J. Towns, NLANR/NCSA
J. Ferguson, NLANR/NCSA
D. Frederick, SDSC
G. Myers, NASA IPG
February 2003

Grid User Services Common Practices

Status of this Memo

This memo provides information to the Grid community regarding support practices for Grid environments. It is a GGF community practice document. It does not define any standards or technical recommendations. Distribution is unlimited.

Copyright Notice

Abstract

As Grid environments develop, a variety of support functions will be needed, analogous to the support functions found in computer center helpdesks, software support organizations, and application development services. This document surveys some of the current and planned practices in some developing distributed environments and recommends practices as appropriate for various elements of the stated support model. In addition, recommendations are provided as to how to best support users and applications in these nascent environments.

This document will require regular review and updating as Grids develop and mature and support requirements change.

This document does not address the support issues for use of specific resources imbedded within the Grid environment nor the entire Grid itself, but addresses the use and support of a particular Grid computing environment. Moreover, although this document does point out the need in various areas to define the security practices to be used in a particular Grid environment, it does not advocate the use of particular policies or technologies to implement those policies.

This document is closely related to another document under development by the Grid User Services Research Group intended to define requirements for services, information and tools in order to enable applications and their support in Grid environments. The Grid User Service Research Group also intends to produce other documents addressing in much more depth many of the issues raised here.

Contents

1 A Support Model

As a basis for outlining the practices, a model of support is given. The elements of this support model are based on the current practices and expected needs for Grid environments. This is a straightforward mapping of a general support model of a modern high-performance computing center to a Grid environment.

1.1 Elements of a Support Model

Here we delineate various elements of a support model. Each of these elements is expanded upon in later sections of this document.

1.1.1 User Information and Tools

Important information resources and tools must be provided in order to enable the use of a Grid environment. These range from basic online documentation to information about the current status of resources in the Grid environment and the Grid infrastructure itself to debugging and performance analysis tools. As on integral part of this, the methods of delivery of these information resources and tools must be included.

1.1.2 Service-Level Agreements

It is important for the organization or collection of organizations providing the Grid environment to appropriately set the shared expectations for the users of these environments and those providing support. A clear statement that accurately delineates these expectations for both the users and support operations in a Grid computing environment is therefore critical.

1.1.3 User Accounts and Allocation Procedures

Currently, all users need to obtain an account and authorization to use specific resources within any Grid environment. Accounts for users typically take the form of logins for individuals on specific resources. This is primarily an artifact of the process by which Grids are being created; they are typically the aggregation of pre-existing resources under sufficiently separate control such that they have had independent processes for establishing accounts. While the umbrella organization providing the basis for establishing the Grid environment helps to unify some of these issues, there are still many implications for users, and, in fact, these processes are still evolving. Processes for these actions must be clearly delineated. In addition, capabilities for account management, both at the PI level and at the resource level need to be provided.

1.1.4 Education and Training

The users of Grids need to be educated and trained in their use. Ideally, such users will not have to learn the individual nuances of using all of the various resources within the Grid environment. In practice, this goal may be difficult to achieve, so the need for instruction on some "local" issues for resources on the Grid will likely need to be maintained. Nonetheless, what is new to the majority of users is the distributed Grid environment and, just as documentation of this is needed, training -- both on line and in person -- is required to develop a user community fluent in the use of the environment.

1.1.5 Help Desk Process

No support function would be complete without the core of staff providing day-to-day assistance to the users of the resources and services available. A well-understood process for the submission and handling of user contacts is required and must deal with the process followed to take a user query from its inception to resolution and address the levels of required support to affect this. This function is typically supported by an effective trouble ticket system.

1.1.6 Support Staff Information and Tools

The support staff must have at their disposal a number of "tools of the trade" and information resources to effectively provide support to the user community. These include a knowledge base to draw upon, information about the status and scheduling of resources and Grid services, tools to assist in the diagnosis of problems reported and appropriate levels of access to resources to operate effectively.

1.1.7 Measuring Success

A support group needs some way to determine success or failure of problem solving and support methods. This is seldom an easy task because it can be largely subjective. While qualitative information is a more useful indicator of the success of the support organization, it is more difficult to get. Frequently, this information can be obtained from various forms of user feedback. Many organizations collect quantitative metrics, which are fairly easy to obtain but say little about the quality of an organization. Effective measures must be in place to advance the support functions. More research needs to be done into methods for developing effective and accurate indicators of the performance of support groups.

1.2 Current Support Models in Use

Included in Appendix A of this document are the descriptions of current and planned practices in currently developing Grid environments. This is certainly not intended to be all-inclusive, but to give a flavor of current activities.

2 User Information and Tools

2.1 Providing and Disseminating Information

A clear need exists to disseminate information to the users and applications developers. Here we outline the information and modes that are seen as most important and most effective.

2.1.1 Types of Information for Users and Support Staff

Frequently, users can make use of various resources to accomplish the task they have at hand, but they need the ability to decide which resources they should target. In addition, this information allows users to know something about the state of the execution of a particular task or set of tasks.

It is equally important, if not more important, for support staff to have access to this information. This allows them not only to assist users in selecting resources but also to determine what has gone wrong when there is a problem with the execution of a task or set of tasks. The following is a list of information considered to be of great importance to make available to users and support staff in order to support the use of Grid environments. The list in not intended to detail all the information needed but rather to give a sense of the types of information. In general, the greatest level of detail possible is required.

The list is divided into two general categories:

- **Quasi-static information:**
  - Grid-connected resource information, software, and Grid services
    - Specification/configuration
    - Access/availability/use policies
  - Infrastructure information
    - Connectivity information between any set of resources
      - Latency/Bandwidth of pipes
      - Feature set (QoS, etc.)
    - Access/availability/use policies
  - Software
    - Availability on resources

- **Dynamic information:**
  - Grid-connected resource information, software, and Grid services
    - Up/down
    - Availability of, or "load" on, a resource
      - Job information (queue status information)
      - Availability interrupts
    - Resource component status information (e.g., disk available, memory free)
  - Infrastructure information
    - Link status (up/down)
    - Current measured available latency/bandwidth/packet loss and so forth.
      - Availability interrupts

All the facets of the Grid environment with which the user will come into contact must be documented at a level that provides an adequate understanding of their function and use. This information changes slowly over time as the environment develops. These are the information resources that users will use to understand how to operate in the environment, how to develop their applications and how to actually use various resources. The following is a representative set of documentation that is required:

- **Access:**
  - Overall Grid environment documentation
  - Applying for an account
  - Obtaining an allocation of resources
    - Management of allocations
  - Security model and authentication processes
  - Service level agreements
- **Application development:**
  - APIs for developing Grid-based applications
    - APIs available (installed)
    - User and reference manuals
- **Software tools:**
  - Debugging tools
  - Performance tools
- **Application execution:**
  - Usage policies and procedures
    - Job submission and monitoring
  - Scheduling and meta-scheduling

2.1.2 Method(s) of Disseminating Information

In recent years, methods of delivery of information to end-users have evolved. It is expected that this will continue to be true in various ways.

The de facto standard for delivery of user documentation in Grid environments, and computing environments in general, has become the Web. The reasons are many, the most compelling of which is that users are using interfaces with Web browsers available as part of the environment to make use of Grid environments. There is little reason to believe that the Web will not continue to be the preferred method of content delivery for this type of information for quite some time. As the Grid computing environment interfaces develop, extensions to this notion will be required. Most notably, wireless devices are becoming more commonplace, and delivery of Web content to these devices requires special considerations. Nonetheless, this is the preferred method. Another significant advantage is the ability to provide search capabilities on the content of each document and across documents. Special considerations should be made in the development of the online materials to support effective search capabilities.

Hardcopy materials come into play in the support of users and applications in two cases. First, some independent software vendors provide documentation only in hardcopy form. Indeed, this situation was historically true for support documentation in computing centers in the past, but the demand for such hardcopy documentation is rapidly decreasing.

Second, some users prefer hardcopy documentation, particularly of reference documents. Hence, it is considered beneficial, though not critical, that indexed, formatted, printable versions of documentation be made available in addition to the on-line forms, when reasonable.

2.2 Portals

Portals are rapidly evolving to be a common interface to Grid-based services for communities of users. These are seen as an important development to further the usability of Grid services and resources.

