



# 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

# 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.

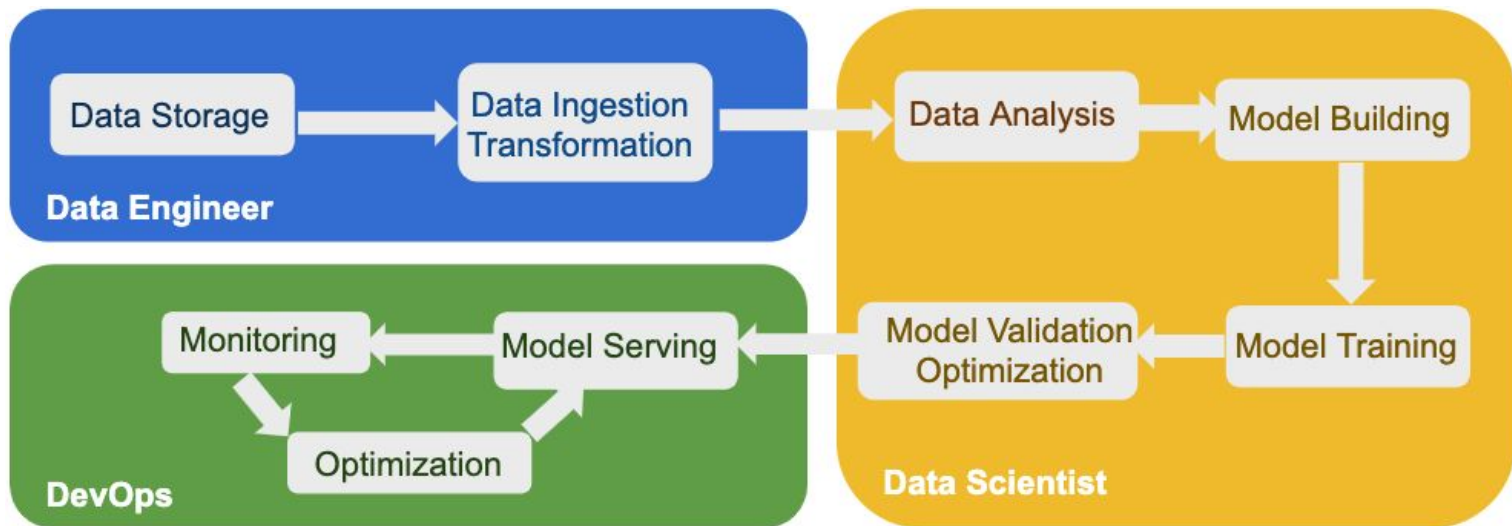


## **Customer Support Data**

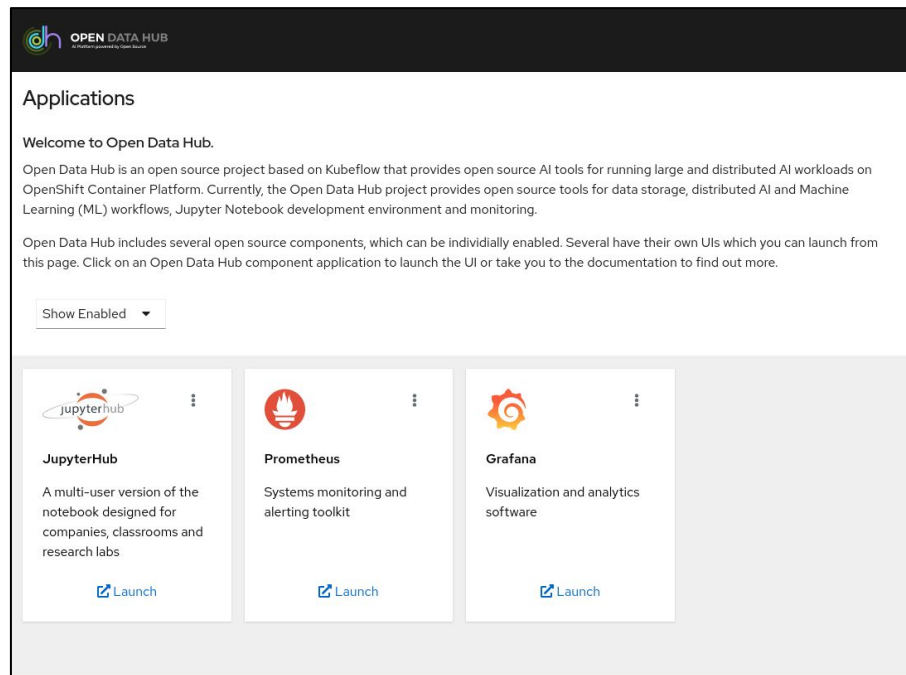
