Data Science Services – the simple way

Red Hat OpenShift Data Science introduction for the rest of us

Stefan Bergstein
Principal Chief Architect
+49 172 570 3774
stefan.bergstein@redhat.com

Red Hat

What we'll discuss today

- What is MLOps?
- ML Model Development
- ML Model Serving
- Offer a MLOps service to your Data Science teams
- Wrap-up and Q&A

What is MLOps?

This section includes:

What is MLOps:

- The AI Journey
- Typical roles
- MLOps with open source

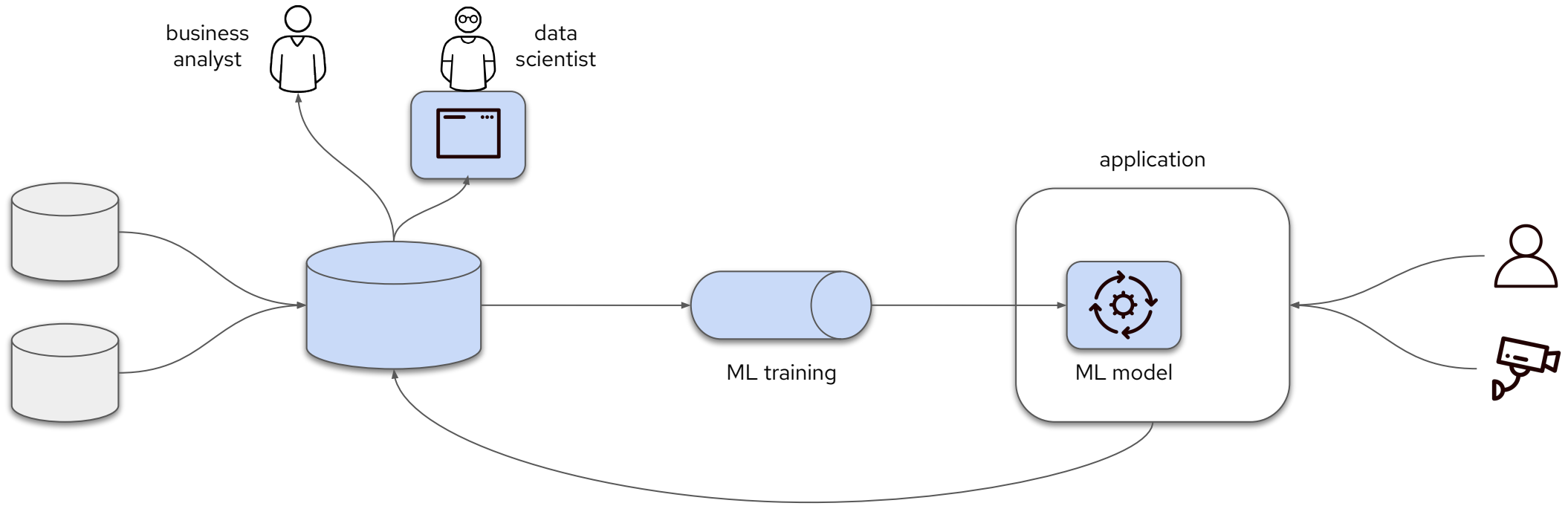
The AI Journey

Business Intelligence

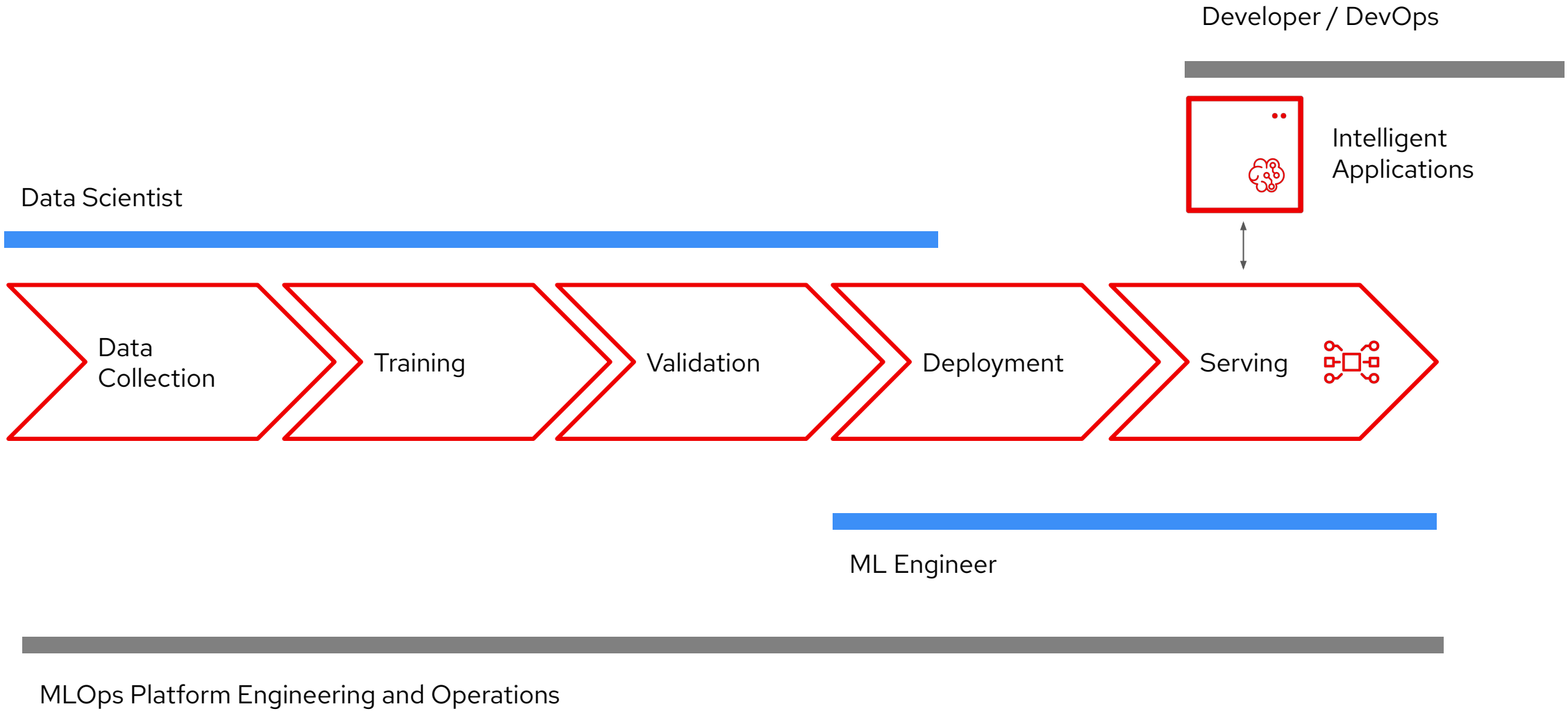
Data Science

Machine Learning

Intelligent Applications

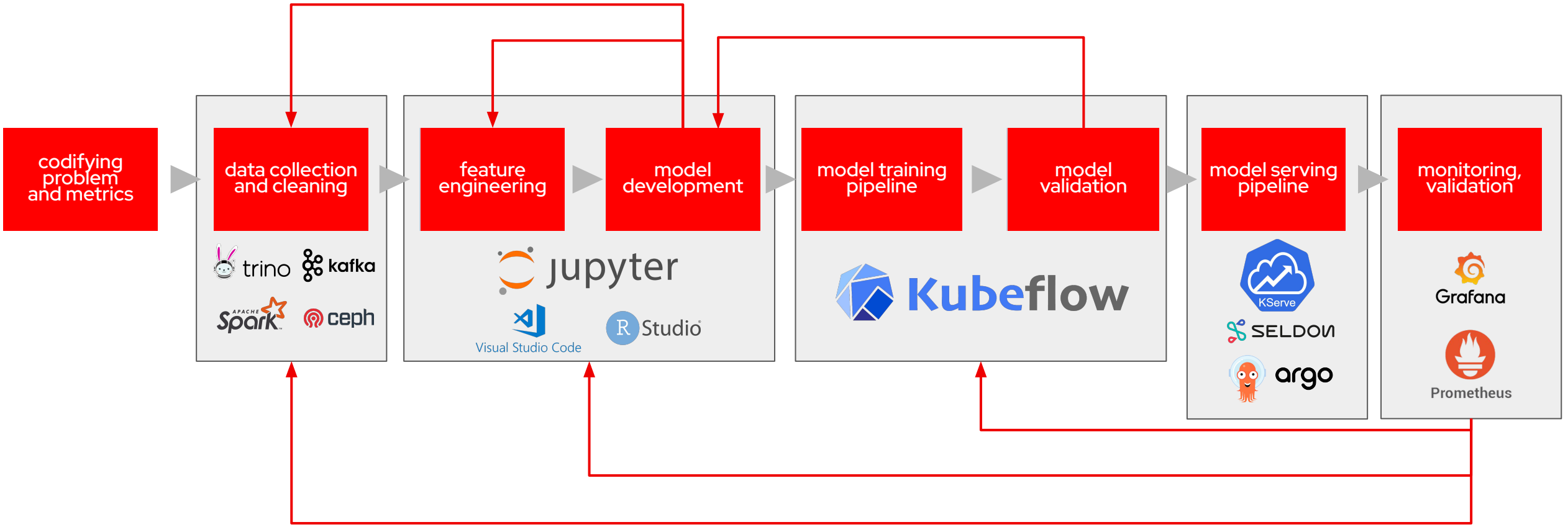


Arrows denote direction of data flow.



Simplified view on the roles and responsibilities

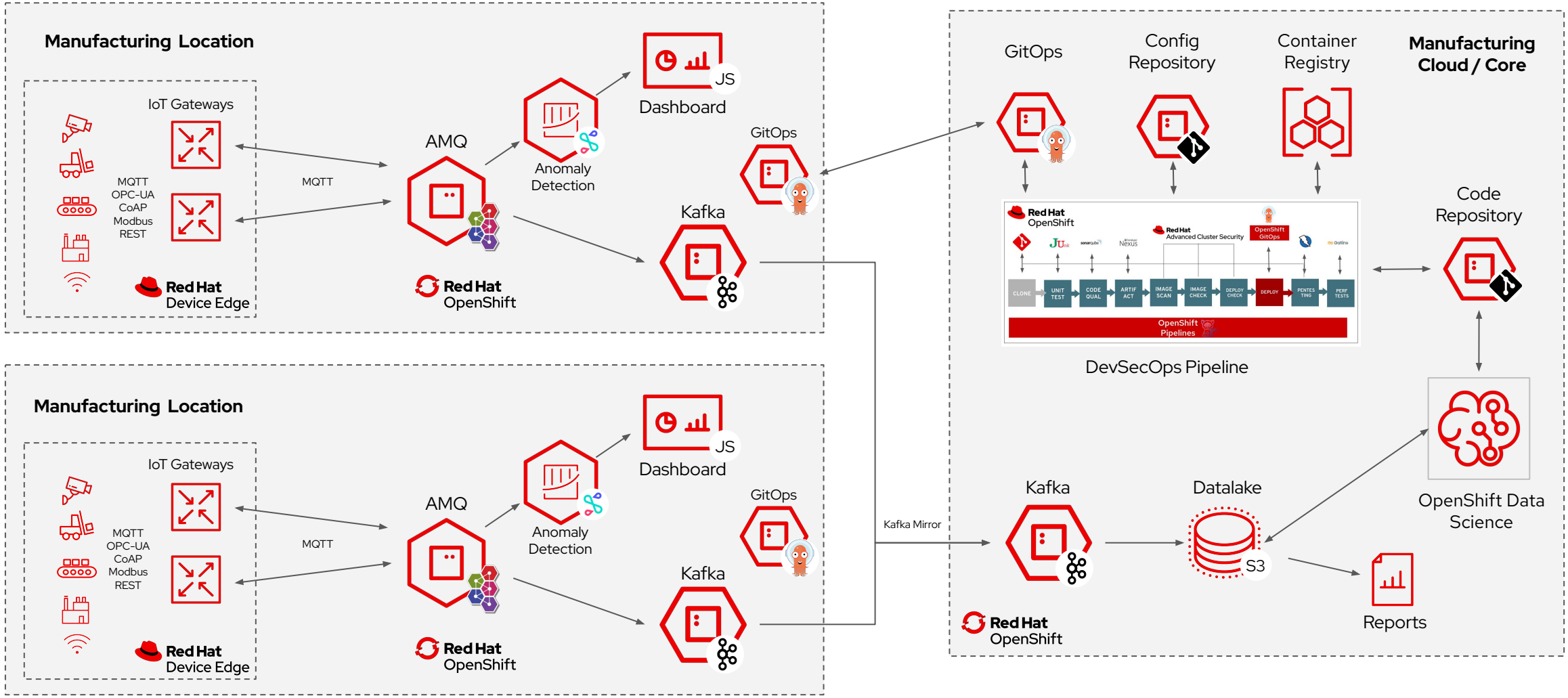
MLOps with open source



Red Hat

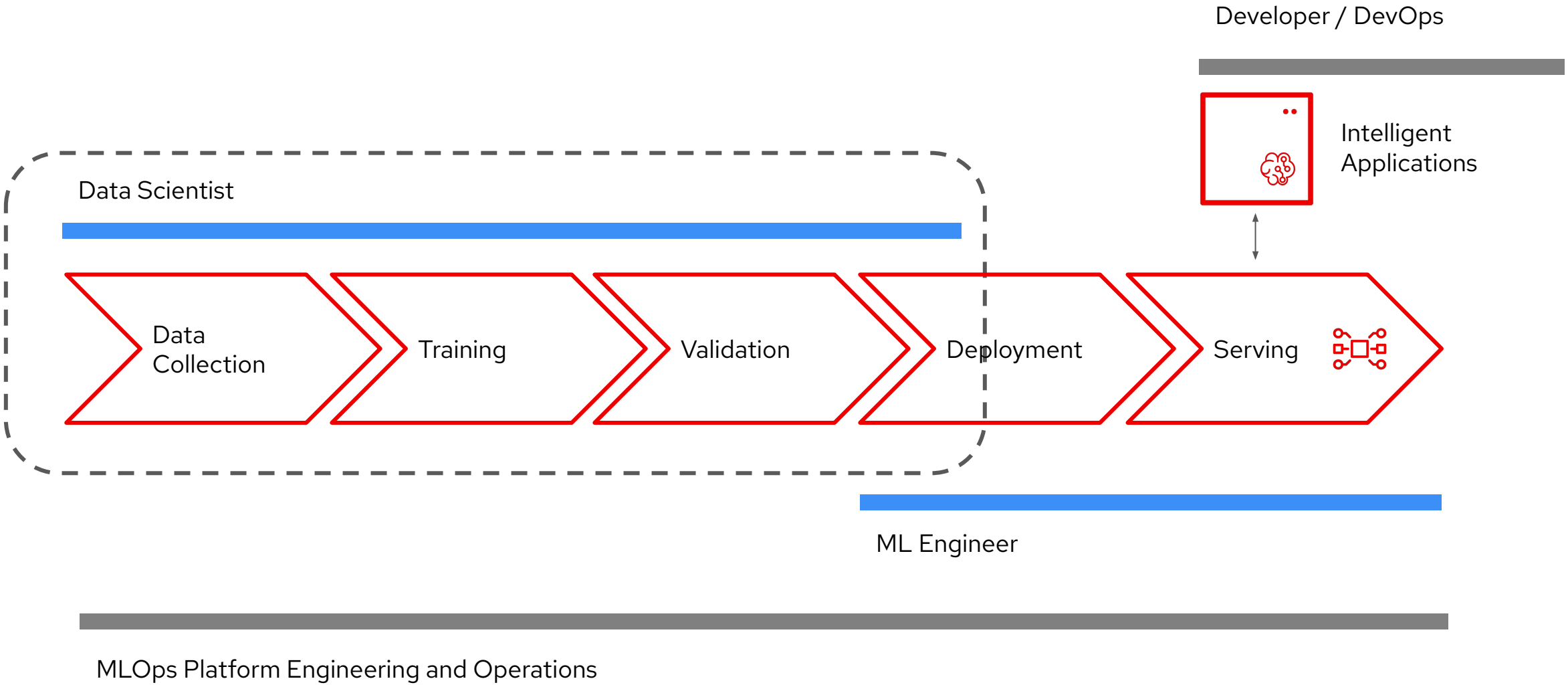
Example for ML based condition monitoring on manufacturing production lines

