White Paper

on

Effective User Services for

High Performance Computing (HPC)

TeraGrid Science Advisory Board

May 2009

Contents

1 HPC User Services …………………………………………………...… 3

2 Characteristics of HPC User Services ………………………………….. 4

3 Essential Elements of an HPC User Services Program ………………… 5

3.1 Routine User Services ………………………………………….. 7

3.2 Effective Use of Resources …………………………………….. 9

3.3 Project Domain Specific Support ………………………………. 9

3.4 User-Developed Software Support ……………………………... 10

3.5 Communications, Education, and Outreach ……………………. 11

4 Social Web and HPC User Services ……………………………………. 12

5 Conclusions …………………………………………………………….. 13

6 Appendix: TeraGrid Science Advisory Board …………………………. 14

1 HPC User Services

This white paper describes our views on the **User Services** that are needed to effectively support users of the National Science Foundation's (NSF) open-science High Performance Computing (HPC) environments including the TeraGrid and the upcoming eXtreme Digital (XD) systems. The purpose of the white paper is to provide NSF program directors, both within the Office of Cyberinfrastructure (OCI) and the domain science directorates, with a basic understanding of the importance of user services within HPC organizations. It is intended to provide background for more critical evaluation of the user services portions of HPC proposals, such as are being considered for the XD program. This white paper was developed by the TeraGrid Science Advisory Board (SAB; see Appendix), a group of domain scientists who appreciate the importance of HPC for advancing and accelerating research in science and engineering. We drew on our years of experience in using digital resources to conduct basic research, including using NSF and other HPC systems, in order to provide our recommendations for an effective HPC User Services program.

This paper is focused on High Performance Computing, because that has been the iconic resource provided by the TeraGrid and its predecessors. We recognize that HPC is only one of several high-end digital services that can and should be provided by TeraGrid and XD. The provision of a communications grid based on high-speed networks, data storage, archive, and management services, visualization, and linkages to special high-end equipment that require large-scale computational support are also essential components. Nevertheless, a full evaluation of the range of user services that may be needed to support resources other than HPC is beyond the scope of this paper.

In this paper we use the phrase **User Services** to describe a wide range of activities that computer system resource providers (i.e. supercomputer centers) can perform to help users make effective use of computer systems. In some cases, User Services may refer to something as simple as helping users login to a computer system or change their passwords. In other cases, User Services may involve very sophisticated, high level, support provided by HPC resource providers to scientific users. An effective User Services program should benefit both the computer resource providers as well as the computer users. For computer resource providers, User Services should ensure that costly computer resources are used safely, securely, efficiently, and are utilized to their full capability for a wide range of sciences and users. For the computer system users, User Services may ensure that users efficiently and effectively access the full range of capabilities of a computer system, gaining the full advantage for their scientific research while resolving quickly any issues that may arise and minimizing any adverse impacts on research productivity.

HPC User Services have much in common with the User Services provided by any academic or business computer center. Some User Services, such as helping a user log into their account for the first time, must be provided by all User Services programs. However, the very high performance computers, software, and networks, enabling work that cannot be done on lesser machines, attract a wide and diverse set of scientists particularly a very high-end, strongly computationally -oriented scientific user group who require special considerations when setting up a HPC User Services Program to make the most of the HPC systems and push the high-end of computational science. In this paper, we discuss our views on what services should be provided by a User Services program that is designed to support scientific HPC users.

2 Characteristics of HPC Users

HPC User Services should provide many of the User Services provided by any large business or academic computer organization. For example, nearly all User Services programs must help users log into their accounts, change their passwords, and run commonly-used application programs. Despite these commonalities, it is important to recognize that scientific HPC users have specific characteristics that are atypical of computer users in general and that will influence the design of an HPC User Services program.

The group of scientific HPC users who use the largest part of any system are typically:

- very technically savvy;
- willing to invest extensive effort to utilize seldom-used capabilities of a computer resource;
- likely to discover hardware or system software problems, but are almost never responsible for repairing hardware problems or debugging system software;
- using software written by themselves or scientific colleagues, so they need HPC software development environments;
- very interested in trying new versions of software to obtain the latest features;
- likely to push the limits of the capabilities of the computer resources, routinely attempt leading edge, at times bleeding edge, computations;
- unlikely to expect to be charged for support; and
- constantly in flux as scientists become new HPC users and also as HPC resources change with time.

