BRIDGES
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE
A Big Big Data Platform

John Urbanic, Parallel Computing Scientist

© 2017 Pittsburgh Supercomputing Center
The Shift to Big Data

**New Emphases**

- **Social networks and the Internet**
- **Video**
- **Collections**
- **Legacy documents**
- **Environmental sensors: Water temperature profiles from tagged hooded seals**

**Examples:**

- Pan-STARRS telescope
  - [http://pan-starrs.ifa.hawaii.edu/public/](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)
- 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
- Environmenta sensors: Water temperature profiles from tagged hooded seals

**Other resources:**

- [Library of Congress stacks](https://www.flickr.com/photos/danlem2001/6922113091/)
- [Get Involved/Bioblitz Insects](http://www.horniman.ac.uk/get_involved/blog/bioblitz-insects-reviewed)
Challenges and Software are Co-Evolving

- Structured Data
- Statistics
- Optimization (numerical)
- Calculations on Data
- Scientific Visualization
- Unstructured Data
- Machine Learning
- Optimization (decision-making)
- Natural Language Processing
- Video
- Sound
- Image Analysis
- Graph Analytics
- Information Visualization
Motivating Use Cases

Data-intensive applications & workflows
Gateways – the power of HPC without the programming
Shared data collections & analyses: cross-domain analytics
Deep learning
Graph analytics, machine learning, genome sequence assembly, and other large-memory applications
Scaling beyond the laptop
Scaling research to teams and collaborations
In-memory databases
Optimization & parameter sweeps
Distributed & service-oriented architectures
Data assimilation from large instruments & Internet
Leveraging an extensive software collection

Research areas that haven’t used HPC
Nontraditional HPC approaches to fields such as the physical sciences
Coupling applications in novel ways
Leveraging large memory and high bandwidth
Objectives

• Bring HPC to nontraditional users and research communities.

• Allow high-performance computing to be applied effectively to big data.

• Bridge to campuses to streamline access and provide cloud-like burst capability.

• Leveraging PSC’s expertise with shared memory, Bridges has 3 tiers of large, coherent shared-memory nodes: 12TB, 3TB, and 128GB.

• Bridges implements a uniquely flexible environment featuring interactivity, gateways, databases, distributed (web) services, high-productivity programming languages and frameworks, and virtualization, and campus bridging.
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* offers interactivity through a combination of virtualization for lighter-weight applications and dedicated nodes for more demanding ones.
Gateways and Tools for Building Them

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

- Extensive leveraging of databases and polystore systems
- Great attention to HCI is needed to get these right

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
Virtualization and Containers

- Virtual Machines (VMs) enable flexibility, security, customization, reproducibility, ease of use, and interoperability with other services.
- User demand is for custom database and web server installations to develop data-intensive, distributed applications and containers for custom software stacks and portability.
- Bridges leverages OpenStack to provision resources, between interactive, batch, Hadoop, and VM uses.
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.

High-Productivity Programming

Supporting languages that communities already use is vital for them to apply HPC to their research questions.
Bridges’ large memory is great for Spark!

*Bridges* enables workflows that integrate Spark/Hadoop, HPC, and/or shared-memory components.
Deep Learning Frameworks on *Bridges*

- Caffe
- PyTorch
- Keras
- TensorFlow
- Theano
- NVIDIA DIGITS
13 RSM nodes, each with 2 NVIDIA Tesla K80 GPUs
32 RSM nodes, each with 2 NVIDIA Tesla P100 GPUs
800 HPE Apollo 2000 (128 GB) compute nodes
20 “leaf” Intel® OPA edge switches
6 “core” Intel® OPA edge switches: fully interconnected, 2 links per switch
4 MDS nodes
2 front-end nodes
2 boot nodes
8 management nodes
Intel® OPA cables

