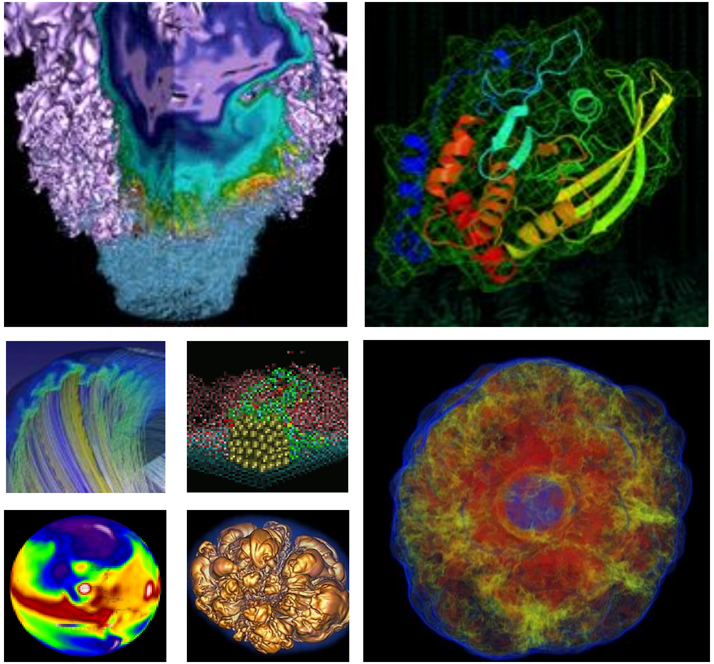
NERSC Data Strategy



Debbie Bard
Lead for Data Science Engagement

ASCAC meeting, March 2019

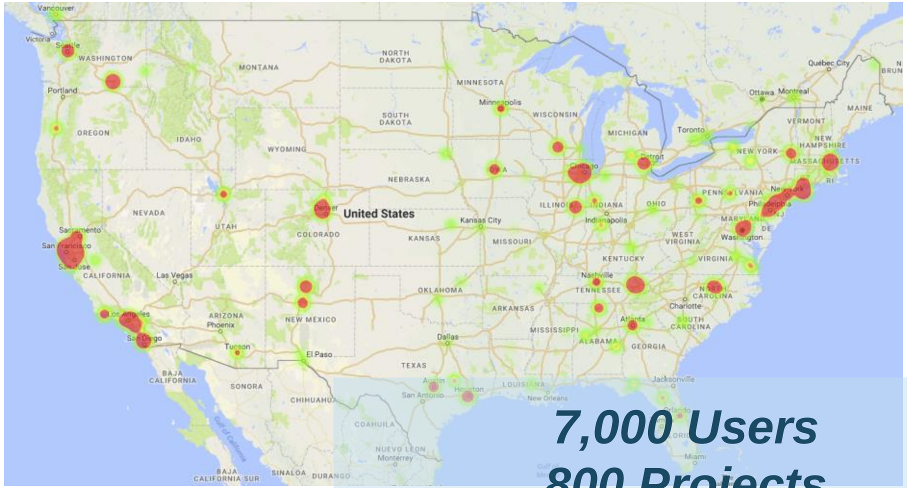
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

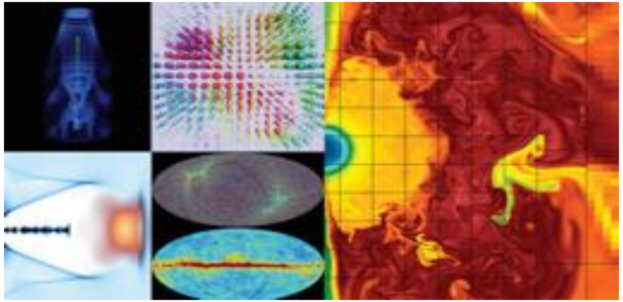
NERSC is the mission HPC and data facility for the Office of Science

U.S. DEPARTMENT OF ENERGY | Office of Science

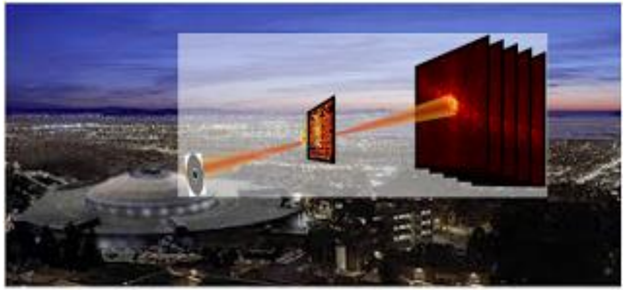
Largest funder of physical science research in U.S.



