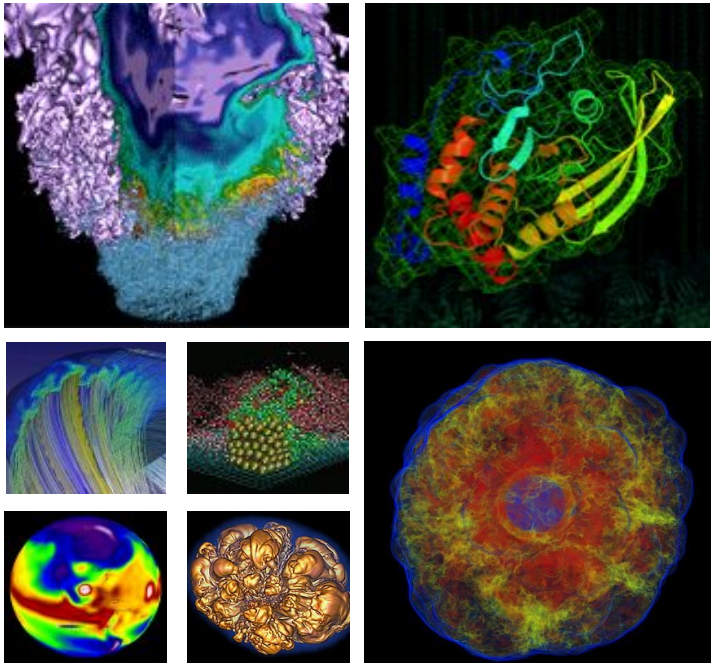
NERSC-9



Katie Antypas, Brian Austin, Brandon Cook

NERSC User Group

Nov. 1, 2018

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

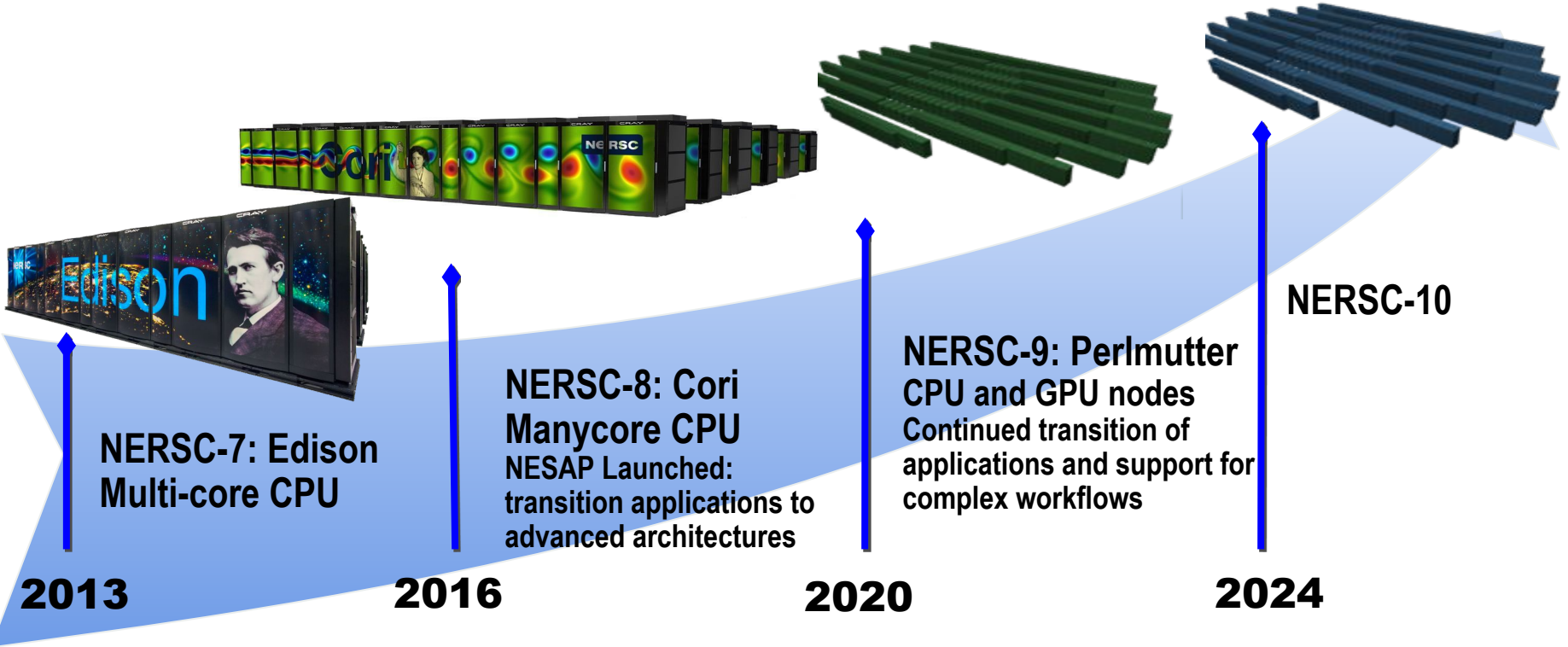
NERSC-9: A System Optimized for Science

- **Cray Shasta System providing 3-4x capability of Cori system**
- **First NERSC system designed to meet needs of both large scale simulation and data analysis from experimental facilities**
  - Includes both NVIDIA GPU-accelerated and AMD CPU-only nodes
  - Cray Slingshot high-performance network will support Terabit rate connections to system
  - Optimized data software stack enabling analytics and ML at scale
  - All-Flash filesystem for I/O acceleration
- **Robust readiness program for simulation, data and learning applications and complex workflows**
- **Delivery in late 2020**

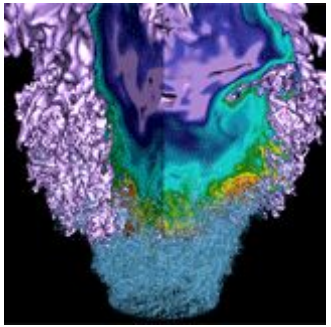
NERSC-9 will be named after Saul Perlmutter

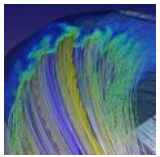
- Winner of 2011 Nobel Prize in Physics for discovery of the accelerating expansion of the universe.
- Supernova Cosmology Project, lead by Perlmutter, was a pioneer in using supercomputers combine large scale simulations with experimental data analysis
- Login “saul.nersc.gov”

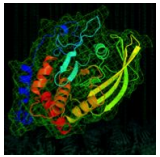
NERSC Systems Roadmap

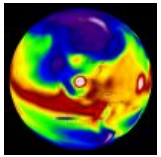


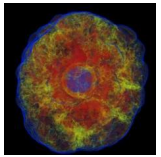
NERSC-9 Architecture

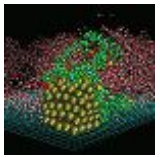












NERSC

From the start NERSC-9 had requirements of simulation and data users in mind

- All Flash file system for workflow acceleration
- Optimized network for data ingest from experimental facilities
- Real-time scheduling capabilities
- Supported analytics stack including latest ML/DL software
- System software supporting rolling upgrades for improved resilience
- Dedicated workflow management and interactive nodes

