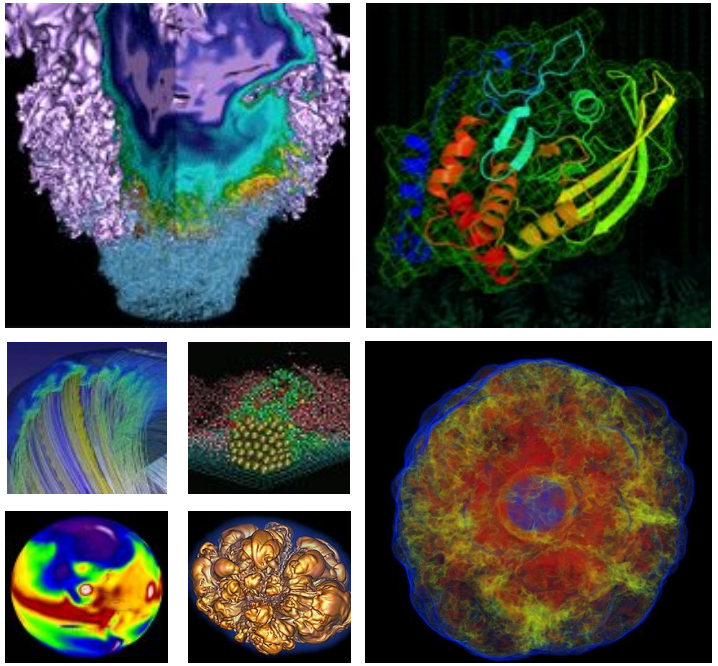
NERSC Data Ecosystem Overview



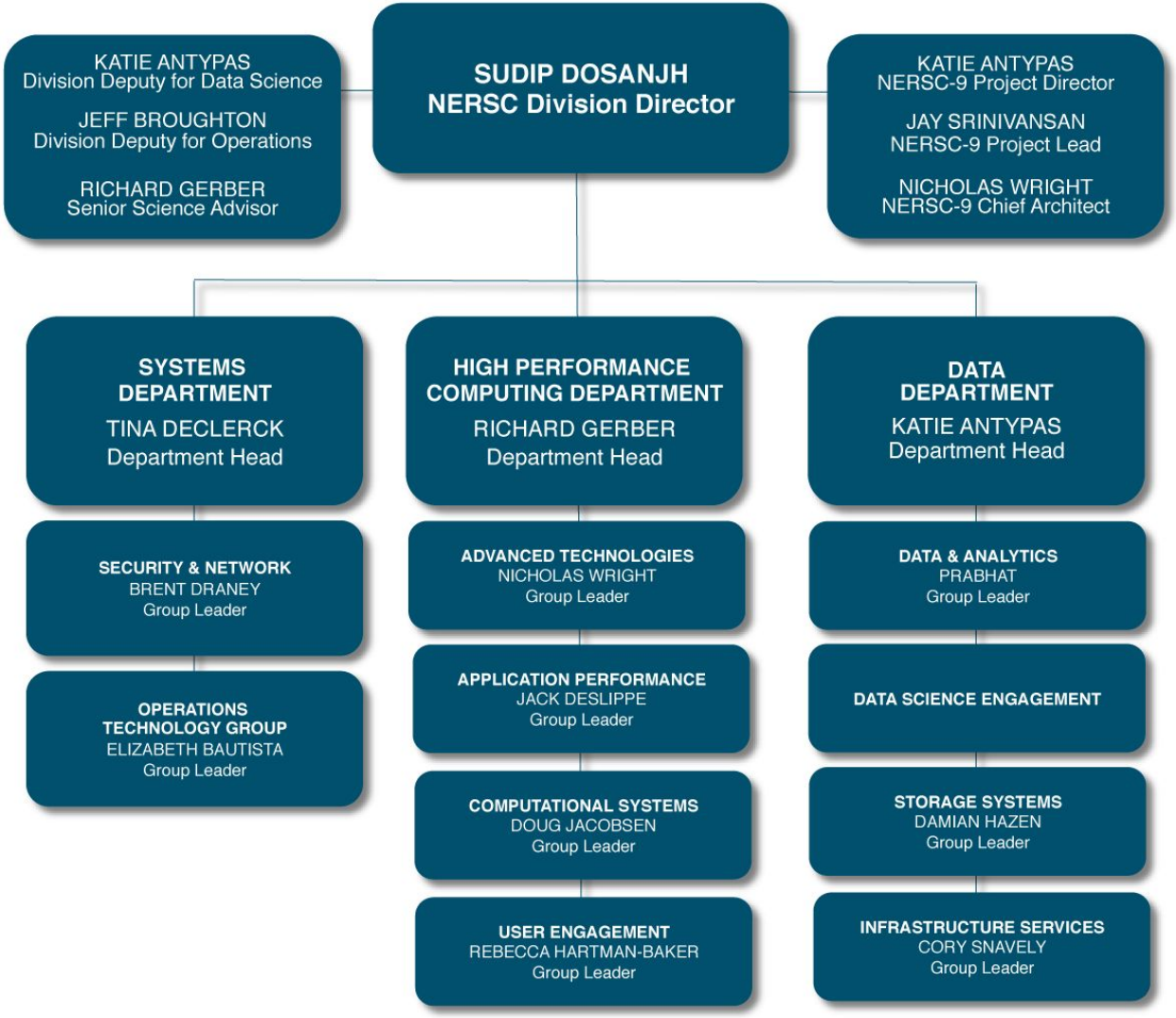
Prabhat

Data & Analytics Services Group

January 25, 2019

Afternoon Schedule

1:40 pm	Data Transfer	Shreyas Cholia
2:10 pm	File Systems + Burst Buffer	Wahid Bhimji
2:30 pm	I/O Best Practices	Quincey Koziol
2:50 pm	Break	
3:10 pm	Python and Jupyter	Rollin Thomas
3:30 pm	Machine Learning	Mustafa Mustafa
3:50 pm	Shifter	Shane Canon
4:10 pm	End	


KATIE ANTYPAS
Division Deputy for Data Science
JEFF BROUGHTON
Division Deputy for Operations
RICHARD GERBER
Senior Science Advisor
SUDIP DOSANJH
NERSC Division Director
KATIE ANTYPAS
NERSC-9 Project Director
JAY SRINIVANSAN
NERSC-9 Project Lead
NICHOLAS WRIGHT
NERSC-9 Chief Architect
SYSTEMS DEPARTMENT
TINA DECLERCK
Department Head
HIGH PERFORMANCE COMPUTING DEPARTMENT
RICHARD GERBER
Department Head
DATA DEPARTMENT
KATIE ANTYPAS
Department Head
SECURITY & NETWORK
BRENT DRANEY
Group Leader
OPERATIONS TECHNOLOGY GROUP
ELIZABETH BAUTISTA
Group Leader
ADVANCED TECHNOLOGIES
NICHOLAS WRIGHT
Group Leader
APPLICATION PERFORMANCE
JACK DESLIPPE
Group Leader
COMPUTATIONAL SYSTEMS
DOUG JACOBSEN
Group Leader
USER ENGAGEMENT
REBECCA HARTMAN-BAKER
Group Leader
DATA & ANALYTICS
PRABHAT
Group Leader
DATA SCIENCE ENGAGEMENT
STORAGE SYSTEMS
DAMIAN HAZEN
Group Leader
INFRASTRUCTURE SERVICES
CORY SNAVELY
Group Leader

Cori Brings HPC & Data Together

Gerty Cori: Biochemist and first American woman to win a Nobel Prize in science

Phase I: 2388 x 32-core Intel Xeon "Haswell" 128 GB DDR4
Phase II: 9688 x 68-core Intel Xeon Phi "KNL" 96 GB DDR4 + 16 GB MCDRAM

Production Data Stack

Capabilities	Technologies
Data Transfer + Access	globus online, GridFTP, jupyter, !m, python django, newt
Workflows	FireWorks, TaskFarmer
Data Management	HDF, netCDF, ROOT Data Analysis Framework, mongoDB, MySQL, PostgreSQL
Data Analytics	python, Spark, TensorFlow, K, julia, Caffe, PYTORCH
Data Visualization	VisIt, ParaView

Cori's Data-Friendly Features

12 32-core Haswell 500 GB Login Nodes

Jupyter Notebook Node

768 GB "bigmem" Haswell Nodes

Pipeline/Workflow Management Nodes

Serial Queue
Shared-node Queue
Transfer Queue

Containerized Environments

Real-Time Queues for Co-Scheduling w/Experiments

Streaming Data to Compute Nodes

Interactive Queue: 64 Nodes x 4 Hours

External Network Access to/from Compute Nodes

Cray DataWarp: Burst Buffer for I/O acceleration

Asks

- **Please engage with NERSC staff members**
  - Provide feedback, critique
  - Tell us about interesting science problems!

Questions? Comments?