INTERNAL

SAP AI Core hands-on exercises

AI220-EXERCISE02-BYOM_SOUND_BASED_PDM_WITH_SAP_AI_LAUNCHPAD

This document will guide you step-by-step through the process of training and implementing an AI model in SAP AI Core by using SAP AI Launchpad as a client.

www.sap.com/contactsap

THE BEST RUN SAP®

Table of Contents

DISCLAIMER

The information shared in this document is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. All functionality presented here is subject to change and may be changed by SAP at any time for any reason without notice.

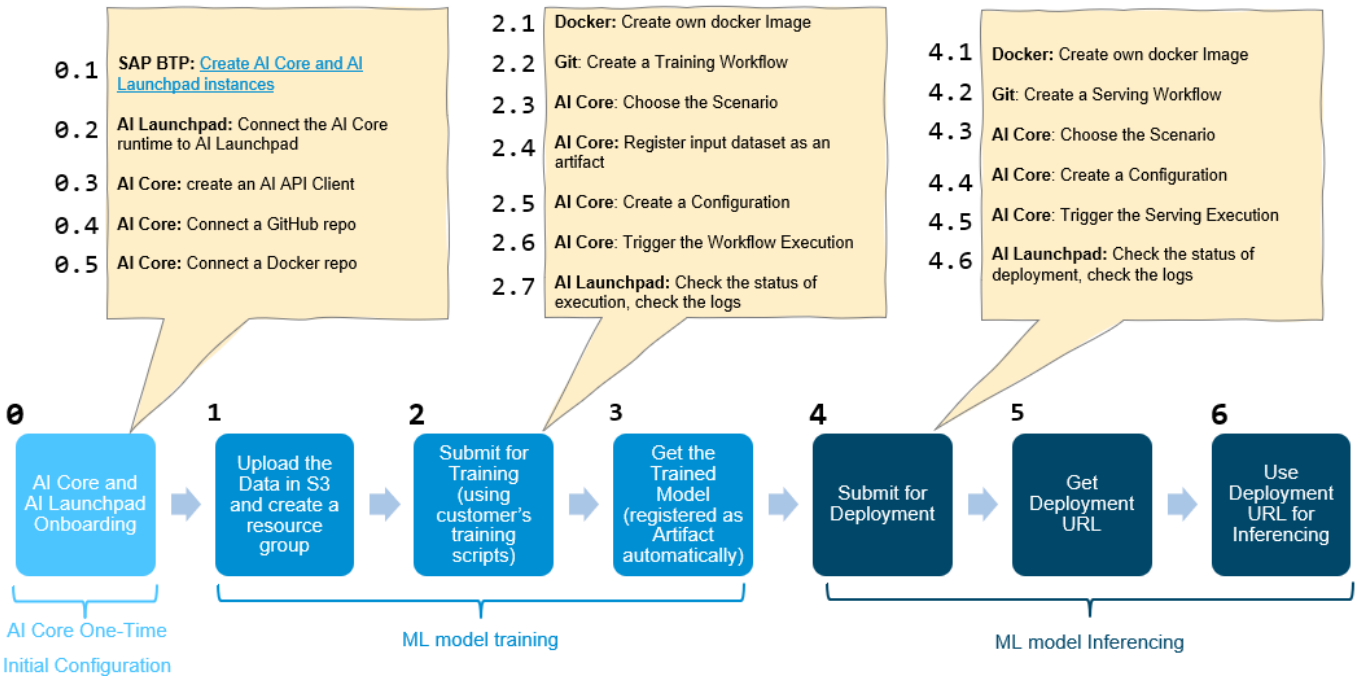
OBJECTIVE

The objective of this exercise is to show that it is possible to start a training or a deployment also from SAP AI Launchpad without passing through a Jupyter notebook.

For this exercise we assume the one-time configurations are already done and that we need to retrain our model for some reason (for instance because we have changed the dataset, we have changed the architecture of the model, we are not satisfied with the previous training, etc…) and to deploy it again.

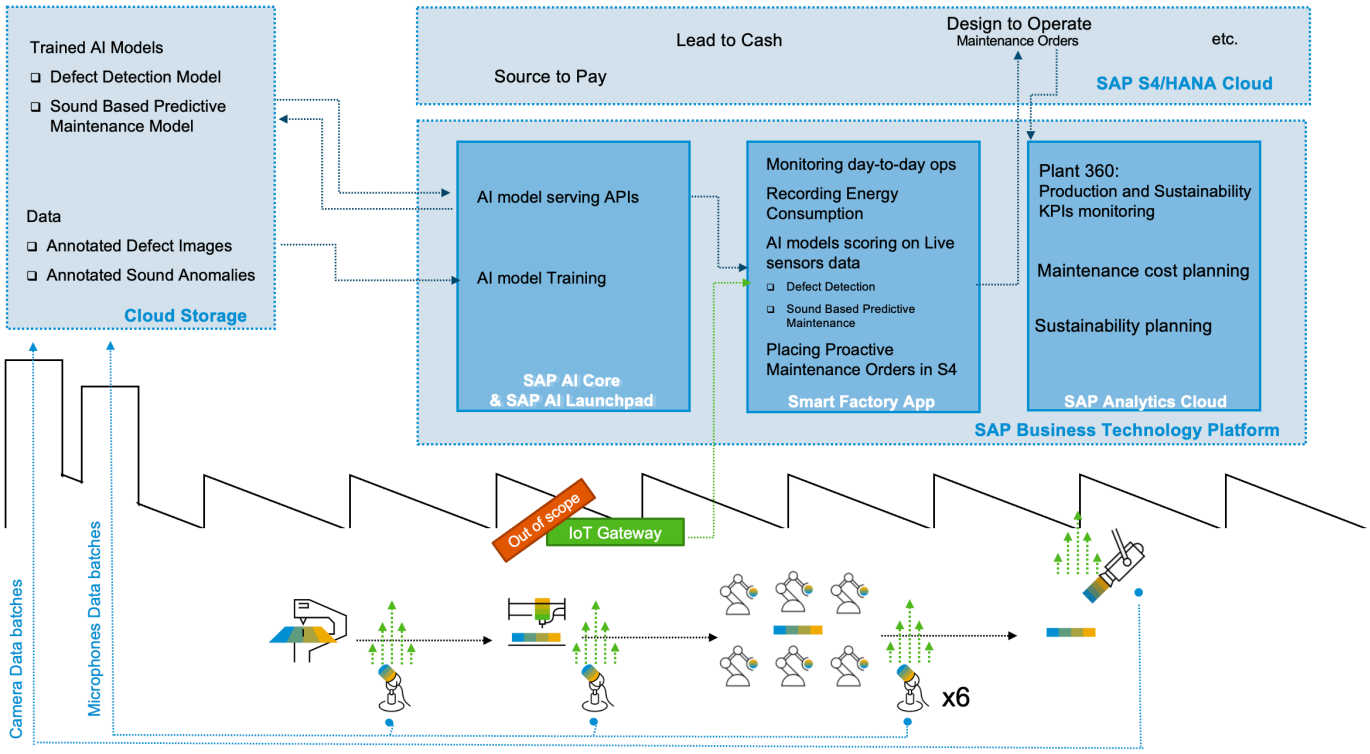
So we are going to reuse the Docker images that we have already built and loaded in the Docker repo, the application and the resource group we have already created.

From the point of view of the ML workflow reported below and with the assumptions we have made, our exercise starts at step 2.



SCENARIO

This exercise follows the scenario you were introduced to in the demo. Bagnoli & Co manufactures Light Guide Plates (LGPs) used in LED panels. An SAP Partner proposed to implement a Smart Factory Application on top of SAP BTP to improve manufacturing operations and raise awareness on sustainability KPIs. Machine condition monitoring based on sound classification was implemented to replace time-based preventive maintenance with proactive (and predictive) maintenance and to prevent downtimes. Sounds are acquired along the production line with several microphones. A Deep Learning model is then built, trained and deployed to classify the sounds, identify anomalies and therefore achieve proactive maintenance of the production line (see image below).



