

P202 - Full Pass7 Reprocessing

status: **Complete**

last update: 07 Nov 2013

This page is a record of the configuration and execution of the P202 reprocessing project, full reprocessing from DIGIs using Pass7 analysis code. This project involves reprocessing with [Pass7 classification trees](#) and up-to-date alignment/calibration data. This task will read DIGI files and emit RECON, MERIT, GCR and CAL ROOT files, and the standard array of FITS files. It will be a CPU-intensive and storage-intensive enterprise requiring months of elapsed time and of order 0.7 Pbyte of storage. At the time of this task beginning, there will be about 20,000 science runs in Fermi (3.5 years accumulation).

To avoid occupying a new 0.7 PB of disk space, the plan is to remove old RECON files once they have been reprocessed. This is a shell game that involves some amount of buffer space and then waiting until the new RECON file has been created and (to some extent) validated before removal. The old RECON files will be retained on tape in the HPSS system and they will be available via xroot (but with some delay as these large files are staged in). In addition, old CAL files will be removed from disk **without** being stored on tape.

The name "P202" derives from the word "processing" and the **initial** file version to be used for the output data products, e.g., r0123456789_v202_merit.root.

"New generation" tasks (using SCons builds, rewritten task scripts, common python scripts, etc.)

- [P202-ROOT](#) - This [task](#) reads DIGI and produces reprocessed RECON + CAL + GCR + MERIT + FILTEREDMERIT (photons) + ELECTRONMERIT
- [P202-FITS](#) - This [task](#) reads MERIT and produces FT1 (photons) + EXTENDEDFT1 + LS1 (merit-like FITS file for photons) + EXTENDEDLS1 + ELECTRONFITS file
- [P202-LEO-ROOT](#) - This [task](#) performs the same function as P202-ROOT, but with the 200 L&EO data taken summer 2008.

Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v202.

XROOT location and file naming

Location template:

```
/glast/Data/Flight/Reprocess/<reprocessName>/<dataType>
```

Locations for P202:

```
/glast/Data/Flight/Reprocess/P202/recon
/glast/Data/Flight/Reprocess/P202/cal
/glast/Data/Flight/Reprocess/P202/gcr
/glast/Data/Flight/Reprocess/P202/merit
/glast/Data/Flight/Reprocess/P202/filteredmerit
/glast/Data/Flight/Reprocess/P202/electronmerit
/glast/Data/Flight/Reprocess/P202/ft1
/glast/Data/Flight/Reprocess/P202/extendedft1
/glast/Data/Flight/Reprocess/P202/electronft1
/glast/Data/Flight/Reprocess/P202/ls1
/glast/Data/Flight/Reprocess/P202/extendedls1
```

File naming:

Data Type	aka	Send to FSSC	Naming template
RECON		No	r<run#>_<version>_<dataType>.root
CAL		No	r<run#>_<version>_<dataType>.root
GCR		No	r<run#>_<version>_<dataType>.root
MERIT		No	r<run#>_<version>_<dataType>.root
FILTEREDMERIT		No	r<run#>_<version>_<dataType>.root
ELECTRONMERIT		No	r<run#>_<version>_<dataType>.root
ELECTRONFT1		No	gll_el_p<procVer>_r<run#>_<version>.fit
EXTENDEDFT1		No	gll_xp_p<procVer>_r<run#>_<version>.fit

FT1	LS-002	Yes	gll_ph_p<procVer>_r<run#>_<version>.fit
EXTENDEDLS1		No	gll_xe_p<procVer>_r<run#>_<version>.fit
LS1	LS-001	Yes	gll_ev_p<procVer>_r<run#>_<version>.fit

Note: 'procVer' is a field added to the file name (and the keyword "PROC_VER" in the primary header) added to the FFD 5/12/2010. Ref: http://fermi.gsfc.nasa.gov/ssc/dev/current_documents/Science_DP_FFD_RevA.pdf

Examples:

```
/glast/Data/Flight/Reprocess/P200/recon/r0239557414_v202_recon.root
/glast/Data/Flight/Reprocess/P200/cal/r0239557414_v202_cal.root
/glast/Data/Flight/Reprocess/P200/gcr/r0239557414_v202_gcr.root
/glast/Data/Flight/Reprocess/P200/merit/r0239557414_v202_merit.root
/glast/Data/Flight/Reprocess/P200/filteredmerit/r0239557414_v202_filteredmerit.root
/glast/Data/Flight/Reprocess/P200/electronmerit/r0239557414_v202_electronmerit.root
/glast/Data/Flight/Reprocess/P200/extendedft1/gll_xp_p202_r0239559565_v202.fit
/glast/Data/Flight/Reprocess/P200/ft1/gll_ph_p202_r0239559565_v202.fit
/glast/Data/Flight/Reprocess/P200/electronft1/gll_el_p202_r0239559565_v202.fit
/glast/Data/Flight/Reprocess/P200/extendedls1/gll_xe_p202_r0239559565_v202.fit
/glast/Data/Flight/Reprocess/P200/ls1/gll_ev_p202_r0239559565_v202.fit
```

DataCatalog location and naming

Logical directory and group template:

```
Data/Flight/Reprocess/<reprocessName>:<dataType>
```

Note that the <dataType> field (following the colon) is a DataCatalog 'group' name, and file names are of the form r<run#>.