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



Simulations at scale



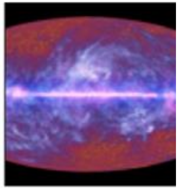
Data analysis support for DOE's experimental and observational facilities

U.S. DEPARTMENT OF ENERGY | Office of Science

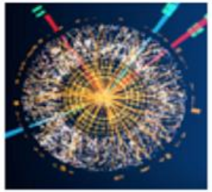
BERKELEY LAB

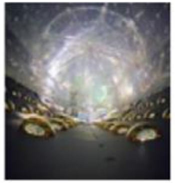
NERSC supports a large number of users and projects from DOE SC's experimental and observational facilities

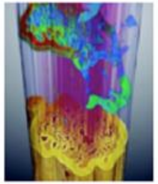

Palomar Transient Factory Supernova


Planck Satellite Cosmic Microwave Background Radiation

Star Particle Physics

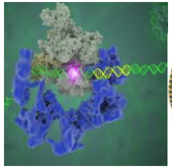

Atlas Large Hadron Collide

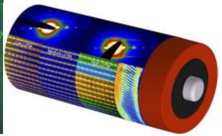

Dayabay Neutrinos

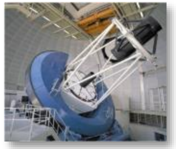

ALS Light Source


LCLS Light Source

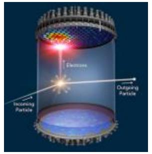

Joint Genome Institute Bioinformatics

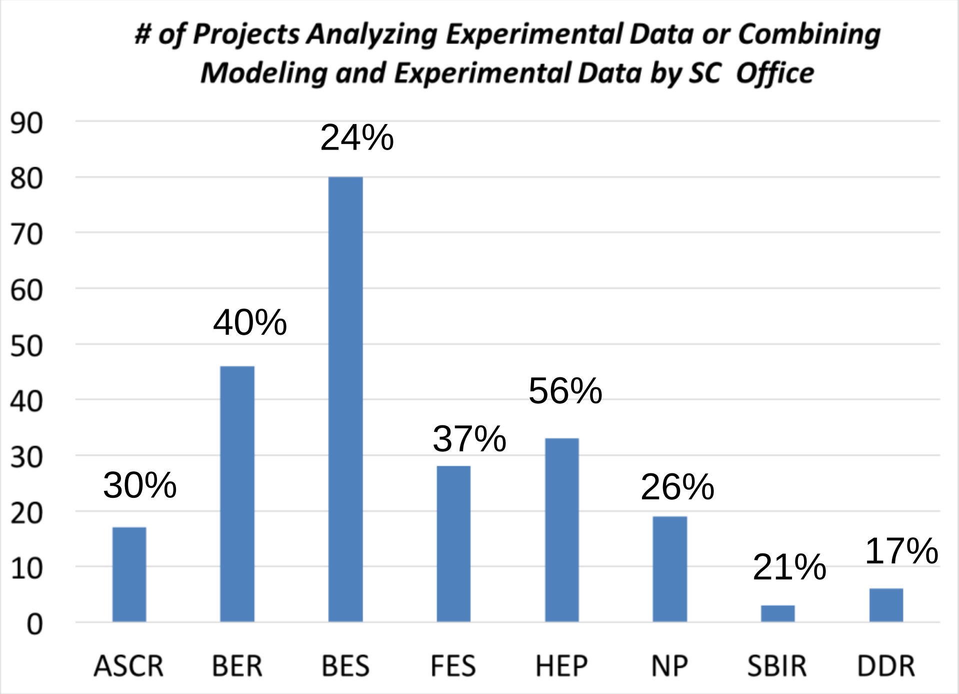

Cryo-EM


NCEM


DESI


LSST-DESC


LZ



~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

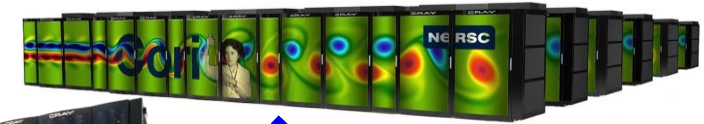
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

