

# NERSC Site Update

## National Energy Research Scientific Computing Center

Lawrence Berkeley National Laboratory

Richard Gerber

NERSC Senior Science Advisor

High Performance Computing Department Head



Cray XC40 system

9,600 Intel Xeon Phi (KNL 7250) @ 1.4 GHz

2,000 Intel Dual-Socket Haswell @ 2.3 GHz

Cray Aries 3-level dragonfly network connects KNL and Haswell nodes

NVRAM Burst Buffer 1.8 PB, 1.5 TB/sec

30 PB Lustre scratch, >700 GB/sec I/O

**Cori KNL is running well and is being used productively by a DOE Office of Science researchers**

- KNL nodes installed in late 2016
- Early users Nov. 2016 – Feb. 2017
- All users enabled – March 2017
- Production computing – July 2017

**Follow on system (NERSC 9) will be announced in late 2018**

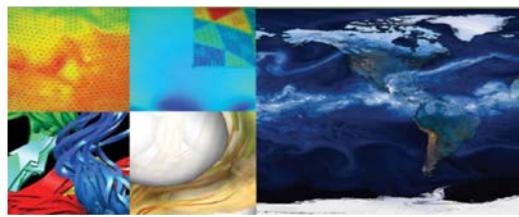
# NERSC: the Mission HPC Facility for DOE Office of Science Research



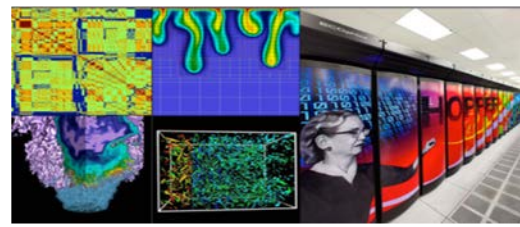
U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

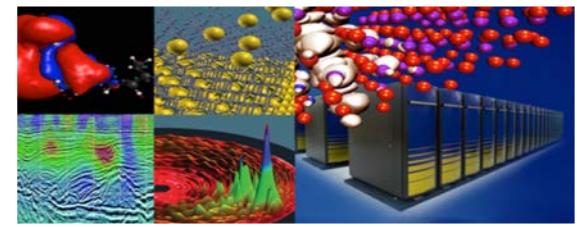
Largest funder of physical  
science research in the U.S.



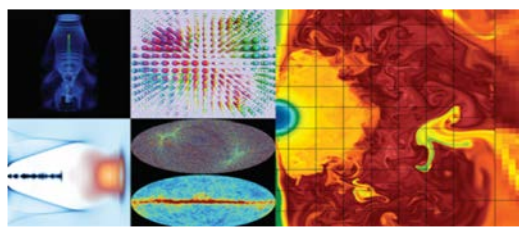
Bio Energy, Environment



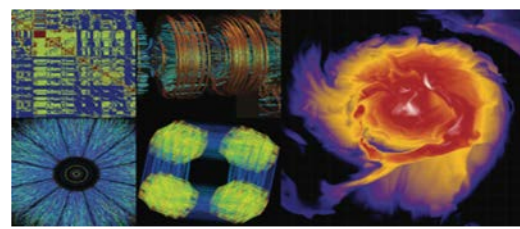
High Performance Computing



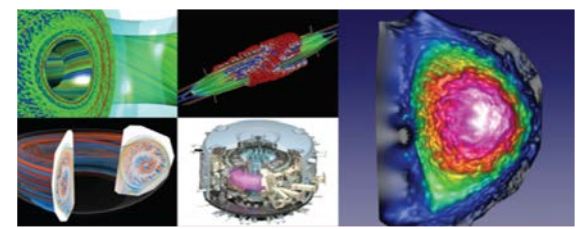
Materials, Chemistry, Geophysics



Particle Physics, Astrophysics



Nuclear Physics



Fusion Energy, Plasma Physics

7,000 users, 800 projects, 700 codes, 50 states, 40 countries, universities & national labs

