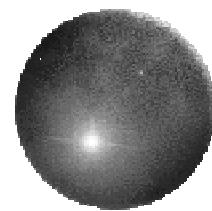




Gabriele Garzoglio for
Armando Fella, Stefano Belforte and
Donatella Lucchesi

SAM-Grid durable files for CDF: a status





What's on the plate

1. Usage of durable files for during processing
 - Belforte and Lucchesi skimming and merging efforts
 2. Usage of durable files after processing
 - sam_upload tools (Armando Fella)
-
- Both mechanisms are inspired to durable file management à la DZero MC via JIM.
 - The future direction is making `sam_upload` more generic, so that it can be used for processing AND post-processing



1) Skimming and merging

- The idea
 - Produce skimmed files running on the CAF
 - Store the files in a SAM durable area
 - Run a merging job at the durable area
 - Store permanently the results
 - Cleanup
- The challenges
 - Bringing up the infrastructure and putting together a set of client scripts
 - Doing metadata right (lineage etc.)
 - Recovery datasets
 - Scalability issues

2) *sam_upload*

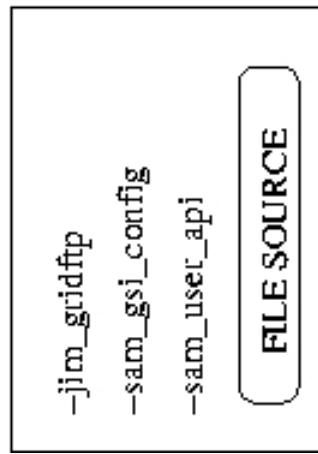
- The idea

- Files of interest to the community (e.g. mc)
are on a disk AND not in SAM
- Store the files in a durable SAM location
- Validate/process them using SAM
- Decide which ones to delete and which
ones to store permanently
- Clean up

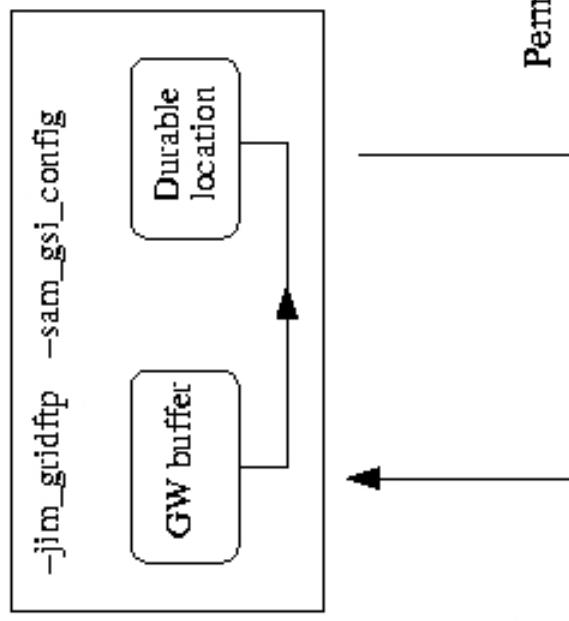


2) Architecture

Sam Client



Sam Station (Stager)



Durable Upload

Permanent Upload



2) *Status of the tool*

- Available in kits as sam_upload
- Depends on other tools (sam_gsi_config, jim_gridftp, CDFSamStore)
- Needs tailoring to work (should be simplified)
- Needs an infrastructure to work (durable area, gridftpd at the durable area, sam station, etc.)
- Reasonably good instructions:
http://www.pi.infn.it/~fella/sam_upload/index.htm
- Needs beta testers

Possible future directions

- Make sam_upload more portable
- Change Belforte/Luchesi's scripts to use sam_upload
- Strengthen metadata for durable files
- Automate rescue dataset creation/error recovery mechanisms
- Strengthen the mechanism
- Strengthen the mechanism
- Spread the word