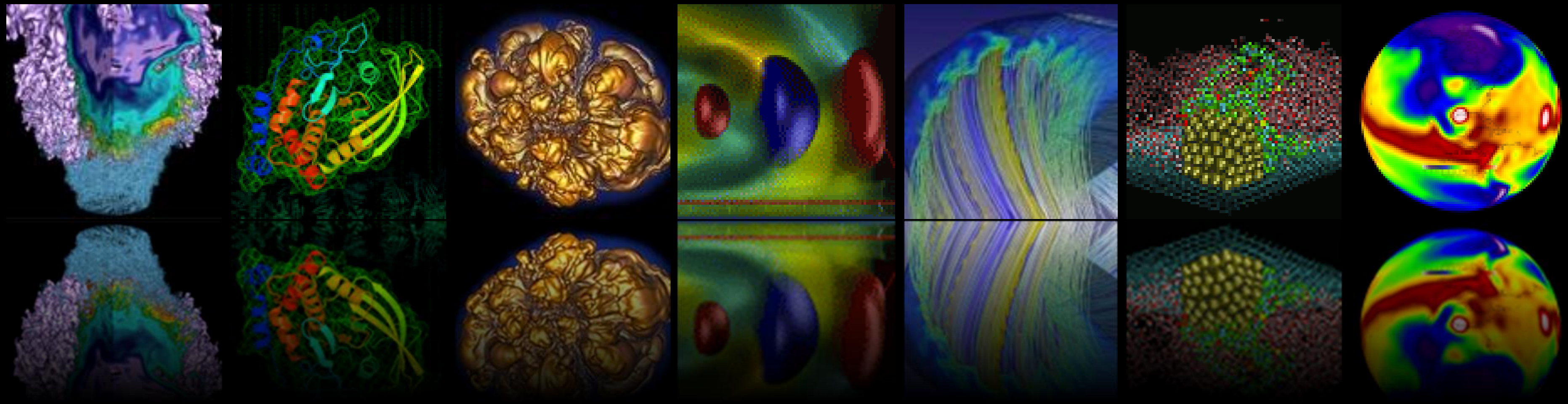


A Superfacility for Data Intensive Science

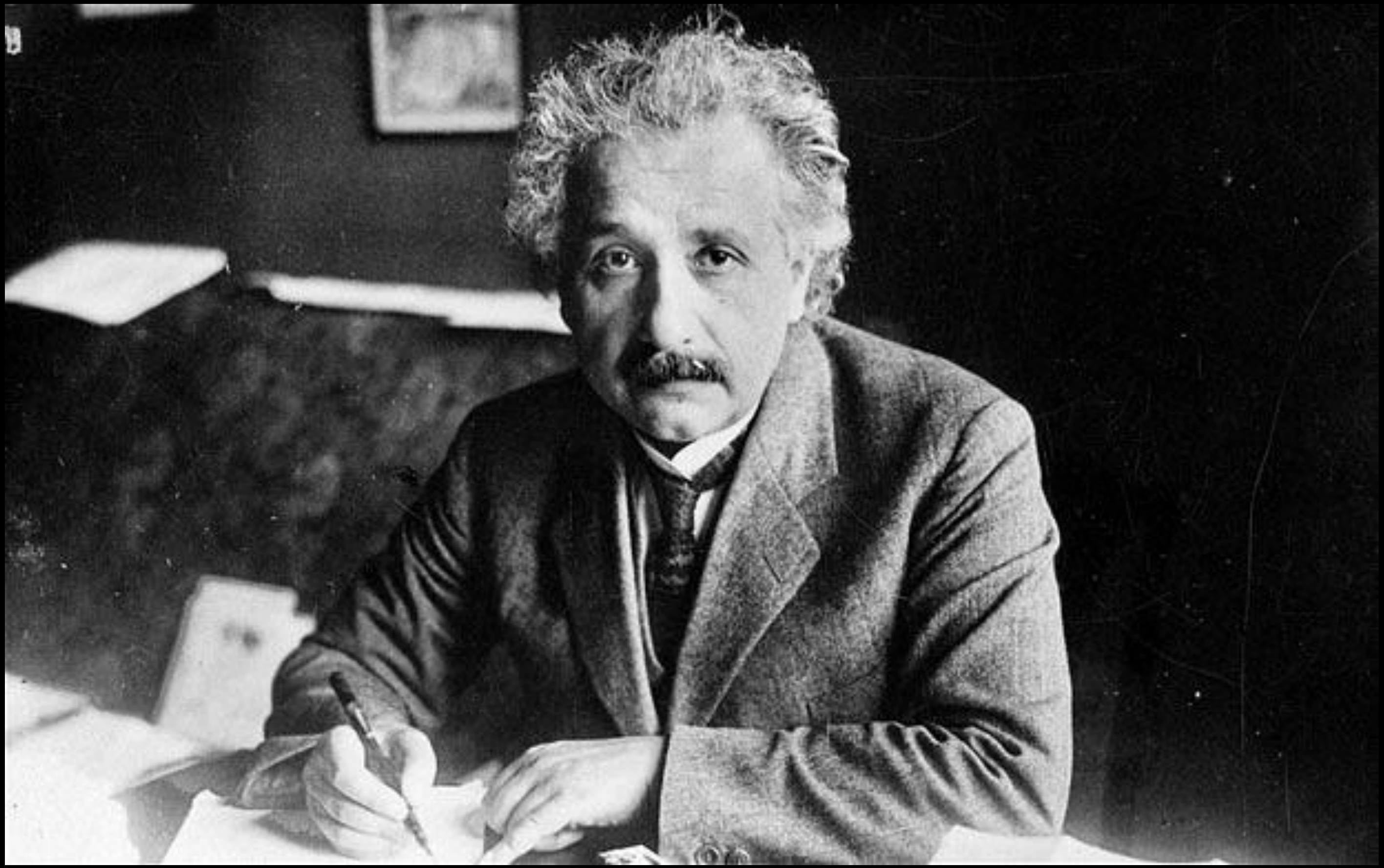
Kathy Yelick

Associate Laboratory Director for Computing Sciences
Lawrence Berkeley National Laboratory
Professor of Electrical Engineering and Computer Sciences
University of California at Berkeley

