



**XSEDE**  
Extreme Science and Engineering  
Discovery Environment



# BRIDGES

A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE

Connecting Researchers, Data & HPC

Nick Nystrom • Director, Strategic Applications & Bridges PI • [nystrom@psc.edu](mailto:nystrom@psc.edu)  
July 1, 2015

# The Shift to Big Data

## *New Emphases*



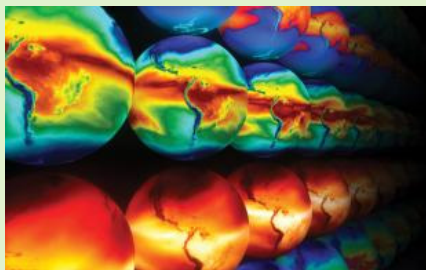
Pan-STARRS telescope

<http://pan-starrs.ifa.hawaii.edu/public/>



Genome sequencers

(Wikipedia Commons)



NOAA climate modeling

[http://www.ornl.gov/info/ornlreview/v42\\_3\\_09/article02.shtml](http://www.ornl.gov/info/ornlreview/v42_3_09/article02.shtml)



Social networks and the Internet



Video

Wikipedia Commons



Library of Congress stacks

<https://www.flickr.com/photos/danlem2001/6922113091/>



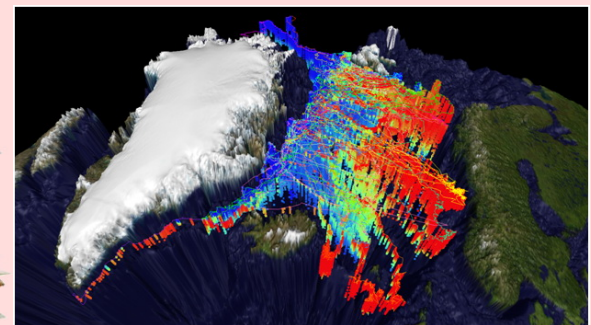
Collections

Horniman museum: [http://www.horniman.ac.uk/get\\_involved/blog/bioblitz-insects-reviewed](http://www.horniman.ac.uk/get_involved/blog/bioblitz-insects-reviewed)



Legacy documents

Wikipedia Commons



Environmental sensors: Water temperature profiles from tagged hooded seals

[http://www.arctic.noaa.gov/report11/biodiv\\_whales\\_walrus.html](http://www.arctic.noaa.gov/report11/biodiv_whales_walrus.html)

