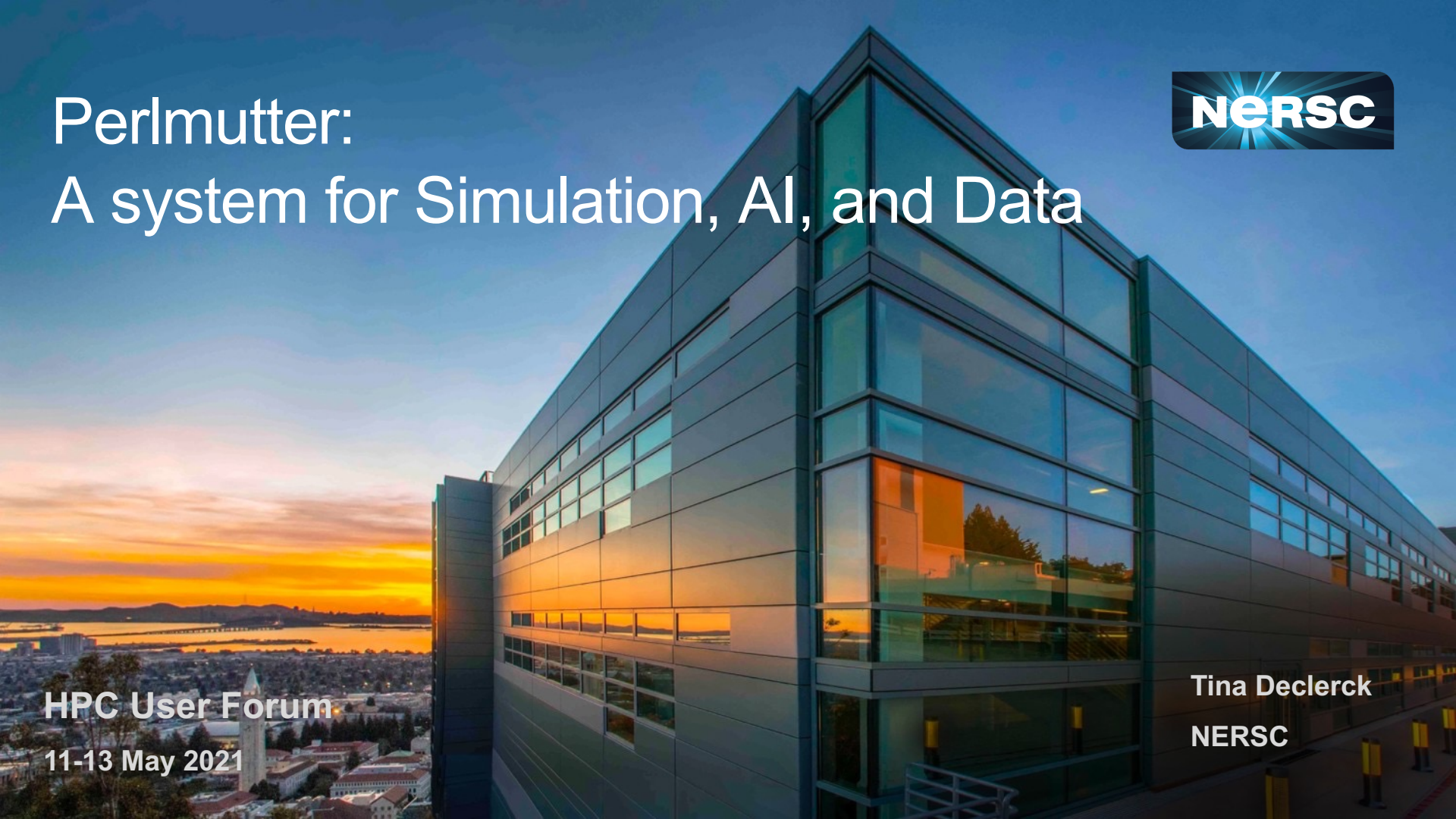
Perlmutter:
A system for Simulation, AI, and Data



