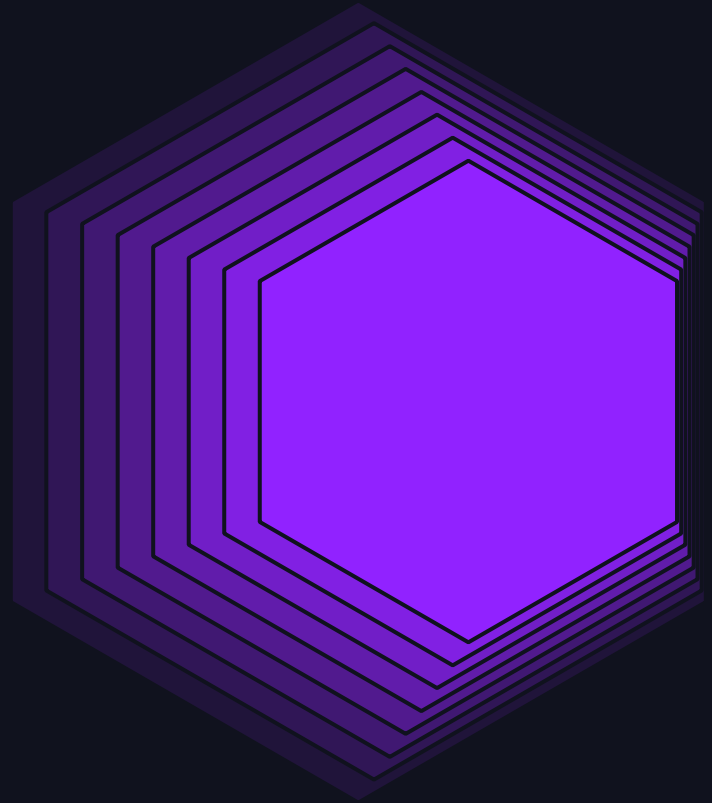


SciCLOps: Databricks Quick Start, Powered by DABs



Tara Enright, Data Engineer at 84.51°
Connor Brown, Cloud Engineer at 84.51°

Introductions

Tara Enright

- 5 years experience in ML model development, Data Science, ML/Data Engineering
- 1+ years MLOps and DevOps

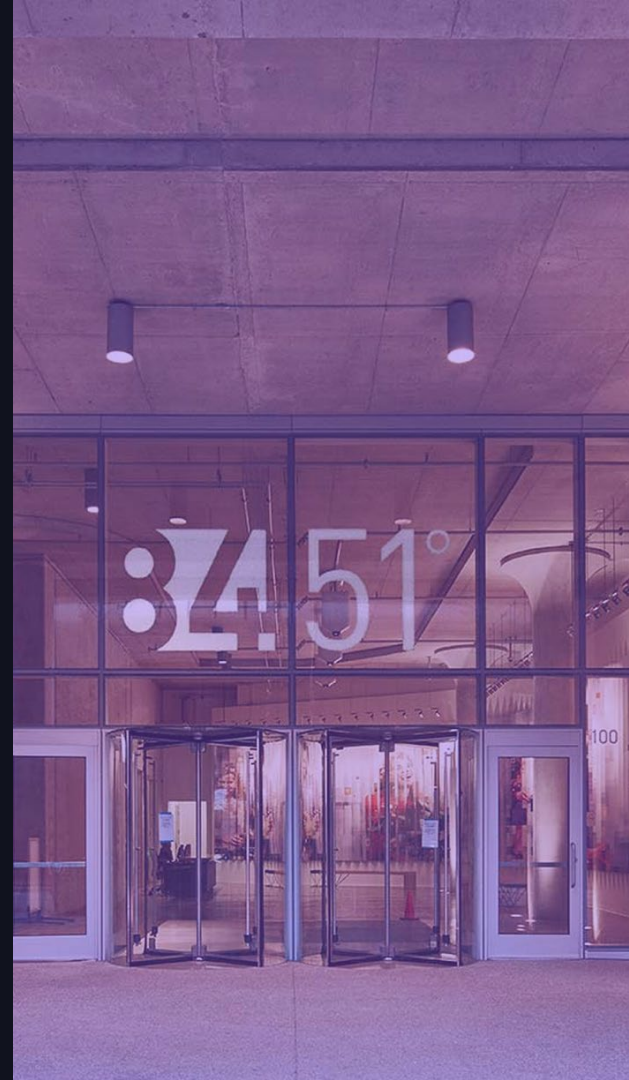
Connor Brown

- 7+ years experience in DevOps/Security/Platform Engineering
- Spent a year working on the MLOps team on Platform Engineering
- Recently moved to the AI Acceleration team

About 84.51°

“We are a **retail data science, insights and media company**. We help The Kroger Co., consumer packaged goods companies, agencies, publishers and affiliates **create more personalized and valuable experiences** for shoppers across the path to purchase.

