The ZIDGrid: a versatile production environment for scientific computing

Reinhard Bischof*, Gabriel Esterhammer†, Michael Fink*, Wolfgang Kausch† and Peter Öttl†

Abstract.   *One of the most fascinating aspects of Grid computing is the utilization of resources for production which are normally used only part of the time (e.g. PCs in student labs). At the University of Innsbruck the High Energy Physics group initiated a co-operation with the ITCenter of the University (ZID) to make the PC-labs (with 290 PCs) accessible for scientific research groups for jobs with lower hardware requirements as a supplement to the computing resources operated by the ZID. An additional important aspect of the ZIDGrid is the usage as a testbed for middleware development within the AustrianGrid.*

In this paper we discuss the technical details and present some use cases for production and middleware development on this Grid environment.

1.   Introduction

1.1.   The history and the aims of the ZIDGrid

In 2003 the High Energy Physics group at the University of Innsbruck started a co-operation with the ZID within a pilot Grid project with the aim to make PCs in student's labs available for scientific computing during idle times. At this time the Institute of Engineering Mechanics had already announced their interest in such a supplement to the production clusters.

It was decided to restrict the Grid usage of the student labs to the time when the labs are closed for the students and not to try to harvest cycles during normal opening hours, because the machines are shut down or rebooted rather often.

In contrast to other projects which aim at the utilization of PC-labs during idle times such as PIRANHA [1] or WINZIG [2], one main aspect of the ZIDGrid right from the beginning was to provide a Grid testbed for the testing of middleware developed within the AustrianGrid Project. For this special purpose it is necessary to have a flexible system which allows starting several Grid Clusters with different setup details using the pool of available machines.

*ZID, University of Innsbruck, Austria

†HEPHY, University of Innsbruck, Austria

1.2. The ZIDGrid and its powerful local colleagues

The ZIDGrid is intended to be a supplement to the production environment for scientific computing maintained by the ZID Innsbruck. There are two SGI Altix350 machines, one from the AustrianGrid Project with 16 CPUs and 16 GB RAM and one for local usage with 32 CPUs and 128 GB RAM. The cluster of the High Performance Computing Group (LEO1) offers 206 CPU-Cores and 16 TB of disk storage and the ZID Compute Cluster provide another 44 CPUs. Several research groups are performing parameter studies with low hardware requirements and these jobs can be run on ZIDGrid resources to minimize the time needed for the production and to reduce the load on the production clusters.

2. Details

2.1. The Hardware

Currently the ZIDGrid consists of up to 292 PCs in 14 PC-labs. The hardware varies from machines with 1.8 GHz Pentium4 CPUs and 256 MB of RAM to machines with 2.8 GHz Pentium4 CPUs and 512 MB RAM. These PCs are dual boot machines (Windows XP and Fedora Core 3) with Linux as default boot. The network controllers support wake up on lan. After the normal opening hours for interactive usage the PCs are automatically shut down and can be restarted by wake up on lan servers if they are needed for the ZIDGrid. After the time reserved for Grid usage is finished the machines are shutdown. The home directories for the Grid users are exported via NFS and have a disk quota from 1.2 GB up to 12 GB, the machines offer up to 20 GB of local disk space.

2.2. The ZIDGrid clusters

Currently each of the 14 PC labs is a unique Grid cluster, several rooms can be divided into smaller clusters whenever the middleware tests make it necessary to have a larger number of clusters. The properties of each cluster are defined in a configuration file containing e.g. the name of the machine acting as master and the name of the local resource management. In order to make a direct comparison between different resource management systems possible, each cluster can be started either with Torque and Maui as scheduler, Condor or Sun Grid Engine. The master machine runs the necessary Grid daemons such as GridFTP, Globus-gatekeeper and the information system. The information system (MDS) uses the generic information provider [3] to collect status information of the cluster. The status of each cluster is sent to a central LDAP server and could be inserted into the AustrianGrid BDII. The Ganglia monitoring tool is used to get a summary of the load on all clusters and to quickly check which nodes did not start. The accounting log files of the resource management are collected on a central server. Keyboard or mouse activitiy is not checked because the clusters are used only during times when the rooms are closed for interactive usage.

