Hewlett Packard Enterprise

Operationalization for the Machine Learning Lifecycle live demonstration

Contents

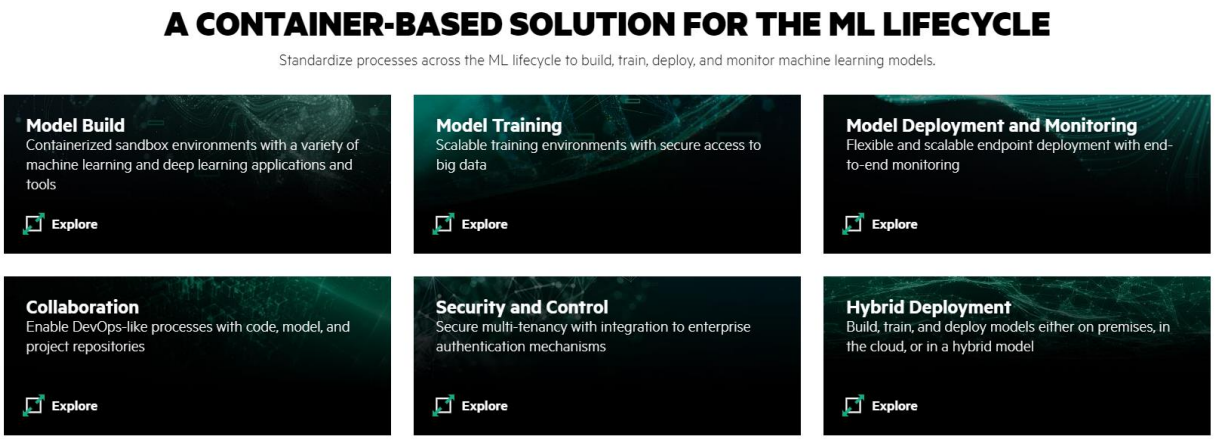
Overview and Introduction

Description and Objective

This demo provides an overview of HPE Machine Learning (ML) Ops solution, powered by HPE servers and HPE Container Platform. HPE ML Ops solution supports every stage of machine learning lifecycle at enterprise scale on containerized distributed clusters — data preparation, model build, model training, model deployment, collaboration, and monitoring.

The demo environment provides access to:

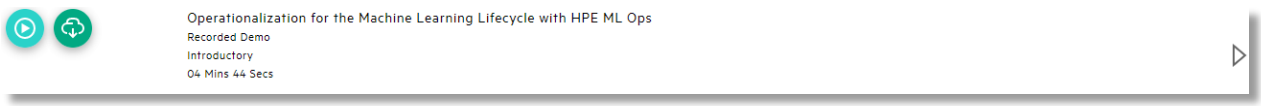
- HPE Container Platform (formerly BlueData EPIC v5.0)



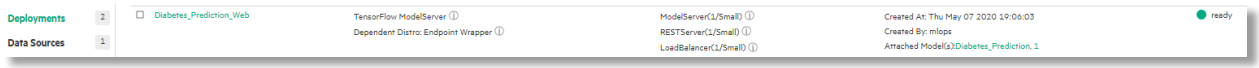
Proceed with caution:

HPE Container Platform (EPIC) demonstration does not include any HA and/or recovery mechanisms to protect against hardware failures and non-reversible actions.

Note: Please use Operationalization for the Machine Learning Lifecycle with HPE ML Ops recorded video as backup solution.



You can stop and cleanup all running cluster(s) excepted Diabetes_Prediction_Web cluster.



How to schedule and access your demo

1. Request your live demo via HPE Demonstration Portal
   www.hpe.com/demos/hpedemos

2. Download demonstration guide

3. **launch** it now or **Schedule** your demo

LAUNCH DEMO

Gray out when demo is in use.
Need to wait or schedule for later

(*) use **Schedule** option to ensure demo availability at the time of your customer event.

Option1: Launch demo now (not available for all demos)

Launch Demo Now

1. Launch your Demo after having specified demo type:

Customer Event
Training / Classroom
Dry Run / Self-Paced Instruction

2. Demo environment is getting configured (wait a few seconds)

OnDemand – TEST-OnDemandDemoviaCitrix
Please wait for some time , getting demo environment ready

3. Citrix demo session opens automatically

* Citrix receiver plugin to be installed the first time

4. Demo is also available from HPE Demo Portal User dashboard

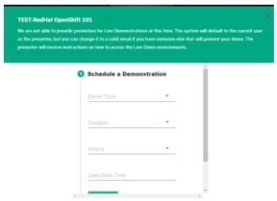
Homepage
User Dashboard
Site Catalog
Get Help

My Upcoming Demonstration(s)

Event ID: 13669
Event Subject: HPE ProLiant for Azure Stack Live Tour (#9009)
When: Tue May 15 2018 17:00:00 GMT+0200 (Romance Daylight Time)
Duration: 1 Hour(s)
Status: Approved

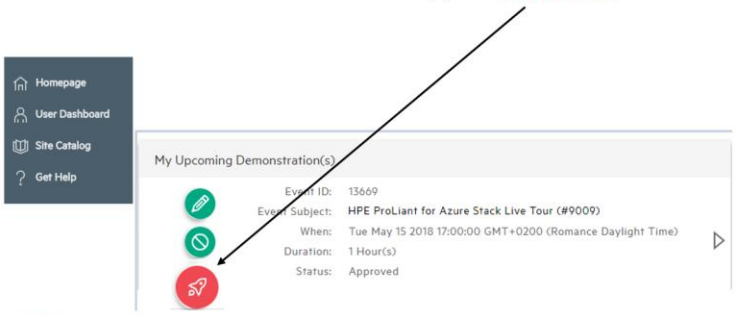
Hewlett Packard Enterprise

Option2: Scheduled a demonstration



1. Scheduled Demo will be approved and will be ready at the time you choose:
   Presenter Access Instructions email is sent (15 minutes prior the demo)
   with explanation how to connect.

2. Demo will directly be accessible from HPE Demo Portal User dashboard
   Launch button will appear **at the time** of the demo



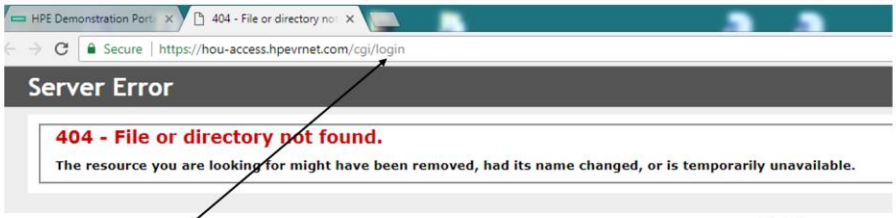
4. Run your demo (open Citrix session)
   * Citrix receiver plugin to be installed the first time

Hewlett Packard Enterprise

Optional

In case of problem …

When trying to launch either too early or a second time a demo via launch button the following error may occur (404 error or Invalid credentials).



4. Run your demo (open Citrix session)
   * Citrix receiver plugin to be installed the first time

Change and run URL, replacing « login » by logout

https://xxx-access.hpevrnet.com/cgi/logout

And launch again your demo from « my upcoming demonstrations »
In HPE Demo Portal User dashboard :

Hewlett Packard Enterprise

Configuration Diagram

Target Audience

All potential existing or new customers that are looking for solution to deploy distributed AI, machine learning (ML), and big data analytics running on Docker containers either on-premises or in the cloud..

HPE field personal and/or partners that need hands-on experience with HPE ML Ops platform will also find it valuable.

Requirements for Presenters

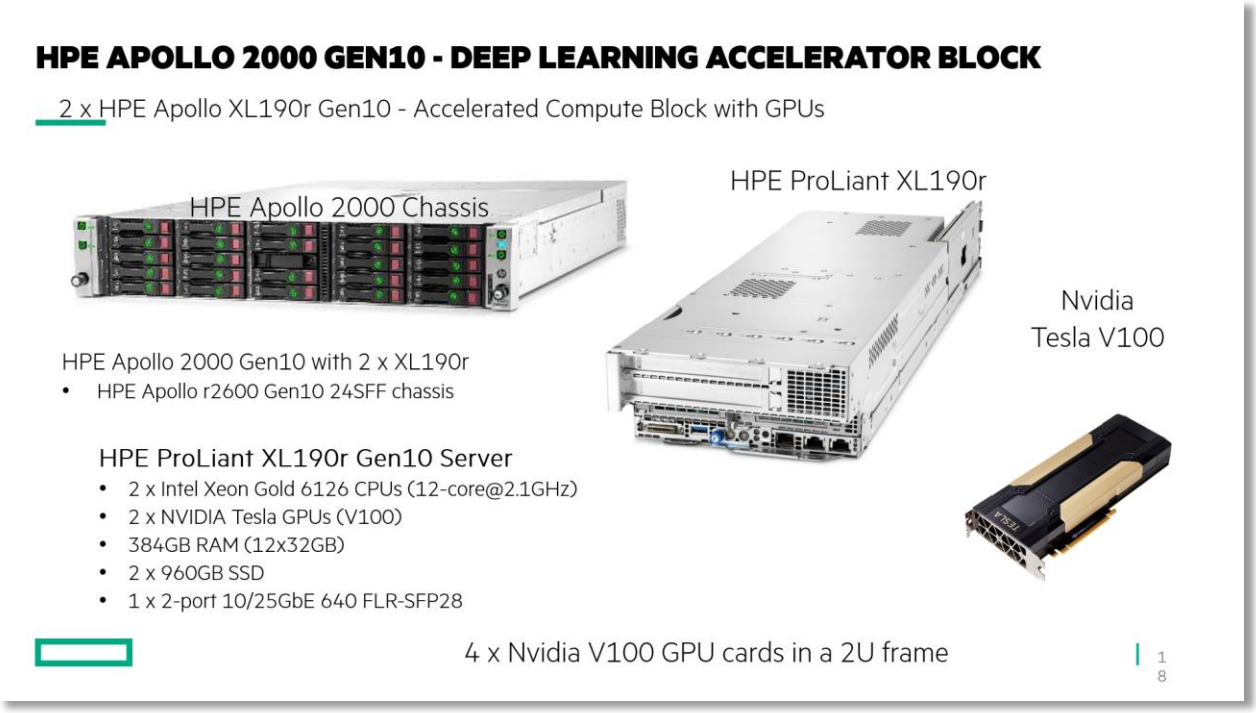
Presenters should have some knowledge of HPE Container Platform (EPIC) software including Elastic Platform for Big Data Analytics servers.

The HPE Advantage

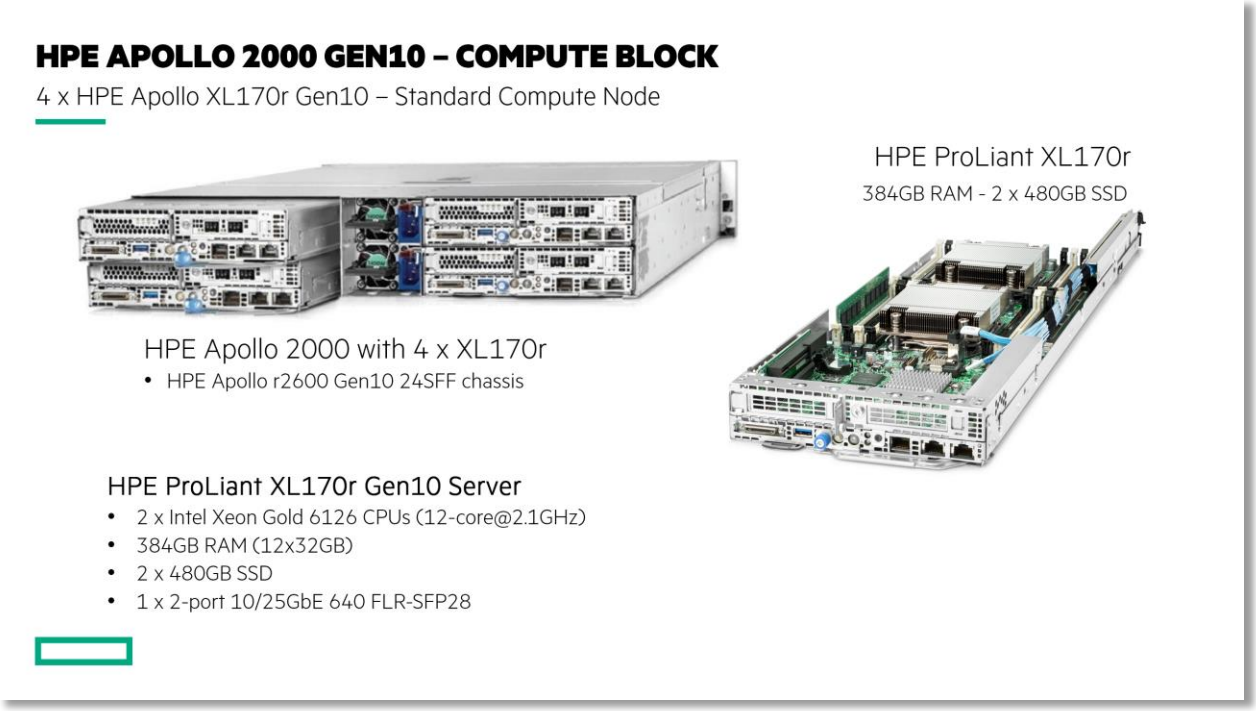
HPE acquires leading provider of container-based software solution that offers an as-a-service experience for AI, machine learning, and analytics in the enterprise.. We believe this differentiated hardware EPA + software solution offer will help IT customers deploy BigData Analytics and AI in minutes to support more efficiently data science teams deriving insights from data, instead of waiting for infrastructure.

HPE EPA demonstration platform Compute Blocks

HPE Deep Learning Accelerator Block with Nvidia GPU:



HPE Apollo 2000 Gen10 Standard Compute Block:



Demo Use Cases

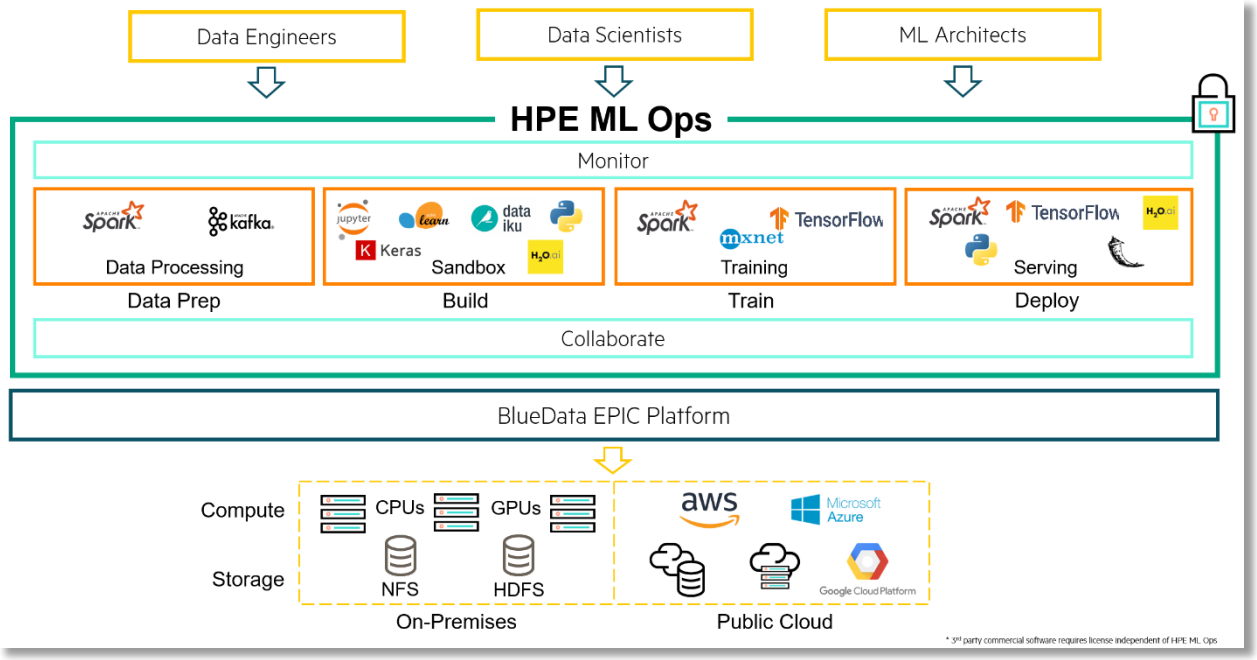
Operationalization for the Machine Learning Lifecycle

Much like pre-DevOps software development, data science organizations still spend a significant amount of time and effort when moving projects from development to production. Model version control and code sharing is manual and there is a lack of standardization on tools and frameworks, making it tedious and time-consuming to productize machine learning models.

HPE Machine Learning Ops (HPE ML Ops) extends the capabilities of the HPE Container Platform and brings DevOps-like agility to enterprise machine learning. With the HPE ML Ops solution, enterprises can implement DevOps processes to standardize their ML workflows.

HPE ML Ops provides data science teams with a platform for their end-to-end data science needs with the flexibility to run their machine learning or deep learning (DL) workloads on-premises, in multiple public clouds, or a hybrid model and respond to dynamic business requirements in a variety of use cases.

HPE ML Ops solution Architecture



Access HPE ML Ops demonstration platform as mlops user

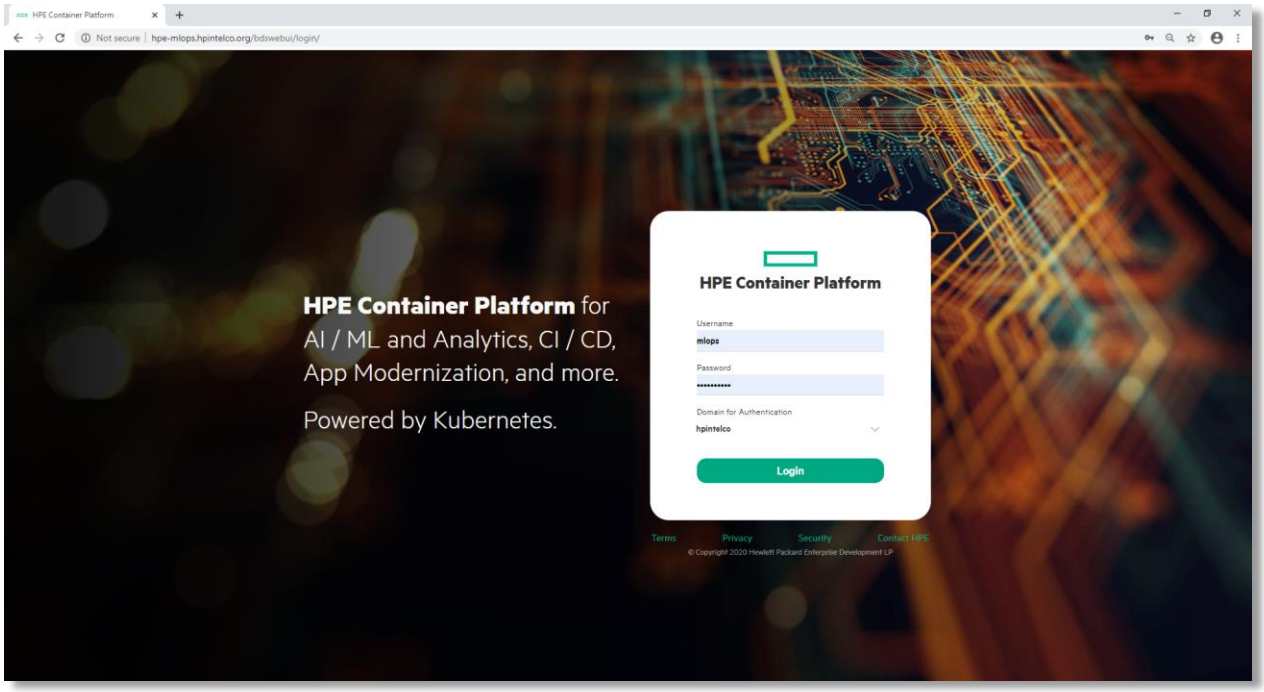
Environment	Web UI	Username
HPE ML Ops	https://hpe-mlops.hpintelco.org	mlops

Once connected to the Jumphost station, open in your browser (chrome) the web ui login console, fill the logon credentials provided on the BlueData_EPIC html page and click “Login”.

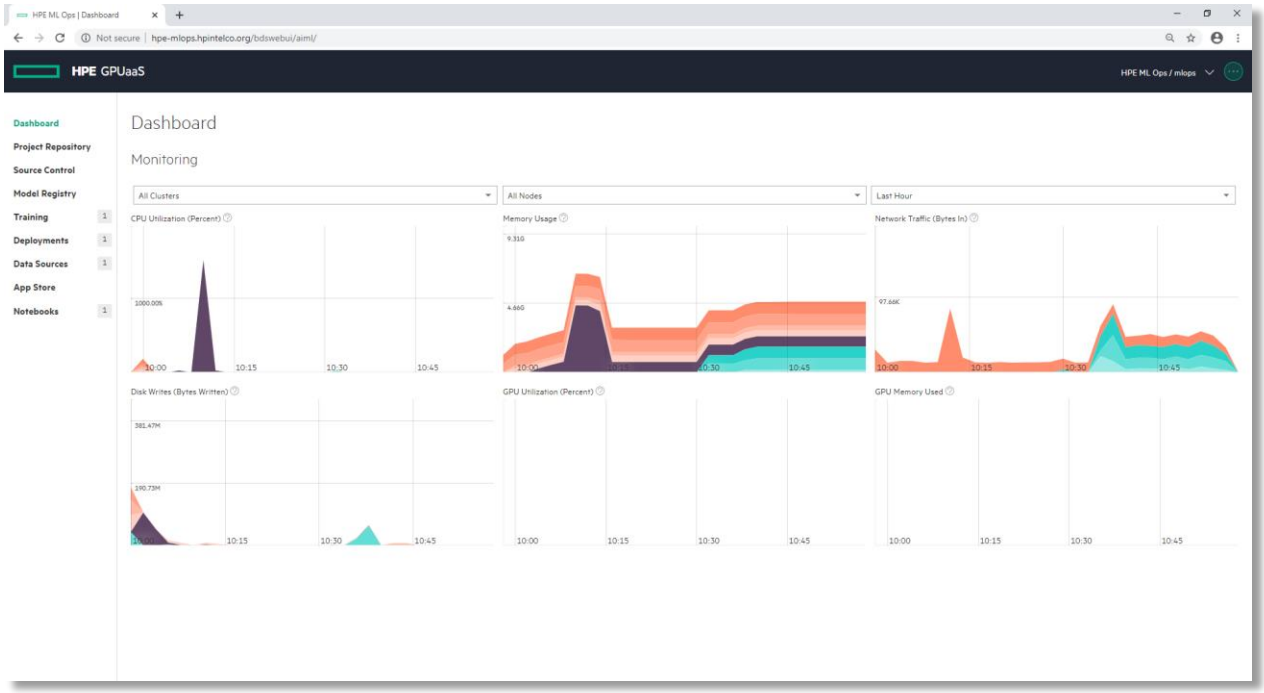
This link will start Chrome and open an html document with links and credentials!

Discover HPE ML Ops platform capabilities

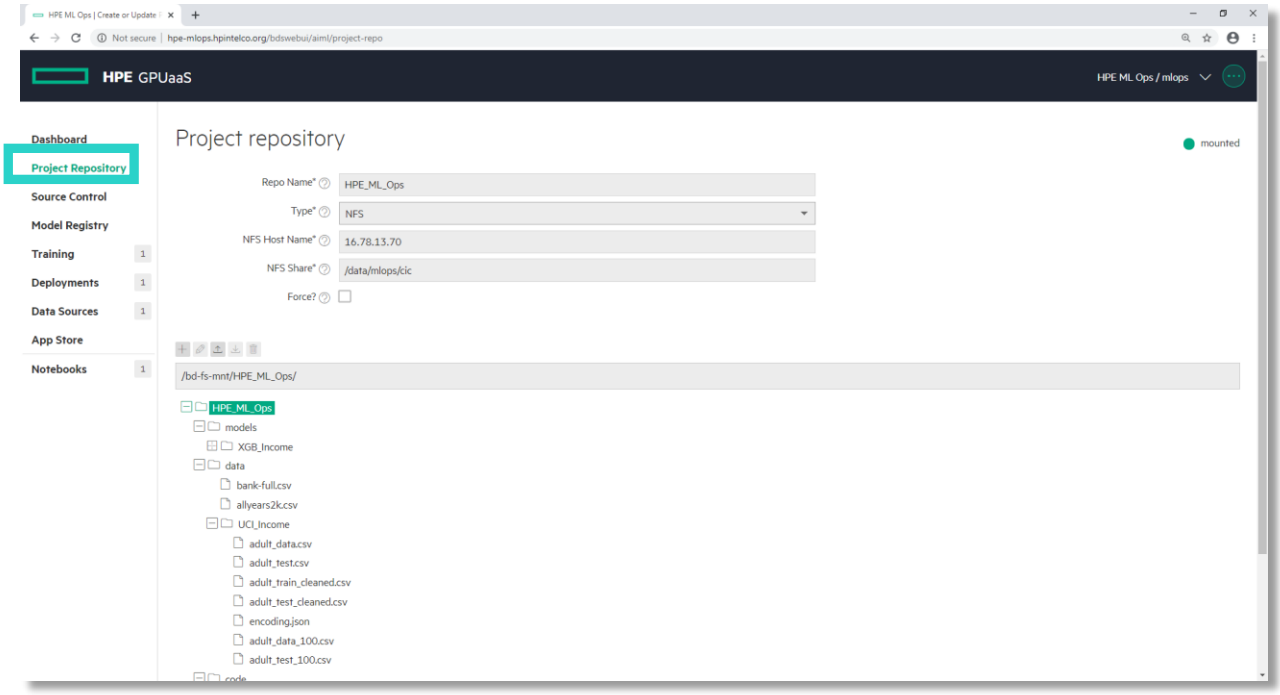
1. HPE Container Platform Login Page: Login in to the WebUI as **mlops** user: http://hpe-mlops.hpintelco.org



2. HPE Container Platform User Dashboard: See the monitoring

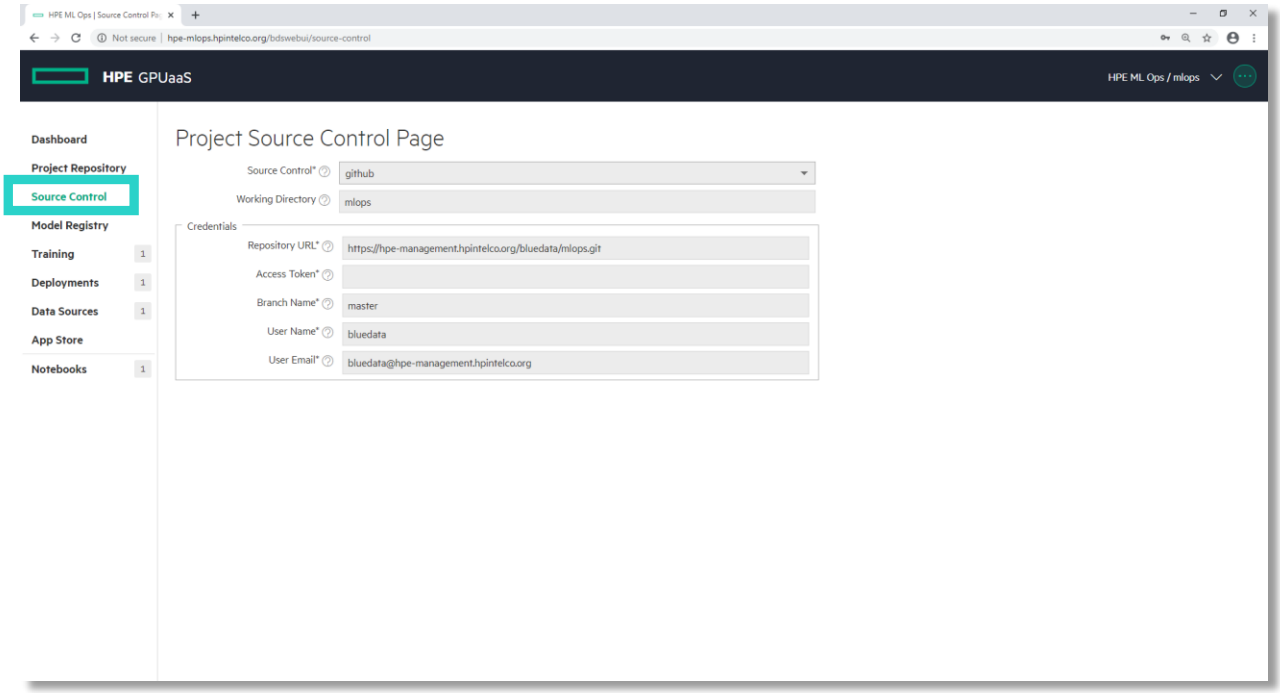


3. Select Project Repository from the left menu



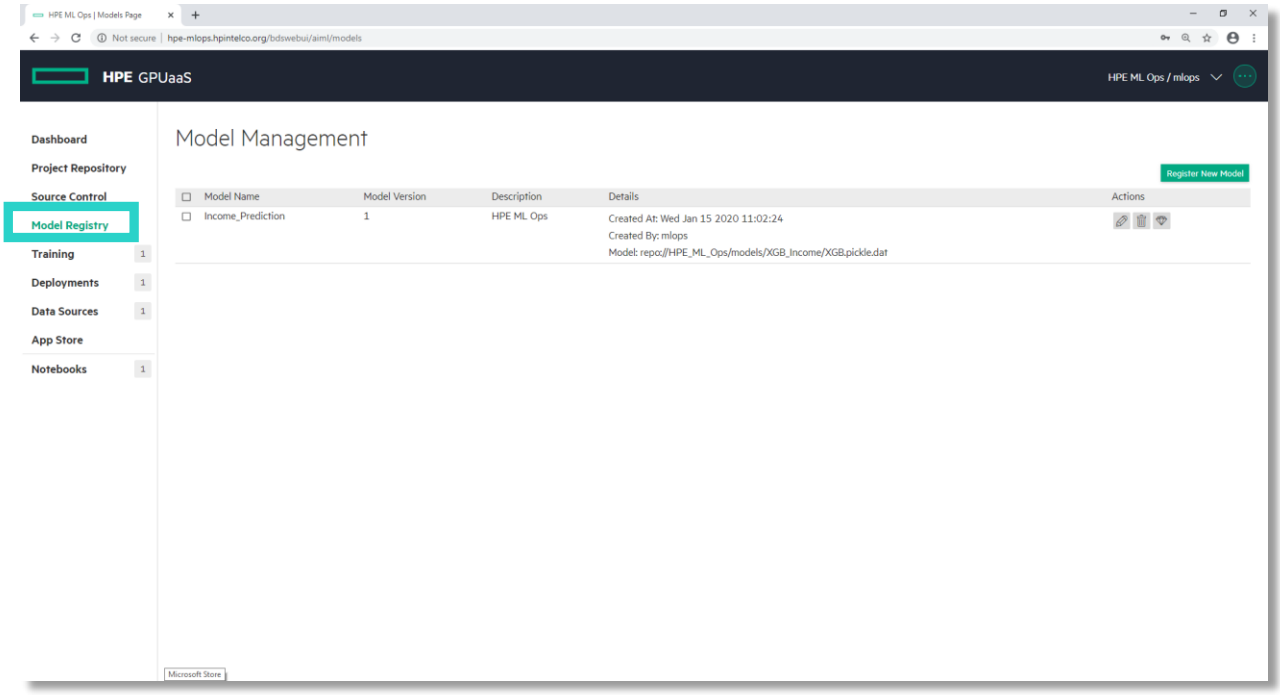
The **Project Repository** screen displays information about the project repository that was created by the Project Administrator for the current AI/ML project.

