DATA MANAGEMENT

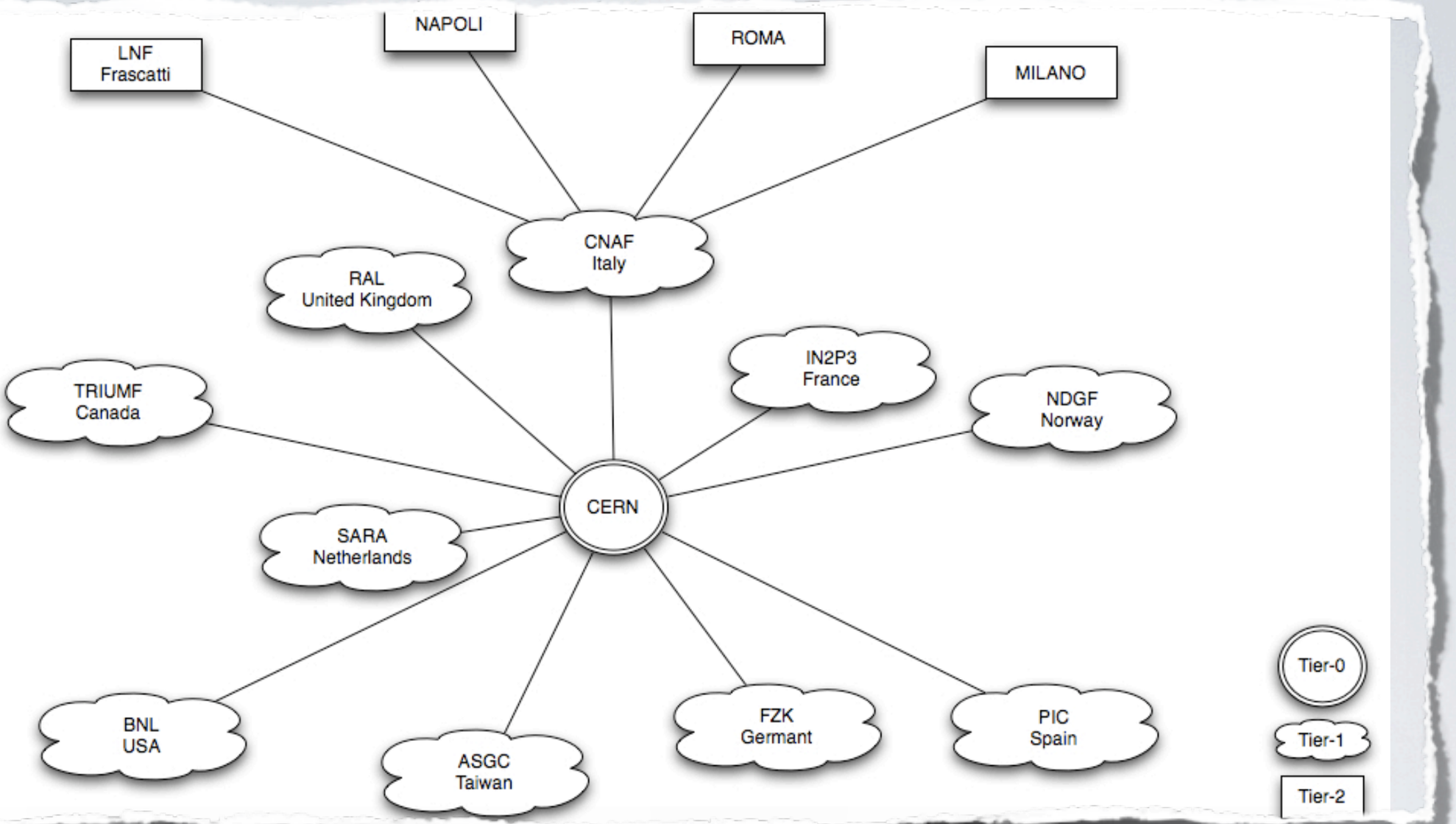
Pedro Salgado

BROOKHAVEN
NATIONAL LABORATORY

- what's at stake
- what's the problem
- how ATLAS is doing it
- numbers

WHAT

- LHC = 15PB/year
- primary backup - CERN (T0)
- 2 copies somewhere else (T1)