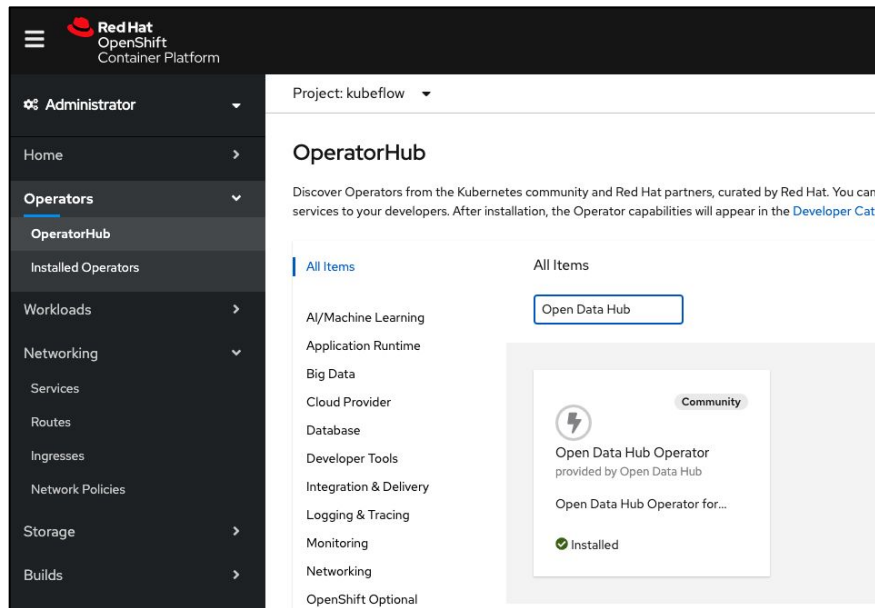
Storage of customer data like **SOSReports, customer feedback**, etc.

# What is Open Data Hub

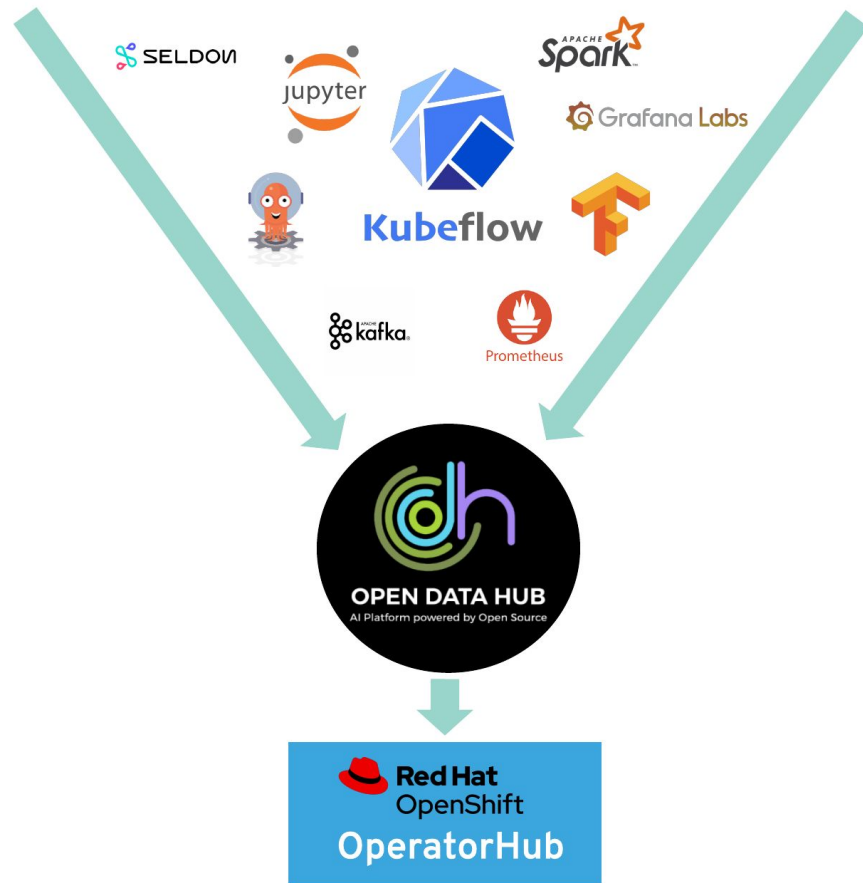
**Open Data Hub** is an open source community AI/ML platform on Openshift



