Braga - Portugal
24th - 28th May 2010

# PT Tier 2 Readiness

## IBERGRID Conference

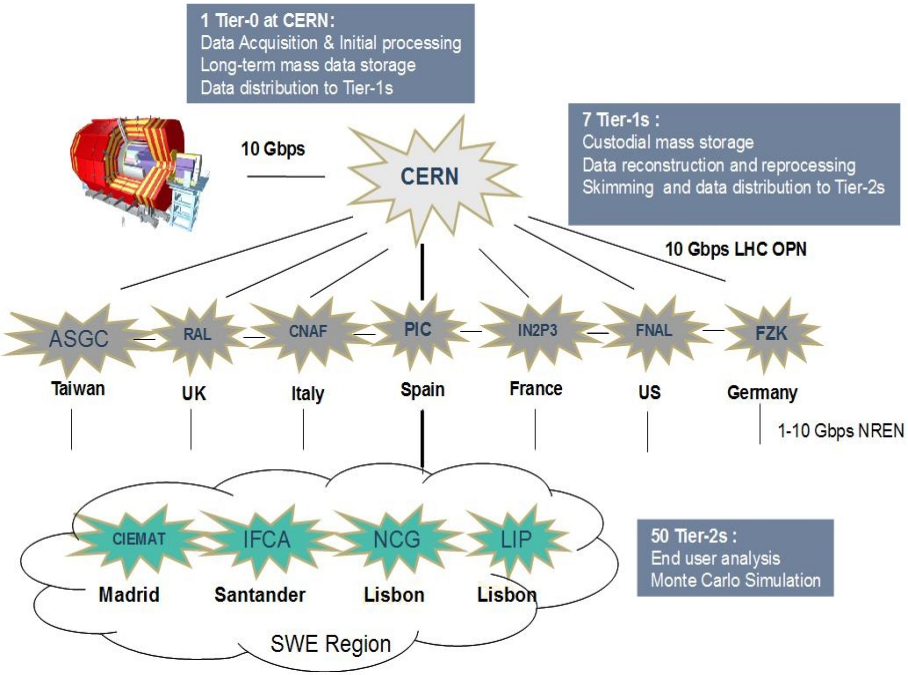
G.Barreira, G.Borges, M.David, N.Dias, H.Gomes, J.Gomes, J.Martins, M.Oliveira

# Introduction

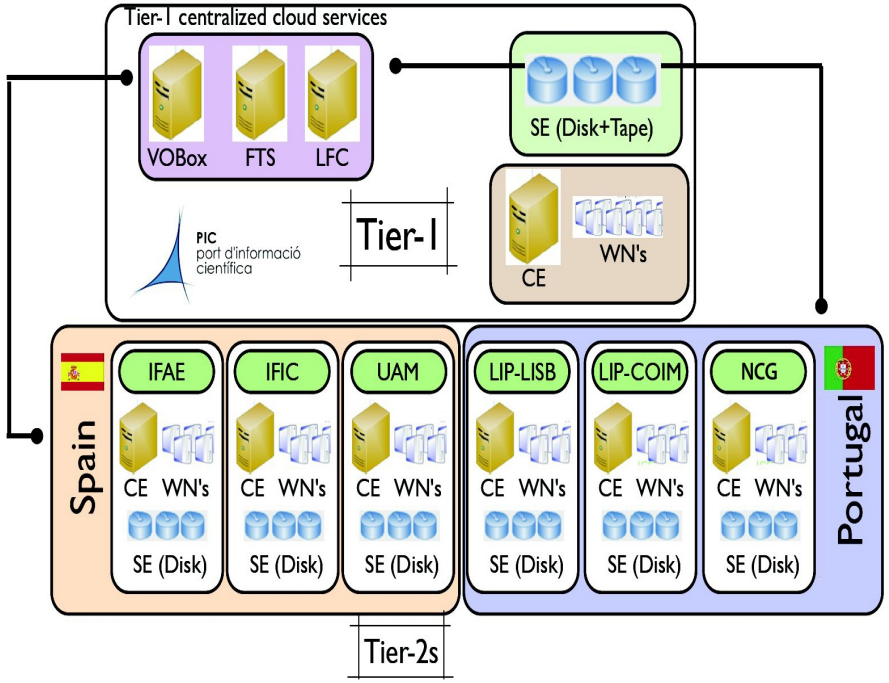
- **The Portuguese WLCG Tier-2 is an association of 3 different sites**
  - LIP-LISBON and LIP-COIMBRA
    - Laboratory for Instrumentation and Experimental Particle Physics
      - Two branches of the same institution
    - Researchers deeply involved in CMS and ATLAS detectors construction
    - The same degree of commitment is now addressed to preparing, building and maintaining the local LHC computing grid infrastructure
  - NCG-INGRID-PT
    - **INGRID main node for grid computing**
    - **Contribution of the Portuguese Grid Initiative for the fulfillment of the Portuguese responsibilities assumed for the WLCG MoU**
  - All 3 sites are operated by the same **LIP** team

# PT T2 ATLAS and CMS Topology

- CMS
  - LIP-LISBON
  - NCG-INGRID-PT




- ATLAS
  - LIP-LISBON
  - LIP-COIMBRA
  - NCG-INGRID-PT



# T2 Commitments

- **Computing and Storage pledges**
  - **CMS**
    - **200 TB / 3200 HEPSPEC06**
  - **ATLAS**
    - **231 TB / 3200 HEPSPEC06**
      - **ATLAS committed 31 TB from its Tier-3 to be used by the Tier-2**
  - **Presently well above the initial commitments**
    - **Dynamic resources which will be taken by PT NGI in a near future**

**Table 1.** ATLAS and CMS storage and computing pledges for the PT Tier-2

| Storage (TB) | LIP-Lisbon | LIP-Coimbra | NCG-INGRID-PT | Total |
|---|---|---|---|---|
| CMS T2 | 75. | - | 125. | 200. |
| ATLAS T2 | 67. | 67. | 97 | 231. |
| | | | | |
| CPU (HEPSPEC06 [8]) | LIP-Lisbon | LIP-Coimbra | NCG-INGRID-PT | Total |
| CMS T2 | 469 | - | 2731 | 3200 |
| ATLAS T2 | 950 | 950 | 1300 | 3200 |

