OAK RIDGE National Laboratory

Balancing Load in More Ways than One

Verónica G. Melesse Vergara
Paul Peltz
Nick Hagerty
Christopher Zimmer
Reuben Budiardja
Dan Dietz
Thomas Papatheodore
Christopher Coffman
Benton Sparks

ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

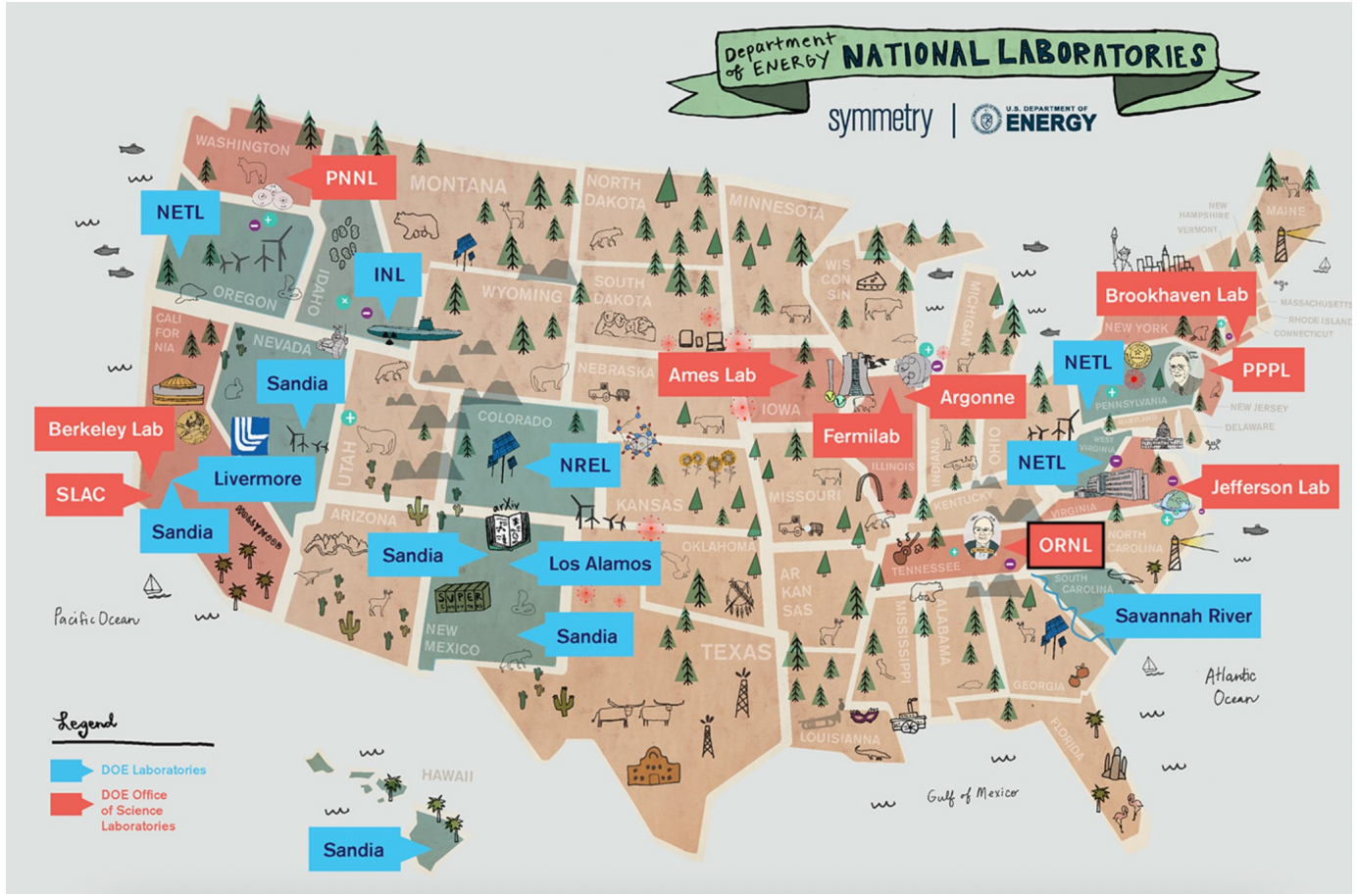
Outline

- DOE National Labs
  - ORNL
  - Leadership Computing Facilities
  - Computing at ORNL
- NCRC Partnership
  - NCRC Systems
- C5 Deployment
  - Acceptance Testing (AT)
  - OLCF Test Harness
  - AT Results
- Challenges and Lessons Learned
- Scaling study of OLCF applications
- Conclusions & Future Work

OAK RIDGE National Laboratory

DOE National Laboratories

U.S. Department of Energy National Laboratories in a nutshell



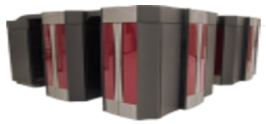
- The U.S. Department of Energy (DOE) has 17 national laboratories across the country
- Two separate offices:
  - National Nuclear Security Administration (NNSA)
  - Office of Science (SC)
- The national laboratories complex tackles critical scientific challenges of our time
- Serve as the leading institutions for scientific innovation
- Provide advanced instruments and user facilities to the scientific community across the globe

Leadership Computing Facilities

- Collaborative, multi-lab DOE initiative (2 centers / 2 architectures)
- Mission: Provide an ecosystem that enables capability computing opportunities to solve the most challenging problems.
- Administer and support two highly competitive user allocation programs
  - Innovative and Novel Computational Impact on Theory and Experiment (INCITE)
  - ASCR Leadership Computing Challenge (ALCC)
  - Computational allocations typically 100x larger than generally available in university, laboratory, and industrial (scientific and engineering) environments.

