FORSCHUNGSZENTRUM JÜLICH GmbH
Zentralinstitut für Angewandte Mathematik
D-52425 Jülich, Tel. (02461) 61-6402

Technical Report

Core D-Grid Infrastructure

Thomas Fieseler, Wolfgang Gürich

FZJ-ZAM-IB-2007-09

July 2007
(last change: 19.07.2007)

Preprint: to be published

Core D-Grid Infrastructure

T. Fieseler, W. Gürich

Zentralinstitut für Angewandte Mathematik, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

e-mail: t.fieseler@fz-juelich.de, w.guerich@fz-juelich.de

Abstract

D-Grid is a German implementation of a grid, granted by the German Federal Ministry of Education and Research. In this paper we present the Core D-Grid which acts as a condensation nucleus to build a production grid infrastructure. The main difference compared to other international grid initiatives is the support of three middleware systems, namely LCG/gLite, Globus, and UNICORE for compute resources. Storage resources are connected via SRM/dCache and OGSA-DAI. In contrast to homogeneous communities, the partners in Core D-Grid have different missions and backgrounds (computing centers, universities, research centers), providing heterogeneous hardware from single processors to high performance supercomputing systems with different operating systems. We present methods provided by the Core D-Grid to integrate these resources and services for the infrastructure like a point of information, centralized user and VO management, resource registration, software provisioning, and policies for the implementation (firewalls, certificates, user mapping).

1 Introduction to D-Grid

In September 2005, the German Federal Ministry of Education and Research started six community grid projects and an integration project to build up a sustainable grid infrastructure in Germany. More grid projects in the same context will follow to join the common infrastructure. The initial community projects are high energy physics (HEPCG) [1], astrophysics (AstroGrid-D) [2], medicine and life sciences (MediGRID) [3], climate research (C3Grid) [4], engineering (InGrid) [5], and humanities (TextGrid) [6]. The first additional partner project that joined the evolving infrastructure is energy meteorology (WISENT) [7].

The community projects are heterogeneous concerning the scientific field, the structure of the community, the structure and size of data being processed, the type of grid software in use, and the experience with grid computing. Despite all differences these communities are united by their common interest in grid methods for the solution of their scientific computation challenges. Some communities like high energy physics have wide experience with grid computing (HEPCG, AstroGrid-D), while others are just starting to apply the grid approach to their computational tasks (TextGrid). Some of the communities which already applied grid computing intensively have a strong affinity to use a certain middleware (HEPCG / gLite, AstroGrid-D / Globus), while communities with less experience are still open in the choice of the middleware. The

requirements of the communities in the grid middleware are highly variable, e.g. in applications of the HEPCG or AstroGrid-D comparatively few but very large data transfers are needed, while applications of TextGrid tend to have many transfers of small data sets.

In order to build up a common basic grid infrastructure for these heterogenous grid communities the integration project has been started. The goal of the integration project is to build up a general, sustainable grid infrastructure, the Core D-Grid, first as a testbed and later as the productive environment for the grid communities.

In this paper, the structure of the integration project, its partners and resources, the supported middleware, and methods to integrate the resources into the Core D-Grid are presented. Services for the infrastructure like a point of information, a centralized user and VO management, a centralized resource management, software provisioning, and policies for the implementation (firewalls, certificates, user mapping) are described.

2 D-Grid integration project

Partners and resources

The integration project started with seven partners who contribute their own compute and storage resources and three associated partners without a contribution of resources, but technical know-how and a strong interest in grid computing. The background and the working areas of these partner institutions are highly heterogeneous, as the resources they initially contributed: Partners are the computing centers of the national research centers and of universities; the compute resources vary from large supercomputers with several TFlops peak performance to small clusters with only a few CPUs, and have different operation systems like AIX, Solaris and various Linux flavors.

