The WLCG services required by ALICE at the T2 sites: evolution of the CREAM-system, VOBOXES and monitoring

Dagmar Adamova,
Patricia Mendez Lorenzo,
Galina Shabratova

For ALICE

WLCG Workshop, London 2010

Outline

- WLCG T2 center in Prague
- History of our experience with WLCG and AliEn services:
  - Vobox
  - CE
  - Monitoring issues
- Concluding remarks

HEP Computing in Prague: the farm GOLIAS

- **A national computing center for processing data from various HEP experiments**
  - Located in the Institute of Physics in Prague
  - Officially started in 2004, basic infrastructure already in 2002
- **Certified as a Tier2 center of LHC Computing Grid** (praguelcg2) Collaboration with several Grid projects.
- **April 2008, WLCG MoU signed for Golias** (ALICE+ATLAS)
- **Excellent network connectivity:** Multiple dedicated 1Gb connections to collaborating institutes
- **Provides computing services** for ATLAS + ALICE, D0, Solid state physics, Auger, Star ...
- **Started in 2002 with:**
- 32 dual PIII 1.2GHz, 1 GB RAM, 18 GB SCSI HDD, 100 Mb/s Ethernet rack servers .... (29 of these decomissioned in 2009)
- Storage - disc array 1TB: HP server TC4100

History: 2002 -> 2010

Current numbers

- 1 batch system (torque + maui)
- moved from PBSPro due to licensing costs in 2009
- 2 main LHC VOs: Atlas, Alice
  - FNAL's D0 (dzero) user group
- Other EGEE VOs: Auger, Star
- 336 nodes
- 2630 cores (approx. 20500 HEP Spec ~= 5000kSI2K)
- 520TB on disk servers (DPM or NFSv3)

- ..... → featuring e.g.
  - 65x IBM iDataPlex x360 nodes
- - 2x Xeon E5520 with HT => 16 Cores
- 9x Altix XE 340 (twin nodes)
- - 2x Xeon E5520 without HT => 8 cores
- All mentioned water cooled

- Cooling problems long - term solution
- . In 2009 we installed new water cooling infrastructure (STULZ CLO 781A 2x 88 kW)
- . Since 2009 all new computing hardware must be equiped with water cooling ...

Software

- Installation via PXE, tftp, redhat kickstart
- . - No disk images
- . We use CFengine for node configuration
- . Various history graphs
- . - Ganglia + Munin + in house developed, rrd-based scripts
- . Nagios is the main alarm source
- . - various in house scripts, connected to syslog-ng
- . Users in LDAP

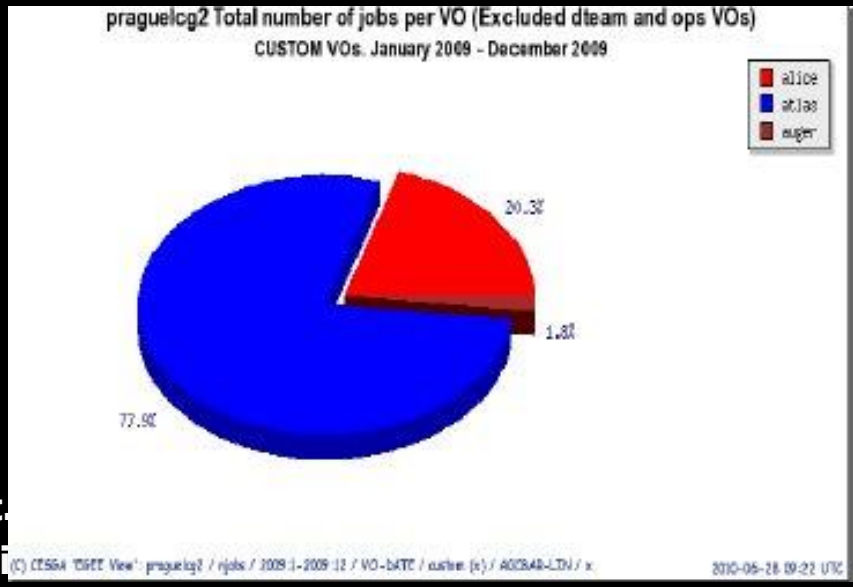
- conserver - IPMI consoles aggregator
- . IPMIView - supermicro graphical IPMI client
- . In house developed hardware database
- . MRTG+WeatherMap+FlowTracker
- . - network load + visualization
- . Many configuration files with node description
- . - We want to change node metadata at one place
- . - Project Deska

Grid services

- 60 WNs: SL4 32bit, used mainly by D0
- . All other WNs: SL5 64bit, gLite 3.2
- . Site BDII, lcg-CE, creamCE, 2x Alice vobox
  - running on virtual machines
- . Atlas Frontier Squid
- . MONBox
- . DPM SE (head on SL4, disk nodes SL5)

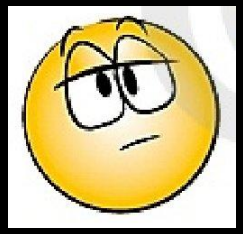
→ Alice SE (pure XROOTD) installed at a displace dept.

– 1Gb/s E2E link. 3 machines, SLC5 64bit, 60TB capaci



Our life with WLCG sevices ....

.... a sentimental journey ?



General remarks ....

