Introduction

Our passion for HPC stems in part from its diverse application. Many fields currently necessitate the use of complex simulations, whether it be to model the fluid flow and heat transfer around an object, decipher and match genomic sequences, or calculate the gravitational force in between neighboring stars. Another example includes the collection of personal data centered about people's online presence. This ever-growing store requires efficient computational and data analysis methods that have and continue to transform our everyday lives. Applications for HPC are growing and expanding into more diverse fields, ensuring that it will be a much-needed skill for the foreseeable future.

Our team is a cross-institutional team formed by the University of Texas at Austin-based Texas Advanced Computing Center composed of members from a variety of fields with one thing in common: a desire to learn more about high performance computing. **Rainier Ababao** is a third year University of Texas at Austin student double majoring in computer science and biology whose interests lie in bioinformatics and natural language processing. **Joe Garcia** is a third year UT Austin computational physics who hopes to learn more about the hardware and software used to perform complex systems simulations. **Josh Granberry** is a third year Texas State University at San Marcos computer science major interested in learning about distributed systems and their role in big data analytics and engineering as he prepares for an internship with Rackspace's Cloud Big Data team. **CJ Kim** is a UT Austin third year electrical and computer engineering major interested in computer architecture, operating systems, and the performance comparisons of traditional single-CPU systems to HPC clusters. **Joseph Voss** is a second year UT Austin mechanical engineering major specializing in the field of mechatronics with an interest in software development. **Joe Zhao** is a second year UT Austin student majoring in mathematics with a concentration in probability, statistics, and data analysis, who hopes to gain a better understanding of the analysis of enormous data sets using HPC technologies. As an aside, Texas State University is proudly represented for the first time at SCC this year.

Strength of Team

None of the current team members have participated in the Student Cluster Competition previously. Every student joined because of their interest in HPC and the application of it to solve real-world problems. Each of us are excited and want to learn as much about this field as we can. We see this competition as a great way to dive into HPC to learn and hone a valuable skill set that will benefit our future careers.

Our team deliberately includes members from many different academic backgrounds. Our majors not only come from the field of computer science, but also mechanical engineering, electrical & computer engineering, computational physics, and computational biology. This variety of disciplines will add a broad wealth of knowledge and alternative perspectives for our team to draw upon as we prepare and compete, exposing each of us to new fields of science.

We consider our diverse exposure to computing concepts, software, and hardware commonly utilized in high performance computing as one of our team's strengths. Roughly half of us have taken parallel computing courses which utilized TACC's Stampede supercomputer as their teaching platform, and the other half will be taking a course in the future. This gives experienced members the opportunity to solidify their knowledge while sharing what they've learned with the less experienced members, who in turn will have a head start for their future courses.

Strength of Diversity

Having students from diverse communities allows us to approach problems from multiple perspectives, leading us to innovative solutions. With respectful team members, the dialectic that forms as a result of different viewpoints can be incredibly productive and bring us greater insight into the challenges related not only to this competition but also to the scenarios that play out in life beyond the university setting.

Efforts made during the team selection process included reaching out to a diverse set of communities at two neighboring universities. Our advisors contacted several professors within different colleges and visited technical classes that contained a variety of students with different majors. Furthermore, the call for SCC16 team selection was posted in college and department-wide emails. The team consists of students from the University of Texas, and a non-traditional student from Texas State who had a culinary arts background prior to computer science exposure. One of the team members from the University of Texas is an underrepresented minority student within the region.

Strength of Hardware and Software Approach

For this competition we plan to have a heterogeneous cluster with seven CPU-only nodes and five GPU-accelerated nodes. Each of these nodes will be contained in a Dell PowerEdge R720 server rack, and outfitted with two Intel Xeon E5-2650v4 series Broadwell-EP processors. This server idles on a low amount of power and will give us a large amount of space to add in disk drives and GPUs. This particular Broadwell processor has 105 W of power consumption at peak, and represents a good balance between power and performance. In particular, it provides the highest available memory bandwidth and QPI bandwidth in the processor series in a fairly constrained power envelope, and many of the proposed applications seem to have high memory bandwidth requirements. The GPU-accelerated nodes will be outfitted with one NVIDIA Tesla K80 each. While AMD has come a long way in recent years with OpenCL support, some of the applications in this competition still only support CUDA, making an NVIDIA chip preferable. These GPUs will more than triple the floating point operations per second and memory bandwidth, with the trade-off of significantly increasing the power usage. We plan to mitigate this power consumption by alternating which nodes are running at maximum power for certain applications, while restricting or even idling other nodes.