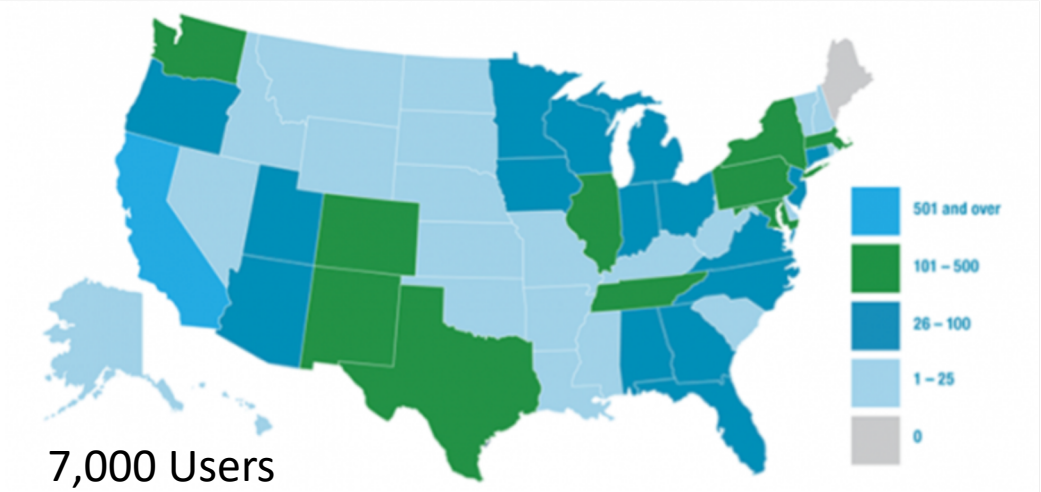
HPC User Forum

11-13 May 2021

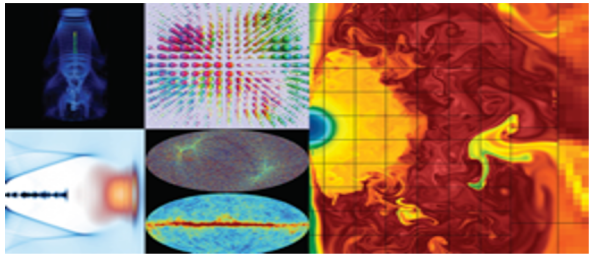
Tina Declerck

NERSC

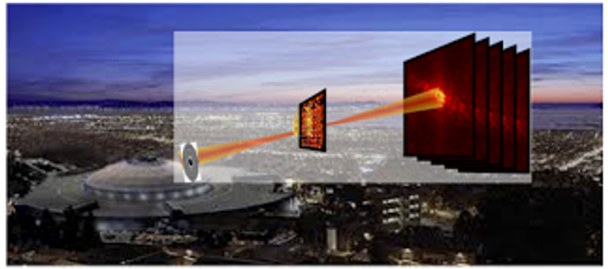
NERSC: Mission HPC facility for the DOE Office of Science



7,000 Users
800 Projects
700 Codes
50 States
40 Countries
~2000 publications per year



Simulations at scale

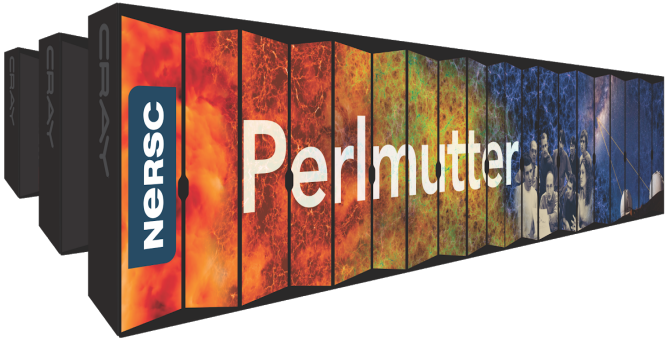


Data analysis support for DOE's experimental and observational facilities

Photo Credit: CAMERA

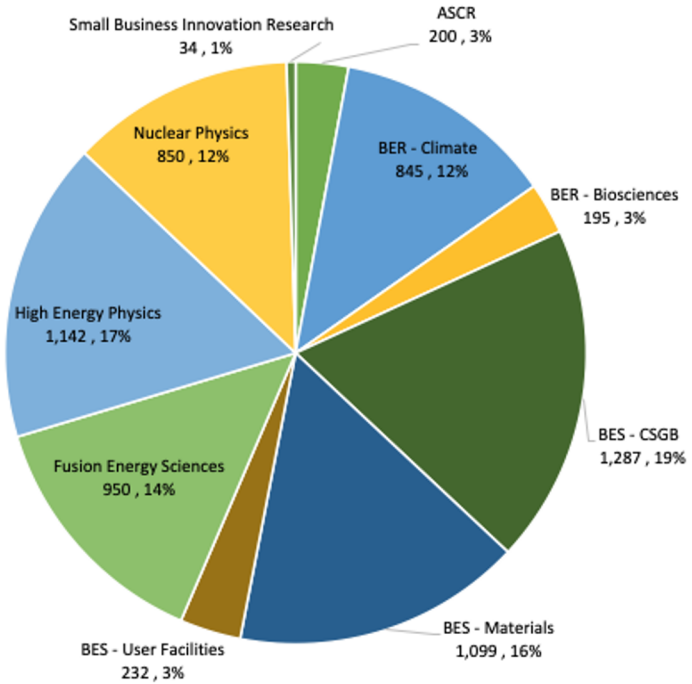
NERSC has a dual mission to advance science and the state-of-the-art in supercomputing

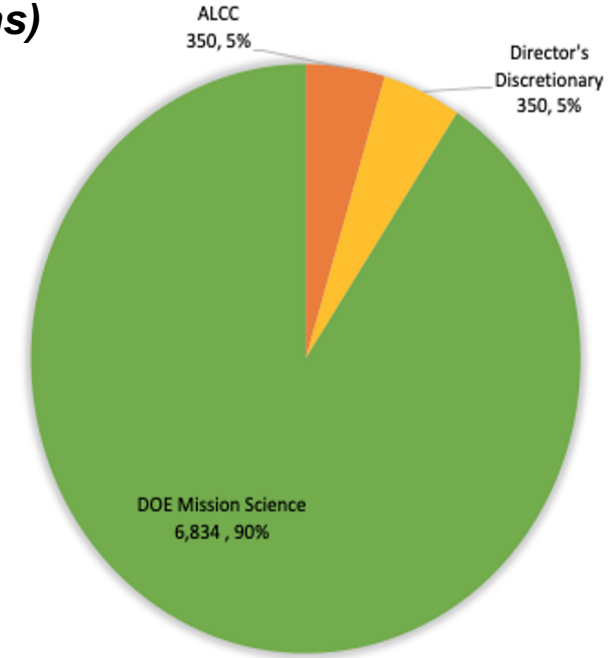
- We collaborate with computer companies years before a system's delivery to deploy advanced systems with new capabilities at large scale
- We provide a highly customized software and programming environment for science applications
- We are tightly coupled with the workflows of DOE's experimental and observational facilities – ingesting tens of terabytes of data each day
- Our staff provide advanced application and system performance expertise to users



