Data Science Model Lifecycle Explained

Audrey Reznik
Sr Principal Software Engineer
Red Hat

Hema Veeradhi
Sr Software Engineer
Red Hat

Oindrilla Chatterjee
Sr Data Scientist
Red Hat

Speakers

Audrey Reznik

Sr Principal Software Engineer

Red Hat

Hema Veeradhi

Sr Software Engineer

Red Hat

Oindrilla Chatterjee

Sr Data Scientist

Red Hat

Red Hat

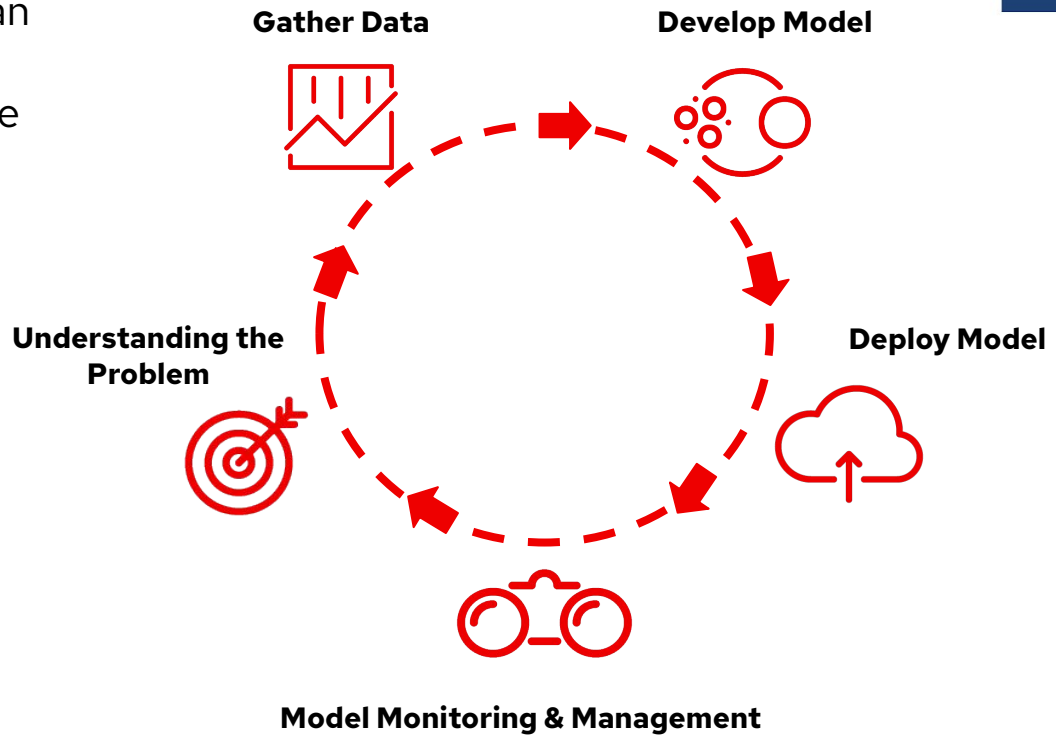
Agenda

- What is a Data Science Model lifecycle?
- Data Engineering & Its importance
- Data Science Model Lifecycle
  - onPrem with Open Data Hub (ODH)
  - in the public (or hybrid) cloud with Red Hat OpenShift Data Science (RHODS)

Red Hat

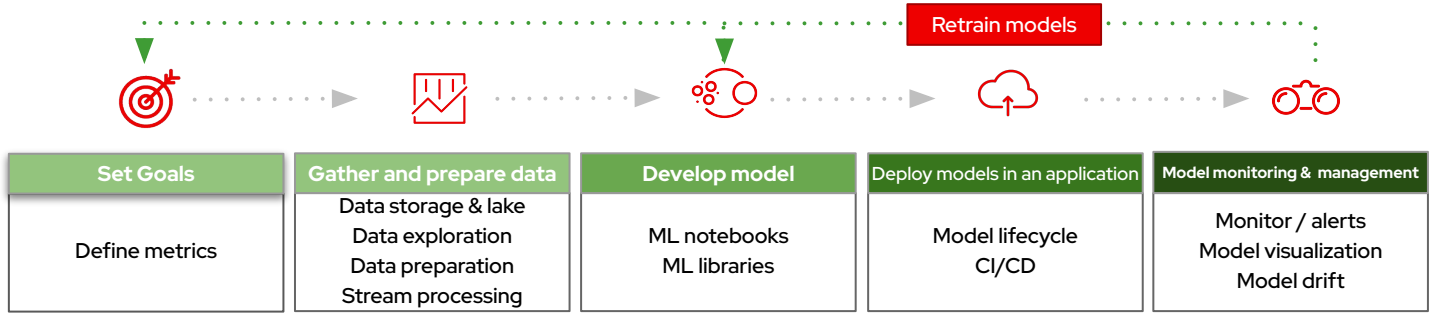
What is a Data Science Model lifecycle?

