NERSC overview

Sudip Dosanjh
Director, NERSC

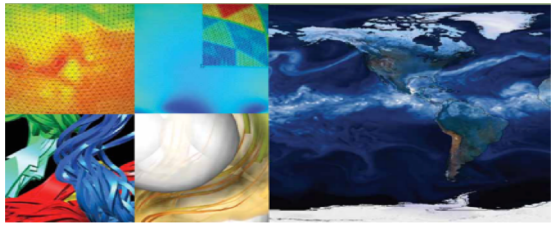
Outline

- NERSC Overview
- Operational highlights from a challenging year
- Perlmutter coming this year
- Expanded support for data

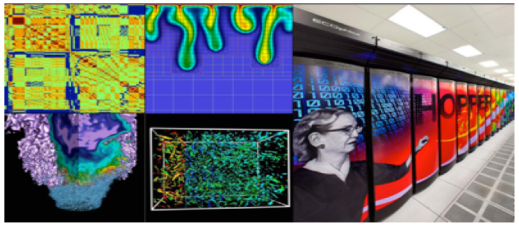
NERSC: the Mission HPC Facility for DOE Office of Science Research

U.S. DEPARTMENT OF ENERGY | Office of Science

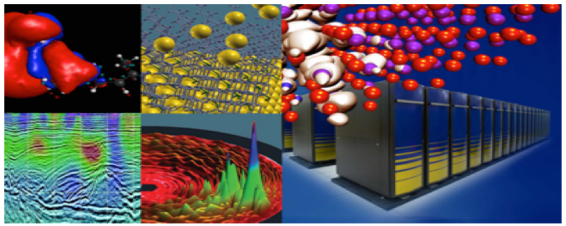
Largest funder of physical sciences research in the U.S.