Additionally, there are three other groups of HPC users: those who are just starting out, those who wish to be moderate users and those who prefer science gateways or other mechanisms that deal with some of the complexities of HPC systems for the user. We are putting outside the scope of this report the many scientists that have scientific problems that could benefit from HPC, but perceive the complexity of the systems and the time investment to learn the system to be too high to merit the unknown potential rewards and science gateways are perceived to not address their science problem. All of these groups need support that is quite different from the high-end scientific HPC users. In general, they are not as technically savvy as the current typical HPC user, and in some cases do not wish to be, specifically with HPC systems, although they may have ample understanding of how to use and program computers; some are not willing to invest as much effort as the current typical HPC user into using the resources; they are likely to be stopped by HPC problems, rather than working to identify them in detail; they often use software developed (and installed) by others, particularly that which is also available on non-HPC systems; they often perceive current software to be adequate for their science and are not concerned if the software has a newer version; and they are unlikely to push the limits of the computer resources; although they are doing work that they could not do elsewhere. Like the high-end scientific HPC users, they are unlikely to expect to be charged for support.

HPC User Services needs to support all groups of users, as the high-end users are the main users today, but the latter groups are fast growing, and can both increase in number as well as move into the former category tomorrow. Science is well served by an HPC User Services that contributes toward the goals of “wide” as well as “deep.”

3 Essential Elements of an HPC User Services Program

As described above, scientific HPC User Services extend over a range of types that often go beyond what is typically provided in a business or academic computing organization. It is important that a customer service model be adopted that addresses the full range of types of service, with multiple pathways for feedback to and from users, and a hierarchical, web-based documentation system that enables users to quickly and accurately identify the service(s) they require. A simple categorization of the types of services needed by an HPC user services organization is shown in Table 1. This is not intended as an organizational structure for User Services, but rather as an identification of key responsibilities within an HPC User Services program.

Routine User Services	Provide user account management, allocation information, add/remove users, storage quota, hardware and software documentation, and issue resolution using a trouble ticket system (triage, transition, tracking and resolution).
Effective Use of the Resources	Help users identify appropriate hardware and software for their needs. Provide users with information on how to make best use of cyberinfrastructure. Improve performance users obtain from hardware and software.
Project- and Domain-Specific Support	Support for users from a specific project and or a specific domain. Link to resources in appropriate software engineering or scientific communities that enable this support.
User-Developed Software Support	Support for user-developed software, including a robust development environment, analysis of performance, debugging, and porting.
Commonly-Used Software Support	Support for commonly used software, including installation, and working with key users, analysis of performance, debugging, and porting user problems to the software or the system.
Communications, Education, Outreach	Communication of system and support capabilities to current and potential users, as well as examples of science done. Communication of issues and concerns from users to appropriate staff. Communication of results to funding agencies and scientific communities. Educational and training programs for users.

Table 1: Key responsibilities of an effective scientific HPC User Services program

3.1 Routine User Services

Routine User Services for HPC systems are similar but not identical to those required for most academic computer systems. HPC User Services should provide standard user support functions such as adding new accounts, changing passwords, deleting accounts.

Since routine services are fairly standard within education and industry, we will not elaborate extensively on what is required. HPC systems have some additional routine management responsibilities such as allocation management that other shared use computer system may not have. Importantly, since HPC resources are likely to be distributed over a high-speed communications grid, there should be a single point of entry to initiate, track and resolve such routine user services.

Here, we propose a few principles that may help define an effective user support for routine services.

Life cycle: All routine user services go through a series of steps, including analysis of requirements and readiness, triage, transition, tracking and resolution. *Analysis* can determine what a user, whether a novice or an experienced user, needs from a high-end computing environment and how ready a given user may be to efficiently and effectively use the resources. *Triage* is the process whereby a quick determination is made about where a given user services request should be directed. For example, the first order of business is to distinguish between routine user support and technical support. Routine user support provides information and services to users to help them access and use computers productively. Technical support is a level of user support that focuses on higher-level troubleshooting and problem solving. Technical support deals with difficult and complex problems users encounter. Making this distinction early in the process is essential to effectively handling requests. Another matter that can be resolved in the triage step is determining whether a user's issues are associated with moving from a local university computing environment to an HPC environment. Such users are not necessarily novices to computing, but have reached the point where their local environments cannot support their needs. *Transition* refers to the process of moving a user services request to the appropriate section or individual within the user services organization. Routine user support should be handled at the level of a help desk where individuals who are knowledgeable about the accounting system, login security, etc., while technical support should be addressed by software engineers or other trained professionals who have more detailed and technical knowledge of the HPC systems, software, data management tools, and connections to the grid. *Tracking* provides a way for both the user support organization and users to monitor the status and progress of their user services request. There are many examples of successful tracking systems (e.g. FedEx). *Resolution* is the final disposition of a user services request that may or may not satisfy the user's need. Naturally, feedback from the user services organization to the user and vice versa is critical at this stage.

