Testing Heterogeneous Computing Nodes for Grid Computing

Eamonn Kenny, Brian Coghlan, John Walsh, Stephen Childs, David O' Callaghan, Geoff Quigley

Department of Computer Science, Trinity College Dublin, Ireland.
email: [ekenny,coghlan,walshj1,childss,ocalladw,gquigle]@cs.tcd.ie
phone: (+353 1) 608 1536, *fax:* (+353 1) 677 2442

Abstract

A heterogeneous implementation of the current LCG2/EGEE grid computing software is supported in the Grid-Ireland infrastructure. The porting and testing of the current software version of LCG2 is presented for IRIX 6.5, Red Hat 7.3, Red Hat 9 and Fedora Core 2. Three benchmarks are run on each platform and the results for each are compared.

1 Introduction

Grid-Ireland uses the LCG[6] and EGEE[2] grid middleware, which only supports assumed reference ports to Red Hat 7.3, Microsoft Windows and Scientific Linux 3 (SL3). Unfortunately this is a very restrictive situation, counter to the original heterogeneous ethos of grid computing. As a result of our interest in heterogeneity, we at Trinity College Dublin began porting to non-reference platforms in October 2003. Subsequently EGEE have finished porting the current LCG2 grid implementation to Scientific Linux on 32-bit and 64-bit architectures. Below we report on our progress with non-reference ports.

2 Porting for Heterogeneity

2.1 Current Porting Status

The LCG2/EGEE software components are shown in the form of a dependency graph in Figure 1. Grid-Ireland wished, in the first instance, that the porting of the LCG2 software to other platforms would focus on the ability to execute Globus and EDG jobs on worker nodes, and that replica management (both Java and C/C++ based), R-GMA and VOMS would be supported.

To avail of the base functionality requires Globus and various EDG support packages. Since Globus 2.4.3 is known to have many bugs, the University of Wisconsin-Madison corrects these and packages all the necessary components as part of the Virtual Data Toolkit (VDT) [11]. We have assisted Maarten Litmaath in CERN to port VDT-1.1.14 to IRIX and Fedora Core 2. A Red Hat 9 port is already provided by VDT [10]. This Red Hat 9 port is used under

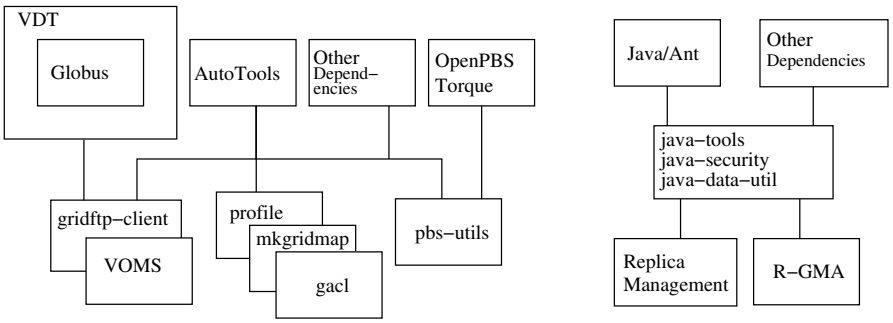

Fig. 1: LCG2/EGEE software components

SL3 without any recompilation of the RPMs. A port exists for Globus to Mac OS X (Darwin) and AIX but the VDT version must be ported to both of these platforms. In the near future CERN will provide a 64-bit AMD port of VDT which we will then port to AIX and Darwin. This will then allow globus and EDG job submission to both platforms.

Grid-Ireland also wished that MPI, replica management and the OpenPBS client be provided on each worker node. In some cases Torque might be required since newer versions of operating systems are not always provided for in OpenPBS. Also the R-GMA information system producer and consumer APIs and the VOMS client were required.

