CERN Data Archive

Bit-level integrity and Data Archiving, Sharing and Exchange

Bit Rot

Visit of German Science Foundation 15/10/2013

Germán Cancio
CERN IT Department

DPHEP

WLCG

RESEARCH DATA ALLIANCE

Outline

- Data integrity and bit-level preservation
- Our Use Case: CERN Data Archive & Distributed Storage
- Bit-level preservation and reliability activities
  - What overall reliability to expect?
  - Media Verification, Repacking, Replication
  - What to do when we lost the data?
  - Issues with inter-archive sharing/exchange
- Conclusion

Data Integrity and Bit-level preservation

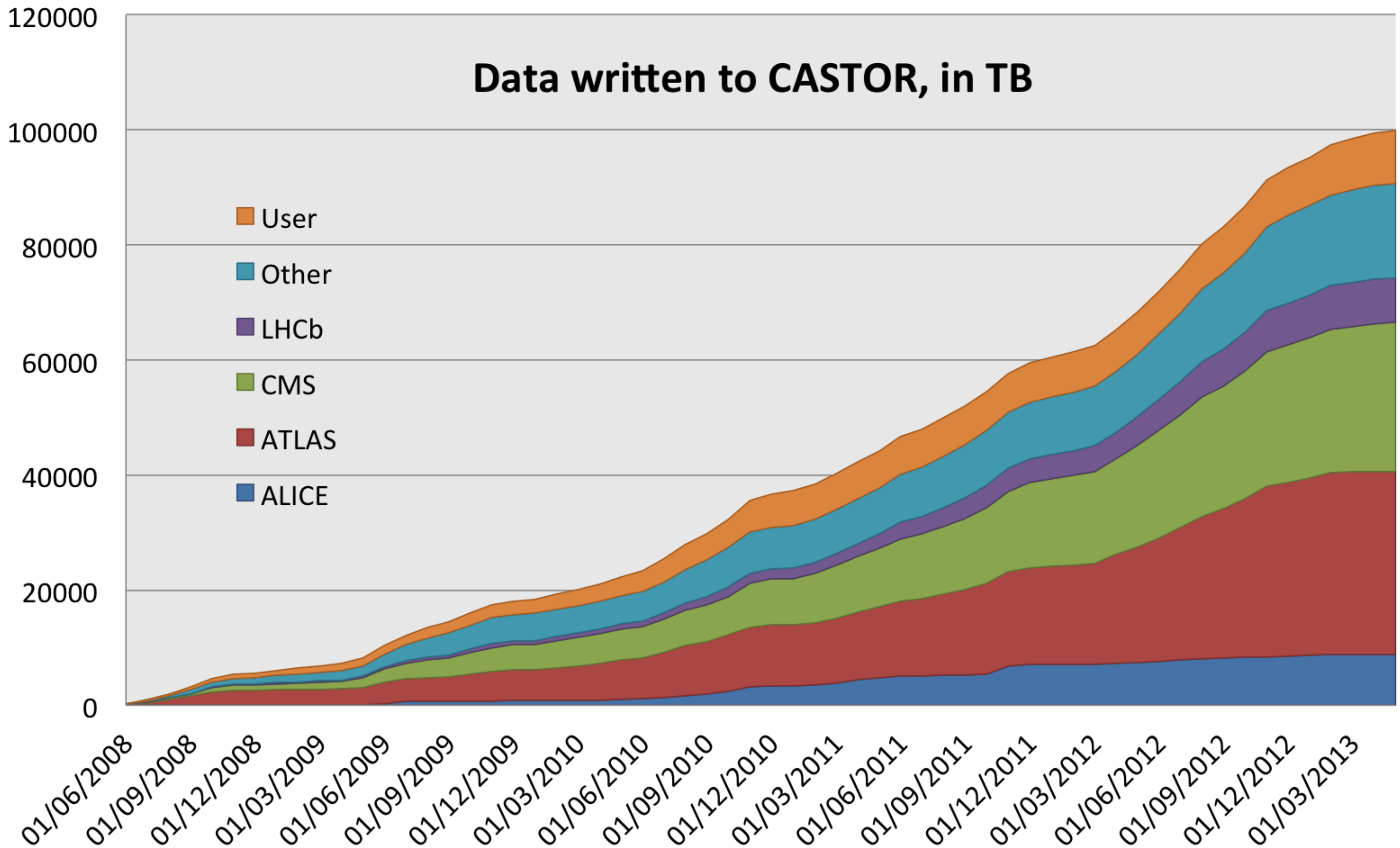
- Data Integrity is an essential component of a Collaborative Data Infrastructure
- Bit-level storage preservation mechanisms required for ensuring Data Integrity of each participating instance / site / VO

- **bit rot** (*uncountable*) is unavoidable!
  - ***(computing) The putative tendency of content in storage to become corrupt over time.***
- Reasons for bit rot:
  - Corruptions due to OS / firmware problems / bugs
  - Wear out and breakage (tape snap, disk drive fail, RAID controller)
  - Undetected bit flip during data transfers / storage
  - Accidents/disaster (water leak, fire,…)
  - Media or (more often): hardware obsolescence
  - Security break-ins
  - ….

Our Use Case (1): CERN Data Archive

Data at CERN:
- ~100PB of data on tape; 250M files
- Up to 5 PB new data per month
- Multiple User Communities (WLCG/HEP, EGI, ..)
- **Aged up to 30y; stored ad eternum**

Infrastructure at CERN:
- 52’000 tapes
- 9 Robotic libraries
- 110 tape drives

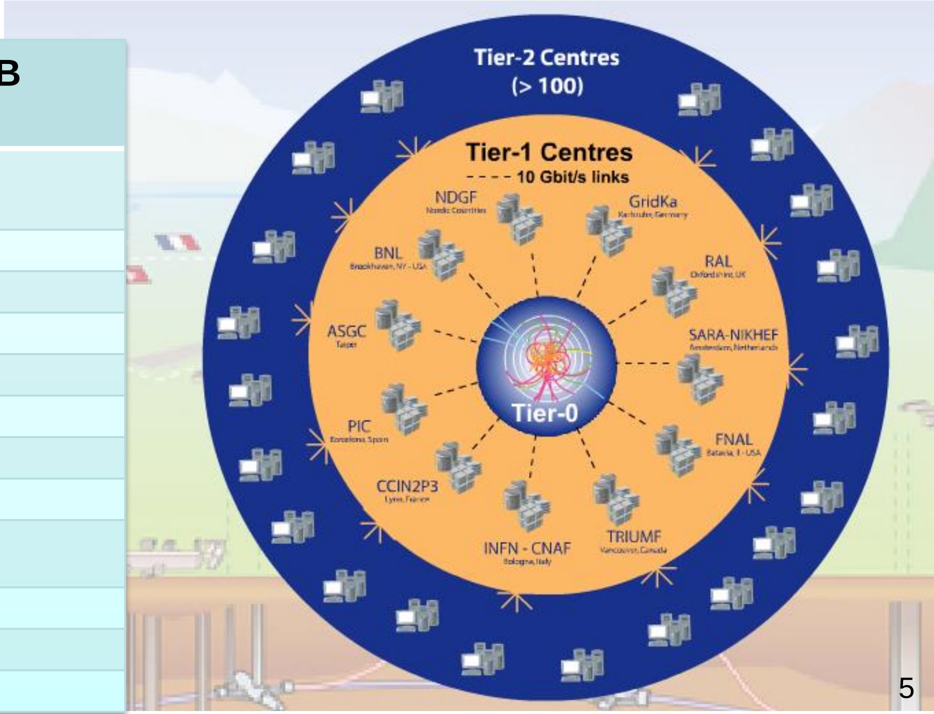


Our Use Case (2): Distributed Environment

- LHC Computing Grid: Distributed computing infrastructure to provide the production and analysis environments for the LHC experiments
  - Infrastructure leveraged by multiplicity of projects and services in the wider scientific community, such as EGEE, European Grid Infrastructure (EGI), US Open Science Grid (OSG)
  - Applications ranging from Astronomy over Earth and Life Sciences or Multimedia to Finance
- Data Replication and local Archiving plays a core role
  - Grid paradigm for distributed computing; each Tier-1 centre replicates > 10% of overall data
  - Each T1 site provides a logically independent archive system, independent technology choices
  - >> 200PB archived in Tier-0+Tier-1 - peaks of 13GB/s in and 25GB/s out between Collaborating Sites

