IBEX GPU Job Submission

Amr Radwan
HPC Applications Specialist
Supercomputing Core Laboratory

IBEX GPU Nodes Specs

Sl. No	GPU Architecture	Available Number Of Nodes	Available GPU Cards Per node	GPU Memory Per Card	Usable Node Memory
1	Kepler: K40m	3	8	12GB	246GB
2	Fermi: GTX1080TI	8	4	12GB	246GB
3	Fermi: GTX1080TI	4	8	12GB	366GB
4	Pascal: P100	6	4	16GB	246GB
5	Pascal: P6000	2	2	22GB	246GB
6	Volta: V100	8	4	32GB	366GB

https://www.hpc.kaust.edu.sa/ibex/gpu_nodes

IBEX GPU Nodes Specs

sinfo --partition=batch --format="%n %f" | fgrep gpu| grep -v nogpu |sort

```
sinfo --partition=batch --format="%n %f" | fgrep gpu| grep -v nogpu |sort
dgpu502-17-l ibex2017,nolmem,cpu_intel_e5_2670,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_tesla_k40m,tesla_k40m
dgpu502-17-r ibex2017,nolmem,cpu_intel_e5_2670,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_tesla_k40m,tesla_k40m
dgpu502-21-r ibex2017,nolmem,cpu_intel_e5_2670,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_tesla_k40m,tesla_k40m
dgpu502-25 ibex2017,nolmem,cpu_intel_e5_2699_v3,gpu,mpi_intel,intel_gpu,local_200G,local_400G,local_500G,gpu_p100,p100
dgpu502-29 ibex2017,nolmem,cpu_intel_e5_2699_v3,gpu,mpi_intel,intel_gpu,local_200G,gpu_p100,p100
dgpu502-33 ibex2017,nolmem,cpu_intel_e5_2699_v3,gpu,mpi_intel,intel_gpu,local_200G,gpu_p100,p100
dgpu502-37 ibex2017,nolmem,cpu_intel_e5_2699_v3,gpu,mpi_intel,intel_gpu,local_200G,gpu_p100,p100
gpu601-02 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-03 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-04 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-05 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-06 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-07 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-08 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
gpu601-09 dragon,ibex2018,nolmem,cpu_intel_gold_6142,intel,gpu,intel_gpu,local_200G,local_400G,local_500G,gpu_v100,v100
```

GPU Jobs Constraints

- glogin/vlogin: Compile GPU code
- sbatch from glogin runs on arbitrary nodes; not a GPU compute node

  [dgpu105-09:~]$ salloc -t 00:10:00

  salloc: Pending job allocation ...

  salloc: Granted job allocation …

  salloc: Nodes ds503-01 are ready ….

- **Cluster is very heterogeneous**

GPU Jobs Constraints

There are three basic ways to ask for GPUs.

- You want a specific count of a specific model of GPU
- You want a specific count of any type of GPU
- You want a specific count of some subset of available types (e.g. any with > 8 GB of memory)

https://www.hpc.kaust.edu.sa/ibex/slurm_constraints

GPU Jobs Constraints

Allocate specific model of GPU:

```
# Request 2 P100 GPUs.
[hanksj@dm511-17:~]$ srun --pty --time=1:00 --gres=gpu:p100:2 bash -l
[hanksj@dgpu703-01:~]$ nvidia-smi
Tue Jun 27 14:33:53 2017
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 375.26                 Driver Version: 375.26                    |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla P100-PCIE...  On   | 0000:02:00.0     Off |                    0 |
| N/A   31C    P0    25W / 250W |      0MiB / 16276MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla P100-PCIE...  On   | 0000:03:00.0     Off |                    0 |
...
```

https://www.hpc.kaust.edu.sa/ibex/slurm_constraints

GPU Jobs Constraints

Allocate any model of GPU:

```
# Request 1 GPU of any kind
[hanksj@dm511-17:~]$ srun --pty --time=1:00 --gres=gpu:1 bash -l
[hanksj@dgpu502-01-r:~]$ nvidia-smi
Tue Jun 27 14:37:29 2017
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 375.26                 Driver Version: 375.26                    |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla K20m          On   | 0000:09:00.0     Off |                    0 |
| N/A   27C    P8    19W / 225W |      0MiB /  4742MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
...
```

GPU Jobs Constraints

Allocate subset of available GPU types:

```
# Request 1 GPU of type p100 OR tesla_k40m OR p6000
[hanksj@dm511-23:~]$ srun --pty --gres=gpu:1 --constraint="[p100|tesla_k40m|p6000]" -
-time=1:00:00 bash -l
[hanksj@dgpu703-01:~]$ nvidia-smi
Thu Jun 29 07:58:06 2017
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 375.26                 Driver Version: 375.26                    |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla P100-PCIE...  On   | 0000:02:00.0     Off |                    0 |
| N/A   32C    P0    25W / 250W |      0MiB / 16276MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla P100-PCIE...  On   | 0000:03:00.0     Off |                    0 |
| N/A   33C    P0    25W / 250W |      0MiB / 16276MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
...
```

https://www.hpc.kaust.edu.sa/ibex/slurm_constraints

GPU Jobs Constraints

- Example to GPU batch job

```
#!/bin/bash
#SBATCH -N 1
#SBATCH --partition=batch
#SBATCH -J MyJob
#SBATCH -o MyJob.%J.out
#SBATCH -e MyJob.%J.err
#SBATCH --time=01:30:00
#SBATCH --mem=100G
#SBATCH --gres=gpu:2
#SBATCH --constraint=[p100]

#run the application:
Myapp
```

- You can create your job by using:
  https://www.hpc.kaust.edu.sa/ibex/ibex-job-generator

FYI - Cluster nodes are a shared resource

Don't ssh to nodes you haven't allocated!

- Cluster nodes are a shared resource

Tell Slurm what you need:

- `srun --pty --time=10 --gres=gpu bash -l`
- `man sbatch`
  - `--exclusive`
  - `--nodelist`
  - `--exclude`

Acknowledgments to Glendon R. Holst


جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Thank You