CMS Computing Model.
- Data Distribution and Data Access-

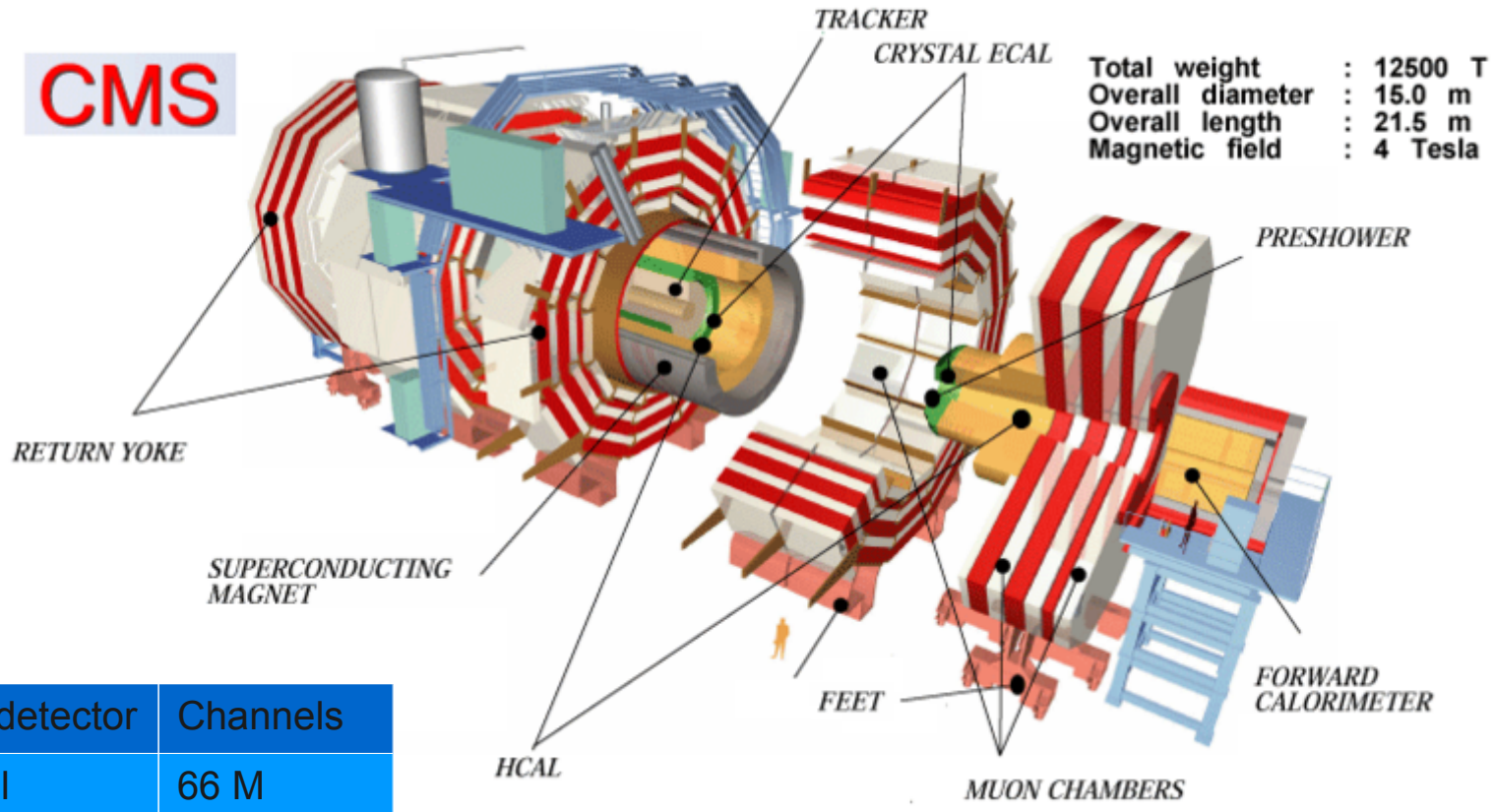
GridKa School 2010
- HEP Track -

Christoph Wissing
(DESY)

HELMHOLTZ ASSOCIATION

DESY

CMS Detector.



Subdetector	Channels
Pixel	66 M
Si Strip	9.3 M
Muons	890 k
Preshower	140 k
ECAL	76 k
HCAL	9 k

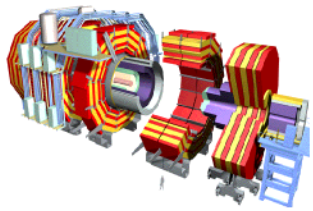
RAW data:
1.5 MB/ev
2 MB/ev (MC)

RECO data:
500 kB/ev

AOD data:
150 kB/ev

Logging: 300 Hz

CMS Dataflow.