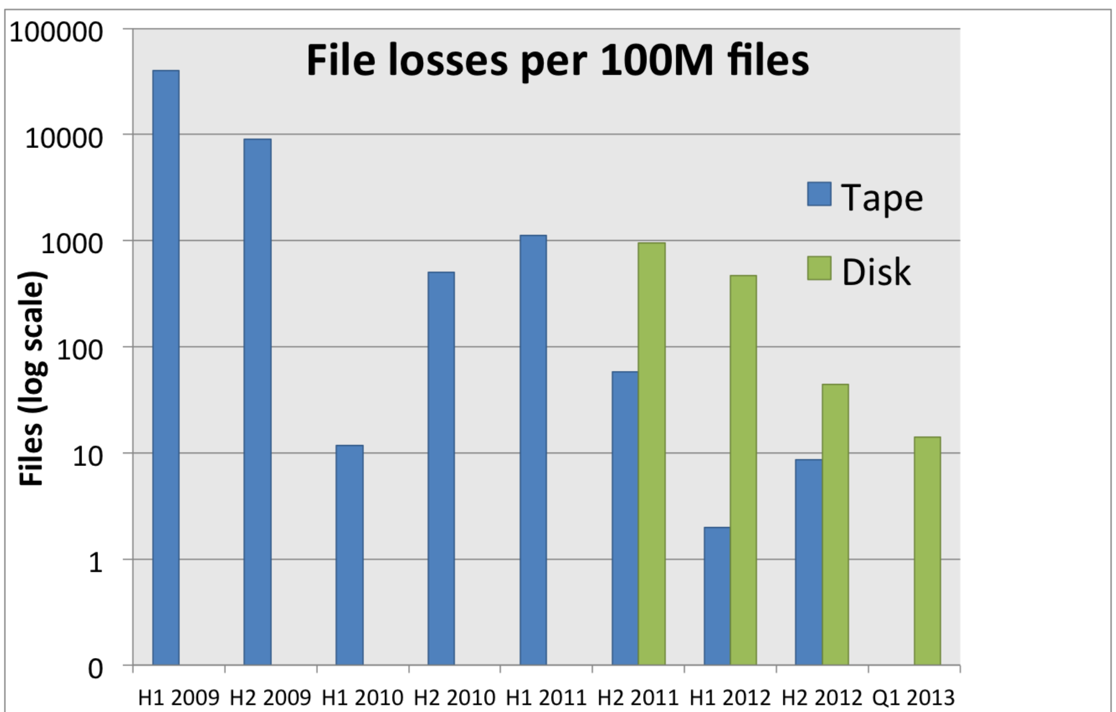
Monthly data transfers - T0 and T1's

Tier	Receive PB (M files)	Serve PB (M files)
CERN	4.5 (4.9)	4.8 (3.1)
UK (RAL)	1.4 (2.4)	1.4 (2.1)
Germany (FZK)	1.4 (2.8)	1.2 (2.0)
USA (Fermilab)	1.5 (0.4)	1.0 (0.36)
France (Lyon)	1.2 (3.0)	1.3 (2.1)
Italy (INFN)	1.1 (1.8)	1.3 (1.5)
NL (NIKHEF)	0.8 (2.2)	0.6 (1.0)
Spain (PIC)	0.6 (1.2)	0.7 (1.9)
USA (BNL)	0.4 (0.4)	0.9 (1.7)
Taipei(ASGC)	0.6 (1.3)	0.5 (0.6)
Canada(TRIUMF)	0.5 (1.4)	0.4 (0.8)
Nordic(NorduGrid)	0.4 (0.5)	0.3 (0.7)



What overall reliability to expect?

- Bit rot -> file loss on both disk and tape is unavoidable and needs to be factored in at all stages
- Good news: it has been getting better
  - And will get even better with standards such as SCSI-4 logical block protection
  - But (y)our data grows fast as well
- Last year's persistent (tape) data loss: ~ 0.000032%
- Tape reliability still ~O(1) higher than disk-pool system
  - Note: single tape copy vs. 2 redundant copies on disk

