Cygnus supercomputer

Osamu Tatebe
tatebe@cs.tsukuba.ac.jp
Center for Computational Sciences,
University of Tsukuba

Cygnus supercomputer

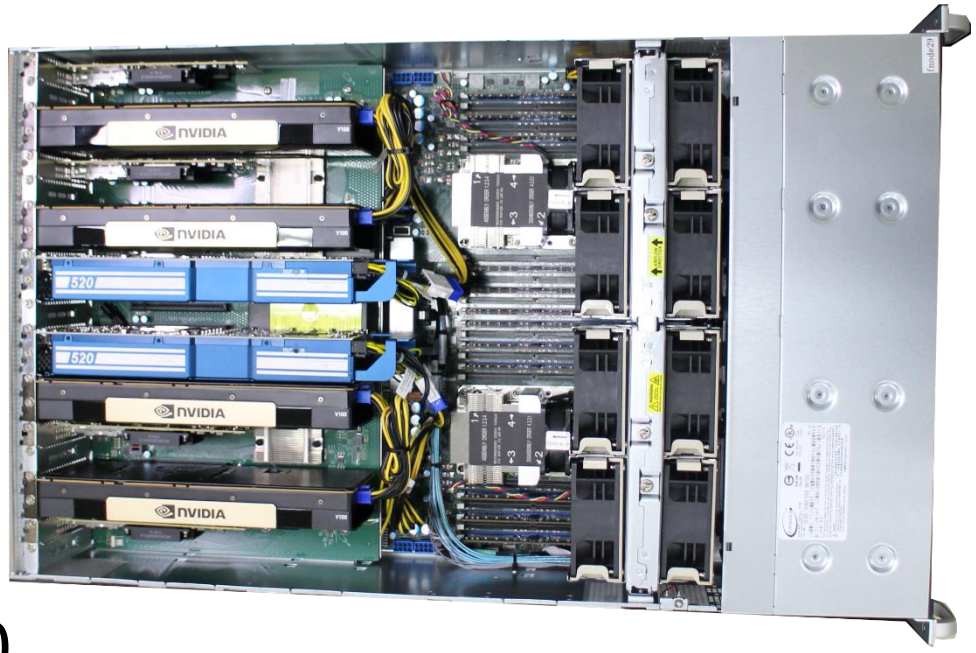
- Operational from April 2019
- 2.3 PFLOPS, 2.4 PB storage
- 78 nodes (32 nodes include FPGA cards)
  - 2 x 2.6GHz 12c Xeon Gold — 2 TFlops
  - 4 x Tesla V100 GPU (32GB HBM2) — 28 TFlops
  - (2 x Stratix10 FPGA)
  - 192GB DDR4 memory
  - 3.2TB NVMe SSD
- Interconnects 4 x InfiniBand HDR100 — 400 Gbps
  - Full-bisection network
  - (FPGA networks 400 Gbps)

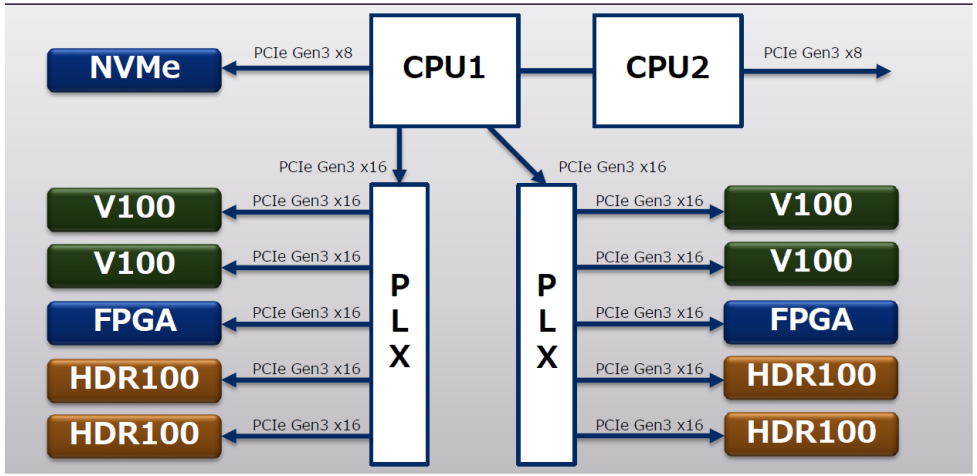
Cygnus supercomputer

Location: Center for Computational Sciences, Computer Building

Compute nodes

- 2 x 2.6GHz 12core Xeon Gold 6126
- 192GB DDR4
- 4 x NVIDIA Tesla V100 32GB HBM2
- 4 x InfiniBand HDR100
- 3.2TB NVMe SSD





