INSTALLING AND OPERATING A GRID INFRASTRUCTURE AT DESY

A. Gellrich, A. Campbell, U. Ensslin, M. Ernst, P. Fuhrmann, V. Gülzow, B. Lewendel, R. Mankel, S. Padhi, M. de Riese, M. Vorobiev, K. Wrona, DESY, Hamburg, Germany

F. Brasolin, INFN Bologna, Italy

J. Ferrando, University of Glasgow, Scotland

Ch. Wissing, University of Dortmund, Germany

Abstract

DESY is one of the world-wide leading centres for research with particle accelerators and synchrotron light. At the hadron-electron collider HERA three experiments are currently taking data and will be operated until 2007. Since end of August 2004 the *DESY Production Grid* is operated on basis of the recent LCG-2 release. Its Grid infrastructure is used for all DESY Grid activities, including national and international Grid projects. The HERA experiments are adapting their *Monte Carlo* production schemes to the Grid.

INTRODUCTION

The HERA experiments at DESY [1] have been taking data for more than a decade and are planning to continue running until 2007. Their data are managed by the DESY computer centre which stores data volumes of the order of 1 PB and is the home for around 1000 CPUs.

The role of DESY for the HERA experiments is –in LCG terminology - as Tier-0 and Tier-1 centre for data recording and production and as analysis centre. For the HERA-II program it is essential to use Grid resources at remote sites, starting with *Monte Carlo* (MC) production, where the applications are ready to use the Grid infrastructure. In cooperation with the University of Hamburg DESY supports the Large Hadron Collider (LHC) experiment CMS in the LHC Grid (LCG).

DESY started Grid activities with the installation of an EDG1.4 based Grid testbed in 2003. In April 2004 a second Grid testbed was set up, exploiting LCG-2 middleware. As a consequence DESY joined a number of national and international Grid projects. Since August 2004 a production-grade Grid infrastructure is being operated at DESY, using the LCG-2_2_0 middleware release. The *DESY Production Grid* is part of the LCG TestZone. It supports the LCG *Virtual Organizations* (VO) 'cms' and 'dteam' and hosts and manages VOs for the DESY experiments and groups, among them: 'hone' and 'zeus' for the HERA experiments H1 and ZEUS, 'ildg' for the Lattice QCD community, and 'ilc' for the international linear collider detector groups.

2. GRID ACTIVITIES AT DESY

EGEE@DESY

DESY participates in the EU-project *Enabling Grid for E-SciencE* (EGEE) [3] as a member of the German / Switzerland federation. EGEE started on April 1st, 2004 for a first two-year period of a four-year programme. It aims on building a Grid infrastructure in Europe which is made available to scientists of various areas. Two pilot application domains have been selected. One is HEP with LCG in particular. DESY participates in the Service Area (SA1) and will bring in its experience and know-how in computing infrastructure with elaborated data management and analysis functionality.

The *DESY Production Grid* was part of the Grid infrastructure presented to the participants of the *GridKa School '04* end of September 2004.

D-GRID@DESY

DESY is a founding member of the German D-GRID initiative [4]. The D-GRID project will presumably start on January 1st, 2005, with a research and development programme for a national e-science infrastructure. Exploiting the experiences and know-how in data management, DESY will play a leading role in the *HEP-Community Projects* as well as in the *Integration Project.*

For the HEP-community DESY manages and hosts together with the University of Wuppertal and FZ Karlsruhe a German-wide *catch-all* VO and provides core services to the German HEP community. It is planned to organize the universities and institutes in a *German Grid for High Energy Physics* (GHEP) and to set up a VO 'ghep'.

ILDG@DESY

The International Lattice DataGrid (ILDG) [5] was started with the aim of making gauge field configurations available to an international group of scientists using Grid technologies. Each configuration submitted to ILDG will consist of a set of meta-data and a set of binary files.

In the context of the German *Lattice Forum* (LATFOR) [6], DESY is setting up a Data Grid testbed, exploiting its Grid infrastructure, in particular the catalog services.