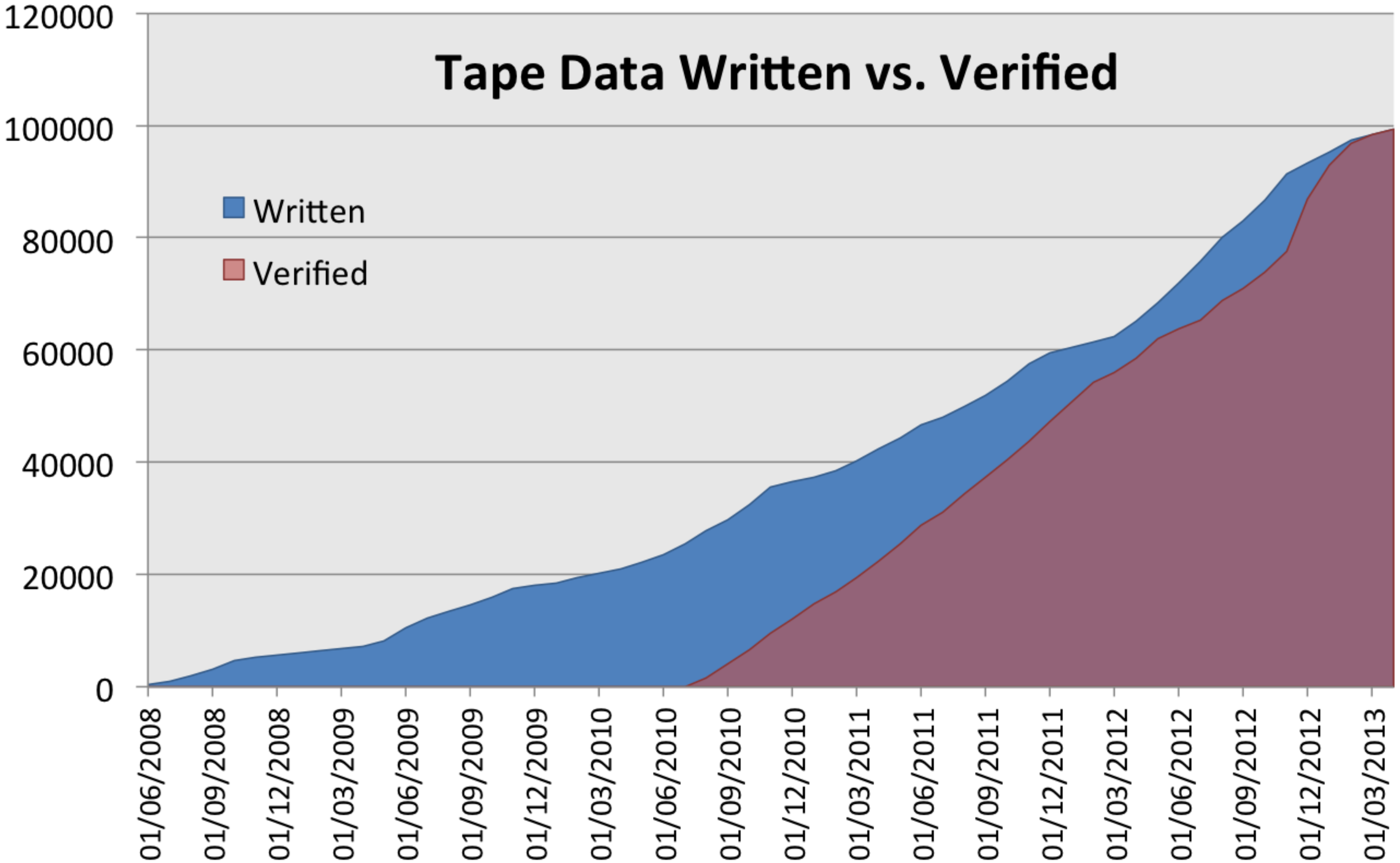


Active Archive Verification

- Data in the archive cannot just be written and forgotten about.
  - Q: can you retrieve my file?
  - A: let me check… err, sorry, we lost it.
- Proactive and regular verification of all (hot + cold) archive data required
  - Ensure cartridges can be mounted
  - Check data can be read+verified against metadata (checksum/size, …)
  - Do not wait until media migration to detect problems
- Several commercial solutions available on the market
  - Difficult integration with our storage systems
  - Not always check *your* metadata
- Recently implemented and deployed a background scanning engine:
  - Read back all newly filled tapes
  - Scan the whole archive over time, starting with least recent accessed tapes

Verification: first round completed!

- Up to 10-12 drives (~10%) for verification @ 90% efficiency
- Turnaround time: ~2.6(!!) years @ ~1.26GB/s

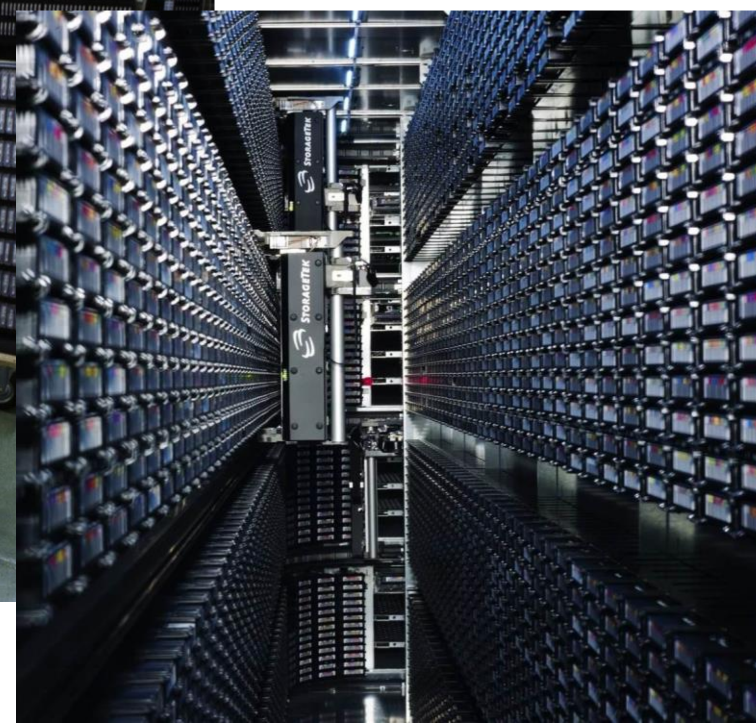


Media Repacking (1)

- Mass media migration or “repacking” required for
  - Higher-density media generations, and / or
  - Higher-density tape drives (reusing media)
  - Liberating space - tape library slots
- Media itself can last for 30 years, but not the infrastructure!

1970’s – GB’s

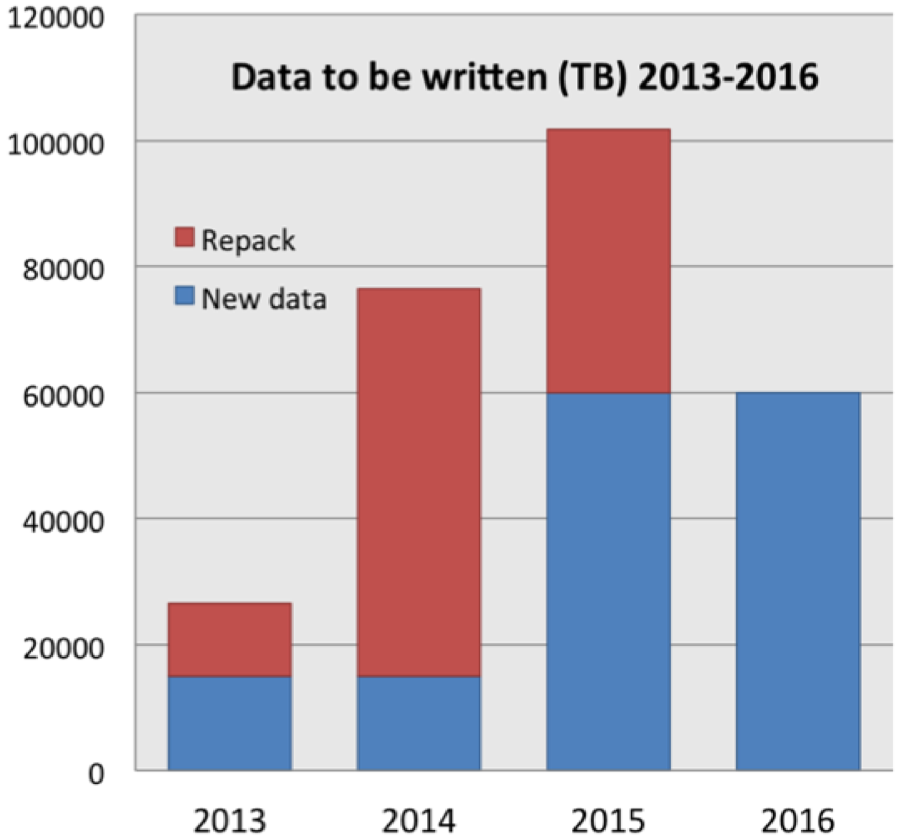
1990’s – TB’s



Today – PB’s

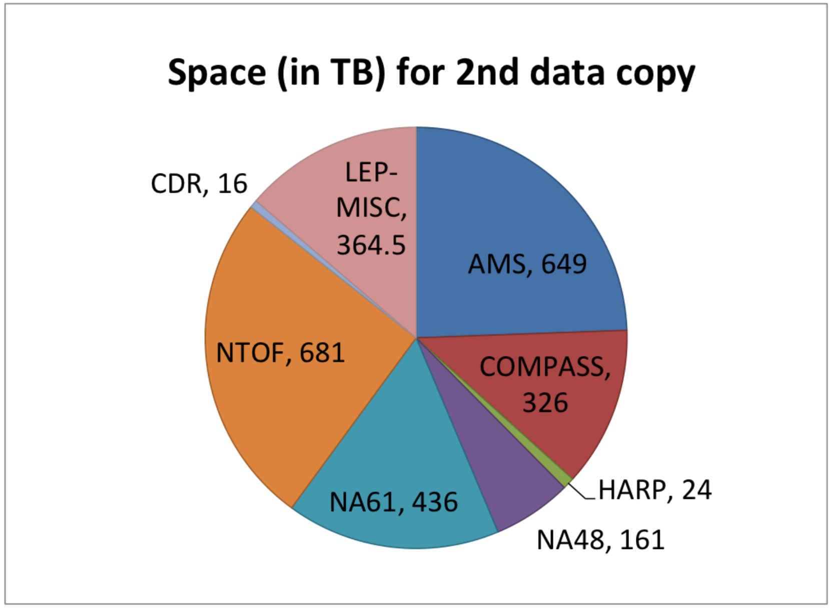
Media Repacking (2)

- Mass media migration or “repacking” required for
  - Higher-density media generations, and / or
  - Higher-density tape drives (reusing media)
  - Liberating space - tape library slots
- Media itself can last for 30 years, but not the infrastructure!
- Repack exercise is **proportional** to the **total size of archive** - and **not** to the fresh or active data
  - Scale infrastructure accordingly!
- Next Repack run @ CERN ~ Q4 2013 – 2015
  - ~100PB to migrate from over 52’000 cartridges
- Repack rates comparable to LHC data rates…
  - Over 3 GB/s sustained
  - Must be compatible with “regular” data taking (60PB/year in 2015++)
  - requires ~ 50% of drive capacity for 2 years
  - (… and some manpower to run it!)



