OPEN DATA HUB

AI Platform powered by Open Source

AI & ML on OpenShift with Open Data Hub

Audrey Reznik
Data Scientist, Red Hat OpenShift Data Science team

Artificial Intelligence and Machine learning on **OpenShift** with

Open Data Hub

Open Data Hub is how **Red Hat** does Artificial Intelligence and Machine learning internally on OpenShift

Red Hat

Today **Open Data Hub** internally runs AI/ML workloads

Application Logs

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

Cluster Metrics

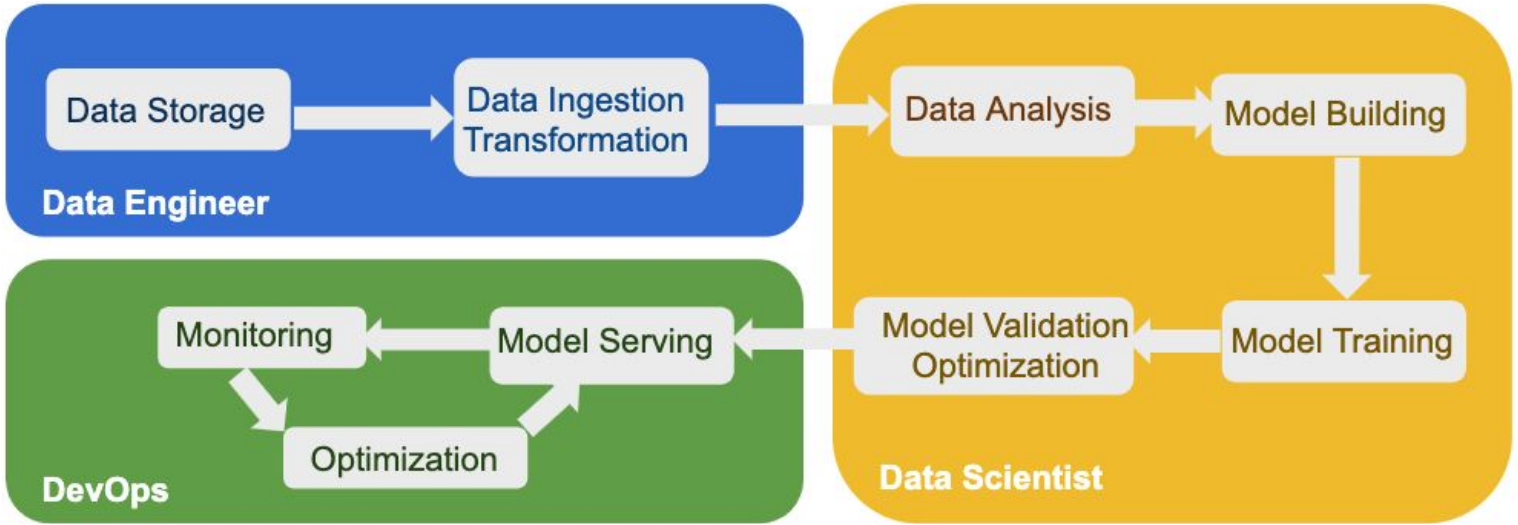
Operational metrics from OpenShift clusters. AIOps is engaged here.

