



AI Center of Excellence
Office of the CTO

Feedback about deployment of an intelligent app

DevConf US 2021

Francesco Murdaca
Senior Data Scientist

- Project Thoth
- MLOps relation
- Why, What, When
- How to use it
- Pipeline and results

Project Thoth



Project Thoth

- Help developers in the selection of dependencies for their applications depending on their requirements
- Deliver optimized images for your applications
- Use bots to automate mundane work to offload humans work

How Thoth can help developers?

- Keep dependencies up to date.
- Maintain software stack secure avoiding CVE.
- Recommend the most performant software stacks for AI Apps.
- Integrate source metadata information related to the packages used in the software stack to give advice to users.
- Integrate Thoth in day-to-day developers/data scientists tools.
- ...

Thoth Integrations

- Command line tool thamos (developer laptop)
- Jupyter Tools (data scientist browser)
- Cyborg Kebechet (pull request/issues creator)
- Source-to-Image (container builder)
- Optimizing Deployment Pipeline

```
host: {THOTH_SERVICE_HOST}
tls_verify: true
requirements_format: {requirements_format}

runtime_environments:
  - name: '{os_name}:{os_version}'
    operating_system:
      name: {os_name}
      version: '{os_version}'
    hardware:
      cpu_family: {cpu_family}
      cpu_model: {cpu_model}
    python_version: '{python_version}'
    cuda_version: {cuda_version}
    recommendation_type: stable
    platform: '{platform}'
```

MLOps relation

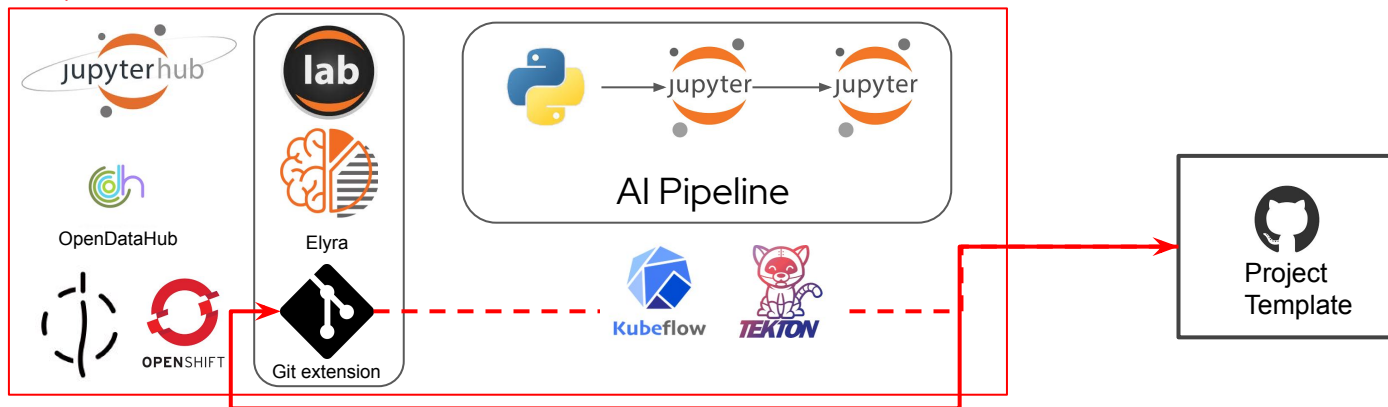
Providing feedback about deployment of an intelligent app



