Connect

Red Hat

ChatOCP - Intelligent application development on OpenShift

Juozas Auskalnis
Solution Architect

Red Hat

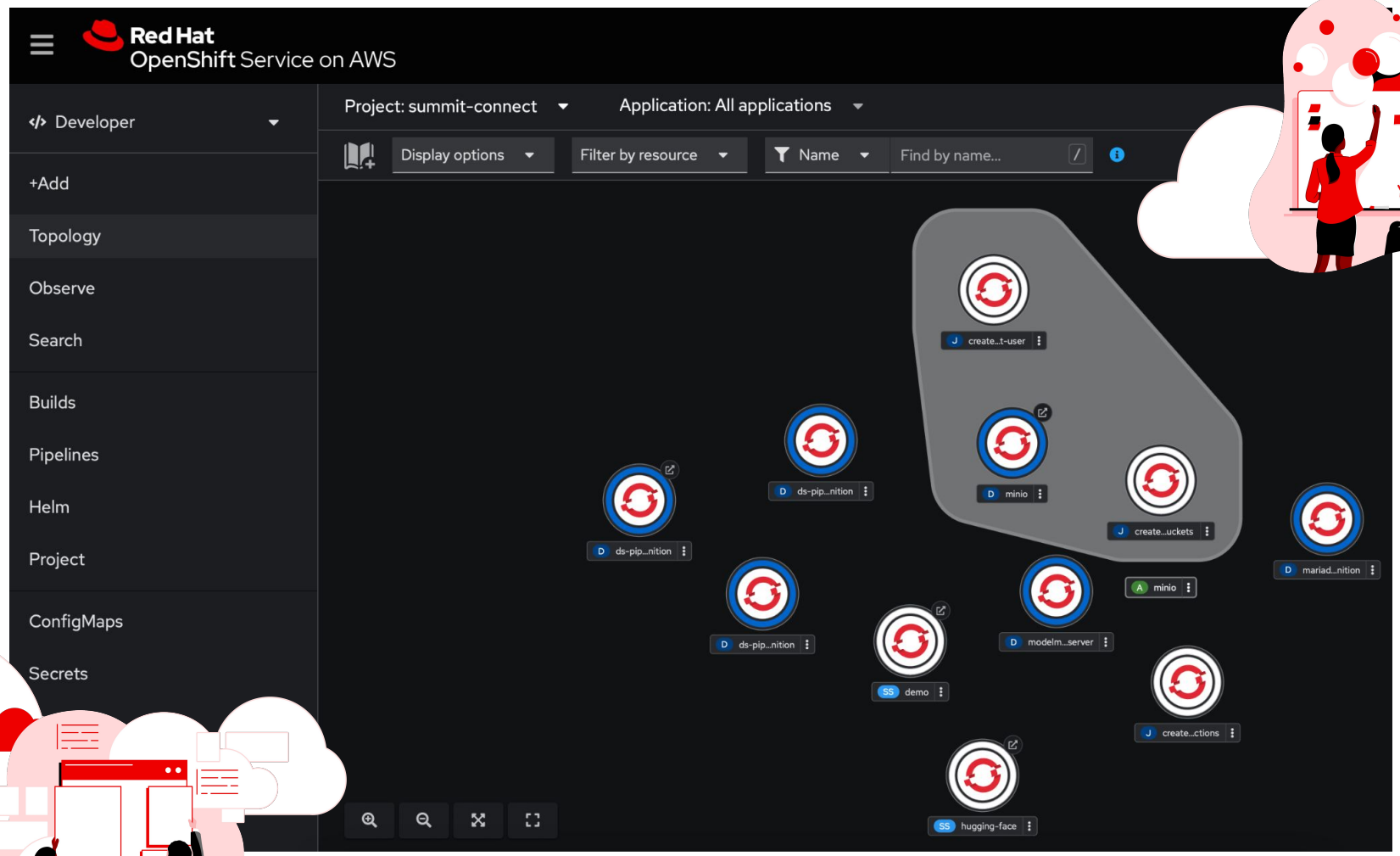
- Platform
- Development tools
- Storage
- OpenShift Data Science
- Security

Red Hat

- Disconnected operations across teams that create deployment delays.
- Increasing complexity of integrating AI solutions with in-place infrastructure.
- The need to continuously (re)train models, auditing requirements and monitoring of model performance.
- As generative AI has emerged, organizations are now facing new challenges related to pre-built models, including:
    - Licensing and ownership.
    - Regulatory and locality restrictions.

Red Hat

secure, supported & stable platform



Red Hat

secure, supported & stable platform

Start quickly, we manage it for you

Managed Red Hat OpenShift services

aws	Azure	IBM Cloud	Google Cloud
Red Hat OpenShift Service on AWS[1]	Azure Red Hat OpenShift	Red Hat OpenShift on IBM Cloud[1]	Red Hat OpenShift Dedicated[2]