Naming examples:

```
Data/Flight/Reprocess/P202:RECON r0239557414
Data/Flight/Reprocess/P202:CAL r0239557414
Data/Flight/Reprocess/P202:GCR r0239557414
Data/Flight/Reprocess/P202:MERIT r0239557414
Data/Flight/Reprocess/P202:FILTEREDMERIT r0239557414
Data/Flight/Reprocess/P202:EXTENDEDFT1 r0239557414
Data/Flight/Reprocess/P202:FT1 r0239557414
Data/Flight/Reprocess/P202:ELECTRONFT1 r0239557414
Data/Flight/Reprocess/P202:EXTENDEDLS1 r0239557414
Data/Flight/Reprocess/P202:LS1 r0239557414
```

Data Sample

The currently defined data sample (as of May 2012) for P202 reprocessing includes:

First run	239557414 (MET), 2008-08-04 15:43:34 (UTC)	
Last run	354923690 (MET), 2012-03-31 21:54:48 (UTC)	
Total runs	20,229	
Total input DIGI events	44,125,679,961	
Total RECON events	44,125,679,961	
Total CAL events	44,125,679,961	
Total GCR events	44,125,679,961	
Total MERIT events	44,125,679,961	all "events"
Total FILTEREDMERIT events	6,291,424,926	selected photon event classes
Total ELECTRONMERIT events	90,904,582	all electron events

Generation of FITS files is a second step in the reprocessing and has only been run on the first year of data. Stay tuned...

Total EXTENDEDFT1/LS1 events	6,291,424,926	selected photon event classes
Total LS1 (FSSC selection) events	1,325,204,821	event classes (bits) 0,2,3,4 (transient, source, clean, ultraclean)
Total FT1 (FSSC selection) events	189,323,074	event classes (bits) 2,3,4 (source, clean, ultraclean)
Total disk space used	762.4 TB	
Total effective disk footprint	43.7 TB	after removal of old RECON and CAL files

NOTE: One run, 242429468, of type TrigTest was declared 'good for science' and has been included.

Bookkeeping

1. (This page): Define ingredients of reprocessing (processing code/configuration changes)
2. Processing History database: <http://glast-ground.slac.stanford.edu/HistoryProcessing/HProcessingRuns.jsp?processingname=P202>
 - a. List of all reprocessings
 - b. List of all data runs reprocessed
 - c. Pointers to all input data files (-> dataCatalog)
 - d. Pointers to associated task processes (-> Pipeline II status)
3. Data Catalog database: <http://glast-ground.slac.stanford.edu/DataCatalog/folder.jsp>
 - a. Lists of and pointers to all output data files
 - b. Meta data associated with each output data product

P202-ROOT

Status chronology

- 2/13/2012 - begin trials with final calibration and alignments from Leon; 5 runs reprocessed
- 2/14/2012 - trials continue with blocks of 15, 20, 25 and 50 runs reprocessed (each run generates ~20 batch jobs)
- 2/16/2012 - begin trickleStream production. Initial config:

```
=====
TRICKLE PARMS
=====
task = P202-ROOT
maxRuns = 19172
firstStep = setupRun
steps = [['/processRun processClump', 1500, 20], ['mergeClumps', 70, 1]]
maxStreamsPerCycle = 20
timePerCycle = 300
=====
```

- 2/21/2012 - One clump reprocessed with pointer to new mySQL DB (stream 710.0)
- 2/22/2012 - 776 runs complete. Pausing task.

S/W component	modification	status
FILTEREDMERIT TCut	CTBClassLevel>0 changed to ((FT1EventClass & 0x00003EFF)!=0)	done 3/14/2012

- 3/15/2012 - resume task. New goal is 1-year of data (~5600 runs)
- 3/31/2012 - 1-year complete (5600 runs). There have been a few nasty problems which need to be fixed before continuing.

S/W component	bug fix	status
New ROOT version	5-min 'transaction timeout' triggered by xroot data server reboot	done 4/3/2012
New GlastRelease	1) include new ROOT version (above); 2) exit with non-zero RC on ROOT write error	done 4/5/2012, GR 17-35-24-rp04 (or -rp07)
New GPL_TOOLS ?	check size/checksum of file written to xroot with known size/checksum	pending

Tuned xroot on new Dell servers	silent file truncation when volume fills up JIRA	done 4/4/2012 (100 MB min space limit -> 100 GB; file system space check cadence changed from 10 min to 2 min)
New xroot client tools	complain when xroot data server fails on write	done 4/3/2012, v3.1.1
New TSkim	1) new ROOT version (above); 2) complain on ROOT write errors	done 4/5/2012, v08-02-01
New xroot redirector	required step toward enabling HPSS staging	done 4/3/2012, v3.1.1

Note also that the FILTEREDMERIT files contain 42 more events than the EXTENDEDFT1 files; they should be identical.

- 4/5/2012 - resume task. New goal is entire science dataset.
- 4/10/2012 - Unknown 'glitch' may have caused a few 100's of jobs to crash and take sulky46 along with them.
- 4/11/2012 - due to possible overload of sulky46/u18 writing a lot of core files, have introduced one change to processClumps.py: prepend "ulimit -c 0;" to gleam command to disable all core file generation. This starts approx with run 7605 (+/-).
- 4/12/2012 - 10:40pm lightening strikes SLAC power lines. Site-wide power outage. Stream 7795 was the last stream submitted prior to the outage.
- 4/15/2012 - Batch farm back in operation, resume task...
- 5/9/2012 - major pipeline issue...shut down pipeline and allow to drain (due to tomorrow's major outage)
- 5/10/2012 - 13:40 outage over.
 - Update GR from 17-35-24-rp04 to 17-35-24-rp07 in which the only change is replacing the 5-minute xroot time-out with 8 hours. This change effective with stream 14314 and previously failed pieces of four other runs: 14247.6, 14273.23, 14274.8, 14231.9.
 - Leon advises that as of today, calibrations are valid only thru ~15 Dec 2011 (run 345574915) - which is somewhere around stream 18,400. He asks Sasha to produce more up-to-date calibs.
- 5/18/2012 - all calibrations now valid through 6 May 2012. No need to pause P202 task.
- 5/28/2012 - 15:30 Complete (through 31 March 2012)
 - Data Catalog summary:

