A GRID PORTAL BASED TOOL FOR THE DEPLOYMENT AND MANAGEMENT OF GRID INFRASTRUCTURES

By

Fernando José Cintrón-González

A project submitted in partial fulfillment of the requirements for the degree of

MASTER IN ENGINEERING

in

COMPUTER ENGINEERING

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS

2007

Approved by:

Nayda G. Santiago, Ph.D Member, Graduate Committee	Date
José F. Vega, Ph.D Member, Graduate Committee	Date
Wilson Rivera, Ph.D President, Graduate Committee	Date
Silvestre Colón, M.S. Representative of Graduate Studies	Date
Isidoro Couvertier, Ph.D Chairperson of the Department	Date

Abstract of Project Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Master in Engineering

A GRID PORTAL BASED TOOL FOR THE DEPLOYMENT AND MANAGEMENT OF GRID INFRASTRUCTURES

By

Fernando José Cintrón-González

2007

Chair: Wilson Rivera
Major Department: Electrical and Computer Engineering Department

The deployment of grid infrastructures is a time consuming and arduous task, therefore, the goal of this work was to establish an easy, “out of the box”, grid portal environment that allows administrators to go through installation/configuration procedures as fast possible. This project comprehended the development and integration of a set of administrative tools into a grid portal environment. It was designed to automate and help grid administrators in the deployment and configuration of services on the grid. The work conducted here shows to be highly expandable, opening opportunities for new emerging grid configuration tools through grid portals and contributes as the starting point for future developers.

Resumen de Proyecto Presentado a Escuela Graduada
de la Universidad de Puerto Rico como requisito parcial de los
Requerimientos para el grado de Maestría en Ingeniería

UNA HERRAMIENTA PARA EL MANEJO E INSTALLACION DE INFRAESTRUCTURAS "GRID"A TRAVÉS DE UN PORTAL "GRID"

Por

Fernando José Cintrón-González

2007

Consejero: Wilson Rivera
Departamento: Departamento de Ingeniería Eléctrica y Computadoras

El proceso de instalación de una infraestructura "grid" es una tarea ardua que consume tiempo, por lo cual el principal objetivo de este proyecto fue establecer un ambiente sencillo, fácil y listo para usar, que permite a los administradores llevar a cabo los procesos de instalación ágilmente. El proyecto abarcó el desarrollo e integración de un grupo de herramientas administrativas en un portal "grid", diseñadas para automatizar y asistir a los administradores en la configuración de los servicios en el mismo. El trabajo realizado puede ser ampliamente adaptado, abriendo oportunidades para la creación de nuevas herramientas de configuración emergentes a través de portales "grid", a la vez que contribuye sentando las bases para futuros desarrolladores.

I want to dedicate this work to my family, who always has been there for me, giving me support in my life decisions.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor Wilson Rivera for his guidance and support toward the completion of this project, and for always showing confidence in me.

I appreciate the effort of Kennie Cruz who was involved in all the development aspect of this project, and the contribution of other PDC Lab members.

My especial thanks to all my friends, especially Laura E. Sánchez, Angel Villalain, Nestor Heredia, and Pablo Robles who always were there for me during my graduate studies.

This material is based upon work supported by the Puerto Rico Development Company (PRIDCO) through the Center for Industrial Software Development (InduSoft).

TABLE OF CONTENTS

LIST OF FIGURES

Figure page

LIST OF ABBREVIATIONS

OGSA	Open Grid Services Architecture.
SOAP	Simple Object Access Protocol.
WSRF	Web Services Resource Framework.
WSDL	Web Services Description Language.
JSP	Java Server Pages.
OGCE	The Open Grid Computing Environments.
SILC	Service Integration Life Cycle.
JSR	Java Specification Request.

CHAPTER 1
INTRODUCTION

Grid computing [1] involves coordination, storage and networking of resources across dynamic and geographically dispersed organizations in a transparent way for users. The Open Grid Services Architecture (OGSA) [2], based upon standard Internet protocols such as SOAP (Simple Object Access Protocol) and WSDL (Web Services Description Language), is becoming a standard platform for grid services development. Operational grids based on these technologies are feasible now, and a large number of grid prototypes are already in place (e.g. Grid Physics Network (GridPhyN)[1] and Teragrid[2] among many others).

The deployment of grid infrastructures is a very time consuming, and demanding activity that requires detailed knowledge of protocols and grid component interactions. Grid portals, on the other hand, hide this complexity via easy-to-use interfaces, creating gateways to computing resources. An effective grid portal provides tools for user authentication and authorization, application deployment, configuration and application execution, and management of distributed data sets.

1.1 Contribution of the Project

Recently, there have been significant advances in grid portal technologies and development of scientific grid interfaces [3, 4]. Grid portals have been used as web

[1] GridPhyN web site: http://www.griphyn.org

[2] Teragrid web site: http://www.teragrid.org

interfaces that allows users to access grid facilities through a web browser. However these do not address the need of grid administrators at the time of grid deployment.

It is a fact that the grid deployment processes is a time consuming and arduous task. Therefore, the goal of this project is to establish an easy "out-of-the box"grid portal environment that allows administrators to go through installation and configuration procedures as fast as possible.

This project makes use of an existing grid portal environment that allows the aggregation of portlets interfaces. Making use of portlet technology and combining it with configuration tools, we are setting the bases for developers to continue aggregating configuration tools which will lead to a more robust and fast deployment process.

1.2 Document Organization

This document is organized as follows: Chapter 2, discusses concepts and tools related to the development and configuration of grid infrastructures and grid based-applications. Chapter 3, provides a description of the proposed grid administration tool kit. Chapter 4, is intended to get future developers to have an understanding of the development process, and how to enhance this project. Finally, conclusions and future work are presented in Chapter 5.

CHAPTER 2
CONCEPTS AND TOOLS

In this section we describe the basic concepts and tools related to the development and configuration of grid infrastructures, grid portals and grid-based applications. First, we describe the Globus toolkit as the deployment platform. Second, we introduce the portlet technology that allows the creation of dynamic content and grid portals. Third, we discuss two grid portal frameworks, namely Gridsphere and OGSE. Finally, we discuss a number of grid management middlewares.

2.1 Globus Toolkit

The Globus Toolkit[1] is a community-based, open-architecture, open-source set of services and software libraries that support grid applications [2, 5]. The toolkit includes software services and libraries for resource monitoring, discovery, management, and additionally provides services for security and file management.

The Grid Security Infrastructure (GSI) supports single sign-on and delegation. Single sign-on is a mechanism to allow users to authenticate once by creating a proxy credential. This credential is used by client programs to authenticate with any remote service on the user's behalf [2]. Delegation is the process that allows the creation and communication of delegated proxy credentials to a remote service. By delegating proxy credentials, remote services can use them to act on the user's behalf. For user authentication, GSI employs the standard of X.509 user certificates.

[1] GT4 web site: http://globus.org/

With the use of the Globus Myproxy credential manager server (later discussed on section 2.4.1), users can obtain proxy certificates extensions of the X.509 user's certificate.

The grid infrastructure needs to deal with the transfer and management of distributed data. Globus provides two components for the movement of data: the GridFTP and the Reliable File Transfer (RFT) services. The GridFTP is composed of a server and command line client tool to transfer data fast and securely. The Reliable File Transfer is a service that provides the functionality of a “job scheduler”to transfer data, combined with the GridFTP protocol. With RFT the transfer state of the data is stored in a reliable storage. If a failure occurs, the service will have access to the state where the transfer was just before the failure occurred and retake the transfer process where left. Besides these two components, Globus provides the Replica Location Service (RLS) to keep track of replicas of files on the grid. RLS is implemented as a server where users or services register files, at creation time, and later make queries to find where the replicas are located. RLS can be deployed as a centralized server or as multiple servers. The later case will increase the scalability of the service, making possible to store more replicas as compared to using just a single centralized server.

Globus provides two services to manage resources: MDS and GRAM. The Meta Directory Service (MDS) provides information discovery and system configuration access, which includes server configuration, network status, and replicated datasets locations. The Grid Resource Allocation Management (GRAM) is a protocol that provides reliable, secure remote creation and management of arbitrary computations [2]. GRAM implements mechanisms to submit, monitor, and cancel jobs on grid computing resources. GRAM provides job-management service to monitor and control the job life cycle by communicating with the available local resources schedulers. It makes uses of Globus existing infrastructures, like GSI and GridFTP , to provide security and data management.

2.2 Portlet Specification

In an effort to establish a standard to guide developers in the creation of web portal and portlet applications, Sun Microsystems Inc. and IBM Corporation collectively lead the request for the Portlet Specification, today known as the Java Specification Request (JSR 168)[6]. This document defines the java portlet technology, its environment and the related concepts. It describes in detail all the portlet functionality, where they coexist, and how they are used on portals. Next, we will give a review of some of the concepts defined in the JSR 168.

2.2.1 Portals

A portal is a web based application that allows the integration of content from different sources within a web page [6]. It provides a personalized environment that allows remote authenticated users, through single sign on, to access services and resources of a certain organization. One of the most attractive feature of portals is the presentation of different set of portlets within a single web page. Portals can be user-oriented in the way that the same portal can present different set of portlets for different users.

