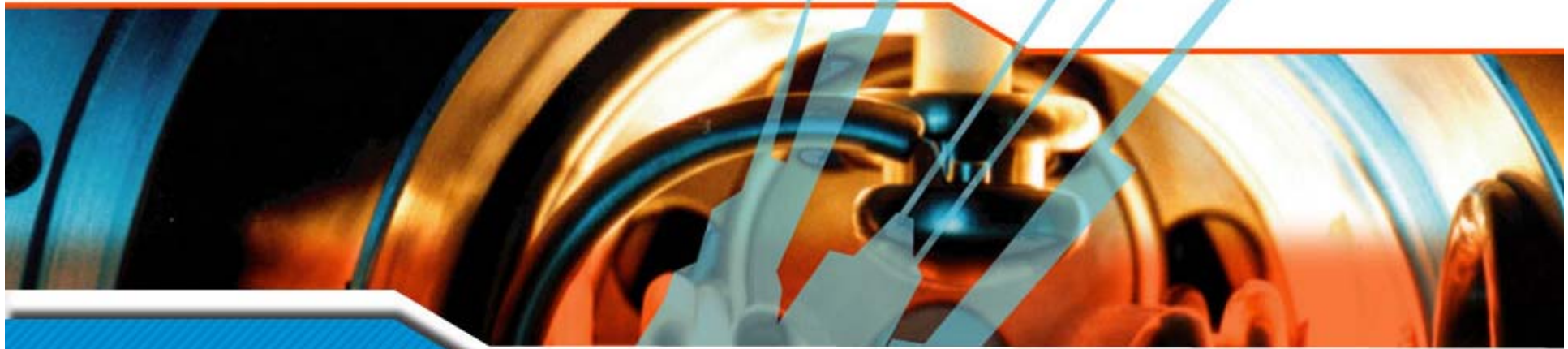




Overview of day-to-day operations

Suzanne Poulat



Overview



- Human resources
- Operation group role
- Services during Out of Office Hours
- Monitoring
- Tickets and alarms
- Documentation
- Conclusion



Human resources in Operation group



- 8 people in the Operation group (7.8 FTE)
 - 2 for contributing to operations of global and national grids
 - 4 for site operation (3.8 FTE)
 - From June 2009 : 2 technicians instead of 2.8 FTE operators



Operation group role



■ Operation

- Monitor the availability of all services (storage, batch, grid services...)
- Configure and regulate the system for optimizing the usage
- Coordinate actions during service interruptions
- Overall coordination of the « on call » service
- Other tasks: Monitoring and day-to-day management of tape libraries, Create and manage user accounts and AFS disk space...

■ Tight cooperation between operation and support groups



Services during Out of Office Hours



- On site night security guard from 6PM to 8AM and weekends
 - no computing actions : Alerting and Messaging
 - 1 on-duty person for facilities management
 - 1 on-duty computing engineer (evenings, weekends)
 - Corrective actions if possible (documentation, Training)
 - else call an expert ... if available
- Result is a « Best effort » coverage



Monitoring



- NAGIOS
- Monitoring tools for BQS's farms
- Dashboard of storage services
 - HPSS
 - dCache
 - Tape and drive monitoring with StorSentry



Monitoring with NAGIOS



- NAGIOS has replaced NGOP in March 2009 to monitor the services
 - Batch
 - Storage : HPSS, dCache, AFS, Xrootd, SRB
 - Grid : CE, SRM, TOP BDII
 - Databases
 - Others : Tape libraries, Saphir (privileges and location of services)

Current Network Status

Last Updated: Thu Oct 22 18:17:45 CEST 2009
Updated every 90 seconds
Nagios® 3.0.6 - www.nagios.org
logged in as: centre

[View Service Status Detail For All Host Groups](#)
[View Host Status Detail For All Host Groups](#)
[View Status Overview For All Host Groups](#)
[View Status Grid For All Host Groups](#)