NERSC Systems: present and future

NERSC-7:
Edison
Multicore CPU

2013

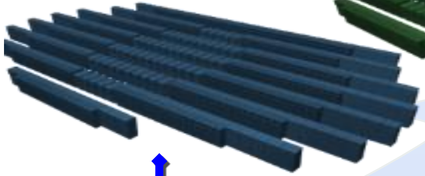


NERSC-8: Cori
Manycore CPU
NESAP Launched: transition applications to advanced architectures

2016

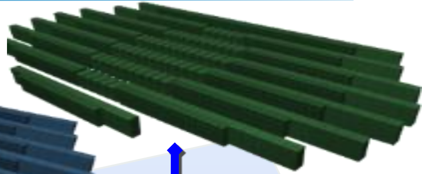
NERSC-9:
CPU and GPU nodes
Continued transition of applications and support for complex workflows

2020



NERSC-10:
Exa system

2024



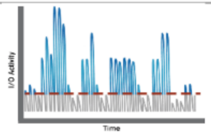
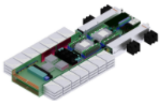
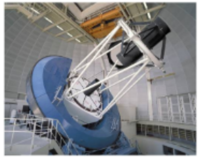
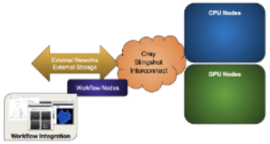
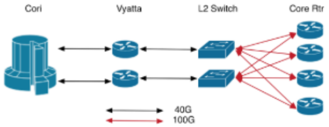
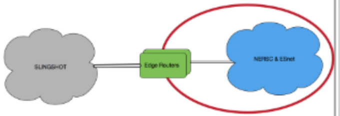
NERSC-11:
Beyond Moore

2028

NERSC-9: 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 network for Terabit-rate connections to system
- Optimised data software stack enabling analytics and Machine Learning at scale
- All-flash file system for accelerated IO

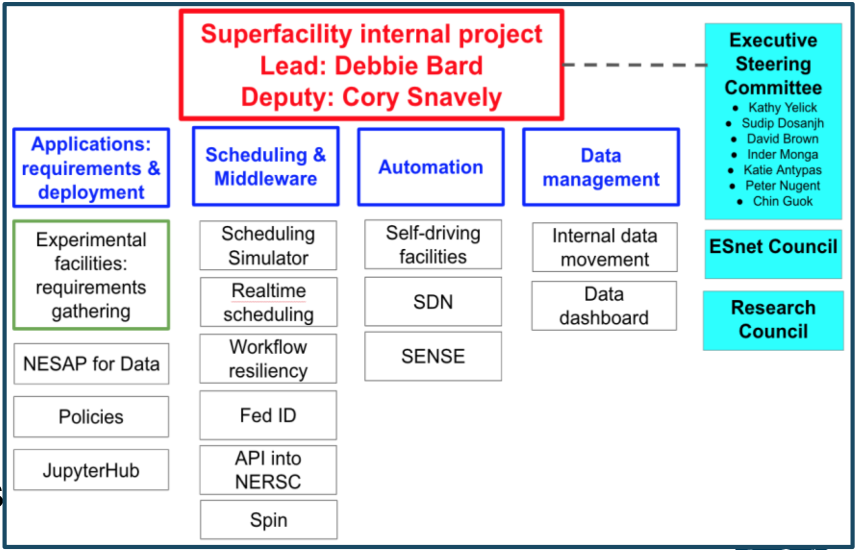
Data Features	Cori experience	N9 enhancements
I/O and Storage	Burst Buffer 	All-flash file system: performance with ease of data management 
Analytics - Production stacks - Analytics libraries - Machine learning	User defined images with Shifter NESAP for data New analytics and ML libraries	 Production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

Projectizing data science support at NERSC

Aim: Provide coherent support for experimental science by coordinating and managing cross-facility tasks at LBNL, including teams across NERSC, ESnet and CRD.

