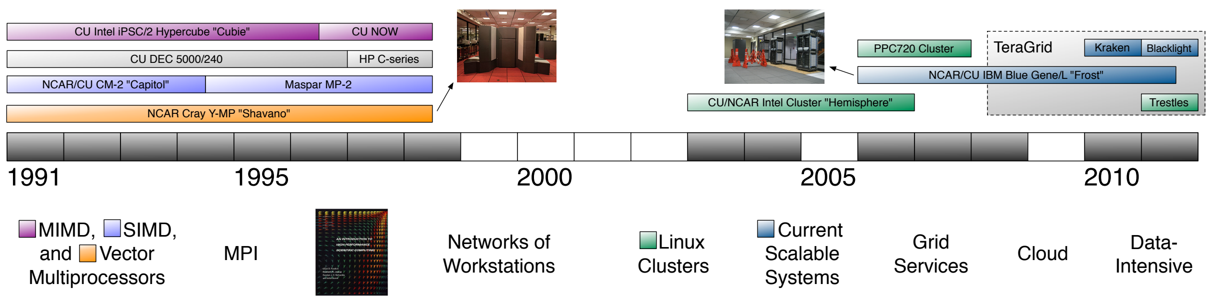
Practical Cyberinfrastructure for High-Performance and Parallel Computing Education

Matthew Woitaszek, Henry M. Tufo, and Elizabeth R. Jessup
University of Colorado, Boulder
National Center for Atmospheric Research

The University of Colorado's High-Performance Scientific Computing (HPSC) course was established in 1991. The course introduces undergraduate and graduate students to the theory and practice of scientific computing with emphasis on parallel software development, performance and scalability analysis, and the influence of architectures and topologies on program design, implementation, and performance.

Course Platforms and Technologies

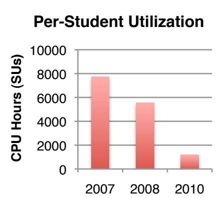


In recent years, the course's composition and presentation have been updated to emphasize practical hands-on experience with today's platforms and technologies of high-performance scientific computing.

TeraGrid Resources

POWERED BY TeraGrid

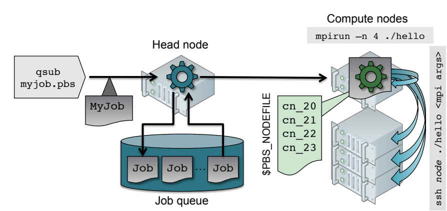
- Goal: **Provide students with direct experience on current NSF cyberinfrastructure**
- All exercises and projects have been conducted exclusively with TeraGrid resources since 2008
- Diverse resources for:
  - comparative performance analysis
  - shared memory programming
  - data-intensive workflow experimentation
  - Grid technology demonstrations



Students experience TeraGrid as real investigators on a shared allocation. Usage depends on course content (e.g., a computational midterm) and student projects, but has generally decreased as students now use laptops for debugging and the course explores a broader range of topics.

Linux Cluster Construction

- Goal: **Reinforce operating system concepts that allow students to effectively use Linux**
  - obtain experience with useful Linux tools
  - understand how cluster systems work
  - become "power users" of shared systems
- Conducted using cloud-based resources
- Exercises:
  - turn networked nodes into a cluster
  - install and configure MPI, Torque, Ganglia
  - benchmark cluster and monitor performance



Students learn cluster architecture and experience configuring the software used to launch and run parallel programs.

Conclusions and Continuing Work

- These practical technology experiences support computational thinking skills for diverse disciplines.
- For HPSC, we are exploring high-performance computing beyond batch scheduling: multi-site workflows, data-centric computing, cloud-based resources, and new parallel programming techniques.
- Our use of these concepts and technologies in this elective course demonstrates their applicability to the systems-centric undergraduate computer science curriculum; we expect migration to core courses.

Acknowledgements

Thanks to the other HPSC instructors and recent teaching assistants: Lloyd Fosdick, Carolyn Schauble, Maria Murillo, Jason Cope, Theron Voran, Paul Marshall, Ian Karlin, Ben Barrow, Tim Dunn, and Cory Zachman. Thanks to Theron Voran for assistance with the poster and Guy Cobb for the cluster construction figure. The picture of Shavano is from the CISL Supercomputer Image Gallery at http://www.cisl.ucar.edu/computers/gallery/cray. The HPSC course was developed with support from National Science Foundation (NSF) award CDA-9017953 and the University of Colorado. Computational resources and support were provided by the University of Colorado, NSF TeraGrid, NSF MRI grants CNS-0421498, CNS-0420873, CNS-0420985, NSF sponsorship of the National Center for Atmospheric Research, and the IBM Shared University Research program.