

ISS Computing Facility

1

Status report on clustering and Grid services

Ionel STAN

Institute of Space Science

P.O. Box: MG-23, RO 077125, Bucharest-Magurele ROMANIA

<http://www.space-science.ro>

Fifth Romania Tier 2 Federation “Grid, Cloud & High Performance Computing Science”
25-27 October 2012

ISS Computing

2



Cluster	Number of servers	Logical CPUs
ISS-ALICE	40	388
RO-13-ISS	4	32
PlanckGrid	14	112
RoSpaceGrid	26	208
Total	88	740

Fifth Romania Tier 2 Federation "Grid, Cloud & High Performance Computing Science
25-27 October 2012

Rocks Clusters

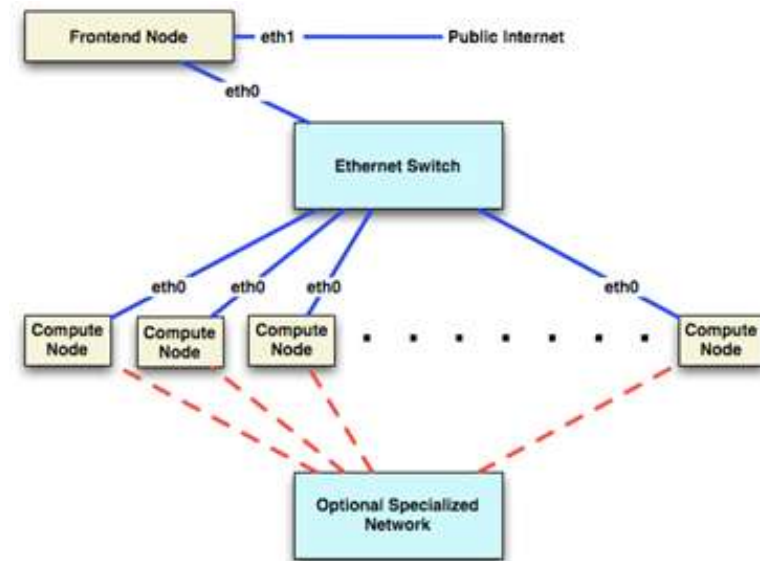
- was chosen due to ease of deployment, management, maintenance and flexibility , facile extensibility of software packages via software entities known as “Rolls”.
- is more than a middleware. It is a system composed of open source operating system that is based on a Red Hat flavor - CentOS (which comes with a modified interface for Anaconda for setup of cluster parameters and services) and a database with a corresponding set of scripts that simplify the process of node deployment, authorization of services and cluster monitoring.
- Installation can be customized by using so-called software Rolls. These Rolls expand system capacity by integrating effortlessly and automatically software packages and management mechanisms used by the Rocks administration tools, simplifying installation and configuration over a large number of computing nodes. Rocks Clusters brings a large number of rolls, such as: Torque, SGE, Condor, Ganglia, Java, HPC, etc.

CLUSTERING TECHNOLOGIES IN ISS DATA CENTER

4

Rocks Clusters

- used in more than 1800 clusters worldwide
- the model used is the server (frontend) - client model in which the central server will contain all necessary central clustering services that will serve a number of clients types like the computing nodes (main computing resources) and the login nodes – the nodes for user interaction with the cluster.
- On the computing nodes the first Ethernet interface will be used for the private network but Rocks have the possibility of setting up over a large cluster of nodes of other various private network used for dedicated usage like high speed networking or special storage sharing (10 Gbps, Infiniband).
- Rocks use the well-known and established tools for clustering, “Maui” and “TORQUE”.
- “Maui” - We use a priority based node allocating policy that would give priority to the most unoccupied servers
- “TORQUE” - We use several queues for the jobs to be submitted with restrictions in place for memory consumption (3584 MB resident memory and 4096B virtual memory) and wall time (36 hours).



GRID SERVICES AND MIDDLEWARE DEPLOYED IN DATA CENTER

AliEn – Alice Environment

- is a lightweight Grid framework which is built around Open Source components using the Web Services model. It has been initially developed by the ALICE collaboration (ALICE Offline Project) as a production environment for the simulation, reconstruction, and analysis of physics data in a distributed way.

EMI – European Middleware Initiative

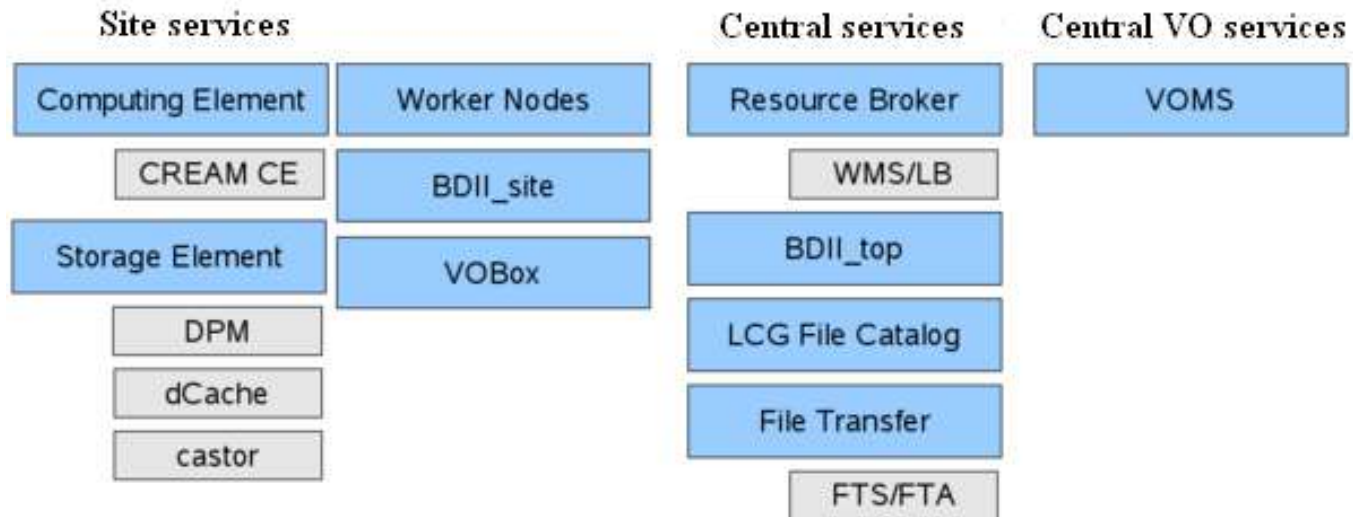
- is a Service Oriented Grid middleware providing services for managing distributed computing and storage resources and the required security, auditing and information services. Target server platform is Red Hat Linux or any binary compatible distribution, such as SL and CentOS
- EMI is a close collaboration of three major middleware providers, ARC, gLite and UNICORE, and other specialised software providers like dCache.
- The products, managed in the past by these separate providers, and now developed, built and tested in collaboration, are for deployment in EGI as part of the Unified Middleware Distribution - UMD

