A close-up of a computer program

AI-generated content may be incorrect.

**Table of Contents**

[1.0 Overview and Goals of Lab 1](#_Toc206236688)

[2.0 Lab prerequisites 2](#_Toc206236689)

[3.0 Machine Learning for z/OS 6](#_Toc206236690)

[3.1 Login to Machine Learning for z/OS (MLz) 6](#_Toc206236691)

[3.2 The IBM Machine Learning for z/OS main dashboard 8](#_Toc206236692)

[4.0 Import a Model into MLz 9](#_Toc206236693)

[4.1 MLz sample model 9](#_Toc206236694)

[4.2 Click on the Models tab on the MLz main dashboard 9](#_Toc206236695)

[4.3 Import model into MLz 10](#_Toc206236696)

[5.0 Deploy Your Model 13](#_Toc206236697)

[5.1 Using the Action Pulldown Menu 13](#_Toc206236698)

[5.2 Selecting the Deploy option 14](#_Toc206236699)

[5.3 Completing the Create deployment form 15](#_Toc206236700)

[5.4 Finding the model deployment ID 16](#_Toc206236701)

[6.0 Use python to access scoring endpoint via Jupyter Notebook 18](#_Toc206236702)

[6.1 Sign In to Jupyter Notebook 19](#_Toc206236703)

[6.2 Upload the sample python program to Jupyter Notebook 19](#_Toc206236704)

[6.3 Using Jupyter Notebook to load cars.ipynb file that contains our initial python code 20](#_Toc206236705)

[6.4 Updating code to talk to our scoring endpoint 21](#_Toc206236706)

[7.0 Summary 22](#_Toc206236707)

# 1.0 Overview and Goals of Lab

This lab will explore the process of importing a model and deploying a model on to z/OS, using that model in real-time from a python program via the scoring server endpoint.

Overall steps of the Lab:

1. Import the cars-test-model.tar.gz model into Machine Learning for z/OS (MLz for z/OS).
2. Deploy the cars-test-model.tar.gz model into Machine Learning for z/OS (MLz for z/OS) to obtain the scoring server endpoint.
3. Use a python script via Jupyter Notebook running on z/OS to query the scoring server for the model and quickly obtain result from the model.
   1. Use the REST API interface and JSON to interact with scoring server endpoint

# 2.0 Lab prerequisites

We will be using a Remote Desktop environment to access our z/OS system located behind an IBM firewall. To access the Remote Desktop, you will need to enter the following URL:

https://techzone.ibm.com/my/workshops/student/689f829646c8a302aef248d9

**A screenshot of a computer

AI-generated content may be incorrect.**

In the text field for the password/access code, enter: **aionz**

Now, you will be in a screen that displays your assigned environment:

A screenshot of a computer

AI-generated content may be incorrect.

Scroll down until you locate the blue rectangle with under the text VM Remote Console:

A screenshot of a computer

AI-generated content may be incorrect.

To start your remote desktop session, click on the link in the blue rectangle.

You will see a black screen pop-up and on top of the black area your will see buttons for Reboot, Shutdown, and Ctrl+Alt+Del in red:

A screenshot of a phone

AI-generated content may be incorrect.

Click on the Ctrl+Alt+Del red button:

A screenshot of a computer screen

AI-generated content may be incorrect.

Now, you will see the Windows Remote Desktop. By default, the Administrator user is selected and you will be prompted for a password.

Enter the password, **IBMDem0s**

A screenshot of a computer

AI-generated content may be incorrect.

In your Remote Desktop, in the Windows Search text area, type Cisco,

A computer screen with a blue background

AI-generated content may be incorrect.

Now, click on Cisco Secure Client, the Cisco Secure Client application will start.

A computer screen with a lock and text

AI-generated content may be incorrect.

If Cisco AnyConnect was previously set up for you, you will already see the correct server, **asa003b.centers.ihost.com**, on the field next to the Connect button. In that case, just click on Connect to establish the VPN connection to IBM.

\*\*\* **If successfully connected, you can skip the Cisco AnyConnect set up and move on to the “3.0 Machine Learning for z/OS” section. \*\*\***

If the server next to the Connect button is not **asa003b.centers.ihost.com**, replace it with: asa003b.centers.ihost.com,

Click Connect after replacing the VPN site name with asa003b.centers.ihost.com to view the Cisco Any Connect sign-on window that requests the username and password,

A screenshot of a computer

AI-generated content may be incorrect.

