supports synchronization of SAML or SSO groups to Domino organizations

Admin lessons here - https://learn.dominodatalab.com/admin-training/900302

ACCESSING THE AUDIT TRAIL

It is often required to know who logged into the system, either Domino or Keycloak, and who made changes.





ORGANIZATIONS IN DOMINO

Domino lets users create Organizations which can be used to:

- Permission projects to many users at once and easily add/remove collaborators from multiple projects
- Assign *owner* level permissions to a group of users
- Share compute environments
- Restrict hardware tiers to specific groups (controlled by Admins)

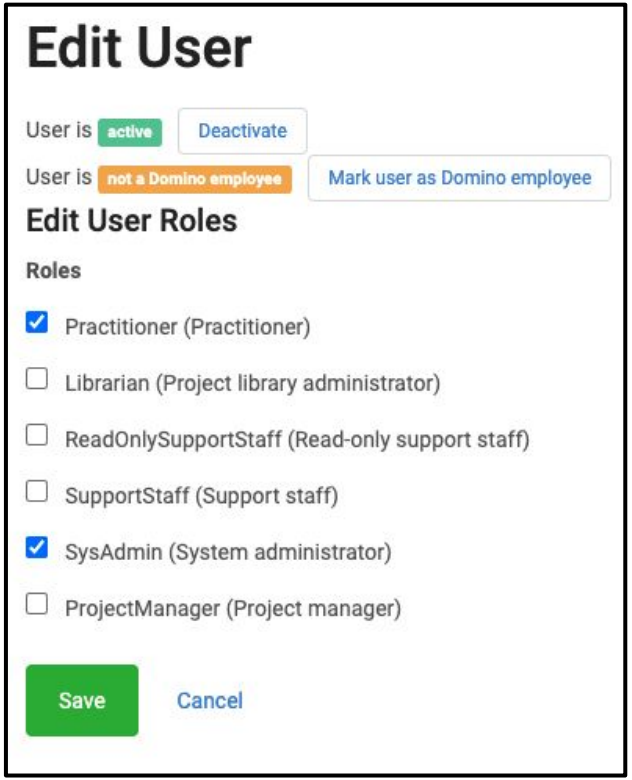
Lesson here - https://learn.dominodatalab.com/admin-training/899436

SYSTEM ROLES ARE MANAGED BY ADMINISTRATORS

There are various levels of system level roles

SysAdmins can manually edit the roles of other users to give elevated system privileges

Lesson here - https://learn.dominodatalab.com/admin-training/900255



LICENSE MANAGEMENT

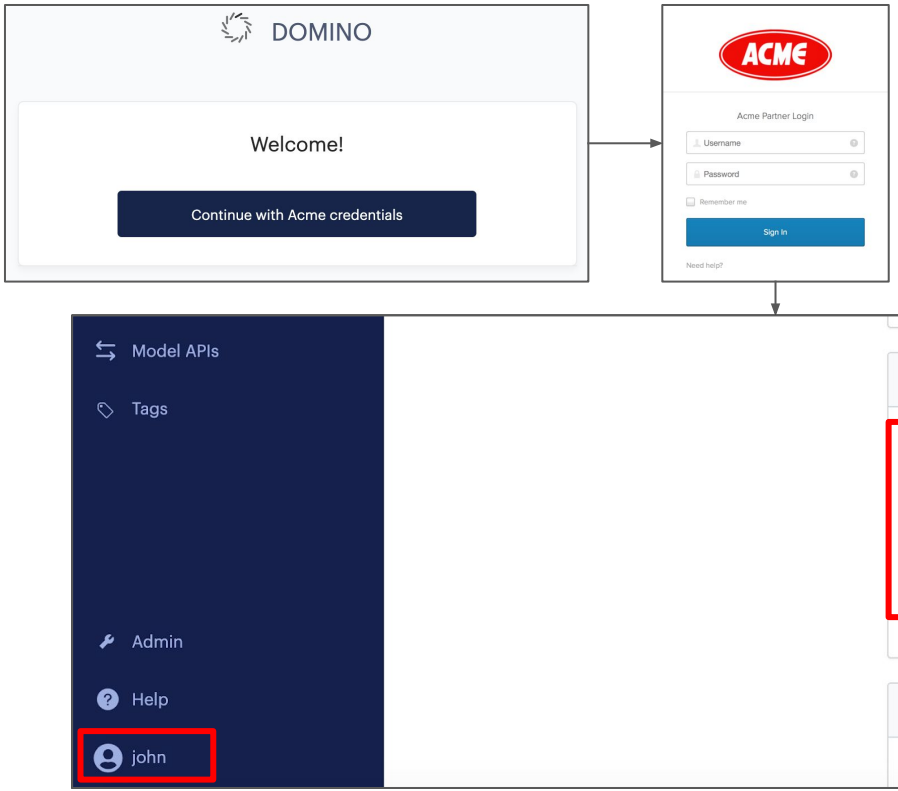
There are now global user roles (in addition to Admin-level roles)

