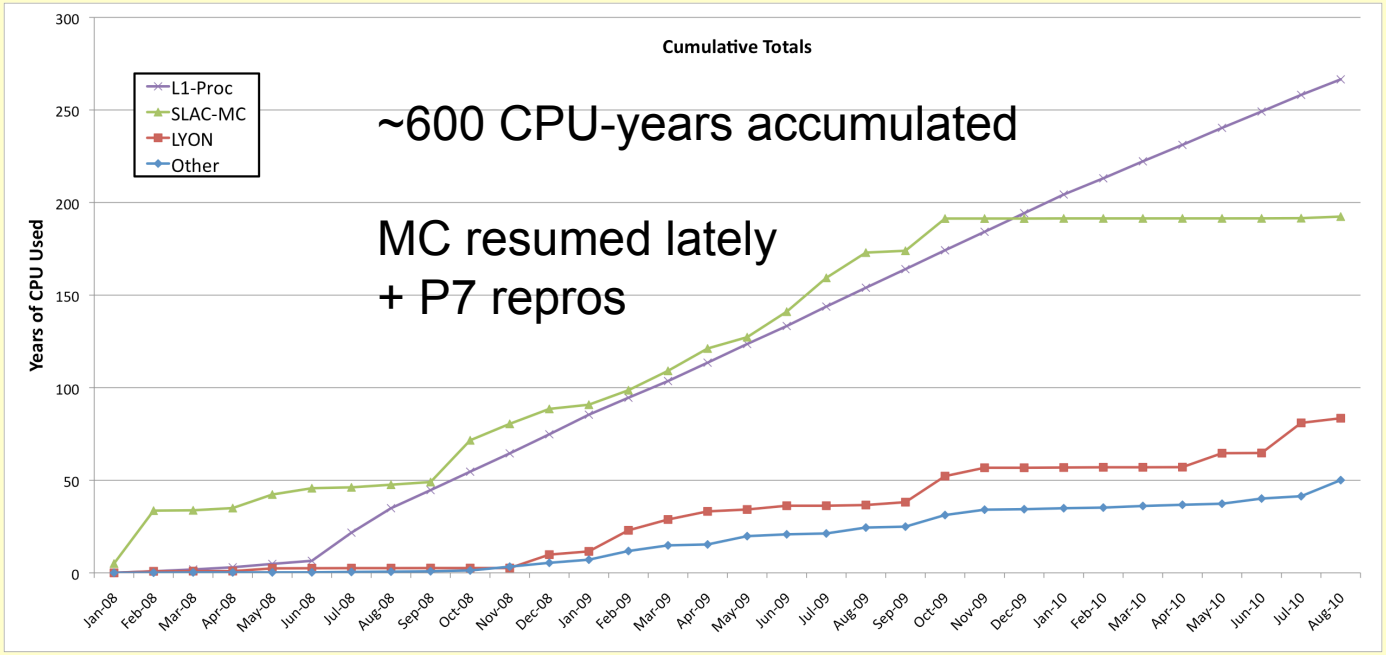
Collaboration Computing

Richard Dubois
richard@slac.stanford.edu

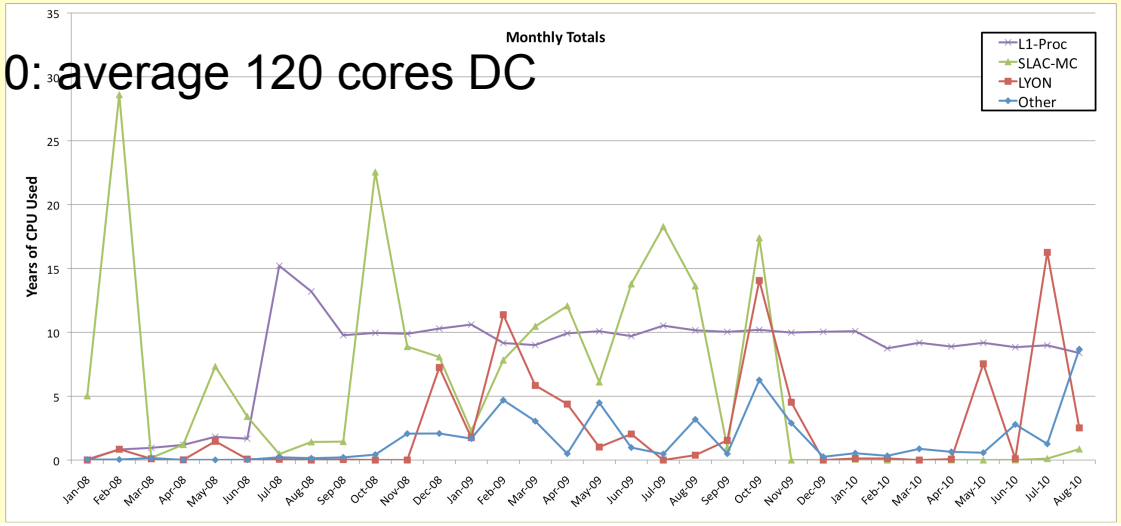
Current Resource State

- Resources
  - SLAC
    - 1600 cores + 1.06 PB disk + 1.25 PB tape – 168 TB disk free
    - We have to return 96 TB to BABAR for their kind loan during this year
    - 6 TB/wk used by L1 Processing
  - Lyon
    - 600 cores, used for MC
- Level 1 processing + ASP
  - Up to 900 cores to turn around downlink promptly – recon + monitoring etc
  - ASP not noticeable
  - L1 will exhaust disk in ~11-12 weeks
- Simulations have been fairly quiet as Pass7 is being validated.
  - Pass8 starting to ramp up
- Pass7 reprocessings run by Tom Glanzman (many thanks to Tom!!)
  - Merit reprocessing very I/O intensive and have to be limited so as not to overwhelm the xrootd servers

