HPE Private Cloud AI Live Demo

HPE Geneva Customer Innovation Center

Introduction & Demonstration platform

HPE Private Cloud AI is a turnkey AI solution built on NVIDIA AI computing. It includes the hardware, software, and services required to deploy AI applications quickly and efficiently while keeping private data safe and accessible.

Powered by HPE AI Essentials, HPE Private Cloud AI provides a single interface for managing one or more solutions across an organization regardless of system location

The purpose of this demonstration is to propose a Live Tour of the HPE Private Cloud AI solution, management from HPE GreenLake and the main features enables through HPE AI Essentials.

The demonstration platform is a Small HPE Private Cloud AI, its architecture is detailed on the right side.

HPE GreenLake Platform

NVIDIA Networking
Nvidia SN4600M Switches (100GbE Networking)

3 x Control Nodes (PCAI and K8 Control Plane)
HPE DL325 servers

HPE GreenLake for File Storage
Powered by HPE Alletra MP

2 x AI-optimized Nodes
HPE DL380a servers

Shared PCAI environment and automatic cleanup

The HPE Private Cloud AI environment is a shared demonstration platform.

Automatic cleanup scripts are run on a daily basis to remove users and cleanup resources.

If there is a need to secure something you have created, please contact the Demo Portal team:

contacthpedemoportal@hpe.com

Demonstration environments

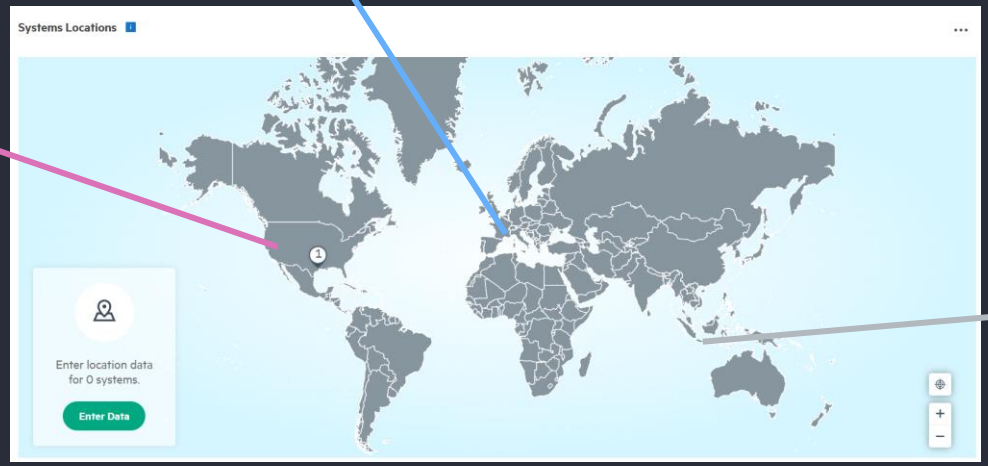
Status of the demonstration platforms September 2025:

A new Private Cloud AI instance has been deployed in Houston (US WEST), in version 1.8.0. It is the preferred platform to demonstrate as it is in the latest version.

The Geneva PCAI instance (EU Central) is deployed in version 1.6, and is still available as a backup platform in case there are issues with the Houston one.

HPE Private Cloud AI in the Houston CIC
Region US West

HPE Private Cloud AI in the Geneva CIC
Region EU Central



A 3rd HPE Private Cloud AI instance is available in the Singapore CIC.
It can be booked in a separate demo here

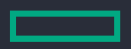
Demonstration menu

Status of the demonstration platforms September 2025:

	Geneva EU Central	Houston US West
1. Manage the HPE Private Cloud AI from the HPE GreenLake Platform	✓	✓
2. Access HPE AI Essentials to deploy and manage your AI workloads	✓	✓
3. Run an imported app : Open WebUI with Test Llama model Open WebUI with RAG (HPE QuickSpecs)	✓	✓
4. Deploy an app through Solution Accelerators	✓	⊘ Deprecated in v1.8
5. Visual Language Model (VLM) Imported solution / Traffic Analysis demo	✓	✓
6. HPE Machine Learning Inference Software MLIS	✓	✓
7. PCAI Monitoring in OpsRamp	✓	✓

1. HPE GreenLake Platform

Access for IT Administrators



1. Accessing the HPE GreenLake Platform

The demonstration is very easy to access directly from your own browser.
Once the demonstration booked, you are granted access to the main HPE GreenLake tenant for all demonstrations, the workspace « DreamCompany »

You will have access for 2 hours, if you need a longer access, don't hesitate to book the demo again.

Open your preferred browser to

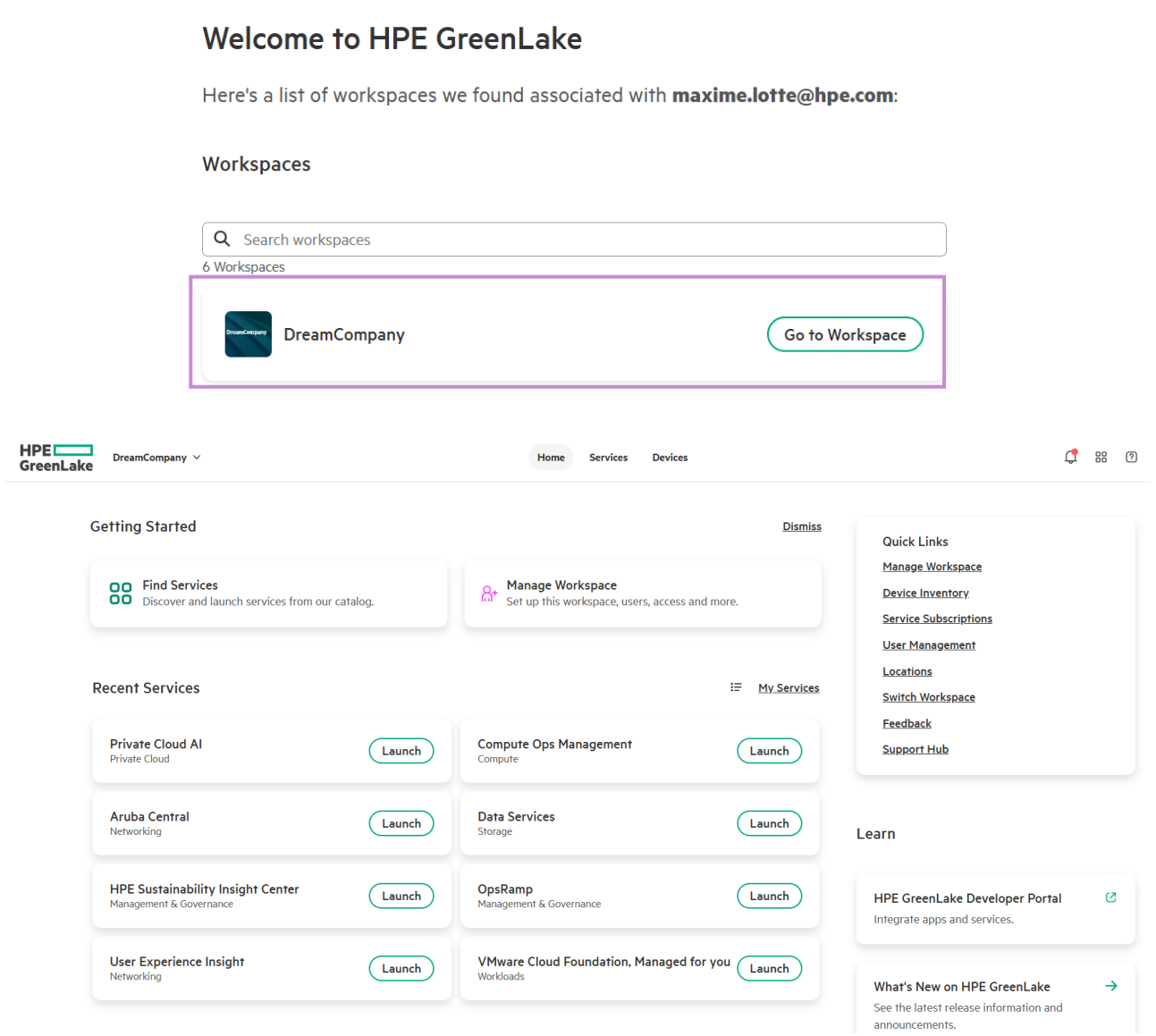
https://common.cloud.hpe.com/

Log in with your own email address.

Select the Workspace DreamCompany

You are connected to the HPE GreenLake Platfom.

Note: this demonstration also gives you access to all other services if you need to demonstrate other features.



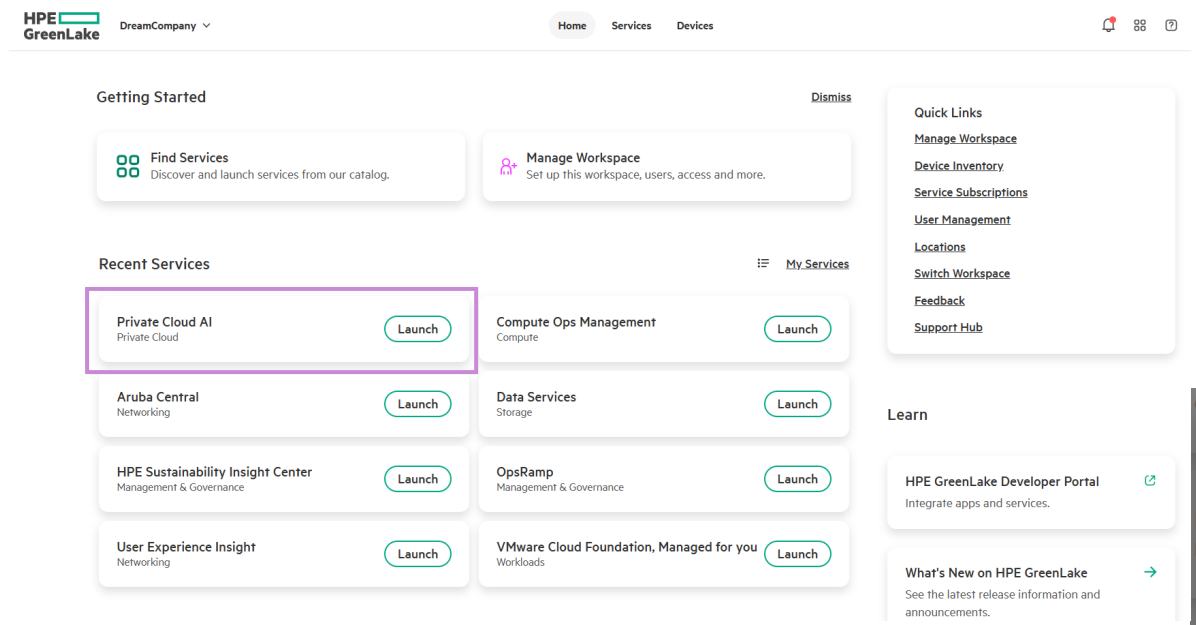
1. Accessing PCAI in the HPE GreenLake Platform

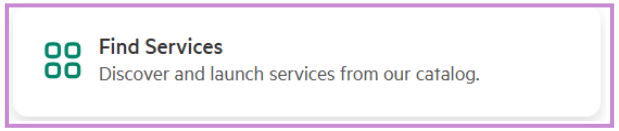
Once there, you can open the Private Cloud AI Service.

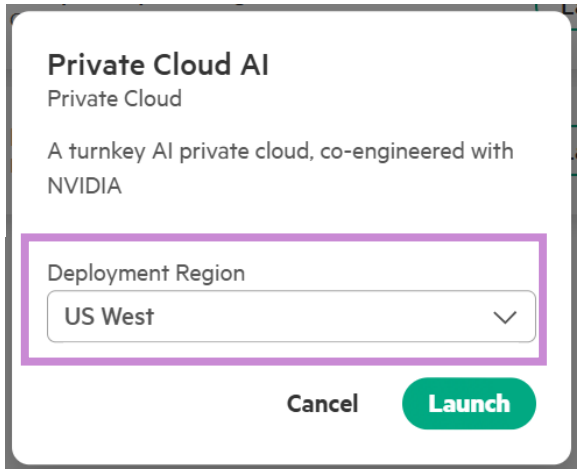
You can find it:

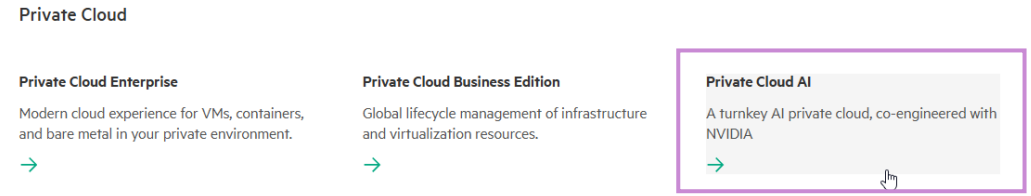
- In the Recent Services if it was recently used:
  Select region US West and click Launch. The Private Cloud AI instance in region EU is the instance always available, with all use-cases deployed and documented.

