Argonne NATIONAL LABORATORY

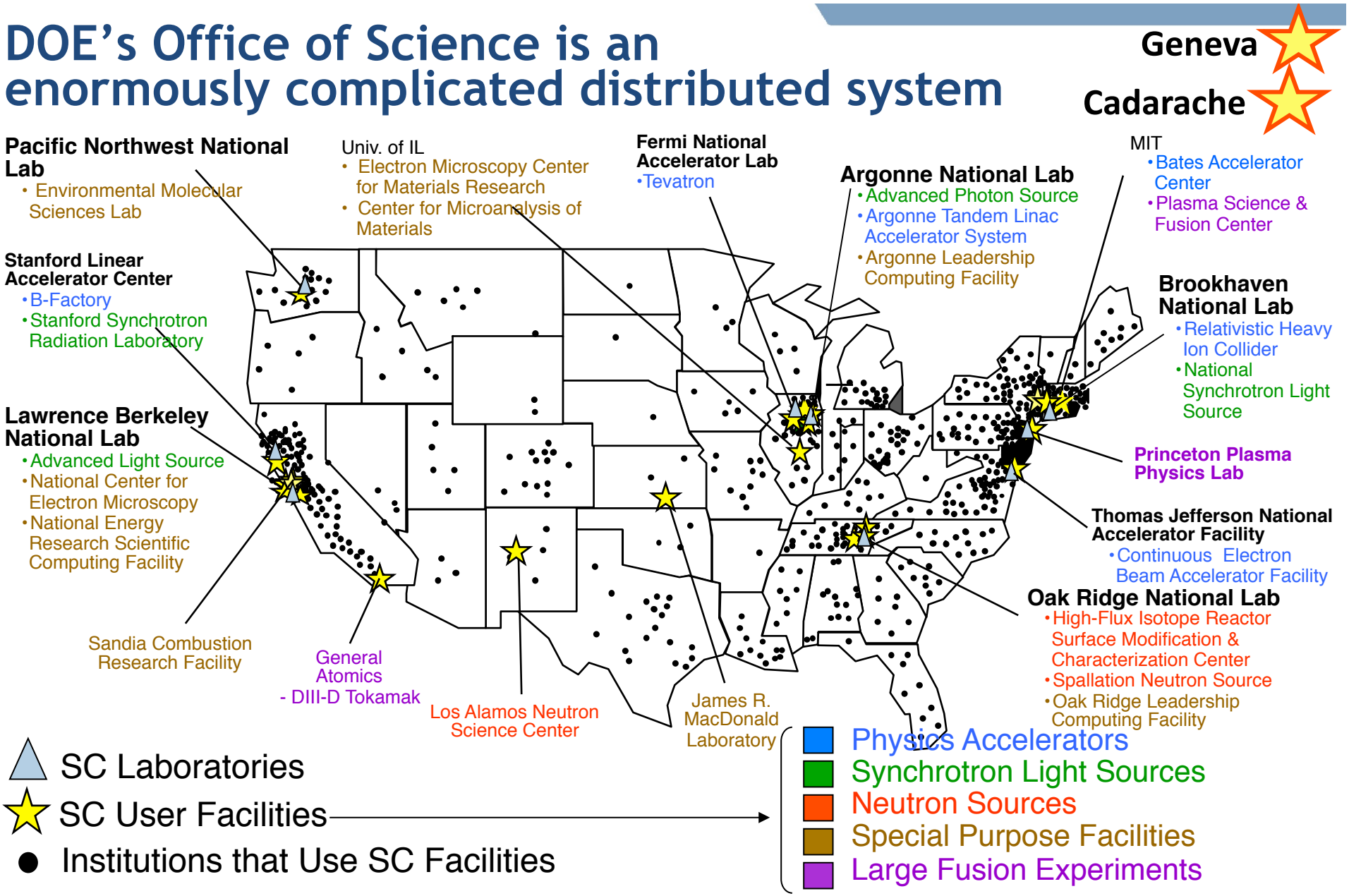
Distributed Science, Data Deluge, and SaaS

Raj Kettimuthu

Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY

DOE's Office of Science is an enormously complicated distributed system

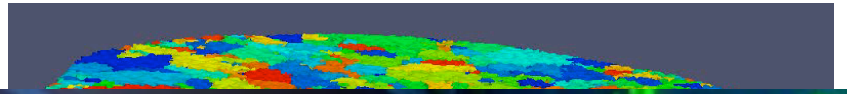


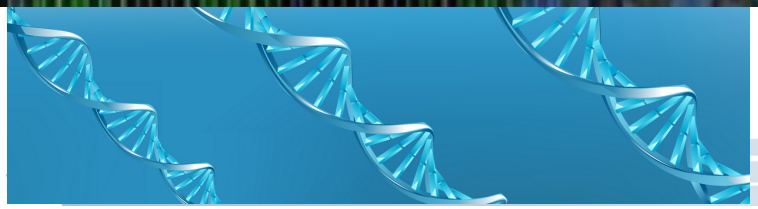
$5B/yr supports **26,000** investigators at **300** academic institutions and all DOE laboratories; **27,000** researchers use scientific user facilities

