

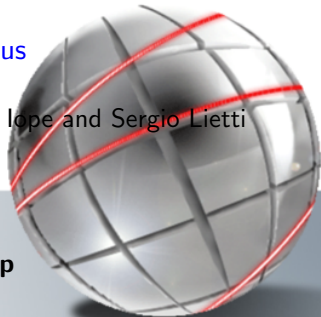
# *São Paulo Regional Analysis Center*

<http://www.sprace.org.br>

## T2\_BR\_SPRACE Site Status

Marco Dias,

Allan Szu, Gabriel Winckler, Jadir Silva, Rogerio Iope and Sergio Lietti



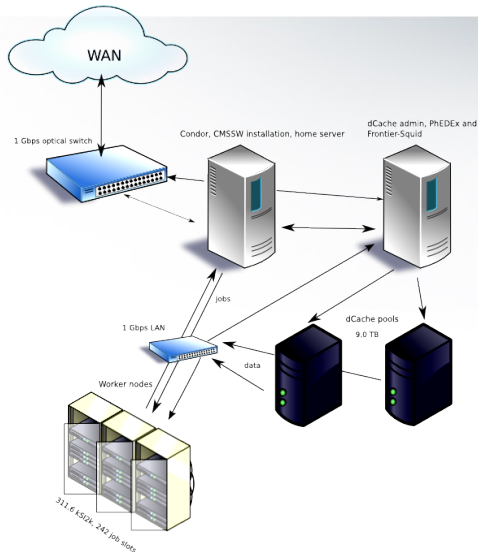
## US CMS Tier-2 Workshop

Fermilab-Illinois

Mar 08, 2010

# Farm Overview

- SPRACE facilities are fully dedicated to CMS and DZero collaborations:



# CPU, Hardware and Data Storage



- 82 worker-nodes, 222 batch slots ( 400 kSI2K), but only 178 of those batch slots are available for CMS (64 bit). Currently our amount of resource deployment, only for CMS, was calculated in 1256,32 HS06.

	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2010</b>
<b>CPUs (cores)</b>	50	116	222	350

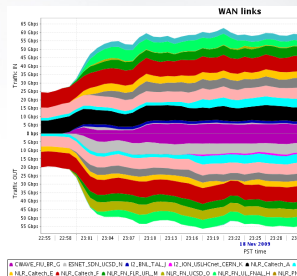
- One dCache server, namely a Intel Xeon 5130, 2.00GHz, 4 cores with 4GB mem, running almost all necessary services.
- Two dCache pools, in a RAID 5 arrangement. At time being 6.91 TB occupied with CMS data; 1.56 TB still available for hosting.
- Links commissioned with all US CMS tier-2, T2\_BR\_UERJ and T3\_MX\_Cienvstav.

# Location and Infrastructure



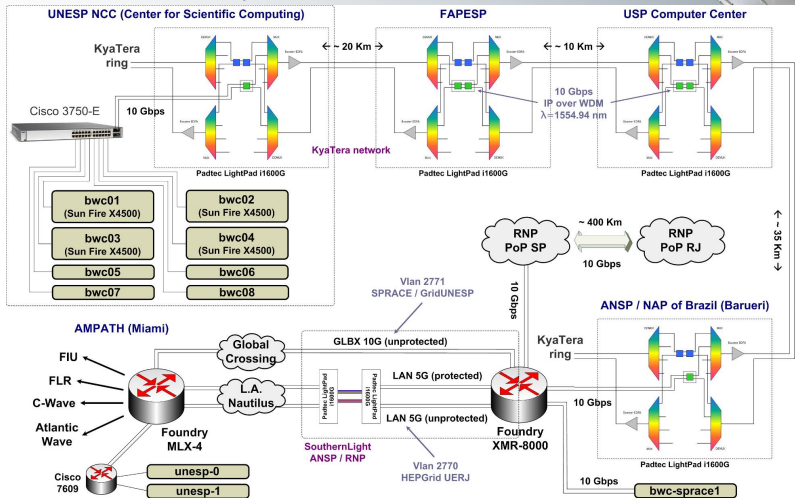
- From Feb 2004 to Aug 2009
  - ▶ Located in a single room at USP Physics Institute building.
  - ▶ Several problems with air-conditioning and power outages.
  - ▶ No room to grow, network link limited to 1 Gbps.
- Now, from September 2009
  - ▶ Located inside GridUnesp datacenter, on UNESP campus.
  - ▶ Plenty of space and appropriate cooling.
  - ▶ Generator/UPS backup.