- **Vobox:**
- **Job Agents are submitted from the local vobox**
- **Provides file system access to the experiment software area (shared with WNs)**
- **A WLCG service developed in 2006 (NOT woms aware)**
- **Beginning of 2008 - a new vobox version was deployed: gLite3.1 vobox (woms aware)**
- **2009 - gLite3.2-vobox deployment in connection with the creamCE strategy**

- **CE:**
- **ALICE deprecated the LCG-RB in November 2008**
- **Using WMS for production (required voms aware proxies)**
- **Most of 2009 - hybrid configuration:**
- **gLite-CE & creamCE (→ 2 independent voboxes)**

History - 2006
– the beginning of my lcgadmin career

- Jun: ALICE vobox set up by Patricia and I took over
- fixing problems with the vobox proxy (unwanted expirations)
- changing the RBs used by the JAs
- my first alien services set-up: problems with my registration in AliEn (aliprod, lcgadmin) - was fixed with the help of Predrag
- problems with our local LFC
- problem with my credentials ruining the JA submission via vobox
- problems with setting all the proxies corect
- ☺ in the end - successful participation in PDC'06

- first appearance of the phenomenon of alicesgm $HOME sweeping:
  - accidental start of an automatic grid cleanup process caused a cleaning also in /home/alicesgm
  - as a result, the AliEn services on the vobox crashed

- Nov: First problems with the fair-share of the local batch system (then, PBSPro)

History - 2007

- started with alien central services crash, but fixed immediately.
- problems with the default vobox-proxy time length: too short in Prague
- problem with the proxy registration in the vobox
- in general, during the PDC'07 lots of manual work with the Alice jobs:
  - problems with job submission due to malfunctions of the default RB – was changed occasionally
- as a result, the failover submission was configured at the sites. Still, sometimes the list of 3 default RBs had to be changed also

- → Some monitoring issues:
- Apr: GoogleMap incorporated in the MonALISA for the Alice sites status map
- Jun: site efficiency monitoring - dashboard announced by Pablo
- Aug: the SAM implementation for the alarm system ready

History - 2008

- started with jobs crashing right after landing in Prague: a local CE failure: was constatnly overloaded, transfer onto a stronger machine helped
- migration to glite3.1 vobox through the first months of the year - problems with installation on 64bit machines
- Feb: gLite3.1 vobox on SLC4
- newly required specificication of the voms role during the registration of the vobox proxy: took some time to get it work (the voms extension issue)
- Mar: problems with the proxy renewal service: identified a wrong content of the file /opt/glite/etc/vomses/alice-voms.cern.ch
- Apr: upgrade of the local CE serving ALICE to lcg-CE 3.1 and shortly after transfer to a new and stronger machine
- Jul: proxy repository in Prague corrupted – fixed - all renewed

History – 2008 (2)

- Sep: included in SAM vobox test failure automatic notification
- repeating problems with job submission through RB's: in October the site re-configured for the WMS submission
- Oct: problems with the voms extension renewal, solved promptly by Patricia

- → Some monitoring issues:
- Apr: farmnagios discovered a full /tmp on the Alice vobox
- May: vobox memory usage monitoring set on
- Aug: new generic UI to visualize the current vobox status through dashboard

History - 2009

- **Big events:**
  - **installation and tuning of the creamCE**
  - **starting the second (cream) vobox**

- **hybrid state for a part o the year:**
- **- glite vobox and WNs: 2.6.9-89.0.9.EL.cernsmp**
- **- cream vobox: 2.6.9-78.0.1.EL**
- **- cream WNs: 2.6.18-128.7.1.el5**

- **Apr installed the first version of our creamCE and a GridFTP server. Plans for the new vobox affiliated to the creamCE and the Torque batch.**
- **Apr installation of a second vobox, submitting JAs directly to the new creamCE. - on 64bit, connected with the local Torque cluster (WNs on 64bit)**

- **May the Prague creamCE put into production**

History – 2009 (2)

- May, Jun solving problems arising from the existence of 2 voboxes running on different OS and sending JAs to different batch systems: PBS on 32bit and Torque on 64bit. Solved by running 2 separate systems with separate SW areas and 2 independent PackMans.

- Jun 44444 error for gliteCE/cluster: publication problem in local BDII.
- - Fixed locally.

- Jul another problem with creamCE: 'related to glexec', got fixed by cleaning in the directory
- /opt/glite/var/cream_sandbox/alicesgm on the creamCE
- Jul update of the package glite-security-lcmaps:
- → ruine mapping and the job submission via creamCE was failing. Re-configuration with yaim fixed that

- Aug announcement of the 1st SL5 vobox

History – 2009 (3)

- Aug/Sept Alien reporting lots of running (but in fact already finished) jobs. The cause was in the creamCE reporting finished as running. Fixed with the developers.
- Aug again the problem with creamCE: 'related to glexec', got fixed by cleaning in the directory
- /opt/glite/var/cream_sandbox/alicesgm on the creamCE

- Sep: a problem with discrepancies between the number or R/W jobs reported by the creamCE, local BDII and AliEn (fixed by AliEn experts)

- Nov: Problems with the job submission via creamCE after its upgrade. Fixed in collaboration with the developpers

History – 2009 (4)

- Dec a new machine (Intel Xeon E5420) configured as an Alice cream vobox on gLite3.2 and SL5.4/64bit. Intended mainly as a test, not to ruin the production with an untested system
- → running with 2 cream oboxes
- Dec 44444 error for creamCE - creamCE publication problem in BDII. Fixed locally.

- Dec the old glite vobox disabled. Since then, alice jobs submitted only to the creamCE.