Customer Support Data

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

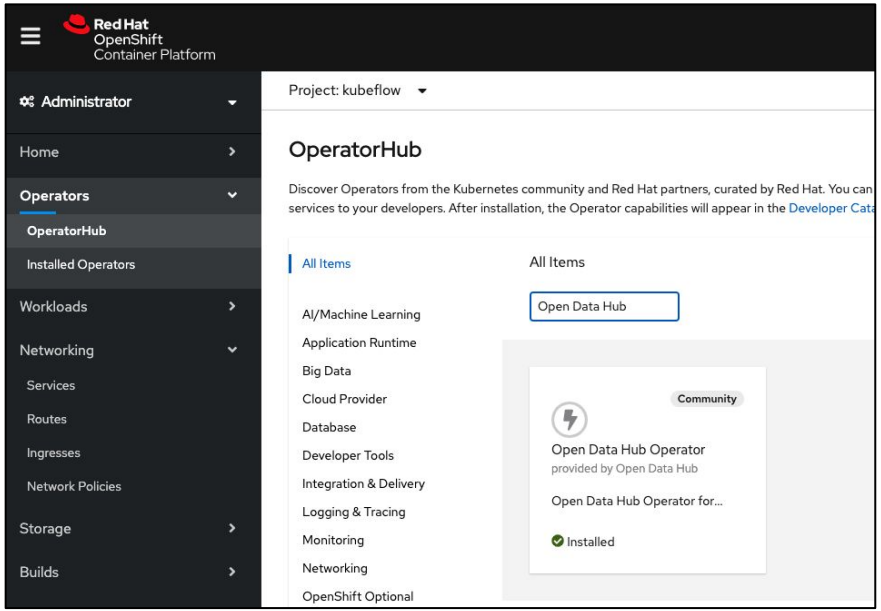
What is Open Data Hub

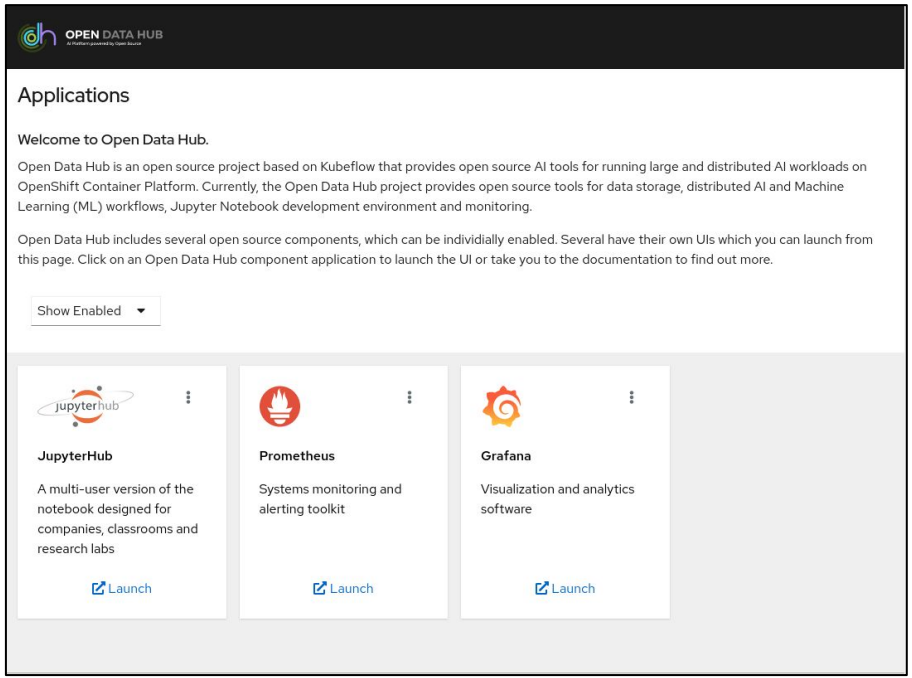
Open Data Hub is an open source community AI/ML platform on Openshift



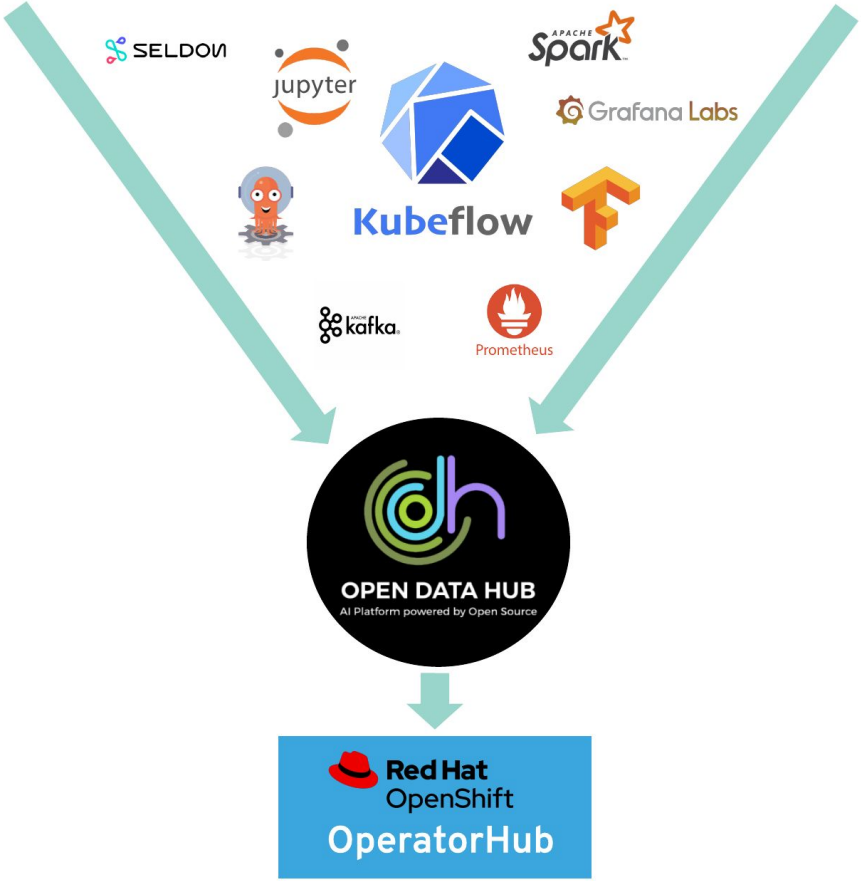
Red Hat

Open Data Hub is an **Operator** installed from OpenShift **OperatorHub**





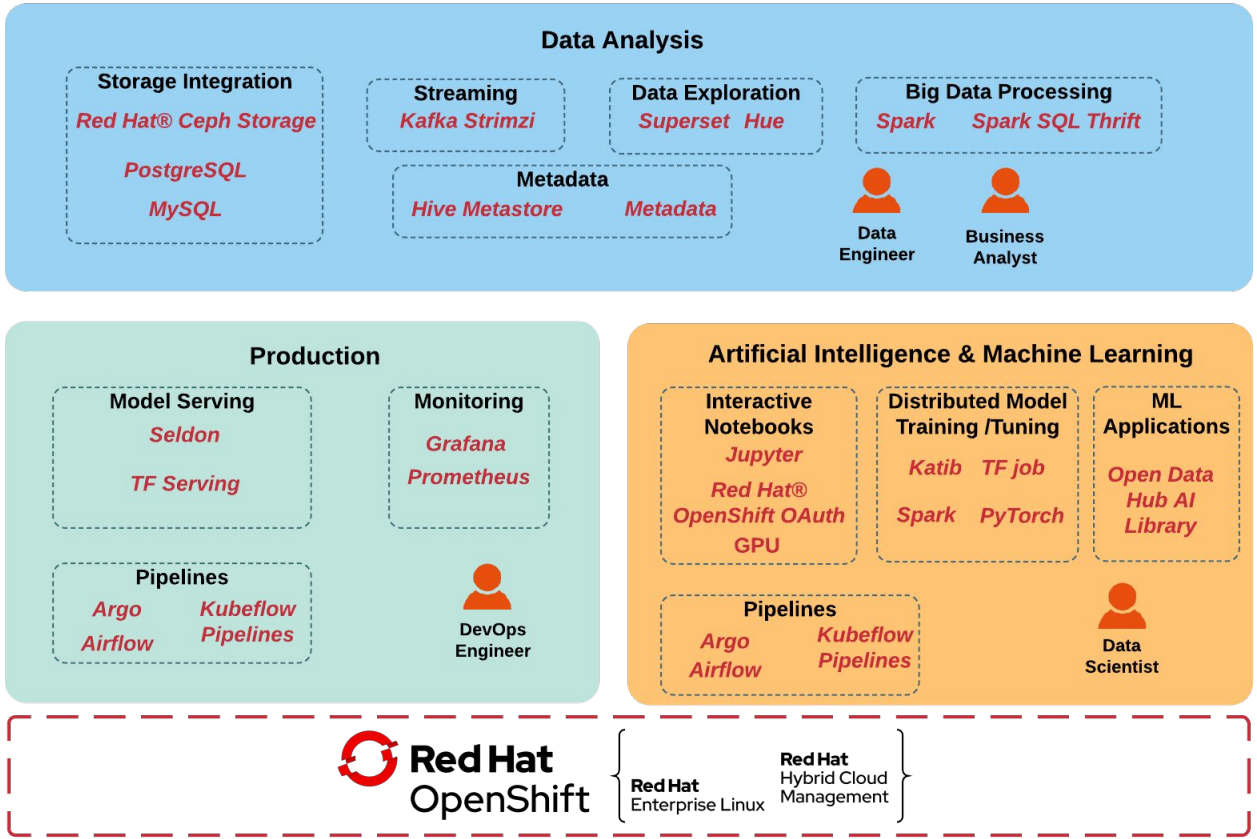
Open Data Hub integrates open source projects into an **end-to-end AI/ML platform** o OpenShift