Oak Ridge Leadership Computing Facility

- OLCF part of the National Center for Computational Sciences (NCCS) has successfully delivered seven leadership-class systems since 2004
- Frontier is system number seven and provides an increased capability of over 80,000x
- Large part of success has been strong user partnerships to scale & refactor codes/methods
- Partnering has been essential to delivering science in a rapidly changing computational environment



Phoenix X1
• X1e

2004

Jaguar XT3
• Dual core upgrade

2005

Jaguar XT4
• Quad core upgrade

2007

Jaguar XT5
• 6 core upgrade
• 16 core upgrade

2008

Titan XK7
• GPU upgrade

2012

Summit

2018

Frontier

2022

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

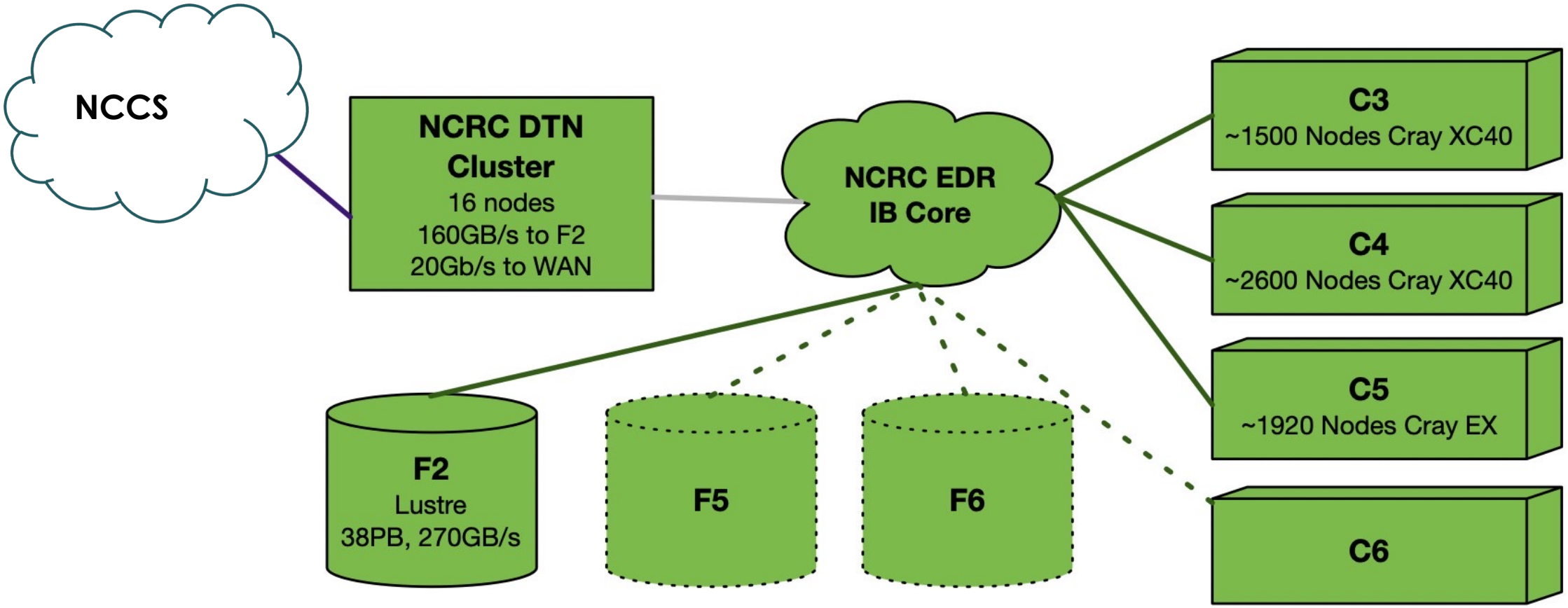
NCRC Partnership

But NCCS supports multiple programs!

National Climate-Computing Research Center

- The NCRC is a strategic partnership between U.S. Department of Energy and the National Oceanic and Atmospheric Administration (NOAA)
- Managed and operated by the NCCS at ORNL since 2009
- NCRC has deployed 4 compute systems since the start of the partnership and 4 file systems
- Most recent compute system deployment is C5

NCRC Compute and Storage Resources



OAK RIDGE National Laboratory

C5 Deployment

C5 Deployment

Phase I: August 2022

1,728 compute nodes

- Two 64-core AMD EPYC 7H12 CPUs
- 256 GB RAM
- Two Mellanox ConnectX-5 NICs

Phase II: February 2023

Expanded the system to 1,920 compute nodes

C5 Acceptance Testing

Vendor Test

POST

Per-node health check:

- HPL
- Stream

Vendor-provided network health checks

Contractual benchmarks execution

Functionality Test

System Administration

Reliability and Serviceability

Network Health

Programming Environment

Performance Test

NCRC Benchmarks

- CM4
- ESM4
- SHiELD
- Spear
- UFS

Stability Test

Continuous execution of simulated realistic workload using OLCF Test Harness

Completion criteria:

- All jobs that complete must produce correct answers
- Must perform within predetermined threshold of expected runtime variability

