VSCC2020 Team 15: Dabo's Tech Support

CLEMSON UNIVERSITY — Dabo's Tech Support

DELL Technologies

CLEMSON COMPUTING & INFORMATION TECHNOLOGY

Team Introductions:

High-performance computing (HPC) has long been a priority and topic of research across several departments and colleges at Clemson University. Clemson students and faculty regularly use HPC to revolutionize their field of study. Clemson has assembled a diverse and highly competitive team for this year's Student Cluster Competition under the mentorship of Dr. Jon C. Calhoun.

John Hollowell

Major: Computer Science
Application Lead: LINPACK
Intern at LPL Financial for network security and data loss prevention. Skilled in network and server administration.

Nik Heitzeg

Major: Computer Engineering
Application Lead: HPCG
Involved in machine learning research groups. Background in network management and automation projects.

Cavender Holt

Major: Computer Engineering
Application Lead: Gromacs
Undergraduate Research in lossy compression and object detection. Interned at Amazon Web Services during summer 2020.

Griffin Dube

Major: Computer Engineering
Application Lead: MemXCT
Undergraduate research studying exascale applications at ORNL and optimizing SZ lossy compressor. Interned at Delta Airlines for parallel programming.

Sarah Placke

Major: Computer Engineering
Application Lead: CESM
Undergraduate research in fault tolerance. First generation Japanese American college student working with the Missile Defense Agency.

Sansriti Ranjan

Major: Computer Engineering
Application Lead: IO-500
Involved in Clemson IEEE Robotic team. Experienced in running IO-500 benchmarks. Learning about HPC and ML.

Team Preparation:

As a team, we:

- Attend weekly webinars to learn more about the applications we are running and become more familiar with Azure.
- Test the competition applications on a variety of Azure resources in order to determine the best configuration for running.
- Hold weekly meetings to share our progress on application optimization and competition strategy.
- Utilize Clemson's Palmetto cluster to quickly iterate and test applications and environments

Cloud Configuration and Software Description:

Hardware Configuration:					
Application List	**ND40rs v2[1]**	**E64d v4[2]**	**NC24r*[3]**	**NC24rs* v2[4]**	**L64s v2[5]**
LINPACK	✓	✓	✓	✓	✓
HPCG	✓	✓	✓	✓	✓
IO-500	✓	✓	✓	✓	✓
CESM	✓	✓	✓	✓	✓
Gromacs	✓	✓	✓	✓	✓
MemXCT-GPU	✓	✗	✓	✓	✗
MemXCT-CPU	✓	✓	✓	✓	✓

Legend:	
Most Performant	✓
Performant	✓
Compatible	✓
Not Compatible	✗

Software Versions:	
gcc	10.1
cuda	10.2.89
cmake	3.17.4
openMPI	4.0.3

Software Stack:	
OS	CentOS 7
PFS	BeeGFS
Scheduler	PBS Pro

1 - NVIDIA Tesla V100 (32 GB of GPU mem. ea.), Intel Xeon Platinum 8168, 672 GiB of RAM
2 - Intel® Xeon® Platinum 8272CL, 504 GiB of RAM, Memory Optimized
3 - NVIDIA Tesla K80 (48 GiB of GPU mem. ea.), Intel Xeon E5-2690 v3, 224 GiB of RAM
4 - NVIDIA Tesla P100 (64 GiB of GPU mem. ea.), Intel Xeon E5-2690 v4, 448 GiB of RAM
5 -AMD EPYC 7551, 512 GiB of RAM, Storage Optimized

- In order to maximize use of our cloud budget, we will spin up separate instances based on each application's needs.
- Our software stack mirrors Clemson's Palmetto Cluster so that we can easily adapt to running on our cloud instances.
- With our choices of both compute and memory optimized VMs, we are well equipped to handle the mystery application.

Why will we win?

- We have a diverse group with complimentary experience in fields such as HPC and cloud computing.
- Our application leads have backgrounds suitable for handling the unique aspects of each competition application.
- Our team have been working together for over a year and have a combined total 11 years of HPC experience, improving our ability to communicate and collaborate while striving to achieve a goal.
- As a first-year team, we are prepared to be "All In" (a Clemson moto) and bring our best skills to win.