Purpose-built Intel® Omni-Path Architecture topology for data-intensive HPC

Bridges Virtual Tour:
https://www.psc.edu/bvt
## Node Types

<table>
<thead>
<tr>
<th>Type</th>
<th>RAM</th>
<th>Phase</th>
<th>n</th>
<th>CPU / GPU / other</th>
<th>Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESM</td>
<td>12TB(^b)</td>
<td>1</td>
<td>2</td>
<td>16 × Intel Xeon E7-8880 v3 (18c, 2.3/3.1 GHz, 45MB LLC)</td>
<td>HPE Integrity Superdome X</td>
</tr>
<tr>
<td></td>
<td>12TB(^c)</td>
<td>2</td>
<td>2</td>
<td>16 × Intel Xeon E7-8880 v4 (22c, 2.2/3.3 GHz, 55MB LLC)</td>
<td>HPE ProLiant DL580</td>
</tr>
<tr>
<td>LSM</td>
<td>3TB(^b)</td>
<td>1</td>
<td>8</td>
<td>4 × Intel Xeon E7-8860 v3 (16c, 2.2/3.2 GHz, 40 MB LLC)</td>
<td>HPE ProLiant DL580</td>
</tr>
<tr>
<td></td>
<td>3TB(^c)</td>
<td>2</td>
<td>34</td>
<td>4 × Intel Xeon E7-8870 v4 (20c, 2.1/3.0 GHz, 50 MB LLC)</td>
<td>HPE ProLiant DL580</td>
</tr>
<tr>
<td>RSM</td>
<td>128GB(^b)</td>
<td>1</td>
<td>752</td>
<td>2 × Intel Xeon E5-2695 v3 (14c, 2.3/3.3 GHz, 35MB LLC)</td>
<td>HPE ProLiant DL580</td>
</tr>
<tr>
<td>RSM-GPU</td>
<td>128GB(^b)</td>
<td>1</td>
<td>16</td>
<td>2 × Intel Xeon E5-2695 v3 + 2 × NVIDIA Tesla K80</td>
<td>HPE Apollo 2000</td>
</tr>
<tr>
<td></td>
<td>128GB(^c)</td>
<td>2</td>
<td>32</td>
<td>2 × Intel Xeon E5-2683 v4 (16c, 2.1/3.0 GHz, 40 MB LLC) + 2 × NVIDIA Tesla P100</td>
<td>HPE Apollo 2000</td>
</tr>
<tr>
<td>DB-s</td>
<td>128GB(^b)</td>
<td>1</td>
<td>6</td>
<td>2 × Intel Xeon E5-2695 v3 + SSD</td>
<td>HPE ProLiant DL360</td>
</tr>
<tr>
<td>DB-h</td>
<td>128GB(^b)</td>
<td>1</td>
<td>6</td>
<td>2 × Intel Xeon E5-2695 v3 + HDDs</td>
<td>HPE ProLiant DL360</td>
</tr>
<tr>
<td>Web</td>
<td>128GB(^b)</td>
<td>1</td>
<td>6</td>
<td>2 × Intel Xeon E5-2695 v3</td>
<td>HPE ProLiant DL360</td>
</tr>
<tr>
<td>Other(^a)</td>
<td>128GB(^b)</td>
<td>1</td>
<td>16</td>
<td>2 × Intel Xeon E5-2695 v3</td>
<td>HPE ProLiant DL360, DL380</td>
</tr>
<tr>
<td>Gateway</td>
<td>64GB(^b)</td>
<td>1</td>
<td>4</td>
<td>2 × Intel Xeon E5-2683 v3 (14c, 2.0/3.0 GHz, 35MB LLC)</td>
<td>HPE ProLiant DL380</td>
</tr>
<tr>
<td></td>
<td>64GB(^c)</td>
<td>2</td>
<td>4</td>
<td>2 × Intel Xeon E5-2683 v3</td>
<td>HPE ProLiant DL380</td>
</tr>
<tr>
<td>Storage</td>
<td>128GB(^b)</td>
<td>1</td>
<td>5</td>
<td>2 × Intel Xeon E5-2680 v3 (12c, 2.5/3.3 GHz, 30 MB LLC)</td>
<td>Supermicro X10DRi</td>
</tr>
<tr>
<td></td>
<td>256GB(^c)</td>
<td>2</td>
<td>15</td>
<td>2 × Intel Xeon E5-2680 v4 (14c, 2.4/3.3 GHz, 35 MB LLC)</td>
<td>Supermicro X10DRi</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>281.75TB</strong></td>
<td></td>
<td><strong>908</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Other nodes = front end (2) + management/log (8) + boot (4) + MDS (4)
\(^b\) DDR4-2133
\(^c\) DDR4-2400
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

