Experience operating multi-PB disk storage systems

CASTOR and EOS

Why & how
efficiency, inefficiency and costs

Luca Mascetti
CERN IT-DSS

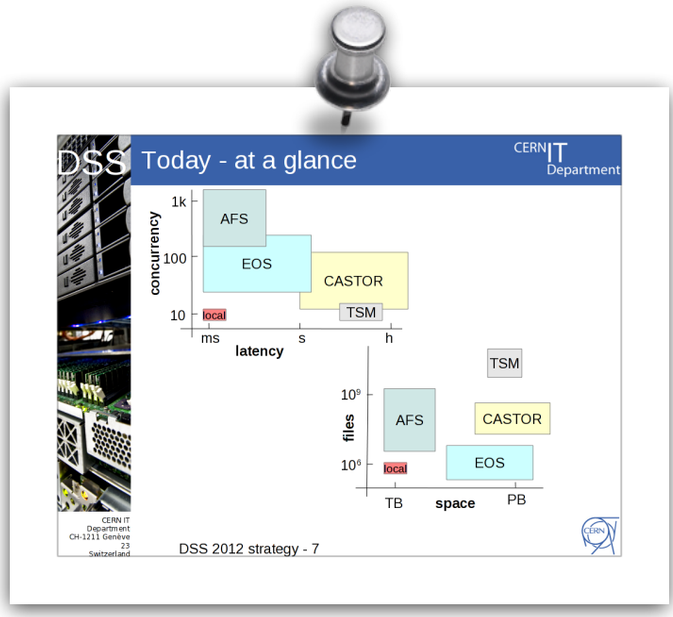
Outline

- **CASTOR and EOS**
  - **Why and how?**
- What works
  - performance, availability, reliability
  - support
  - shuffling hardware
- What does not really work
  - Draining
  - Balancing and FSCK
  - ...
- (In)efficiency and costs
- Summary

CASTOR and EOS - Why?

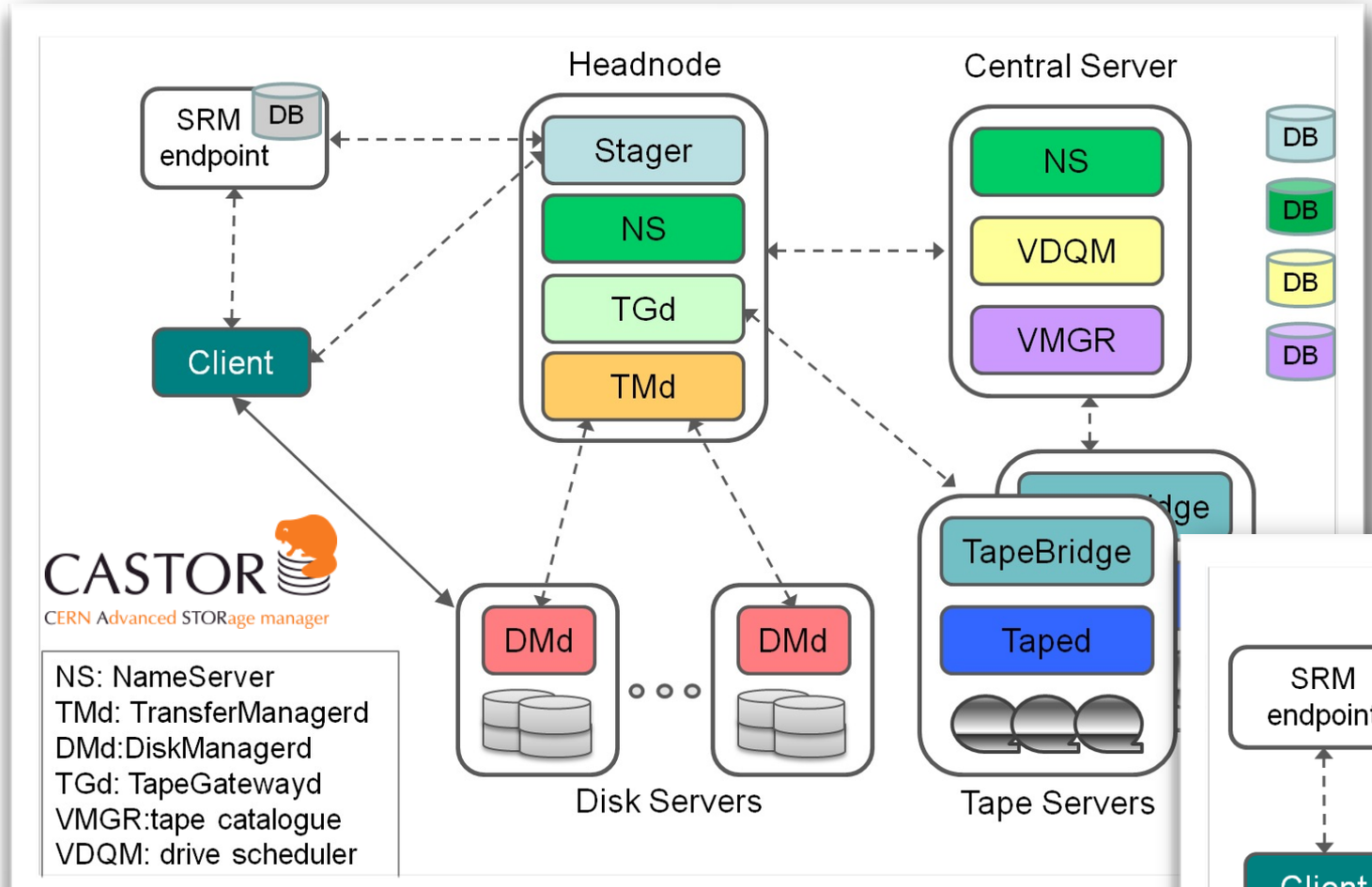
CERN IT Department

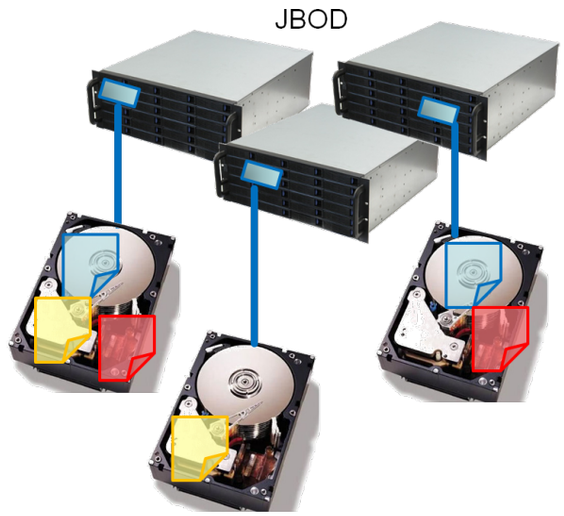
- CASTOR
  - "Physics data storage" requirements for WLCG-T0
  - .. and CDR (central data recording) for various non-LHC experiments
  - .. and local LHC and non-LHC analysis storage
- Desired properties
  - Theory: Big files (but not too big), long-lived, custodial, experiment-related, non-confidential, sequential access from few readers
  - Practice: 0-size/temp. data/user backups, random & parallel access from many readers
- 2011 Strategic decision
  - split T0 activity (CASTOR) from analysis (EOS)
  - slowly remove all diskonly pools from CASTOR
- EOS goals
  - low-latency and tunable reliability (multiple copies)
  - cheap (hw + ops)