Part 1

Science is poised for transformatio

Old School Scientists: The Lone Scientist



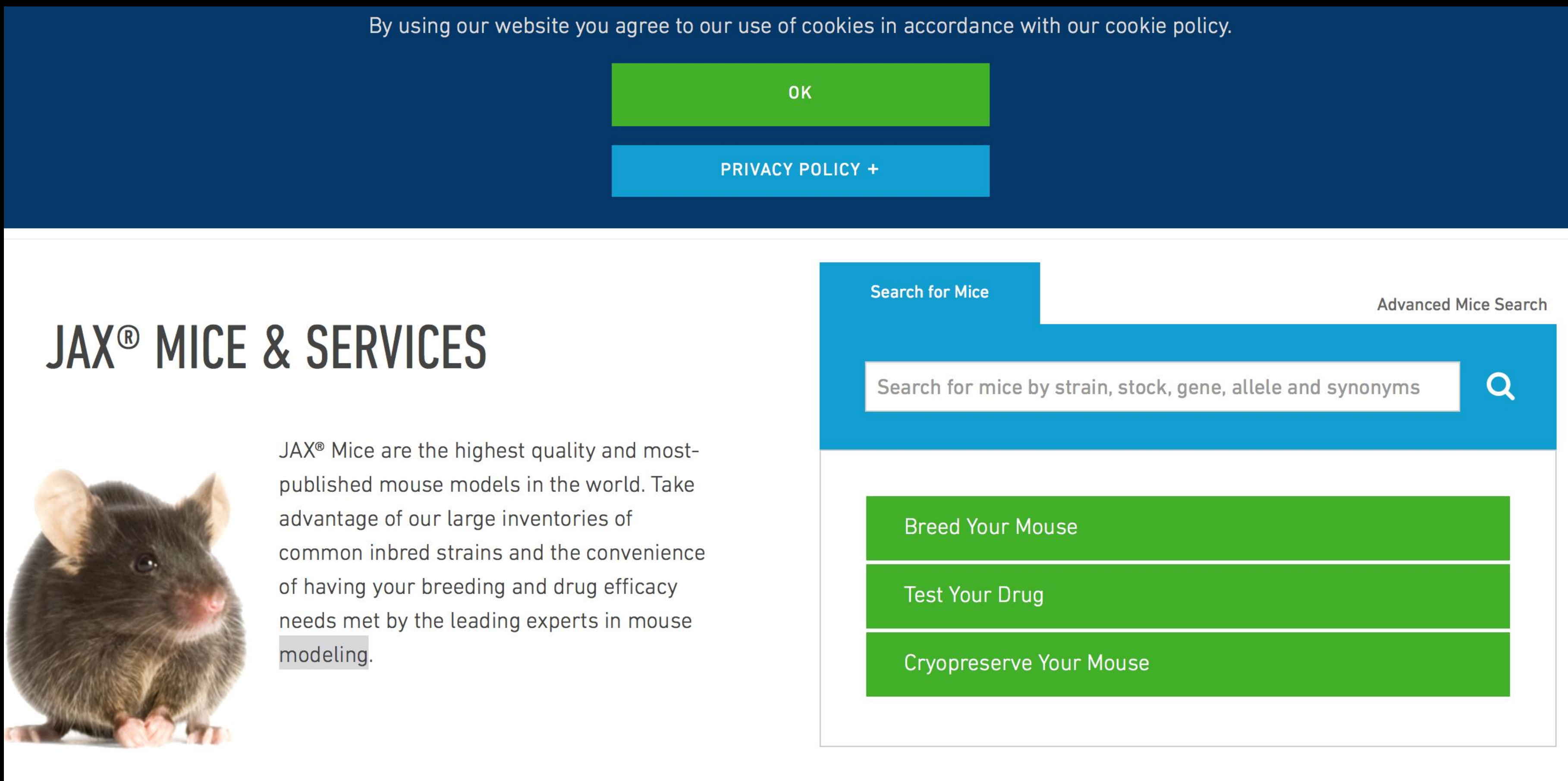
Team Science

New Scientists

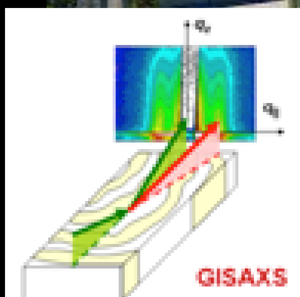
17-year-old Brittany Wegner creates breast cancer detection tool that is 99% accurate on a minimally invasive, previously inaccurate test.

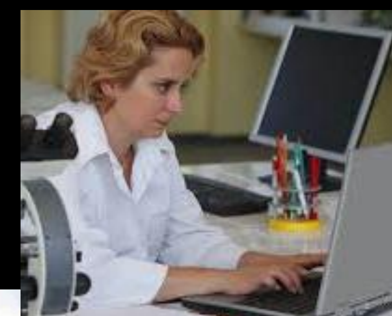
Machine Learning + Online Data + Cloud Computing

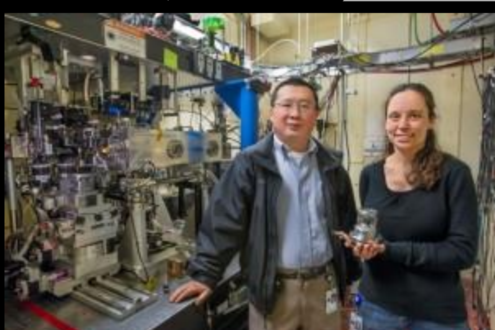
Experimental Science is Changing



Old School Scientific Workflow



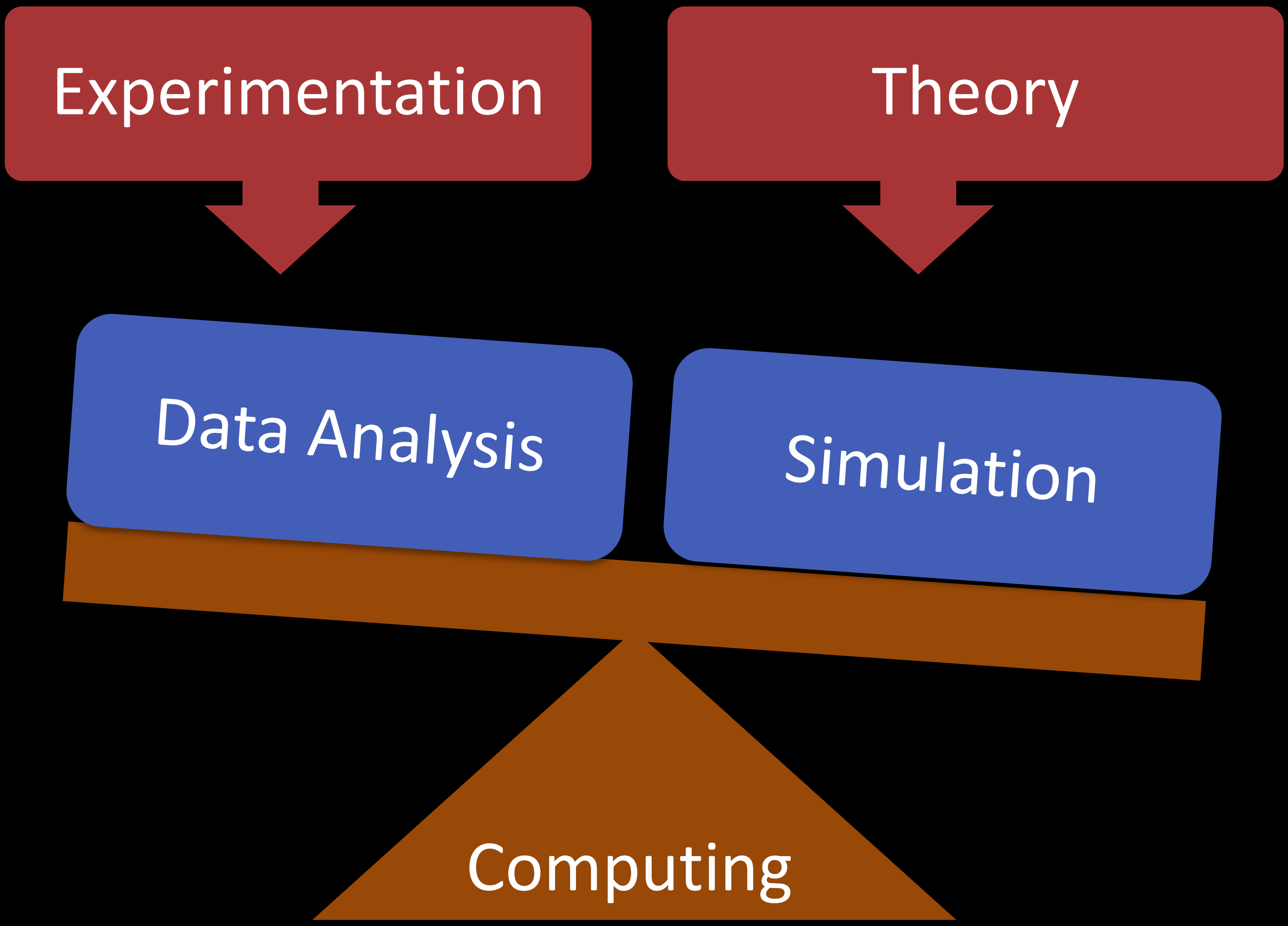




Computing, experiments, networking and expertise in a "Superfacility" for Science

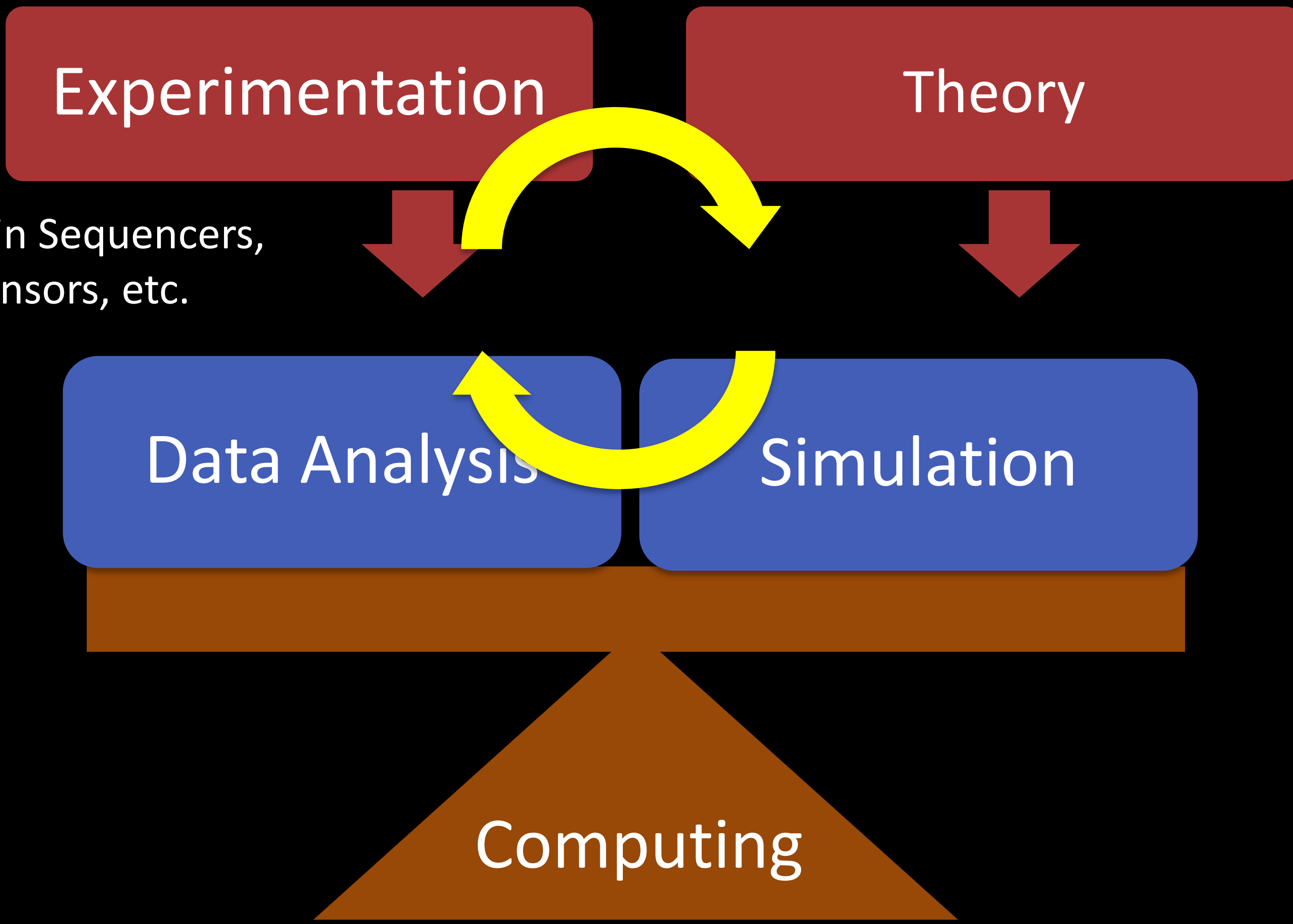
Liu et al, "Fast printing and in situ morphology …". Adv Mater. 2015