Prompt Reconstruction

TIER-0

CAF

Calibration
Express-Stream
Analysis

CASTOR

600MB/s

50-500MB/s

TIER-1

TIER-1

TIER-1

Re-Reconstruction
Skims

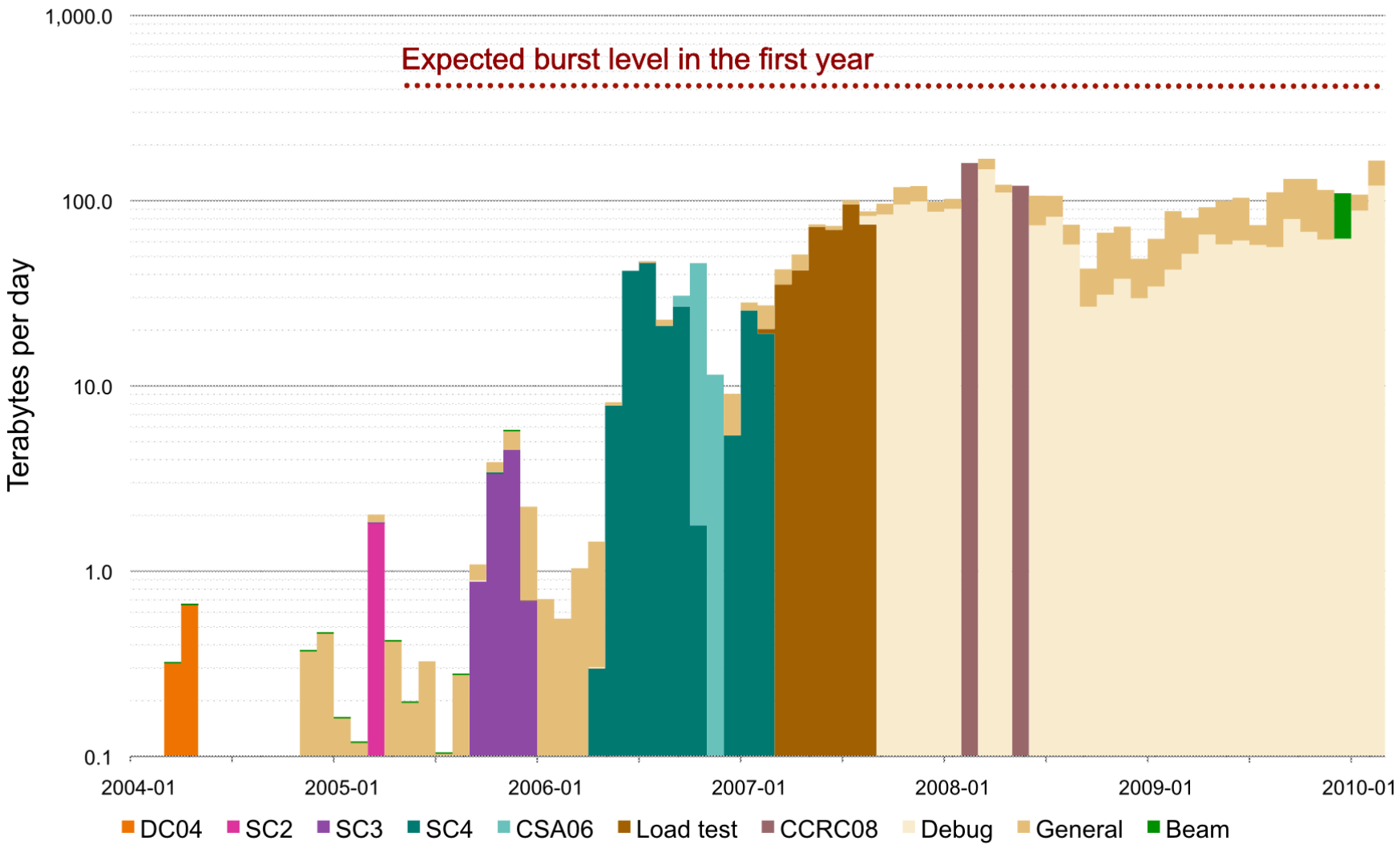
50-500MB/s

~20MB/s

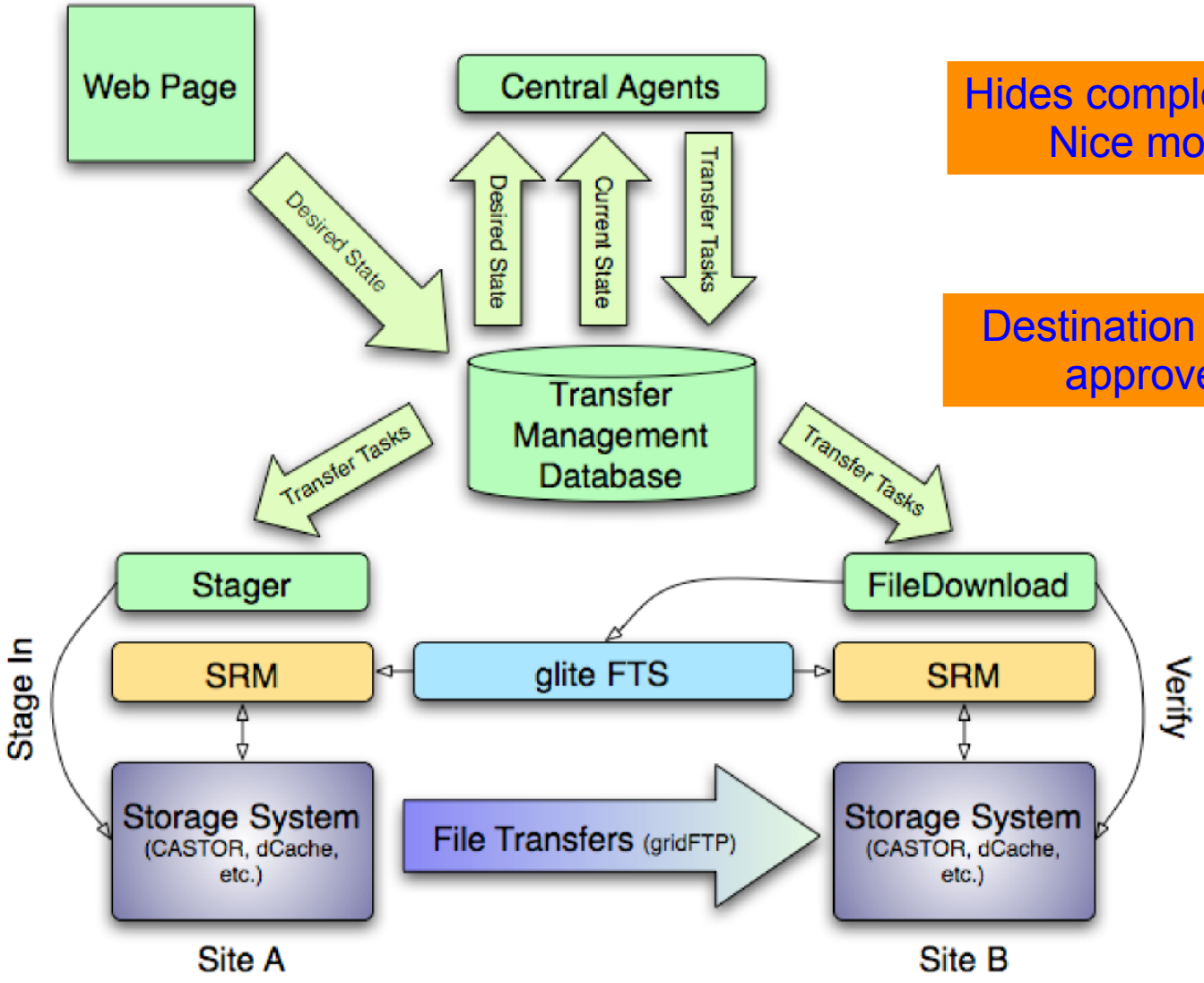
Simulation
Analysis

TIER-2

History of CMS Data Transfers.



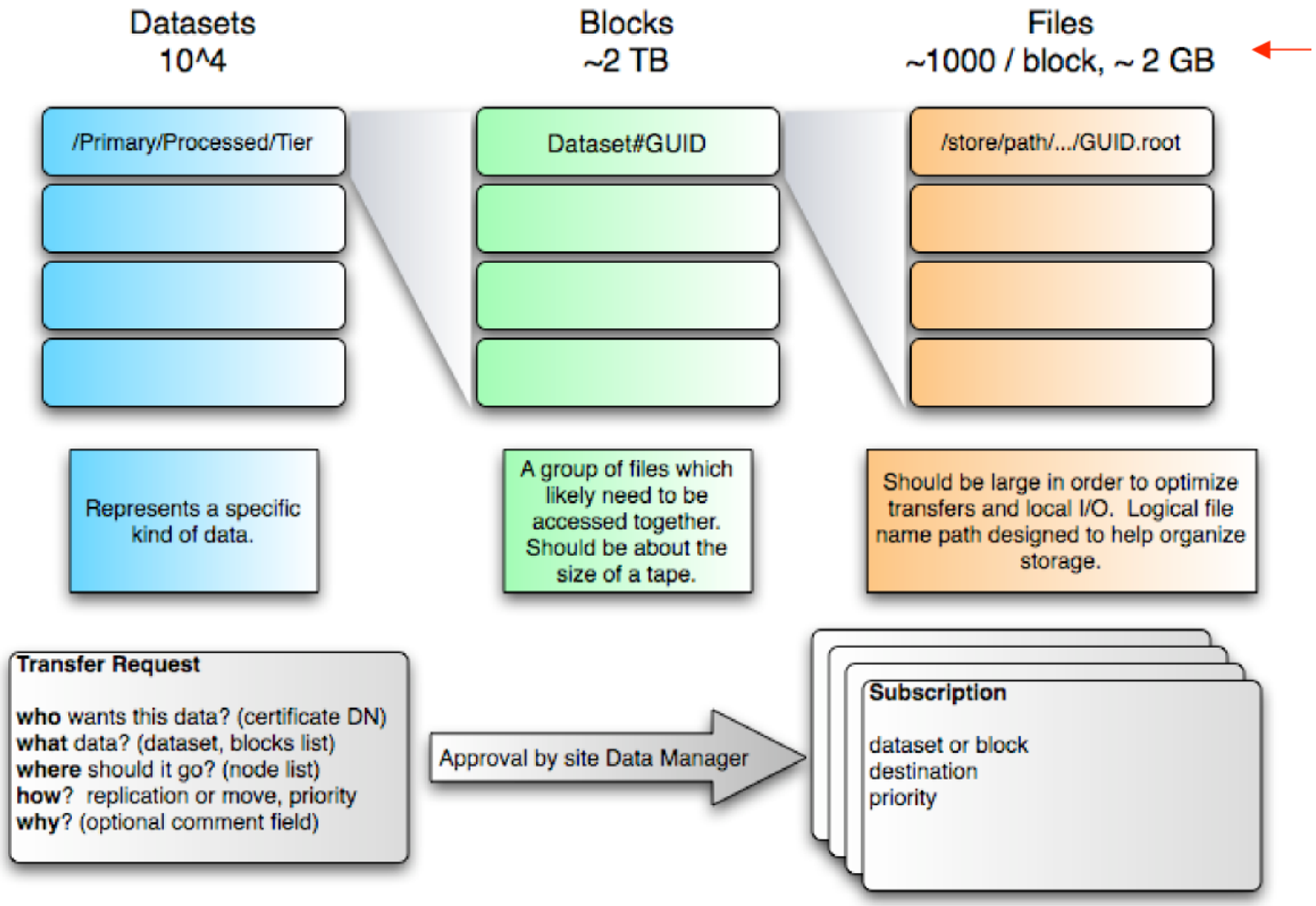
CMS Data Transfer Tool: Phedex.