OLCF Test Harness

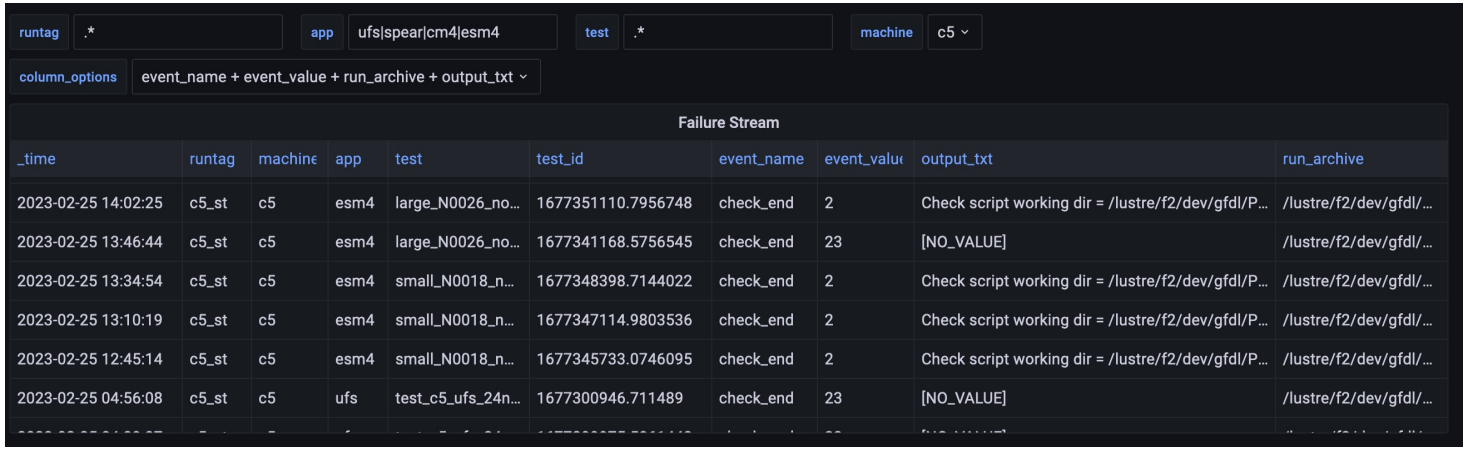
https://github.com/olcf/olcf-test-harness

- Developed in 2007 for Jaguar's acceptance testing by Arnold Tharrington at ORNL
- Python-based framework to simulate user workloads on an HPC system such that they are:
  - launched and executed in the same way a user would execute them
  - run continuously on the system without requiring manual actions by staff
  - uniquely identified and tracked
  - easily expanded with additional workloads due to requiring a low-level of effort
- Recently upgraded in preparation for the next release

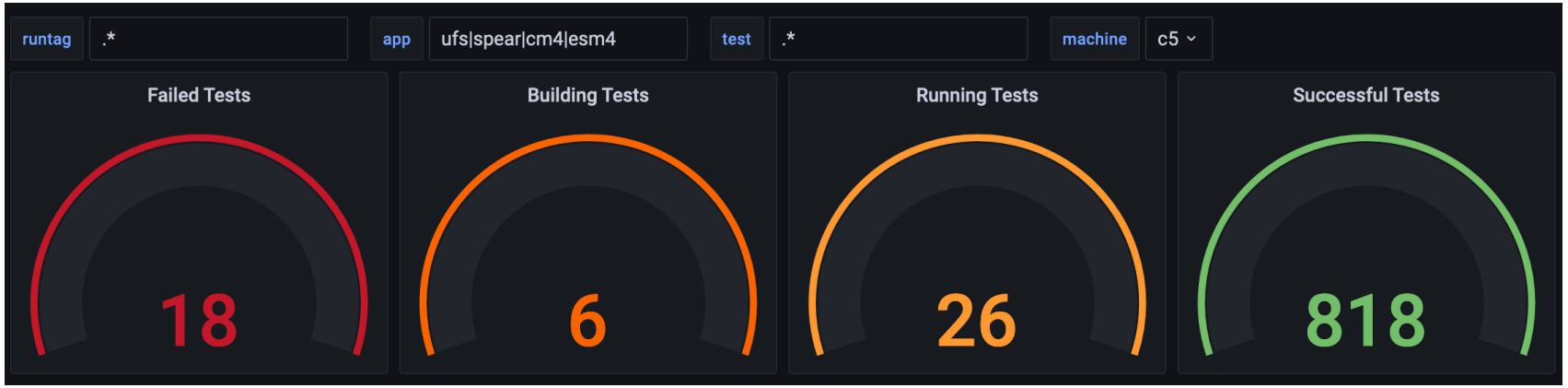
OLCF Test Harness (cont'd)

https://github.com/olcf/olcf-test-harness

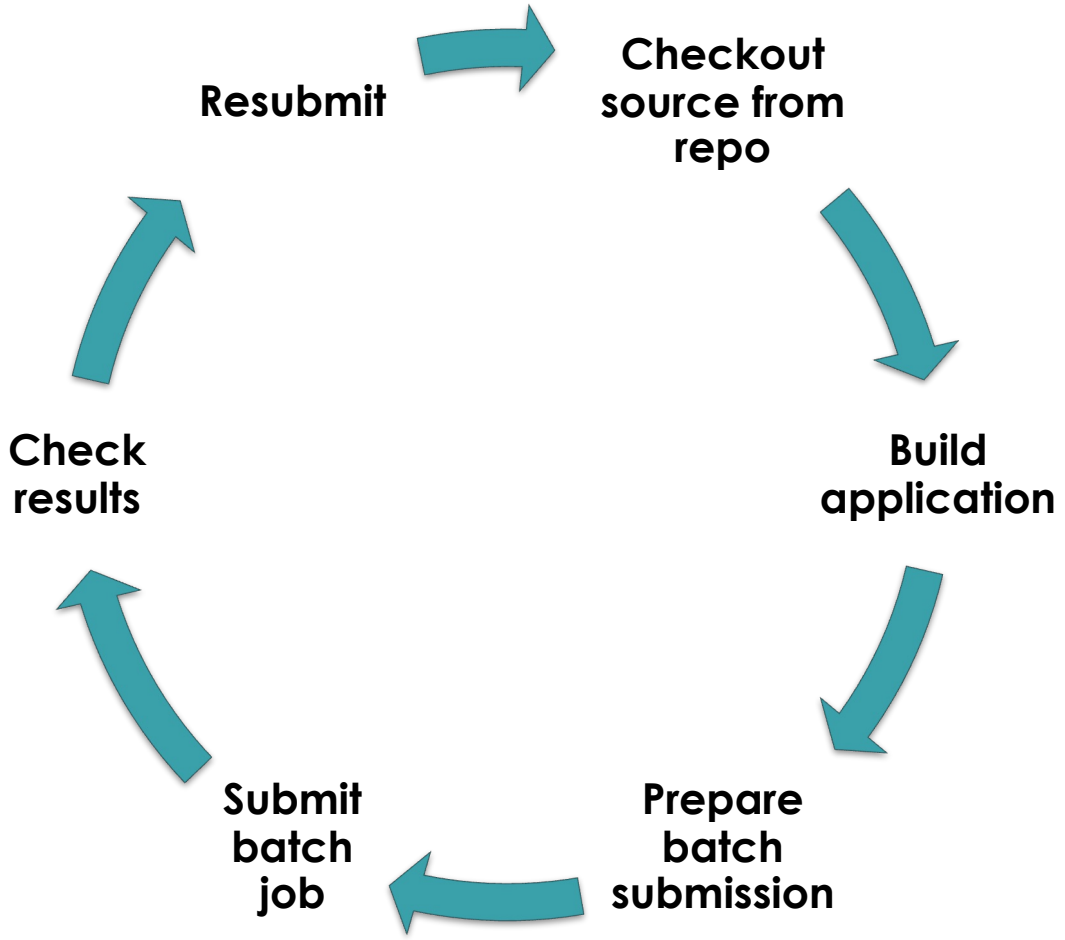
- Includes new features to improve:
  - Failure classification
  - Correctness checks
  - Performance checks
  - Additional logging
  - Monitoring using InfluxDB and Grafana