Red Hat

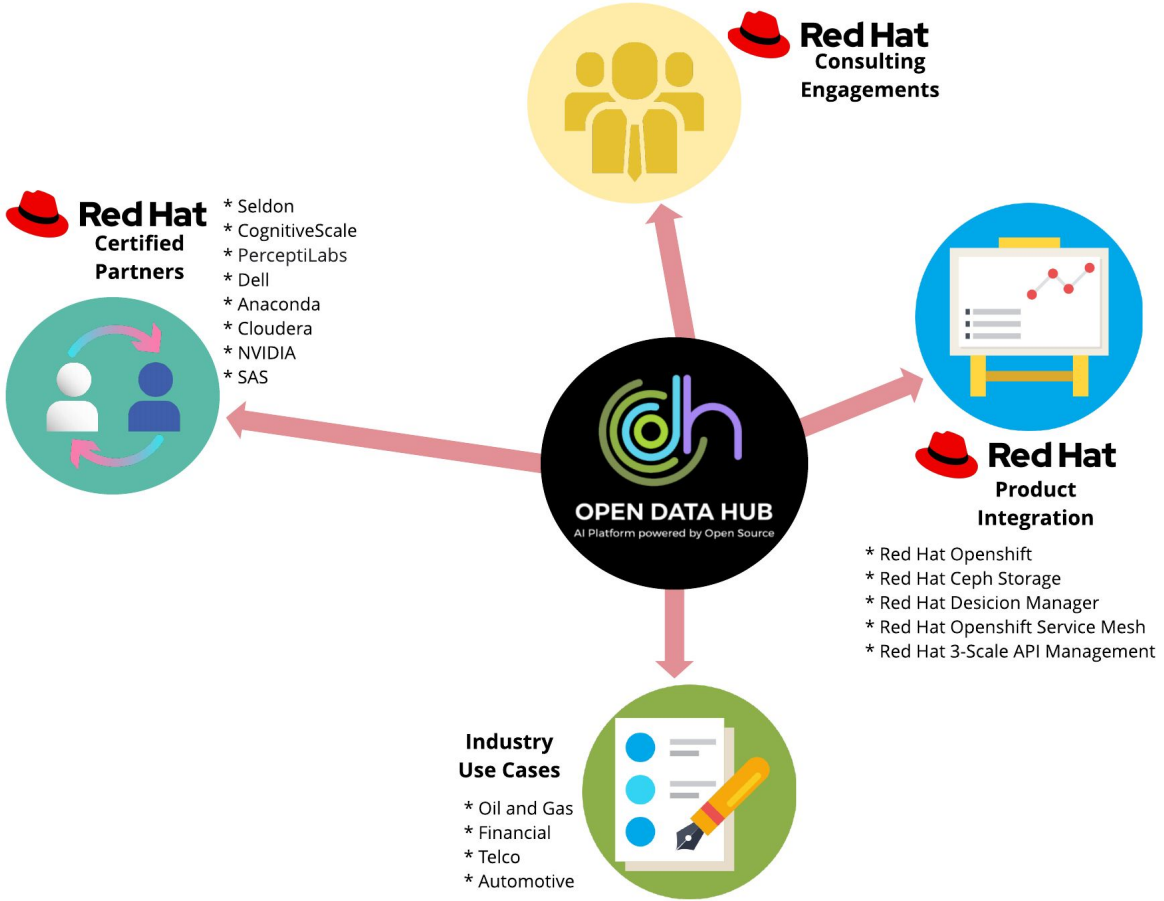
OPEN DATA HUB

AI Platform powered by Open Source



Red Hat

Open Data Hub comes with an ecosystem provided by **Red Hat** and **Certified Partners**