Intra-Archive Replication

- By default, only one copy of a file is stored on tape.
- If justified and funded for, second copies can be generated on different tapes (or even different libraries / physical locations)
- Typically the case for “small” VO’s where data is stored only at CERN
  - Caveat: Everything gets “small” as media capacity growths
- Around 2.6PB of additional space (3% of total space)
- Replica failure / recovery is transparent to user/application

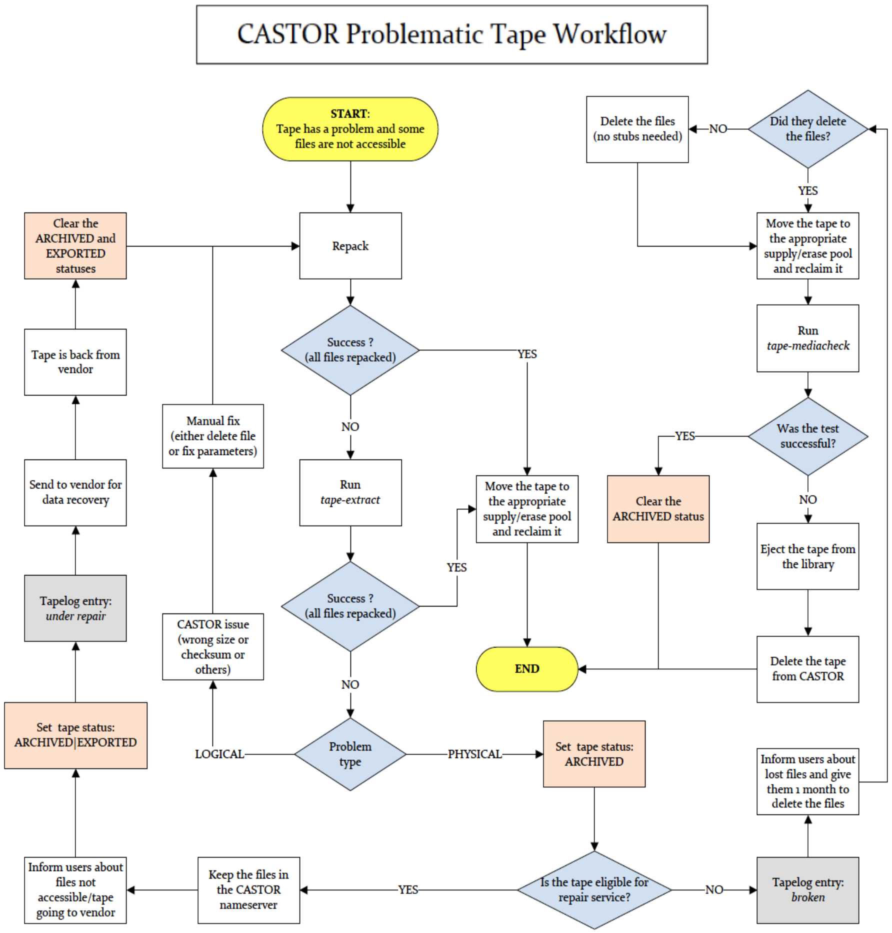


File inaccessible workflow

- Elaborated and well-established workflow for dealing with media problems
  - Several repair levels (local to vendor)
  - Different latency (hours to months to.. forever)

But..

- No ‘real’ standard exists for dealing with inaccessible files (whether temporarily or persistently lost)
  - Storage systems may answer differently when being asked for broken files
- There is no standard “push” mechanism
  - storage system to broadcast that it has detected inaccessible data
- Typical procedure: identify file owners, send mail(!) with list of inaccessible files
  - Works well for large / active VO’s and fresher data sets
  - Legacy VO’s (growing!) are most problematic …
    - Sometimes e-mail recipient doesn’t exist any longer!
    - Or doesn’t feel responsible any longer



Issues with inter-archive Sharing/Exchange

Bit-level integrity preservation across different (logical) sites / storage archives becomes even more complex

- The case for LHC data (each Tier-1 holds avg ~10% of LHC data)
- … and many other VO's at smaller scale

- Each site acts independently: potentially uses different storage technology/software; applies different policies for bit-level data preservation (ie verification, migration, intra-site replication)
- Cross-site data replication, consistency checking (and recovery) is organised at the application level provided by the different virtual organisations
  - Part of "data management" which is best done by the data owner (VO's)
- Complexity as multiple dimensions to consider
  - Effort ~ to (Applications/VO's) x (Storage Systems) x (Sites)

- Fostering of inter-site collaborations for common practices, tools, services and interfaces for bit-level preservation and inter-site data recovery at the archive level

Conclusions

- Bit rot is unavoidable and hits archives “every day”.
- Ensuring data integrity on a PB-EB-scale, long-term archive is an (expensive) active and regular task. The effort is proportional to the total archive size.
- Well-established mechanisms (need to be) deployed for active data verification and migration at an intra-archive level.
- Opportunities arise for establishing collaborations fostering inter-site common practices, tools, or even standard interfaces and services for bit-level preservation and inter-site data recovery.

CERN

Reserve slides

The CERN Data Centre in Numbers

CERN IT Department

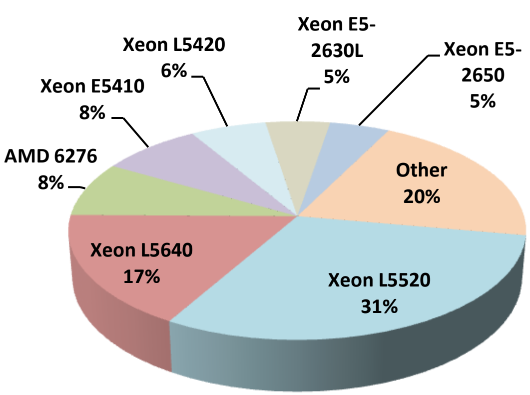
- Data Centre Operations (Tier 0)
  - 24x7 operator support and System Administration services to support 24x7 operation of all IT services.
  - Hardware installation & retirement
    - ~7,000 hardware movements/year; ~1800 disk failures/year
  - Management and Automation framework for large scale Linux clusters

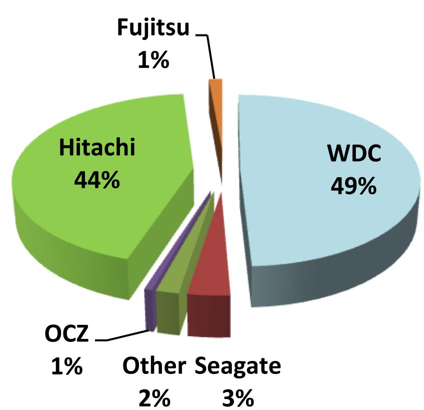
Racks	**1127**
Servers	**10,070**
Processors	**17,259**
Cores	**88,414**
HEPSpec06	**744,277**