# Focus on Science



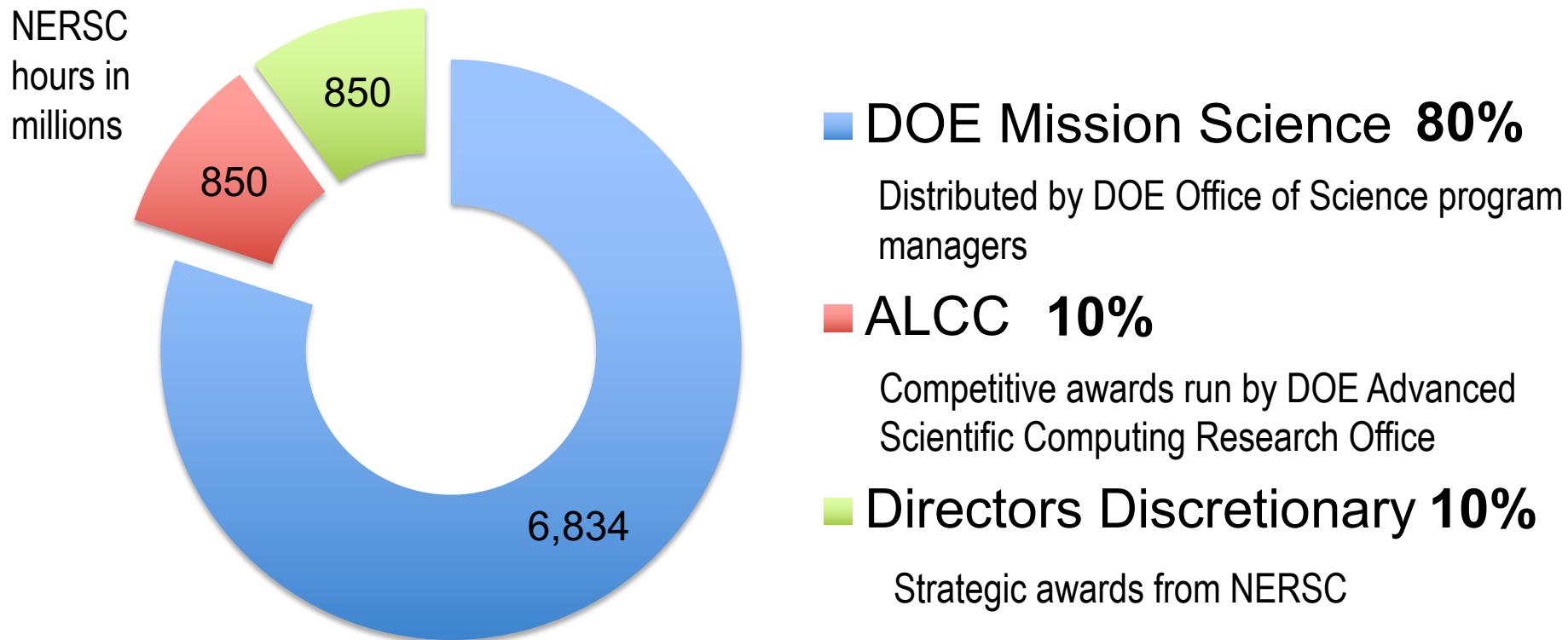
NERSC supports the broad mission needs of the six DOE Office of Science program offices

Systems procured on based on performance of real-world proxies for NERSC workload

Supports high-impact simulation and data science

>2,400 refereed publications in 2017





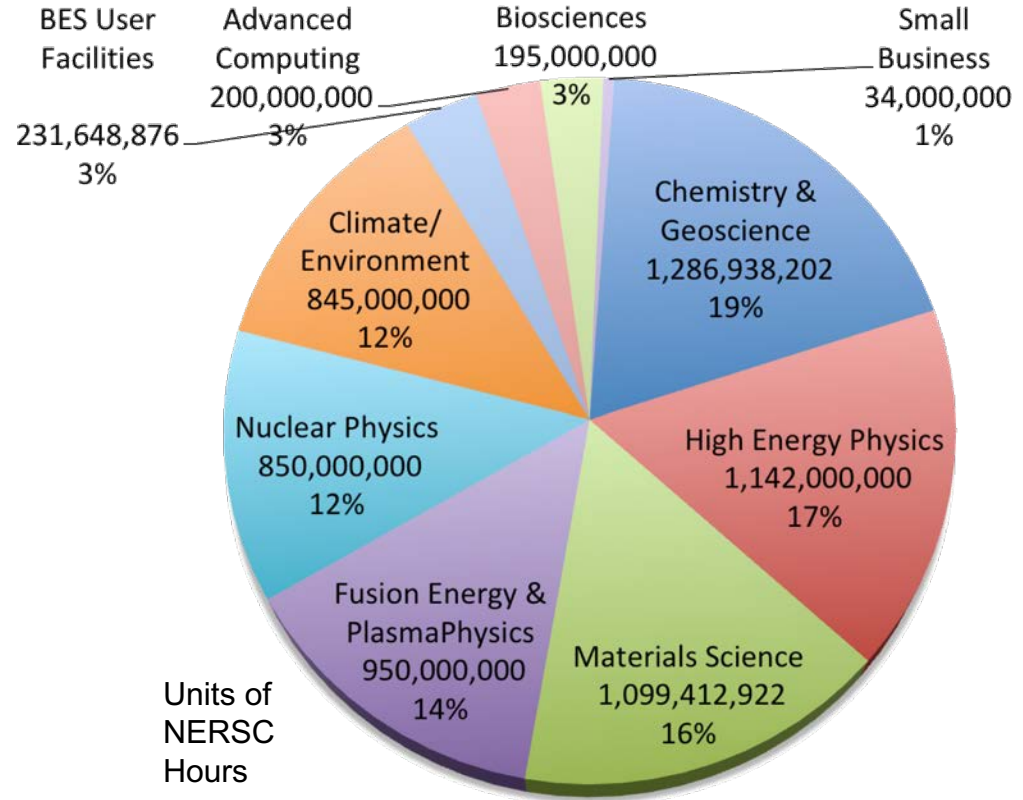
# DOE Mission Science Allocation Breakdown 2018



