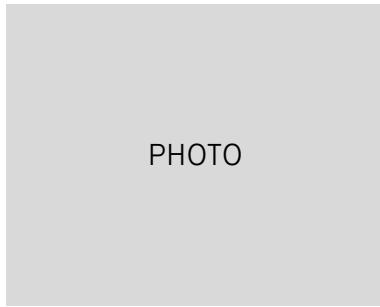


Data Engineering and Machine Learning on OpenShift

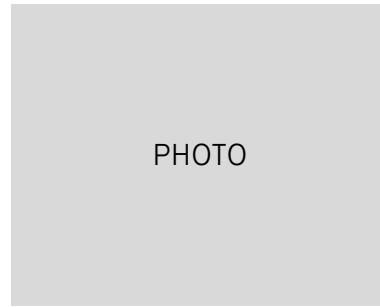
Agenda

- Introductions
- How Red Hat Sees AI
- Why AI and Machine Learning on OpenShift?
- AI as a Service
- Data Engineering on OpenShift
- Running GPU Workloads
- AI/ML Software Partners
- Lab Logistics and Setup
 - Presenter (System Admin persona) - Installing Open Data Hub (ODH) via ODH operator
 - Attendees (Data Engineers and Project Owners)
 - Instantiating ODH in their OpenShift projects
 - JupyterHub, Kafka, Spark, Ceph, Prometheus, Grafana, Seldon Inferencing
- Lab - Analytics and Data Science (Data scientists persona)
 - Spark: Using Spark to access and manipulate data from two data sources
 - TensorFlow: Using TF to create models and serving models using Seldon
 - Creating Jobs on OpenShift

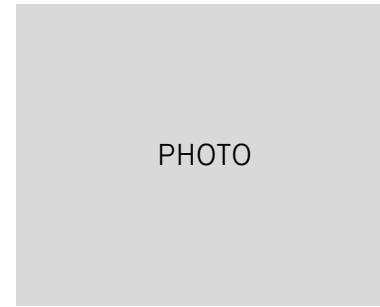
Introductions



PHOTO



PHOTO



PHOTO

Name:
Role/team:
Where you're from:

Name:
Role/team:
Where you're from:

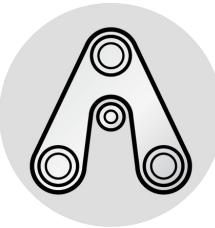
Name:
Role/team:
Where you're from:

How Red Hat Sees AI



AI as a Workload

Represents a workload requirement for our **platforms** across the hybrid cloud.



AI Accelerates Open Source Development

Applicable to Red Hat's existing core business in order to increase **Open Source** development and production **efficiency**.



AI Enhances RH Products & Services

Valuable to our customers as specific services and product capabilities, providing an **Intelligent Platform** experience.



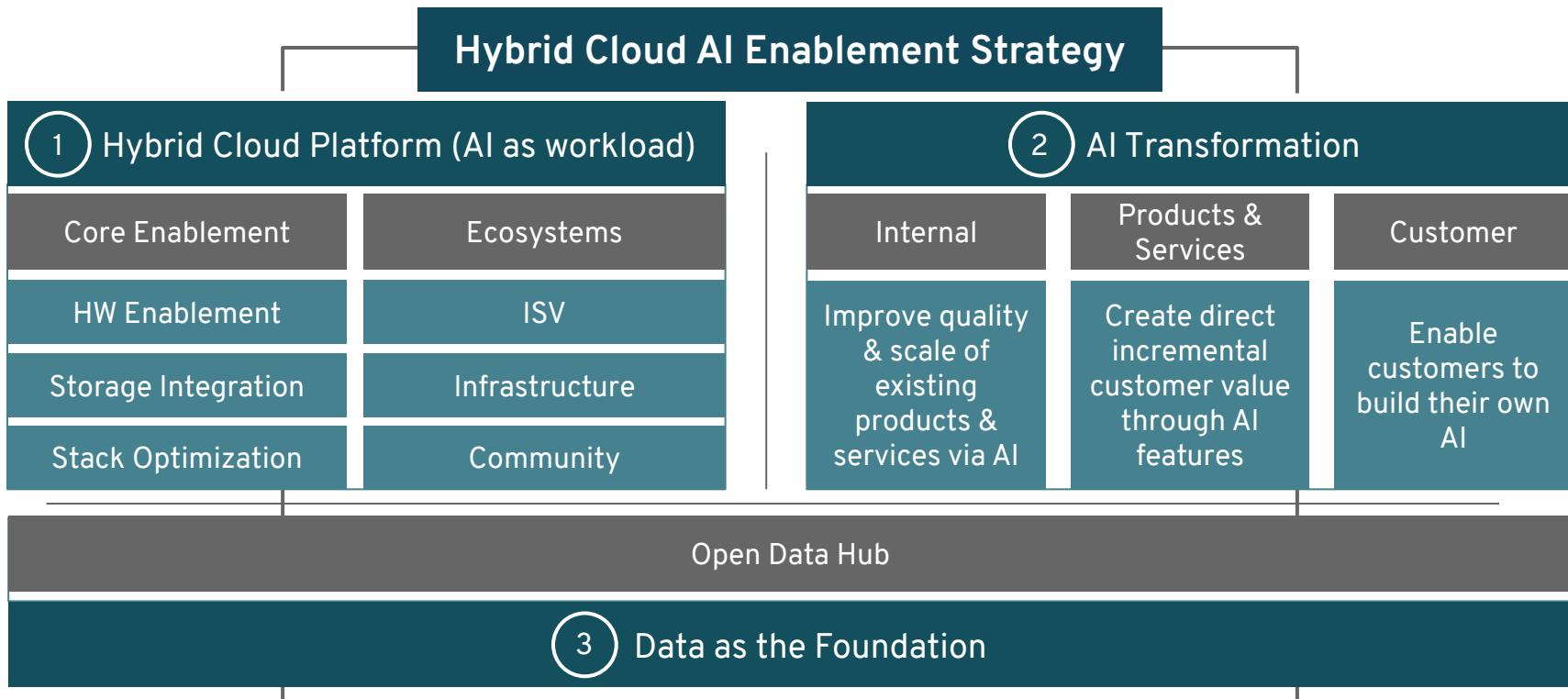
AI Enables Customers

to build **Intelligent Apps** using Red Hat products as well as our broader partner ecosystem.

010110
101010

Data as the Foundation

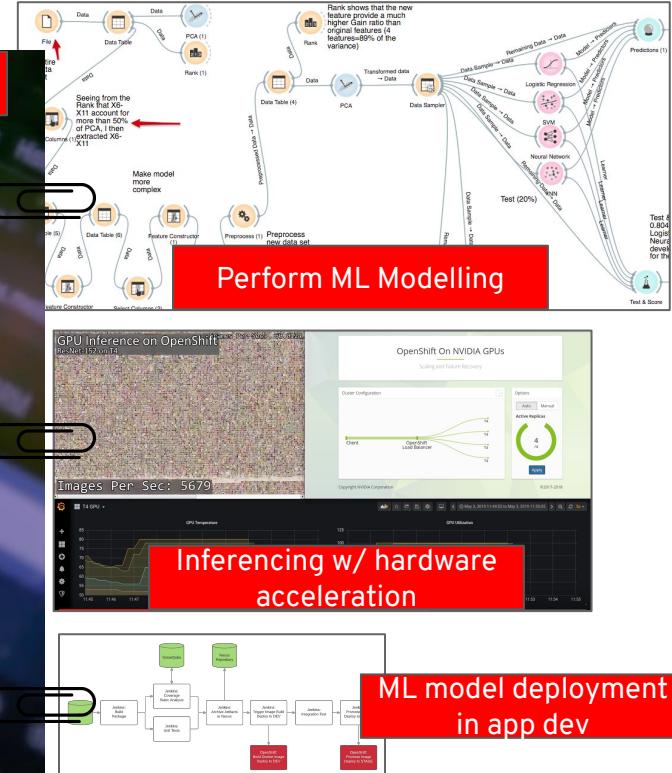
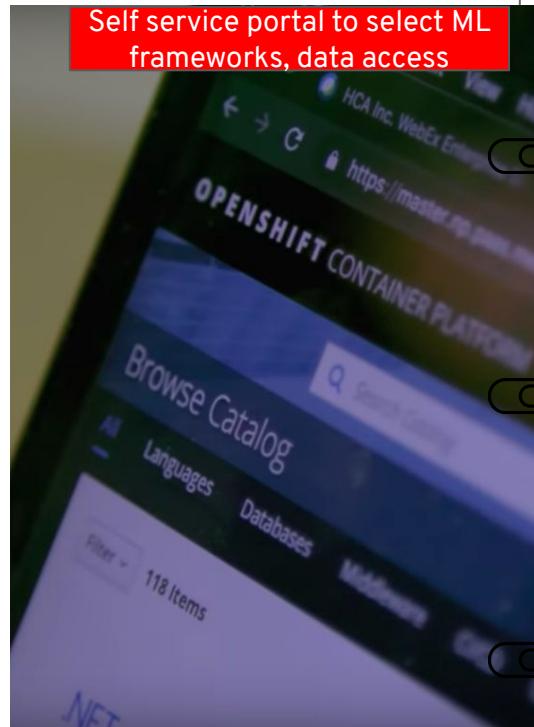
AI at Red Hat



