



Red Hat OpenShift AI Self-Managed 2.6

**OpenShift AI tutorial – Fraud detection
example**

Red Hat OpenShift AI Self-Managed 2.6 OpenShift AI tutorial - Fraud detection example

Legal Notice

Copyright © 2024 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux® is the registered trademark of Linus Torvalds in the United States and other countries.

Java® is a registered trademark of Oracle and/or its affiliates.

XFS® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js® is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack® Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

Abstract

Step-by-step guidance for using OpenShift AI to develop and train an example model in Jupyter Notebooks, deploy the model, integrate the model into a fraud detection application, and refine the model by using automated pipelines.

Table of Contents

CHAPTER 1. INTRODUCTION	3
1.1. ABOUT THE EXAMPLE FRAUD DETECTION MODEL	3
1.2. BEFORE YOU BEGIN	3
CHAPTER 2. SETTING UP A PROJECT AND STORAGE	4
2.1. NAVIGATING TO THE OPENSHIFT AI DASHBOARD	4
2.2. SETTING UP YOUR DATA SCIENCE PROJECT	5
2.3. STORING DATA WITH DATA CONNECTIONS	7
2.3.1. Creating data connections to your own S3-compatible object storage	8
2.3.2. Running a script to install local object storage buckets and create data connections	11
2.4. ENABLING DATA SCIENCE PIPELINES	14
CHAPTER 3. CREATING A WORKBENCH AND A NOTEBOOK	16
3.1. CREATING A WORKBENCH AND SELECTING A NOTEBOOK IMAGE	16
3.2. IMPORTING THE TUTORIAL FILES INTO THE JUPYTER ENVIRONMENT	19
3.3. RUNNING CODE IN A NOTEBOOK	23
3.3.1. Try it	24
3.4. TRAINING A MODEL	25
CHAPTER 4. DEPLOYING AND TESTING A MODEL	27
4.1. PREPARING A MODEL FOR DEPLOYMENT	27
4.2. DEPLOYING A MODEL	28
4.3. TESTING THE MODEL API	31
CHAPTER 5. IMPLEMENTING PIPELINES	33
5.1. AUTOMATING WORKFLOWS WITH DATA SCIENCE PIPELINES	33
5.1.1. Create a pipeline	33
5.1.2. Add nodes to your pipeline	35
5.1.3. Specify the training file as a dependency	36
5.1.4. Create and store the ONNX-formatted output file	37
5.1.5. Configure the data connection to the S3 storage bucket	38
5.1.6. Run the Pipeline	41
5.2. RUNNING A DATA SCIENCE PIPELINE GENERATED FROM PYTHON CODE	43
CHAPTER 6. CONCLUSION	46

CHAPTER 1. INTRODUCTION

Welcome!

In this tutorial, you'll learn how to incorporate data science and artificial intelligence and machine learning (AI/ML) into an OpenShift development workflow.

You'll use an example fraud detection model to complete the following tasks:

- Explore a pre-trained fraud detection model by using Jupyter Notebooks.
- Deploy the model by using OpenShift AI model serving.
- Integrate the model into a real-time fraud detection application.
- Refine and train the model by using automated pipelines.

And you'll do all of this without having to install anything on your own computer, thanks to [Red Hat OpenShift AI](#).

1.1. ABOUT THE EXAMPLE FRAUD DETECTION MODEL

The example fraud detection model mentors credit card transactions for potential fraudulent activity. It analyzes the following credit card transaction details:

- The geographical distance from the previous credit card transaction.
- The price of the current transaction, compared to the median price of all the user's transactions.
- Whether the user completed the transaction by using the hardware chip in the credit card, entered a PIN number, or for an online purchase.

Based on this data, the model outputs the likelihood of the transaction being fraudulent.

1.2. BEFORE YOU BEGIN

[Set up your Red Hat OpenShift AI environment](#)

If you don't already have an instance of Red Hat OpenShift AI, find out more on the [developer page](#). There, you can spin up your own account on the **free OpenShift Data Science Sandbox** or learn about installing on **your own OpenShift cluster**.

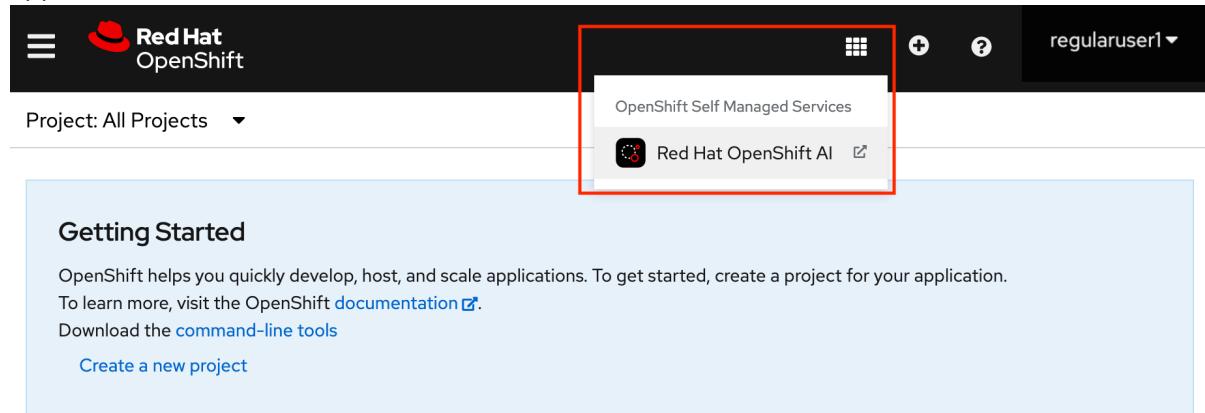
If you're ready, [start the tutorial!](#)

CHAPTER 2. SETTING UP A PROJECT AND STORAGE

2.1. NAVIGATING TO THE OPENSIFT AI DASHBOARD

Procedure

- After you log in to the OpenShift console, access the OpenShift AI dashboard by clicking the application launcher icon on the header.



- When prompted, log in to the OpenShift AI dashboard by using your OpenShift credentials. OpenShift AI uses the same credentials as OpenShift for the dashboard, notebooks, and all other components.



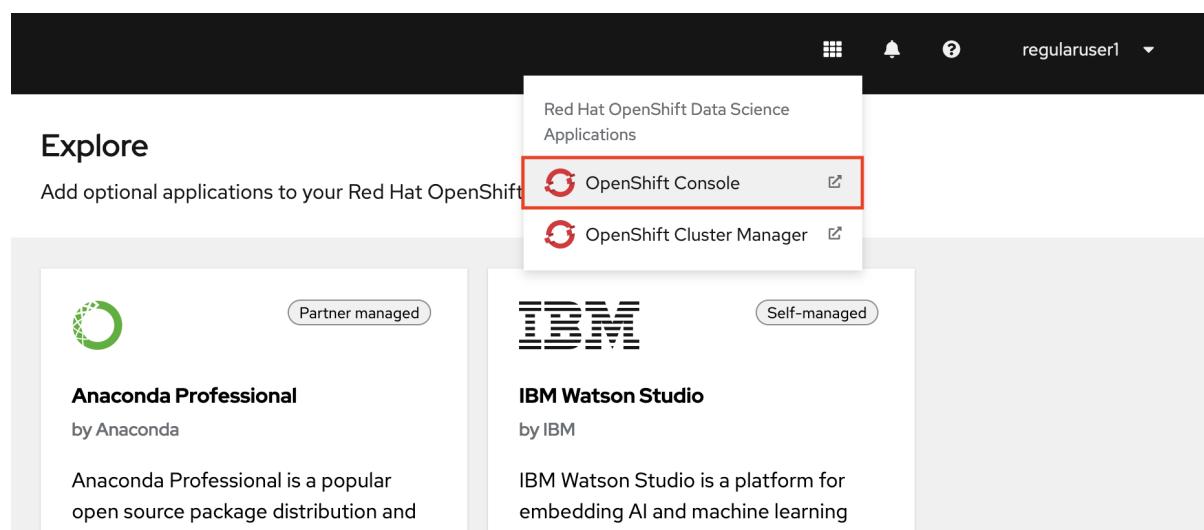
The OpenShift AI dashboard shows the status of any installed and enabled applications.

- Optionally, click **Explore** to view other available application integrations.

The screenshot shows the 'Explore' section of the OpenShift AI dashboard. On the left, a sidebar menu is open with 'Enabled' selected, showing options like 'Data Science Projects', 'Model Serving', and 'Resources'. The main area is titled 'Explore' with the sub-instruction 'Add optional applications to your Red Hat OpenShift Data Science instance.' It lists several application integrations:

- Anaconda Professional** (Partner managed) by Anaconda: Described as a popular open source package distribution and management experience optimized for commercial use.
- IBM Watson Studio** (Self-managed) by IBM: Described as a platform for embedding AI and machine learning into your business and creating custom models with your own data.
- iuvvter** (Red Hat managed): A smaller card with a logo and the name.

Note: You can navigate back to the OpenShift console in a similar fashion. Click the application launcher to access the OpenShift console.



For now, stay in the OpenShift AI dashboard.

Next step

[Setting up your data science project](#)

2.2. SETTING UP YOUR DATA SCIENCE PROJECT

Before you begin, make sure that you are logged in to **Red Hat OpenShift AI** and that you can see the dashboard:

Note that you can start a Jupyter notebook from here, but it would be a one-off notebook run in isolation. To implement a data science workflow, you must create a data science project. Projects allow you and your team to organize and collaborate on resources within separated namespaces. From a project you can create multiple workbenches, each with their own Jupyter notebook environment, and each with their own data connections and cluster storage. In addition, the workbenches can share models and data with pipelines and model servers.

Procedure

1. On the navigation menu, select **Data Science Projects**. This page lists any existing projects that you have access to. From this page, you can select an existing project (if any) or create a new one.

The screenshot shows the 'Data science projects' page. On the left, a sidebar has 'Data Science Projects' highlighted with a red box. The main content area says 'No data science projects yet.' and includes a 'Create data science project' button.

If you already have an active project that you'd like to use, select it now and skip ahead to the next section, [Storing data with data connections](#). Otherwise, continue to the next step.

2. Click **Create data science project**
3. Enter a display name and description. Based on the display name, a resource name is automatically generated, but you can change it if you'd like.

The dialog box is titled 'Create data science project'. It has three input fields: 'Name *' (display name), 'Resource name *' (resource name, derived from the display name), and 'Description'. All three fields contain the text 'Fraud detection'. At the bottom are 'Create' and 'Cancel' buttons.

Verification

You can now see its initial state. There are five types of project components:

- **Workbenches** are instances of your development and experimentation environment. They typically contain IDEs, such as JupyterLab, RStudio, and Visual Studio Code.
- A **Cluster storage** is a volume that persists the files and data you're working on within a workbench. A workbench has access to one or more cluster storage instances.
- **Data connections** contain configuration parameters that are required to connect to a data source, such as an S3 object bucket.
- **Pipelines** contain the Data Science pipelines that are executed within the project.
- **Models and model servers** allow you to quickly serve a trained model for real-time inference. You can have multiple model servers per data science project. One model server can host multiple models.

Next step

[Storing data with data connections](#)

2.3. STORING DATA WITH DATA CONNECTIONS

For this tutorial, you need two S3-compatible object storage buckets, such as Ceph, Minio, or AWS S3:

- **My Storage** - Use this bucket for storing your models and data. You can reuse this bucket and its connection for your notebooks and model servers.
- **Pipelines Artifacts** - Use this bucket as storage for your pipeline artifacts. A pipeline artifacts bucket is required when you create a pipeline server. For this tutorial, create this bucket to separate it from the first storage bucket for clarity.

You can use your own storage buckets or run a provided script that creates local Minio storage buckets for you.

Also, you must create a data connection to each storage bucket. A data connection is a resource that contains the configuration parameters needed to connect to an object storage bucket.

