

# **SAMGrid Status Report**

**Adam Lyon, 31 May 2005 GDM**

## **1 Project Drivers, Scope and Milestones**

### **1.1 Inter-VO Grid Access to Fermilab Resources [COMMON]**

Parag has prepared a document detailing the design and integration of standard grid VO services (VOMS and VOMRS) with DØ and is working on its implementation (under "*SAMGrid: Deploy on Fermigrid, LCG & OSG*" project). See [http://www-d0.fnal.gov/computing/grid/doc/Samgrid\\_VO\\_Proposal.pdf](http://www-d0.fnal.gov/computing/grid/doc/Samgrid_VO_Proposal.pdf). Parag has also been in communication with CDF to learn how they utilize VO services and if the work he is doing can be useful to them.

### **1.2 Fermilab Computing and Strategic Storage Resources Accessible to the Open Science Grid [COMMON]**

Gabriele Garzoglio and Andrew Baranovski (under "*SAMGrid: Deploy on Fermigrid, LCG & OSG*" project) have guided the Fermilab CMS cluster using OSG infrastructure through installation and certification. It is now certified for DØ reconstruction and is undergoing certification for merging. Andrew is giving a presentation at the OSG Applications Meeting at SLAC in early June on SAMGrid/OSG efforts.

### **1.3 Minimize Operational and Deployment Loads [COMMON]**

See other sections for details on the SAMGrid DH and JS deployment projects.

### **1.4 DØ Reprocessing [DØ]**

Gabriele, Andrew and Parag have been supporting the DØ Reprocessing effort's use of SAMGrid for job and data management (under "*SAMGrid: Thick Job Manager Development*" and "*SAMGrid JS Deployment to Production [DØ]*" projects). Westgrid and in2p3 are still the main contributors of reprocessed events. As of May 23, 200 million events have been reprocessed (out of 1 billion desired). 100 million done remotely. The LCG integration has overcome some major hurdles. The SAMGrid forwarding node can now submit jobs to the LCG grid, but jobs are failing when the sandbox is transferred. A fix is under test. SAMGrid and OSG integration is covered in section 1.2.

## **1.5 Large Pick Events [DØ]**

As mentioned last month, DØ would like to protect the SAM cache from large pick events projects. The current implementation of groups and quotas does not work in the d0mino-less station configuration (DØ now has many cache disks on many nodes). Igor Mandrichenko (on loan to the SAMGrid team) has been communicating with DØ to design an alternative cache resource strategy. Igor is working out some final details and will start implementation within the SAM station framework.

## **1.6 Current Stable Operation [DØ]**

- Members of the SAMGrid team rotate a weekly expert shift. Issues are first filtered by the DØ SAM Shifter.
- The issue tracker has been in use for the past month at DØ. Overall, it works quite well but it is not without its annoyances. I will be working with Marc Mengel to come up with fixes. Perhaps will introduce it to CDF later in June.
- The "MIS" project is ongoing. Steve Sherwood and Lauri Carpenter are charged with load testing and deployment. Sinisa is working on a prototype Information Service (meant to reduce SAM station reliance on the central database).

## **1.7 User Analysis using Grid Resources [DØ]**

Resources cannot be devoted to this driver at this time due to the DØ Reprocessing. Though of course work and investigations for the Reprocessing may be applied to learning how to handle user analysis jobs submitted via the JS part of SAMGrid.

## **1.8 Production use of SAM for MC [CDF]**

Gabriele, Valeria and Andrew have a system ready for testing, including merging as per CDF's merge specifications. They are awaiting CDF personnel to do the testing. The main risk here is again CDF's commitment to using SAMGrid for MC production (there apparently is competition from another system).

## **1.9 Production Use of SAM on CAF & DCAF [CDF]**

This driver is still one of the major sources of activity in the SAMGrid group.

- The hardware has been deployed and is in use. A nameserver migration must be performed so that remote sites can utilize the new location of the nameserver. This migration has been planned and fully documented. We are waiting for CDF to implement it.
- Alan Sill, Robert Illingworth, Sinisa and Lauri have released a "frozen" version of the SAM client software. The frozen client puts the python interpreter and all client libraries into one executable. This client is

significantly faster (by over an order of magnitude!!!) than the non-frozen client (python libraries spread out over many directories). The frozen client also is much easier to deploy and use in environments where a different version of python is used by the experiment. Lauri has also made many speed up enhancements to the command builder of the client.