This project will:

- Deploy large scale computing and storage resources at NERSC;
- Define policies that support data science workloads;
- Provide reusable building blocks for experimental scientists to build pipelines;
- Provide scalable infrastructure to launch services;
- Provide expertise on how to optimize pipelines



U.S. DEPARTMENT OF ENERGY | Office of Science

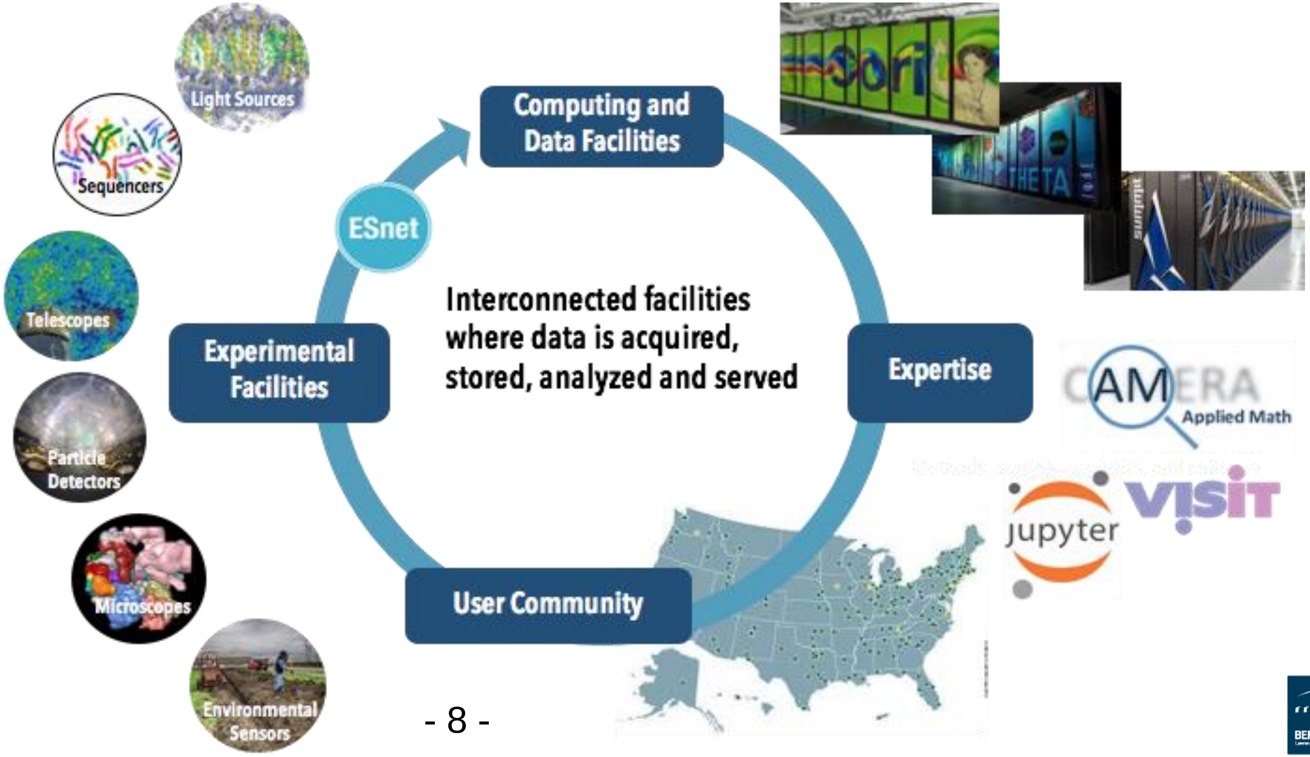
BERKELEY LAB

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

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

BERKELEY LAB

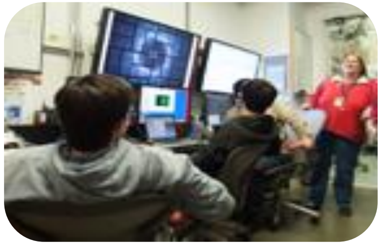
Computing Sciences
Strategic Plan



U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

LBNL CS Strategic Plan: Supporting the Superfacility Model



User Engagement

Engage with experimental, observational and distributed sensor user communities to deploy and optimize data pipelines for large-scale systems.

