Application Profiling and Power Management for the Student Cluster Competition

Alex Ballmer\textsuperscript{+}, David Ghiru\textsuperscript{+}, Ryan Mitchell\textsuperscript{+}, Ryan Prendergast\textsuperscript{+}, Hasan Rizvi\textsuperscript{+}, Iva Veseli\textsuperscript{+}, William Scullin\textsuperscript{+}, Ben Allen\textsuperscript{+}, Ioan Raicu\textsuperscript{+}

\textsuperscript{1}Department of Computer Science, Illinois Institute of Technology, Chicago, IL
\textsuperscript{2}Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, IL

Our Team

- **Alexander Ballmer**: a 4th-year Computer Science undergraduate at IT. He was a member of the IT SCC teams in both 2014 and 2015. He focuses on the cloud components and the team's system administration.
- **David Ghiru**: a 4th-year Computer Science student at IT. He has interned twice with IBM Technolo-gies LLC, and has also done research in the IT DataLab. This is his first time as an SCC official contestant. He is the team’s cloud component expert.
- **Ryan Mitchell**: a senior at Adlai E. Stevenson High School. He is majoring in Computer Science and has attended hackathons and worked extensively with Arduino. He is responsible for the mystery application.
- **Ryan Prendergast**: a senior at Maine South High School. He is majoring in Computer Science and has attended hackathons and worked extensively with Arduino. He is responsible for the mystery application.
- **Hasan Rizvi**: a 3rd-year IT Computer Science undergraduate with an Ap-plied Behavior Analysis minor. He has a background in Linux server administration, and virtual machines. He is the team’s system admin and benchmarking specialist, as well as being responsible for the mystery application.
- **Iva Veseli**: a 4th-year University of Illinois School of Business student majoring in Computer Science.

Hardware

- **Chassis**: Intel Wolf Pack with Calyos Heat Pipe Coolers
- **CPU**: Intel Xeon Platinum 8180 2 sockets/node
- **Memory**: 384 GB DDR4-2666
- **Storage**: Intel NVMe SSDs 2/node
- **Accelerators**: NVIDIA Tesla V100 8
- **Interconnect**: Intel OmniPath

Preparation Strategy

We started preparing as a team in June. Throughout the summer, we worked three days a week, with pairs of team members focusing on each of the competition’s applications. We obtained a deep understanding of the applications, their structure, and their input files.

We modeled the scalability of each application and determined the most efficient suite of parameters. Experiments included:

- Intra-node threading through OpenMP and MPI
- Inter-node scaling through MPI
- CPU frequency scaling

Some team members focused on other competition aspects. The cloud component focused on:

- Determined which node types are the most efficient per dollar
- Determined which applications to run on the cloud

Visualization tools were MiBayes, LAMMPS, and Born

10 16x16 LED squares on the side of our cluster for system monitoring

Power Management

The primary constraint in the competition is a total cluster power consumption limit of 3000 Watts. To stay below this limit while maintaining high application performance, we have:

- Set a hard limit of 3000 Watts over the entire cluster using Intel DCM
- Profiled application power consumption to inform scheduling decisions such that:
  - We can modify parameters to save energy without significantly impacting performance
  - Power intensive applications can be run concurrently with those that use less energy
  - Established system monitoring (including power usage)

Application Optimization

Pre-Compilation Optimizations

- Compiler usage (GNU vs Intel)
- Compiler optimization flags, such as O2, 03, fp-model, -fmax-model, etc.
- Static vs dynamic compilation of binaries

Run Time Optimizations

- Intra-node scalability:
  - Determining the ideal number of cores per node to run on
  - Whether to use pure MPI parallelism or a combination of MPI+threading (via OpenMP)
- Inter-node scalability:
  - Deciding whether each application scales well across multiple nodes
- CPU frequency scalability, subject to power constraint (3 kW)
  - Determining whether it is more efficient to maximize node count with a slower clock speed, or to maximize clock speed with fewer nodes

We used Allinea Map and Intel iVume to profile applications and identify resource bottlenecks.

- Allinea Performance Reports identified the primary bottlenecks (CPU, MPI communications, or I/O)
- Profilers showed us which functions took up most of the runtime

Other School

The cloud component:

- Visualization:
  - We modeled the scalability of each application and determined which node types are the most efficient per dollar.
  - We performed an OS scalability study for our hardware.
  - The current task is integrating MiBayes with the cloud component.

- Hardware:
  - Calyos coolers use liquid evaporation to efficiently cool each processor, reducing amount of fans in the system.
  - The Intel 8080’s two AXI-512 FMA units/core gives us an advantage when running highly vectorized applications.
  - Our RAM fills all six memory channels for optimal memory bandwidth and performance.
  - NVMe SSDs using 3D XPoint memory allow for low-latency, high bandwidth IO.
  - OmniPath’s on-load method of transferring data offers more performance.

Why We Will Win

Most of our team members have been working on the applications since June 2017.

- Each application has been studied extensively and profiled by at least two team members.
- We also have three team members responsible for the cloud component, and several working on visualization.
- We have practiced with an 8-hour mock competition including interviews and a ‘mystery application’ (WRF).
- We have had excellent support from our advisors, backup team members, and sponsors who have provided their experience and advice to help us prepare well.
- We have a steady supply of candy to fuel us throughout the competition.