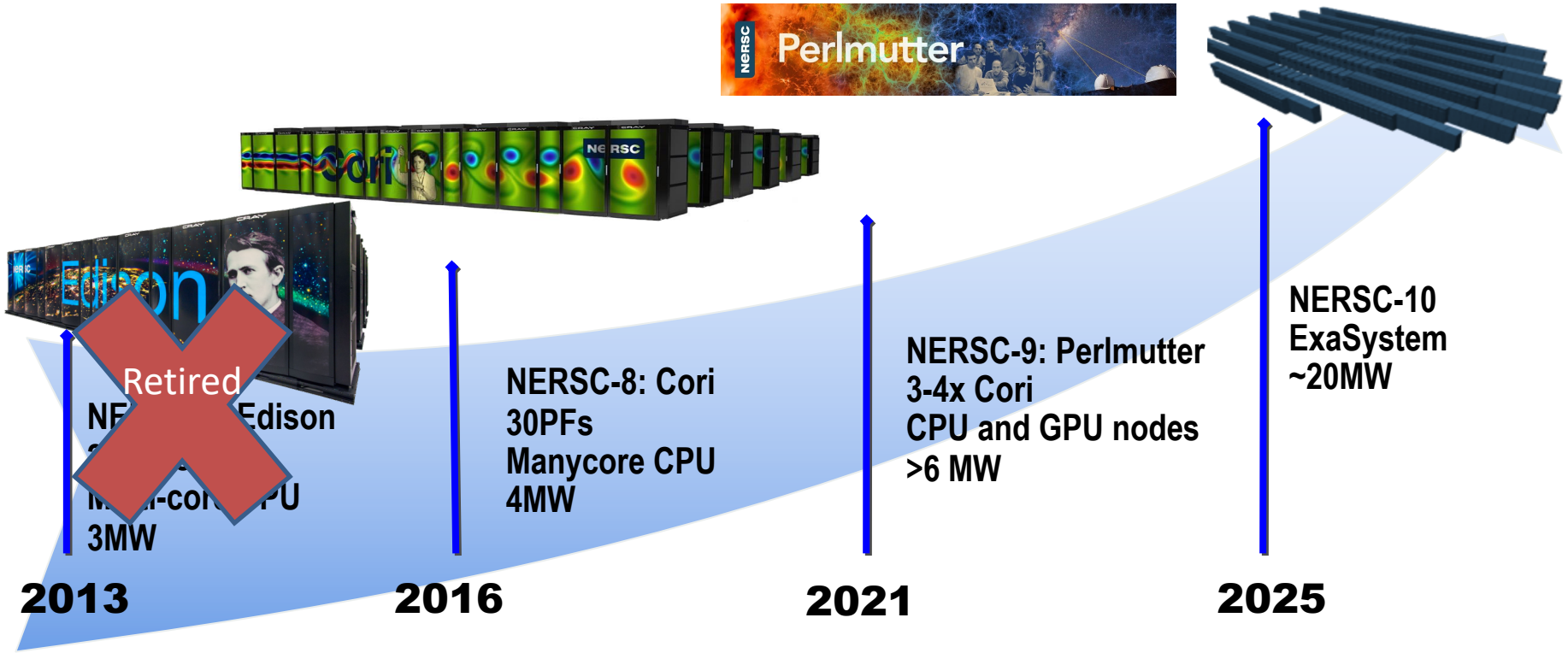
NERSC Directly Supports Office of Science Priorities

