

Introduction to Red Hat OpenShift AI

MLOps Platform for artificial intelligence/
machine learning (AI/ML) use cases

Eoin Crosbie
Technical Partner Enablement Manager



Red Hat OpenShift AI

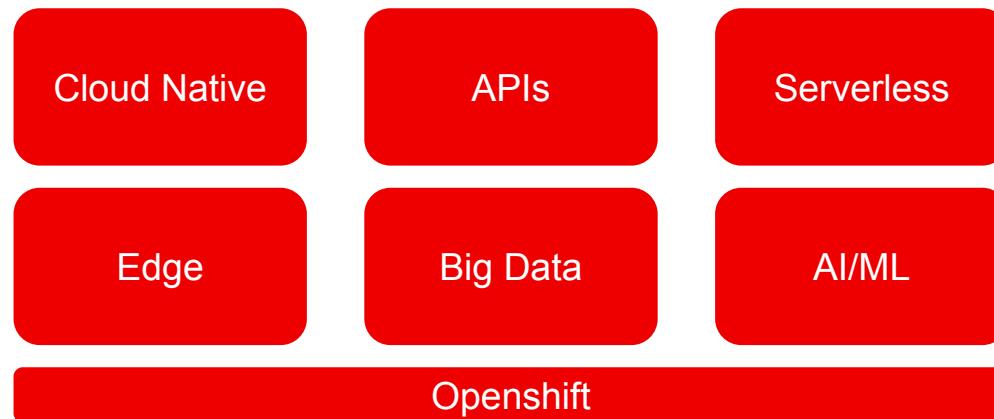
1. Why

2. What

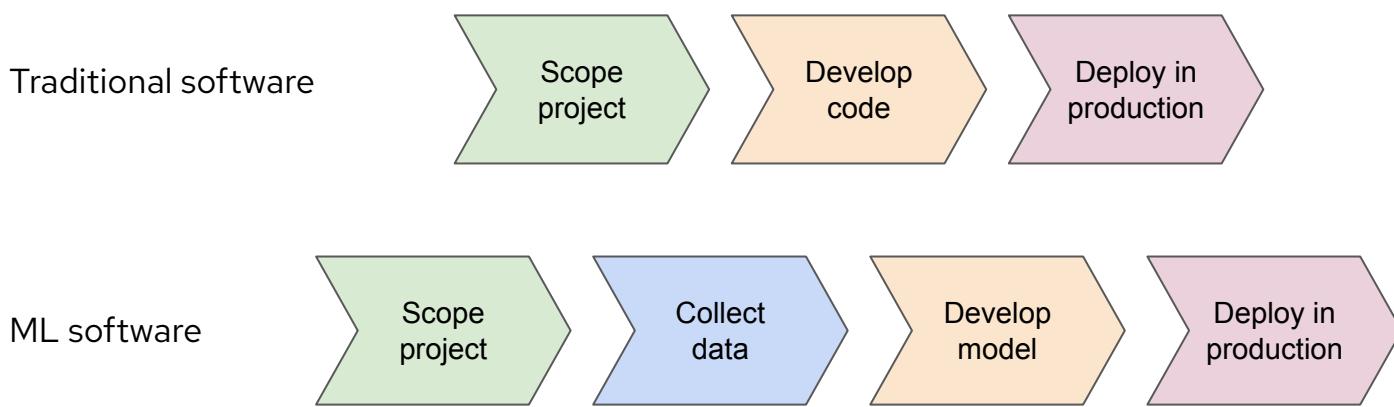
3. Try

WHY



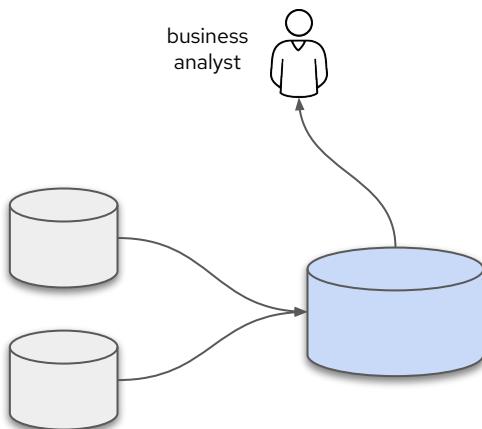