2.2.2 Portlets

Portlets are a Java technology web components managed by a portlet container, which is in charge of process requests and generate dynamic content [6]. They are used in portals to present user interfaces. In a portal context, portlets represent the link between web users and services. Portlets also can be user-oriented in the way that a given portlet can present different information for different users. For example the GridFTP portlet, which allows the transfer of a file between two given computer nodes, displays a list of files located in the computer nodes that correspond to the web authenticated user's account only.

Java portlets are similar to java servlets, but in contrast, portlets can not send redirects or errors to browsers directly, forward requests or write arbitrary markup to an output stream. Another difference with servlets is that portlets rely on portal specific infrastructure functions such as access to user profile information, and also they depend on an environment maintained by the portlet container.

2.2.3 Portlet Container

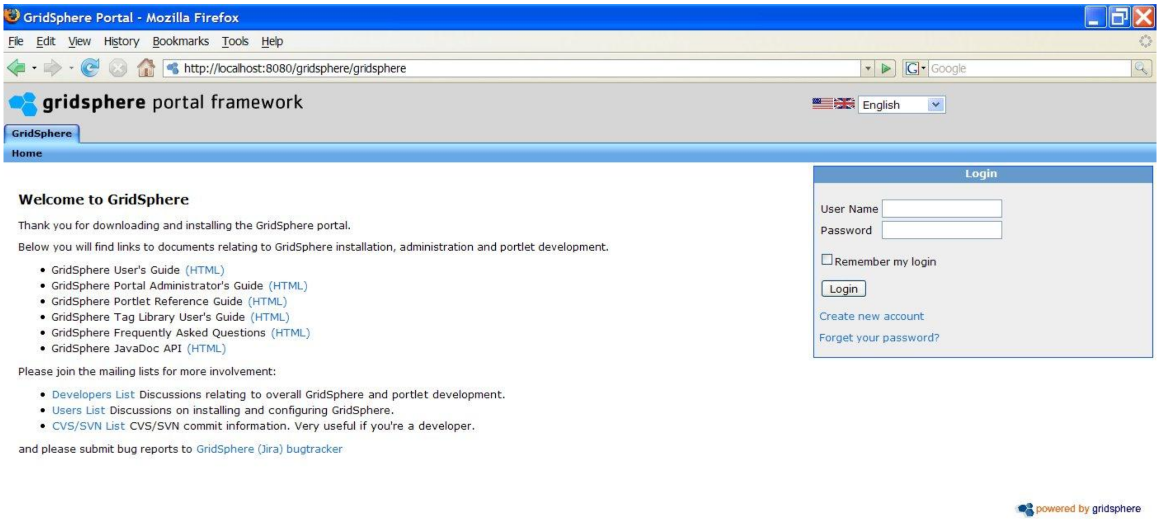
A portlet container provides a runtime environment for portlets, according to the java portlet specification. In this environment portlets are instantiated, used and destroyed. The portlet container is not a stand-alone container like the servlet container, instead, it is implemented on top of it. It reuses the functionality provided by the servlet container.

2.3 Gridsphere

GridSphere[2] is an open source project, designed to help developers quickly set-up a portal environment [7]. It combines both, the portlet container and the portal framework, which only leaves to portal developers the task of customizing and editing the presentation layer, as shown in figure 2–1. Based on the JSR 168 portlet specification standard, developed with the java programming language, and runs on the Tomcat servlet container. Gridsphere comes with its own JSP user interface tag library, allowing developers to easily design and deploy new portlets.

Gridsphere enables developers to easily deploy and package new portlet web applications, to be launched and administered within its own portlet container. Adding more portlets enhances the portal functionality and at the same time makes the grid portal fit according to the grid infrastructure, and address user needs.

[2] GridSphere web site: http://www.gridsphere.org


GridSphere Portal - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:8080/gridsphere/gridsphere
Google
gridsphere portal framework
English
GridSphere
Home
Login
User Name
Password
□ Remember my login
Login
Create new account
Forget your password?
powered by gridsphere

Welcome to GridSphere

Thank you for downloading and installing the GridSphere portal.

Below you will find links to documents relating to GridSphere installation, administration and portlet development.

- GridSphere User's Guide (HTML)
- GridSphere Portal Administrator's Guide (HTML)
- GridSphere Portlet Reference Guide (HTML)
- GridSphere Tag Library User's Guide (HTML)
- GridSphere Frequently Asked Questions (HTML)
- GridSphere JavaDoc API (HTML)

Please join the mailing lists for more involvement:

- Developers List Discussions relating to overall GridSphere and portlet development.
- Users List Discussions on installing and configuring GridSphere.
- CVS/SVN List CVS/SVN commit information. Very useful if you're a developer.

and please submit bug reports to GridSphere (Jira) bugtracker

Figure 2–1: Gridsphere portal

As a framework it allows the full administration of the portal with a built-in support for Role Based Access Control (RBAC) separating users into: guests, users, admins and super users. Having these groups of users permits to establish different access rights depending on the user roles inside the portal. One attractive feature is the independent user accounts configuration, which promotes an environment in which administrators controls "who access what" in the portal context.

As a portal, it is accessed through a web browser, giving mobility to users, enabling them to work from anywhere without the need of installing additional third-party software on their machines.

2.4 OGCE

The Open Grid Computing Environment (OGCE)[3] is a project that fosters collaborations and sharable components among portal developers worldwide. One of its tasks is the development of a set of portlets that facilitate the set up of a basic Grid Computing Portal. This set of portlets is contained in the OGCE release

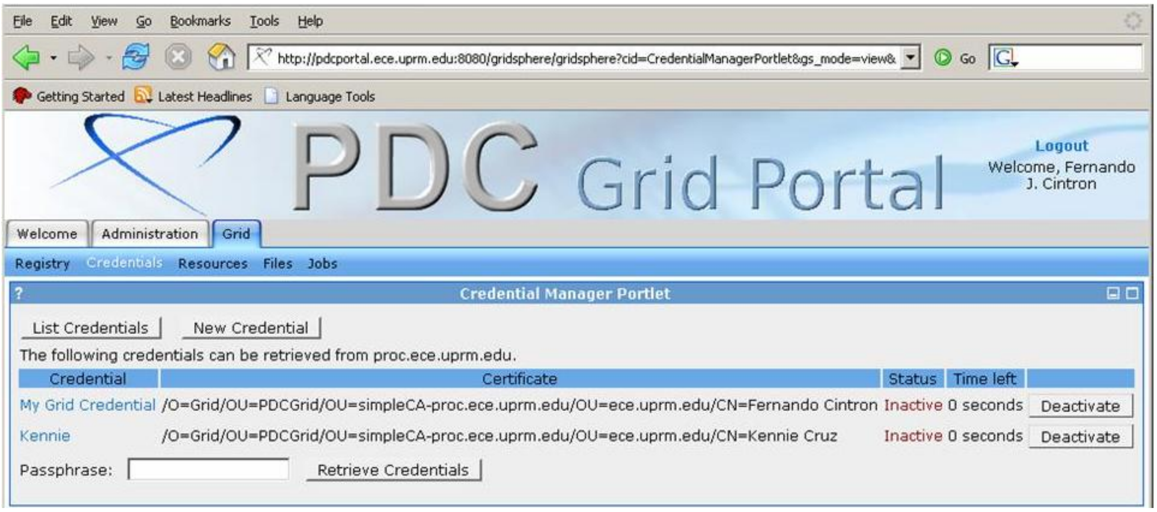
[3] OGCE web site: http://www.collab-ogce.org

2, which gives support for MyProxy credential management, Grid File Transfer Protocol (GridFTP), job submission through Grid Resource Allocation Manager (GRAM)[4] , among others.

In the next sub-sections we present descriptions of the portlets included in the OGCE release 2, and how they are related to grid services.

2.4.1 Credential Manager

The Credential Manager Portlet enables a user to retrieve a grid proxy credential from the Myproxy credential manager server, figure 2–2.



The following credentials can be retrieved from proc.ece.uprm.edu.

Credential	Certificate	Status	Time left	
My Grid Credential	/O=Grid/OU=PDCGrid/OU=simpleCA-proc.ece.uprm.edu/OU=ece.uprm.edu/CN=Fernando Cintron	Inactive	0 seconds	Deactivate
Kennie	/O=Grid/OU=PDCGrid/OU=simpleCA-proc.ece.uprm.edu/OU=ece.uprm.edu/CN=Kennie Cruz	Inactive	0 seconds	Deactivate

Passphrase:

Figure 2–2: Credential Manager portlet

MyProxy [8] is a service provided by the Globus Tool Kit 4, that stores user credentials[5] . Users or Grid administrators can store their credentials in the MyProxy service, and then login (obtain a proxy credential) on any system on the network.

