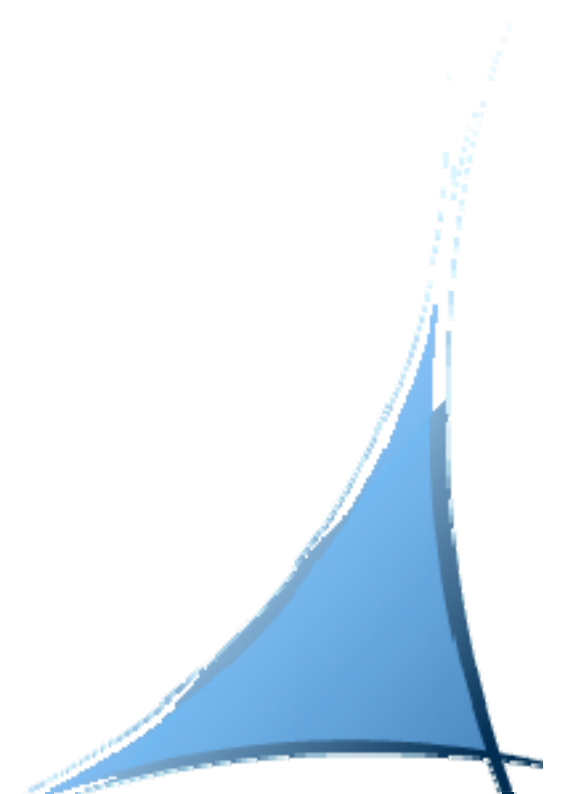
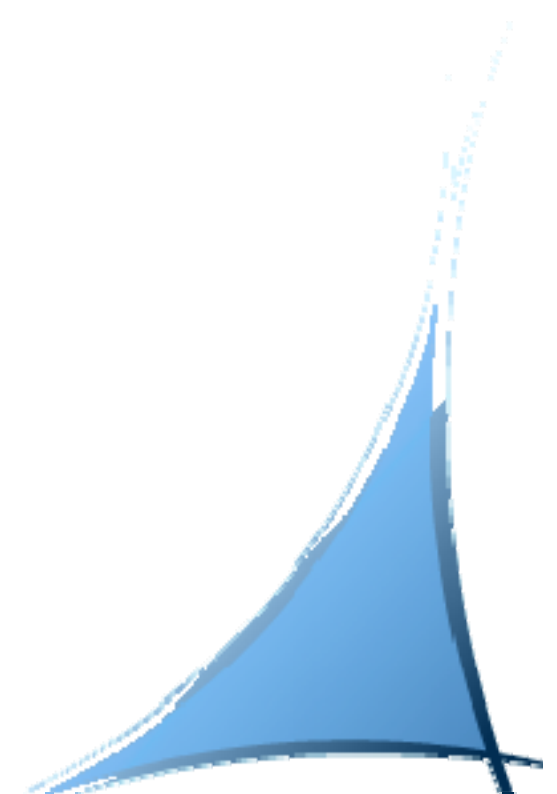


