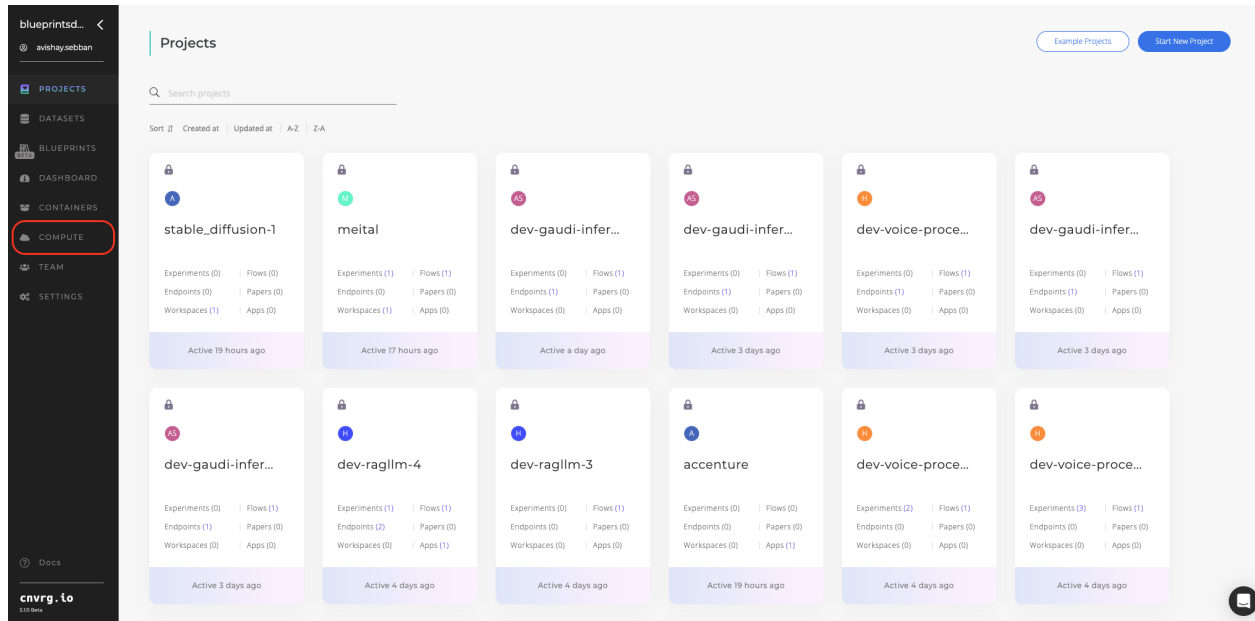


Workshop run stable diffusion in cnvrg.io: a step by step guide with Gaudi

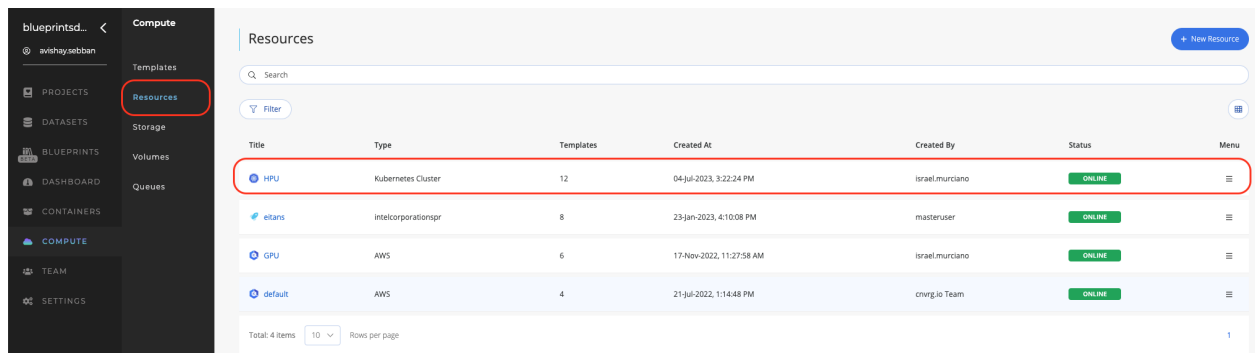
cnvrg.io is a machine learning platform that streamlines the development and deployment of machine learning models. It provides tools for version control, collaboration, model training, and deployment, allowing data scientists and machine learning engineers to focus on building and improving models, rather than worrying about infrastructure and other technical details.