- A *Lite User* role (the Practitioner box is *unchecked*)
    - These users can be given access to Launchers and published Apps but will be unable to start a workspace, create projects, etc
- A *Practitioner* role
    - Has the same capabilities as a *User* today

By default, a new user will be a *Practitioner*

- This can be changed by Admins to default to *Lite Users*

ROLE AND ORGANIZATION SYNCHRONIZATION

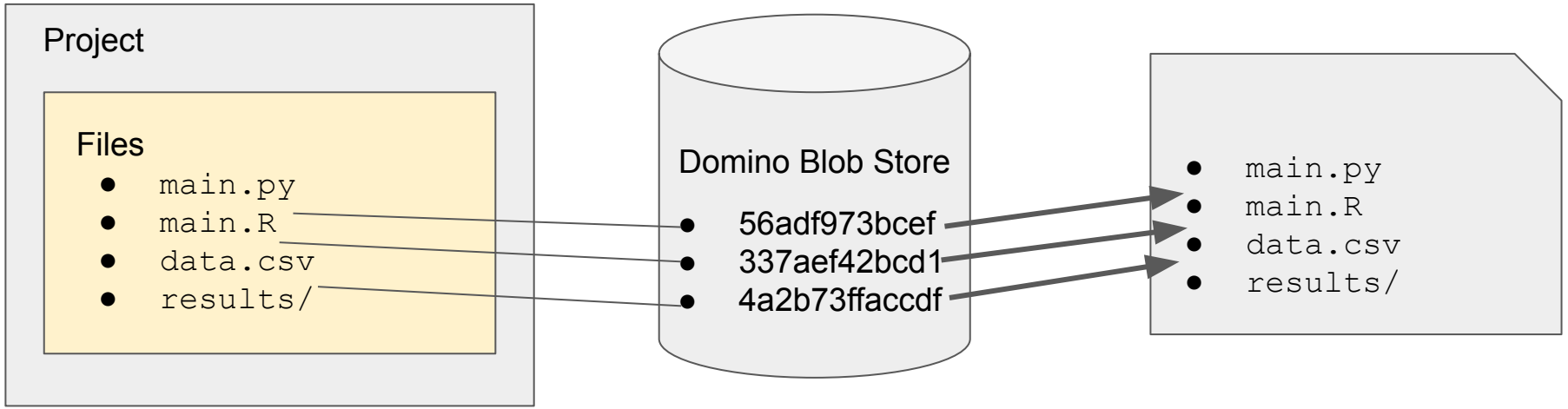


Organization membership and system level roles can be managed from an IdP, *instead of Domino*

Upon first login, the user will be in the appropriate organizations, have access to their organizations' projects, and have their proper system level role.

Data Management

DOMINO FILE SYSTEM



- All *files* in Domino projects are stored in the Domino blob store.
- Domino copies those files into the run pod when user starts an execution.
- When files are created or modified, new copies are uploaded to the Domino blob store.
- *All blobs are kept forever.*

FULL DELETE

- All blobs in Domino are kept forever by default.
- In case a deletion of a file is required to comply to law, admins have the ability to full delete all revisions of the file from the blob store.
- Extreme caution should be exercised!!!
  - Domino only stores one unique copy of the contents of a file.
  - If two projects share the exact same file contents and the blob is deleted from one project, both projects will be affected!

