Tools for GRID deployment of CDF offline & SAM data handling systems for Summer 2004 computing

A. Kreymer[1], A. Baranowski[1], V. Bartsch[2], S. Belforte[3], M. Burgon-Lyon[4], G. Garzoglio[1], R. Herber[1], R. Illingworth[1], R. Kennedy[1], U. Kerzel[5], M. Leslie[2], L. Loebel-Carpenter[1], A. Lyon[1], W. Merrit[1], F. Ratnikov[6], R. St. Denis[4], A. Sill[7], S. Stonjek[2,3], I. Terekhov[1], J. Trumbo[1], S. Veseli[1], S. White[1]

[1]Fermi National Accelerator Laboratory, [2]University of Oxford, [3]Istituto Nazionale di Fisica Nucleare, [4]University of Glasgow, [5]Universität Karlsruhe, [6]Rutgers University, [7]Texas Tech University
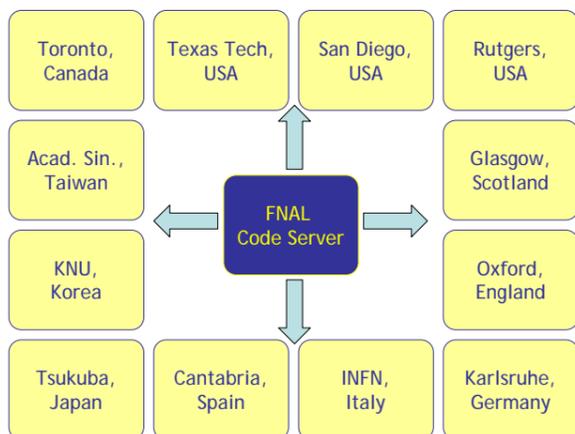
Introduction

The Fermilab CDF Run-II experiment is now providing official support for remote computing, which has provided approximately 35% of the total CDF computing capacity during the Summer of 2004.

Approach

We face the challenge of unreliable networks, time differences, and remote managers with little experience with this particular software.

The approach we have taken has been to separate the data handling components from the main CDF offline code releases by means of shared libraries, permitting live upgrades to otherwise frozen code. We now use a special 'development lite' release to ensure that all sites have the latest tools available. We have put substantial effort into revision control, so that essentially all active CDF sites are running exactly the same SAM code.


Monitoring Tools

We have created light weight testing and monitoring tools to assure that these sites are in fact functional when installed. These are used in conjunction with externally developed tools such as Ganglia.

SAM TV allows a user or SAM Shifter to see the details of any registered SAM Station, the projects presently running, the file transfers taking place, and the status of each process.

DCAF monitoring allows all the batch queues for each DCAF site to be viewed, giving usage plots for
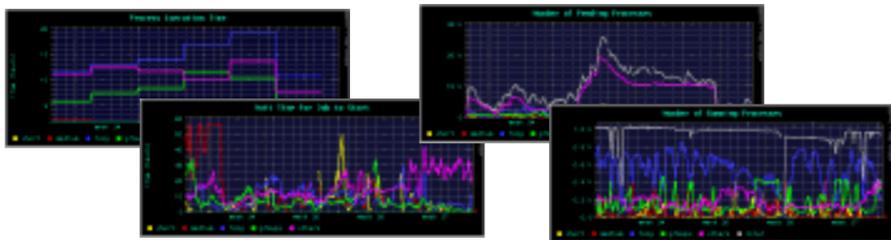
DCAF using FBSNG and Condor

The original CAF (CDF Analysis Farm) was controlled using FBSNG (Fermilab Batch System Next Generation) based software. A new version that uses Condor as the underlying batch system is now the standard for new DCAF (Distributed CAF) installations.

Deployment of SAM

We have made the first deployment of the SAM data handling system outside its original home in the D0 experiment. We have deployed to about 20 remote CDF sites. We are distributing and configuring both client code within CDF code releases, and the SAM servers to which the clients connect. Complete station installations, which once took days, are now performed in minutes. Simplification of this process was a crucial step for deployment.

SAM installation on a new cluster requires only a local SAM account. Root access is not required, removing headaches for both the remote system administrators and the CDF users.

Reuse Potential

The tools developed for CDF can be used to install SAM servers for D0 and other experiments. As new experiments, such as MINOS, expand their use of SAM the development and deployment of common tools become more important.

Further Information

The following links show the suite of CDF offsite tools in action and give information on how to set up a remote site:

Overview of which stations are running
http://cdfdb-prd.fnal.gov/sam_local/SamAtAGlance/

SAM TV for individual station, project and file monitoring:
http://ncdf151.fnal.gov:8520/samTV/current/samTV.html

DCAF queues monitoring links and instructions for setting up a remote site:
http://cdfkits.fnal.gov/DIST/doc/DCAF/

CDF SAM Home Page:
http://projects.fnal.gov/samgrid/cdf/cdf.html