Hides complexity of FTS
Nice monitoring

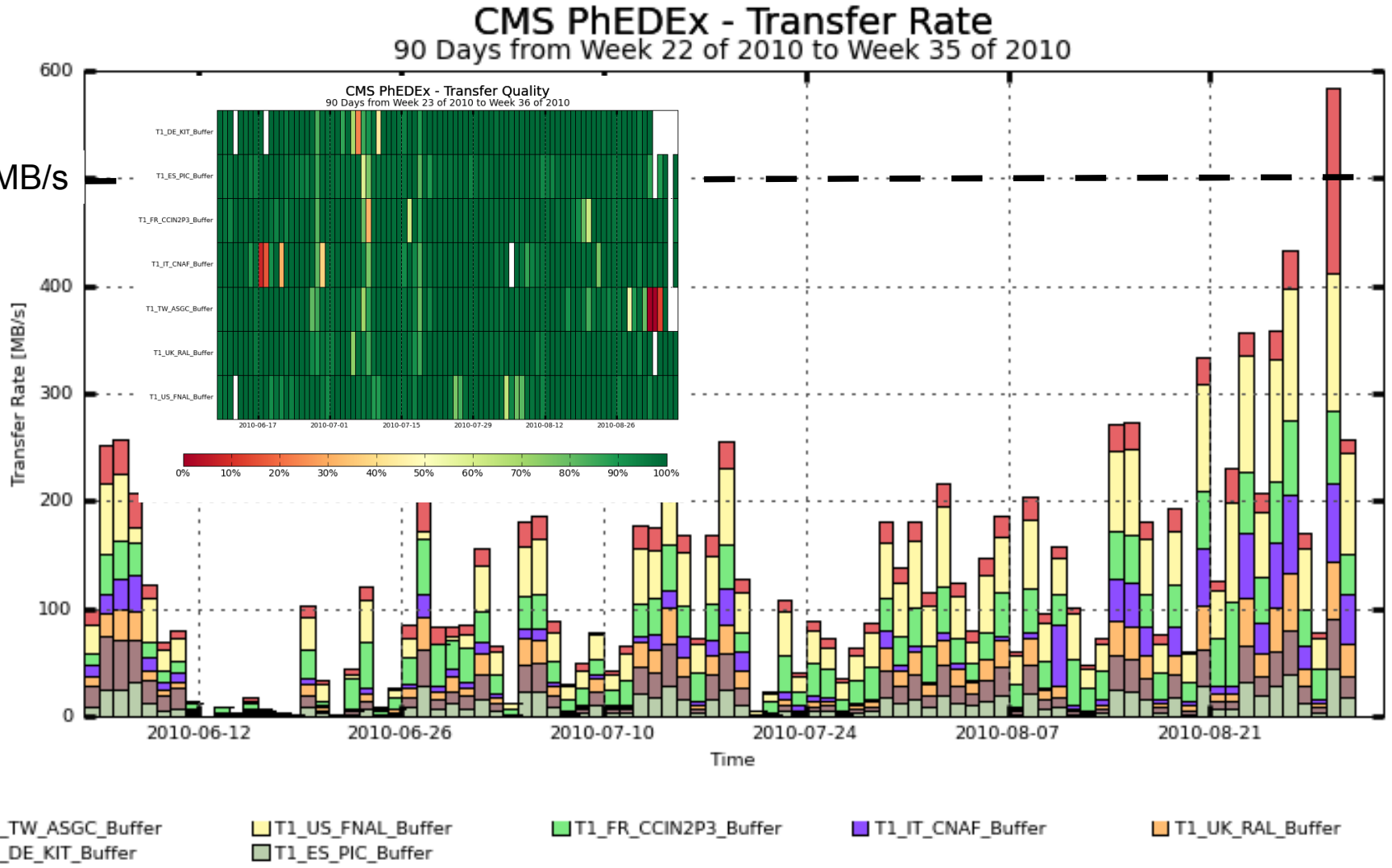
Destination "data manager" approves requests

Data Organisation.

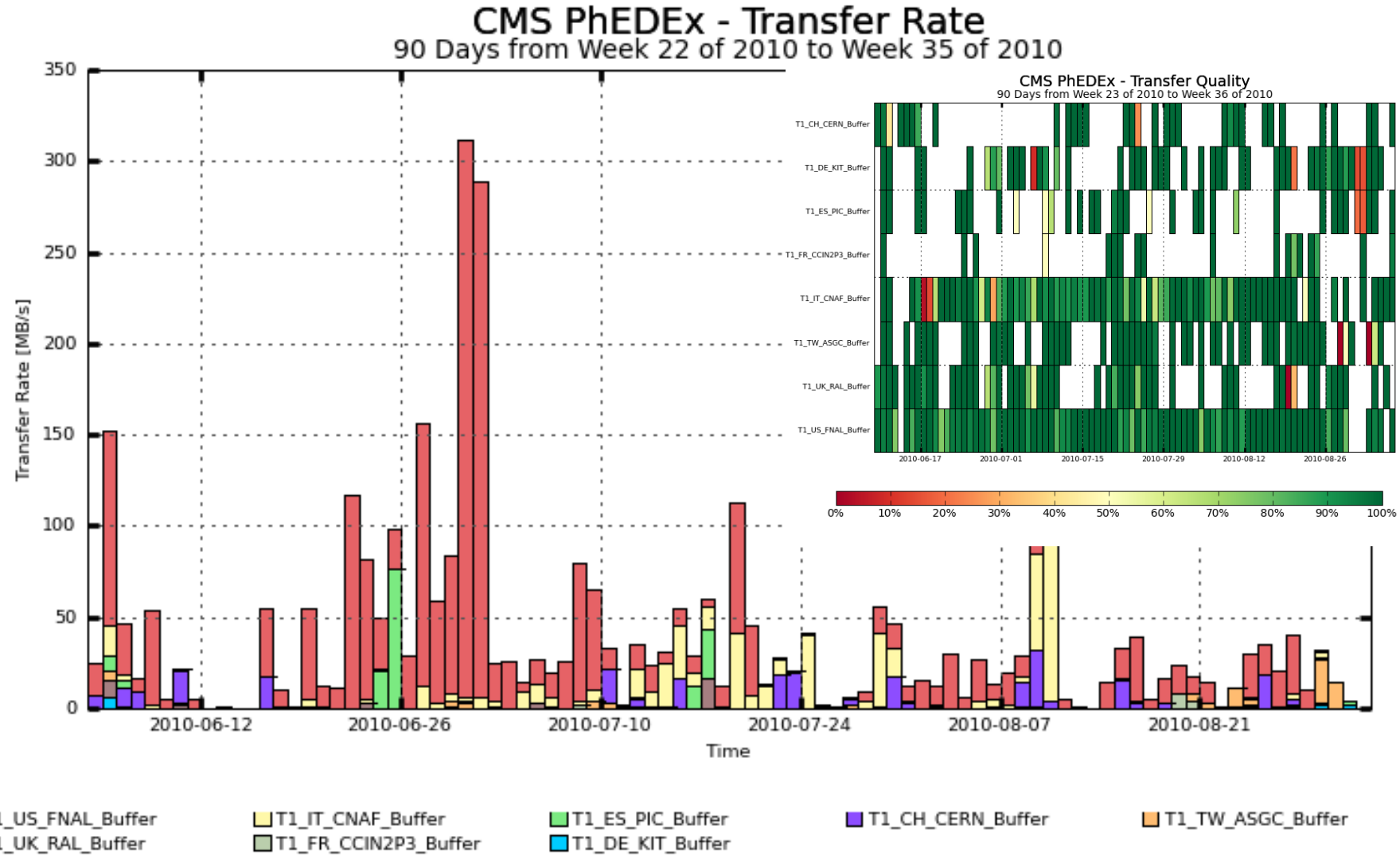


T0 Export.