History - 2010

- again, the new year started with problems, this time with the local creamCE: solved quickly

- Jan: the new sl5 cream vobox registered in the myproxy list of trusted nodes
- Feb the old sl4 cream vobox only as a backup
- Mar the backup cream vobox re-installed on gLite3.2 and SL5.4/64bit

- May creamCE 1.6 / gLite 3.2/sl5 64bit installed in Prague
- --- in the beginning, the info-system of the new CE was not working properly – fixed locally

- → ☺ we were the first T2 where CREAM1.6 was tested with CERN (Patricia) credentials

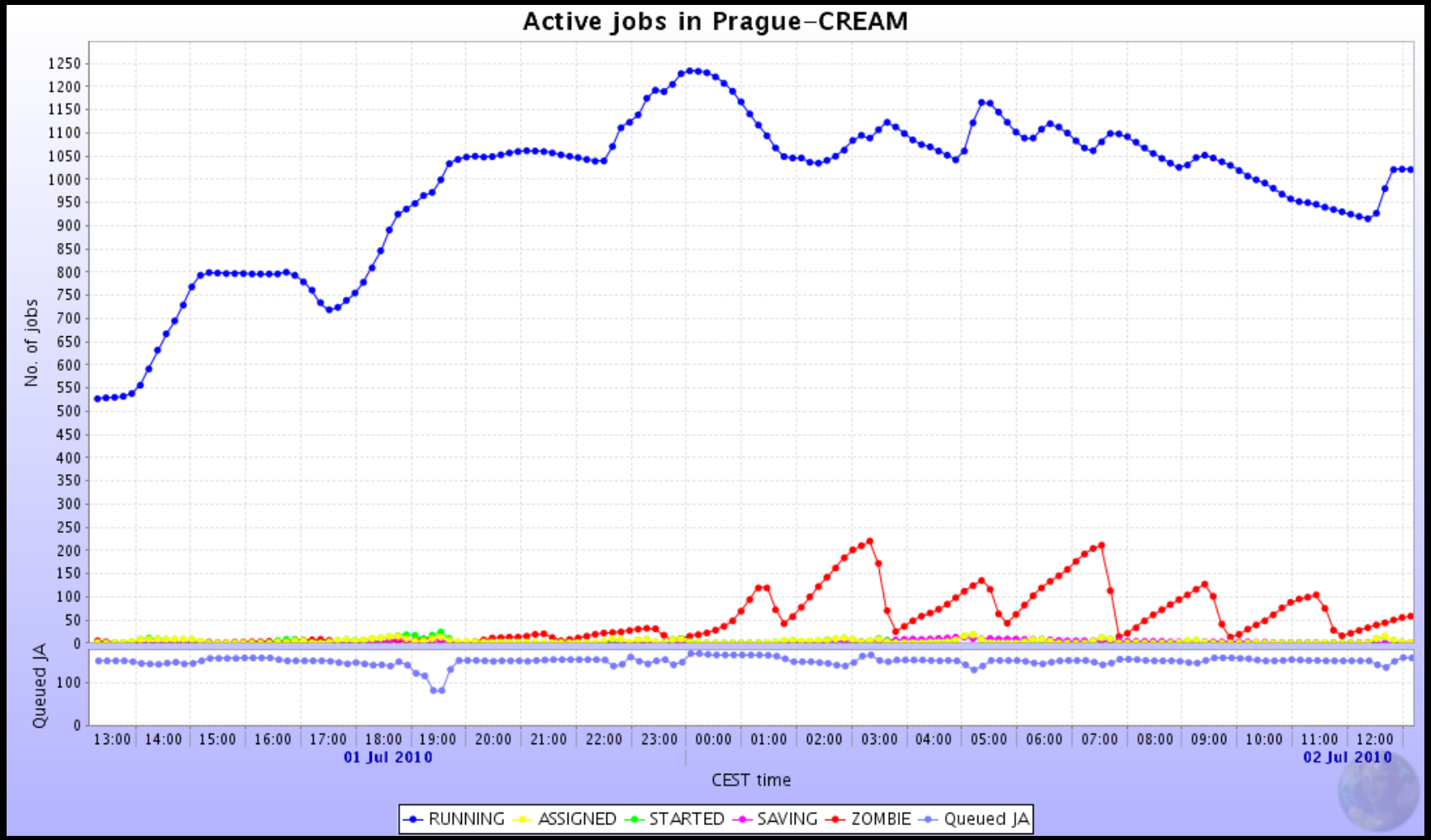
- ☺ Now the job submission and processing works OK. At the moment, we have 1280 running and 150 queued jobs
-

Monitoring issues

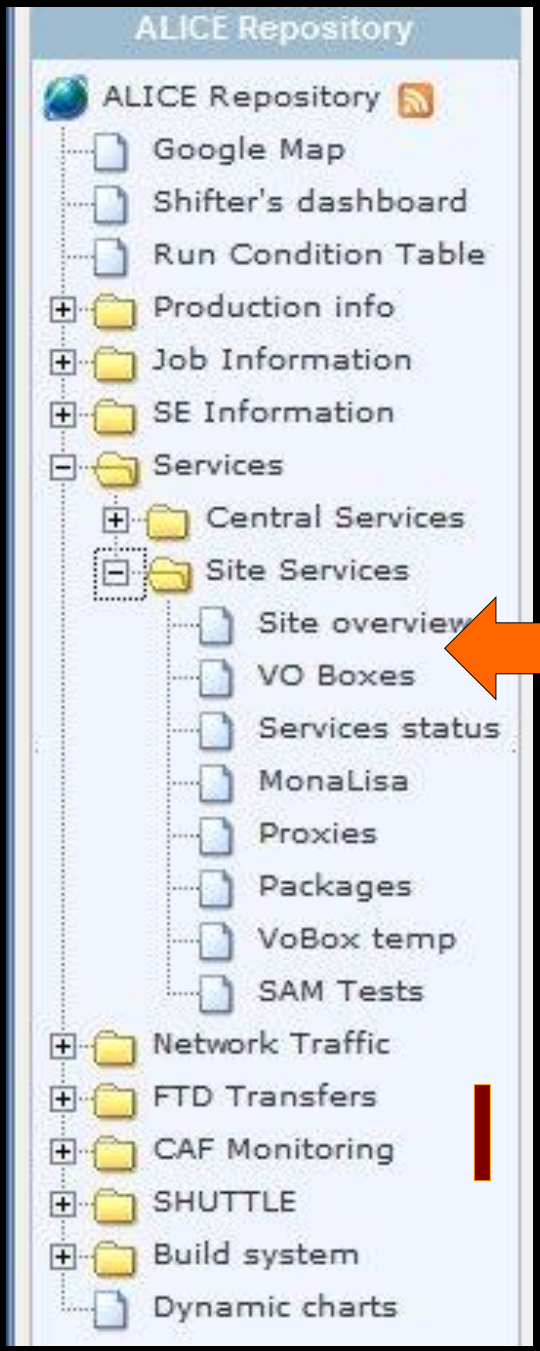
- Although being aware of the existence and occasionally using the WLCG monitoring tools (Dashboard, Gstat, Nagios .....)