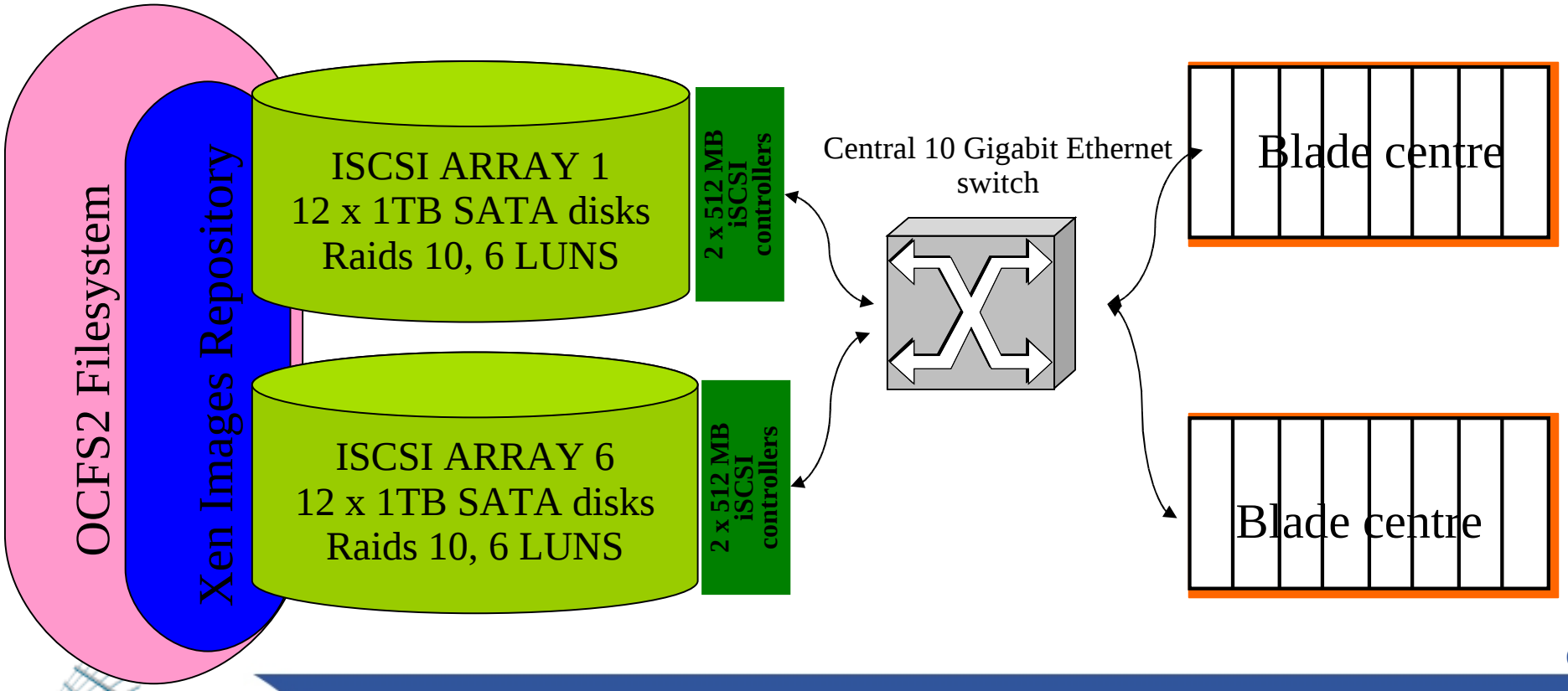
# Infrastructure: Virtualization (I)

- Solution based on Xen Virtual machines
  - Encapsulation
  - Multiple environments and flavors
  - Flexible framework for persistent VMs enabling resilient service provisioning
  - Testing Testbeds

- To host gLite and other basic services
  - DNS, Web Servers, Monitoring tools (Nagios, Ganglia, ...)
  - Local Grid Services
    - CEs, BDIIs, MONBOX, UIs ...
  - Core Grid Services
    - PX Servers, VOMS Servers, LFCs, WMSs, TOP-BDIIs, ...

