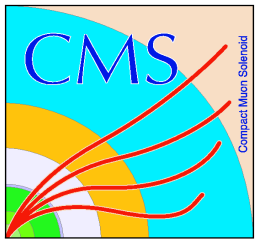
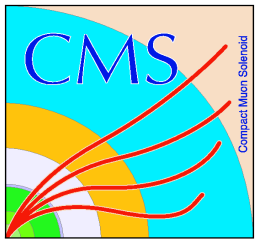


CHEP 2006 Mumbai

PhEDEx – high-throughput data transfer management system

Jens Rehn, CERN

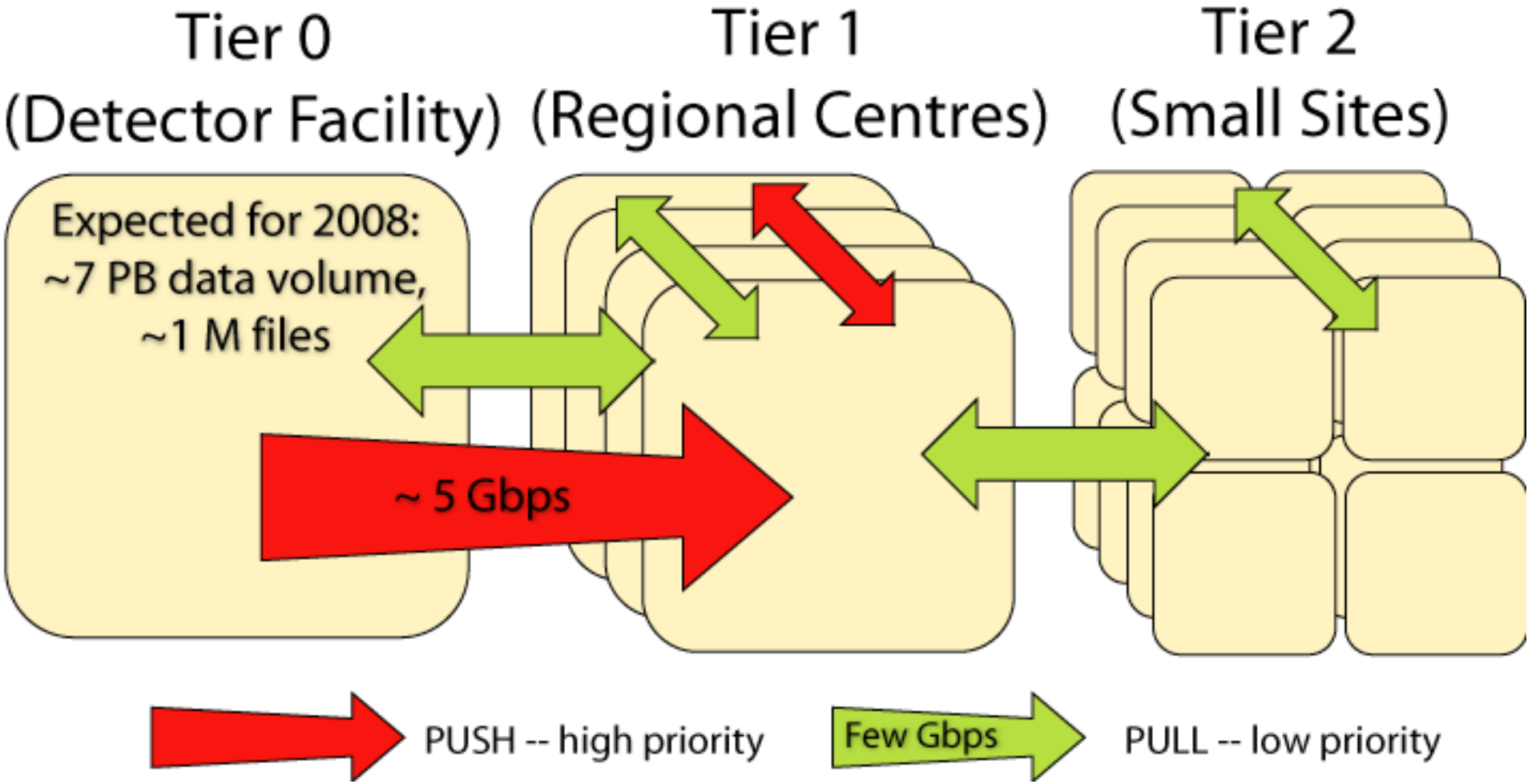
On behalf of numerous
PhEDEx contributors
and the CMS collaboration



