

SAMGrid Status Report

Adam Lyon, 12 April 2005 GDM

1 Predefined Project Drivers, Scope and Milestones

The following sections detail drivers and projects listed on the V0.13 (12/14/04) spreadsheet.

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

Parag Mhashilkar has recently joined the SAMGrid team. Along with providing some relief for Gabriele, has been charged with working on migrating VO management from the SAM based methods to VOMS, as part of the "*SAMGrid: Deploy on Fermigrid, LCG & OSG*" project. Over the next few weeks he will be preparing a document detailing the work needed and investigating issues. The goal is to remove the gridmap and VO information from the SAM database. Risks are how easy it will be to accommodate the extra information SAMGrid keeps about users (e.g. group membership) that would perhaps not be included in the VO database and how to handle the registration process (the SAM registration will need to be replaced).

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

Gabriele Garzoglio (under "*SAMGrid: Deploy on Fermigrid, LCG & OSG*" project) has been helping Joel Snow with the installation of SAMGrid software and cluster configuration on the Fermilab CMS farm. Joel has reached the testing phase.

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 and Andrew 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). Gabriele has been assisting various remote sites with operations and testing. Several sites including WestGrid and IN2P3 are actively reprocessing now. The DØ Reprocessing Group is certifying other sites. Gabriele performed scalability tests, running 700 concurrent jobs on one farm. 99.9% of the jobs went through successfully. Failures were due to overloading an underpowered node hosting the XMLDB. Testing will continue. Andrew has been working on SAMGrid metrics with regard to the

reprocessing. He has written a small system for keeping track of job efficiency on the various remote sites and has worked on station improvements to improve efficiency overall.

1.5 Large Pick Events [DØ]

The late last year reconfiguration of fnal-cabsrv1 and fnal-cabsrv2 as well as the modifications to the *pick-events* script allows for large pick event jobs to run on any SAMGrid station without quickly turning over the central DØ cache located at FNAL. DØ, however, would like to protect part of the cache from the Pick Events activities. 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) is investigating alternatives and fixes.

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.
- Judy Nichols and Marc Mengel have implemented an Issue Tracker for SAMGrid operations. It has been deployed at DØ and will be deployed at CDF after some experience is gained and they have instituted a shift structure. The issue tracker allows issues to be submitted via the web and e-mail (through the d0sam-admin mail address, which users already know). Issue updates and resolutions are automatically sent to the submitter and shifters. Shifters and other experts can update issues via e-mail and the web.
- The "MIS" project is ongoing (Sinisa, Matt Leslie [CDF graduate student], Adam). The SAMGrid specific monitoring implementation via CORBA messages is about a month or two from completion (held up due to Adam's new management responsibilities). The Information Service is meant to reduce SAM station reliance on the central database. As Matt is leaving soon to graduate, Sinisia is working to make a prototype in short order.

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]

Though there is no specific project marked for SAMGrid for this driver, but clearly the "*SAMGrid: JS Deployment to production (CDF)*" project applies. Valeria and Gabriele have been working on installing SAMGrid at several CDF remote sites (problems at those sites beyond their control have been a

stumbling block) as well as implementing tools for merging. They have been incorporating suggestions from user feedback as well as testing and packaging the merging tools. Once the latter is accomplished, the users can try the system. The main risk here is 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 one of the major sources of activity in the SAMGrid group.

- Working with Jerry Guglielmo, Ashutosh, Amber, Krzysztof, Rick Snider & Doug Benjamin we have a hardware and deployment plan (see Jerry's talk at the next GDM).
 - The hardware plan details the location of the CORBA nameservice, DB servers, and web servers needed for SAMGrid operations and monitoring. The hardware has been deployed and installation and configuration are underway (to be finished by the end of this week).
 - A "CDF SAM Users Group" has been formed to advise on compatibility issues (to ensure that no old functionality is missed) and documentation issues.
- Work within the SAMGrid team is already underway on DB server fail over (Steve White), the database browser (Randolph), verification of integration with AC++ and the SAMGrid C++ API (Krzysztof, Valeria), and performance improvements in SAM v6 (Lauri, Robert).
- The SAM station can already interoperate with dCache in a SAM specific way. The need for prefetching will be assessed after the initial deployment.
- Krzysztof and Doug Benjamin (CDF SAM Liaison/Duke Postdoc) will perform load tests with the entire CAF during a dCache upgrade down day.