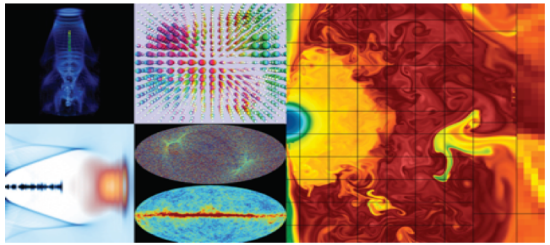
Bio Energy, Environment



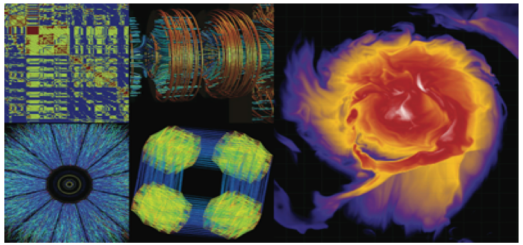
Computing



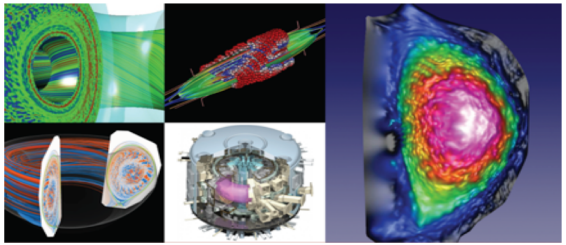
Materials, Chemistry, Geophysics



Particle Physics, Astrophysics

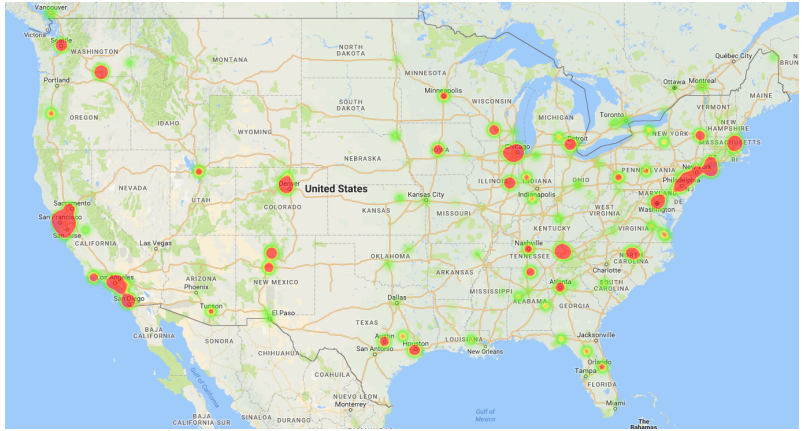


Nuclear Physics

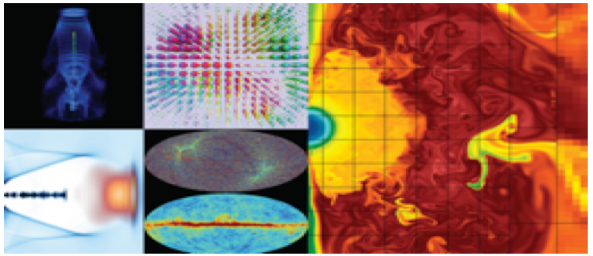


Fusion Energy, Plasma Physics

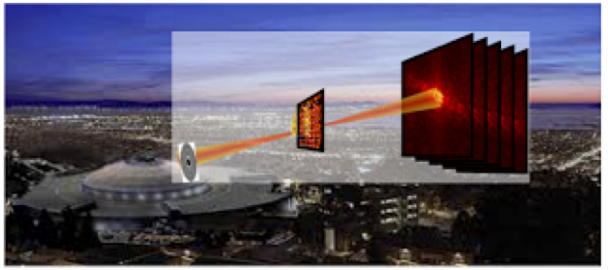
NERSC supports a broad user base



7,000 Users
800 Projects
700 Codes
~2000 publications per year



Simulations at scale

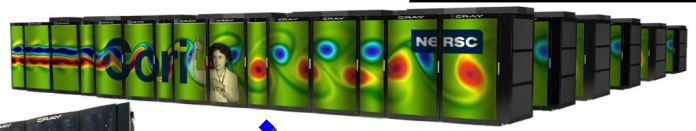


Data analysis support for DOE's experimental and observational facilities

NERSC Systems Roadmap

NERSC-7: Edison Multicore CPU

2013



NERSC-8: Cori Manycore CPU

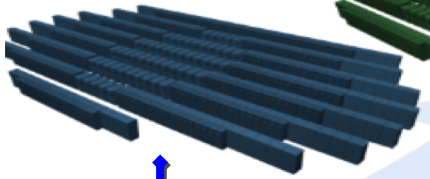
NESAP Launched: transition applications to advanced architectures

2016

NERSC-9: CPU and GPU nodes

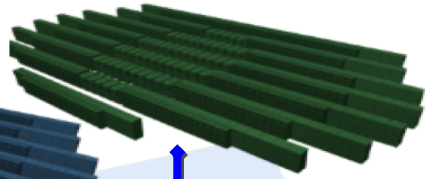
Continued transition of applications and support for complex workflows

2020



NERSC-10: Exa system

2024



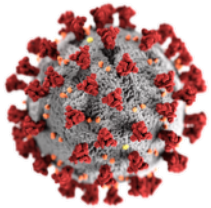
NERSC-11: Beyond Moore

2028

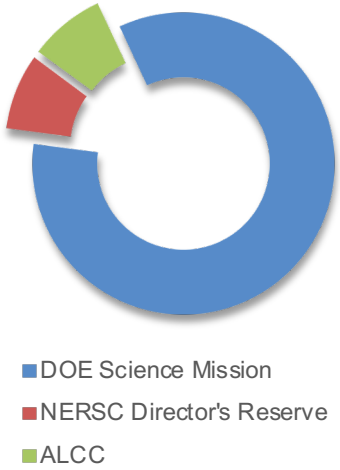
Operational Highlights from a Challenging Year

- NERSC kept systems up and running for users overcoming the challenge of shelter at home orders
  - Still performed maintenance, installed new hardware, responded to hardware failures
- NERSC Users Slack Channel
- MFA kept systems secure despite cyber attacks on HPC centers
- Utilization on Cori KNL improved by ~10% through a coordinated effort (>500M extra hours delivered)
  - Scheduler enhancements
  - Training and office hours
  - Policy changes