Outline

- Introduction to PhEDEx
- Performance and scalability
  - In production environment
  - During recent service challenge
  - In a testbed environment
- Analysis of transfer related problems
- Conclusions and outlook

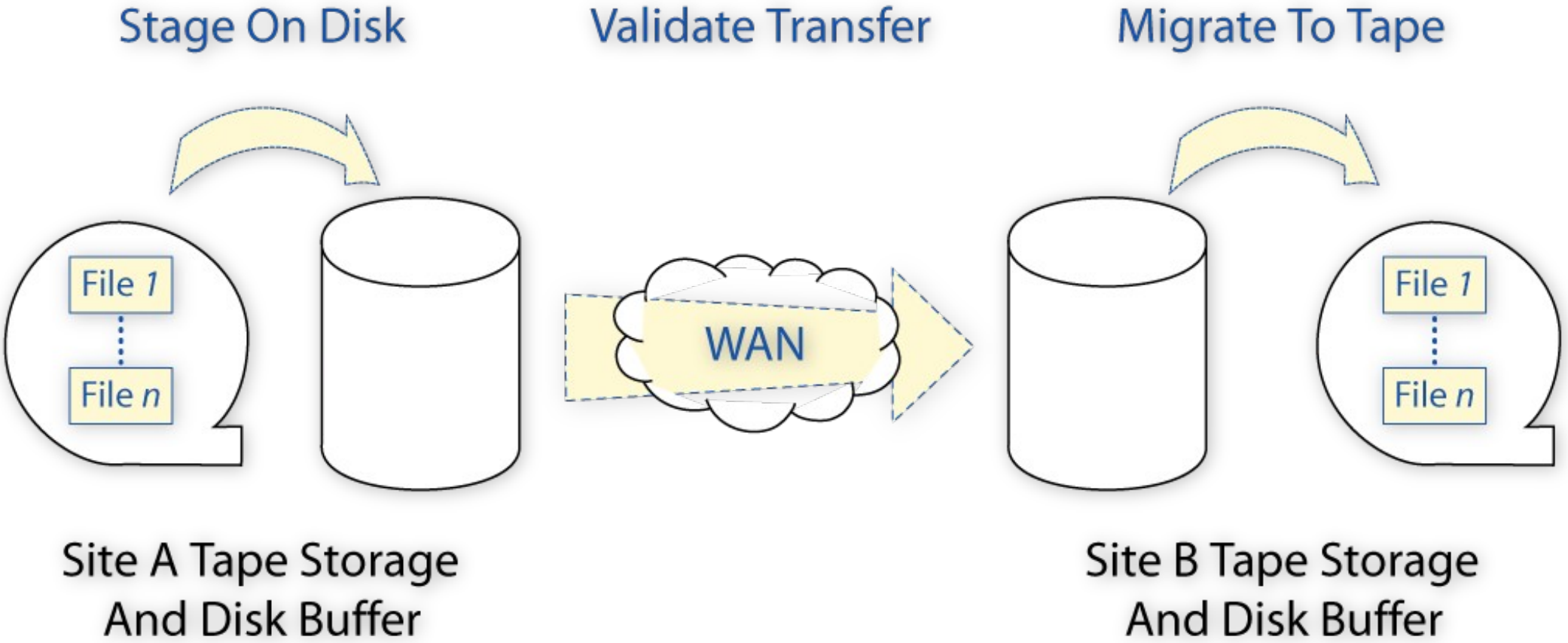
Tiered data flow



See CMS Data Management, P. Elmer (Plenary - 445)

