PIC
port d'informació
científica

# STEP'09 Tier-1 Centres Report

WLCG STEP'09 Post-Mortem Workshop

CERN, 9 July 2009

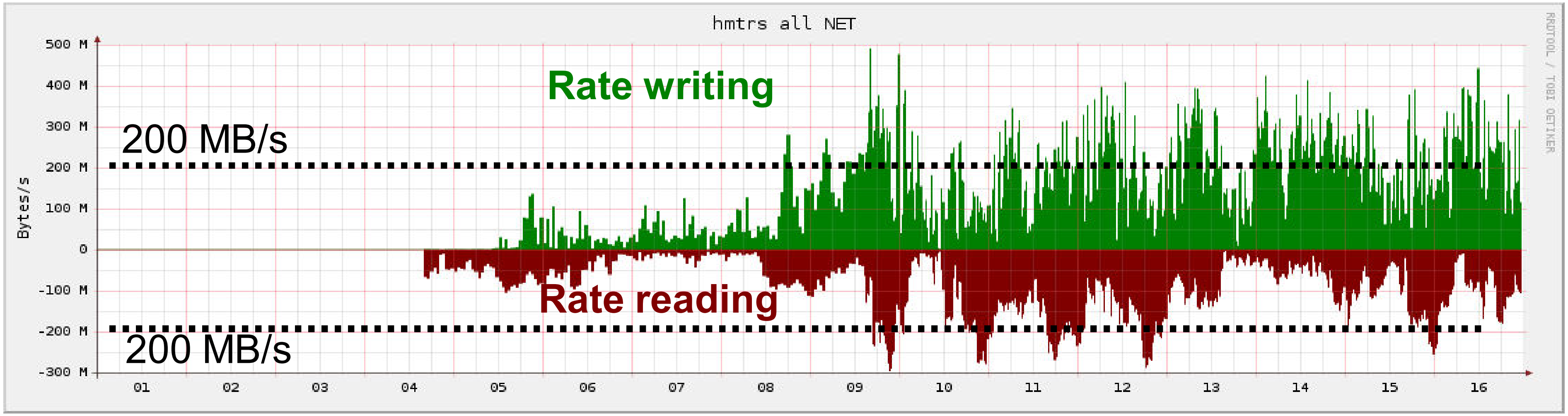
G. Merino, PIC

# General comments

- **NDGF:** All systems performed as designed, many reaching 99% eff.
- **RAL:** Very smooth operation. Calm atmosphere. Low load on most services.
- **CNAF:** No relevant issues. General stability for all the period.
- **NL-T1:** dCache performed OK and stable, delivering 1 GB/s sustained.
  - MSS performance issues, known before STEP09.
- **BNL:** All Tier-1 Services stable over the entire course of STEP09.
- **PIC:** Services were stable, performing and with high availability.
- **ASGC:** Good data transfer performances and prestage rates.
  - Some issues with reprocessing uncovered, that were then fixed.
- **FNAL:** After solving some issues the 1sr days, had a successful running.
- **IN2P3:** Demonstrated that MSS is able to deliver the data needed for the reprocessing of both ATLAS and CMS.
- **FZK:** Severe SAN problems affecting the MSS. Initially decided not to participate in STEP09 tape activities, but finally did with "emergency" setup.
  - Need to repeat STEP09 tape activities to test the system under full load.
- **TRIUMF:** Succeeded in STEP09 data distribution and reprocessing.