Data
Scientist

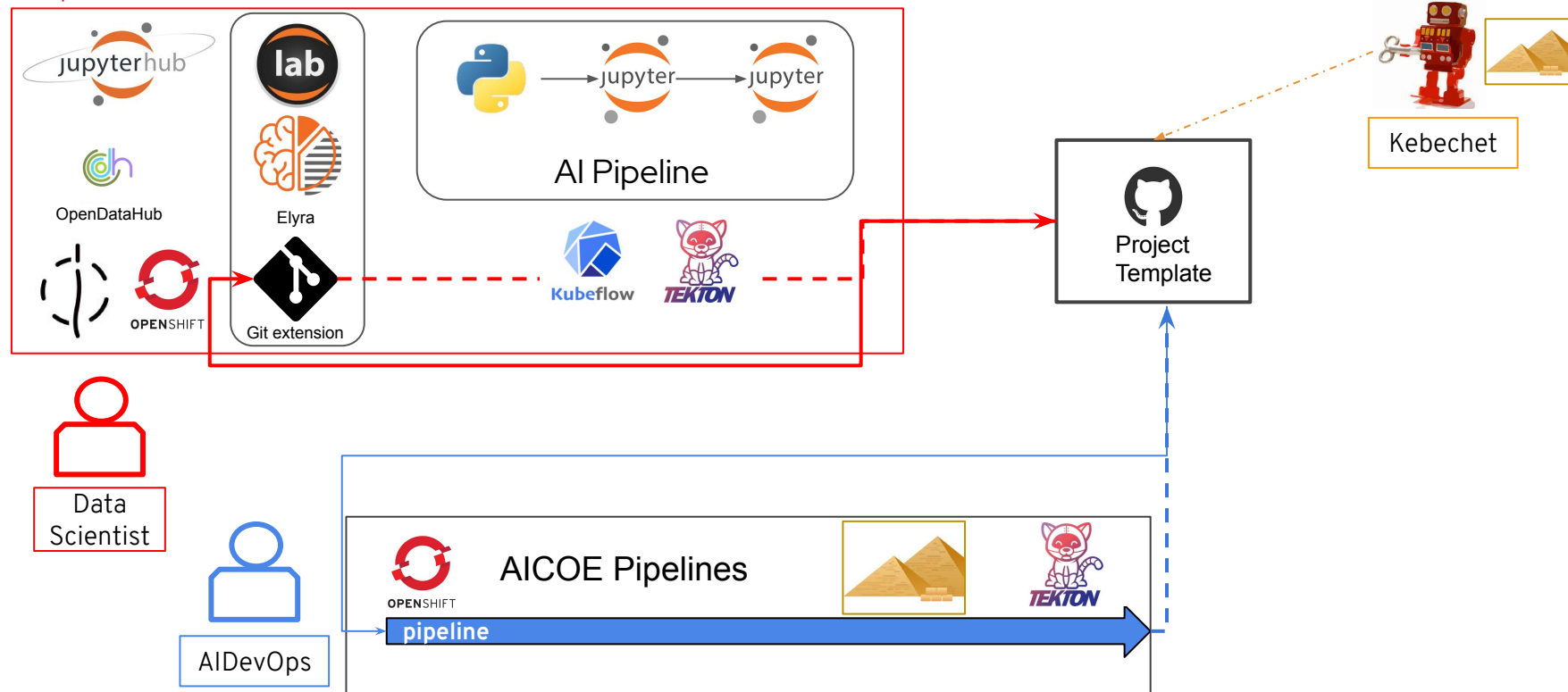


Providing feedback about deployment of an intelligent app

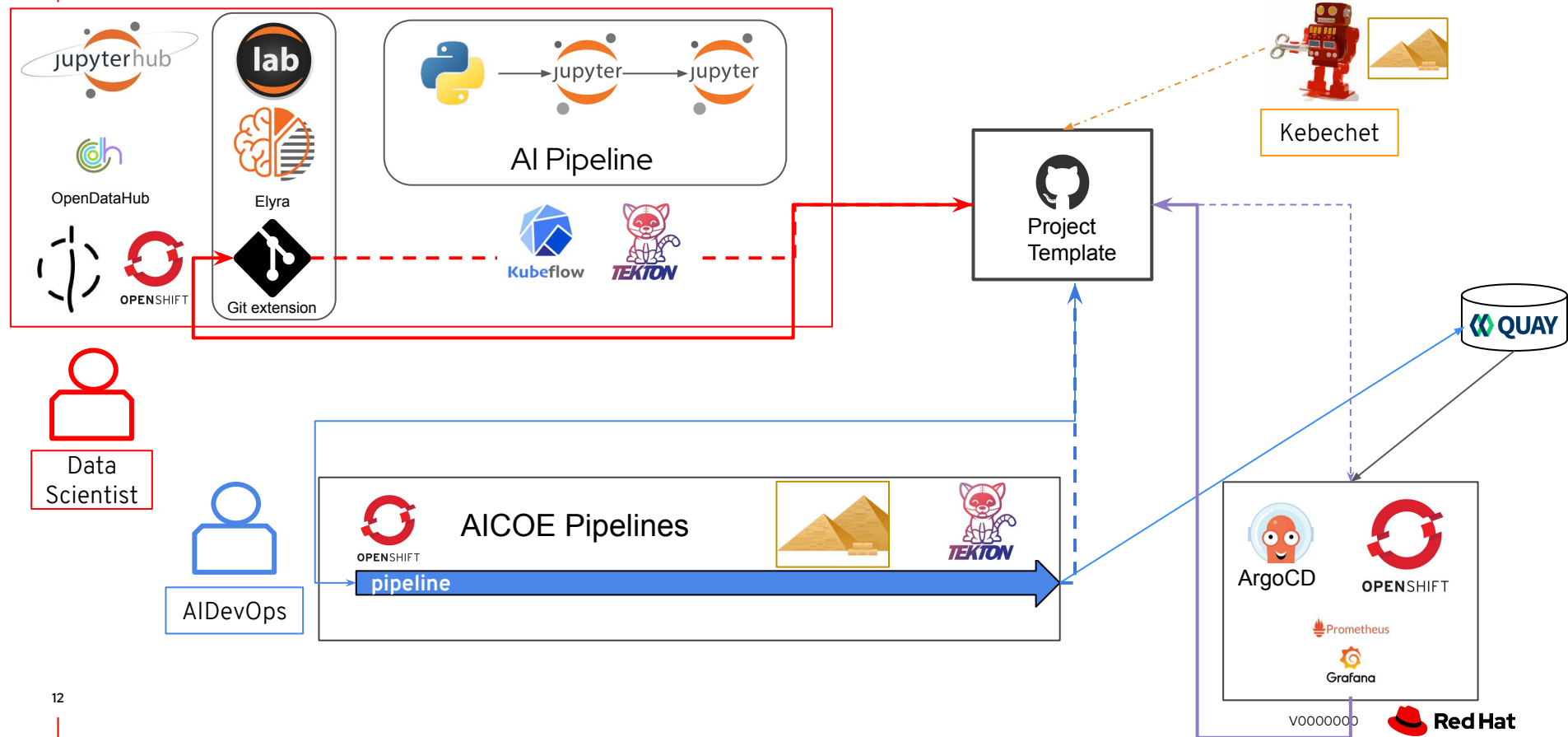


Data
Scientist

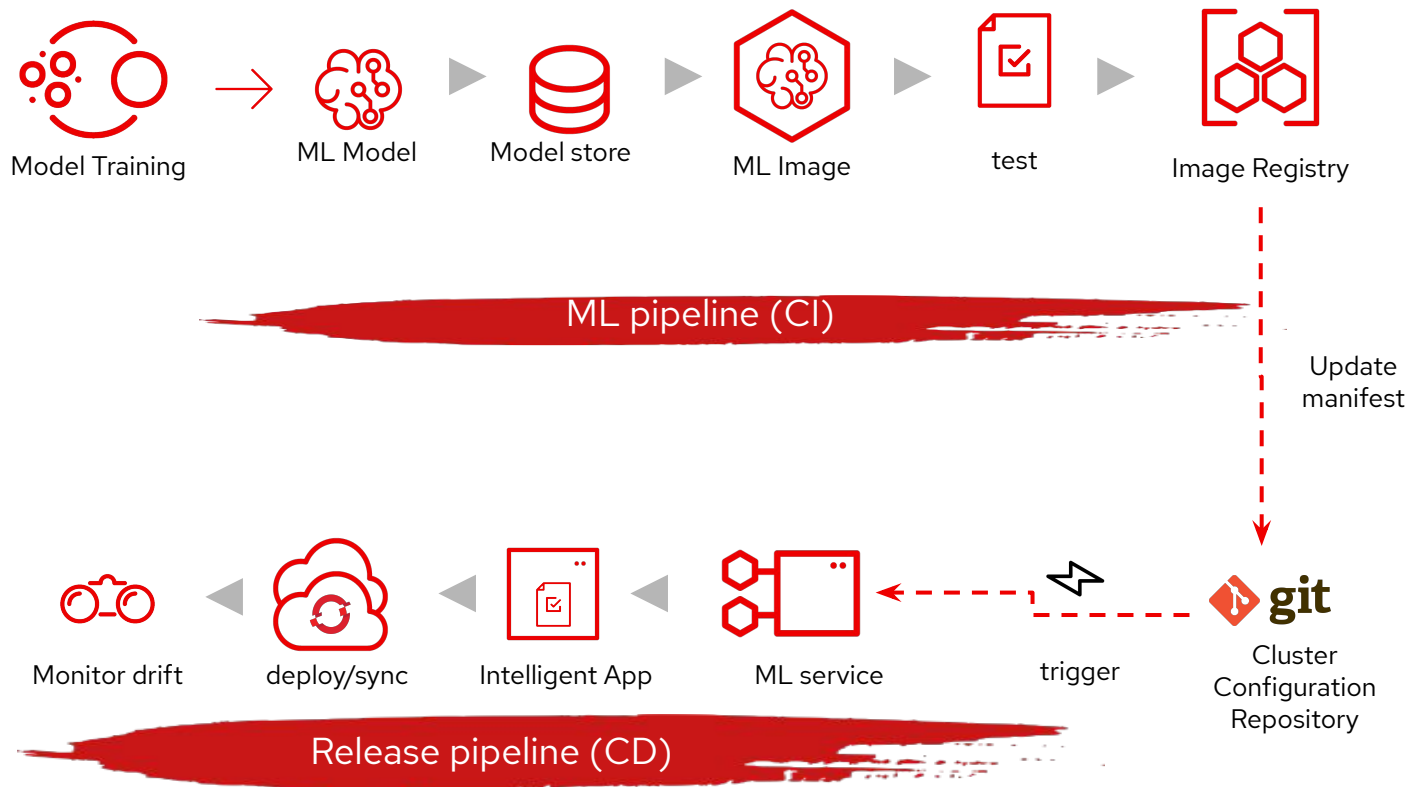
Providing feedback about deployment of an intelligent app