[4] GRAM is a service provided by Globus Toolkit 4

[5] MyProxy web site: http://grid.ncsa.uiuc.edu/myproxy/

MyProxy is used for user authentication on grid portals, enabling authenticated users to work with grid resources remotely. A MyProxy server must be installed and configured on a designated host server, in order to expedite Proxies to users.

2.4.2 File Transfer

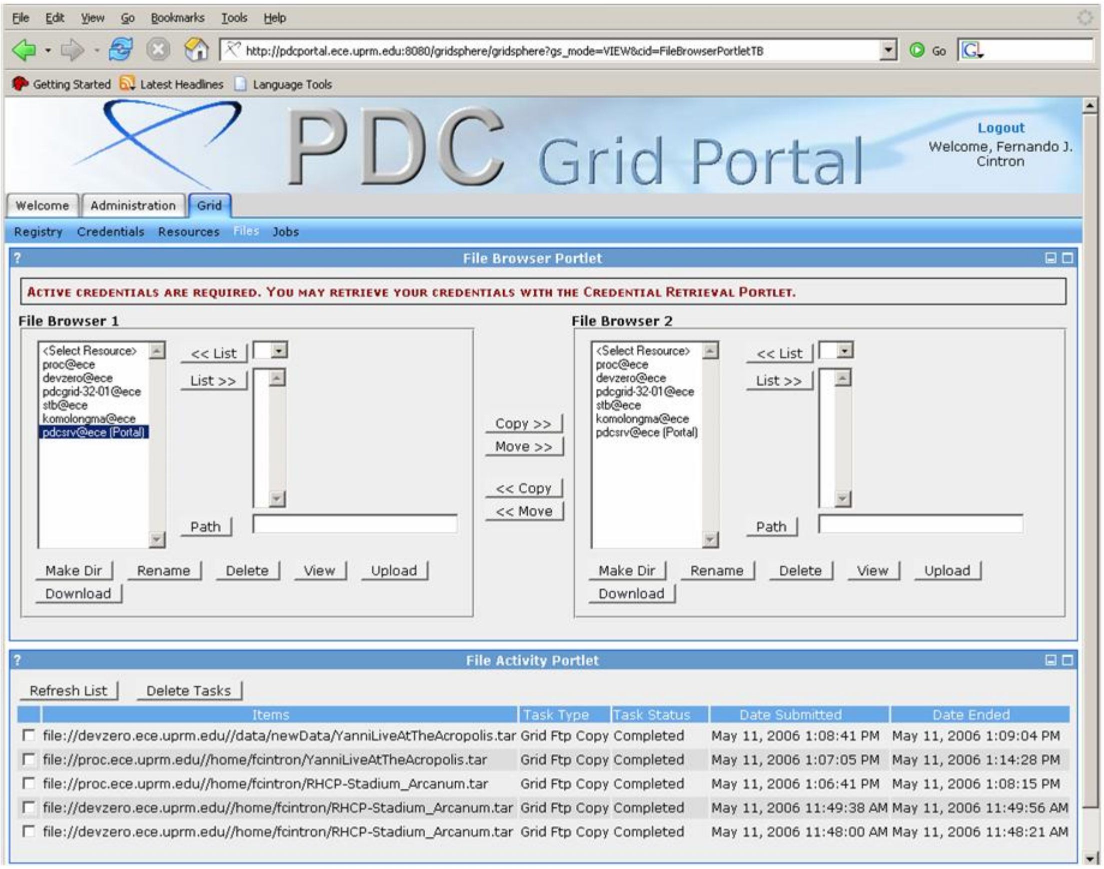
The File Transfer portlet makes use of the Grid File Transfer Protocol (GridFTP), which is part of the services included in the Globus Toolkit 4. It has a user friendly interface, figure 2–3, that allows users to browse their files on a remote machine via the GridFTP service. This portlet allows users to perform third party file transfers, meaning that users can select two independent machine nodes and carryout file transfers between them, and monitor the operation from the machine the user is accessing the portal. Users also have the option to download from and upload files to the remote machine.

2.4.3 Job Submission

The job submission portlet allows the portal user to submit jobs to remote resources through the Globus GRAM (Grid Resource Allocation Manager) service. The portlet allows to specify the job parameters, submit it, and watch the job state information, figure 2–4.

2.4.4 Dependencies

All of the above mentioned portlets from section 2.4 need to be configured prior to be used. They all have something in common, they need to have information of which are the resources (grid nodes) on the grid that provide their target service. The Credential Manager portlet need to know how to reach Myproxy host server. The File Transfer needs a list of all the nodes containing the GridFTP service, and so the Job Submission needs to know which is the GRAM host server. This configuration is achieved through a file called “Resources.xml”, in which the grid administrator establishes the host servers and the services they provide. The OGCE group of portlets displays the Resources.xml file through another portlet where the



ACTIVE CREDENTIALS ARE REQUIRED. YOU MAY RETRIEVE YOUR CREDENTIALS WITH THE CREDENTIAL RETRIEVAL PORTLET.

	Items	Task Type	Task Status	Date Submitted	Date Ended
☐	file://devzero.ece.uprm.edu//data/newData/YanniLiveAtTheAcropolis.tar	Grid Ftp Copy	Completed	May 11, 2006 1:08:41 PM	May 11, 2006 1:09:04 PM
☐	file://proc.ece.uprm.edu//home/fcintron/YanniLiveAtTheAcropolis.tar	Grid Ftp Copy	Completed	May 11, 2006 1:07:05 PM	May 11, 2006 1:14:28 PM
☐	file://proc.ece.uprm.edu//home/fcintron/RHCP-Stadium_Arcanum.tar	Grid Ftp Copy	Completed	May 11, 2006 1:06:41 PM	May 11, 2006 1:08:15 PM
☐	file://devzero.ece.uprm.edu//home/fcintron/RHCP-Stadium_Arcanum.tar	Grid Ftp Copy	Completed	May 11, 2006 11:49:38 AM	May 11, 2006 11:49:56 AM
☐	file://devzero.ece.uprm.edu//home/fcintron/RHCP-Stadium_Arcanum.tar	Grid Ftp Copy	Completed	May 11, 2006 11:48:00 AM	May 11, 2006 11:48:21 AM

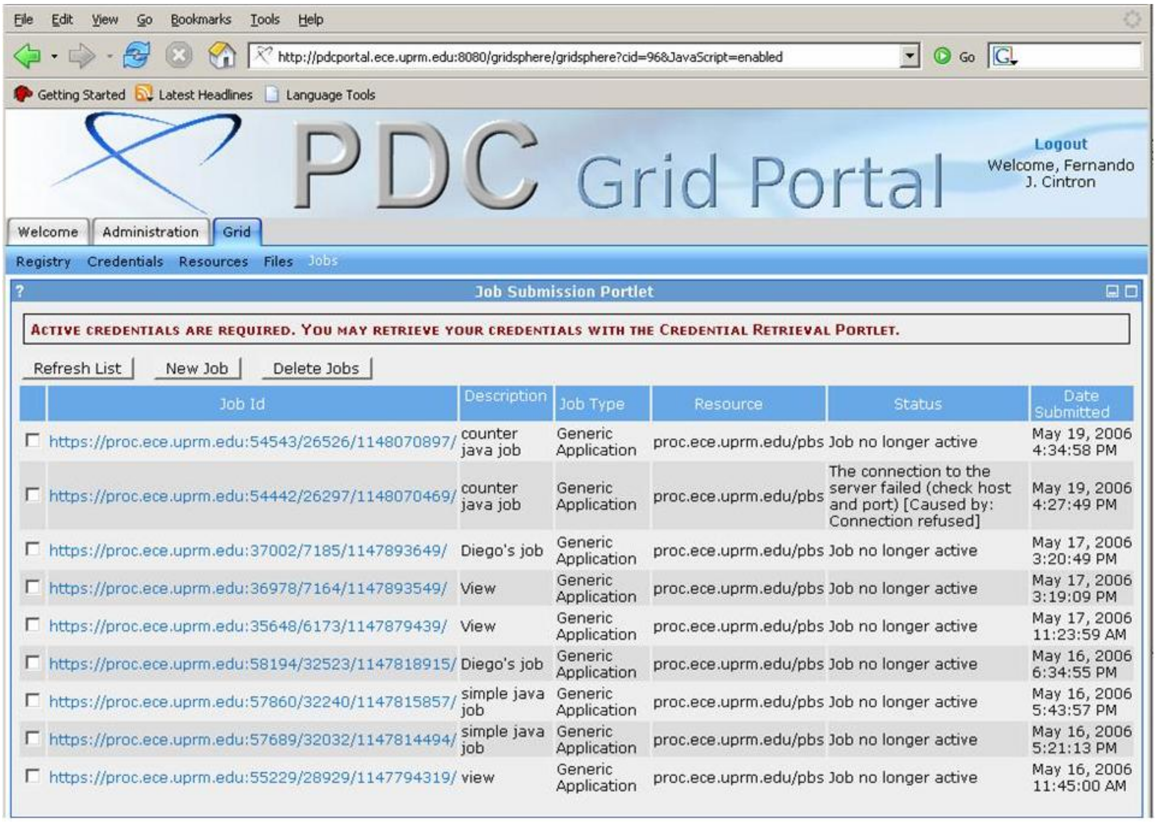
Figure 2–3: File Transfer portlet

administrator can view the grid nodes and services each one provides (relying on the resources file), figure 2–5.

