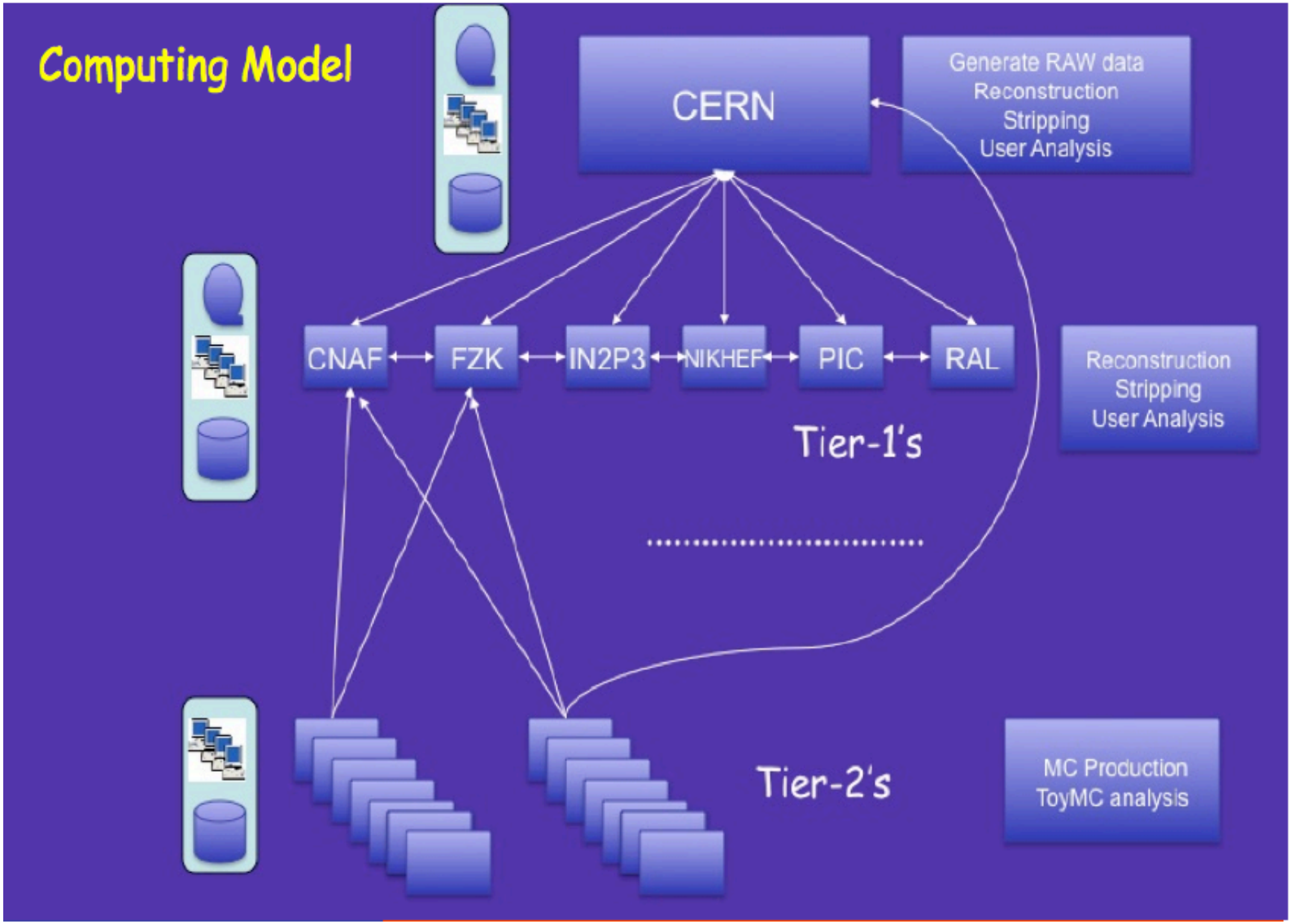
LHCb computing

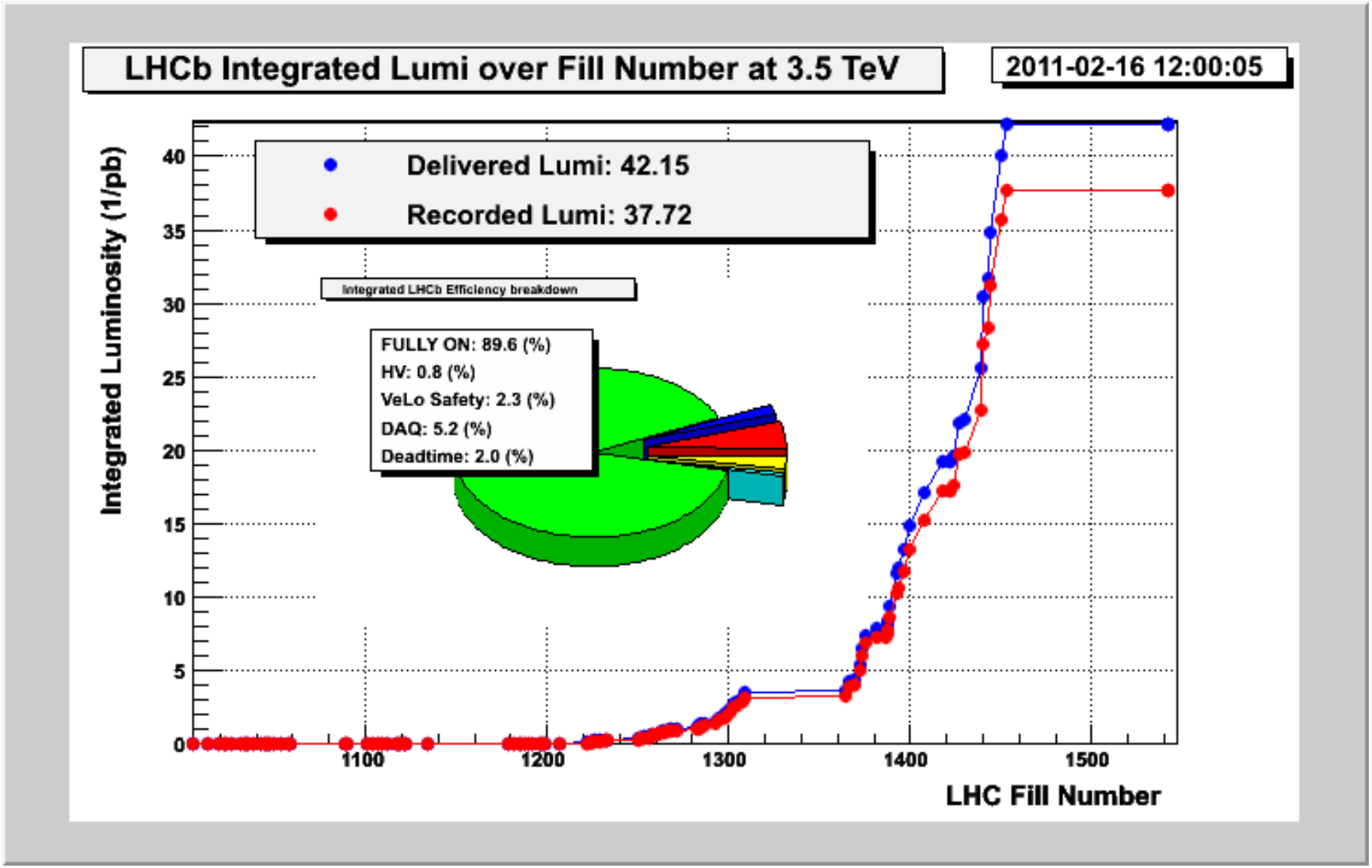
A. Carbone

INFN-Bologna

Computing model



Integrated luminosity



Computing activities

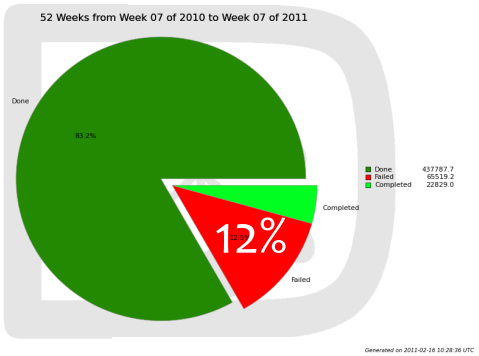
- Production activities
  - Simulation, reconstruction, stripping, WG analysis (µDST)
  - Use DIRAC and the LHCb Production System
- User analysis
  - Data and MC analysis
  - For testing, use local resources (including local batch system)
  - For large datasets, use Grid Computing
- Toy MC and fits
  - Use Grid Computing for large samples
- Non-Grid user analysis
  - Mostly interactive analysis on local clusters (Tier-3), desk/lap-top (Tier-4)

Data flow

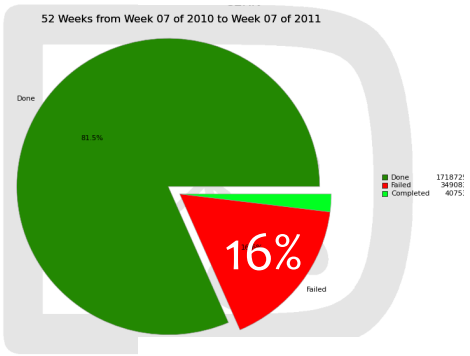
- As soon as the raw data are recorded by the detector they are sent to Tier-1's+CERN-CAF to be reconstructed