  ![Database Logos](examples)

• Dedicated web server nodes
  – Enable distributed, service-oriented architectures
  – High-bandwidth connections to XSEDE and the Internet
Bridges’ GPUs are accelerating both deep learning and simulation codes

**Phase 1:** 16 nodes, each with:
- $2 \times$ NVIDIA Tesla K80 GPUs (32 total)
- $2 \times$ Intel Xeon E5-2695 v3 (14c, 2.3/3.3 GHz)
- 128GB DDR4-2133 RAM

**Phase 2:** +32 nodes, each with:
- $2 \times$ NVIDIA Tesla P100 GPUs (64 total)
- $2 \times$ Intel Xeon E5-2683 v4 (16c, 2.1/3.0 GHz)
- 128GB DDR4-2400 RAM

**Kepler architecture**
- 2496 CUDA cores (128/SM)
- 7.08B transistors on 561mm$^2$ die (28nm)
- 2x24 GB GDDR5; 2x240.6 GB/s
- 562 MHz base – 876 MHz boost
- 2.91 Tf/s (64b), 8.73 Tf/s (32b)

**Pascal architecture**
- 3584 CUDA cores (64/SM)
- 15.3B transistors on 610mm$^2$ die (16nm)
- 16GB CoWoS® HBM2 at 720 GB/s w/ ECC
- 1126 MHz base – 1303 MHz boost
- 4.7 Tf/s (64b), 9.3 Tf/s (32b), 18.7 Tf/s (16b)
- Page migration engine improves unified memory
- 64 P100 GPUs $\rightarrow$ 600 Tf/s (32b)
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 (OPA)**

- *Bridges* is the first production deployment of Omni-Path.
- Omni-Path connects all nodes and the shared filesystem, providing *Bridges* and its users with:
  - 100 Gbps line speed per port;
  - 25 GB/s bidirectional bandwidth per port
  - Observed <0.93μs latency, 12.36 GB/s/dir
  - 160M MPI messages per second
  - 48-port edge switch reduces interconnect complexity and cost
  - HPC performance, reliability, and QoS
  - OFA-compliant applications supported without modification
  - Early access to this new, important, forward-looking technology
- *Bridges* deploys OPA in a two-tier island topology developed by PSC for cost-effective, data-intensive HPC.
Example: Causal Discovery Portal
Center for Causal Discovery, an NIH Big Data to Knowledge Center of Excellence

Web node
- VM
- Apache
- Tomcat
- Messaging

Execute causal discovery algorithms
Omni-Path

Analytics:
- FGS and other algorithms, building on TETRAD

LSM Node
(3TB)

ESM Node
(12TB)

Database node
- VM
- MySQL
- Other DBs

Pylon filesystem
- TCGA
- fMRI
- ...

Internet

Browser-based UI
- Prepare and upload data
- Run causal discovery algorithms
- Visualize results

• Authentication
• Data
• Provenance

• Prepare and upload data
• Run causal discovery algorithms
• Visualize results

TCGA
fMRI
...

Memory-resident datasets

Pylon filesystem

Carnegie Mellon
Yale
PSC
BRIDGES

A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE
Some of the Deep Learning Projects Using Bridges

Deep Learning of Game Strategies for RoboCup, Manuela Veloso (CMU)

Automatic Building of Speech Recognizers for Non-Experts, Florian Metze (CMU)

Automatic Evaluation of Scientific Writing, Diane Litman (U. of Pittsburgh)

Image Classification Applied in Economic Studies, Param Singh (CMU)

Exploring Stability, Cost, and Performance in Adversarial Deep Learning, Matt Fredrikson (CMU)