Not only the portlets need configuration, also the grid services do. Grid services need to be deployed (installed and configured) on the grid nodes. Currently this is a task that is addressed manually by the grid administrator, “node by node”.

2.5 Related projects

As new computing problems emerge, more researchers face the need of tools that enable them to conduct research. Dealing with long installations and configuration tasks, may delay the main purpose of the research. This situation have awaken the interest of researchers and developers to produce solutions at a reachable level for people with other expertise areas. Some work has been focused on the


File Edit View Go Bookmarks Tools Help
http://pdcportal.ece.uprm.edu:8080/gridsphere/gridsphere?cid=96&JavaScript=enabled
Go
Getting Started
Latest Headlines
Language Tools
PDC Grid Portal
Logout
Welcome, Fernando J. Cintron
Welcome
Administration
Grid
Registry
Credentials
Resources
Files
Jobs
?

Job Submission Portlet

ACTIVE CREDENTIALS ARE REQUIRED. YOU MAY RETRIEVE YOUR CREDENTIALS WITH THE CREDENTIAL RETRIEVAL PORTLET.

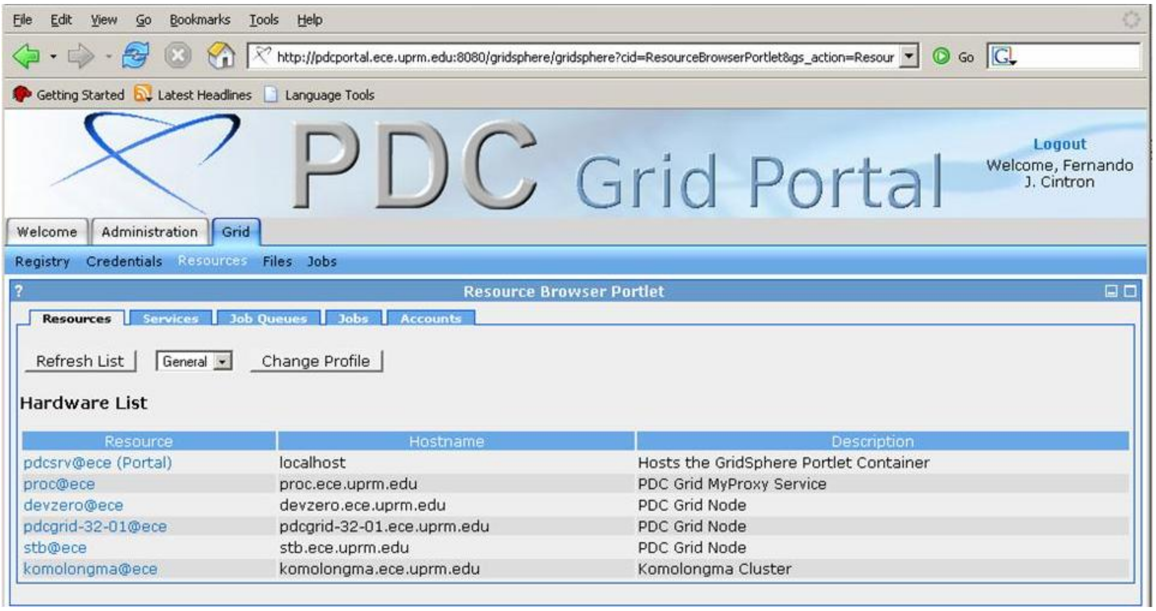
Refresh List | New Job | Delete Jobs

	Job Id	Description	Job Type	Resource	Status	Date Submitted
□	https://proc.ece.uprm.edu:54543/26526/1148070897/	counter java job	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 19, 2006 4:34:58 PM
□	https://proc.ece.uprm.edu:54442/26297/1148070469/	counter java job	Generic Application	proc.ece.uprm.edu/pbs	The connection to the server failed (check host and port) [Caused by: Connection refused]	May 19, 2006 4:27:49 PM
□	https://proc.ece.uprm.edu:37002/7185/1147893649/	Diego's job	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 17, 2006 3:20:49 PM
□	https://proc.ece.uprm.edu:36978/7164/1147893549/	View	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 17, 2006 3:19:09 PM
□	https://proc.ece.uprm.edu:35648/6173/1147879439/	View	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 17, 2006 11:23:59 AM
□	https://proc.ece.uprm.edu:58194/32523/1147818915/	Diego's job	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 16, 2006 6:34:55 PM
□	https://proc.ece.uprm.edu:57860/32240/1147815857/	simple java job	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 16, 2006 5:43:57 PM
□	https://proc.ece.uprm.edu:57689/32032/1147814494/	simple java job	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 16, 2006 5:21:13 PM
□	https://proc.ece.uprm.edu:55229/28929/1147794319/	view	Generic Application	proc.ece.uprm.edu/pbs	Job no longer active	May 16, 2006 11:45:00 AM

Figure 2–4: Job Submission portlet

creation of grid portals. Wang et al [9] focuses on the design aspects of a successful Grid portal implementation, and recognizes the existence of portals for different targeted audiences. Similarly, [10] exposed how grid portals can serve as hosts for grid applications from the web service perspective providing custom and dynamic user interfaces.

However, the work found related to the management of the grid deployment process is limited and scarce. Similar projects that deal with the management of existing grid services are found. GCM [11] for example, is a tool with the goal of hiding heterogeneous aspects from end-users and developers by going through the necessary tasks for the management of computational job submission in heterogeneous grid environments.



Hardware List

Resource	Hostname	Description
pdcsrv@ece (Portal)	localhost	Hosts the GridSphere Portlet Container
proc@ece	proc.ece.uprm.edu	PDC Grid MyProxy Service
devzero@ece	devzero.ece.uprm.edu	PDC Grid Node
pdcgrid-32-01@ece	pdcgrid-32-01.ece.uprm.edu	PDC Grid Node
stb@ece	stb.ece.uprm.edu	PDC Grid Node
komolongma@ece	komolongma.ece.uprm.edu	Komolongma Cluster

Figure 2–5: Grid Resources portlet

The Grid User Management System (GUMS) is a Grid Identity Mapping Service [12]. Identity Mapping is a mechanism to identify users on a Grid when no Grid credential is used natively. In these cases, the GRID credential (e.g. Certificates) for each incoming job must be associated with an appropriate site credential (e.g. user accounts). The GUMS server is in charge of performing the GRID credential mapping with the site credential and communicates it with the gatekeepers. GUMS also provides mechanisms for a client gatekeeper, to query the state of the GUMS server. GUMS is a mechanism that address management of the authentication/security on the grid.

Ali et al [13] recognize the substantial advance of grid computing, but also make a remark on the existing barriers for its adaptation on the “day-to-day“ computer environment. Their approach is to migrate grid computing into what they define the Grid Operating System (Grid OS). They conclude that the Grid OS approach overcomes problems such as transparent resource discovery, “grid enable” desktop applications, provides enough security, be network aware, and interoperability with existing grid middleware. However, they recognize that for such interoperability with grid middleware some interfaces are required.

The related work presented here goes from the use of grid portal as a prominent user interface for grid infrastructures, the use of tools to configure single services, a service for the user's credential mapping, to the migration of grid computing into a Grid OS architecture. Although these works address certain needs, there is a demand for a set of administrative tools that go through the grid middleware deployment process.

CHAPTER 3
A GRID ADMINISTRATION TOOLKIT

This project focuses on the development and integration of a set of administrative tools into a grid portal environment. This chapter provides a technical description of our grid testbed resources, a description of the proposed grid administration tools and its components, and finally, a discussion on how this set of administrative tools are integrated into a Grid In-A-Box (GIAB) distribution, developed at the Parallel and Distributed Computing Laboratory (PDCLab) at the University of Puerto Rico, Mayaguez campus.

3.1 PDC Lab Grid testbed

The computational resources available on the grid test bed (figure 3–1) include an IBM xSeries Linux cluster with 64 nodes, dualprocessor at 1.2GHz, 53GB of memory and 1TB of storage; eight (8) IA-64 Itanium servers, dual processor at 900 MHz, each with 8GB of memory and 140GB of SCSI Ultra 320 storage; two (2) IA-32 Pentium IV servers, dual processor at 3 GHz, each with 1GB of memory and 120GB of ATA-100 storage; one (1) IA-32 Pentium III server, dual processor at 1.2 GHz with 2GB of memory and 140Gb of SCSI Ultra 160 storage; one (1) IA-32 Xeon server, dual processor at 2.8 GHz, L2 Cache 1MB with 1GB of memory and one 230 GB RAID of storage (STB Server); and two (2) PowerVault storage with 8TB.

