National Tsing Hua University Team In SC24 Student Cluster Competition

Chun-Mu Weng, Tung-Yu Hsieh, Chih-Yu Hsieh, Zhan-Yi Lin, Chen-Ai Pai, Hao-Tien Yu, Jerry Chou

Department of Computer Science, National Tsing Hua University, HsinChu, Taiwan

Architecture

Cluster Overview

Compute Node x3:
- high performance CPU
- 2 equipped with accelerating card

InfiniBand NIC:
- 200Gbps bandwidth HDR IB
- maximize IO performance

Gigabit Ethernet:
- 25Gbps bandwidth with MT4127
- used for management and monitoring

Hardware

CPU: Intel Xeon® Platinum 8458P
- 48 cores per processor
- 2 CPUs per node

GPU: Tesla H100 (2 nodes)
- 4 GPUs per GPU node
- 2TB/s memory bandwidth
- 350 Watts power consumption

NVLink Bridges:
- Support communication between 4x H100
- Each pair provides 600GB/s of bandwidth

RAM DDR4 4800 MHz 64GB x16:
- Not using all DIMM slots for max freq

SSD Samsung 7.68TB x2:
- RAID 0 for best performance

Software

OS:
- Ubuntu 24.04 LTS

Package Manager: Environment Modules, Spack
- convenient for managing environments
- save time on compilations
- provide more flexible options

Compiler:
- GNU Compiler Collection 13.2.0
- Intel oneAPI-2024.2.0

GPU Software:
- CUDA 11.8, 12.4
- HPC-SDK 24.5

MPI:
- OpenMPI 4.1.7
- Intel oneAPI-2024.2.0

InifiniBand:
- rdma-core 51

Optimization

General Methodology

Profiling:
- Intel® VTuneTM Amplifier, NVIDIA® Profiling Tools
- identify the hotspots of an application
- provides and analyzes time usage details

Code improving:
- vectorize with SIMD, offload to accelerators
- reduced communications and computations
- cache optimization, branch prediction, data alignment

Compile: enable modern compiler optimization
Tuning: mpitune, CPU affinity, NUMA binding

LINKPACK & MLPerf

LINKPACK:
- settings tuning
- automatically optimizing and running

MLPerf:
- try parameters tuning in config file and research the meanings of different parameters
- find balance between power and performance
- set proper bindings to reduce communication overhead

NAMD

- CPU and GPU affinity binding
- try different parameters / compiler suite
- exploit GPUs by GPU-resident mode
- collab with expertises in computational chemistry
- profile to know the computation pattern in different test cases
- checkpoint to prevent unexpected error
- choose best libraries such as FFTW libraries to make the most use of them

ICON

- adopt process binding
- leverage UCC, NCCL
- try coding optimization
- conduct profiling
- tune namelist parameters: nproma, sub_nproma
- enable GPU
- experiment parallel file system

Reproducibility Challenge

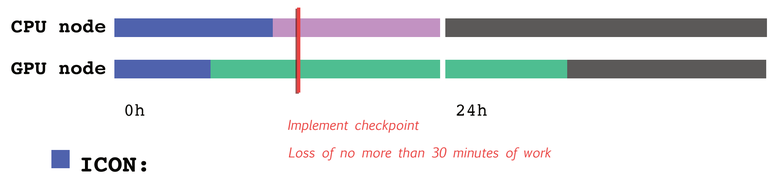
- try NFS and BeeGFS in the paper
- evaluate the improvement of a new software implementation designed to address specific pain points
- use DataLife, a tool proposed in the paper "Data Flow Lifecycles for Optimizing Workflow Coordination," to assess specific applications.
- compare the performance difference between the original software and the improved version based on the pain points identified through tools provided by the paper's authors
- conduct reproducibility experiments on a smaller-scale cluster, focusing on modified experimental objectives

Preparation and Tactics

Praparation

In order to get familiar with the cluster environment and the SCC selected applications, we meet our advisor and coach every week to discuss in the problems we encountered.
Moreover, additional HPC applications are selected to be installed and optimized for our server every few weeks. We are fully prepared and confident that we are ready for the fight!

Scheduling in 48h



ICON:
- complete evaluation and start writing report
- can overlap with reproducibility challenge

NAMD:
- can be accelerated with GPU
- can overlap with reproducibility challenge

Reproducibility: use CPU mostly

Mystery:
- run after detailed research and analysis

Controlling

NFS over RDMA:
- allowing machines to mount a disk partition on a remote machine
- leverage IB RDMA

Dynamic Frequency Scaling:
- developing tools to scale CPU frequency on demand

P-State Driver:
- finer grain frequency management with CPPC

CGroups resource management:
- prioritizing critical applications

Power Shut-off:
- implementing checkpoints on all applications

Monitoring

Grafana:
- visualizing real-time server metrics
- making decision quickly

IPMI:
- reports server power consumption, temperature
- Controls fan speed

Notifications:
Push notifications to members' phones
- on power limit events
- on program completions
- on execution failures

Diversity and Collaboration

Diversity Team

Our team consists of five CS majors and one student from Interdisciplinary Program of Technology and Art. All team members are first-time participants. Our team members are diverse in terms of education, gender, orientation, and motivation. One-third of the members are traditionally considered to be underrepresented in STEM fields.

Our goal is to create an SCC team that can represent all communities in Taiwan. We are proud to have these individuals from different backgrounds and provide their skills.

Collaboration

Chun-Mu Weng
- Major: Computer Science
- Skillset: System Administration
- Working on: System Management, Reproducibility

Zhan-Yi Lin
- Major: Interdisciplinary Program of Tech and Art
- Skillset: Parallel Computing, CUDA Programming
- Working on: Architecture, ICON

Chan-An Pai
- Major: Computer Science
- Skillset: LLM, Model Inference
- Working on: MLPerf, Reproducibility

Chih-Yu Hsieh
- Major: Computer Science
- Skillset: Parallel Computing
- Working on: ICON

Tung-Yu Hsieh
- Major: Computer Science
- Skillset: Algorithm
- Working on: MLPerf, NAMD

Hao-Tien Yu
- Major: Computer Science
- Skillset: LLM, Agent
- Working on: NAMD, Reproducibility

Supporter

Our School

National Tsing Hua University is a comprehensive research university offering a full range of degree programs in science, technology, engineering, humanities, social science, and management. It consistently ranks as one of the premier universities in East Asia.
https://www.nthu.edu.tw/

Our Sponsor

Quanta Cloud Technology
https://www.qct.io/

National Applied Reasearch Labs
https://www.narlabs.org.tw/