

AI/ML with OpenShift

Turning your AI/ML lifecycle into standard software development practices with containers and Kubernetes

Tushar Katarki
AI/ML on OpenShift
Product Manager
Red Hat

Robert Sedor
Chief Architect
Red Hat

Agenda

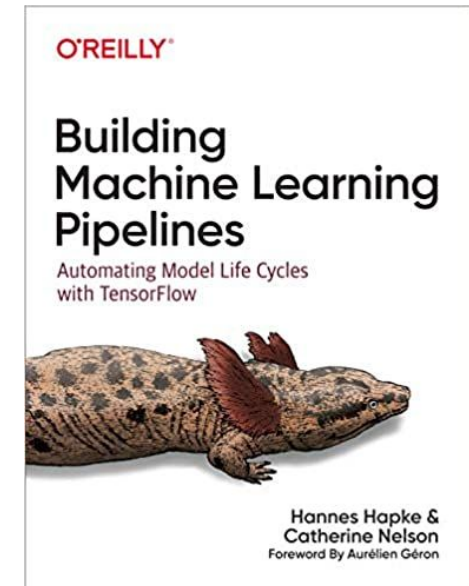
- Level Setting on AI/ML
- Data Science: Models and Jupyter
- OpenShift for Data Science
- Customer Success Stories
- How to move forward

Slides and recording will be posted below - may take a couple of days

<https://red.ht/LiveAI-ML>

Raffle - Book on ML

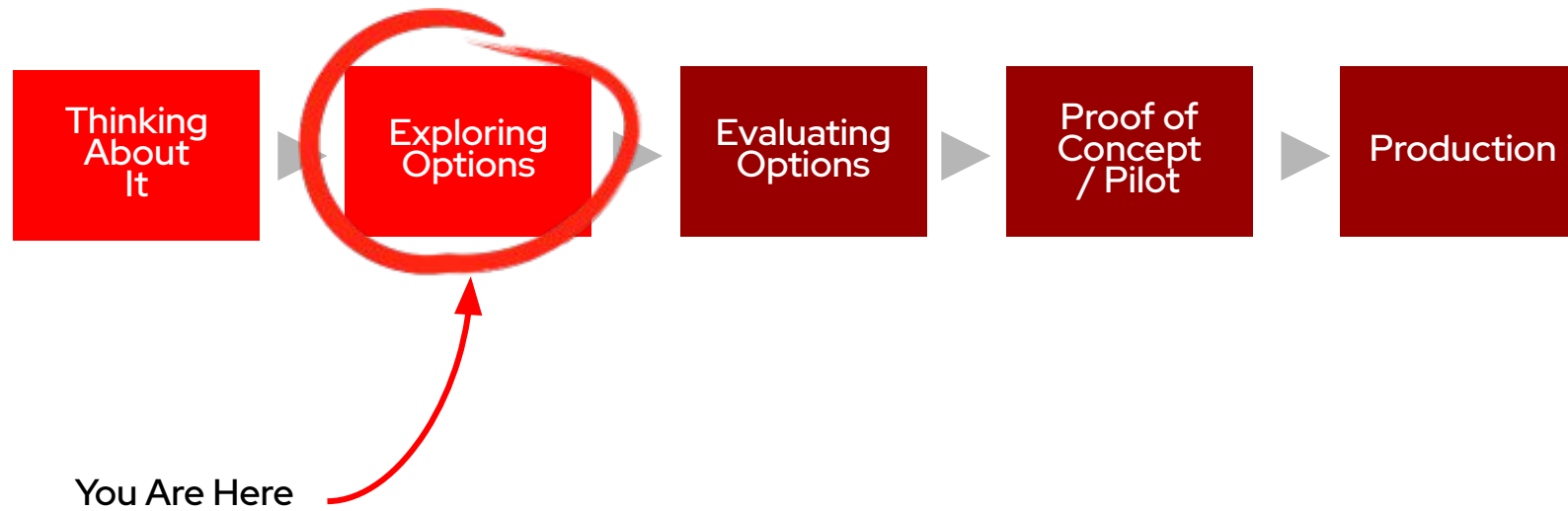
(will be announced at the end of the Webinar!)



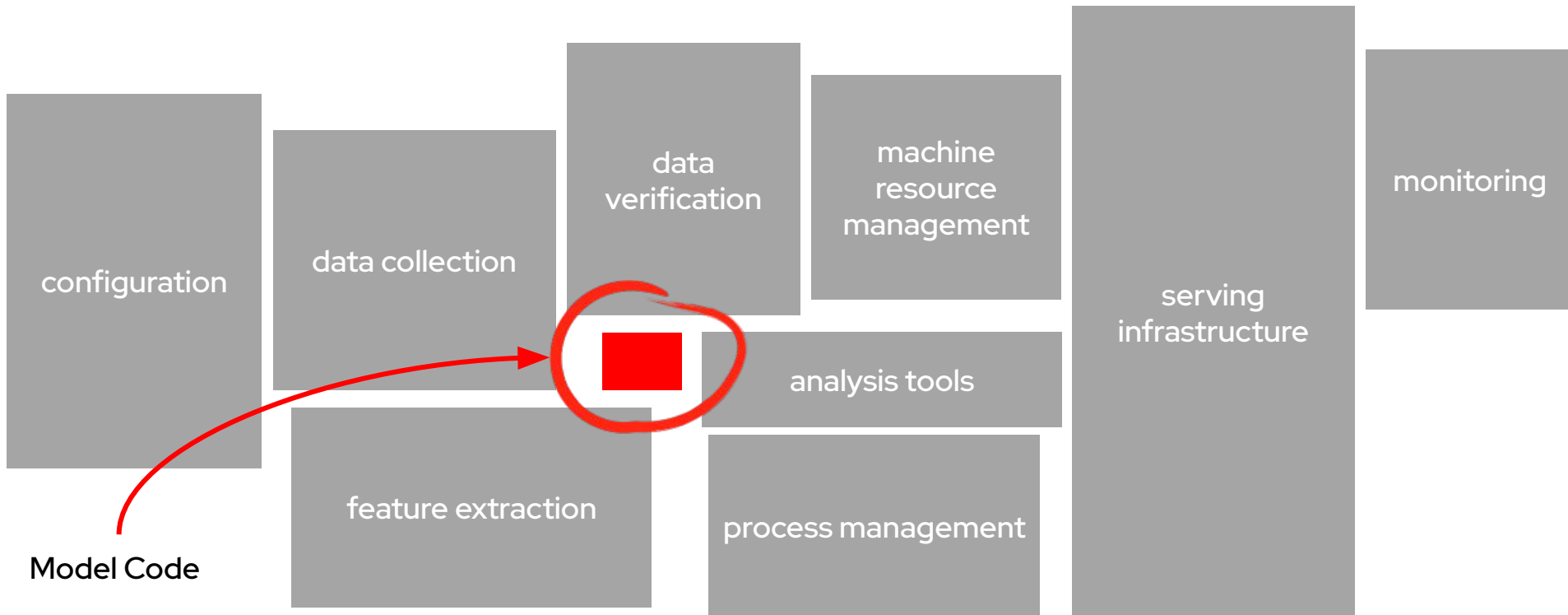
<https://www.amazon.com/Building-Machine-Learning-Pipelines-Automating/dp/1492053198>

Level setting

The intelligent application journey

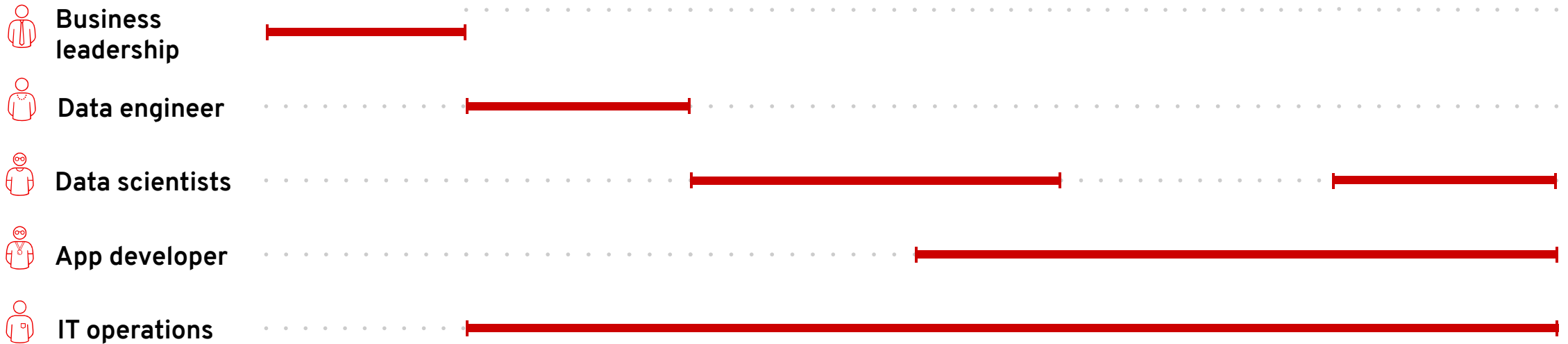
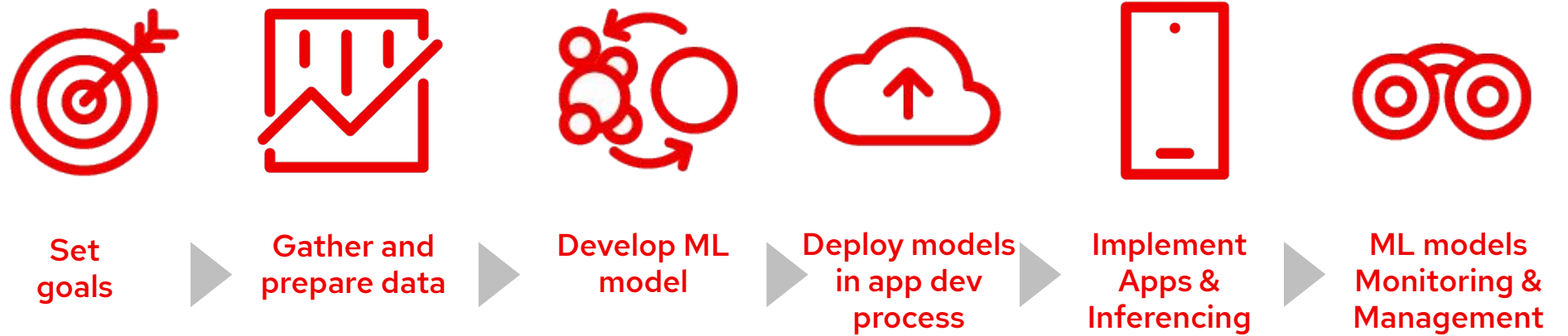


Intelligent applications are distributed systems



(Adapted from Sculley et al., "Hidden Technical Debt in Machine Learning Systems." NIPS 2015)