"A data science life cycle is an **iterative set of steps** you take to deliver a data science product/ML model"



Red Hat

Data Science Model Lifecycle steps & Personas



Red Hat

Data engineering

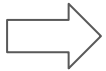
Get and store
the data

ceph

trino

Red Hat

Data engineering



Get and store the data

Filter the data

ceph

trino

Red Hat

Data engineering

Get and store the data

Filter the data

Enrich the data

ceph

trino

Red Hat

Data engineering

Get and store the data

Filter the data

Enrich the data

Find Insights and Visualize

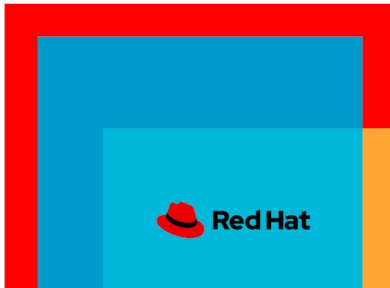
ceph

trino

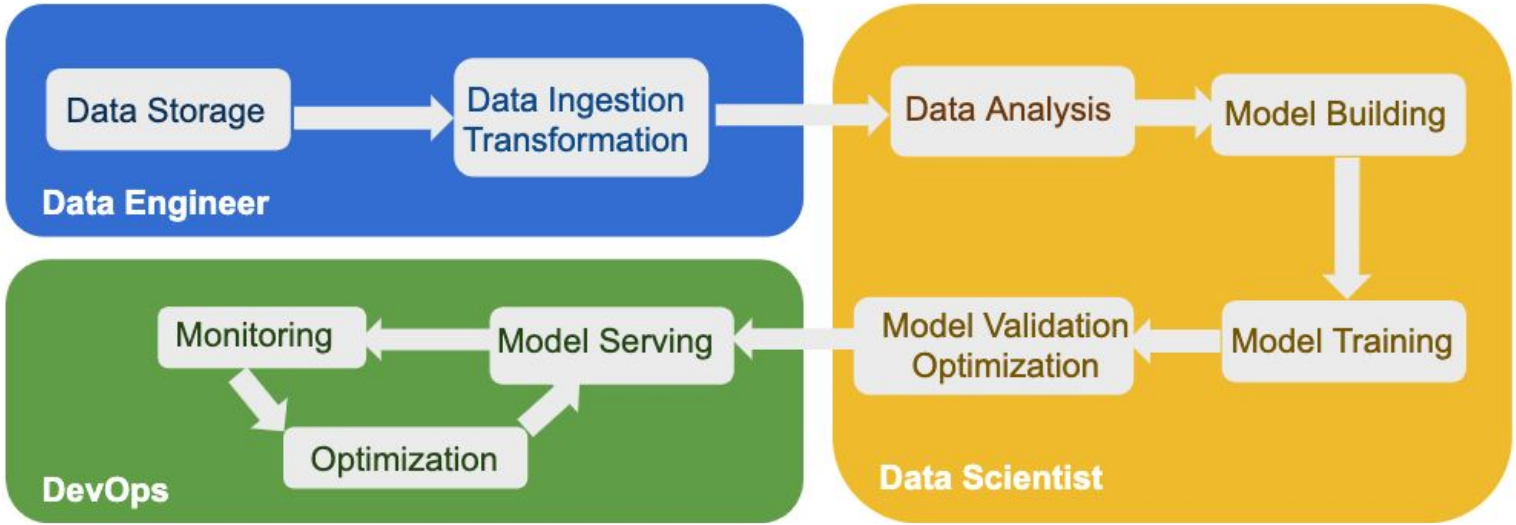
Superset

Red Hat

Questions so far...