- ..... the everyday life of a site Alice liaison officer would be impossible and unthinkable without the Alice MonALISA monitoring system
  - Constantly developping over the years mentioned
  - Tailored to cover all the Alice specific needs
  - Covering everything from the distributed sites hardware or the proxies renewal status until the RAW data QA Analysis trains status, end users resources quotas or the Shuttle preprocessors
  - All the charts refreshing automatically
  - Everything accessible from one basic menu

Everyday's first check .....



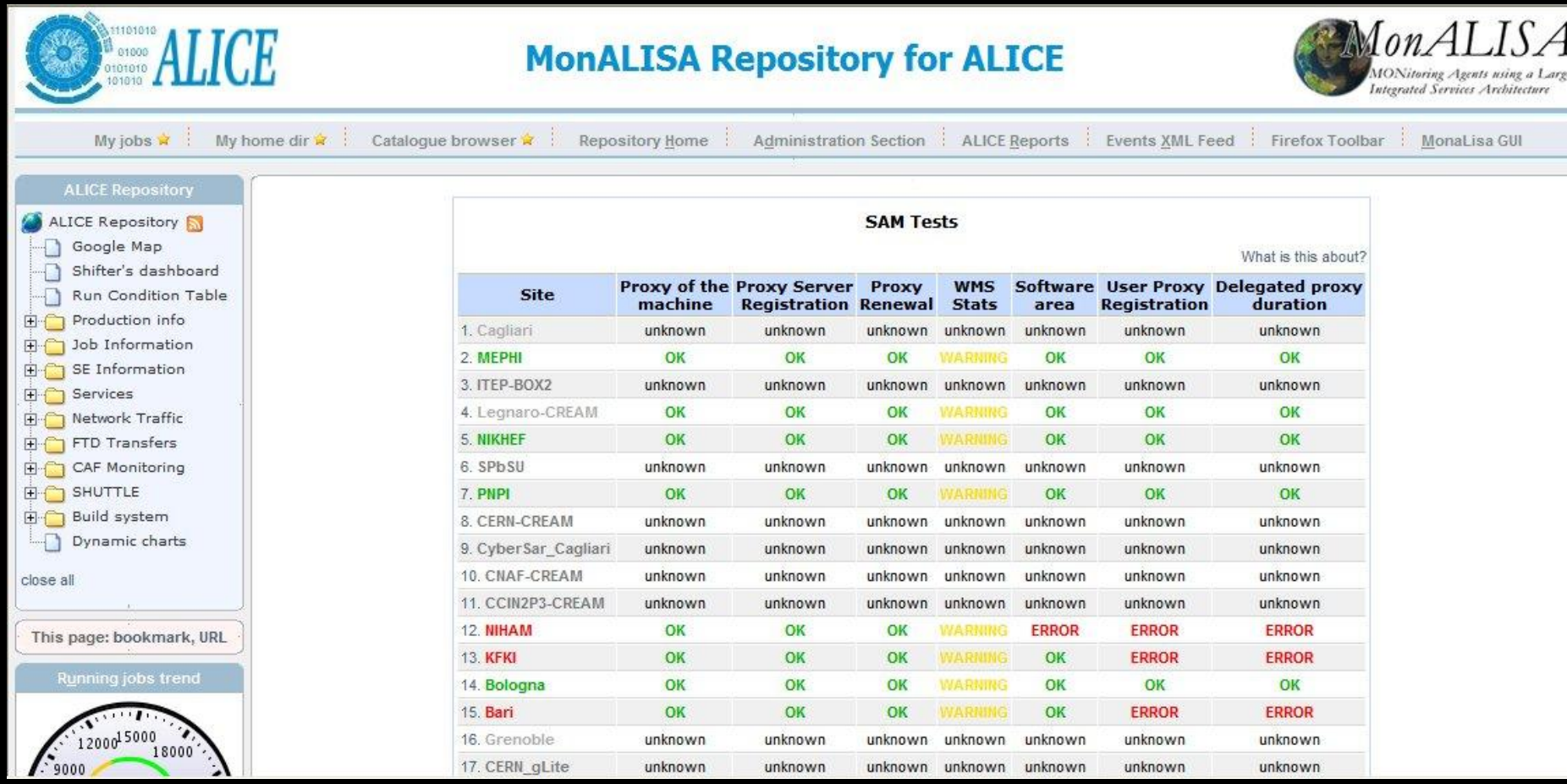
Examples of charts: the main menu



Sites services

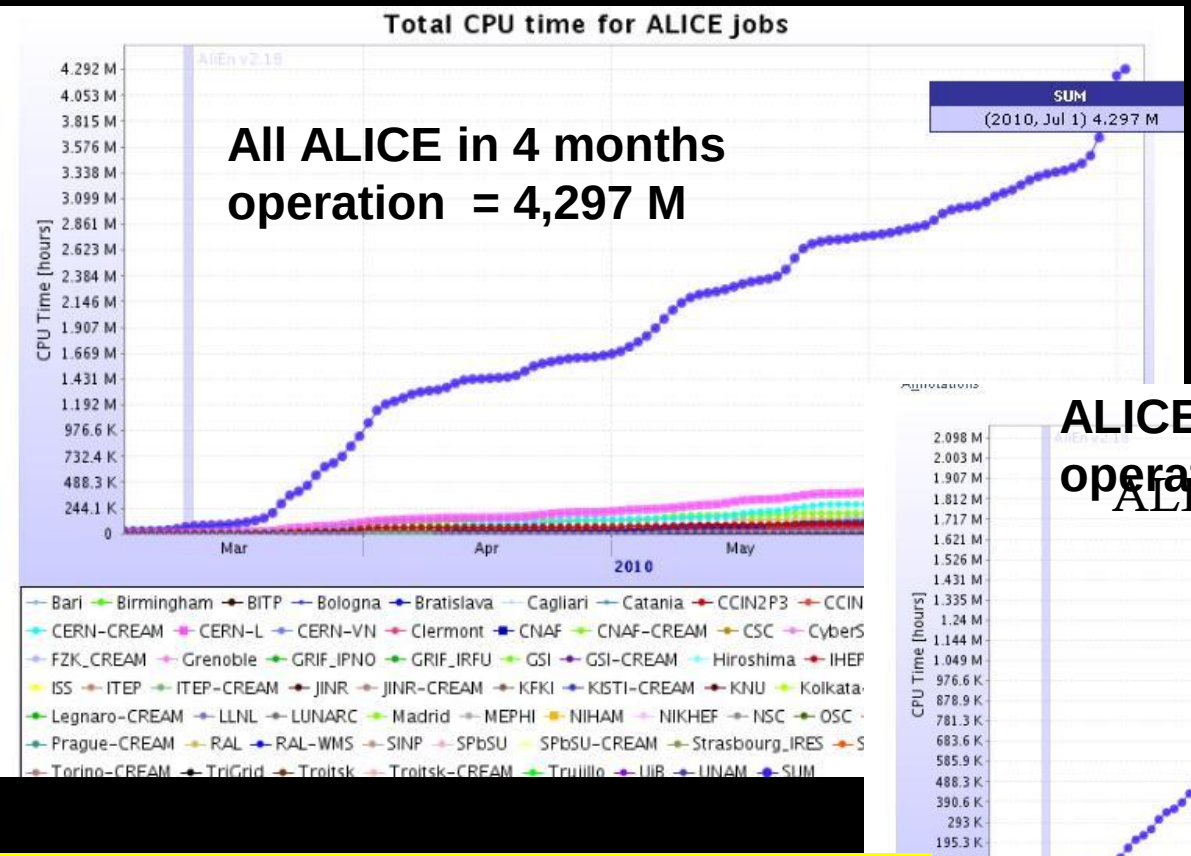
Running trenss

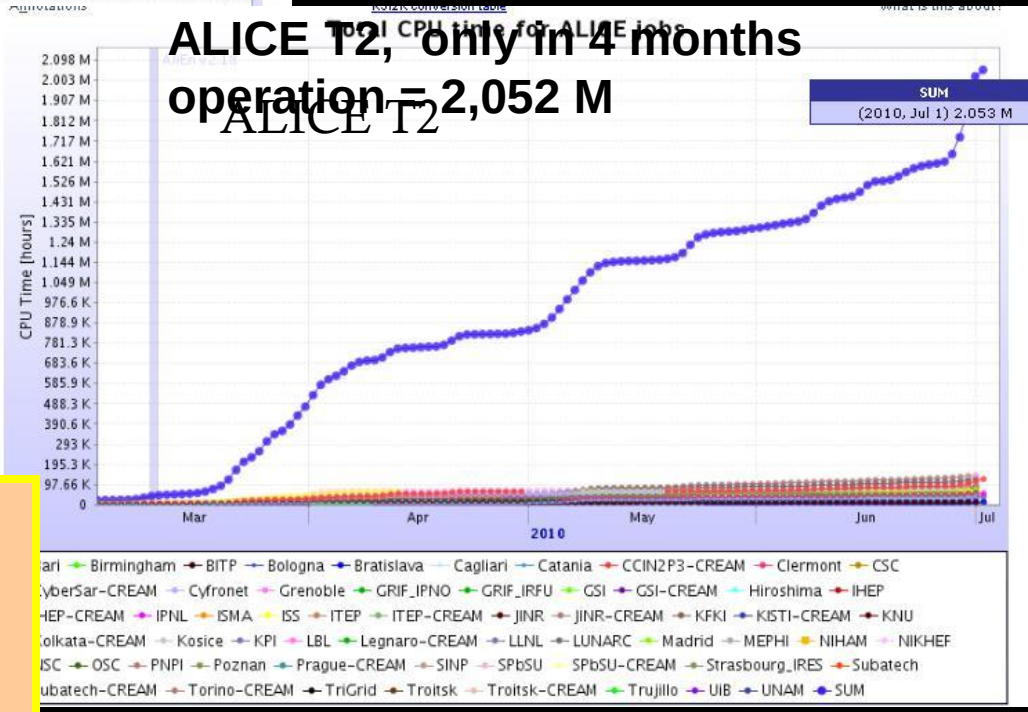