ENVIRONMENT ACCESS

Steps	Explanation
SAP provisioned environments	SAP have provisioned the following: • SAP AI Core • SAP AI Launchpad • AWS S3 data storage bucket
Prerequisites	You have completed exercise 2 available in the bootcamp repository: https://github.com/SAP-samples/btp-ai-sustainability-bootcamp/tree/main/src/ai-models/predictive-maintenance/exercises
	Your trainer will allocate each student a unique user number. This number is used when you save or create objects in the application. In these training instructions, the user number is referred to as uXXXX. Please, refer to your ID in order to reuse the objects you have already created in the exercise 2 with Jupyter notebooks: application, resource group, scenarios, executables, …

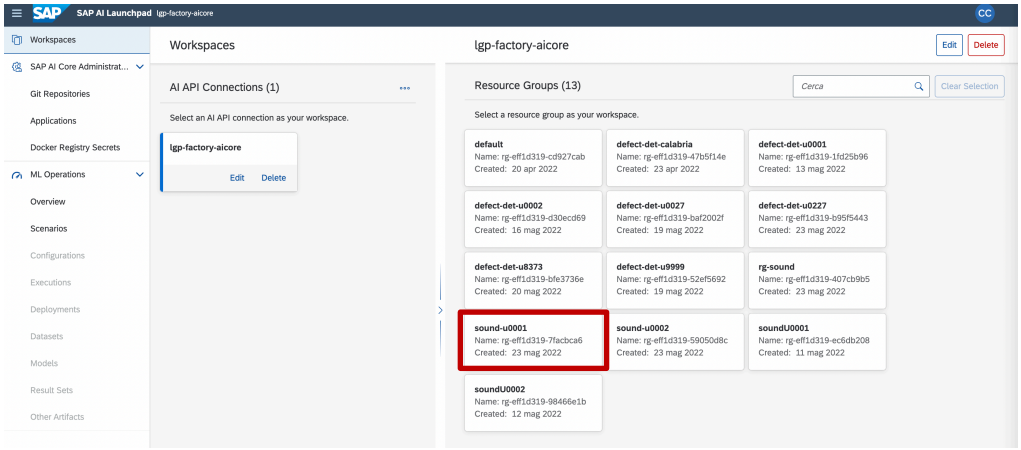
STEP 2 – SUBMIT THE TRAINING

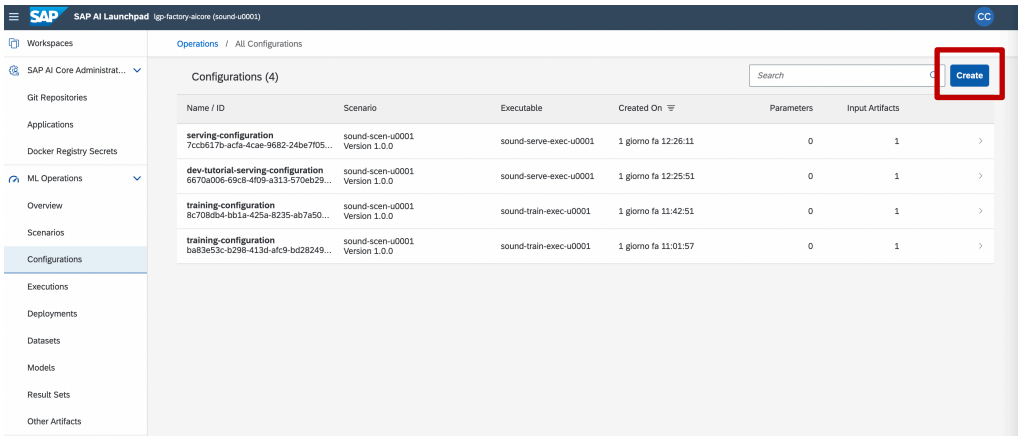
Please, remember that it is assumed that you have performed step 0 and 1 already during Exercise 2. You will start directly with step 2 and will use SAP AI Launchpad exclusively.

It is assumed that substeps 2.1, 2.2 and 2.4 have already been performed, while substep 2.3 is basically included in substep 2.5.

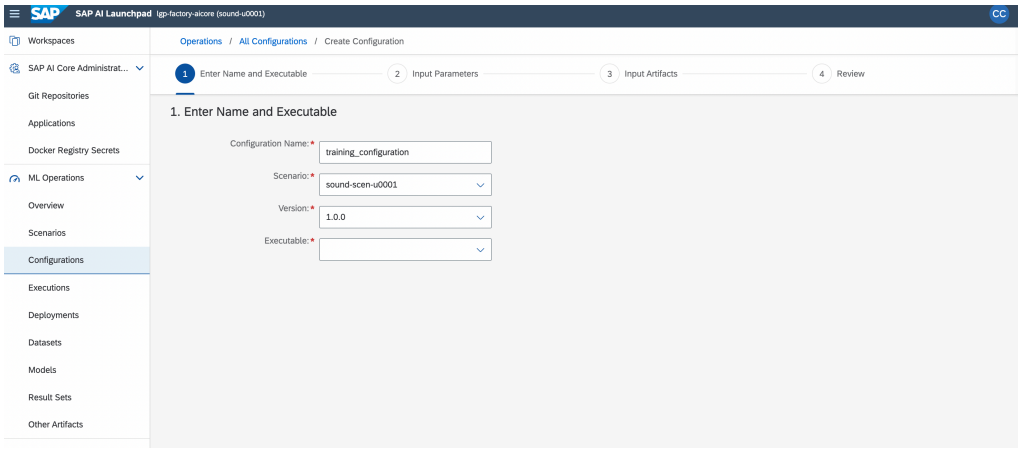
The main reasons for retraining a model are:

- The dataset has changed, e.g. because new data have been added;
- The model architecture has changed or other parts of the code have changed;
- The training templates have changed:
    - Different resource plan;
    - Different docker secret;
    - Different pipeline in the number of steps and container used;

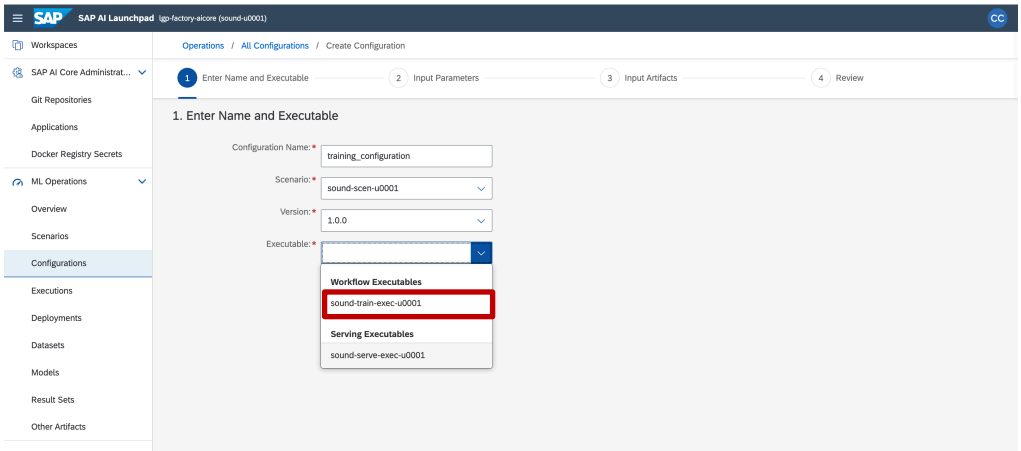
Steps	Explanation + Screenshots
Access SAP AI Launchpad	Please, refer to the instructions reported in the Teams channel: SAP AI Launchpad in System Access 🚀
Choose your resource group from Workspaces in SAP AI Launchpad	
2.5 Create a training configuration	Once you have selected your resource group, move to the left and from the ML Operations menu choose Configurations. There you will find all the previous configurations you have created during exercise 2 with the second Jupyter notebook. You will create a new configuration. Click on the Create button at the top right of the page.