Host Status Totals

Up	Down	Unreachable	Pending
91	0	0	0
All Problems		All Types	
0		91	

Service Status Totals

Ok	Warning	Unknown	Critical	Pending
176	0	0	1	0
All Problems		All Types		
1		177		

Status Summary For All Host Groups

Host Group	Host Status Summary	Service Status Summary
VOBOX (vobox-group)	3 UP	3 OK
ACSL5 (acsls-group)	2 UP	2 OK
AFS (afs-group)	22 UP	22 OK
BQS (bqs-group)	3 UP	14 OK
CE (ce-prod-group)	7 UP	21 OK
CRL (crl-group)	1 UP	1 CRITICAL : 1 Unhandled
DATABASE (db-group)	3 UP	10 OK
DCACHE (dcache-group)	2 UP	4 OK
DIVA (diva-group)	1 UP	5 OK
IRODS (irods-group)	2 UP	4 OK
MAGO (mago-group)	1 UP	No matching services
TEMPERATURE (misc machines)	1 UP	2 OK
NAGIOS (nagios-group)	1 UP	2 OK
RFIO (rfio-group)	3 UP	6 OK
SAPHIR (saphir-group)	1 UP	1 OK
SITE BDII (site-bdii-group)	2 UP	10 OK
SRB (srb-group)	3 UP	13 OK
SE SRM (srm-group)	2 UP	2 OK
TMS (tms-group)	1 UP	1 OK
TOP BDII (top-bdii-group)	1 UP	2 OK
XRootd (xrootd-group)	23 UP	46 OK

Nagios monitoring :
Web site interface



Monitoring tools for BQS farms (1)



Detect slow and stalled jobs, show overload of a service, problem with users' code or disfunctioning worker.

Users are automatically notified and their jobs are deleted after a delay.

For CMS's jobs, an alarm is sent to CMS support staff when the number of « slow » jobs is too high.

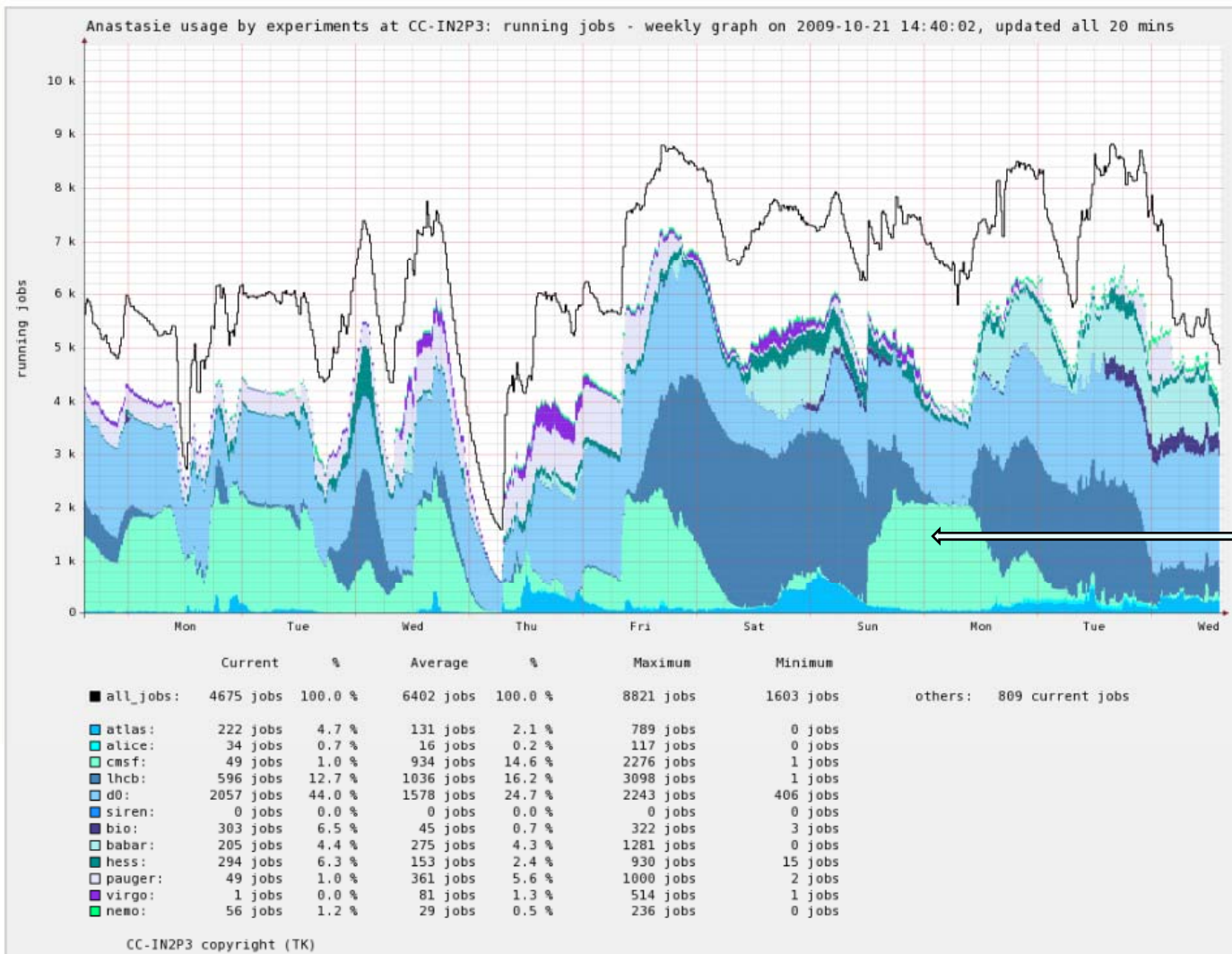
[Readme](#)