2020 Allocation Breakdown (Hours Millions)

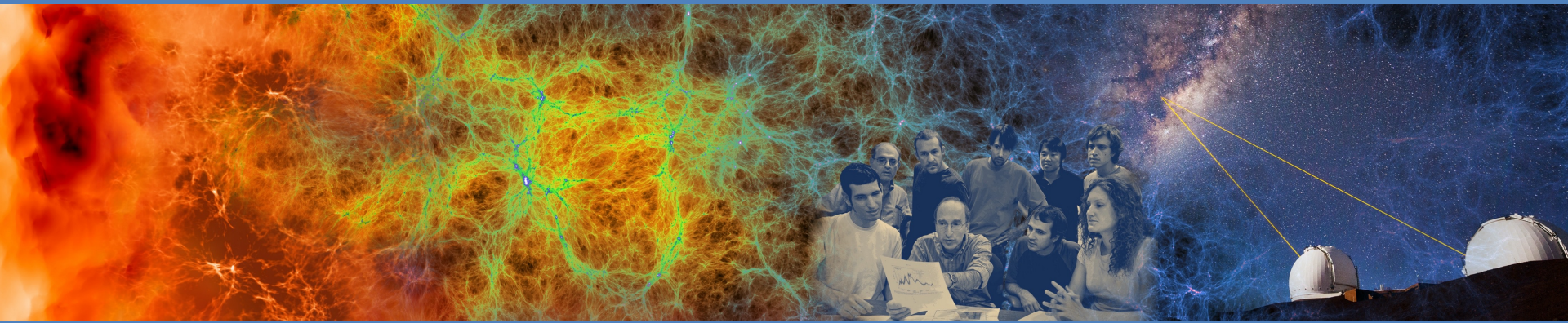




NERSC Systems Roadmap



NERSC-8: Cori

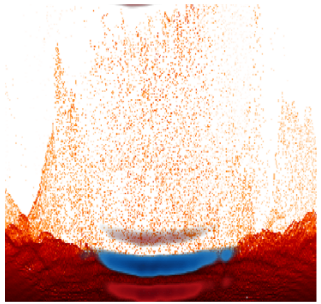


Cori: Designed for Simulation and Data

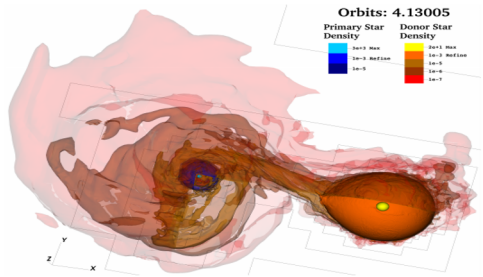
- Cray XC System - heterogeneous compute architecture
  - 9600 Intel KNL compute nodes, >2000 Intel Haswell nodes
- Cray Aries Interconnect
- NVRAM Burst Buffer, 1.6PB and 1.7TB/sec
- Lustre file system 28 PB of disk, >700 GB/sec I/O
- Investments to support large scale data analysis
  - High bandwidth external connectivity to experimental facilities from compute nodes
  - Virtualization capabilities (Shifter/Docker)
  - More login nodes for managing advanced workflows
  - Support for real time and high-throughput queues
  - Data Analytics Software
- ***New this year: GPU rack integrated into Cori***

Cori Hours used for Science

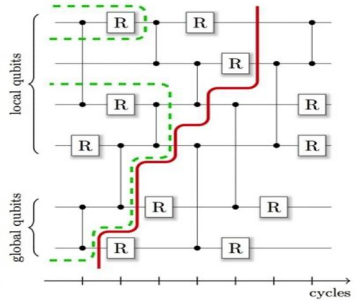
	2019 Target	2019 Actual	2020 Target	2020 Actual
Total Usage	8.1 B	8.8 B	7.5 B	8.0 B
Capability Usage	≥ 25%	28.6%	≥ 25%	39.76%



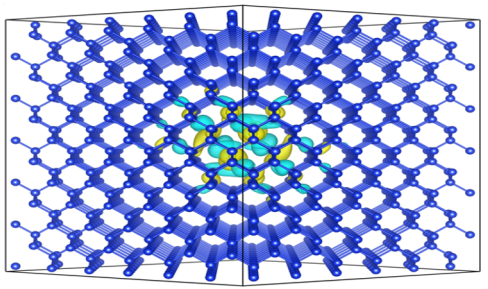
Large Scale Particle in Cell Plasma Simulations