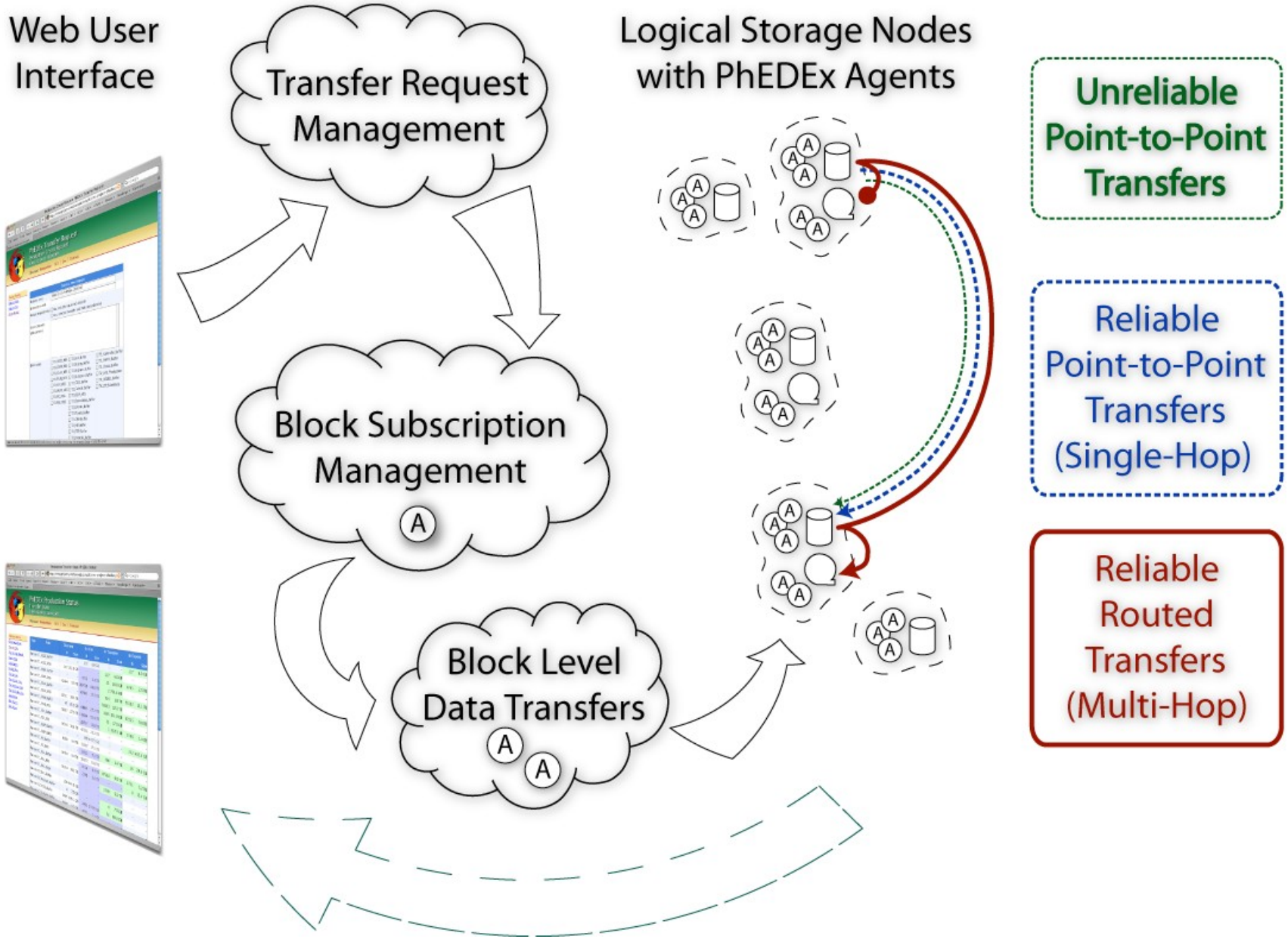
See Distributed Data Management in CMS, A. Fanfani (DEPP - 360)