Traditional software vs. ML software

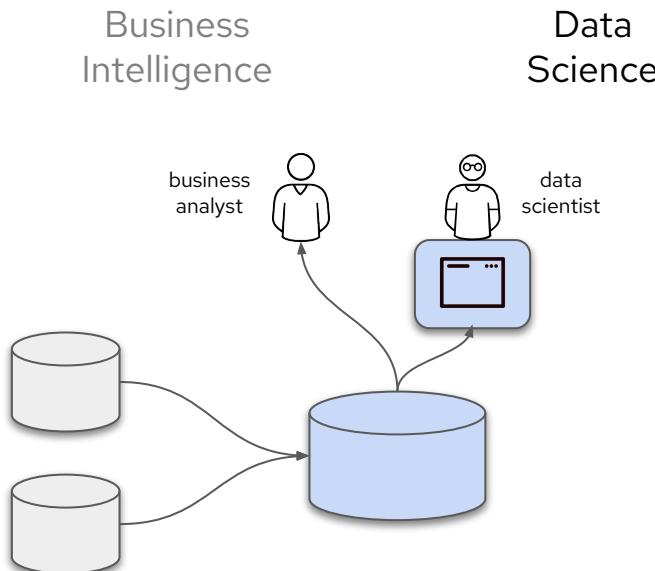


The AI Journey

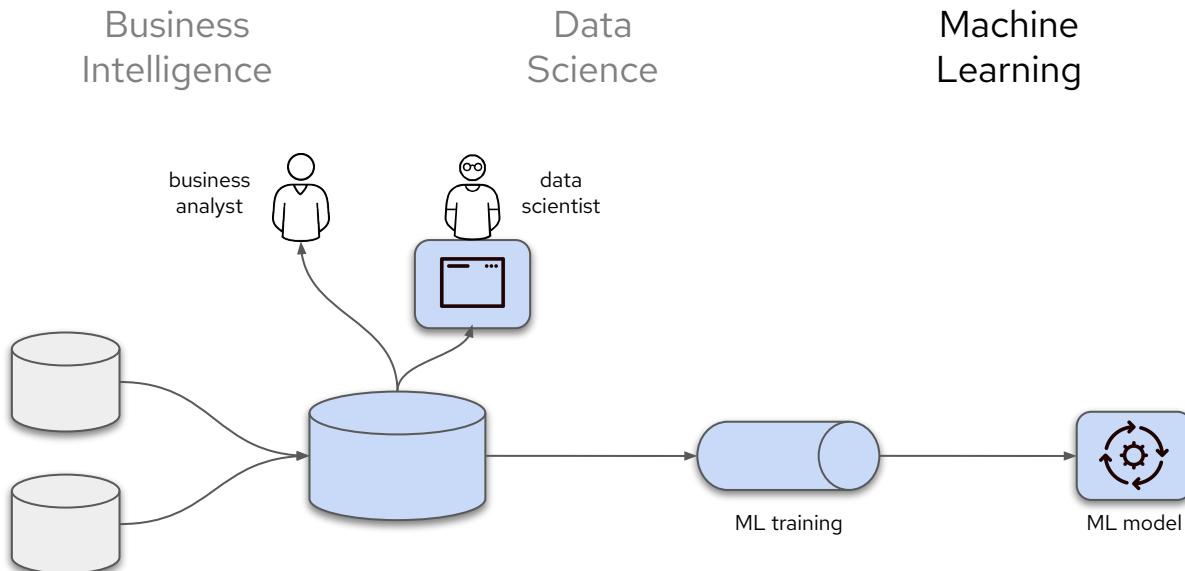
Business
Intelligence



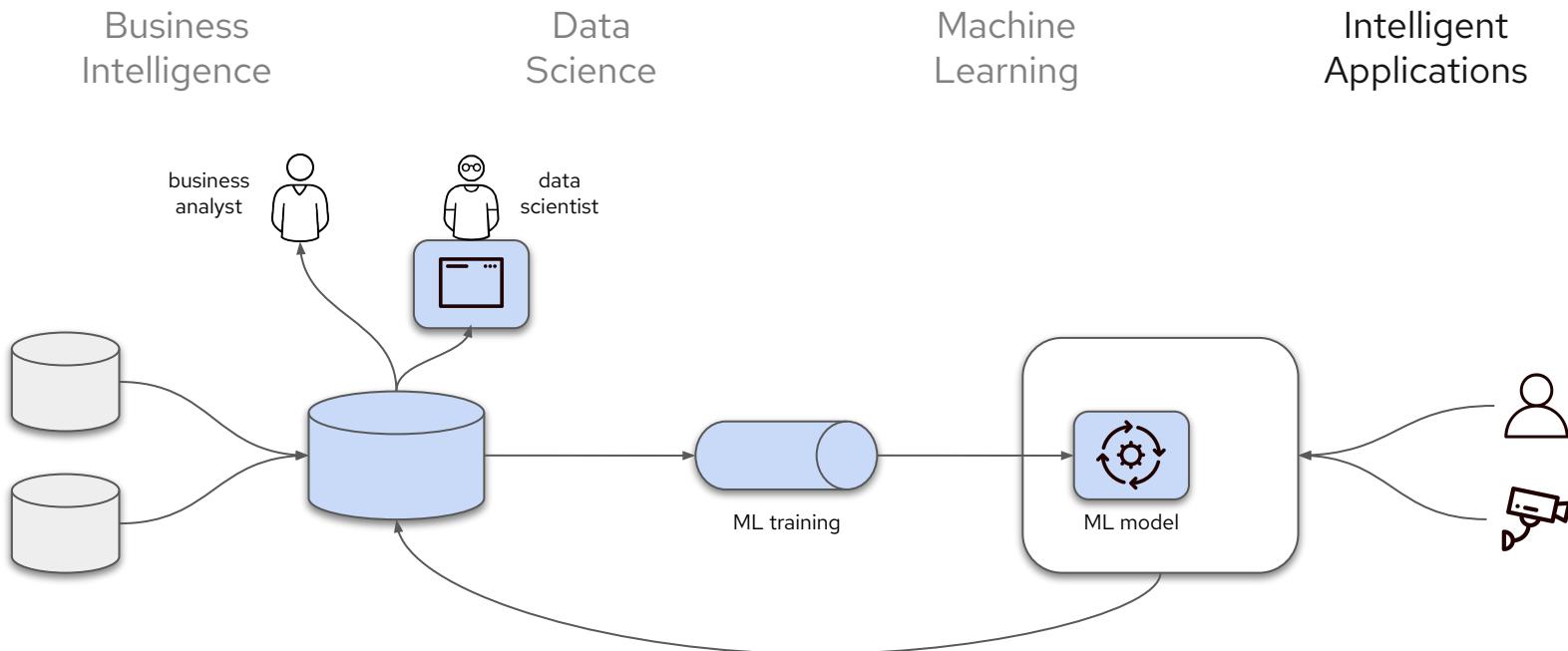
The AI Journey



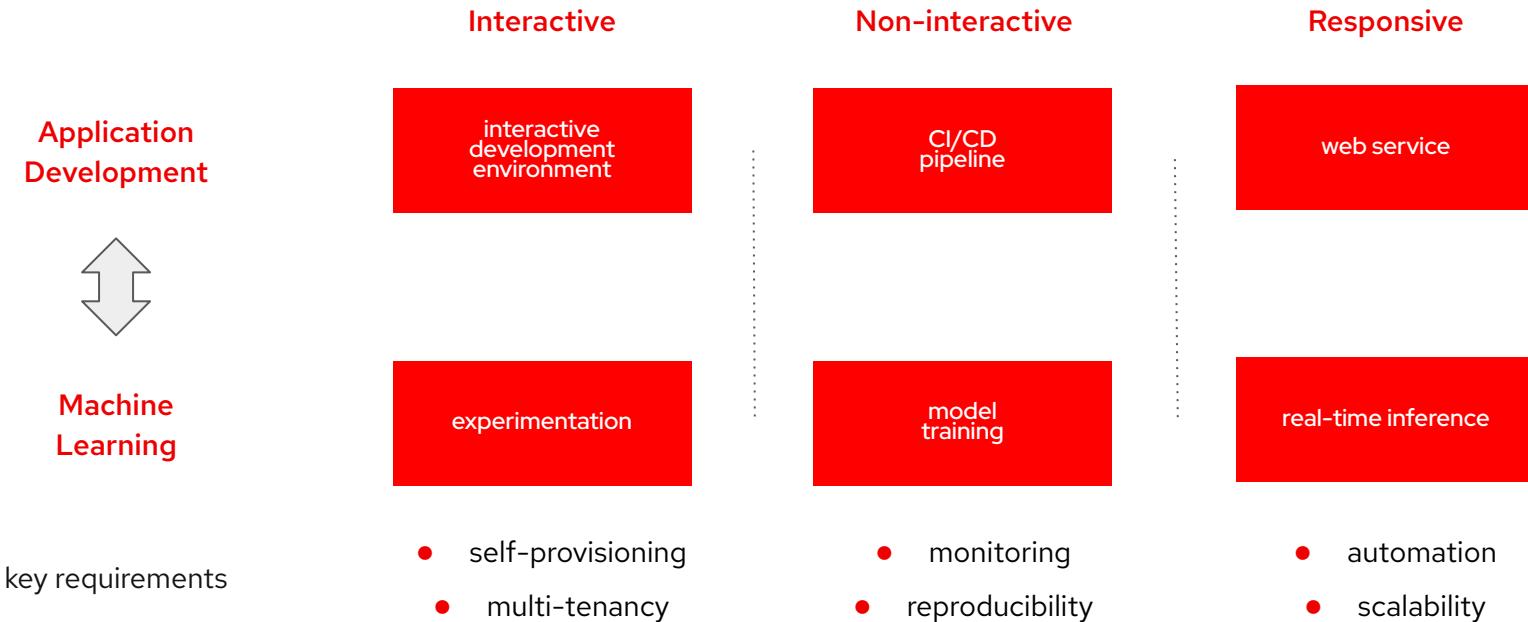
The AI Journey



The AI Journey

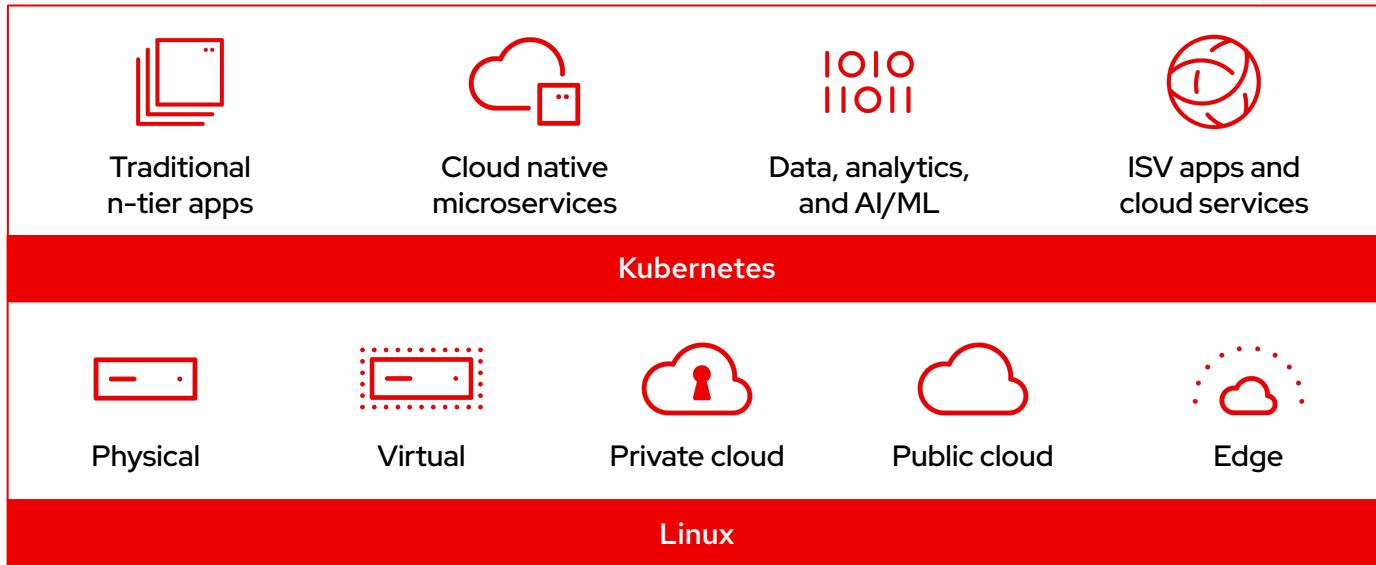


Workload Types



Open hybrid cloud

Red Hat helps customers build, run, and manage applications everywhere



Our AI/ML strategy



AI workload support

Support **AI workload requirements** on Red Hat platforms

e.g., *hardware acceleration, GPU Operator*



Platform for AI-enabled apps

Provide a consistent, hybrid cloud **application platform for customers** to build, train, and deploy AI-enabled applications

