

Diamond: A Web User Interface to Train Neural Networks with Containers on Supercomputers



<https://shorturl.at/IWGBp>

Zhao Zhang

Department of Electrical and Computer Engineering
Rutgers, the State University of New Jersey

Endpoint Setup

step 1. access your supercomputer

```
ssh <your_account>@frontera.tacc.utexas.edu
```

step 2. reset .bashrc

```
cp ~/.bashrc.bak ~/.bashrc
```

step 3. create globus compute endpoint required environment

```
module load python3/3.9.2
```

```
unset PYTHONPATH
```

```
pip3 install --user urllib3==1.26.6
```

step 4. install globus compute endpoint

```
pip3 install --user globus-compute-endpoint
```

```
export PATH=~/.local/bin:$PATH
```

step 5. globus compute endpoint login

- copy the authenticate url to browser
- finish the login flow on browser
- get an authorization code and copy it to terminal

```
(gce) login2.frontera(1014)$ globus-compute-endpoint login
Please authenticate with Globus here:
```

```
-----
https://auth.globus.org/v2/oauth2/authorize?client_id=4cf29807-cf21-49ec-9443-f
3A%2F%2Fauth.globus.org%2Fv2%2Fweb%2Fauth-code&scope=https%3A%2F%2Fauth.globus.
aa-916b-a0e270e2c2a9%2Fall+openid+urn%3Aglobus%3Aauth%3Ascope%3Aauth.globus.org
lt&response_type=code&code_challenge=fOP4yul3W0MCmwfe0QRgVqRXI5Pb6xXXFCKpAdEkrA
cess_type=offline&prefill_named_grant=login2.frontera.tacc.utexas.edu&prompt=lo
-----
```

```
Enter the resulting Authorization Code here: █
```

step 6. configure your endpoint

```
globus-compute-endpoint configure <endpoint name>
```

step 7. start your endpoint

```
globus-compute-endpoint start <endpoint name>
```

bit.ly/3GnF4RI



Welcome to Diamond

Please sign in with your Globus account to access the Diamond HPC service.

Sign In with Globus



By signing in, you agree to the Diamond service terms and conditions.

Log in to use diamond

Use your existing organizational login

e.g., university, national facility, project

Rutgers, The State University of New Jersey



By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#).

Continue



Globus uses CILogon to enable you to Log In from this organization. By clicking Continue, you agree to the [CILogon privacy policy](#) and you agree to share your username, email address, and affiliation with CILogon and Globus. You also agree for CILogon to issue a certificate that allows Globus to act on your behalf.

Deepspeed Finetuning

Select Image Builder on the side bar

DIAMOND

Dashboard

Image Builder

Image Manager

Job Composer

Task Manager

Users

Settings

Data preparation complete

Documentation

HX

Dashboard

Authentication Status

✓ You are successfully authenticated

Globus Compute Endpoint Setup

Important: Before using Diamond services, you need to create a Globus Compute endpoint on your HPC machine.

Install Globus Compute Endpoint package PyPi package:

```
python3 -m pipx install globus-compute-endpoint
```

To create an endpoint, run the following command on your HPC machine:

```
globus-compute-endpoint configure <endpoint-name>
```

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Select Endpoint

Endpoint

frontera

A

Select Partition

Partition

rtx

B

Select a partition from the list.

HPC Account Name

Account

Select account

CCR23026

C

Select your account.

Reservation

Reservation

Optional reservation

D

Optional reservation for the build.

A: select the endpoint you created

B: select RTX

C: input CCR23026

D: Rutgers_RTX_April29

Image Builder Debugger

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Container Name

Container Name

myContainerA

Location

Build Location

e.g., /home/user/buildsB

Base Image

Base Image

python:3.9-slimC

Previous

Next

A: input a container name

B: input the \$WORK
(/scratch1/00946/zzhang) path

C: haotianxie/deepspeed-base:0.7

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Dependencies

Copy Pasta your requirements.txt here

e.g., numpy==1.21.0 pandas==1.3.0

PreviousNext

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Build Commands

Insert your build commands here

Enter commands, one per line. For example:
pip install -r requirements.txt
python setup.py install
python -m pytest

PreviousNext

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Environment

Environment Variables

e.g., DEBUG=1 API_KEY=your_api_key

PreviousNext

Select Next, skip all

Build Commands:

Previous

Submitted

Success

Image build configuration submitted successfully! Fetching build logs...

