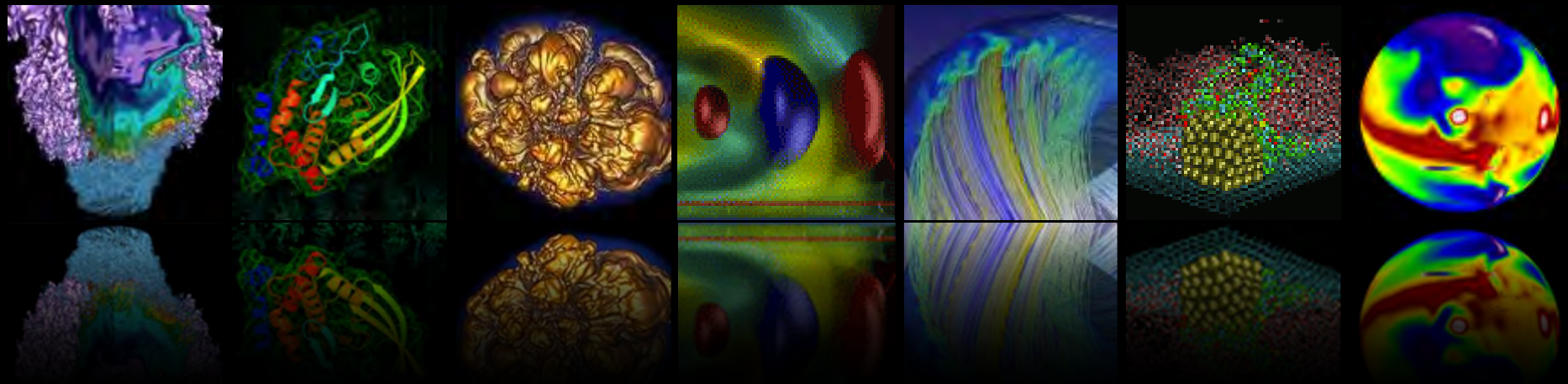


SPOT Suite: Scalable, autonomous data management and scientific workflows for X-Ray Light Sources

Craig E. Tull, Physics and X-Ray Science Lead

Computational Research Division

Lawrence Berkeley National Laboratory

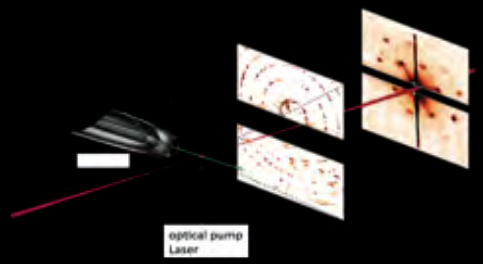
BERKELEY LAB

U.S. DEPARTMENT OF ENERGY

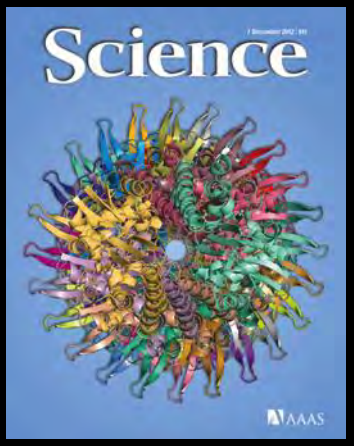
BES Facilities serve 16,000 users/yr in Materials, Biology, Energy, Medicine, …

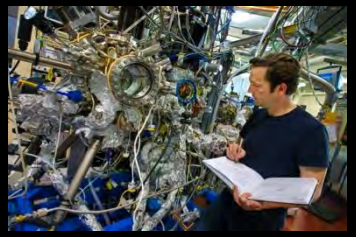
- **Virtually every area of science and techn are taking advantage of Lightsources, etc.**
- **The ALS user base is expanding to new a and includes more 1st timers who cannot affo investment in learning hardware & softwa**
- **Data volumes are exploding:**
  - Lightsources are getting brighter
  - Detectors are getting faster
  - Beamlines are automating
- **New mathematical techniques, new architectures, and even new paradigms (eg. Neuromophic, Quantum) are being developed or researched.**

Experimental Science Workflows of the past



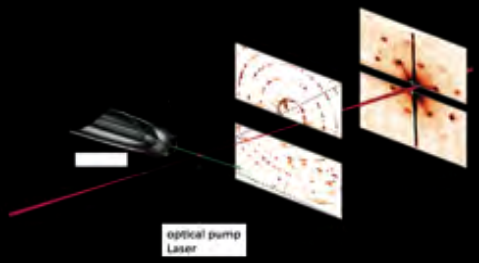
Experiment





Beamline User

Modern Experimental Science Workflows



Experiment

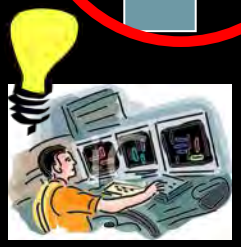
Data Pipeline

HPC Storage and Compute

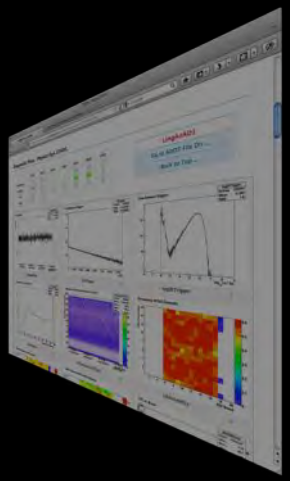
Networks, workflows, and automation

New experiment

Prompt Analysis Pipeline

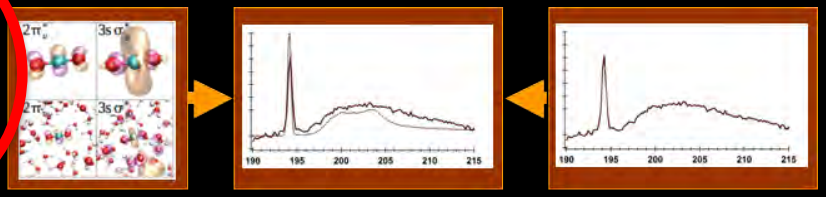


Beamline User



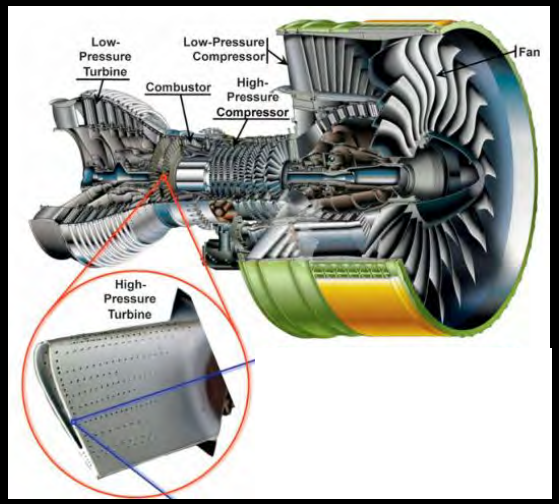
Science Gateway

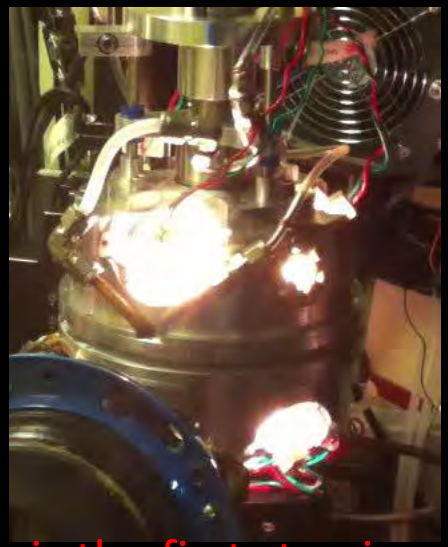
simulate → compare ← measure

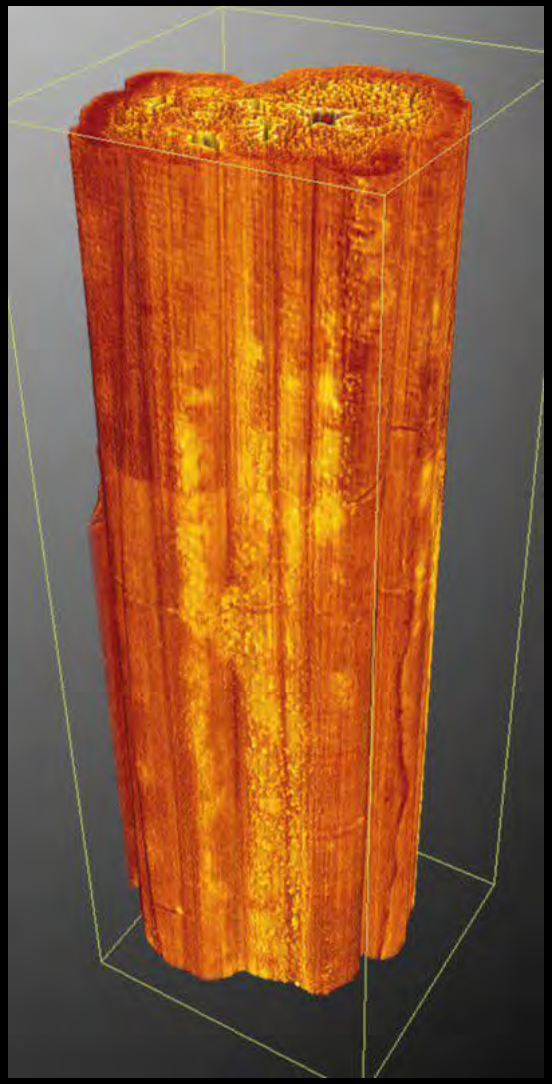


Simulation and Analysis Framework

Automated Processing: real-time QA & monitoring of in-situ, time-resolved experiments.



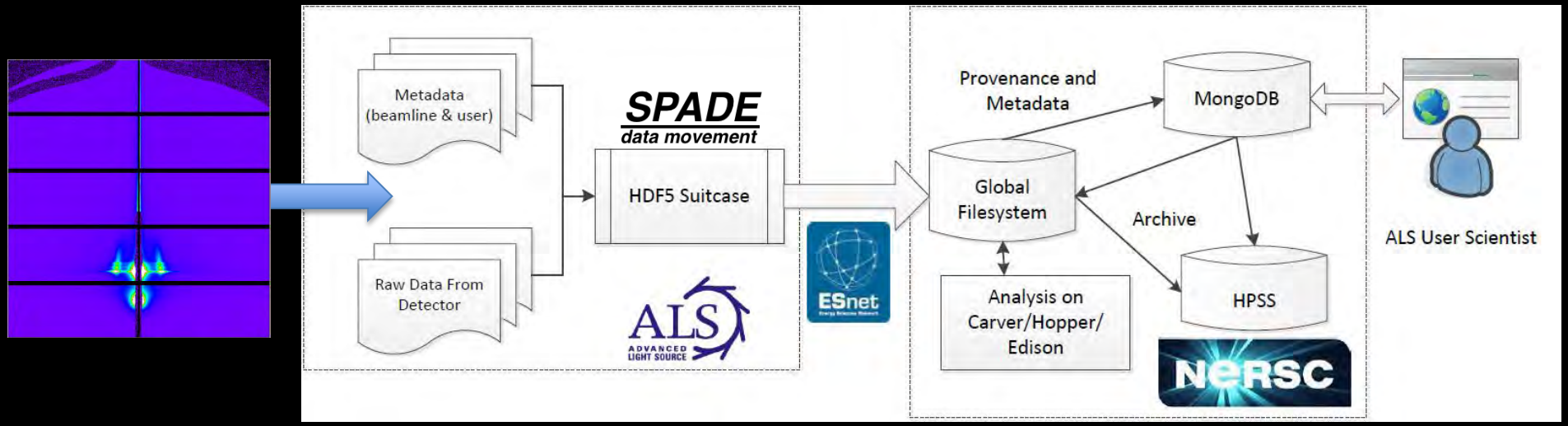




"Not processing your data in real time is the first step in not processing your data at all" – Peter Zwart, LBNL

HA Bale, RO Ritchie, et al, Nature Materials, 2013, 12(1) 40-46

SPOT Suite: Integration of ALS, ESnet, and NERSC into a proto-super-facility.