e.g., *Red Hat OpenShift AI*



AI-enabled platforms

Use **AI models, tools, and services** to accelerate **adoption** of existing Red Hat products and services

e.g., *Red Hat Ansible Lightspeed, Red Hat Developer Hub*

2024 AI Awards for OpenShift AI

MERIT AWARDS 2024 Telecom

Congratulations to the Merit Awards Winners

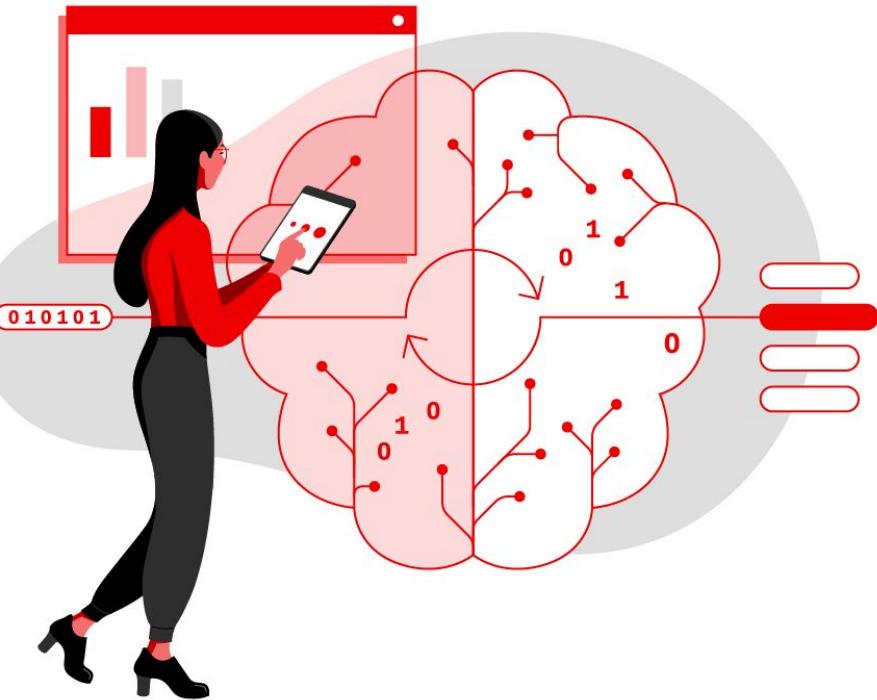
AI: Technology

Gold: Red Hat



2024 Finalist - Artificial Intelligence Excellence Awards

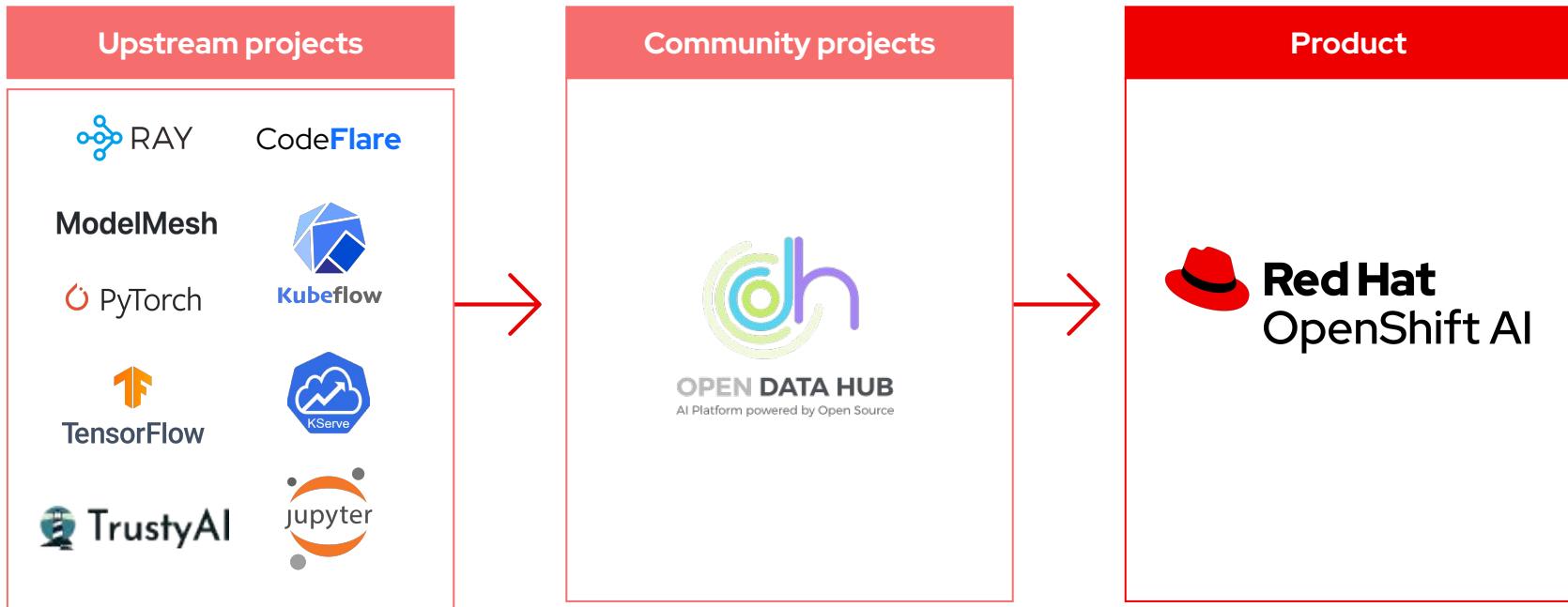
WHAT



Red Hat OpenShift AI

An AI-focused platform that provides tools to train, tune, serve, monitor and manage AI/ML experiments and models.