# Mass Storage Systems

# MSS performance: IN2P3

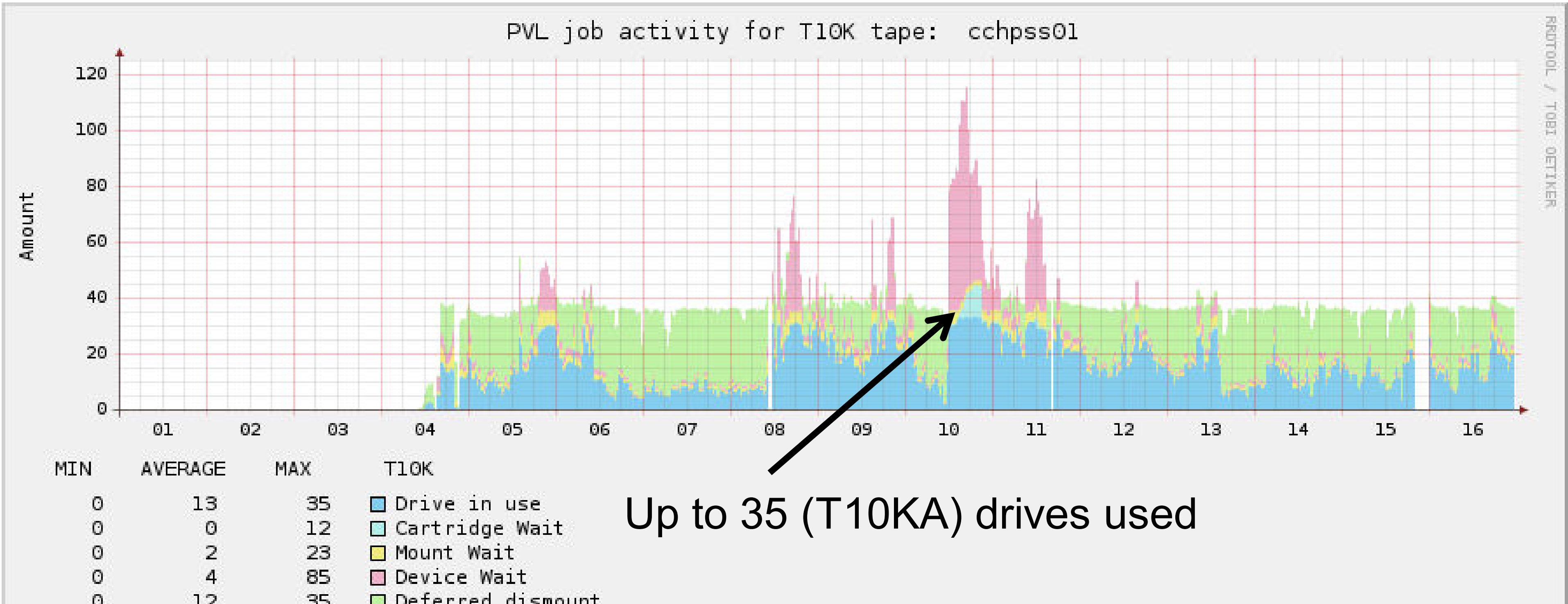


Rate writing

200 MB/s

Rate reading

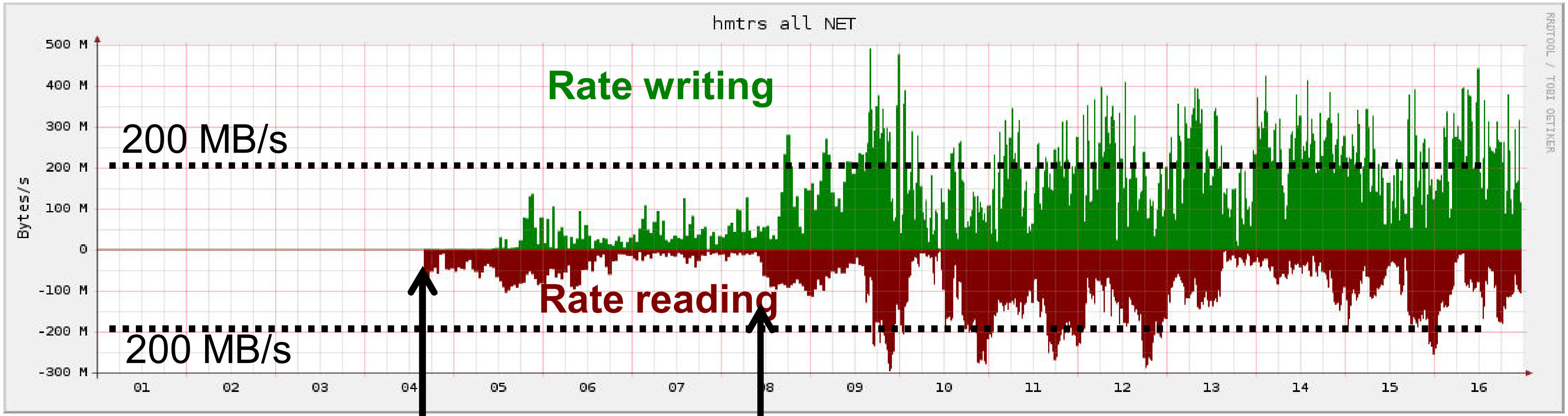
200 MB/s

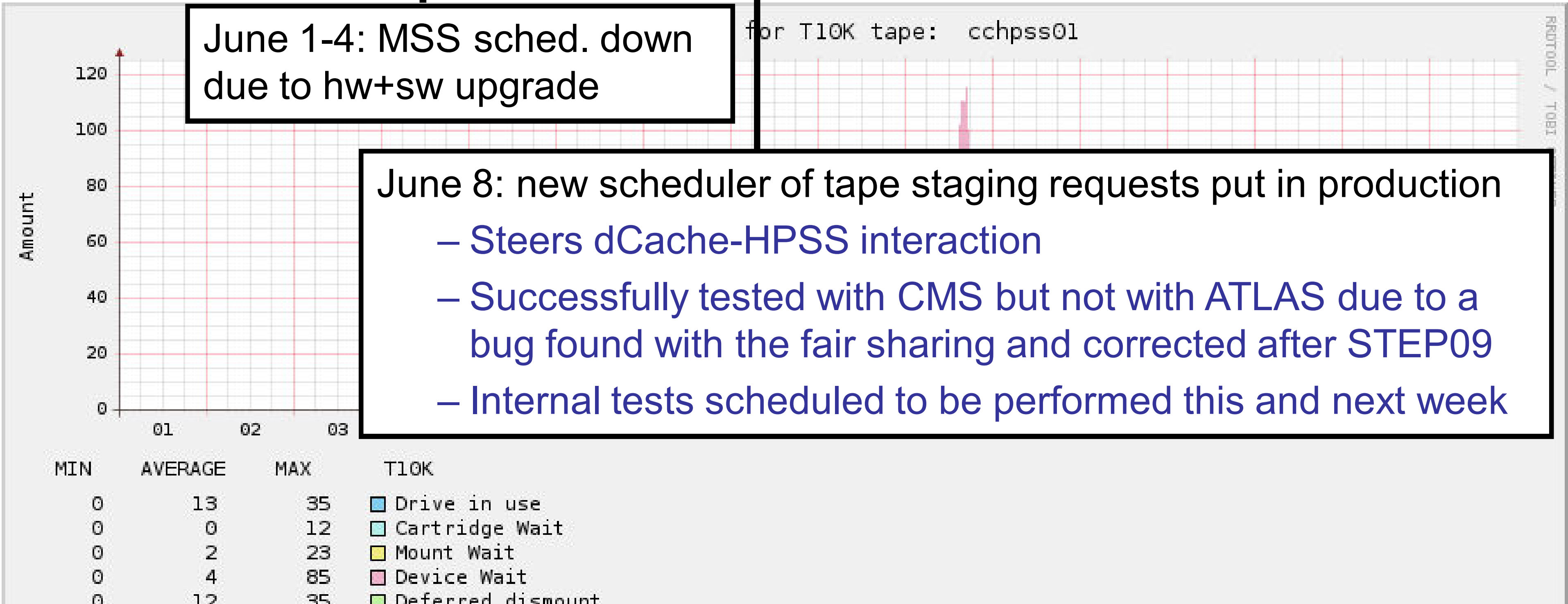


| MIN | AVERAGE | MAX | T10K |
|---|---|---|---|
| 0 | 13 | 35 | □ Drive in use |
| 0 | 0 | 12 | □ Cartridge Wait |
| 0 | 2 | 23 | □ Mount Wait |
| 0 | 4 | 85 | □ Device Wait |
| 0 | 12 | 35 | □ Deferred dismount |

Up to 35 (T10KA) drives used

# MSS performance: IN2P3





| MIN | AVERAGE | MAX | T10K |
|---|---|---|---|
| 0 | 13 | 35 | 🟦 Drive in use |
| 0 | 0 | 12 | □ Cartridge Wait |
| 0 | 2 | 23 | 🟨 Mount Wait |
| 0 | 4 | 85 | 🟥 Device Wait |
| 0 | 12 | 35 | 🟩 Deferred dismount |

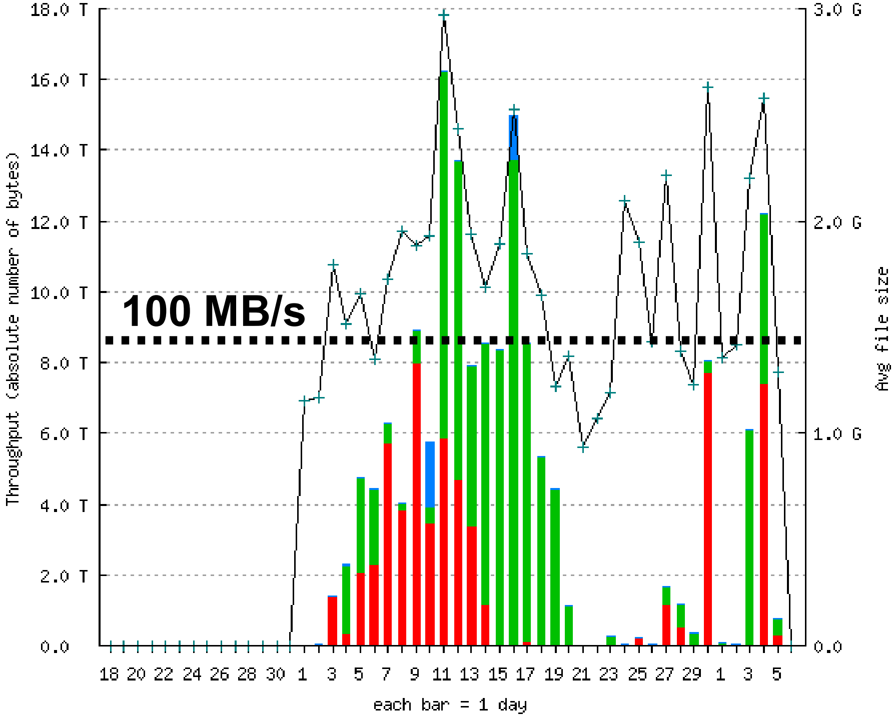
June 1-4: MSS sched. down due to hw+sw upgrade

June 8: new scheduler of tape staging requests put in production

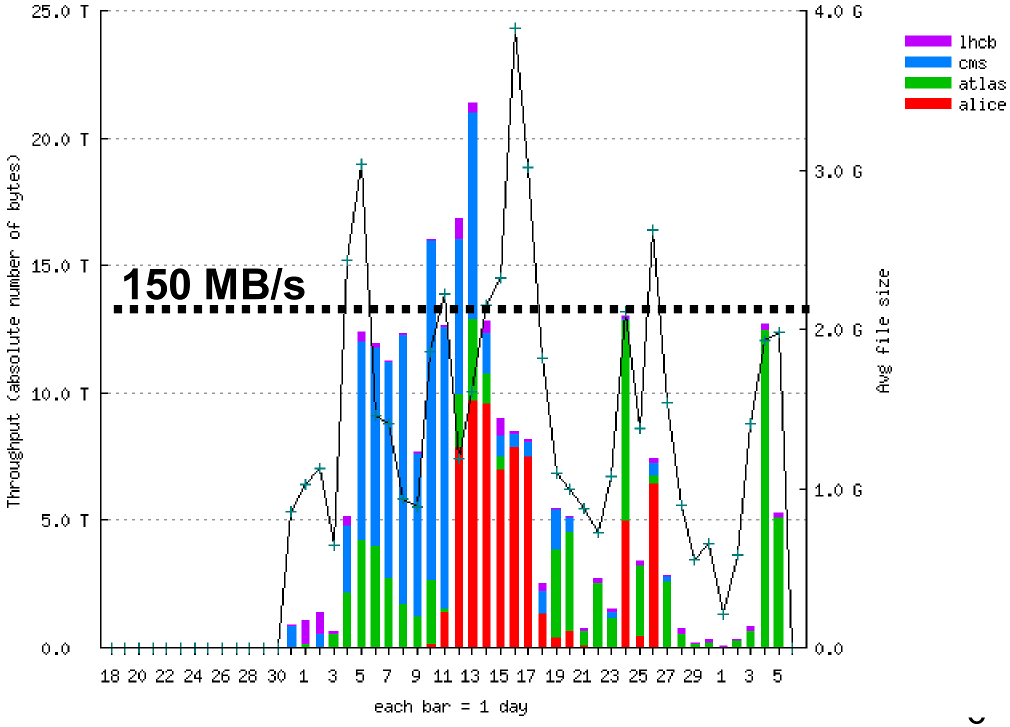
- Steers dCache-HPSS interaction
- Successfully tested with CMS but not with ATLAS due to a bug found with the fair sharing and corrected after STEP09
- Internal tests scheduled to be performed this and next week