GRID SERVICES AND MIDDLEWARE DEPLOYED IN DATA CENTER

6

EMI – European Middleware Initiative

- One of the central services in the EMI middleware is the information system consisting of BDII_site and BDII_top services, all available grid services and their status being listed in this system



GRID SERVICES AND MIDDLEWARE DEPLOYED IN DATA CENTER

EMI – European Middleware Initiative

- April 2012: update our site from gLite 3.2 to EMI 1 (Kebnekaise)
- July 2012: update our site from EMI 1 to EMI 2 (Matterhorn)

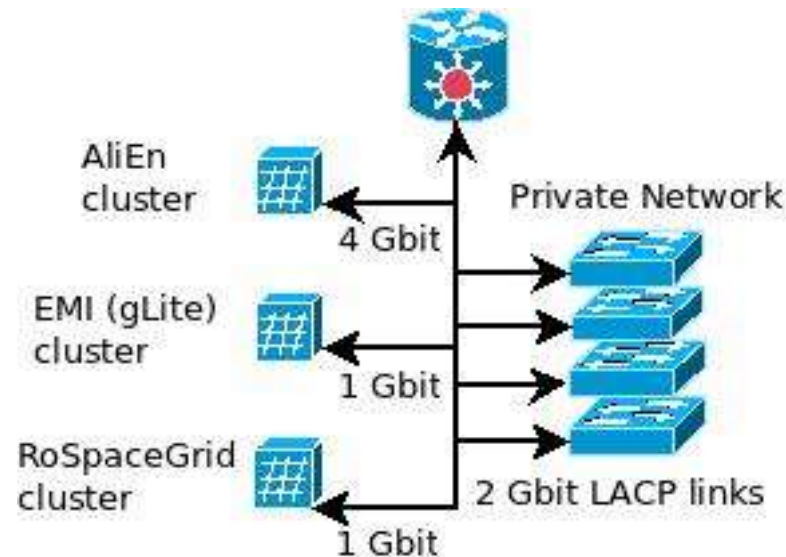
XRootD storage

- The XRootD project aims at giving high performance, scalable fault tolerant access to data repositories of many kinds. The typical usage is to give access to file-based ones. It is based on a scalable architecture, a communication protocol, and a set of plugins and tools based on those.
- XRootD software framework is a fully generic suite for fast, low latency and scalable data access, which can serve natively any kind of data, organized as a hierarchical filesystem-like namespace, based on the concept of directory.

HARDWARE AND TOPOLOGY OF COMPUTING FACILITY

8

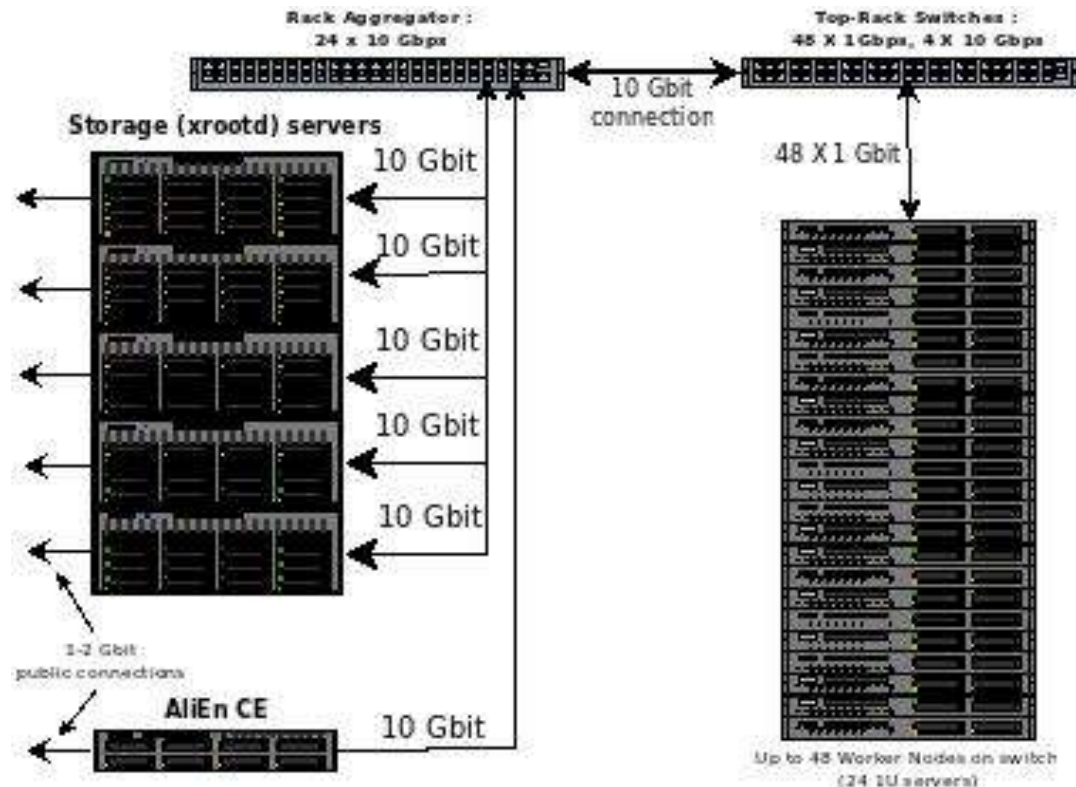
- Our hardware is mainly comprised of SuperMicro machines that were chosen for the great resource density/price ratio. For computing nodes we use Twin servers which give us densities of 4 sockets/1U and for the storage we use servers with 24, 36 drives and JBOD cases with 45 drives in 4U of rack space.
- Generic schematic of ISS computing facility :



HARDWARE AND TOPOLOGY OF COMPUTING FACILITY

9

- The AliEn cluster has at his core a 10 Gbps aggregating switch which is connected to the top-of-rack switch of the computing nodes. In the aggregating switch are connected the private interfaces of the storage node, a topology which give a high bandwidth connection between worker nodes and storage with very little oversubscribing.



