Bologna
INFN
Istituto Nazionale di Fisica Nucleare

Status of CMS activities in *SC3-phase2* at the INFN-CNAF Tier-1

D. Bonacorsi
(INFN-CNAF Tier-1 and CMS experiment)

GDB
Oct 11th-12th, 2005 - CNAF, Bologna

T0→ INFN T1 transfers (within SC3)

- CNAF T1
  - SRM/Castor via srmcp (whenever possible - see next slide)
    - in throughput phase we developed the ability to switch fast to…:
      - to any different base transfer-mechanism (srmcp, g-u-c, lcg-cp, …) if needed
      - to different infrastructure (e.g. 'production' one instead of 'SC3' one)
    - hope now to exercise and challenge the 'Service' - and not the aforementioned skills - in the official SC3 infrastructure
  - Transfers triggered by PhEDEx
    - *SC3 instance* coexisting with *Prod instance*
    - full PhEDEx set-up at INFN T1, Castor MSS part also deployed
    - LCG local file-catalogue (POOL MySQL)
    - data publishing in PubDB v4.0.4 (*SC3 instance*)
  - Work going on in parallel:
    - FTS installed with LCG260 → PhEDEx/FTS integration in progress
- ~ 4.3 TB of data so far (~1/2 of which in just one day - see next slide)
  - publishing procedure and job submission steps in progress…

Before actual 'Service Challenge'…

… one first needs to help in debugging what's crucial to work at the T0

- CMS@Tier-1 offered its collaboration to the LCG-SC community
  - since mid-Sep05: debugging + daily reports from CMS@T1 to several actors:
    - CNAF core staff, J.Casey, O.Barring, B.Couturier, E.Martelli, A.Hirstius, V.Bahyl, .. (+ SRM experts, Castor experts, PhEDEx experts, …): thanks to all.

- Why? Because operations affected by:
  - crashes and downtimes of Castor-2 Oracle DB instance at CERN:
    - 21 Sep 19:27, 22 Sep 03:20, 22 Sep 15:00, 24 Sep 02:15: all properly addressed
  - 27 Sep: SRM bug found, major Castor-2 upgrade, fixed rpms deployed 28 Sep
  - 28 Sep: most gridftp transfers failed from castorgridsc
    - Diagnosis: transfer requests hangs/time-outs due to net config change on ia32 nodes
  - 3 Oct: problem partly understood
    - Diagnosis: "network interface misconfiguration": actions brought 40% of CERN disk-servers back into a usable state, others disabled to grant immediate&safe ops restart
  - 4 Oct: transfer requests stuck from castorgridsc
    - Cause: internal state errors when copying files among nodes, i.e. copy ok but update of internal state failed. Corrected by hand, fast fix tested and roll-out transparently
  - 6 Oct: disk failure on the main Castor DB server holding the Castor namespace and tape volume repo. Hot-swappable, replaced

Note: all these are not PhEDEx-related issues: also seen from c.l. srmcp's, g-u-c's, …

