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:
Tufts University
Lehigh University
Howard University
University of Iowa
Purdue University
Harvey Mudd College
University of Delaware
Old Dominion University
Georgia State University
George Mason University
Louisiana State University
Michigan State University
Oklahoma State University
Kennesaw State University
Ohio Supercomputer Center
Pennsylvania State University
University of Nebraska-Lincoln
University of Texas at El Paso
San Diego Supercomputer Center
University of Houston - Clear Lake
University of California, Los Angeles
North Carolina A&T State University
Yale Center for Research Computing
National Center for Supercomputing Applications
University of Tennessee, Knoxville - National Institute for Computational Sciences
Who am I?

John Urbanic
Parallel Computing Scientist
Pittsburgh Supercomputing Center

What I mostly do:

Parallelize codes with

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

Primarily for XSEDE platforms. Mostly to extreme scalability.
<table>
<thead>
<tr>
<th>Date</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 6-9</td>
<td>Summer Boot Camp</td>
</tr>
<tr>
<td>August 15</td>
<td>HPC Monthly Workshop: OpenMP</td>
</tr>
<tr>
<td>September 12-13</td>
<td>HPC Monthly Workshop: Big Data</td>
</tr>
<tr>
<td>October 3-4</td>
<td>HPC Monthly Workshop: MPI</td>
</tr>
<tr>
<td>November 7</td>
<td>HPC Monthly Workshop: OpenACC</td>
</tr>
<tr>
<td>December 5-6</td>
<td><strong>HPC Monthly Workshop: Big Data</strong></td>
</tr>
<tr>
<td>January 9</td>
<td><strong>HPC Monthly Workshop: OpenMP</strong></td>
</tr>
<tr>
<td>February 7-8</td>
<td><strong>HPC Monthly Workshop: Big Data</strong></td>
</tr>
<tr>
<td>March 6</td>
<td><strong>HPC Monthly Workshop: OpenACC</strong></td>
</tr>
<tr>
<td>April 3-4</td>
<td><strong>HPC Monthly Workshop: MPI</strong></td>
</tr>
<tr>
<td>May 1-2</td>
<td><strong>HPC Monthly Workshop: Big Data</strong></td>
</tr>
<tr>
<td>June 4-7</td>
<td>Summer Boot Camp</td>
</tr>
<tr>
<td>August 7</td>
<td><strong>HPC Monthly Workshop: OpenMP</strong></td>
</tr>
<tr>
<td>September 5-6</td>
<td><strong>HPC Monthly Workshop: Big Data</strong></td>
</tr>
<tr>
<td>October 2-3</td>
<td><strong>HPC Monthly Workshop: MPI</strong></td>
</tr>
<tr>
<td>November 6</td>
<td><strong>HPC Monthly Workshop: OpenACC</strong></td>
</tr>
<tr>
<td>December 4-5</td>
<td><strong>HPC Monthly Workshop: Big Data</strong></td>
</tr>
</tbody>
</table>
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
  - This is the Wide Area Classroom
  - so raise your expectations
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>Welcome</td>
</tr>
<tr>
<td>11:25</td>
<td>Intro To Big Data</td>
</tr>
<tr>
<td>12:00</td>
<td>Hadoop</td>
</tr>
<tr>
<td>12:30</td>
<td>Intro to Spark</td>
</tr>
<tr>
<td>1:00</td>
<td>Lunch Break</td>
</tr>
<tr>
<td>2:00</td>
<td>Spark</td>
</tr>
<tr>
<td>3:30</td>
<td>Spark Exercises</td>
</tr>
<tr>
<td>4:30</td>
<td>Spark</td>
</tr>
<tr>
<td>5:00</td>
<td>Adjourn</td>
</tr>
</tbody>
</table>

**Wednesday, December 6**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:00</td>
<td>Machine Learning: A Recommender System</td>
</tr>
<tr>
<td>1:00</td>
<td>Lunch break</td>
</tr>
<tr>
<td>2:00</td>
<td>Deep Learning with Tensorflow</td>
</tr>
<tr>
<td>4:30</td>
<td>A Big Big Data Platform</td>
</tr>
<tr>
<td>5:00</td>
<td>Adjourn</td>
</tr>
</tbody>
</table>
We do this all the time, but...

- This is a very ambitious agenda.
- We are going to cover the guts of a semester course.
- We may get a little casual with the agenda.
- Three reasons we can attempt this now:
  - Tools have reached the point (Spark and TF) where you can do some powerful things at a high level.
  - We are going to assume you will use your extended access to do exercises. Usually this is just a bonus.
  - Worked last time.
Resources