In the menu, you assign a name to the new configuration and choose a scenario from the ones available (please refer to the one identified by your personal ID). Also choose the version (in this exercise there should be just one available - the one you specified in the templates).

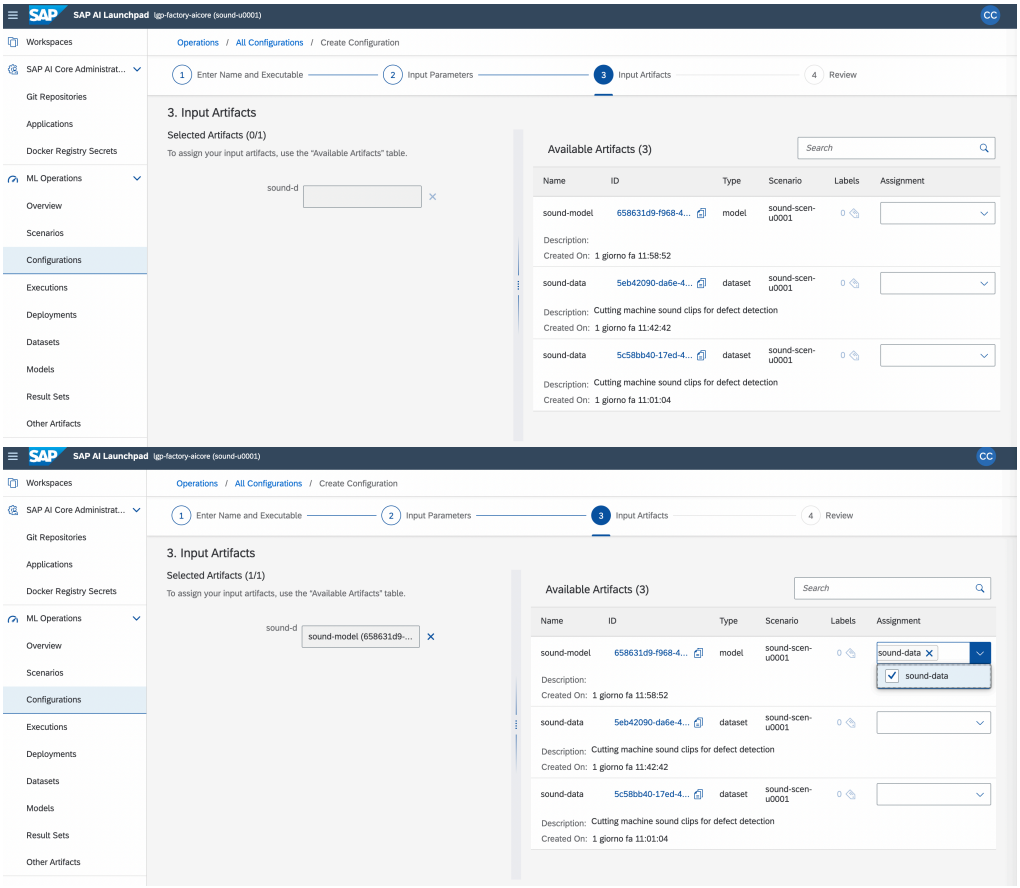


Then you choose the executable. Since you are going to submit a training, choose the one belonging to the training pipeline.

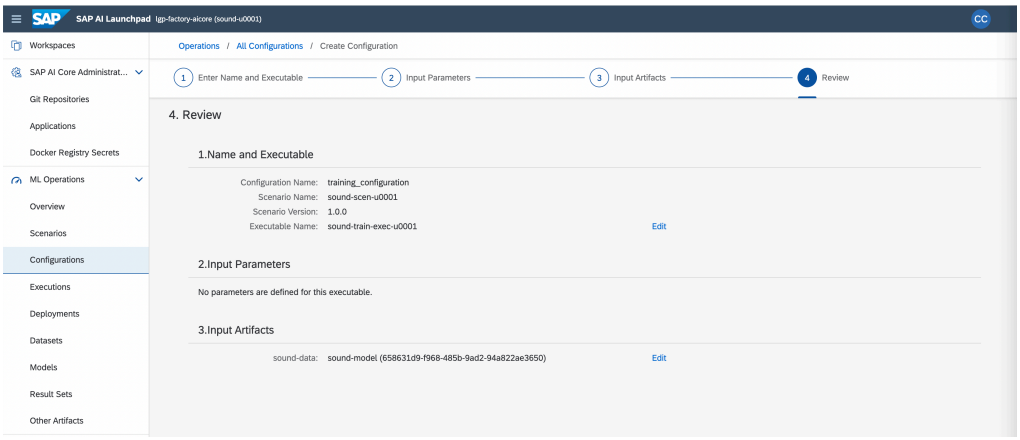


The next step of the creation would be to choose the parameters for your executable. In this exercise, you do not need any parameter so you can click

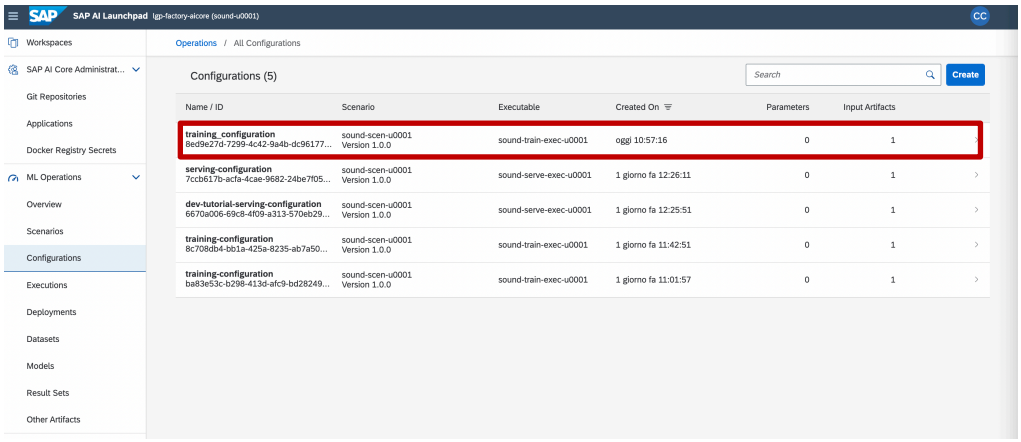
Next and move to the selection of the input artifact from the ones available (from those previously registered for your resource group).



Click Review to review what you have done so far. If you are satisfied you can click create.

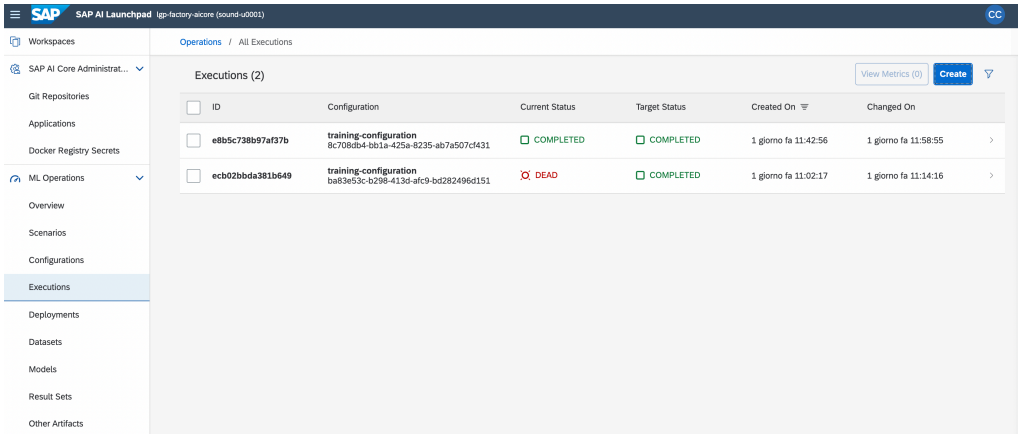


Now the new configuration will appear under Configurations.