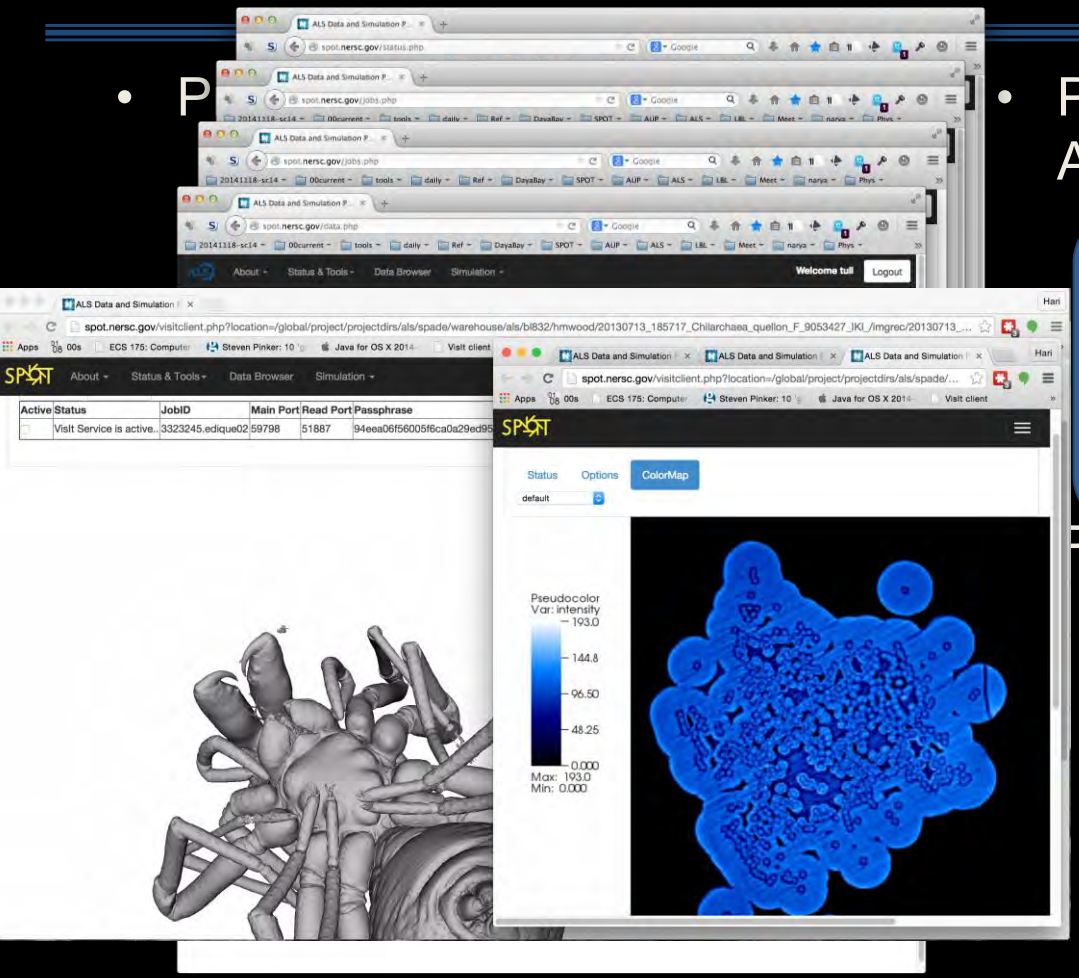


- **Computing Research Division** – Large Data & Workflows
- **Advanced Light Source** – Science & Detectors
- **Material Science Division / Molecular Foundry** – Theory
- **ESnet** – National Networking
- **NERSC** – Large-scale Scientific Computing Facility

CETull@lbl.gov - 14 April 2016

SPOT Suite: From research project to production capability.

- P

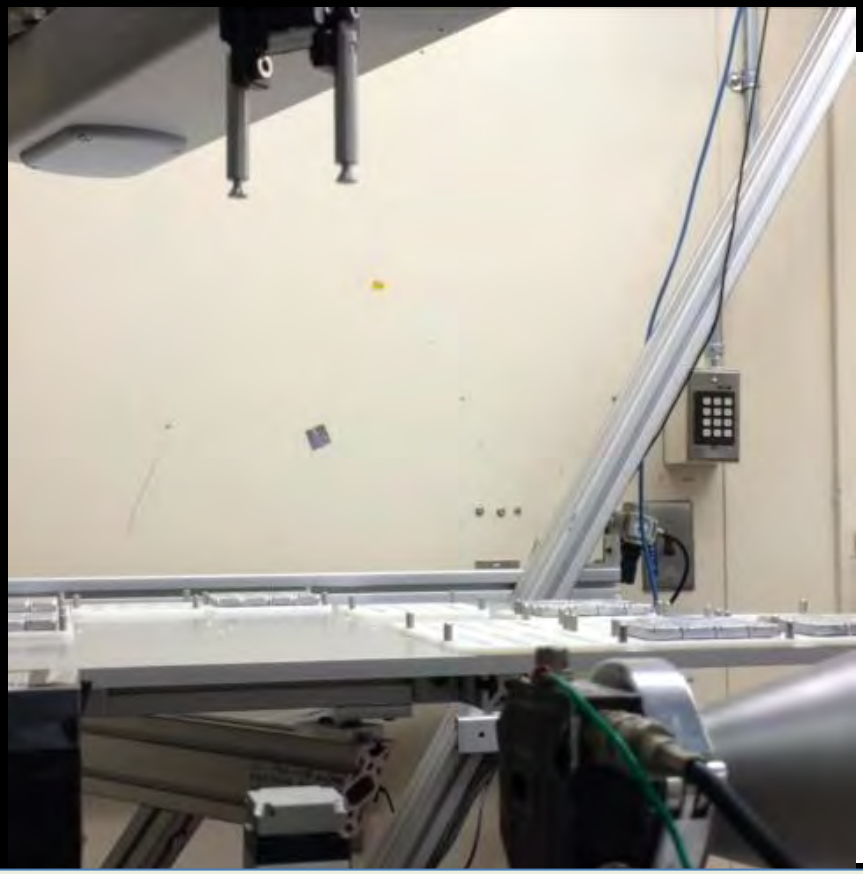


- Pseudo-production running at ALS beamlines:

➢ 24/6 Operation	(Jun2017)
➢ >365,000 Datasets	
➢ ~250 Beamline Users	
➢ >3 PB Data Stored	
➢ >5 million Jobs at NERSC	

- Portal interfaces for:
  - Data monitoring & control
  - Job monitoring
  - Job inspection
  - Data browsing and searching
  - Browser-based remote viz
  - … metadata, provenance, algorithms, theory-simulation, etc.

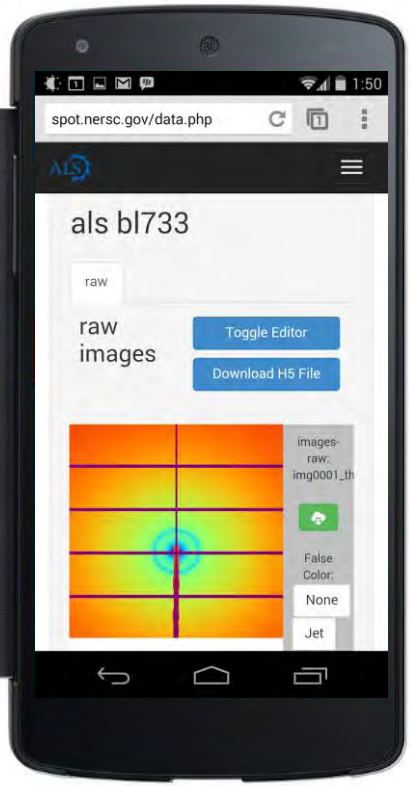
Remote experiments now a reality.



25mar2014:

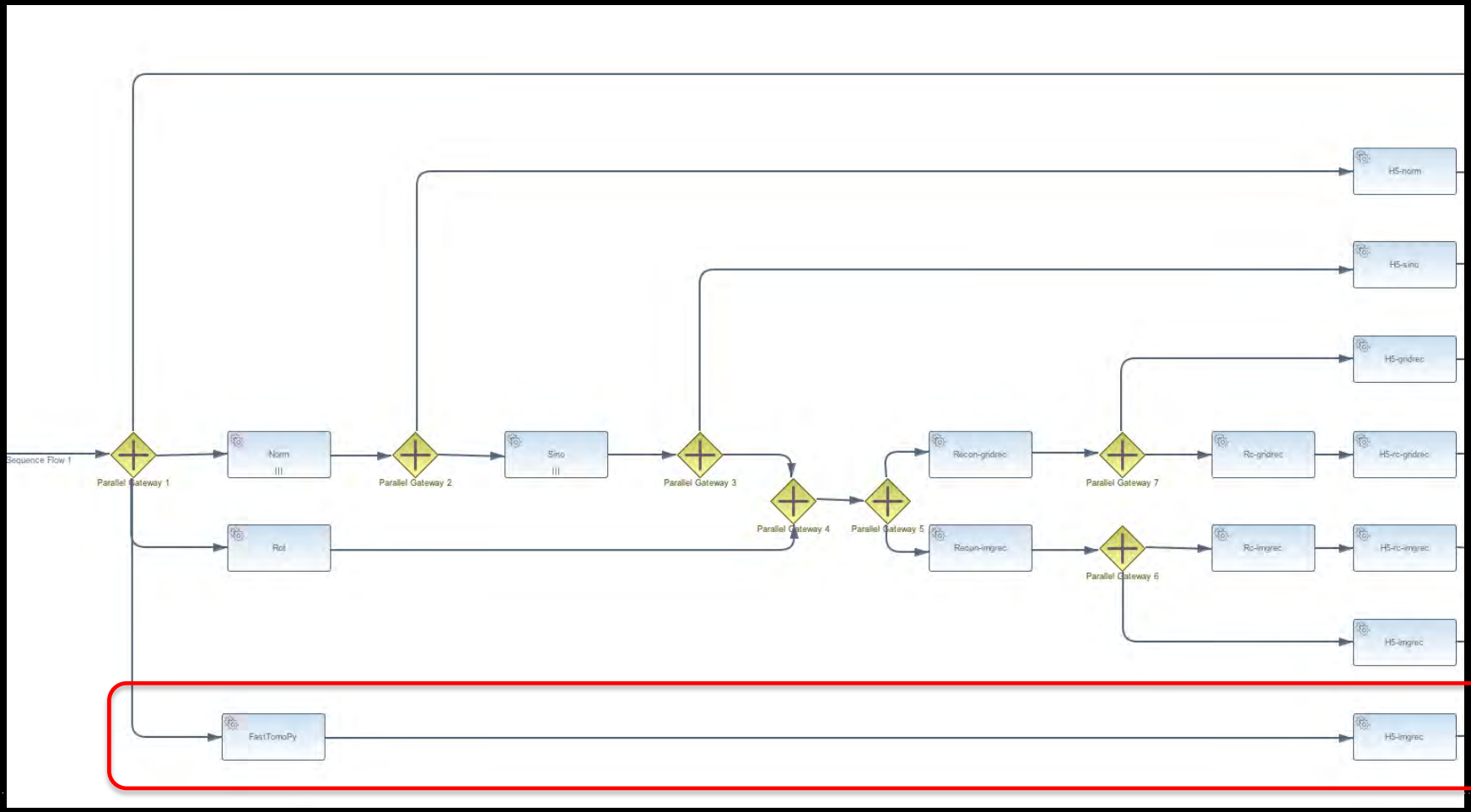
UK scientists conduct remote experiment using new BL 7.3.3 robot and SPOT.