Powered by cutting-edge science, we utilize first-party retail data from over 62 million U.S. households sourced by the Kroger Plus loyalty program to fuel a more customer-centric journey using 84.51° Insights, 84.51° Loyalty Marketing and our retail media advertising solution, Kroger Precision Marketing.”



Here's What we Hope you Learn!

What should I take away to apply at my company?

- How do we enable many Data Scientists/Data Engineers/ML Engineers in Databricks with limited DevOps resources?
- What does a CI/CD workflow for Databricks resources look like?
 - Hint: Check out Databricks Asset Bundles!
- How can I take advantage of new Databricks features to improve my workflows?

DISCLAIMER

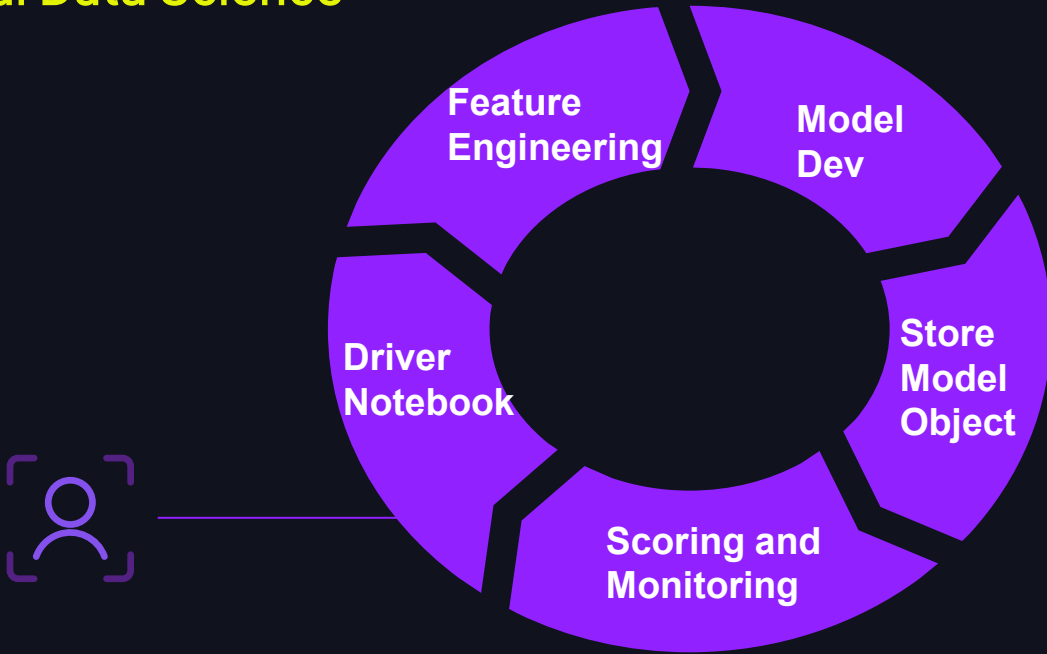
Our demo is focused on ML... but anything deployable to Databricks is supported by SciCLOps

- MLOps Patterns
- DLT Pipelines
- Databricks Job Workflows
- Standardized Reporting Jobs/Dashboard Jobs
- Future patterns at 84.51°