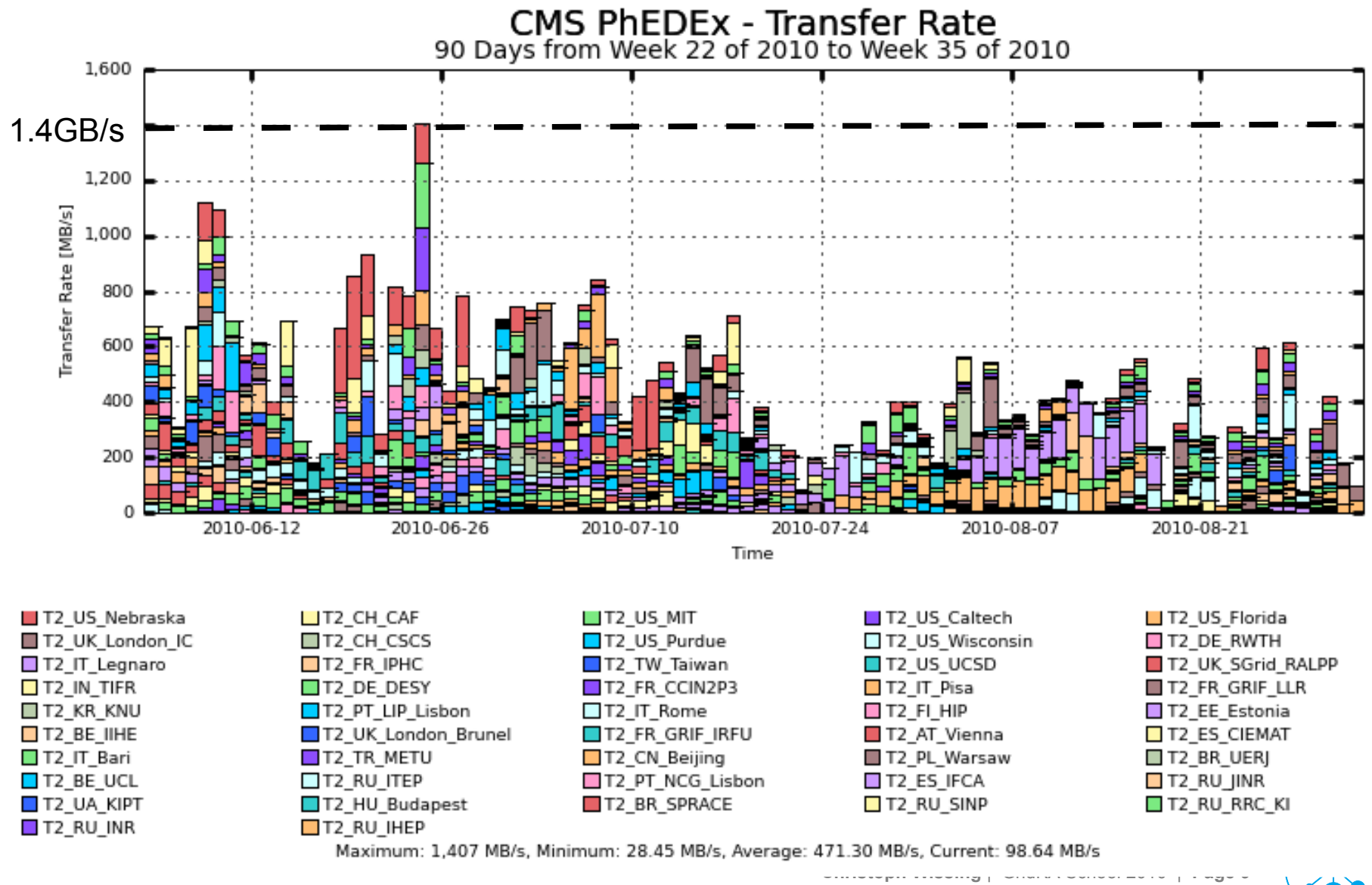
500MB/s



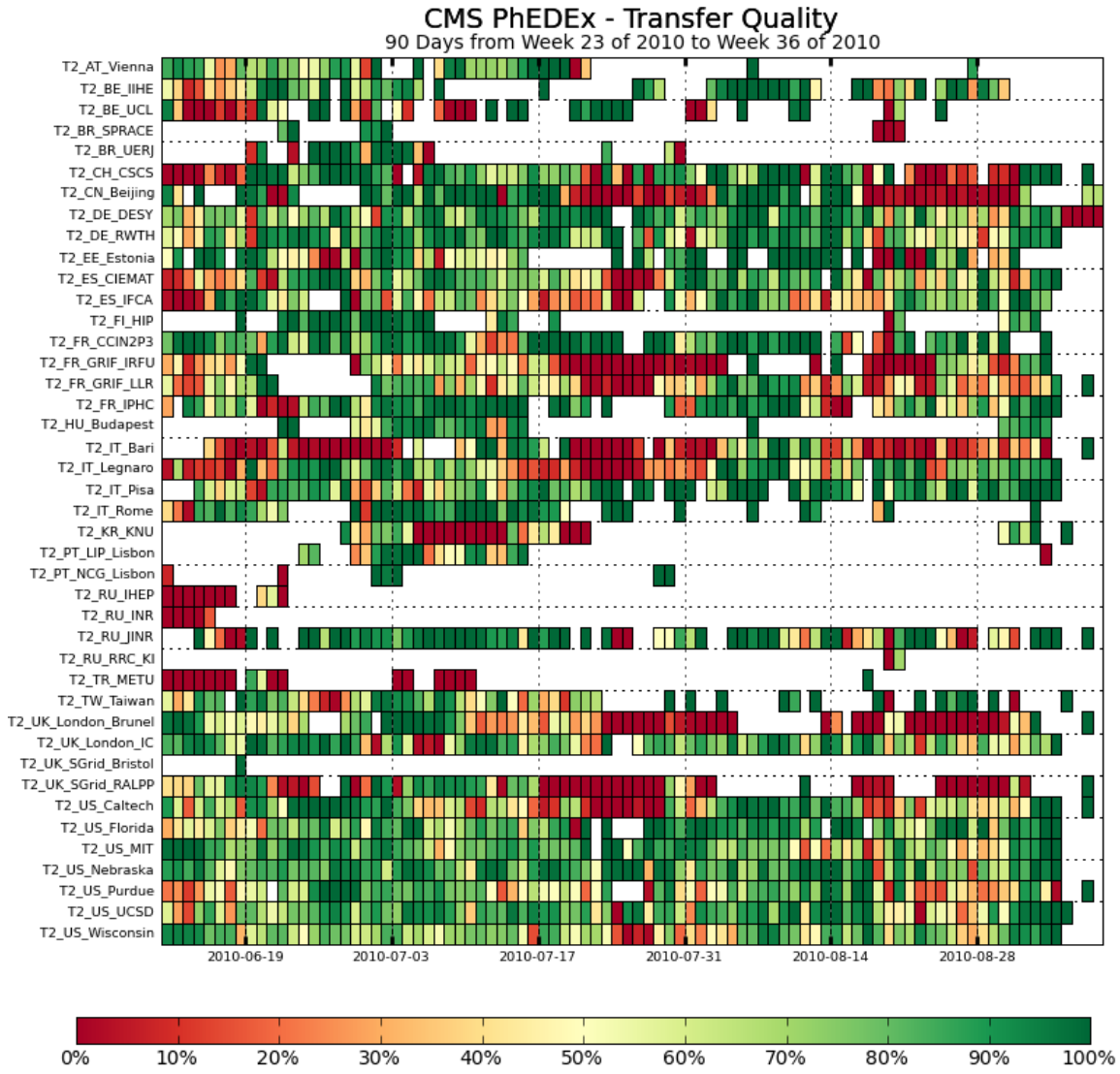
Tier-1 to Tier-1 Transfers.



Tier-1 to Tier-2 Transfers.



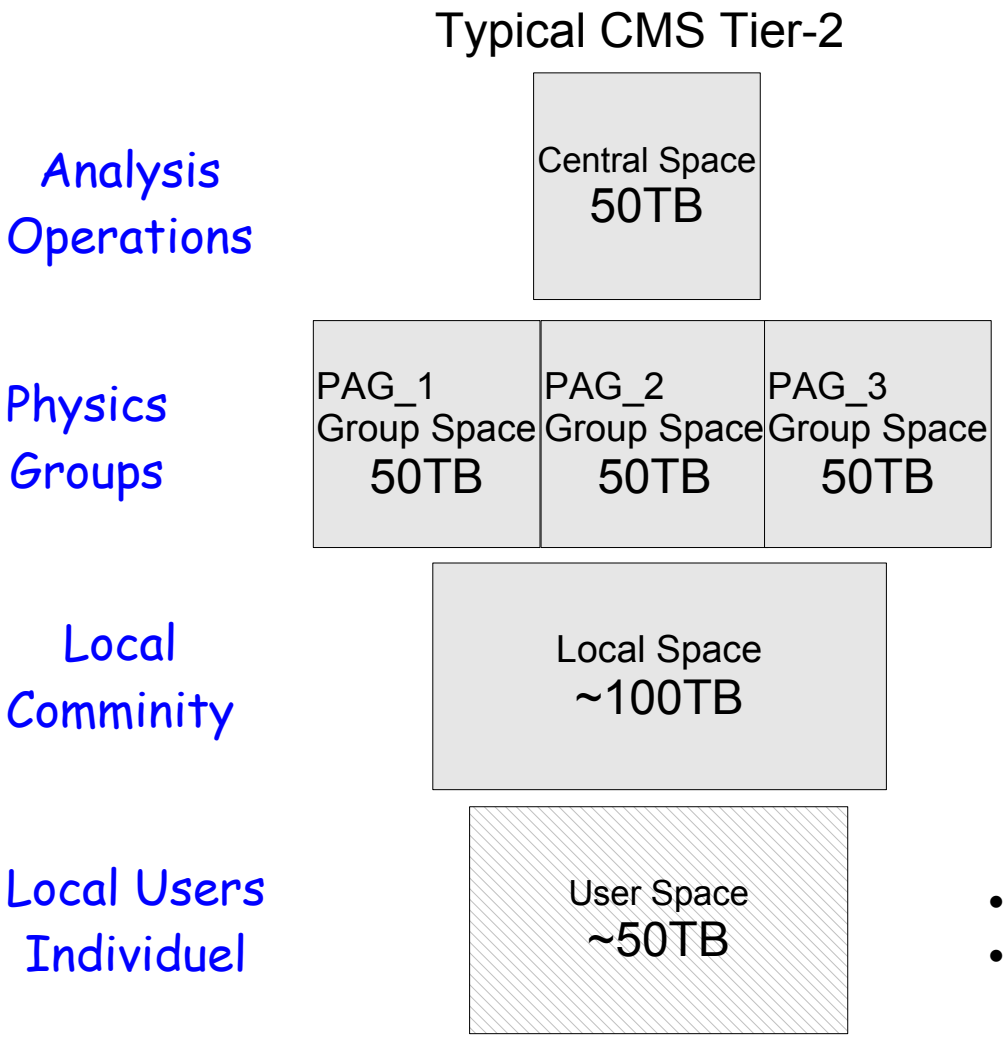
Tier-2 to Tier-2 Transfers.



CMS allows T2 to T2

Not as green as Transfers from/to Tier-1

Organisation of Tier-2 Storage.