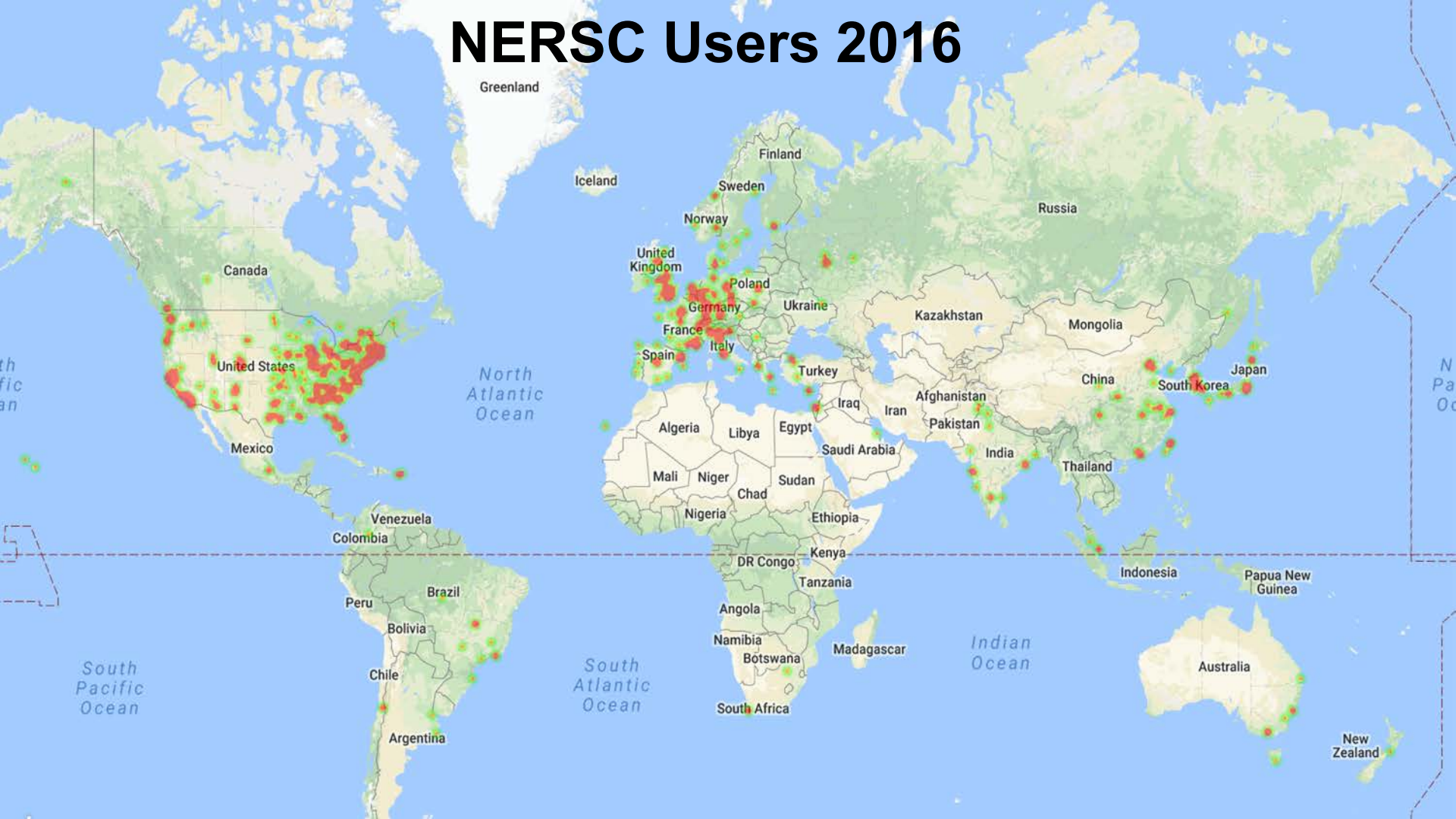
NERSC's diverse workload comes from its mission to support all areas of science within the Department of Energy

NERSC is not able to select and support only codes that run well on its systems

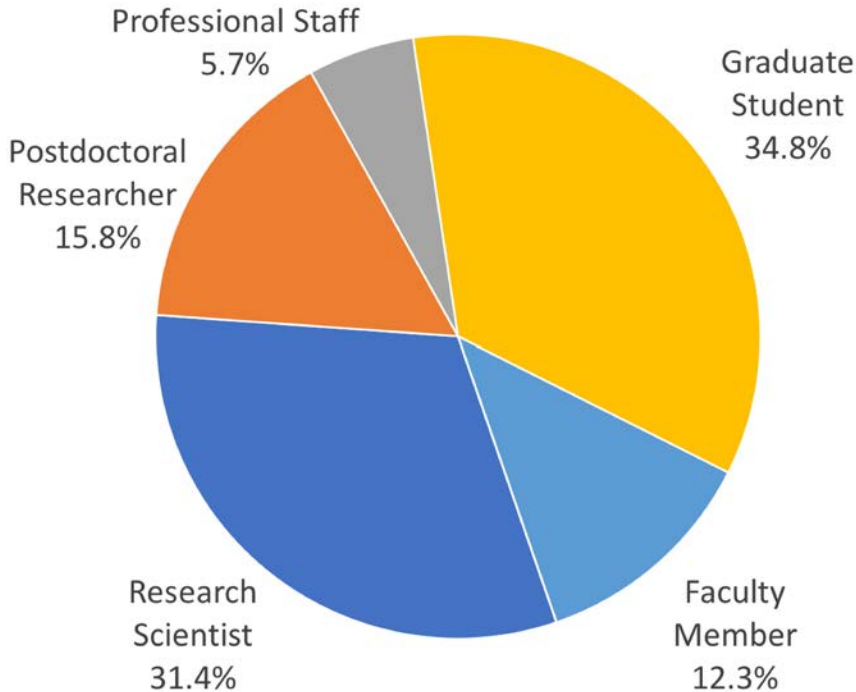
Workload mix is driven by science priorities that require HPC