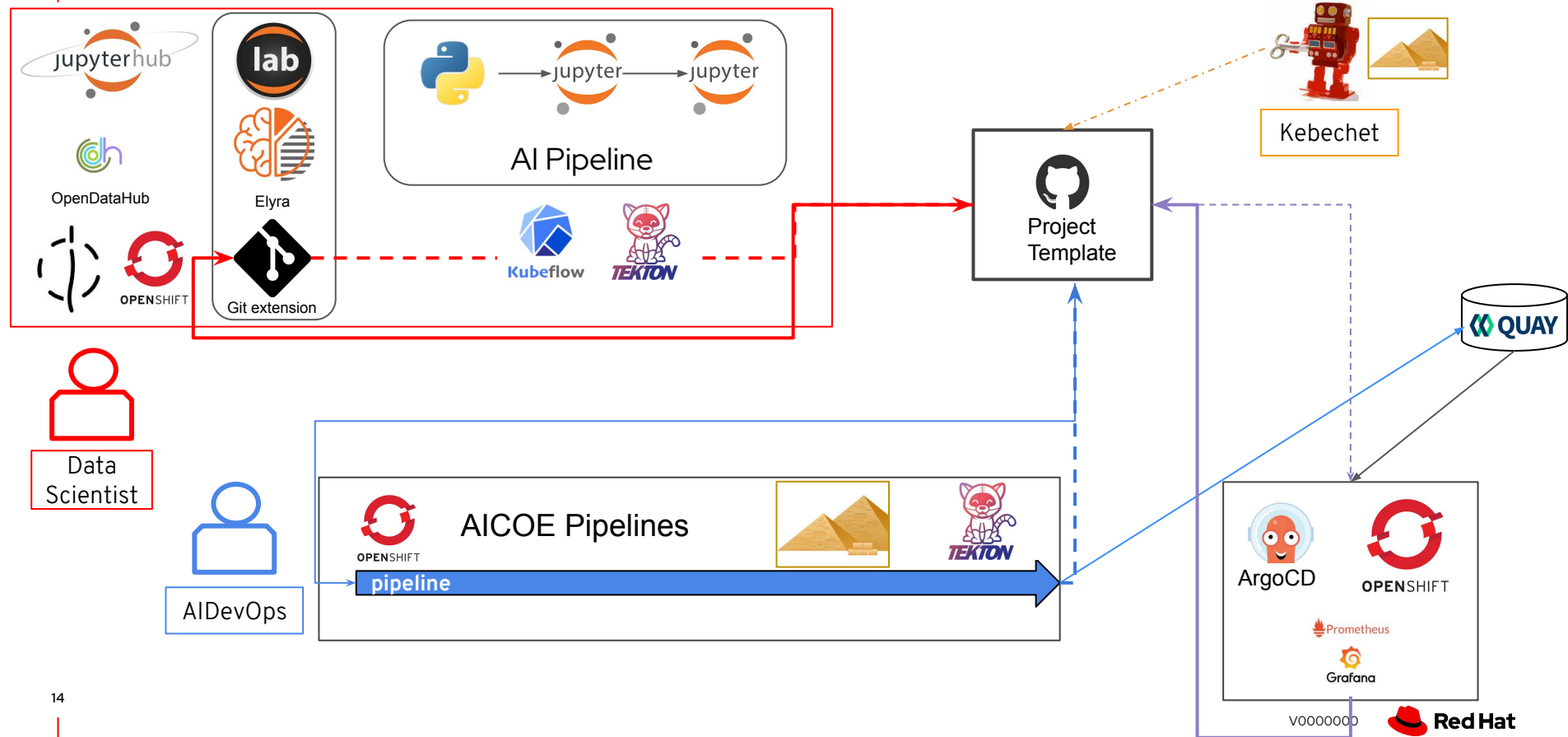
Providing feedback about deployment of an intelligent app



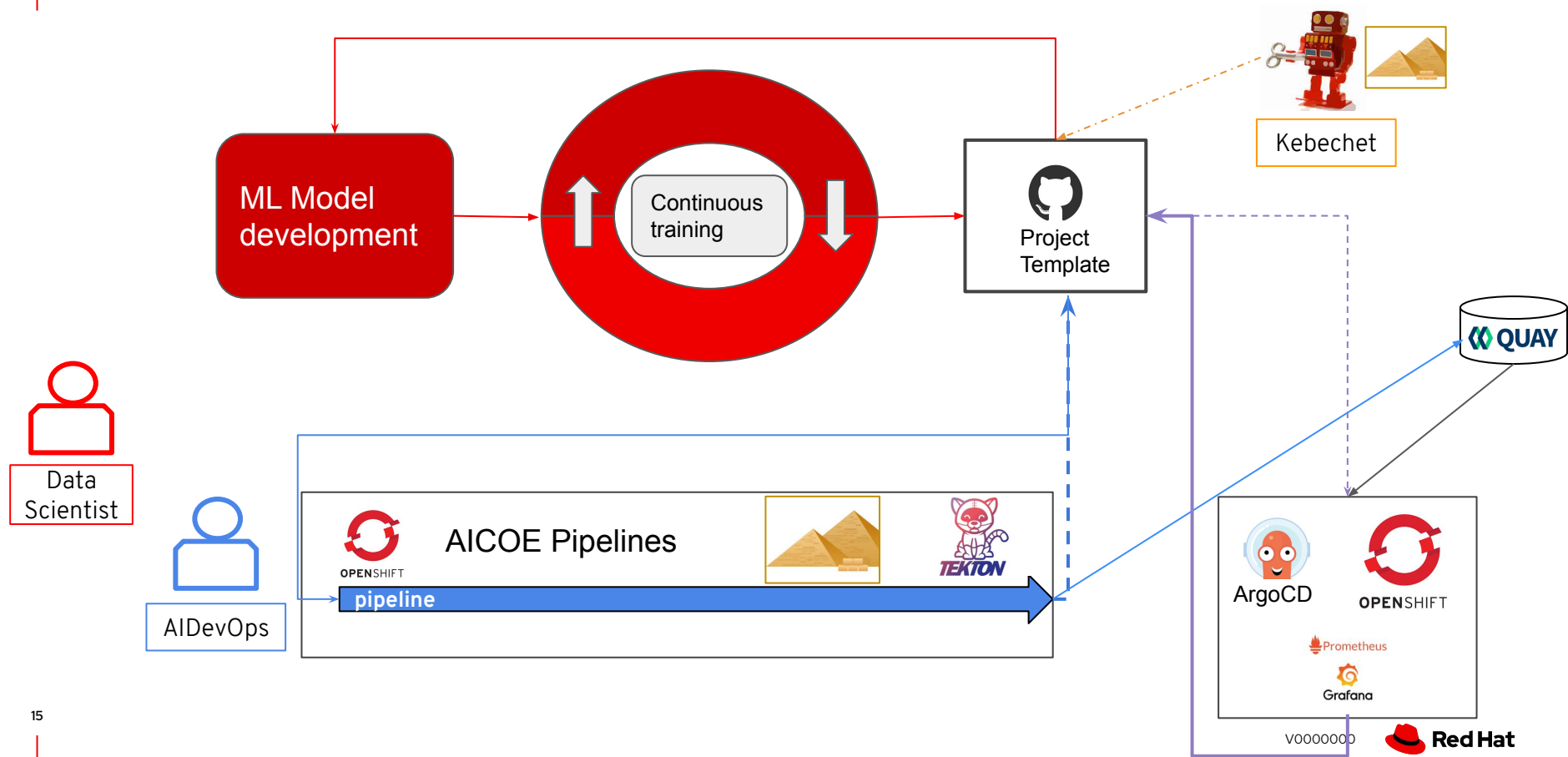
Providing feedback about deployment of an intelligent app