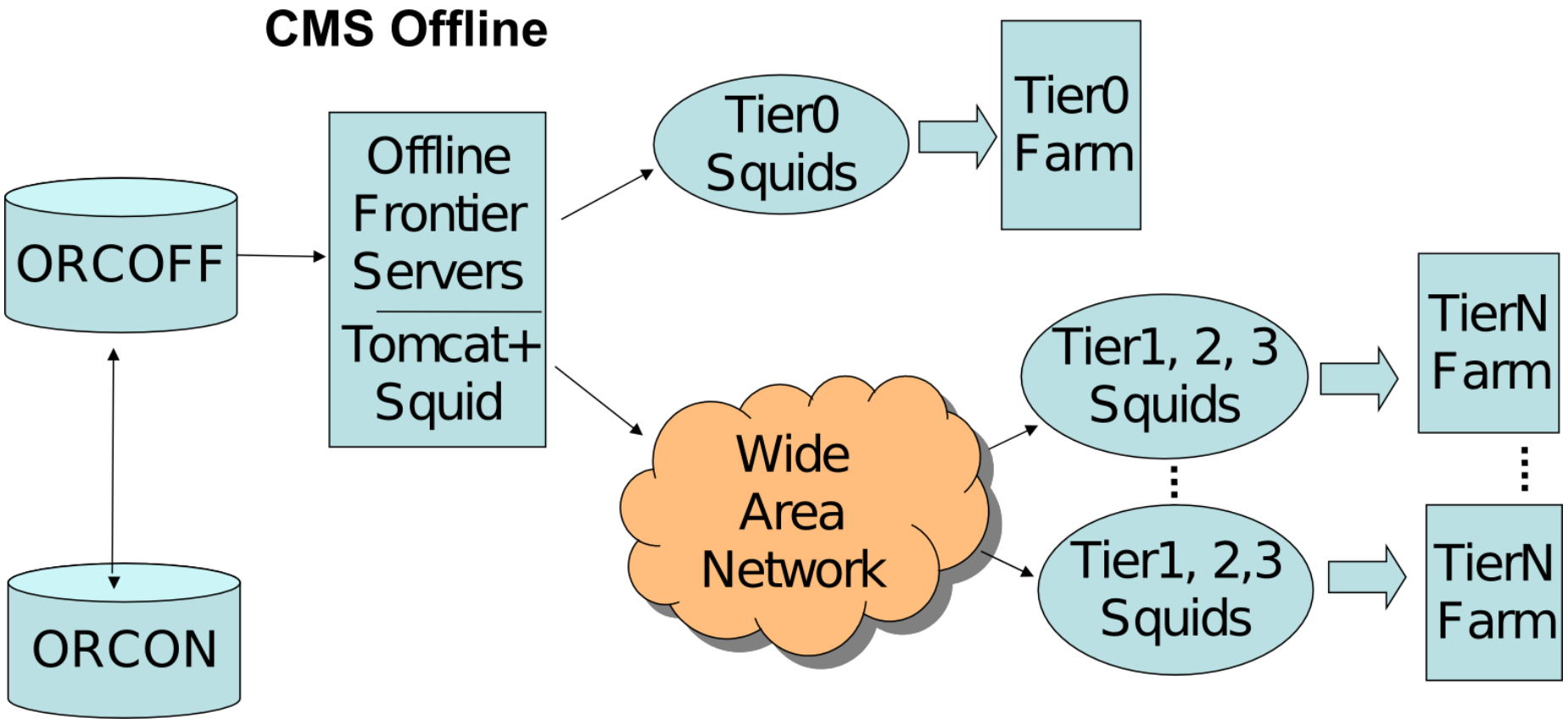
- Total Central Space: 2.2PB
- Good distribution of analysis samples
- Number of replicas ~ attractivity

- Each group supported by at least 2 Tier-2 sites
- Total space used by groups: 2.5PB

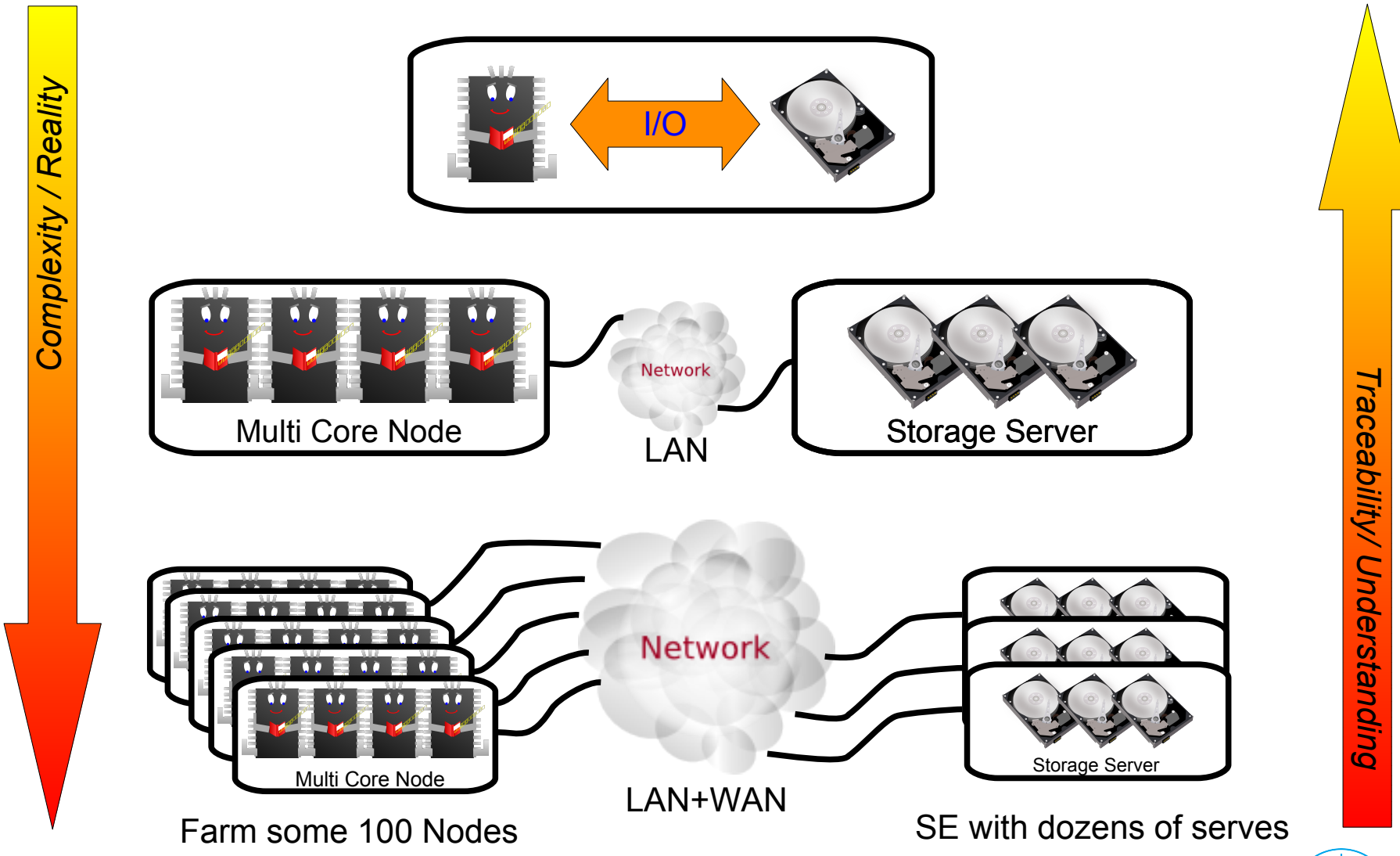
- Users assigned to a “home Tier-2”
- 1-1.5TB per user (suggested by CMS)

Conditions Data Distribution.

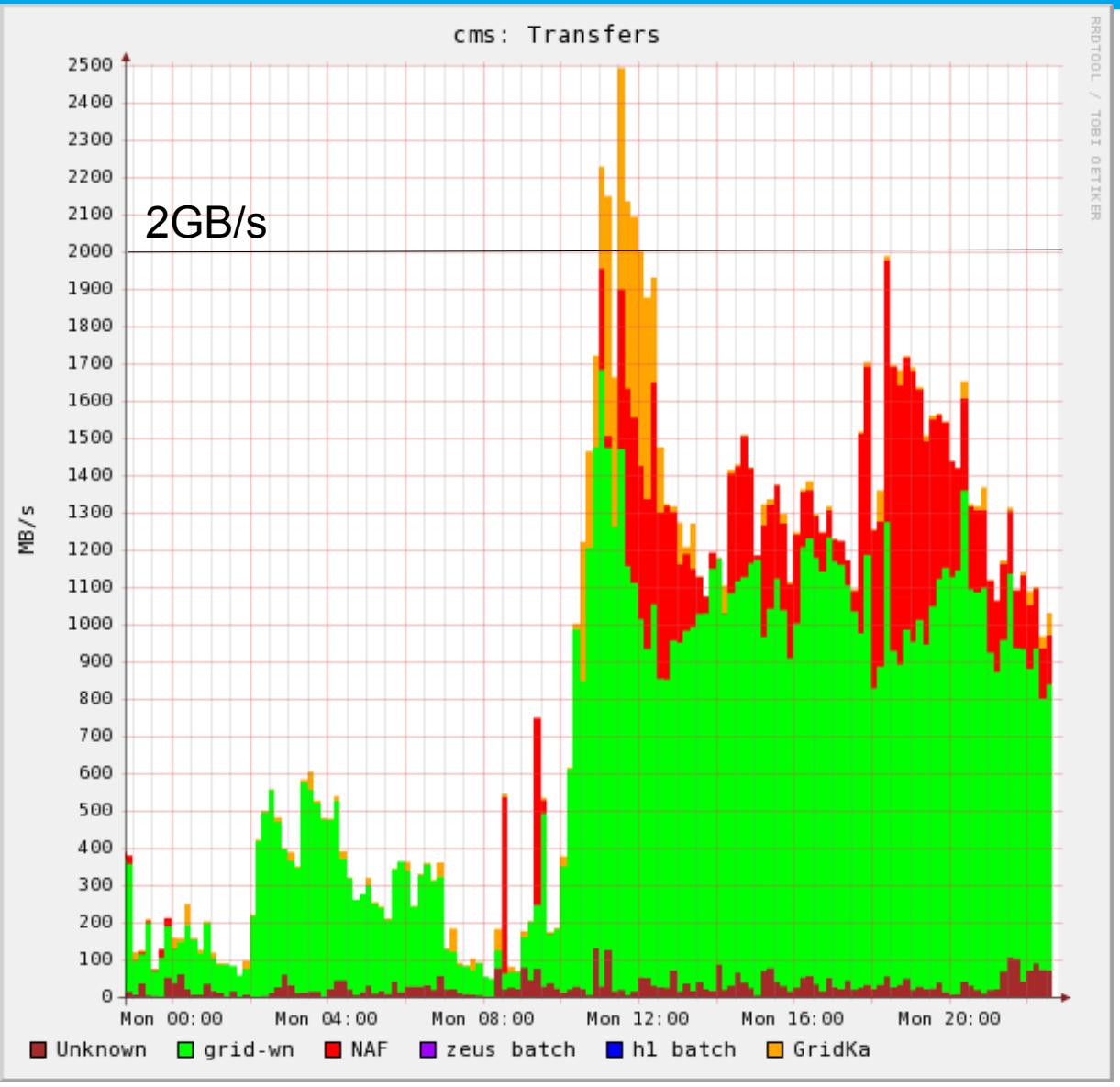
HTTP Squid Caches
- Rather easy to operate
- Good scaling: Larger sites run several squids