Your local TAs

Questions from the audience

On-line talks

bit.ly/XSEDEWorkshop

Copying code from PDFs is very error prone. Subtle things like substituting “-” for “-“ are maddening. I have provided online copies of the codes in a directory that we shall shortly visit. I strongly suggest you copy from there if you are in a cut/paste mood.
16 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
42 HPE ProLiant DL580 (3TB) compute nodes
12 HPE ProLiant DL380 database nodes
6 HPE ProLiant DL360 web server nodes
4 MDS nodes
2 front-end nodes
2 boot nodes
8 management nodes
Intel® OPA cables
20 Storage Building Blocks, implementing the parallel Pylon storage system (10 PB usable)
4 HPE Integrity Superdome X (12TB) compute nodes...
... each with 2 gateway nodes
4 HPE ProLiant DL580 (3TB) compute nodes
20 “leaf” Intel® OPA edge switches
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>12TBb</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>12TbC</td>
<td>2</td>
<td>2</td>
<td>16 × Intel Xeon E7-8880 v4 (22c, 2.2/3.3 GHz, 55MB LLC)</td>
<td></td>
</tr>
<tr>
<td>LSM</td>
<td>3TBb</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>3TBc</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></td>
</tr>
<tr>
<td>RSM</td>
<td>128GBb</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 Apollo 2000</td>
</tr>
<tr>
<td>RSM-GPU</td>
<td>128GBb</td>
<td>1</td>
<td>16</td>
<td>2 × Intel Xeon E5-2695 v3 + 2 × NVIDIA Tesla K80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>128GBc</td>
<td>2</td>
<td>32</td>
<td>2 × Intel Xeon E5-2683 v4 (16c, 2.1/3.0 GHz, 40MB LLC) + 2 × NVIDIA Tesla P100</td>
<td></td>
</tr>
<tr>
<td>DB-s</td>
<td>128GBb</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>128GBb</td>
<td>1</td>
<td>6</td>
<td>2 × Intel Xeon E5-2695 v3 + HDDs</td>
<td>HPE ProLiant DL380</td>
</tr>
<tr>
<td>Web</td>
<td>128GBb</td>
<td>1</td>
<td>6</td>
<td>2 × Intel Xeon E5-2695 v3</td>
<td>HPE ProLiant DL360</td>
</tr>
<tr>
<td>Othera</td>
<td>128GBb</td>
<td>1</td>
<td>16</td>
<td>2 × Intel Xeon E5-2695 v3</td>
<td>HPE ProLiant DL360, HPE ProLiant DL380</td>
</tr>
<tr>
<td>Gateway</td>
<td>64GBb</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>64GBc</td>
<td>2</td>
<td>4</td>
<td>2 × Intel Xeon E5-2683 v3</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>128GBb</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>256GBb</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></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>281.75TB</strong></td>
<td><strong>908</strong></td>
<td></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
Getting Time on XSEDE

https://portal.xsede.org/web/guest/allocations
Getting Connected

- The first time you use your account sheet, you must go to apr.psc.edu to set a password. You may already have done so, if not, we will take a minute to do this shortly.

- We will be working on bridges.psc.edu. Use an ssh client (a Putty terminal, for example), to ssh to the machine.

- If you are already an active Bridges user, then to take advantage of the higher-priority training queue we are using for this workshop you will have to change to the training group account that is also available to you:
  - newgrp tr561bp
  - You can see what groups you are in with the “id” command, and which group you are currently using with “id -gn”

- You will want to use the training group today. With hundreds of us on the machine, the normal interact access time might leave you waiting for a bit.
Getting Connected

At this point you are on a login node. It will have a name like “br001” or “br006”. This is a fine place to edit and compile codes. However we must be on compute nodes to do actual computing. We have designed Bridges to be the world’s most interactive supercomputer. We generally only require you to use the batch system when you want to. Otherwise, you get your own personal piece of the machine. For this workshop we will use

`interact`

to get a regular node of the type we will be using with Spark. You will then see name like “r251” on the command line to let you know you are on a regular node. Likewise, to get a GPU node, use

`interact -gpu`

This will be for our Tensorflow work tomorrow. You will then see a prompt like “gpu32”.

Some of you may follow along in real time as I explain things, some of you may wait until exercise time, and some of you may really not get into the exercises until after we wrap up tomorrow. It is all good.
We have hundreds of packages on Bridges. They each have many paths and variables that need to be set for their own proper environment, and they are often conflicting. We shield you from this with the wonderful modules command.

You can load the two packages we will be using as

**Spark**

```
module load spark
```

**Tensorflow**

```
module load tensorflow/1.1.0
source $TENSORFLOW_ENV/bin/activate
```

The Tensorflow one is atypical and reflects the complexities of its installation. If you find either of these tedious to repeat, feel free to put them in your .bashrc.
Editors

For editors, we have several options:

- *emacs*
- *vi*
- *nano*: use this if you aren’t familiar with the others

For this workshop, you can actually get by just working from the various command lines.
Programming Language

- We have to pick something
- Pick best domain language
- Python
- But not “Pythonic”
- I try to write generic pseudo-code
  - If you know Java or C, etc. you should be fine.

Warning!

Several of the packages we are using are very prone to throw warnings about the JVM or some python dependency. We’ve stamped most of them out, but don’t panic if a warning pops up here or there.

In our other workshops we would not tolerate so much as a compiler warning, but this is the nature of these software stacks, so consider it good experience.
Our Setup For This Workshop

After you copy the files from the training directory, you will have:

/BigData
   /Clustering
   /MNIST
   /Recommender
   /Shakespeare

Datasets, and also cut and paste code samples are in here.
Let’s get the boring stuff out of the way now.

- Log on to apr.psc.edu and set an initial password if you have not.
- Log on to Bridges.
  
  \texttt{ssh username@bridges.psc.edu}

- Copy the Big Data exercise directory from the training directory to your home directory.
  
  \texttt{cp -r ~training/BigData .}

- Edit a file to make sure you can do so. Use \texttt{emacs}, \texttt{vi} or \texttt{nano} (if the first two don’t sound familiar).

- Start an interactive session.
  
  \texttt{interact}