Red Hat's AI/ML engineering is 100% open source



Red Hat OpenShift AI

Hybrid MLOps platform

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

Now available as fully managed cloud service or traditional software product on-prem or in the cloud!



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.



Lifecycle Management

Create repeatable data science pipelines for model training and validation and integrate them with devops pipelines for delivery of models across your enterprise.



Increased capabilities / collaboration

Create projects and share them across teams. Combine Red Hat components, open source software, and ISV certified software.

AI for the Open Hybrid Cloud



Enterprise grade hybrid AI and MLOps platform

Train, serve, monitor and manage the lifecycle of AI/ML models and applications, from experiments to production

 **Red Hat**
OpenShift

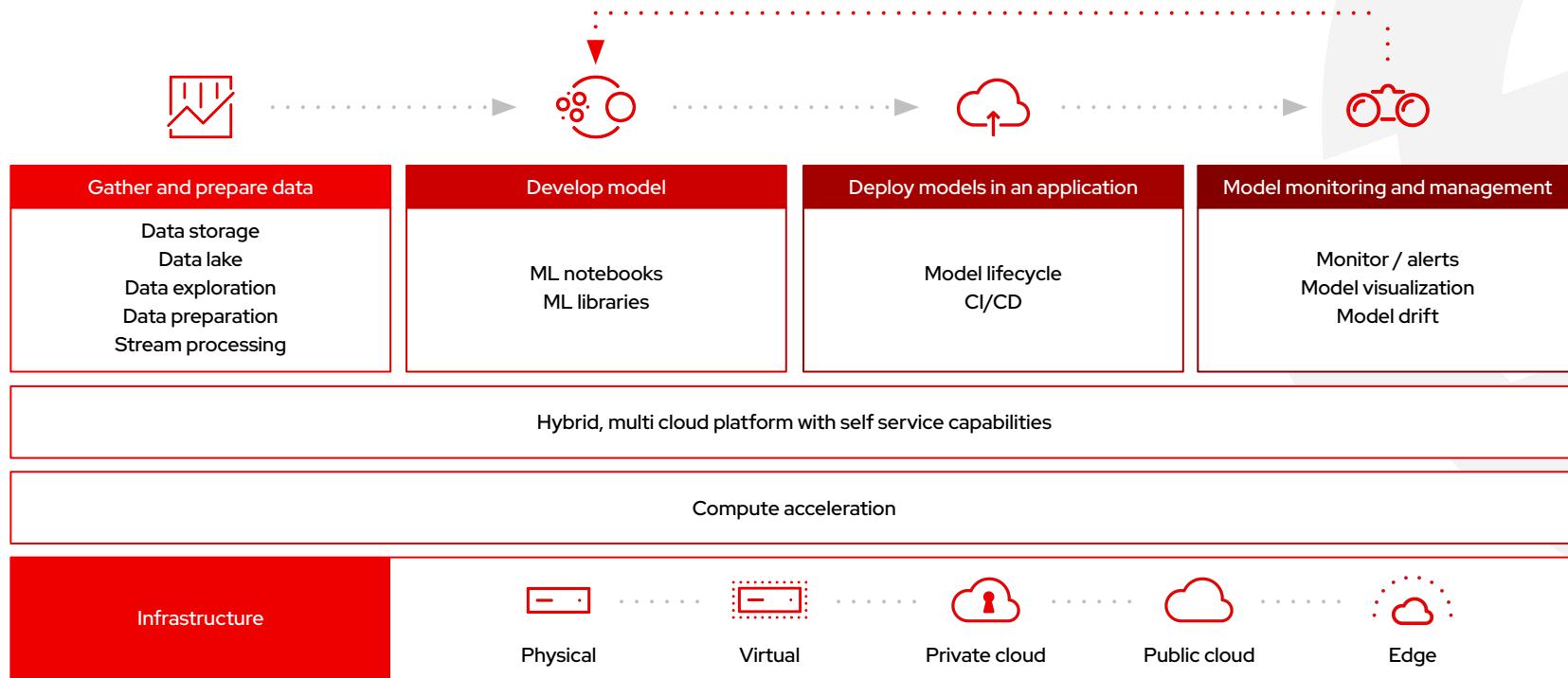
 **Red Hat**
OpenShift AI

Red Hat OpenShift AI

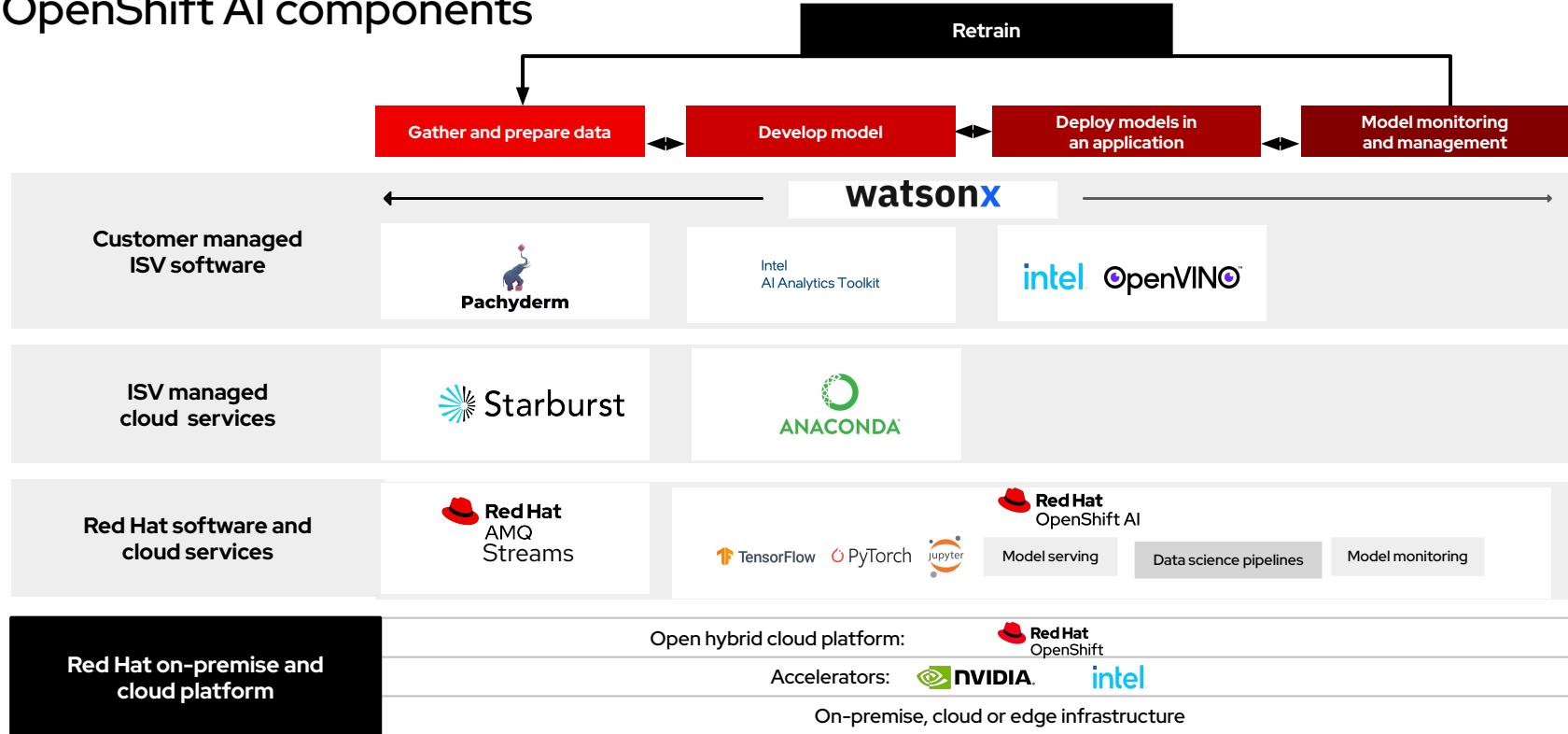
Expands upon the proven capabilities of Red Hat OpenShift and Red Hat OpenShift Data Science, to:

- ▶ Provide a unified platform for data scientists, application developers and IT Ops
- ▶ Scale to handle workload demands of foundation models (volume of data, duration of training run, size of model, acceleration required, and scalability).
- ▶ Deliver consistency, ease-of-use, and cloud-to-edge deployment options.
- ▶ Underlying platform for training, serving and tuning foundation models for IBM watsonx.ai and Red Hat Ansible Lightspeed with Watson Code Assistant

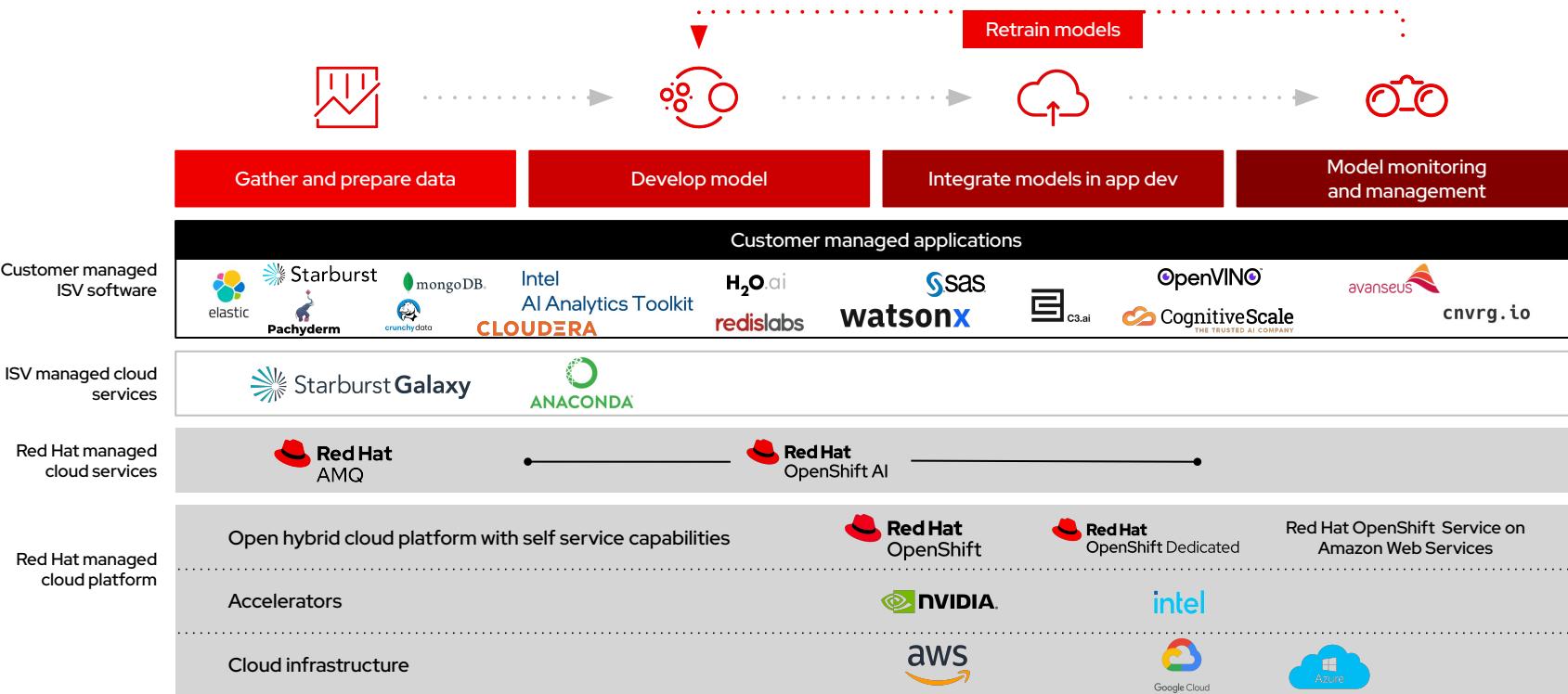
Conceptual machine learning architecture