_time	runtag	machine	app	test	test_id	event_name	event_value	output_txt	run_archive
2023-02-25 14:02:25	c5_st	c5	esm4	large_N0026_no...	1677351110.7956748	check_end	2	Check script working dir = /lustre/f2/dev/gfdl/P...	/lustre/f2/dev/gfdl/...
2023-02-25 13:46:44	c5_st	c5	esm4	large_N0026_no...	1677341168.5756545	check_end	23	[NO_VALUE]	/lustre/f2/dev/gfdl/...
2023-02-25 13:34:54	c5_st	c5	esm4	small_N0018_n...	1677348398.7144022	check_end	2	Check script working dir = /lustre/f2/dev/gfdl/P...	/lustre/f2/dev/gfdl/...
2023-02-25 13:10:19	c5_st	c5	esm4	small_N0018_n...	1677347114.9803536	check_end	2	Check script working dir = /lustre/f2/dev/gfdl/P...	/lustre/f2/dev/gfdl/...
2023-02-25 12:45:14	c5_st	c5	esm4	small_N0018_n...	1677345733.0746095	check_end	2	Check script working dir = /lustre/f2/dev/gfdl/P...	/lustre/f2/dev/gfdl/...
2023-02-25 04:56:08	c5_st	c5	ufs	test_c5_ufs_24n...	1677300946.711489	check_end	23	[NO_VALUE]	/lustre/f2/dev/gfdl/...



OLCF Test Harness (cont'd)



OAK RIDGE National Laboratory

C5 Acceptance Test Results

C5 Acceptance Test Results

Hardware and System Administration

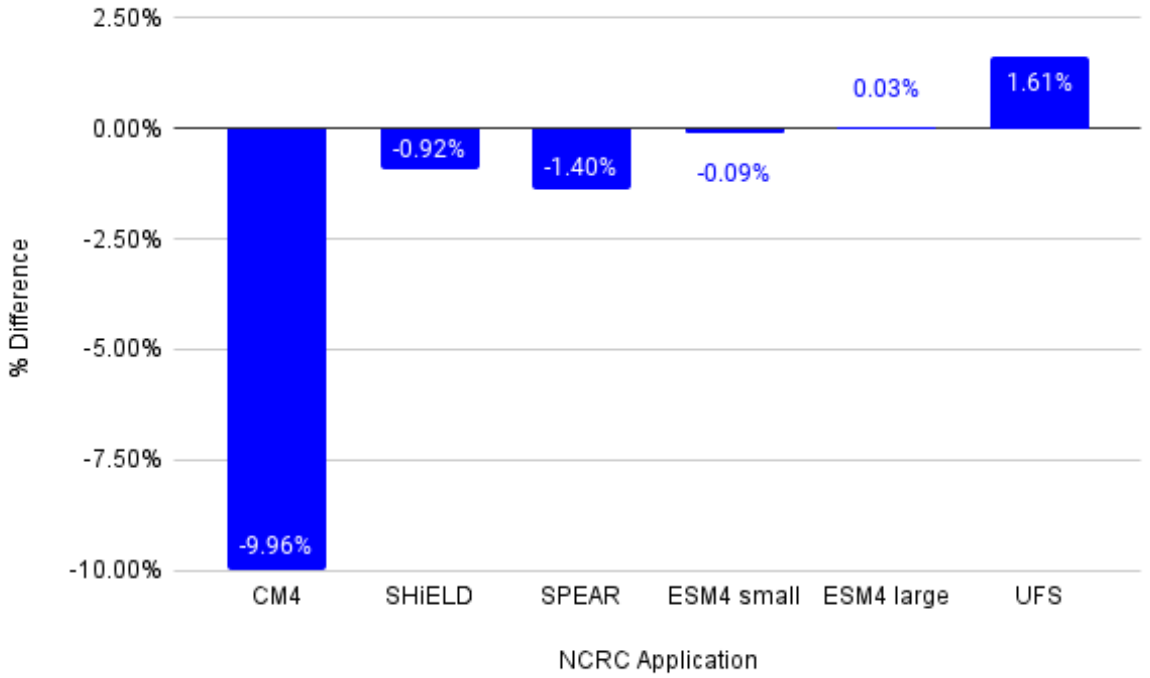
- Leveraged previous work done on ORNL's first HPE/Cray EX to accelerate testing timeline
- Encountered several issues with Lustre and the Slingshot fabric
  - Intermittent performance issues on F2
  - Nodes failing to mount F2
  - Observed DNE1 lock contention

Network Testing

- MPI Graph:
  - 1800 nodes with 4MB messages demonstrates
  - Ave: 8.6 GB/s/NIC
  - Max: 11.6 GB/s/NIC
- GPCNet:
  - 10 processes per node across 1800 nodes
  - C5 handles adversarial congestion scenarios well
  - Latency sensitive tests show virtually no impact when the all-to-all workload is running within the system

