

Building a distributed software environment at UCL utilising a switched light path

V. Bartsch¹, N. Pezzi¹, M. Lancaster¹
¹University College London

July 5, 2006

Abstract

The amount of data produced in high energy experiments like CDF (Collider Detector at Fermilab) [1] requires to distribute the computing and the data for the scientists analysing data of the experiments. Often remote computing resources are shared throughout several high energy experiments, although the underlying grid software differs from experiment to experiment. Here the situation is described for the computing set up at UCL (University College London) for both the local users and all users for the experiment CDF. Grid software and techniques developed for the CDF experiment are used and a dedicated switched light path between Fermilab and UCL is utilised.

1 Introduction

CDF is a particle physics experiment investigating the fundamental nature of matter. It is presently taking data from proton anti-proton collisions at the Tevatron at Fermilab, which is located just outside Chicago in the USA. The experiment currently produces approximately 1PB of raw data per year and will continue to do so until 2009. Analysis of this data is underway by almost 800 physicists located at 61 institutions in 13 countries across 3 continents. The amount of raw data and the need to produce secondary reduced datasets have required new distributed storage and analysis. Grid systems based on DCAF [2] and SAM [4] have been developed and deployed during the last year and the focus is now to make the DCAF systems interoperable with other grid

systems, e.g. LCG [5] sites. The aim is that 50% of CDF's CPU and storage requirements will be provided by institutions remote from Fermilab. In order to effectively utilise this distributed computing network it is necessary to have high speed point to point connections, particularly to and from Fermilab, which have a bandwidth significantly higher than commonly available. To this end, as part of the ESLEA [6] project, the use of a dedicated switch light path from Fermilab to UCL in the UK has been optimised.

This paper describes the experiences utilising grid middleware in a switched lightpath environment for the CDF experiment.

2 CDF's Data Handling and Analysis Systems

The analysis model of data in high energy physics is highly sequential and is generally carried out on dedicated Linux PC analysis farms. These farms may be shared between high energy physics experiments. Since there are still several approaches to implement a world wide computing grid it becomes more and more important to design grid systems in a way that they are interoperable with each other. The CCC grid cluster at UCL [7] is a CERN Tier2 center which utilises the LCG approach to the grid. CDF however chose to use a combination of the SAM and the DCAF system, which are explained subsequently, as a grid system.

The core capabilities of the grid enabled data handling system SAM are bookkeeping

of metadata and locations of data files and transfers of the files to the nodes to process them. In order to be used at a remote site several services need to be running remotely which are summarised as a SAM station. The SAM station communicates with a central service, the so-called database server via CORBA in order to get the location of a file and is able to transfer the file with various transfer mechanisms including gridftp. The new location is reported back to the central database. Fig.1 shows the data consumption of all SAM stations last year. The major amount of data has been consumed at FNAL because SAM is used both onsite and remotely. The SAM stations at UCL have started to import data beginning of this year due to a downtime of the UK light link autumn until winter last year.

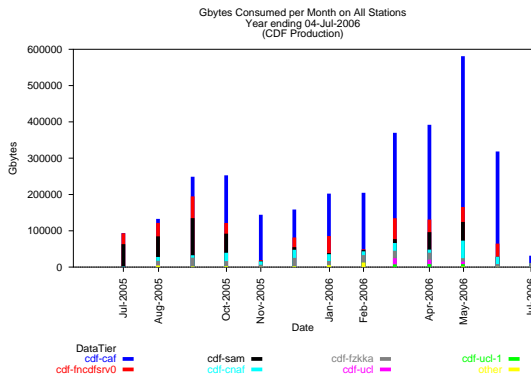


Figure 1: *Data consumption of all SAM stations last year. cdf-caf, cdf-fncdsv0 and cdf-sam are local to Fermilab, cdf-cnaf is located at Bologna (Italy), cdf-fzkka at Karlsruhe (Germany), cdf-ucl and cdf-ucl-1 at London (United Kingdom).*

DCAF systems provide authorization mechanisms (currently kerberos) and job handling mechanism to run CDF jobs in case the CDF software is NFS mounted on all worker nodes. In order to run on LCG clusters the job is submitted to the globus gatekeeper at the LCG headnode and uses during runtime the Condor [8] glidein mechanism by which one or more grid resources temporarily join a local Condor pool. The so-called GlideCaf is described in more detail at [3]. The features of this pool

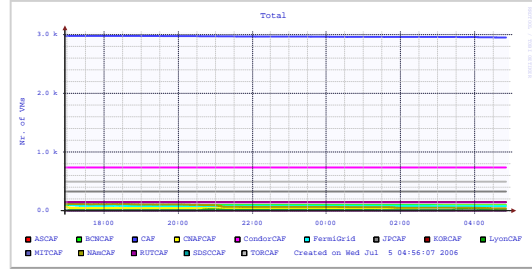


Figure 2: *Total number of virtual machines (VMs) at the DCAFs during the last year: The main one are CAF, CondorCAF, FermiGrid at FNAL. JPCAF, KorCAF and ASCAF are located in Asia, NAmCAF, TORCAF and MIT-CAF in the US/Canada, CNAFCAF in Italy.*

are changed so that it appears like any DCAF cluster. In order to utilise a 200PC cluster at UCL this technique will be used, it has already been tested at a smaller cluster at UCL.