Stellar Merger Simulations with Task Based Programming

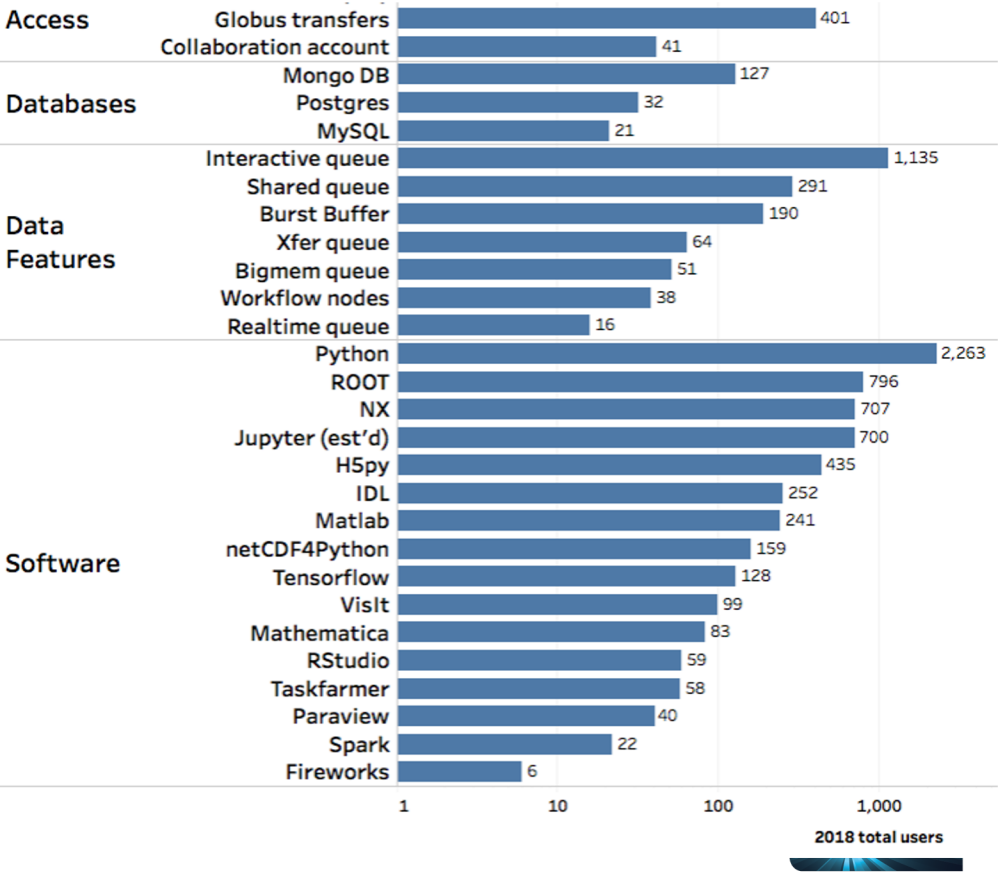


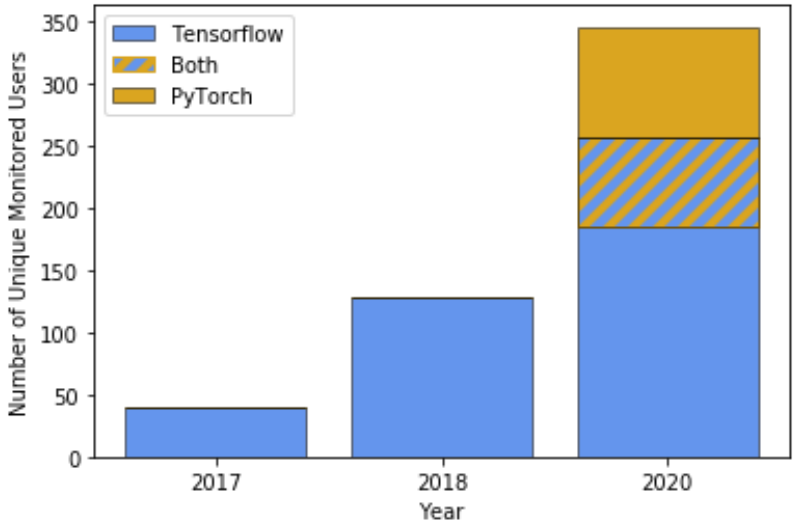
Largest Ever Quantum Circuit Simulation



Largest Ever Defect Calculation from Many Body Perturbation Theory > 10PF

Strong Adoption of Data Software Stack





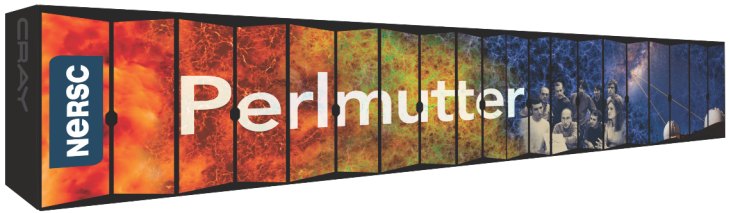
NERSC 9: Perlmutter

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 NERSC supercomputers combine large scale simulations with experimental data analysis
- Login “saul.nersc.gov”

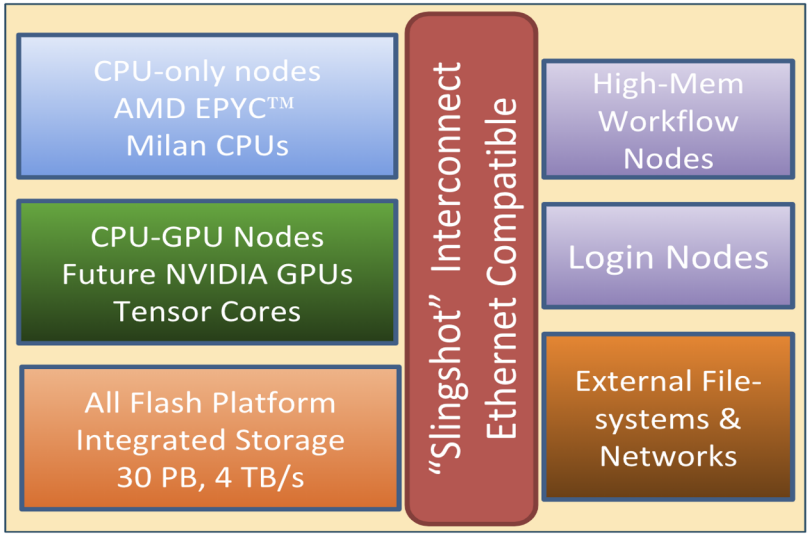
Perlmutter Capabilities

- Cray Shasta System providing 3-4x capability of Cori
- Large CPU-only partition provides capability similar to Cori
  - 2 Milan CPUs w/256GiB DDR4
- GPU nodes: 4 NVIDIA A100 GPUs each w/Tensor Cores, NVLink-3 and High-BW memory + 1 AMD “Milan” CPU
  - Over 6000 NVIDIA A100 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



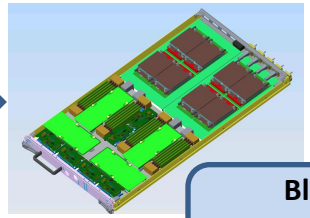
From the start NERSC-9 had requirements of simulation and data users in mind