Admin lessons here - https://learn.dominodatalab.com/admin-training/899259

ACTIVITY: FULL DELETE A FILE

1. Sign into the temporary training Domino instance as an Admin: https://admin-rev.workshop.domino.tech/
2. **Delete** the model.R file from the quick-start project
3. Now, restore the model.R file
4. Now, **full delete** the model.R file

DATASETS VS PROJECT FILES

Project Files:

- Files are synced whenever a code session is launched or stopped
- Large or many files can slow down these launch/sync times
- Limits based on project volume size
- Can only import data by importing entire project
- Automatically versioned

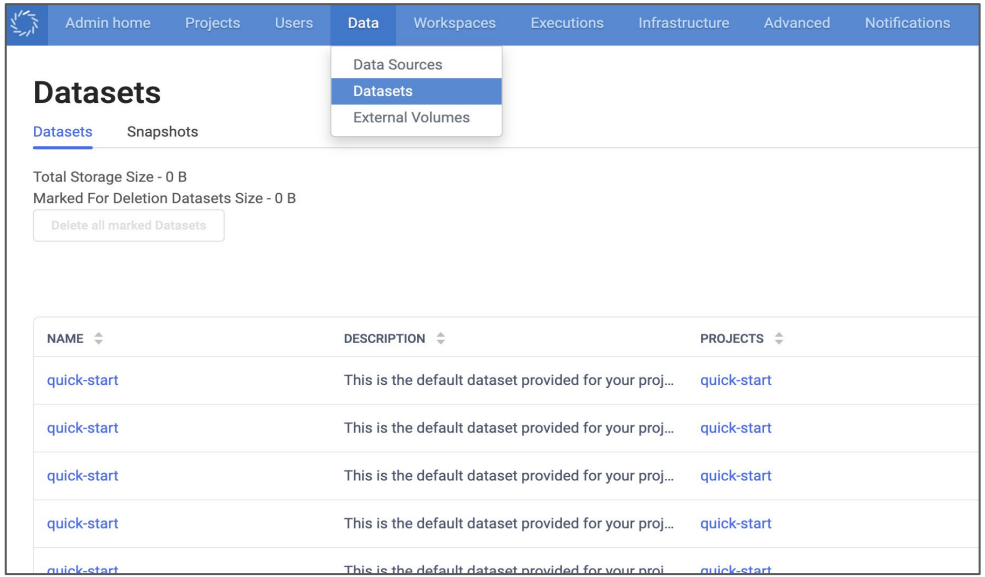
Datasets:

- Stored in NFS/EFS
- Mounted to the machine for use during executions, no copying
- No copying means no additional wait for launching and syncing
- Large file size and large numbers of files allowed
- Manually versioned

DOMINO

DATASET ADMINISTRATION RESPONSIBILITIES

1. Periodically check the Datasets administration interface
2. Monitor and track storage consumption
3. Set limits on usage per-Dataset
4. Handle deletion of Dataset snapshots



Admin lessons here - https://learn.dominodatalab.com/admin-training/880025

ACTIVITY: DELETE A MARKED SNAPSHOT

1. Sign into the temporary training Domino instance as an Admin: https://admin-rev.workshop.domino.tech/
2. Delete the snapshot-1 that's marked for deletion

EXTERNAL DATA VOLUME OVERVIEW

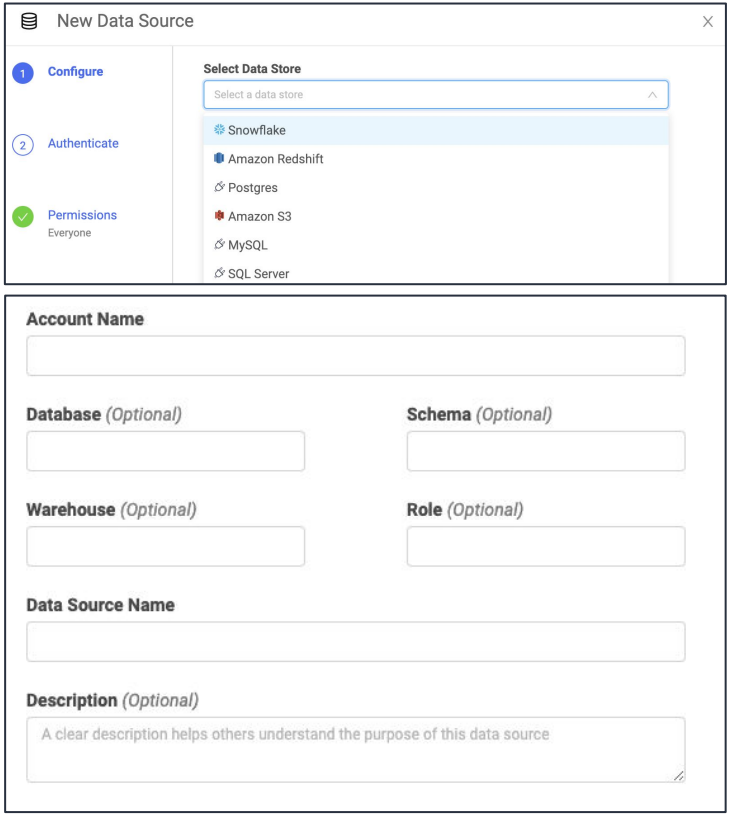
- Domino projects can be setup to access external data volumes
- This can only be done by Admins
- Requires access to the Kubernetes cluster supporting Domino and the Admin page
- Access is controlled separately from projects - can be limited to specific users or organizations