Work packages

The D-Grid integration project (DGI) is divided into the following four work packages:

1. Basic grid software
2. Setup and operation of the D-Grid infrastructure
3. Network and security
4. Project management

Basic grid software

In the software section of the integration project, the basic grid middleware and further basic grid software is packaged and made available for the resource providers, grid developers, and grid users. Unlike other large grid projects as EGEE, which are mainly based on a single middleware, the requirements of the diverse D-Grid projcts are too different to rely on a single grid middleware. Therefore, three middleware systems for compute resources are supported in

the software stack of the integration project: LCG/gLite [8, 9], Globus (version 4) [10], and UNICORE [11] (version 5). For storage resources, SRM/dCache /[12] and OGSA-DAI [13] are supported by the integration project. Furthermore, GridSphere [14] is provided to implement portal solutions, the Grid Application Toolbox [15] is supported for application level programming.

Setup and operation of the D-Grid infrastructure

The second section of the integration project is the setup and operation of the Core D-Grid, which acts as a condensation nucleus to build a production grid infrastructure. The infrastructure of the Core D-Grid is described in detail in chapter 3.

Network and security

In the third part, networking issues, security aspects, and firewewalls are covered. D-Grid is based on the X-WiN network which is run by the Deutsche Forschungsnetz (DFN), who is coordinating this work package. Some of the partners already have 10 Gbit/s connections to other partners. The extension of the network infrastructure according to the upcoming requirements of partners and communities is coordinated in this work package. Alternative transport protocols are tested, compared to standard TCP, and optimized. Security aspects of grid middleware and firewalls are considered and advice is given to the partners in the Core D-Grid and the D-Grid community projects. Furthermore, grid security aspects like authentication and authorization are investigated in this part.

Project management

The last section covers the leadership and coordination of all four parts of the infrastructure project and the coordination of the collaboration with the D-Grid community projects. Furthermore, dissemination and legal and organizational questions are part of this package in order to create a sustainable infrastructure for e-science in Germany.

3 Core D-Grid infrastructure

For the operation of the Core D-Grid, different infrastructure components are required like a certificate infrastructure, a concept to install the three middlewares on one machine, a user and resource management system, resource monitoring, user support, and a point of information.

Certificates

The security of all three middleware systems is based on PKI and X.509 certificates. In Germany there are two certificate authorities for grid certificates, the Deutsche Forschungsnetz (DFN) [16] and the Forschungszentrum Karlsruhe (GridKA) [17] which have been accredited by the EUGridPMA [18]. For both certificate authorities many registration authorities have been approved. All

partners of the Core D-Grid and the community projects have setup registration authorities to enable an easy access of users and administrators to grid certificates. Since most of the D-Grid projects are parts of international communities, foreign grid user certificates issued by any certificate authority accredited by EUGridPMA [18] and IGTF [19] are accepted.

Resources financed by additional funding

The hardware which has initially been provided by the partners of the DGI was highly heterogeneous. The installation of grid middleware on less frequent platforms (e.g. Globus on AIX) and the integration of this hardware into the upcoming grid infrastructure was complicated but helpful to gain experience with different systems. At the end of 2006, the German Federal Ministry of Education and Research decided to invest additional funds for compute and storage resources located at partner sites of the Core D-Grid and the D-Grid community projects to serve as an additional incentive of the upcoming infrastructure. The additional funding was combined with the obligation to install all three middlewares for compute resources (gLite, Globus, UNICORE) in parallel on each of the new compute resources. All of the compute-nodes of these clusters (about 20 clusters have been aqcuired) must be accessible via each of the three middlewares. Furthermore, at least one of the two middlewares for storage access (SRM/dCache, OGSA-DAI) must be installed on the storage resources. Access to these additional resources must be granted to all virtual organizations (VOs) of D-Grid.

Reference installation

The request to install the complete middleware stack of the DGI on a single resource presently is a very demanding challenge, since the different middleware systems partially have very restrictive and mutually exclusive requirements (e.g. Scientific Linux for gLite worker-nodes (WN) and even more restrictive Scientific Linux 3.0.x for the compute element (CE) on the one hand, the most up-to-date packages for Globus 4.0.x on the other hand). Since any solution to this problem is highly complicated, a reference installation, realizing the simultaneous installation of all supported middlewares systems has been set up [20]. This reference installation demonstrates how to run the different middleware systems on the same machine with access to all compute-nodes by each of the three middlewares for compute resources (see figure 1).

