



XSEDE

Extreme Science and Engineering  
Discovery Environment



# BRIDGES

A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE

A Big Big Data Platform

John Urbanic, Parallel Computing Scientist

# The Shift to Big Data



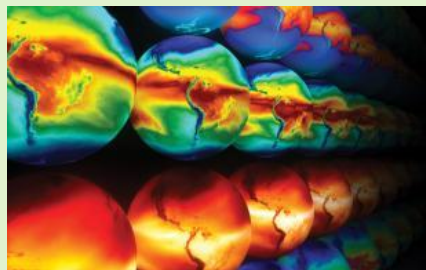
Pan-STARRS telescope

<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)

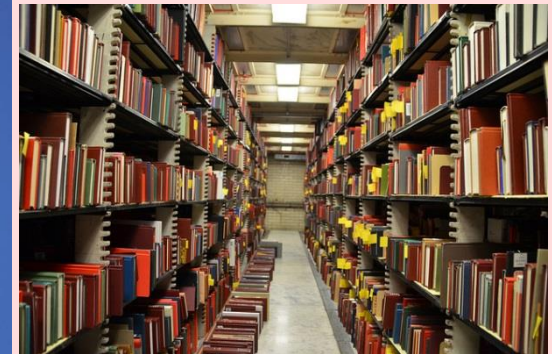


Social networks and the Internet



Video

Wikipedia Commons



Library of Congress stacks

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



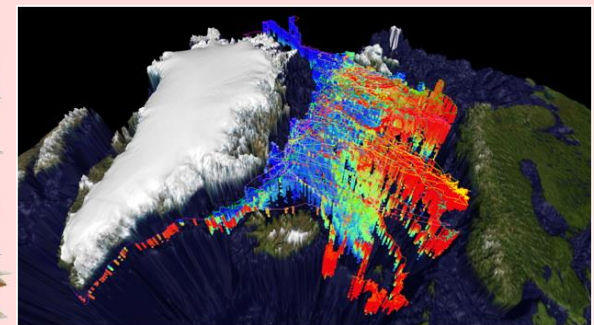
Collections

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)