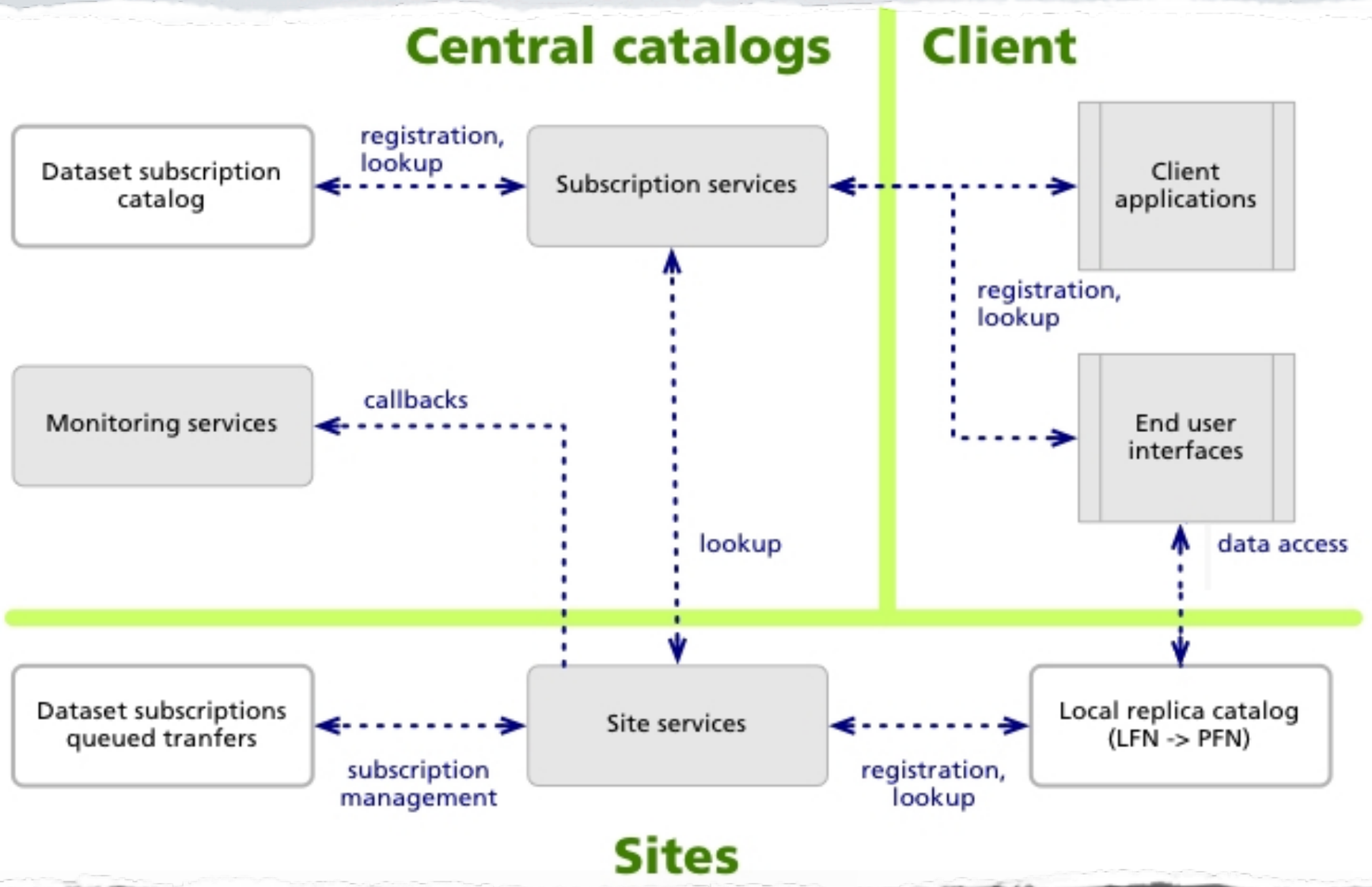
TIERS OF ATLAS

PROBLEM

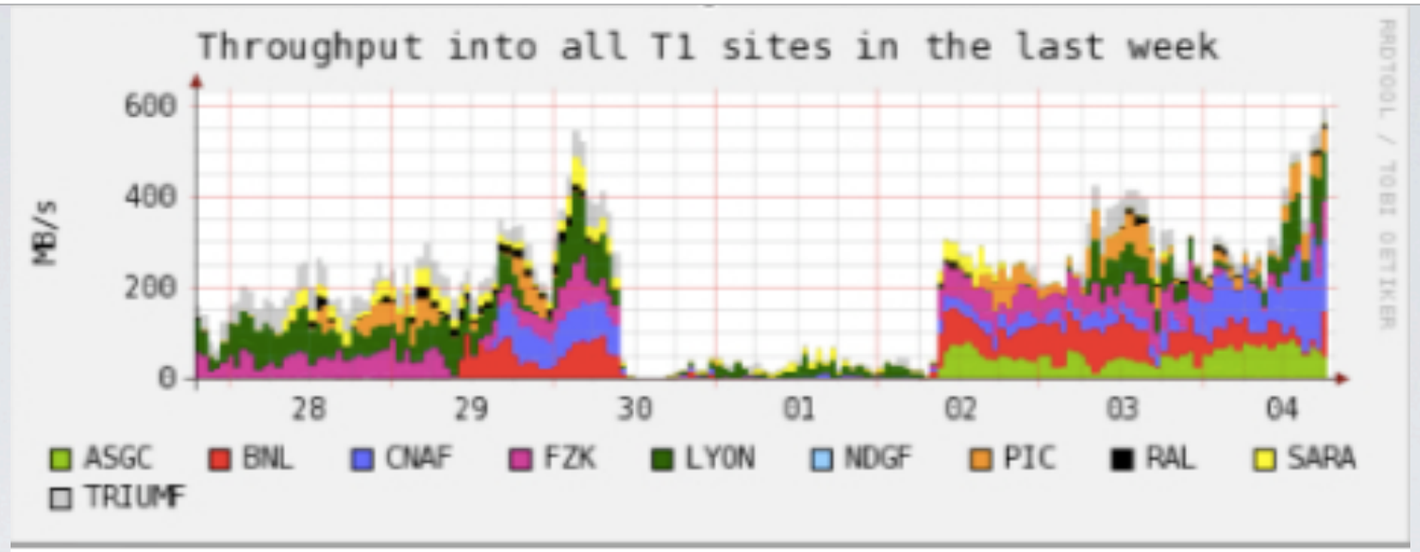
- petabyte-scale
- distributed nature
- scalability
- reliability