Admin lesson here - https://learn.dominodatalab.com/admin-training/880027

DOMINO DATA SOURCES

Domino data sources include connectors for:

- Snowflake
- Amazon S3 and Redshift
- MySQL
- Microsoft SQL Server (MSSQL)
- PostgreSQL
- Oracle Database
- Azure Data Lake Storage
- Google Cloud Storage



https://learn.dominodatalab.com/admin-training/1196453

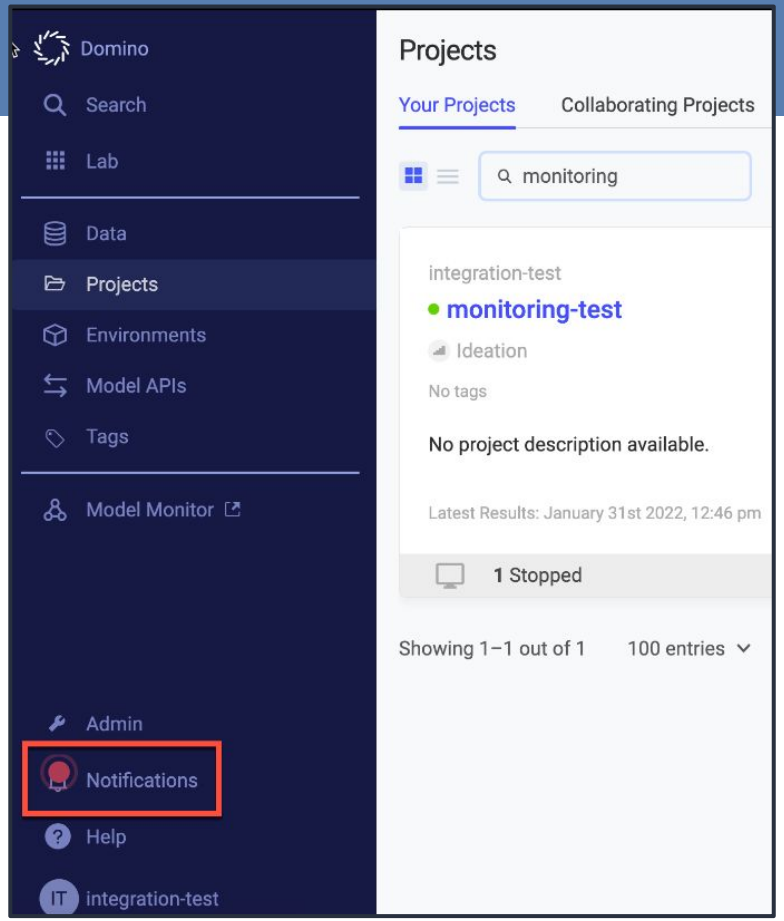
DOMINO

Monitoring and Security

DOMINO NOTIFICATIONS

Admins can create notifications that will appear in the Domino navigation bar for users. The indicator color matches criticality:

- Default
- Critical



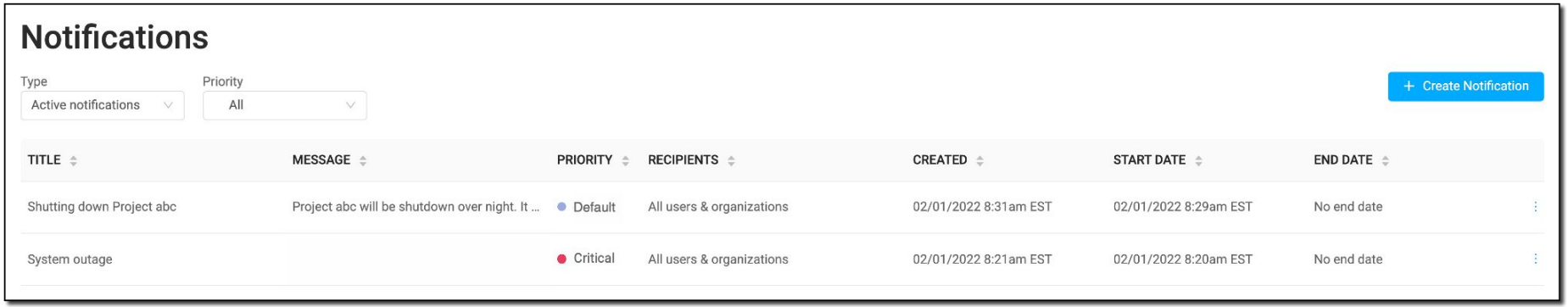
DOMINO NOTIFICATIONS

Create and manage notifications in the Admin Panel > Notifications page

Choose recipients (users or organizations), priority, start and (optional) end times

Can be created via the Domino APIs

Limits set via a central config

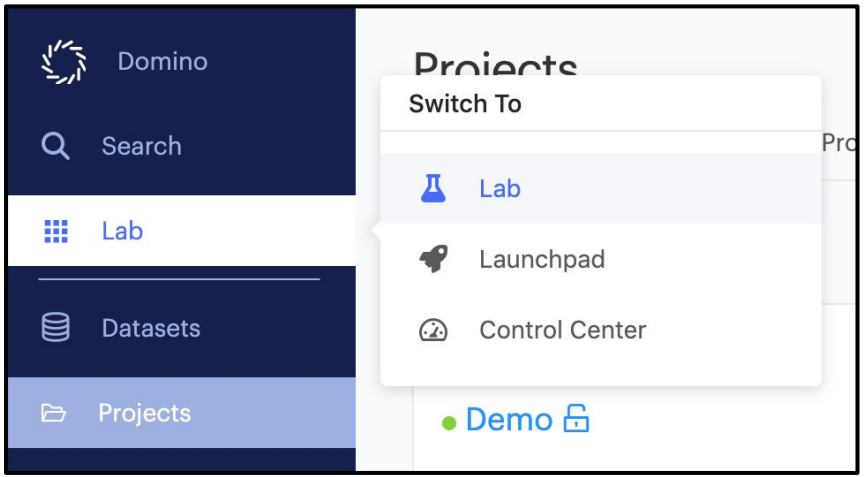


CONTROL CENTER

Data science leaders and administrators can navigate to the Control Center.