# 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 on  
OpenShift

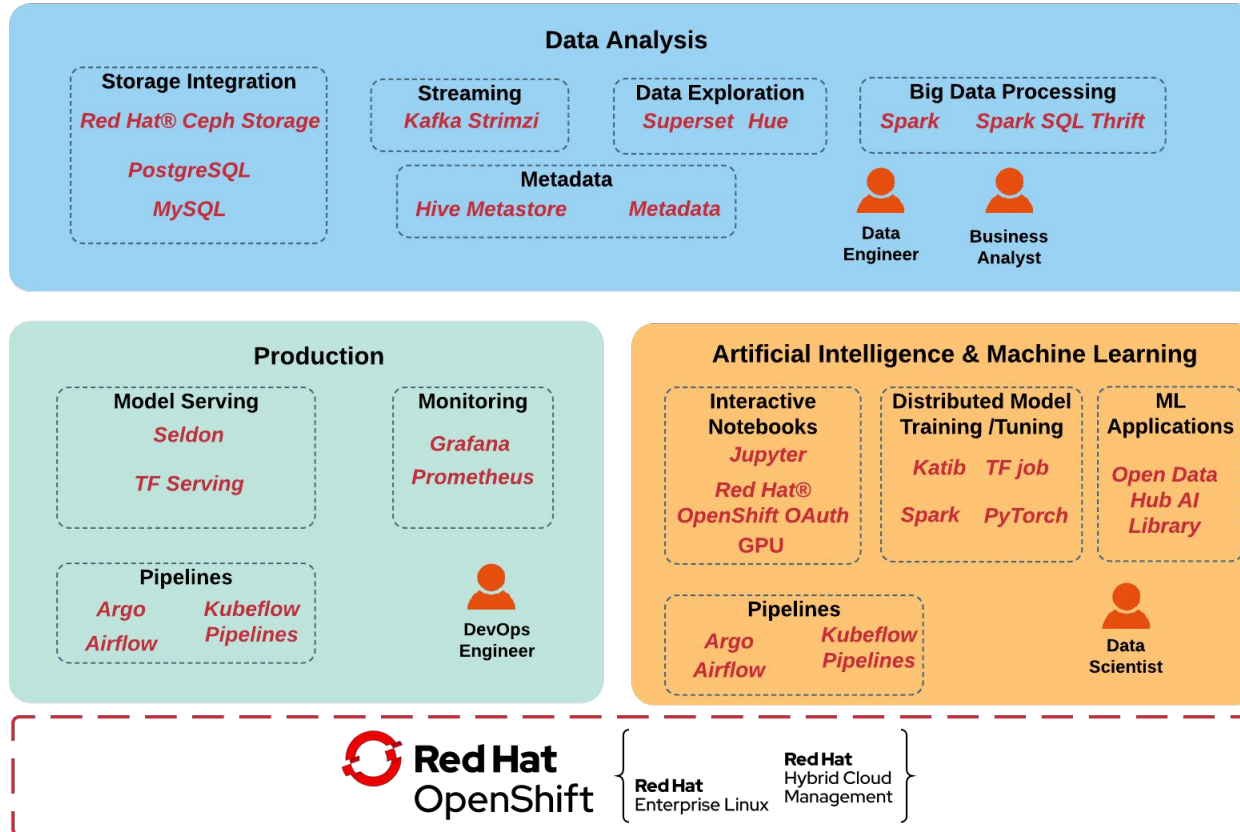




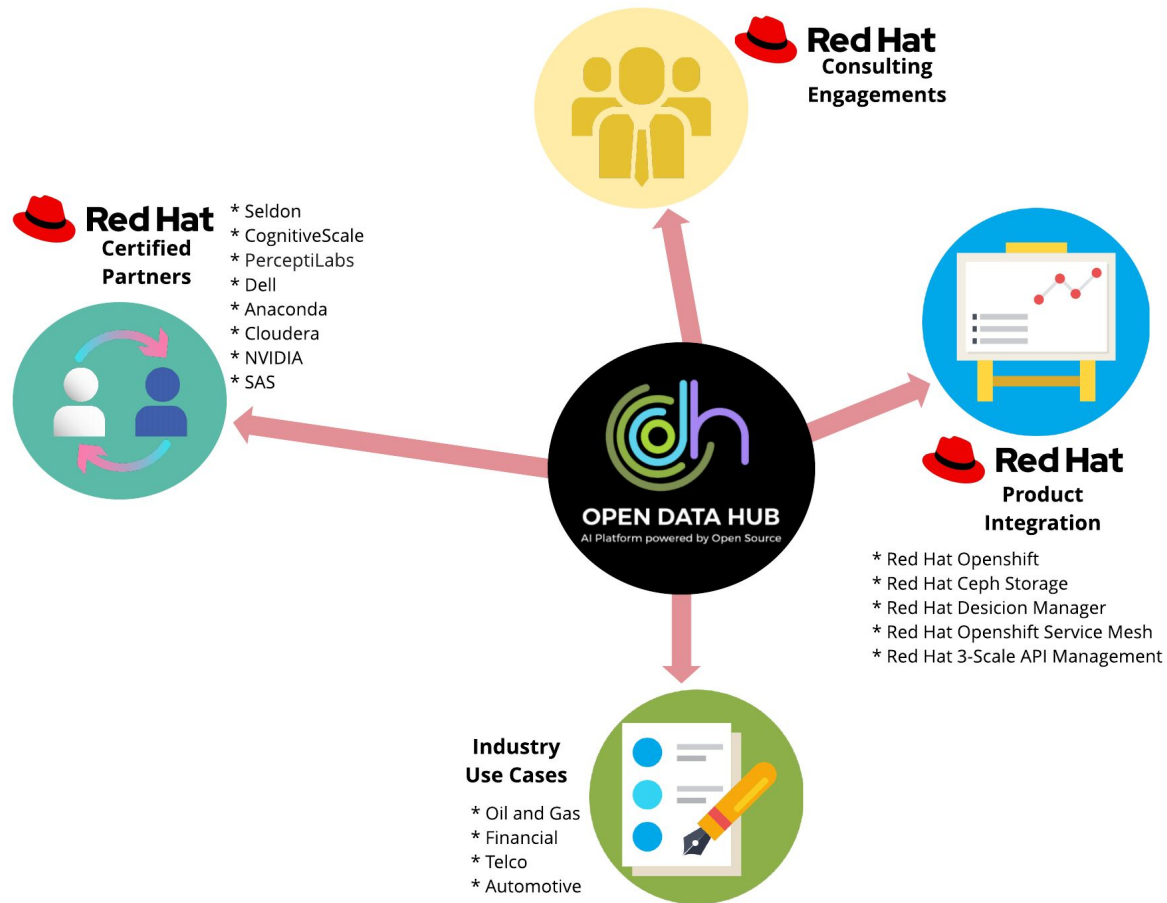


## OPEN DATA HUB

AI Platform powered by Open Source



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