Data Lifecycle

Mange the generation, movement and analysis of data for scalability, efficiency and usability. Enable data reuse and search to increase the impact of experimental, observational and simulation data.

Automated Resource Allocation

Deliver a framework for seamless resource allocation, calendaring and management of compute, storage and network assets across administrative boundaries.

Computing at the Edge

Design and deploy specialised computing devices for real-time data handling and computation at experimental and computational facilities.

U.S. DEPARTMENT OF ENERGY | Office of Science
ESnet ENERGY SCIENCES NETWORK
CRD
NERSC
BERKELEY LAB

Engagement across the Office of Science

NERSC

User Engagement

Engage with experimental, observational and distributed sensor user communities to deploy and optimize data pipelines for large-scale systems.

ESnet
ENERGY SCIENCES NETWORK

CRD

CAMERA
Applied Math

ECP
EXASCALE COMPUTING PROJECT

Argonne NATIONAL LABORATORY
Leadership Computing Facility

OAK RIDGE National Laboratory
LEADERSHIP COMPUTING FACILITY

SLAC NATIONAL ACCELERATOR LABORATORY

Los Alamos NATIONAL LABORATORY

Pacific Northwest NATIONAL LABORATORY

Jefferson Lab
Thomas Jefferson National Accelerator Facility

BROOKHAVEN NATIONAL LABORATORY

Fermilab

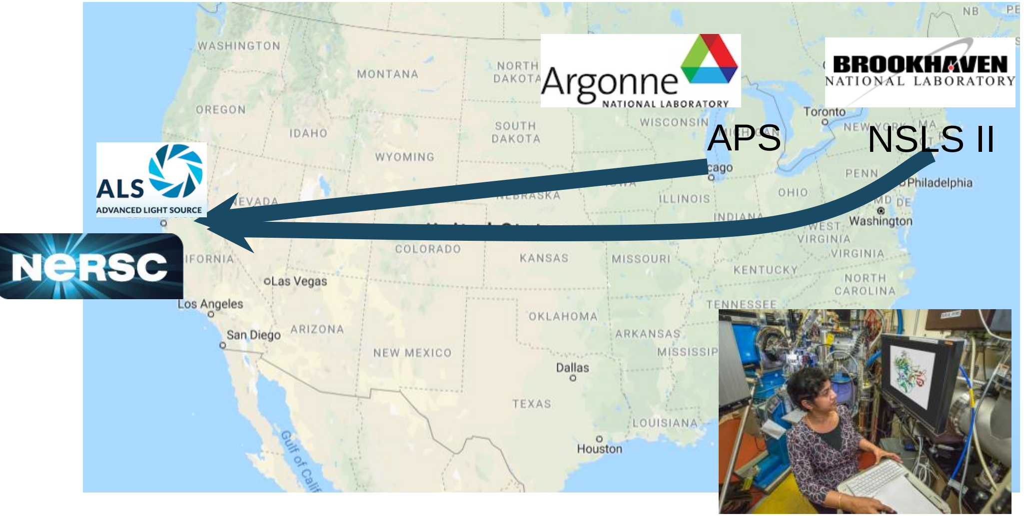
Sandia National Laboratories

NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Lawrence Livermore National Laboratory

FICUS
Facilities Integrating Collaborations for User Science

Engagement: DOE synchrotron light sources



Support multiple coherent & full-field experiments using high frame rate 2D detectors.