Using the cnvrg.io platform, it is easy to run a stable diffusion model and generate your own custom image.

1. Login to your account.
2. Once you've entered cnvrg.io, click on resource on the left side



3. In cnvrg.io you are able to connect any kind of workload to the platform. Make sure you have Gaudi HPU integrated in your environment.



4. Create a template to assign the relevant number of CPU MEMO and HPU.

Make sure you choose the relevant resource to assign for.

Compute Templates

The available Compute Templates for Workspaces, Experiments, and model endpoints.

Xeon4

eilans

60 CPUs 100 GB

StableDiffusion10CPU

default

10 CPUs 20 GB

0 jobs

stable

eilans

16 CPUs 32 GB

0 jobs

SPR-large

eilans

4 CPUs 16 GB

4 jobs

smaller

default

0.5 CPUs 2 GB

0 jobs

smalli

HPU

1 CPUs 2 GB

0 jobs

small

default

2 CPUs 16 GB

0 jobs

small

GPU

1 CPUs 2 GB

0 jobs

+ Add Compute Template

HPU

eilans

GPU

default

5. Provide input title name - Gaudi Small

Insert the provided details.

CPU: 16

Memory: 32

HPU: 2

CPU

16

Memory

32

GB

GPU

HPU

2

Hugepages

Mi

Shared Memory

GB

Node Labels

Comma-separated key-value pairs

Node Taints

Comma-separated key-value pairs

PERMISSIONS

Public

Public

All users have access

Private

Select the users that can use this template (Administrators have permissions to all compute resources)

Select users


Save


And click on save.


6. Bring your own container


Click on **Add Image**


Registries


cnvrg


Intel


Habana


NVIDIA

Docker Hub

quay

habana ga...































