



TEST AND INTEGRATION - STATE OF THE ART

WP3 New Grid Services and Tools

Document Filename:	CG-3.5-SRS-0020-2-0-StateTheOfArt
Work package:	WP3 New Grid Services and Tools
Partner(s):	PSNC, CYFRONET, UAB, CSIC
Lead Partner:	CSIC
Config ID:	Task 3.5 State of the Art Public Document
Document classification:	PUBLIC

Abstract: This document gives a brief description about the projects and works related to the Test and Integration Process of the whole set of tools of the WP3 middleware. The main goal of this is to deliver an official prototype version of the middleware which will be used by WP1 and WP2 (Applications) and will be prepared with the aim of running on an interactive Grid framework (Testbed supported by WP4).



Delivery Slip

	Name	Partner	Date	Signature
From	WP3	CSIC	29-05-2002	Santiago González de la Hoz
Verified by				
Approved by				

Document Log

Version	Date	Summary of changes	Author(s)
1-0-A	31-May 2002	First version	Santiago González de la Hoz, Fco. Javier Sánchez Martínez

CONTENTS

1 INTRODUCTION.....	4
1.1 PURPOSE.....	4
1.2 DEFINITIONS, ACRONYMS, AND ABBREVIATIONS.....	4
1.3 REFERENCES.....	4
1.4 CROSSGRID – TEST AND INTEGRATION.....	4
2 SURVEY OF PROJECTS AND WORKS.....	5
2.1 OBTAIN CROSSGRID SOFTWARE.....	6
2.2 SUPPORT.....	7
2.2.1 Website.....	7
2.2.2 WP3 Middleware Installation.....	7
Bugzilla.....	8

1 INTRODUCTION

1.1 PURPOSE

The purpose of this document is to give a short description of the current state of projects and works aimed at giving the user (WP1, WP2) easy access to the tools and services developed by the WP3. The main emphasis is put on products that give access to all of the software, application kits and information of the WP3 release version.

1.2 DEFINITIONS, ACRONYMS, AND ABBREVIATIONS

CVS	Concurrent Version System
RPM	Red Hat Package Manager
NFS	Network File System
LCFG	Local ConFiGuration system (tool developed at the Edinburgh University)
DHCP	Dynamic Host Configuration Protocol

1.3 REFERENCES

- [1] "Description of Work". CrossGrid Project technical Annex, 2001
http://www.cyf-kr.edu.pl/crossgrid/CrossGridAnnex1_v31.pdf
- [2] EDG User's Guide. DataGrid-06-TED-0104-1-03
<http://marianne.in2p3.fr/datagrid/documentation/>
- [3] EDG Installation Guide. DataGrid-06-TED-0105-1-25
<http://marianne.in2p3.fr/datagrid/documentation/>

1.4 CROSSGRID – TEST AND INTEGRATION

The main objectives of Task 3.5 are deliver all information about possible errors detected while the system tests and required corrections, and prepare an official prototype version of the whole set of tools of the WP3 middleware. This middleware will be used by other WP's, especially WP1 and WP2. Task 3.5 requires that the local grid Testbed will be available in Month 6 (with a certification authority, registration, authority, monitoring system and access to data storage system) and the full grid will be available in Month 12 [1].

2 SURVEY OF PROJECTS AND WORKS

The Grid makes widely distributed computing resources transparently available to the end-user. As well as purely computational resources, these include data storage and networking. The CrossGrid collaboration builds software components which enable the access to these resources and the CrossGrid Testbed is the vehicle for testing of the CrossGrid software by end-user.

In the Testbed there are a handful of different type of machines which provide various services to the grid [2]. The following describe these various machines and where they fit into the overall Testbed:

User Interface

This machine runs the User Interface software which allows the end-user to interact with the DataGrid testbed. This is typically the machine the end-user logs into to submit jobs to the grid and to retrieve the output from those jobs.

Computing Element (Service)

A computing element consists of one gatekeeper node and one or more worker nodes. Together these provide for the user some computation resources.

Gatekeeper

This is the frontend of a computing element. This node handles the interaction with the rest of the grid environment accepting jobs, dispatching them for execution, and returning the output. This provides a uniform interface to the computational resources it manages.

Worker Node

These nodes sit behind a gatekeeper and are typically managed via a local batch system. The details of this are hidden from the end-user by the gatekeeper; however, these are the nodes on which user computations are actually performed. Consequently, the end-user software is installed on these nodes. These nodes do not run any DataGrid daemons, but do have a few APIs for accessing information.

Storage Element

These nodes provide uniform access to large storage spaces. The storage element may control large disk arrays, mass storage systems and the like. This element hides the details of the backend storage systems and provides a uniform interface to the grid user.

Resource Broker

These machines accept jobs from users (via the User Interface), match the jobs' requirements to the available resources, and dispatch them.

The test and Integration process begins with preparation of the infrastructure by Task 3.5 and ends with the release of an integrated WP3 system by the Integration Team. This paper contains examples of products to access to CrossGrid software and lead site administrators through the whole installation, configuration and customisation process. This document describes specially products used by the DataGrid project [3].