- If not, you can find it by clicking Find Services. Scroll down to the Private Cloud Category, select Privae Cloud AI, Click Launch, select region US West and click Launch again









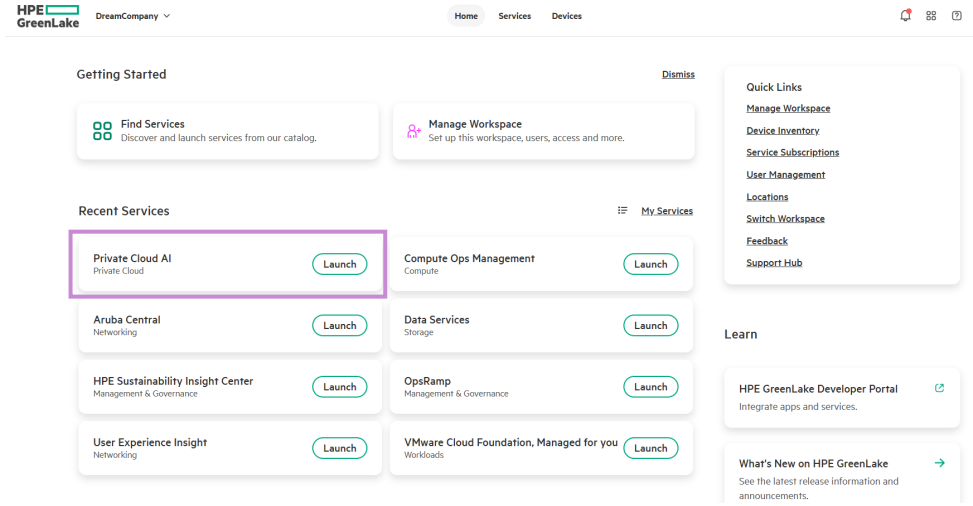
1. Accessing PCAI: Main demo environment / Backup demo environment

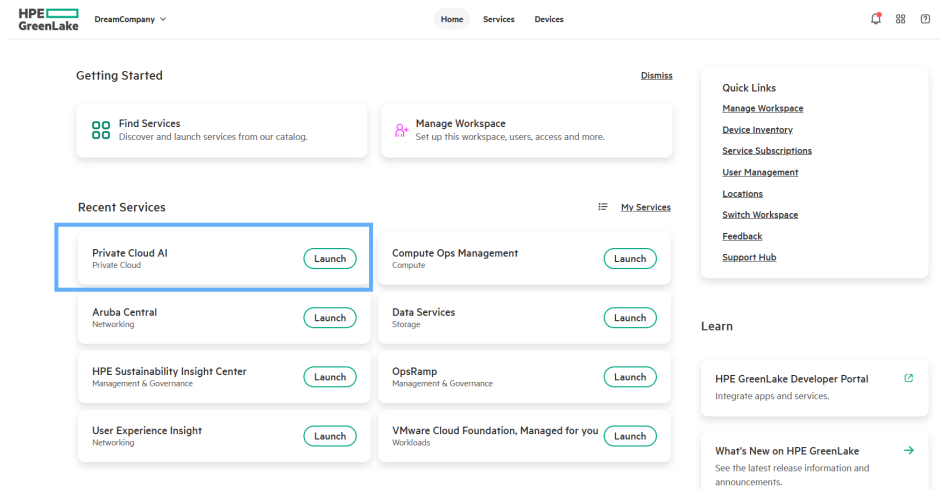
Important: In case of an issue with the HPE Private Cloud AI instance in West US region (in the HPE CIC in Houston):

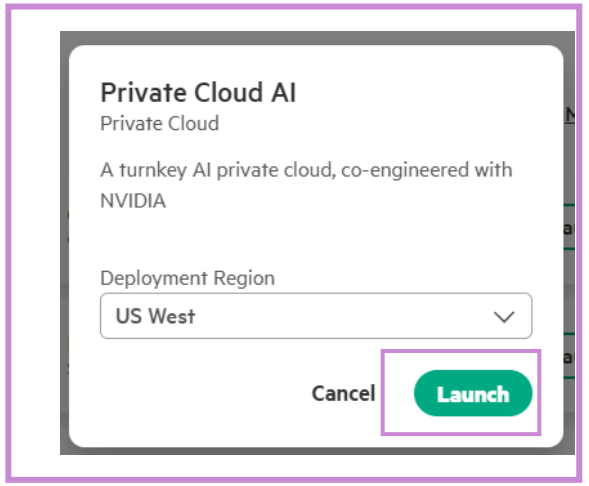
Come back to the HPE GreenLake homepage, and open Private Cloud AI in the EU region to demonstrate on the HPE Private Cloud AI in Geneva, as a backup demo platform.

The HPE Private Cloud AI in the US West region is deployed in the latest version, and has all use-cases implemented, it should be your preferred platform to demonstrate.

Use the EU region Private Cloud AI instance as a backup in case there is an issue with the US instance.

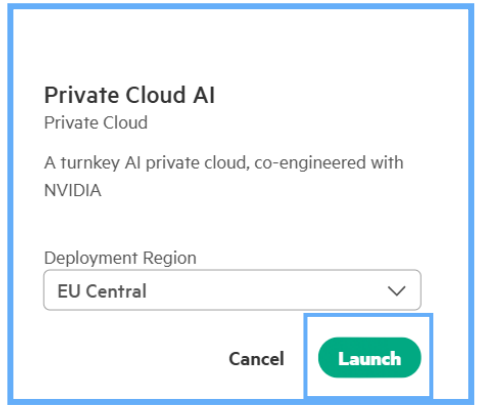






Main Demo Region

Backup Demo Region



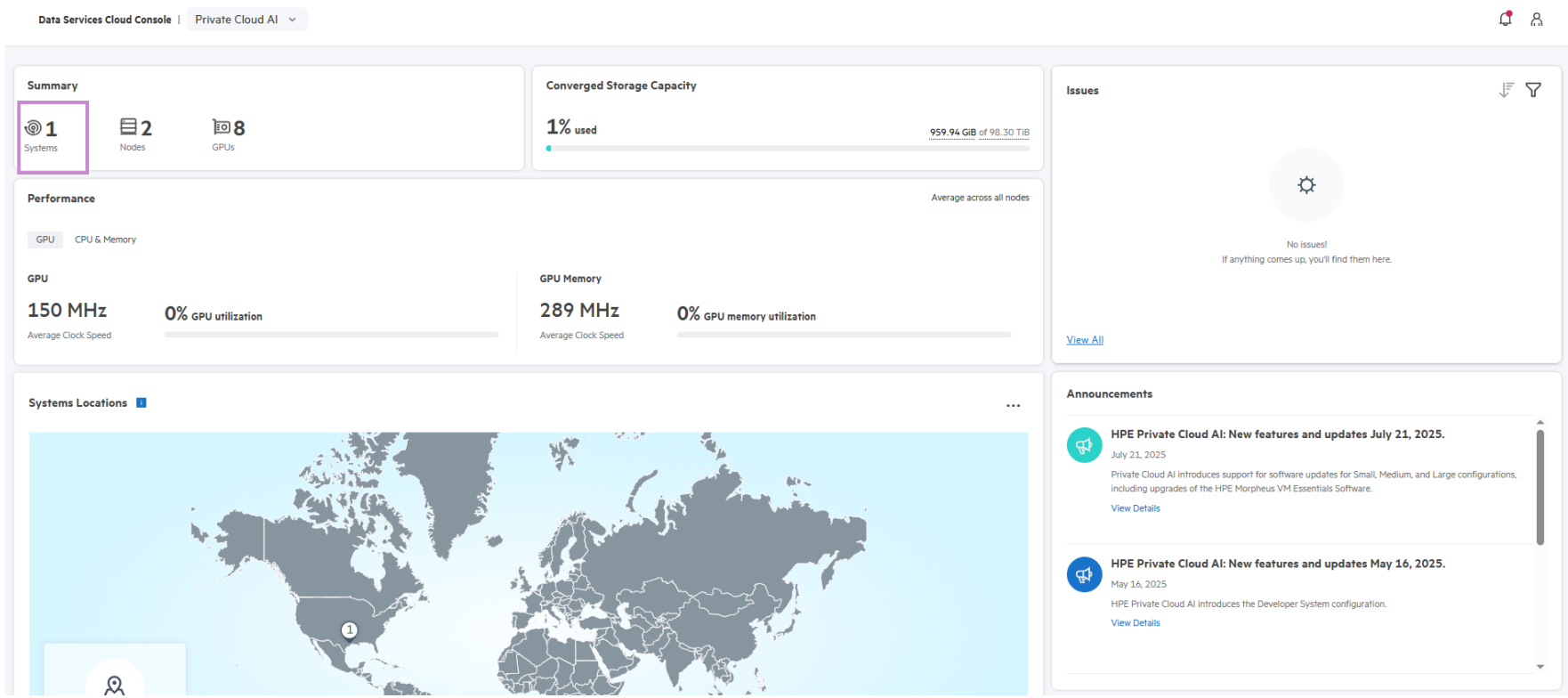
1. Managing PCAI in the HPE GreenLake Platform

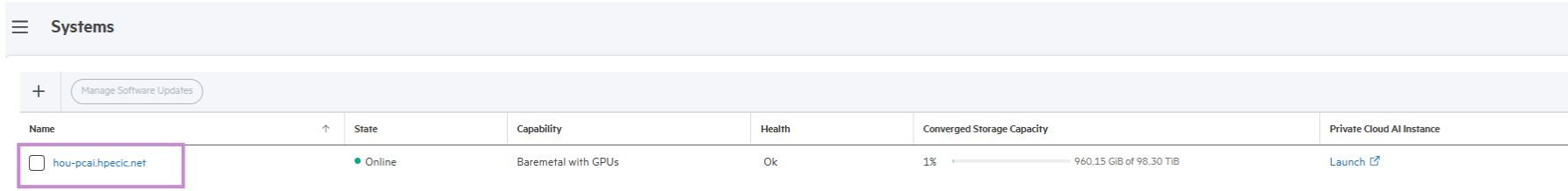
Here you have a global view of the different Private Cloud AI platforms you have deployed in the US West region.

For this demo there is 1 Private Cloud AI available, in Houston. You can see it on the map.

Click 1 System to show this PCAI.

Click the System hou-pcai.hpecic.net to open the system.

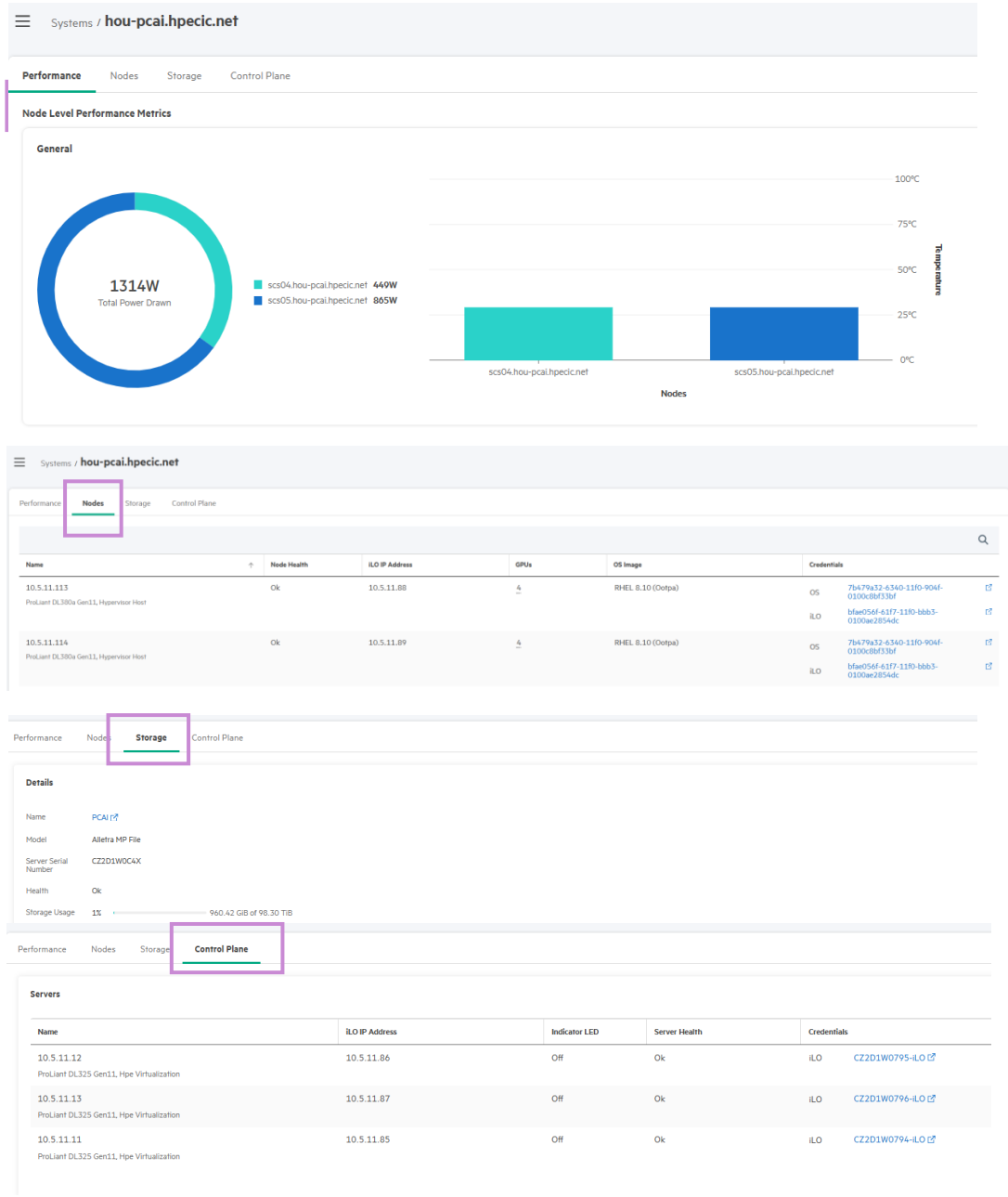




1. Managing PCAI in the HPE GreenLake Platform

Here you will find information about the different parts of the system:

- Performance of the nodes: GPU, CPU usage and power used
- Information and status of the Nodes
- Information and status of the Storage Alletra MP
- Information and status of the Control Plane (3 ProLiant Servers in the PCAI rack)

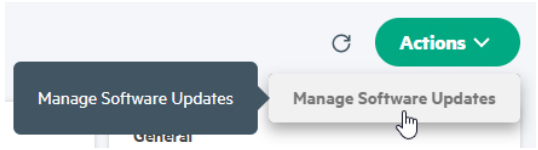


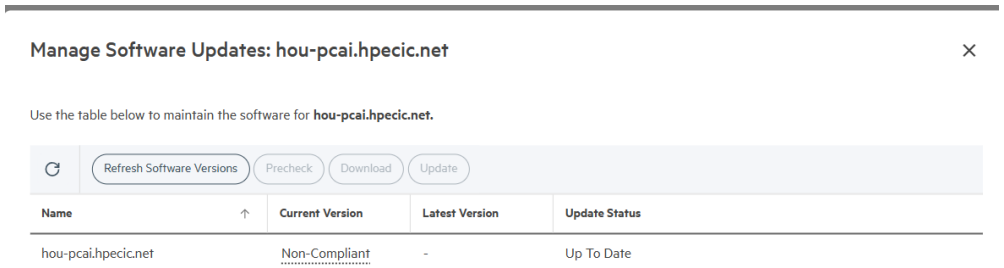
1. Managing PCAI in the HPE GreenLake Platform

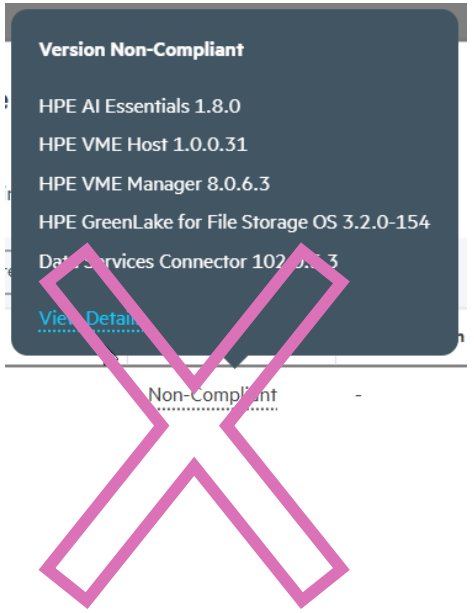
You also have an Actions button in the top right corner, allowing you to check for updates to the Software stack.

