

GLAST Large Area Telescope

File Versions in L1

Warren Focke

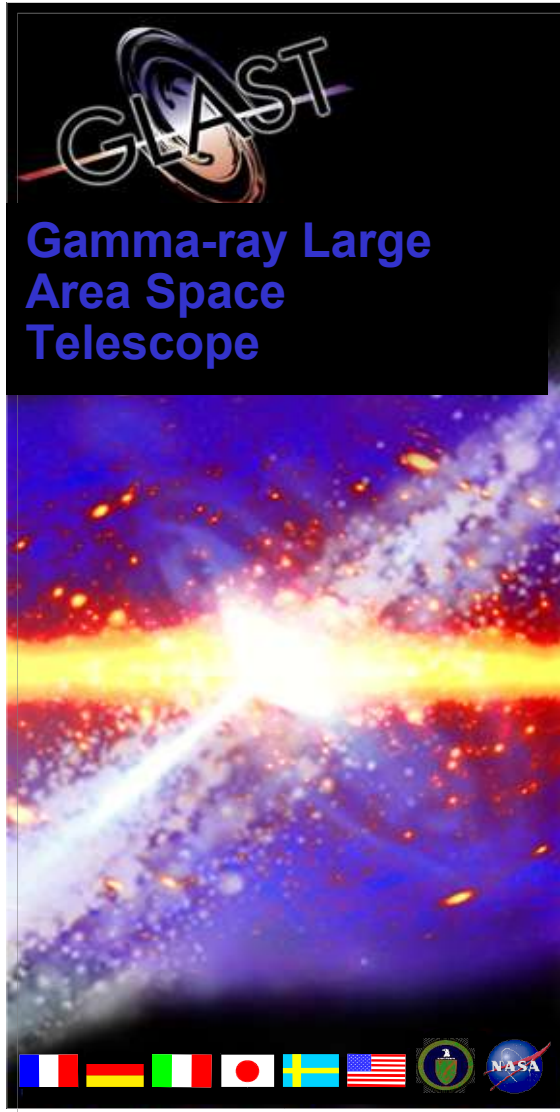
SLAC

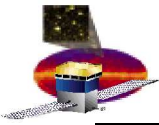
I&T Science Verification Analysis and Calibration

Engineering Physicist

focke@slac.stanford.edu

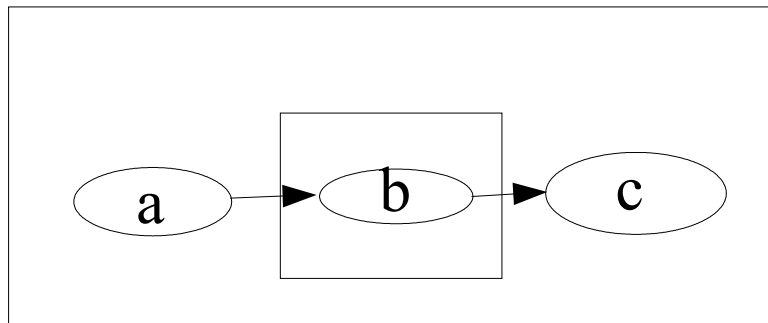
650-926-4713



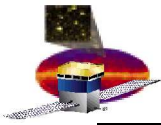


The Problem

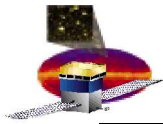
- **L1 processes that read a file need to know its version**
 - even if the upstream process that produced it was rolled back multiple times
 - and previous file versions were produced by a different task version
- **Currently only have versions on run-level files**
 - implementation is grotty (lots of tiny text files)
 - might be able to do better with data catalog's help



- b needs to know versions from a
- c needs to know versions from a and b
- the run may have been processed by previous task versions, file versions must carry on



- **Versions on chunks and crumbs would be nice**
 - use **xroot** for buffering
 - use **CEs** instead of merging
- **Lots of chunks & crumbs**
 - **more load on whatever keeps the versions**
- **Tony and Dan have been reluctant to allow data cat access from batch**
- **904 cores limits number of concurrent jobs**
- **Calibrations DB handles a similar load**
- **So maybe this could go in a separate set of tables?**



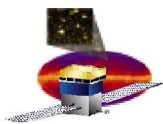
By the book

runs				
run_pk	run_id	recon_version	digi_version	ft1_version

chunks				
chunk_pk	chunk_id	run_fk	recon_version	digi_version

crumbs			
crumb_pk	crumb_id	chunk_fk	recon_version

- **select crumbs.recon_version from crumbs, chunks, runs where runs.run_id = runid and chunks.run_fk = runs.run_pk and chunks.chunk_id = chunkid and crumbs.chunk_fk = chunks.chunk_pk and crumbs.crumb_id = crumbid**
- **with runkey as (select run_pk from runs where run_id = runid), chunkkey as (select chunk_pk from chunks where run_fk = runkey and chunk_id = chunkid) select recon_version from crumbs where chunk_fk = chunkkey and crumb_id = crumbid**
- **select chunks.recon_version from chunks inner join runs on chunks.run_fk = runs.run_pk where runs.run_id = runid and chunks.chunk_id = chunkid**



Faster?

runs				
run_pk	run_id	recon_version	digi_version	ft1_version

chunks				
chunk_pk	run_id	chunk_id	recon_version	digi_version

crumbs				
crumb_pk	run_id	chunk_id	crumb_id	recon_version

- **No joins**