ML Lifecycle

Machine Learning Lifecycle

Traditional Data Science

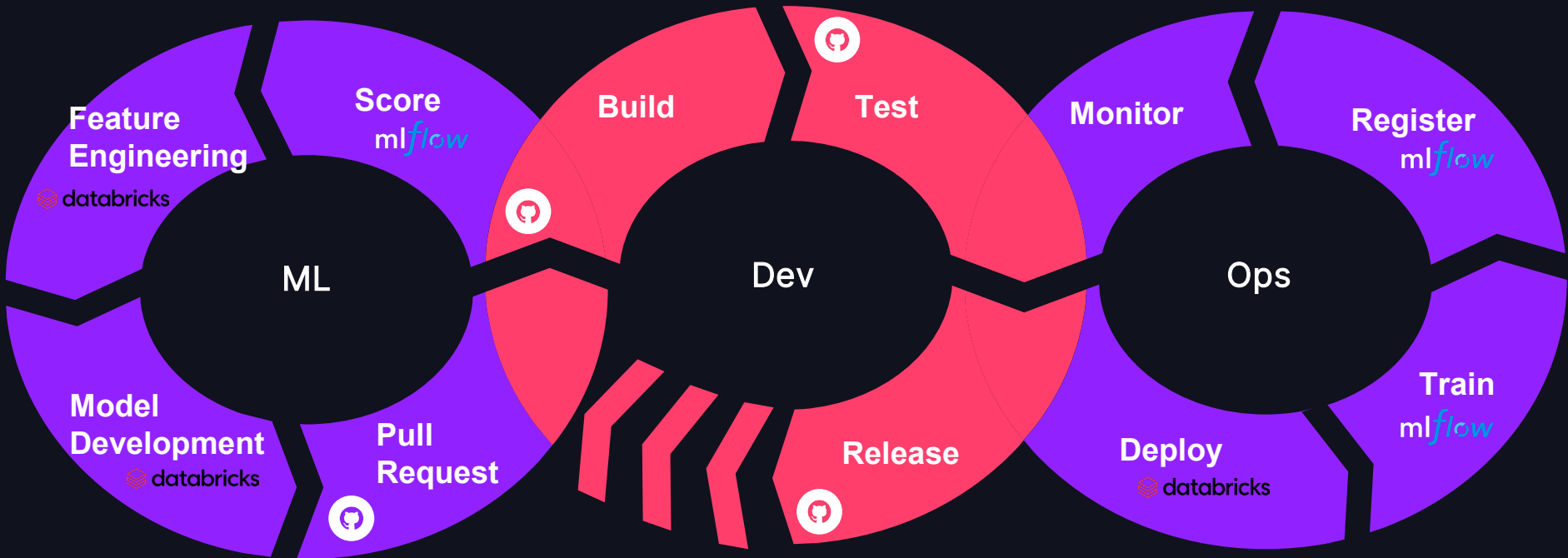


Engineering effort: Estimated 2-3 weeks to production



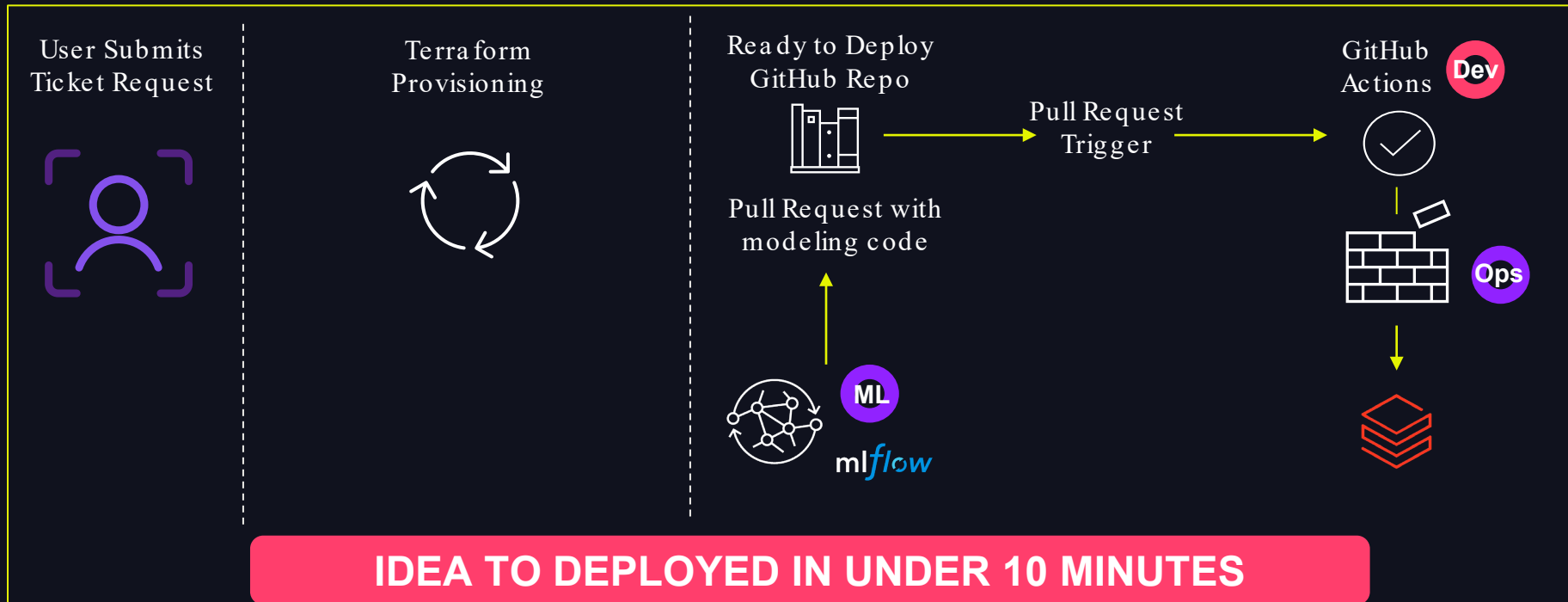
Machine Learning Lifecycle

DevOps Style



SciCLOps

Science Creation Lite Operations: A curated out-of-box DevOps reusable templated workflow to discover, build, and maintain production-grade science and software



SciCLOps: How does it work?

SciCLOps

How does SciCLOps work?