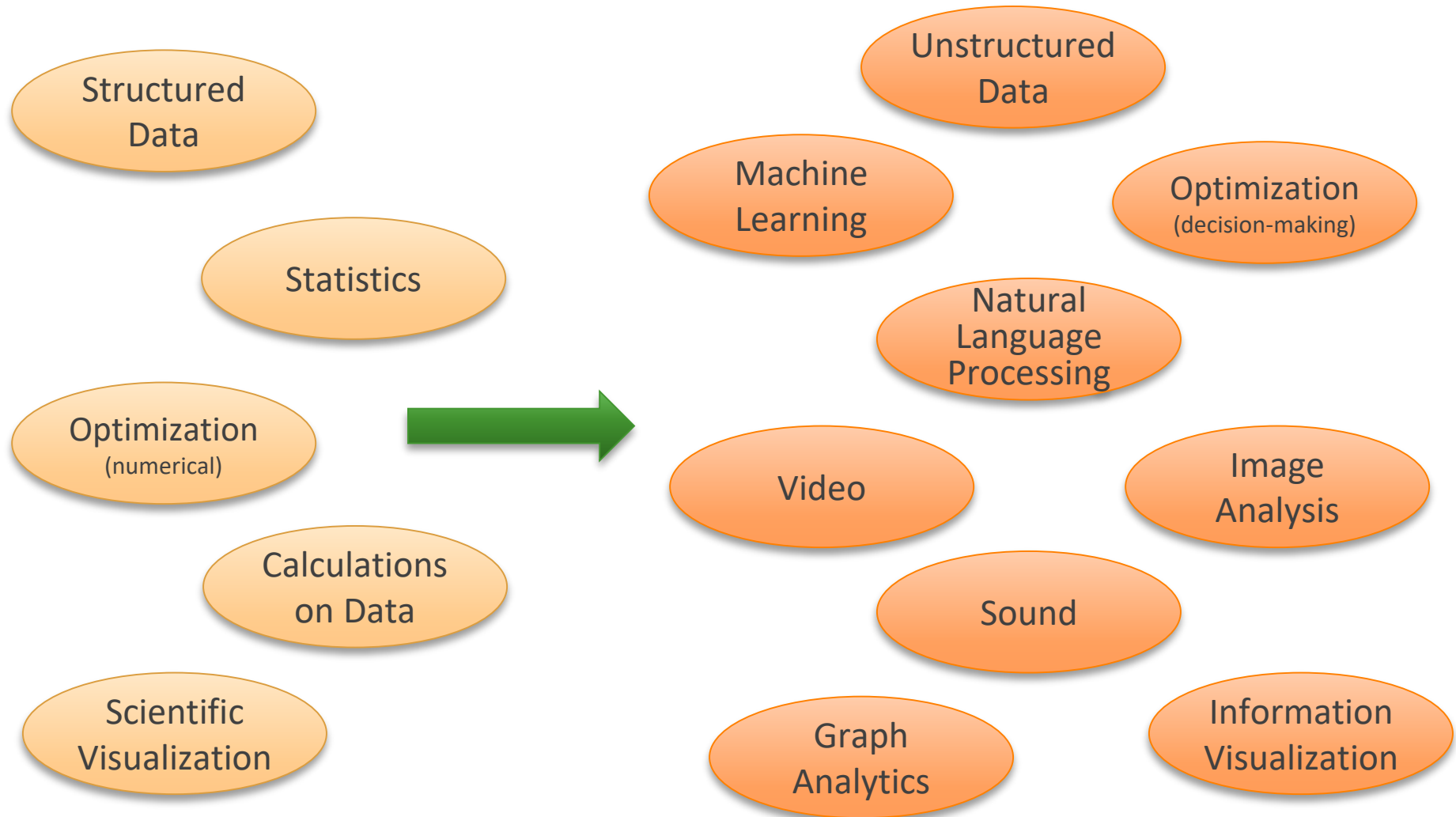


**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE



**PITTSBURGH  
SUPERCOMPUTING  
CENTER**

# Challenges and Software are Co-Evolving



# 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



# 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.



**BRIDGES**  
A PITTSBURGH SUPERCOMPUTING CENTER RESOURCE

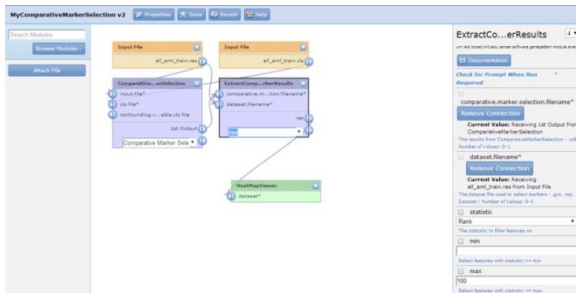




# Gateways and Tools for Building Them

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](http://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.



# High-Productivity Programming

Supporting languages that communities already use is vital for them to apply HPC to their research questions.

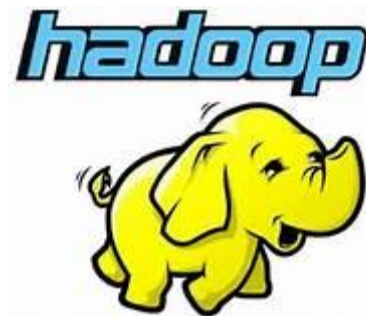




# Spark, Hadoop & Related Approaches

Bridges' large memory is great for Spark!