The ML lifecycle and responsibilities

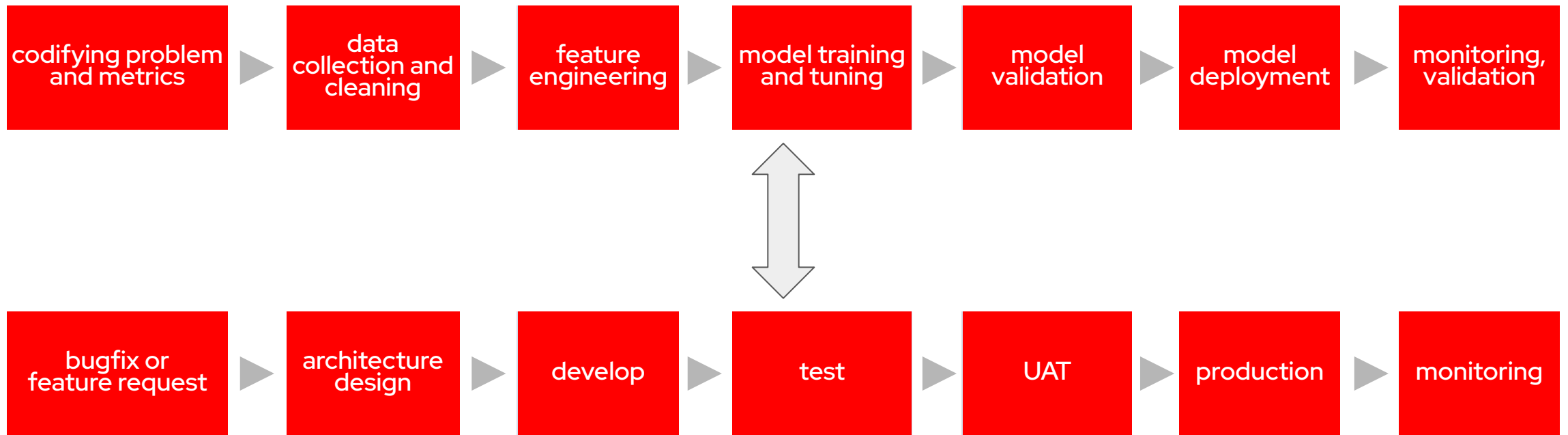


The ML lifecycle in technical detail



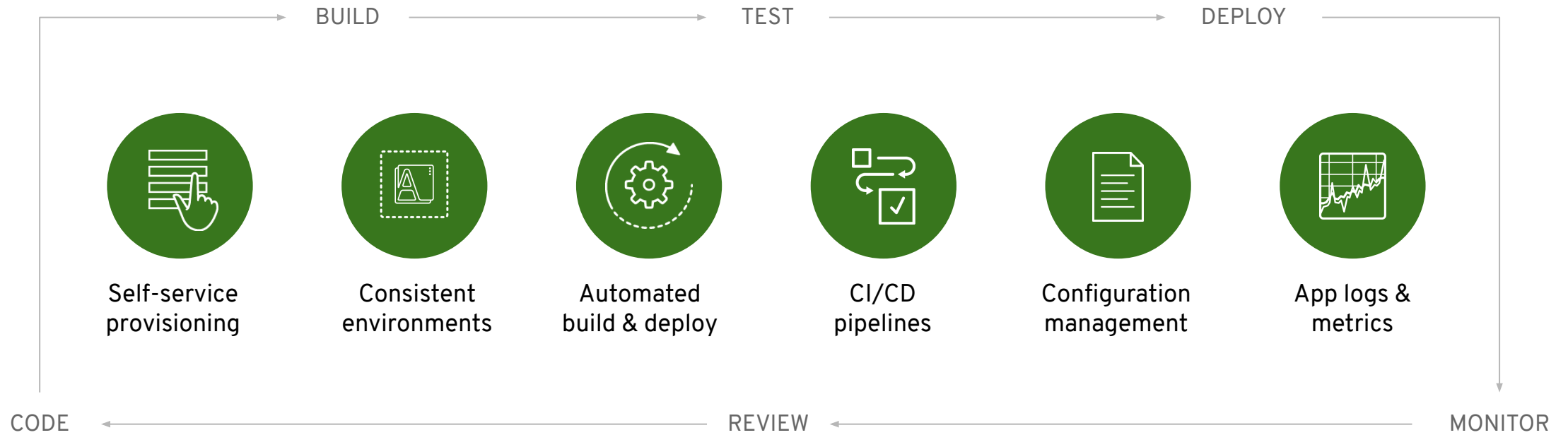
You have seen this before

It's just software development

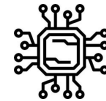
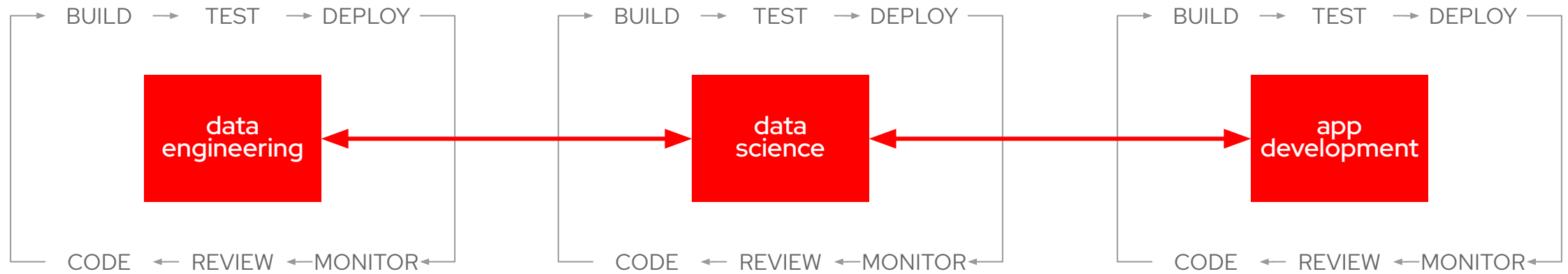


What is DevOps?

Where can I buy it?



One platform for all the domains



Physical

Virtual

Private cloud

Multi-Arch

Public cloud

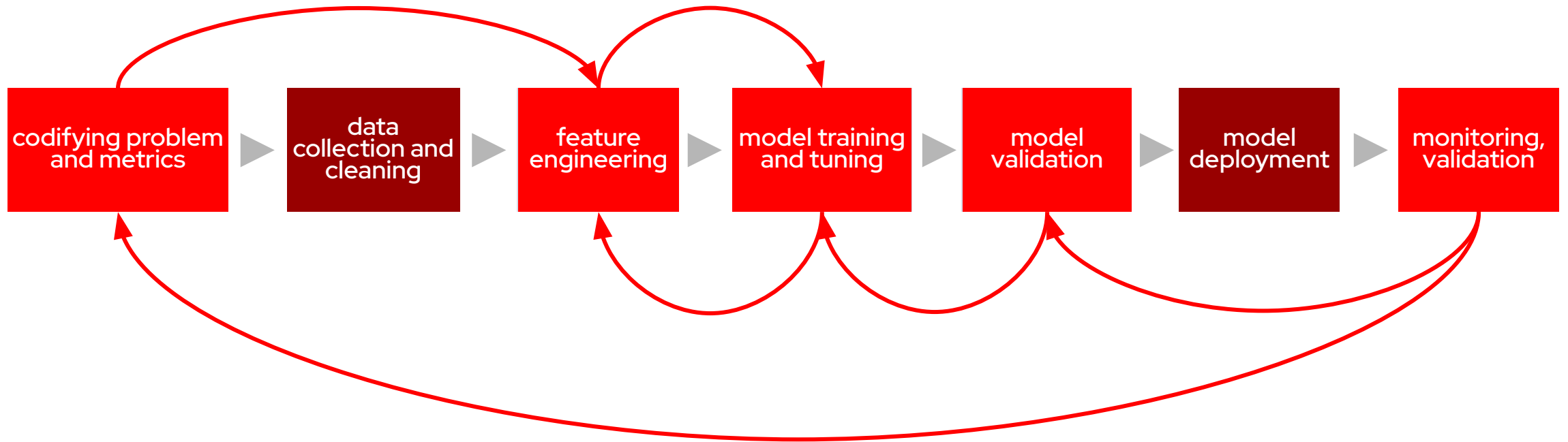
Managed cloud
(Azure, AWS, IBM, Google)

Data science: Models and Jupyter

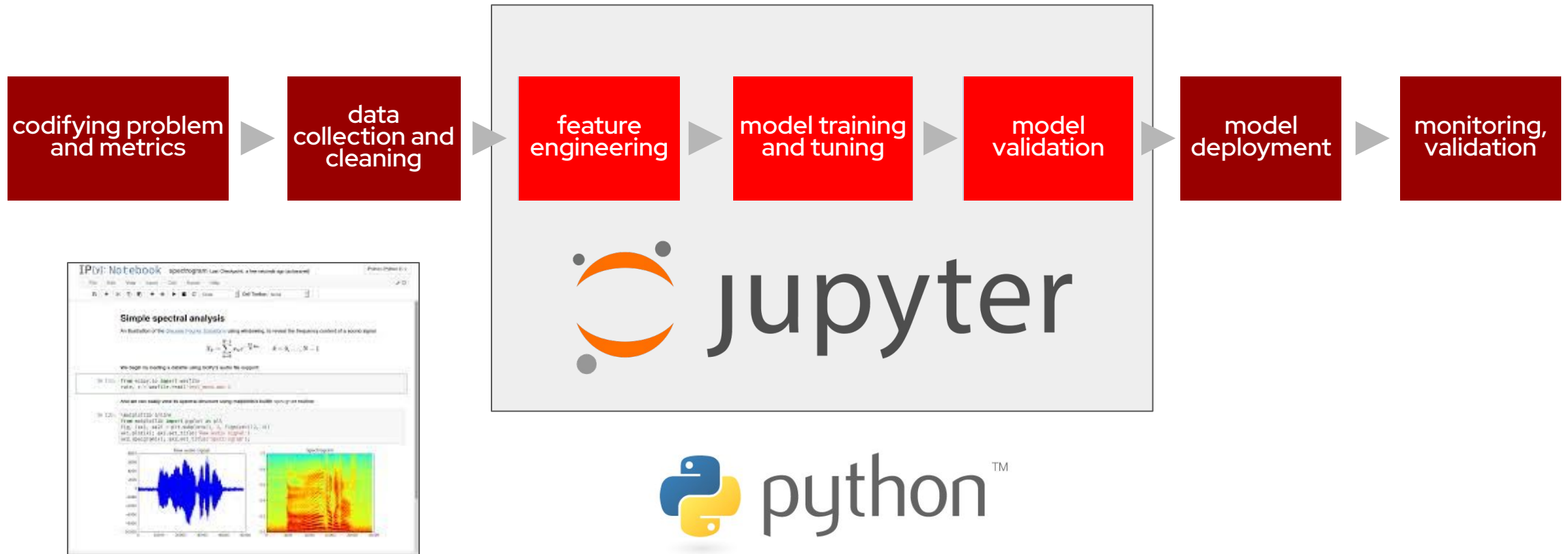
The data scientist's domain



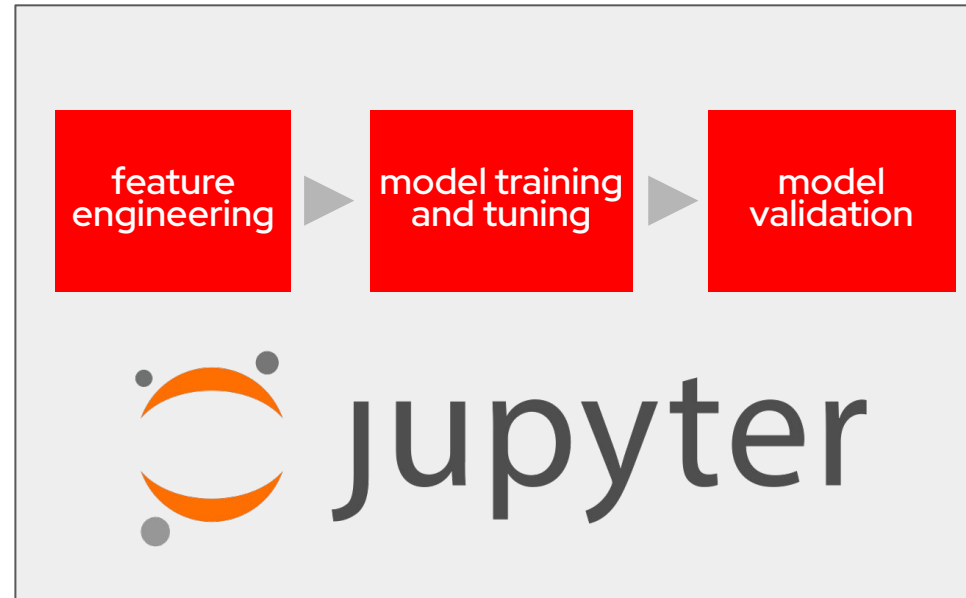
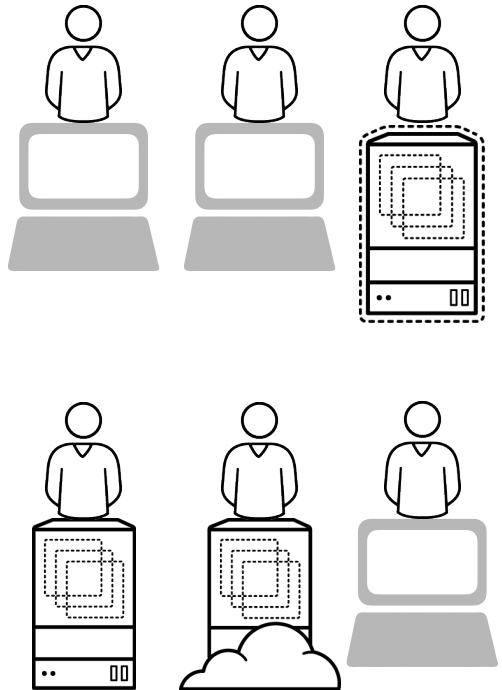
Data Science is iterative



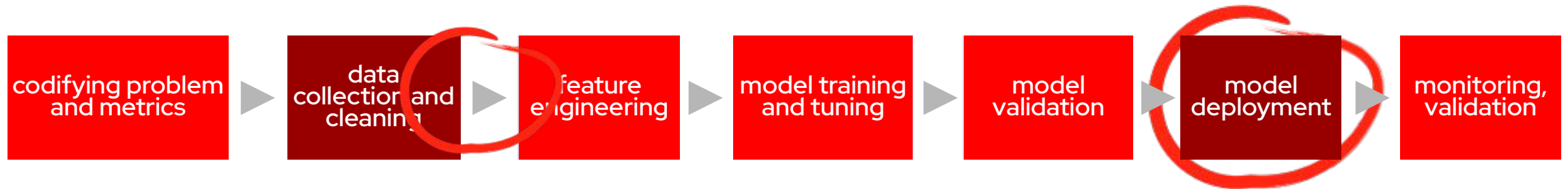
Jupyter(Hub/Notebooks) in data science



Everyone does the same thing... differently



Two primary bottlenecks



- Getting started
- Getting into production

The two bottlenecks

The data scientist's dilemma

Getting Started:

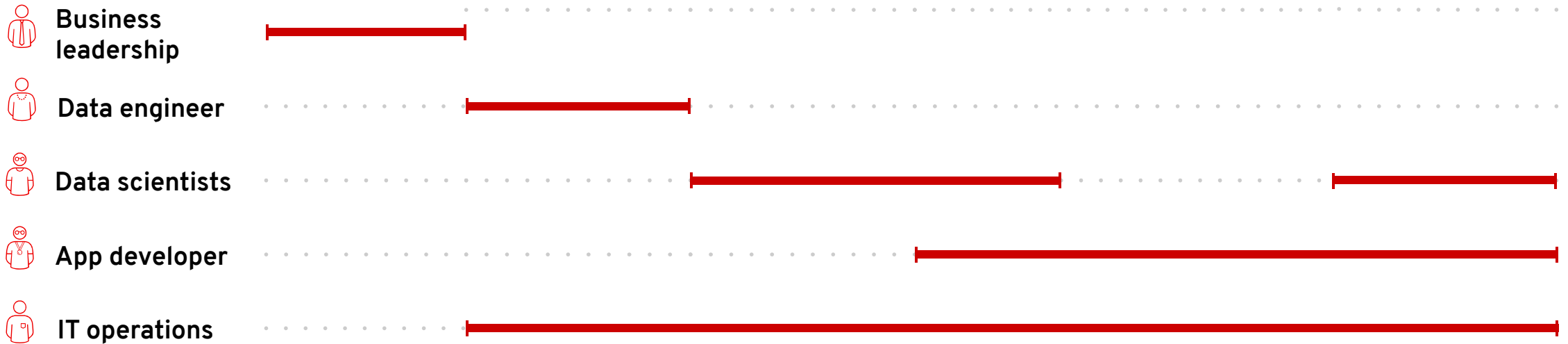
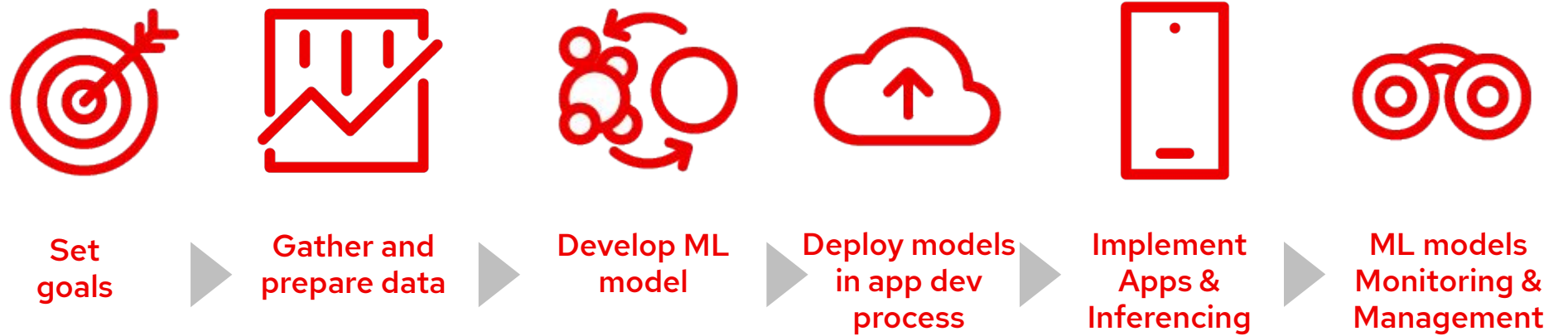
- ▶ Self service
- ▶ Repeatable/shareable environments
- ▶ Repeatable/shareable experiments
- ▶ Access to specialty resources (GPU, etc)

Getting into production:

- ▶ Self service
- ▶ Ease of instrumentation
- ▶ Access to specialty resources (GPU, etc)
- ▶ (Automated) scale-out

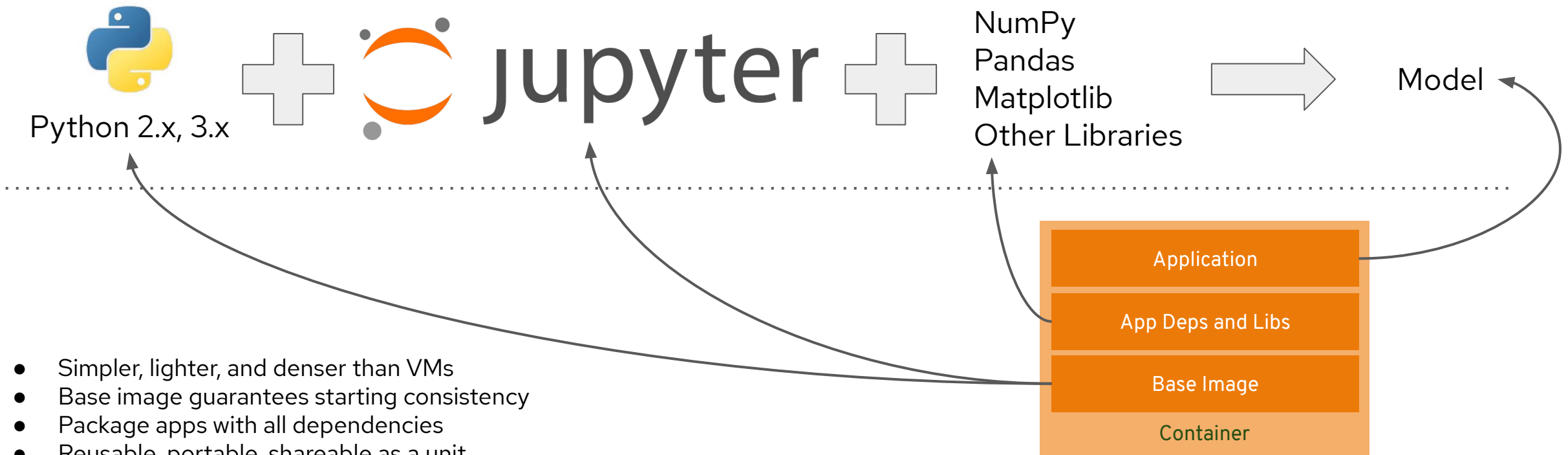
OpenShift for data science: Getting started

The ML lifecycle and responsibilities



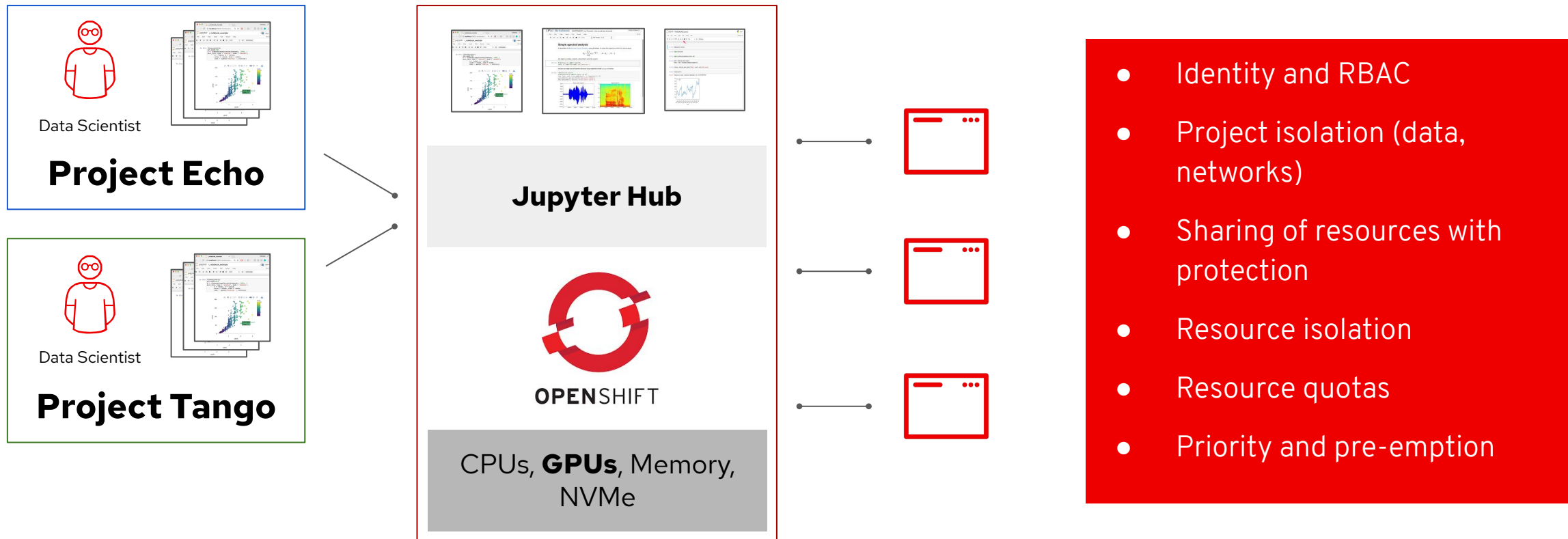
Containers for models

Reproducible and shareable environments for building, training and serving



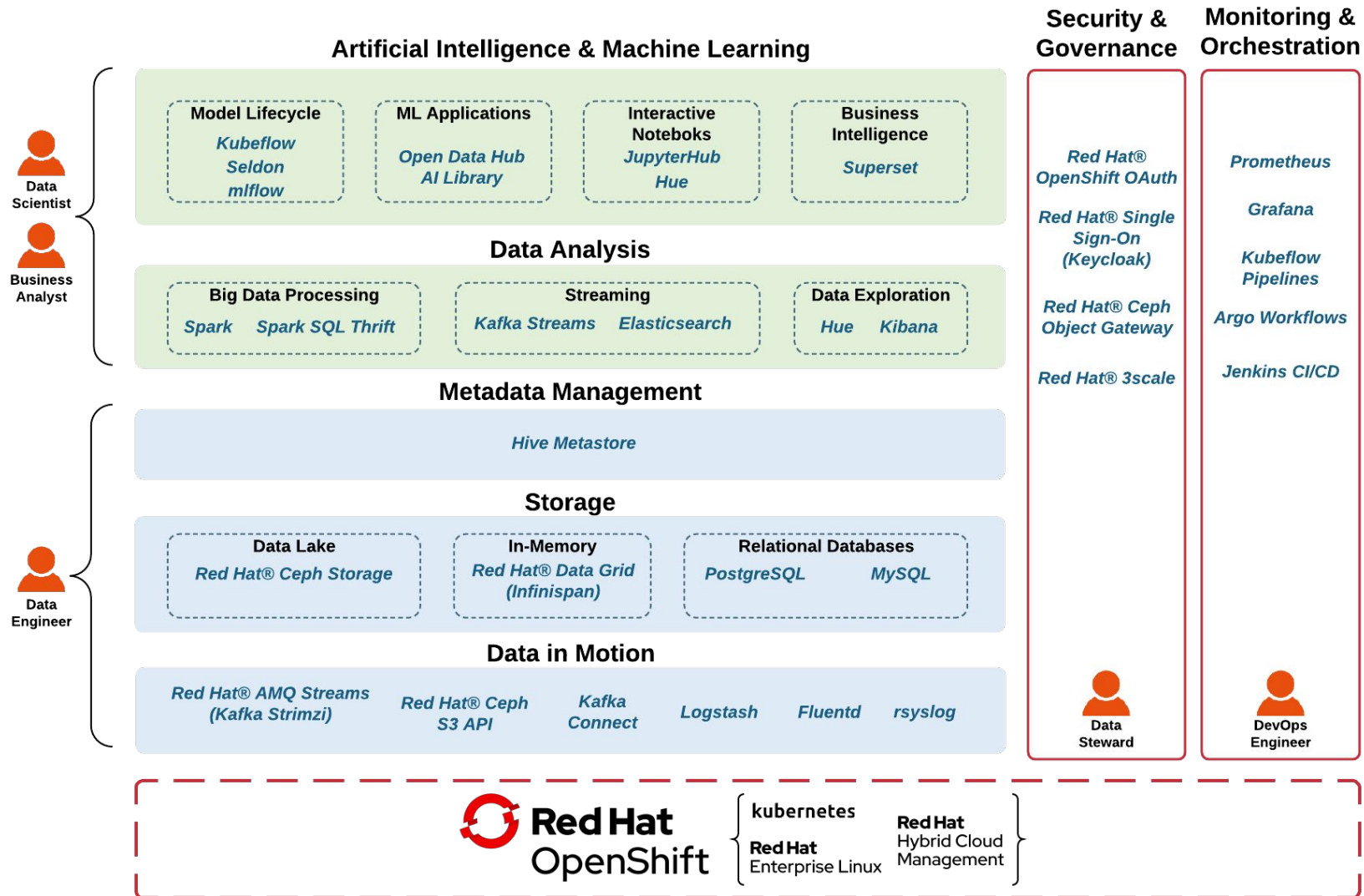
- Simpler, lighter, and denser than VMs
- Base image guarantees starting consistency
- Package apps with all dependencies
- Reusable, portable, shareable as a unit