Disks	**79,505**
Raw disk capacity (TiB)	**124,660**
Memory modules	**63,326**
Memory capacity (TiB)	***298***
RAID controllers	**3,091**

Tape Drives	**120**
Tape Cartridges	**52000**
Tape slots	**66000**
Data on Tape (PB)	**92**
High Speed Routers	**29**
Ethernet Switches	**874**
10 Gbps/100Gbps ports	**1396/74**
Switching Capacity	**6 Tbps**
1 Gbps ports	**27984**
10 Gbps ports	**5664**

IT Power Consumption	**2392 KW**
Total Power Consumption	**3929 KW**





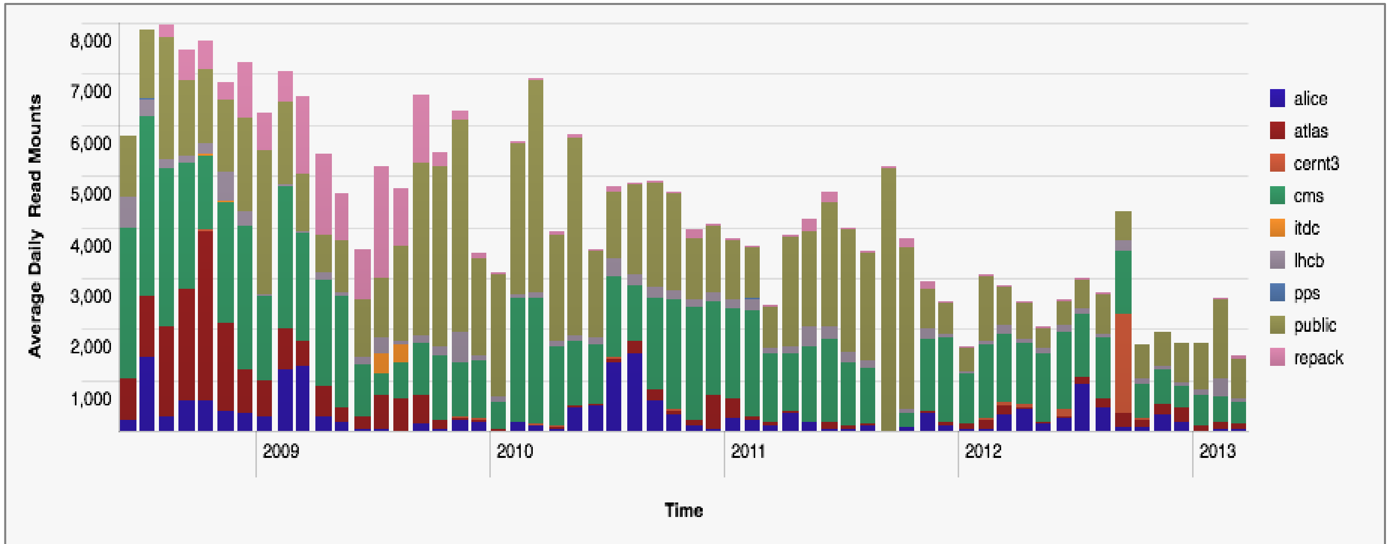
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

A sample case

- Some data in our archive system has been migrated over the past 22 years
- A user came to IT asking for the contents of a Wylbur data disk (the mainframe installed in 1976)
- The image of the disk had been copied to CERNVM and compressed (tersed) before sending to the archive
  - Terse was VM's equivalent of zip
- The archive system had been migrated from CERNVM to AIX to Linux
- The archive data had been migrated from IBM 3480, IBM 3490, STK 9940 and IBM 3592
- Terse was not available on Linux so data could not be read
- Luckily, we found an open source 'terse' 16bit OS/2 program and ported it to 64bit Linux to allow us to recover the files

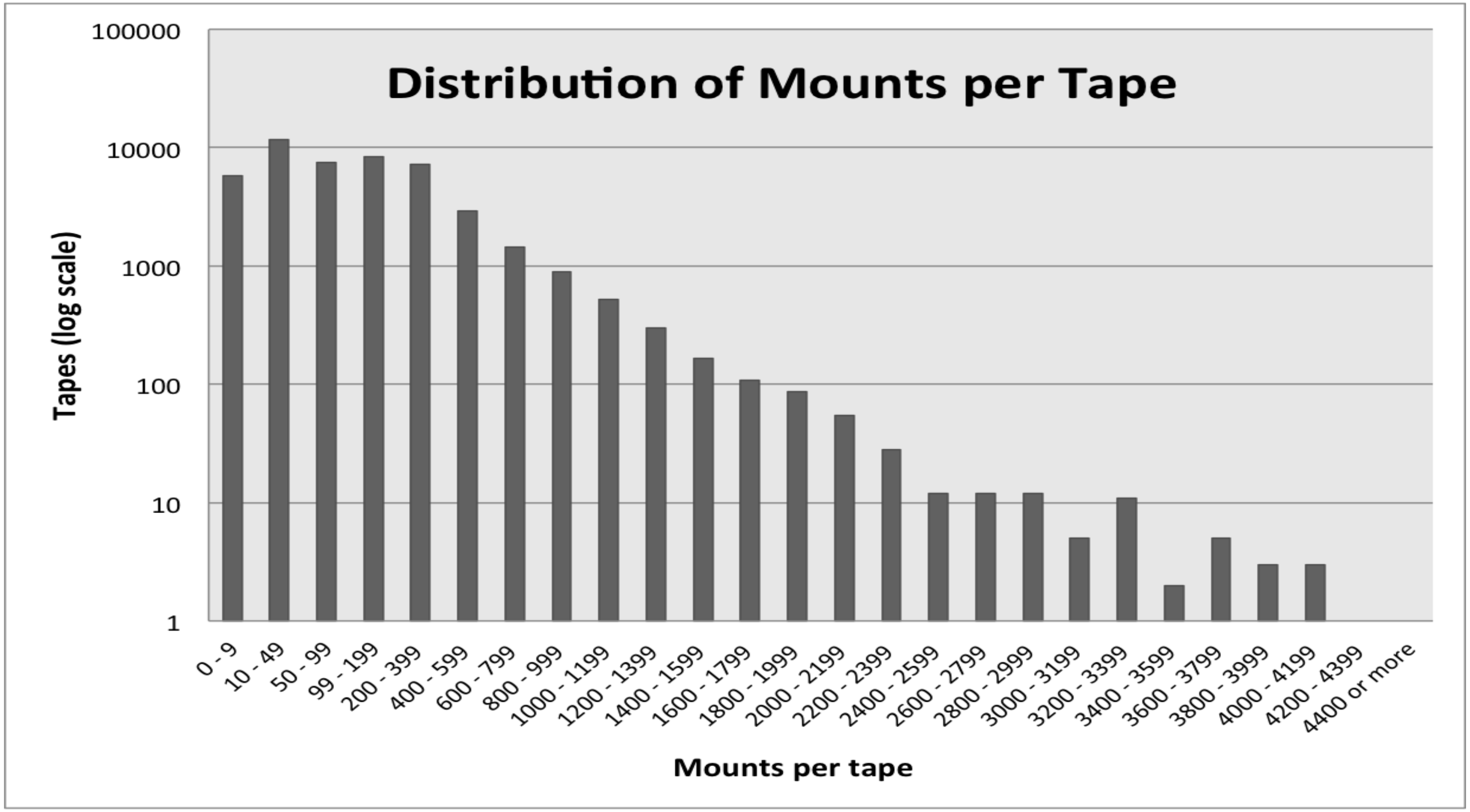
Maximising media longevity (1)