The ZIDGrid software is exported via NFS for all machines in the domain uibk.ac.at and any set of Linux PCs can be turned into a grid cluster using a "gridify" script.

2.3. The startup and shutdown

The time tables with the opening hours of the rooms for the students are maintained manually via a web interface. This information is converted into tables containing the time intervals when the rooms

are open for Grid users. Two wake up on lan servers (for the two different locations) check these time tables for the clusters and start the clusters when necessary. The Grid opening hours can be looked up on the ZIDGrid homepage [5], it is planned to insert this information into the MDS.

During the boot of the PCs a Grid startup script is executed which looks up the hostname in a list to find out if the machine is part of the ZIDGrid and to which cluster it belongs at the moment. According to the configuration file of the cluster the startup script creates the files which have to be local on disk (e.g. for the local resource management) and starts all needed daemons depending on the type of the machine (master or worker node). In addition a cron job is established which checks the cluster's time table every 5 minutes and performs a shutdown when the time for the Grid usage is finished. Jobs which are not finished in time are terminated and repeated during the next Grid time window.

2.4. Job submission

Jobs can be submitted to the ZIDGrid Clusters using the commands of the Globus toolkit or the commands of the local resource management after login to the master with glogin [6], using Condor-G on the AustrianGrid portal machine of the ZID or using the ZIDGrid web interface (figure 1).

The ZIDGrid web interface uses the Gridsite software for authentication and starts parameter studies after the ZIDGrid resources become available. The parameter study is defined in a text file which is uploaded by the user together with a proxy certificate via https. Large output files can be generated on the local disk space of the worker node and transferred using Gridftp to the AustrianGrid portal machine or to an own Gridftp server. The syntax for the job definition file will be changed to the JDL-Syntax (job description language) of the EGEE project for parametric jobs [7].

3. Use Cases

The usability of the ZIDGrid is currently limited to applications which need less than 512 MB of RAM due to the current hardware equipment of the PC-labs; for future renewals and enlargement of the equipment the second usage as production cluster for scientific computing will be taken into account. The application software can be installed in the NFS-shared home directory or in the case of large installations in a dedicated area on the central file server, which is exported to all ZIDGrid machines.

3.1. High Energy Physics

The High Energy Physics group (HEPHY) at the University of Innsbruck collaborates with the ATLAS-Experiment [8] at CERN. The simulation of the response of the ATLAS detector to events in the detector is a very time consuming process, which is mainly done on the computing resources of the LCG (Large Hadron Collider Computing Grid). The group participates in LCG with the HEPHY-UIBK-Cluster [9].