Scope of the talk is CASTOR DISK and EOS

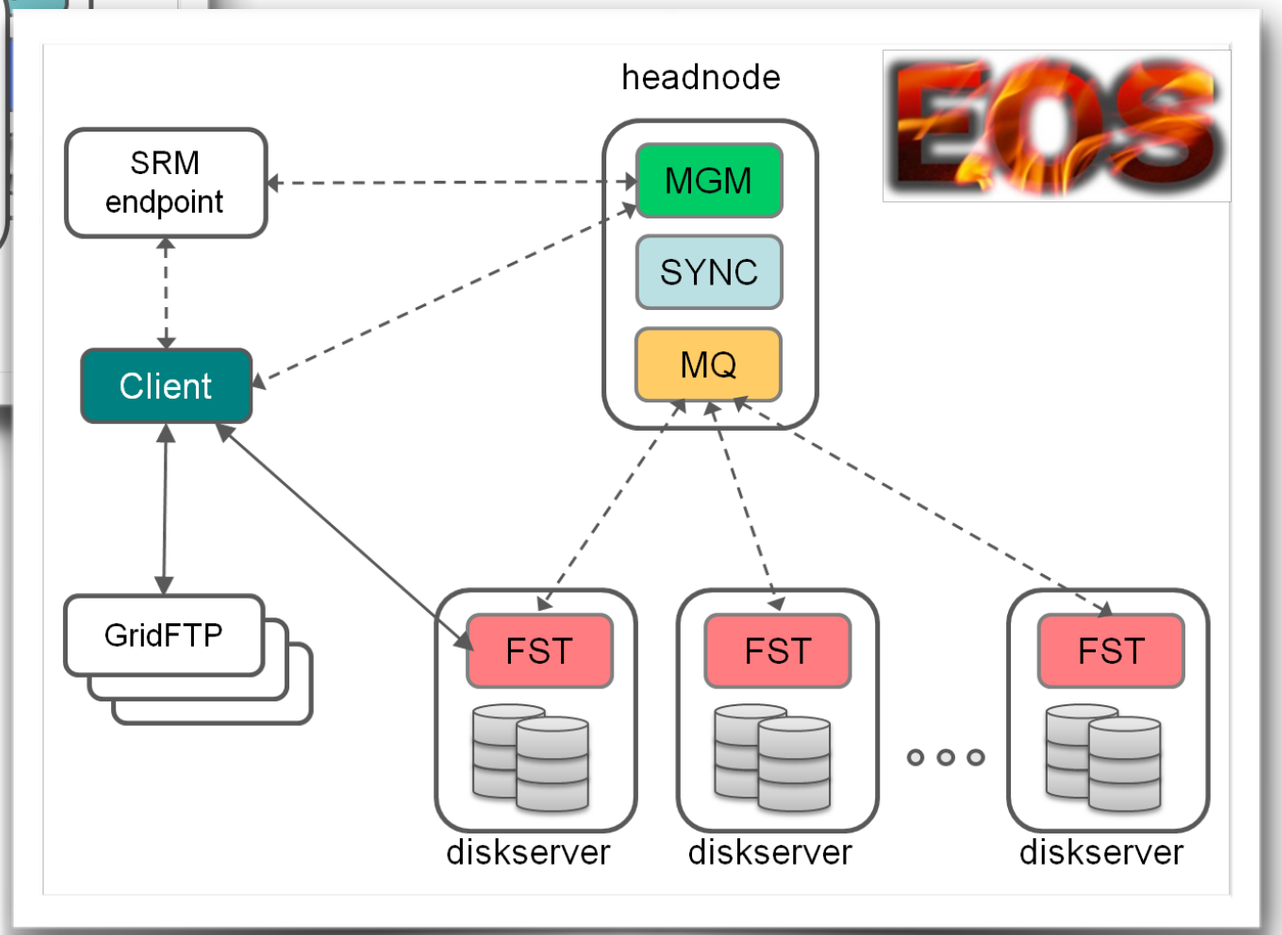
CASTOR and EOS architecture





CASTOR and EOS are using the same commodity hw what change is the layout of the disks