One way we plan to increase efficiency is by requiring our head node to take part in computational work. With the head node performing computations, we utilize its cores for calculations and don't waste power with it idling. It is possible in this configuration that the head node will experience high input/output operations on disk. To minimize the overall effect of this I/O load and to be able to handle any sequential I/O that may be required, we will equip a 1.2 TB SSD on PCIe drive provided by Intel.

Each node will have a relatively small but fast hard drive that will be used for local storage of the operating system and applications themselves. Each node will also host a Mellanox EDR Infiniband network card, which then connects them all to a single unmanaged IB switch. A Gigabit Ethernet network will be used in order to manage the workload via ssh across the different nodes. Two GigE ports are needed for the head node, and one for each compute node. In terms of memory, each node will have about 256 GB of DDR4 2400 RAM. This configuration provides full memory bandwidth by populating each of the 4 memory channels available for each Intel Broadwell-EP socket.

We chose to use a heterogenous cluster like this because of the diversity of applications in this year's competition. In our preliminary investigation, we have found that there are highly optimized versions of some of the applications. HPL and HPCG have efficient versions for NVIDIA GPUs, and some password recovery software has support for both AMD and NVIDIA GPUs, with a significant improvement in performance per watt with respect to a pure CPU version. ParaView is an interesting case, because depending on the given workload and workflow required the effort could lean mostly on the CPU or on the GPU, since it seems that ParaView support for GPUs is restricted to the rendering part of the processing pipeline. Even with the possibility for GPU acceleration with ParaView, multi-GPU support may be limited. For this reason we decided to outfit each accelerated node with only one GPU to avoid idling hardware and draining power during one sixth of the total applications. From the description of ParConnect it seems that no support for GPUs is available in the current code, so consideration has been given to the benefit of having additional nodes for flexibility and additional memory footprint for this application. Finally, there is the issue of the mystery application. This could be something that runs exceptionally well with GPUs, in which case our

system will be competitive; or it could be something that has no GPU support, in which case we still have a good number of high powered CPU nodes to complete any workloads associated with it.

To be more power efficient during the competition, the applications which run faster on GPUs will be run on the accelerated nodes, while the other applications can be put on the CPU nodes where they will use less power. The team has secured two Geist metered power strips similar to those used in the competition thanks to one of the sponsors, and these will be used use to test and optimize application placement and power consumption before the competition. Power changes are expected on the exhibition floor when compared to the machine room where our cluster will be accessible before the competition starts. We will use as much time as possible during competition set-up to recalibrate our expectations on the floor and get as close to maximum allowed consumption (without going over!) as possible.

We are well aware that the system we propose, at peak, would consume more power than allowed in the competition. We plan to idle several nodes during our HPL runs to avoid excessive power consumption. We do not expect to run into this issue during the application part of the competition simply because applications do not typically run at system peak performance and the system should fit within the allocated budget under those conditions. Even idling multiple nodes, the presence of the GPUs in the system should allow us to complete HPL with a competitive number, essentially doubling the performance of last year's HPL winners.

The base operating system we plan to use is the CentOS Linux distribution. This is a stable enterprise version of Linux, proven to be highly successful for HPC clusters across the globe. It is open source and can be obtained easily. Additionally, it is highly configurable and extendable, enabling us to tailor it for the individual applications. For compiler technology we will use the Intel HPC software. Because we are using Intel hardware these compilers will provide an extra level of optimization which will increase the efficiency of each application run. A license for this software can be readily obtained as Intel offers free licenses to students. We plan to use MVAPICH2 as our MPI distribution because it has extremely low latency, and TACC's good relationship with The Ohio State University developer team will allow us to learn of any possible runtime optimizations that may be useful in the competition. We will also test the Intel MPI library in case some of the applications are faster when using it. Using the Intel suite of tools will also allow us to investigate on-node and network performance for each application using the Intel VTune profiler and the Trace analyzer, which will help us decide how best to assign applications to nodes. We will also take advantage of optimized numerical libraries when possible, including the Intel Math Kernel Libraries.