Able to assess experimental data on train to Zurich via mobile interface.



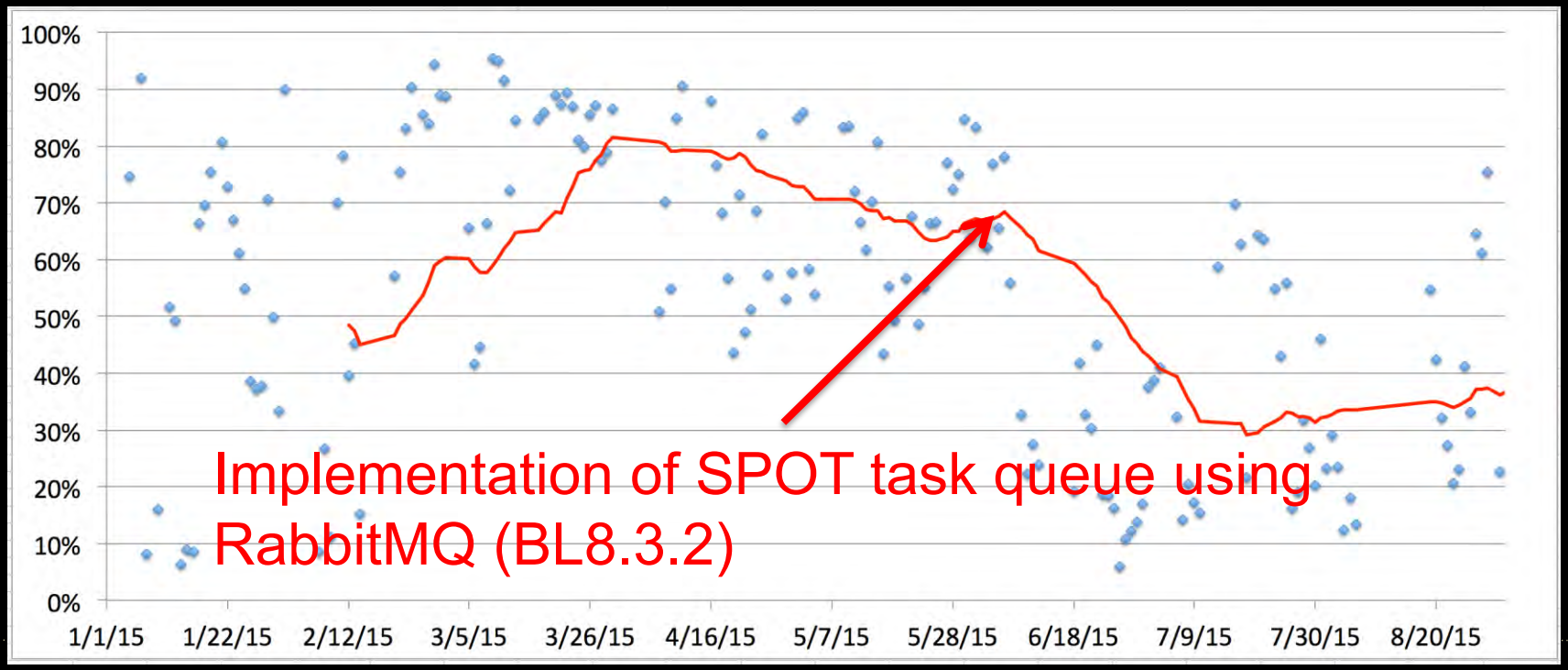
From: Alessandro Sepe as2237@cam.ac.uk -- **Actually, I did not feel any difference between a standard beamtime and this NERSC remotely accessed beamtime, which is quite an extraordinary result.**

In a complex workflow, not all paths are of equal value for streaming feedback.



X-SWAP: Extreme-Scale Scientific Workflow Analysis and Prediction - Instrumented NERSC workflow provides lever for optimizing throughput.

- ALS beamline 8.3.2 (Tomography) queue wait time dropped from 60-70% to 30% of total turn-around time for jobs.
- We can optimize by deploying more master/worker pods.



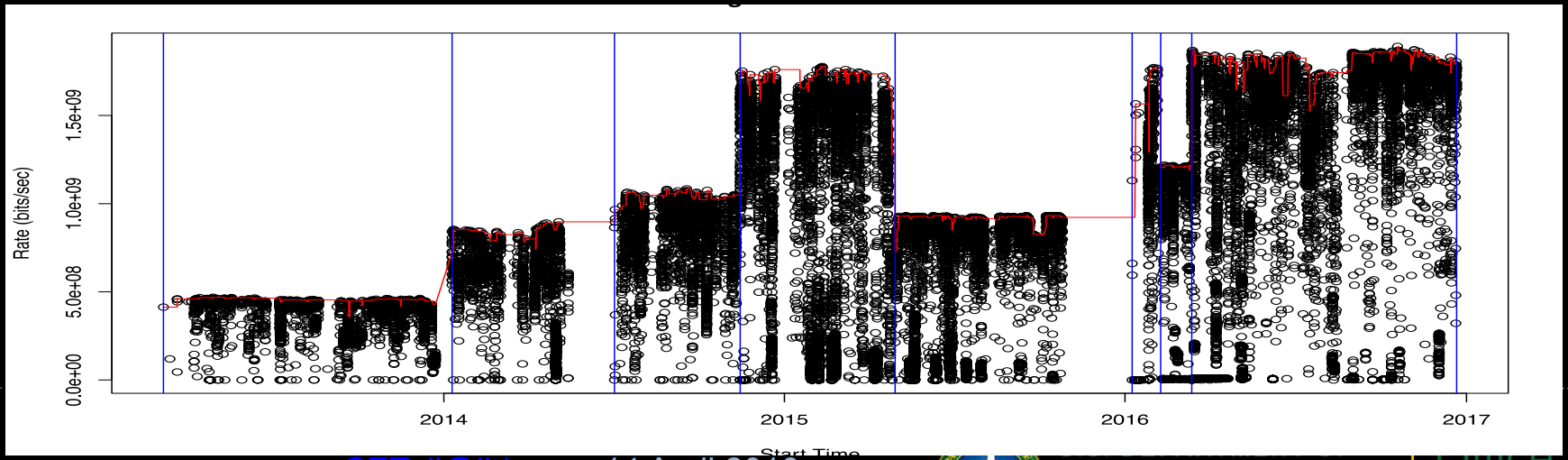
X-SWAP: Instrumenting and modeling to minimize workflow branch latency.

- SPOT Tomographic processing is a DAG of 54 graph nodes.
- Fast feedback on a small subset of data is sufficient for QA.
- Introduce a new DAG branch (Fast TomoPy)
- First feedback reduced from ~16 minutes to ~2
- Trade-off quality & completeness.

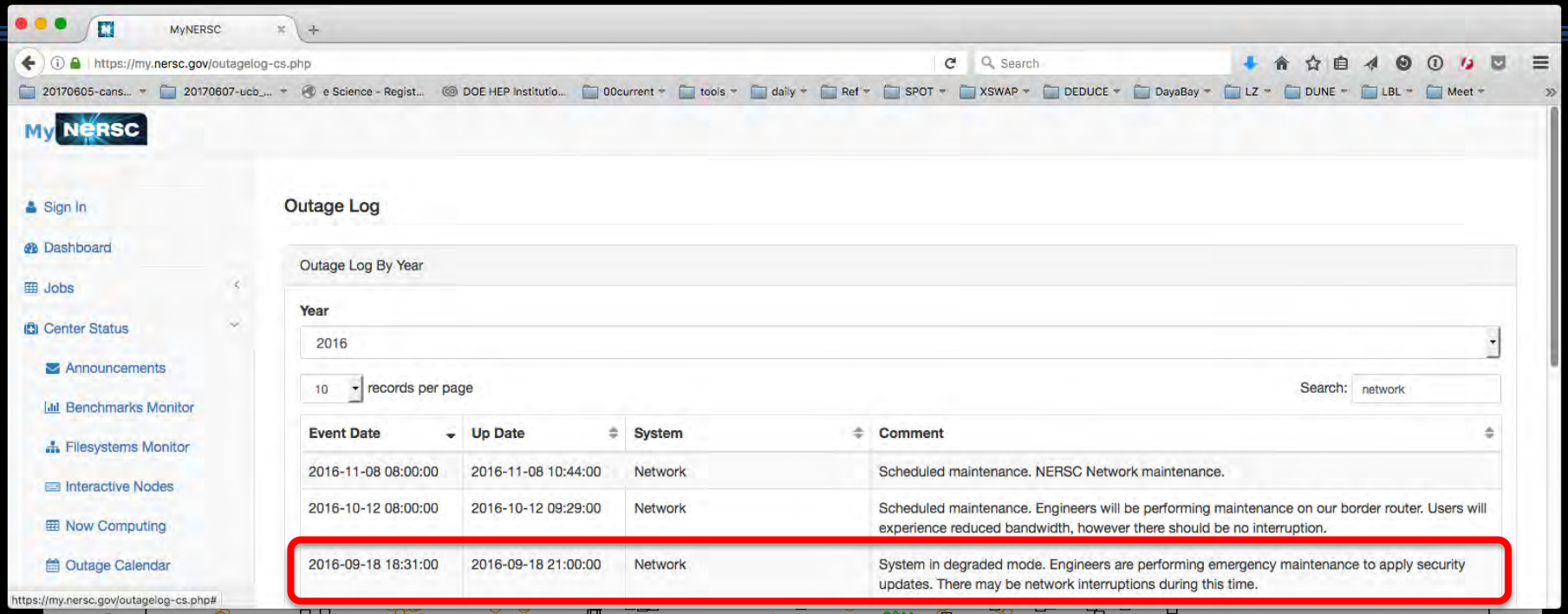
CETull@lbl.gov - 14 April 2016

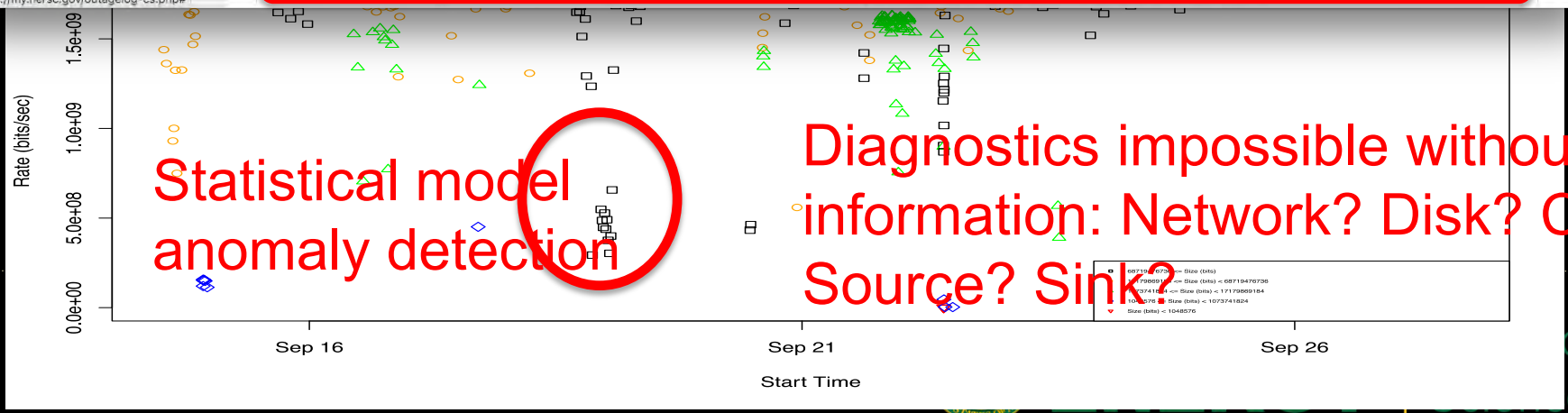
ALS: X-SWAP Modeling and real-time change-point detection

- Distributed workflows in the real world prone to many changes. Need monitoring that identifies changes and anomalies.
  - Input to Alerts and Alarms and/or dynamic shifting of resources.
- Applied statistical, ”science-neutral” modeling.
  - Matched distribution shapes and magnitudes
  - Data File size and time-of-day dependencies
  - Detects baseline changes & anomalies in realtime with sliding time-window model.