4. Select Source Control from the left menu



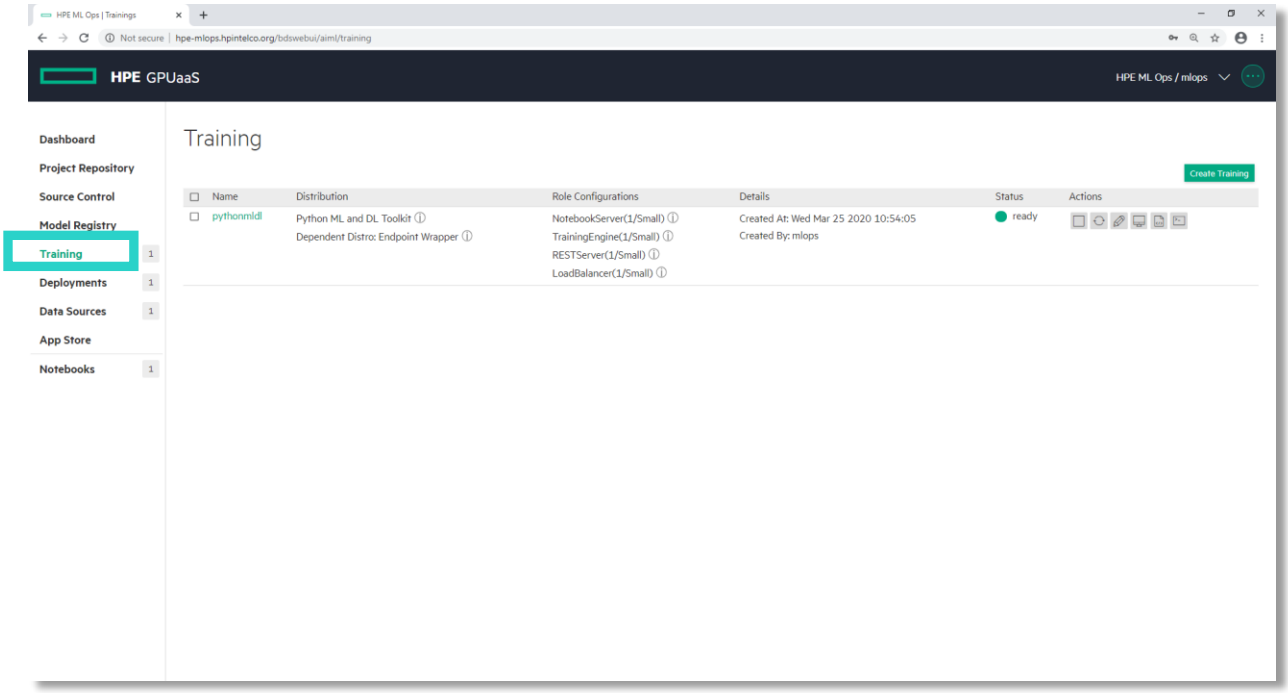
The **Project Source Control Page** screen allows you to specify a cloud-based source control service to hold notebooks for this project. Notebook clusters in this AI/ML project will retrieve notebooks from the source control location specified on this page.

5. Select Model Registry from the left menu



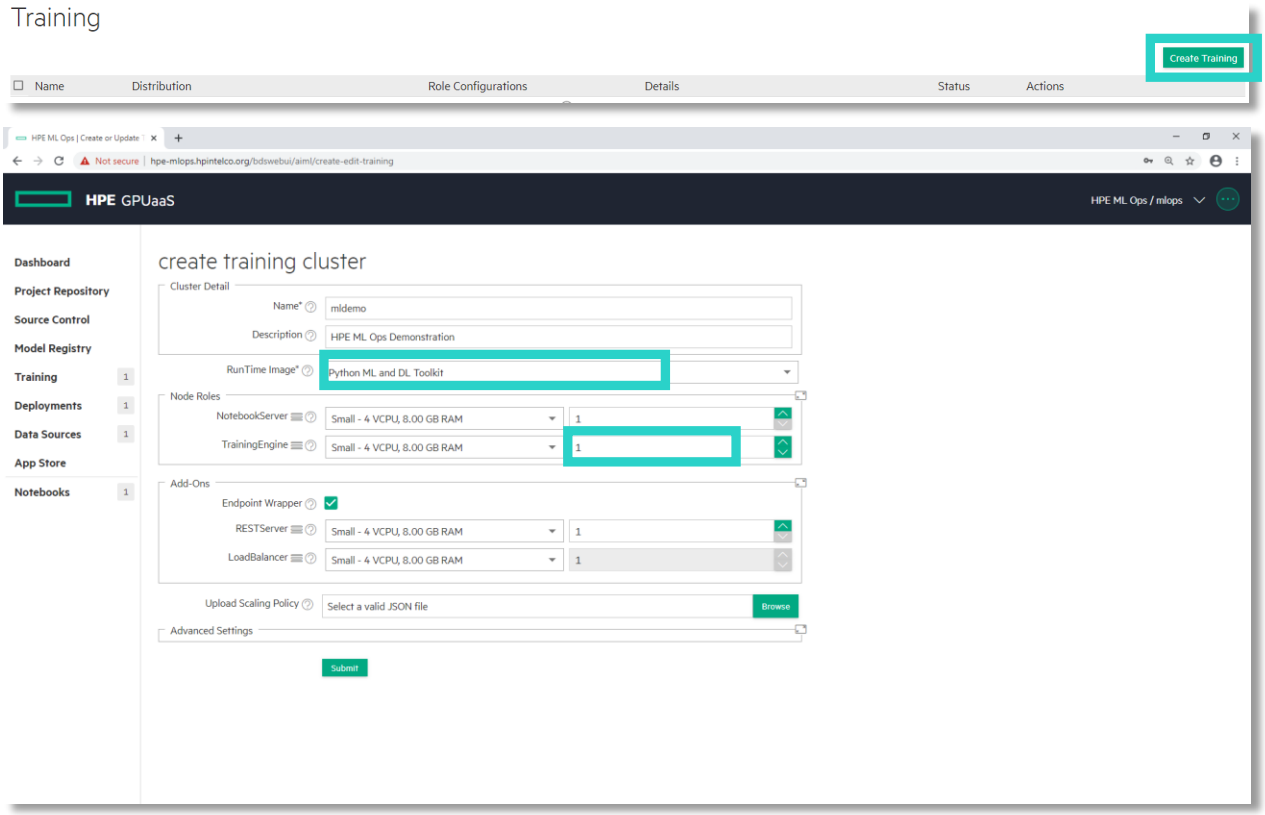
The **Model Management** screen displays all of the models that have been registered for the current project.

6. Select Training from the left menu

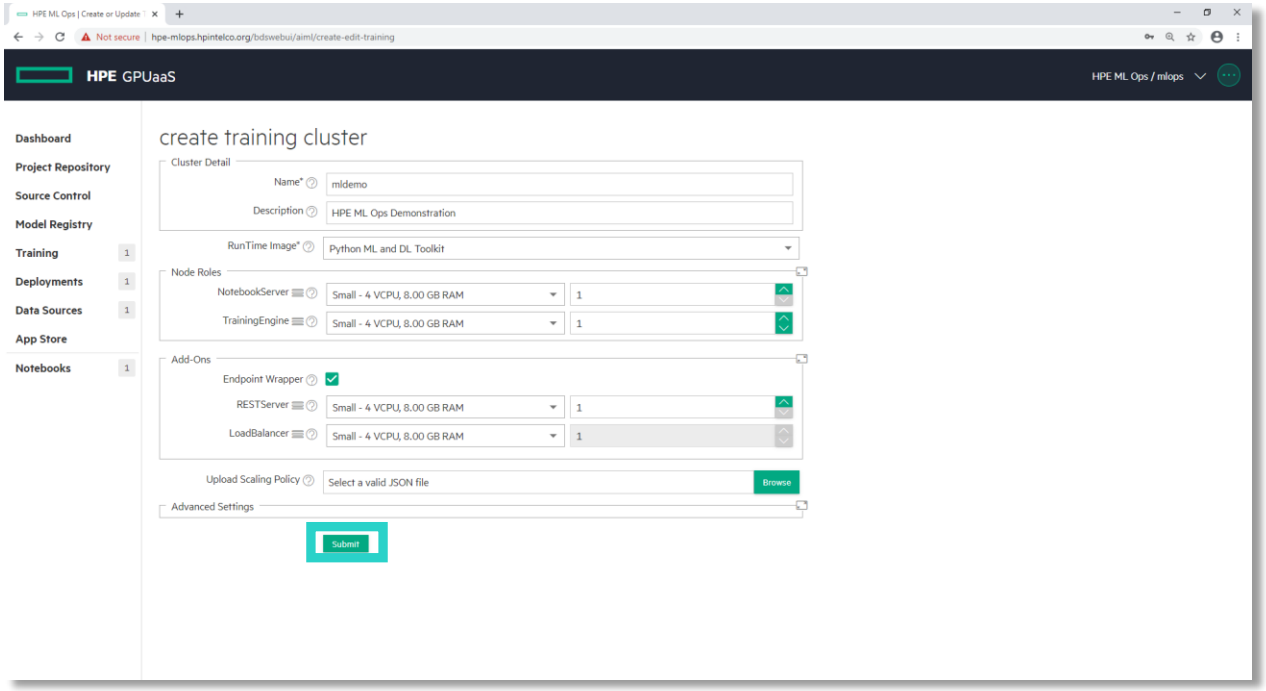


Create your HPE ML Ops Training cluster

1. Select Create Training

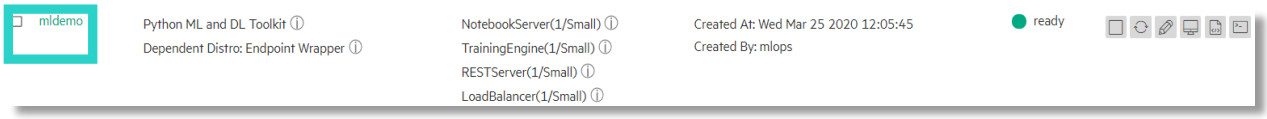


Don't forget to add a Training Engine and update the RunTime Image and Click on Submit

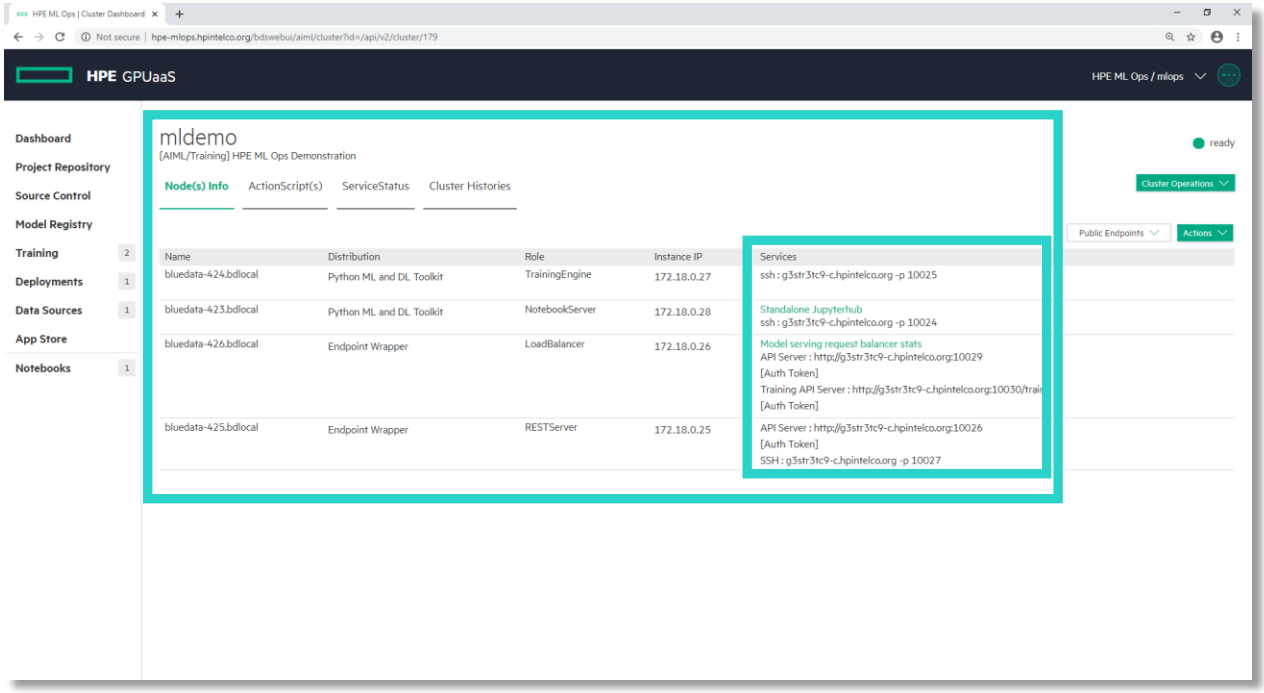


2. Check that the cluster is ready

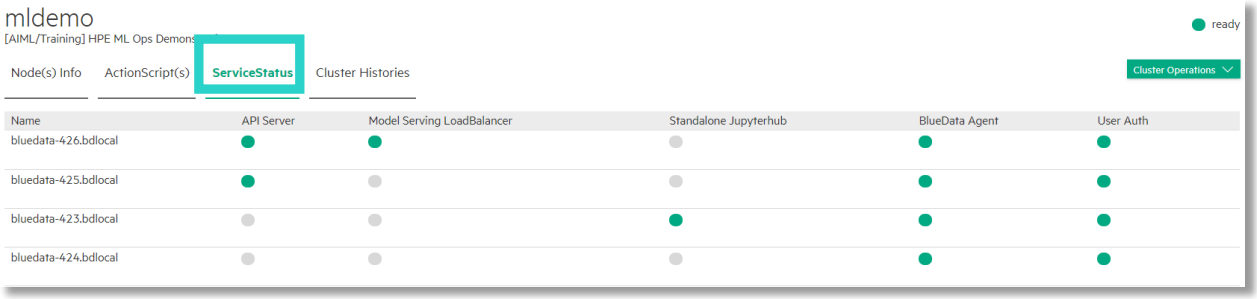
3. Open Training Cluster details (Click on mldemo)



4. Show the cluster nodes and services. (NotebookServer / TrainingEngine / …)

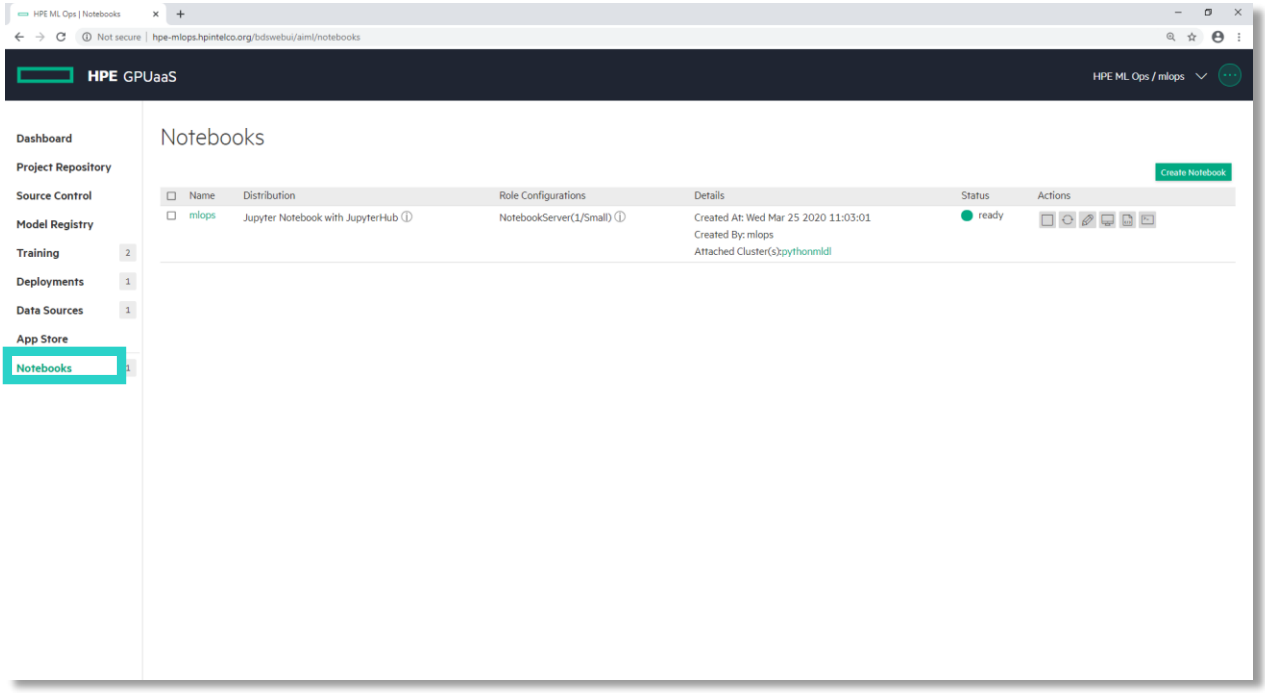


5. Check the Services Status of your cluster (Application / System services)

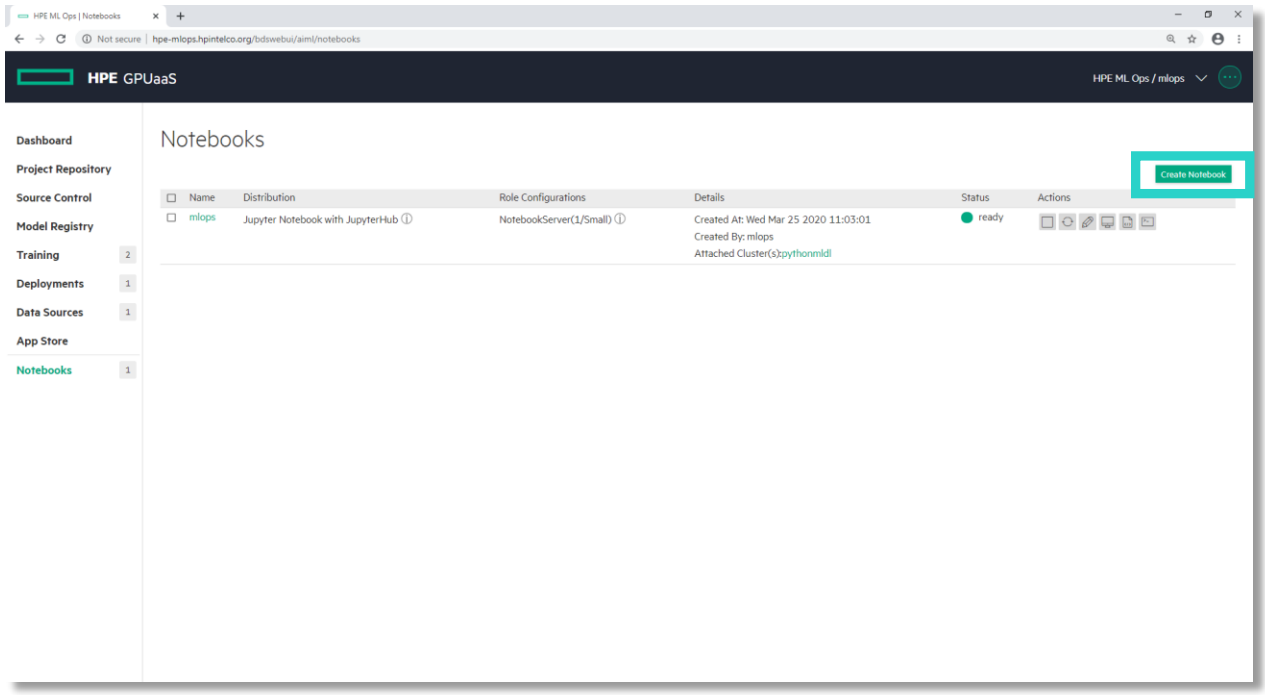


Develop/Create Jupyter Notebooks

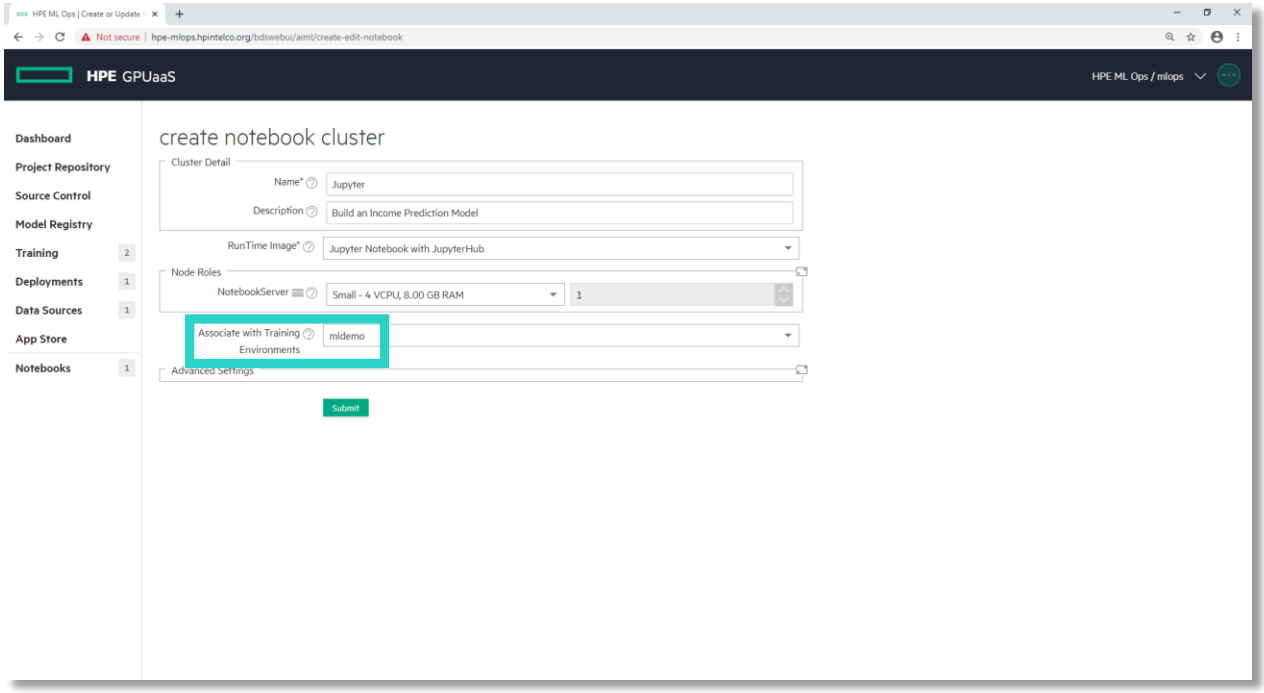
1. Select Notebooks from the left menu



2. Create a Jupyter Notebook (Click on Create Notebook)

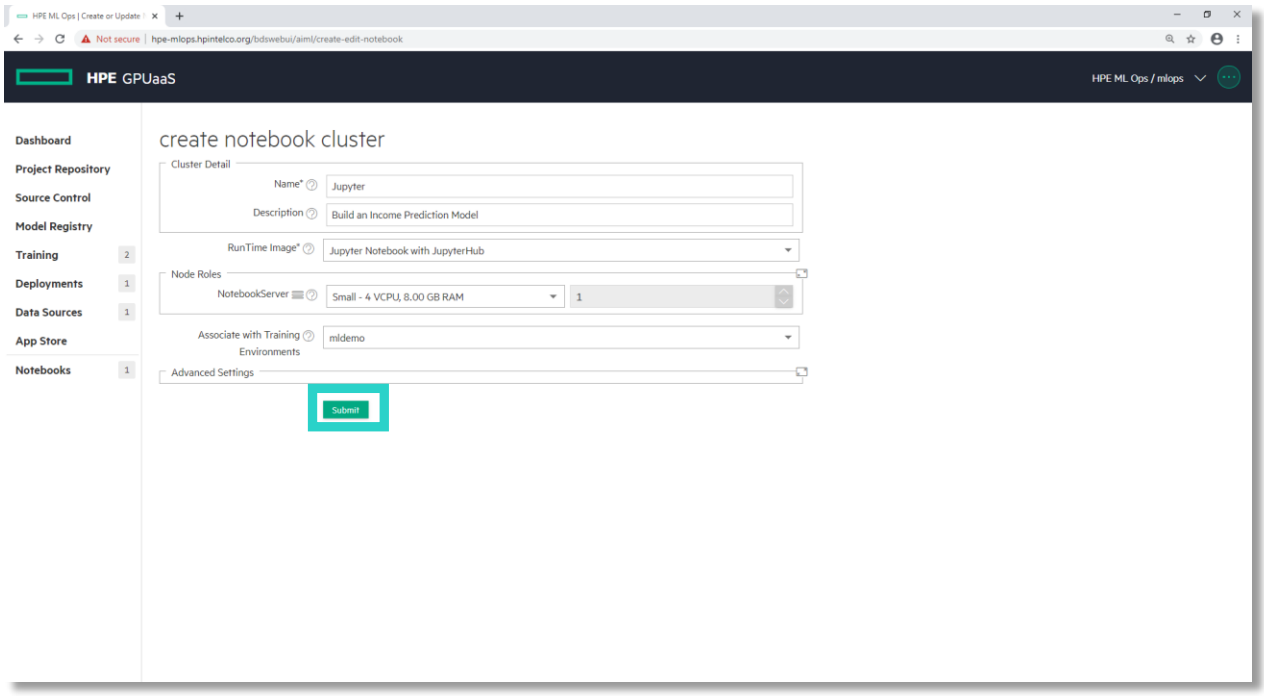


3. Associate the Notebook with your Training cluster

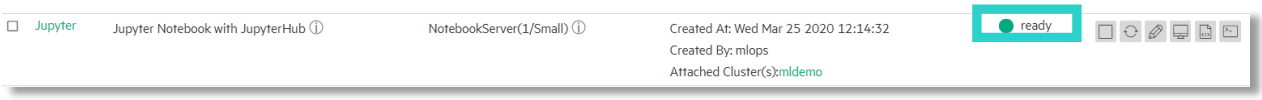


Don't forget to select Training Engine, setup flavor and update the Jupyter RunTime Image

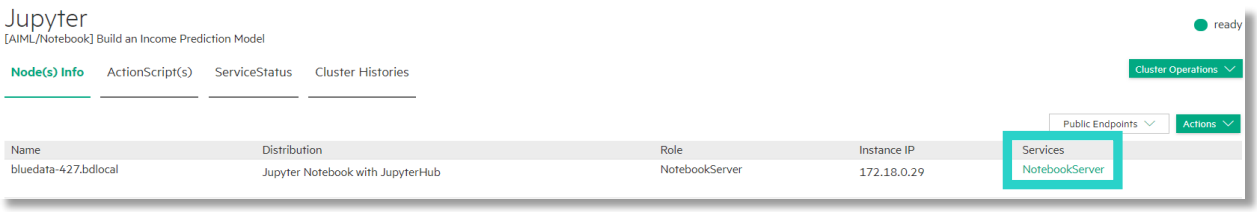
4. Create Notebook Cluster (Click on Submit)



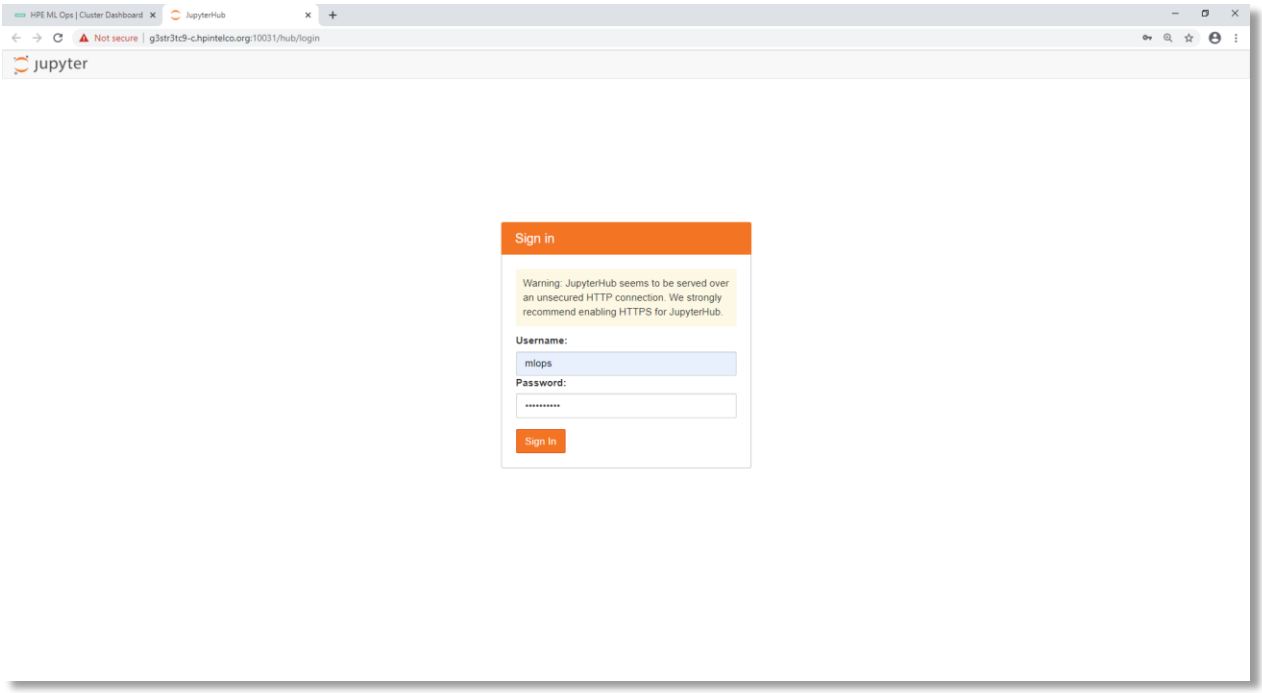
5. Make sure that the status of your notebook is ready



6. Open JupyterHub web ui (Click on NotebookServer)



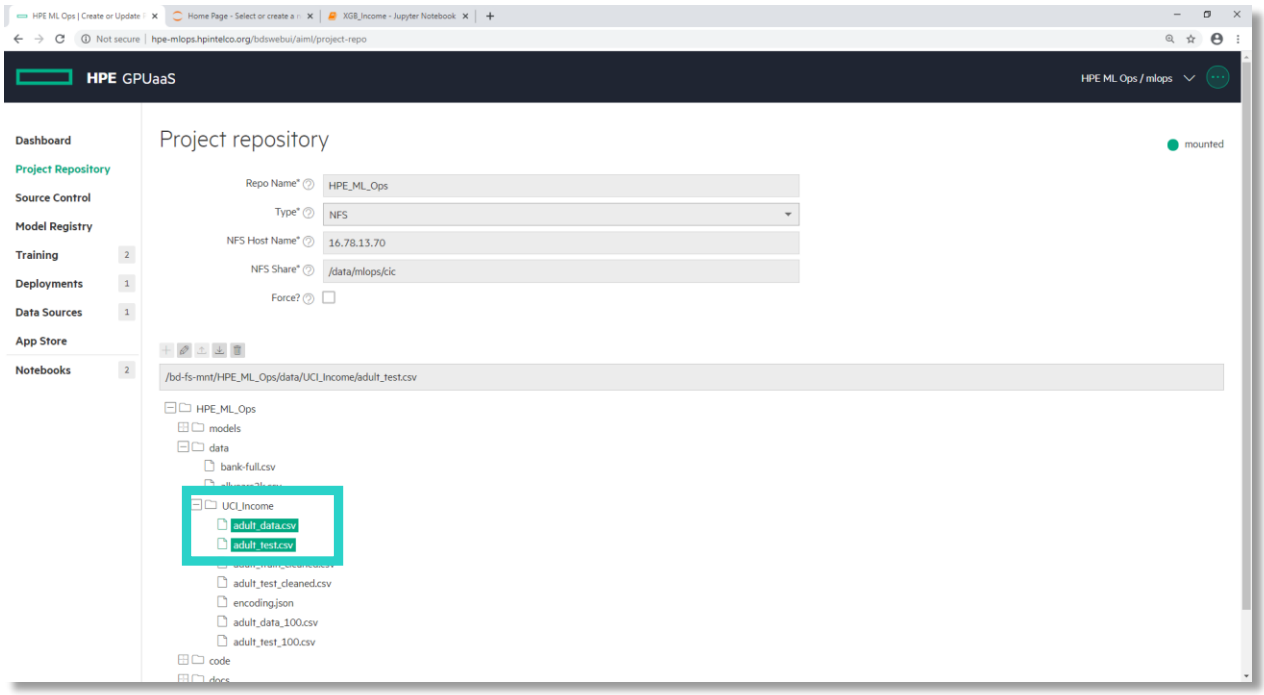
7. Login in to the Jupyter WebUI as **mlops** user



Income Prediction Demo Scenario

This tutorial uses a model to classify a person's income as being either less than or equal to $50,000 or more than $50,000.

