PROCEEDINGS OF SCIENCE

PoS(HEP2005)395

CDF Way to the Grid

Subir Sarkar[*][†]

INFN-CNAF & INFN-Roma1, Italy

E-mail: subir.sarkar@cern.ch

Igor Sfiligoi

INFN-LNF, Frascati & Fermilab

E-mail: Igor.Sfiligoi@lnf.infn.it

Higher instantaneous luminosity of the Tevatron Collider and improved detector and trigger efficiencies require large increases in computing power for the CDF experiment in order to be able to meet future needs of data analysis and MC production. So far, CDF has been relying on a set of Analysis Farms (CAFs) with dedicated resources of commodity nodes managed as Condor pools, with a small CDF specific software layer on top of Condor. However, it has long been understood that significant expansion of the dedicated pools will be impossible in the future, and that CDF must move towards utilizing shared Grid resources. We have extended the existing CDF computing model by incorporating the Condor glide-in mechanism that can create dynamic Condor pools on top of existing batch systems, without requiring any additional software installation. All the monitoring tools supported on the dedicated CAFs, including semi-interactive access to the running jobs and detailed monitoring, have been preserved. CDF is also building an interface over the gLite tools for a broader access of resources and looking into other possibilities. In this paper, we will discuss about various CDF Grid initiatives and how they fit into the CDF plans.

International Europhysics Conference on High Energy Physics
July 21st - 27th 2005
Lisboa, Portugal

[*] Speaker.

[†] Contributing Authors: Stefano Belforte <*stefano.belforte@ts.infn.it*>, Armando Fella <*armando.fella@pi.infn.it*>, Shih-Chieh Hsu <*schsu@fnal.gov*>, Daniel Jeans <*daniel.jeans@roma1.infn.it*>, Elliot Lipeles <*lipeles@fnal.gov*>, Donatella Lucchesi <*donatella.lucchesi@pd.infn.it*>, Mark Neubauer <*msn@fnal.gov*>, Francesco Delli Paoli <*francesco.dellipaoli@pd.infn.it*>, Frank Wuerthwein <*fkw@fnal.gov*>

1. Introduction

The CDF experiment in the Tevatron Run II has been collecting Physics data since 2000. Over the years, improvements in accelerator running have resulted in a higher instantaneous luminosity, while that of detector and trigger conditions have steadily increased the data taking efficiency. CDF has so far collected about 1 PB of data. Analysis of a rapidly growing volume of data and a constant need of larger samples of Monte Carlo (MC) events require consistent growth of computing resources over time.

Thus far, in addition to the Central production and analysis farms at Fermilab (CAFs [1]), CDF has been successfully maintaining its own global computing environment with dedicated resources, in the form of 10 Decentralized Analysis Farms (dCAFs) and a few devoted MC production farms distributed over three continents. About 50% of the total CDF computing ($\sim$ 1.15 MSI2K) is done at remote sites.

However, expansion of the dedicated computing pools has virtually become impossible. Many sites have expressed intent to migrate towards shared resources. In fact, there are several sites, both in the LHC Computing Grid (LCG [2]) in Europe as well as the Open Science Grid (OSG [3]) in the US, with shared resources that could be available to CDF immediately, if the computing framework were designed to take advantage of it. A large fraction of these Grid resources is expected to be available to CDF before the LHC era starts, that will hopefully provide the required expansion of computing power for CDF. It is essential that the CDF computing model adapts to the Computing Grid paradigm and builds Grid enabled dCAFs, preserving all the CDF/CAF specific features such that the users do not experience a steep migration curve.

CDF is working on the following areas of development, ordered in terms of maturity, in order to gradually integrate its computing model into the Grid computing,

- CAF over Condor glide-in [4] extensions (GlideCAF)
- CAF built on top of the gLite [5] tools (LcgCAF)
- SAMGrid [6], direct use of the Grid tools etc.

2. GlideCAF

A GlideCAF is a simple extension of a regular dCAF with a slight modification of the Condor configuration and an addition of a glide-in factory, that create a virtual private Condor pool out of shared resources. Once a pool is established, it looks and behaves like a dedicated one, and allows sophisticated monitoring, both interactive and web based, as well as user and group scheduling policies to work in a manner identical to a dedicated Condor farm.

The virtual private pool is maintained by the glide-in factory, a process that keeps submitting Condor glide-ins to the Computing Element (CE) of the Grid site. On successful authentication, the CE delegates the glide-ins to the underlying batch system. Once a glide-in starts in one of the worker nodes of the local batch system, it talks back to the Headnode and joins the Condor pool. At this point, a regular CDF job can be started on that worker node using the CDF scheduling policies, as if it were a dedicated resource.

There are a number of points in favour of the Condor glide-in approach. The glide-in mechanism does not depend on the local batch system of a Grid site that makes deployment of a GlideCAF rather easy. When a glide-in starts on the Grid resource, it can perform a set of sanity checks before reporting back to the Headnode and pulling a user job. This way, broken nodes can be guarded against by the glide-ins themselves and no user jobs will ever be lost. Policy-making and monitoring also remain as good as a dedicated Condor pool since we gain complete control over the shared resources.

The first production GlideCAF [7] accesses the shared resources available at the Italian Tier1 centre at CNAF, Bologna. At the moment, the farm has a small share of dedicated resources (∼ 100 KSI2K) and a maximum of about 1.6 MSI2K accessible from the common pool. Figure 1 shows the usage pattern of the farm for the last six months in order to point out the turn-around with the Condor glide-in approach. Another GlideCAF has recently moved into production at San Diego (SDSC).