ALS: X-SWAP missing data





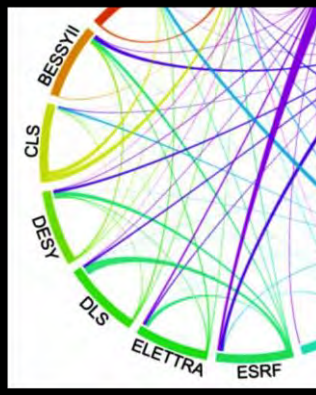
Statistical model anomaly detection

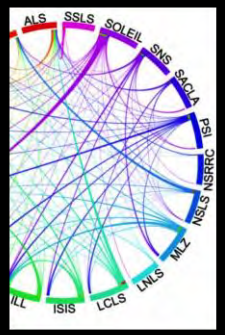
Diagnostics impossible without fa information: Network? Disk? CPU Source? Sink?

DEDUCE: Distributed Dynamic Data Analytics Infrastructure for Collaborative Environments – ALS Data Fusion/Federation & Change Detection

- **Data Fusion**: Spade-driven suitcasing of data & metadata is currently hand-crafted for each beamline/facility.
- DEDUCE Decision Triggers must analyze ancillary data as well as main instrument.
- **Data Federation**: ALS users are also users of many other beamlines, light sources, neutron sources, electron microscopes, …
- Can DEDUCE detect scientificly meaningful differences between sample measures?
- **Digital Twins**: Simulate experiment (before or during data taking) – Expensive
- Can DEDUCE predict from data when simulations yield significant scientific insight?

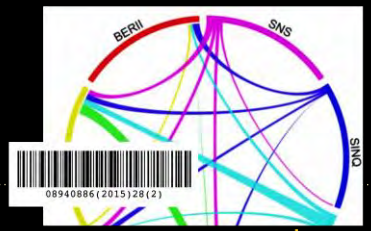
SRN





Information Technology/

Large-Scale Data Handling



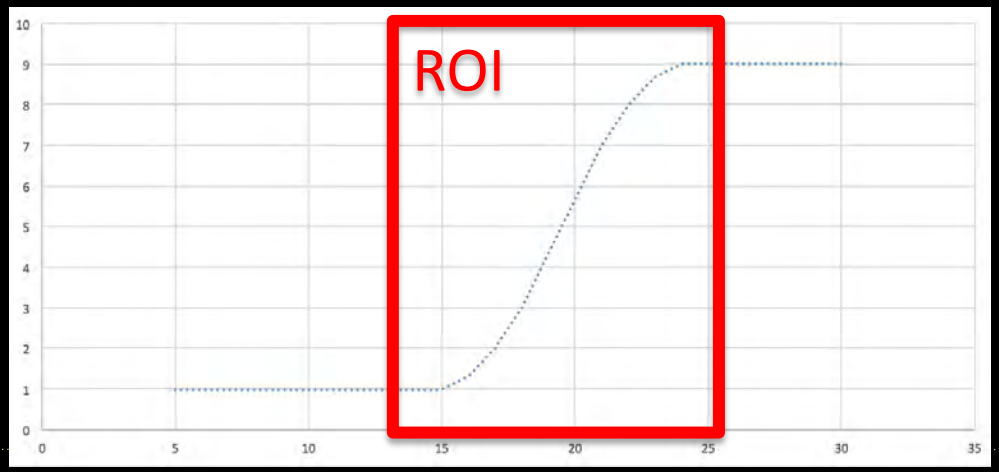
CETull@lbl.gov - 14 April 2016

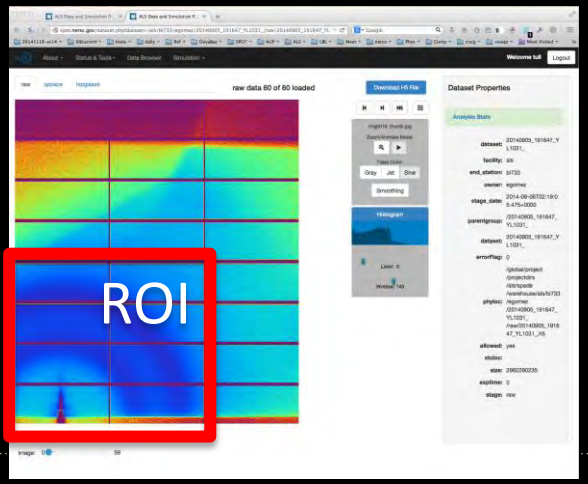
U.S. DEPARTMENT OF ENERGY | Office of Science

DEDUCE: Change-triggered decisions

- Printable Organic Photo-Voltaics: Experiment generate 36,000 images in 3 days => ½ Titan-year to analyze. But, only a fraction of data are a) Region of Interest for science b) fundamentally different from prior experiments c) Successful
  - All data transferred and analyzed – Impossible.
  - Change detection (time and phase-space) cuts data x100
  - Change detection (results wrt prior experiment)
- Transfer*, store*, and analyze only data of scientific interest

  * All data will be transferred and stored until exhaustive validation.



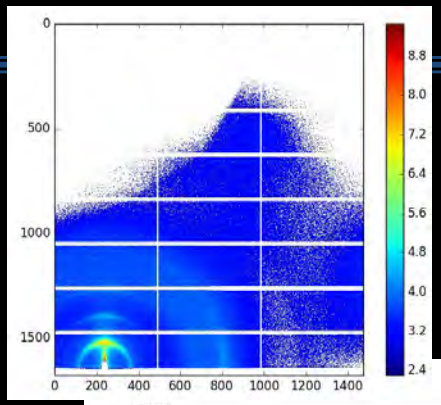


CETull@lbl.gov - 14 April 2016

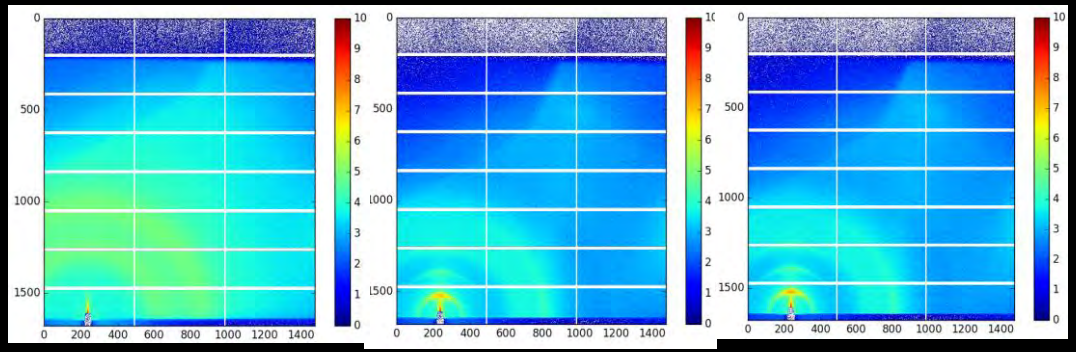
Detecting Change in a Sequence of Light Source Images

Scientific Goal: In a sequence of scattering images, determine when (which frames) and where in the frame significant changes are observed, e.g. beginning of crystallization

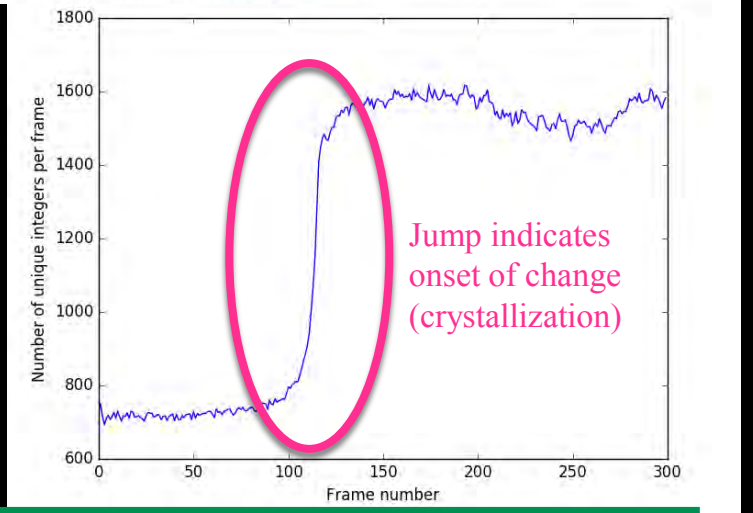
- DEDUCE question: How can we determine *in situ* the temporal and spatial changes that occur in a sequence of images with the goal to reduce the number and size of stored images?
  - Store all images completely, analyze images later
- Data changes and tools that represented the data change well: number of unique integers to detect temporal onset of change; Gaussian smoothing to detect spatial change
- Descriptive statistics (mean, median, etc.) were not useful in this case