During the competition, the applications will be run simultaneously using separate sets of nodes to complete different applications. The applications which can benefit from GPU acceleration such as password cracking and HPL will be run on the GPU accelerated nodes, while the regular CPU nodes will be used to run applications like ParConnect which are memory instead of compute bound. This allows us to have each application run in its best suited environment, and gives us a certain degree of flexibility during the competition. To manage the cluster we plan to use Rocks Cluster to deploy, manage, and troubleshoot our separate nodes. We choose Rocks because it simplifies the many tasks involved in setting up, provisioning, and maintaining a cluster, and it is open source.

The workflow will be controlled manually instead of using workload managers such as Slurm. The reason for this is the small number of separate tasks that need to be managed and the efficiency needed for this competition. To run the six applications effectively, we believe a hands on approach will be less error-prone and require less overhead than a workload manager. By doing so we can inspect the output from each job at run time to ensure that we use our compute time to complete the most workloads.

Team Preparation

Both UT Austin and Texas State offer many courses that help lay the foundations of working in a high performance computing environment. These include but are not limited to Introduction to Scientific

Programming, Parallel Computing for Scientists and Engineers, Parallel Programming, Data Structures and Algorithms, and various domain-based scientific computing classes in the physics, biology, and mathematics departments. Additionally, TACC offers several training courses that cover topics including MPI and OpenMP as well as parallel programming concepts. Coming from many scientific and engineering-related major backgrounds, each team member has taken at least one of these courses. Our collective programming experience includes languages used heavily in the HPC space, such as C, Fortran, Python, and Bash.

In addition to this formal coursework, our advisors at TACC have given us a tremendous amount of advice and support while encouraging us to perform independent research on the competition benchmarks and presenting our findings to our team. In the weeks leading up to this proposal, our team has met once a week in order to thoroughly understand the given applications while also gaining knowledge about the needed hardware/software to run these applications. Not only have we been learning about using Stampede, VirtualBox VM environments, and profiling tools for efficiency analysis, we've toured the server room, been introduced to the hardware components of a node, and learned about the different hardware and software configuration options that can be chosen for our cluster.

We will continue to have regular preparation meetings throughout the summer leading up to the competition in the fall. Some of these meetings will be dedicated to testing and experimenting with the applications on TACC supercomputing systems. Others will focus on specific HPC subjects pertinent to the team, e.g. process affinity, file systems. Once the hardware becomes available, we will begin power testing and exploring application management strategies. Another exercise that we will go through will be a complete recovery from node shutdown, including tests for accessibility, file system availability, and daemons that should be running.

Strength of Vendor Partner/ Institution Relationship

Our team is partnered with the University of Texas at Austin-based Texas Advanced Computing Center (TACC). Travel and per diem costs to Salt Lake City will be supported by TACC and the team has secured the commitment of Dell, Intel, Mellanox, and Geist as hardware suppliers.

Questions about the architecture may be directed to the following points of contact:

- Cyrus Proctor (cproctor@tacc.utexas.edu) from TACC
- Jared Carl (Jared_Carl@Dell.com) from Dell
- Brian Dietrich (Brian.Dietrich@Intel.com) from Intel
- Gilad Shainer (Shainer@Mellanox.com) from Mellanox
- Glen Stewart (gstewart@geistglobal.com) from Geist

Given our past history with vendor partners over the five years that we have attended this competition, we look forward to continuing strong ties with Dell, Intel, and Mellanox. We are excited to work with Geist this year who will be providing power monitoring equipment for us to test our hardware. We are confident that our team will be able to procure the hardware it needs for a competitive system.

We expect to utilize CPUs from the Broadwell-EP line which was officially launched March 31st of this year. We do not anticipate any issue with procuring the chips but in the event that we cannot, we would fall back to Haswell-based CPUs. For GPUs, we are currently in contact with NVIDIA vendors to procure already commercially available server-grade general purpose GPUs (K80). In the unlikely case a sponsorship with NVIDIA cannot be secured, we will obtain server-grade GPUs from our principal sponsor, Dell.