Multiple pathways of information: Routine user support typically is supplied through several avenues. The following are a minimal set:

- A series of web pages that provides users with information about the computer resources and types of support provided and responses to frequently asked questions (FAQ)
- A telephone contact line which connects a user to a support person
- A user support email address to which users can submit help requests

Help users help themselves. Users are often willing to perform routine tasks themselves if they can be provided with the tools. For example, a web site that enables a user to perform a routine function may off-load tasks from the user services support group and provide users a direct way to solve their own issues. As another example, if detailed documentation about hardware and software is made available to users, many HPC users will use such a resource.

Provide very basic support to new and naïve user. The users of gateways and the new HPC users may need handholding initially. The fact that these users need more support than experienced users, but may only use a small fraction of the TeraGrid resources, leads to a challenging problem of how to best allocate of user support.

Show status of user reported problems. Tracking a user services request should take place over the entire life cycle. Users who report problems more likely to be understanding and patient with resource providers as long as they know their problem has not been forgotten. A user support system that communicates activity, or expected wait time, reassures users their issue has not been forgotten. Several commercial providers of time-sensitive services have developed systems for tracking individual requests and resolutions (e.g., Amazon or FedEx) – the best practices from these industries should be examined and adopted where appropriate.

Optimize the most common tasks first. An effective user support system will identify the most common user support requests. These most common interactions should then be optimized. For example, most users will need to request a password reset from time to time. Making password resets a simple procedure is a good place to optimize.

Provide examples. Explanations, particularly written documentation, regarding system usage is often more helpful if working examples are provided. Examples that show working examples of environment files, job submissions scripts, data transfers and other system usages are particularly helpful to reasonably savvy HPC computer users.

Provide an escalation path. All user problems are not equal. An effective HPC User Services program will provide clear procedures that enable users to escalate a problem when it is critical.

3.2 Effective Use of Resources

HPC systems should be well integrated cyberinfrastructure that includes computer processing, data storage, high performance networks, and extensive collections of software. Compounding the challenge of supporting users in such complex computing environments is the fact that HPC system users frequently push the limits of every aspect of an HPC system. Inevitably, HPC system users need help extracting a little more performance out of some aspect of the HPC system.

HPC User Services should provide not only routine user support, but must also help users use the available resource effectively. If the computing environment is heterogeneous, such as the TeraGrid system, such support begins by helping to match users to the most appropriate resource.

An extensive array of User Services offerings are commonly needed to help users make effective use of systems. Training classes, performance optimization tools, performance engineering and application tuning, as well as individual program or project meetings may be provided to help users make effective use of resources.

We distinguish these types of User Services from routine services because effective use of resources typically requires a more sophisticated and experienced support staff. Routine support helps users use the system. Effective use is an optimization step.

3.3 Project- and Domain-Specific Support

In the NSF HPC computing environment, it is in the best interests of resource providers and the scientific users (as well as the NSF, the nation, and the world) for the HPC system to contribute to the development of significant scientific results. An effective HPC User Services program will proactively support new users to encourage their continued use. It will also work with existing productive groups in order to sustain and accelerate their progress.

HPC systems usage is typically restricted to users and groups who request allocations of limited resources. The allocation requests tends to limit the total number of users to a fraction of what typical academic computer systems (at large universities) must support.

In our experience, it is very helpful to users if the HPC User Services program assigns a support person to each new allocation award. Through personal interactions, the support person can quickly build an understanding of the level of support required by the users. Brand new HPC users can be quickly distinguished from very experienced (but new to the system) users.