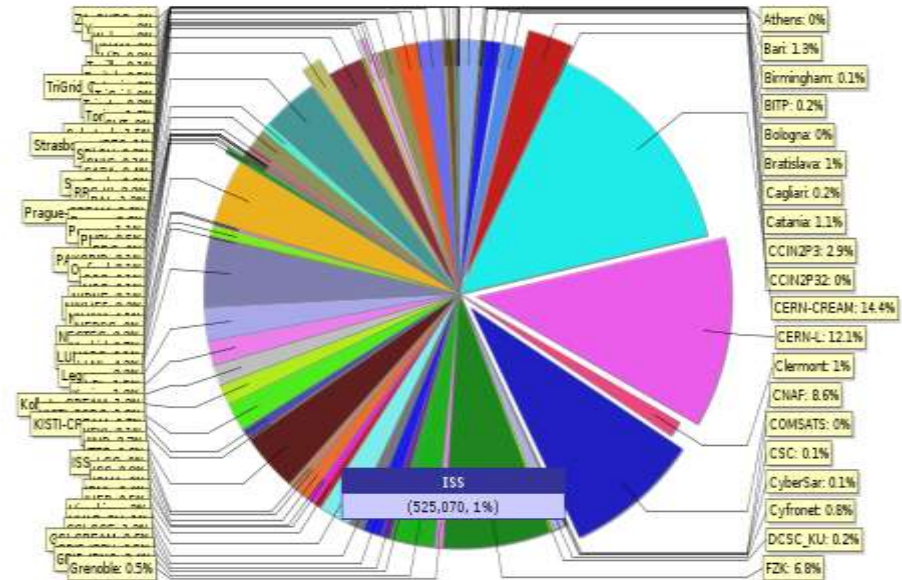
High utilization of what is provided ISS-ALICE

10

Jobs history (5 years!)



1% of ALICE computing (last year period)



<http://alimonitor.cern.ch>

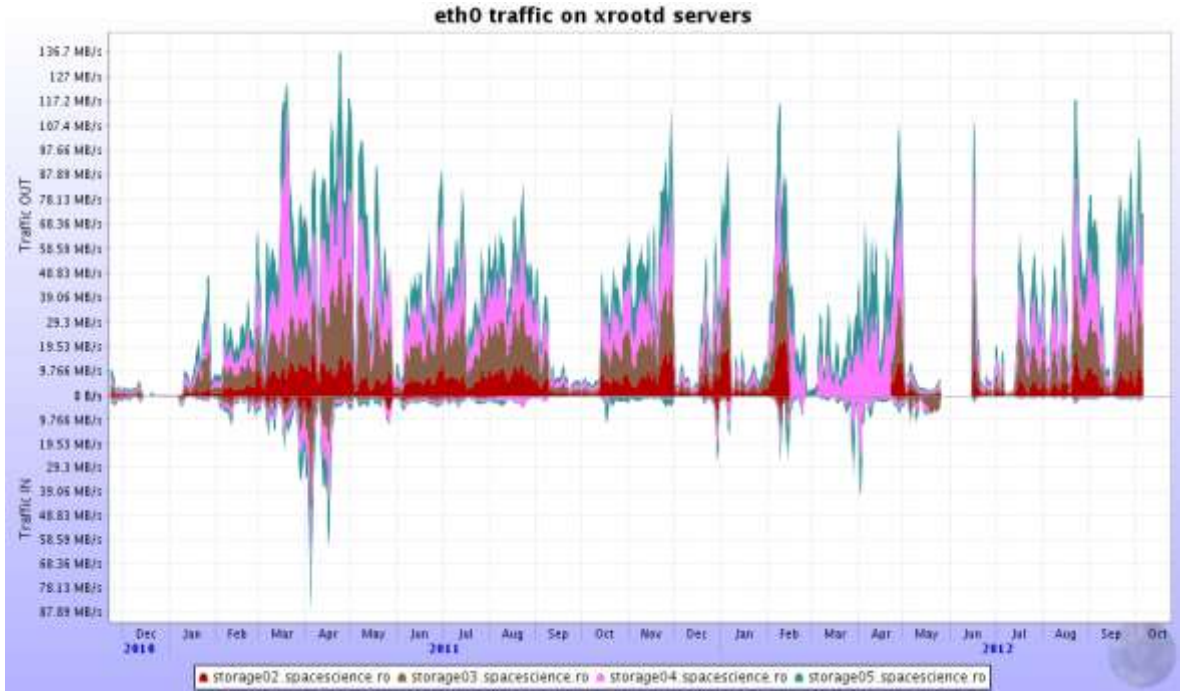
- high utilization by ALICE user analyses





High utilization of what is provided ISS-ALICE

High storage traffic:



Last 3 years

Traffic IN					
Series	Last value	Min	Avg	Max	Total
1. storage02.spacescience.ro	135.6 KB/s	0.148 KB/s	463.1 KB/s	90.8 MB/s	25.43 TB
2. storage03.spacescience.ro	416.4 KB/s	0.151 KB/s	1.206 MB/s	90.99 MB/s	67.92 TB
3. storage04.spacescience.ro	428.5 KB/s	0.135 KB/s	1.355 MB/s	99.84 MB/s	76.31 TB
4. storage05.spacescience.ro	395.9 KB/s	0.139 KB/s	1.291 MB/s	101.5 MB/s	72.72 TB
Total	1.344 MB/s		4.305 MB/s		242.4 TB

Traffic OUT					
Series	Last value	Min	Avg	Max	Total
1. storage02.spacescience.ro	7.566 MB/s	56.19 B/s	5.46 MB/s	107.7 MB/s	307 TB
2. storage03.spacescience.ro	22.07 MB/s	56.39 B/s	10.92 MB/s	117.4 MB/s	615 TB
3. storage04.spacescience.ro	22.23 MB/s	47.81 B/s	12.38 MB/s	117.8 MB/s	697.3 TB
4. storage05.spacescience.ro	20.98 MB/s	44.89 B/s	10.91 MB/s	117.5 MB/s	614.4 TB
Total	72.85 MB/s		39.68 MB/s		2.181 PB