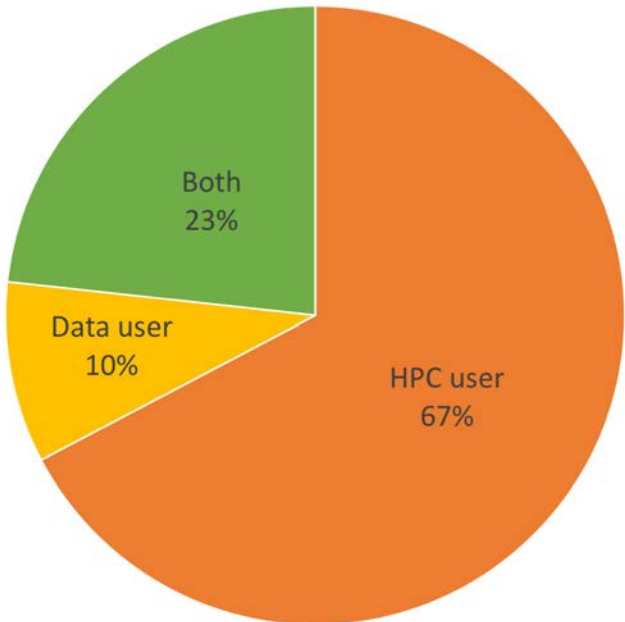
# NERSC Users 2016



### Hours Used 2017



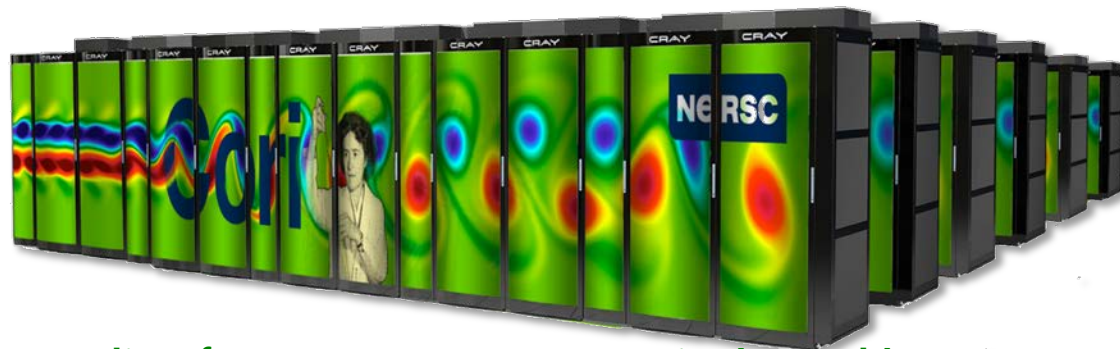
### User Type from User Survey





## Cori

9,600 Intel Xeon Phi "KNL" manycore nodes  
2,000 Intel Xeon "Haswell" nodes  
700,000 processor cores, 1.2 PB memory  
Cray XC40 / Aries Dragonfly interconnect  
30 PB Lustre Cray Sonexion scratch FS  
1.5 PB Burst Buffer



#8 on list of Top 500 supercomputers in the world



## Edison

5,560 Ivy Bridge Nodes / 24 cores/node  
133 K cores, 64 GB memory/node  
Cray XC30 / Aries Dragonfly interconnect  
6 PB Lustre Cray Sonexion scratch FS

- Enable state-of-the-art HPC at scale for all areas of science,
- For users of all levels of experience and knowledge,
- Including new communities with non-traditional data and machine learning needs
- Lead the community to exascale and beyond

# Cori: Challenges and Opportunities for NERSC Users on the Path to Exascale



## Edison (“Ivy Bridge”):

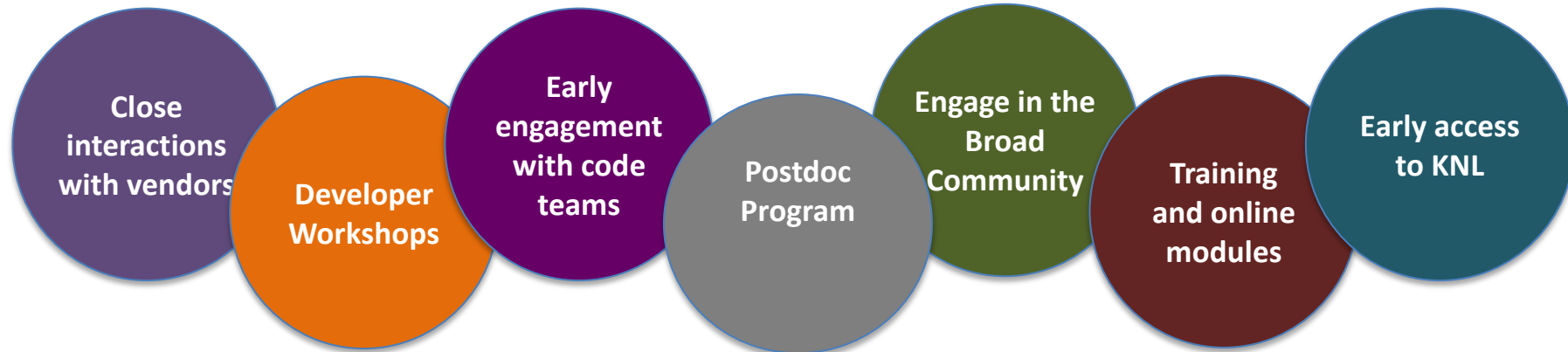
- 5,576 nodes
- 24 physical cores per node
- 48 virtual cores per node
- 2.4 GHz
  