- RAID-1 for CASTOR
- JBOD with RAIN for EOS

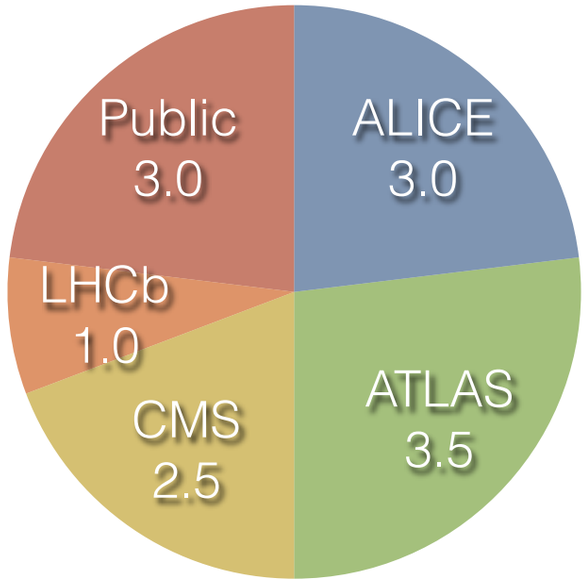


Few numbers

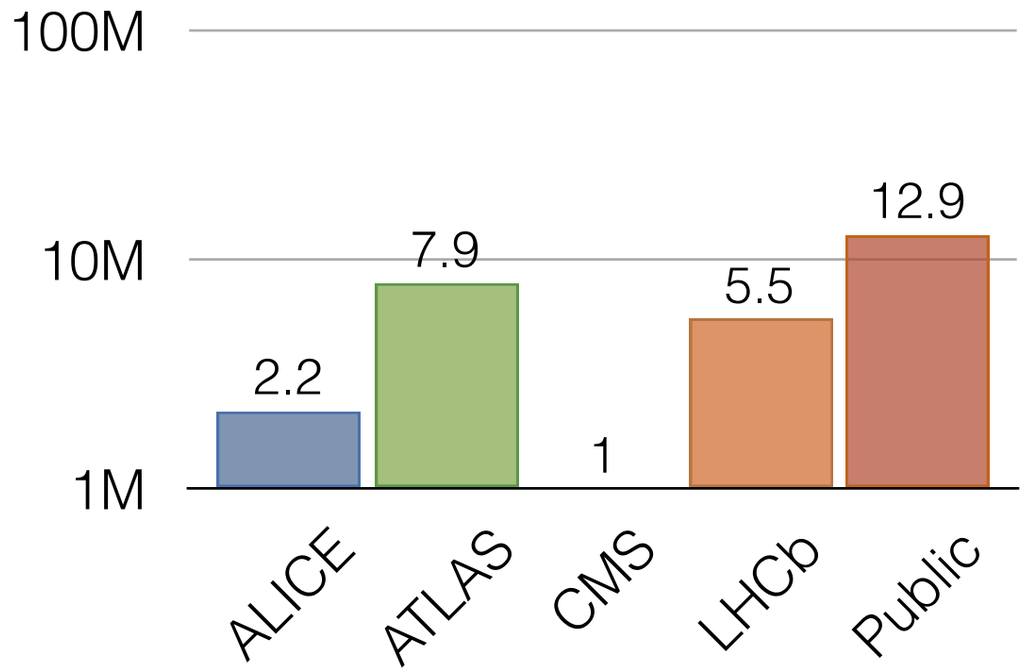
CASTOR
CERN Advanced STORage manager

Total installed disk capacity : 13.0PB

Installed (usable) disk capacity



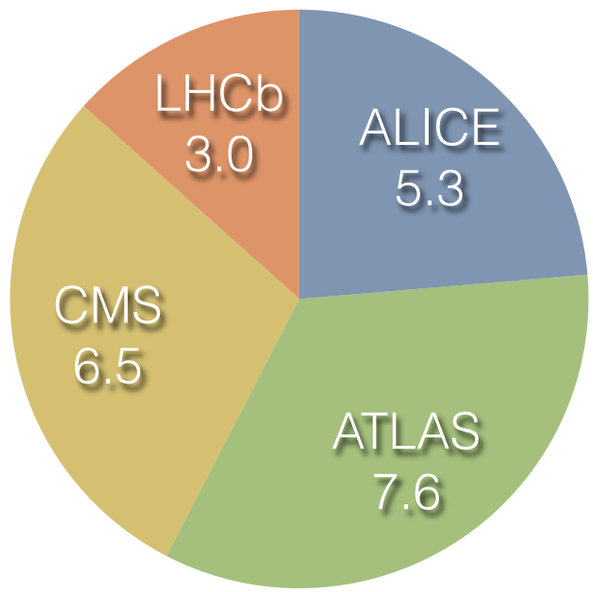
Number of Staged Files



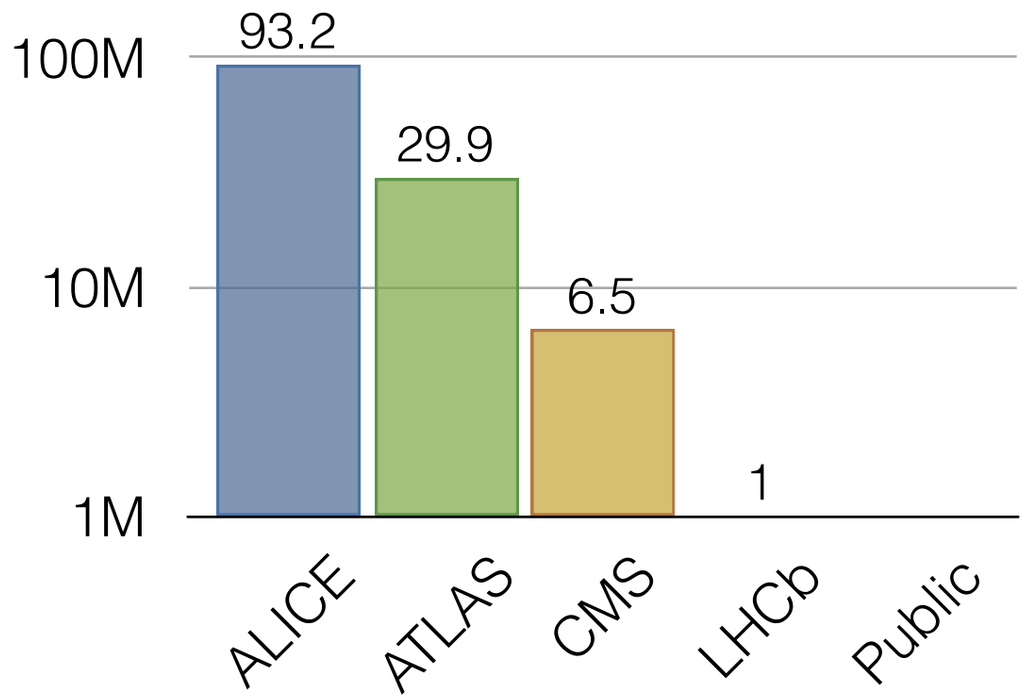
EOS

Total installed capacity : 22.4PB

Installed (usable) capacity



Number of Files

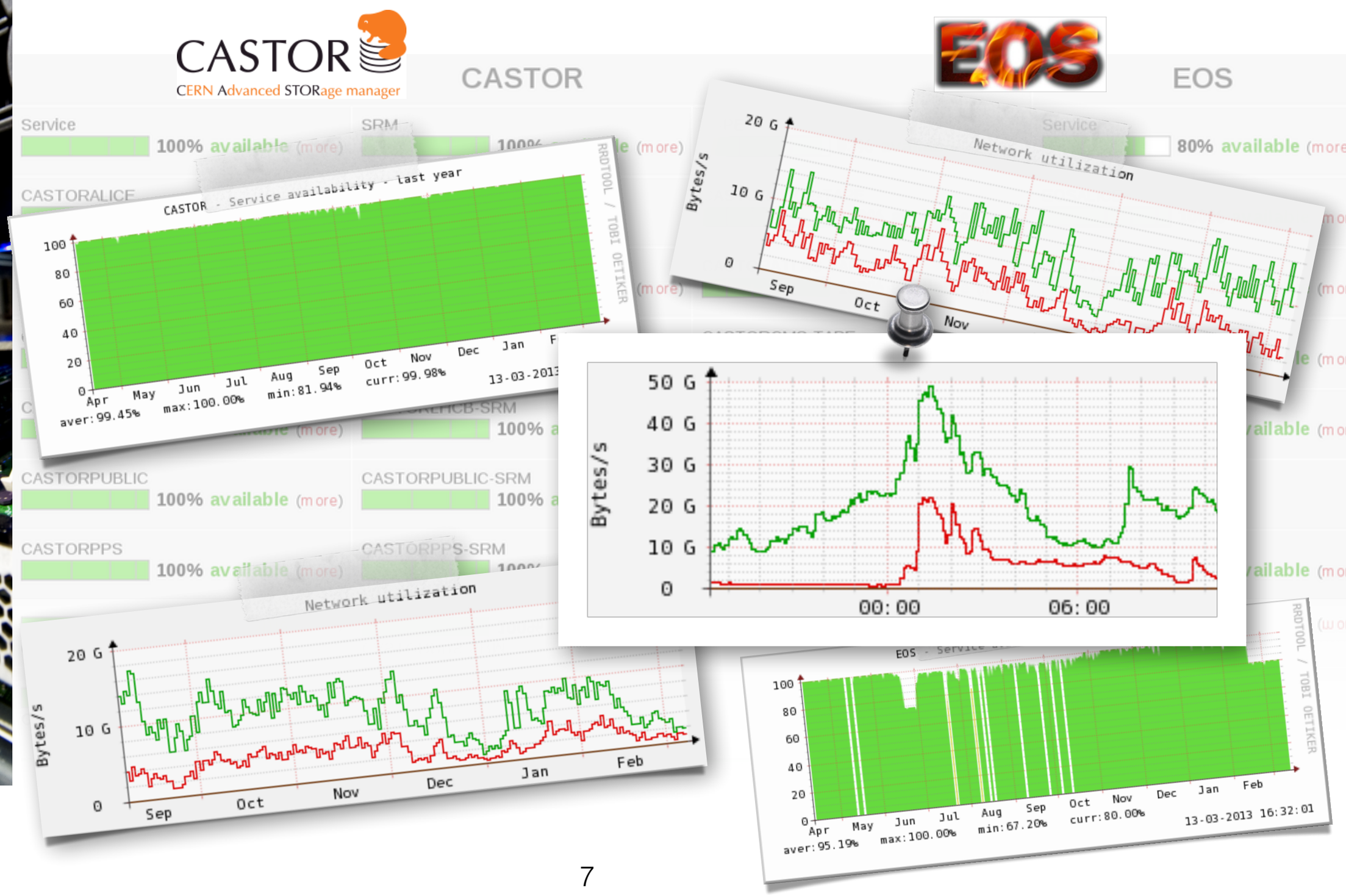


Outline

- CASTOR and EOS
  - Why and how?
- **What works**
  - **performance, availability, reliability**
  - **support**
  - **shuffling hardware**
- What does not really work
  - Draining
  - Balancing and FSCK
  - ...
- (In)efficiency and costs
- Summary