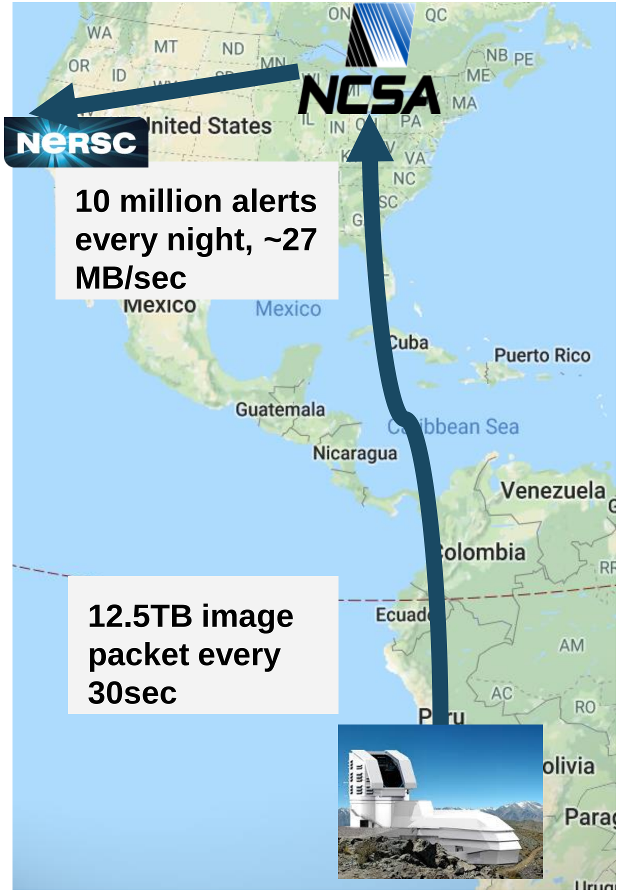
Increasing demands in data volume and computation.

- *50-200MB/s, 30-60 TB raw data per week per detector today*
- *GPUs increasingly used for data processing*

HPC computing and data needs:

- Real-time computing for fast feedback
- Streaming data into compute nodes
- Automated data movement, archiving and retention
- Easy data sharing across multiple facilities and communities

Engagement: LSST DESC



Explain Dark Energy through multiple science probes: Galaxy catalogs, supernovae, lensing

Survey covers the whole sky every few nights using 3.2 Gpix camera built by DOE.

- *10M alerts/night*
- *15 PB catalog data (~0.5 EB total data)*

LSST DESC Dark Energy Science Collaboration

HPC computing and data needs:

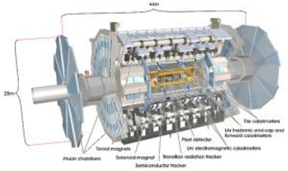
- Large-scale simulation production
- Real-time analysis of streaming data
- Jupyter for data analysis across sites
- Automated data movement, archiving and retention
- Easy data sharing across multiple facilities and communities

Data science at scale

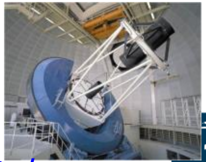
NERSC

- **Big Data Software Stack**
  - Big Data Center
    - NERSC/Intel/Cray/IPCC collaborations
    - Production-level big data software stack that can be used to solve leading scientific challenges at full HPC scale

- **NESAP program for Cori & Perlmutter**
  - NESAP for Simulations (13 projects)
  - NESAP for Data (currently 7 projects)
  - NESAP for Learning (currently 6)



FICUS Facilities Integrating Collaborations for User Science



Machine Learning at Extreme Scale

Scientific data is typically large and complex

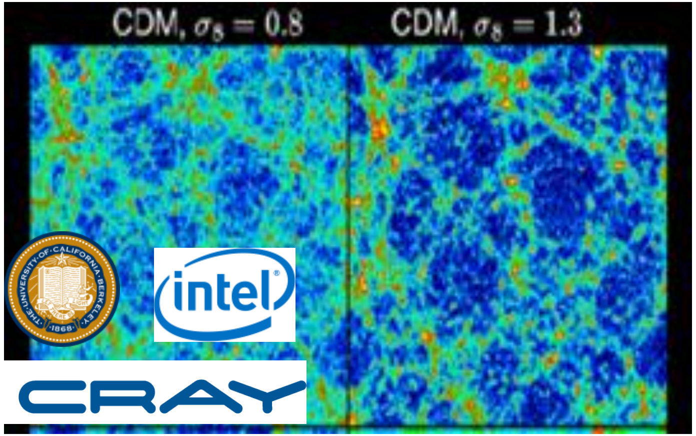
- Harder to find optimal hyperparameters
- Need lots of prototyping and model evaluation