- Classic HSM systems: If user file is not on disk -> recall it from tape ASAP
  - Experiment data sets can be spread over hundreds of tapes
  - Many tapes get (re)mounted but files read is very low (1-2 files)
- Mounting and unmounting is the highest risk operation for tapes, robotics and drives.
  - Mechanical (robotics) failure can affect access to a large amount of media.
- Technology evolution moves against HSM:
  - Bigger tapes -> more files -> more mounts per tape -> reduced media lifetime
- Try to avoid using tape in random access; increase disk caches -> reduced tape mount rates
  - Use tape for ARCHIVING “colder” data, separate disk system for serving data to users



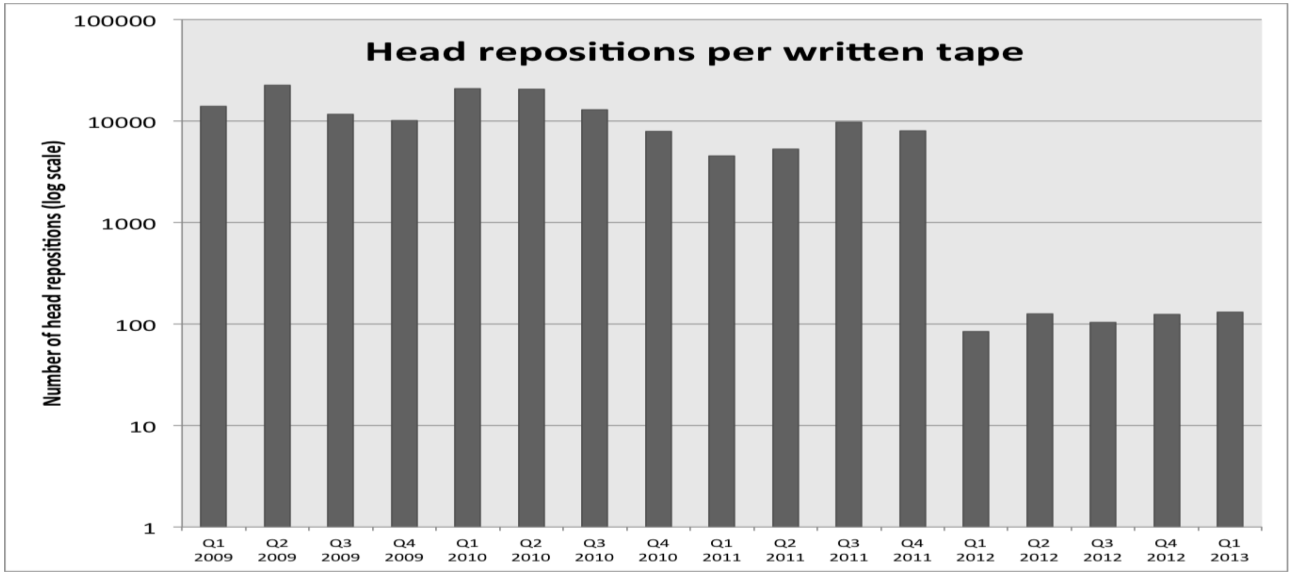
Maximising media longevity (2)

- Monitoring needed for identifying tapes mounted "too frequently" to avoid media wear
  - Automated decommissioning of media with mounts above threshold
  - Tape gets disabled and ticket generated for media repacking



Avoiding “shoe-shining”

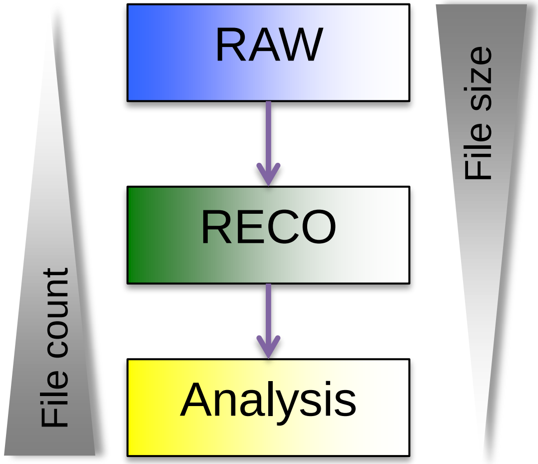
- Media wear also happens when writing small files to tape
  - By default, tape flushes buffers after close() of a tape file -> stop motion and rewind to end of last file (“head reposition”)
  - CASTOR uses ANSI AUL as tape format: 3 tape files per CASTOR file!
  - Performance (and media life time) killer in particular with new-generation drives (higher density -> more files)
- Can be avoided by using file aggregations (requires tape format change)
- Alternative found: logical (or “buffered”) tape marks
  - Prototyped by CERN, now fully integrated in Linux kernel
  - Synchronize only every 32GB worth of data
- Reduced number of head repositions from ~10000/tape to ~100/tape



LHC Data Types and Access Patterns

1. Detector Data Acquisition (DAQ) raw data
   - Sensor results from collision events
   - Writing of continuous streams of RAW format data when LHC is running
   - Packed into large files (1-10GB) stored ad eternum