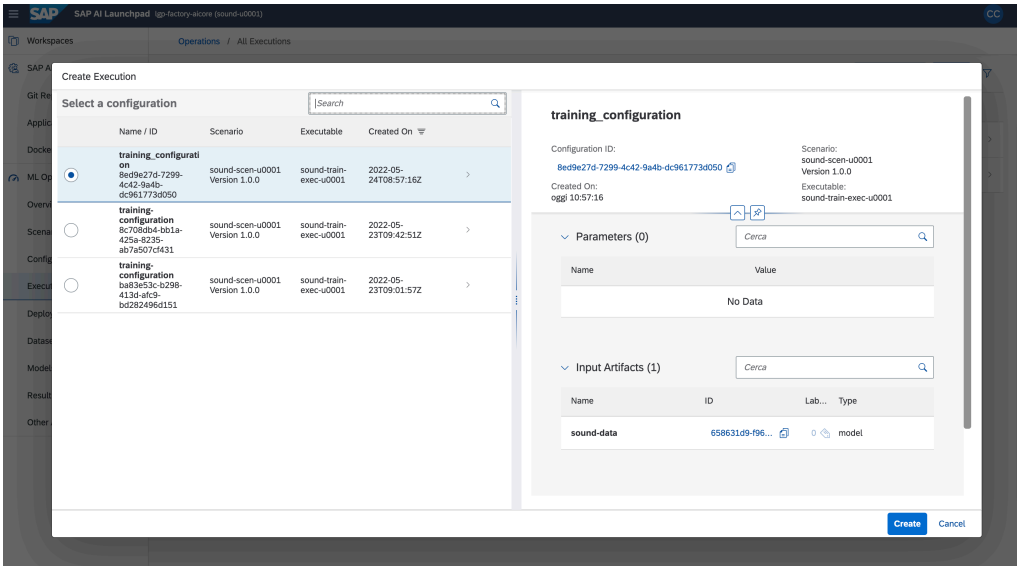


2.6 Trigger the training execution

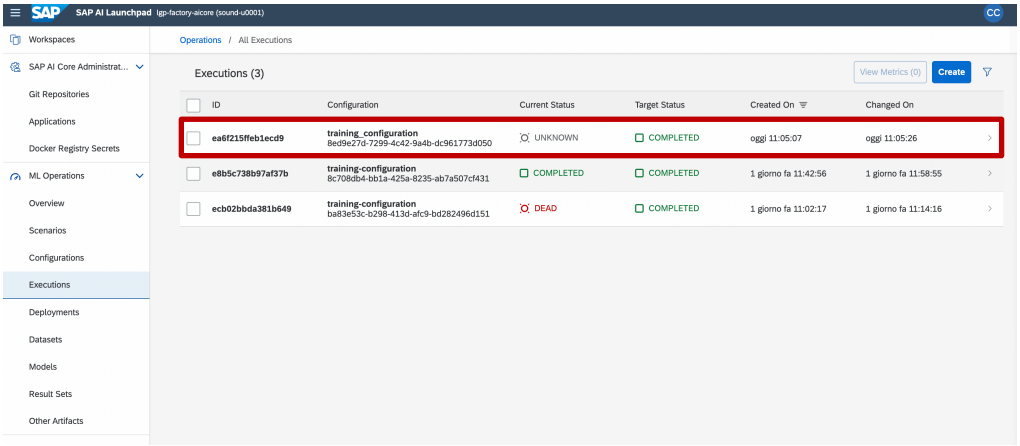
Once the configuration is created, you can proceed and submit a training execution. You need to move to Executions where you will find the list of all the previous executions.



You need to click on the Create button, at the top right of the page, and a menu will appear where you can select which configuration you want to start. For this exercise, select the one you have just created (you could also choose an old configuration if needed).



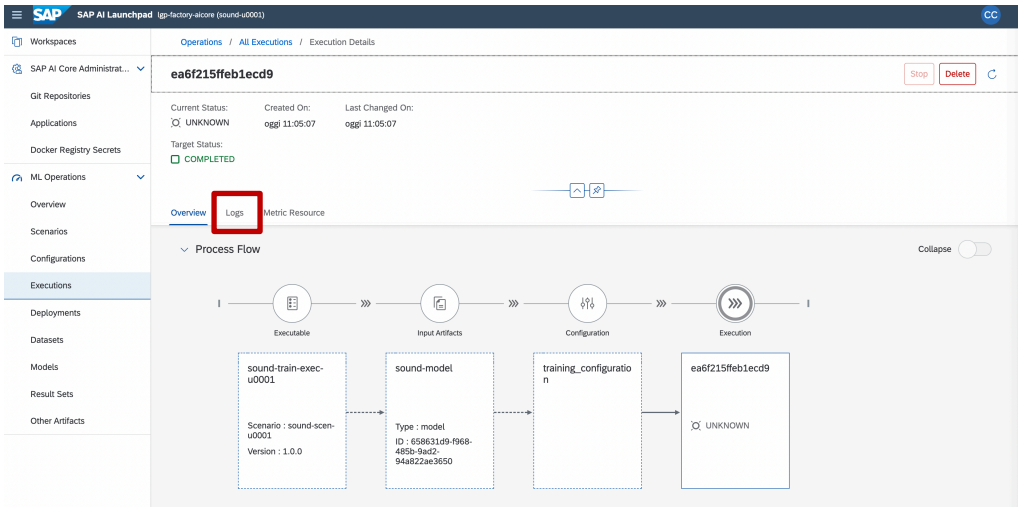
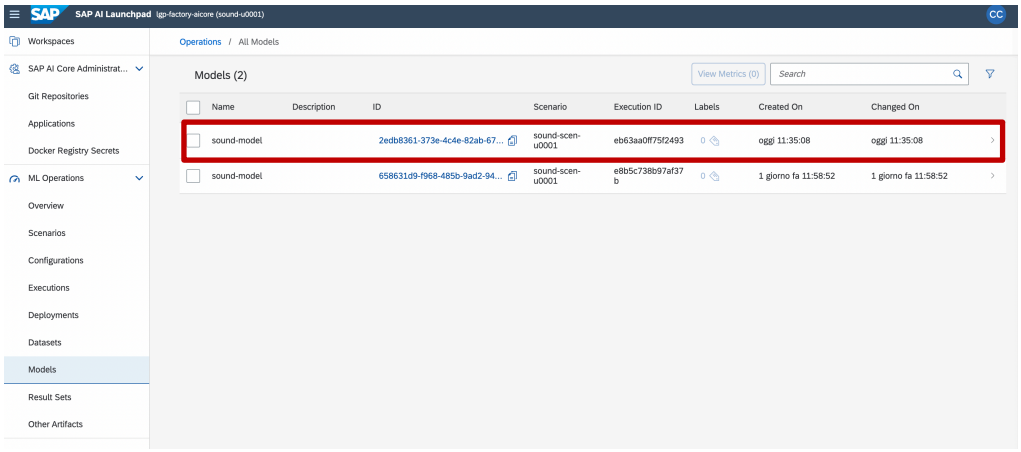
At this point, you need to click Create to start the training and it will appear under Executions.