- After the reconstruction, the Stripping process is performed
  - pre-selections provided by analysis working groups are applied in order to
    - write on disk different streams where each contains similar selected events, such as B-hadron, V0 decays, charm decays, etc…
    - reduce the data sample to a handy size in order to perform a finely tuned analysis
- The stripping can be performed several times on the same datasets (according to the availability of resources) if pre-selection algorithms change.

Tier-1 site efficiency – Feb 2010- Feb. 2011

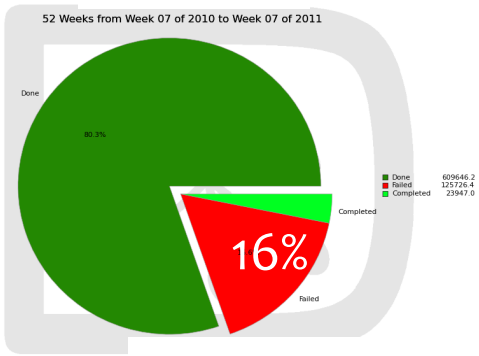
CNAF



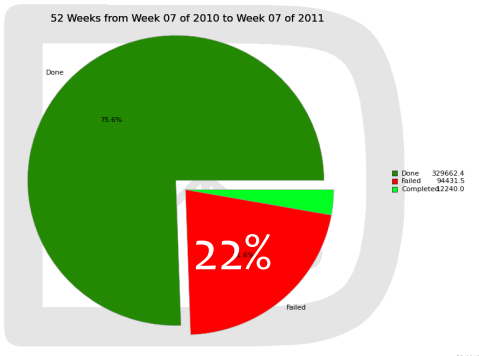
CERN



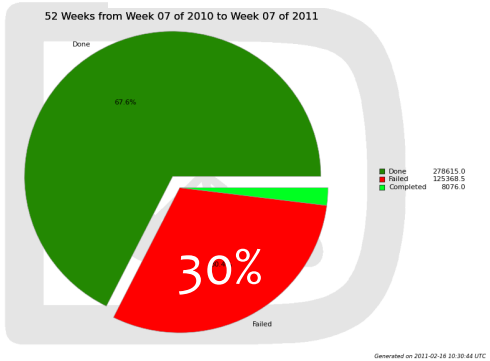
GRIDKA



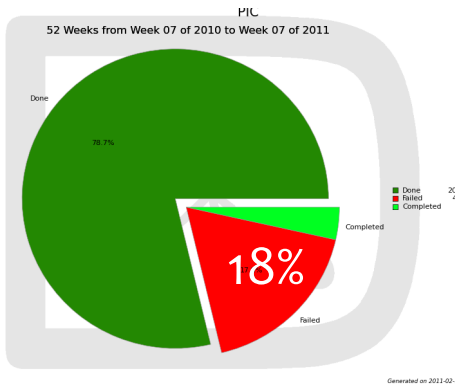
IN2P3



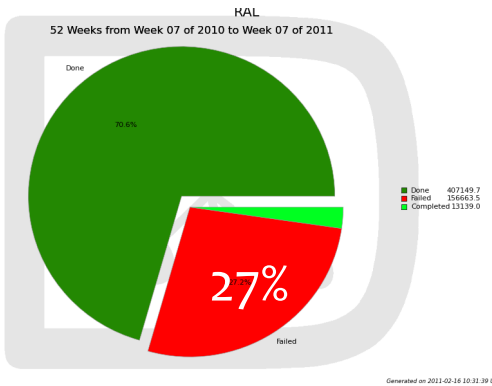
NIKHEF



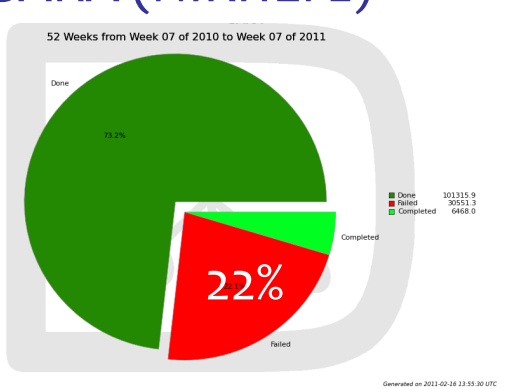
PIC



RAL



SARA (NIKHEF2)

