

## Keynote Talk

# Scaling Up Data Intensive Scientific Applications to Campus Grids

Douglas Thain  
University of Notre Dame  
Notre Dame, IN, USA  
dthain@nd.edu

### Abstract

A campus grid aggregates computing and storage capacity from many clusters, desktops, and servers to provide a cost-effective large scale computing facility for an entire institution. Many universities today operate campus grids on the scale of thousands to tens of thousands of nodes. However, the people that might derive the most benefit from campus grids are generally accustomed to using single computers or conventional clusters. Making the jump from a single cluster to a campus grid can be very challenging, because the user or programmer must contend with a host of performance and reliability challenges that only arise in large scale systems.

To solve this problem, we advocate the use of high level abstractions. An abstraction is a framework that represents a very narrow category of problems, but is both robust and scalable across very large problem sizes. We have developed a suite of abstractions and deployed them on campus grids at the University of Notre Dame, Purdue University, and the University of Wisconsin to attack large scale problems in fields such as biometrics, bioinformatics, data mining, and economics. With this approach, we have enabled end users to efficiently harness hundreds to thousands of CPUs without requiring them to learn a radically new programming model.

**Categories & Subject Descriptors:** C.2.4 [Computer Communication Networks] Distributed Systems; D.1.3 [Programming Techniques] Concurrent Programming

**General Terms:** Algorithms, Design, Performance

### Bio

Prof. Douglas Thain is an Assistant Professor of Computer Science and Engineering at the University of Notre Dame. He directs the Cooperative Computing Lab, an enterprise that connects computer scientists with researchers in other scientific fields to solve new problems on large distributed systems. He received the M.S. and Ph.D. in Computer Sciences from the University of Wisconsin, where he contributed to the Condor distributed computing project. He received a B.S. in Physics from the University of Minnesota. Prof. Thain is the recipient of an NSF CAREER award.