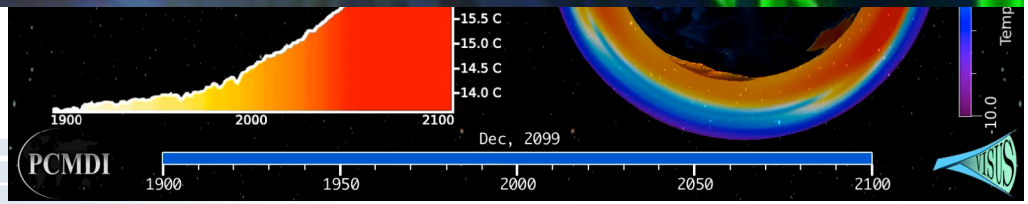
Data Deluge

Cosmology

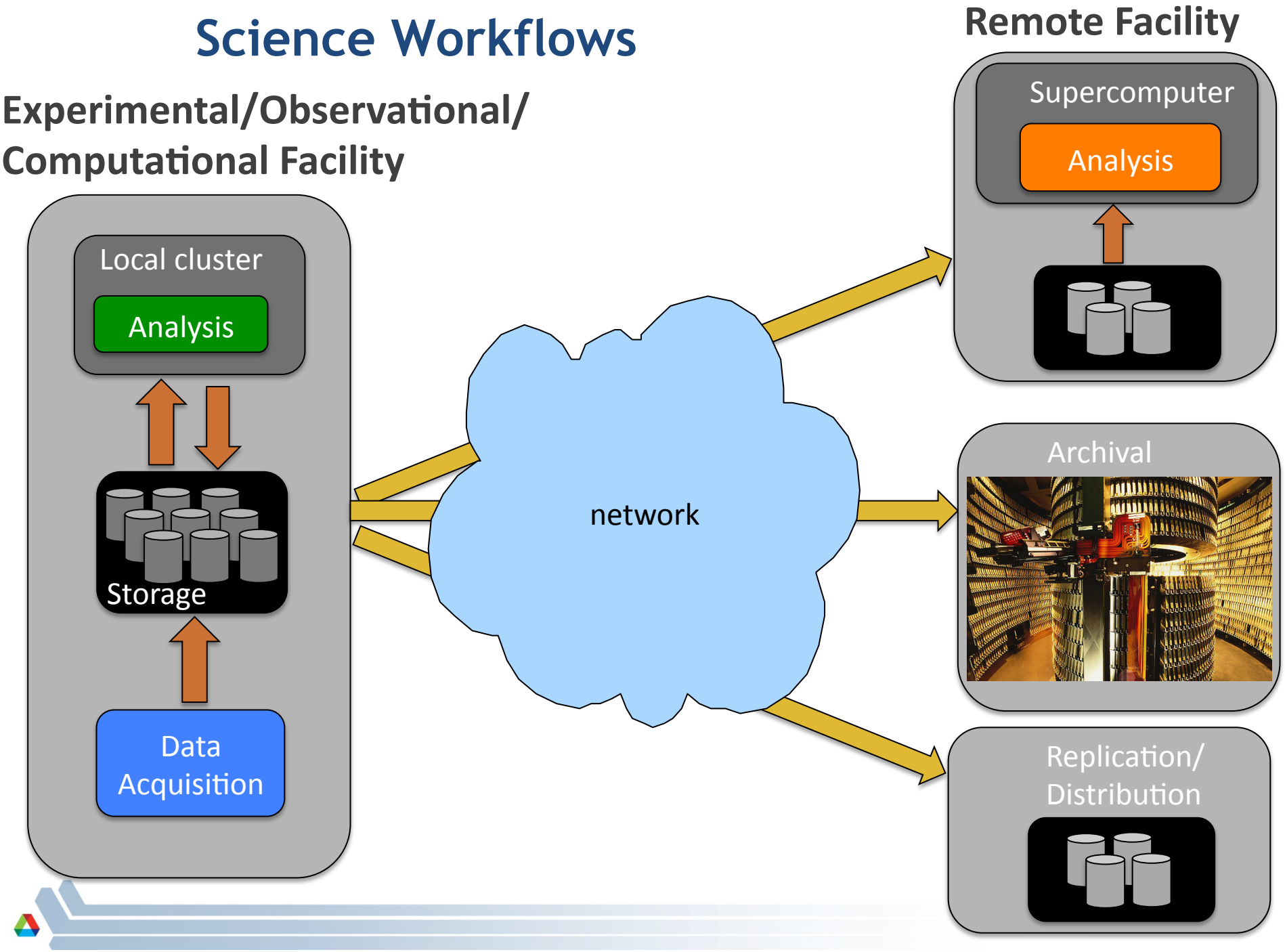
Light Source Facilities



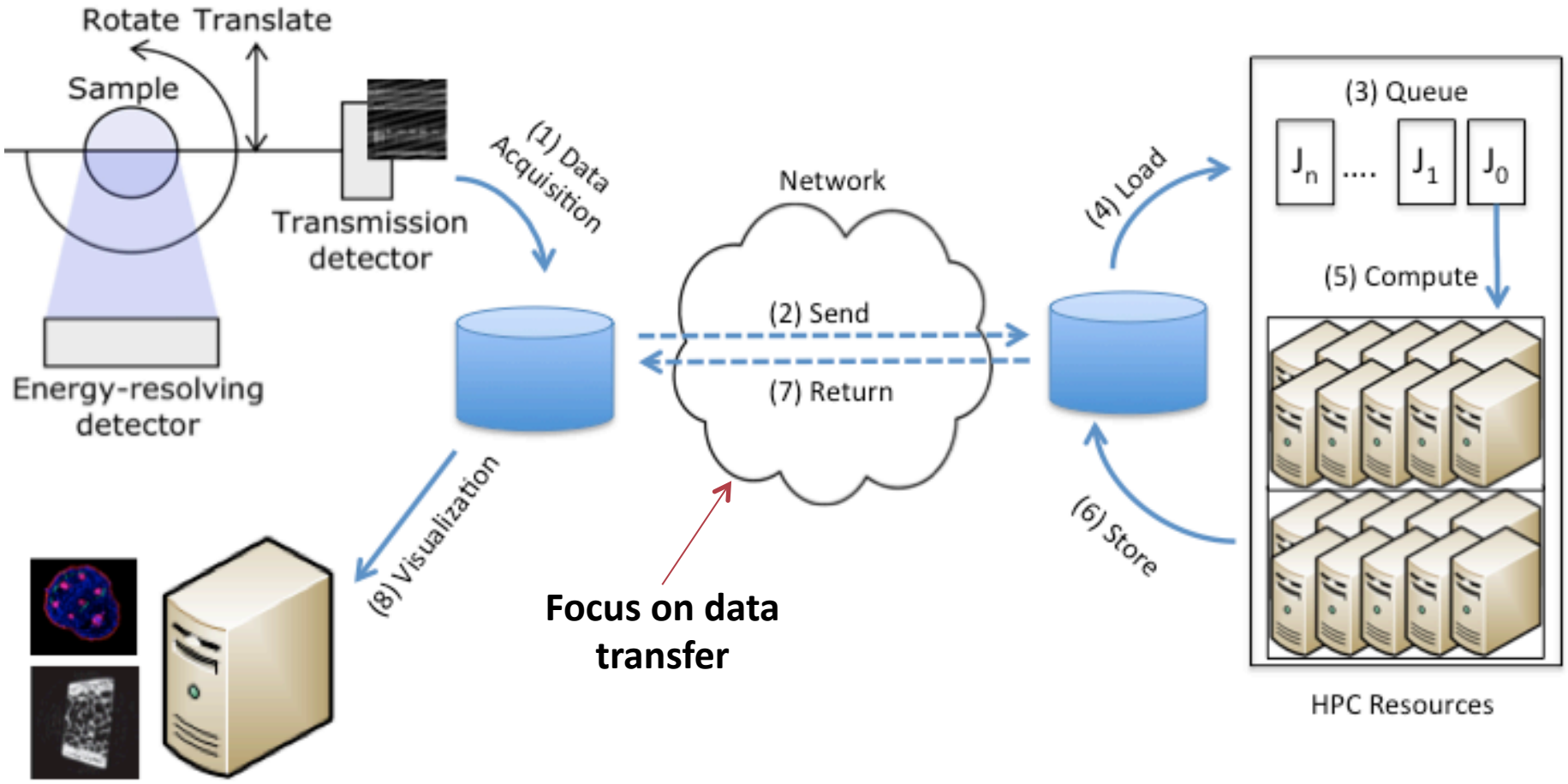




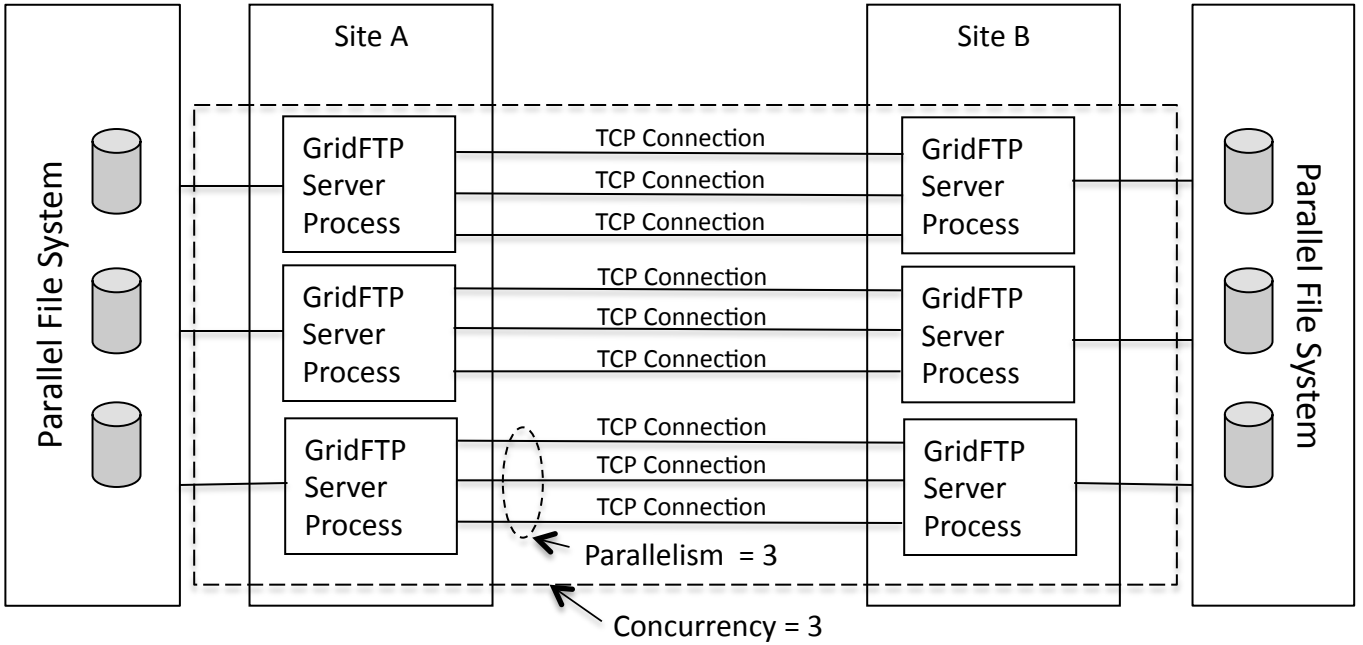