- While the client will be frozen, there is little advantage to freezing the station (mostly C++, not python) and the DB server. Problems have been cropping up in the DB Server due to CDF's use case of handling huge datasets (tens of thousands of files in one dataset -- DØ tends to use datasets of less than 1000 files). Lauri has recently discovered a speed up which will hopefully solve many of the large dataset problems. This fix will be in testing.
- There do seem to be some conflicts with the frozen client and the CDF CAF submission scripts. These are being worked out.
- Randolph Herber's Data Base Browser has been updated to handle SAM browsing and is in use at CDF and MINOS. It will soon be introduced to DØ.
- I am still awaiting a description of the planned full CAF load test from CDF.

### ***1.10 Production Use of SAM on Reconstruction Farm [CDF]***

CDF has changed details of this deployment many times (now calling it "Phase 2"). They are currently using the "prototype" or "phase 1" SAM farm for production reconstruction. The phase 2 farm will use dCache read and write pools. Some dCache problems are under investigation and must be solved before they migrate.

### ***1.11 Deployment of SAM [MINOS]***

Art Kreymer is installing SAMGrid products and configuring components on MINOS systems for production. He has now installed the frozen client. I know of no issues with this deployment.

### ***1.12 V6 Deployment at DØ***

In keeping with the experiment schedule, v6 is now the default SAM station and client for clued0 (the desktop cluster), though the v5 station still exists. The frozen client is in testing. For the June down day (June 7th), the v5 station will be removed and half of the main analysis farm will be upgraded to v6. In July, the other half with the large cache disks will be upgraded to v6. Though the cache disk servers will not be upgraded until July, we will upgrade them anyway in June and immediately downgrade them -- this will be done to test the v6 DB servers against station restart problems involving large number of cached files.

### **1.13 Enth Distributed Dimensions [SBIR II]**

Matt Varanicar is putting more manpower on the Enth integration project.

### **1.14 Packaging and Configuration**

See sections 1.9 and 1.12.

### **1.15 Request System**

Steve Sherwood has joined Lauri to finish the MC request system for both DØ and CDF. The system is now in final testing.

### **1.16 Adam's new family addition**

No baby yet (due 6/8). Doctors are now checking my wife three times per week (ugh!). Baby furniture is in place and ready. Our dog has no idea that she will soon no longer be the baby of the family.

### **1.17 Other projects not started**

Station-SRM integration (on hold - prototype completed); SQL builder; RCP to SAM parameters (CDF request); Process bookkeeping (CDF/DØ design request); JIM Job Brokering.

## **2 Main Project Milestones**

- SAM on CDF CAF (need to define and execute CAF load test)
- SAM on CDF Reco Farm (need to migrate to "phase 2" farm with dCache)
- SAMGrid v6/v7 (frozen) on DØ systems (clued0 is using v6 station and client; analysis farm migrates in June and July)
- SAMGrid LCG/OSG interoperability (CMS Farm using OSG infrastructure is being merge certified; can submit jobs to LCG via forwarding node - sandbox problems are being worked out)
- SAMGrid for CDF MC (ready for testing by CDF)
- Packaging/configuration Freezing (deployed - being tested by CDF and DØ)
- MC Request system (in final testing)
- MIS server used for SAMGrid specific monitoring with SAM HDTV (estimate one month)

## **3 Effort**

Fermilab CD effort is 9.5 FTE's as of April 2005

- 100%: Andrew, Gabriele (effective), Lauri, Sinisa, Parag, Krzysztof, Valeria, Steve Sherwood (new hire)
- 50%: Randolph, Adam, Steve White, Robert, Igor Mandrichenko

Operations support

- Art, Dehong, Liz Buckley-Geer

April Effort Reporting (9.75 FTE total from team)

- 6 FTE Core development
- 1.5 FTE Deployment to production
- 1.5 FTE Operations
- 0.75 FTE Project Management
- 0 FTE Outreach

May Effort Reporting (9.5 FTE total from team)

- 5 FTE Core development
- 2.25 FTE Deployment to production
- 1.25 FTE Operations
- 0.75 FTE Project Management
- 0.25 FTE Outreach

## 4 Risks

The main risks, as I see them, are

- Issues with SAM DH brought up by CDF full CAF testing. We still do not have a testing plan from CDF (but should get one soon). Problems with DBServers are looking fixable, but are we hitting some python limit?
- Scope/Feature creep. Jerry has been instrumental in keeping the deployment task list on track.
- For SAMGrid, the risk is delivery of an interoperable solution for OSG/LCG. The reliance on external middleware that has not been production-battle tested is a risk.
- CDF's use of SAMGrid for MC production or another system.