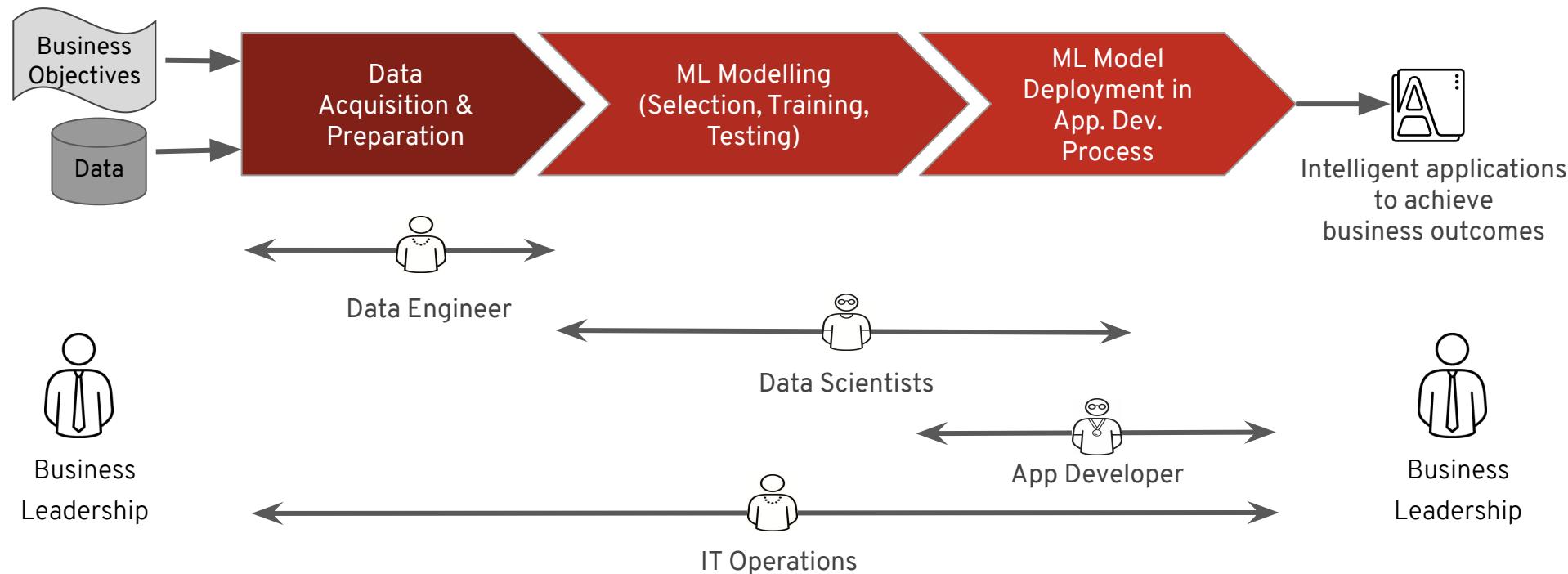
Why AI and Machine Learning on OpenShift?

What Does a Data Scientist Want?

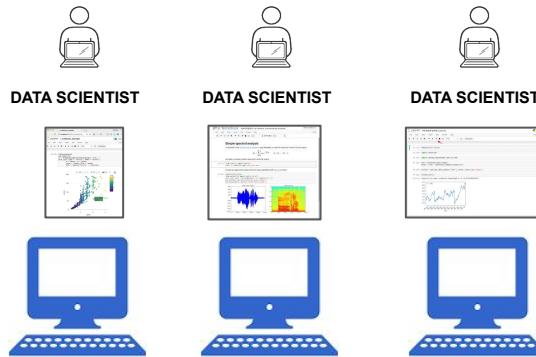
As a Data Scientist, I want a “self-service cloud like” experience for my Machine Learning projects, where I can access a rich set of modelling frameworks, data, and computational resources, share and collaborate with colleagues, and deliver my work into production with speed, agility and repeatability to drive business value!



Machine Learning Pipeline



Current Situation and Challenges

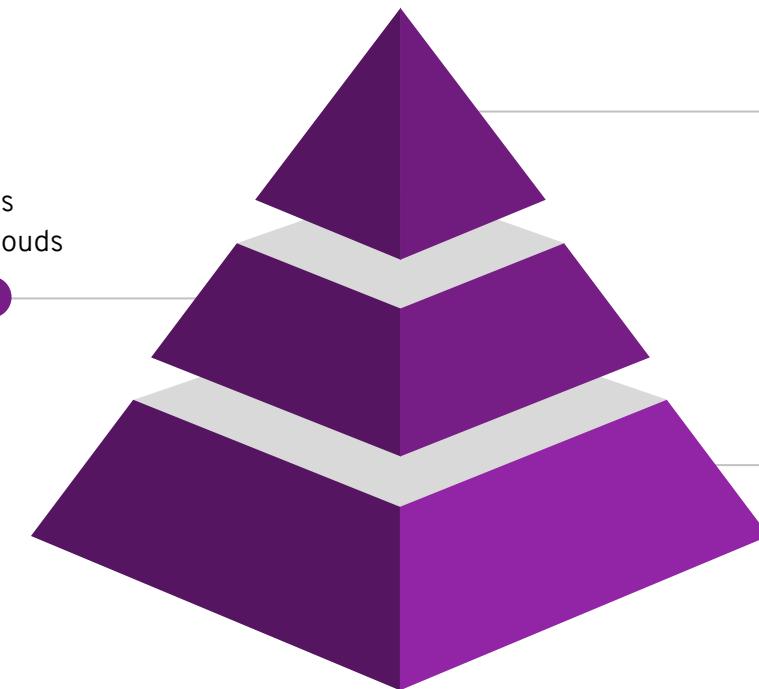


- Team(s) of Data Scientists and Developers
- Sharing and collaboration, if any, is difficult, manual, error prone and take time
- Access to limited non-shared resources means modeling takes long times or can't achieve desired accuracy
- Delivering into models into production is a challenge

Why Containers & Kubernetes in Hybrid Cloud for AI/ML workloads?

Portability & flexibility for ML powered apps

- Develop/deploy ML apps across data center, edge, and public clouds
- Offer ML-as-a-service



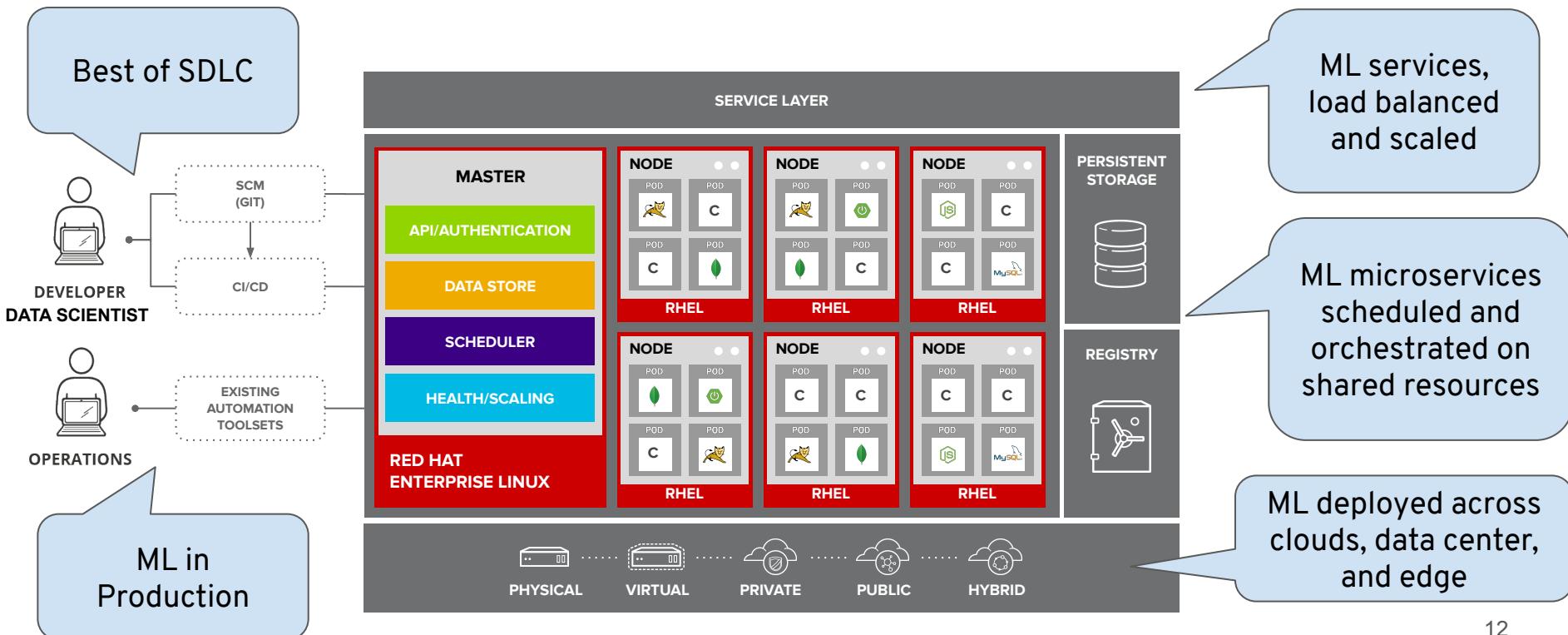
Red Hat products & services help solve additional challenges

- Automation, CI/CD drive collaboration
- Boost productivity
- Data access, prep, & governance
- Apps lifecycle management & operations

Agility across the ML pipeline

- Automated install and provisioning
- Autoscaling
- GPU acceleration, scaling, security, uptime

Why OpenShift And Cloud Platforms for ML Workloads?