# OPEN DATA HUB

AI Platform powered by Open Source

## Demo


# Log into your OpenShift account

Log in to your account

Username \*

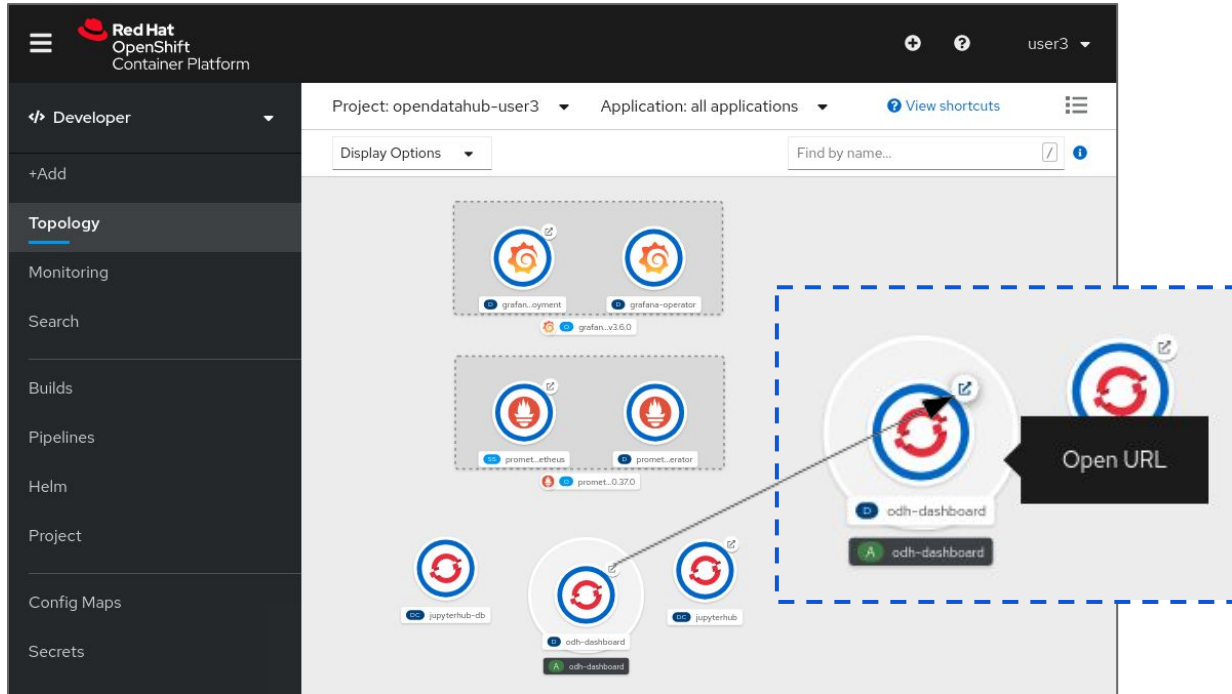
Password \*

Log in

**Red Hat**  
OpenShift Container Platform

Welcome to Red Hat OpenShift Container Platform.