What works

- I/O performance and availability

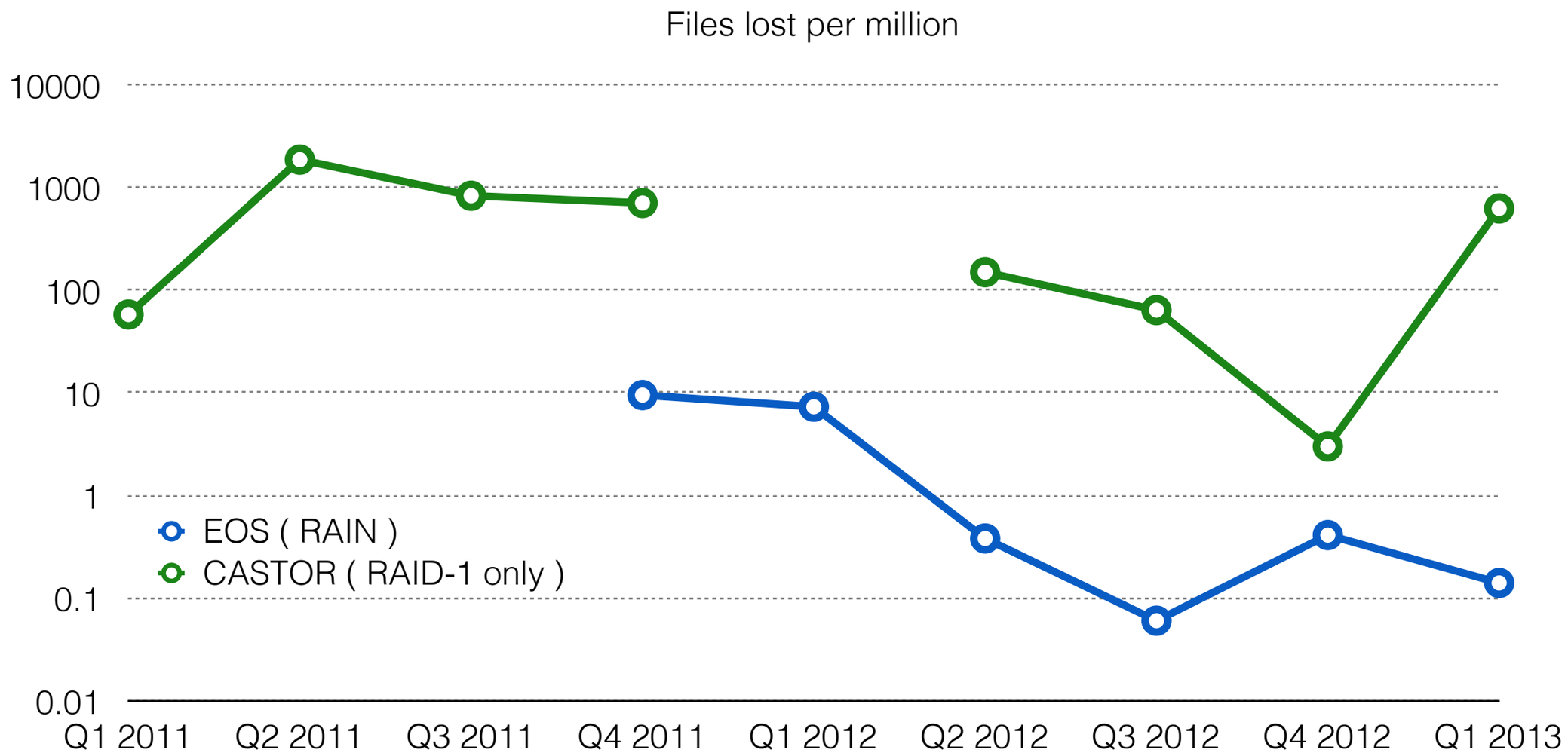


CERN
IT Department
L.Mascetti

DSS Reliability... Good Enough

CERN IT Department

- File loss is not nice but unavoidable with a certain probability
  - RAID-1 does not protect against controller or machine problem, filesystem corruption and finger trouble
  - typically important files can be recovered from offsite
- In case of backup (CASTOR) the tape reliability is helping the disk one