Open Data Hub is an open source community AI/ML platform on Openshift

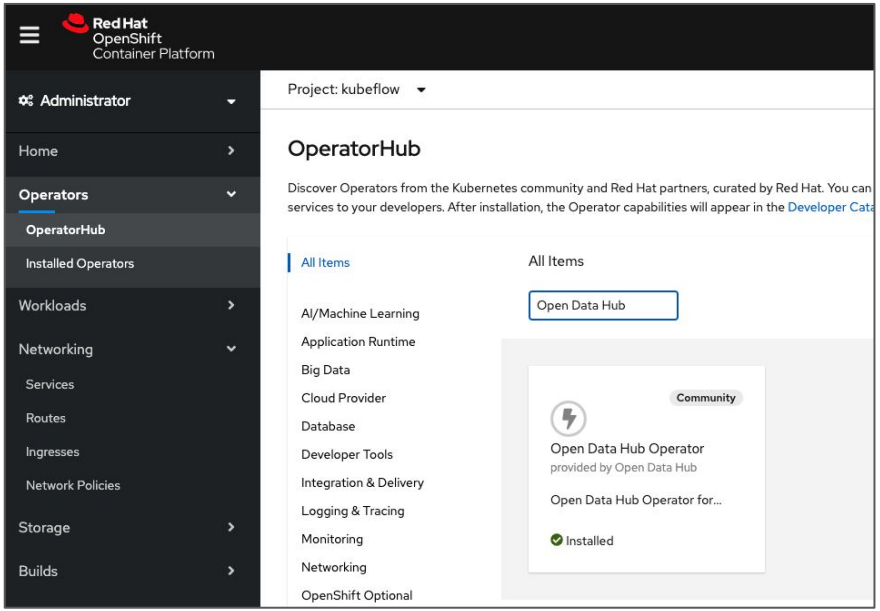


Red Hat

Open Data Hub is an **Operator** installed from OpenShift **OperatorHub**

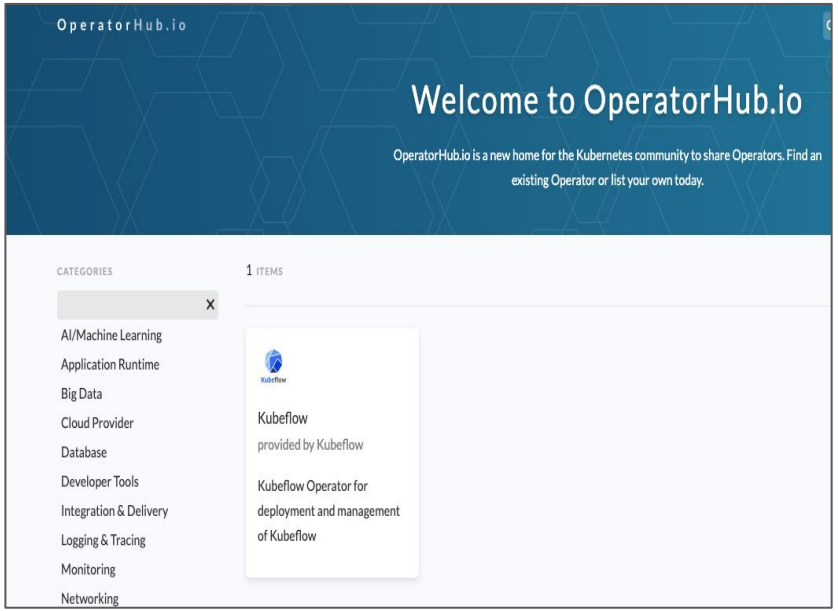
