High Performance Computing in a Small College Environment: Tools, Techniques and Resources

Oklahoma Supercomputing Symposium

October, 2007

Paul Gray, University of Northern Iowa

Charlie Peck, Earlham College

How your IT department can help you

- Two approaches, active and passive. For active they will be involved with:
  - system administration
  - network administration
  - database administration (possibly)
  - performance analysis (possibly)
- Repeatability, stability, sensitivity to upgrades, and all the other details that scientists care about.
- For passive they give you a sub-net, bandwidth (widely varying needs), and advice now and again.
- Both active and passive approaches benefit greatly from the involvement of students in all aspects of the work: support, science, etc.

- Software
  - Why use open source software? Looking under the hood is very important in this context. Consider verification, validation, collaborative development, low TCO, high value, open standards, open science.
  - System software, science software
  - Bootable Cluster CD - ad-hoc lab clusters, fixed clusters
- Site logistics for your own cluster
  - Power
  - Cooling and humidity
  - Environmental Monitoring - temperature, humidity, water, power; WeatherDuck
  - Systems Monitoring - availability, services, capacities; Ganglia, Nagios, MRTG, Cacti
  - Access

Most important aspect in support of High Performance Computing in a Small College Environment: Communication

- Communicate with your colleagues in other disciplines. Likely you'll find that your colleagues are doing HPC in isolation.

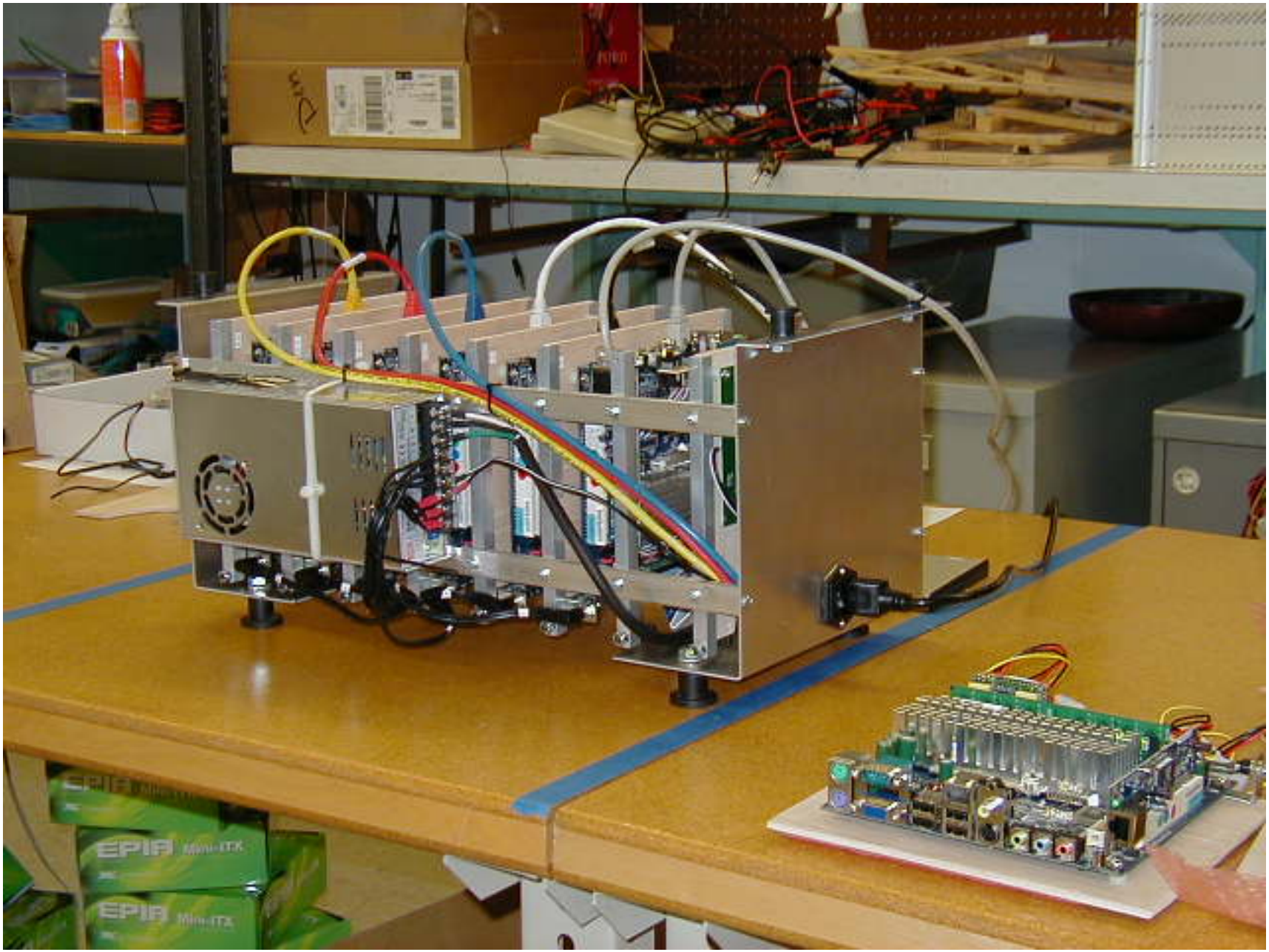
- Communicate with your Deans, Provosts, and Presidents. Likely they do not know the scope and scale of HPC inside of their own walls. Nor are they aware of the impediments in place at the institution that preclude moving things to the next level.