[About...](#)

133 Slow Jobs

UserName	JobName	GroupName	Worker	RequestedRsc	CurrentRsc	CPUTime	CPULimit	StartOfJob	EndOfJob
atlagrid	lcg1014163543-15346	atlas	ccwl1260	u_dcache_atlas	u_dcache_atlas	10012	600000	10/14/2009 16:38:13	-
augermgr	Imp-2009_10_15	pauger	ccwl9070	oracle	oracle	16646	36000	10/15/2009 01:01:54	10/18/2009 15:13
				u_srb_pauger	u_srb_pauger				
				u_sps_pauger	u_sps_pauger				
cms050	lcg1015184038-14679	cmsf	ccwl0479	u_dcache_cmsf	u_dcache_cmsf	1244	3801600	10/15/2009 18:42:30	-
aligid	lcg1015094950-21833	alice	ccwl1354	-	-	40	3801600	10/15/2009 09:50:51	-
aligid	lcg1014185140-27867	alice	ccwl1401	-	-	40	3801600	10/15/2009 07:06:55	-
aligid	lcg1014185118-29679	alice	ccwl1430	-	-	40	3801600	10/15/2009 07:06:55	-

Monitoring tools for BQS farms (2)



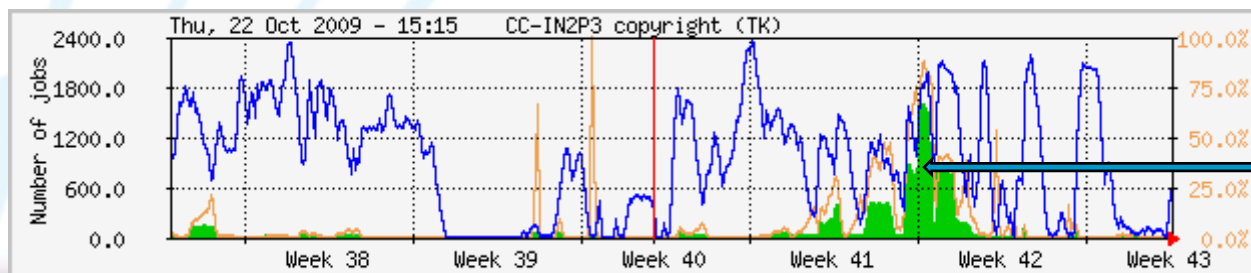
CMS jobs



Remarks (in October)