2.2.1 General Grid Portals

Users, particularly users of distributed environment, often benefit from having access to an interface that executes many of the actions that a user would otherwise have to complete for each individual resource. A general Grid portal also provides a central location for access to the various online information and documentation of interest to the Grid user community, with an integrated presentation. A basic Grid computing environment portal should provide support services to users in two general categories:

- **Information services**
  - Quasi-static and dynamic
  - Accounting and allocation information
  - Helpdesk
  - Training
- **Interactive services**
  - Helpdesk problem submission
  - Knowledge base searching
    - FAQ
  - Web-based access to resources
    - File browsing
    - Job submission
    - Account management
  - Development environments

2.2.2 Application Portals

A number of efforts are under way to build graphical interfaces to applications that, in the past, were accessed and used through command-line-driven interfaces. In particular, there is an increase in development of Web-based interfaces, or application portals. These application portals fall into two general categories: interfaces for specific applications developed within a research group, and interfaces for more broadly used applications such as community codes or ISV applications.

Application portals developed for the use of specific research groups will certainly allow them to be more effective in using resources within a Grid environment. Typically, however, they are useful only to those groups' activities. More general application portals are being developed (e.g., a GAUSSIAN98 portal), and such interfaces should be adopted and made available via general Grid portals for use by interested members of the user community. Not only do such application portals make the use of resources easier, but a well-constructed application portal also typically reduces the number of errors users might make in the process of making use of the applications. Thus, the portal allows the researcher to be more productive and have a better experience, while lowering the impact of supporting such applications.

3 End-User Service Level Expectations

One of the most difficult issues in providing good support and in giving users a good experience with that support is managing their expectations. To complicate matters, currently most users of developing Grid environments have no formal contractual arrangement with the providers of services and support within the Grid environment. As such, there are rarely any well-defined agreements on the shared expectations the users of these environments and those providing support can count on. A clear statement that accurately delineates these expectations for both the users and support operations in a Grid computing environment is therefore critical. The following issues must be delineated for the users:

- Who is supported?
- What is supported?
- When is it supported?
- What the commitment is to acknowledge problem reports?
- What the commitment is to solve problem reports?

Grid user service-level agreements must be arranged among cooperating sites providing services and support within the Grid environment. The establishment of such agreements, through a specific and well-documented mechanism such as a memorandum of understanding, must be part of the generic arrangement among sites, as with security and accounting. Ideally, user accounts should not be authorized without this arrangement and the establishment of the necessary minimum Grid user services infrastructure.

Service-level agreements should delineate user services goals from the user perspective and be agreed to by all participating sites. Areas covered should include the following support services infrastructure:

- Consulting/Technical Support
  - Mechanisms for contacting support
    - Web problem report forms
    - Email
    - Phone contacts during specified times of the day
  - Percentage of user problems to be resolved within x working days
  - Priority to be given to problems not resolved within x working days
  - Mechanisms for users to track problem report

- Documentation – provide accurate, complete information on
  - Grid resources and services
  - Use of the Grid computing environment, particularly resource access and security
  - Software development
  - Software optimization
  - Allocation procedures

- Training
  - Software development for Grid systems
  - Software optimization
  - Software performance measurement

- User Services Performance Metrics
  - User surveys
  - Other user feedback, formal and informal
  - Support contacts / trouble ticket statistics
  - Annual summaries of metrics made available to users

- System Resource and Grid Environment Notices
  - Timely notice of regularly scheduled system downtimes
  - Notice of major system downtimes for upgrades, and the like, “X” days in advance

While this list is not exhaustive, it does provide insight as to the level of detail a common understanding must support.

4 User Accounts and Allocation Procedures

Clearly, the policies under which the Grid environment will operate will evolve, but they must be well defined as early as possible. This section addresses a number of questions and issues prospective users must deal with when trying to work in a Grid computing environment.

4.1 Trust

One of the most difficult issues in dealing with the creation of accounts to access multiple resources in an emerging Grid environment is the establishment of a trust relationship between sites and a formalization of that trust that minimizes impact on the user community. An effective means of accomplishing this trust relationship has been through the establishment of a Public Key Infrastructure (PKI). The establishment of a Certificate Policy (CP) is the basis of these trust relationships and provides for either the creation of a trusted Certificate Authority (CA) or the enlistment of an existing CA for issuance of certificates. Given the common agreement to the

CP, participating sites can reliably accept certificates issued by the CA to allow for authentication to local resources.

Certainly, many other issues remain, but this basis for trust allows many of them to be addressed in a relatively straightforward and logical manner. For example, a PKI does not, by itself, establish a single sign-on capability but does make it possible in a sensible way. In addition, a PKI lays the foundation for trust relationships between Grid environments. This, in turn can allow users access to resources in other Grid environments without needing to go through an account or certificate acquisition process.

We note that this trust relationship does not address the issue of authorization to use a resource. It simply provides a mechanism for authentication.

4.2 Acceptable Use

As users begin to explore the possible resources and services in a Grid environment, they must be guided by a clear acceptable use policy (AUP) for each resource or service or for collections of these. Typically, such statements for the use of resources exist and address issues in the context of an isolated site. These must be reviewed and extended to address the acceptable use of resources and services provided to the grid environment for Grid users.

4.3 Account Acquisition Process

A user must acquire an account in order to be able to access resources within a Grid environment. As Grid environments develop and the policies surrounding access and accounts evolve, the exact processes by which a user obtains an account in any particular environment will change. The consensus is that these environments should develop single sign-on capabilities. Given a PKI for example, it is possible to develop an environment that does not require individual local accounts on the resources to which users have access. Currently, however, users typically must have local accounts at each site participating in a Grid environment.

Thus a number of issues must be clearly documented so that users understand what they are required to do in order to access the resources of the Grid environment. In particular, the mechanics of requesting an account must be clearly defined. The user should be able to obtain an account through a centralized account management system, although it might also be possible to obtain accounts within the Grid computing environment from any of the participating sites individually. Either of these is possible, but they have distinct implications on the account management process within the Grid environment. The policy must be decided early.

Additionally, in some environments separate accounts are created either for individuals or for all users associated with a specific project. The process by which these accounts are created must be well defined.

The user needs to know whether an individual account needs renewal in addition to any possible renewal or extension of an allocation of resources.

No clear “best practice” exists in the area of security requirements in order to establish an account for an individual; indeed, significant variations occur in different Grid computing environments. Issues to be decided upon include the following:

- Do account requests require individual signatures be on file?
- Do account requests require fingerprinting of the individual?
- Are security/background checks required?
- What information is required to perform these checks?
- Does the extent of these checks (level of detail) vary according to citizenship?

4.4 Resource Allocation Process

Significant variations also exist in the processes by which resources are allocated in different emerging Grid environments and are largely based on the historical artifacts of the allocation processes on individual resources. These processes range from allocations based on contractual arrangements, to the assignment of resources purchased for specific purposes, to peer review of requests for resources.

Often there are restrictions on who might request an allocation of resource. For resources in an open environment, allocation is frequently tied to their institutional affiliation. If contractual arrangements exist providing for resources for specific purposes, or if resources have been obtained for specific activities, the mechanisms by which an individual obtains access to an allocation of those resources must be known and documented. If there is a selection or peer review process, the requests and proposal requirements must be stated, along with the criteria against which these requests will be judged. Certainly other mechanisms might exist or come to exist, such as openly providing resources to any paying customer. In any case, the policies and procedures need to be well understood and documented.

Exactly what resources and in what units the allocation of resources is made must be clearly defined. This information will often have implications on how the use of those resources is measured and quantified. If, for example, the use of computational resource not only is measured in the CPU hours accumulated but also includes a memory residency factor, this must be taken into account when requesting resources.

4.5 Allocation Management

The users, and in particular the principal investigators on specific projects, must understand the mechanisms by which the expenditure of allocations of resources is measured. In addition, they must be provided with the tools necessary to obtain information regarding the status of their allocation of resources.

Practice varies in how allocations of resources are charged. Frequently, allocations are decremented either by charging the allocation the amount of resource requested by the user for any particular action or by metering the actual use of resources and charging the allocation the measured amount. In other cases, the allocation of resources is a specified fraction of a particular resource. In these cases, usage can be tracked via either of the previously mentioned processes, but the effort is to balance the use of resources among those assigned fractions thereof.