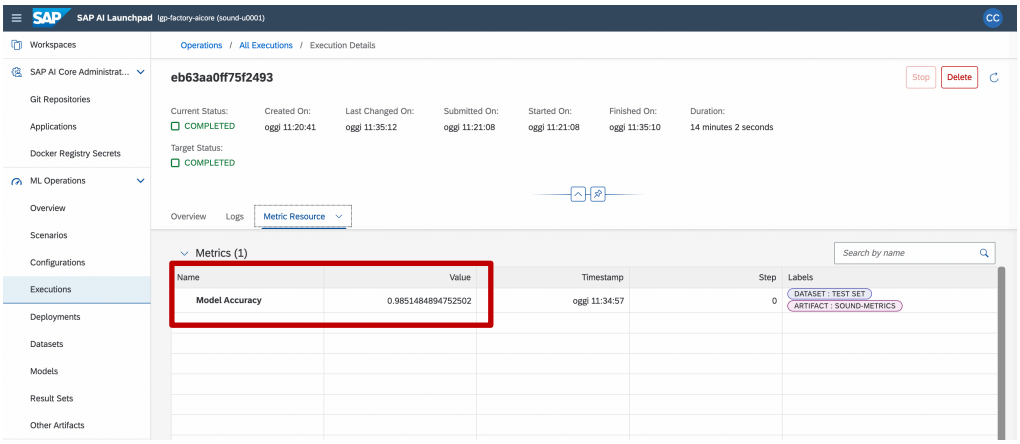


2.7 Check the status of the execution

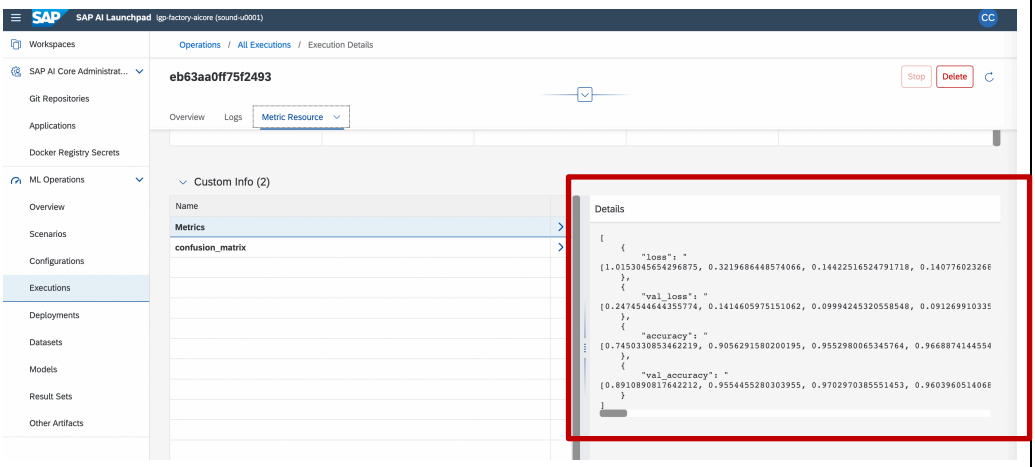
This step can be performed in SAP AI Launchpad by clicking on the running execution. A page similar to the one shown below will appear.

From there you can check the logs for example.

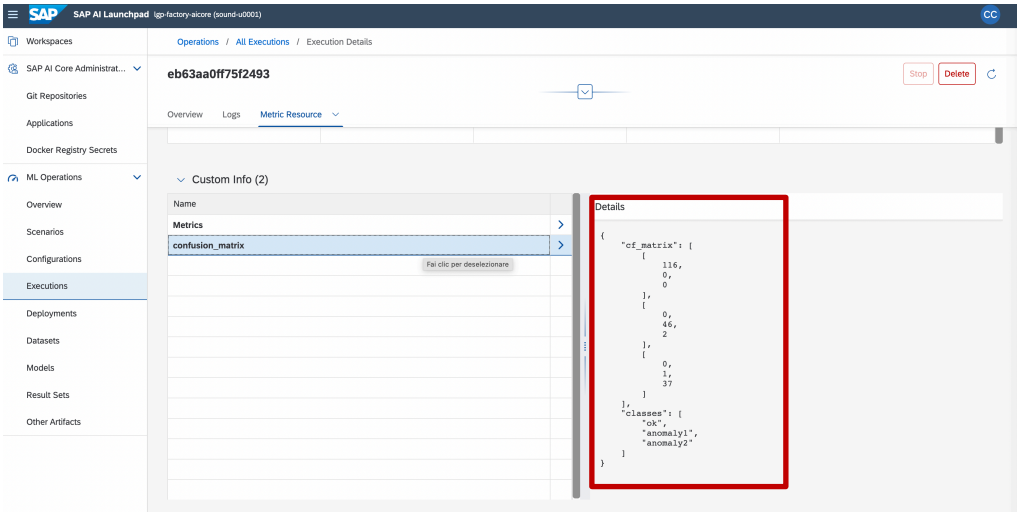
	
3 Get the trained model	When the training is completed, a new model is produced and saved in the AWS S3 bucket. It is automatically registered as an artifact in SAP AI Core, and it will appear under Models. 
Checking the loss function trend and the metrics	The loss function and the chosen metrics values are defined in the training python code and there they are required to be stored in SAP AI Core. It is possible to consult them also through SAP AI Launchpad. It is enough to go under Executions, select the desired (completed) execution and then go in the metrics tab. Here it is possible to check the value of the metrics like the overall accuracy of the model obtained by applying the trained model to an independent test sample (created internally in the training code from the original input dataset).



Additional info has been stored as custom info that can be also checked: we have saved in SAP AI Core all the loss function and accuracy values for each iteration/epoch for both the training and validation samples.



Moreover we have saved the final confusion matrix obtained by applying the trained model to an independent test sample.