Name	Type	Files	Events	Size	Created (UTC)	Links
CAL	Group	20229	44,125,599,595	128.7 TB	25-Jan-2012 00:53:31	Files
ELECTRONMERIT	Group	20229	90,904,582	205.7 GB	25-Jan-2012 00:53:32	Files
FILTEREDMERIT	Group	20229	6,291,396,710	5.3 TB	25-Jan-2012 00:53:29	Files
GCR	Group	20229	44,123,014,456	942.7 GB	25-Jan-2012 00:53:31	Files
MERIT	Group	20229	44,125,679,961	35.4 TB	25-Jan-2012 00:53:30	Files
RECON	Group	20229	44,123,612,977	590.0 TB	25-Jan-2012 00:53:33	Files

There are discrepancies to track down!

Turns out to be three problematic runs/streams:

- 272707024/5723 - I/O prob, corrupt files, entire stream rolled back
 - 279108810/6847 - xroot transient access prob., re-registered in dataCat
 - 284813327/7848 - xroot transient access prob., re-registered in dataCat
- Final trickleStream configuration:

```
=====
TRICKLE  PARMS
=====
task = P202-ROOT
maxRuns = 20229
firstStep = setupRun
steps = [['/processRun processClump', 2000, 21], ['mergeClumps', 200, 1]]
maxStreamsPerCycle = 20
timePerCycle = 300
-----DEBUG-----
maxCycles = 0
chatter = False
dryRun = False
=====
```

- 5/31/2012 - Cleanup and summary
 - Rolling back all or part of the three runs above solved the discrepancies in # events. New dataCatalog tally looks like this:

Name	Type	Files	Events	Size	Created (UTC)	Links
CAL	Group	20229	44,125,679,961	128.7 TB	25-Jan-2012 00:53:31	Files
ELECTRONMERIT	Group	20229	90,904,582	205.7 GB	25-Jan-2012 00:53:32	Files
FILTEREDMERIT	Group	20229	6,291,396,711	5.3 TB	25-Jan-2012 00:53:29	Files

GCR	Group	20229	44,125,679,961	942.7 GB	25-Jan-2012 00:53:31	Files
MERIT	Group	20229	44,125,679,961	35.4 TB	25-Jan-2012 00:53:30	Files
RECON	Group	20229	44,125,679,961	590.0 TB	25-Jan-2012 00:53:33	Files

- Total run time for 20,229 runs was ~74 days (or about 273 runs/day reprocessed). This includes periods of changing trickleStream configuration as we figured out how much load we could safely put on the system.
- 6/5/2012 - Three streams rolled back and minor code changes for cleanup (see FITS chronology below for details)
- 8/10/2012 - Update task for a block3 of backfill (1 Apr 2012 - 31 July 2012) and restart reprocessing.
- 8/22/2012 - backfill complete
- 10/6/2012 - Rollback the following seven streams to fix apparently corrupt MERIT files.

Stream	Run	
3345	259101994	<- found by FSSC
4122	263571912	
4707	266893978	
13927	319436826	
16181	332306548	<- found by FSSC
17430	339161346	
17479	339408141	

- 10/8/2012 - Update task for block4 of backfill (1051 new runs for a total of 23,141) and start reprocessing.

First run of block 4	365473283	2012-08-01 00:21:20 UTC
Last run of block 4	371258376	2012-10-06 23:19:33 UTC

- 10/15/2012 - Block 4 complete. One problem with run 22240 (see below), rolled back successfully.
- 12/13/2012 - Update task for block5 of backfill (1001 new runs for a total of 24,142)

First run of block 5	371264424	2012-10-07 01:00:21 UTC
Last run of block 5	376959687	2012-12-11 23:01:24 UTC

Status of Calibrations: unknown

- 12/14/2012 - Status of dataCatalog prior to beginning block 5 catch-up:

Name	Type	Files	Events	Size
CAL	Group	23141	50,454,195,283	146.6 TB
ELECTRONFT1	Group	23141	N/A	9.6 GB
ELECTRONMERIT	Group	23141	103,221,645	233.1 GB
EXTENDEDFT1	Group	23141	7,252,996,601	662.6 GB
EXTENDEDLS1	Group	23141	7,252,996,601	1.1 TB
FILTEREDMERIT	Group	23141	7,252,996,601	6.1 TB
FT1	Group	23141	229,126,932	21.5 GB
GCR	Group	23141	50,454,195,283	1.0 TB
LS1	Group	23141	1,549,133,207	251.7 GB
MERIT	Group	23141	50,454,195,283	40.4 TB
RECON	Group	23141	50,454,195,283	673.4 TB

- 12/18/2012 - block 5 complete

Name	Type	Files	Events	Size
CAL	Group	24142	52,616,161,016	152.8 TB
ELECTRONMERIT	Group	24142	107,563,959	242.8 GB
FILTEREDMERIT	Group	24142	7,578,468,156	6.4 TB
GCR	Group	24142	52,616,161,016	1.1 TB
MERIT	Group	24142	52,616,161,016	42.1 TB
RECON	Group	24142	52,616,161,016	702.0 TB

- 3/25/2013 - prepare for block 6 backfill

First run of block 6	376965268	2012-12-12 00:34:25
Last run of block 6	385777036	2013-03-24 00:17:13

Block 6 represents 25,682 runs, an increase of 1540 runs over block 5.