A mechanism by which users and principal investigators can track the usage of an allocation must be provided. Currently, the most useful way to provide this information is through a secure Web form. This information must be retrieved in real time from a current database of information. Further, it has been found to be most useful to automatically notify users and principal investigators that their allocation of resources has been nearly expended or expired. This is best done with a short series of notifications leading up to the allocations of resources being depleted or the allocation period expiring.

5 Education and Training

Users of Grids will need to be educated and trained in their use. Ideally, such training will mean the user will not have to learn the individual nuances of using all of the various resources within the Grid environment. In practice, this goal may be difficult to achieve, so instruction on some "local" issues for resources on the Grid will likely need to be maintained.

The target audience for Grid user training is the researchers and engineers who wish to use resources on the Grid. This audience would be expected to be somewhat familiar with the basic concepts of computing, and using computers, though the audience could range from secondary school students all the way to highly skilled scientists. While the target audience will have a wide age and knowledge range, its range of skill in Grid technologies is likely to be narrower.

Training for "new" users of Grids will consist of several topics. Key Grid use concepts such as how to access resources, security infrastructure considerations, how to track resource usage, and how to schedule resources will all build on a basic class of how the Grid works.

Support staff will need to be regularly trained about the new resources and services being added to the Grid environment. This specialized training, which will likely come directly from the developers of these new capabilities, should be synchronous ("live") events for all support staff on the Grid who would be able to attend. Archives of such events would be maintained for new support staff. New support staff will also need to make themselves familiar with all the new user training materials available.

New user training will need to be modularized, so that these "new" users can skip concepts with which they are familiar and concentrate on those areas that they need more knowledge. All of these modules could be put together for synchronous ("live") training sessions for true beginners, though the preferred method for delivering this training for new users will be asynchronously via online training modules.

Live training via the classroom, or in combination with a distance learning capability that leverages collaborative technologies such as the Access Grid, will have its place as well. Live events would most often be used to explain new concepts or services. When the new concept or service is captured in an asynchronous training module, live events dealing with that particular concept/service would be cut back.

6 Help Desk Process

Grid user support would be incomplete without the core of staff providing day-to-day assistance to the users of the resources and services available. A well-understood process is required for the submission and handling of user contacts. Specifically, the process must be able to track a user queries from inception to resolution and must address the levels of required support to effect this.

6.1 Query Acquisition and Tracking

User queries, or problem reports, are submitted in a variety of ways. Given that the Grid environment support organization is often distributed, electronic (Web-based) submission of these issues is preferred. This allows for some level of automated triage of the reported problem and more timely resolution. Still, users must have the capability to make contact by phone, by email, or in person. The support staff must then have an interface for entering these contacts for proper tracking.

A necessary piece of the infrastructure required to support Grid user support activities is an effective ticketing system. The ticketing system should be a Web-based helpdesk utility for issuing and tracking support issues or “tickets.” Support staff must be able to access the ticketing system using a standard Web browser, though a database engine of some variety will likely power the system. While commercial helpdesk software is readily available, none tested to date is sufficiently flexible to provide the necessary infrastructure for support in a wide-ranging Grid environment. The ticketing infrastructure needs to be scalable and provide inter-site and inter-helpdesk support capabilities. Indeed, any trouble ticket system therefore must smoothly and seamlessly integrate with any existing trouble ticket system at any collaborating sites.

An additional ticketing system feature, which would make it scalable, is for tickets to be created, assigned, and routed by any consultant. No central authority is necessary. Each consultant may, upon seeing a new email from a user, create a ticket and begin working on the problem. With support staff being geographically distributed, this decentralized workflow model is ideal.

Individual submissions from users or informational submissions from other sources are referenced by ticket number. Either a central "clearinghouse" function or a distributed support model will be found at major computational centers. In the former case, a single point of contact is allowed to create service requests (or tickets); in the latter case, a number of groups are allowed to create tickets within the same tracking system.

Consultants create a ticket after receiving email or a phone call from a user. Tickets are then dispatched or assigned to various groups depending on their local responsibilities. Tickets may, however, be assigned directly to the dispatching group. The ability for consultants or systems staff to assign tickets directly to themselves is an important flexibility that needs to be provided by the ticketing system. Consultants should typically solve user problems without having to route a ticket to a specialized technical group, and so assigning tickets directly to consultants removes a layer from the support infrastructure; a separate staff for routing and assigning tickets is unnecessary.

Queries should be recorded at the time of submission and updated with each addition to the problem status, until the query is resolved. Following is a minimal set of information to be maintained for each ticket:

- User information and submission data user name
    - User id and associated user information
    - Time of submission
    - Nature of problem

- Work logs to journal progress
    - Assignments/reassignments to groups/persons
    - Log all work done
    - Actions done for a service request in the ticketing system

- Resolution
    - Summary of resolution
    - Notification to the user
    - Notification to the ticket creator

While support staffs should be given universal access to the ticketing system, it is usually considered inappropriate to allow users such access. In some cases, there may be some good reasons for allowing this type of access. At the current time, however, no one has resolved the negative issues surrounding giving users this type of access. User access to the ticketing system could provide "status of ticket" information, if not more detailed information on the ticket, but it is frequently ill advised to give users complete access to the ticket system, or even complete access to all information associated with a particular ticket.

All support staff need to have access to all levels of information regarding trouble tickets. Implementations may differ on having restricted access information spaces relied on by the ticket system. Access to data and to what level that access should extend is a policy decision that needs to be decided upon by those managing the relationships creating the Grid environment.

6.2 Problem Resolution

While the details of resolving any particular problem will be heavily dependent on the nature of that problem, a number of practices facilitate this resolution.

6.2.1 Access to User Applications, Code, and Data

The project PI may grant access to private user data space, or individual users may grant more limited access to files. In general, it is assumed that user support staff do *not* have superuser privileges on the systems and environments they support

6.2.2 Access to Implement Changes in the Environment

A policy decision must be made as to the level of access that support staffs have to implement changes in the environment. Frequently, such changes are restricted to system operators and administrators. This must be decided early in the process; if such access is restricted, the process by which such changes are requested and made must be clearly defined.

6.3 Tiered Support Issues and Problem Escalation

It is not uncommon that a problem must be resolved either by the involvement of staff in other parts of the overall virtual organization (frequently crossing institutional boundaries) or by external entities such as hardware or software vendors who might be either on- or off-site from the relevant resource(s) related to the problem. Policy must be defined and implemented within the ticketing system to support the hand-off of tickets to other groups within the overall organization.

In addition, a clearly defined escalation policy is essential. This policy must reflect the support commitment made to users in the Grid environment so as to meet the expectations set. The escalation policy must also be attuned to the management chain of the overall organization and not to any particular organization participating in the Grid environment. We note that the nature of the virtual organization, particularly how tightly coupled the participating institutions are, will have a significant impact on the particular policy developed.

7 Support Staff Information and Tools

Various resources must be available to support staff to help them provide timely and accurate answers to user queries. Two major categories are identified here: information resources and tools of the trade.

7.1 Information Resources

Information resources needed to assist in the determination and resolution of problems may be divided into several categories as outlined below. These resources are used to expand on the expertise support staff members will have garnered over their career.

- **Knowledge base** indicating the expertise of the support staff members. This may be written or may, in the case of smaller organizations, simply exist as a mental construct that is built through time and interaction. In either case, this knowledge base is the primary resource available to individual support staff members. In a Grid environment it will extend beyond the boundaries of any single location, suggesting an electronic version that is regularly updated as members of the support staff gain expertise, staff members come and go, and new technologies requiring support arise.

- **System status**. In traditional support organizations a mechanism usually exists to obtain system status, whether it be a phone call to operations staff or a software mechanism. Slightly more sophisticated mechanisms are required in a Grid environment. Standard methods such as pinging a system, add valuable information for sites that do not provide twenty-four hour coverage. Ideally, the operational support infrastructure of the Grid

environment will provide services that will be made available to Grid user support staff to have a better understanding of the state of the resources and services of the Grid.