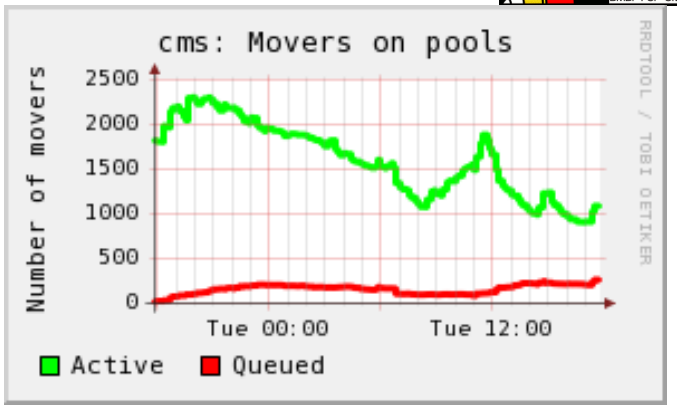


Data Access Pattern.



Oberserving Data Accesses (High Level View).

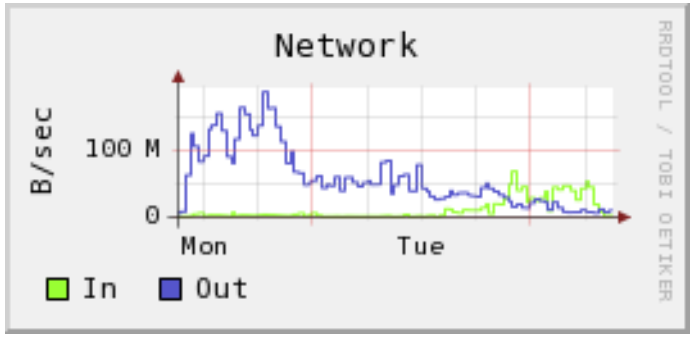


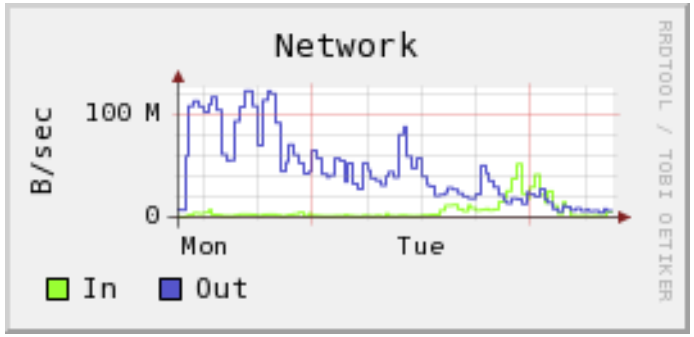


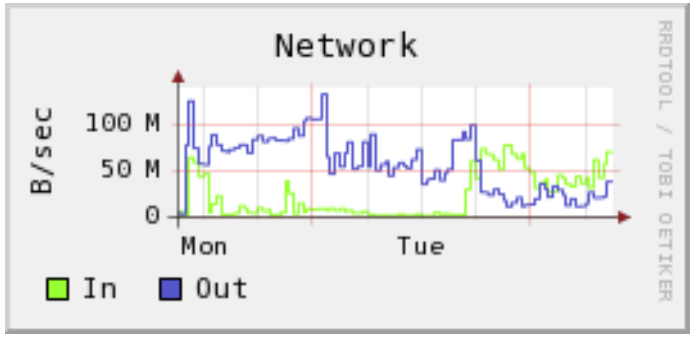
- Typical Monday morning
  - Over 2000 mover
  - WAN + LAN transfer
- Assume 20 servers
  - 100MB/s each
  - 10 GigaBit should be faster
- It's all not easy

Oberserving Data Accesses (Lower Level View).

Typical Networkserver:

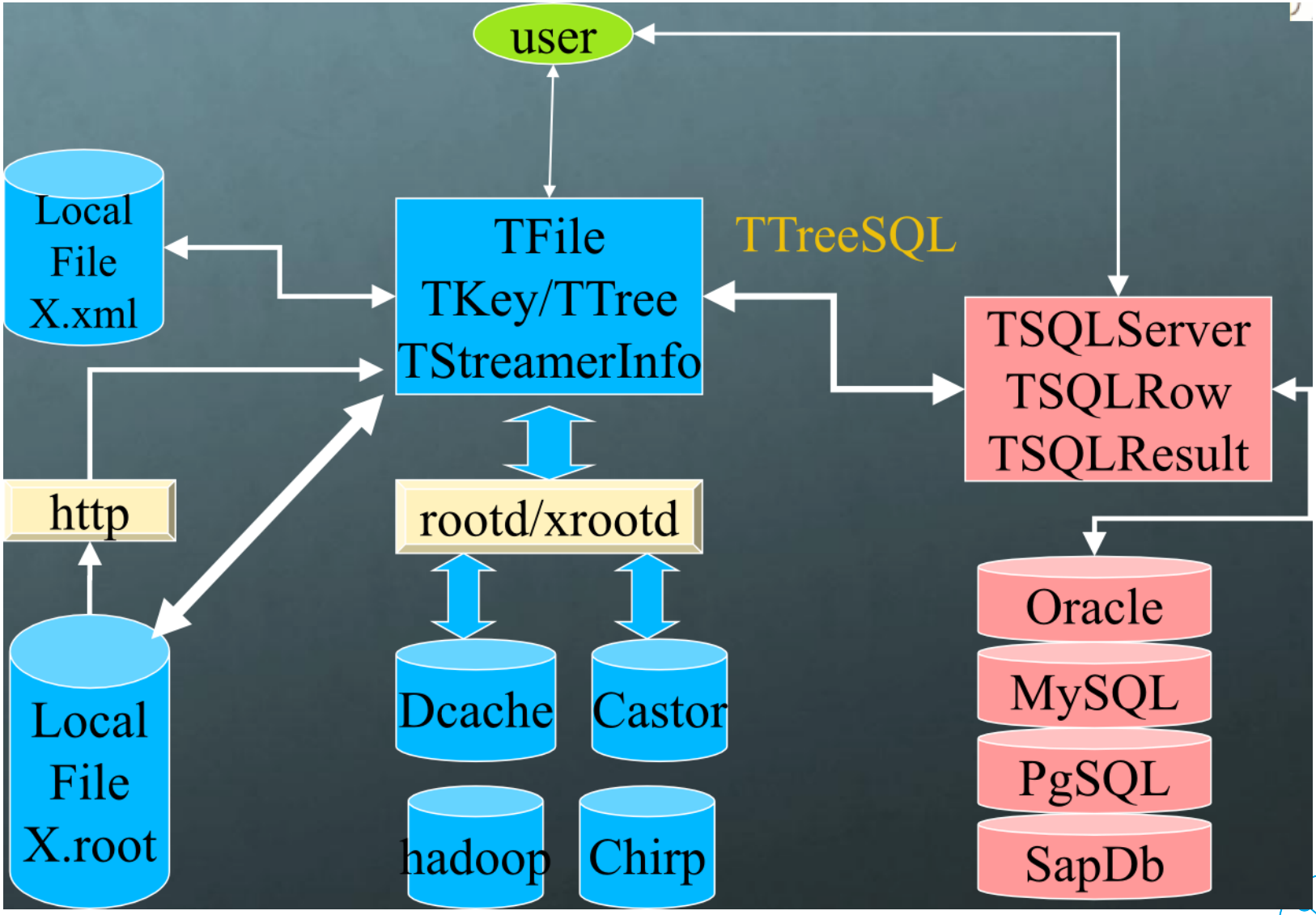




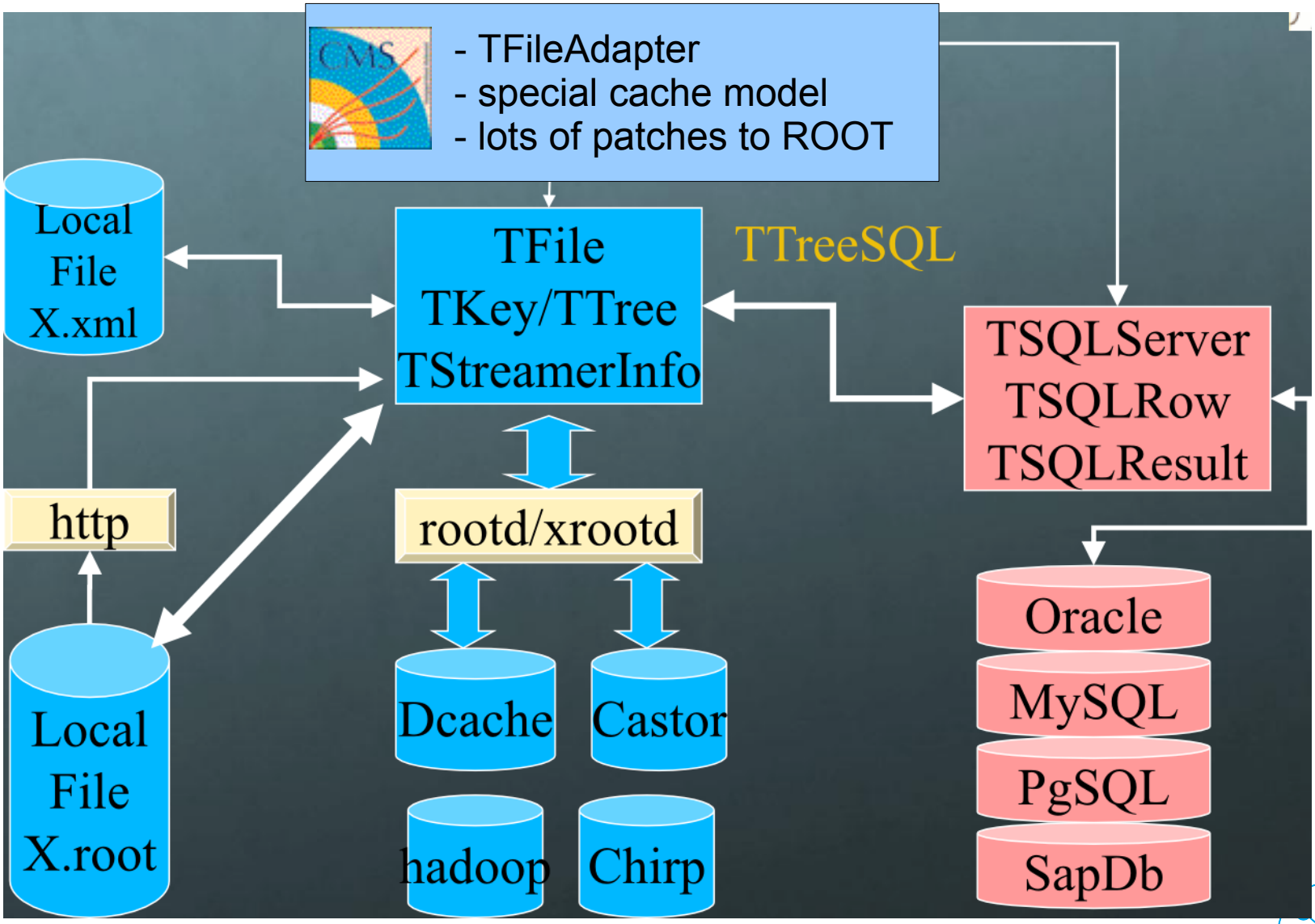


- 1GE Networking
  - ~100MB/s expected
  - “Easy” to saturate
- 10GE Networking
  - Up 1GB/s should be possible
  - Not easy to saturate
  - Bottleneck now somewhere else
    - Underlaying storage system
    - Sensitive to RAID setup
- Difficult to tune for all use cases
  - Streaming more easy
  - Seeking due to random access is expensive

ROOT I/O is rather complex....



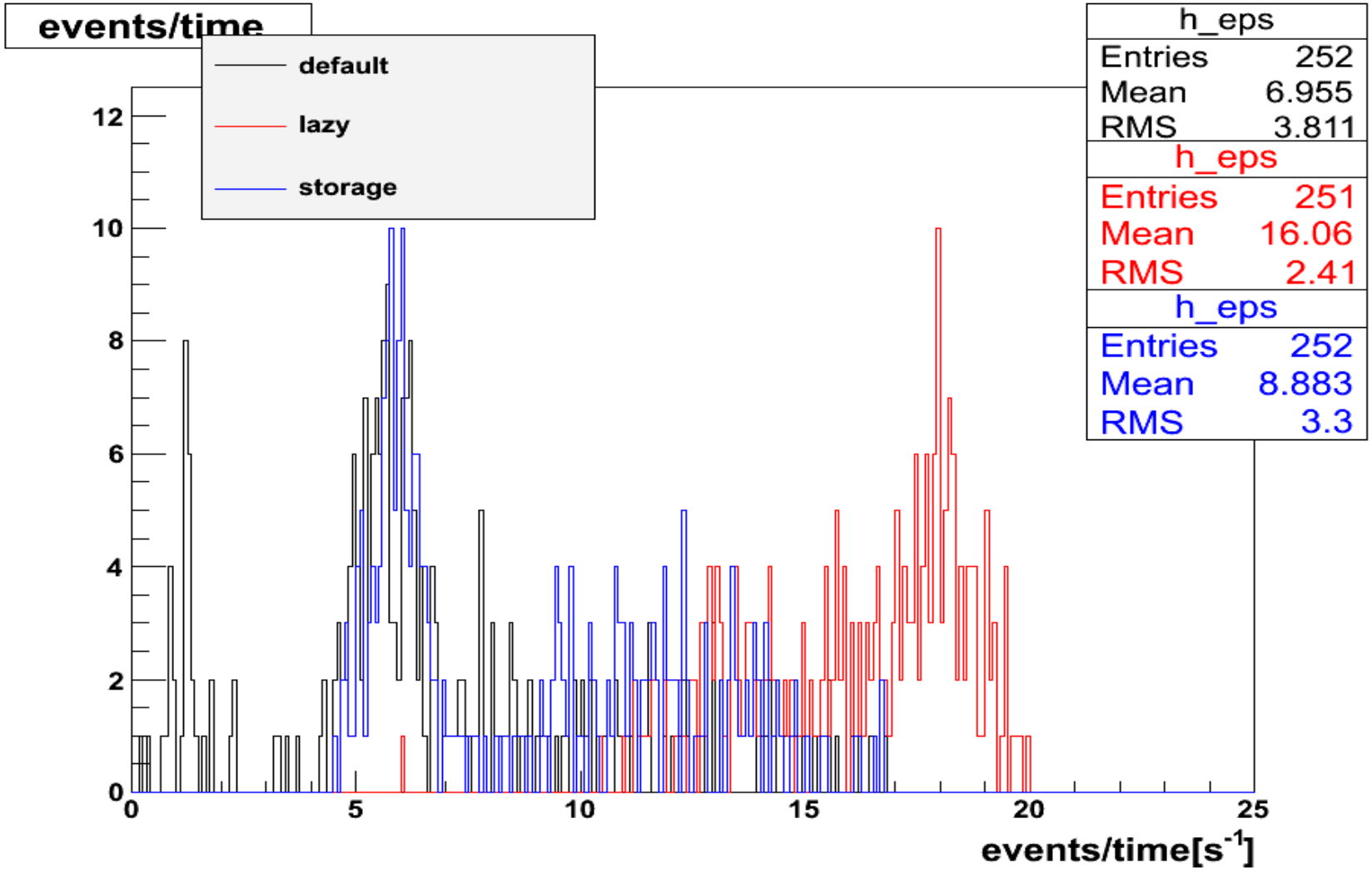
...but not complex enough for CMS.



I/O Layer Steering.

- Various parameters in TFileAdapter
- CacheHint
  - Application only: ROOT does the caching
  - “Lazy download”
    - Chunks of several MBs copied to locat scratch disk
    - Reading and seeking on local disk
  - Storage only: ROOT only drives the caching, but actually does not
- ReadHint
  - Direct-unbuffered
  - Read-ahead-buffered
    - Buffer size can be configured
- Big parameter space
- No set fits for all storage solutions
- Each job type likes other settings

