

SAM Project Status

Robert Illingworth

SCD/REX

14 March 2012

Topics

- Adding files
 - File Transfer Service (FTS)
- Getting them back
 - SAM projects
 - SAM ART integration
 - IF data handling utility layer (IFDH)

Uploading files

- File Transfer Service originally designed for uploading NOvA raw data
 - Intended to be flexible enough for other purposes and experiments
- Features
 - Adding metadata to SAM database
 - Copying files from initial location to BlueArc/Enstore etc
 - Tar up small files for tape archiving (expect to replace this with Enstore file aggregation)
 - Erase initial file only after it's satisfied that it has been properly archived
 - Forward files from remote location (like Ash River)

FTS status web page

FTS status for novadaq-ctrl-datadisk-02.fnal.gov

Generated at 2012-03-14 10:51:01 CDT ([refresh](#))

Summary

FTS: OK	FSS: OK	Stager: OK
---------	---------	------------

Completed: 58170	Error: 97	Pending: 185
------------------	-----------	--------------

- Recent completed transfers

Time	File name	Destination
raw		
2012-03-14 09:35:18 CDT	ndos_r00013574_s10_t02.raw	novadata:/nova/data/rawdata/NDOS/000135/13574/02
2012-03-14 09:35:17 CDT	ndos_r00013574_s10.raw	novadata:/nova/data/rawdata/NDOS/000135/13574/all
2012-03-14 09:05:17 CDT	ndos_r00013574_s09_t02.raw	novadata:/nova/data/rawdata/NDOS/000135/13574/02
2012-03-14 09:05:16 CDT	ndos_r00013574_s09.raw	novadata:/nova/data/rawdata/NDOS/000135/13574/all
2012-03-14 08:05:17 CDT	ndos_r00013574_s08_t02.raw	novadata:/nova/data/rawdata/NDOS/000135/13574/02
merged-log		
2012-03-13 19:04:01 CDT	log_merge_ids_3569323_3569550_1331675622.tar	/pnfs/nova/archived_logs/2012/03
merged-raw		
2012-03-13 07:03:31 CDT	raw_merge_ids_3569268_3569489_1331626492.tar	/pnfs/nova/rawdata/NDOS/2012/03

+Errors (97 hidden)

+Pending (185 hidden)

Beyond NOvA online

- FTS is now running for NDOS raw data and log files
- Although it was originally requested by NOvA online, it is designed to be flexible
- Should be usable for anyone with a SAM database
 - Will need to implement a metadata extractor for new file types

SAM file retrieval

- Advantages of running jobs with SAM
 - Datasets defined by metadata
 - No need to explicitly give file locations
 - Easier to run outside FermiGrid (D0, CDF, OSG remote sites)
 - Track progress of jobs
 - Automatically distribute files across jobs processing the same dataset

Defining datasets

MINERvA Demo Dataset Definition Editor

Selection Criteria

Data Set:
Group: minerva

Data Tier:

Version:

Time Range: to
(format: 2011-05-09 or 2011-05-09T23:46:04)

Run Number: Subrun %

Quantity:

Operators

Data Set Definition (Dimensions query):

(you may also edit this query string directly to add custom fields to your query)

Save Definition

Name:

User:

Group: minerva

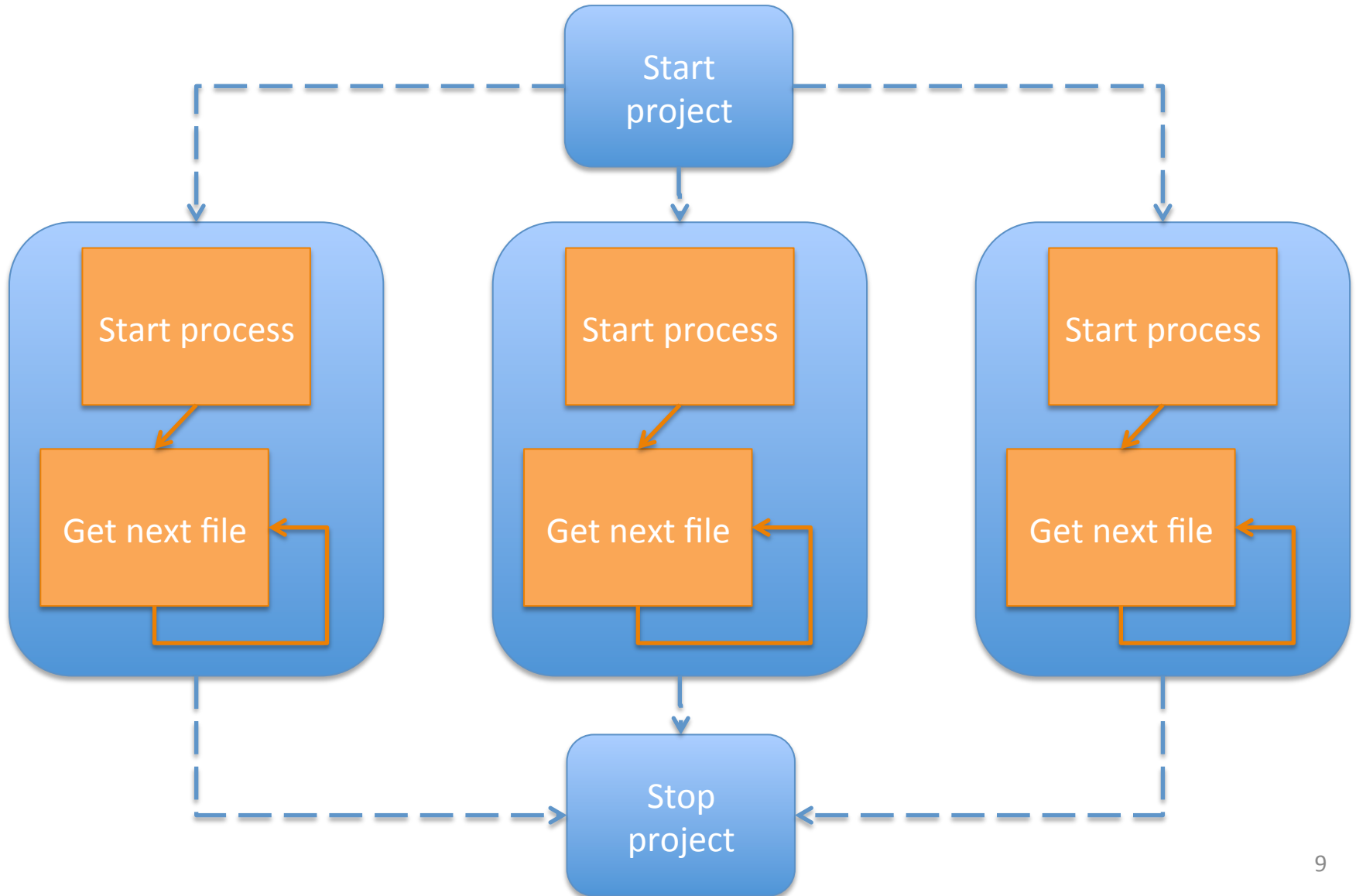
Datasets can have an arbitrary name but should not include spaces or special characters (underscores and dashes are permitted)

- Web editor for defining datasets
- Customized by experiment so can handle experiment specific metadata

Processing with SAM

- Start task (a SAM project) to retrieve a given dataset
- Job connects to project; requests a new file
- SAM returns file location
- Job does its work; releases file back to SAM (marks file as processed in database)
- Job requests new file; repeats until no files left

Parallel SAM jobs



SAM Web Service

- Original (Run II) SAM has dedicated shell/python and C++ client
 - C++ obtrusive to integrate into existing framework
 - Python requires dragging sizable client code around
- For IF we have developed a service that only requires something that can do HTTP requests
 - <https://cdcvcs.fnal.gov/redmine/projects/sam-web/wiki>

ART integration

- Discussions with ART developers about integrating SAM with ART framework
- Two main topics
 - SAM file retrieval
 - Metadata

ART file retrieval

- Implementing SAM projects within ART should be relatively easy using SAM Web interface
- Initial set up will be responsibility of job wrapper scripts
- ART will handle get next file/release file loop
- Since SAM only provides the incoming file location, an extra step is needed to get file to the local machine (see the next part of this talk)

ART metadata

- We want to ensure that ART output files contain all relevant metadata about the file
- The idea is that files should be “self-describing”; there should be no need for auxiliary data that you can lose
- Importing metadata into SAM requires you to run a simple program on the file that pulls out the relevant metadata and converts it to the format SAM needs (this step can be handled by the FTS)

IF Data Handling Utility Layer

- What is it?
 - Packaging: API/command-line
 - One stop Shopping for Data Handling
 - Small, Fast
- Packaging
 - Simple, small C++ api
 - Command line utility
 - Python library

IFDH functionality

- File transfer
 - Use cpn, srm,... as appropriate
- Logging
 - UDP syslog based, for reporting what a job is doing
- SAM web client
 - Query SAM metadata
 - Run SAM projects as just described

IFDH Packaging

- No complicated external dependencies
- Small enough to be easily distributable
 - Command line executable (< 300 kB)
 - C++ shared library (<300 kB)
 - Python module (750 kB)

Conclusions

- Work is on going to make SAM more effective for IF experiments
- New components (FTS, SAM Web, IFDH) becoming available
- Need users to test them out and get them into real use