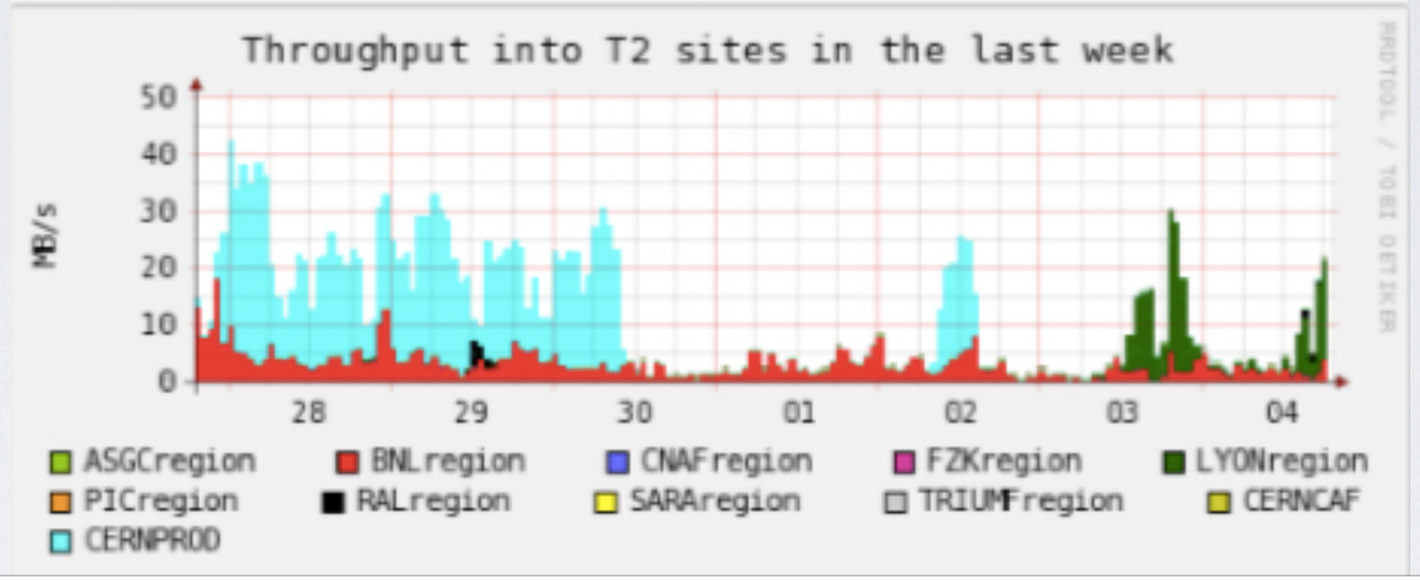
SOLUTION

- data set based bookkeeping [1]
  - data set 1..* files (2005)
  - container 1..* data sets (2008)
- software that implements this: DQ2 [2, 3, 4, 5]


Central catalogs
Client
Dataset subscription catalog
registration, lookup
Subscription services
Client applications
registration, lookup
Monitoring services
callbacks
End user interfaces
lookup
data access
Dataset subscriptions queued tranfers
subscription management
Site services
registration, lookup
Local replica catalog (LFN -> PFN)
Sites

