LHCb experiment operations report

Elisa Lanciotti and Stefan Roiser

(IT-ES-VOS)

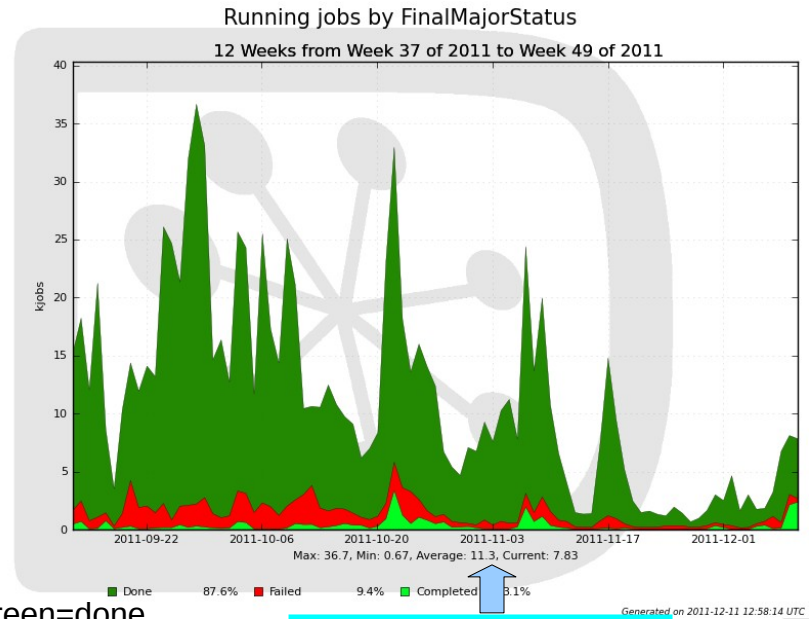
on behalf of LHCb grid Operations Team

GDB 14.12.2011

Overview

- LHCb operations during last quarter
  - Reprocessing campaign
- Next activities
  - MC campaign just started
  - Re-stripping before the re-start of data taking (~March)
- Summary of issues and tickets
- Summary

