Status report of ISS Grid activities

Liviu IRIMIA, Ionel STAN, Adrian SEVCENCO

Institute of Space Science
P.O. Box: MG-23, RO 077125, Bucharest-Magurele ROMANIA
http://www.spacescience.ro

RO-LCG 2016 'Grid, Cloud, and High-Performance Computing in Science
26-28 October 2016

ISS Computing

Cluster	Number of servers	Core
ISS-ALICE	44	416
RO-13-ISS	8	128
PlanckGrid	16	144
RoSpaceGrid	50	784
Total	118	1472

New ISS Computing Infrastructure

➔ Designed for high density computing (Hot Aisle, InRow cooling)

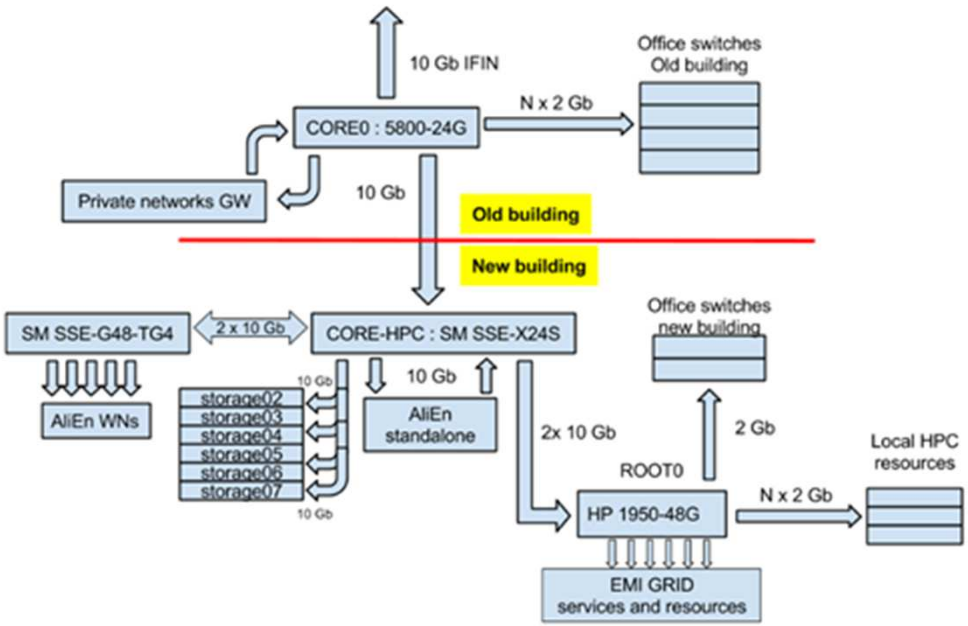
➔ Scalable solution for future investments

➔ UPS Power : 48 kVA (with N+1 redundancy power units)

➔ Cooling capacity : 80 kW installed (2N capacity redundancy)

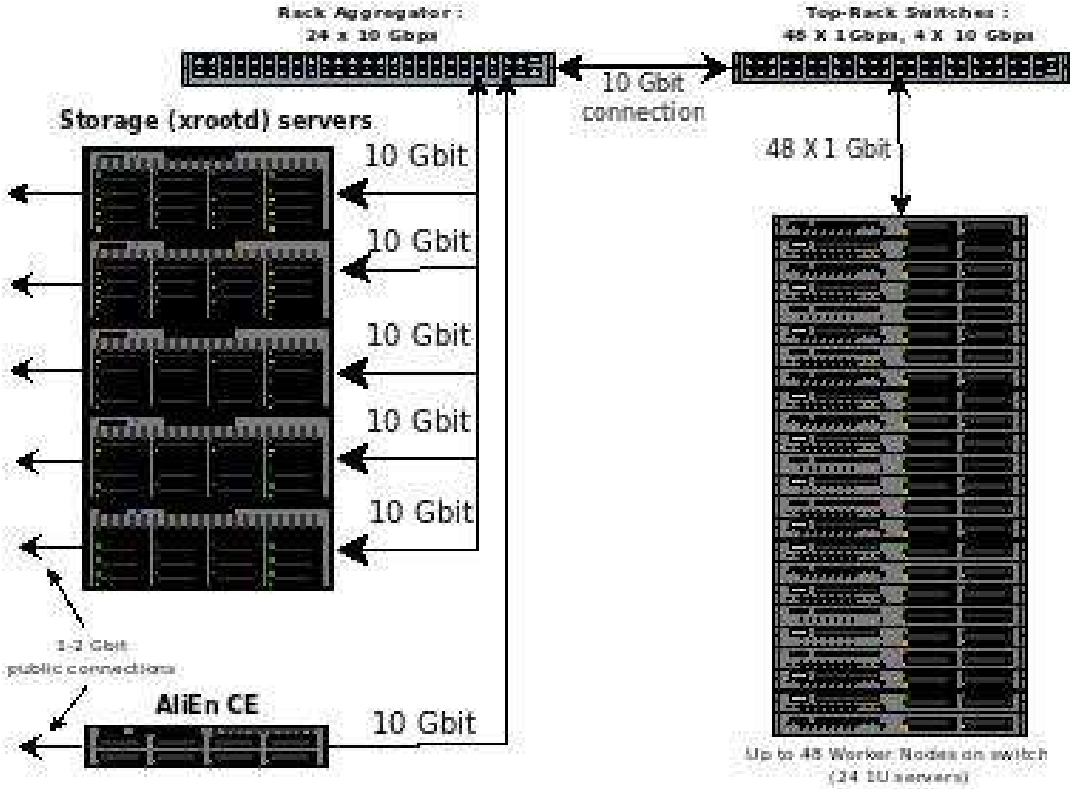
HARDWARE AND TOPOLOGY OF COMPUTING FACILITY