Service Portal

"I want a project"

- Users fill out a form providing details about the kind of data access they need, name their project, etc.
- The service portal uses the automation engine API to create the infrastructure



Automation (Terraform)

"I am getting a project"

- Creates infrastructure for the user
- Applies updates to the infrastructure (day two operations)
- Provides an overview of all environments



GHA* Reusable Workflow

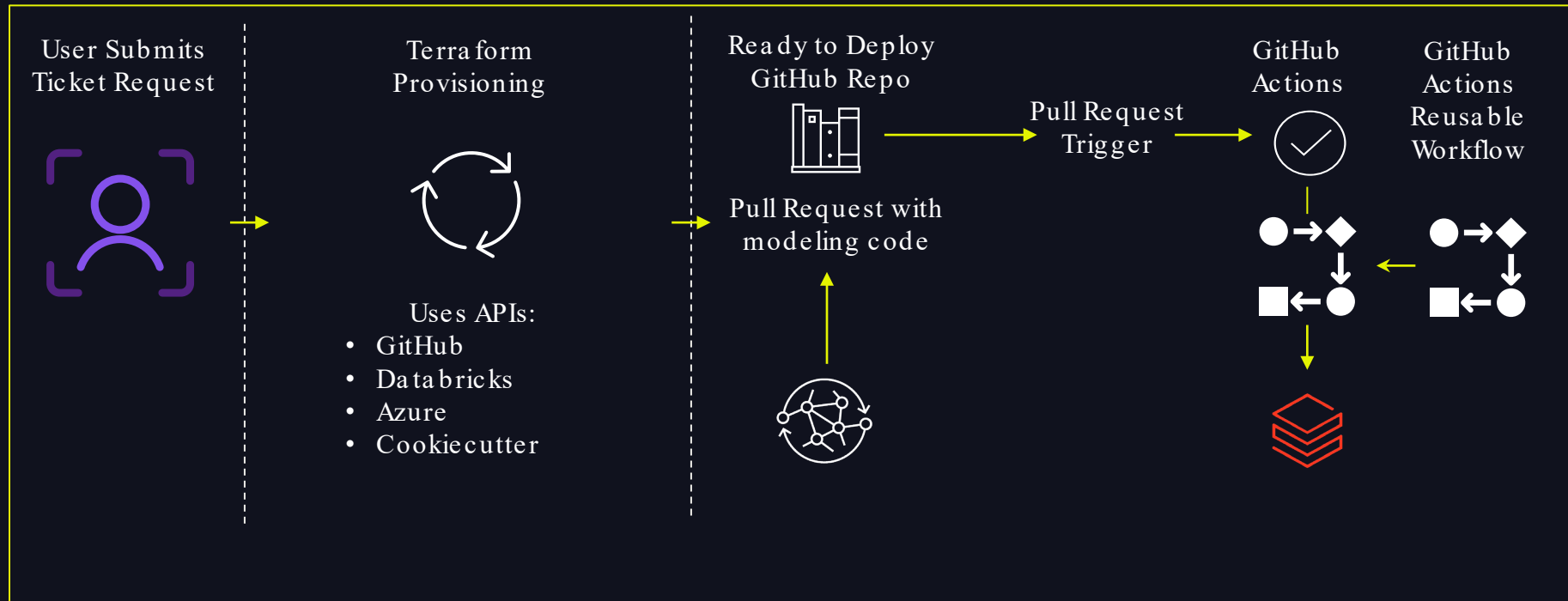
"I want to deploy my project"

- Deploys starter code and future commits to Databricks (via DABs)
- Allows the SciCLOps team to update the workflow with minimal user disruption

* = *GitHub Actions*

SciCLOps

How does SciCLOps work?



SciCLOps

Day One and Day Two

“Day One”

Project Provisioning

- We target less than ten minutes from form submission to project availability
- Customers are often deployed to Databricks literally on “day one”

“Day Two”

Project Maintenance and Deprovisioning

- We refer to updates and maintenance of projects provisioned using SciCLOps as “day two”
- Example: Key rotation, new SciCLOps features (such as DABs)
- Destroying a project also counts as “day two”, such as for demo or sandbox projects

Demo: SciCLOps on "Day 1"

Service Portal

Home > Service Catalog > Cloud > SciClops

Search Catalog

SciClops

SciClops

Add a SciClops configuration to a Managed Team's Gigabricks Environments.

