The OSU College of Engineering DGX System for Advanced GPU Computing

Mike Bailey
mjb@cs.oregonstate.edu

OSU’s College of Engineering has 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

BTW, you can also use the rabbit machine:
ssh rabbit.engr.oregonstate.edu
It is a good place to write your code and get it to compile.
It is not a good place to do the final run of your code.

How to SSH to the DGX Systems

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

flip3 151%  ssh submit-c.hpc.engr.oregonstate.edu

submit-c 142%  module load slurm
Type this right away to set your path correctly
### How to Check on the DGX Systems

#### Submitting a CUDA job to the DGX Systems using Slurm

```bash
#!/bin/bash
#SBATCH  -J  MatrixMult
#SBATCH  -A  cs475-575
#SBATCH  -p  class
#SBATCH  --gres=gpu:1
#SBATCH  -o  matrixmul.out
#SBATCH  -e  matrixmul.err
#SBATCH  --mail-type=BEGIN,END,FAIL
#SBATCH  --mail-user=joeparallel@oregonstate.edu
/usr/local/apps/cuda/cuda-10.1/bin/nvcc -o matrixMul matrixMul.cu
./matrixMul
```

#### System Information

<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</td>
<td>7-00:00:00</td>
<td>2</td>
<td>drain</td>
<td>compute-4-[3-4]</td>
</tr>
<tr>
<td>share*</td>
<td>up</td>
<td>7-00:00:00</td>
<td>1</td>
<td>mix</td>
<td>compute-2-6</td>
</tr>
<tr>
<td>sharegpu</td>
<td>up</td>
<td>7-00:00:00</td>
<td>1</td>
<td>mix</td>
<td>compute-dgxs-1, compute-gpu</td>
</tr>
<tr>
<td>dpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>1</td>
<td>drain</td>
<td>compute-dgxs-2-2</td>
</tr>
<tr>
<td>dpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>5</td>
<td>mix</td>
<td>compute-dgxs-1, compute-gpu</td>
</tr>
<tr>
<td>class</td>
<td>up</td>
<td>1-00:00:00</td>
<td>4</td>
<td>mix</td>
<td>compute-dgxs-1</td>
</tr>
<tr>
<td>class</td>
<td>up</td>
<td>1-00:00:00</td>
<td>2</td>
<td>idle</td>
<td>compute-dgxs-1, [1-2-3]</td>
</tr>
<tr>
<td>dpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>1</td>
<td>mix</td>
<td>compute-gpu</td>
</tr>
<tr>
<td>gpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>1</td>
<td>down</td>
<td>compute-gpu</td>
</tr>
<tr>
<td>dpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>3</td>
<td>mix</td>
<td>compute-gpu</td>
</tr>
<tr>
<td>dpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>1</td>
<td>mix</td>
<td>compute-gpu</td>
</tr>
<tr>
<td>dpx2</td>
<td>up</td>
<td>7-00:00:00</td>
<td>2</td>
<td>idle</td>
<td>compute-gpu</td>
</tr>
</tbody>
</table>
```

#### Check on the queues