Here the plaform is up to date. Hovering the version will show more details. It mentions on this one *'Non-Compliant'* as it is a demonstration instance of PC AI.

Do not trigger updates on the PC AI environment as this is a shared demonstration platform.







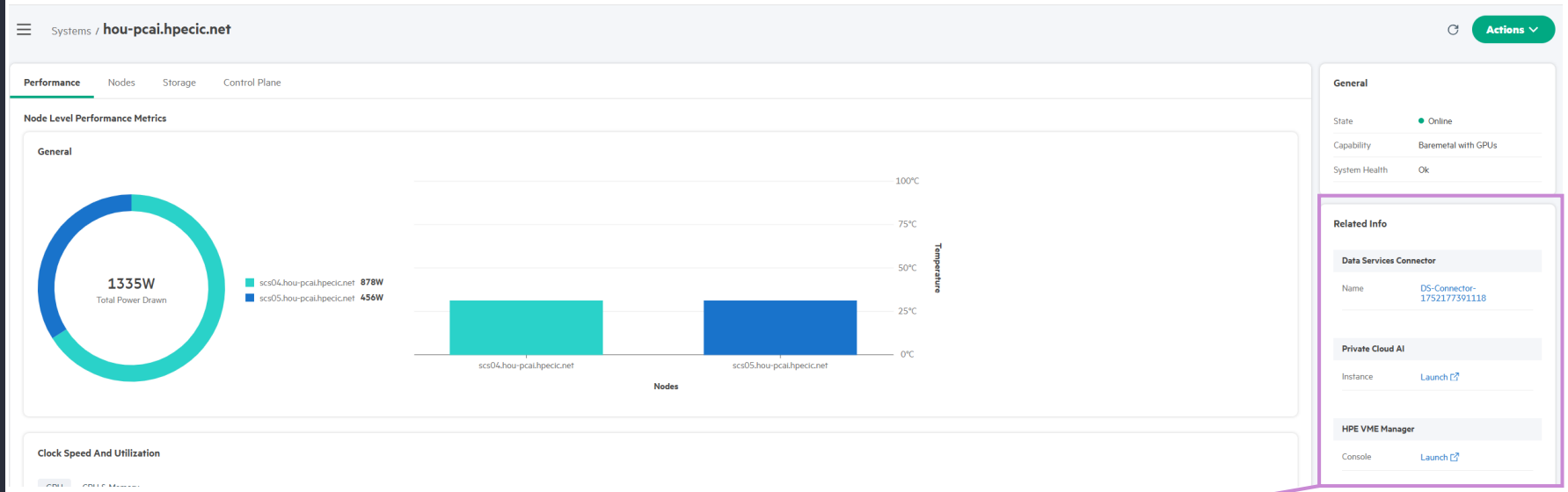
Do not trigger Updates !

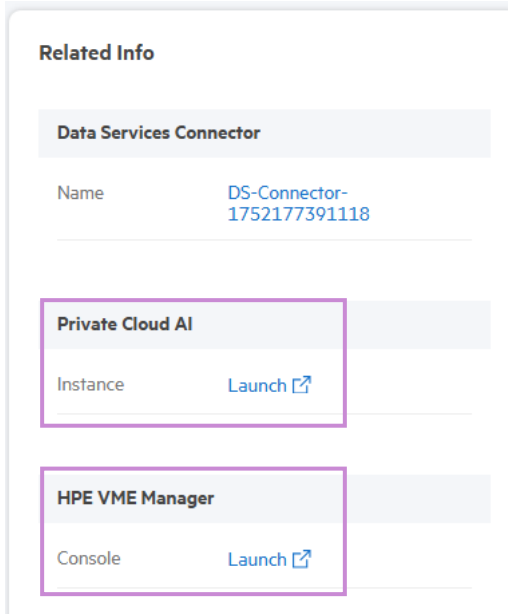
1. Accessing the PCAI endpoints from the HPE GreenLake Platform

Finally, the HPE GreenLake Platform shows you the main endpoints needed for the PC AI Platform. You can find them on the right side:

- Private Cloud AI: Will launch the HPE AI Essentials dashboard, to start deploying AI workloads on the platform. This is described in details in the next part of the demo guide. Click this link to continue the demonstration.

- HPE VME Manager: This will open the HPE VME console to manage the Control Plane. This is not covered in this version of the demonstration, as it is not central to the PC AI Solution. But it is interesting to see that HPE VM Essentials is running the Private Cloud AI Control Plane.





2. HPE AI Essentials

Access for IT Operators and Data Scientists

2. Accessing HPE AI Essentials

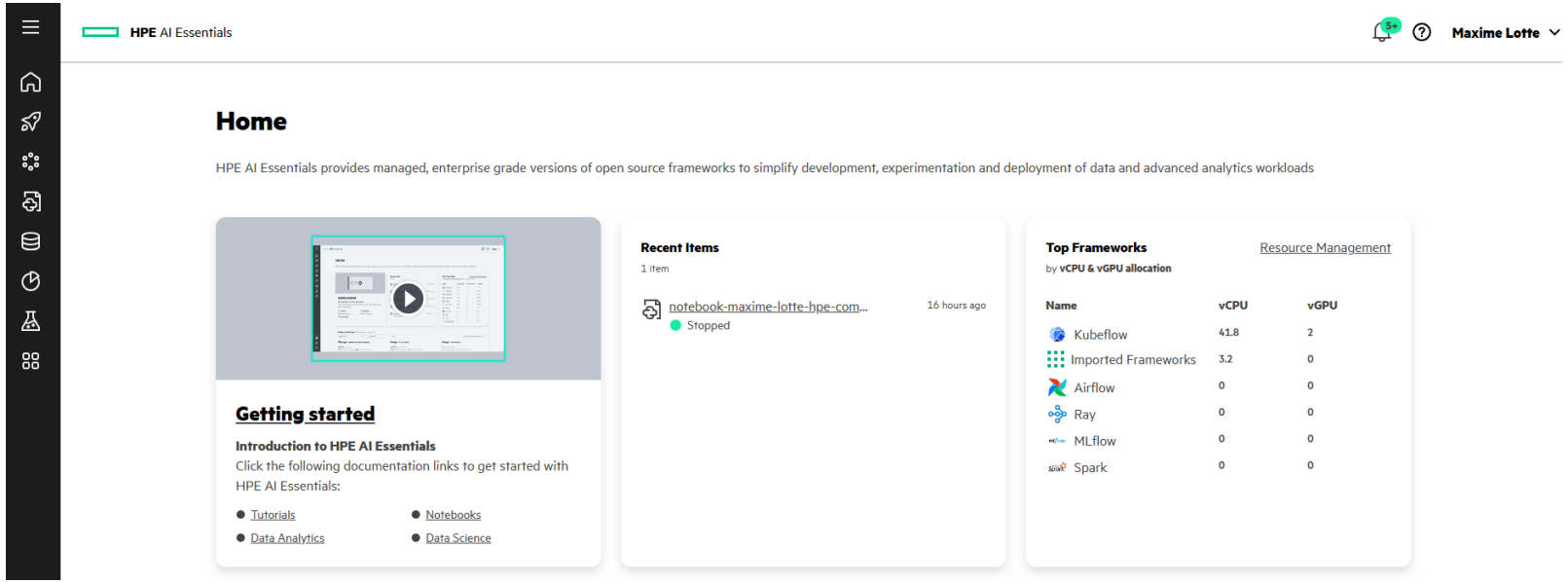
Click the Private Cloud AI link in HPE GreenLake. Another tab opens, and you are connected with SSO to HPE AI Essentials part of HPE Private Cloud AI.

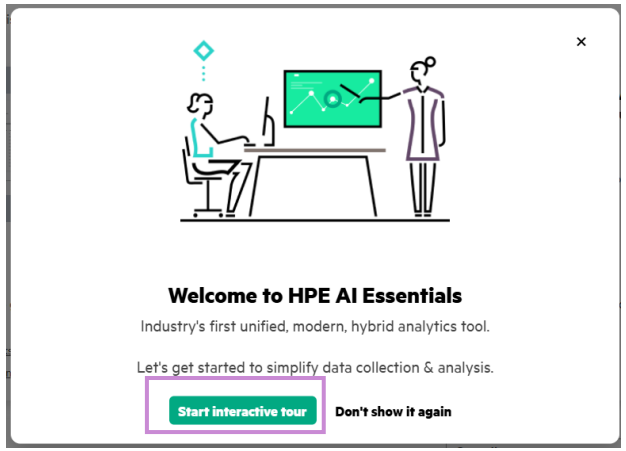
Note: You may experience a sporadic SSO issue when connecting to HPE AI Essentials. Just close the browser tab and try again if this issue happens.

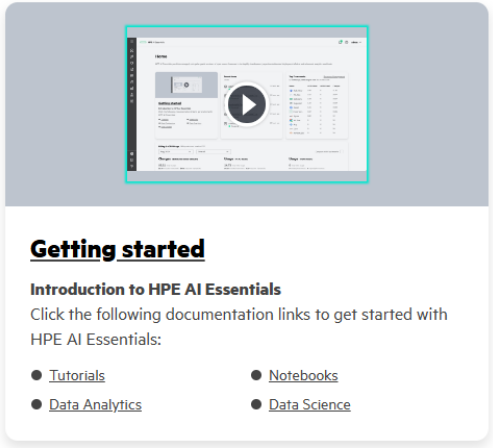
The Interactive Tour will help you navigate the different menus, and the video Introduction to HPE AI Essentials provides a quick explanation of all features proposed by the platform.

If it is your first time connecting to HPE PC AI, the video can be very useful.

We will also have a tour of these features in the next parts of this demonstration guide.







2. Quick tour of HPE AI Essentials

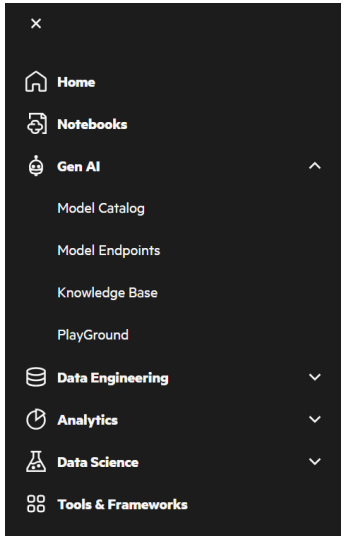
Let’s start the tour of HPE AI Essentials, to explain all capabilities enabled by the platform.

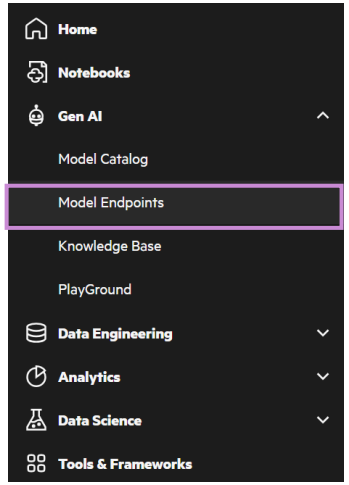
To start, go to Gen AI, Model Catalog.

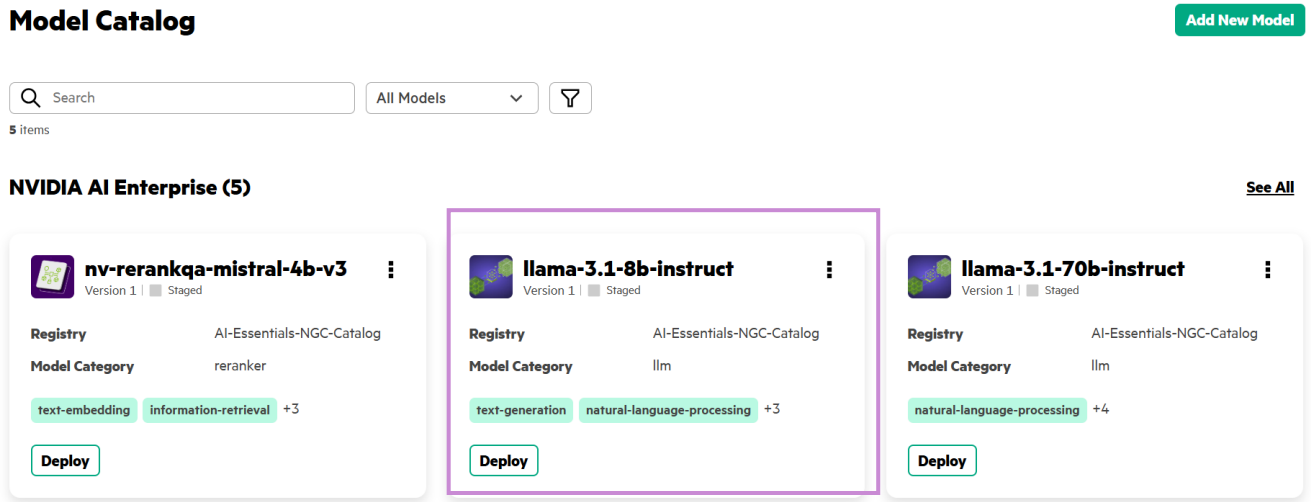
You can see there the pre-installed models that can be quickly deployed on Private Cloud AI to server AI applications. You can see for example the llama-3.1 model.

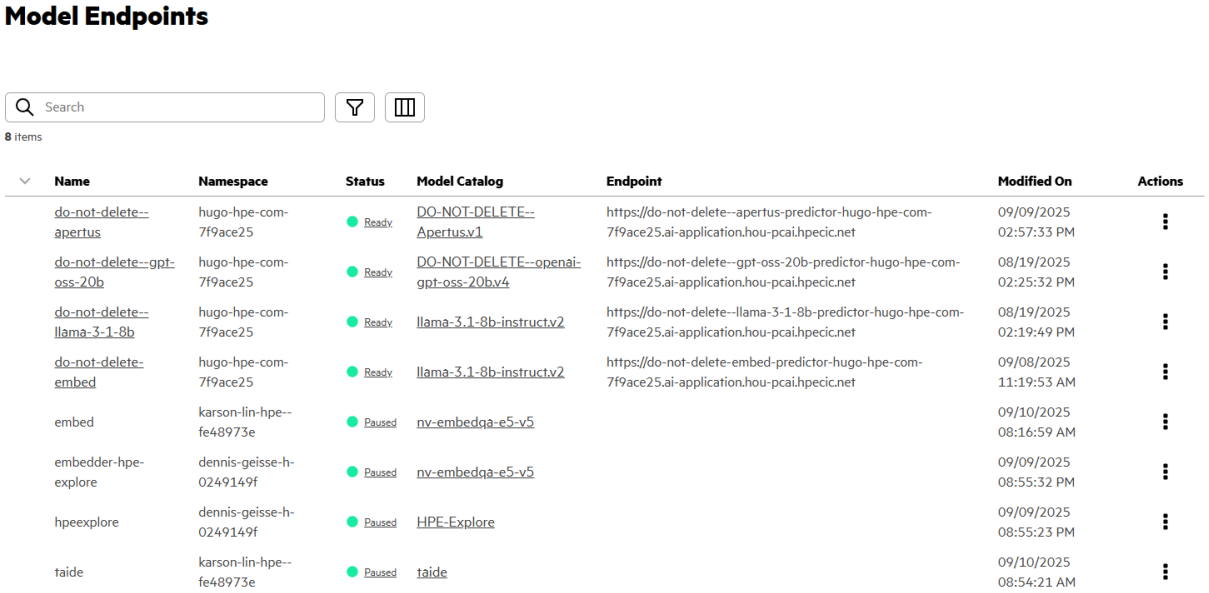
Please do not stop or delete models that start with « do-not-delete »

Next you can go to Model Endpoints, where you can see the deployed models, serving the AI Applications.







