



Instituto Nazionale di Fisica Nucleare  
Centro Nazionale per la Ricerca e  
Sviluppo nelle Tecnologie Informatiche e  
Telematiche

# **CASTOR2@CNAF**

## **Status and plans**

**Dejan Vitlacil**  
**Stefano Dal Pra**  
on behalf of  
**Tier1 Storage**

# Overview



- Storage & CNAF
- Setup
- History
- Future
- SRM
- Top 5

# Storage classes

- Implementation of 3 (2) Storage Classes needed for LHC
- Disk0Tape1 (D0T1) → CASTOR
- Disk1Tape0 (D1T0) → GPFS/StoRM
- *Disk1Tape1 (D1T1) → GPFS & TSM / StoRM*

# CASTOR2

## ■ Production Instance

- 4+1 core servers (stager, lsf, dlf, name-server) all SLC4 32-bit
- Name server, Stager Oracle RAC (SRMv.2.2 DB & Repack, Single Instance)
- 1 server for ACSLS (library management)
- 2 Tape libraries: STK L5500 (1 PB on-line), SL8500 (~ 2PB ; 6PB soon)
- 38 Tape servers, to reduce to 30
  - 6 9940b drives,
  - 4 LTO2 drives,
  - 8 T10000a, to be dismissed
  - 20 T10000b, to put into prod
- 5 servers srm v. 1.1
- 6 servers srm v. 2.2
- 45 disk-servers all XFS (nightly defragmentation) SLC4 32 and 64-bit
- 2 xrootd servers (peer and manager)

## ■ Tools, Monitoring and Alarms

- QUATTOR for configuration and installation
- LEMON for fabric monitoring (want to try LAS)
- A bunch of “Hand-made” tools for some Castor parts (by CERN / RAL)
- NAGIOS for proactive alarms
  - E-mail notification (SMS in progress)
  - Exclusion of problematic servers from service (via dns alias)

# CASTOR 2.1.7-17 deployment

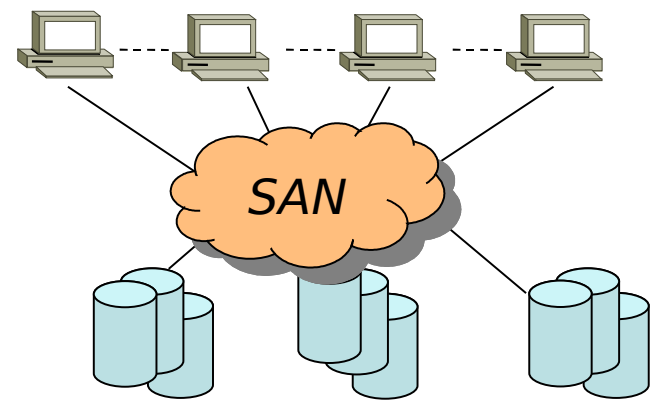


- Core services are on machines with SCSI disks, hardware RAID1, redundant power supplies SLC4 32 bit
- tape servers and disk servers have lower level hardware, like WN



~ 45 CASTOR disk servers attached to a SAN full redundancy FC 2Gb/s or 4Gb/s connections (dual controller HW and Qlogic SANsurfer Path Failover SW or Vendor Specific Software)

- STK L5500 silos (5500 slots, 200GB cartridges, capacity ~1.1 PB ) + SL8000 silos (7000 slots, 500GB/1TB cartridges, actual capacity ~2 PB )
- 38 tape drives, 3 Oracle databases (DLF, Stager, Nameserver) on ORACLE Real Application Cluster
- LSF plug-in for scheduling
- SRM v2 (2 front-ends), SRM v1 (phasing out)



38 tape servers

STK FlexLine 600, IBM FastT900, EMC Clarion

# CASTOR Setup

CASTOR core services v2.1.7-17 on 5 machines



Castor-4 (castorstager):  
rhtserver, stager, rtcpcclientd,  
MigHunter, RecHunter, expertd



Castor-6 (castorscheduler):  
Master LSF, jobManager,  
rmMasterDaemon



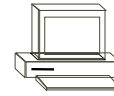
Castordlf (castor): dlfserver,  
repackserver



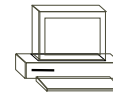
castorns01: nsdaemon,  
vmgrdaemon, vdqmserv,  
Cupvdaemonv



castor-8: old nsdaemon, To  
become castorns02



SRM v2.2. Oracle DB  
(Single Instance)



