TEXAS STATE UNIVERSITY SAN MARCOS

TEAM TEXAS

TEXAS The University of Texas at Austin

R. Ababao • J. Garcia • J. Granberry • C. Kim • J. Voss • J. Zhao

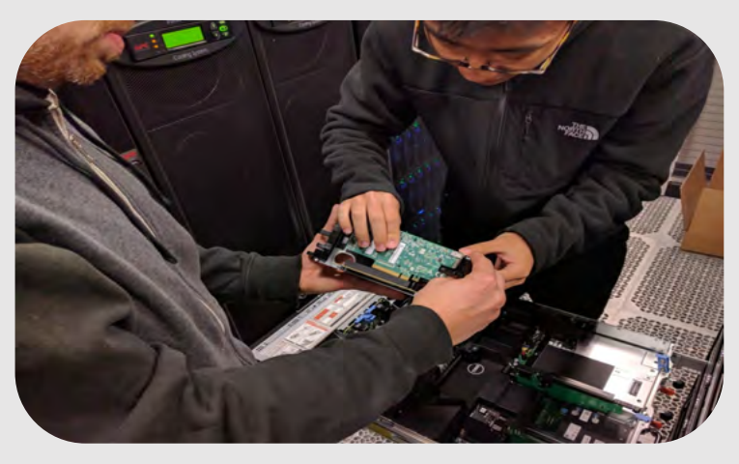
ABOUT THE TEAM

Our advisors contacted several professors/professional groups at both universities to talk to their classes/members and solicit students for the Student Cluster Team. These groups contained a variety of students with different majors and backgrounds. From these applicants, the final team was selected. Each member chose an application to specialize in to help with equal division of labor.

Rainier Ababao	: Computer Science and Biology	: Parconnect
Joe Garcia	: Computational Physics	: Password
Josh Granberry	: Computer Science	: HPCG
CJ Kim	: Electrical and Computer Engineering	: Password
Joseph Voss	: Mechanical Engineering (Mechatronics)	: HPL
Joe Zhao	: Mathematics (Statistics)	: ParaView

PREPARATION

- Met on a weekly basis to discuss and plan strategies to tackle the competition applications
- Used TACC supercomputer clusters to learn and familiarize ourselves with HPC tools including compilers and MPI stacks as well as HPC applications and scientific libraries
- Attended courses and seminars taught by TACC staff in topics such as scientific programming, parallel computing, and HPC Python
- Assembled a beta version of our cluster from scratch to build knowledge and test applications before final hardware was delivered and configured
- Developed scripting and software tools to make each other's lives easier



Inserting Infiniband switch

Inserting node into server rack

Turning on the full cluster for the first time

Installing Geist monitors for power analysis

HARDWARE CONFIGURATION

Mellanox InfiniBand switch provides an excellent way for us to share data between our compute nodes during the different applications' execution, minimizing the communication time of each application.