Strategizes for Running and Optimization:

Benchmarks:

LINPACK - John Hollowell

- A dense matrix multiplication benchmark that makes heavy use of GPU computer and memory.
- Highly tunable for a variety of hardware.
- The homogeneous nature of cloud resources allows simpler tuning and faster end results (better optimization of P and Q).
- Utilize high performance GPU nodes with large GPU Memory capacity to reduce data transfer costs and provide the optimal environment.
- Optimize the split of CPU and GPU use to maximize the total performance while not hindering GPUs waiting on CPUs.
- Test performance of optimized binaries from Nvidia and Intel against building from standard source.
- Support from CCIT who has run HPL on the Palmetto cluster for TOP500 runs.

HPCG

- Developed as a more relevant benchmark [compared to HPL] for practical HPC applications as HPCG has a memory access rate proportional to N, meaning performance is strongly influenced by memory bandwidth.
- Solves a system of sparse linear equations arising from a three-dimensional partial differential equation model using the conjugate gradient method
- The first CLI argument determines the local memory mesh for each rank (nx, ny, nz), which represents `X*Y*Z` points per MPI rank. The second determines ranks in the processor mesh (npx, npy, npz). This determines how effectively the ranks communicate
- For optimization, we explore both Jagged Diagonal and ELLPACK on GPUs to achieve efficient memory accesses.
- Explore message aggregation to avoid multiple sends between nodes

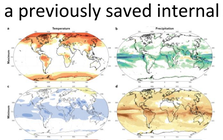
IO-500 - Sansriti Ranjan

- The IO-500 has been created to compare storage systems. It relies on community benchmarks while tracking storage performance.
- The community benchmarks are simulated based on IOEasy, IOHard, MdEasy, MdHard and Find.
- There are four phases for calculating the IOR and Mdtest scores - produce phase, consume phase, find phase and calculating the bandwidth and iops.
- Produce phase consists of write and create tests whereas the consume phase consists of read and stats test. Read and write is performed on IOR while create and stats is performed on MD. Find phase only runs the find operation. The Bandwidth comprises the four IOR Easy/Hard read or writes while iops comprises the mdtest, find.
- Using Lsv2-series instances with direct-mapped MVMe SSD storage for the fastest storage.

Applications:

CESM - Sarah Placke

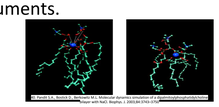
- CESM is a numerical simulation that models how Earth's climate may respond to change; it is used to model potential past, present, and future Earth systems.
- Due to the nature of the competition, we focus on optimizing particular statistics from the produced timing table such as the overall model cost and throughput.
- Because our instances utilize multiple processors, we focus on optimizing parallelization and configuration files instead of command line execution flags.
- We utilize prior testing to understand where typical throughput congestion occurs during execution. More resources are allocated to the discovered areas.
- We use restart files to overcome timing restrictions. These allow CESM to start from a previously saved internal state.



Gromacs - Cavender Holt

Gromacs is a molecular dynamics package mainly designed for simulations for of proteins, lipids, and nucleic acids.

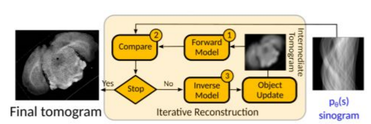
- Compile Gromacs with `-DGMX_SIMD` for non-bonded and bonded force calculations, PME, and neighbor searching computations.
- Compile with `-DGMX_MPI=on` enabled to allow openMP scaling of threads per rank and creation of separate ranks for PME calculations as PME requires global communication, a bottleneck for performance.
- Offload computation to GPU particularly for short range non bonded interactions as a GPU can perform these type of computations much more efficiently.
- Generally we compile for all the features we intend to employ and tune these features at runtime through the extensive set of available command line arguments.



MemXCT - Griffin Dube

Memory-Centric X-Ray CT Reconstruction with Massive Parallelism

- We focus on selecting instances to most accurately reproduce performance and strong/weak scaling results.
- To reproduce KNL performance and scaling, we use E64d v4 instances with Hyper-Threading and AVX-512.
- Replication of single-node performance will take place on instances with V100, P100 and K80 GPUs (ND40rs v2, NC24r*, and NC24rs* v2 respectively).
- Strong and weak scaling will take place on NC24r* instances, providing the most similar GPU performance to Blue Waters.
- To determine the best configuration for our system, we run a sweep of the tile, block, and buffer size parameters.



M. Hidayetoğlu et al. (2019)