3.2 Grid Portal Specifications

The PDC Lab grid portal, figure 3–2, is based on the Gridsphere portal/portlet container framework, mounted over an Apache Tomcat servlet container on a Linux

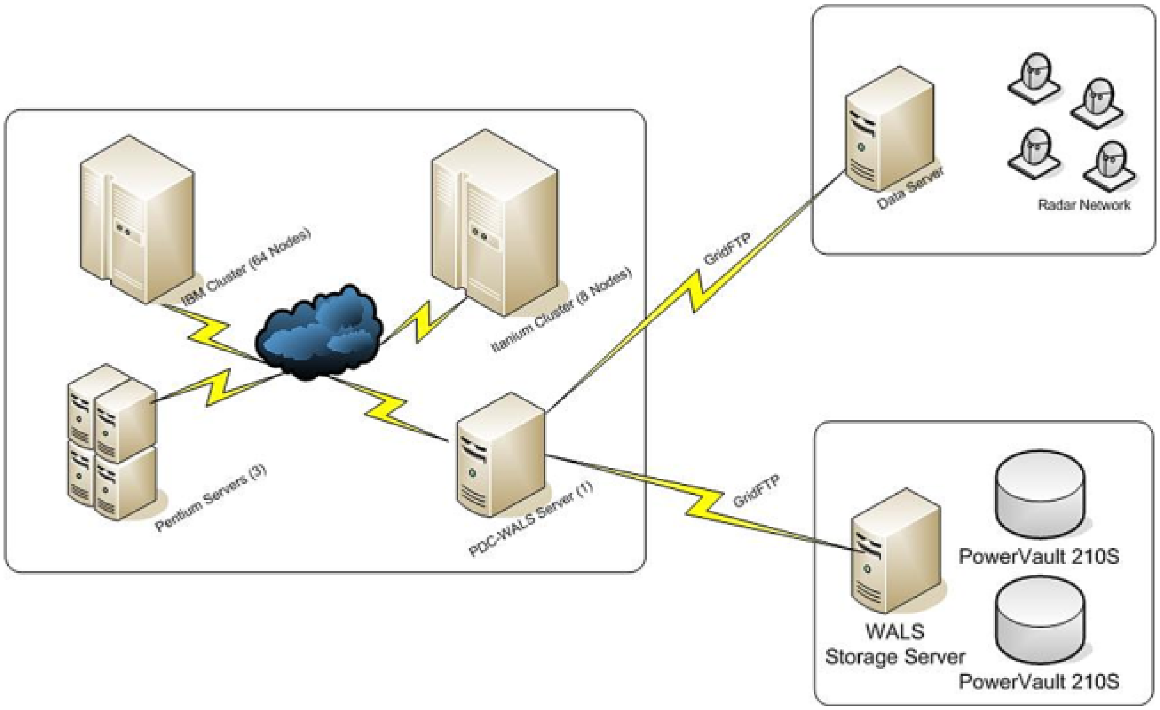


Figure 3–1: PDC Lab Grid Test Bed

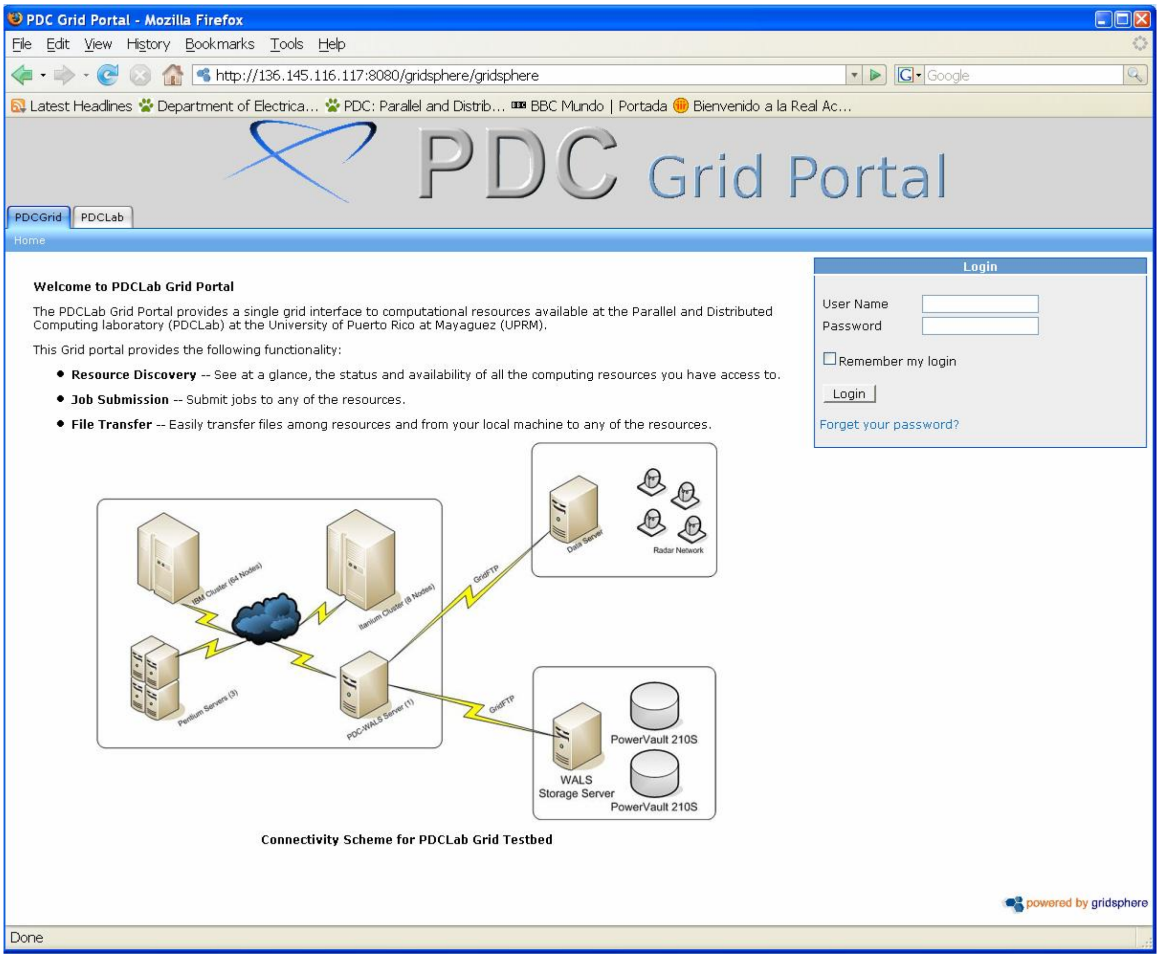
CentOS 4 configured machine. It is developed using Java SDK 1.4.2 release 9 and deployed with Ant 1.6.5.

3.3 GIABLinux

Grid In A Box Linux (GIABLinux) is a system installation package developed, at the PDCLab. It is based on Linux CentOS 4 and all the necessary services to build a grid based on the GT4. It comes with our pre-configured grid portal and allows the administrators to access the System Admin Tools to make further deployments and configurations of services.

3.4 System Administration

The administrative tools are integrated to the grid portal interface allowing administrators to work within the portal. The set of administrative tools can be viewed as a group of portlets placed together under a group in the grid portal.



Welcome to PDCLab Grid Portal

The PDCLab Grid Portal provides a single grid interface to computational resources available at the Parallel and Distributed Computing laboratory (PDCLab) at the University of Puerto Rico at Mayaguez (UPRM).

This Grid portal provides the following functionality:

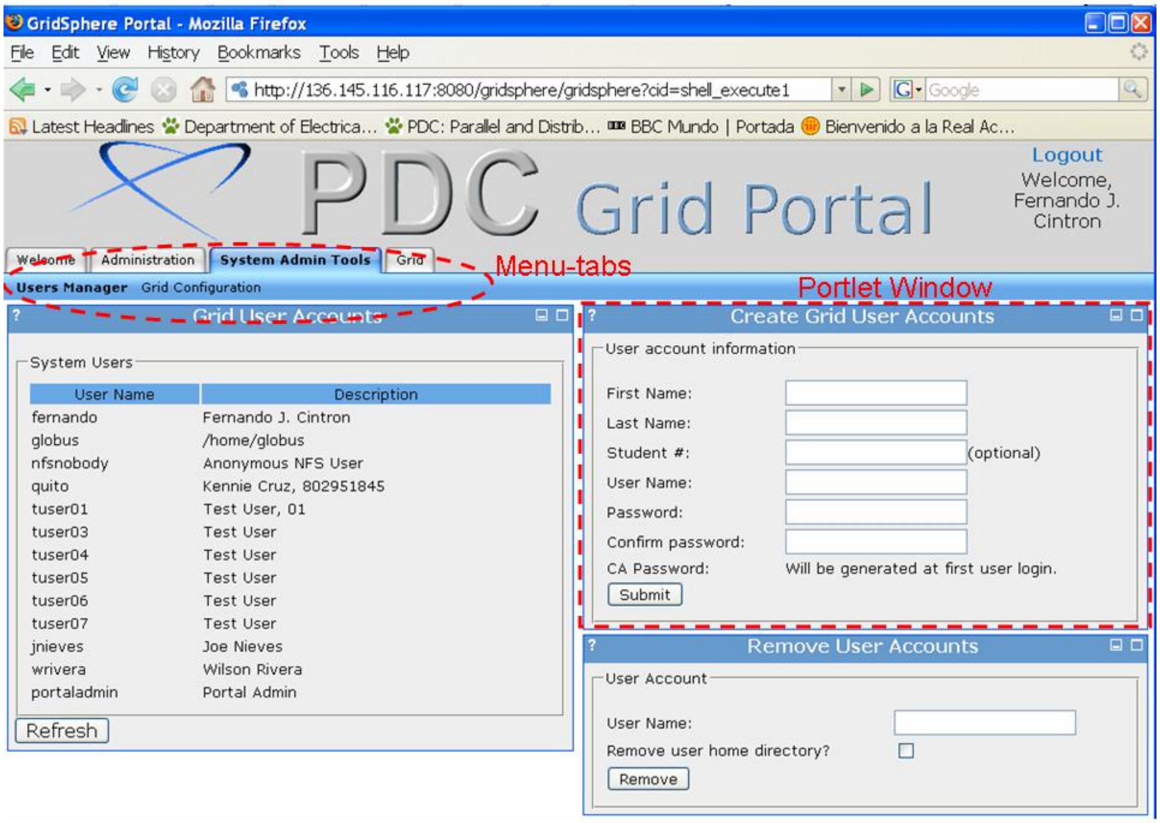
- **Resource Discovery** -- See at a glance, the status and availability of all the computing resources you have access to.
- **Job Submission** -- Submit jobs to any of the resources.
- **File Transfer** -- Easily transfer files among resources and from your local machine to any of the resources.