# MSS performance: FZK

Rate read from tape
8-10 LTO4 drives

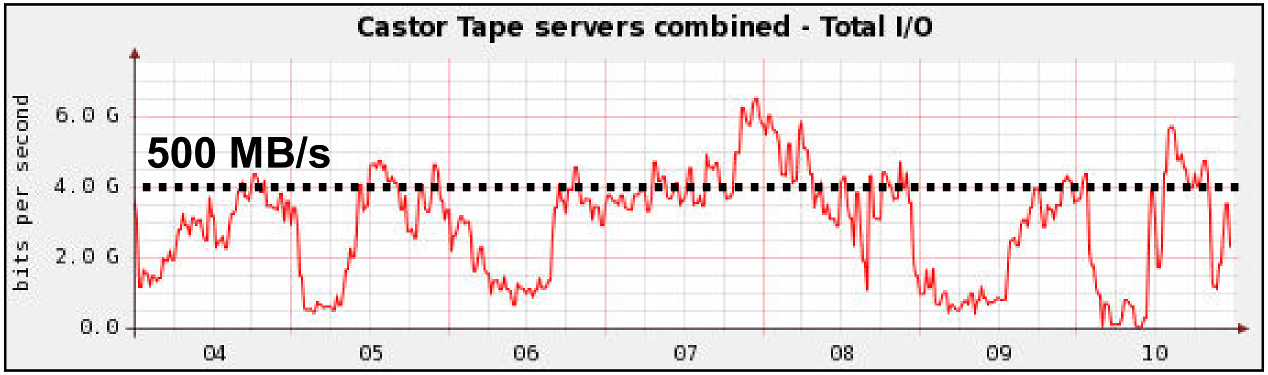


Rate written to tape
6-8 LTO3 drives



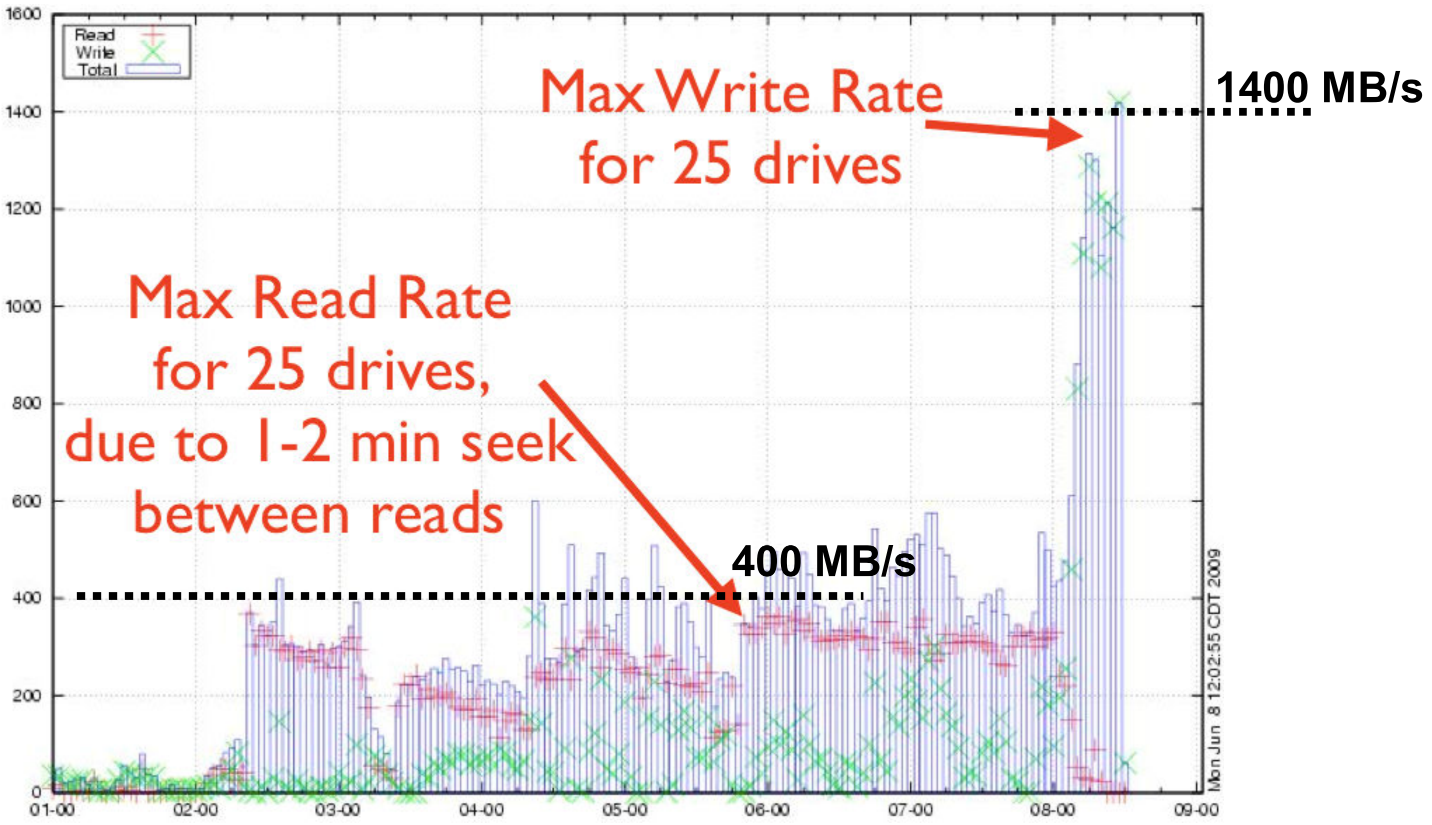
# MSS performance: RAL

- Tape system worked well. Sustained 500 MB/s during peak load.
- Using up to 15 T10KA drives: 4 ATLAS, 4 CMS, 2 LHCB, 5 shared.



- Typical average rate of 35MB/s per drive (1 day average)
  - Lower than expected (looking for nearer 45MB/s)
- On CMS instance, modified write policy gave > 60MB/s but reads more challenging to optimise.

# MSS performance: FNAL



Max Write Rate for 25 drives

1400 MB/s

Max Read Rate for 25 drives, due to 1-2 min seek between reads

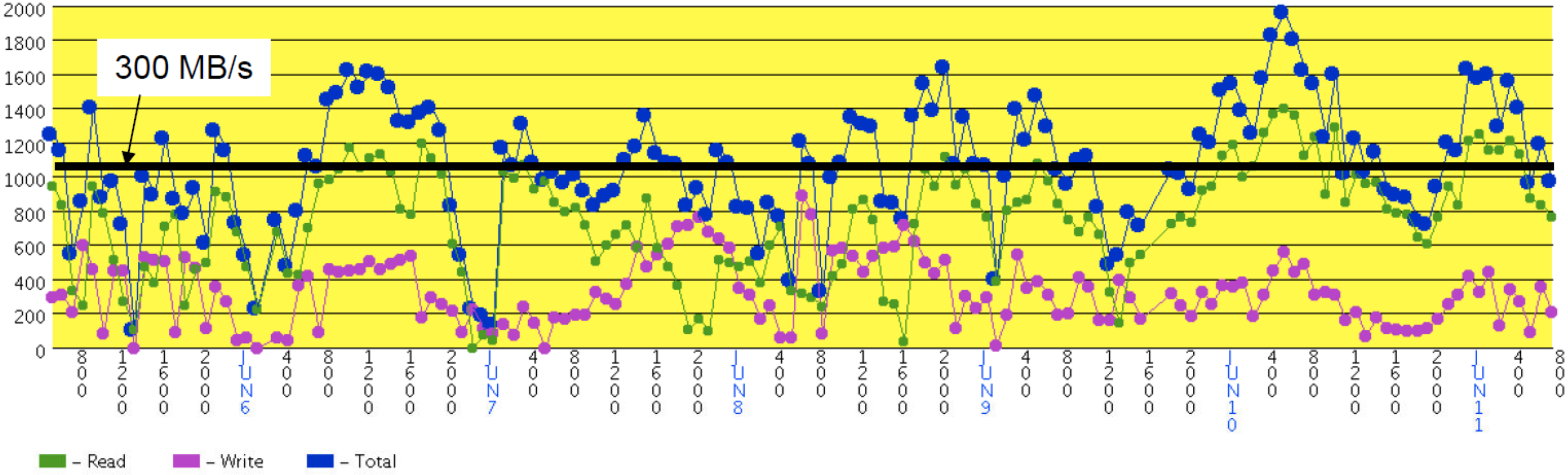
400 MB/s

Issue: ~1 min delay noticed between reads - learnt that this is "normal" for non-adjacent files in LTO (seek time)

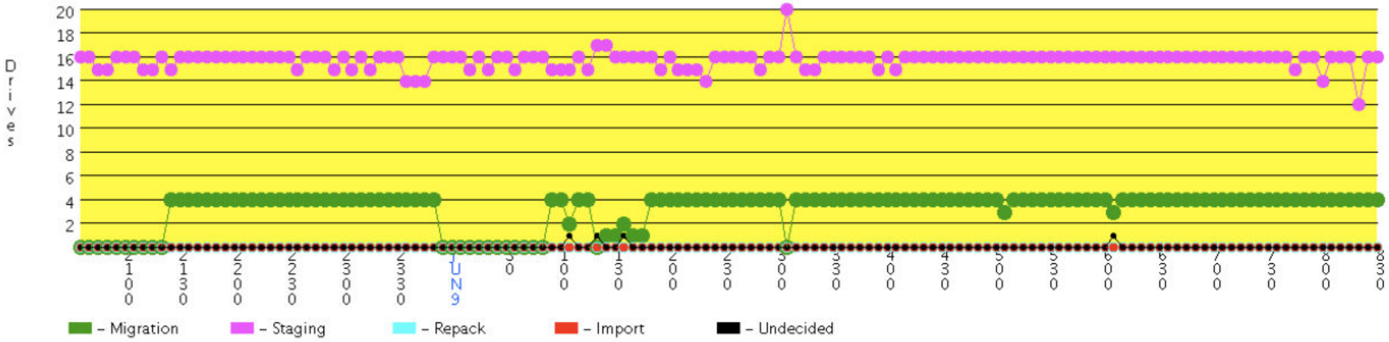
- Lesson learnt: good file organization on tape is important.