Examples of charts: SAM tests



Site	Proxy of the machine	Proxy Server Registration	Proxy Renewal	WMS Stats	Software area	User Proxy Registration	Delegated proxy duration
1. Cagliari	unknown	unknown	unknown	unknown	unknown	unknown	unknown
2. MEPHI	OK	OK	OK	WARNING	OK	OK	OK
3. ITEP-BOX2	unknown	unknown	unknown	unknown	unknown	unknown	unknown
4. Legnaro-CREAM	OK	OK	OK	WARNING	OK	OK	OK
5. NIKHEF	OK	OK	OK	WARNING	OK	OK	OK
6. SPbSU	unknown	unknown	unknown	unknown	unknown	unknown	unknown
7. PNPI	OK	OK	OK	WARNING	OK	OK	OK
8. CERN-CREAM	unknown	unknown	unknown	unknown	unknown	unknown	unknown
9. CyberSar_Cagliari	unknown	unknown	unknown	unknown	unknown	unknown	unknown
10. CNAF-CREAM	unknown	unknown	unknown	unknown	unknown	unknown	unknown
11. CCIN2P3-CREAM	unknown	unknown	unknown	unknown	unknown	unknown	unknown
12. NIHAM	OK	OK	OK	WARNING	ERROR	ERROR	ERROR
13. KFKI	OK	OK	OK	WARNING	OK	ERROR	ERROR
14. Bologna	OK	OK	OK	WARNING	OK	OK	OK
15. Bari	OK	OK	OK	WARNING	OK	ERROR	ERROR
16. Grenoble	unknown	unknown	unknown	unknown	unknown	unknown	unknown
17. CERN_gLite	unknown	unknown	unknown	unknown	unknown	unknown	unknown

CPU resources usage by T2s





T2s in 4 months were using

2,052/4,297 = 47.7% of the whole ALICE CPU time

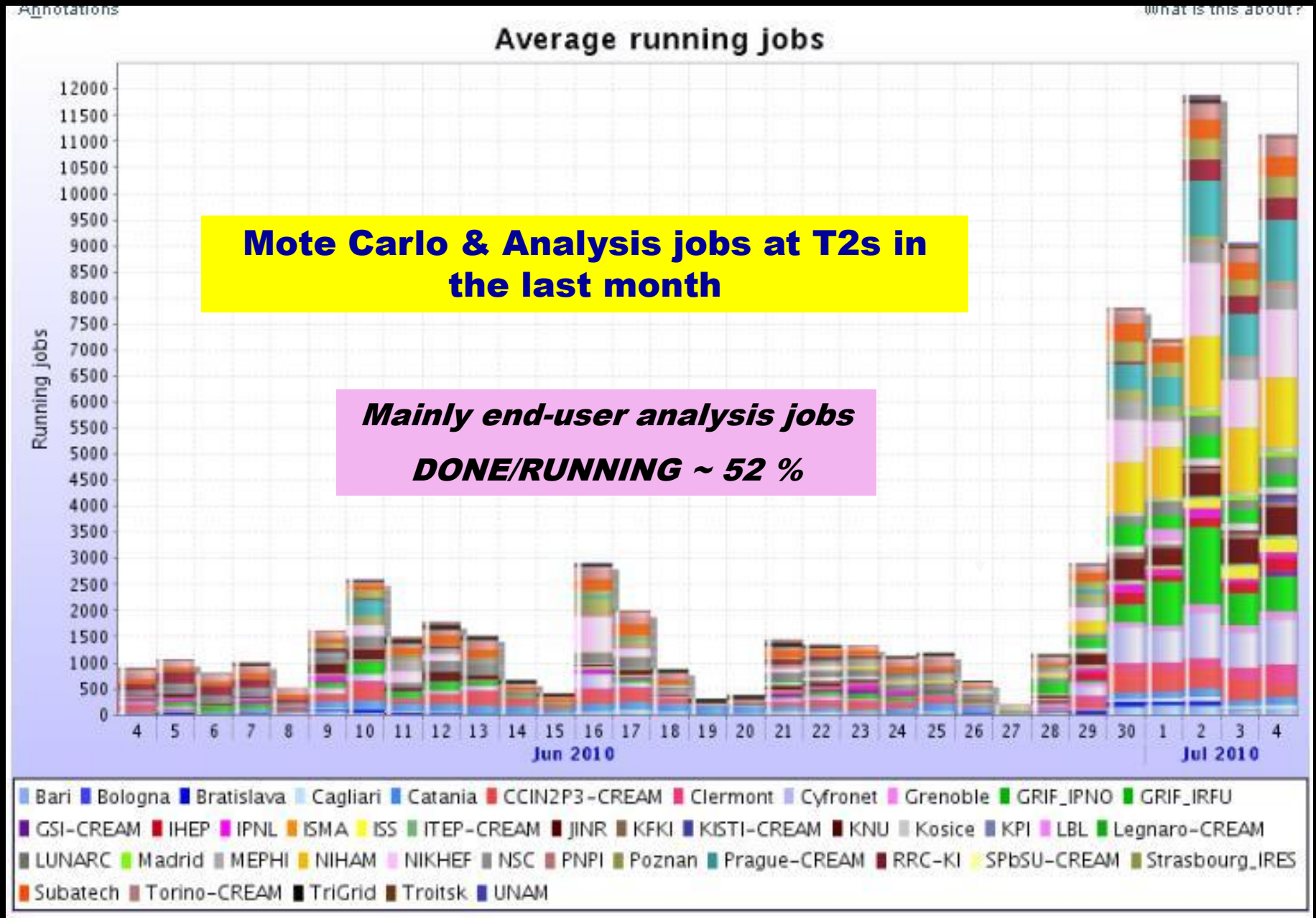
DISK resources usage by T2s

From 7448.6 TB allocated at the whole ALICE,
3558.9 TB allocated at T2s,
Used till today 3127.3 TB by the whole ALICE,
1062 TB used at T2s