At the moment there is no requirement by Grid-Ireland for the workload management system (WMS) but it appears that there is logging of WMS events from worker nodes to the resource broker. This logging activity can be disabled but we would prefer to retain this desirable feature. WMS consists of many modules, but it might be able to be refactored to delineate those specific to the event logging, so that just this functionality needs to be ported. It should be noted that if the whole of WMS were ported successfully then almost everything will have been ported because it depends on so many other packages. At the moment job submission is possible without having any WMS software on the worker node.

In the case of Fedora Core 2, Queen Mary's College London would like us to provide a LCG 2.3 port of the computing element (CE), storage element (SE), user interface (UI), network monitor (NM) and worker node (WN). In it's present state the Fedore Core port builds all but 3 modules. Pending issues are:

1. WMS requires the building of boost-g3 from source RPMs. boost-g3 gives problems because it requires python-2.2. Python-2.3 can not be removed easily from Fedora because so many packages rely on it. Other problems exist due to a requirement to use gcc-3.3.3 under Fedora Core. Once boost-g3 can be built the rest of the port should be straight forward.
2. The LCG monitoring module requires that perl-5.8.0 be installed temporarily under Fedora to build the RPM, which is unacceptable. The code has not been upgraded for new version of perl.

3. The Replica management service (RMS) has issues related to the use of the gcc-3.3.3 compiler.

Recently we have helped INFN successfully port VOMS to all versions of Linux that we are currently considering. In January 2005 be have started work on the gLite build system since this should be used to port R-GMA to IRIX, AIX, Fedora Core 2, Red Hat 9 and Darwin. This is not fully set up yet, but should be operating soon under Fedora Core 2 and SL3.

There are a number of on-going issues, but we have successfully ported the functionality for job submission to Fedora Core 2, IRIX 6.5.14 and 6.5.17m, and Red Hat 9. We also plan to do this for a number of other platforms if the need arises within Grid-Ireland. Debian might be a possible candidate.

2.2 Autobuilding the Non-Reference Ports

Early in the porting effort it became obvious that it would be necessary to make an exact replica of the DataGrid Red Hat 7.3 software repository so that any teething problems could be solved incrementally as the code was ported to other platforms. At first the whole European DataGrid repository was obtained from the developer of the build software, Yannick Patois[3][7], of IN2P3. This software was used to maintain a Red Hat 6.2 and 7.3 repository of all the EDG software up to November 2003. The build software is now used to maintain the CrossGrid repository at FZK and the LCG2/EGEE repositories at CERN. CERN is using the auto-building tool to support Scientific Linux 3 (SL3) on both 32-bit and 64-bit architectures, as well as the Red Hat 7.3 version. Since April 2004 we are in constant contact with CERN with regard to porting the worker-node software to five other platforms. Currently we maintain a full copy of the EDG, LCG2 and some CrossGrid modules on a Red Hat 7.3 repository.

A number of CVS repositories are used to build all the necessary software for a worker node. The head version of VOMS is obtained from INFN's own repository. The whole of LCG2 is extracted using CVS checkouts directly from CERN's lcgware repository. The CrossGrid software is obtained by directly copying the CVS repository to a local repository. Edg-build then contacts this local repository to perform the nightly builds. The RAL repository of R-GMA will also need to be added soon, since LCG2 no longer maintain the most recent version of R-GMA.

Recently the port to Mac OS X running Darwin was initiated. Darwin requires the installation of *fink* to provide the tarballs necessary to start the autobuild process. Because it is a 64-bit architecture, completing the port is not possible until the 64-bit port of VDT is available.

Because different repositories are being used to build different pieces of software, this means that there must be a number of different build machines, one for each repository. We have designed additional configuration scripts within edg-build for each type of CVS repository and also different start-up scripts for the cron jobs for each repository type. This makes the build process very easy to maintain and easy to port to other machines.

In Figure 2 the current status of the build system can be seen. This figure is a snapshot of the Grid-Ireland auto-build web page [1] at the start of February 2005. The results change quite regularly as new ports are completed.