High level architecture



Red Hat

ML Model Development

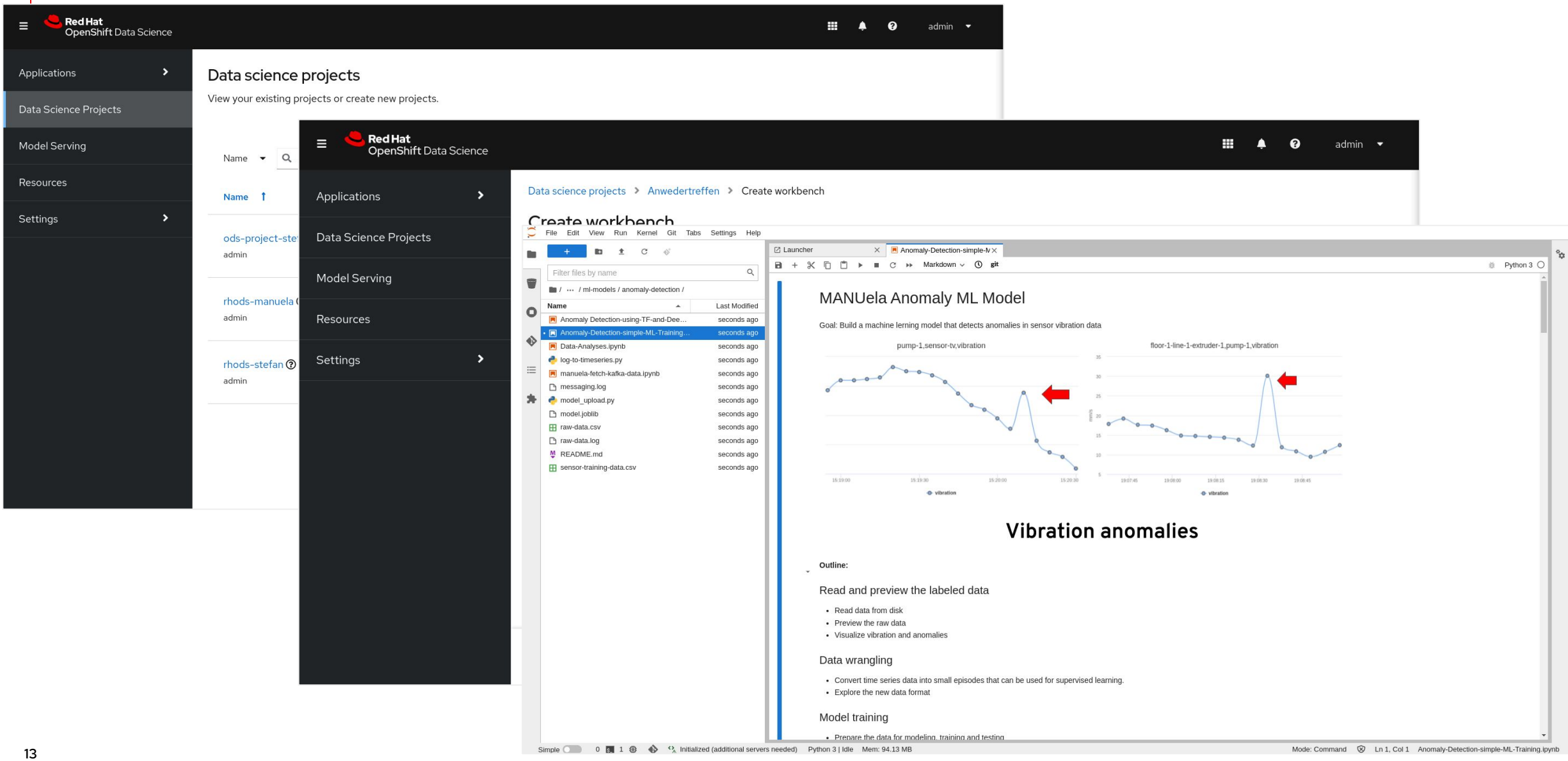

Developer / DevOps
Intelligent Applications
Data Scientist
Data Collection
Training
Validation
Deployment
Serving
ML Engineer
MLOps Platform Engineering and Operations

This section includes:

ML Model Development

- Data Scientist's Workbench
- Model Training
- ML Pipelines

Demo


Red Hat OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
View your existing projects or create new projects.
Name
Name
ods-project-ste
admin
rhods-manuela
admin
rhods-stefan
admin
Red Hat OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects > Anwedertreffen > Create workbench
Create workbench
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ ml-models / anomaly-detection /
Name
Last Modified
Anomaly Detection-using-TF-and-Dee… seconds ago
Anomaly-Detection-simple-ML-Training… seconds ago
Data-Analyses.ipynb seconds ago
log-to-timeseries.py seconds ago
manuela-fetch-kafka-data.ipynb seconds ago
messaging.log seconds ago
model_upload.py seconds ago
model.joblib seconds ago
raw-data.csv seconds ago
raw-data.log seconds ago
README.md seconds ago
sensor-training-data.csv seconds ago
Launcher
Anomaly-Detection-simple-M
Markdown
git
Python 3
MANUela Anomaly ML Model
Goal: Build a machine lerning model that detects anomalies in sensor vibration data
pump-1,sensor-tv,vibration
floor-1-line-1-extruder-1,pump-1,vibration
vibration
vibration
Vibration anomalies
Outline:
Read and preview the labeled data
Read data from disk
Preview the raw data
Visualize vibration and anomalies
Data wrangling
Convert time series data into small episodes that can be used for supervised learning.
Explore the new data format
Model training
Prepare the data for modeling, training and testing
Simple
Initialized (additional servers needed)
Python 3 | Idle
Mem: 94.13 MB
Mode: Command
Ln 1, Col 1
Anomaly-Detection-simple-ML-Training.ipynb

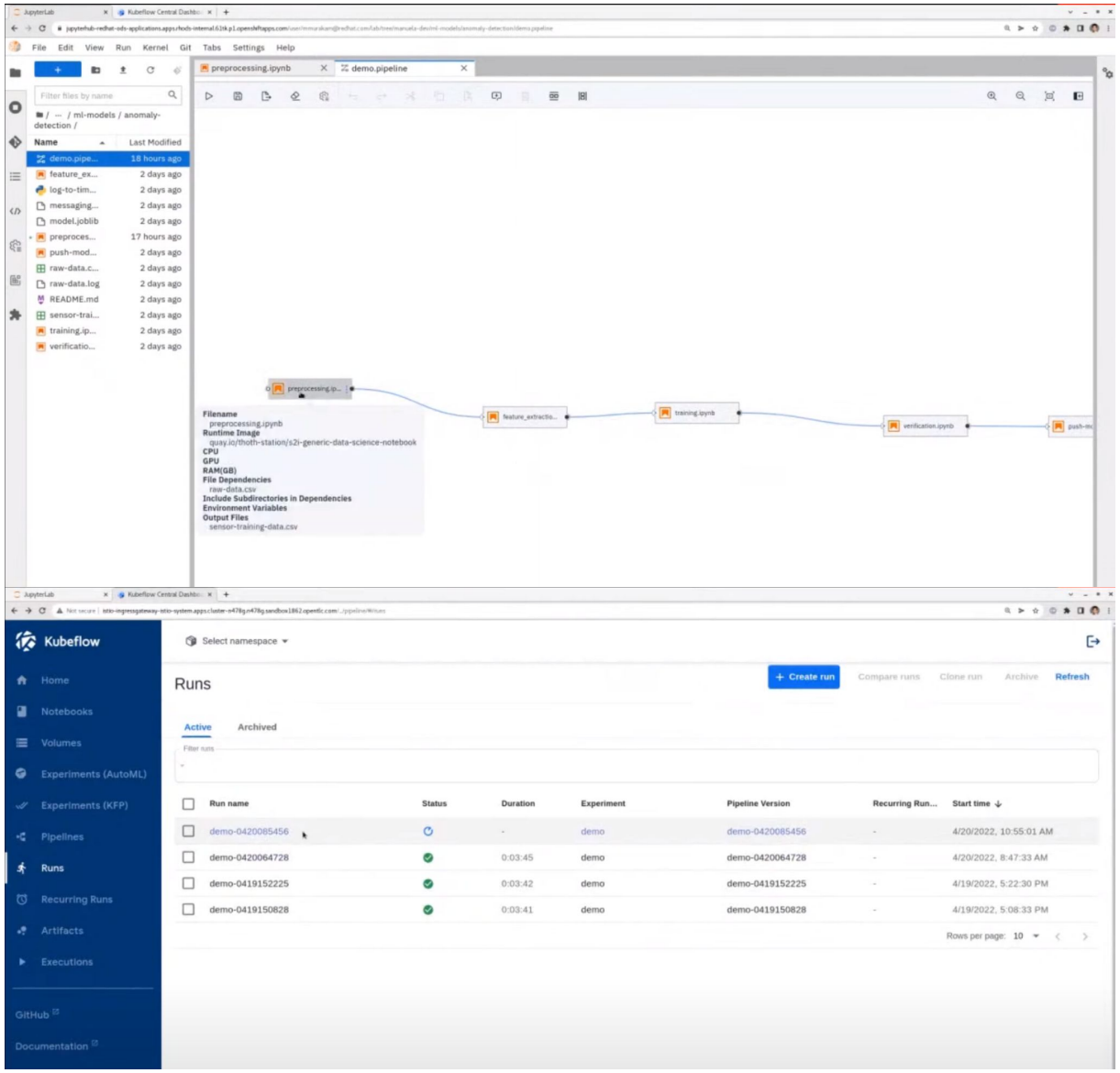
Red Hat

Elyra pipelines

- Build, run, and share machine learning pipelines
- Built on top of JupyterLab
- Use Git to manage the versions of your pipelines

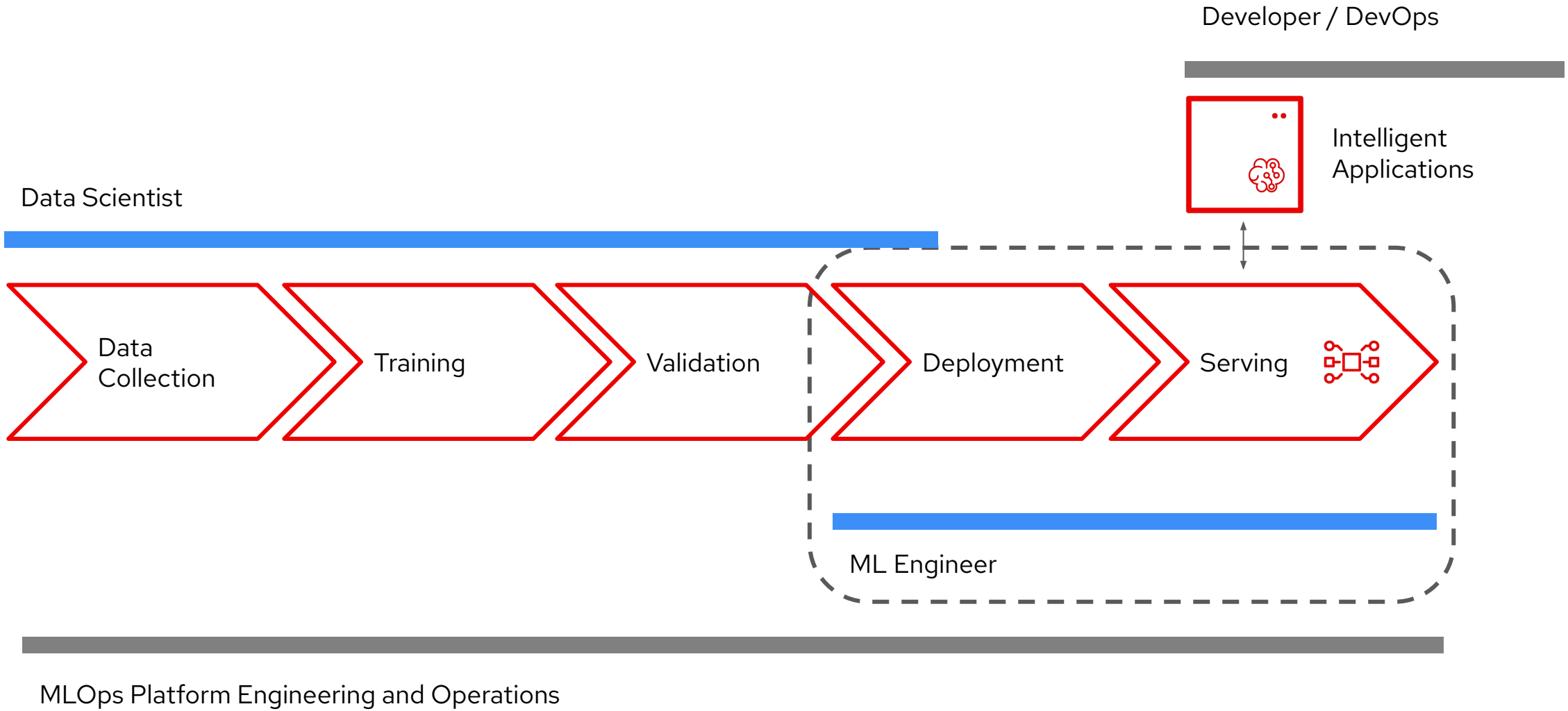
Kubeflow Pipelines

- Build, deploy, and manage end-to-end machine learning workflows
- Execution engine for Elyra pipelines
- Based on Tekton



Red Hat

ML Model Serving

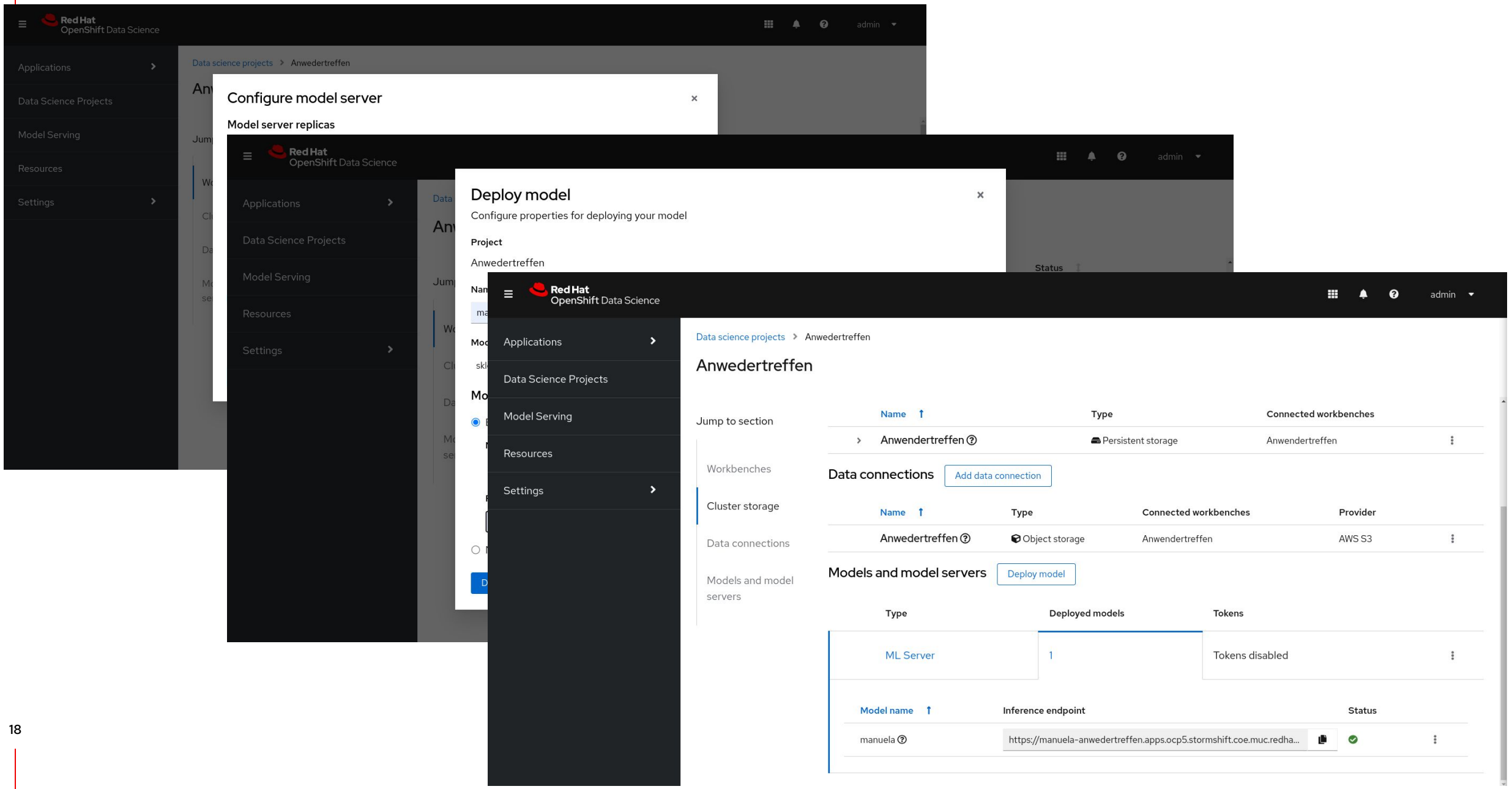

Developer / DevOps
Intelligent Applications
Data Scientist
Data Collection
Training
Validation
Deployment
Serving
ML Engineer
MLOps Platform Engineering and Operations

This section includes:

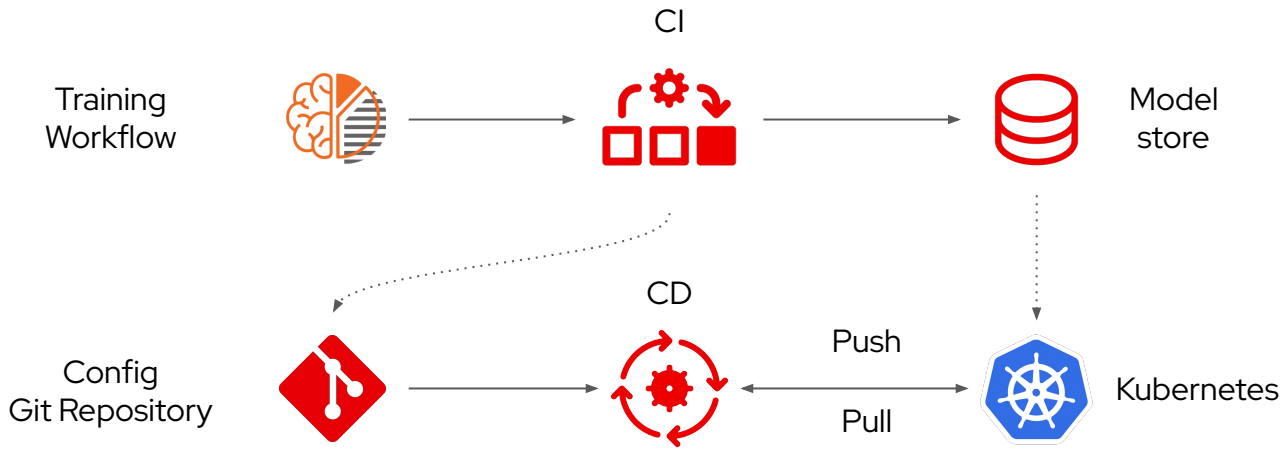
ML Model Serving

- Model Serving with Red Hat OpenShift Data Science
- Utilize the model in the manufacturing example