The Control Center consists of

1. Compute and Spend [admins only]
2. Project Portfolio
3. Assets



Admin lessons here - https://learn.dominodatalab.com/admin-training/900544

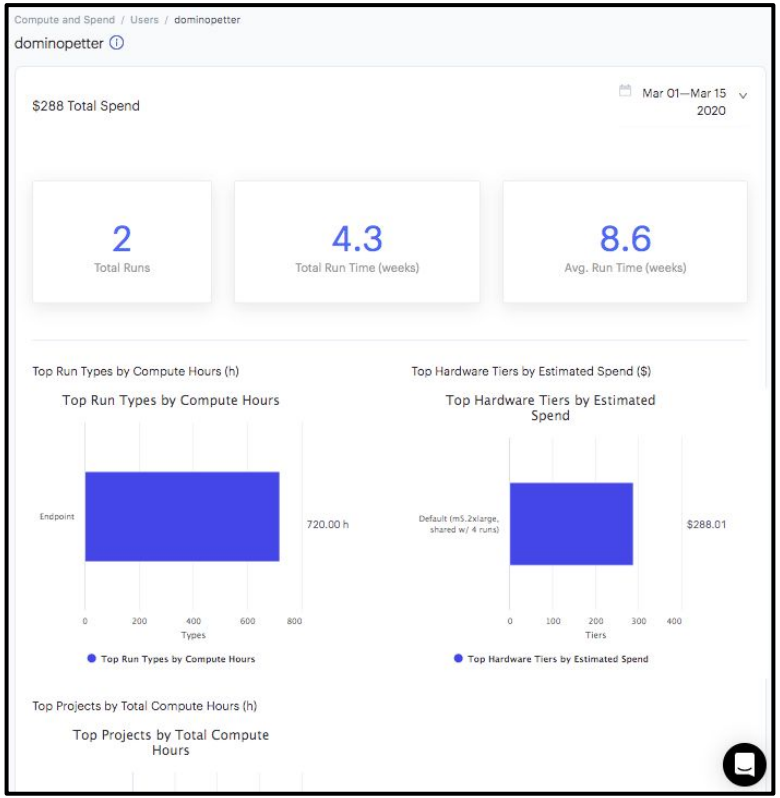
COMPUTE AND SPEND

Dashboard of overall spending on the compute grid.

Enables interactive investigation of compute costs by user.

Only administrators can access the Compute and Spend section.

This can also be [exported with an API]

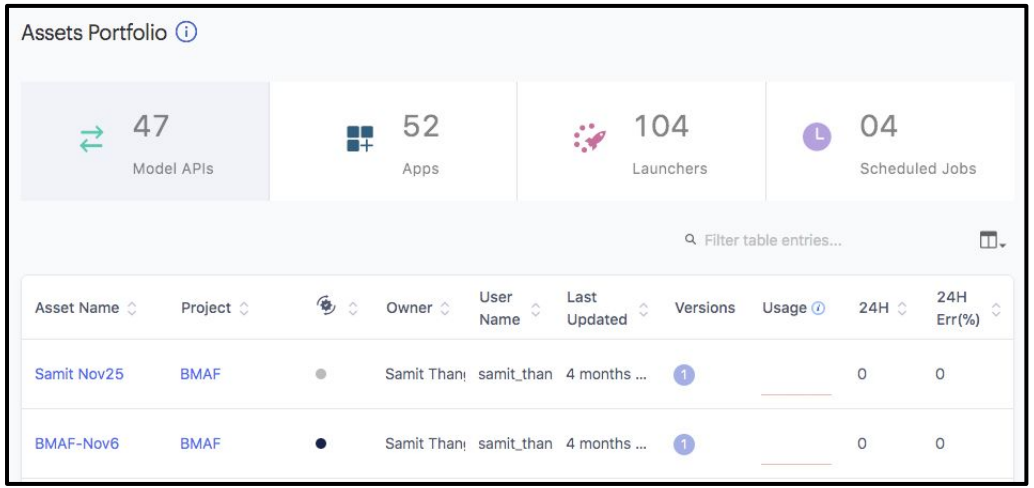


ASSETS

The Assets view will allow users to see all assets that they have access to.

Admins will see all assets.

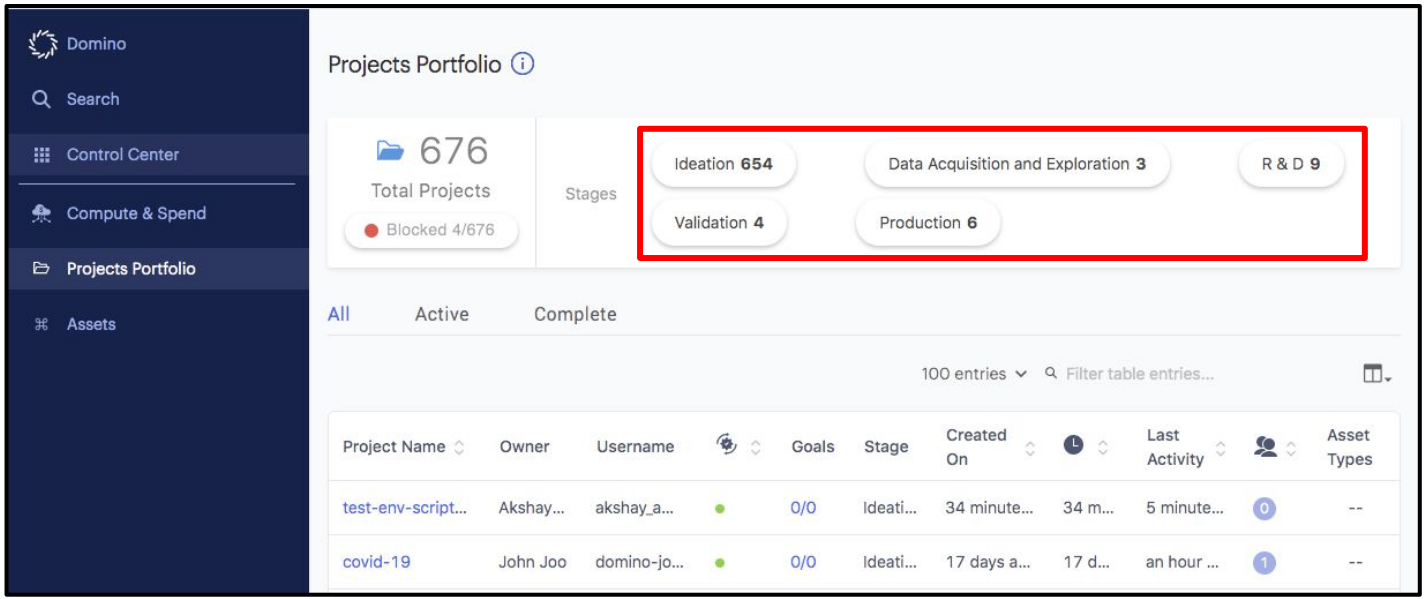
Project managers will see all assets owned by users in their Organizations.