- 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

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

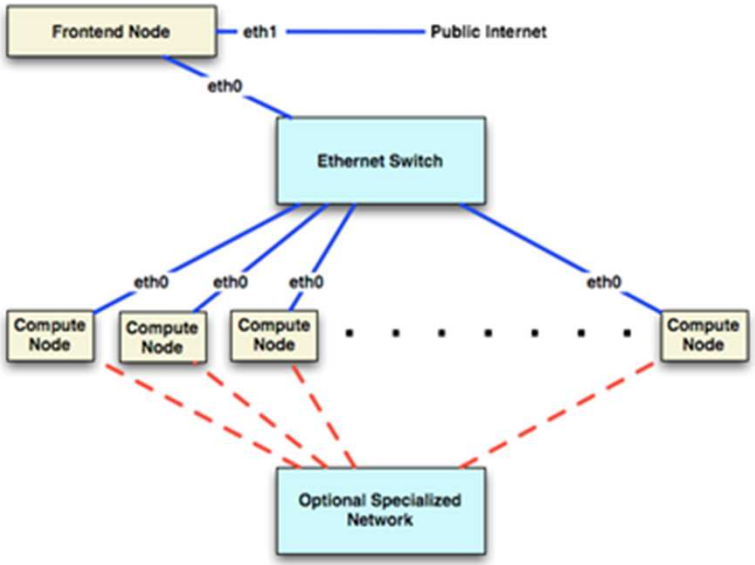


Rocks Clusters

- easy deployment, management, maintenance; flexible
- facile extensibility of software packages via software entities known as “Rolls”.
- it’s a system based on the Red Hat – CentOS flavor
- comes with a database of scripts that simplifies node deployment procedures
- customizable installation via Rolls.

Rocks Clusters

- wide spread use : present in more than 1800 clusters worldwide
- uses the server – client model
- the central server will contain all necessary central clustering services
- on the computing nodes the first Ethernet interface will be used for the private network



- Rocks uses 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

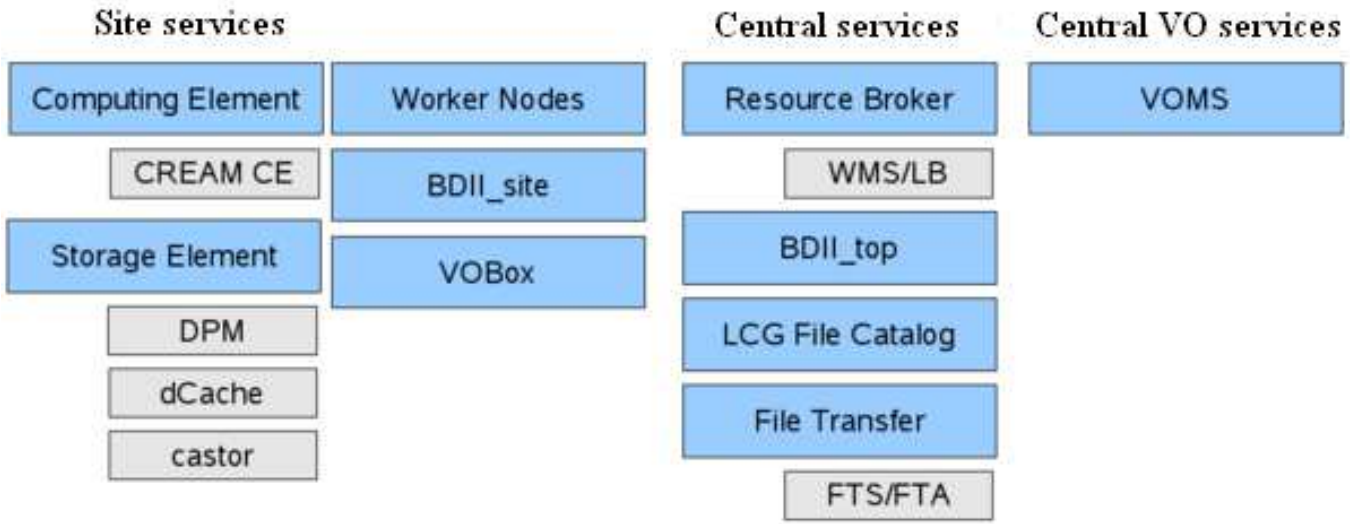
- is a lightweight Grid framework which is built around Open Source components using the Web Services model. It has been initially developed by the ALICE collaboration (ALICE Offline Project) as a production environment for the simulation, reconstruction, and analysis of physics data in a distributed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