In addition to the participation in the official ATLAS-Production on LCG and the LCG usage for the groups research, the HEPHY group tested the usability of the ZIDGrid for High Energy Physics applications. The ATLAS-software package (~5GB) was installed on the NFS-exported software area in a similar way as software is installed by software administrators of the virtual organization ATLAS on the supporting LCG-sites. One of the tests was a full detector simulation for 10000 events split
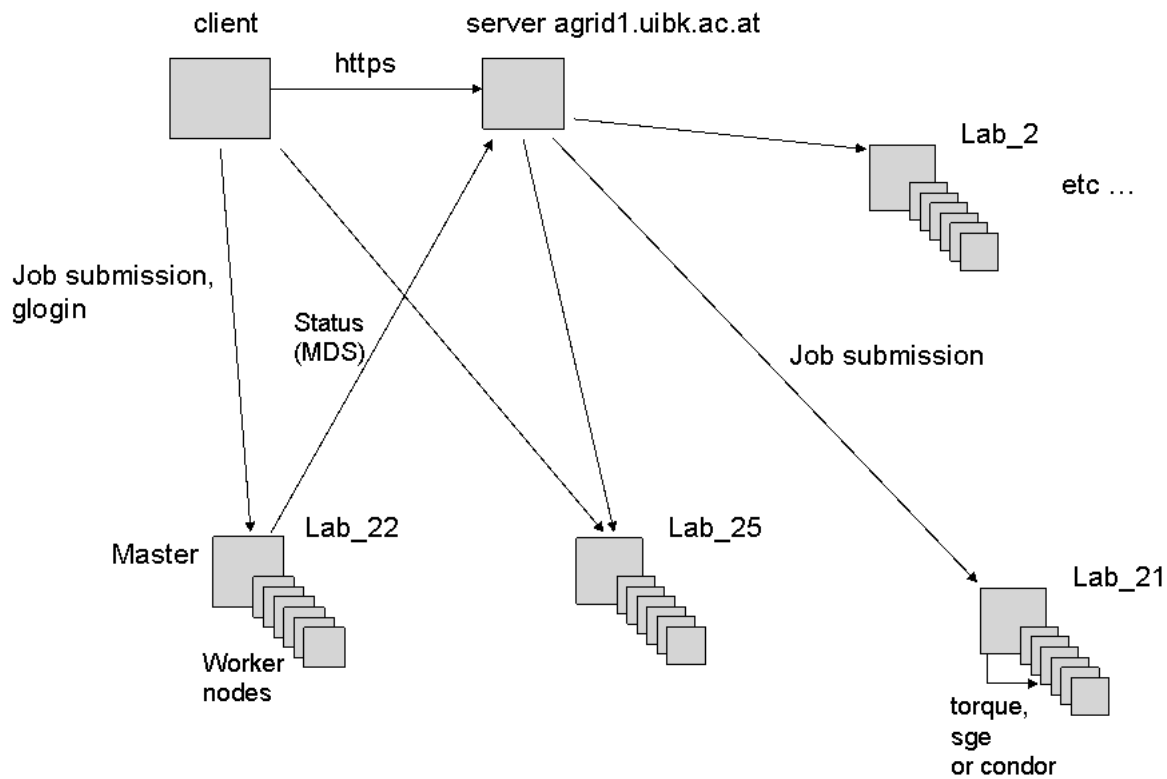
Figure 1. Job submission

into 100 jobs executed over a weekend. The job submission was done using the web interface for the ZIDGrid, which allows the definition of parametric jobs; in this case the arguments (e.g. input file) for all 100 jobs were defined in a text file which was uploaded together with a proxy certificate and a shell script via https. As soon as the ZIDGrid resources were available, the defined jobs were sent automatically to the resources.

The jobs copied their input file (~600MB) from a local storage element of the HEPHY group using GridFTP directly onto the worker node and stored the result (~100 MB) on the same storage element for further analysis.

The first try of this production showed a black hole effect, which had already been observed by ATLAS during the first production on the DataGrid testbed caused by misconfigured worker nodes. In our case a single machine did not have enough disk space for the input file, the job soon crashed and a lot of other jobs were attracted by the node. On another node the disk space containing the software package was not mounted. In the retry of the production the shell script for the jobs compared the time consumption of the job to the expected value, going to sleep for several hours if the result was negative thus isolating the node for some time.

A disadvantage during these tests was the unavailability of the LCG-specific commands for file transfers, which have meanwhile been installed on the ZIDGrid. Probably the HEPHY group will not use

the ZIDGrid for larger productions as they have access to the large LCG resources.

3.2. Institute of Engineering Mechanics

The Institute of Engineering Mechanics was the first institution at the University of Innsbruck interested in using the PC-labs for scientific computing, as they need a large number of Monte Carlo simulations for their research in Stochastic Structural Mechanics and Reliability and for a considerable part of these simulations the currently available RAM on the ZIDGrid machines is sufficient.

