CSCS

Swiss National Supercomputing Centre

Scheduling GPUs With SLURM

Stephen Trofinoff—CSCS
HPC-CH
Basel, Switzerland
27-October-2011

Scheduling GPUs With SLURM

SLURM—a (relatively) “simple” open-source resource management system.

Three primary SLURM objectives:

1) Allocate exclusive/non-exclusive access to resources to users

2) Provide framework for starting, executing and monitoring of work on these allocations

3) Use queues to manage contention

Scheduling GPUs With SLURM

Modify **slurm.conf**:

Add entry for the gres type (e.g. GresType=gpu)
Add name of GPU family as a feature of Node
Add “Gres=gpu:[n]” where n is the # of GPUs
NodeName=compute22 Feature=”Fermi” Gres=gpu:1

Create **gres.conf:**

Name=gpu File=/dev/nvidia0
CPUs=... List of CPUs with GPU access (optional)

Scheduling GPUs With SLURM

User specifies the number of GPUs needed per node with “**--gres=gpu...**”

For example:

```
sbatch -N 2 -n 4 –gres=gpu
sbatch -N 2 -n 4 –gres=gpu:1
sbatch -N 2 -n 4 –gres=gpu:2
```

Scheduling GPUs With SLURM

Use “--constraint” to limit the type of GPU

```
sbatch -N 2 -n 4 –gres=gpu:1 –constraint=”Fermi”
sbatch -N 2 -n 4 –gres=gpu:1 –constraint=”Fermi|geforce”
```

Scheduling GPUs With SLURM

Can select nodes based upon GPU memory

Trickier than specifying the GPU family

Specify “gpu_mem” as an additional GRES

```
--gres=gpu,gpu_mem:2000
```

“2000” signifies we need AT LEAST 2000MB of GPU memory

Scheduling GPUs With SLURM

Configuring GPU memory:

1) Add a line to gres.conf such as

```
Name=gpu_mem Count=2048
```

For “gpu_mem” count is interpreted as # of MB

2) Append similar clause to NodeName line in slurm.conf

```
NodeName=... gres=gpu:1,gpu_mem:2048
```

3) Append “gpu_mem” to GresTypes line in slurm.conf

Scheduling GPUs With SLURM

Future work = add GPU accounting:

- Number of GPUs requested by job
- Number of GPUs allocated to a job

Will necessitate addition of several database fields

Will necessitate modification of sacct command and possible others

Accounting aides in determining whether the machine is being properly utilized

Scheduling GPUs With SLURM

Q & A