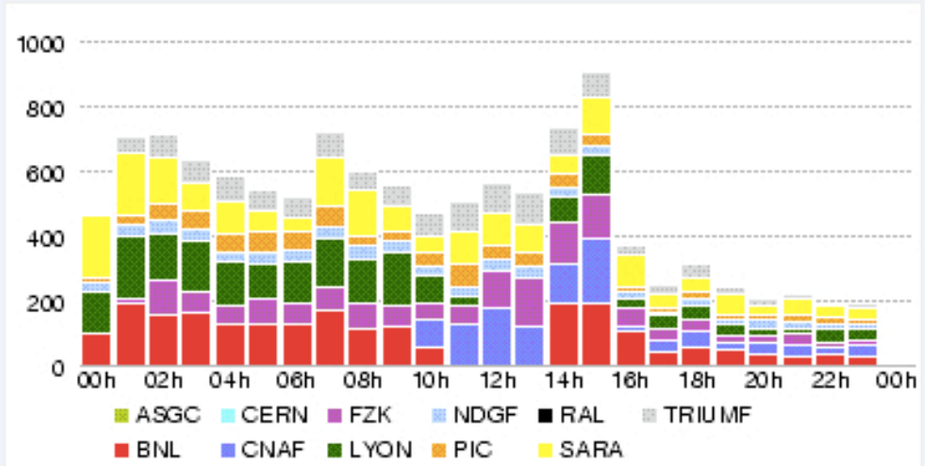
DQ2



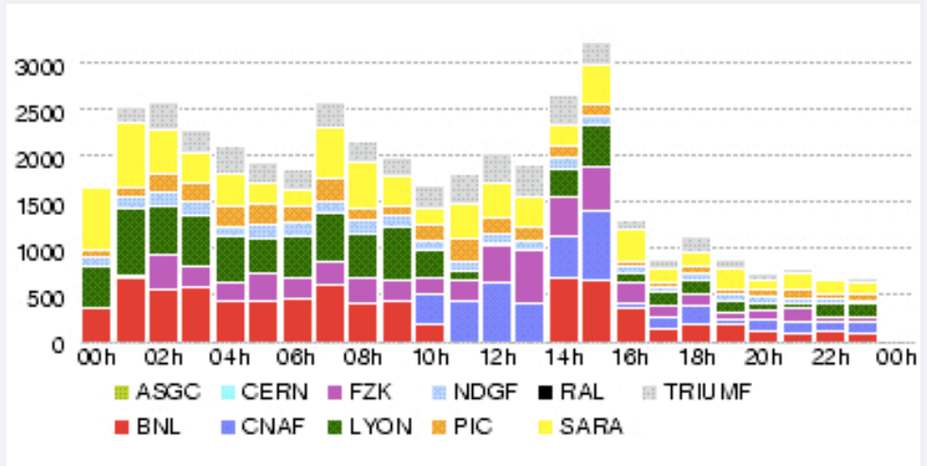


4 October 2006

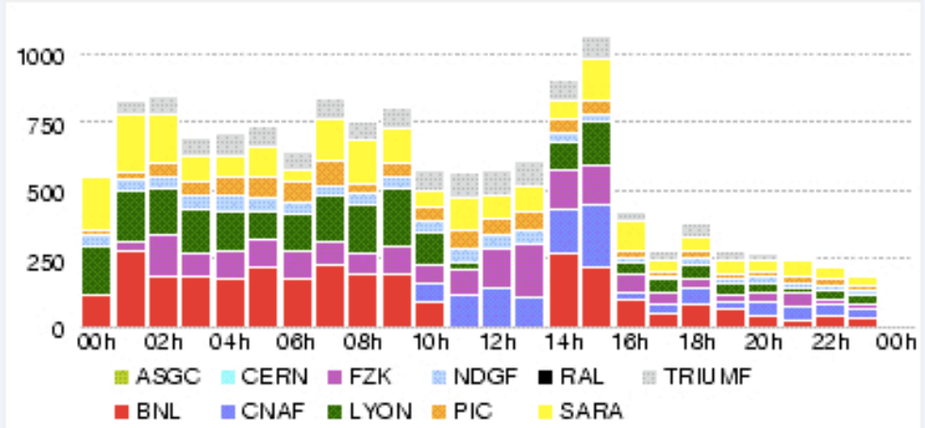
Throughput (MB/s)



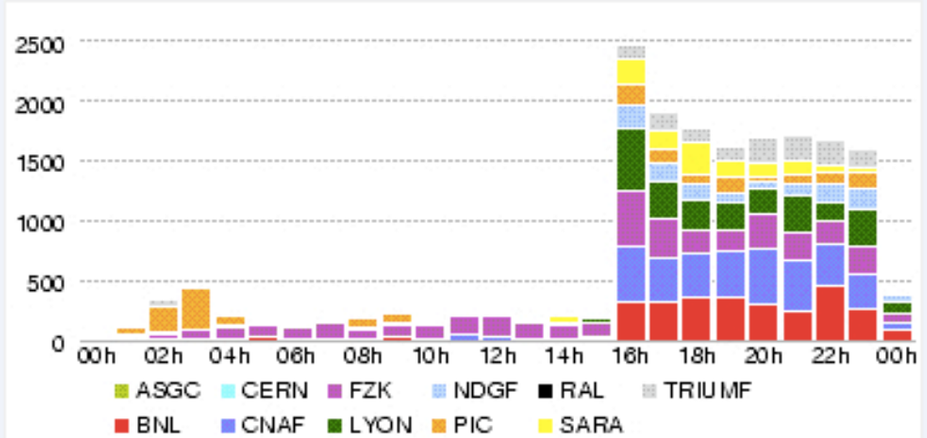
Data Transferred (GBytes)