You manage it, for control and flexibility

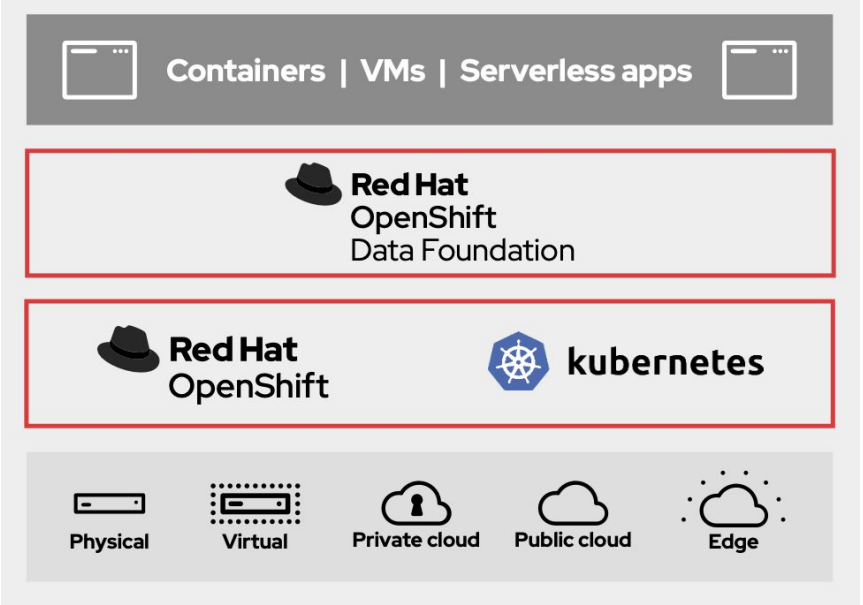
Self-managed Red Hat OpenShift

Red Hat OpenShift Platform Plus	**Red Hat** OpenShift Container Platform	**Red Hat** OpenShift Kubernetes Engine	On public cloud, or on-premises on physical or virtual infrastructure[3]

Red Hat

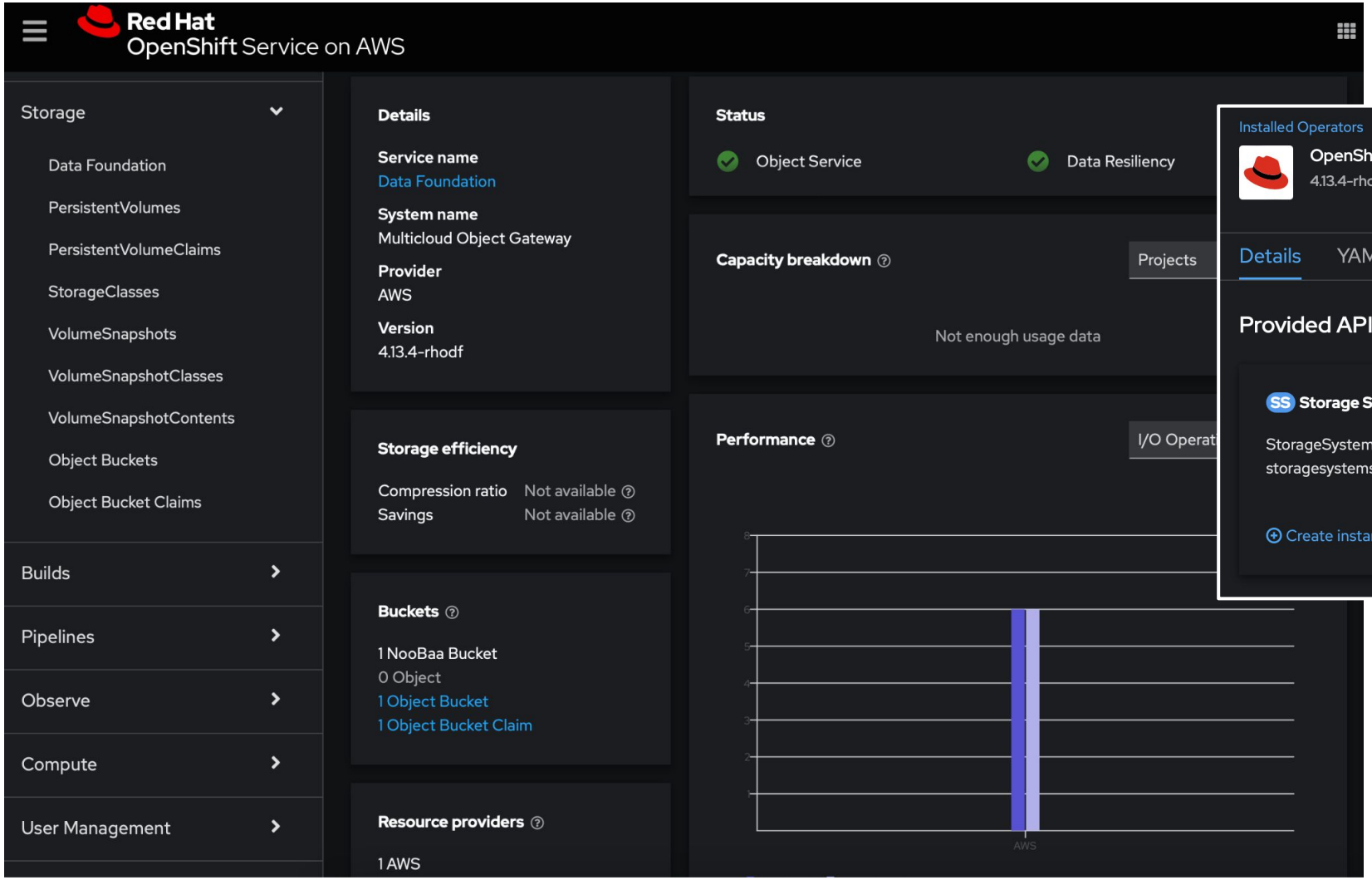
object storage everywhere

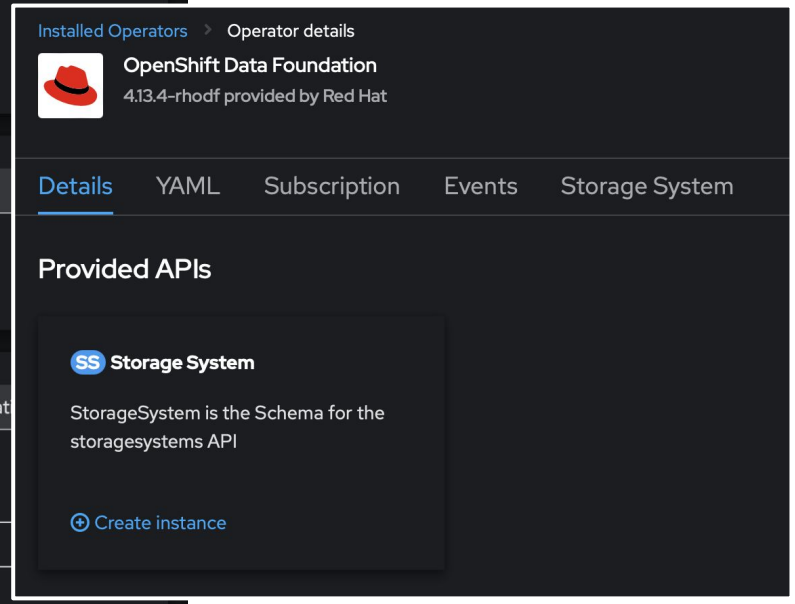
- OpenShift Data Foundation provides the functionality that organizations need to support applications and workloads in one complete solution—fully integrated and delivered with Red Hat OpenShift.
- Multicloud object gateway exposes a consistent Amazon S3 compatible endpoint to developers and applications.
- OpenShift Data Foundation clusters support end-to-end encryption including encryption with user-provided keys.