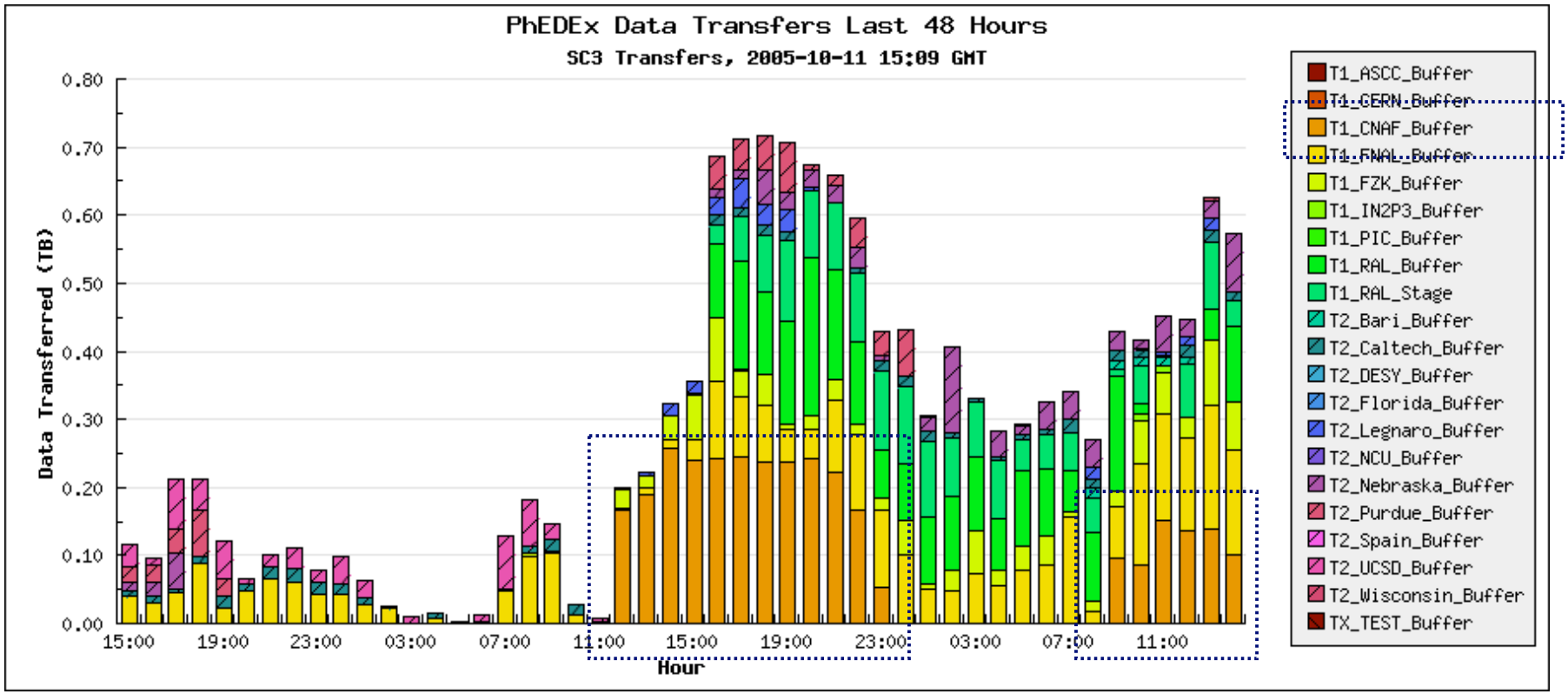
T0→ INFN T1 transfers (within SC3) [2/2]

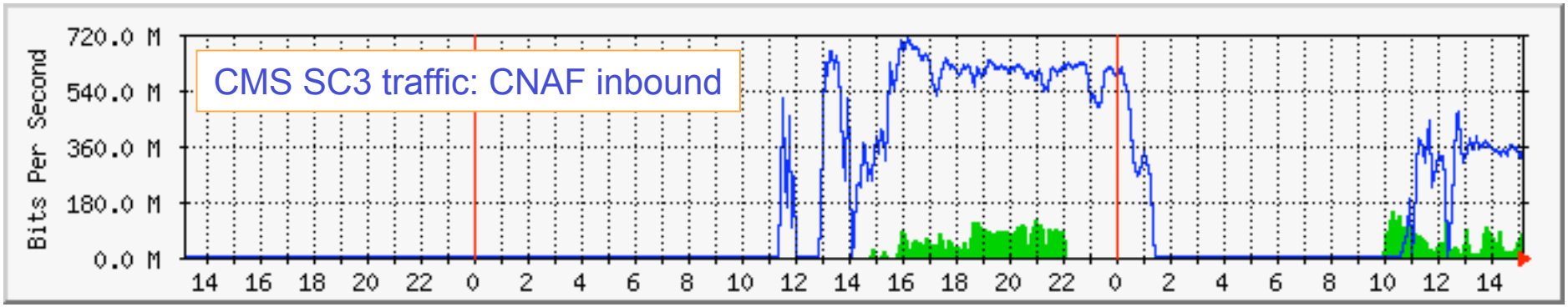
- 6 Oct: another SRM bug at CERN (affecting 10% of the requests). Fix+roll-out
  - to remove SRM from the problem picture, I decided to step back from srm and debug pure g-u-c transfers from castorgridsc
  - turned out to be the most relevant test: g-u-c tests revealed 100% failures in CMS transfers to Tier-1.. it worths more in-depth testing… which yielded to:
- 7 Oct: Problem understood
  - Diagnosis: access list and routing needed mods: machine used to trigger third-party transfers was not within LHCOPN
    - it was a known fact since June-July, but functionality was there in the past, also in SC3-phase1, and in second-half Sep05, and stopped working sometime ~ begin.October..
  - similar problems experienced by RAL T1 and others in the same days