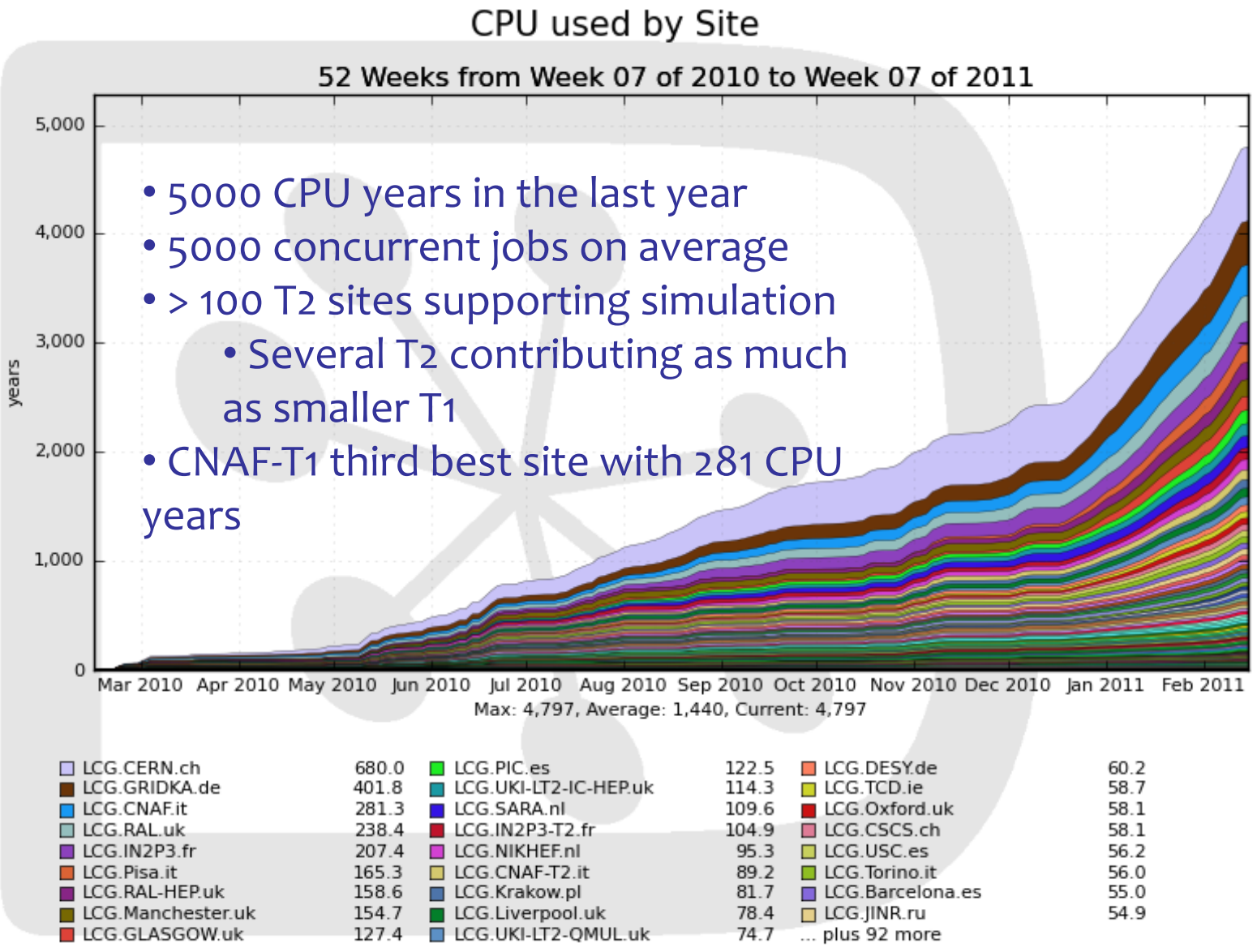


CNAF best Tier-1, failure rate 12%.

NIKHEF, SARA, RAL e IN2P3 show a failure rate higher than 20%

NB: the user failures are included, but they are uniformly distributed amongst the sites

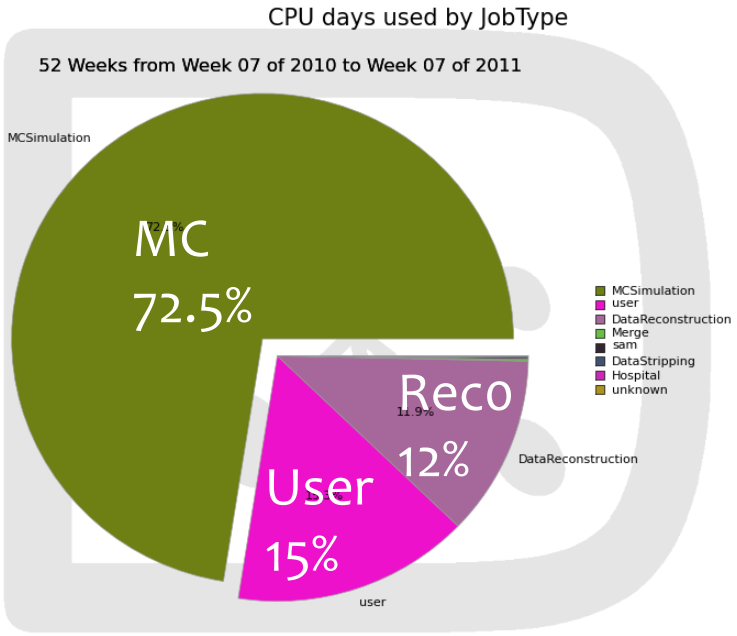
Cumulative CPU usage by site: last year

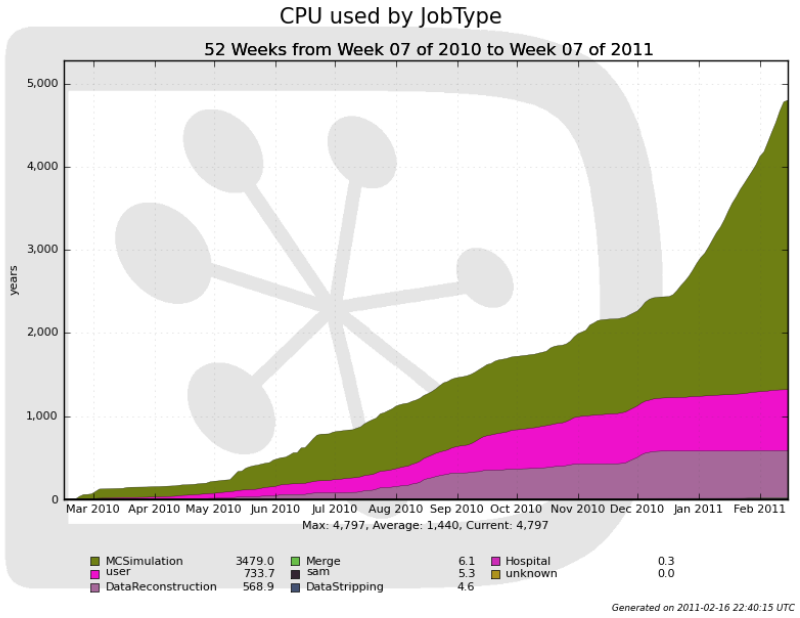


- 5000 CPU years in the last year
- 5000 concurrent jobs on average
- > 100 T2 sites supporting simulation
  - Several T2 contributing as much as smaller T1
- CNAF-T1 third best site with 281 CPU years