✅ Yes

OPENSHIFT



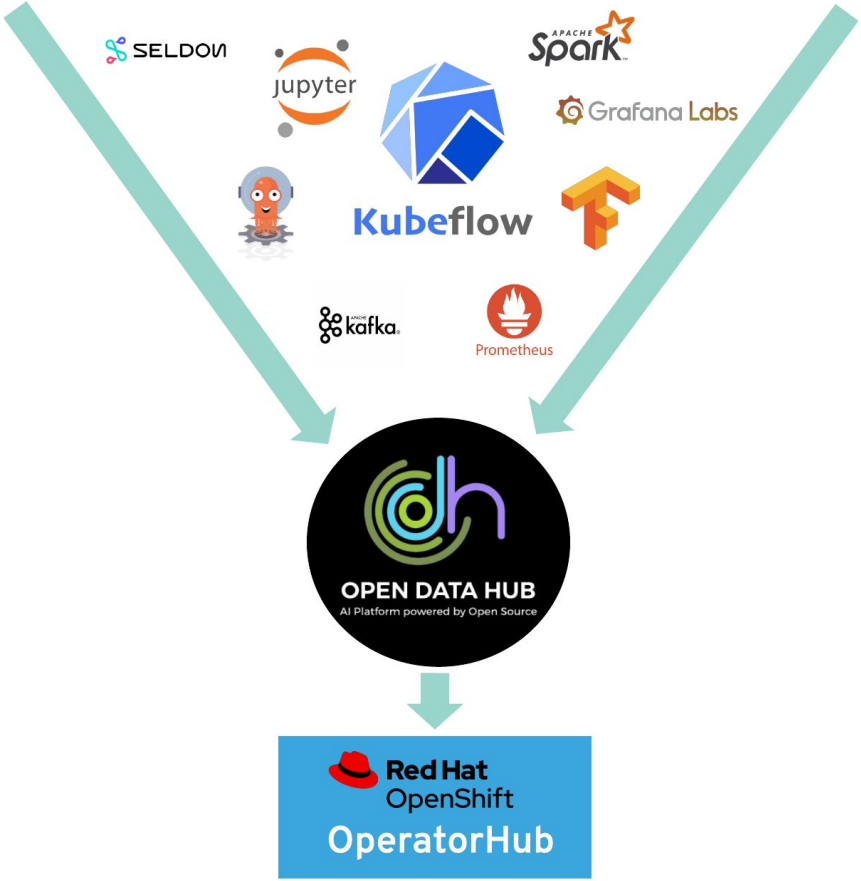
❌ No

kubernetes



Red Hat

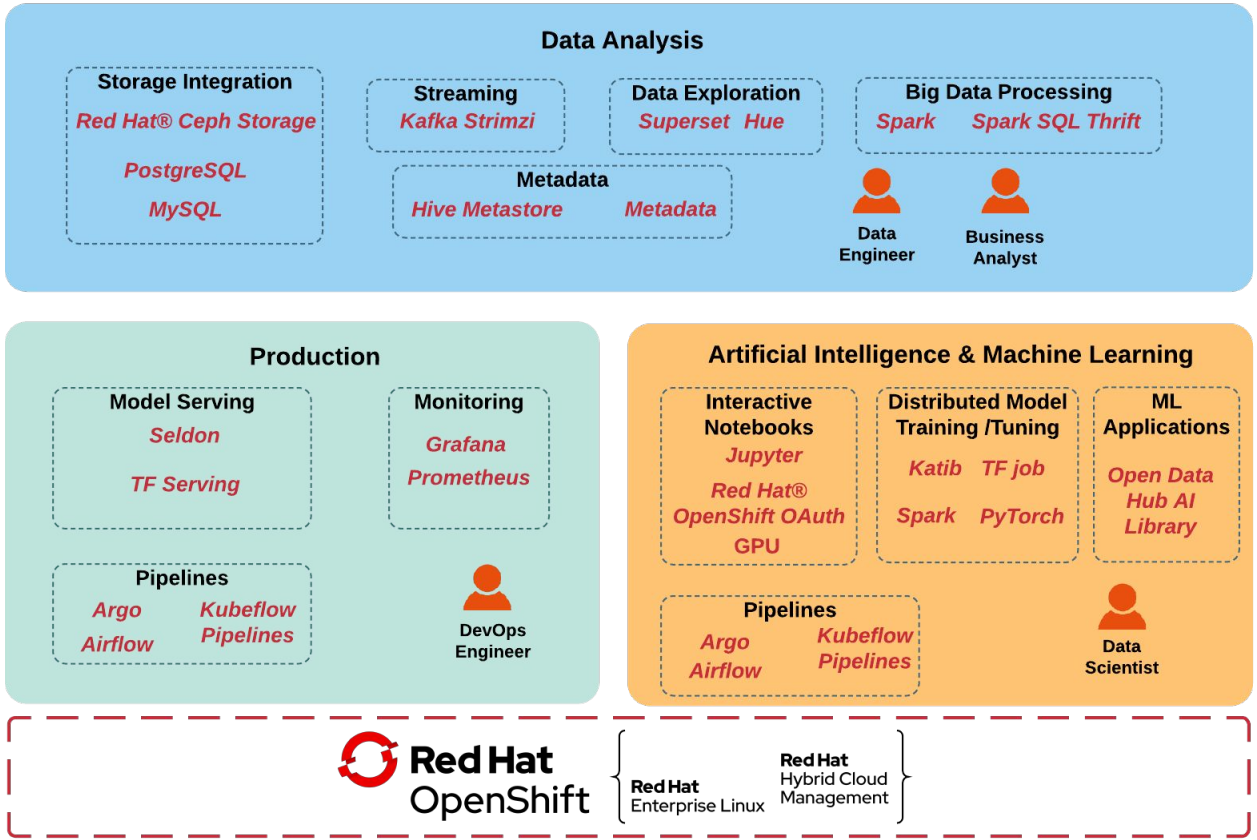
Open Data Hub integrates open source projects into an **end-to-end AI/ML platform** on OpenShift