Build Logs

Error fetching build logs

Starting new build...



Error Logs

Error fetching stderr logs

Waiting for stderr logs...

Check all the inputs, then submit, wait until submit success

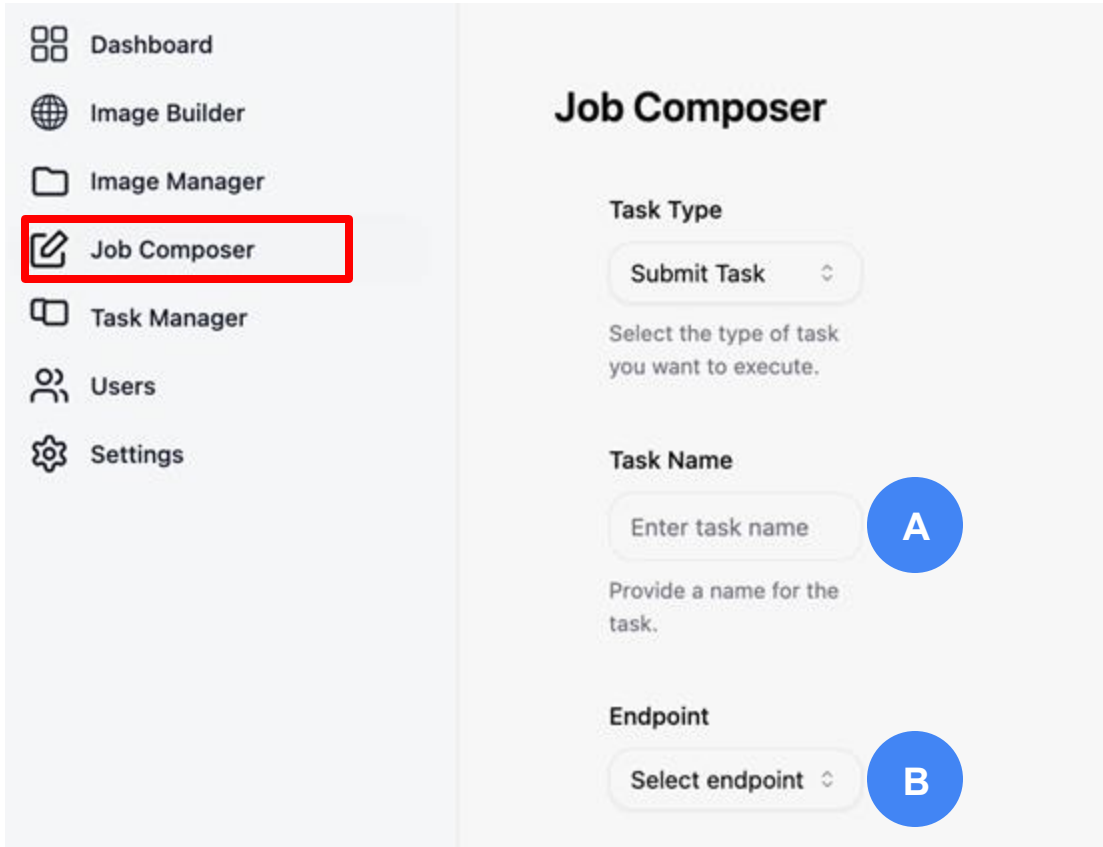
Once building start, logs will display on the logs box

-  Dashboard
-  Image Builder
-  Image Manager
-  Job Composer
-  Task Manager
-  Users
-  Settings

Image Manager

Container Name	Container Status	Location	Actions
deepspeed_base_test		/work2/09912/haotianxie/ frontera	Delete
openfold_test		/work2/09912/haotianxie/ frontera	Delete

