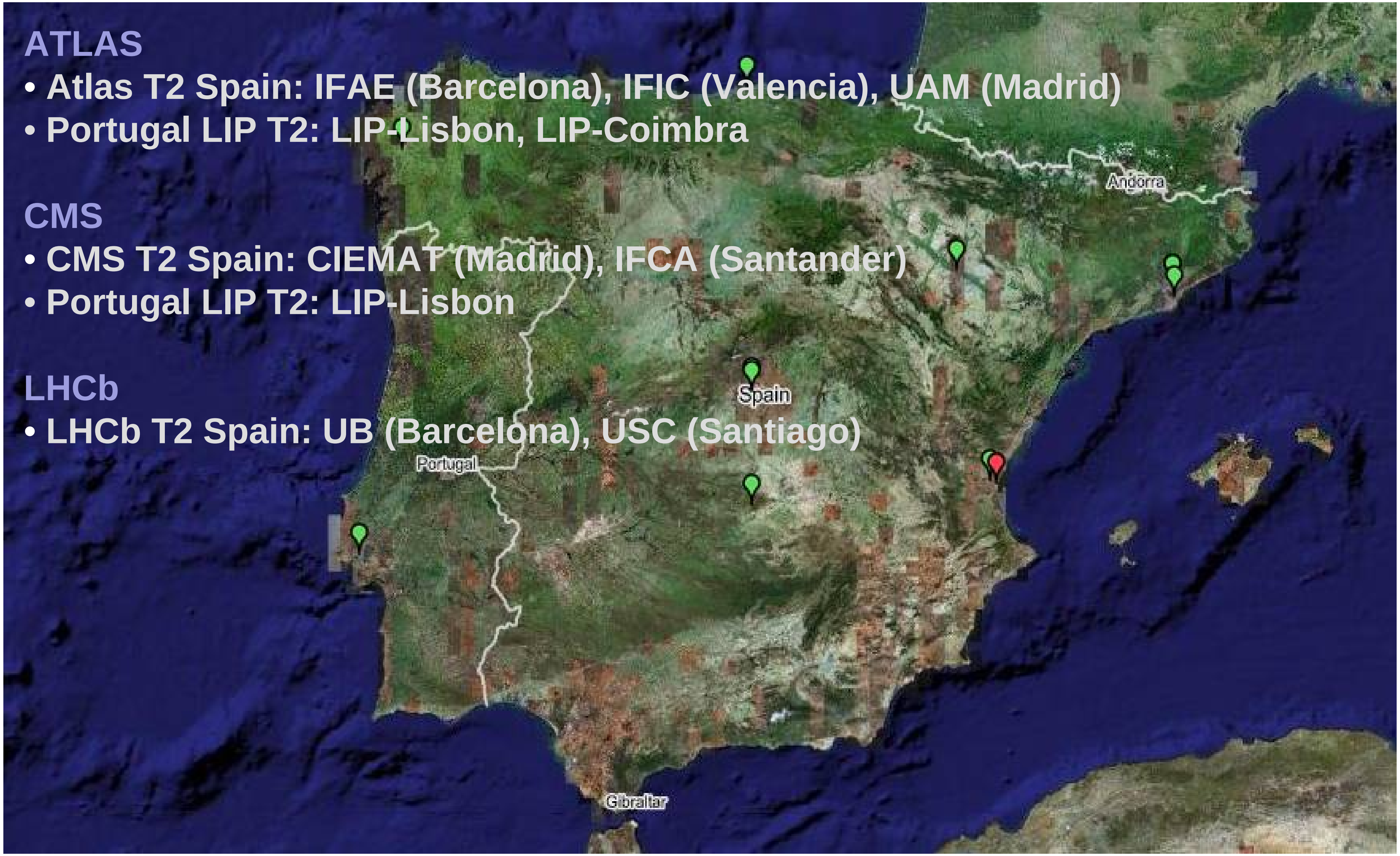
Report from PIC Tier-1 and associated Tier-2s

G. Merino, wLCG workshop, CERN 24-01-2007

Tier-2s associated to PIC



ATLAS

- **Atlas T2 Spain: IFAE (Barcelona), IFIC (Valencia), UAM (Madrid)**
- **Portugal LIP T2: LIP-Lisbon, LIP-Coimbra**

CMS

- **CMS T2 Spain: CIEMAT (Madrid), IFCA (Santander)**
- **Portugal LIP T2: LIP-Lisbon**

LHCb

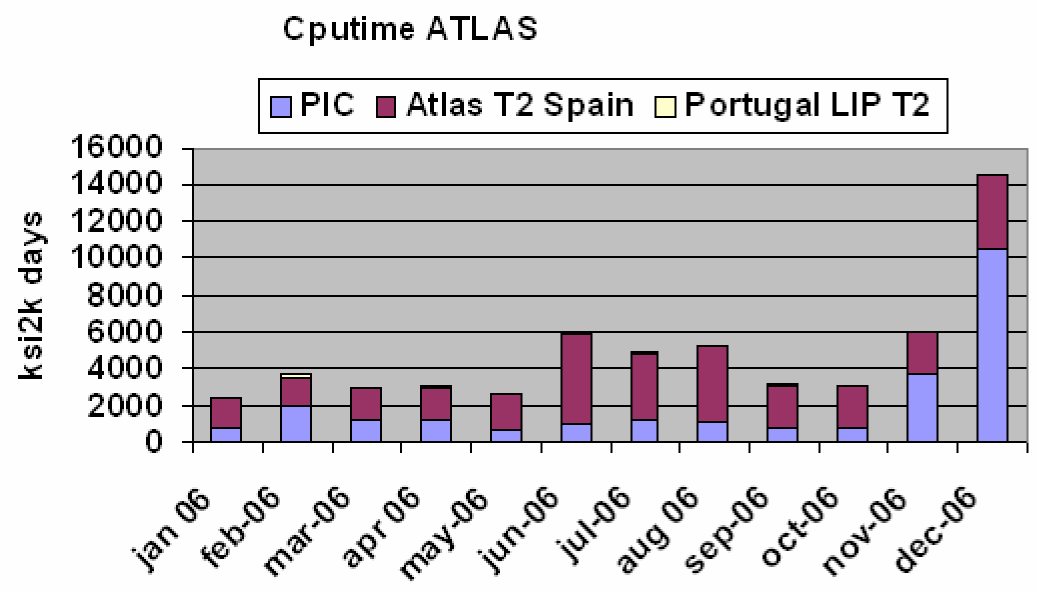
- **LHCb T2 Spain: UB (Barcelona), USC (Santiago)**