2. Event Reconstruction
   - Reading of RAW data files, mostly streaming
   - Writing of RECO data files, size ~1GB

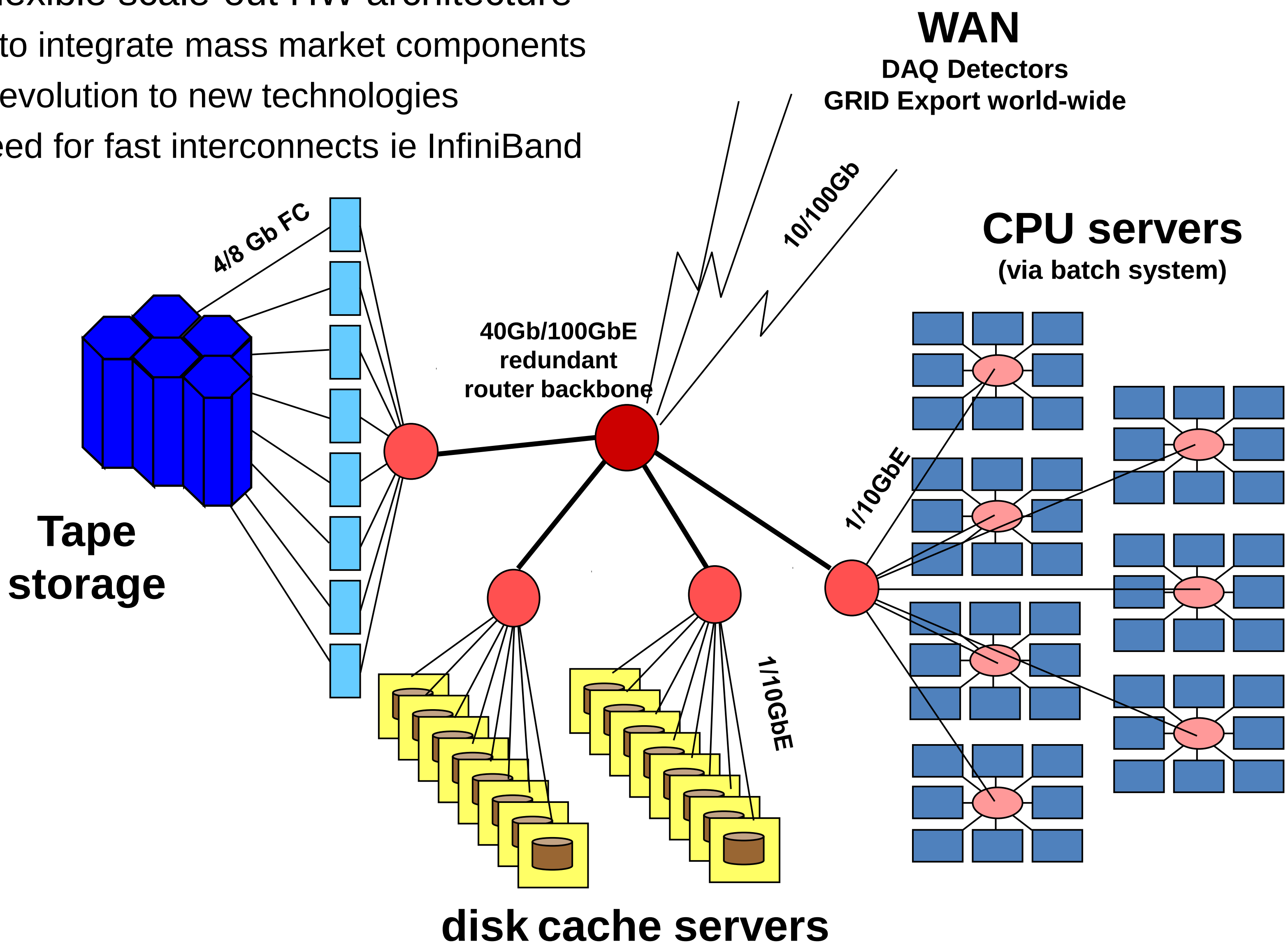


3. WAN Data Export to Grid Laboratories
   - Sequential reading of RAW and RECO data sent via WAN over the whole world

4. Data Analysis
   - Analysis by O(1000) end-users, skimming over O(TB) RECO data sets, mostly random I/O
   - O(M) small files written, high volatility
   - Bulk activity supposed to happen outside CERN

Hardware Architecture at CERN

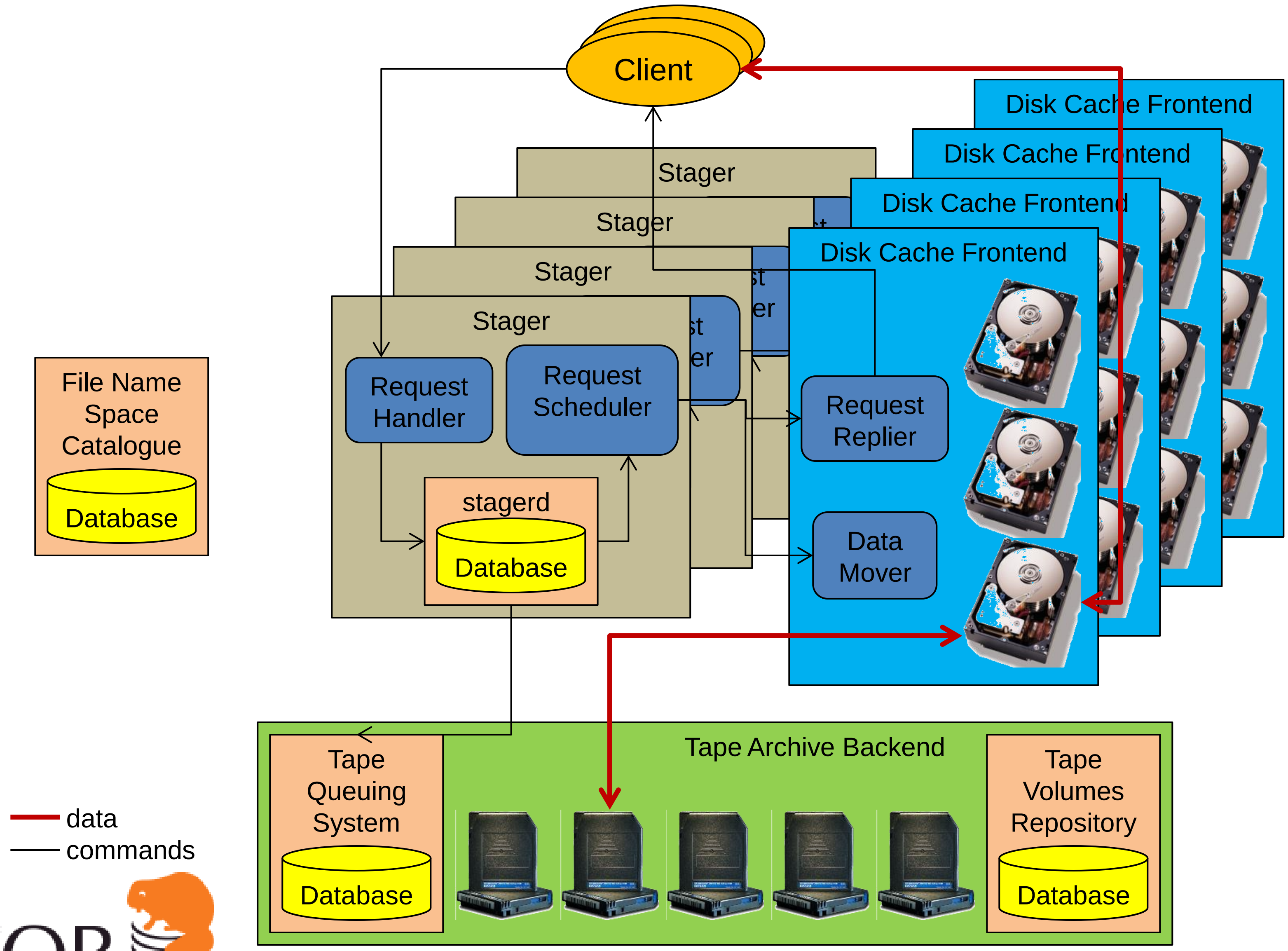
- Simple, flexible scale-out HW architecture
  - Easy to integrate mass market components
  - Easy evolution to new technologies
  - No need for fast interconnects ie InfiniBand



CERN Data Archive SW - CASTOR

- CERN Advanced STORage manager
  - Hierarchical Storage Management (HSM) system
    - Front-end disk and back-end tape layer
  - Long-lived storage of O(PB) data and O(M) files
  - File based with POSIX like hierarchical name space
- Made @ CERN
  - Used also at other Physics Research sites
- Original design dates ~14 years ago
  - In prod since 2001, many incarnations, data imported from previous CERN solutions (over 30 years old data!)

CASTOR SW Architecture



CASTOR
CERN Advanced STORage manager

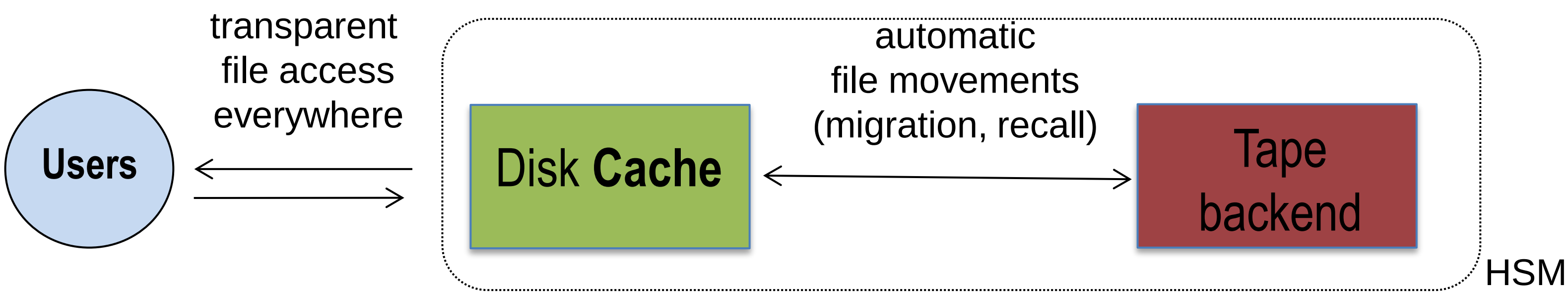
HSM model limitations

- Enforcing “traffic lights” not sufficient…
  - Unhappy users as Analysis Use Case not really working for them
  - …. but we didn’t expect much Analysis to happen at CERN!
- Neither is increasing disk caches
  - Even if 99% of required data is on disk, mount rates can be huge for remaining 1%

