Data Engineering using Open Data Hub on OpenShift Hybrid Cloud

Devconf US, 2019

Anish Asthana, Software Engineer
Landon LaSmith, Senior Software Engineer
Juana Nakfour, Senior Software Engineer

Red Hat

Intelligent Applications

Introduction to **Kubernetes** and **OpenShift**

What is **Open Data Hub**

Workshop Components and Details

Intelligent Applications Development

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

Data **Pipeline**

Machine Learning **Pipeline**

Application **Pipeline**

Intelligent Applications for Hybrid Cloud

Portability

Agility

Scale

Resilience

Self Healing

Automatable

Kubernetes

Containers can also create *Chaos*

Manage: Security, Resources, Applications, at SCALE

Kubernetes is the *de facto* container platform for the **hybrid cloud**

Openshift Container Platform

Enterprise Kubernetes Application Platform

Open source version of OpenShift which is **okd.io**

Built-in Services and Features

Security: RBAC, strict security policies, authentication and authorization

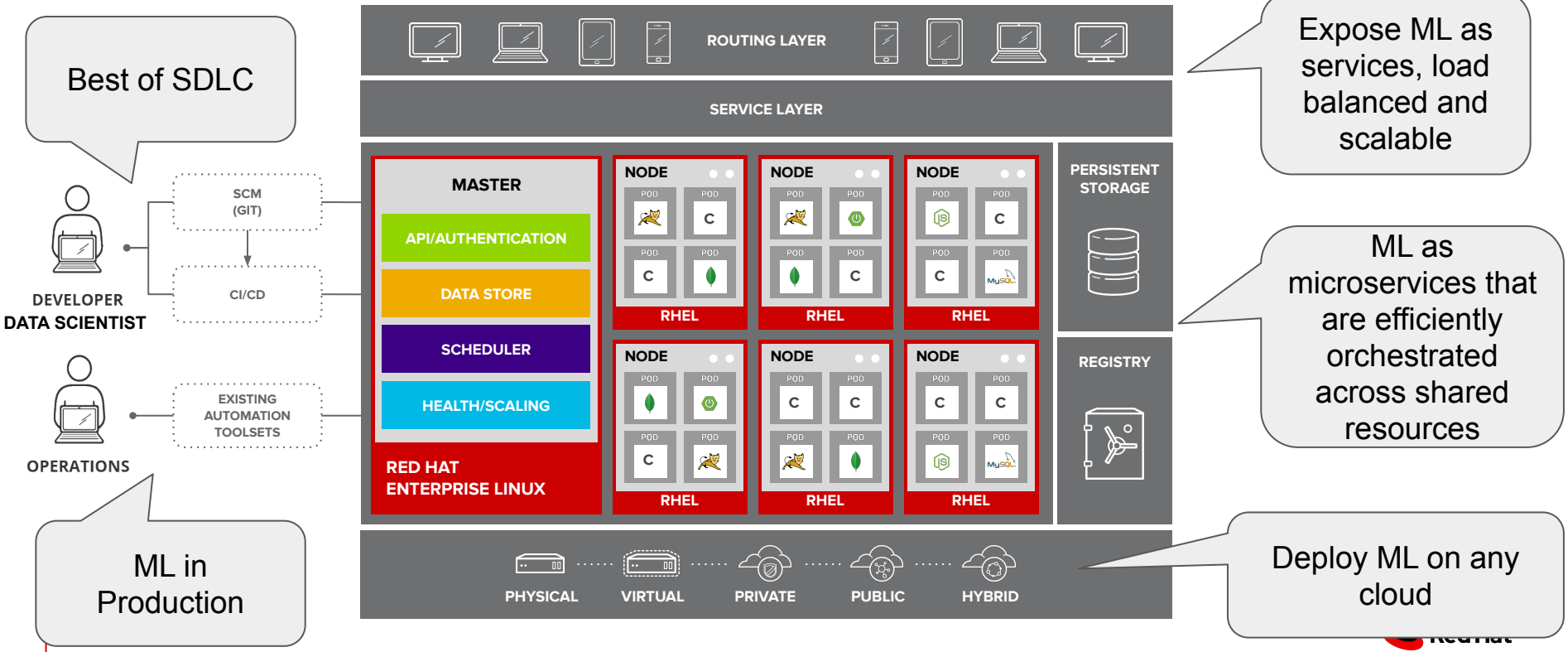
CI/CD with **Jenkins**

Enhanced **Developer Experience**: S2i, Imagestreams, Portal

Enhanced **DevOps** Experience: **Logging** stack based on EFK (ElasticSearch, Fluentd, Kibana) and **Monitoring** based on Prometheus