Data set is available from https://archive.ics.uci.edu/ml/datasets/Adult. This data set is a spreadsheet with approximately 32,000 rows of training data that was acquired from the 1994 Census database.



Data set is in the Project Repository

The features (columns) in this spreadsheet that are used to train the model are:

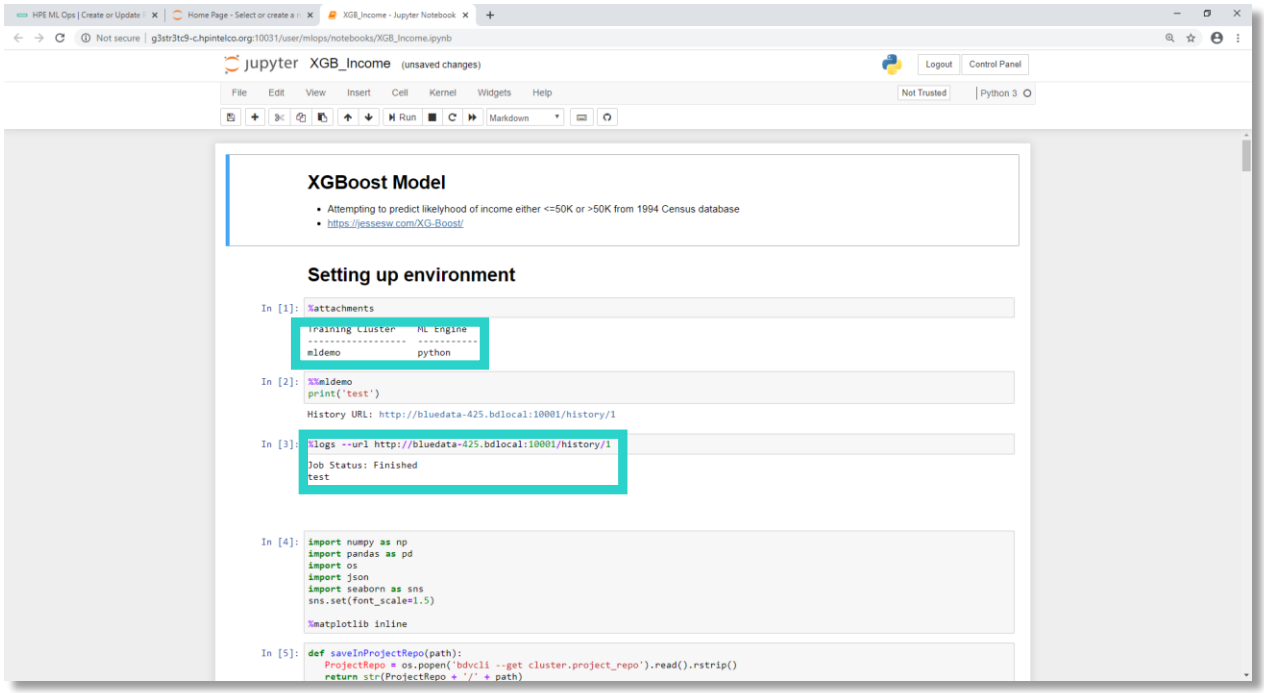
- age
- workclass
- fnlwgt (the number that the census believes represents the population)
- education
- education_num (number representation of education)
- marital_status
- occupation
- relationship
- race
- sex
- capital_gain
- capital_loss
- hours_per_week
- native_country

The last column indicates the income classification of that individual.

Load XGB_Income.ipynb notebook

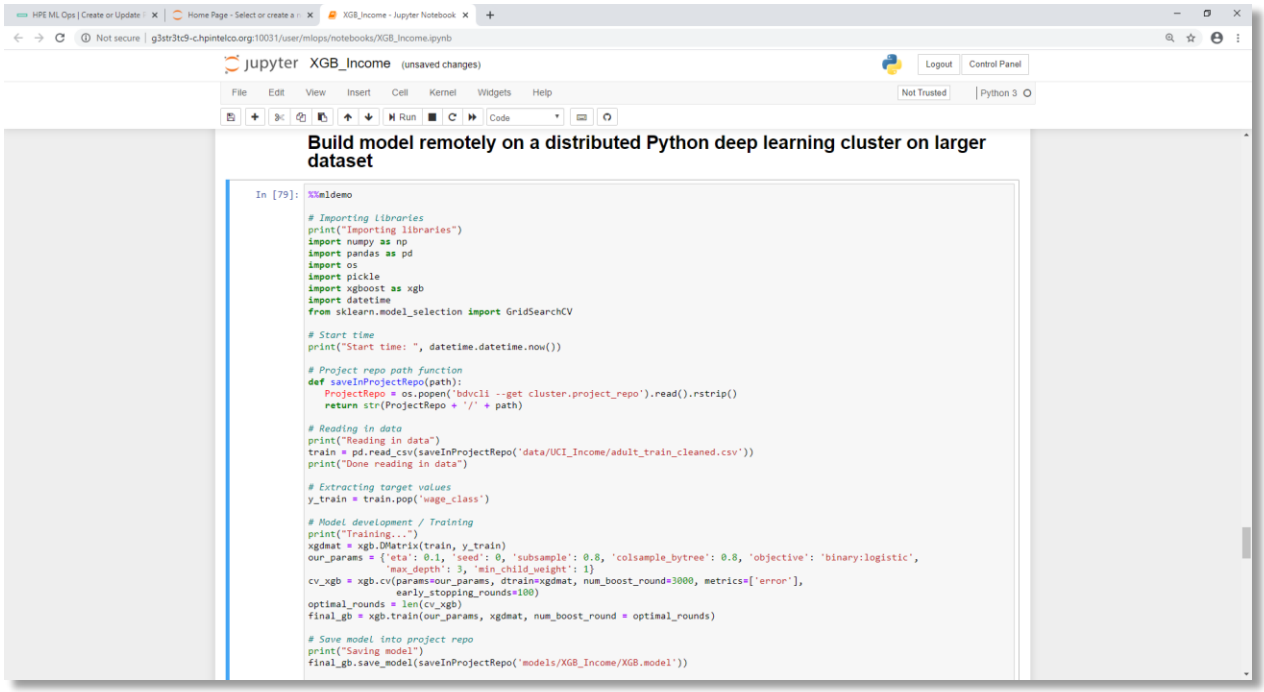
1. Load XGB_Income.ipynb notebook

Open the `XGB_Income.ipynb` file, and then run each cell individually. This generates cleaned .csv files, an encoded file, and model files. The Notebook contains detailed comments and explanations.



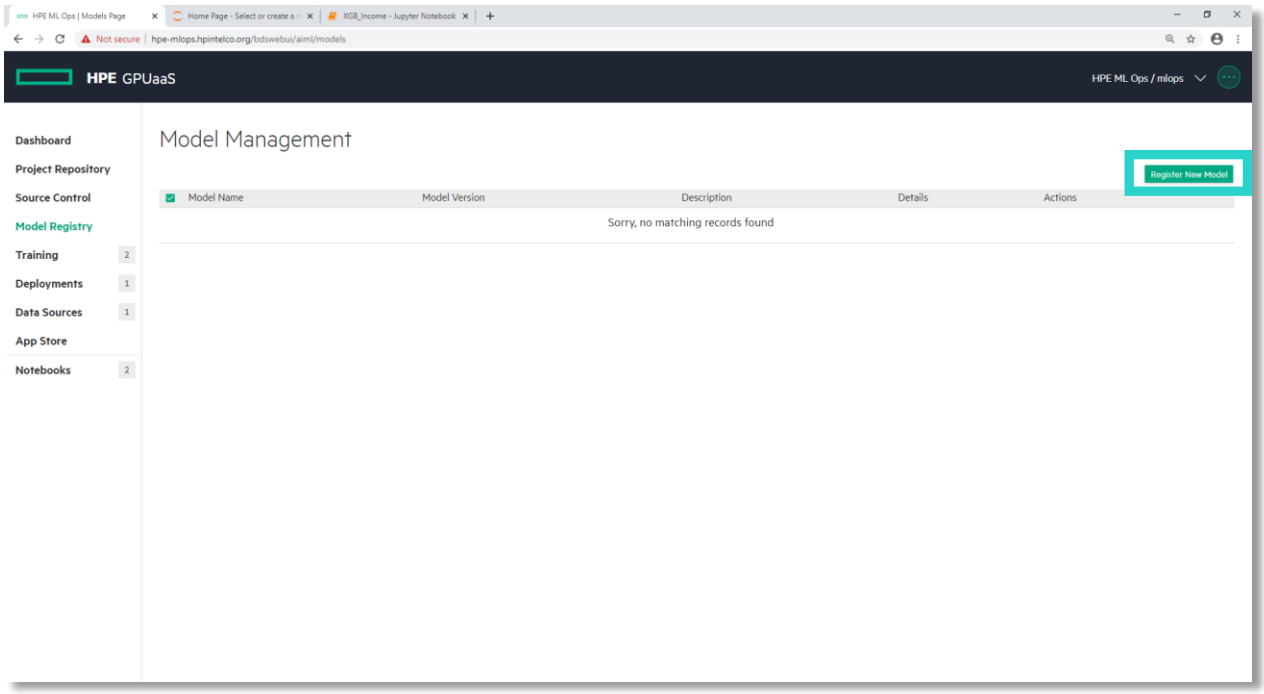
(Caution: Update the variable according to your Training Cluster including History urls.)

Once the model is tuned to the optimal parameters, run the cell on a remote Training cluster (where there are potentially more resources to train a model on a larger dataset).

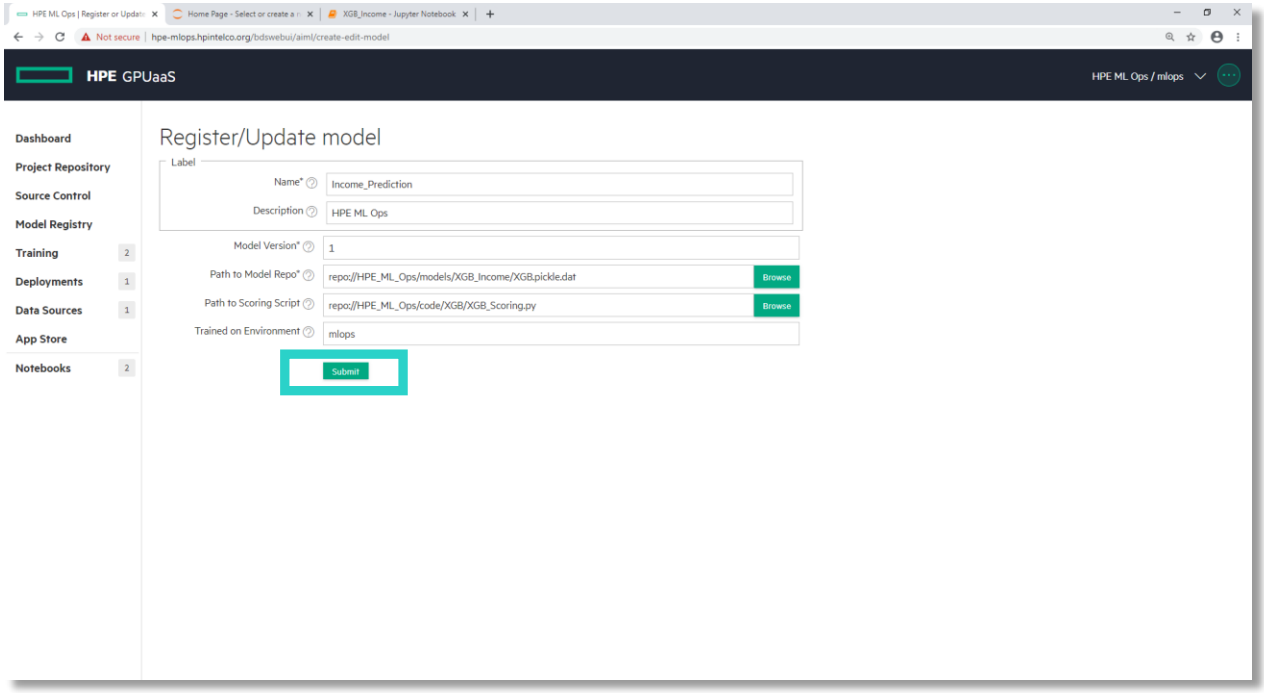


Register and Deploy the Model

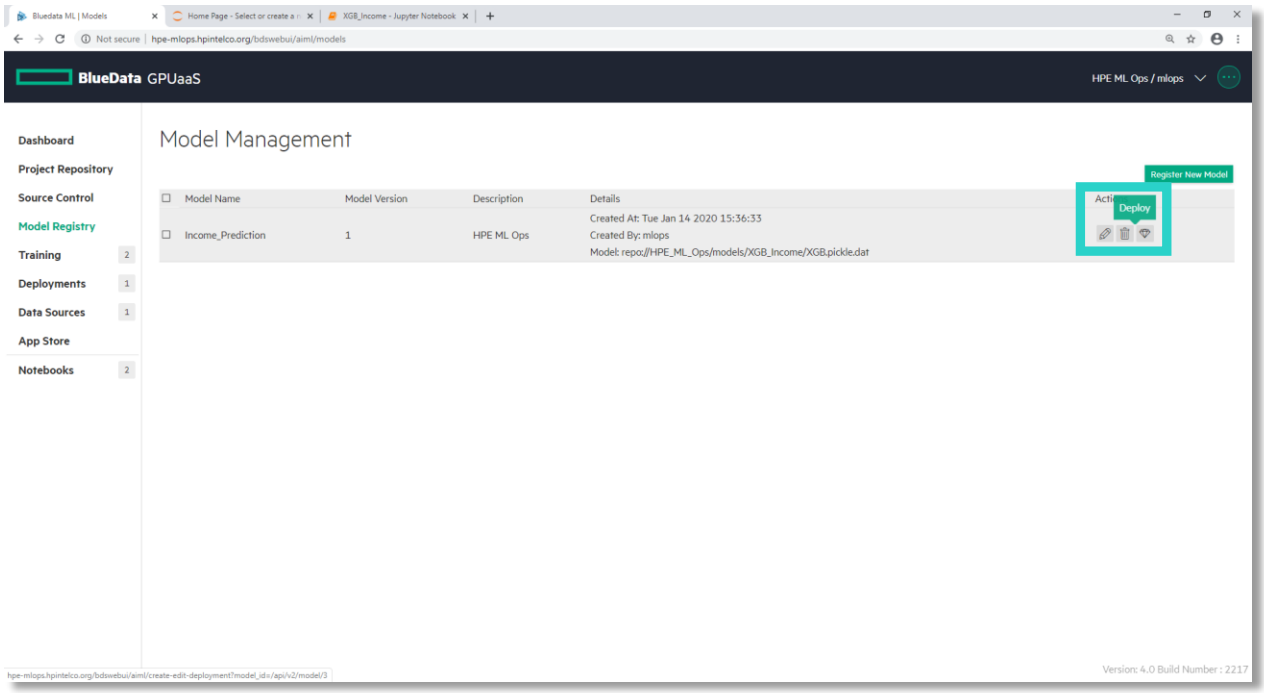
1. Select Model Registry from the left menu and click on Register New Model



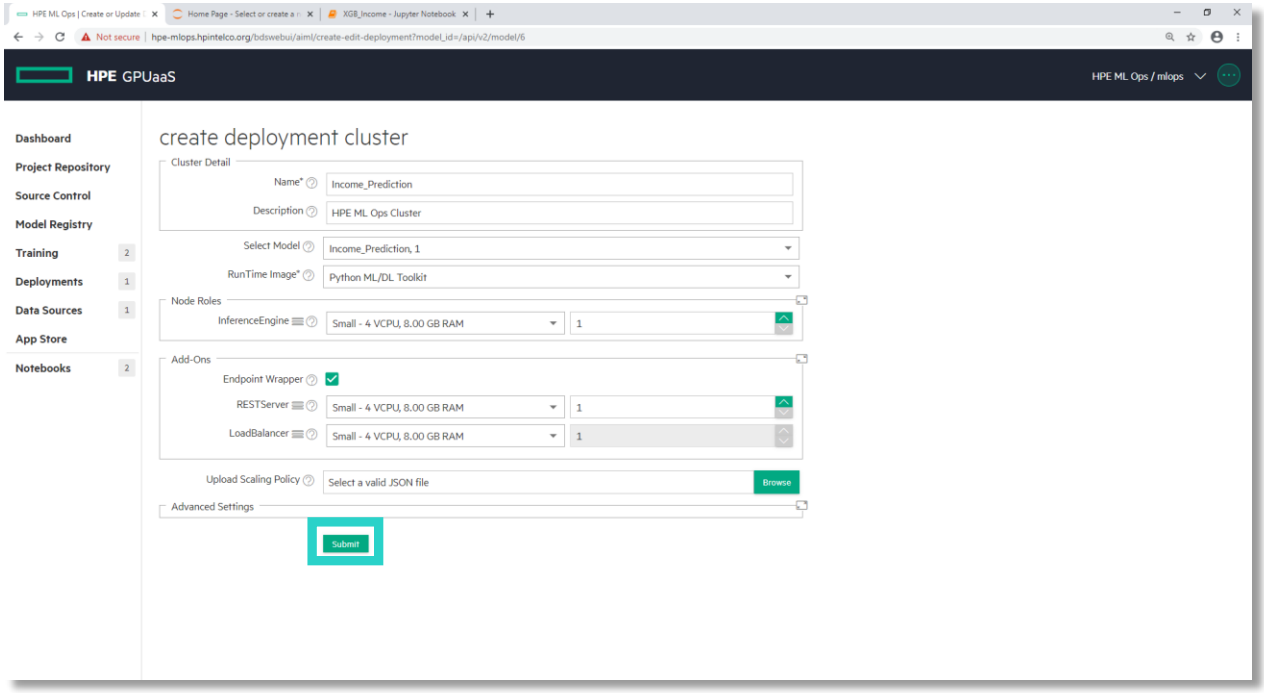
2. Register New Model using the information displayed below.



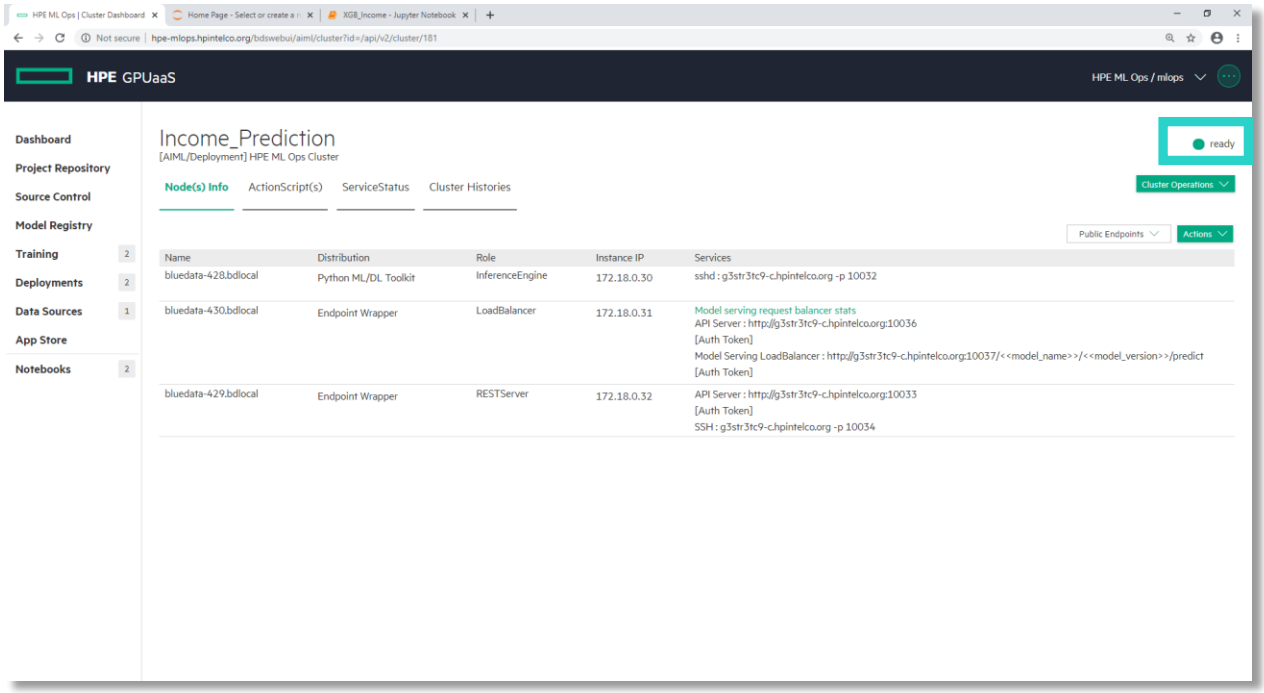
3. Deploy your registered model (Click on Deploy)



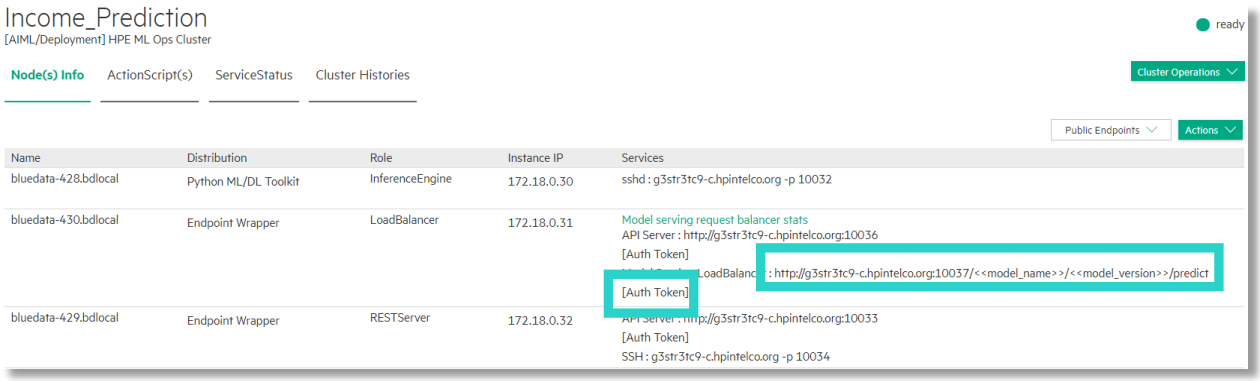
4. Create a Deployment Cluster for Income Prediction (Click on Submit)



5. Income Prediction Cluster is now Ready



6. Generate your Prediction Request



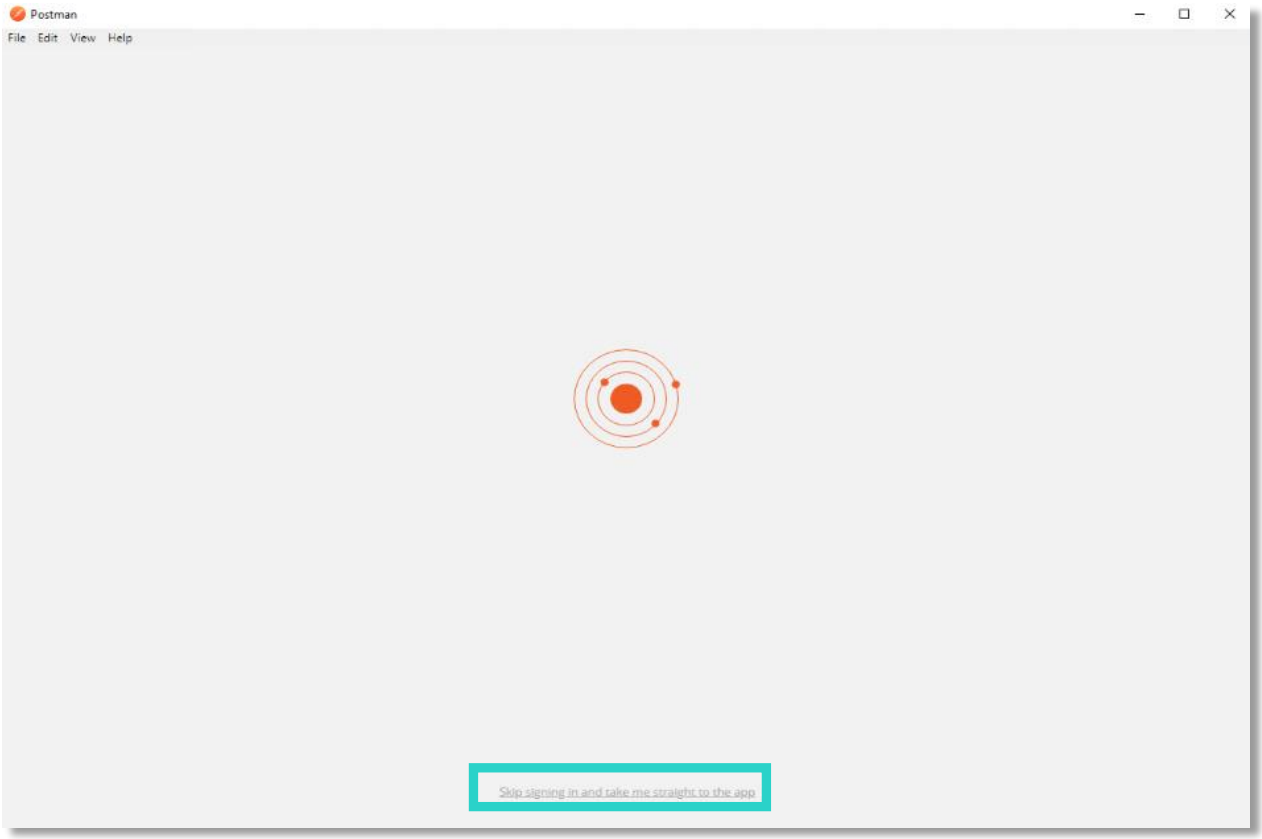
Copy the link and check update the port + model version

http://g3str3tc9-c.hpintelco.org:10037/Income_Prediction/1/predict
[Auth Token] Click to copy the token

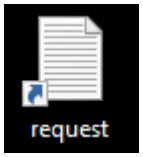
Run the Income prediction using Postman

This link will start Postman!

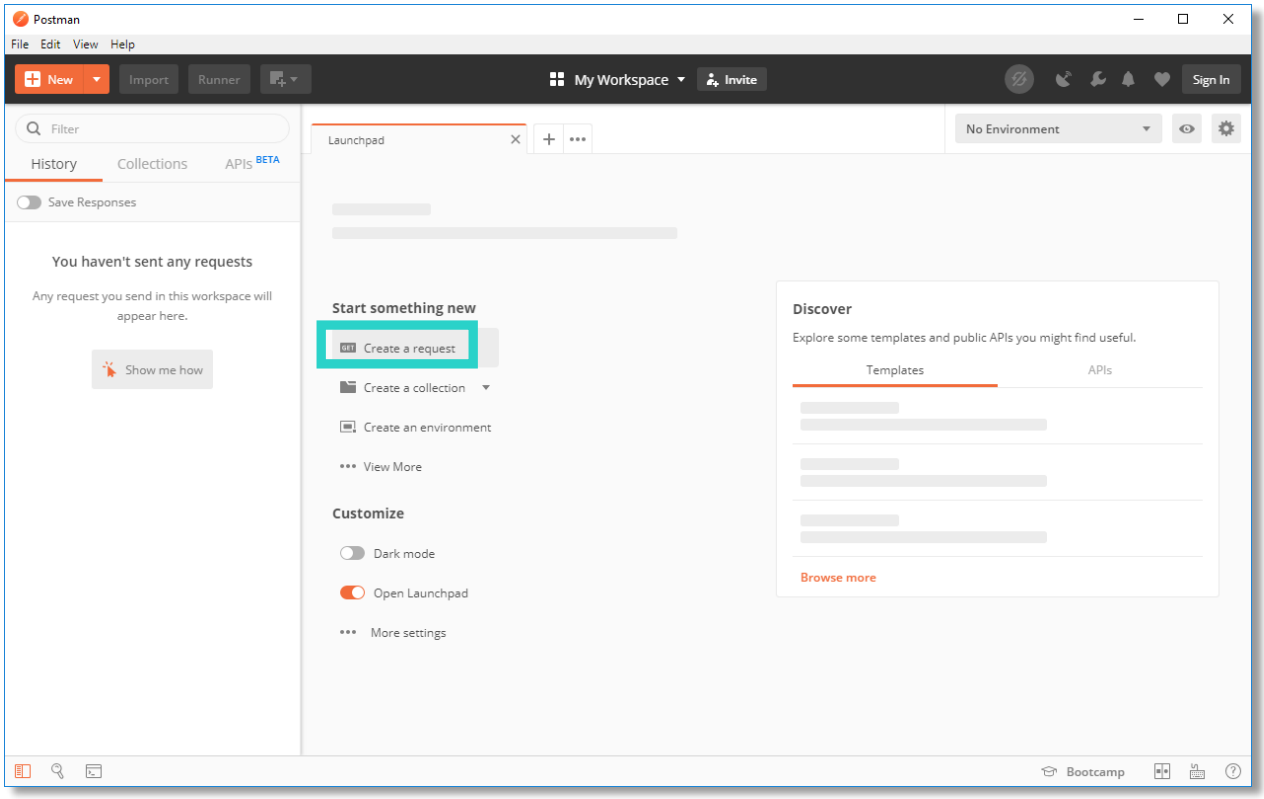
1. Select Skip signing in to go directly to the app



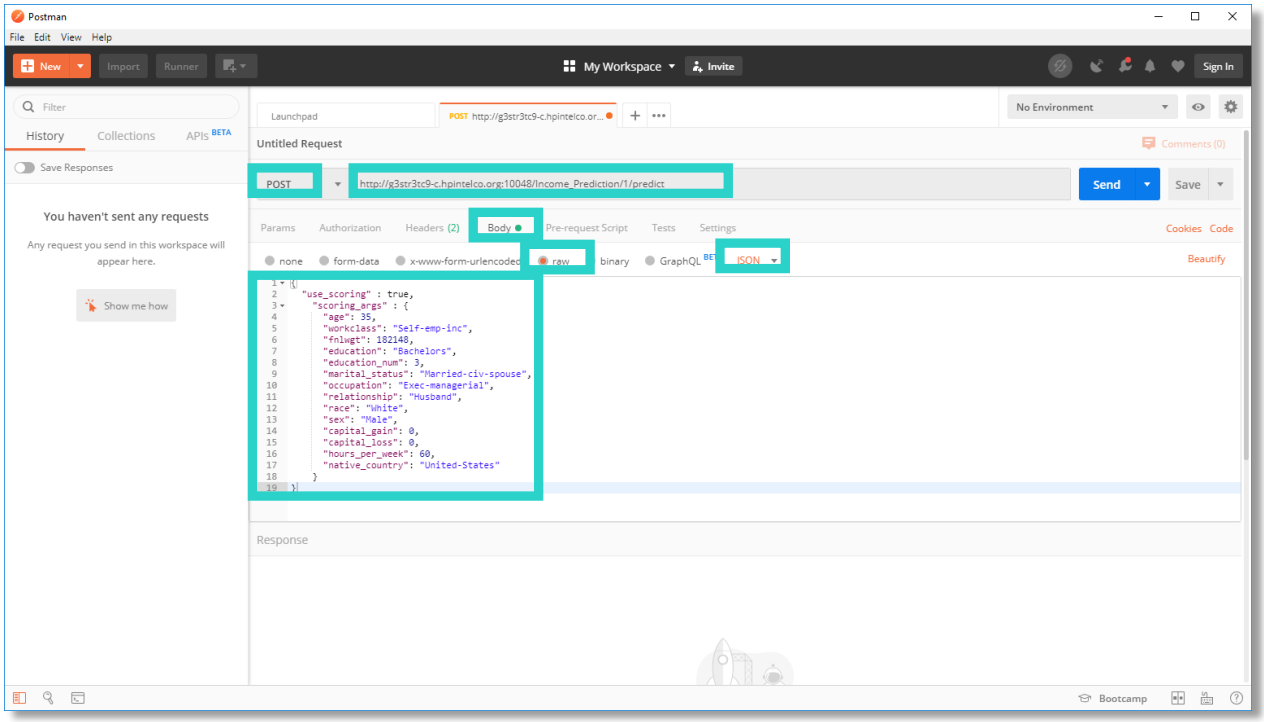
2. Copy the Income Json request in Postman body tab: Click on below icon