You have two options for this tutorial, depending on whether you want to use your own storage buckets or use a script to create local Minio storage buckets:

- If you want to use your own S3-compatible object storage buckets, create data connections to them as described in [Creating data connections to your own S3-compatible object storage](#).
- If you want to run a script that installs local Minio storage buckets and creates data connections to them, for the purposes of this tutorial, follow the steps in [Running a script to install local object storage buckets and create data connections](#).

2.3.1. Creating data connections to your own S3-compatible object storage



NOTE

If you do not have your own s3-compatible storage, or if you want to use a disposable local Minio instance instead, skip this section and follow the steps in [Running a script to install local object storage buckets and create data connections](#).

Prerequisite

To create data connections to your existing S3-compatible storage buckets, you need the following credential information for the storage buckets:

- Endpoint URL
- Access key
- Secret key
- Region
- Bucket name

If you don't have this information, contact your storage administrator.

Procedures

Create data connections to your two storage buckets.

Create a data connection for saving your data and models

1. In the OpenShift AI dashboard, navigate to the page for your data science project.
2. Under **Components**, click **Data connections**.
3. Click **Add data connection**

The screenshot shows the Data Science Projects interface for a 'Fraud detection' project. The left sidebar has 'Data Science Projects' selected. The main area shows sections for 'Workbenches', 'Cluster storage', and 'Data connections'. The 'Add data connection' button for 'Data connections' is highlighted with a red box.

- Fill out the **Add data connection** form and name your connection **My Storage**. This connection is for saving your personal work, including data and models.

Add data connection

Name *
My Storage

Access key *
yourccesskey

Secret key *
.....

Endpoint *
https://storage-endpoint.storage-cluster.com:1234

Region
us-east-1

Bucket
my-storage

Connected workbench
No available workbenches

Add data connection **Cancel**

- Click **Add data connection**

Create a data connection for saving pipeline artifacts

**NOTE**

If you do not intend to complete the pipelines section of the tutorial, you can skip this step.

1. Click **Add data connection**
2. Fill out the form and name your connection **Pipeline Artifacts**

Add data connection

Name *
Pipeline Artifacts

Access key *
yourcesskey

Secret key *
..... (eye icon)

Endpoint *
https://storage-endpoint.storage-cluster.com:1234

Region
us-east-1

Bucket
pipeline-artifacts

Connected workbench
No available workbenches ▼

Add data connection Cancel

3. Click **Add data connection**

Verification

Check to see that your data connections are listed in the project.

Data connections Add data connection

Name	Type	Connected workbenches	Provider	⋮
My Storage ②	Object storage	No connections	AWS S3	⋮
Pipeline Artifacts ②	Object storage	No connections	AWS S3	⋮

Next step

[Creating a workbench](#)

2.3.2. Running a script to install local object storage buckets and create data connections

For convenience, the provided script installs two data connections (and associated secrets) and two Minio buckets as s3-compatible storage. The script creates a random user and password for security. This script is based on the instructions for installing Minio in this [guide](#).



IMPORTANT

The storage buckets that this script creates are **not** meant for production usage.



NOTE

If you want to connect to your own storage, see [Creating data connections to your own S3-compatible object storage](#).

Prerequisite

You must know the OpenShift resource name for your data science project so that you run the provided script in the correct project. To get the project's resource name:

In the OpenShift AI dashboard, select **Data Science Projects** and then hover your cursor over the ? icon next to the project name. A text box appears with information about the project, including its resource name:

Data science projects

View your existing projects or create new projects.

The screenshot shows the OpenShift AI Data Science Projects interface. At the top, there is a search bar labeled "Find by name" and a blue button "Create data science project". Below the search bar, there is a table header with columns "Name" and "Created". A tooltip is displayed over the "Name" column of the first row, which contains the project name "Fraud detection". The tooltip content is: "Resource names and types are used to find your resources in OpenShift." It also shows the "Resource name" field with the value "fraud-detection" and a clipboard icon, and the "Resource type" field with the value "Project".



NOTE

The following procedure describes how to run the script from the OpenShift console. If you are knowledgeable in OpenShift and can access the cluster from the command line, instead of following the steps in this procedure, you can use the following command to run the script:

```
oc apply -n <your-project-name/> -f https://github.com/rh-aiservices-bu/fraud-detection/raw/main/setup/setup-s3.yaml
```

Procedure

- In the OpenShift AI dashboard, click the application launcher icon and then select the **OpenShift Console** option.

Data science projects

View your existing projects or create new pr

Name Find by name Create data science project Launch Jupyter

Name	Workbench	Status	Created
Fraud detection <small>regularuser1</small>	-	-	11/9/2023, 1

- In the OpenShift console, click + in the top navigation bar.

Project: All Projects

Topology

Select a Project to view the topology or [create a Project](#).

- Select your project from the list of projects.

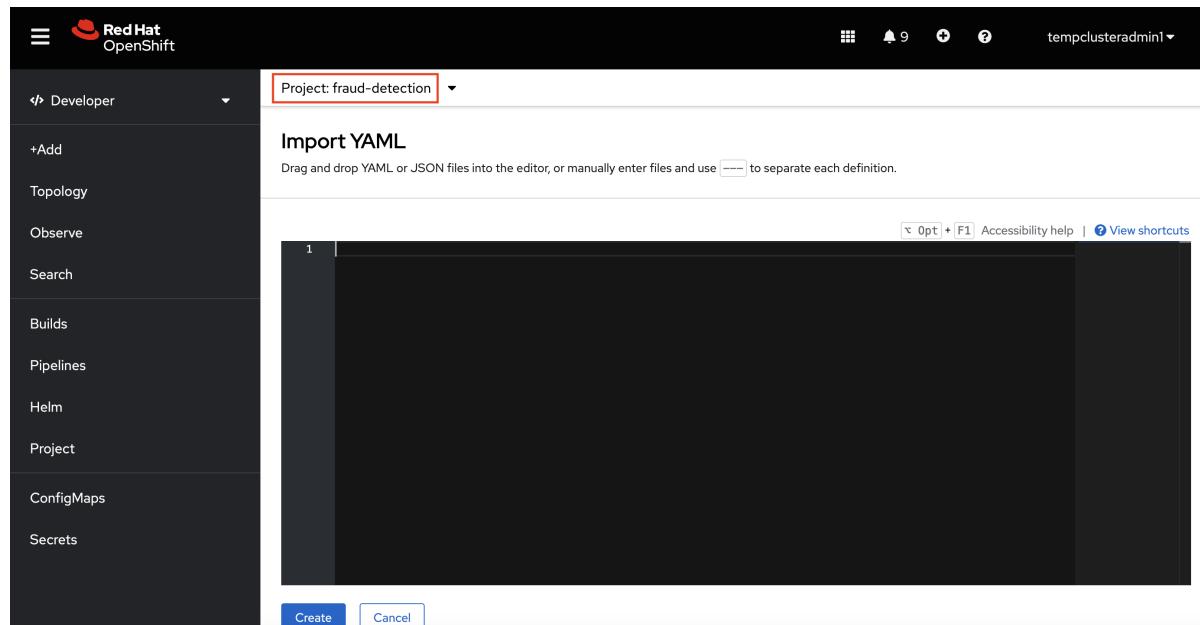
Project: All Projects

Select project...

Show default projects

Status	Requester	Created
Active	No requester	May 11, 2023, 5:32 AM

- Verify that you selected the correct project.



5. Copy the following code and paste it into the **Import YAML** editor.

Note: This code gets and applies the **setup-s3-no-sa.yaml** file.

```
---
apiVersion: v1
kind: ServiceAccount
metadata:
  name: demo-setup
---
apiVersion: rbac.authorization.k8s.io/v1
kind: RoleBinding
metadata:
  name: demo-setup-edit
roleRef:
  apiGroup: rbac.authorization.k8s.io
  kind: ClusterRole
  name: edit
subjects:
  - kind: ServiceAccount
    name: demo-setup
---
apiVersion: batch/v1
kind: Job
metadata:
  name: create-s3-storage
spec:
  selector: {}
  template:
    spec:
      containers:
        - args:
            - -ec
            - |-echo -n 'Setting up Minio instance and data connections'
            - oc apply -f https://github.com/rh-aiservices-bu/fraud-detection/raw/main/setup/setup-s3-no-sa.yaml
          command:
```

```

- /bin/bash
image: image-registry.openshift-image-registry.svc:5000/openshift/tools:latest
imagePullPolicy: IfNotPresent
name: create-s3-storage
restartPolicy: Never
serviceAccount: demo-setup
serviceAccountName: demo-setup

```

6. Click **Create**.

Verification

You should see a "Resources successfully created" message and the following resources listed:

- **demo-setup**
- **demo-setup-edit**
- **create s3-storage**

Next step

[Creating a workbench](#)

2.4. ENABLING DATA SCIENCE PIPELINES



NOTE

If you do not intend to complete the pipelines section of the workshop you can skip this step and move on to the next section, [Create a Workbench](#).

In this section, you prepare your tutorial environment so that you can use data science pipelines.

Procedure

1. In the Data Science dashboard, navigate to **Data Science Pipelines → Pipelines**.
2. Click **Create pipeline server**.

Name	Type	Connected workbenches	Provider
My Storage	Object storage	No connections	AWS S3
Pipeline Artifacts	Object storage	No connections	AWS S3

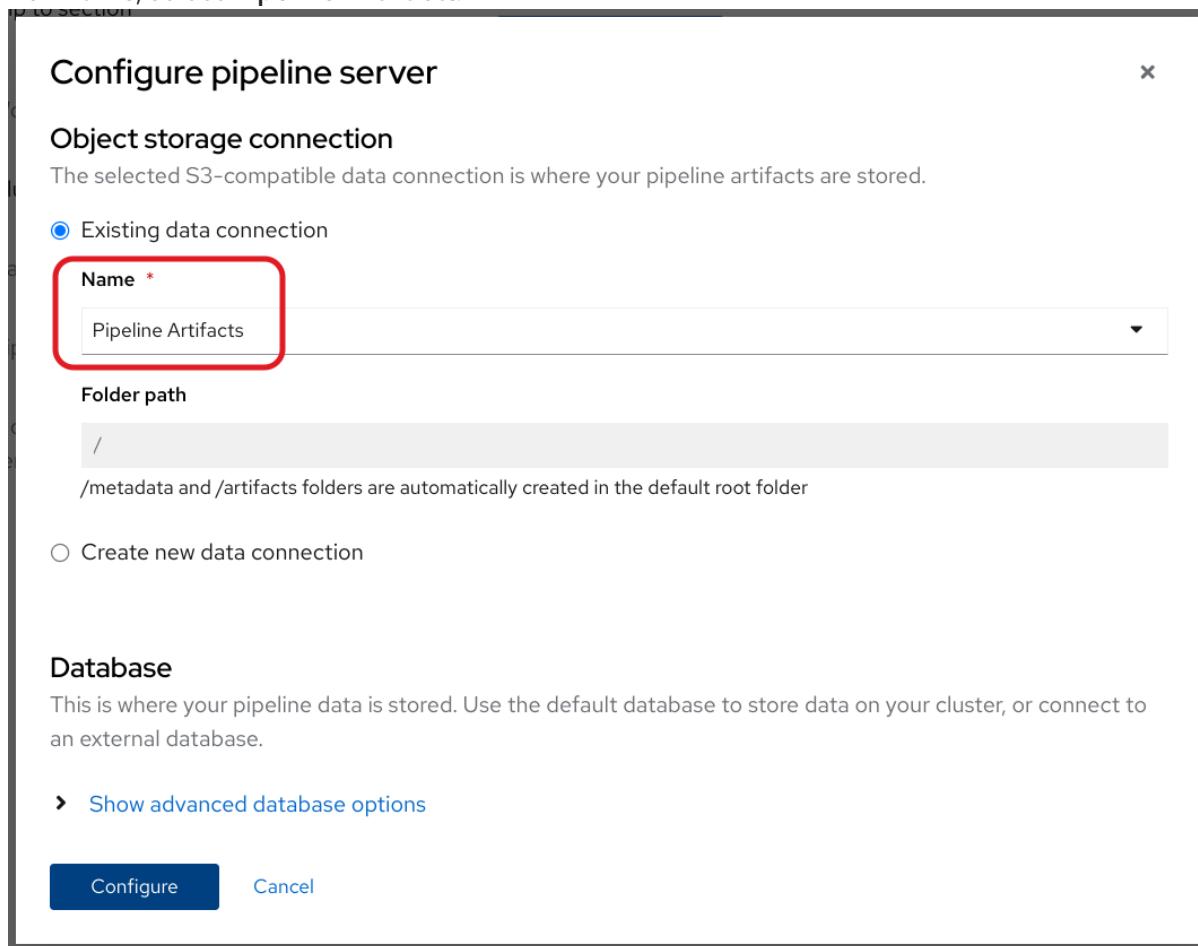
No pipeline server
To import a pipeline, first create a pipeline server.
[Create a pipeline server](#)

Models and model servers [Add server](#)

No model servers
Before deploying a model, you must first add a model server.