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



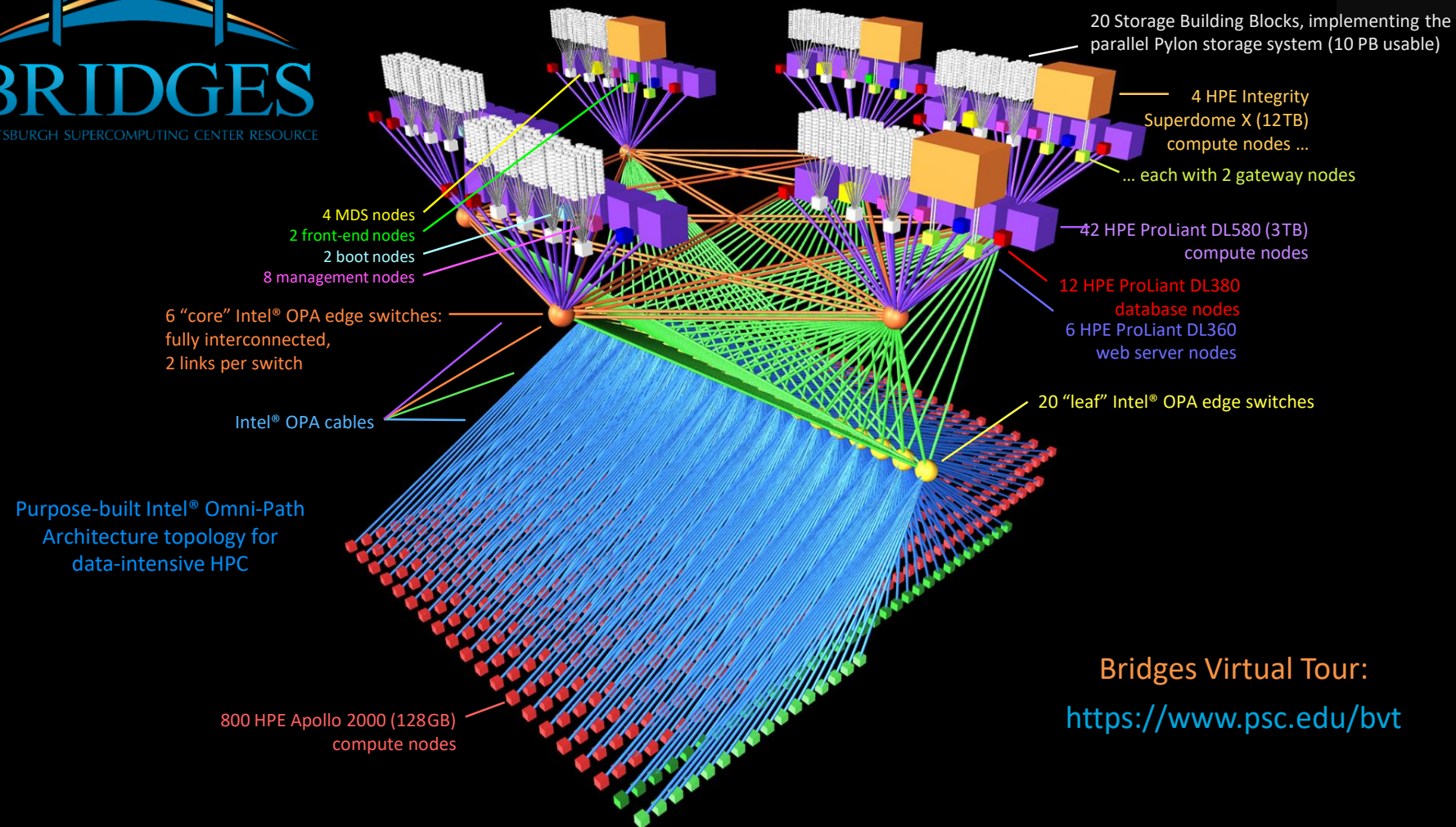
# Deep Learning Frameworks on *Bridges*

Caffe

PYTORCH



theano



Bridges Virtual Tour:  
<https://www.psc.edu/bvt>

This tool depicts the current plan for PSC's Bridges supercomputing system. Details are subject to change.