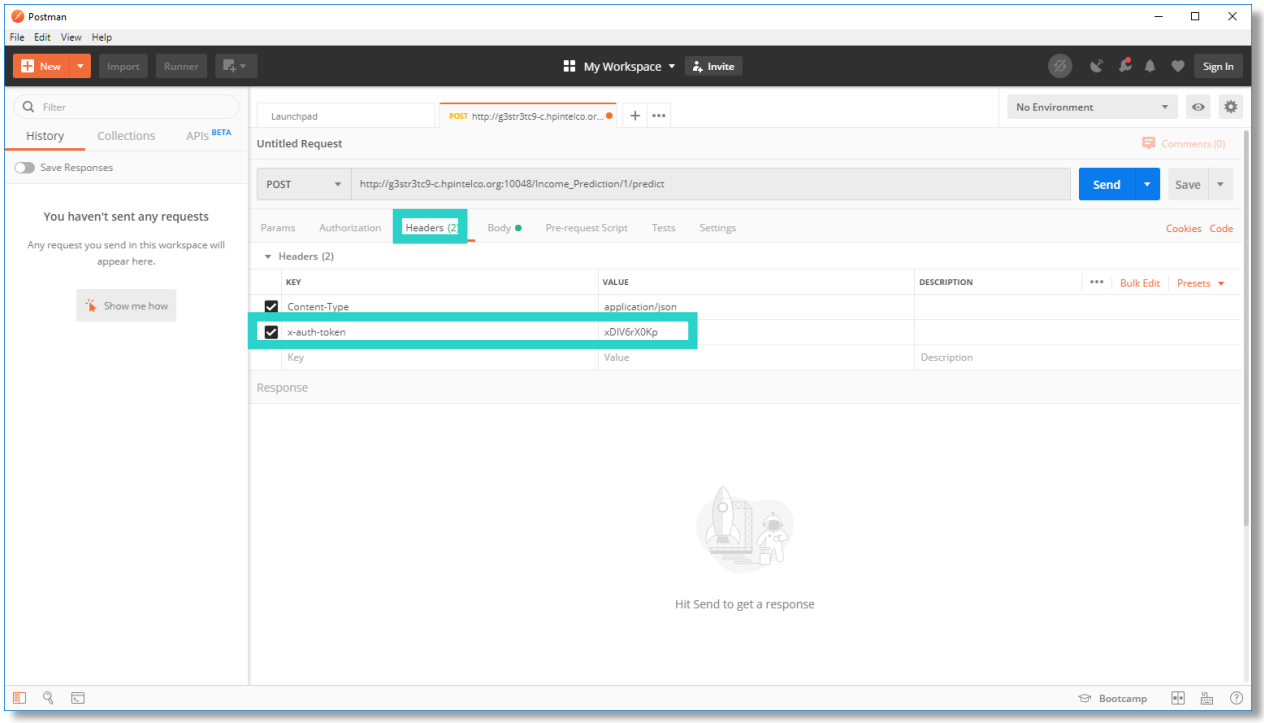
3. Create a new Request



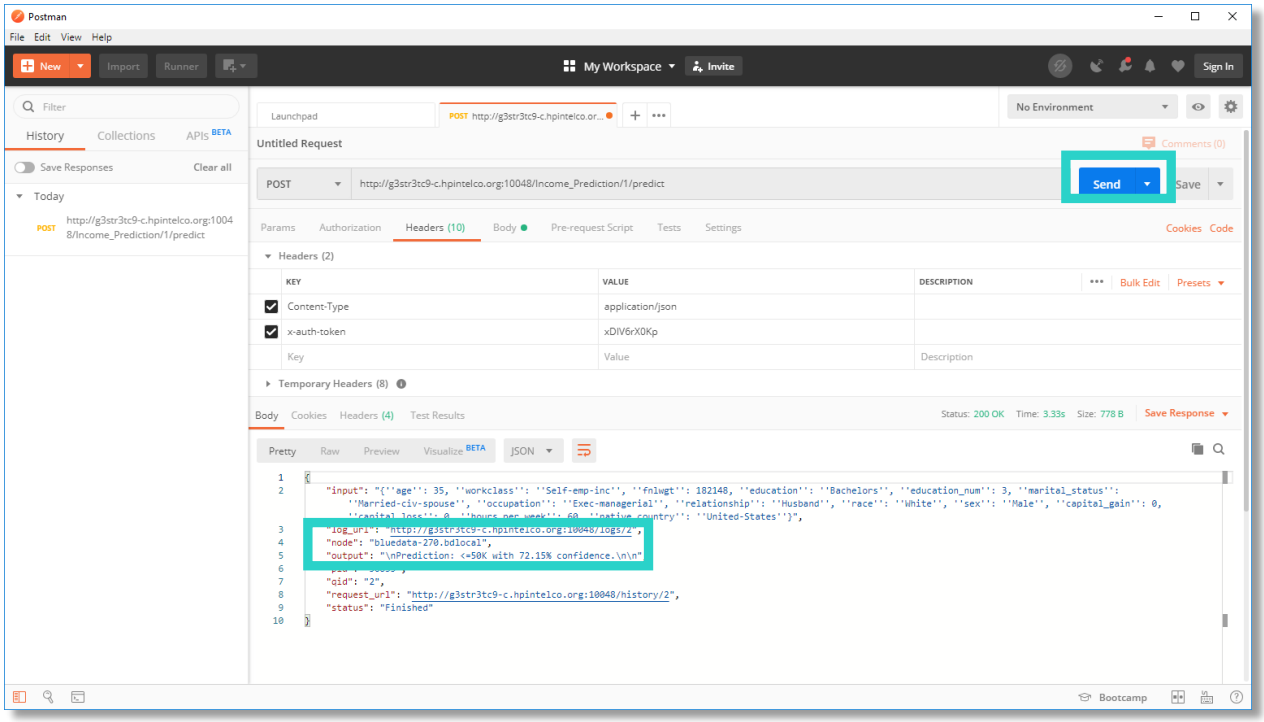
4. Fill in the Body tab to match the image below



5. Fill in the Header tab with your token



6. Send the post request to the cluster and get the income prediction result

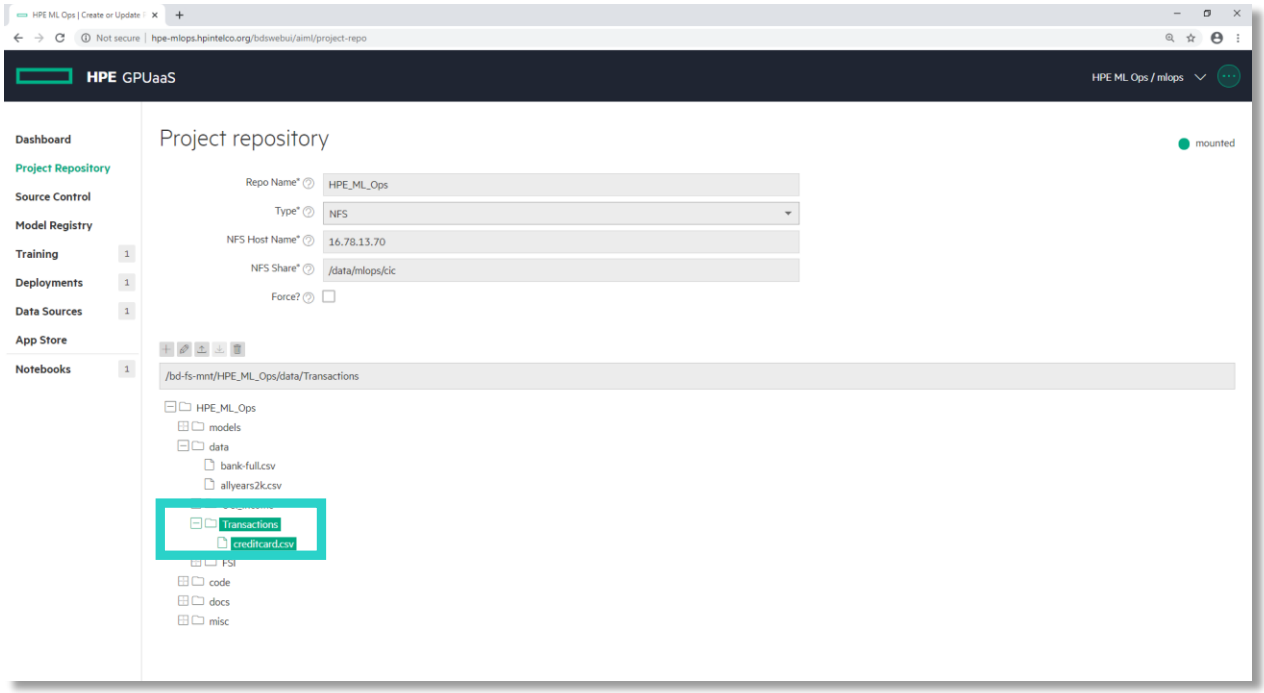


Fraud Detection Demo Scenario

This use case demonstrates using HPE ML Ops and XGBoost to detect credit card fraud.

Data set is in the Project Repository

This dataset contains the credit card transactions that occurred during a period of two days with 492 frauds out of 284,807 transactions. All the variables in the dataset are numerical. The time column contains the seconds elapsed between the first transaction in the dataset.



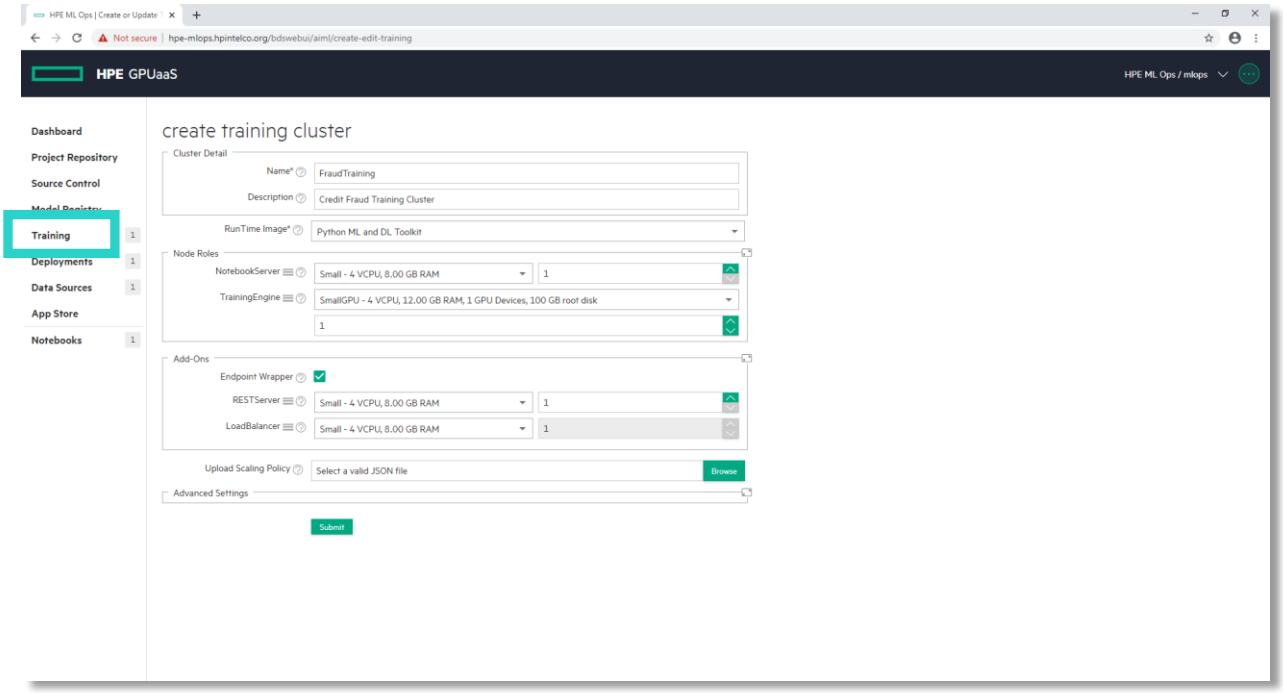
mlops project member (data scientist)

As a project member of ML Ops project, the following are typical tasks a data scientist would perform:

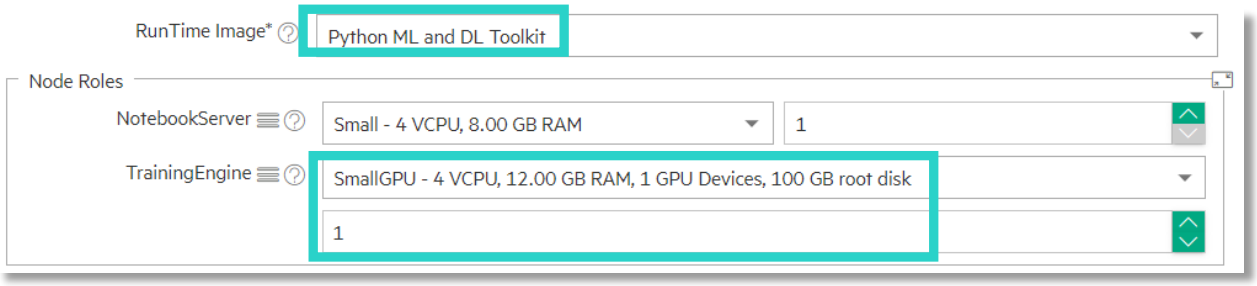
1. Create Training Cluster
2. Create personal Jupyter Notebook
3. Fraud data discover
4. Train XGBoost with CPU
5. Train XGBoost with GPU
6. Register a model
7. Create deployment cluster
8. Fraud prediction with PostMan

Create your HPE ML Ops Training cluster

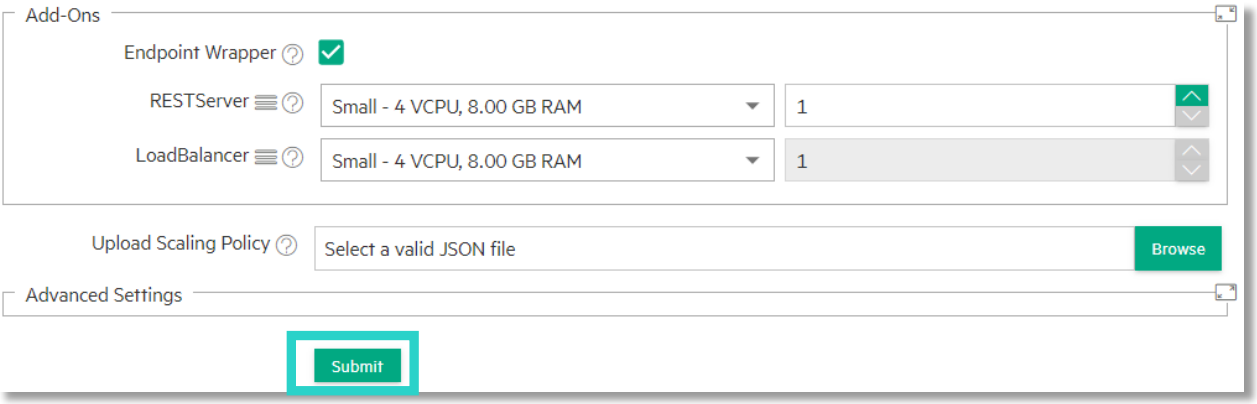
1. Select Create Training



Don't forget to add a Training Engine with GPU flavor and update the RunTime Image

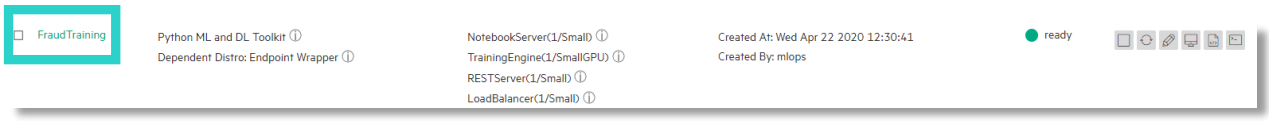


2. Click on Submit

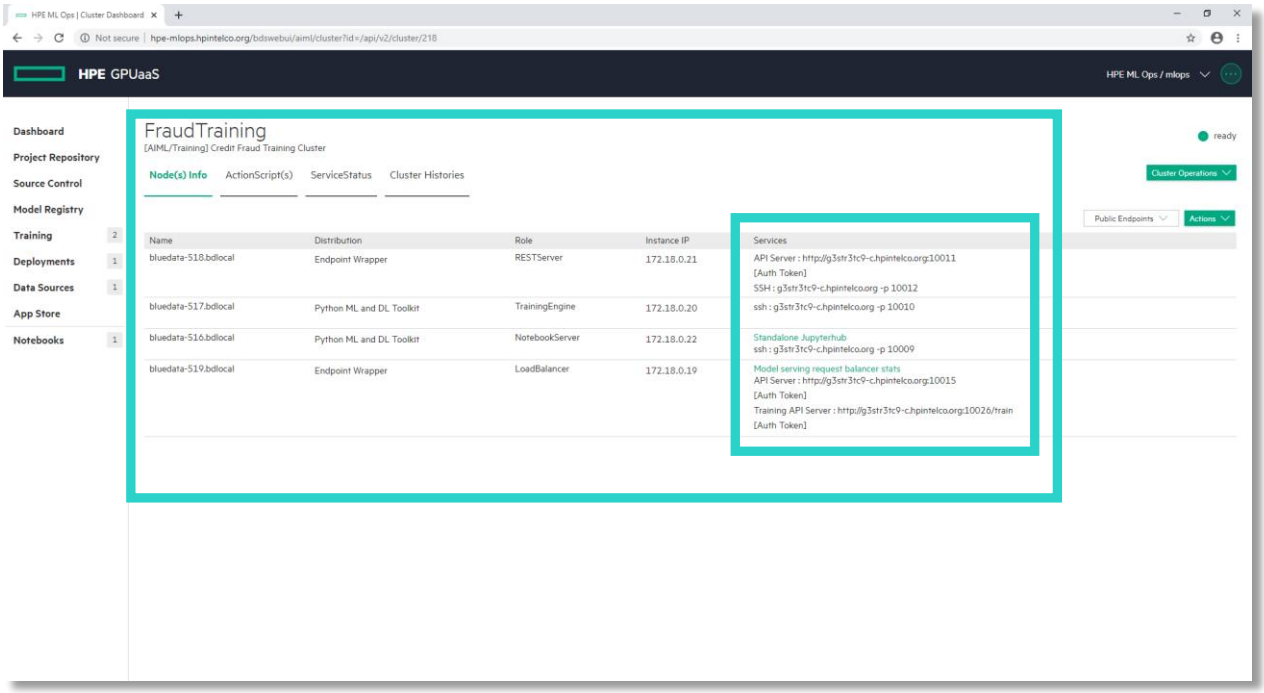


3. Check that the cluster is ready

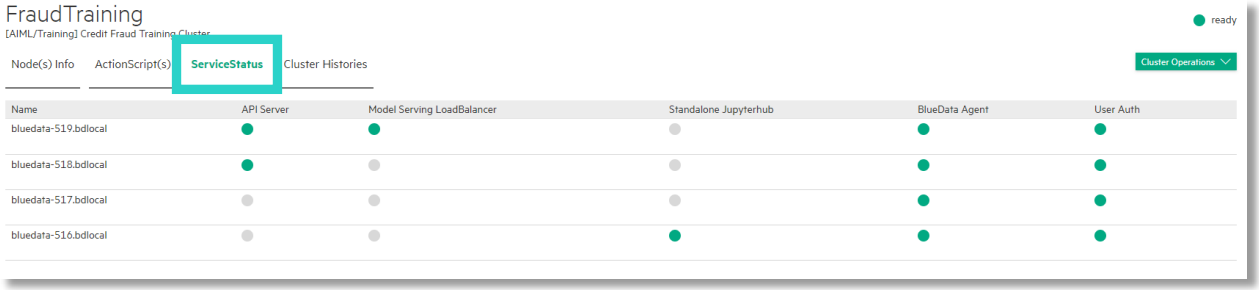
4. Open Training Cluster details (Click on FraudTraining)



5. Show the cluster nodes and services. (NotebookServer / TrainingEngine / …)

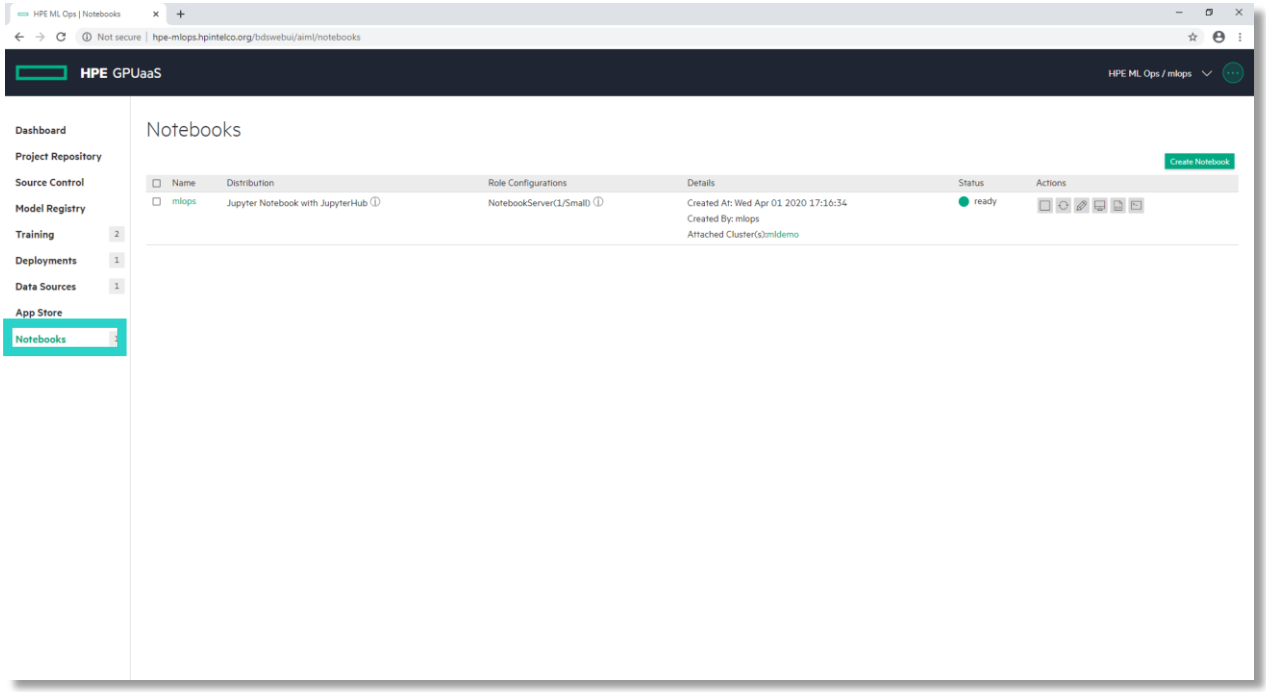


6. Check the Services Status of your cluster (Application / System services)

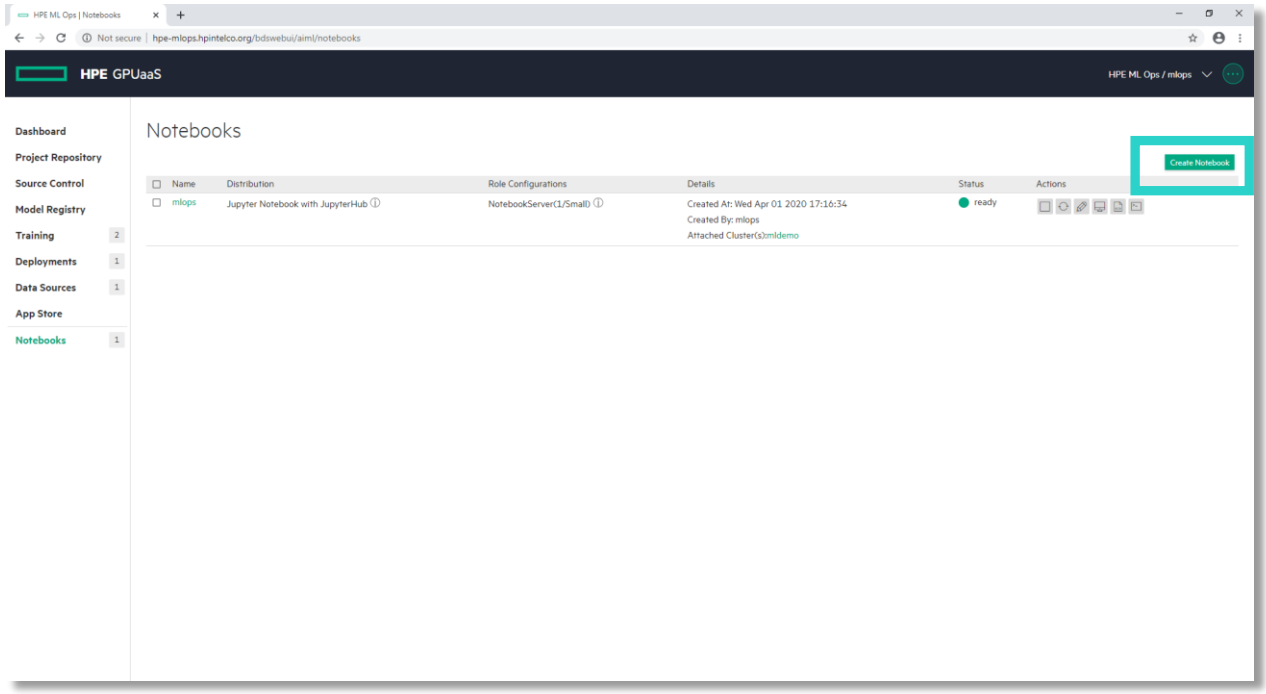


Develop/Create Jupyter Notebooks

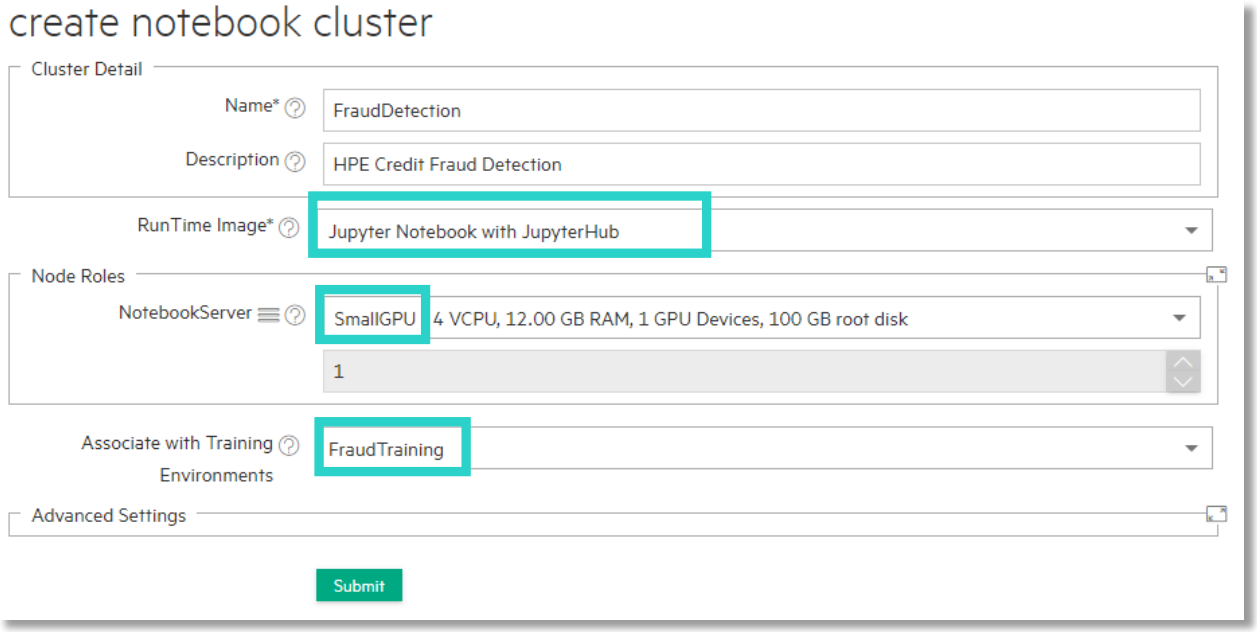
1. Select Notebooks from the left menu



2. Create a Jupyter Notebook (Click on Create Notebook)

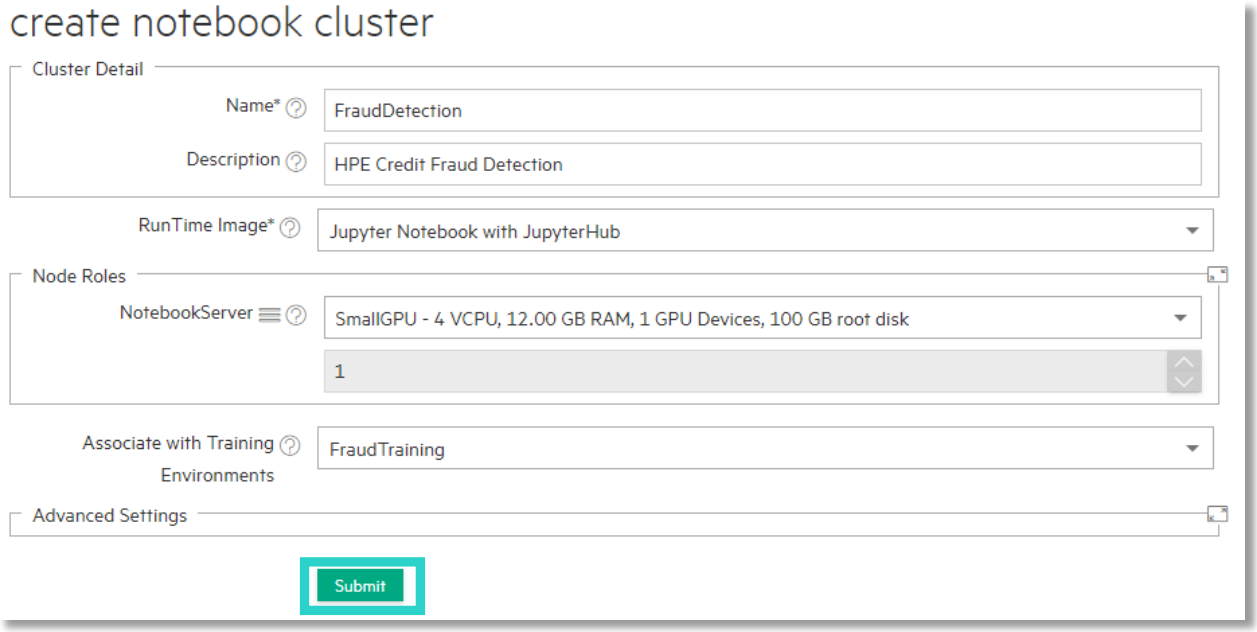


3. Associate the Notebook with your Training cluster

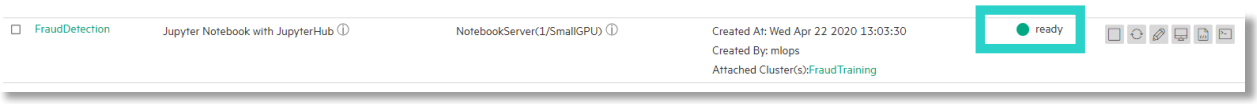


Don't forget to associate with Training Engine, setup GPU flavor and update the RunTime Image v2

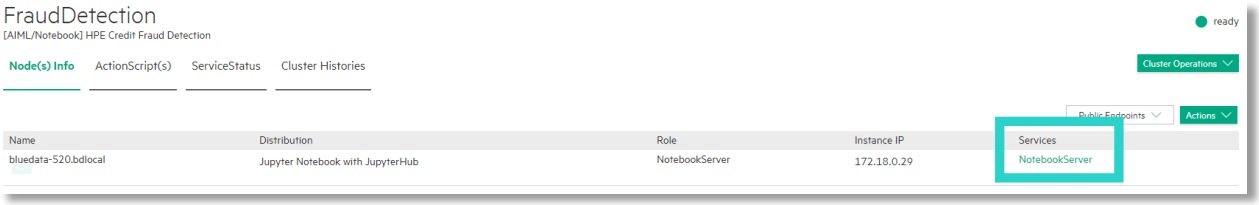
4. Create Notebook Cluster (Click on Submit)