System software

- OS: CentOS 7
- Compiler： Intel, PGI, GCC
- MPI： Open MPI, MVAPICH2-GDR, Intel MPI
- Library： Intel MKL, cuDNN, NCCL
- Job management： NEC NQSV

Storage

- Home (NFS)
  - /home/EDU1/<user_name>
  - Just for log in. Basically not much use.
- Work area (Lustre)
  - **/work/EDU1/<user_name>**
  - Files and executables required for parallel computation
- Scratch area (within a compute node)
  - /scr

Compilation of MPI programs

- OpenMPI

```
% module load openmpi/gdr/4.0.3/gcc8.3.1-cuda11.2.1      # GCC
% module load openmpi/gdr/4.0.3/intel19.1.3-cuda11.2.1   # Intel
% module load openmpi/gdr/4.0.3/pgi20.4-cuda11.0         # PGI
```

- Compiler drivers
  - mpicc, mpicxx, mpif77, mpif90
- Refer to module avail for MVAPICH GDR and Intel MPI
- User Guide
  - https://www2.ccs.tsukuba.ac.jp/ccswiki/cygnus/index.php?Cygnus%20User%20Guide
  - http://www2.ccs.tsukuba.ac.jp/ccswiki/cygnus

Available queue

- Batch queue
  - edu-1 maximum 1 hour for 1-node job
  - edu-2 maximum 30 minutes for 2-node job
  - edu-4 maximum 15 minutes for 4-node job
  - edu-8 maximum 7.5 minutes for 8-node job
- Interactive queue
  - debug maximum 1 hour for up to 2 nodes

Using multilanes

- Each compute node connected with 4 lanes of InfiniBand HDR100
- # lanes for each process can be specified by the environment variable
  - mpirun … -x UCX_MAX_RNDV_LANES=4 …
  - mpiexec … -genv MV2_NUM_HCAS 4 …
    # 4 lanes for each process
- For details, refer to Cygnus user guide

Batch script for Pure MPI

```
#!/bin/bash
#PBS -A EDU1
#PBS -q edu-2
#PBS -b 2
#PBS -l elapstim_req=00:30:00
#PBS -T openmpi
#PBS -v NQSV_MPI_VER=gdr/4.0.3/gcc8.3.1-cuda11.2.1

module load openmpi/$NQSV_MPI_VER
mpirun ${NQSII_MPIOPTS} -np 48 -npernode 24 -x
UCV_MAX_RNDV_LANES=4 ${PBS_O_WORKDIR}/prog
```

For MPI+OpenMP (1 process with 24 threads for each node)

```
#!/bin/bash
#PBS -A EDU1
#PBS -q edu-2
#PBS -b 2
#PBS -l elapstim_req=00:30:00
#PBS -T openmpi
#PBS -v NQSV_MPI_VER=gdr/4.0.3/gcc8.3.1-cuda11.2.1
#PBS -v OMP_NUM_THREADS=24

module load openmpi/$NQSV_MPI_VER
mpirun ${NQSII_MPIOPTS} -np 2 -npernode 1 --bind-to none -x
UCV_MAX_RNDV_LANES=4 ${PBS_O_WORKDIR}/prog
```

For MPI+OpenMP (2 processes with 12 threads for each node)

```
#!/bin/bash
#PBS -A EDU1
#PBS -q edu-2
#PBS -b 2
#PBS -l elapstim_req=00:30:00
#PBS -T openmpi
#PBS -v NQSV_MPI_VER=gdr/4.0.3/gcc8.3.1-cuda11.2.1
#PBS -v OMP_NUM_THREADS=12

module load openmpi/$NQSV_MPI_VER
mpirun ${NQSII_MPIOPTS} -np 4 -npernode 2 --bind-to socket -x
UCV_MAX_RNDV_LANES=2 ${PBS_O_WORKDIR}/prog
```

Job submission

qsub <script>	Submit a job
qstat	Check submitted jobs
sstat	Check execution start time
qdel <job ID>	Remove a job
qlogin -A EDU1 -q debug	Request an interactive job