# Infrastructure: Virtualization (II)

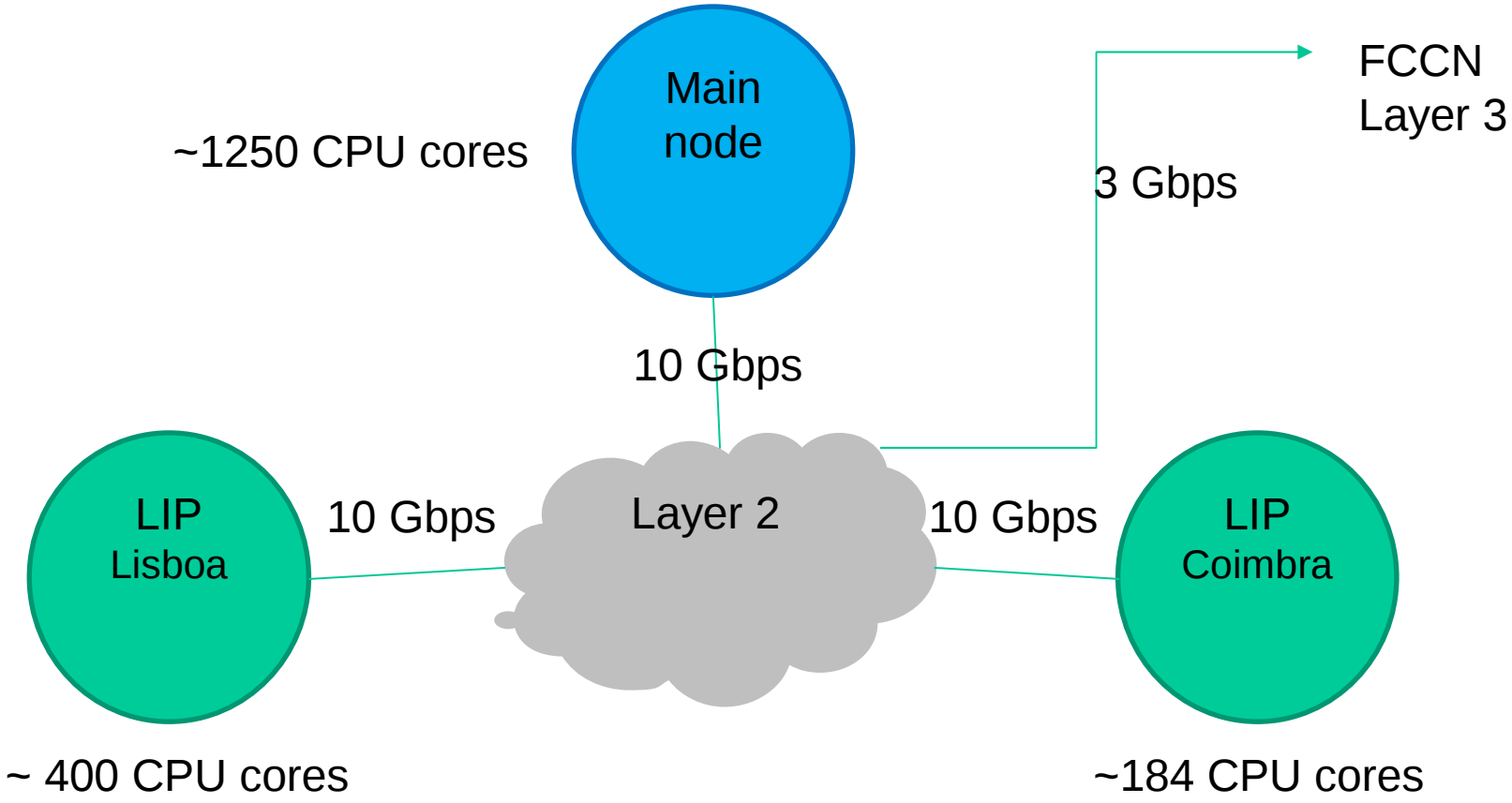
- Redundant services distributed between 2 blade centres
  - Xen images available Storage accessible via Internet SCSI (iSCSI)
  - Controlled by OCFS2 shared cluster file system...



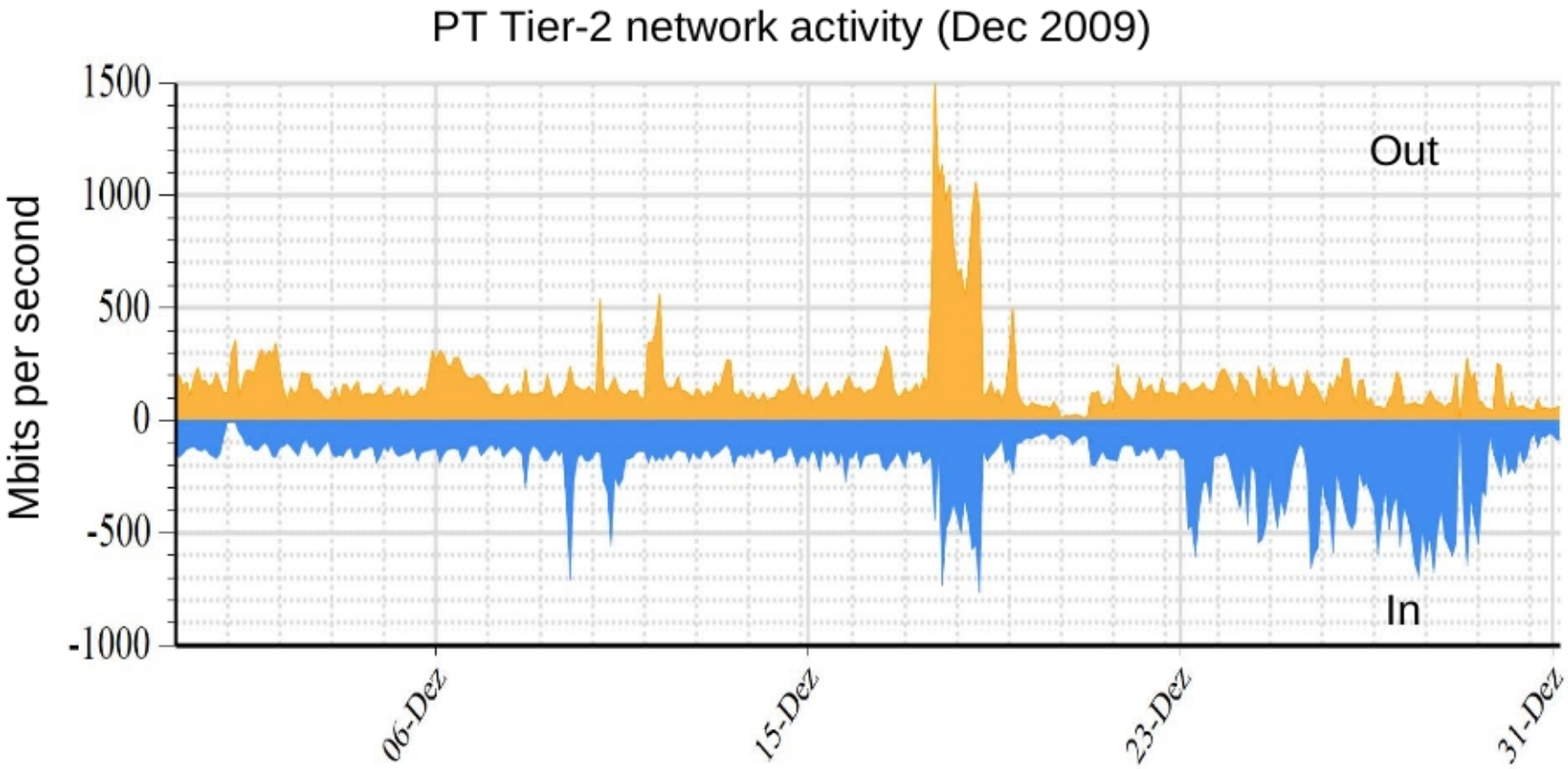
# Infrastructure: Networking (I)