- 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

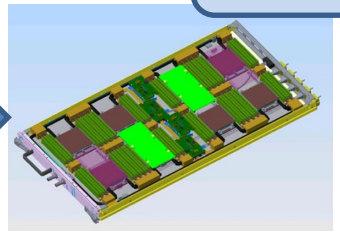


Perlmutter at a glance

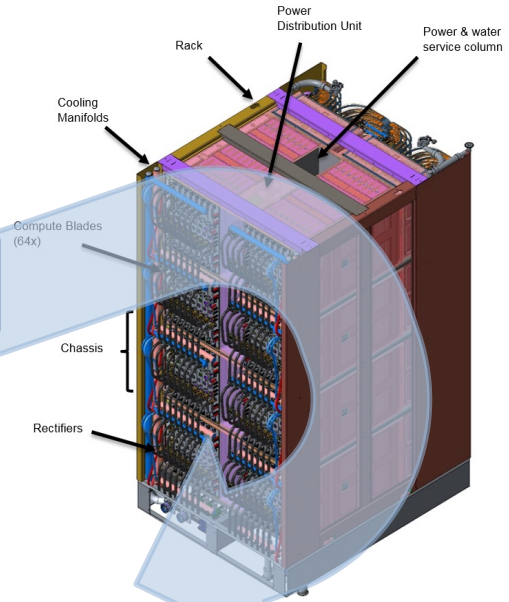
NVIDIA ”A100" GPU Nodes
4x GPU + 1x CPU (>75 TF)
160 GiB HBM + DDR
4x 200G High Performance NICs



AMD "Milan" CPU Node
2x CPUs
> 256 GiB DDR4
1x 200G High Performance NIC



Blades
2x GPU nodes or
4x CPU nodes



Centers of Excellence
Network
Storage
App. Readiness
System SW

Perlmutter system
12 GPU racks
12 CPU racks

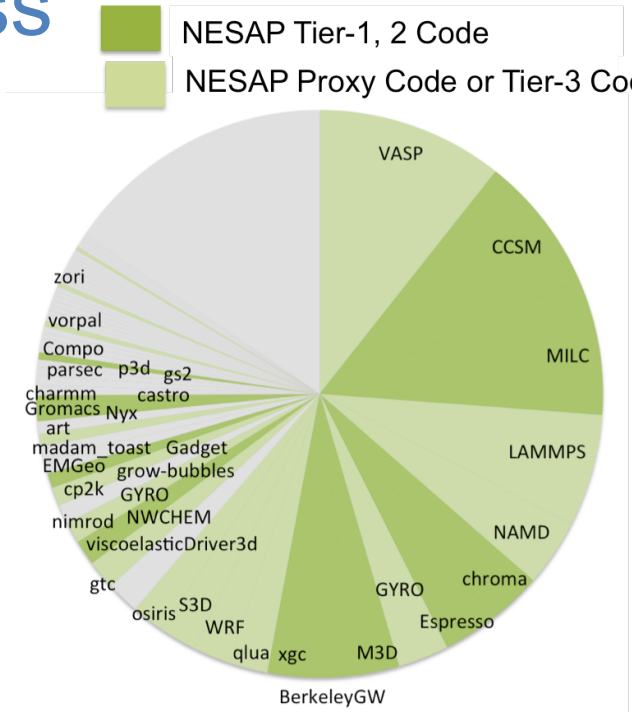
NESAP: Application Readiness

NESAP is NERSC's Application Readiness Program for preparing our workload for new systems.

Strategy: Partner with application development teams and vendors to port & optimize key applications of importance to the Office of Science. Share lessons learned with with NERSC community via documentation and training.

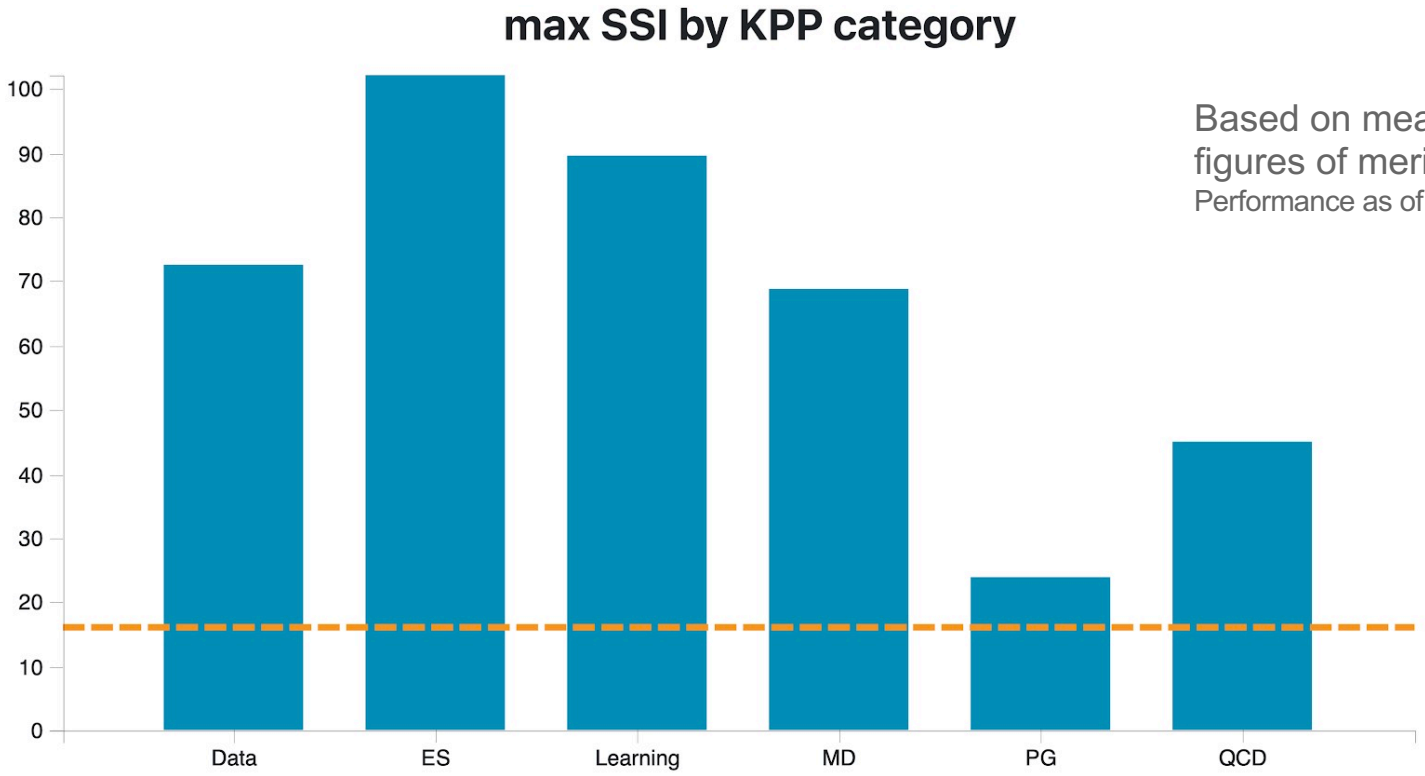
Resource Available to Teams: NERSC Staff technical liaisons, performance postdocs, access to vendor application engineers, hackathons, early access to hardware (GPU nodes on Cori and Perlmutter)