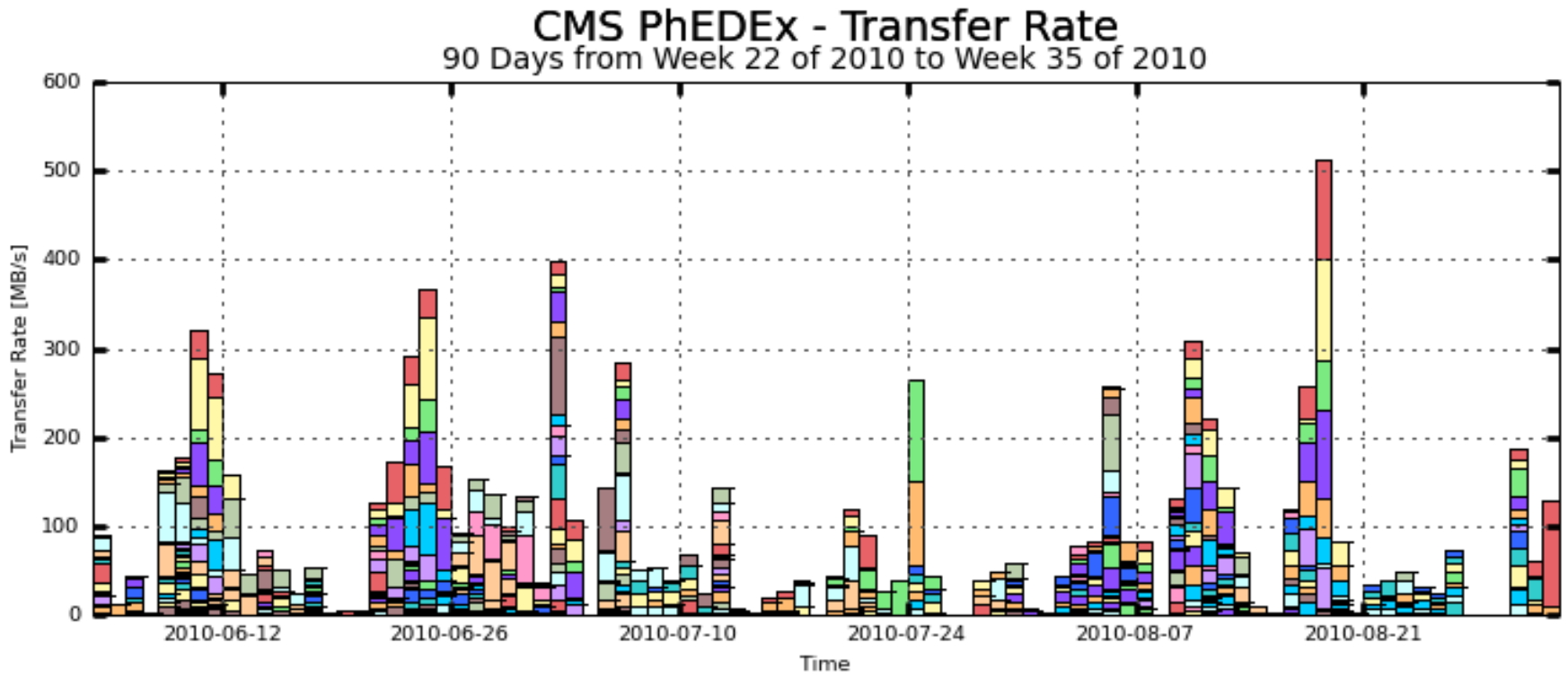
CacheHint Parameter Matters.



Summary.

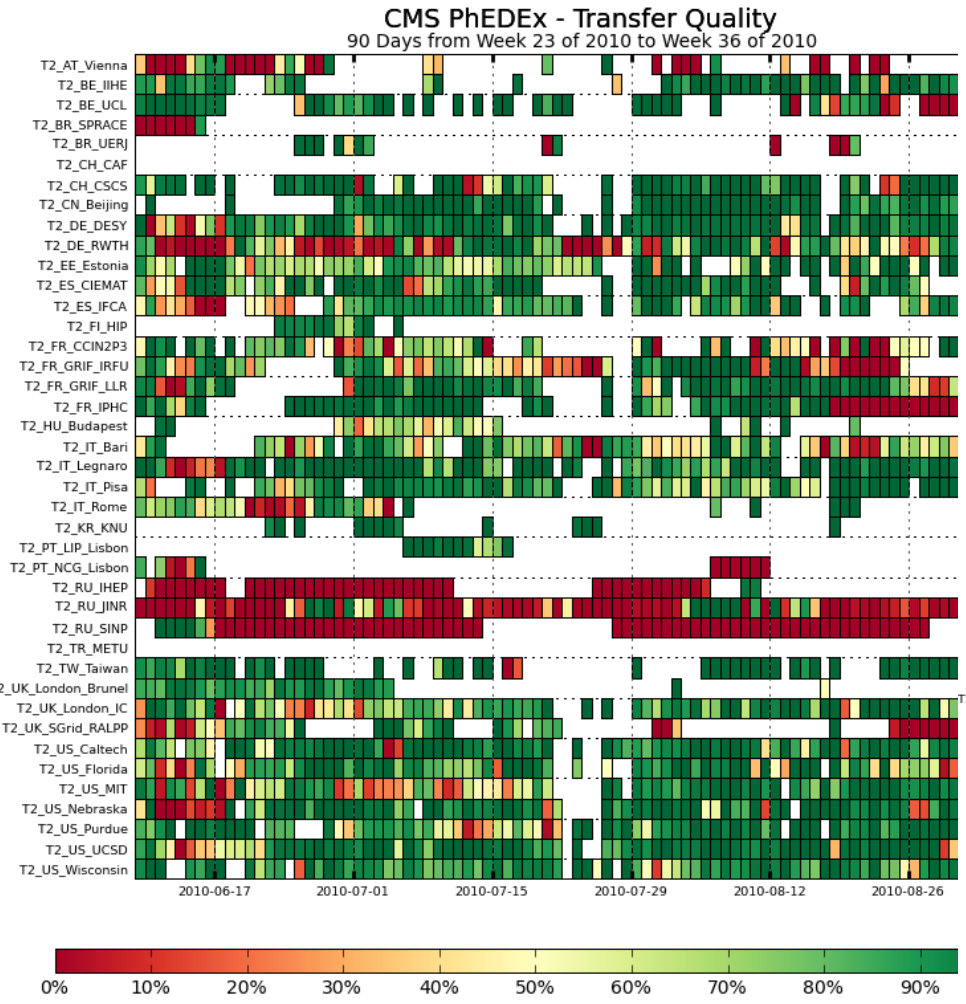
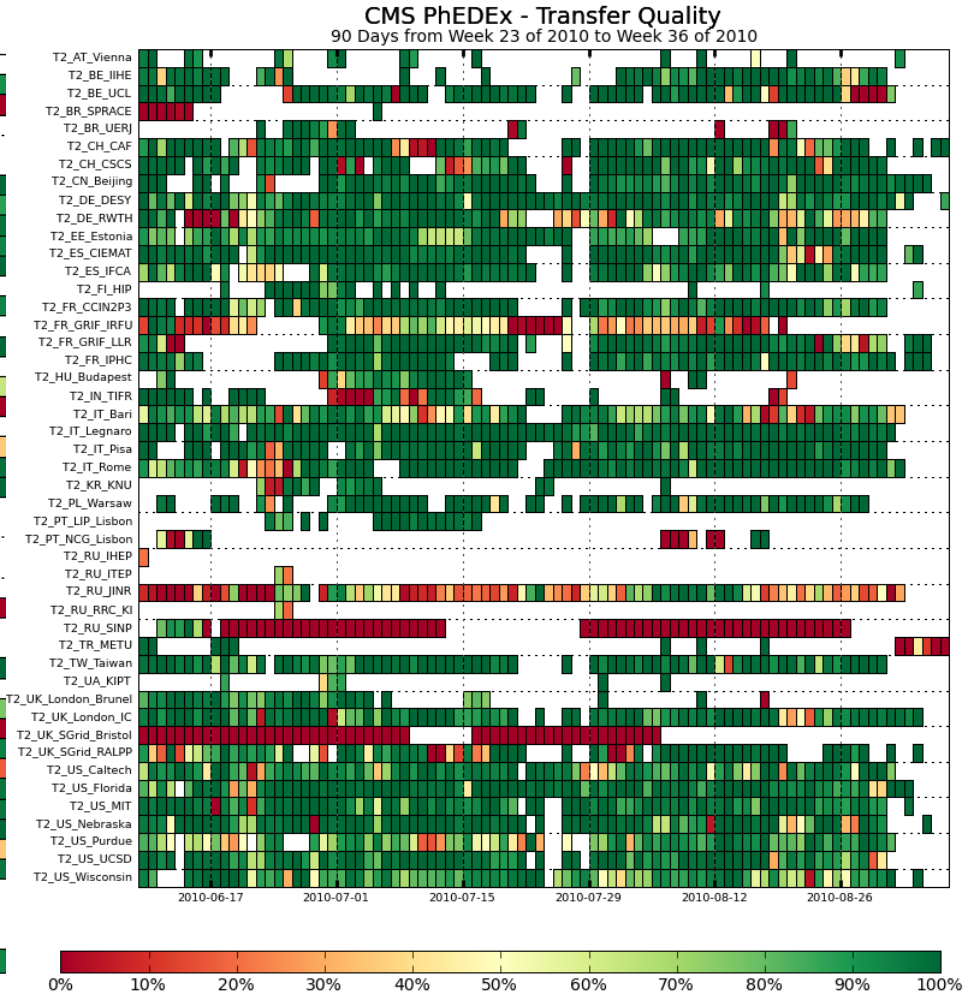
- Experiments move around Petabytes routinely
  - Exercised over many years in several data challenges
  - Experiment specific tools hide weaknesses of the middleware
  - Rates and volumes of the Computing Models are met
    - Real challenges (with real data) still to come
- Access to conditions data
  - So far no scaling issues seen
- Efficient local data access
  - Topic only in the very recent time
  - Difficult on complex infrastructures
  - Will be a big topic for the coming years
- Analysis of LHC data will be also a technical challenge

Tier-2 to Tier-1 Transfers (MC upload)



More Quality Plots

Tier-2 to Tier-2 (by Destination)



Tier-1 to Tier-2 (by Destination)