Red Hat

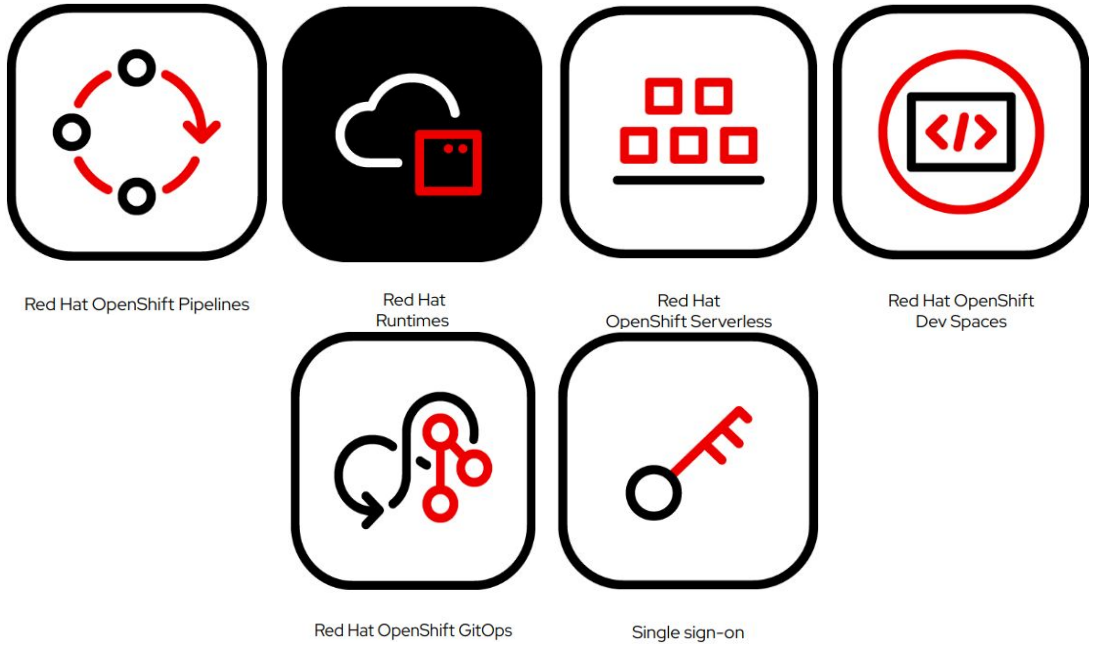
object storage everywhere



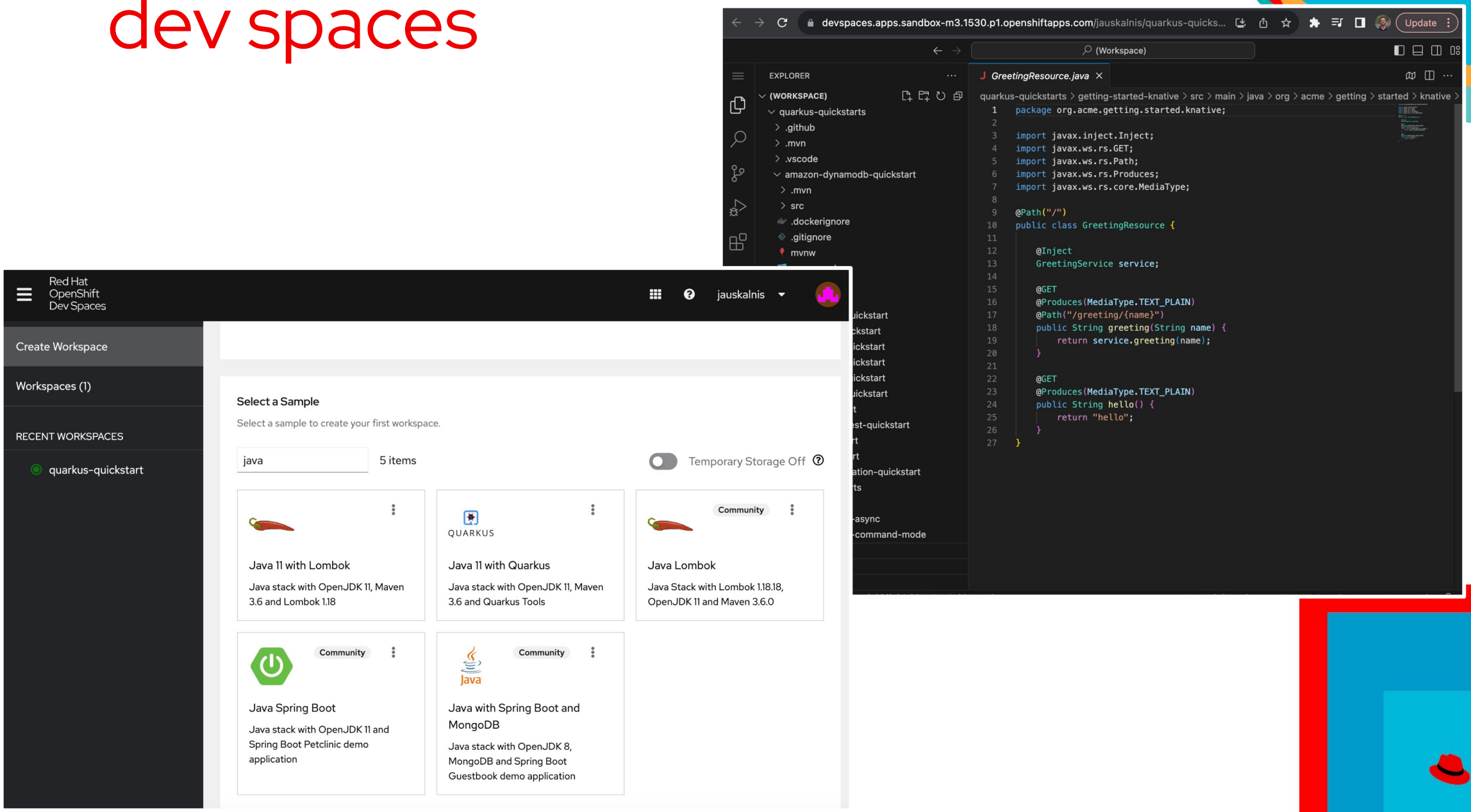


Red Hat

developer tools comes with OpenShift



dev spaces



Red Hat

what is ODS?

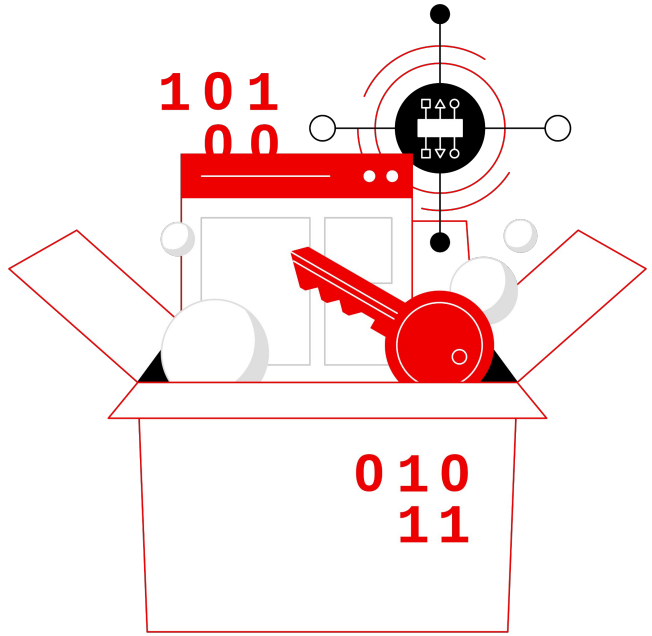
Red Hat OpenShift Data Science provides the ability to build, train, deploy and monitor models on premise, in the public cloud or at the edge.