Current status of P202 dataCatalog has not changed since 12/18/2012 (see above).

- 4/1/2013 - block 6 complete. Xroot scratch cleaned up, new CAL & RECON removal lists conveyed to Wilko.
- 5/2/2013 - prepare for block 7 backfill

First run of block 7	385782758	2013-03-24 01:52:35
Last run of block 7	389089696	2013-05-01 08:28:13

Block 7 contains 26263 runs, an increase of 581 runs.

DataCatalog report before start of block 7:

Name	Files	Events	Size
CAL	25682	55,969,490,601	162.4 TB
GCR	25682	55,969,490,601	1.2 TB
MERIT	25682	55,969,490,601	44.8 TB
RECON	25682	55,969,490,601	746.4 TB

- 5/6/2013 - block 7 complete
 - bulk of runs completed within 2 days, one extra day for stragglers
 - DataCatalog report after completion of block 7:

Name	Files	Events	Size
CAL	26263	57,227,320,767	166.0 TB
ELECTRONMERIT	26263	116,799,950	263.6 GB
FILTEREDMERIT	26263	8,284,037,323	7.0 TB
GCR	26263	57,227,320,767	1.2 TB
MERIT	26263	57,227,320,767	45.8 TB
RECON	26263	57,227,320,767	763.1 TB

- the number of events in the unfiltered ROOT files agrees with checkRunList script operating on the input runList.
- xroot scratch cleaned up
- list of RECON and CAL files to cleanup sent to Wilko
- 6/20/2013 - Tale of two runs

Stream	Run	Comment
25232	383219654	Truncated run (~9 min), recovered, rolled back
26263	338868584	mysteriously appeared in most recent genRunFile cycle, had to append to end of runList

What happened? Warren says this run is perfectly normal. Could have the "Intents" changed? This single orphan run, tacked onto the end of block7 (run 389089696) and will be known as "block 8" (one new run and one updated run).

- 6/27/2013 - Gear up for block 9 backfill, through 6/25/2013.
 - Last run is 393895214 (2013-06-25 23:20:11 UTC)
 - Regenerate runList.txt, and move run 338868584 to after run 389089696 to preserve stream<->run correspondence. Note that the runList is now OUT OF ORDER, and the setupRun.py script has been changed to NOT automatically order its internal list of runs.
 - block 9 contains 27114 runs and 59,049,646,168 digi events.
 - 851 new runs to reprocess
- 6/28/2013
 - Discover problem with mergeClumps when it runs on a bullet (RHEL6-64) machine. Halt trickleStream after stream 26687. Make two code changes:
 - mergeClumps.py - remove env setup for ST (it is not needed)
 - config.py - update GPLtools to enable use of /lustre scratch area on bullets
 Then test five rollbacks: streams 26264-26268.
- 6/30/2013
 - block 9 basically complete Saturday morning (6/29/2013) except for five stalled jobs, rolled back. Then two merge steps took a very long time to complete.
- 7/1/2013

- Current ROOT file generation situation after block 9. Number of files and events constitutes a level 0 consistency test.

Name	Files	Events	Size
CAL	27114	59,049,646,168	171.2 TB
ELECTRONMERIT	27114	120,405,572	271.6 GB
FILTEREDMERIT	27114	8,562,732,063	7.2 TB
GCR	27114	59,049,646,168	1.2 TB
MERIT	27114	59,049,646,168	47.3 TB
RECON	27114	59,049,646,168	787.2 TB

- Available xroot space = 147.0 TB

- 8/13/2013
 - 12:00 Begin block 10 backfill, 759 runs in the range 393900935 through 398081853 (2013-08-13 10:17:30). Once complete, 27873 runs will have been reprocessed.
- 8/16/2013 - block 10 complete (see below for datacatalog content)
- 9/17/2013
 - 13:33 Begin block 11 backfill, 539 runs in the range 398087830 through 401106634 (2013-09-17 11:48:16). Once complete, 28412 runs will have been reprocessed.
- 9/19/2013
 - 08:30 block 11 complete. The bulk of runs completed within 14 hours, but stragglers, failed/terminated jobs, due to transient problems, bad batch machines, etc., delayed completion until this morning. DataCatalog reports a total of 61,827,628,706 events in {CAL,MERIT, RECON,GCR} and 8,991,163,530 events in FILTEREDMERIT.
 - xroot scratch space cleaned up.
- 10/4/2013
 - Prepare for block 12 backfill. Warren suggests ending with run 402560477 (2013-10-04 06:21:14 UTC). First run of block 12 is 401112810. There are 263 runs in this backfill block which will bring the reprocessed run total to 28675.
- 10/15/2013
 - Prepare for block 13 backfill. Warren suggests ending with run 403510814, 170 new runs.

	Run	Run start time (UTC)	Task Stream
Block 13 start	402566464	2013-10-04 08:01:01	28675
Block 13 end	403510814	2013-10-15 06:20:11	28844

- 11/05/2013
 - Today P7REP goes public.
 - Prepare for block 14 backfill – the ultimate backfill for P7REP. Last run = 405329691 (2013-11-05 07:34:48)
 - 313 new runs to reprocess...