AI/ML on OpenShift Momentum is Strong



Connected Drive &
Autonomous Driving



Data driven diagnosis
of “Sepsis”



Digital banking with
personalized services



Autonomous Driving

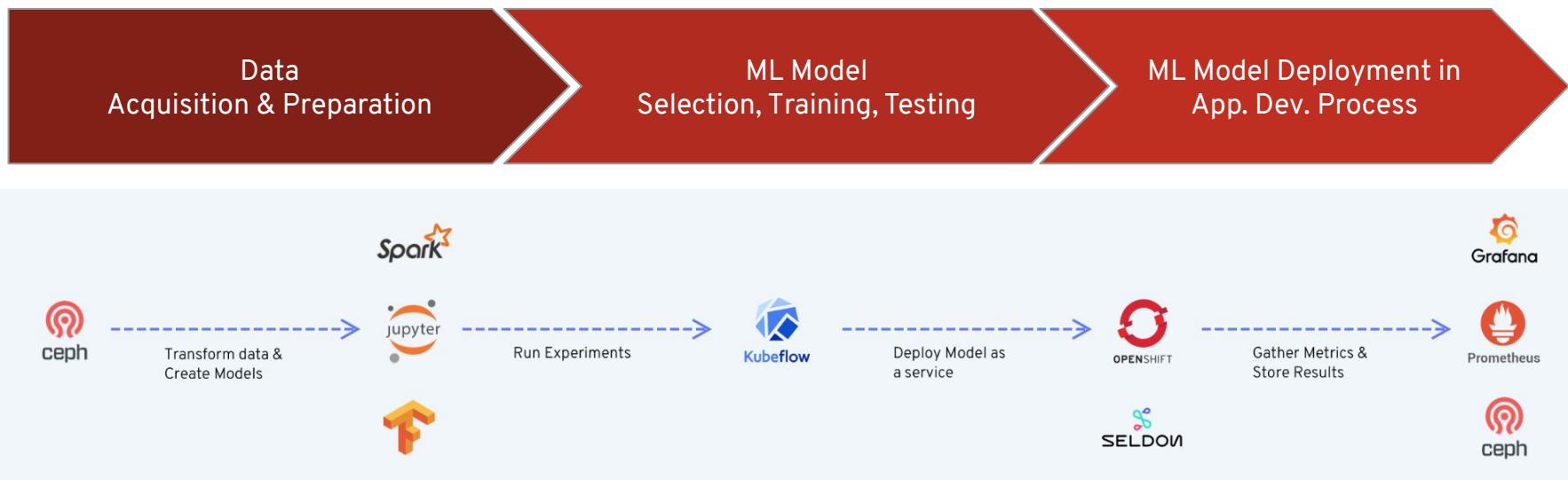


Oil & Gas Exploration

AI as a Service

The Open Data Hub Project

- OpenDataHub.io
- Meta-operator that integrates best open source AI/ML/Data projects
- Blueprint architecture for AI/ML on OpenShift
- Red Hat's internal Data Science and AI platform
- Open Data Hub Architecture: <https://opendatahub.io/docs/architecture.html>





Open Data Hub v0.4 Operator

Available Now at OpenDataHub.io



Prometheus

- Monitoring and alerting toolkit
- Used to diagnose problems



Grafana

- Analytics platform for all metrics
- Query, visualize and alert on metrics



SELDON

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



- Unified analytics engine
- Large-scale data access



- Multi-user Jupyter
- Used for data science and research



ceph

- Distributed Object Store
- S3 Interface



STRIMZI
kafka

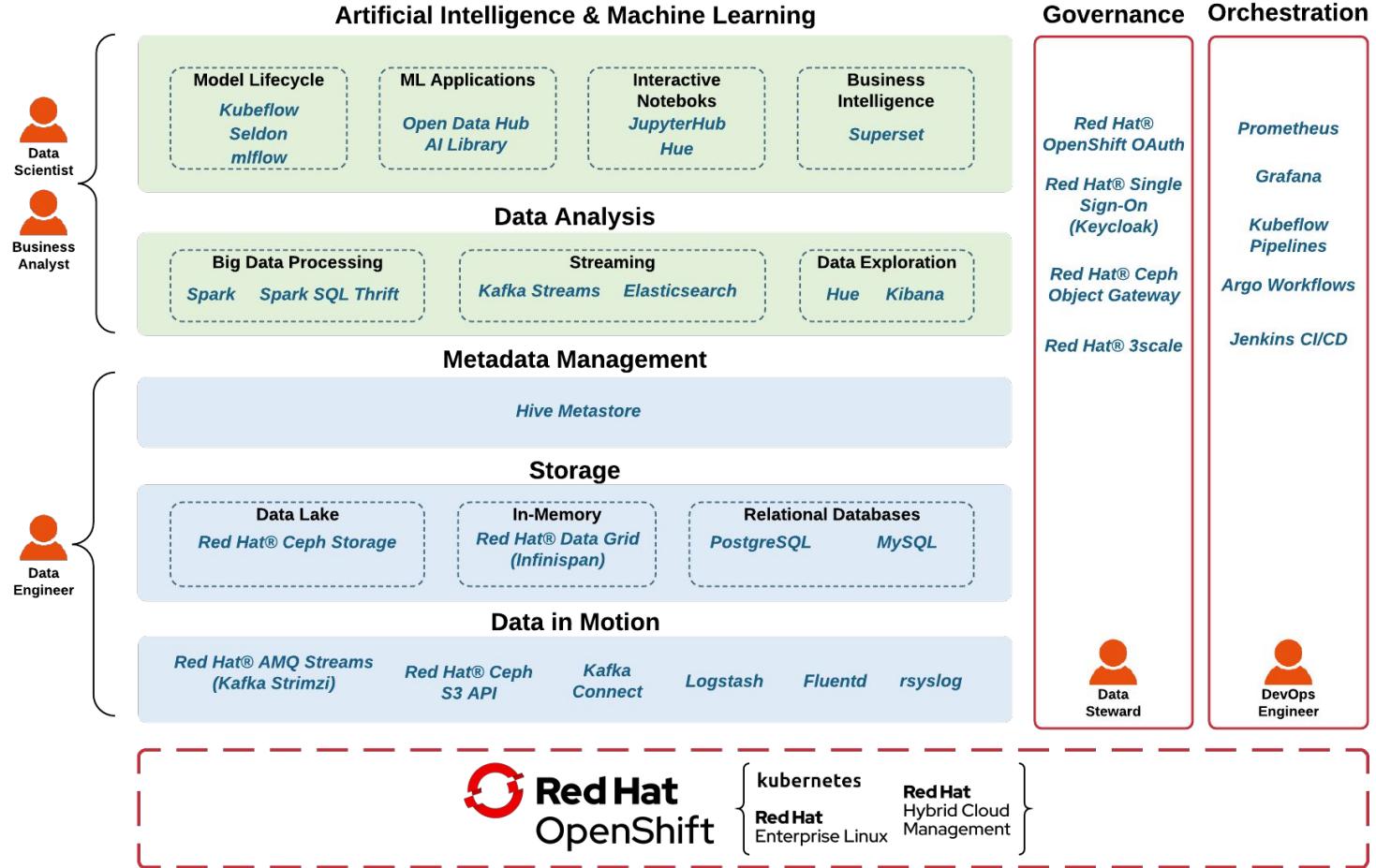
- Distributed event streaming
- Pub/Sub Messaging



argo

- Container-native workflow engine
- Declaratively deploy ML pipelines and models