Science Workflows



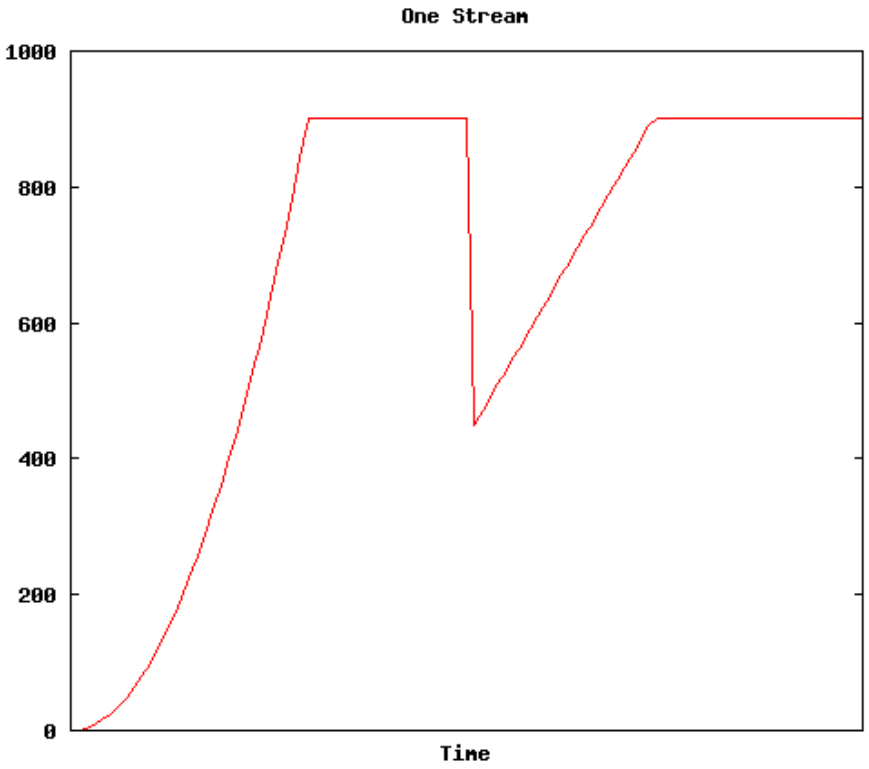
Science workflow - current

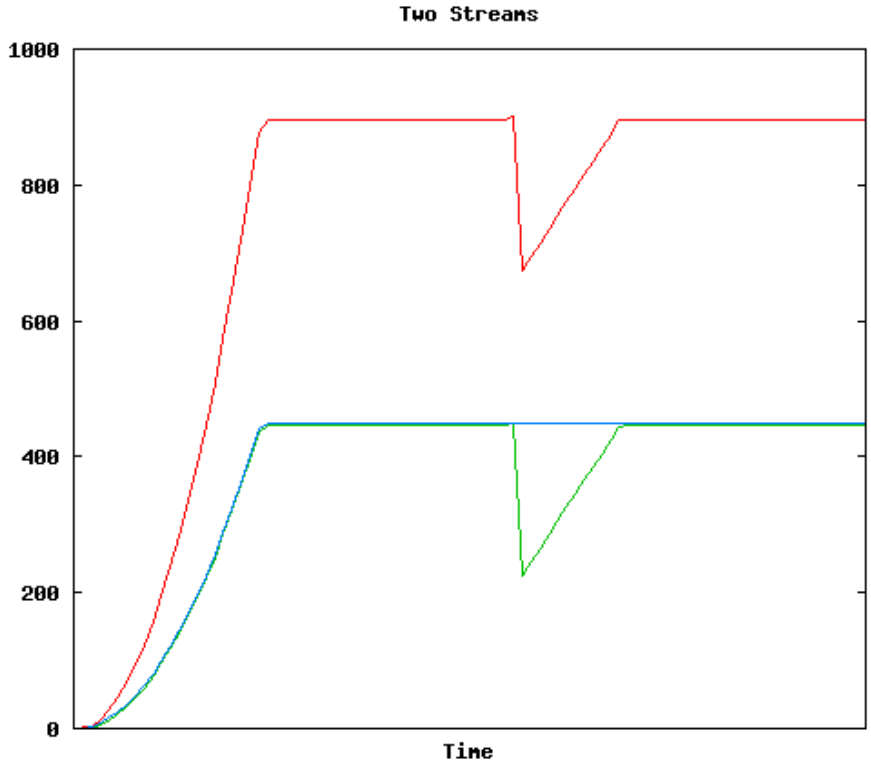


GridFTP – parallelism, concurrency, and multi-node data movement