	Run	Run start time (UTC)	Task Stream
Block 14 start	403516762	2013-10-15 07:59:19	28845
Block 14 end	405329691	2013-11-05 07:34:48	29157

- 8/19/2014
 - User reported problem with an event inside a ~195s interval with LIVETIME=0 in run 395891323 (2013-07-19 01:48:40 UTC). Problem traced to a bad FT2SECONDS file. Reprocessing took place 14 Aug 2013 using v002 FT2 file. Then on 19 Aug 2013, the FT2 file was updated to v003. M.E. has rebuilt a proper FT2 file in the Reprocess/P202 area, v203 and will roll back entire stream 27487 to rebuild ROOT files. To get this rollback to run, the following steps were necessary:
 - config.py - change ft2Selection from 'P105' to 'P202' (commonTools/repTools.py:findFT2()) already knows about P202, so no changes needed there.)
 - commonTools/setupGR.sh - change svsopts from 'redhat5' back to 'redhat4' (TEMPORARILY) for processClump step
 - commonTools/setupSkimmer.py - a series of reversion:
 - revert to skimmer 08-02-01
 - revert to rootVersion = 'v5.26.00a-gl6'
 - revert glastBuild from redhat5-i686-32bit-gcc41 to redhat4-i686-32bit-gcc34 to get libraries for mergeClumps step.
 - revert to old ROOTSYS definition that includes compiler version as last path item
 - The mods to setupGR.sh and setupSkimmer.py have been preserved as OLD-setupGR.sh and OLD-setupSkimmer.py in the commonTools directory

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P202-ROOT
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/index.jsp

GlastRelease	17-35-24-gr17 and 17-35-24-rp04 (SCons RHEL4-32 build)
Run Selection	based on a modified "standard" selection, see https://confluence.slac.stanford.edu/display/SCIGRPS/Official+LAT+Datasets (((sIntent=="nomSciOps" sIntent=="nomSO_noSk_noCno_optGccc_allEna" sIntent=="nomSciOps_diagEna" (sIntent=="nomSciOps_Emin5MeV"&&RunMin>242070455) nRun==242429468) && (RunQuality != "Bad" is_null (RunQuality))) sIntent=="nadirOps")
s/c data	"standard" Public Release 2 https://confluence.slac.stanford.edu/display/SCIGRPS/Official+LAT+Datasets
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P202-ROOT/config/runList.txt
photonFilter	CTBParticleType==1 && ((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1 in evtClassDefs
electronFilter	CTBParticleType==1
Code Variants	redhat4-i686-32bit-gcc34 (Optimized)
jobOpts	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P202-ROOT/config/doRecon.txt
Output Data Products	RECON, GCR, CAL, MERIT, FILTEREDMERIT, ELECTRONMERIT

Timing and Scaling

- processClump
 - with 1300 jobs completed, the average time to run varies by processor type from 220 min (hequ) to 370 min (boer).
 - ..
 - with nearly 10,000 runs complete, the plots appear below:
 - ..
- mergeClumps
 - with 42 jobs completed, the average time to run varies by processor type from 5-30 minutes.
 - ..

Load balancing

trickleStream parameters (see above).

P202-FITS

This task generates all desired FITS data products.

Status chronology

- 3/2/2012 - Define block 1 as the 776 runs in P202-ROOT block 1. Configure trickleStream and begin (14:08)
- 3/31/2012 - Define block 2 as 5600 runs. Reconfig trickleStream and begin (18:05)
- 4/01/2012 - Block 2 complete (most of the 4824 jobs completed in about six hours w/1000 job limit).
- 5/31/2012
 - Discover stream 5599 (run 271999199) requires rollback - new MERIT file version (v203->v205).
 - Regenerate runlist with 20,229 input MERIT files.
 - Minor config change: twoClumpMin=False (formerly True, but caused unnecessary extra processClump dummy jobs)
- 6/5/2012 - Final cleanup
 - Five runs are responsible for discrepant event tallies:

Run	Stream	Reason	Action
239557414	0	bad TCut	rollback ROOT + FITS
241599746	352	bad processing order	rollback FITS
245403855	1019	bad start time	change start time in runList and rollback
332661583	16244	silent root errors	rollback ROOT
339081502	17416	silent root errors	rollback ROOT

- Code changes:

directory	script	modification
-----------	--------	--------------

commonTools/00-01-00	repTools.py	added new getKey() function to extract #events from FITS event files
P202-FITS/config	config.py	added os.environ['HEADASNOQUERY']='true' to enable 'flist' to run in batch
P202-FITS/config	makeFITS.py	add diagnostic print of #evts in FITS files for each processing sub-step

- After these actions, the dataCatalog tallies are now consistent:

Name	Type	Files	Events	Size	Created (UTC)	Links
ELECTRONFT1	Group	20229	0	8.5 GB	02-Mar-2012 00:06:07	Files
ELECTRONMERIT	Group	20229	90,904,582	205.7 GB	25-Jan-2012 00:53:32	Files
FT1	Group	20229	189,323,074	17.8 GB	02-Mar-2012 00:06:06	Files
LS1	Group	20229	1,325,204,821	215.3 GB	02-Mar-2012 00:06:08	Files
EXTENDEDFT1	Group	20229	6,291,424,926	574.7 GB	02-Mar-2012 00:06:09	Files
EXTENDEDLS1	Group	20229	6,291,424,926	1,020.1 GB	02-Mar-2012 00:06:09	Files
FILTEREDMERIT	Group	20229	6,291,424,926	5.3 TB	25-Jan-2012 00:53:29	Files
MERIT	Group	20229	44,125,679,961	35.4 TB	25-Jan-2012 00:53:30	Files
RECON	Group	20229	44,125,679,961	590.0 TB	25-Jan-2012 00:53:33	Files
GCR	Group	20229	44,125,679,961	942.7 GB	25-Jan-2012 00:53:31	Files
CAL	Group	20229	44,125,679,961	128.7 TB	25-Jan-2012 00:53:31	Files

Note that the number of events in ELECTRONFT1 files are not currently tallied by the dataCatalog.