Exascale Requirements Reviews 2015-2018

First time users from DOE experimental facilities broadly included

NERSC

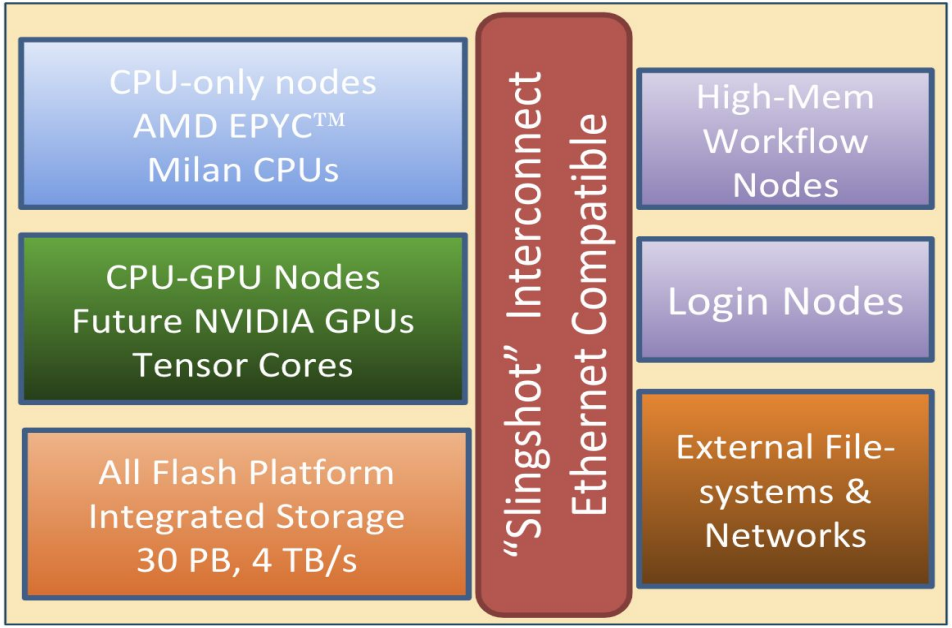
ENERGY

NERSC-9 System Optimized for Science

- In order to meet science requirements and mission need, accelerator technology was essential
- System contains large number of CPU-only nodes for applications that are not yet ready
- Significant fraction of NERSC workload can now use GPUs
  - GPU programming has matured
  - Improved software environment
  - Increases in GPU memory capacity improve programmability
- System well balanced between network and FLOPS
- All-flash filesystem for I/O acceleration

Perlmutter: A System Optimized for Science

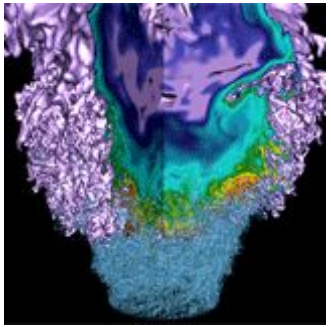
- GPU-accelerated and CPU-only nodes meet the needs of large scale simulation and data analysis from experimental facilities
- Cray “Slingshot” - High-performance, scalable, low-latency Ethernet-compatible network
- Single-tier All-Flash Lustre based HPC file system, 6x Cori’s bandwidth
- Dedicated login and high memory nodes to support complex workflows

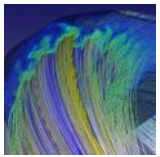
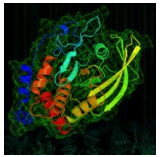
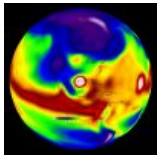
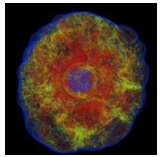
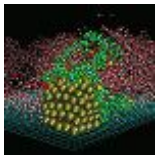


Compute Node Details

- **CPU only nodes**
  - Next Generation AMD CPUs
  - CPU only cabinets will provide approximately same capability as ***full*** Cori system (~8B hours) > 4000 nodes
  - Efforts to optimize codes for KNL will translate to NERSC-9 CPU only nodes
- **CPU + GPU nodes**
  - Next Generation NVIDIA GPUs with Tensor cores, high bandwidth memory and NVLINK-3
  - Unified Virtual Memory for improved programmability
  - 4 to 1 GPU to CPU ratio
  - (> 16B hours)

NERSC Exascale Science Application Program (NESAP)



NERSC

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Application Readiness for Perlmutter

- NERSC Exascale Science Application Program (NESAP)
- 25 Simulation, Data-analytics and Learning Applications
  - Additional Tier 2 applications
- up to 17 postdoctoral fellows
- Deep partnerships with every SC Office area
- Leverage vendor expertise and hack-a-thons
- Knowledge transfer through documentation and training for all users
- Optimize codes with improvements relevant to multiple architectures

Transitioning From KNL to AMD Processors

Codes optimized on Xeon Phi (KNL) will run well on Perlmutter

Many KNL architecture features are present on Perlmutter CPUs

- Many-Core
- MPI+OpenMP Programming Model Will Continue

Easier Onramp to "Many-Core" with Perlmutter CPUs than with KNL

- More Traditional Cores
- Single Memory Technology

GPU Transition Path for CPU Apps

NESAP for Perlmutter will extend activities from NESAP for Cori

1. Identifying and exploiting on-node parallelism - threads + vector
2. Understanding and improving data-locality within the cache-memory hierarchy

What's New?