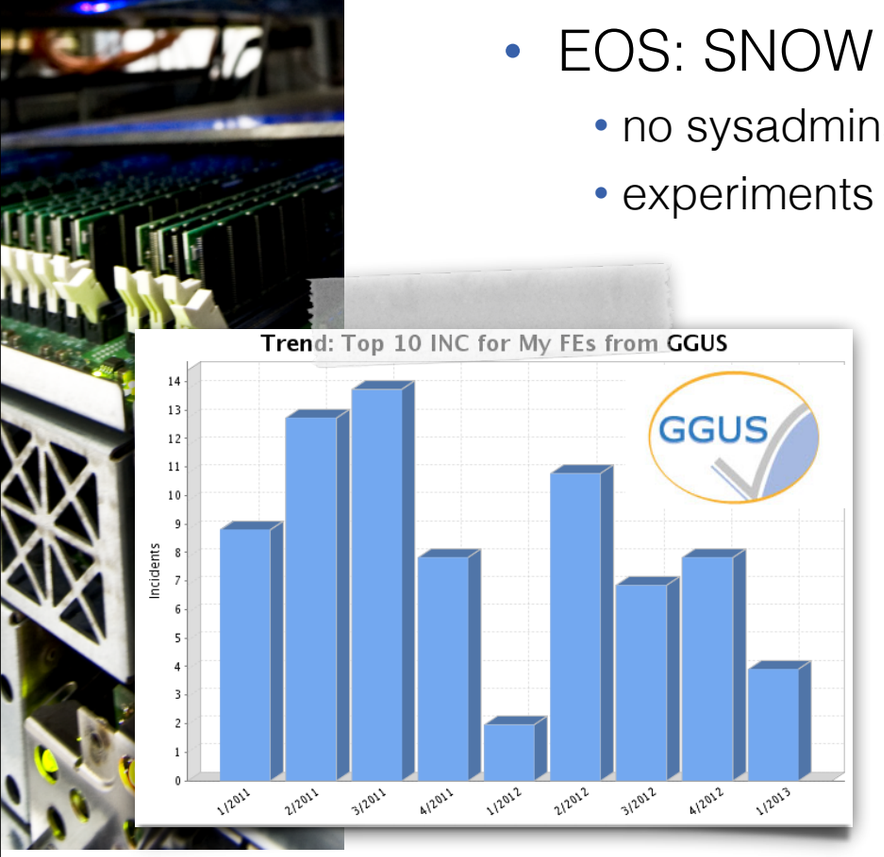


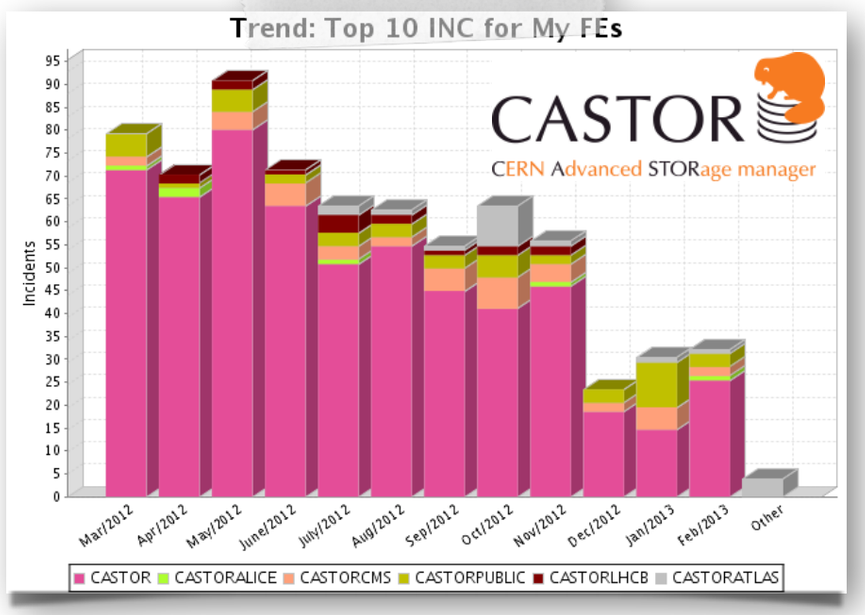
What works: support

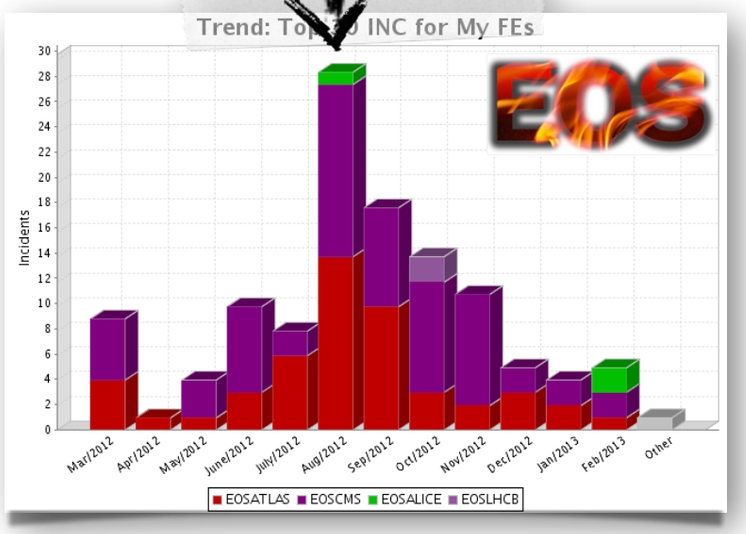
Experiments and Users (reasonably) happy.

- GGUS ticket rate ~3/month (alarms + some team ticket)
  - GGUS higher priority (T0 data involved)
- CASTOR: SNOW ticket rate ~60/month
  - of which a good number of machines and sysadmins
- EOS: SNOW ticket rate ~10/month
  - no sysadmins tickets
  - experiments handle directly majority users' issues

Wait... What was going on?

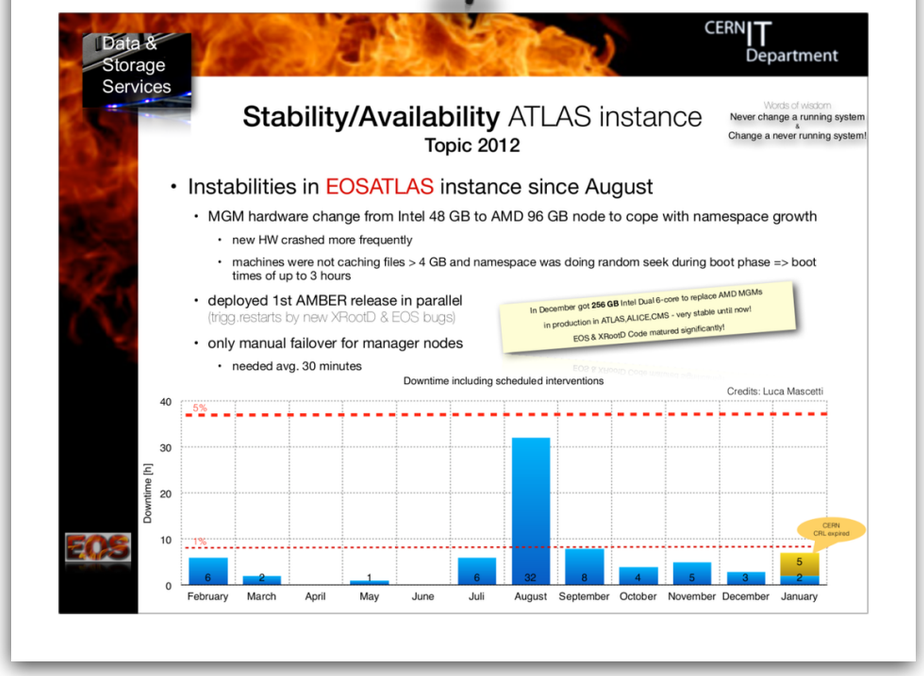






August Instabilities (causes)

