

Campus Grids at Indiana University

Note: Indiana University and Purdue University propose coordinated talks, starting with Purdue University and followed by Indiana University. This will allow each university to discuss their own campus grids, and leave a few minutes at the end of the second talk to discuss statewide grid efforts. A description of IU's campus grid activities follows below, followed by statewide grid efforts.

Indiana University Grid efforts

Population

IU campus grid efforts support local and national users, as follows:

- Computational biology users. We have deployed parallel versions of important bioinformatics applications on a Condor flock that encompasses our student computing labs, and deliver this as a production resource for users of these particular applications. This facility serves a relatively small number of users who use a tremendous amount of computer resources.
- Local researchers. We maintain a grid of high performance computing applications that includes facilities at IU Bloomington and the Indiana University – Purdue University Indianapolis campus. Local researchers from both campuses run across this intercampus grid, providing resilience in case of disaster and also better service for users on each campus. Overall our high performance computer grids serve more than 1,000 local users who use these facilities routinely.
- National research community. Indiana University is part of the TeraGrid and participates in the Open Science Grid. We serve users from the national user community who are not part of IU through these grids and their local instantiations at IU. We have generated accounts on our grid facilities for well over a thousand users.

We provide user education and training through a number of vehicles. A key issue is simply interesting people in using campus grids. We host periodic symposia (www.i-light.org for more information) which focus on researchers explaining how their own research programs have benefited from use of grid computing. In addition, we provide regular training sessions, online help, and a small and very popular booklet entitled “IU's advanced information technology resources: the least every researcher needs to know.”

Applications

- Our Condor pool runs three very popular bioinformatics applications: BLAST, fastDNAMl, and MEME
- Our campus research grid runs a very large number of applications – literally hundreds. The applications are concentrated in life sciences, chemistry, and physics
- National users of our campus grids tend to focus on chemistry and physics applications

Infrastructure

- Software Stack

- Our resources in general are heterogeneous, and we use multiple software distribution and packaging tools to manage a very diverse suite of software. Synchronization of software in compliance with multiple grid research efforts (especially the TeraGrid and the Open Science Grid) requires significant human effort, even though we use a variety of tools – in some case specific to particular projects – within a given pool of resources.
- Our Condor pool runs Condor on windows, with an IU-developed parallel communications library called SMBL – Simple Message Brokering Library – which manages a subset of the MPI standard communications between Condor workers in a way that enables MPI applications to tolerate a constantly shifting pool of compute resources.
- Our campus research grid generally includes the TeraGrid CTSS (Common TeraGrid Software Stack), based on Globus.
- Our facilities involved in the Open Science Grid run OSG-compliant installations of Condor on Linux

Account management

- Any member of the university research community is eligible to receive accounts on our systems, and there are no usage limits or allocations. (Jobs are prioritized on many resources via a fair-share scheme). Users of our campus grids include faculty, staff, graduate students, and a small number of undergraduates who have faculty sponsors supervising their research projects. Local users are managed via an IU-created account management system.
- We also support users within the TeraGrid and Open Science Grid communities, with accounts managed by the facilities pertinent to each.

Usage Tracking

- We track usage, and use it for local management and national reporting. We do not reveal any detailed user information in an 'identifiable' way, but we do use usage information as a way to plan our support activities. Users that use extremely large amounts of CPU time are invited to receive consulting help with code optimization. We do report usage on a department-by-department basis as a matter of public record. This helps the academic units of the university understand the value they receive from computing grids on campus.
- We also collect and report usage via project-specific mechanisms for the TeraGrid and the Open Science Grid, which are reported back to the relevant contact points within these projects.
- We track usage on a system-by-system basis, or an an architecture-by-architecture basis, which allows us to manage system heterogeneity (except within the Condor pool, where CPU heterogeneity is simply ignored at present and all we track is CPU hours). We track CPU usage, disk usage, and archival tape usage.

Network Connectivity

- We have local campus networks that are primarily 100 baseT ethernet, and it is via this network that the Condor pool is connected. At each campus (Bloomington and Indianapolis) we have large Force10 Gigabit ethernet switches, connected to each other and national networks. The Force10 switches function essential as a backplane for each campus grid. The Indianapolis and Bloomington campus grids are connected

to each other by multiple 10GigE connections, and are connected to the TeraGrid network via a 20 Gbit connection. We have two 1 GigE connections to Abilene/Internet2. As a result we have relatively few network bottlenecks.

Security

- We use kerberos authentication within our research grid generally. For our Condor pool we use ads authentication.

Management

- There are significant technical issues in terms of managing our campus grids, especially as regards security. These simply require a considerable amount of expertise and effort to manage. For example, one of the principles of operation of the TeraGrid is that there be no firewalls between TeraGrid resources and the TeraGrid network. For this reason, we have dual network connections (one to the TeraGrid, one to the campus network) for each system connected to the TeraGrid, and IU firewall policy is enforced between such TeraGrid-connected resources and other campus resources.
- Administrative issues: While our campus grids are significant in size and widespread, most of the resources are controlled by the central campus IT organization (University Information Technology Services - UITS). We have collaborative relationships with other resource providers and participants in the grid which enable effective participation of such resources in our campus grids. In some cases, such as the widely used Flybase, the IU Department of Biology runs servers connected to Abilene and the general national and international research community, while UITS operates a mirror server which is directly connected to the TeraGrid.
- Political issues. We have had relatively few 'political' issues to deal with, in large part thanks to a perception that UITS, which operates the majority of the campus grid efforts, is diligent in trying (and succeeding) to meet the needs of local researchers.

Other issues

- Indiana University operates a production local grid and participates in two national grid projects. IU's unusual status of having two large research campuses, connected by university-owned optical fiber, provides very interesting opportunities for grids that span multiple campuses. For example, we have a storage grid that includes distributed data movers and mirrored data silos on IU's two core campuses, providing disaster resilience and reliable protection of data in case of natural disaster.
- Indiana University and Purdue University operate a statewide network called the Ilight network. This network, jointly owned by Purdue and Indiana universities, connects the Purdue campus in West Lafayette, the Indiana University Purdue University joint campus in Indianapolis, and the Indiana University campus in Bloomington. This statewide network provides extremely important intercampus grid possibilities, as well as disaster resilience capabilities for both universities.
- Indiana University (alone and in collaboration with Purdue University) are implementing production and experimental campus grids that are delivering important research services today, as well as providing research and development platforms for the grids of tomorrow.