The ILC and the Grid

A. Gellrich*

Deutsches Elektronen-Synchrotron (DESY)
Notkestr. 85, 22607 Hamburg, Germany

Already today major computing resources for HEP are available in the Grid only. For the ILC project two Virtual Organization were founded which are supported by many Grid sites world-wide. The ILC community has started to use Grid resources for *Monte Carlo* production, data analysis, and data storage. The development of specific tools for ILC has recently been initiated.

1 Introduction

Driven by the computing needs of the four LHC experiments, a global computing infrastructure is under construction, the so-called *world-wide LHC Computing Grid* (WLCG) [2]. It incorporates compute and date storage resources that are distributed all over the world. Access to the resources is provided by means of a service infrastructure. People, institutions, and resources are collected within *Virtual Organizations* (VO) that are based on common sharing rules. Users authenticate against the Grid services by way of personal certificates and are authorized to use resources within a VO.

The European part of WLCG is operated in the context of the EU-project *Enabling Grids for E-SciencE* (EGEE) project [3]. EGEE brings together scientists and engineers from more than 240 institutions in 45 countries world-wide to provide a seamless Grid infrastructure for e-Science that is available to scientists 24 hours-a-day. Conceived from the start as a four-year project, the second two-year phase started on 1 April 2006, and is funded by the European Commission. Currently EGEE is planning for a third period, starting in 2008.

Within the EGEE project two VOs for the ILC project [5] were founded and officially registered: *'ilc'* and for the CALICE collaboration [6] *'calice'*. Both VOs are supported by a large number of sites within the EGEE.

This paper complements a talk given at LCWS2007 in the 'LCWS: Sim/Reco' parallel session [1].

2 The Grid

2.1 Grid Infrastructure

The EGEE Grid infrastructure consists of Grid services and Grid resources. The compute and data storage resources are contributed by the participating Grid sites and are distributed globally. Sites operate *Computing Elements* (CE), which act as logical entry points to clusters of *Worker Nodes* (WN), and *Storage Elements* (SE), which provide interfaces to data storage systems.

Core Grid services are *VO Membership Services* (VOMS), *VO Membership Registration Service* (VOMRS), *LCG File catalogues* (LFC), *Workload Management Systems* (WMS), and *Resource Brokers* (RB). Those services must exist at least once per VO.

*http://www.desy.de/~gellrich

2.2 Grid Middleware

As Grid software – called *middleware* which mediates between the user's application and the operating system – the EGEE development *gLite* [9] is deployed. *gLite* provides software to implement the Grid services as well as client software to communicate with the Grid services. Usually the user's workgroup servers or desktops deploy this *User Interfaces* (UI) client software.

3 The ILC VOs

DESY hosts the ILC VOs *'calice'* and *'ilc'* as a part of its EGEE Grid infrastructure. The membership to the VOs is managed by VOMS [12]. User registration is handled by VOMRS [13]. In order to register files, LCG File Catalogues are operated for both VOs. DESY also operates Workload Management Systems and Resource Brokers for the VOs.

3.1 Communications within ILC

The following communication channels are available for ILC. DESY operate the mailing lists

- `calice-vo-support(@)desy.de`
- `calice-vo-users(@)desy.de`
- `ilc-vo-support(@)desy.de`
- `ilc-vo-users(@)desy.de`

A web page with instructions to site administrators and users is available [7].

3.2 Testbeam Data

DESY enabled its mass storage system as a Storage Element, including a large tape library, for the ILC VOs. Read and write access to the data is possible by way of Grid tools.

The CALICE collaboration [6] decided a year ago to use DESY mass storage as the main repository for testbeam data. To date (October 2007) a total of $30TB$ of data was transfered through the Grid to DESY. It is planned to also transfer testbeam data from FNAL via the Grid to DESY.

A similar approach has recently been chosen by the detector development group of EUDET-JRA1 [10]. To date (October 2007) a total of $1.3TB$ have been transfered to DESY via the Grid.

All CALICE and EUDET-JRA1 (Grid) users benefit from the presence of the testbeam data in the Grid which allows seamless access within hours after data transfer world-wide.