habana

Add Registry

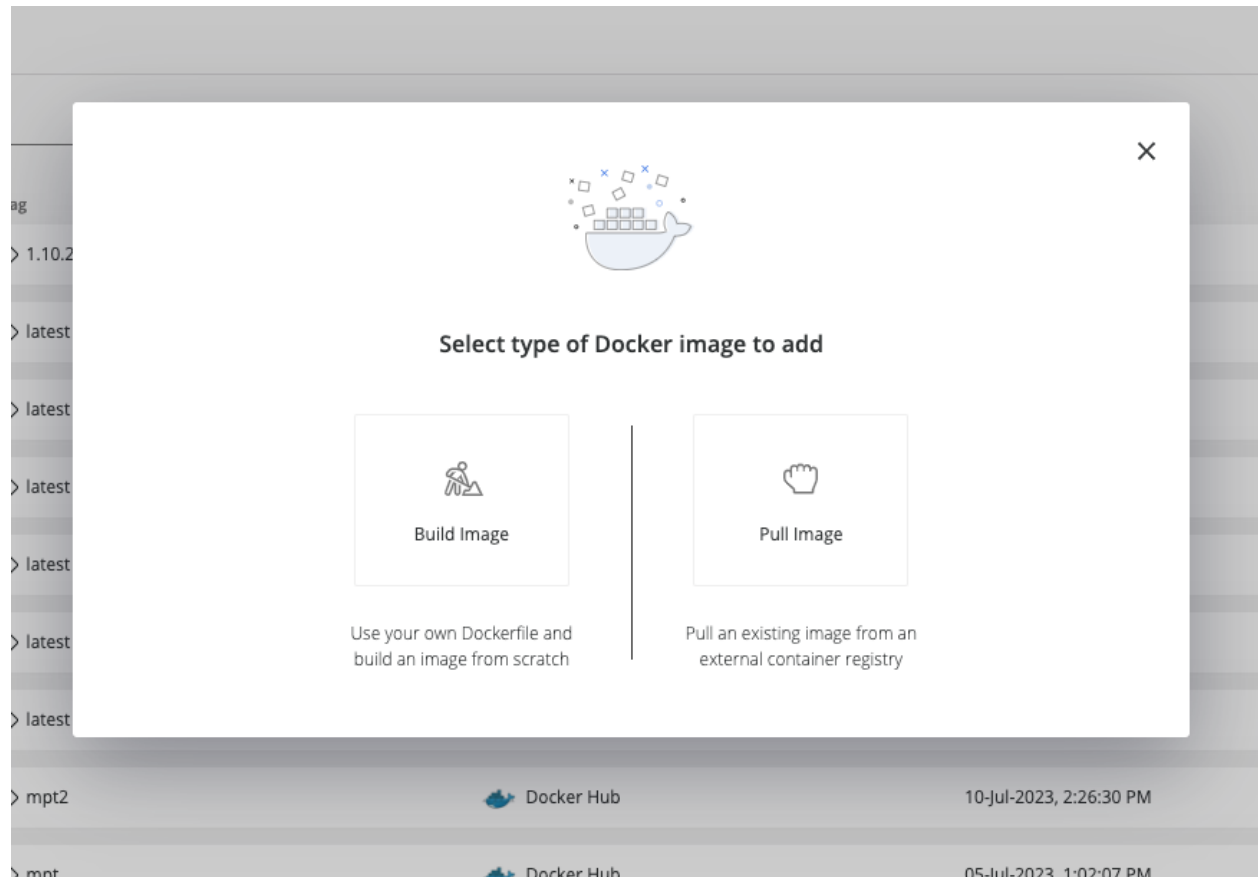
Images

Search Images

↓ Created At | Registry | Repository | Created By

Status	Repository name	Tag	Registry Name	Created at	Created by	
●	 pytorch-training-habana	1.10.2-hpu-py38-synapseai1.4.1-ubuntu20.04	 habana	12-jul-2023, 8:46:56 AM	avishay.sebban	
●	 gaudi-docker/1.10.0/ubuntu22.04/h...	latest	 Habana	12-jul-2023, 8:25:40 AM	avishay.sebban	
●	 gaudi-docker/1.10.0/amzn2/habana...	latest	 Habana	12-jul-2023, 7:44:23 AM	abhay.saini	
●	 gaudi-docker/1.10.0/ubuntu20.04/h...	latest	 Habana	12-jul-2023, 7:29:43 AM	abhay.saini	
●	 ui/native/gaudi-docker/1.10.0/ubuntu...	latest	 Habana	12-jul-2023, 7:25:42 AM	abhay.saini	
●	 ui/native/gaudi-docker/1.10.0/ubuntu...	latest	 Habana	12-jul-2023, 7:21:50 AM	abhay.saini	
●	 gaudi-docker/1.10.0/ubuntu20.04/h...	latest	 Habana	10-jul-2023, 2:29:15 PM	abhay.saini	
●	 harindercnvr/llm	mpt2	 Docker Hub	10-jul-2023, 2:26:30 PM	harinder.mashiana	
●	 harindercnvr/llm	mpt	 Docker Hub	05-jul-2023, 1:02:07 PM	harinder.mashiana	
●	 heziraha/convege_python3.8	spix_entry	 Docker Hub	22-jun-2023, 7:45:25 AM	harshul.thakur	

And choose **Pull image**



Provide the relevant information:

Image:

<https://vault.habana.ai/gaudi-docker/1.10.0/amzn2/habanalabs/pytorch-installer-2.0.1:latest>

choose the relevant **Registry:** Habana

Pull Image


Change logo 

Image URL
vault.habana.ai/https://vault.habana.ai/gaudi-docker/1.10.0/amzn2/habanalabs/pytorch-installer-2.0.1:latest

Registry
Habana

Repository
https://vault.habana.ai/gaudi-docker/1.10.0/amzn2/habanalabs/pyt

Tag
latest

README

Cancel Add

Click on the **Add** button.

7. Start your project - let's bring the code into a working project.

Projects

Search projects

Sort IT Created at Updated at A-Z Z-A

Example Projects Start New Project

stable_diffusion-1
Experiments (0) | Flows (0)
Endpoints (0) | Papers (0)
Workspaces (1) | Apps (0)
Active 20 hours ago

meital
Experiments (1) | Flows (1)
Endpoints (0) | Papers (0)
Workspaces (1) | Apps (0)
Active 18 hours ago

dev-gaudi-infer...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active a day ago