UMD/EMI –Unified Middleware Distribution/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
- is a close collaboration of three major middleware providers, ARC, gLite and UNICORE, and other specialized 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

- One of the central services in the UMD/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



UMDUI – User interactive tool

Umdui - ISS Proxy LCG Info Manage Jobs File Management

Proxy Management

Grid-proxy-init | Grid-proxy-info | Grid-voms-info

Grid-voms-init | Grid-proxy-destroy

Command output:

```
subject  : /DC=RO/DC=RomanianGRID/O=ISS/CN=Liviu IRIMIA/CN=1977948870
issuer   : /DC=RO/DC=RomanianGRID/O=ISS/CN=Liviu IRIMIA
identity : /DC=RO/DC=RomanianGRID/O=ISS/CN=Liviu IRIMIA
type     : RFC 3820 compliant impersonation proxy
strength : 1024 bits
path     : /tmp/x509up_u501
timeleft : 0:00:00

subject   : /DC=RO/DC=RomanianGRID/O=ISS/CN=Liviu IRIMIA/CN=1977948870
issuer    : /DC=RO/DC=RomanianGRID/O=ISS/CN=Liviu IRIMIA
identity  : /DC=RO/DC=RomanianGRID/O=ISS/CN=Liviu IRIMIA
type      : RFC3820 compliant impersonation proxy
strength  : 1024
path      : /tmp/x509up_u501
timeleft  : 00:00:00
key usage : Digital Signature, Key Encipherment, Data Encipherment
```

Grid-proxy-init dialog

Certificate Password: ••••••••••••••••

Close Execute

UMDUI – User interactive tool

Umdui - ISS | Proxy | LCG Info | Manage Jobs | File Management

General informations

LCG-info | LCG-infosites

Command output:

```
Avail Space(kB)  Used Space(kB)  Type      SE
----------------------------------------
    1968384708        628175  SRM     grid02.spacescience.ro
   88709068147    30812943865  SRM     seau.spacescience.ro
```

Infosites Dialog

VO: dteam

Filter: Filter(Optional)...

- **Site** ☐
- **Verbose** ☐
- **CE** ☐
- **SE** ☑
- **Space** ☐
- **closeSE** ☐
- **tag** ☐
- **lfc** ☐
- **wms** ☐
- **site names** ☐

Close | Execute

Future applications – Monitoring tools

ISS-CORE-0 | ISSAF | **ProCurve2810-48G_-_CORE1** | ProCurve2810-48G_-_CORE2 | ProCu

Description	Admin status	Operation status	In discards	In errors	Out discards	Out errors
1	up	up	0	6	164866	0
2	up	down	0	0	0	0
3	up	down	0	0	0	0
4	up	down	0	0	0	0
5	up	down	0	0	160	0
6	up	down	0	0	0	0
7	up	up	0	1	6739	0
8	up	down	0	0	0	0
9	up	down	0	9	54287	0
10	up	down	0	0	0	0
11	up	down	0	0	0	0
12	up	up	0	0	160305	0
13	up	down	0	0	0	0
14	up	down	0	0	0	0
15	up	down	0	0	0	0
16	up	down	0	1	22251	0
17	up	up	0	98	170637	0
18	up	down	0	0	0	0
19	up	up	0	1	170646	0
20	up	up	0	0	159117	0
21	up	up	0	0	167051	0

Future applications – Monitoring tools

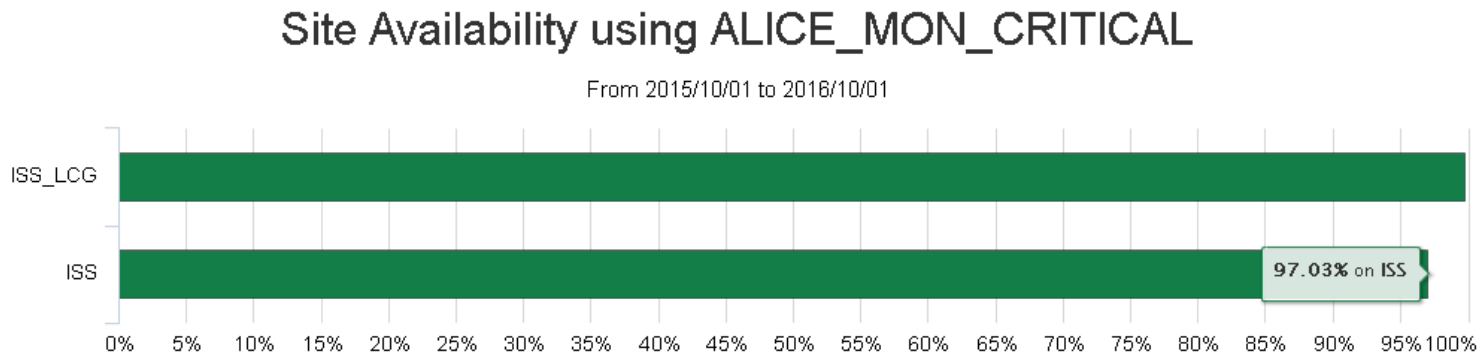
172.20.0.51 | **172.20.0.52** | 172.20.0.53 | 172.20.0.54 | 172.20.0.55 | 172.20.0.56