- The core network equipment at all sites has been replaced by non-blocking layer 2 / layer 3 switches
  - All ports can operate simultaneously at higher speed without packet loss
- The local networks have been divided in different VLANS
  - Separates local traffic from grid traffic
  - Scalability, security, easier management
- WLAN links
  - L2 connectivity between the 3 resource centers @ 10 Gbps
  - Geant connectivity 3 Gbps for the whole cloud

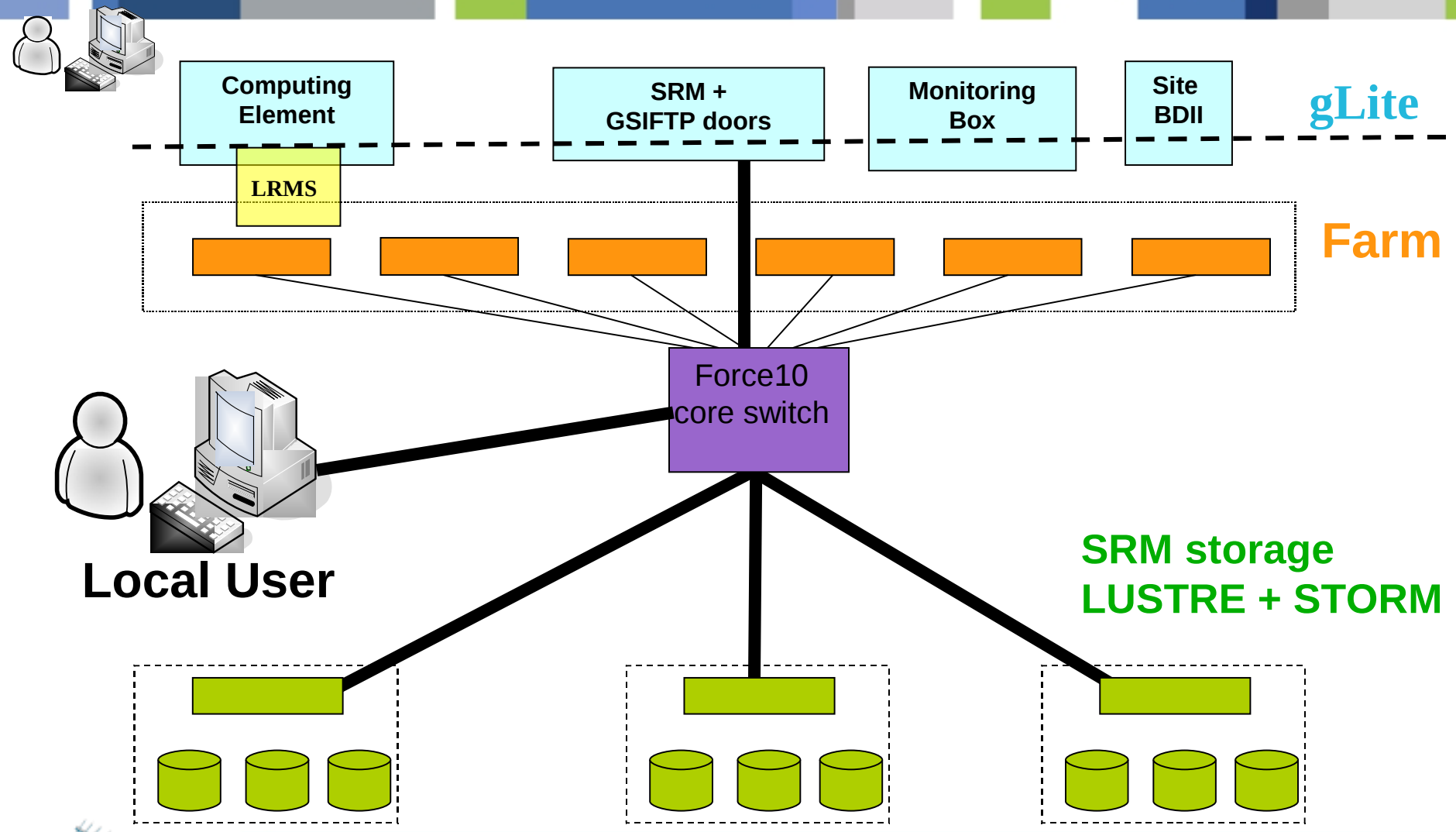
# Infrastructure: Networking (II)