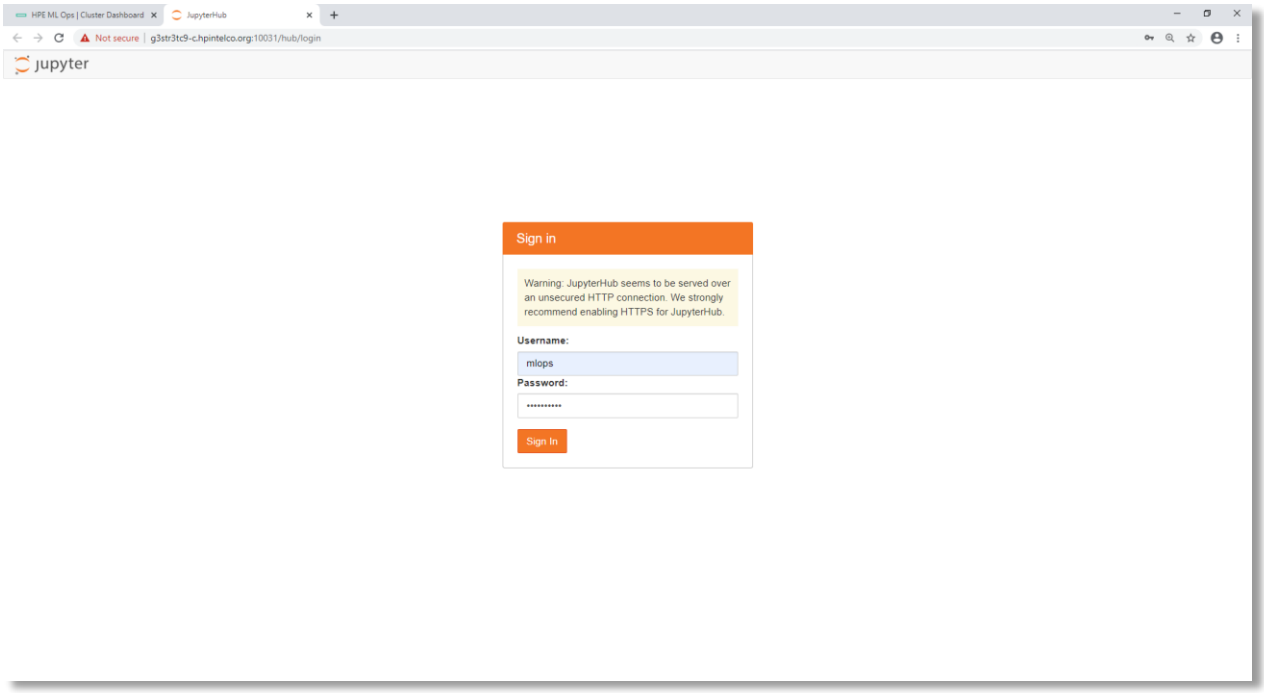
5. Make sure that the status of your notebook is ready



6. Open JupyterHub web ui (Click on NotebookServer)

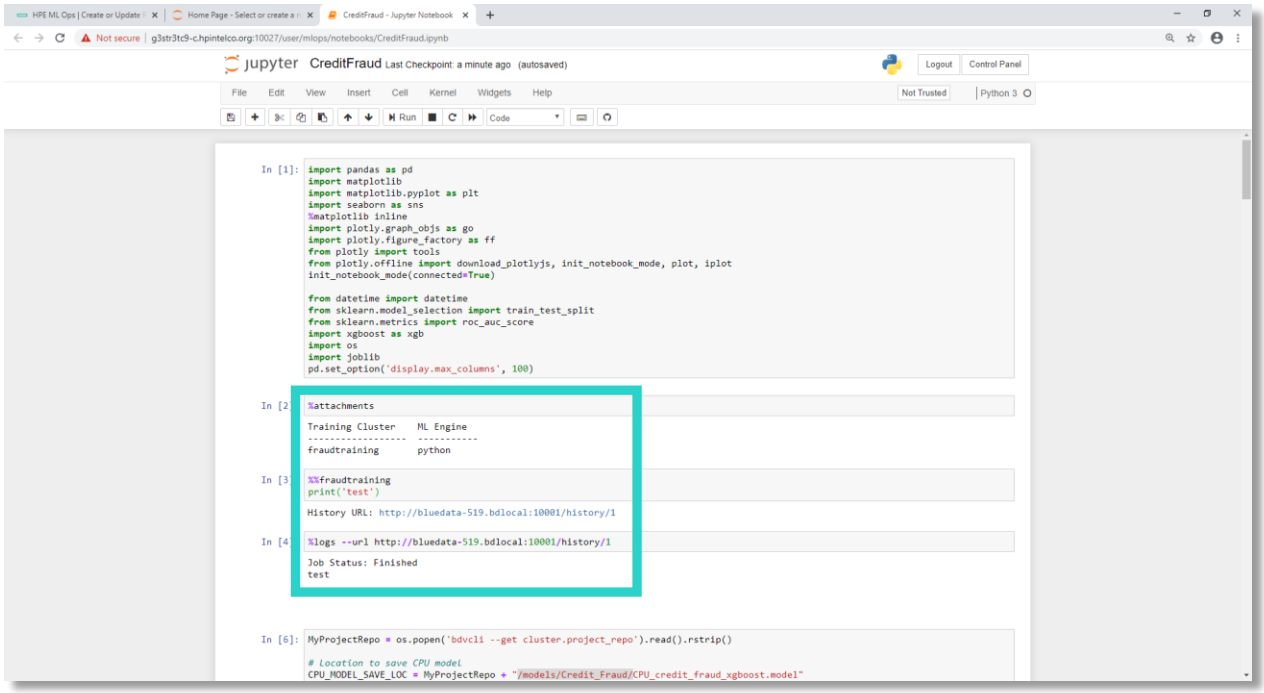


7. Login in to the Jupyter WebUI as **mlops** user



8. Load CreditFraud.ipynb notebook

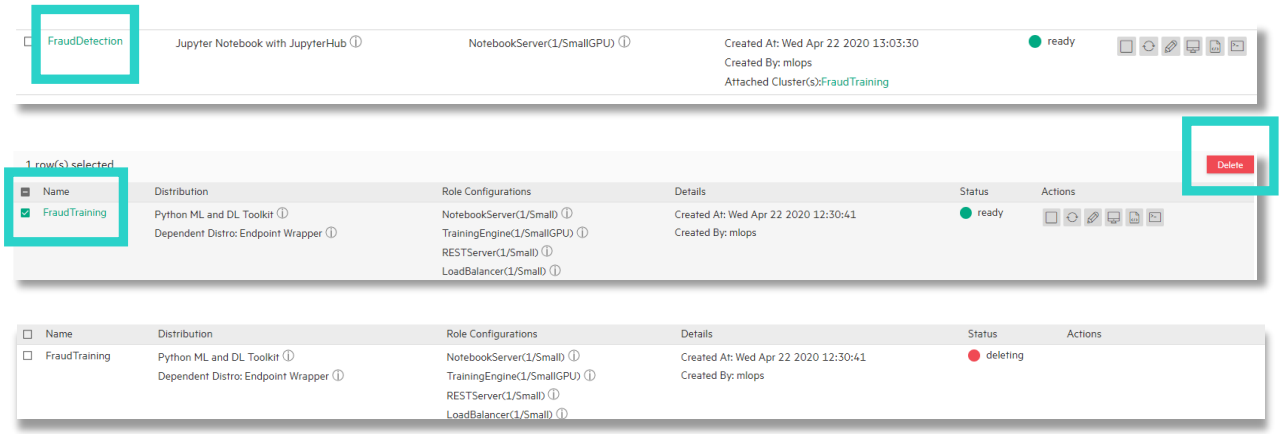
Open the `CreditFraud.ipynb` file, and then run each cell individually. This generates model files using CPUs and GPUs. The Notebook contains detailed comments and explanations.



(Caution: Update the variable according to your Training Cluster including History urls.)

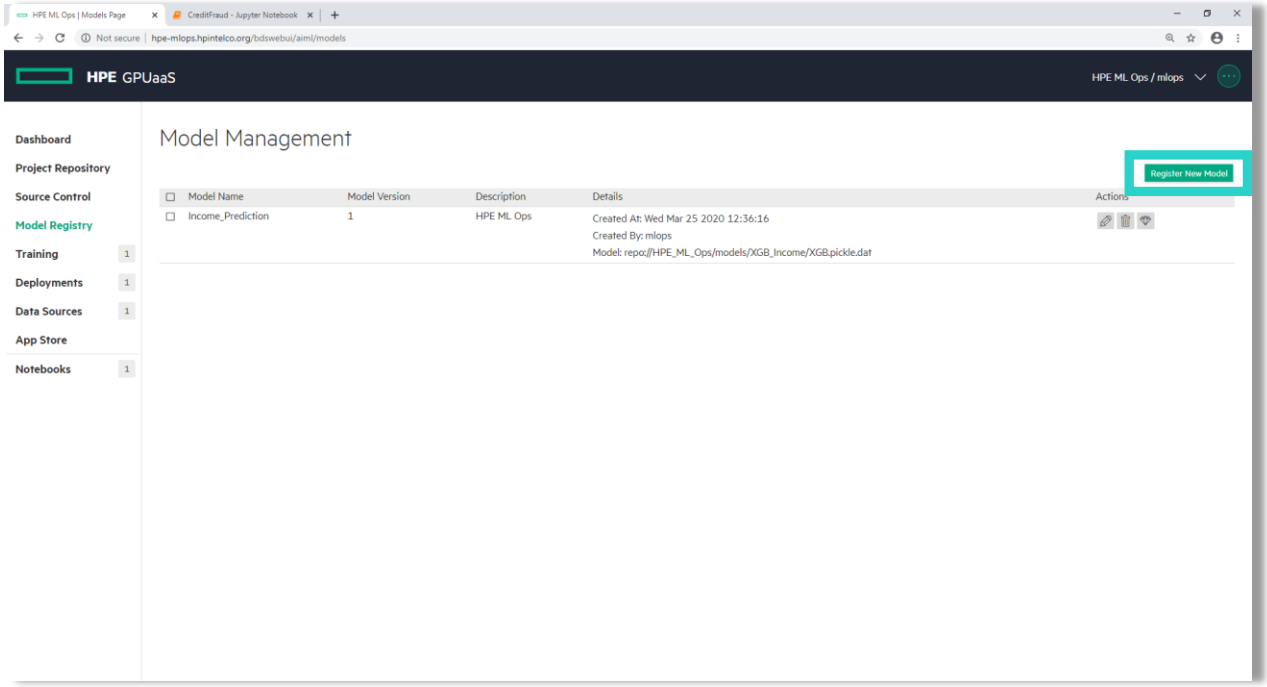
Once the model is tuned to the optimal parameters, run the cell on a remote Training cluster (where there are potentially more resources to train a model on a larger dataset).

9. Cleanup your Notebook and Training cluster (Needed to release GPU resources)

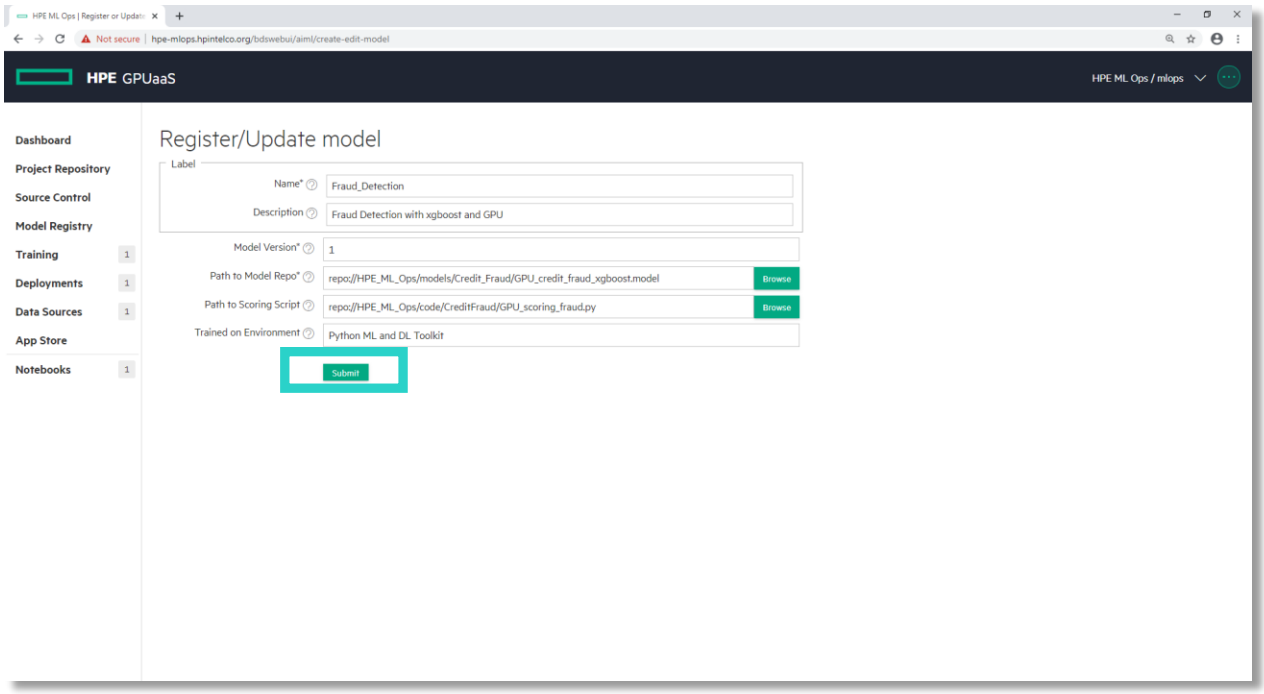


Register and Deploy the Model

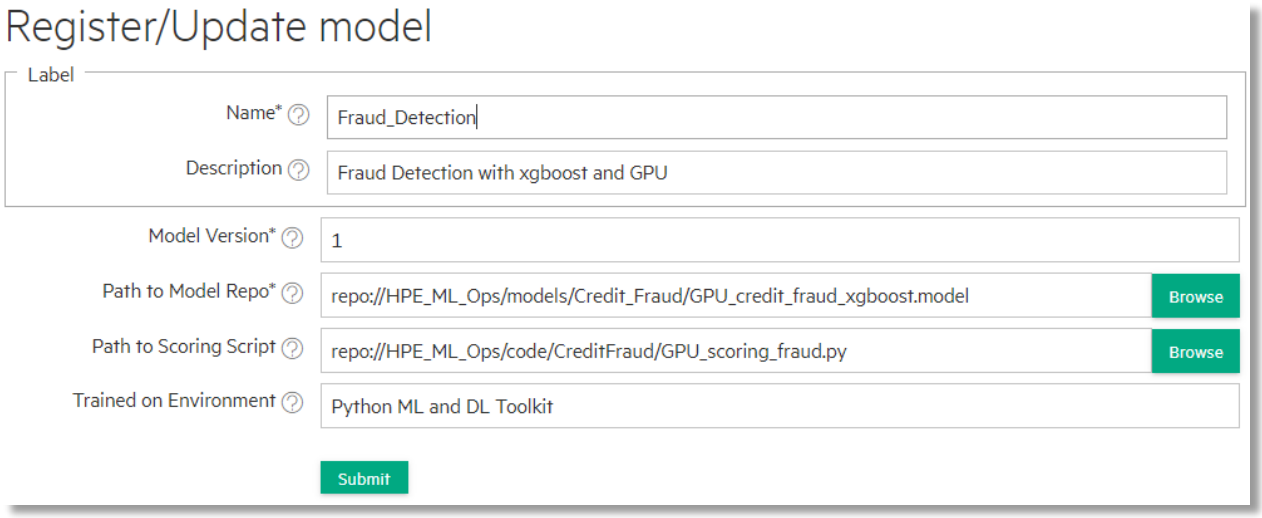
1. Select Model Registry from the left menu and click on Register New Model



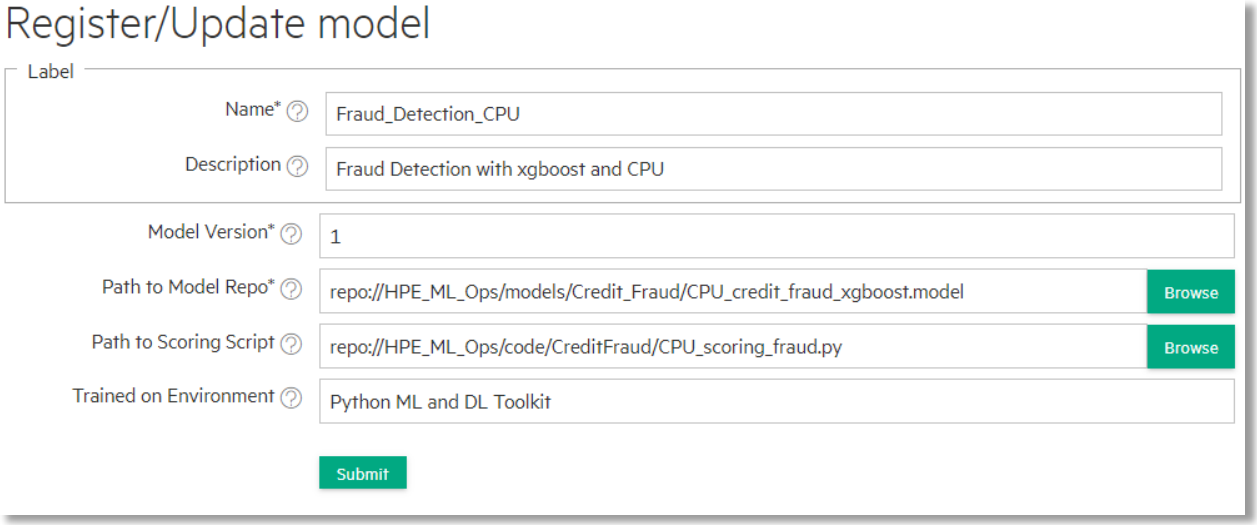
2. Register New Model using the information displayed below. (GPU / CPU option)



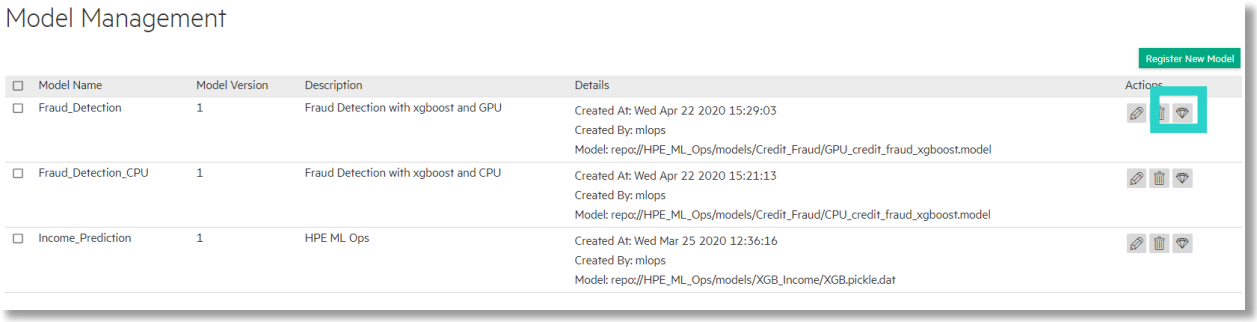
3. GPU option



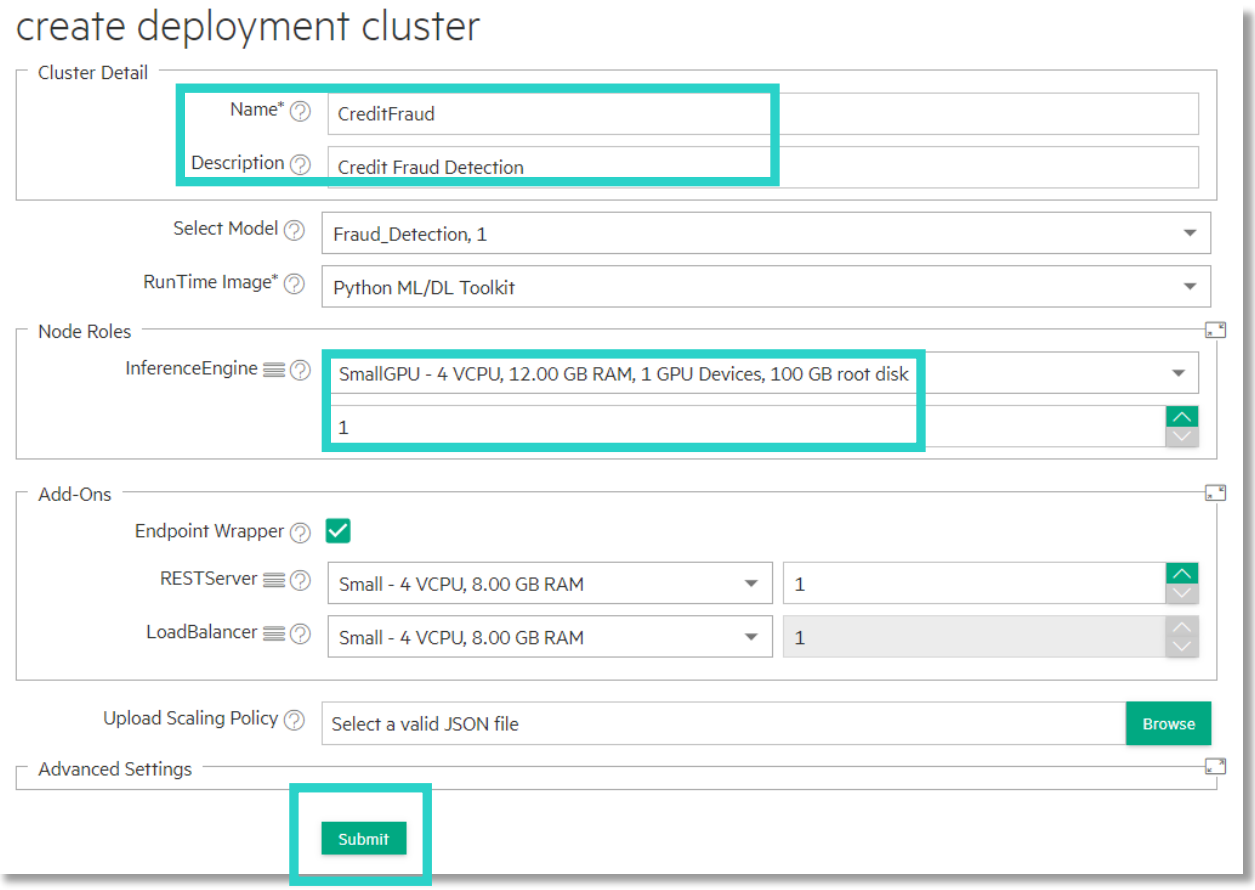
4. CPU option



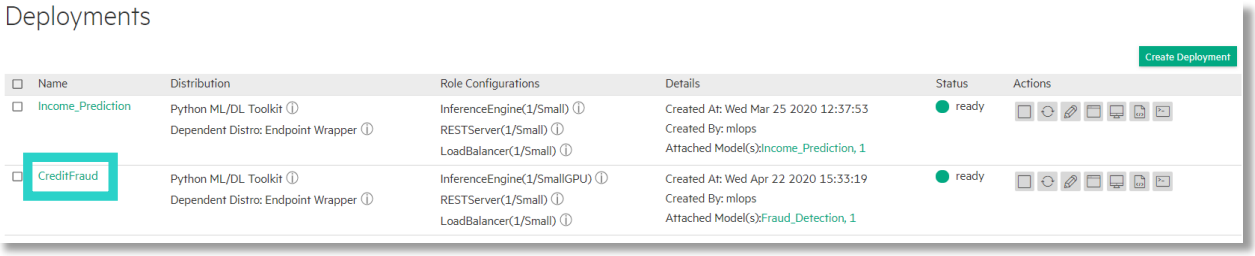
5. Deploy your registered model (Click on Deploy)

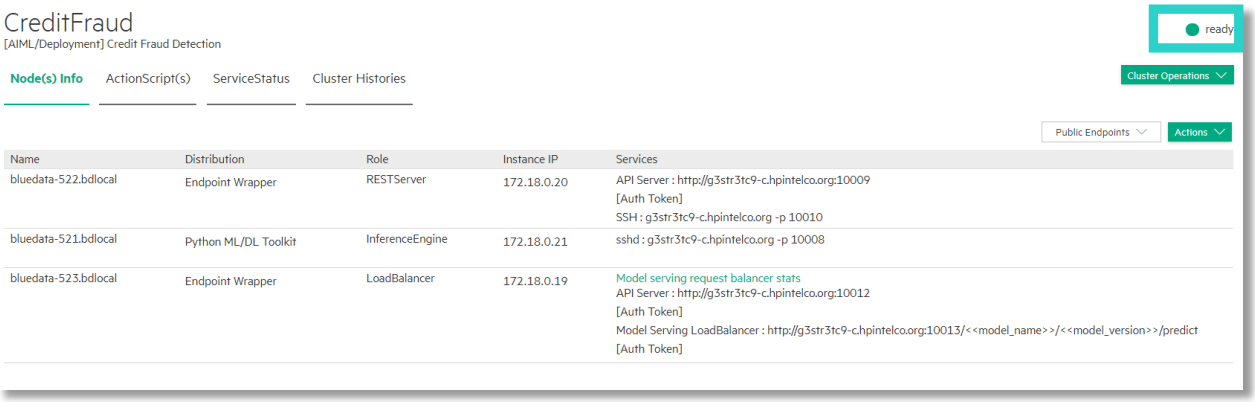


6. Create a Deployment Cluster for Credit Fraud Prediction (Click on Submit)

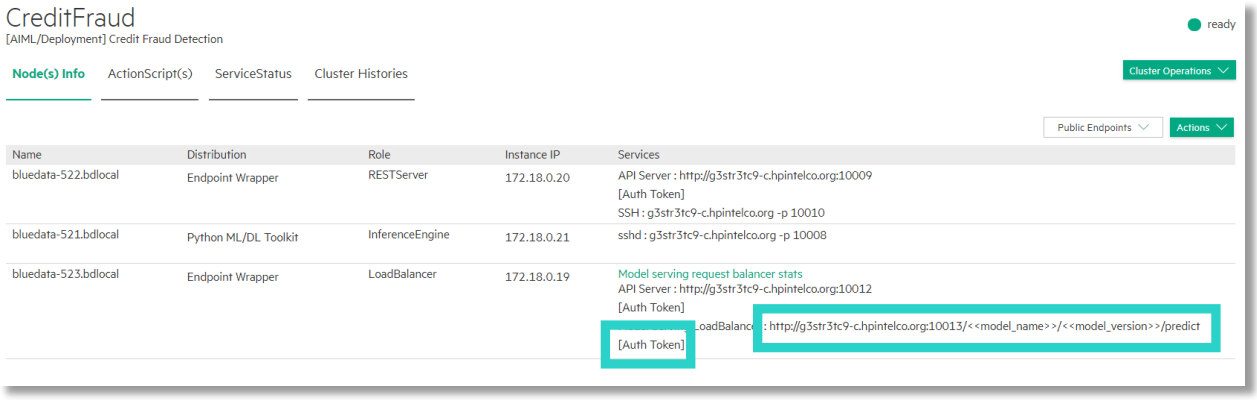


7. Credit Fraud Prediction Cluster is now Ready





8. Generate your Prediction Request



Copy the link and check update the port + model version

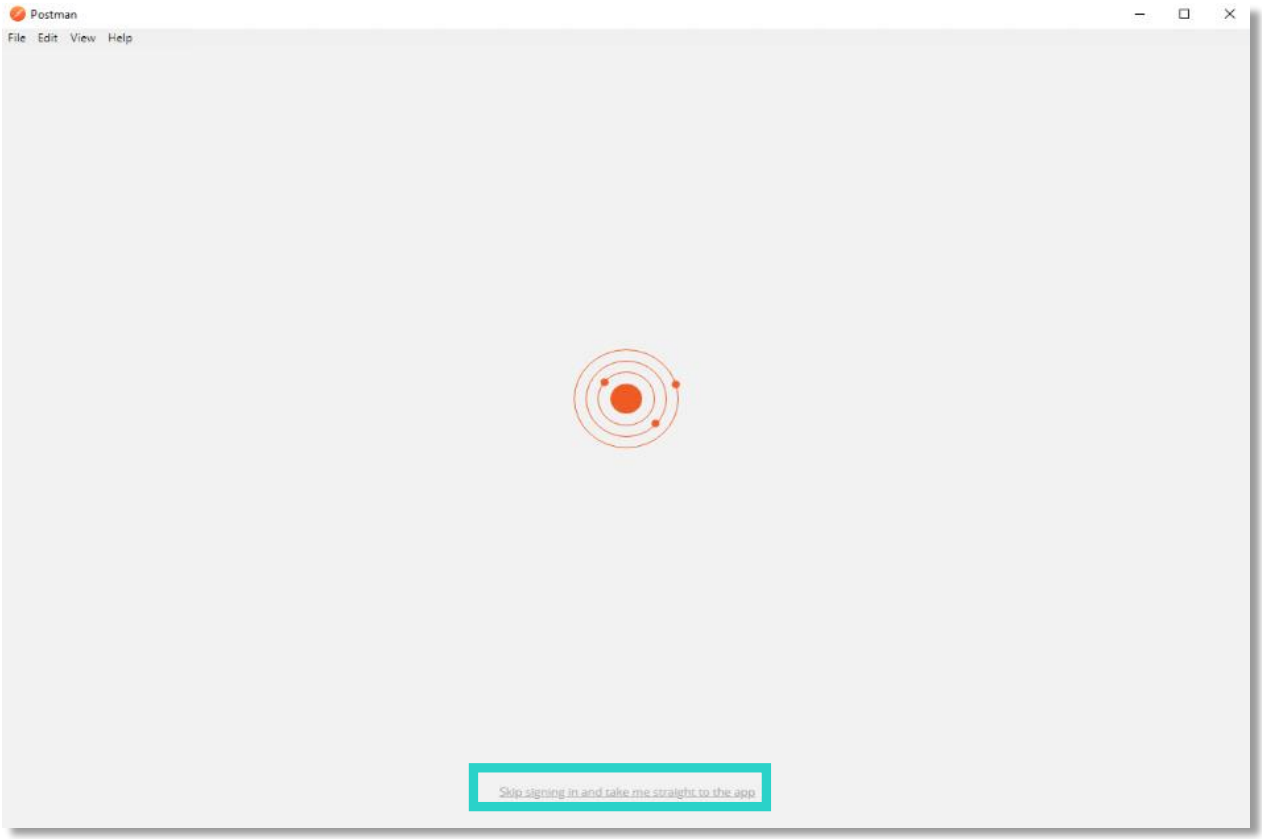
http://g3str3tc9-c.hpintelco.org:10013/Fraud_Detection/1/predict
[Auth Token] Click to copy the token

```
Non-Fraudulent Request

{
 "use_scoring" : true,
 "scoring_args" : {
"transaction": "0,-1.359807134,-0.072781173,2.536346738,1.378155224,-
0.33832077,0.462387778,0.239598554,0.098697901,0.36378697,0.090794172,-0.551599533,-0.617800856,-0.991389847,-
0.311169354,1.468176972,-0.470400525,0.207971242,0.02579058,0.40399296,0.251412098,-0.018306778,0.277837576,-
0.11047391,0.066928075,0.128539358,-0.189114844,0.133558377,-0.021053053,149.62,0",
 "execCount": 1
 }}

Fraudulent Request

{
 "use_scoring" : true,
 "scoring_args" : {
  "transaction": "406,-2.3122265423263,1.95199201064158,-1.60985073229769,3.9979055875468,-0.522187864667764,-
1.42654531920595,-2.53738730624579,1.39165724829804,-2.77008927719433,-2.77227214465915,3.20203320709635,-
2.89990738849473,-0.595221881324605,-4.28925378244217,0.389724120274487,-1.14074717980657,-2.83005567450437,-
0.0168224681808257,0.416955705037907,0.126910559061474,0.517232370861764,-0.0350493686052974,-
0.465211076182388,0.320198198514526,0.0445191674731724,0.177839798284401,0.261145002567677,-
0.143275874698919,0,0",
"execCount": 1
}}
```

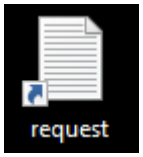
Run the Fraud Detection prediction using Postman

This link will start Postman!