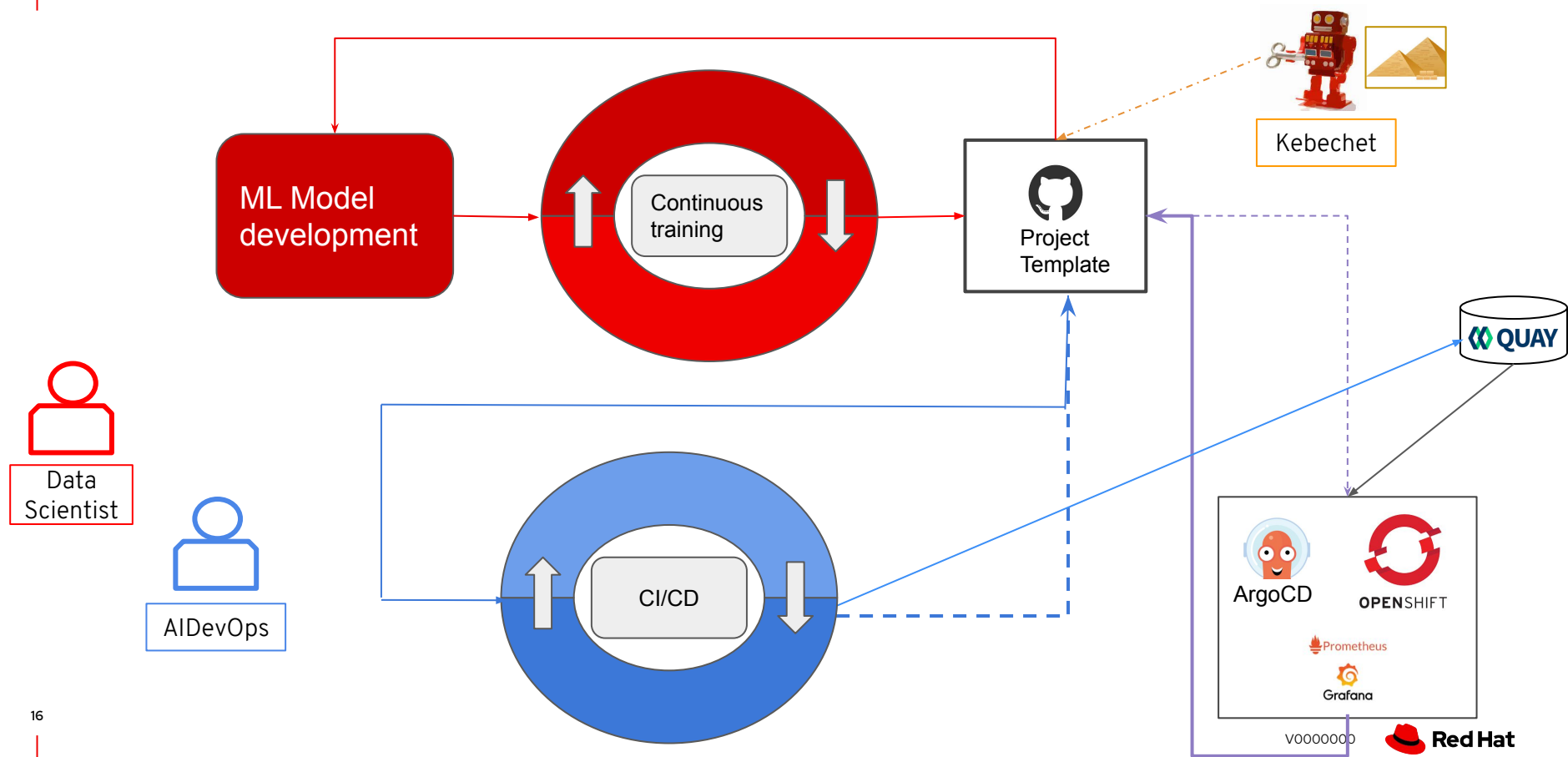
Providing feedback about deployment of an intelligent app



