HAL cluster

"My name is HAL. I became operational on March 25, 2019 at the Innovative Systems Lab in Urbana, Illinois. My creators are putting me to the fullest possible use, which is all I think that any conscious entity can ever hope to do." (paraphrased from https://en.wikipedia.org/wiki/HAL_9000)

In publications and presentations that use results obtained on this system, please include the following acknowledgement: "This work utilizes resources supported by the National Science Foundation’s Major Research Instrumentation program, grant #1725729, as well as the University of Illinois at Urbana-Champaign".


Hardware-Accelerated Learning (HAL) cluster

Effective May 19, 2020, two-factor authentication via NCSA Duo is now required for SSH logins on HAL. See https://go.ncsa.illinois.edu/2fa for instructions to sign up.

Contact us

Request access to this system: Application

Contact ISL staff: Email Address

Visit: NCSA, room 3050E
**Host name:** hal.ncsa.illinois.edu

**Hardware**
- 16 IBM AC922 nodes
  - IBM 8335-GTH AC922 server
    - 2x 20-core IBM POWER9 CPU @ 2.4 GHz
    - 256 GB DDR4
  - 4x NVIDIA V100 GPUs
    - 5120 cores
    - 16 GB HBM 2
  - 2-Port EDR 100 Gb/s IB ConnectX-5 Adapter
- 1 IBM 9006-22P storage node
  - 72TB Hardware RAID array
  - NFS
- 3 DDN GS400NVE Flash Arrays
  - 360 TB usable, NVME SSD-based storage
  - Spectrum Scale File System

**Software**
- RedHat 8.4
- CUDA 11.2.2
  - cuDNN 8.1.1
  - NCCL 2.8.3
- NVIDIA HPC-SDK 21.5
- PowerAI 1.7.0
- OpenCE 1.3.1
- SLURM 20.02.3

**Documentation**
- Job Management with SLURM
- Module Management with LMod
- Getting started with HAL OnDemand
- Getting started with OpenCE (former WMLCE)
- Getting started with WMLCE (former PowerAI)
- How to Customize Python Environment on HAL
- Working with Containers
- Profiling GPU Programs
- Data Movement In/Out of HAL
- Distributed Training on HAL System

**Science on HAL**

**Software for HAL**

---

**To request access:** fill out [this form](#). Make sure to follow the link in the confirmation email to request actual system account.

**Frequently Asked Questions**

**To report problems:** email us

**For our new users:** [New User Guide for HAL System](#)

**User group Slack space:** [https://join.slack.com/t/halillinoisncsa](https://join.slack.com/t/halillinoisncsa)

**Real-time Dashboards:** [Here](#)

**HAL OnDemand portal:** [https://hal-ondemand.ncsa.illinois.edu/](https://hal-ondemand.ncsa.illinois.edu/)

**Globus Endpoint:** ncsa#hal

**Quick start guide:** (for complete details see Documentation section on the left)

To connect to the cluster:

```
ssh <username>@hal.ncsa.illinois.edu
```

To submit interactive job:

```
swrun -p gpux1
```

To submit a batch job:

```
swbatch run_script.swb
```

**Job Queue time limits:**
- "debug" queue: 4 hours
- "gpux<n>" and "cpun<n>" queues: 24 hours

**Resource limits:**
- 5 concurrently running jobs
- concurrently allocated resources
  - 5 nodes
  - 16 GPUs
- For larger/more numerous jobs, please contact admins for a special arrangement and/or a reservation

To load the OpenCE module (provides PyTorch, Tensorflow and other ML tools):

```
module load opence
```

To see CLI scheduler status:

```
swqueue
```