

# Storage Infrastructure at the INFN LHC Tier-1



INFN – CNAF is the main Italian LHC data centre and one of the few primary level (Tier-1s) LHC data centers in the world. LHC experiments will produce up to tens of petabytes of data which need to be stored at the various LHC computing centers for reconstruction and analysis.

Different data management policies are provided, identified by 3 distinct Storage Classes:

- **Disk0-Tape1 (D0T1):** data migrated to tape and deleted from disk when the staging area is full. Disk and tape space is managed by the system → **CASTOR** (testing **GPFS/TSM/StoRM**)
- **Disk1-Tape0 (D1T0):** data always available on disk, never migrated to tape, never deleted by the system. Space management is demanded to the experiment → **GPFS/StoRM**
- **Disk1-Tape1 (D1T1):** large buffer on disk with a tape back-end. No garbage collector, space is managed by the experiment.

Moreover, an Oracle clustered database infrastructure is deployed for relational data → **CASTOR** (moving to **GPFS/TSM/StoRM**)

## CASTOR Deployment

Sun Blade v100 with 2 internal IDE disks with software RAID1 running ACSLS 7.0 OS Solaris

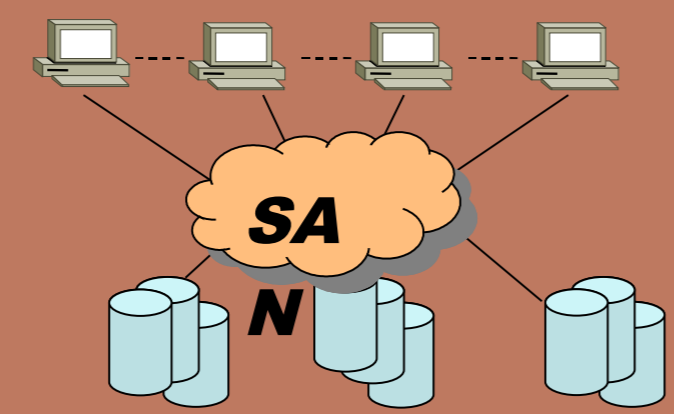


Core services are on machines with SCSI disks, hardware RAID1, redundant power supplies

Tape servers and disk servers have lower level hardware, like WNs

~ 40 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)  
16 tape drives, 3 Oracle databases  
LSF plug-in for scheduling  
SRM v2 (2 front-ends), SRM v1 (phasing out)  
15 tape servers



