

Technical Note: Kinds of cluster building/ grid environment with different resources for distributed computing in biomedicine

Fred Viezens¹

¹ Institute of Biometry and Medical Informatics, Otto-von-Guericke-University, Magdeburg, Germany

Abstract

The German Biomedical Grid Projects at the University Medical Centers in Magdeburg and Göttingen have developed environments for distributed computing in medicine, especially with heterogeneous computing resources in different infrastructures. At the beginning it was a stand-alone computer pool installation, it was translated into virtualized environment in a sub network and finally the installation with different types of worker nodes. The paper shows the approach and steps for the possibility of reproducing these experiments.

Keywords: *Intra-Grid, Virtualization, Hybrid Computers, High Performance Computing, Grid Computing.*

1. Introduction

For some biomedical applications [1] it is necessary to use more computing power as available with own capacities in a short time slot. Solutions are cluster systems or computing grids (see Fig. 1). Mostly of such distributed computer systems are very homogeneous in terms of architecture.



Fig. 1 Grid is a virtual computer architecture that's able to distribute process execution across a parallel infrastructure [2].

The described methods make the usage of heterogeneous systems possible. The solutions are also interesting in the point of possibility to move into different networks, so that internal business structures are not compromised.

Therefore the usage of latest applications in the field of biomedical imaging or the so-called “- omics” processing fields are possible [3] [4] [5]. Parts of investments for provider of computing infrastructure are refundable in such a way.

2. Development Steps

2.1 Grid Node in a Local Area Network as Client-Server Structure

The easiest way would be the solution with a Live-CD based on a Knoppix Distribution, which included all web services/ components of a Grid Middleware (see Fig. 2).

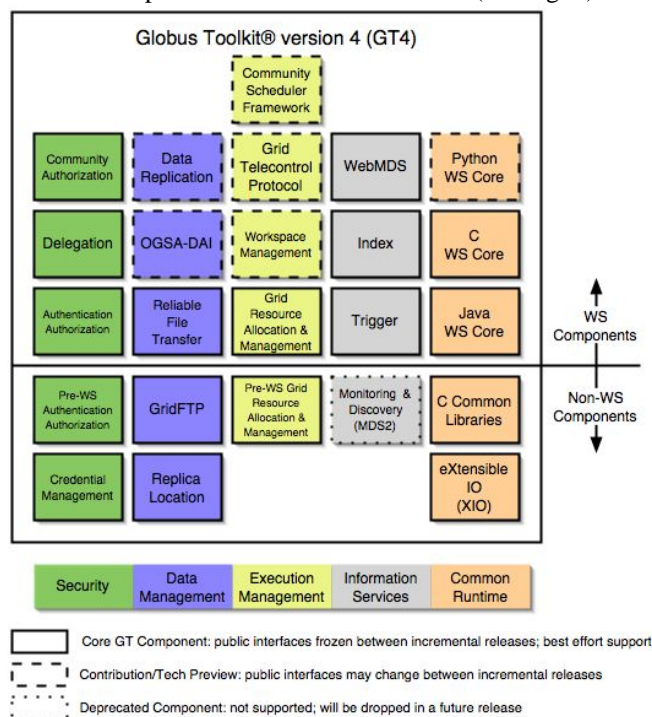


Fig. 2 Components of the Globus Toolkit v4 [6].

Such an environment recognizes existing components of a computer system or local network. In an existing Local Area network (LAN) based on client-server architecture, a grid system (Globus Toolkit 4) file structure can be rolled out and an IT system for distributed computing is ready to go. Assumption is that all computers are connected with network cable, allowing a Preboot eXecution Environment (PXE) boot process. With Instant-Grid [7] such a solution was experimentally implemented in a Computer Pool, the upper limit is a number of 256 computers inclusive front-end grid server (see Fig. 3).