Raid-172.20.0.52 | Volume-172.20.0.52 | Events-172.20.0.52

Number	Name	Disks	State	Total Capacity	Free capacity	HDD capacity	HDD channels
ARC-1280							
0	"raid0"	12	"Normal"	12000000	0	1000000	"1.2.3.4.5.6.7.8.9.10.11.12"
1	"raid1"	12	"Normal"	12000000	0	1000000	"13.14.15.16.17.18.19.20.21.22.23.24"

Global Performance

SITE AVAILABILITY



http://wlcg-sam-alice.cern.ch

STORAGE AVAILABILITY

Statistics

Link name	Data		Individual results of writing tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
ISS::FILE	01 Oct 2015 22:14	01 Oct 2016 20:14	4271	116	97.36%	97.37%

http://alimonitor.cern.ch

Statistics per site – RO-13-ISS

Done jobs

RO-13-ISS Total number of jobs by SITE and VO.
VOs. September 2015 - September 2016.

The following table shows the distribution of Total number of jobs grouped by SITE and VO.

Total number of jobs run by SITE and VO						
SITE	alice	auger	dteam	ops	Total	%
RO-13-ISS	332,141	2,259	27	92,730	427,157	100.00%
Total	332,141	2,259	27	92,730	427,157	
Percentage	77.76%	0.53%	0.01%	21.71%		

http://accounting.egi.eu

Normalised CPU time (HEPSPEC06)

RO-13-ISS Normalised CPU time (HEPSPEC06) by SITE and VO.
LHC VOs. September 2015 - September 2016.

The following table shows the distribution of Normalised CPU time (HEPSPEC06) grouped by SITE and VO (only information about **LHC VOs** is returned).

Normalised CPU time [units HEPSPEC06.Hours] by SITE and VO			
SITE	alice	Total	%
RO-13-ISS	20,428,712	20,428,712	100.00%
Total	20,428,712	20,428,712	
Percentage	100.00%		

http://accounting.egi.eu

Statistics per site – RO-13-ISS

Total CPU time

RO-13-ISS Total CPU time used by SITE and VO.

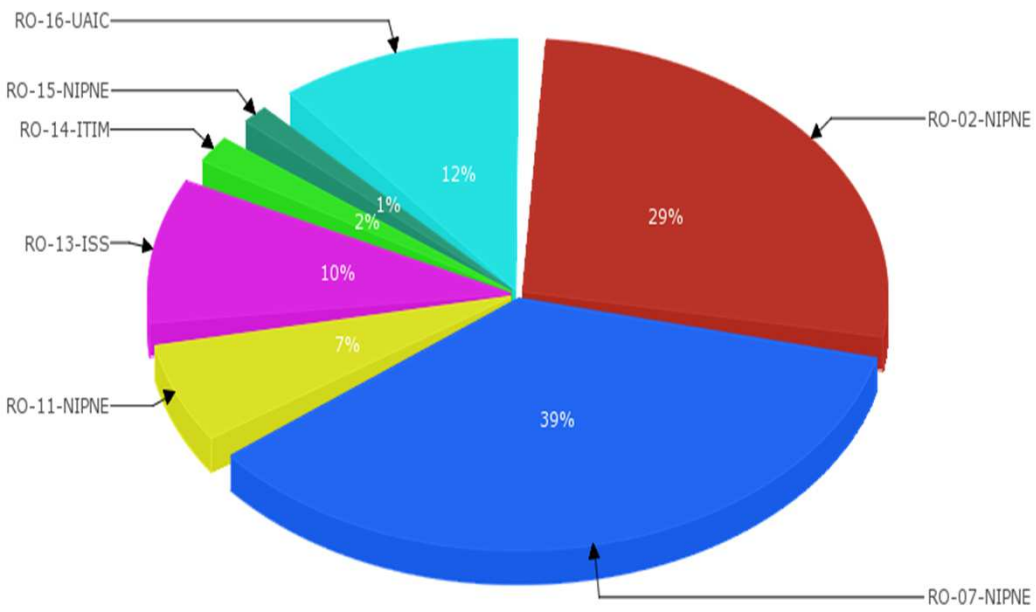
LHC VOs. September 2015 - September 2016.

The following table shows the distribution of Total CPU time used grouped by SITE and VO (only information about **LHC VOs** is returned).