Traditional HEP data replication

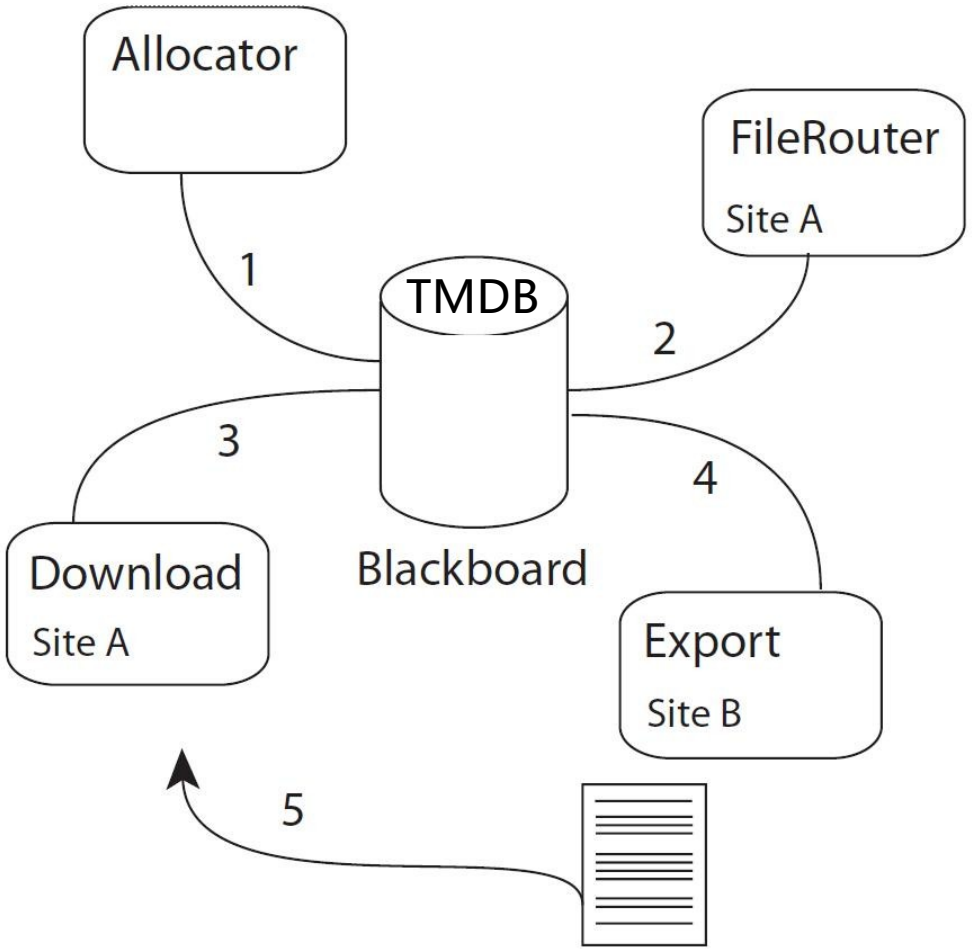


- Each step done by hand
- Manpower-intensive
- Feasible only for small amount of files