The TeraGrid computer resources are physically located at several, widely separated, computer centers and there is a large diversity in the computing hardware and software. It is not uncommon for users to gravitate towards a specific system, which they may view as their "home" system. Once familiar with the characteristics of a particular system, users are often reluctant to move to a new site, and relearn the characteristics of a new system because of the difficulty to do so and the optimization of code to get the most from a particular set of system characteristics.

Because the useful life span of an HPC system is only a few years, it is important to help users move from one system to another to take advantage of new system characteristics and capabilities. As in the case of new awards, identifying a user services person with each allocation can be very helpful in easing user transition to new computing environments.

The concept of associating a user services person with multiple projects in the same domain is an appealing idea. Such a person might facilitate communications between groups using similar computing techniques so users can help each other and groups can exchange knowledge and effective practices. Technologies such as standardized data storage formats and visualization technologies are obvious opportunities for technology transfers that can benefit HPC users.

The personal interactions between users and HPC centers are effective in helping users on specific projects with specific deadlines. User Services representatives with responsibility for supporting specific projects can play an effective role in coordinating special system access, such as reservations or large temporary storage needs.

3.4 User-Developed Software Support

Most HPC users write their own software and this has a significant impact on the HPC User Services. In addition to supporting use of existing commercial software tools, HPC User Services are often called on to help users build, modify, debug, and optimize custom software. Furthermore, HPC software is typically parallel software which adds to the complexity and challenge of working with the code.

The consequences for HPC User Services are quite significant. To begin, this implies that many, perhaps most, users must have access to a software development environment, including editors, compilers, linkers, debuggers, and performance monitors, that at least simulates the behavior on the largest systems. Multiple programming languages must be supported. Multiple compilers from multiple vendors are often required. Multiple versions of parallel libraries such as MPICH and OpenMP are often required. Moreover, the software is dynamic and continually evolving.

The most significant implication of this fact may be that a very high level of software expertise is required within HPC User Services. Software engineers with HPC skill sets are difficult and expensive to find and retain. Software specialists who can work effectively with both HPC codes and scientific users across multiple domains are even rarer. Computer scientists who support HPC software development often have doctorates in computer science or other scientific fields.

Granted the specialized skills needed to support HPC software developers, it is important to note that HPC user service groups will typically contain two categories of staff positions. One type of User Services staff will deal with routine user issues, such as resetting passwords. The highly skilled HPC software developers should operate separately from the routine User Services positions. These highly trained and experienced people, while capable of resetting passwords, should not be bogged down with such routine responsibilities.

Software development support by HPC User Services also involves setting appropriate system access to such things as parallel debuggers and queues which support interactive debugging. A significant number of software development tools including source code version management, code profilers, and performance improvement tools must be provided and supported by User Services to support the scientific HPC community. User Services must also support and understand the broad range of scientific domains the HPC community represents.

3.5 Communications, Education, and Outreach

We will combine several valuable HPC User Services into a collective category we will call Communications, Education, and Outreach (CEO).

An HPC User Services program should provide at least three types of communications. First, for an organization like the TeraGrid, User Services should communicate the capabilities of the available computer resources to potential and current users. A second type of communications that should be provided by a HPC User Services will describe the benefits to science research and education, especially by examples of user successes and important results. These types of communication help establish the value of HPC systems to scientific groups, and funding organizations, and are essential to widening the use of HPC systems within and across scientific communities. These communications should be provided in multiple mediums, forums, and forms including print, web, email and others. A third type of communication provides a two-way path for information flow with users, including potential users, and the HPC providers.

Educational and training programs are an important element of an effective HPC User Services program. While earlier we called for personalized interactions between User Services and users, classroom- or web-based, guided, or directed training of users on how to make effective use of HPC systems will be required for larger audiences. The trainers may be User

Services staff. Because many of the TeraGrid resource providers are on college campuses, it may be possible for an HPC User Services group to leverage education and training capabilities from host institutions.

Outreach efforts, like the communications mentioned above, targeted to a desired audience, are an effective way to broaden interest and use of HPC resources. Outreach efforts are essential to involving underrepresented groups in HPC activities, and building interest and establishing the value of HPC results in social, business, and governmental communities. The User Services group should be involved in outreach efforts, and they must be capable of supporting a wide and diverse group of scientists at the routine and more in-depth level once they become new moderate or high-end users, or users of science gateways. Otherwise, outreach efforts pique interests that cannot be satisfied, these new users leave and/or waste their allocation, and their science research, education and workforce development that could benefit from HPC is indefinitely postponed or never done.