Red Hat updates the platform and integrated AI tooling (like Jupyter Notebooks, PyTorch, and TensorFlow libraries).

Using the model serving and monitoring tools built into Red Hat OpenShift Data Science, models are container-ready, which makes it easier to integrate them into an intelligent app.

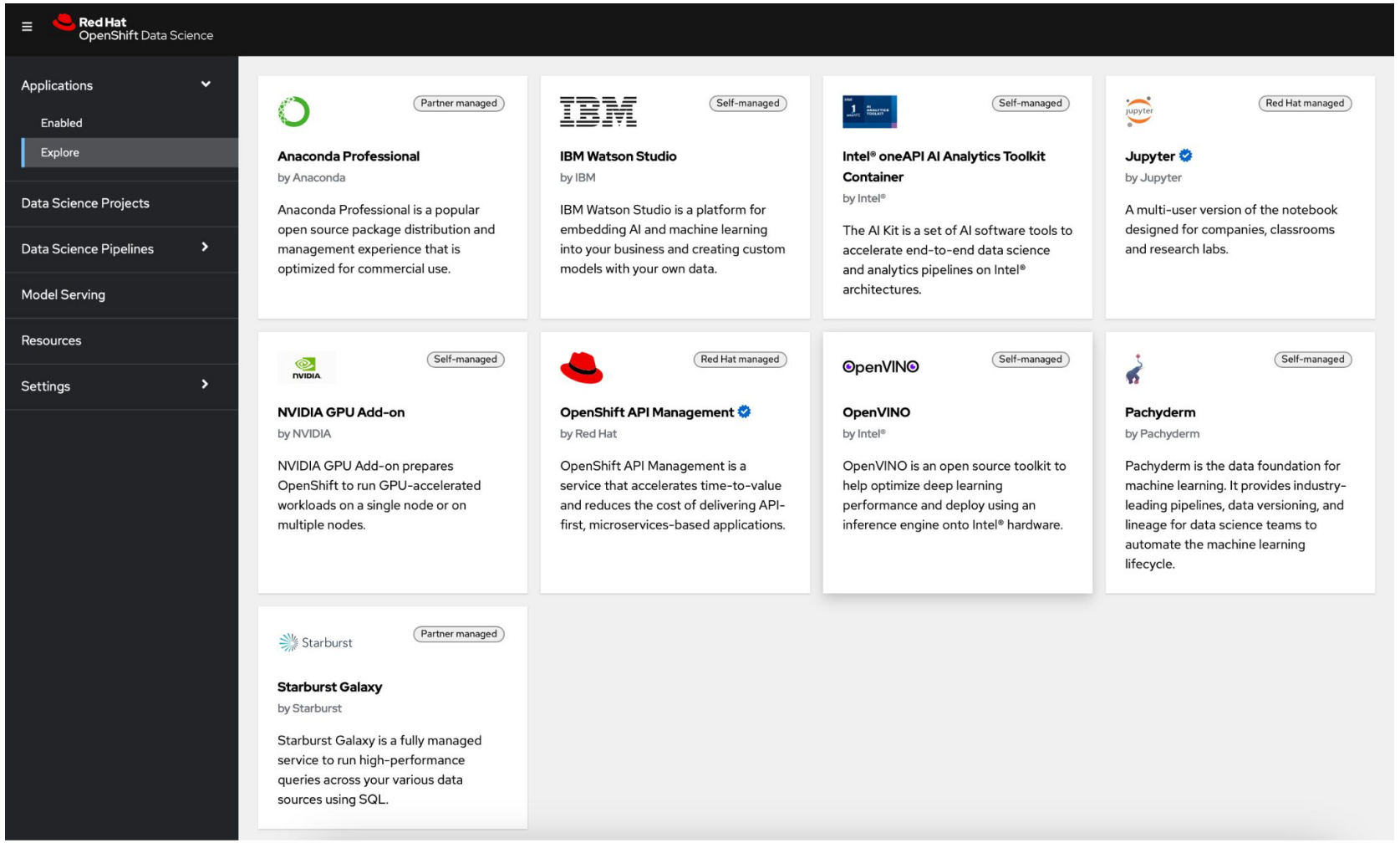
Models can be rebuilt and redeployed as part of a continuous integration/continuous development process based on changes to the source notebook.