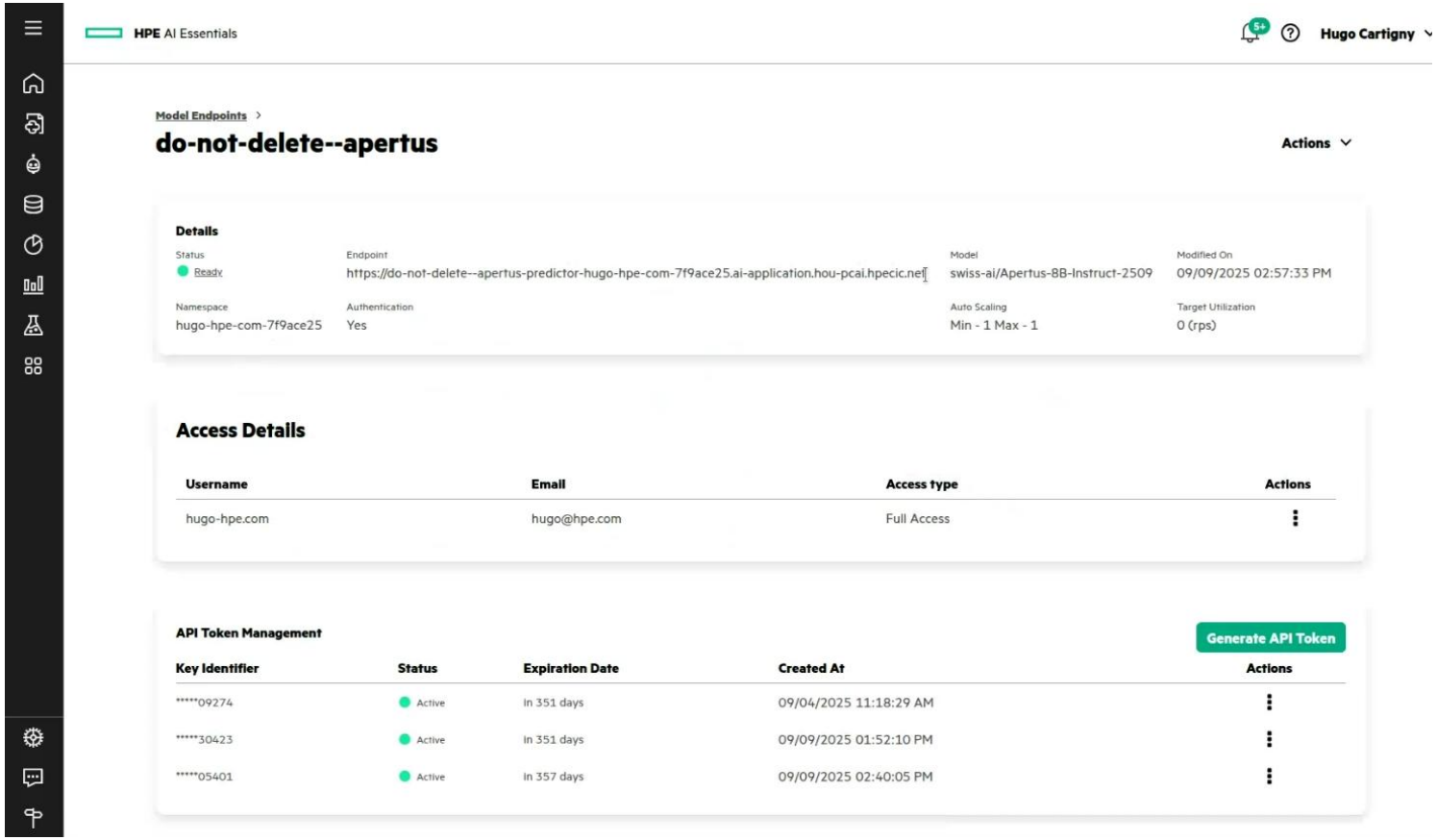


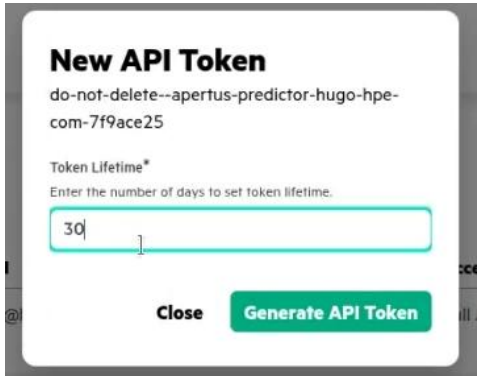
2. Quick tour of HPE AI Essentials

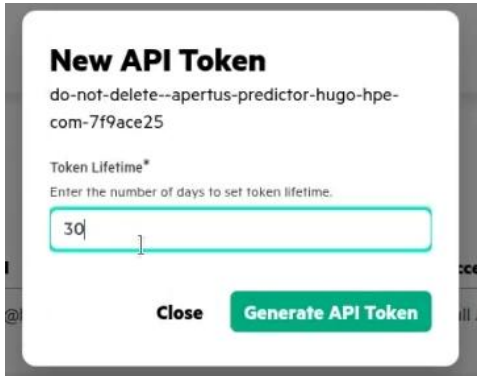
If you select a model, you will see more details:

Endpoint: URL for applications to use the model

Generate API Token: Token needed for applications to identify when accessing the model.







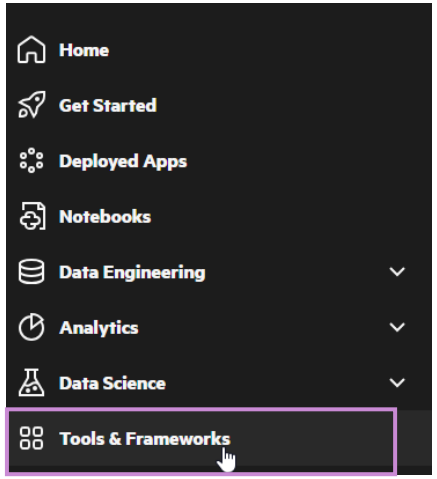
2. Quick tour of HPE AI Essentials

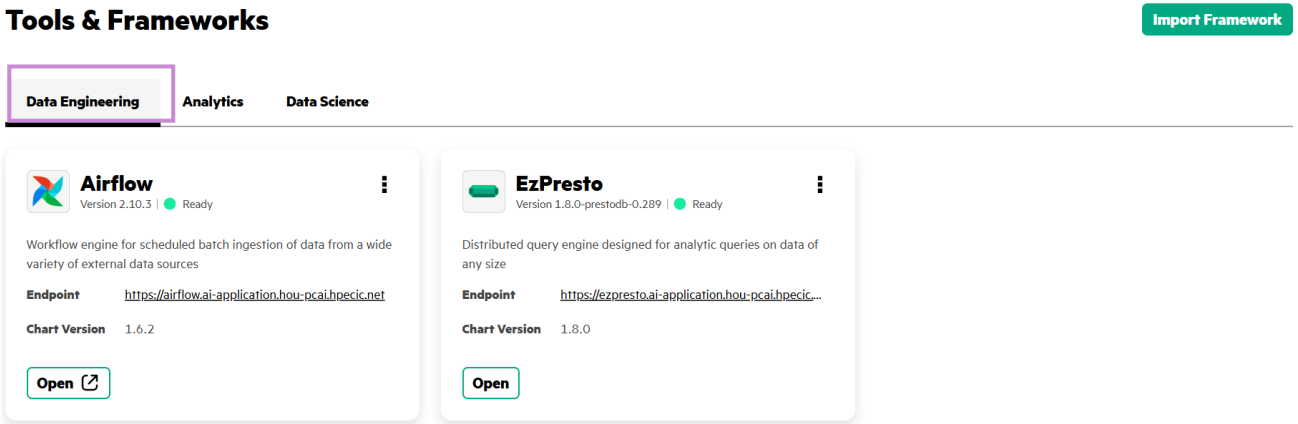
HPE AI Essentials also includes a curated set of Open Source applications to enable enterprises to build, deploy & operationalize analytics in GenAI applications.

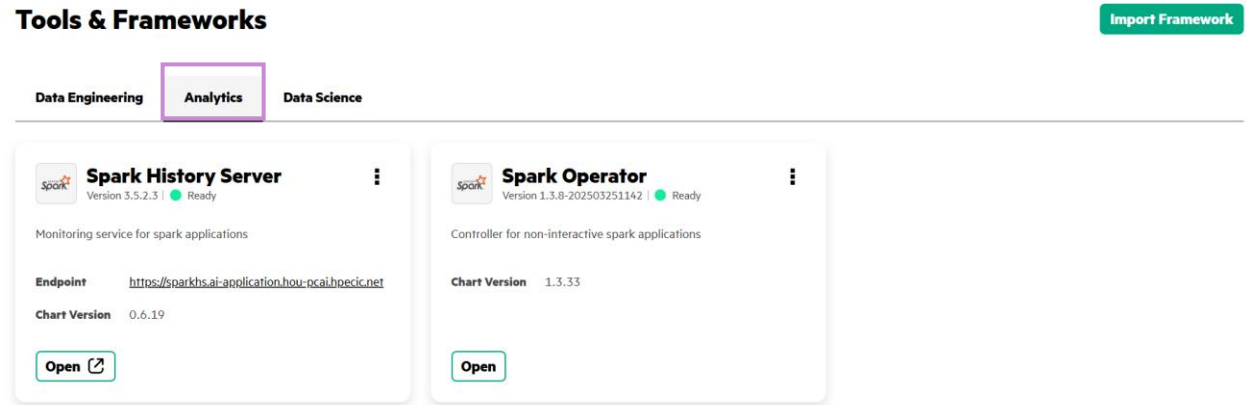
These Tools can be found in Tools & Frameworks, In the different tabs.

Data engineers have tools to design and implement data pipelines for ML and AI workloads, including

- Apache Airflow to connect a data set and automate data pipelines.
- Apache Spark to run large scale analytics workloads.







2. Quick tour of HPE AI Essentials

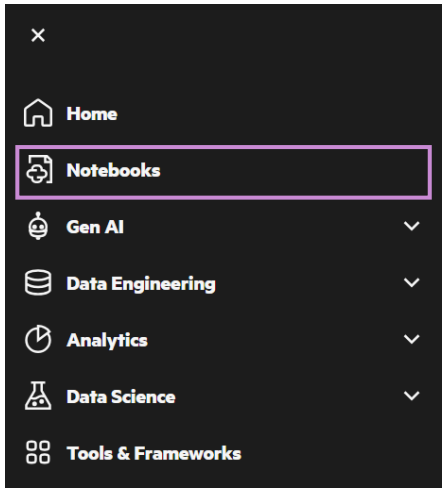
Notebooks:

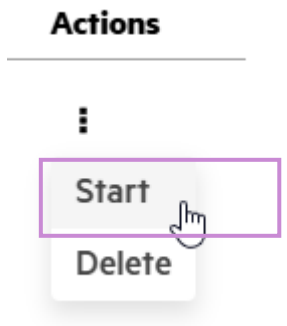
Every PC AI User is allocated a secure JupyterLab environment to create notebooks, develop scripts, run experiments, and deploy applications. All within their own private workspace

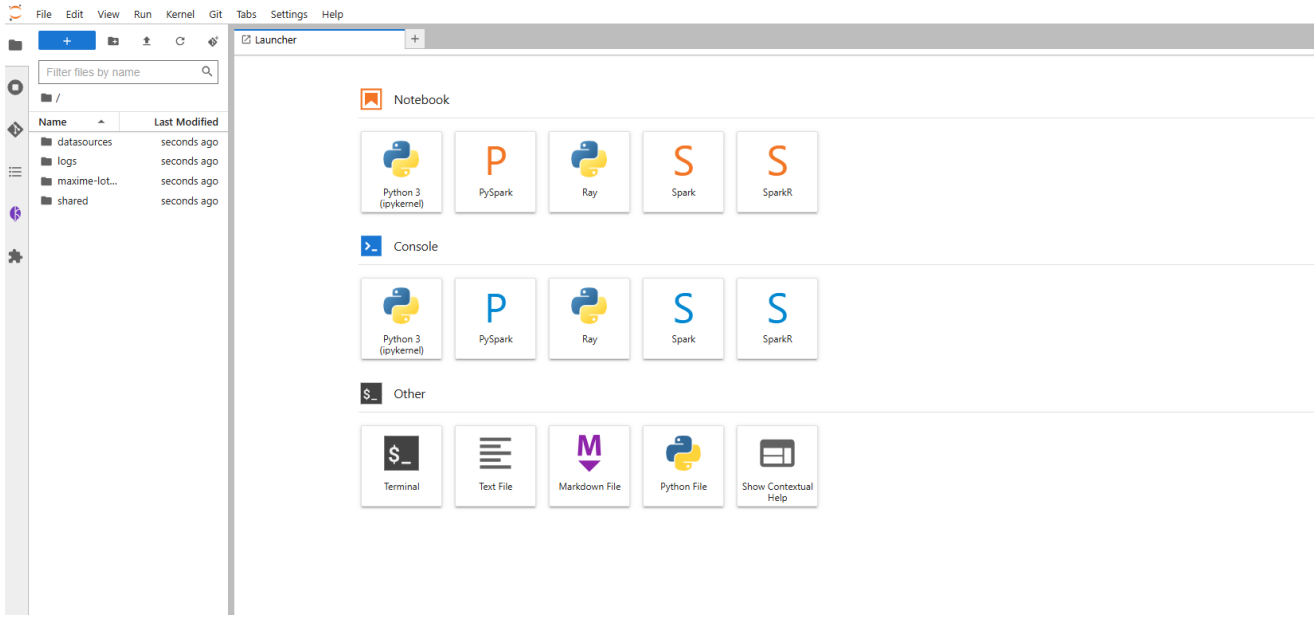
Notebooks automatically connect to the full suite of integrated tools and frameworks, and come pre-packaged with additional popular frameworks.