Jupyter-Notebook-as-a-Service



Open Data Hub: A reference implementation of the possible

Open Data Hub: A community effort



Open Data Hub: Projects and Community



*

Prometheus

- Monitoring and alerting toolkit
- Used to diagnose problems



*

Grafana

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



- 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



- Distributed Object Store
- S3 Interface



- Distributed event streaming
- Pub/Sub Messaging



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

Ecosystem Partnerships

Strategic partnerships within AI/ML ecosystem

AI/ML Lifecycle



Data Governance & Security



Data Processing



Data Analytics



Databases



AI Ops



Infrastructure Partners



Hardware Acceleration

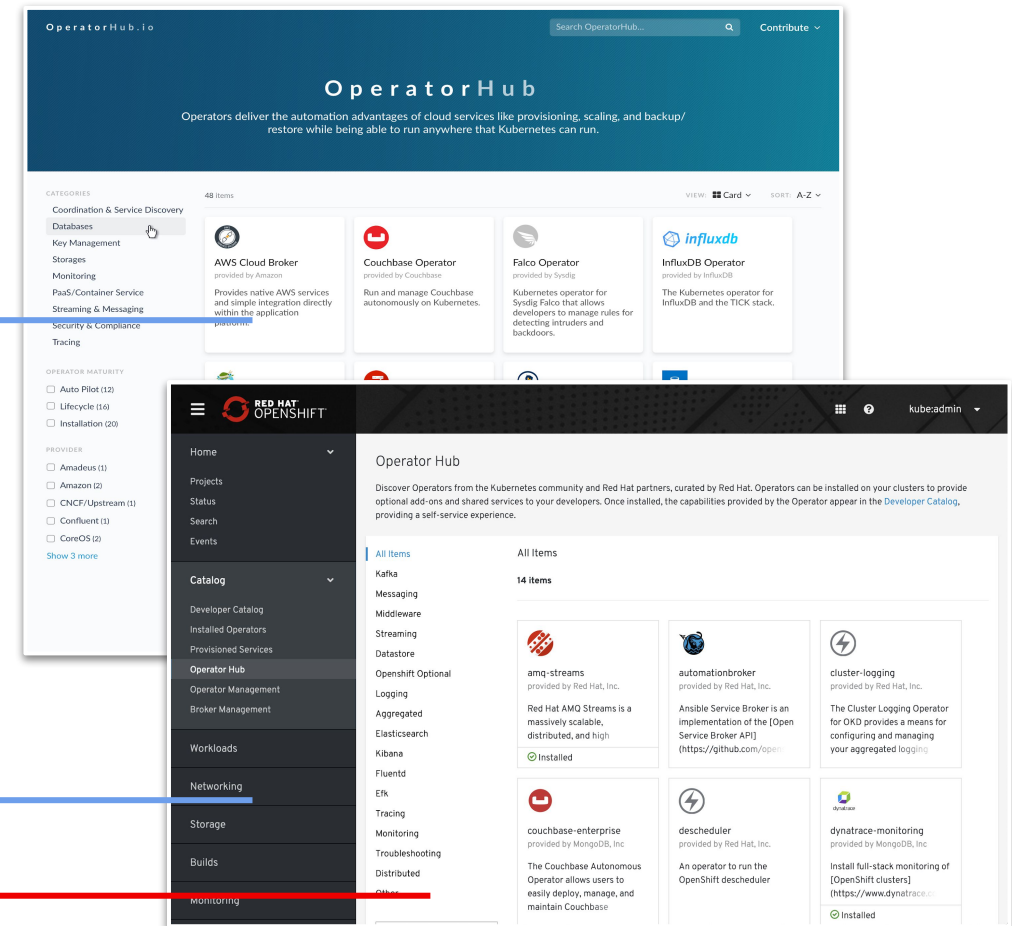


OperatorHub and certified Operators

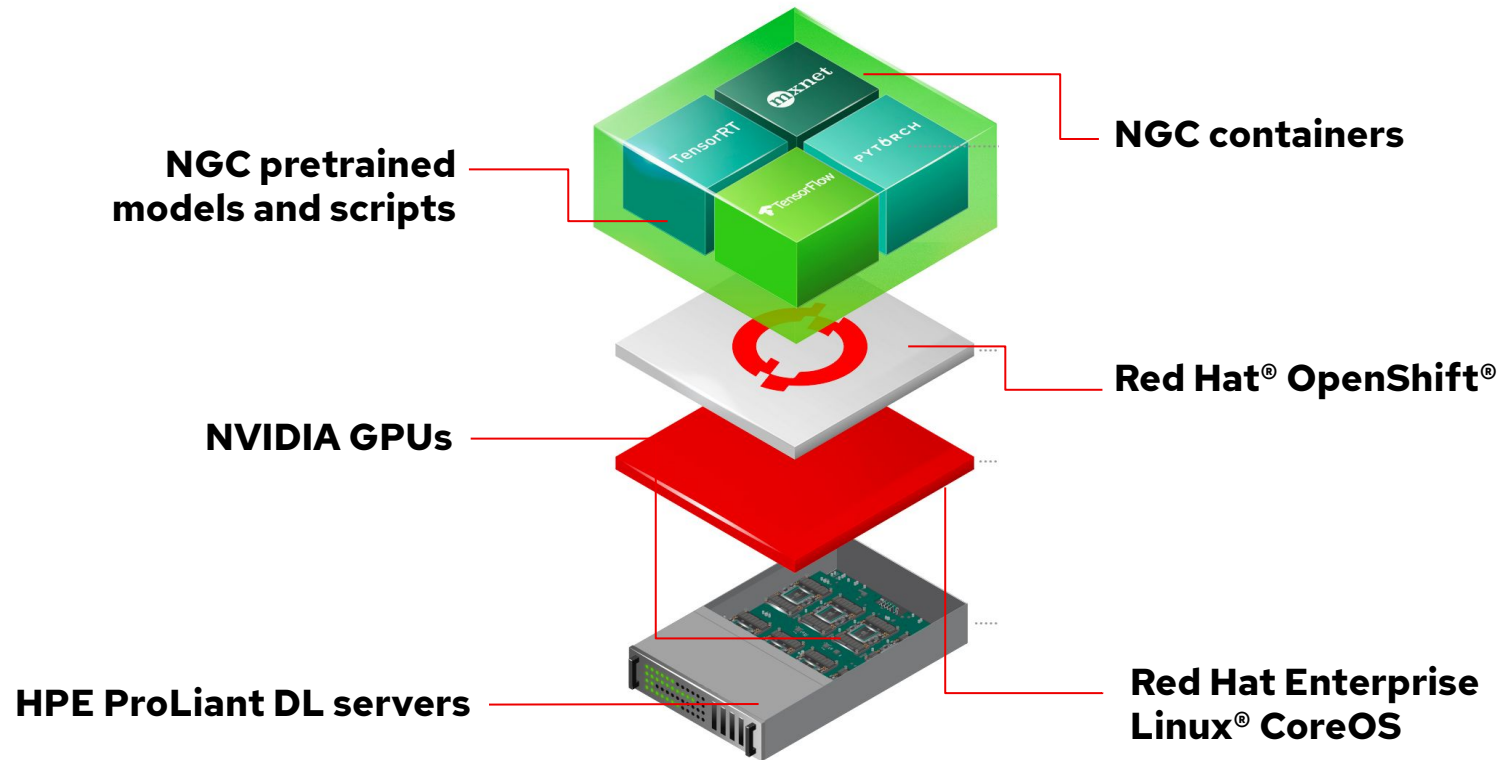
- Kubernetes Operators for life deployment and lifecycle management
- Red Hat Marketplace for ISV operators
- OpenShift Operator Certification
- OperatorHub integrated into OpenShift 4

COMMUNITY OPERATORS

OPENSIFT CERTIFIED OPERATORS



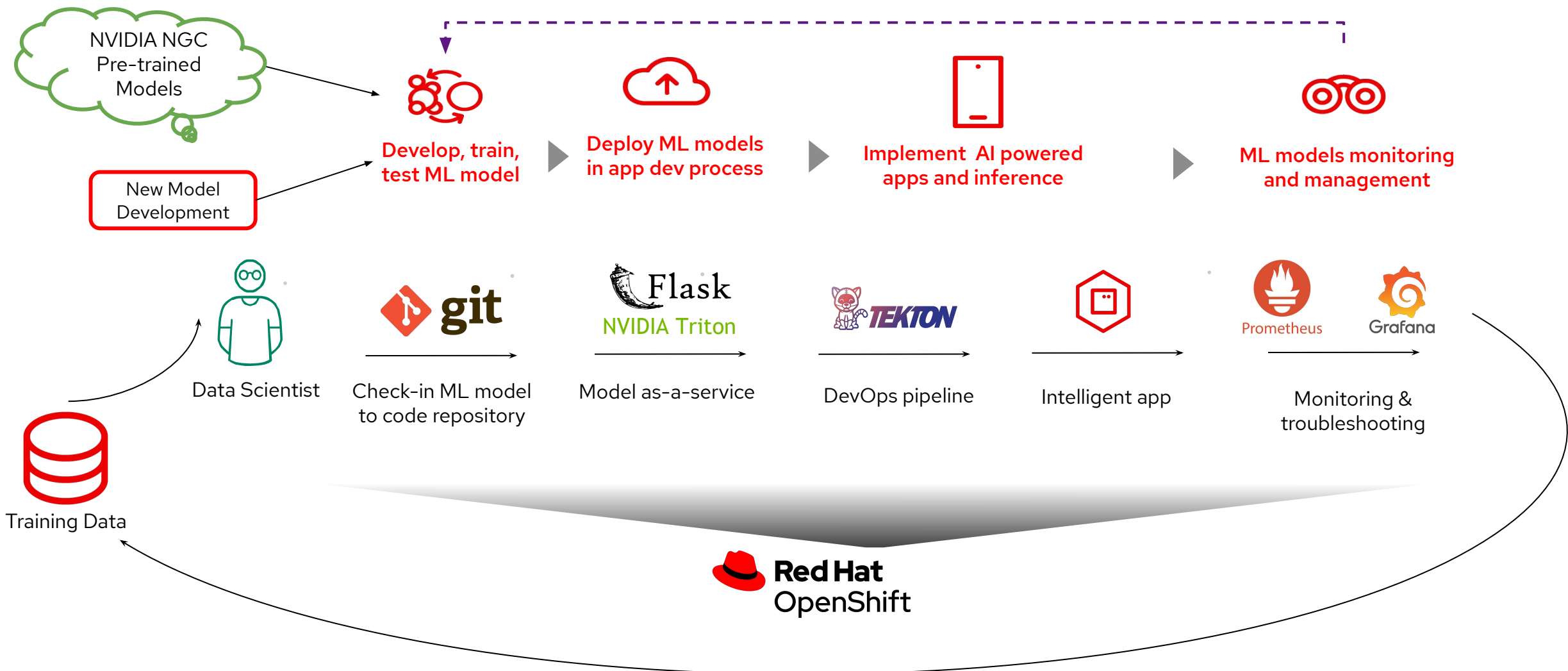
Accelerated AI Platform by HPE, NVIDIA, and Red Hat