Risks are unforeseen problems may crop up in the full CAF load test. We also need much better documentation. Though the help of a technical writer is seen as necessary, the time scale for such work would be beyond the deployment schedule.

Jerry will detail milestones, schedule, and status in his talk at the next GDM.

1.10 Production Use of SAM on Reconstruction Farm [CDF]

A hardware plan has been worked out. The style of SAM station to be used is nearly identical to that used by the CAF. Unless there are station problems, the involvement of the SAMGrid team is limited to station installation and configuration advice.

1.11 Deployment of SAM [MINOS]

Art Kreymer is installing SAMGrid products and configuring components on MINOS systems for production. I know of no issues with this deployment.

2 Other Projects

These drivers/projects are not explicitly listed on the spreadsheet.

2.1 V6 Deployment at DØ

The v6 deployment at DØ must be worked into the experiment schedule. Test DB servers have already been deployed. In the past month, Sinisa and Paul Russo installed a v6 station on clued0 (the desktop cluster). Beta-testers are using the station and Sinisa is collecting feedback. So far he has easily fixed a few glitches that were found. Once the testing is satisfied, v6 will be deployed on one of the analysis clusters for production. If that works well, then the other analysis station will convert to v6. We hope to have this deployment finished on the order of two months [dictated by experiment schedule]. Risks are finding more people to do the testing.

2.2 Enth Distributed Dimensions [SBIR II]

(Driver is Grid interoperability) This project is in progress and is 2-3 weeks behind schedule, as some of the special SAM dimensions have been more difficult to implement than predicted.

2.3 Packaging and Configuration

(Driver is minimize operations and make deployments timely) Alan Sill has taken the "python freezing" project to near deployment stage. Alan can take a SAM python distribution and "freeze" it into one executable (instead of ~100 python scripts). The advantages of this system are

- Individual Python scripts do not have to be imported over NFS (import over NFS is very, very slow for DØ). Instead, only the one executable file needs to be accessed
- Deployment is simple - only a few files to deploy (CDF Reco Farm wants this)
- SAM operating under a different python version is simple (problem for DØ). The executable incorporates the version of Python that SAM uses and does not interfere with other versions

Alan is here at Fermilab this week and will interact with SAMGrid developers in order to make "freezing" a step in their deployment process.

2.4 Request System

Lauri is working on an improved "Request" system asked for by both CDF and DØ to track MC requests. Estimated completion time is 2 weeks.

2.5 Management Initiatives

Since becoming SAMGrid project manager, I have taken steps to improve management of the project and team. We are now working within a six week "assessment cycle" where we set goals for a six week time period, monitor those goals throughout the period, and make a final assessment at the end of the period. Such a system should improve planning and reduce the interruptions that affected performance previously. We also now have a weekly SAMGrid planning meeting, where the entire team meets to discuss the six week goals and status and here reports from team members on design ideas and implementations.

2.6 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; MISWeb enhancements (probably to be replaced by Randolph's DB Browser)

3 Main Project Milestones

- SAM on CDF CAF (see Jerry's talk at next GDM)
- SAM on CDF Reco Farm (see Jerry's talk at next GDM)
- SAMGrid v6 on DØ systems (estimate 2 months)
- SAMGrid LCG/OSG interoperability (estimate 1-2 months for prototypes)
- SAMGrid for CDF MC (estimate 1 month)
- Packaging/configuration Freezing (estimate 3 weeks)
- MC Request system (estimate 2 weeks)
- MIS server used for SAMGrid specific monitoring with SAM HDTV (estimate 1-2 months)

4 Effort

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

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

Operations support

- Art, Paul Russo, Liz Buckley-Geer

February Effort Reporting (9.75 FTE total from team)

- 4 FTE Core development
- 3 FTE Deployment to production
- 1.25 FTE Operations

- 1 FTE Project Management
- 0.5 FTE Outreach

March Effort Reporting (9.25 FTE total from team)

- 4 FTE Core development
- 2 FTE Deployment to production
- 1.25 FTE Operations
- 1.75 FTE Project Management
- 0.25 FTE Outreach

Now that there is a deployment plan, the Deployment to production fraction should increase in April.

5 Risks

The main risks, as I see them, are

- Issues with SAM DH brought up by CDF full CAF testing. Given previous testing and use at DØ, I believe that the chance of a big issue coming up is not large.
- 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.
- The vision of the future. The future of SAMGrid development beyond CDF/DØ/MINOS deployment and the DØ reprocessing is up in the air. Discussions need to start soon on how to proceed and what future forms the project will take.