# Infrastructure: Networking (III)

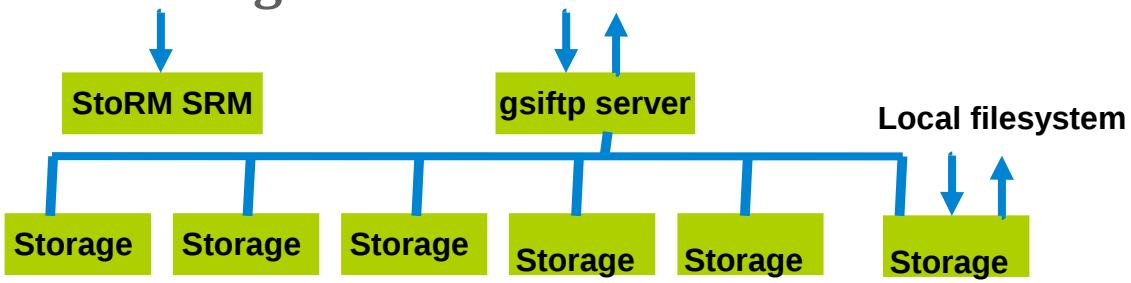


# Tier-2 resources topology

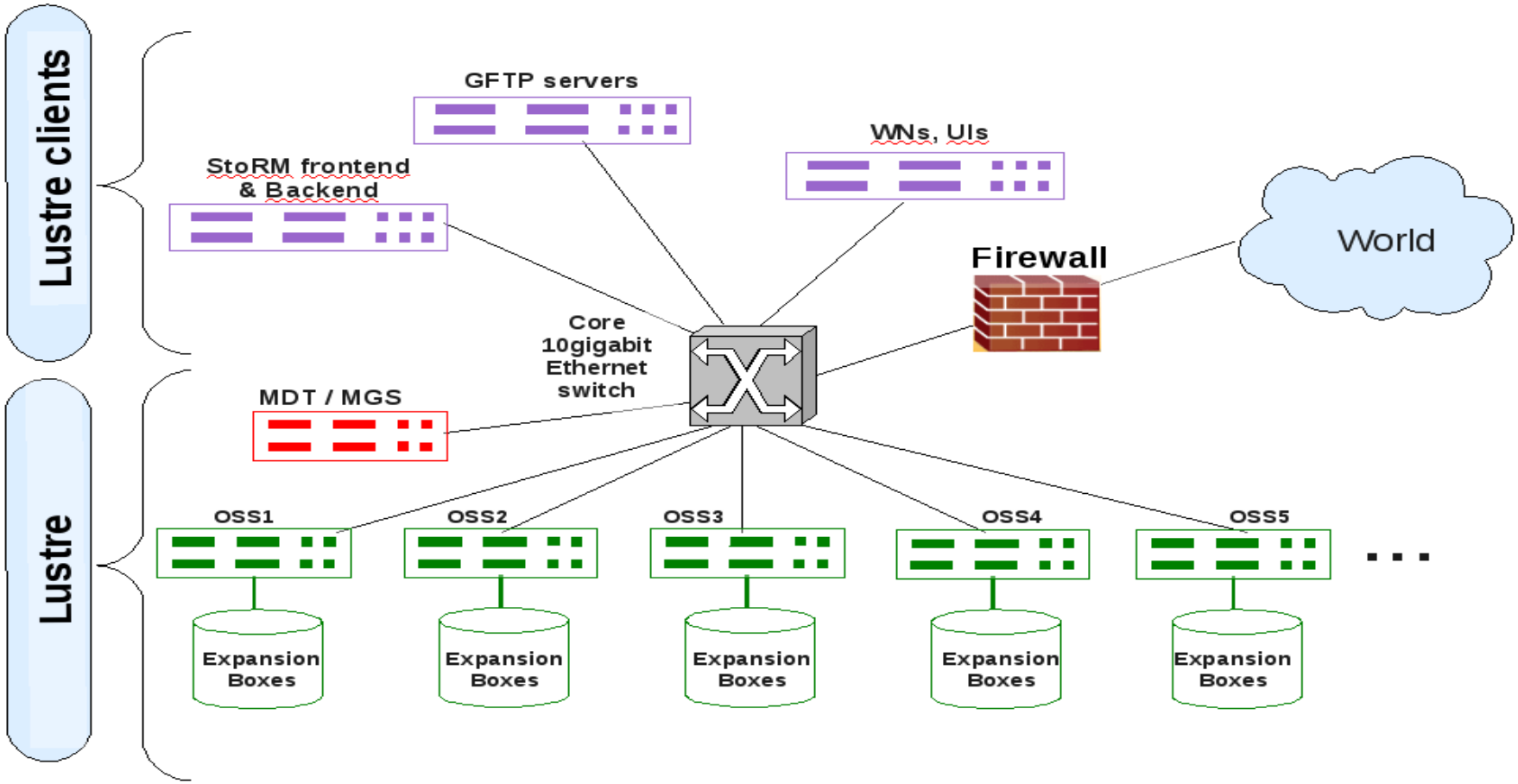


# Infrastructure: Storage (I)

- StoRM using Lustre as the underlying filesystem
- Lustre is a High Performance filesystem mostly used in HPC
  - Creation of POSIX filesystems across multiple servers
    - A file can stored only in one server or stripped across several
- StoRM decouples the SRM services from the filesystem itself
  - Lustre filesystem can be mounted in any (non-grid) Linux box, as long as the appropriate kernel modules are loaded
  - The filesystem “IS” the data catalogue
    - Avoids the use of DBs



# Infrastructure: Storage (II)



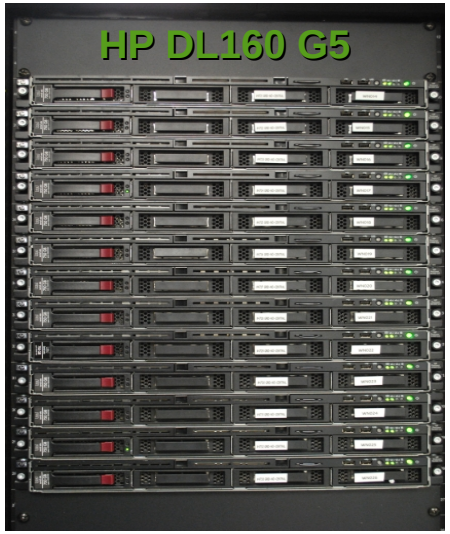
# Infrastructure: Computing (I)

- One single FARM at each site
  - Better optimization and management
  - Tried and used several LRMS
    - Torque+Maui (LIP-Coimbra), SGE (LIP-Coimbra / NCG)
  - Heterogeneous (blade, servers,...)
    - Hardware (AMD, Intel, several generations, ...)
    - Administrative (different groups and purposes, grid vs non-grid)

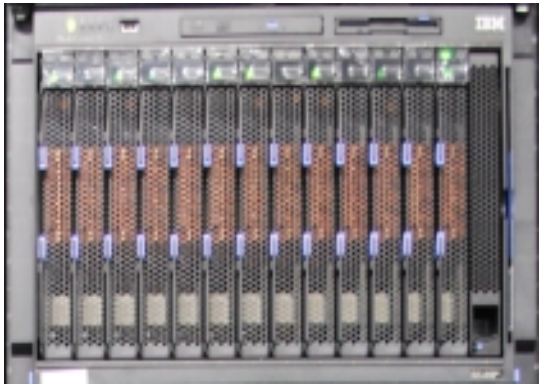
- PT WLC Resources
  - Dual CPU machines with 4 cores each, 3 GB/core, Running SL5
  - @ NCG-INGRID-PT
    - IBM and HP blade solutions hosting from 12 to 14 blades
  - @ LIP-Lisbon / @ LIP-Coimbra
    - Mostly discrete but powerful HP Linux boxes