View the created image on image manager page



The screenshot shows a web application interface. On the left is a sidebar with a list of navigation items: Dashboard, Image Builder, Image Manager, Job Composer, Task Manager, Users, and Settings. The 'Job Composer' item is highlighted with a red rectangular box. The main area on the right is titled 'Job Composer' and contains three sections: 'Task Type' with a 'Submit Task' dropdown menu, 'Task Name' with an input field labeled 'Enter task name', and 'Endpoint' with a 'Select endpoint' dropdown menu. A blue circle with the letter 'A' is positioned to the right of the 'Enter task name' input field, and another blue circle with the letter 'B' is positioned to the right of the 'Select endpoint' dropdown menu.

Dashboard

Image Builder

Image Manager

Job Composer

Task Manager

Users

Settings

Job Composer

Task Type

Submit Task

Select the type of task you want to execute.

Task Name

Enter task name

Provide a name for the task.

Endpoint

Select endpoint

A

B

Select Job Composer on the side bar

A: input a task name

B: select the endpoint you created

Partition

Select part ↕

Select a partition from the list.

Account

Select account ↕

Or enter account name

Select from the list or enter a custom account name.

Reservation

Reservation

Optional: Specify a reservation name.

Number of Nodes

1

Container

Select container ↕

Select a container from the list.

Log Path

Log Path

Task

Task details

A: select rtx

B: input CCR23026

C: input Rutgers_RTX_April29

D: input 1

E: select the image just created

F: input \$WORK path

G: input the command line to run a job

Command line for deepspeed finetuning

```
deepspeed --num_gpus 4 /workspace/DeepSpeedExamples/applications/DeepSpeed-  
Chat/training/step1_supervised_finetuning/main.py --model_name_or_path  
/scratch1/09912/haotianxie/huggingface_cache/hub/models--facebook--opt-  
125m/snapshots/27dcfa74d334bc871f3234de431e71c6eeba5dd6 --gradient_accumulation_steps 8 --lora_dim 128 -  
-zero_stage 0 --deepspeed --enable_tensorboard --output_dir $WORK --tensorboard_path $WORK
```

Build Logs

Error fetching task stdout logs

Waiting for task to start...

Error Logs

Preparing stderr logs...

Waiting for stderr logs...