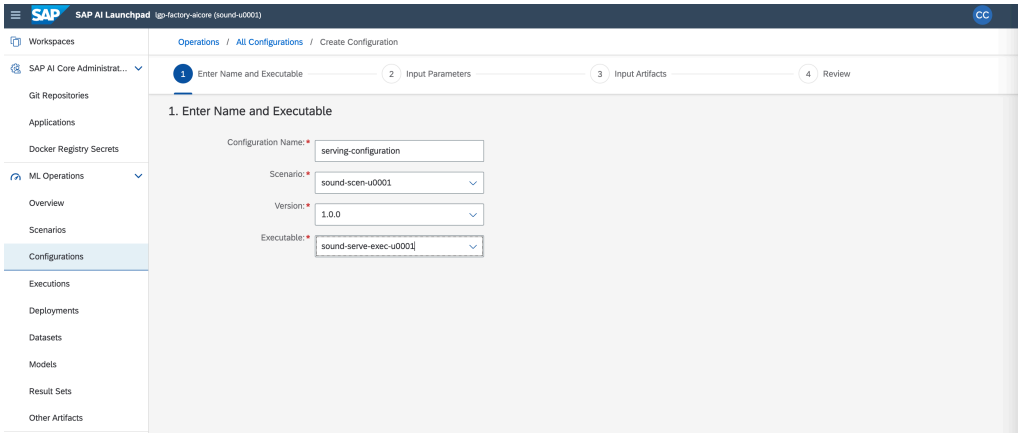
STEP 4 – SUBMIT THE DEPLOYMENT

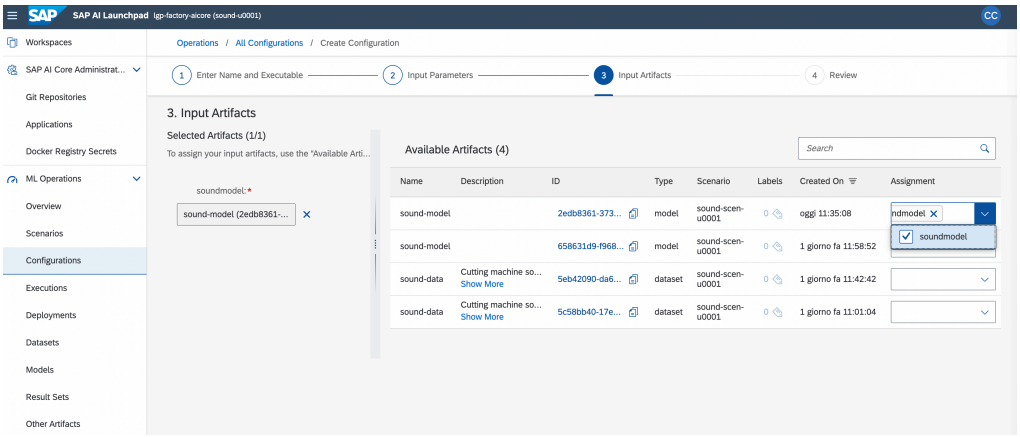
In this section it is assumed you have already performed a training and generated a model, or that a model was loaded into the AWS S3 bucket and registered as an artifact in SAP AI Core.

It is assumed substeps 4.1 and 4.2 have been already performed, while substep 4.3 is basically included in substep 4.4.

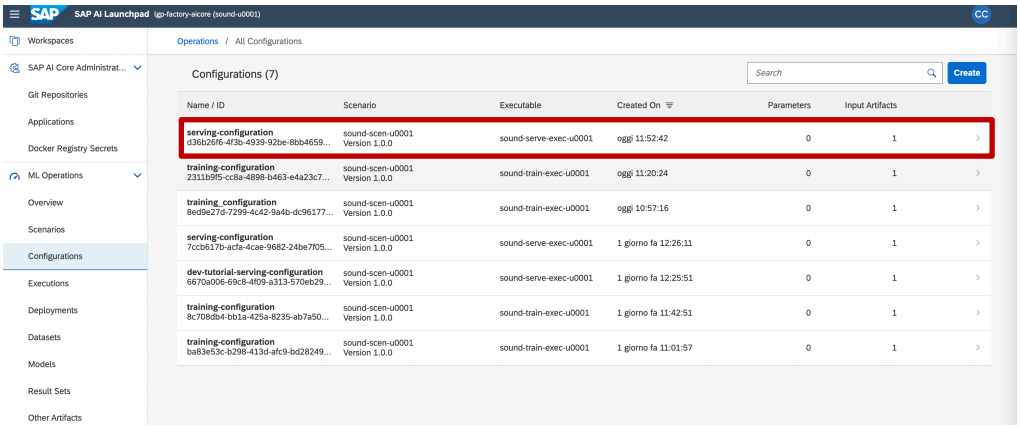
The main reasons why you might need to start a new deployment are:

- The model has changed so it has been retrained;
- The serving applications has changed;
- The resource plan has changed in the serving template.

4.4 Create a serving configuration	To create a serving configuration, you need to repeat the same operations you did in substep 2.5: • Go to Configurations; • Click the Create button; • In the menu choose a name for the configuration; • Choose a scenario from the ones available; • Choose an executable (this time choose the one corresponding to the serving part); • Choose the parameters if they are defined for your executable. Here, in this exercise, no parameters have to be passed.  Then you choose the model to use as an input artifact for the serving part. Use the one you have just trained.

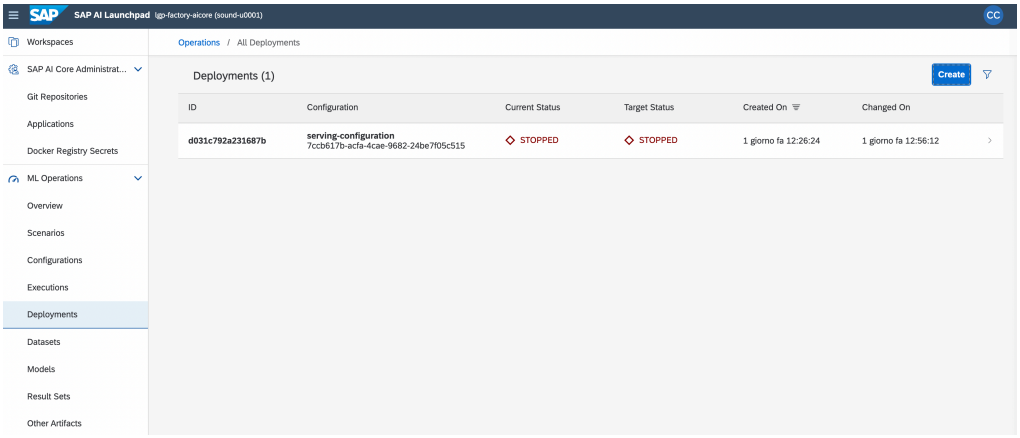


Once you are happy with the configuration, you can click Create and then the new serving configuration will appear under Configurations.

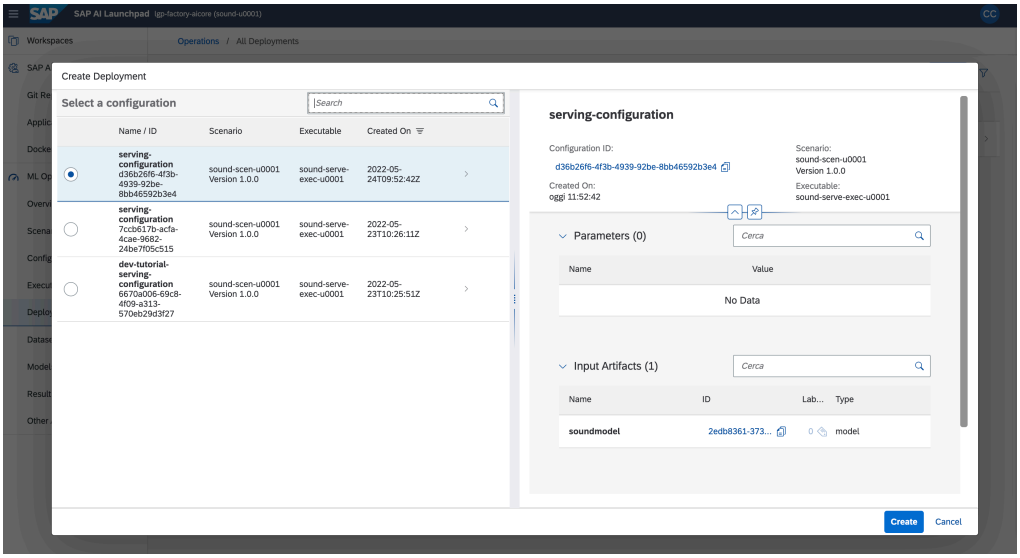


4.5 Trigger the serving execution

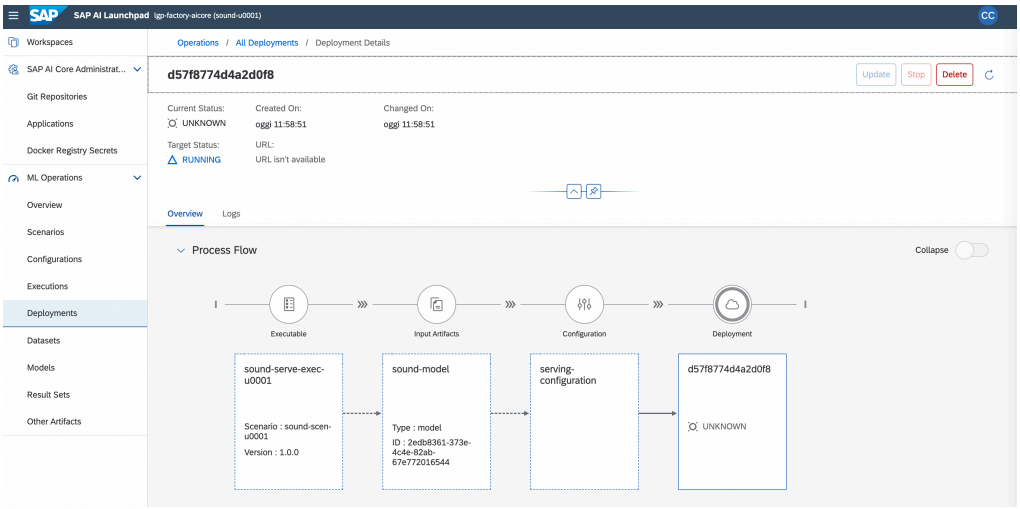
In order to start the deployment, you need to go to Deployments. There you will find the list of past deployments, or running deployments.