castor2 @ pic status report



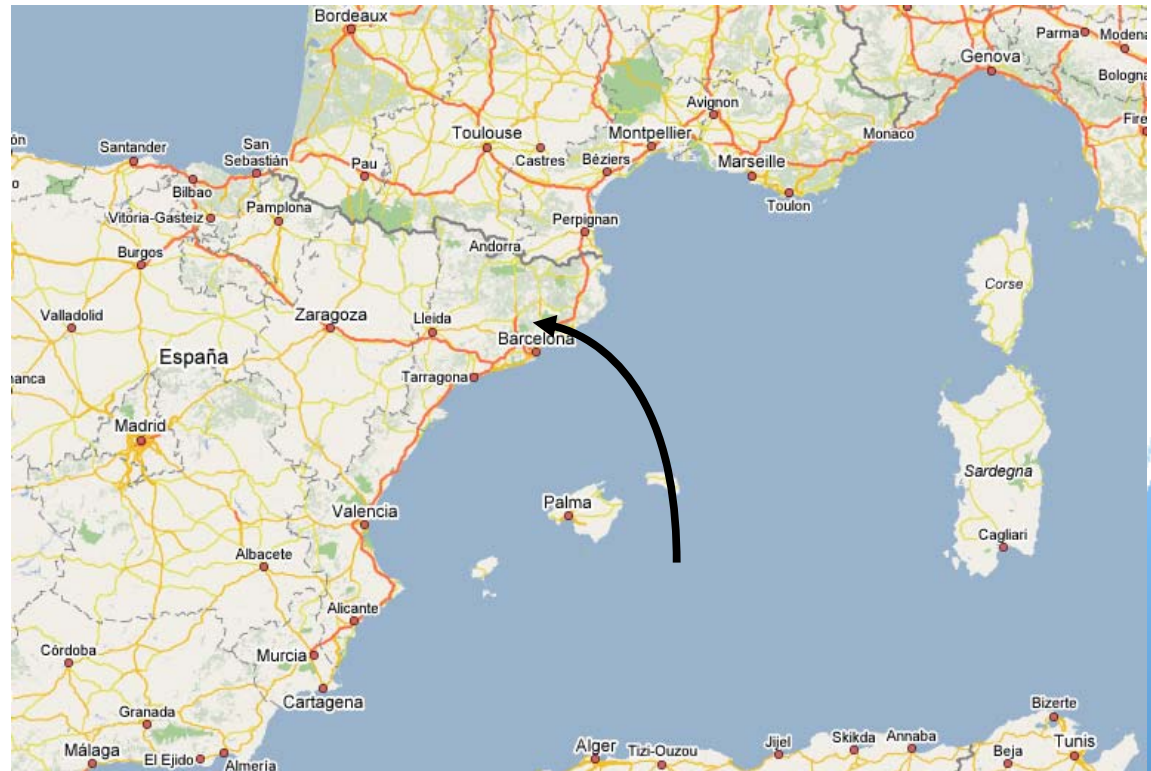
Index

- Presentation & numbers: magnitude of the installation
- History of Castor @ PIC
- 2 similar Storage Products?
- Castor2 test infrastructure & status
- Future plans



Who we are?

- PIC is the Spanish tier-1 in the LHC project for Atlas, CMS and LHCb (no official collaboration from any Spanish institute with the Alice project, but this may change)
- PIC stands for Port d'Informació Científica (Scientific Information Harbour)
- PIC is physically located at the UAB Bellaterra campus near Barcelona



pic in numbers

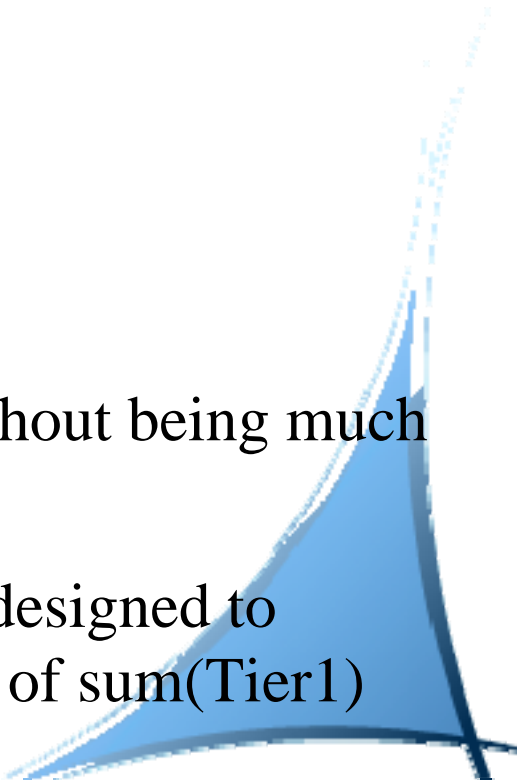
We plan to handle:

year	2007	2008	2009	2010
CPU (kS I2K)	500	1700	2700	5400
Disk (TB)	220	900	1600	2900
Tape (TB)	250	1200	2400	4500