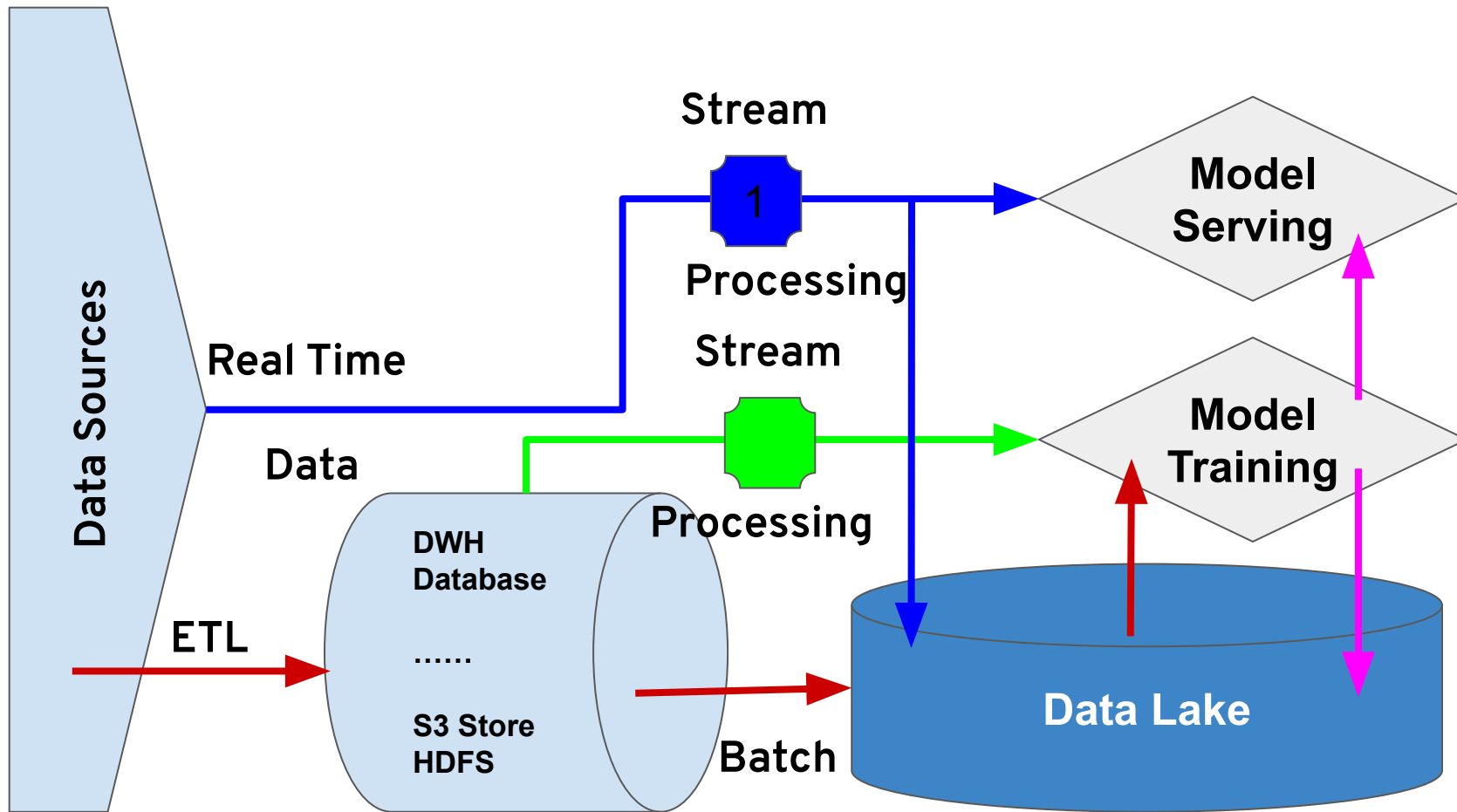
Deployed and Managed by the NVIDIA GPU operator on OpenShift

Backed by [Reference Architectures](#) and enterprise support from HPE, NVIDIA, and Red Hat

Fast Tracking AI projects with NVIDIA NGC & GPUs, and Red Hat OpenShift

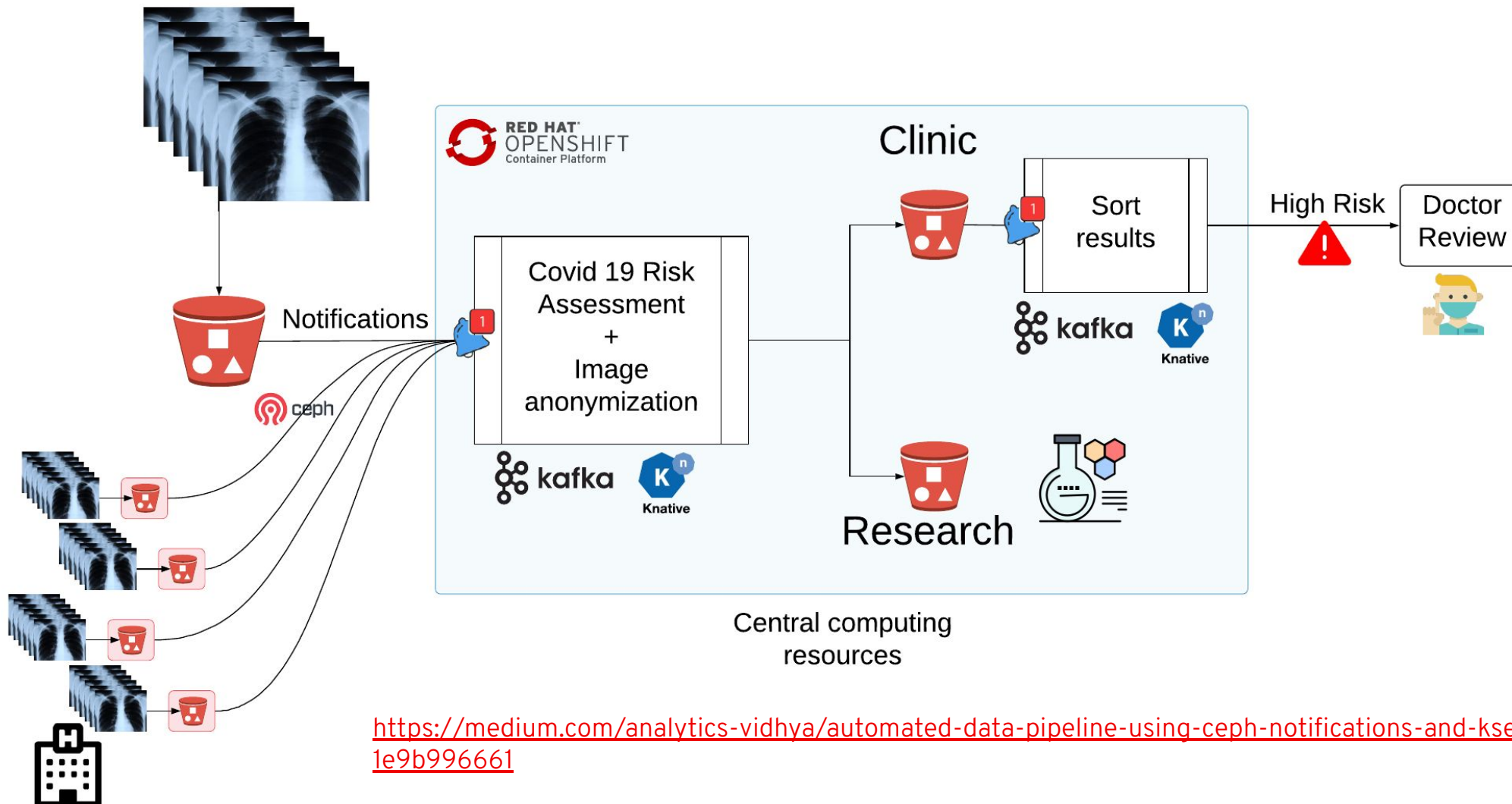


Data Pipelines Scenarios



- Many data sources
- Many data pipelines and situations can exist
- Machine Learning should be tap into all these data pipelines

Covid19 Assessment Pipeline

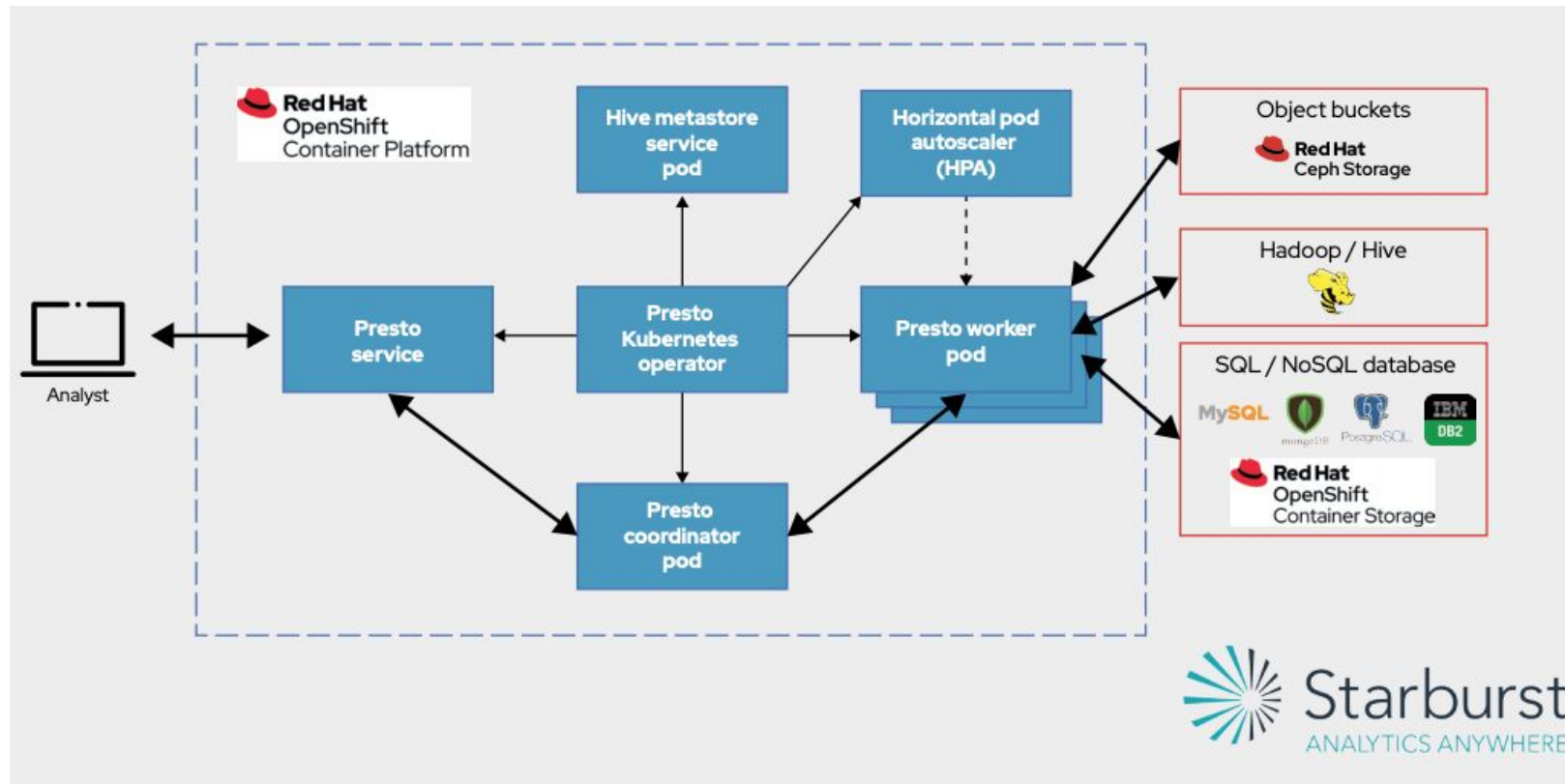


<https://medium.com/analytics-vidhya/automated-data-pipeline-using-ceph-notifications-and-kserveing-5e1e9b996661>

<https://github.com/guimou/dataprep-code/tree/master/examples/xray>

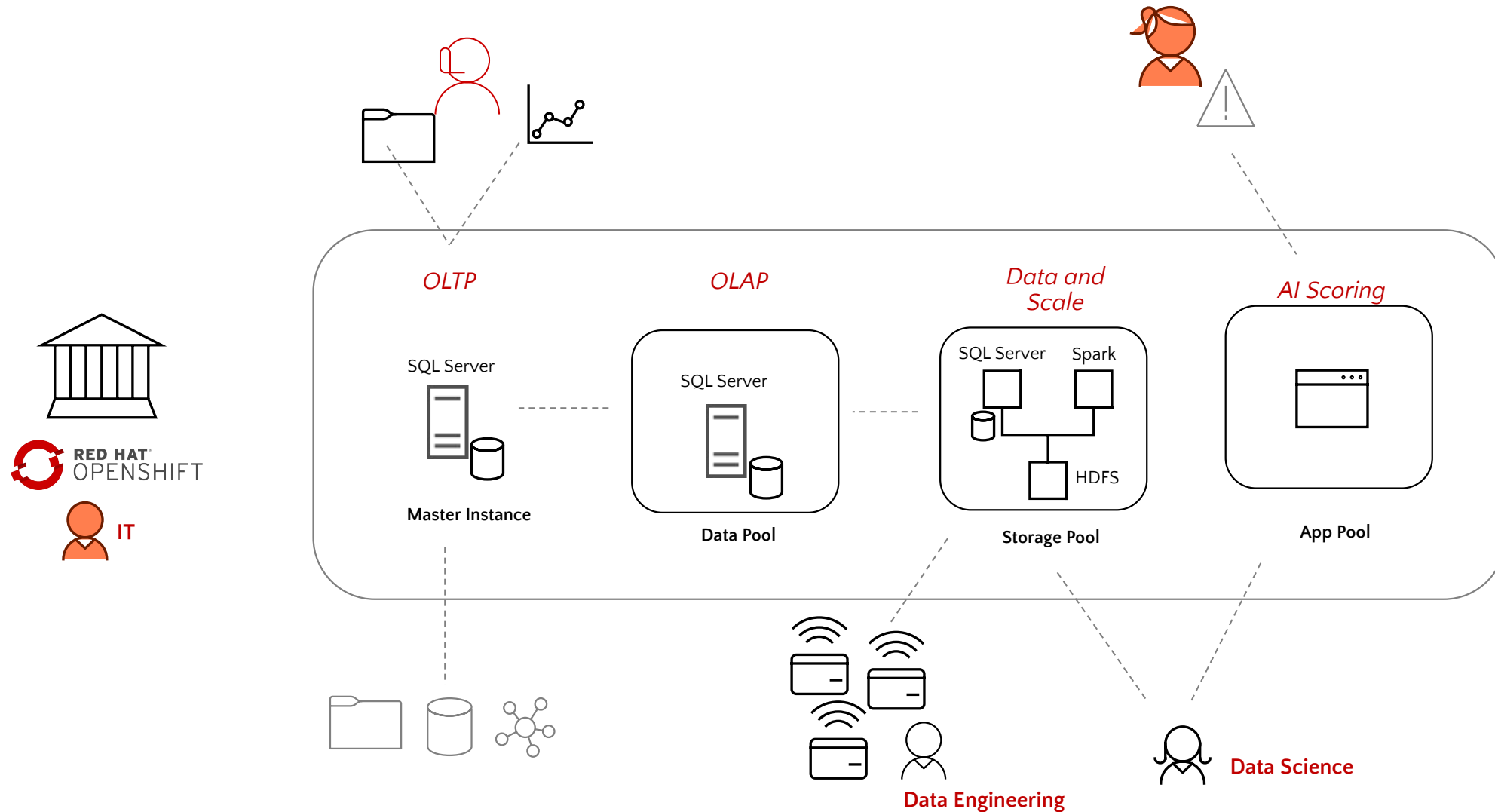
Starburst Enterprise for Presto on Red Hat OpenShift

