Tsinghua University Team
SC21 Virtual Student Cluster Competition

Mingshu Zhai, Zeyu Song, Kaiyuan Rong, Yanyu Ren, Yuxi Zhu, Juncheng Cao, *Wentao Han and Jidong Zhai*

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

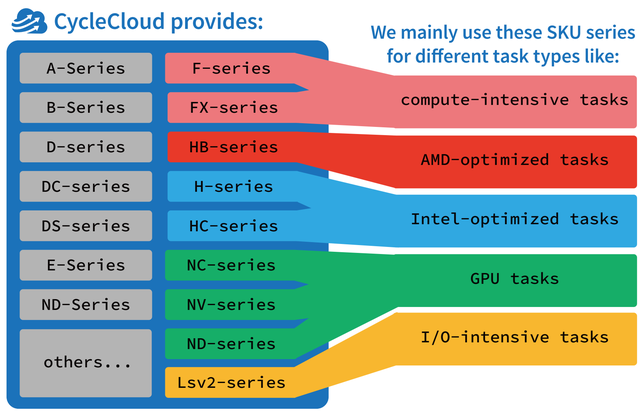
we will win powered by

fully-utilized cloud resources

cloud overview

- ✓ various computing hardwares for different HPC workloads
- ✓ monitoring and visualization toolchain
- ✓ built-in templates and custom images
- ✓ autoscaling on top of commonly used schedulers

SKUs choices



- ■ *F-series & FX-series* :
  - · Intel® Xeon® Platinum 8272CL processors for Fsv2
  - · high CPU-to-memory ratio
  - · the best value in price-performance in the Azure portfolio
- ■ *HB-series*:
  - · 120 AMD EPYC™ 7003-series processor cores for HBv3
  - · up to 200 Gb/sec Mellanox® HDR InfiniBand
  - · the best performance, scalability, and consistency for CPU computing
- ■ *H-series & HC-series*:
  - · 44 Intel® Xeon® Platinum 8168 processor cores for HC44rs
  - · high CPU frequency with large memory per core
  - · the best for applications that rely on Intel® toolchain
- ■ *NC-series & NV-series & ND-series*:
  - · various single, multiple, or fractional NVIDIA® Tesla® GPUs
  - · ability to fully utilize GPU performance
  - · the best for compute-intensive GPU-accelerated applications
- ■ *Lsv2-series*:
  - · 19.2TB (10x1.92TB) NVMe SSD M.2 device for L80sv2
  - · up to 3.8M Read IOPS and 20GBps throughput
  - · the best for single node I/O-intensive tasks

software choices

- · ***CycleCloud Ubuntu 18 (with Azure tuned kernel)***:
  - · friendly to HPC applications with native supports
  - · support for Hyper-V™ NVMe Direct technology
- · ***Resource manager***:
  - · instantiate the VMs directly instead of through schedulers
  - · use the official CycleCloud cli and web interfaces to orchestrate
- · ***Spack***: flexible package manager for HPC
  - · manage multiple versions of compilers, utilities and libraries
  - · Intel® Compilers 19/20/21; GCC 6.3/9.2/10.2
  - · Intel® MPI 2019/20/21; Mellanox® HPC-X; OpenMPI 1.10.7/3.1.6/4.0.5
  - · NVIDIA® CUDA 8/9/10/11; cuDNN 5/6/7

we will win for

specialized optimizations

general methodology

Profiling: find out the hotspots of an application
- · ARM® Allinea MAP, Intel® VTune™ Amplifier, NVIDIA® Profiling Tools

Code improving:
- · overlapping communication and computation
- · cache optimization, data alignment, branch prediction

Compiling: with recent optimizing compilers and tune options

Tuning: mpitune, CPU affinity, NUMA binding, etc

LINPACK & HPCG

Tune program input parameters:
- · fully understand and tune the problem size for each possible scale

Test different GPU cards to choose the most efficient SKU

Use all possible GPU nodes for close to linear scalibility

IO500

Use a custom file system called MadFS:
- · a highly-scalable distributed file system for metadata operations
- · inspired by GekkoFS but rewritten in Rust

Use storage-optimized node with NVMe storage

Tuned paramaters: process numbers, file size, files per process...

Cardioid

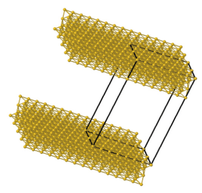
Background understanding:
- · mechanism of cardiac electrophysiology simulation
- · motivation and design of Niederer Benchmark

Parallel strategy:
- · choose between MPI processes and OpenMP threads
- · carefully decide GPU usages for higher efficiency

Quantum ESPRESSO

Underlying principles:
- · relationship between the Schrödinger equation and density-functional theory
- · iterative solution of KS equations

Tuned performance:
- · different building options : MKL, scalapack, GPU support
- · cross-node communication