- Efficient, transparent HSM file access requires complex placement policies as well as in-depth understanding of data access patterns for each use-case
  - “Easy” to do for controlled data flows such as data acquisition from experiments or data export via WAN
  - “Hard” to “impossible” for 1000’s of autonomous end-users looking at different data in different ways

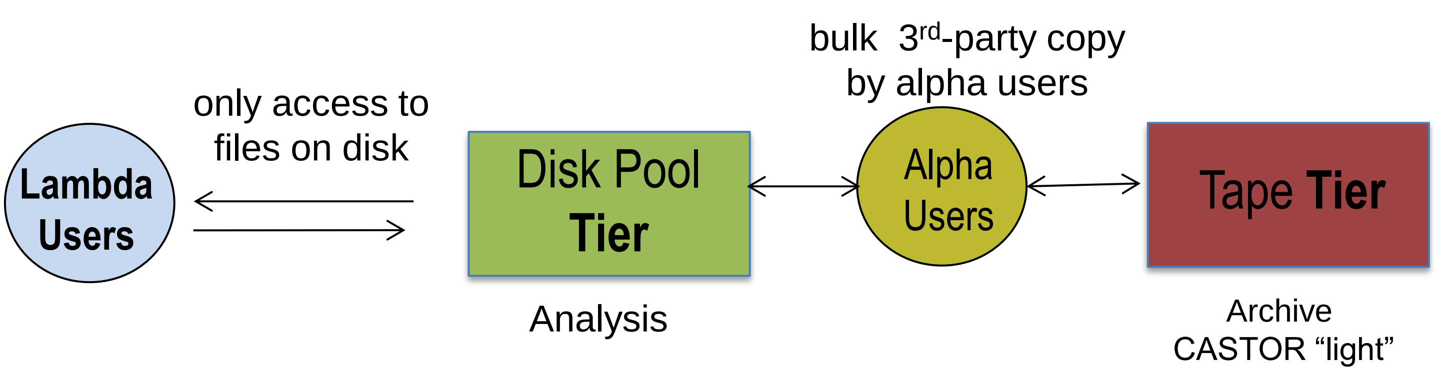
Model Transition

Move away from “transparent”, file/user based HSM



Model change from HSM to more loosely coupled Data Tiers

- Separate Analysis from other Use Cases
- Introduce a new (decoupled) system for random-access data analysis
- Tape access limited to privileged “alpha” users who manage the disk pools
    - Data “management” is better done by the data owner (experiment) who has the knowledge about data and access patterns



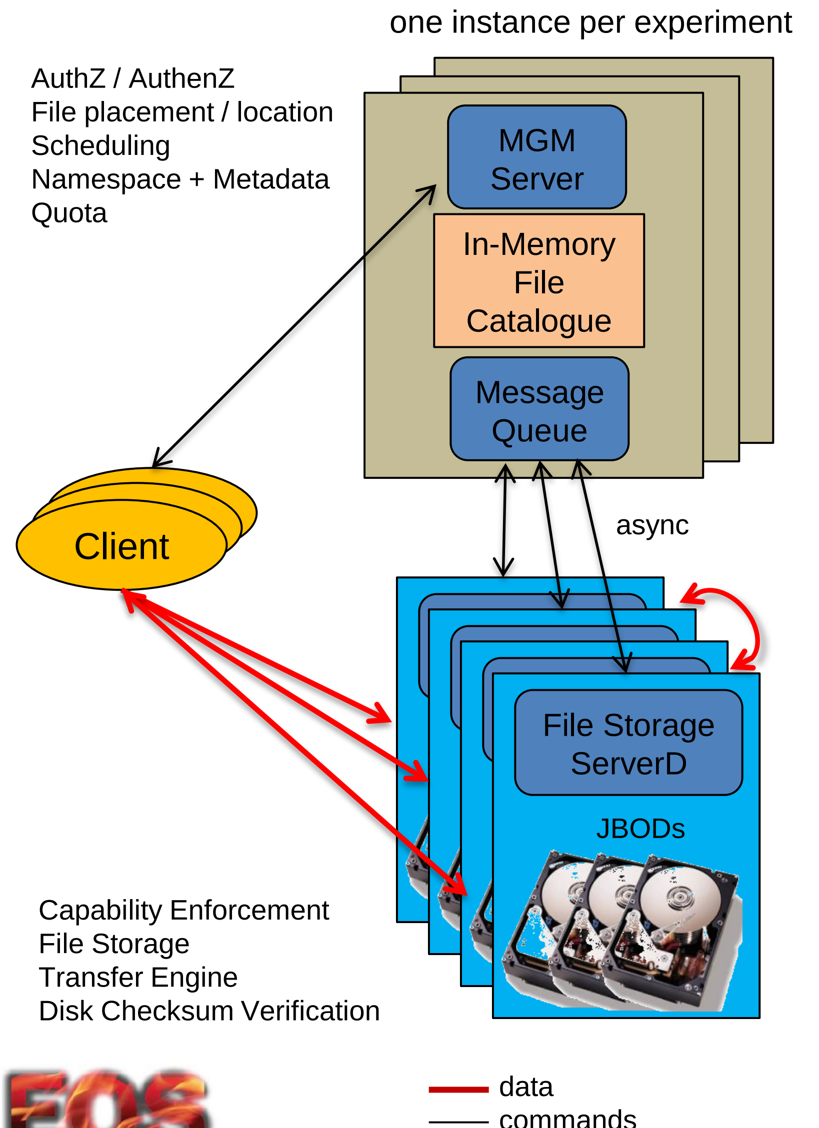
New Analysis Disk Pool System: EOS

- Evaluated several shared disk pool systems and decided to develop our own: EOS
  - Based on a well-known HEP data analysis framework (XROOT)

- Focus on Analysis Use Case
  - High-performance access: Low-latency (ms), high-concurrency O(10K) open/s; POSIX "light" interface
  - Scalable to O(1000) users, O(1B) files, O(10K) disks
  - Load-balancing and no single point of failure
  - Adopting ideas from Hadoop, Lustre, et al.

- No Tape – neither HSM nor Backup
  - Tunable reliability on commodity disk HW – tunable file replication (2-5) across network (JBOD)
  - Built-in media verification (load-adaptative checksum scanning)
  - Asynchronous replacing and growing of disks and servers w/o downtime

EOS Architecture and Usage Numbers



(two replicas per file)

Has Tape A Future at CERN?

- Tapes have gotten a “bad reputation” for some Use Cases
  - Slow in Random Access mode
  - Inefficient for writing small files
  - Bottlenecks for media verification/repack
  - .. comparable cost per raw (Peta)Byte as hard disks!

- But if used correctly, advantages prevail:
  - Faster (2x) than disk in sequential access mode
  - Orders of magnitude more reliable than disks
  - No power required to preserve the data
  - Less physical volume required per (Peta)Byte
  - Nobody can delete hundreds of PB in minutes
  - Ideal for “colder” data – historical data, as “freshest” data being most active

- Bottom line: if not used for random access, tapes have a clear role also in the future for ARCHIVING