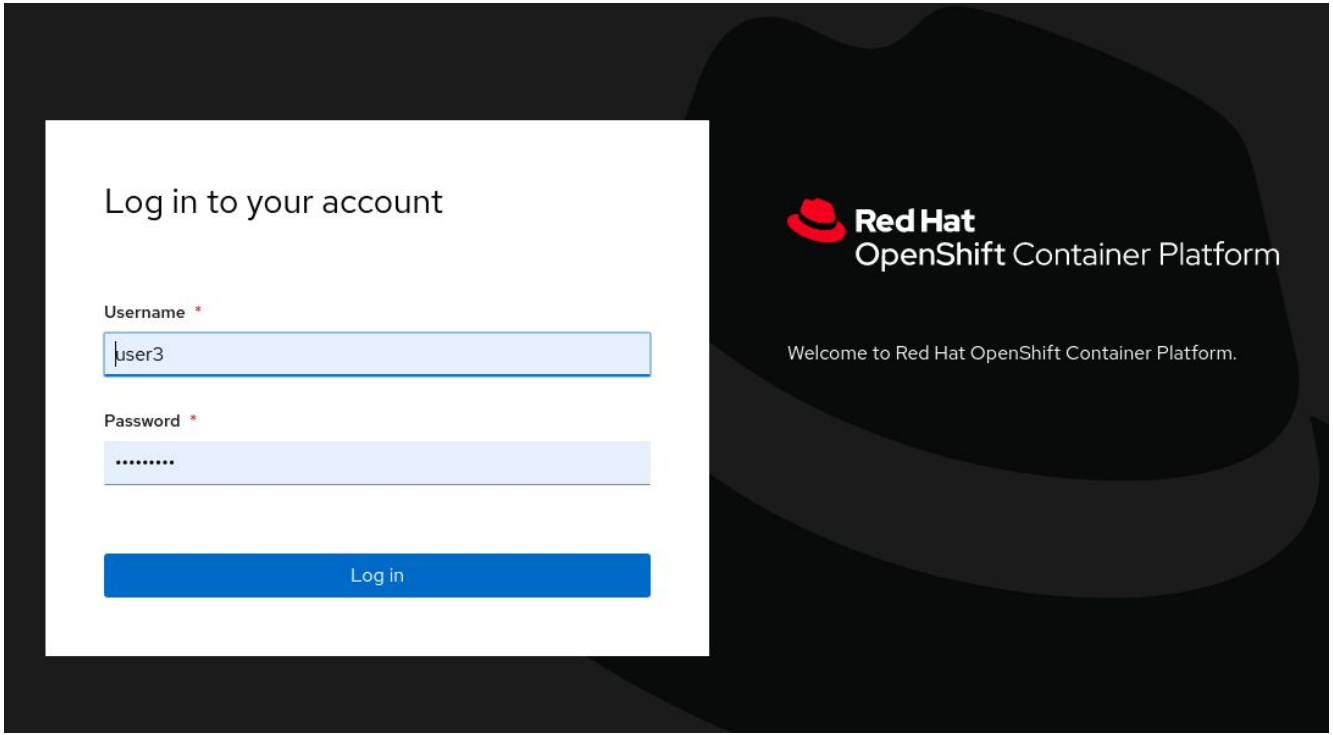
Red Hat

OPEN DATA HUB

AI Platform powered by Open Source

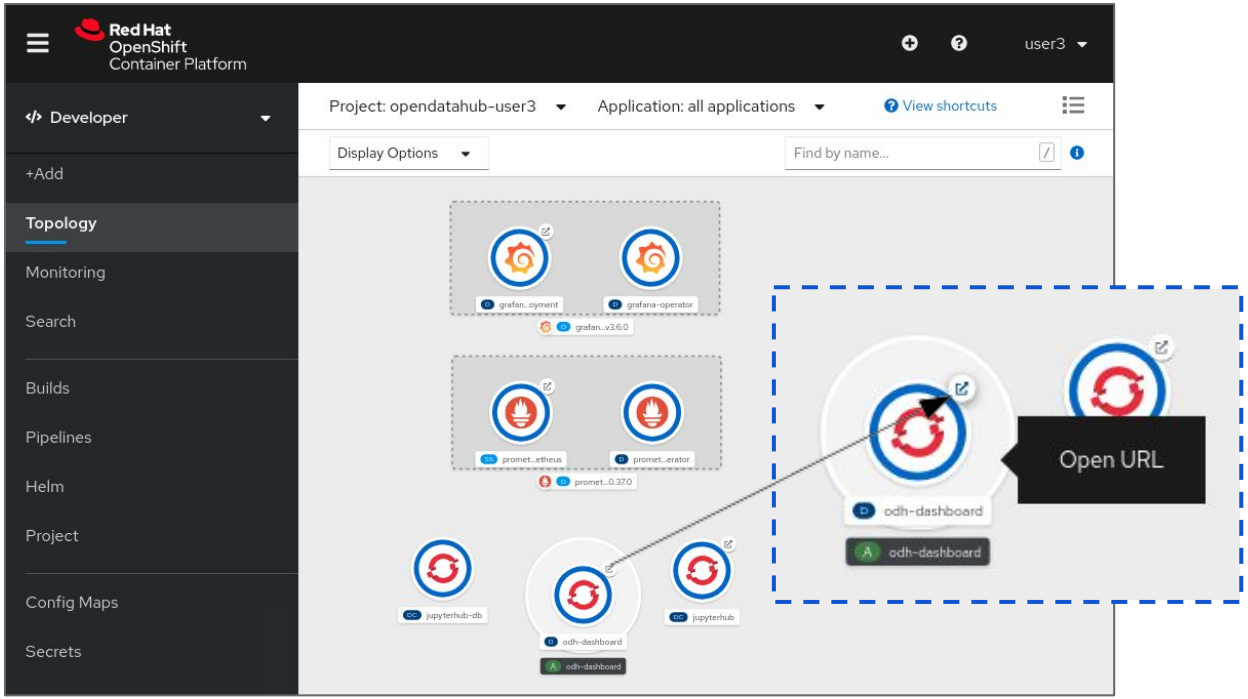
Demo

Log into your OpenShift account



Red Hat

Proceed to the Open Data Hub Dashboard



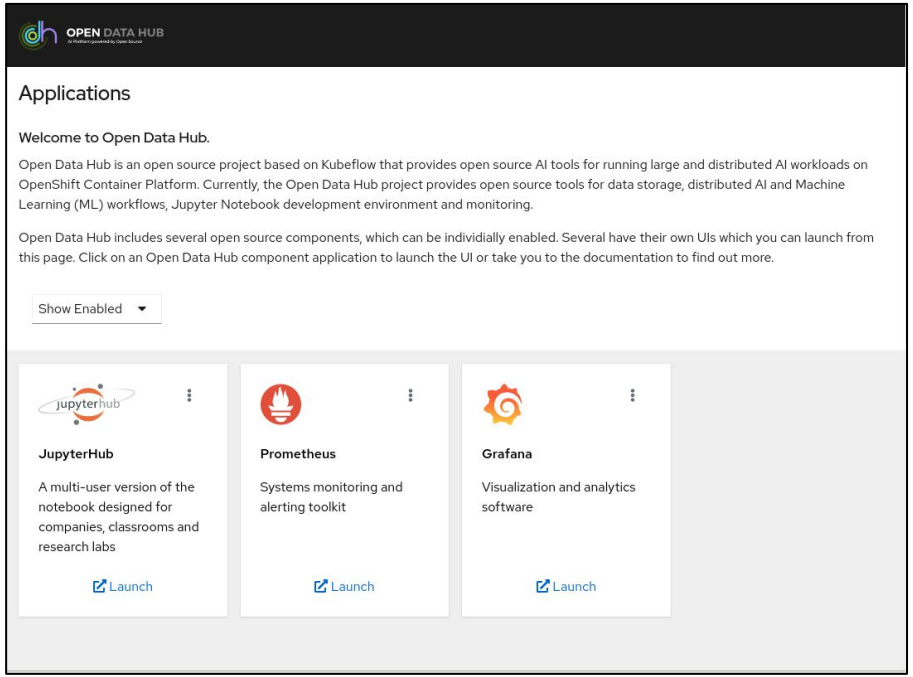
We are logged in as a developer

To navigate use the left panel navigation bar

Proceed to the Open Data Hub dashboard by clicking on the odh-dashboard operator “Open URL” button

Red Hat