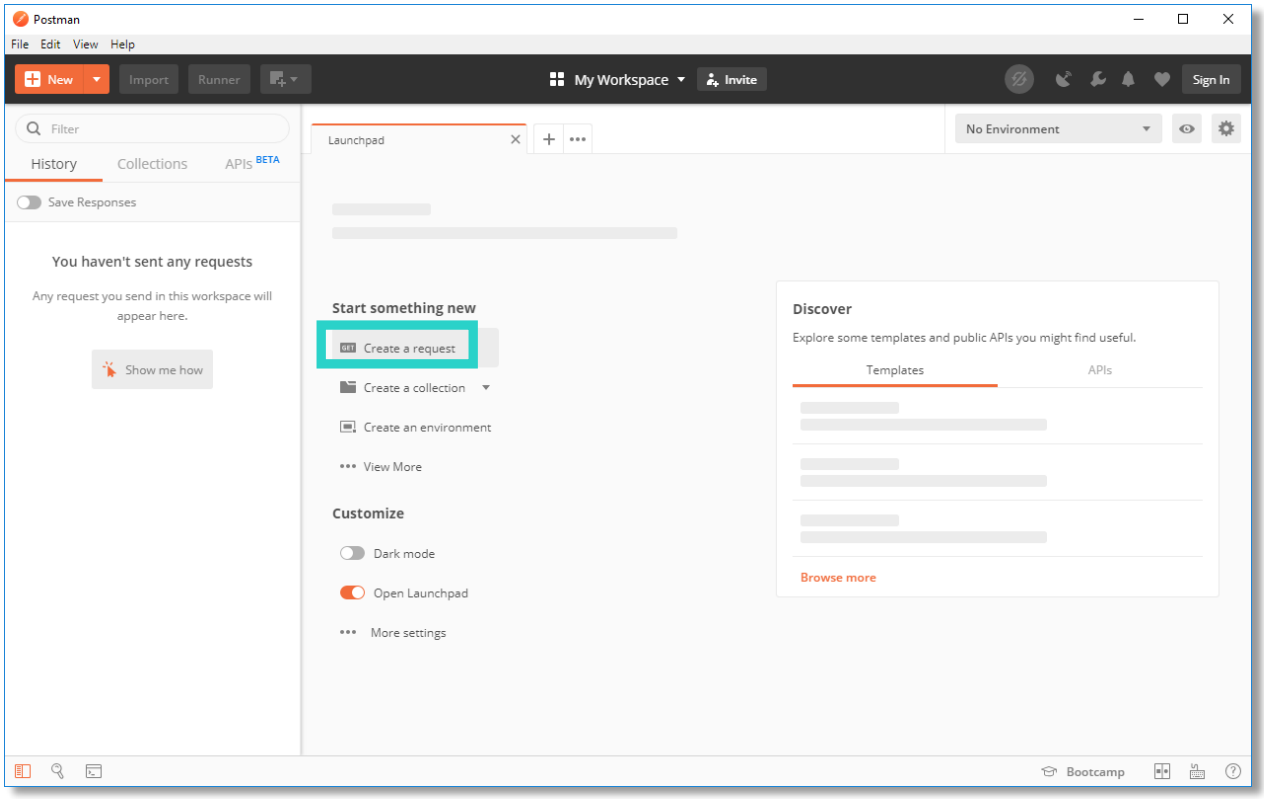
9. Select Skip signing in to go directly to the app



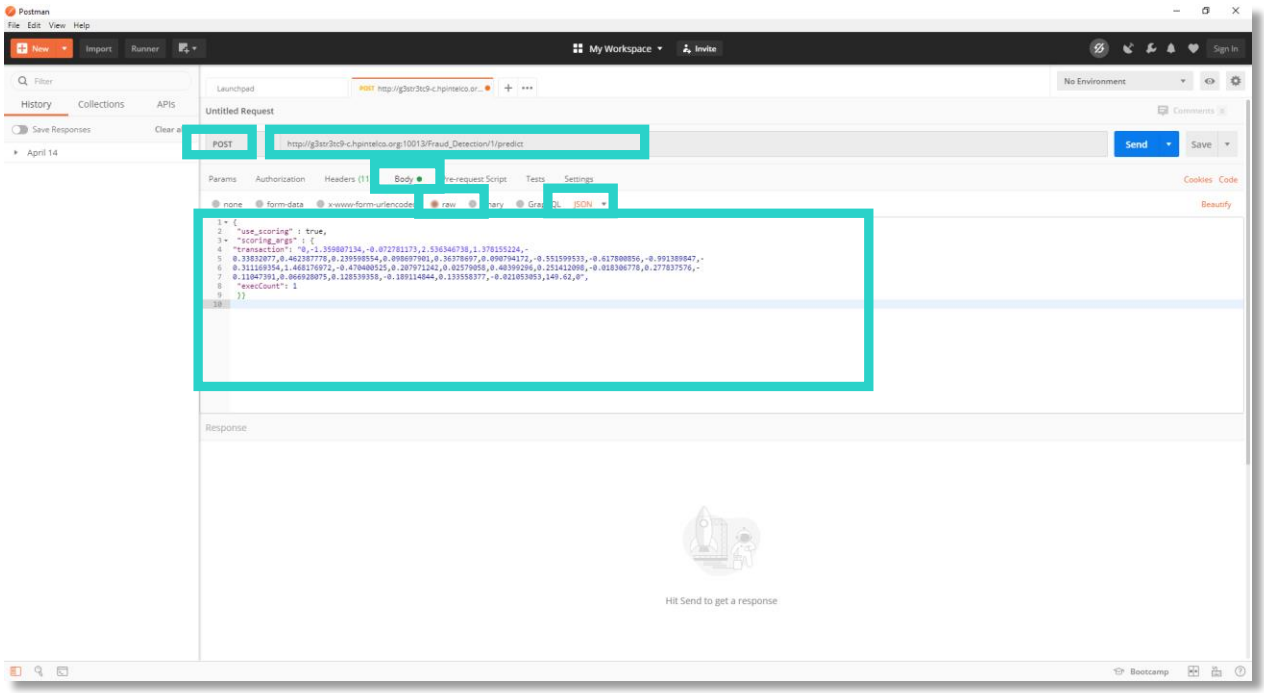
10. Copy the Json request in Postman body tab: Click on below icon



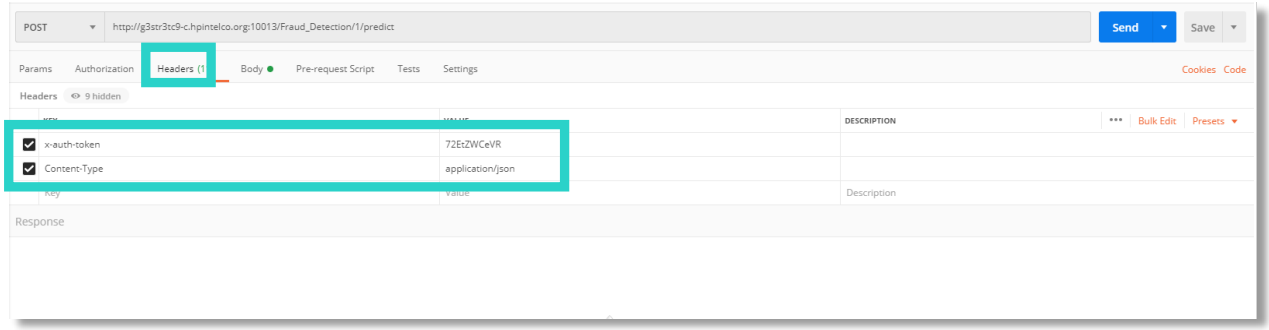
11. Create a new Request



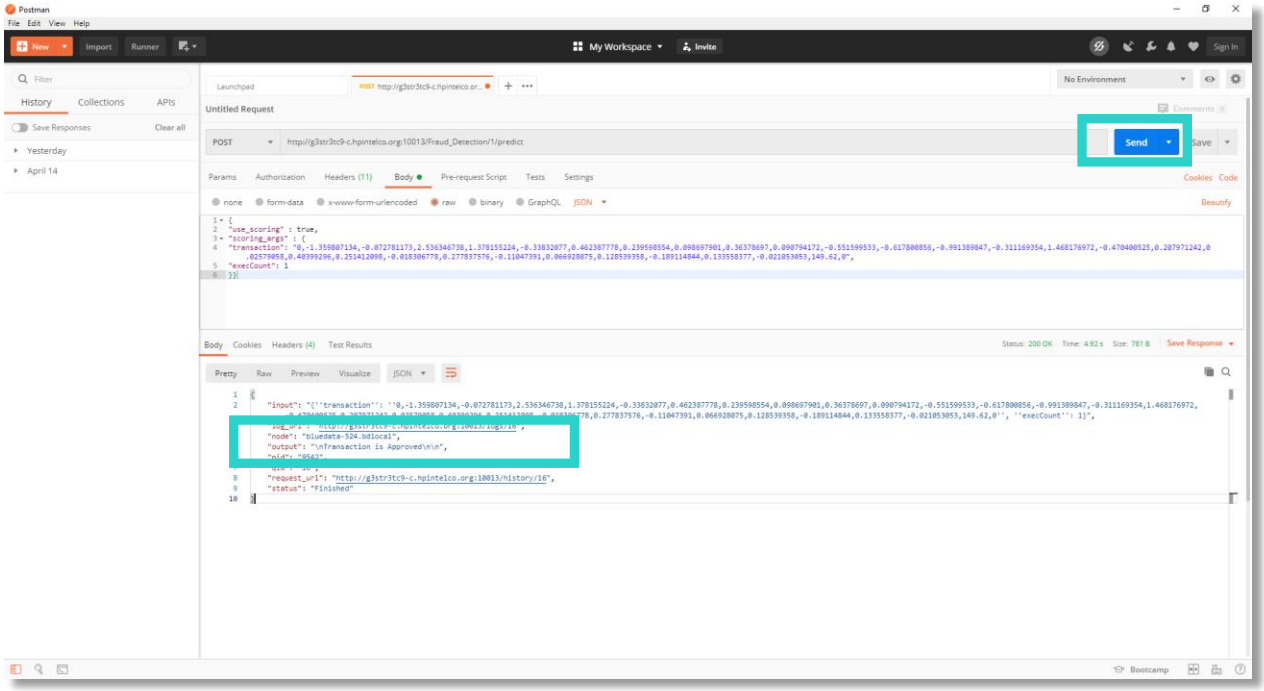
12. Fill in the Body tab to match the image below



13. Fill in the Header tab with your token



14. Send the post request to the cluster and get the Credit Fraud prediction result

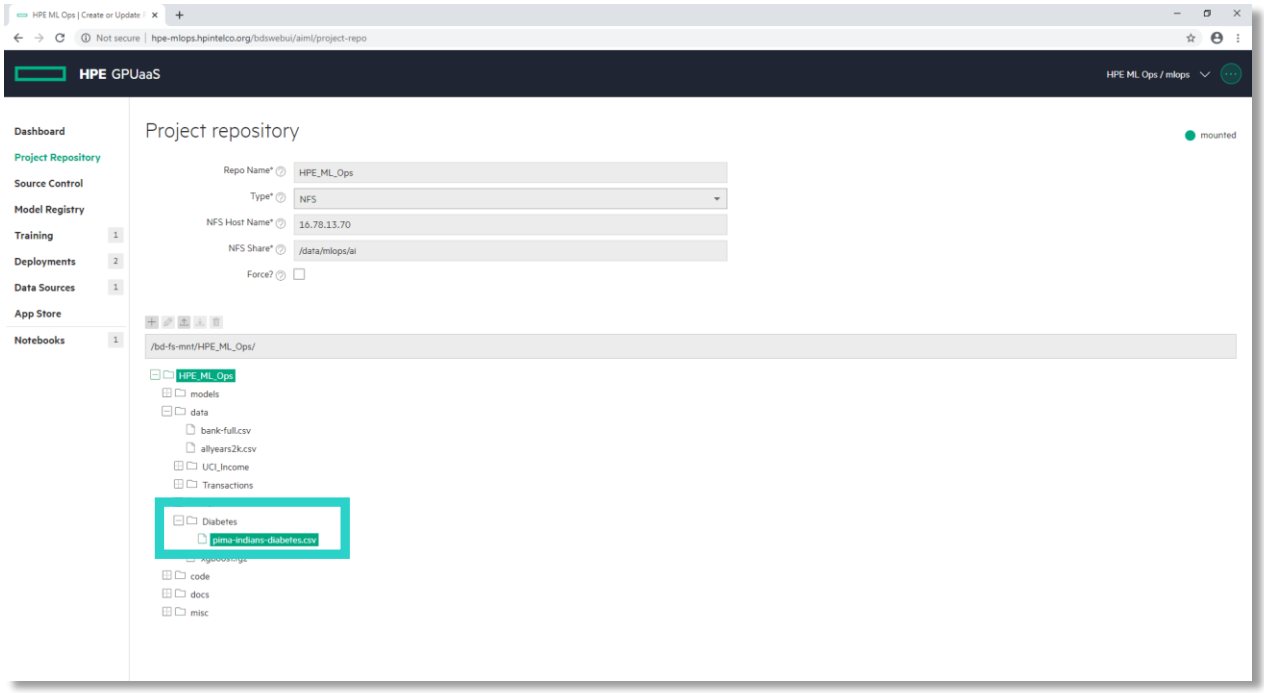


15. **Fraudulent Request**

Diabetes Prediction Demo Scenario

This tutorial generates a model that classifies whether a person has diabetes or not taking into account a dataset that contains data from female patients who were at least 21 years old and of Pima Indian heritage...

Data set is in the Project Repository

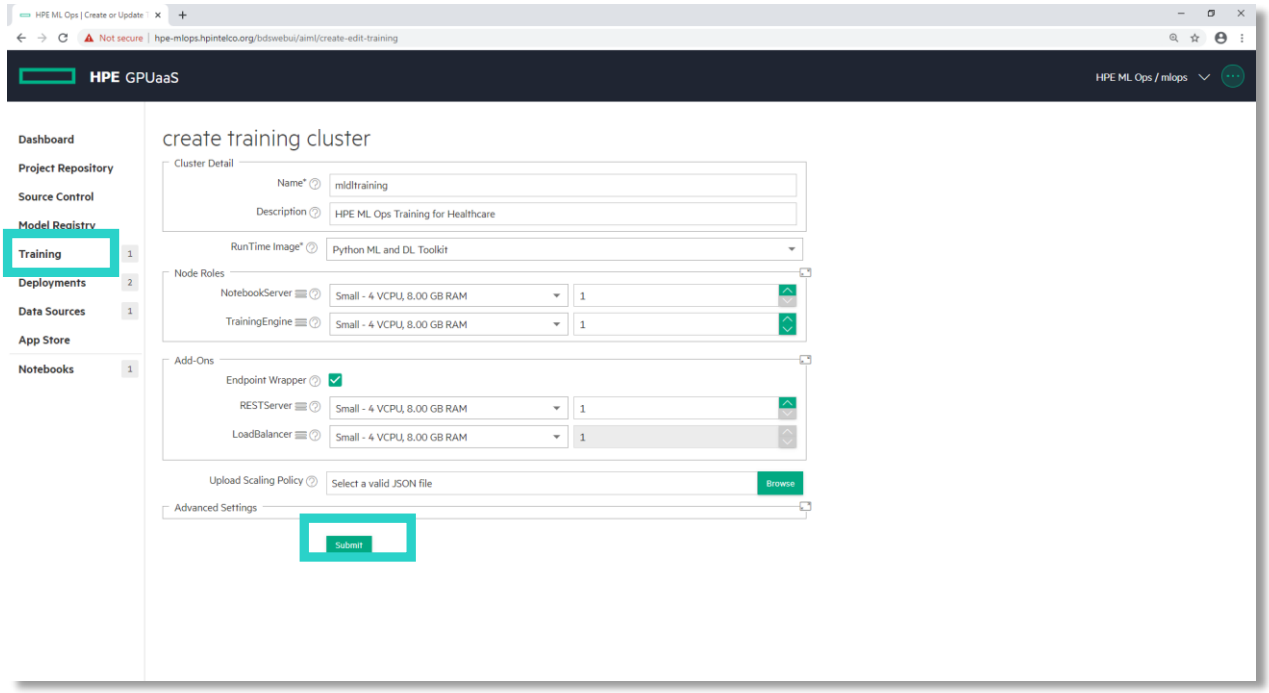


The features (columns) in this spreadsheet that are used to train the model are:

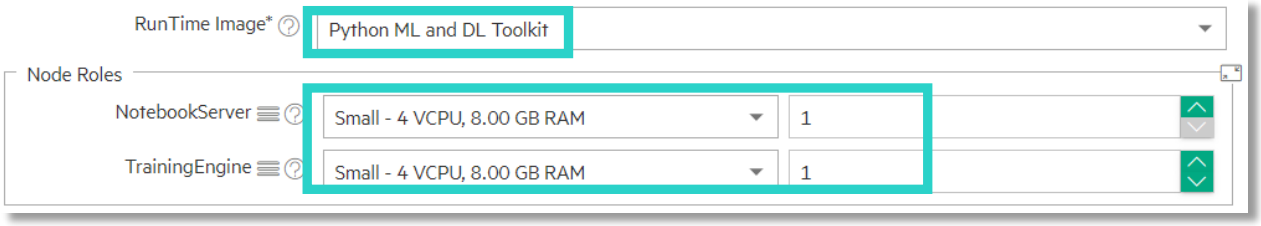
- Pregnancies
- Glucose — The blood plasma glucose concentration after a 2 hour oral glucose tolerance test.
- BloodPressure — Diastolic blood pressure (mm/HG).
- SkinThickness — Skinfold thickness of the triceps (mm).
- Insulin — 2 hour serum insulin (mu U/ml).
- BMI — Body mass index (kg/m squared)
- DiabetesPedigreeFunction: Function that determines the risk of type 2 diabetes based on family history.
- Age
- Outcome — whether the person is diagnosed with type 2 diabetes (1 = yes, 0 = no).

Create your HPE ML Ops Training cluster

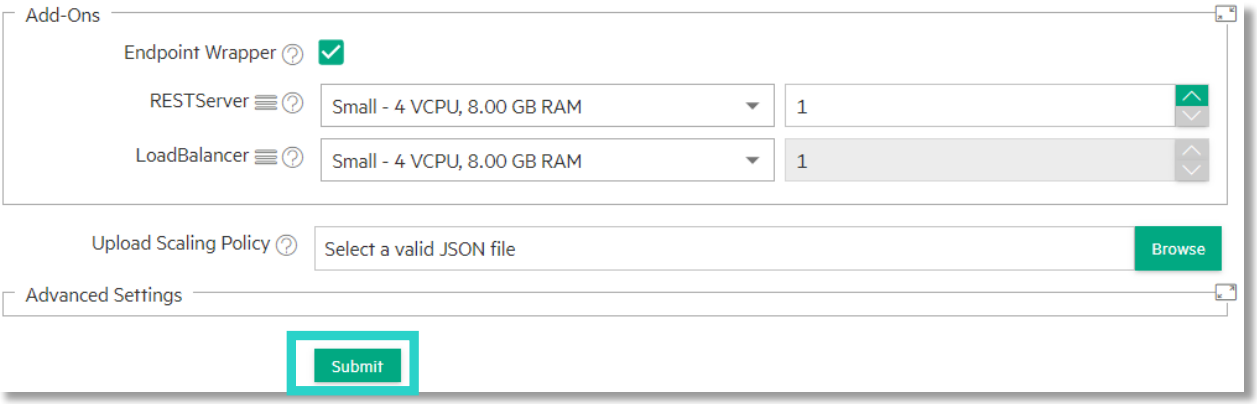
1. Select Create training cluster



Don't forget to add a Training Engine with CPU flavor and update the RunTime Image



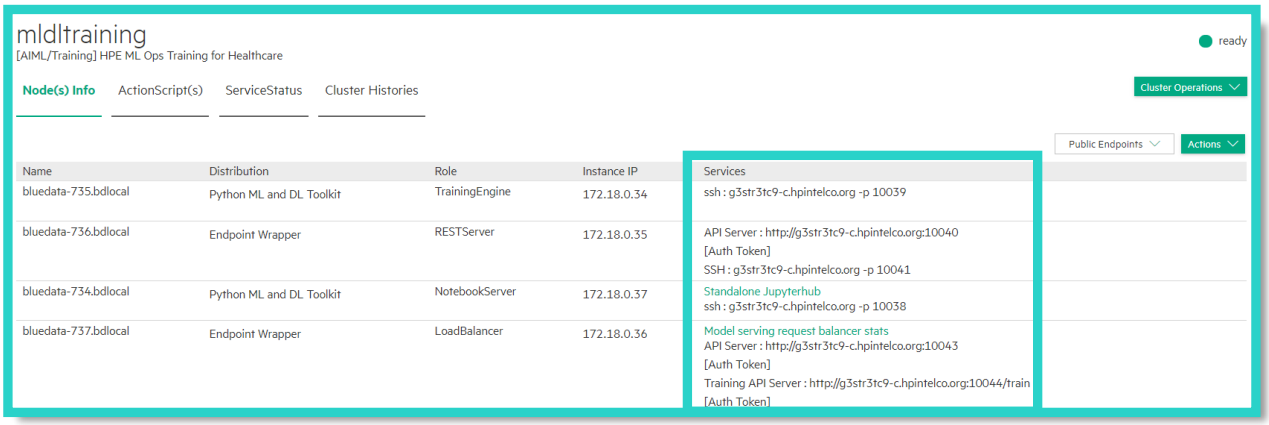
2. Click on Submit



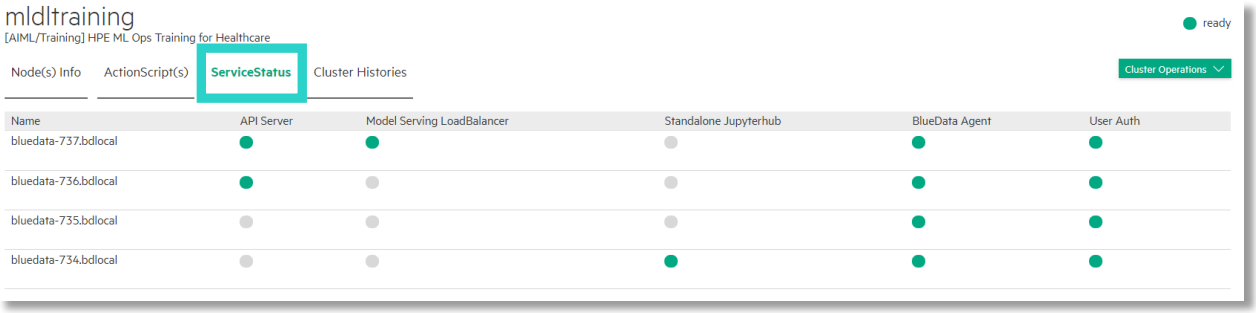
3. Check that the cluster is ready

4. Open Training Cluster details (Click on mldltraining)

5. Show the cluster nodes and services. (NotebookServer / TrainingEngine / …)

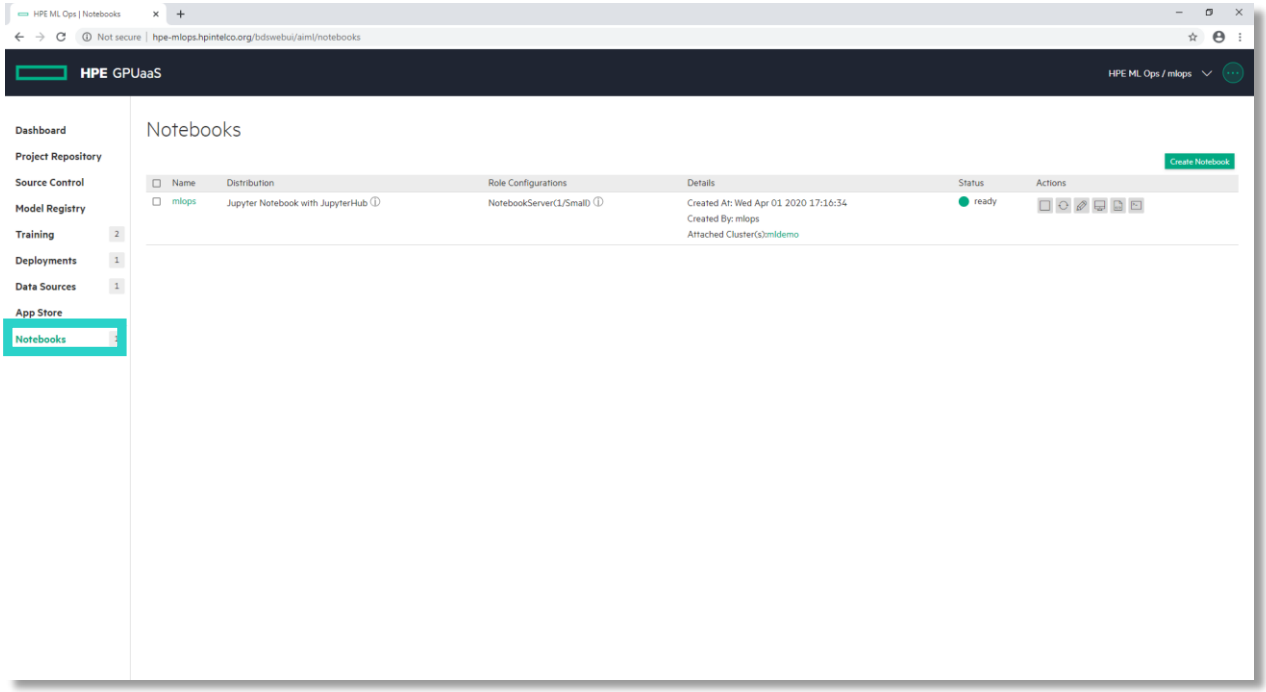


6. Check the Services Status of your cluster (Application / System services)

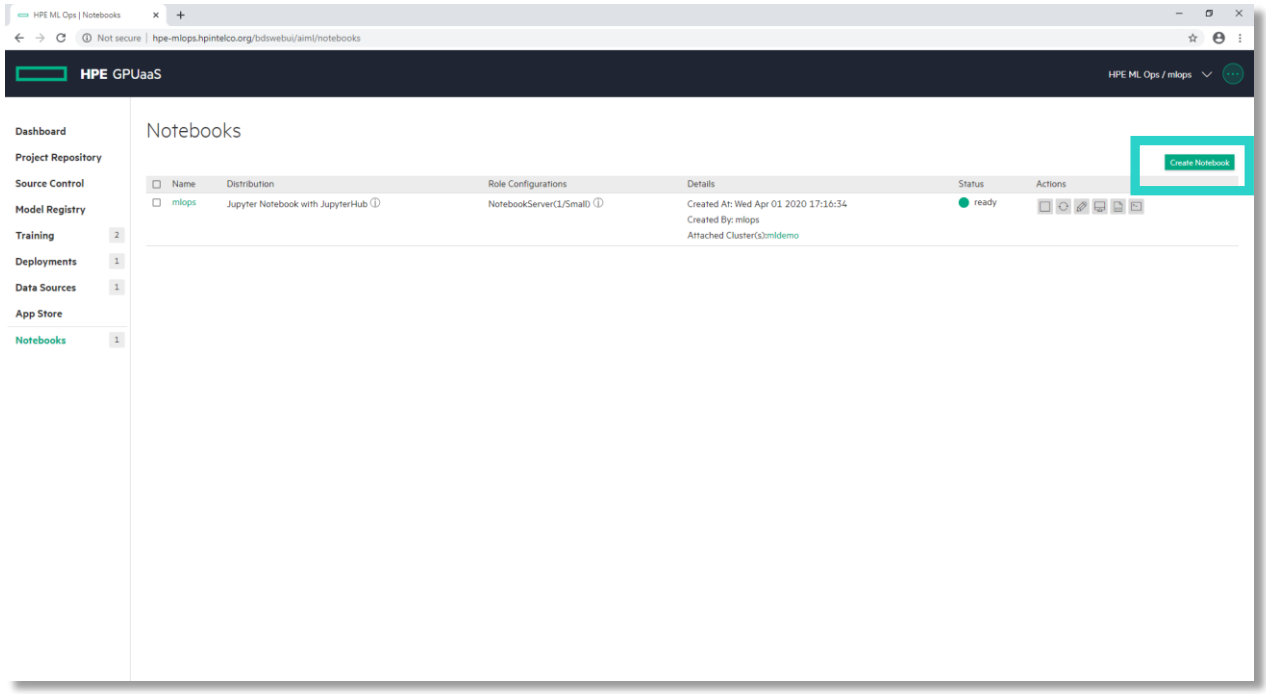


Develop/Create Jupyter Notebooks

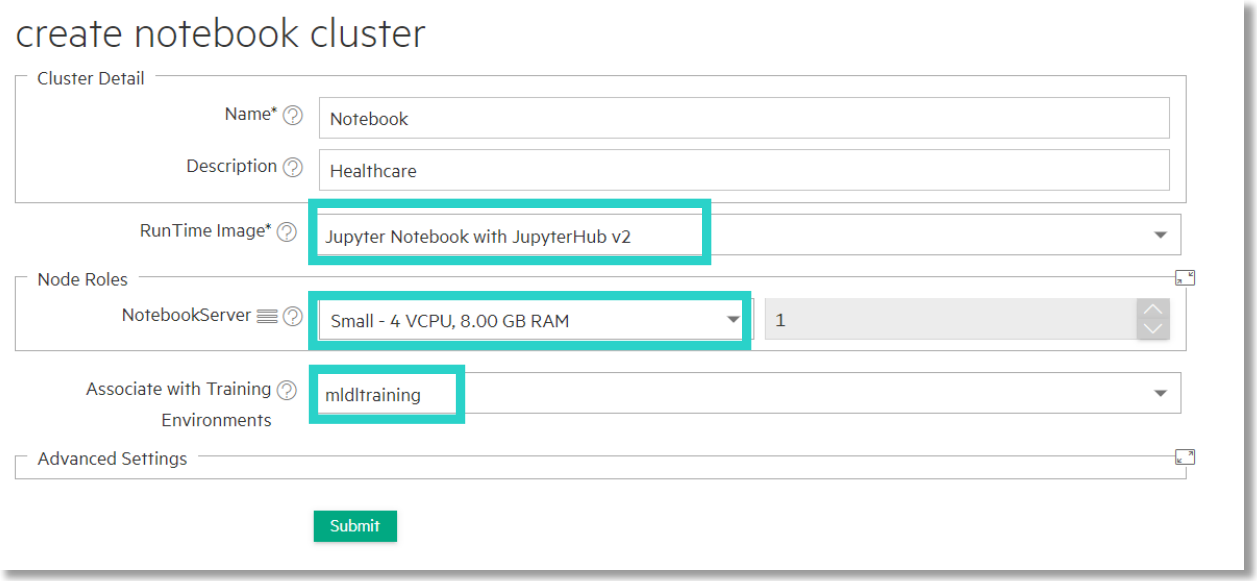
1. Select Notebooks from the left menu



2. Create a Jupyter Notebook (Click on Create Notebook)

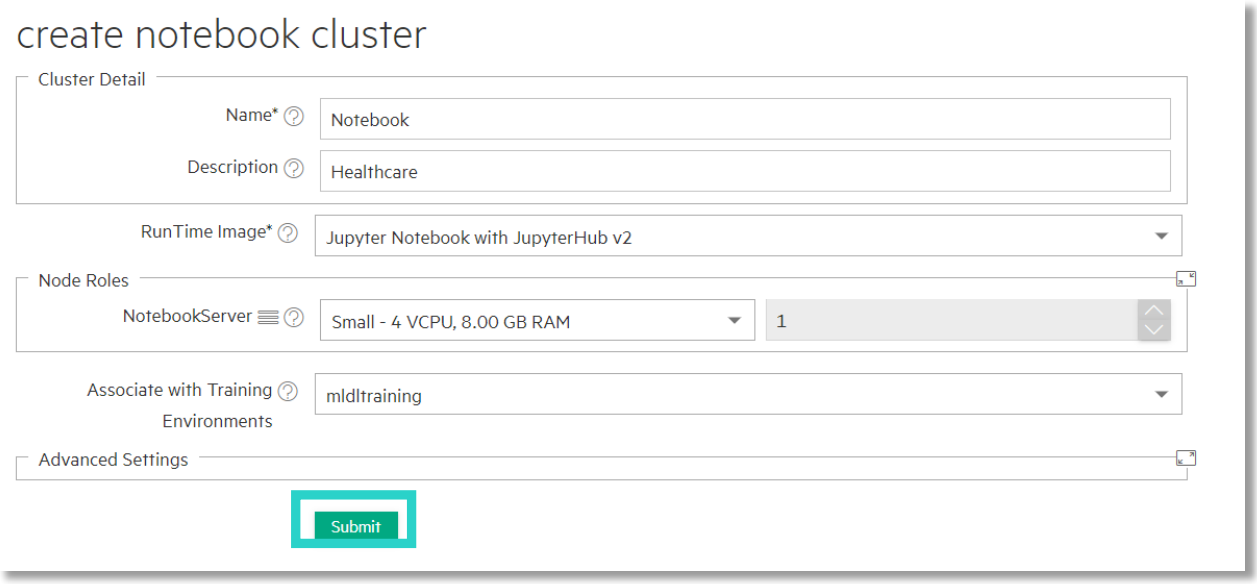


3. Associate the Notebook with your Training cluster

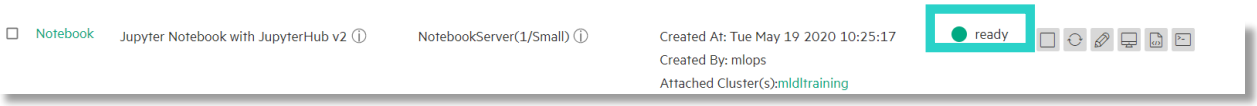


Don't forget to associate with Training Engine, setup CPU flavor and update the RunTime Image v2

