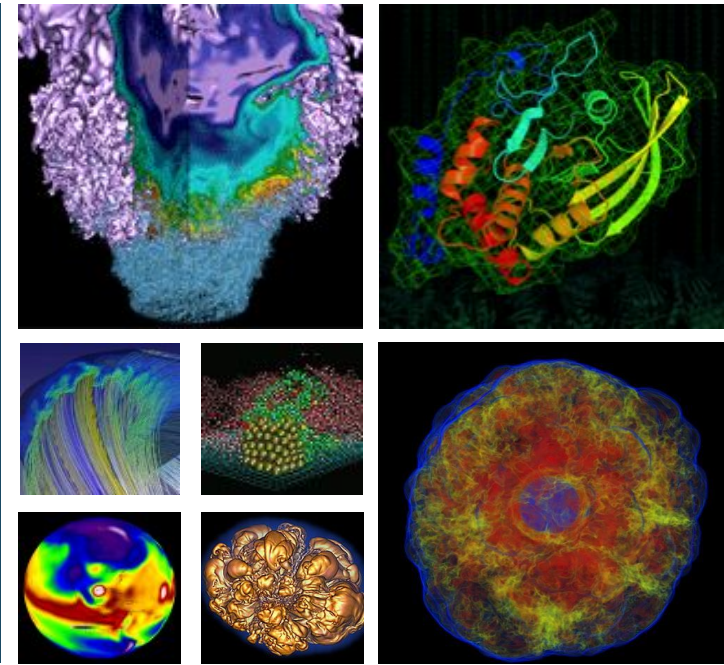
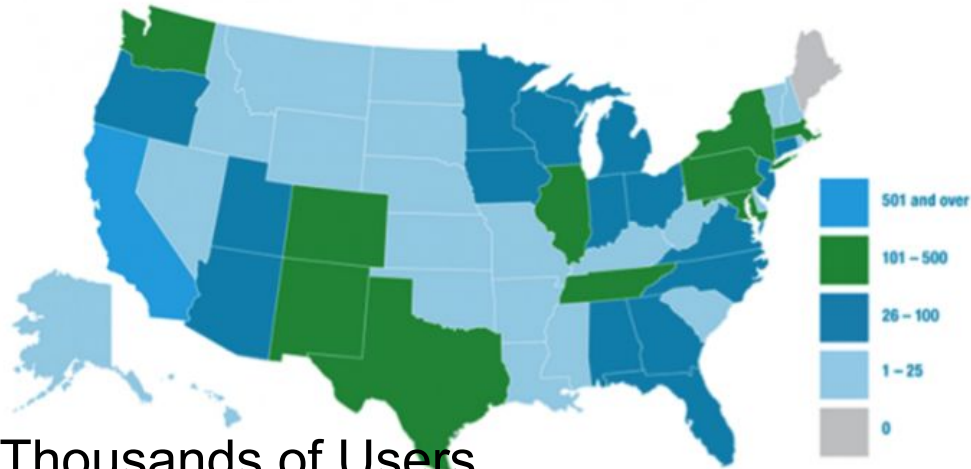