Tiers Capacity

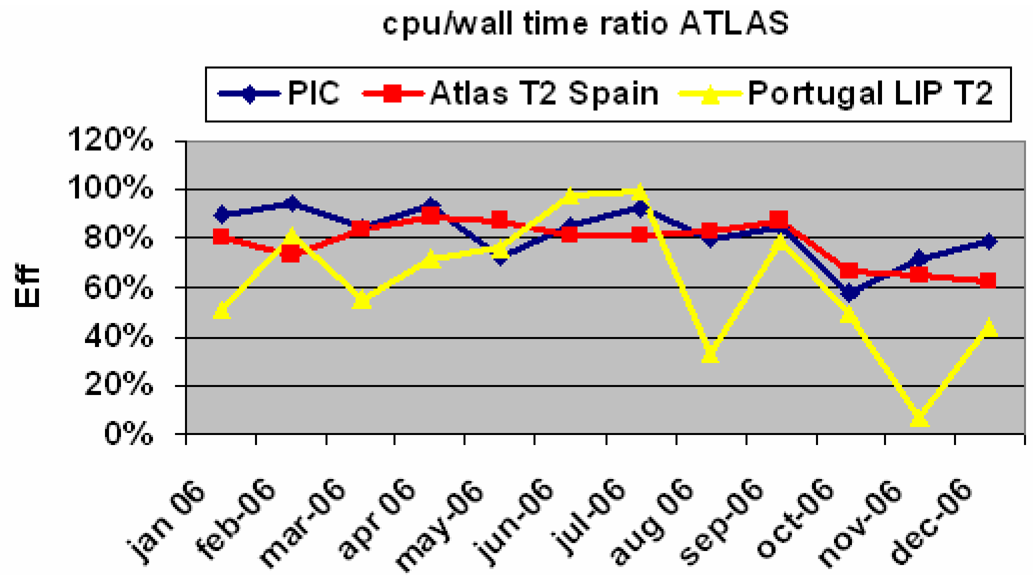
	Atlas	CMS	LHCb		*Current*	2007	2008	2009	2010
PIC	~5%	~5%	~6,5%	**cpu (ksi2k)**	*600*	501	1654	2647	5381
				disk (TB)	*69*	218	845	1578	2878
				tape (TB)	*140*	243	1149	2425	4473
Atlas T2 Spain	~5%			**cpu**	*410*	117	875	1349	2577
				disk	*40*	63	387	656	1107
CMS T2 Spain		~5%		**cpu**	*340*	380	760	1280	2260
				disk	*51*	65	210	420	665
LHCb T2 Spain			~6,5%	**cpu**	*200*	200	300	750	750
				disk	*1*	1	1	1	1
Portugal LIP T2	x	x		**cpu**	*25*	500	750		
				disk	*4*	84	130		

- Spain did not sign the MoU yet. These 2007-2010 pledges are new estimations derived from Oct-2006 new exp. requirements
  - No major problems foreseen for providing capacity for Jul-2007
  - 2007-2008 ramp up is really big (>3x, specially in disk)
  - Deployed capacity tries to fit cpu/disk/tape ratios from exp
- New pledges for Portugal still not available