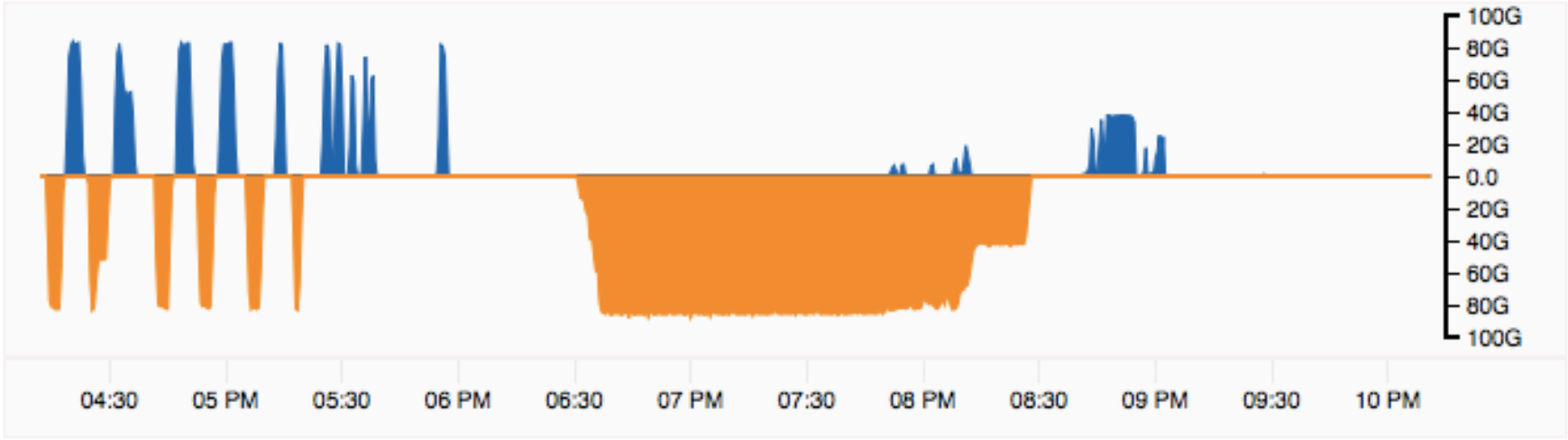


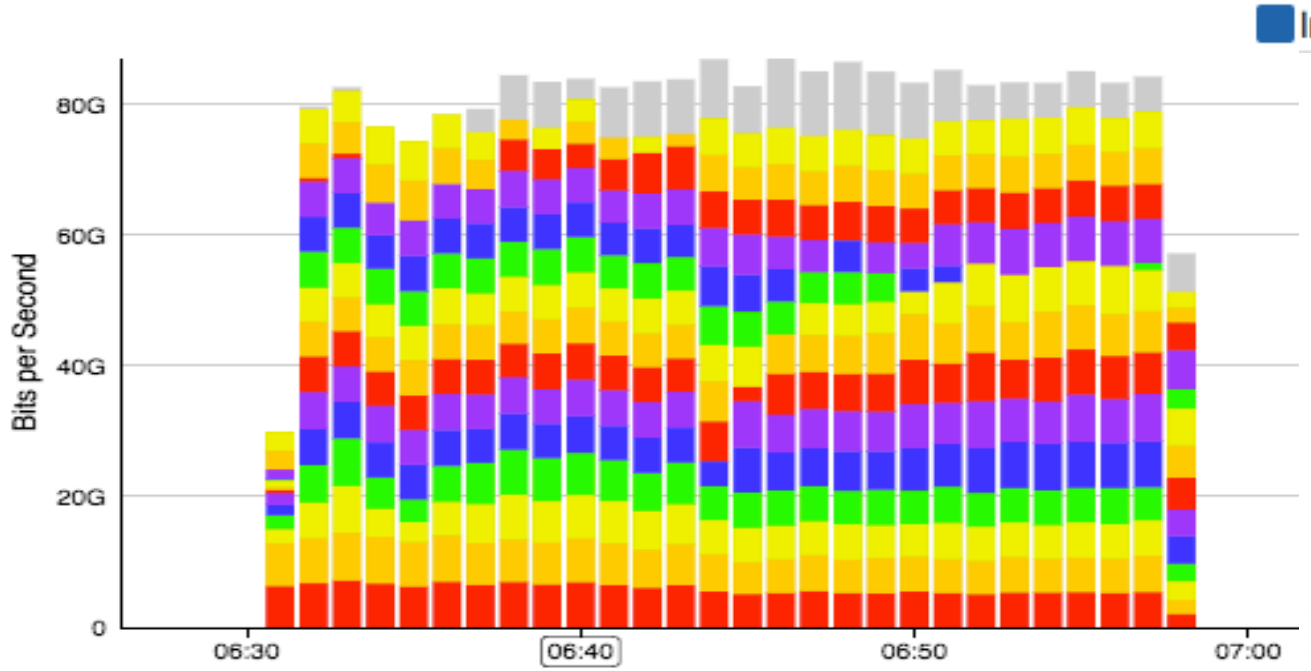
Parallel TCP streams





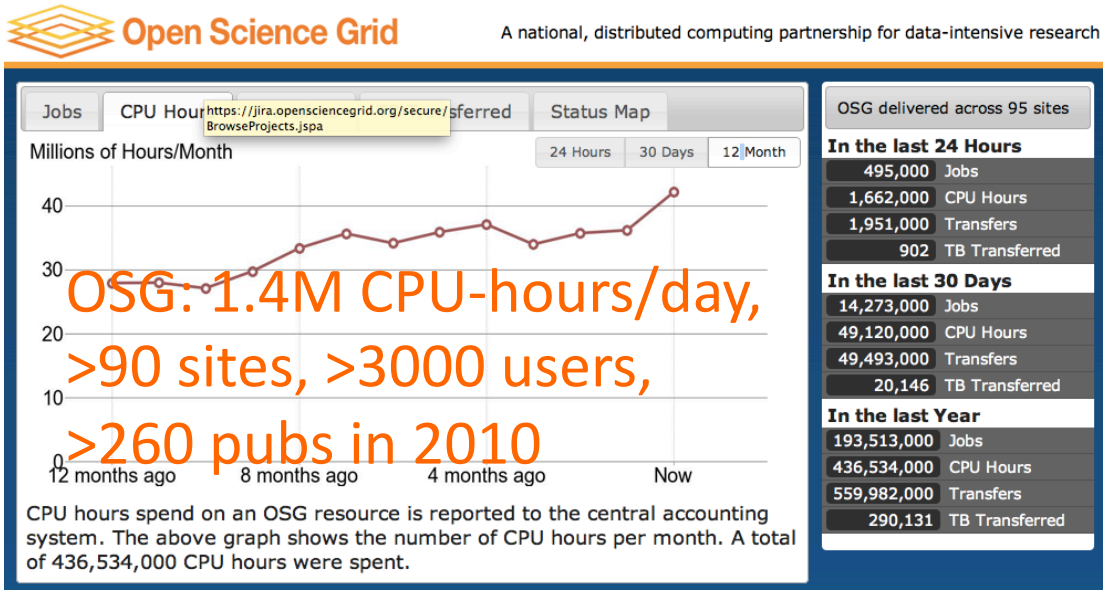
85 Gbps Sustained Disk-to-Disk over 100 Gbps Network, Ottawa—New Orleans





