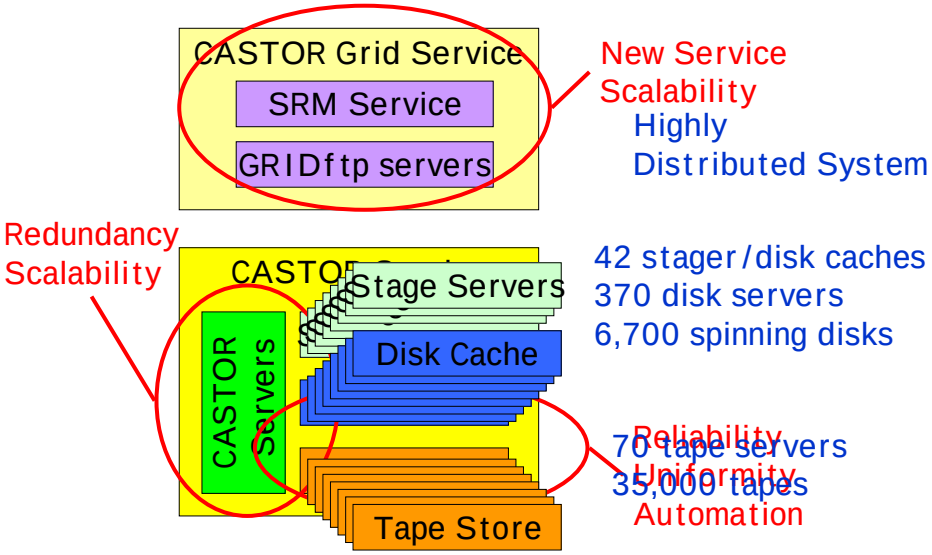
A selection of slides to illustrate issues raised at the CASTOR external operations meeting

CASTOR external operation/Barcelona

Data Services team:
Vladimír Bahyl, Hugo Caçote, Charles Curran,
Jan van Eldik, David Hughes, Gordon Lee,
Tony Osborne, Tim Smith

CERN Managed Storage

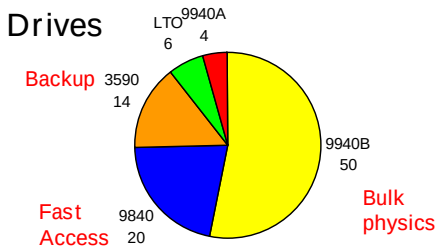


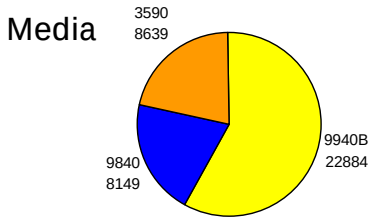
CASTOR Service

- Running experiments
  - CDR for NA48, COMPASS, Ntof
    - Experiment peaks of 120MB/s
      - Combined average 10TB/day
      - Sustained 10MB/s per dedicated 9940B drive
    - Record 1.5 PB in 2004
  - Pseudo-online analysis
- Experiments in the analysis phase
  - LEP and Fixed Target
- LHC experiments in construction
  - Data production / analysis (Tier0/1 operations)
  - Test beam CDR

Tape Service

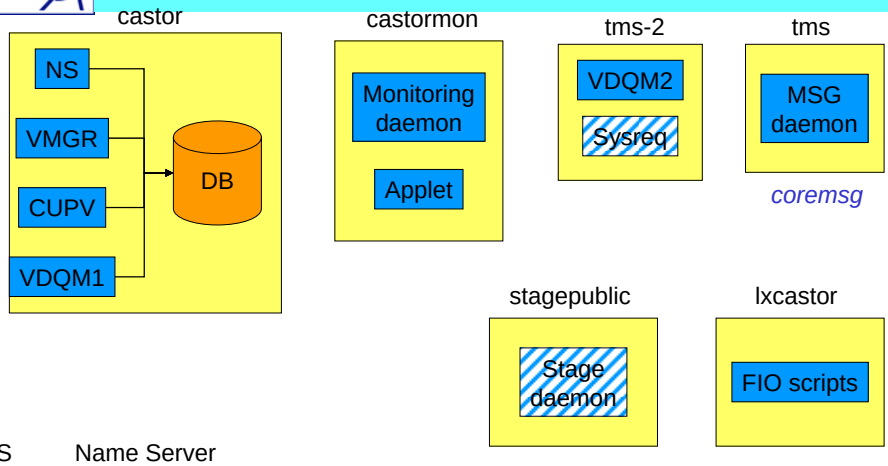
- 70 tape servers (Linux)
  - (mostly) Single FibreChannel attached drives
- 2 symmetric robotic installations
  - 5 x STK 9310 Silos in each