Nvidia and Intel AI tools unlock high-performance training.



Red Hat

what is inside ODS?



Red Hat

how we use ODS?

Red Hat OpenShift Data Science — juozas

- Applications
- Data Science Projects
- Data Science Pipelines
- Model Serving
- Resources
- Settings

Data science projects > Summit Connect

Summit Connect

Components | Permissions

Jump to section: Workbenches, Cluster storage, Data connections, Pipelines, Models and model servers

Workbenches

Create workbench

Name	Notebook image	Container size	Status	
Demo	TensorFlow	Small	Stopped	Open
Hugging Face	PyTorch	Medium	Running	Open

Cluster storage

Add cluster storage

Name	Type	Connected workbenches
Demo	Persistent storage	Demo
Hugging Face	Persistent storage	Hugging Face

Data connections

Add data connection

Name	Type	Connected workbenches	Provider
My Storage	Object storage	Hugging Face	AWS S3
Pipeline Artifacts	Object storage	No connections	AWS S3

Pipelines

Import pipeline

Pipeline name	Last run	Last run status	Last run time	Created
Hugging Face Created with Elyra 3.15.0 pipeline editor using `untitled2.pipeline`.	Hugging Face-1104094346	Running	-	3 minutes ago
Runs Hugging Face-1104094346		Running	-	3 minutes ago

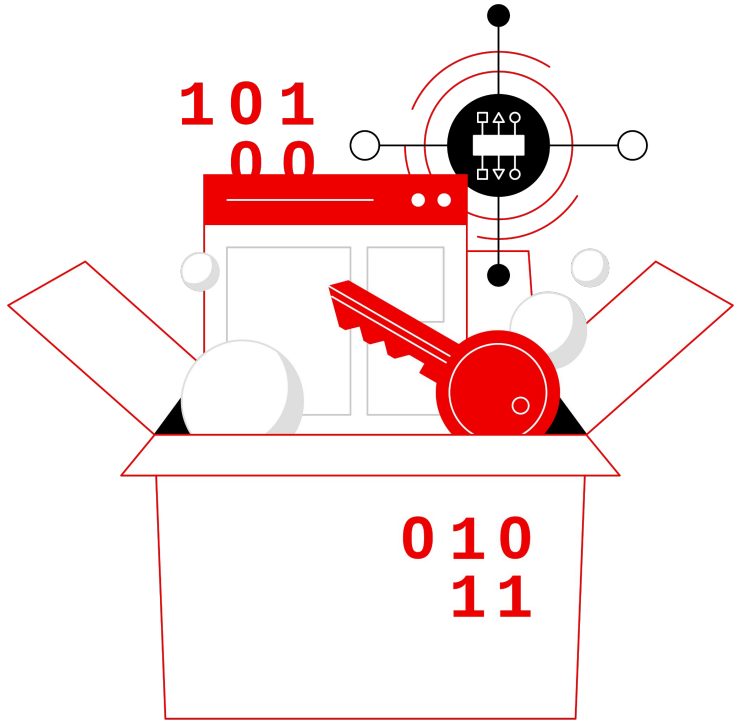
Models and model servers

Add server

Model Server Name	Serving Runtime	Deployed models	Tokens	
Custom Model Server	Triton runtime 23.01 - added on 20230815	4	Tokens disabled	Deploy model

Red Hat

workbenches



Notebook image

- ○ Minimal Python 2023.1 ⓘ
  Python v3.9
  › Versions
- ● CUDA 2023.1 ⓘ
  CUDA v11.8, Python v3.9
  › Versions
- ○ TensorFlow 2023.1 ⓘ
  CUDA v11.8, Python v3.9, TensorFlow v2.11
  › Versions
- ○ Standard Data Science 2023.1 ⓘ
  Python v3.9
  › Versions
- ○ PyTorch 2023.1 ⓘ
  CUDA v11.8, Python v3.9, PyTorch v1.13
  › Versions
- ○ TrustyAI ⓘ
  Python v3.9

Red Hat

storage and data connections

Red Hat OpenShift Data Science

juozas

Applications
Data Science Projects
Data Science Pipelines
Model Serving
Resources
Settings

Data science projects > Summit Connect

Summit Connect

Components | Permissions

Jump to section
- Workbenches
- Cluster storage
- Data connections
- Pipelines
- Models and model servers

Workbenches

Create workbench

Name	Notebook image	Container size	Status	
Demo	TensorFlow	Small	Stopped	Open
Hugging Face	PyTorch	Medium	Running	Open

Cluster storage

Add cluster storage

Name	Type	Connected workbenches
Demo	Persistent storage	Demo
Hugging Face	Persistent storage	Hugging Face

Data connections

Add data connection

Name	Type	Connected workbenches	Provider
My Storage	Object storage	Hugging Face	AWS S3
Pipeline Artifacts	Object storage	No connections	AWS S3