Federated queries across multiple software-defined data sources



SQL Server Big Data Cluster – Fintech

OLTP, OLAP, and Fraud Detection



Cloudera CDP Private Cloud



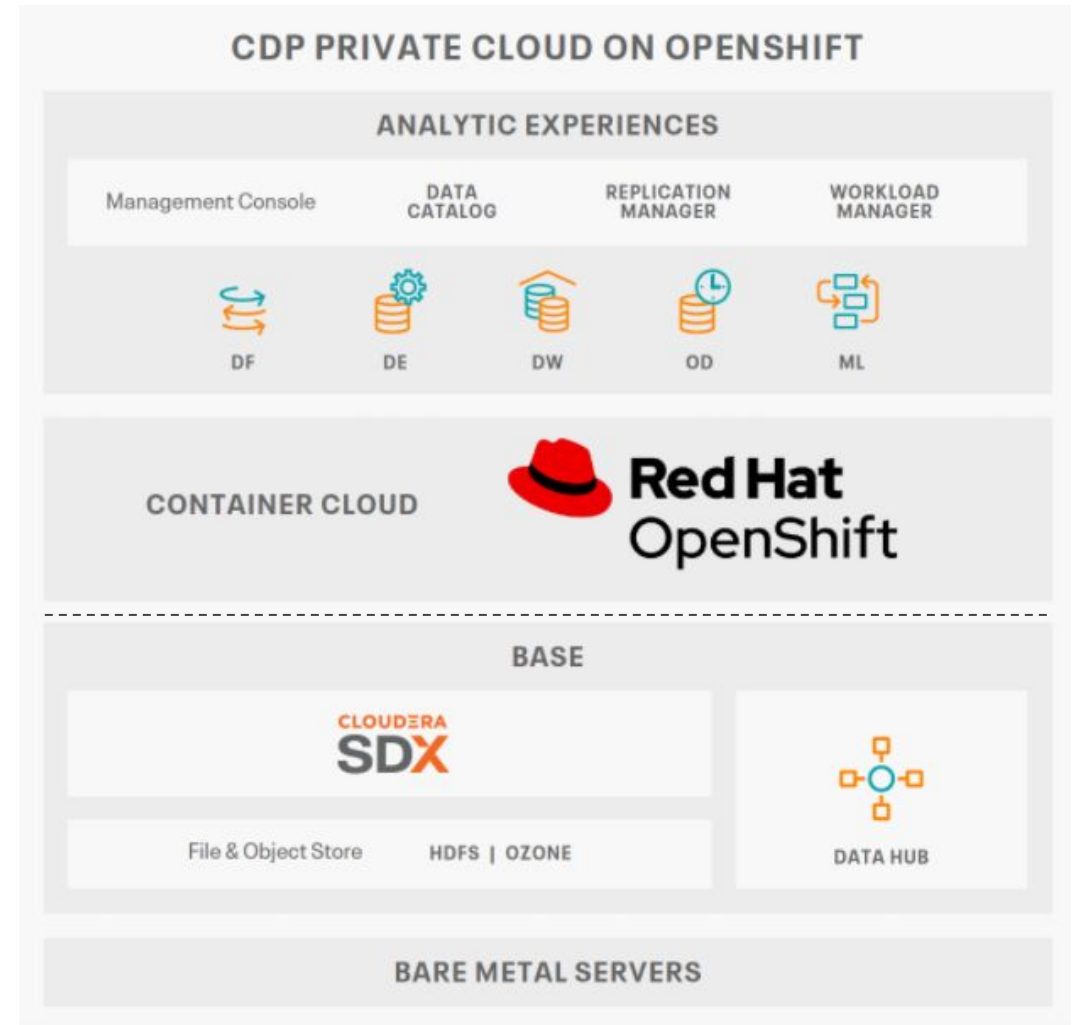
New set of data analytics applications
Featuring use-case optimized interfaces