NERSC COVID-19 Support Overview

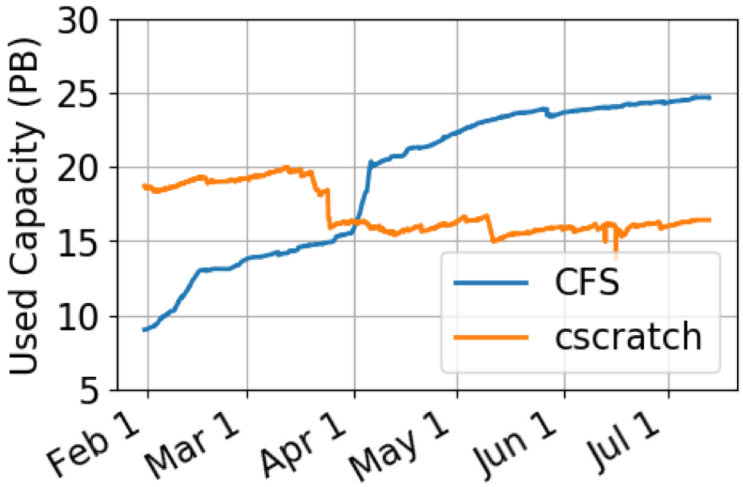


- Supporting COVID-19 Research
  - National COVID-19 HPC Consortium Member
  - ECP Project ExaLearn
  - Strategic collaborations
- Rapid project creation and staff technical liaisons
- Large-memory nodes added to Cori & Spin
  - 40 16 core AMD processors (3GHz, 2TB/node) added to Cori
  - 16 32 core processors (2.35GHz, 2TB/node) added to Spin
- Allocations from NERSC Director's Discretionary Reserve
  - 110 M Hours (1.5 M node hours)



Community File System Launched!

- **More space:** ~60PB
- **Default quota increases:** from 1TB to 20TB
- **Better quota management for subprojects in a repo:** PIs can request separate directories and have individual quotas for them
- **New allocation model:** Quotas are granted by DOE as part of the ERCAP process
- **File system features**
  - Faster rebuilds from distributed raid
  - End to end checksums (ensure data integrity)
  - Subblocks allow more efficient use of capacity especially for small files



Iris Launched in Dec. 2019, replacing 20 years with NIM

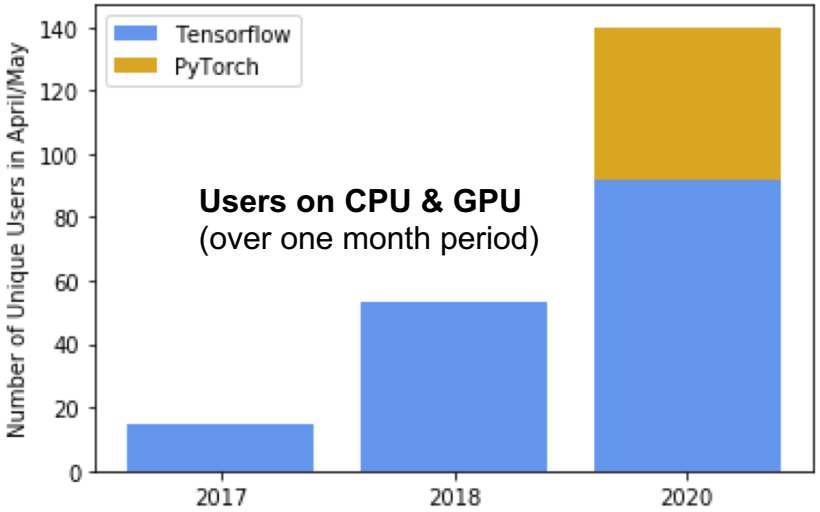
Benefits to users and staff

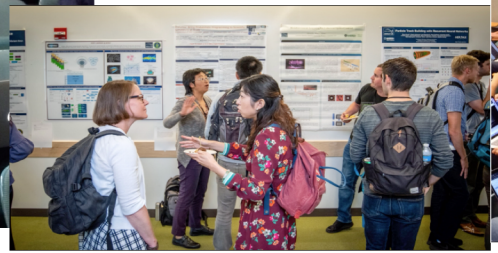
- Job-level accounting detail
- Responsive, intuitive UI
- More self-service features for users
- Easier reporting for staff and DOE
- Real-time updates
- Weekly (Agile) software releases