Demo

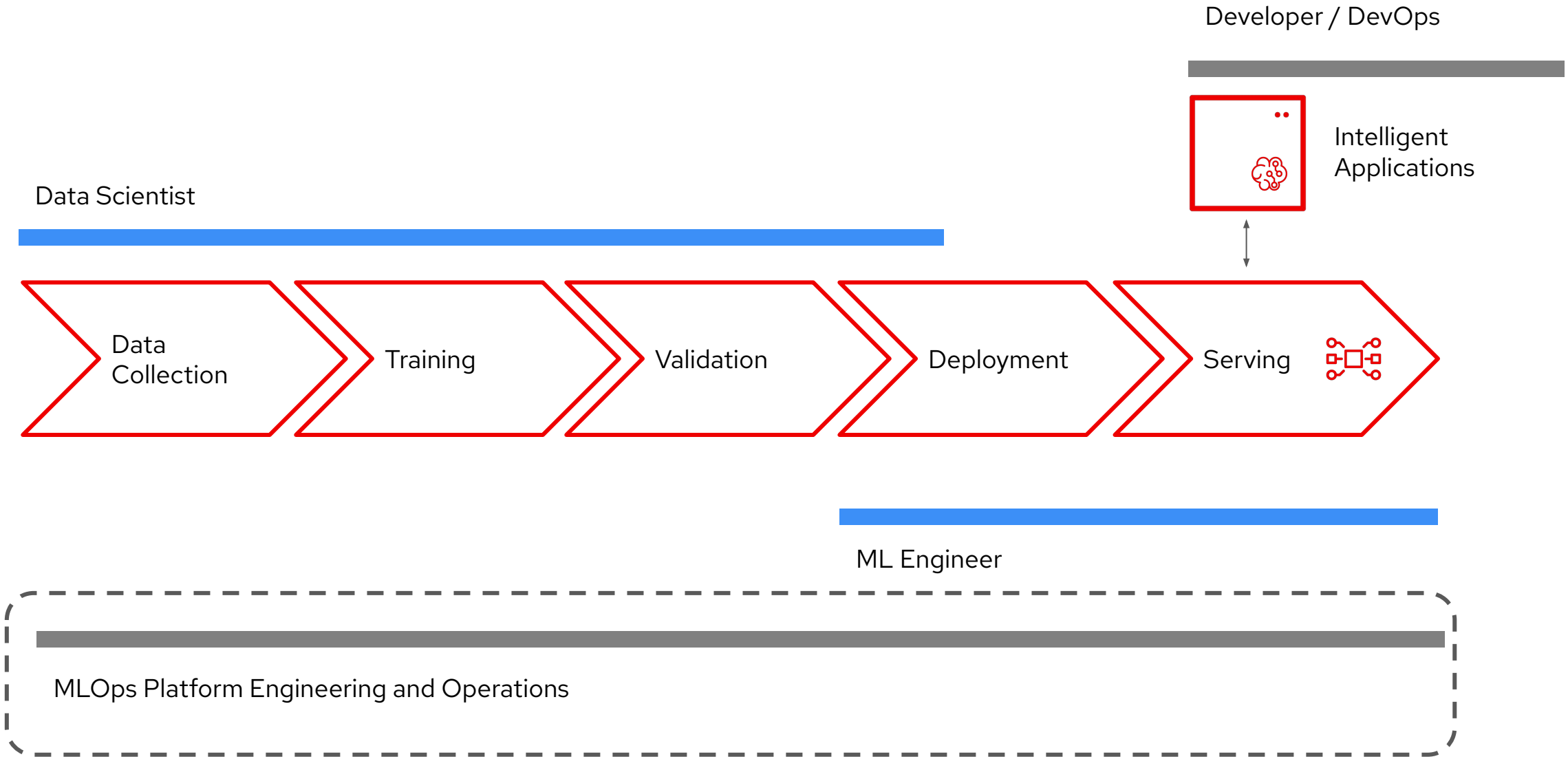

Red Hat OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
Anwedertreffen
Configure model server
Model server replicas
Deploy model
Configure properties for deploying your model
Project
Anwedertreffen
Jump to section
Workbenches
Cluster storage
Data connections
Models and model servers
Name
Type
Connected workbenches
Persistent storage
Data connections
Add data connection
Provider
Object storage
AWS S3
Models and model servers
Deploy model
Deployed models
Tokens
ML Server
1
Tokens disabled
Model name
Inference endpoint
Status
manuela
https://manuela-anwedertreffen.apps.ocp5.stormshift.coe.muc.redha...

Delivering models through GitOps



Red Hat

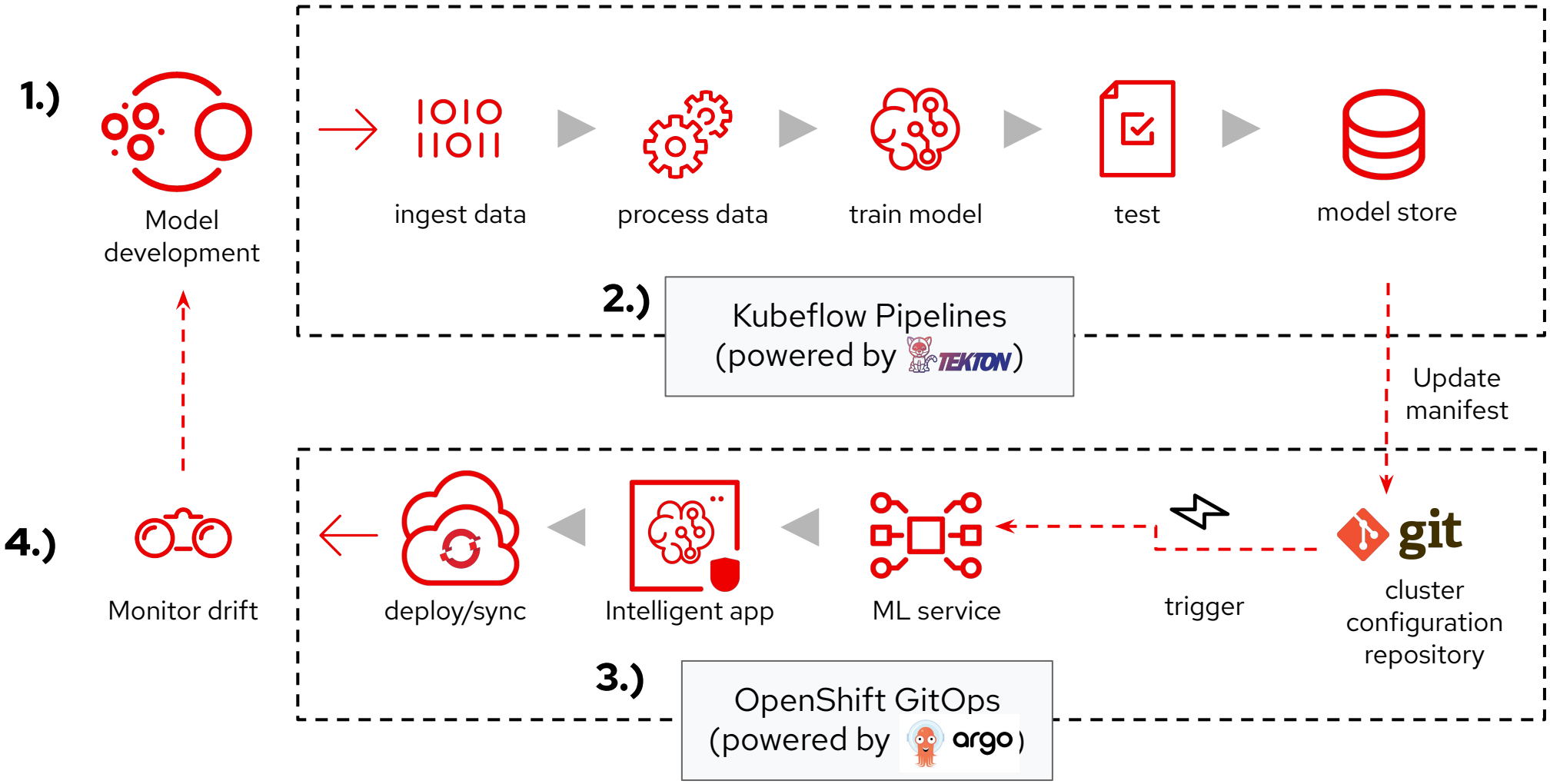
Offer a MLOps service to your Data Science teams


Developer / DevOps
Intelligent Applications
Data Scientist
Data Collection
Training
Validation
Deployment
Serving
ML Engineer
MLOps Platform Engineering and Operations

This section includes:

- Offer a MLOps service to your Data Science teams
  - MLOps with Red Hat OpenShift
  - Open Data Hub
  - Cloud service and self-managed components
  - Warp-up

MLOps with Red Hat OpenShift



Open Data Hub

An open source MLOps suite

- Multi-tenant data science platform
- Self-service workspaces

- Preinstalled machine learning libraries
- Custom stack can be integrated

Elyra

- AI pipeline editor
- Define workflows through Jupyter

Kubeflow Pipelines

- Machine learning workflow orchestration
- Experiment tracking

Kserve ModelMesh

- Deploying machine learning models as micro-services
- Pre-built inference servers

- Monitoring stack
- Native integration with OpenShift

Red Hat

Bringing AI/ML open source to the enterprise

Upstream Projects

Community Project

Product

Red Hat

Cloud service and self-managed components

	Gather and prepare data	Develop model	Integrate models in app dev	Model monitoring and management
Customer managed ISV software	Pachyderm	Intel AI Analytics Toolkit	IBM Watson	intel OpenVINO
ISV managed cloud services	Starburst Galaxy	ANACONDA		
Red Hat managed cloud services	Red Hat OpenShift Streams for Apache Kafka	Red Hat OpenShift Data Science: TensorFlow, jupyter, PyTorch, model serving	Source to image	Red Hat OpenShift API Management
Red Hat managed cloud platform	Open hybrid cloud platform with self service capabilities	Red Hat OpenShift Container Platform	Red Hat OpenShift Dedicated	Red Hat OpenShift Service on Amazon Web Services
	Accelerators	NVIDIA	intel	
	Cloud infrastructure	aws	Google Cloud	

Red Hat

Red Hat OpenShift Data Science

Hybrid MLOps platform

Collaborate within a common platform to bring IT, data science, and app dev teams together

Now GA as fully managed cloud service and self managed on-prem

Model development

Conduct exploratory data science in JupyterLab with access to core AI / ML libraries and frameworks including TensorFlow and PyTorch using our notebook images or your own.

Model serving & monitoring

Deploy models across any cloud, fully managed, and self-managed OpenShift footprint and centrally monitor their performance.

Hardware acceleration

Accelerate your data science experiments through the use of CPU and GPU acceleration on Red Hat OpenShift.

Partner ecosystem

Optional integrated ISVs to complement core platform capabilities with access to the full partner ecosystem with over 30 AI/ML technology partners.

Red Hat

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

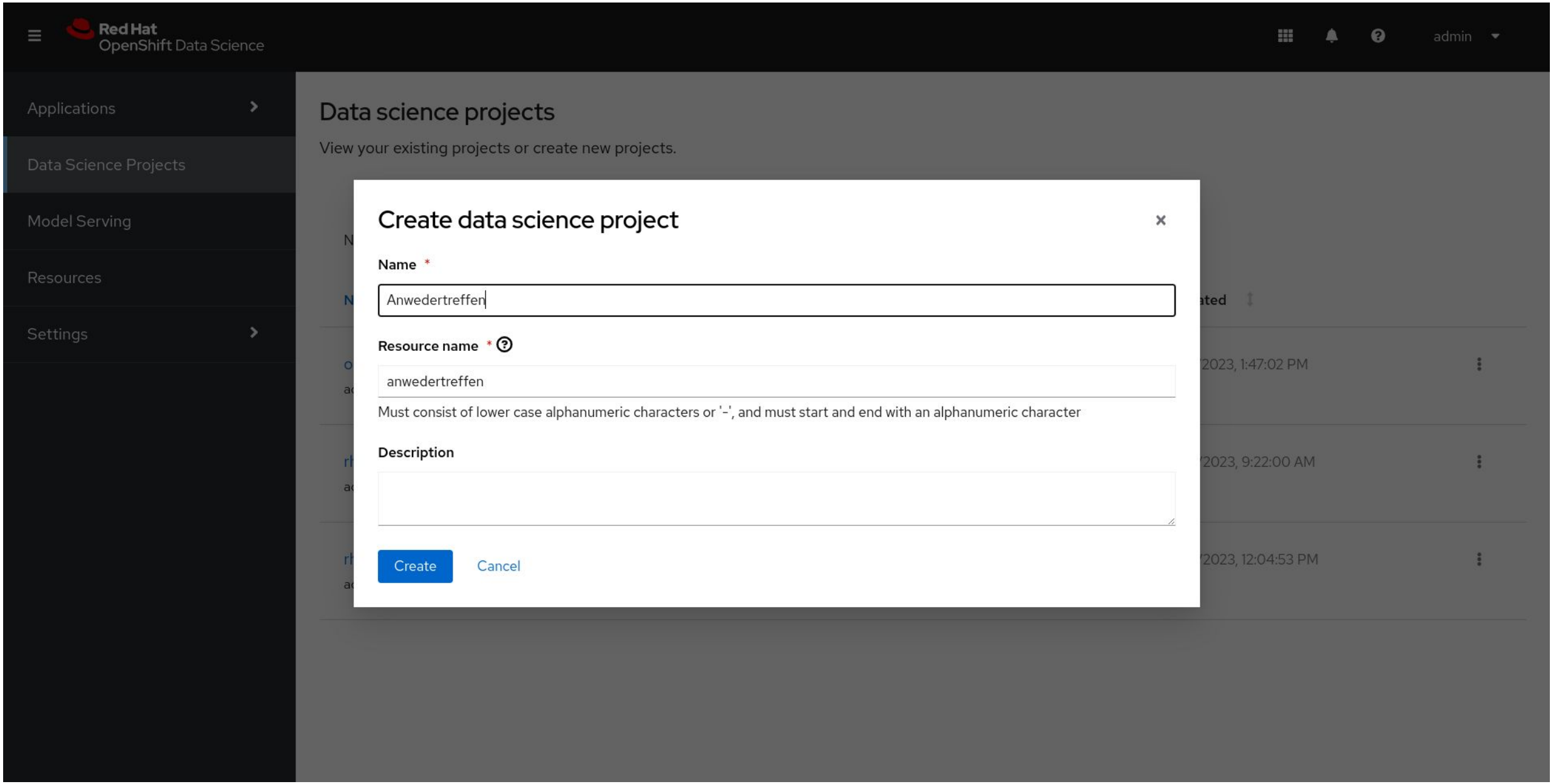
twitter.com/RedHat

Red Hat

Screenshots

This section includes:

- Red Hat OpenShift Data Science - Demo Part 1


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
View your existing projects or create new projects.
Create data science project
Name *
Anwedertreffen
Resource name *
anwedertreffen
Must consist of lower case alphanumeric characters or '-', and must start and end with an alphanumeric character
Description
Create
Cancel

Applications

Data Science Projects

Model Serving

Resources

Settings

Anwedertreffen

Jump to section

Workbenches

Cluster storage

Data connections

Models and model servers

Workbenches

Create workbench

No workbenches

To get started, create a workbench.

Cluster storage

Add cluster storage

No storage

Choose existing, or add new on cluster storage.

Data connections

Add data connection

Applications
Data Science Projects
Model Serving
Resources
Settings

Create workbench

Configure properties for your workbench.

