Lab 3 GPU Cluster

Parallel Programming 2025/10/17

Taiwania 2 Cluster

Manual: https://man.twcc.ai/@twccdocs/doc-twnia2-main-zh?type=view

Login node: In01.twcc.ai

- You can only submit job and compile program on login node
- Do NOT run program on it

GPU: NVIDIA V100 * 8 per node

You will use 1 GPU in this lab and hw3

Job submission using srun

```
srun [options] ./executable [args]
```

Options:

- -N NODES: the number of nodes to run the job
- -n PROCESSES: the number of total processes to launch
- -c CPUS: CPUS is the number of cpus available to each process
- -t TIME: The time limit in "minutes" or "minutes:seconds"
- -J NAME: The name of the job. Will be displayed on squeue
- -A PROJECT_ID: The ID of the project. In this course, the project ID is ACD114118

Job submission using srun

```
srun [options] ./executable [args]
```

Options:

--gpus-per-node GPUS: Number of GPUs per node

Example:

srun -N 1 -n 1 --gpus-per-node 1 -A ACD114118 -t 1 ./executable

No Mining

Educational use only

Compile & Run

- Load cuda module: module load cuda
- Compiler: nvcc (a gcc compiler wrapper with CUDA support)
 - o nvcc [options] <inputfile>
 - E.g., nvcc hello_world.cu -o hello_world
- Execute
 - \$ srun -N 1 -n 1 --gpus-per-node 1 -A ACD114118 -t 1 hello_world

Useful Commands

- nvidia-smi
 - Check for the available GPUs and utilization
- nvprof
 - Nvidia kernel profiler
- ncu
 - Nsight Compute. A newer version of the kernel profiler.
- cuda-memcheck
 - o cuda-memcheck <gpu_exe>
 - Reports illegal memory accesses. Used for debugging

NVIDIA-SMI

- NVIDIA System Management Interface program
- You can query details about
 - a. gpu type
 - b. gpu utilization
 - c. memory usage
 - d. temperature
 - e. clock rate
 - f. ...

NVIDIA-SMI 535.161.08 Driver					Version: 535.161.08 CUDA Version: 12.2		
	ame emp	Perf	Persiste Pwr:Usago				Uncorr. ECC Compute M. MIG M.
	esla 26C	V100-SXM2-32GE P0	43W /	0n 300W 	00000000:DC:00.0 Off 0MiB / 32768MiB	 0% 	Default N/A
Process GPU	ses: GI ID	CI PIC	Туре	Proces	s name		GPU Memory Usage

Profiler

- nvprof
 - Showing kernel execution time, memory usage, and API calls.
- Nsight Compute (ncu)
 - o kernel-level analysis
 - o memory throughput, occupancy, and warp efficiency.

cuda-memcheck

- This tool checks memory errors of your program, and it also reports hardware exceptions encountered by the GPU. These errors may not cause program to crash, but they could result in unexpected program behavior and memory misusage.
- Useful for debugging
- Error types
 - https://docs.nvidia.com/cuda/archive/11.0/cuda-memcheck/index.html

Lab 3 Assignment

Google form

Complete the google form before 2025/10/20 23:59

https://forms.gle/gNtQyJ5vkB88RbUd8