Open Data Hub Blueprint



Upstream/Community Projects in the AI/ML Space



NVIDIA NGC
GPU optimized
and curated



Kubeflow

ML toolkit and lifecycle
Kube



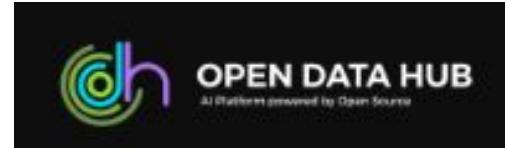
Tensorflow



P Y T O R C H



Others

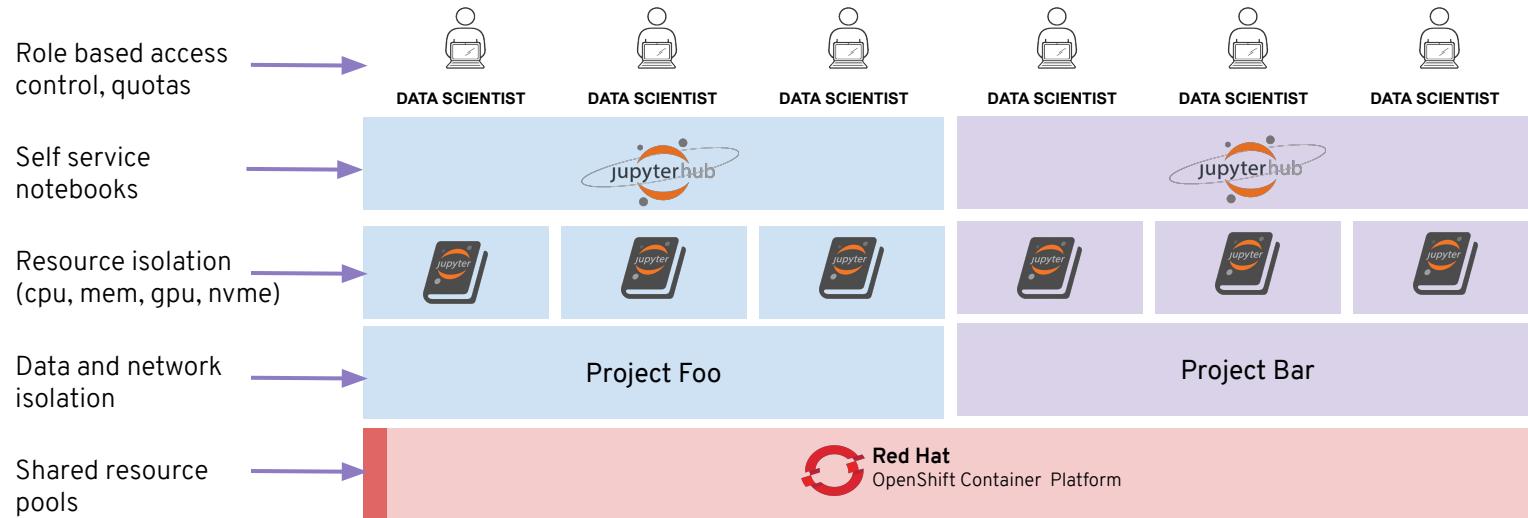


ML-as-a-service platform based on
OpenShift, Ceph, Kafka, JupyterHub,
Spark,

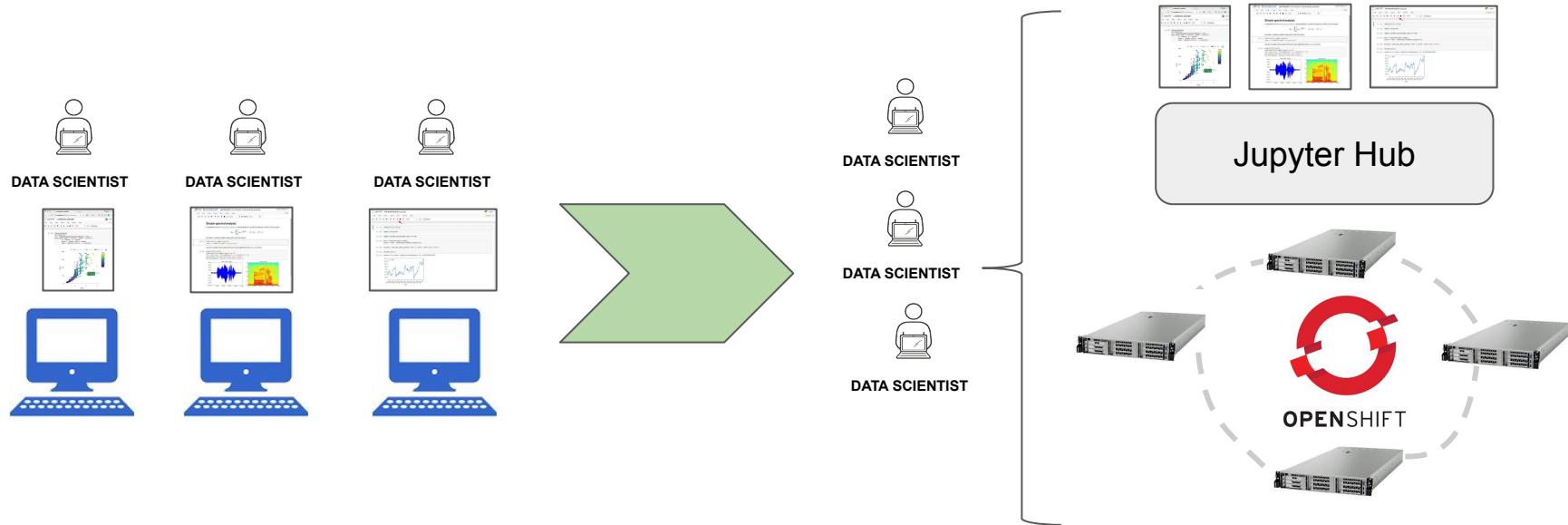


Home for k8s community to share
operators for various apps/tools

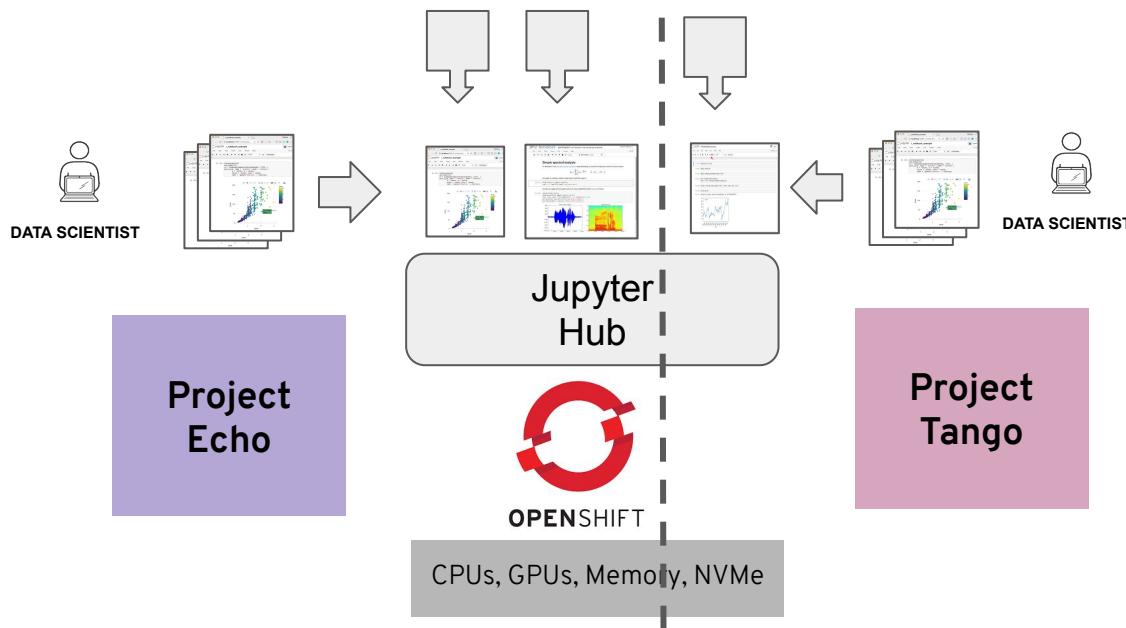
Jupyter Notebooks as a Service



Pooling and Sharing Resources with JupyterHub on OpenShift



Jupyter as a Service Multi-tenancy

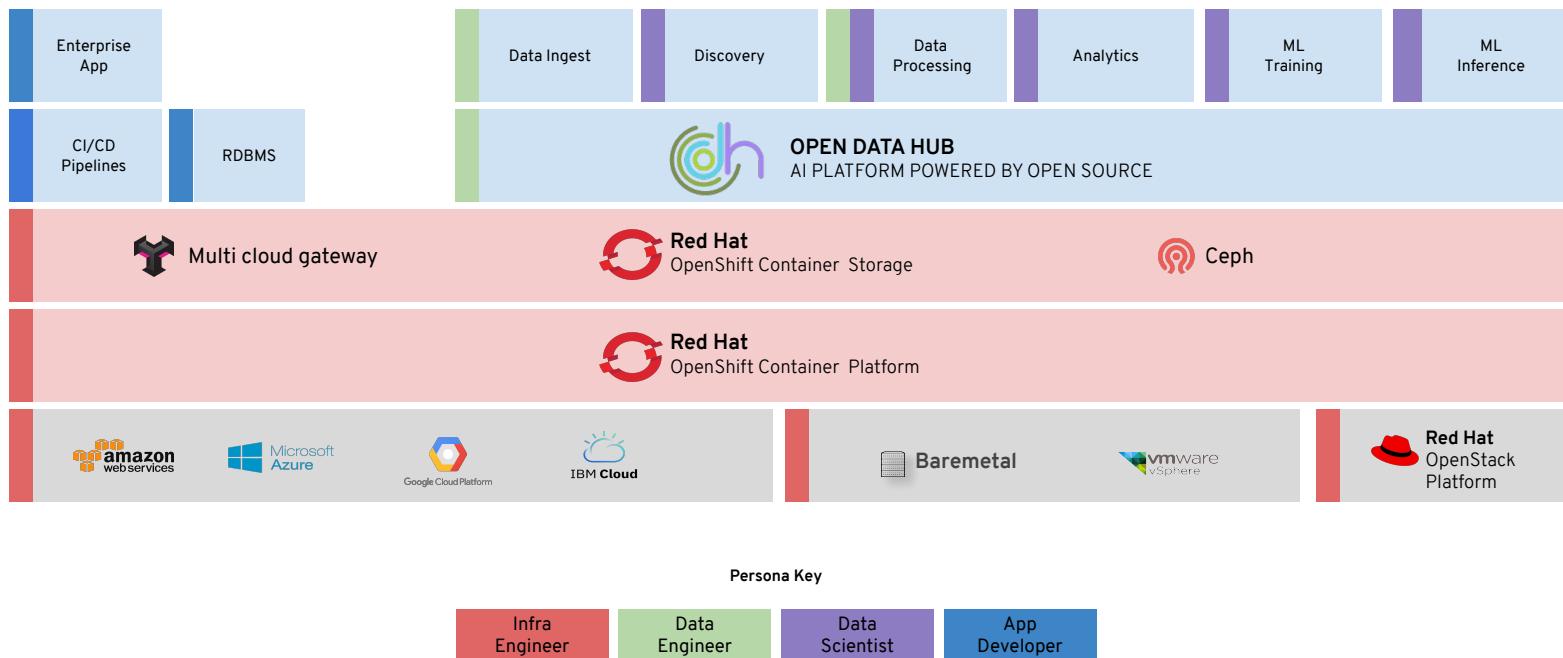


- Identity and Role Based Access