APEL Accounting ATLAS

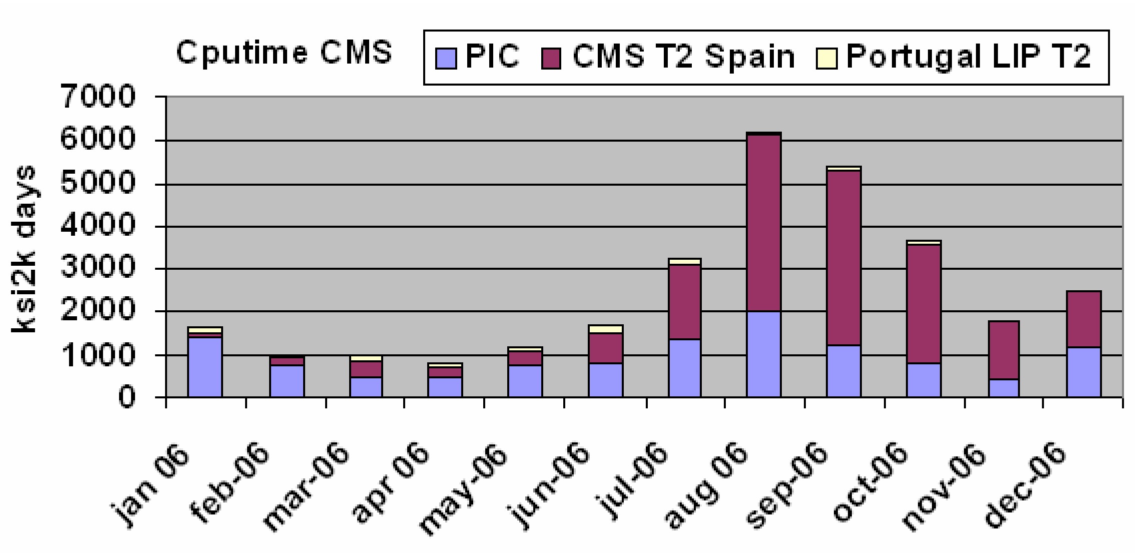


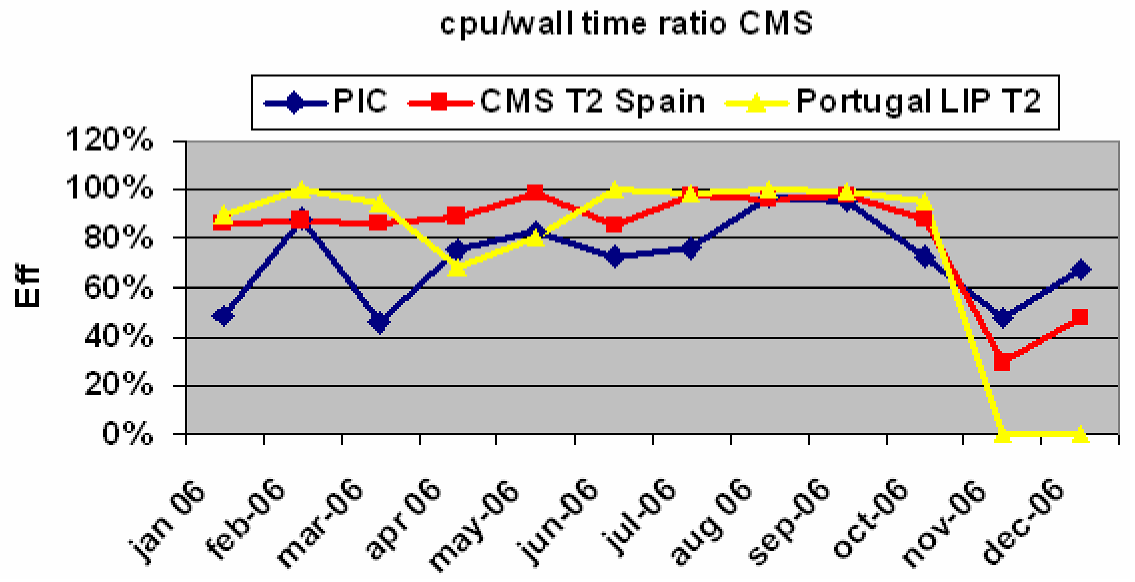
	Total cput (ksi2k*days)
PIC	25187
Atlas T2 SP	32111
PT LIP T2	608



- CPU eff. stayed >70% for most of the year
- October effect understood (condor glide_ins)
- LIP-Portugal yet to enter in the Atlas production machinery

APEL Accounting CMS

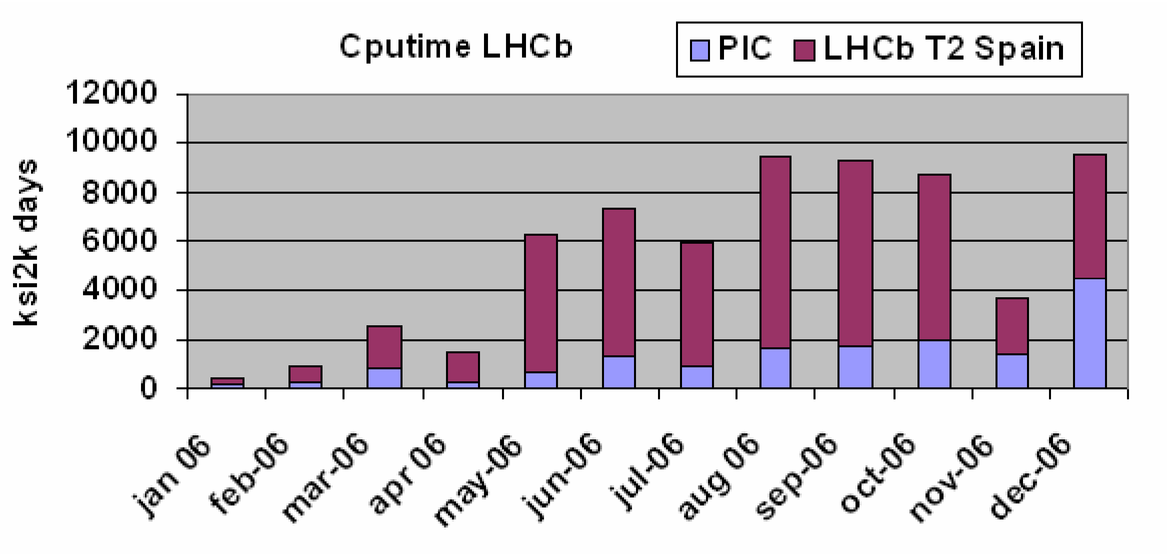




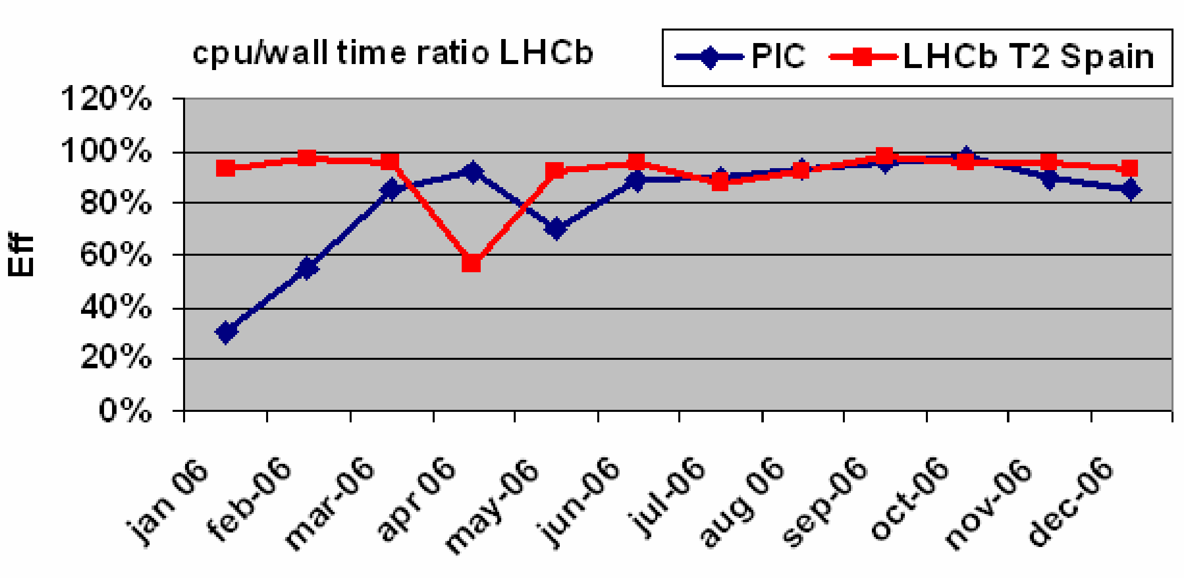
	Total cput (ksi2k*days)
PIC	11755
CMS T2 SP	17061
PT LIP T2	1060