Running on a container cloud
Fast provisioning & scaling, efficient, simple

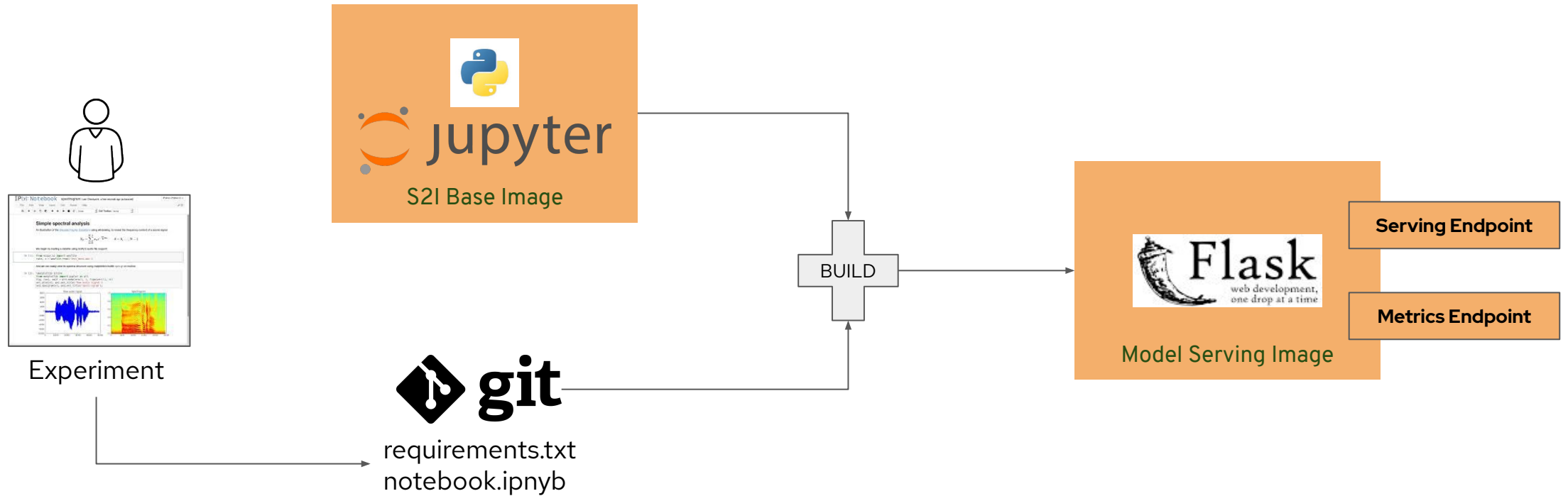


With access to a shared data lake
That is secured and governed



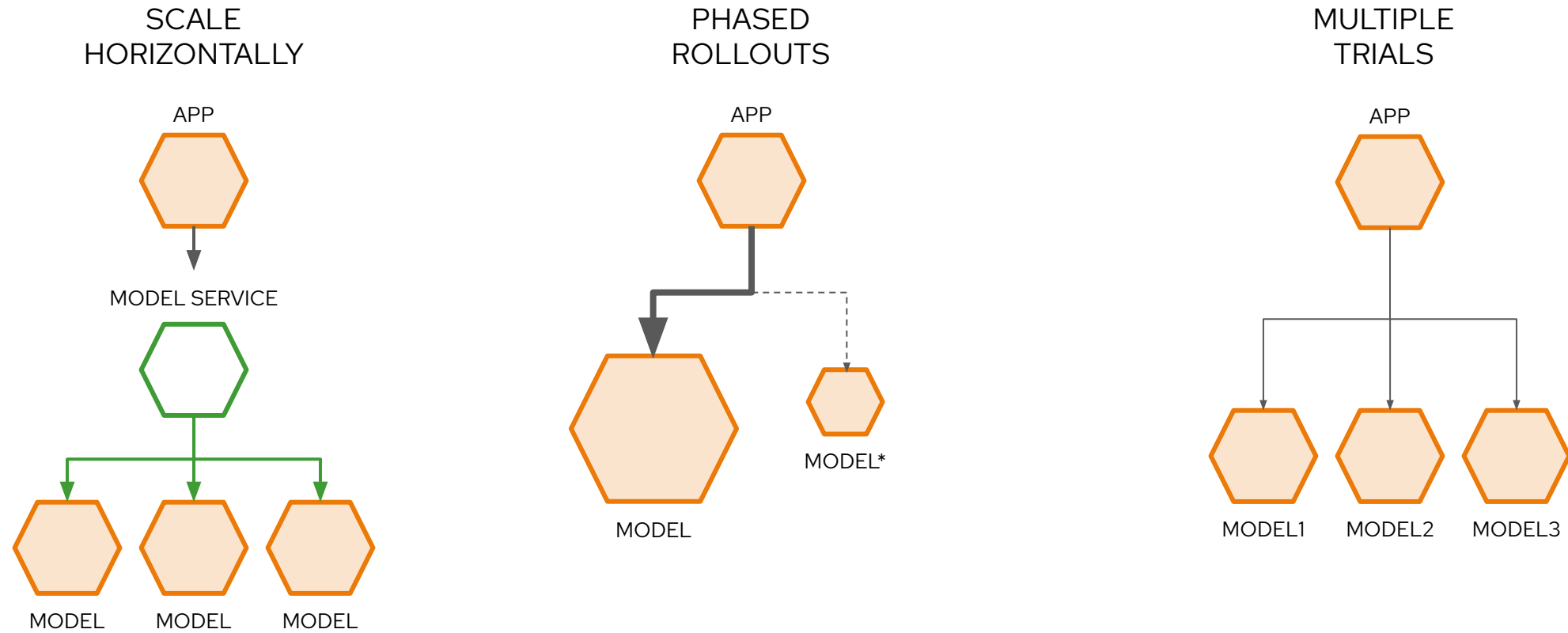
Source-to-Image for models

Automating build, deployment, APIs and instrumentation



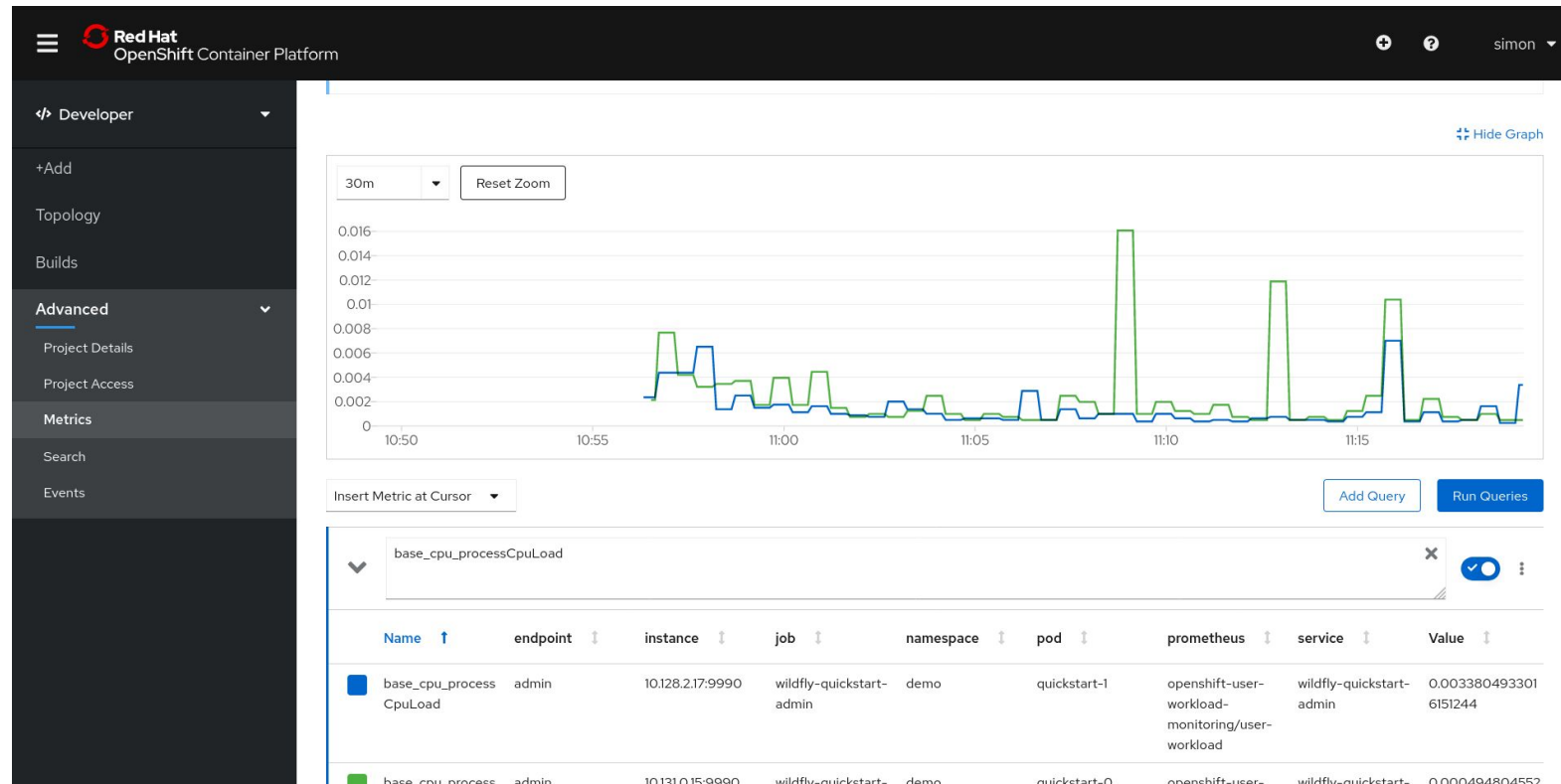
Models as stateless microservices

The power of Kubernetes

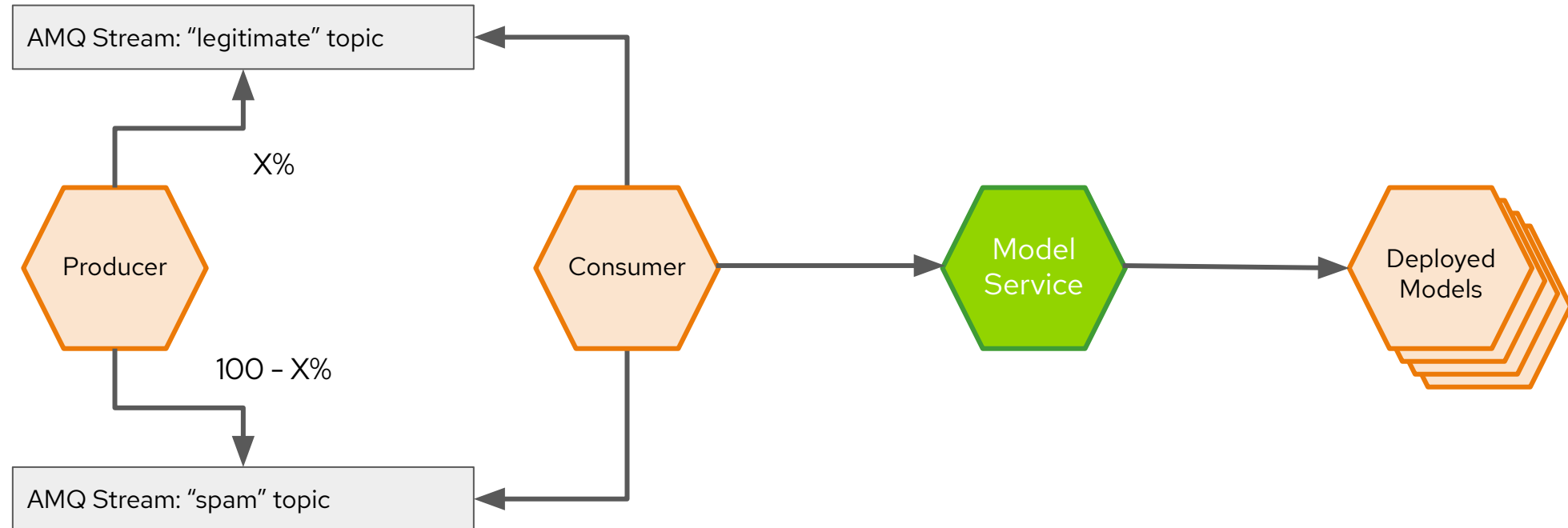


Monitoring performance and drift

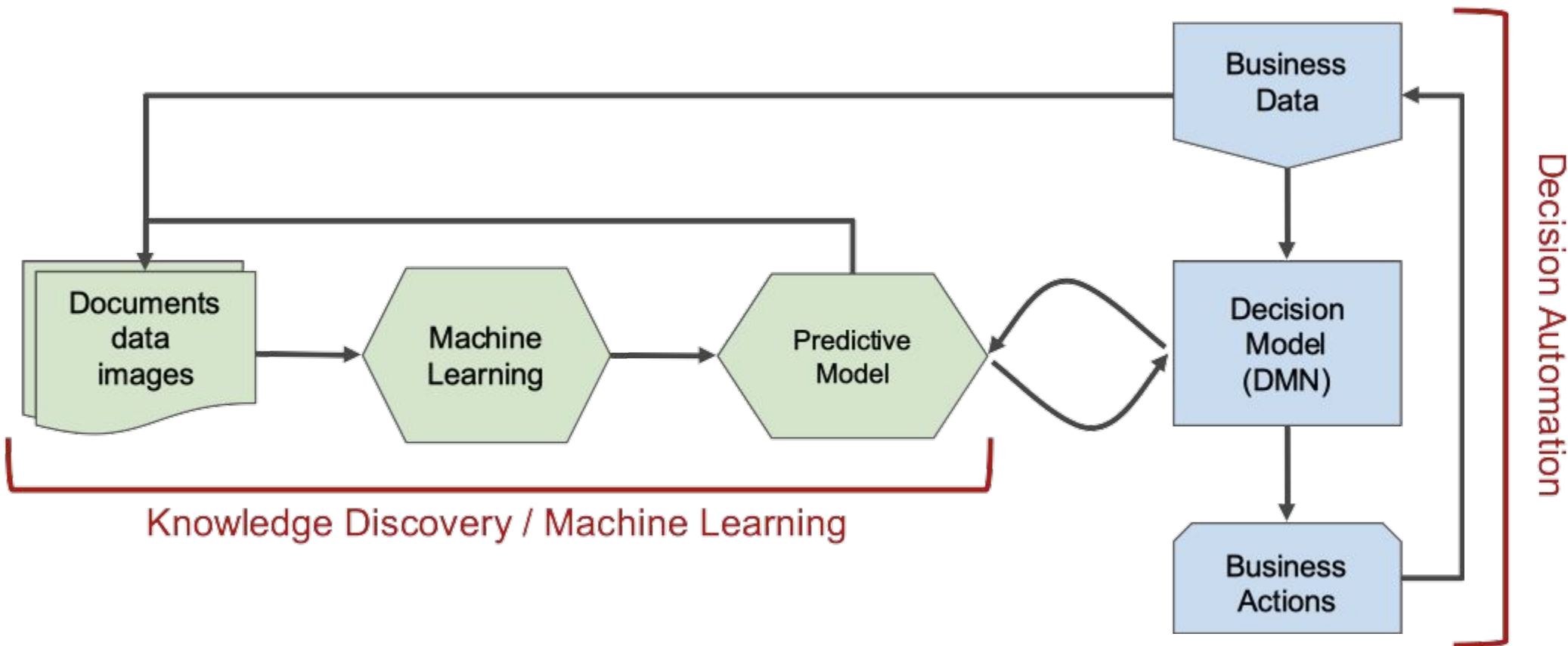
Keeping track of models over time



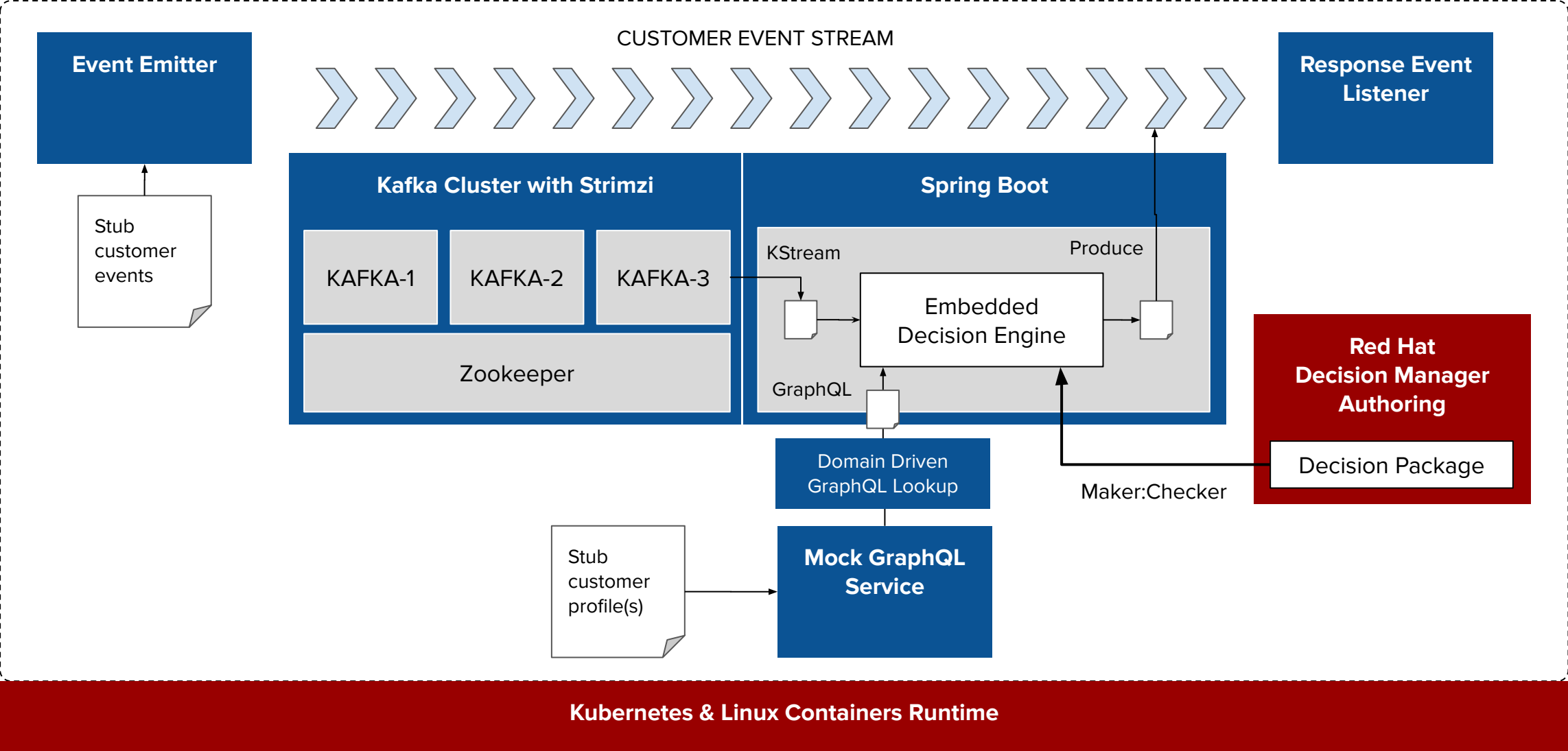
Intelligent spam detection - app architecture



Predictive Decision Modeling - Decisions driven by AI / ML



Example Scenario TOPOLOGY: EVENT CLOUD DECISIONS



AI/ML success stories



Partner to bring
COVID-Net radiography
screening AI to hospitals

Red Hat Press Release (11/16/20): [DarwinAI and Red Hat Team Up to Bring COVID-Net Radiography Screening AI to Hospitals, Using Underlying Technology from Boston Children's Hospital](#)

