

SAM @ IF

Adam Lyon

CD/REX

NUCOMP 2011-09-14

What is SAM (Reminder)

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

Delivers files (tape/cache) to jobs or you

Stores files to tape or other locations

Manages all this with some sanity

Used by CDF, DØ, and MINOS

Focusing on Minerva and NOvA

[See May 2011 NuCOMP]

Improving the interface to SAM

Currently have C++ and Python APIs and a command line interface

SAM commands are unfamiliar and complex

TWO IMPROVEMENTS:

samweb – an HTTP based system for simple integration with experiment's framework

samfs – A system using FUSE for easy interactive access (makes SAM look like a file system)

samweb

CD/REX has started on the *samweb* project

A lightweight HTTP interface to a subset of SAM functionality needed to run data handling for jobs

Goal: Make the interface between the experiment framework and SAM simpler – actions are possible with libCurl or similar HTTP library

Project lead is Robert Illingworth

Participants are Marc Mengel, Mike Kirby and others

Stakeholders are you and the Art team

samweb project

See <https://cdcv.s.fnl.gov/redmine/projects/sam-web/wiki>

SAM Web services

	2011-8					2011-9					2011-10					2011-11				
	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47			
Feature 1668: Design file retrieval interfaces	<div><div></div> Resolved 100%</div>																			
Idea 1672: Investigate Gaudi Integration	<div><div></div> Assigned 0%</div>																			
Feature 1664: Implement base web server	<div><div></div> Resolved 100%</div>																			
Feature 1665: Implement test framework	<div><div></div> New 70%</div>																			
Feature 1670: Implement project interface	<div><div></div> Assigned 80%</div>																			
Feature 1669: Implement locate files	<div><div></div> Assigned 100%</div>																			
Feature 1673: Integrate with Gaudi	<div><div></div> New 0%</div>																			
Milestone 1671: Deploy system	<div><div></div> New 0%</div>																			

Implementing SAM commands:

- locate
- start project
- establish consumer
- get next file
- release file
- commit process
- end project

Information returned as:

human readable text

JSON

XML (perhaps)

SAM locate demo:

sam locate foo

sam locate MN 000...root

Interfacing to *samweb*

Minerva uses the Gaudi (LHCb) Framework

Little in-house expertise

File stream handling in Gaudi is quite opaque

Decision: Use the job wrapper script to communicate with SAM and pass file information into Gaudi configuration file

Impact: An instance of Gaudi runs over one file (wrapper script can loop over files by starting Gaudi multiple times)

**Output file meta-data handling & recovery mechanism:
Needs to be determined**

Interfacing to *samweb*

NOvA (and mu2e, g-2) uses ART

We have had extensive discussions with the ART team for integrating *samweb* into ART. They will write a SAMDH plugin to do data handling from within ART.

Meta-data will be stored in the output file itself (!!!)

Recovery dataset generation is trivial (we will provide a script to run over collection of output files)

Dealing with small files

Tape storage demands large files (e.g. > 1 GB)

**The file limit has been a problem for IF experiments.
Typically, files are *tarred* before uploading to tape**

Retrieval of the individual files is difficult.

After *samweb*, will start a project for transparent handling of tar files. SAM will,

- Given a file name, determine the tar file where it resides**

- Retrieve that tar file (*.tgz*)**

- Unpack the tar file in a cache area**

- Deliver the desired file to the job/application**

- Smartly handle subsequent requests for other files
in the same *.tgz***

Project still needs to be written, hope to deliver early 2012

samfs

Filesystem like access for SAM

Start project in early 2012

Summary

Three projects:

***samweb* for easy interfacing with experiment framework**
Project is underway, deliver 10/2011

Transparent tar file handling
Project to start late 2011, deliver 2/2012 (estimated)

***samfs* for easy interactive access to files in SAM**
Project to start early 2012, deliver mid 2012