Massachusetts Green Team @ SC21

Boston Linux & Unix - April 21, 2021

Introduction to the Team - History

- Started in 2011, competed at SC11 and SC12
  - 2nd Place
- Called Team Chowdah
- Nvidia Tesla K10s
  - Cutting edge at the time!

Introduction to the Team - History

- Golden Days
- SC14, ASC15, ISC16, SC16, ASC16, SC17, SC18
- Various universities were represented each year
    - After 2018, everything just kinda stopped...

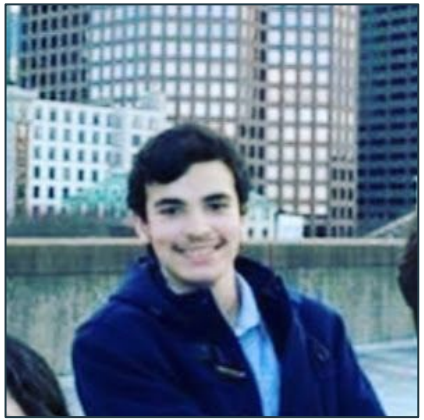
Episode VI: Return of the Massachusetts Green Team

Introduction to the Team

David Shen, Computer Science, BC’22

IO500 Filesystems

New Competitor



Carlton Knox, Computer Engineering, BU’23

Application - Cardioid

New Competitor

Richard Kumahia, Computer Engineering, UML’22

Application - Quantum Espresso

New Competitor

Introduction to the Team

Howie Chen (Po Hao Chen), Computer Science, BU'23

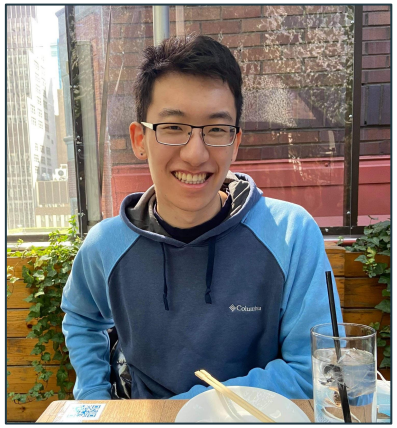
Benchmarking Lead

Returning

Michael Klein, Computer Science, BU'24

Application - Quantum Espresso

New Competitor



Ben Li, Electrical & Computer Engineering, BU'22

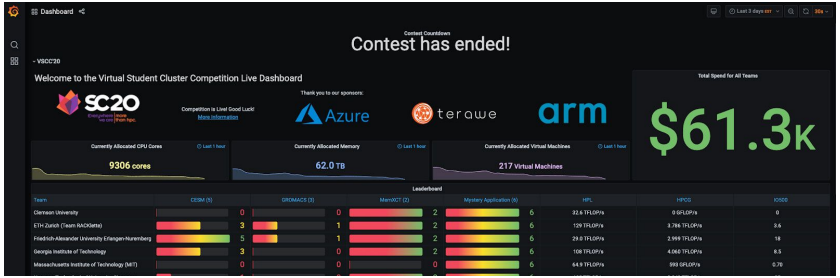
Applications Lead

Returning

The Student Cluster Competition @ SC

- Multi-day event - Early November
  - Part of the Supercomputing Conference
- International representation
  - 6 students per team, all undergraduate
  - 16-18 teams
- Build a supercomputing cluster
  - 3 Benchmarks, 3-4 Applications, Reproducibility Challenge
  - Challenge - Optimize workloads to achieve highest score
- Great opportunity for students to learn about HPC and network

