The OSU College of Engineering DGX System for Advanced GPU Computing

Mike Bailey
mjb@cs.oregonstate.edu

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
OSU’s College of Engineering bought six Nvidia DGX-2 systems

Each DGX server:
• Has 16 NVidia Tesla V100 GPUs
• Has 28TB of disk, all SSD
• Has two 24-core Intel Xeon 8168 Platinum 2.7GHz CPUs
• Has 1.5TB of DDR4-2666 System Memory
• Runs the CentOS 7 Linux operating system

Overall compute power:
• Each V100 NVidia Tesla card has 5,120 CUDA Cores and 640 Tensor Cores
• This gives each 16-V100 DGX server a total of 81,920 CUDA cores and 10,240 Tensor cores
• This gives the entire 6-DGX package a total of 491,520 CUDA Cores and 61,440 Tensor Cores
Performance Comparison with one of our previous Systems

DGX2 vs. Rabbit for Monte Carlo Calculations

- DGX-2’s Tesla
- Rabbit’s Titan Black

# of Monte Carlo Trials vs. MegaMonteCarloTrials/Second
How to SSH to the DGX Systems and Check on Them

ssh submit-c.hpc.engr.oregonstate.edu

**submit-c 142%** module load slurm

Do this to set your path correctly

- ssh over to a DGX submission machine -- submit-a and submit-b will also work

**submit-c 143%** sinfo

<table>
<thead>
<tr>
<th>PARTITION</th>
<th>AVAIL</th>
<th>TIMELIMIT</th>
<th>NODES</th>
<th>STATE</th>
<th>NODELIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>share*</td>
<td>up 7-00:00:00</td>
<td>4</td>
<td>idle compute-0-[4-7]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgx2</td>
<td>up 7-00:00:00</td>
<td>2</td>
<td>down* compute-dgx2-[4-5]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgx2</td>
<td>up 7-00:00:00</td>
<td>1</td>
<td>drng compute-dgx2-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgx2</td>
<td>up 7-00:00:00</td>
<td>1</td>
<td>mix compute-dgx2-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgx2</td>
<td>up 7-00:00:00</td>
<td>1</td>
<td>idle compute-dgx2-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gpu</td>
<td>up 7-00:00:00</td>
<td>1</td>
<td>idle compute-gpu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cbee</td>
<td>up 7-00:00:00</td>
<td>2</td>
<td>idle compute-b-2d,compute-i-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>student</td>
<td>up 7-00:00:00</td>
<td>4</td>
<td>idle compute-dgx2-1,compute-dgxs-[1-3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dgx</td>
<td>up 7-00:00:00</td>
<td>1</td>
<td>idle compute-dgxs-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eecsc</td>
<td>up 7-00:00:00</td>
<td>6</td>
<td>idle compute-1-[8-13]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**submit-c 144%** squeue

<table>
<thead>
<tr>
<th>JOBID</th>
<th>PARTITION</th>
<th>NAME</th>
<th>USER</th>
<th>ST</th>
<th>TIME</th>
<th>NODES</th>
<th>NODELIST(REASON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>dgx2</td>
<td>zsh yangyil</td>
<td>R 2-20:12:52</td>
<td>1</td>
<td>compute-dgx2-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>417</td>
<td>dgx2</td>
<td>zsh nguyehu5</td>
<td>R 1-10:04:09</td>
<td>1</td>
<td>compute-dgx2-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>423</td>
<td>dgx2</td>
<td>bash nguyenkh</td>
<td>R 1-06:31:52</td>
<td>1</td>
<td>compute-dgx2-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>424</td>
<td>dgx2</td>
<td>bash nguyenkh</td>
<td>R 1-06:26:56</td>
<td>1</td>
<td>compute-dgx2-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>426</td>
<td>dgx2</td>
<td>mlnt lehoang</td>
<td>R 22:00:03</td>
<td>1</td>
<td>compute-dgx2-2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

System Information

Check on the queues
Submitting a job to the DGX Systems using Slurm

Create a shell file

submit.bash:

```
#!/bin/bash
#SBATCH -J ArrayMult
#SBATCH -A cs475-575
#SBATCH -p class
#SBATCH --gres=gpu:1
#SBATCH -o arraymul.out
#SBATCH -e arraymul.err
#SBATCH --mail-type=BEGIN,END,FAIL
#SBATCH --mail-user=jparallel@oregonstate.edu
./arrayMul
```

Note: A single dash (-) is used for a single character flag
A double dash (--) is used for a word (more than a single character) flag

This is the queue name we use for classes

Double dash

Submit the job described in your shell file

submit-c 143% sbatch submit.bash
Submitted batch job 474
Submitting a job to the DGX Systems using Slurm

submit-c 144% **srun** –A cs475-575 –p student –gres=gpu:1 --pty bash
bash-4.2$ **/arrayMul**

Open an interactive shell on the DGX

Double dash