dev-gaudi-infer...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

dev-voice-proce...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

dev-gaudi-infer...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

dev-gaudi-infer...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

dev-ragllm-4
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

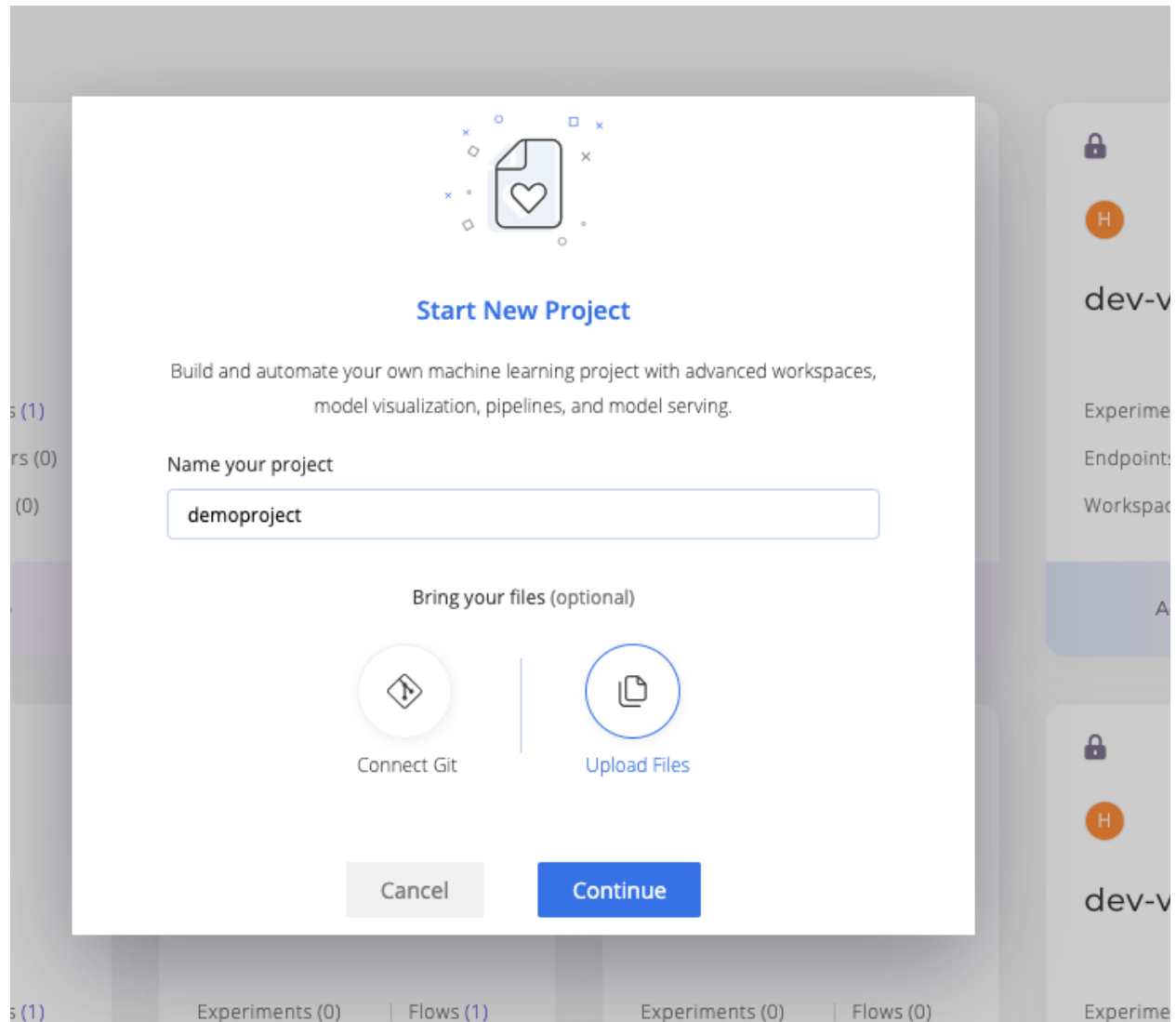
dev-ragllm-3
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

accenture
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

dev-voice-proce...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

dev-voice-proce...
Experiments (0) | Flows (1)
Endpoints (1) | Papers (0)
Workspaces (0) | Apps (0)
Active 3 days ago

Click on the project on the left side, and “Start new project” Name your project(for example : **demoproject**) and click submit.



8. Click on the blue 'new workspace' button to start a jupyterlab workspace so we can easily create, edit and run code.

The screenshot displays the 'dev-gaudi-inference-10' dashboard. On the left is a dark sidebar with navigation options: PROJECTS, DATASETS, BLUEPRINTS, DASHBOARD, CONTAINERS, COMPUTE, TEAM, and SETTINGS. The main content area has a header 'dev-gaudi-inference-10' and a sub-header with counts: Experiments (0), Endpoints (1), Workspaces (0), and Apps (0). Below this is an 'Activity Feed' table with columns for user, action, and time. The feed shows four entries: 'abhay.saini created endpoint inference-1021.', 'abhay.saini ran Version 1 of dev-Gaudi Inference.', 'abhay.saini created flow dev-Gaudi Inference.', and 'abhay.saini saved Version 1 of dev-Gaudi Inference.', all dated 09:59. A 'Project was created' entry is also present. To the right of the activity feed is a 'Running Now' section with a table showing the title 'inference-1021' and status 'Running Now'. Further right is a 'Default Project Settings' section with fields for 'Compute' (medium) and 'Image' (cmrgv5.0), along with a 'Link to git repository' button. A red box highlights a 'NEW WORKSPACE' button in the top right corner.