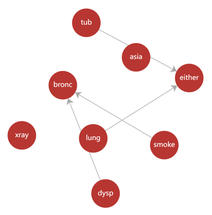


visualization of the AUSURF benchmark

Reproducibility Challenge (ramBLe)

comprehensive understanding of the original paper and source code

Experiments:
- · tested under different libraries and configuration
- · use existing dataset



a bayesian network of the ASIA dataset constructed by ramBLe

Mystery Application

Different compiling methods: thanks to ***Spack***

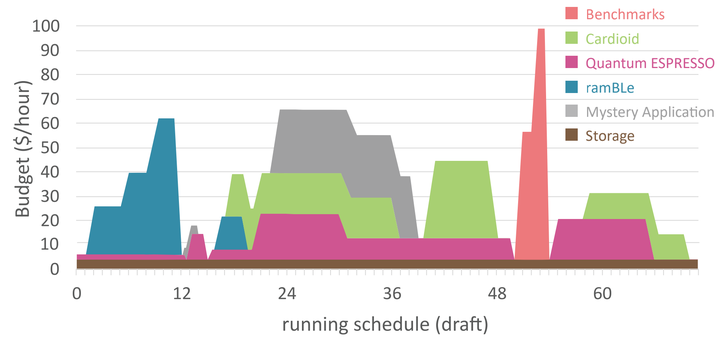
Resource preservation: budget co-scheduling with other applications

Analysis: scalability, communication-intensity and GPU-friendliness

we will win since we have

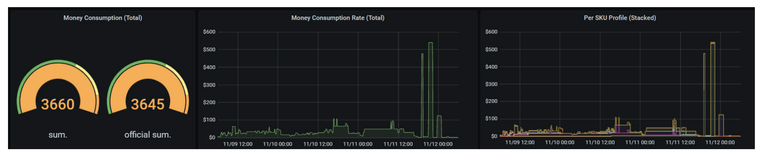
sophisticated tactics

scheduling in 72h allotted time & budget



- ■ *Benchmarks*:
  - · allocate spare budge in case of unexpected failure
  - · take advantage of our experience last year
  - · run on a large scale of nodes to take advanatge of cloud scalablity
- ■ *Cardioid*:
  - · determine resource usage by task size
  - · need both CPU nodes and GPU nodes
- ■ *Quantum ESPRESSO*:
  - · examine the input file for load balancing
  - · test various machine to find the most cost-efficient SKU type
  - · run actual tasks on CPU and GPU nodes
- ■ *Reproducibility Challenge (ramBLe)*:
  - · high priority for performance evaluation to start report writing
  - · finish scalability tests first to gradually release nodes
  - · only needs CPU nodes
- ■ *Mystery Application*:
  - · run after thoroughly analysis and careful optimization
  - · decide the SKU based on analysis
  - · choose run time based on optimization and scoring rubric

controlling over budget consumption



Realtime money consumption with Grafana
An alarm bot deployed to team slack

working together on campus

SC20 VSCC team members and advisors

we will win with a team of

diversity and collaboration

efforts to broaden participation in under-represented communities

✓ **Acting as short-time voluntary teachers to introduce computer science and supercomputing among young students in ethnic minority areas**
- · in Qinghai province
- · during summer vacation

lecture of supercomputering at high school in Qinghai province

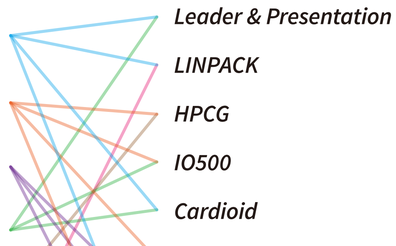
create a diverse team

- ✓ ***Birthplace diversity (particularly in China)***:
  - · members from 6 different provinces, with different cultural backgrounds
- ✓ ***Diversity of experience***:
  - · different years of study: experienced HPC contestants and newbies
- ✓ ***Diversity of personal interest***:
  - · took different courses: architecture, algorithm, OS, AI, database, etc.
  - · interest in different fields: math modelling, deep learning, network, etc.
  - · have different social works: opensource community, volunteer teacher, co-founder of student writing association, etc.
  - · different extracurricular activities: cyalume dance, drawing, badminton, international style of ballroom dancing, etc.
  - · plan on different future: doctoral/master study, work, startup, etc.
- ✓ ***Encourage females, sophomores and students with interdisiplinary knowledge to join us***

collaboration based on different majors and skill sets

- · ***Mingshu Zhai***: **Computer Science**
  - · team leader, graph mining
- · ***Zeyu Song***: **Computer Science**
  - · molecular dynamics, ML
- · ***Kaiyuan Rong***: **Statistics & CS**
  - · computer system and architecture
- · ***Yanyu Ren***: **Economics & CS**
  - · algorithm, GPU scheduling
- · ***Yuxi Zhu***: **Computer Science**
  - · program auto-tuning
- · ***Juncheng Cao***: **Computer Science**
  - · networking and automation

...More backup members in Chemistry, Physics and other majors...



Tsinghua SC21 VSCC team

the whole team (with backup members !)

we will win with

the support of

our advisors

The team led by them has achieved 14 international champions in student supercomputing competitions at SC, ISC and ASC.
- · ***Prof. Jidong Zhai***
  - · research interests: compiler and performance evaluation
- · ***Dr. Wentao Han***
  - · research interests: computer systems and parallel computing

our school

- · **Tsinghua University**
  - · Motto: Self-Discipline and Social Commitment
  - · Spirit: Actions Speak Louder than Words

our sponsors

BITMAIN intel Mellanox TECHNOLOGIES

Corresponding author: zhaijidong@tsinghua.edu.cn