Ibergrid 2010. Braga (Portugal), May 24-27 2010.
Iberian Grid Infrastructure Conference.

Operational Experience Running the CIEMAT Grid Site

Nicanor Colino, Antonio Delgado Peris, José M. Hernández, Francisco Javier Rodríguez Calonge, Juan José Rodríguez Vázquez

Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain.

Agenda

- Introduction
- Activities and results
- Logical site setup
- Physical site setup and hardware studies
- Best practices
- Conclusions

The CIEMAT grid site

- The CIEMAT-LCG2 grid site
  - At the CIEMAT institute, High Energy Physics department
  - There are other (younger) grid resource centers at CIEMAT
- Continued participation in WLCG, EGEE, NGI (EGI)
- Support of several VOs
  - Biomed, Calice, CMS, CTA, Fusion, Magic, NGI's
  - CMS is the most important one, reflecting local community
    - CIEMAT is part of a federated Tier-2 with IFCA (Cantabria)
- Consistently satisfied VOs quality requirements
  - One of the top CMS sites in terms of availability, reliability and service delivered to the experiment

Running a production grid site

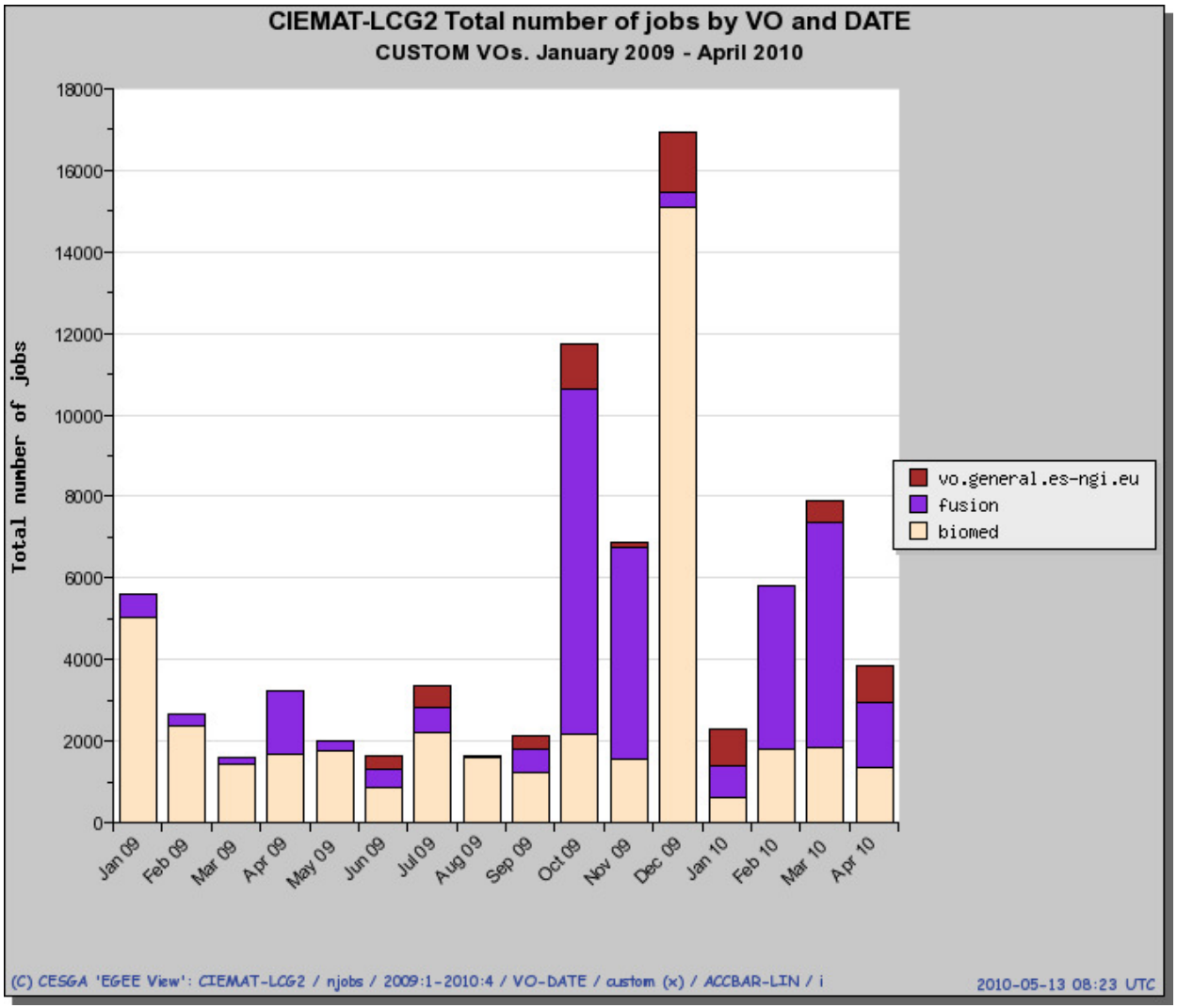
- Challenge: keep on offering same service quality
  - Increasingly demanding VO requirements
    - Failure tolerance, complexity, volume
  - Limited (not increasing) manpower
- Evolution from *early adopter* to *production center*
  - Deeper understanding of the technology
    - Configuration and deployment of grid services
    - Selection and tuning of resources and network settings
  - Improved internal operational procedures
    - More efficient administrative efforts
- This article describes the (up to now) outcome of this evolution for the CIEMAT case