- **System scheduling information.** Understanding scheduling policies, activity schedules, and other scheduling information allows the support person to determine conflicts or other sources of problems. In a distributed environment such as the Grid, having access to this information becomes even more important.

- **FAQs.** Assuming there is some form of problem tracking system, the ability to query closed tickets provides a valuable resource for the support staff in identifying frequently asked questions on a variety of topics. Depending on the quality of the answers given for these questions, a good FAQ, beneficial to users and support staff alike, can be developed.

7.2 Problem Determination Tools

In current Grid environments few tools are available to enable support staff and users to identify problems in their program. Current practice is still largely one of isolating a program to a single system where some tools exist, eliminating everything that can be found in that environment, and then experimenting to determine what is wrong in the distributed environment. One of the challenges of the Grid environment will be identifying or developing a new set of tools to provide at least the following capabilities:

- Debugging
- Performance monitoring
- Process/job tracking

Current practice at large centers frequently provides limited or even full root access to support staff in order to enable them to act as the user to identify problems in the user's environment. In the Grid environment this approach becomes more difficult because the Grid may span different organizations with completely different access and security models. We need to understand how to provide access by the support staff to the user's environment. The tools that are currently in use such as **`lsu`**, **`sudo`**, and actual root access will not suffice in the grid environment. This is another topic that will be included in the requirements document.

Most systems and batch schedulers provide some type of logging that support personnel can review to locate the time and cause of a problem. Event-tracking timelines can identify how problems build up to a point of failure. This type of investigation can easily be provided in an environment where a centralized control authority is in place. Here again, the Grid environment presents a challenge because of the decentralized authority and access control. This presents yet another topic for our requirements document.

In a multisystem environment, being able to check on the status of the systems without having to call someone has been found to be very useful, both for support staff and for users. This is no less important in the Grid environment. Several convenient tools have been developed to provide information such as system load, queue status and disk space utilization.

The design and development of a generic Web-based Grid user support portal may provide a useful part of the basic infrastructure for Grid user services. With support for a minimum set of features and with built-in extensibility, such a portal would make it easier for providing user support.

8 Measuring Success

In a good support model, a support group needs some way to determine success or failure of problem-solving and support methods. Quantitative statistics provide some information useful in determining the quality of the support provided. For example, they can measure success in metrics such a mean problem resolution times or adherence to target thresholds such as resolution of a specific percentage of problem reports within a specific number of hours. Another useful measure provided by a statistical review of trouble tickets comes from "repeat events," which can be used to identify areas that need more attention (e.g., in the form of additional or improved on-line information, user training, or additional training for support staff to be better prepared to handle a frequent class of problems).

Qualitative information is often a more useful indicator of the success of the support organization. Frequently, this information can be obtained from various forms of user feedback (e.g., in the form of email or anonymous forms). The biggest issue is that these types of activities require a lot of effort and must be carefully designed. Moreover, a careful balance must be struck in obtaining sufficient evaluation information and having the solicitation of such information become intrusive or disruptive to the users.

Here we briefly review various qualitative techniques for evaluating success.

8.1 Surveys

Surveys provide a periodic, ongoing method for users to provide feedback information. In addition to hardcopy, surveys can take the form of electronic mail to email lists or as Web-based forms. Surveys may be done on the phone and may be a formal list of questions or an informal information-gathering effort. Results of surveys may then be used to improve documentation or services such as training, both for the end user and support staff.

Broad surveys of the users, conducted yearly via on-line forms, can provide an overall sense of the effectiveness of the support activities for the moderate and small users of the Grid environment, who typically are responsible for the majority of the questions and problem reports addressed to the support organization. To reach the heavy users of the Grid environment, direct contact is worthwhile. These users typically have more difficult issues to deal with, and feedback from this group will focus on more advanced support requirements.

8.2 User Groups

User groups may meet in real time (in face-to-face meetings) or as asynchronous virtual communities (via mailing list or Web forums). They can be used to broadly comment on services or focus on specific issues. In-depth discussions on topics of interest to the user community helps to assess the effectiveness of the support organization and provides valuable information about problems confronted in the Grid environment. The larger the organization, the more attractive a face-to-face meeting becomes to better encourage participation by the users. Such face-to-face meetings also allow for conducting training sessions with the user community to keep users abreast of new technologies and developments in the Grid environment.

8.3 Accountability

The ultimate measure of success of a support organization is the productivity of the community of users they support. This is often difficult to measure and the direct impact that the support organization has is frequently even more difficult to assess. Accountability means placing the users needs first and taking responsibility for determining a solution. Accountability should be geared to quality rather than quantity.

9 Summary and Conclusions

Effective user services in a Grid environment are essential for the success of the Grid. We are taught by history that users of computing resources will stop trying to use a resource if they are frustrated while attempting to use those resources. Builders of Grid infrastructures are attempting to make Grids as easy to use as possible, but Grids will not be "simplistic" in their operation for the foreseeable future. A user service organization prepared to respond to the problems of Grid users is key

In this document we have identified the major components of the process of supporting user in a Grid community. This document outlines the common practices known and expected for Grid environments at this time and recommends elements required to provide a solid suite of support services to the Grid user community, based on the current and planned practices in developing distributed environments (see the Appendix). Current practices in user services in a Grid environment will be a moving target in the immediate future, and one can envision several revisions to this document.

10 Author Contact Information

John Towns
National Laboratory for Applied Network Research (NLANR)
National Center for Supercomputing Applications (NCSA)
605 E Springfield Avenue
Champaign, IL 61820
jtowns@nlanr.net
ph:+1-217-244-3228
fx: +1-217-244-2909

11 Intellectual Property Statement

12 Full Copyright Notice

developing Grid Recommendations in which case the procedures for copyrights defined in the GGF Document process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the GGF or its successors or assigns.

13 Appendix A: Current Practices in Support

Below are statements defining the current user support practices of several different organizations. These are given exactly as provided by each of the organizations. We note that the primary source of implementation differences in various support organizations has been rooted in higher-level policy decisions regarding the relationships among the participating organizations or, more commonly, in requirements dictated by organizations or agencies providing the funding to support these Grid efforts.

13.1 The NASA Information Power Grid (IPG) Support Model

The domain of the Information Power Grid Support Model addresses the problem of supporting a computational grid spread over a diverse geographical area involving many independent organizations with autonomous control of a portion of the resources that comprise the Grid at large. This model attempts to address the issues involved for both support staff and the end user of the Grid.

The initial document developed to address this task focused on the procedures necessary to track problems via "trouble tickets" and addressed some of the issues involved with communication between support organizations at different sites. That document was considered to be a near-term solution to be implemented between the few initial participants, Ames, Langley, and Glenn Research Centers, with the understanding that a potentially much larger community of participants would necessarily have to be accounted for in the final model. The experience gained from the first several quarters work with these procedures has led to the IPG Support Model, which is largely based on that initial document.

13.1.1 Primary Approach

The primary approach to this task is based on the fundamental goals of the IPG testbed implementation team:

A. To provide grid functionality without adversely affecting users
B. To provide a set of standards that would make becoming a part of the grid easier
C. To provide an infrastructure that would make adding functionality to the grid easier
D. To design policies that would allow each site to maintain its independence

13.1.2 Underlying Requirements

Due to several factors, the underlying requirements have changed slightly from those layed out in the "Trouble Ticket Procedures" document. One factor is the disparity between levels of support available at each site. Some sites provide minimal user support for this project, ie. system administration and specific task support, while others provide 24x7 and second level support staff. This leads to the realization that in the Grid environment, this is going to be the case. Not all participants will have the resources to provide more than minimal support. At the same time, support for specific aspects of the Grid, ie. Globus, CORBA, Legion, etc. may be provided by a site remote from the user, thereby creating the need to discover the user's request for help in their area.

The primary change to the "model" is the need for a centralized tracking capability, with access from remote support staff. This supports the goal of having decentralized support for system administration and specific topic areas, while at the same time providing a central repository for tracking trouble tickets, and maintaining support information.