Usage of Tier-1 resources

Feb 2010-Feb. 2011





Generated on 2011-02-16 22:39:19 UTC

MC

User

Recostruction

Analysis

- Batch analysis
  - Batch processing runs on the Grid at the Tier-1's
    - Submission to DIRAC from Ganga
  - Output data uploaded to the Tier-1 storage
    - MicroDST or Ntuples
  - Dedicated USER area at all sites
    - about 450 TB in total are kept permanently on disk
- Interactive analysis
  - Interactive clusters (Tier-3), desk/lap-tops (Tier-4)
  - These are NOT meant as part of the Grid
  - Data transfers up to users, as well as bookkeeping

Ganga

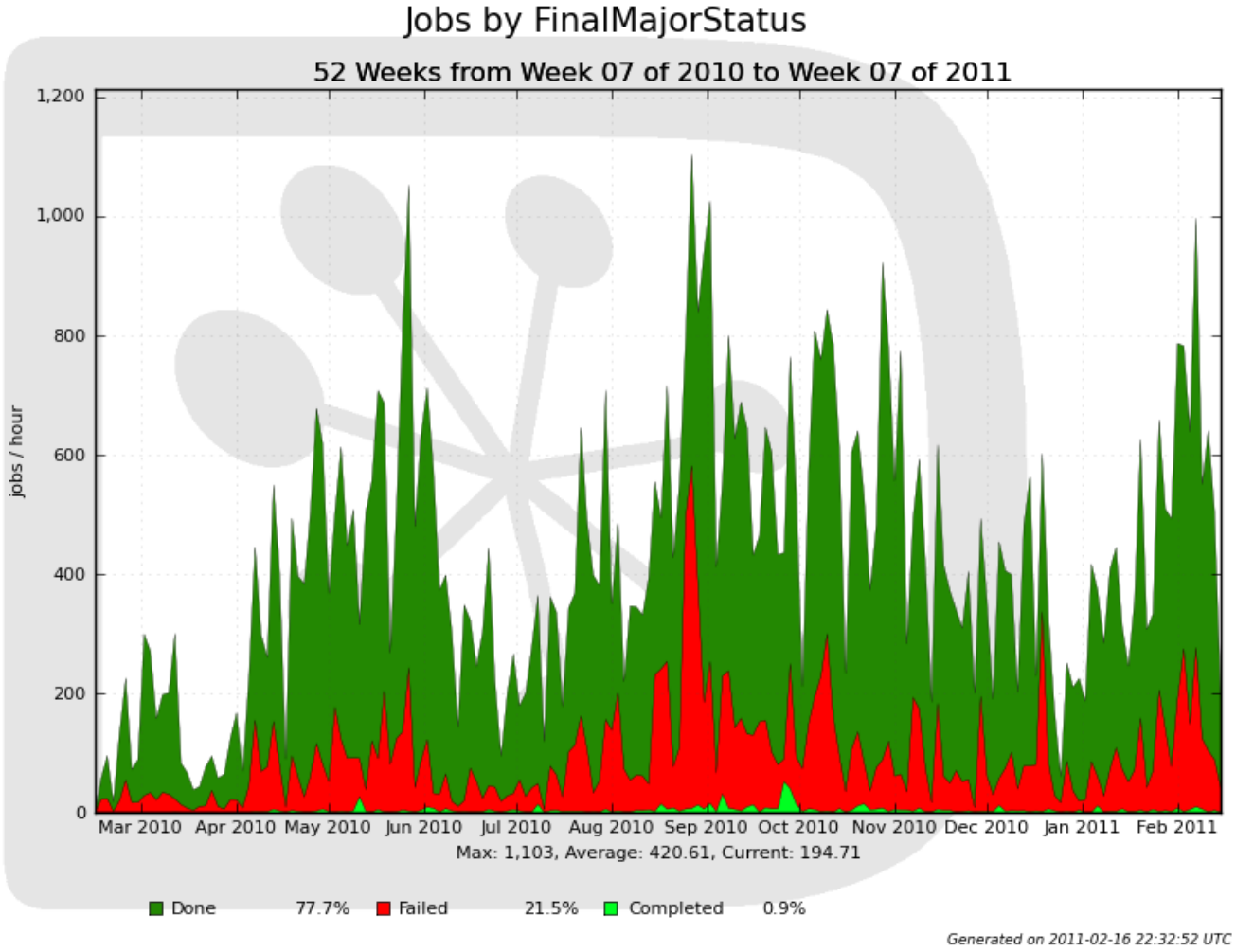
- The batch analysis in LHCb is handled by Ganga
- Ganga is:
  - an Atlas/LHCb joint project
  - an application allowing a user to perform the complete life cycle of a job
  - Build – Configure – Prepare – Monitor – Submit – Merge – Plot
  - Run on the local machine (interactive or in background), submit to batch systems (LSF, PBS, SGE, Condor) or Grid systems (LCG, gLite, NorduGrid
  - Jobs look the same whether they run locally or on the Grid

