

Analytics and Machine Learning with Open Source Infrastructure

Devconf US, 2019

Anish Asthana, Software Engineer

Juana Nakfour, Senior Software Engineer

Creating Intelligent Applications for the Hybrid Cloud

Intelligent Applications

Make extensive use of data with machine learning technology (models) to deliver rich, adaptive and personalized experiences for users.

Developing Intelligent Applications

Data Pipeline

Machine Learning Pipeline

Application Pipeline

Intelligent Applications for Hybrid Cloud

Portability

Agility

Scale

Resilience

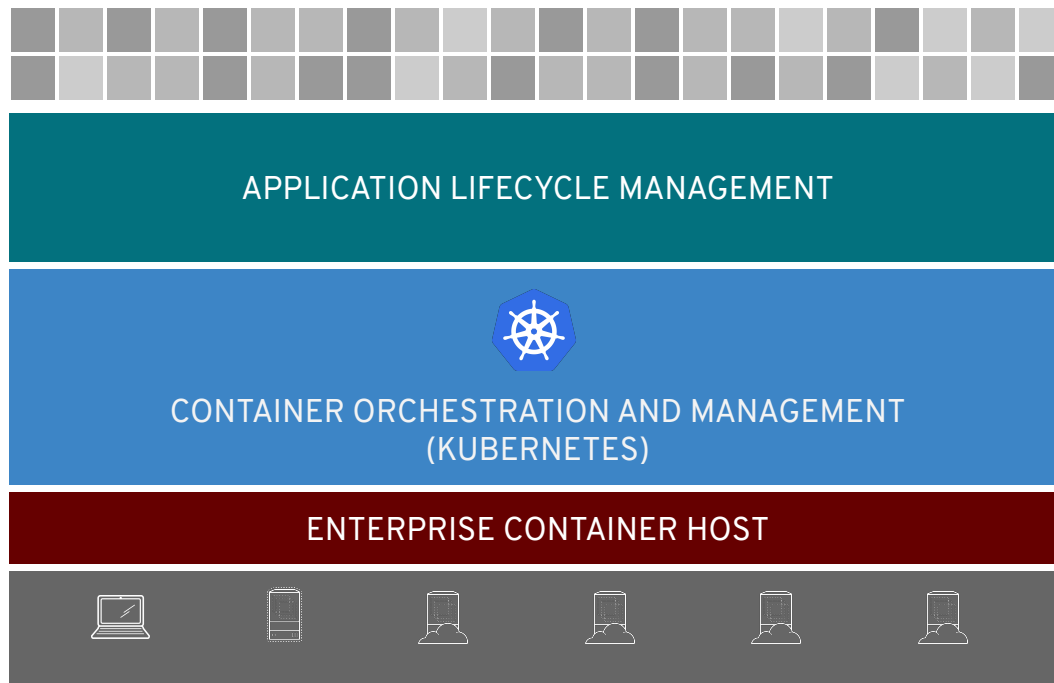
Self Healing

Automatable

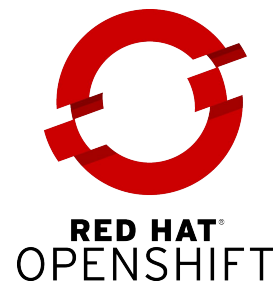
Intelligent Applications

Make extensive use of data with machine learning technology (models) to deliver rich, adaptive and personalized experiences for users.

OpenShift - Distro of Kubernetes



ANY
CONTAINER



ANY
INFRASTRUCTURE

OpenShift Architecture

Best of SDLC



DEVELOPER
DATA SCIENTIST

SCM
(GIT)