- 8 double precision ops/cycle
  
- 64 GB of DDR3 memory
- 2.5 GB per physical core
  
- ~100 GB/s Memory Bandwidth

## Cori (“Knights Landing”):

- 9,600 nodes
- 68 physical cores per node
- 272 virtual cores per node
- 1.4 GHz
  
- 32 double precision ops/cycle
  
- 16 GB of fast memory (.25/core)
- 96GB of DDR4 memory (1.5/core)
- Fast memory has 400 - 500 GB/s
- No L3 Cache

Partner closely with ~20 application teams and apply lessons learned to broad NERSC user community: proxies for 50% of workload

NESAP activities include:



## Code Speedups Via NESAP (per node):

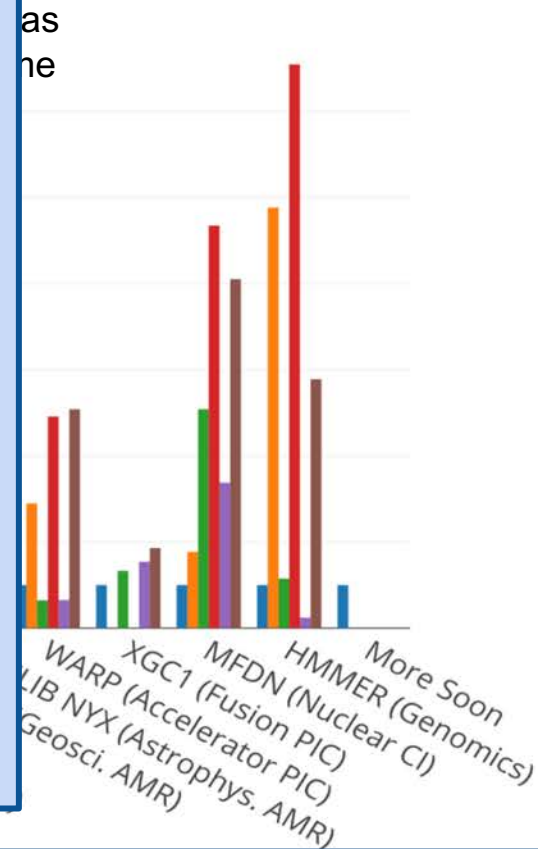
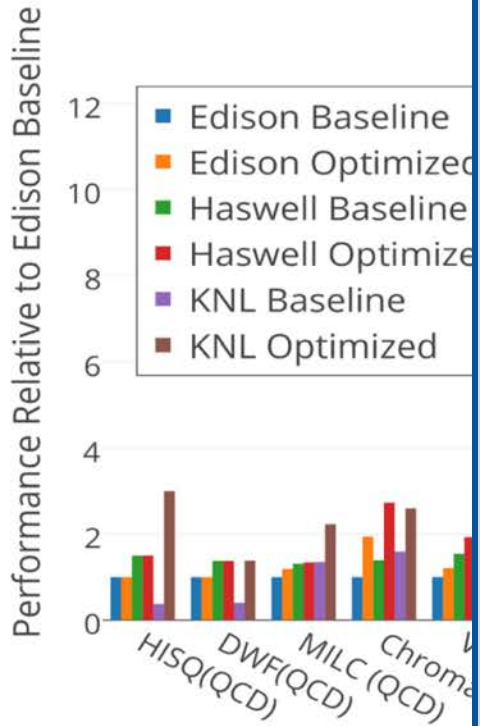
### KNL / Haswell Performance Ratio

Baseline Codes	0.7	(KNL is slower)
Optimized Codes	1.2	(KNL is faster)
KNL Optimized / Haswell Baseline	<b>2.5</b>	

### KNL / Ivy-Bridge (Edison) Performance Ratio

Baseline Codes	1.1	(KNL is faster)
Optimized Codes	1.8	(KNL is faster)
KNL Optimized / Edison Baseline	<b>3.4</b>	

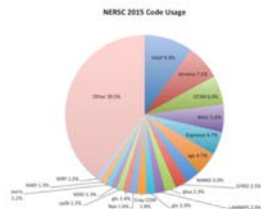
"Business as Usual" : Recompile and run on Haswell:  
**Haswell baseline/Ivy Bridge baseline ~ 1.5**