	Iris	NIM
Lines of custom code	42k	680k
Report speed*	500 ms	7.5 sec
AY Cutover tickets**	169	229

A growing ML workload at NERSC

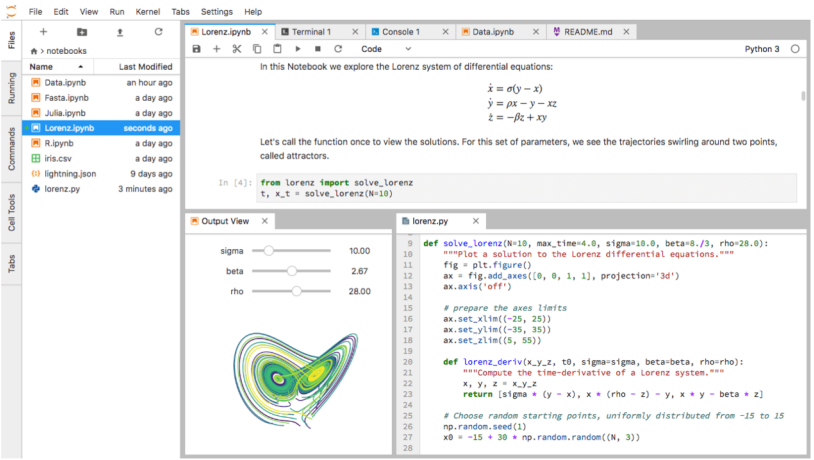
- **Outreach and Training events at NERSC and in the community:**
  - **ECP All Hands Meeting**
  - **SC 19 (and coming at SC20)**
  - **CUG'19 Tutorial**
  - **ISC Tutorial**
  - **Deep Learning for Science School**
  - **Monterey Data Conference**





Jupyter Notebooks for HPC

- **Jupyter growing in popularity at NERSC and broader community**
- **NERSC's goal - Enable exploratory data analytics, deep learning, workflows, through Jupyter on HPC systems**
- **Over 700 unique users at NERSC**
- **New features in 2019-2020:**
  - **Access to Cori compute nodes**
  - **Access to Cori GPUs**



Data 8: Foundations of Data Science, Fall 2018, Zellerbach Hall

Facility Upgrade Construction

- 12.5 MW Power/Cooling Expansion
  - 100% increase in B59 power
  - 167% increase in computer power
- New Scope:
  - 5 electrical substations
  - 3 cooling towers
  - 3 heat exchangers
  - 6 pumps
  - 3 air handling units
  - 29 electrical distribution panels
  - Enhanced backup power
  - Maintains energy efficiency

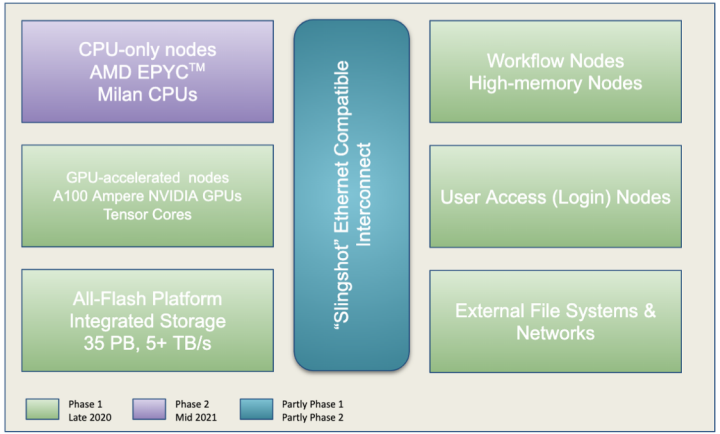



Outline

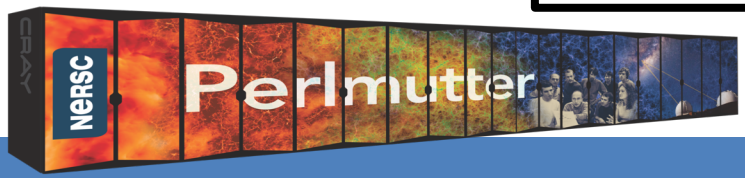
- NERSC Overview
- Operational highlights from a challenging year
- **Perlmutter coming this year!**
- Expanded support for data