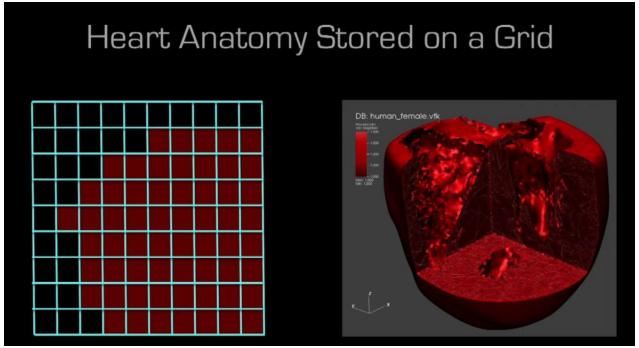
Summary of SC20 + What We Learned



- SC20 online event
  - Worked with Microsoft Azure
    - AMD EPYC 2nd Gen (Rome), Nvidia V100
  - Lots of benchmark testing with multiple VM SKUs, but couldn't run most applications
    - HPL (High Performance LINPACK) - 65 TFlops
    - HPCG - 593 GFlops
    - IO500 - 0.7
    - Mystery Application 6/6
- Scripts to get benchmarks setup -> more time for applications
  - Submit something for everything
  - Better distribution of system resources to workloads

SC21 Info

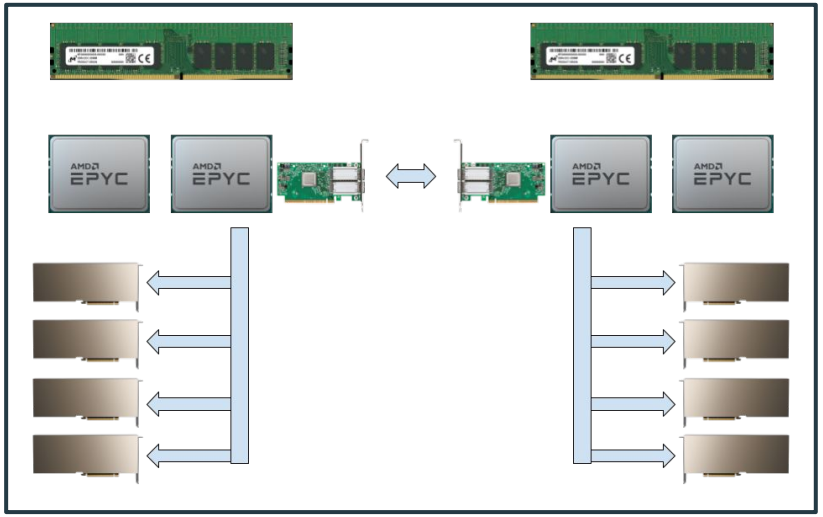
- In-person competition
- Same benchmarks as previous year
  - LINPACK, HPCG, IO500 - team has experience
- New applications
  - Cardioid
  - Quantum Espresso
- Variable power limit
  - 2000W - 4000W



https://github.com/LLNL/cardioid - "Cardiac multiscale simulation suite"

https://www.quantum-espresso.org/ - "An integrated suite of Open-Source computer codes for electronic-structure calculations and materials modeling at the nanoscale."