- Short jobs (CPU < 30s) :
 - T1 prod : 25%
 - T2 prod : 30%
- Use of memory (Long class)
 - 0.6% used more than 2GB
 - 8.5% used more than 1.5GB
- Slow jobs (all cms users)

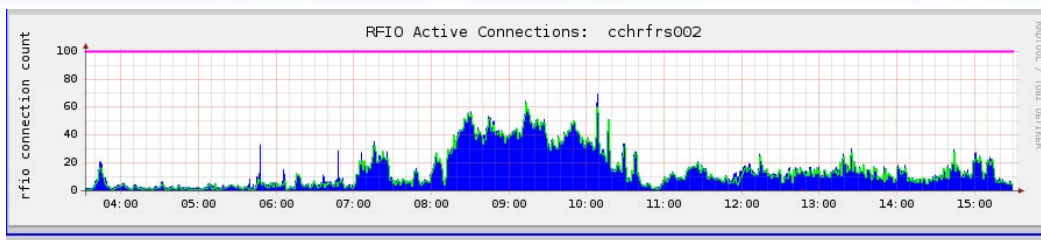


Blue = running jobs
Green = slow

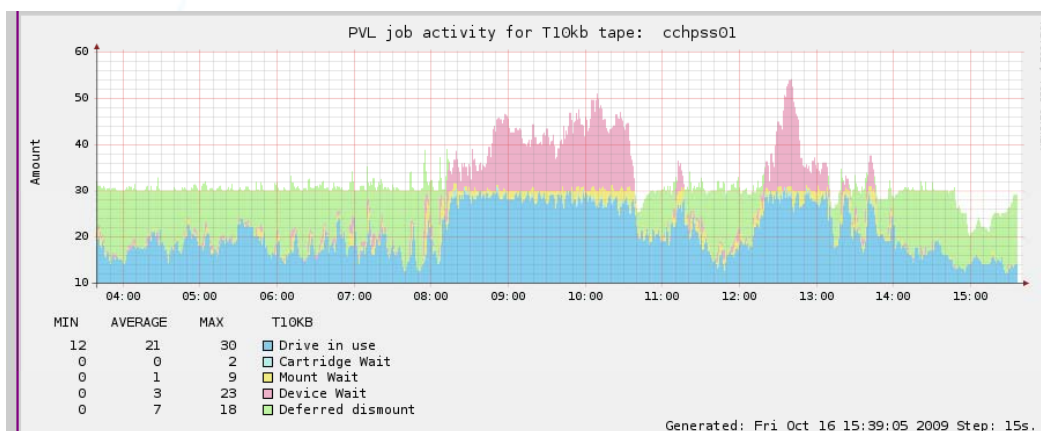
Problem with WMS

Storage services (1)