# Future Advanced Computing Ecosystem



Katie Antypas  
NERSC Division Deputy  
August 17, 2020

# NERSC is the mission High Performance Computing facility for the DOE Office of Science

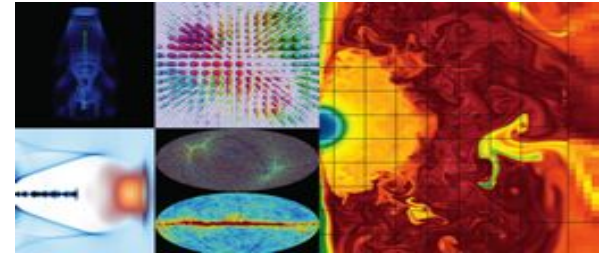


Thousands of Users

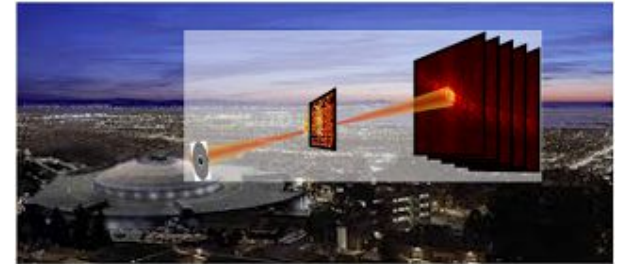
800 Projects

700 Codes

2000 NERSC citations per year

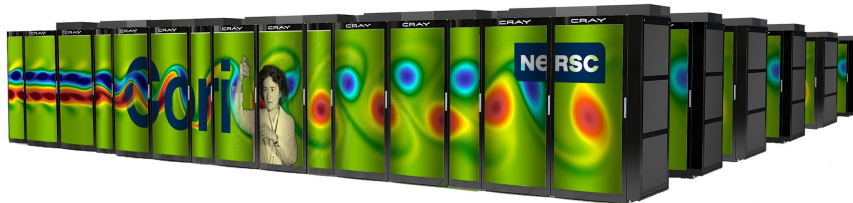


Simulations at scale



Data analysis support for  
DOE's experimental and  
observational facilities

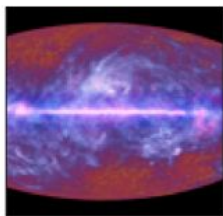
Photo Credit: CAMERA



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



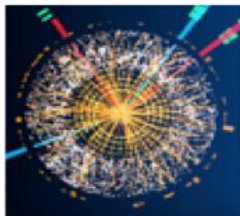
Palomar Transient  
Factory  
Supernova



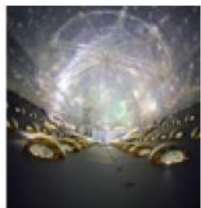
Planck Satellite  
Cosmic Microwave  
Background  
Radiation