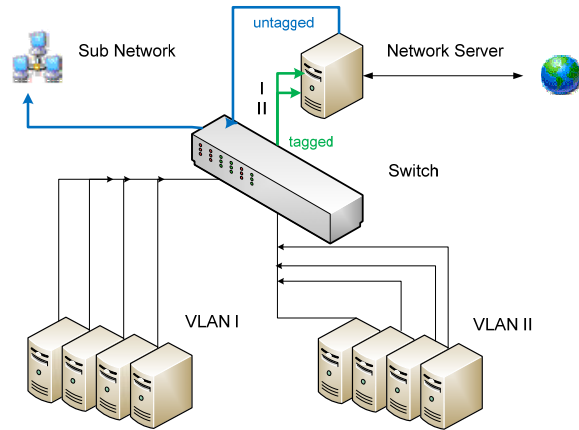


Fig. 3 VLAN/ NAT configuration [8].

The first booting computer is grid head node who manages all of the required grid services and distributed processes (jobs) to the other computing nodes in the network. All computers in such a network have then been booted, are Worker Nodes in terms of compute processing. Distributed Computing is now being possible or better the grid node is ready for compute and job/ resource allocation. A disadvantage is the temporary kind of this solution, in case of system restart, all input data, computed data and system preferences are deleting and nothing will be/ is stored permanently. One practice should be the academic education in the area of computer science and medical informatics. In laboratory practise [9] initial skills for the grid community can be acquired. The logically further development is virtualization and adaptation to transform a stand-alone internal grid into an external/ global grid node with separation from the internal network.

2.2 Virtualization of Grid Node - Integrated in Network Structures with external Connection

The first solution (I) was installed on a high dimensioned server (main storage > 64 GB) with technology of Virtual Machines (VMware) included three computing systems. The choice of operating system is not so important by

booting with Live - CDs, but in this special case one preference is necessary – server or head node was arranged with two Network Interface Cards (NIC). One NIC is used for internal addressing with Dynamic Host Configuration Protocol (DHCP) and the other NIC realized the external communication over Network Address Translation (NAT) - separation of each communication way. Next numeration shows the final Start Sequence:

1. boot front end and DHCP server; assign *eth0* as network interface
2. boot clients, related information via Trivial File Transfer Protocol (TFTP)
3. second NIC *eth1* of the front end activate over the configuration menu of VMware
4. input *pump -i eth1* (obtain IP address for internet access) in the server shell

To solve the automatic IP assignment of the DHCP server during the boot process of the clients, the host configuration file in the directory */etc/hosts* was adjusted and via command *ifconfig* the correct IP number has been re-defined [10] with an editor program so that was linked on the respective machines. An interesting aspect of the experiment was the possibility of moving the machines in a Storage Area Network (SAN). Furthermore there could be constructed subnets and operated within or outside a Demilitarized Zone (DMZ) of the institution.

Practise Example

Applications so jobs can be send in Instant-Grid with the service Web Services Grid Resource Allocation and Management (WS GRAM); kept simple with the command line program **globusrun -ws**. In the example would pass only one parameter:

```
globusrun -ws -submit -f job.xml
```

Thereby constitute the file job.xml a job description with following structure:

```
< job >
< executable > /bin/cat < /executable >
< argument > /etc/motd < /argument >
< stdout > /tmp/stdout < /stdout >
< /job >
```

These job calls the program under `/bin/cat` with `/etc/motd` as parameter and the file `/tmp/stdout` as standard output file. It will be read and stored the message of the day. Alternative would be the following job call:

```
globusrun - ws - submit - c
/bin/cat /etc/motd - so /tmp/stdout
```

Analogue to (stdout) let be redirected the standard input file (stdin) and standard error output file (stderr). It is possible to specify multiple tags argument. Then it can be send a job, but it runs only on the front end. By sending jobs to a different computer can be used the command-line parameter `-F`:

```
globusrun - ws - submit - f job.xml - F
https : // ${KNOTEN} : 8443 / wsr / services
/ManagedJobFactoryService
```