Resource accounting and analysis priorities

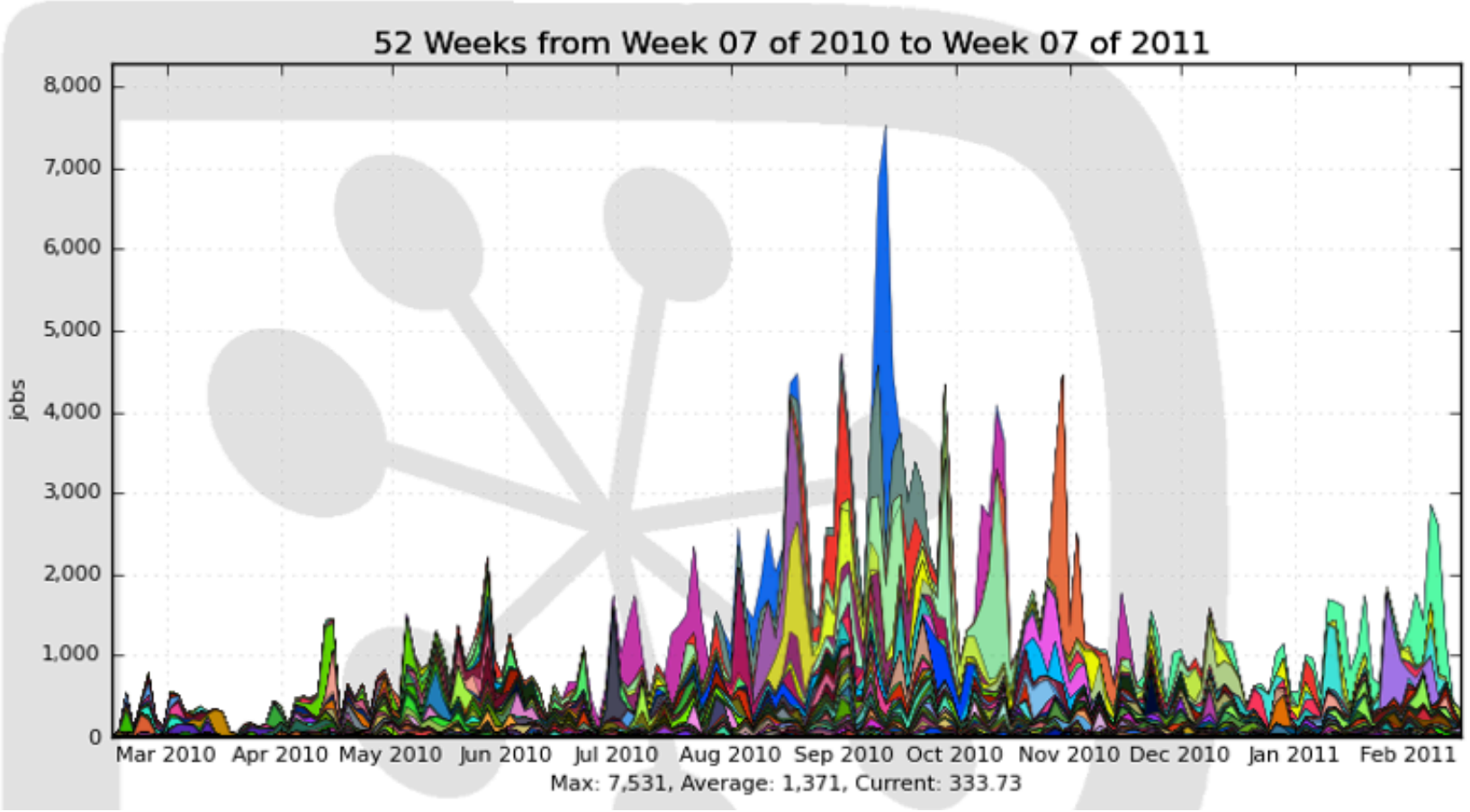
- Priorities
  - Assigned for groups of users
    - Groups are defined by the PPG (Physics Planning Group)
    - Currently a single generic group exists: lhcb_user
  - Group priorities will be defined by the PPG
- Accounting
  - Grid CPU usage is accounted for by DIRAC
    - Group accounting, user accounting
  - No CPU quota, but regular checks for excessive usage
    - Individual priority may be affected by heavy usage
  - Storage accounting for user spaces
    - User quotas managed by DIRAC
    - No hard limit for the moment
    - Mail sent to the user when approaching the quota
      - Asking for clean up of old datasets