- CPU eff. stayed >70% for most of the year
- November effect understood (mostly analysis jobs after CSA06. Lots of disk access)
- LIP-Portugal yet to enter in the CMS production machinery

APEL Accounting LHCb



	Total cput **(**ksi2k*days)
PIC	15677
LHCb T2 SP	49930



- CPU eff. stayed >70% for most of the year
  - Jan,Feb dip at T1 due to PBS problem + small statistics
- LHCb-T2-SP: Despite being a small T2, delivers quite high capacity

Batch / Storage Technologies

		Batch		Storage		
		TorqueMaui	SGE	DPM	dCache	Castor
PIC		prod.			prod.	prod.
Atlas T2 SP	IFAE	prod.			prod. (*)	
	IFIC	prod.			testing	prod.
	UAM	prod.			prod.	
CMS T2 SP	CIEMAT	prod.		testing	testing	prod.
	IFCA	prod.		prod.		
LHCb T2 SP	UB	prod.				
	USC	prod.				
PT LIP T2	LIP-Lisbon		prod.			
	LIP-Coimbra	*planned*		*planned*		

- (*) IFAE using Fermilab implementation of the SRM interface to a Unix FS
- PIC currently using dCache for disk and Castor1 for tape. Castor2 still in testing mode. Need to take decision in the next weeks

Network Connectivity to GÉANT

	Current	Planned
PIC	1Gbps	10Gbps (expected May 2007)
Atlas T2 Spain	1Gbps (at the 3 centres)	10Gbps (expected by 2008)
CMS T2 Spain	622Mbps (IFCA) 2,5Gbps (CIEMAT)	2,5Gbps (IFCA this year)
LHCb T2 Spain	100Mbps (UB) 2Gbps (USC)	
Portugal LIP T2	100Mbps	1Gbps (expected by 2007)

Manpower

	Operations	Experiment-specific
PIC	12,5	3
Atlas T2 Spain	4,5	6
CMS T2 Spain	5	4
LHCb T2 Spain	1	1
Portugal LIP T2	8	2

Schedule Downtimes in 2006

		SD days	Nr. SDs
PIC		9,9	5
Atlas T2 SP	IFAE	6	4
	IFIC	6	5
	UAM	13,3	8
CMS T2 SP	CIEMAT	10,2	12
	IFCA	4	7
LHCb T2 SP	UB	32,6	6
	USC	23,9	5
LIP-Lisbon		37	6

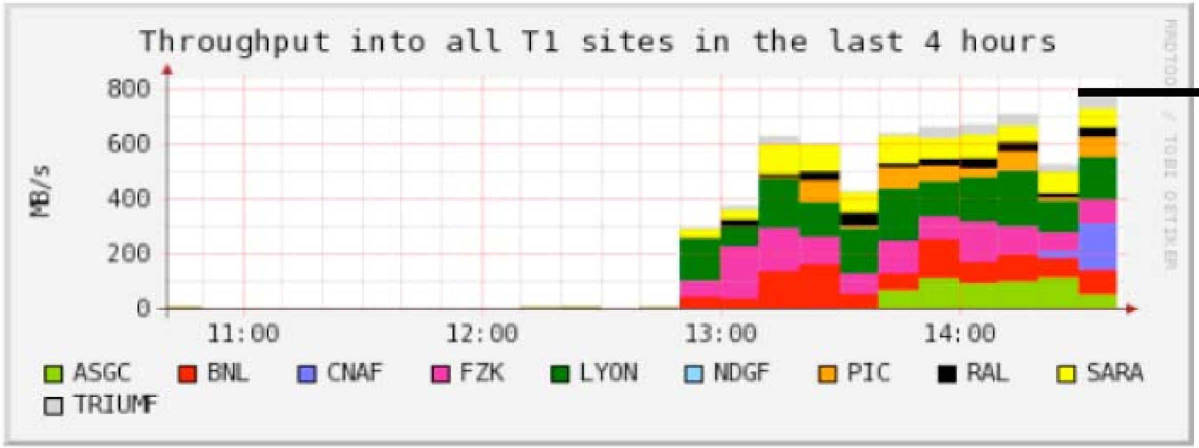
- LIP-Lisbon: upgrade to dCache-1.7 + major network intervention
- UB: Unexpected power cut and major site reconfiguration
- USC: Unexpected power cut (coinciding with gLite upgrade!)
- PIC: ~30% electrical maintenance (rest gLite-3 in May and PBS security threat in Nov)

Contribution to ATLAS LCG activities

- Distributed Monte Carlo production
  - Coordination of MCprod shift system
  - Active follow-up of MCprod status (job failure reasons analysis, ...)
- Distributed Analysis
  - Test of torque/maui config. for job prioritisation
  - Deployment of GANGA for user analysis
- Distributed Data Management
  - Participation in the DDM Operations Team
  - Good results in the ATLAS “DDM challenges”
    - Tier-0 Scale Test, July-2006
    - T1-T2 Functional Tests, Sep-Oct-2006