- ▶ This new location and infrastructure was used to join the C-Wave team on Bandwidth Challenge Award, during the SuperComputing 2009, achieving a transmission rate of 6 + 6 Gbps (both directions, disk2disk)



# Location and Infrastructure (cont)

- Also remarkable, we have established a new record on data transmission between the two hemispheres: 8,26 + 8,26 Gbps (both directions, mem2mem).



# Location and Infrastructure (cont)

- SPRACE new location



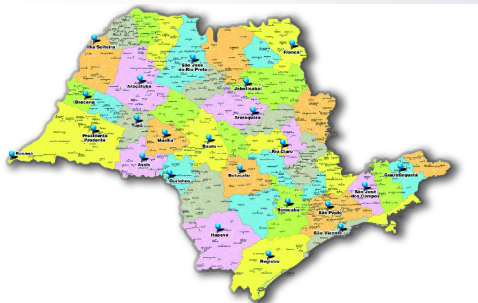
# Software and Operations: Developments over the year (2009)



- OSG 1.2.4
  - ▶ Using GUMS, moved to a separated node from our gatekeeper.
- Condor 7.4.0
  - ▶ We are using a distinct installation from OSG package. Using a heterogeneous setup (32 and 64 bit), to supply the non-CMS demand. 32 bit WNs will be retired in a very near future.
- dCache
  - ▶ Upgraded to the Golden Release edition.
  - ▶ A failure in our RAID setup reduced our storage capacity.
  - ▶ Migration to Chimera without any problems so far.
- OS and Cluster Administration
  - ▶ A new server dedicated to authentication and monitoring issues was deployed.
  - ▶ All worker nodes upgraded to SL 5.3.

# Use of GridUNESP for CMS

- UNESP (São Paulo State University) is a set of 24 campi spread over the state of São Paulo. GridUNESP is the first shared university campus cyberinfrastructure being built in Brazil.



- We are planning to use its resource as part of SPRACE, in the sense that CMS VO jobs will be accounted as part of SPRACE for WLCG.
- A Certification Authority is being deployed to supply the demand coming from local reserachers and regional research groups.



## Use of GridUNESP for CMS (cont)



- **Processing power:** 8 clusters (Sun/Intel) 9,2 MSI2k; theoretical peak FLOPS performance, summing up the 8 distributed clusters is 33,3 TFlops.
- **Central Cluster:**
  - ▶ 256 worker nodes: dual quad-core Intel Xeon processors (E5440 Harpertwon, 2.83 GHz, 12MB L2 Cache), corresponding to 2048 batch slots, 16 GB mem.
  - ▶ 4 head nodes quad quad-core Intel Xeon processors (E5440 Harpertwon, 2.83 GHz, 12MB L2 Cache), 32 GB mem.
- **Secondary clusters:**
  - ▶ 16 worker nodes: dual quad-core Intel Xeon processors (E5440 Harpertwon, 2.83 GHz, 12MB L2 Cache), corresponding to 128 batch slots, 16 GB mem.
  - ▶ 2 head nodes dual quad-core Intel Xeon processors (E5440 Harpertwon, 2.83 GHz, 12MB L2 Cache), 16 GB mem.