Agenda

- Introduction
- Activities and results
- Logical site setup
- Physical site setup and hardware studies
- Best practices
- Conclusions

Continued VO activity

- Non-CMS jobs running continuously at the site

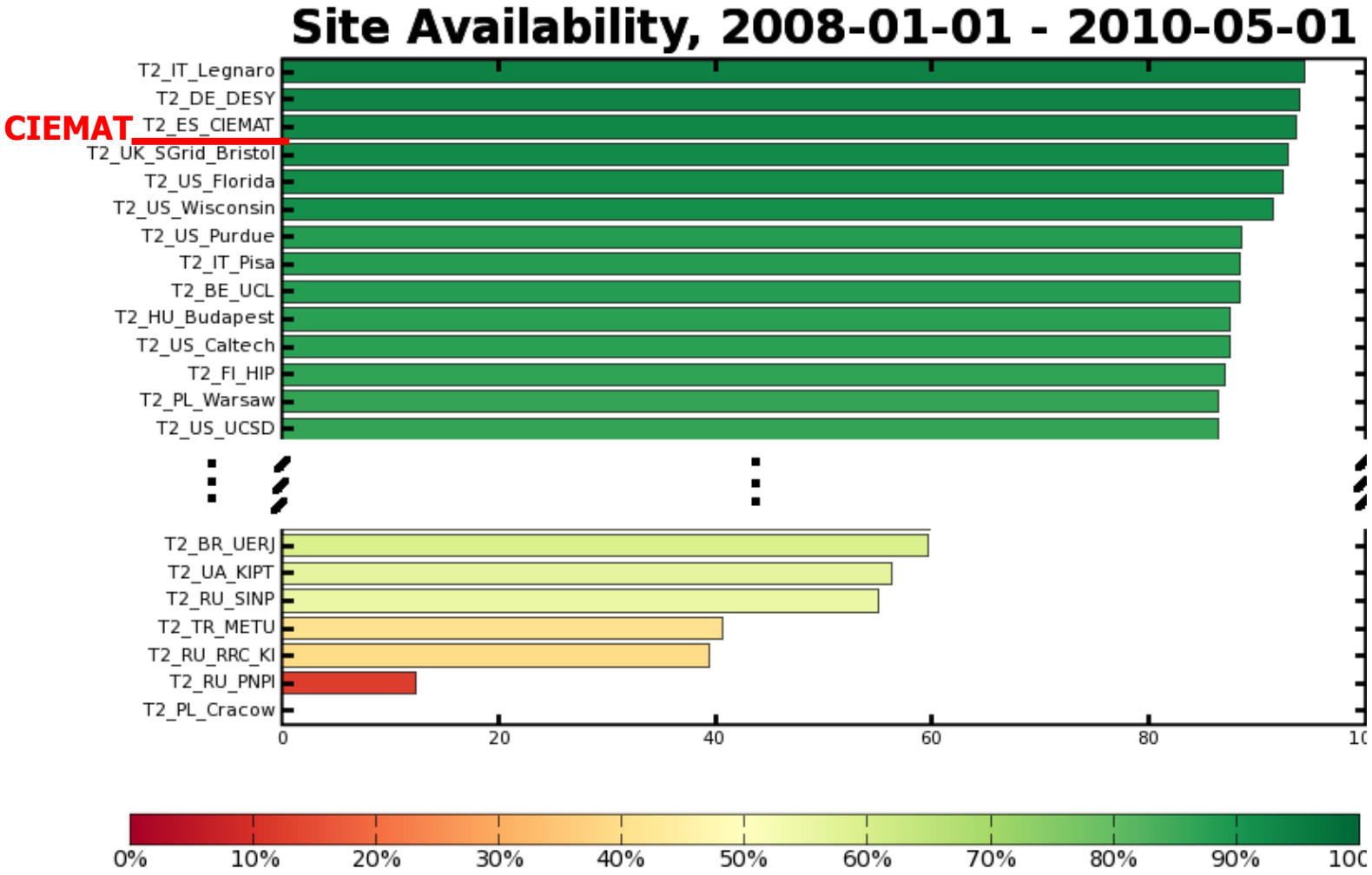


Availability and reliability

- CIEMAT site (ops VO)
  - Info from: http://gvdev.cern.ch/GVPC/Excel/
  - From January 2009 to April 2010
    - Availability: 95%
    - Reliability: 97%
- CMS VO tests
  - Info from: http://ardadashboard.cern.ch/cms/
  - From January 2008 to April 2010 (28 months)
    - Availbility: ~94%
    - Reliability: ~96 %
  - Third best CMS Tier-2 in that period

CMS T2s availability ranking

- Third site in T2s ranking



Computing capacity provided

- As a CMS Tier-2, CIEMAT must contribute
  - Processing capacity: central production, analysis jobs
  - Expectations are to provide 2.5% of CMS needs
- During 2009, CIEMAT has run 3,600 jobs/day
  - 1.3 million jobs (19.5 million HEPSPEC06 hours)
  - 700,000 production jobs (8% of CMS)
  - 600,000 analysis jobs (5% of CMS)
- Higher success rates than CMS averages