Notebooks also include pre-packaged tutorials, with predictive analysis examples to use as references when developing models.

You may need to start your Notebook to be able to open it, you can do this in Actions, Start







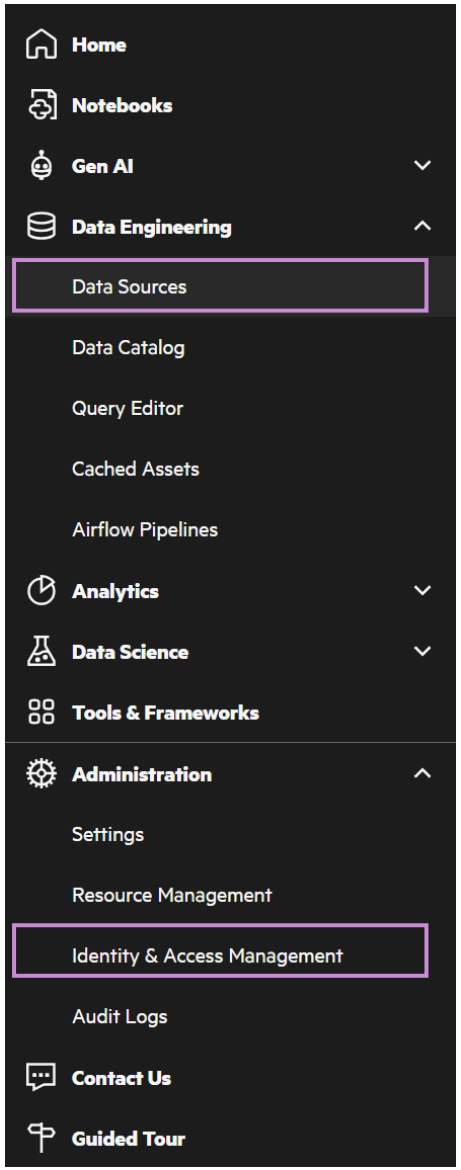
2. Quick tour of HPE AI Essentials

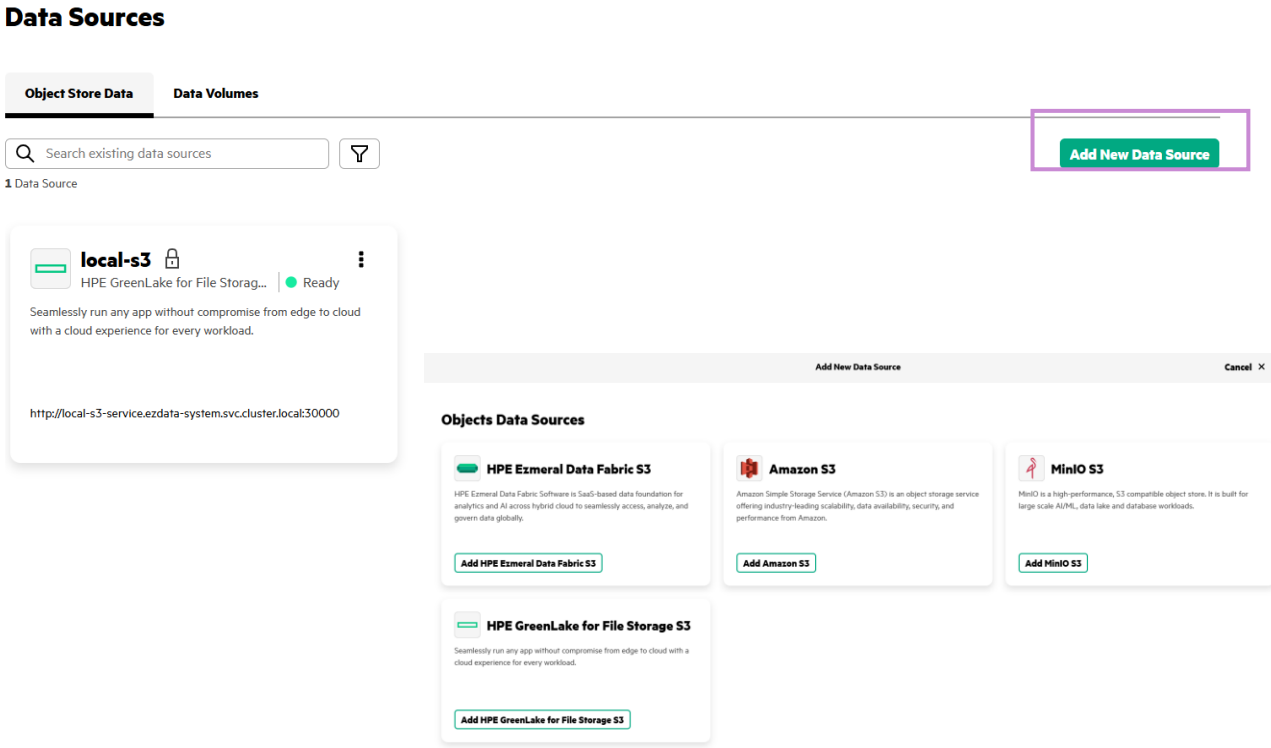
Data Sources:

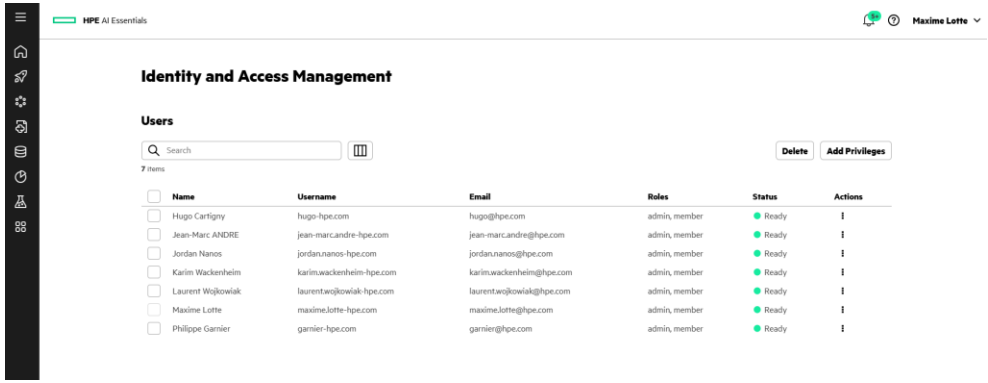
HPE AI Essentials utilizes HPE GreenLake for File Storage as the default File & Object storage.

However, PC AI Administrators can also connect external Data Sources, including relational Data bases, S3 object Storage, and Data Volumes in File Storage.

Once Data Sources are connected, an administrator can grant permissions on data through Identity & Access Management







2. Quick tour of HPE AI Essentials

Administration:

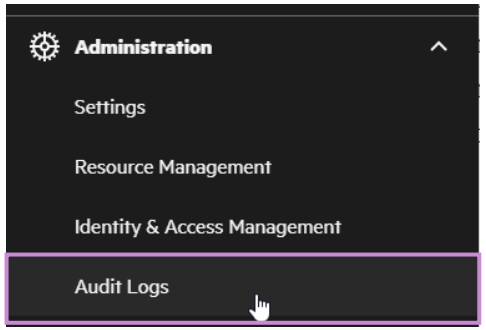
Administrators can audit user operations and manage GPU Resources.

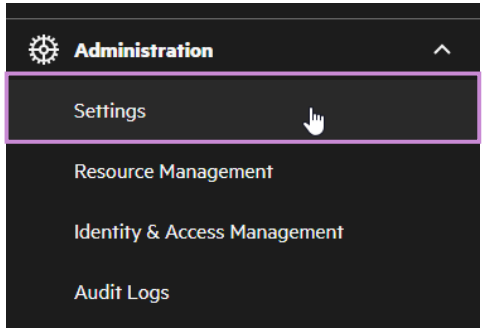
On the Settings page, Admins can view cluster config, add endpoints, and modify timeouts.

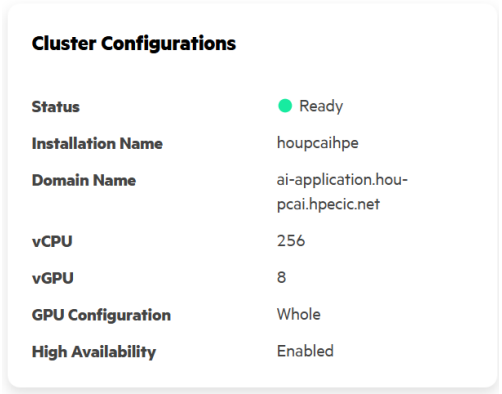
On the Tools & Frameworks tab, administrators can configure the installed tools & frameworks, or install the included tools and frameworks that are not installed by default, such as Superset and Feast.

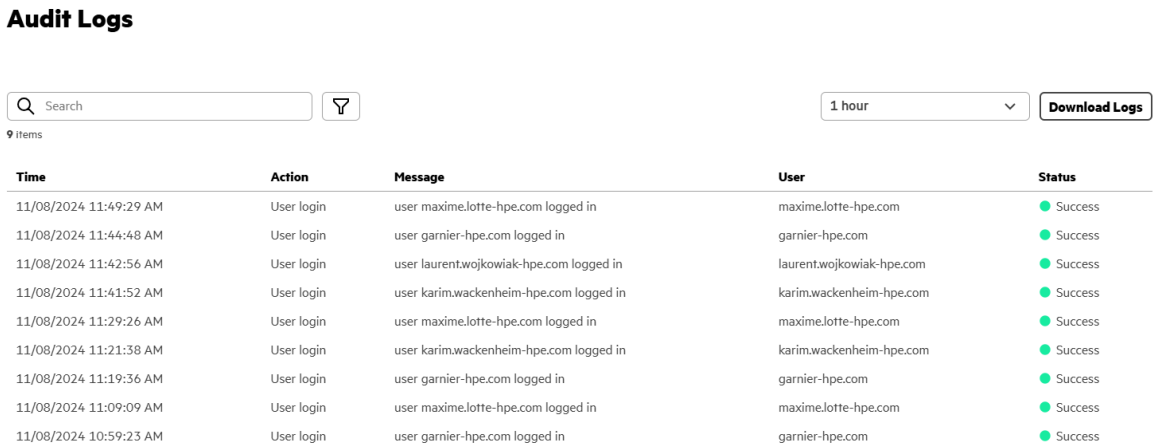
This concludes the HPE AI Essentials overview.

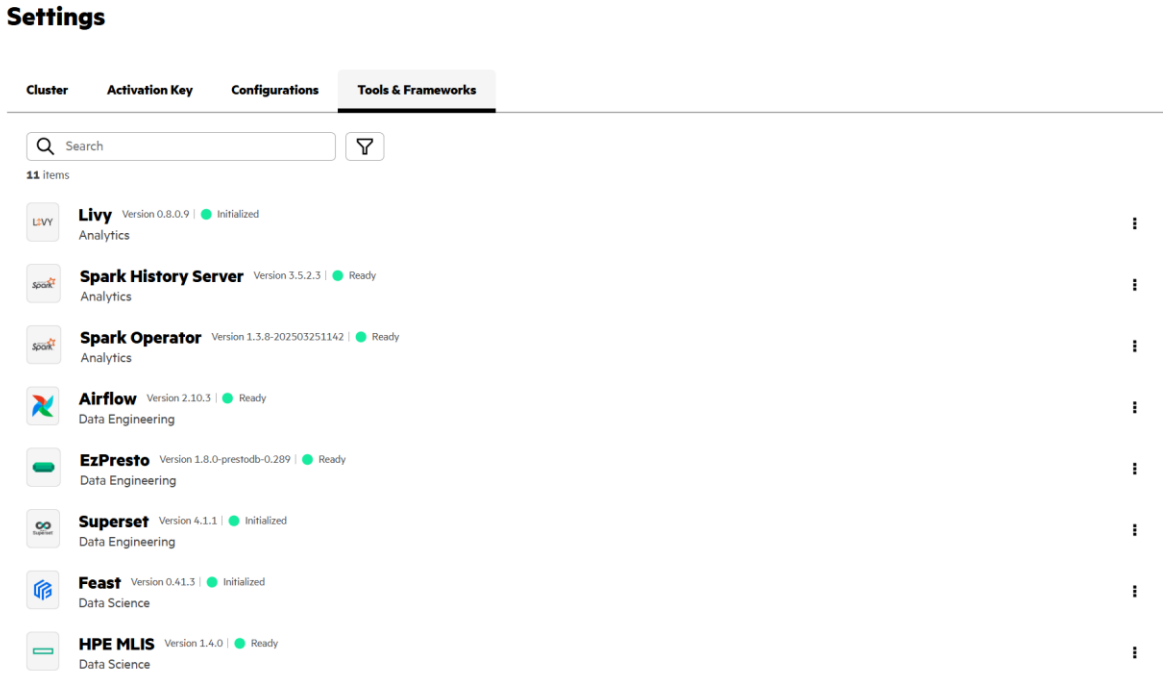
Let's move to the next step of the demonstration.











3. Run an imported application to start interacting with models

3. Run an imported application

It is also possible in PC AI to import your own developed application, to run on the platform.

An application has been already imported, to demonstrate this feature.

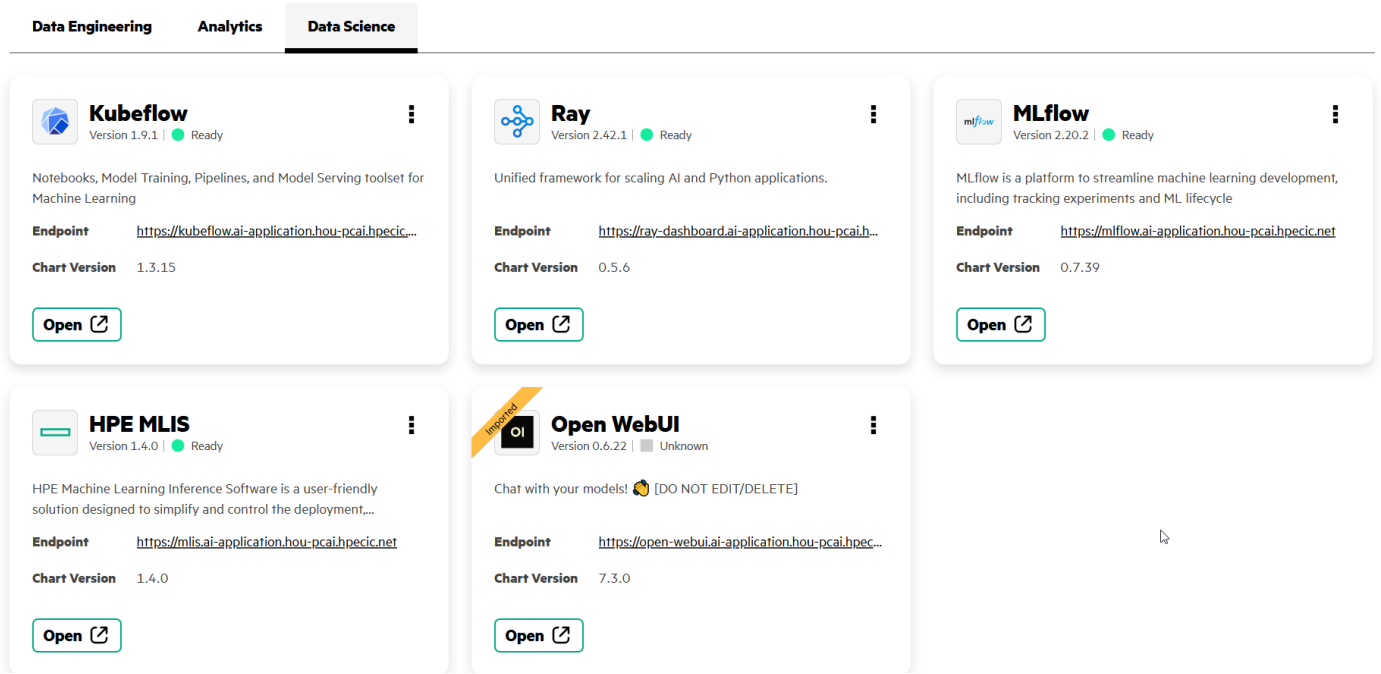
Go back to Tools & Frameworks, go to Data Science. You will find there the application Open WebUI.