Red Hat

OpenShift Architecture



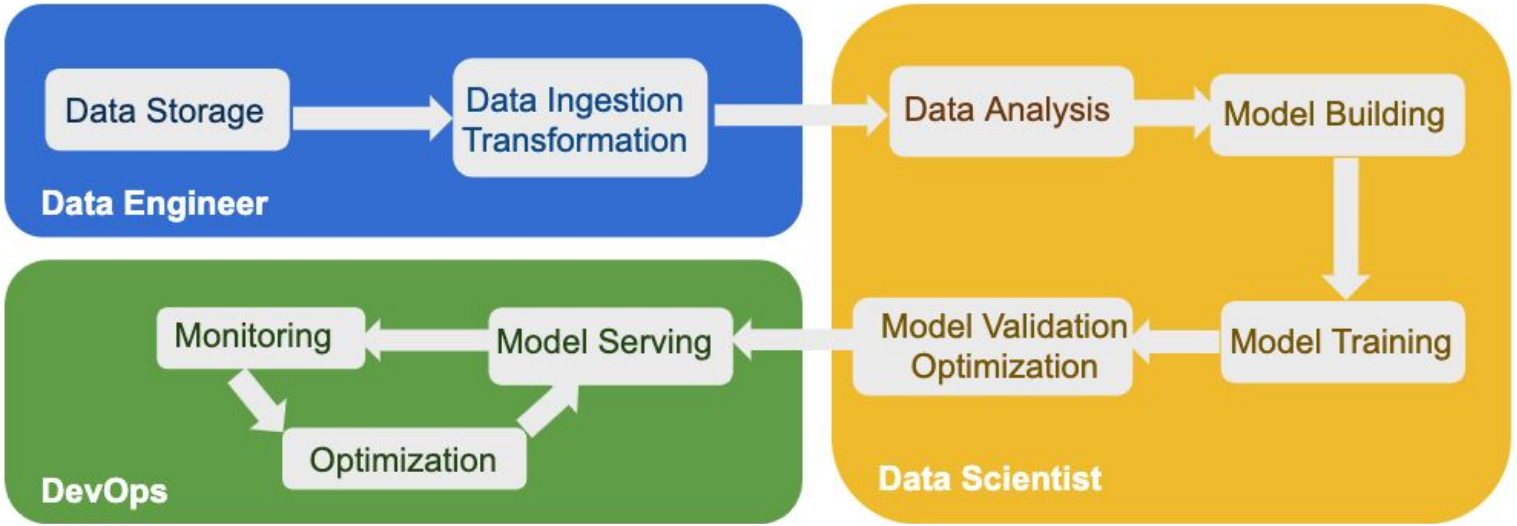
Open Data Hub

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

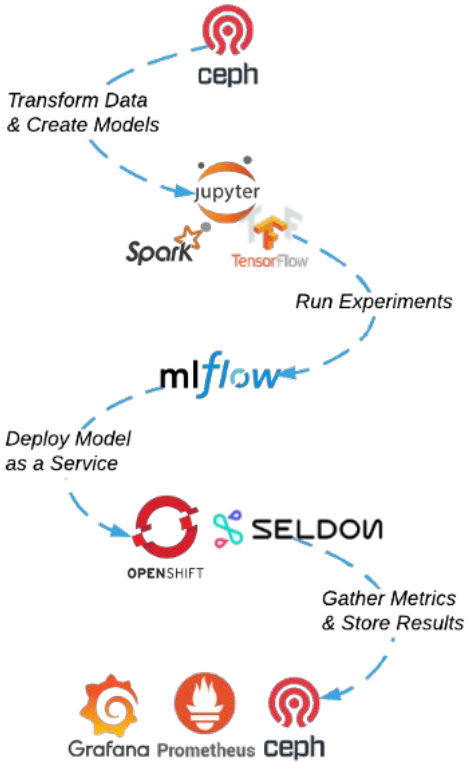
Open Data Hub Project

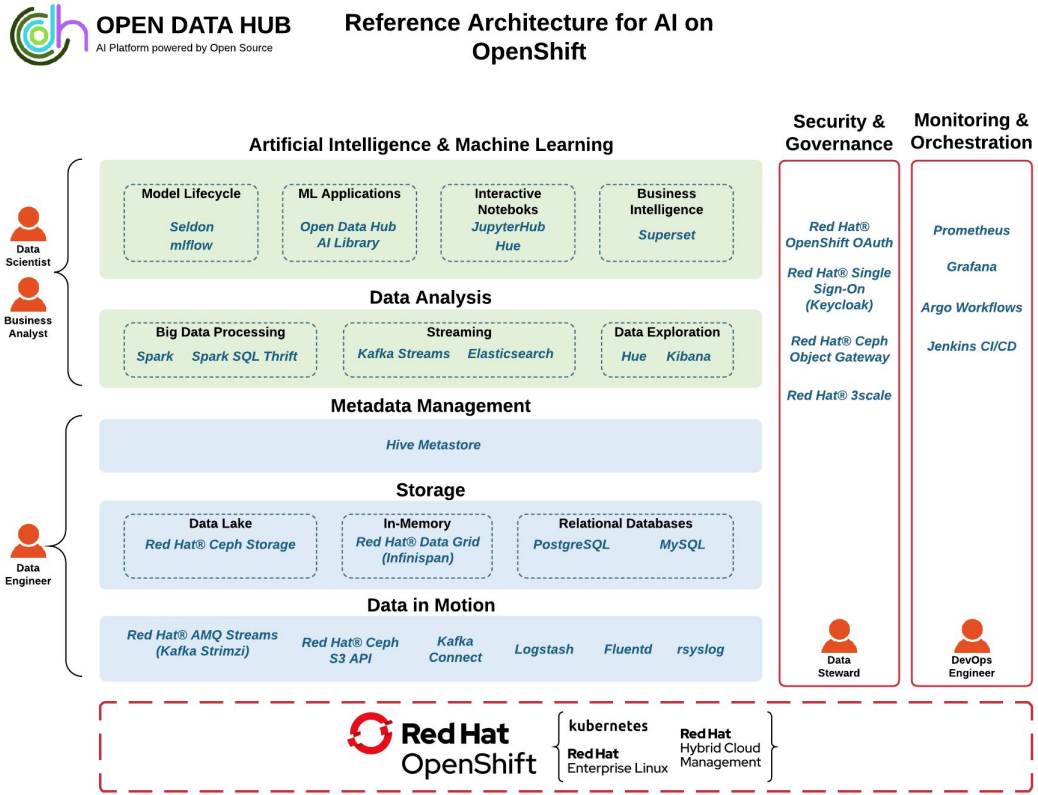
Collaborate on a Data & AI platform for the Hybrid Cloud - https://opendatahub.io/

- Meta-Project to integrate Open Source projects into a practical service oriented solution.
- Red Hat's internal Data Science and AI platform.
- ODH Reference Architecture: https://opendatahub.io/arch.html