Jump to section

Name and description
Notebook image
Deployment size
Environment variables
Cluster storage

Name *

Anwendertreffen

Description

Notebook image

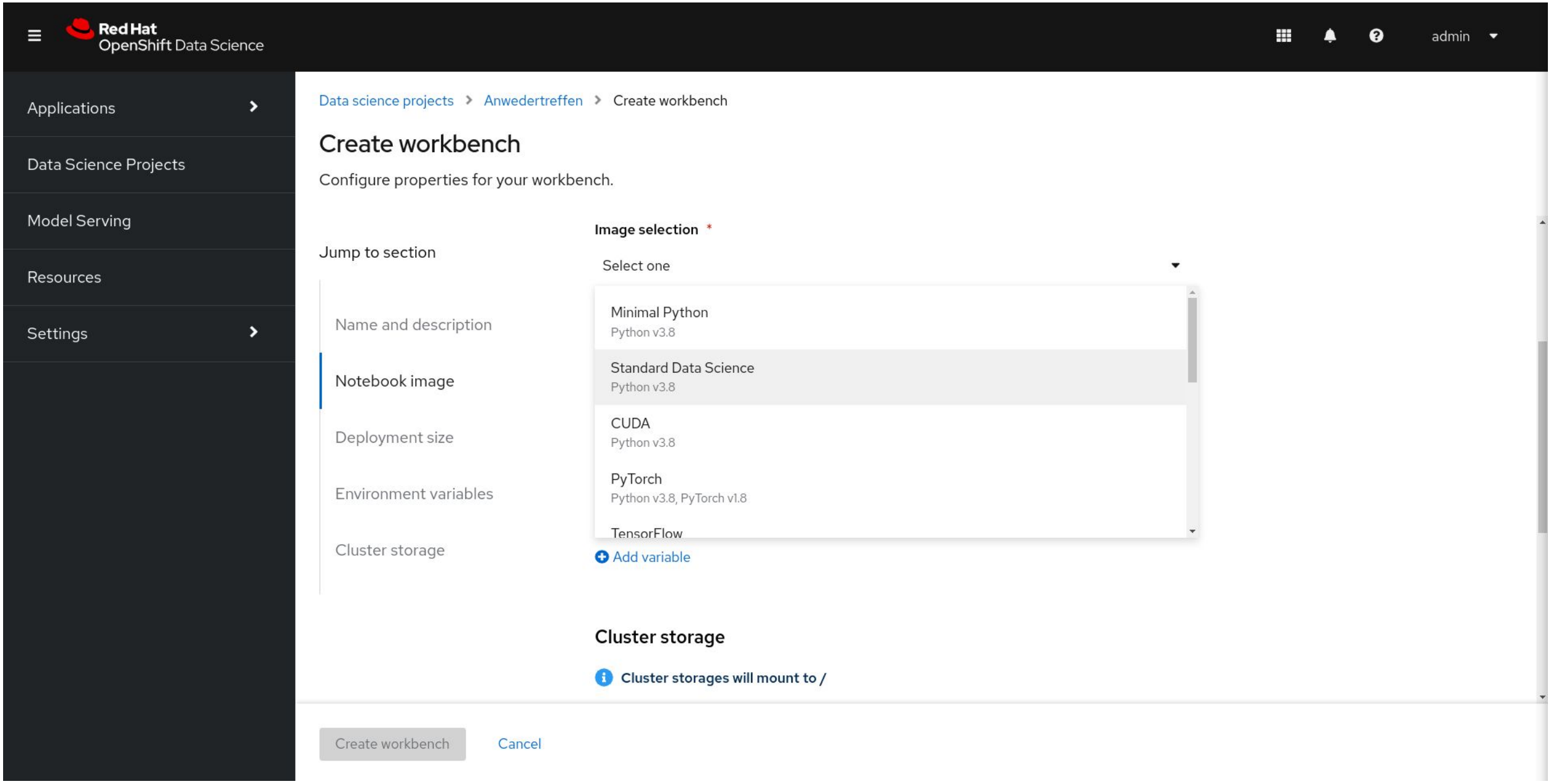
Image selection *

Select one

Minimal Python
Python v3.8

Standard Data Science
Python v3.8

Create workbench Cancel


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects > Anwedertreffen > Create workbench
Create workbench
Configure properties for your workbench.
Jump to section
Name and description
Notebook image
Deployment size
Environment variables
Cluster storage
Image selection *
Select one
Minimal Python
Python v3.8
Standard Data Science
Python v3.8
CUDA
Python v3.8
PyTorch
Python v3.8, PyTorch v1.8
TensorFlow
Add variable
Cluster storage
Cluster storages will mount to /
Create workbench
Cancel

Applications

Data Science Projects

Model Serving

Resources

Settings

Create workbench

Configure properties for your workbench.

Jump to section

Name and description

Notebook image

Deployment size

Environment variables

Cluster storage

Cluster storage

Cluster storages will mount to /

Create new persistent storage

This creates a storage that is retained when logged out.

Name *

Anwendertreffen

Description

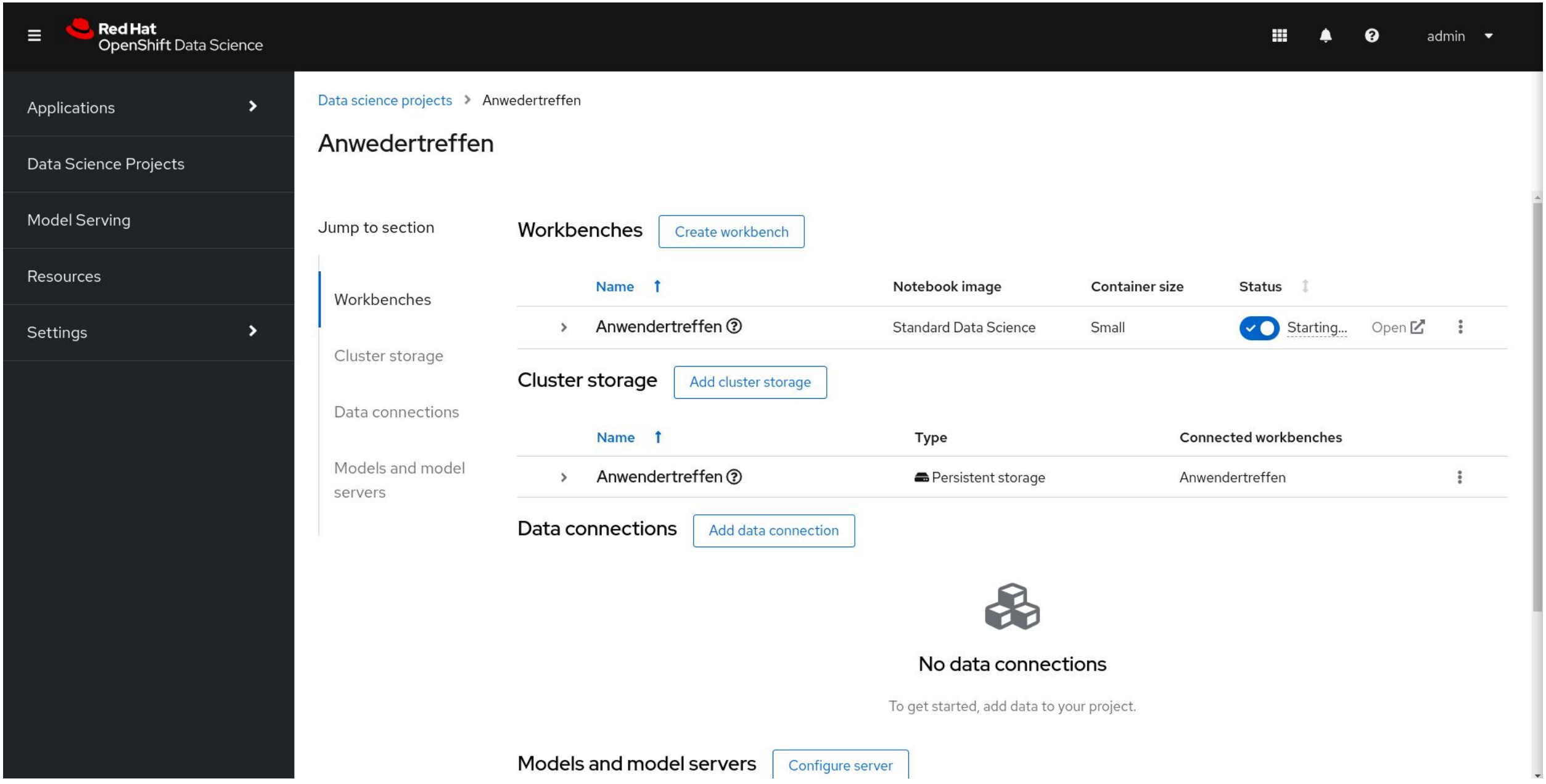
Persistent storage size

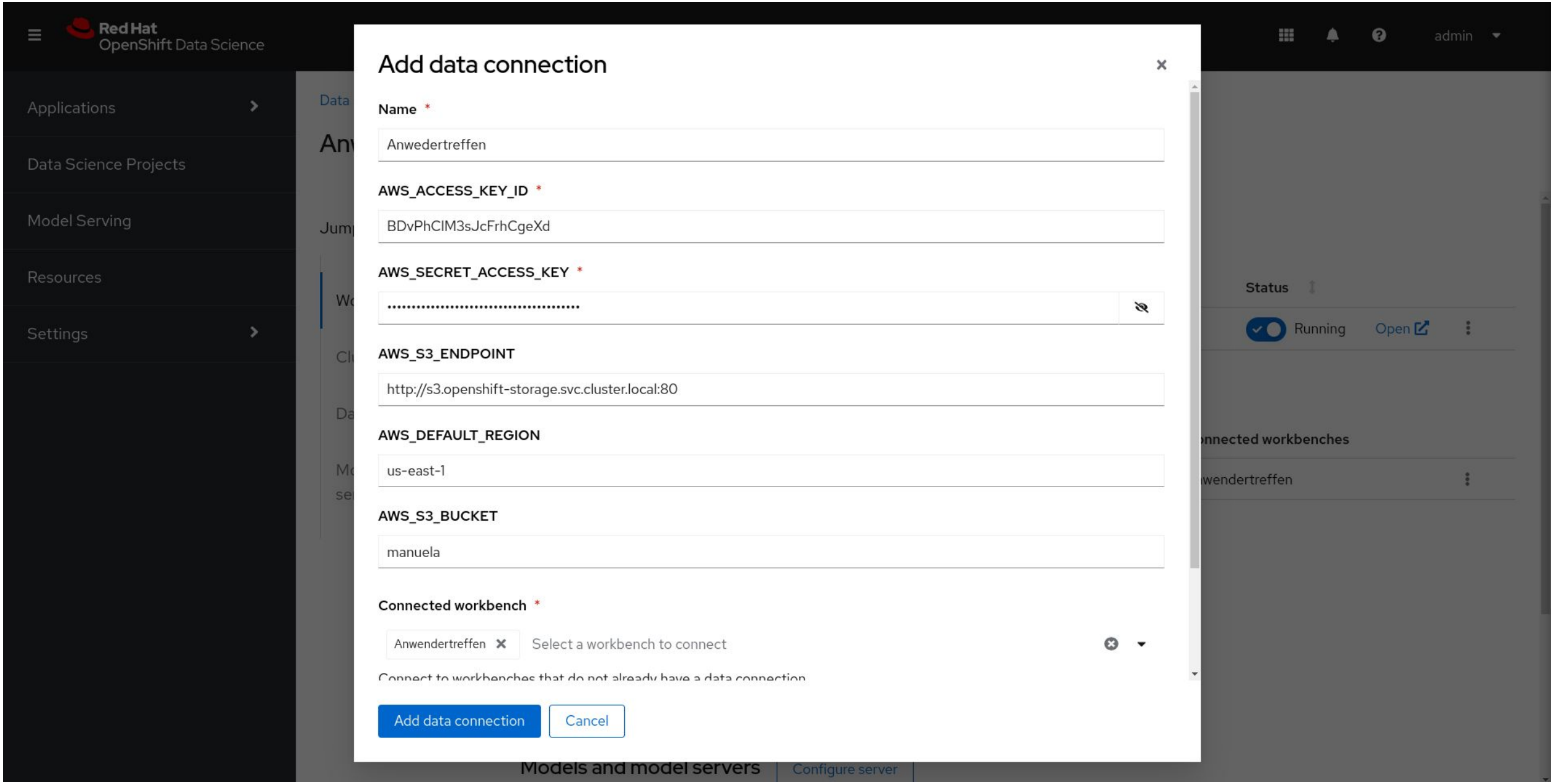
− 20 + GiB

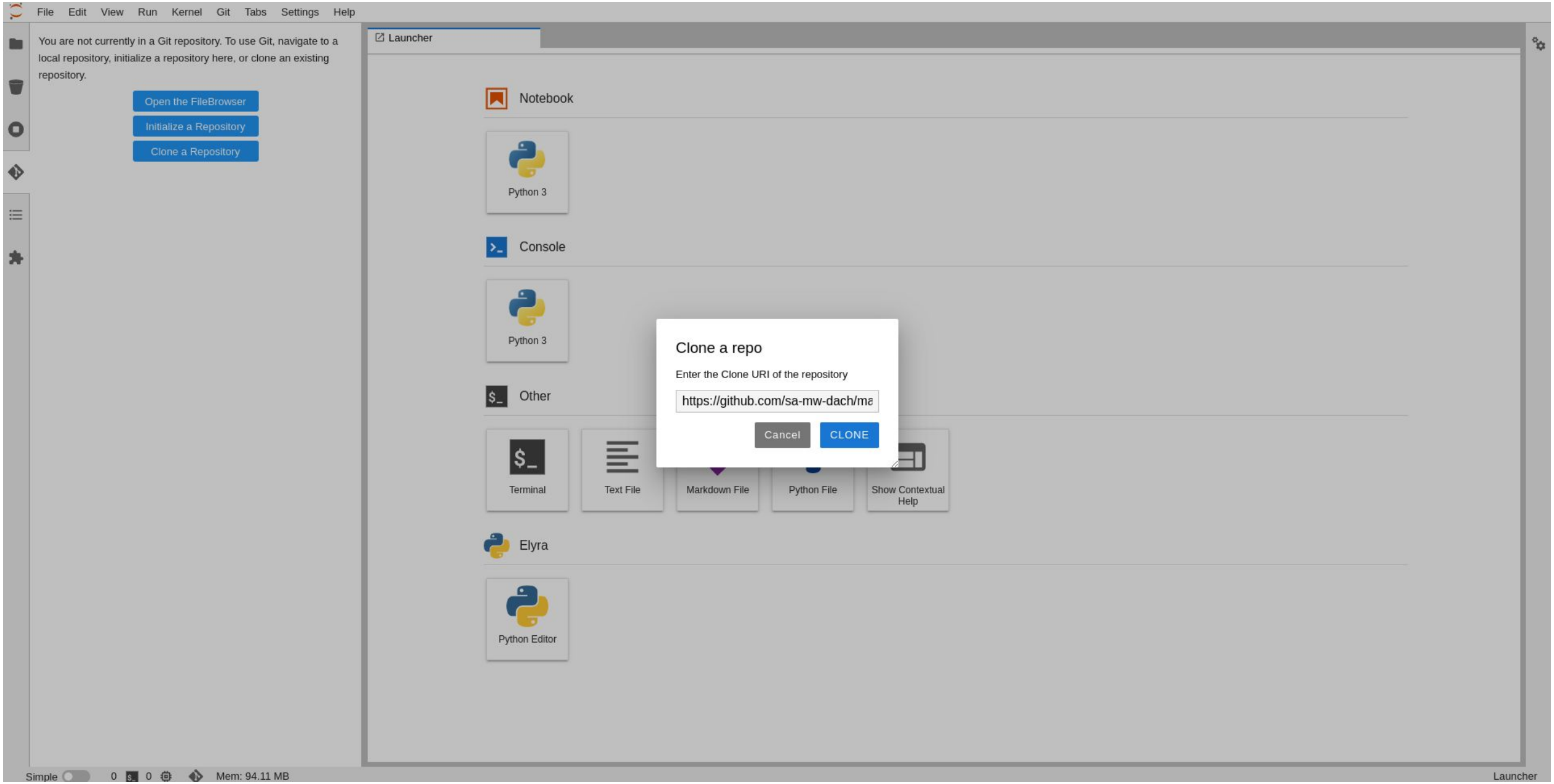
Use existing persistent storage

This reuses a previously created persistent storage.

Create workbench Cancel


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
Anwedertreffen
Anwedertreffen
Jump to section
Workbenches
Cluster storage
Data connections
Models and model servers
Workbenches
Create workbench
Name
Notebook image
Container size
Status
Anwendertreffen
Standard Data Science
Small
Starting...
Open
Cluster storage
Add cluster storage
Name
Type
Connected workbenches
Anwendertreffen
Persistent storage
Anwendertreffen
Data connections
Add data connection
No data connections
To get started, add data to your project.
Models and model servers
Configure server


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Add data connection
Name *
Anwedertreffen
AWS_ACCESS_KEY_ID *
BDvPhCIM3sJcFrhCgeXd
AWS_SECRET_ACCESS_KEY *
AWS_S3_ENDPOINT
http://s3.openshift-storage.svc.cluster.local:80
AWS_DEFAULT_REGION
us-east-1
AWS_S3_BUCKET
manuela
Connected workbench *
Anwendertreffen
Select a workbench to connect
Add data connection
Cancel
Status
Running
Open
Models and model servers
Configure server


File Edit View Run Kernel Git Tabs Settings Help
You are not currently in a Git repository. To use Git, navigate to a local repository, initialize a repository here, or clone an existing repository.
Open the FileBrowser
Initialize a Repository
Clone a Repository
Launcher
Notebook
Python 3
Console
Python 3
Other
Terminal
Text File
Markdown File
Python File
Show Contextual Help
Elyra
Python Editor
Clone a repo
Enter the Clone URI of the repository
https://github.com/sa-mw-dach/ma
Cancel
CLONE
Simple
0
0
Mem: 94.11 MB
Launcher

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/ ... / ml-models / anomaly-detection /

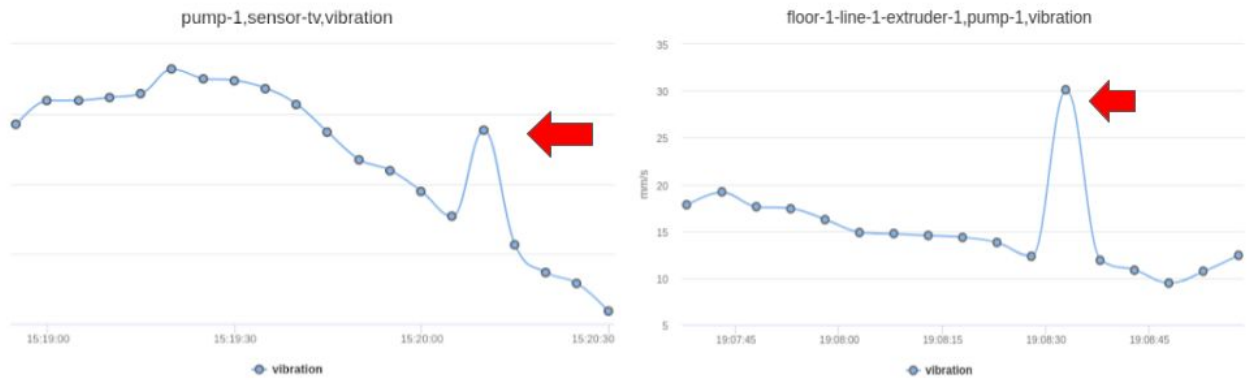
Name	Last Modified
Anomaly Detection-using-TF-and-Dee…	seconds ago
Anomaly-Detection-simple-ML-Training…	seconds ago
Data-Analyses.ipynb	seconds ago
log-to-timeseries.py	seconds ago
manuela-fetch-kafka-data.ipynb	seconds ago
messaging.log	seconds ago
model_upload.py	seconds ago
model.joblib	seconds ago
raw-data.csv	seconds ago
raw-data.log	seconds ago
README.md	seconds ago
sensor-training-data.csv	seconds ago