For each of the grid middleware systems (LCG/gLite, Globus, UNICORE, OGSA-DAI, SRM/dCache) the reference installation provides a dedicated so-called head-node for the installation of the server side of the middleware. The operation system of the head-nodes for Globus, UNICORE, and OGSA-DAI is SLES 10. The OS of the SRM/dCache head-node is Scientific Linux 4, whereas the OS of the head-node for the LCG/gLite compute element (CE) is Scientific Linux 3 (32bit) which is running on a Xen virtual machine under SLES 10. On the Globus head-node Globus Toolkit 4 is installed and the UNICORE head-node runs TSI, NJS, and UUDB. On the LCG/gLite head-node, the LCG-CE variant (production version) of the compute element is used. OGSA-DAI is installed together with Globus Toolkit 4 on the OGSA-DAI head-node. The

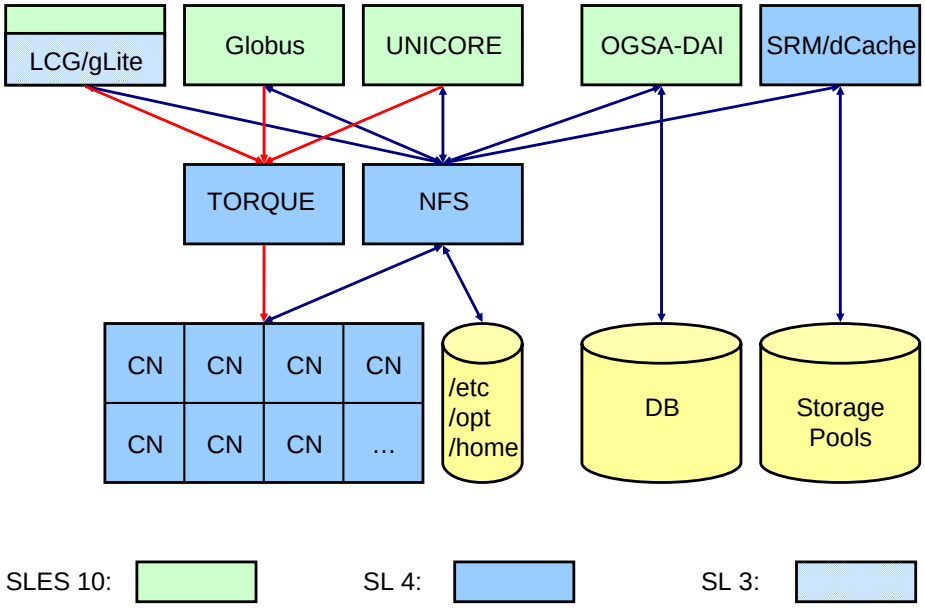

Figure 1: Architecture of the reference installation. Jobs can be submitted via the head-nodes for LCG/gLite, Globus, and UNICORE to the TORQUE batch system which can access all compute-nodes (CN). Storage resources can be be accessed by the head-nodes for OGSA-DAI and SRM/dCache. The NFS node provides a common file system for configuration files (e.g. CA certificates), software, and home directories. The color of the nodes denotes the operating system.

SRM/dCache head-node runs dCache 1.0.7 of the LCG distribution. Two further dedicated nodes are used for NFS which exports common directories like the certificates of the certificate authorities, the gLite user interface (UI) software, and the home directory for grid users, and a node for the server of the batch system TORQUE 2.1.6. The batch system server node is running under Scientific Linux 4. All three middleware installations for compute resources (LCG/gLite, Globus, UNICORE) connect to the same batch system. Therefore, all compute-nodes of the cluster can be access by all the middleware systems. Another special node is dedicated for interactive use and can be accessed remotely by D-Grid developers and users via GSI-SSH or UNICORE-SSH. This node has the same environment (OS and configuration) as the compute-nodes and can be used for software development and testing purposes. All other nodes are compute-nodes running under Scientific Linux 4. On the compute-nodes, the client part of the batch system (TORQUE) and the gLite worker-node (WN) software are installed. To submit further jobs within a running grid job, the gLite user interface (UI) can be accessed from the compute-nodes via NFS.

Specially pre-configured packages of the middlware for the use within D-Grid are provided to ease the middleware installation and configuration for the partners. The recipients of the financial support for resources do not have to follow the exact way of the installation of the reference system. But even if the individual installations may differ according to the requirement of the local environment of the resource providers, the functionality must be the same as in the reference

installation (access by all middlewares, access for users of all D-Grid VOs).