Operations at all sites



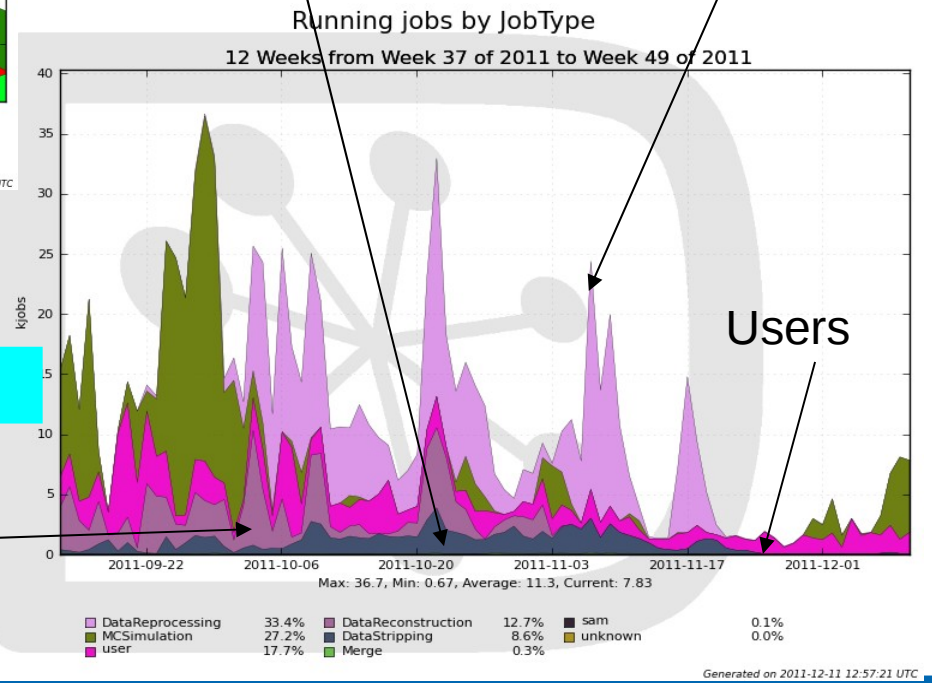
green=done
red=failed

By final status

Running jobs at all sites since September

Stripping

Reprocessing

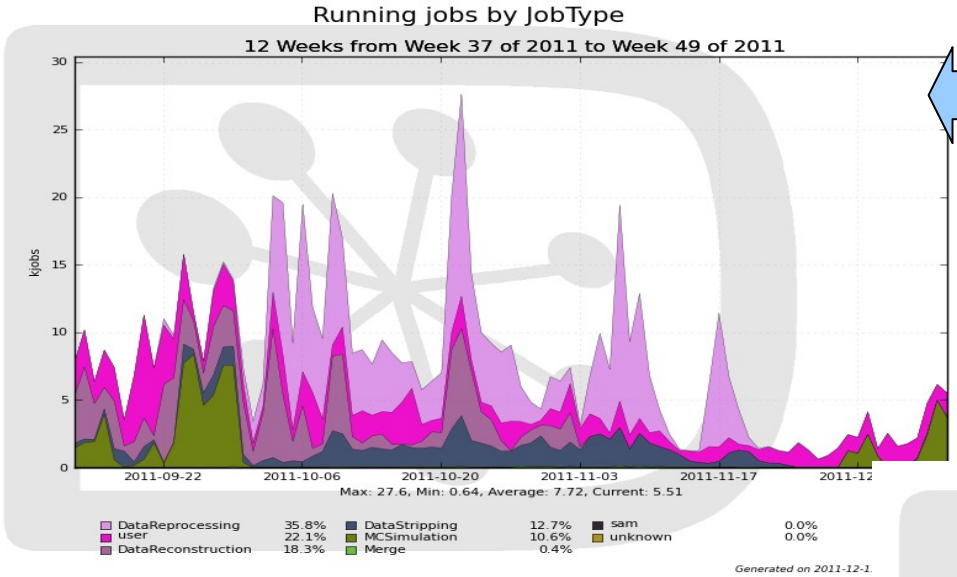


Users

By job type