***Project Owners**

***Managed Team AAD Group**

***SciCLOps Deploy Name (Name of the team's SciCLOps Project)**

***GitHub Team Name**

***GitHub Reviewers**

***Gigabricks Environment**

gbx1

gbx2

gbxphil

gbxpiil

Automation Engine: Day 1

The screenshot displays the Databricks Automation Engine interface for a workspace named "sciclops-sciclops-dais-demo". The breadcrumb path is "8451LLC / Workspaces / sciclops-sciclops-dais-demo / Runs / run-ao6eWXg9WNL39rWf". The run ID is "ws-An9x3JWQ3CS5yWwY". The run is currently in the "Planning" phase, as indicated by the "CURRENT" and "Planning" buttons. The run was triggered via ServiceNow Automation. Key metrics shown include: Estimated cost change (0.00), Plan duration (Less than a minute), and Resources to be changed (0). The run log shows that the run was triggered by user "connorbrown-8451" a few seconds ago. The current step is "Plan running", which started a few seconds ago. The log entry shows "Agent pool Automation" with agent "Agent az-lix-vs-tfc-auto-000002". A "Download raw log" button is available. Below the log, there are three pending steps: "Waiting for resources...", "Cost estimation pending", and "Apply pending". At the bottom, there is a "Cancel run" section with a "Cancel run" button and an "Add comment" button.

8451LLC / Workspaces / sciclops-sciclops-dais-demo / Runs / run-ao6eWXg9WNL39rWf

sciclops-sciclops-dais-demo Force unlock + New run

ID: ws-An9x3JWQ3CS5yWwY [Add workspace description](#)

Running Resources 0 Terraform =>1Z7Q Updated a few seconds ago

Run triggered via ServiceNow Automation CURRENT Planning

Estimated cost change Plan duration Resources to be changed

Less than a minute

connorbrown-8451 triggered a run from UI a few seconds ago Run Details

Plan running a few seconds ago

Started a few seconds ago

Agent pool Automation Agent az-lix-vs-tfc-auto-000002 Download raw log

Waiting for resources...

Cost estimation pending

Apply pending

Cancel run
You can cancel this run to stop it from executing.

Cancel run Add comment

Support Terms Privacy Security Accessibility © 2024 HashiCorp

Automation Engine: Day 1


8451LLC / Workspaces / sciclops-sciclops-dais-demo / Runs / run-ao6eWXg9WNL39rWF


sciclops-sciclops-dais-demo Force unlock + New run


ID: ws-An9xJUNQ3C5bYwJy [Add workspace description.](#)


Running Resources 0 Terraform >=1.7.0 Updated a few seconds ago


Run triggered via ServiceNow Automation CURRENT Applying

Estimated cost change	Apply duration	Resources to be changed
None	Less than a minute	+83 ~0 -0 

 connorbrown-8451 triggered a run from UI a few seconds ago Run Details

 **Plan finished** a few seconds ago Resources: 83 to add, 0 to change, 0 to destroy


 **Cost estimation finished** a few seconds ago Resources: 0 of 82 estimated - \$0.00/mo - +\$0.00

 **Apply running** a few seconds ago ^

Started a few seconds ago

Agent pool Automation Agent az-fox-vs-tlc-auto-000002 Download raw log


Waiting for resources...

 connorbrown-8451 a few seconds ago

Run confirmed

Cancel run
You can cancel this run to stop it from executing.

Cancel run Add comment

Support Terms Privacy Security Accessibility  © 2024 HashiCorp



sciclops-sciclops-dais-demo

Lock

+ New run

ID: ws-An9xJUNQ3C5bYwly

[Add workspace description.](#)

Unlocked Resources 118 Terraform >=1.7.0 Updated a few seconds ago

Run triggered via ServiceNow Automation

CURRENT

Applied

Estimated cost change

None

Plan & apply duration

3 minutes

Resources changed

+83 -0 -0

- connorbrown-8451 triggered a run from UI a few seconds ago Run Details
- Plan finished** a few seconds ago Resources: 83 to add, 0 to change, 0 to destroy
- Cost estimation finished** a few seconds ago Resources: 0 of 82 estimated - \$0.00/mo - +\$0.00
- Apply finished** a few seconds ago Resources: 83 added, 0 changed, 0 destroyed

connorbrown-8451 a few seconds ago

Run confirmed

Comment

Leave feedback or record a decision.

Add comment



Demo: SciCLOps on "Day 2"

General

Access

Collaborators and teams

Team and member roles

Code and automation

Branches

Tags

Rules

Actions

Webhooks

Capilot

Environments

Codespaces

Pages

Custom properties

Security

Code security and analysis

Deploy keys

Secrets and variables

Integrations

GitHub Apps

Email notifications

Autolink references

Environments / Configure Development

Deployment protection rules

Configure reviewers, timers, and custom rules that must pass before deployments to this environment can proceed.

Required reviewers

Specify people or teams that may approve workflow runs when they access this environment.

Wait timer

Set an amount of time to wait before allowing deployments to proceed.

Jira

Custom protection rule provided by Jira.

Enable additional custom rules with GitHub Apps Beta

[Learn about existing apps](#) or [create your own protection rules](#) so you can deploy with confidence.

