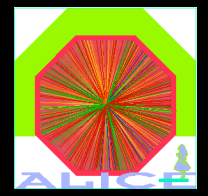


Access to storage in ALICE

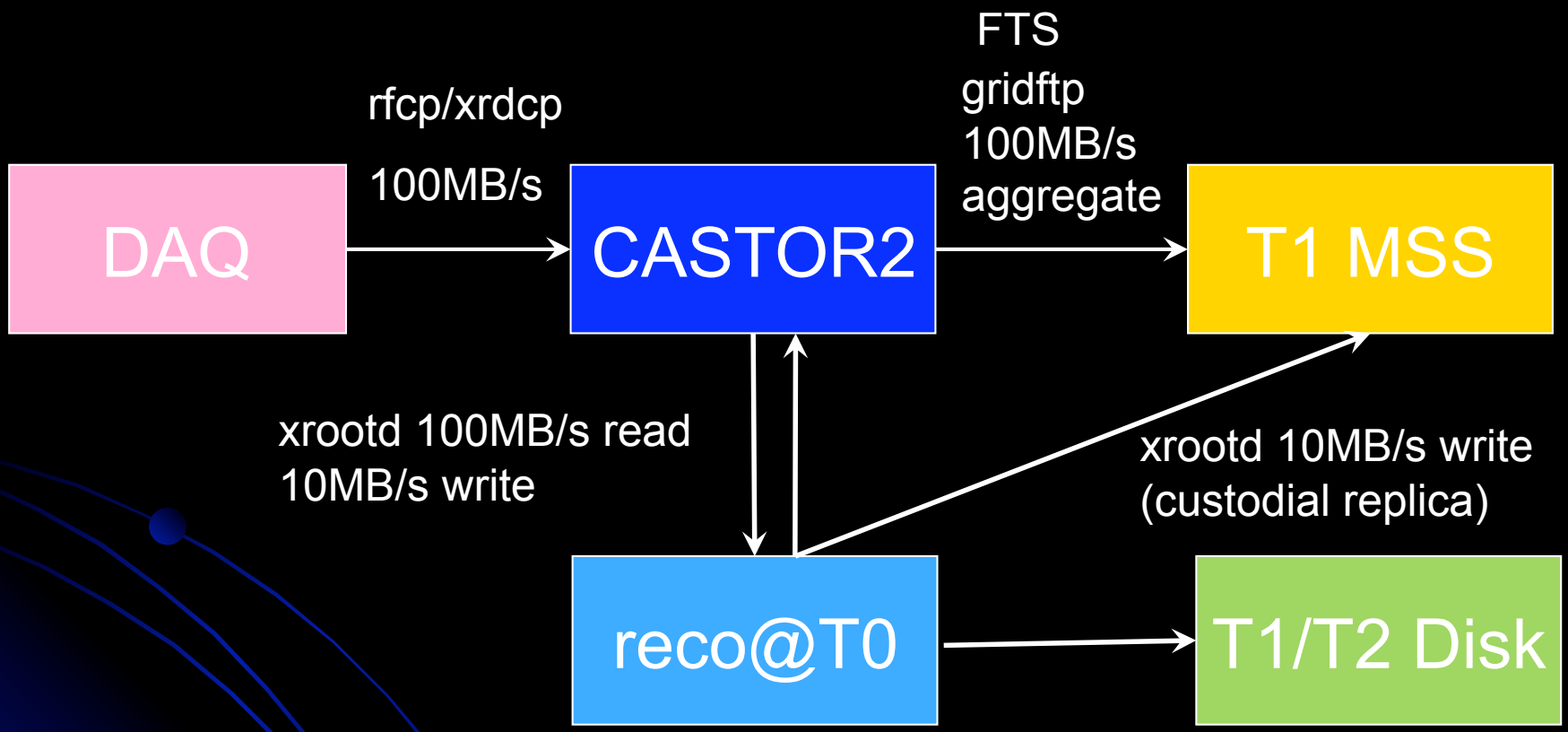
CCIN2P3, March 4, 2009

Rates

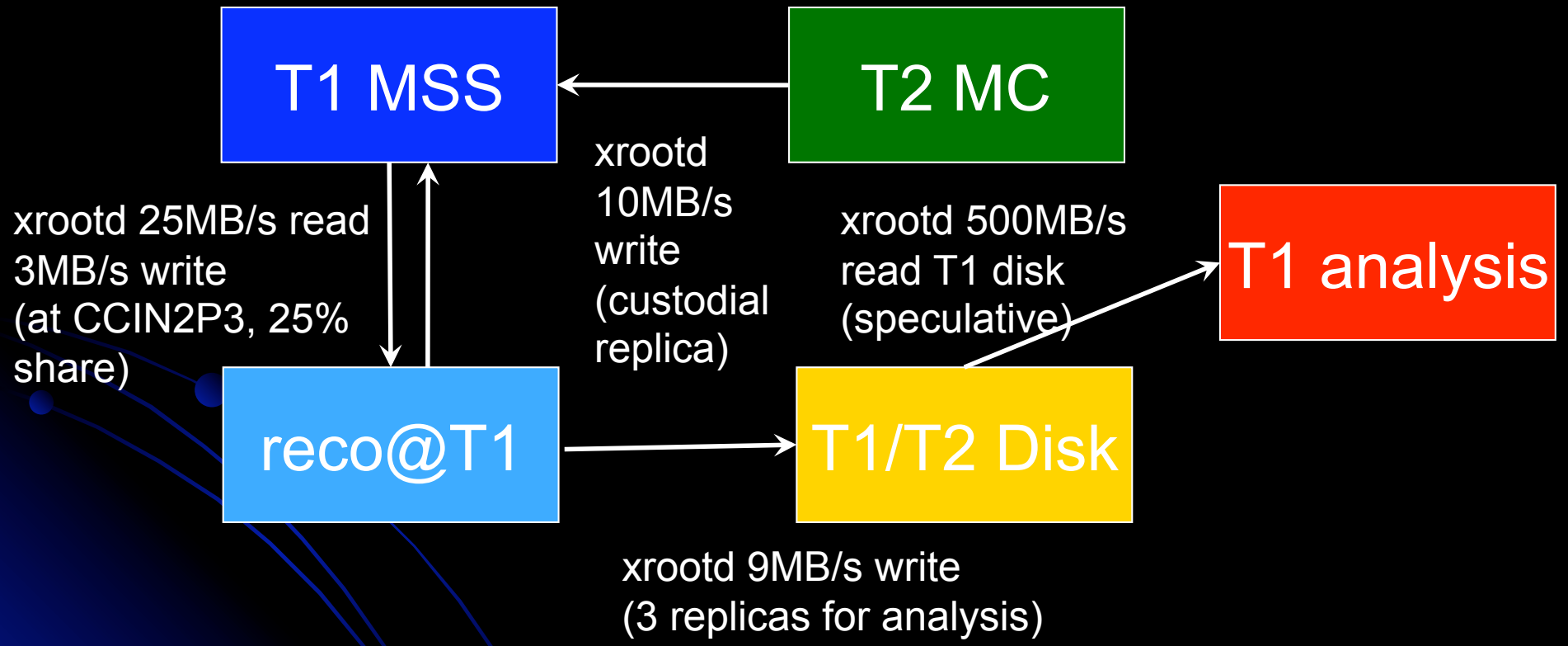
- p+p data taking, steady state

- A+A data taking – 2GB/sec to T0 MSS, all other rates remain unchanged

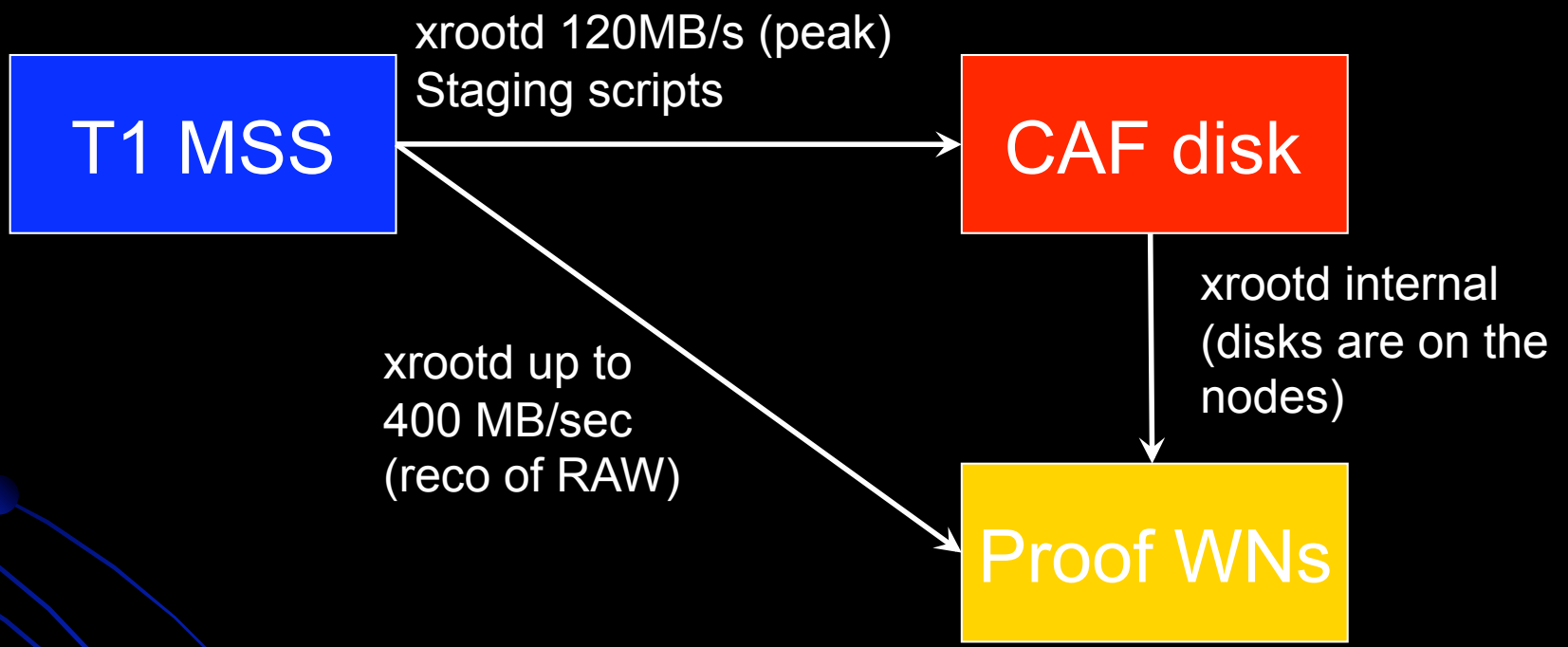
T0 storage for data taking, replication and processing



T1 storage for RAW processing, batch analysis and T2 custodial



CERN AF storage



Zoom in on CCIN2P3

- Storage managed by 2 virtual SEs
  - T1D0 (MSS)
    - CCIN2P3 setup -> /pnfs/in2p3.fr/data/alice/T1D0
  - T0D1 for disk
    - CCIN2P3 setup -> /pnfs/in2p3.fr/data/alice/disk
- Both currently under dCache
- Common requirement for all storage
  - xrootd-enabled
  - visible on WAN

Zoom in on CCIN2P3 (2)

- What is tested
  - T1D0 for RAW replication
  - T1D0 for (limited) reconstruction
- Recall of data from tape is an **open issue**
  - 'prepare to get' is a serial, per client, request
    - Multiple tape mounts, large overhead, inefficient use of tape drives
  - At CERN – CASTOR-specific parallel recall of multiple files
    - Only applicable for centrally managed operations, like reconstruction

dCache storage for analysis tasks

- General experience on T0D1
  - Tested at 2 T2s and 2 T1s
    - Performance **strongly** depends on the level of dCache expertise
  - Number of data movers – bottleneck for multi-file access
    - Clients blocked, frequent timeouts
    - Tuning of storage is a long and tedious process – no generic rules, mostly trial and error
  - Not very positive overall