Fig. 2 shows the total number of virtual machines (VMS) at each CAF worldwide. The main resources (about 4kVMS) can still be found at FNAL, about 2k VMs are located outside of FNAL. Some of the DCAFs are not using the LCG submission mechanism but dedicated resources (e.g. ASCAF and JPCAF), others use the glidein mechanism, for example LyonCAF and FermiGrid. One can see that those resources claim not to have any CPUs at all, but when a job gets submitted they are executing through the Globus gatekeeper and therefore running on a cluster which does not appear in the monitoring. Fig.3 shows the usage of the DCAFs throughout the last year. One can see that the central resources show a much higher usage than the remote DCAFs. This is partly due to the fact that the import of necessary data from FNAL takes long, partly due to downtimes of the resources. The UCL glidein CAF is not yet officially monitored.

3 Utilisation of the UK/Star Light switched light link

Typical CDF secondary datasets are presently 1-50 TB in size. The CPU resources required to repeatedly analyse such datasets will ex-

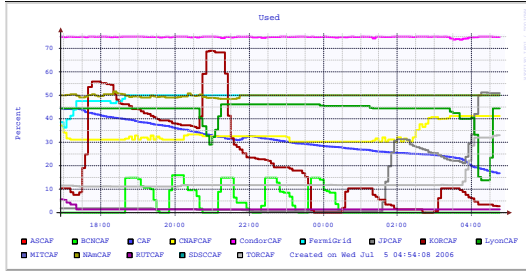


Figure 3: Usage of the DCAFs in percentage during the last year.

ceed those available at Fermilab and so the datasets need to be distributed to centers in Europe and Asia to facilitate the use of their CPU resources. Typical transfer rates from Fermilab to Europe (UCL) using the standard network are approximately 25 Mbit/sec (for multiple streams). The transfer of a single dataset would take months. Therefore most DCAFs are specializing in dedicated branches of physics and the according datasets. UCL has the possibility to transfer data over the dedicated switched light path from Fermilab (using the US Starlight network[9]) to UCL in the UK (using the UKLight network[10]) with a bandwidth of 1 Gbit/sec. During the last year several integrity and bandwidth tests have been performed which showed that network equipment in the path between the CDF storage area at Fermilab to the Starlight network was causing an unexpected slowdown and therefore these network switches needed to be replaced. After this replacement sustained data transfer rates from the CDF storage at Fermilab to the disks at UCL have been at a maximum of 550 Mbit/sec at a 2 hours average. Typical data transfer rates are shown at Fig. 4. Therefore it is possible for the users to import datasets on the fly rather than the site administrator caring for the transfer.

4 Conclusion

The software setup allows users from UCL and CDF to utilize the computing environment of the local cluster at UCL. A further extension of the services of the Tier-2 center at UCL in

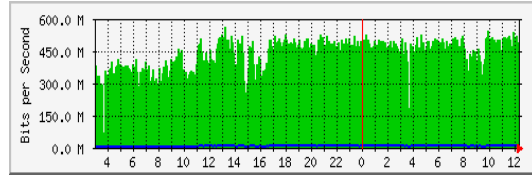


Figure 4: Snapshot of the data transfer rate at the UK light link during one day. A sustained rate of 550 Mbit/sec is feasible.

on the way. The dedicated switched light link between Fermilab and the UK delivers a sustained data transfer rate from disk to disk of 550 Mbit/sec and therefore allows the users to choose the data to analyse without high delay due to data transfer. This is unique compared to other remote data analysis centers of CDF which allow to run only on the data which was copied due to the decision of the site administrators.

References

- [1] <http://www-cdf.fnal.gov/>
- [2] <http://cdfcaf.fnal.gov>
- [3] GlideCaf - a late binding approach to the Grid, I. Sfiligoi et al., CHEP06 conference proceedings
- [4] <http://d0db-prd.fnal.gov/sam/>
- [5] <http://lcg.web.cern.ch/LCG/>
- [6] <http://www.eslea.uklight.ac.uk>
- [7] <http://wiki.gridpp.ac.uk/wiki/UCL-CCC>
- [8] <http://www.cs.wisc.edu/condor/>
- [9] <http://www.startap.net/starlight/>
- [10] <http://www.uklight.ac.uk>