PhEDEx data replication

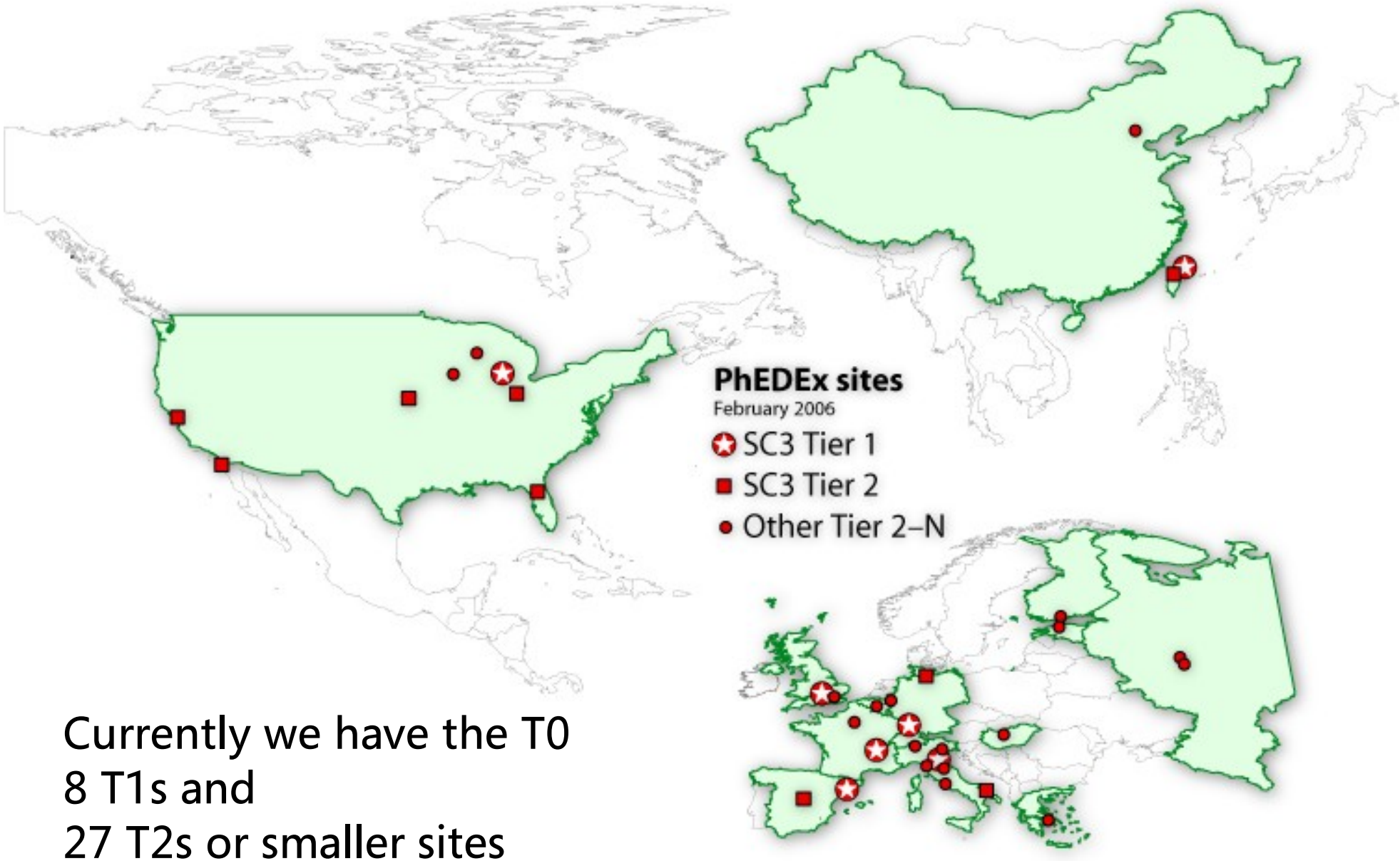


PhEDEx file replication



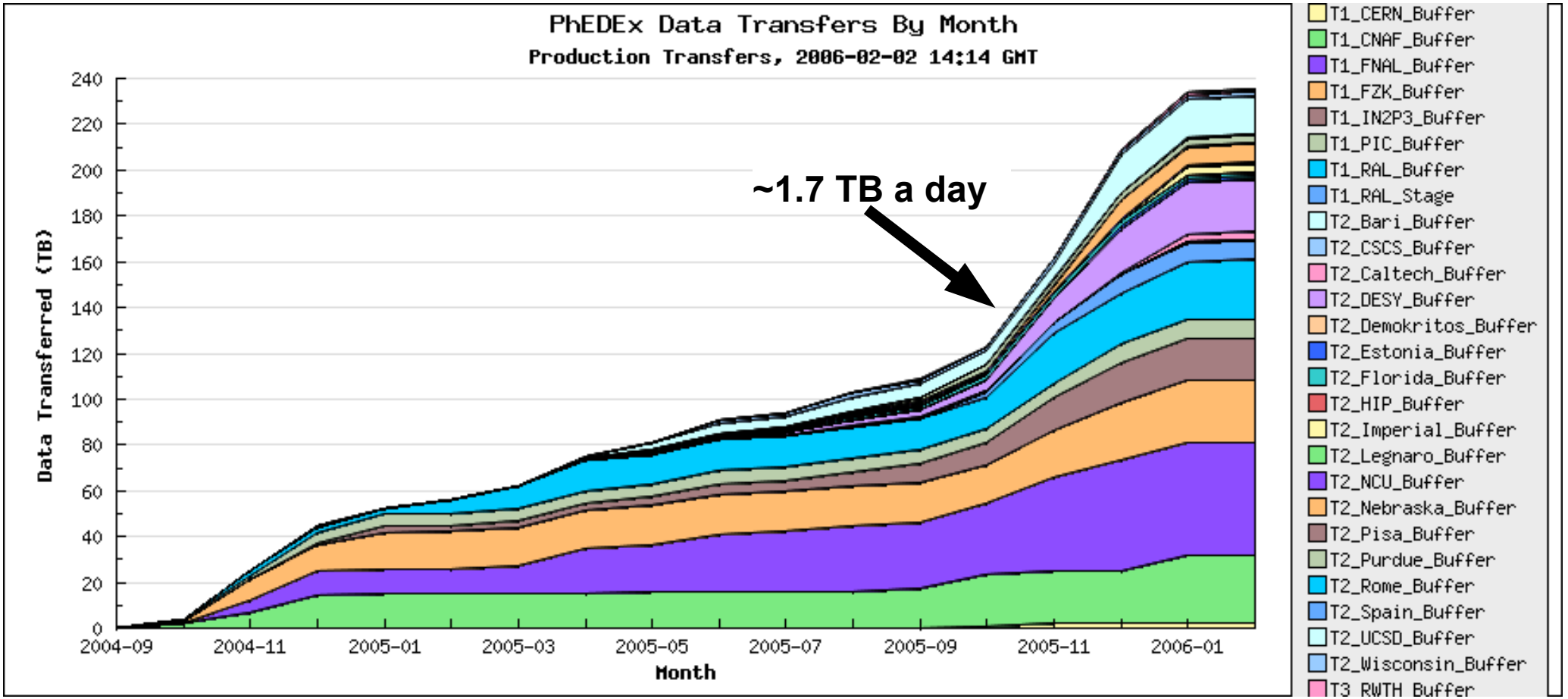
1. Allocator: allocate files to destinations
2. FileRouter: determines closest replica
3. Download: marks files „wanted“ from site B
4. Export: initiate staging and provide contact information
5. Download: transfer file