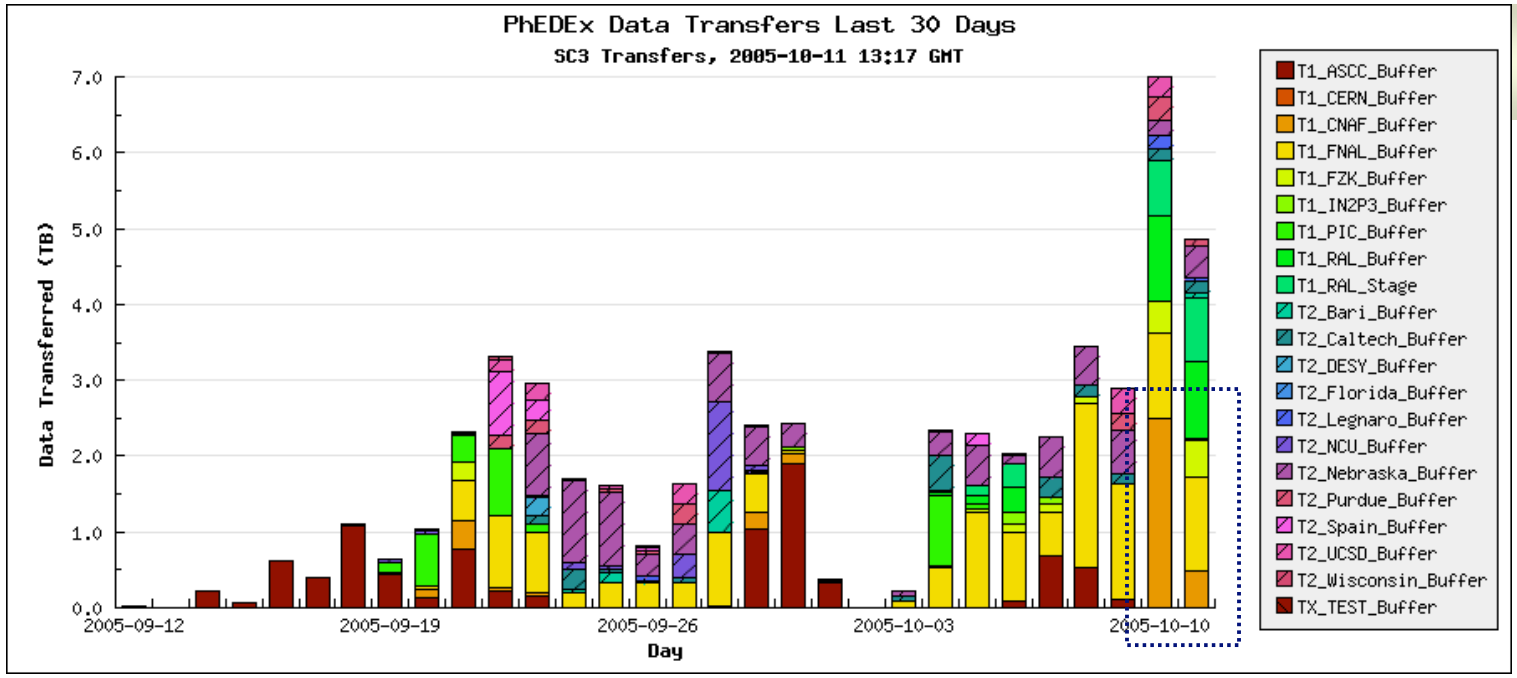
Also: INFN T1 → INFN T2s transfers

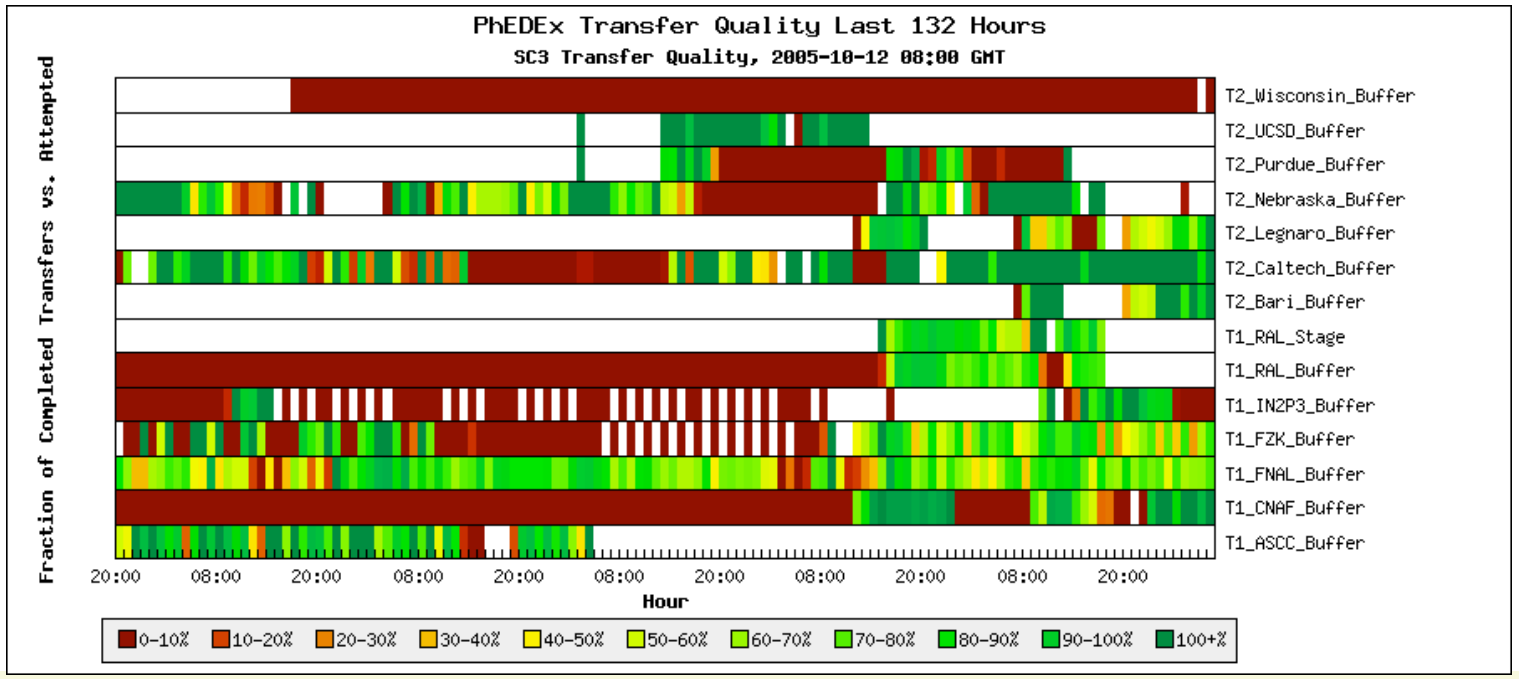
- Past: affected by castorgridsc not efficiently delivering data to INFN T1
  - … and by T1 helping on debugging transfer failures
- Now: main problem found its diagnosis, and a first solution
  - … roll-out, and INFN T1 is catching up quite fast
    - smooth transfers from castorgridsc to CNAF Castor staging area on SC3 infrastructure
      - ~4.3 TB transferred since beginning, *~2.5 TB in 12h (>~70 MB/s sustained)*
    - data migration to Castor MSS running
      - ~1.3 TB already on tape, *now optimising Castor set-up in real working conditions*

T0→ INFN T1 CMS transfer rates









all areas should look predominantly dark green,,.