Job Script:

```
01 : #!/bin/sh
02 :
03 : HOSTS = cat /usr/bin/machines
04 : OUTDIR = /clusterwork/mo
05 :
06 : for HOST in $HOSTS
07 : do
08 : cat << EOF > job.xml
09 : < job >
10 : < executable > /bin/hostname < /executable >
11 : < stdout > $OUTDIR/hostnames < /stdout >
12 : < factoryEndpoint >
13 : < wsa : Address >
14 : https : // $HOST : 8443 / wsr / services / ManagedJobFactoryService
15 : < /wsa : Address >
16 : < /factoryEndpoint >
17 : < /job >
18 : EOF
19 : echo "starting job on $HOST"...
20 : globusrun - ws - submit - f job.xml
21 :
22 : done
23 :
24 : echo "ready.now displaying $OUTDIR/hostnames :"
25 : cat $OUTDIR/hostnames
```

Script 1 Grid-Job executing on available Worker Nodes.

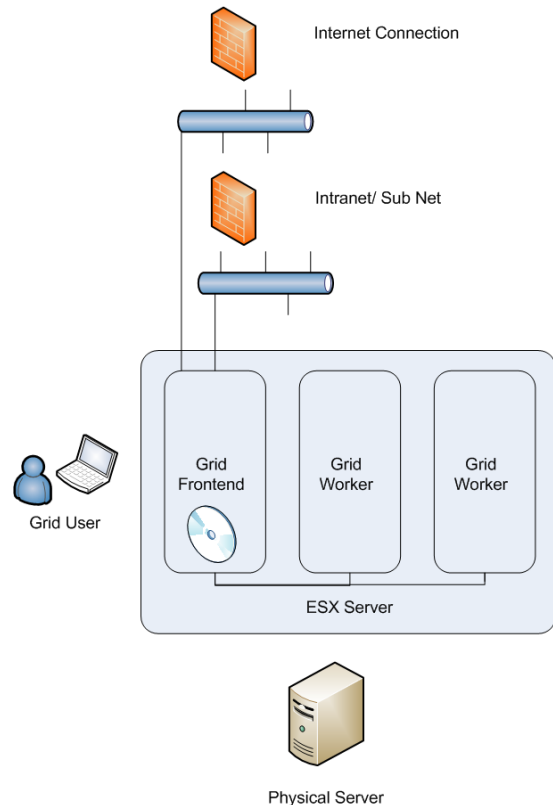


Fig. 4 Virtualization with ESX Server and bootable Instant-Grid Image.

The advantage of virtualization is the possibility to store each update of the grid system or the proceeded data, e.g. downloading and installing of additional software packages for grid services application development like portal software (Gridsphere [11]).

2.3 Integration of an Apple XServer Cluster into a Grid with heterogeneous Operating Systems

Software stack for grid middleware Globus Toolkit with Linux operating system was installed at existing Enclosure of Blade Technology with more than one hundred processor cores and storage capacity of approximate twenty Tera Byte. At that Apple XServers with more than sixty processors and the BSD - Unix based Mac OS X 10.5 should be included. In a Ratio of 1:1 it isn't able to install the equal software stack on both architectures. The conventional installation process aborted with error messages.

```
ibmi-mac63925:admin root# cd /var/root/maui-3.2.6p19/
ibmi-mac63925:maui-3.2.6p19 root# ./configure
checking for gcc... gcc
checking for C compiler default output... a.out
checking whether the C compiler works... yes
checking whether we are cross compiling... no
checking for suffix of executables...
checking for suffix of object files... o
checking whether we are using the GNU C compiler... yes
checking whether gcc accepts -g... yes
checking for gcc option to accept ANSI C... none needed
checking whether ln -s works... yes
configure: WARNING: Unsupported OS: Darwin, attempting build with OPSYS=_LINUX
configure: WARNING: Resource Manager not specified - attempting build with pbs
configure: error: can't find pbs-config or libpbs.a
```