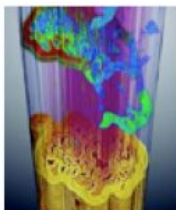
Alice  
Large Hadron Collider



Atlas  
Large Hadron Collider



Dayabay  
Neutrinos



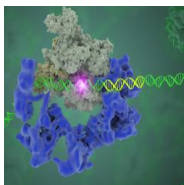
ALS  
Light Source



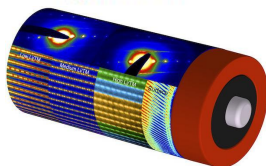
LCLS  
Light Source



Joint Genome Institute  
Bioinformatics



Cryo-EM

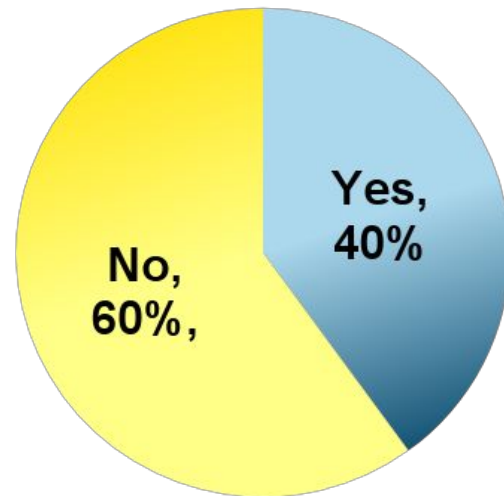


NCEM

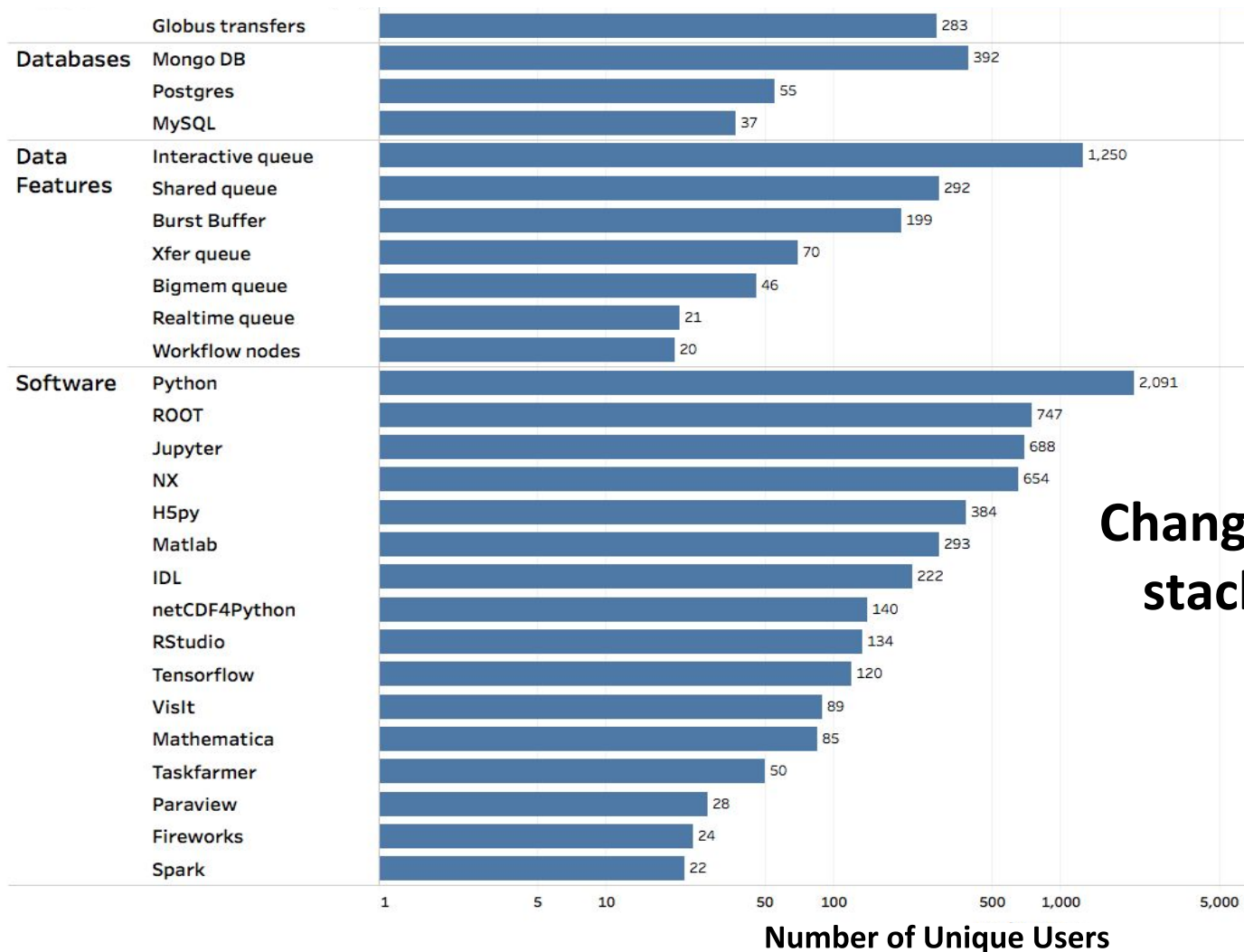


DESI

## 2018 Call for Proposals Question

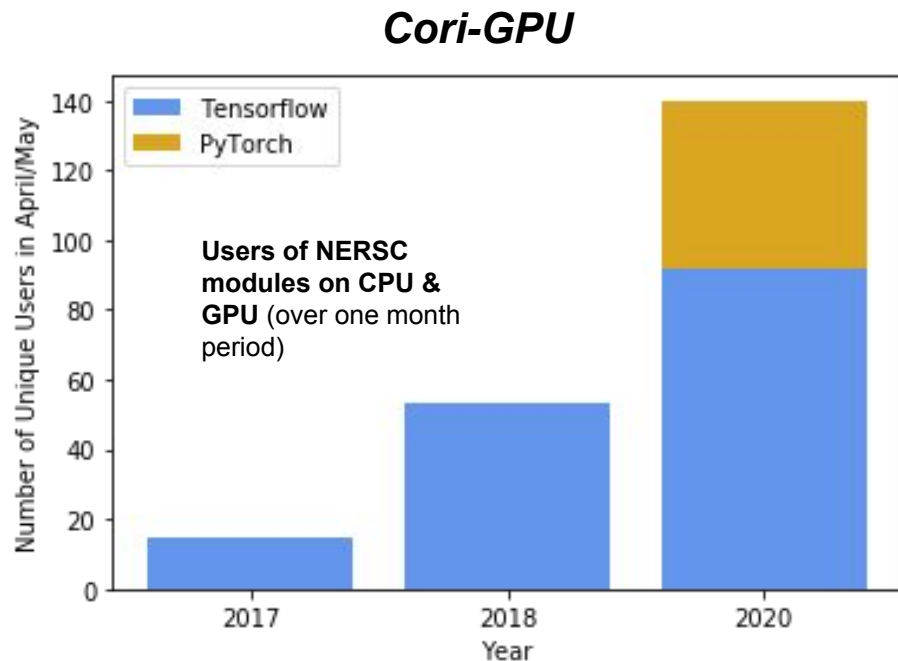