Therefore, the following requirements apply:

A. User questions and requests will be centrally tracked using a common problem reporting system. Support will be provided by the group responsible for the problem area regardless of their geographic location. In other words, decentralized support with centralized tracking. (this does not preclude local tracking as well)
B. Deviation from local support procedures will only occur to resolve Grid-related issues. Otherwise, each site retains locally established processes to support their user community.
C. Each site must designate POCs responsible for resolving cross-site issues. In addition, each site is responsible for notifying all participants of any change in POCs.
D. If necessary, user initiated cross-site issues that cannot be resolved by the problem area support group, will be addressed at a collective venue for resolution.

13.1.3 The Support Model

Based on the above information, the IPG Support Model can be outlined as follows:

A. A centralized tracking system, Remedy, currently maintained by the NAS Systems Division at Ames Research Center shall be used to track Grid related user and development questions/problems/issues.
B. Participating Grid sites shall be given access to Remedy, and membership in appropriate support groups.
C. Support for specific problem/question topics shall be distributed based on the organization handling the particular topic.
D. Grid users may make requests for assistance in any of several methods:
    1. By phone to any available Grid Support Organization
    2. By email to any published Grid Support email address
    3. Via the web interface to Remedy (still in development)
    4. Via email from a Grid Support web site
E. All requests shall be entered as tickets in Remedy by the receiving support personnel, specifying the "problem area" appropriate for the call.
F. Local system administration, and support of all non-Grid resources shall be handled by the local staff of the Participating Site.
G. The Remedy system provides means to transfer tickets between groups when needed. Problems requiring escalationie. those that cannot be resolved between support groups, shall be brought to the attention of the group at large via the IPG Engineering meeting, the Support Model meeting, or some as yet to be determined meeting for this purpose.

13.1.4 Records

The Remedy system keeps a record of all transactions for each "ticket" submitted. From these records, any required documentation, metrics, or reports may be generated. Each Participating Site will have access to these records.

13.1.5 Future Plans

The Support Model is subject to changes deemed necessary by the Grid support community. One possibility being considered is to share "tickets" between problem tracking systems that may be in use at the various participating sites. In addition, a web based interface to the Remedy system is being tested for suitability. Access to this data would be made available via the IPG User Portal when it becomes available.

13.1.6 Appendix

This appendix is intended to provide support information for the IPG Support Model. Included are the breakdown of IPG related Remedy Support groups, and the owner of each list; the POC list

for each participating site, along with POC for specific topic areas; and the local support methodology of each participating site.

13.1.6.1 Support Group Breakdown

At the time of this writing, the following is the agreed upon break-down of Remedy Support Groups based on IPG Tasks. Each group has an associated email group to which all appropriate support personel are a member.

Subtask Number	Mail List Name	Name	Owner(s)
2.0	Grid Information Services	infoservices	Judith Utley
8.0	Condor Integration	ipg-condor	Eric Langhirt
9.0	Cluster Integration	clusters	Allen Holtz
13.0	Portal Development	ipg-portal-support	George Myers
18.0	User Guide/Web Documentation	ipg-documentation	George Myers Pam Walatka
23.0	System Testing	ipg-testing	Ray Turney
24.0	CORBA Integration	ipg-corba	Alan Liu
25.0	Legion Integration	ipg-legion	Greg Cates

Other topics shall be added as needed. Some topics not listed here are already supported by existing NAS support groups.

13.1.6.2 POC Lists

Again, each site is to provide a list of designated POCs responsible for resolving cross-site issues and notifying participants of any change in POC. It is up to each site whether this requirement is to be met by designating individuals, mail lists, etc. or by a combination of contact methods. At a minimum each site must provide a POC for local:

- System resources
- Job management system

And, where applicable, local support of:

- Middleware
- Metacomputing Directory Service (MDS)
- Certificate Authority (CA)

For the IPG testbed implementation, the following POC lists have been established:

GRC

User and Direct Technical Support

Sharp Administration: ipg-admin@.grc.nasa.gov
(sharp.lerc.nasa.gov)
Aeroshark Administration: ipg-admin@.grc.nasa.gov
(aeroshark.lerc.nasa.gov) (linux cluster)
CORBA support: corba-support@.grc.nasa.gov
LSF Support: lsf-support@grc.nasa.gov

ICASE

User and Direct Technical Support

```
Pizza Ovens Administration:  larc-sn0@rogallo.larc.nasa.gov
(oven0[0-3].icase.edu)

LaRC
User and Direct Technical Support

Rogallo Administration:  larc-sn0@rogallo.larc.nasa.gov
(rogallo.larc.nasa.gov)
Whitcomb Administration:  larc-sn0@rogallo.larc.nasa.gov
(whitcomb.larc.nasa.gov)
MDS-larc Administration:  larc-sn0@rogallo.larc.nasa.gov
(mds-larc.larc.nasa.gov)

NAS

User Support:  support@nas.nasa.gov
  1-800-331-USER

Direct Technical Support
(Only for peer-to-peer use)

Evelyn Administration:  sn0admin@nas.nasa.gov
(evelyn.nas.nasa.gov)
PBS Support:  ipg-tech@nas.nasa.gov
Globus Support:  ipg-tech@nas.nasa.gov
CORBA Support:  ipg-tech@nas.nasa.gov
Condor Support:  ipg-tech@nas.nasa.gov
MDS-arc Administration:  ipg-admin@nas.nasa.gov
CA-arc Administration:  ipg-admin@nas.nasa.gov
```

13.2 The Alliance Virtual Machine Room (VMR) Support Model

13.2.1 Goals

The first important support goal is providing a single point of access to consulting support. No user should be confused where to start looking for support. This is especially important for later major phases of the Virtual Machine Room (VMR) when a user may not know on which physical machine their job is running, or where their data is physically stored.

The second important goal is that the VMR support network and infrastructure be completely transparent to the user. A scientist experiencing a problem with a VMR resource and prevented from making progress with his science, should only know that he has requested assistance and that a knowledgeable and helpful consultant has responded. They should not have to assist their own support by routing their concern from site to site until finding the most appropriate person to help them.

13.2.2 The Alliance Virtual Consulting Office

Taking into account the desire to provide user support “on demand” during the hours when people actually work, while at the same time recognizing the limitations imposed by the nature of the Alliance and its being geographically distributed, we propose a two-tier model for providing user support for the VMR, and the establishment of the Alliance Virtual Consulting Office (VCO).

The first tier of the VCO would be a 24 by 7, immediate response team trained to provide a "useful" level of support for all VMR resources. The NCSA Consulting Group and the NCSA Technology Management Group (TMG) would together make up this immediate response team, with the consultants providing support for high-performance issues and TMG staff providing operations support. During business hours consultants would answer user questions about compilers, parallel programming, the Globus software infrastructure, and the like. TMG staff would handle questions about system availability, network performance, and related operations issues on a 24 by 7 basis.

It is unrealistic and inefficient to expect support staff to master the details of all VMR resources, and so defining the "useful" level of support provided by the first tier of the VCO is a challenge. NCSA consultants should be capable of answering basic questions about compilers, parallel programming libraries, debugging tools, performance tools, and the like for all VMR resources. NCSA consultants will also be fully trained to provide detailed support for those issues not directly related to a specific site, such as the common interfaces for job submission and access to mass storage. NCSA TMG staff will require training on monitoring the systems and networks comprising the VMR.

The second tier of the VCO would be made up of the support staff from the other VMR sites. Issues that cannot easily be resolved by the NCSA consultants or TMG staff, and which involve particular resources at a site, would be routed to the support staff at that site.

While this model for VMR user support does not provide for expert support available 24 hours a day—and so might not immediately help the user in the eastern U.S. having trouble with a system in Maui—it does provide for some level of support at any time of day. We assume that a quick response from a perhaps "non-expert" support person, followed by a later response from an expert, is more beneficial to a user than no response until much later.

Some Alliance partners not providing computational resources to the VMR may still offer their support staff to contribute to the VMR support effort. It's expected that this type of distributed support would be extremely helpful to the NCSA support groups providing the top tier of VMR support, especially with general issues not relating to the details of a specific VMR resource.

Like any consulting office or "helpdesk", users would be able to contact the VCO in a number of ways: email sent to consult@alliance.edu would route directly to the VCO, a single phone number would be published for those wanting to speak directly with a VCO consultant, and VMR documentation would be available at a single comprehensive web site2. Additionally, a link on the web portal interface would point all users directly to the VCO.