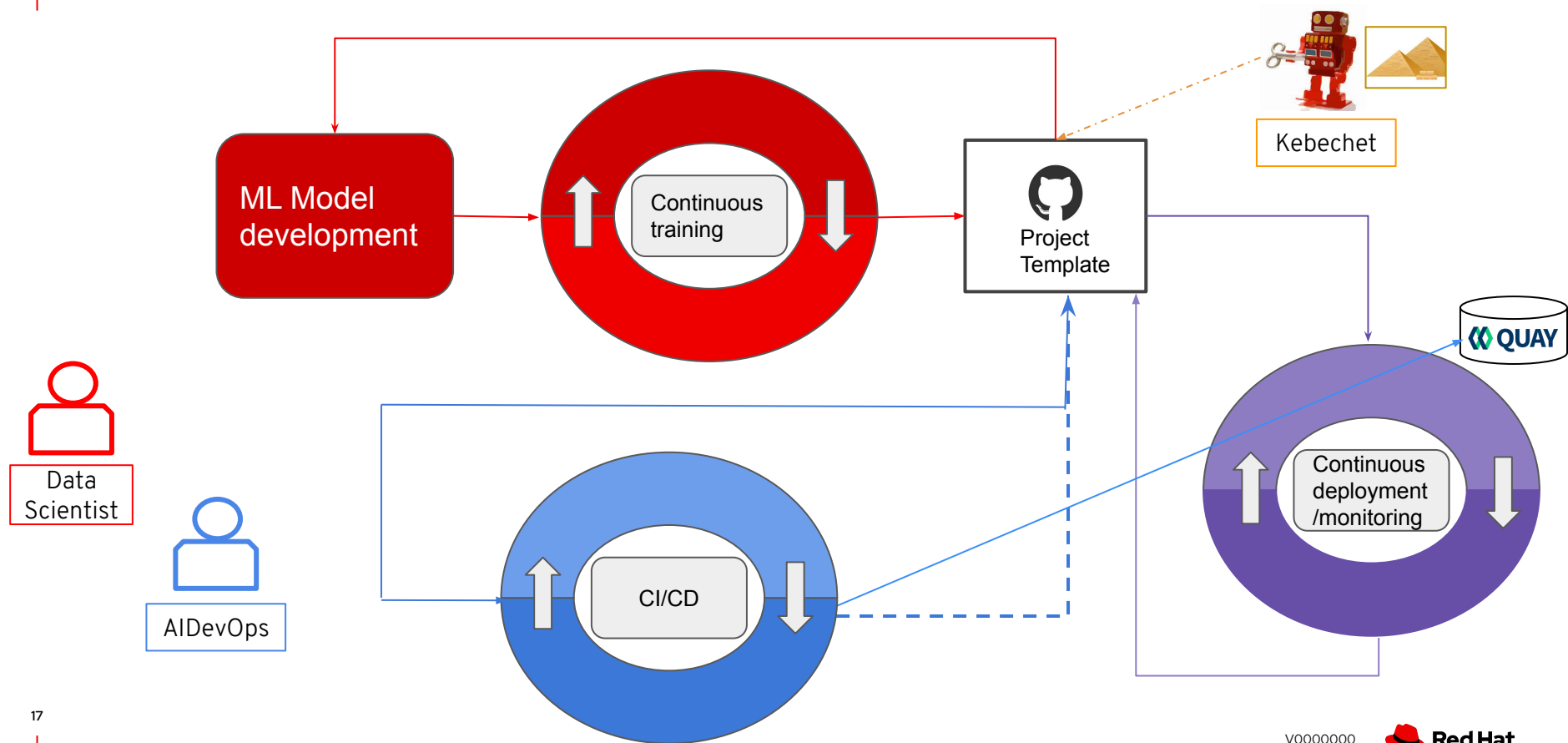
Providing feedback about deployment of an intelligent app



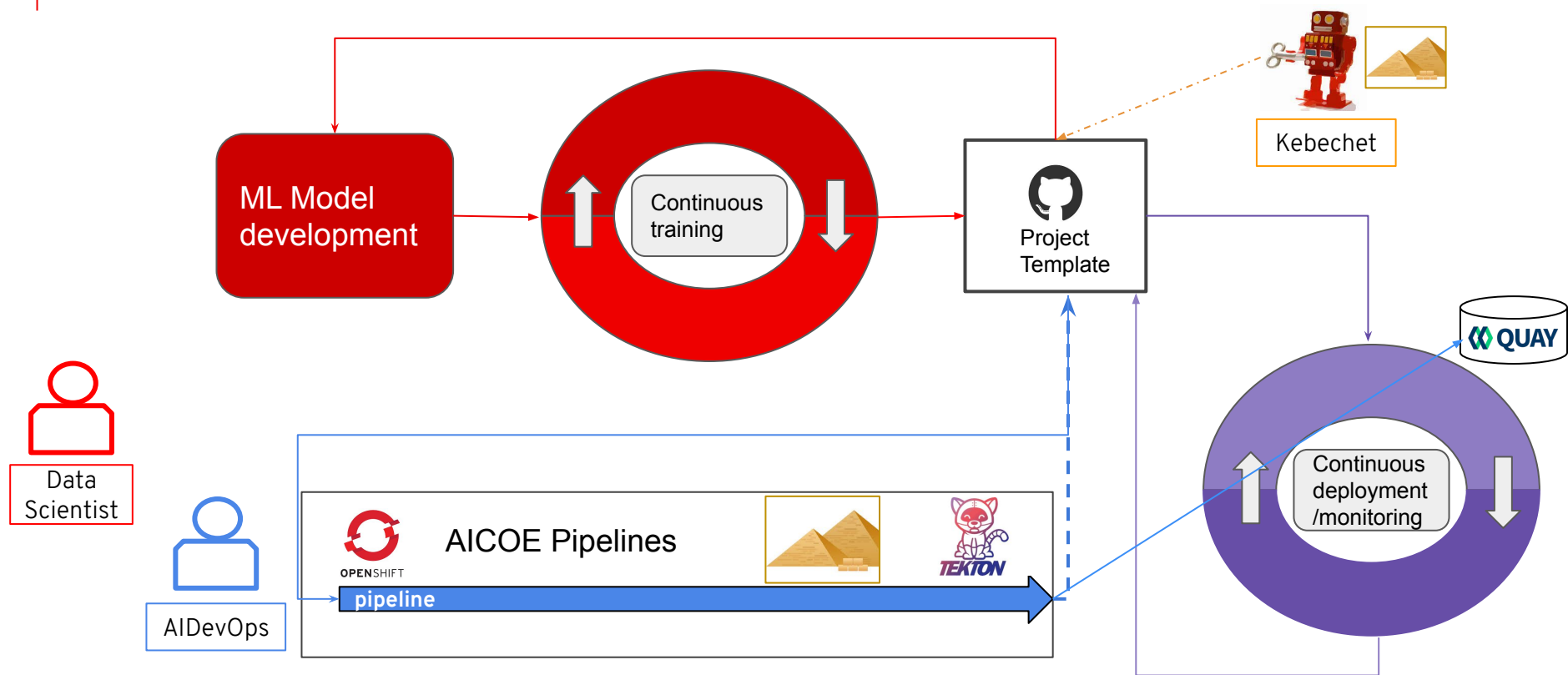
Providing feedback about deployment of an intelligent app