Pipelines

Import pipeline

Pipeline name	Last run	Last run status	Last run time	Created
Hugging Face Created with Elyra 3.15.0 pipeline editor using `untitled2.pipeline`.	Hugging Face-1104094346	Running	-	3 minutes ago
Runs Hugging Face-1104094346		Running	-	3 minutes ago

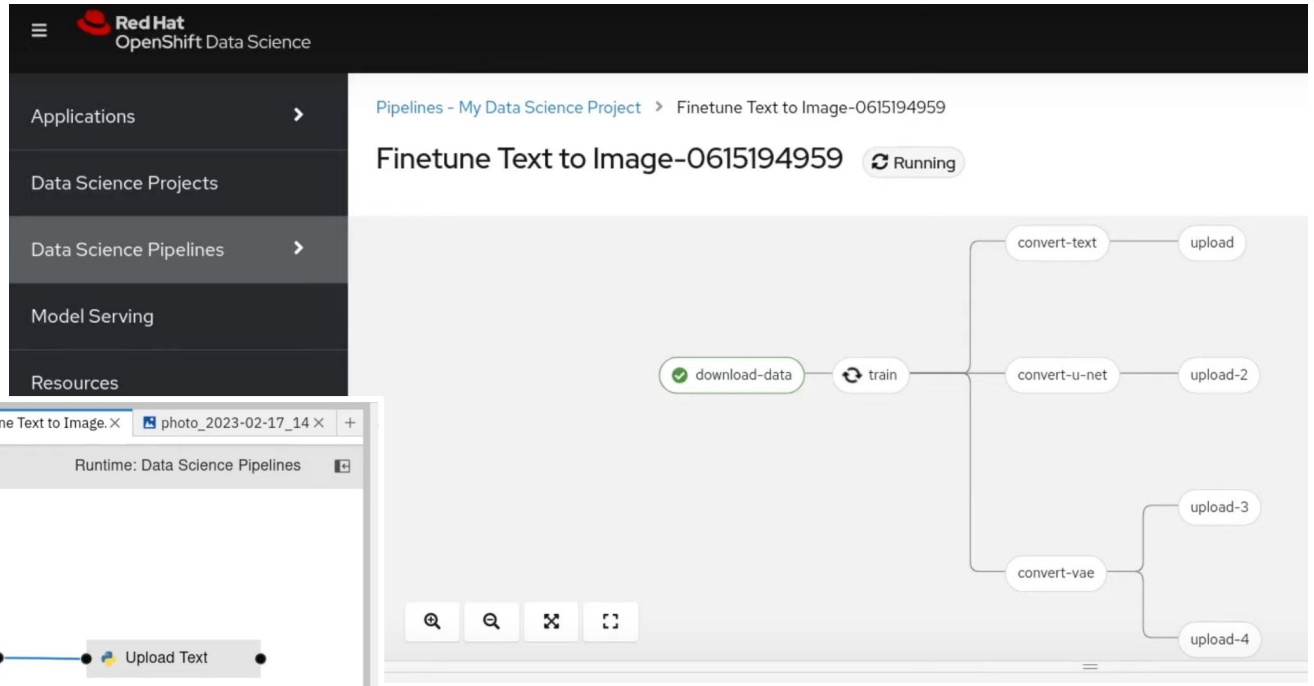
Models and model servers

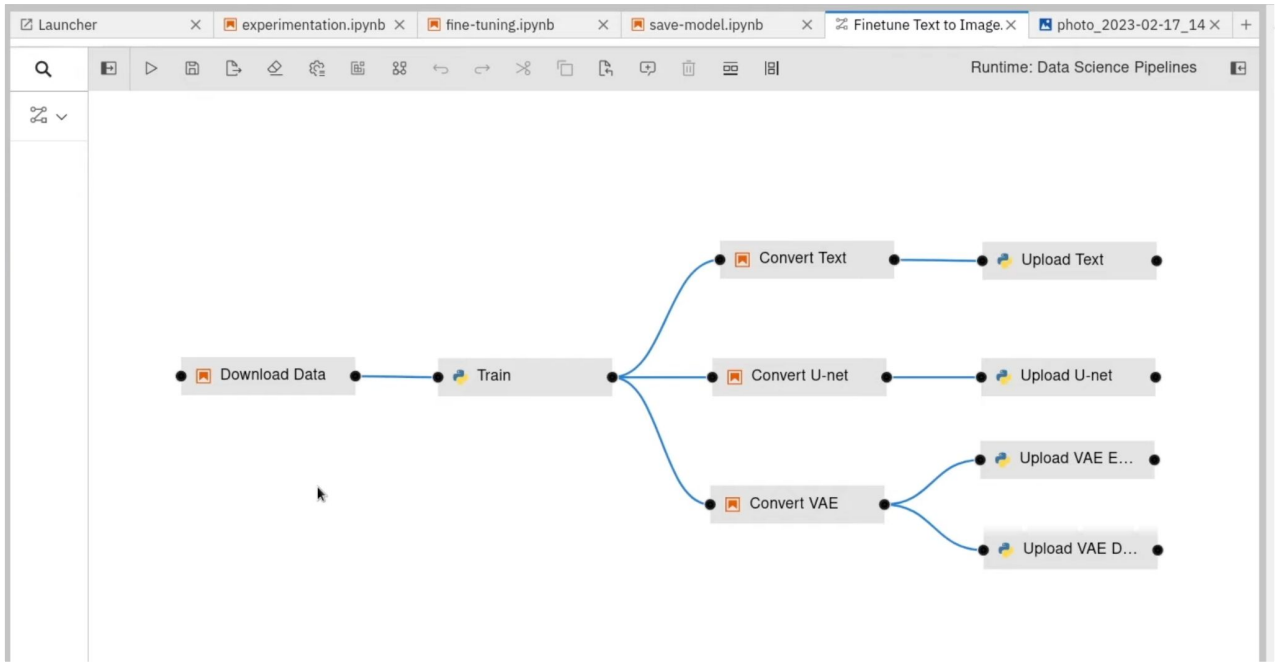
Add server

Model Server Name	Serving Runtime	Deployed models	Tokens	
Custom Model Server	Triton runtime 23.01 - added on 20230815	4	Tokens disabled	Deploy model