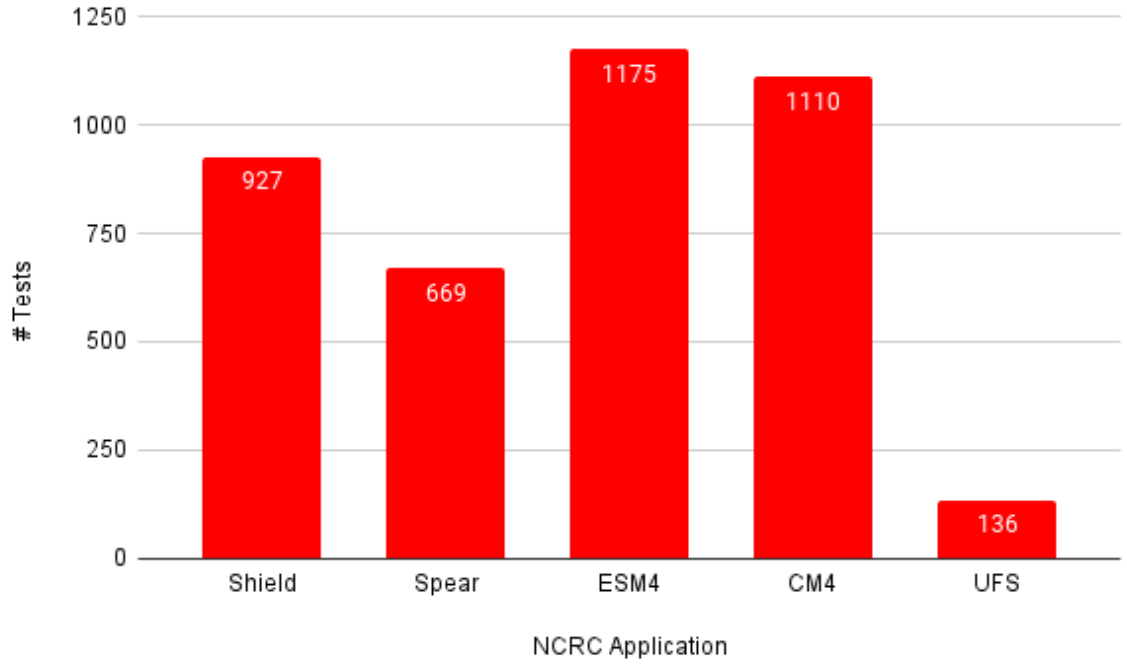
C5 Acceptance Test Results

- Performance testing
  - Contractual requirement
  - Two scenarios:
    - Single copies of each benchmark
    - Multiple copies of each to fill C5
  - HPE submitted results were compared to those obtained by ORNL
  - Multiple copies scenario proved to be challenging

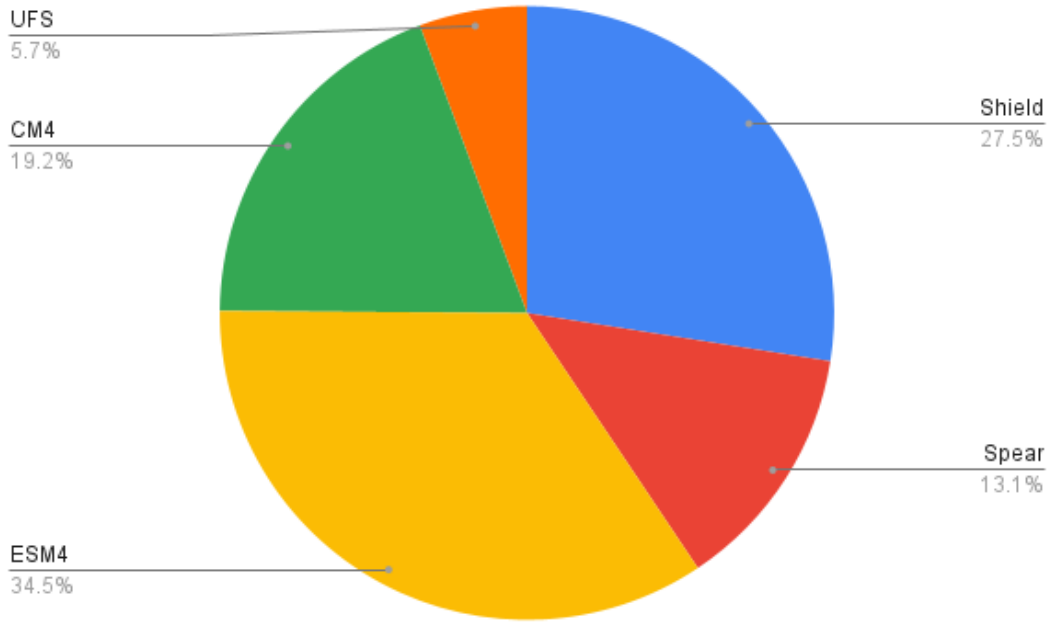


C5 Acceptance Test Results

- Stability testing included all NCRC benchmarks
- Ran for10-days continuously and executed >4,000 jobs