Experiments	Endpoints	Workspaces	Apps
0	1	0	0

Activity Feed
abhay.saini created endpoint inference-1021.
abhay.saini ran Version 1 of dev-Gaudi Inference.
abhay.saini created flow dev-Gaudi Inference.
abhay.saini saved Version 1 of dev-Gaudi Inference.
Project was created

Running Now
inference-1021

Default Project Settings
Compute - medium
Image - cmrgv5.0
Link to git repository

Make sure to choose the relevant **compute** and the right container **image**.



Change type ▾

New Jupyter Workspace

Title

myworkspace

Compute



1 HPU.Gaudi Small ×



Datasets



Image

gaudi-docker/1.10.0/amzn2/habanalabs/pytorch-installer-2.0.1:latest



harindercnvrng/llm

Docker Hub



gaudi-docker/1.10.0/ubuntu20.04/habanalabs/pytorch-installer-2.0.1

Habana



ui/native/gaudi-docker/1.10.0/ubuntu20.04/habanalabs/pytorch-ins...

Habana



ui/native/gaudi-docker/1.10.0/ubuntu20.04/habanalabs/pytorch-ins...

Habana



gaudi-docker/1.10.0/ubuntu20.04/habanalabs/tensorflow-installer-t...

Habana



gaudi-docker/1.10.0/amzn2/habanalabs/pytorch-installer-2.0.1

Habana



latest

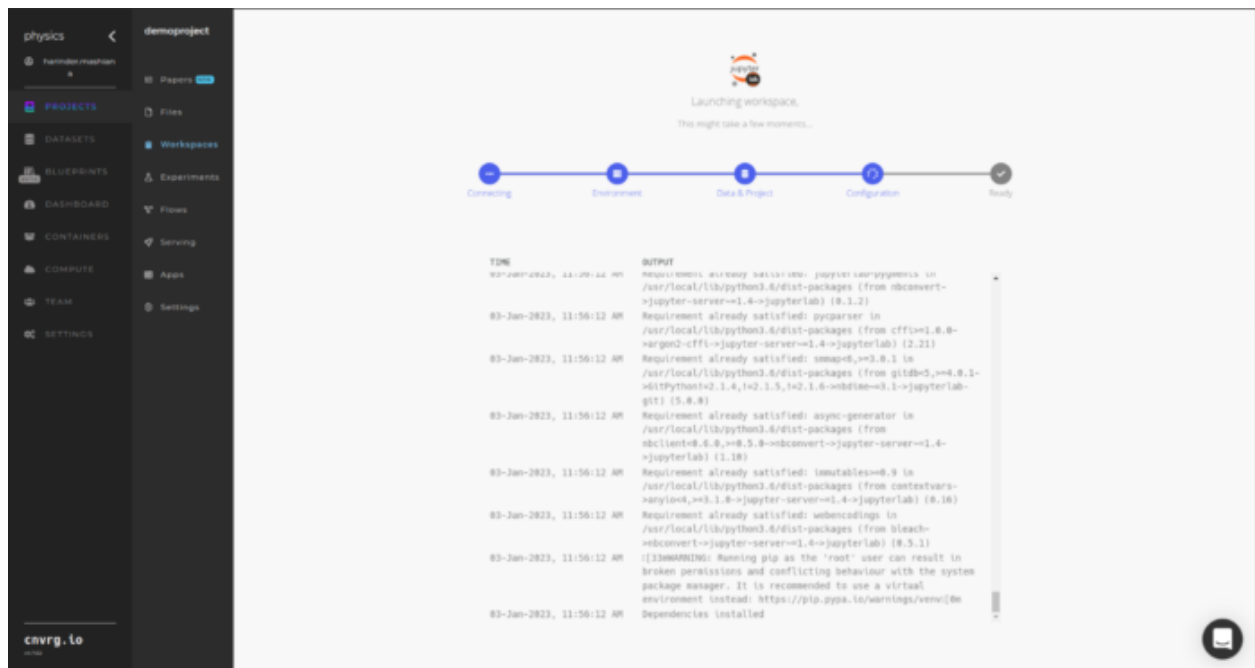


gaudi-docker/1.10.0/ubuntu22.04/habanalabs/pytorch-installer-2.0.1

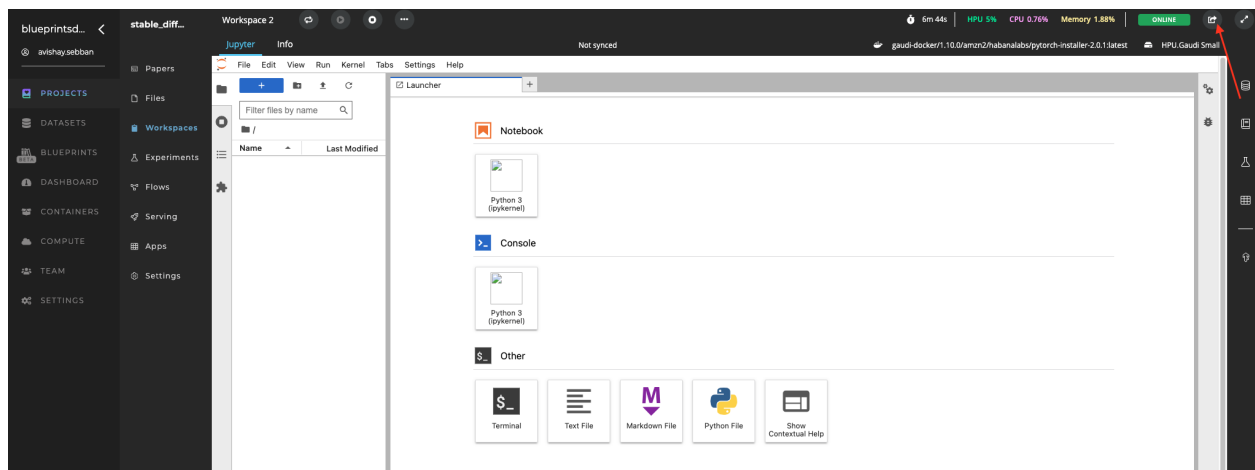
Habana



9. Next, the workspace will start loading and you will see a screen like this:



Once the workspace is up please click on “**open in a new tab**” icon



10. Now click on the **terminal** option to start a terminal so you can clone the git repository for stable diffusion.