Simulation: 14 application teams
Data: 9 applications
Learning: 5 applications



NESAP covers a broad swath of the NERSC workload. The pie chart shows a breakdown of applications used at NERSC in 2017-18.

NESAP Application Performance Improvements



Based on measured application figures of merit for 6

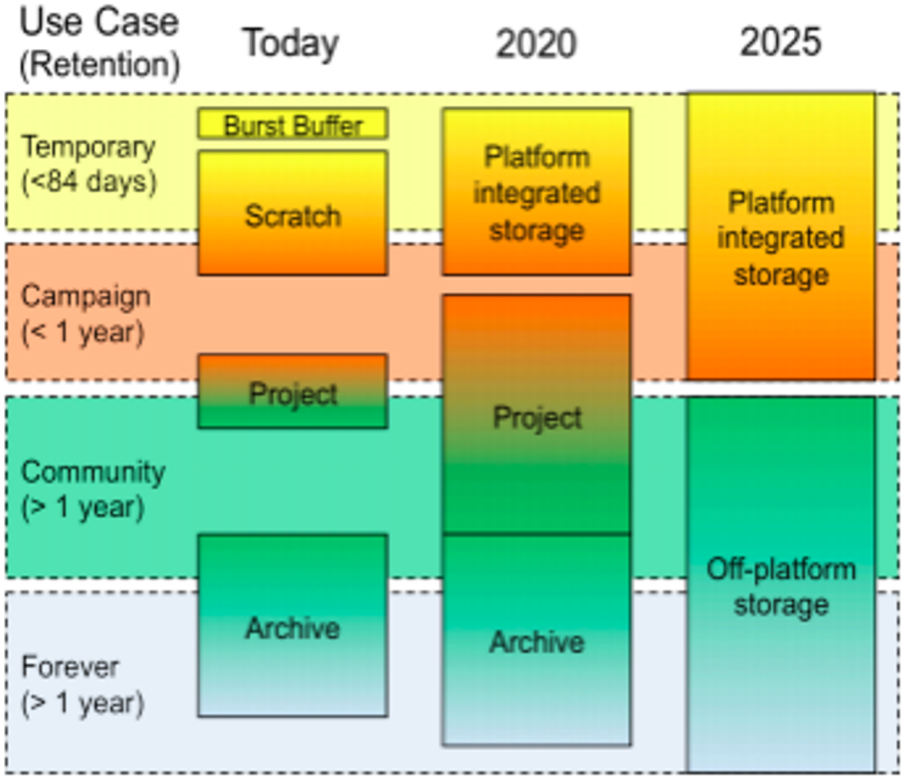
Performance as of January 2021

Additional NERSC Capabilities

Storage 2020 Community File System

Project filesystem replacement

- 75 PB available to users in FY2020
- 150 - 300 PB by Perlmutter deployment



Enabling Edge Services with Spin

Challenge

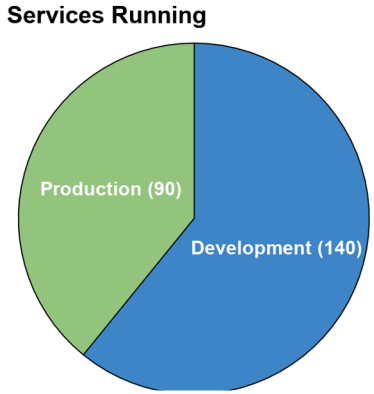
- Workflows often require additional edge services (DBs, APIs, Portals) to achieve their science.