Red Hat

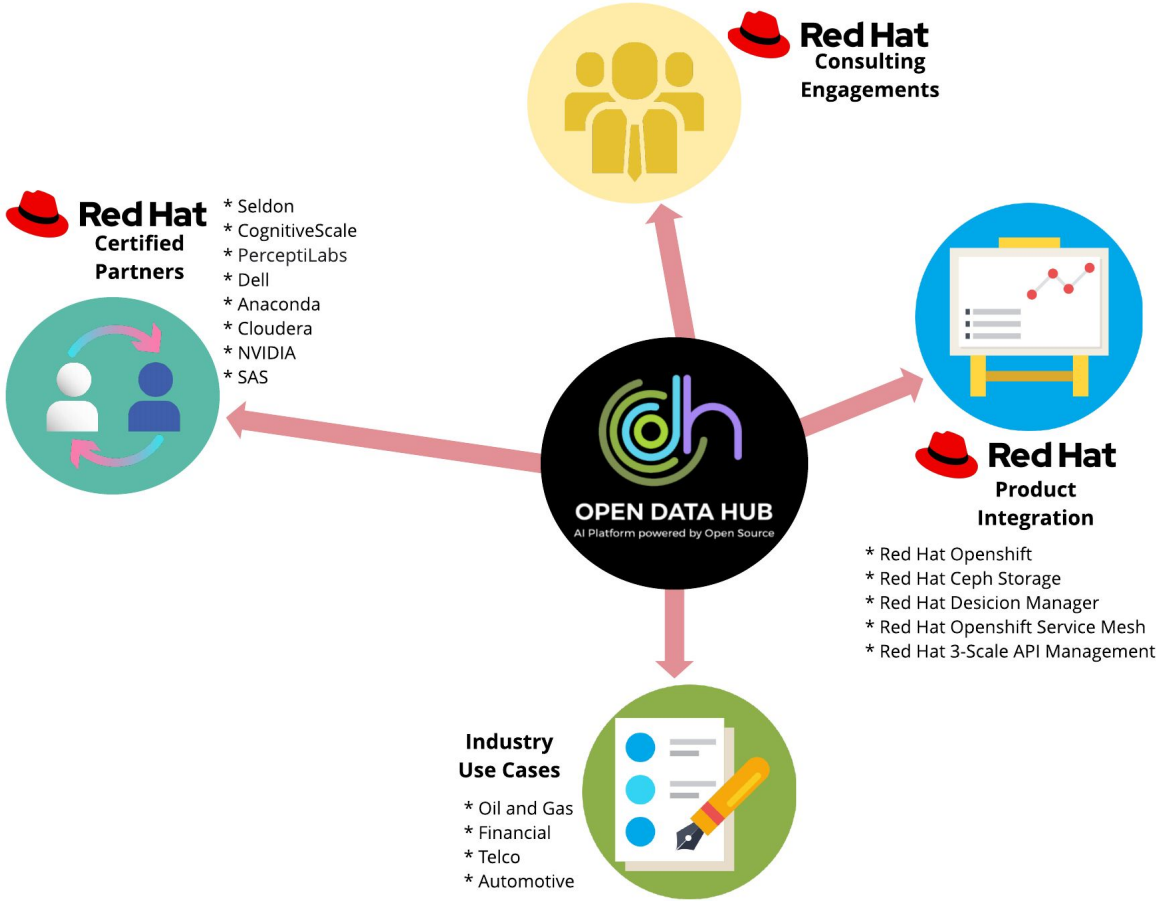
OPEN DATA HUB

AI Platform powered by Open Source



Red Hat

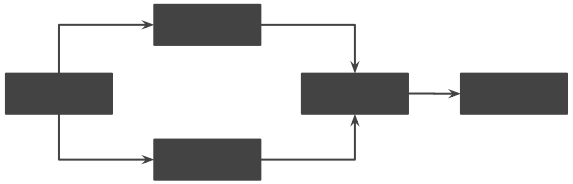
Open Data Hub comes with an ecosystem provided by **Red Hat** and **Certified Partners**



Red Hat

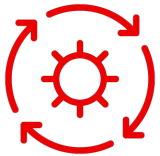
AI/ML Pipelines

A reproducible workflow that can be used to automate repetitive steps in a machine learning workflow.



Red Hat

Why AI/ML Pipelines



Automating Manual Tasks

Better Reproducibility

Monitoring and Tracking

Red Hat

Tools for AI/ML Pipelines

Elyra

Notebook pipelines visual editor

Kubeflow Pipelines

Build and deploy scalable ML workflows

Based on Open Data Hub and Operate First

