



# CNAF readiness

Luca dell'Agnello  
INFN-CNAF

Victoria – September 1 2007



# Hardware resources (1)

- Requested resources for 2007 (nearly) all in place
  - CPU power installed and available
    - 1300 KSI2k pledged for LHC available (out of a total of 3000 KSI2k)
  - Just delivered the last part of storage
    - In production at the end of Q3
    - At present 320 TB for LHC instead of 500 (out of a total of 800 TB installed)
  - Additional tape drives and tapes ordered
- Tenders for 2008 resources starting soon

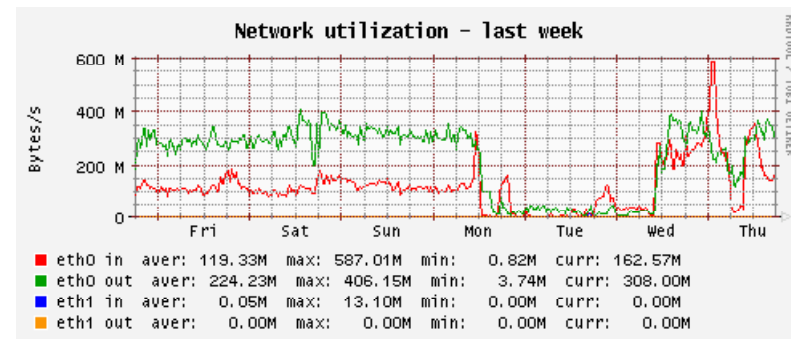


# Hardware resources (2)

- During 2008 CNAF computer room will be restructured to allow distribution of electric power up to 4 MW and the installation of up to 11 APC high density islands with water cooling
  - Preliminary: a new chiller added (end of August)
  - First step: 2 high density islands + 1 single row (56 APC racks) will be installed in a presently not used part of the computing center
  - All newest hw will be moved in this new location
  - Older hw will be upgraded to increase capacity and save power
    - All older CPU resources replaced with quad-core servers
    - All older storage systems with newer mounting 1 TB disks
- During 2008 will be able to install in 56 racks :
  - CPU: 7.5 M SpecInt2K
  - Disco: 3 PByte
  - Tape: 4.5 PByte
- After the end of the upgrade (Q1 2009) space for 70 additional racks (not all installed at once) will be available to install all the needed resources for the following years (beyond 2010)

# MSS

- Tape library presently installed (SL 5500) contains up to 1 PB on-line
- A larger tape library required
  - To contain up to 10 PB on-line
  - It will be acquired during 2008
  - Tendering phase started last July
- CASTOR installed for T1Dx SC
  - Improvement in stability with the July upgrade from v. 2.1.1-9 to v. 2.1.3-15
  - Still missing features (e.g. repack)
  - Still problems to be solved
    - Tape recall is an issue (10% failures in CMS August “exercise”)
    - Work-around in place and plan to implement several gc policies (September)
  - Plan to differentiate stagers for larger experiments and disk pools basing on type of use (more hw resources needed)

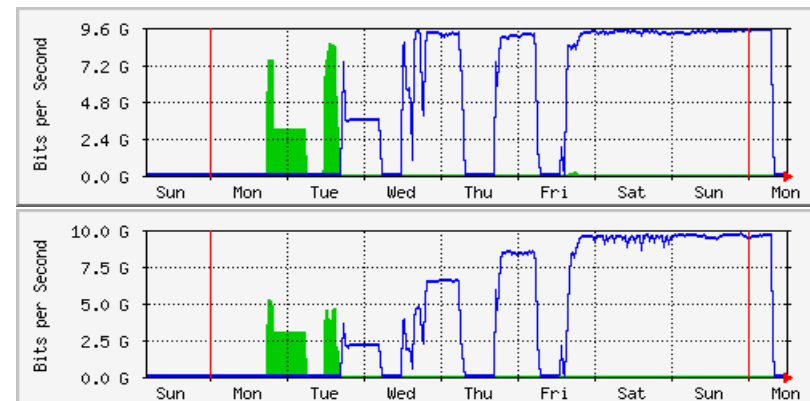
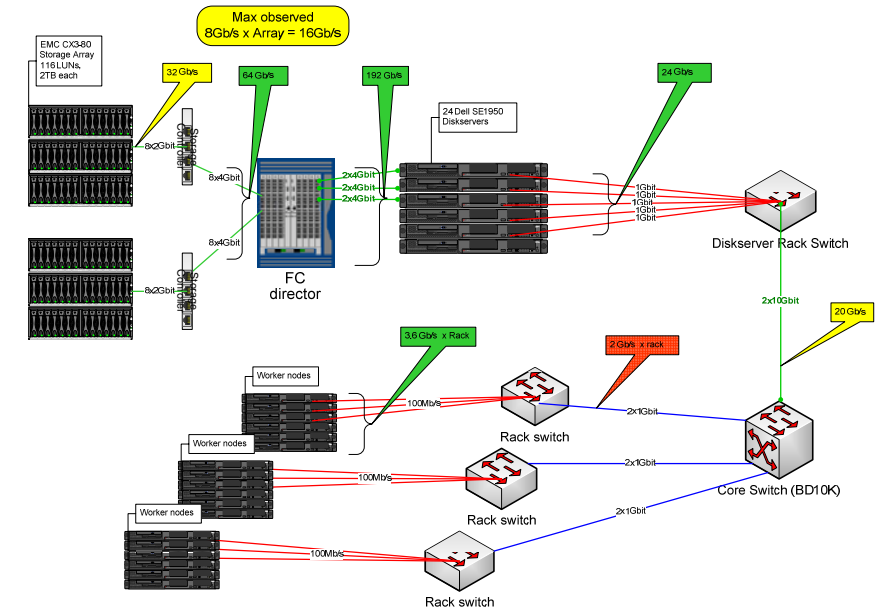