Perlmutter: A System Optimized for Science

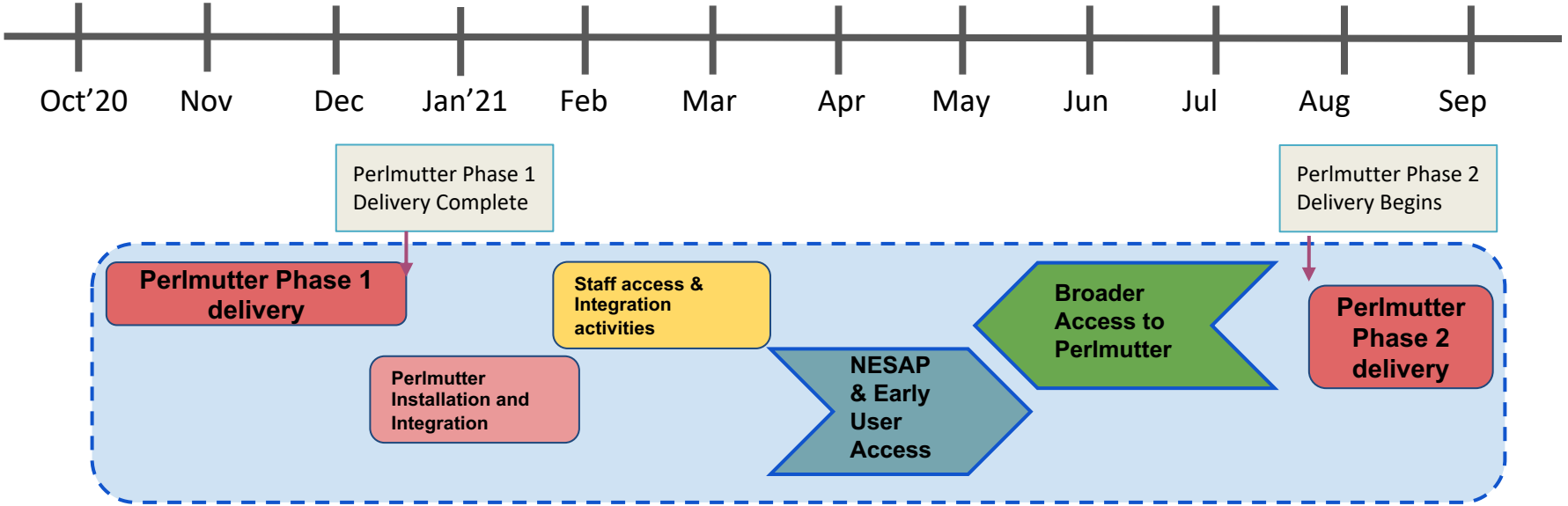
- Cray Shasta System providing 3-4x capability of Cori
- GPU-accelerated and CPU-only nodes meet the needs of large scale simulation and data analysis from experimental facilities
  - Large CPU-only partition providing capability similar to Cori
  - Support for complex workflows using compute, storage and networking resources
  - Optimized data software stack enabling analytics and ML at scale
- GPU nodes: 4 NVIDIA A100 “Ampere” GPUs each w/Tensor Cores & NVLink-3 and High-BW memory + 1 AMD “Milan” CPU
  - Over 6000 NVIDIA Volta-Next GPUs
  - Unified Virtual Memory support improves programmability
- Cray “Slingshot” - High-performance, scalable, low-latency Ethernet-compatible network
  - Capable of Terabit connections to/from the system
- Single-tier All-Flash Lustre based HPC file system
  - 6x Cori’s bandwidth
  - Cray’s ClusterStor E1000 system



