

# Final Architecture Proposal for Peking University

## Supercomputing Team

### Hardware Configuration

According to coordination with the vendor partner, the final cluster configuration that the Peking University Supercomputing Team will bring to the SC19 Student Cluster Competition is as follows:

GPU Node (4 of these)

A GPU node features the following key hardware components:

- CPU: 2x Intel Xeon Gold 6240 (TDP 150W each)
- GPU: 3x NVIDIA Tesla V100 32GB (TDP 250W each)
- Chassis fans, miscellaneous components, CPU idle: 30W

*Note: The GPU will draw at most 120W via frequency limitations. Furthermore, the CPU nodes (described below) will be completely idle when running GPU tasks.*

CPU Node (1 of these)

A CPU node features the following key hardware components:

- CPU: 2x Intel Xeon Gold 6240 (TDP 150W each)
- Chassis fans, miscellaneous components, CPU idle: 30W

*Note: The GPUs in the GPU nodes will stay idle (~20W each) when the whole cluster is under full CPU load.*

Supplementary Devices

- InfiniBand EDR switch: 100W
- GigEthernet switch: 10W

Platform choice motivation:

The team has picked the above configuration due to general availability as products from our vendor. The Xeon Gold + NVIDIA Tesla platform features reasonable performance and cost-efficiency. The choice to bring more GPUs and limiting their power to below TDP is due to the experience that GPU's power efficiency tends to degrade as the running power approaches TDP.

Final power estimation:

GPU full load:  $(150*2+120*3+30)*4+30*1+100+10=2900W$

CPU full load:  $(150*2+20*3+30)*4+(150*2+30)*1+100+10=2000W$

## Software used

The (incomprehensive) list of software deployed on the competition cluster is as follows. The reason that each bundle has been chosen, if considered significant, follows.

- CentOS 7.7 x86\_64 with EPEL repositories
  - Freely available Linux distribution that features enterprise-level stability. Wide adoption in the HPC community. Excellent support from HPC-related software tools.
  - EPEL features extra packages commonly used by software development.
- NVIDIA CUDA 10.1 toolkit
  - A necessity to compile and run GPU programs.
- PGI Compilers Community Edition
  - OpenACC for fast assessment and porting of possible applications.
- Mellanox OFED with UCX and OpenMPI 4.0
  - Official driver + user libraries for Mellanox EDR IB cards. Highly optimized vendor-provided communication libraries.
- Intel Parallel Studio XE
  - Highly optimized code generation via ICC for Intel processors.
  - Fast math routines via MKL.
- OpenZFS + Lustre for storage
  - RAIDZ to provide redundancy (thus reliability of data) in case of power shutoff.
  - Supports RDMA networks for high-performance shared network filesystem.