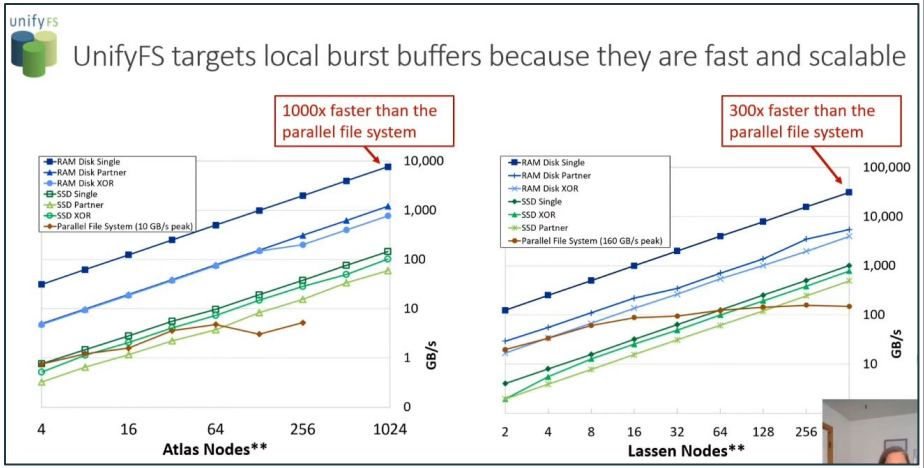
Architecture Proposal



- Hardware
  - 2 nodes
    - 2x AMD EPYC Milan
    - 4x Nvidia A100
    - 32GB DDR4 ECC/socket - 64GB/node
    - Infiniband
- Software
  - Distributed File System - GekkoFS, UnifyFS
  - AMD Optimized C/C++ Compiler <-> Nvidia CUDA-X libraries
    - Provides optimized support for workload dependencies

Distributed File Systems

- Local Node Storage Burst Buffers
  - Large increase in performance vs. parallel file systems
    - No contention!
    - Scales very well
  - Not so good when files need to be shared or producer/consumer applications (CESM)
  - GekkoFS, UnifyFS



Kathryn Mohror (LLNL), *UnifyFS: A filesystem for burst buffers*

Nvidia A100

- Reigning champion for HPC applications
- In comparison with the V100...
  - Per core clock ⬇ core count ⬆
  - More memory, higher memory bandwidth and memory clock speed
- Running HPL on GPU instead of CPU
  - Why didn’t we do this last year?

AMD EPYC Milan

Up to 64 processor cores per socket, Improved CPU speed up to 3.7 GHz, and more ….

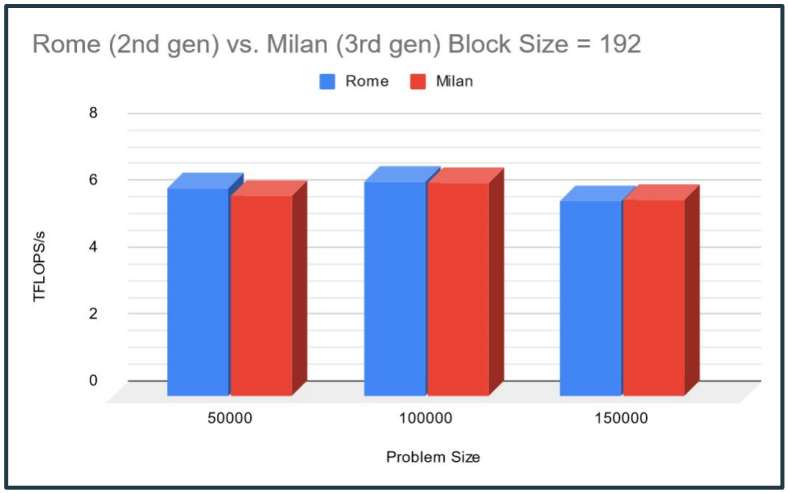
L3 Unified Cache:

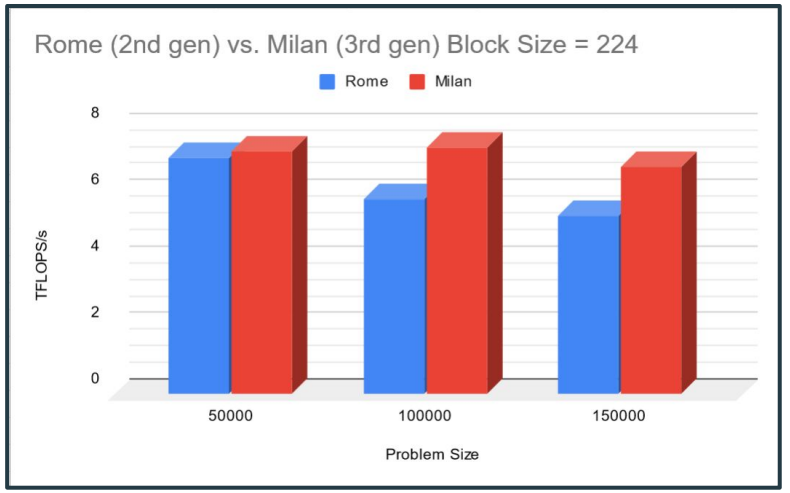
- Rome’s L3 cache boundary is every 4 cores and 16 MB
- Milan’s L3 cache boundary is every 8 cores and 32 MB