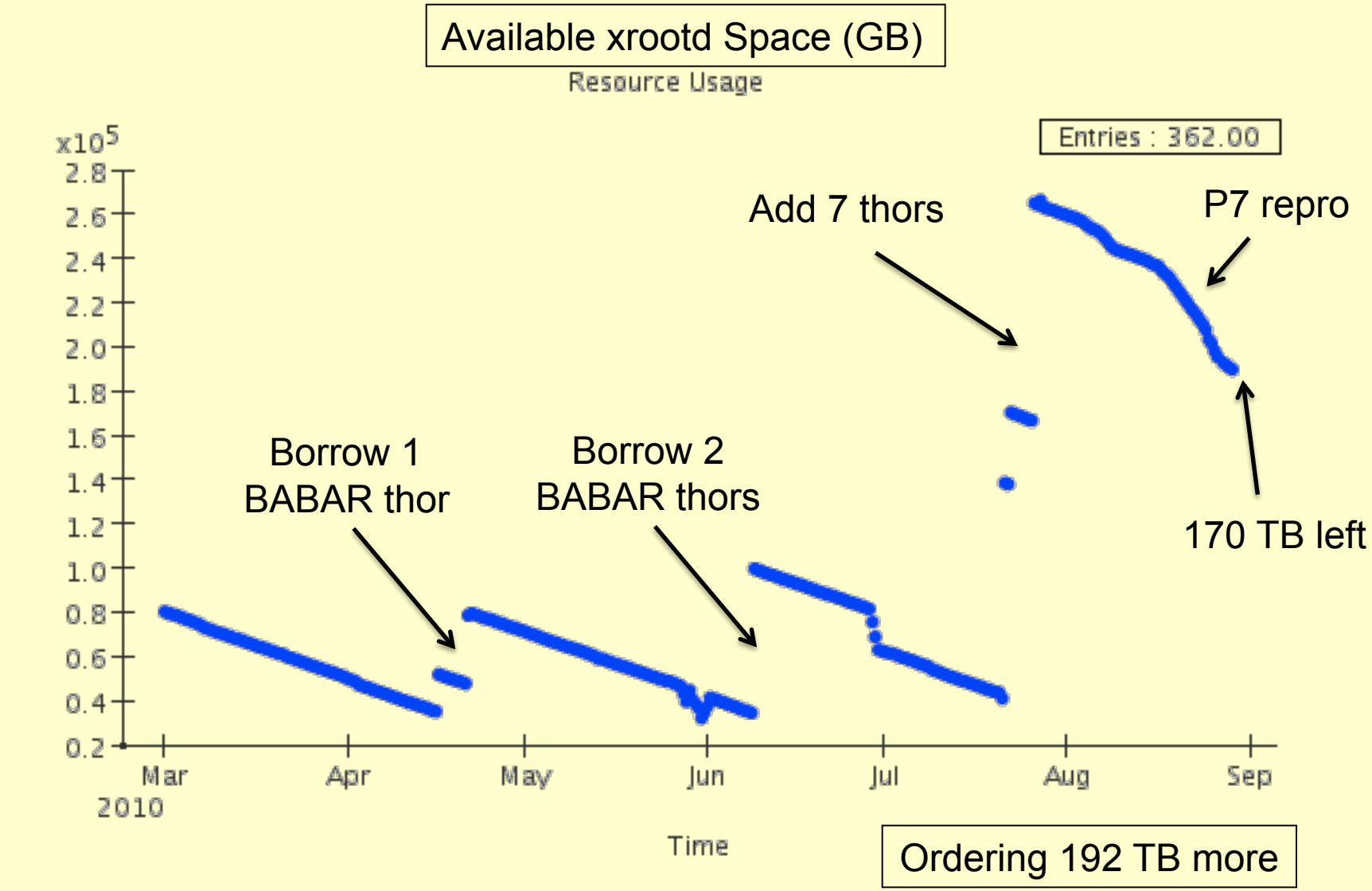
1.5 Yrs in CPU Cycles



L1 flat-lined at 10: average 120 cores DC



6 Months in Disk



Review of Computing Model

- Current model is:
  - Latest versions of files on disk
  - Older versions on tape (hopefully rarely if ever read)
- But: experience with the new silo shows transfer rates much higher than anticipated 2+ years ago
  - Can now transfer 30-50 TB/day from tape
  - No longer a significant bottleneck for reprocessing
  - Tape is 1/3 (or better) the price of disk
  - Review the model
    - We concluded we should test a model reminiscent of previous ones
      - Keep a few-month buffer of recon etc files
      - Have a working buffer for retrieving ~10 days' worth of files during a reprocessing
      - Rely on the tape copy to retrieve files
      - Make a 2nd tape copy
    - We will test this model over the next 6 months

Resource Prediction for 2011

- Accounting for purged L1 files, assume 25 TB/mo
  - 360 TB/yr
- Add 50 TB for MC
- Total 410 TB disk

- Assume 1 full reprocessing of all data to date
  - 1200 TB total
  - Need ~1600 TB of new tape by end of 2011
  - Storage costs are budgeted for via the IFC

- have we plateaued on cores?
  - Routine use is fine
  - What about spikes for reprocessing?
    - We should investigate cloud computing

Oracle ate Sun: we're getting heartburn

- We had a scare from Oracle this year
  - 6 months from start of order to delivery
  - Lost good pricing – got one-off good deal (10% higher than previous)
- Started the order for next 200 TB