# Regular GPU Nodes

*Bridges' GPUs are accelerating both deep learning and simulation codes*

Phase 1: 16 nodes, each with:

- **2 × NVIDIA Tesla K80 GPUs (32 total)**
- 2 × Intel Xeon E5-2695 v3 (14c, 2.3/3.3 GHz)
- 128GB DDR4-2133 RAM

Phase 2: +32 nodes, each with:

- **2 × NVIDIA Tesla P100 GPUs (64 total)**
- 2 × 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<sup>2</sup> die (28nm)  
2×24 GB GDDR5; 2×240.6 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<sup>2</sup> die (16nm)  
16GB CoWoS<sup>®</sup> 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 → 600 Tf/s (32b)



# Bridges Hardware

Bridges-DL

Type	RAM	#	CPU / GPU / SSD	Server
ESM	12 TB <sup>b</sup>	2	16 × Intel Xeon E7-8880 v3 (18c, 2.3/3.1 GHz, 45MB LLC)	HPE Integrity Superdome X
	12 TB <sup>c</sup>	2	16 × Intel Xeon E7-8880 v4 (22c, 2.2/3.3 GHz, 55MB LLC)	
LSM	3 TB <sup>b</sup>	8	4 × Intel Xeon E7-8860 v3 (16c, 2.2/3.2 GHz, 40 MB LLC)	HPE ProLiant DL580
	3 TB <sup>c</sup>	34	4 × Intel Xeon E7-8870 v4 (20c, 2.1/3.0 GHz, 50 MB LLC)	
RSM	128 GB <sup>b</sup>	752	2 × Intel Xeon E5-2695 v3 (14c, 2.3/3.3 GHz, 35MB LLC)	HPE Apollo 2000
RSM-GPU	128 GB <sup>b</sup>	16	2 × Intel Xeon E5-2695 v3 + 2 × NVIDIA Tesla K80	
	128 GB <sup>c</sup>	32	2 × Intel Xeon E5-2683 v4 (16c, 2.1/3.0 GHz, 40MB LLC) + 2 × NVIDIA Tesla P100	
GPU-AI16	1.5 TB <sup>d</sup>	1	16 × NVIDIA V100 32GB SXM2 + 2 × Intel Xeon Platinum 8168 + 8 × 3.84 TB NVMe SSDs	NVIDIA DGX-2 delivered by HPE
GPU-A8	192 GB <sup>d</sup>	9	2 × Intel Xeon Gold 6148 + 2 × 3.84 TB NVMe SSDs	HPE Apollo 6500 Gen10
DB-s	128 GB <sup>b</sup>	6	2 × Intel Xeon E5-2695 v3 + SSD	HPE ProLiant DL360
DB-h	128 GB <sup>b</sup>	6	2 × Intel Xeon E5-2695 v3 + HDDs	HPE ProLiant DL380
Web	128 GB <sup>b</sup>	6	2 × Intel Xeon E5-2695 v3	HPE ProLiant DL360
Other <sup>a</sup>	128 GB <sup>b</sup>	16	2 × Intel Xeon E5-2695 v3	HPE ProLiant DL360, DL380
Gateway	64 GB <sup>b</sup>	4	2 × Intel Xeon E5-2683 v3 (14c, 2.0/3.0 GHz, 35MB LLC)	HPE ProLiant DL380
	64 GB <sup>c</sup>	4	2 × Intel Xeon E5-2683 v3	
	96 GB <sup>d</sup>	2	2 × Intel Xeon	
Storage	128 GB <sup>b</sup>	5	2 × Intel Xeon E5-2680 v3 (12c, 2.5/3.3 GHz, 30 MB LLC)	Supermicro X10DRi
	256 GB <sup>c</sup>	15	2 × Intel Xeon E5-2680 v4 (14c, 2.4/3.3 GHz, 35 MB LLC)	
Total	286.5 TB	920		

- a. Other nodes = front end (2) + management/log (8) + boot (4) + MDS (4)  
b. DDR4-2133  
c. DDR4-2400  
d. DDR4-2666

# The Heart of Bridges-DL: NVIDIA Volta



NVIDIA Tesla V100 SXM2 Module  
with Volta GV100 GPU

*New Streaming Multiprocessor (SM) architecture*, introducing Tensor Cores, independent thread scheduling, combined L1 data cache and shared memory unit, and 50% higher energy efficiency over Pascal.