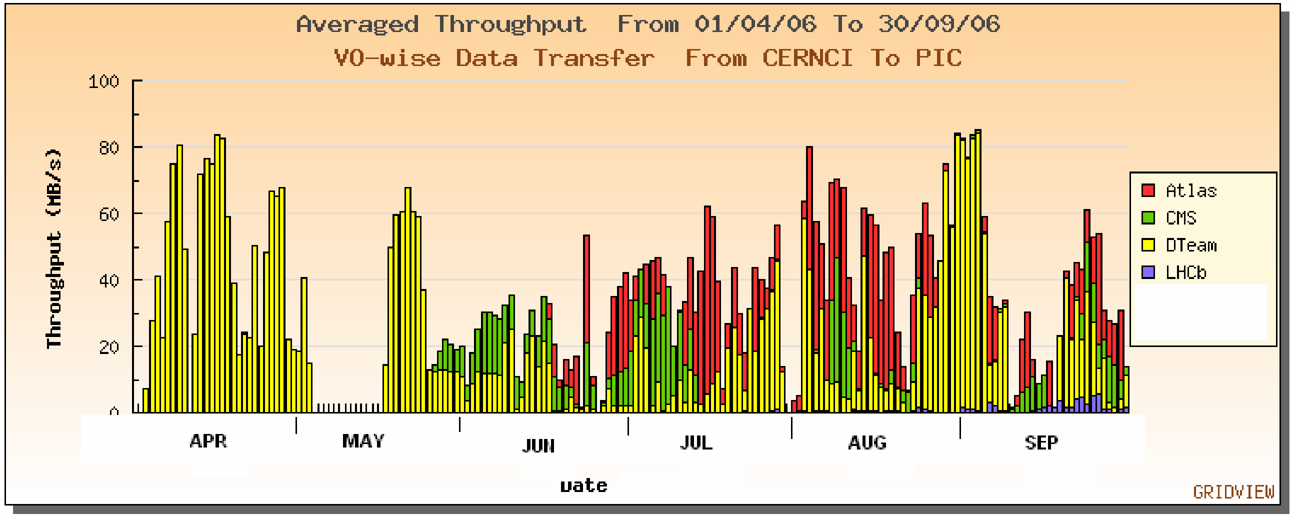
ATLAS Data Transfers Tests at PIC

- T0 scaling test July-2006



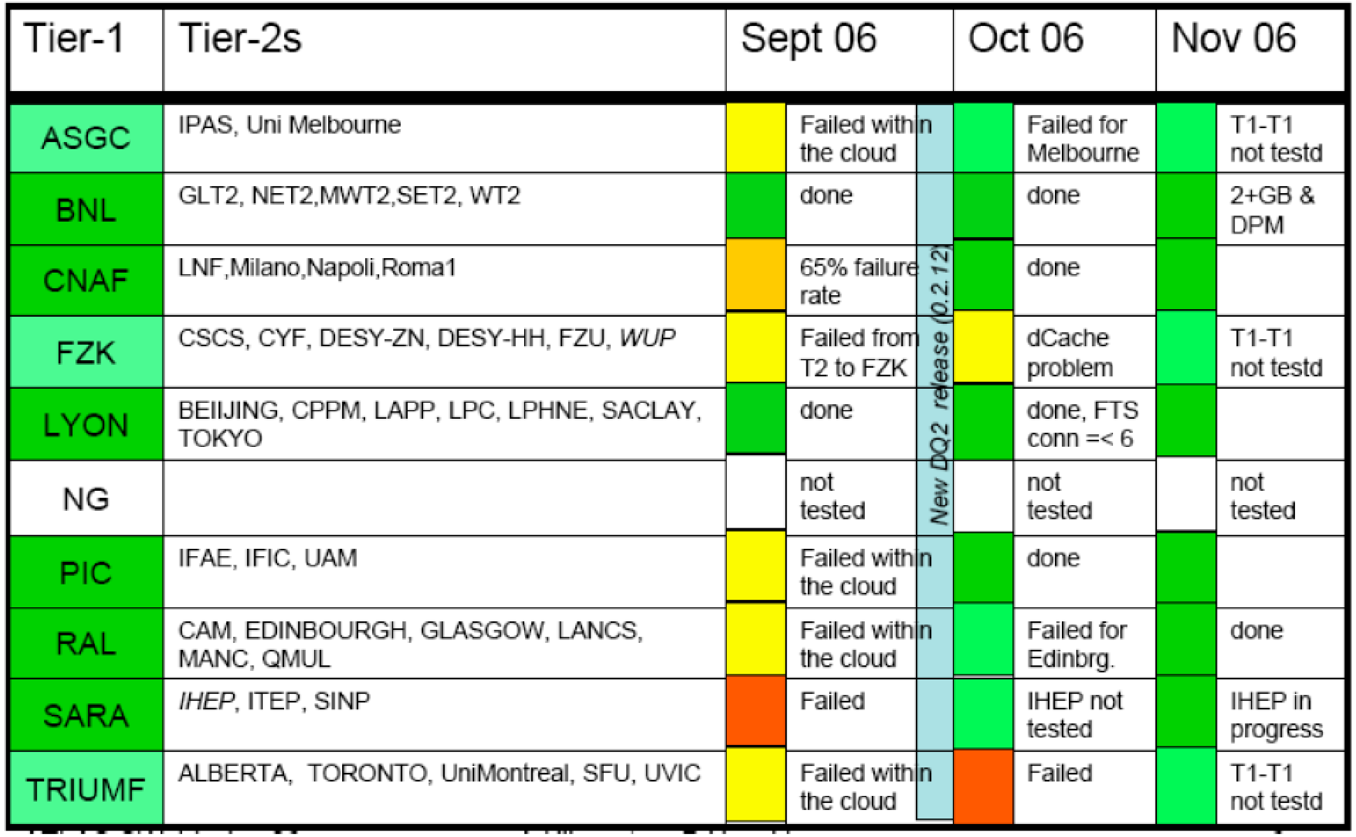
700MB/s (9T1s)

- Transfer load T0→T1 continued during summer (together with CMS)...