Connectivity Scheme for PDCLab Grid Testbed

Figure 3–2: PDC Lab Grid Portal

Groups are presented to users as “menu-tabs”panes by Gridsphere. These “menu-tabs”can be further divided and organized (figure 3–3), allowing grouping of related portlets.

Following the menu-tabs schema, the set of administrative tools contain a section for grid user account management. The grid user accounts are the ones used on the grid nodes for users to login through a command line interface and work. User without accounts could not be able to store and manage data, obtain credential certificates, and submit jobs. The grid user account management section allows the easy creation of grid accounts by administrators through the grid portal interface, instead of having to go through the command line interface.


GridSphere Portal - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://136.145.116.117:8080/gridsphere/gridsphere?cid=shell_execute1
Google
Latest Headlines
Department of Electrica...
PDC: Parallel and Distrib...
BBC Mundo | Portada
Bienvenido a la Real Ac...
PDC Grid Portal
Logout
Welcome, Fernando J. Cintron
Welcome
Administration
System Admin Tools
Grid
Menu-tabs
Users Manager
Grid Configuration
Portlet Window
Refresh
Submit
Remove

Grid User Accounts

System Users

User Name	Description
fernando	Fernando J. Cintron
globus	/home/globus
nfsnobody	Anonymous NFS User
quito	Kennie Cruz, 802951845
tuser01	Test User, 01
tuser03	Test User
tuser04	Test User
tuser05	Test User
tuser06	Test User
tuser07	Test User
jnieves	Joe Nieves
wrivera	Wilson Rivera
portaladmin	Portal Admin

Create Grid User Accounts

User account information

First Name:

Last Name:

Student #: (optional)

User Name:

Password:

Confirm password:

CA Password: Will be generated at first user login.

Remove User Accounts

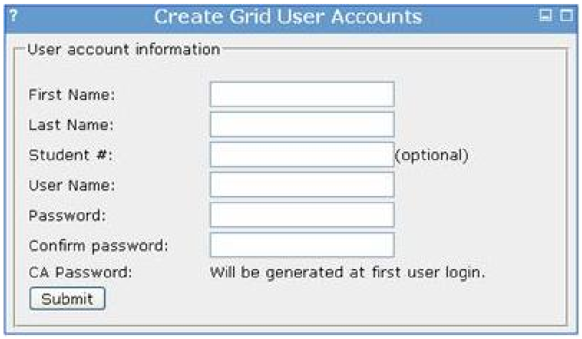
User Account

User Name:

Remove user home directory? □

Figure 3–3: Menu-tabs panes and portlets windows

The grid user account management contains three portlets: create account, remove account, and show accounts. The Create User Accounts portlet receives the user account details as input by an administrator, figure 3–4. It then validates this information making sure every required field is filled. It also verifies that the user remembers its password by making him/her entering it twice, to compare it. When validation is complete, the portlet object utilizes a GIAB client object to establish a connection with the GIAB configuration agent and sends the information for the account creation on the grid. The GIAB configuration agent creates the user account and replies back to the GIAB client object with the creation status. The portlet object obtains the creation status from the GIAB client object and displays it on the portlet user interface.



Create Grid User Accounts

User account information

First Name:

Last Name:

Student #: (optional)

User Name:

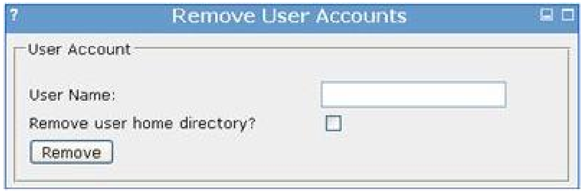
Password:

Confirm password:

CA Password: Will be generated at first user login.

Figure 3–4: Create User's Accounts portlet

The Remove User Account, figure 3–5, behaves in the same manner as the create user account portlet. It receives the input information for the account desired to be removed, then with the GIAB client object it sends the information to the GIAB configuration agent, which takes charge for the rest of the process.



Remove User Accounts

User Account

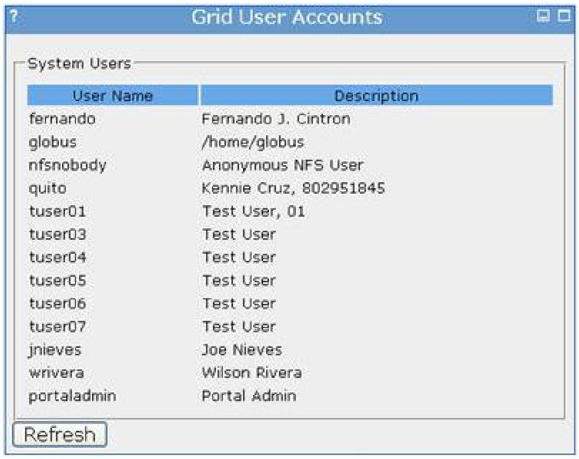
User Name:

Remove user home directory? □

Figure 3–5: Remove User's Accounts portlet

The Show User Accounts is a portlet that displays the existing accounts on the system, figure 3–6. This portlet helps grid administrators identify the user accounts and serves as confirmation method when working with the add/remove user accounts portlets. When a user account is created or deleted, the Show User Accounts reflects this change in the system.

Foreseeing potential growing of the grid portal configuration tools, we have added a menu-tab section (figure 3–7) for the grid services configuration. Within this section the administrator is able to configure Globus Simple CA and generate host and user certificates, configure Myproxy server, GridFTP, Globus service container,



User Name	Description
fernando	Fernando J. Cintron
globus	/home/globus
nfsnobody	Anonymous NFS User
quito	Kennie Cruz, 802951845
tuser01	Test User, 01
tuser03	Test User
tuser04	Test User
tuser05	Test User
tuser06	Test User
tuser07	Test User
jnieves	Joe Nieves
wrivera	Wilson Rivera
portaladmin	Portal Admin