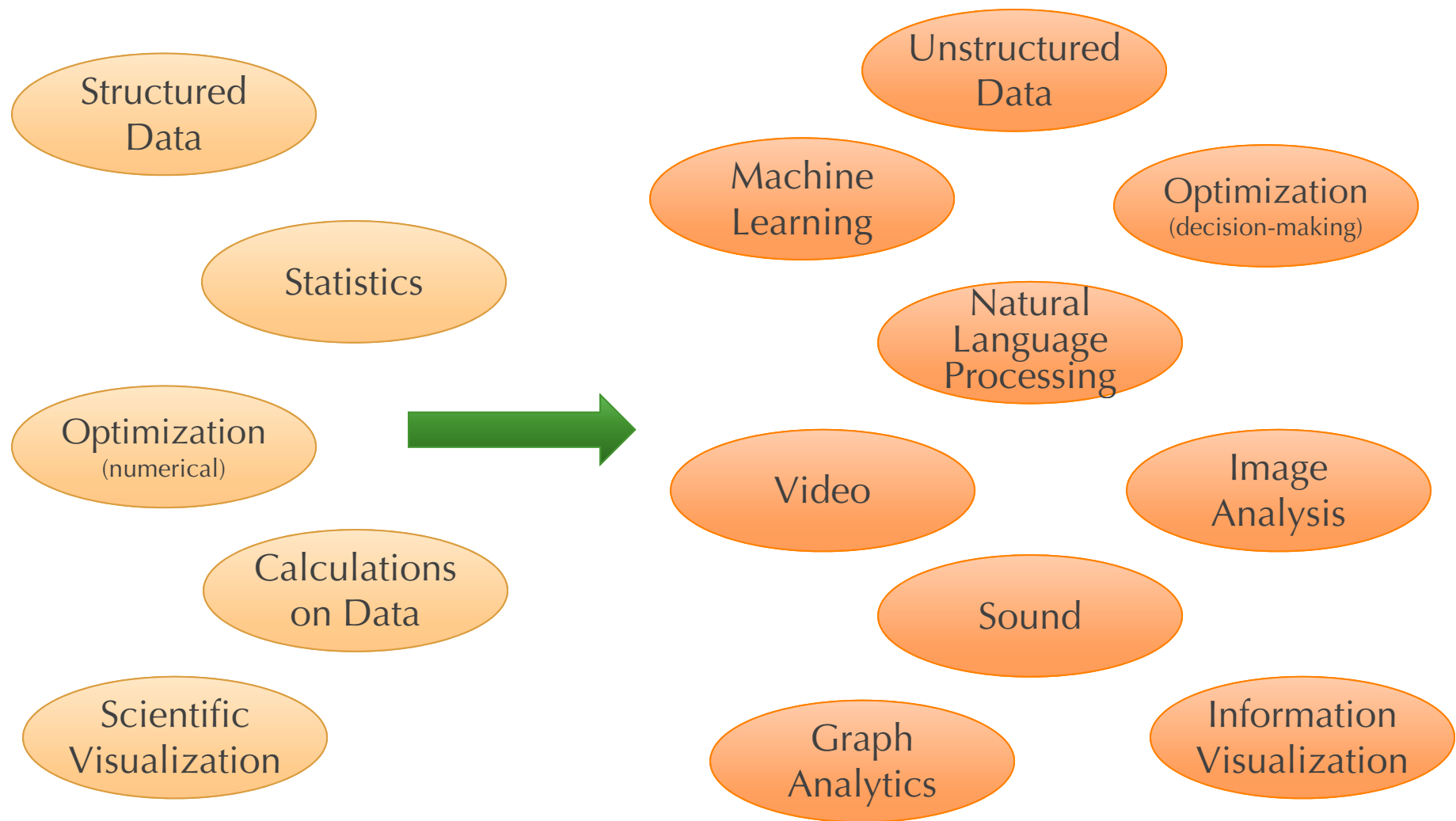


**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE



**PITTSBURGH  
SUPERCOMPUTING  
CENTER**

# Algorithms and Applications Have Also Changed







Pittsburgh is a city of bridges: from its history in steel to its leadership in computer science and biotechnology, between diverse neighborhoods housing its many universities, and at PSC, from science-inspired national cyberinfrastructure to researchers' breakthroughs.

*Bridges* will be a new XSEDE resource that will integrate advanced memory technologies to empower new communities, bring desktop convenience to HPC, connect to campuses, and intuitively express data-intensive workflows.



**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE





The \$9.65M *Bridges* acquisition is made possible by National Science Foundation (NSF) award #ACI-1445606:

*Bridges: From Communities and Data to Workflows and Insight*



HP is delivering *Bridges*



**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE



# Disclaimer

The following presentation conveys PSC's current plan for the *Bridges* supercomputer. Details are subject to change.

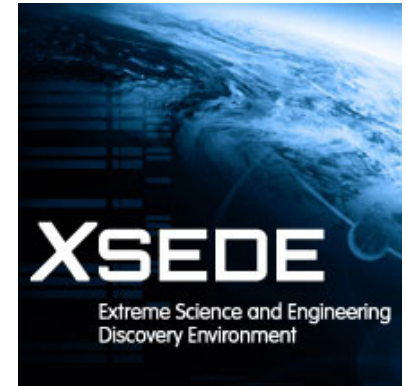


**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE



# An Important Addition to the National Advanced Cyberinfrastructure Ecosystem

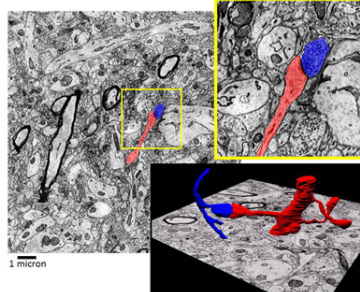
*Bridges* will be a new resource on XSEDE and will interoperate with other XSEDE resources, Advanced Cyberinfrastructure (ACI) projects, campuses, and instruments nationwide.