Site	All jobs	Monitoring	Production	Analysis
All T2s	76% (9%)	85% (6%)	83% (8%)	63% (11%)
CIEMAT	84% (4%)	97% (1%)	90% (3%)	67% (8%)

Data storage and transfer

- CIEMAT storage can hold up to 250 TBs of data
  - Locally produced data (prior to T1 upload)
  - External data for local analysis (comes in bursts)
- Transfer rates
  - Download rate
    - >100 MB/s continuously (10 TB/day) from T1 (PIC)
    - ~80 MB/s avg. from all T1s (avg. for all T2s is 73 MB/s)
  - Upload rate
    - >50 MB/s (CMS requirement is 10 MB/s)
    - Foreseen network equipment upgrade and topology reconfiguration will make much higher rates possible
  - During 2009, received 300 TB of data and exported 160 TB

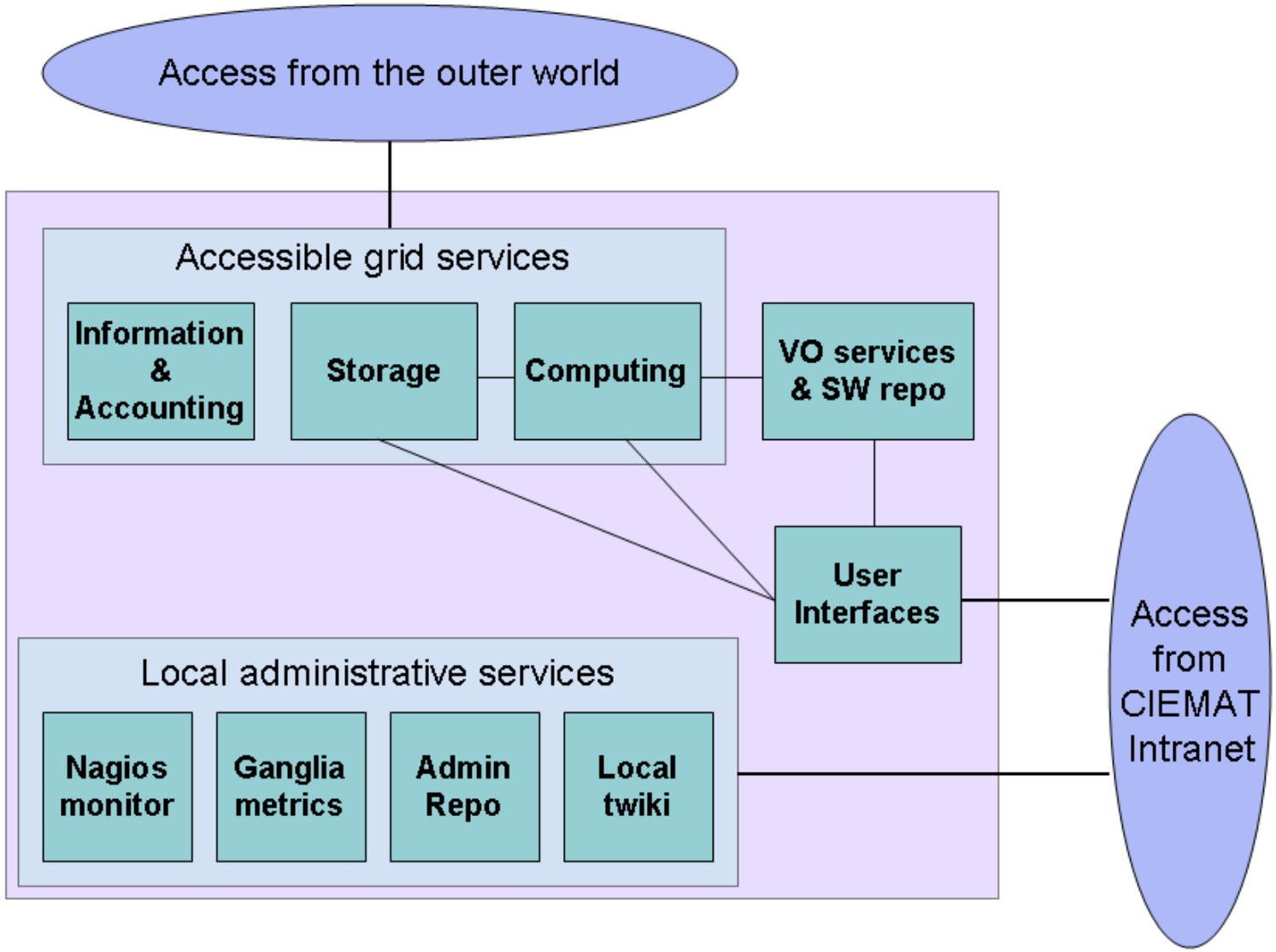
Other commitments

- Support local community of CMS physicists
  - Dedicated computing share
  - Resources for development and interactive analysis
  - User support and advice
- Support 3 analysis groups (those of more interest to us)
  - Storage capacity (50 TB per group)
  - Increased priority for privileged users
- Important testing exercises
  - STEP09, June: sustained transfer rates with T1s
  - October analysis exercise: analysis workflows
    - Including special group priorities and transfers among T2s