ATLAS T1-T2 Functional Tests

- T1-T2 functional tests (Sep and Oct 2006)
  - Problems in september (LFC overloaded)
  - Ok in October

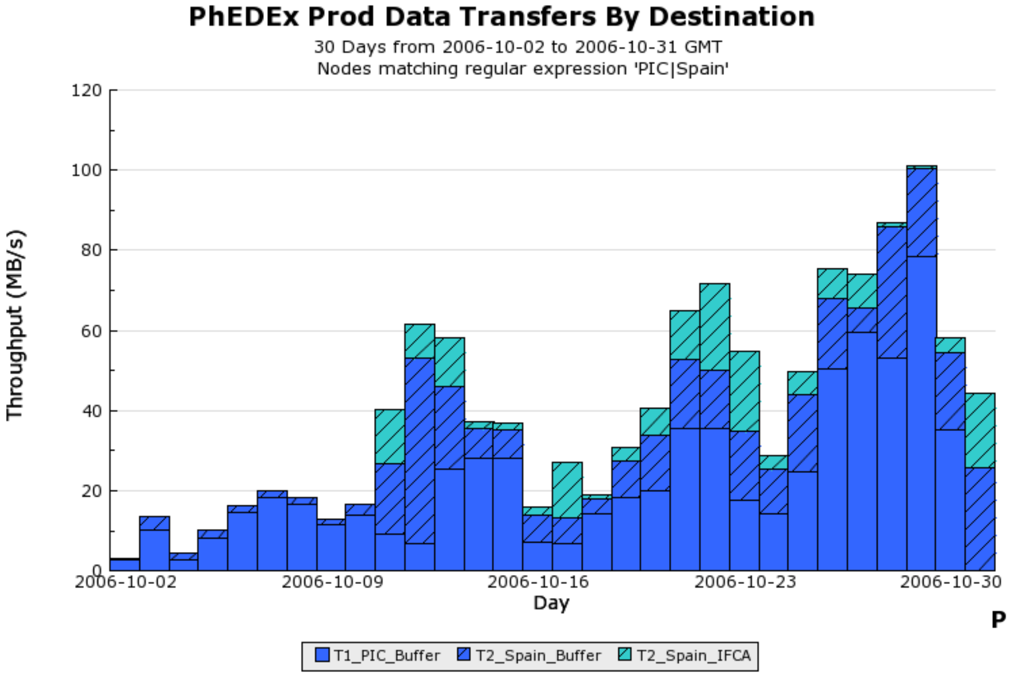


Tier-1	Tier-2s	Sept 06			Oct 06		Nov 06	
ASGC	IPAS, Uni Melbourne	🟨	Failed within the cloud	*New DQ2 release (0.2.12)*	🟩	Failed for Melbourne	🟩	T1-T1 not testd
BNL	GLT2, NET2,MWT2,SET2, WT2	🟩	done		🟩	done	🟩	2+GB & DPM
CNAF	LNF,Milano,Napoli,Roma1	🟧	65% failure rate		🟩	done	🟩	
FZK	CSCS, CYF, DESY-ZN, DESY-HH, FZU, *WUP*	🟨	Failed from T2 to FZK		🟨	dCache problem	🟩	T1-T1 not testd
LYON	BEIIJING, CPPM, LAPP, LPC, LPHNE, SACLAY, TOKYO	🟩	done		🟩	done, FTS conn =< 6	🟩	
NG		⬜	not tested		⬜	not tested	⬜	not tested
PIC	IFAE, IFIC, UAM	🟨	Failed within the cloud		🟩	done	🟩	
RAL	CAM, EDINBOURGH, GLASGOW, LANCS, MANC, QMUL	🟨	Failed within the cloud		🟩	Failed for Edinbrg.	🟩	done
SARA	*IHEP*, ITEP, SINP	🟥	Failed		🟩	IHEP not tested	🟩	IHEP in progress
TRIUMF	ALBERTA, TORONTO, UniMontreal, SFU, UVIC	🟨	Failed within the cloud		🟥	Failed	🟩	T1-T1 not testd

Contribution to CMS CSA06

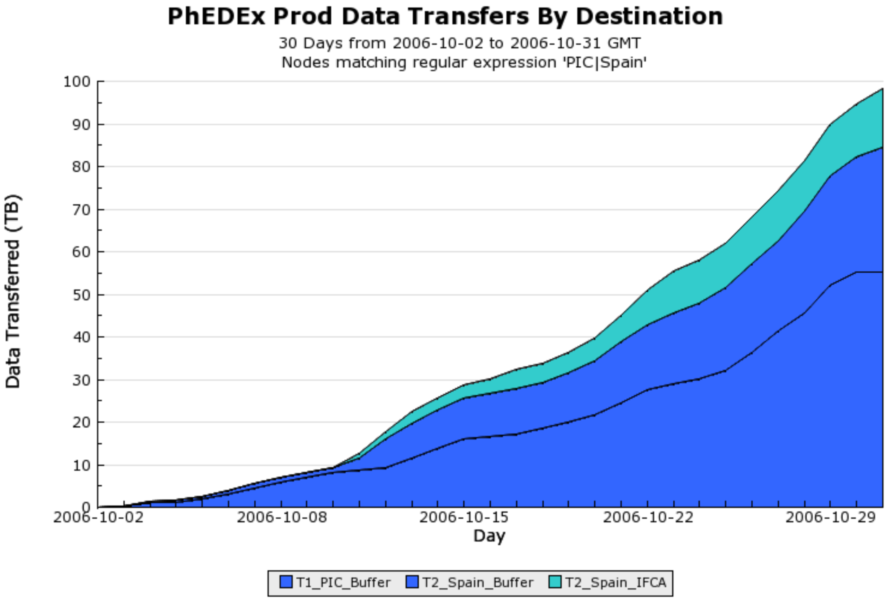
- Participation of PIC Tier-1 and Spanish Tier-2 (CIEMAT/IFCA) in CSA06 CMS computing challenge in all data- and workflows with excellent results:
  - Data Transfers T0→PIC→T2_Spain running backlog-free, with high efficiency
  - Bursty transfers T0→PIC, PIC→T2s and PIC→T2_Spain successfully exercised
    - Essentially saturating the available network bandwidth
  - Skimming and re-reconstruction workflows successfully run at PIC at a large scale
    - Reading calibration/alignment constants via local Frontier cache
  - User Data analysis over skimmed data run at T2_Spain
    - Peak of ~7000 user jobs/day
  - Alignment workflow successfully exercised at T2_Spain