Communication from users and potential users to the HPC providers must be facilitated. This naturally includes the system for tracking problem reports and their resolution (see section 3.1), but it also should accommodate the flow of information from sophisticated users. For example, many HPC systems are deployed with a break-in period open to "friendly users", and the knowledge about new systems obtained by these early adopters can be invaluable to both the User Services staff and the other users who will use the resources.

An important function of User Services is to collect, analyze and document the process of providing services to users. Best practices in defining metrics, collecting data, capturing anecdotal information, and synthesizing the information to inform the process and make it increasingly efficient and effective are critically needed. This knowledge also forms the basis for the instructional material to be used to prepare the next generation of user support experts.

4 Social Web and HPC User Services

Up to this point, our comments have focused primarily on HPC User Services capabilities that we have used and found effective. Now, we will comment briefly on untried or immature approaches that may have an important role in future HPC User Services.

Social web phenomena are changing how people with common interests and goals interact with each other. Web 2.0 tools such as blogs, wikis, web forums and other social tools have changed the web from passive "brochure-ware" to a collaborative environment. Tools like email subscription lists, RSS feeds, and news aggregators help people filter an overwhelming volume of information to find information of particular interest. And open-source collaborative software development shows what is possible when distributed groups of technical people work together.

In our view, these technologies, if properly applied, hold significant promise for HPC Users Services. Appropriately deployed and applied, social web tools could help HPC users find each other, identify others with common interests or problems, communicate their interests and specialties, and collaborate on finding answers, resolving problems, or performing research. The danger of improper advice flowing among users is relatively small, but safeguards should be put in place to ensure that authoritative information eventually becomes prevalent (viz. Wikipedia).

While none of us is a specialist in social web technologies, we believe that forward looking HPC User Services should investigate how these new tools might be used by HPC User Services to establish a well-connected, interactive, HPC user community that uses the systems to the fullest. We would like to encourage HPC User Services organizations to evaluate and selectively apply these new tools with the goal of establishing and integrating the User Service group into a dynamic user community that is willing and capable of helping other members of the community.

5 Conclusions

The organization competencies that are needed to construct and operate leading-edge HPC computing resources include many highly technical computer science skill sets. These organization competencies may also require specific knowledge across a broad set of scientific domains. Finally, the organizational competencies that are needed to conduct an effective HPC User Services program include quite different, but complementary, people-oriented skills. The challenge of building an effective User Services program involves assembling an HPC resource provider organization that recognizes, supports, synergizes, and encourages all of these communities within a single organization.

6 Appendix: TeraGrid Science Advisory Board

The TeraGrid Science Advisory Board (SAB) provides advice to the TeraGrid Forum and the NSF TeraGrid Program Director on a wide spectrum of scientific and technical activities within or involving the TeraGrid. The SAB considers the progress and quality of these activities, their balance, and the TeraGrid's interactions with the national and international research community, with the ultimate aim of building a more unified TeraGrid and enhancing the progress of those aspects of academic research and education that require high-end computing. The SAB provides advice on future TeraGrid plans, identifies synergies between TeraGrid activities and related efforts in other agencies, promotes the TeraGrid mission and its activities in the national and international community, and provides help in building and expanding the TeraGrid community. The SAB consists of scientists from the university community, and other laboratories and institutions, as appropriate. The members represent broad scientific disciplines that make use of high-end computation in their research and serve rotating terms of three years in duration. The Chair of the TeraGrid Forum, the Director of the TeraGrid Grid Infrastructure Group, and the NSF TeraGrid Program Director serve as ex-officio, non-voting members. The 2009 members of the SAB are listed below:

Eric Chassignet	Florida State University
Thomas Cheatham	University of Utah
Gwen Jacobs	Montana State University
Dave Kaeli	Northeastern University
James Kinter (chair)	Center for Ocean-Land-Atmosphere Studies George Mason University
Luis Lehner	Louisiana State University
Michael Macy	Cornell University
Phil Maechling	University of Southern California
Alex Ramirez	Hispanic Association of Colleges and Universities
Nora Sabelli	SRI International
Pat Teller	University of Texas, El Paso
P. K. Yeung	Georgia Institute of Technology
Cathy Wu	Georgetown University