Completed File Transfers



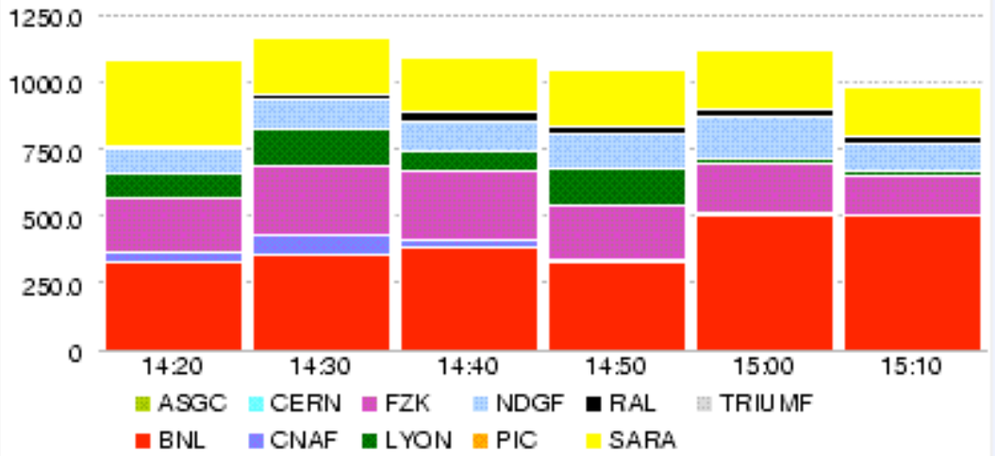
Total Number Transfer Errors



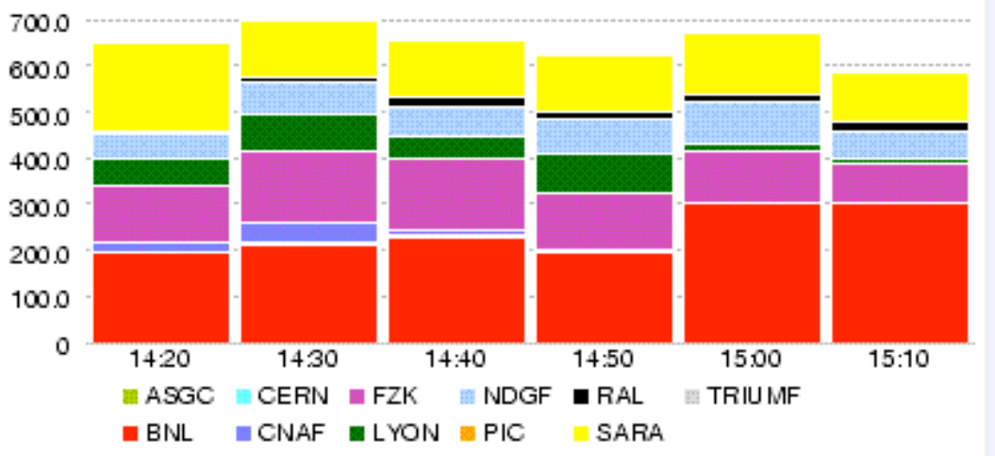
29 May 2007 (day)

29 May 2007 (hour)

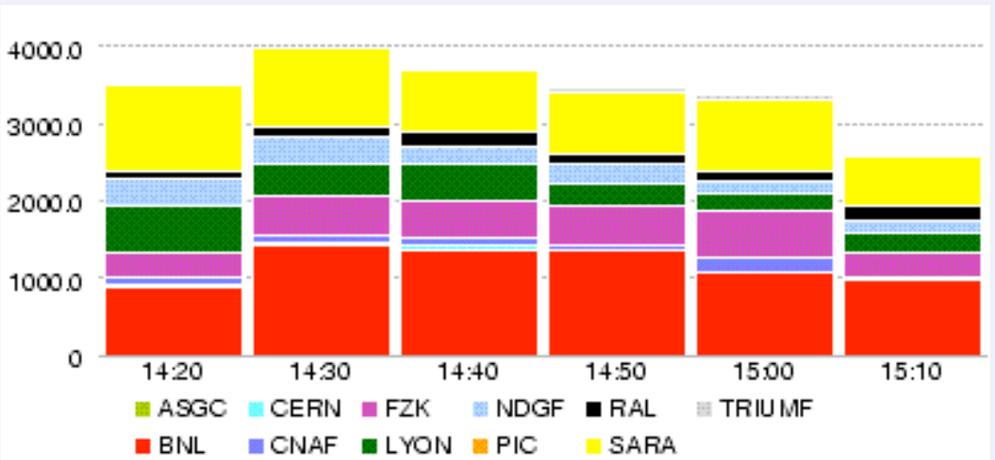
Throughput (MB/s)