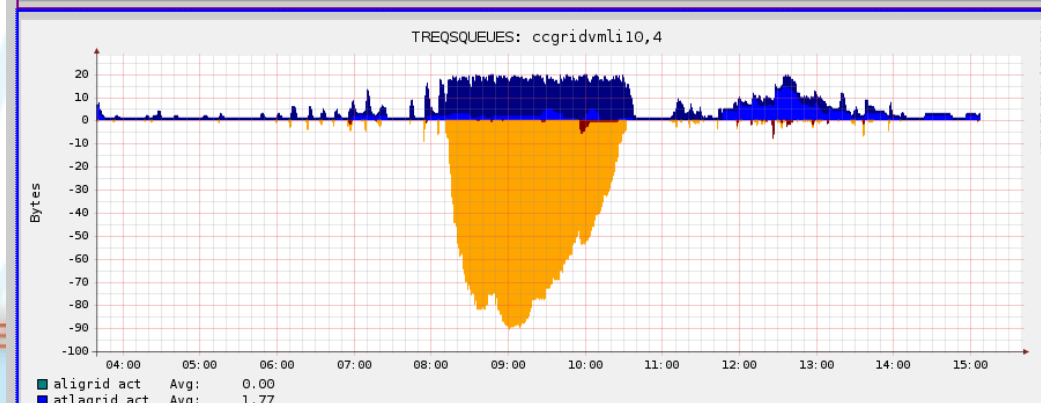
HPSS dashboard



Number of connexions
on rfio server
dedicated to dCache



T10K -B
tape drive
usage



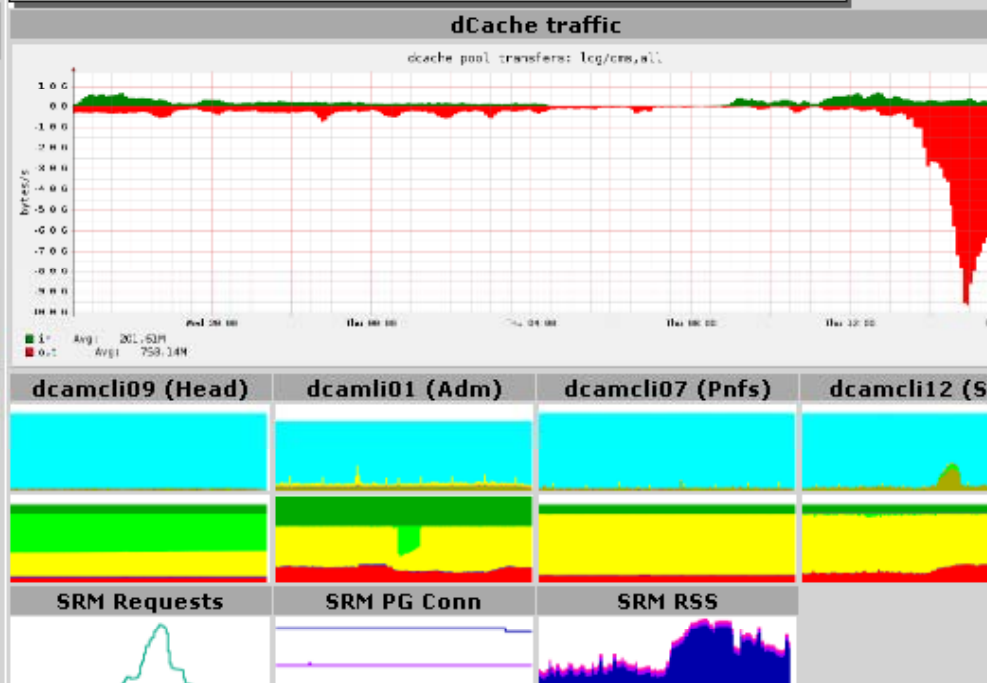
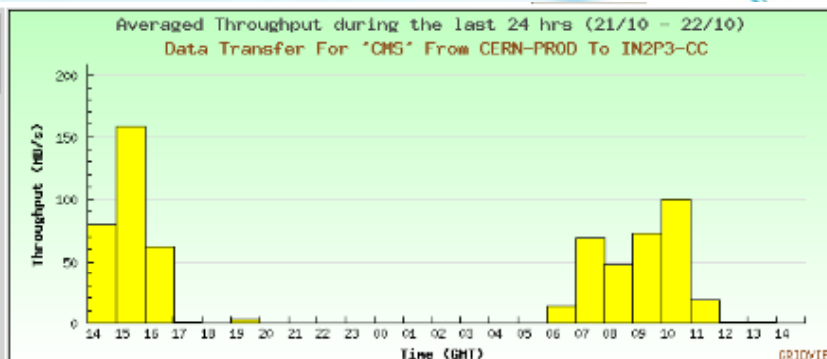
Treqs activity
for files on
TK10-B tapes