*Tensor Cores accelerate deep learning training and inference*, providing up to 12× and 6× higher peak flops respectively over the P100 GPUs currently available in XSEDE.

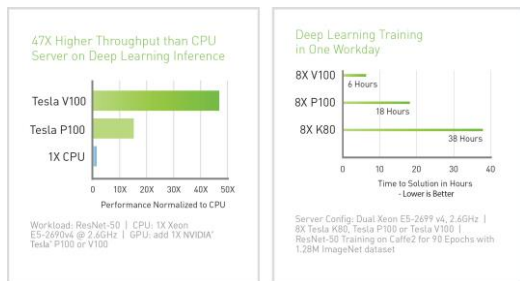
*NVLink 2.0 delivering 300 GB/s total bandwidth per GV100*, nearly 2× higher than P100.

*HBM2 bandwidth and capacity increases: 900 GB/s and up to 32GB.*

*Enhanced Unified Memory and Address Translation Services* improve accuracy of memory page migration by providing new access counters.

*Cooperative Groups and New Cooperative Launch APIs* expand the programming model to allow organizing groups of communicating threads.

*Volta-Optimized Software* includes new versions of frameworks and libraries optimized to take advantage of the Volta architecture: TensorFlow, Caffe2, MXNet, CNTK, cuDNN, cuBLAS, TensorRT, etc.



Training ResNet-50 with ImageNet:

V100 : 1075 images/s<sup>a</sup>

P100 : 219 images/s<sup>b</sup>

K80 : 52 images/s<sup>b</sup>

a. <https://devblogs.nvidia.com/tensor-core-ai-performance-milestones/>

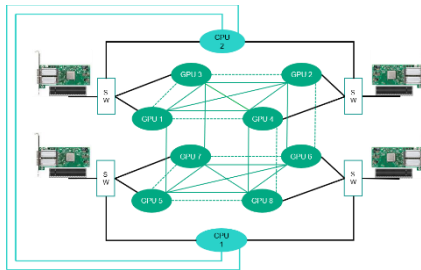
b. <https://www.tensorflow.org/performance/benchmarks>



# Balancing AI Capability & Capacity: HPE Apollo 6500



HPE Apollo 6500 Gen10 Server



HPE Apollo 6500 Gen10  
hybrid cube-mesh topology

*Bridges-DL adds 9 HPE Apollo 6500 Gen10 servers*

*Each HPE Apollo 6500 couples 8 NVIDIA Tesla V100 SXM2 GPUs*

– 40,960 CUDA cores and 5,120 tensor cores

*Performance:* 1 Pf/s mixed-precision tensor, 125 Tf/s 32b, 64 Tf/s 64b

*Memory:* 128 GB HBM2, 7.2 TB/s aggregate memory bandwidth

*2 × Intel Xeon Gold 6148 CPUs and 192 GB of DDR4-2666 RAM*

– 20c, 2.4–3.7 GHz, 27.5 MB L3, 3 UPI links

*2 × 4 TB NVMe SSDs* for user and system data

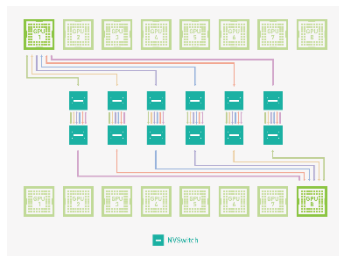
*1 × Intel Omni-Path host channel adapter*

*Hybrid cube-mesh topology* connecting the 8 V100 GPUs and 2 Xeon CPUs, using NVLink 2.0 between the GPUs and PCIe3 to the CPUs