User jobs, last year

- 78.5% success rate
  - ~50% of failures due to user mistakes



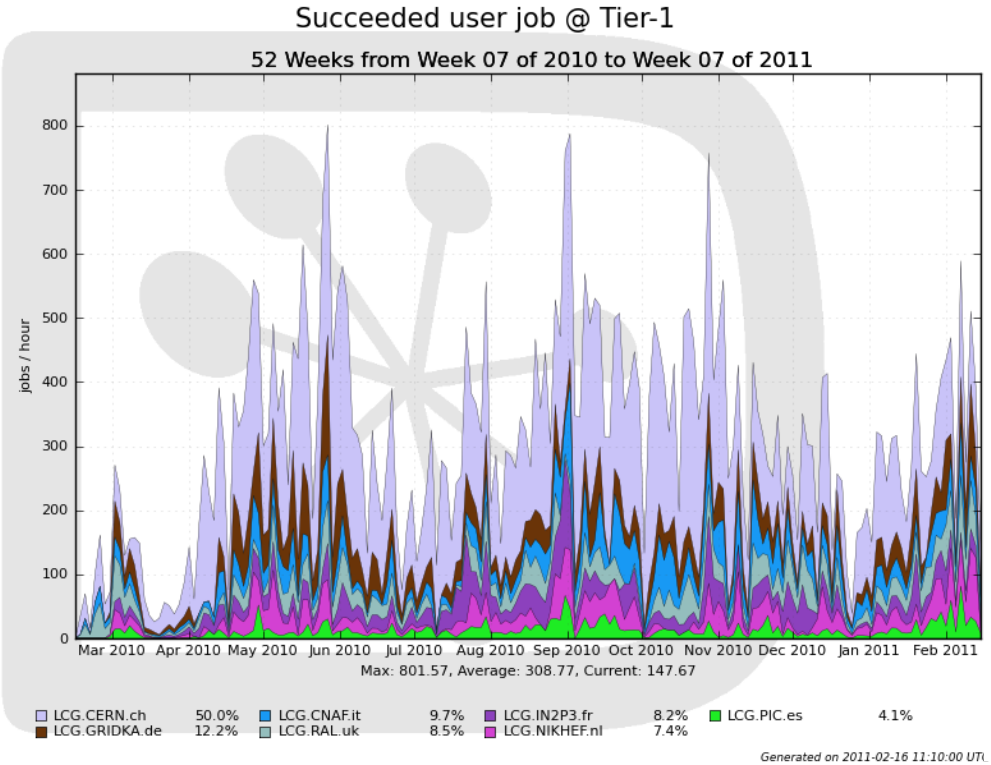
User jobs last year by user

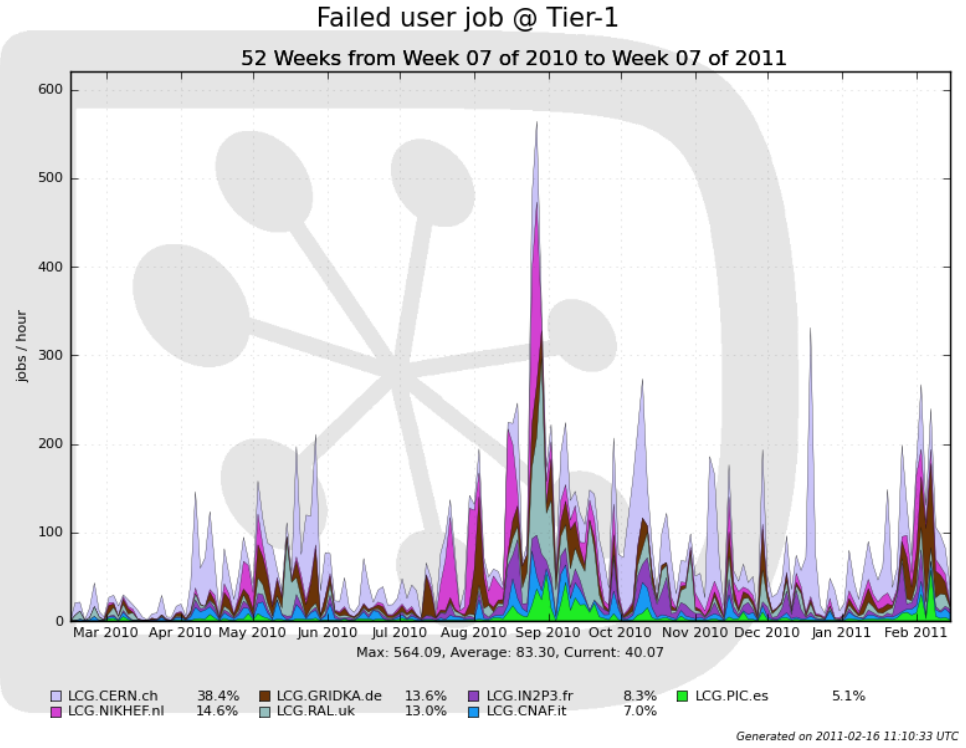


~ 320 unique users

User jobs by site (Tier-1)

Feb 2010-Feb. 2011





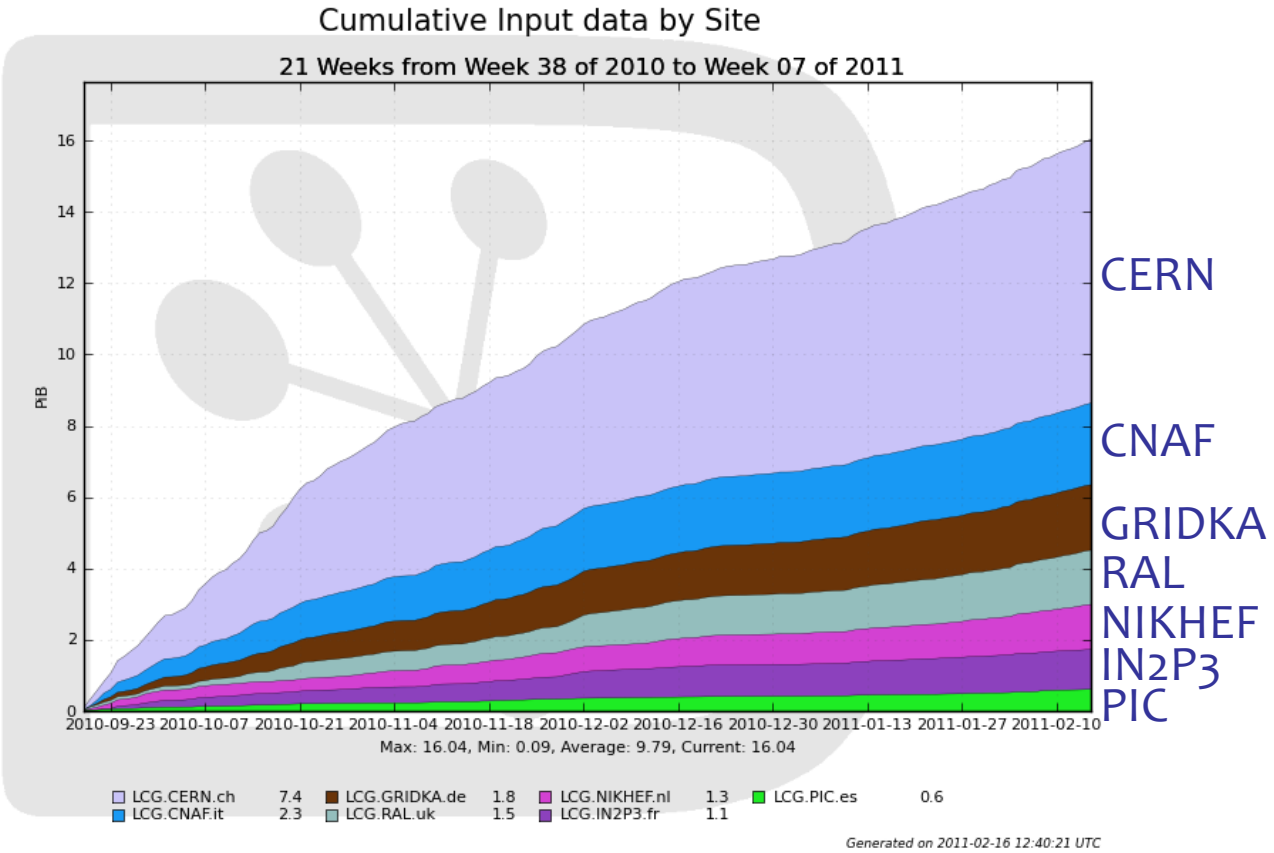
10% of succeeded analysis user jobs executed at CNAF