## Examples:



High-throughput genome sequencers



Reconstructing brain circuits from high-resolution electron microscopy



Social networks and the Internet

**Data Infrastructure Building Blocks (DIBBs)**

- Data Exacell (DXC)
- Other DIBBs projects

**Other ACI projects**




Carnegie Mellon University's Gates Center for Computer Science



Temple University's new Science, Education, and Research Center







# Meeting Research Needs – *Examples*

Scaling research questions beyond the laptop

From individual researchers to teams & collaborations

Cross-domain analyses

Shared data collections & related analysis tools

Workflows – HPC power without the programming

Large-memory applications & in-memory databases

Optimization & parameter sweeps

Powerful collections of application & tools

Modern, widely-used software environments



**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE





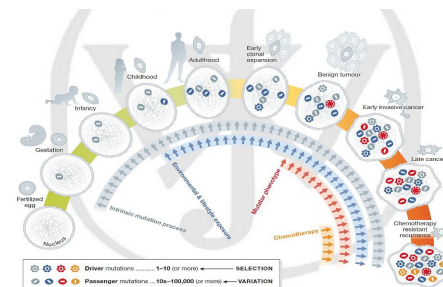
# Potential Applications (*Examples*)

- Finding causal relationships underlying various diseases
- Assembling large genomes and metagenomes
- Analysis of financial markets and policies
- Improving the effectiveness of organ donation networks
- Recognizing events and enabling search for videos
- Understanding how the brain is connected from EM data
- Addressing societal issues from social media data
- Analyzing large bodies of work in the digital humanities
- Agent-based modeling for epidemiology
- Cross-observational analyses in astronomy & other sciences
- Data integration for history, political science & cultural studies

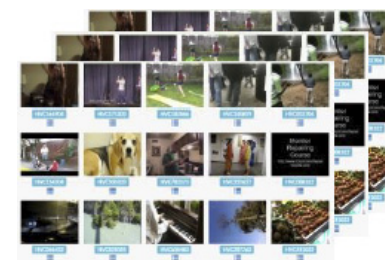


# Objectives and Approach

- Bring HPC to nontraditional users and research communities.
- Allow high-performance computing to be applied effectively to big data.
- Bridge to campuses to ease access and provide burst capability.
- Leveraging PSC's expertise with shared memory, *Bridges* will feature 3 tiers of large, coherent shared-memory nodes.
- *Bridges* will leverage its large memory for interactivity and to seamlessly support applications through virtualization, gateways, familiar and productive programming environments, and data-driven workflows.



EMBO Mol Med (2013) DOI: 10.1002/emmm.201202388:  
*Proliferation of cancer-causing mutations throughout life*

Alex Hauptmann et. al.: *Efficient large-scale content-based multimedia event detection*

# User-Friendly HPC & Data Analytics

- **Interactivity** is the feature most frequently requested by nontraditional HPC communities and for doing data analytics and testing hypotheses.
- **Gateways and tools for gateway building** will provide easy-to-use access to Bridges' HPC and data resources.
- **Database and web server nodes** will provide persistent databases to enable data management, workflows, and distributed applications.
- **High-productivity programming languages & environments** will let users scale familiar applications and workflows.
- **Virtualization** will allow users to bring their particular environments and provide interoperability with clouds.



# Interactivity

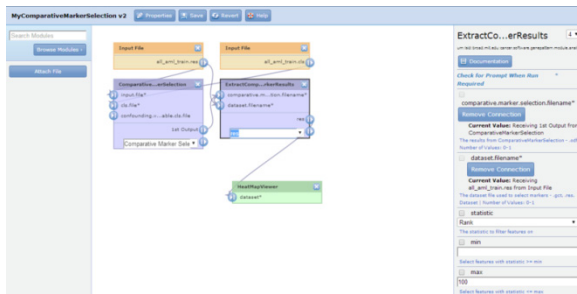
- *Interactivity is the feature most frequently requested by nontraditional HPC communities.*
- Interactivity provides immediate feedback for doing exploratory data analytics and testing hypotheses.
- *Bridges* will offer interactivity through a combination of virtualization for lighter-weight applications and dedicated nodes for more demanding ones.



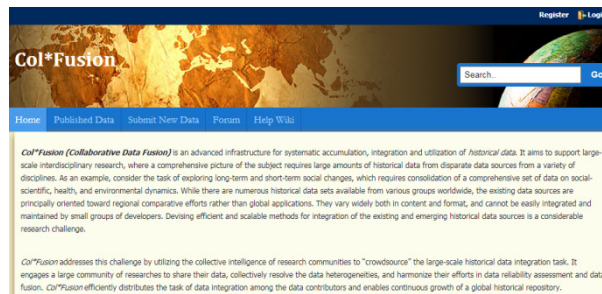


# Gateways and Tools for Building Them

Gateways will provide easy-to-use access to *Bridges'* HPC and data resources, allowing users to launch jobs, orchestrate complex workflows and manage data from their web browsers.