Phased delivery
1st phase: End CY2020
2nd phase: Spring CY2021



Perlmutter Phased Timeline



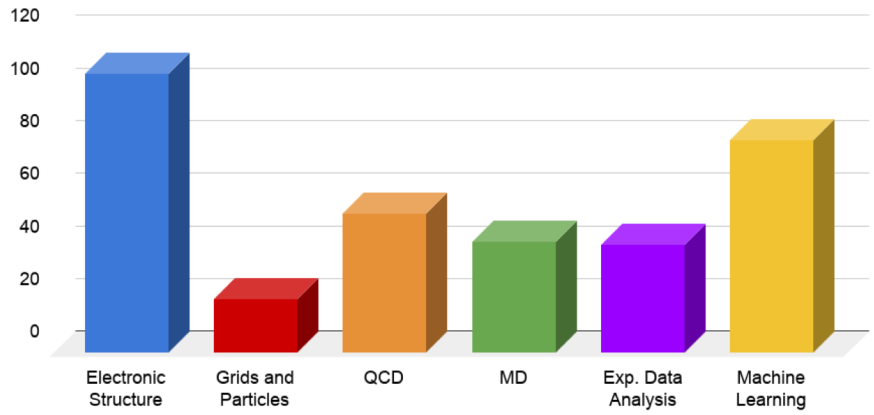
NESAP (NERSC Exascale Science Applications Program)

Goal:

Partner with Cray/NVIDIA and ~25 Teams (broad range across workload) at Deep Level to Prepare Apps for Perlmutter.

Disseminate Lessons Learned to NERSC Community Through Documentation, at Training Events and Community Hackathons

GPU Community Hackathons



Higher is Better

Projected Speedups on Perlmutter over Edison for Top NESAP Apps in Algorithmic Areas.

Includes Software Improvements from NESAP.

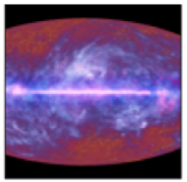
GPU For Science Days

Outline

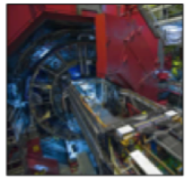
- NERSC Overview
- Operational highlights from a challenging year
- Perlmutter coming this year
- **Expanded support for data**

NERSC already supports a large number of users and projects from DOE SC's experimental and observational facilities

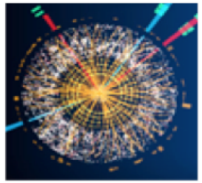
Palomar Transient Factory Supernova



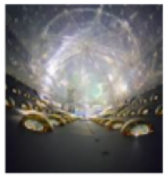
Planck Satellite Cosmic Microwave Background Radiation