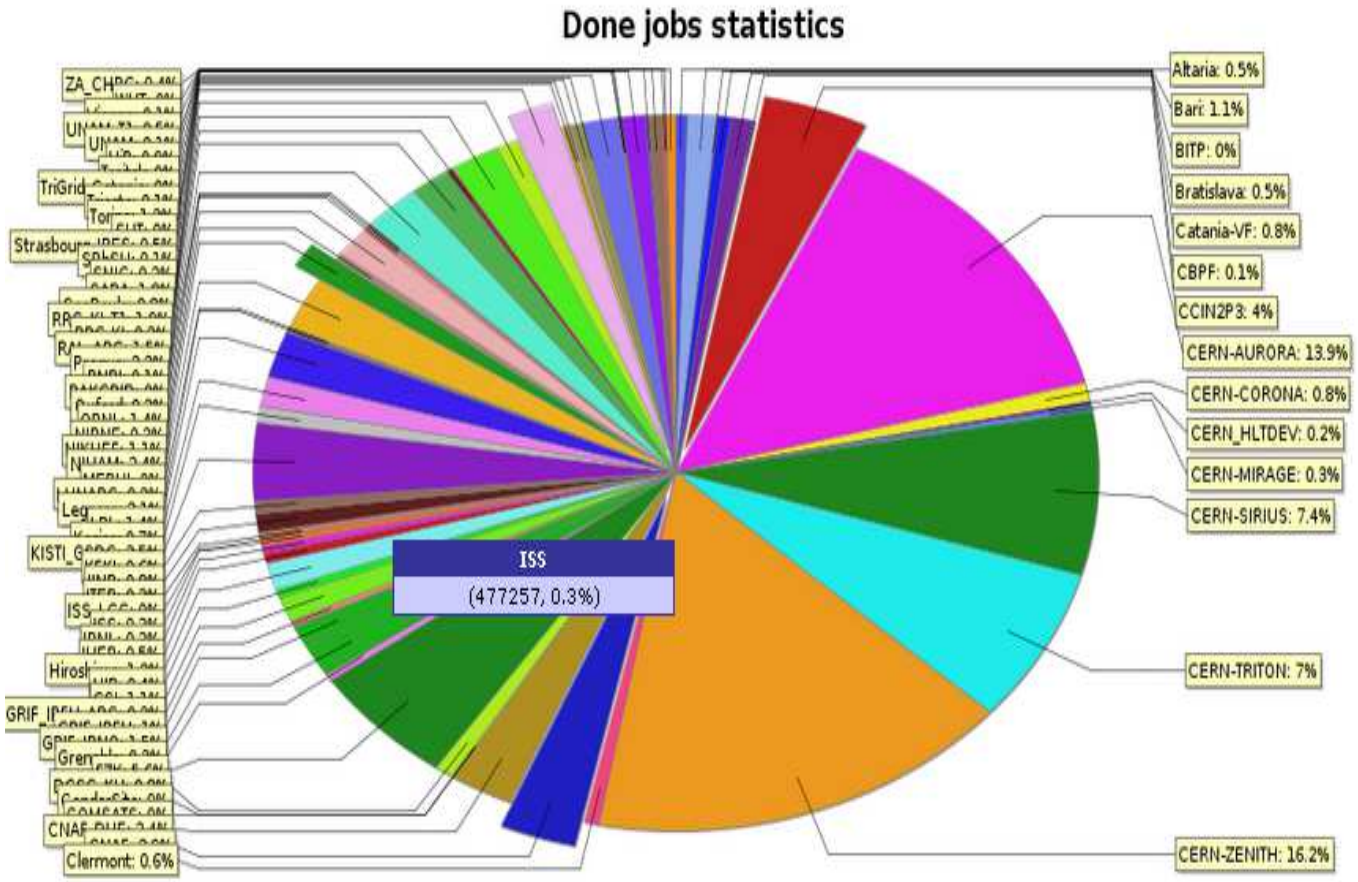
Total CPU time used [units Hours] by SITE and VO			
SITE	alice	Total	%
RO-13-ISS	2,592,475	2,592,475	100.00%
Total	2,592,475	2,592,475	
Percentage	100.00%		

Developed by CESGA 'EGI View': / sumcpu / 2015:9-2016:9 / SITE-VO / lhc (x) / GRBAR-LIN / l