Storage services (2)

dCache Monitoring



p2p Active/Queued	Current activity Pool movers Active/Queued	HPSS Active/Queued	
p2p client:	cms-analysis:12/0	Staging:	
cms-xferout:5/0	pgroup-cms-prod:125/92	Migration:	
p2p server:			
dCache Billing (last 3 hours)		SRM Billing (last 3 hours)	
Action	Errors/total	Action	Errors/total
Pool movers	3698	copy	2
Door requests	2236	put	1149
	866	get	1099
		ls	204
			1203
Alarms			
Component			Alarm
SAM tests/cms/SRMv2			ok
BDII/glueservicestatus			running
cleaner backlog			8572
Stuck Hops (for > 6 hours)			82
old connections on: dcap-cdcacsn105Domain			3





Tapes drives monitoring



- Done with StorSentry
- Regular reports on tapes and drives quality
- Drives can be changed before damaging tapes
- Tapes can be changed before losing data



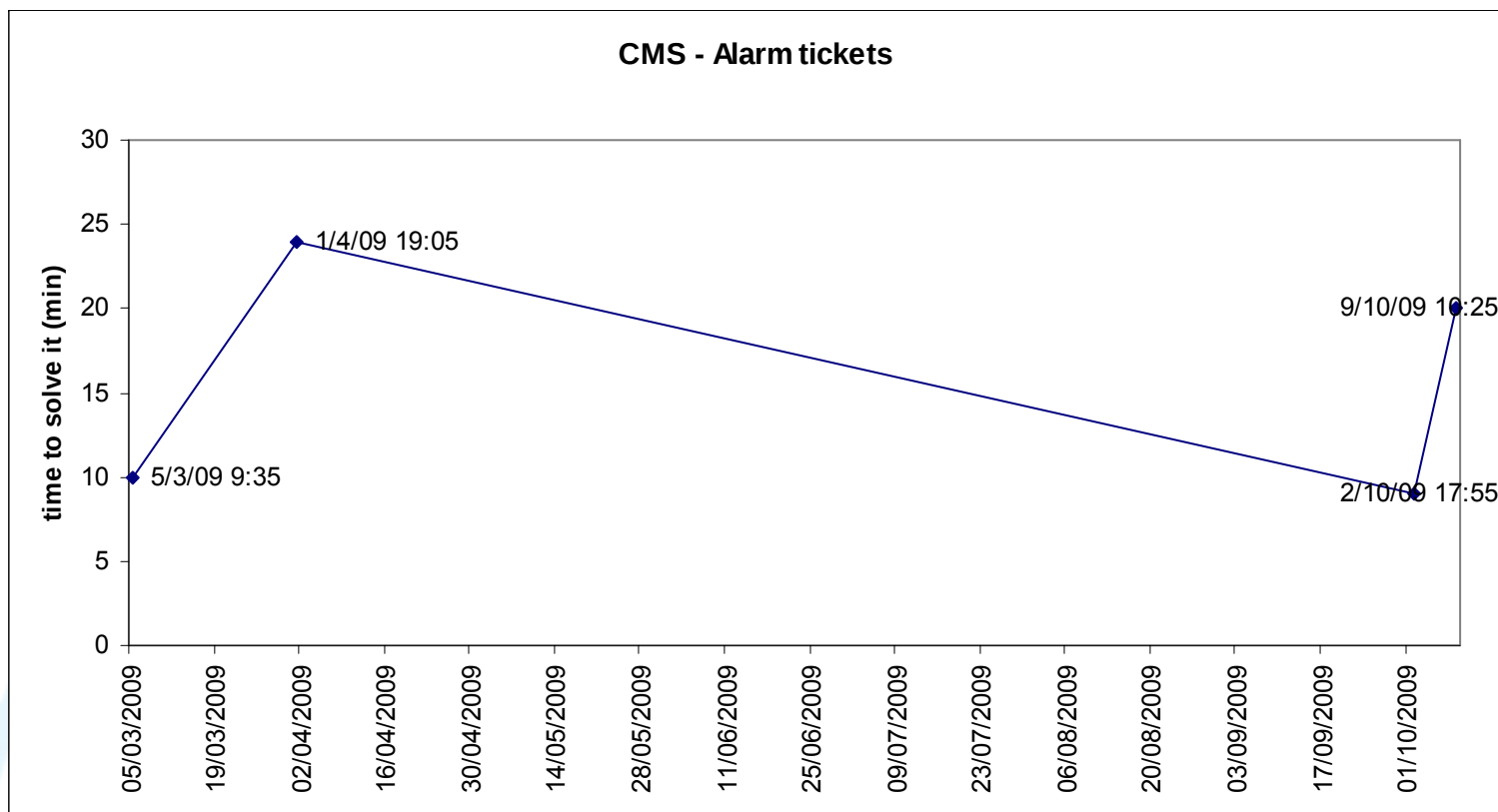
Tickets and Alarms



- Integration of GGUS tickets into the XHELP interface: all the treatment is done at XHELP level and synchronised with GGUS interface.
- Only one Team ticket for CMS in August
- Setup the procedure for fast answers to the LHC Alarm tickets
 - Ticket is acknowledged by a robot
 - Mail to Operation team and on-duty person
 - SMS to on-duty person
 - Answer to this mail automatically updates the ticket



CMS Alarm tickets (Tests)





Documentation



- Documentation system includes :
 - Documentation on services
 - Actions on services in case of interruptions
 - Communication procedures
 - All Grid services integrated to general documentation
- Each documentation has several levels :
 - Experts
 - Recipes for first level repair or restart of services by on-duty people
- Generalization of the use of electronic logbooks to follow up the activities