# Maximum DL Capability: NVIDIA DGX-2



NVIDIA DGX-2



NVIDIA DGX-2 with NVSwitch internal topology

*Couples 16 NVIDIA Tesla V100 SXM2 GPUs*

- 81,920 CUDA cores and 10,240 tensor cores

*Performance:* 2 Pf/s mixed-precision tensor, 251 Tf/s 32b, 125 Tf/s 64b

*Memory:* 512 GB HBM2, 14.4 TB/s aggregate memory bandwidth

*2 × Intel Xeon Platinum 8168 CPUs* and *1.5 TB of DDR4-2666 RAM*

- 24c, 2.7–3.7 GHz, 33 MB L3, 3 UPI links

*2 × 960 GB NVMe SSDs* host the Ubuntu Linux OS

*8 × 3.84 TB NVMe SSDs* (aggregate ~30 TB) for user data

*8 × Mellanox ConnectX adapters* for EDR InfiniBand & 100 Gb/s Ethernet

The *NVSwitch* tightly couples the 16 V100 GPUs for capability & scaling

- Each of the 12 NVSwitch chips is an 18×18-port, fully-connected crossbar
- 50 GB/s/port and 900 GB/s/chip bidirectional bandwidths
- 2.4TB/s system bisection bandwidth

# The Center for Causal Discovery (CCD)

## An NIH Big Data to Knowledge (BD2K) Center of Excellence

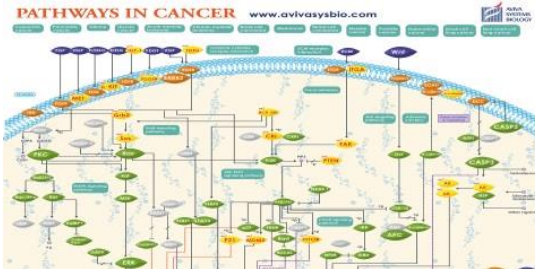
*To help discover valid, novel, and significant causal relationships in big biomedical data that lead to new insights in health and disease.*

- Develop highly efficient causal discovery algorithms that can be practically applied to very large biomedical datasets
- Conduct projects addressing 3 distinct biomedical questions (cancer driver mutations, lung fibrosis, brain causome) as a vehicle for algorithm development and optimization
- Disseminate causal discovery algorithms, software, and tools
- Train data scientists and biomedical investigators in the use of CCD tools
- Train data scientists and biomedical investigators to collaborate in the discovery of causality

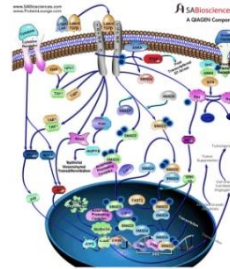
Supported by the NIH National Human Genome Research Institute under award number 5U54HG008540 (\$11M).

Yale

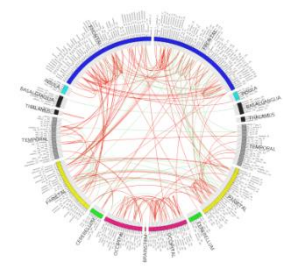
## Driving Biomedical Projects



Discover cell  
signaling networks  
in cancer



Discover the mechanisms of  
disease onset and  
progression in chronic  
obstructive pulmonary  
disease and idiopathic  
pulmonary fibrosis