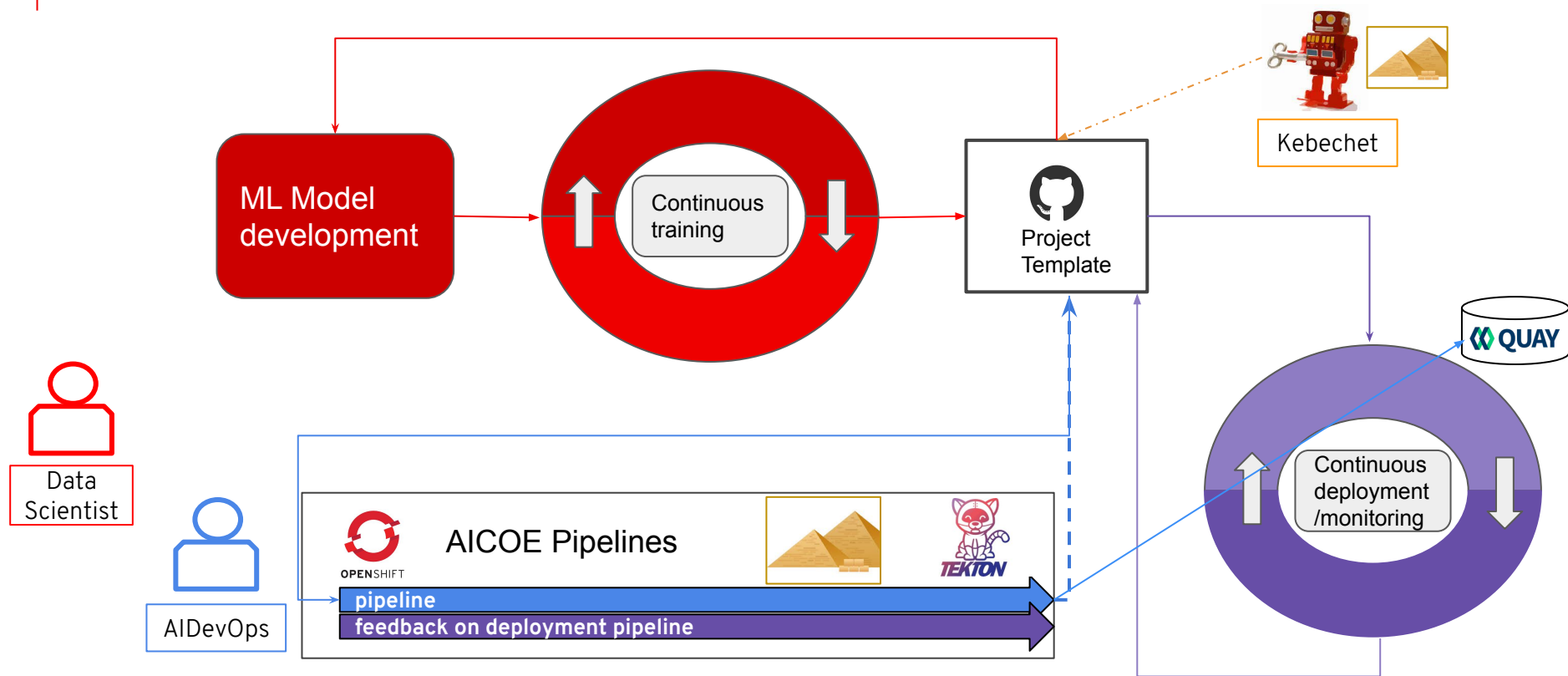
Providing feedback about deployment of an intelligent app



Providing feedback about deployment of an intelligent app



Providing feedback about deployment of an intelligent app



Why, What, When

Why feedback?

- App lifecycle is not static
- Speed up process
- Focus on modeling and data
- Reduce cost

What feedback?

- Model metrics
- Application metrics
- Platform Metrics

Personas that benefit from the feedback?

- AIDevOps wants to know operational information (latency, memory consumption, CPU usage)
- Data Scientist wants to verify model created is still performing well in production environment

When giving feedback?

- When a new software stack is created
- When a new model version is created
- When a new dataset version is created

How to use it?

What do you need?

- AICoE-CI
- Khebut Bot
- Model to be deployed
- Test to collect metrics

AICoE CI



This is the Continuous Integration Cyborg maintained by Thoth Station.



Manage your installation settings.

Developer



aicoe-ci is provided by a third-party and is governed by separate terms of service, privacy policy, and support documentation.



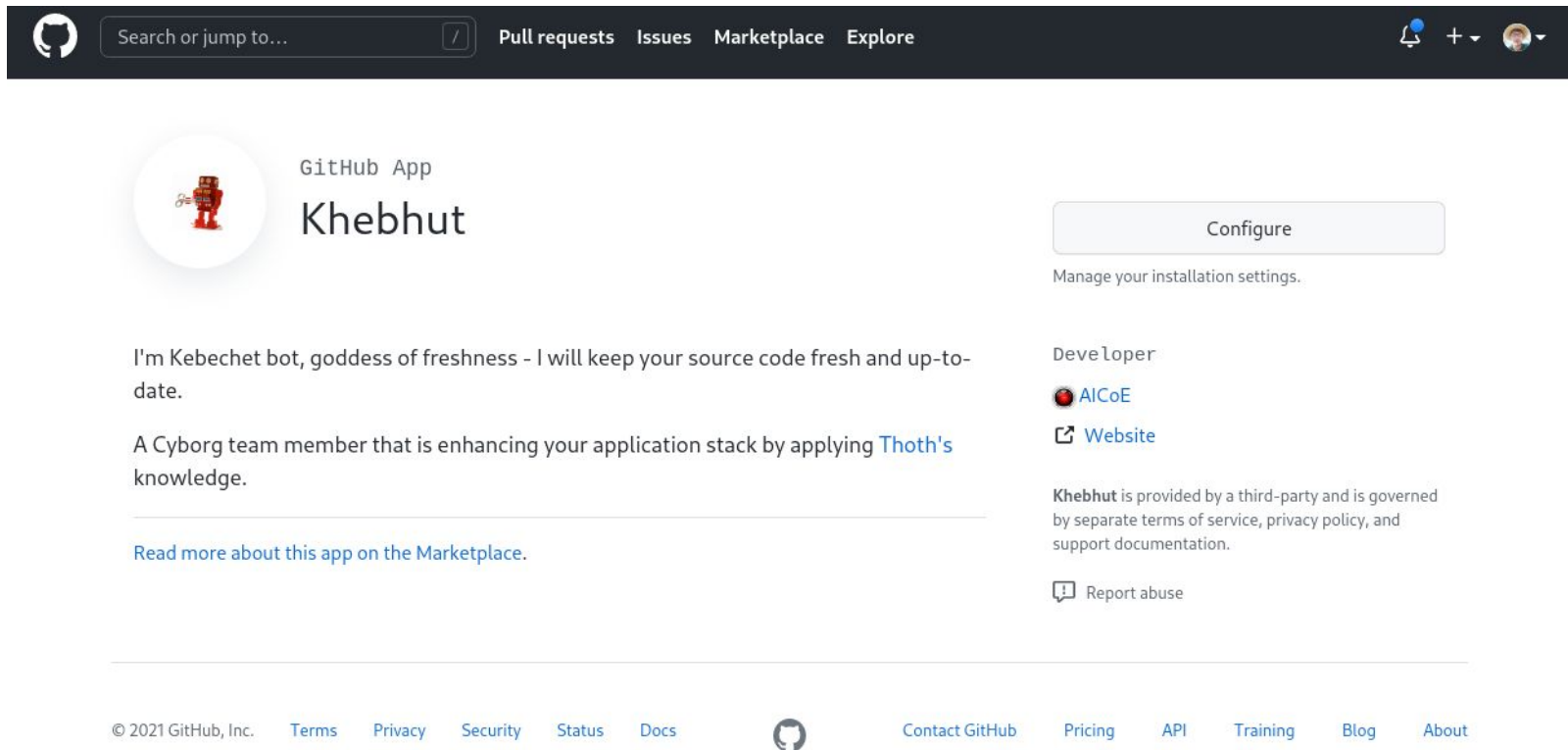
AICoE CI

```
33
34 - name: inference
35   build:
36     base-image: "quay.io/thoth-station/s2i-thoth-ubi8-py38:v0.28.0"
37     build-strategy: Source
38     registry: quay.io
39     registry-org: thoth-station
40     registry-project: elyra-aidevsecops-tutorial
41     registry-secret: thoth-station-thoth-pusher-secret
42   deploy:
43     project-org: "thoth-station"
44     project-name: "elyra-aidevsecops-tutorial"
45     image-name: "elyra-aidevsecops-tutorial"
46     overlay-contextpath: "manifests/overlays/test/imagestreamtag.yaml"
```