- Project Isolation (Data, Networks)
- Sharing of Resources with Protection
- Resource Isolation
- Resource Quotas
- Priority and Preemption

All this is possible with OpenShift 3.11 and OpenShift 4.1 (and later)

Data Engineering on OpenShift

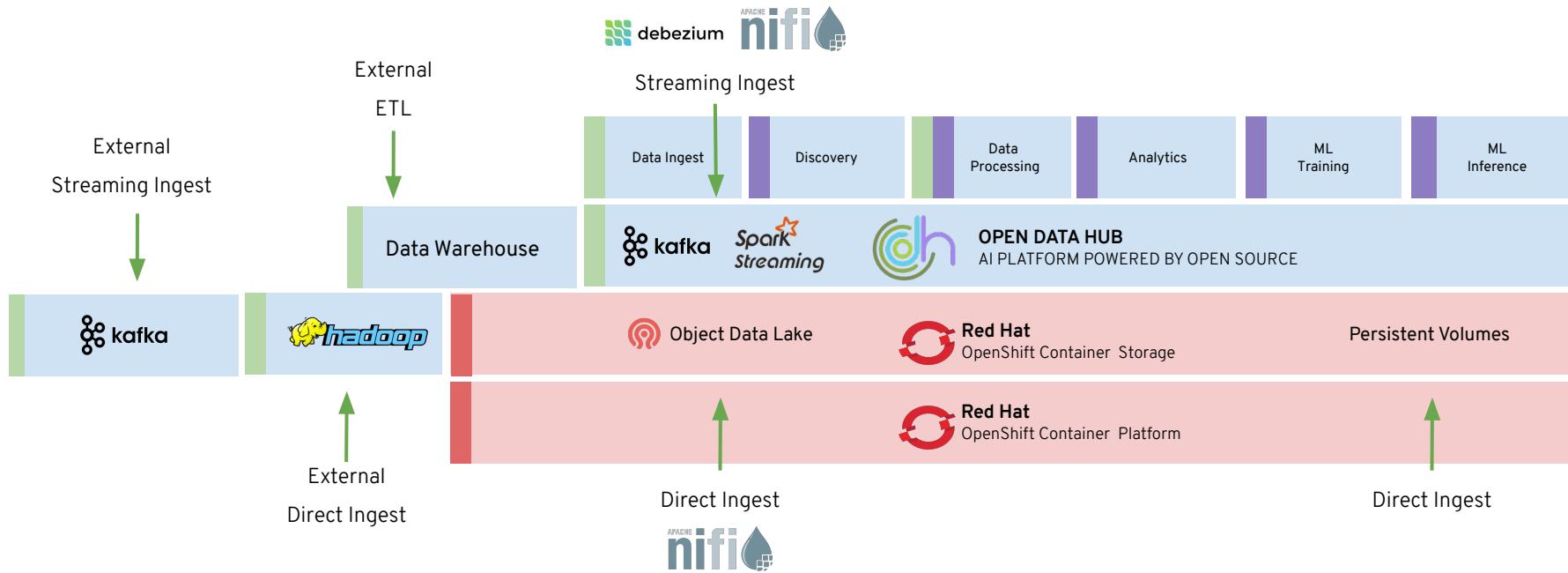
Data Pipelines



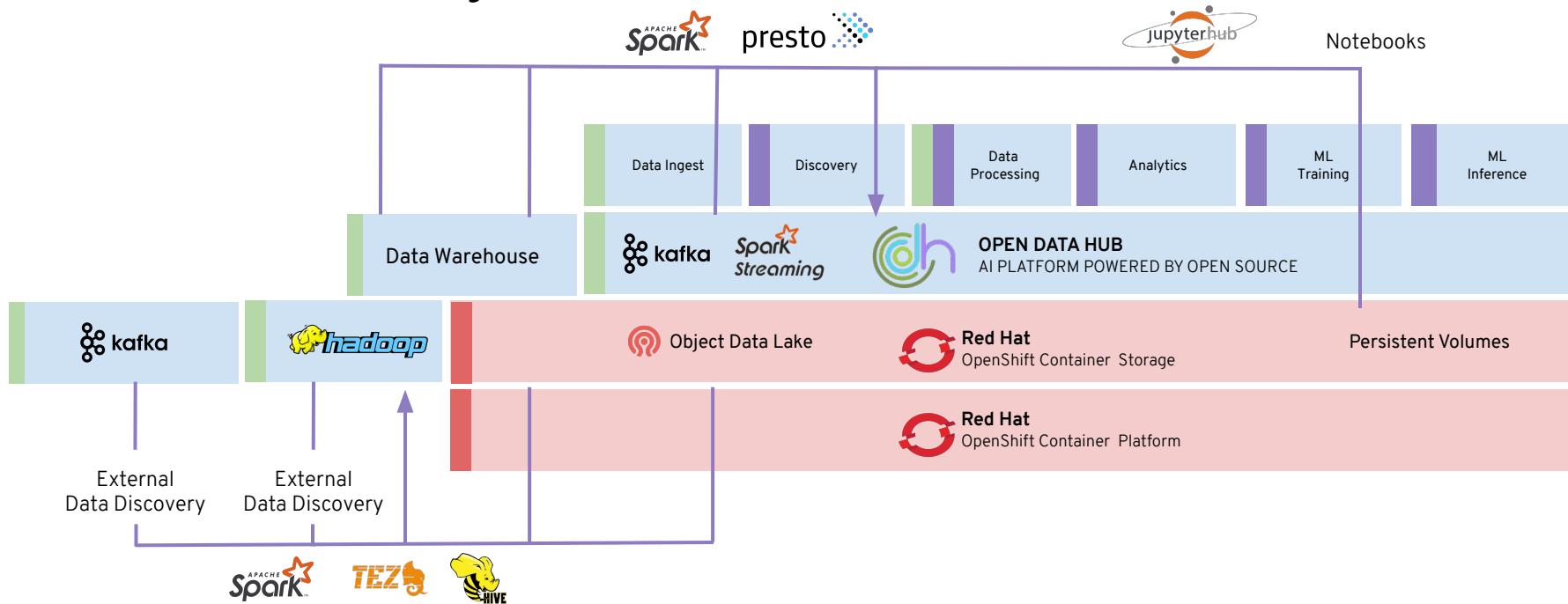
Data Pipelines

- Multiple **data sources** (apps, sensors, devices, satellites, infrastructure)
- Multiple **data pipelines** exist in every organization
- **Data pipelines** should continually evolve to improve
 - Discoverability
 - Cleanliness
 - Scalability
- **Machine Learning** should be used to derive value from data pipelines in ways that would not otherwise be possible with analytics alone