User and VO management

With an increasing number of communities, virtual organizations (VOs), and resources, an automated management of users and VOs on one side and of the resources on the other site are required to operate the grid infrastructure. The creation of a new VO is not an automated process in the present state of the integration project. Presently, this constraint is not a real problem, since the number of VOs is still manageable (7 VOs for the community projects and 2 VOs for testing and administration purposes). The creation of a new VO must be agreed between the managements of the community project to which the VO is related and of the integration project; one or more representatives of the new VO must nominated etc. For each VO, an own instance of a virtual organization membership registration service (VOMRS) [21] server is installed and configured on a centralized D-Grid management server. A new D-Grid user must find a VO which is appropriate for his field of research. One of the representatives of the VO in question must agree with the membership of the new member. If these requirements are fulfilled the new member can register to the VOMRS server of the VO and will obtain access to all resources of this VO.

Resource management

In order to be integrated into the D-Grid infrastructure, each resource has to be registered at the grid resource registry service (GRRS) server, which has been developed within the D-Grid integration project (see figure 2). During the registration process of a resource at the GRRS server, all relevant information of the compute or storage resource and its administrators is collected and the grid server certificate of the resource is uploaded to the GRRS and stored in its database. For compute and storage resources which have several middlewares (gLite, Globus, UNICORE, SRM/dCache, OGSA-DAI) installed simultaneously, the resource has to be registered for each middleware with the grid server certificate of the head-node of the respective middleware. For the administration of the resources, a client (dgridmap) is distributed to the resource providers which must be run regularly on the resource. For resources with more than one middleware, the dgridmap client must executed on the head-nodes for each of the middleware systems. The client contacts the GRRS server of the D-Grid resource management, authorizing itself with the grid server certificate for the respective resource (head-node). On the server side, the VOs which are allowed to access the resource are determined as entries in the GRRS database. The corresponding VOMRS servers of these VOs are queried to provide the user information (DNs of the grid user certificates, ID of the user in this VO, etc.) of the members of the VOs. The result is the user mapping for the corresponding resource in the format which is appropriate for the middleware (e.g. grid-mapfile for Globus, entries for uudb_admin for UNICORE). The naming convention of the unix accounts in the user mapping is *ppvvnnnn*, where *pp* is a prefix (default *dg*) which can be changed by the administrator according to the local requirements, *vv* is a short-cut for the VO of the user entry, and *nnnn* is a number which is unique for this combination of user DN and VO. A member

of a VO thus is mapped to accounts with name *ppvvnnnn* on all the resources of this VO, apart from the prefix *pp* which may vary at the provider sites.