13.2.3 NCSA Ticketing System

A necessary piece of the infrastructure needed to support the VCO is the NCSA Ticketing System (NTS). NTS is a web-based helpdesk utility for issuing and tracking support issue or "tickets." Support staff access NTS using a standard web browser, though a Sybase database engine powers the system. The NCSA Information Resources Group (IRG) designed, coded, and deployed NTS, which has been in production at NCSA for over a year. While commercial helpdesk software is readily available, the NCSA IRG group determined through experimentation and testing that none is sufficiently flexible to provide the necessary infrastructure for support at NCSA and within the Alliance. In addition, the IRG group required that the NCSA ticketing infrastructure be scalable and provide inter-site and inter-helpdesk support capabilities. Indeed, any trouble ticket system for the Alliance VCO must smoothly and seamlessly integrate with any existing system at any of the VMR sites.

Consultants create a NTS ticket after receiving email or a phone call from a user. Tickets are then dispatched or assigned to various groups at NCSA or the Alliance, such as the Systems Group or the High Performance Data Management Group. Tickets may, however, be assigned

directly to the dispatching group. NCSA currently has two dispatching groups: the NCSA Consulting Group and the Technology Management Group. That consultants or TMG staff can assign tickets directly to themselves is an important flexibility provided by NTS. Consultants typically solve user problems without having to route a ticket to a specialized technical group, and so assigning tickets directly to consultants removes an unnecessary layer from the support infrastructure; a separate staff for routing and assigning tickets is unnecessary.

An additional NTS flexibility, making it particularly scalable to the VCO, is that tickets can be created, assigned, and routed by any consultant. No central authority is necessary. Each consultant may, upon seeing a new email from a user, create a ticket and begin working on the problem. With VMR support staff being geographically distributed, this decentralized workflow model is ideal.

13.2.4 Specialized Support for Common Software

The VMR provides the opportunity to develop new ways to support scientific software. The first step is an online database of all scientific software available throughout the Alliance. This software repository will maintain important information about each package or library: included are version level, vendor contact, local coordinator for the package, and a pointer to instructions for using the software at the local site.

We also have the opportunity to provide effective Alliance-wide distributed support for some of these scientific packages and libraries. For example, the chemistry community requires help from Alliance support staff. Often this support is less a question of interacting with the local computing environment, and is more a question of helping the user interact with Gaussian98 (or some other chemistry application) to solve a particular science problem. There are Ph.D.-level computational chemists at most of the Alliance resource partner sites. These computational chemists often have different expertise and differing familiarity with the chemistry packages. Using the NCSA Ticketing System, we are establishing a VMR chemistry support group. Relevant tickets will be dispatched to the VMR chemistry group, and the most appropriate chemistry support staff throughout the Alliance will investigate the tickets. We see this distributed support mechanism as an opportunity to provide even better scientific support to the chemistry community. We anticipate expanding and scaling the discipline-specific distributed support to include math libraries and tools and eventually support of structural engineering and CFD codes.

13.2.5 Supporting The New Technologies

The Alliance VMR and the web portal interface expose users to new technologies like the Globus Metacomputing Toolkit and web portal technologies such as XML. Some fraction of scientists will desire to exploit these technologies directly to further enhance their HPC environment, and will naturally turn to support staff for assistance. While developers might provide some level of support to those scientists extending the VMR using tools the developers have produced, it is unlikely they will have the resources to support not only Alliance VMR users but other grid users as well.

Should then VMR consultants be trained to some level to support VMR infrastructure such as Globus or XML? Most likely yes, all consultants should have some familiarity and be able to provide some level of support, if nothing more than pointing users to the appropriate documentation. A more efficient strategy is to train a group of specialists able to provide specific support for infrastructure pieces of the VMR, in much the same way that professional chemists support chemistry applications across the Alliance, as detailed above. Such specialists might work closely with developers and provide support for users of other grid environments as well. Should “grid computing” become ubiquitous, we expect that commercial companies will be formed to meet the demand for specialized “grid consulting”.

13.2.6 Desktop Data Sharing

Providing support for users accessing the VMR through the web portal interface is especially challenging. Traditionally, users connect to a host using standard tools like telnet and a simple line terminal, and when they encounter a problem they can simply copy the plain text session output and email it to the consultants. With a web portal interface, however, users cannot simply email a copy of what they are seeing. Clearly, providing support for VMR web portal users requires a different approach.

Desktop sharing, also called data sharing or data conferencing, promises to provide the new approach necessary for supporting VMR web portal users. Data conferencing systems allow consultants to interactively work with a user and directly see what the user is seeing. Desktop sharing technology is fast maturing and on some platforms is already ubiquitous, being provided as part of the operating system. The ITU T.120 data conferencing standard is considered robust and is “incorporated” into the H.323 standard for audio, video, and data communications across IP-based networks, better known as internet desktop videoconferencing. Many different videoconferencing products are available from well-established vendors such as PictureTel, Intel, and Microsoft. Each of these solutions provide data conferencing based on the T.120 standard, allowing people to collaborate and share desktops even if they are using systems from different vendors.

Other web based data conferencing solutions exist separate and apart from videoconferencing solutions. Services such as WebEx[9] allow people to data conference “on demand” using a Java enabled web browser and without having to have previously installed or configured dedicated videoconferencing software. Currently WebEx runs on Microsoft Windows, Apple Macintosh, Linux, and Solaris platforms, and user and consultant need not be running on the same platform in order to data conference.

The benefits of data conferencing for providing user support are many. During a data conferencing session a consultant can directly see what the user is seeing, and can easily pickup details and clues that the user might have missed or disregarded as not important. Most data conferencing solutions allow for some level of dynamic interaction so that the parties connected not only see what the other person sees but can also, with appropriate permission, take control and manipulate the remote desktop. In this way it is much easier for a consultant to “become” the user and directly investigate the problem “inside” the user's environment. This approach should significantly reduce the time spent exchanging email, providing access to files, and checking environment variables.

The ability to multicast and have more than two parties data conference allows others like system administrators to join the discussion directly, further enhancing the level of support. Other applications of data conferencing for consultants include direct demonstration of visually enabled software and tools like graphical debuggers. Direct demonstration of such tools is much more efficient and powerful than simply typing instructions out and emailing them. In the future, data conferencing sessions might be recorded and then presented to a user for later reference.

Although data conferencing technologies promise exciting new ways for consultants to support users and in particular VMR users, security and privacy issues inherent when sharing a desktop need to continually be addressed and monitored as the technologies mature.

13.2.7 VMR Tools for Consultants

A necessary part of the VMR software infrastructure will be a set of tools consultants can use to investigate the details of any job running on any particular host. Included should be tools for querying the state of batch queues and batch hosts, gathering detailed process information, inquiring about pending jobs, and other common utilities usually found as part of a job batch system. A single common interface should allow consultants to investigate the details of a job

regardless of what system a job is running on or what batch manager runs on a particular host. In addition tools should be available allowing consultants to query and investigate the details of the global queuing and job routing infrastructure built on top of Globus.

While these types of tools will be provided at some level for all users, it is also helpful for support staff to have “hooks” not necessarily available to general users. These specialized tools might allow consultants access to things such as job submission transcripts, detailed batch manager queries, individual system and network logs, and the like.

13.3 The NPACI Scientific Computing Services Model

The NPACI 2000 Program Plan summarizes our goals in designing/developing and fielding NPACI User Services:

> “…Providing nationally recognized support in consulting, documentation, and training in coordination with partners. Develop measures of customer satisfaction and apply the results of those measures to improving support…”

In addition to the teraflops IBM SP, there are computers from HP, SUN and Cray as part of the computational resources of NPACI. The computer systems designs vary from vector to vector-scalar to various forms of parallel architectures. In addition to these compute servers, there is a powerful and complex archival storage systems, such as HPSS, DMF and ADSM for users. The design of the Partnership is one with a Major Resource Site - SDSC - and Resource Partners – currently, Caltech, University of Michigan and University of Texas, Austin, that provide/maintain mid-range and/or diverse HPC architectures. The Resource Partners are geographically dispersed, but via the Internet and NPACI user-interface infrastructure, constitute the NPACI “distributed machine room" (DMR).