CMS data distribution network

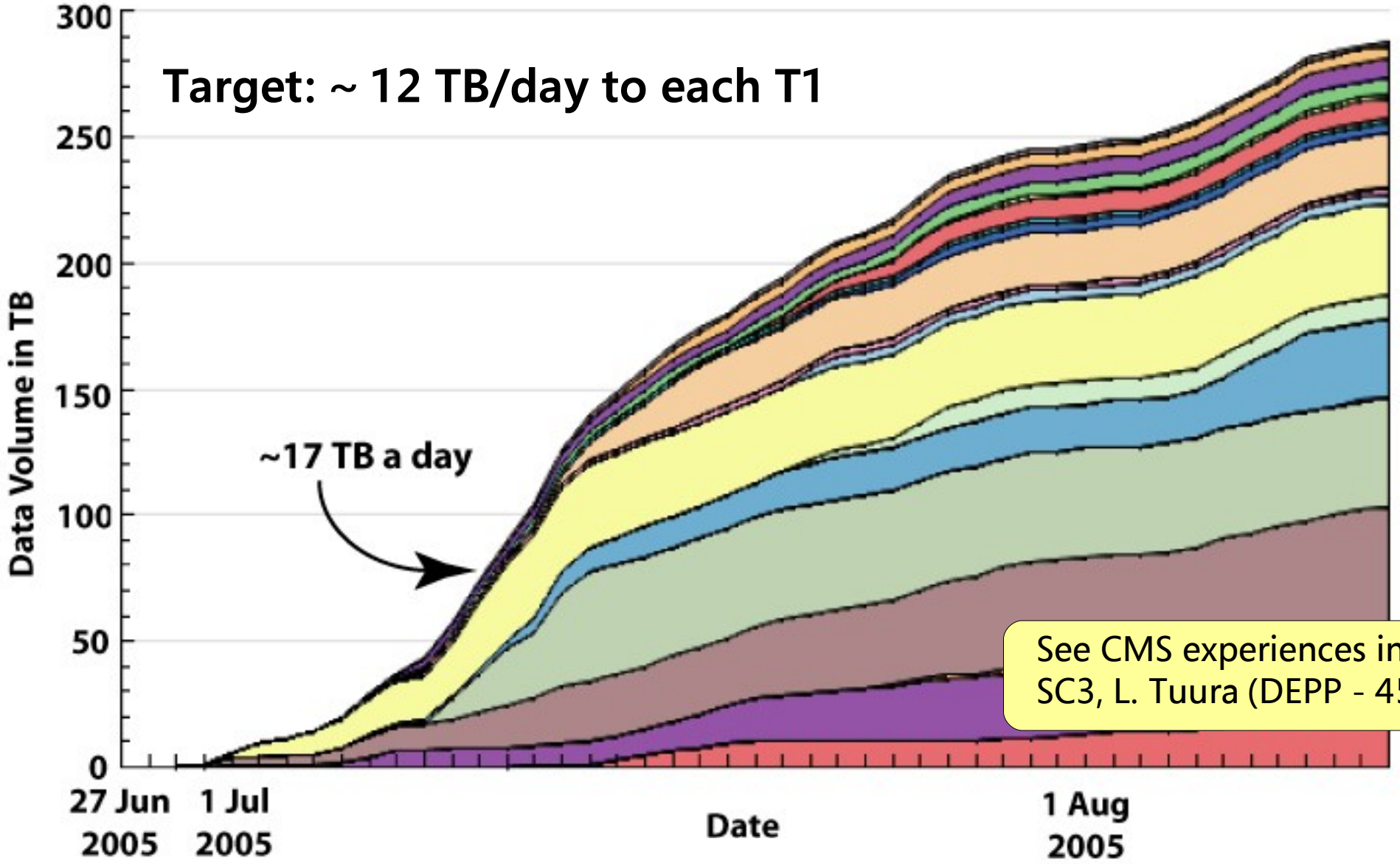


Currently we have the T0
8 T1s and
27 T2s or smaller sites

PhEDEx transfer volume
Last 17 months



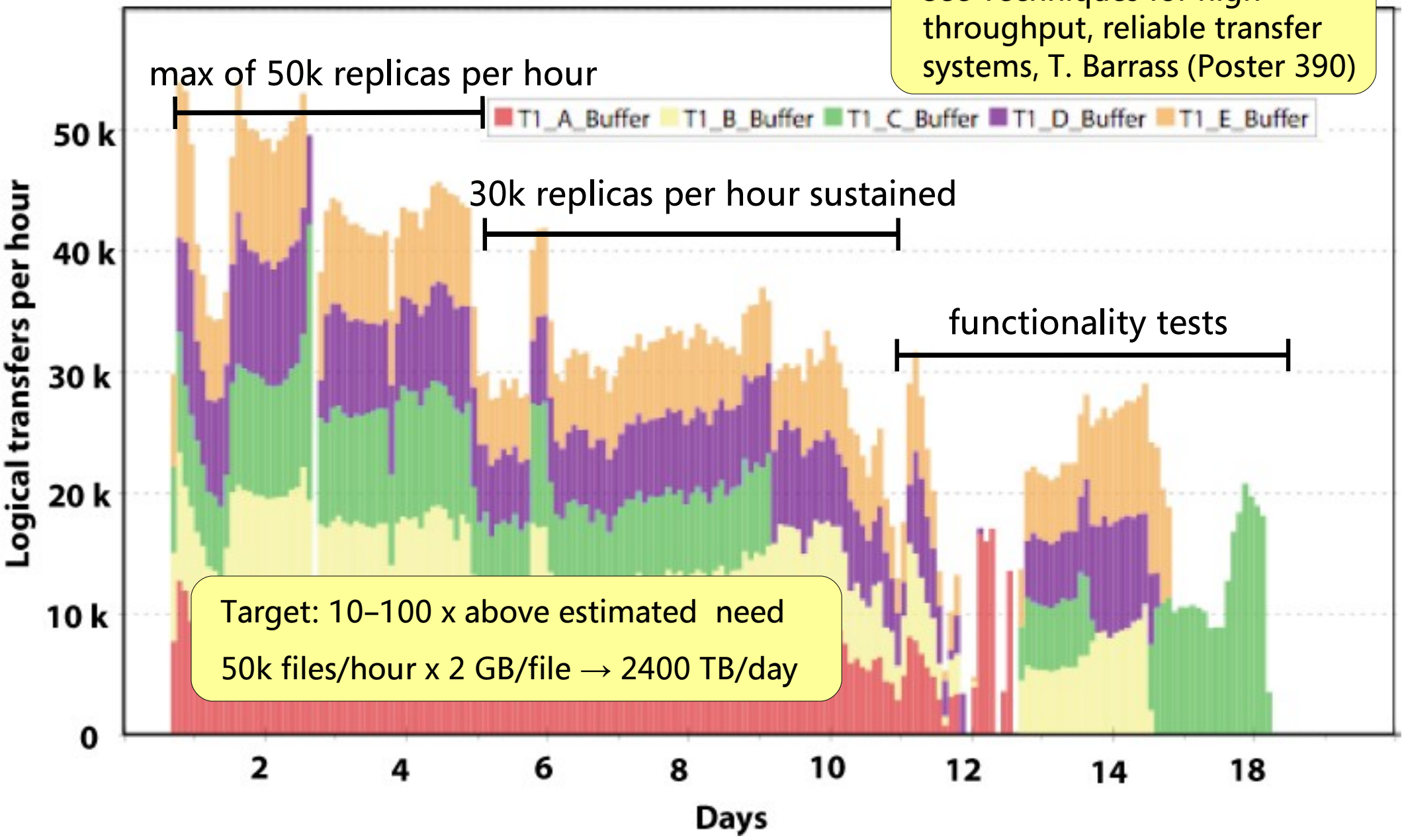
PhEDEx in LCG SC3 throughput phase



See CMS experiences in LCG SC3, L. Tuura (DEPP - 453)

PhEDEx scalability exercise

See Techniques for high-throughput, reliable transfer systems, T. Barrass (Poster 390)

