

CSBase: A Framework for Building Customized Grid Environments

M. J. de Lima, C. Ururahy, A. L. de Moura, T. Melcop
M. N. dos Santos, B. O. Silvestre, V. Q. dos Reis
C. Cassino, R. Cerqueira
Tecgraf/PUC-Rio
Rio de Janeiro, Brazil
csbase@tecgraf.puc-rio.br

Grid middleware provides general mechanisms to integrate resources and to present them to their end-users. This is a convenient feature of the current grid middleware, since it allows its (re)use in different situations. However, this general approach usually involves complex user interfaces. CSBase is a framework that supports the development of customized grid systems, presented to the end-user through a friendly user interface that abstracts the complexities of a grid computational environment.

All CSBase facilities are presented to the end-user through a project-centric web interface, which represents a virtual desktop where data files are organized per project. Through this desktop, the user can install *algorithms* — domain-specific non-interactive programs that run on remote *execution hosts*. When a user installs a new algorithm, it is dynamically integrated to the environment. CSBase builds the user interface through which the algorithm's parameters are provided, with no need for recompilation.

CSBase desktop allows users to request the execution of a selected algorithm on a remote host. Monitoring facilities allow users to monitor an algorithm execution and, if needed, to interrupt it. Users can also monitor CPU and memory usage. The collected information can be presented both textually and through several types of graphs and charts. Clusters and host groups can be monitored as a whole. Based on the presented information, the user can select an adequate machine for the execution of an algorithm. This selection can also be delegated to the system, which will base its decision on information gathered from the execution hosts.

The user of a CSBase instance organizes her working area by creating *projects*: hierarchical structures that store related user files. By arranging her work in projects, the user is able to define different working spaces. A user project can be either private to its creator or shared by different users. CSBase also offers a publishing facility through which an individual user's file can be made available to other users.

An important feature of the framework is its access control mechanism. Within a CSBase instance, storage areas, algorithms, and execution hosts are protected against unauthorized access. User access rights are defined by specific permissions; the system administrator is responsible for creating those permissions and assigning them to individual users or user groups. Through the system administration interface, the administrator can perform tasks such as adding new users to the system, creating and assigning access per-

missions, and allocating storage space to individual projects.

CSBase environment can spread over several geographic locations. Each of them maintains a CSBase Server, sharing resources locally. A central CSBase Server is responsible for administrative tasks, maintaining a global administration data repository.

One of the most important goals of CSBase is to facilitate the development of domain-specific applications and to enable their integration to the grid. The end-users of these applications can thus benefit from all CSBase built-in facilities, customized to their own needs. This means that an application can access and modify user projects and their contents, start the execution of a specific algorithm, or make use of any other internal service. The core of the CSBase framework is an infra-structure layer for client/server foundation. At the client side, CSBase provides classes to develop new applications. At the server side, CSBase offers a wide range of services and allows the creation of new ones.

CSBase resulted from projects developed by Tecgraf/PUC-Rio (a laboratory from PUC-Rio's Computer Science Department) in partnership with Petrobras (Brazilian Oil Company). One of these projects, WebSintesi, integrates applications that support the analysis and processing of seismic data. These applications demand a high processing power, handling seismic data files that comprise hundreds of gigabytes. WebSintesi's first version was installed about two years ago. Nowadays it has more than 1,000 users, about 30 algorithms, and several execution hosts in different platforms. It has a central server located in Rio and 6 servers located in different Brazilian states. Connected to the central server are 8 execution hosts: 5 multiprocessed IRIX machines, 1 SP2 with 16 nodes, and 2 beowulf Linux clusters running PBS, one with 1,008 processors and the other with 250 processors.

Currently, we are investigating the evolution of CSBase in many directions, including better support for grid orchestration and scheduling mechanisms. CSBase is also adopted in CSGrid, a grid workbench for scientific applications in advanced networks and is about to become the first multi-institutional usage of CSBase. CSGrid is available for download at <http://www.tecgraf.puc-rio.br/csbase/csgrid>.

We would like to thank the support and funding of Petrobras/Brazil, RNP (National Education and Research Network) and the CSBase development team.