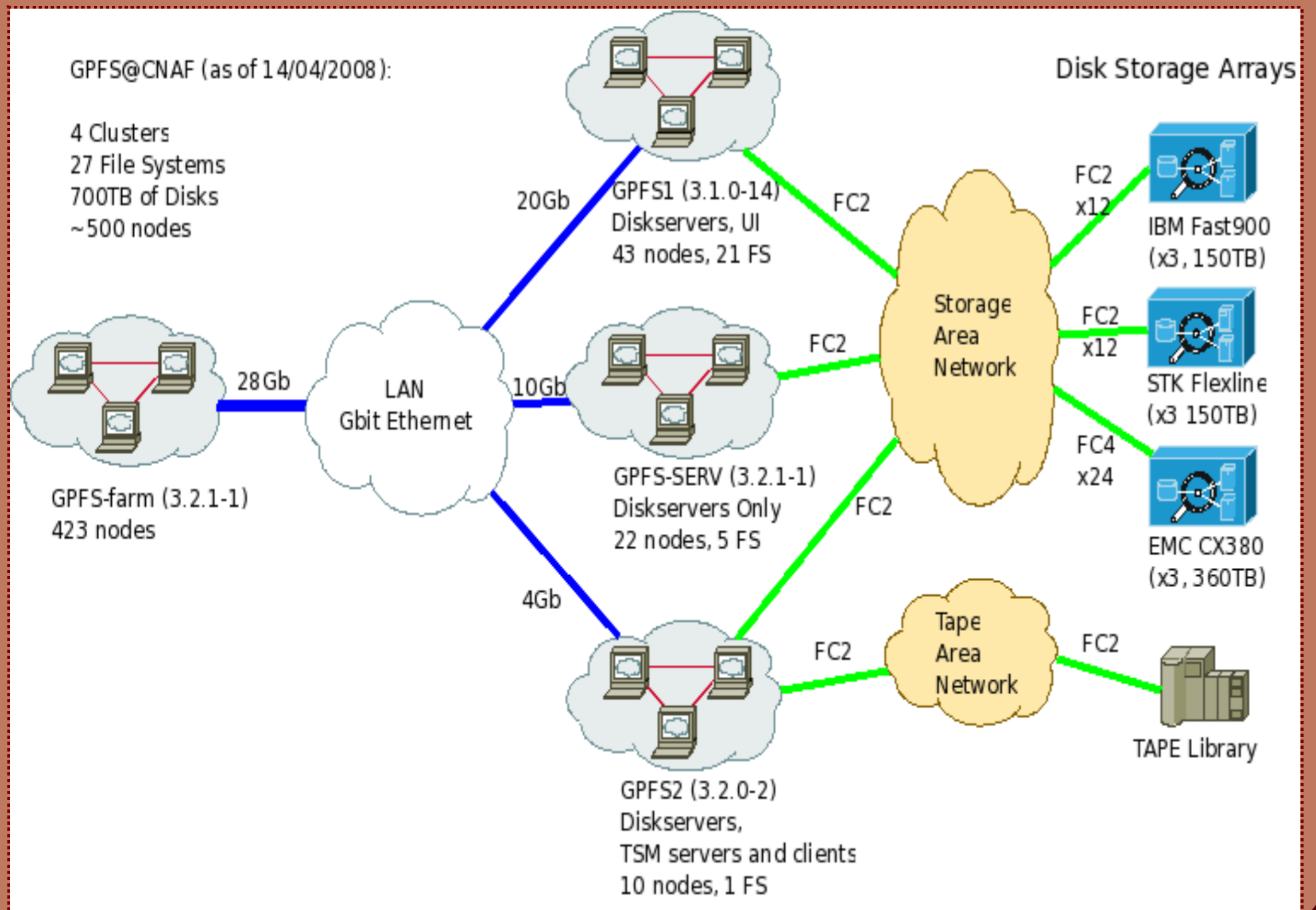
STK FlexLine 600, IBM FastT900

## GPFS Deployment

GPFS 3.2 is the IBM high-performance cluster file system. It greatly reduces administrative overhead and provides redundancy on the level of IO server failure.

After extensive tests in our SAN environment, GPFS in demonstrated robustness and high performances, in particular it showed better performance, as compared to CASTOR, dCache and Xrootd solutions.

Access to remote cluster file system by WNs has proven to be as efficient as the local one.



## StoRM

StoRM (STORage Resource Manager) is an implementation of Storage Resource Manager (SRM) standard interface version 2.2. It is designed to support guaranteed space reservation and direct access (native POSIX I/O call), as well as other standard Grid access protocols.