OS Type	Version	VDT	Basic	VOMS	RGMA	RM	GFAL
Redhat	7.3	RPMS	RPMS	RPMS	RPMS	RPMS	RPMS
Redhat	9.0	RPMS	RPMS	RPMS	RPMS	RPMS	RPMS
Fedora Core	2	RPMS	RPMS	RPMS	RPMS	RPMS	RPMS
SGI	6.5.14	tarball	tarball	tarball	tarball	tarball	tarball
AIX	5.2L	tarball	tarball	tarball	tarball	tarball	tarball
Darwin	10	tarball	tarball	tarball	tarball	tarball	tarball
SL	3	RPMS	RPMS	RPMS	RPMS	RPMS	RPMS

Colour	Meaning
	To be started
	Started
	Done

Fig. 2: Auto-build Results for Worker Nodes

3 Deployment of Non-Reference Ports

Currently our download repository contains only the Fedora Core 2 worker node RPMs. They can be deployed using yum, but we have found yum to be very slow in finding package dependencies, so we plan to migrate to apt. An empty RPM is provided which depends on all the others, so performing an install of this one package will install the whole worker node. Post installation is achieved using the guidelines for LCG2 worker node installation[8]. At the moment this must be done by hand, but we plan to automate this step. In the case of IRIX where only tarballs are available, copying the */opt/* directory from one IRIX machine to another may be sufficient to create a working worker node, but this needs to be tested, particularly to see if libraries are missing or stored in other directories. One way to avoid this might be to create *tardist* files of all packages and ensure that all dependencies are accounted for on the new worker node. Currently every worker node undergoes testing within what Grid-Ireland calls *TestGrid*. This is a pre-deployment site that is used to test the durability of each non-reference port. Currently TestGrid contains 5 non-reference architectures and will possibly contain more in the future.

4 Benchmarking Results

Presently EDG job submission is possible for the following non-reference platforms: SGI IRIX 6.5, Fedora Core 2, Red Hat 7.3 and 9 as described in [5], where preliminary results using a fast fourier transform (FFT) were used to show differences in computational speed between different architectures. A routine such as a FFT cannot be independently used as a benchmark since it gives no explicit information about I/O, CPU, caching, floating point or disk write speed.

CrossGrid GridBench benchmarks[9][4] developed at the University of Cyprus provide more precise benchmarking. Such benchmarks are not described in this paper since a collaboration between TCD and UCY to work on the issues of heterogeneity was agreed after the initial submission of this paper.

OS Type	Version	CPU Speed	iterations	$\bar{x}$	σ
Red Hat	9	2.8GHz	400	8.65	0.36
Red Hat	7.3	2.8GHz	400	8.39	0.53
Fedora Core	2	2.8GHz	400	9.6	0.34
IRIX	6.5.14	400MHz	100	10.52	0.01

Tab. 1: FFT EDG Job Submission Results

OS Type	Version	flops-1	flops-2	flops-3	flops-4	$\bar{x}$	σ
Red Hat	9	482.5173	508.6425	755.2024	897.5966	137.14	0.17
Red Hat	7.3	505.2518	513.4918	767.2589	922.5997	134.76	0.12
Fedora Core	2	446.5512	469.2480	696.9718	829.5849	148.93	0.19
IRIX	6.5.14	140.5242	127.0460	222.0428	329.9174	212.04	0.10

Tab. 2: Flops Job Submission Results

OS Type	Version	Dhrystones	$\bar{x}$	σ
Red Hat	9	3987240	25.050	0.143
Red Hat	7.3	4239084	23.729	0.404
Fedora Core	2	3766057	27.118	0.398
IRIX	6.5.14	527945	189.435	0.025

