



筑波大学計算科学研究センター CCS HPCサマーセミナー 「Cygnusの利用法」

建部修見

tatebe@cs.tsukuba.ac.jp

計算科学研究センター



Cygnusシステム

- 2019年4月より運用開始
- 筑波大学計算科学研究センターが管理・運用
- 2.3 PFLOPS、ストレージ2.4 PB
- 78ノード(内32ノードはFPGAノード)
 - 2 x 2.6GHz 12c Xeon Gold 2 TFlops
 - 4 x Tesla V100 GPU (32GB HBM2) 28 TFlops
 - (2 x Stratix10 FPGA)
 - 192 GB メモリ
 - 3.2 TB NVMe SSD
- ノード間ネットワーク 4 x IB HDR100 400 Gbps
 - フルバイセクション
 - (FPGA間ネットワーク 400 Gbps)



Cygnusシステム全体

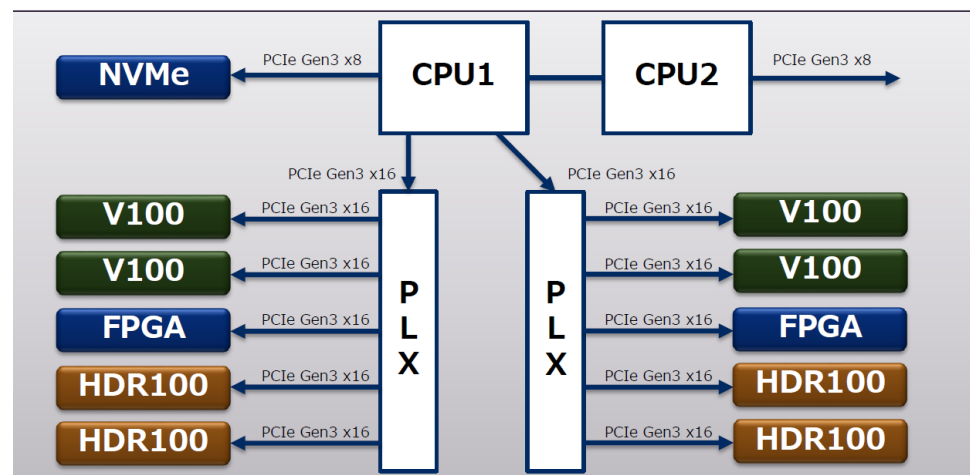
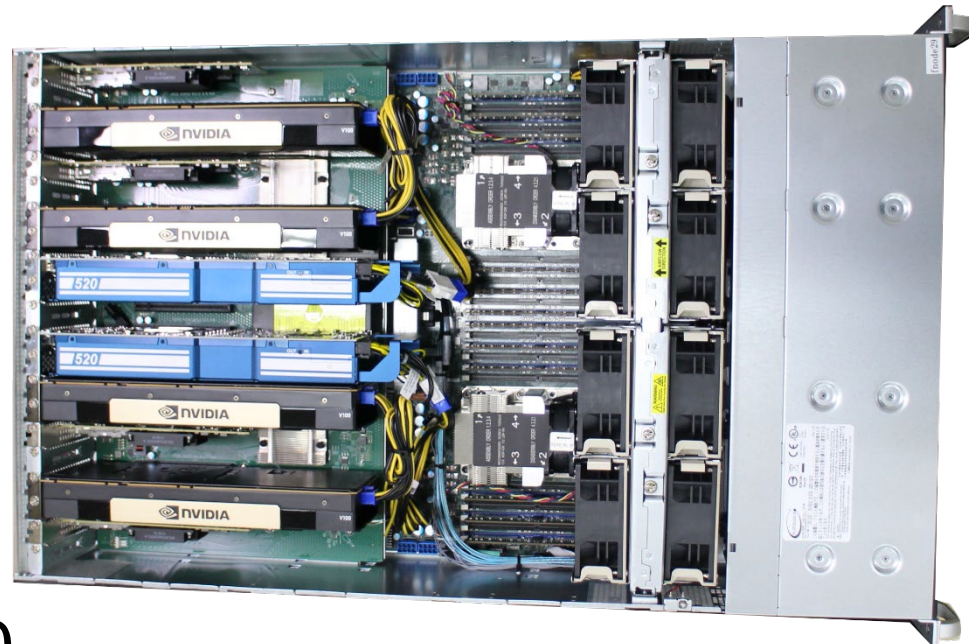


設置場所：筑波大学計算科学研究センター・計算機棟



計算ノード

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





Cygnusシステムのソフトウェア

- ソフトウェア環境
 - OS: CentOS 7
 - コンパイラ: Intel, PGI, GCC
 - MPI: Open MPI, MVAPICH2-GDR, Intel MPI
 - ライブラリ: Intel MKL, cuDNN, NCCL
 - ジョブ管理: NEC NQSV



ストレージ

- ホーム領域 (NFS)
 - /home/EDU2/<ユーザ名>
 - 基本あまり使わない
- ワーク領域 (Lustre)
 - **/work/EDU2/<ユーザ名>**
 - 計算に必要なファイル、実行ファイル
- スクラッチ領域 (計算ノード内)
 - /scr



MPIプログラムのコンパイル

- 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
```

- コンパイラドライバ

- mpicc, mpicxx, mpif77, mpif90

- MVAPICH GDR, Intel MPIはmodule availを参照

- 利用の手引き

- <http://www2.ccs.tsukuba.ac.jp/ccswiki/cygnus>



利用可能キュー

- バッチキュー
 - edu-1 1ノードジョブ最大1時間
 - edu-2 2ノードジョブ最大30分
 - edu-4 4ノードまでのジョブ最大15分
 - edu-8 8ノードまでのジョブ最大7.5分
- 会話キュー
 - debug 2ノードまで最大1時間



マルチレールの利用

- 各ノードはIB4本で接続
- プロセス当たり利用するレールの数は環境変数で指定
 - `mpirun ... -x UCX_MAX_RNDV_LANES=4 ...`
 - `mpiexec ... -genv MV2_NUM_HCAS 4 ...`
#1プロセスで4レール利用する場合
- 詳細はCygnusシステム利用ガイドを参照



Pure MPIのバッチスクリプト

```
#!/bin/bash
#PBS -A EDU2
#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 ${NQSV_MPI_OPTS} -np 48 -npernode 24 -x
UCV_MAX_RNDV_LANES=4 ${PBS_O_WORKDIR}/prog
```



MPI+OpenMP (各ノード1プロセス24スレッド)

```
#!/bin/bash
#PBS -A EDU2
#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 ${NQSV_MPI_OPTS} -np 2 -npernode 1 --bind-to none -x
UCV_MAX_RNDV_LANES=4 ${PBS_O_WORKDIR}/prog
```



MPI+OpenMP (各ノード2プロセス12スレッド)

```
#!/bin/bash
#PBS -A EDU2
#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 ${NQSV_MPI_OPTS} -np 4 -npernode 2 --bind-to socket -x
UCV_MAX_RNDV_LANES=2 ${PBS_O_WORKDIR}/prog
```



ジョブの投入

qsub スクリプト

ジョブの投入

qstat

投入したリクエストの確認

sstat

実行開始時間の確認

qdel ジョブID

ジョブの削除

qlogin -A EDU2 -q debug 会話リクエスト投入