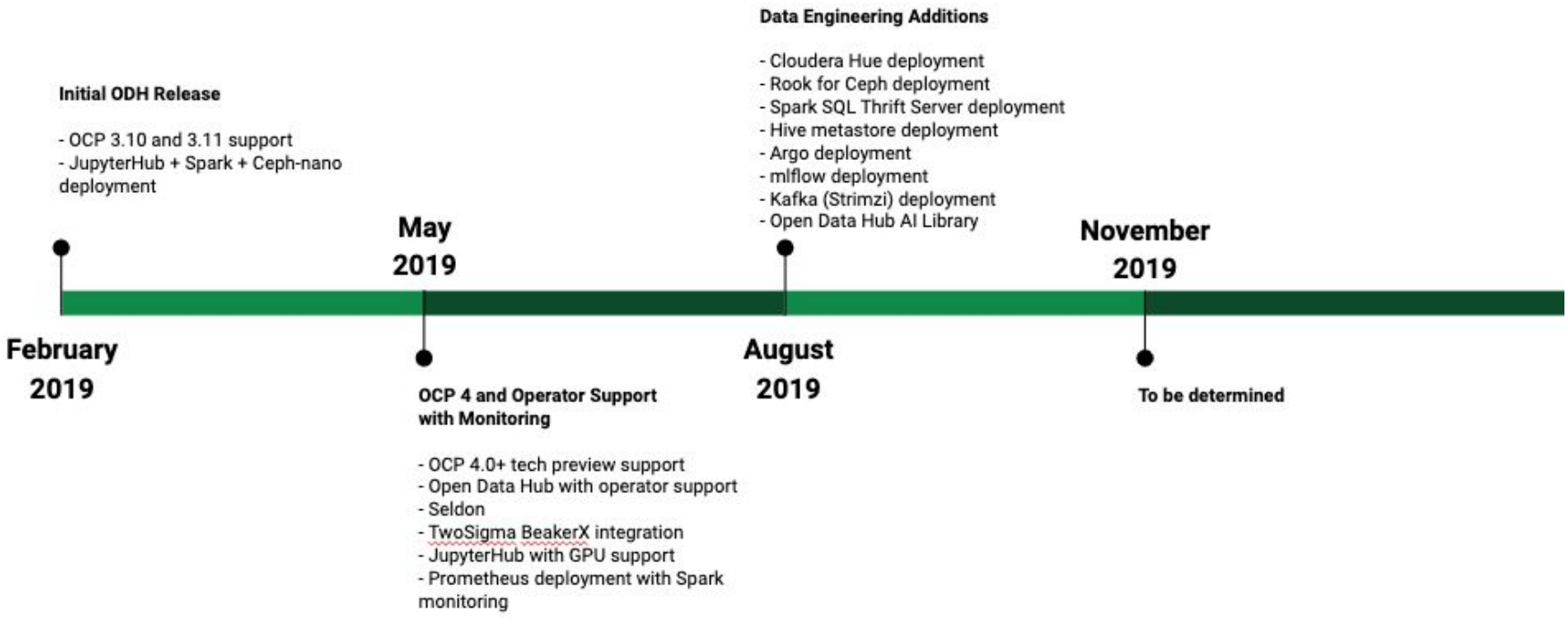
Open Data Hub Architecture





Red Hat

Open Data Hub Roadmap



ODH Core Deployment 0.4

Available Now at OpenDataHub.io

- Monitoring and alerting toolkit
- Records numeric time series data
- Used to diagnose problems

- 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

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

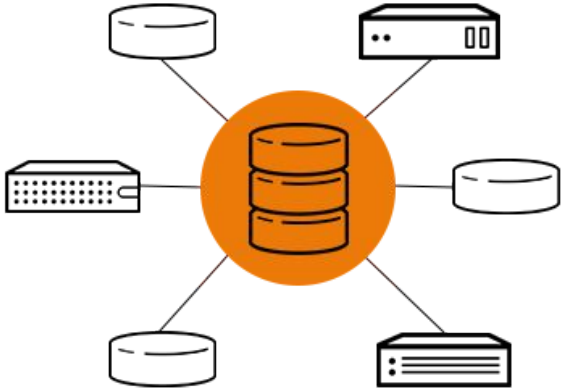
- Multi-user Jupyter
- Used for data science and research

- Distributed Object Store
- S3 Interface

- Distributed event streaming
- Pub/Sub Messaging

Red Hat

- Open Source Distributed Storage for Object, Block and File.
- Data is replicated for fault-tolerance and self healing
- Rook is an open source orchestrator for distributed storage systems running in Openshift/Kubernetes



Massachusetts Open Cloud (MOC)

Led by Boston University, the MOC is a collaborative effort among BU, Harvard, UMass Amherst, MIT, and Northeastern University, as well as the Massachusetts Green High-Performance Computing Center (MGHPCC) and Oak Ridge National Laboratory (ORNL).

It is supported by a broad alliance of industry partners, including Red Hat.

Internal Data Hub

PnT DevOps

Applications in the product release pipeline store their runtime logs in our system. These groups are also engaged for anomaly detection.

Telemeter

Operational metrics from OpenShift clusters. AIOps is engaged here.