Use the usernames and passwords provided for access to the IBM VPN network.

After successfully establishing a VPN connection, you can proceed to the next section.

# 3.0 Machine Learning for z/OS

## 3.1 Login to Machine Learning for z/OS (MLz)

MLz is the flagship AI platform for z/OS. MLz:

* Provides a GUI that allows for model management, deployment and administration
* Supports hosting models using many different runtimes including Spark ML, Python ML, PMML scoring, and ONNX scoring
* Provides a model repository database, a Jupyter AI training environment, **RESTful endpoints for inference calls**, and high-performance integrated scoring capabilities for CICS and IMS applications
* Provides micro-batching capabilities to maximize inferencing efficiency by allowing for multiple requests to be bunched together
* Will also optimize the model to leverage the new Integrated Accelerator for Analytics (AIU) whenever possible and if available

We are using the full MLz 3.2 Enterprise Edition in our lab today.

Let’s login to our MLz and get ready to **import** our model.

In the URL field of your web browser, type the following:

<https://129.40.117.161:9888>

Note that this is an actual z/OS IP address as MLz runs as a z/OS executable. MLz also provides its own Jupyter framework and we will use it in this lab. Although we will not export any models in this lab, the framework provides the ability to export models directly into MLz from the notebook.

After entering the URL in your browser, you will end up in the following MLz Log in screen:

A screenshot of a computer

AI-generated content may be incorrect.

For both Username and Password fields, type in MLz login credentials provided for this lab all in lowercase. Click the **Log in** button.

