Journal of Physics: Conference Series **396** (2012) 032001 doi:10.1088/1742-6596/396/3/032001

Eurogrid: a new glideinWMS based portal for CDF data analysis

S.Amerio[a,1], **D.Benjamin**[b], **J.Dost**[e], **G.Compostella**[a,c,d], **D.Lucchesi**[a,c], **I.Sfiligoi**[e]

[a] INFN Padova, [b] Duke University, [c] University of Padova, [d] Now with Max-Planck-Institut fuer Physik (MPG), [e] University of California San Diego

E-mail: silvia.amerio@pd.infn.it

Abstract. The CDF experiment at Fermilab ended its Run-II phase on September 2011 after 11 years of operations and 10 fb^{-1} of collected data. CDF computing model is based on a Central Analysis Farm (CAF) consisting of local computing and storage resources, supported by OSG and LCG resources accessed through dedicated portals. At the beginning of 2011 a new portal, Eurogrid, has been developed to effectively exploit computing and disk resources in Europe: a dedicated farm and storage area at the TIER-1 CNAF computing center in Italy, and additional LCG computing resources at different TIER-2 sites in Italy, Spain, Germany and France, are accessed through a common interface. The goal of this project is to develop a portal easy to integrate in the existing CDF computing model, completely transparent to the user and requiring a minimum amount of maintenance support by the CDF collaboration. In this paper we will review the implementation of this new portal, and its performance in the first months of usage. Eurogrid is based on the glideinWMS software, a glidein based Workload Management System (WMS) that works on top of Condor. As CDF CAF is based on Condor, the choice of the glideinWMS software was natural and the implementation seamless. Thanks to the pilot jobs, user-specific requirements and site resources are matched in a very efficient way, completely transparent to the users. Official since June 2011, Eurogrid effectively complements and supports CDF computing resources offering an optimal solution for the future in terms of required manpower for administration, support and development.

1. Introduction

The CDF experiment collected 10 fb^{-1} of data until the Tevatron accelerator shutdown in September 2011. During its operations, the CDF computing architecture evolved from the initial model based on dedicated farms to using grid-based resources. The current CDF computing model[1] is based on a central farm located in Fermilab and accessed through the CDFGrid portal, while Open Science Grid (OSG) and LHC Computing Grid (LCG) resources are accessed through the dedicated portals NamGrid and Eurogrid respectively. An experiment specific package, the Central Analysis Farm (CAF) software, provides the users with a uniform interface to farms on different grid sites. The CAF software, based on Condor batch system [2][3], provides tools for submitting, managing and monitoring batch jobs.

[1] Silvia Amerio was supported by a Marie Curie International Outgoing Fellowship within the 7^{th} European Community Framework Programme

The Eurogrid portal allows to access a dedicated farm at CNAF Tier-1 and additional LCG computing resources at different Tier-2 sites in Italy, Spain, Germany and France. In the past Tier-1 resources were accessed through a dedicated Condor based portal, CNAFCaf, and LCG resources through a gLite-based one, LcgCAF [4]. In 2011, in preparation for the end of CDF data taking, a reorganization of the computing model in Europe was planned. The aim of this was to consolidate the different resources used by the experiment and to simplify as much as possible their usage and support, while retaining their full available capabilities. Resources accessed through CNAFCaf and LcgCAF were then merged in a single portal, using a pilot based workload management system, the glideinWMS [5][6] as backend. glideinWMS comprises two main elements: the Factory and the Frontend; the Factory is a service that submits properly configured pilot jobs (Condor glideins) to a static list of grid sites supporting CDF Virtual Organization (VO). The Frontend acts as a client of the Factory: when user jobs are waiting in the central task queue of Eurogrid, it notifies the Factory and asks for glidein submission. The Frontend knows nothing about the glideins or grid sites; it only has to match user job requirements to the sites attributes published by the Factory. In Eurogrid, CDF frontend interacts with the Factory at the University of California San Diego (UCSD).

glideinWMS was chosen for Eurogrid for different reasons: its underlying batch system is based on Condor as the CAF, so the integration with CAF code was seamless; additionally, the manpower required for maintenance on the experiment side is minimum: job submission to the grid is under the responsibility of the Factory at UCSD, while CDF is responsible for the Frontend only. Finally, the use of pilot jobs effectively masks site related errors and malfunctions from the users.