Upstream code enhanced with operational excellence

Open Data Hub

Community driven upstream meta-project demonstrating AI/ML platform on Red Hat OpenShift comprised of open source projects

Operate First

Subset Open Data Hub operated at scale for community and university audiences to infuse operational excellence

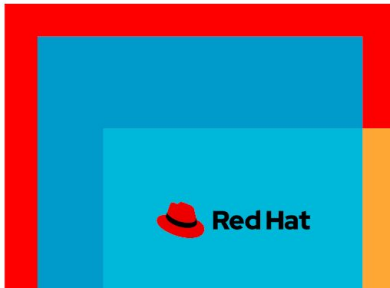
Red Hat OpenShift Data Science

Subset of Operate First delivered as a cloud service on Red Hat OpenShift Managed on Amazon Web Services with optional ISV offerings

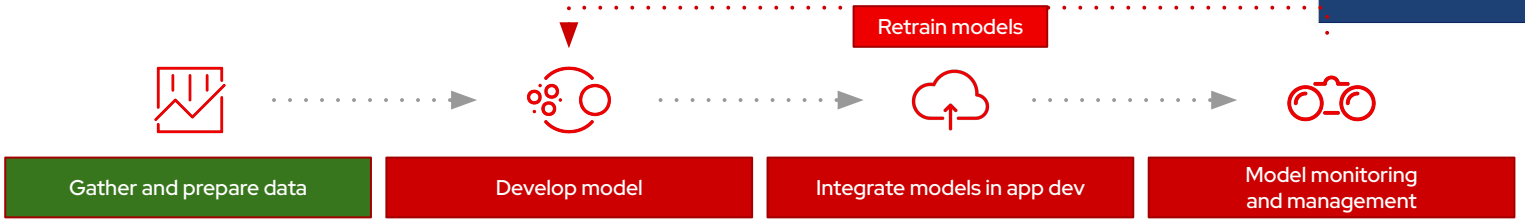
Open Data Hub: https://opendatahub.io
Operate First: https://openinfralabs.org/ and https://www.operate-first.cloud/

Red Hat

Questions so far...