PROJECT PORTFOLIO

This interface allows you to quickly digest the state of work in your projects.

A project's status is set within the project.



SECURITY OVERVIEW

There are many ways to control security in your Domino, including:

- Limiting user actions
- Setting up 2FA and other protocols
- Configuring content security policies
- Disabling public projects
- Auditing users

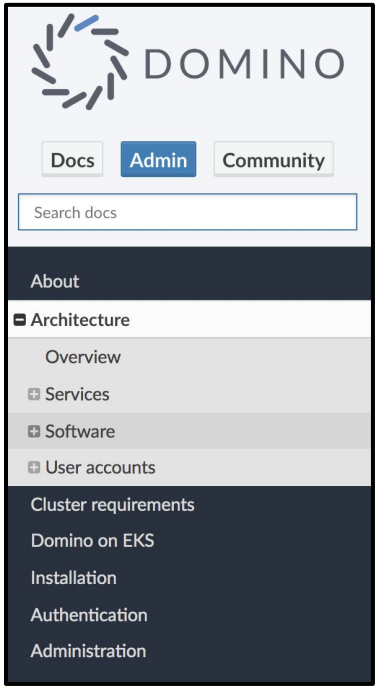
Admin lesson here for more details - https://learn.dominodatalab.com/admin-training/900805

SETUP CHECKLIST AND BEST PRACTICES

See the checklist here for questions to ask your data science team and initial setup tasks - https://learn.dominodatalab.com/admin-training/894574

This lesson covers best practices for setting up new users - https://learn.dominodatalab.com/admin-training/917575

TECH SUPPORT > GETTING HELP



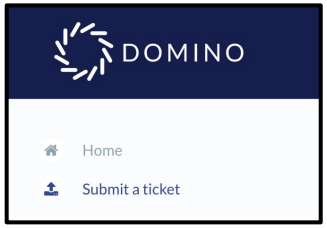
Admin Docs

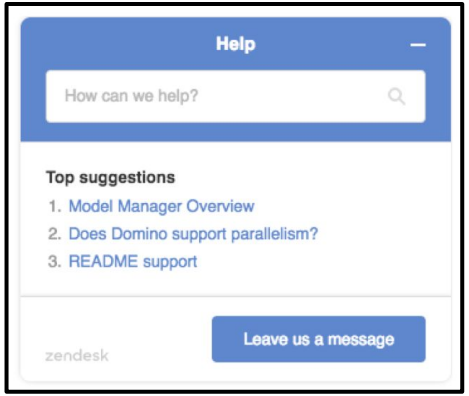
Use the search bar if you can't find what you are looking for.
https://admin.dominodatalab.com/

The Support Team

For more complex issues, you can reach out to our Support team in one of the following ways:

- Submit a request from the Help Center directly
- Leave a message on the help widget
- Email support@dominodatalab.com





Admin Course Summary

Great Job! You've completed this course!
The main learning points from this course are:

Domino Architecture, Components, and Configuration.

Managing the Domino Environment & Infrastructure.

How to Manage Domino Storage.

Optimization and Troubleshooting.

Domino Monitoring, Data Management, and Security.

Thank you!