Is the primary role of this project to analyze data from experimental facilities; Or combined models and simulations with experimental data?



**Changing software stack at NERSC**

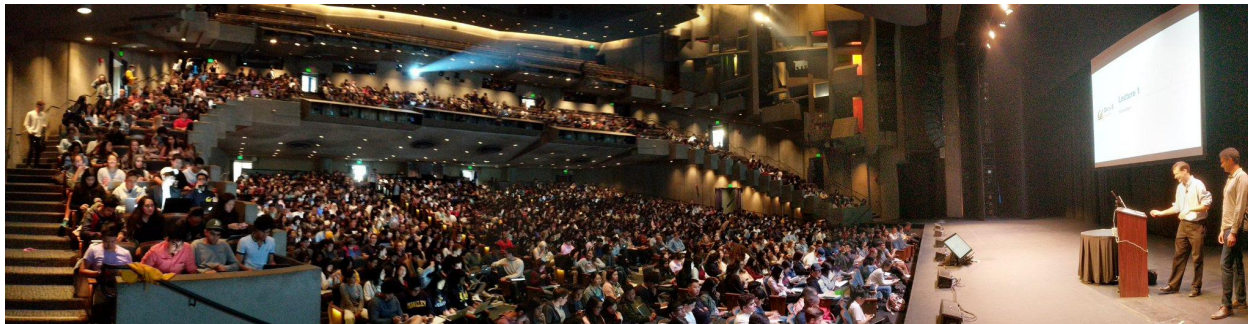
# Growing AI and DL workload



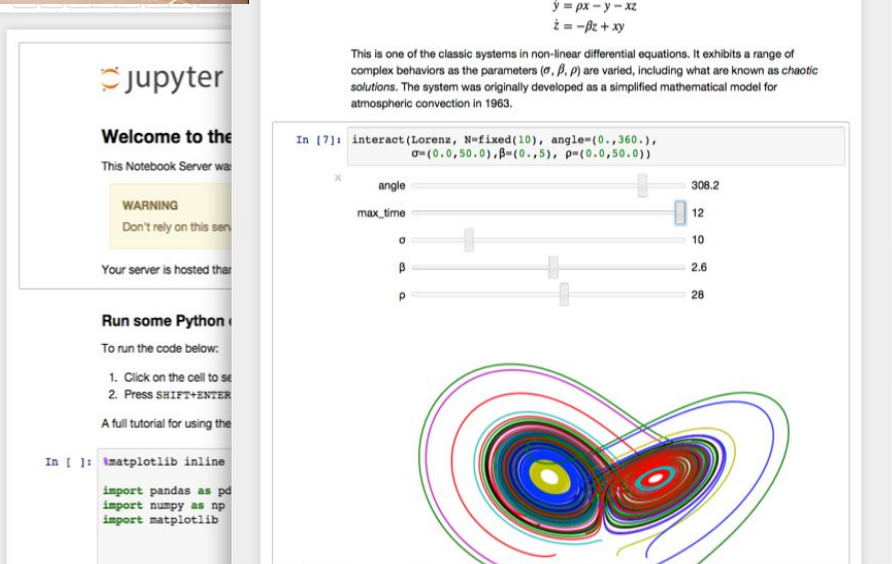
- Deep learning (DL) users/workload growing
- [Cori-GPU](#) development system useful indicator for Perlmutter
- More than 75% of the compute hours are DL users
- Users routinely run distributed across multiple GPUs