The Experimental Physics Institute of the University of Hamburg, located on the DESY site, participates in the CMS experiment. DESY supports the University Institute in installing and operating a Grid infrastructure for the LCG [7]. These CMS Grid activities are incorporated in the LCG-2 based Grid infrastructure at DESY. The DESY Grid computing resources, including computers of the University Institute, are part of the LCG-2 TestZone.

3. GRID DEVELOPMENTS AT DESY

The goal of the dCache project [8], a joint effort between DESY and FermiLab, is to provide a system for storing and retrieving huge amounts of data, distributed among a large number of heterogeneous server nodes, under a single virtual file system tree with a variety of standard access methods. Depending on the Persistency Model, dCache provides methods for exchanging data with backend (tertiary) Storage Systems (HSM) as well as space management, pool attraction, dataset replication, hot spot determination and recovery from disk or node failures. Connected to a tertiary storage system, the cache simulates unlimited direct access storage space. Data exchanges to and from the underlying HSM are performed automatically and invisibly to the user. File system namespace operations may be performed through a standard nfs(2) interface.

A dCache-based Storage Element, which includes an *Information Provider* exploiting the GLUE schema, was developed for LCG-2 and is in use at DESY.

4. GRID INFRASTRUCTURE AT DESY

The Grid Infrastructure at DESY is based on the most recent LCG-2 middleware LCG-2_2_0. It contains all elements to make up a complete Grid, including a Resource Broker (RB) with Information Index (BDII) and Proxy Server (PXY), catalog services such as a Local Replica Catalog (LRC) and a Meta Data Catalog (MDC), the management of Virtual Organizations (VO), User Interfaces (UI), Computing Elements (CE) with Worker Nodes (WN), and a dCache-based [8] Storage Element (SE) with the attached DESY mass storage facilities. The *DESY Production Grid* exploits a total of 27 dual-CPU servers of which 17 nodes are set up as Worker Nodes. The nodes are equipped with two Intel Pentium 4 XEON / 2.8 GHz CPUs, 2 GB of DDRAM, an IDE system disk of 80 GB, an IDE data disk of 200 GB, and are housed in rack-mounted 1U boxes. The nodes are connected to DESY's 10 Gbit/s backbone via Gigabit-Ethernet. The WAN connection to the world is a G-WIN 1 Gbit/s link. The dCache-based Storage Element (SE) provides access to the entire DESY data space of 0.5 PB. See figure 1.

The concept of incorporating computing resources of the various groups into a common homogenous system (*federating resources*) will be realized by way of separate CEs submitting to one farm. The CEs then act as batch slave servers and submit jobs to the Worker Nodes by means of a central batch (PBS) server. Since information on the computing resources is published to the outside world by the Computing Elements, the various VOs gain an explicit visibility.

Installation Experiences

The first testbed was based on EDG1.4. The installation was carried out manually on standard DESY workstation, exploiting a SuSE7.2 based Linux system. It turned out that such an installation approach is possible. The usage of the site standard DESY Linux 4/5 was beneficial with respect to all aspects of system updates and maintenance, including security. However, the EDG as well as the LCG middleware is tailored for RedHat Linux 7.3 which causes much effort to set up and install a Grid infrastructure on a SuSE-based system.

The second Grid testbed as well as the recent *DESY Production Grid* were set up using LCFGng which turned out to be a useful installation and maintenance tool. The User Interfaces are equipped with NIS/YP and AFS provide the DESY users with their home directories and standard login procedures.

Operational Aspects

The Globus MDS based Information Service, especially the EDG BDII as used in the first EDG-based testbed, appeared as a constant source of problems. The LCG-2 middleware replaced the Globus MDS and BDII scheme with the so-called LCG-BDII. This step has simplified the Information Provider hierarchy and led to more stability, however, without solving the fundamental scalability problems of the LDAP-based system. Experiences with LDAP-free Information Providers such as R-GMA have not been gained at DESY as of now.

