

SAMGrid Status Report

Adam Lyon, 29 November 2005 GDM

1 Project Drivers, Scope and Milestones

These are taken from the FY06 SAMGrid Budget Document

1.1 Continuing Activities

1.1.1 Continue Smooth Operations

We continue to make bug fixes to SAM DH (data handling) and SAMGrid software as needed. The SAMGrid team rotates a week long primary and secondary on call expert shifter duty.

We are beginning to realize some extreme inefficiency exist in the SAM DB server. For example, a simple SAM metadata query takes less than one second for the database to fulfill, but takes 45 seconds for the information to be returned to the user. We are beginning to tackle these problems with a review of the marshalling system. Steve White has also found and fixed other extreme inefficiencies (e.g. checking dimension name validity whenever a query is performed even if the dimension query is in the DB and already passed a validity test).

1.1.2 Complete full deployment of SAM DH at CDF

- SAM DH is used by CDF for their CAF analysis farm in production (YAY!)
- SAM is being integrated in their CAF restart system (restarting a job segment when the worker node dies). Additional functionality has been added to SAM and CAF software. We are awaiting feedback from CDF.
- SAM DH access from Root has been completed. The CDF Framework component that handles SAM can be loaded into Root and accessed as a "*TSam*" object. This version of code, however, has not been released to the collaboration yet. Waiting on testing and passing of the entire release.
- 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.

1.1.3 DØ MC & SAMGrid

Full information may be found at http://www-d0.fnal.gov/computing/grid/mc_todo.html . Much work has been done to

integrate JIM and the DØ MC process, including adding monitoring for the MC group.

1.1.4 LCG & OSG Integration

Most of the work is on the LCG side. See <http://www-d0.fnal.gov/computing/grid/SAMGridLCGStatus.html> for the project status. A 400 node test (spread out among many sites) was performed. Job efficiency (fraction of started jobs that completed) was about 70%, but that was mostly due to one site that was throwing away jobs for a period of time. That site has been fixed. The job start up time was slow (on the order of an hour) due to moving data from the SAM DH station to the site. Work is being done now to speed up this transfer by making use of LCG storage elements. Also, new nodes are being deployed at sites so that they can have one production forwarding node and a development node for testing.

1.1.5 Integrate with VOMS/VOMRS

A design document on the sam-design mailing list is out describing the interaction between the SAM database and VOMRS. Awaiting comments and Parag's return from his wedding/honeymoon.

1.1.6 SQLBuilder

Design decisions for the SQLBuilder are finished and integration of the SQLBuilder with Corba are nearly complete.

1.1.7 Upgrade from v5 to v7

CDF is already at SAM v7. DØ analysis farms are running v7 and most collaborators use the v7 client. But there are still many DØ tools that rely on v5 SAM.

- All of the MC scripts and MCRunJob work within v5. Progress is being made on an MCRunJob in v7 given that we now have a version of SAM that runs in the latest version of Python.
- MC and the Farms are still using v5 request systems (the Farm request system hasn't even been maintained). They need to upgrade to the new v7 Request system. The migration involves a database migration (they need to use different tables). Coordination is needed between SAM and the experiment.
- Online is using v5 SAM. Some special components of the SAM client used by online were never ported to v7 (due to a former developer's objections). We will be meeting with the DØ online folks to figure out what to do.

Gavin Davies and Laurent Duflot are heading up this effort from the DØ end.

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. We need to merge the "MIS" branch of the SAM code with the mainline branch (this week). Deployment can then begin.

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 are now developing integration testing in the TH.

1.2 Moving forward with new technology (new activities)

1.2.1 Integrate SAMGrid with v6/7 compatible Run Job

Working to remove the application particulars of a job out of SAMGrid and put into RunJob. One problem has been Python versions. SAM uses an old version of Python (2.1) lacking many advanced features. We have recently tested our packaging ("freezing") system with the latest version (2.4) with much success. We are going to start deploying SAM software that uses the latest Python.

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.

1.3 Providing new capabilities

1.3.1 SAM DH and Condor Glide in

Initial discussions are beginning to determine requirements.

1.3.2 SAM Edge Service prototype

A student at U. Wisconsin is working on deploying SAM stations "on the fly" at remote sites. He has been experimenting with the CDF "init_sam" script. He is meeting with us shortly to go over plans.

1.3.3 SAM usage of SRM capable storage elements

We have begun planning and design of the SRM and SAM DH 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 November Effort Reporting)

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

In August, our effort was 9.5 FTE.

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

3 Risks

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.