# Infrastructure: Computing (II)

HP DL160 G5



IBM LS22



HP BL160 G6

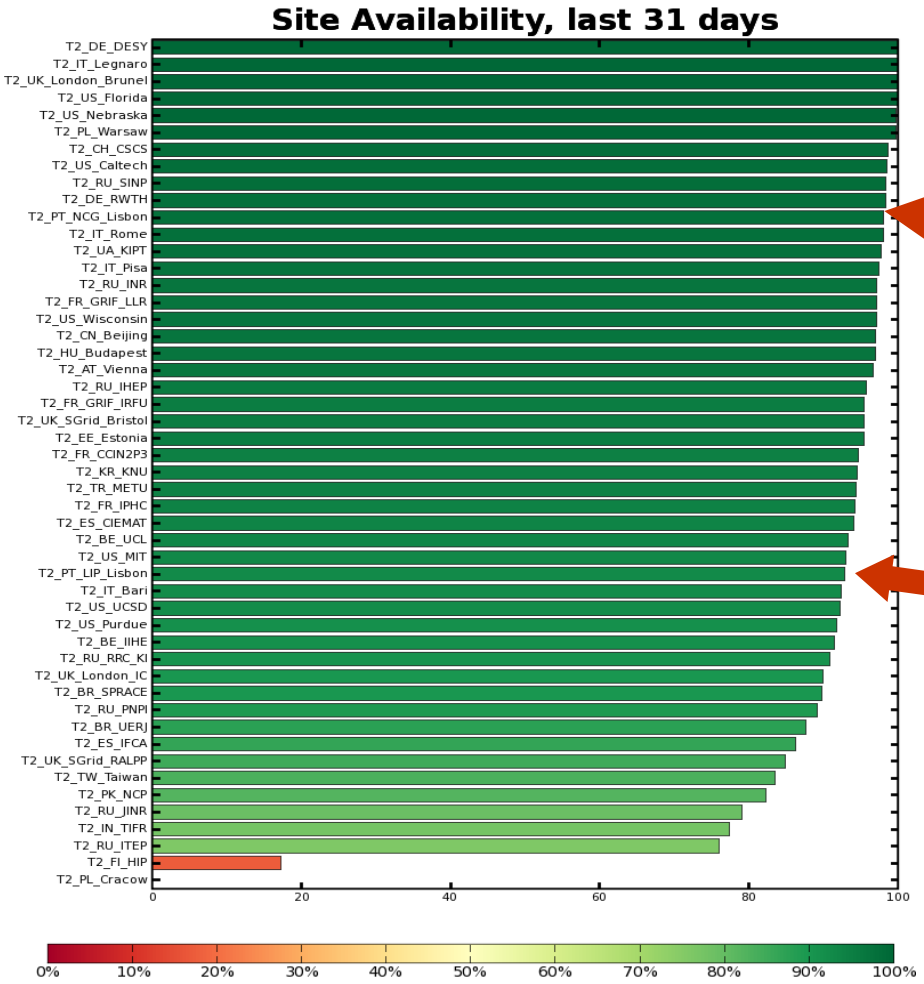
**Table 2.** Benchmark results for the PT-Tier-2 resources. The reported values represent the average results of several consecutive measurements, and their errors represent the average dispersion of those measurements.

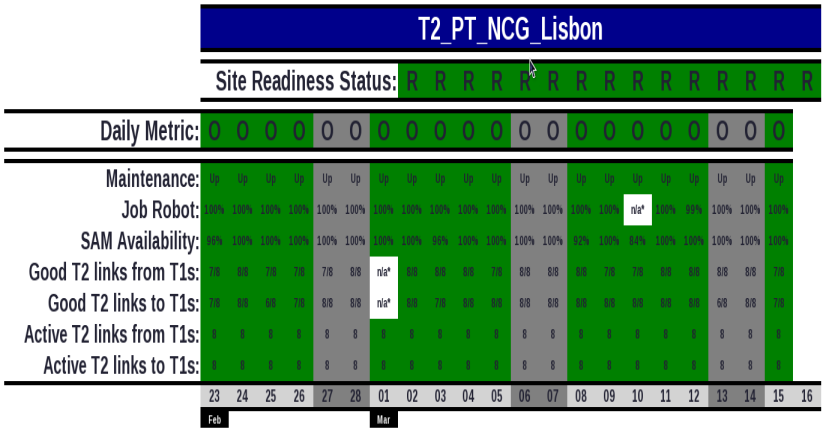
| Model | Processor | OS | Arch | HEPSPEC06/Core | SI00/HEPSPEC06 |
|---|---|---|---|---|---|
| IBM LS22 | AMD Opteron 2356 | SL5 | X86_64 | 6.83±0.01 | 208 |
| IBM HS21 | Intel Xeon E5420 | SL5 | X86_64 | 8.20±0.03 | 244 |
| HP DL160 G6 | Intel Xeon E5540 | SL5 | X86_64 | 13.14±0.02 | 186 |
| HP BL460c G6 | Intel Xeon X5550 | SL5 | X86_64 | 14.15±0.04 | 189 |
| HP DL160 G5 | Intel Xeon E5472 | SL5 | X86_64 | 10.21±0.02 | 227 |
| HP DL160 G5 | Intel Xeon E5472 | SL4 | i386 | 9.49±0.04 | 248 |