3. In the **Configure pipeline server** form, select **Existing data connection**

4. For **Name**, select **Pipeline Artifacts**.



5. Leave the database configuration as the default.

6. Click **Configure**.

Verification

Check the **Pipelines** page. Pipelines are enabled when the **Pipeline server actions** option appears and the **Create pipeline server** button no longer appears.

The screenshot shows the 'Pipelines' page. It has a 'Pipelines' tab and an 'Import pipeline' button. Below is a large '+ New Pipeline' button with a plus sign icon. The text 'No pipelines' is displayed, followed by 'To get started, import a pipeline.'

Next step

[Automating workflows with data science pipelines](#)

[Running a data science pipeline generated from Python code](#)

CHAPTER 3. CREATING A WORKBENCH AND A NOTEBOOK

3.1. CREATING A WORKBENCH AND SELECTING A NOTEBOOK IMAGE

A workbench is an instance of your development and experimentation environment. Within the workbench you can select a notebook image for your data science work.

Prerequisite

- You created a **My Storage** data connection as described in [Storing data with data connections](#).

Procedure

1. Navigate to the project detail page for the data science project that you created in [Setting up your data science project](#).
2. Click the **Create workbench** button.

3. Fill out the name and description.

Name *	<input type="text" value="Fraud Detection"/>
Description	<input type="text" value="My Fraud Detection workbench"/>

Red Hat provides several supported notebook images. In the **Notebook image** section, you can choose one of these images or any custom images that an administrator has set up for you. The **Tensorflow** image has the libraries needed for this tutorial.

4. Select the latest **Tensorflow** image.

Notebook image

Image selection *

TensorFlow ▾

Version selection *

2023.2 (Recommended) ▾

Hover an option to learn more information about the packages included.

5. Select a small deployment.

Deployment size

Container size

Small ▾

6. Leave the default environment variables and storage options.

Environment variables

[+ Add variable](#)

Cluster storage

Cluster storage will mount to /

Create new persistent storage

This creates storage that is retained when logged out.

Name *

Fraud Detection

Description

Persistent storage size

- 20 + GiB

Use existing persistent storage

This reuses a previously created persistent storage.

7. Under Data connections, select Use existing data connection and select My Storage (the object storage that you configured previously) from the list.

Data connections

Use a data connection

Create new data connection

Use existing data connection

Data connection *

My Storage



8. Click the Create workbench button.

Create workbench

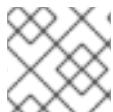
Verification

In the project details page, the status of the workbench changes from Starting to Running.

Workbenches

[Create workbench](#)

Name	Notebook image	Container size	Status	
Workshop ⓘ My Workshop Workbench	TensorFlow	Small	Running	Open



NOTE

If you made a mistake, you can edit the workbench to make changes.

Workbenches

[Create workbench](#)

Name	Notebook image	Container size	Status	
Workshop ⓘ My Workshop Workbench	TensorFlow	Small	Running	Open
Workbench storages Workshop 0Gi 20Gi Mount path: /	Packages TensorFlow v2.11 Tensorboard v2.11 Boto3 v1.26	Limits 4 CPU, 8Gi Memory	Requests	Edit workbench Delete workbench

Next step

[Importing the tutorial files into the Jupyter environment](#)

3.2. IMPORTING THE TUTORIAL FILES INTO THE JUPYTER ENVIRONMENT

The Jupyter environment is a web-based environment, but everything you do inside it happens on **Red Hat OpenShift AI** and is powered by the **OpenShift** cluster. This means that, without having to install and maintain anything on your own computer, and without disposing of lots of local resources like CPU, GPU and RAM, you can conduct your Data Science work in this powerful and stable managed environment.

Prerequisite

You created a workbench, as described in [Creating a workbench and selecting a Notebook image](#).

Procedure

- Click the **Open** link next to your workbench. If prompted, log in and allow the Notebook to authorize your user.

Workbenches

[Create workbench](#)

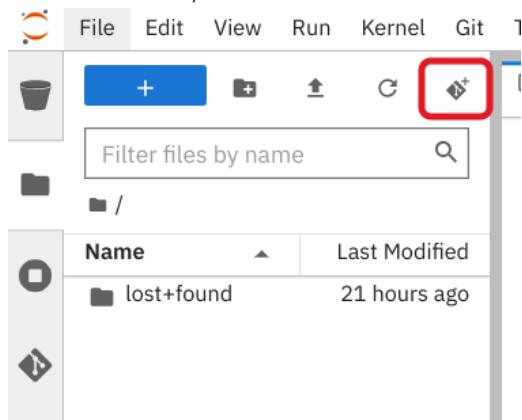
Name	Notebook image	Container size	Status	
Fraud Detection ⓘ My fraud detection workbench	TensorFlow	Small	Running	Open

Your Jupyter environment window opens.

This file-browser window shows the files and folders that are saved inside your own personal space in OpenShift AI.

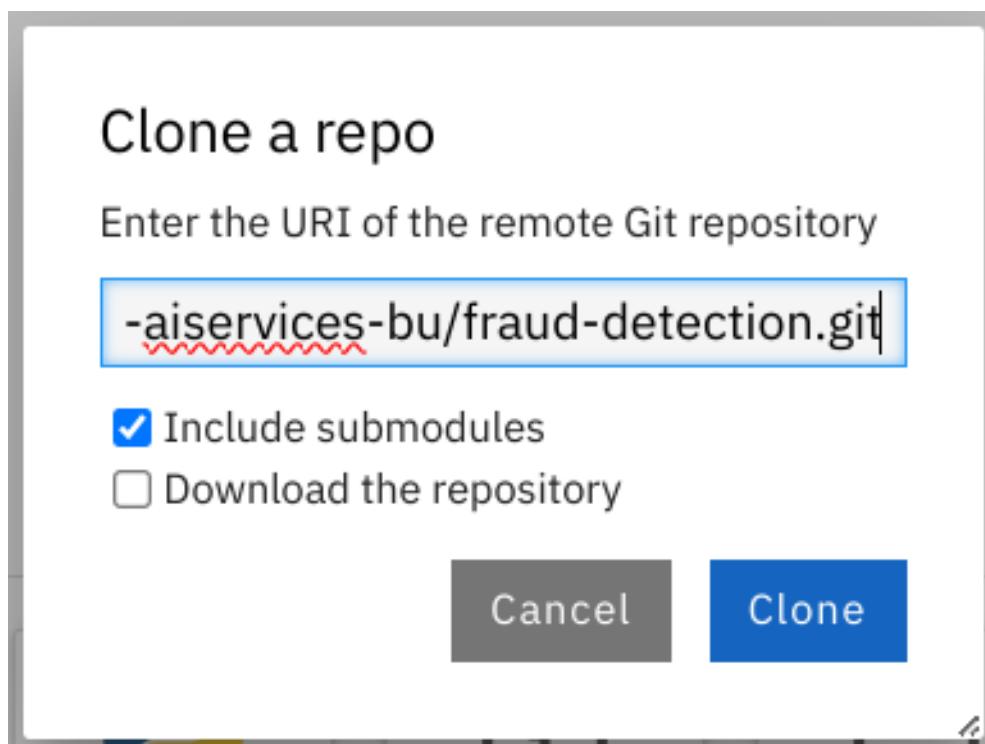
2. Bring the content of this tutorial inside your Jupyter environment:

- a. On the toolbar, click the **Git Clone** icon:



- b. Enter the following tutorial Git **https** URL:

```
https://github.com/rh-aiservices-bu/fraud-detection.git
```

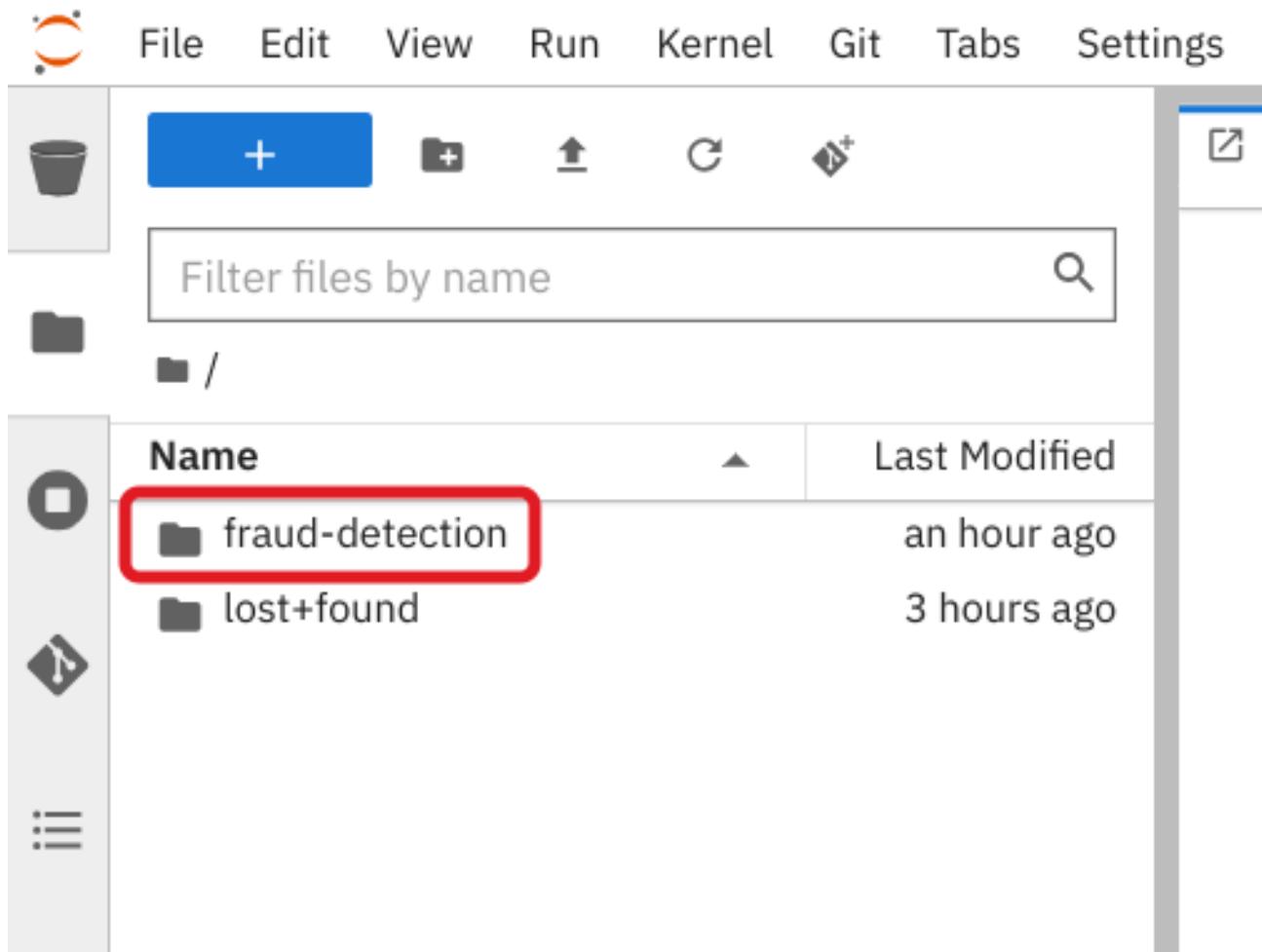


- c. Check the **Include submodules** option.