Gaussian smoothing eliminates camera artifacts

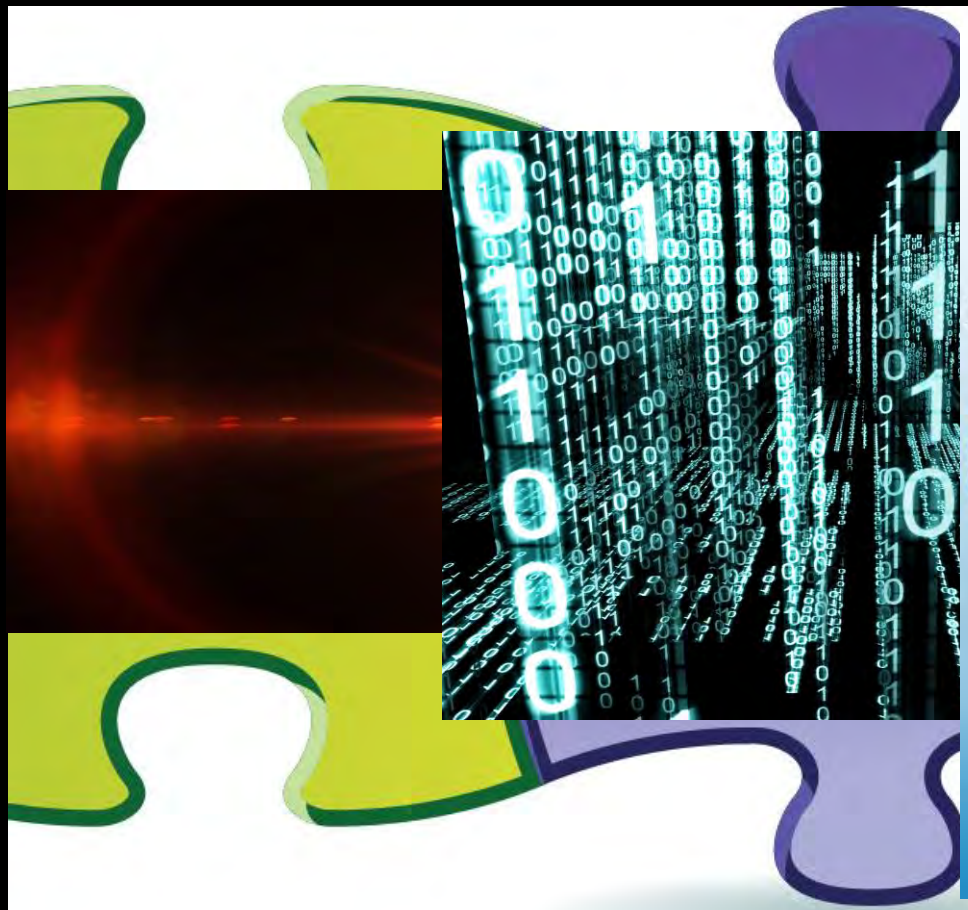


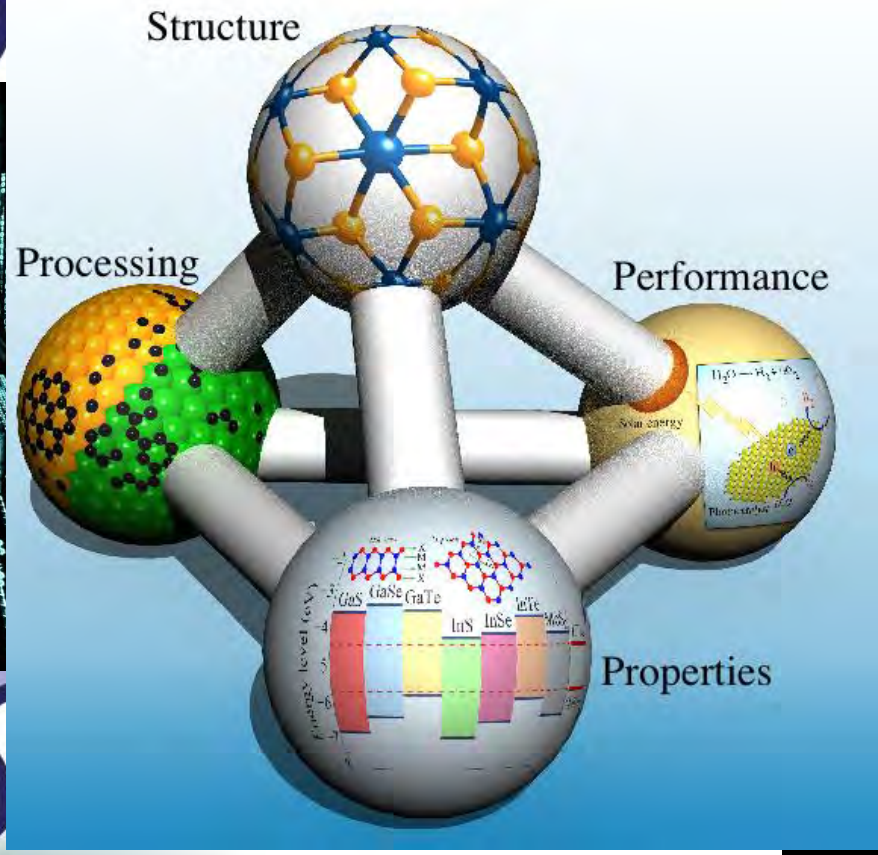
First frame | Onset of crystallization | Last frame



Significance & Impact: We were able to detect significant temporal and spatial changes of image sequences. Potentially, fewer and smaller images have to be stored at NERSC

“SuperFacility” – Seamless integration of Synchrotron & Supercomputer.





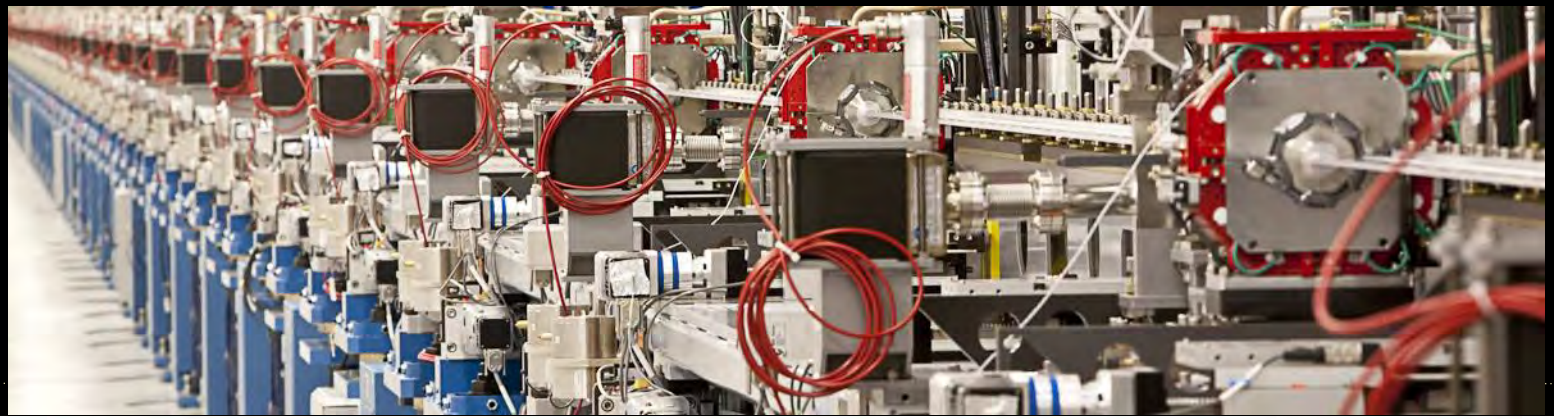
CETull@lbl.gov - 14 April 2016

U.S. DEPARTMENT OF ENERGY | Office of Science

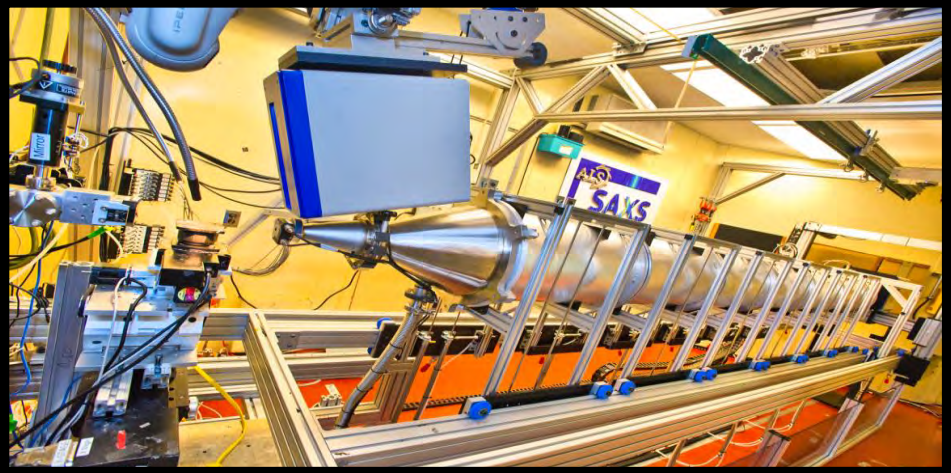
OPV Data Demo: Can this model be extended to illustrate the Super-Facility?

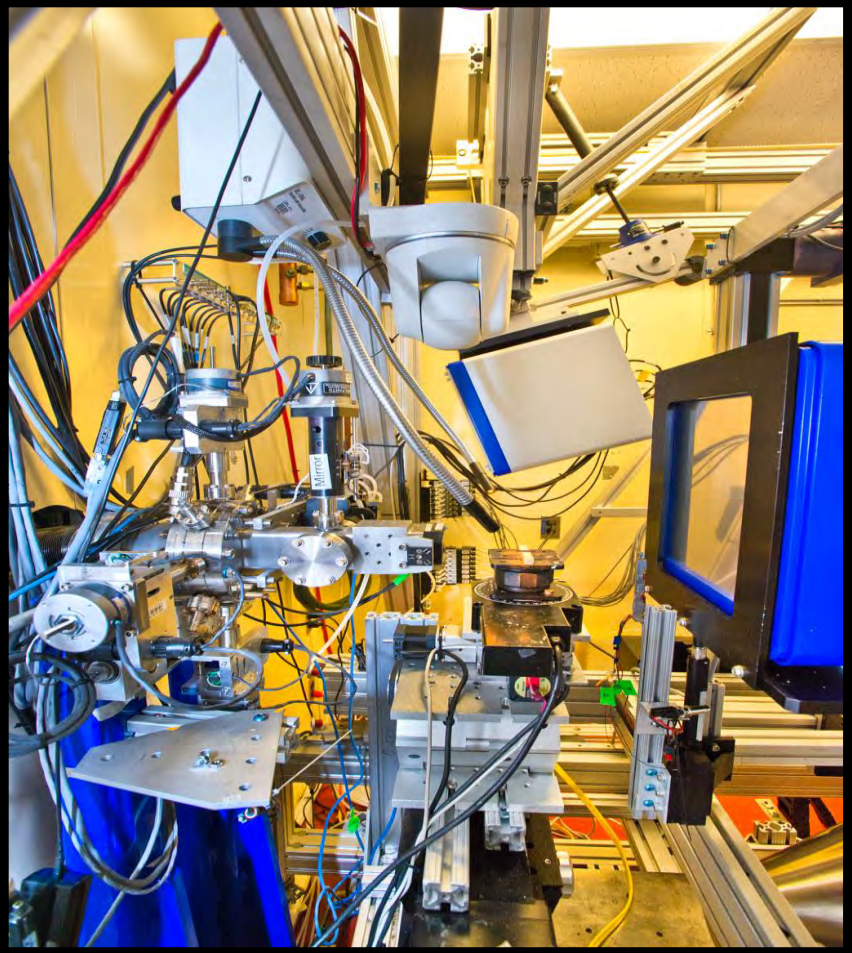
- Printing BCP Organic PhotoVoltaics & Realtime Analysis