Allow administrators to bypass configured protection rules

Save protection rules

Deployment branches and tags

Limit which branches and tags can deploy to this environment based on rules or naming patterns.

No restriction

Environment secrets

Add environment secret

Secrets are encrypted environment variables. They are accessible only by GitHub Actions in the context of this environment by using the [secrets context](#).

Name	Last updated
DATABRICKS_HOST	12 minutes ago
DATABRICKS_JOB_CLUSTER_POLICY_ID	11 minutes ago
DATABRICKS_TOKEN	12 minutes ago
DATABRICKS_TOKEN_MT_SP	12 minutes ago
MGDTH_CLIENT_ID	13 minutes ago

sciclops-sciclops-dais-demo

Force unlock

+ New run

ID: ws-An9xJUNQ3C5bYwWY

[Add workspace description.](#)

Running Resources 118 Terraform >1.7.0 Updated a few seconds ago

feat: Add new secret called DABS_ROCKS

CURRENT

Planning

Estimated cost change

Plan duration

Resources to be changed

Less than a minute

connorbrown-8451 triggered a run from GitHub a few seconds ago (failed to detect changed files)

Run Details

Plan running a few seconds ago

Started a few seconds ago

Agent pool Automation Agent az-lnx-vs-tfc-auto-000001

Filter resources by address...

Filter by action

Show data sources

Terraform 1.7.5

Download raw log

Cost estimation pending

Apply pending

Cancel run

You can cancel this run to stop it from executing.

Cancel run

Add comment

sciclops-sciclops-dais-demo

Force unlock

+ New run

ID: ws-An9xJUNQ3C5bYwWY

[Add workspace description.](#)

Running Resources 118 Terraform >1.7.0 Updated a few seconds ago

feat: Add new secret called DABS_ROCKS

CURRENT

Cost estimated

Estimated cost change
None

Plan duration
1 minute

Resources to be changed
+12 -0 -0

connorbrown-8451 triggered a run from GitHub a few seconds ago (failed to detect changed files) Run Details

Plan finished a few seconds ago Resources: 12 to add, 0 to change, 0 to destroy

Cost estimation finished a few seconds ago Resources: 0 of 94 estimated - \$0.00/mo - +\$0.00

Apply pending

Please review the following changes before continuing:

To create
+ 12

Choosing "Confirm & apply" below will execute the above changes.
Please [review the plan output](#) before proceeding.

Confirm & apply Discard run Add comment

sciclops-sciclops-dais-demo

Force unlock

+ New run

ID: ws-An9xJUNQ3C5bYwWY

[Add workspace description.](#)

Running Resources 118 Terraform >1.7.0 Updated a few seconds ago

feat: Add new secret called DABS_ROCKS

CURRENT

Applying

Estimated cost change
None

Apply duration
Less than a minute

Resources to be changed
+12 -0 -0

connorbrown-8451 triggered a run from GitHub a few seconds ago (failed to detect changed files)

Run Details

Plan finished a few seconds ago

Resources: 12 to add, 0 to change, 0 to destroy

Cost estimation finished a few seconds ago

Resources: 0 of 94 estimated - \$0.00/mo - +\$0.00

Apply running a few seconds ago

Started a few seconds ago

Agent pool Automation Agent az-lnx-vs-tfc-auto-000000

Download raw log

Waiting for resources...

connorbrown-8451 a few seconds ago

Run confirmed

Cancel run

You can cancel this run to stop it from executing.

General

Access

Collaborators and teams

Team and member roles

Code and automation

Branches

Tags

Rules

Actions

Webhooks

Copilot

Environments

Codespaces

Pages

Custom properties

Security

Code security and analysis

Deploy keys

Secrets and variables

Integrations

GitHub Apps

Email notifications

Autolink references

Environments / Configure Development

Deployment protection rules

Configure reviewers, timers, and custom rules that must pass before deployments to this environment can proceed.

Required reviewers

Specify people or teams that may approve workflow runs when they access this environment.

Wait timer

Set an amount of time to wait before allowing deployments to proceed.

Jira

Custom protection rule provided by Jira.

Enable additional custom rules with GitHub Apps Beta

[Learn about existing apps](#) or [create your own protection rules](#) so you can deploy with confidence.

Allow administrators to bypass configured protection rules

Save protection rules

Deployment branches and tags

Limit which branches and tags can deploy to this environment based on rules or naming patterns.

No restriction

Environment secrets

Add environment secret

Secrets are encrypted environment variables. They are accessible only by GitHub Actions in the context of this environment by using the [secrets context](#).