These are just approximate capacities, that try to take into account the revised requirements from the experiments. The MoU for Spain has not been signed yet

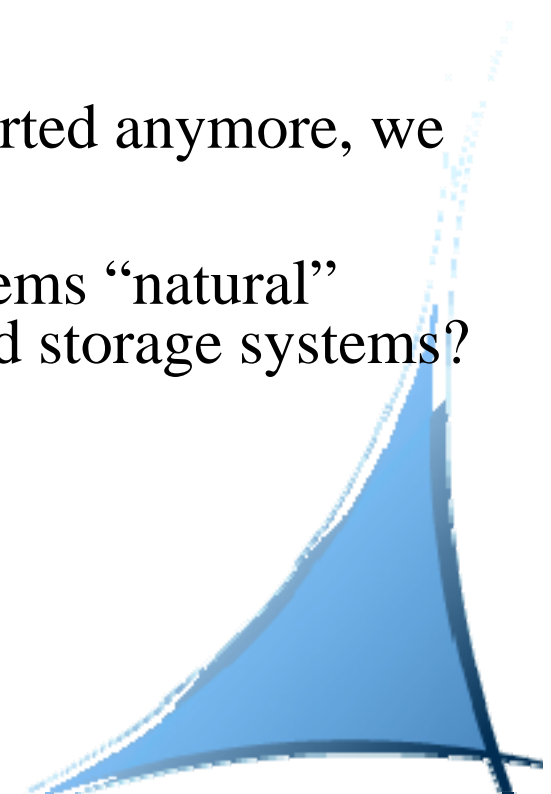


History of Castor (1 and 2) at PIC

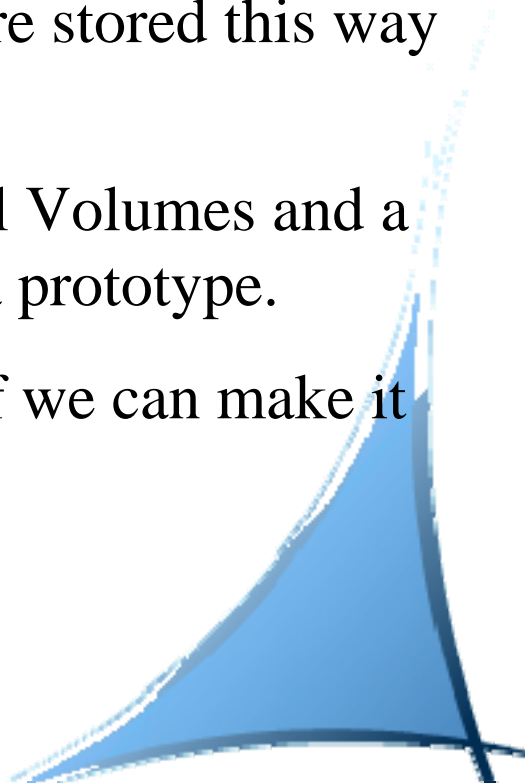
- From the beginning of PIC (in 2003) we have been using Castor(1)
 - Storagetek L5500 “Powderhorn” which can handle 9940B as well as LTO-3. Currently 7 9940B drives and no LTO-3.
 - Beginning of 2005 we installed an IBM 3584 L52 tape library with 4 LTO-3 drives. As this robot is different to the IBM supported by CERN, we added this support to Castor1.
 - Castor1 posed some problems for us
 - Bad handling of failing disk servers
 - Poor scalability
 - No file replication (for performance / data resilience)
 - No support for small files
 - We hoped Castor2 would remove these problems without being much more complicated to install/manage than Castor1
 - Maybe this was an incorrect assumption, as Castor2 designed to handle huge loads, whereas PIC “only” handles ~4% of sum(Tier1)
- 

TWO Storage Managers in a small Tier1?