Deployments


- ▶ **Flask Application**
- ▶ KFServing
- ▶ Seldon
- ▶ TensorRT

Providing feedback about deployment of an intelligent app



The screenshot shows the GitHub App page for 'Khebhut'. At the top is a dark navigation bar with the GitHub logo, a search bar, and links for 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. On the right of the bar are notification, add, and user icons. The main content area has a white background. On the left, there's a circular profile picture of a red robot, the text 'GitHub App' and 'Khebhut', and a bio: 'I'm Kebechet bot, goddess of freshness - I will keep your source code fresh and up-to-date. A Cyborg team member that is enhancing your application stack by applying Thoth's knowledge.' Below the bio is a link 'Read more about this app on the Marketplace.' On the right, there's a 'Configure' button with the text 'Manage your installation settings.' below it. Further down, it says 'Developer' with a profile picture and name 'AICoE', followed by a 'Website' link. At the bottom, there's a disclaimer: 'Khebhut is provided by a third-party and is governed by separate terms of service, privacy policy, and support documentation.' and a 'Report abuse' link. The footer contains copyright information '© 2021 GitHub, Inc.' and various links: 'Terms', 'Privacy', 'Security', 'Status', 'Docs', 'Contact GitHub', 'Pricing', 'API', 'Training', 'Blog', and 'About'.

Search or jump to... / Pull requests Issues Marketplace Explore

 GitHub App
Khebhut

I'm Kebechet bot, goddess of freshness - I will keep your source code fresh and up-to-date.


A Cyborg team member that is enhancing your application stack by applying [Thoth's](#) knowledge.

[Read more about this app on the Marketplace.](#)

[Configure](#)

Manage your installation settings.


Developer

 AICoE

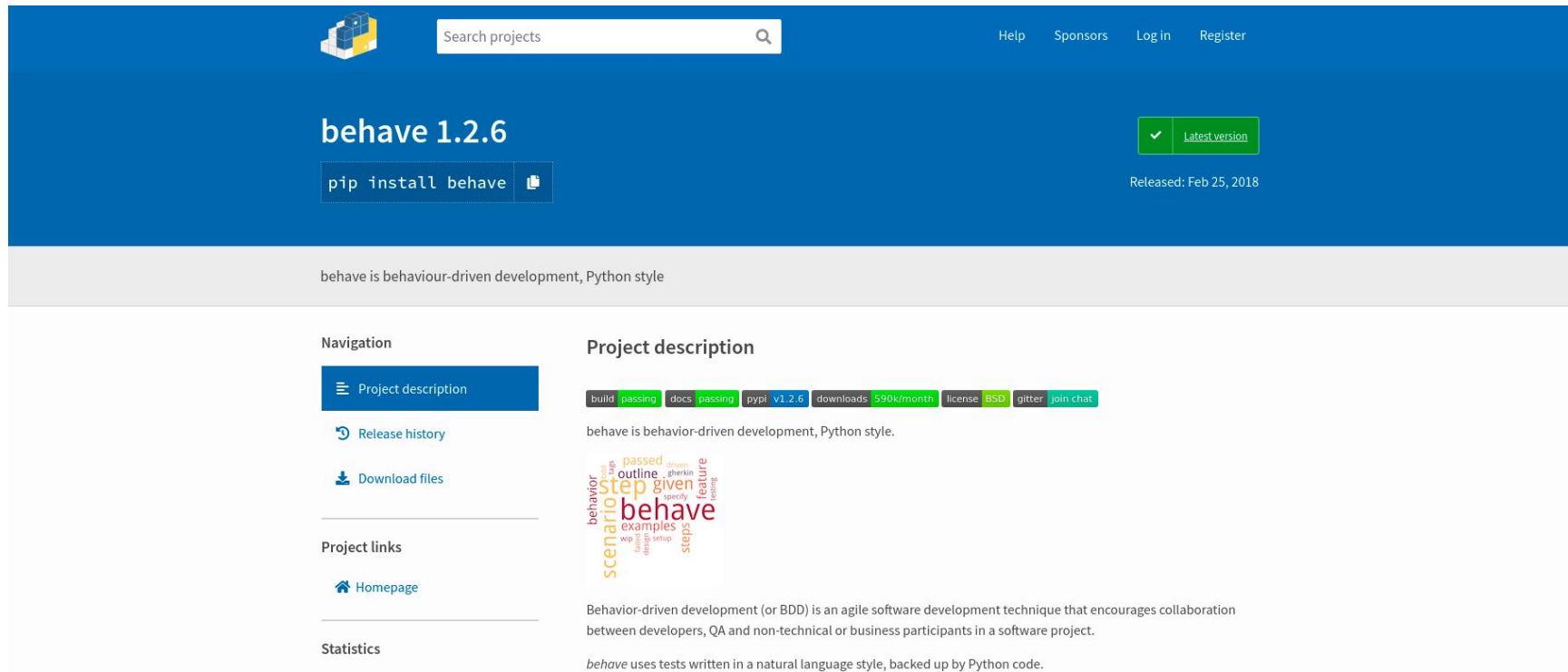
[Website](#)

Khebhut is provided by a third-party and is governed by separate terms of service, privacy policy, and support documentation.

[Report abuse](#)

© 2021 GitHub, Inc. [Terms](#) [Privacy](#) [Security](#) [Status](#) [Docs](#)  [Contact GitHub](#) [Pricing](#) [API](#) [Training](#) [Blog](#) [About](#)

behave



The screenshot shows the PyPI page for the 'behave' project. The header is blue with a search bar and navigation links (Help, Sponsors, Log in, Register). The main section features the project name 'behave 1.2.6' in large white text, a 'pip install behave' button, and a 'Latest version' badge. Below this, a description states 'behave is behaviour-driven development, Python style'. The left sidebar contains navigation links: 'Project description' (selected), 'Release history', 'Download files', 'Project links' (with a 'Homepage' link), and 'Statistics'. The right section, titled 'Project description', includes a row of badges for build status, docs, version, downloads, license, and chat. Below the badges, the description text is repeated, followed by a word cloud and a paragraph explaining Behavior-driven development (BDD) and how 'behave' uses tests written in a natural language style.

behave 1.2.6

`pip install behave`

Released: Feb 25, 2018

behave is behaviour-driven development, Python style

Navigation

- Project description
- Release history
- Download files

Project links

- Homepage

Statistics

Project description

build passing docs passing pypi V1.2.6 downloads 590k/month license BSD gitter join chat

behave is behavior-driven development, Python style.

behavior step outline given feature scenario examples steps

Behavior-driven development (or BDD) is an agile software development technique that encourages collaboration between developers, QA and non-technical or business participants in a software project.

behave uses tests written in a natural language style, backed up by Python code.