Eurogrid was designed to be the best solution for the future of CDF computing in Europe in terms of transparency for the user, manpower required for development and maintenance and job efficiency.

2. Eurogrid architecture

In the following the main elements of Eurogrid architecture will be described. The outline of the new portal is presented in Fig. 1.

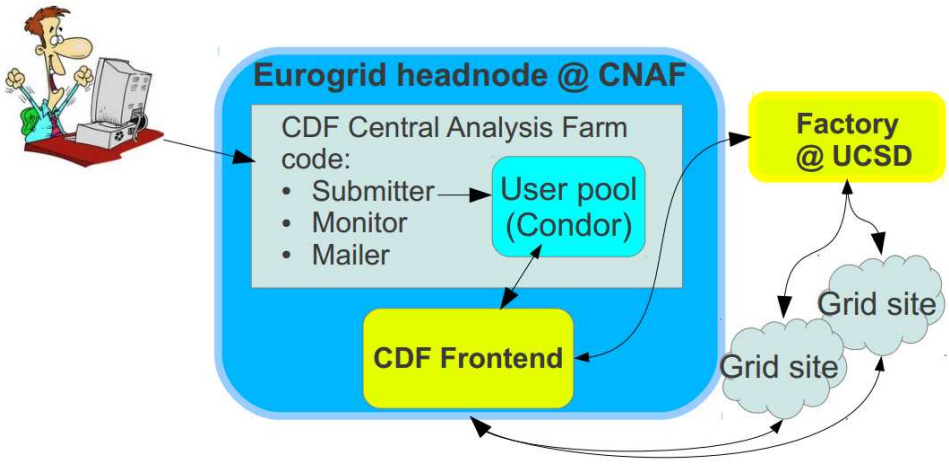

Figure 1. Scheme of Eurogrid portal.

2.1. Job submission and execution
The submission of a job is initiated by a custom client software distributed to all collaborators together with the CDF software, where users specify the requirements for their jobs such as analysis farm, job output location, method for data access (if the submitted job requires access to CDF data), job duration and whether CDF analysis software is needed. Upon user request, the job is submitted to the Eurogrid headnode, located at CNAF computing center, where a daemon called Submitter receives the job, prepares it to be executed by Condor and finally submits it to the local Condor pool. The glideinWMS Frontend polls the Condor pool and contacts periodically the Factory and requests the submission of glideins until the local Condor pool queue is empty. Glideins are sent by the Factory to a predefined list of grid sites accepting jobs by the CDF VO. Thanks to a custom glideinWMS configuration, a subset of the requirements of the job is enforced already at a Factory level: for example, if all the jobs in the local Condor pool queue require access to data, glideins are submitted only to those sites where data access is possible. The same happens for the requirement of the CDF analysis software. This approach guarantees an optimal usage of the resources and prevents pilot jobs to run on worker nodes that are not suitable for the tasks requested by the users. When glideins finally run, they validate the node (i.e. check available disk space, libraries, software availability), start the necessary Condor daemons and publish the details of the worker node to the headnode of the Condor pool using custom Condor ClassAds. Thanks to this setup, the worker node becomes part of the Condor pool. Using Condor matchmaking, the headnode dispatches jobs to the available worker nodes. User jobs run under the supervision of a CDF job wrapper called CafEXE, that fetches the job tarball via http and runs the job. Since job executables and libraries are usually large, the usage of http allows some caching through proxies at the grid sites, allowing a better usage of the available bandwidth. When the job is done, CafEXE tars up the output files and sends them back to a user-defined data server, usually via kerberized ftp.

2.2. Authentication
Before submitting any job, the user needs to authenticate via Kerberos [7] as a member of the CDF collaboration in the FNAL realm. Upon receiving a job request and authenticating the user, Eurogrid headnode contacts the Fermilab Kerberized Certification Authority (FKCA) using the Kerberos credentials forwarded by the user to obtain a X.509 proxy certificate for that principal. The Virtual Organization Membership Service (VOMS [8]) at CNAF is then contacted to attach the correct VOMS role extensions to the proxy before submitting to the grid. Jobs are submitted and executed with user credentials. During job execution kerberos ticket is renewed at fixed intervals by CafEXE.