# Proceed to the Open Data Hub Dashboard

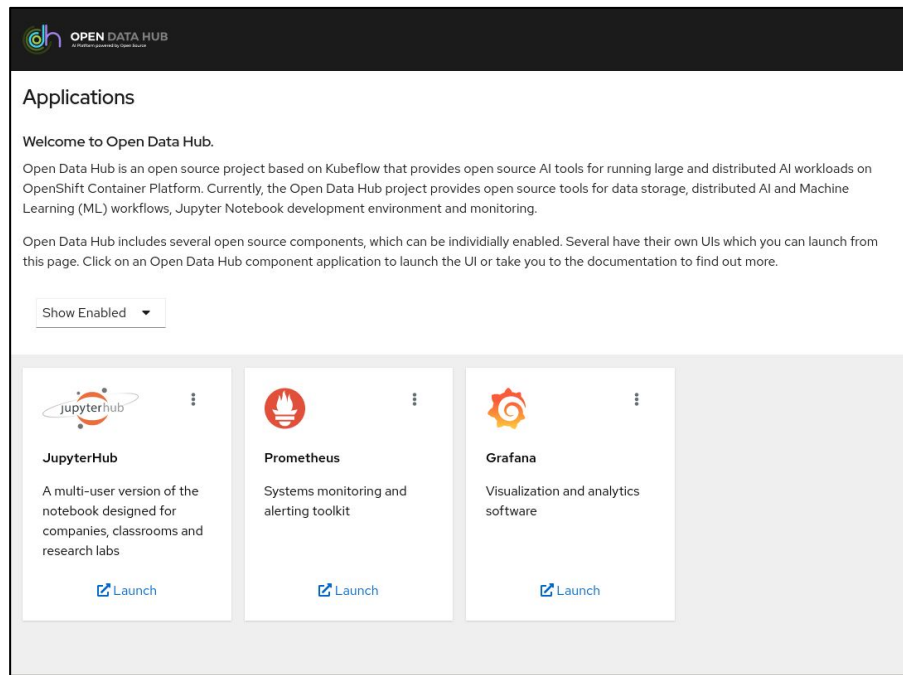


We are logged in as a developer

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

# Using ODH

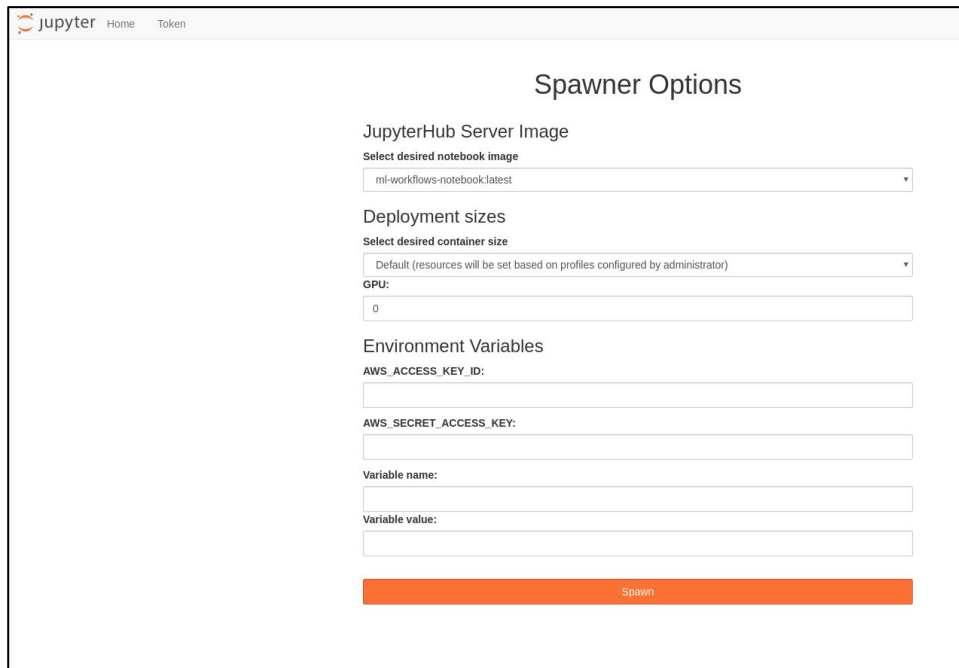


ODH contains a number of tools that you can use to build, manage and deploy your models.

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

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

# Jupyter Hub Spawner Options



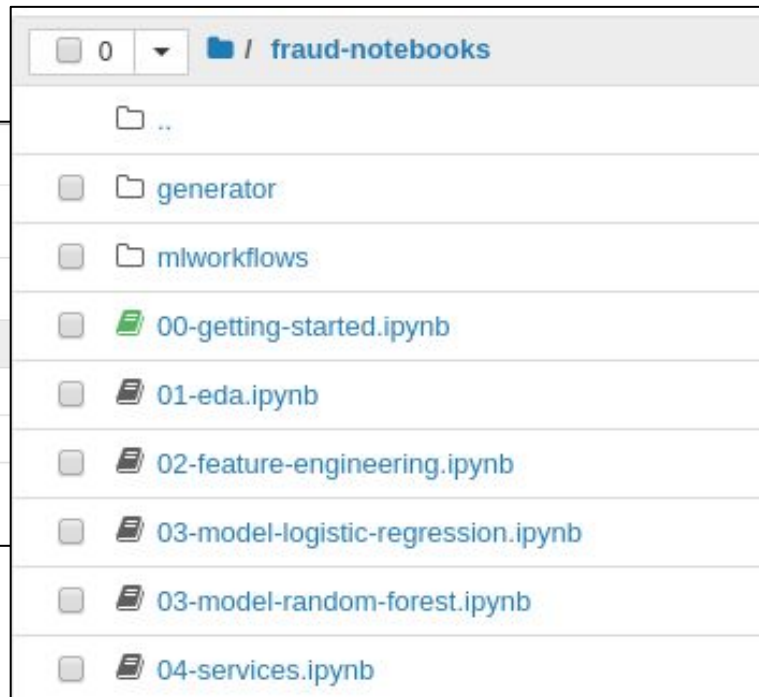
The screenshot shows the Jupyter Hub web interface. At the top, there's a navigation bar with the Jupyter logo, 'Home', and 'Token'. The main heading is 'Spawner Options'. Below this, there are several sections: 'JupyterHub Server Image' with a dropdown menu showing 'ml-workflows-notebook:latest'; 'Deployment sizes' with a dropdown menu showing 'Default (resources will be set based on profiles configured by administrator)'; 'GPU:' with a text input field containing '0'; 'Environment Variables' with fields for 'AWS\_ACCESS\_KEY\_ID:', 'AWS\_SECRET\_ACCESS\_KEY:', 'Variable name:', and 'Variable value:'. At the bottom, there is an orange 'Spawn' button.

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.

# Your Jupyter Notebook(s)





# Putting a Model into production using pipelines

The screenshot shows the OpenShift Pipelines console interface. At the top, it indicates the project is 'opendatahub-user1'. Below this, there's a breadcrumb 'Pipelines > Pipeline Details' and a 'Tech Preview' badge. The main title is 'pipeline' with a 'PL' icon and an 'Actions' dropdown menu. A tab bar at the top of the main content area includes 'Details' (selected), 'YAML', 'Pipeline Runs', 'Parameters', and 'Resources'. The 'Pipeline Details' section shows a visual representation of the pipeline with two tasks: 'build-pipeline-im...' and 'deploy-pipeline-...'. Below this, the 'Name' is 'pipeline', the 'Namespace' is 'opendatahub-user1', and there are no labels. The 'Tasks' section lists two tasks: 'simple-model-s2i (build-pipeline-image)' and 'kn (deploy-pipeline-kn-service)'. At the bottom, it shows '0 Annotations' and 'Created At' as 'Dec 3, 12:26 am'.