# Jupyter: reproducible, sharable computing, now at scale



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



The image shows a Jupyter Notebook interface with a title bar that reads "jupyter Lorenz Differential Equations (autosaved)". The notebook contains the following text:

**Exploring the Lorenz System**

In this Notebook we explore the [Lorenz system](#) of differential equations:

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy\end{aligned}$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters  $(\sigma, \beta, \rho)$  are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

In [7]: `interact(Lorenz, N=Fixed(10), angle=(0.,360.),  
O=(0.0,50.0),beta=(0.,5),p=(0.0,50.0))`

The interface shows sliders for the parameters: angle (0.0 to 360.0), max\_time (0.0 to 12),  $\sigma$  (0.0 to 10),  $\beta$  (0.0 to 2.6), and  $\rho$  (0.0 to 28). Below the sliders is a 3D plot of the Lorenz attractor, showing its characteristic butterfly shape with multiple colored trajectories.

Below the notebook, a "Welcome to the Jupyter Notebook Server" message is visible, along with a "WARNING" box stating "Don't rely on this server" and "Your server is hosted at".

Run some Python code

To run the code below:

1. Click on the cell to select it
2. Press SHIFT+ENTER

A full tutorial for using the Jupyter Notebook is available at [http://jupyter.org](#)


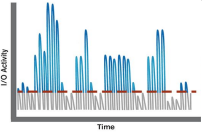
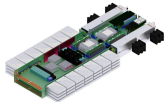


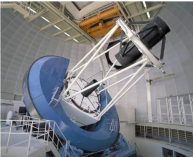

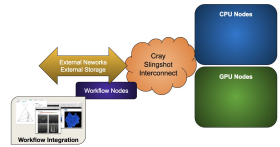
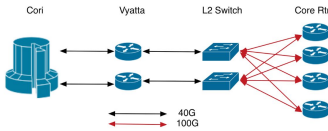
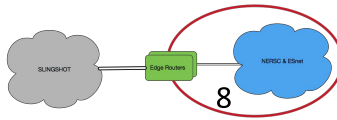
In [ ]: `matplotlib inline  
import pandas as pd  
import numpy as np  
import matplotlib`

# NERSC-9: A System Optimized for Science



- **Cray Shasta System providing >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
  - All-Flash filesystem for I/O acceleration
  - Cray Slingshot high-performance network will support Terabit rate connections to system
  - Optimized data software stack enabling analytics and ML at scale
  - Dedicated workflow and data management nodes
  - Real-time and co-scheduling capabilities
- **Robust readiness program**
- **Delivery in 2 phases, GPU cabinets in late 2020**



Data Features	Cori System	
<b>I/O and Storage</b>	<b>Burst Buffer</b> 	<b>All-flash file system: performance with ease of data management</b> 
<b>Analytics</b> <ul style="list-style-type: none"> <li>- <b>Production stacks</b></li> <li>- <b>Analytics libraries</b></li> <li>- <b>Machine learning</b></li> </ul>	<div>  <p>User defined images with Shifter NESAP for data</p> </div> <div> <p>New analytics and ML libraries</p>  </div>	<div>  <p>Complex workflow support</p> </div> <p>Optimised analytics libraries and deep learning frameworks supported at scale</p>
<b>Workflow integration</b>	 <p>Real-time queues</p>	 <p>SLURM co-scheduling Workflow nodes integrated</p>
<b>Data transfer and streaming</b>	<b>SDN</b> 	<p>Slingshot ethernet-based converged fabric</p> 



# In Summary:

---

- HPC Workloads are ***changing***: workflows are becoming more ***complex***, and ***need*** to be accommodated on HPC systems.
- Scientists need support for ***data management, data analysis, sharing and curation*** to enable their science: ***Machine learning*** is transforming science
- Much of this new robust ecosystem, now being deployed on HPC systems, originated ***outside*** the HPC community
- HPC Community has both an ***opportunity*** and a ***need*** to partner with existing communities and develop new capabilities to advance the software ecosystem for science.

*"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."*

The Networking and Information Technology Research and Development  
(NITRD) Program

**Mailing Address:** NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

**Physical Address:** 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674,  
Fax: 202-459-9673, Email: [nco@nitrd.gov](mailto:nco@nitrd.gov), Website: <https://www.nitrd.gov>