- 8/24/2012 - Configure and run backfill through 31 July 2012
- 10/7/2012 - Rollback the following seven streams to fix corrupt FITS files. The makeFT1 app received a large number of ROOT error, but terminated normally. The resulting FT1/LS1 files had multiple symptoms, including one discovered by the FSSC during transfer to them: one EVENT_ID was repeated ~1000 times. Scan of log files uncovered an additional five mergeClumps with these errors.

```
%INFO: 20120331:20:34:32 - makeFITS(run)/line-127 - Running: makeFT1
----- start commentary -----
About to run [time makeFT1 rootFile=/scratch/glastmp/P202-FITS/3345/r0259101994_v202_merit.root fitsFile=
/scratch/glastmp/P202-FITS/3345/gll_xp_p202_r0259101994_v202.fit TCuts=/afs/slac.stanford.edu/g/glast
/ground/releases//volume04/evtClassDefs/00-19-04/data/pass7.6_Extended_cuts_L1 dict_file=/afs/slac.
stanford.edu/g/glast/ground/releases//volume04/evtClassDefs/00-19-04/data/FT1variables tstart=259101996.
933000 tstop=259106272.085000 file_version=202 tempRootFile=/scratch/glastmp/P202-FITS/3345/dummy.root
xml_classifier=/afs/slac.stanford.edu/g/glast/ground/releases//volume04/evtClassDefs/00-19-04/xml
/EvtClassDefs_P7V6.xml evtclsmmap=FT1EventClass chatter=4 debug=yes] at Sat Mar 31 20:34:32 2012
----- start log -----
This is makeFT1 version ScienceTools-09-27-01
applying TCut: ((FT1EventClass & 0x00003EFF)!=0) && (EvtElapsedTime >= 259101996) && (EvtElapsedTime <=
259106273)
Warning in <TClass::TClass>: no dictionary for class FileHeader is available
Warning in <TClass::TClass>: no dictionary for class RootObj<int> is available
R_unzip: error in header
Error in <TBasket::ReadBasketBuffers>: fNbytes = 15550, fKeylen = 86, fObjlen = 31912, noutot = 0,
nout=0, nin=11010118, nbuf=7864444
Error in <TBranch::GetBasket>: File: /scratch/glastmp/P202-FITS/3345/r0259101994_v202_merit.root at byte:
459100837, branch:Vtx2LongDoca, entry:510338, badread=0, nerrors=1, basketnumber=1
[...]
```

Stream	Run	
3345	259101994	<- found by FSSC
4122	263571912	
4707	266893978	
13927	319436826	
16181	332306548	<- found by FSSC
17430	339161346	
17479	339408141	

- 10/15/2012 - Reconfigure for block 4 backfill and begin trickleStream.

- Discover run 22240 makeFT1 contains error messages as the mentioned last week (above). Rolling back the entire chain of processing, starting with P202-ROOT seemed to do the trick.
- 10/17/2012 - block4 backfill complete.
- 12/18/2012 - Prepare for block 5. Before:

Name	Type	Files	Events	Size
EXTENDEDFT1	Group	23141	7,252,996,601	662.6 GB
EXTENDEDLS1	Group	23141	7,252,996,601	1.1 TB
FT1	Group	23141	229,126,932	21.5 GB
LS1	Group	23141	1,549,133,207	251.7 GB

- 12/19/2012 - block5 backfill complete. Final DataCatalog numbers:

Name	Type	Files	Events	Size
EXTENDEDFT1	Group	24142	7,578,468,156	692.3 GB
EXTENDEDLS1	Group	24142	7,578,468,156	1.2 TB
FT1	Group	24142	240,949,332	22.6 GB
LS1	Group	24142	1,621,773,636	263.5 GB

- 3/6/2013 - Many changes!
 - New task created (version 0.9 -> 1.0)
 - Update to run native on RHEL5-64 and RHEL6-64
 - New Interstellar Emission Model, v2r0 -> v3r0, provided by Luigi Tibaldo.
 - Update ScienceTools from 09-27-01 to 09-31-01
 - Update IRFS from P7*V6 to **P7REP_V10** (= CLEAN or SOURCE)
 - This means that rather than rolling-back, the task starts from scratch.
 - New evtClassDiffs version: 00-19-05, which changes P7V6 -> P7REP, the new designation.
 - Reference list of all existing FITS files (with old diffuse model) generated and stored here: /nfs/farm/g/glast/u38/Reprocess-tasks/P202-FITS/config/task-v0.9

```
diffmodel_p7rep_clean_v10_reduced.xml
diffmodel_p7rep_clean_v10.xml
diffmodel_p7rep_source_v10_reduced.xml
diffmodel_p7rep_source_v10.xml
gal_p7rep_v10_v1.fits
gal_p7rep_v10_v1_reduced.fits
iso_p7rep_clean_v10_back_v1.txt
iso_p7rep_clean_v10_front_v1.txt
iso_p7rep_clean_v10_v1.txt
iso_p7rep_source_v10_back_v1.txt
iso_p7rep_source_v10_front_v1.txt
iso_p7rep_source_v10_v1.txt
```

- 3/7/2013 - Run off 10 test runs (starting at beginning of mission with task version 1.0)
- 5/30/2013 - Green light given on new IRFs and diffuse model.
 - Changes to config.py:
 - New diffuse model: /afs/slac/g/glast/ground/GLAST_EXT/diffuseModels/v4r0
 - New IRFs: P7REP_SOURCE_V15,P7REP_CLEAN_V15
 - Awaiting a new ScienceTools release...
- 6/10/2013 - Several ST updates, and renaming of the diffuse model files. All appears good to go.
- 6/11/2013 - final tweaks to diffuse model v4r0 files; bump task version 1.0 -> 1.1 (delete version 1.0); and start...
- 6/14/2013 - block 1 complete (26,263 runs: 239557414 through 389089696; MET 2008-08-04 15:43:33 through 2013-05-01 08:28:13)