- Multi-Facility Imaging and Analysis Pipelines for Large X-ray Data



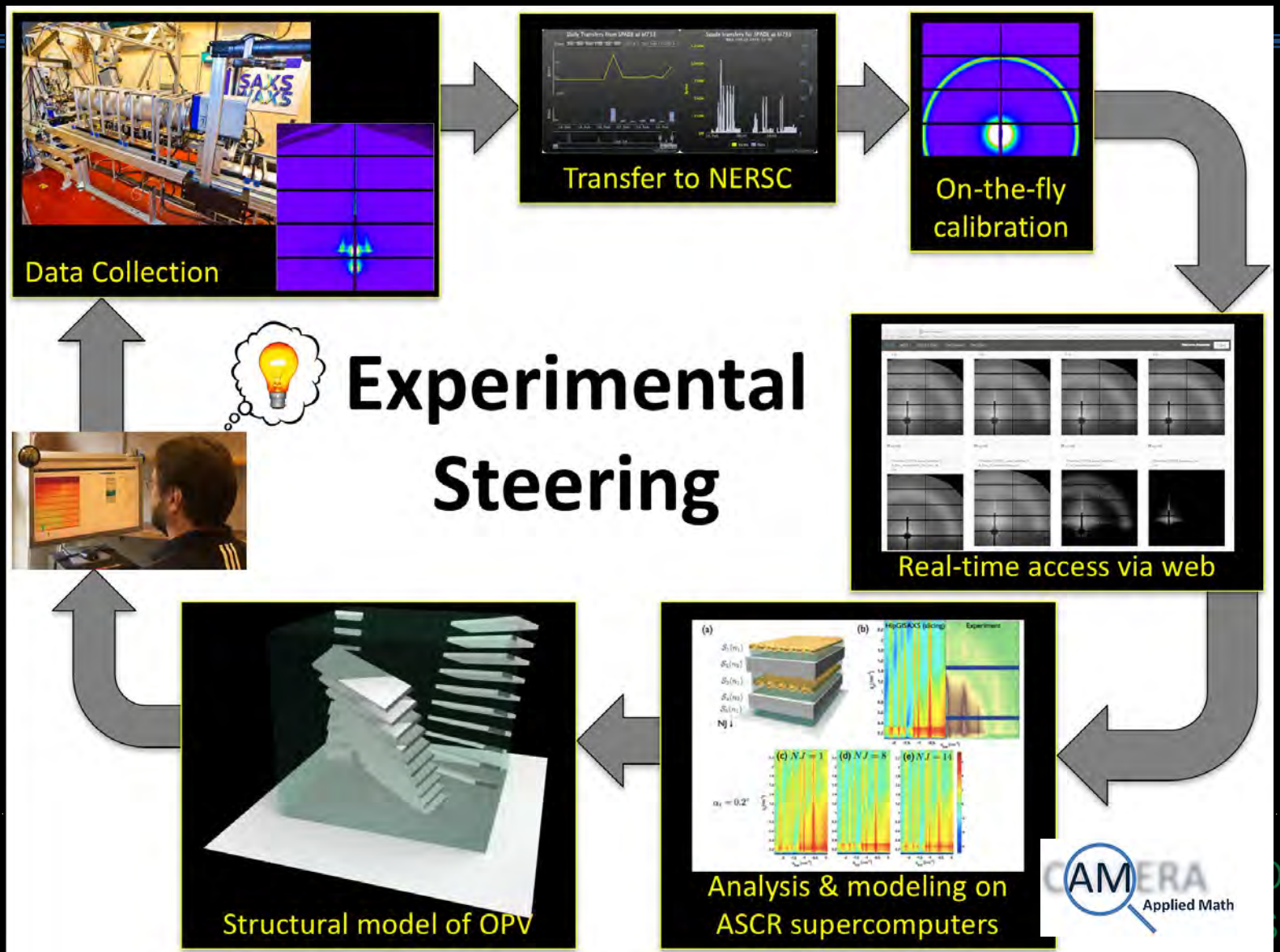
Beamline 7.3.3 SAXS/WAXS/GISAXS





CETull@lbl.gov - 14 April 2016

GISAXS "SuperFacility" Demo Data Flow



CAMERA Applied Math

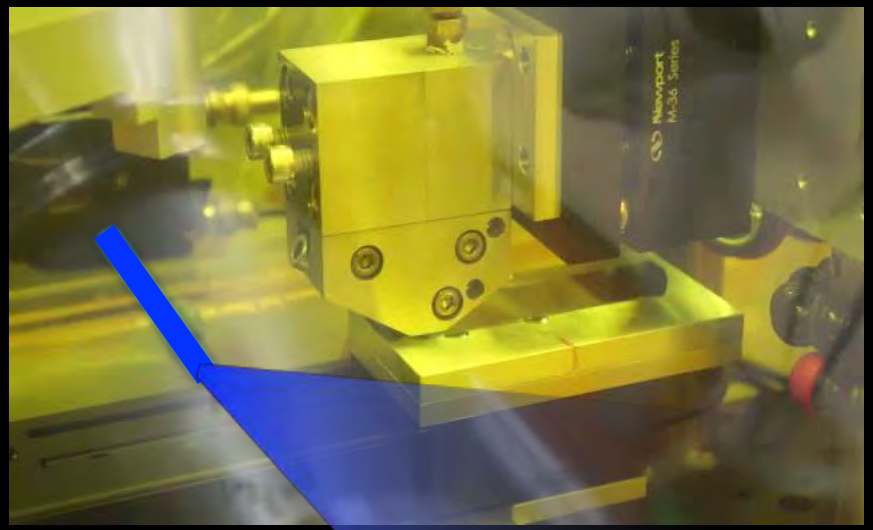
Office of Science

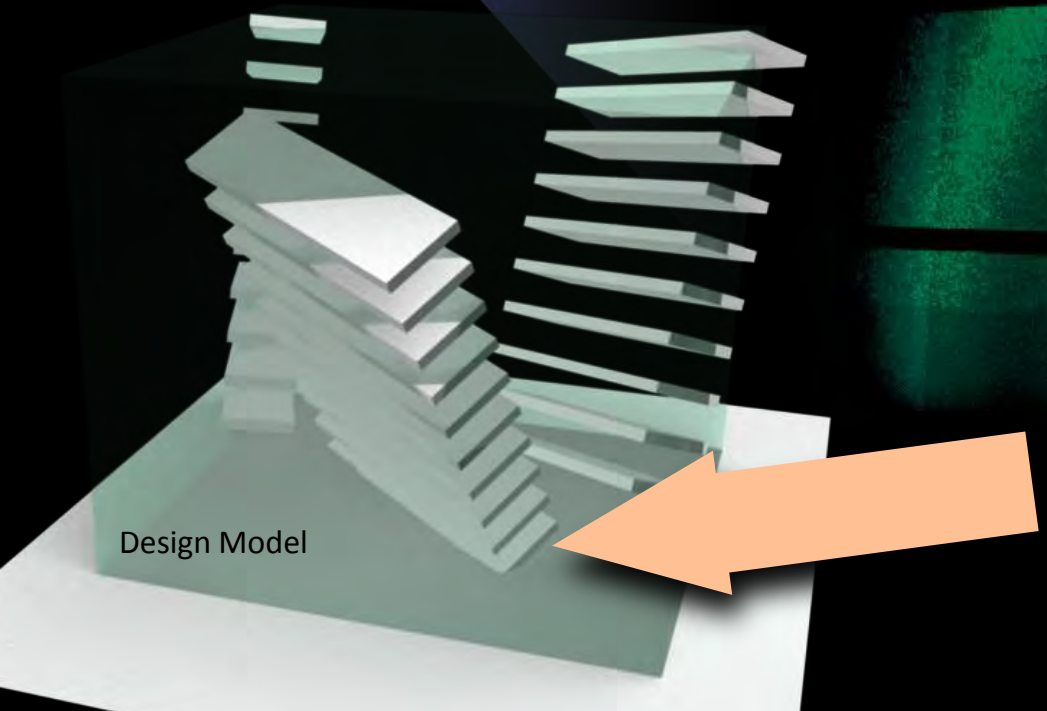
ALS-NERSC-Titan Super-Facility Demo: Real-time analysis for Scattering Data

- Typical ALS chain + dpdak + HipRMC on NERSC reservation
- Co-schedule ALS beamtime & Titan reservation
- Typical ALS chain + dpdak => Spade + Globus Online => ORNL (FW)
- Sentry launches HipGISAXS – 8000 nodes and feeds results back IRT
- HipGISAXS runs **live** during beamtime (see next)
- Integrated Super-Facility Dashboard – **not** production quality

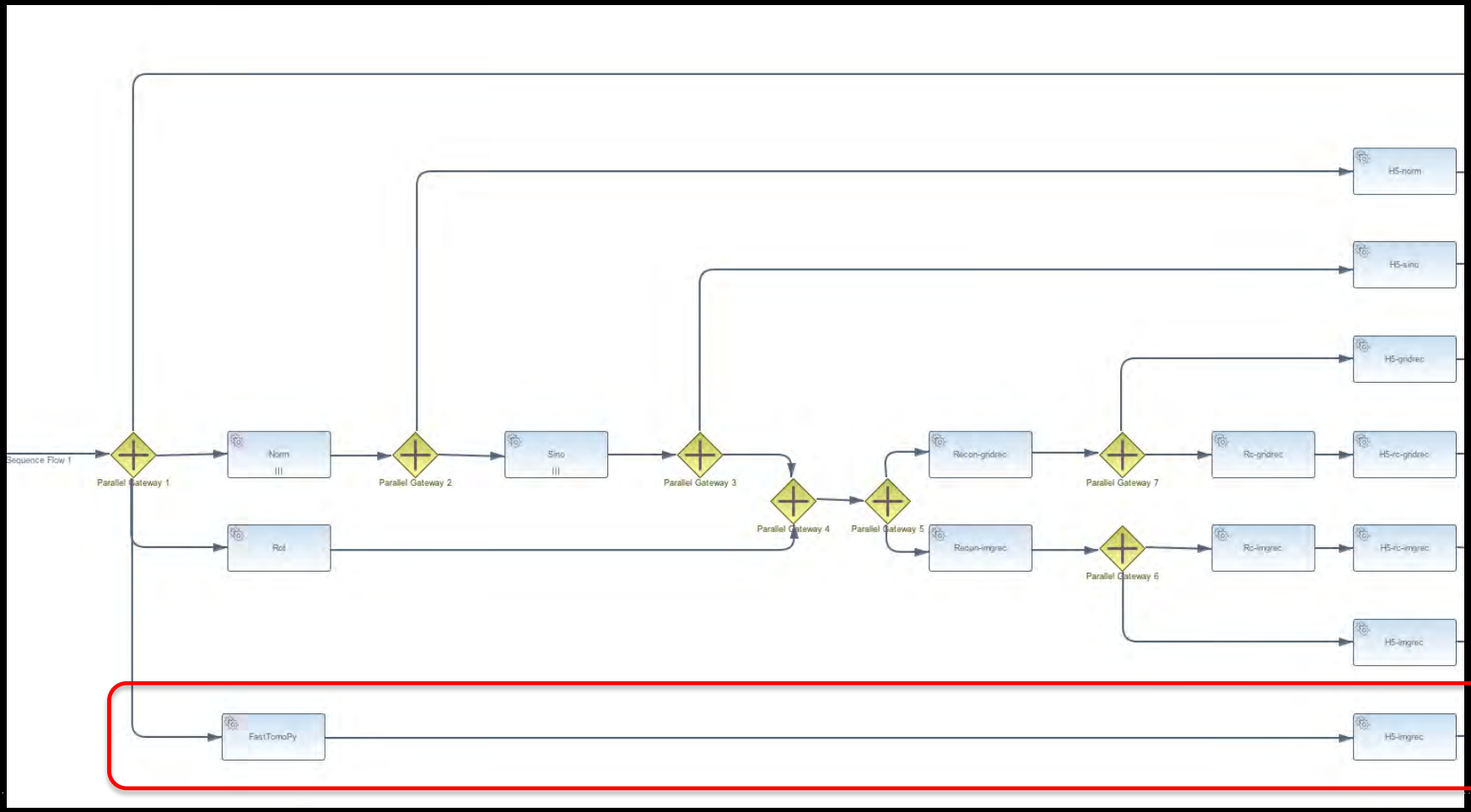
run on some of the fastest computer in the world

critical for QA & in-situ, time-
resolved experiments



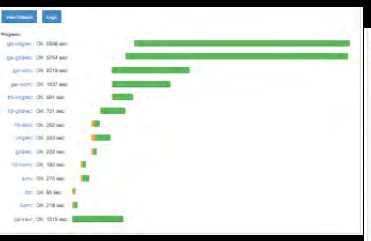


In a complex workflow, not all paths are of equal value for streaming feedback.