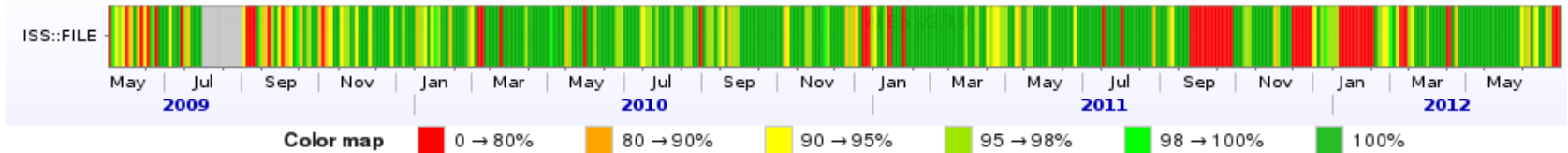
High utilization of what is provided ISS-ALICE

High availability - 88.68%
 Success ratio - 88.47%
 High usage - 97.96%

Disk storage elements

ISS																		
SE Name	AliEn SE	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage	Version	Functional tests		Last day tests		Demotion	
SE Name	AliEn name	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage	Version	add	get	Last OK test	Successful	Failed	factor
1. ISS - FILE	ALICE::ISS::FILE	136.1 TB	113.4 TB	22.65 TB	83.35%	2,291,056	FILE	140.5 TB	137.6 TB	2.865 TB	97.96%	v3.1.0			05.07.2012 16:04	7	5	38.39%
Total		136.1 TB	113.4 TB	22.65 TB		2,291,056		140.5 TB	137.6 TB	2.865 TB								

AliEn SEs availability



Statistics

Statistics

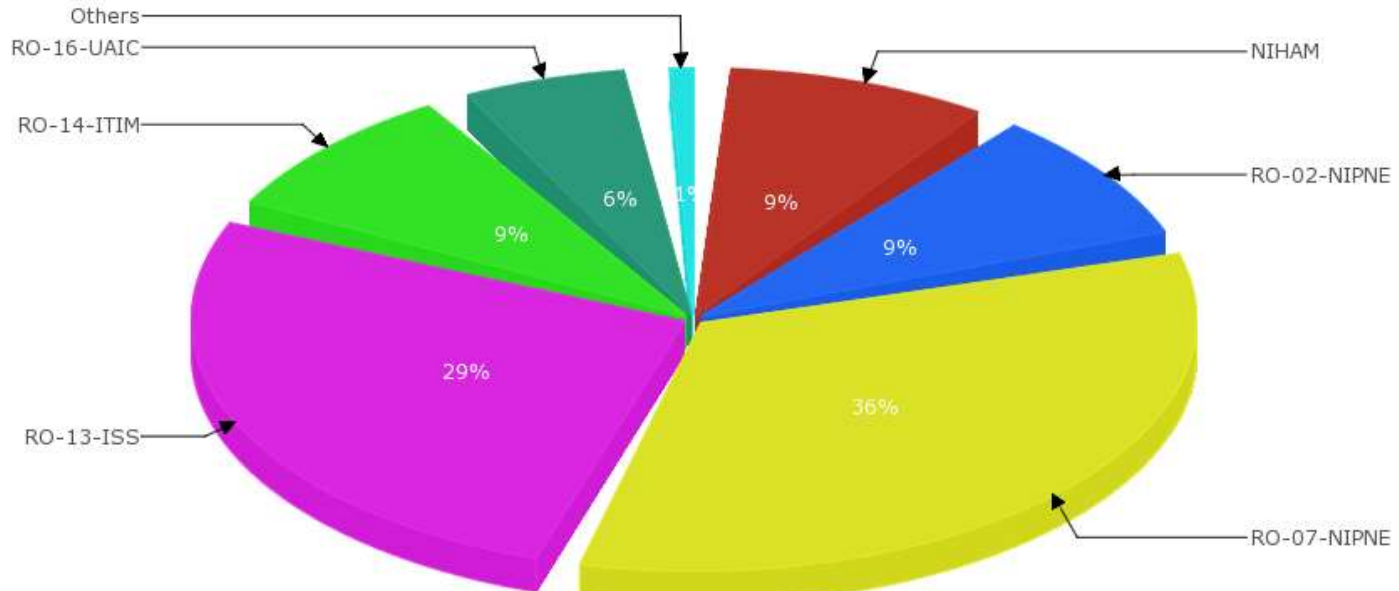
Link name	Data		Individual tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
ISS::FILE	01 May 2009 16:10	30 Jun 2012 16:02	11555	1506	88.47%	88.68%



RO-13-ISS Statistics

Total number of jobs run by SITE and VO						
SITE	alice	atlas	lhcb	Total		
NIHAM	1,501,239	0	0	1,501,239	9.25%	
RO-02-NIPNE	0	1,500,227	0	1,500,227	9.24%	
RO-07-NIPNE	162,111	5,856,719	102,718	5,921,548	36.40%	
RO-09-NIPNE	0	0	75,029	75,029	0.46%	
RO-13-ISS	4,713,792	0	0	4,713,792	29.84%	
RO-14-ITIM	0	1,508,966	0	1,508,966	9.30%	
RO-15-NIPNE	0	0	70,707	70,707	0.44%	
RO-16-UAIC	14,103	927,947	0	942,050	5.80%	
Total	6,391,246	8,993,859	246,464	16,233,568		
Percentage	39.37%	59.10%	1.53%			