Interactive pipeline creation in GenePattern (Broad Institute)



Col\*Fusion portal for the systematic accumulation, integration, and utilization of historical data, from <http://colfusion.exp.sis.pitt.edu/colfusion/>



Download sites for MEGA-6 (Molecular Evolutionary Genetic Analysis), from [www.megasoftware.net](http://www.megasoftware.net)

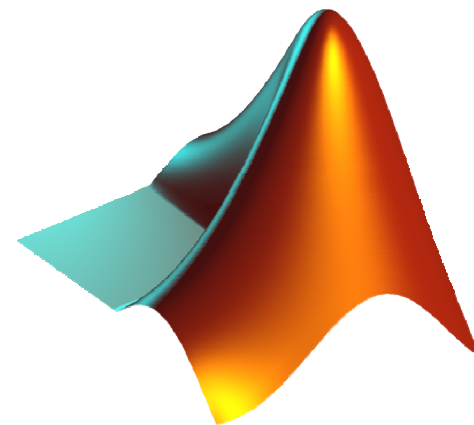
# Virtualization and Containers

Users will be able to import their entire computational environments into *Bridges*, conferring ease of use, reproducibility, and interoperability with cloud services.



# High-Productivity Programming

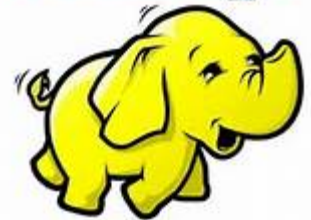
*Bridges* will feature high-productivity programming languages and tools.



# Hadoop Ecosystem

- *Bridges* will provide acceleration for Hadoop applications running on its 128GB nodes.
- Large memory will be great for Spark.

**hadoop**



**Cassandra**



**APACHE  
HBASE**

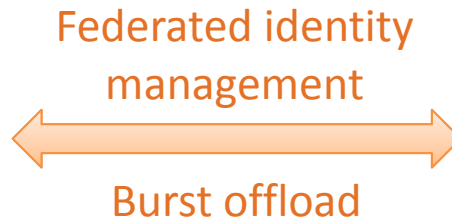


**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE





# Campus Bridging



[http://www.temple.edu/medicine/research/RESEARCH\\_TUSM/](http://www.temple.edu/medicine/research/RESEARCH_TUSM/)

- Through a pilot project with Temple University, the *Bridges* project will explore new ways to transition data and computing seamlessly between campus and XSEDE resources.
- **Federated identity management** will allow users to use their local credentials for single sign-on to remote resources, facilitating data transfers between *Bridges* and Temple's local storage systems.
- **Burst offload** will enable cloud-like offloading of jobs from Temple to *Bridges* and vice versa during periods of unusually heavy load.



**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE

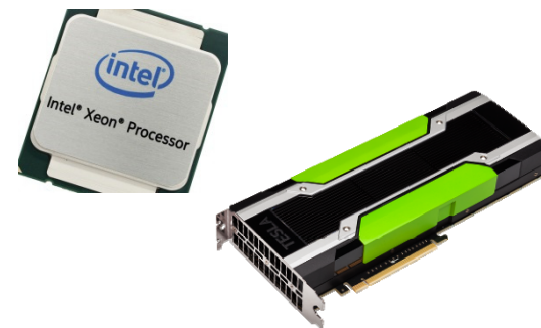


# High-Performance, Data-Intensive Computing

- 3 tiers of large, coherent shared memory nodes

Memory per node	Number of nodes	Example applications
12 TB	Several	Genomics, machine learning, graph analytics, other extreme-memory applications
3 TB	Tens	Virtualization and interactivity including large-scale visualization and analytics; mid-spectrum memory-intensive jobs
128 GB	Hundreds	Execution of most components of workflows, interactivity, Hadoop, and capacity computing

- The latest Intel® Xeon® CPUs
- NVIDIA® Tesla® dual-GPU accelerators



# Database and Web Server Nodes

- Dedicated database nodes will power persistent relational and NoSQL databases
  - Support data management and data-driven workflows
  - SSDs for high IOPs; RAIDed HDDs for high capacity



- Dedicated web server nodes
  - Enable distributed, service-oriented architectures
  - High-bandwidth connections to XSEDE and the Internet

# Data Management

- *Pylon*: A large, central, high-performance filesystem
  - Visible to all nodes
  - Large datasets, community repositories (~10 PB usable)
- Distributed (node-local) storage
  - Enhance application portability
  - Improve overall system performance
  - Improve performance consistency to the shared filesystem
- Acceleration for Hadoop-based applications