So 47% of all Disk resources are allocated at T2s.
30% of these are already used

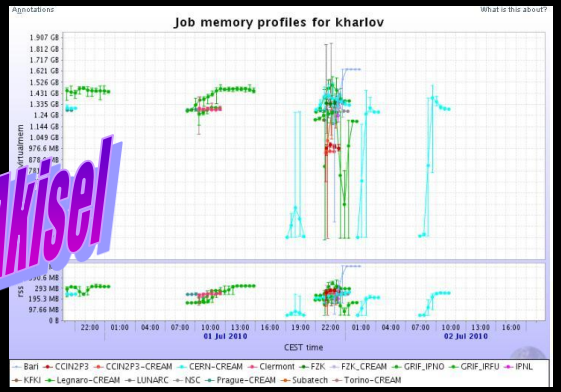
18. GSI - SE									-	-	to
19. Hiroshima - SE									2 TB	75.26 TB	20100510-1509_dbg
20. IHEP - SE									5 TB	34.48 TB	20100510-1509_dbg
21. IPNL - SE									7 TB	10.6 TB	20100510-1509_dbg
22. ISS - FILE									6 TB	114.5 TB	20100510-1509_dbg
23. ITEP - SE									2 TB	39.85 TB	20100510-1509_dbg
24. JINR - SE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	36%	1,053,620	File	149.1 TB	20.78 TB	128.3 TB	20100510-1509_dbg
25. KFKI - SE	ALICE::KFKI::SE	39.34 TB	6.869 TB	32.47 TB	17.46%	153,769	File	36.38 TB	11.37 TB	25 TB	20100510-1509_dbg
26. KISTI - SE	ALICE::KISTI::SE	28 TB	7.948 TB	20.05 TB	28.38%	162,594	File	28.04 TB	9.356 TB	18.69 TB	20100510-1509_dbg
27. Kolkata - SE	ALICE::Kolkata::SE	73.24 TB	720.7 GB	72.54 TB	0.961%	30,296	File	-	-	-	
28. Kosice - SE	ALICE::Kosice::SE	21.48 TB	3.729 TB	17.76 TB	17.36%	77,961	File	20.92 TB	9.881 TB	11.04 TB	20100510-1509_dbg
29. LBL - SE	ALICE::LBL::SE	20 TB	1.291 TB	18.71 TB	6.456%	15,621	File	33.64 TB	3.503 TB	30.14 TB	20100510-1509_dbg
30. Legnaro - SE	ALICE::Legnaro::SE	45 TB	25.71 TB	19.29 TB	57.13%	1,146,408	File	42.75 TB	37.85 TB	4.902 TB	20100510-1509_dbg
31. LLNL - SE	ALICE::LLNL::SE	97.66 TB	0.507 GB	97.66 TB	0.001%	78	File	-	-	-	
32. Madrid - DPM	ALICE::Madrid::DPM	44 TB	59.06 GB	43.94 TB	0.131%	3,156	SRM	-	-	-	
33. MEPHI - SE	ALICE::MEPHI::SE	18.2 TB	1.894 TB	16.31 TB	10.41%	26,093	File	18.19 TB	14.99 TB	3.202 TB	20100510-1509_dbg
34. NDGF - DCACHE	ALICE::NDGF::DCACHE	204.6 TB	56.42 TB	148.2 TB	27.57%	757,378	srm	-	-	-	
35. NIHAM - FILE	ALICE::NIHAM::FILE	855 TB	50.11 TB	804.9 TB	5.861%	6,454,079	File	854.9 TB	42.77 TB	812.1 TB	20100510-1509
36. PNPI - SE	ALICE::PNPI::SE	38.2 TB	1.452 TB	36.75 TB	3.801%	76,516	File	38.2 TB	1.82 TB	36.38 TB	20100115.1117_dbg
37. Poznan - SE	ALICE::Poznan::SE	30 TB	1.921 TB	28.08 TB	6.403%	43,676	File	29.53 TB	3.526 TB	26 TB	20100510-1509_dbg
38. Prague - SE	ALICE::Prague::SE	59.03 TB	9.015 TB	50.02 TB	15.27%	130,345	File	59.03 TB	10.9 TB	48.14 TB	20100510-1509_dbg
39. RRC-KI - SE	ALICE::RRC-KI::SE	202.6 TB	18.79 TB	183.8 TB	9.275%	552,387	File	182.3 TB	42.02 TB	140.3 TB	20100510-1509_dbg
40. SaoPaulo - DPM	ALICE::SaoPaulo::DPM	0	0.36 GB	-	-	37	SRM	-	-	-	
41. SARA - DCACHE	ALICE::SARA::DCACHE	435 TB	237.3 GB	434.8 TB	0.053%	16,187	srm	-	-	-	
42. SPbSU - SE	ALICE::SPbSU::SE	22.66 TB	1.101 TB	21.56 TB	4.859%	62,649	File	-	-	-	
43. Strasbourg_IRES - SE	ALICE::Strasbourg_IRES::SE	10.9 TB	1.931 TB	8.969 TB	17.71%	49,939	File	3.638 TB	3.567 TB	73.71 GB	20100510-1509_dbg
44. Subatech - SE	ALICE::Subatech::SE	195 TB	79.69 TB	115.3 TB	40.86%	2,549,456	File	270.3 TB	138.8 TB	131.5 TB	20100510-1509_dbg
45. Torino - SE	ALICE::Torino::SE	65.4 TB	43.55 TB	21.85 TB	66.59%	1,384,682	File	206.7 TB	149.4 TB	57.29 TB	20100510-1509_dbg
46. Troitsk - SE	ALICE::Troitsk::SE	20 TB	7.695 TB	12.31 TB	38.47%	137,895	File	0	0	0	20100115.1117_dbg
47. Trujillo - SE	ALICE::Trujillo::SE	48.4 TB	413.3 GB	48 TB	0.834%	8,424	File	46.6 TB	657.2 GB	45.95 TB	20100510-1509_dbg