- d. Click **CLONE**.

Verification

Double-click the newly-created folder, **fraud-detection**:



In the file browser, you should see the notebooks that you cloned from Git.

The screenshot shows a file browser interface with the following details:

- Toolbar:** File, Edit, View, Run, Kernel, Git, Tabs, Settings.
- File Operations:** + (Create), folder (New Folder), upload (Upload), refresh (Refresh), and a search icon.
- Search Bar:** Filter files by name.
- Current Path:** / fraud-detection /
- Table View:** A list of files and folders in the current directory, sorted by Name (ascending).
- Columns:** Name (sorted), Last Modified.
- Content:**

Name	Last Modified
artifact	2 hours ago
assets	2 hours ago
data	2 hours ago
models	2 hours ago
pipeline	2 hours ago
setup	2 hours ago
utils	an hour ago
workshop	2 hours ago
0_sandbox.ipynb	2 hours ago
1_experiment_train.ipynb	2 hours ago
2_save_model.ipynb	2 hours ago
3_deploy_model.ipynb	2 hours ago
4_rest_requests.ipynb	an hour ago
5_grpc_requests.ipynb	an hour ago
6 Train Save.pipeline	2 hours ago
7_get_data_train_upload....	2 hours ago
LICENSE	2 hours ago
README.md	2 hours ago

Next step

[Running code in a notebook](#)

or

[Training a model](#)

3.3. RUNNING CODE IN A NOTEBOOK



NOTE

If you're already at ease with Jupyter, you can [skip to the next section](#).

A notebook is an environment where you have *cells* that can display formatted text or code.

This is an empty cell:

```
In [ ]:
```

This is a cell with some code:

```
In [ ]: def print_some_text(entered_text):
    print('This is what you entered:' + entered_text)

my_text = 'Hello world!'
print_some_text(my_text)
```

Code cells contain Python code that you can run interactively. You can modify the code and then run it. The code does not run on your computer or in the browser, but directly in the environment that you are connected to, **Red Hat OpenShift AI** in our case.

You can run a code cell from the notebook interface or from the keyboard:

- **From the user interface:** Select the cell (by clicking inside the cell or to the left side of the cell) and then click **Run** from the toolbar.



- **From the keyboard:** Press **CTRLENTER** to run a cell or press **SHIFTENTER** to run the cell and automatically select the next one.

After you run a cell, you can see the result of its code as well as information about when the cell was run, as shown in this example:

```
In [2]: def print_some_text(entered_text):
    print('This is what you entered:' + entered_text)

my_text = 'Hello world!'
print_some_text(my_text)
executed in 9ms, finished 14:47:30 2021-04-13
This is what you entered:Hello world!
```

When you save a notebook, the code and the results are saved. You can reopen the notebook to look at the results without having to run the program again, while still having access to the code.

Notebooks are so named because they are like a physical *notebook*: you can take notes about your experiments (which you will do), along with the code itself, including any parameters that you set. You can see the output of the experiment inline (this is the result from a cell after it's run), along with all the notes that you want to take (to do that, from the menu switch the cell type from **Code** to **Markdown**).

3.3.1. Try it

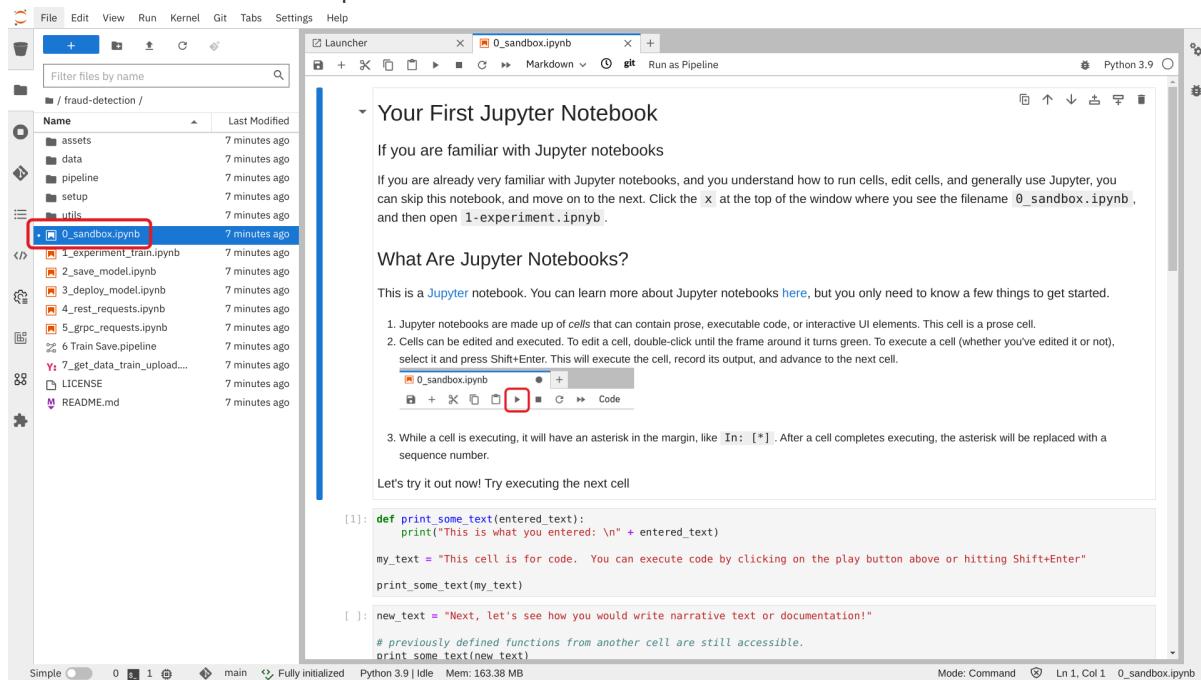
Now that you know the basics, give it a try!

Prerequisite

- You have imported the tutorial files into your Jupyter environment as described in [Importing the tutorial files into the Jupyter environment](#).

Procedure

1. In your Jupyter environment, locate the **0_sandbox.ipynb** file and double-click it to launch the notebook. The notebook opens in a new tab in the content section of the environment.



2. Experiment by, for example, running the existing cells, adding more cells and creating functions. You can do what you want - it's your environment and there is no risk of breaking anything or impacting other users. This environment isolation is also a great advantage brought by OpenShift AI.
3. Optionally, create a new notebook in which the code cells are run by using a Python 3 kernel:
 - a. Create a new notebook by either selecting **File → New → Notebook** or by clicking the Python 3 tile in the Notebook section of the launcher window:



You can use different kernels, with different languages or versions, to run in your notebook.

Additional resource

To learn more about notebooks, go to [the Jupyter site](#).

Next step

Training a model

3.4. TRAINING A MODEL

Now that you know how the Jupyter notebook environment works, the real work can begin!

In your notebook environment, open the **1_experiment_train.ipynb** file and follow the instructions directly in the notebook. The instructions guide you through some simple data exploration, experimentation, and model training tasks.

```

File Edit View Run Kernel Git Tabs Settings Help
+ 1_experiment_train.ipynb + Python 3.9
Filter files by name
/ fraud-detection /
Name Last Modified
artifact 10 minutes ago
assets 40 minutes ago
data 40 minutes ago
models 9 minutes ago
pipeline 40 minutes ago
setup 40 minutes ago
utils 40 minutes ago
O 0_sandbox.ipynb seconds ago
I 1_experiment_train.ipynb seconds ago
2_save_model.ipynb 40 minutes ago
3_deploy_model.ipynb 40 minutes ago
4_rest_requests.ipynb 40 minutes ago
5_grpc_requests.ipynb 40 minutes ago
6 Train Save.pipeline seconds ago
Y 7_get_data_train_upload... 18 minutes ago
LICENSE 40 minutes ago
README.md 40 minutes ago

Experiment
Install Python Dependencies
[ ]: !pip install onnx onnxruntime seaborn tf2onnx
Now we can import those dependencies we need to run the code
[ ]: import numpy as np
import pandas as pd
from keras.models import Sequential
from keras.layers import Dense, Dropout, BatchNormalization, Activation
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.utils import class_weight
import tensorflow as tf
import onnx
import pickle
from pathlib import Path

Load the CSV data which we will use to train the model. It contains the following fields:


- distancefromhome - The distance from home where the transaction happened.
- distancefromlast_transaction - The distance from last transaction happened.
- ratiotomedianpurchaseprice - Ratio of purchased price compared to median purchase price.
- repeat_retailer - If it's from a retailer that already has been purchased from before.
- used_chip - If the (credit card) chip was used.
- usedpinnumber - If the PIN number was used.
- online_order - If it was an online order.
- fraud - If the transaction is fraudulent.


[ ]: Data = pd.read_csv('data/card_transdata.csv')
Data.head()
[ ]: # Set the input (X) and output (Y) data.

```

Next step

[Preparing a model for deployment](#)

CHAPTER 4. DEPLOYING AND TESTING A MODEL

4.1. PREPARING A MODEL FOR DEPLOYMENT

After you train a model, you can deploy it by using the OpenShift AI model serving capabilities.

To prepare a model for deployment, you must move the model from your workbench to your S3-compatible object storage. You use the data connection that you created in the [Storing data with data connections](#) section and upload the model from a notebook. You also convert the model to the portable ONNX format. ONNX allows you to transfer models between frameworks with minimal preparation and without the need for rewriting the models.

Prerequisites

- You created the data connection **My Storage**.

Data connections

- Use a data connection
- Create new data connection
- Use existing data connection

Data connection *

✖ ▾

- You added the **My Storage** data connection to your workbench.

Workbenches		Create workbench		
Name	Notebook image	Container size	Status	
Workshop ⓘ My Workshop Workbench	TensorFlow	Small	Running	Open ⋮
Workshop storages Workshop 0Gi	Packages TensorFlow v2.11 Tensorboard v2.11 Boto3 v1.26	Limits 4 CPU, 8Gi Memory	Requests	Edit workbench Delete workbench

Procedure

1. In your Jupyter environment, open the **2_save_model.ipynb** file.
2. Follow the instructions in the notebook to make the model accessible in storage and save it in the portable ONNX format.

```

File Edit View Run Kernel Git Tabs Settings Help
+ Filter files by name
/fraud-detection/
Name Last Modified
assets a day ago
images an hour ago
models an hour ago
utils 2 hours ago
0_sandbox.ipynb an hour ago
1_experiment.ipynb an hour ago
2_save_model.ipynb 27 minutes ago
3_deploy_model.ip... a day ago
4_rest_requests.ip... an hour ago
5_grpc_inference.i... an hour ago
LICENSE a day ago
README.md a day ago

Save the Model

To save this model to use from various locations, including other notebooks or serving the model, we need to upload it to s3 compatible storage.

[1]: import os
import boto3
import botocore

aws_access_key_id = os.environ.get('AWS_ACCESS_KEY_ID')
aws_secret_access_key = os.environ.get('AWS_SECRET_ACCESS_KEY')
endpoint_url = os.environ.get('AWS_S3_ENDPOINT')
region_name = os.environ.get('AWS_DEFAULT_REGION')
bucket_name = os.environ.get('AWS_S3_BUCKET')

session = boto3.Session(aws_access_key_id=aws_access_key_id,
aws_secret_access_key=aws_secret_access_key)

s3_resource = session.resource(
    's3',
    config=botocore.client.Config(signature_version='s3v4'),
    endpoint_url=endpoint_url,
    region_name=region_name)

bucket = s3_resource.Bucket(bucket_name)

def upload_directory_to_s3(local_directory, s3_prefix):
    for root, dirs, files in os.walk(local_directory):
        for filename in files:
            file_path = os.path.join(root, filename)
            relative_path = os.path.relpath(file_path, local_directory)
            s3_key = os.path.join(s3_prefix, relative_path)
            print(f"{file_path} -> {s3_key}")
            bucket.upload_file(file_path, s3_key)

[2]: upload_directory_to_s3("models", "models")
models/volov5n/model.onnx -> models/volov5n/model.onnx

```

Simple 2 dev Fully initialized Python 3.9 | Idle Mem: 631.75 MB Mode: Command Ln 3, Col 45 2_save_model.ipynb

Verification

When you have completed the notebook instructions, the **models/fraud/model.onnx** file is in your object storage and it is ready for your model server to use.