# MSS performance: BNL

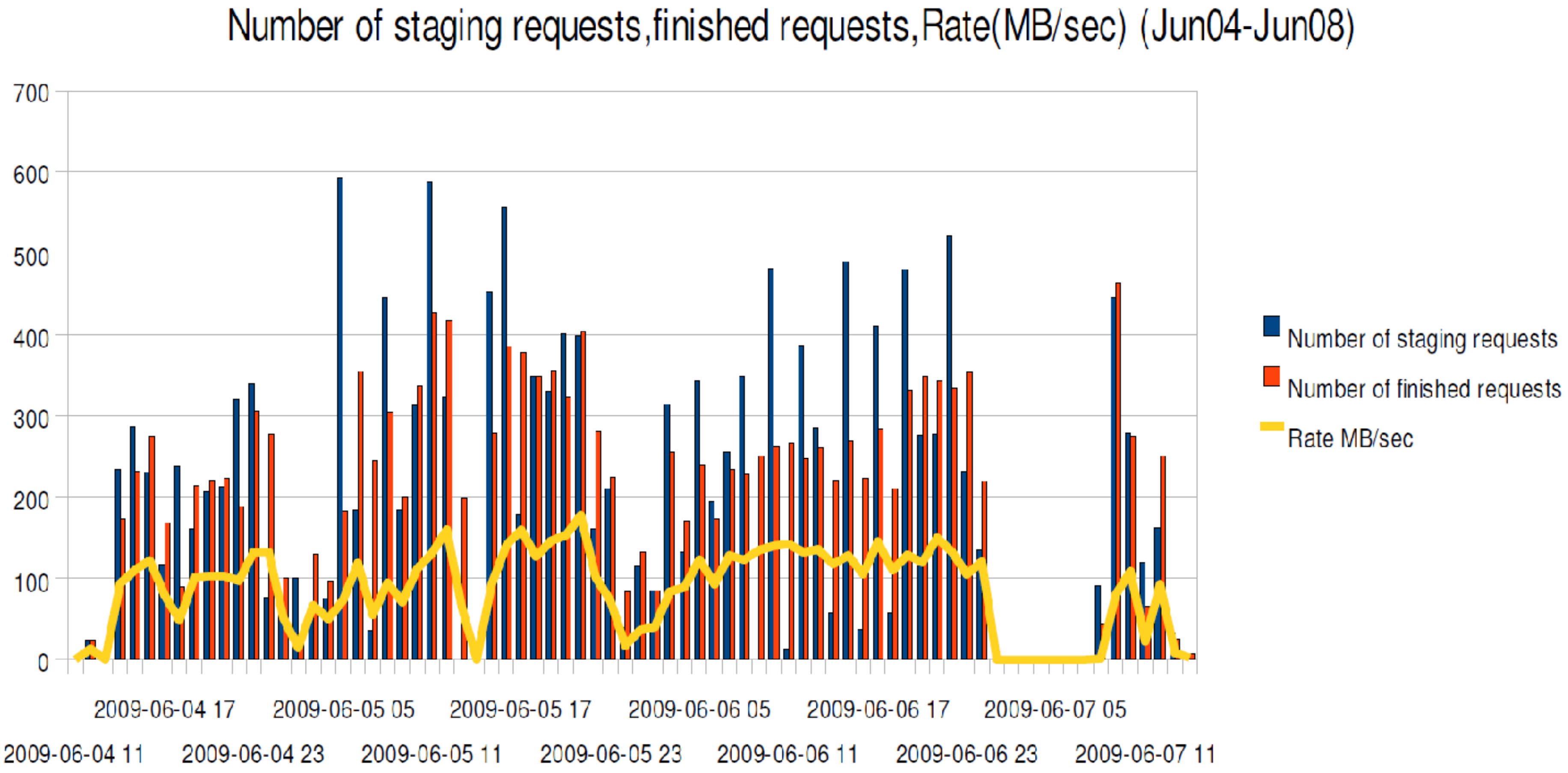
Gigabytes transferred per hour (**blue** is total=r+w)



Count of the 20 LTO4 drives used: 16 read + 4 write

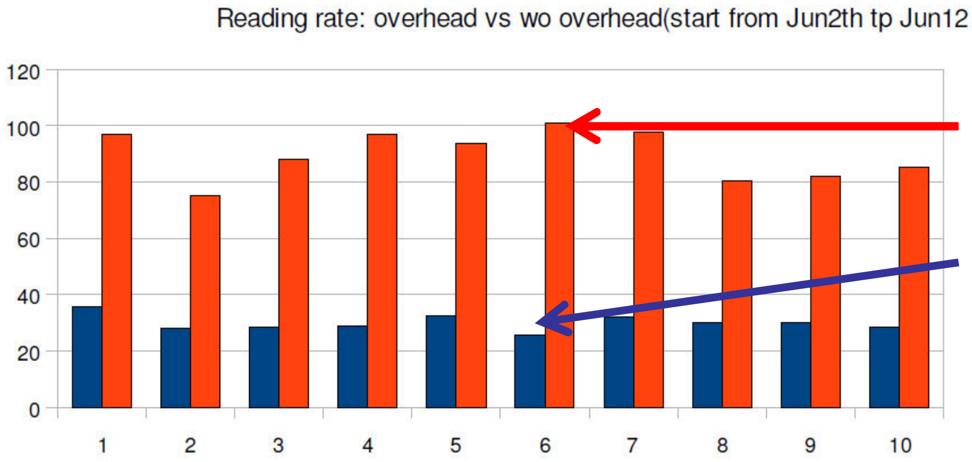


# MSS performance: TRIUMF



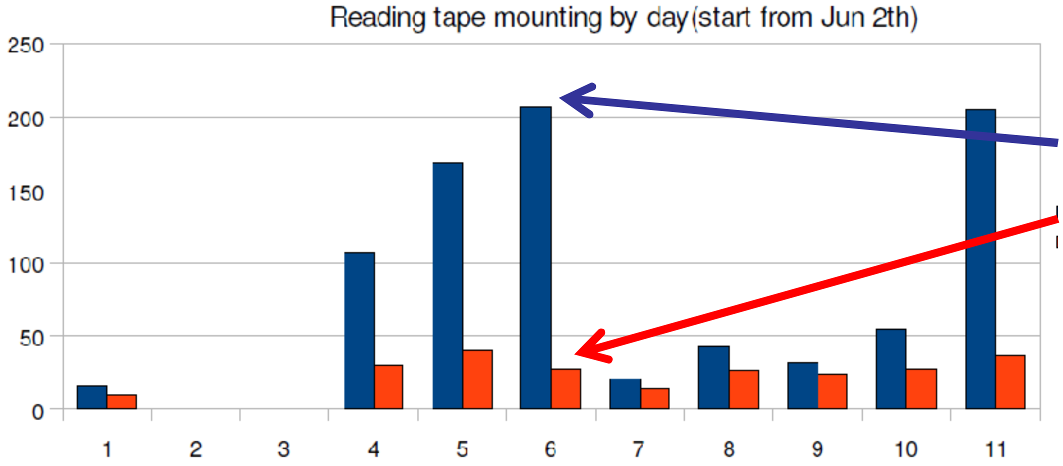
Using 8 LTO4 total: 2 dedicated for writing and 6 for reading.

# MSS performance: TRIUMF


Reading rate: overhead vs wo overhead(start from Jun2th tp Jun12

Read rates per drive: 80-100 MB/s without overheads

Down to ~30 MB/s overall rate, with mount, seek, etc, overheads


Reading tape mounting by day(start from Jun 2th)

