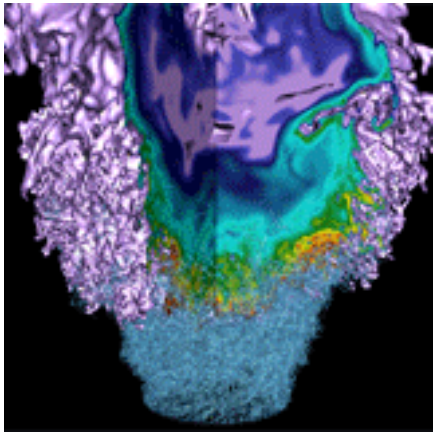
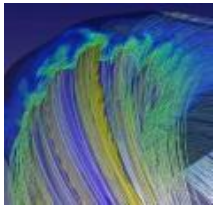
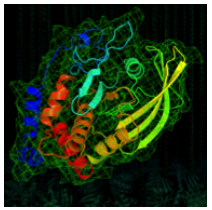
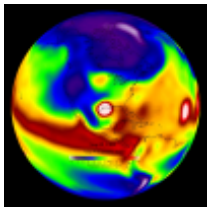
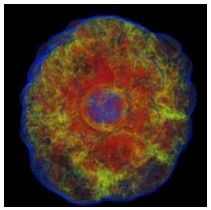
Data and Analytics Services Overview

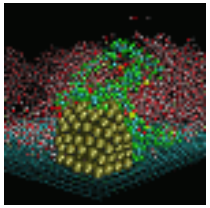












Wahid Bhimji

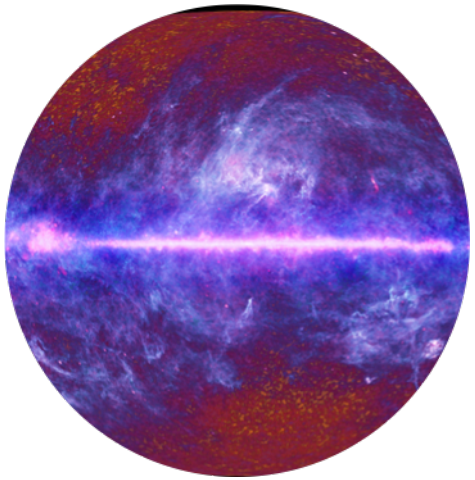
Data and Analytics Services

August 13th, 2015

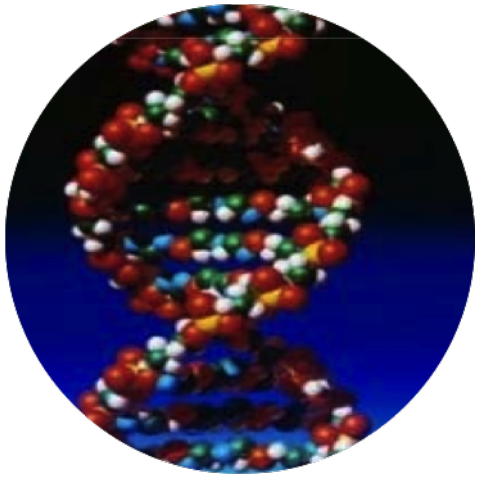
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