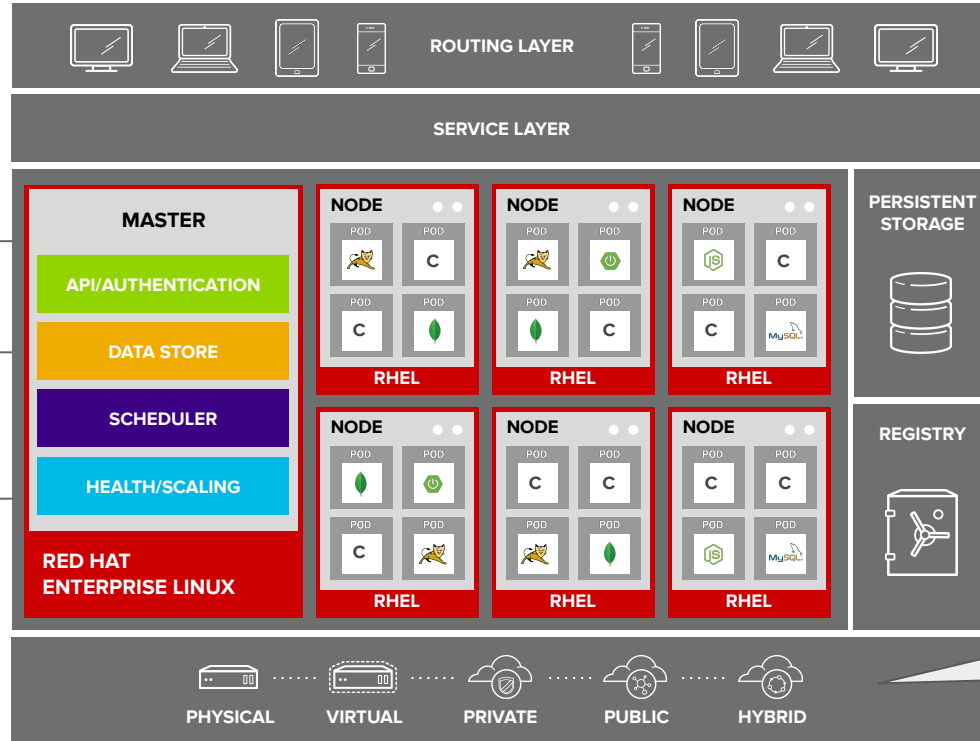
CI/CD



OPERATIONS

EXISTING
AUTOMATION
TOOLSETS

ML in
Production



Expose ML as
services, load
balanced and
scalable

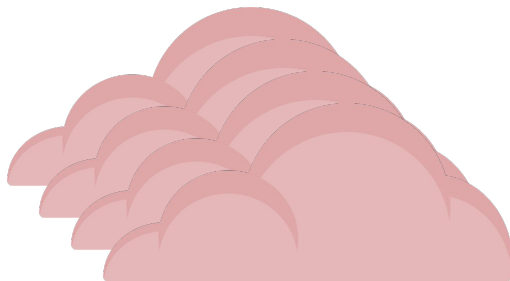
ML as
microservices that
are efficiently
orchestrated
across shared
resources

Deploy ML on any
cloud

Open Data Hub

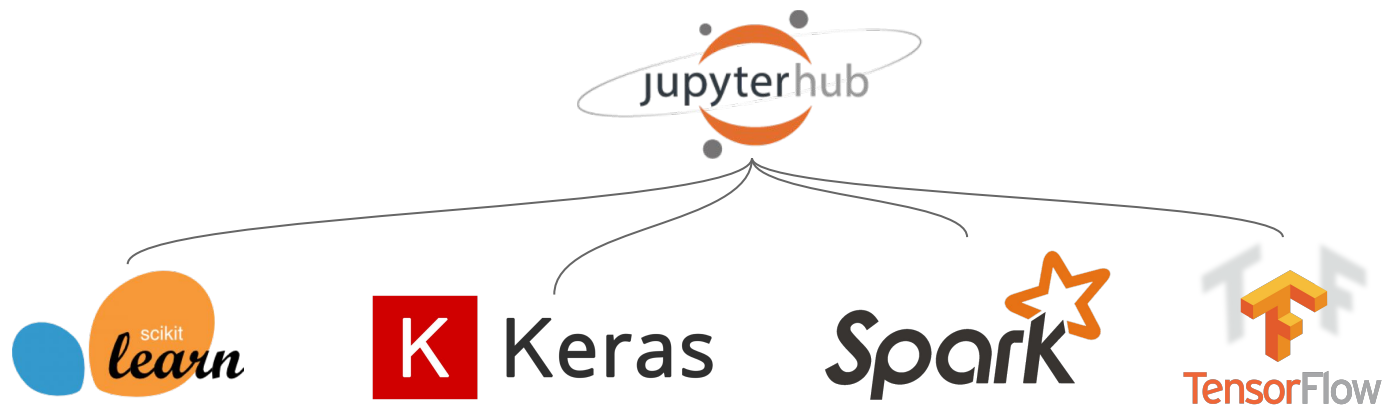
Meta open source project that brings in multiple open source technologies for data and machine learning pipelines on to Kubernetes (OpenShift) to create Intelligent Applications for hybrid cloud

Open Data Hub



Benefits of Data Science on Containers

- Data scientists can choose community provided images with exchangeable tool sets to get their job done.
- Collaboration and sharing of data and work

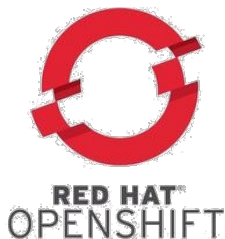


Benefits of Data Science on Containers

- CI/CD process for data science
- Growing ecosystem of data science products

Core Deployment

Available Now at [OpenDataHub.io](https://opendatahub.io)



- Container platform
- Certified Kubernetes
- Hybrid cloud



- Unified, distributed storage
- RESTful gateway
- S3 and Swift compatible



- Unified analytics engine
- Large-scale data
- Runs on Kubernetes



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

Add-Ons

Available Now at [OpenDataHub.io](https://opendatahub.io)

Open Data Hub AI Library



Prometheus

- Part of Open Data Hub
- Set of deployed pre-defined AI models available to use
- Monitoring and alerting toolkit
- Records numeric time series data
- Used to diagnose problems



Grafana

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



- Deploying machine learning models on Kubernetes
- Expose models via REST and gRPC
- Full model lifecycle management



Running on
OpenShift

Running at Red Hat

Artificial Intelligence & Machine Learning

Model Lifecycle

Seldon
MLFlow

ML Applications

Open Data Hub
AI Library

Interactive
Notebooks
JupyterHub
Hue

Business
Intelligence
Superset

Data Analysis

Big Data Processing

Spark *Spark SQL Thrift*

Streaming

Kafka Streams *Elasticsearch*

Data Exploration

Hue *Kibana*

Metadata Management

Hive Metastore

Storage

Data Lake

Red Hat® Ceph Storage

In-Memory

Red Hat® Data Grid
(Infinispan)

Relational Databases

PostgreSQL *MySQL*

Data in Motion

Red Hat® AMQ Streams
(Kafka Strimzi)

Red Hat® Ceph
S3 API

Kafka
Connect

Logstash

Fluentd

rsyslog



Security & Governance

Red Hat®
OpenShift OAuth

Red Hat® Single
Sign-On
(Keycloak)

Red Hat® Ceph
Object Gateway

Red Hat® 3scale

Monitoring & Orchestration

Prometheus

Grafana

Argo Workflows

Jenkins CI/CD



Red Hat

Link to Workshop

Resources

Thank You



linkedin.com/company/red-hat



youtube.com/user/RedHatVideos



facebook.com/redhatinc



twitter.com/RedHat