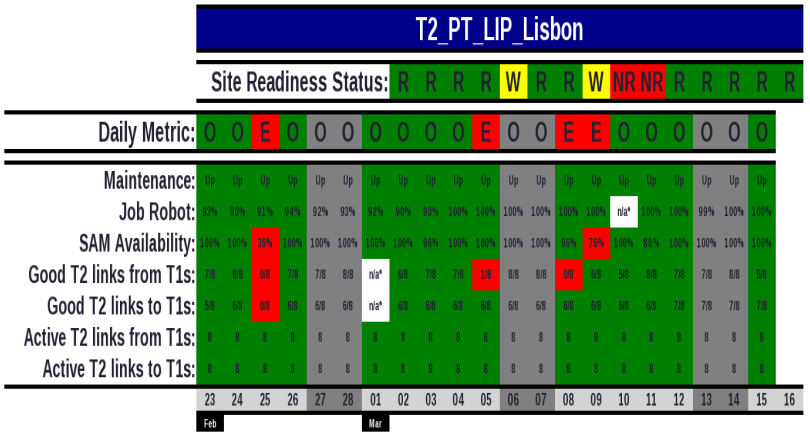
# Performance (I)

- Long inactivity periods are bad for distributed computing infrastructures
- Both ATLAS and CMS exercise their infrastructures
  - Check the readiness status of the participating sites
  - Spot operational problems
  - Maintaining the data flows active and prepared for massive data distribution
- Evaluation techniques to check the fulfillment of WLCG MoU
  - Sites must maintain a sustainable success rate