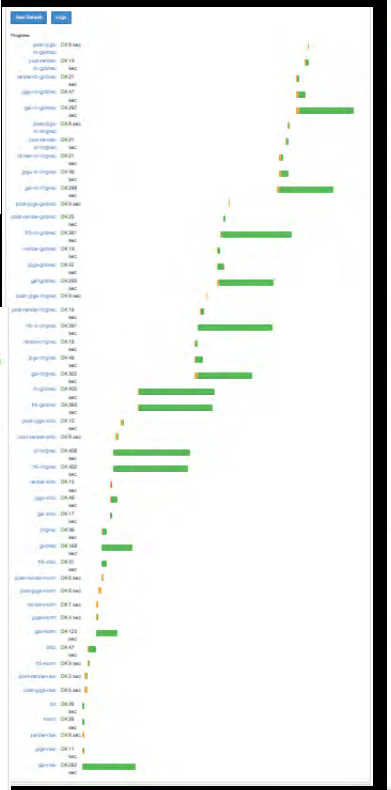
Evolution of BL832 Workflow & Throughput

10/2013

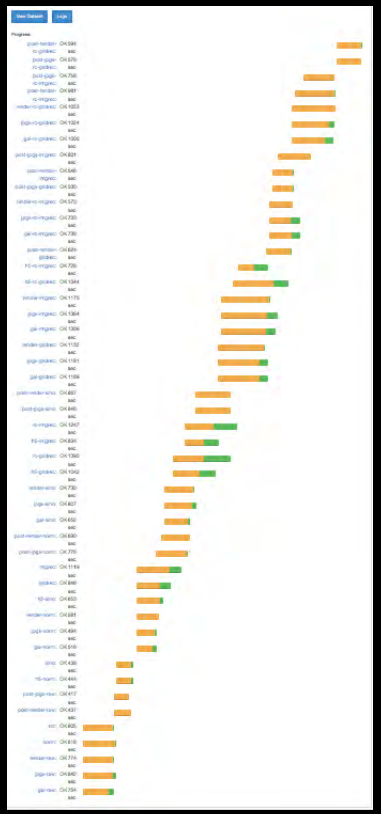


04/2014

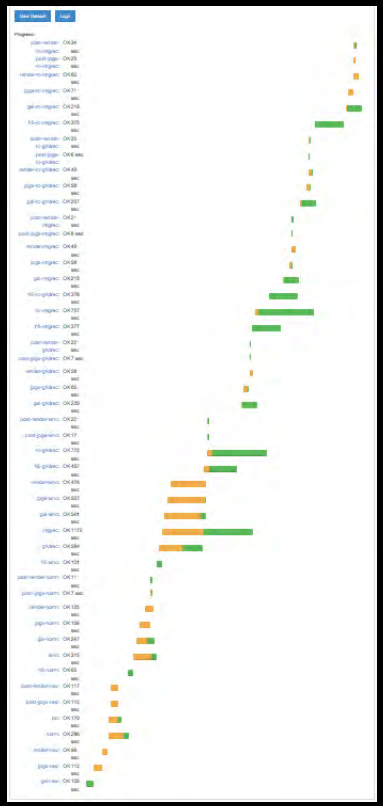
10/2014



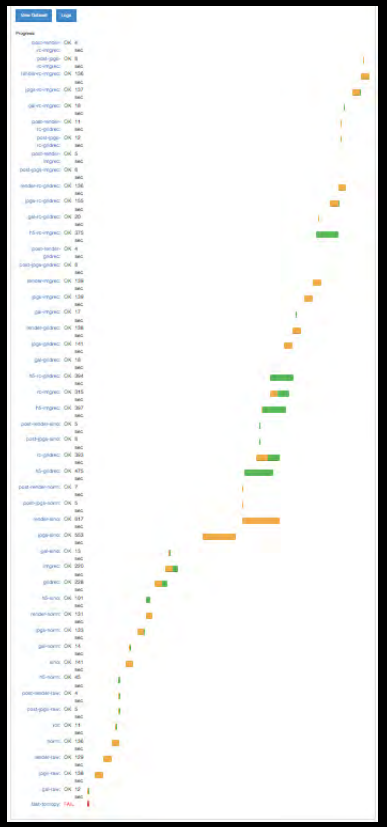
04/2015



10/2015



Now



CETull@lbl.gov - 14 April 2016

Xi-cam: Scattering data reduction and visualization

Scientific Achievement

Development of a community-maintainable platform for new analysis and visualization techniques in SAXS/GISAXS, extensible for other characterization methods

Significance and Impact

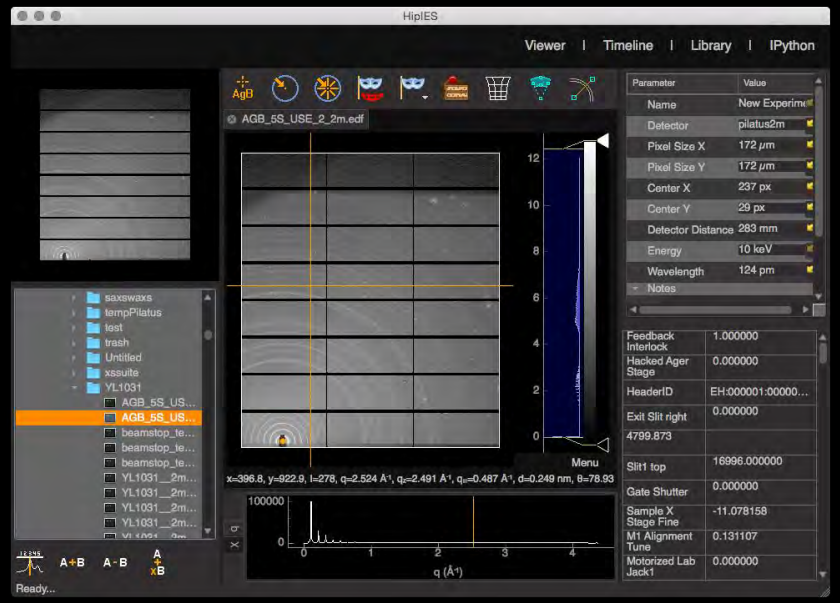
New visualization techniques extract further information from scattering data. Open source modular design allows collaborative development and resource consolidation across synchrotron facilities.

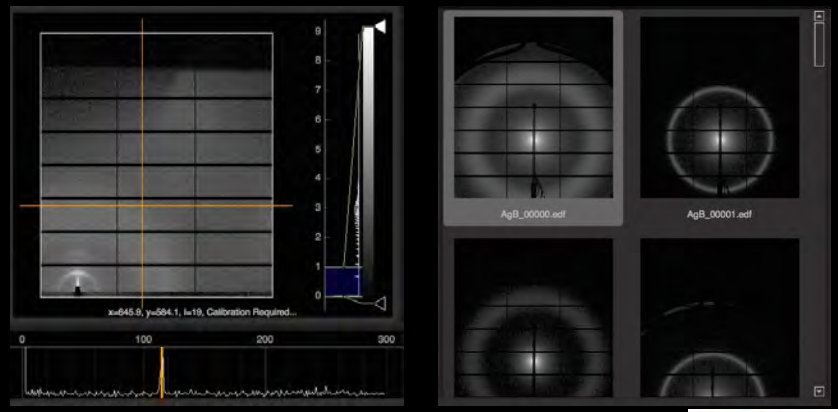
Research Details

Remote processing with HPC for real-time high data rate analysis

NERSC/SPOT remote data access for high-volume data archive retrieval

Highly interactive responsive design improves user control and accessibility





The project is in collaboration with CAMERA

* Pandolfi, R., Venkatakrishnan, Kumar, D., Hexemer, A. *(under preparation)*

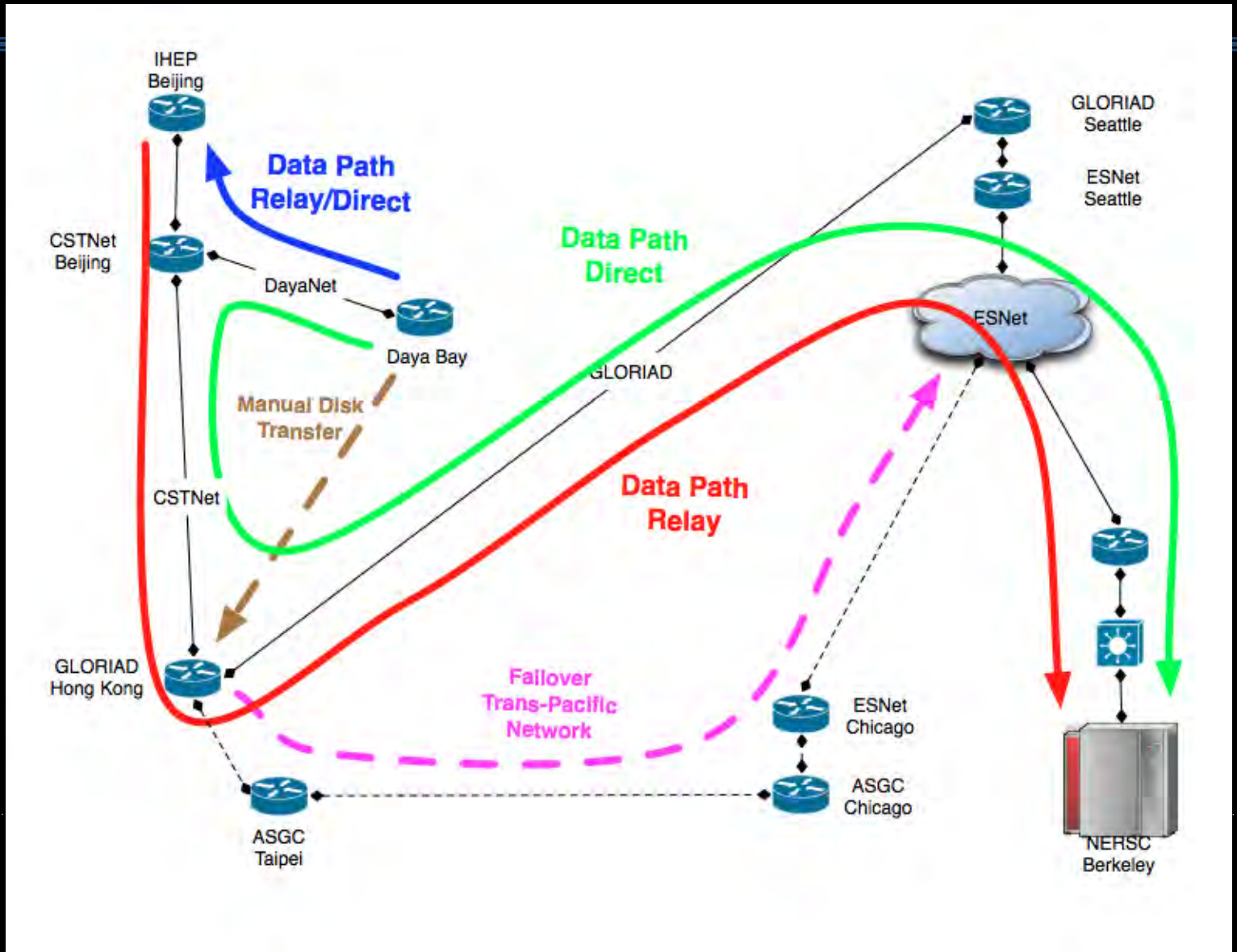
SPOT Suite presents interfaces usable from a wide variety of tools and applications from the wider BES community – BES Analytics Ecosystem

Ecosystem will transform the way researchers do science

- **Automation of workflows & dataflows alread enhancing the capabilities of beamlines**
- **Composible, API-based ecosystem allows contributions from many sources – CAMERA, BES facilitie**
- **ASCR/cloud, lightsource, and custom resou bound together with networking & workflow**
- **Advanced mathematical techniques, machin learning, Software Defined Networking, new architectu**

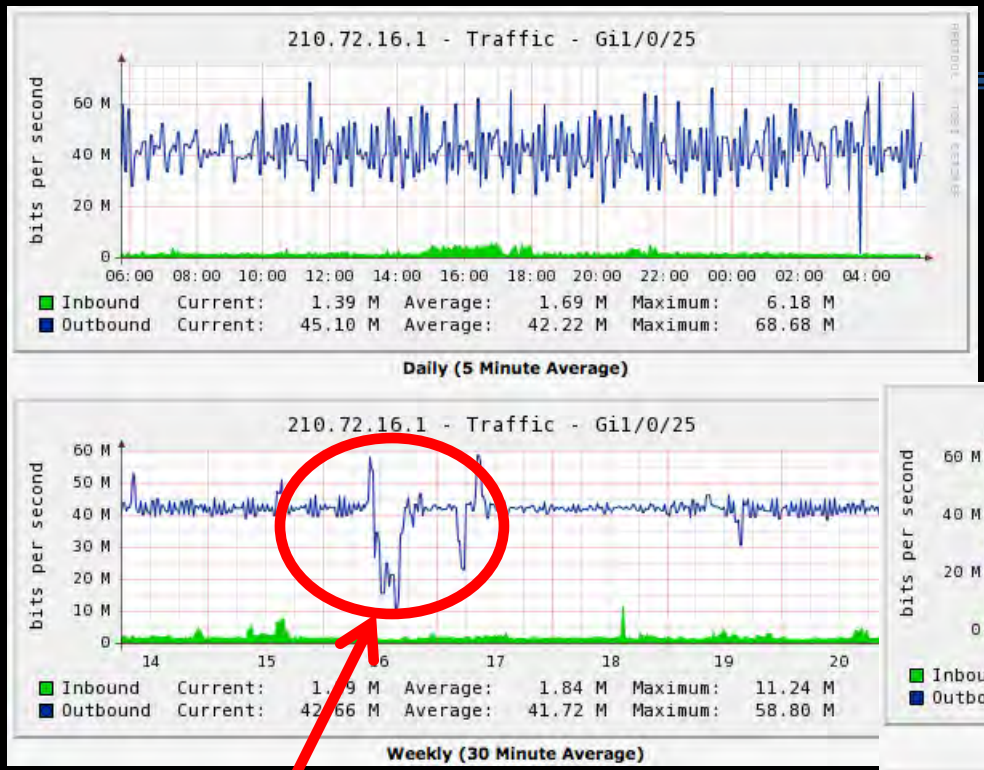
- **Bespoke workflow systems for individual facilities and beamlines are much less effective than scientific**

