

SAM Project Revised Organization

17 September 2003

1. Current Status

The SAM project has been operating with two teams, Core and Grid, with the project management focussing on the Core team, and the Grid team having a separate manager, Igor Terekhov. The Core team has had full responsibility for continuing production operations in the D0 systems (central and remote) and for attempting to integrate CDF. The Grid team has worked on preparing a product to permit remote job submission and monitoring.

Let us describe the current status. The Core team has worked with the D0 experiment to bring a set of experiment shifters fully into the operations model for SAM. This model is now working rather well. The documentation and support for users is at an acceptable level; the installation of remote SAM stations in 'friendly' configurations for D0 proceeds in a reasonable amount of working time (elapsed real time is of course determined by other factors related to time zone differences and priority at either end given to the install). The performance of the SAM stations on the central D0 systems has supported the level of use through two major conference cycles, Winter 03 and Summer 03, without breaking down. Two long-awaited improvements to the core SAM software, a more general batch interface and a more performant user API, have been designed, coded, and deployed to production. The dcache copy has been added to the samcp options, although not deployed in production anywhere yet. Other changes to permit operation with HPSS cache have been added and deployed at Lyon.

Integration of CDF into the project has had major obstacles to overcome, in that the experiment's ways of working had evolved with a very different system. The transition to a different system, more metadata-driven, is just not so easy, and requires full commitment from the experiment's offline software management. The first prerequisite for merging, the agreement on a common schema for the event/file catalog, has been achieved. A stop-gap version of the SAM dbserver to function with the new schema is available for testing. A more complete adaptation of the dbserver to the new schema, making full use of the changes in the data_files table, is being coded now and is expected to go into testing, modulo other calls on the team's time, in 1-2 months. This adaptation will also of course include the migration to OmniORB and the cleanup and redesign of the IDL interfaces. Operational experience from the previous few years has been fully utilized in this adaptation, with the aim of making any future changes much easier to implement. From the CDF side, there are very encouraging signs of real integration and deployment becoming possible. The project to switch online recording from the old DFC to the SAM catalog is complete. The experiment has

agreed to drop the requirement of file deletion, a concept that would be very hard to implement in the SAM context. The effort from the CDF side is no longer interrupted by serious breakdowns in their current system. There is assignment of CDF effort to testing and configuration management tasks, and serious testing of the SAM dcache deployment is beginning now. The Bird review recommendation is that the CDF transition to SAM should proceed aggressively.

The Grid team has deployed version 1.0 of its job submission and monitoring product JIM for testing at multiple sites. It currently permits use of various job submission mechanisms: runjob, sam submit, CDF CAF. JIM uses Condor and Globus products from the standard Grid middleware world, and has well-established contact with the Condor team at Wisconsin.

2. Next steps

Obviously, there is still major work on hand to finish the dserver project, complete the testing for CDF deployment and go into production for CDF, and to deploy JIM software in its first production task, using it for simulation job submission for D0. There is also the continuing commitment to keep systems operational and functional for D0 (and CDF when the time comes). Nevertheless, we can now look ahead to the next big development issues.

The next major functional developments needed for the project as a whole are, not ordered in any specific way:

1. Revisit the caching requirements for data handling and plan how SAM should adapt to meet those requirements
2. Revisit the configuration requirements for SAM and begin to generalize them. This has two subprojects:
 - a. Adapt to full Linux operation (requires introducing new configurations for running the SAM servers, possibly new communication methods among servers, failover strategies,...)
 - b. Adapt to a wider range of installation constraints (consider the impact at remote sites of installation privileges, products used, etc., and work toward reducing this impact)
3. Improve monitoring of the system's functioning, locally and remotely. The necessity of doing this is very evident in trying to make it possible for the shifters to diagnose problems in the current operation.

Each of these development areas can be considered as Grid-related. Just as the dserver adaptation is driven both by the new functional requirements and also by the desire to improve maintainability, so should these larger development plans also be driven by the additional constraint of modularizing the system along lines that make it possible both to incorporate products of other Grid-like data handling

systems and to offer its own products to provide well-defined functionality to other projects.

III. Reorganization for the next phase

How should the project proceed with these new major developments? There are two guiding principles for them – first, to keep the production systems operating well at all times, and to make sure the new functionalities come in properly phased, without large disruptions but also without long delays between upgrades; and second, to make sure that we are good Grid citizens, participating actively rather than passively in defining data handling requirements and interfaces that meet real HEP needs, and also in testing and providing Grid middleware. Keeping both these principles firmly in mind is best served by returning the project to a single team, committed to both principles, and moving forward on both fronts together. So the proposal for SAM project management is to have the co-leaders reporting to the experiments and the Division as in the current management document, but to have one project team, called the SAMGrid team. Its task list for development would be organized under the 3 major subheads above (and would of course still contain all the support and operations tasks in the currently two separated task lists as well).

Proposed organization table:

Project co-leaders: (set priorities, define requirements, evaluate progress)

Wyatt Merritt CD/DØCA

Rick St. Denis U Glasgow/CDF

Project technical leads: (provide technical designs to requirements, organize schedules for implementation and deployment)

Sinisa Veseli CD/DØCA

Rob Kennedy CD/CDFCA

It is proposed to have a unified weekly operations meeting, for which the agenda is produced by the technical leads. A daily touch-base meeting is also proposed. The weekly meeting would include operational communication with external team collaborators (Condor, GridPP, ...). There would also be a weekly design meeting, for which the agenda is produced by the project leaders. Working group meetings on specific design implementations would report back to the design meeting.

Communication with the outside grid community is important to satisfying the second of the two guiding principles, and cannot be limited to just the groups we are directly collaborating with. We will handle this communication by raising general grid issues in the design meetings, and by giving individual team members rotating assignments to attend grid meetings or research particular grid products or strategies. We have to be careful, however, to limit the impact of the

flood of grid-related activities since we have a long detailed task list and limited personnel. It will be a priority of the project management to keep this delicate balance, and in particular to maintain close communication with US CMS grid efforts, which we hope will lead in appropriate areas to establishing direct collaboration on specific subprojects.