Realization of the DMR is a moving target due to technological advances as well as programmatic changes, and requires not only hardware (machines, network) and software, but also User Support Services designed to provide the necessary consulting infrastructure. The outline of our current NPACI User Support structure is described in this document. The model combines both distributed and centralized resources organized in a manner that attempts to take advantage of the geographic dispersion of Resource Sites.

13.3.1 Requirements

NPACI users must see a uniform User Support Interface, independent of Resource Site, whether SDSC or NPACI Resource Partner. This interface should "shield" the users, as much as possible, from the respective underlying organizational structure (e.g., system and network administration, etc.,) of the various institutions providing compute resources. It is important that user problem reports be assigned and tracked to ensure timely response.

- NPACI User Services should be organized around the concept of supporting the DMR.
- User support services coverage must be available across the entire continental US working day – from 8:00 AM Eastern Time, through 5:00 PM Pacific Time. The burden of support duties will be shared across NPACI Resource Sites
- .Users should what appears to them as one NPACI Contact Point – one phone number, one e-mail address, one web interface.
- NPACI User Services shall consist of:
- User Consulting
  - Help Desk
    - Web-based interface for NPACI users
    - e-mail
    - telephone

- User Training
  - Workshops, seminars, distance-training
- User Contact/Updates
  - Mailing lists
  - Web pages
  - Resource Status Updates
- Common User Environment
  - Security
  - Common Login scripts
  - NPACI User HotPage
- User Documentation
  - Web-based Machine Resource User Guides
- User Allocations and Account Services
- Allocations database
- Applications database
- Resources database
- Support Staff Tools and Services
- Training of support staff
- Tools for accessing/updating allocations database
- Tools for accessing/updating usage database for all production machines
- Methods for Performance Evaluation

13.3.2 The NPACI Support Model

Using the requirements given in the previous section, the Support Model for NPACI User Services can be outlined as follows:

- Remedy Help Desk software is used to provide central user problem report tracking. There are two GUIs associated with Remedy, one for users that provides web-based problem submittal; the other gives NPACI Consultants at all Resource Sites access to user problem reports and the capability to respond, update, assign and re-assign ticket responsibility.
  - Central cgi-bin scripting support for maintenance/update of web interface
  - Central support at SDSC for customization/maintenance/update of Remedy scripts
    - Support includes:
      - Interface for user problem characterization by machine, problem type, etc.
  - Resource Partner sites given access to Remedy, and staff membership in appropriate mailing lists
- Use of common Help Desk software allows tracking of individual tickets as well as maintenance and updating of a ticket database that can be mined for:
  - FAQs
  - Updates for User Guides and other docs
  - Candidate subjects for future Training sessions
  - Summary reports of user tickets
- NPACI Users are able to obtain assistance by one or more methods:
- By web interface to Remedy
- By phone to the Central POC
- All requests shall be entered as Remedy tickets by the answering support staff, identifying the "problem area", etc.
      - By email to Central POC email address
        - All requests shall be entered as Remedy tickets by the answering support staff, identifying the "problem area"
- Common web-site for access to all NPACI services
  - www.npaci.edu – initial web POC

- User Training
  - Workshops, seminars, distance-training are provided by SDSC and (optionally) the Resource Partners
  - Web-based materials, examples, etc.
- User Contact/Updates
  - Mailing lists
    - NPACI users normally subscribe to npaci-news for general updates on machine status, workshops, etc.
  - Web pages
    - npaci-news mailings are archived on separate web page
  - Resource Status Updates
    - "Live" resource status provided by NPACI HotPage
- Common Distributed User Environment
  - NPACI User HotPage
    - Provides seamless uniform user environment
    - Secure User logins (available soon)
    - Developing complete user interface functionality (available later this year)
  - Security - common security infrastructure set up at all Resource Sites - ssh and/or kerberized logins required
  - Common Login scripts
    - Common Unix login/shell scripts provided to give users uniform environment
- User Documentation
  - Web-based Machine Resource User Guides -common format at NPACI level
    - Local User Guides available at Resource Partners
- User Allocations and Account Services
  - Complete information on application process available online
  - All forms available online
- Web-enabled Allocations database for use by Support staff
- Web-enabled Applications database accessible to users, providing links to documents, etc.
  - Web-enabled Resources database accessible to users with complete, up-to-date resources descriptions
  - Account database updated daily allows accounts to be monitored
- Support Staff Tools and Services
- Web-enabled Allocations database for use by Support staff
    - Query tools for providing queries for specific user accounts, as well as summary information
- Inter-site issues resolved by NPACI Resource Working Group (RWG) composed of key staff from all Resource Sites
- System administration, network administration, etc., is handled by staff at the respective Resource Sites
- Tools for accessing/updating usage database for all production machines
  - Query tools for providing queries for specific user accounts, jobs, as well as summary information
- Training of support staff
- User Feedback Methods
  - User Survey(s)
    - Annual NPACI User Survey
  - Annual All Hands Meeting Sessions on user issues
    - Incorporated several suggestions from this years session
  - User Advisory Committee

13.3.3 Future Plans

As user needs change and hardware and software technologies advance, so too, must User Support. Annual evaluation partly based on critical user input from the yearly NPACI User Survey, NPACI User Advisory Committee and annual NPACI All Hands Meeting (AHM), will be used to identify areas for improvements/changes.

13.3.4 Mailing Lists

This appendix summarizes NPACI mailing lists with user services related functions.

List Name (@npaci.edu)	Function	Membership
npaci-consulting	General notices to NPACI Consultants	Consulting staff at Resource Sites
consulting-affiliates	Contacting Academic Associates institutions	AA consulting staff with local HPC support staff
npaci-news	General notices of interest to NPACI users	All NPACI users
npaci-services resources-wg	Discussions related to NPACI resources inter-site issues	Selected staff from Resource Sites
training	Notices for NPACI Training coordinators	Training Coordinators at Resource Sites

13.4 The Department of Defense (DoD) Aeronautical Systems Center (ASC) Major Shared Resource Center (MSRC) Support Model

13.4.1 Goals

The ASC MSRC is one of four MSRCs in the DoD High Performance Computing (HPC) Modernization Program (HPCMP). Each MSRC has been designated to provide a complete, robust HPC environment to DoD Science and Technology (S&T) and Developmental Test and Evaluation (DT&E) users. The MSRC environment includes a full range of resources such as hardware, software, data storage and archiving, scientific visualization, high speed networking interfaces to the Defense Research and Engineering Network (DREN), a supporting infrastructure, and expertise in computational and computer sciences and HPC systems. The MSRCs provide the largest share of HPC support to the DoD community.

Selection criteria used to select the four MSRC sites included the following:

- Impact on DoD Research and Development (R&D) goals and benefits to S&T and DT&E Programs,
- HPC experience,
- Existing HPC infrastructure,
- Personnel,
- Proactive user services,
- Physical facility,
- Site and Service/Agency management commitment,
- Cost efficiency and leveraging,
- Ability to support classified and unclassified processing,
- Ability to satisfy immediate requirements, and
- Ability to complement existing DoD HPC centers.

The MSRCs support the ten Computational Technology Areas (CTAs) that have been identified as major thrust areas for DoD R&D and DT&E. They are:

- Computational Structural Mechanics (CSM),
- Computational Fluid Dynamics (CFD),
- Computational Chemistry and Materials Science (CCM),
- Computational Electromagnetics and Acoustics (CEA),
- Climate/Weather/Ocean Modeling (CWO),
- Signal/Image Processing (SIP),
- Forces Modeling and Simulation/C4I (FMS),
- Environmental Quality Modeling and Simulation (EQM),
- Computational Electronics and Nanoelectronics (CEN), and
- Integrated Modeling and Testing (IMT).

The goals of the ASC MSRC include:

- Establishing "worldclass" capabilities that apply high performance computation toward solving DoD problems.
- Ensuring military advantage and warfighting superiority on the 21st century battlefield through the use of high performance information technologies.
- Strengthening national prominence and preeminence by advancing critical technologies and expertise in high performance computing.

Additional information is available at our public website: http://www.asc.hpc.mil.