Old School HPC: only for Simulation

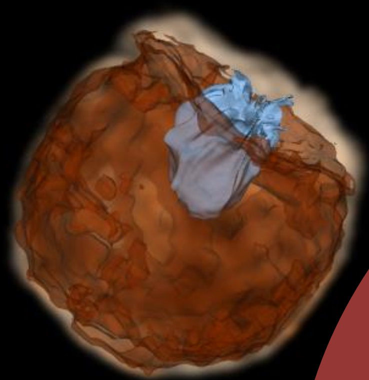


HPC is equally important in experimentation

Growth in Sequencers, CCDs, sensors, etc.

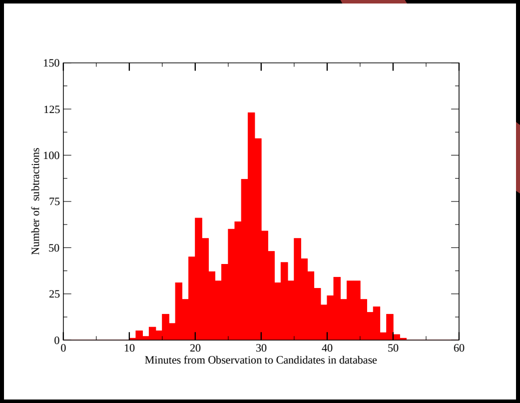


Integration of Simulation and Observational Science



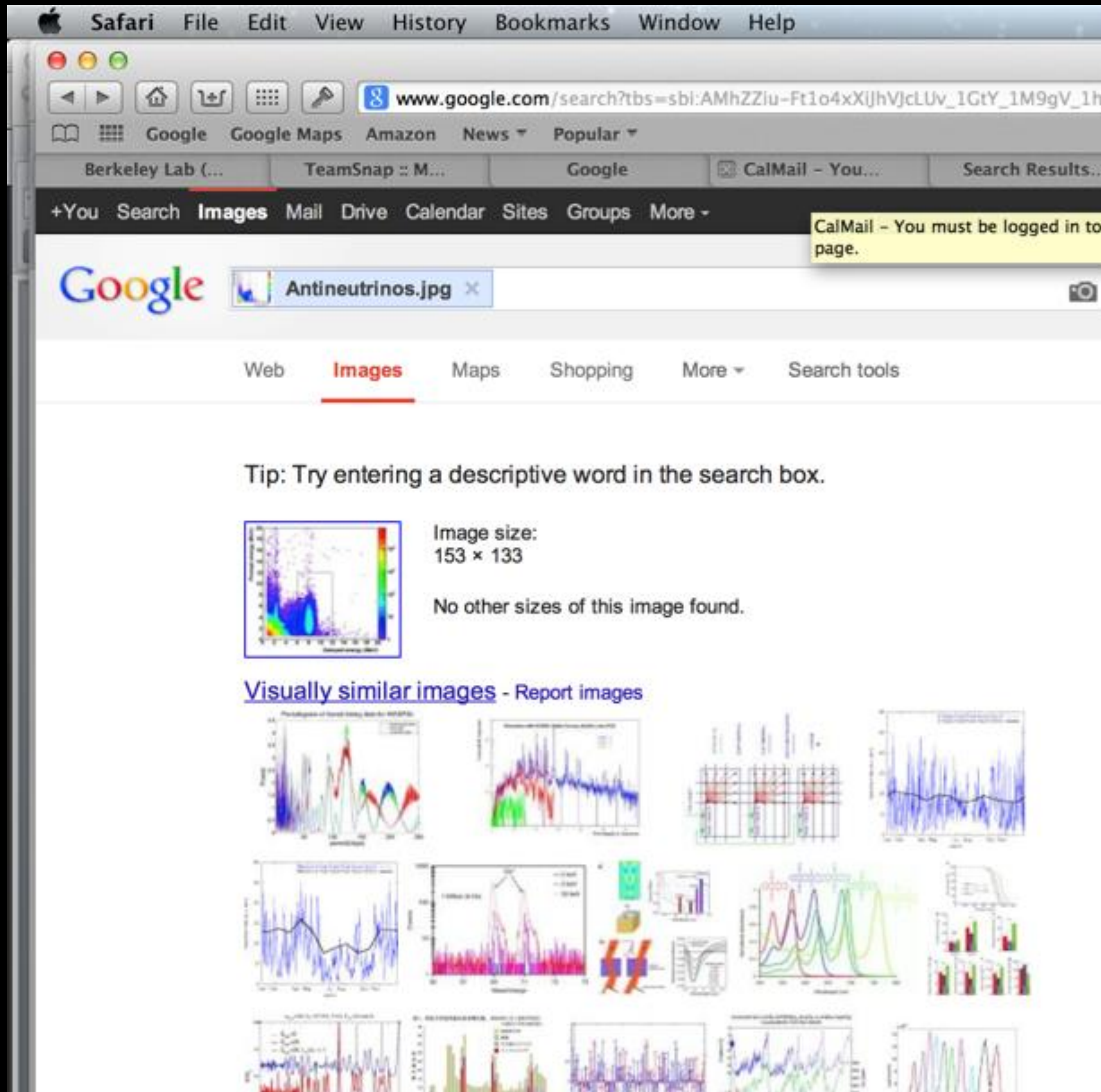
Intermediate Palomar Transient Factory

- **Nightly images transferred**
- **Subtractions, machine learning**
- **Candidates in database in < 5 minutes**
- **Simulations aid in interpreting data**



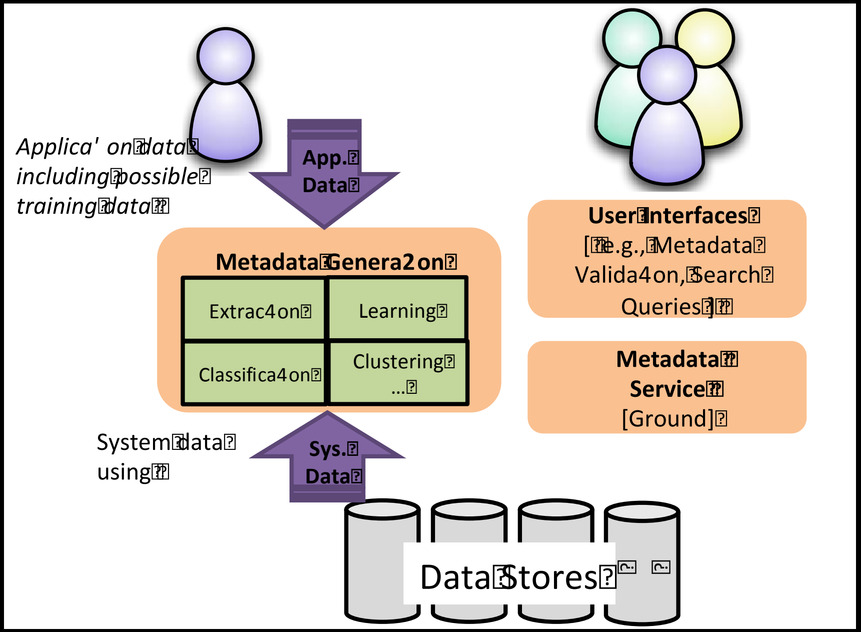
Yi Cao, *et al*. (2015) *Nature*,
"A strong ultraviolet pulse from
a newborn Type Ia supernova"