Launcher | Anomaly-Detection-simple-M

Markdown | git | Python 3

MANUela Anomaly ML Model

Goal: Build a machine lerning model that detects anomalies in sensor vibration data



Vibration anomalies

Outline:

Read and preview the labeled data

- Read data from disk
- Preview the raw data
- Visualize vibration and anomalies

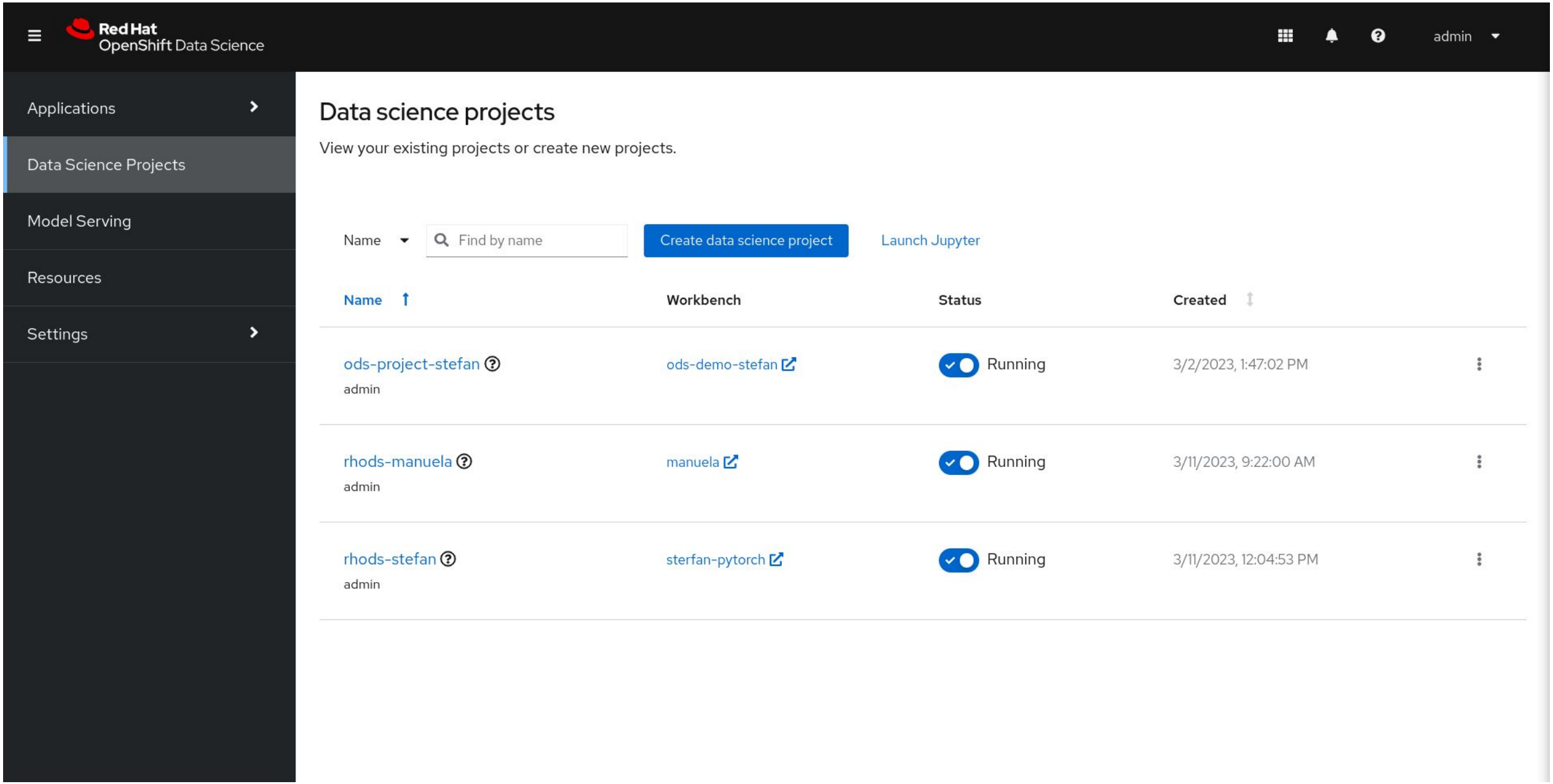
Data wrangling

- Convert time series data into small episodes that can be used for supervised learning.
- Explore the new data format

Model training

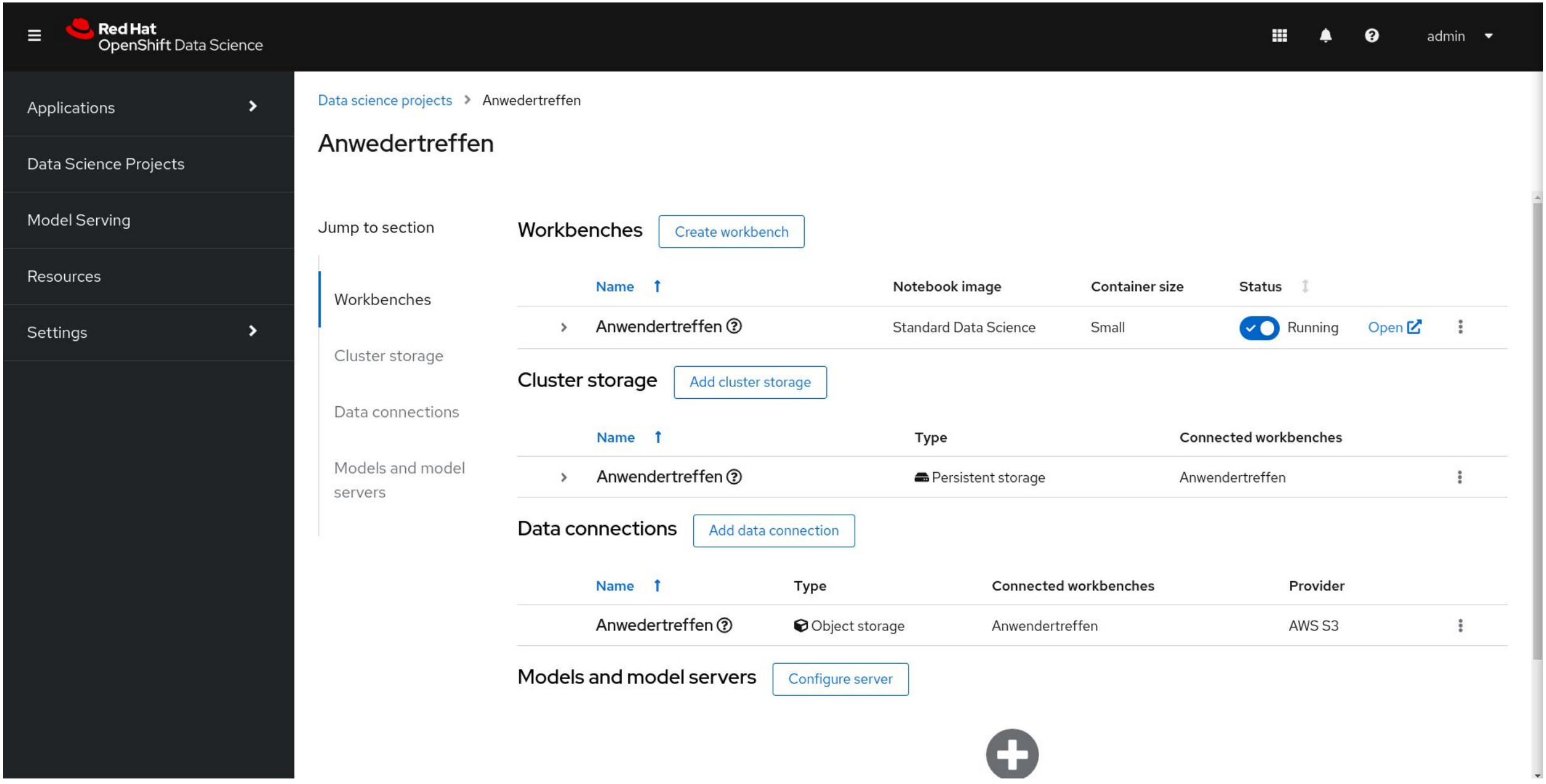
- Prepare the data for modeling, training and testing

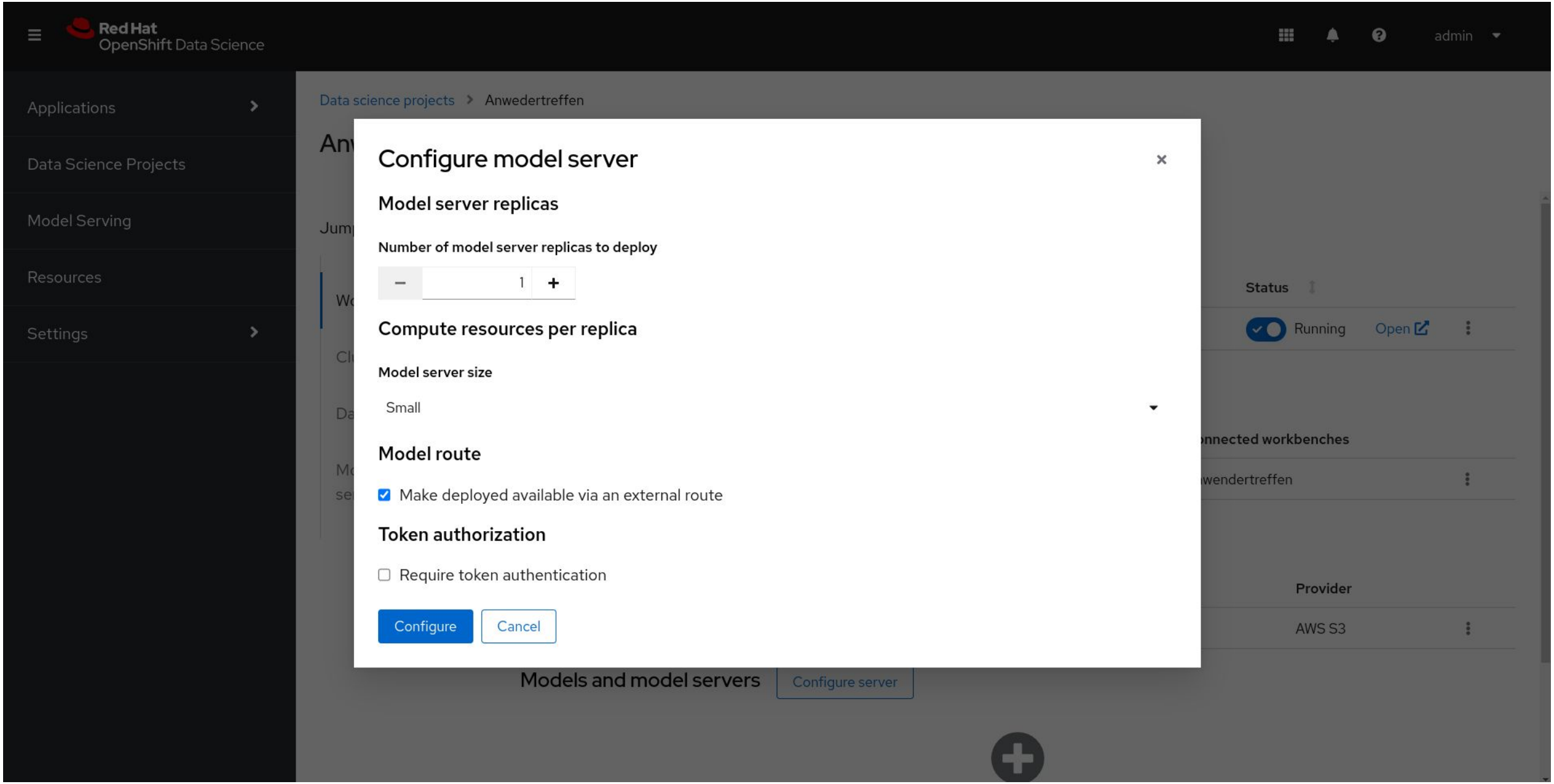
Simple 0 1 Initialized (additional servers needed) Python 3 | Idle Mem: 94.13 MB Mode: Command Ln 1, Col 1 Anomaly-Detection-simple-ML-Training.ipynb


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
View your existing projects or create new projects.
Name
Find by name
Create data science project
Launch Jupyter
Name
Workbench
Status
Created
ods-project-stefan
admin
ods-demo-stefan
Running
3/2/2023, 1:47:02 PM
rhods-manuela
admin
manuela
Running
3/11/2023, 9:22:00 AM
rhods-stefan
admin
sterfan-pytorch
Running
3/11/2023, 12:04:53 PM

This section includes:

- Red Hat OpenShift Data Science - Demo Part 2


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
Anwedertreffen
Anwedertreffen
Jump to section
Workbenches
Cluster storage
Data connections
Models and model servers
Workbenches
Create workbench
Name
Notebook image
Container size
Status
Anwendertreffen
Standard Data Science
Small
Running
Open
Cluster storage
Add cluster storage
Name
Type
Connected workbenches
Anwendertreffen
Persistent storage
Anwendertreffen
Data connections
Add data connection
Name
Type
Connected workbenches
Provider
Anwedertreffen
Object storage
Anwendertreffen
AWS S3
Models and model servers
Configure server


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
Anwedertreffen
Configure model server
Model server replicas
Number of model server replicas to deploy
1
Compute resources per replica
Model server size
Small
Model route
Make deployed available via an external route
Token authorization
Require token authentication
Configure
Cancel
Status
Running
Open
Provider
AWS S3
Models and model servers
Configure server

Applications

Data Science Projects

Model Serving

Resources

Settings

Anwedertreffen

Jump to section

Workbenches

Cluster storage

Data connections

Models and model servers

Name	Notebook image	Container size	Status	
Anwendertreffen	Standard Data Science	Small	Running	Open

Cluster storage

Add cluster storage

Name	Type	Connected workbenches
Anwendertreffen	Persistent storage	Anwendertreffen