- 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:

# Running the OCP Pipeline

Project: opendatahub-user4

Pipeline Runs > Pipeline Run Details

**PLR pipeline-emzfvl** Running

Details | YAML | Logs

Pipeline Run Details

build-pipeline... deploy-pipe...

Name  
pipeline-emzfvl

Status  
Running

Namespace  
NS opendatahub-user4

Labels  
tekton.dev/pipeline=pipeline

Annotations  
1 Annotation

Created At  
less than a minute ago

Owner  
No owner

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.

Project: opendatahub-user4

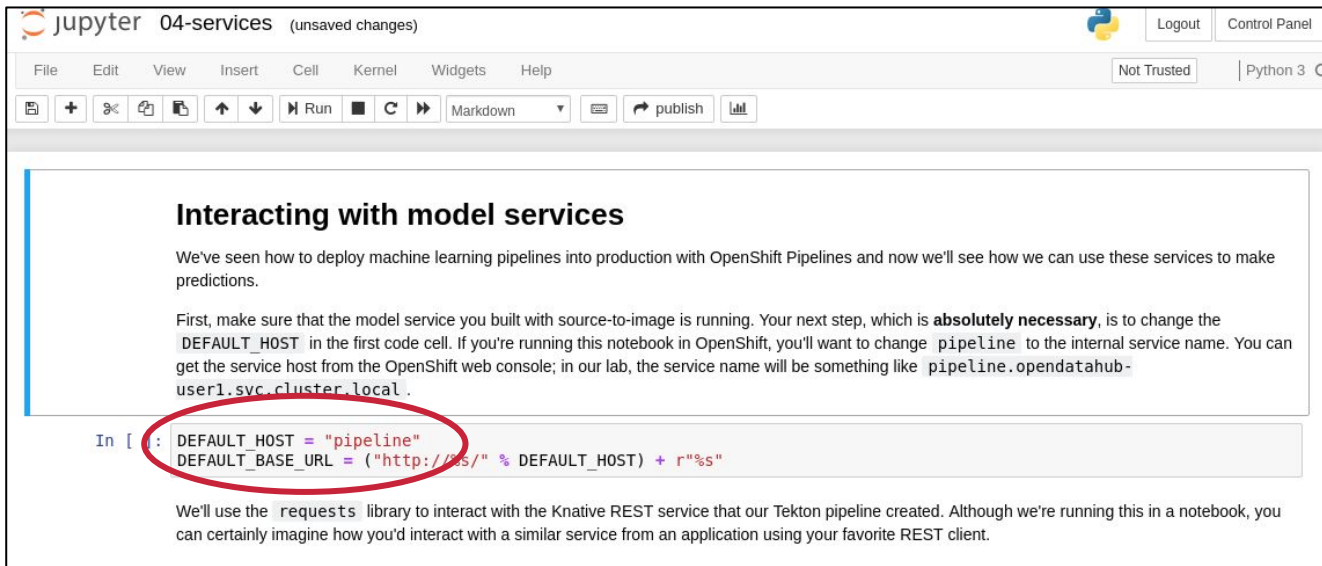
Pipelines Tech Preview Create Pipeline

Filter Name Search by name...

Name	Last Run	Task Status	Last Run Status	Last Run Time
PL pipeline	PLR pipeline-emzfvl		✓ Succeeded	a minute ago

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

# Interacting with Model Services



The screenshot shows a Jupyter Notebook interface with the title '04-services (unsaved changes)'. The notebook contains a section titled 'Interacting with model services'. The text explains that the goal is to use OpenShift Pipelines to deploy machine learning pipelines into production. It instructs the user to ensure the model service is running and to change the `DEFAULT_HOST` in the first code cell. It provides an example of a service name: `pipeline.opendatahub-user1.svc.cluster.local`. Below this, a code cell is shown with the following code:

```
In [ ]: DEFAULT_HOST = "pipeline"
        DEFAULT_BASE_URL = ("http://s/" % DEFAULT_HOST) + r"%s"
```

The code cell is circled in red. Below the code cell, the text explains that the `requests` library will be used to interact with the Knative REST service. It also mentions that the user is running this in a notebook and can imagine how they would interact with a similar service from an application using their favorite REST client.