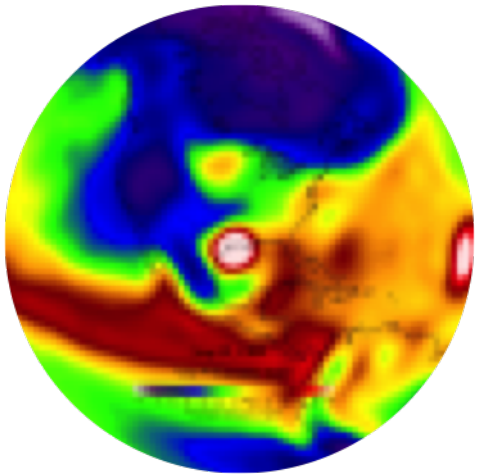
DOE Facilities are Facing a Data Deluge



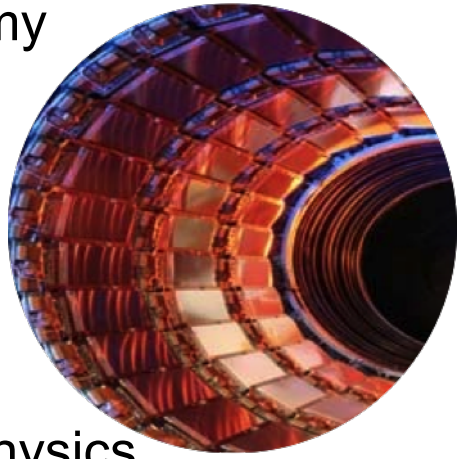
Astronomy



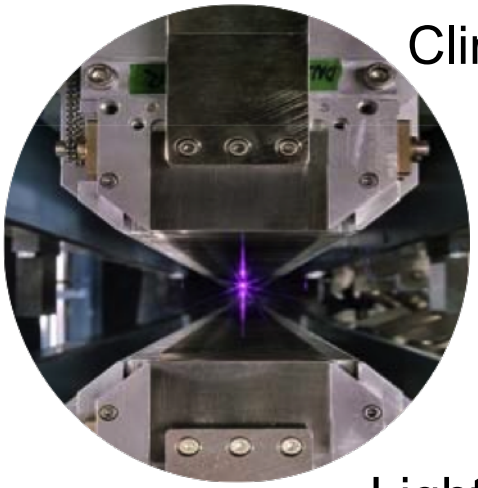
Genomics



Climate



Particle Physics



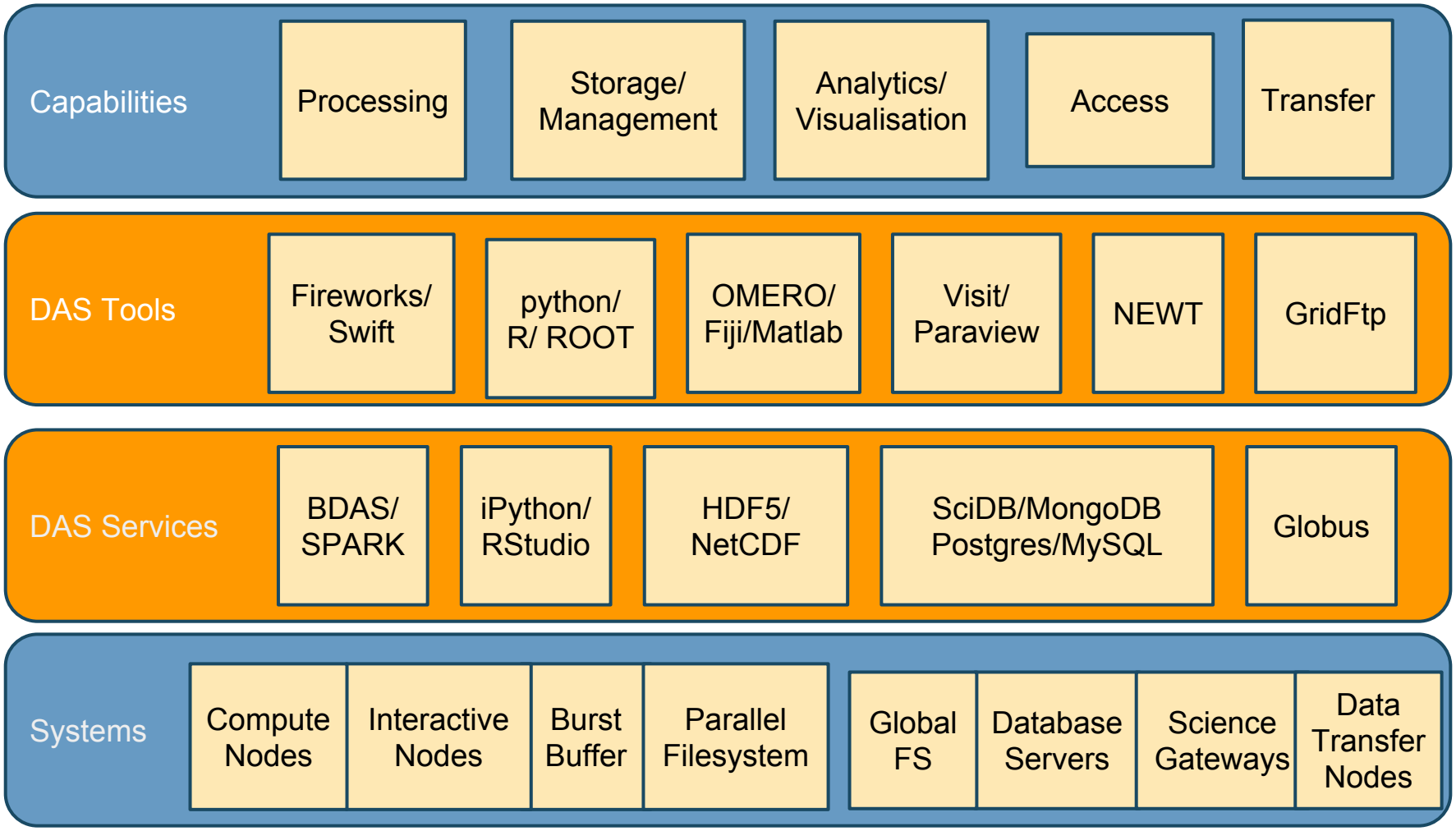
Light Sources

NERSC Data and Analytics Services Group

DAS Team Goal: "Enable Data-Centric Science at Scale"

- **Big Data Software**
  - Broad ecosystem of capabilities and technologies
  - Research and evaluate
  - Customize and optimize for NERSC/HPC platforms
  - Deploy and maintain
- **Engaging NERSC Users**
  - Broad user base support
  - 1-1 in-depth engagement

NERSC Data and Analytics Stack



DAS Team Member		Technology Areas
Debbie Bard		Workflows, Cori Phase1
Wahid Bhimji		Databases, Cori Phase1
Shreyas Cholia		Gateways, Web, Grid, Databases
Joaquin Correa		Imaging, Analytics (Matlab/Python),
Lisa Gerhardt		SciDB, Spark
Annette Greiner		Gateways,UI, Web,
Burlen Loring		Vis
Jeff Porter		Data Transfer, Grid
Prabhat		Group Leader, I/O Libraries
Oliver Ruebel		Vis, Analytics (Python)
Michael Urashka		Web
Dani Ushizima		Analytics (*R*)

NERSC Systems

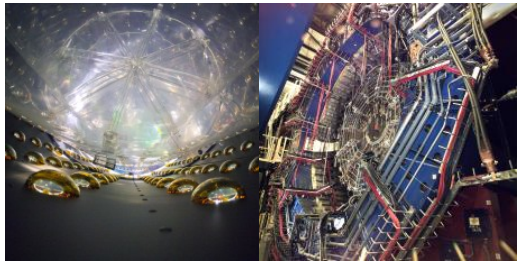
Compute Intensive

Data Intensive

Carver

Retiring end-Sept

Genepool



PDSF

The Cori System

- **Cori will transition HPC and data-centric workloads to energy efficient architectures**

Image source: Wikipedia

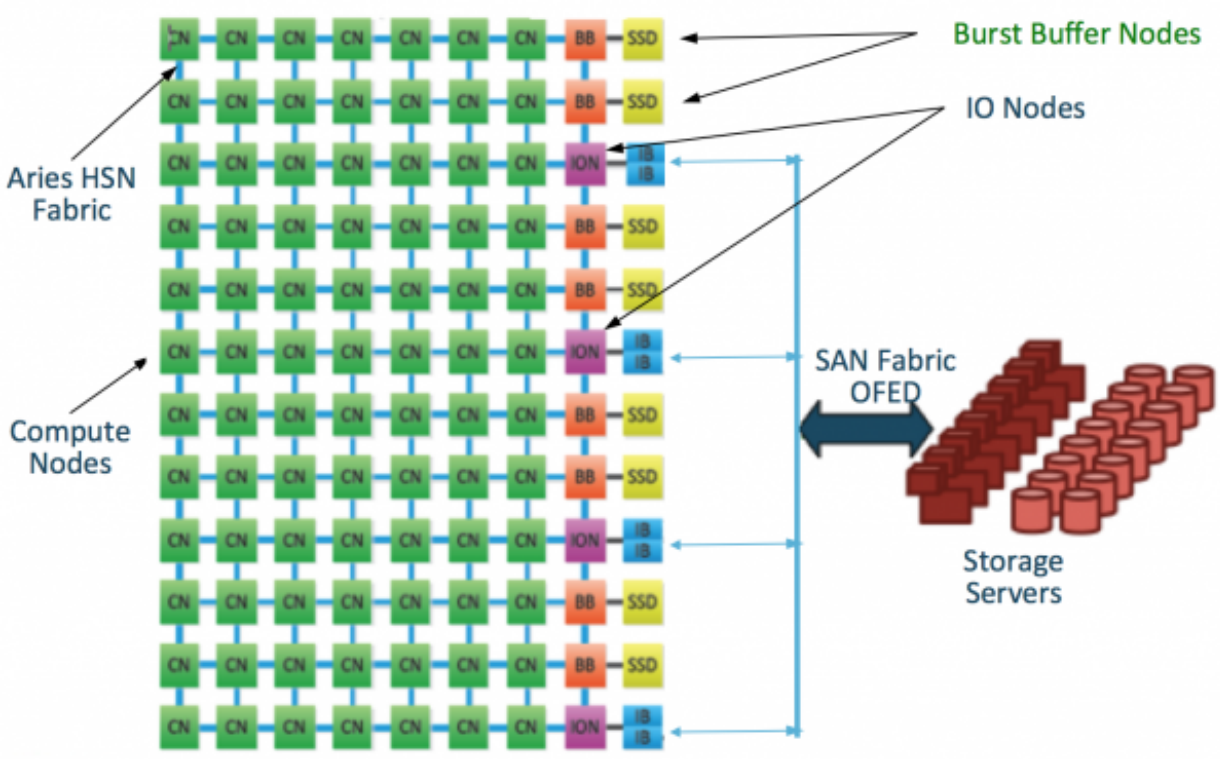
System named after Gerty Cori, Biochemist and first American woman to receive the Nobel prize in science.

Cori Configuration

- **Over 9,300 'Knights Landing' compute nodes**
  - 72 cores per node
  - High bandwidth on-package memory
  - Coming mid-2016
- **Phase 1 (coming fall 2015): ~1,400 'Haswell' compute nodes will be 'data partition'**
- **Aries Interconnect**
- **Lustre File system (arriving 2015)**
  - 28 PB capacity, >700 GB/sec peak performance
- **NVRAM "Burst Buffer" for I/O acceleration**
  - ~1.5PB capacity, ~1TB/s (half with Phase 1)

Burst Buffer

https://www.nersc.gov/users/computational-systems/cori/burst-buffer/



- **IO improvements: high bandwidth** reads and writes, e.g. checkpoint/restart; **high IOP/s** (input-output operations per second), e.g. non-sequential table lookup; **out-of-core** applications
- **Workflow performance improvements: coupling applications,** using the BB as interim storage; Optimizing node usage by changing node concurrency part way through a workflow (using a persistent BB reservation)
- **Analysis and Visualization: In-situ / in-transit; Interactive** (using a persistent BB reservation)

- Integrated with SLURM batch system
- BB Early User Program (deadline Aug 23rd) https://www.nersc.gov/users/accounts/allocations/burst-buffer-early-user-program/

Other Data Features on Cori P1

- High memory per node (128 GB/node) and some nodes with more
- Many login nodes for advanced workflows and analytics
- Outbound internet connection for e.g. external DBs
- Cross mounting of Cori and Edison filesystems
- Queues (in addition to debug, regular, premium and low priority):
  - 'Realtime' and 'Killable': real-time data ingestion/analysis
  - 'Throughput' and 'Serial' for long-running and data-intensive workloads
- Docker-like containers (shifter): custom software (and performance)

NERSC Users: How to get help?

- **Documentation:**
  - https://www.nersc.gov/users/data-analytics/
- **Cori:**
  - https://www.nersc.gov/users/computational-systems/cori/

- **Routine startup/troubleshooting questions:**
  - Routed to DAS via trouble ticket system: consult@nersc.gov

- **In-depth 1-1 collaborations:**
  - e-mail wbhimji@lbl.gov or prabhat@lbl.gov

Afternoon Agenda

1:00 p.m.	Data and Analytics Services Overview	Wahid Bhimji
1:30 p.m.	Moving and Sharing Data	Shreyas Cholia / Jeff Porter
2:00 p.m.	Databases and I/O	Wahid Bhimji
2:30 p.m.	Data Analytics	Joaquin Correa
3:00 p.m.	Break	
3:15 p.m.	Visualization	Burlen Loring
4:00 p.m.	Workflows	Debbie Bard
4:45 p.m.	Next Steps and Wrap-up (and optional tour of machine room)	Wahid Bhimji

Questions ?