https://techcommunity.microsoft.com/t5/azure-compute/hpc-performance-and-scalability-results-with-azure-hbv3-vms/ba-p/2206471

https://www.microway.com/knowledge-center-articles/detailed-specifications-of-the-amd-epyc-milan-cpus/

Benchmarks!!!

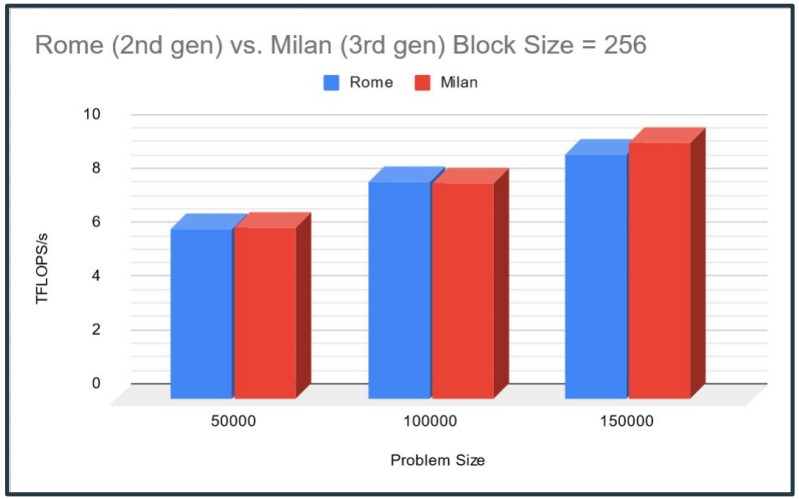




5-node clusters, 120 cores/node, basic techniques, openmpi-4.0.5, P=24 Q=25

Benchmarks!!!

On the Azure VMs, Milan has half as many blocks, 2 times as many cores/boundary and 2 times as much L3 cache per boundary! This decreases the probability of cache misses significantly.




```
The matrix A is randomly generated for each test.
The following scaled residual check will be computed:
    ||Ax-b||_oo / ( eps * ( || x ||_oo * || A ||_oo + || b ||_oo ) * N )
The relative machine precision (eps) is taken to be            1.110223e-16
Computational tests pass if scaled residuals are less than                16.0

T/V                N    NB     P     Q               Time                 Gflops
--------------------------------------------------------------------------------
WR11C2R4      200000   256    24    25             534.09             9.9859e+03
HPL_pdgesv() start time Fri Apr  9 06:26:36 2021

HPL_pdgesv() end time   Fri Apr  9 06:35:30 2021

--------------------------------------------------------------------------------
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)=   1.71190436e-03 ..... PASSED
================================================================================
```

5-node clusters, 120 cores/node, basic techniques, openmpi-4.0.5, P=24 Q=25

Benchmark - Just Numbers

Block Size	Problem Size	Rome	Milan
192	50000	6.23	6.01
192	100000	6.42	6.39
192	150000	5.85	5.86
Block Size		Rome	Milan
224	50000	7.1	7.33
224	100000	5.87	7.44
224	150000	5.37	6.86
Block Size		Rome	Milan
256	50000	6.32	6.39
256	100000	8.09	8.03
256	150000	9.11	9.55

5-node clusters, 120 cores/node, basic techniques, openmpi-4.0.5, P=24 Q=25

Acknowledgements

Microway

BU Ignite Council

Kurt Keville

SC21 Organization Team

MGHPCC