NGI_RO Total number of jobs per SITE

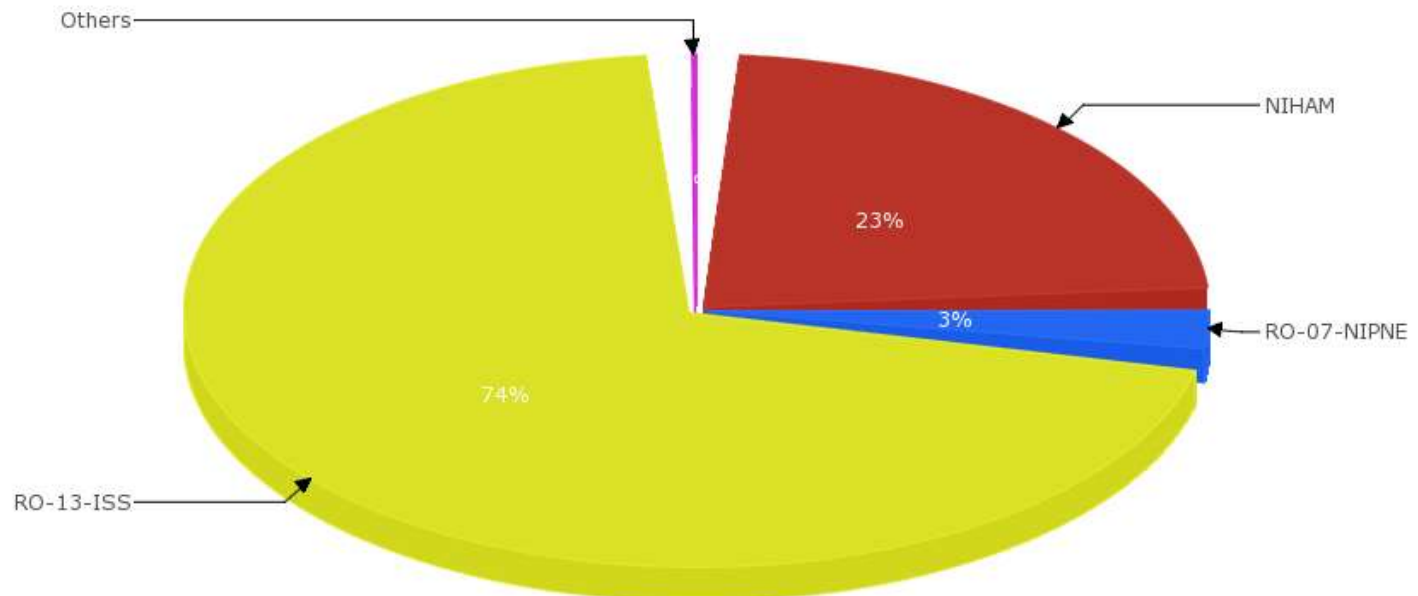


RO-13-ISS Statistics

14

Total number of jobs run by SITE and VO				
SITE	alice	Total	%	
NIHAM	1,501,239	1,501,239	23.49%	
RO-07-NIPNE	162,111	162,111	2.54%	
RO-13-ISS	4,713,792	4,713,792	73.75%	
RO-16-UAIC	14,103	14,103	0.22%	
Total	6,391,245	6,391,245		
Percentage	100.00%			

NGI_RO Total number of jobs per SITE



Monitoring solution - MonALISA

15



MonALISA Repository for ISS @ ALICE



[Repository Home](#) | [Administration Section](#) | [Events XML Feed](#) | [MonaLisa GUI](#)

ISS

- ISS Repository
 - Networking
 - WAN
 - Uplink
 - LAN
 - alien
 - Storage
 - hrootd status
 - hrootd traffic
 - Jobs
 - Activity
 - Accounting
 - Current jobs
 - Infrastructure
 - Datacenter sensors
 - UPSes Status
 - Switches
 - Switches Status
 - Switches Sensors
 - Servers status
 - AliEn cluster
 - gLite cluster
 - Xrootd storage
 - Various machines
 - vobox.spacescience.ro
 - Repository info

close all

This page: bookmark, URL

Running Jobs trend

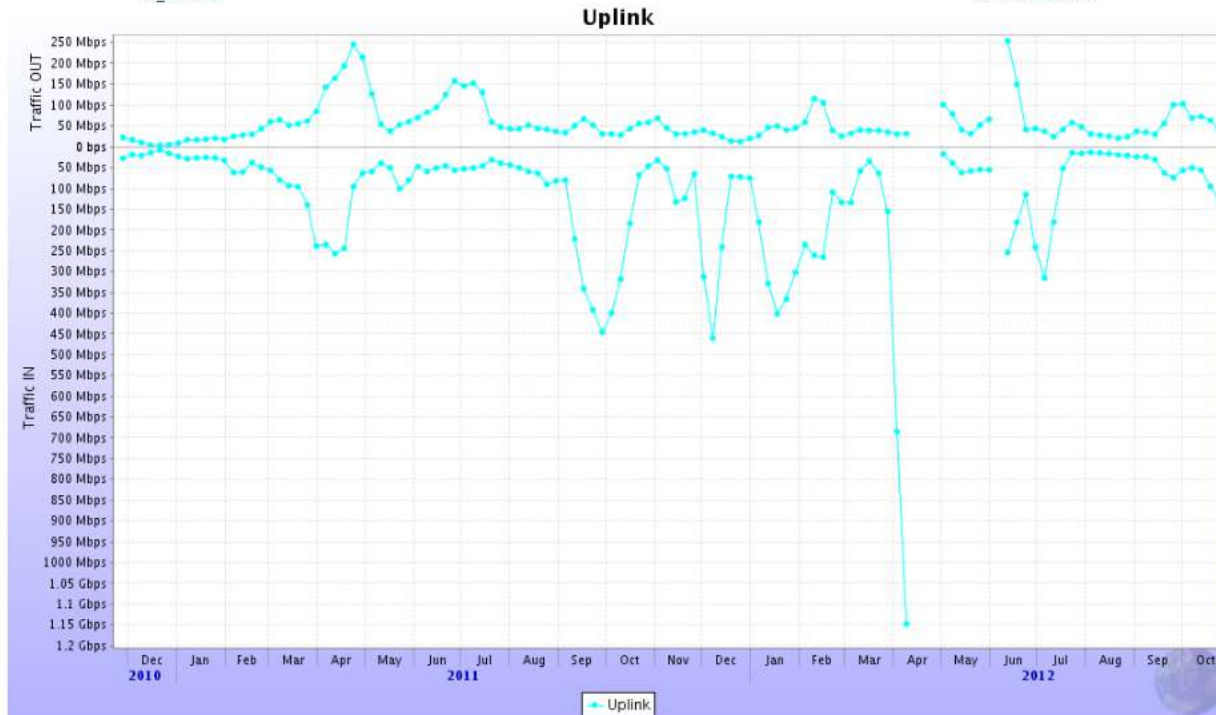


Series Options

Interval selection: last 3 years or « 2009-10-25 22:0 - 2012-10-24 23:00 » Plot

Charts: OUT IN (check all | uncheck all)

