Experience using grid tools for CDF Physics

M. Burgon-Lyon[a], A. Stan Thompson[a], Igor Terekhov[b], Richard St.Denis[a], Gabriele Garzoglio[b], Stefan Stonjek[c], Parag Mhashilkar[b], and Vijay Murthi[b].

[a]University of Glasgow, Scotland, United Kingdom

[b]Computing Division, Fermi National Accelerator Laboratory, Batavia, IL, USA

[c]University of Oxford, England, United Kingdom, based at Fermi National Accelerator Laboratory

We discuss our experience configuring the grid fabric at various Institutions participating in the CDF experiment, and its use for Monte Carlo production, data reconstruction and secondary data set analysis using SAM (Sequential Access via Metadata) and JIM (Job Information Management) with Condor grid tools.

1. Introduction

SAM (Sequential Access via Metadata) [1] is a mature data handling system that has been in use at Fermilab since 1997. Originally developed for the D0 experiment, it is now used by other experiments including CDF (Collider Detector at Fermilab) [2]. JIM (Job and Information Management) [3] is a grid extension to SAM. A remote JIM installation is comprised of one or more of the following components: client, submission, monitoring and execution.

A client component allows users to submit jobs and is configured to forward job submissions to a specific submission component. This submission component then queues incoming jobs pending a decision from the broker. A centralised remote broker decides on which site the job will run and communicates this information to the submission component. Given the remote broker's decision, the submission component forwards the job to the appropriate execution location. On receiving a job, the execution component runs the job using the local batch system and returns the output to the submission component. The monitoring component allows the user to keep track of job progress. SAM and JIM combined make up SAMGrid [9,6].

2. Deployment

The hardware used for testing comprised two Dell clusters, one located at the University of Glasgow whilst the other at the University of Oxford, used in combination with a submission site at Fermilab. The Fermilab site submitted jobs to ScotGrid[8]. When the initial installations were completed ScotGrid comprised 59 dual 1GHz Pentium III worker nodes and 5TB of disk located at the University of Glasgow. This resource was recently expanded with an additional 28 dual 2.4GHz Xeon nodes. The two identical Dell clusters comprise 9 worker nodes and over 8TB of disk.

Initially the Dell cluster at Glasgow University was given a complete JIM (client, submission, monitoring and execution site) and SAM installation. The same installation was completed on a front-end to ScotGrid and the Dell cluster at Oxford University. The Glasgow University setup is currently being altered to entail a single submission site and a single monitoring site. Both clusters will continue to provide client and execution site software. This change makes better use of the available resources by reducing the amount of software running.

The SAM data handling component required a samserver service certificate to operate. In addition to this JIM also required a host certificate. The Dell cluster required a port to be opened on

ScotGrid to allow monitoring of the ScotGrid execution site. Submissions made to the ScotGrid front-end go to the submission queue on the Dell cluster.

3. Components of SAM

SAM was written in response to the data processing challenge faced initially by D0 and subsequently CDF [7] by optimising the use of data storage and delivery resources such as tape mounts, drive usage and network bandwidth.

SAM stations are a set of servers that coordinate local file management, interacting with the central database. Each running SAM station has a local cache for retrieved files. The stager daemon is a local cache manager. When a user submits a script to a SAM station, the SAM station instanciates a project master that is responsible for coordinating access to the files to the consumers. The stager locates the files and starts transfer of the files to the local cache. Once the files begin copying, a consumer is registered to the project master to use the files requested in the script.

SAM keeps track of the locations of each file. New files, such as the output of Monte Carlo simulation, can be stored in SAM, usually in Enstore [5] the mass storage system at Fermilab, and delivered via SAM to any SAM station that requests them. In addition to Data Management (data transfer and replica management), SAM also allows the description of data and the booking of processes using the metadata catalog.

4. Components of JIM

JIM can be divided into two logical parts. Job Handling is the transfer of the job from the client software, via the submission software to an execution site and also transfer of the output sandbox back to the submission software. Monitoring and Information is the resource information, replica catalogue and logging and bookkeeping. These elements use SAM for data handling.

The JIM job handling client may submit jobs to the request broker, via a submission site, which uses the Condor match making service to automatically select the job execution location[10]. This has been extended to use extra criteria for the match. A site's resources are described using XML with subsequent projections onto the Condor Classified Advertisements (ClassAd) framework. Condor-G middleware has been enhanced for JIM to enable scheduling of data-intensive jobs with flexible resource description using the submission component. The job would then be run on the local batch system of the selected execution site. SAM may be used to retrieve large volumes of data for the execution site.

JIM distinguishes grid level scheduling (selection of an execution site) from local scheduling (distribution of the job within the cluster). JIM also distinguishes between structured jobs (where the details are known to grid middleware), from unstructured jobs (where the whole job is mapped onto a single cluster) [4].

5. Grid Job Types

SAMGrid will be used mainly to process three types of grid submission. These are analysis, Monte Carlo simulation and reconstruction.