low rate of failures at CNAF with respect to the other sites

50% of the analysis performed outside CERN

Input data user jobs

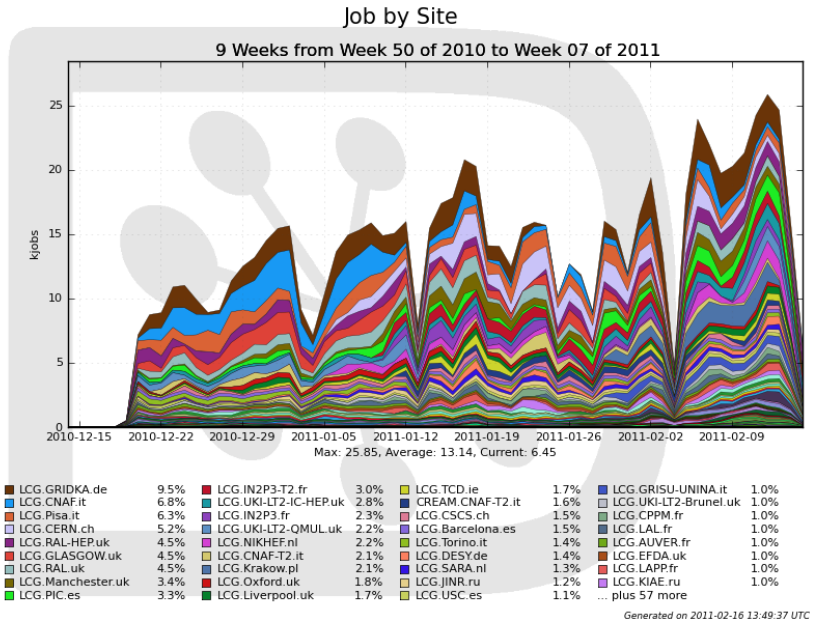
Sept. 2010-Feb. 2011

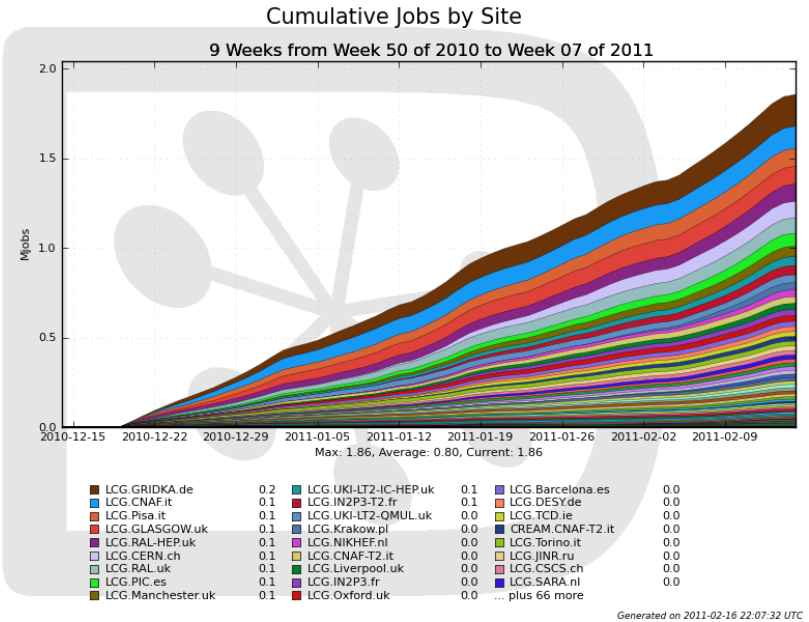


2.3 PiB of data read at CNAF since September 2010. Best site after CERN

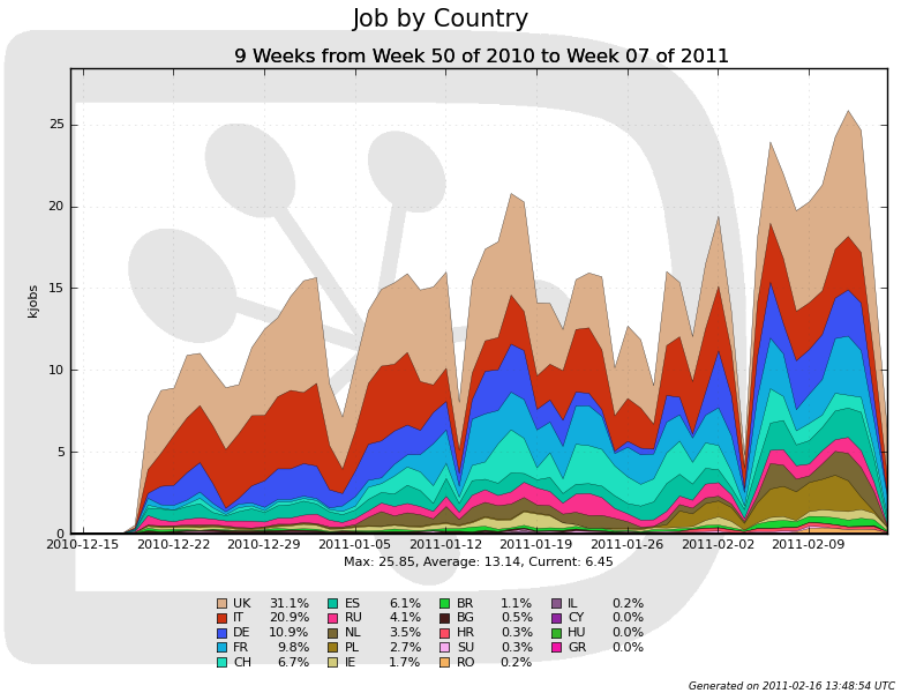
MC production

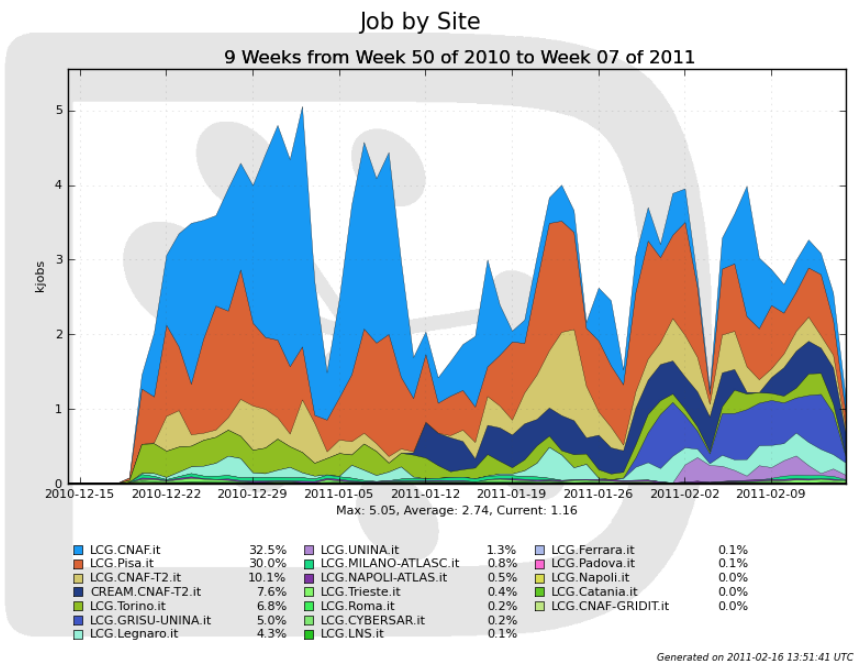
- Started a massive MC production since mid-December in preparation for the winter conferences
  - completed 2M of jobs, corresponding to ~10^9 events