```bash
submit-c 143% 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>3923</td>
<td>mime4</td>
<td>jayasurw</td>
<td>R</td>
<td>1-10:32:19</td>
<td>1 compute-e-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3925</td>
<td>mime4</td>
<td>20me</td>
<td>R</td>
<td>16:21:03</td>
<td>1 compute-e-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3876</td>
<td>share</td>
<td>chukwuk</td>
<td>R</td>
<td>1-21:36:45</td>
<td>1 compute-2-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3881</td>
<td>dpx2</td>
<td>halle</td>
<td>R</td>
<td>8:59:45</td>
<td>1 compute-b-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3904</td>
<td>dpx2</td>
<td>changyu</td>
<td>R</td>
<td>13:47:36</td>
<td>1 compute-dgxs-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3965</td>
<td>dpx2</td>
<td>mishrash</td>
<td>R</td>
<td>5-16:48:09</td>
<td>1 compute-dgxs-2-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3876</td>
<td>share</td>
<td>chukwuk</td>
<td>R</td>
<td>1-23:36:45</td>
<td>1 compute-2-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3971</td>
<td>nerhp</td>
<td>dionne</td>
<td>R</td>
<td>8:59:45</td>
<td>1 compute-h-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3881</td>
<td>dgx2</td>
<td>bash</td>
<td>R</td>
<td>1-22:50:44</td>
<td>1 compute-dgx2-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3965</td>
<td>dgx2</td>
<td>bash</td>
<td>chenju3</td>
<td>R</td>
<td>13:47:36</td>
<td>1 compute-dgxs-4</td>
<td></td>
</tr>
<tr>
<td>3965</td>
<td>dgx2</td>
<td>bash</td>
<td>chenju3</td>
<td>R</td>
<td>13:47:36</td>
<td>1 compute-dgxs-4</td>
<td></td>
</tr>
<tr>
<td>3645</td>
<td>dgx2</td>
<td>mishrash</td>
<td>R</td>
<td>5-16:48:09</td>
<td>1 compute-dgxs-2-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3585</td>
<td>dgx2</td>
<td>mishrash</td>
<td>R</td>
<td>5-16:48:09</td>
<td>1 compute-dgxs-2-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3583</td>
<td>dgx2</td>
<td>mishrash</td>
<td>R</td>
<td>6-17:34:00</td>
<td>1 compute-dgxs-2-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3583</td>
<td>dgx2</td>
<td>mishrash</td>
<td>R</td>
<td>6-18:26:44</td>
<td>1 compute-dgxs-2-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3645</td>
<td>dgx2</td>
<td>mishrash</td>
<td>R</td>
<td>5-16:48:09</td>
<td>1 compute-dgxs-2-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3585</td>
<td>dgx2</td>
<td>mishrash</td>
<td>R</td>
<td>6-17:34:00</td>
<td>1 compute-dgxs-2-3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Submitting a Loop

submitloop.bash:

```bash
#!/bin/bash
#SBATCH -J MatrixMul
#SBATCH -A cs475-575
#SBATCH -p class
#SBATCH --gres=gpu:1
#SBATCH -o matrixmul.out
#SBATCH --mail-type=BEGIN,END,FAIL
#SBATCH --mail-user=joeparallel@oregonstate.edu

for t in 1 2 4 8 16 32
do
    /usr/local/apps/cuda/cuda-10.1/bin/nvcc -DNUMT=$t -o matrixMul matrixMul.cu
    ./matrixMul
done
```

These 5 lines are actual bash code

Display: Displays the latest output added to matrixmul.err. Keeps doing it forever. Control-c to get out of it.

Results for Multiplying two 1024x1024 Matrices: Varying the CUDA Block Size

GigaFlops during Matrix Multiplication

NUMT (Each CUDA block was actually NUMTxNUMT threads)

Submit an OpenCL job to the DGX System using Slurm

submit.bash:

```bash
#!/bin/bash
#SBATCH -J MatrixMult
#SBATCH -A cs475-575
#SBATCH -p class
#SBATCH --gres=gpu:1
#SBATCH -o printinfo.out
#SBATCH --mail-type=BEGIN,END,FAIL
#SBATCH --mail-user=joeparallel@oregonstate.edu

g++ -o printinfo printinfo.cpp /usr/local/apps/cuda/cuda-10.1/lib64/libOpenCL.so.1.1 -lm -fopenmp
./printinfo
```
Here’s what `printinfo` got on one graphics card on the DGX System

Number of Platforms = 1
Platform #0:
   Name = 'NVIDIA CUDA'
   Vendor = 'NVIDIA Corporation'
   Version = OpenCL 1.2 CUDA 11.2.153'
   Profile = 'FULL_PROFILE'
Number of Devices = 1
Device #0:
   Type = 0x0004 = CL_DEVICE_TYPE_GPU
   Device Vendor ID = 0x10de (NVIDIA)
   Device Maximum Compute Units = 80
   Device Maximum Work Item Dimensions = 3
   Device Maximum Work Item Sizes = 1024 x 1024 x 64
   Device Maximum Work Group Size = 1024
   Device Maximum Clock Frequency = 1530 MHz
Device Extensions:
   cl_khr_global_int32_base_atomics
   cl_khr_global_int32_extended_atomics
   cl_khr_local_int32_base_atomics
   cl_khr_local_int32_extended_atomics
   cl_khr_fp64
   cl_khr_byte_addressable_store
   cl_khr_id
   cl_khr_gl_sharing
   cl_khr_device_attribute_query
   cl_khr pragma unroll
   cl_khr_resource_query
   cl_khr_copy_cofs
   cl_khr_create_buffer

For reference, rabbit’s graphics card has 15 Compute Units.