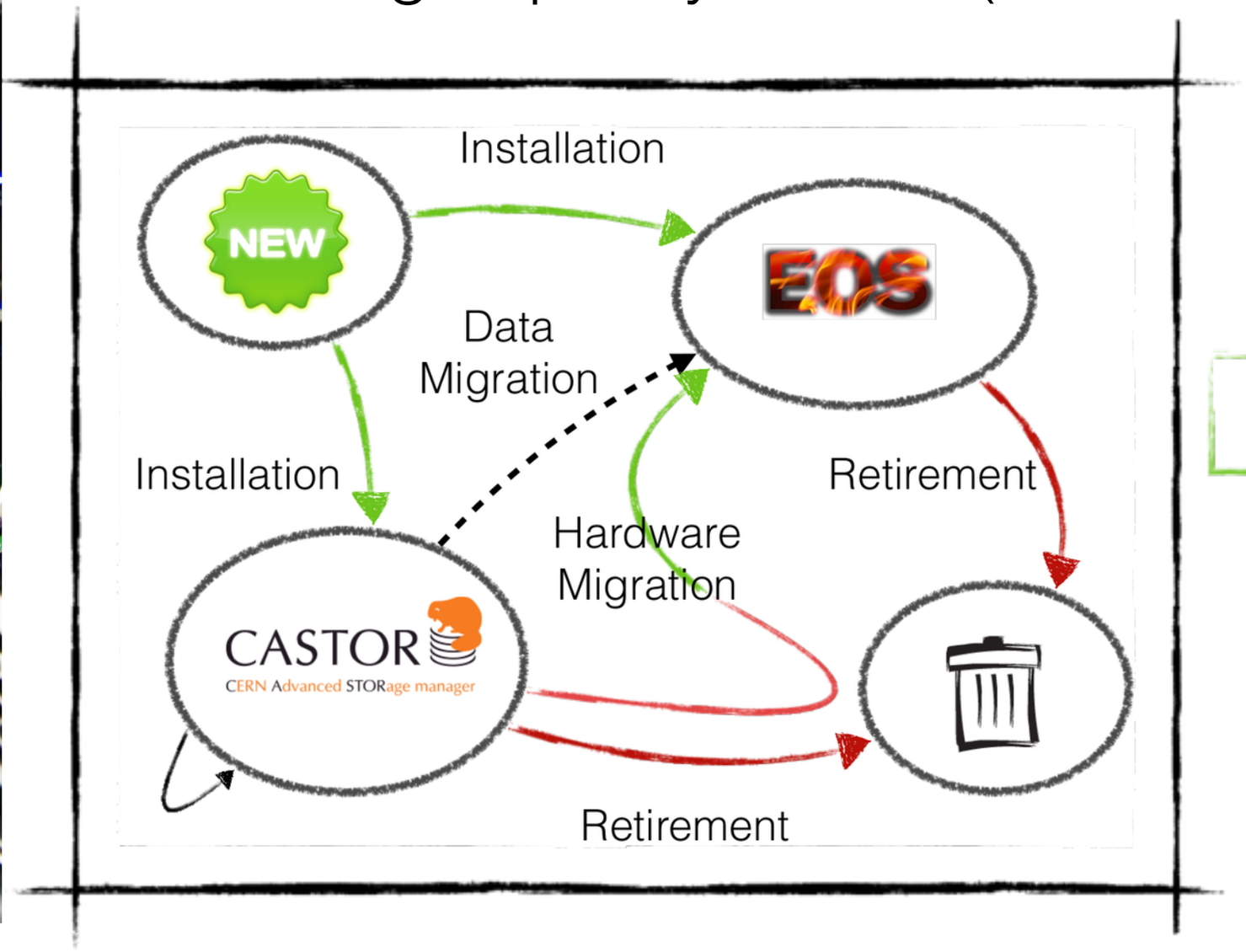
- Multiple crashes
  - GSI Auth bug (XrootD)
  - Retry bug (XrootD)
- Unable to compact namespace
  - Increase of restart time
- New + Unstable headnode hardware
  - several no_contact+reboots (7)
  - 96GB to check every time
  - AMD NUMA Layout = no disk cache
- Solution: software update + new headnode

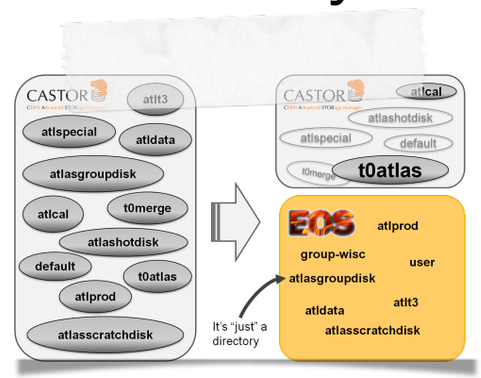


Understood and fixed

What works

- shuffling hardware.. why?
  - simplify CASTOR by reducing pool numbers
  - moving capacity to EOS (faster than hw lifecycle)





Installation

EOS: 1 step
- fully automated

CASTOR : 2 steps
- installation
- registration

Outline

- CASTOR and EOS
  - Why and how?
- What works
  - performance, availability, reliability
  - support
  - shuffling hardware
- **What does not really work**
  - **Draining**
  - **Balancing and FSCK**
  - **...**
- (In)efficiency and costs
- Summary

Drain is a pain...

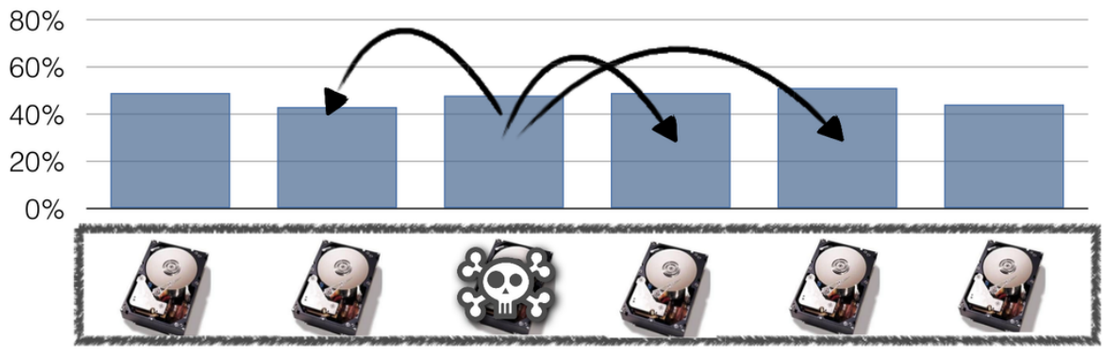
- delicate procedure
  - require to move all files present
  - moment of truth (things going wrong during time)

- "bottom of the barrel" problem
  - checksum/size discrepancy
  - dark data

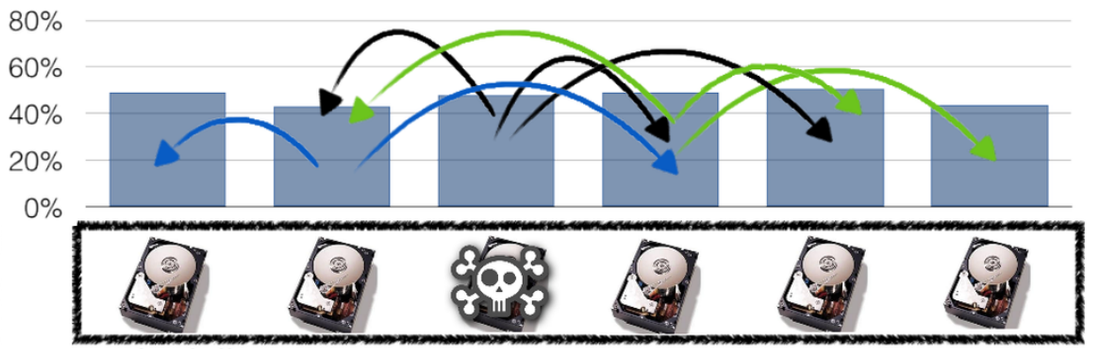
- require manual effort
  - recover data
  - clean up metadata
  - declare data loss

Drain is a pain...