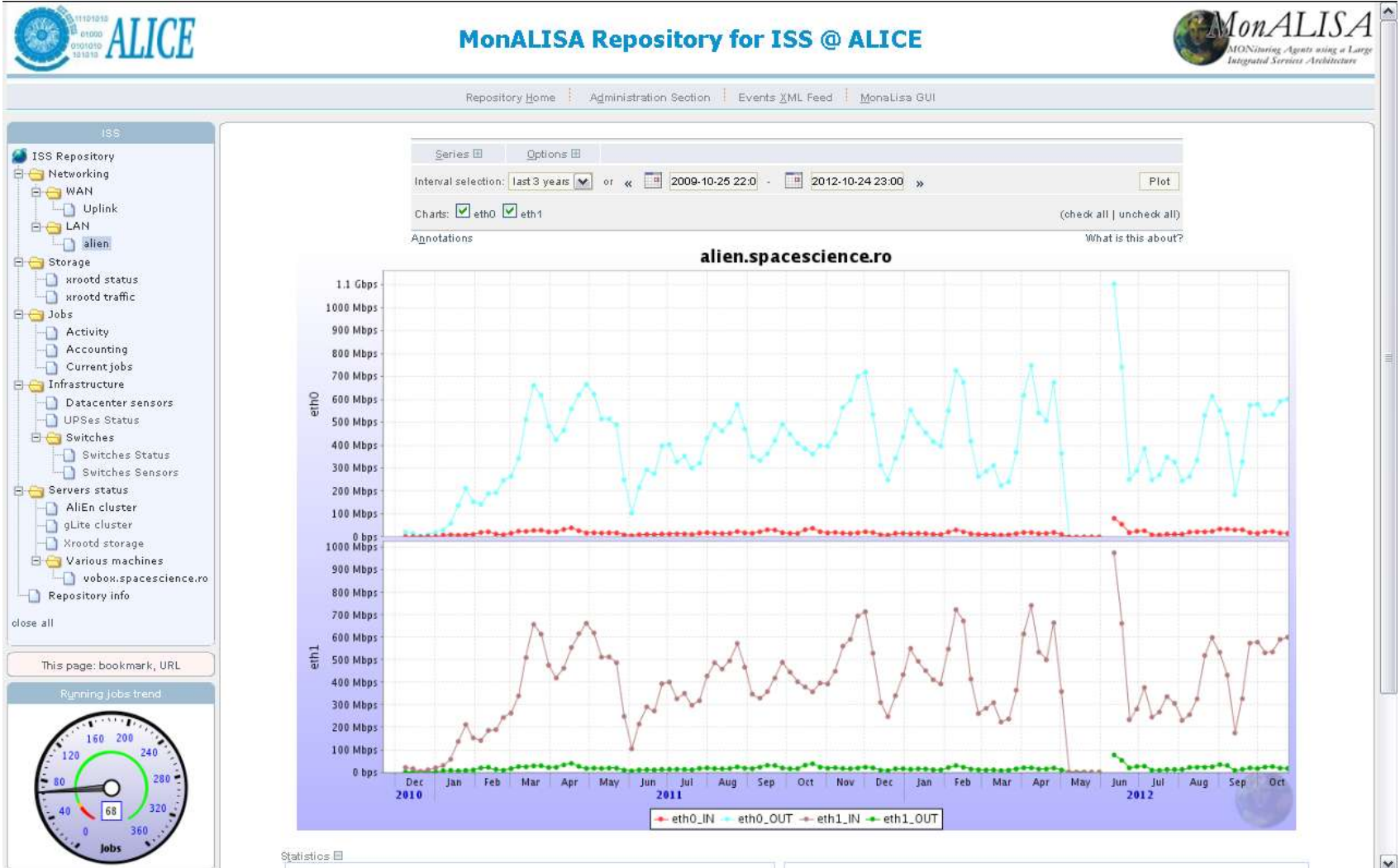
Annotations What is this about?



Statistics

Monitoring solution - MonALISA

16



Monitoring solution - MonALISA

17

ISS Repository

- Networking
 - WAN
 - Uplink
 - LAN
 - alien
- Storage
 - xrootd status
 - xrootd traffic
- Jobs
 - Activity
 - Accounting
 - Current jobs
- Infrastructure
 - Datacenter sensors
 - UPSes Status
 - Switches
 - Switches Status
 - Switches Sensors
- Servers status
 - AliEn cluster
 - gLite cluster
 - Xrootd storage
 - Various machines
 - vobox.spacescience.ro
- Repository info