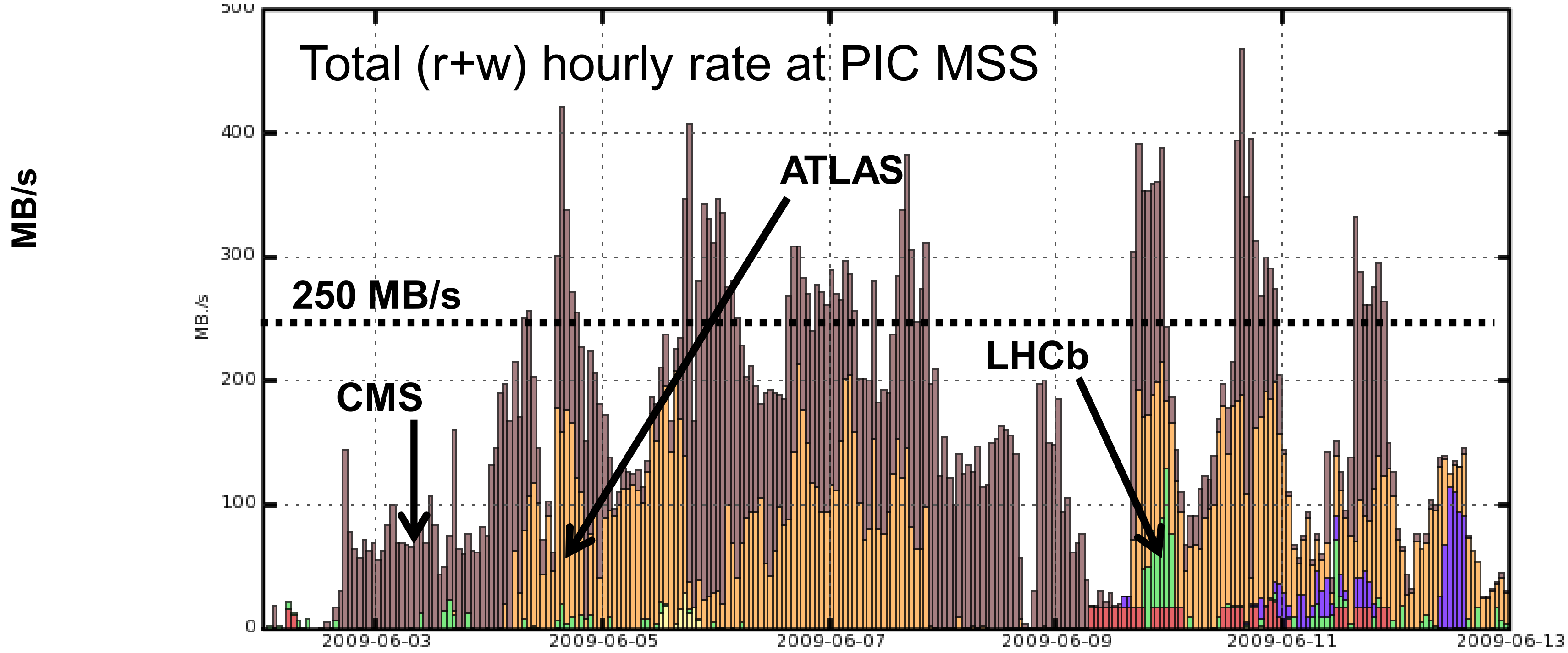
Nr of mounts ~200

Nr tapes read ~25

Specific issue at TRIUMF: nprestage=0 in Panda (conf mistake?)

Tape system never got a long queue of requests to be able to optimally sort it to minimise tape mounts

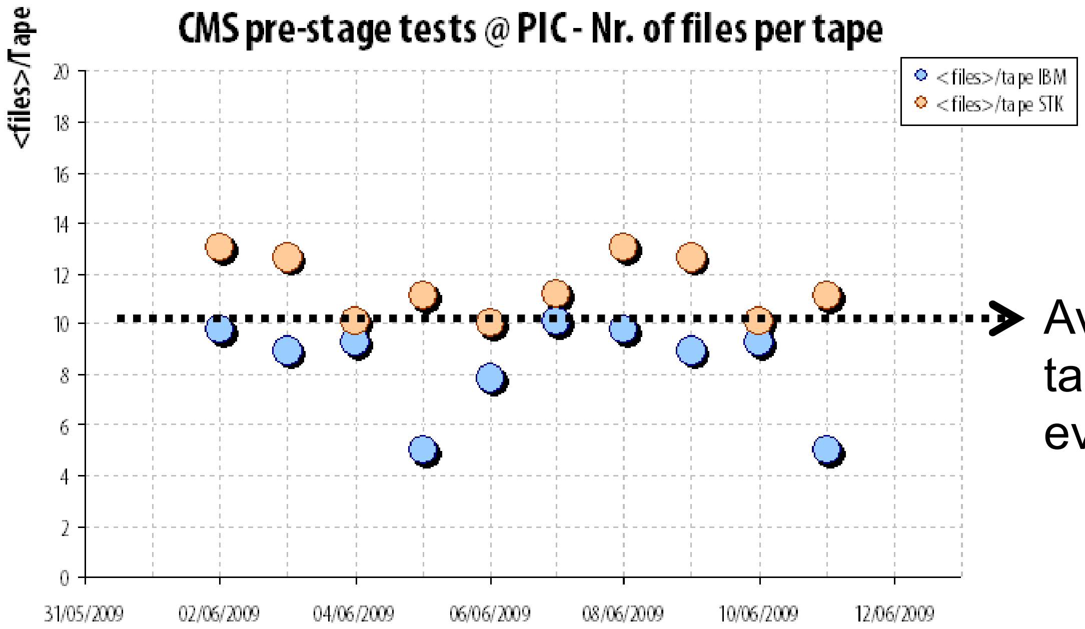
# MSS performance: PIC



- Up to 16 drives used: 7 LTO3 + 9 LTO4
- Average performance per drive seen ~30 MB/s (both read & write)
  - Read: Understanding is that this is dominated by seek times.
  - Write: Dominant effect still being investigated.

# MSS performance: PIC

- CMS pre-stage tests daily stage bulk of files to process next day
  - Example at PIC: Average 4.2 TB (1500 files) per day

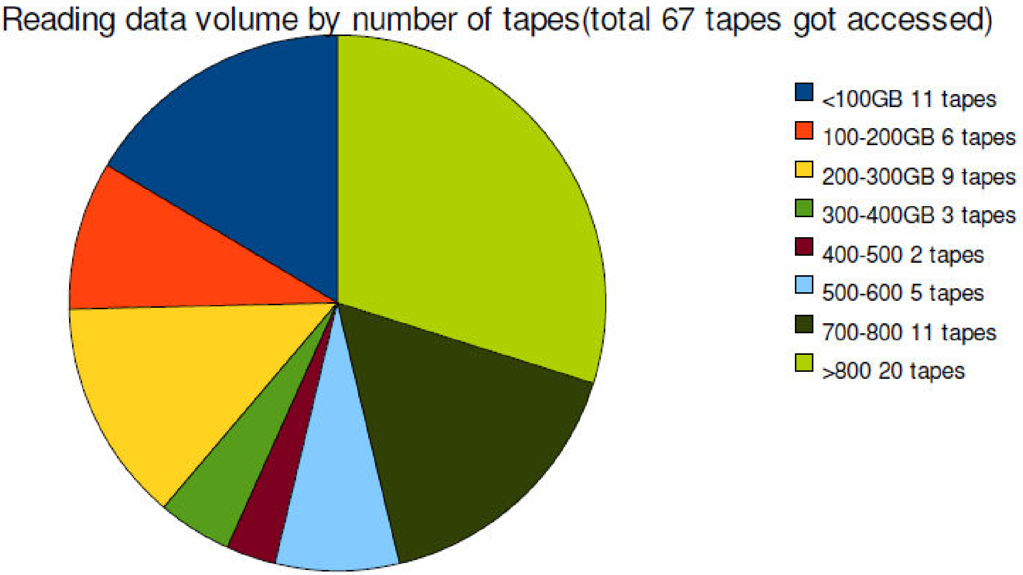


Average 10 files read per tape cartridge mounted every day

- About 140 different tape cartridges accessed in total - 70% of those tapes were mounted every day (>90% every day but one)
- Room for improvement: maximising the nr. files read per tape (ideally read full tape!) can push up stage rates x3

# MSS performance: TRIUMF

Nice plot from TRIUMF, showing that for most of the tapes, almost the full tape is read in the end.



Again: indication that there is room for improvement, if reading full tapes in one go could push rates up to x3.

Provided that experiments store “similar” data on the same tape, would it make sense that sites retrieve the full tape when one file is touched and then “pass” to the experiments the file list just staged so that processing jobs can be triggered by frameworks?

# MSS issues

- **IN2P3:**
  - Need to improve monitoring to view data transfer activity per experiment (WAN: import/export and LAN: Storage → WNs)
    - Site tools but also global ones (GridView?)
  - Competition to access tape by CMS T2s seen – modification of the storage configuration ongoing to avoid T2-T1 interferences.
- **NL-T1:** Performance issues with the MSS: lack of tape drives and tuning of new components (see detailed SIR on the web)
  - 12 new drives installed after STEP'09 and will be brought online over the next month.
  - Tuning of MSS cache during STEP'09 is finished, unexplained low performance during limited period still under investigation.
- **RAL:** Some problems with CASTOR tape migration (and robotics). Handled satisfactorily and fixed. Did not impact experiments.
- **ASGC:** Big ID problem (known bug in CASTOR version used) made up to 50% of ATLAS repro jobs fail - fixed in new CASTOR version.