- Communicate with your facilities management about appropriate environmental conditions for HPC resources. Heat kills.

- Communicate with your peers about the science enabled in your classroom which is based upon HPC resources. Enjoy the "theater" of HPC.

- Get out in the community, attend NCSI/SC Education/TeraGrid workshops to learn more about the curricular materials, techniques, and resources available to you for teaching and research.

LittleFe - A portable platform for teaching HPC and Parallel/Distributed computing


EPIA Mini-ITX
EPIA Mini-ITX

- Many institutions and teaching environments do not have access to parallel platforms for parallel and distributed computing education. Teaching key concepts such as speedup, efficiency, and load balancing are much more effectively done on a parallel platform.

- LittleFe is a complete 6 node Beowulf style portable computational cluster. LittleFe weighs less than 50 pounds, easily and safely travels via checked baggage on the airlines, and sets-up in 10 minutes wherever there is a 110V outlet and a wall to project an image on.

- By leveraging the Bootable Cluster CD project, and its associated curriculum modules, LittleFe makes it possible to have a powerful ready-to-run computational science and HPC educational platform for about $3,000.

General Resources

Beowulf mailing list, archives, How-Tos, *etc.* -
`http://www.beowulf.org`

IEEE Task Force on Cluster Computing -
`http://www.ieeetfcc.org`

ClusterMonkey - `http://www.clustermonkey.net`

Computing in Science and Engineering, IEEE CS and the AIP -
`http://www.computer.org/cise`

The National Computational Science Institute -
`http://www.computationalscience.org`

Computational Science Education Reference Desk (CSERD) - `http://www.shodor.org/refdesk/Resources`

The Krell Institute - `http://www.krellinst.org`

NCSA's on-line tutorials (MPI, cluster computing, etc.) - `http://webct.ncsa.uiuc.edu:8900`

Robert Brown's book - `http://www.phy.duke.edu/~rgb/Beowulf/beowulf.php`

Cluster Software Resources

DIY vs off-the-shelf - flame war ensues...

- Open Cluster Group (OSCAR) - `http://openclustergroup.org`
- Platform ROCKS - `http://www.platform.com/Products/Platform.OCS`
- The Bootable Cluster CD - `http://bccd.cs.uni.edu`

User Software Resources

NetLib - `http://www.netlib.org`

GNU (GNU's Not Unix) - Free Software Foundation - `http://www.gnu.org`

- Science packages - `http://www.gnu.org/directory/science`
- Math packages - `http://www.gnu.org/directory/science/math`

OSDN - Open Source Developers Network

- SourceForge - `http://sourceforge.net`

- FreshMeat - `http://freshmeat.net`
- Slashdot - `http://slashdot.org`

Google

Your peers

Evaluating Software

- CiteSeer - `http://citeseer.ist.psu.edu`
- Google's scholar - `http://scholar.google.com`
- FAQs
- Mailing list archives
- Your peers

Scientific Software Packages

GNUmp

GSL

Molecular Dynamics

- GROMACS - `http://www.gromacs.org`
- NAMD - `http://www.ks.uiuc.edu/Research/namd`

Linear Algebra

- LAPACK - `http://www.netlib.org/lapack`

- Trilinos - `http://software.sandia.gov/trilinos`
- High-Performance BLAS (Goto) - `http://www.tacc.utexas.edu/resources/software/`

PETSc (PDE and ODE) - `http://www-unix.mcs.anl.gov/petsc/petsc-as` (parallelized with MPI)

FFTW, Fastest Fourier Transformation in the West - `http://www.fftw.org` (parallelized with MPI)

N.B. Cite the tools you use in your papers, most have a description (and BibTex source) for how they would like it done. Reference counting helps the developers to justify continued support and makes it easier for people like you to identify good tools.