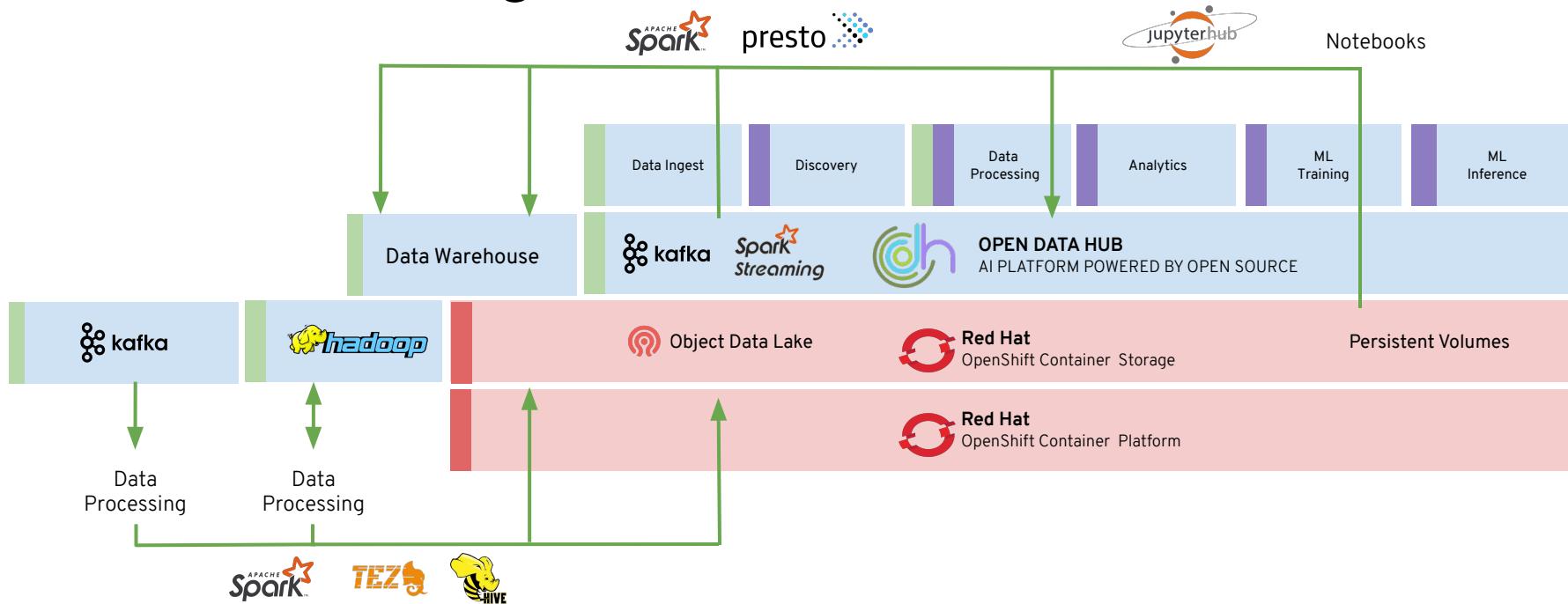
Data Ingest



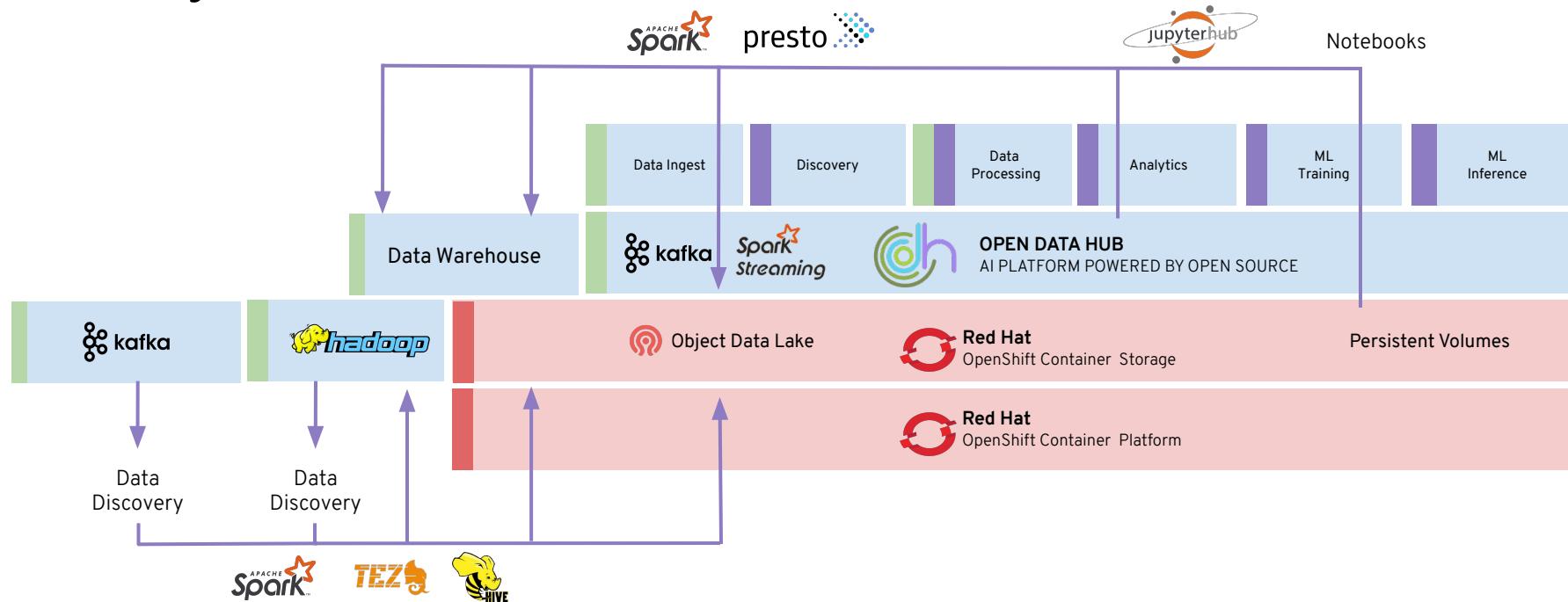
Data Discovery



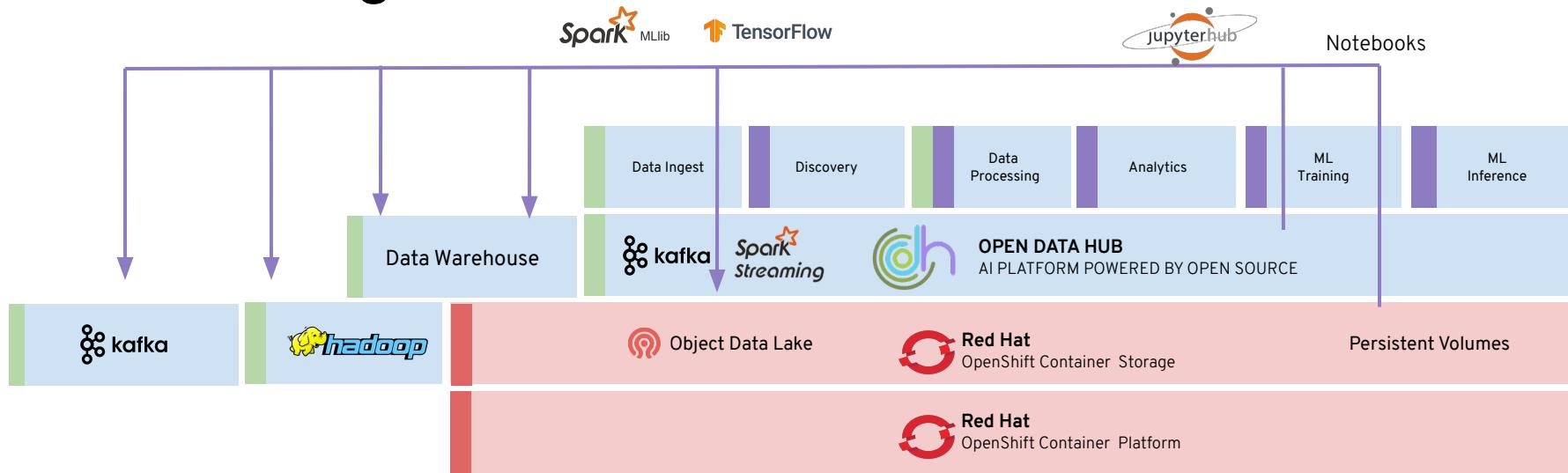
Data Processing



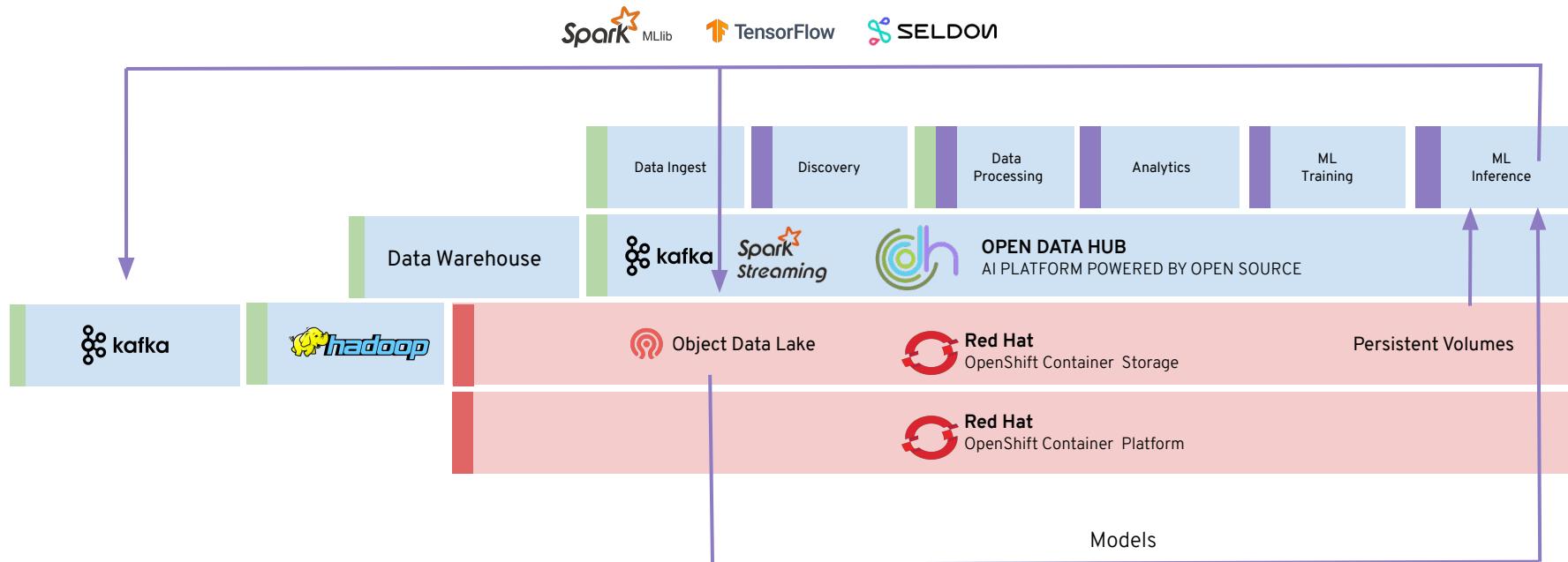
Analytics



ML Training



ML Inference



Mapping Data Pipeline Activities



Data Access with OpenShift Container Storage

Persistent Volumes

- Fast and dynamic provisioning
- Regional durability
- RWO, RWX
- Provided by OCS 3 and OCS 4

Object Storage

- Common API across hybrid cloud (S3)
- Provided by OCS 4
- Scale Massively
 - Billions of objects
 - Petabytes of data