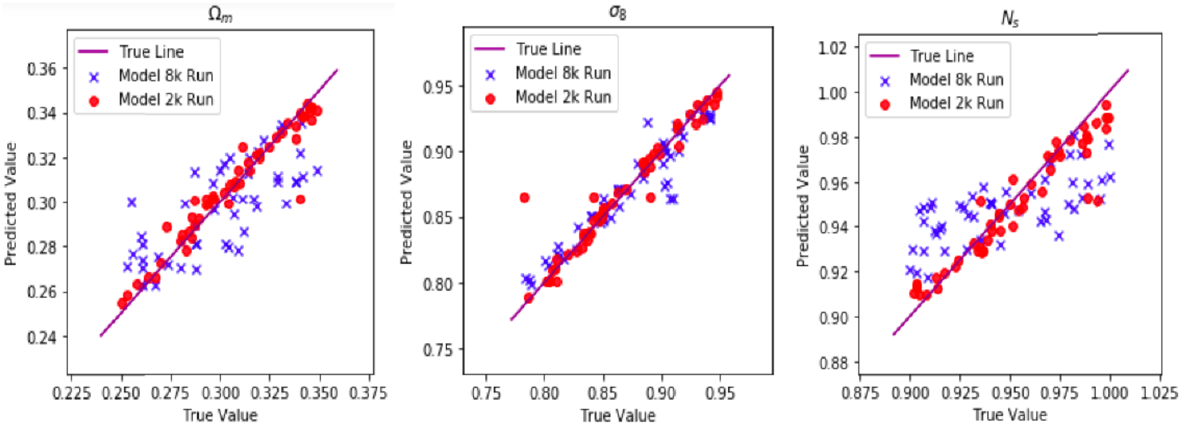
Key metric: *time to scientific insight*

- Don’t want to wait for days to train a single model
- Fast turnaround of ideas and exploration

→ use supercomputers to scale machine learning algorithms for superfast training

ML at scale: determining fundamental constants of cosmology





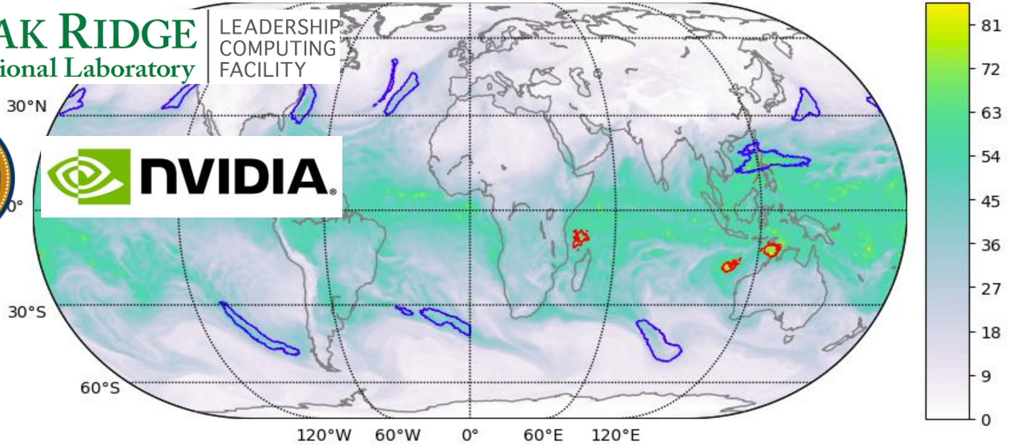
- Achieved unprecedented accuracy in cosmological parameter estimation from the matter distribution in 3D simulation boxes.
- **Scaled to 8192 CPU nodes; 20min training time; 3.5PF sustained performance.**
- Largest application of TensorFlow on CPU-based system with fully-synchronous updates.

U.S. DEPARTMENT OF ENERGY | Office of Science

https://arxiv.org/abs/1808.04728

BERKELEY LAB

ML at scale: Characterising Extreme Weather in a Changing Climate



- High quality segmentation results to identify extreme weather events.
- Network scaled out to 4560 Summit nodes (27,360 Volta GPUs).
- **60min training time, 0.99 EF sustained performance in 16-bit precision.**
- **Largest application of TensorFlow on GPU-based system, first Exascale DL application.**

https://arxiv.org/abs/1810.01993

Thank You