Boston Children's Hospital and DarwinAI COVID-Net radiography screening AI

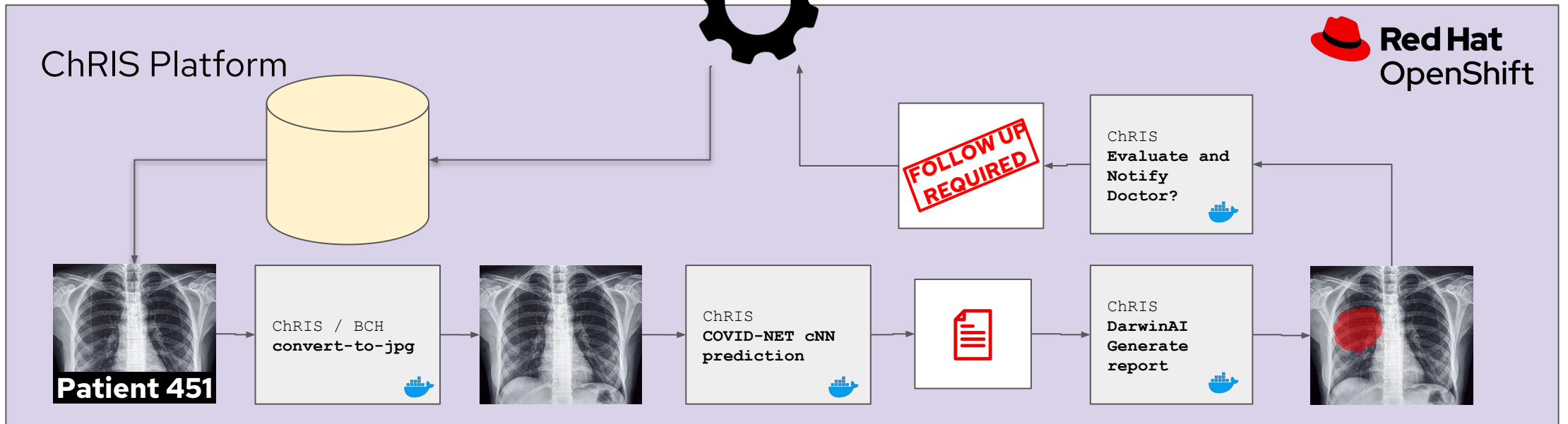
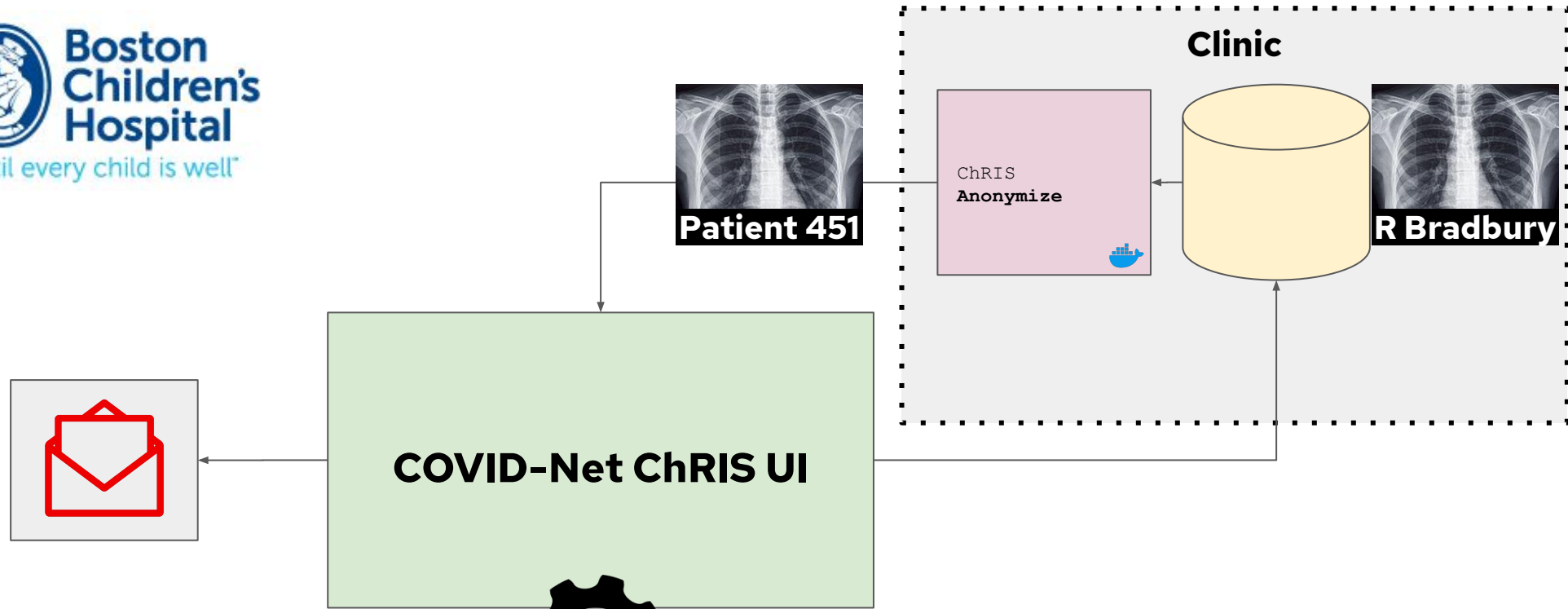
The challenge:

COVID-Net, a suite of deep neural networks for COVID-19 detection and risk assessment developed by DarwinAI and University of Waterloo's Vision and Image Processing Lab, is continuing to evolve at a rapid rate as the understanding of the disease evolves. These tools are rather technical and perhaps not very doctor-friendly in their raw form, and because of that, there has been a growing need to make COVID-Net more readily accessible to front-line clinicians around the world.

The solution:

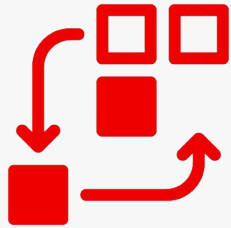
By working with Boston Children's Hospital and Red Hat, the COVID-Net project has been able to use the BCH's CHRIS platform infrastructure and Red Hat OpenShift to build a container-based workflow to easily deploy the tools into the hands of clinicians on the front lines. Work started early Summer 2020. Within a few days, the COVID-Net suite was fully containerized on the CHRIS platform.





RBC Machine Learning – The challenge

AI/ML projects took up to 2 months to get off the ground, slowing research and new offerings



Workload orchestration and self-service access to GPU resources was manual and slow



AI/ML platforms were hard to build and required significant resources



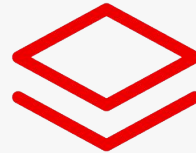
Security and compliance requirements were hard to meet

The solution

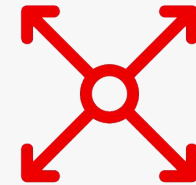
Accelerated, more secure and flexible AI/ML infrastructure on Red Hat OpenShift, NVIDIA



ML applications and services
deployed in containers



Consistent private cloud platform for
bank's AI/ML research center



NVIDIA Operator
simplifies use of GPUs

The Goal

Digital transformation of BMW Group from Cars to Devices and Data

Provide a superior **connected car experience** for 12.5 million drivers

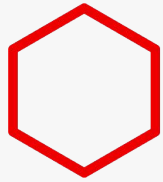
Develop driving algorithms and analyze data to support **autonomous vehicle project**

Adopt a **DevOps culture** and leverage cloud to improve efficiency, application release cycles

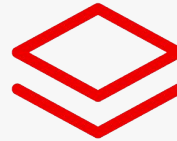


The Solution

Cloud native platform lets developers focus on building apps



New ConnectedDrive backend run on Red Hat OpenShift and OpenShift Dedicated (Managed Service on AWS)



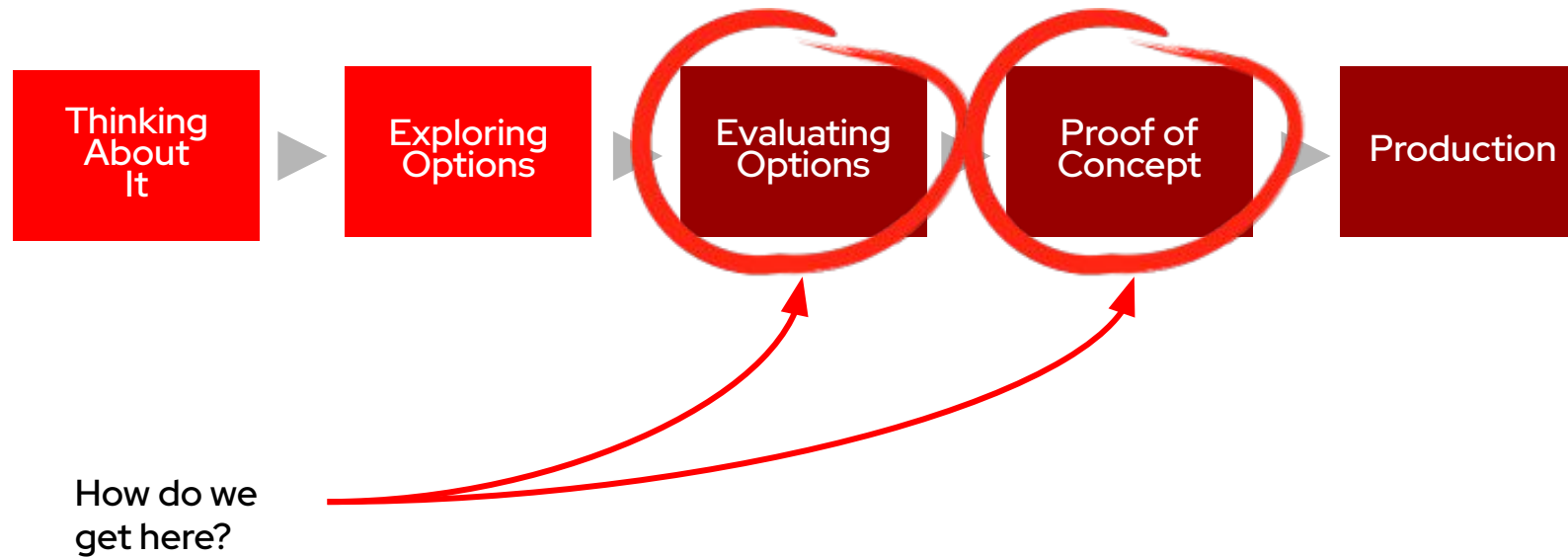
High performance D3 Platform to analyze massive amounts of global test fleet data



Cloud native DevOps platform build on Red Hat OpenShift

Where Next?

The intelligent application journey



AI/ML discovery session

A one day, no-cost planning session



GOAL:

To understand the customer's business drivers and technical use cases to propose an Open AI/ML architecture.



DETAIL:

Discussion guided by Red Hat AI/ML Consulting Architect

Attendees from Data Science team, Business decision makers, Engineering, Operations, and Application Development

Deep dive into customer use cases



RED HAT PROVIDES:

Our vision for organization-wide AI/ML adoption

Tailored proposal to solve customer use cases

High-level adoption roadmap

Red Hat AI/ML service offerings

Who attends the AI/ML discovery session?

Customer

- Data scientists
- Data engineers
- IT operations
- Intelligent App developers
- Business sponsors

Red Hat

- AI Center of Excellence (leads)
- Red Hat Services
- Open Innovation Labs
- Solutions Architect (optional)
- Account Team (optional)

AI/ML Architecture Review

Vendor lock-in and high costs of scaling AI/ML applications



Open and flexible architecture for AI/ML applications

Inputs

- Customer AI/ML use cases
- AI/ML platform requirements

Scope

- Red Hat OpenShift
- 1 week

Activities

- Day 1: Customer use cases deep dive
- Day 2: AI/ML architecture review workshop
- Day 5: Live demo in customer lab
- Mentoring and roadmap creation

Key contacts

- Neeraj Kuppam - Services Portfolio Team
- William Benton - AI Center of Excellence

Outcomes

- AI/ML platform strategy
- AI/ML architecture and roadmap
- Live demo in customer lab



Red Hat Services Solution: Open AI/ML Platform

Why

- Accelerate your AI/ML project using the Open Data Hub architecture
- Create a pilot platform for AI/ML use cases
- Bring data scientists, app developers and operations teams together
- Experience DevOps for machine learning best practices
- Build an AI powered application prototype

What

- AI/ML Consulting + Open Innovation Labs Residency
- Infra: Open Data Hub + OpenShift + Open Innovation Labs tools

Who

- Neeraj Kuppam, Will Benton, Patrick Carney, Eric Murphy





Check out our AI/ML success stories

- [RBC](#)
- [HCA Healthcare](#)
- [Boston Children's Hospital](#)
- [BMW Group](#)
- [ExxonMobil](#)
- [NEC](#)
- [KBTG](#)
- [Discover Financial Services](#)
- [Ministry of Defense, Israel](#)

eBook (8 pages) – [Top consideration for building a production-ready AI/ML environment](#)

Websites

- <https://openshift.com/ai-ml>
- <https://learn.openshift.com/ai-machine-learning/>
- <https://openshift.com/nvidia>

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