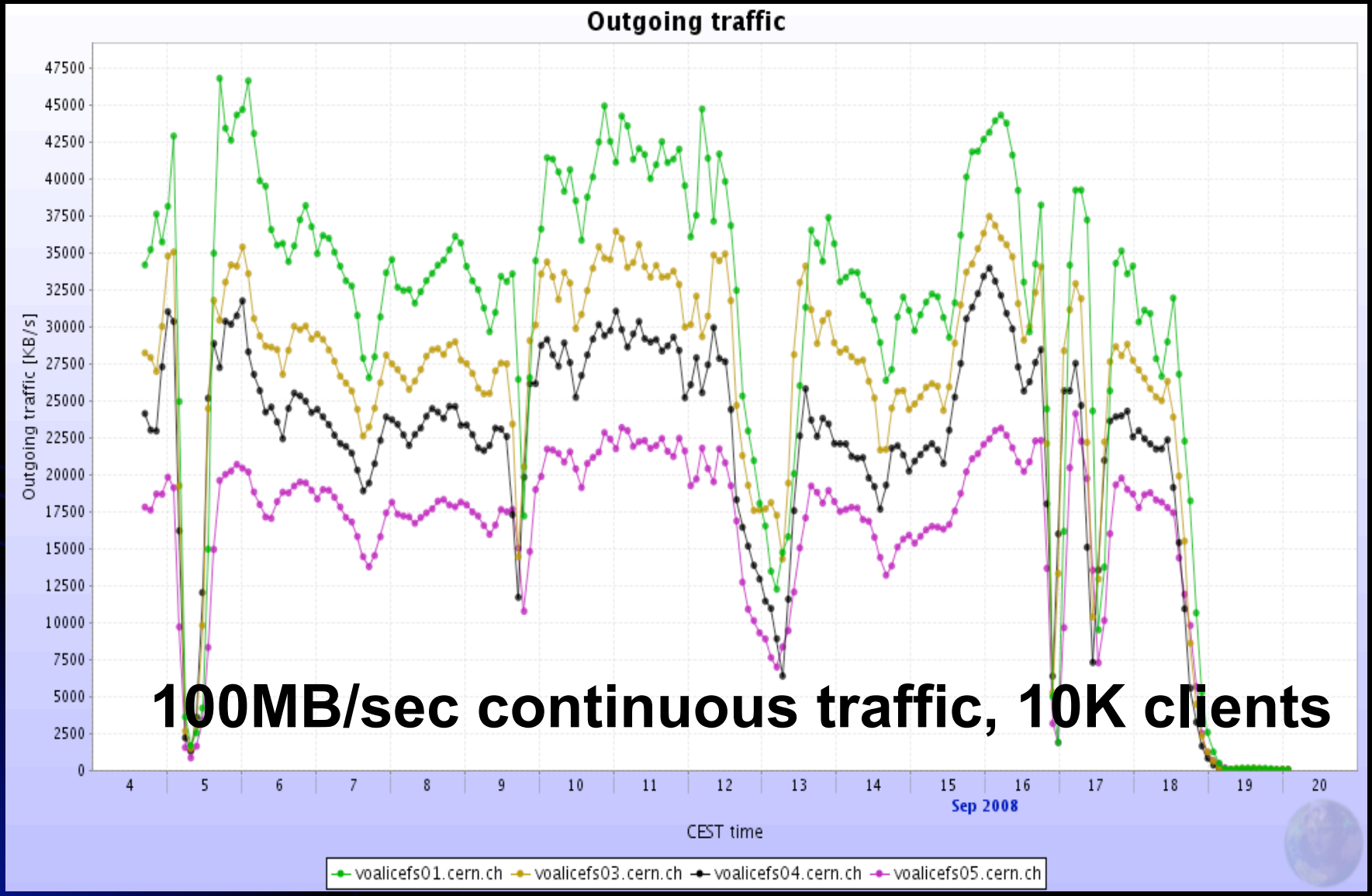
T1D0 disk buffer sizes

- MSS buffer – large is good
  - Allows for data to be ‘hot’ for longer
  - Less tape recalls
  - Ideally as large as the tape pool
  - ALICE CERN pool is 500TB = 50 days of data taking + reconstruction ESDs
    - It is xrootd-enabled (plugin), also good for analysis