Enabling Robust Image Understanding Using Deep Learning, Adriana Kovashka (U. of Pittsburgh)

Optimal Data Representation for Deep Learning for Computational Chemistry, Garrett Goh (Pacific Northwest National Laboratory)

Petuum, a Distributed System for High-Performance Machine Learning, Eric Xing (CMU)

Deep Learning the Gene Regulatory Code, Shaun Mahony (Penn State)

Developing Large-Scale Distributed Deep Learning Methods for Protein Bioinformatics, Junbo Xu (Toyota Technological Institute at Chicago)

Education Allocation for the Course Unstructured Data & Big Data: Acquisition to Analysis, Dokyun Lee (CMU)

Deciphering Cellular Signaling System by Deep Mining a Comprehensive Genomic Compendium, Xinghua Lu (U. of Pittsburgh)

Quantifying California Current Plankton Using Machine Learning, Mark Ohman (Scripps Institution of Oceanography)

Automatic Pain Assessment, Michael Reale (SUNY Polytechnic Institute)

Learning to Parse Images and Videos, Deva Ramanan (CMU)

Deep Recurrent Models for Fine-Grained Recognition, Michael Lam (Oregon State)
Some of the Deep Learning Projects Using Bridges

Live Song Identification Using Semantic Features, Timothy Tsai (Harvey Mudd College)

Inverse Graphics Engines for Visual Inference, Ioannis Gkioulekas (CMU)

Development of a Hybrid Computational Approach for Macroscale Simulation of Exciton Diffusion in Polymer Thin Films, Based on Combined Machine Learning, Quantum-Classical Simulations and Master Equation Techniques, Peter Rossky (Rice U.)

Summarizing and Learning Latent Structure in Video, Jeff Boleng (CMU)

Machine Learning for Medical Image Analysis, Mai Nguyen (UCSD)

Deep Learning for Drug-Protein Interaction Prediction, Gil Alterovitz (Harvard Medical School/Boston Children’s Hospital)

CMU course Deep reinforcement Learning, Aikaterini Fragkiadaki (CMU)

Course 11-364: Introduction to Deep Learning, James Baker (CMU)

Deep Recurrent Models for Fine-Grained Recognition, Michael Lam (Oregon State University)

ARIEL: Analysis of Rare Incident-Event Languages, Ravi Starzl (CMU)

Aarti Singh, Deep Purple: Deep Purposeful Learning of Complex Dynamic Systems (CMU)

Deep Learning for Genomic Sequence Associated Study, Zhi Wei (New Jersey Institute of Technology)

Learning to Parse Images and Videos, Deva Ramanan (CMU)

Preparing Grounds to Launch All-US Students Kaggle Competition on Drug Prediction, Gil Alterovitz (Harvard Medical School/Boston Children’s Hospital)

Modeling Enzymatic Carbohydrate Decomposition, Heather Mayes (U. of Michigan)
Gaining Access to *Bridges*

*Bridges* is allocated through XSEDE: [https://www.xsede.org/allocations](https://www.xsede.org/allocations)

- **Starter Allocation**
  - Can request *anytime*
  - Up to 50,000 core-hours on RSM and GPU (128GB) nodes and/or 10,000 TB-hours on LSM (3TB) and ESM (12TB) nodes
  - Can request XSEDE ECSS (Extended Collaborative Support Service)

- **Research Allocation (XRAC)**
  - Appropriate for larger requests; can request ECSS
  - Can be up to millions to tens of millions of SUs
  - Quarterly submission windows: *March 15–April 15, June 15–July 15*, etc.

- **Coursework Allocations**
  - To support use of *Bridges* for relevant courses

- **Community Allocations**
  - Primarily to support gateways

Up to 10% of *Bridges’* SUs are available on a discretionary basis to industrial affiliates, Pennsylvania-based researchers, and others to foster discovery and innovation and broaden participation in data-intensive computing.
For Additional Information

Project website: www.psc.edu/bridges

Questions: bridges@psc.edu

*Bridges* PI: Nick Nystrom
   nystrom@psc.edu

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

Project Manager: Robin Scibek