Daya Bay – "typical" HEP experiment

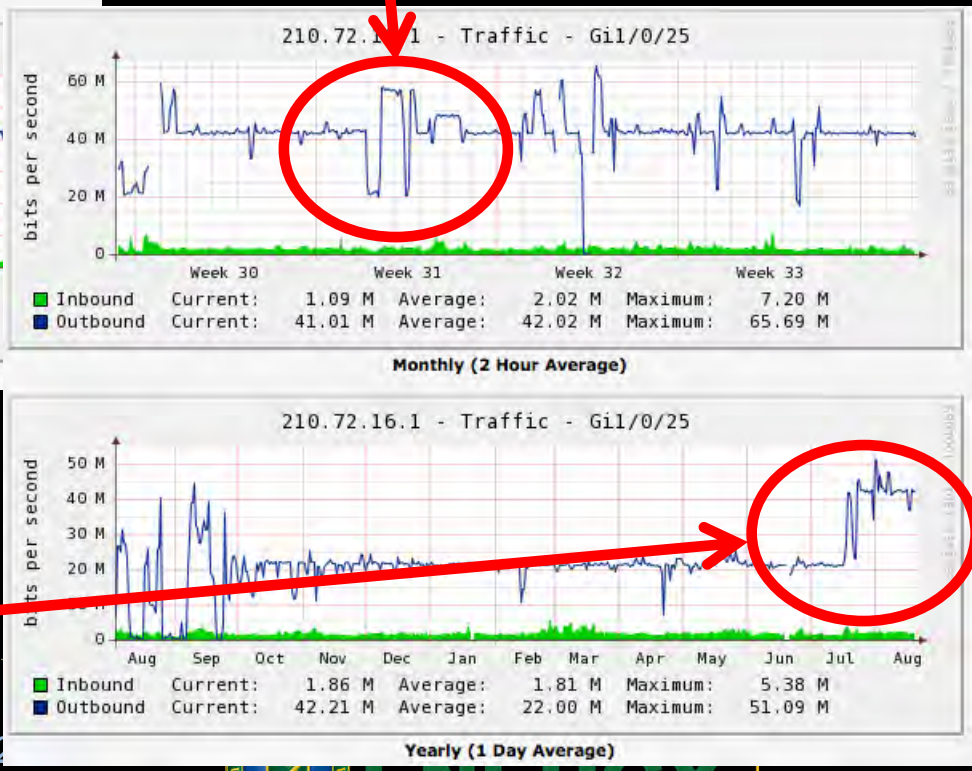


Office of Science

DYB SPADE data rate is stable



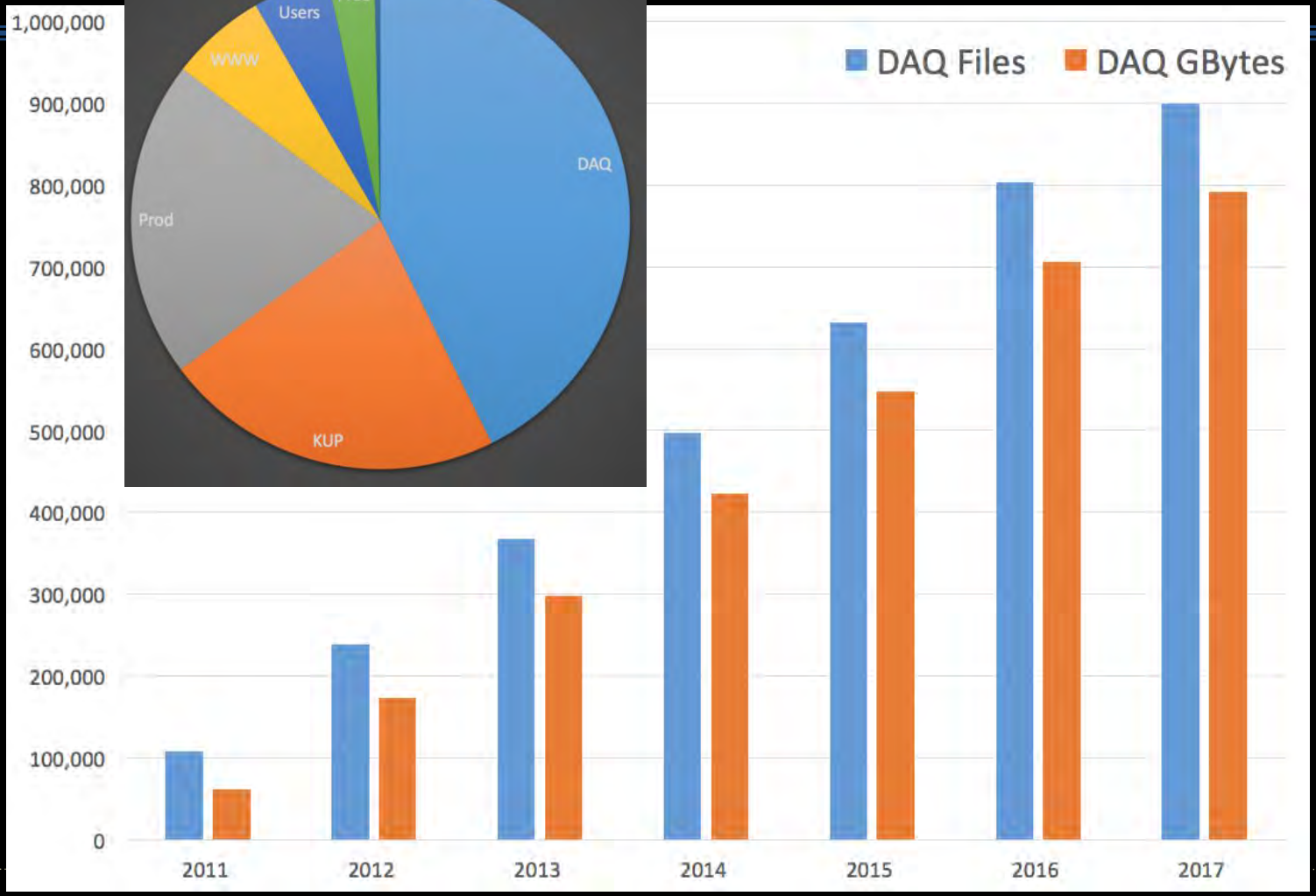
Automatic recovery from temporary outages.



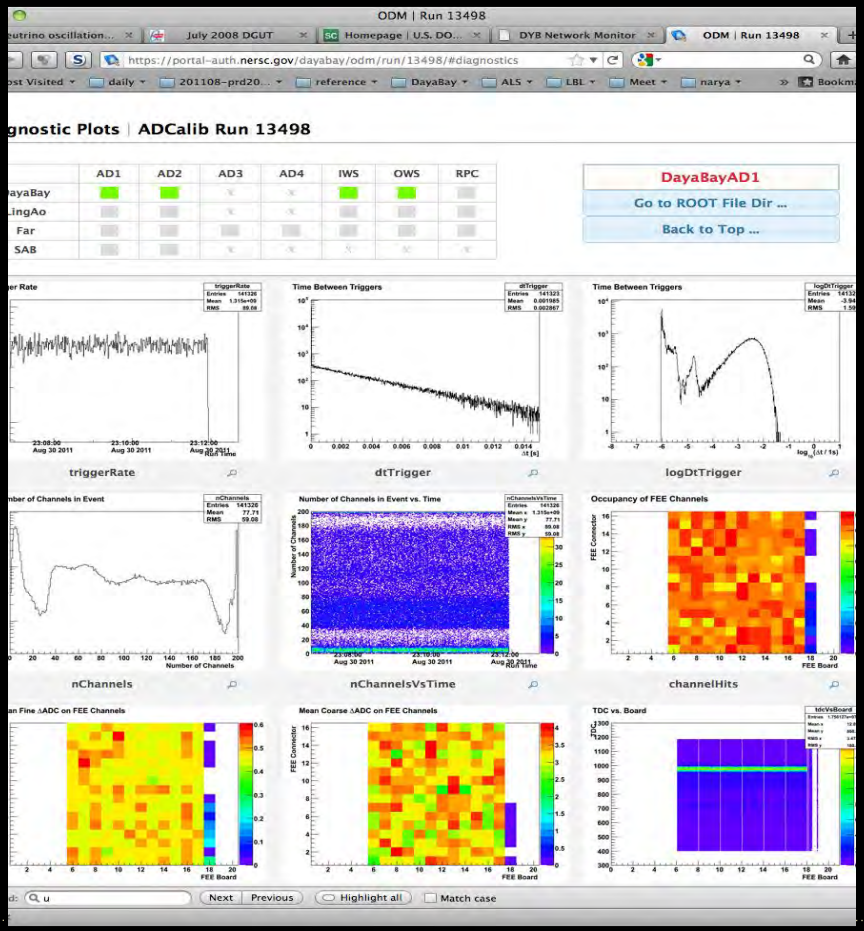
Friday Calibration

Change from “Relay” to “Direct” mode.

DYB Raw data 900TB (Sep2017) and 4[illegible]ge.



Science Data Gateway-based ODM



- Realtime analysis of data at NERSC
- Data transferred over 4 national & international research networks.
- Automated KUP (Keep Up Production)
- NuWa production job using DBI and Offline DB execute in ~1.5 Hours.
- Offline, DCS, and DAQ DBs aggregated with onsite ELog entries and output of NuWa.
- Available globally to collaborators.
- Reference plots and IHEP plots used to validate consistency.
- Data and Root files can be downloaded from GFS/HPSS.
- Best tool for immediate feedback on physics quality and used by onsite shifters and QA (offsite) shifters.
- Basis of Remote Shift system.

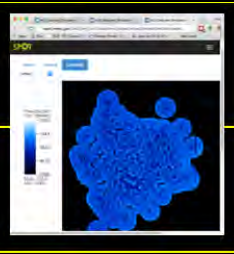
Presentation Layers

SPOT Services

Compute Resources

Applications

SPOT Web Portal Interface



Xi-CAM Local GUI

SPOT Presentation Templates

Xi-CAM Plugins

SPOT Metadata & Provenance DB

SPOT Task Queue & Workflow Manager

SPOT Data Management & Resource Matching

SPOT Computer Resource API (NEWT-based)

SPOT Resource APIs @ NERSC

SPOT Archive

SPOT Warehouse

SPOT Workers

SPOT Resource APIs @ OLCF

SPOT Warehouse

SPOT Workers

...

SPOT Application APIs

CAMERA Algorithms & Executables

HipGISAXS

Image Analysis

Reconstruction

...

3rd Party Algorithms & Executables

TomoPy

DPDAK

...

SPADE used for production orchestration of network data movement

- SPADE developed in IceCube, used in Daya Bay & ALS
- Underlying protocols: scp, bbcp, gridftp, Globus Online, RDMA?
- Highly Configurable: push, pull, relay, local
- Integrated warehouse, catalog, monitoring; Highly instrumented