Red Hat

pipelines





Red Hat

model serving

Red Hat OpenShift Data Science

juozas

Applications
Data Science Projects
Data Science Pipelines
Model Serving
Resources
Settings

Data science projects > Summit Connect

Summit Connect

Components | Permissions

Jump to section
- Workbenches
- Cluster storage
- Data connections
- Pipelines
- Models and model servers

Workbenches

Create workbench

Name	Notebook image	Container size	Status	
Demo	TensorFlow	Small	Stopped	Open
Hugging Face	PyTorch	Medium	Running	Open

Cluster storage

Add cluster storage

Name	Type	Connected workbenches
Demo	Persistent storage	Demo
Hugging Face	Persistent storage	Hugging Face

Data connections

Add data connection

Name	Type	Connected workbenches	Provider
My Storage	Object storage	Hugging Face	AWS S3
Pipeline Artifacts	Object storage	No connections	AWS S3

Pipelines

Import pipeline

Pipeline name	Last run	Last run status	Last run time	Created
Hugging Face Created with Elyra 3.15.0 pipeline editor using `untitled2.pipeline`.	Hugging Face-1104094346	Running	-	3 minutes ago
Runs Hugging Face-1104094346		Running	-	3 minutes ago

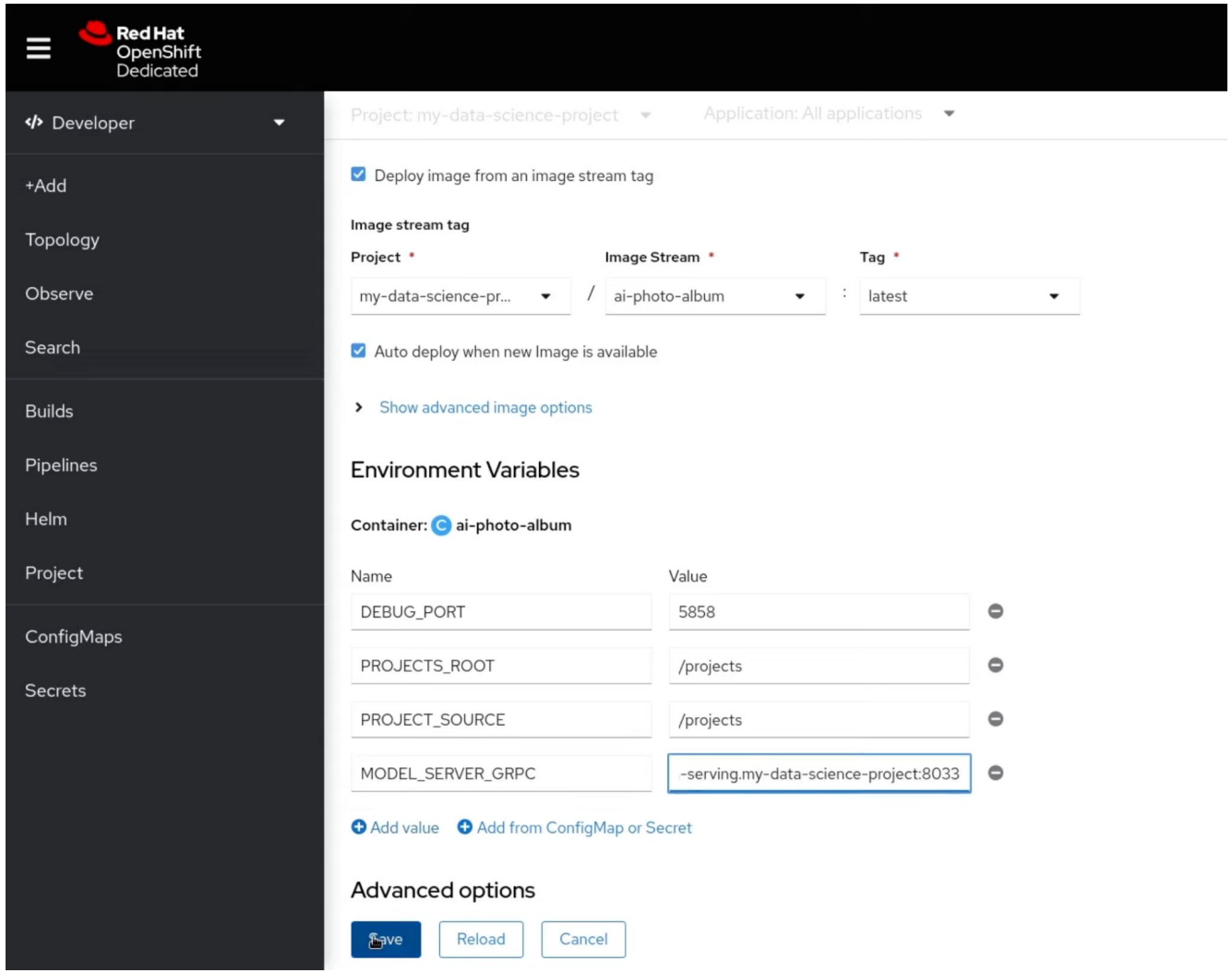
Models and model servers

Add server

Model Server Name	Serving Runtime	Deployed models	Tokens	
Custom Model Server	Triton runtime 23.01 - added on 20230815	4	Tokens disabled	Deploy model