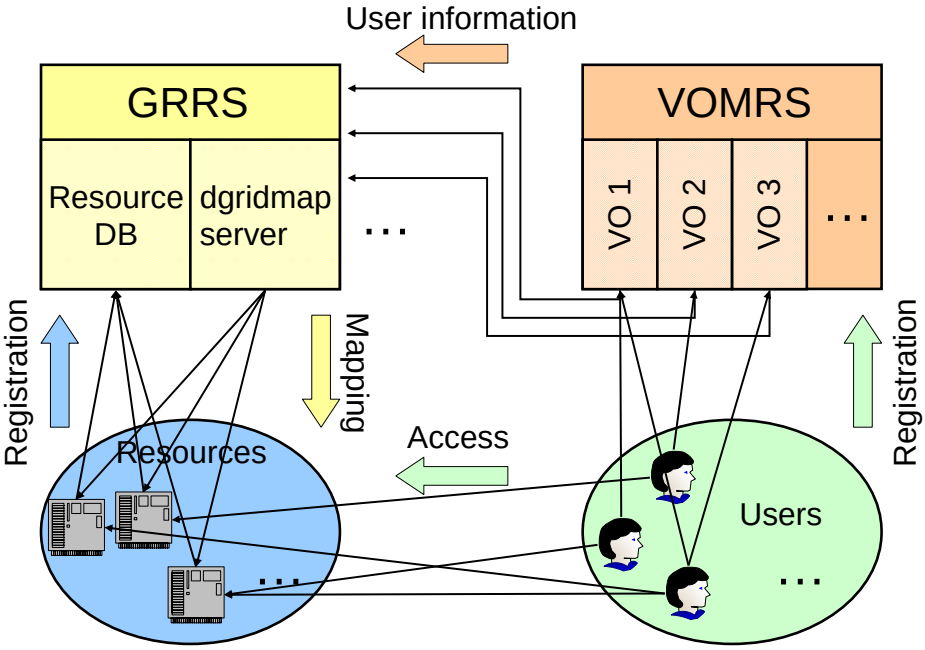

Figure 2: Structure of the D-Grid user, VO, and resource management. In order to obtain access to the grid resources, users must register to a VO using the VOMRS server for this VO. New resources must register at the GRRS server to be integrated into the resource management. The user mapping for the resources is generated by the GRRS server which in turn obtains the information about uers of the VOs from the VOMRS servers.

Monitoring

On the LCG/gLite resources of the Core D-Grid infrastructure, LCG/gLite functional site tests (SFT) are performed. D-Grid users can inspect the SFT reports and find out which resources are available.

For UNICORE, the monitoring abilities of the UNICORE client can be used, i.e. D-Grid users can observe which resources are registered at the gateway and if the user has access to a resource with his certificate. Within the UNICORE client, additional information about the resources, as number of nodes, the number of CPUs per node, memory etc., and the jobs of the user can be gathered.

The Globus resources are monitored with MDS. On each Globus resource, the MDS4 software and a sensor transmitting additional information like the geographical coordinates of the site, the schedule of maintenance periods etc. have been installed. Each D-Grid resource provider is running a MDS index server for the site, collecting the MDS information of all resources of this site. The site index servers upload their information to a central D-Grid Web-MDS server, where D-Grid users can obtain the monitoring information in a hierarchical view, according to the organization levels of D-Grid. Furthermore, the

resources can be displayed in a topological map.

User support

A trouble ticket system similar as the system of the EGEE project has been installed to organize the user support. In the user support center, tickets are handled and forwarded to the next level of the user support, depending on the area of the user request. For community specific requests, each community must setup and operate an own consulting process. The partner sites must run a user support for requests concerning their site, and the integration project operates a user support for grid middleware specific and grid infrastructure specific requests.

Point of information

The point of information (POI) is divided into different sections, a section with general information about D-Grid and the community projects [22], a section with information about the integration project [23], a user portal [24], and a provider portal [25]. The user portal is intended as a first starting point for D-Grid users. Users can find information about middleware services, i.e. installation and usage of grid middleware clients of the middleware systems that are supported within D-Grid, about the resources which can be accessed in the Core D-Grid, information how the user can get access to the resources like grid user certificates, creation of a new virtual organization, membership in an existing virtual organization, the status of the resources (Globus: WebMDS, LCG/gLite: SFT) and a link to the trouble ticket system of the user support. In the provider portal, resource providers can find the information that is needed to integrate a new resource into the Core D-Grid, as information about grid server certificates, installation and configuration of the grid middleware servers according to the reference installation, information about ports to be opened in the firewalls, integration of the compute or storage resource into the grid resource registry service (GRRS), and the integration of the resource into the D-Grid monitoring system.

4 D-Grid integration project 2

The first phase of the integration project ends in September of 2007. A subsequent second phase of the integration project (DGI-2) is planned with a duration three years. While the integration of new resources was the major task of the Core D-Grid during the first phase, the second phase will be more focused on the consolidation and extension of the developing infrastructure. The services of the Core D-Grid as VO and user administration, resource management (GRRS), user support, monitoring etc. will be further improved and a failsafe infrastructure will be established with redundant servers for the core services. Furthermore, accounting solutions which have been acquired within the integration project, will be integrated into the Core D-Grid infrastructure. Improved versions of policies of grid users, virtual organizations, and resource providers will be developed and published. In this phase an increasing number of new

grid communities is expected to join the infrastructure and therewith a large number of new virtual organizations and resources is expected to be integrated.

Acknowledgements

The D-Grid integration project is completely funded by the German Federal Ministry of Education and Research. The integration project involves a large number of colleagues who all contribute to the project and the development of the Core D-Grid infrastructure.

References

[1] High Energy Physics Community Grid (HEPCG) documentation, http://documentation.hepcg.org.