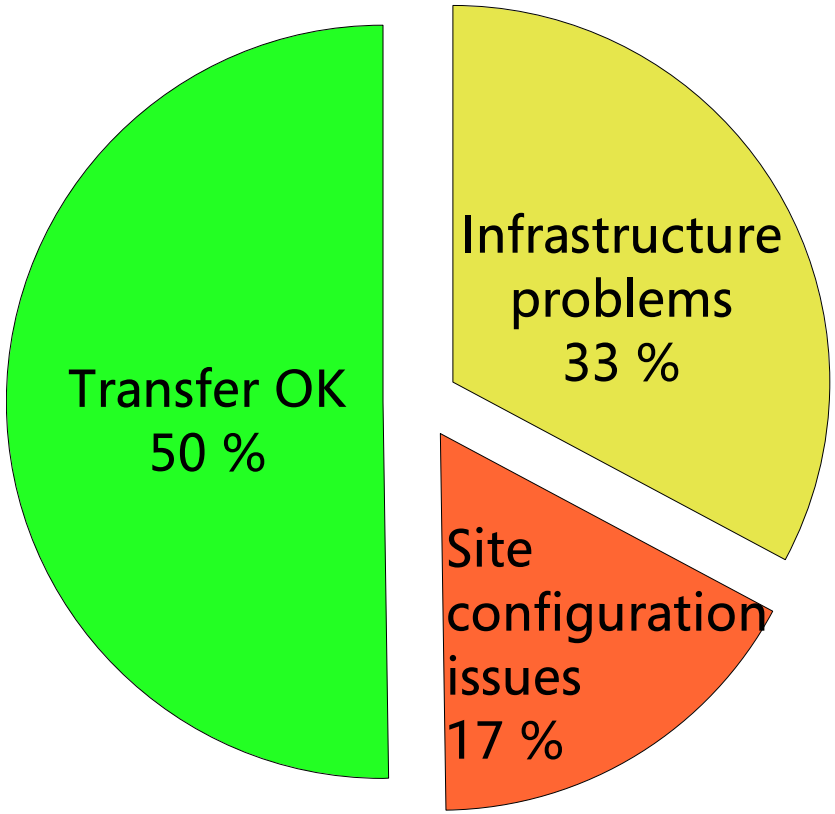


Target: 10–100 x above estimated need

50k files/hour x 2 GB/file → 2400 TB/day

Reliability: impossible odds?

Case study: transfer level

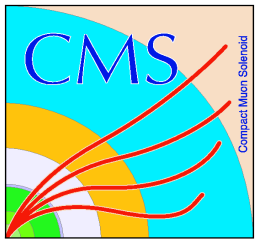


- High failure rate on new SRM/storage infrastructure
- 50% of the transfers successful on the first try
- Main problems
  - Configurations changed or wrong at sites
  - Problems related to network or storage infrastructure

Reliability: against the odds!

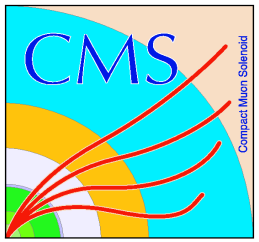
Case study: after PhEDEx failure recovery

- All failures recovered, eventually
- Files retransferred
- No data lost :-)
- Recovery fully automatic
  - Absolute must: in 2007 CMS will transfer ~2-10k files per day
  - Manual recovery infeasible: 1 ‰ permanent error rate ≃ 2 hrs daily maintenance



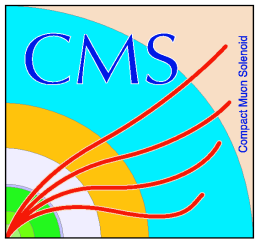
PhEDEx deployment

- Each site runs agents close to their storage
  - Modest resource requirements
  - Usually hosted on CMS-dedicated server
  - PhEDEx and tool installation with XCMSi
  - Underlying transfer utilities like srmcp, fts, etc.
  - Grid services: certificates and proxy renewal
  - Configuration: site registration and site specific settings
- Operated by local CMS community, in close communication with site's adminstrators



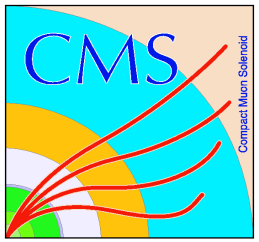
Summary and outlook

- PhEDEx provides
  - Reliable and scalable data distribution on the grid
  - Flexibility to use any file replication tool, especially grid tools
  - Real life monitoring through a web status display
- Plans
  - More and improved web based data management tools
    - data subscriptions, transfer requests, agent management, deployment
  - Support transfers for physics groups and individual physicists
    - decentralisation of central database
- Hope to bring many new sites onboard :-)



Useful links and contacts

- PhEDEx project web page:
  - http://cern.ch/cms-project-phedex
  - Links to documentation, monitoring and CVS repository
- PhEDEx mailing list:
  - cms-phedex-developers@cern.ch
  - Shortly hn-cms-phedex@cern.ch



Related talks and posters

- Techniques for high-throughput, reliable transfer systems: break-down of PhEDEx design – T. Barrass (Poster-390)
- CMS experience in LCG SC3 – L. Tuura (DEPP-453)
- CMS Data Management – P. Elmer (Plenary-445)
- Distributed Data Management in CMS – A. Fanfani (DEPP-360)
- Italien Tiers hybrid infrastructure for large scale CMS data handling and challenge operations – D. Bonacorsi (DEPP-288)