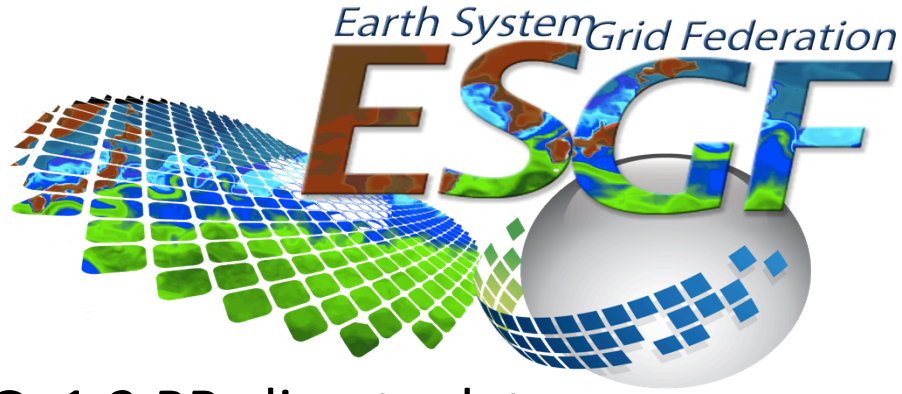
GridFTP underpins many science projects & facilities

LIGO: 1 PB data in last science run, distributed worldwide



OSG: 1.4M CPU-hours/day, >90 sites, >3000 users, >260 pubs in 2010

Robust production solutions
Substantial teams and expense
Sustained, multi-year effort
Application-specific solutions, built on common technology