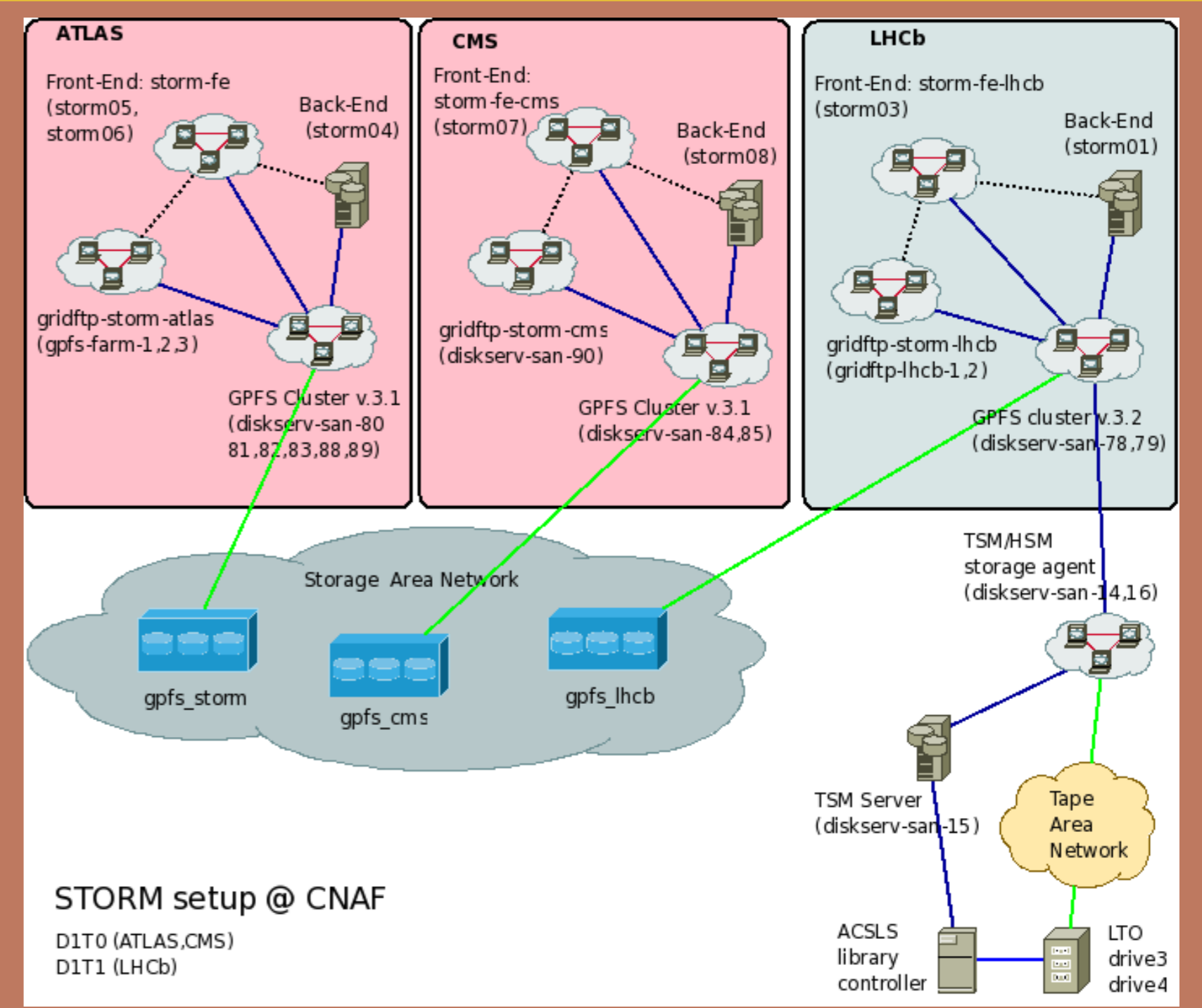
StoRM can take advantage of high performance storage systems based on cluster file system such as GPFS file system from IBM and Lustre from Sun Microsystems.

## D1T0 & D1T1@CNAF using StoRM/GPFS/TSM

Main idea is to combine new features of GPFS (v.3.2) and TSM (v.5.5) with SRM (StoRM), to provide transparent GRID-friendly HSM solution.