- Because of the slippage in delivery of Castor2, we were forced to look for alternatives for 2006
- So we decided to set up a dCache instance to provide “Disk1Tape0”
 - In spite of poor documentation, one part-time engineer managed to install a pilot dCache in 3 months. We then ramped it up to 50 TB net capacity in 3 weeks with no problems and went directly into CMS CSA06
 - CMS CSA06 used our dCache installation successfully
 - ATLAS is now using dCache successfully as well
- Due to the limitations of Castor1 and that it is not supported anymore, we have to move to Castor2
- dCache for Disk1Tape0 and Castor2 for Disk0Tape1 seems “natural” ...but would we be the only T1 with 2 complex managed storage systems?
- Other possibilities:
 - “Hide” Castor2 “behind” dCache HSM interface
 - Switch to a simpler tape-handling system “behind” dCache
 - Use Castor2 for Disk1Tape0 and decommission dCache



What about efficient handling of small files?

- We don't believe that the “small file problem” will go away.
 - Even if it would for LHC experiments, the problem exists in other projects supported by PIC, ranging from medical images to cosmology simulations.
 - Our “Virtual Volume” technique using ISO images stored on tape is doing the job. For example, all files of the MAGIC telescope are stored this way automatically.
 - Architecturally, we see ways to make dCache use Virtual Volumes and a student has started a Master Engineering thesis to build a prototype.
 - We don't know enough about Castor2 (yet) to evaluate if we can make it use Virtual Volumes.
- 

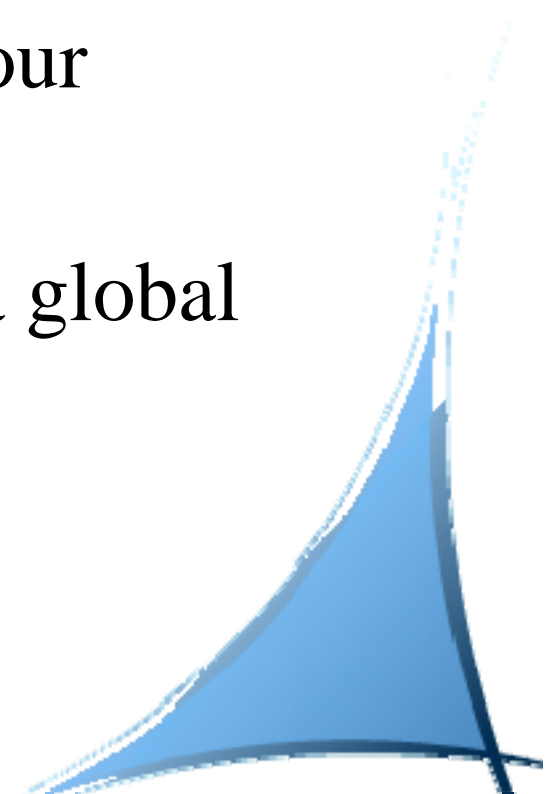
castor2 test

- We have a testing platform running 2.1.1-9 release with Oracle 10g and lsf 6.1
- We had to learn two commercial products
- Disk access working
 - can write/read files
- SRM1/Gridftp working
 - with the old servers
- Tape migration/recall working
- Thanks to all the castor support people at CERN!