ESG: 1.2 PB climate data delivered to 23,000 users; 600+ pubs

All build on GridFTP and other Globus Toolkit software

Small science struggling

More data, more complex data

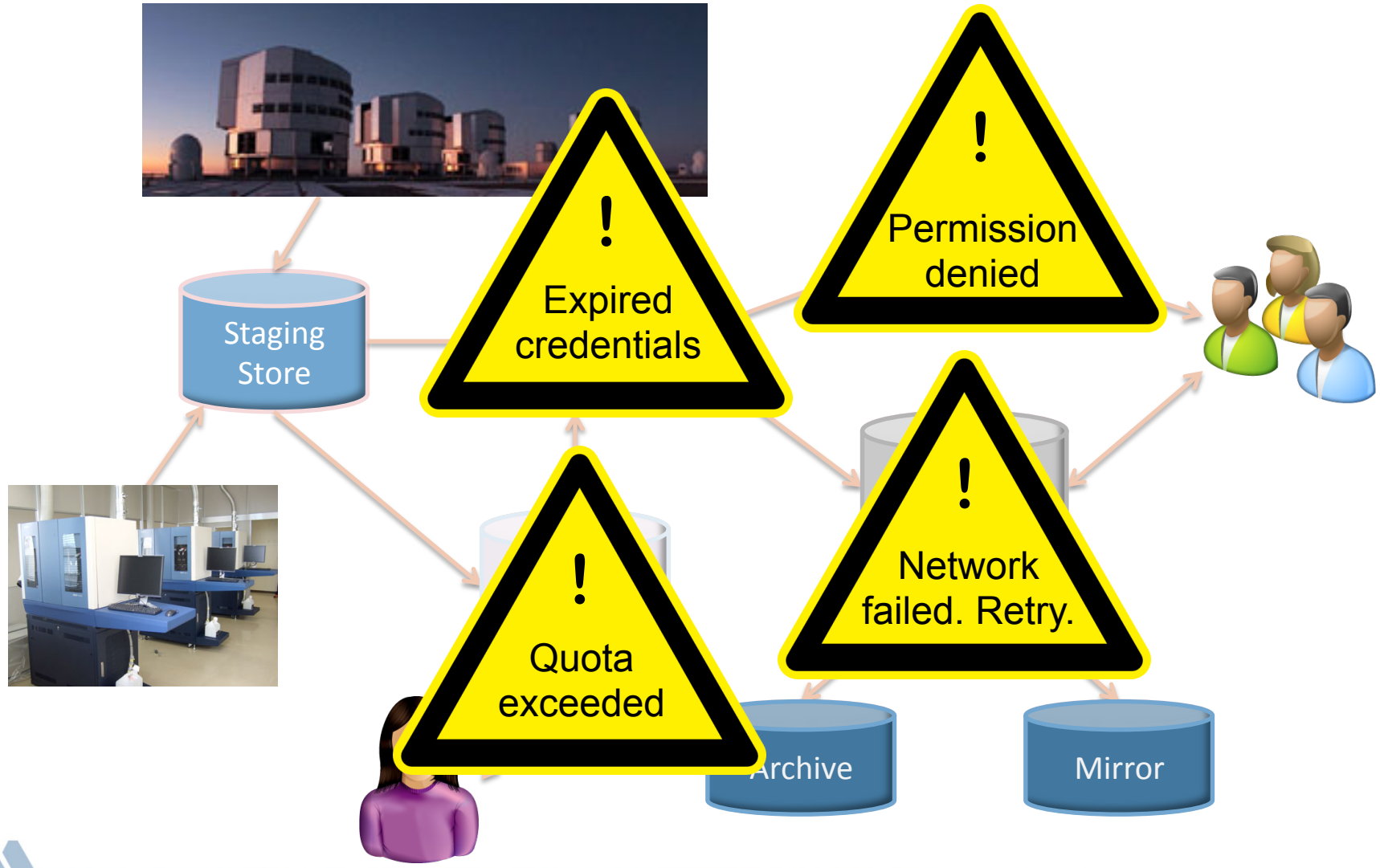
Ad-hoc solutions

Inadequate software, hardware

Poor performance

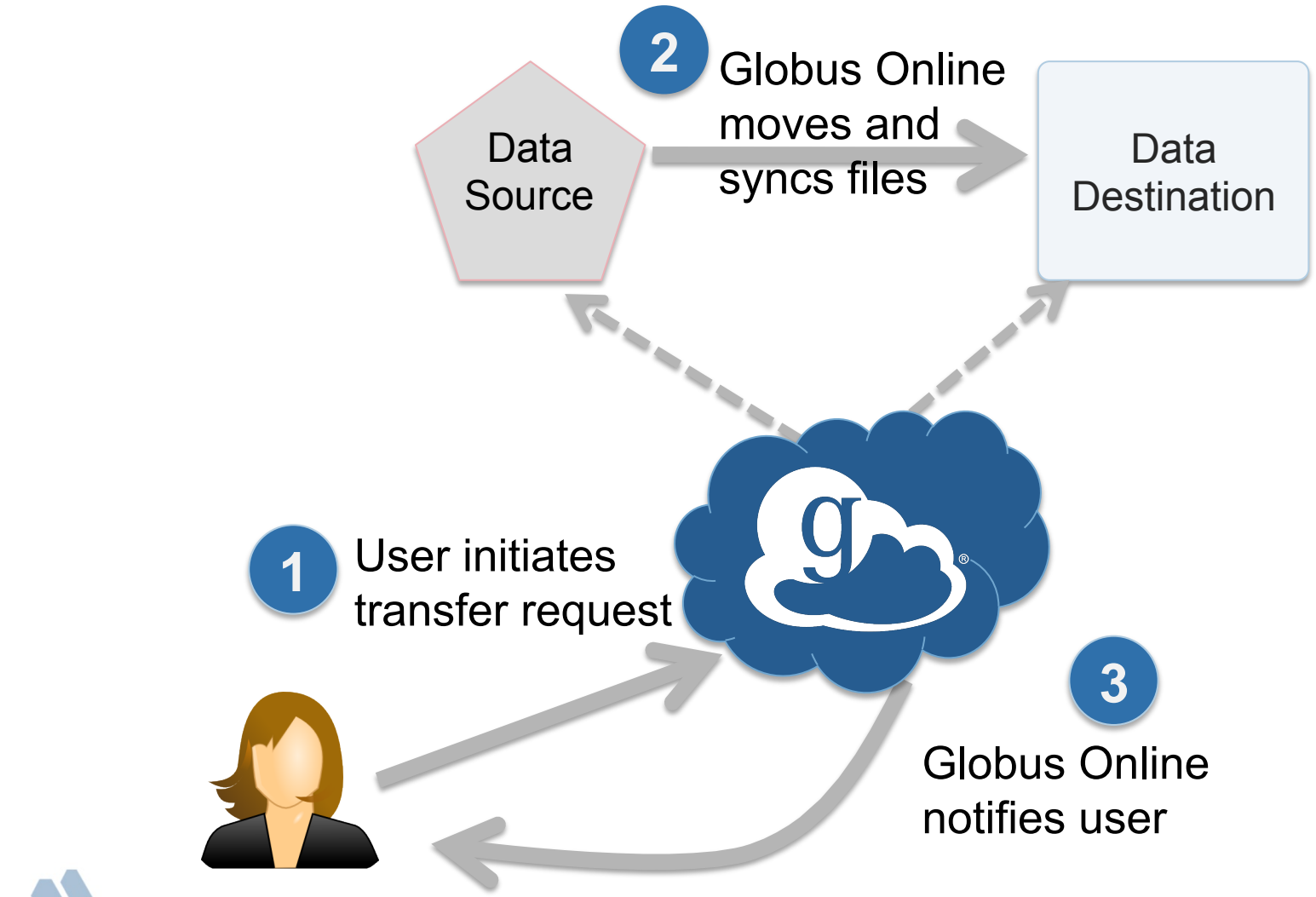
Moving data should be easy …

…but it's hard and frustrating



SaaS

Globus Online



globus

Products Pricing Developers Support Log In

Research data management simplified.

175,462,132,924 MB TRANSFERRED

Researchers

Focus on your research, not IT problems. We make it easy to move, manage, and share big data.

LEARN MORE

GET STARTED

Resource Providers

Globus gives you more control over your data infrastructure, while providing excellent ease-of-use for your researchers.

LEARN MORE

GLOBUS SUBSCRIPTIONS

Our Users

Researchers and resource providers are our greatest inspiration and we love it when they say nice things about Globus.

USER QUOTES

CASE STUDIES

Fast, Reliable, Secure File Transfer

UPCOMING EVENTS

Subscriptions for Non-Profit Research and Education

If you wish to use Globus in a commercial setting, you must have a commercial subscription.

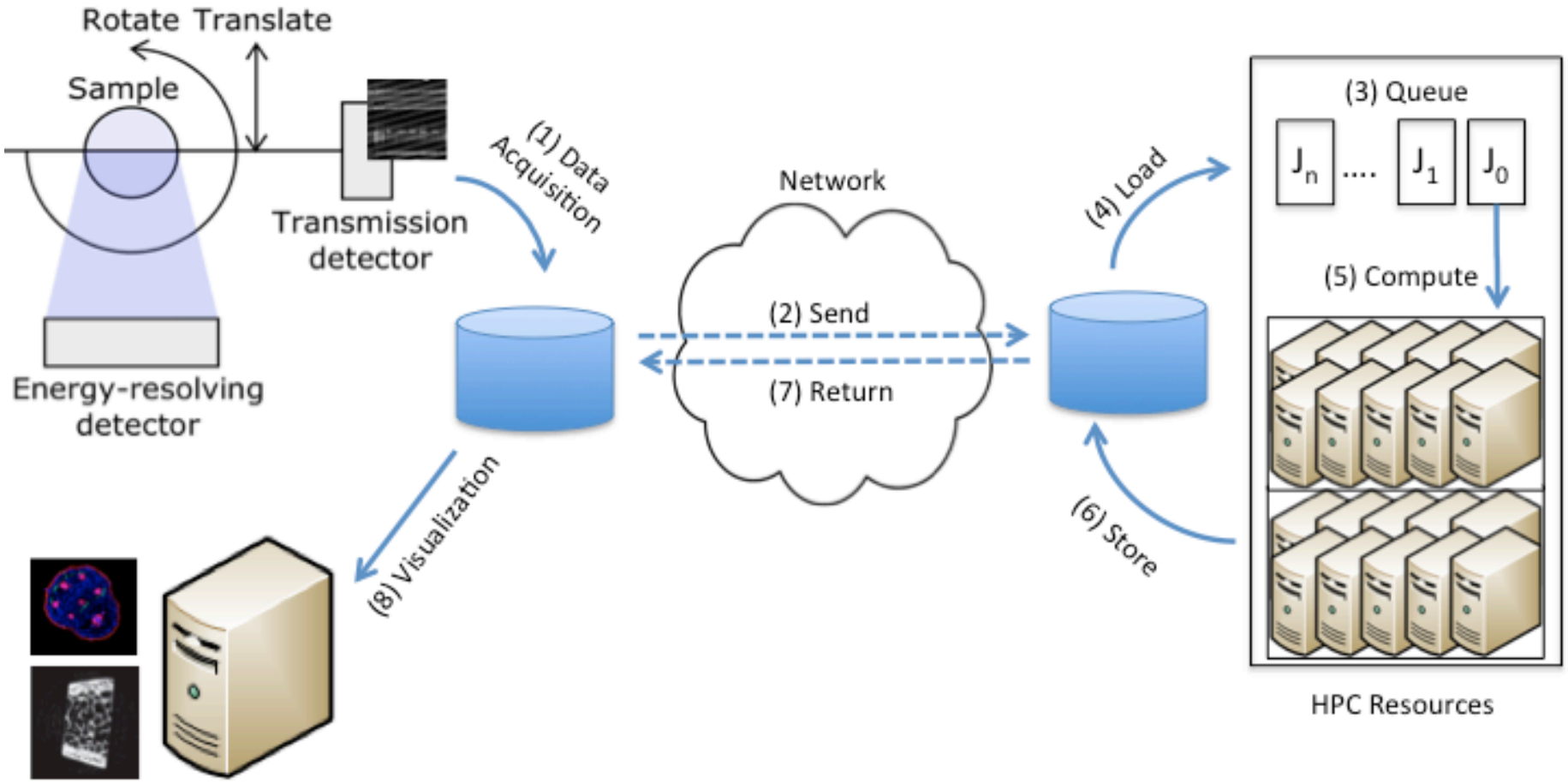
Features (click ℹ for description)	**Basic - Free**	**Starter Subscription**	**Standard Subscription**
Transfer level ℹ	Unlimited	Unlimited	Unlimited
User level ℹ	Unlimited	Unlimited	Unlimited
Managed endpoints ℹ	None	1	Unlimited
Management console ℹ	–	✓	✓
Usage reports ℹ	–	✓	✓
Support for Globus Connect, Web, CLI ℹ	–	✓	✓
Shared endpoints ℹ	–	✓	✓
Globus Plus users ℹ	–	–	✓
Data publication ℹ	–	–	✓
Application integration support ℹ	–	–	✓
HTTPS support *(coming soon)* ℹ	–	–	✓
Support service level ℹ	–	Monday-Friday, 9am-5pm Central; 1-business day response	
Named support contacts ℹ	–	1	5
Pricing	Free	Contact us for subscription pricing	