Name server Oracle DB  
(Oracle RAC)



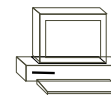
Stager Oracle DB  
(Oracle RAC)

1 SRMv1 endpoints, DNS balanced:  
• srm://castorsrm.cr.cnaf.infn.it:8443  
(used for disk pools with tape backend)  
**HOPE TO BE DISSMISSED SOON !**

2 SRMv2 endpoint:

srm://srm-v2.cr.cnaf.infn.it:8443

srm://srm-v2-cms.cr.cnaf.infn.it:8443



Repack Oracle DB  
(Single Instance)

## Castor setup (3)

<b>svcClass</b>	<b>exp</b>	<b>diskPool</b>	<b>diskServers</b>	<b>size (TB)</b>
<b>alice</b>	<b>ALICE</b>	<b>alice1</b>	<b>10</b>	<b>25.5</b>
<b>cms</b>	<b>CMS</b>	<b>cms1</b>	<b>15</b>	<b>93.22</b>
<b>atlas</b>	<b>ATLAS</b>	<b>atlas1</b>	<b>4</b>	<b>25.42</b>
<b>atlas_pin</b>	<b>ATLAS</b>	<b>atlas2</b>	<b>3</b>	<b>14.61</b>
<b>lhcb</b>	<b>LHCb</b>	<b>lhcb1</b>	<b>1</b>	<b>2.7</b>
<b>lhcb_raw</b>	<b>LHCb</b>	<b>lhcb_raw1</b>	<b>2</b>	<b>8.84</b>
<b>argo</b>	<b>ARGO</b>	<b>argo1</b>	<b>6</b>	<b>35</b>
<b>ams</b>	<b>AMS</b>	<b>ams1</b>	<b>1</b>	<b>3.53</b>
<b>pamela</b>	<b>PAMELA</b>	<b>pamela1</b>	<b>1</b>	<b>3.59</b>
<b>magic</b>	<b>MAGIC</b>	<b>archive1</b>	<b>2</b>	<b>22.19</b>
<b>babar</b>	<b>BABAR</b>			
<b>lvd</b>	<b>LVD</b>			
<b>virgo</b>	<b>VIRGO</b>			
<b>cdf</b>	<b>CDF</b>			

- Probably will change as soon as new storage arrive 1.6 PB (not all for Castor)

# CASTOR blues

Latest Upgrade: 2.1.7-17 (sep. 2008 - Today)

- Good integration with GridFTPv2
- More stability
- Better repack behaviour, Smarter diskpool usage
- Better overall performances
- Upgrade to 2.1.7-24 (mainly bugfixes) to be scheduled



## Our next Castor steps

- Get rid of SRM v1.1 as soon as possible (non LHC exp.)
- Move SRMv2 (1.3.x to 2.1-x) in progress, on new x86\_64 bit SLC4 machines
- CIP (Castor Information Provider) – (installation ongoing)
- New tape drives:
  - 20 drives (T10000B; 1TB,)
  - 20 tape servers (64-bit SLC4)
  - 10000 slots
  - Repack from older media (9940B, LTO2, T10000A) to T10000B scheduled soon
- Try to integrate Castor Lemon Sensor
- Try Lemon LAS (Lemon Alarm System)
- Develop a few more tools to keep Castor under control

## SRM at CNAF

We still run srmv1 (non LHC exp.)

We still run srmv2.2 (rpm: srmv2-1.3-28) lack of  
x86\_64 HW

Problems:

- DB deadlocks
- BE crashes
- Great help from Shaun DeWitt
- We'd like to implement a solution with multiple and mixed FE/BE as soon as possible (also multiple BE will be possible) => need the srmv2.7.1

## Wish list

- CIP (Castor Info Provider)
  - We wish it to be maintained and distributed directly by the Castor development team.
- Stream and Migration Control (implement better policies)
- Castor Admin and Monitoring tools
  - News? Still hand-written sql queries for obtaining some info?
  - News? Castor Lemon sensor still missing; is there any distributable plugin?
- Will support for older version (2.1.7-x) be granted until LHC re-start
- Reorganization of out of date information in nameservers and stagers services (nsenterclass, svcClass, ... )