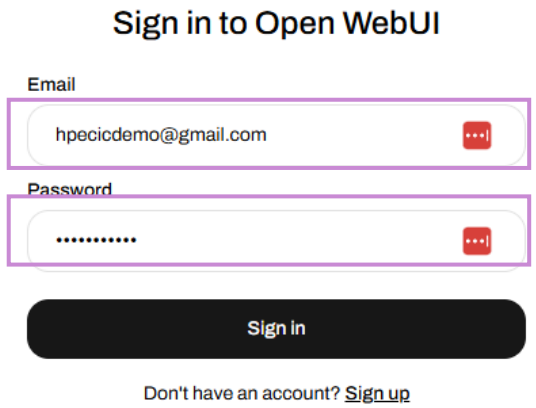
Click Open, and log in the application with:

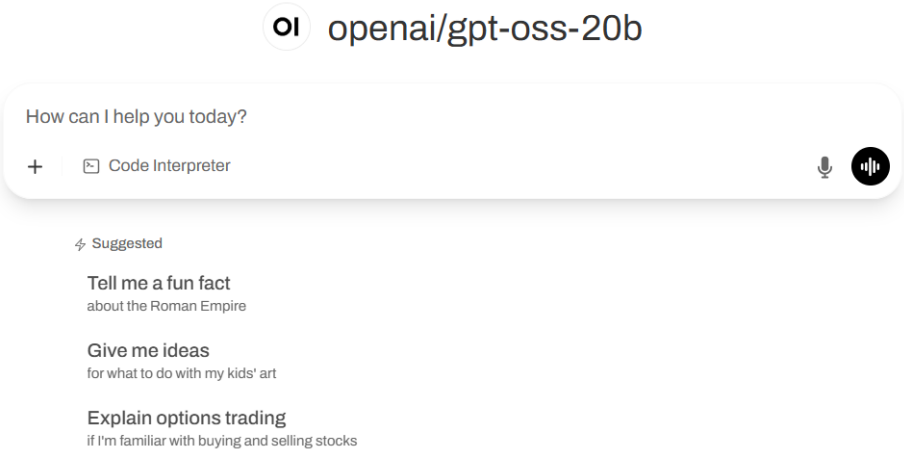
hpecicdemo@gmail.com

ITAdmin1234

You are logged in the application, that is a Chat-GPT like engine, used to test the different LLM Models.







3. Run an imported application

In the top left corner, you can select which model you would like to interact with. Here you can see multiple options available.

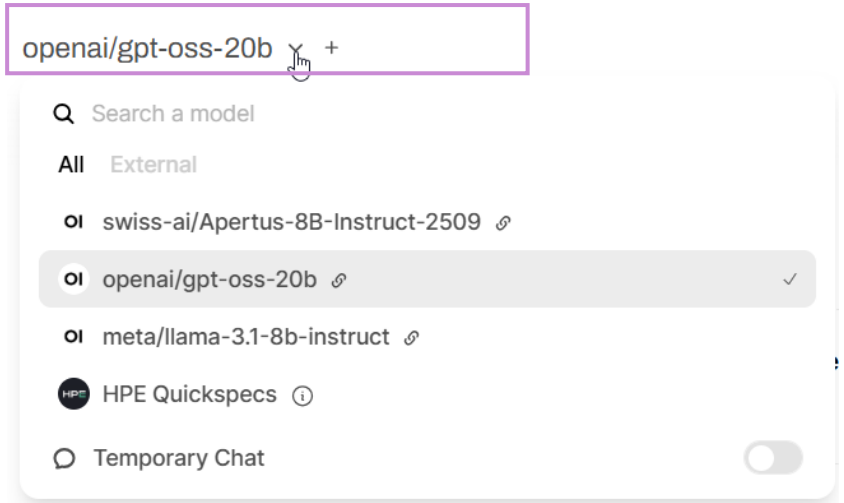
Apertus is the model pushed by Swiss AI, trained on the Alps Supercomputer provided by HPE to Swiss AI.

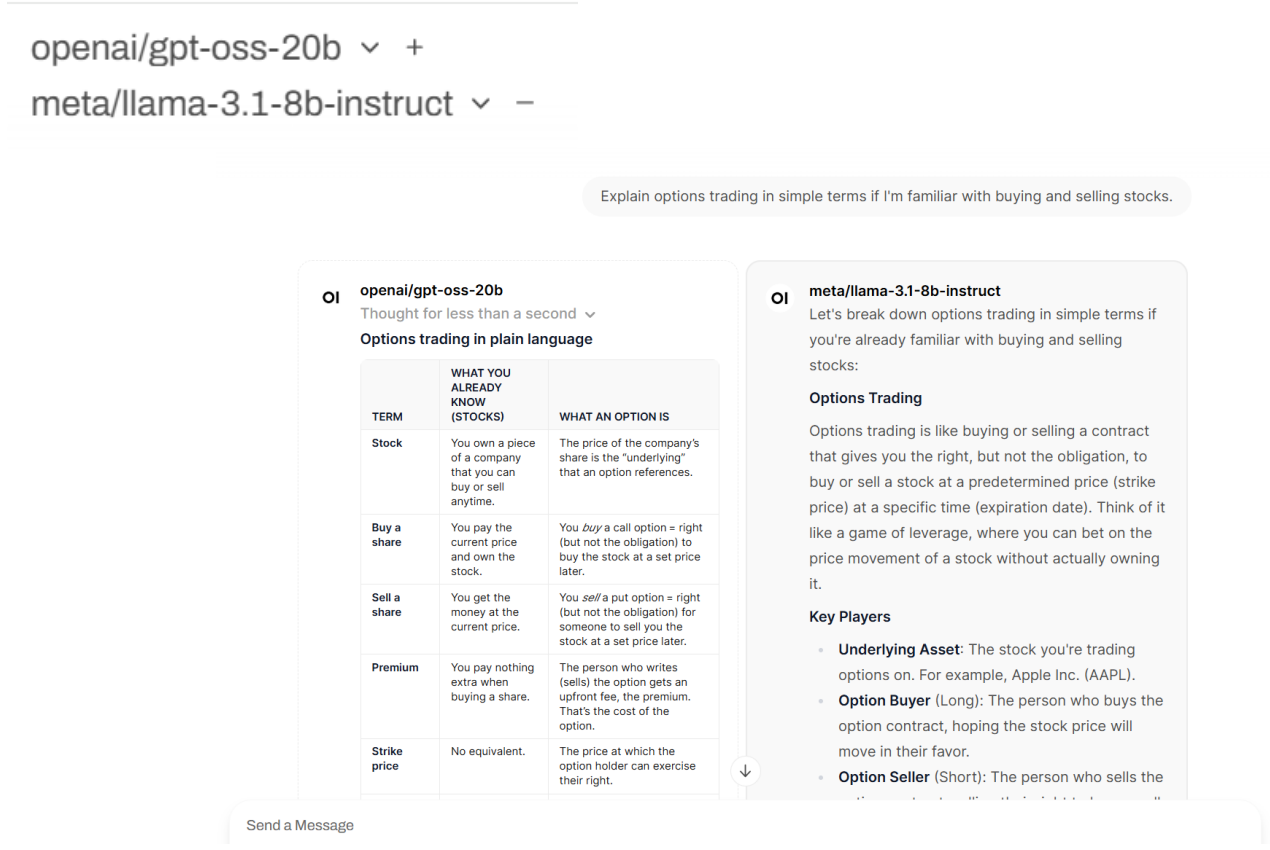
GPT-oss: is the Chat-GPT model hosted on-premises. This model has reasoning capabilities and is unique in that regard.

Llama-3.2-Vision: Is a multi-modal model and supports text and image as prompts.

Clicking the + button will allow you to add a second model, and to see the responses that 2 models provide for the same prompt:

Here we see the result for 2 models (OpenAI and llama) for the same prompt:

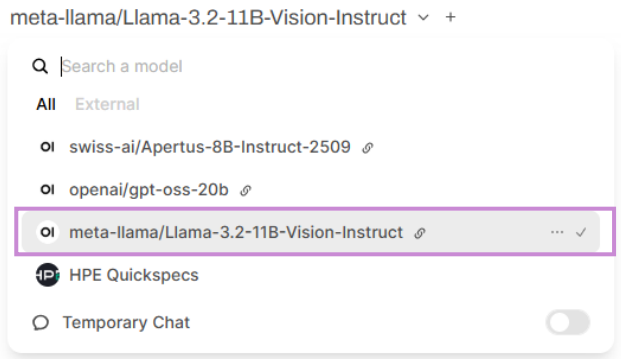


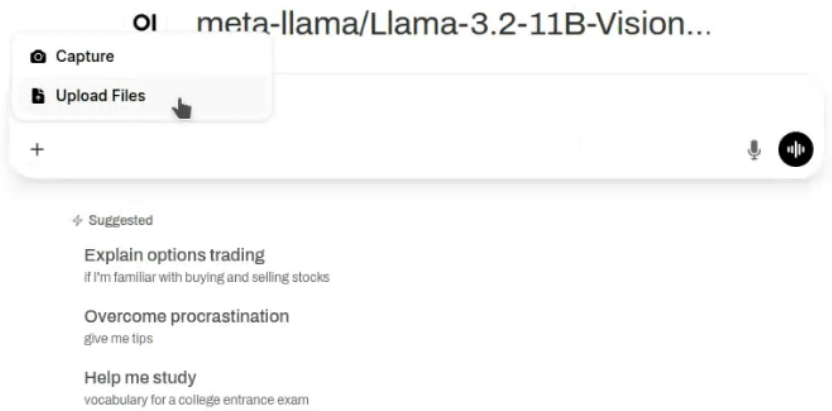


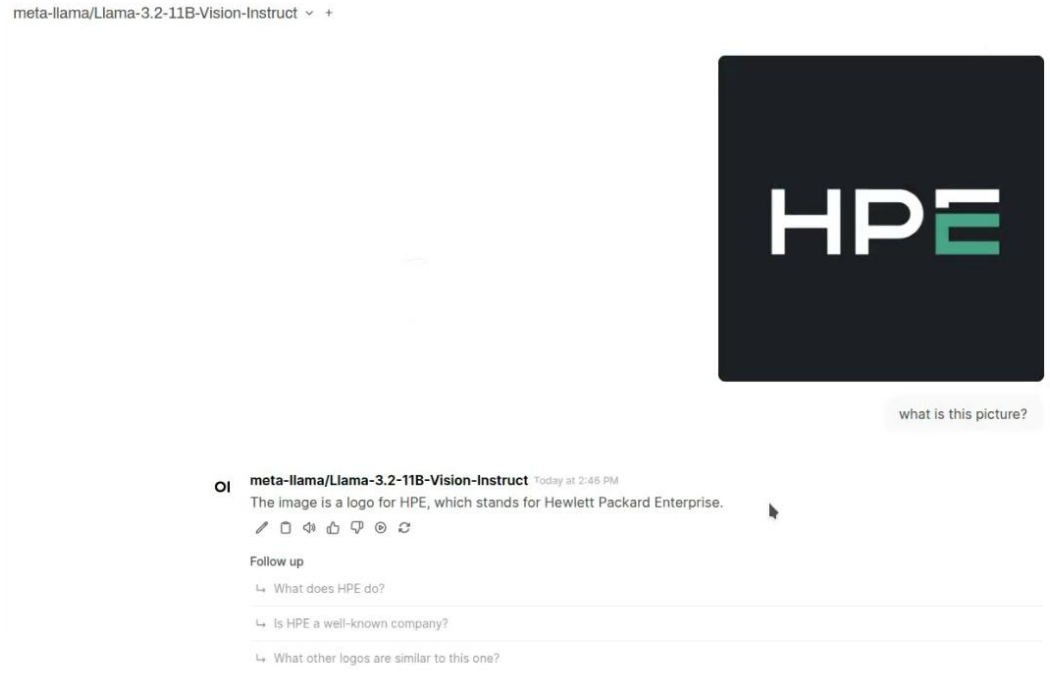
3. Run an imported application

Another test you can run is the Llama-3.2 model with Vision. This model will allow you to upload a file to analyze it.

Upload a file of your choice, here for example the HPE logo, and you can start a conversation from here as you can see in the screenshot.