Current status of INFN CMS T2s in SC3

- Bari T2
  - SRM/dCache v1.6.5-2, srmcp transfers
    - 1 admin node + 1 pool node with 3.3 TB (can go up to ~5 TB)
  - PhEDEx
    - *SC3(/Prod) instance*, LCG local file catalogue (POOL-LFC-MySQL) - PubDB v4.0.4 (*SC3 instance*) - CNAF-Bari bandwidth saturated with PhEDEx transfers
  - PhEDEX/LFC-v.1.3.7 integration done and operational in SC3 now
  - ~700 GB so far
    - publishing and job submission tests in progress

- Legnaro T2
  - SRM/DPM v1.3.7 (in LCG v2.6.0)
    - 1 host as SRM/DPM/DPNS server (dual PIII, 1,3 GHz, 1GB RAM) - 1 ds with 2 arrys (2 TB each) in the DPM pool (dual Xeon 2.8 GHz, 4GB RAM, 2 ctrl 3-ware 12 HD 250GB RAID-5) - DPM client on UI/WNs - another ds not yet online, waiting for disk capacity estimated needs in SC3-phase2
  - PhEDEx
    - *SC3(/Prod) instance*, LCG local file catalogue (POOL-MySQL) - PubDB v4.0.4 (*SC3 instance*), needs dataset completion to publish
  - ~300 GB so far
    - forced to a slower ramp because of ongoing debugging on access to DPM-hosted data from CMS apps (rfio-dpm and rfio-castor compatibility issue)

Beyond data transfer..

Service exercise is not only data distribution…

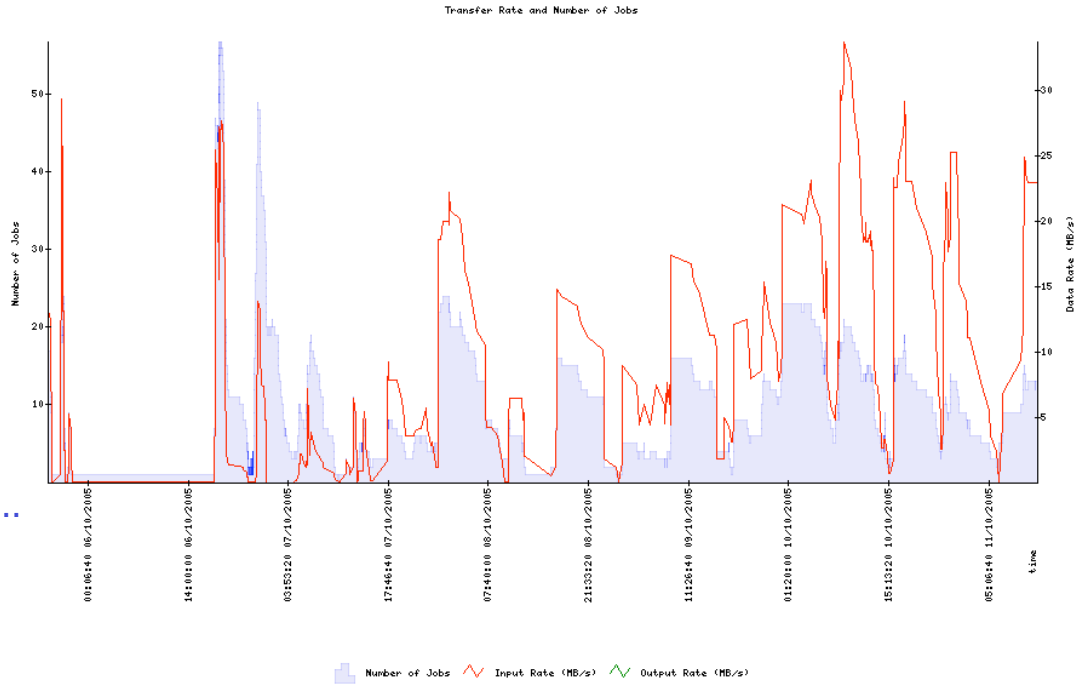
- simultaneous data import/export/access at T1/T2s

❖ primary needs for INFN CMS Tiers (my rough list.. please refer to CMS SC3 talks by L.Tuura):

- CMS sw installation at sites CEs (coherent sw env accessible from WNs)
- data distribution (reliable transfers of complete datasets) - MSS also (at the T1)
- data publishing, job creation/submission/output-retrieval
- monitoring

All this started:

- safe transfers from the T0 will feed CMS INFN SC3 sites with data to proceed with next steps…



❖ secondary needs: optional: … please address primary needs first! …

- not all the above are sequential, but ALL are needed