Screenshot 1 Error Message Maui (scheduler) Installation.

```
-rw-r--r-- 1 root admin 1820 16 Apr 18:35 torque_server.sh
ibmi-mac63925:installation root# rpm -ihv torque-*.rpm
Preparing... [100%]
1:torque [33%]
/var/local/tmp/rpm-tmp.58431: line 1: /bin/grep: No such file or directory
2:torque-client [67%]
3:torque-server [100%]
/var/local/tmp/rpm-tmp.75238: line 1: /sbin/chkconfig: No such file or directory
error: %post(torque-server-2.1.6-1crl_2dgrid_sl4.x86_64) scriptlet failed, exit
status 127
ibmi-mac63925:installation root# rpm -ihv maui-*.rpm
ibmi-mac63925:installation root# rpm -ihv maui-*.rpm
Preparing... [100%]
1:maui [33%]
2:maui-client [67%]
3:maui-server [100%]
/var/local/tmp/rpm-tmp.72778: line 2: /sbin/chkconfig: No such file or directory
/var/local/tmp/rpm-tmp.72778: line 3: /sbin/chkconfig: No such file or directory
error: %post(maui-server-3.2.6p19-snap1171482917_1dgrid_sl4.x86_64) scriptlet fa
iled, exit status 127
ibmi-mac63925:installation root#
```

Screenshot 2 Error Message Maui and Torque (batch system) with RPM.pkg support.

With the aid of the application packages *Xcode*, *Fink* and *RPM* for Mac OS X was it possible to compile and execute torque, a part of this stack. Torque is the batch system also used in the enclosure. With the Scheduler (*maui*) at the grid head node of the enclosure computing jobs are able to run on both platforms with the especially environment. The IP Addresses of the Apple XServers have to be written at the enclosure grid node in the *pbs_mom*-file (Portable Batch System). The *pbs_server* listened at signals of the *pbs_mom* clients running on the xservers. The *pbs_server* have to written in *var/spool/torque* at client side. Logging messages are written in */spool/pbs/mom.logs*. With the command line instruction *pbsnodes -a* (server side) the log report shows the status information of all worker nodes. The attribute *opsys* showed in this case Linux or Darwin (available nodes with the kind of Operating System) and respectively the machine name of the worker nodes.

3. Results

At the network file system (NFS) two installations of an image processing application [12] in the case Linux and Mac OS compilations are stored. With bash scripts queried arguments, such as:

```
$OS = uname a | cut -d" " | f1
or = uname -a | awk (print $1)
if ($OS -/ - "Darwin")
./set_env_osx
fi else
./set_env_linux
```

or arguments (see 2-3) could depending vary availability the Linux or the Mac OS version to execute the scheduled job. The following arguments are possible:

```
qsub option argument (nodes/ opsys)
nodes = machine name
opsys = Linux/ Darwin
```

The arguments (see 2) realized a mapping between application, resources and operating system. Numeral 3 shows an example for running an interactive job on the particular platform. Note: This is only possible under the grid user; the root user can't operate this.

```
qsub I (interactive) q (job queue) l arguments
```

The following terminal output shows such a workflow, a script to calculate a statistic program written in R:

```
fviezens@medinfogrid9b:~> qsub -I -q dgiseq -l nodes = ibmi - mac63930
qsub : waiting for job 225.medinfogrid9b to start
./mnt/opt/mac/r_skript
qsub : job 225.medinfogrid9b ready
./mnt/opt/mac/r_skript
ibmi - mac63930:~ fviezens$ ./mnt/opt/mac/r_skript
```

Some computer centers adopt this technology with a small dimensioned Unix Server to managed a cluster heterogeneously nature with job submission on command line.

4. Conclusions

It is possible to create grid computing environments in each IT Infrastructure, e.g. intra-, extra- and global IT architectures. The logical and physical disconnection allows distributed computing also in such sensitive areas like the health care sector. The possibility of sharing and processing data as an additional benefit in points medical care and research opens new ways for collaboration and utilization/ provision existing compute resources.