Analysis typically takes a large number of files, referenced as a dataset, as input. These files are located and copied by SAM to the execution site, where the analysis job will run. The output of an analysis job is typically very much smaller than the input dataset and can be downloaded using the web interface to SAMGrid. Several test analysis grid jobs have been successfully run using SAMGrid. The choice of execution site was determined by the amount of requested data already present at a site in some cases, and selected in the input file in others.

Monte Carlo simulation involves the generation of events which are passed through a detector simulation and then reconstructed. The events are generated using a defined generator and parameters which are passed to the execution site using the input sandbox. These typically have no input data files and one large (1GB) output data file. This file is stored in SAM. The output sandbox available online only contains the log files from the simulation, not the actual output.

Files in SAM are typically grouped into

datasets. Once in SAM the file can be copied to any machine that has SAM installed. A user can do this using the Dataset Definition Editor [11].

Reconstruction involves the delivery of a file containing real data to the execution site. After processing, the output file is stored as a new file in SAM.

6. Testing Approach And Problems

The main problem encountered when testing the CDF part of SAMGrid was stability due to version control. On installing the SAM and JIM software at the remote sites at Glasgow and Oxford Universities the versions would work together. However as the broker software was upgraded at Fermilab problems were encountered with the configuration at remote sites. In part this was due to ongoing development for two different experiments, with differing priorities. As the JIM software has matured, it has become more stable.

To ensure that all components of SAM and JIM were installed and configured correctly, the following checks were carried out on each installation. Firstly the batch system, in this case PBS for the three execution sites, was tested. This was done by running a small job using the PBS command `qsub`, which was submitted on both the headnode, and on a worker node. Next the SAM station software for execution sites were tested with three steps. The first was a SAM command to check the station was working. This was followed by three scripts to ensure that multiple files could be retrieved from SAM. Two analysis jobs were then run using SAM, one using a dataset on the local SAM cache, the other using a retrieved dataset.

JIM was then tested with a very simple submission. By selecting the station on which the job would run in the `.jdl` job submission file, an analysis test job that uses a dataset stored in the local SAM cache was submitted to JIM. Likewise another analysis test job that used a retrieved dataset was submitted.

For client only installations, a simple JIM submission, followed by both types of analysis test job as JIM submissions were used.

7. Further Work

In January 2004, a CDFGrid workshop was held at the University of Florida proceeding the CDF collaboration meeting. During this workshop installations of distributed Central Analysis Farm (dCAF) batch processing system, SAM Station data handling system and Distributed Cache (dCache) were begun on PC's at eleven institutions, including three in Asia and four in Europe. Six of these sites have committed to usage of their facilities for all of CDF (mostly in the form of Monte Carlo simulation) for Summer 2004. Since the workshop the number of running sites has reached twenty.

The next step is to install the JIM software at these institutions, allowing SAMGrid to be used. Another workshop has been arranged for 1st April 2004 for this purpose. The target is to have 25% of CDF computing from external resources by June 2004 and 50% by June 2005. It is intended that all CDF computing will be on CDF Grid by April 15th 2004. The CAF software with which many CDF users have become familiar uses the same user interface regardless of whether it is configured for a single cluster or as a grid tool. This removes the need to train and encourage users to migrate to the new tool set.

The D0 experiment has made progress running Monte Carlo simulations using SAMGrid with recent success rates at 85-90%. CDF will move to the latest JIM code that incorporates the changes made for D0 and expect to achieve similar results. As Monte Carlo is a priority for both experiments, the development team are focusing on improving these success rates for both experiements. Future work will include further testing of Monte Carlo simulation, Analysis and Reconstruction physics jobs in addition to enhancing monitoring for bottle-neck detection, load analysis and load balance efficiency.

REFERENCES

1. SAM home page `http://fcdfsun1-prd.fnal.gov/sam/`

2. CDF home page http://www-cdf.fnal.gov/
3. JIM home page http://www-d0.fnal.gov/computing/grid/
4. G. Garzoglio, I. Terekhov, A. Baranovski, S. Veseli, L. Lueking, P. Mhashikar, V. Murthi, The SAM-Grid Fabric Services, Advanced Computing and Analysis Techniques in Physics Research (ACAT) 2003.
5. http://www-isd.fnal.gov/enstore/
6. A. Baranovski, *et al.*, The SAM-GRID project: architecture and plan, Advanced Computing and Analysis Techniques in Physics Research (ACAT) 2002.
7. I.Terekhov, Meta-Computing at D0, Advanced Computing and Analysis Techniques in Physics Research (ACAT) 2002.
8. http://www.scotgrid.ac.uk/
9. A. Baranovski, G. Garzoglio, L. Lueking, D. Skow, I. Terekhov, R. Walker, SAM-GRID: A System Utilizing Grid Middleware and SAM to Enable Full Function Grid Computing, Beauty 2002.
10. A. Baranovski, *et al.*, Computing for High Energy Physics (CHEP) 2003, La Jolla, CA.
11. Dataset Definition Editor http://fcdfsun1-prd.fnal.gov/sam_project_editor/