During summer 2004 they started with several hundred million Monte Carlo simulations for their project "Benchmark Study on Reliability Estimation in Higher Dimensions of Structural Systems" [10]. They continued in July 2006 with a reliability analysis of a finite element model of a satellite (INTErnational Gamma Ray Astrophysics Laboratory). 1 million simulations were done in 100 compiled Matlab jobs, each job running for approximately 10 hours. The output files were written locally on disk and copied from the worker nodes to their file server for further (less time consuming) analysis. In this case the usage of the ZIDGrid made it possible to finish the computation needed for a conference paper in time.

3.3. Department of Statistics

During the Christmas holidays 2005 and the semester holidays in February 2006 the ZIDGrid Clusters were heavily used by a production of the Department of Statistics (Faculty of Economics & Statistics), consisting of 6700 compiled Matlab jobs (Monte Carlo Simulation for Spatial Econometrics) with a run time per job from 4 to 40 hours, consuming in total 70.000 CPU-hours (calculated time on an actual server machine) [11]. Long running jobs were submitted when ZIDGrid Clusters were available for several days. The production on the ZIDGrid covered 60% of the complete task and the rest was done on the cluster of the High Performance Computing group. The supplementary usage of the ZIDGrid made it possible to finish the needed computations for a paper in time.
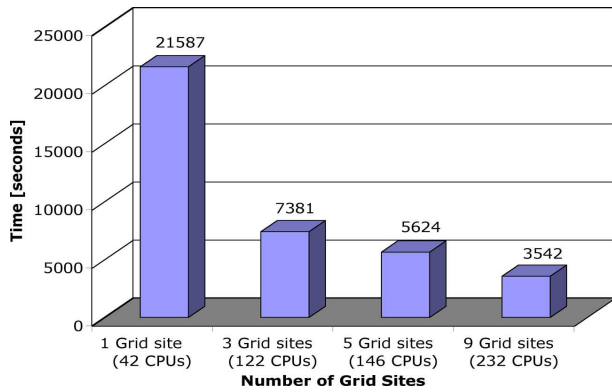
3.4. Middleware development (DPS)

The ZIDGrid resources are used by the Distributed and Parallel Systems Group who are developing ASKALON [12], a Grid Application Development and Computing Environment. They are using the ZIDGrid as part of the AustrianGrid as testbed for middleware development. One of their experiments [13] showed how powerful movie rendering using POV-Ray can be done on the Grid and how much the ZIDGrid can contribute to that (figure 2). The experiment used 9 heterogeneous Grid sites of the AustrianGrid with 232 CPUs. For this experiment the resources of the ZIDGrid have been split up into 4 Grid sites with a total of 86 CPUs (Pentium 4, 1.8 GHz, 512MB RAM). Providing 37% of the CPUs the ZIDGrid was able to render 29% of the movie.
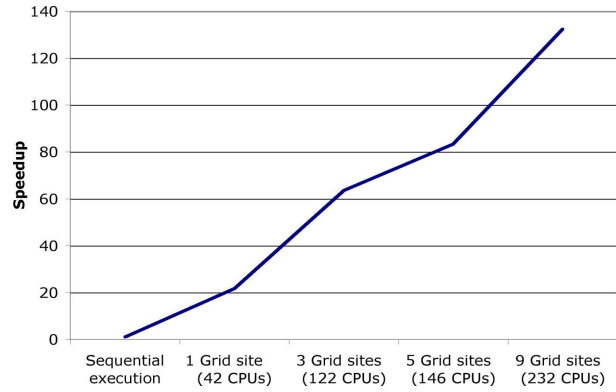
4. Ongoing and future work