Acknowledgments

This work was supported from the German Federal Ministry of Education and Research (BMBF) in the Projects Instant-Grid, MediGRID and MedInfoGRID, grants 01AK807, 01AK803 and 01G07016.

References

- [1] MediGRID-Applikationsportal, "MediGRID-Applikationsportal," 2006.
- [2] pug, "Rechenleistung soll wie Strom aus der Steckdose fließen - Göttinger Zentrum will innovative Netzwerktechnologien für Grid Computing entwickeln," in Göttinger Tageblatt, vol. 119. Göttingen, 2008, pp. 22.
- [3] F. Viezens, "Grid-Computing in der Biomedizin," in Grid-Computing in der Biomedizinischen Forschung – Datenschutz und Datensicherheit, vol. 90, Medizinische Informatik, Biometrie und Epidemiologie, U. Sax, Y. Mohammed, F. Viezens, and O. Rienhoff, Eds. München: Urban&Vogel, 2006, pp. 56-62.
- [4] R. Beisse, M. Bettag, H. Gassen, W. Höppner, F. Koch, S. Nikol, D. Schmidt, D. Schnorr, A. Schrattenholz, M. Schumacher, and W. Siebert, Medizin im 21.Jahrhundert, Laubach, E.,Mau, F.,Mau, Th. ed: Springer-Verlag Berlin Heidelberg New York, 2002.
- [5] S. Kottha, K. Peter, T. Steinke, J. Bart, J. Falkner, A. Weisbecker, F. Viezens, Y. Mohammed, U. Sax, A. Hoheisel, T. Ernst, D. Sommerfeld, D. Krefting, and M. Vossberg, "Medical Image Processing in MediGRID," presented at German e-Science Conference, Baden-Baden, 2007.
- [6] GT4, "The Globus Toolkit 4 Programmer's Tutorial," 2006.
- [7] C. Boehme, A. Félix, B. Neumair, and U. Schwarzmann, "Instant-Grid: Demonstration, Entwicklung und Test von Grid-Anwendungen," in GWDG-Nachrichten, vol. 29, Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen ed, 2006, pp. 5-13.
- [8] T. Rings, A. Aschenbrenner, J. Grabowski, T. Kalman, G. Lauer, J. Meyer, A. Quadt, U. Sax, and F. Viezens, "An Interdisciplinary Practical Course on the Application of Grid Computing," presented at 1st Annual IEEE Engineering Education Conference – The Future of Global Learning in Engineering Education (EDUCON 2010), Madrid, Spain, 2010.
- [9] T. Rings, F. Viezens, J. Meyer, and A. Aschenbrenner, "Ein interdisziplinäres Grid-Anwenderpraktikum basierend auf Instant-Grid," GWDG-Bericht, pp. 19-28, 2009.
- [10] F. Viezens, A. Barz, and K. Lorberg, "Pseudonymisierung von Daten in einem Grid," presented at 52. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie., Augsburg, 2007.
- [11] GridSphere, "GridSphere," 2006.
- [12] Oxford, "FMRIB Software Library," Analysis Group FMRIB, 2009.

Consulting and Software Development. 2001-2002 Research Assistant at the Otto-von-Guericke-University Magdeburg, Medical Faculty. 2003 Research Assistant in the MBR Computing Centre GmbH, Magdeburg. 2004-2005 Research Assistant at the Otto-von-Guericke University Magdeburg, Faculty of Mechanical Engineering. 2006-2008 Research Assistant at the Georg-August University Göttingen, Department of Medical Informatics at the University Medical Center. Since 2008 Research Assistant at the Otto-von-Guericke University Magdeburg, Medical Faculty. His Research Interest are located on Security Mechanisms, Smartcard Technology, Distributed Computing, Service Oriented Architecture and Computer-Integrated Manufacturing.

Fred Viezens is a member of the IEEE (German Section), studied 1988-1993 Informatics at the Technical University „Otto-von-Guericke“, Magdeburg. 1994-1997 Lecturer at the DEKRA Academy GmbH, Area: Logistics. 1997-2001 Freelance Activity, IT