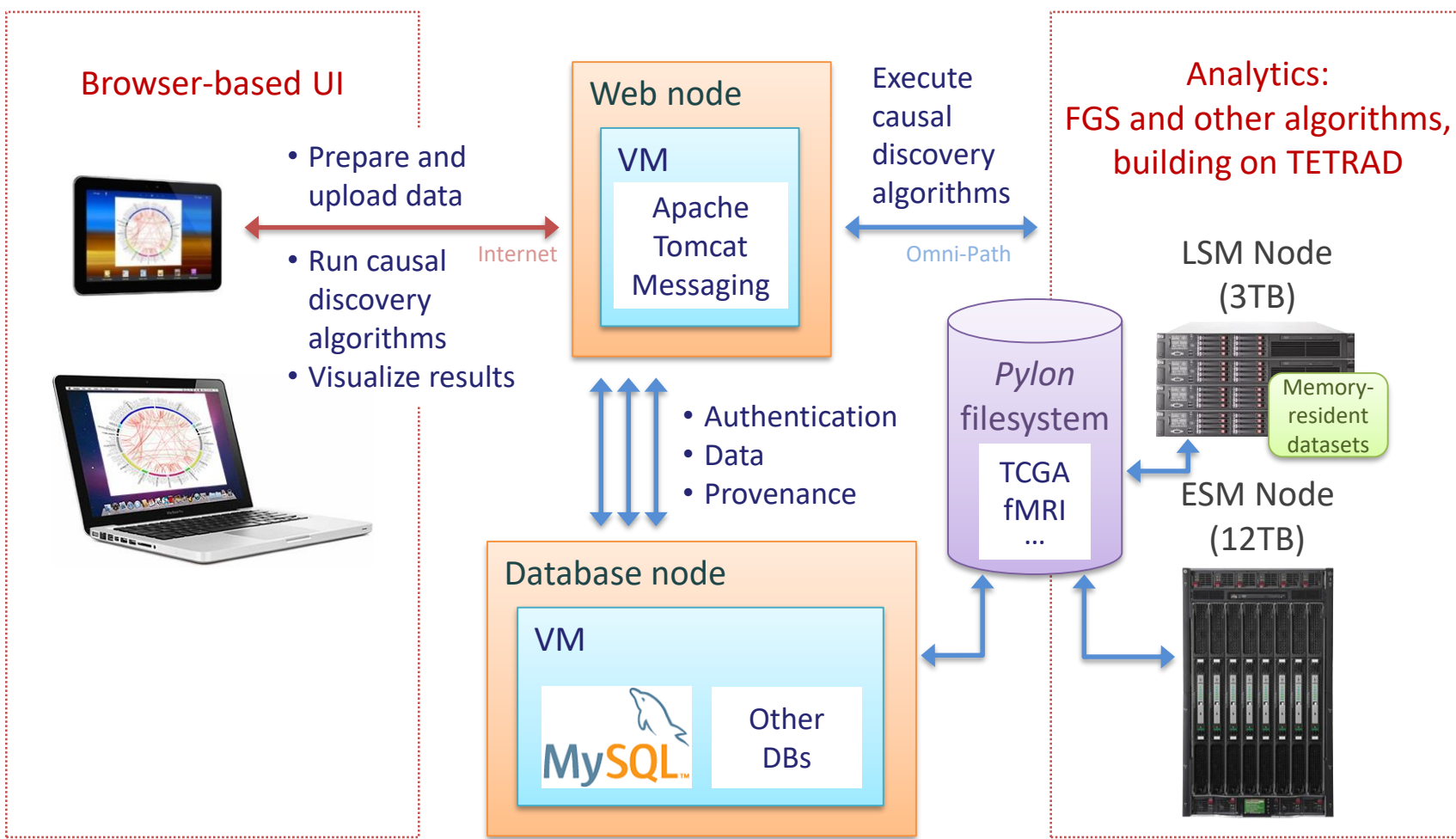
Discover the functional  
(causal) connectivity of  
regions of the human brain  
from fMRI data

*Courtesy Greg Cooper (Pitt)*


Yale

# Example: Causal Discovery Portal

Center for Causal Discovery, an NIH Big Data to Knowledge Center of Excellence



# Accessing *Bridges*

	Open Research				Industry
	Pittsburgh Research Computing Initiative (PRCI)				PSC Corporate Program
		Startup	Research	Education	
Cost	No charge	No charge	No charge	No charge	Cost recovery
CPU-hours	50k	50k	Up to $\sim 10^7$	Up to $\sim 10^6$	Up to $\sim 18M$
GPU-hours	2500	2500	Up to $\sim 10^5$	Up to $\sim 10^4$	Up to $\sim 180k$
GPU-AI hours	1500	1500	Up to $\sim 10^5$	Up to $\sim 10^4$	Up to $\sim 69k$
TB-hours	1000	1000	Up to $\sim 10^4$	Up to $\sim 10^4$	Up to $\sim 137k$
Developer	-	Yes	Yes	(Yes)	Yes
Accepted	Any time	Any time	Quarterly	Any time	Any time
Awarded	$\sim 1-2$ days	$\sim 1-2$ days	Quarterly	$\sim 1-3$ days	ASAP