[2] AstroGrid-D, German Astronomy Community Grid (GACG), http://www.gac-grid.de.

[3] Grid Computing for Medicine and Life Sciences (MediGRID), http://www.medigrid.de.

[4] Collaborative Climate Community Data and Processing Grid (C3Grid), http://www.c3grid.de.

[5] Innovative Grid Technology in Engineering, http://www.ingrid-info.de.

[6] Modular platform for collaborative textual editing – a community grid for the humanities (TextGrid) http://www.textgrid.de.

[7] Scientific network for energy meteorology (WISENT), http://wisent.offis.de.

[8] J. Knobloch and L. Robertson. LHC Computing Grid. The LCG TDR Editorial Board, December 2006, http://lcg.web.cern.ch/LCG/tdr/LCG_TDR_v1_04.pdf.

[9] R. Berlich, M. Kunze, and K. Schwarz. Grid Computing in Europe: From Research to Deployment. Proceedings of the 2005 Australasian workshop on Grid computing and e-research, Newcastle, New South Wales, Australia, Vol.44, pp. 21 - 27, 2005.

[10] I. Foster. Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp. 2-13, 2006.

[11] A. Streit, D. Erwin, T. Lippert, D. Mallmann, R. Menday, M. Rambadt, M. Riedel, M. Romberg, B. Schuller, and P. Wieder. UNICORE - From Project Results to Production Grids. Grid Computing: New Frontiers of High Performance Computing, L. Grandinetti ed., Elsevier, pp. 357 - 376, 2005..

[12] T. Perelmutov, D. Petravick, E. Corso, L. Magnoni, J. Gu, O. Barring, J.-P. Baud, F. Donno, M. Litmaath, S. De Witt, J. Jensen, M. Haddox-Schatz, B. Hess, A. Kowalski, and C. Watson. The Storage Resource Manager Interface Specification. December 2006, http://sdm.lbl.gov/srm-wg.

[13] K. Karasavvas, M. Antonioletti, M. Atkinson, A. Hume, M. Jackson, A. Krause, and C. Palansuriya. Introduction to OGSA-DAI Services. Lecture Notes in Computer Science, Springer-Verlag LNCS 3458, pp. 1-12, 2005.

[14] J. Novotny, M. Russell, and O. Wehrens. GridSphere: A Portal Framework for Building Collaborations. Concurrency & Computation-Practice & Experience. Vol.16(5), pp. 503-513, 2004.

[15] G. Allen, K. Davis, K.N. Dolkas, N.D. Doulamis, T. Goodale, T. Kielmann, A. Merzky, J. Nabrzyski, J. Pukacki, T. Radke, M. Russell, E. Seidel, J. Shalf, and I. Taylor. Enabling Applications on the Grid: A GridLab Overview. International Journal of High Performance Computing Applications, Vol.17(4), pp. 449-466, 2003.

[16] Deutsches Forschungsnetz (DFN), https://www.dfn.de, Deutsches Forschungsnetz – Public Key Infrastructure (DFN-PKI), https://www.pki.dfn.de.

[17] Grid Computing Centre Karlsruhe (GridKa) Deutsches Forschungsnetz (DFN), https://www.dfn.de, Deutsches Forschungsnetz – Public Key Infrastructure (DFN-PKI), https://www.pki.dfn.de.

[18] European Policy Management Authority for Grid Authentication (EUGridPMA), http://www.eugridpma.org.

[19] International Grid Trust Federation (IGTF), The Grid's Policy Management Authority, http://www.gridpma.org.

[20] Reference installation of the D-Grid integration project, http://www.d-grid.de/index.php?id=298, or http://www.d-grid.de/providerportal → 'Bereitstellung von Ressourcen' → 'Grid-Middleware (Server)' → Refernz-Installation.

[21] Virtual Organization Membership Registration Service (VOMRS), http://www.uscms.org/SoftwareComputing/Grid/VO.

[22] D-Grid Initiative, http://www.d-grid.de.

[23] D-Grid Integration project (DGI), http://dgi.d-grid.de.

[24] D-Grid User Portal, http://www.d-grid.de/userportal.

[25] D-Grid Provider Portal, http://www.d-grid.de/providerportal.