Globus by the numbers

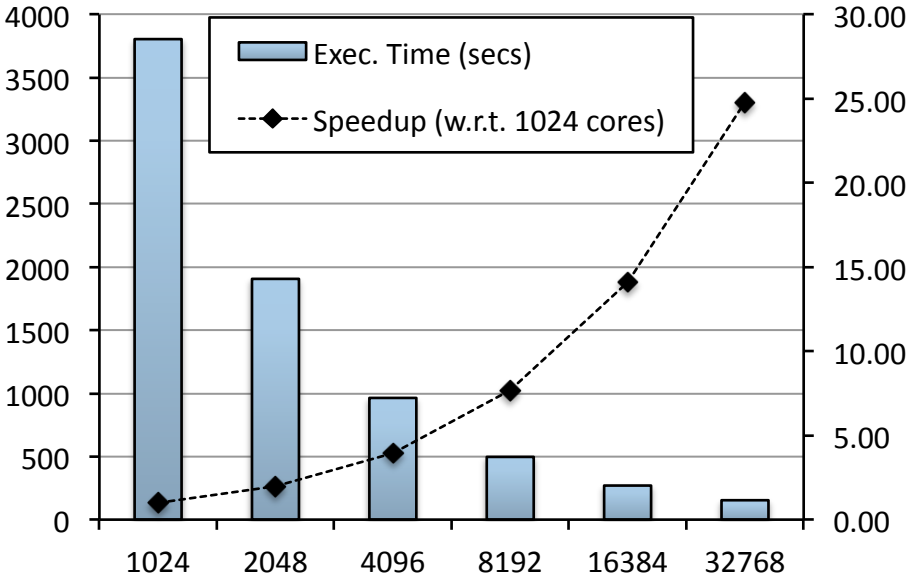
4 major services	**175 PB** transferred	**30 billion** files processed	**43,000** registered users
13 national labs use Globus	**2,500** active endpoints	**450+** active daily users	**99.9%** uptime
35+ institutional subscribers	**1 PB** largest single transfer to date	**3 months** longest continuously managed transfer	**130** federated campus identities