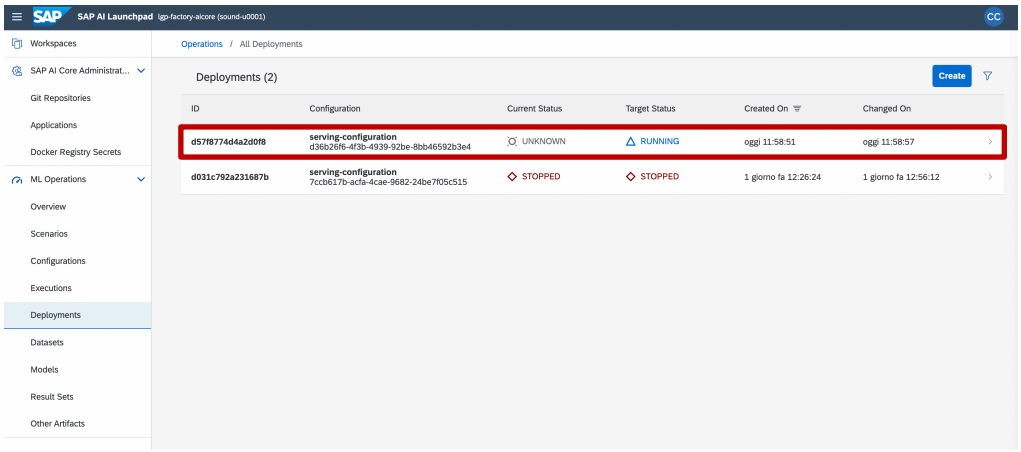
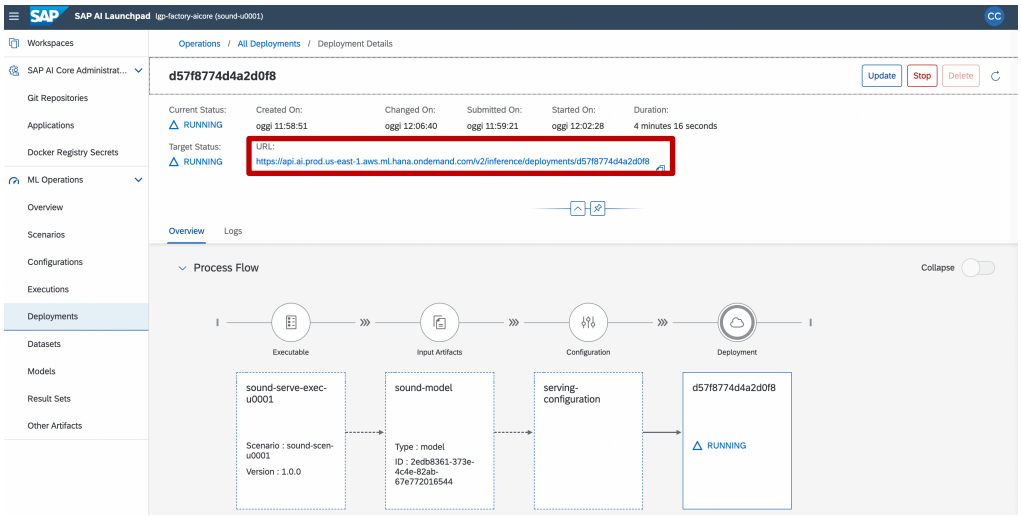
Click on the Create button at the top right of the screen to create your deployment. From the left side screen, choose the serving configuration you want to use. In this exercise, choose the one you have just created (you can also reuse an old configuration if required).

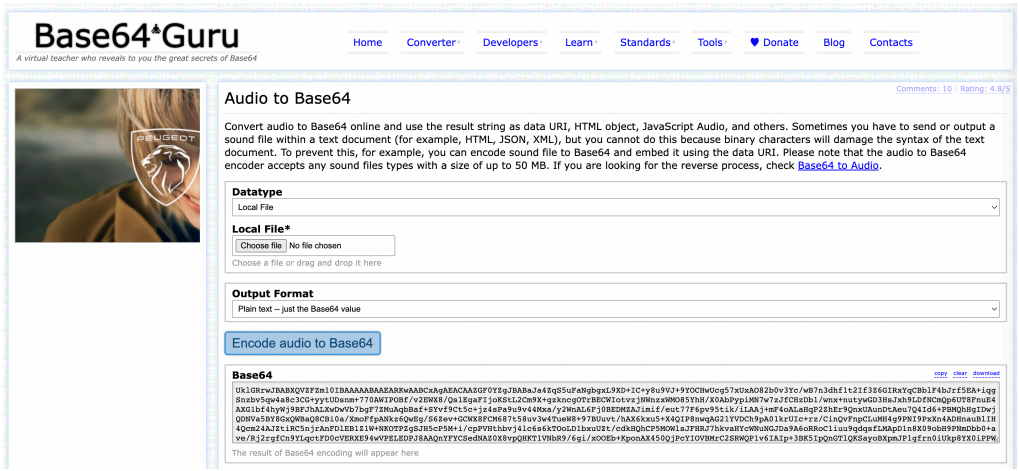


Click Create and then the deployment will start on all the nodes you have specified in the serving templates. This operation will take few minutes to complete.



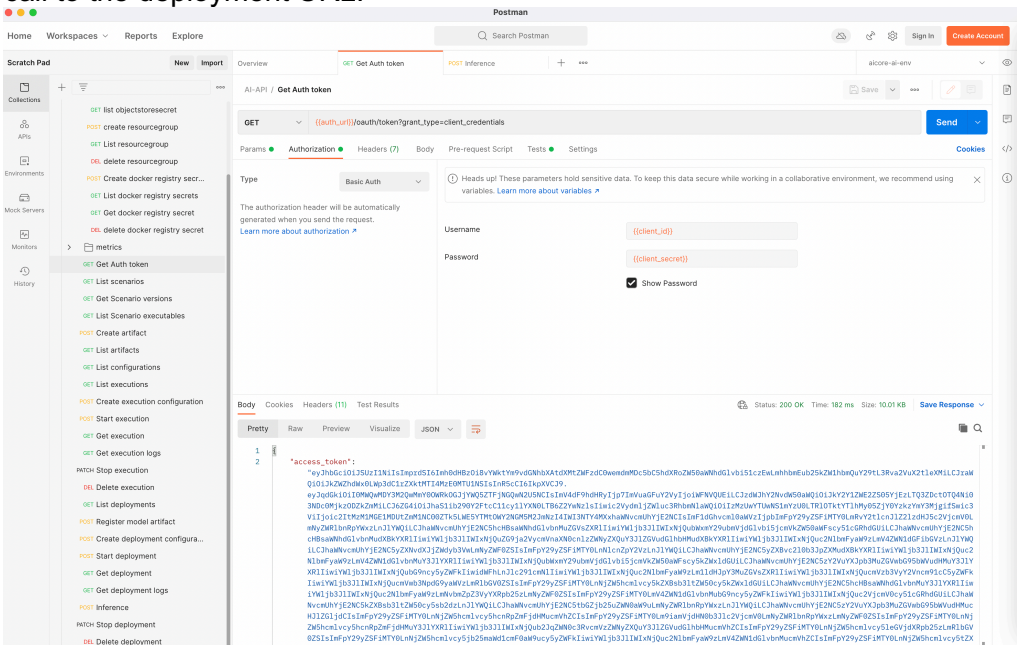
The deployment will appear immediately under Deployments.