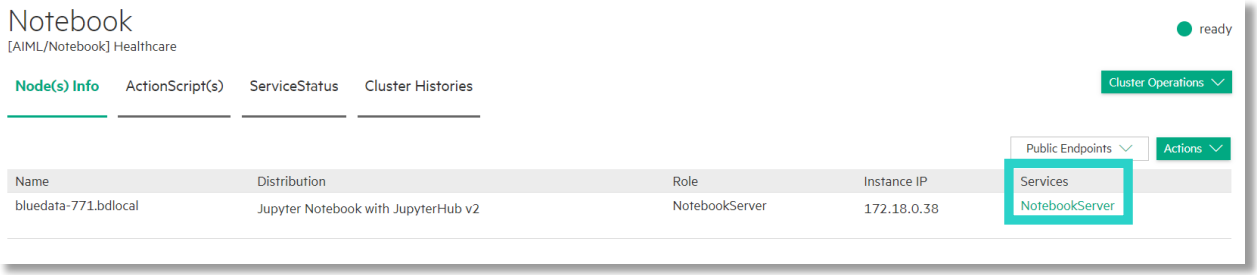
4. Create Notebook Cluster (Click on Submit)



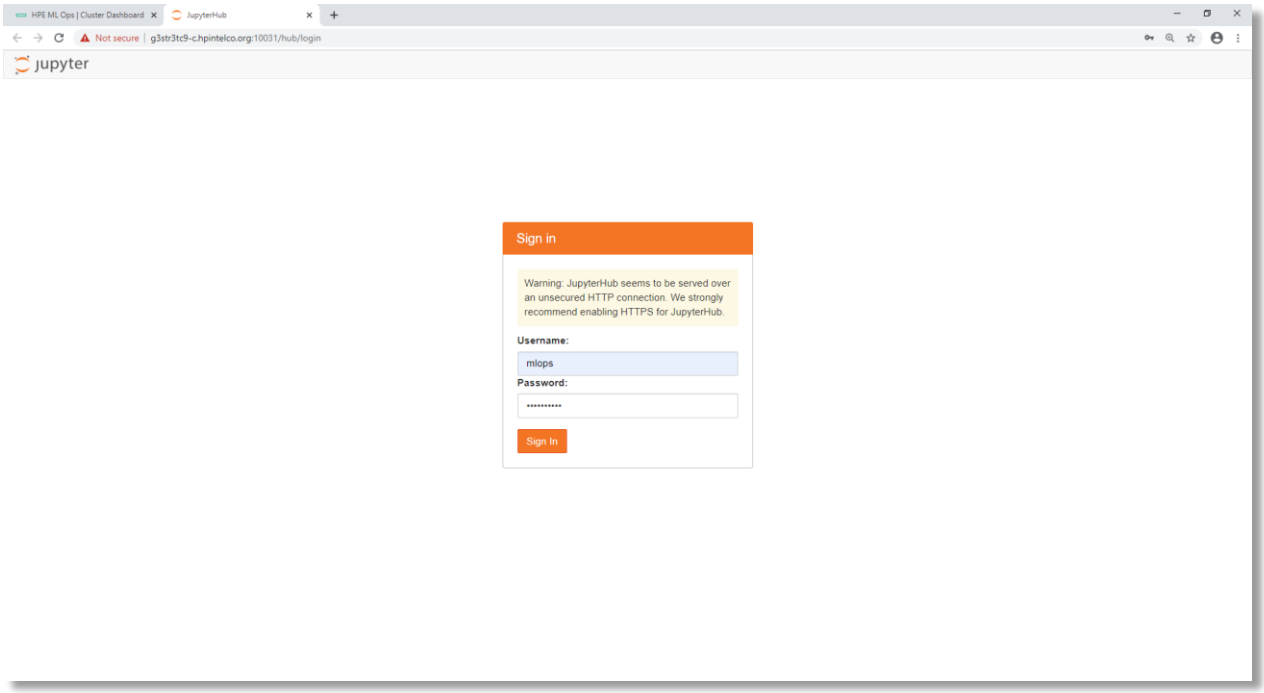
5. Make sure that the status of your notebook is ready



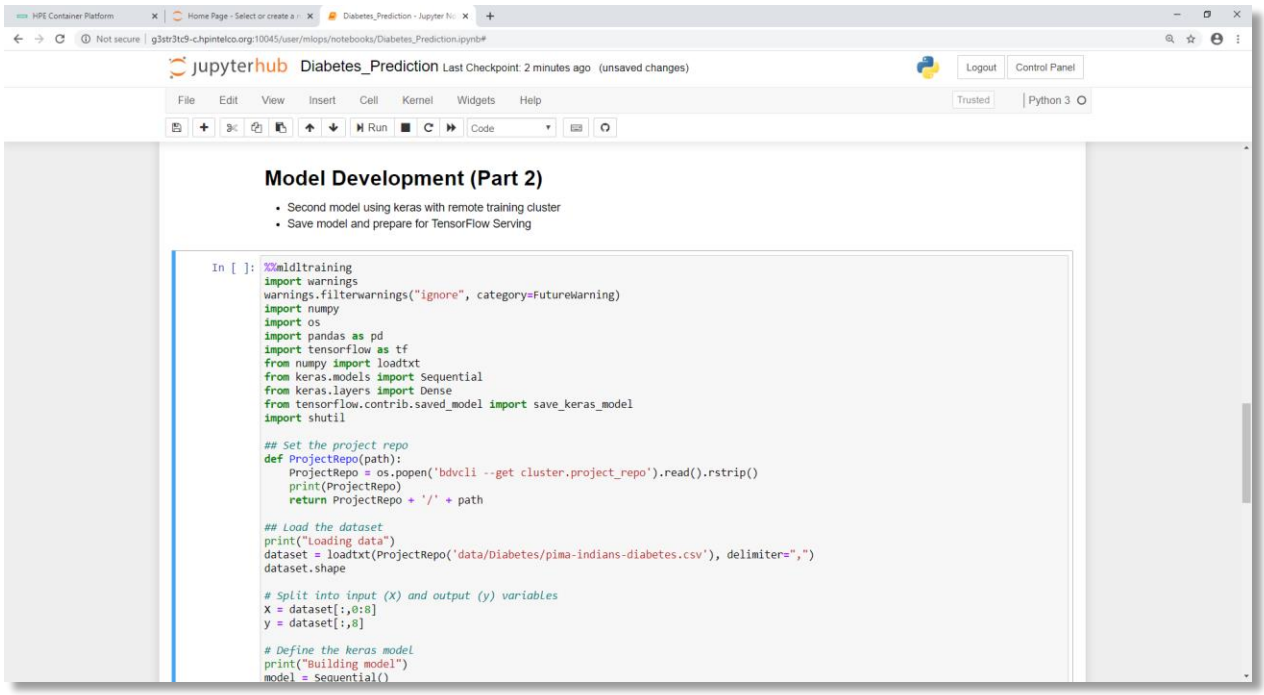
6. Open JupyterHub web ui (Click on NotebookServer)



7. Login in to the Jupyter WebUI as **mlops** user



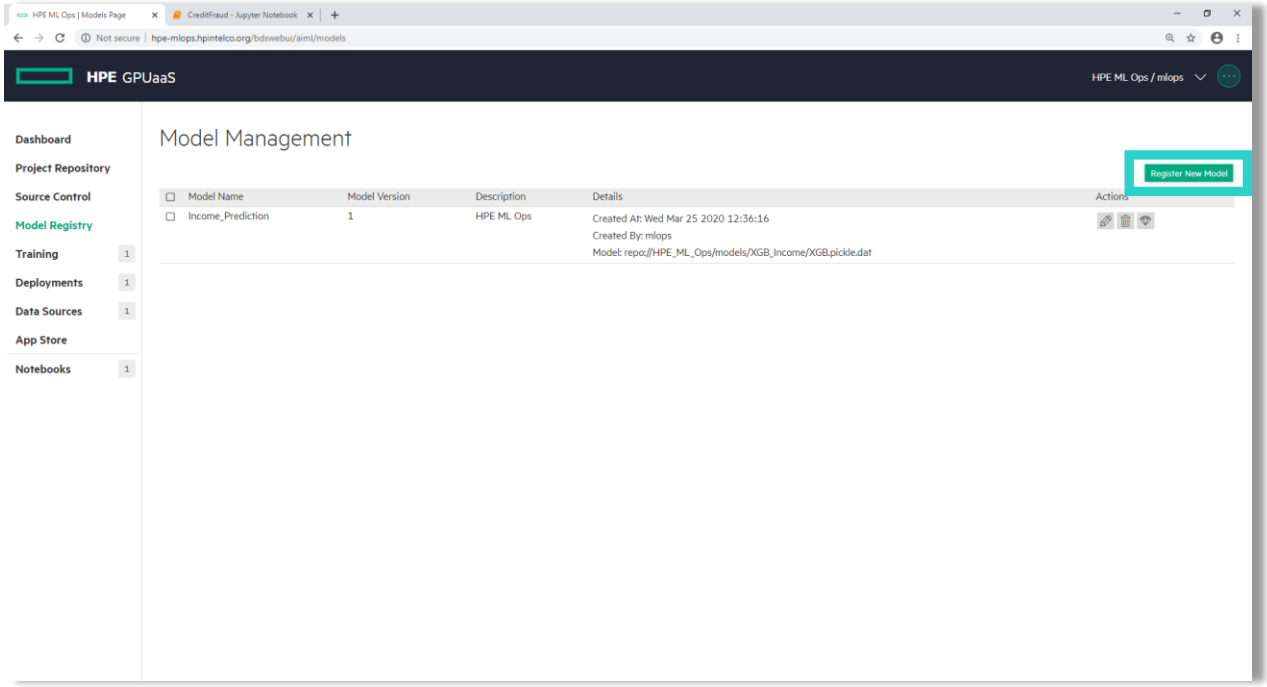
8. Open the Diabetes_Prediction.ipynb file, and then run each cell individually.



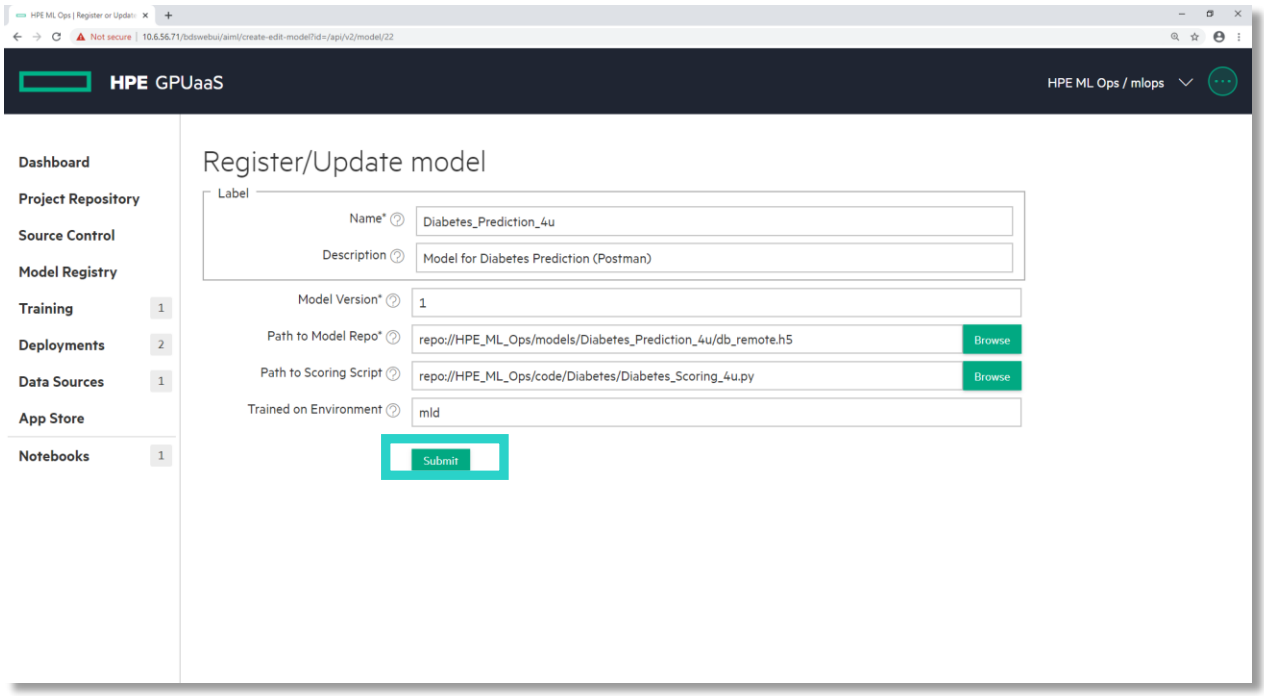
(Caution: Update the variable according to your Training Cluster including History urls.)

Register and Deploy the Model

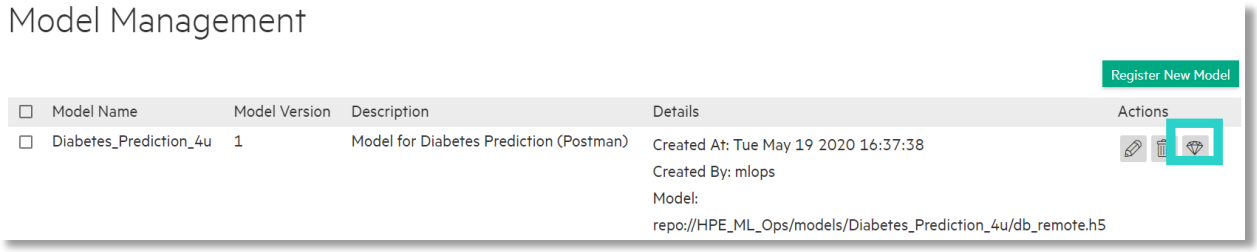
1. Select Model Registry from the left menu and click on Register New Model



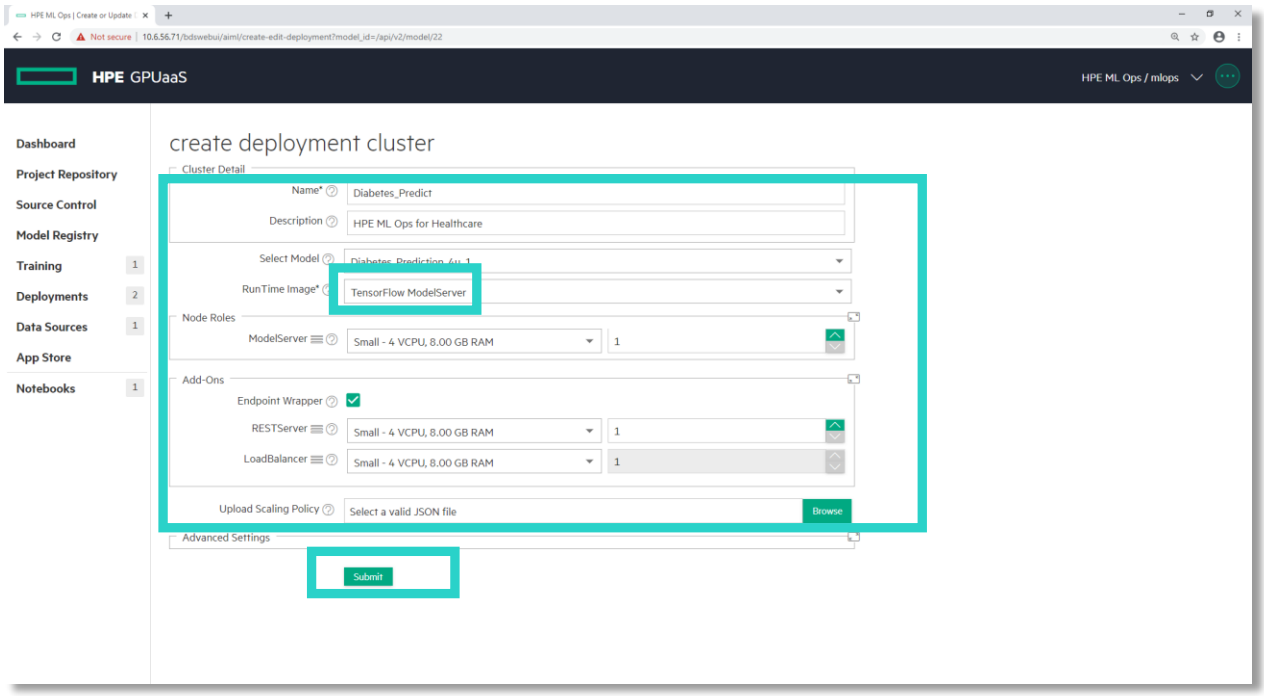
2. Register New Model using the information displayed below.



3. Deploy your registered model (Click on Deploy)

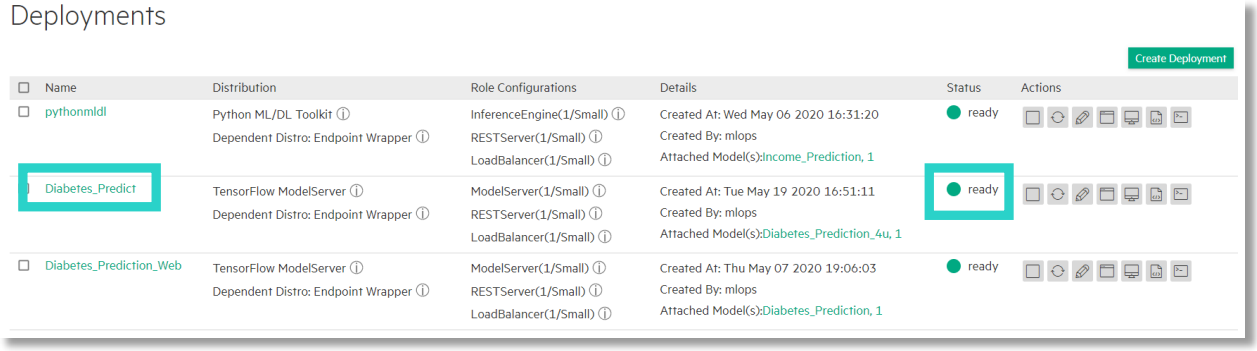


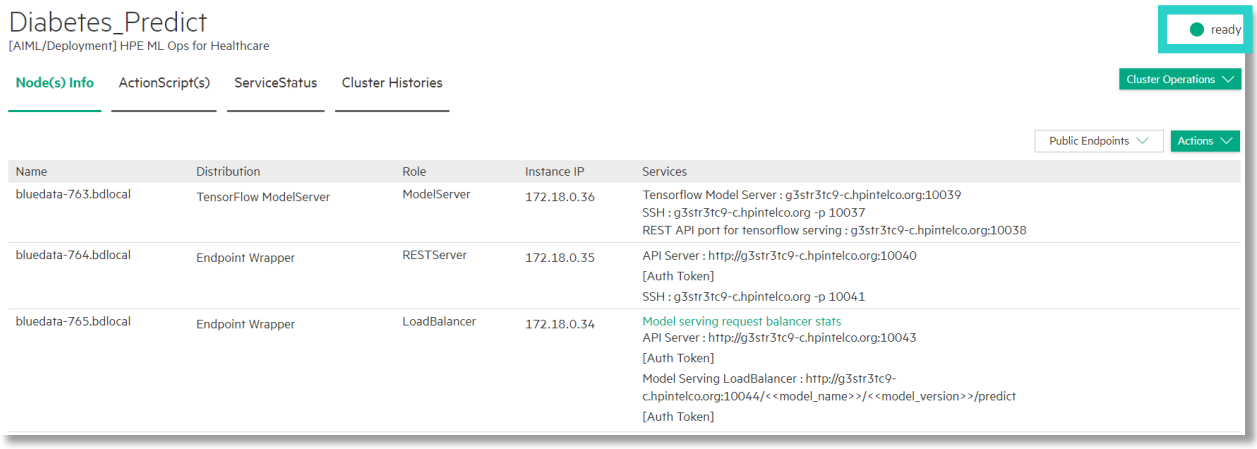
4. Create a Deployment Cluster for Diabetes Prediction (Click on Submit)



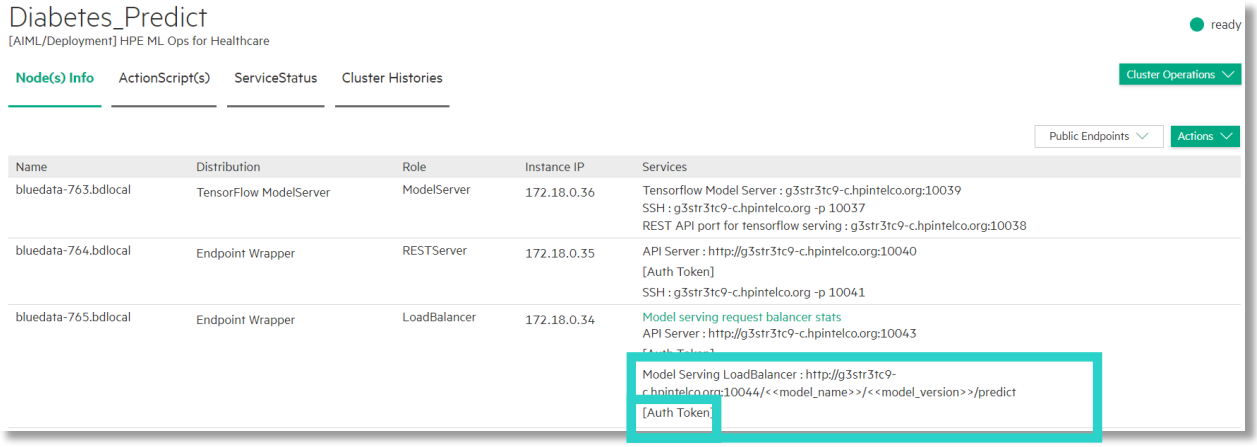
Don't forget to update the RunTime Image to TensorFlow ModelServer

5. Diabetes Prediction Cluster is now Ready





6. Generate your Prediction Request



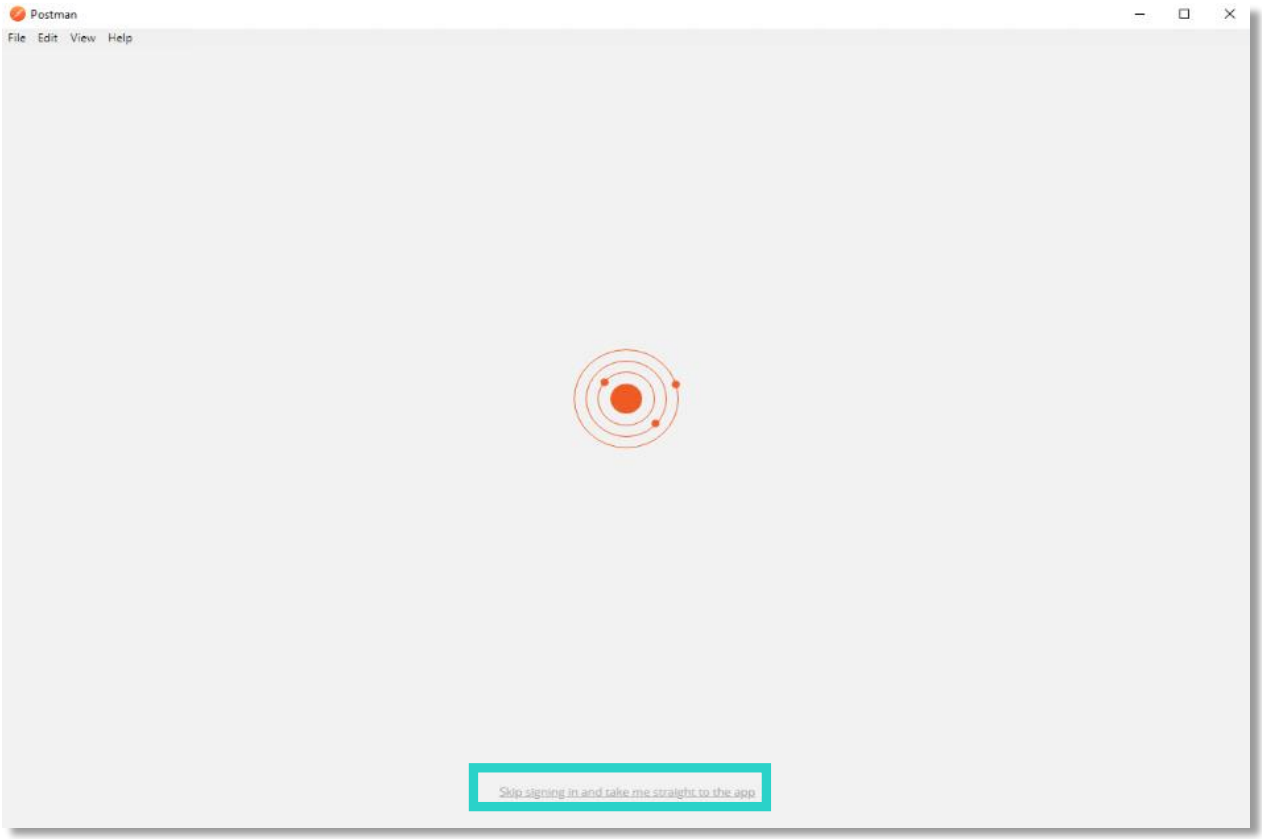
Copy the link and check update the port, model name and version

http://g3str3tc9-c.hpintelco.org:10044/Diabetes_Prediction_4u/1/predict
[Auth Token] Click to copy the token

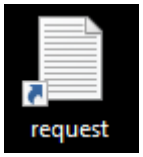
Run the Diabetes prediction using Postman

This link will start Postman!

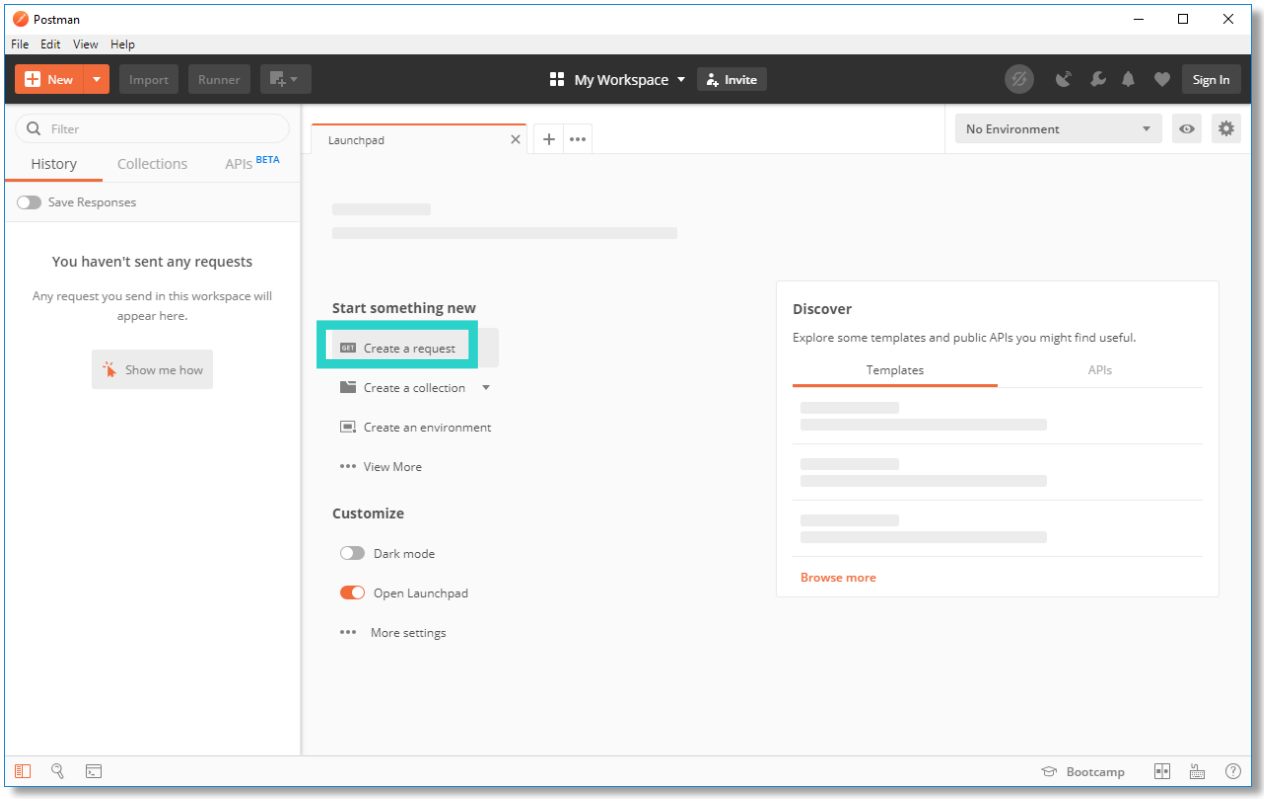
1. Select Skip signing in to go directly to the app



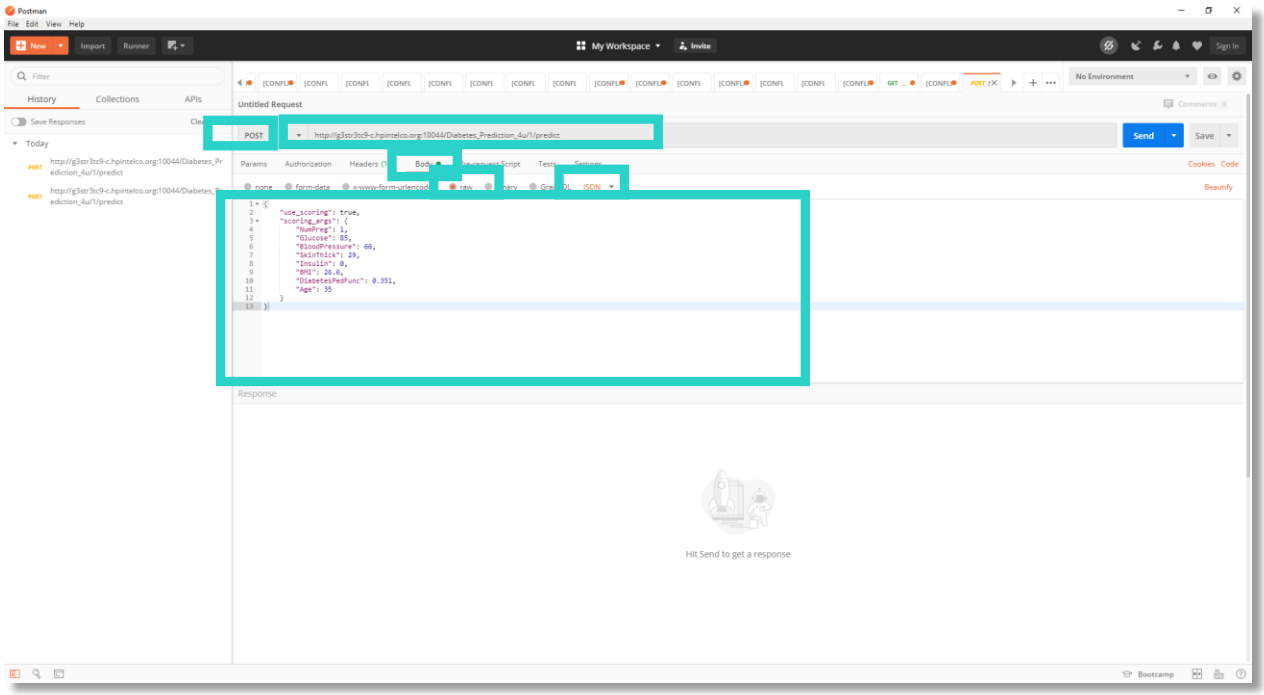
2. Copy the Json request in Postman body tab: Click on below icon



