Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nuclear Instruments and Methods in Physics Research A 559 (2006) 31–34
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

Grid technology in production at DESY

A. Gellrich*, M. Ernst, P. Fuhrmann, V. Gülzow, M. de Riese

Deutsches Elektronen-Synchrotron, Notkestr. 85, 22607 Hamburg, Germany

Available online 6 December 2005

Abstract

DESY operates a complete and independent LCG-2 Grid infrastructure. It provides all services needed to run a Grid. Among the supported VOs are the global HERA VOs H1, ILC, and ZEUS, which are hosted at DESY, and the LHC VOs ATLAS and CMS. Since November 2004 the Grid infrastructure is used by the HERA VOs to produce Monte Carlo events at collaborating LCG sites.

PACS: 07.05.Bx; 29.85.+c; 07.05.Tp

Keywords: Grid; Computing infrastructure; Monte Carlo

1. Introduction

DESY is one of the world-wide leading centers for research with particle accelerators and a center for research with synchrotron light. The hadron-electron collider HERA houses three experiments which will be operated until 2007 [1]. The HERA experiments H1 and ZEUS at DESY face a growing demand for *Monte Carlo* (MC) events after the recent luminosity upgrade of the HERA collider. Traditionally, the DESY experiments performed MC production in outside institutes of the collaborators, whereas DESY concentrated on the data taking on site and the analysis work. Those outside institutes operated local batch farms to which MC production requests were sent. In this context ZEUS [2] developed a sophisticated MC production system on the basis of FUNNEL [3] which efficiently exploits spare CPU cycles on various computers of the collaborators. With the start of the building of a computing infrastructure for the LHC experiments at CERN and in particular the LHC Computing Grid (LCG) [4], most of the HEP institutes moved their computing resources to the Grid. Batch computing, e.g. for the MC production for the HERA experiments, is not widely supported anymore. On the other hand, much more resources become available via the Grid. The HERA experiments benefit from these resources given that suitable Virtual Organizations (VO) with the necessary core Grid services are available.

Grid activities at DESY are carried out in the context of the EU-project *Enabling Grids for E-SciencE* (EGEE) [5] which started in April 2004 for a period of two years and currently plans for a second period from April 2006 on. The main objective of EGEE is to build a permanent Grid infrastructure that can serve a broad spectrum of applications reliably and continuously. As a partner in the German/Swiss federation, DESY participates in the Service Area (SA1) and provides computing and storage resources to EGEE. Very recently, the German initiative to set up a national e-science infrastructure D-GRID [6] started. DESY plays a leading role in HEP community and in the Integration Project (IP).

In this paper we will describe in detail the DESY Grid infrastructure in the context of the DESY Grid activities and present operation experiences and future plans.

2. The DESY Grid infrastructure

DESY started to set up a Grid infrastructure [12] based on the LCG-2 middleware in early 2004 with the first LCG-2 release. The goal was to maintain a complete, independent, production-grade core Grid site which allows to set up VOs for the HERA experiments and DESY groups. By deploying the most recent versions of the LCG-2

*Corresponding author.
E-mail address: Andreas.Gellrich@desy.de (A. Gellrich).

0168-9002/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.nima.2005.11.101

middleware, DESY was given a spot on the worldwide map of active LCG-2 sites. Currently, two compute farms are part of the Grid infrastructure at DESY with 70 and 34 CPUs (70 kSPECINT2000 and 14 kSPECINT2000). The available disk space is 5 TB and DESY's tertiary storage system provides 0.5 PB. The system is in production since November 2004.

2.1. VO Support

For the H1 [7] and ZEUS [2] VOs were founded for which registration is available with the LCG registrar team. Both VOs are supported world-wide (see Table 2). The *International Lattice Data Grid* (ILDG) [14] is an initiative of the lattice QCD (LQCD) community, aiming on the interchange of data which are the result of very costly calculations on high-performance super computers. For the nationally acting LATFOR [8] group a VO '*ildg*' was set up [15]. The *International Linear Collider* community (ILC) [13] has recently started to use the Grid for MC productions. Together with some of the UK sites a VO '*ilc*' was founded which is hosted at DESY. For the German CMS community a dedicated VO '*dcms*' was installed. The DESY Grid infrastructure also supports the VOs for ATLAS and CMS. A complete list of all supported VOs is shown in Table 1.