Data connections

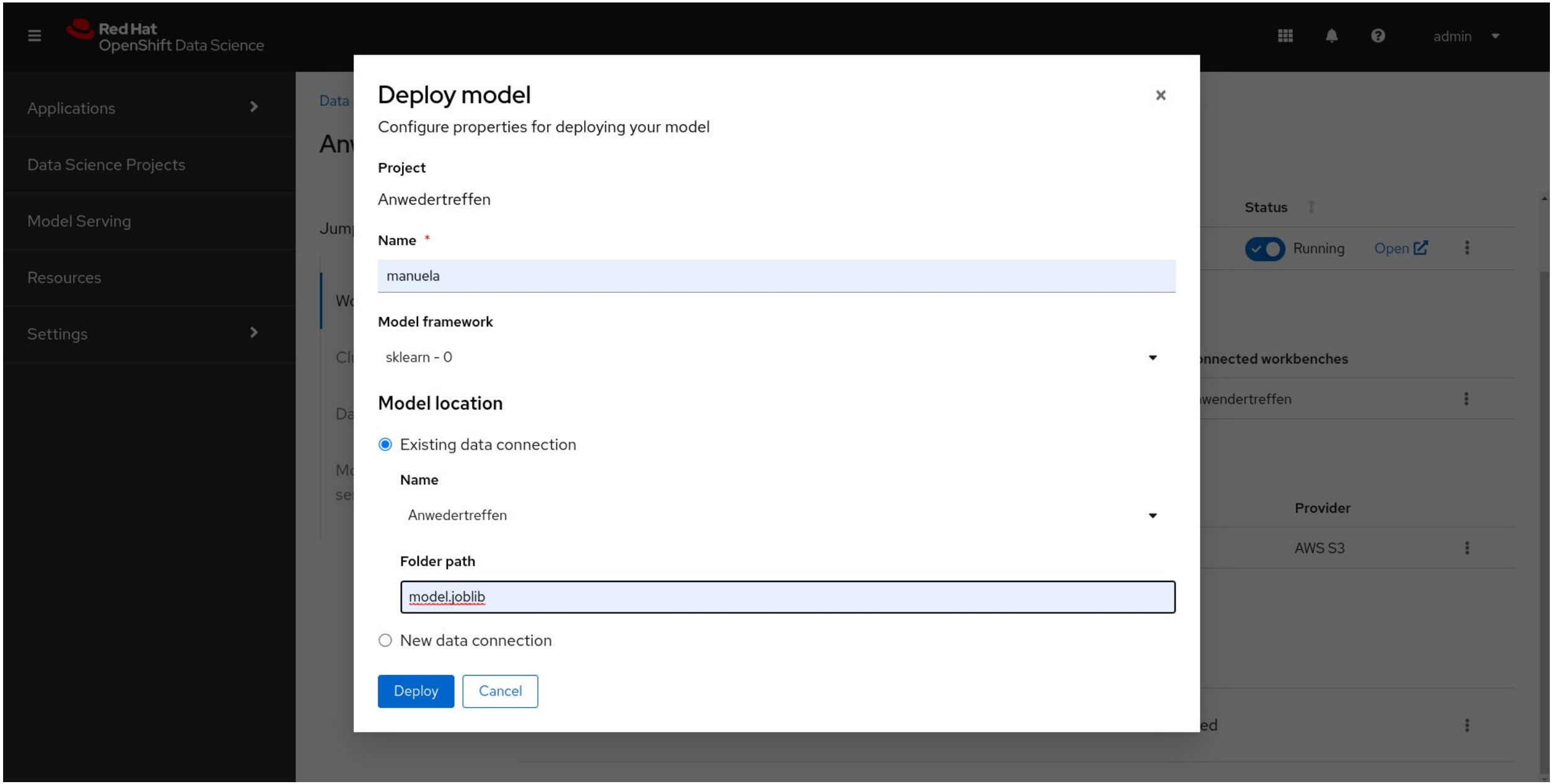
Add data connection

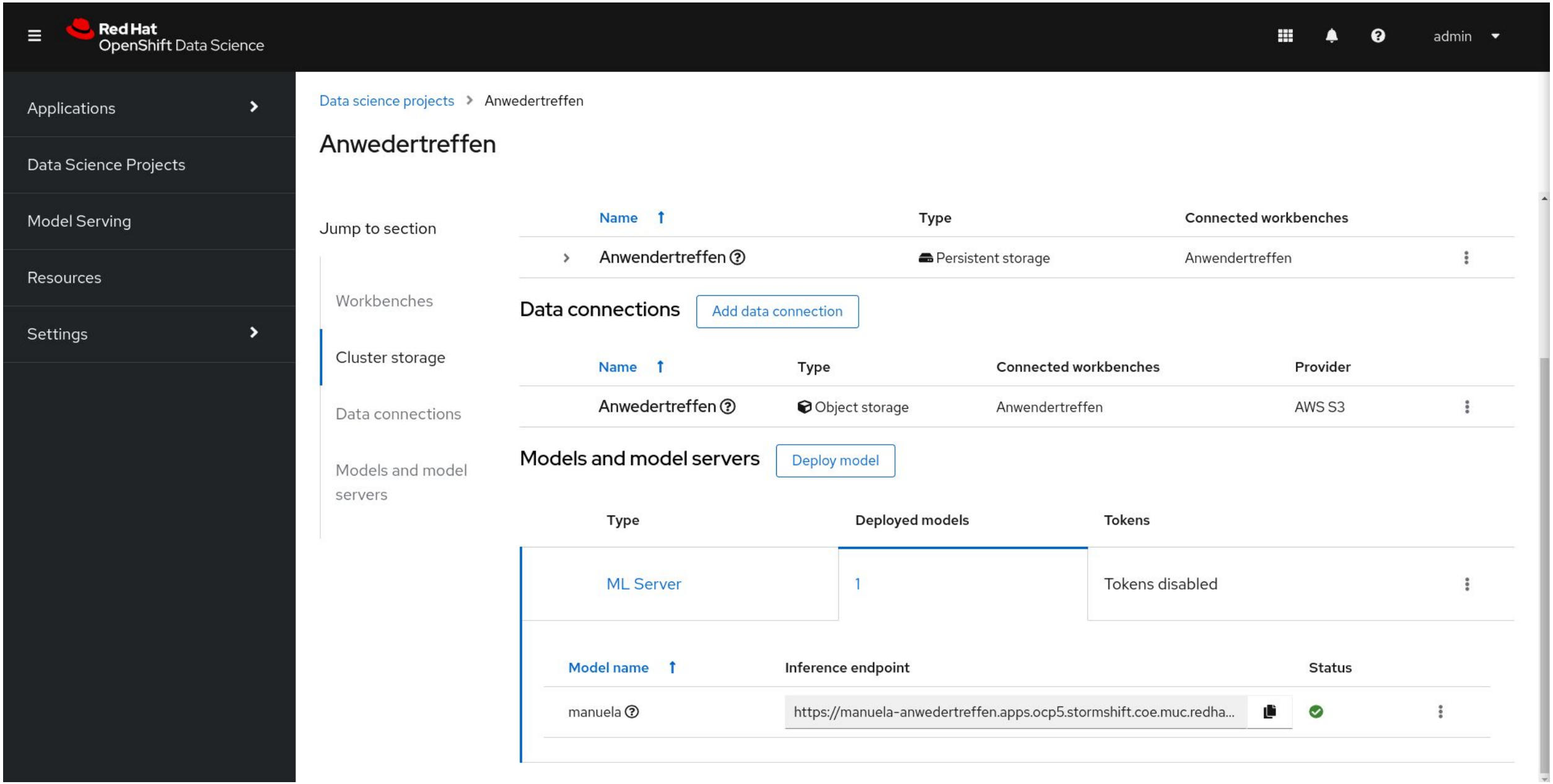
Name	Type	Connected workbenches	Provider
Anwedertreffen	Object storage	Anwendertreffen	AWS S3

Models and model servers

Deploy model

Type	Deployed models	Tokens
ML Server	0	Tokens disabled


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data
Status
Running
Open
Provider
AWS S3
Deploy model
Configure properties for deploying your model
Project
Anwedertreffen
Name *
manuela
Model framework
sklearn - 0
Model location
Existing data connection
Name
Anwedertreffen
Folder path
model.joblib
New data connection
Deploy
Cancel


Red Hat
OpenShift Data Science
admin
Applications
Data Science Projects
Model Serving
Resources
Settings
Data science projects
Anwedertreffen
Anwedertreffen
Jump to section
Workbenches
Cluster storage
Data connections
Models and model servers
Name
Type
Connected workbenches
Anwendertreffen
Persistent storage
Anwendertreffen
Data connections
Add data connection
Name
Type
Connected workbenches
Provider
Anwedertreffen
Object storage
Anwendertreffen
AWS S3
Models and model servers
Deploy model
Type
Deployed models
Tokens
ML Server
1
Tokens disabled
Model name
Inference endpoint
Status
manuela
https://manuela-anwedertreffen.apps.ocp5.stormshift.coe.muc.redha...

https://manuela-anwedertreffen.apps.ocp5.stormshift.coe.muc.redhat.com/v2/models/manuela/infer

Save

POST https://manuela-anwedertreffen.apps.ocp5.stormshift.coe.muc.redhat.com/v2/models/manuela/infer

Send

Params Authorization Headers (9) Body Pre-request Script Tests Settings Cookies

none form-data x-www-form-urlencoded raw binary GraphQL JSON Beautify

```
{
    "inputs": [
        { "name": "predict",
        "shape": [1, 5],
        "datatype": "FP32",
        "data": [1.0, 2.0, 1.0, 2.0 , 12.0 ]
        }
    ]
}
```

Body Cookies (1) Headers (6) Test Results Status: 200 OK Time: 186 ms Size: 357 B Save Response

Pretty Raw Preview Visualize JSON

```
{
    "model_name": "manuela__isvc-a8e7b6283b",
    "outputs": [
        {
            "name": "predict",
            "datatype": "INT64",
            "shape": [
                1
            ],
            "data": [
                1
            ]
        }
    ]
}
```

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

Red Hat

TO-DO:

- Prep demo flow with Manuela Model (and recording?)
  - Create workbench
  - Clone repo
  - Train model
  - Save mode in S3 Bucket S3

- What is in it for me?
  - OpenShift Services Owner
  - Data Scientist / ML engineer
  - App Developer of smart applications

Notes

https://docs.google.com/presentation/d/1Tvir9Fots1QRgNsLnHfcDt32mgLF_FP9c8J5vfZeYjM/edit#slide=id.gd65fe1b395_1_251

Red Hat OpenShift Data Science_Messaging - final - New Kind

https://docs.google.com/presentation/d/1Tvir9Fots1QRgNsLnHfcDt32mgLF_FP9c8J5vfZeYjM/edit#slide=id.gd65fe1b395_1_251

CY23 RHODS Focus - AI/ML COP -
https://docs.google.com/presentation/d/1NXkkeLFgwOX6Ciljdb7zQMSHw91ldN98HAJpklsk45E/edit#slide=id.g1f096d90674_0_1751

OpenShift Coffee Break: MLOps with OpenShift - YouTube
https://www.youtube.com/watch?v=5zHumreNS_k&t=1705s

AI/ML with Red Hat for Siemens - Google Slides
https://docs.google.com/presentation/d/1G3wCU2GJfsELsw7XnRwKJtGHeBFeSYyalJ5ZvRmIWwI/edit#slide=id.g10487018f2c_0_719

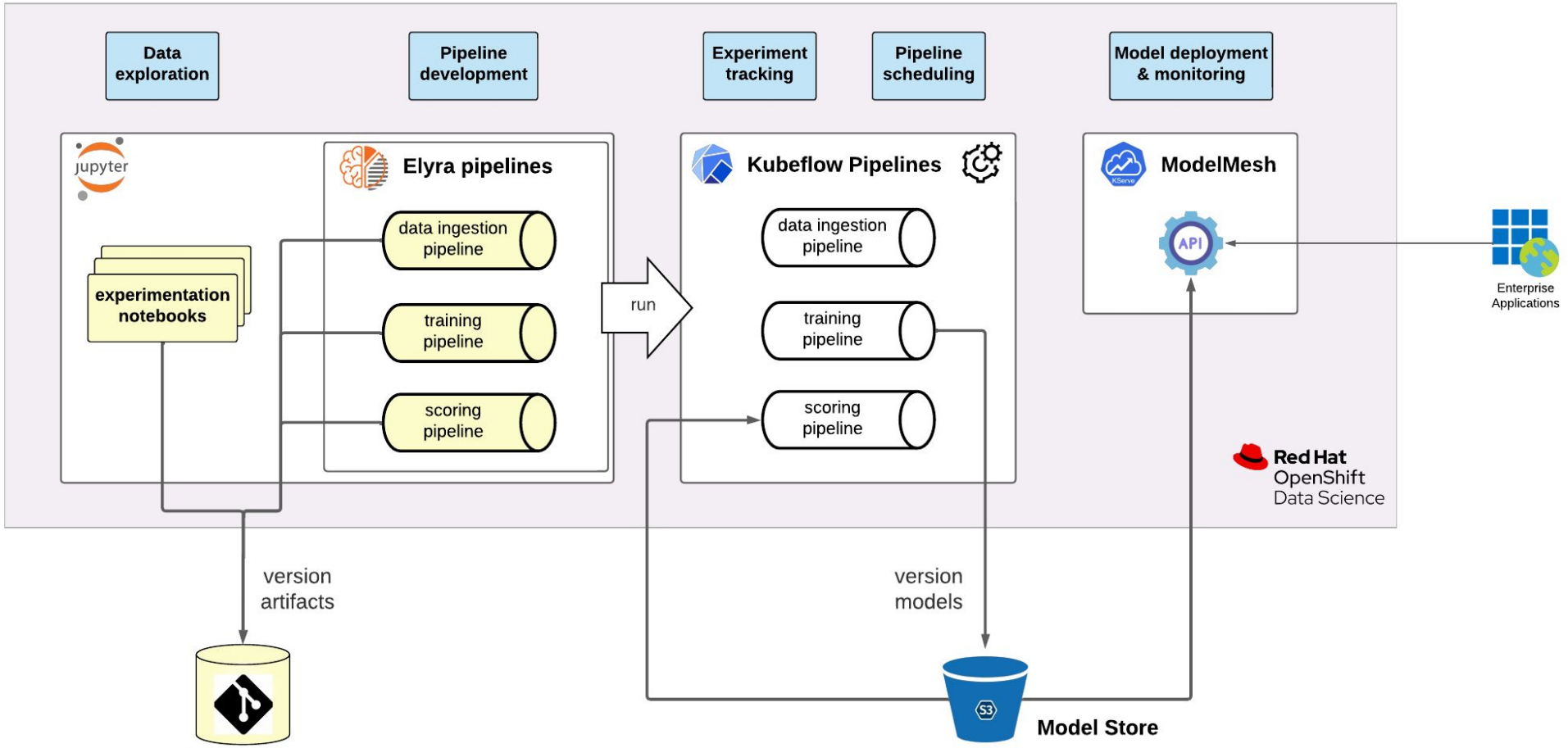
Red Hat OpenShift Data Science 101 deck - Google Slides
https://docs.google.com/presentation/d/1YKlaYhb8Du8clmEt3wMEWZ8F2dONp3Jbetpr2r47ego/edit#slide=id.g16c6c4e6195_9_0

RHODS - Anwendertreffen 03-2023 - Google Slides
https://docs.google.com/presentation/d/1ARDPhjqzd0_OYsLbsesg8e4WUZIBD01T1Wv3hz0m4wg/edit#slide=id.g21949895a79_0_1186

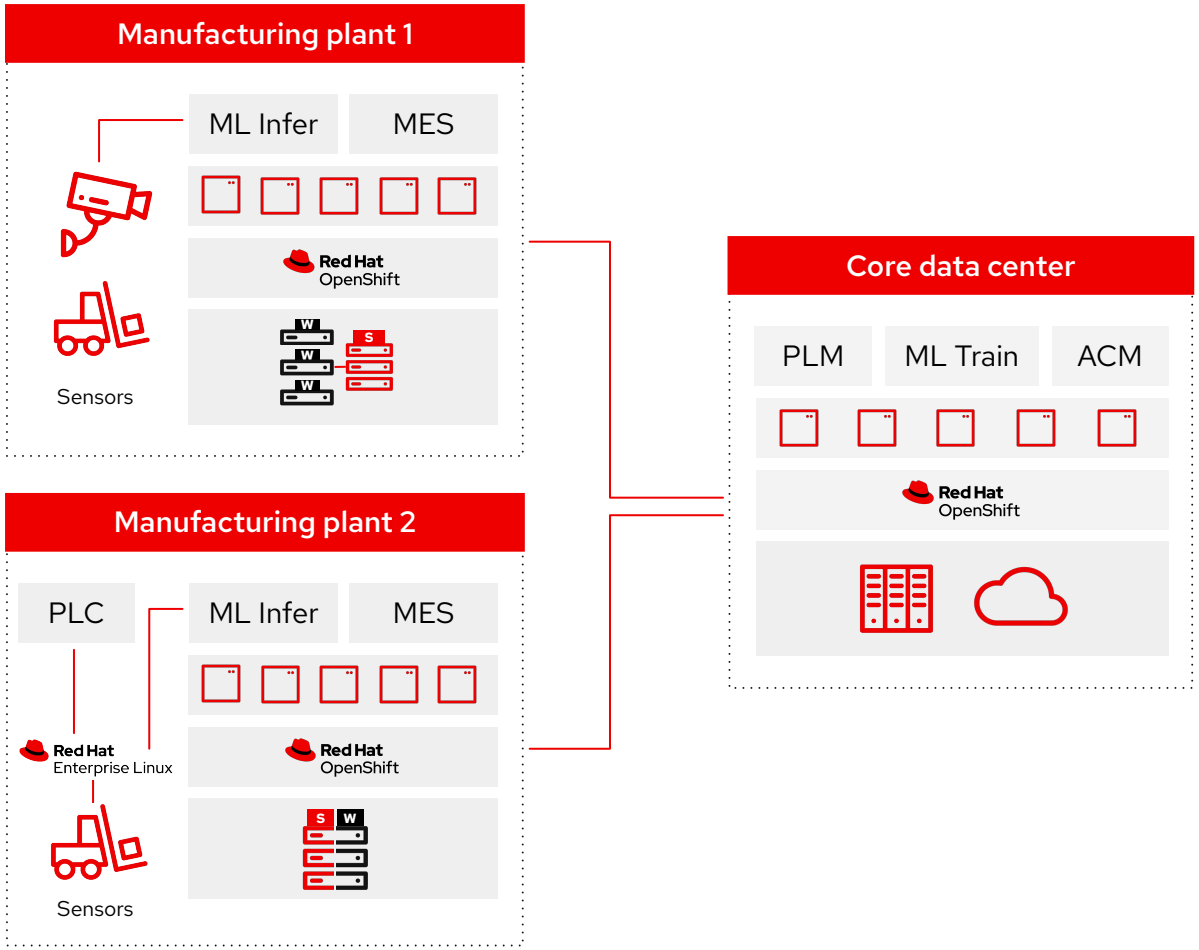
TO-DO:

- Prep demo flow with Manuela Model (and recording?)
  - Create model server and deploy model
  - Basic test with CURL / POSTMAN
  - **["Update" Manuela to use Model**