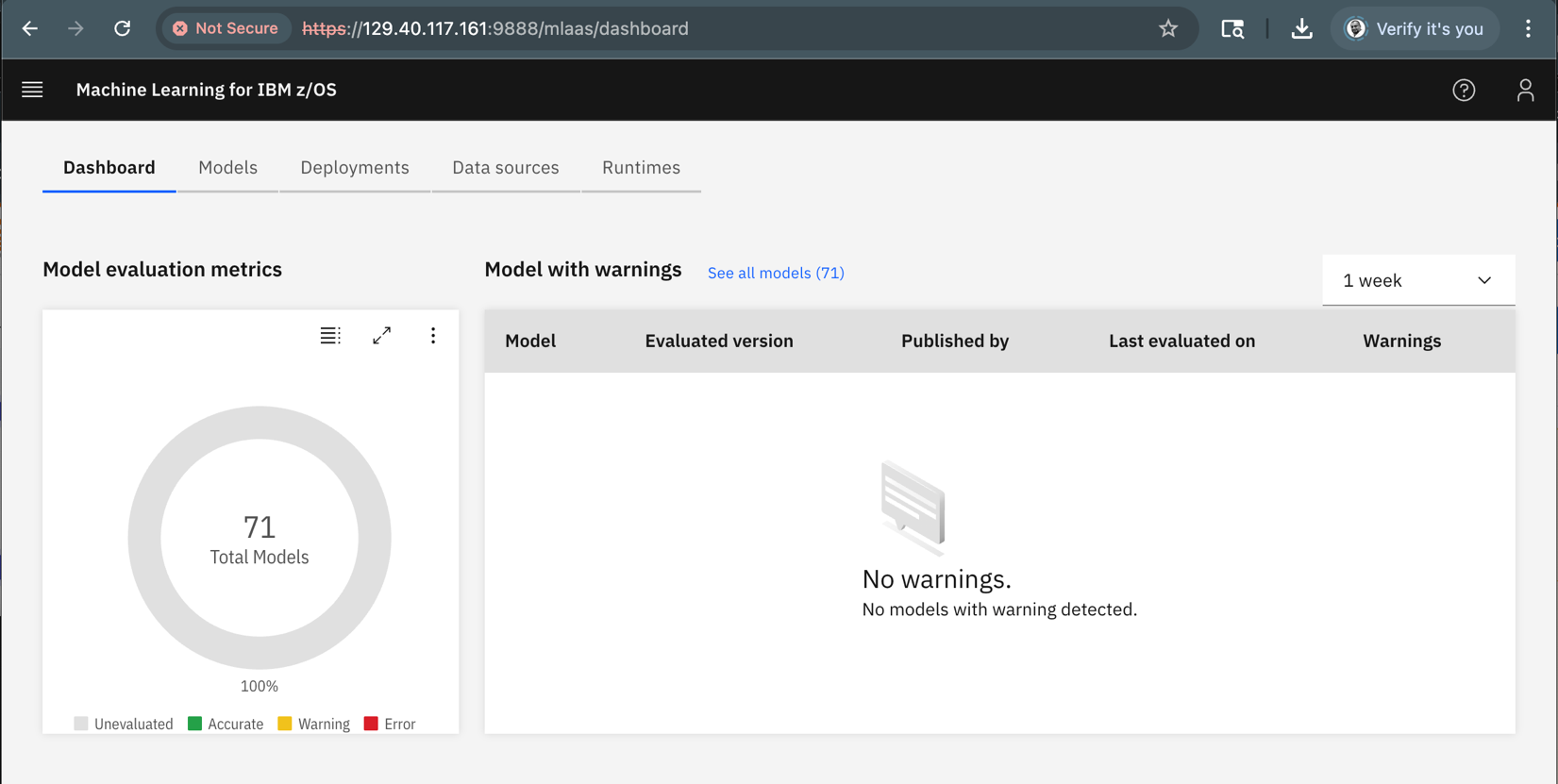
**NOTE:**

**The username for MLz and Jupyter Notebook has the following format: aizos0xx where xx are digits assigned to each student.**

**The password for all usernames for MLz and Jupyter Notebook is z17aionz all in lowercase.**

## 3.2 The IBM Machine Learning for z/OS main dashboard

## 



The Machine Learning for z/OS dashboard is a web-based interface use for managing and administering Machine learning for the z/OS platform. The main dashboard provides the tools necessary for model management, user management and system administration.

**NOTE: The files needed for this workshop including this handout, the cars-test-model.tar.gz, and the cars.ipynb are available in the following gitHub repository:**

[**https://www.github.com/WalterManrique/cars-weight-model-onz**](https://www.github.com/WalterManrique/cars-weight-model-onz)

# 4.0 Import a Model into MLz

## 4.1 MLz sample model

The model that we will use in this workshop is a sample model that is included in MLz 3.2 and it can be found in the $IML\_INSTALL\_DIR/alnsamp directory under z/OS. The **cars-test-model.tar.gz** contains the sample model. It is a sample **regression** model that can be used to predict the weight of a car.

We have already downloaded the model to your Windows environment. It is in the same folder together with the **cars.ipynb** program that we will use later in the workshop.

## 4.2 Click on the Models tab on the MLz main dashboard

You will then be presented with the Models Table Screen.