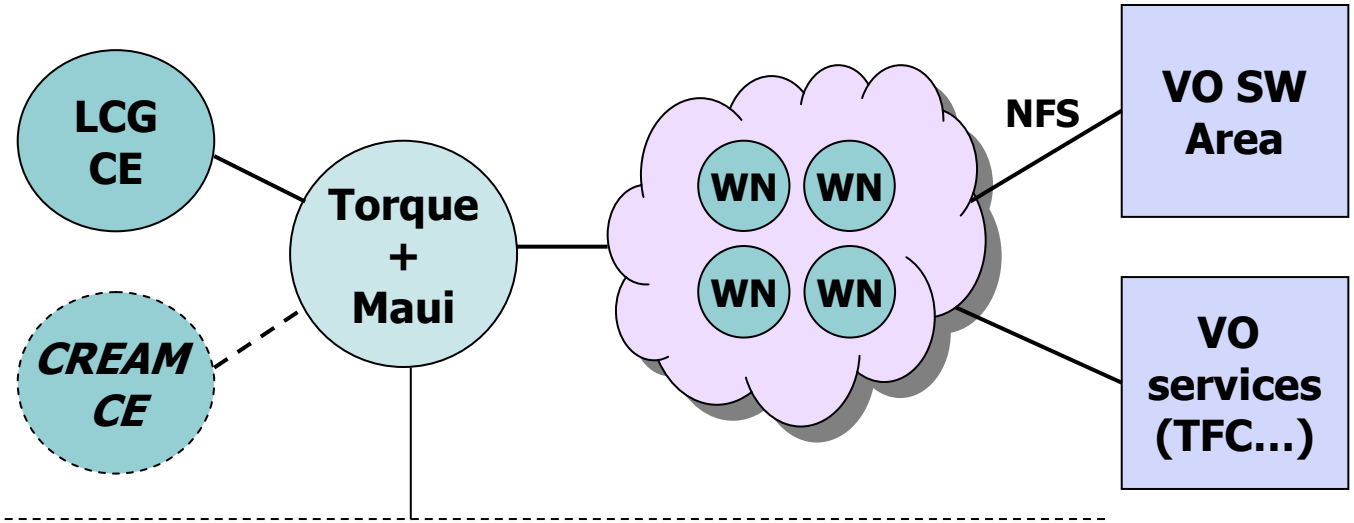
Agenda

- Introduction
- Activities and results
- Logical site setup
- Physical site setup and hardware studies
- Best practices
- Conclusions