13.4.2 Support Model

The ASC MSRC has a Service Center based on a three tier system. It is similar to the medical triage system with the assignment of priorities. The most critical problems are screened and the proper level of technical expertise is applied. The first tier (Help Desk), receives and resolves requests whenever possible. The desk is considered the collective "receptacle" and "clearing house" for the ASC MSRC. It resolves approximately 70-80% of the service requests received. If the service request cannot be resolved at this level, then it is sent to the second tier which involves the Technicians and Application Managers. The third tier consists of the System Analysts, Vendors, and Academia. Upon resolution the service ticket is returned to the Help Desk, an email is sent to the requestor, and the service ticket is CLOSED.

13.4.3 Systems Management and Reporting Tool

The Systems Management And Reporting Tool (SMART) is the central tool being used at the Aeronautical Systems Center (ASC) Major Shared Resource Center (MSRC) at Wright Patterson Air Force Base. SMART is used to coordinate the four separate functions performed at the ASC MSRC: application processing, allocation and utilization of system time, service center tracking, and inventory. Each of the four subsystems is separately named to eliminate confusion and aid in controlling access to each system. These subsystems are described in detail in the sections below.

13.4.3.1 Application Processing System (APS)

In order to receive access to the ASC MSRC you must fill out an application. All applications are processed through User Services using the Application Processing System (APS). APS tracks the personal information about an individual, such as their name and address, as well as the security information, such as passwords, SecurID card numbers and login names. An e-mail module has been developed to automatically send correspondence to new and potential users based upon what actions the User Services personnel perform. For example, once an application

has been accepted a Welcome letter is sent to the new user alerting them of their login name and the rules for using the ASC MSRC. Separate mailings are done for passwords and other protected information. Logs are kept for each email that is sent.

13.4.3.2 Machine Allocation and Utilization Database (MAUD)

Once an application has been accepted, a user is given an allocation: an amount of time they may use each system. Their allocation time and system utilization are tracked via the Machine Allocation and Utilization Database (MAUD). MAUD enables the system administrators to increase or decrease allocation time, receive reporting on the amount of time allocated for each system, and automatically mail reports to other individuals. MAUD also enables system administrators and selected users to view system utilization, run utilization reports and see statistics on time allocated vs. time utilized for given periods of time.

13.4.3.3 Service Request System (SRS)

All service center tracking is done using the Service Request System (SRS). SRS assigns each new service request a unique number. Once a service request has been entered, e-mail is automatically sent to the user who initiated the service request. The e-mail details the service request and prompts the user to contact the helpdesk if the problem has not been accurately reported. Help Desk staff will then assign a technician to resolve the service request. Another e-mail is automatically sent to the technician detailing the service request and providing information about the user, such as their name, login name and email address. All e-mail correspondence is logged into the database and can easily be retrieved by querying on the service request number.

Automatic e-mails are also sent when a problem is resolved. A copy of the resolution is e-mailed to the user along with a request that they “test” the resolution within 24 hours. If a user does not dispute the resolution, the service request is closed automatically after 24 hours of being resolved.

The SRS provides extensive tracking capabilities including:

- tracking all communication between the user and Help Desk (face-to-face, incoming phone calls, outgoing phone calls, e-mails, etc.),
- providing statistics on the amount of time a service request is unresolved,
- detailing the types of service requests received (hardware, software, unable to logon on, etc.)

providing statistics on how frequently the problem is resolved without a technician being assigned.

13.4.3.4 The Purchasing and Records Keeping System (PARKS)

With all of the different hardware and software being used at the ASC MSRC it was necessary to design an inventory module. The Purchasing and Records Keeping System (PARKS) is designed to track an item from purchase request to packing slip. This module is used as a “receiving” module – it is not designed to create purchase requests or purchase orders, but to log them once they have been received as inventory.

13.4.4 Points of Contact (POCs)

ASC MSRC POCs
MANAGERS, SUPPORT, TECHNICIANS AND SYSTEM ADMINISTRATORS

SECURITY

NETWORK SECURITY:

MSRC ENVIRONMENT SECURITY:
SECURITY MANAGERS:

USER SERVICES CONTACTS

USER SERVICES MANAGEMENT:
(Employment, Tours, User Services Issues, SW Acquisitions, etc.)
SECRETARIAT:
(Configuration Control Boards, Remote Installation Petitions, Memorandum Of Understandings, Software Working Group, etc.)
ACCOUNTS CENTER:
(Project Tracking, User Accounts & Access, NAC/Clearance issues)
SERVICE CENTER (1-888-677-2272):
(Management, Company Policy Issues, etc.)
(Message Of The Day (MOTD), Kerberos Support, User Problem Solving & Guidance, Password Resets & Unexpires, "Frontline" for User Services)
(Government PEM, Policy Issues, S/AAA for Internal Accounts)
(Accounting, Reports, Charts, etc.)

SYSTEMS SUPPORT

SYSTEM ADMINISTRATION MGT:
(Systems Metrics, System Administration Issues, HW Acquisitions, etc.)
O2K:
(Application Analyst(s))

(System Administration)
(Accounting Questions)

IBM:
(Application Analyst(s))

(System Administration)

COMPAQ:
(Application Analyst(s))
(System Administration)
SCI-VIS:
(Management/System Administration)
(Application Analyst)
(System Administration)
(Government PEM)
SS1:
(All Problems)
SUN:
(All Problems)

MSRC GENERAL SUPPORT

OPERATIONS:
(Management, System Administrative Liaison, Ops Policies and Controls, etc.)
(System Ops, System Ops Messages, Machine Monitoring, Queue Monitoring, After-Hours Building Monitors)
(UPS Power Backup Unit Systems Monitor, Building Access, etc.)
LOCAL WORKSTATION SUPPORT:
MSRC SOFTWARE INSTALLS:

WTS:
(Client)
PC SUPPORT:
(PC Repairs)
(Equipment Assignment/Upgrade)
PBS:
(Administration/Accounting Problems)

HAFS:
(All Problems)

KDC SERVER:

KERBEROS:
UAS:
(System Administration)
NETWORK:
(Troubleshooting, Tracing, Routers)

(DREN Specialist, Classified Network)
ARCHIVE (MSAS):
(*All Problems*)
(User Support POC)
BACKUPS & FILE RESTORES

SPT09:
(Licensing)

SAS*(Mail)*/PRINT SERVERS:
TADE DEVELOPMENT ENVIRONMENT:
ODBS1:
(Oracle/Database, Service Tickets, Desk & Accounts Screens)
(*System Administration*)
(*Web Applications*)

WEBSITE:
(*Webmaster*)
(*System Administration*)
(*MOTDs*)
(Content/Sustainment)
PET TRAINING:
(Classroom System Administration)
(Classes/Registration)
(NT System Administration)

13.4.5 Software Supported

The ASC MSRC maintains over 100 commercial-off-the-shelf software products for use in a variety of disciplines. Visualization, analysis, and programming tools are available for use with the CTAs of CCM, CEA, CEN, CFD, and CSM. Each of these products is installed, maintained, and managed by our on-site staff of application managers. These packages are managed through the ASC MSRC Configuration Control Board.

13.4.6 Future Plans

The ASC MSRC is part of the DoD HPCMP Computational Grid Initiative. The goal of this effort is to create within four years an operational and stable DoD HPCMP Computational Grid (HPCMP Grid) comprising a variety of distributed computing, storage, and visualization resources. The establishment of the HPCMP Grid will be accomplished by close collaborations among various DoD HPCMP Shared Resource Centers (SRCs) and by involving real end-users at various stages of the initiative so that the HPCMP Grid is fully oriented around the users. In addition, the initiative will leverage the experiences and extensive knowledge gained by other major computational grid projects to build better computational tools that will permit the DoD scientists and engineers to collaborate and share information and resources. The requirements that are identified will be tested in a testbed environment to insure that the functional needs of the user and the SRCs can be met using the available technologies.

The HPCMP Grid will be deployed in three distinct phases. The first phase will be a prototype phase developed outside the production environment with local staff performing end-user activities. In the second phase, a testbed grid will be built within a production environment with both staff and pioneer users working together so that the infrastructure is oriented around the end-users, and is easy to use. The pioneer users will be a selected set of end-users who will also act as advocates to help promote the ease-of-use of the fully functional grid. The third and final phase will expand the testbed into a production environment. At this stage, the HPCMP Grid is operational and stable, and fully accessible to all DoD HPCMP users