Name ↕↑	Last updated		
DATABRICKS_HOST	13 minutes ago		
DATABRICKS_JOB_CLUSTER_POLICY_ID	12 minutes ago		
DATABRICKS_TOKEN	13 minutes ago		
DATABRICKS_TOKEN_MT_SP	13 minutes ago		
MGMTM_CLIENT_ID	14 minutes ago		

Demo

User Experience:

Register New

Model Version

DABs Experience - PyCharm

The screenshot displays the PyCharm IDE interface. On the left, a project tree shows a directory structure including 'dais_ml_workflow_demo' and 'model_train_register'. The main editor window shows a Python file named '02_train_and_register_model.py'. A yellow box highlights a dictionary definition for 'high_level_model_tags' with the following code:

```
high_level_model_tags = {
    "business_unit": f"{business_unit}",
    "project_name": f"{project_name}",
    "model_name": f"{model_name}",
    "original_developer_email": f"{original_developer_email}",
    "current_owner_email": f"{current_owner_email}",
    "model_type": f"{model_type}",
    "model_package": f"{model_package}",
    "platform": f"{platform}",
    "dummy_model": f"{dummy_model}",
    "model_factory": f"{model_factory}",
    "dais_presentation": "yes",
}
```

Below the editor, a terminal window shows the execution of a Databricks CLI command:

```
(venv) [redacted] dais_ml_workflow_demo % databricks bundle deploy -t dev
Uploading bundle files to /Repos/[redacted]/.bundle/dais_ml_workflow_demo/my-envs/dev...
Deploying resources...
Updating deployment state...
Deployment complete!
(venv) [redacted] dais_ml_workflow_demo %
```

DABs Experience - Databricks Jobs

Deployment resources all listed under the development mode and user's name or ID

Name	Tags	Created by
[dev T123] dais_data_pull	T123	Tara Enright
[dev T123] dais_scoring_feature_monitoring	T123	Tara Enright
[dev T123] dais_train_and_register_model	T123	Tara Enright

Workflows > Jobs > [dev T123] dais_train_and_register_model

Connected to Databricks Asset Bundles

Runs Tasks

Job details

- Job ID: 123456789
- Creator: Tara Enright
- Run as: Tara Enright
- Tags: T123
- Description: Add description
- Lineage: No lineage information for this job.

No job runs were found for this job within the last 60 days

Run now



DABs Experience - Models

Two listed models under schema 'dais_demo' – the team's model and my personal one from a local deployment



Catalog Explorer [Send feedback](#)

Catalogs > sciclops_dev > sciclops_dev.dais_demo

Overview Details Permissions

Filter models

Tables 9 Volumes 0 Models 2

Name	Owner	Created at
dais_ml_workflow_demo	Team Service Principal ID	2024-05-14 11:00:30
dev_ T123 _dais_ml_workflow_demo	T123	2024-05-14 18:35:19

Catalogs > sciclops_dev > dais_demo > dev_ T123 _dais_ml_workflow_demo

Overview Details Permissions

Description

DAIS ML Workflow Demo

About this model

Owner: Tara Enright

Tags: business_unit: Platform Teams(Data and AI)

dais_presentation: yes dummy_model: False model_factory: False

model_package: XGBoost model_type: Regression

platform: Azure (Databricks)

project_name: California Housing Prices

Serve this model

Versions

Status	Version	Time registered	Tags	Aliases	Registered by	Comment
✓	Version 1	2024-05-14 18:39:55	training_win... : 2024... training_win... : 2024... workflow_versi... : lat...	@ champion	T123	



Bonus DABs!

How did we turn on a development mode?


Targets allow a user to...


update variables by environment


Turn on env deployment modes

```
bundle:  
  name: "dais_ml_workflow_demo"  
  
variables:  
  WORKING_DIRECTORY:  
    description: 'The working directory of the bundle in Databricks.'  
    default: "my_working_directory"  
  
workspace:  
  file_path: ${var.WORKING_DIRECTORY}  
  # It is not recommended to change the below parameter  
  state_path: /Users/${workspace.current_user.userName}/.bundle/${bundle.name}/${bundle.environment}/state  
  
targets:  
  dev:  
    mode: development  
  
resources:  
  registered_models:  
    my_model_name:  
  
  experiments:  
    my_experiment_name:  
  
  jobs:  
    my_training_job_definition:  
  
      job_clusters:  
        - job_cluster_key: my_training_jobs_cluster_definition
```


SciCLOps






 **dais_ml_workflow_demo** Internal Watch 2

 **dais** had recent pushes 2 seconds ago Compare & pull request

 main 1 Branch 1 Tags t Add file <> Code





-  **Review required**
At least 1 approving review is required by reviewers with write access. [Learn more about pull request reviews.](#)
-  **Some checks haven't completed yet** Hide all checks
1 in progress check
-   **model_train_register / tag_model_with_workflow_version / get_latest_release_tag (push)** *In p...* Details
-  **Merging is blocked**
Merging can be performed automatically with 1 approving review.