Next step

[Deploying a model](#)

4.2. DEPLOYING A MODEL

Now that the model is accessible in storage and saved in the portable ONNX format, you can use an OpenShift AI model server to deploy it as an API.

OpenShift AI multi-model servers can host several models at once. You create a new model server and deploy your model to it.

Procedure

1. In the OpenShift AI dashboard, navigate to **Models and model servers**.
2. Click **Add server**.

The screenshot shows the left sidebar with 'Data Science Projects' selected. The main area has two sections: 'Pipelines' and 'Models and model servers'. The 'Pipelines' section shows 'No pipelines' with a 'Import pipeline' button. The 'Models and model servers' section shows 'No model servers' with a red box around the 'Add server' button.

Name	Type	Connected workbenches
My Storage	Object storage	Fraud Detection
Pipeline Artifacts	Object storage	No connections

3. In the form:

- Fill out the **Model server name**, for example **Model Server**.
- Select **OpenVINO Model Server**.
- Leave the other fields with the default settings.

Add model server

Model server name *
Model Server

Serving runtime *
OpenVINO Model Server

Model server replicas

Number of model server replicas to deploy
- 1 +

Compute resources per replica

Model server size
Small

Model route

Make deployed models available through an external route

Token authorization

Require token authentication

Add **Cancel**

4. Click **Add**.

5. In the **Models and model servers** list, next to the new model server, click **Deploy model**.

Models and model servers		Add server			
Model Server Name	Serving Runtime	Deployed models	Tokens		
Model Server	OpenVINO Model Server	0	Tokens disabled	Deploy model	:

6. In the form:

- Fill out the **Model Name** with the value **fraud**.
- Select the server that you created (for example, **Model Server**).
- Select **Existing data connection My Storage**.
- Enter the path to your uploaded model: **models/fraud/model.onnx**

Deploy model

Configure properties for deploying your model

Model Name *
fraud

Model server
Model Server

Model framework (name - version) *
onnx - 1

Model location

Existing data connection

Name *
My Storage

Path *
/ models/fraud/model.onnx
Enter a path to a model or folder. This path cannot point to a root folder.

New data connection

Deploy **Cancel**

Verification

Wait for the model to deploy and for the **Status** to show a green checkmark.

Models and model servers		Add server		
Model Server Name	Serving Runtime	Deployed models	Tokens	
Model Server	OpenVINO Model Server	1	Tokens disabled	<button>Deploy model</button> ::
Model name ↑	Inference endpoint		Status	
fraud ⓘ	Internal Service		✓	::

Next step

[Testing the model API](#)

4.3. TESTING THE MODEL API

Now that you've deployed the model, you can test its API endpoints.

When you created the model server, you did **not** create a route for external access to the API and you did not protect it with an authentication token. By default, if you do not specify external access, the model server provides an internal endpoint with no authentication.

You can communicate directly with this internal service in the same way that an application in your project would. An easy way to test it is from a notebook in the same project.

Procedure

1. In the Data Science dashboard, navigate to the project details page and scroll down to the **Models and model servers** section.
2. Take note of the model's resource name (API endpoint name) and the internal service's grpcURL and restURL. You need this information when you test the model API.

Model name	Resource names and types are used to find your resources in OpenShift.	Status
fraud ⓘ	Resource name fraud	✓
Resource type	InferenceService	::

The screenshot shows the ModelHub interface. On the left, there's a sidebar with 'Models and Pipelines' and a search bar. In the center, a modal window titled 'Internal Service can be accessed inside the cluster' displays configuration details for the 'fraud' model:

- grpcUrl**: grpc://modelmesh-serving.fraud-detection:8033
- restUrl**: http://modelmesh-serving.fraud-detection:8008
- url**: grpc://modelmesh-serving.fraud-detection:8033

To the right of the modal, the main interface shows the 'Model' tab selected. It lists the 'fraud' model with a status indicator (green checkmark) and a 'Deploy model' button.

3. Return to the Jupyter environment and try out your new endpoint. You'll try REST API calls in [4_rest_requests.ipynb](#) and gRPC requests in [5_grpc_requests.ipynb](#).

The screenshot shows a Jupyter Notebook interface with several open files. The left sidebar shows a file tree with various notebooks and scripts. The main area is titled 'REST Inference' under a 'Setup' section. It contains the following code snippet:

```
deployed_model_name = "fraud"
rest_url = "http://modelmesh-serving:8008"
infer_url = f"{rest_url}/v2/models/{deployed_model_name}/infer"
```

Below this, under 'Request Function', is another code snippet:

```
import requests

def rest_request(data):
    json_data = {
        "inputs": [
            {
                "name": "dense_input",
                "shape": [1, 5],
                "datatype": "FP32",
                "data": data
            }
        ]
    }

    response = requests.post(infer_url, json=json_data)
    response_dict = response.json()
    return response_dict['outputs'][0]['data']

data = [0.3111400080477545, 1.9459399775518593, 1.0, 0.0, 0.0]
prediction = rest_request(data)
```

The notebook tabs at the top show '4_rest_requests.ipynb' and '5_grpc_requests.ipynb' as being run.

Next step

[Automating workflows with data science pipelines](#)

[Running a data science pipeline generated from Python code](#)

CHAPTER 5. IMPLEMENTING PIPELINES

5.1. AUTOMATING WORKFLOWS WITH DATA SCIENCE PIPELINES

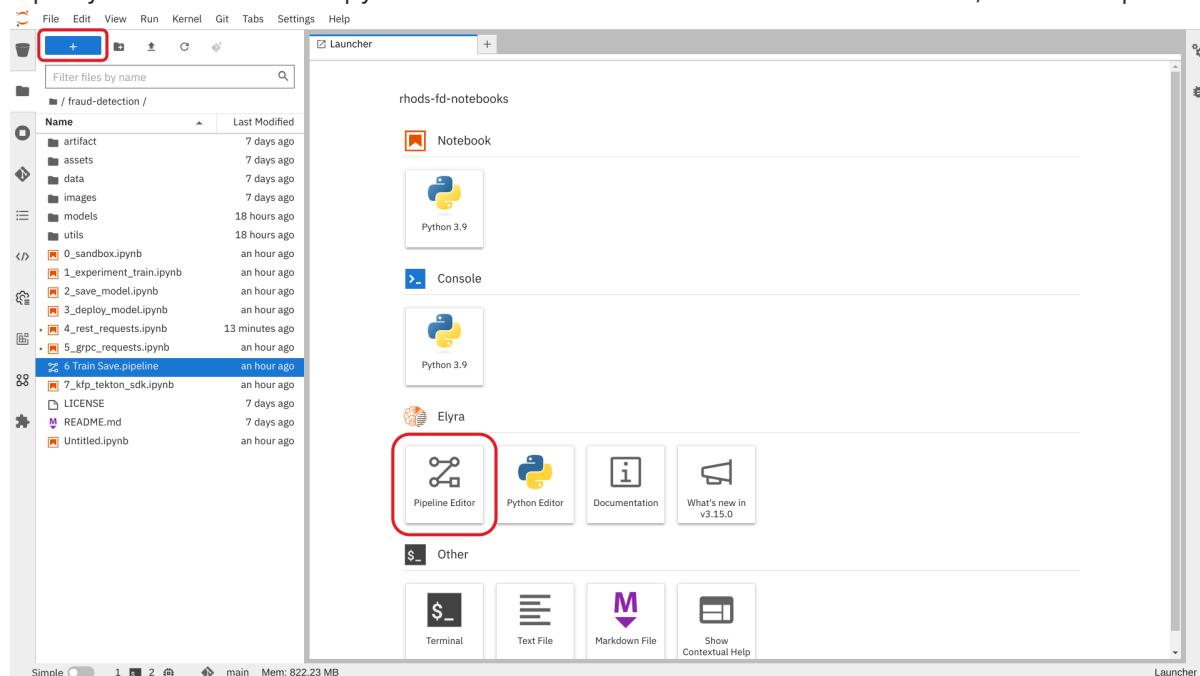
In previous sections of this tutorial, you used a notebook to train and save your model. Optionally, you can automate these tasks by using Red Hat OpenShift AI pipelines. Pipelines offer a way to automate the execution of multiple notebooks and Python code. By using pipelines, you can execute long training jobs or retrain your models on a schedule without having to manually run them in a notebook.

In this section, you create a simple pipeline by using the GUI pipeline editor. The pipeline uses the notebook that you used in previous sections to train a model and then save it to S3 storage.

Note: Your completed pipeline should look like the one in the **6 Train Save.pipeline** file.

5.1.1. Create a pipeline

1. Open your workbench's JupyterLab environment. If the launcher is not visible, click **+** to open it.



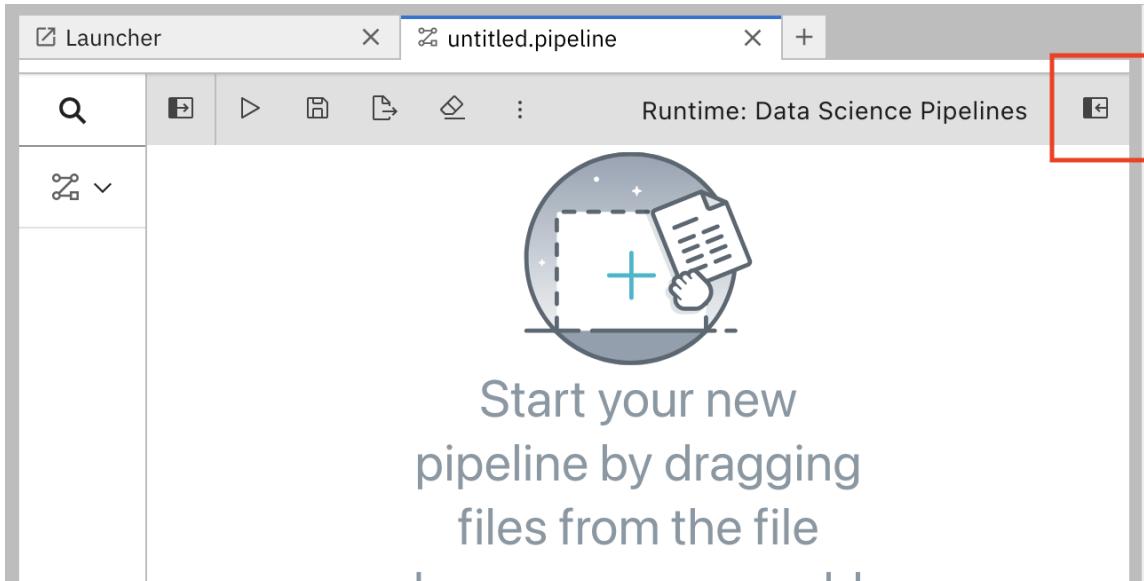
2. Click **Pipeline Editor**.