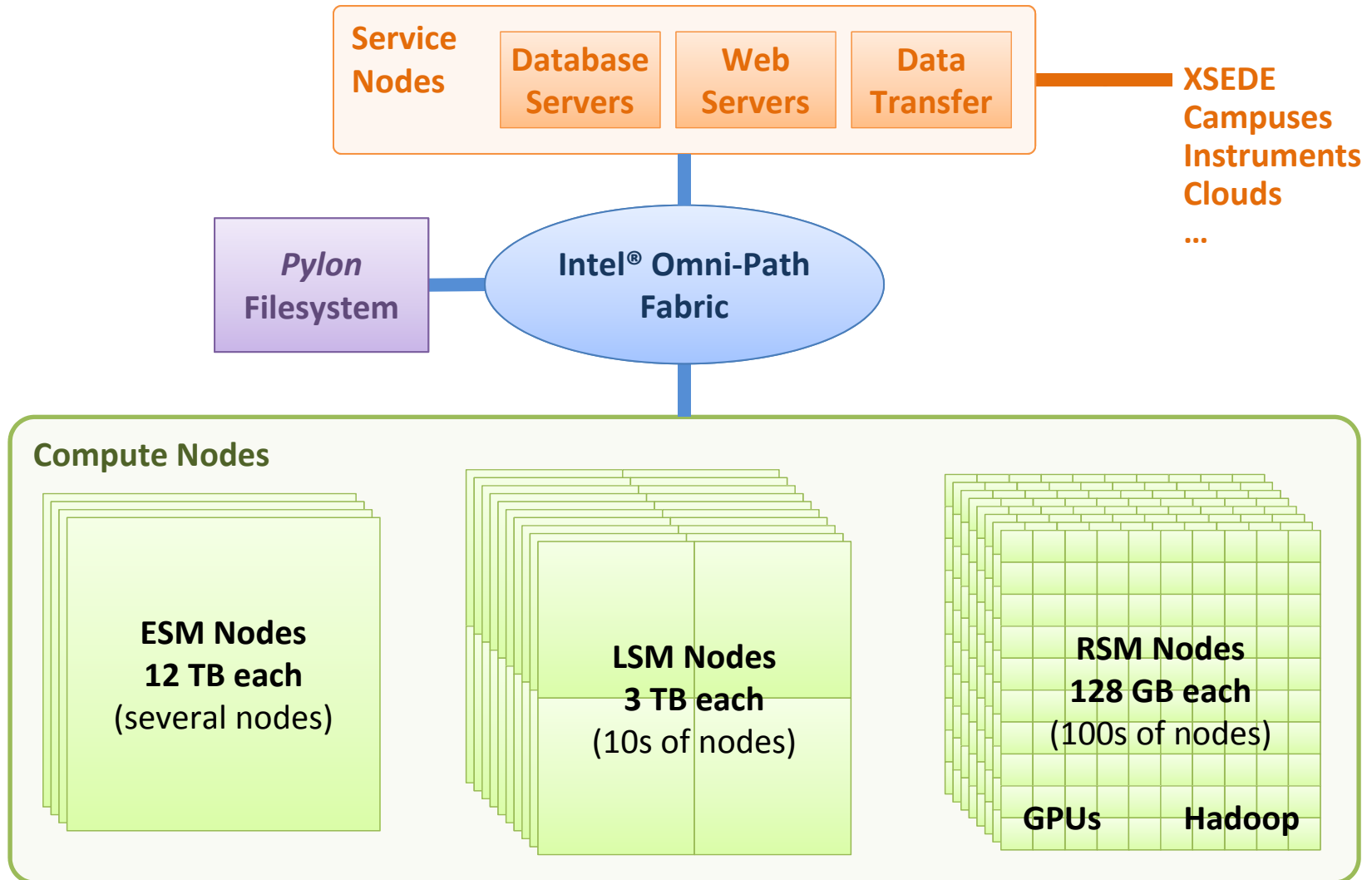


# Intel® Omni-Path Architecture Fabric

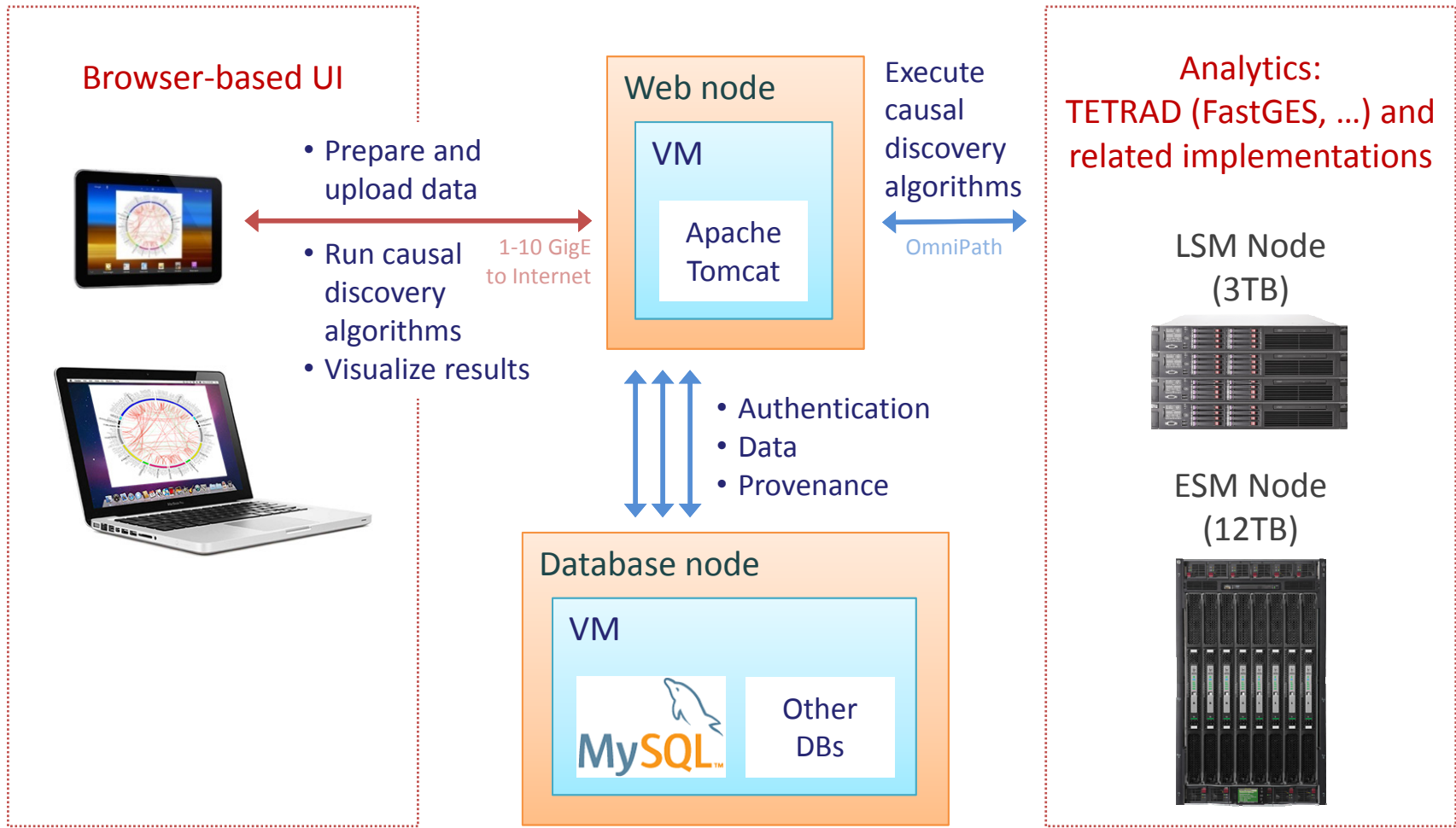
- Omni-Path will connect all nodes and the shared filesystem, providing *Bridges* and its users with:
  - The highest-bandwidth internal network
  - Valuable optimizations for MPI and other communications
  - Early access to this new, important, forward-looking technology



# High-Level Architecture



# Example: Causal Discovery



# *Bridges Target Schedule*

- **Acquisition**
  - Begins December 2014
  - Construction to begin October 2015
    - Will allow for including important new technologies
  - Early user period in late 2015
- **Production**
  - January 2016





# For Additional Information

Project website: [www.psc.edu/bridges](http://www.psc.edu/bridges)

*Bridges* PI: Nick Nystrom  
[nystrom@psc.edu](mailto:nystrom@psc.edu)



Co-PIs: Michael J. Levine  
Ralph Roskies  
J Ray Scott

Project Manager: Janet Brown



**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE