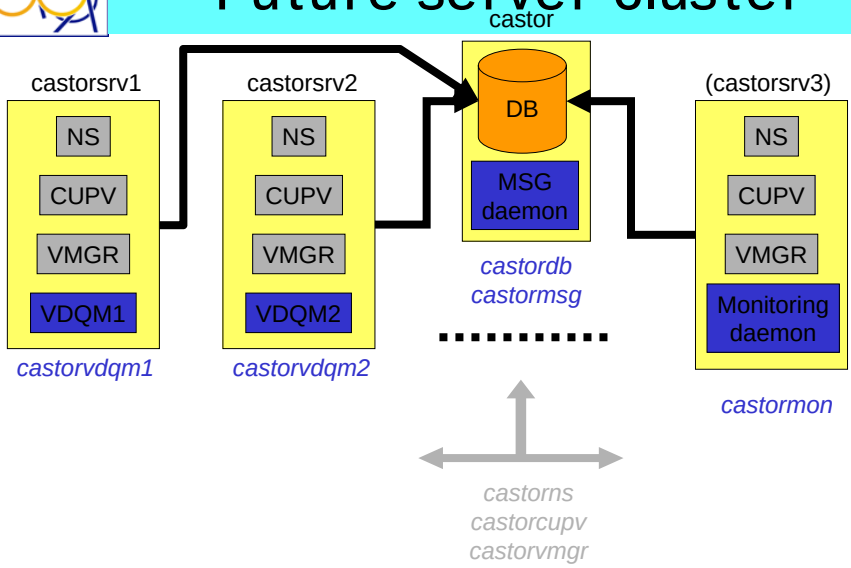
- CASTOR Central Services Cluster
  - From a talk by Vlado Bahyl

Today



NS	Name Server
VMGR	Volume Manager
CUPV	User Privilege Daemon
VDQM	Volume Drive Queue Manager

Future server cluster



- Tape Server performance in CASTOR production
  - From a talk by Jan van Eldik

- Study of tape access patterns, as done today by LHC experiments
- Based on log file analysis
- Results are preliminary…

infrastructure

- 2 robotic installations
- 70 tapeserver + tapedrive combinations:
  - 50 9940 200GB tapes (experiment data)
    - **30 MB/s drive speed**
  - 20 9840 20GB tapes (user data)
    - 10 MB/s drive speed