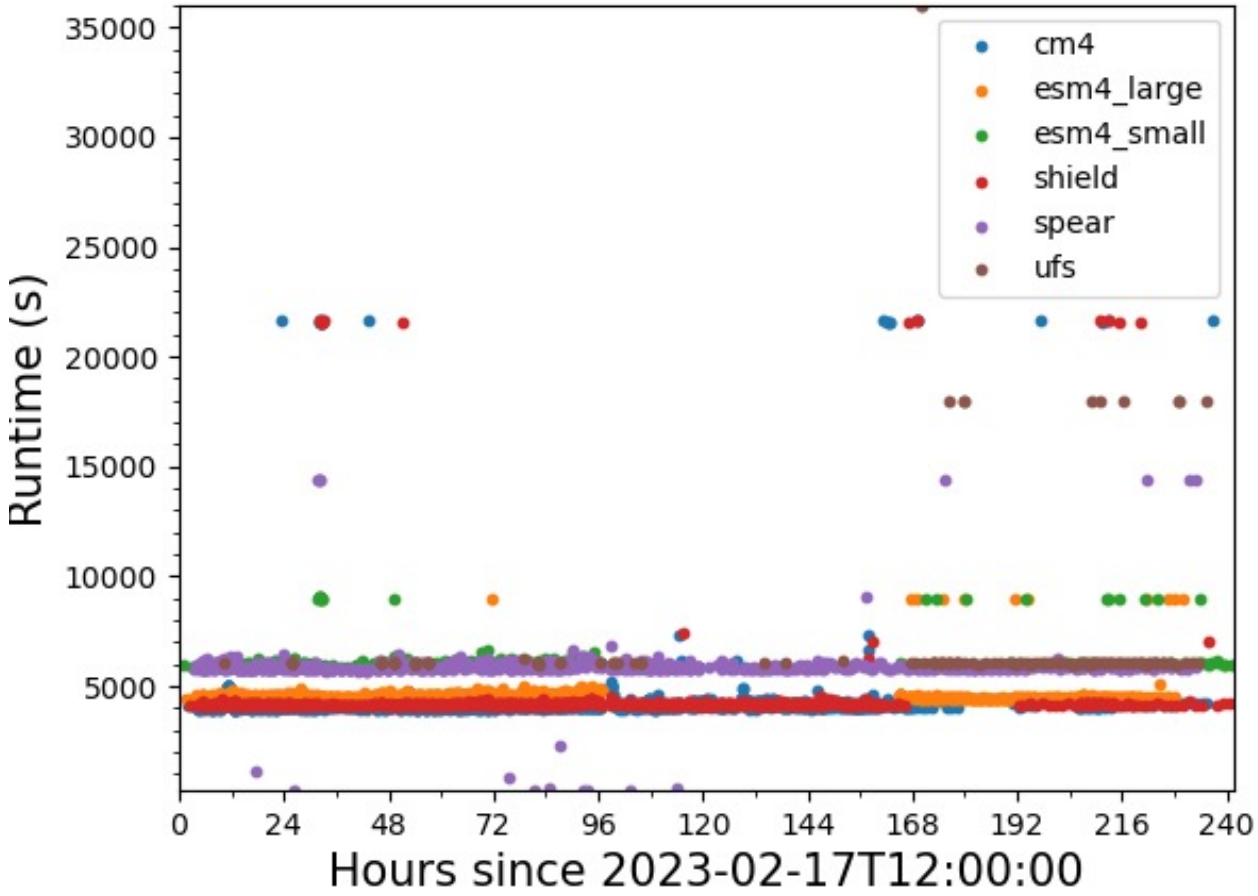
Jobs executed per Application



Failures per Application

C5 Acceptance Test Results (cont'd)

- Stability testing included:
  - Runtime variability analysis to understand performance of individual codes on a fully loaded system
  - Jobs exceeding 30% runtime variability criteria classified as performance failures
  - Some jobs saw performance 2X expected values



OAK RIDGE National Laboratory

Challenges and Lessons Learned

C5 Acceptance Testing Challenges

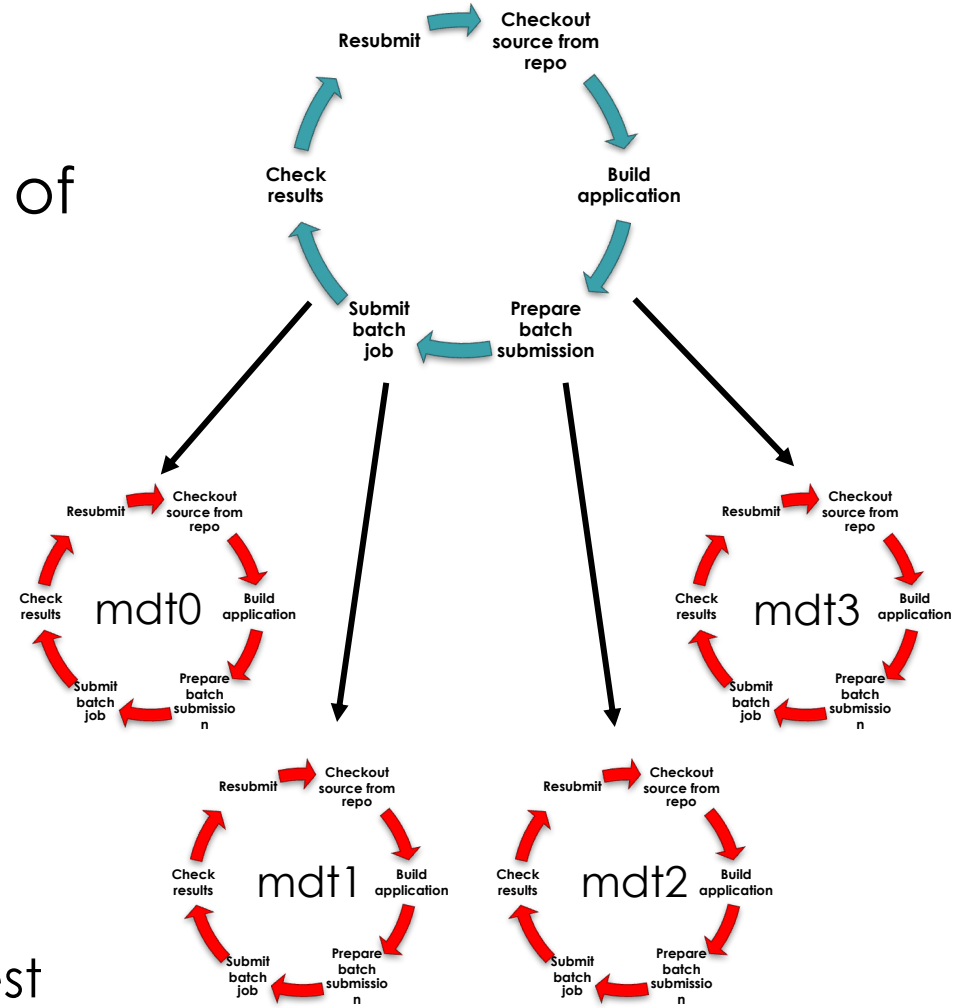
- Originally, planned to run all components and run acceptance as we always do
- But then… we began impacting production workloads running against F2
- Back to the drawing board…