1. Heterogeneous compute elements
2. Identification and exploitation of even more parallelism
3. Emphasis on performance-portable programming approach:
    - Continuity from Cori to NERSC-10

NERSC

NVIDIA

GPU Programming Models

We will support and engage our user community where their apps are:

CUDA: MILC, Chroma, HACC ...

CUDA FORTRAN: Quantum ESPRESSO, StarLord (AMREX)

OpenACC: VASP, E3SM, MPAS, GTC, XGC ...

Kokkos: LAMMPS, PELE, Chroma ...

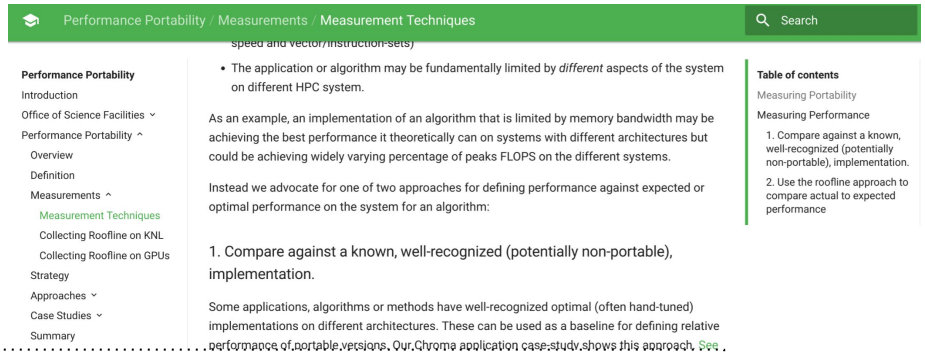
Raja: SW4

Engaging around Performance Portability

NERSC is working with PGI/NVIDIA to enable OpenMP GPU acceleration

NERSC Hosted Past C++ Summit and ISO C++ meeting on HPC.

NERSC Will Pursue Membership in 2018



NERSC is leading development of performanceportability.org

Doug Doerfler Leading Accepted Performance Portability Workshop at SC18. and 2019 DOE COE Perf. Port. Meeting

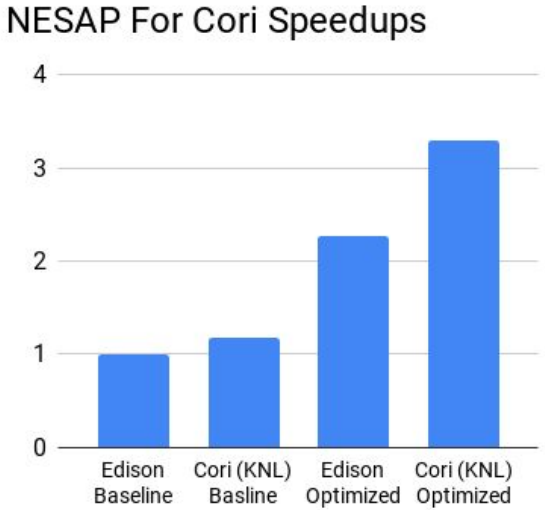
NESAP

Simulation ~12 Apps	Data Analysis ~8 Apps	Learning ~5 Apps

- 6 NESAP for Data apps will be continued. Additional apps focused on experimental facilities.
- 5 ECP Apps Jointly Selected (Participation Funded by ECP)
- Open call for proposals.Reviewed by a committee of NERSC staff, external reviewers **and input from DOE PMs.**
  - **App selection will contain multiple applications from each SC Office and algorithm area**
  - **Additional applications (beyond 25) will be selected for second tier NESAP with access to vendor/training resources and early access**

NESAP call for proposals

- Resources available to awardees
  - 1 Hackathon Session Per Quarter
    - NERSC, Cray, NVIDIA Engineer Attendance
  - Cray/NVIDIA Engineer Time Before and After Sessions
  - NESAP PostDocs (NERSC will hire up to 17)
  - NERSC Application Performance Specialist Attention
  - General Programming, Performance and Tools Training
  - Early Access (Perlmutter and GPU testbed)