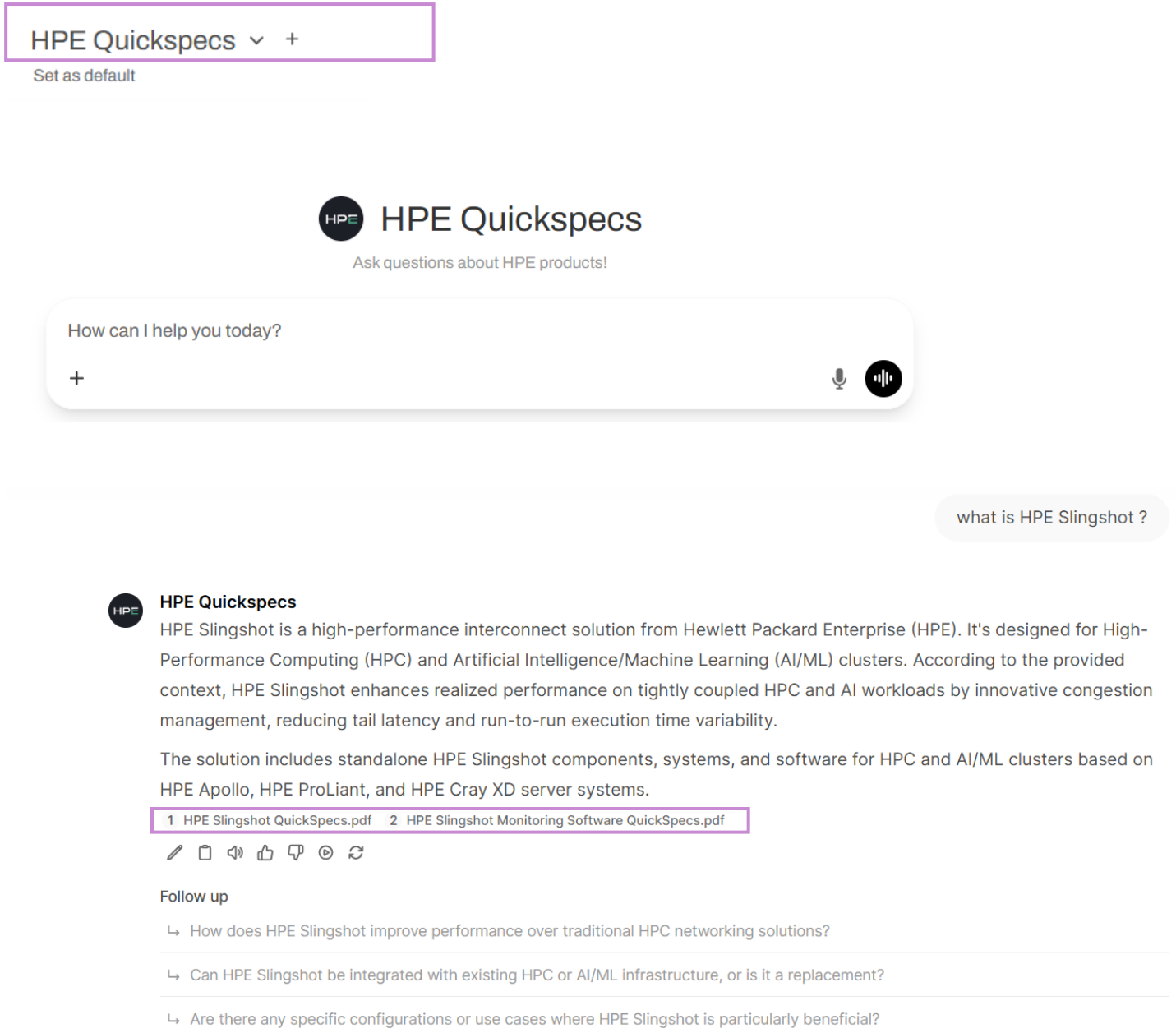




3. Run an imported application

Finally, you can the HPE Quickspecs, to interact with a model enhanced with RAG, with specific information on HPE products & technology.

Here the answers will leverage the additional context provided, and the model will mention the source for the answer that was formed.

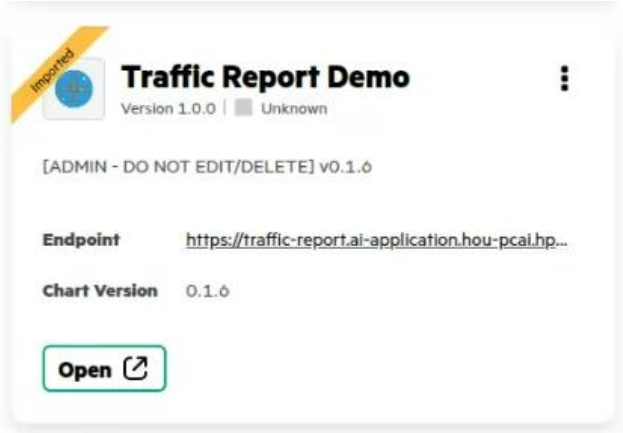


4. Additional demonstration use-cases

4.1 Traffic Analysis

A new application has been deployed to demonstrate AI in action on HPE PC AI.

This demo is the Traffic Report Demo, that you can open from the Tools & Frameworks, Data Science :

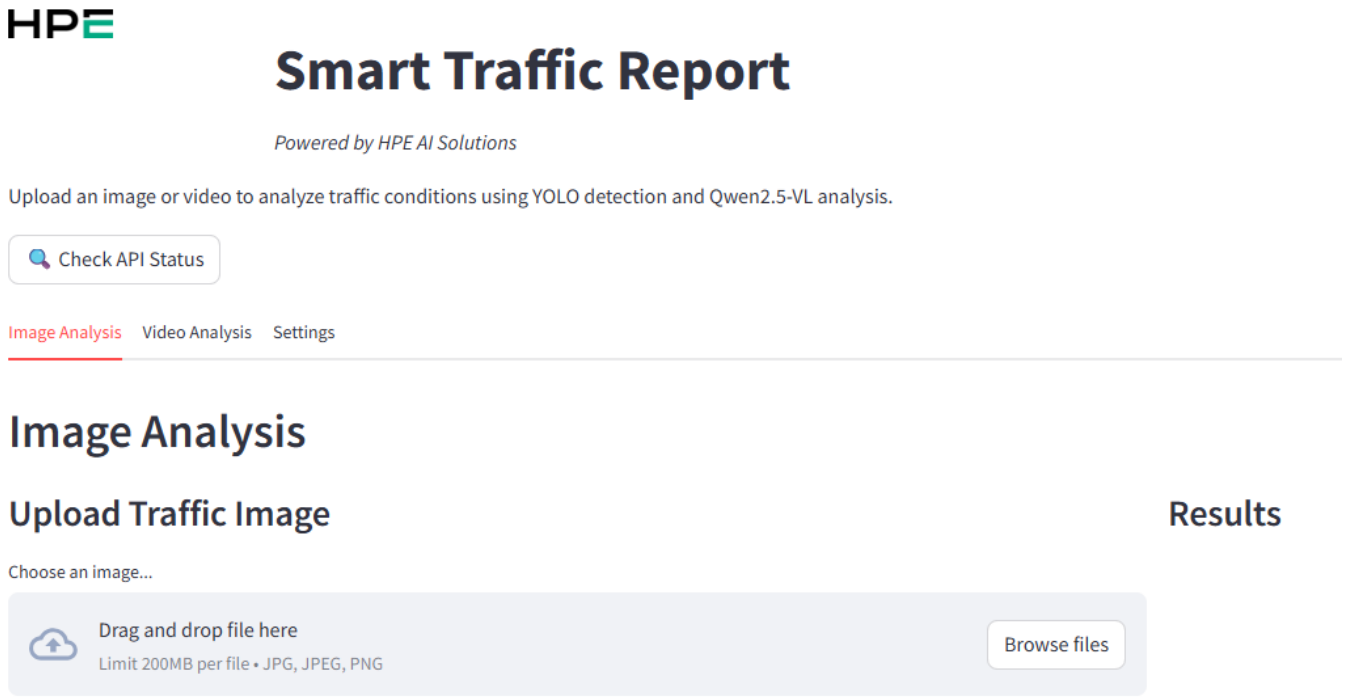


Here we leverage 2 models to achieve the objects identification in the image, and then the analysis of the traffic.

Model YOLO v8: defines bounding boxes around identified elements, with a confidence index on what has been identified.

Llama-Multimodal: we send to this model the source image and the output from the YOLO model. And we ask the Llama model to analyse and describe the model.

Let's see this in action.



4.1 Traffic Analysis

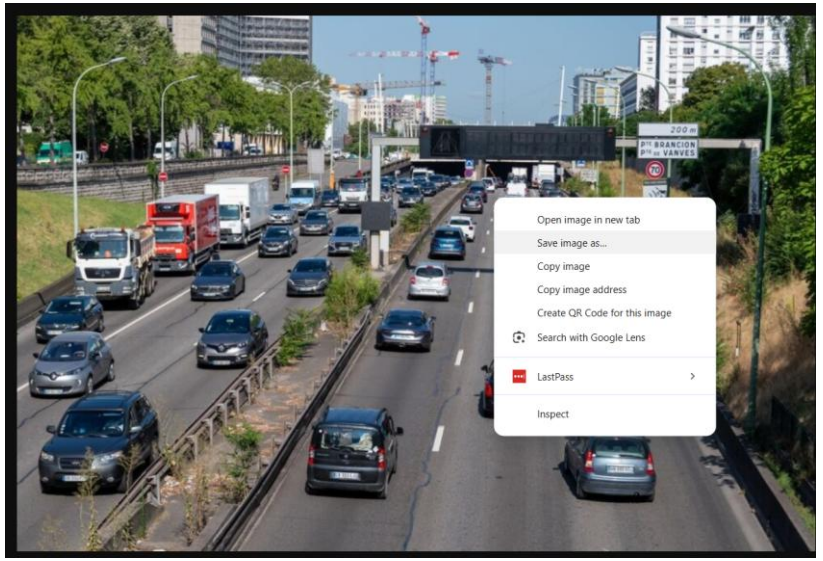
You will need pictures to upload to the model.

You can easily download 2 images that work well for this demo, here:

https://hdpgvaanimateddemos.hpecic.net/images/659362.jpeg

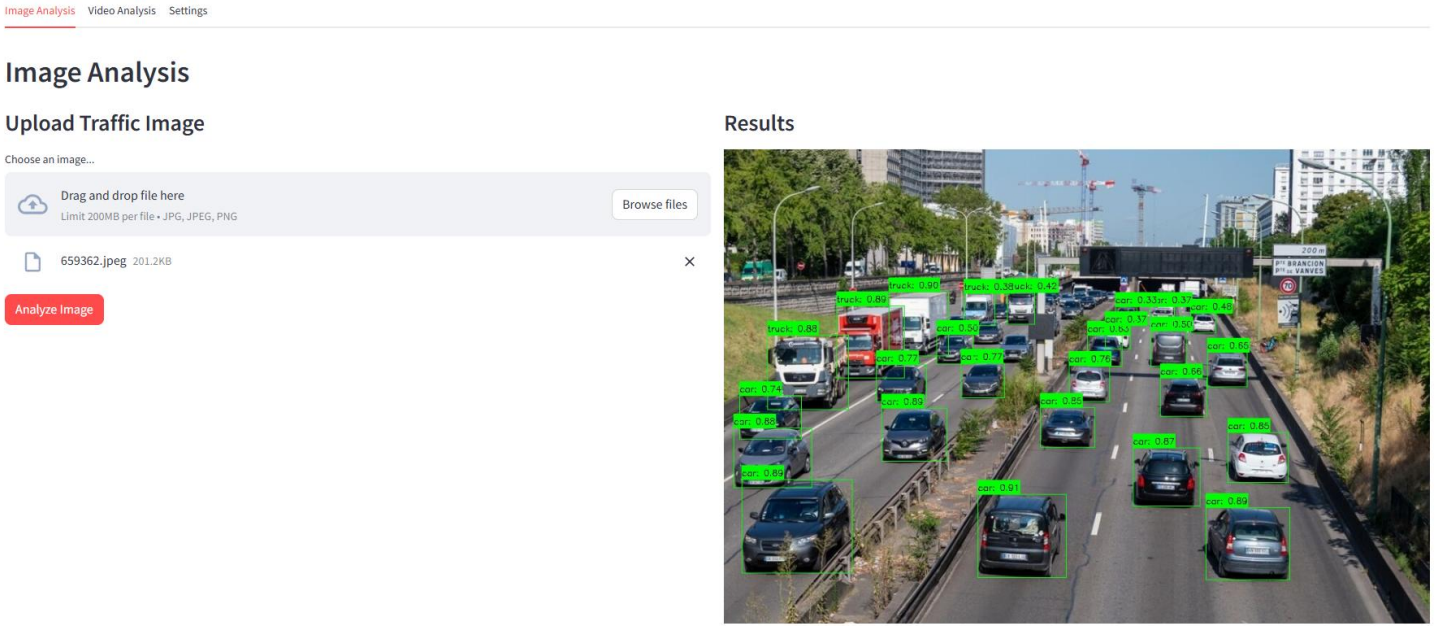
https://hdpgvaanimateddemos.hpecic.net/images/14951887_061424-wls-i355-rollover-crash-10p-vo-vid.jpg

Save the images to your PC.



Then you can upload the image.

First you will see the result of the Image objects recognition by the YOLO model, green rectangles with the confidence index for each object detected.



4.1 Traffic Analysis

Below the picture, you will see the analysis of the picture, the AI report, formed by the second model, the Llama model.

You can test as well with the second picture, the results are very interesting as well, with a Safety Assessment and Recommendations.

Detected Vehicles

- **Car:** 21
- **Truck:** 5

AI Report

- **Traffic Status: Normal**
- **Traffic Flow: Moderate**
- **Incident Analysis: No incidents or unusual situations observed.**
- **Safety Assessment: No safety concerns or violations observed.**
- **Recommendations: Maintain normal traffic flow.**

Image Analysis

Upload Traffic Image

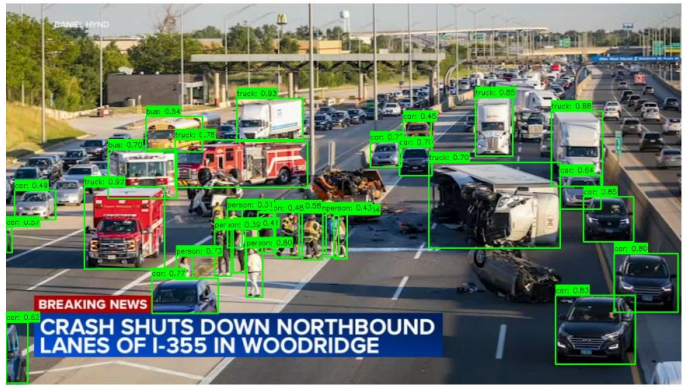
Choose an image...

Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG
Browse files

14951887_061424-wls-i355-rollover-crash-10p-vo-vid.jpg 0.6MB

Analyze Image

Results



Detected Vehicles

- **Truck:** 6
- **Car:** 11
- **Person:** 10
- **Bus:** 2

AI Report

- Based on the provided image, here is a concise analysis of the traffic scene:
- **Traffic Status: ABNORMAL**
- **Traffic Flow: HEAVY**
- **Incident Analysis: A multi-vehicle crash has occurred in the northbound lanes of I-355 in Woodridge, involving 6 trucks, 11 cars, and 2 buses. The incident has caused significant disruptions to traffic flow.**
- **Safety Assessment: There is a high risk of injury or fatality due to the severity of the crash and the number of vehicles involved.**
- **Recommendations: Emergency services should be dispatched immediately to assist with the incident, and traffic authorities should consider diverting traffic around the affected area to minimize further congestion and ensure public safety.**

4. HPE PC AI monitoring in OpsRamp

4. Monitor PCAI with OpsRamp

Accessible from HPE GreenLake Platfom, the OpsRamp service provides IT Operations and Monitoring capabilities.