CIEMAT logical site setup

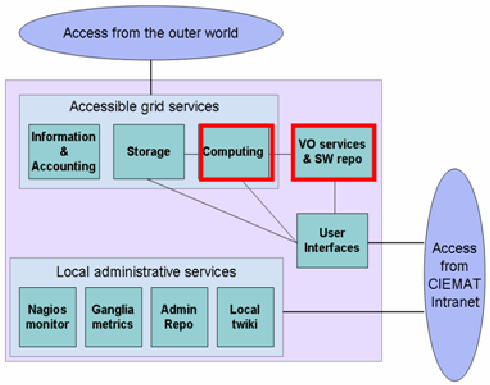


Computing resources

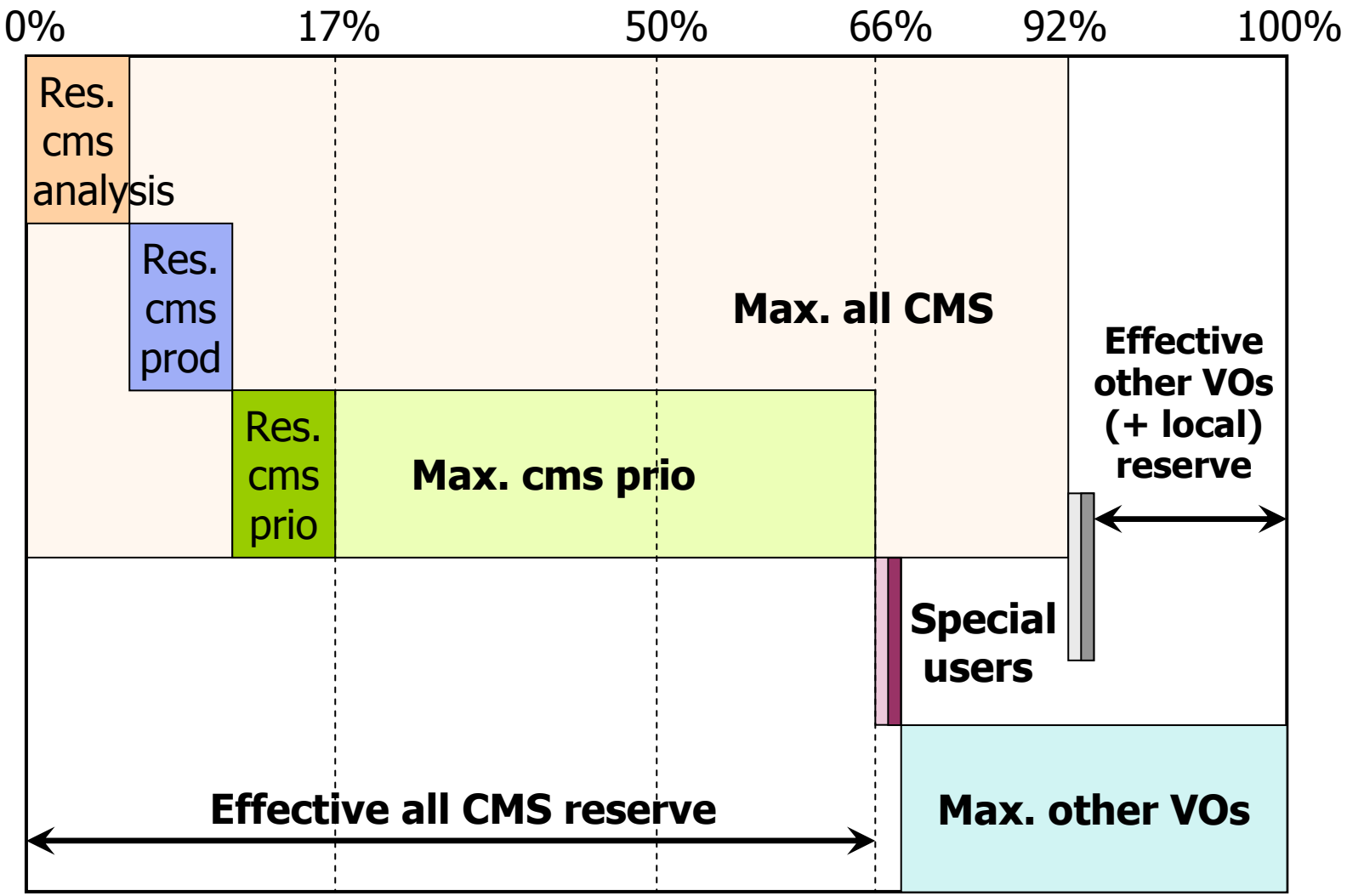


Maui configuration

- Fair share targets at 3 levels:
  - QoS (cms – local – other)
    - >> Group (prod – analysis)
      - >> User (default per user)
- Static high priority:
  - Special monitoring and SW install users (*sgm*)
- Standing reservations and job limits ➔ Next slide

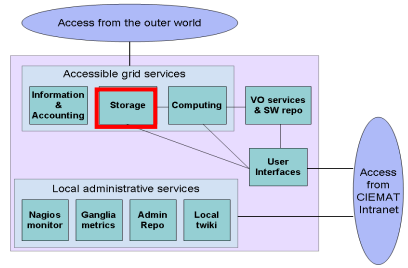


Maui reservations and limits



Storage resources

Namespace ←→ Pool group (isolated spaces)



/pnfs/ciemat.es/data/ …

Main dCache servers

SRM

Chimera

dCache admin

/cms/prod	**cms prod** **Permanent**
/cms/prod/scratch **/cms/scratch**	**cms prod, cms (rest)** **Volatile**
/cms	**cms (rest)** **Permanent**
/cms/store/group	**cms prio** **Permanent**
/<VO>	**VO users** **Permanent**
/ops **/dteam**	**ops, dteam users** **Volatile**

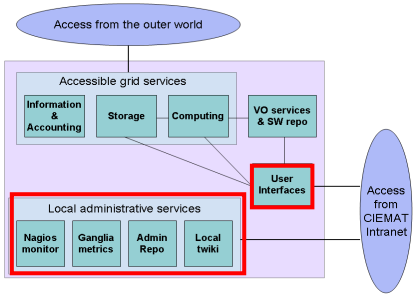
ACLs required

… group/muon: cmsmuon

… group/ewk: cmsewk

Other resources

- User Interfaces
  - Shared NFS personal area (all UIs look the same)
  - Direct access to VO environment, SW repository, services
  - Access to SE (including mounted namespace)
  - ➔ User code development and debugging
  - ➔ Grid submission once code is ready