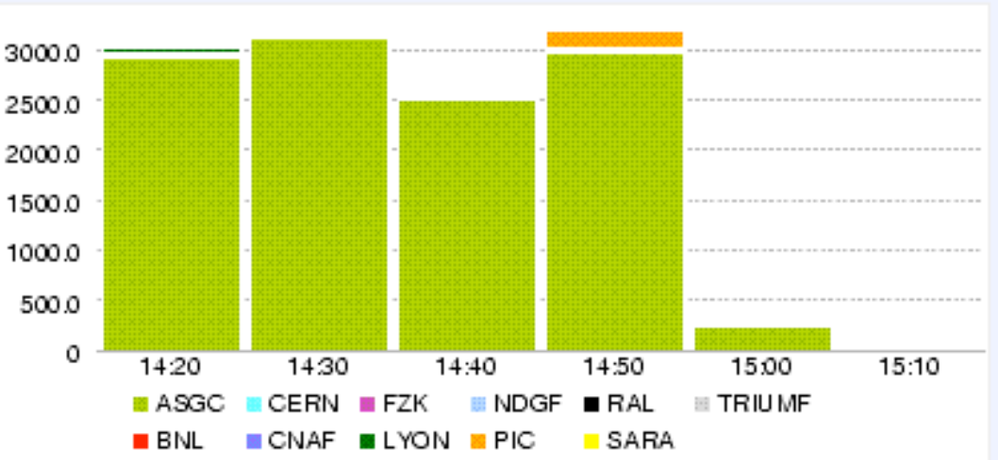
Data Transferred (GBytes)



Completed File Transfers

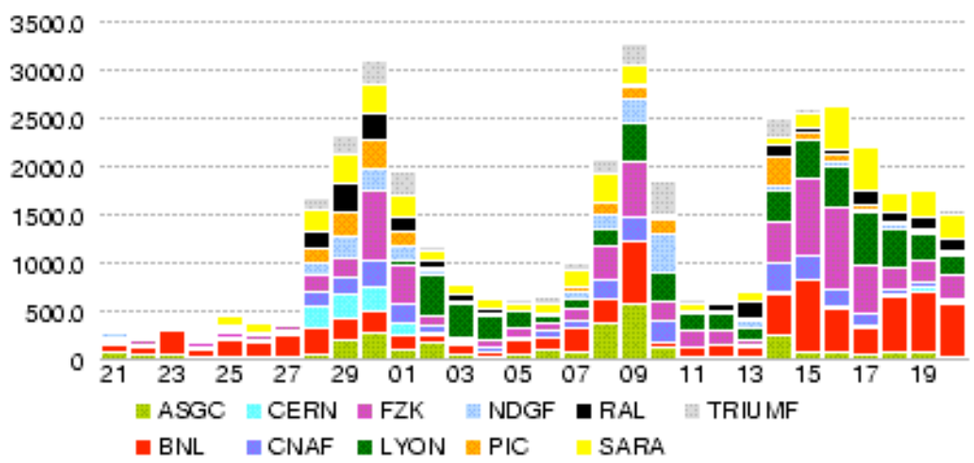


Total Number Transfer Errors

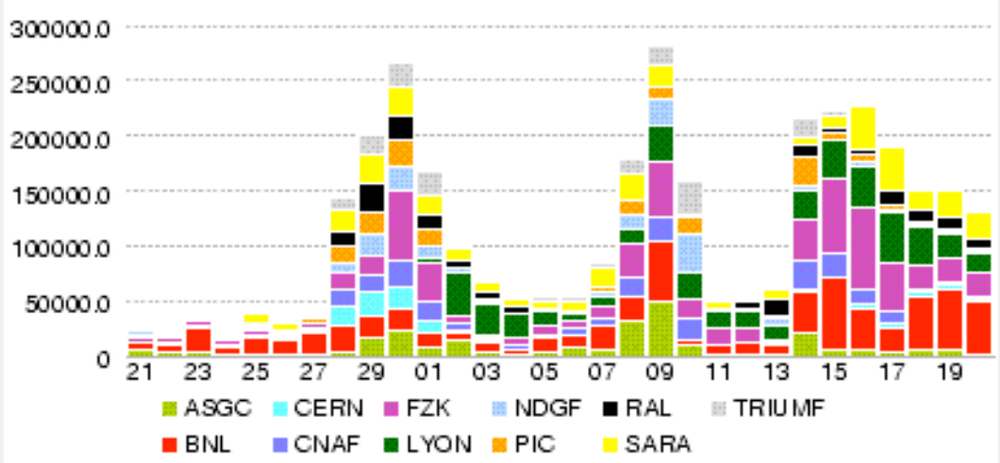


today

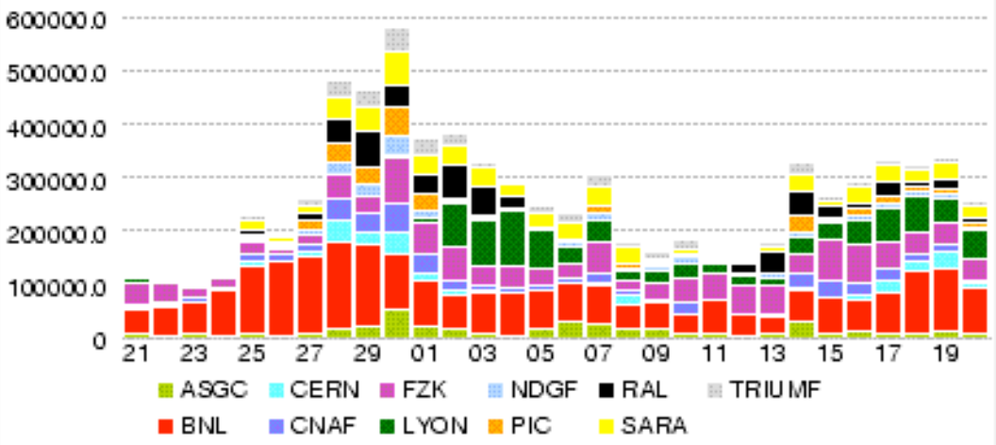
Throughput (MB/s)