Tab. 3: Dhrystones Job Submission Results

Three benchmark modules: fft, flops and dhrystones were compiled on each architecture using the GNU compiler: gcc-3.2.2, except for Fedora which uses gcc-3.3.3. A compilation option of -O2 optimisation was used for Redhat 9, Redhat 7.3, Fedora Core 2 and IRIX 6.5 platforms. One benchmark executable along with a suitable job description language (JDL) file, is required for the job submission to a single workernode to obtain the benchmark results. To achieve consistent load between submissions, each job submission must be staggered in time. The results in Tab. 1, Tab. 2 and Tab. 3 were computed on the 4 worker nodes. In the case of Linux platforms each has identical hardware specifications. Each benchmark was run using a script that both timed and ran the benchmark

10 times. The results were then used to obtain a mean computation time ($\bar{x}$), and standard deviation (σ) for each benchmark.

5 Conclusions

The base worker node port of the LCG2/EGEE grid software for Globus and EDG job submission is now completed for Fedora Core 2, Red Hat 9 and IRIX. A 64-bit AMD version of VDT plus some manual post-configuration should be all that is required to achieve job submission to Darwin and AIX.

The results presented in Tab. 1, Tab. 2 and Tab. 3 show conclusively that Fedora Core 2 is slower for computational performance. In the case of integer computations it is markedly slower (see dhrystones in Tab. 3). In the case of the IRIX times it is difficult to compare the performance since the CPU speed is so much slower than that of the Linux systems. It is worth noting that the load on the IRIX is far more consistent as shown by the standard deviation results (σ) for all 3 benchmarks.

The results for IRIX are presented to show that job submission is possible for a number of applications rather than to be used as a comparison with Linux job submission.

6 Acknowledgements

We would like to thank IBM, Dell and DIAS for sponsoring us with machines to perform the software ports, and Science Foundation Ireland for funding this effort. Most of all we would like to thank the deployment group in CERN and INFN for all their help in porting to each platform.

References

1. Autobuild. *TCD Autobuild Web-page.* http://grid.ie/autobuild, November 2004.
2. EGEE. Enabling grids for e-science in europe. http://www.eu-egee.org/, November 2004.
3. Ariel Garcia, Marcus Hardt, Yannick Patois, and Ulrich Schwickerath. *Collaborative Development Tools.* Cracow Grid Workshop, October 2003.
4. Eamonn Kenny, Brian Coghlan, George Tsouloupas, Marios Dikaiakos, John Walsh, Stephen Childs, David O'Callaghan, and Geoff Quigley. *Heterogeneous Grid Computing: Issues and Early Benchmarks.* Submitted to ICCS 2005, December 2004.
5. Eamonn Kenny, Brian Coghlan, John Walsh, Stephen Childs, David O'Callaghan, and Geoff Quigley. *Heterogeneity of Computing Nodes for Grid Computing.* Accepted and to appear in EGC 2005, February 2005.
6. LHC. Large hadron collider computing grid project. http://lcg.web.cern.ch/LCG/, November 2004.
7. Yannick Patois. *Datagrid configuration management and build conventions.* http://datagrid.in2p3.fr/d6.2/DataGrid-ConfMngmt-BuildConv.html, 2003.

8. Antonio Retico, Alessandro Usai, Oliver Keeble, and Guillermo Diez-Andino. *LCG Grid INfrastructure Support.* http://grid-deployment.web.cern.ch/grid-deployment/gis/lcg-2_2_0/LCG2InstallNotes.pdf, August 2004.
9. George Tsouloupas and Marios D. Dikaiakos. Gridbench: A tool for benchmarking grids. In *Proceedings of the 4th International Workshop on GridComputing (GRID2003)*, pages 60–67, Phoenix, AZ, November 2003. IEEE.
10. VDT. *Versions for RH7.3 and RH9.0.* http://www.cs.wisc.edu/vdt/vdt_rpms/1.2.1/stable/, September 2004.
11. VDT. Virtual data toolkit. http://www.cs.wisc.edu/vdt/, November 2004.