Contribution to CMS CSA06

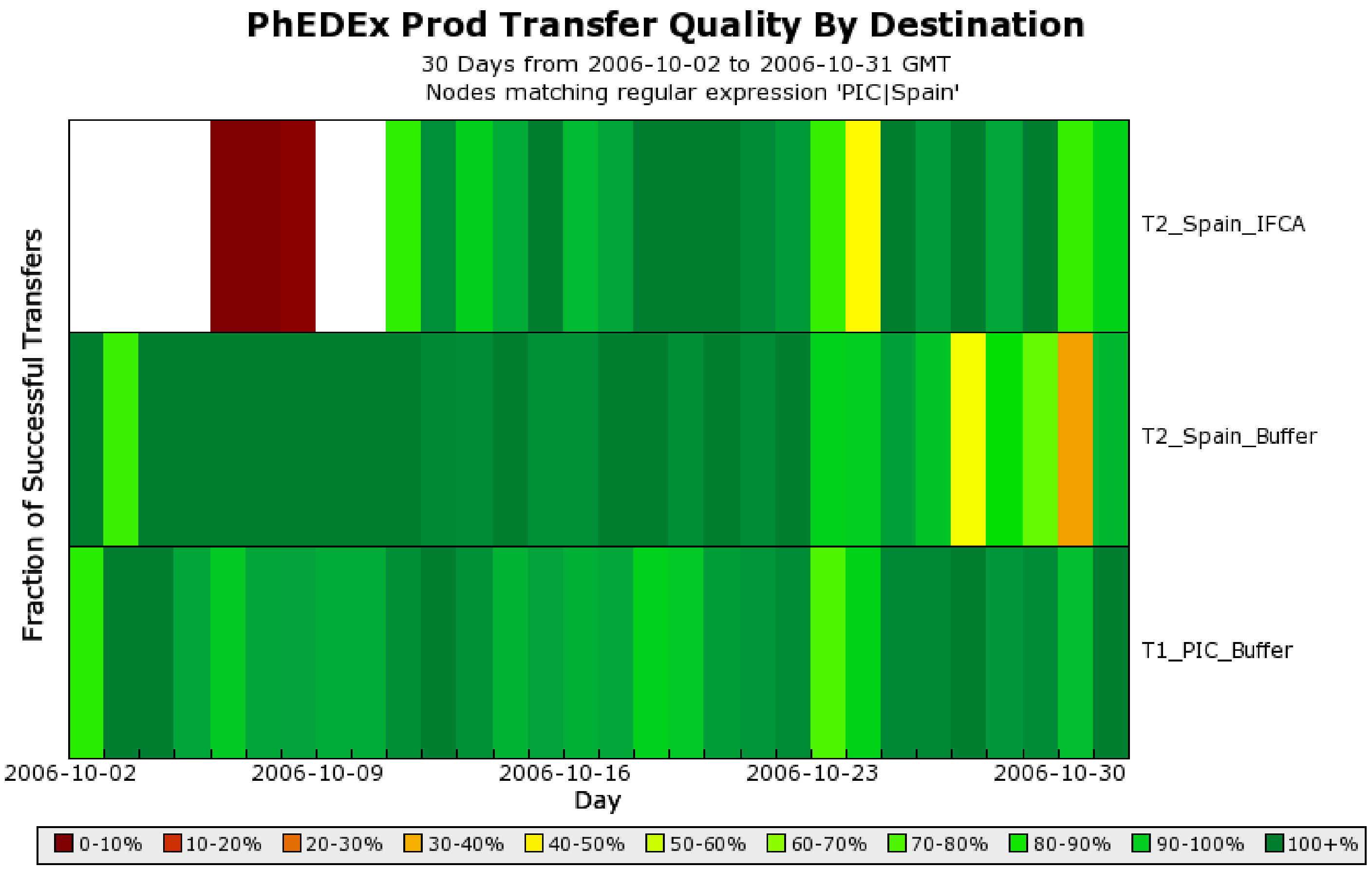


- About 20-80 MB/s T0→T1
- About 20-40 MB/s T1→T2
- Both concurrent and sustained during several days