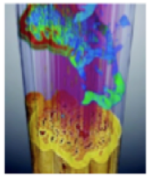
Alice Large Hadron Collider



Atlas Large Hadron Collider



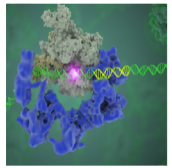
Dayabay Neutrinos



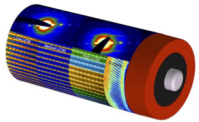
ALS Light Source

LCLS Light Source

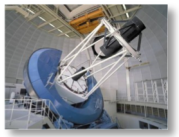
Joint Genome Institute Bioinformatics



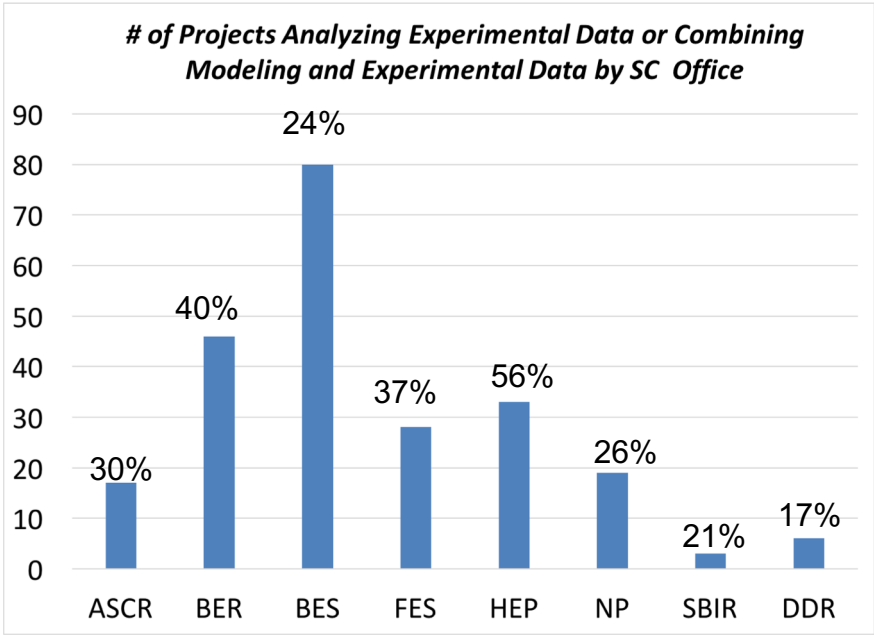
Cryo-EM



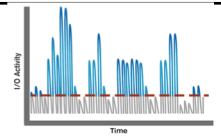
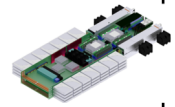
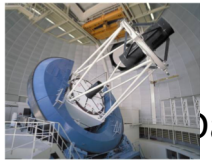
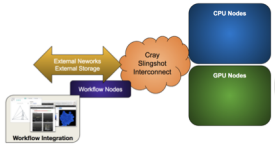
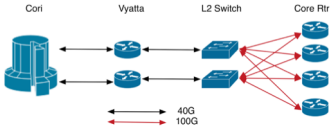
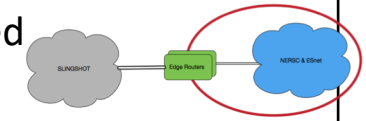
NCEM



DESI



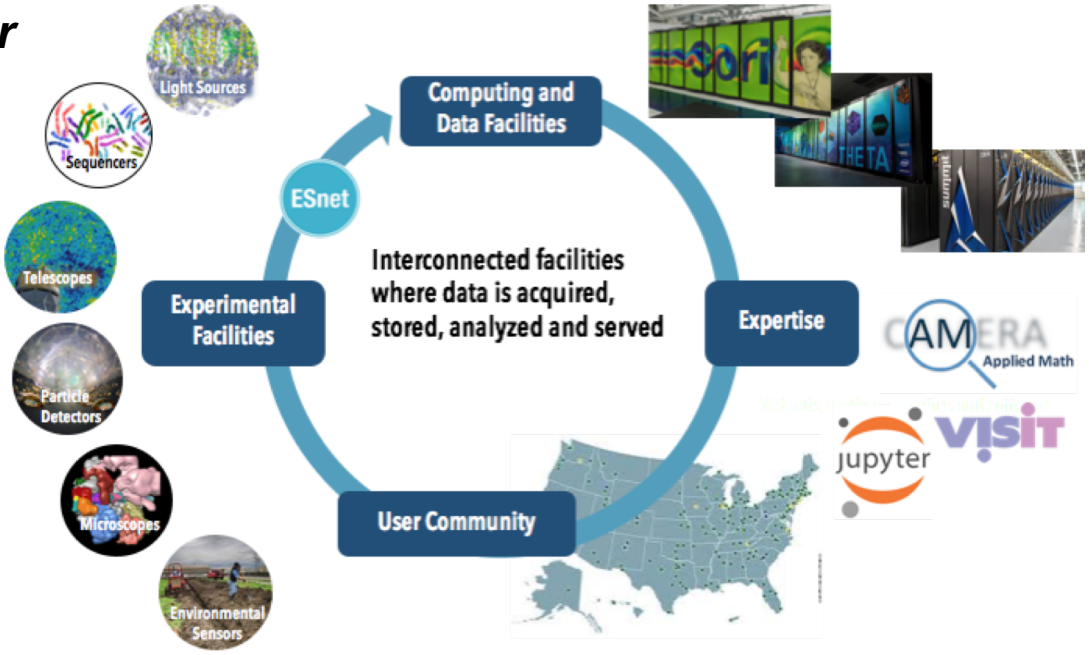
~35% (235) of ERCAP projects self identified as confirming the primary role of the project is to 1) analyze experimental data or; 2) create tools for experimental data analysis or; 3) combine experimental data with simulations and modeling