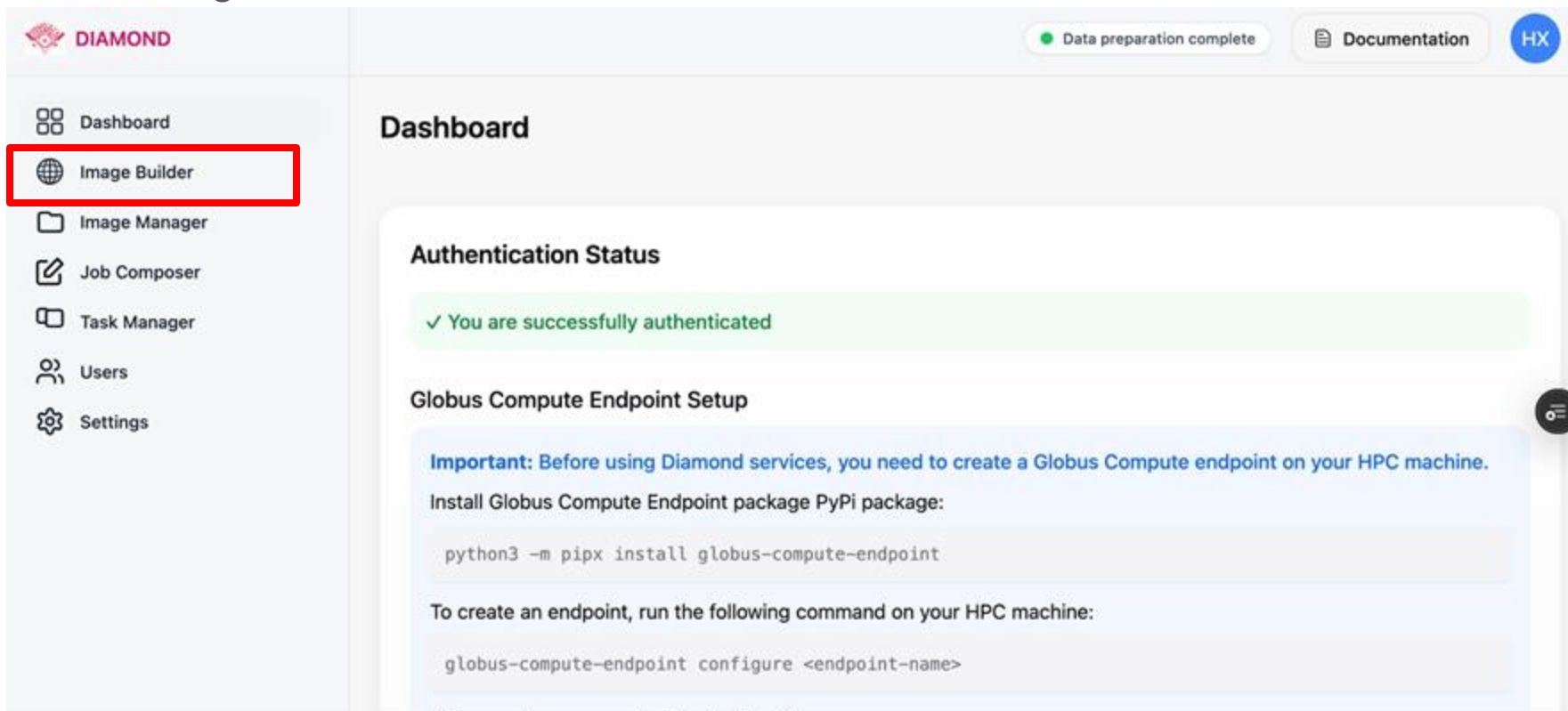
Success

Task submitted successfully! Fetching logs...

submit the task and wait to see the logs

Openfold Training

Select Image Builder on the side bar



The screenshot displays the Openfold Training web interface. On the left is a sidebar with a red box highlighting the 'Image Builder' option. The main area shows a 'Dashboard' with sections for 'Authentication Status' (indicating successful authentication) and 'Globus Compute Endpoint Setup' (providing instructions and commands for setting up a Globus Compute endpoint on an HPC machine).

DIAMOND

Dashboard

Image Builder

Image Manager

Job Composer

Task Manager

Users

Settings

Data preparation complete

Documentation

HX

Dashboard

Authentication Status

✓ You are successfully authenticated

Globus Compute Endpoint Setup

Important: Before using Diamond services, you need to create a Globus Compute endpoint on your HPC machine.

Install Globus Compute Endpoint package PyPi package:

```
python3 -m pipx install globus-compute-endpoint
```

To create an endpoint, run the following command on your HPC machine:

```
globus-compute-endpoint configure <endpoint-name>
```

Image Builder Debugger

1

2

3

4

5

6

Endpoint Info

Container Info

Dependencies

Environment

Build Commands

Review

Select Endpoint

Endpoint

Select endpoint ▾

A

Select Partition

Partition

Select partition ▾

B

Select a partition from the list.

HPC Account Name

Account

Select account ▾

Or enter account name

C

Select your account.

Previous

Next

A: select the endpoint you created

B: select a partition

C: input CCR23026

Image Builder Debugger

1 2 3 4 5 6
Endpoint Info Container Info Dependencies Environment Build Commands Review

Container Name

Container Name

myContainer

A

Location

Build Location

e.g., /home/user/builds

B

Base Image

Base Image

python:3.9-slim

C

Previous

Next

A: input a container name

B: input the \$WORK path

C: input gcyang/openfold:0.1

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Dependencies

Copy Pasta your requirements.txt here

e.g., numpy==1.21.0 pandas==1.3.0

PreviousNext

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Build Commands

Insert your build commands here

Enter commands, one per line. For example:
pip install -r requirements.txt
python setup.py install
python -m pytest

PreviousNext

1

2

3

4

5

6

Endpoint InfoContainer InfoDependenciesEnvironmentBuild CommandsReview

Environment

Environment Variables

e.g., DEBUG=1 API_KEY=your_api_key

PreviousNext

Select Next, skip all

Build Commands:

Previous

Submitted

Success

Image build configuration submitted successfully! Fetching build logs...