Old School Scientific Data Search



Automated Search, Meta-Data Analysis, and On-Demand Simulation

Automated metadata extraction using machine learning

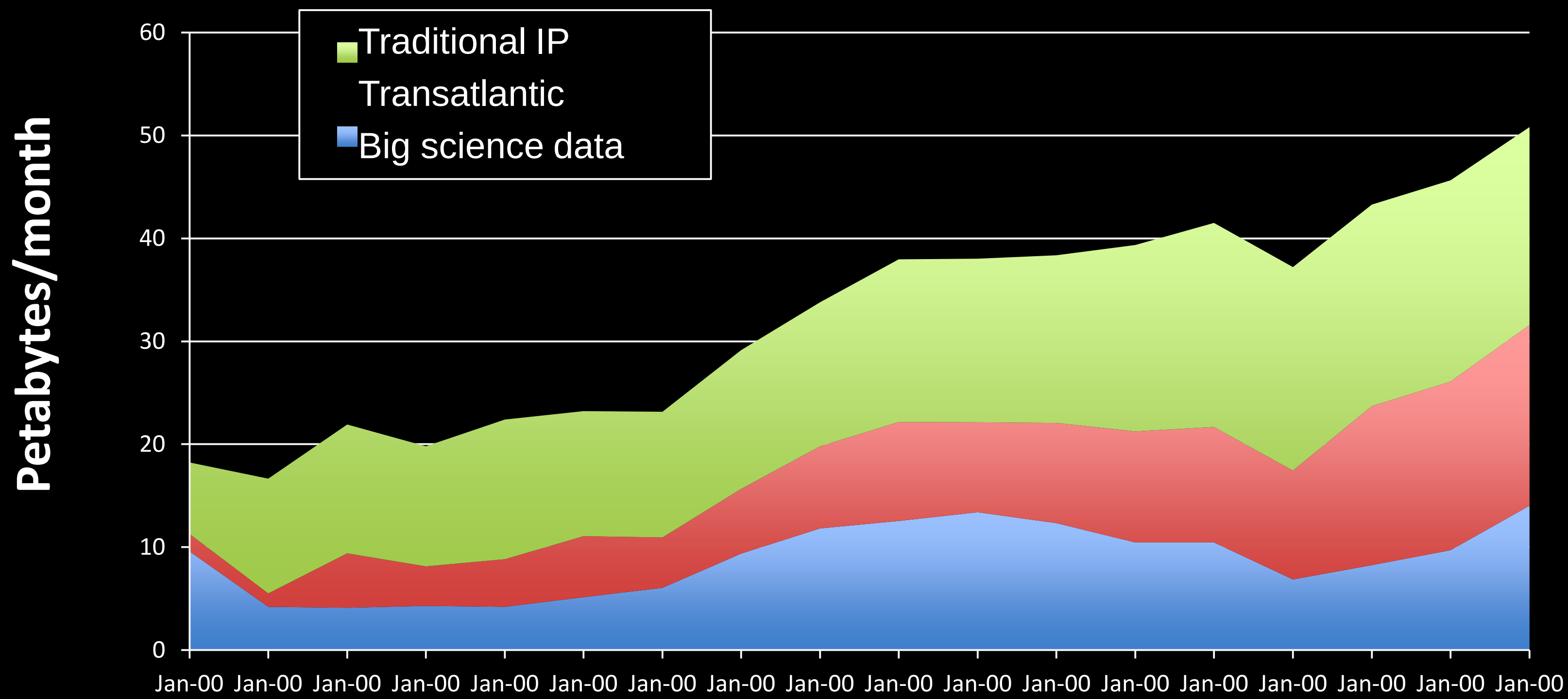


Jobs submitted by "bots" based on queries; algorithms extract informatics for design

Part 2

ASCR Facilities need to adapt

ESnet: Exponential data growth drives capacity

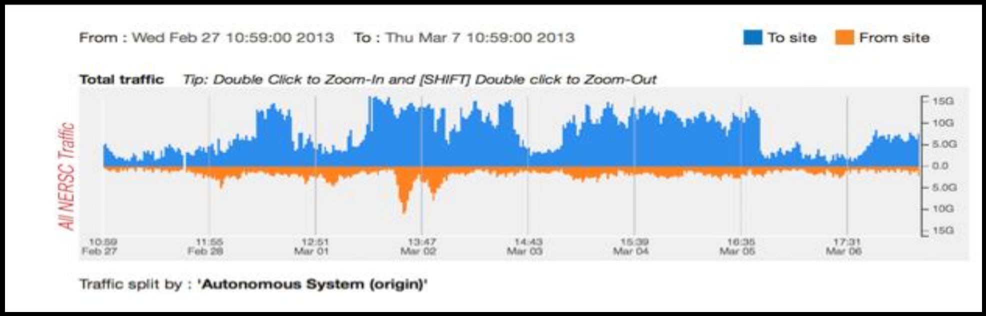


Science DMZ to deliver bandwidth to the end users

OSCARS for bandwidth reservation

100 Exabytes/year by 2024!

ESnet: Discovery Unconstrained by Geography



LCLS/NERSC/Esnet Superfacility demo for Photosystem II

→ 3x network traffic

ESnet-6 Upgrade Options trade off risk and capability

Software Defined Networking

- Programmable switches may improve cost and speed
- Adapt lower level network layers for major science flows

Packet Optical

- Combine hardware for packetization/routing with optical transport
- Lower cost

Current Architecture

- Keep packet and optical separate with current fixed routing tables
- Known technology

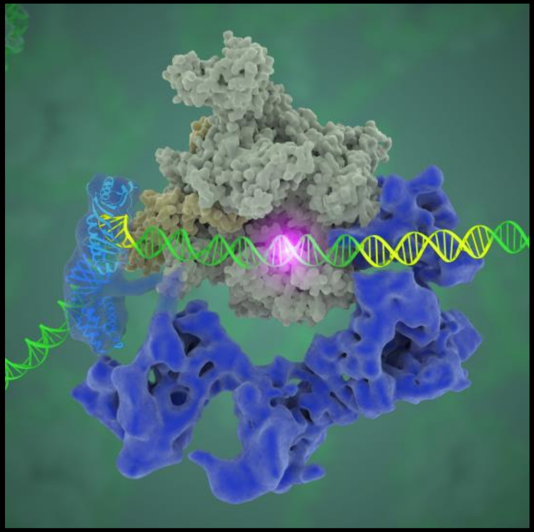
Network performance enables efficiency of centralized computing

Systems configured for data-intensive science

NERSC Cori has data partition (Phase 1, Haswell) pre-exascale (Phase 2, KNL preproduction)
WAN-to-Cori optimized for streaming data: 100x faster from LCLS to Cori and Globus to CERN

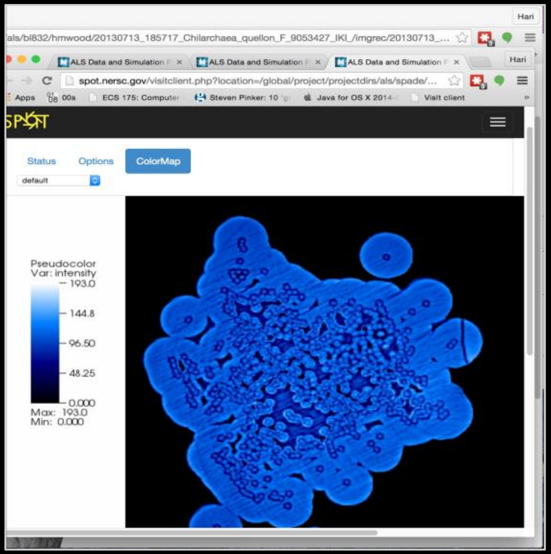
Real-time queue prototyped at NERSC

- **In 1998 dedicated hardware; now prototype queue on Cori**
- **<1% of NERSC allocation**
- **Cryo-Em, Mass spec, Telescopes, Accelerator, Light sources**