2.1 OBTAIN CROSSGRID SOFTWARE

The package repository, hosted by FZK (Karlsruhe) and a backup copy located at IFIC (Valencia), will provide access to the packaged Globus, CrossGrid, and required external software. In the DataGrid project all software is packaged as source binary RPM files. A central CVS repository, intended mainly for developers, maintains the sources of the DataGrid code.

All of the packages are available directly through a Web Interface, but this is convenient only for downloading individual packages. The most convenient access is a set of configuration files. These files contain link to various subsets of the recommended RPMs. The "wget" command can be use to retrieve all of the referenced RPMs. You can get "wget" from rpmfind.net (<http://rpmfind.net/linux/RPM/redhat/6.2/i386/wget-1.5.3-6.i386.html>). More recent version of "wget", in particular the one distributed with RH7.2, does not work correctly.

- RPM is the RPM Package Manager. It is an open packaging system available for anyone to use. It allows users to take source code for new software and package it into source and binary form such that binaries can be easily installed and tracked and source can be rebuilt easily. It also maintains a database of all packages and their files that can be used for verifying packages and querying for information about files and/or packages. Permission is granted to use and distribute RPM royalty free under the GPL (GNU General Public License).

More complete documentation is available on RPM in the book by Ed Bailey, Maximum RPM. That book is available for download or purchase at www.redhat.com.

- CVS is a version control system. Using it, you can record the history of your source files. For example, bugs sometimes creep in when software is modified, and you might not detect the bug until a long time after you make the modification. With CVS, you can easily retrieve old versions to see exactly which change caused the bug. This can sometimes be a big help.

You could of course save every version of every file you have ever created. This would however waste an enormous amount of disk space. CVS stores all the versions of a file in a single file in a clever way that only stores the differences between versions.

CVS also helps you if you are part of a group of people working on the same project. It is all too easy to overwrite each others' changes unless you are extremely careful. Some editors, like GNU Emacs, try to make sure that the same file is never modified by two people at the same time. Unfortunately, if someone is using another editor, that safeguard will not work. CVS solves this problem by insulating the different developers from each other. Every developer works in his own directory, and CVS merges the work when each developer is done.

For more information on downloading CVS and other CVS topics, see:

<http://www.cyclic.com/>

<http://www.loria.fr/~molli/cvs-index.html>

- GNU Wget is a free software package for retrieving files using HTTP, HTTPS and FTP, the most widely-used Internet protocols. It is a non-interactive commandline tool, so it may easily be called from scripts, cron jobs, terminals without Xsupport, etc.

Wget has many features to make retrieving large files or mirroring entire web or FTP sites easy, including:

- Can resume aborted downloads, using REST and RANGE

- Can use filename wild cards and recursively mirror directories

- NLS-based message files for many different languages

- Optionally converts absolute links in downloaded documents to relative, so that downloaded documents may link to each other locally

- Runs on most UNIX-like operating systems as well as Microsoft Windows

- Supports HTTP and SOCKS proxies

- Supports HTTP cookies

- Supports persistent HTTP connections

- Unattended / background operation

- Uses local file timestamps to determine whether documents need to be re-downloaded when mirroring

- GNU wget is distributed under the GNU General Public License.

2.2 SUPPORT

2.2.1 Website

The main WP3 Website has to contain documentation, contact information, the bug-reporting system, links to the source and packages repositories as well as links to other sites. An example of this is the Website of WP6 in the DataGrid project (<http://marianne.in2p3.fr/>).

2.2.2 WP3 Middleware Installation

The DataGrid project is installing and configuring automatically a set of machines (a farm) from scratch, using LCFG (<http://datagrid.in2p3.fr/distribution/datagrid/wp4/installation/doc/>). This tool is useful not only if you have a large number of nodes, but also if you want to automate your Testbed installation and configuration. It is also possible to perform automatic upgrading/downgrading operations.

LCFG needs a central server (LCFG server) from which LCFG clients fetch configuration files and RPMs (via HTTP and a NFS export).

On the LCFG server both HTTP and DHCP server are needed, and a disk partition is shared with the clients via NFS. The clients, during the boot process, take the IP number via DHCP and the configuration via HTTP.

In DataGrid project LCFG configuration templates and tools (LCFG objects) have been developed in order to install and configure the four types of Testbed machines: Computing

Element (CE), Worker Node (WN), Storage Element (SE) and User Interface (UI) (for the Resource Broker and the Network Monitoring machine RPM lists will exist but a manual configuration will be needed).

- The Dynamic Host Configuration Protocol (DHCP) is an Internet protocol for automating the configuration of computers that use TCP/IP. DHCP can be used to automatically assign IP addresses, to deliver TCP/IP stack configuration parameters such as the subnet mask and default router, and to provide other configuration information such as the addresses for printer, time and news servers.

- **Bugzilla**

The bug-tracking system used in DataGrid is Bugzilla (<http://marianne.in2p3.fr/datagrid/bugzilla>), it is intended to facilitate the reporting and fixing of bugs in the European DataGrid software. This includes the DataGrid's distribution of the Globus2 system; confirmed bugs in Globus will be forwarded to the Globus team.

This system is not intended to track bugs in application software, that is, use and experimental software running on the Grid.

The Bugzilla database is publicly available and can be searched by anyone. However, reporting bugs requires a valid e-mail address. Note that Bugzilla uses cookies to keep track of your account data, so your browser must have cookie support enabled.