SAMGrid Status Report

Adam Lyon, 10 January 2005 GDM

See the last section for implications of the DØ Grid Goals.

1 Project Drivers, Scope and Milestones

These are taken from the FY06 SAMGrid Budget Document. Updates are based on the December 2005 report.

1.1 Continuing Activities

1.1.1 Continue Smooth Operations

DB Server Improvements: Steve has fixed other thread safety issues in the DBServer. He has had no time to address speed issues.

1.1.2 Complete full deployment of SAM DH at CDF

- SAM on the farm is STILL using v6 (not the frozen v7). The CDF farm group needs to update their scripts and do testing. I understand they are manpower limited for this change, but it needs to be done.
- SAM is being integrated in their CAF restart: This is still proceeding slowly due to the DØ Refixing (see pertinent section below).
- CDF wants to be able to transfer a file out of SAM (e.g. a Root file) to any node running GridFTP. v5 SAMGrid already has a facility to do this that does not work in v7. Andrew has made updates to the v7 code and is in testing. A problem with the dCache gridFTP door was discovered – dCache experts are being consulted – UPDATE - solutions are being considered (a meeting is necessary).

1.1.3 DØ MC & SAMGrid & *REFIXING*

In late December, a calibration bug was discovered in the latest DØ fixing pass. Approximately 1.5B events need to be "refixed". DØ has started an intense project to refix the data quickly.

The first refixing pass is to refix critical skims needed for analyses going to the winter conferences. The reconstruction farm, analysis farm, a large farm at IN2P3, the CMS farm and some other offsite locations are participating in this first pass.

Running the refixing offsite required some changes to SAMGrid and mc_run_job (we had never run a fixing process with SAMGrid before - the logistics are somewhat different from MC and Reprocessing). Andrew and

Parag along with DØ mc_run_job experts worked over the holidays to get SAMGrid working. We have starting running refixing on the CMS farm (thanks to Ian for his technical support). We typically achieve processing of 600 files per day (~600 GB).

We are also submitting jobs to generic (non-DØ specific) LCG resources. We are currently running at Clermont-Ferrand, Prague, and NIKEF. We have achieved these submissions in the past few days and have processed about 200 files so far.

The next step is to submit to generic OSG sites. We believe this can work with some minor changes to SAMGrid, but it is still be evaluated. The lack of a broker on OSG means we will have to direct jobs by hand, but that should not be a problem. We already have an OSG forwarding node in place and it is currently being configured. We hope to try submitting to OSG sites by the end of this week.

The second phase of the refixing effort is to refix the entire 1.5B event unskimmed sample. This will require a large amount of resources (it is essentially all of the I/O of the reprocessing effort [especially the output] squeezed into 6 weeks). We are currently evaluating the infrastructure at Fermilab to support such an effort (do we have enough SAM stations to allow prestaging at remote sites? Can we handle the rate of files coming into FNAL for storage?). We expect to make heavy use of LCG and OSG resources (probably 10-20% of the refixing).

For this second phase, we expect to have others operate the refixing machinery, hopefully freeing up Andrew and Parag.

1.1.4 LCG & OSG Integration

See above for use of LCG and OSG for the refixing.

1.1.5 Integrate with VOMS/VOMRS

Parag has been working on the refixing effort.

1.1.6 SQLBuilder

Before the holidays we finalized the new Dimensions language syntax. It is fully backwards compatible with the old dimensions language and adds new wanted features such as child_of, parent_of, raw_of. Randolph is starting to implement.

1.1.7 DØ upgrade from v5 to v7

Much of this effort has been put on hold due to the DØ refixing effort. Dehong has ported the v5 online code to v7. The MC group is evaluating the new MC request system.

1.1.8 Rewrite broken groups and quotas for SAM managed cache

The human resources for this project have been retasked (Igor M.).

1.1.9 Deploy new SAM Data Handling Monitoring

A test version of "SamHDTV" is operational and working correctly. The MIS and mainline station codes were merged and final testing will begin before a test deployment. -- Still testing.

1.1.10 Testing

All of the DB Server unit tests have been wrapped in our SAM Test Harness, making testing new releases much easier. We also need to make specific tests on the client in order to test the affects of Python 2.4. These tests are being written.