- O(50TB) moved into the T1 and O(50TB) into the T2



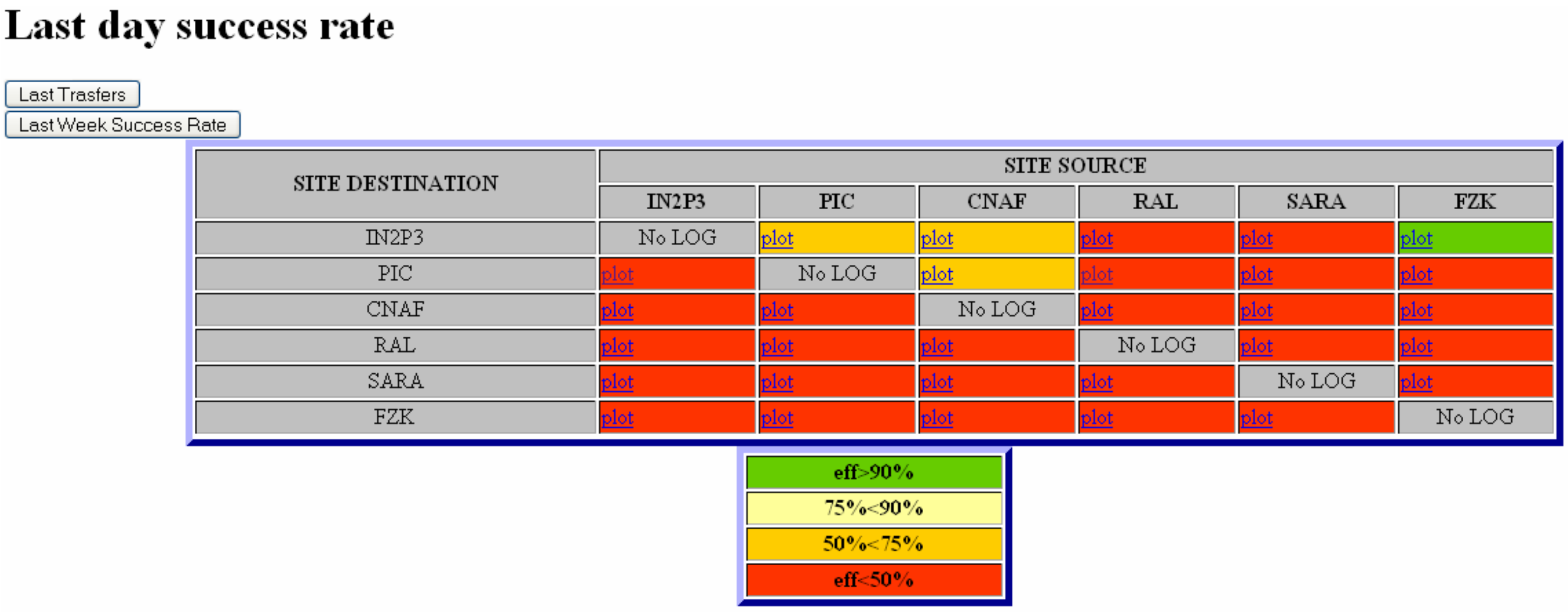
Contribution to CMS CSA06



- Transfer quality nicely monitored by CMS
- Transfers proceeded with few errors during the exercise

Contribution to LHCb LCG activities

- Successful participation in the SC4 test of massive online reconstruction at the Tier-1
  - About 20TB of *digi* files transferred to PIC and reconstructed online
- Participation in the T1-T1 continuous transfer monitoring
  - ISSUE: very difficult to get stable/reliable transfers for long periods of time



Last day success rate

SITE DESTINATION	SITE SOURCE					
	IN2P3	PIC	CNAF	RAL	SARA	FZK
IN2P3	No LOG	🟧 plot	🟧 plot	🟥 plot	🟥 plot	🟩 plot
PIC	🟥 plot	No LOG	🟧 plot	🟥 plot	🟥 plot	🟥 plot
CNAF	🟥 plot	🟥 plot	No LOG	🟥 plot	🟥 plot	🟥 plot
RAL	🟥 plot	🟥 plot	🟥 plot	No LOG	🟥 plot	🟥 plot
SARA	🟥 plot	🟥 plot	🟥 plot	🟥 plot	No LOG	🟥 plot
FZK	🟥 plot	🟥 plot	🟥 plot	🟥 plot	🟥 plot	No LOG

🟩 eff>90%
🟨 75%<90%
🟧 50%<75%
🟥 eff<50%

Hardware procurement - WNs

- Procurement of substantial amount of WNs took place last year at various sites
- The final configuration choice depends on several parameters, but key figures are ksi2k/€ and ksi2k/Watt
- Depending on the date of the purchases, sites ended up with different configurations, mainly:
  - 2 x Opteron 270 Dual Core 2.0 GHz (~before summer purchases)
  - 2 x Intel Xeon 5160 3GHz Dual Core (~after summer purchases)
- The specs in ksi2k of one option aprox. doubles the other
  - Experiments starting to look at exp-specific benchmarks
  - First results from LHCb suggest an over-estimation of ~20-25% in the relative power of both CPUs
- This might be an issue, given that the requirements → procurement process is very much based on si2k benchmark

Issues: deploying robust services

- Use as much robust hw as possible
  - Dual power supply, RAID hot-swap HDs, lots of CPU and RAM ...
- Difficult to deploy services in High Availability mode
  - lcg-CE: trying to deploy two CEs publishing identical queues info
    - If one CE dies, all the jobs it manages are lost
    - Is there any other recipe?
    - Will this “dual-CE” recipe apply to the glite-CE?
  - SRM: Using DNS load-balancing among various hosts
    - Automatic DNS switch not yet implemented
    - DNS switch will take time to propagate. Are there other HA recommended configurations?
  - FTS/LFC: HA at the level of the DB.
    - Still studying best way to deploy the server front-end in HA mode. DNS load-balancing + switch is a recommended option?

Issues

- Deploying an old OS (SL3) makes it difficult to install the services on the new hardware (controllers, ...)
- Concern at Tier-2s about the powerful hw needed to run certain services
  - E.g. CE, dCache-admin, MON seem to prefer dual-proc and 4GB RAM
- Sometimes, disk space for a VO at a site happens to be full
  - Need tools to ease the “purging” of the space (LFC-SRM consistency check, detect very-old-obsolete files ...)
  - If CPU slots are still available, can the experiments make use of them redirecting output to another site?
- Users are sometimes reluctant to adopt GRID technologies for their analysis
  - Some T2s are implementing User Support services to help on this

Issues

- ATLAS Tier-2 concerned about the CPU inefficiency (installed vs. delivered CPU capacity) observed
  - Currently trying to understand the origin of this effect
- ATLAS also starting to have a look into monitoring job failure rates and failure modes at the sites
  - Having fast access to this monitoring information can be very useful for the sites
  - Need to get a consistent and digested view
    - currently atlas-prodsys and dashboard do not really match
    - failure mode pie chart is still too complicated
- LHCb sees the T1-T1 transfers are not reliable enough during long periods of time
- CMS concerned by the fact that the tools for job prioritisation are not yet there and everything still needs to be done “by hand”

Summary

- The PIC Tier-1 and associated Tier-2s have been presented
  - 3 Federated Tier-2s in Spain
  - 1 Federated Tier-2 in Portugal
- Weekly operations meetings in the context of the EGEE-SWE federation keep the base infrastructure coordination at a good level
  - Eg, all sites reporting APEL accounting since long time
- Experiment-specific T1-T2 coordination meetings being set up now with monthly initial periodicity
  - Will help in getting LIP integrated inside ATLAS and CMS LCG activities
  - Enter in a continuous-test mode as soon as we can
- Contribution of the sites to the LCG activities has been up to now satisfactory, given the resources
- A number of issues have been presented