Status Of the Bonn EUDET Computing Cluster

K. Desch[*], M. Killenberg[*], S. K. Nderitu[*], P. Wienemann[*], R. Zimmermann[*]

December 4, 2008

Abstract

In this report the current status of the EUDET computing cluster at the university of Bonn is outlined. The cluster has one Computing Element, one batch head–node and 8 Worker Nodes. In addition there is a Storage Element managed by dCache. There is also a grid User Interface and a machine on which site monitoring services run. The cluster has been set up in the physics institute's main computing room where it's mounted in one of the racks. An appropriate redundant cooling plant has been built maintaining the room at constant working temperature. The cluster is grid enabled and utilizes WLCG grid middleware.

[*]University of Bonn, Bonn, Germany

1 Introduction

For analysis of prototype data or detailed simulations of detector effects, a large amount of computing resources is required. To this end the EU funded the equivalent of 40 CPU cores to be used for ILC at a grid enabled site. The Bonn Physics institute added a file server for completeness of the installation. These machines are currently set up as a grid computing cluster for the ILC and CALICE virtual organizations (VO). Other VOs to be supported later on are ATLAS as well as the mandatory DTEAM and OPS VOs, which are included for testing and validation purposes of grid sites.

In the following sections a brief description of each of these services is given. Details of the gLite middleware set up on all services is beyond the scope of this document. For this we refer the reader to the general gLite documentation [1], especially the user guide. Simplified explanations will be given where necessary for clarity, however.

2 Infrastructure

2.1 Computer room

The institute's computer room had a new (redundant) 200kW cooling plant installed to match the increasing demand for additional computing power. The EUDET cluster had to be relocated during this upgrade and moved back after it was completed in the fall of 2008. The room has ten 19" Racks with 48U each available for the institute's experimental high energy physics groups. Currently only one of the racks is dedicated to the EUDET computing cluster, but additional racks are reserved for the grid site's additional ATLAS dedicated hardware due to arrive before the end of the year.

2.2 Networking

To simplify maintenance and for security reasons only necessary nodes or services are directly exposed to the grid and the internet. This includes all the nodes running middleware services as well as a machine providing outbound connectivity for all the nodes providing computing power (e.g. Worker Nodes) which are located in a private IP network, as sketched in Figure 1.

3 Site Resources and Services

3.1 The Computing Element

The Computing Element (CE) consisting of the 10 EUDET fundet servers of type Sun-Fire X2200M2, each with 2 dual–core CPUs at 2.2GHz and 2GB of RAM per core, GB–Ethernet and a service processor for remote management.

Conceptually it provides the site's computing power to the grid as a whole in a black–box fashion. Internally it consists of a local resource management system (LRMS) and the
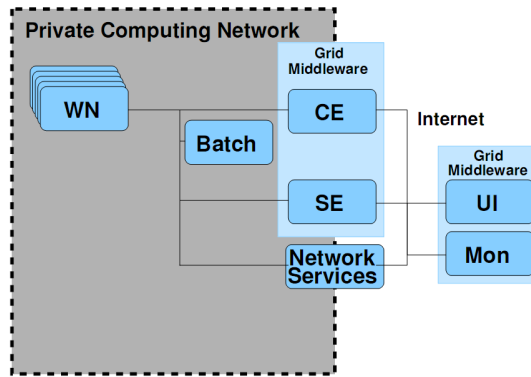
Figure 1: Schematic setup of all machine types

actual CE middleware (Grid Gate service). The LRMS itself usually consists of some kind of head or master node providing scheduling, job management and accounting features and a (possibly large) number of Worker Nodes (WNs) that do the actual computational work.

The software deployed in Bonn is the default version of the middleware in production for the Worldwide LHC Computing Grid (WLCG) [2], i.e. gLite 3.1. For the LRMS the open source Torque Resource Manager and Maui Cluster Scheduler by Cluster Resources Inc. [3] are used. The base system software is 32 bit Scientific Linux 4 [4], maintained by CERN and Fermilab.

While in principle it's possible to have the CE middleware and the LRMS head node running on one machine, installing them separately makes for a much more flexible and maintainable setup (e.g. allows replacing a CE without disturbing the LRMS). This approach was also chosen for the Bonn grid site, because it simplifies interaction of the dual–homed CE middleware host with the LRMS, which is completely contained in the private IP network.

3.2 The Storage Element

For the Storage Element (SE) a 4U file server of type ICO Servermaster 451 is used. The file server has 3 TB disk space configured as hardware raid 5 on 32 bit Scientific Linux 4. It is running on dCache 1.8 [5] and gLite 3.1 middleware. All dCache services are currently running on this one machine. That is, the Storage Resource Manager (SRM), the GridFTP doors, the Pool Manager and the PNFS database. It is planned to distribute these services to 3 or 4 machines to improve performance and for scalability purposes. The 3 TB are partitioned into 3 pools and are allocated to all five supported VOs. The authorizations are managed using dCache gPlazma VO role mapping policies. For data access four main protocols are used. For copying operations between different SEs across the grid, for example, srmcp, lcg-cp and globus-url-copy can be used, while dccp can be used locally. There are are also other data management techniques all within the LCG project framework, e.g. the File Transfer Service (FTS) which is used for bulk data transfers between sites.

Figure 2: Photo of the SE , CE and WNs inside the 19" rack

As illustrated in Figure 1 running jobs have direct access to the storage while all authorized users can access the storage from a User Interface after obtaining a grid proxy certificate.

3.3 The Information System

Installed on the same node as the CE is the so called Site BDII which is part of the grid information system. It collects information on the resources currently provided by the CE and SE as well as static site info and is queried regularly by a Top Level BDII to make the information available to the grid.
Additional accounting and user level information is collected in a complementary information system (R-GMA) on another node called MON box. This service is currently installed on a standard PC provided by the institute.

3.4 The User Interface

The User Interface (UI) provides all grid tools necessary for job and file management as well as access to the grid information services to the users of all VOs supported by a site. Since the ILC and CALICE VOs are not part of the WLCG, some configuration files provided by DESY (see [6] for more info) had to be installed in addition to a default WLCG setup.
The UI, set up on another institute provided PC has been in use by the local ILC group for some time already.

4 Cluster Installation and Management

Basic system installation is currently done using Kickstart (see e.g. [7]) and simple shell scripts. For the configuration of the middleware YAIM [8] is used.
A move to a more powerful cluster management system like QUATTOR [9] is foreseen and work on this has already started.

5 Conclusion

A grid enabled computing cluster has been set up in Bonn and all components are working. The grid site is currently undergoing final tests before certification. Once this step is passed it will be available to users of the ILC and CALICE VOs. Additional computing and storage resources procured for use by the ATLAS VO will be available on a best effort basis as soon as they have been integrated in the current setup.

Acknowledgement

References

[1] gLite homepage: http://glite.web.cern.ch/

[2] WLCG technical hompage: http://lcg.web.cern.ch/LCG/

[3] Cluster Resources Inc. homepage: http://www.clusterresources.com

[4] Scientific Linux homepage: https://www.scientificlinux.org/

[5] The dCache Desy Project. homepage: http://www.dcache.org/

[6] DESY's Grid for ILC webpage: http://grid.desy.de/ilc/

[7] Kickstart manual for Red Hat Enterprise Linux 4: http://www.redhat.com/docs/manuals/enterprise/RHEL-4-Manual/en-US/System_Administration_Guide_/Kickstart_Installations.html

[8] YAIM homepage: https://twiki.cern.ch/twiki/bin/view/EGEE/YAIM

[9] QUATTOR homepage: http://www.quattor.org