Electronic Logbooks

CCIN2P3

Support

support-general Logbook du support	75	jeu 10 sep 2009 10:58:49 CEST
support-atlas Logbook du support Atlas	97	mer 14 oct 2009 16:27:05 CEST
support-cms Logbook du support CMS	88	lun 19 oct 2009 16:07:01 CEST
support-alice Logbook du support Alice	14	mar 13 oct 2009 12:49:57 CEST
support-lhcb Logbook du support LHCb	51	lun 19 oct 2009 12:49:56 CEST
support-astro Logbook du support Astro	0	-

Exploitation

exploitation Logbook de l'exploitation	2597	lun 19 oct 2009 17:57:47 CEST
astreinte Logbook de l'astreinte	432	lun 19 oct 2009 06:53:35 CEST
evenements Événements généraux liés à l'Opération	54	mer 07 oct 2009 13:53:51 CEST
problemes-VO Logbook des problemes recuents des VO's LHC	115	lun 19 oct 2009 14:45:35 CEST

Services

HPSS Logbook d'HPSS	1276	lun 19 oct 2009 15:45:19 CEST
dCache Logbook de dCache	340	lun 19 oct 2009 17:44:30 CEST
logon Logbook de logon@cc	2568	lun 19 oct 2009 14:58:38 CEST
Analyse Logbook de la mise en place de la ferme d'analyse	28	mar 06 oct 2009 16:49:15 CEST

Suivi des opérations

operations-HPSS Logbook des opérations sur HPSS	23	ven 07 août 2009 14:22:49 CEST
operations-dCache Logbook des opérations sur dCache	132	lun 19 oct 2009 14:51:44 CEST
operations-LFC Logbook des opérations sur LFC	8	mar 28 jui 2009 16:06:50 CEST
operations-FTS Logbook des opérations sur FTS	35	mar 06 oct 2009 16:46:37 CEST
operations-CE Logbook des opérations sur les CEs	22	mar 08 sep 2009 12:05:10 CEST



Conclusion



- During the past year :
 - Big improvements on procedures and monitoring tools
- Future :
 - Cross-monitoring with SYMOD
 - Training for Operation people in Quality procedures
(*Information Technology Infrastructure Library*)