

Welcome to the XSEDE Big Data Workshop

John Urbanic
Parallel Computing Scientist
Pittsburgh Supercomputing Center

Who are we?

Your hosts:

Pittsburgh Supercomputing Center

Our satellite sites:

University of Utah

Purdue University

Stanford University

Harvey Mudd College

Wayne State University

Georgia State University

University of Notre Dame

Louisiana State University

Kennesaw State University

Ohio Supercomputer Center

Pennsylvania State University

University of Illinois at Chicago

University of Nebraska - Lincoln

University of Houston - Clear Lake

San Diego Supercomputer Center

North Carolina A&T State University

University of Puerto Rico - Mayaguez

National Center for Supercomputing Applications

XSEDE

Extreme Science and Engineering
Discovery Environment

Who am I?

John Urbanic
Parallel Computing Scientist
Pittsburgh Supercomputing Center

Parallelize codes with

- MPI, OpenMP, OpenACC, Hybrid
- Big Data, Machine Learning

Mostly for XSEDE platforms. Mostly to extreme scalability.

XSEDE

Extreme Science and Engineering
Discovery Environment



XSEDE HPC Monthly Workshop Schedule

- November 3 *HPC Monthly Workshop: Big Data*
- December 3 *HPC Monthly Workshop: OpenACC*
- January 20 *HPC Monthly Workshop: OpenMP*
- February 9,10 *HPC Monthly Workshop: MPI*
- March 8 *HPC Monthly Workshop: OpenACC*
- April 5 *HPC Monthly Workshop: Big Data*
- May 10 *HPC Monthly Workshop: OpenMP*
- June 14-17 *HPC Monthly Workshop: HPC Summer Boot Camp*
- August 9 *HPC Monthly Workshop: Big Data*
- September 7,8 *HPC Monthly Workshop: MPI*
- October 4 *HPC Monthly Workshop: OpenMP*
- November 1 *HPC Monthly Workshop: Big Data*
- December 6 *HPC Monthly Workshop: OpenACC*
- January 16 *HPC Monthly Workshop: OpenMP*
- **February 10** ***HPC Monthly Workshop: Big Data***

HPC Monthly Workshop Philosophy

- Workshops as long as they should be.
- You have real lives...
 - in different time zones...
 - that don't come to a halt.
- Learning is a social process
 - This is not a MOOC

Agenda

Tuesday, November 1st

- 11:00 Welcome
- 11:25 Intro To Big Data
- 11:45 Hadoop
- 1:00 Lunch Break
- 2:00 Spark
- 4:15 A Big Big Data Platform: Bridges
- 5:00 Adjourn

20 Storage Building Blocks, implementing the parallel *Pylon* filesystem (~10PB) using PSC's SLASHFS nodes

2 front-end nodes
2 boot nodes
8 management nodes
6 "core" Intel® OPA edge switches: fully interconnected, 2 links per switch

Intel® OPA cables

800 HPE Apollo 2000 (128GB) compute nodes

4 HPE Integrity Superdome X (12TB) compute nodes

42 HPE ProLiant DL580 (3 TB) compute nodes
12 HPE ProLiant DL380 database nodes
6 HPE ProLiant DL360 web server nodes

20 "leaf" Intel® OPA edge switches

32 RSM nodes with NVIDIA next-generation GPUs
16 RSM nodes with NVIDIA K80 GPUs

Purpose-built Intel® Omni-Path topology for data-

<http://staff.psc.edu/nystrom/b>

Resources

Your local TAs

Questions from the audience

On-line talks

bit.ly/XSEDE-Workshop

Getting Time on XSEDE

XSEDE

Extreme Science and Engineering
Discovery Environment

<https://portal.xsede.org/web/guest/allocations>