NGI_RO Total CPU time used per SITE

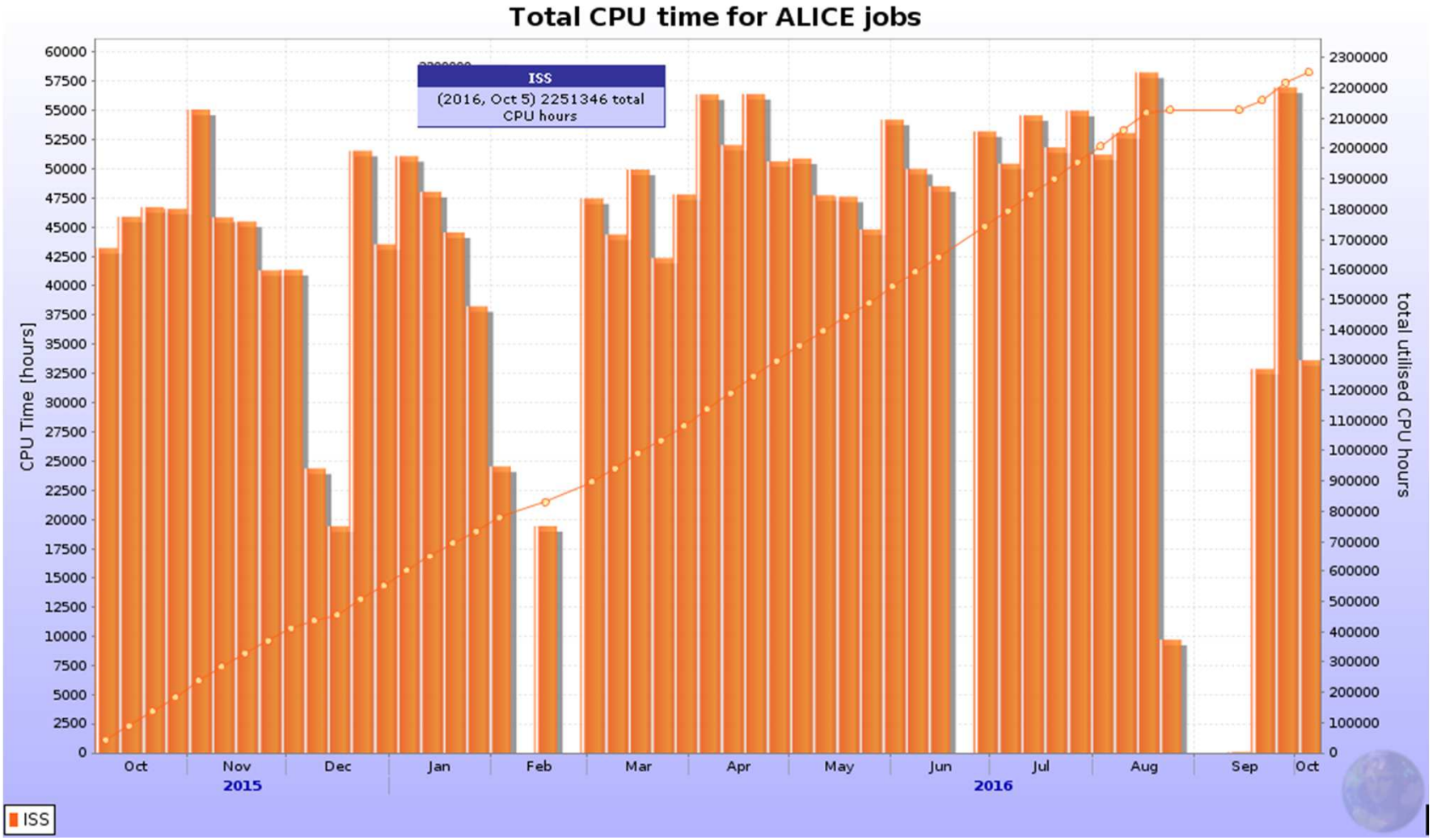


Statistics per site – ISS-Alice

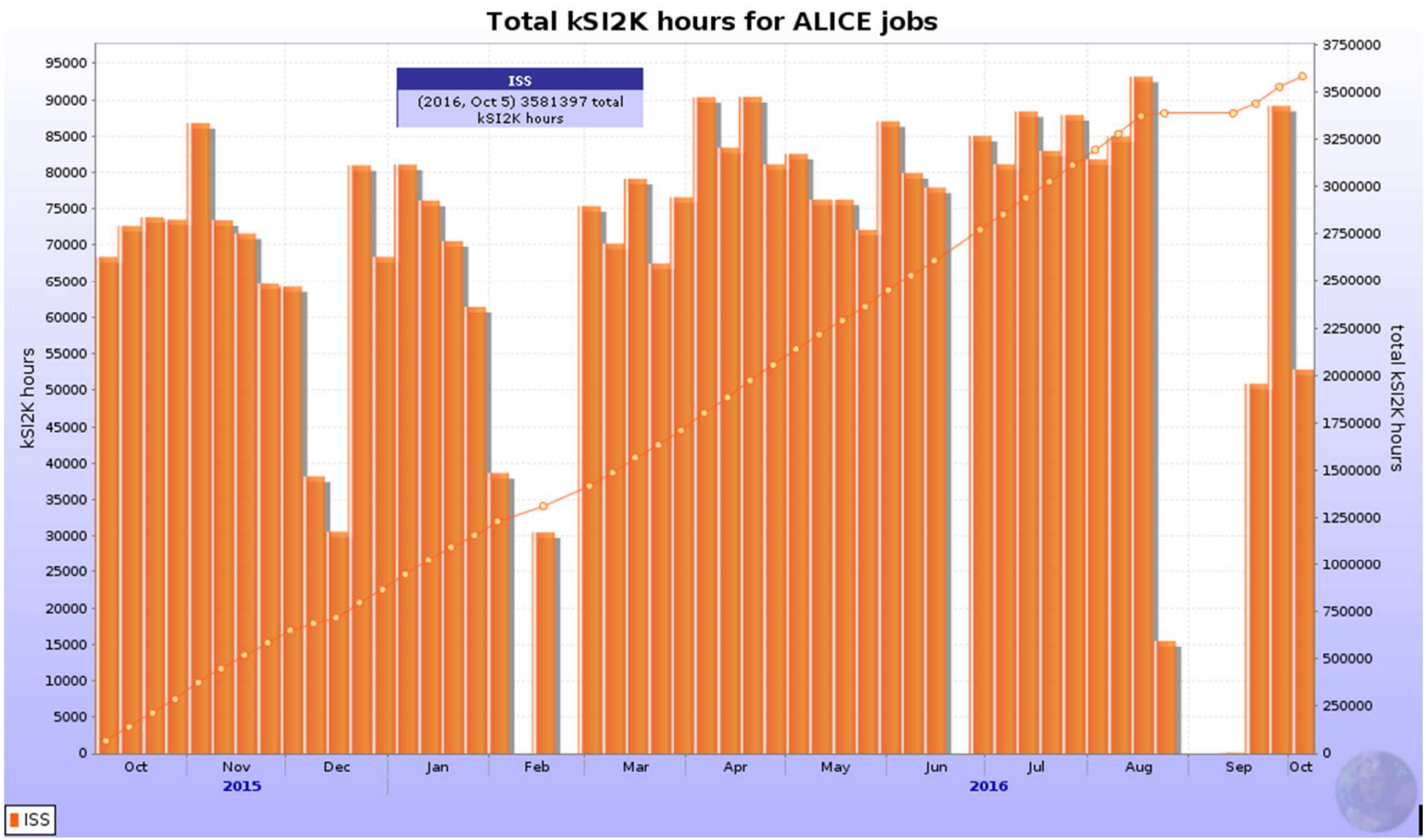


http://alimonitor.cern.ch

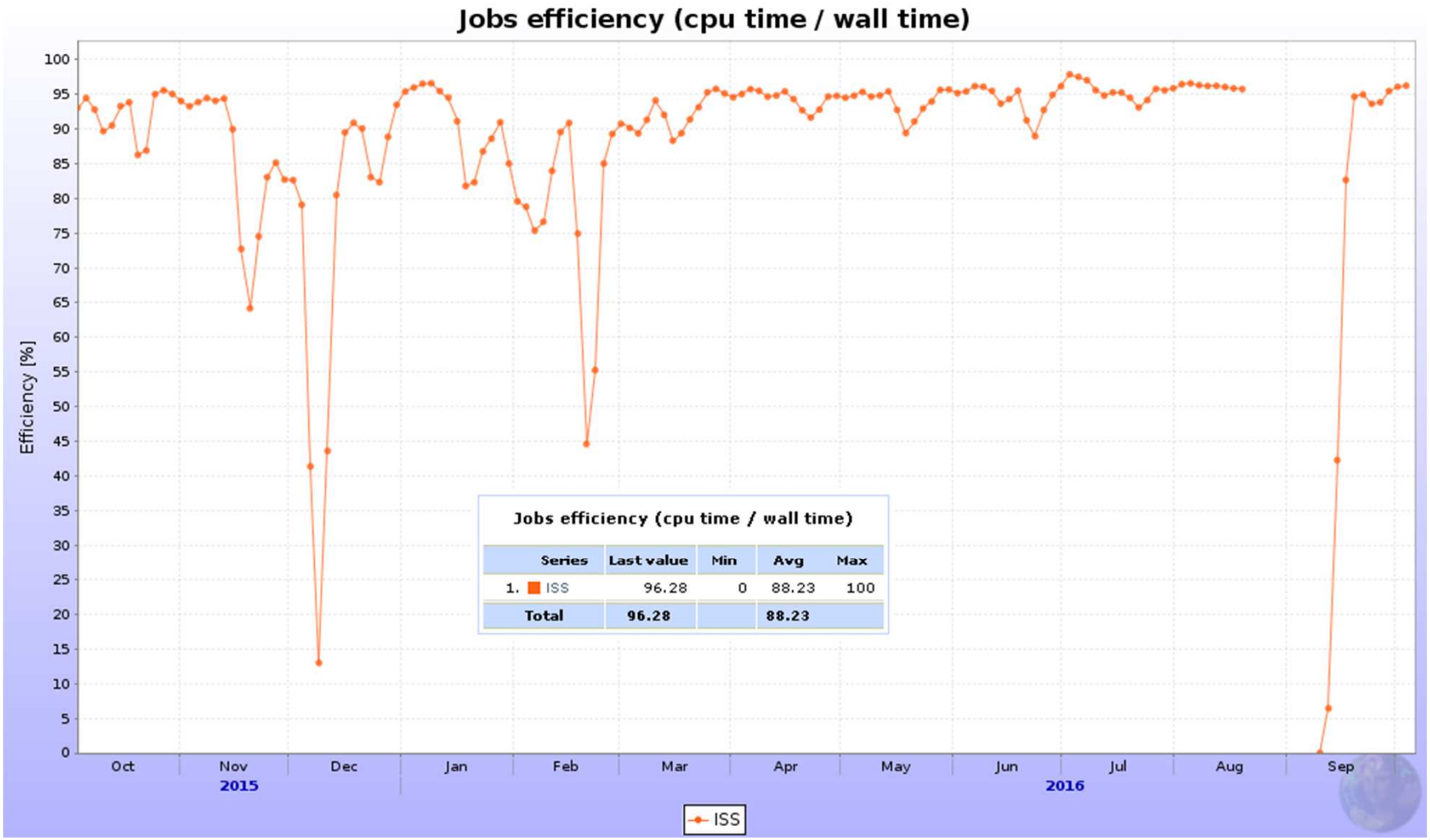
Statistics per site – ISS-Alice



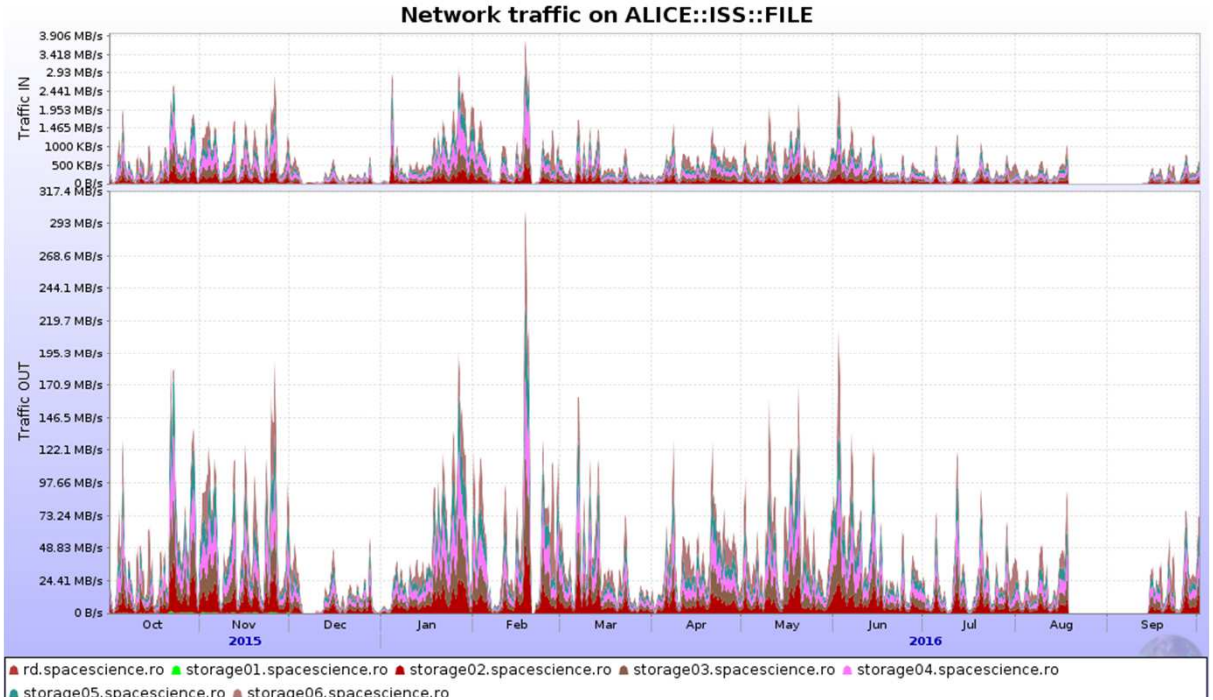
Statistics per site – ISS-Alice



Statistics per site – ISS-Alice



Storage performance



Traffic IN

	Series	Last value	Min	Avg	Max	Total