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

# Model making its predictions

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

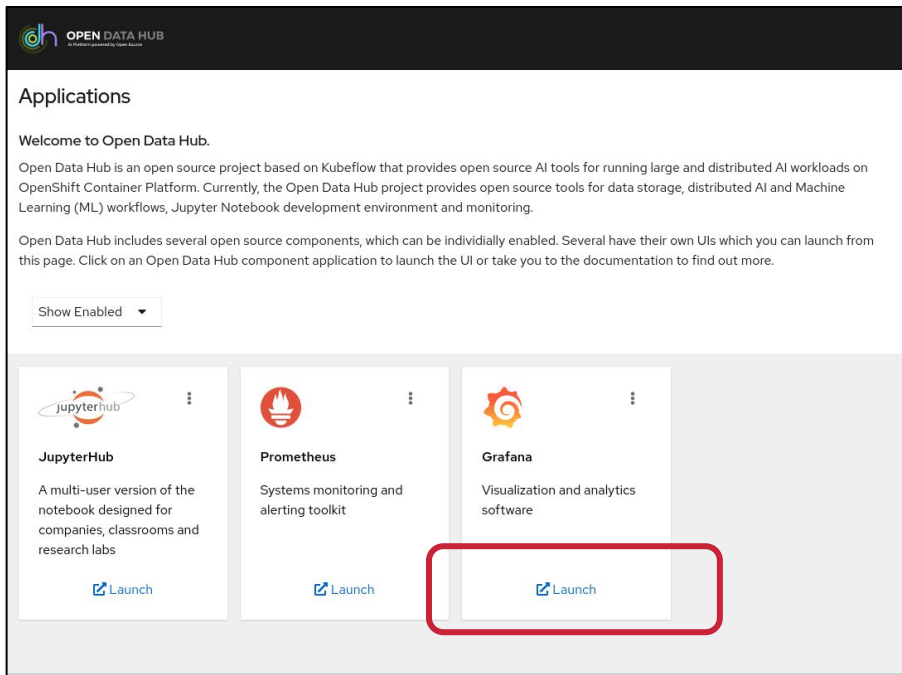
```
In [5]: sample = data.sample(200)
sample["predictions"] = score_transactions(sample)
sample
```

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	5788	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 x 9 columns

# Go back to ODH & launch Grafana



**OPEN DATA HUB**  
A Kubernetes-powered Open Source


## Applications

Welcome to Open Data Hub.

Open Data Hub is an open source project based on KubeFlow that provides open source AI tools for running large and distributed AI workloads on OpenShift Container Platform. Currently, the Open Data Hub project provides open source tools for data storage, distributed AI and Machine Learning (ML) workflows, Jupyter Notebook development environment and monitoring.

Open Data Hub includes several open source components, which can be individually enabled. Several have their own UIs which you can launch from this page. Click on an Open Data Hub component application to launch the UI or take you to the documentation to find out more.


Show Enabled ▾



**JupyterHub**

A multi-user version of the notebook designed for companies, classrooms and research labs


[Launch](#)



**Prometheus**

Systems monitoring and alerting toolkit

[Launch](#)



**Grafana**

Visualization and analytics software

[Launch](#)

# Examining Fraudulent & Legitimate transactions in a Grafana dashboard



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



# OPEN DATA HUB

AI Platform powered by Open Source

## AI & ML on OpenShift with Open Data Hub

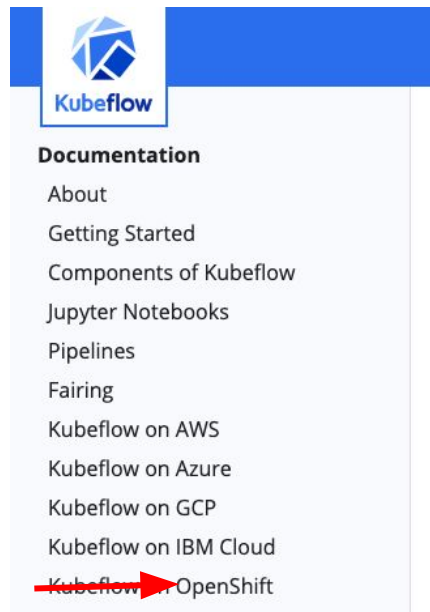
For more information on Open Data Hub:

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](https://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**





# OPEN DATA HUB - Current Releases