Cryo-EM: Image classification
Nogales Lab

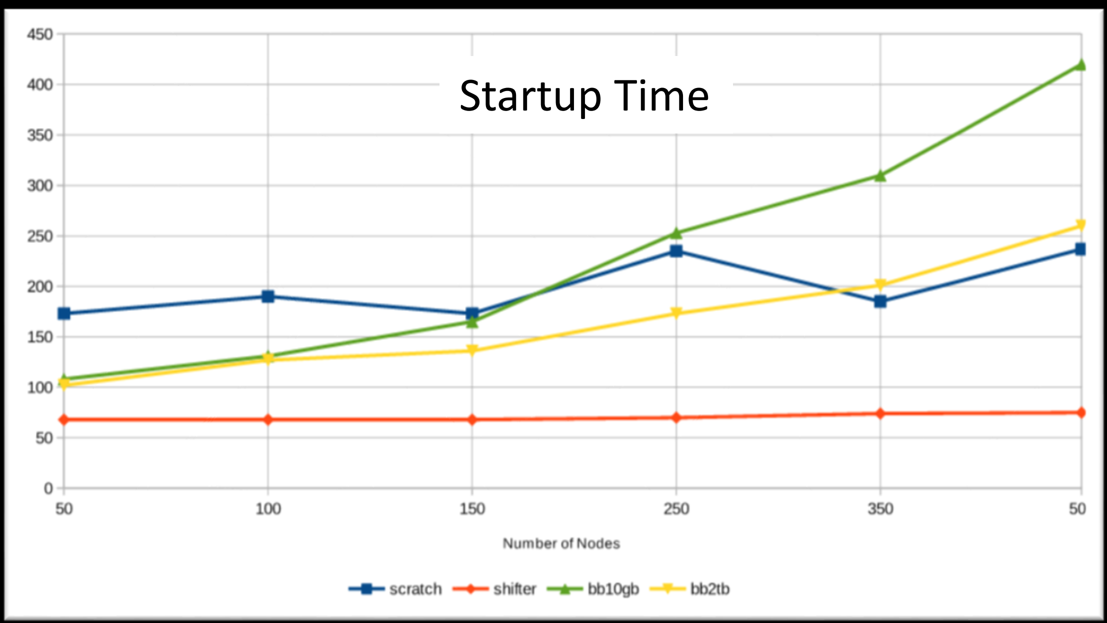
PTF: Image subtraction pipeline



ALS: 3D Reconstruction,
rendered on SPOT web portal

Containers for HPC Systems

- Data analysis pipelines are often large, complex software stacks
- NERSC Shifter (with Cray), supports containers for HPC systems
- Used in HEP and NP projects
  (ATLAS, ALICE, STAR, LSST, DESI)



SHIFTER docker


The Register®
Biting the hand that feeds IT
DATA CENTER SOFTWARE NETWORKS SECURITY INFRASTRUCTURE DEVOPS BUSINESS HARDWA

Data Center ▸ HPC

Cray hoists Docker containers into supercomputers

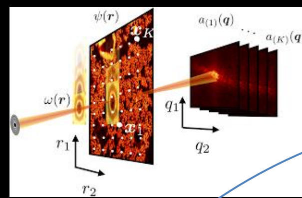
Productivity gains without performance hits

18 Nov 2015 at 00:01, Drew Cullen 25 35

Part 3

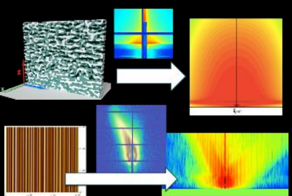
ASCR Research challenges are substantial

CAMERA: Math for the Facilities



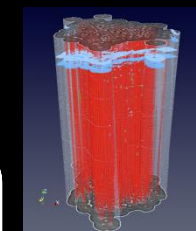
Designing mathematical algorithms to allow real-time analysis next to the equipment

Real-time streaming ptychography—ALS, delivered to NSLS2, LANL, BESSY,



SFM/TEM + GISAXS

Multi-modal: Building the math that fuses information from multiple experiments

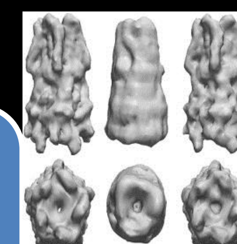


New algorithms to transform manual into automatic analysis

Automatic image processing for the ALS/GE

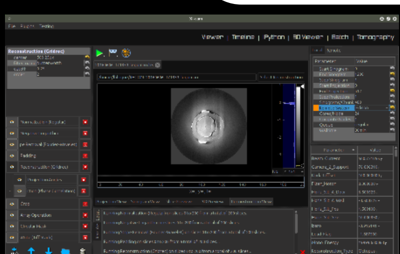
CAMERA workshop on Tomography: Joint with APS, ESRF, DIAMOND, LNLS, LLNL, SSRL,….,

Compare and integrate multiple analysis tools



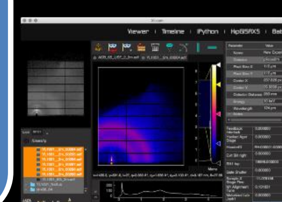
Inventing new math and models to match new acquisition technologies