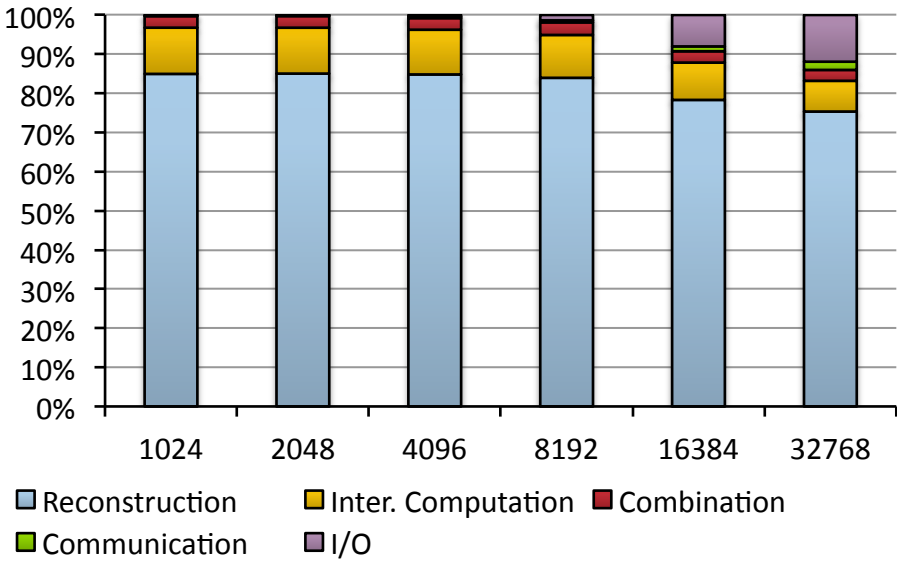
Research challenges

Science workflow - current

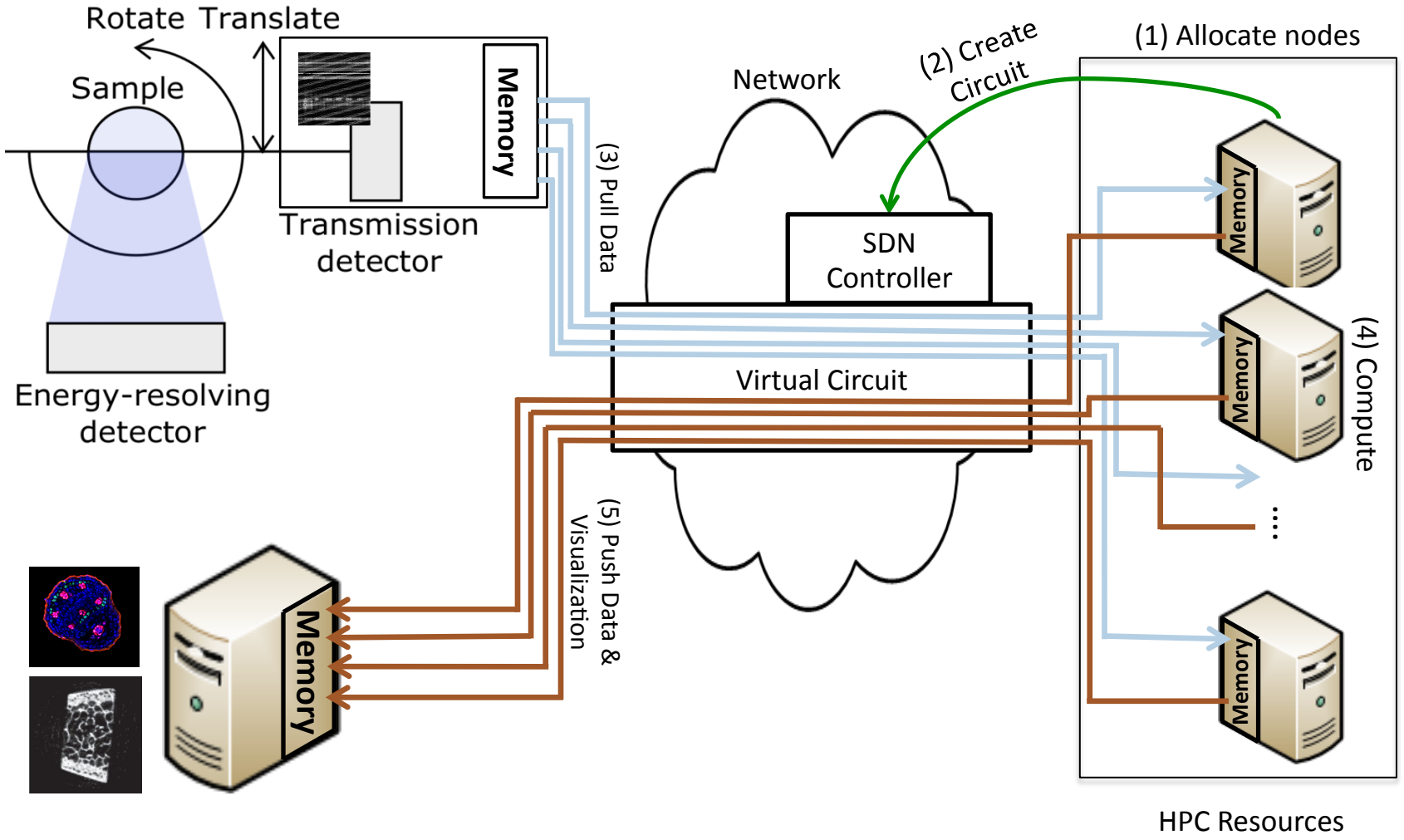


Iterative tomographic reconstruction





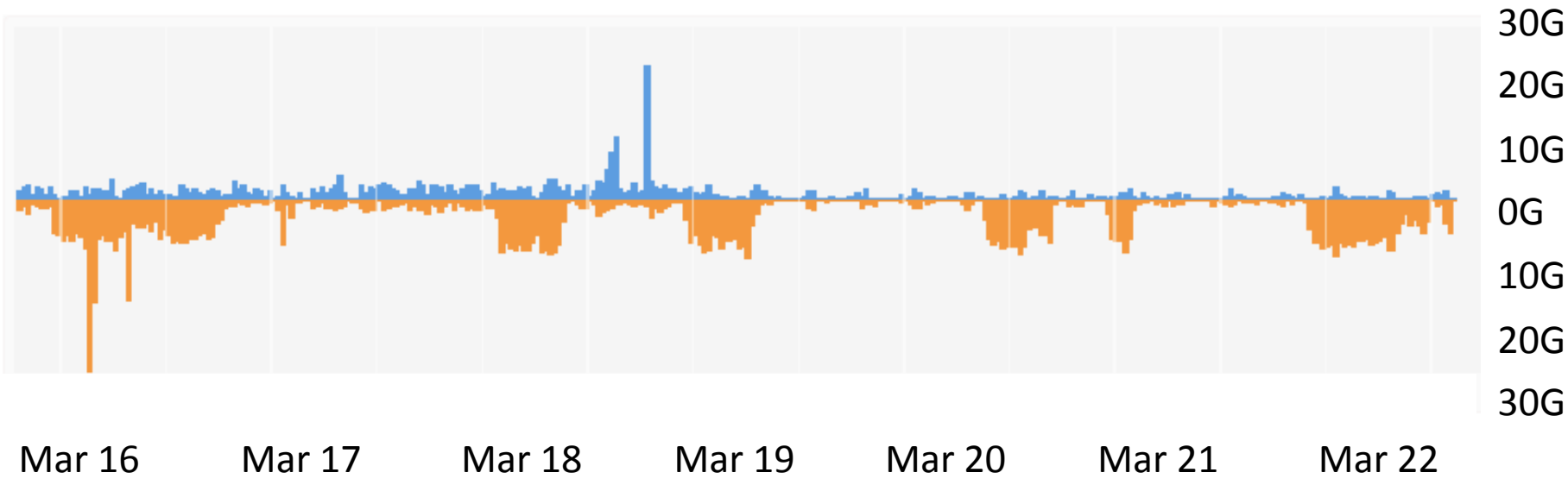
Science workflow - future



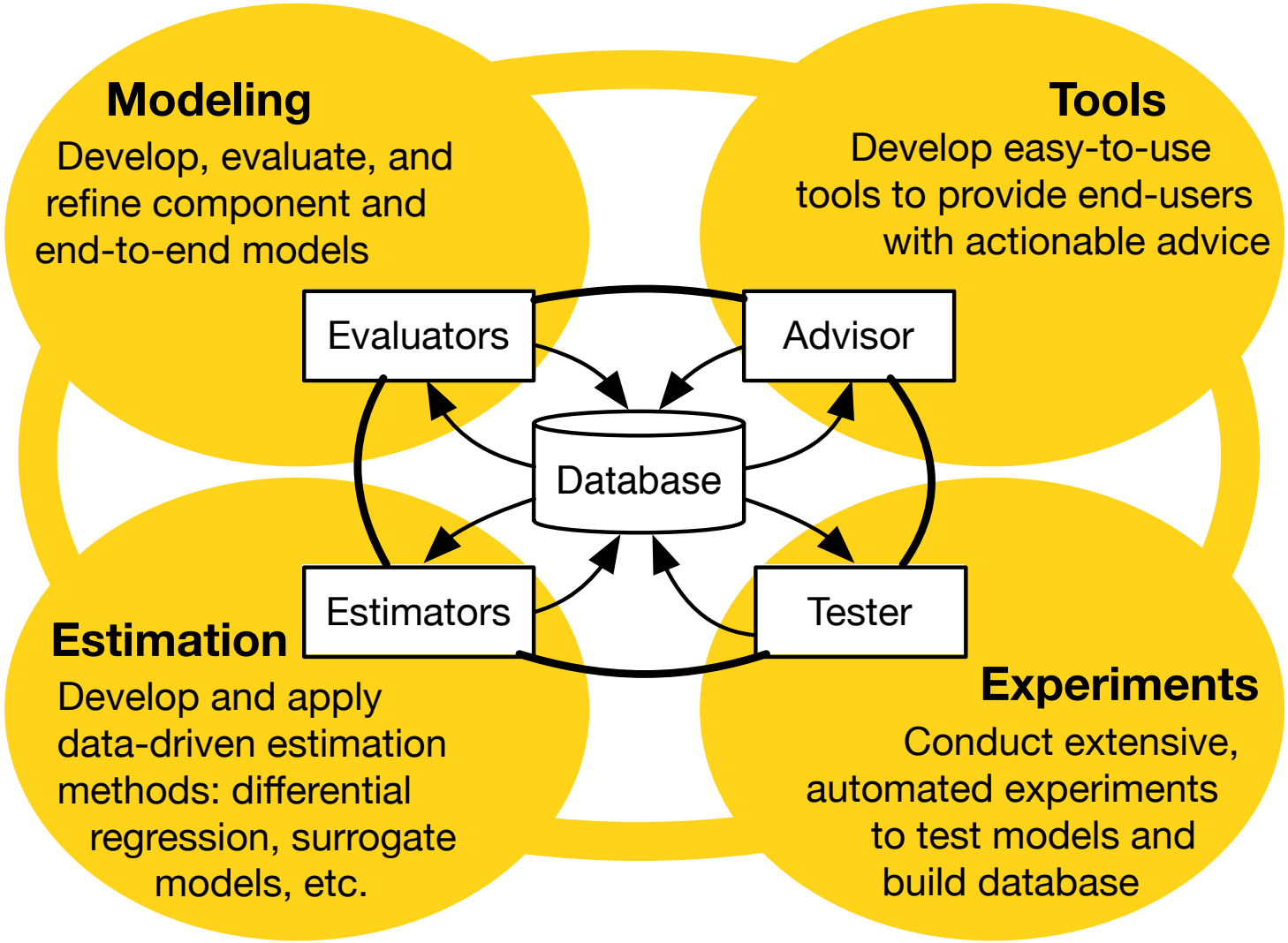
On-demand computing in supercomputers

- Real-time queue?
  - Higher charge rate
  - NERSC has modest number of nodes allocated
- Steal resources that service the batch queue?
  - Preempt batch jobs
  - Low-priority queues?
    - Lower charge rate
- How to keep the average utilization of the system high?
  - Bursty real-time loads
- Checkpointing
  - User or system?
  - Restart overhead

Peak vs. average network utilization



Robust Analytical Models for Science at Extreme Scale (RAMSES) project



Questions