2.3. Monitoring
Users can monitor their jobs via a web monitor or via a command-line tool. A monitoring daemon running on the headnode periodically requests information from the glideins. The information, exported in XML format, is then published in a user-friendly way by the web monitor. The web monitor (Fig. 2) publishes detailed information about all jobs, running and completed. Users can check the status of their own jobs but also have a general overview of the status of the farm.

Users can also interactively monitor their job via a command-line tool which directly queries another monitoring daemon running on the headnode. This daemon implements unix-like commands like `ls,ps,tail,...` using the interactive monitoring implementation available in the glideinWMS: each glidein starts 2 different Condor virtual machines on the worker node, one is used to actually run the user job, the other is kept idle and waiting for "high priority" querying/monitoring jobs that the headnode sends on-demand when users access the command-line monitoring tool. These queries are, unlike the web interface, authenticated by a Kerberos

certificate, so that only the owner of the job can ask for specific job information and take actions on the job itself.

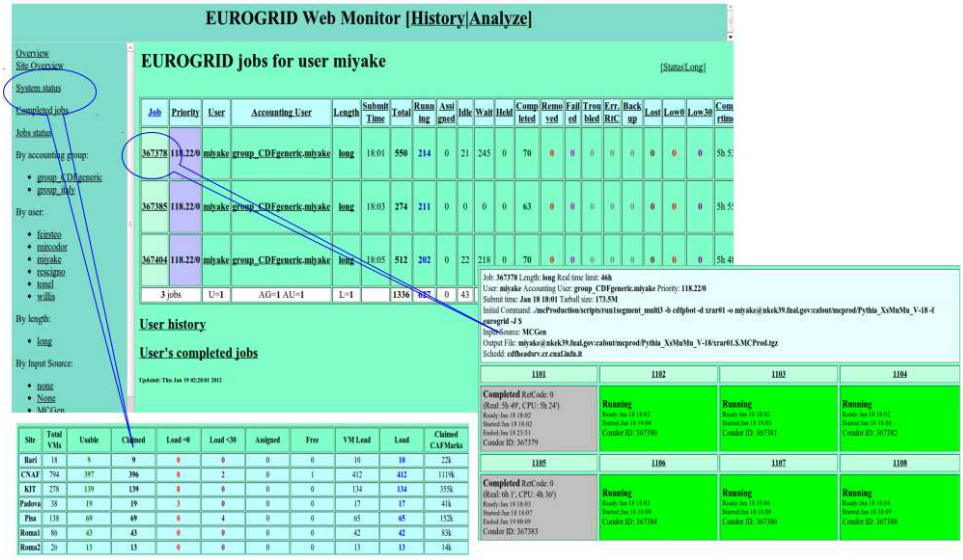

Figure 2. Eurogrid web monitor.

2.4. Database access and code distribution

Eurogrid offers users access to CDF offline code, data and online database.

The latest CDF offline software releases are available to users via AFS [9], in Eurogrid sites where AFS clients are installed.

A subset of CDF data is copied on disk at CNAF and is accessed through Sequential Access Model (SAM) [10], the data catalog used by the collaboration. A SAM station is installed at CNAF and allows users to access data as they do at Fermilab using CDFGrid.

Monte Carlo simulation jobs require information about run conditions, configuration, trigger, luminosity, alignment and calibrations; this information is stored on a dedicated Oracle database at Fermilab. CDF analysis programs access the database information through a multi-tier web-based system, FroNTier [11]. When a CDF Monte Carlo job is running on a worker node, database queries are translated into http requests by the FroNTier client which has been shipped to the worker node within the Monte Carlo tarball. Requests can be routed through the Squid proxy server installed at CNAF, which provides a caching for frequent queries of database information.

2.5. Matching resources and user needs

Users may have different requirements for their jobs: a specific operating system, the need to access data, availability of CDF code, etc., and it is fundamental that job submission is maintained as simple as possible and uniform among different sites. glideinWMS, thanks to the pilot-job approach, easily allows to match users needs with available resources, in a way that is totally transparent for users.