  - We got a price quote last week: 50% higher – again
  - We are lobbying (again) for the 50% discount, but may not get it
  - Looking at an alternate vendor (Berkeley Communications - LCLS had used one of their systems this year)
  - A potential risk is whether Oracle will allow running Solaris on non-Sun hardware in the future…

OS Migration

- RHEL3 is gone – baseline now RHEL4
- SLAC Batch farm is now RHEL5 64 bit
  - We run RHEL4 32 bit builds on those nodes
- Would be easier if we can drop mac tiger Science Tools builds
- We need help from FSW to port OBF code to RHEL5
  - FSW promises progress report in November
  - MUST have a resolution by late 2011 when RHEL4 is retired

SCons Status

- **Reminder of What**
  - **Replacement for CMT – package management & build tool for multiple OSes**
  - **SCons is python based, so more easily configurable and we're back in (more) control of our destiny**
  - **Turned into a tougher nut to chew than we anticipated… windows is the bugaboo**
- **Why?**
  - **Support from CMT disappeared and we found we could not build for newer compilers**
- **ScienceTools has been the first target for SCons**
  - **Now working for linux and mac**
    - **ScienceTools ready to go – L1 & ft2util is the current hang-up for turning off CMT builds**
- **GlastRelease in progress**
  - **Gaudi upgrade done – now closing the loop on our package changes (Heather)**
- **See Joanne's nice update at the Pass8 workshop:**

https://confluence.slac.stanford.edu/download/attachments/92181543/P8-SCons.pdf

Tips for Newbies (reminders for the rest)