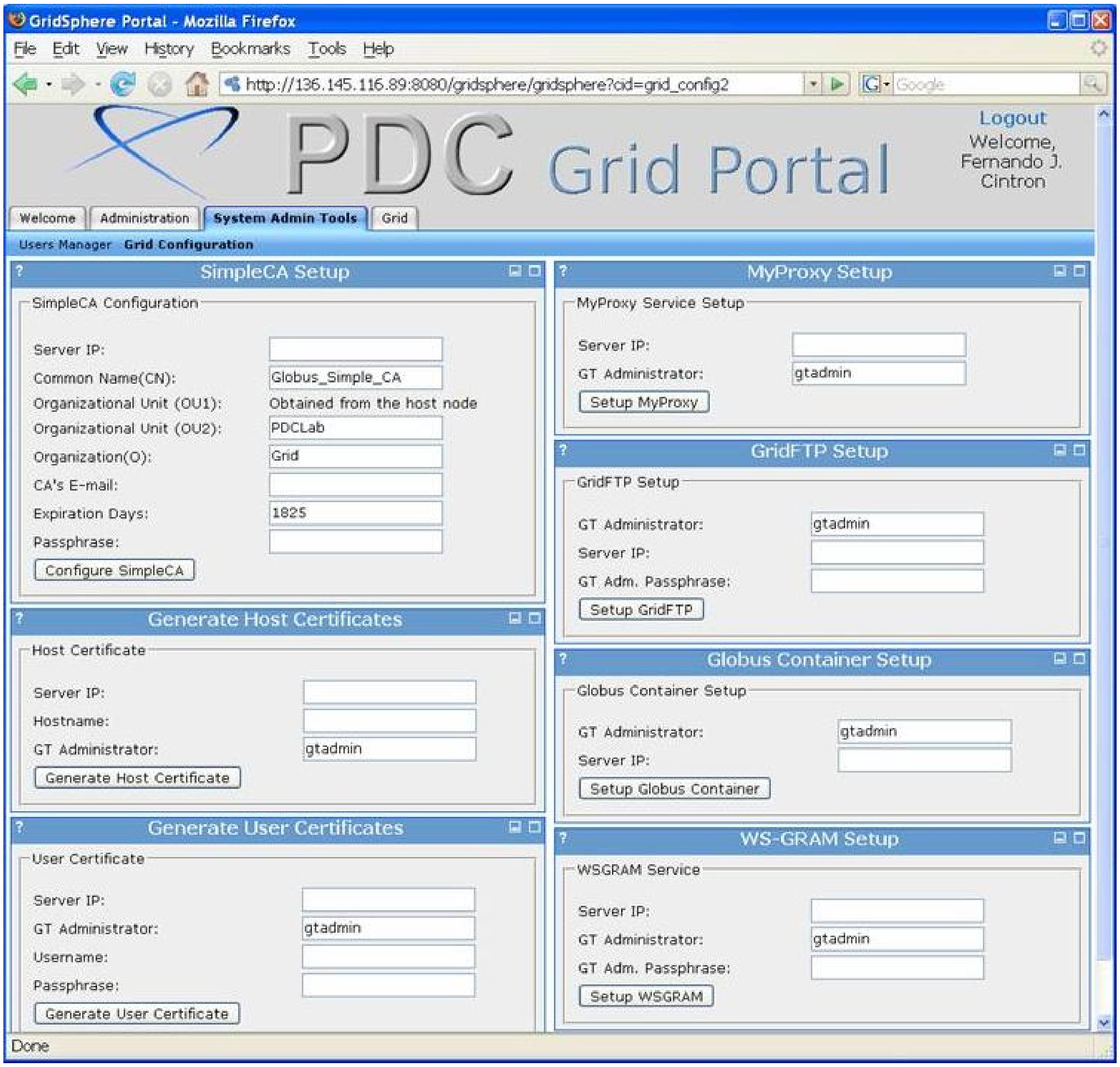
Figure 3–6: Show User's Accounts portlet

and WS-GRAM. This section is expected to be complemented and enhanced with future contributions related to the automated configuration of grid infrastructures.

3.5 GIAB Client/Agent Interaction

The GIAB client is a custom developed class, invoked by the portlets to communicate with the GIAB agent. It establishes the connection with the GIAB agent and server, sends the commands from the portlet, and communicates back the status of the request. The configuration command is a keyword that represents the action required followed by the arguments needed to perform this action. If we take, for example, the Remove User Account configuration portlet, the configuration command looks like *uRem username*, where *uRem* is the keyword telling the GIAB agent that a user account need to be removed. The second word in the command tells which user account to remove.

The GIAB configuration agent that takes charge of launching configuration scripts on the grid nodes. It receives the configuration commands from the GIAB client, interprets these commands, executes them, and replies back with the status of the action. During interpretation, the GIAB agent translates the configuration command into a Linux command. The Linux command can be either a system command, or in a more detailed task, it can launch a configuration script for a


GridSphere Portal - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://136.145.116.89:8080/gridsphere/gridsphere?cid=grid_config2
Google
PDC Grid Portal
Logout
Welcome, Fernando J. Cintron
Welcome
Administration
System Admin Tools
Grid
Users Manager
Grid Configuration
Configure SimpleCA
Generate Host Certificate
Generate User Certificate
Setup MyProxy
Setup GridFTP
Setup Globus Container
Setup WSGRAM
Done

SimpleCA Setup

SimpleCA Configuration

Server IP:
Common Name(CN): Globus_Simple_CA
Organizational Unit (OU1): Obtained from the host node
Organizational Unit (OU2): PDCLab
Organization(O): Grid
CA's E-mail:
Expiration Days: 1825
Passphrase:

Generate Host Certificates

Host Certificate

Server IP:
Hostname:
GT Administrator: gtadmin

Generate User Certificates

User Certificate

Server IP:
GT Administrator: gtadmin
Username:
Passphrase:

MyProxy Setup

MyProxy Service Setup

Server IP:
GT Administrator: gtadmin

GridFTP Setup

GridFTP Setup

GT Administrator: gtadmin
Server IP:
GT Adm. Passphrase:

Globus Container Setup

Globus Container Setup

GT Administrator: gtadmin
Server IP:

WS-GRAM Setup

WSGRAM Service

Server IP:
GT Administrator: gtadmin
GT Adm. Passphrase:

Figure 3–7: Grid Services Configuration section

given service. The configuration scripts are written to configure a specific service or execute a chain of tasks. Usually they are written by a service developer who has a complete understanding of the service and the requirements to deploy it on the grid.

Figure 3–8 presents a flow diagram of interaction between the configuration portlet and GIAB agent. The process starts when the administrator accesses a configuration portlet on the grid portal. The portlet creates an instance object of the GIAB client class, which in turn establishes a connection with the GIAB agent. The GIAB agent launches the required task script to accomplish the request.

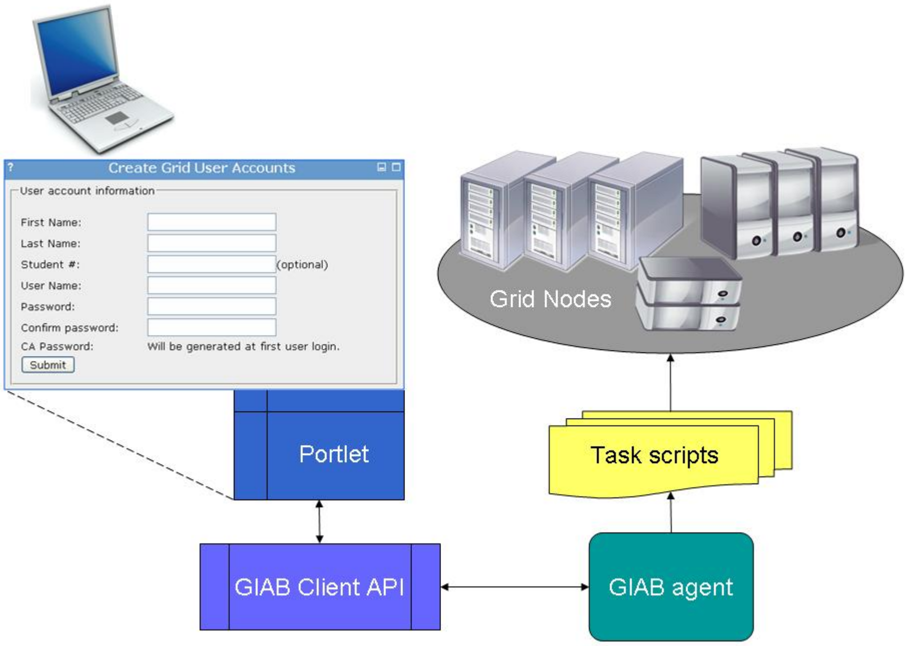


Figure 3–8: Portlet and GIAB Agent Interaction

3.6 Workflow Scenario

Now that we have presented all the pieces of this project, let us expose a case scenario that illustrates how they bind together. First, we want to deploy a grid and we count with the system resources identified on section 3.1. One option is to start with a portion of these resources, deploy a grid using just this subset of resources and add more later as needed. The next step is to install the GIABLinux bundle on each of these resources. Installing GIABLinux is a straight forward task which only requires the system administrator to boot the system from the bundle CD. During installation the system administrator has to choose at least one node to host the grid portal, this is an option prompted before the bundle starts the installation process on the node where it is running.

Having all the grid infrastructure requirements in place, the next task is to start configuring grid services, and managing grid user accounts and grid portal user accounts. All of these tasks, can be accomplished by the system administrator accessing the grid portal.

In summary, the deployment process can be viewd as: Resources identification, GIABLinux bundle installation, and grid configuration and management from the grid portal. After these steps are completed, grid users can start accessing the grid portal, obtain their proxy credentials and have access to services through the portal user interfaces (portlets). From this point after, the work of grid administrators is to add more systems to the grid as needed by following the steps mentioned above, and deploy newly developed services with their corresponding portlet configurations.

CHAPTER 4
DEVELOPMENT PROCESS

The Grid Administration Toolkit is not final, meaning that each time a new service is developed, a new portlet interface has to be aggregated for its configuration and deployment. This chapter is dedicated to familiarize the reader and future developers with the required steps for contributing to this project and enhance the proposed configuration tool.