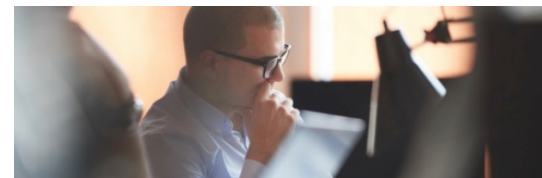
# It Takes a Connected Community



NESAP: 20 codes, >50% workload



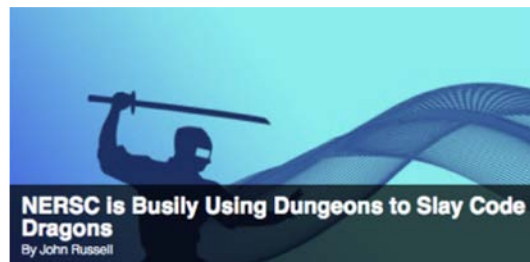
NERSC Application  
Performance Team  
Staff & Postdocs



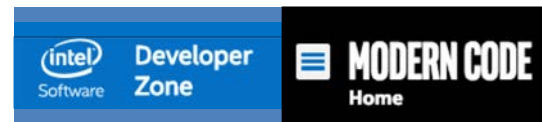
Robust training programs  
~100s –1,000s of users