SciCLOps Github Actions

add tag and up model version #20

Summary

Jobs

- train-score-model
 - ✓ Dump Github Context
 - ✓ Set Jobs to Run
 - Create Semantic Version
 - ✓ **Deploy to Development**
 - Destroy Bundle in Development
 - Destroy Bundle in Test
 - Destroy Bundle in Staging
 - Destroy Bundle in Production
 - Overwrite setup.cfg File
 - **Run dais_data_pull in Development**

Triggered via push 1 minute ago

Status: **Queued** Total duration: - Artifacts: -

pushed a5d4d4c dais

train_register_monitor.yml
on: push

- t... / ... / Bump Semantic Versi...
- ✓ t... / ... / **Deploy to Develop...** 17s
- train-s... / ... / **Deploy to Testing**
- train-score-model / ... / deploy
- Matrix: train-score-model / Ru...
 - 0/1 jobs completed
 - Show all jobs
- Matrix: train-score-model / Ru...
 - Waiting for pending jobs
- tra... / **Dump Github Context** 0s
- train-score-model / ... / destroy
- train-score-model / ... / destroy
- train-score-model / ... / destroy
- train-score-model / ... / destroy
- train-score-model / ... / destroy



SciCLOps Dev Deployment to DBX

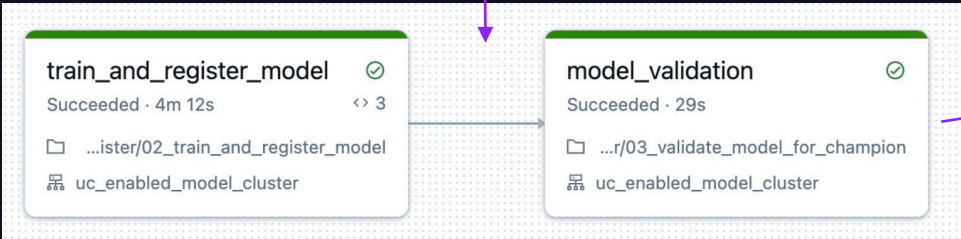
Workflows

Jobs Job runs Delta Live Tables

Search: dais [x] Only jobs owned by me 6 jobs found Reset filters

Name	Tags	Created by
[dev] T123 dais_data_pull	T123	Tara Enright
[dev] T123 dais_scoring_feature_monitoring	T123	Tara Enright
[dev] T123 dais_train_and_register_model	T123	Tara Enright
dais_data_pull		Team Service Principal ID
dais_scoring_feature_monitoring		Team Service Principal ID
dais_train_and_register_model		Team Service Principal ID

SciCLOps auto-kicks off jobs!



Filters [x] Search: is:pr is:open

1 Open ✓ 3 Closed

add tag and up model version ✓

#4 opened last week by [redacted] · Review required



10 minutes later...

SciCLOps - Approved to Production

🔔 requested your review to deploy to Production Review deployments

train_register_monitor.yml
on: push



Review pending deployments ✕

requested by [redacted] in model_train_register #23

Production [redacted]

Leave a comment:

Reject Approve and deploy 1



Learnings

What made this a success?

- Community of Practice established with SciCLOps users for sharing demos, thoughts, etc.
- Inner Source contributions
- Team with diverse talent
- Datadog integration for monitoring usage of the GitHub Actions Reusable Workflows
- SciCLOps testing suite deployed from Terraform module
- DABs + Terraform Cloud

THANK YOU

84.51° SciCLOps Team

- Matt Willis
- Jacob Ford
- Colin Donovan
- Ian Listopad
- Caleb Goodreau
- Jingling Zhang
- Vanisha Bafna

84.51° Key Collaborators

- Kyle Brooks
(Original developer)
- Caitlin Casar
- Ken Meyers
- Greg Anderson
- Mattias Moser
- Chris Mathers

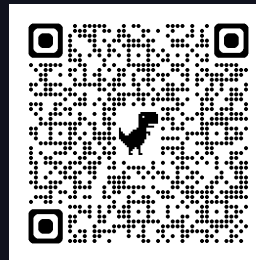
Databricks DABs Team

- Pieter Noordhuis
- Saad Ansari
- Theresa Hammer

Our Other Talks!

Check out more from 84.51°

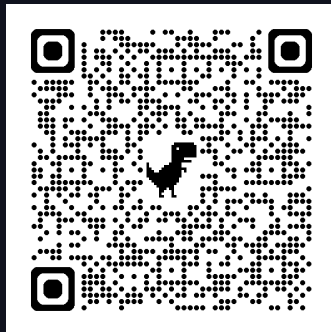
Learn more about DABs!



Sponsored by: Alation |
Delivering Data
Products in a Data Mesh
at Kroger



Unlocking Data Value:
84.51°'s Journey with
Databricks Unity Catalog



Seamless Flow: Evolving
From Batch to Streaming
Data Flows Using DLT