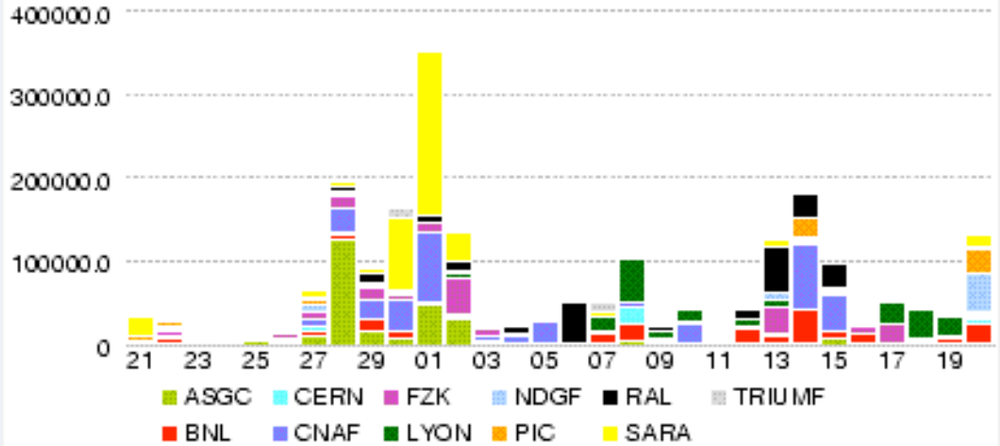
Data Transferred (GBytes)