Mapping OpenShift Container Storage Services



Notebooks



Broker Cluster



Streaming Cluster



Batch Cluster



Model Serving

RWX PVs

RWO PVs

Checkpoint to S3

Sink to S3

RWX PVs



Metastore



Consumer Cluster



Ad-hoc Cluster



Interactive Query



Training, Serving

RWO PVs

Sink to S3

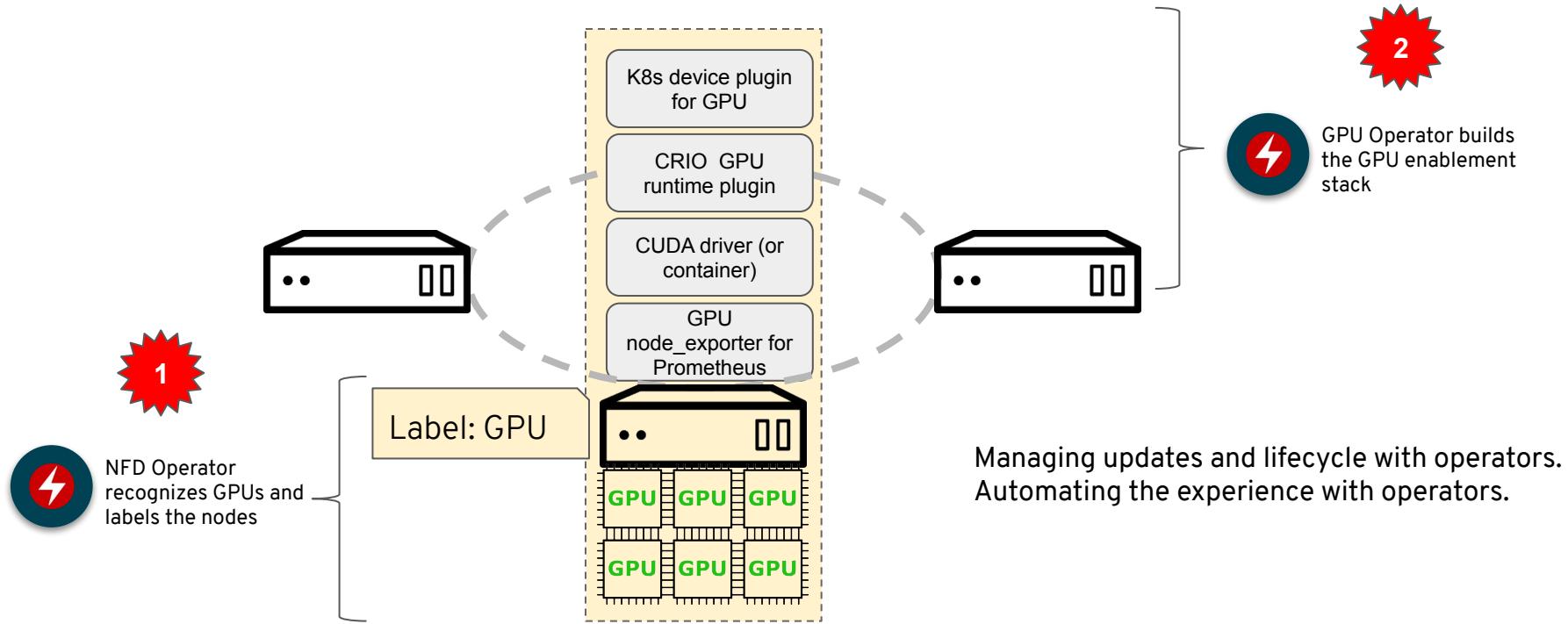
Source / Sink to S3

Source / Sink to S3

Source from S3

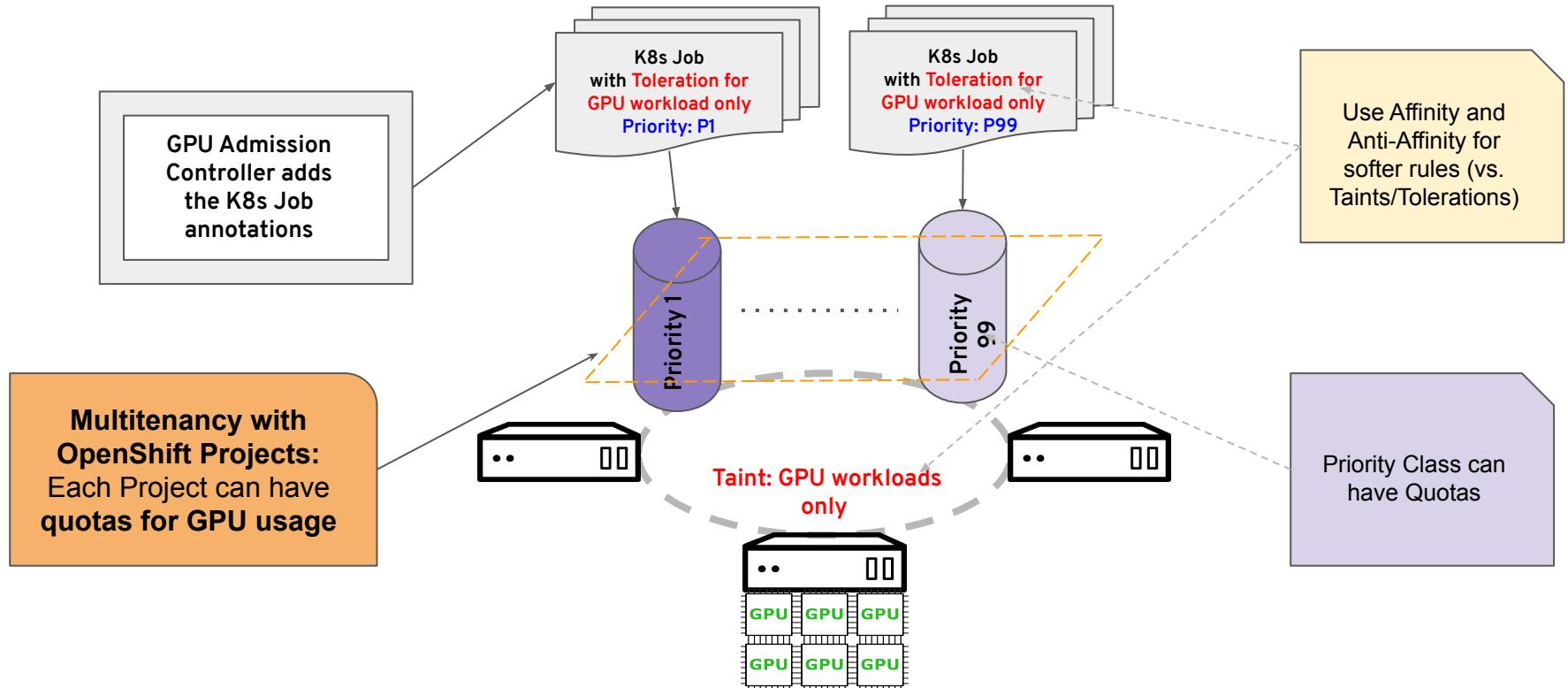
Running GPU Workloads

Enablement of GPUs as a service on OpenShift



GPU supported in OpenShift 3.11 and OpenShift 4.1 with RHEL7 only on GPU nodes; NFD and GPU operator are in roadmap