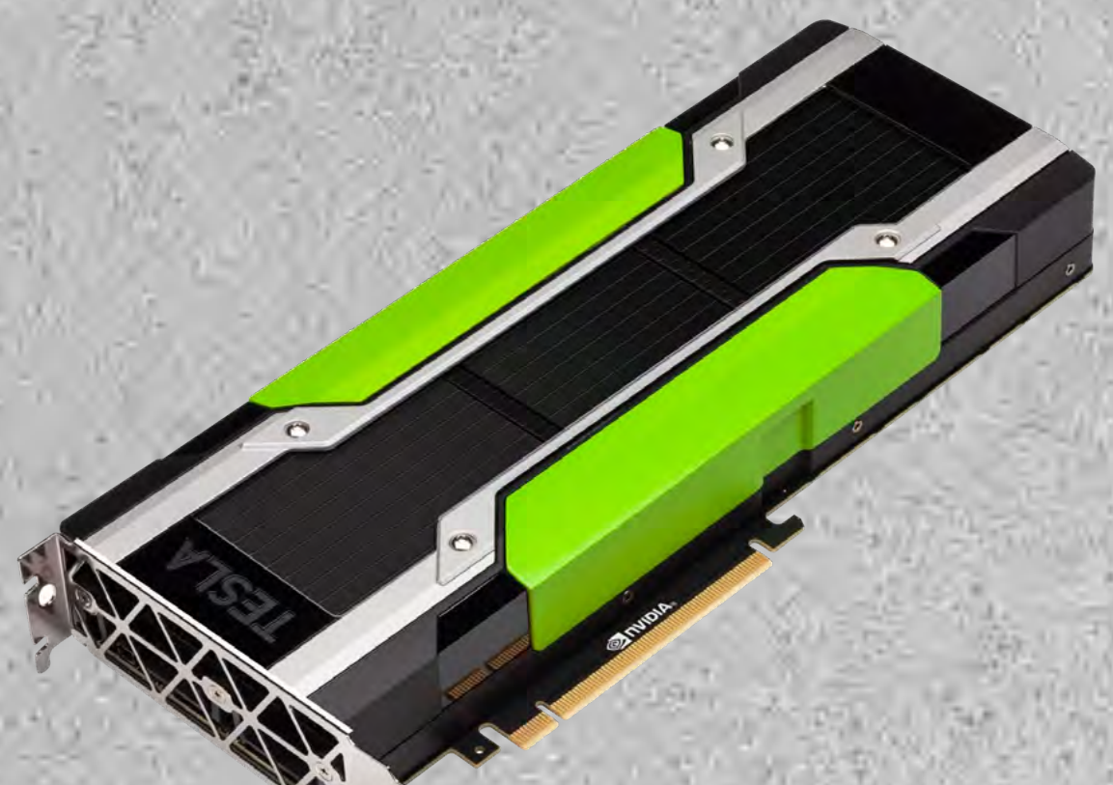
- 1 Dell PowerEdge R730 (Head Node) - 210-ACXU
- 5 Dell PowerEdge R730s - 210-ACXU
- 6 Dell PowerEdge R430s - 210-ADLO
- 1 DLink DGS-101D 16 port Gigabit Ethernet Switch + 12 Ethernet cables
- 1 Mellanox EDR Switch + 12 EDR cards + 12 EDR cables
- 24 Intel Xeon E5-2660 v4 (Broadwell-EP)
- 6 NVIDIA Tesla P100s
- 2 800 GB Intel DC P3700 NVMe PCIe 3.0 SSDs RAID 0 (/scratch)
- 2 300GB 15K RPM SAS HDDs RAID 1 (/home)

APPLICATION STRATEGIES

- **HPL**: Our team is using NVIDIA-provided binaries optimized to run on this GPU architecture; Tuning has been done on 3 different systems to find the important options and their values which produce the best results on our cluster
- **HPCG**: The benchmark will be run with 6 GPUs using a problem configuration chosen specifically for our cluster after several rounds of testing
- **Password**: MD5 hashes will be broken on GPU-centric nodes and BCrypt hashes on CPU-centric nodes for the best performance; When a hash cracking job takes longer than a certain threshold, the job is added to a priority queue and the next job is executed
- **ParConnect**: This will be run on the minimum number of nodes to meet the competition requirements for MPI scaling and Allinea analysis; In our case, the GCC build yielded better timing results than the ICC build
- **Paraview**: We have 6 different versions of paraview installed in case a certain version is inadequate; We plan on using pvserver and the GUI trace analyzer to generate scripts first and then run them directly on the cluster with pvbatch

OPTIMIZING STRATEGIES

- Analyze each application's input files and split the work depending on optimized settings determined through previous iterative test runs
- For certain applications which require GPU usage for the best performance, manipulating the clock rates of the GPUs can allow for more efficient power management; this can be leveraged when running the applications and compared to maximum clock speeds if the power budget allows
- Split the work depending on CPU only or GPU only, which will increase performance and reduce the run times
- Prior to competition, we ran several cases with multiple settings which allowed us to estimate the application power usage; from that, we decided which settings that would perform the best while remaining under the power budget
- The ability to work on TACC's systems (e.g. Maverick) several months prior to arrival of any hardware gave our team a jump start on preparation for this competition



HARDWARE AND SOFTWARE SELECTION

- Shared file systems /home and /scratch via NFS over IB: allows cluster I/O to be available on all nodes
- 2x800GB Intel PCIe NVMe SSDs in RAID 0 (striped) boost head node I/O performance
- NVIDIA P100 GPUs provide significant performance per watt: helps limit the power used during the competition
- OpenHPC repository is utilized to give flexible control over selecting different HPC configuration options and software applications desired in the cluster
- Intel compilers coupled with Intel hardware provides an extra level of optimization which increases the efficiency of each application

SOFTWARE CONFIGURATION

- CentOS7 GNU/Linux distribution
- NFS (version 3) to share directories between nodes
- OpenHPC yum repository and configurations
- Mellanox Infiniband drivers; CUDA 8.0 SDK
- Intel and GNU compilers

CentOS 7 openHPC

NVIDIA® DELL GEIST Future Thinking • Solutions Today TACC bp intel® Mellanox® TECHNOLOGIES