OpenShift AI components



... and integrating our partner ecosystem



Dashboard user interface

The screenshot displays the Red Hat OpenShift Data Science dashboard interface. The left sidebar contains navigation links: Applications (Enabled, Explore), Data Science Projects, Model Serving, Resources, and Settings.

Explore Section: This section allows users to add optional applications to their instance. It includes four cards:

- Anaconda Professional** (Partner managed) by Anaconda: A popular open-source package distribution and environment management system.
- IBM Watson Studio** (Self-managed) by IBM: A platform for embedding AI and machine learning.
- Intel® oneAPI AI Analytics Toolkit Container** (Self-managed) by Intel®: A multi-user version of the notebook designed for companies, classrooms, and research labs.
- Jupyter** (Red Hat managed) by Jupyter: A multi-user version of the notebook designed for companies, classrooms, and research labs.

Enabled Section: This section lists the currently enabled applications. It includes four cards:

- Intel® oneAPI AI Analytics Toolkit Container** (Self-managed): An AI Kit for accelerating end-to-end data science and analytics pipelines on Intel® architectures.
- Jupyter** (Red Hat managed): A multi-user version of the notebook designed for companies, classrooms, and research labs.
- NVIDIA GPU Add-on** (Self-managed): NVIDIA GPU Add-on prepares OpenShift to run GPU-accelerated workloads on a single node or on multiple nodes.
- OpenVINO** (Self-managed): OpenVINO is an open source toolkit to help optimize deep learning performance and deploy using an inference engine onto Intel® hardware.

A "Launch application" button is located at the bottom of the Enabled section.

Dashboard resources

The screenshot shows the Red Hat OpenShift Data Science dashboard. On the left, a sidebar menu includes sections for Applications (Enabled, Explore), Data Science Projects, Model Serving, Resources (selected), and Settings. The main content area is titled "Resources" and displays a grid of learning resources. The grid has two columns of headers: "All Items" and "All Items". It includes search and filter tools like "Sort by name" and "15 of 50 items". The resources are categorized by application: Jupyter, NVIDIA, and OpenVINO. Each resource card includes a thumbnail, title, author, duration, and status (e.g., Quick start, In Progress). A "Continue" or "Open" button is also present.

All Items	All Items
 Creating a Jupyter notebook by Jupyter Quick start 5 minutes In Progress	 Creating a Machine Learning Model using the NVIDIA GPU Add-on. by NVIDIA GPU Add-on Quick start 5 minutes Install GPU add-on
This quick start will walk you through creating a Jupyter notebook. Continue	 Open
 Getting started with Pachyderm concepts by Pachyderm Tutorial 15 minutes	 How to install Python packages on your notebook server by Jupyter How-to 15 minutes
 Open	 Open
 OpenVINO	How to serve a model using OpenVINO Model Server by OpenVINO How-to 10 minutes

What differentiates us



Hybrid cloud

Deploy models in containerized format for intelligent apps on-premise or in cloud



Easy to manage

Simple configurations on a secure and proven platform, that you can scale up or down with low effort



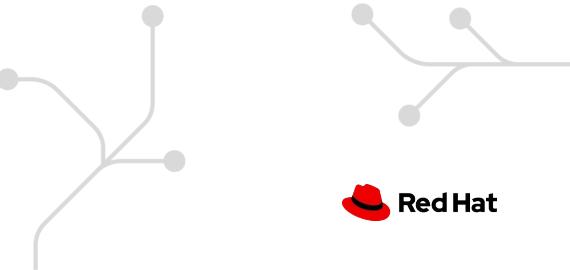
Collaborate

Collaborate on a common, extensible platform to bring IT, data science and application development teams together