Using ODH



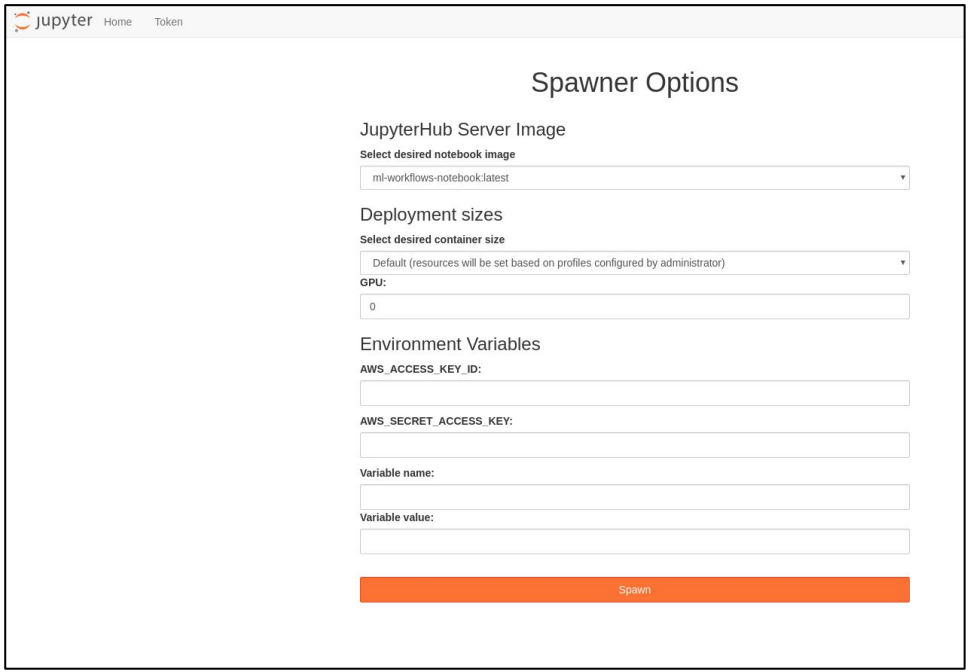
ODH contains a number of tools that you can use to build, manage and deploy your models.

Let’s take on the Role of a DS and work on a fraud detection model.

Click on the “JupyterHub” card to open JuypterHub and being programming!

Red Hat

Jupyter Hub Spawner Options



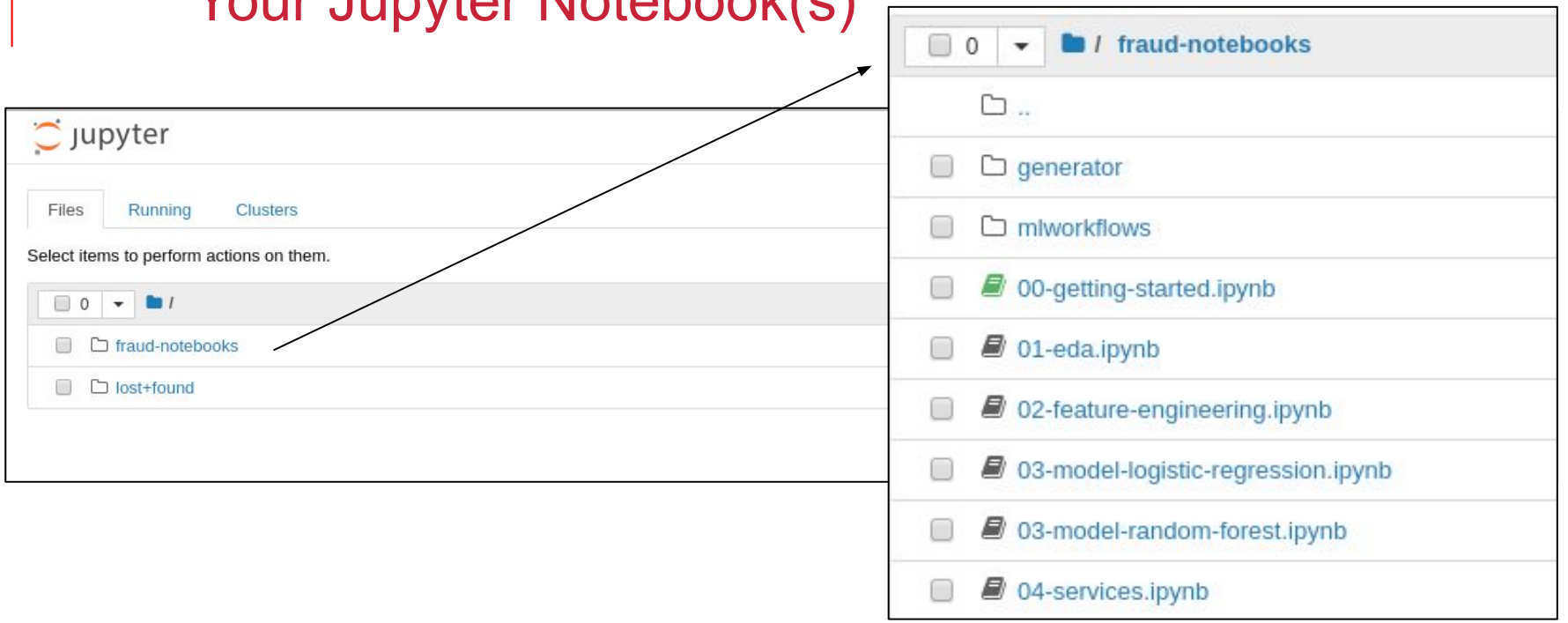
We are spawning a Jupyter Notebook image that we can use to deploy a **fraud detection** model.

