TEAM Peking University, China

Yun Liang (advisor), Xiaolin Wang (advisor), Lei Yang, Wenbin Hou, Zhenxin Fu, Zhuohan Li, Yihua Cheng, Yifan Wu

{ericlyun,wxl,yangl1996,catchyrime,fuzhenxin,zhuohan123,chengyihua,evanwu}@pku.edu.cn

About Peking University

Peking University is a comprehensive national key university in China. It ranks as one of the best universities in China and a top university in Asia and worldwide. Here, top Chinese students seek academic excellence and dedication to the society.

HPC in Peking

High performance computing matters to the research of many institutes in Peking University. An HPC Platform administrated centrally by school can greatly boost our efficiency. For example, the application Paraview last year, Born and MrBayes this year will achieve enormous improvement in analysis with an HPC platform.

To meet the demand of high performance computing, the PKU HPC Platform was constructed and opened to the whole school in 2016. It is liquid cooled, with amazingly high PUE (Power Usage Effectiveness). Still in attempt as it is, it enhanced working condition for researchers at Peking University, bringing access to everyone in need.

About Our Team

We have seniors, juniors and sophomores. This is sophomores' first time to participate in the Student Cluster Competition.

YANG Lei Senior
Team leader and coordinator
Expert in networking and sys-admin
Focus: Benchmark, system administration

FU Zhenxin Senior
Research interests in Natural Language Processing
Focus: Reproduction Task (Tersoff)

CHENG Yihua
Sophomore
Expert in Parallel Computing
Focus: Born

Prof. LIANG Yun
Advisor, School of EECS
Research Interests: High performance computing, compilation and architecture.
Publication in Top-tier conferences: HPCA, ISCA, MICRO, FPGA, DAC, etc.

HOU Wenbin Senior
Expert in CUDA and sysadmin
Focus: MrBayes, Born

LI Zhuohan Junior
Expert in Machine Learning
Won medals in ACM/ICPC
Focus: Born

WU Yifan Sophomore
Expert in interdisciplinary computing
Focus: Reproduction Task (Tersoff)

Prof. WANG Xiaolin
Advisor, School of EECS
Research interests : High performance computing, algorithm design.
Publication in Top-tier conferences: TACO, USENIX ATC, ICS, etc.

Our Preparation

Self-study

Freshmen and sophomores joined our HPC discussion group months before the release of the SC17 SCC applications. Juniors and seniors offered help on parallel programming to them.

For each of the applications, we tried several different approaches to achieve the best performance. To improve the performance of applications, we learnt CUDA and OpenACC, read through each line of the source code, and delved into algorithms behind the tasks. We organized and reconstructed the source code, eventually reaching our present code performance. For Born, we ran both the CPU and the GPU version, worked seperately with CUDA and OpenACC, and finally reached a boost in speed of over 30 times that of the original version.

Teamwork

During preparation, every team member was assigned an application to study. Weekly discussions and seminars were held to discuss the progress and blockage of each application, as well as to introduce interested topics in HPC to our junior members.

Support from our venders

We have benefitted a lot from our collaboration with OMNISKY and NVIDIA. OMNISKY provided us with high-end cluster and devices, and hosted our machine during preparation. NVIDIA have provided vendor-optimized benchmark applications, toolchains and math libraries so that we can fully optimize our applications. We have also learnt a lot of skills about optimizing scientific applications and HPC system maintenance from the collaboration.

Special Thanks

OMNISKY 睿天

NVIDIA

Hardware and Software Config

Hardware Configuration

SuperMicro SYS-4028GR-TR2 Server

CPU	2×Intel Xeon E5-2690 v4	14 Cores, 2.6GHz
GPU	10×NVIDIA Tesla V100	
MEM	512GB	DDR4
HDD	Intel 750 Series Seagate HDD	400GB PCI-e SSD 4TB SATA

The server we are using is a brand-new model from SuperMicro with 10 NVIDIA GPUs. It has 10 GPU slots attached to the same root complex,making peer-to-peer data exchange between GPUs much faster than that on a traditionalbalanced configuration.

Software Configuration

This year we are going to use Ubuntu 16.04. The high GPU density architecture enables us to, for the first time, use only a single computing node, which saves the energy overhead of network interface cards (NICs). Also, this year's new cloud platform resource gives us extra confidence in choosing this single-node architecture.

ubuntu

CUDA C/C++ Runtime Libraries CUDA Profilers

allinea PERFORMANCE REPORTS

Optimizations and Strategies

Benchmark

HPL and HPCG benchmark programs are provided by our sponsors OMNISKY and NVIDIA. The binaries are fully optimized for our hardware. We also fine-tuned the parameters on our cluster using a home-brewed tuning script.

Born

Born is born with a simple computation pattern with little data dependency, making it suitable for GPU acceleration.
We have managed to build a multi-GPU version of Born, which is multiple orders of magnitude faster than the CPU version.

MrBayes

MrBayes can be parallelized on multiple GPUs. We have managed to implement a working multi-GPU version of MrBayes, making full use of our cluster.
What's more, we build our own MPI implementation from scratch for lower data communication overhead on our single-node cluster.

Running Strategies

We have profiled all the applications to gain a deep understanding of their computation and memory-access patterns. With this knowledge we can estimate the runtime of each application. Then we will schedule the applications to maximize the resource utilization.

Reproduction Task

The Tersoff Multi-body Potential application comes with codes for different hardwares, including CPU and GPU.
Apart from our cluster, we also reproduced the results on the cloud platform, both CPU and GPU verison.
The results from both hardware platforms are checked against the original paper, in order to make sure our results align with the paper.

Mystery Application

With the combination of our high-GPU-density single-node cluster and the flexible cloud platform, we are confident that our hardware can handle mystery applications of any kind. For applications that are optimized for CPU, we will start a cloud cluster instance and run the application in the cloud. For applications that are optimized for GPU or easily parallelizable, we will run them on our GPU cluster.

Power Management

Onboard SYS-4028GR server there is a powermeter, which can be read from IPMI locally or remotely to help us monitor the machine.
We have developed tools to control the frequency of GPU and CPU, in order to make full use of the power budge.
A customized tool is built to record the power consumption in a round-robin database (rrd) and push power overage alerts to team members' phones, should a power overage occur.

We Will Win!

- We have passionate members who worked together for the competition!
- Each application is well studied and tweaked for months, so we have a good understanding of different applications for the competition.
- Our members are veterans in areas like Parallel Computing, System Administration, Power Management and Computer Networking
- Well-designed hardware and software configurations are prepared, and we have fully exploited the performance of V100.
- Support from big-name vendors like OMNISKY and NVIDIA.