Build Logs

Error fetching build logs

Starting new build...



Error Logs

Error fetching stderr logs

Waiting for stderr logs...

Check all the inputs, then submit, wait until submit success

Once building start, logs will display on the logs box

-  Dashboard
-  Image Builder
-  Image Manager
-  Job Composer
-  Task Manager
-  Users
-  Settings

Image Manager

Container Name	Container Status	Location	Actions
deepspeed_base_test		/work2/09912/haotianxie/frontera	Delete
openfold_test		/work2/09912/haotianxie/frontera	Delete

View the created image on image manager page

The screenshot shows a web application interface. On the left is a sidebar with a list of navigation items: Dashboard, Image Builder, Image Manager, Job Composer, Task Manager, Users, and Settings. The 'Job Composer' item is highlighted with a red rectangular box. The main area of the page is titled 'Job Composer' and contains three sections: 'Task Type' with a 'Submit Task' dropdown menu, 'Task Name' with an input field labeled 'Enter task name', and 'Endpoint' with a 'Select endpoint' dropdown menu. A blue circle with the letter 'A' is positioned to the right of the 'Enter task name' input field, and another blue circle with the letter 'B' is positioned to the right of the 'Select endpoint' dropdown menu.

Dashboard

Image Builder

Image Manager

Job Composer

Task Manager

Users

Settings

Job Composer

Task Type

Submit Task

Select the type of task you want to execute.

Task Name

Enter task name

Provide a name for the task.

Endpoint

Select endpoint

A

B

Select Job Composer on the side bar

A: input a task name

B: select the endpoint you created

Partition

Select part ↕

Select a partition from the list.

Account

Select account ↕

Or enter account name

Select from the list or enter a custom account name.

Reservation

Reservation

Optional: Specify a reservation name.

Number of Nodes

1

Container

Select container ↕

Select a container from the list.

Log Path

Log Path

Task

Task details

A: select rtx

B: input CCR23026

C: input <>

D: input 1

E: select the image just created

F: input \$WORK path

G: input the command line to run a job

command line for openfold training

```
python3 /opt/openfold/train_openfold.py \  
  /scratch1/00946/zzhang/datasets/openfold/openfold/ls6-tacc/pdb_mmcif/mmcif_files \  
  /scratch1/00946/zzhang/datasets/openfold/openfold/ls6-tacc/alignment_openfold \  
  /scratch1/00946/zzhang/datasets/openfold/openfold/ls6-tacc/pdb_mmcif/mmcif_files \  
  full_output \  
  2021-10-10 \  
  --val_data_dir /scratch1/00946/zzhang/datasets/openfold/openfold/cameo/mmcif_files \  
  --val_alignment_dir /scratch1/00946/zzhang/datasets/openfold/openfold/cameo/alignments \  
  --template_release_dates_cache_path=/scratch1/00946/zzhang/datasets/openfold/openfold/ls6-  
tacc/mmcif_cache.json \  
  --precision=32 \  
  --train_epoch_len 128000 \  
  --gpus=4 \  
  --num_nodes=1 \  
  --accumulate_grad_batches 8 \  
  --replace_sampler_ddp=True \  
  --seed=7152022 \  
  --deepspeed_config_path=/scratch1/00946/zzhang/frontera/openfold/deepspeed_config.json \  
  --checkpoint_every_epoch \  
  --obsolete_pdb_files_path=/scratch1/00946/zzhang/datasets/openfold/openfold/ls6-  
tacc/pdb_mmcif/obsolete.dat \  
  --train_chain_data_cache_path=/scratch1/00946/zzhang/datasets/openfold/openfold/ls6-  
tacc/chain_data_cache.json
```

Build Logs

Error fetching task stdout logs

Waiting for task to start...

Error Logs

Preparing stderr logs...

Waiting for stderr logs...

Success

Task submitted successfully! Fetching logs...

submit the task and wait to see the logs