Few examples of xrootd use in ALICE

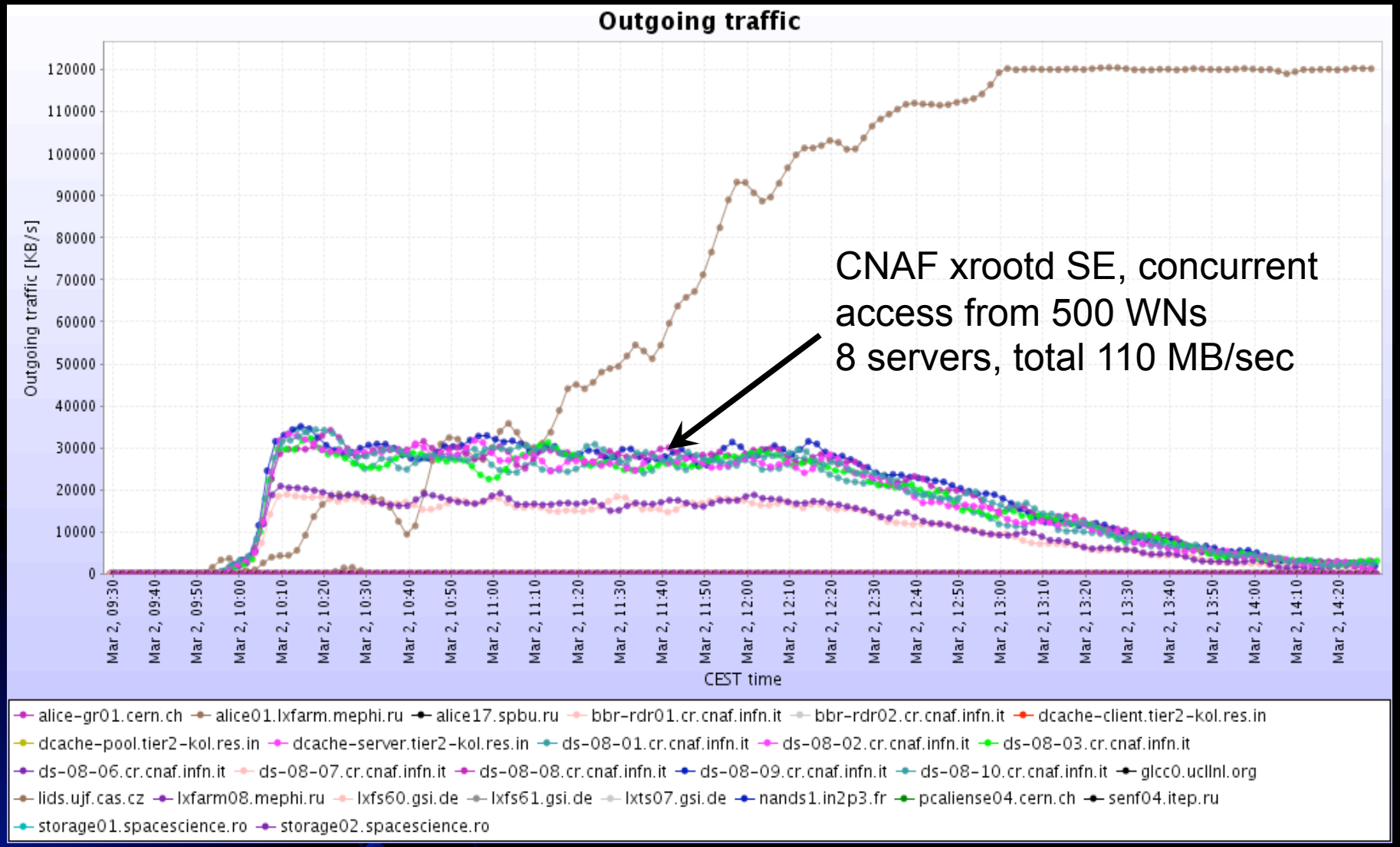
- Conditions data repository
  - ALICE conditions data are ROOT files annotated in the AliEn catalogue
  - Populated from various online DBs and runtime detector tasks
  - Primary copy on xrootd storage servers at CERN (5x, 30 TB total)
  - Accessed by **all** MC and reconstruction jobs on the Grid