CASTOR
CERN Advanced STORage manager

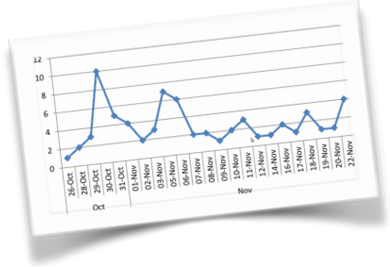


EOS



- Draining is a manual decision: CANBEMIGR vs diskonly vs recall
- drain machinery not very robust
  - stuck/interrupted draining jobs
  - generation of FAILED copies
  - (better in v2.1.14)
- limited in bandwidth by a single box

- Drainig is part of standard automatic operation
  - more robust and faster
- expired draining - tool not perfect
  - e.g. bug cannot drain 0-size files (fixed)



- but.. for both same problem with metadata operations and data recovery

FSCK and Balancing

- FSCK
  - CASTOR: decentralized checksum verification
  - EOS: much better but no autorepair
- Balancing
  - CASTOR
    - not present, box are unevenly filled
    - manual procedure for disks 100% full
  - EOS:
    - tunable balancing inside groups
    - missing balacing between groups
      - useful when instance grows very fast

Outline

- CASTOR and EOS
  - Why and how?
- What works
  - performance, availability, reliability
  - support
  - shuffling hardware
- What does not really work
  - Draining
  - Balancing and FSCK
  - ...
- **(In)efficiency and costs**
- Summary

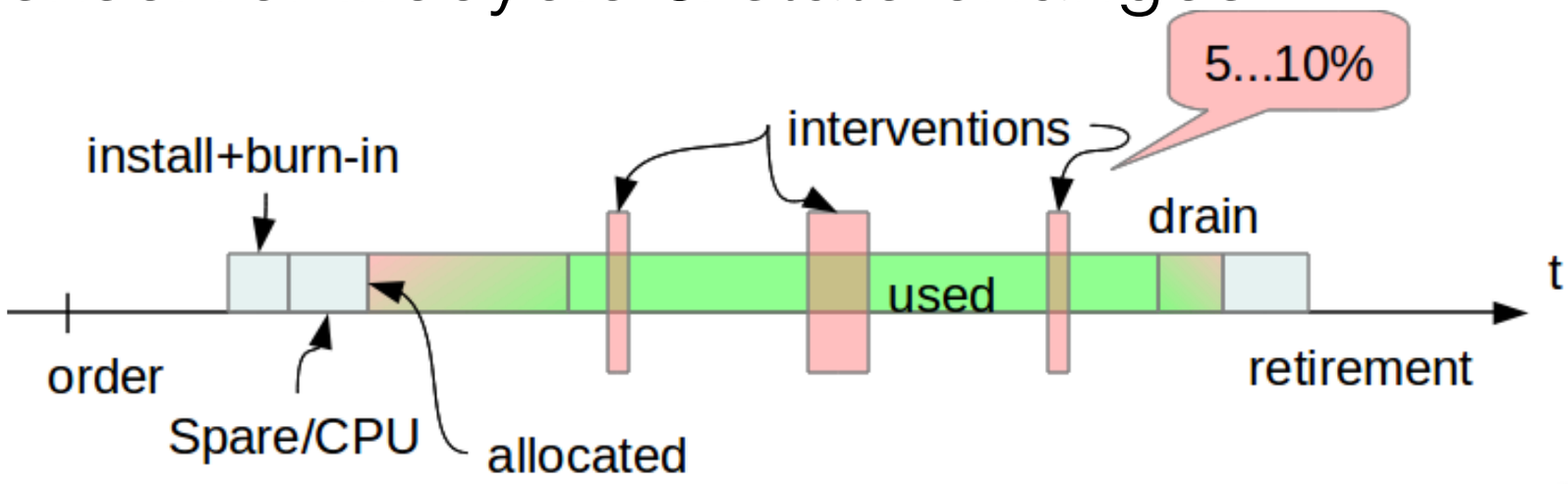
(in)efficiency

- Several “small” inefficiencies will accumulate
  - Machine & human-level
- External usage - guideline: “user decides”.
  - Read same file over & over again
  - Write-only files (“efficient optimization” possible..)
  - 0-size files
  - ad-nauseam replication

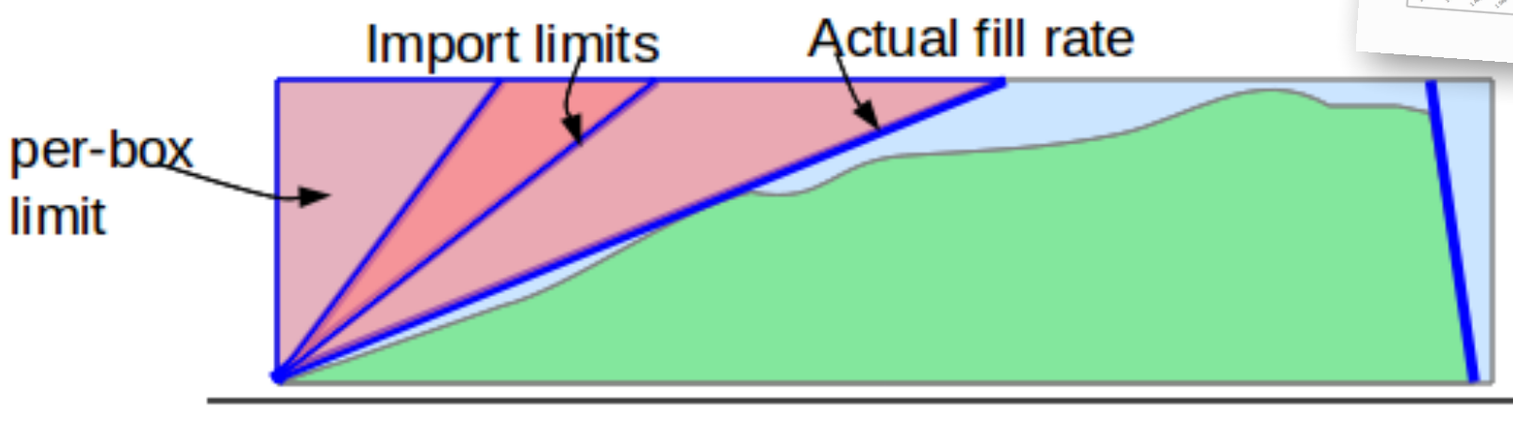
  Note: storage vendors get all excited about “deduplication”. We don't.
- Inefficiency might be OK
  - Conscious trade-offs