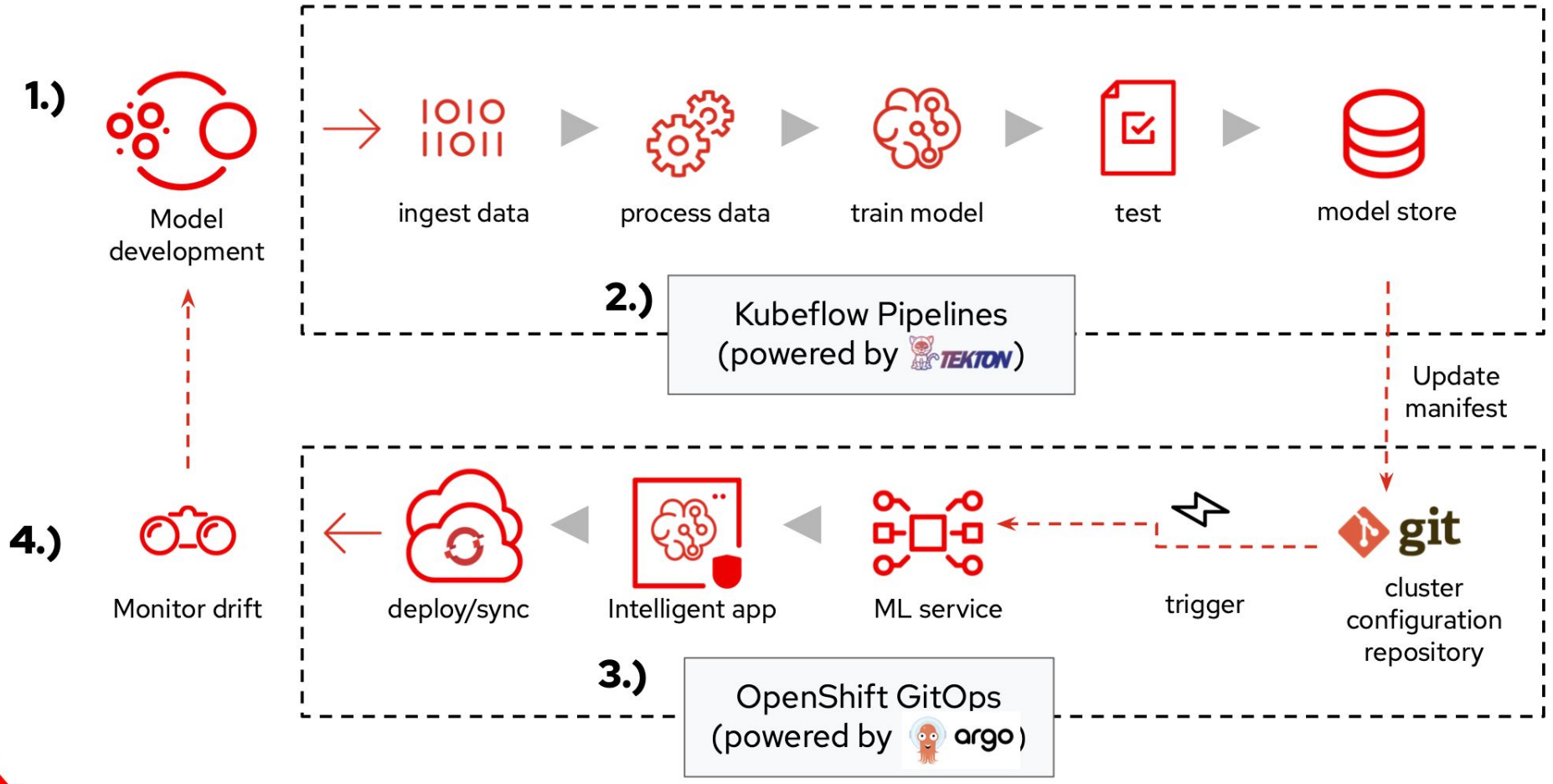
Red Hat

model serving



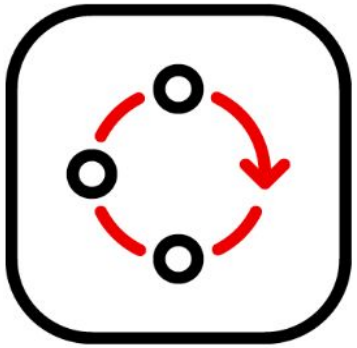
Red Hat

lifecycle



Red Hat

Red Hat
Advanced Cluster
Management
for Kubernetes



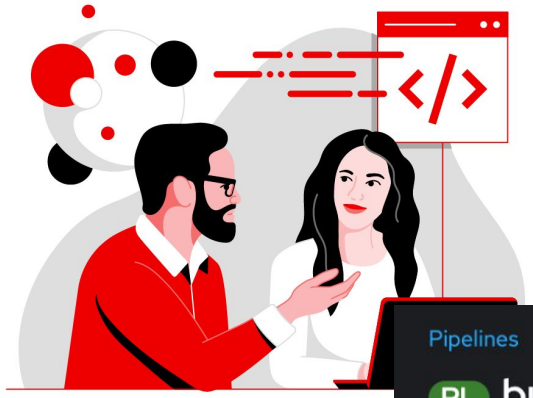
Red Hat OpenShift Pipelines

DevSecOps?

Red Hat OpenShift GitOps

Red Hat
Advanced Cluster
Security
for Kubernetes

Red Hat



Pipelines > Pipeline details

PL build-and-deploy

Actions

Details | Metrics | YAML | PipelineRuns | Parameters | Resources

Pipeline details

Red Hat

Summary

Red Hat OpenShift AI is an AI-focused portfolio that provides tools across the full lifecycle of AI/ML experiments and models and includes Red Hat OpenShift Data Science.

Red Hat OpenShift Data Science is an open source machine learning (ML) platform for the hybrid cloud.

By providing a fully supported environment to establish MLOps best practices, data scientists and developers can rapidly train, deploy and monitor ML workloads and models on-premise and in the public cloud.

https://developers.redhat.com/learn/openshift-data-science

Red Hat

Red Hat Summit

Thank you


linkedin.com/company/red-hat


facebook.com/redhatinc


youtube.com/user/RedHatVideos


twitter.com/RedHat

Red Hat