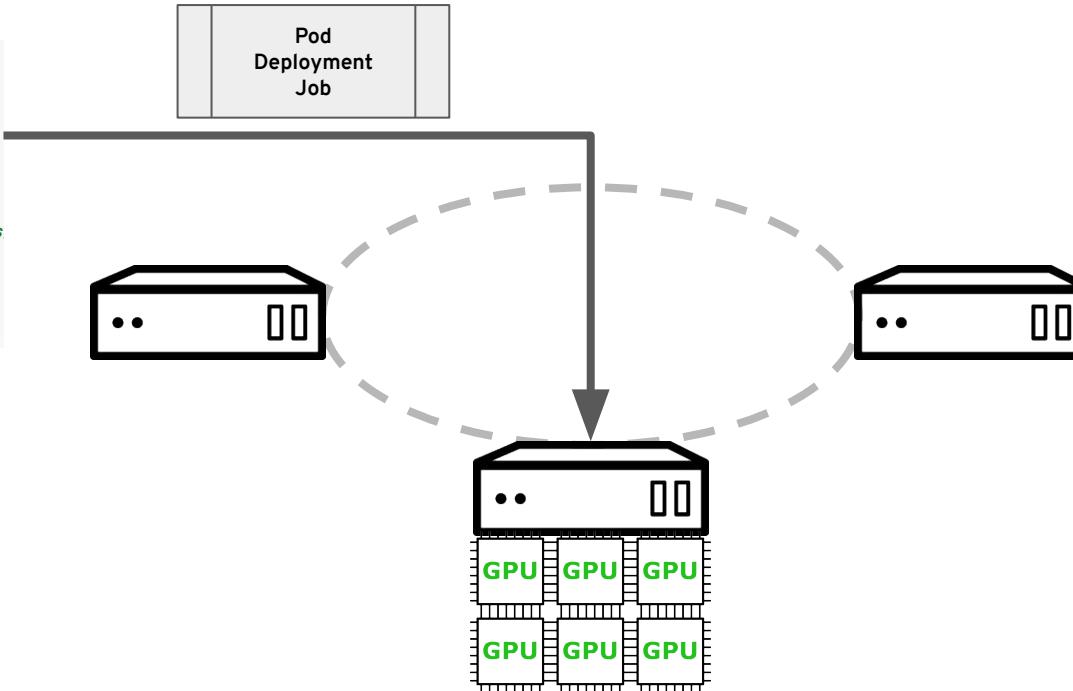
Sharing GPUs in OpenShift for Model Training



Supported in OpenShift 3.11 (except Priority Class Quotas) and OpenShift 4.1 (RHEL7 only on GPU nodes)

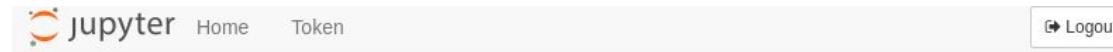
Deploying GPU Workloads onto OpenShift

```
apiVersion: v1
kind: Pod
metadata:
  name: cuda-vector-add
spec:
  restartPolicy: OnFailure
  containers:
    - name: cuda-vector-add
      # https://github.com/kubernetes/kubernetes
      image: "k8s.gcr.io/cuda-vector-add:v0.1"
      resources:
        limits:
          nvidia.com/gpu: 1 # requesting 1 GPU
```



Supported in OpenShift 3.11 (except Priority Class Quotas) and OpenShift 4.1 (RHEL7 only on GPU nodes)

Using GPUs in Jupyter



Spawner Options

JupyterHub Server Image

Select desired notebook image

s2i-spark24-minimal-notebook:3.6

Deployment sizes

Select desired container size

Default (resources will be set based on profiles configured by admin)

GPU:

0

Select number of GPUs to use via UI

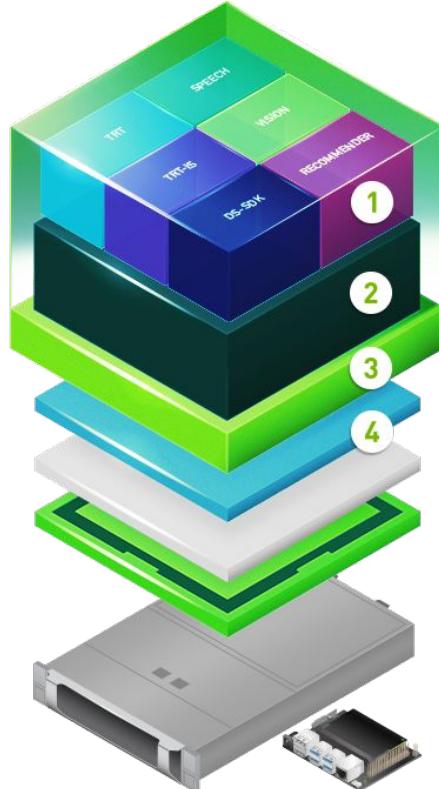
Environment Variables

AWS_ACCESS_KEY_ID:

AWS_SECRET_ACCESS_KEY:

Supported in OpenShift 3.11 (except Priority Class Quotas) and OpenShift 4.1 (RHEL7 only on GPU nodes)

NVIDIA + Red Hat Joint Collaboration



1. NGC Containers
2. Kubernetes
3. Automation and autoscaling
4. CUDA Drivers for RHEL and RHEL CoreOS (RHCOS)

Also, joint development of support for HPC:

- GPUDirect RDMA - direct I/O between GPU and 3rd party peer devices
- SR/IOV and InfiniBand - sharing of single PCIe devices by multiple VM

Accelerated AI EAP - Partnership with NVIDIA

Objective: Drive mainstream adoption of AI across teams



General Enablement Work - RHEL, OpenShift GPU Enablement

- DGX line, OEMs, Cloud Providers

OpenShift 4 Pilot - Accelerated AI EAP

- OpenShift 4 on Bare-metal HW with NVIDIA GPUs
- NVIDIA NGC Containers (AI/ML libraries CUDA optimized)
- Request form: <https://www.openshift.com/accelerated-ai>

Summit Keynote Demo: <https://youtu.be/FUu4kMc0PL8?t=1448>

Offering Blog: <https://blogs.nvidia.com/blog/2019/05/09/red-hat-openshift-gpu-kubernetes/>

Technical Overview: <https://devblogs.nvidia.com/gpu-support-ai-workloads-openshift4/>

Documentation:

<https://docs.nvidia.com/datacenter/kubernetes/openshift-on-gpu-install-guide/index.html>

AI/ML Software Partners

Software Vendor Partner Program

Mission: With partners, make best-in-class AI/ML hybrid/multi-cloud ready for mainstream and enterprise markets

- Primary Focus on OpenShift - kubernetes-powered multi/hybrid-cloud AI/ML platform
- Hardware Ecosystem for Performance - GPUs
- SW Ecosystem Enablement
 - Projects w/ Vendor Support - Certified vendors w/ commercial support
 - Balance Platforms & Components - Provide choice
- Cloud-Like Experience - Operators, operatorhub.io, embedded operatorhub
- Certifications with Red Hat Partner Connect (for Container and Operator certifications)

Via:

- AI Reference Architecture - Open Data Hub - RH Ref Arch for AI for upstream/products/partners
- AI Use Case Definitions - define core use cases - RH + Partners Enable these use cases



Red Hat

PARTNER CONNECT

Red Hat software certifications

RED HAT ENTERPRISE LINUX

Build on a trusted foundation

Red Hat Enterprise Linux 8 (RHEL 8) is the intelligent operating system providing you and your customers a scalable, consistent and supported foundation for the enterprise hybrid cloud. Discover how RHEL 8 benefits you and your customers and learn more about our new RHEL Application Certification.

CONTAINERS

Modernize your applications

Embrace containers as an efficient way to build, distribute and deploy applications at scale. Explore our Red Hat Container Certification with the new Red Hat Universal Base Image, providing your Red Hat OpenShift customers application deployment flexibility in a cloud platform of their choice.

OPENSHIFT OPERATORS

Manage at scale

Enable automation at scale with Operators and make the leap to cloud native infrastructure. Understand how our Red Hat OpenShift Operator Certification combines the advantages of Kubernetes Operators with the strength of Red Hat OpenShift and Ansible Automation.

Who Are We Working With Today?



Just to name a few...

THE LAB - LOGISTICS

1. WIFI for labs:

SSID: <enter>

Password: <enter>

2. This presentation: <enter>

Full URL: <enter>

3. Get a attendee number for the lab:

<enter instructions here>

4. Address for OpenShift web console: <enter>

Full URL: <enter>

Login to OpenShift web console

Username: user<number>

Password: <enter>

Instructions:

<https://tinyurl.com/y3m2wrtz>

Or full URL

<https://gitlab.com/opendatahub/data-engineering-and-machine-learning-workshop/blob/master/doc/for-attendees.md>

Thank you



OpenDataHub.io