PnT Data & Analytics: Grokket

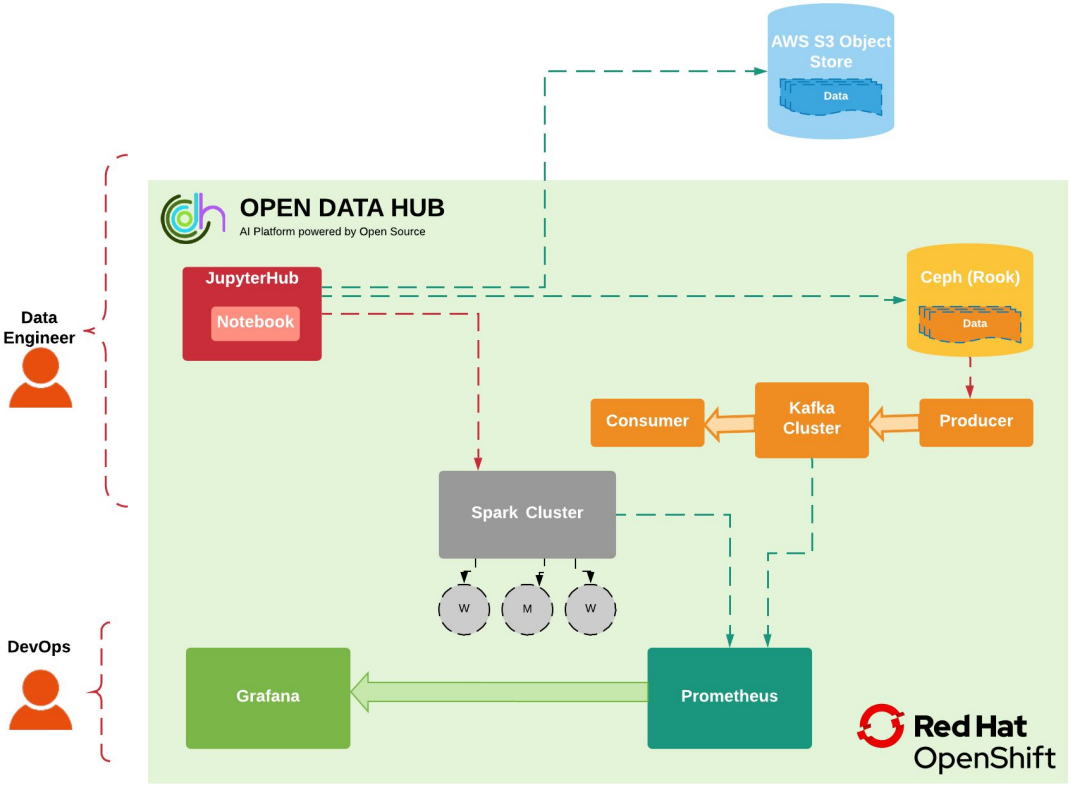
Automated data science on insights data

Customer Insights

Storage of customer data like SOSReports, customer feedback, etc.

Red Hat

Workshop High Level Components



Workshop Workflow

1. How to deploy the Open Data Hub on an OpenShift cluster
2. How to perform some data wrangling using Spark and Jupyter Notebooks in a hybrid cloud environment
3. How to send data with Kafka, and monitor the health of the ODH environment with Prometheus and Grafana

Red Hat

Workshop Guide

Workshop Guide: https://bit.ly/33Ckx3U

OpenShift Cluster: https://bit.ly/2OVuHcu

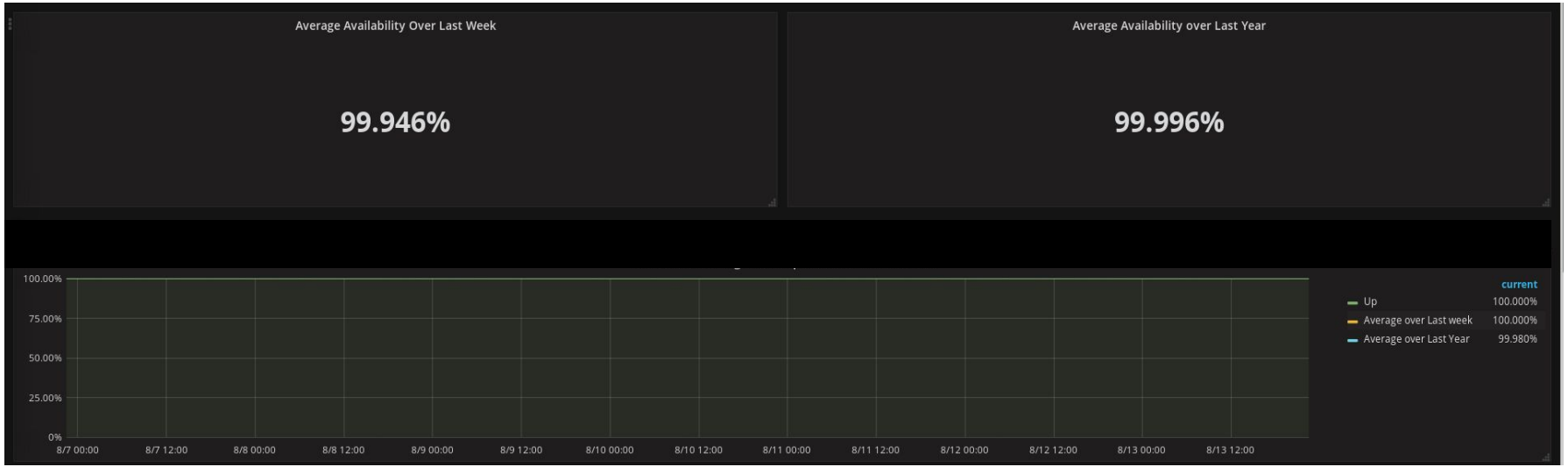
Username: user<number>

Password: r3dh4t1!

Red Hat

Internal Monitoring

Internal Monitoring



Red Hat

Resources

Open Data Hub Community: https://opendatahub.io/

ODH Operator: https://gitlab.com/opendatahub/opendatahub-operator

Red Hat

Thank You

linkedin.com/company/red-hat

youtube.com/user/RedHatVideos

facebook.com/redhatinc

twitter.com/RedHat

Red Hat