Table 1
VOs supported by DESY

Global LHC VOs (hosted by LCG)		
ATLAS	*atlas*	ATLAS Experiment
CMS	*cms*	CMS Experiment
DTEAM	*dteam*	Deployment Team
Global VOs (hosted by DESY)		
H1	*hone*	H1 Experiment at HERA
ILC	*ilc*	International Linear Collider community
ZEUS	*zeus*	ZEUS Experiment at HERA
Regional VOs (hosted by DESY)		
CALICE	*calice*	
DCMS	*dcms*	German CMS Community
GHEP	*ghep*	German HEP community
ILDG	*ildg*	International Lattice Data Grid group
Regional VOs (hosted by SCAI)		
DECH	*dech*	German/Swiss EGEE Federation
Local VOs (hosted by DESY)		
BAIKAL	*baikal*	Baikal Collaboration
DESY	*desy*	DESY Test VO
HERAB	*herab*	HERA-B Experiment at HERA
HERMES	*hermes*	HERMES Experiment at HERA
ICECUBE	*icecube*	IceCube Collaboration

Table 2
International support of the global DESY VOs

ILC VO *ilc*		
DESY-HH	DESY Hamburg	Germany
DESY-ZN	DESY Zeuthen	Germany
RHUL-LCG2	Royal Holloway University London	UK
IC-LCG2	Imperial College London	UK
QMUL-ESCIENCE	Queen Mary University London	UK
RAL-PP-LCG2	Rutherford Appleton Laboratory	UK
H1 VO *hone*		
BHAM-LCG2	University of Birmingham	UK
DESY-HH	DESY Hamburg	Germany
DESY-ZN	DESY Zeuthen	Germany
RAL-LCG2	Rutherford Appleton Laboratory	UK
RAL-PP-LCG2	Rutherford Appleton Laboratory	UK
UNI-DORTMUND	University of Dortmund	Germany
ZEUS VO *zeus*		
CYFRONET-LCG2	Krakow	Poland
DESY-HH	DESY Hamburg	Germany
DESY-ZN	DESY Zeuthen	Germany
IC-LCG2	Imperial College London	UK
INFN-*	Various INFN sites	Italy
OXFORD-01-LCG2	Oxford University	UK
RAL-LCG2	Rutherford Appleton Laboratory	UK
RAL-PP-LCG2	Rutherford Appleton Laboratory	UK
SCOTGRID-GLA	Scottish Grid Service	UK
SNS-PISA	SNS Pisa	Italy
TAU-LCG2	Tel Aviv University	Israel
TORONTO-LCG2	University of Toronto	Canada
UAMLCG2	Universidad Auto'noma de Madrid	Spain
UCL-CCC	London's Global University	UK
UCL-HEP	London's Global University	UK
UWMadisonCMS	University of Wisconsin Madison	USA
WEIZMANN-LCG2	Weizmann Institute of Science	Israel

2.2. Conceptual aspects

The DESY Grid infrastructure runs all services which are needed to operate a complete Grid. The following items were considered:

- In full analogy to the production system a testbed is available to test the installation and functionality of new releases.
- The services are associated with node types in the LCG nomenclature.
- For each node type a separate machine was installed.
- Testbed and production system have similar node type layouts.
- The current system deploys an LDAP-based VO services.
- A VOMS server is available for testing. It is planned to deploy VOMS as soon as all LCG-2 Grid sites provide the VOMS client functionality.
- Service nodes which contain unique informations are regularly back-uped.

- Monitoring of high level functionalities such as job submission and output retrieval as well as individual services is needed.

We distinguish *core services* and *site resources*:

- *Exactly* one instance of the VO or VOMS service, which provides the users belonging to a VO, and the replica location services, which consists of the replica and the Metadata catalog (CAT), are unique per VO. We call them **Central Core Services**.
- *At least* one instance of the Resource Broker (RB), MyProxy server (PXY), and the Information Index (BDII) are needed. We call them **Distributed Core Services**.
- The *site resources* are delivered by the Computing Elements (CE) and Storage Elements (SE) of the sites which also run Grid Information (GIIS) and monitoring services (R-GMA).
- User connect via User Interfaces (UI) which are widely distributed.

Data management is a key aspect of Grid computing in HEP. In cooperation with Fermilab, DESY has developed a Storage Element (SE) which consists of dCache as the core storage system and an implementation of the Storage Resource Manager (SRM). Access to the entire DESY data space of 0.5 PB is provided by a dCache-based SE [9].