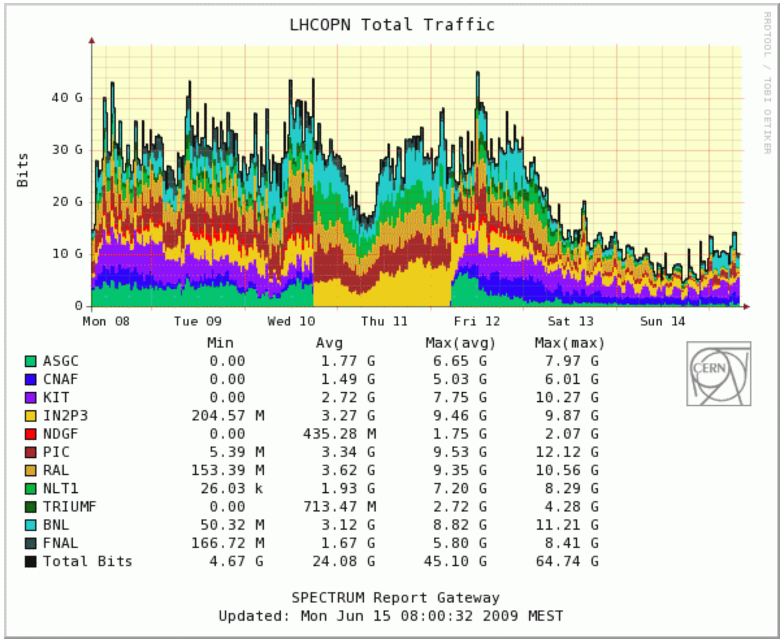
# FZK storage problems

- See detailed SIR on the web
- Necessary setup changes (SAN, drives, dCache pool layout) started well before STEP09, but not solved before the start.
  - Decided first not to take part in STEP09, but on 29th May changed decision to participate with Plan-B:
  - Plan-B: back to old SAN setup. Only 2 drivesVO. Joining with a not-enough tested system ended up creating more problems...
    - Configuration problem of a dCache pool caused problems for CMS (data overwritten on tape read pool before it could be read by jobs)
    - Hw failure of tape server node – ATLAS tape activities delayed.
- CMS availability suffered from hw problems: faulty disk controller.
- ATLAS gridftp servers overloaded
  - Analysis jobs use gridftp to access local SE - Traffic via NAT: bottleneck => Nr of connections to gridftp doors increase => gridftp doors overloaded => FTS transfers also affected.

# Wide Area Network

# Network incident 10-12 June

- Fibre cuts affecting GÉANT and USLHCnet circuits.
- T0-T1 links affected: ASGC, CNAF, FZK, NDGF, TRIUMF, FNAL, BNL
- The LHCOPN traffic was automatically rerouted on other links:
  - ASGC –via dedicated backup link
  - CNAF –via CCIN2P3/FZK
  - FZK – via CCIN2P3
  - NDGF – via NL-T1
  - TRIUMF A+B – via BNL



| | Min | Avg | Max(avg) | Max(max) |
|---|---|---|---|---|
| 🟩 ASGC | 0.00 | 1.77 G | 6.65 G | 7.97 G |
| 🟦 CNAF | 0.00 | 1.49 G | 5.03 G | 6.01 G |
| 🟪 KIT | 0.00 | 2.72 G | 7.75 G | 10.27 G |
| 🟨 IN2P3 | 204.57 M | 3.27 G | 9.46 G | 9.87 G |
| 🟥 NDGF | 0.00 | 435.28 M | 1.75 G | 2.07 G |
| 🟫 PIC | 5.39 M | 3.34 G | 9.53 G | 12.12 G |
| 🟧 RAL | 153.39 M | 3.62 G | 9.35 G | 10.56 G |
| 🟩 NLT1 | 26.03 k | 1.93 G | 7.20 G | 8.29 G |
| 🟩 TRIUMF | 0.00 | 713.47 M | 2.72 G | 4.28 G |
| 🟦 BNL | 50.32 M | 3.12 G | 8.82 G | 11.21 G |
| ⬛ FNAL | 166.72 M | 1.67 G | 5.80 G | 8.41 G |
| ⬛ Total Bits | 4.67 G | 24.08 G | 45.10 G | 64.74 G |

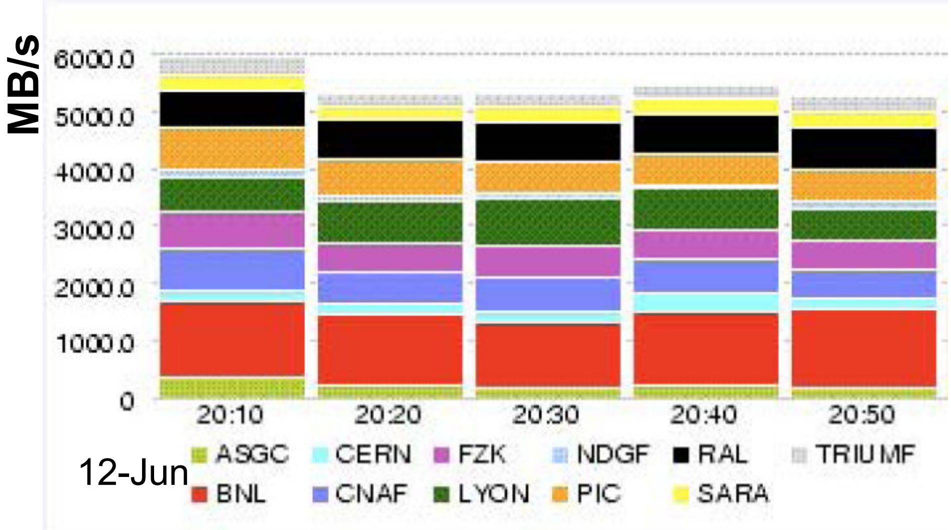
SPECTRUM Report Gateway
Updated: Mon Jun 15 08:00:32 2009 MEST

☑ The backup paths worked successfully and no sites lost connectivity, although there was degradation in the bandwidth available.

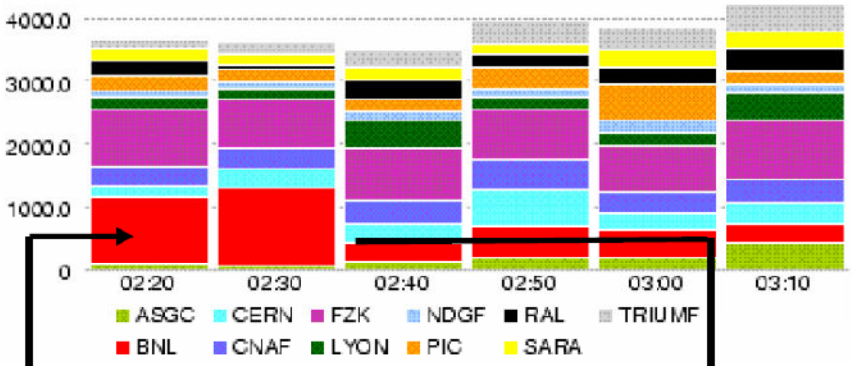
# WAN (FTS) issues: BNL

BNL reached very high WAN transfer rates:

1GB/s import from CERN/T1s + 500 MB/s export to T2s



Issue: ATLAS T0 pool got saturated at 3 GB/s (output rate) BNL was not getting data at nominal rate at times => leading to backlog.



- Other sites still observing decent data rates.
- ATLAS+LCG working on a fix (proposed: dynamic FTS active transfers adjustment).

# WAN (FTS) issues: TRIUMF

- Backlog for transferring data to T2s at the beginning of STEP09.
  - Fine-tune needed per T2

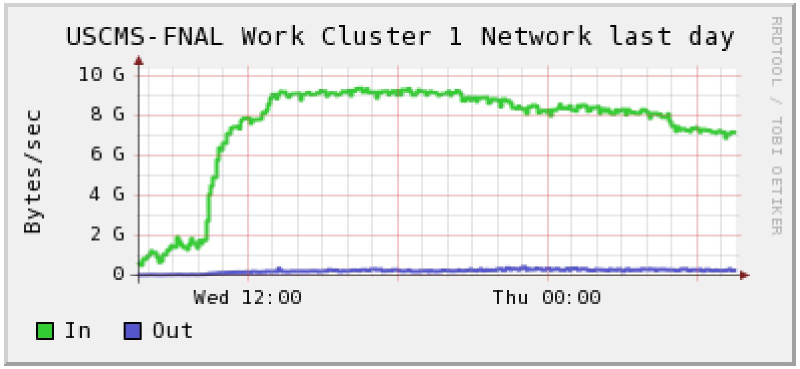
| **Site** | **Nfiles** | **Nstreams** |
|---|---|---|
| SFU | 3 | 1 |
| Alberta | 8 | 7 |
| UVic | 1 | 1 (reduced to avoid overloading dCache at the T2) |

- Transfers between T1s often timed out with original FTS channel parameters at the beginning of STEP09
  - The timeout parameter was increased to higher value for most of T1 channels.
- Lots of transfer failures to CNAF. Problem identified only after STEP09.
  - One of the the 10G network cards had MTU=9000 (jumbo frames). Affected CNAF only because traffic through NREN. It is now set to 1500 like the other cards.