3.3 ILC Application Software on the Grid

The WLCG is close to finishing the transition from Scientific Linux 3 (SL3) to SL4 as the Operating System on the Worker Nodes. Most of the HEP software – also in ILC – is supported for SL4.

In HEP, geometry and calibration data are typically stored in central database systems which leads to severe bottlenecks when many jobs try to access them concurrently. For ILC the tool set *MyGridDB* [17] was developed which decouples production jobs from the central ILC database. Necessary database entries are simply dumped into files and stored in the Grid. At run-time, jobs create local database instances with these files.

4 Perspectives

So far, the Grid has been used by ILC on a basic level only. Data transfers as well as data analyzes and *Monte Carlo* production have been carried out by way of simple scripts which utilize the generic gLite command suite. In order to make most efficient use of the Grid resources high level user tools will be needed. As a product of the LCG developments, a number of elaborated tools are available that might well be applicable to the ILC needs. Among those tools are GANGA [14] and DIRAC [15]. A working group was set up recently to evaluate those tools for ILC.

Up to now, all software, that is needed to run ILC jobs, including the libraries of the persistency framework /em LCIO [16], are either submitted with the job or are downloaded from at run-time. The WLCG Grid infrastructure foresees a mechanism to install VO-specific software on the various sites locally. This mechanism includes a software *tag* which is published by way of the information system. This *tag* can be evaluated by the Workload Management System to ensure the presence of the requested software on the picked Computing Element.

In EGEE, sites usually provide Grid resources to supported VOs on one infrastructure. Resource distribution is performed by virtual shares rather than dedicating hardware directly to VOs. This ansatz allows to very efficiently utilize resources and makes it easy to add more. Many Grid sites have already committed a fraction of their Grid resources to ILC. With the increase of resources in the Grid, more resources will be available for ILC computing too.

To date two instances of the VO *'ilc'* live concurrently in EGEE and in the American *Open Science Grid* (OSG) [4]. Methods to synchronize the members of the VOs are being developed. Again, it is planned to benefit from WLCG developments on interfaces between those two Grids.

5 Summary

Grid computing is a strategic technology for the future. It is about virtualization of computing resources, such as compute cycles and data storage, in a *global* context. Already now large computing resources outside the Grid are not available anymore in HEP.

For the ILC project two VOs were set up which use the existing EGEE/WLCG infrastructure; all core Grid services are hosted at DESY. Various sites world-wide support the ILC VOs and provide compute and data storage resources to the ILC VOs.

Though tools for the ILC *Monte Carlo* mass production and user analysis as well as data transfer are still rudimentary, considerable amounts of resources have already been exploited by ILC.

Recently, working groups within ILC were founded to evaluate and adapt software tools and frameworks to the needs of ILC.

ILC strongly benefits from the existing Grid infrastructure with its increasing resources as well as from products which have been developed for LHC.

Acknowledgments

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

EGEE-II is a project funded by the European Commission with the contract number INFSO–RI–031688C.

References

[1] Slides: http://ilcagenda.linearcollider.org/contributionDisplay.py?contribId=233&sessionId=76&confId=1296

[2] http://cern.ch/lcg/

[3] http://www.eu-egee.org/

[4] http://www.opensciencegrid.org/

[5] http://www.linearcollider.org/

[6] http://polywww.in2p3.fr/activites/physique/flc/calice.html

[7] http://grid.desy.de and http://grid.desy.de/ilc/

[8] http://www.dcache.org/

[9] http://glite.web.cern.ch/glite/

[10] http://www-zeus.desy.de/~haas/eudet-jra1/

[11] http://www.desy.de/

[12] https://grid-voms.desy.de:8443/voms/calice/ and https://grid-voms.desy.de:8443/voms/ilc/

[13] https://grid-voms.desy.de:8443/vo/ilc/vomrs/

[14] http://ganga.cern.ch/

[15] http://dirac.cern.ch/

[16] http://lcio.desy.de/

[17] http://www.desy.de/~hhoelbe/