Looking at legitimate & fraudulent transactions

Accept the defaults and choose “Spawn” to continue.

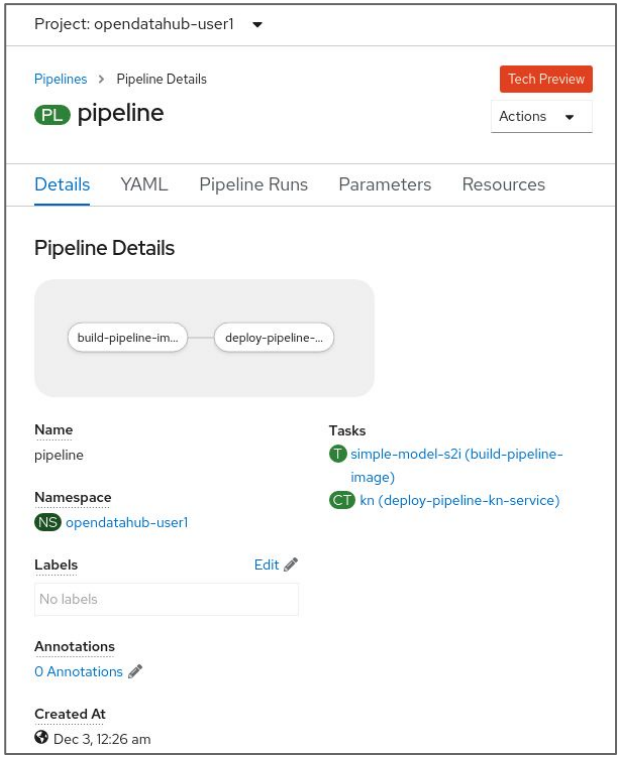
Red Hat

Your Jupyter Notebook(s)



Red Hat

Putting a Model into production using pipelines

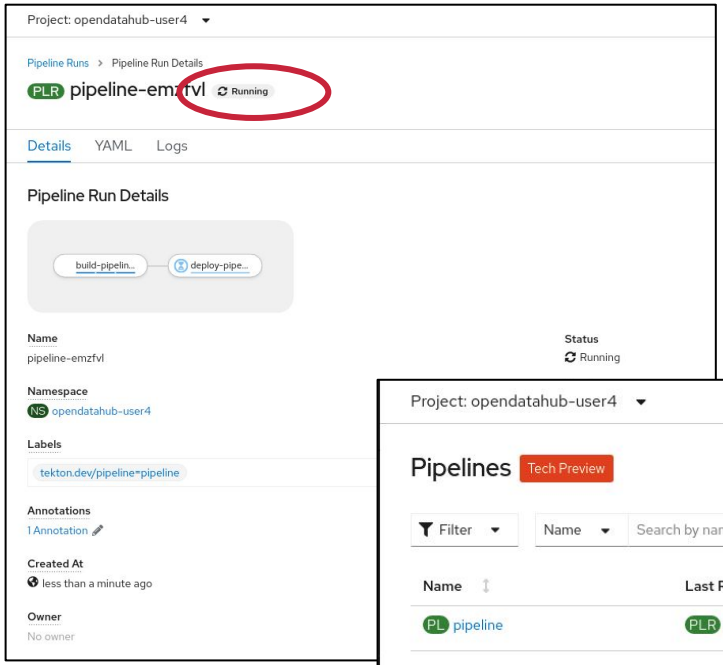


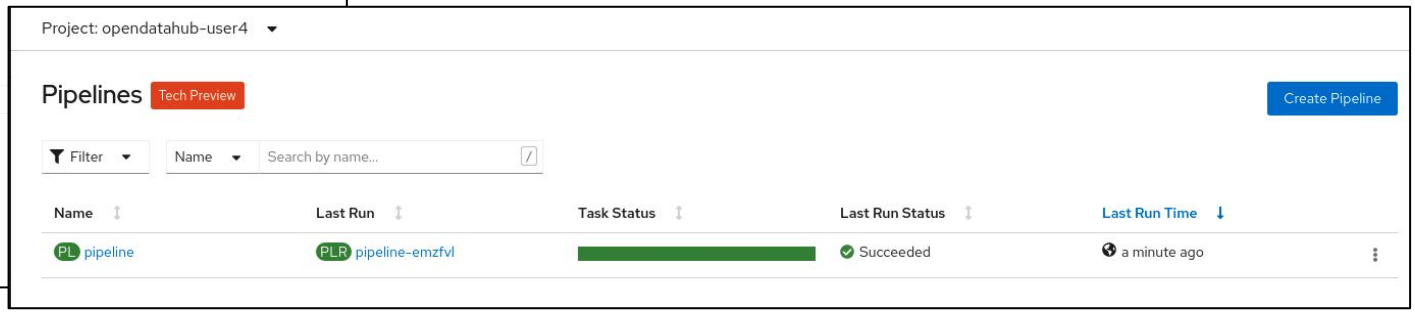
- Deploy machine learning pipelines into production with OpenShift pipelines and see how we can use these services to make predictions