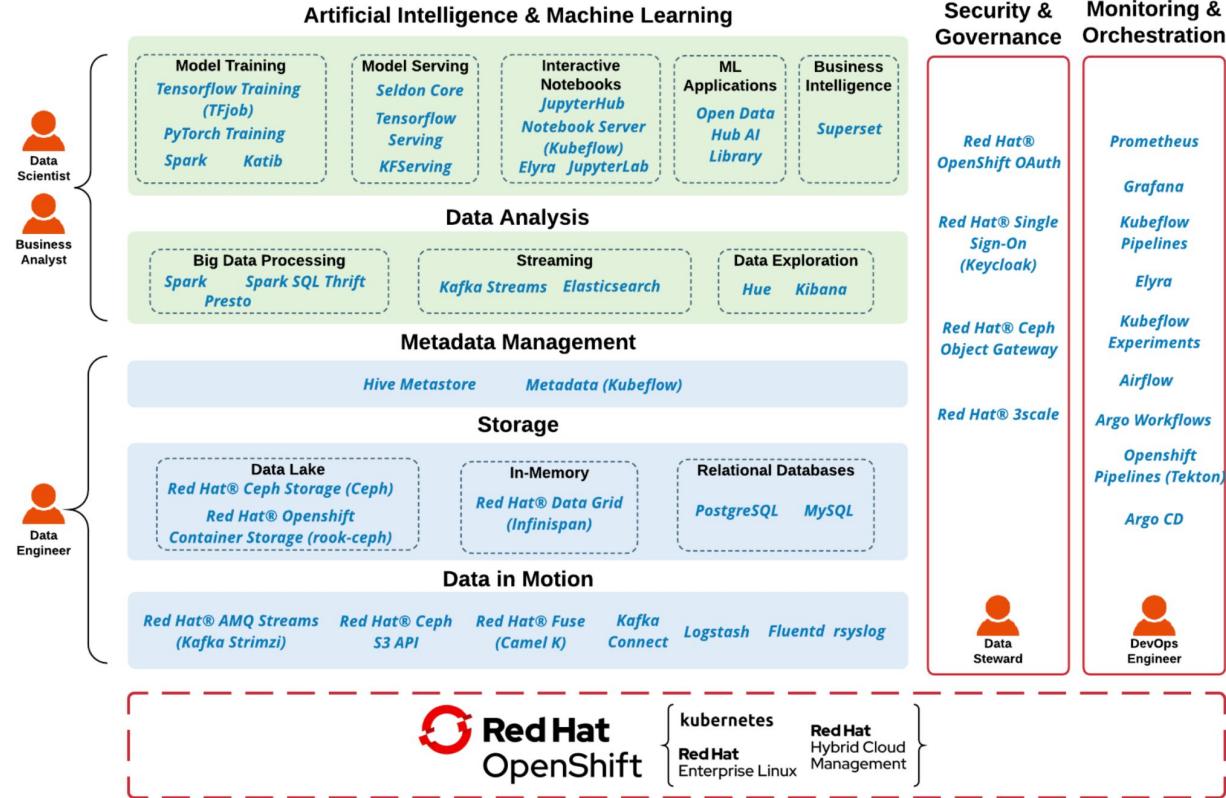
Open Source

Red Hat tracks changes and fixes to open source AI/ML tooling and enables customer access to upstream innovation





Reference Architecture for AI on OpenShift



Timeline



2H '24

Next

- ▶ **MLOps**
 - Data Science Pipelines v2 incl. experiment tracking
 - Model registry v1
 - Data drift detection, bias detection, AI explainability
 - Model deployments to edge locations
- ▶ **Model development**
 - Feature store (Tech Preview)
 - VS Code workbench image (GA)
- ▶ **Platform/integrations**
 - RHOAI / RHEL AI integration v1
 - Support AMD GPUs
 - Support Intel Gaudi 2/3 devices
 - Vector DB partner solution - Elasticsearch
 - NVIDIA NIM integration
 - Enhance data connections

1H'25

Future

- ▶ **MLOps**
 - Model registry storage solution & tighter integration with DS Pipelines
 - Model serving enhancements - eg. inference graphs, canary rollouts
 - More OOTB monitoring visualizations, explainability UI
- ▶ **Model development**
 - Local IDE plugins (eg. VS Code)
 - Enhance model tuning experience
 - RS Studio workbench image (GA)
 - Enhance gen AI guardrails capabilities
 - Enhance distributed workloads integration with DS Pipelines
- ▶ **Platform/integrations**
 - Enhance RHOAI / RHEL AI integration
 - Expand admin UI configuration capabilities
 - Enhance public API & GitOps support
 - Expand accelerator support - specific devices TBD

TRY

Getting started with Red Hat OpenShift AI

Free

\$

\$\$

Sandbox environments

- ▶ Developer portal
red.ht/rhods_sandbox
- ▶ Support for OpenShift AI only
(Limited access to ISV software)
- ▶ Single-user environment
- ▶ Small cluster size only
- ▶ 30-day maximum
- ▶ No cost for infrastructure, OpenShift,
OpenShift AI

Trial environments

- ▶ 60 day trial (self-managed w/ OpenShift Container Platform included as option)
red.ht/rhods_60_trial
- ▶ Support for OpenShift AI with ISV software trials available from partners
- ▶ Multi-user environment
- ▶ Full range of cluster sizes supported
- ▶ If interested in managed cloud service option, check out Level Up ROSA program

Paid environments

- ▶ Support for OpenShift AI with ISV software available from partners
- ▶ Multi-user environment sized based on customer needs
- ▶ Annual and On-demand Hourly Costs per vCPU (managed cloud) and Yearly (managed and self managed)

- Dev Sandbox

<https://developers.redhat.com/products/red-hat-openshift-ai/overview>

- License Plate Detection

<https://redhat-scholars.github.io/rhods-lp-workshop/rhods-lp-workshop/index.html>



Red Hat OpenShift AI

Partner-Led
Workshop-As-A-Service
ecrosbie@redhat.com

RHOAI Workshop - 29th May
<https://connect.redhat.com/en/training/emea/course-catalog>



- ▶ **Implemented interactive lecture and lab environment** for computer scientists and engineers based on Red Hat OpenShift Data Science
- ▶ **Currently over 300 users** including over 100 concurrent
- ▶ **Integrates with the Boston University online textbook material,** also authored using the Red Hat OpenShift Data Science
- ▶ **Fast time to solution:** cloud services environment enabled BU to configure and deploy in December for classes that started in January
- ▶ **Lowers cost:** auto-scales based on demand; enables bursty interactive use cases at optimized cost



Red Hat OpenShift AI



[Learn more ▶ redhat.com](#)



[Contact us ▶](#)

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](https://www.linkedin.com/company/red-hat)

 [youtube.com/user/RedHatVideos](https://www.youtube.com/user/RedHatVideos)

 [facebook.com/redhatinc](https://www.facebook.com/redhatinc)

 twitter.com/RedHat