Recently gLite software for user interface and worker node functionality has been installed for the ZIDGrid. The HEPHY group set up a resource broker node and a LCG-computing element, which will be used to test if the gLite middleware can be used for job submission to the ZIDGrid.
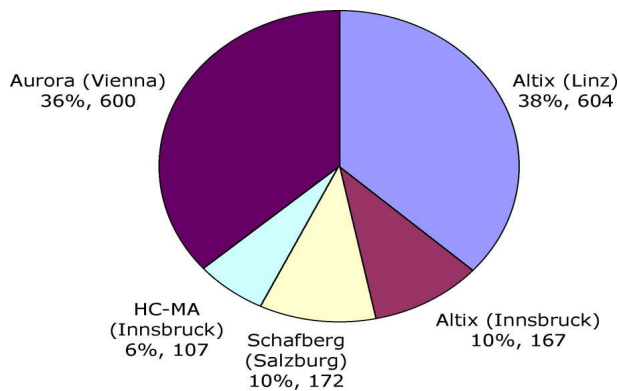
The HEPHY group started a co-operation with the Faculty of Biology to support computation for genetic research on the ZIDGrid and on EGEE-resources. Currently a co-operation between ZID
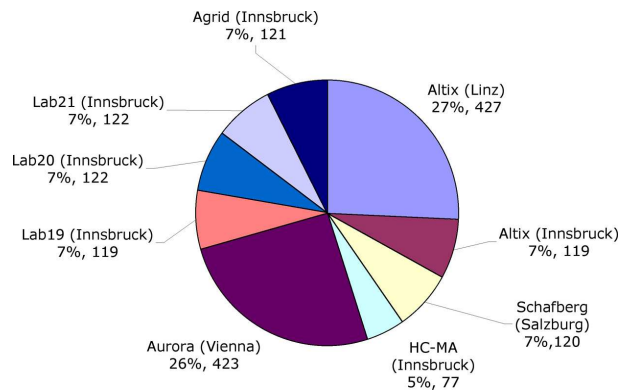
(a) Scalability.

(b) Speedup.

(c) Job distribution, 5 sites.

(d) Job distribution, 9 sites.

Figure 2. POV-Ray Experimental Results.

and the MCI (Management Center Innsbruck) is in preparation in order to make PC-labs at the MCI (approx. 200 CPUs) available for the AustrianGrid.

5. Conclusion

Several research groups at the University of Innsbruck have successfully used ZIDGrid clusters for large scale productions especially for time-consuming Monte Carlo simulations. In addition the flexibility of the setup allowed the usage as a testbed for middleware development.

We have found the ZIDGrid to be an interesting supplement to the production computing environment at the University of Innsbruck and will enhance the support for users to intensify the usage of the ZIDGrid resources.

6. Acknowledgement

References

[1] PIRANHA homepage http://wwwcs.uni-paderborn.de/pc2/index.php?id=72

[2] WINZIG description http://www.zid.tuwien.ac.at/zidline/zl14/grid.html

[3] Generic Information Provider homepage
http://lfield.home.cern.ch/lfield/cgibin/wiki.cgi?area=gip&page=documentation

[4] Gridsite homepage http://www.Gridsite.org/

[5] ZIDGrid homepage http://agrid.uibk.ac.at/zidgrid

[6] glogin homepage http://www.gup.uni-linz.ac.at/glogin/

[7] jdl parametric job
http://trinity.datamat.it/projects/EGEE/wiki/wiki.php?n=JDL.JDLTypes#parametric

[8] ATLAS-homepage http://atlasexperiment.org/

[9] Reinhard Bischof, Peter Oettl, Dietrich Liko and Gregor Mair, Large Scale Production on
the LCG seen by HEPHY-UIBK, 1st AustrianGrid Symposium, Österreichische Computer
Gesellschaft, Vienna 2006

[10] http://mechanik.uibk.ac.at/Benchmark_Reliability/

[11] M. Larch, G. Tappeiner, J. Walde, Performance Contest Between MLE and GMM for Huge
Spatial Autoregressive Models, Journal of Statistical Computation and Simulation, in press

[12] Askalon Homepage http://www.askalon.org/

[13] Francesco Nerieri, Mumtaz Siddiqui, Jürgen Hofer, Alex Villazon, Radu Prodan, Thomas
Fahringer, Using a Heterogeneous Service-oriented Grid Infrastructure for Movie Rendering,
1st AustrianGrid Symposium, Oesterreichische Computer Gesellschaft, Vienna 2006