- Go back to the main openshift console, select “pipelines” and you will see that there is this pipeline already built for you.
  - Click on the pipeline and you will see the pipeline details:

Red Hat

Running the OCP Pipeline





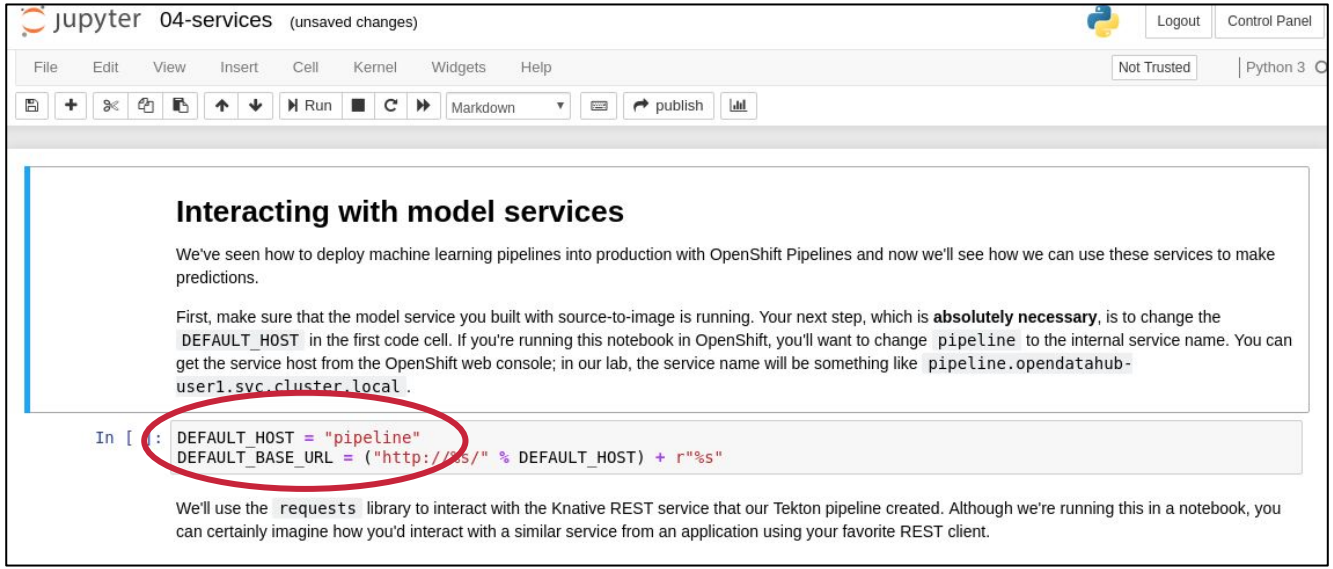
Once the pipeline is finished we have a model service (rest service) which is built with source to image (s2i).

At this point we want to obtain the pipeline service url. We'll be using a request library in python to interact with our rest service.

Service url: pipeline-opendatahub-user1.apps.cluster-3cc9.3cc9.example.opentlc.com

Red Hat

Interacting with Model Services



Once the service is running, we will copy the generated url and paste it into the 04-services.ipynb notebook

Red Hat

Model making its predictions

```
In [4]: data = pd.read_parquet("fraud-cleaned-sample.parquet")

In [5]: sample = data.sample(200)
        sample["predictions"] = score_transactions(sample)
        sample
```

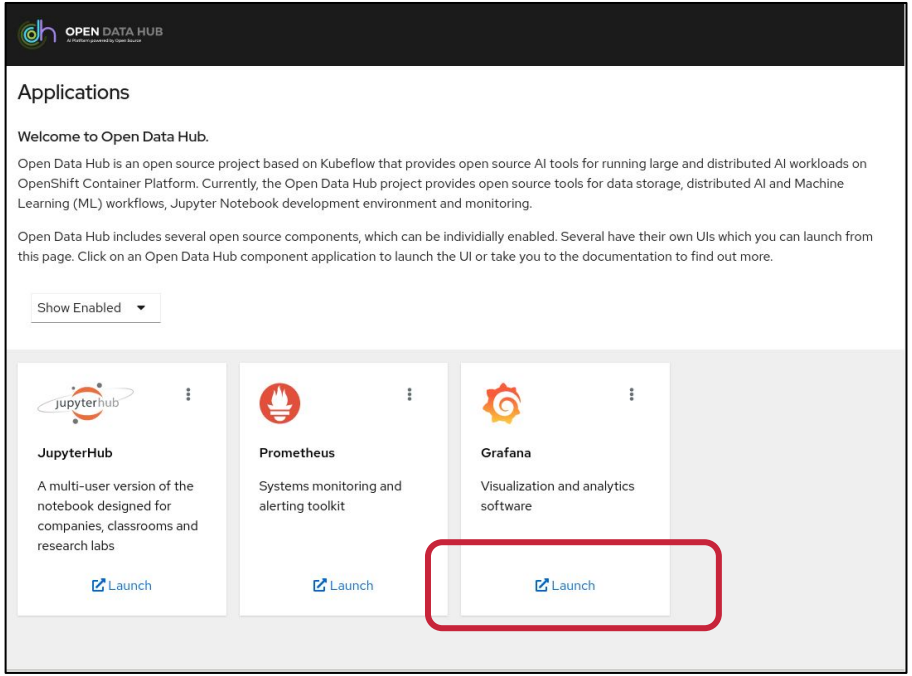
Out[5]:

	timestamp	label	user_id	amount	merchant_id	trans_type	foreign	interarrival	predictions
45575650	1648146992	legitimate	9101	20.020000	654	online	True	13023.0	legitimate
28870926	1631713489	legitimate	5768	756.200012	13321	online	True	5796.0	legitimate
28717980	1584909456	legitimate	5738	18.400000	12721	online	True	8271.0	legitimate
879985	1606740723	legitimate	175	86.769997	14941	chip_and_pin	False	13527.0	legitimate
27853238	1597581774	legitimate	5564	3.190000	11473	online	True	39813.0	legitimate
...	...	...	...	...	...	...	...	...	...
7130223	1619350159	legitimate	1427	19.700001	6696	contactless	False	7404.0	legitimate
47590587	1645056492	legitimate	9517	25.940001	770	contactless	False	7201.0	legitimate
39919463	1596042240	legitimate	7972	27.010000	6200	online	False	7606.0	legitimate
45432525	1633592676	legitimate	9077	11.540000	12201	contactless	False	49416.0	legitimate
9590754	1652401133	legitimate	1918	370.140015	18095	online	False	5897.0	legitimate