Math, CS, Computational  
Science



NESAP, Intel, Cray



# But ... Not All Projects Have Adopted KNL

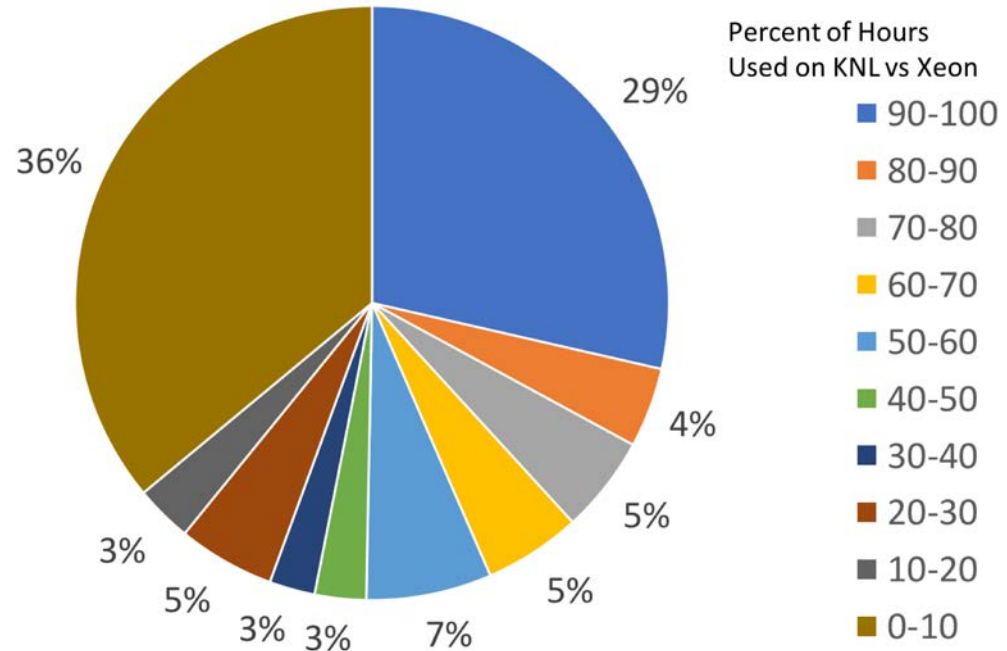


Some legacy codes perform poorly on KNL

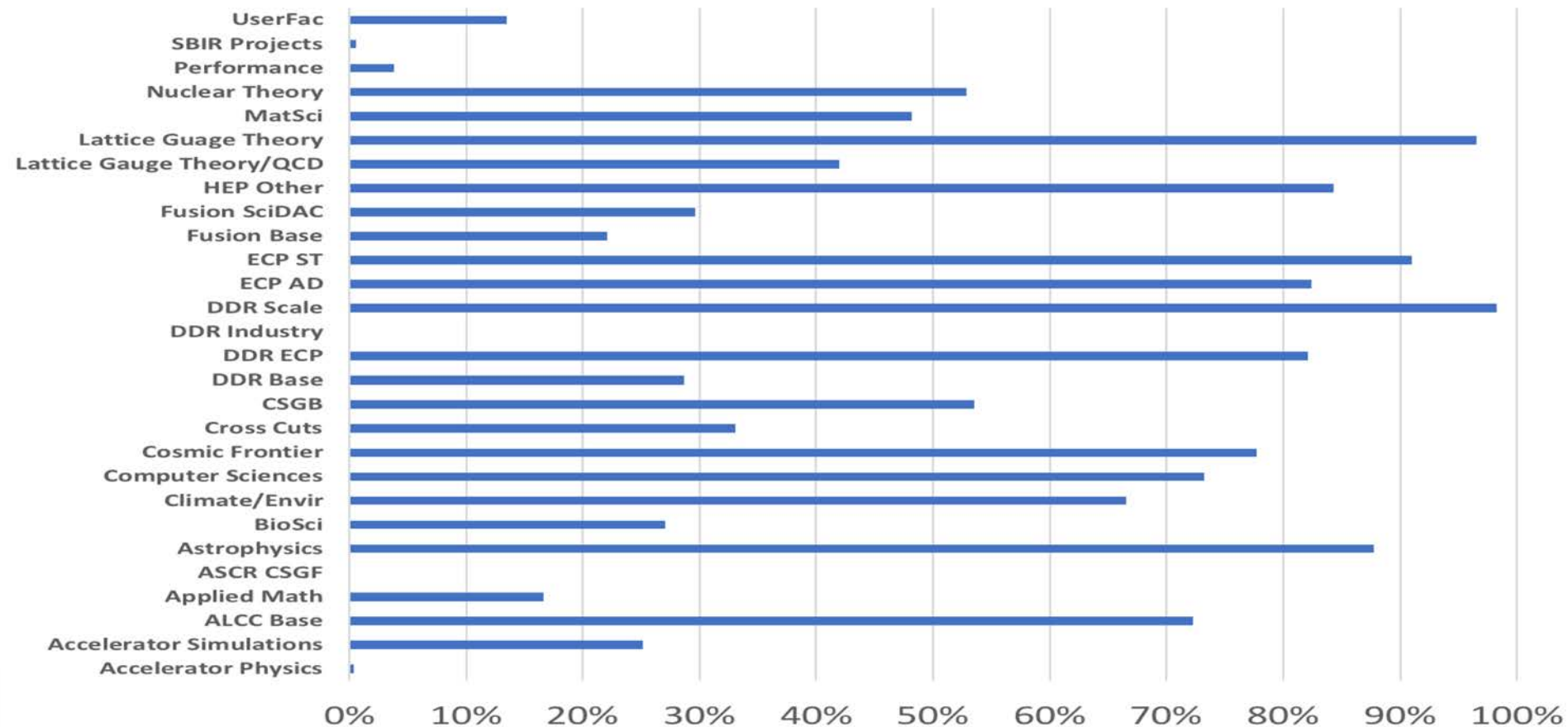
Average performance of our “unoptimized well written” codes on KNL vs. Haswell dual socket: 70%

Some single-threaded codes see 500% slowdown

Percent of NERSC Projects 2018



# Percent of 2018 NERSC Usage on KNL





- Goal for 2018: have 50% of projects using KNL for >50% of their computing at NERSC.
- We are exploring languages, programming models, libraries, frameworks and working with standards committees, tool vendors and Exascale Computing Project to make advanced architectures useful to the broad community
- NESAP 2 targeting NERSC 9 (2020) system
- Big Data Center & NESAP for data

# NERSC-9 System in 2020



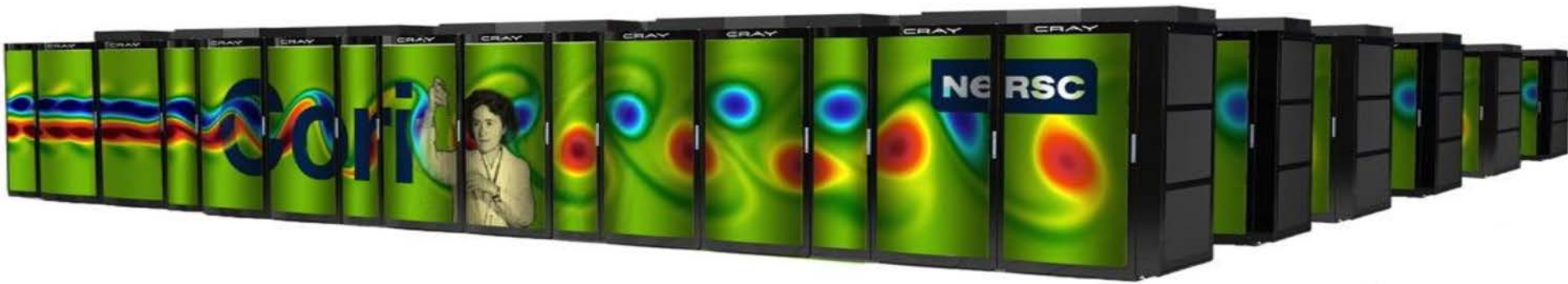
1. Provide 3-4X capability of Cori
2. Meet needs of simulation and data analysis use cases including:
  - a. Complex workflows
  - b. Analytics and machine learning at scale
  - c. Support for experimental facilities workflows
3. Prepare users for exascale and more specialization / heterogeneity

*System will be announced in 2<sup>nd</sup> half 2018*

# NERSC Systems Timeline

2007/2009	NERSC-5	Franklin	Cray XT4	102/352 TF
2010	NERSC-6	Hopper	Cray XE6	1.28 PF
<b>2014</b>	<b>NERSC-7</b>	<b>Edison</b>	<b>Cray XC30</b>	<b>2.57 PF</b>
<b>2016</b>	<b>NERSC-8</b>	<b>Cori</b>	<b>Cray XC</b>	<b>30 PF</b>
<b>2020</b>	<b>NERSC-9</b>		Selection underway	<b>~100-150 PF</b>
<b>2024</b>	<b>NERSC-10</b>			<b>1EF</b>