Completed File Transfers

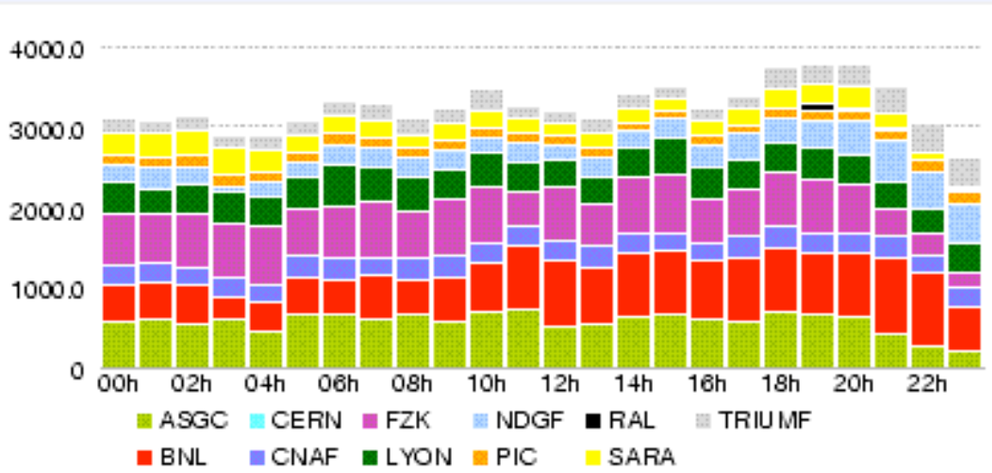


Total Number Transfer Errors

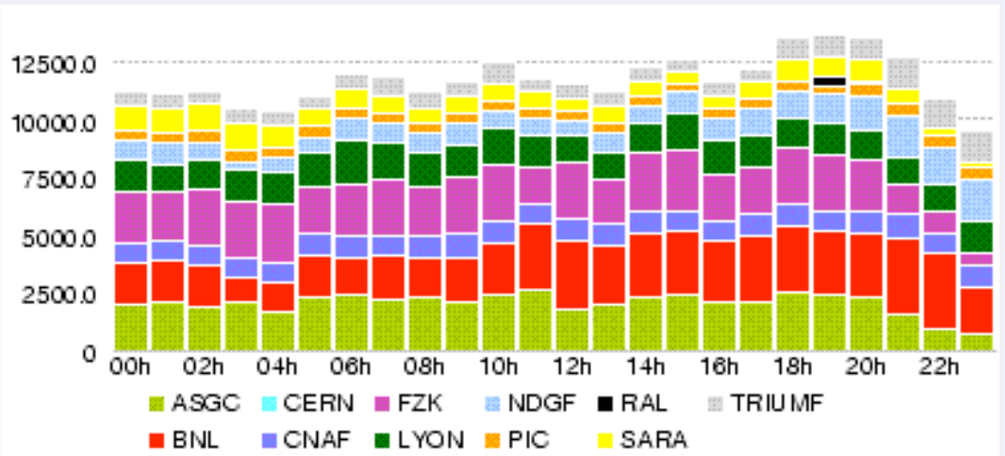


last month

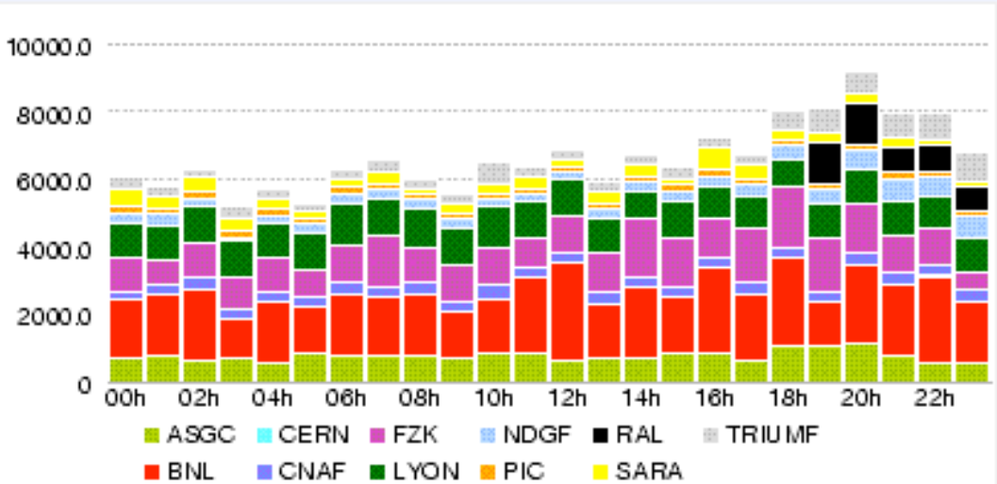
Throughput (MB/s)



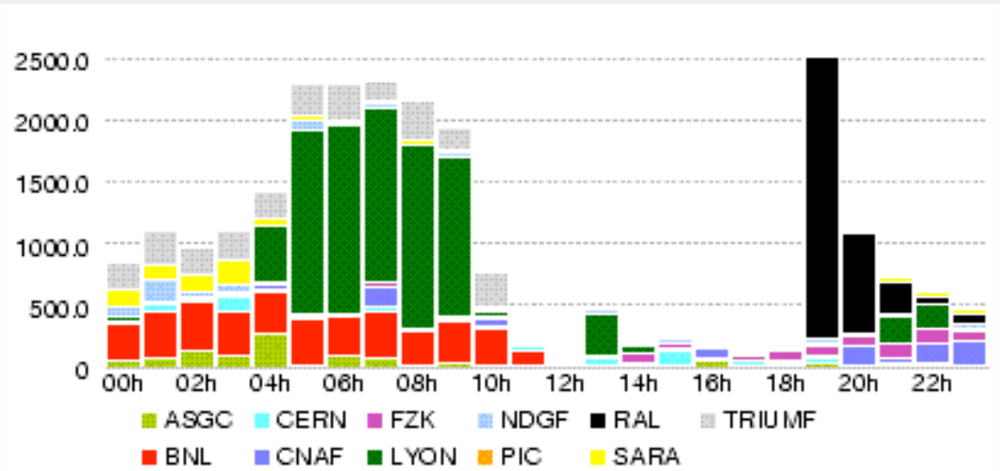
Data Transferred (GBytes)



Completed File Transfers



Total Number Transfer Errors



9 September 2009 "best day"

STORAGE @ BNL

- 4 PB (and counting)
- finely tuned storage
- redundancy
- high availability
- operations

BIBLIOGRAPHY

- [1] ATLAS Computing Group. ATLAS computing technical design report published on http://atlas-proj-computing-tdr.web.cern.ch/atlas-proj-computing-tdr/PDF/Computing-TDR-final-July04. pdf, July 2005.

- [2] Mario Lassnig, Miguel Branco, David Cameron, Benjamin Gaidioz, Vincent Garonne, Birger Koblitz, Massimo Lamanna, Ricardo Rocha, and Pedro Salgado. Managing ATLAS data on a petabyte-scale with DQ2. In Journal of Physics: Con- ference Series, Bristol, England. Institute of Physics Publishing.

- [3] DQ2 Central Catalogues published on https://twiki.cern.ch/twiki/bin/view/Atlas/DonQuijote2CentralCatalogs, January, 2009.
- [4] DQ2 Container Catalog use cases published on https://twiki.cern.ch/twiki/bin/view/Atlas/DonQuijote2ContainerCatalog, September, 2008.

- [5] Pedro Salgado. Specification and implementation of a data warehousing system for the ATLAS' distributed data management system. Faculdade de Engenharia da Universidade do Porto, July, 2008. http://pedrosalgado.beanstalkapp.com/mieic_report/browse/trunk/index.pdf