Fluctuation scattering and single particle imaging for the LCLS



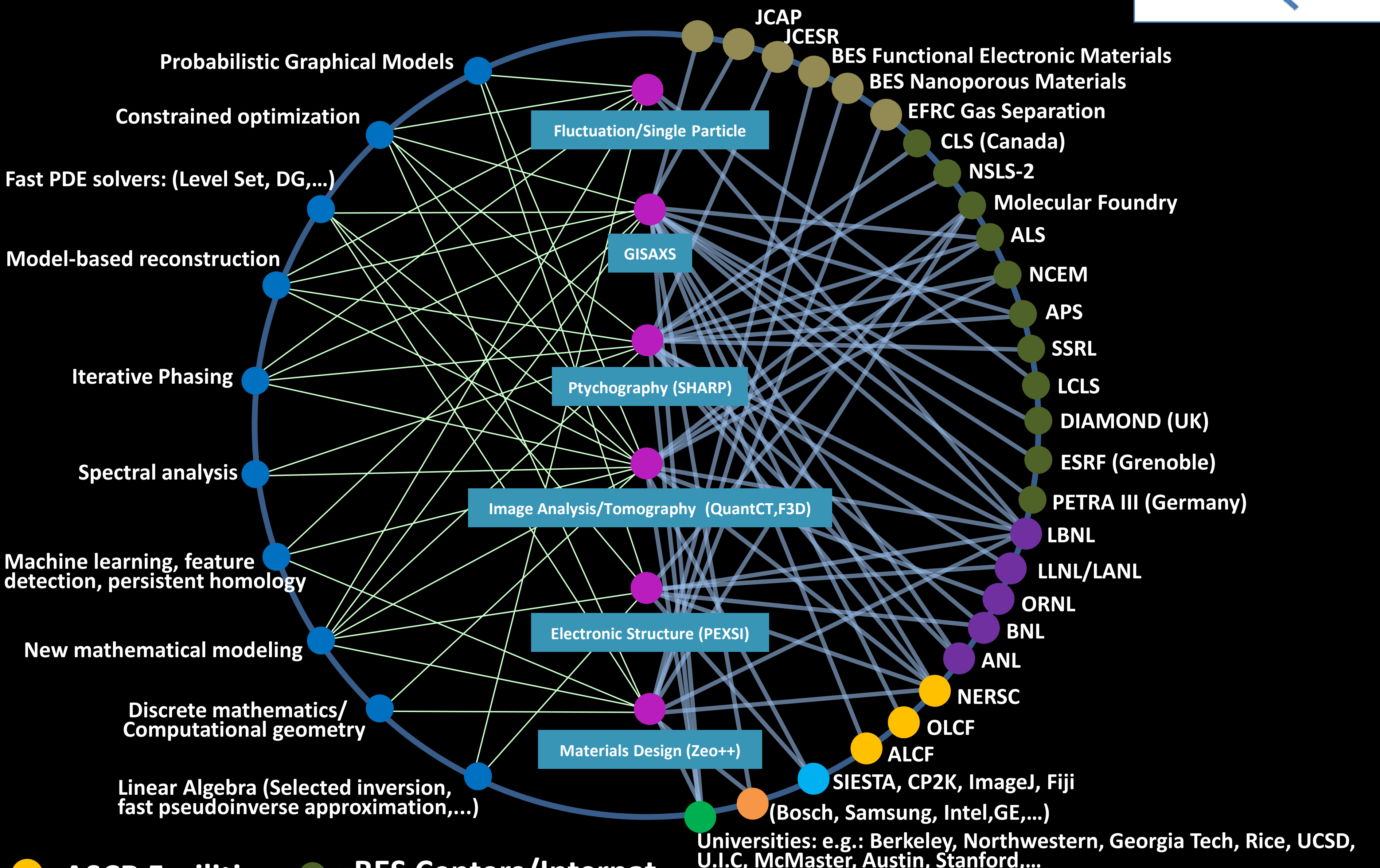
Cultural and Sociological Challenges

Robust and reliable codes and data flow: workflow environments



Workflow and access to remote supercomputers: XiCAM for ALS, SSRL, APS, NSLS2

CAMERA: Making the connections



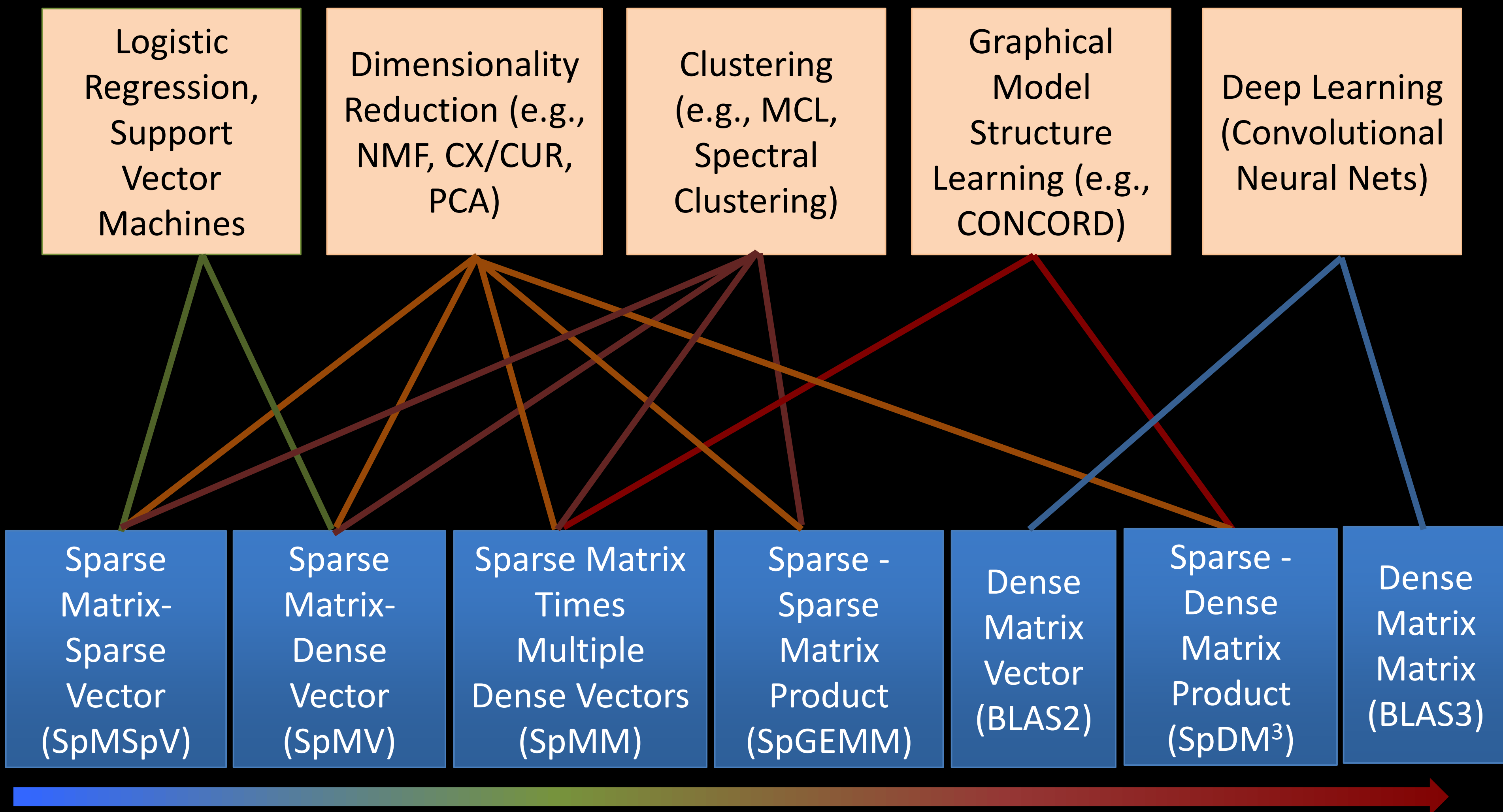
=ASCR Facilities =BES Centers/Internat.

=BES (and other) Facilities =LABS =3[Rd] Party Codes =Industry =Universities

Analytics vs. Simulation Kernels:

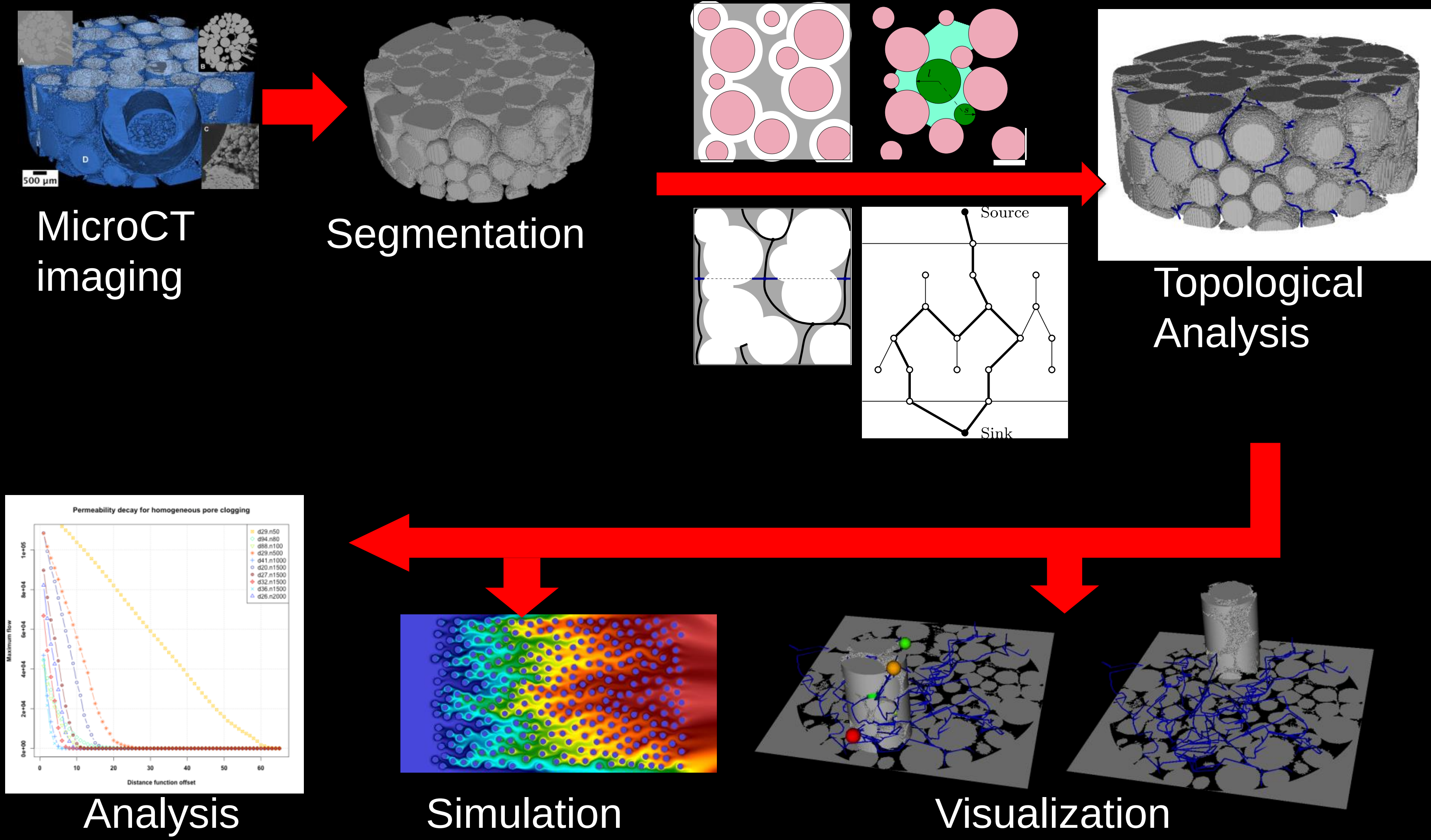
7 Giants of Data	7 Dwarfs of Simulation
Basic statistics	Monte Carlo methods
Generalized N-Body	Particle methods
Graph-theory	Unstructured meshes
Linear algebra	Dense Linear Algebra
Optimizations	Sparse Linear Algebra
Integrations	Spectral methods
Alignment	Structured Meshes