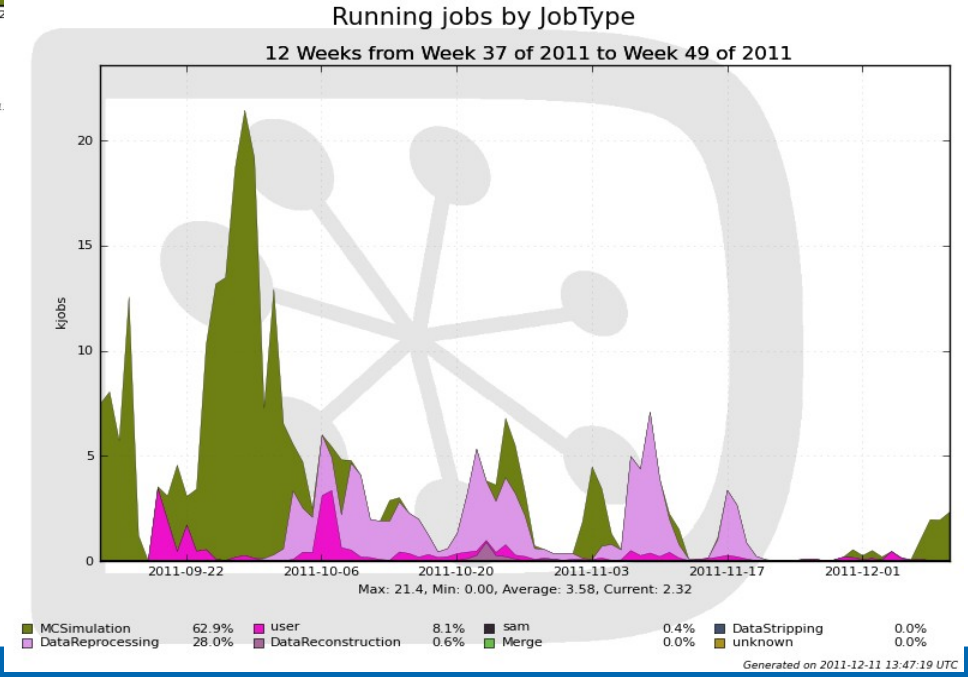
prompt reconstruction

Operations split by T0/T1 and T2



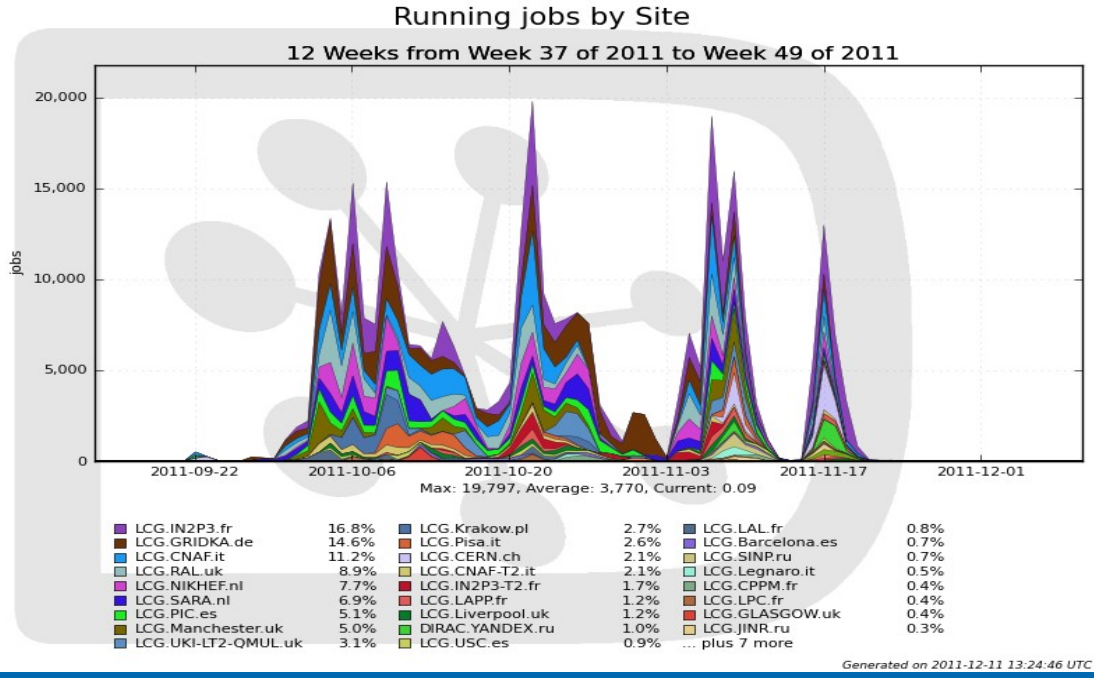
T0 and T1s:
Data reprocessing 36%
Users 22%
Data reconstruction 18%
Stripping 13%
MC 10%