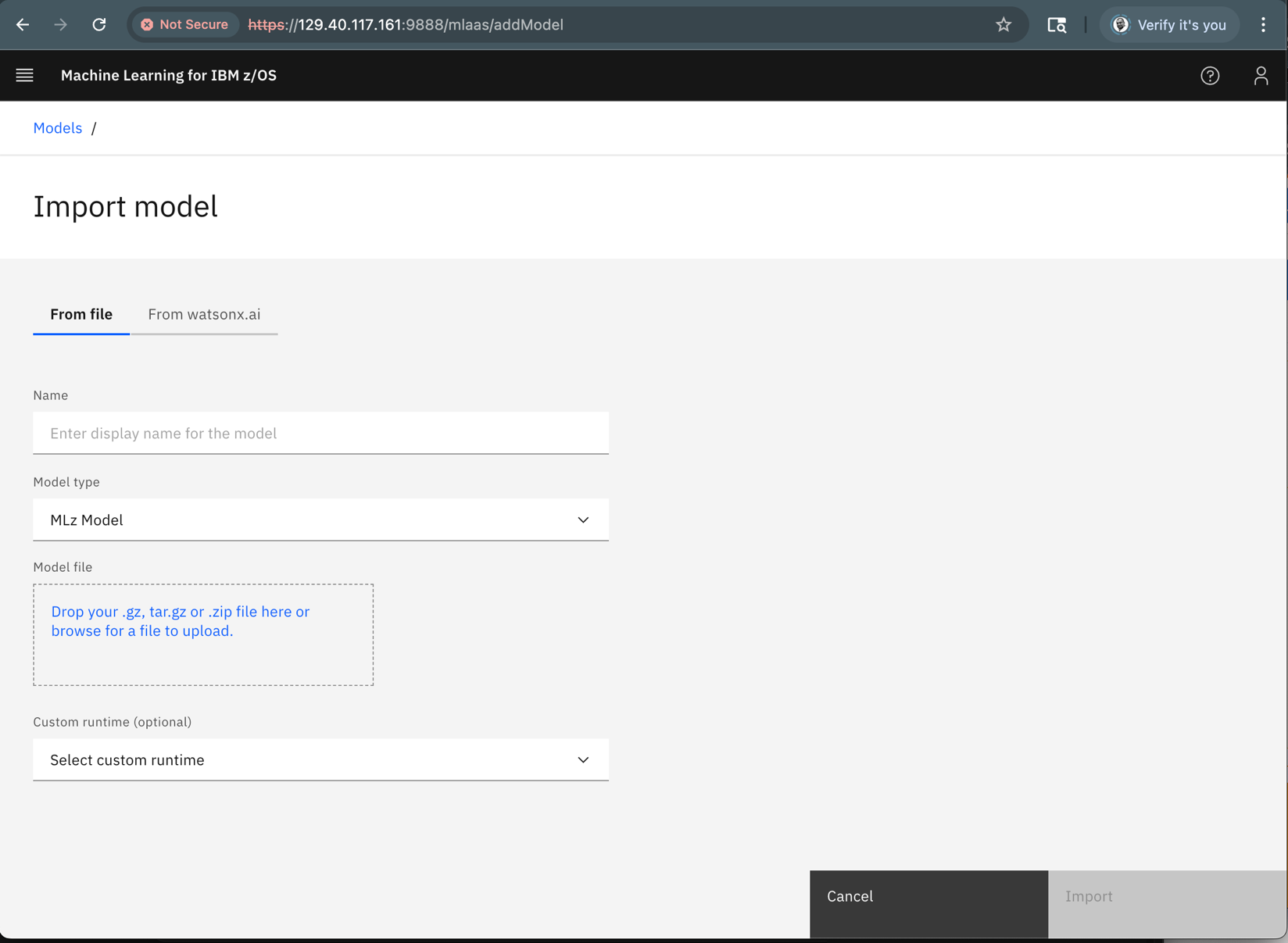
A screenshot of a computer

AI-generated content may be incorrect.

The Models Table Screen has blue button with the “Import model” text. We will click on it to import our model in the next section.

## 4.3 Import model into MLz

Click on the blue “Import model” button to navigate to the Import Model screen.



* On this screen, click the “**From file”** tab if not already selected
* Enter the name for the new model in the **Name** field. To avoid name clashing with other lab participants, we suggest naming your model **cars<nn>,** where <nn> represents the last two digits of the **aizos0*xx*** userid that you used to login to MLz earlier. For example, if your userid was aizos001, you would name your model cars01
* There is no need to update the **Model type** field
* In the **Model file** field, drag the file named cars-test-model.tar.gz from the Downloads folder to the ‘Model file” field to upload the model source file to MLz. Alternatively, you can click the blue text in the **Model file** box and use the interactive file explorer to search the file manually.

After completing all the previous steps, your screen should look something like this:

A screenshot of a computer

AI-generated content may be incorrect.

Note that the Import button at the bottom right of the browser is now blue.

Click the **Import** blue button in the bottom right-hand corner to import the selected model.

Importing takes the model, compiles it and generates z/OS executable code using the best hardware available.

Once the model is imported, the browser will display the list of models. The latest imported model will be listed on top:

A screenshot of a computer

AI-generated content may be incorrect.

Verify that your model is listed and no errors are reported. There will be a green check if there were no errors.

Now, you are ready to deploy your newly imported model.

# 5.0 Deploy Your Model

By now, you have imported a model into MLz. The model shows a green status which means it is ready to be deployed.

The model needs to be deployed to a scoring service managed by MLz. You can think of the scoring service as the location where the model will run.

There are different types of scoring services and depending on the use case for the deployment of your models, you can have different scoring service configurations. The scoring service be configured,

* on a single z/OS image running in a Liberty server managed by MLz
* for High Availability (HA), which gives it the ability to leverage DDVIPAs on the system to round-robin the inferencing requests across multiple LPARs in your SYSPLEX
* to run inside a Liberty server within a CICS region, configured for normal or HA
* and much more…

Let’s deploy our model. From the list of models, find your model, and find the **Action** column on the right.

## 5.1 Using the Action Pulldown Menu

Click on the 3 stacked dots (vertical ellipsis) to get a list of actions we can perform on this model.

A screenshot of a computer

AI-generated content may be incorrect.

## 5.2 Selecting the Deploy option

Select **Deploy** from the pulldown menu.