Machine Learning Mapping to Linear Algebra

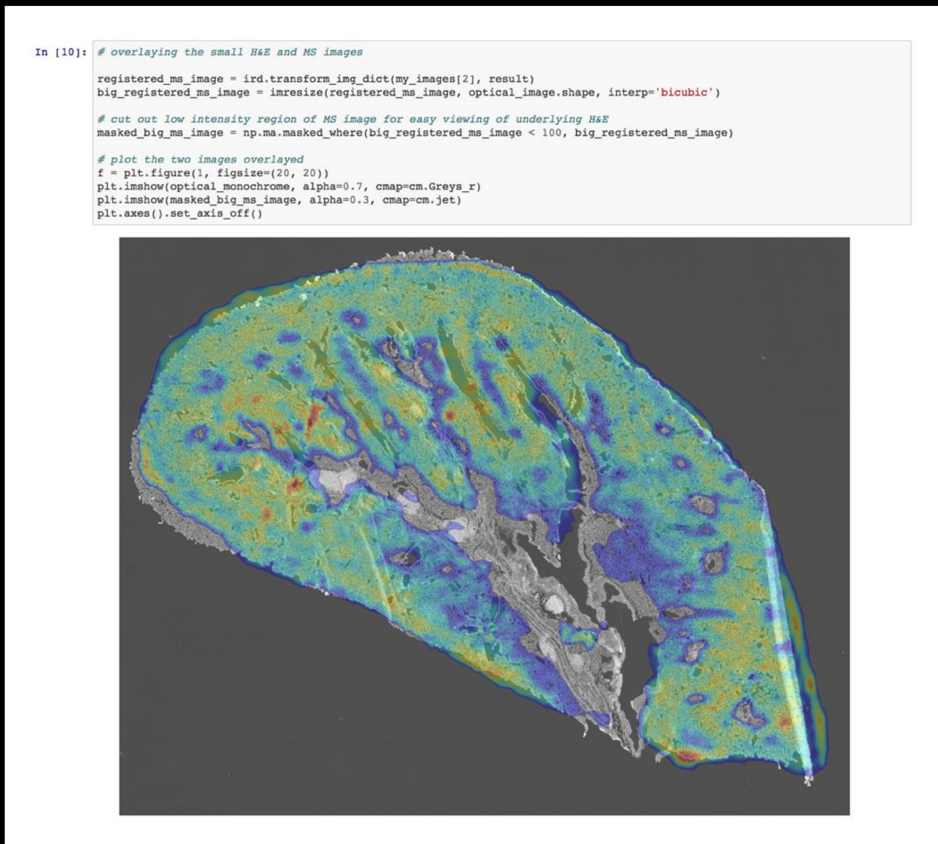


Aydin Buluc

Software implementations at scale in pipeline



Interactive Analytics using Jupyter



Science notebooks through Jupyter (iPython)

- Widely used in science
- Interactive HPC LDRD

Deployed at NERSC:

- >100 users pre-production

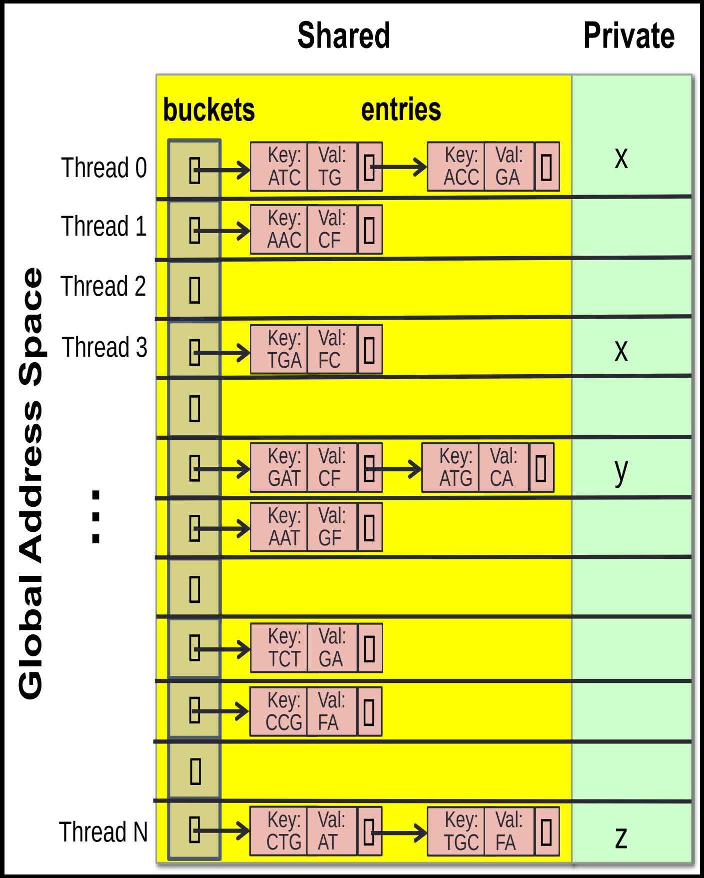
Fernando Perez et al

Random Access Analytics

- Genome assembly ~~"needs shared memory"~~

Global Address Space

- Low overhead communication
- Remote atomics
- Partitions for any structure



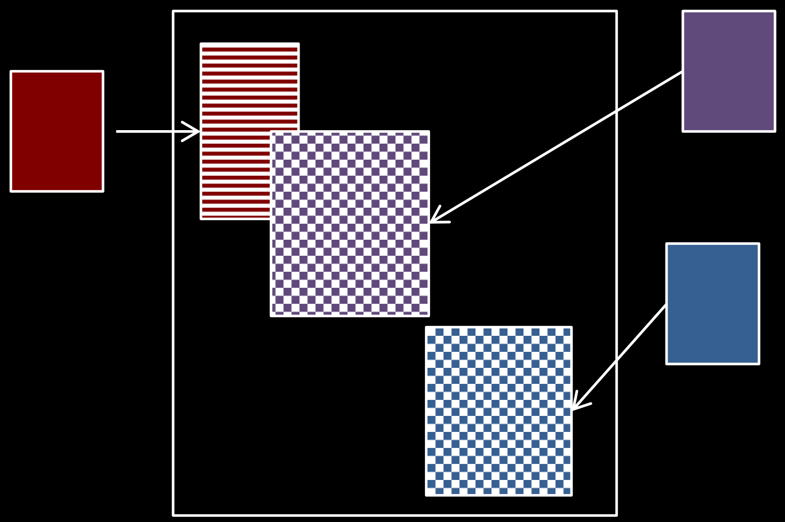
Scales to 15K+ cores

Under 10 minutes for human

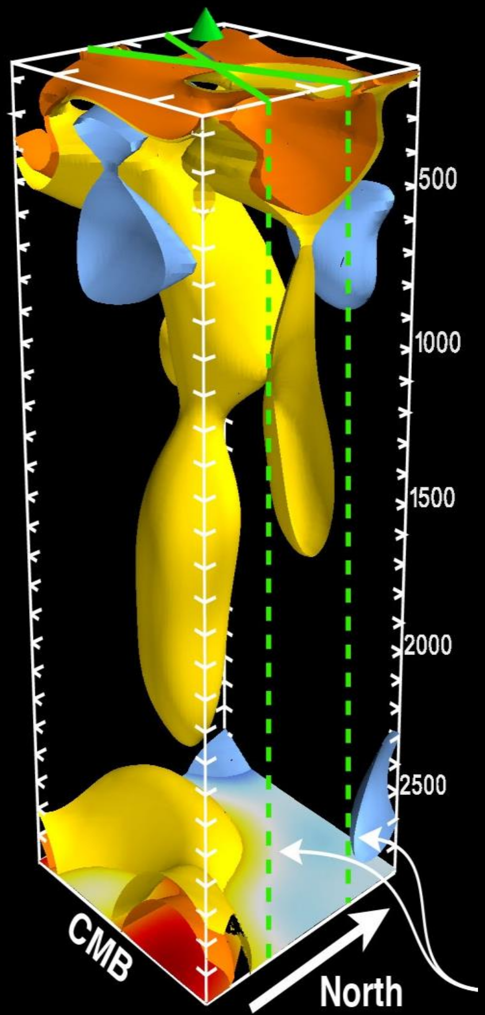
First ever solution

E. Georganas, A. Buluc, J. Chapman, S. Hofmeyr, C. Aluru, R. Egan, L. Oliker, D. Rokhsar, K. Yelick

Data Fusion for Observation with Simulation



- **Unaligned data from observation**
- **One-sided strided updates**



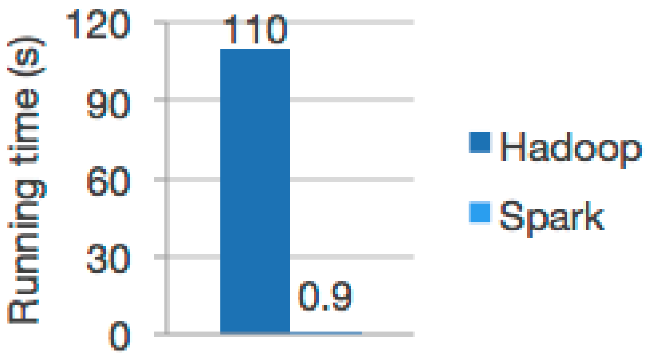
Hawaii hotspot geology

Scott French, Y. Zheng, B. Romanowicz, K. Yelick

Productive Programming

Speed

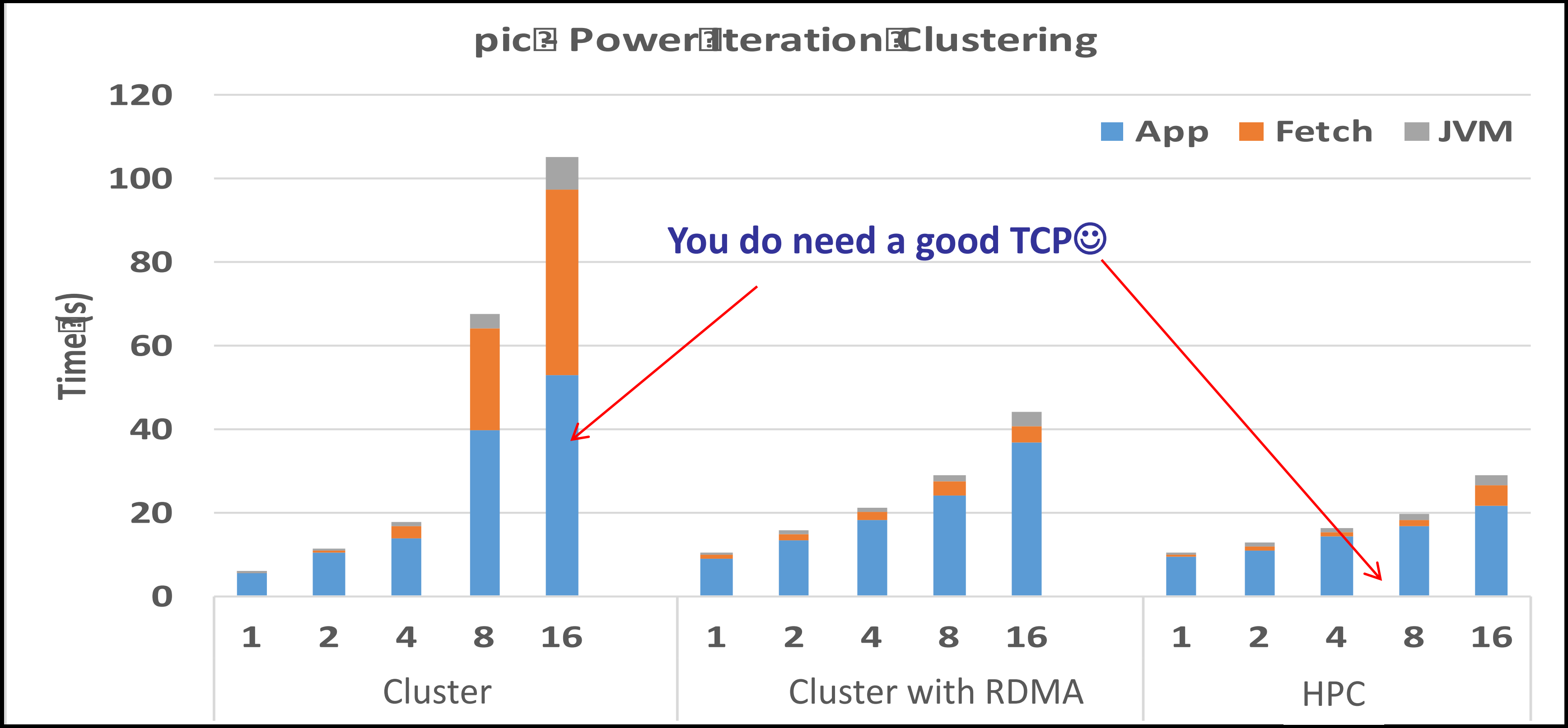
Run programs up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk.



- **High failure rate**
- **Slow network**
- **Fast (local) disk**

And Spark is still 10x+ slower than MPI

SPARK Analytics on HPC

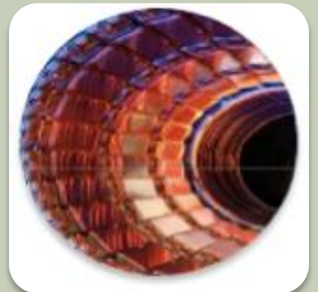


SPARK on HPC vs. clusters

Network, I/O, and virtualization all key to performance

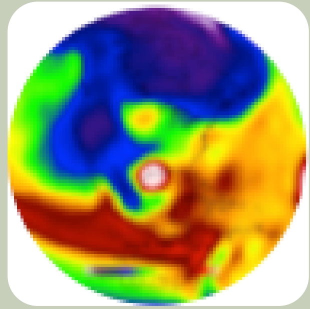
Chaimov, Malony, Iancu, Ibrahim, Canon, Srinivasan

Architectures for Data vs. Simulation

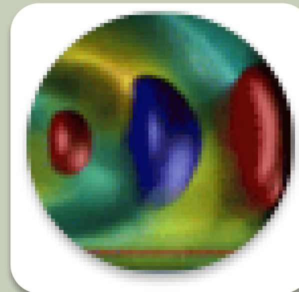


Massive Indepen-dent Jobs for Analysis and Simulation

Compute-Intensive Dense LA for Deep Learning and Simulation



Nearest Neighbor Simulation



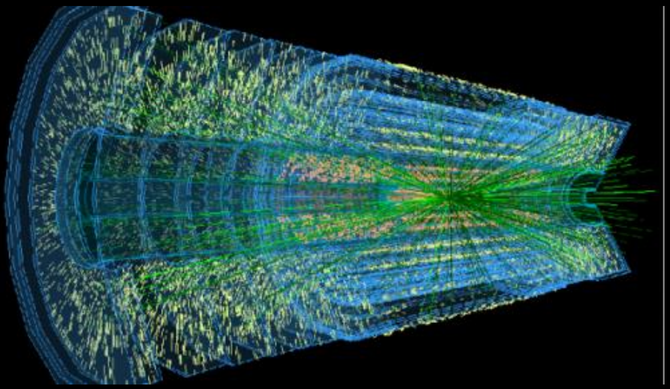
All-to-All Simulation (3D FFTs) and analysis

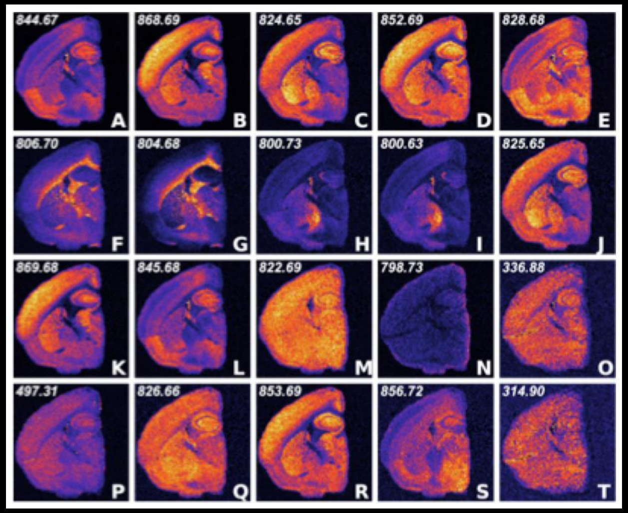
Random access, large data Analysis

Different architectures for simulation? Can simulation use data architectures?

Data processing with special purpose hardware

- General trend towards specialization for continued performance growth
- Data processing (on raw data) will be first in DOE



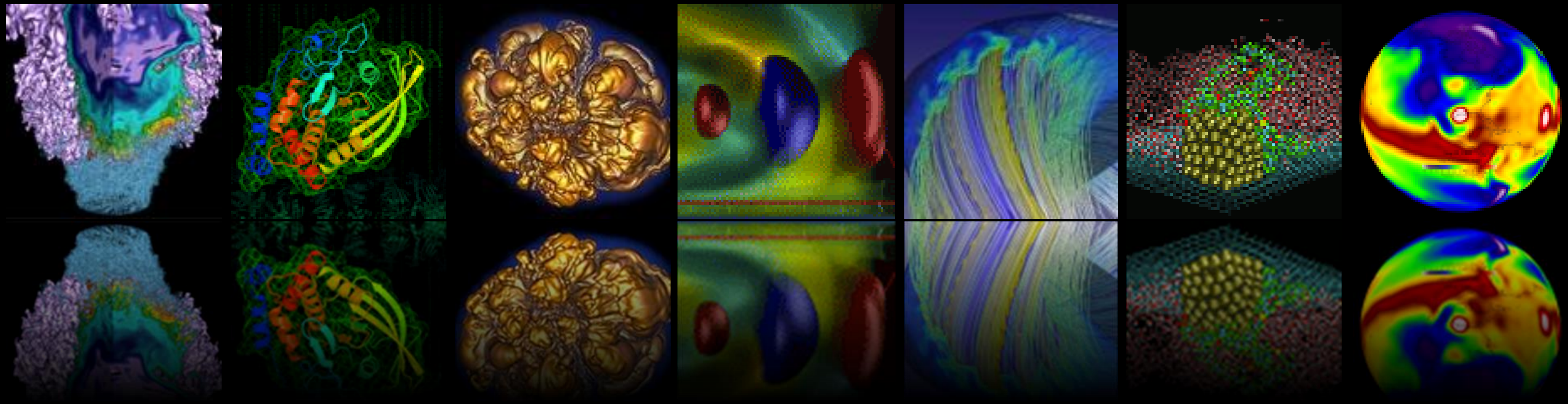


Particle Tracking with Neuromorphic chips

Deep learning processors for image analysis

Computing in Detectors

FPGAS for genome analysis



Extreme Data Science

The scientific process is poised to undergo a radical transformation based on the ability to access, analyze, simulate and combine large and complex data sets.

Superfacility: Integrated network of experimental and computational facilities and expertise