ML workflow with RHODS



Red Hat



Industrial use case

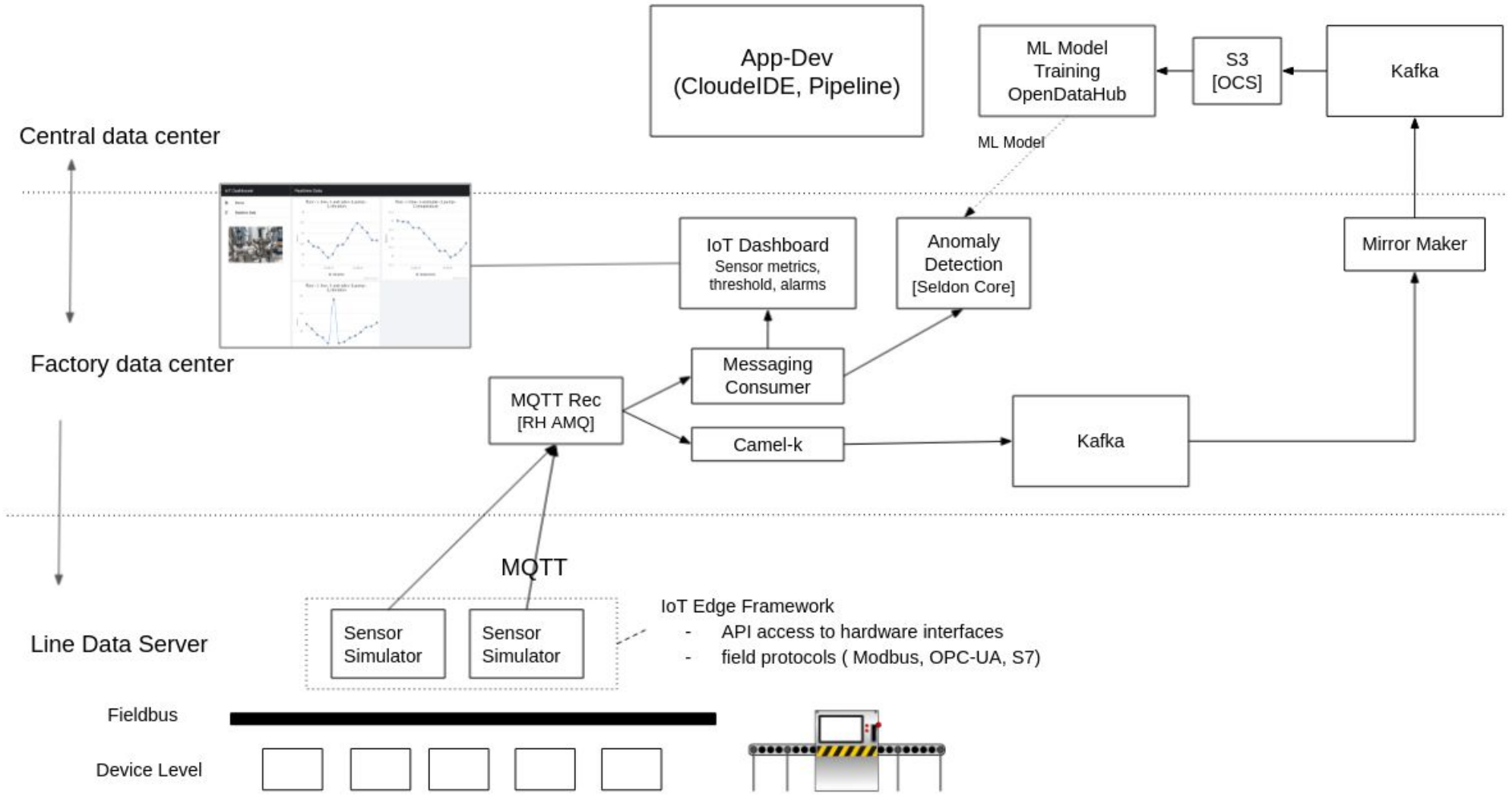
Red Hat OpenShift helping create the smart factory at the industrial edge

- Simplify the deployment and lifecycle management of AI-powered applications
- Accelerate data gathering, preparation, and inferencing tasks
- Consistent development platform and tools
- Turn insights into positive business outcomes faster

Legend:
ML Train: Machine learning training
ML Infer: Machine learning inference
MES: Manufacturing execution system
ACM: Red Hat Advanced Cluster Management for Kubernetes
PLC: Programmable logic controller
PLM: Product lifecycle management

Red Hat

AI at the Factory



Red Hat

Others running AI on Red Hat OpenShift

Boston Children's Hospital

Learn more

BMW GROUP

Learn more

ExxonMobil

Learn more

HCA Healthcare

Learn more

RBC

Learn more

An Roinn Talmhaíochta, Bia agus Mara
Department of Agriculture, Food and the Marine

Learn more

KBTG
KASIKORN BUSINESS-TECHNOLOGY GROUP

Learn more

Public Health England

Learn more

Red Hat

AI/ML consulting services portfolio

A modular approach to customer implementations

Open AI/ML platform
- 4-8 weeks

MLOps Foundation
- 6-12 weeks

Intelligent application development pilot
- 6-12 weeks

Data engineering pilot
- 6-12 weeks

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

facebook.com/redhatinc

youtube.com/user/RedHatVideos

twitter.com/RedHat

Red Hat

MLOps with Red Hat

An Overview

Max Dargatz
Senior Solution Architect
mdargatz@redhat.com

Max Murakami
Specialist Solution Architect
mmurakam@redhat.com

Red Hat

MLOps with Red Hat

- MLOps Platform
- Demo: ML CI/CD Pipeline
- Q&A

Red Hat

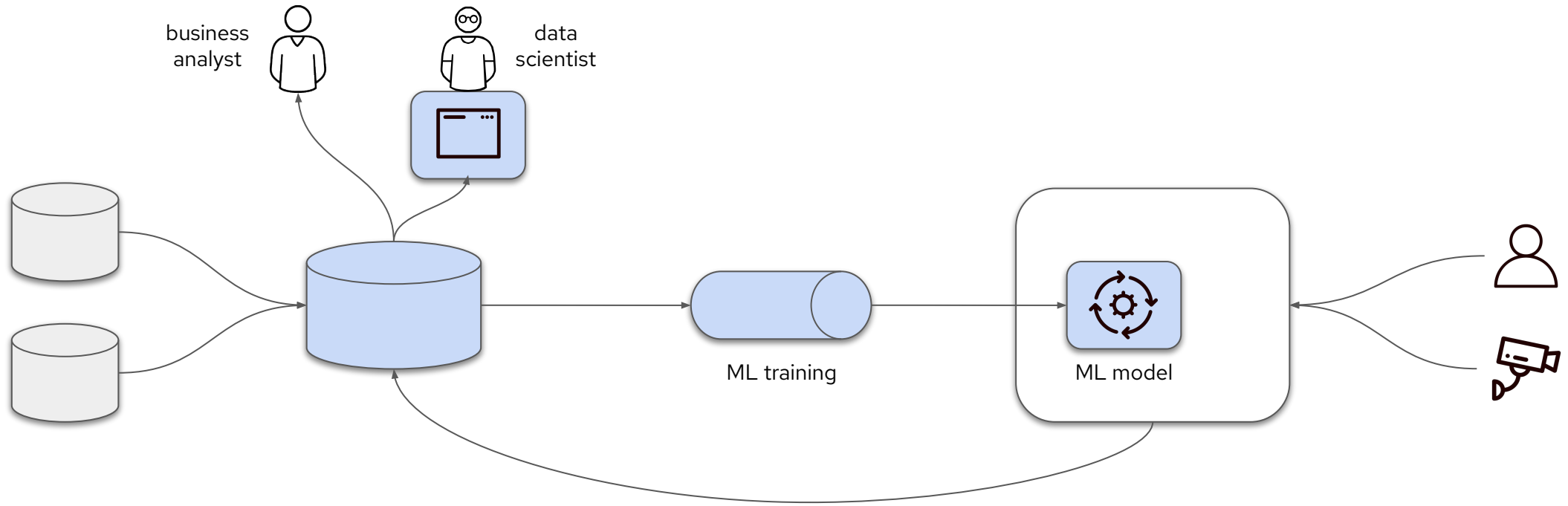
The AI Journey

Business Intelligence

Data Science

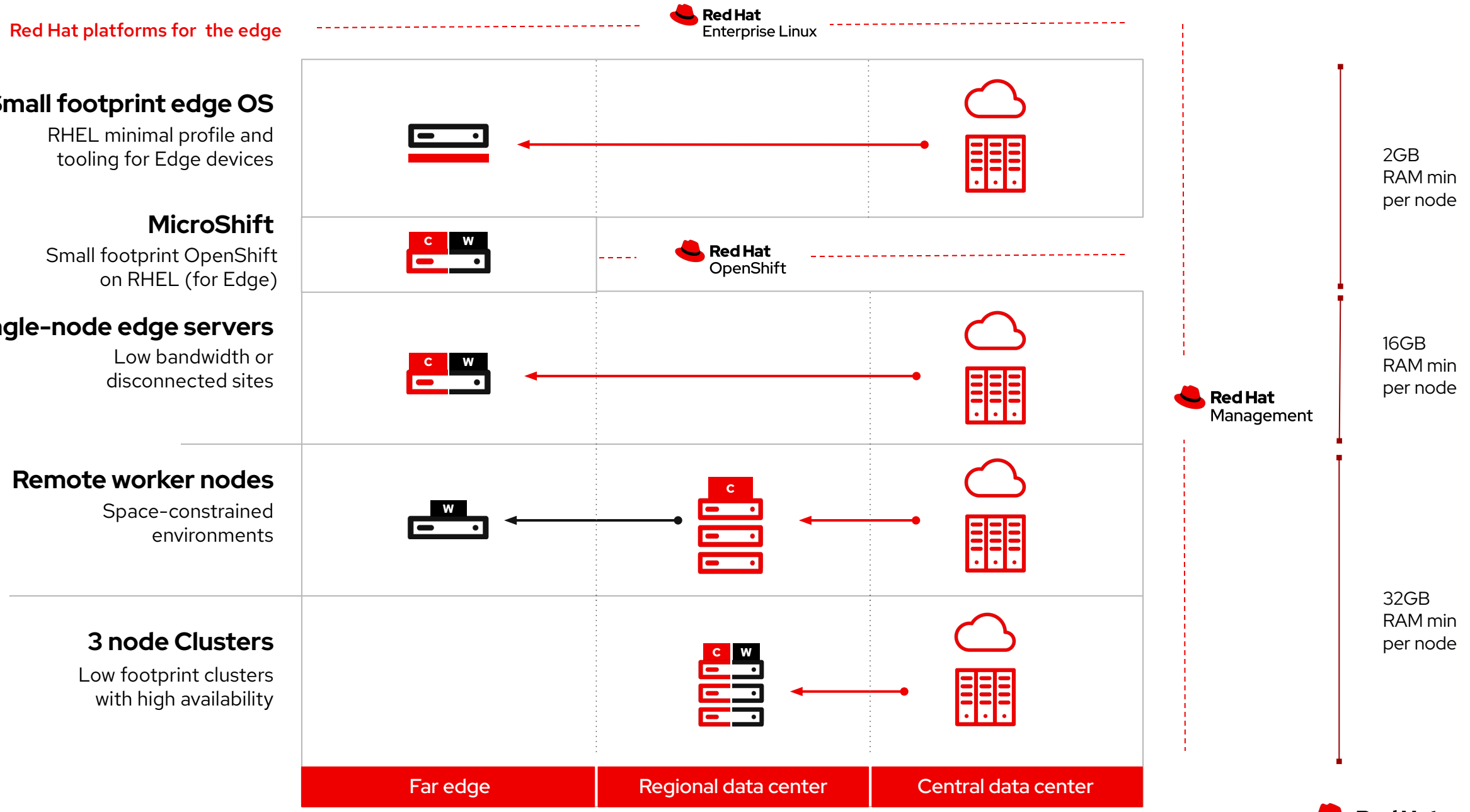
Machine Learning

Intelligent Applications



Arrows denote direction of data flow.

Red Hat platforms for the edge



Cluster management and application deployment

Kubernetes node control

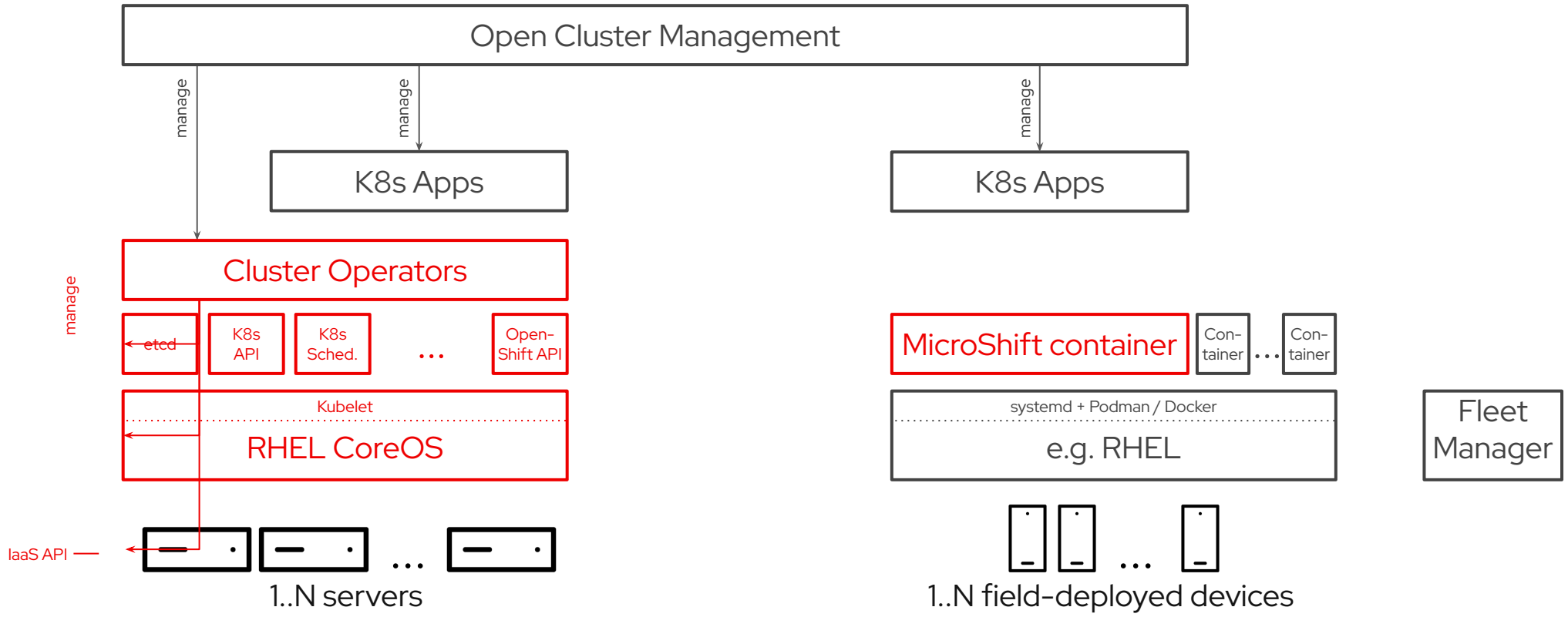
C Control node

W Worker node

Red Hat

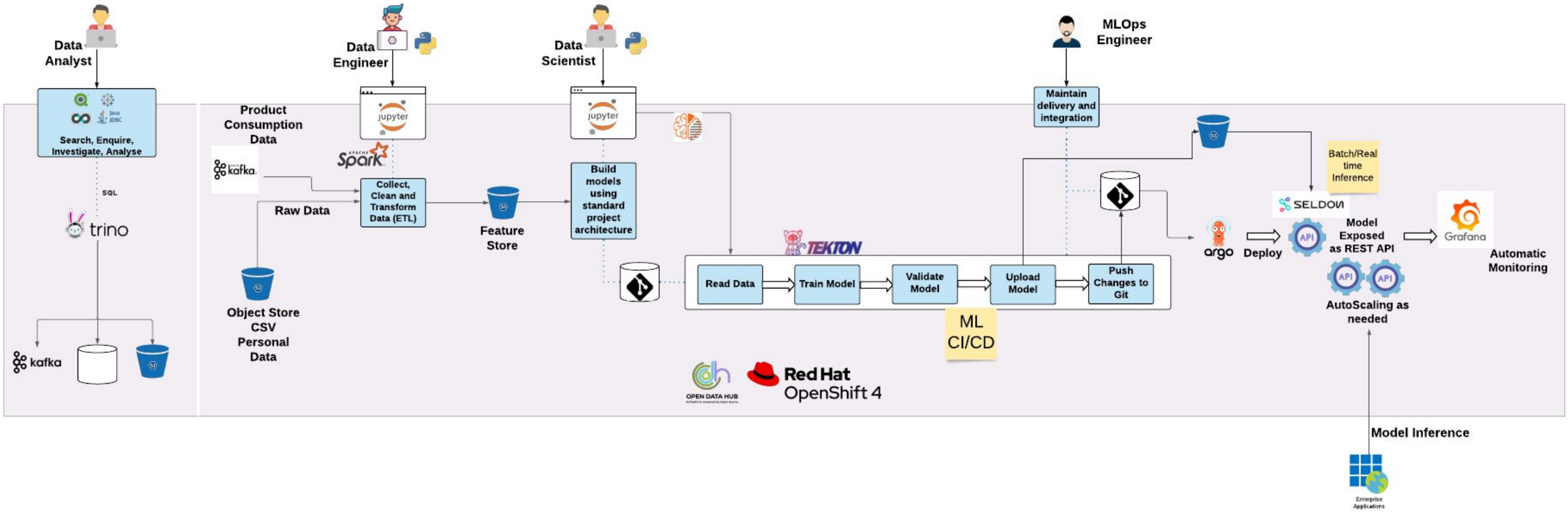
What is MicroShift?