Information Lifecycle Management (ILM) used to order moving of data between disks and tapes

Interface between GPFS and TSM is on our shoulders



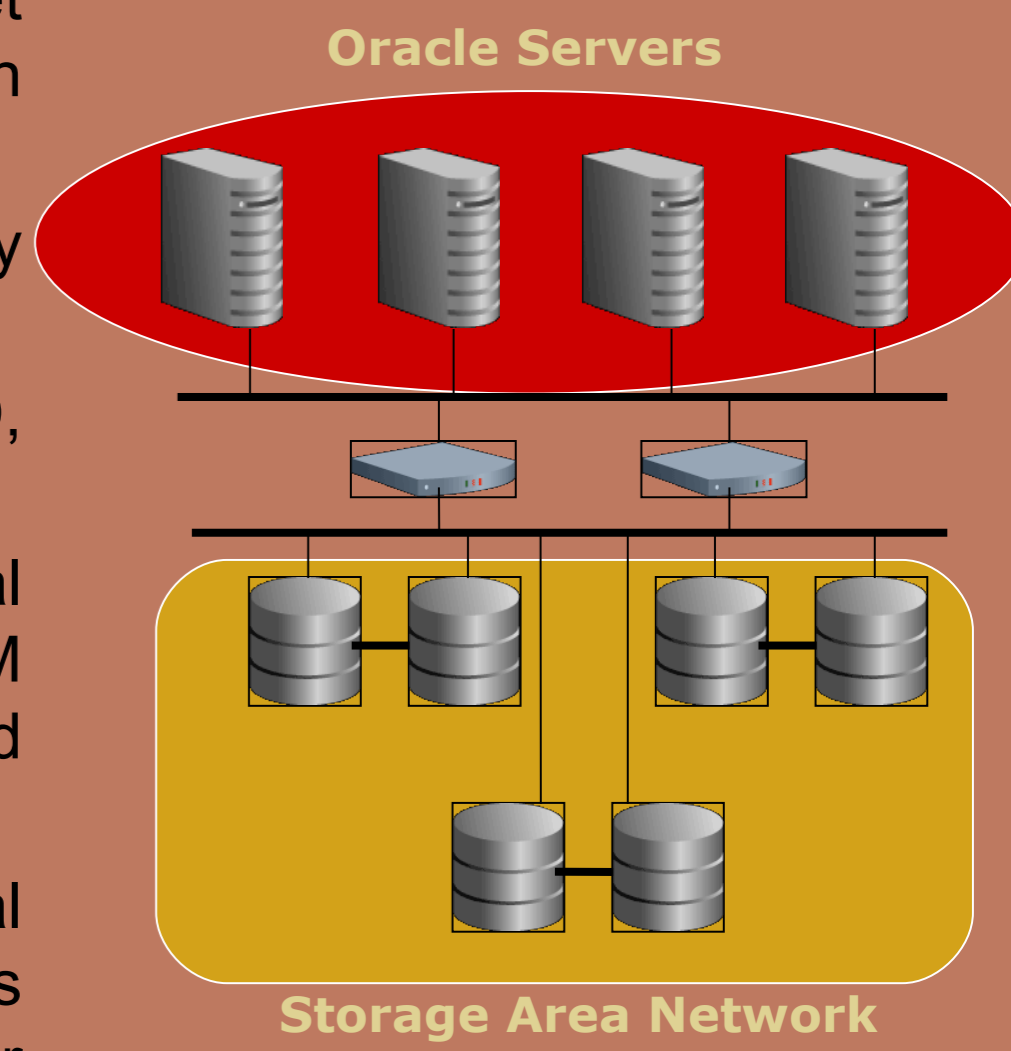
3 STORM instance, 3 major HEP experiments, 2 Storage classes, 12 servers, 200TB of disk space, 3 LTO2 tape drives

## ORACLE Infrastructure

Highly redundant architecture deployed in order to meet performance, scalability and high availability requirements.

Multiple level of high availability technologies adopted:

- H/W Storage level: RAID, Storage Area Network
- S/W Storage level: logical volume manager ASM implements striping and mirroring on Oracle blocks.
- Database level: Oracle Real Application Cluster a database is shared among multiple server implementing load balancing and failover policies.
- WAN level: database replication via Oracle Streams: data are shipped from CERN to Tier-1s databases.



Disaster recovery: backup via Recovery Manager (RMAN) integrated with TSM, 2-days retention on disk, 31-days retention on tape.