5. GRID APPLICATIONS AT DESY

The physics output of HERA-II is the driving force for an increasing demand for MC production. While the paradigms of resource sharing are changing, Grid technologies are applied to account for the demand. The HERA experiments H1 [9] and ZEUS [10] are adapting their MC production schemes to the Grid. The DESY *Production Grid* hosts the VOs 'hone' , 'ilc', 'ildg', and 'zeus', which are now also supported by other collaborators world-wide, as well as some other DESY VOs. More collaborations and groups will follow.

H1

There are mainly three sites involved in the MC production for the H1 experiment, namely DESY, Dortmund, and RAL. Using conventional local batch systems like PBS, more than 300 million events per year can be produced.

Since the needs for MC events are expected to rise significantly for high luminosity running at HERA-II, additional resources for the MC production are needed. In order to include more institutes the MC production will be done based on LCG-2. With such a framework institutes

participating also in an LHC experiment can be included without much effort.

A few thousand test events have been produced already as proof of principal. Presently, tools to manage larger productions are in preparation. With the first version of these tools in place the fraction of events simulated with Grid techniques should rise quickly.

For the H1 experiment DESY hosts the VO 'hone', that is currently supported by the University of Dortmund and RAL.

ILC

In order to enable the usage of Grids for the International Linear Collider (ILC) detector group at DESY, a VO '*ilc*' was set up. The exchange of detector simulation data with outside institutes, exploiting Grid tools, will be the first goal. Catalog Services such as Local Replica Catalog and Meta Data Catalog are provided by the *DESY Production Grid* for the VO '*ilc*'. The dCache-based Storage Element allows presenting all ILC data files which are stored in DESY mass storage facility to the Grid.

A next objective would be the leveraging of MC production with shared resources in the ILC community.

ZEUS

The ZEUS experiment has been pioneering world-wide distributed MC production since 1995 with a tool named *funnel* [11] which successfully produces up to 300 million events per year on more than 20 sites. Since the yearly demand for MC production is expected to approximately double until 2007, the experiment is currently building an integrated MC production system ("ZIMP") that is suited to submit production jobs transparently both to the grid as well as to traditional funnel sites. One key point of this concept is a *gateway* between the integrated production service and the ZEUS Grid, which resides on a dedicated Grid UI, and which generates the required job description language code dynamically. Another essential element is the migration of the required production binaries, steering cards and detector constants into the replica catalog of the VO '*zeus*', in order to avoid submission of this data by way of the sandbox of each job. This concept allows re-using much of the already existing ZEUS-specific software tools for book-keeping, user request portals, monitoring, and statistics.

Besides the core installation at DESY, the *DESY Production Grid*, several other sites are already supporting the VO '*zeus*', including the Rutherford Appleton Lab (RAL), ScotGrid (University of Glasgow) and the Universities of Hamburg, Dortmund and Karlsruhe. The ZEUS MC production suite has been fully commissioned in the LCG-2 production environment, and about half a million events have already been produced on the Grid in total, of which 84000 have been done at the two RAL sites. The Grid-produced event samples have successfully passed the regular *Data Quality Management* (DQM) checks and are already in use for the analysis of real physics data. While this is clearly suited to prove the

principle, a much higher production potential can be expected with full automation of the production chain completed. Several other partner institutions in the ZEUS collaboration have access to local Grid resources which are excellent candidates to become future sites for the ZEUS Grid production.

CONCLUSIONS

As a result of earlier testbed installations, a common DESY Grid infrastructure was successfully set up on the basis of LCG-2_2_0. The *DESY Production Grid* is used for all Grid activities at DESY. It provides generic Grid services.

The usage of a dCache-based Storage Element (SE) in conjunction with a Local Replica Catalogue (RLC) and Meta Data Services (MDS) enables the DESY VOs to adapt MC production schemes to the Grid. In cooperation with other LCG-2 sites a considerable amount of MC events have recently been produced and are in use by the collaborations.