Data Features	Cori experience	N9 enhancements
I/O and Storage	Burst Buffer 	All-flash file system: performance with ease of data management 
Analytics - **Production stacks** - **Analytics libraries** - **Machine learning**	 User defined images with Shifter NESAP for data New analytics and ML libraries 	 Benchmark Production Analytics workflows. Data apps in NESAP at outset Optimised analytics libraries and deep learning application benchmarks
Workflow integration	 Real-time queues	 SLURM co-scheduling Workflow nodes integrated
Data transfer and streaming	SDN 	Slingshot ethernet-based converged fabric 

Superfacility: an ecosystem of connected facilities, software and expertise to enable new modes of discovery

Superfacility@ LBNL: *NERSC, ESnet and CRD working together*

- A model to integrate experimental, computational and networking facilities for reproducible science
- Enabling new discoveries by coupling experimental science with large scale data analysis and simulations



Science Engagements

ALS

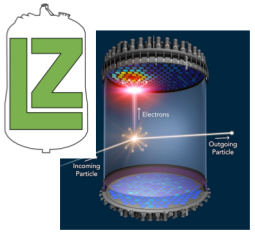
High-rate detectors use NERSC for real-time experimental feedback, data processing/management, and comparison to simulation

LSST DESC Dark Energy Science Collaboration

Processing streaming alerts (from NCSA) for detection of supernova and transient gravitational lensing events

LCLS

High-rate detectors use ESnet and NERSC for real-time experimental feedback and data processing



Next-generation dark matter detection, continuously sending data to NERSC and UK

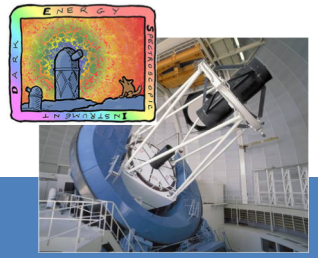
FICUS Facilities Integrating Collaborations for User Science

JGI-EMSL Collaborative Science Call

Complex multi-stage workflow to analyse response of soil microbes to climate change

MOLECULAR FOUNDRY

NCEM: 4D STEM data streamed to NERSC, used to design ML algorithm for future deployment on FPGAs close to detector



Nightly processing of galaxy spectra to inform next night's telescope targets

NUGEX Special Interest Group for Experimental Facilities

- Group formed in April, led by David Lawrence (JNL)
  - A user-organised forum for experimental scientists to discuss how they use NERSC and learn from each other's experience
- The SIG has held almost weekly meetings
  - Topics so far: Best Practices for Experimental Science at NERSC (Bjoern Enders), GlueEX experience @ NERSC (David Lawrence), How the NERSC scheduler works (Lisa Gerhardt), DESI experiment (Stephen Bailey), Workflow Managers at NERSC (Bill Arndt), The STAR Experiment @ NERSC (Michael Poat), Intro to ESnet (Eli Dart), ALICE (Jeff Porter), LCLS-II (Chris O'Grady), JGI Workloads (Bryce Foster)
- Also have a dedicated channel on the NERSC User Slack

Conclusions

- **NERSC has kept systems up and available despite numerous challenges**
- **We are very excited about Perlmutter Phase 1 coming this year**
  - Significant increase in capabilities over Cori
  - NESAP is helping users get ready for GPUs
  - 1st NERSC system designed with data in mind from the very beginning
    - All flash file system, new interconnect, big data stack
- **Demand from Experimental and Observational Facilities is increasing dramatically**
- **NERSC has made a significant investment in data, AI and deep learning**

Backup Slides