A small form-factor OpenShift for field-deployed devices



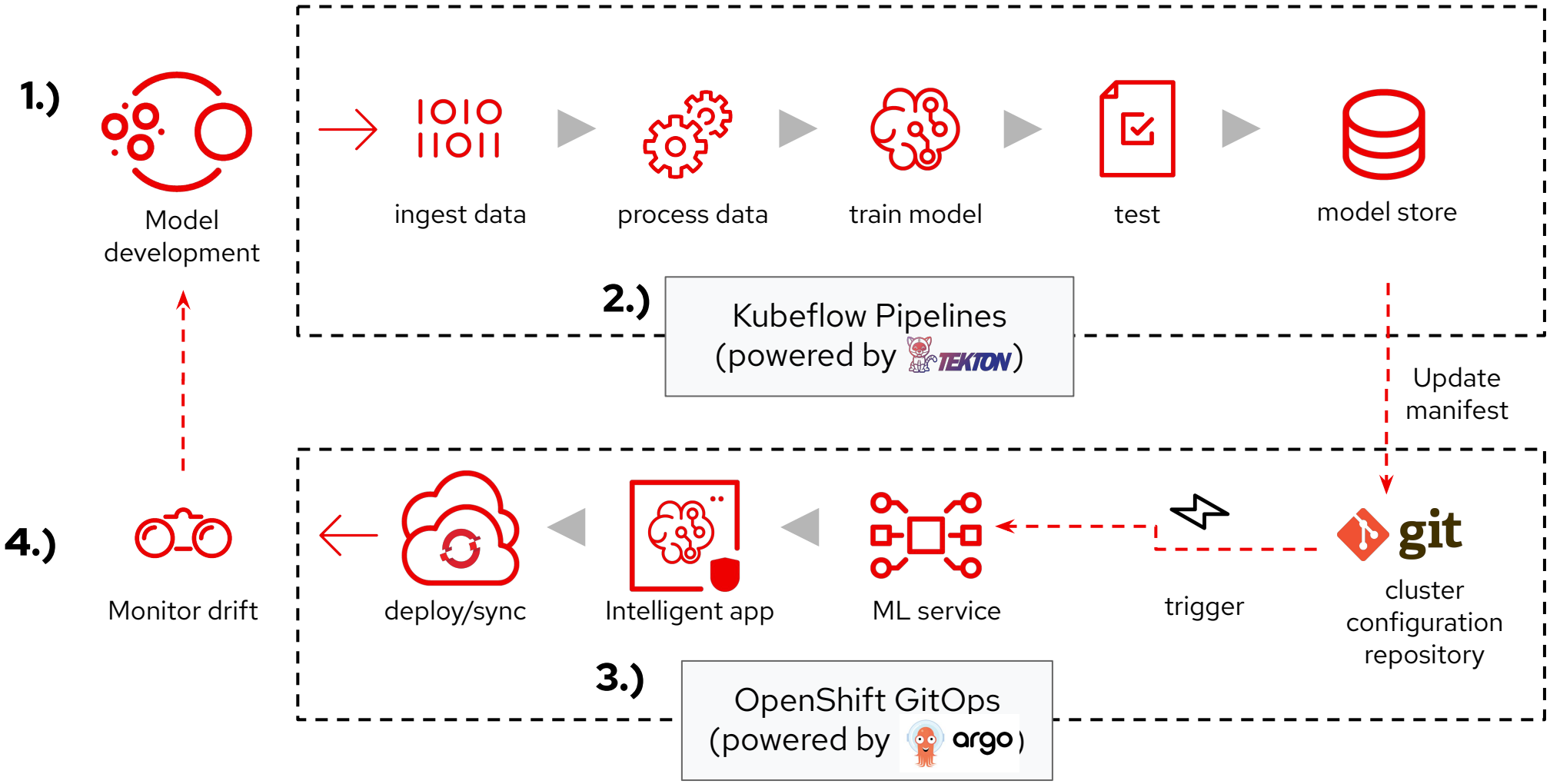
Red Hat

End to End - AI / ML workflow

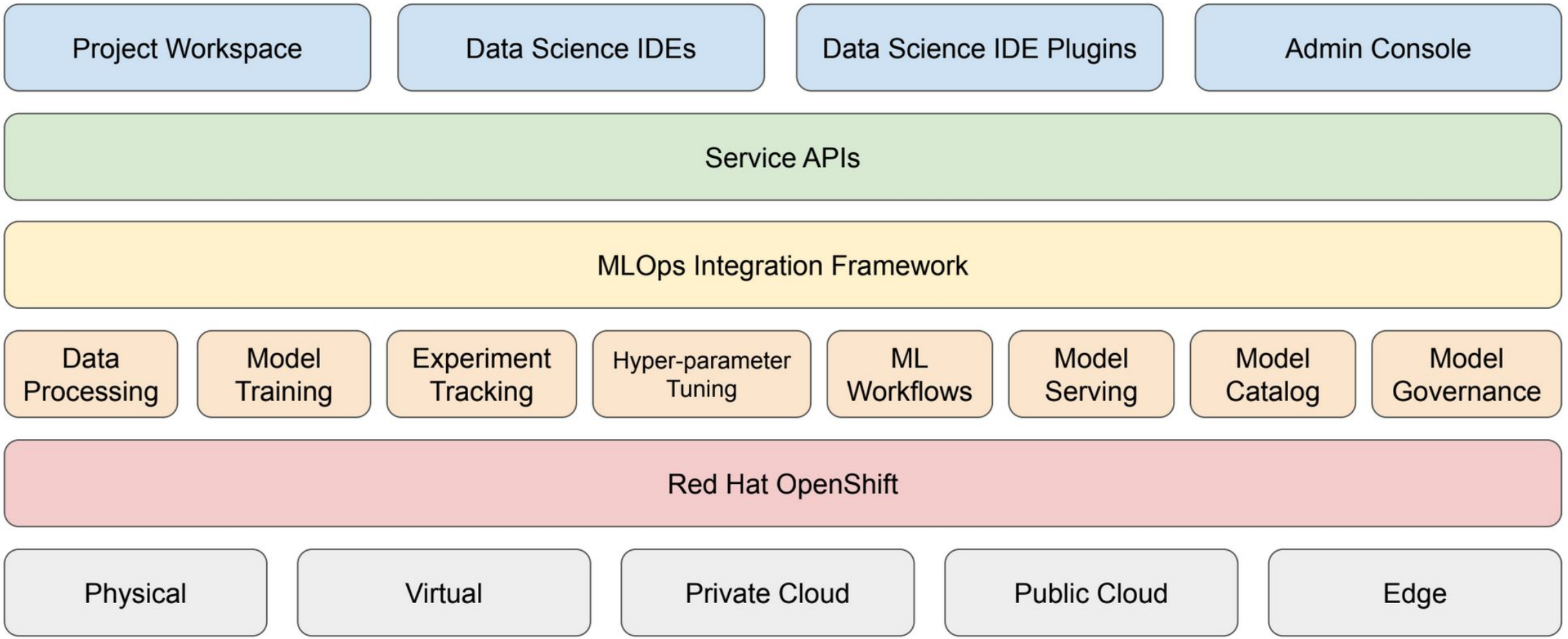


Red Hat

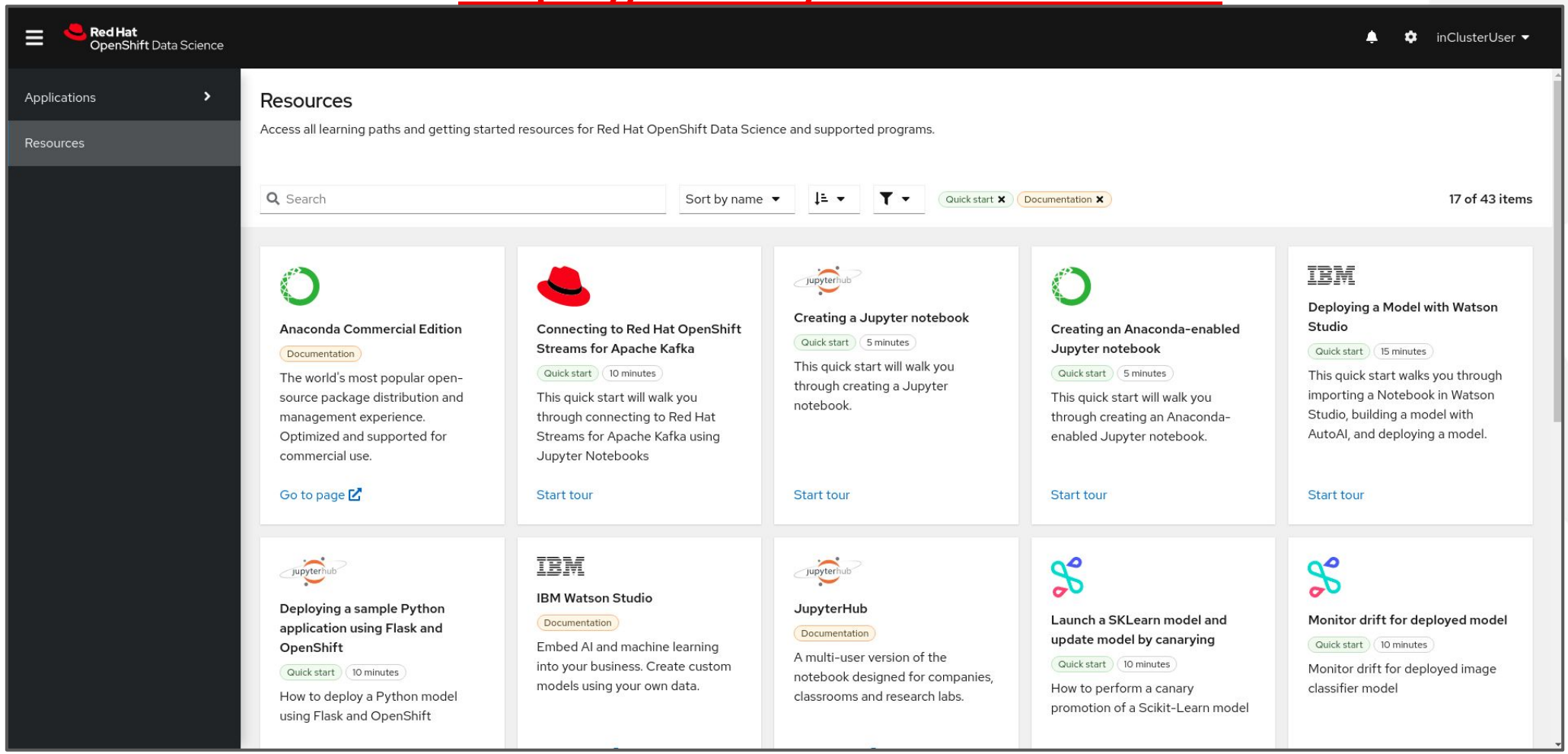
MLOps with Red Hat OpenShift



MLOps Platform Vision



https://red.ht/rhods-sandbox



Red Hat

Bringing AI/ML open source to the enterprise

Upstream Projects

Community Project

Product

Red Hat OpenShift Data Science

Red Hat

Open Data Hub

An open source AI/ML suite

- Popular data science platform
- Self-service Jupyter notebooks

Elyra

- AI pipeline editor
- Define workflows through Jupyter

- Cloud-native pipeline orchestration
- Serverless execution

- Deploying machine learning models as micro-services
- Full model lifecycle management

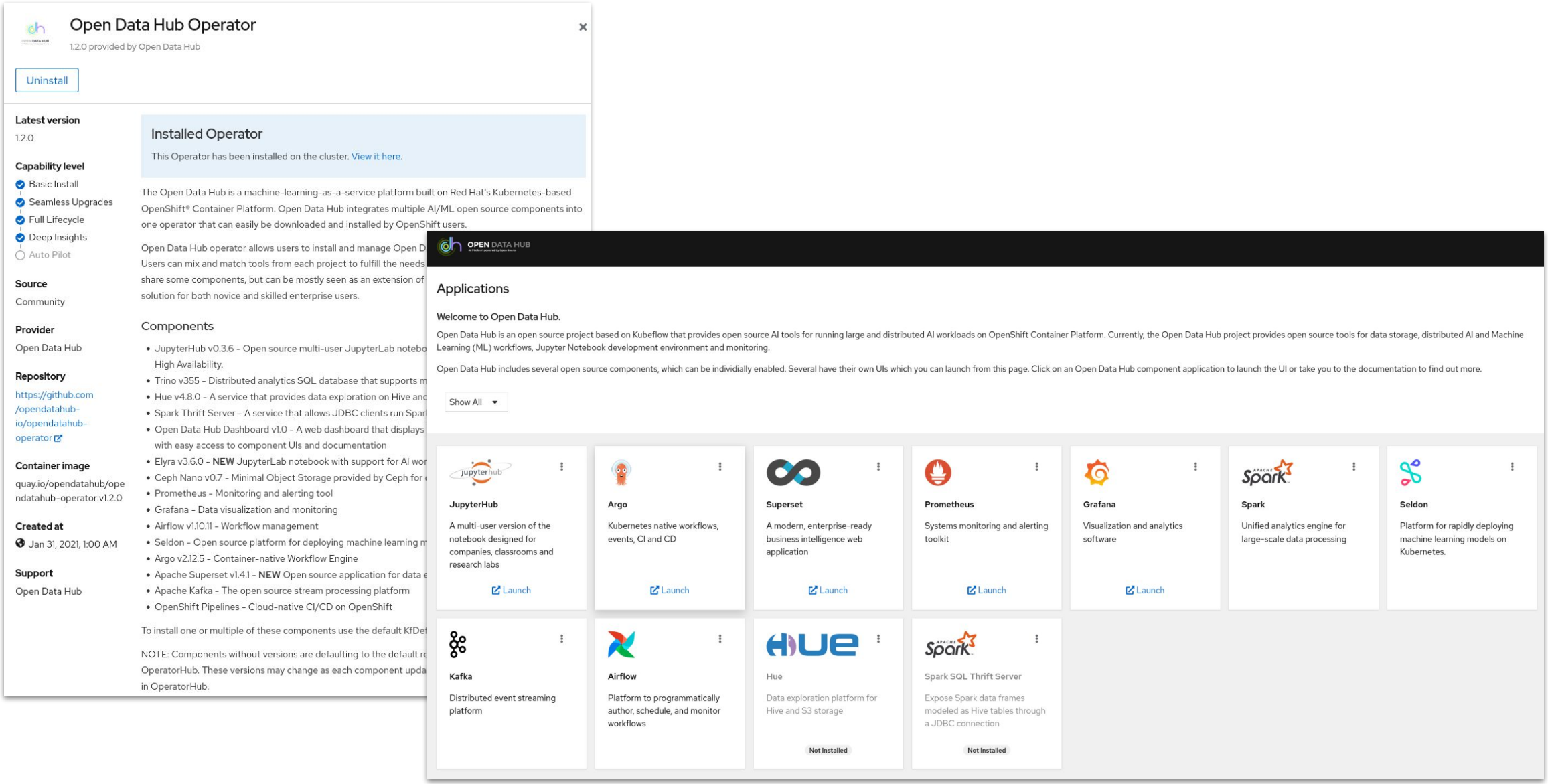
- Distributed object store
- S3 Interface

- Distributed event streaming
- Pub/sub messaging

- Distributed SQL engine
- Integrate heterogeneous data sources

- Data management
- Data lineage & versioning

Red Hat


Open Data Hub Operator
1.2.0 provided by Open Data Hub
Uninstall
Latest version
1.2.0
Capability level
Basic Install
Seamless Upgrades
Full Lifecycle
Deep Insights
Auto Pilot
Source
Community
Provider
Open Data Hub
Repository
https://github.com/opendatahub-io/opendatahub-operator
Container image
quay.io/opendatahub/opendatahub-operator:v1.2.0
Created at
Jan 31, 2021, 1:00 AM
Support
Open Data Hub
Installed Operator
This Operator has been installed on the cluster. View it here.
The Open Data Hub is a machine-learning-as-a-service platform built on Red Hat's Kubernetes-based OpenShift® Container Platform. Open Data Hub integrates multiple AI/ML open source components into one operator that can easily be downloaded and installed by OpenShift users.
Components
JupyterHub v0.3.6 - Open source multi-user JupyterLab
High Availability.
Trino v355 - Distributed analytics SQL database that supports
Hue v4.8.0 - A service that provides data exploration on Hive
Spark Thrift Server - A service that allows JDBC clients run
Open Data Hub Dashboard v1.0 - A web dashboard that displays
with easy access to component UIs and documentation
Elyra v3.6.0 - NEW JupyterLab notebook with support for AI
Ceph Nano v0.7 - Minimal Object Storage provided by Ceph
Prometheus - Monitoring and alerting tool
Grafana - Data visualization and monitoring
Airflow v1.10.11 - Workflow management
Seldon - Open source platform for deploying machine learning
Argo v2.12.5 - Container-native Workflow Engine
Apache Superset v1.4.1 - NEW Open source application for data
Apache Kafka - The open source stream processing platform
OpenShift Pipelines - Cloud-native CI/CD on OpenShift
OPEN DATA HUB
Applications
Welcome to Open Data Hub.
Open Data Hub is an open source project based on Kubeflow that provides open source AI tools for running large and distributed AI workloads on OpenShift Container Platform. Currently, the Open Data Hub project provides open source tools for data storage, distributed AI and Machine Learning (ML) workflows, Jupyter Notebook development environment and monitoring.
Open Data Hub includes several open source components, which can be individually enabled. Several have their own UIs which you can launch from this page. Click on an Open Data Hub component application to launch the UI or take you to the documentation to find out more.
Show All
JupyterHub
A multi-user version of the notebook designed for companies, classrooms and research labs
Launch
Argo
Kubernetes native workflows, events, CI and CD
Launch
Superset
A modern, enterprise-ready business intelligence web application
Launch
Prometheus
Systems monitoring and alerting toolkit
Launch
Grafana
Visualization and analytics software
Launch
Spark
Unified analytics engine for large-scale data processing
Seldon
Platform for rapidly deploying machine learning models on Kubernetes.
Kafka
Distributed event streaming platform
Airflow
Platform to programmatically author, schedule, and monitor workflows
Hue
Data exploration platform for Hive and S3 storage
Not Installed
Spark SQL Thrift Server
Expose Spark data frames modeled as Hive tables through a JDBC connection
Not Installed

Strategic partnerships within AI/ML ecosystem

AI/ML Lifecycle

ANACONDA
C3.ai
CLOUDERA
cnvrg.io
CognitiveScale
THE TRUSTED AI COMPANY
GIGASPACES
H2O.ai
IBM
Labs.ai
NVIDIA
PerceptiLabs
SAS
SELDON
UBIX

Data Governance & Security

ANACONDA
CLOUDERA
Lightbend
CognitiveScale
THE TRUSTED AI COMPANY
Pachyderm

Data Processing

IBM
CLOUDERA
H2O.ai
Lightbend
PerceptiLabs
CONFLUENT
Starburst

Data Analytics

CLOUDERA
SAP
CAMBRIDGE SEMANTICS
Microsoft
SAS
Palantir

Databases

Couchbase
crunchy data
IBM
memSQL
Microsoft
mongoDB.
NUODB
redislabs

AI Ops

Lightbend
PROPHETSTOR
run:ai

Infrastructure Partners

DELL
Hewlett Packard Enterprise
intel
PENGUIN COMPUTING
SUPERMICRO

Hardware Acceleration

intel OpenVINO
NVIDIA
Red Hat