# Job Processing

# LAN: Local SE → WNs rate

- **FNAL:** rate generally at several GB/s (peak at **9 GB/s**) – had around 5000 jobs running
- **BNL: 2.7 GB/s** average transfer rate between disk storage and WNs

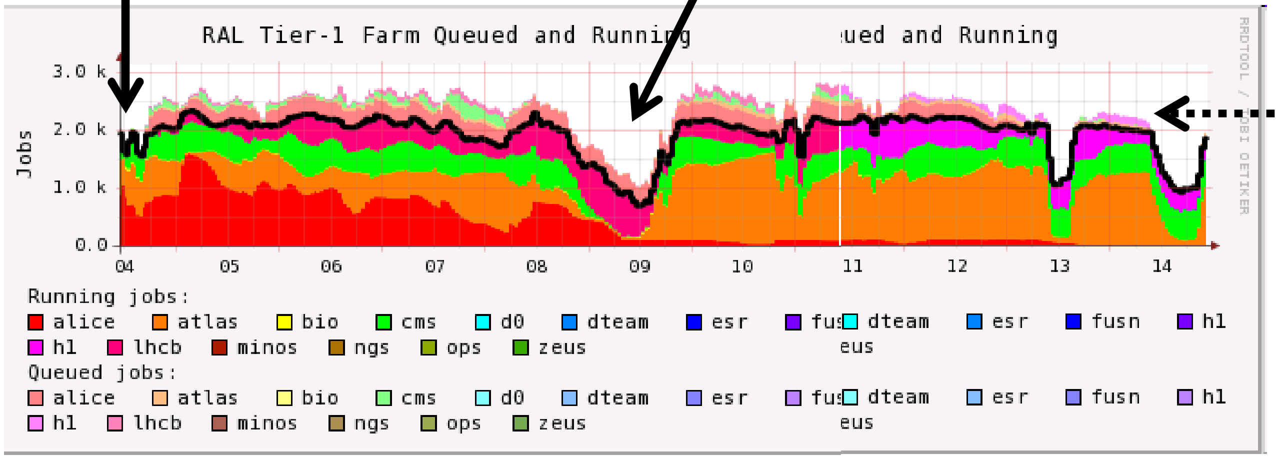


- **RAL:** Batch farm drawing ~375 MB/s during reprocessing. Peaked at **3.7 GB/s** for CMS reprocessing without lazy download – Nr of running jobs around 2000.
- **IN2P3:** Average >1GB/s delivered to WNs, peaks up to **~2.2 GB/s** – Nr of running jobs around 6000.
- **PIC:** Some daily averages >1GB/s, peaks close to max LAN capacity (currently **2.5 GB/s**) – Nr of running jobs around 1400.
- **NL-T1:** Issue - LAN bandwidth from Storage to the WNs too small to accommodate analysis jobs (10 Gbps internal link saturated).
  - Will be fixed over summer together with new procurement of compute and storage resources.

# Batch issues: RAL

Problem: initially, could not entirely fill the system (3GB vmem requirement).

Problem: ATLAS job submission exceeded 32K files on CE.
(thought ATLAS had paused, took a time to spot).



RAL farm typically running >2000 jobs

FairShare not honoured 1st days.
ggressive ALICE submission beats ATLAS.

By 9th June at equilibrium:

| ATLAS, | CMS, | ALICE, | LHCb |
|---|---|---|---|
| 42% , | 18%, | 3%, | 20% |

# Job processing issues: ASGC

- Job scheduling policy - ATLAS T2 MC jobs got more slots than allocated, caused by:
  - ATLAS T1 and T2 jobs submitted by same user - ended up with same priority in the system.
  - T1 jobs failed quickly due to Big ID problem.
  - Consequence: CMS only got its CPU share the last 2 days.

- Job efficiency for CMS reprocessing lower than in other T1s – investigating.

# Job processing issues: NDGF

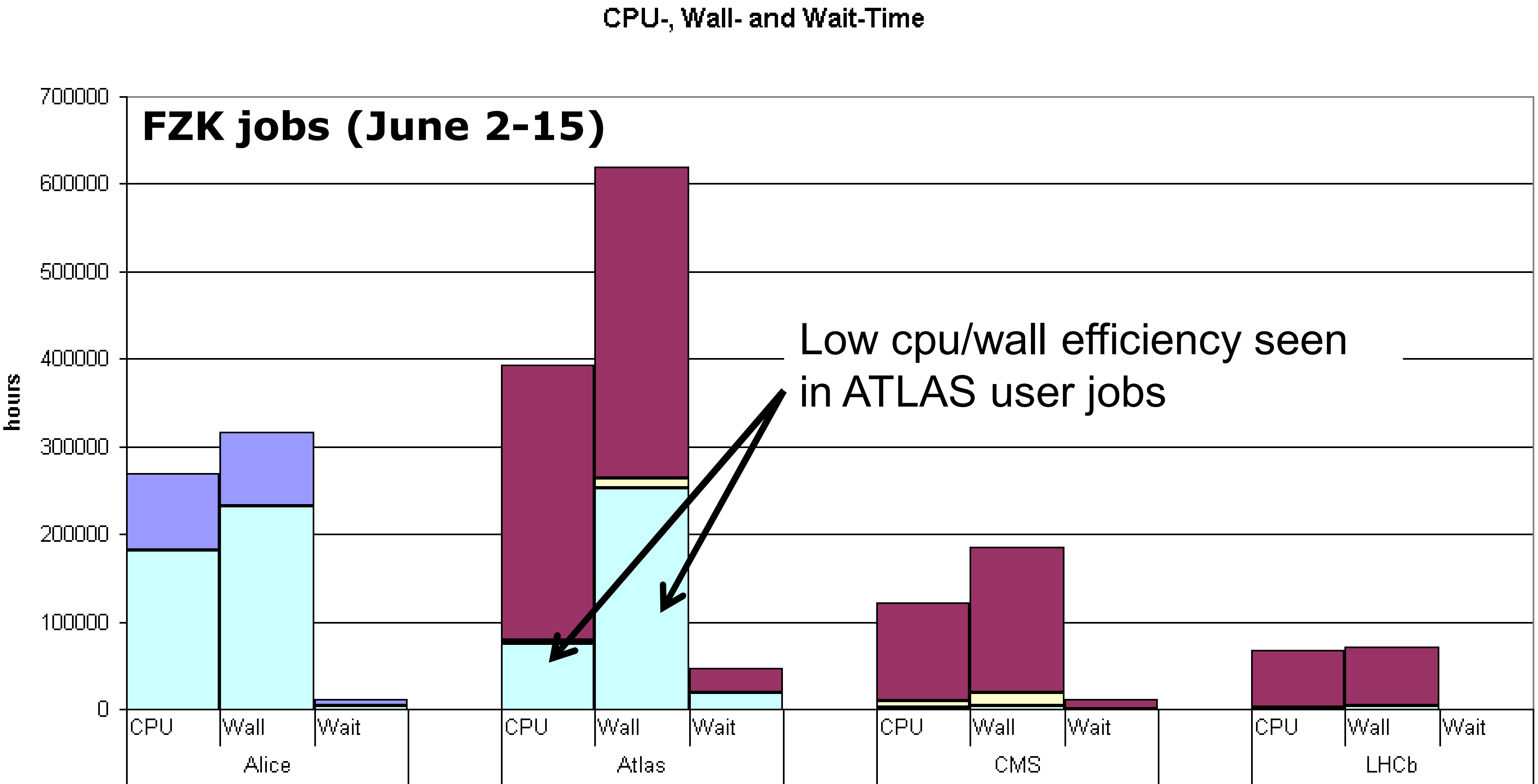
- Short summary: jobs are of very different types…

- Identified 4 types of jobs, with very different requirements, and that have very different efficiency:
  - Low I/O (evgen, simulation) → 100M/12h
  - Medium I/O (digit, reconstruction) → 1G/2-3h
  - High I/O (merge, reprocessing) → 2-4G/30min
  - Highest I/O (user) → 5G/10min
    - **User job requirements are HUGE!**

# Job processing issues: NDGF

- The main issue at NDGF was low throughput (amount of data processed), attributed to various optimisation issues.
  - User analysis jobs caused heavy load on network and file systems on several individual sites, slowing the overall processing.
- In general, different types of jobs require different site optimization (hardware, OS, bandwidth).
  - NDGF does have different sites, but no scheduling optimization was done; and it is next to impossible to do with pilot jobs.
- The distributed dCache setup assumes that the time to transfer data is much shorter than processing time, thus some time ago abandoned the “sending jobs to data” configuration.
  - Learnt that for the analysis jobs it is the other way around, so now have to bring back the old procedure.