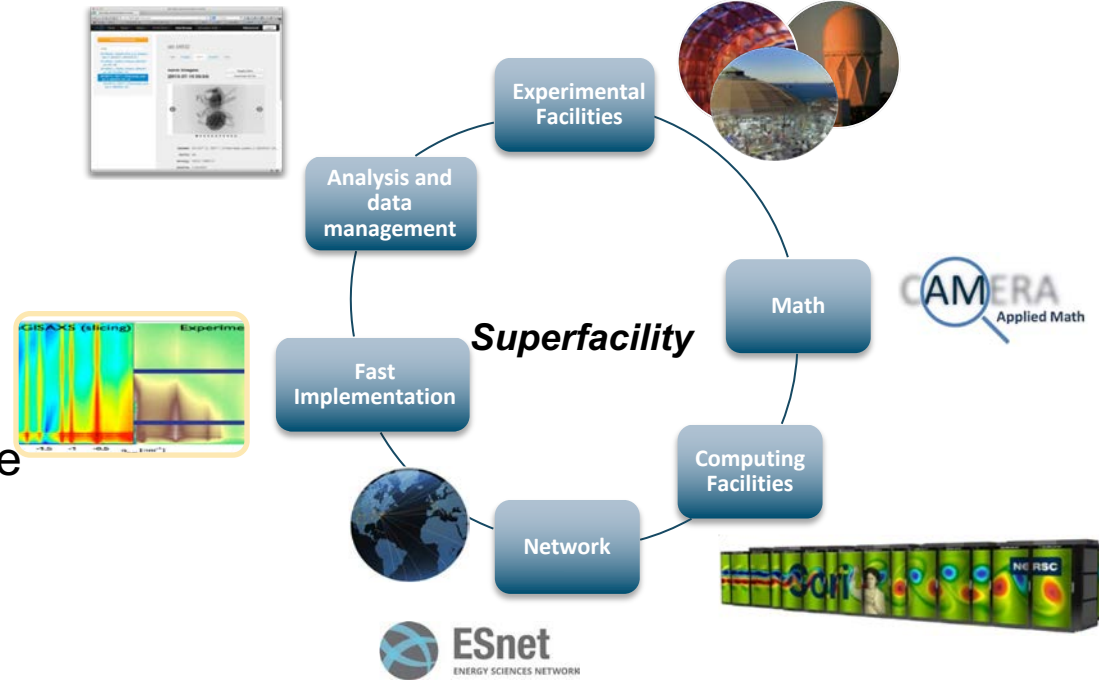
Edison is currently scheduled to retire in ~March 2019 (subject to change)



# Superfacility

A network of connected facilities, software and expertise to enable new modes of discovery

- Deploying large scale computing and storage resources
- Providing reusable building blocks for experimental scientists to build pipelines
- Providing scalable infrastructure to launch services
- Expertise on how to optimize pipelines



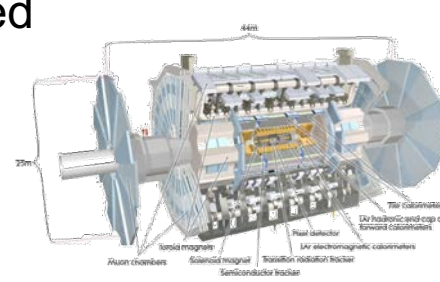
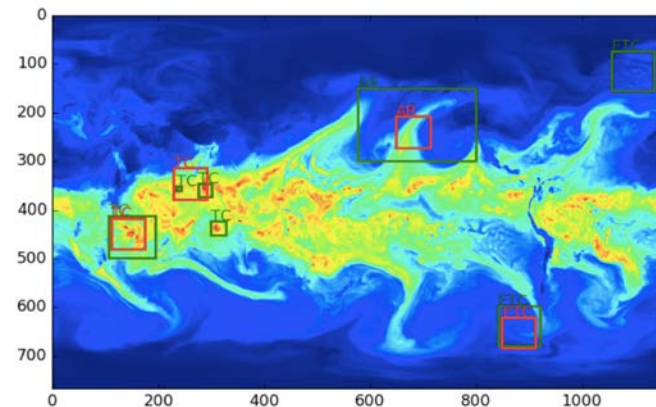
- **Big Data Center**

- Goal: “Enable Data capability applications on Cori; optimize and scale the production data analytics + management stack”
- Collaboration: Intel, Cray, NERSC, IPCC

- **Status**

- Julia (1 PF), Caffe (15 PF) scaled in 2017
- TensorFlow scaled out (GB submission)
- PyTorch and raw Python being targeted next
- Data Management tools are a major bottleneck

- **NESAP for Data**



Cori KNL is running well and is being used productively by a the DOE Office of Science workload

But, there is still more work to do to get the majority NERSC users on Cori KNL and on the path to exascale

NERSC 9 will be coming in 2020 and the architecture will be announced in the second half of 2018

NESAP 2 will begin as soon as NERSC 9 is announced

NERSC is targeting Cori and NERSC 9 for data-intensive workloads

# NERSC at a Glance

A U.S. Department of Energy Office of Science User Facility  
Provides High Performance Computing and Data Systems and Services  
Unclassified Basic and Applied Research in Energy-Related Fields  
7,000 users, 800 different scientific projects  
Located at Lawrence Berkeley National Lab, Berkeley, CA  
Permanent Staff of about 70