MC production



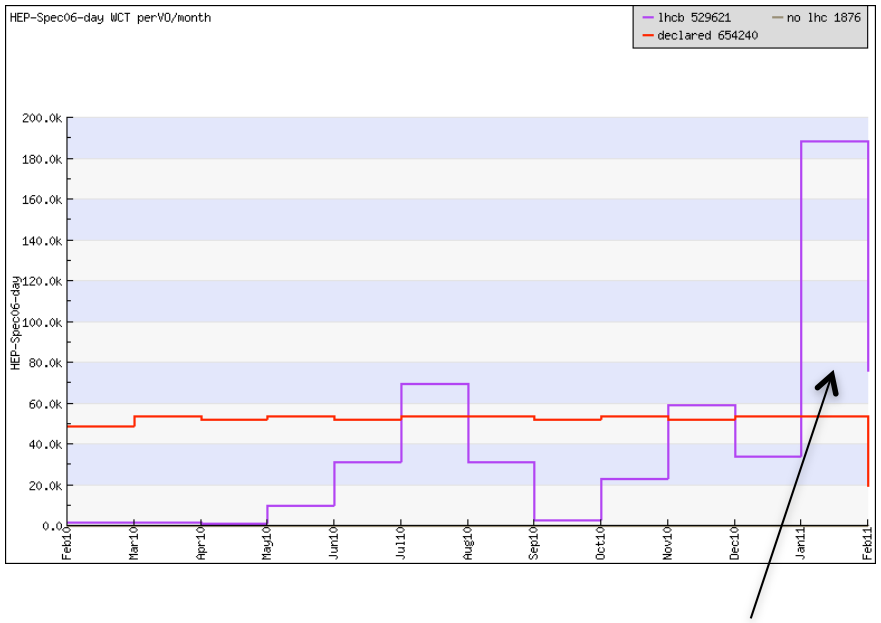


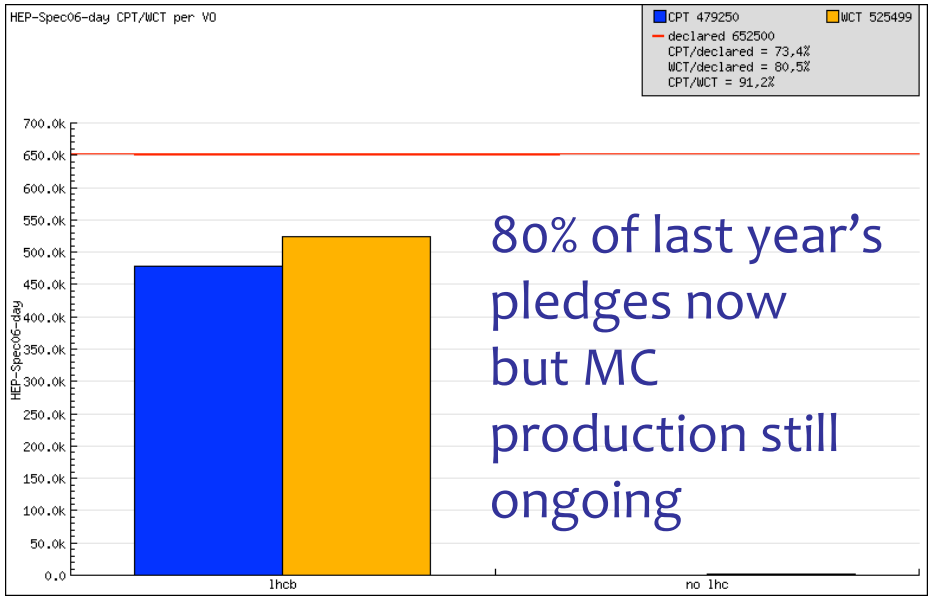
20% of the whole MC production realized in Italy
Second contribution after UK

50% of the Italian fraction done at CNAF T1 (33%)+T2 (17%)
Other relevant contributions from Pisa, Torino, Legnaro

Italian LHCb Tier-2 @ CNAF

Thanks to the MC mass production for the winter conferences the LHCb Tier-2 at CNAF has dramatically increased the average CPU usage





80% of last year's pledges now but MC production still ongoing

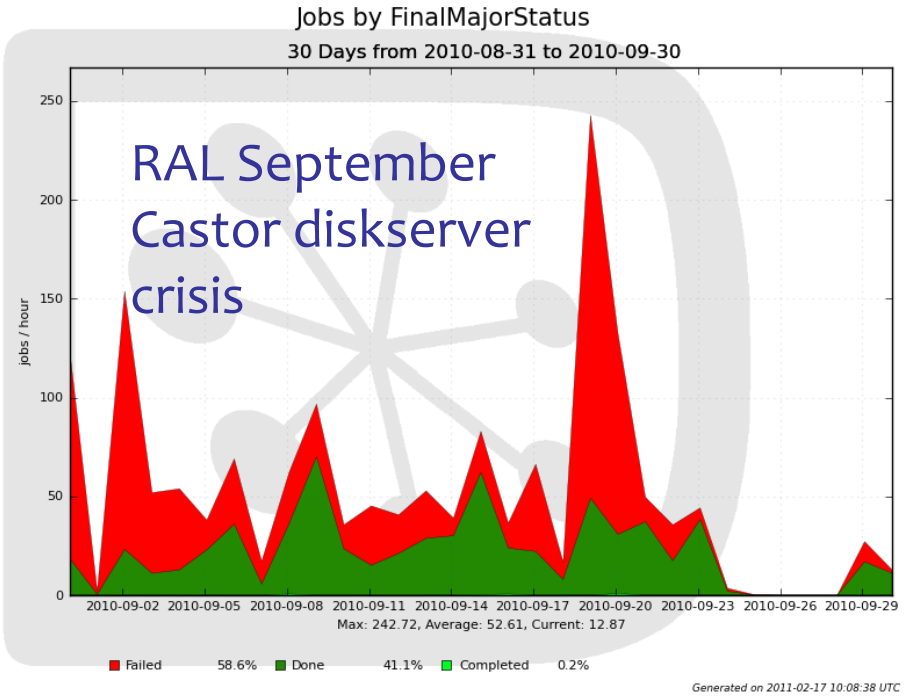
50% of last year's usage made in one month employing Tier-1 free slots during (post-)christmas break

Main operational problems at Tier-1 centres

- Area software availability and responsiveness
  - Showstopper at IN2P3 (AFS)
  - Problematic at all the other Tier-1's
  - Not a major problem at CNAF
    - Clustered NFS on top of GPFS, unique solution of this kind
    - Some rare instabilities however have been observed, but usually solved with little configuration changes
  - CVMFS solution being investigated in all major centres (also at CNAF)

Main operational problems at Tier-1 centres

- Storage system instabilities
  - Crisis after summer due to a rapid increase of user analysis jobs, as soon as real data started to be available



CNAF had no major problems!!!

GPFS diskservers sustained a 100% load at full available bandwidth

By far the most stable storage system in all LHCb Tier-1 centres

Conclusions

- This first year of data taking was not easy for the LHCb computing group
  - Most of the infrastructural problems came from instabilities of the storage systems at Tier-1's, either used for data, software or even condition databases
- CNAF did not have particular problems
  - Some rare instabilities on the software area
    - a couple of times during last year
  - Some hardware failures
    - core network switch
  - Some failures on Oracle servers
    - due to operating system issues → network driver update solved
  - Some LSF instabilities when scaling to thousands of virtual/real worker nodes
  - In contrast to all the other Tier-1 centres, no instabilities related to storage!