1.2 Moving forward with new technology (new activities)

1.2.1 Integrate SAMGrid with v6/7 compatible Run Job

RunJob is still being worked on.

1.2.2 Investigate deployment of SAM redundant information services

This project is Sinisa's Information Service system. No work has been performed in the past month.

1.2.3 Investigate deployment of SAM web services

MINOS has been testing Sinisa's SAM web services prototype. They provided feedback and some bugs have been fixed. The SAM team right now does not have the resources to pursue large scale production testing at the moment.

1.2.4 Investigate use of Enth for data base queries (continuation of SBIR project)

No work has been done. Awaiting Matt's report. Still no report.

1.3 Providing new capabilities

1.3.1 SAM DH and Condor Glide in

Initial discussions are beginning to determine requirements. We will participate in the CDF offsite processing workshop.

1.3.2 SAM Edge Service prototype

The Wisconsin student supposedly has working scripts to do the on the fly deployment, but I have seen no report yet.

1.3.3 SAM usage of SRM capable storage elements

We have begun planning and design of the SRM and SAM DH interface. We had a very fruitful meeting with Timur and now understand the current and future capabilities of the dCache SRM interface.

1.3.4 Implementation of SRM interface around SAM managed cache

No work has been performed yet.

1.3.5 Investigate breakup of SAM data handling services

No serious work has been performed yet.

1.3.6 Investigate SAMGrid for Analysis

No work has been performed yet.

2 Effort

Fermilab CD effort is ~6.0 FTE (as of the December Effort Reporting - no January effort reporting yet)

- 100%: Andrew, Parag, Valeria, Steve Sherwood
- 50%: Randolph, Adam, Steve White, Robert, Krzysztof, Dehong
- 20%: Gabriele

Note that Valeria's guest scientist position ends at the end of December.

Breakdown of effort is below. Note that time off (vacation, sick, holiday) is not included, so the total effort will not match the available effort.

Effort	FTE
Core Development	2.5
Deployment to Production	1.4
Operational Support	1.0
Project Management	0.5
Outreach	0.1
Total	5.5

3 Risks

The risks are unchanged from November.

Some of the previous risks (unreasonable expectations and feature creep) are somewhat under control as we are now bringing related requests to the GDM instead of handling ourselves. A lesson that I'm learning is to always insist on use cases and requirements before any further consideration is made on a request.

Some new risks...

- Human resources: While I think we are in ok shape now to handle the projects we've started, we do not have the resources to start other important projects (Web services deployment, breaking up SAM into services). We are also undertaking some short term rapid projects to get some operational problems out of the way (speed up DB server). I am hoping that completing the CAF restart, monitoring, and DØ migration will free us up a bit.
    - The loss of Valeria will be deeply felt. She has been indispensable in CDF SAM operations and development of some CDF specific SAM components (e.g. their framework interface to SAM). She has also been a vocal supporter of SAM and often a voice of reason.

- Upgrade to Oracle 10g from 9i. It could be great - or a disaster. The speed of many SAM queries worsened when we switched from 8 to 9. We need a good testing program to find problems before 10g goes into production. MINOS has done some preliminary testing and found no problems - but their database is a small fraction of the size compared to CDF and DØ.

- Grid politics – I understand that protocols and interfaces may be changed (e.g. Condor using their own protocols) from what we use currently. SAMGrid will need to keep up.

4 Implications of DØ Goals and Requests

In general, we support the goals of the DØ computing model with regard to SAMGrid.

We are very supportive of the v7 upgrade plans. Unfortunately, the DØ refixing effort has caused us to delay the upgrade of SAMGrid to v7 code. The length of the delay is still in question. I would imagine that Andrew and Parag would be heavily involved in the Refixing effort until January 20th, after which the operations side could be turned over to other people. The refixing should not affect other plans for the v7 upgrade (except that I am busy with the refixing too).

The refixing effort is actually allowing us to meet other DØ goals early. We can now run fixing type jobs with SAMGrid (goal for March 06) on LCG. We will soon be able to submit jobs to OSG, albeit without the convenience of a smart broker.

If the refixing goes smoothly and SAM developers can end their involvement by the end of this month, then the schedule for the DØ goals with regards to SAMGrid would have only suffered a month delay.