### 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](https://en.wikipedia.org/wiki/HAL_9000))

### Hardware-Accelerated Learning (HAL) cluster

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Software</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Host name:</strong> hal.ncsa.illinois.edu</td>
<td>RHEL 7.6</td>
<td>Job management with SLURM</td>
</tr>
<tr>
<td><strong>16 IBM AC922 nodes</strong></td>
<td><strong>CUDA</strong> 10.1.105</td>
<td>Modules management</td>
</tr>
<tr>
<td>- IBM 8335-GTH AC922 server</td>
<td>- cuDNN 7.5.0</td>
<td>Getting started with WMLCE (former PowerAI)</td>
</tr>
<tr>
<td>- 2x 20-core IBM POWER9 CPU @ 2.4GHz</td>
<td>- NCCL 2.4.2</td>
<td>Using Jupyter Notebook on HAL</td>
</tr>
<tr>
<td>- 256 GB DDR4</td>
<td><strong>IBM XLC and IBM XLFORTRAN</strong> 16.1.1</td>
<td>Working with containers</td>
</tr>
<tr>
<td>- 4x NVIDIA V100 GPUs</td>
<td>- PGI Community Edition 19.4</td>
<td></td>
</tr>
<tr>
<td>- 5120 cores</td>
<td>- PowerAI 1.6.0</td>
<td></td>
</tr>
<tr>
<td>- 16 GB HBM 2</td>
<td>- SLURM</td>
<td></td>
</tr>
<tr>
<td>- 2-Port EDR 100 Gb/s IB ConnectX-5 Adapter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 1 IBM 9006-22P storage node</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 72TB Hardware RAID array, NFS-mounted on all nodes via IB EDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Storage upgrade TBD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

To report problems: email us.

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

Real-time system status: [https://hal-monitor.ncsa.illinois.edu:3000/](https://hal-monitor.ncsa.illinois.edu:3000/)

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:

```
srun -p gpux1
```

or

```
srun --partition=gpux1 --pty --nodes=1 --ntasks-per-node=12 - --cores-per-socket=3 --threads-per-core=4 --sockets-per-node=1 --gres=gpu:v100:1 --mem-per-cpu=1500 --time=2:00:00 --wait=0 --export=ALL /bin/bash
```

To submit a batch job:

```
swbatch run_script.swb
```

or

```
sbatch run_script.sb
```

See run_script.swb and run_script.sb for a basic example.

Job Queue time limits:

- "debug" queue: 4 hours
- "gpux<n>" and "cpun<n>" queues: 72 hours

To load IBM Watson Machine Learning Community Edition (former IBM PowerAI) module:

```
module load wmlce
```