Conditions database servers

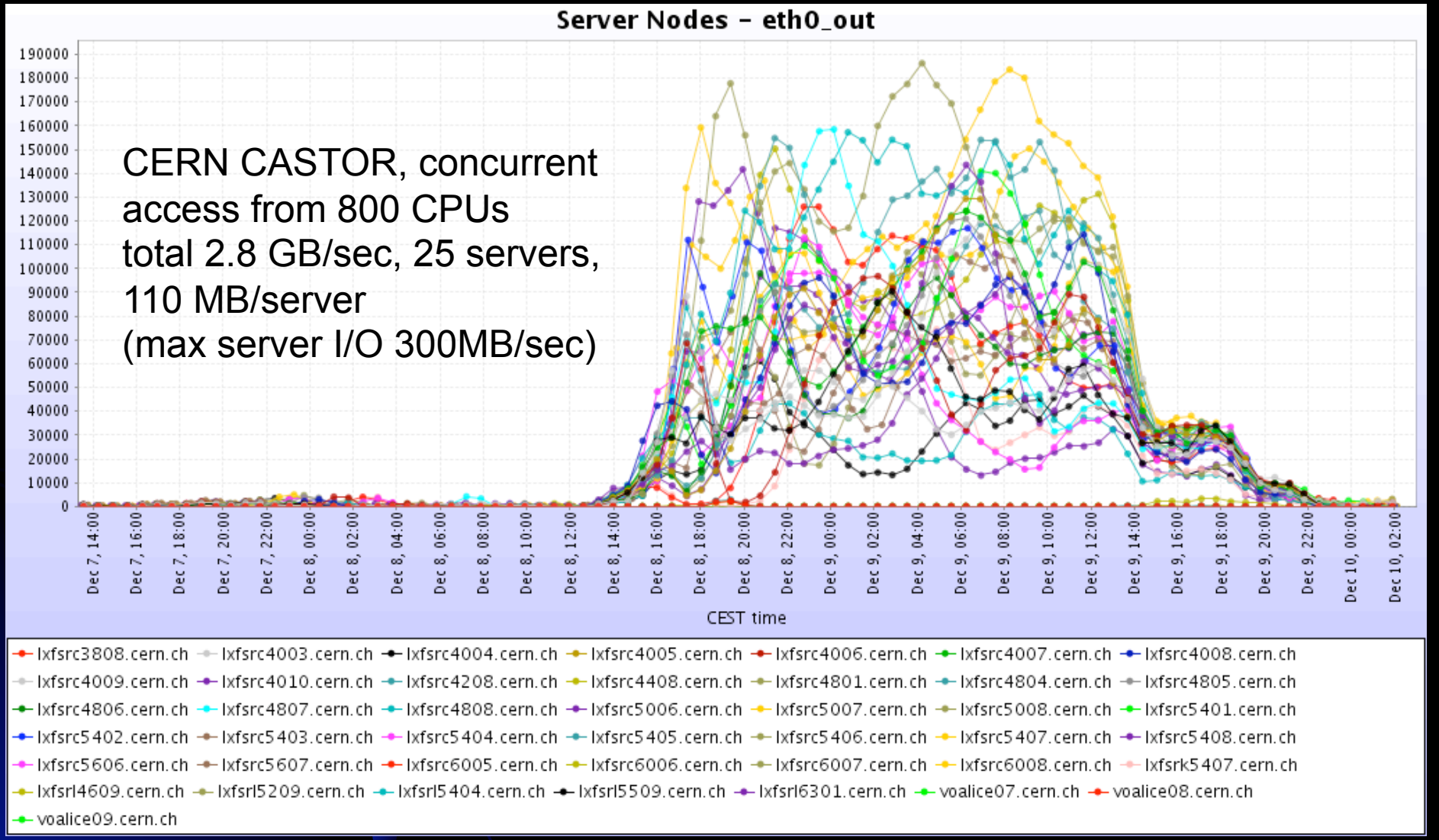


100MB/sec continuous traffic, 10K clients

Analysis servers



Analysis on CASTOR servers



Summary (analysis)

- Analysis on the Grid is quite possible
  - If the data is on disk
- xrootd-enabled storage is fully capable of supporting multi-client access
  - Chaotic analysis up to 800 CPUs tested
  - The limit of the servers was not even approached
  - Stability of xrootd is remarkable
- ALICE would like to get a large xrootd cluster at CCIN2P3 as T0D1 pool
  - Batch user analysis and LAF storage