200 rows × 9 columns

Red Hat

Go back to ODH & launch Grafana



Red Hat

Examining Fraudulent & Legitimate transactions in a Grafana dashboard

And now we have the ability to “visually” monitor our service for fraudulent & legitimate transactions!

Red Hat

OPEN DATA HUB

AI Platform powered by Open Source

AI & ML on OpenShift with Open Data Hub

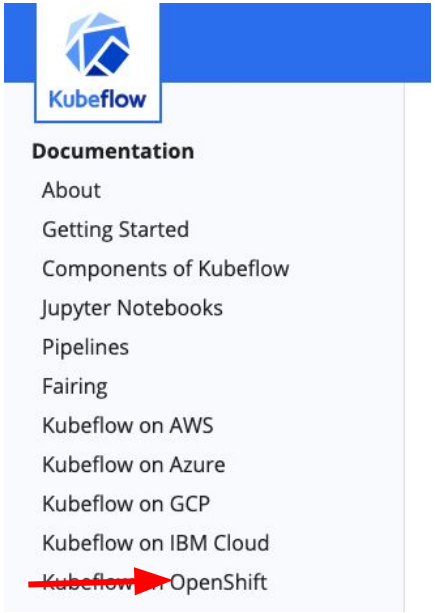
For more information on Open Data Hub:

1. http://opendatahub.io
2. http://Github.com/opendatahub-io

*Huge Thank You to Juana Nakfour for letting me use her ODH slide pack for this presentation.

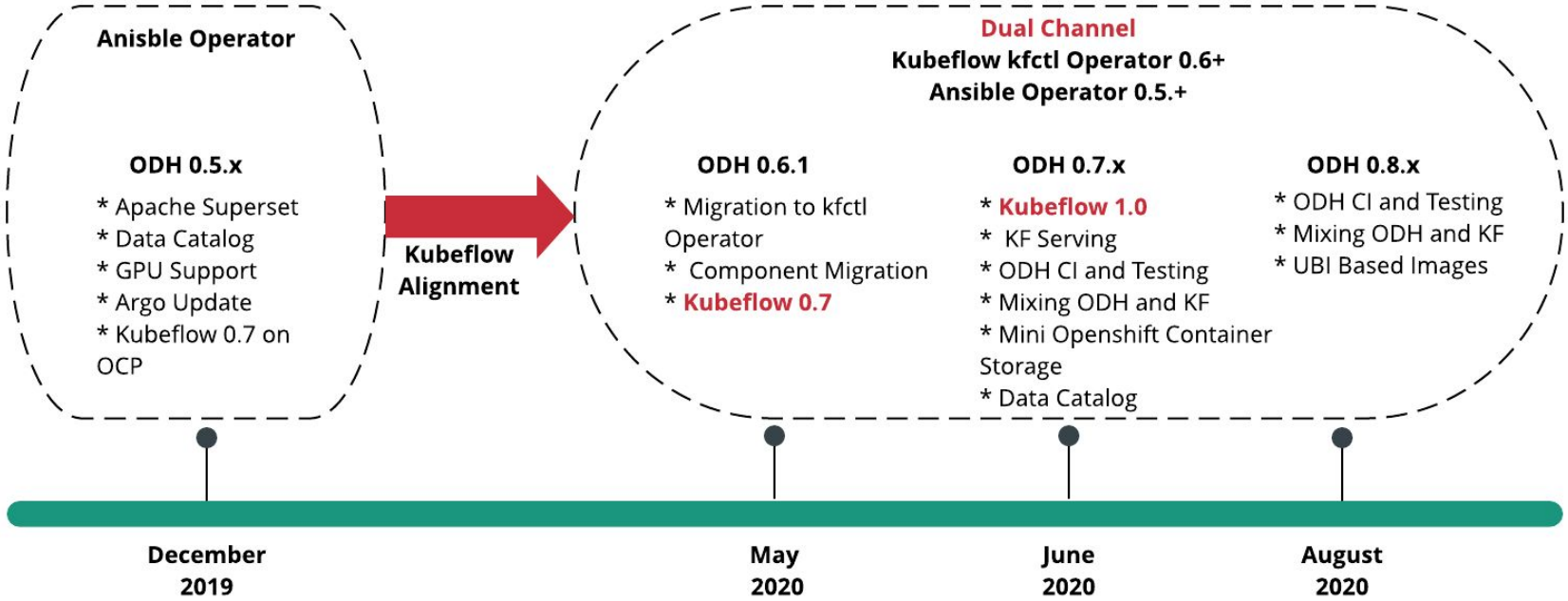
OPEN DATA HUB - Kubeflow

- Kubeflow - A new open source project dedicated to making deployments of machine learning (ML) workflows on Kubernetes **simple, portable and scalable - kubeflow.org**
- It is now integrated into Open Data Hub and runs on **OpenShift**
- Kubeflow brings multiple new AI/ML capabilities and features
  - For **model training** Tensorflow and Pytorch
  - For **model serving** Seldon and KFServing
  - For **pipelines** Kubeflow Pipelines based on Argo
- Contribute **upstream** to Kubeflow, add Openshift as a platform.
- **UBI** as a base image and partner certification
- **One click install** from **Openshift OperatorHub**



Red Hat

OPEN DATA HUB - Current Releases



Red Hat