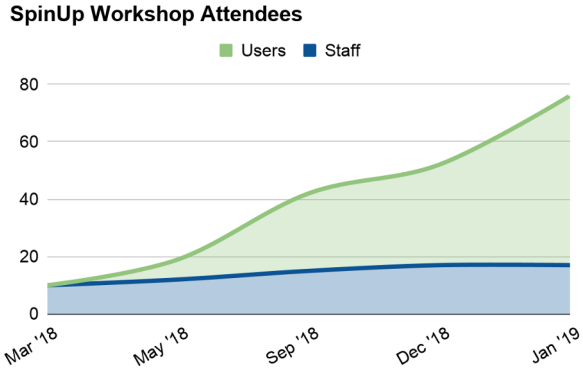
Innovation

- NERSC provides Spin, a multi-tenancy, container-based orchestration system, to support user managed edge services
- NERSC provides the infrastructure, users only concern is to provide their services
- Training and user support were implemented to rapidly on-board projects

Impact and Early Successes

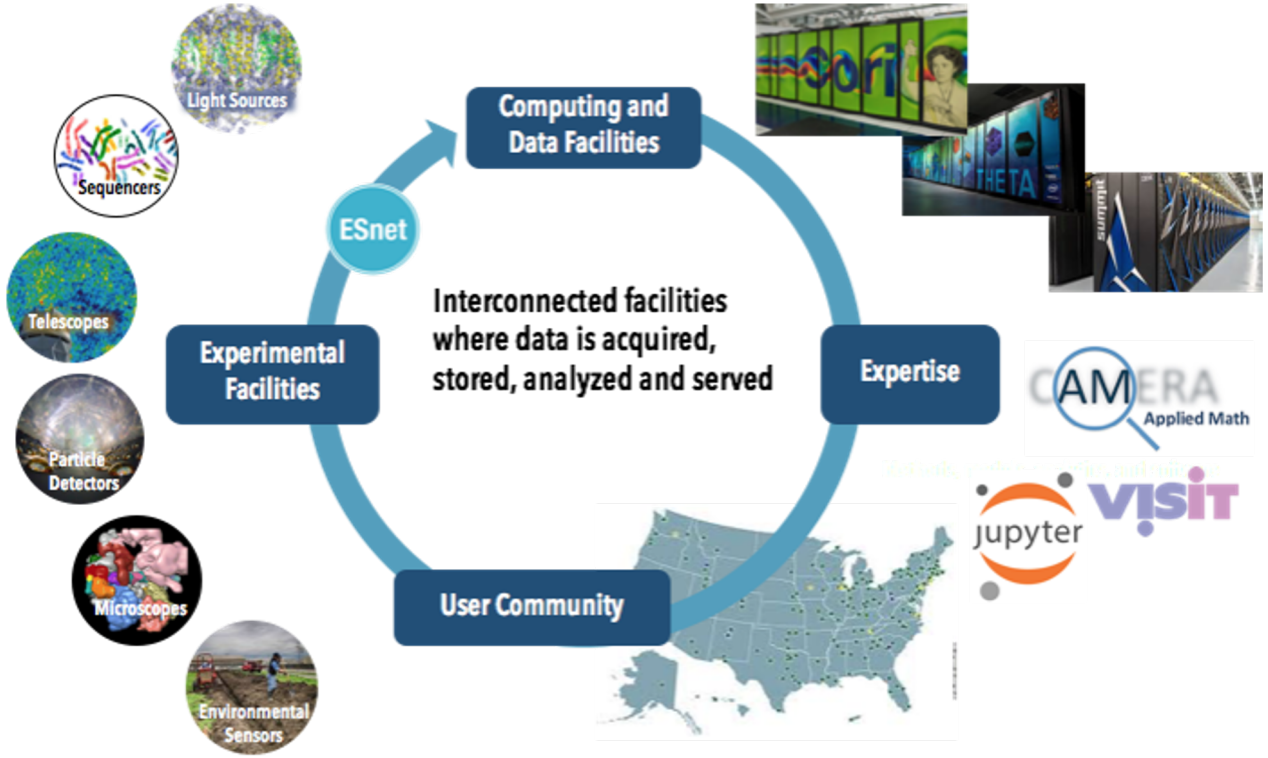
- >70 users have taken training and over 90 services have been deployed in production
- A trained user can bring up a new service in a matter of hours with no staff intervention



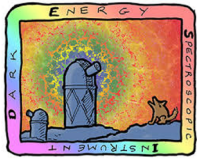


Superfacility: A model to integrate experimental, computational and networking facilities for reproducible science

Enabling new discoveries by coupling experimental science with large scale data analysis and simulations



On-going Engagements with experimental facilities drive our requirements



Experiments operating

Future experiments

Identity and Access Management (IAM)

- **New IAM solution will be built with components from Internet2 TIER project**
- **Benefits for experimental facility workflow users:**
  - Simpler account creation
  - Ability for users to have different roles (data users, shell access, web gateway)
  - Transparent and consistent rules for granting access
  - Easier to activate and deactivate accounts, particularly for large projects with many members ⇒ better security
  - Native federated identity support!

Identity Enrollment & Registry

Group-based access management

NERSC Mission

The mission of the National Energy Research Scientific Computing Center (NERSC) is to accelerate scientific discovery at the DOE Office of Science through high performance computing and data analysis.

Questions?