State/lifecycle inefficiency

- Diskserver lifecycle & state changes



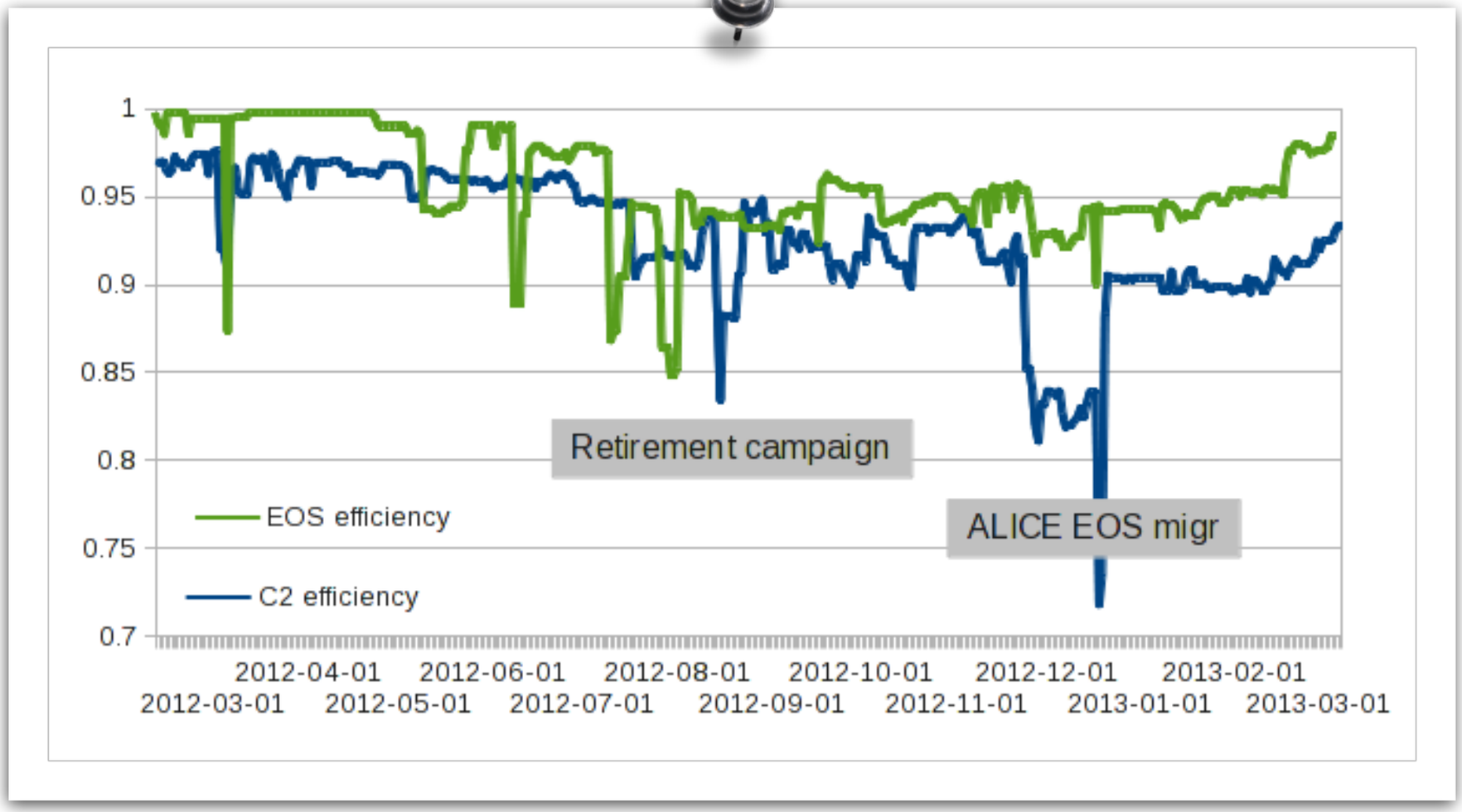
- Diskonly Pool lifecycle and usage



- Better usage of free space?
  - transient replicas? (increase on reliability)

Inefficiency: forgotten boxes...

"no box left behind"?



- CASTOR drains "needs" machines to be non-production
  - trying to mitigate (timeouts)
- 5%-effect..

DSS Current Cost

CERN IT Department

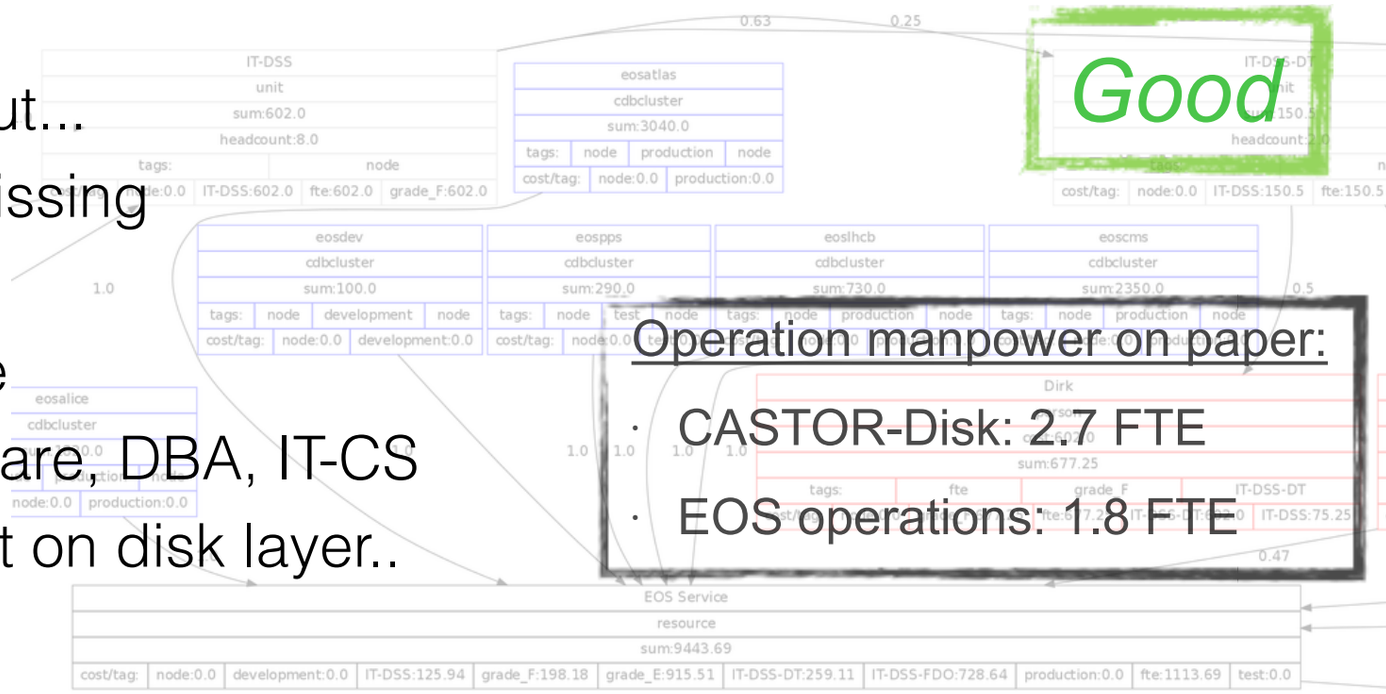
assumptions on:

- prices for HW @ 3years
- electricity cost
- disk operation manpower

	CASTOR (CERN Advanced STORage manager)	EOS
HW+ electricity cost	16.5 CHF/1TBMonth	13.0 CHF/1TBMonth
operation manpower cost	2.7 CHF/1TBMonth	1.3 CHF/1TBMonth
partial "running" cost	**19.2** CHF/1TBMonth	**14.3** CHF/1TBMonth

Amazon S3: "reduced redundancy", Europe, 30PB: **42**US$ / 1TBmonth (no Network, I/O ops)

- Doing OK cost-wise, but...
  - Some manpower missing
    - development
    - sysadmins share
  - ORACLE license share, DBA, IT-CS
  - CASTOR tape effect on disk layer..



- Would like to compare with other HEP labs

ROCK BOTTOM: Raw disk price: **1.9**CHF/1TBmonth..

Outline

- CASTOR and EOS
  - Why and how?
- What works
  - performance, availability, reliability
  - support
  - shuffling hardware
- What does not really work
  - Draining
  - Balancing and FSCK
  - ...
- (In)efficiency and costs
- **Summary**

Summary

- Overall storage “just works”
  - No major issues during last LHC run
  - Lots of “10%s” to be improved on!
- LS1 will be busy
  - (less CDR, more analysis)
  - Federations (xroot/http) - ongoing
  - EOSPUBLIC (AMS, ILC, NAx ..) - now
  - EOS@Wigner / agile puppets - realsooon
- No more CASTOR diskonly pools - LS1
  - Towards 1 tape-backed pool/experiment
  - Later: fewer instances?

Questions ?

Luca Mascetti
CERN IT-DSS