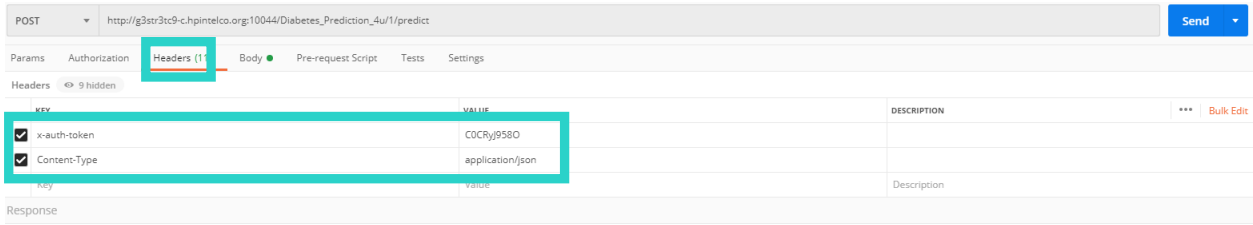
3. Create a new Request



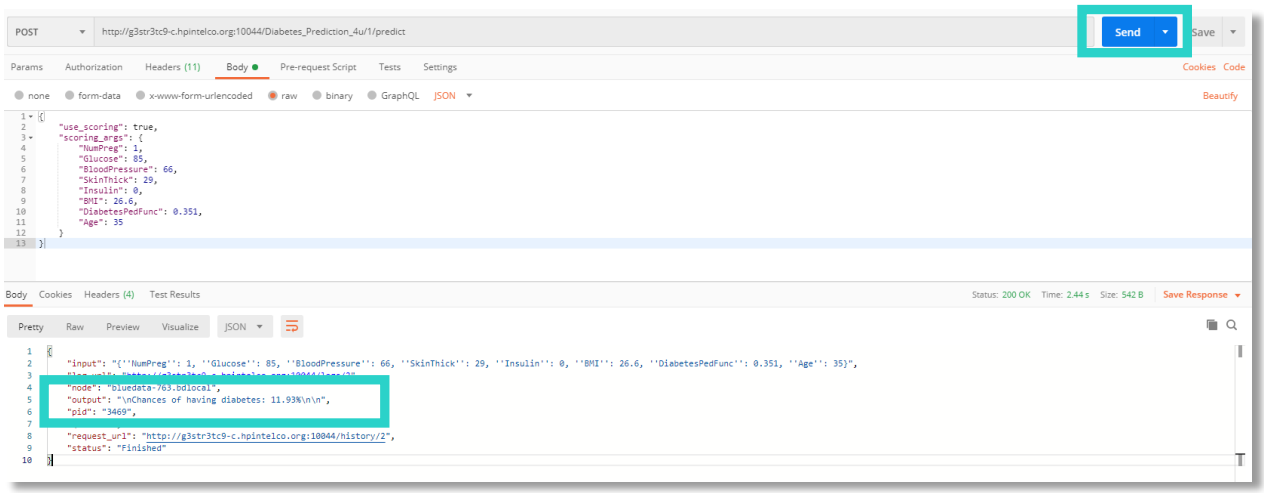
4. Fill in the Body tab to match the image below



5. Fill in the Header tab with your token

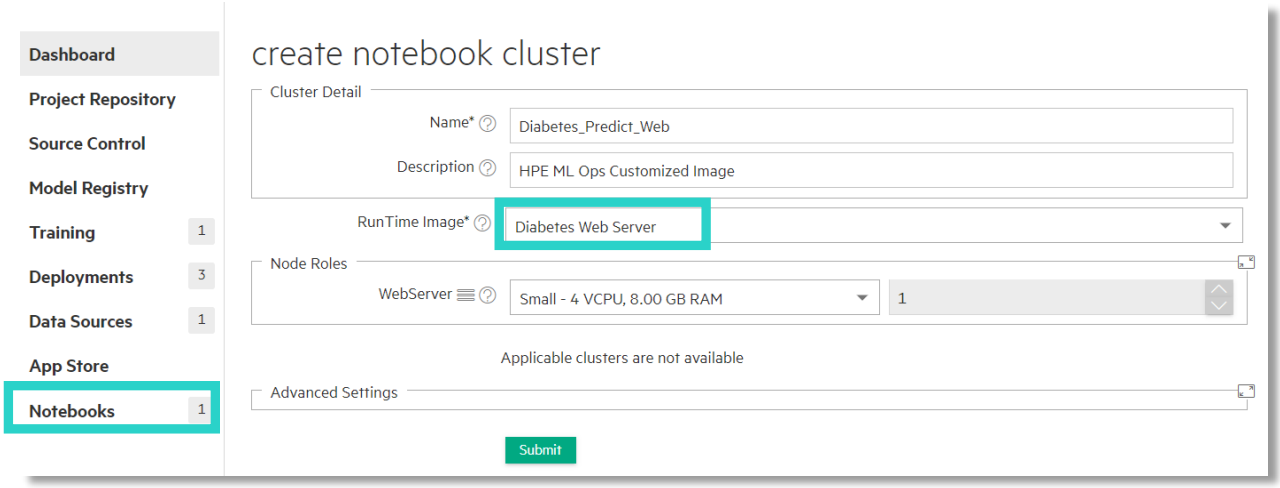


6. Send the post request to the cluster and get the Diabetes prediction result

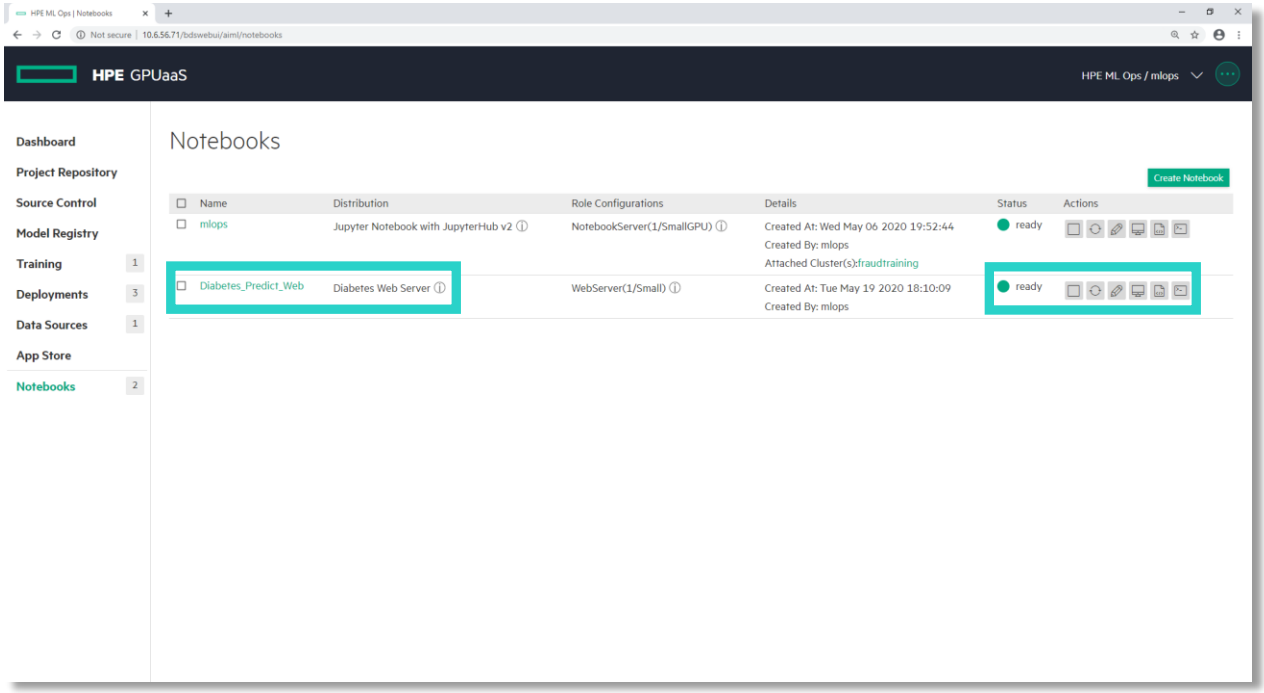


Deploy Customized Notebook Image

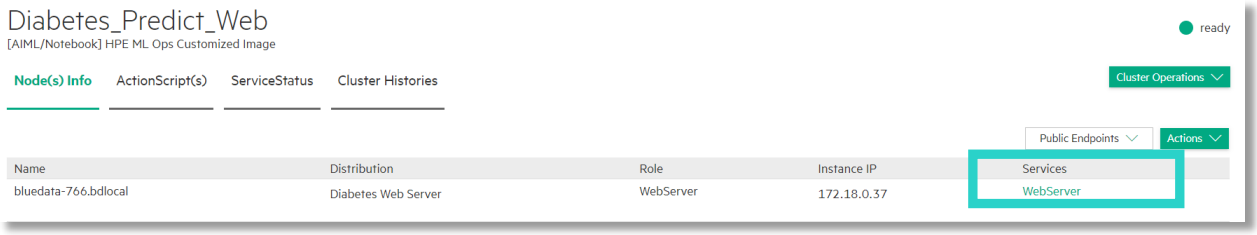
1. Create a new Notebook



2. Diabetes Prediction Notebook is now Ready

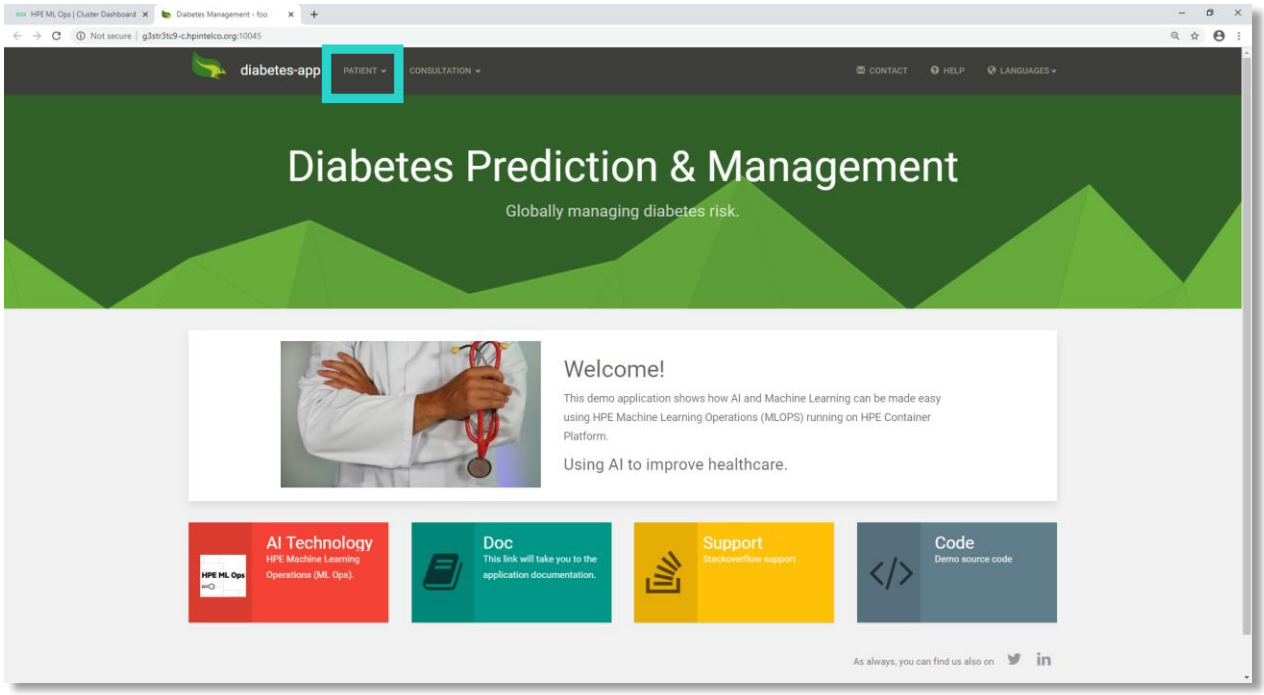


3. Check that the Notebook is ready

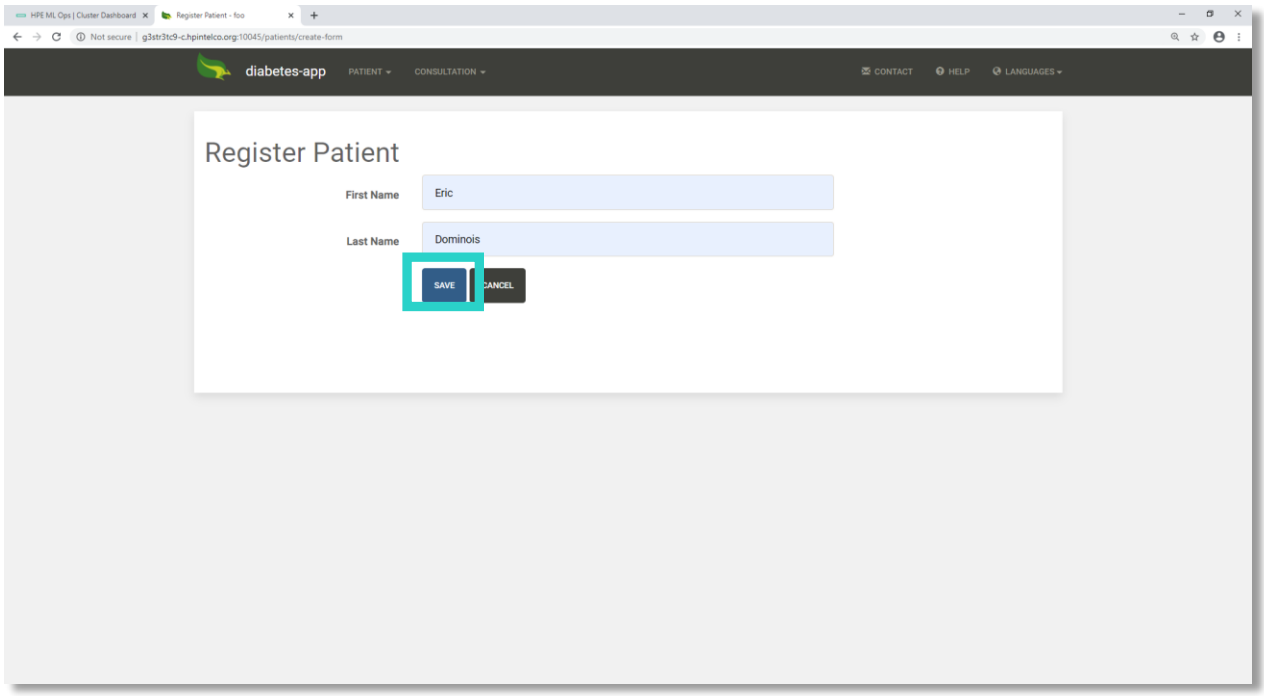


Run the Diabetes prediction using Customized Image

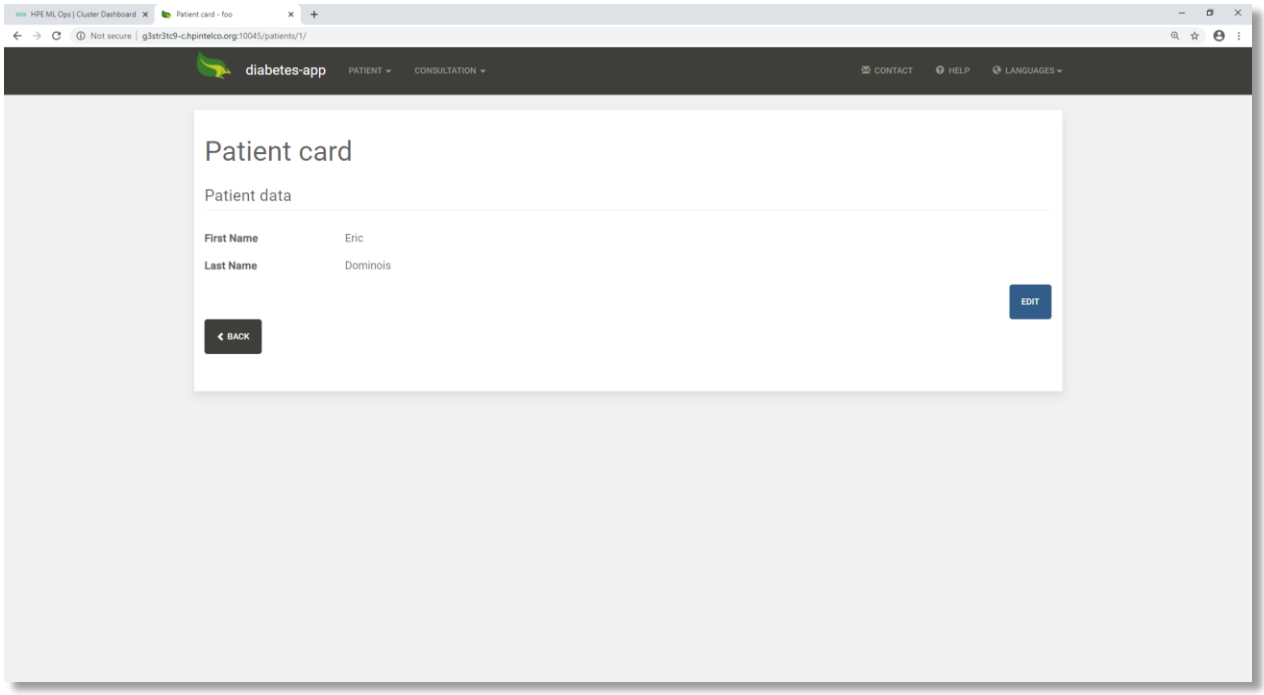
1. Register a new patient



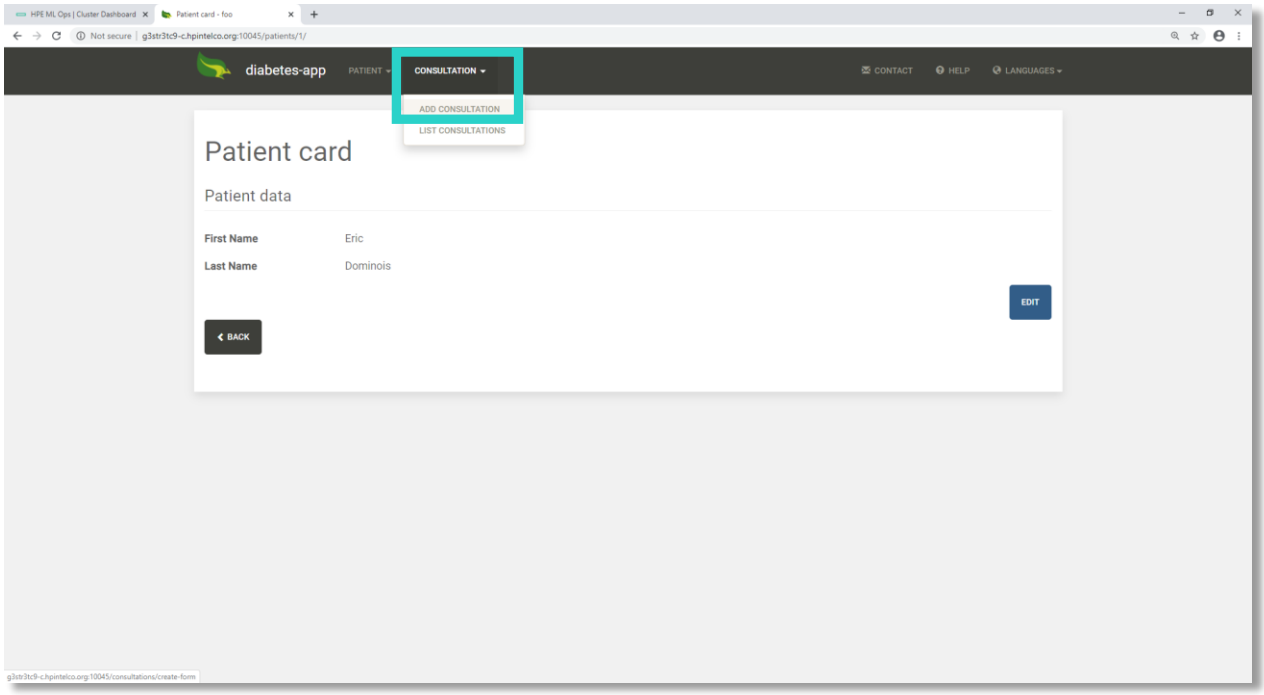
2. Fill in your first and last name and save



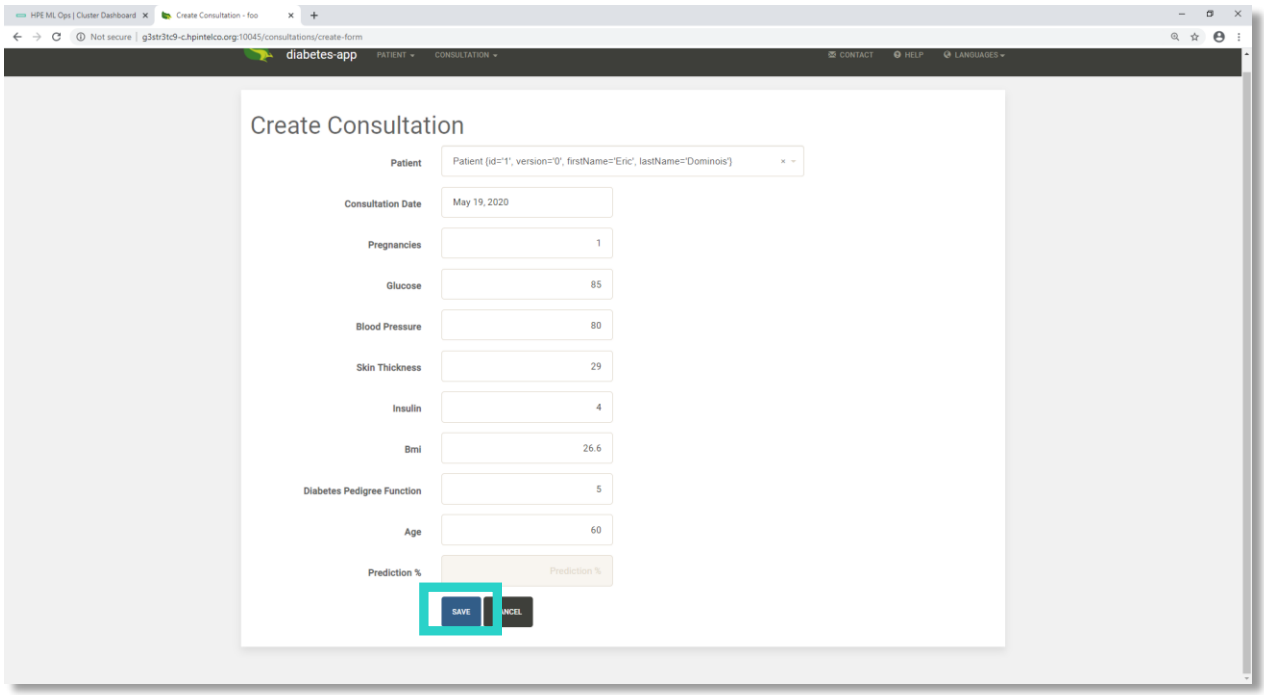
3. Patient card has been created



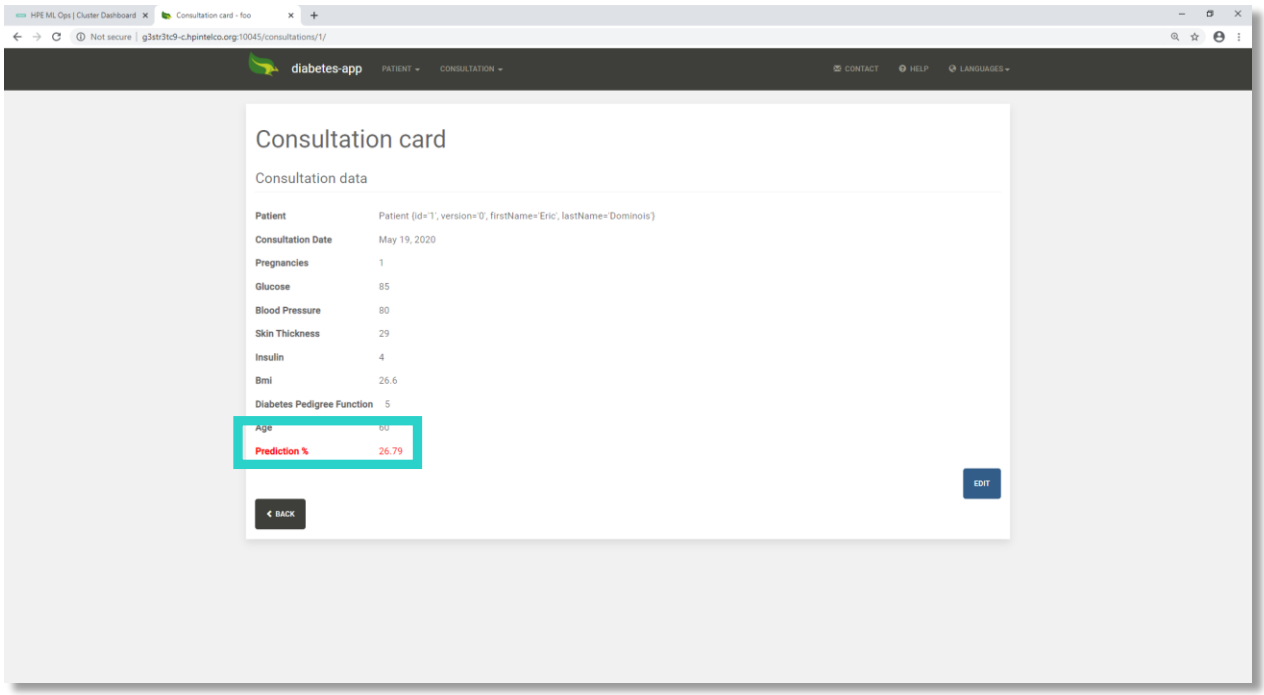
4. Add consultation



5. Create consultation



6. Consultation card has been created with Diabetes prediction score!



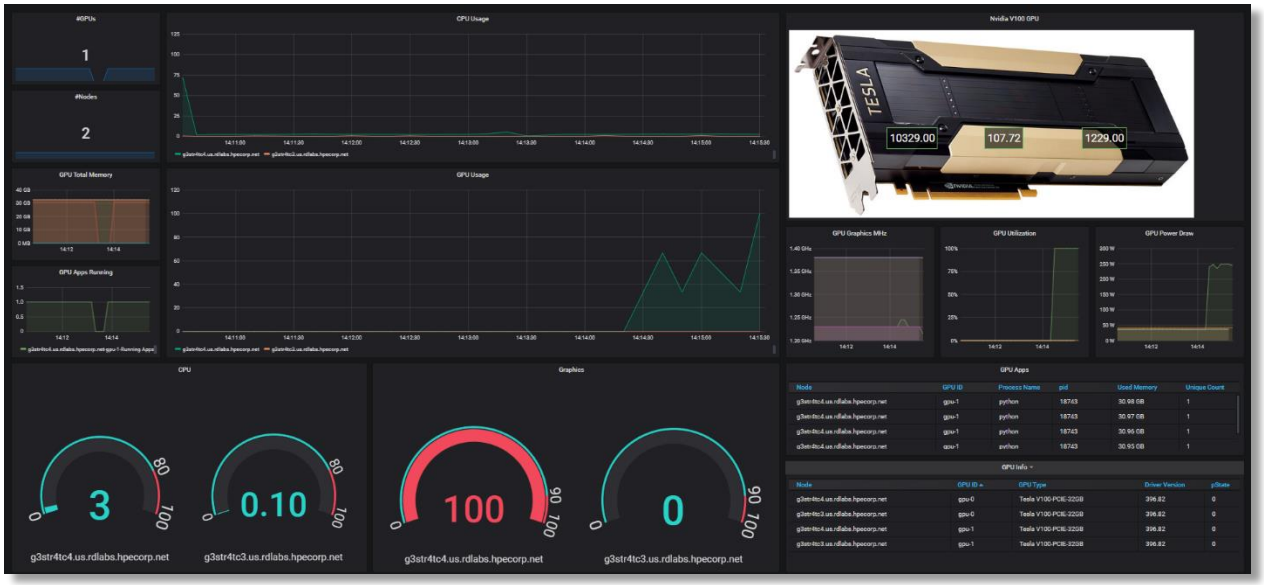
Contributors: ***Terry Chiang, Nanda Vijaydev, and Chris Snow***

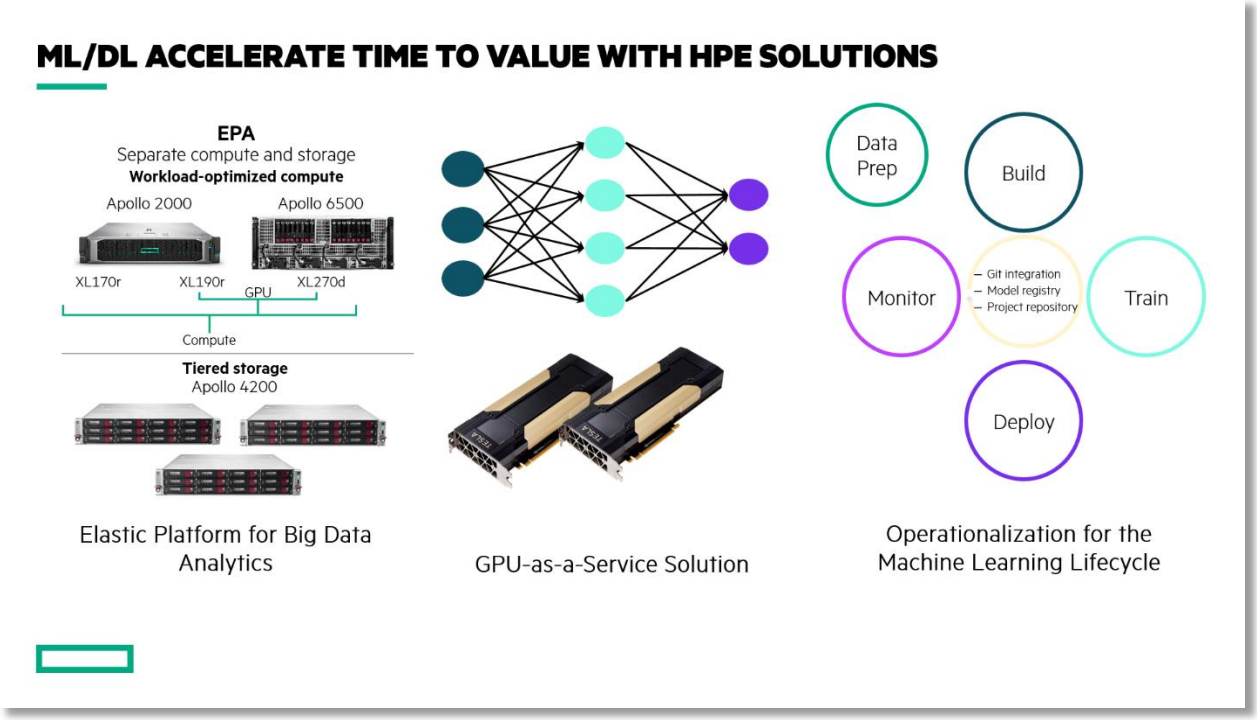
HPE Container Platform GPU Nodes Monitoring

Environment	Web UI	Username
HPE_ML_Ops	http://hpe-gpumon.hpintelco.org	cicdemo@hpintelco.org

Once connected to the Citrix Jumphost, open in your browser (chrome) the web ui login console, fill the user logon credentials provided on the BlueData_EPIC html page and click “Login”.

This link will start Chrome and open an html document with links and credentials!





Thanks!

– End of document –