	
4.6 Check the status of the deployment	From Deployments, it is possible to check on the status of the deployment by clicking on the item in the list. You can check all the logs and once the deployment is completed, you can check the deployment URL. This is the URL that can be used to send inference request to the deployed model. The serving requests will be managed by the web application you have developed. 
6 Use the deployment URL	The model is deployed and the serving applications is running. You can choose just one anomalous sound from the dataset, and convert it to a binary string. For example, use this online service: https://base64.guru/converter/encode/audio

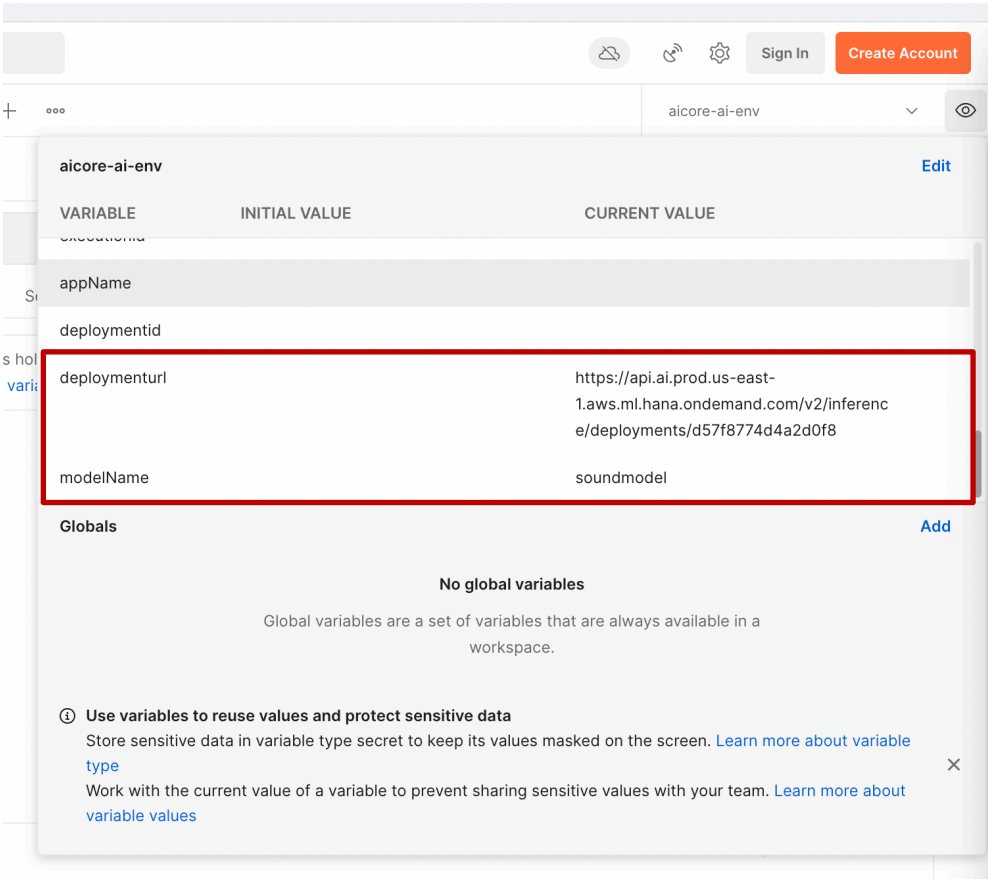


The you can use Postman. Note that Postman should be configured by following the instructions here: https://developers.sap.com/tutorials/ai-core-aiapi-postman-setup.html

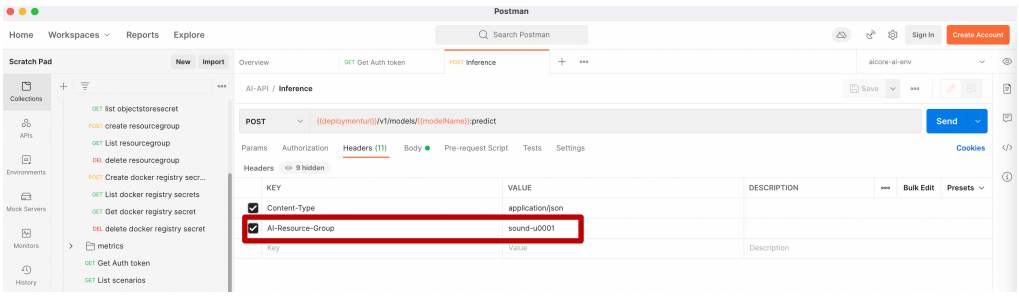
Please, remember to get an authorization token before trying to send an API call to the deployment URL.



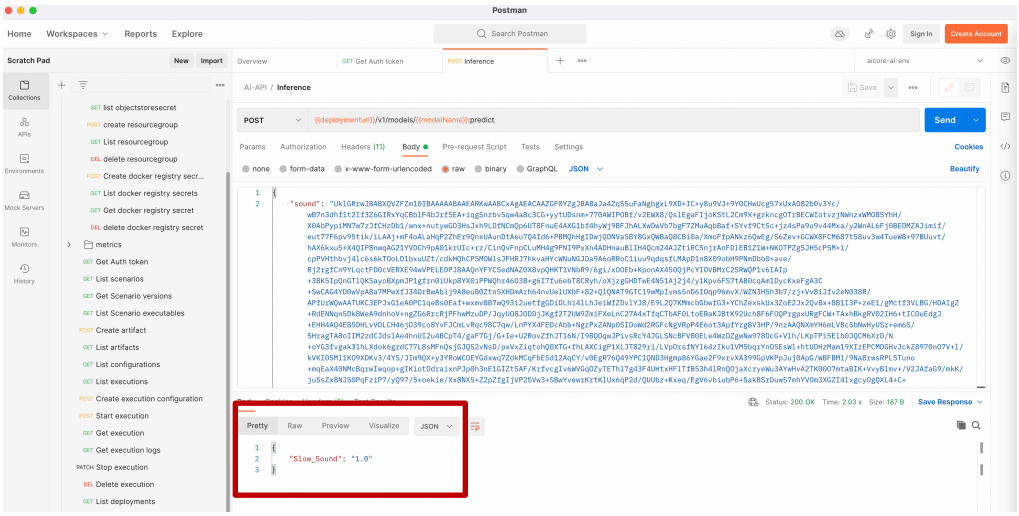
Please also remember to enter the deployment URL in the SAP AI Core environment. The model's name variable must be updated with the name you have given to your model artifact (in our case it is "soundmodel").



Please remember to set the resource group variable in the inference api with the name of your resource group.



Once you have completed this, you are ready to send an api call to the deployment URL. The response will indicate which class the sound belongs to.

	
Stop the deployment	Please remember to stop the deployment in order to reduce costs. This can be done easily from SAP AI Launchpad. In Deployments, choose the deployment you want to stop and click on it. Click on the Stop button at the top right of the new page that opens, and then the execution will be stopped.