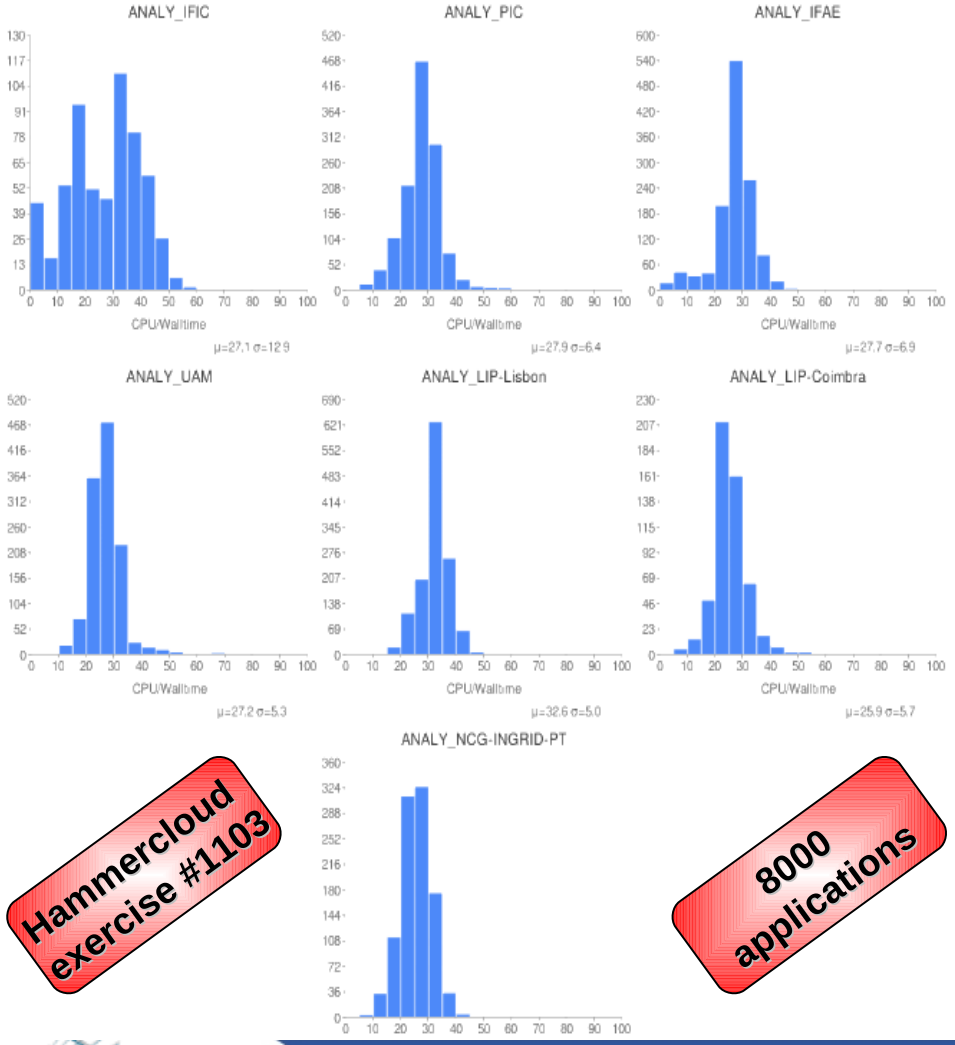
# CMS PT Tier-2 Performance



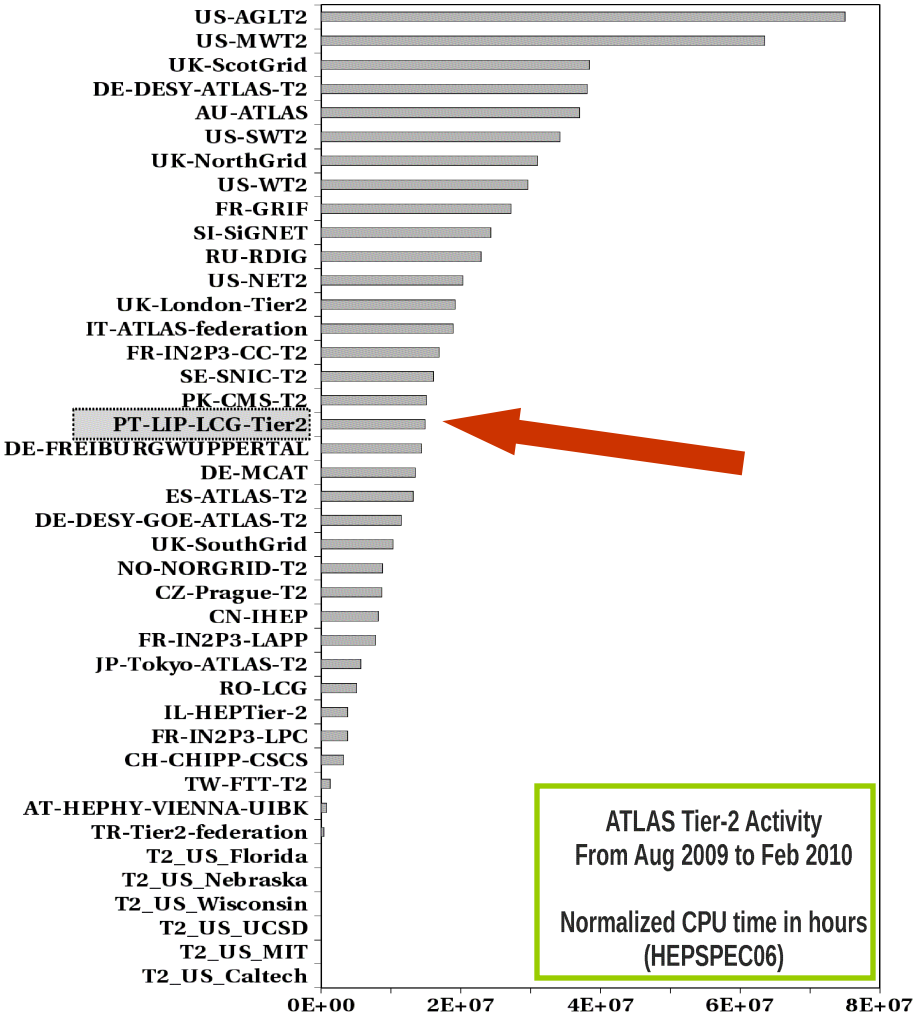
# ATLAS PT Tier-2 Performance

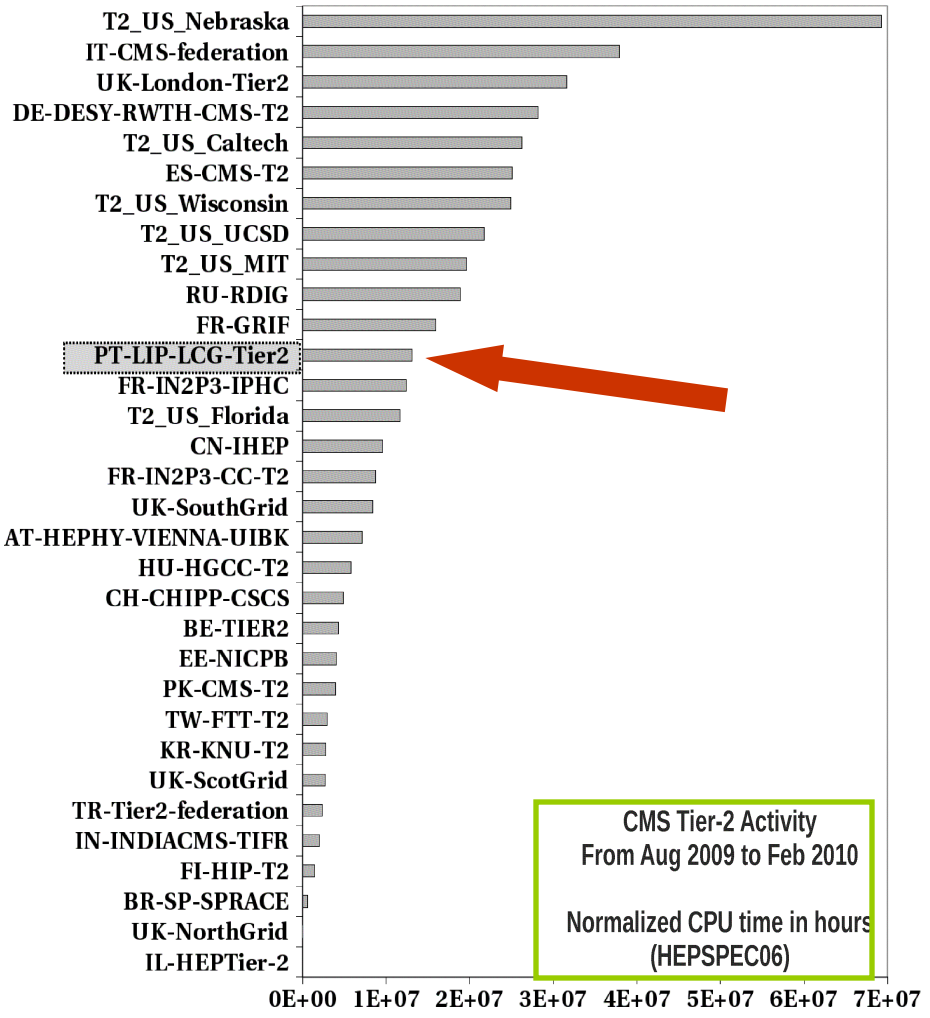



- **High success rate**
  - **< 1% for NCG-INGRID-PT and LIP-Lisbon**
  - **Higher than average for LIP-Coimbra but not site fault**
- **Event Rate**
  - **0.5 Hz, 0.6 Hz and 0.6 Hz for NCG-INGRID-PT, LIP-Lisbon and LIP-Coimbra**
  - **0.5 Hz as average for all Iberian ATLAS sites**
- **Efficiency**
  - **26.0%, 32.6% and 25.9% for NCG-INGRID-PT, LIP-Lisbon and LIP-Coimbra**
  - **28.0% as average for all Iberian ATLAS sites**

LIP Computing

# WLCG PT Tier-2 Performance



# Operational problems (I)

- Communication issues
  - Not easy to pass information to the experiment people
  - Opposite flow from experiments to Tier-2 is always delivered with extreme urgency and deadlines that have to be adhered to
- Information restriction
  - CMS restricts technical information to VO members
- Experiments tools and framework
  - Debug information not always available
  - When available, the debug information isn't clear enough to someone that is not familiar with the application or is not a VO member

# Operational problems (II)

- Hardcoded and static framework
  - New resources are not automatically recognized
    - The (ATLAS) experiment framework do not use the IS
  - Not flexible software conceptualization and architecture
    - We found (the hard way) that a physical NFS mount point was hard-coded in hundreds of places for ATLAS software

- Contradictory demands
  - Immediate migration to SL5
  - Local users still want SL4
  - Manage two sets of resources
    - Load increase

# LIP Computing Team