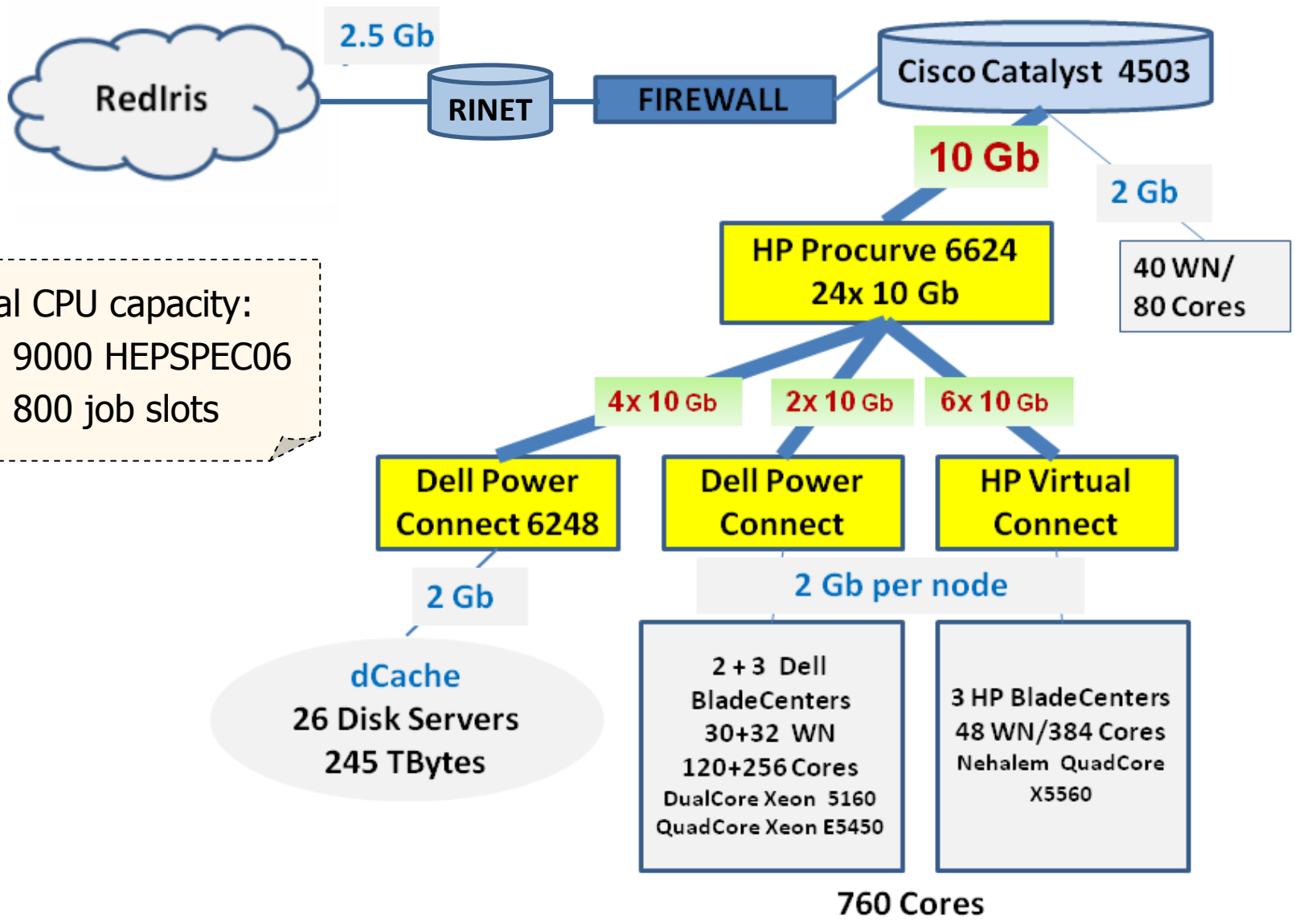


- Administrative services
  - Nagios: monitoring, external tests and home-made probes
  - Ganglia: statistical information about site performance
  - Admin repo: central repository for configuration + remote installation and execution
  - Twiki server: information sharing and documentation

Agenda

- Introduction
- Activities and results
- Logical site setup
- Physical site setup and hardware studies
- Best practices
- Conclusions