C5 Acceptance Testing Challenges

- Concurrent application execution
  - During performance testing, we were required to fill C5 with multiple copies of each application
  - This resulted in hangs and timeouts for both pre-production and production workloads
  - Needed to find a balance between the number of benchmarks actively running on C5 and the load observed on F2 from C3 and C4
  - We split the workloads across multiple metadata targets (MDTs)
  - Split acceptance into subphases that could tolerate interruptions

C5 Acceptance Testing Challenges (cont'd)

- Impact to NCRC production workloads
  - Full workload launched in a single instance of the OTH triggered hangs and locking on F2
    - Switched to 4 instances of the OTH
    - Capped the number of copies per application
  - Determined that individual jobs were generating millions of file opens
    - Single large input deck was reused by many jobs
    - Moved to copying the input on a per-job basis
  - Avoided weeks with heavy production workloads
    - Required reordering of individual acceptance test components



C5 Acceptance Testing Challenges (cont'd)

- Long build times for NCRC applications
  - Part of the OTH workflow requires building each code per job
  - Not feasible for NCRC, since several builds exceeded an hour
    - Created variations of each job that reused a pre-built binary
- New compiler toolchains
  - NOAA leverages the Intel toolchain on C3 and C4
  - Initially explored using the latest Intel compilers provided by Intel oneAPI
  - Issues identified building codes with oneAPI:
    - Led us to recommend Intel classic as default Programming Environment
    - Provided a smoother transition to operations for NCRC users

OAK RIDGE National Laboratory

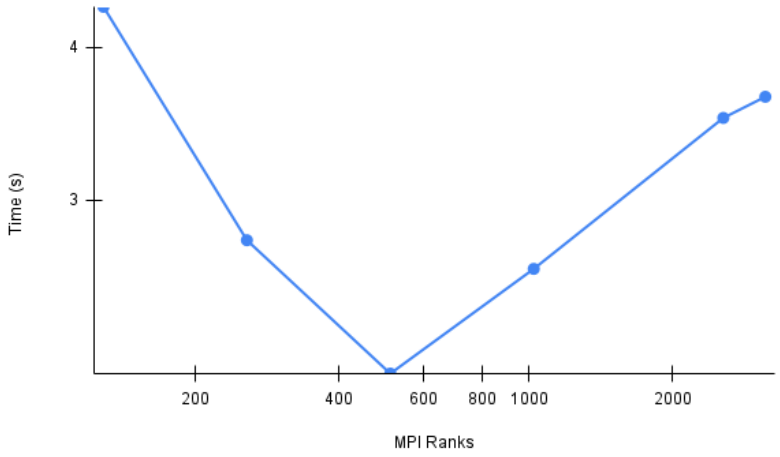
Scaling study of OLCF applications

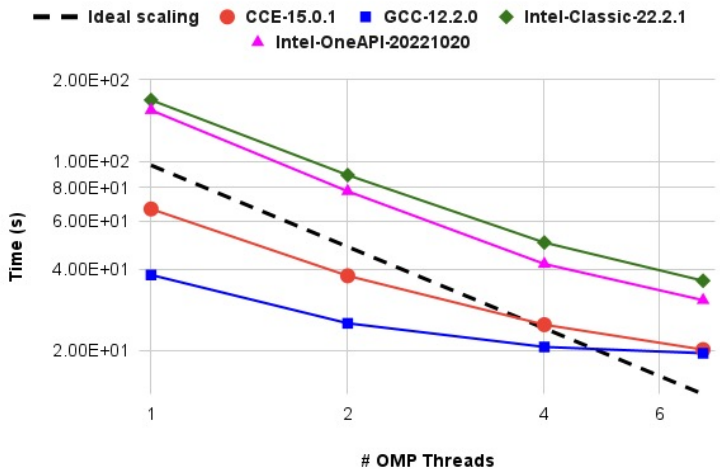
Scaling study of OLCF applications

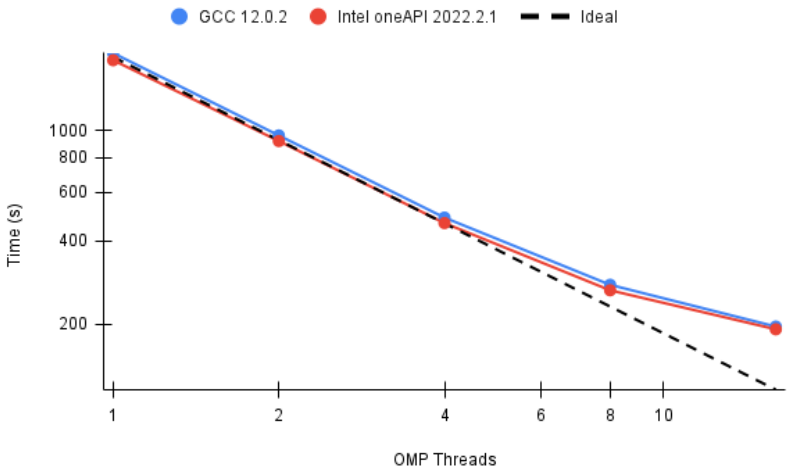
- To evaluate a system as broadly as possible, preferable to work with applications that are well understood internally
- Conducted scaling studies of 4 OLCF mini-applications and codes:
  - minisweep, GenASiS, LAMMPS, LSMS
- Utilized Intel (both intel-classic and intel-oneapi), CCE, and GNU toolchains