Red Hat OpenShift Data Science & our Partner Ecosystem

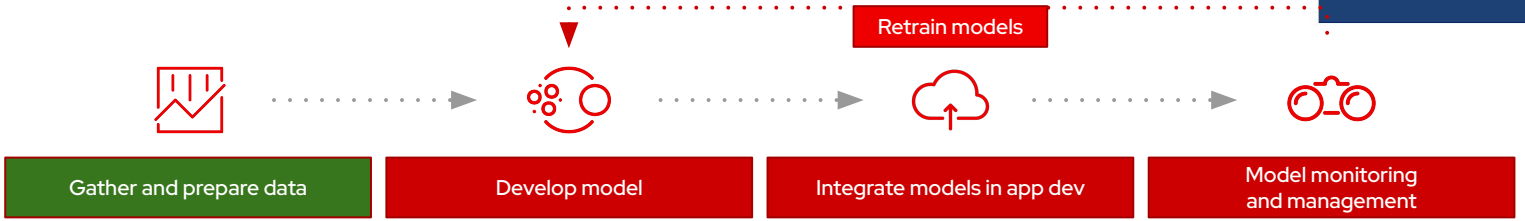


Red Hat managed cloud platform

Open hybrid cloud platform with self service capabilities	Red Hat OpenShift Dedicated	Red Hat OpenShift Service on Amazon Web Services
Accelerators (NVIDIA GPUs)	NVIDIA	intel
Cloud infrastructure	powered by aws	

Red Hat

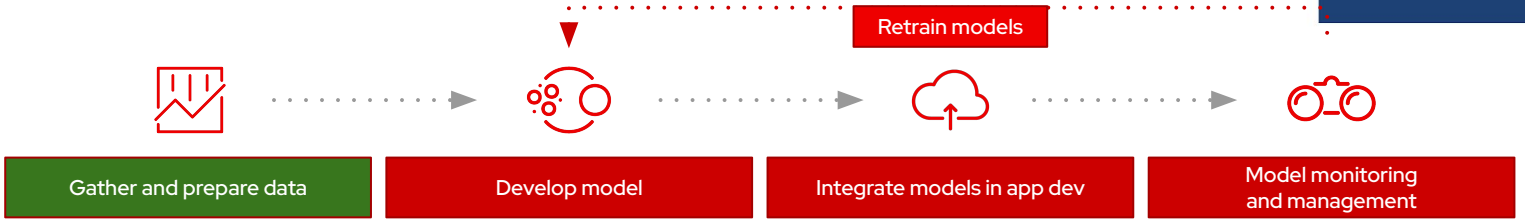
Red Hat OpenShift Data Science & our Partner Ecosystem



Red Hat managed cloud services	Red Hat OpenShift Streams for Apache Kafka	Red Hat OpenShift Data Science	Source to image	Red Hat OpenShift API Management
Red Hat managed cloud platform	Open hybrid cloud platform with self service capabilities		Red Hat OpenShift Dedicated	Red Hat OpenShift Service on Amazon Web Services
	Accelerators (NVIDIA GPUs)		NVIDIA	intel
	Cloud infrastructure		powered by aws	

Red Hat

Red Hat OpenShift Data Science & our Partner Ecosystem



ISV managed cloud services	Starburst Galaxy	ANACONDA. COMMERCIAL EDITION	
Red Hat managed cloud services	Red Hat OpenShift Streams for Apache Kafka	Red Hat OpenShift Data Science	Source to image / Red Hat OpenShift API Management
Red Hat managed cloud platform	Open hybrid cloud platform with self service capabilities	Red Hat OpenShift Dedicated	Red Hat OpenShift Service on Amazon Web Services
	Accelerators (NVIDIA GPUs)	NVIDIA.	intel
	Cloud infrastructure	powered by aws	

Red Hat

Red Hat OpenShift Data Science & our Partner Ecosystem

	Gather and prepare data	Develop model	Integrate models in app dev	Model monitoring and management
Customer managed ISV software	Pachyderm	intel oneAPI, IBM Watson	intel OpenVINO	
ISV managed cloud services	Starburst Galaxy	ANACONDA. COMMERCIAL EDITION		
Red Hat managed cloud services	Red Hat OpenShift Streams for Apache Kafka	Red Hat OpenShift Data Science	Source to image	Red Hat OpenShift API Management
Red Hat managed cloud platform	Open hybrid cloud platform with self service capabilities		Red Hat OpenShift Dedicated	Red Hat OpenShift Service on Amazon Web Services
	Accelerators (NVIDIA GPUs)		NVIDIA.	intel
	Cloud infrastructure		powered by aws	

Red Hat

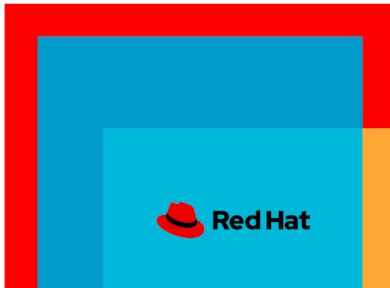
… Red Hat Data Science is award winning for AI/ML