ACKNOWLEDGEMENTS

REFERENCES

[1] http://www.desy.de/
[2] http://grid.desy.de/
[3] http://www.eu-egee.org/
[4] http://d-grid.de/
[5] http://www.lqcd.org/ildg/
[6] http://www-zeuthen.desy.de/latfor/
[7] http://cern.ch/lcg/
[8] http://www.dcache.org/
[9] http://www-h1.desy.de/
[10] http://www-zeus.desy.de/
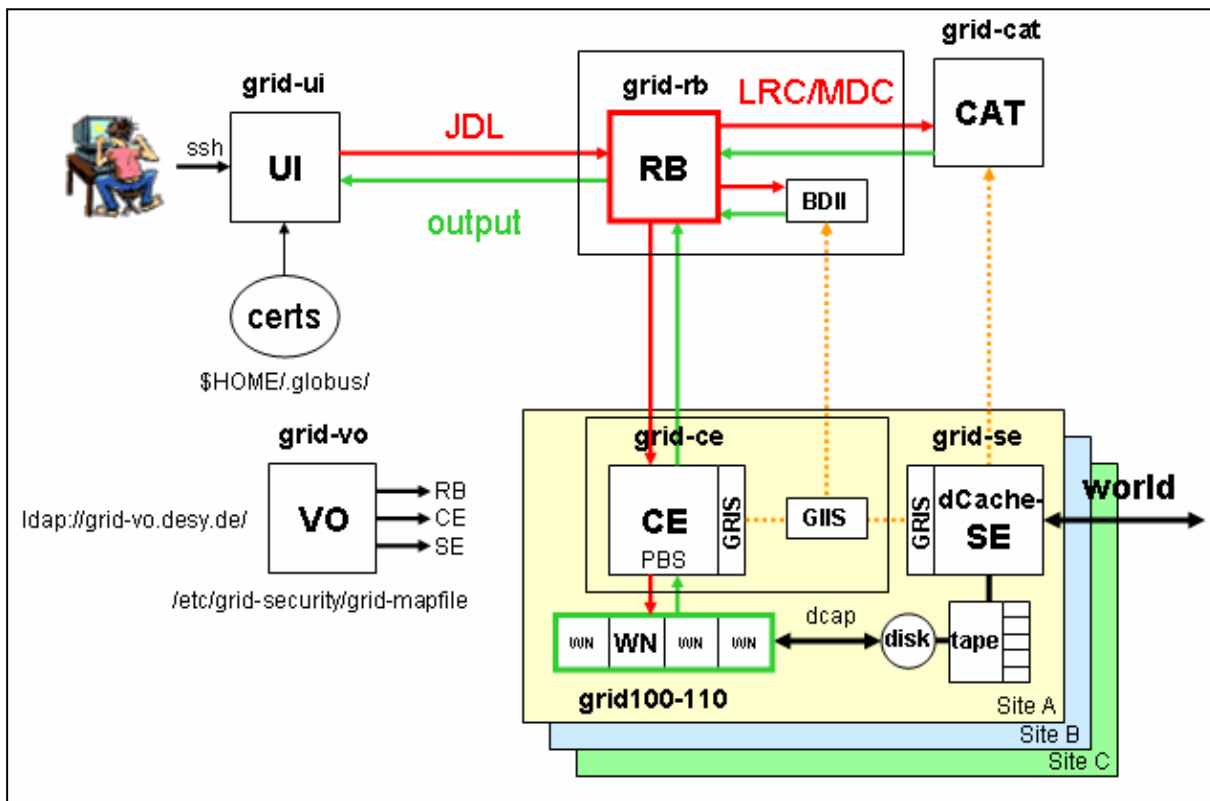[11] http://www-zeus.desy.de/components/funnel/

Figure 1: Schematic view of the *DESY Production Grid*, including all node types. The dCache-based Storage Element (SE) provides access to the entire DESY data space of 0.5 PB.