4.1 Service Integration Life Cycle

The Service Integration Life Cycle (SILC) describes the stages required for the integration of a service into the portal context. It can be described as:

- **Grid Service Development:** This stage encloses all the process for the development of a service to be deployed on the grid. The service can be seen as the mean of addressing a certain need of grid users. Services do not provide an end-user oriented interface; their primary goal is to provide a service on the grid. It is expected that interfaces be developed as methods for accessing the services.
- **Portlet Development:** Now that we have our service, we want to provide an end-user interface to it. This interface must interact with some intermediaries, or with the service directly, to provide the required information for service execution. During this stage the portlet interface is added to the grid portal context, enhancing the portal functionality with the new grid service integration.
- **Intermediary Development:** Sometimes we can not rely on a direct communication between the portlet and the service. It is then when an intermediary is needed. The main task of an intermediary is to receive the information from the

portlet and pass it on to the service. An intermediary can be a client object that establishes the connection with the service or with an agent that will communicate with the service.

Identifying these stages makes possible the designation of groups of developers by stages. There is no specific order in which these stages have to be completed. Developers should use it as an understanding of the implications required to integrate a new service into a portal context.

4.2 Configuration Tool Enhancement

When a new configuration portlet for a service is desired to be available on the proposed configuration tool, the developers will need to deal with the portlet development, configuration scripts writing, and the GIAB agent update. The following sections intend to introduce each of this steps.

4.2.1 Portlet development

Gridsphere provides tutorials and documentation for getting started with JSR 168 portlet development. They can be obtained directly from the web site www.gridsphere.org. Within the Gridsphere development environment there is a preestablished Ant configuration for building new portlets, from the Gridsphere directory by invoking:

ant new-project

This Ant command goal will ask the developers for a series of arguments like the portlet name and the desired new directory name. Then, after the user inputs the arguments, it will generate the basic portlet objects structure in an automatically created directory under the Gridsphere source directory at the projects directory (Gridsphere-directory/projects/new-portlet-directory). The new portlet project will include the references for the Gridsphere classes libraries and the JSP User Interface libraries. Developers can start working from this point by adding the portlet JSP interfaces and editing the portlet class to references this JSP interfaces and perform the desired actions. It is here, during the portlet class development, when the developers will get to use the GIAB client to do the intended new configuration and/or deployment of the service this configuration is for. Also, some pre-created files need to be edited, they are:

portlet.xml File that provides the association of the portlet and the java class it corresponds to. It also sets the windows properties for the portlet. In the event more than one portlet is created under a single project, this file can describe more than one portlet.

layout.xml Provides the organization of where to place the portlet(s) inside the Gridsphere framework. In other words, here the developer can work with the arrangement of the portlets, place them under new tabs-pane, and organize them in multiple columns and rows.

group.xml When portlets are grouped together Gridsphere uses this file to know the portlets in the group and the group properties (e.g. user visibility, establishing who can subscribe to it).

Finally, when the portlet is ready, the developers just need to deploy it into the portal. To deploy the new portlet, there is another defined Ant command goal

ant install

Executed from inside the new portlet directory, it will build and deploy the portlet into the Gridsphere portal running on the servlet container, in this case Tomcat.

4.2.2 Writing Configuration Script

Our methodology for configuring a grid service is through the automation of the manually required steps that usually are carried out by the grid administrator. This automation, is done by writing a script that will do all the required steps to install any dependency and configure the new service. A good understanding of the service intended to be deployed along with its required dependencies is required in order to write a configuration script. There is no specific methodology for the developer to accomplish this step. However a prior agreement has to be made here in terms of which scripting language the grid nodes recognize, to avoid configuration problems at time of deployment.

4.2.3 GIAB Agent Enhancement

Every time a new configuration tool is added, it is required for the GIAB Agents to know what to do with the configuration requests from this new tool. GIAB Agents have to be aware if there are new configuration scripts for this configuration task and how to launch it. To address this, the developer needs to communicate to an administrator with access to the GIAB Agent source code to add the new supported command.

CHAPTER 5
CONCLUSION AND FUTURE WORK

In this project we have developed an easy to use grid portal environment for grid services administration. The main features of the developed portal environment are listed bellow:

1. The portal provides access to grid management services via portlets.
2. The portal portlets are integrated to an installation and configuration package that automates grid management services.
3. The portal uses existing grid portal technologies which facilitates rapid prototyping of new interfaces to grid services.

This work also has introduced the concept of Service Integration Life Cycle (SILC) to formalize the service integration workflow in the proposed environment.

According to our experience, the approximate time for a person with a small background in this field of study to understand and complete the deployment of a grid, composed of two nodes, for first time could be around two weeks. For a person that is very related to the topic and that had go through the deployment process several times, it could take approximately from one to two hours. With the tool presented on this project we successfully went through the configuration process of a grid node in approximately 20 minutes. These times are estimates, and are based on our observations at the Parallel and Distributed Laboratory, at the University of Puerto Rico.

5.1 Future Work

The flexibility of the proposed portal environment allows developers easily to contribute with new management services. Future work includes:

1. The implementation of a system discovery method, in which resources can be discovered after GIAB Linux is fully installed with the GIAB Configuration Agent running. It should be possible to configure new resources with a number of specific services via portlet interfaces.
2. GIAB Configuration Agent needs to be discoverable. It may be registered to the portal host when queried or it could do it periodically. This area need to be studied in order to design a good discovery protocol. A mechanism to alert the GIAB agents of new configuration scripts is also required.

REFERENCE LIST

[1] I. Foster and C. Kesselman. *The grid: blueprint for a future computing infrastructure*. Morgan Kaufmann Publishers, 1998.

[2] I. Foster, C. Kesselman, J. Nick, and S. Tuecke. The physiology of the grid: An open grid services architecture for distributed systems integration. Technical report, Open Grid Service Infrastructure WG, Global Grid Forum, June 2002.

[3] D. Gannon; G. Fox; M. Pierce; B. Plale; G. von Laszewski; C. Severance; J. Hardin; J. Alameda; M. Thomas; J. Boisseau. *Grid Portals: A Scientist's Access Point for Grid Services*. GGF Community Practice document, working draft 1 edition, September 2003.

[4] Gregor von Laszewski; Jarek Gawor; Sriram Krishnan; Keith Jackson. *Grid Computing: Making the Global Infrastructure a Reality*. Wiley, 2003. chapter Commodity Grid Kits - Middleware for Building Grid Computing Environments.

[5] I. Foster. Globus toolkit version 4: Software for service-oriented systems. In Springer-Verlag, editor, *International Conference on Network and Parallel Computing*, volume 3779 of *LNCS*, pages 2–13. IFIP, 2006.

[6] Maintenance lead: Stefan Hepper; Alejandro Abdelnur; Elaine Chien. JSR 168: Portlet Specification. Java Community Process, [web page] http://jcp.org/en/jsr/detail?id=168, October 2003.

[7] Novotny, J.; Russell, M.; Wehrens, O. GridSphere: An Advanced Portal Framework. In *Euromicro Conference, 2004. Proceedings. 30th*, pages 412–419. IEEE, 2004.

[8] J. Novotny, S. Tuecke, and V. Welch. An Online Credential Repository for the Grid: MyProxy. In *Proceedings of the Tenth International Symposium on High Performance Distributed Computing (HPDC-10)*, pages 104–111. IEEE Press, August 2001.

[9] Xiao Dong Wang; Xiaobo Yang; Allan, R.;. Top Ten Questions To Design A Successful Grid Portal. In *Semantics, Knowledge, and Grid, 2006. SKG '06. Second International Conference on*, pages 18–18. IEEE Computer Society, November 2006.

[10] Gannon, D.; Alameda, J.; Chipara, O.; Christie, M.; Dukle, V.; Fang, L.; Farrellee, M.; Kandaswamy, G.; Kodeboyina, D.; Krishnan, S.; Moad, C.; Pierce, M.; Plale, B.; Rossi, A.; Simmhan, Y.; Sarangi, A.; Slominski, A.; Shirasuna, S.; Thomas, T.;. Building grid portal applications from a web service component architecture. In *Proceedings of the IEEE*, volume 93 of *3*, pages 551–563. IEEE, March 2005.

[11] Gabriel E. Lindner, P. and M.M. Resch. GCM: a grid configuration manager for heterogeneous grid environments. *Int. J. Grid and Utility Computing*, Vol. 1(No. 1):pp. 4–12, 2005.

[12] Brookhaven National Laboratory. *Grid User Management System, v.SNAPSHOT*, June 2006. Project Documentation, [web page] http://grid.racf.bnl.gov/GUMS/index.html.

[13] Arshad Ali; Richard McClatchey; Ashiq Anjum; Irfan Habib; Kamran Soomro; Mohammed Asif; Ali Adil; Athar Mohsin. From Grid Middleware to a Grid Operating System. Presented at the 7th IEEE/ACM International Conference on Grid and Cooperative Computing, Changsha, China, [on site] http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0608046, August 2006.

BIOGRAPHICAL SKETCH

Fernando J. Cintón-González was born in May 26, 1982 in Ponce, Puerto Rico. Fernando is the son of José A. Cintrón-García and Olga I. González-Díaz. In December 2004 he received his BS in Computer Engineering from the University of Puerto Rico at Mayaguez. Since January 2005, he is pursuing a master degree in Computer Engineering, at the same institution.