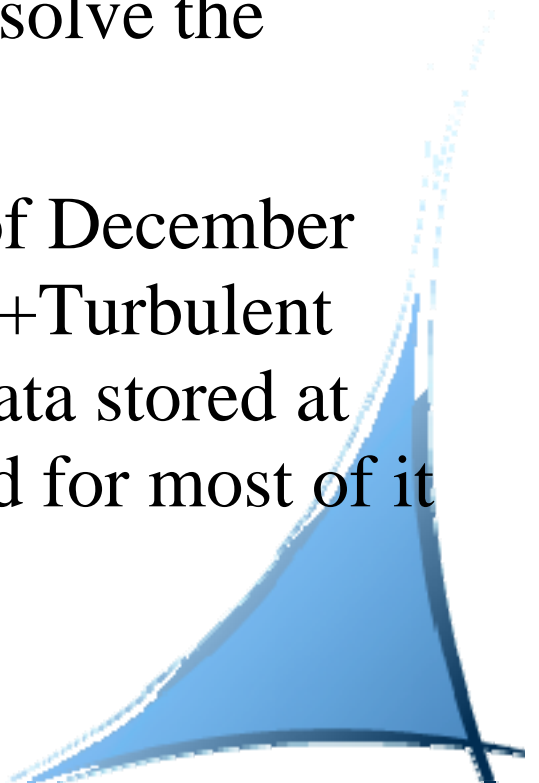


castor2 test - quattor

- Due to lack of quattor expertise, profiles were developed too slowly
- Problems with our quattor infrastructure have forced us to roll back configuration – and use shell scripts
- We expect to end up using quattor for our production system
- PIC management will soon undertake a global review of quattor @ PIC

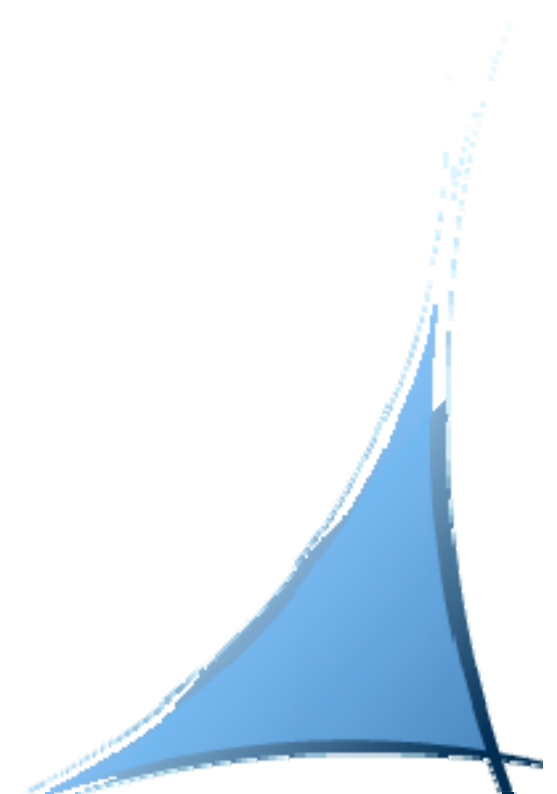


Future plans

- We will start working on the production system once the test infrastructure gives satisfactory results
 - That is planned for January 07 at the latest
 - Production system should be up and running by March 07
 - Early in Q1-2007 we need to understand and solve the “TWO mass storage” dilemma
 - Also need to deal with non-LHC. In fact, as of December 2006, MAGIC+CDF+K2K/T2K+Cosmology+Turbulent flow+Medical image represent a volume of data stored at PIC much larger than the one from LHC! And for most of it we act as the “Tier-0”.
- 

Future plans

- The final layout of the production infrastructure is still work in progress
- The current idea includes
 - 4 machines to run Oracle
 - nsdb, vmgrdb, cupvdb, srmdb?
 - stgdb, srmdb?
 - dlfdb
 - DataGuard of the other 3 machines



Future plans

- 3 machines for the stager
 - srv01 (stager, rh, mighunter, rtcpcd)
 - srv02 (rmmaster, expert, dlf)
 - srv03 (lsf)
 - 1 machine for the core services
 - shared with castor1
 - We have an spare machine as “hardware backup”
 - Currently working on designing some degree of HA on the stager
 - target: total failure of one single machine should not interrupt service
 - Also study if using LEMON is feasible and if it would help us deliver a better service
- 