T2s:
MC 63%
Data reprocessing 28%



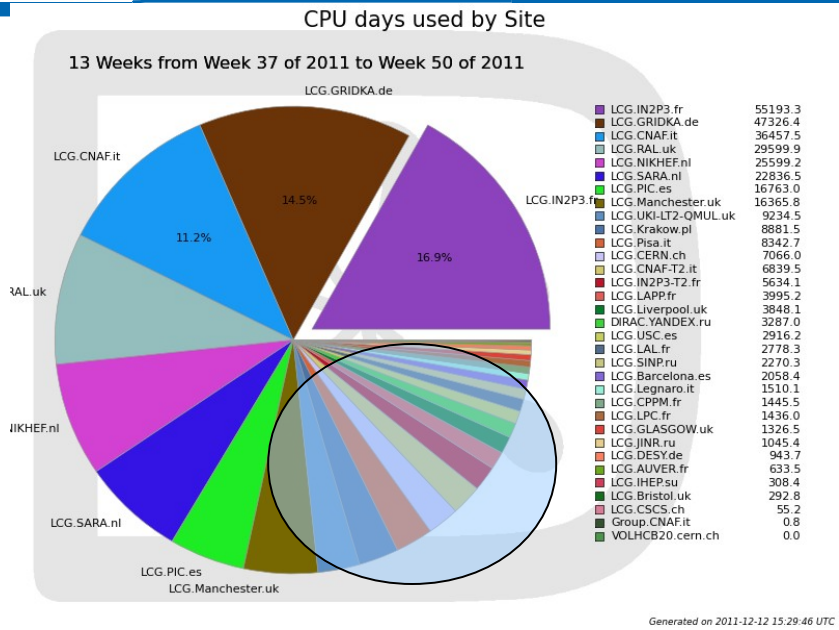
2011 reprocessing

- All 2011 reprocessed i.e. ~800 Tbytes of raw data. 1 fb-1 available for physics
- Started end of September and completed by ~November 20th (two weeks in advance wrt schedule!)
- Main change to computing model: reprocessing running also at Tier2s, 'attached' to some Tier1 storage
- Live reconstruction only at CERN during data taking
- Data stripping, as usual, only at T0/Tiers1



Only data reprocessing jobs

2011 reprocessing (II)



In total Tiers2 have run ~25% of reprocessing jobs. Main contribution from Manchester, QMUL, Krakow, Pisa, Cnaf-t2, In2p3-t2, LAPP, Liverpool, Yandex, USC, LAL...

- About 20 Tier2s selected for their proximity to some T1. The only requirement is to have good network connectivity and available CPU
- The T2 is 'attached' to one T1, from which it can download the input raw data file, run the reconstruction job and then upload to the same T1 the output (SDST) file
- The same T2 cannot then be attached to another T1..

2011 reprocessing: some issues

- Too low ratio of disk spindles/volume of data => disk servers could not sustain the throughput and the number of stripping jobs had to be throttled
  - At CERN mitigated by adding more disks in front of LHCb-Tape service class, which allowed to increase the number of concurrent stripping jobs
  - For next re-stripping, will be tested the possibility of downloading input files locally. Will be carefully tested before to ensure that sites can sustain the download rate
  - This will require more space on the WN disk: currently 10GB local space required in VO card @ CIC portal. But might be retuned to 20GB. Is this acceptable?
- During pre-staging, some problem due to the disk cache size:
  - it would have been useful to have a better monitoring of the cache size in front of tape and its occupancy/free space
  - Also useful to have a clear view of T1D0 pool architecture for dCache sites (put pool, cache pool and read pool separated? Is read possible from put pool?)

Conclusions of reprocessing

- Overall successful exercise
  - Finished 2 weeks in advance
  - 1 fb^{-1} available for physics
- Gave some ideas for the future:
  - More flexibility for production jobs assignment. Will require development effort
  - Stripping jobs: find a more efficient solution for accessing input files. Downloading locally to WN will be tested

Next activities

- Just started MC campaign at the beginning of December
  - Will run mainly at T2s
  - Will be the main activity until re-start of data taking. Overall duration of order of months, probably until late spring
- An additional stripping of data foreseen before the re-start of data taking (March), not to interfere with prompt reconstruction
- Ideas being evaluated:
  - Using the Online Event filter farm (EFF) during first months next year for MC jobs

Data management

- Total storage needed for 2011 reprocessing: 1.7 PB (900 TB tape, 650 TB disk). DSTs, have 4 replicas on disk + 2 archive copies
- Before the next MC campaign, it was necessary to recover some space from Tiers1:
  - Disk space: removed from disk some MC/2010 and reduced the number of replicas from previous pass of stripping (stripping16): gained about 500TB
  - Tape: removed SDSTs of previous reconstructions: gained about 600TB.
- Procedure to identify old data candidate to be removed is totally manual. Envisaged the implementation of a popularity service that will contribute to streamline the procedure
- After data removal, dCache sites have to be asked via GGUS ticket to re-allocate the free space to the new space tokens. A consequence of the migration to a new space token schema last May. It would be highly beneficial if the data in the old space tokens could be migrated to the new ones. Opened a ticket (76561) to ask if it's feasible

Data management (II)

- Current disk space situation at sites for the LHCb-Disk space token (TB):
  - **Important**: this is the space allocated in LHCb-Disk, which does not include space allocated in the 'old' space tokens for dCache sites (see previous slide)

site	CERN	CNAF	GRIDKA	IN2P3	PIC	RAL	SARA
total	1160	540	430	633	288	836	372
free	77	143	170	360	63	365	123

Total free disk space 1300 TB, totally allocated 4.26 PB

Space needed:

- Re-stripping will process all 2011 data, but will require much less space than the one just completed, Stripping17 (disk: 655TB, Tape: 320TB), as only some stripping lines will be produced.