You will now be taken to the **Create Deployment** page.

A screenshot of a computer

AI-generated content may be incorrect.

## 5.3 Completing the Create deployment form

Fill in the following values:

* For **Deployment name**, use the same naming convention as before with your two number suffix, concatenating that number at the end of the word cars.

E.g., **cars<nn>.** If you were assigned, aizos001, this value would be **cars01.**

* For **Deployment type**, the value will be **Online**.
* For **Model version**, there should be only one, so the only selectable value will be **1.**
* **For Scoring service (standalone or cluster),** we are going to deploy this model into a Liberty server that leverages the **WebSphere Optimized Local Adapter** (WOLA) high-speed communications. Select **WOLASCOR** as our scoring service.
* There are some other deployment options but those are beyond the scope of this lab.
* Once you have filled in all the fields, click **Create** in the bottom right-hand corner to create the deployment.

A screenshot of a computer

AI-generated content may be incorrect.

Let’s verify that the model is deployed and that cars<nn> is listed under the Deployments tab.

A screenshot of a computer

AI-generated content may be incorrect.

## 5.4 Finding the model deployment ID

Each model deployed is given a unique deployment ID that is used by the model invoker application to tell MLz which deployed model to use for the scoring request. We will need the deployment model ID when we invoke the model from our python program in the following steps.

To find the deployment ID, click on the name of your deployed model on the deployments page.

A screenshot of a computer

AI-generated content may be incorrect.

After clicking on the deployed model, you will end up in the deployment details page which includes a scoring endpoint. The last qualifier of the scoring endpoint URL is deployment ID.

A screenshot of a computer

AI-generated content may be incorrect.

Copy the endpoint value by clicking on the copy icon next to the scoring point URL and save it.

A screenshot of a computer

AI-generated content may be incorrect.

You can save the scoring point URL in a notepad document, or you can come back to the deployment page to copy the scoring point URL when needed.

# 6.0 Use python to access scoring endpoint via Jupyter Notebook

Now, we are ready to communicate with our deployed scoring model via the scoring endpoint using a REST API call. We will write a simple application in python to query the model. Our python program will execute from a Jupyter Notebook running on z/OS.

Navigate to the Jupyter Notebook by entering the following URL in the address field in your browser.

<https://129.40.117.161:9900/>

After entering the URL listed, you will be in the following Sign In page for the jupyterhub.

A screenshot of a computer

AI-generated content may be incorrect.

## 6.1 Sign In to Jupyter Notebook

For the Username and Password, type in the Jupyter Notebook login credentials provided in lowercase. Click on Sign In button.

Initially, after you sign in, your browser will not list any notebooks.

A screenshot of a computer

AI-generated content may be incorrect.

## 6.2 Upload the sample python program to Jupyter Notebook

Click on the Upward Pointing Arrow to upload your cars.ipynb notebook into the z/OS environment. Use the file explorer to select the cars.ipynb file and click on open.

A screenshot of a computer

AI-generated content may be incorrect.

## 6.3 Using Jupyter Notebook to load cars.ipynb file that contains our initial python code

Now that the cars.ipynb file has been uploaded to z/OS, a new entry will be displayed under the Name column on the left-hand side of your browser. Double click on the filename to access the code. The code will be displayed on the right-hand side of the browser window.

A screenshot of a computer

AI-generated content may be incorrect.

## 6.4 Updating code to talk to our scoring endpoint

At this point, we will follow the instructions on the Jupyter Notebook.

Remember to use the scoring endpoint for your deployed model. Now that the cars.ipynb file has been uploaded to z/OS, a new entry will be displayed under the Name Fill in the following values:

How to execute program after it is updated

# 7.0 Summary

So what did we just do?

1. We imported a model into Machine Learning for z/OS
2. We deployed the model into our Liberty server scoring service with WOLA for ultrafast scoring calls
3. We executed a python program from a Jupyter Notebook running on z/OS to query our deployed model using the deployed model’s scoring endpoint via a REST API call
4. We became familiar with Machine Learning for z/OS

Imagine the possibilities you have with a scoring service natively running on z/OS!

We hope you enjoyed our introductory Lab. If you think your company would be interested in a no-charge AI on IBM Z Discovery workshop, please feel free to drop us a note at aionz@us.ibm.com. Here is an explanation of what are the possibilities for a custom workshop experience to learn about how you can start leveraging AI on the zSystems platform.