close all

This page: bookmark, URL

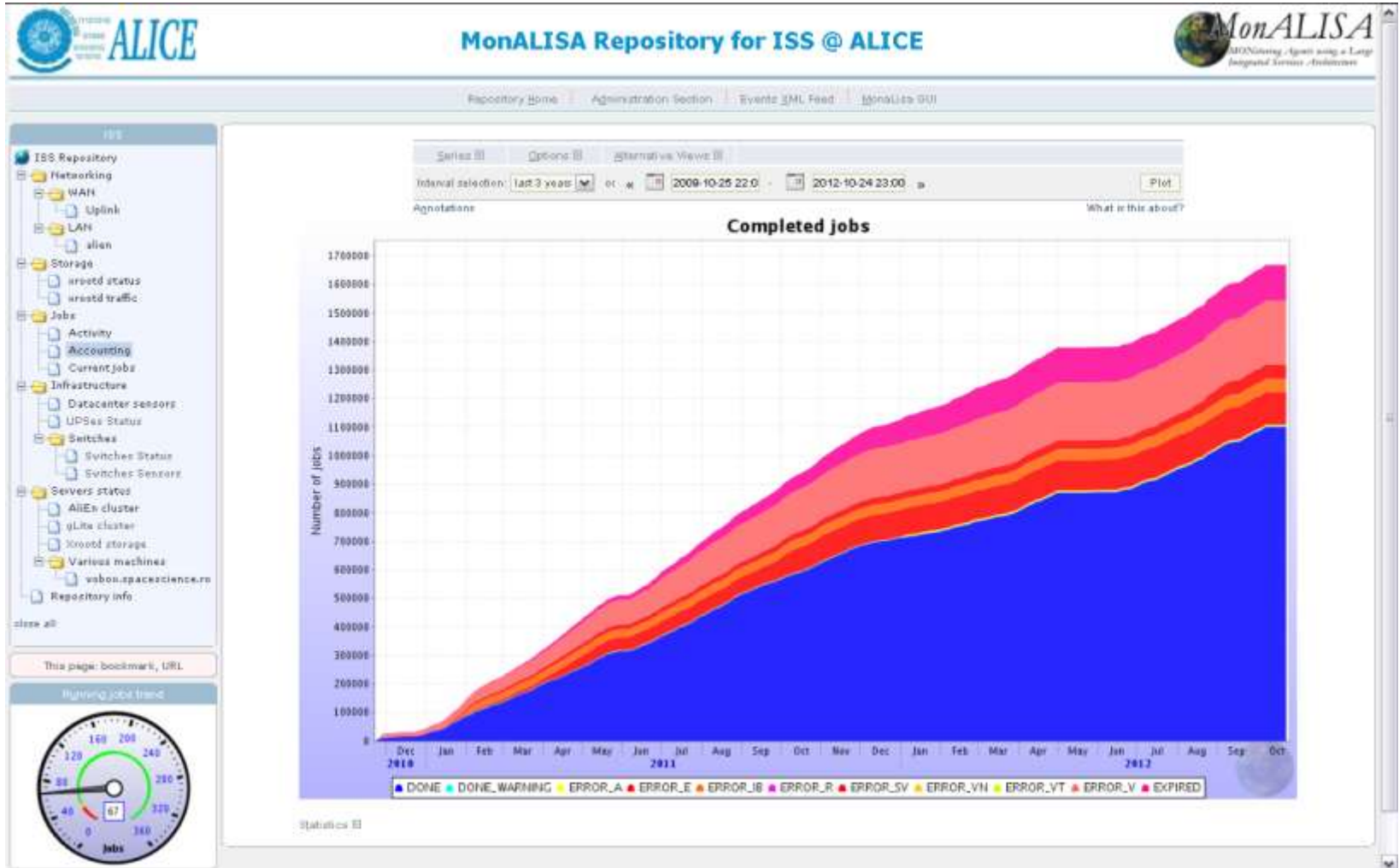
Running jobs trend

Jobs




Monitoring solution - MonALISA

18




Monitoring solution - MonALISA

19



MonALISA Repository for ISS @ ALICE



[Repository Home](#) | [Administration Section](#) | [Events & ML Feeds](#) | [MonALISA GUI](#)

ISS

- ISS Repository
- Networking
 - WAN
 - Uplink
 - LAN
 - alien
- Storage
 - irrored status
 - irrored traffic
- Jobs
 - Activity
 - Accounting
 - Current jobs
- Infrastructure
 - Datacenter sensors
 - UDSax Status
 - Switches
 - Switches Status
 - Switches Sensors
- Servers status
 - ALICE cluster
 - gLite cluster
 - Xrootd storage
 - Various machines
 - vabou.space-science.ro
- Repository info

close all

Xrootd nodes' status


What is this about?

Machines status

Machine	Machine			CPU usage				RAM			Swap		Sockets		xrootd space		
	Load 1	Uptime	User	System	IO	Wait	Nice	Idle	Free	Cache	Total	Used	Total	TCP	Free	Used	Total
1. rd.space-science.ro	0.18	18d 9i 31	4.038	0.11	1.947	0	93.7	2.262 GB	1.903 GB	3.859 GB	0.137 MB	1.999 GB	37	2.056 TB	99.04 TB	101.1 TB	
2. storage01.space-science.ro	0	18d 7i 38	0.14	0.11	0.042	0	99.66	1.823 GB	1.744 GB	1.978 GB	0.41 MB	1.999 GB	10	87.74 GB	4.319 TB	4.405 TB	
3. storage02.space-science.ro	0.19	82d 8i 19	0.021	0.062	1.12	0	98.76	3.389 GB	3.342 GB	3.859 GB	0.18 MB	3.999 GB	18	360.8 GB	17.55 TB	17.9 TB	
4. storage03.space-science.ro	0	82d 8i 19	0.03	0.117	0.371	0	99.29	3.283 GB	3.242 GB	3.859 GB	0.145 MB	3.999 GB	16	817.3 GB	38.6 TB	39.4 TB	
5. storage04.space-science.ro	2774	82d 6i 39	3.38	0.041	0	0	0	5.261 GB	5.061 GB	5.822 GB	0.176 MB	3.999 GB	1048	819.9 GB	38.59 TB	39.39 TB	
Total		0	92d 4i 12	0.032	0.036	0	99.86	3.373 GB	3.329 GB	3.855 GB	0.207 MB	3.999 GB	17	839.1 GB	38.57 TB	39.39 TB	