  ⇨ Rough estimation: disk <100 TB, tape <50 TB

- MC 2011: not easy to foresee the required space, a rough estimate is the size of MC 2010, 1.3PB, with 3 replicas on disk.

  ⇨ Will be produced in stages, at an approximate rate of 200 TB/month

Data management middleware

- **FTS**:

  Problem observed when some transfer fails: transfer cannot be retried because the destination SURL cannot be overwritten. In these case, transfer requests are stuck, and the destination SURL has to be removed manually

  LHCb would need FTS overwrite functionality, implemented in FTS 2.2.8. This is currently deployed as pilot version at CERN and being tested by ATLAS and CMS

- **GFAL**:
  - In touch with developers to give feedback about needed functionality for GFAL2. In principle, all GFAL functionality should be available in GFAL2
  - Rely on GFAL and on its support. At least until GFAL2 will provide equivalent functionality

SLC6

- LHCb not ready for the migration to SLC6
- Core software:
  - LCG-AA not ready yet to support it. Should be working on it at the start of next year
  - Extended platform defined as SLC6, Python2.7, gcc 4.6.2 (is this confirmed?)
- WLCG middleware:
  - WLCG middleware: release for SLC6 foreseen by May
- Dirac framework
  - As long as the WLCG middleware is ready, no other problem

Issues and GGUS tickets

- During last quarter, opened in total 145 tickets, of them 4 alarm tickets, 118 team tickets
- Alarm tickets:
  - 3 concerning SRM down: GRIDKA (75261), IN2P3 (75610), SARA(76629), all timely fixed by sites
  - 11th Dec: an LHCb vo-box at CERN CC not reachable, had to be rebooted (77246). Problem of latest Dirac release, fixed on Mon 12th
- Team tickets:
  - 25 for accessing VO software (more details later)
  - 29 for pilots aborted, at several sites, including SARA (CREAM configuration), NIKHEF (batch system configuration), RAL (CA and host certificate mismatch), CNAF (CE hardware problem), IN2P3 (CREAM configuration), GRIDKA (CREAM problem), CERN (CE overloaded by huge backlog of requests
  - 12 for SRM related issues

Focus on tickets for problems accessing VO software

Team tickets: 25 for accessing VO software:

- Shared area, usually NFS/AFS problems, or quota exceeded: IN2P3 (74334, in Sept before moving to CernVM-FS), GRIDKA (74450), FERRARA (76488), BIFI-IBERGRID (76188), LCG.UFRJ.br, UKI-SOUTHGRID-RALPP (then moved to CernVM-FS), TORINO, UKI-SCOTGRID-DURHAM (76487), JINR-LCG2 (77223), AUVERGRID (76586), WCSS64 (76569) and other Tiers2
- Also some issues with CernVM-FS, usually some misconfiguration that could be quickly fixed: IN2P3 (76022,76515), UKI-LT2-RHUL (76189,77121) , UKI-NORTHGRID-MAN-HEP (75861,76190), DORTMUND (76881), etc..
- Sites are welcome to move to CernVM-FS

CernVM-FS

- Since last report (Sept. GDB) deployed also at CNAF and IN2P3
- Only Tier1 maintaining a shared area is GRIDKA

In general, LHCb has a positive experience with CernVM-FS, sites are encouraged to move to this service.

In addition to the official documentation, some practical instructions for sites that want to move to CernVM-FS (best practice for deployment etc.) are available thanks to UK sites

https://www.gridpp.ac.uk/wiki/UK_CVMFS_Deployment

Summary

- Main activity during last quarter was the 2011 reprocessing: successfully completed
- SLC6:
  - LHCb not ready for the upgrade
- CernVM-FS:
  - deployed at most Tiers1 and many Tier2s
- Next activities:
  - Huge MC campaign just started
  - Additional stripping of 2011 data early next year