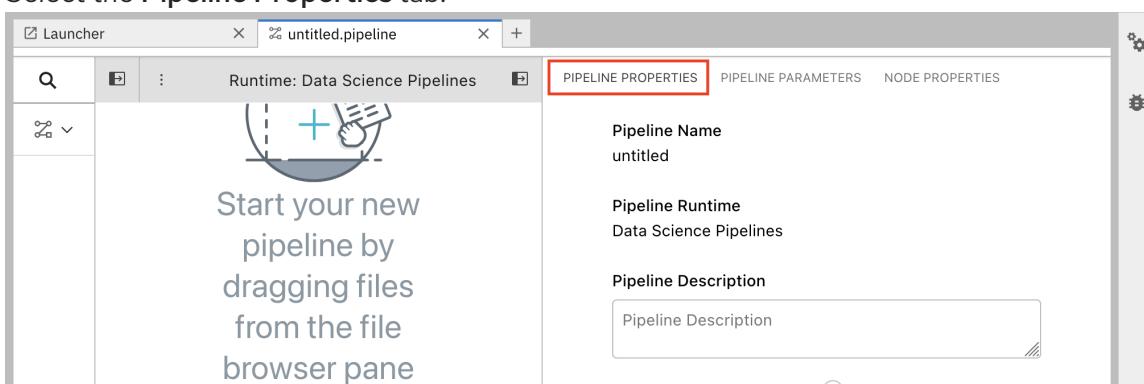
You've created a blank pipeline!

3. Set the default runtime image for when you run your notebook or Python code.

- a. In the pipeline editor, click **Open Panel**



- b. Select the **Pipeline Properties** tab.



- c. In the **Pipeline Properties** panel, scroll down to **Generic Node Defaults** and **Runtime Image**. Set the value to **Tensorflow with Cuda and Python 3.9 (UBI 9)**.

[PIPELINE PROPERTIES](#) [PIPELINE PARAMETERS](#) [NODE PROPERTIES](#)**Kubernetes Tolerations** (?)[Add](#)**Shared Memory Size** (?)**Memory Size (GB)** ▼**Kubernetes Pod Labels** (?)[Add](#)**Generic Node Defaults** (?)**Runtime Image** (?)

TensorFlow with CUDA and Python 3.9 (UBI9) ▾

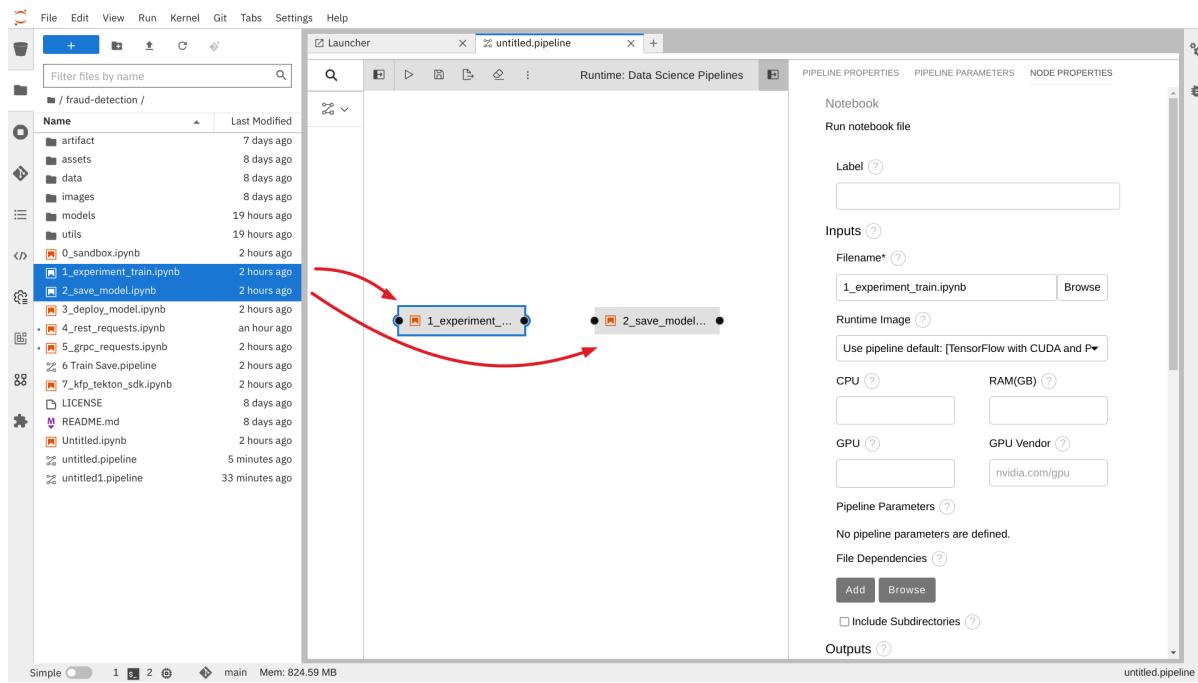
Kubernetes Secrets (?)[Add](#)**Environment Variables** (?)[Add](#)[Refresh](#)**Custom Node Defaults** (?)

4. Save the pipeline.

5.1.2. Add nodes to your pipeline

Add some steps, or **nodes** in your pipeline. Your two nodes will use the **1_experiment_train.ipynb** and **2_save_model.ipynb** notebooks.

1. From the file-browser panel, drag the **1_experiment_train.ipynb** and **2_save_model.ipynb** notebooks onto the pipeline canvas.



- Click the output port of **1_experiment_train.ipynb** and drag a connecting line to the input port of **2_save_model.ipynb**.



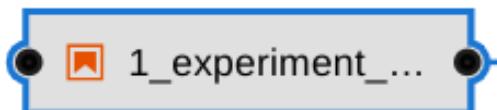
- Save the pipeline.

5.1.3. Specify the training file as a dependency

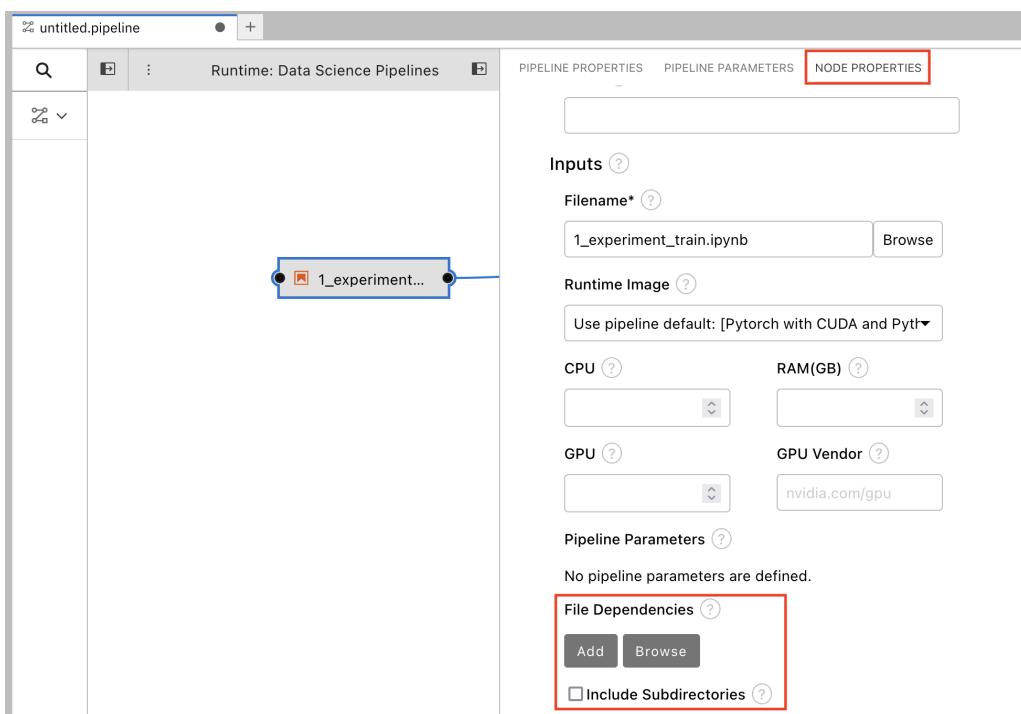
Set node properties to specify the training file as a dependency.

Note: If you don't set this file dependency, the file is not included in the node when it runs and the training job fails.

- Click the **1_experiment_train.ipynb** node.



- In the **Properties** panel, click the **Node Properties** tab.
- Scroll down to the **File Dependencies** section and then click **Add**.



4. Set the value to **data/card_transdata.csv** which contains the data to train your model.
5. Select the **Include Subdirectories** option and then click **Add**.

File Dependencies ?

Add
Browse

Include Subdirectories ?

6. Save the pipeline.

5.1.4. Create and store the ONNX-formatted output file

In node 1, the notebook creates the **models/fraud/model.onnx** file. In node 2, the notebook uploads that file to the S3 storage bucket. You must set **models/fraud/model.onnx** file as the output file for both nodes.

1. Select node 1 and then select the **Node Properties** tab.
2. Scroll down to the **Output Files** section, and then click **Add**.
3. Set the value to **models/fraud/model.onnx** and then click **Add**.

Outputs

Output Files

`models/fraud/model.onnx`

[Remove](#)

[Add](#)

4. Repeat steps 1-3 for node 2.

5.1.5. Configure the data connection to the S3 storage bucket

In node 2, the notebook uploads the model to the S3 storage bucket.

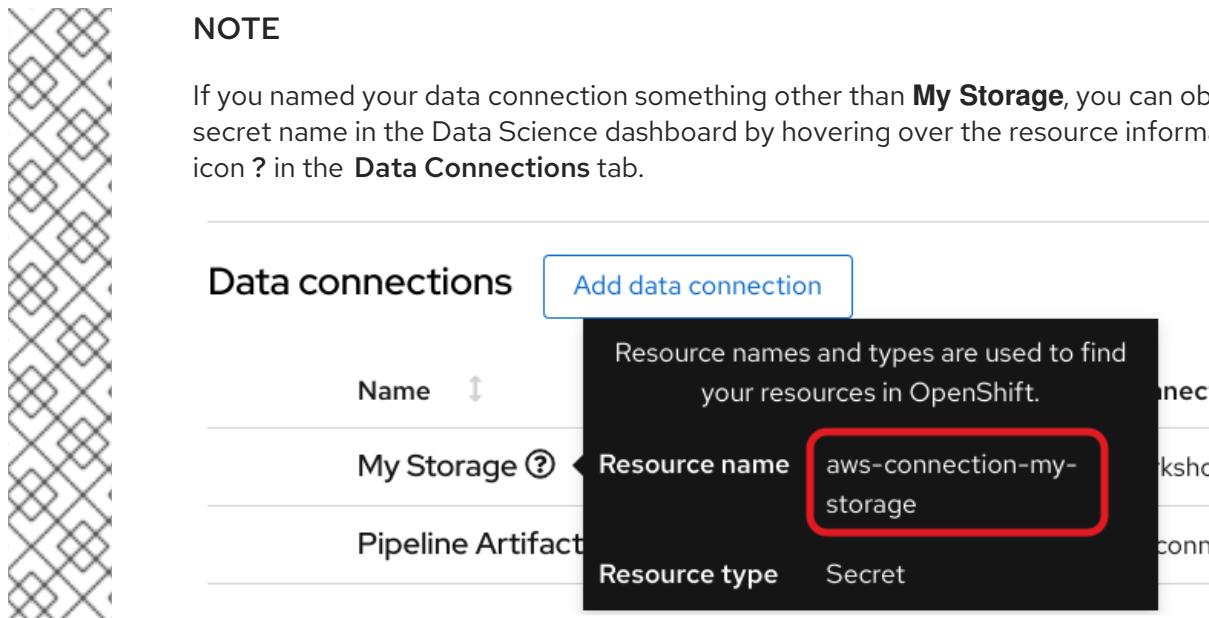
You must set the S3 storage bucket keys by using the secret created by the **My Storage** data connection that you set up in the [Storing data with data connections](#) section of this tutorial.

You can use this secret in your pipeline nodes without having to save the information in your pipeline code. This is important, for example, if you want to save your pipelines – without any secret keys – to source control.

The secret is named **aws-connection-my-storage**.

