SAM Infrastructure

Robert Illingworth

# Introduction

In order for SAM to function for the next 10 years, some infrastructure changes and improvements will be necessary. These changes are discussed in this project plan.

Topic: Figure out what to do with CORBA. Can we remove CORBA from the user facing pieces? Would we want to exorcise CORBA entirely? Is that feasible.

Topic: Correct limitations and add features to the SAM station.

* Original design was to deliver files to a single node. This does not match reality. Many of these enhancements were made for SAMGrid and are not available to normal jobs. Goal: Make these changes usable outside of SAMGrid (e.g. there is no way to determine what server a file is coming from – fine if local storage). – already in progress
* Physically accessing a file and tracking the processing of the file is linked together. (e.g. right now if you are processing the file, SAM assumes you are accessing it). For the case where the file is left on a server and is copied by another system, then the file on the server can be removed during processing. Also, you want to split the monitoring and track the transfer time of the file and not have the processing time included. – already in progress (method in the SAMWeb interface exists, but itis dummied out)

Topic: Dealing with shared storage like Bluearc and NFS. The original design has one stager node per storage disk. This does not work for a large “disk” like the Bluarc. Also want redundancy so that if one stager goes down, SAM can use other still up stagers.

Topic: Add file tracing to SAM for diagnostics.

Topic: Implement a Web Server in the station itself to implement web methods and to inquire the internal state of the station.

Topic: Split the monolithic SAM product into different bits (e.g. Sam station, stager, fss, etc are all in the same package) – already in progress

Topic: Consider dCache with parallel NFS. Probably similar to Bluearc except that you have to do special stuff to stage files in and don’t have to delete anything (since dCache does that).

Topic: Do we need SSL support for file transfers and communication?

Topic: Move some of the station information out of the Oracle DB and into local DB (e.g. no reason that Oracle needs to track every cached file).

Topic: Convert all code to python 2.7 (the client already runs under 2.7). Do not consider python 3 for now.

Topic: Use kerberized access to Oracle. The thing that holds this up is our administrator tools.

Topic: Move away from Oracle? More difficult than getting rid of Corba. Oracle performance is understood.

# Tasks

## Complete splitting the station, enhanced get next file, and notifying transfers

Enhanced get next file and notifying transfers required CORBA changed, so these were done together. Consider backwards compatibility. People: Robert. Duration 4 weeks. Low priority task already in progress.

## Embedding a Web server in the SAM station

Investigate tools and libraries. Decide functionality of the web server. Implement and test. People: Robert. Duration 1 month, Low priority task. Research already started.

## Implement and test Shared Storage solution

Do we want to use Bluearc as cache? Stand alone? Right now plan is to use Bluearc. Dcache 4? Need to fix the one delivery location problem. People: Robert. Duration 1 month. Low priority task. Start: summer

## Encrypted SSL connection to DB

Anil was pushing this. We do care, because the current system is not secure. Research options and possible solutions. Needs a design meeting. People: Robert, Marc. Duration: 1 month, Starting summer

## Migrate to OmniOrb Corba Name service

Right now we use the Orbacus name server, which is unsupported. Migrate to OmniOrb’s name service. People: Robert. Duration: 2 weeks. Start: If needed.

# Risk

Experiments in the future asking for new features or enhancement.

# Conclusion

These topics and topics are necessary to keep SAM running for another decade or more.