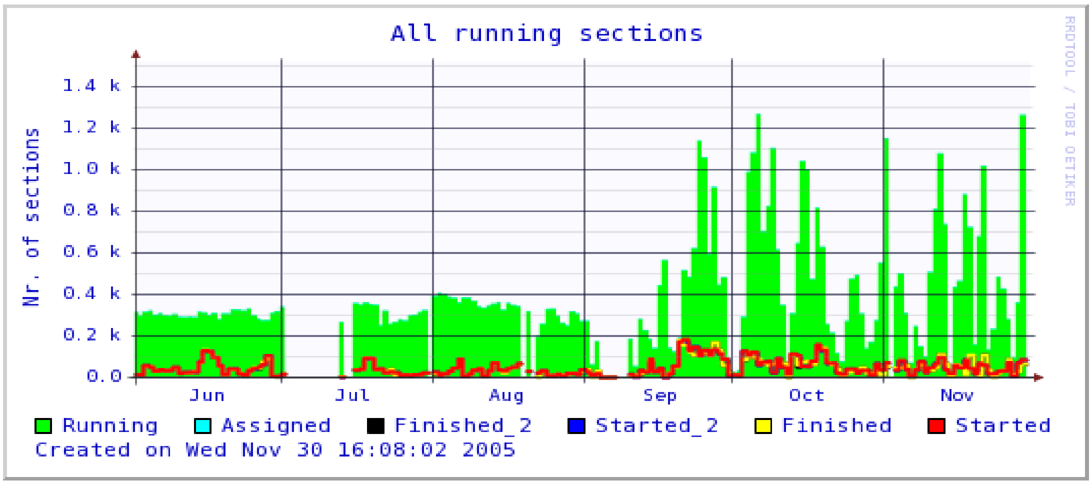

Figure 1: Usage pattern of the CDF farm at CNAF, Bologna during the last six months. The production farm started to access the common pool, using the Condor glide-in mechanism, from September ’05 which is marked by a substantial increase in the resources available to CDF.

In addition to the production farms mentioned above, several other GlideCAFs are in various phases of development at other Tier1 centres accessible to the CDF experiment.

The current incarnation of the GlideCAF has its share of weaknesses also. The major problem is that no firewall is allowed between the Headnode and the worker nodes which essentially means that the Headnode must be installed inside the Grid site domain. In order words, a GlideCAF must be installed for each and every site in order to access its resources. Although this is attainable for major Grid sites, like the Italian Tier1 in Bologna, it is a severe limiting factor for the CDF way to the Grid.

The limitation mentioned above is due to the way the Condor daemons talk to each other. However, the latest version of Condor (v6.7.13) has introduced a proxy mechanism for internal Condor communication which promises to solve the firewall problem. We have just started experimenting with it, but we are confident that it will allow us to create a GlideCAF that will be able to access resources from many Grid sites. The final goal for CDF will be to have a single submission point that will run the CDF jobs on all the available Grid sites.

3. LcgCAF

A Grid standard based approach, that does not suffer from the limitations of the present GlideCAF as discussed in the preceding section, is to make use of the gLite Workload Management System (WMS) that can act as a single point of submission to all the LCG. The concept behind the LcgCAF is to re-implement the CAF infrastructure on top of the gLite middleware while still maintaining the same interface for the end-users.

The LcgCAF accepts user jobs and delegates to the gLite WMS. It also handles authentication/security, job monitoring and data handling tasks. The web based monitoring for the LcgCAF relies mostly on the LCG Logging and Book-keeping information. Interactive monitoring, on the other hand, is based on the Clarens Web Services Framework.

The LcgCAF, in alpha test at present, has been successfully producing CDF MC events over many INFN-Grid sites. It might be noted, however, that the scope of the LcgCAF is limited to LCG sites, at least for the near future.

4. SAMGrid, Direct use of the Grid Tools etc.

SAMGrid is a Fermilab endeavour which is adapting to the Grid world and promises a good integration with the LCG and the OSG sites. However, the user interface is complicated and alien to the CDF users which may somewhat limit its usage in CDF.

Job submission to both LCG and OSG using the Grid tools directly is always an option. Grid tools are general purpose and large scale production of events can be managed reasonably well by a trained few. However, it may not be suited for generic user analysis.

5. Conclusion

Need for more computing power has pushed the CDF experiment to revise its computing model and adapt to the Grid computing paradigm. The first successful step in this direction was to add the Condor glide-in mechanism to a few dedicated Condor pools, which have access to shared resources, to build GlideCAFs. We are also actively pursuing more general approaches for a much wider access of common resources.

References

[1] I. Sfiligoi *et.al.*, The Condor based CDF CAF, Proceedings of the CHEP2004 Conference

[2] The LCG Home Page, http://lcg.web.cern.ch/LCG/

[3] The OSG Home Page, http://www.opensciencegrid.org

[4] The Condor Project Homepage, http://www.cs.wisc.edu/condor

[5] The gLite Home Page, http://glite.web.cern.ch/glite

[6] The SAMGrid Home Page, http://www-d0.fnal.gov/computing/grid

[7] S. Sarkar and I. Sfiligoi, GlideCNAF - A Purely Condor Glide-in Based CAF, CDF Note 7630 (2005)