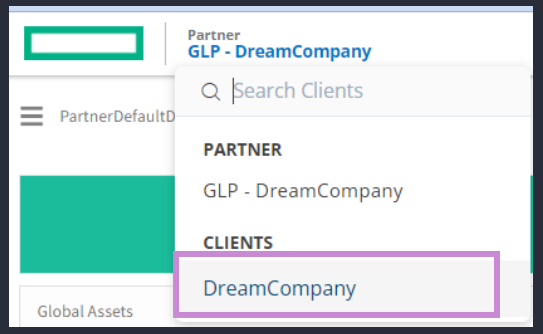
Select OpsRamp and launch

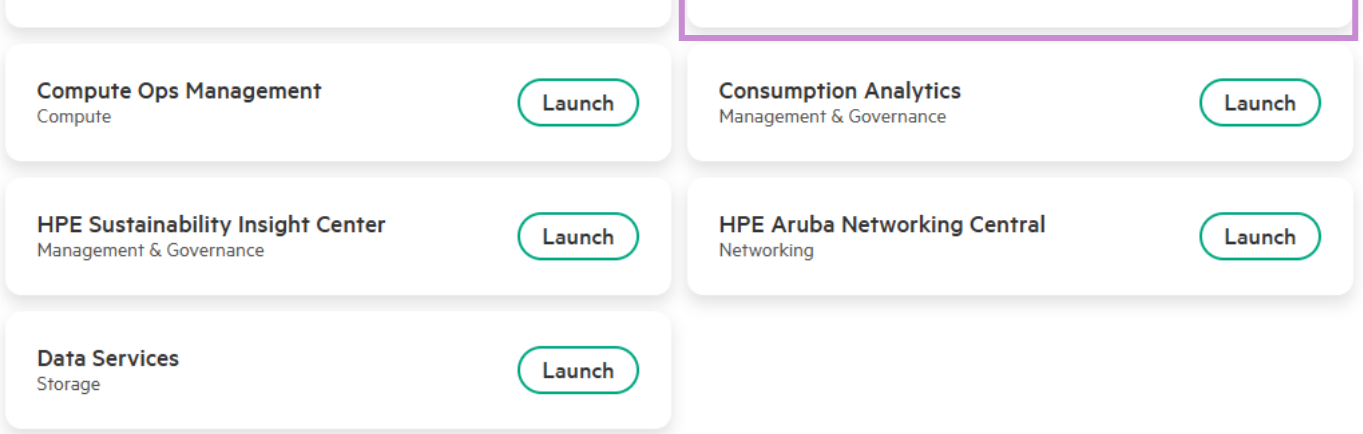
You are connected to OpsRamp as a Partner.

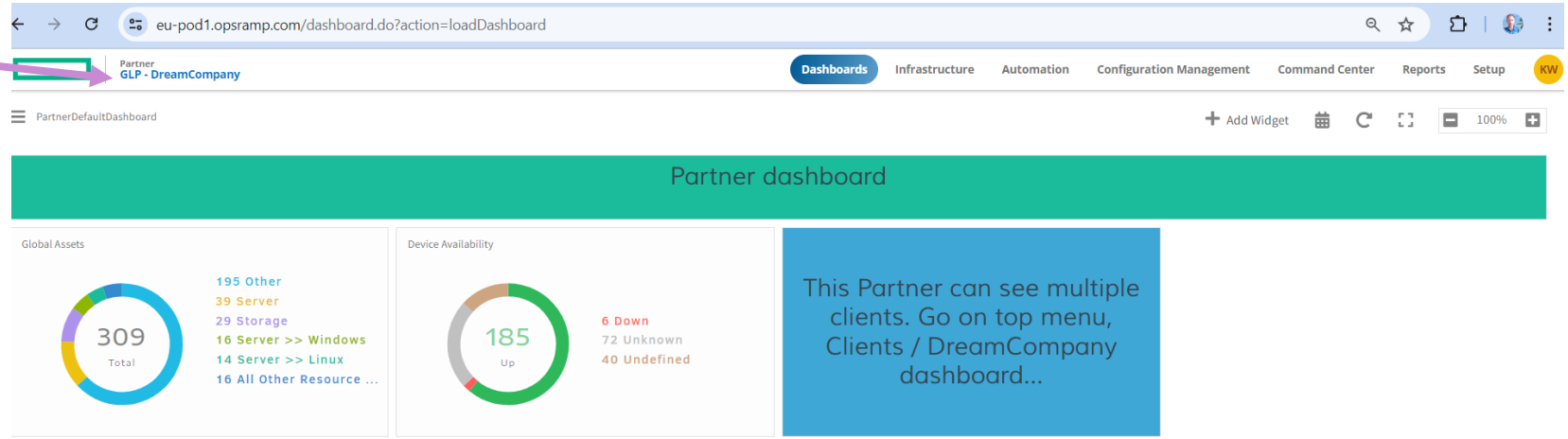
A partner can manage multiple clients.

Go on the top menu

Select DreamCompany client

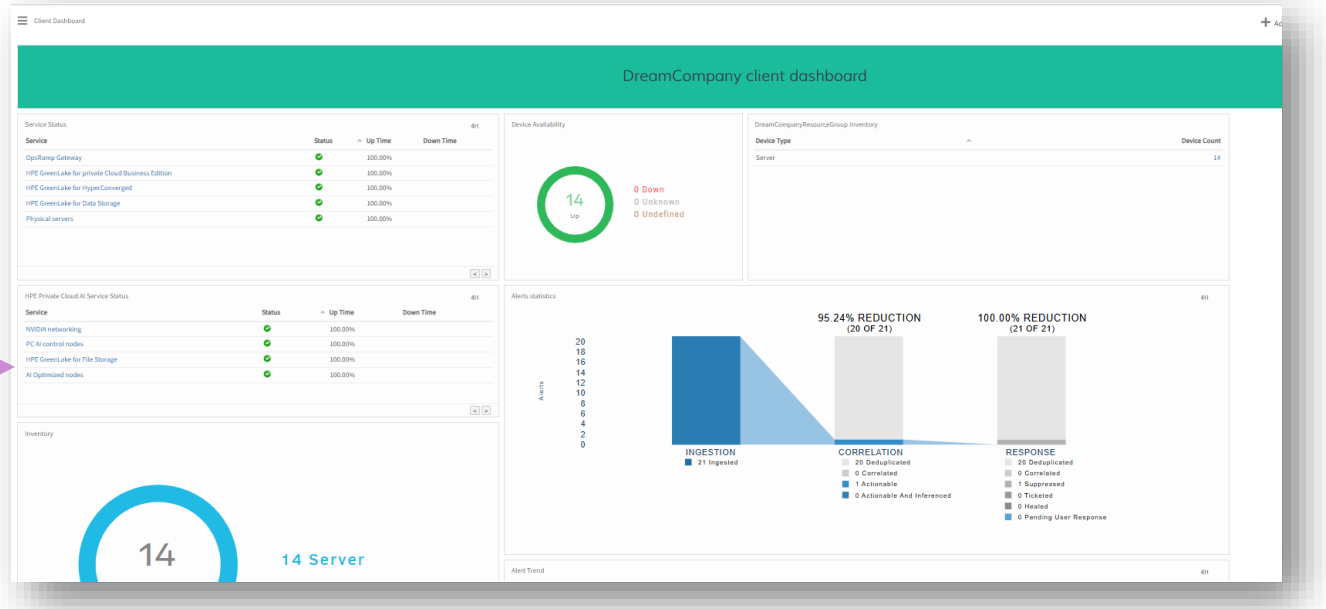






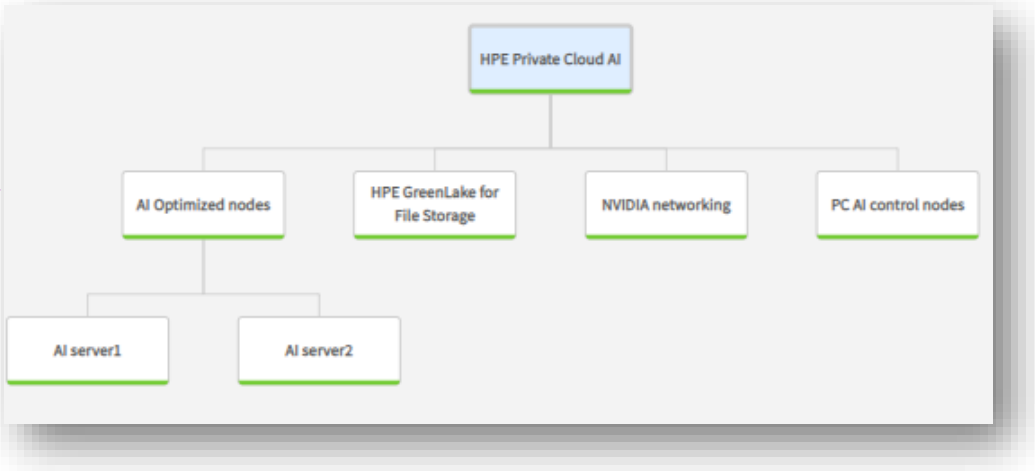
4. Monitor PCAI with OpsRamp

DreamCompany client dashboard shows various services monitored: including HPE private Cloud AI



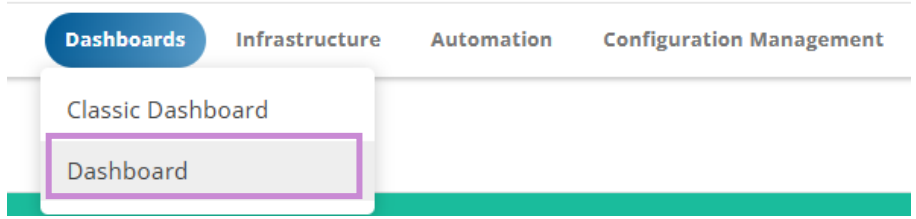
Go in Infrastructure / Service Maps

to view HPE private Cloud AI service and sub services monitored in OpsRamp

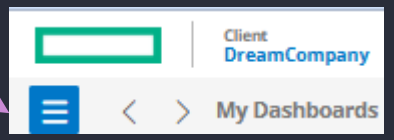


4. Monitor PCAI with OpsRamp

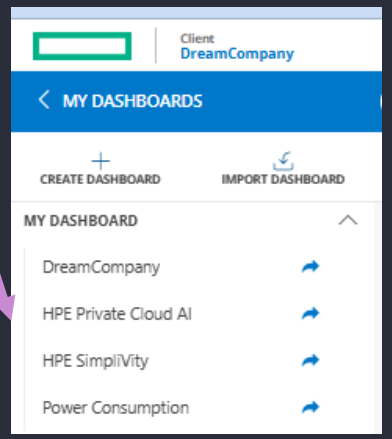
Go to the Dashboard menu



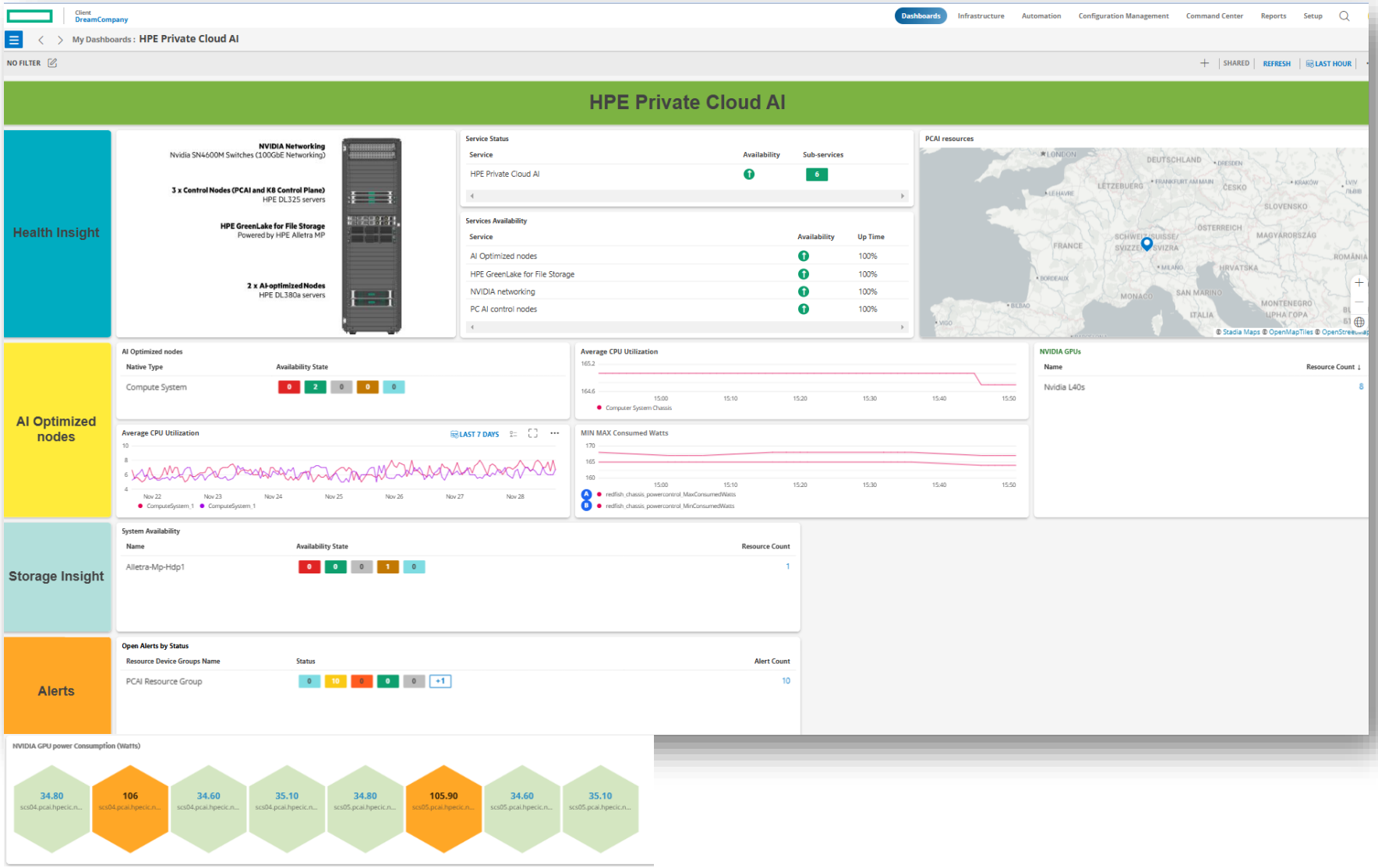
Click on



to see predefined dashboards
Open HPE Private Cloud AI



PCAI Heath, metrics and alerts are shown



Thank you for using the HPE Demonstration Portal !

Do not hesitate to reach out for any improvements or suggestions to this live demonstration:

contacthpedemoportal@hpe.com

Thank You