2.3. Technical aspects

The Grid hardware consists of machines which meet the DESY computer center standards. The following aspects were addressed:

- As far as possible homogeneous hardware was chosen.
- In order to minimize the space consumption, 1U barebone servers with dual-CPUs were installed.
- Rack-mounted machines with an air flow front to back allow for efficient cooling.
- The machines are equipped with single power supplies. Redundant power supplies for the core servers are planned.
- Machines are equipped with E(IDE) disks. Each machine has two disks.
- The machines are connected with 1 Gbit/s copper cables to local switches which are part of the 10 Gbit/s DESY network.
- DESY has a 1 Gbit/s WAN.
- All machines have console connections to a terminal server.
- On all machines Scientific Linux version SL3.04 was installed by means of *Quattor*. For the LCG-2 middleware the LCG-2 tool *yaim* was applied.
- The WNs were completely installed by Quattor with LCG-2 templates.
- The RB was upgraded to 4 GB to avoid swapping. The vast log directory `/var/edgwl` is located on the data disk.
- The WNs are configured to use the directory `/home` as work space, which is located on the data disk.
- In total 10 core service nodes and 35 plus 17 Worker Nodes (WN) are up and running.
- *Ganglia* is used as a system monitoring tool.
- A suite to regularly submit jobs to the DESY site with comprehensive usage of data management features was installed [10].

In the current set-up we use the following hardware. All machines are alike.

- SuperMicro Superserver 6013P-i;
- dual-XEON CPUs 2.8/3.06 GHz;
- 2048 MB ECC DDR-RAM;
- 80 GB/200 GB (E)IDE system/data disk;
- 1000-base FX network interface.

3. Operational experiences

The DESY Grid infrastructure has been successfully used for MC event production by the ZEUS collaboration [11]. Since November 2004 significant parts of the MC production has come from the Grid. H1 has recently finished a management framework to routinely use the Grid. ILC has made first experiences in producing MC events. In all three cases it turned out that data management is the most crucial part of the Grid. In contrast to the CPU cycles, which come from a high number of redundant farm nodes, storage is central. Malfunctioning Computing Elements can easily be recovered by simple job resubmission to another site, whereas a failing Storage Element usually affects a big bunch of jobs. Much emphasis has been put on stabilizing mass storage services by using dCache [9]. The event production systems have been improved to handle failing components by considering alternatives. In the last years DESY rolled out the S.u.S.E. Linux distribution on workgroup servers, farm nodes, and desktops. For the installation of the LCG-2 Grid middleware Scientific Linux version SL3 is (almost) mandatory and is being supported at DESY now. In order to maintain an up-to-date Grid infrastructure, appropriate measures are being taken to regularly upgrade the OS and the Grid middleware. Quattor turned out to be a helpful tool for this purpose and is considered to be used for all compute farms at DESY. Some of the core server nodes hold unique data which cannot be recovered. We therefore plan to upgrade the nodes CAT, VOMS, RB, and CE with hardware which provides mirrored file systems (RAID-5) and redundant power supplies. Thorough monitoring of the overall functionality of the Grid infrastructure (jobs submission and output retrieval; data management) as well

as regular tests of individual services (LDAP-VO-service, GridFTP, SRM, etc.) is needed at all times.

4. Conclusions and outlook

The Grid has become a strategic technology at DESY. By installing and operating a complete production-grade LCG-2 Grid infrastructure, DESY has provided access to additional, permanently increasing computing resources for the HERA experiments and DESY groups. It is planned to continue this effort for the HERA experiments and even more leverage Grid activities of future HEP experiments such as ILC. DESY plans the participation in at least one of the LHC experiments ATLAS and CMS. The Grid activities will be extended by setting up a Tier-2 center for the LHC experiments ATLAS and CMS in collaboration with other German collaborators. Monitoring of the DESY Grid infrastructure is available by means of the web [10].

Acknowledgements

We would like to thank the DESY Computer Center staff for their support in installing and operating the Grid infrastructure.

References

[1] http://www.desy.de/
[2] http://www-zeus.desy.de/grid/
[3] http://www-zeus.desy.de/components/funnel/
[4] http://cern.ch/lcg/
[5] http://www.eu-egee.org/
[6] http://d-grid.de/
[7] http://www-h1.desy.de/
[8] http://www.ifh.de/latfor/
[9] http://www.dcache.org/
[10] http://grid-web.desy.de/
[11] H. Stadie, Monte Carlo Mass production for the ZEUS experiment on the Grid, these proceedings.
[12] http://grid.desy.de/
[13] http://www-flc.desy.de/
[14] http://www.lqcd.org/ildg/
[15] D.Pleiter, ILDG: DataGrids for lattice QCD, these proceedings.