Scaling study of OLCF applications

- minisweep was utilized to run MPI-only experiments
- On both GenASiS and LSMS, we observed poor scaling beyond 4 OMP threads on a single-node
- Similar performance observed with GenASiS using Intel oneAPI vs. Intel Classic

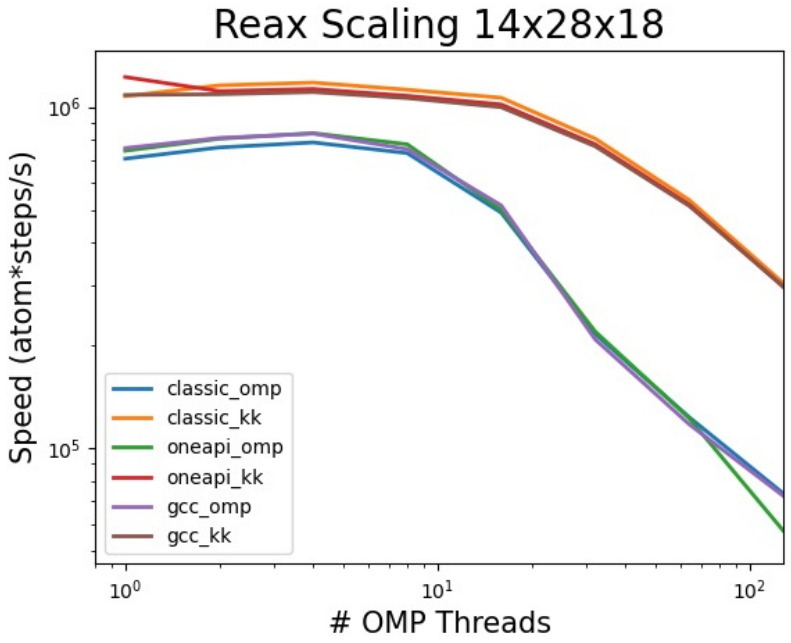


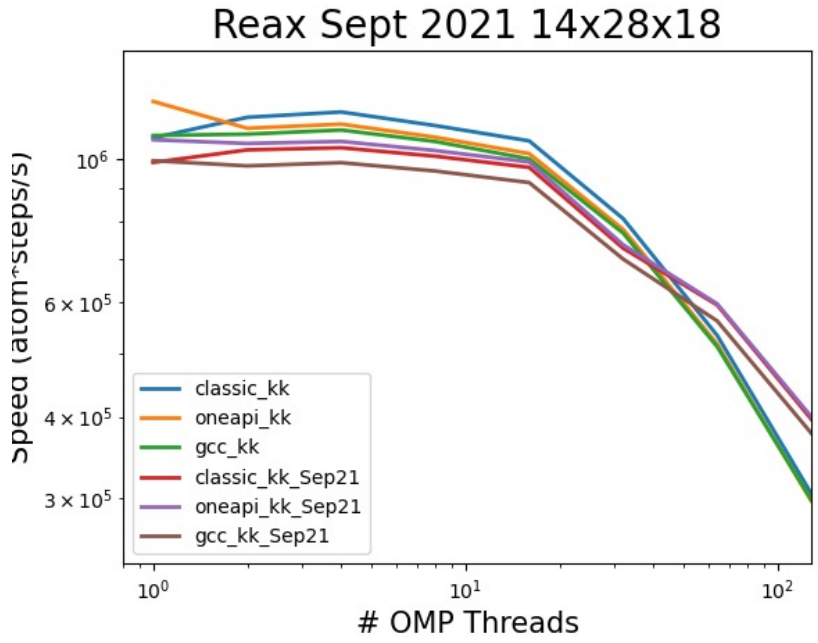




Scaling study of OLCF applications (cont'd)

- All compiler toolchains demonstrated similar scaling behavior for the ReaxFF benchmark using LAMMPS
- Kokkos package achieves better performance than the OpenMP package at all thread counts for all compiler toolchains
- Performance impact between February 2022 and September 2021 releases:
  - At the larger system size, February 2022 outperforms the September 2021 by about 10%
  - At 1 MPI rank / 128 OpenMP threads, September 2021 outperforms February 2022





Conclusions and Future Work

- Every acceptance testing has challenges
  - Both technical and project issues required creative solutions
- Worked closely with HPE and NOAA
  - Addressed critical issues identified preventing full simultaneous utilization of C3, C4, and C5 systems for production.
- Identified gaps in vendor-provided network diagnostic tools
  - Demonstrated with the network health issues identified using GPCNet and mpiGraph that were undetected via network diagnostics.
- Conducted a scalability study to provide broader understanding of the system
  - Highlighted observed differences in functionality and performance
  - Identified a couple of areas worth exploring further including the poor OpenMP scaling
- Having a great team willing to adapt on-the-fly and develop solutions was key to accomplish C5's successful deployment

OAK RIDGE National Laboratory

Acknowledgements

- We would like to also thank Chris Fuson, Don Maxwell, Matt Ezell, Jim Rogers, AJ Ruckman, Tori Robinson, Cathy Willis, Jeff Beckleheimer, Eric Dolven, Lisa Pallotti, Rusty Benson, Frank Indiviglio.
- This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

Interested in these topics and related evaluations?

- Submit your work to the first ever HPCTESTS 2023 Workshop
  - co-located with SC23 in Denver, CO

https://tinyurl.com/hpctests2023

Want to learn more about ORNL?

- Work with us – we are hiring!
  - HPC Engineer, System Acceptance & User Environment group: https://tinyurl.com/hpc-eng-ornl-2023
- Partner with us – apply for time on OLCF systems:
  - Pathways to Supercomputing Initiative: https://www.olcf.ornl.gov/community/pathways-to-supercomputing/
  - INCITE: https://www.anl.gov/article/us-department-of-energys-incite-program-seeks-proposals-for-2024-to-advance-science-and-engineering
- Connect with us:
  - https://twitter.com/OLCFGOV
  - https://www.linkedin.com/showcase/computing-at-ornl/


Questions?
vergaravg@ornl.gov
@verolero86