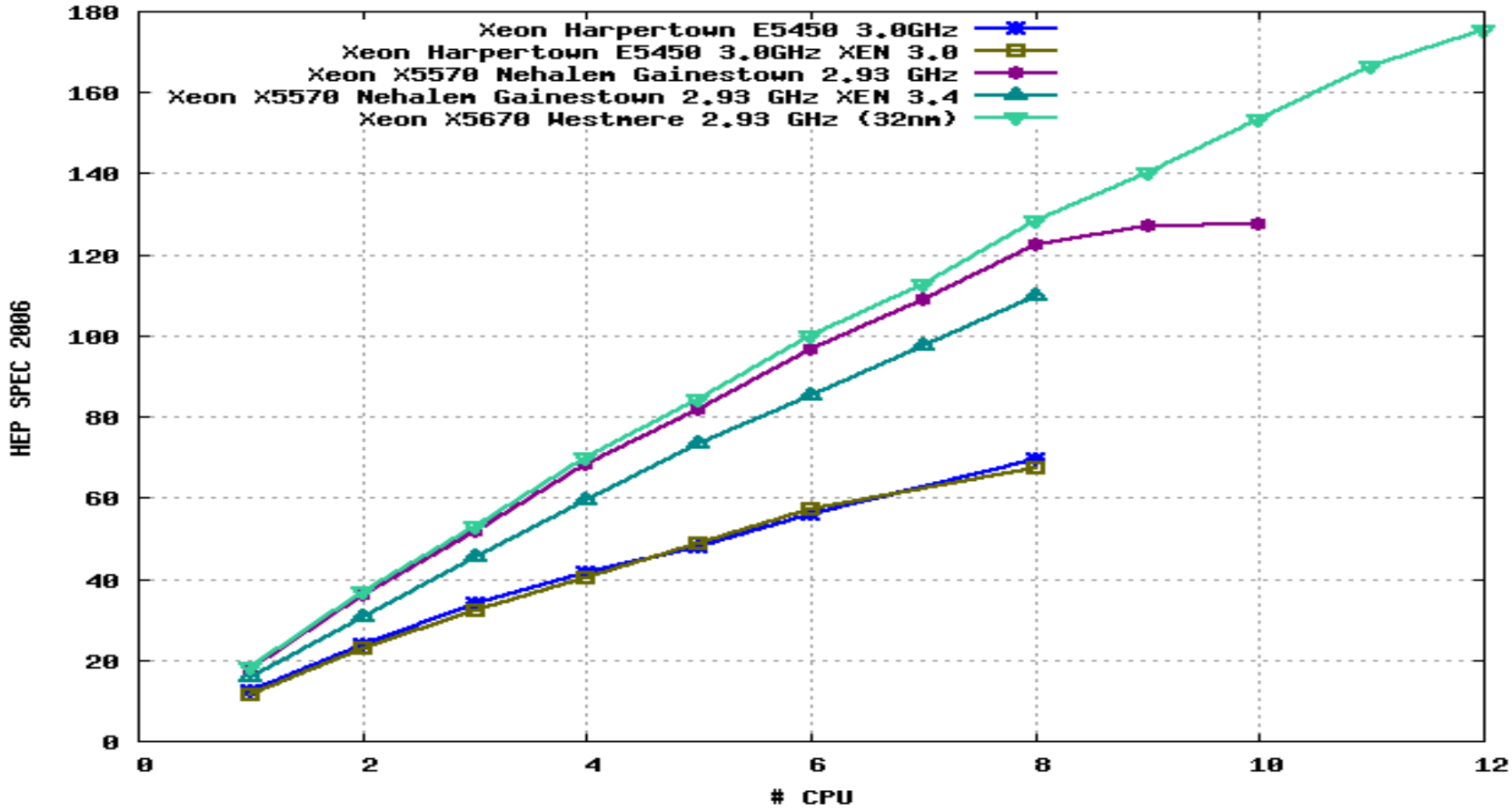
CIEMAT physical site setup



Total CPU capacity:
~ 9000 HEPSPEC06
~ 800 job slots

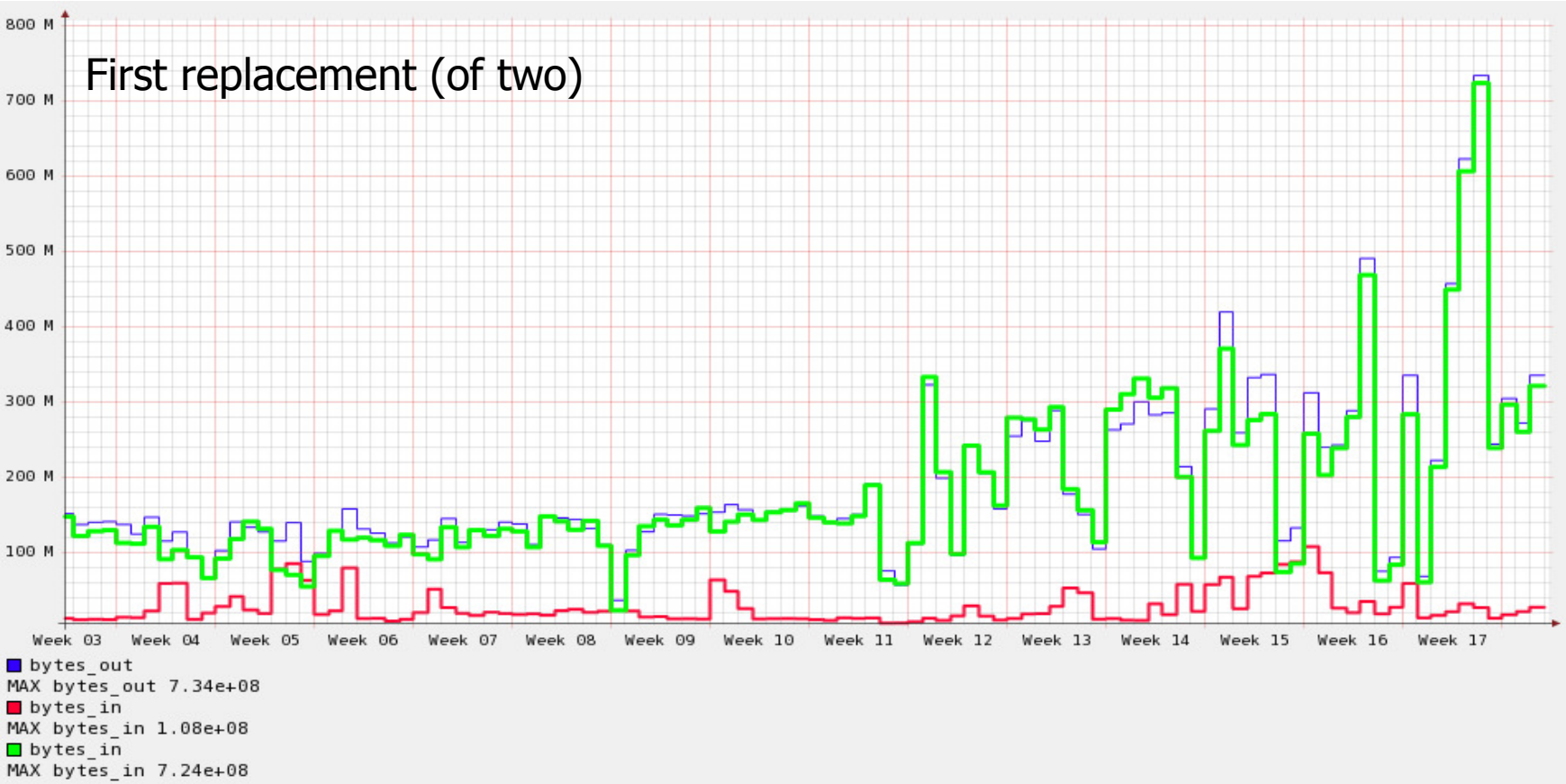
Hardware characterization

- Performance measurements on different hardware
  - SPECs for CPUs, I/O for disk servers (concurrent processes)
  - Example: exclude Virtual Machines for new processors