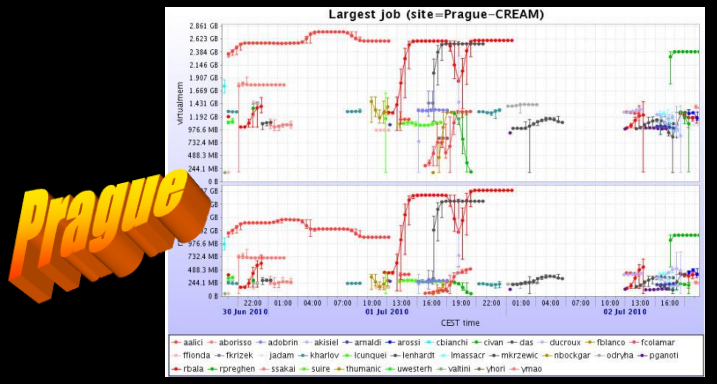
Efficiency of jobs processing at T2s



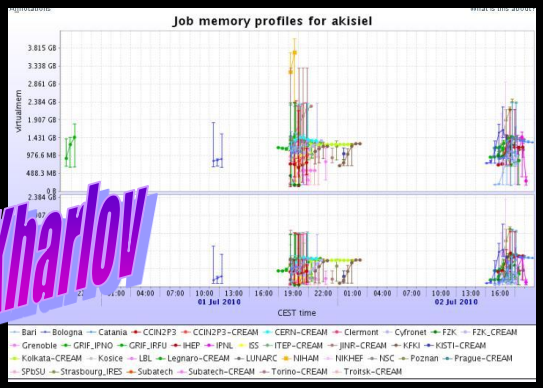
Control of Job efficiency at T2s: *virtual memory distributions*

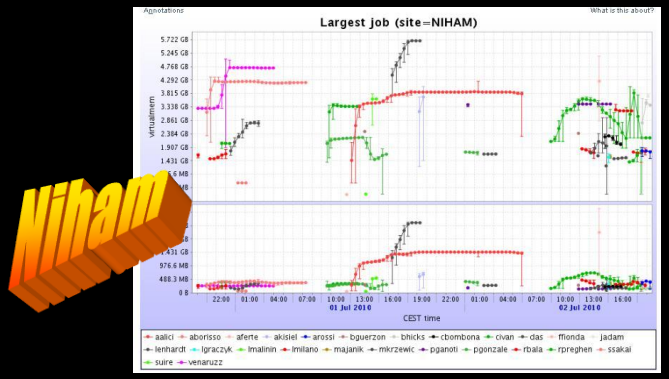
users

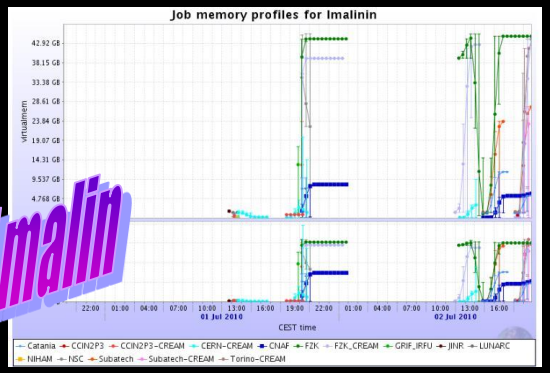


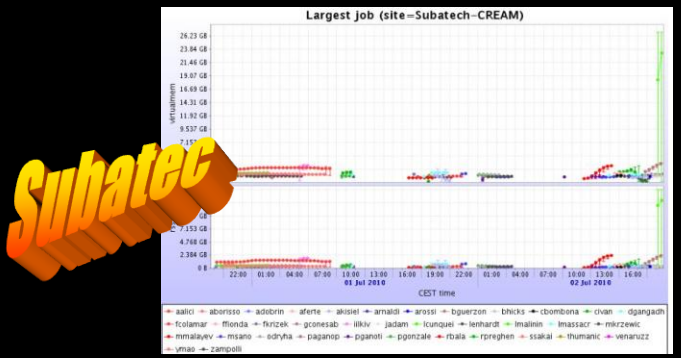


sites









A new item: monitoring with Nagios

- Migration of the sites current SAM infrastructure for the WLCG services monitoring and tracking to Nagios
- ALICE has defined specific tests for the vobox only
- Tests for the CE and the creamCE sensors also running with ALICE credentials

Concluding remarks

- In Prague, the cream services used almost exclusively by Alice
- There were no basic problems with the installation and performance of the version cream 1.6/gLite 3.2 @ SL5/64bit
- The older version: problems with the BLParser
- Some sites are running 2 creamCEs to be able to perform rolling upgrades and strengthen the availability of the CE service
- Services vobox and CE (glite, cream) frequently run on virtual machines
- The continuing development and growing of the MonALISA repository brings sometimes some discrepancies in the chats. But the contact with the developer is very good and the fixes come quickly

- In general, ALICE approach is to lighten the site requirements in terms of services and aims at their high performance

- Sites are supported by a team of experts at any time
- This approach has demonstrated its scalability
- Same team since years with a clear increase of the sites entering in the production

- ***Many thanks to Patricia, Latchezar, Pablo, Costin, Fabrizio and the others!***