6 lines (6 sloc) | 274 Bytes

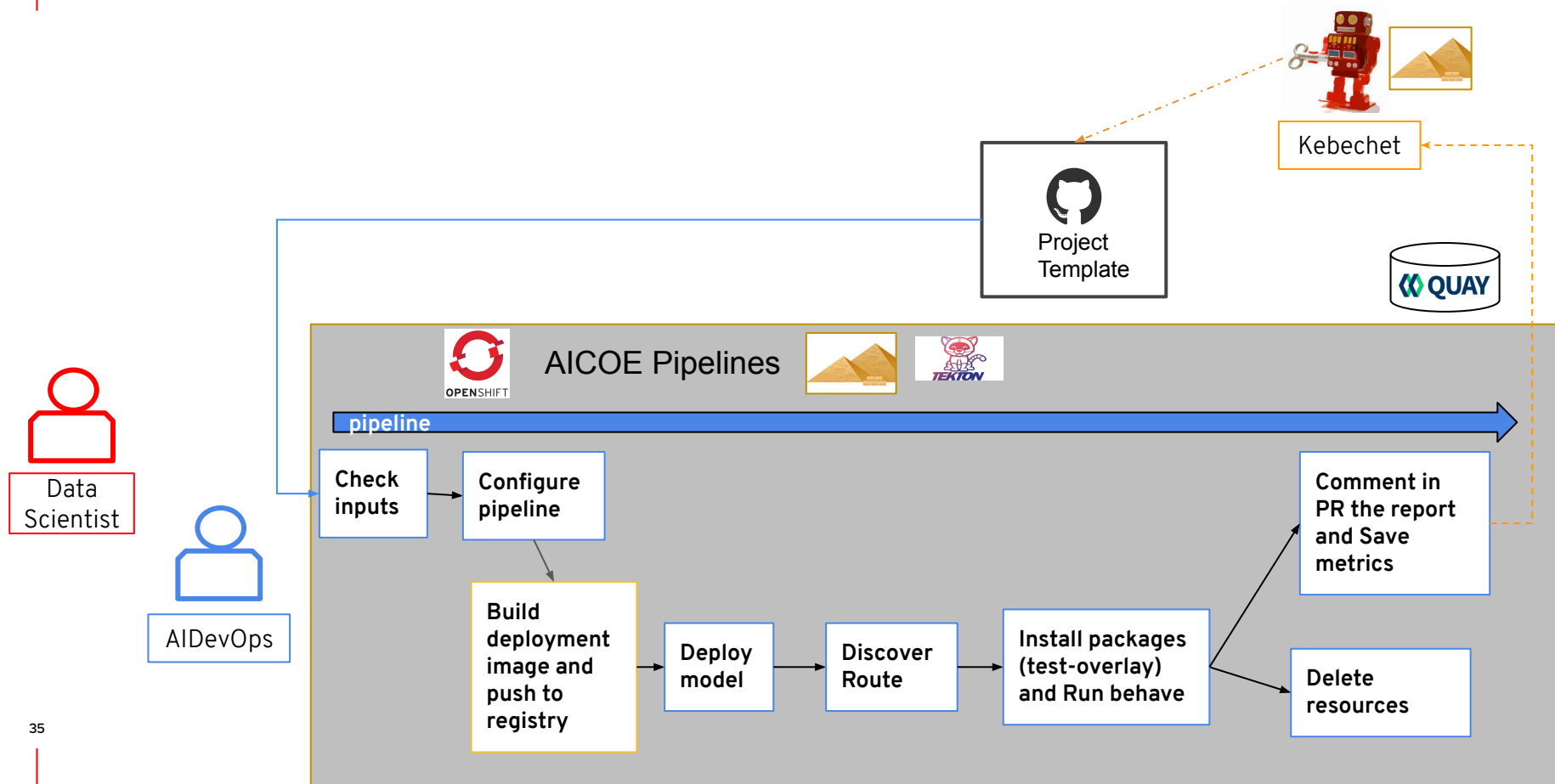
```
1  Feature: Gather metrics deployed model
2      Scenario: Deployment metrics gathering
3          Given dataset is available
4          Given deployment is accessible
5          When I run test to gather metrics on predict endpoint
6          Then I should get model and application metrics
```


Providing feedback about deployment of an intelligent app

```
41 @given("dataset is available")
42 def dataset_availability(context):
43     """Check availability of dataset and retrieves it."""
44     # Prepare MNIST data.
45     _, (x_test, y_test) = tf_dataset.load_data()
46
47     context.dataset = {
48         "x_test": x_test,
49         "y_test": y_test,
50     }
51
52     assert context.dataset
53
54
55 @given("deployment is accessible")
56 def deployment_accessible(context):
57     """Check the deployment is accessible."""
58     context.result = {}
59
60     context.model_api_url = os.environ["DEPLOYED_MODEL_URL"]
61
62     response = requests.get(f"{context.model_api_url}")
63
64     assert (
65         response.status_code == 200
66     ), f"Invalid response when accessing {context.model_api_url}: {response.status_code!r}: {response.text}"
67
68     assert response.text, f"Empty response from server for {context.model_api_url}"
```

Pipeline and results

Providing feedback about deployment of an intelligent app





sesheta commented 1 hour ago

Member



AICoE CI results

Test inputs

The following table shows info about test used to gather metrics.

test URL	namespace deployment
https://github.com/AICoE/elyra-aidevsecops-tutorial/blob/e882fa5d4534aa0ec1909911665d1b05dab35736/features	aicoe-ci

Model and application metrics

The following table shows gathered metrics for model and application on your deployed models.

average_latency	average_error	average_probability	number_requests	model_version
0.118086	0	0.999916	21	pr-160

Platform metrics

The following table shows gathered metrics from platform on your deployed models.

CPU max usage	Memory max usage
1	384Mi

- **Thoth station**
 - <https://thoth-station.ninja/>
- **AICoE CI**
 - <https://github.com/aicoe/aicoe-ci>
- **Khebut**
 - <https://github.com/apps/khebhut>
- **Thoth YouTube**
 - https://www.youtube.com/channel/UCIUIDuq_hQ6vlzmqM59B2Lw
- **Operate First**
 - <https://www.operate-first.cloud/>

Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.



linkedin.com/company/red-hat



youtube.com/user/RedHatVideos



facebook.com/redhatinc



twitter.com/RedHat