- OpenShift Data Science selected as the winner of the 2022 "Overall Data Science Solution of the Year"
- Data Breakthrough Awards - program founded to recognize the data technology innovators, leaders and visionaries in categories, including Data Analytics, Big Data, Business Intelligence, Data Storage and more
- More than 1850 nominations

- Why Red Hat OpenShift Received the EMA Top 3 Award for MLOps
- The Red Hat OpenShift open hybrid cloud platform enables organizations to rapidly spin up new machine learning stacks and move machine learning models through the MLOps pipeline
- Open Data Hub Operator turns OpenShift into a Kubernetes-based machine learning as a service platform
- Red Hat OpenShift Data Science offers a turnkey managed data science environment

Red Hat

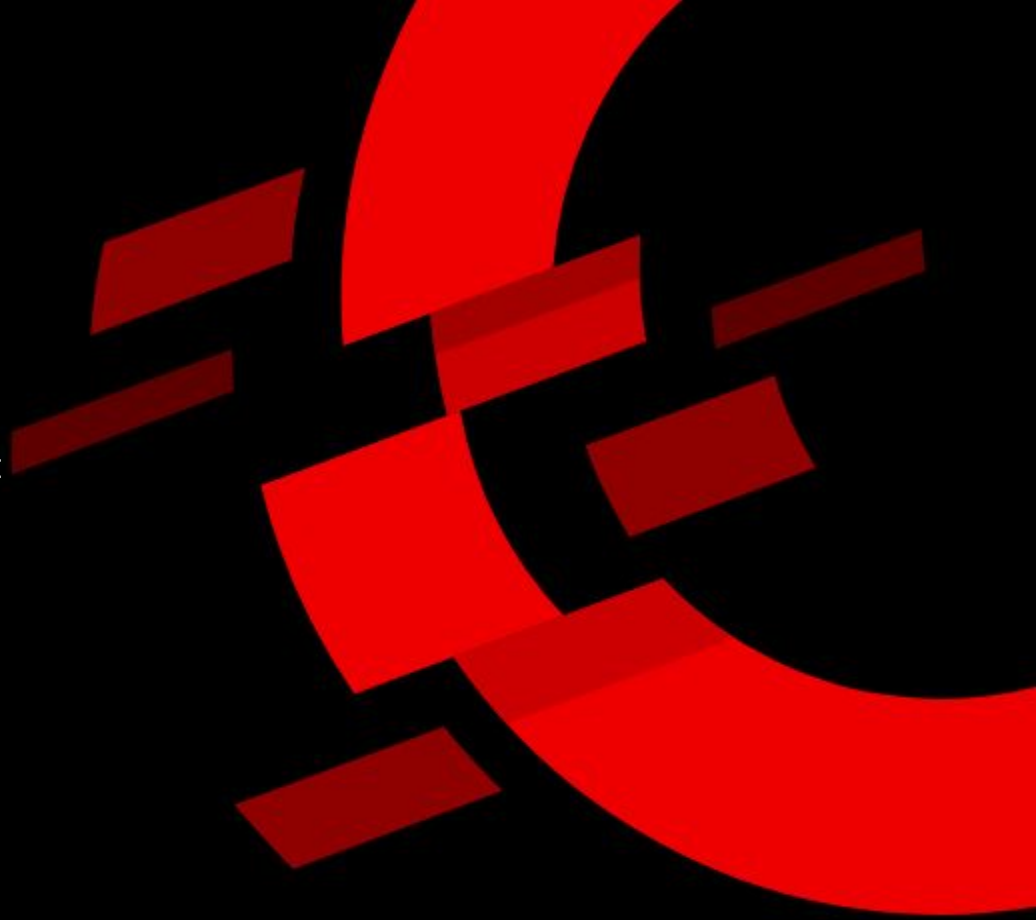
Questions so far...



Explore an open platform and infinite possibilities

Red Hat® OpenShift® 4 represents a significant change in the way that OpenShift clusters are created, deployed, and managed.

BOOK A DISCOVERY SESSION



Red Hat

Red Hat Summit

Thank you

linkedin.com/company/red-hat

facebook.com/redhatinc

youtube.com/user/RedHatVideos

twitter.com/RedHat

Red Hat