In Eurogrid the matching is performed through different steps:

- grid sites are divided into groups according to the availability of CDF software, data access, etc..; this information is set in a configuration file maintained by Eurogrid administrators;

- when a user submits a job, his requests are translated into a list of *desired* sites;
- the Frontend requires glideins to be sent only to sites in the desired sites list;
- matching sites are validated and setup using glideinWMS custom scripts: these scripts untar and setup specific compatibility libraries and kerberos binaries. Site specific environment variables (e.g. CDF software path) are also set.

In this way the site is ready for the requested job before any executable is started on its worker nodes and site-related problems are effectively masked from the job.

3. Performance

Eurogrid is operational since June 2011. Since the beginning, user response has been very good: on average 400 jobs/day are running on the portal, and peaks of 3000 jobs have been reached during the most intense periods of data analysis. About 40 CDF users regularly submit their jobs on Eurogrid, on average 5 users/day. Eurogrid is the most used job submission portal outside Fermilab. Eurogrid has demonstrated to be very reliable: during its first year of operations, it has experienced a negligible rate of failures due to site related errors. There have been only three major downtimes, of which one due to work maintenance on CNAF machines, the others to hardware failure and authentication problems respectively. In all cases, the impact on the users has been minimal: all jobs already running on the worker nodes were successfully completed and the output sent back to the users; those in waiting state were correctly resumed after the downtime. Compared to CNAFCaf and LcgCAF, on average Eurogrid processes the same number of jobs, but with higher efficiency. Moreover, the workload for maintenance is lower and the user's access easier with respect to previous portals.

4. Conclusions

Eurogrid is a new job submission portal for CDF data analysis, to access CDF computing resources at Tier-1 at CNAF and several Tier-2 sites in Italy and other European countries. It is based on glideinWMS, the glidein based workload management system. From the user's point of view, job submission is uniform to other CDF portals and submission to the grid completely transparent. From the administrator's point of view, it is easy to install and maintain even for non-experts. Glideins allow an efficient matching between job requirements and site resources and effectively mask site malfunctions from the job. Since its start in July 2011, the feedback from users has been very positive thanks to excellent performance in terms of reliability and job efficiency.

References

[1] Lucchesi D (Cdf Collaboration), *CDF way to Grid*, 2010 J. Phys.: Conf. Ser. 219 062017 doi:10.1088/1742-6596/219/6/062017
[2] Thain D, Tannenbaum T and Livny M 2005 *Distributed computing in practice: the Condor experience. Concurrency and Computation: Practice and Experience*, vol.17, no.2-4, pp.323-356. doi:10.1002/cpe.938.
[3] Sfiligoi I, Lipeles E, Neubauer M and Würthwein F 2004 The Condor based CDF CAF Proc. of CHEP 2004 CERN Indico Conf. id. 0, Contrib. id. 390.
[4] Delli Paoli F, Fella A, Jeans D, Lucchesi D, et al., *LcgCAF - The CDF portal to the gLite Middleware*, Presented at Computing in High Energy and Nuclear Physics, Mumbay, India, Feb 13-17, 2006, 148, (2006).
[5] Sfiligoi I et al. 2011 *Operating a production pilot factory serving several scientific domains* J. Phys.: Conf. Ser. 331 072031. doi:10.1088/1742-6596/331/7/072031.
[6] Sfiligoi I, Würthwein F, Dost J M, MacNeill I, Holzman B and Mhashilkar P 2011 *Reducing the Human Cost of Grid Computing With glideinWMS* Proc. of CLOUD COMPUTING 2011, The Second International Conference on Cloud Computing, GRIDs, and Virtualization pp. 217-221. ISBN:978-1-61208-153-3.
[7] Neuman, B.C., *Kerberos: an authentication service for computer networks*, Communications Magazine, IEEE, vol.32, issue 9, pp.33-28, doi:10.1109/35.312841
[8] `http://www.globus.org/grid_software/security/voms.php`

[9] http://www.openafs.org
[10] http://projects.fnal.gov/samgrid/WhatisSAM.html
[11] https://twiki.cern.ch/twiki/bin/view/CMS/SquidForCMS