# Job efficiency



**FZK jobs (June 2-15)**

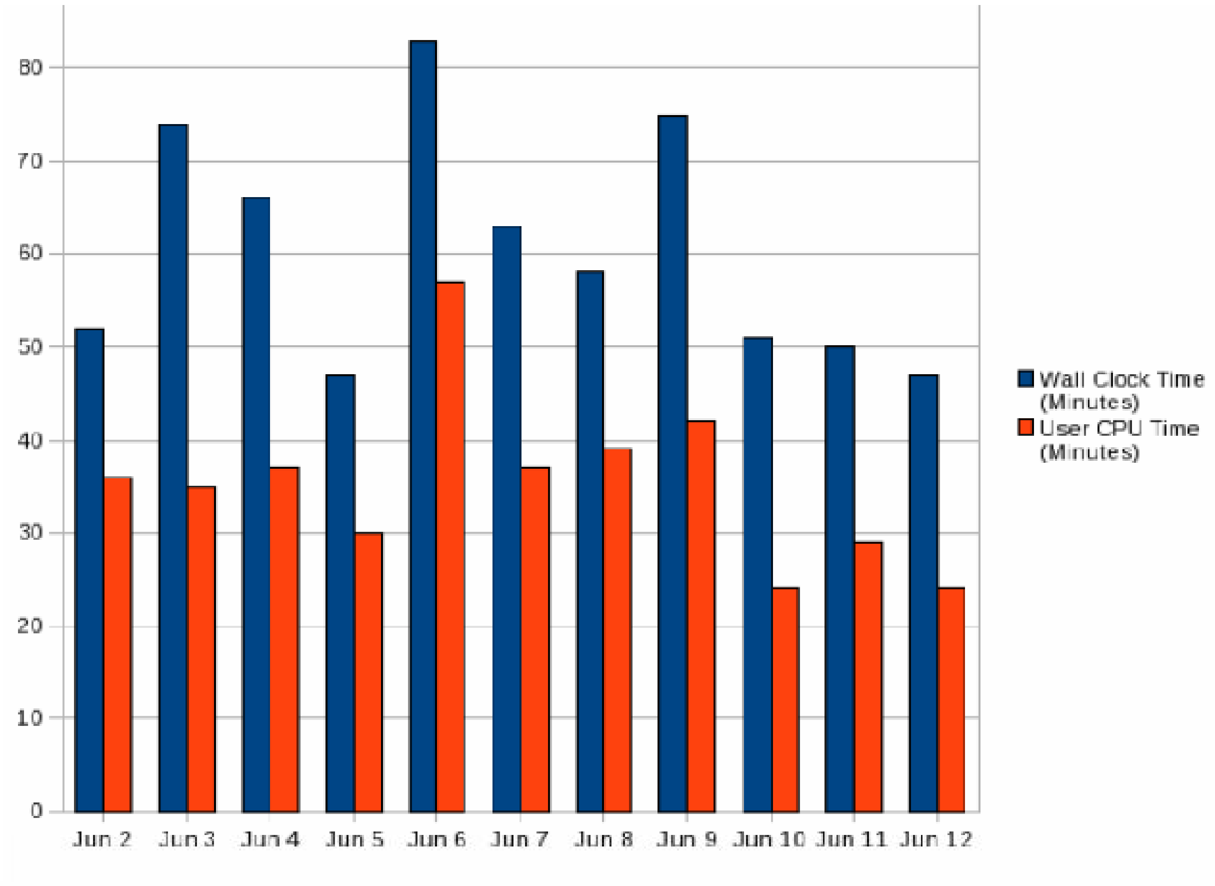
Low cpu/wall efficiency seen in ATLAS user jobs

| | Alice CPU | Alice Wall | Alice Wait | Atlas CPU | Atlas Wall | Atlas Wait | CMS CPU | CMS Wall | CMS Wait | LHCb CPU | LHCb Wall | LHCb Wait |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 🟦 sgm | 88857,98 | 83613,59 | 6796,65 | 11,28 | 81,88 | 25,97 | 9,3 | 119,95 | 19,18 | 2,5 | 9,44 | 1,88 |
| 🟥 prd | 0 | 0 | 0 | 314283,4 | 355984,8 | 27626,61 | 111273,3 | 167088,8 | 10274,16 | 65232,35 | 68654,37 | 104,77 |
| 🟨 DGrid | 0 | 0 | 0 | 4248,36 | 11178,39 | 23,98 | 8413,36 | 14977,69 | 71,01 | 0 | 0 | 0 |
| ⬜ others | 181581,4 | 232612,4 | 4422,99 | 75081,75 | 252956,7 | 18786,17 | 1360,84 | 3544,03 | 61,17 | 2509,56 | 3079,7 | 38,97 |

# Job efficiency

Impact of very I/O intensive jobs also seen…

Average Wall Clock and User CPU Time – U.S. ATLAS Tier-1



# Grid Services

- Several sites report CE, LFC, FTS, BDII services have operated stable, during STEP09 no issues.
- Oracle-3D: no issues.
  - ASGC: wrong 3D listener port delayed the start of ATLAS reprocessing for aprox 1 week.
  - IN2P3, PIC: To my knowledge, Oracle not used to provide conditions to the local jobs.

# Comments

- Information provided by the experiments really appreciated
  - Goals, plans, execution and achievements very well structured and maintained up-to-date daily. Extremely useful.
  - Thanks to the experiments for this effort!
- This is the first time tape systems have been thoroughly exercised. Very valuable experience. Lots of lessons learnt.
  - Tape system requires much synchronization and tuning how requests are submitted (eg allow large tape request queues to allow optimal ordering and minimise nr of mounts).
  - Inefficiency due to seek times found to be non-negligible – room for improving overall drive rates by organising data and recall such that full tapes can be read.
- Very I/O intensive workloads (analysis) seen.
  - Way to control impact on cpu efficiency needs to be studied.
  - Important to plan for the network infr. to scale and cope with this req.
- At most sites, services operating stable and at high performance levels.
  - Sites that found problems and did not meet targets, already addressing the issues and planning for a re-run of test activities.

# Questions/Comments

# WAN transfers

Import rates from T0/T1 tested:

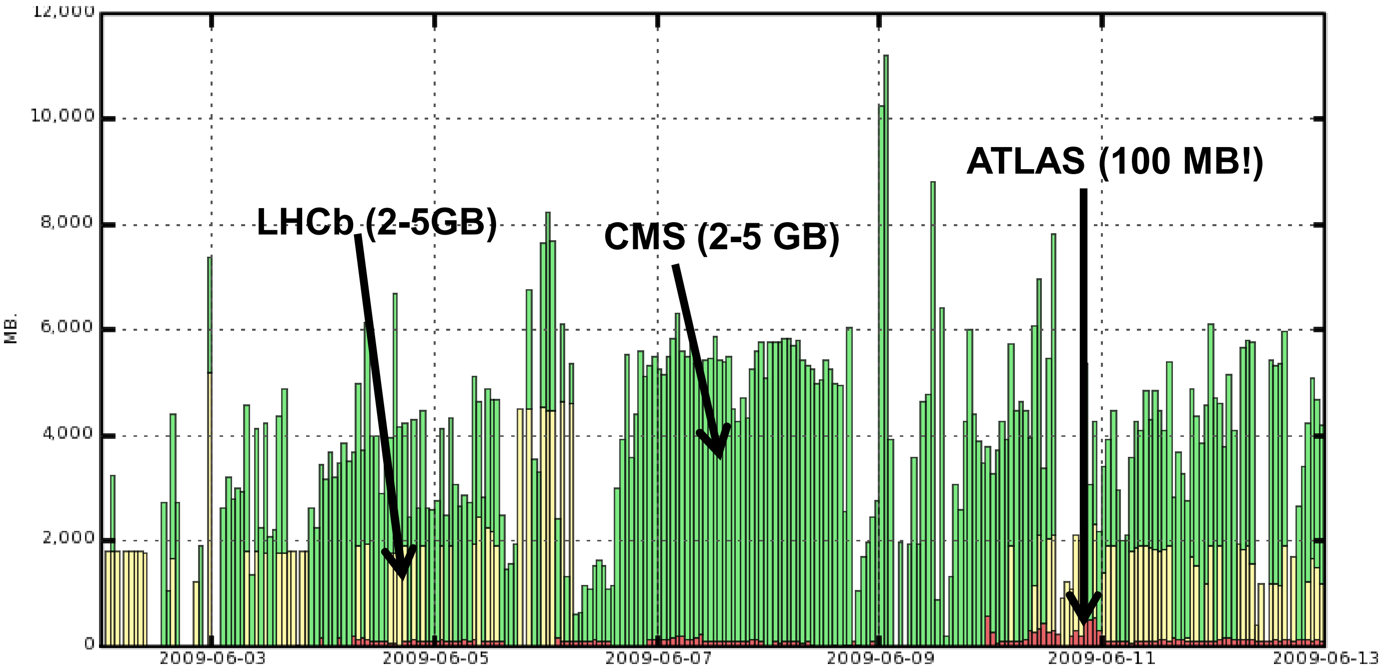
- BNL up to 1000 MB/s
- FNAL up to 470 MB/s
- PIC, RAL up to 440 MB/s
- FZK up to 730 MB/s
- TRIUMF up to 170 MB/s
- …

Export rates to T2s tested:

- FNAL up to 780 MB/s
- RAL up to 750 MB/s
- BNL up to 500 MB/s
- PIC up to 500 MB/s
- FZK up to 350 MB/s
- TRIUMF up to 140 MB/s
- …

# MSS issues: PIC

- Average file sizes seen reading from tape look OK, big enough:
  - Around 1.5 GB for ATLAS, 2.5 GB for CMS and 1.7 GB for LHCb
- For writing, small files seen for ATLAS. Known issue. Will disappear.


MB
LHCb (2-5GB)
CMS (2-5 GB)
ATLAS (100 MB!)