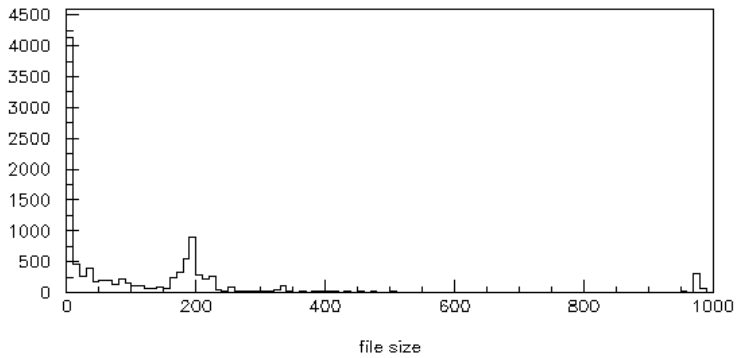
Requests managed by CASTOR

A week in the life of tpsrv115…

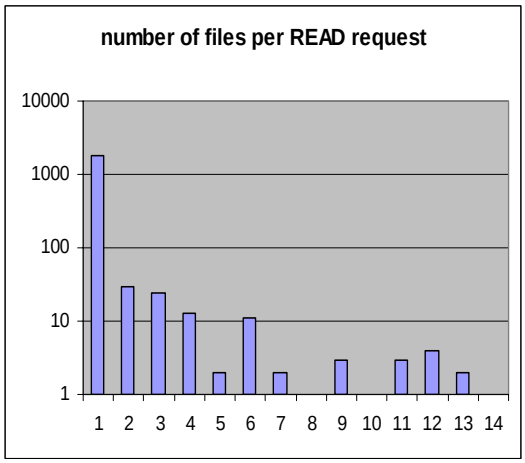
- 9940 experiment data LHC experiments!
- 1912 READ requests
  - 3033 files, 657.1 GB
  - 90% of requests are for 1 file…
- 98 WRITE requests
  - 7883 files, 916.6 GB
  - grouped by CASTOR

~13 requests per hour, 2.8 MB/s

"Mass" Storage System?



efficient reading?



A typical READ request

0 s : Request READ P12796, 1 file
10 s : ACS response
32 s : MIR read
117 s : read 1027 Mbytes in 53 sec = 18.5 MB/s
170 s : Release
171 s : Finish
280 s : Unmount 109 sec

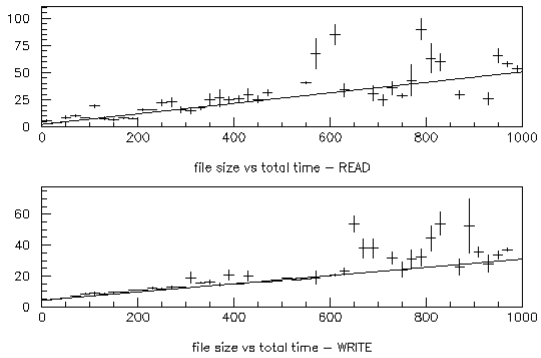
total request time	**280 s**	
reading	**53 s**	**(18.5 MB/s)**
overhead	**227 s**	**(3.5 MB/s)**

A typical WRITE request

```
  0 s ::  Request WRITE P03426, 7 files
 13 s ::  ACS response
 45 s ::  MIR read
 64 s ::      write 1396997840 bytes in 92 sec = 14.5 MB/s
156 s ::      write   988292395 bytes in 37 sec = 25.5 MB/s
193 s ::      write        446464 bytes in   5 sec =   0.1 MB/s
198 s ::      write        417792 bytes in   4 sec =   0.1 MB/s
202 s ::      write        397312 bytes in   4 sec =   0.1 MB/s
206 s ::      write        385024 bytes in   5 sec =   0.1 MB/s
211 s ::      write        344064 bytes in   4 sec =   0.1 MB/s
215 s ::  Release
215 s ::  Finish
272 s ::  Unmount    57 sec
```

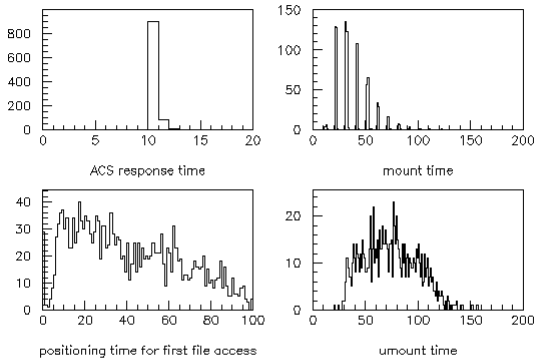
total request time	272 s	
writing	151 s	(15 MB/s)
overhead	121 s	(8.4 MB/s)

Drive speed



Drive speed 30 MB/s observed (more or less)
Note: 4 sec offset for writing tapemarks!

Read requests - overhead



Average overhead: 165 seconds

Faster robotics? Minimize number of mounts!

By the way…

- 1 file per READ request… *pathetic!*
- "new" READ request for mounted tape
  - 65% of requests! Up to 170 requests!
  - Tape must be dizzy for rewinding
  - Must be a bug somewhere…

How to reduce overhead?

- Minimize number of mounts: CASTOR
- Improve grouping of requests: CASTOR
- Stop writing small files to tape…
- Faster robotics