**Aggregate access to CASTOR**

# Storage system (T0D1) tests (1)

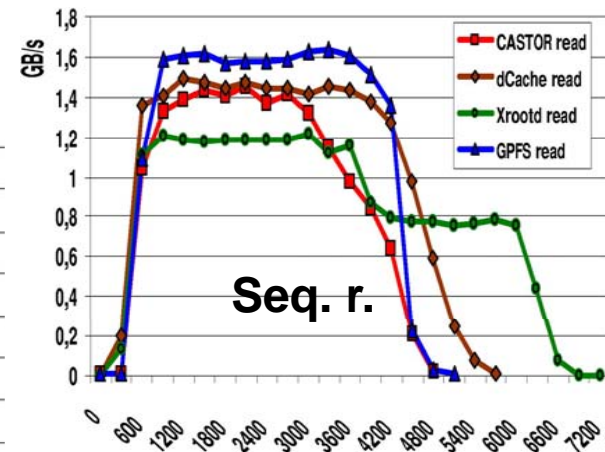
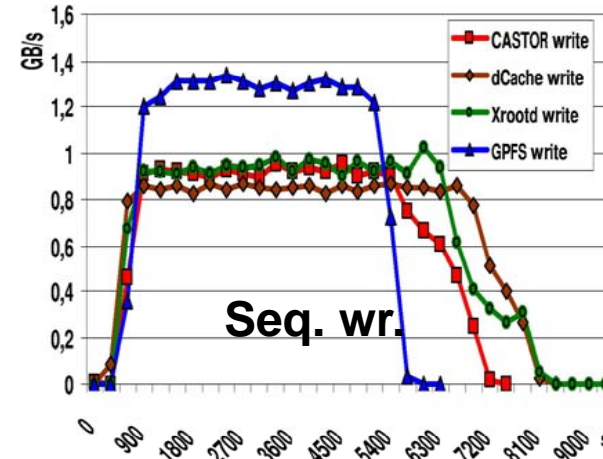
- Large scale validation test last Spring

- Involving ~200 TB of disk and 24 disk-servers
- ~1100 CPU slots (8 wn racks)
- 16 Gbps network available

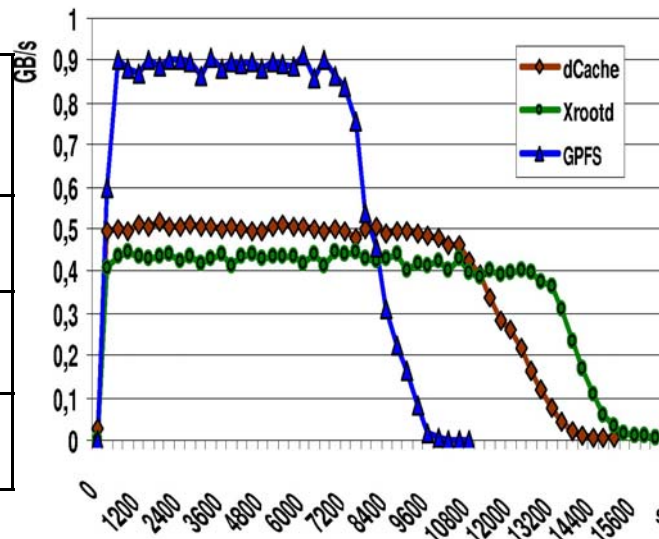


# Storage system (T0D1) tests (2)

- Sequential I/O throughput
  - 5GB files (block size=64K)
- Real analysis case: LHCb DaVinci
  - 1000 analysis jobs running on production farm and accessing data via dCache, Xrootd and GPFS



Test	Total time [s]	Average CPU usage (%)
GPFS	9500	36.43
Dcache	14500	22.69
Xrootd	16000	21.17





# 24 by 7 operations

- At present 24 by 7 operations in place for urgent issues (e.g. cooling malfunction, power interruption)
  - No physical presence on site off business hours
  - Automatic phone calls for alarms to on shift people
    - Alarm system effective ☹️
- Automatic e-mail and SMS alarms for service problems to relevant people
  - All resources are remotely manageable (including power cycling and console access)
  - Effort in making all services redundant (in progress 😊)
  - Best effort service (up to now)



# Cooperation with experiments

- Since the beginning people from experiments are part of our staff
  - both LHC and non LHC experiments
  - Main role is to ease use of T1 to experiments
    - Testing and verifying center operability for their experiment
    - Cope with experiment specific services (e.g. VOBOX management)
    - Explaining particular needs for their experiment (e.g. queue length)
    - Cooperating in resource administration (“simple” operations on batch systems for experiment queues or on CASTOR, FTS etc...)
    - Translate issues of the users ☺
- Model proven to work
  - Experiments without on-site support result in more latency for problem solutions
  - Hiring of new people is on going