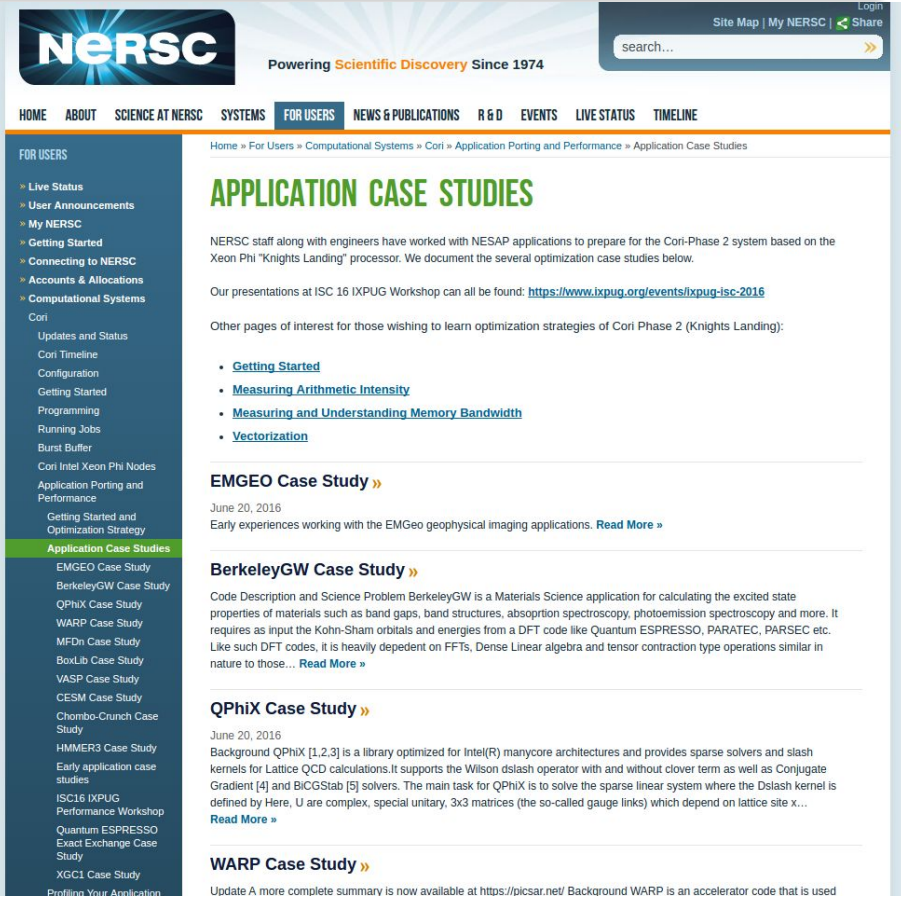
https://nersc.gov/users/application-performance/nesap/perlmutter/

Open Now through December 18, 2018 12:00 noon PST

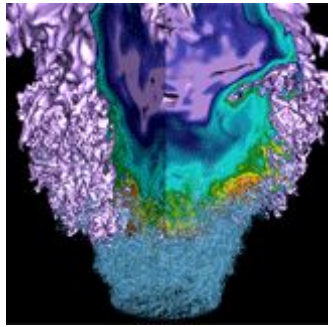
Selections announced in January 2019

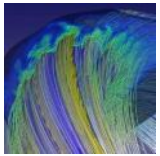
Training, Case Studies and Documentation

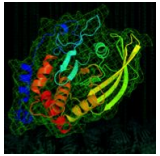
- For those teams not in NESAP, there will be a robust training program
- Lessons learned from deep dives from NESAP teams will be shared through case studies and documentation

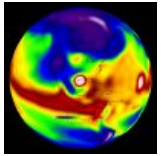


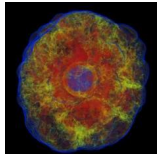
Wrap up

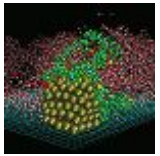












NERSC

NERSC-9 System Deployment Timeline

Milestone	Date
NESAP Call for Proposals	Nov. 2018
GPU Rack on Cori available for NESAP Users	Dec. 2018
NERSC-9 System Delivery	Oct. 2020
System Integration with NERSC Complete	Dec. 2020
Acceptance Testing Begins	Dec. 2020
NESAP Teams on NERSC-9 System	Jan. 2021
All users enabled on NERSC-9 System	Apr. 2021
System Acceptance	Aug. 2021

Summary

- Planning for NERSC-9 started in 2015. We are thrilled to have the contract signed and start putting plans into action
- We welcome any feedback, particularly in how to make NESAP even more successful