The first thing you will want to do when running your script is to verify that your model actually runs on the Gaudi accelerator.

The Gaudi runtime environment includes the [hl-smi](#) tool which reports resource utilization of the Gaudi cores.

Run the following commands

> hl-smi

This is how it looks when you allocate 1 HPU

```
sh-4.2# hl-smi
```

HL-SMI Version:		hl-1.10.0-fw-43.2.0.0					
Driver Version:		1.10.0-416d95e					
AIP	Name	Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC
Fan	Temp	Perf	Pwr:Usage/Cap	Memory-Usage	AIP-Util	Compute	M.
0	HL-205	N/A	0000:90:1e.0	N/A			0
N/A	56C	N/A	109W / 350W	512MiB / 32768MiB	5%		N/A
Compute Processes:							AIP Memory
AIP	PID	Type	Process name				Usage
0	N/A	N/A	N/A				N/A

And this is how it looks when you allocate 2 HPU

```
gaudi-docker/1.10.0/amzn2/habanalabs/pytorch-installer-2.0.1:latest HPU:Medium
```

```
sh-4.2# hl-smi
```

HL-SMI Version:		hl-1.10.0-fw-43.2.0.0					
Driver Version:		1.10.0-416d95e					
AIP	Name	Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC
Fan	Temp	Perf	Pwr:Usage/Cap	Memory-Usage	AIP-Util	Compute	M.
0	HL-205	N/A	0000:10:1d.0	N/A			0
N/A	52C	N/A	102W / 350W	512MiB / 32768MiB	1%		N/A
1	HL-205	N/A	0000:a0:1d.0	N/A			0
N/A	58C	N/A	108W / 350W	512MiB / 32768MiB	4%		N/A
Compute Processes:							AIP Memory
AIP	PID	Type	Process name				Usage
0	N/A	N/A	N/A				N/A
1	N/A	N/A	N/A				N/A

For the subsequent steps, please refer to the detailed instructions provided in the following [link](#).