Name	Files	Events	Size
CAL	26263	57,227,320,767	166.0 TB
DIGIGAP	24200	0	19.0 kB
ELECTRONFT1	26263	0	10.9 GB
ELECTRONMERIT	26263	116,799,950	263.6 GB
EXTENDEDFT1	26263	8,284,002,713	725.9 GB
EXTENDEDLS1	26263	8,284,002,713	1.3 TB
FILTEREDMERIT	26263	8,284,037,323	7.0 TB
FT1	26263	268,810,274	24.2 GB

GCR	26263	57,227,320,767	1.2 TB
LS1	26263	1,782,493,106	289.6 GB
MERIT	26263	57,227,320,767	45.8 TB
RECON	26263	57,227,320,767	763.1 TB

Discrepancy between FILTEREDMERIT and EXTENDED{LS1,FT1}. This turns out to be an issue with tstart/tstop for run 383219654.

- 8/16/2013 - configure block2 backfill: 1610 new runs to reprocess
 - 27873 total runs
 - 60,682,674,790 total events
 - Start run: 239557417 2008-08-04 15:43:37
 - Last run: 398086126 2013-08-13 11:28:46
- 8/21/2013 - block2 complete (after redoing the entire block due to bizarre effects of last Friday's fermi-xrd005 crash). DataCatalog contents as of this morning:

Name	Files	Events	Size	Created (UTC)
CAL	27873	60,682,674,790	175.8 TB	25-Jan-2012 00:53:31
ELECTRONFT1	27873	0	11.5 GB	02-Mar-2012 00:06:07
ELECTRONMERIT	27873	123,494,286	278.5 GB	25-Jan-2012 00:53:32
EXTENDEDFT1	27873	8,811,129,094	772.1 GB	02-Mar-2012 00:06:09
EXTENDEDLS1	27873	8,811,129,094	1.4 TB	02-Mar-2012 00:06:09
FILTEREDMERIT	27873	8,811,129,090	7.5 TB	25-Jan-2012 00:53:29
FT1	27873	289,969,364	26.1 GB	02-Mar-2012 00:06:06
GCR	27873	60,682,674,790	1.3 TB	25-Jan-2012 00:53:31
LS1	27873	1,903,568,484	309.3 GB	02-Mar-2012 00:06:08
MERIT	27873	60,682,674,790	48.6 TB	25-Jan-2012 00:53:30
RECON	27873	60,682,674,790	808.5 TB	25-Jan-2012 00:53:33

- 9/19/2013 - prepare block3 (corresponding to P202-ROOT block11).
- 9/20/2013 - block3 complete. One run from block2 was rolledback due to not using the latest MERIT file (run 383219654). Statistics are now consistent.

.

- 10/7/2013 - prepare block4 (corresponding to P202-ROOT block12).
- 10/8/2013 - block 4 complete

Name	Files	Events	Size	Created (UTC)
ELECTRONMERIT	28675	126,895,805	286.1 GB	25-Jan-2012 00:53:32
FT1	28675	301,166,782	27.1 GB	02-Mar-2012 00:06:06
LS1	28675	1,964,328,038	319.1 GB	02-Mar-2012 00:06:08
EXTENDEDLS1	28675	9,072,793,026	1.4 TB	02-Mar-2012 00:06:09
FILTEREDMERIT	28675	9,072,793,026	7.7 TB	25-Jan-2012 00:53:29
EXTENDEDFT1	28675	9,072,793,026	795.0 GB	02-Mar-2012 00:06:09
MERIT	28675	62,383,754,997	49.9 TB	25-Jan-2012 00:53:30
CAL	28675	62,383,754,997	180.6 TB	25-Jan-2012 00:53:31
GCR	28675	62,383,754,997	1.3 TB	25-Jan-2012 00:53:31
RECON	28675	62,383,754,997	831.0 TB	25-Jan-2012 00:53:33

- 10/17/2013 - prepare block5 backfill (P202-ROOT block13).
- 10/24/2013 - block 5 complete (as of 10/18/2013)

Name	Files	Events	Size
ELECTRONMERIT	28844	127,613,745	287.7 GB
FT1	28844	303,135,818	27.3 GB
LS1	28844	1,976,418,512	321.1 GB
EXTENDEDLS1	28844	9,127,005,732	1.4 TB
FILTEREDMERIT	28844	9,127,005,732	7.7 TB

EXTENDEDFT1	28844	9,127,005,732	799.8 GB
MERIT	28844	62,747,048,217	50.2 TB
CAL	28844	62,747,048,217	181.6 TB
GCR	28844	62,747,048,217	1.3 TB
RECON	28844	62,747,048,217	835.8 TB

- 11/7/2013 - block 6 complete and P7REP was made public two days ago. END OF PROJECT

Final tally by the dataCatalog:

Name	Files	Events	Size	Created (UTC)
CAL	29158	63,449,261,789	183.6 TB	25-Jan-2012 00:53:31
ELECTRONMERIT	29158	128,971,706	290.7 GB	25-Jan-2012 00:53:32
EXTENDEDFT1	29158	9,243,063,549	809.9 GB	02-Mar-2012 00:06:09
EXTENDEDLS1	29158	9,243,063,549	1.5 TB	02-Mar-2012 00:06:09
FILTEREDMERIT	29158	9,243,063,549	7.8 TB	25-Jan-2012 00:53:29
FT1	29158	310,326,817	27.9 GB	02-Mar-2012 00:06:06
GCR	29158	63,449,261,789	1.3 TB	25-Jan-2012 00:53:31
LS1	29158	2,007,123,864	326.1 GB	02-Mar-2012 00:06:08
MERIT	29158	63,449,261,789	50.8 TB	25-Jan-2012 00:53:30
RECON	29158	63,449,261,789	845.0 TB	25-Jan-2012 00:53:33

- 8/20/2014 - Using new MERIT file from P202-ROOT (see log entry above), update fullList.txt and then rollback/regenerate FITS files for run 395891323, stream 27487.