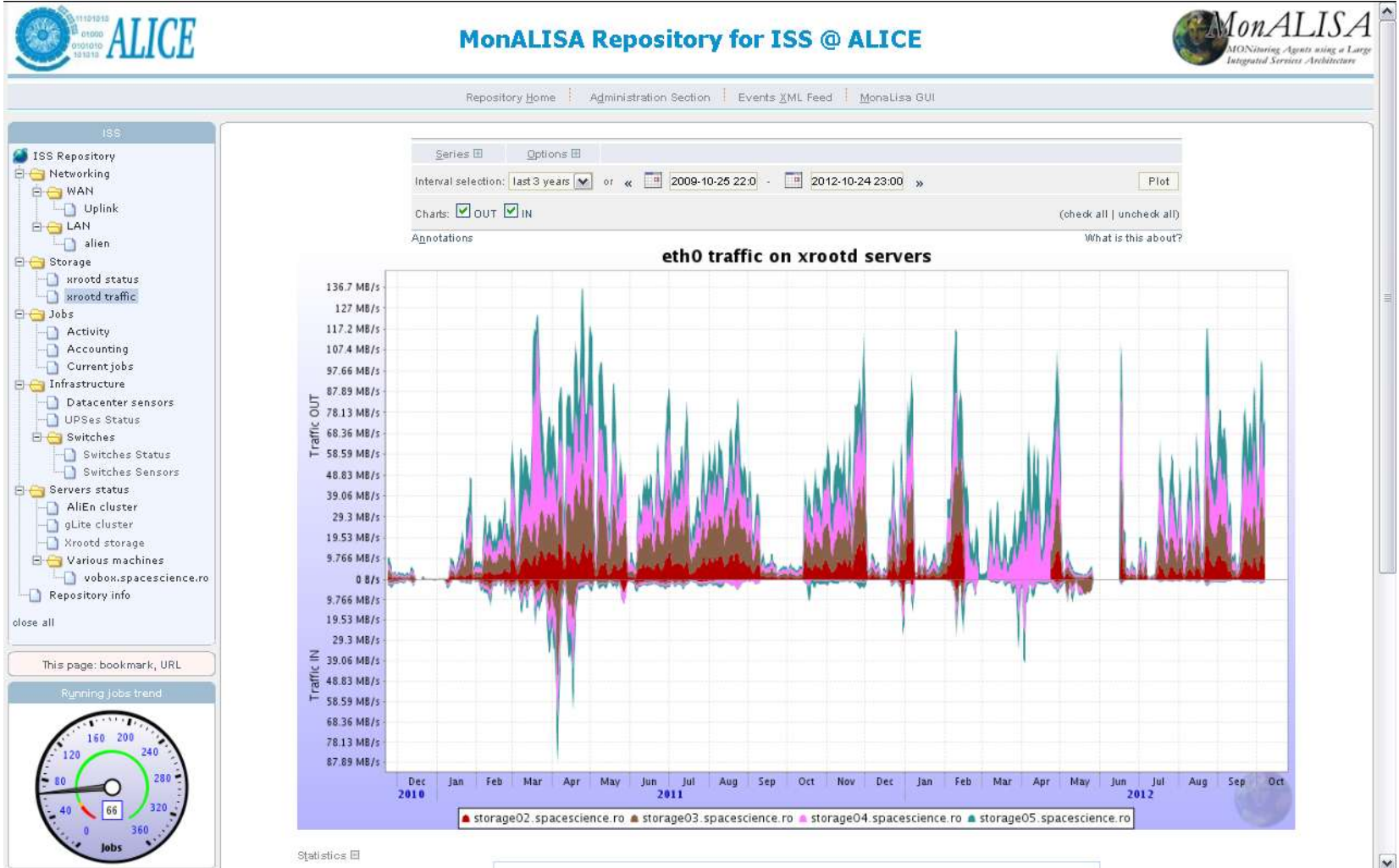
This page: bookmark, URL

Running job trend



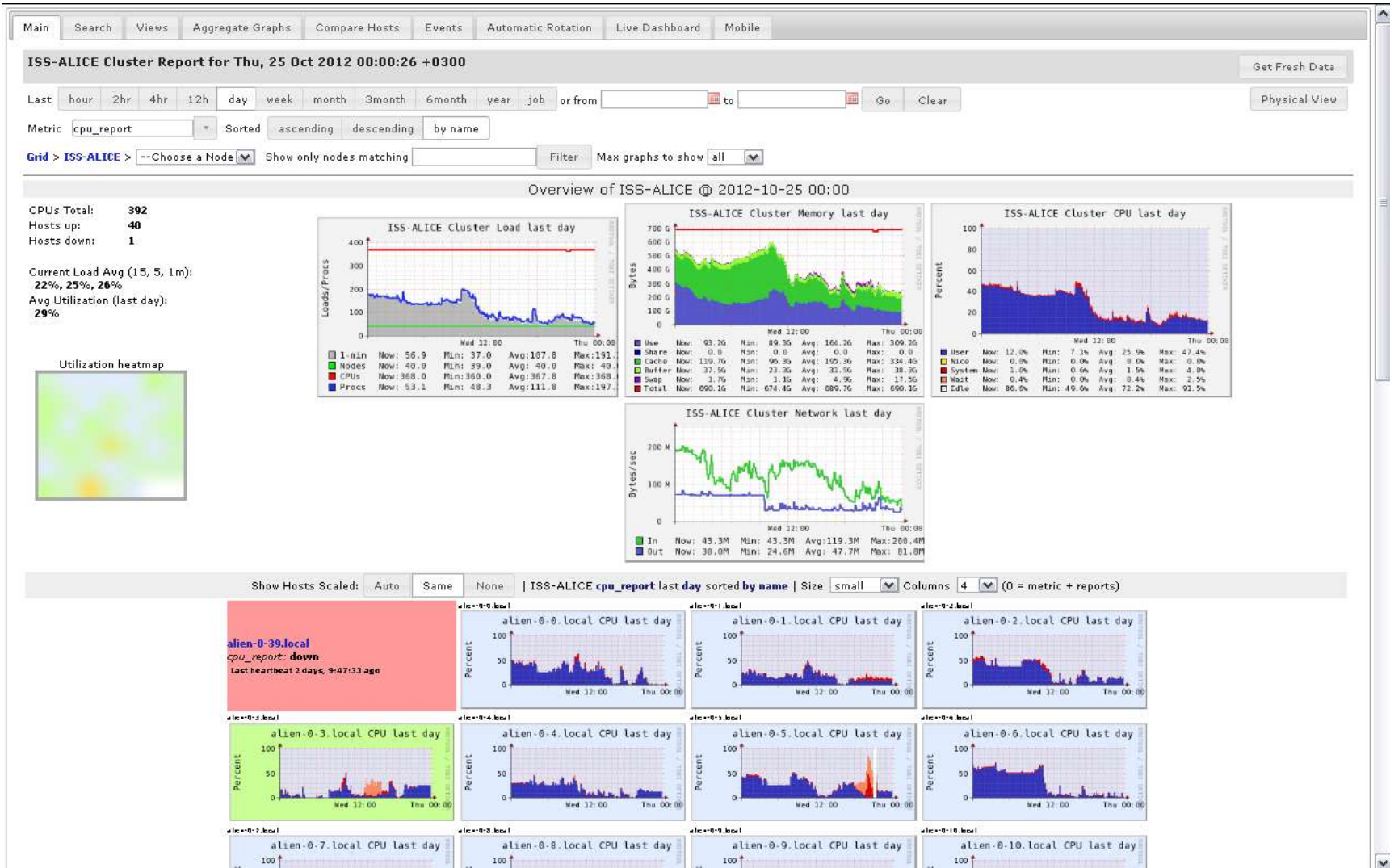
Monitoring solution - MonALISA

20



Monitoring solution - Ganglia

21



Thank you for your attention!