NOTE

If you named your data connection something other than **My Storage**, you can obtain the secret name in the Data Science dashboard by hovering over the resource information icon ? in the **Data Connections** tab.



Data connections		Add data connection
Name	My Storage 	Resource names and types are used to find your resources in OpenShift.
	Resource name	aws-connection-my-storage
	Resource type	Secret
Pipeline Artifact		

The **aws-connection-my-storage** secret includes the following fields:

- **AWS_ACCESS_KEY_ID**
- **AWS_DEFAULT_REGION**
- **AWS_S3_BUCKET**
- **AWS_S3_ENDPOINT**

- **AWS_SECRET_ACCESS_KEY**

You must set the secret name and key for each of these fields.

Procedure

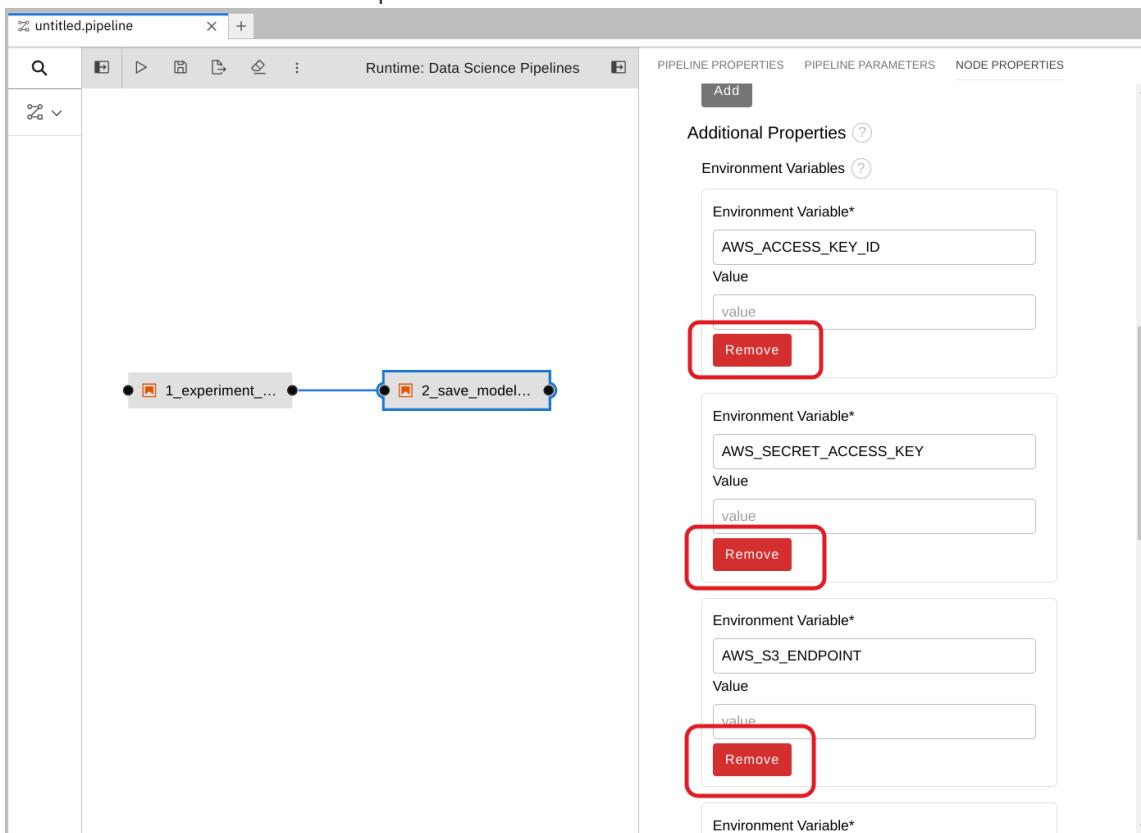
1. Remove any pre-filled environment variables.

- a. Select node 2, and then select the **Node Properties** tab.

Under **Additional Properties**, note that some environment variables have been pre-filled. The pipeline editor inferred that you'd need them from the notebook code.

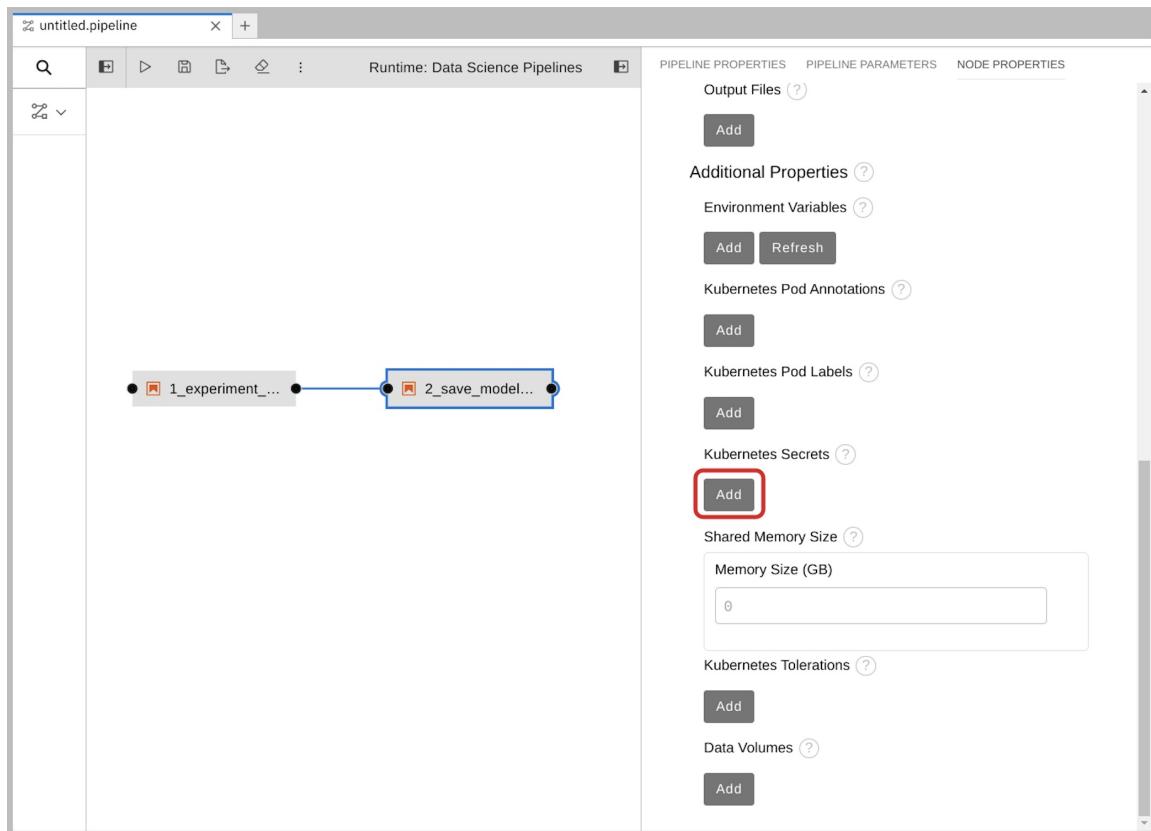
Since you don't want to save the value in your pipelines, remove all of these environment variables.

- b. Click **Remove** for each of the pre-filled environment variables.



2. Add the S3 bucket and keys by using the Kubernetes secret.

- a. Under **Kubernetes Secrets**, click **Add**.



b. Enter the following values and then click **Add**.

- Environment Variable: **AWS_ACCESS_KEY_ID**
- Secret Name: **aws-connection-my-storage**
- Secret Key: **AWS_ACCESS_KEY_ID**

Kubernetes Secrets (?)

Environment Variable*

AWS_ACCESS_KEY_ID

Secret Name*

aws-connection-my-storage

Secret Key*

AWS_ACCESS_KEY_ID

Remove

Add

c. Repeat Steps 3a and 3b for each set of these Kubernetes secrets:

- Environment Variable: **AWS_SECRET_ACCESS_KEY**
 - Secret Name: **aws-connection-my-storage**
 - Secret Key: **AWS_SECRET_ACCESS_KEY**
- Environment Variable: **AWS_S3_ENDPOINT**
 - Secret Name: **aws-connection-my-storage**
 - Secret Key: **AWS_S3_ENDPOINT**
- Environment Variable: **AWS_DEFAULT_REGION**
 - Secret Name: **aws-connection-my-storage**
 - Secret Key: **AWS_DEFAULT_REGION**
- Environment Variable: **AWS_S3_BUCKET**
 - Secret Name: **aws-connection-my-storage**
 - Secret Key: **AWS_S3_BUCKET**

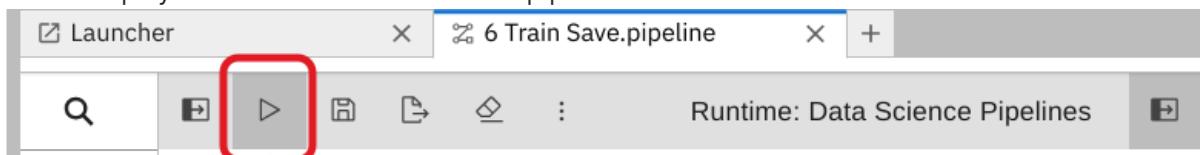
3. Save and Rename the **.pipeline** file.

5.1.6. Run the Pipeline

Upload the pipeline on the cluster itself and run it. You can do so directly from the pipeline editor. You can use your own newly created pipeline for this or **6 Train Save.pipeline**.

Procedure

1. Click the play button in the toolbar of the pipeline editor.

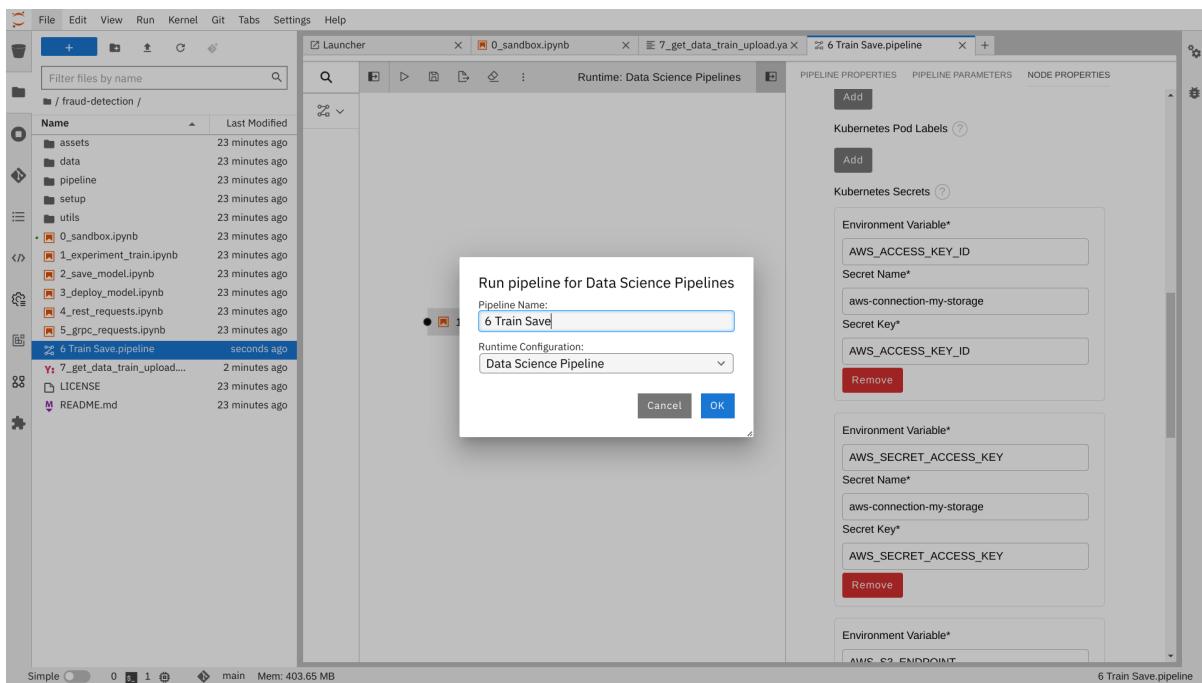


2. Enter a name for your pipeline.
3. Verify the **Runtime Configuration**: is set to **Data Science Pipeline**.
4. Click **OK**.