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P202-FITS
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/task.jsp?task=107152539
Input Data	MERIT (direct from P202-ROOT)
spacecraft data	same as P202-ROOT
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P202-FITS/config/runList.txt
evtClassDefs	00-19-05 (March 2013, changed pass_ver to P7REP)
eventClassMap	EvtClassDefs_P7V6.xml
ScienceTools	09-32-03 (6/7/2013) (but ST may report themselves as 09-32-02 due to RM snafu)
Code Variants	redhat5-x86_64-64bit-gcc41, redhat6-x86_64-64bit-gcc44 (Optimized)
Diffuse Model	based on contents of /afs/slac.stanford.edu/g/glast/ground/GLAST_EXT/diffuseModels/v4r0 (see https://confluence.slac.stanford.edu/display/SCIGRPS/Quick+Start+with+Pass+7)
Diffuse Response	'source' using P7REP_SOURCE_V15 IRF 'clean' using P7REP_CLEAN_V15 IRF
IRFs	P7REP_*_V15, contained within ScienceTools release
Output Data Products	FT1 , LS1 , EXTENDEDFT1 , EXTENDEDLS1 , ELECTRONFT1

Generation of output data products:

Data Product	destination	data content [1]	event selection [1]	makeFT1	gtselect	gtdiffresp	gtmkttime
EXTENDEDFT1	SLAC	FT1variables	((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1	✓	✗	✓	✓
FT1	FSSC+SLAC	FT1variables	'source' and above EVENT_CLASS bits 2,3,4 evclass=2 filtered from EXTENDEDFT1	✗	✓	(inherited)	✓
EXTENDEDLS1	SLAC	LS1variables	((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1	✓	✗	✓	✓

LS1	FSSC+SLAC	LS1variables	'transient' and above EVENT_CLASS bits 0,2,3,4 evclass=0 filtered from EXTENDEDLS1	✗	✓	(inherited)	✓
ELECTRONFT1	SLAC	FT1variables	CTBParticleType==1 pass7.6_Electrons_cuts_L1	✓	✗	✗	✓

[1] /afs/slac/g/glast/ground/releases/volume04/evtClassDefs/00-19-04/data

Note that diffuse response is calculated for 'source' and 'clean' event classes only.

Note on 'Code Variant': The SLAC batch farm contains a mixture of architectures , both hardware (Intel/AMD 64-bit) and software (RHEL5-64, gcc v4.1, etc.). At this time, GlastRelease builds only on RHEL4-32, while ScienceTools builds for RHEL5-32, RHEL5-64.

Timing

-

P202-LEO-ROOT

Status chronology

- 8/8/2012 - Prepare task
- 8/18/2012 - 200 runs complete
- 8/25/2012 - Per Seth and Anders, add five runs and remove two:
 - Add: 238421027, 238489647, 239108423, 239114152, 239208666
 - Remove: 244395837, 244401823
 (Note that removed runs have only been removed from dataCatalog – not from xroot)
 But first, need FT2 files for these five runs!
- 10/19/2012
 - P130-series FT2 files now available for early L&EO period
 - The five runs above were never reprocessed; Seth blesses a new set of 64 runs to take their place
 - Add 64 new runs to runList.txt (generate run list for entire L&EO period, then manually extract the new 64)
 - Begin trickleStream
- 10/20/2012 262 runs complete (Original 200 - two 0244* runs + new 64). Note that the data products from streams 198 and 199 have been de-registered from the dataCatalog but retained in xroot. Current dataCatalog statistics for the P202 L&EO reprocessing:

Name	Type	Files	Events	Size	Created (UTC)
CAL	Group	262	608,752,392	1.7 TB	10-Aug-2012 10:17:29
ELECTRONMERIT	Group	262	1,077,986	2.3 GB	10-Aug-2012 10:17:30
FILTEREDMERIT	Group	262	142,672,239	120.4 GB	10-Aug-2012 10:17:27
GCR	Group	262	608,752,392	13.6 GB	10-Aug-2012 10:17:28
MERIT	Group	262	608,752,392	499.9 GB	10-Aug-2012 10:17:27
RECON	Group	262	608,752,392	8.2 TB	10-Aug-2012 10:17:30

Configuration

Identical with P202-ROOT except for the list of runs to be processed...with one exception: to reprocess the four extra (out-of-order) L&EO runs, disable the event list sort.

Timing

P202 Update Checklist

A checklist for updating a new block of reprocessed data.

	Before
	determine first and last runs to reprocess.
	update genRunFile.csh and generate new list
	run checkRunList.py with new and old run lists
	run tkdiff with new and old run lists

	verify calibration constants are valid for new block
	check if new generation FT2 was introduced mid-block
	update trickleStream.py with new run count
	During
	monitor NFS and xroot performance
	periodically cleanup xroot scratch space
	periodically cleanup old RECON/CAL files (via list to Wilko)
	After
	run log scanner for silent root/xroot failures
	check dataCatalog statistics for consistency
	run xroot scratch cleanup procedure
	provide Wilko with list of old L1 RECON/CAL files to be removed from xroot disk