- Be sure you are using optimised code
- Using the batch system
  - Remember that pfiles can be hidden and that all batch jobs can write to the same .par files!
  - You could be rolling the dice on which version your job uses
  - You could (and do!) cripple afs or the user disk by many hundreds of jobs hitting a couple of files
  - Should set up scripts to use the local batch scratch disk for pfiles
    - Ensures the right .par file and no load on public servers
- User disks at SLAC
  - u31 & u33 are now readonly
  - 9 TB user disk with per-user quotas applied
    - /afs/slac/g/glast/users/<your_account_name>
    - No other user can affect your available space
    - We need to have SCCS create the initial partition; we can then increase it
    - We are also putting group space on this server
    - Web access to user space via “decorator”
      - http://glast-ground.slac.stanford.edu/Decorator/Decorate/users/

Overwhelming the User Disk

- **People are running lots of jobs in parallel now and overloading the user disk**
- **See Workbook for Best Practices Page:**

http://glast-ground.slac.stanford.edu/workbook/pages/installingOfflineSW/usingSlacBatchFarm.htm#bestPractices

Batch Job Listing : http://www.slac.stanford.edu/exp/glast/diskhogs/batchHogs.html

USER/GROUP	JL/P	MAX	NJOBS	PEND	RUN	SSUSP	USUSP	RSV
allafort	-	32	0	0	0	32	0	-
gudlaugu	-	12	0	5	7	0	0	-
guillemo	-	1	0	1	0	0	0	-
hadasch	-	1	0	1	0	0	0	-
kocevski	-	3	0	3	0	0	0	-
ksokolov	-	10	5	5	0	0	0	-
lande	-	810	244	545	21	0	0	-
lemoine	-	6	0	6	0	0	0	-
sanchez	-	9	0	9	0	0	0	-
ttanaka	-	1	0	0	1	0	0	-

Can also use ganglia to see who is abusing the disk!

http://ganglia01.slac.stanford.edu:8080/ganglia/fileservers/?m=load_one&r=hour&s=descending&c=nfs-glast&h=sulky55.slac.stanford.edu&sh=1&hc=3

Tips from Tom in the Meantime

- **1. Create a unique directory in /scratch for your batch job, e.g.,**

  mkdir -p */scratch/<userid>/*${LSB_JOBID}

  2. Define this directory as your $HOME and then go there prior to running any ScienceTools/Ftools/etc.,

  export HOME=*/scratch/<userid>/*${LSB_JOBID}
  cd ${HOME}

  This will automatically take care of PFILES being unique for your job, and not overloading the /nfs user disk with large numbers of opens and closes. Create any new files in $HOME and then copy anything you wish to save at the end of your job.

  3. Cleanup the scratch directory at end of job (after you have copied out anything you want to save),

  rm -rf */scratch/<userid>/*${LSB_JOBID}

  This last step is critical as any scratch files left behind will slowly fill up the /scratch partition and eventually fill it up!

Confluence tips

- You can get rid of the annoying profile pane in people's home spaces by clicking the little arrow in the middle right of the pane
- It is now possible to create confluence pages via a python script
  - Tom has done so for auto-creating reports for the binary RSP process – a weekly summary of the findings
- Don't forget to click "remember me" when you login and it won't ask for your password again
- Work is underway to let confluence check your unix/windows account rather than confluence. Hopefully one less pw to remember!

Data Access

- **LAT data portal: merit, fits skimmers; astro server**

http://glast-ground.slac.stanford.edu/DataPortal/

- **Via the web ("manual")**
  - **Use the LAT astroserver for FT1/FT2**
  - **Can now get merit files based on astro cuts**
- **Use command-line interface**
- **Download Manager**
  - **Java webstart app to grab files (used from Catalogue)**
  - **Programmatic access – can be run locally to grab files via the line mode client**

https://confluence.slac.stanford.edu/display/ds/Command+Line+Download+Manager

Getting at Current & Reprocessed Data

- LAT & FSSC astroservers already concatenate these datasets
- For P6_v1 from the data catalogue (eg for merit files), you have to make 2 selections to get it all. Documented here:

https://confluence.slac.stanford.edu/display/SCIGRPS/LAT+Dataset+Definitions

- Current version of the data is P6_public_v1
  - This dataset consists of the P105 reprocessing of data <271850289, and the LPA output of runs >=271850289.
- We assume Pass7 reprocessing will complete within the next month
  - Will update the definitions page

New in the Workbook

http://glast-ground.slac.stanford.edu/workbook/

- One-stop shopping for most documentation
  - Especially see the SCons doc!
- New since Paris:
  - SCons updates
  - Best practices for SLAC batch
  - GRB Analysis
  - New tips
- Coming Soon
  - big revision for GlastRelease developers
  - Science Tools and Analysis updates