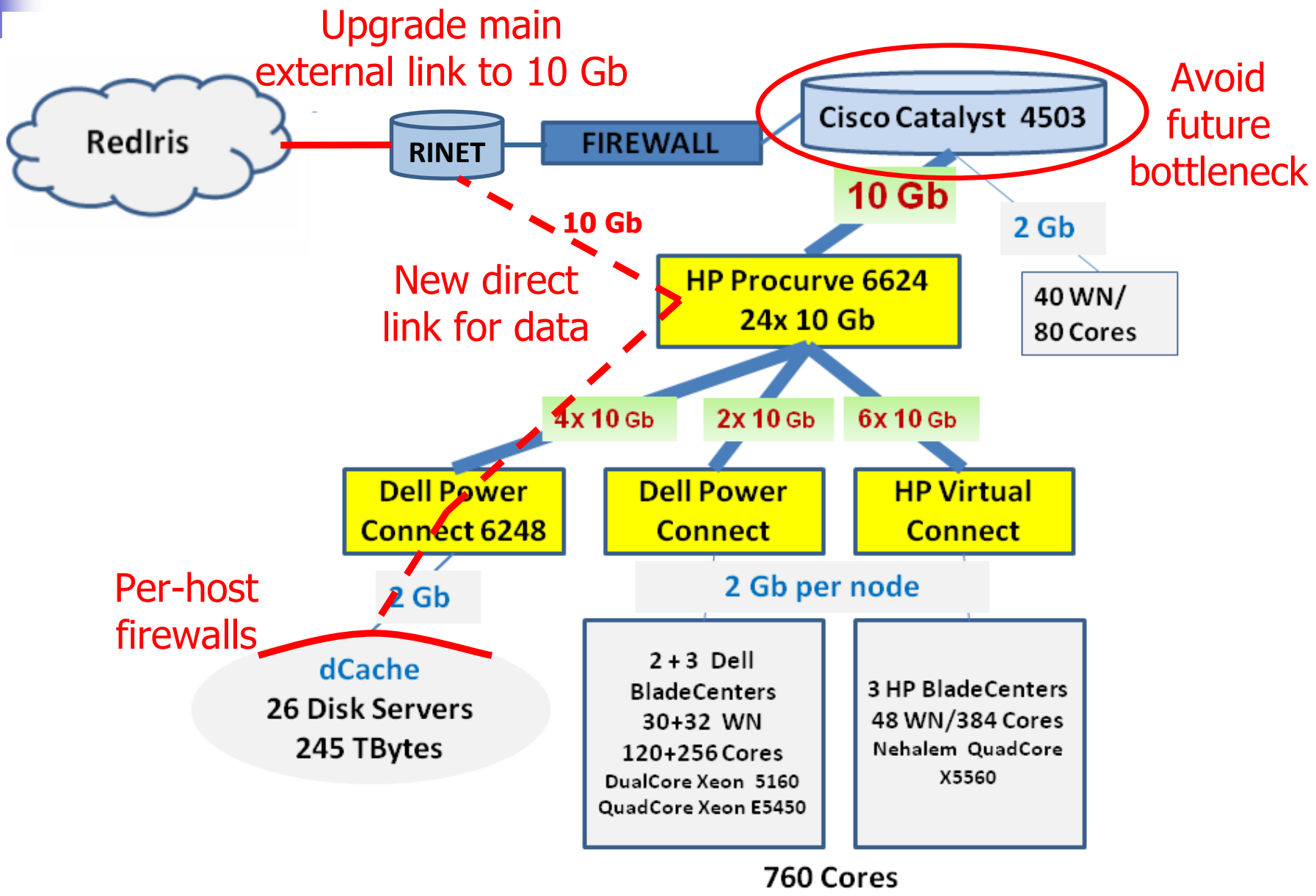


Network reconfiguration

- Passed from 1 Gbps to 12 Gbps in traffic WNs ←→ SE
  - New routers with higher internal bandwidth
  - Bonding in disk servers connections to switches



Future network improvements



Agenda

- Introduction
- Activities and results
- Logical site setup
- Physical site setup and hardware studies
- Best practices
- Conclusions

Best practices

- Document and automate
  - Create procedures for frequent, complex, error-prone tasks
  - Make them scripts when possible
- Keep track of customizations
  - Local configuration always needed, patches often
  - Use some central repository to avoid losing knowledge
- Test as much as possible
  - Prepare upgrades and installs in advance
  - When possible install a new machine and replace old one
  - Deploy test services to play with (…mistrust documentation)

Best practices (II)

- Monitor
  - Keep an eye on all monitoring systems (SAM, VO, local)
  - Make tests warn you actively (e.g.: e-mail)
  - Add a new test for every new problem found
- Listen to the VOs (customers)
  - Try to understand what they are doing
  - Try to anticipate what they plan to do
  - This is an important key for success in the case of CIEMAT and the CMS VO

Agenda

- Introduction
- Activities and results
- Logical site setup
- Physical site setup and hardware studies
- Best practices
- Conclusions

Conclusions

- CIEMAT is a successful grid site of WLCG/EGI
  - Actively used by VOs
  - Complied with expectations and requirements
  - Higher than average figures of merit in many cases
- Keeping quality up requires continuous evolution
  - Enhancing grid services functionality
  - Improving efficiency of resources
  - Easing infrastructure configuration and management
- Hopefully, our experience has taught us something…
  - We have identified several best-practice lessons