## Use of GridUNESP for CMS (cont)



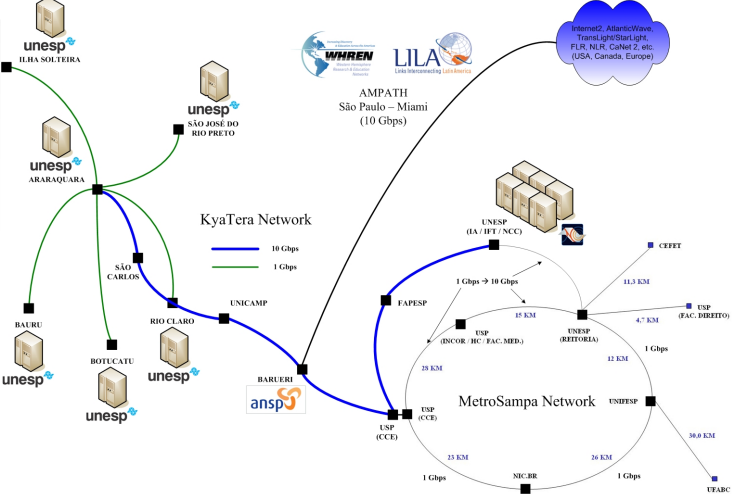
- **Network infrastructure:** 2 x 10 Gbps international connection links between So Paulo and Miami, implemented and jointly operated by ANSP (the state network in So Paulo)-RNP. Currently, between 4 and 5 Gbps of commodity traffic uses these links. 1 Gbps is reserved for routed IP traffic as part of the RedCLARA network, and the remaining capacity is now available for international collaboration.
- **Storage:** Divided between a central and seven secondary clusters:
  - ▶ 4 hybrid data servers (Sun Fire X4500), 24 TB each SAN fibre-channel
  - ▶ 4 Gbps (Sun StorageTek 6140), 36 TB, Interconnected by an fibre-channel 4 Gbps storage network.
  - ▶ SAN fibre-channel 4 Gbps (Sun StorageTek 6140), 12 TB, each one.

	<b>Central Cluster</b>	<b>Secondary Cluster</b>	<b>Total</b>
<b>CPUs (cores)</b>	2,048	896	<b>2,944</b>
<b>KSI2K</b>	6,100	2,600	<b>8,700</b>
<b>Storage (TB)</b>	240	98	<b>338</b>

# Use of GridUNESP for CMS (cont)



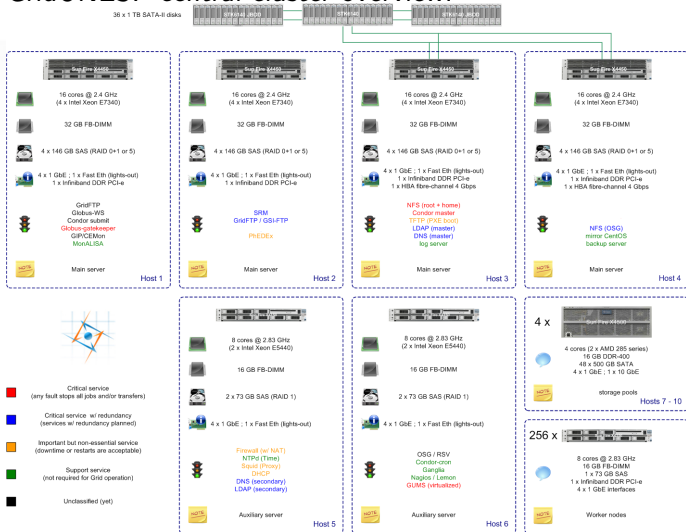
- Network infrastructure diagram:



# Use of GridUNESP for CMS (cont)



## GridUNESP central cluster overview:



## Upgrade Plans for SPRACE



- RFP submitted last week to leading vendors (Dell, HP, IBM, SGI, Sun) for acquisition of new servers with Intel Xeon 5500 series (Nehalem), possibly Intel Xeon 5500 series (Nehalem), 8 cores for server:
  - ▶ Storage: Chimera and other services (8TB RAID arrangement with 4 disks)
  - ▶ Storage: SRM (4 TB RAID arrangement with 4 disks)
  - ▶ Compute: Condor (500GB)
  - ▶ Compute: Gatekeeper (500GB)
  - ▶ Access and local analysis ( 1TB)
- Ordering 16 worker nodes within 2 months.
  - ▶ Intel Xeon series 5500 (Nehalem) with 8 cores for server, 3 GB RAM memory per core, 500GB for disks.
  - ▶ It will increase our processing power in 16 new servers  $\times$  4 cores  $\times$  68 HS06 = 4352 HS06 [estimated from *INFN* homepage 2  $\times$  Hapertown (Intel Xeon X5540, 2.83 GHz) specifications] .
- 3 Sun Fire X4540 are being installed, approx. more 144 TB for dCache (“raw” ). Another two units will be purchased in the middle of this year.

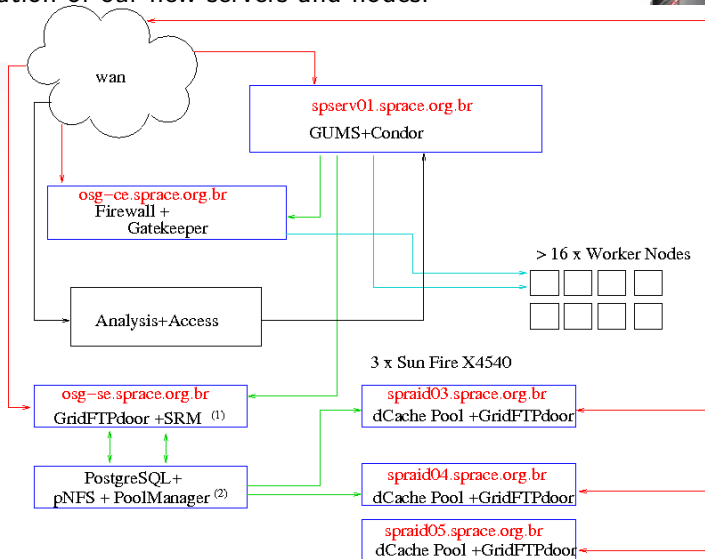
# Upgrade Plans for SPRACE (cont)



- Network Investments
  - ▶ Data center border router: Cisco 6506E with 48 x 1 GbE and 8 x 10GbE ports
  - ▶ data center access switch: Cisco Nexus 5010 with 20 x 10 GbE ports
  - ▶ fully dedicated 10G WDM lambda from the data center to the 10G international links
- Presently we don't have plans to change from dCache to another SE technology (or even start the deployment of another storage element), but we are looking carefully what is going on (working or not) in GridUNESP storage element (currently *Lustre+Bestman*).

# Upgrade Plans for SPRACE (cont)

- Utilization of our new servers and nodes:



# Plans for this year



- Storage
  - ▶ Deployment of our new server.
  - ▶ Several new dCache pools.
- OSG/dCache/Condor/PhEDEx
  - ▶ Version upgrades.
  - ▶ Deployment of PhEDEx and Frontier-Squid servers for GridUNESP site.
  - ▶ Data transfer debugging.
- Administration
  - ▶ New sysadmins will be hired.
  - ▶ Atracting some grad. students for GridUNESP project.
  - ▶ Challenging to keep continuity.
- More interaction with Physics Groups
  - ▶ Nowadays our contact with physics groups with on CMS isn't so constant as we wish.
  - ▶ We intend to associate to Exotica group, and allocate some resources to Heavy Ion group as well.
  - ▶ With more space available in our storage element, our farm will probably be more attractive for users.
  - ▶ Interaction with other analysis groups will probably be more frequent as well