NOTE

If **Data Science Pipeline** is not available as a runtime configuration, you may have created your notebook before the pipeline server was available. You can restart your notebook after the pipeline server has been created in your data science project.



5. Return to your data science project and expand the newly created pipeline.

Pipelines [Import pipeline](#)

Pipeline name	Last run	Last run status	Last run time	Created
6 Train Save Created with Elyra 3.15.0 pipeline editor using 6 Train Save.pipeline .	6 Train Save-1024141957	✓ Completed	3:10	13 hours ago
Runs	6 Train Save-1024141957	✓ Completed	3:10	13 hours ago

6. Click the pipeline or the pipeline run and then view the pipeline run in progress.

Applications > Runs - Workshop > 6 Train Save-1024141957 Actions

6 Train Save-1024141957 Completed

1_expe...train → 2_save_model

Details Input parameters Run output

Name	6 Train Save-1024141957
Pipeline	6 Train Save
Project	Workshop
Run ID	ac15cab1-29b8-422b-b00d-626e9d8c53eb
Workflow name	6-train-save-ac15c
Created at	Tuesday, October 24, 2023 at 10:20:14 AM Eastern Daylight Time

The result should be a **models/fraud/model.onnx** file in your S3 bucket which you can serve, just like you did manually in the [Preparing a model for deployment](#) section.

Next step

(optional) [Automating workflows with data science pipelines](#)

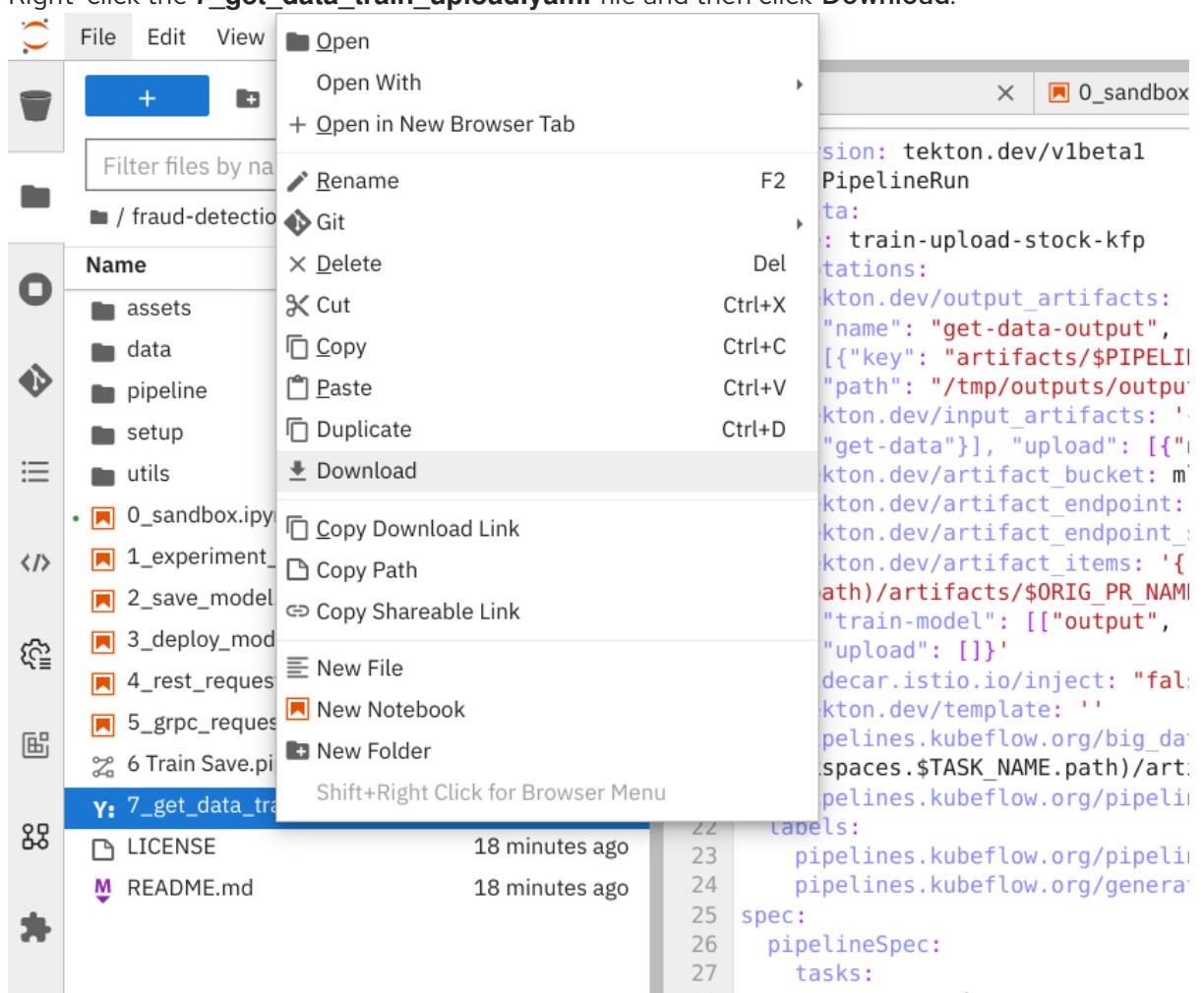
5.2. RUNNING A DATA SCIENCE PIPELINE GENERATED FROM PYTHON CODE

In the previous section, you created a simple pipeline by using the GUI pipeline editor. It's often desirable to create pipelines by using code that can be version-controlled and shared with others. The **kfp-tekton** SDK provides a Python API for creating pipelines. The SDK is available as a Python package that you can install by using the `pip install kfp-tekton~=1.5.9` command. With this package, you can use Python code to create a pipeline and then compile it to Tekton YAML. Then you can import the YAML code into OpenShift AI.

This tutorial does not delve into the details of how to use the SDK. Instead, it provides the files for you to view and upload.

1. Optionally, view the provided Python code in your Jupyter environment by navigating to the **fraud-detection-notebooks** project's **pipeline** directory. It contains the following files:
 - **7_get_data_train_upload.py** is the main pipeline code.
 - **get_data.py**, **train_model.py**, and **upload.py** are the three components of the pipeline.
 - **build.sh** is a script that builds the pipeline and creates the YAML file. The generated **7_get_data_train_upload.yaml** file is located in the **fraud-detection-notebooks** directory.

2. Right-click the **7_get_data_train_upload.yaml** file and then click **Download**.



3. Upload the **7_get_data_train_upload.yaml** file to OpenShift AI.

- In the Data Science dashboard, navigate to your data science project page and then click **Import pipeline**.

The screenshot shows the Red Hat OpenShift AI Data Science dashboard. On the left, there's a sidebar with 'Data Science Projects' selected. Under 'Data Science Pipelines', 'Pipelines' is selected. In the main area, there are sections for 'Workbenches', 'Cluster storage', 'Data connections', and 'Pipelines'. The 'Pipelines' section has a 'Import pipeline' button highlighted with a red box. Below it, a table lists pipelines, including one named '6 Train Save'.

Pipeline name	Last run	Last run status	Last run time	Created
6 Train Save	6 Train Save-I024141957	Completed	3:10	1 day ago

- Enter values for **Pipeline name** and **Pipeline description**.
- Click **Upload** and then select **7_get_data_train_upload.yaml** from your local files to upload the pipeline.

Import pipeline

Project
Fraud Detection

Pipeline name *
Python pipeline

Pipeline description
Pipeline written in Python and converted using kfp-tekton.

```
7_get_data_train_upload.yaml
apiVersion: tekton.dev/v1beta1
kind: PipelineRun
metadata:
  name: train-upload-stock-kfp
  annotations:
    tekton.dev/output_artifacts: '{"get-data": [{"key": "artifacts/$PIPELINERUN/get-data/output.tqz",
```

Import pipeline **Cancel**

- Click **Import pipeline** to import and save the pipeline.
The pipeline shows in the list of pipelines.
- Expand the pipeline item and then click **Create run**.
- On the **Create run** page, enter a **Name**. You can leave the other fields with their default values.

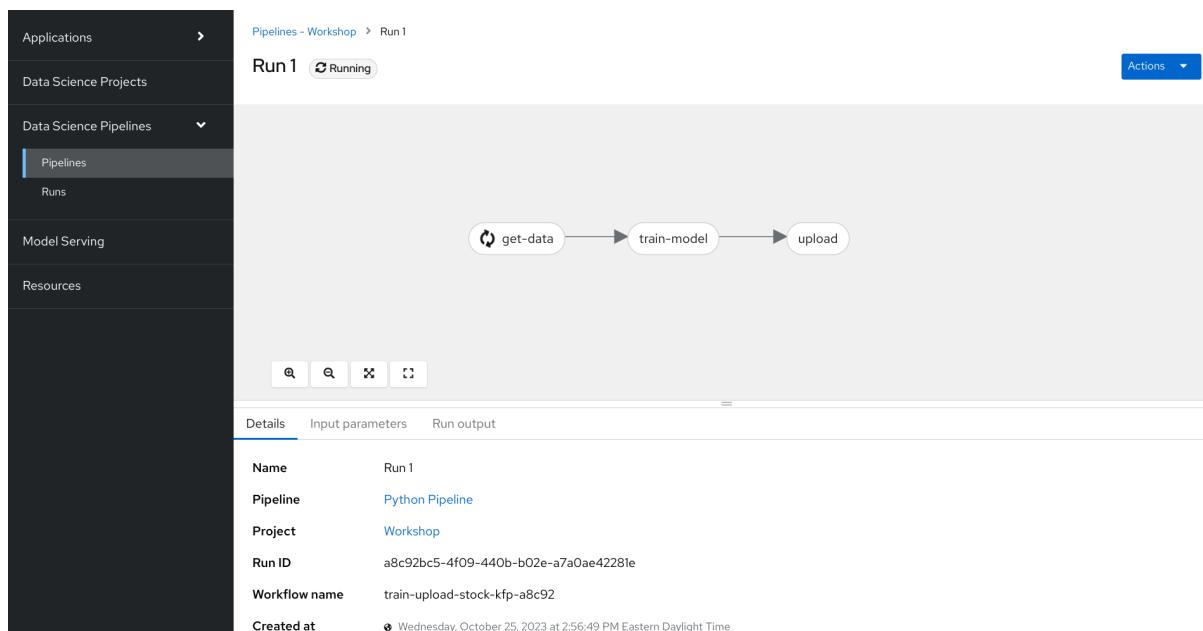
The screenshot shows the 'Pipelines' section of the OpenShift AI interface. On the left sidebar, 'Pipelines' is selected under 'Data Science Pipelines'. The main area is titled 'Run1' and shows the following configuration:

- Name ***: Run1
- Description**: (empty)
- Pipeline**: Python Pipeline
- Run type**: Run once immediately after creation (radio button selected)
- Pipeline input parameters**: (info message: The selected pipeline has no input parameters)

At the bottom are 'Create' and 'Cancel' buttons.

6. Click **Create** to create the run.

A new run starts immediately and opens the run details page.



There you have it: a pipeline created in Python that is running in OpenShift AI.

CHAPTER 6. CONCLUSION

Congratulations!

In this tutorial, you learned how to incorporate data science and artificial intelligence (AI) and machine learning (ML) into an OpenShift development workflow.

You used an example fraud detection model and completed the following tasks:

- Explored a pre-trained fraud detection model by using Jupyter Notebooks.
- Deployed the model by using OpenShift AI model serving.
- Integrated the model into a real-time fraud detection application.
- Refined and trained the model by using automated pipelines.