1.	rd.spacescience.ro	3.744 KB/s	0.116 KB/s	2.806 KB/s	7.172 MB/s	84.62 GB
2.	storage01.spacescience.ro	0.468 KB/s	0.102 KB/s	4.417 KB/s	635.8 KB/s	23.64 GB
3.	storage02.spacescience.ro	108.6 KB/s	0.111 KB/s	86.18 KB/s	12.44 MB/s	2.536 TB
4.	storage03.spacescience.ro	99.95 KB/s	0.107 KB/s	118.8 KB/s	12.84 MB/s	3.495 TB
5.	storage04.spacescience.ro	132.1 KB/s	0.111 KB/s	145.8 KB/s	14.46 MB/s	4.29 TB
6.	storage05.spacescience.ro	122 KB/s	80.58 B/s	114.9 KB/s	16.18 MB/s	3.38 TB
7.	storage06.spacescience.ro	128.1 KB/s	0.206 KB/s	152.5 KB/s	18.01 MB/s	4.486 TB
	Total	**595 KB/s**		**625.4 KB/s**		**18.29 TB**

Traffic OUT

	Series	Last value	Min	Avg	Max	Total
1.	rd.spacescience.ro	2.871 KB/s	0.102 KB/s	2.415 KB/s	78.39 KB/s	72.83 GB
2.	storage01.spacescience.ro	41.26 KB/s	0 B/s	316.5 KB/s	58.69 MB/s	1.654 TB
3.	storage02.spacescience.ro	13.77 MB/s	0.14 KB/s	6.771 MB/s	246.6 MB/s	204 TB
4.	storage03.spacescience.ro	12.53 MB/s	0.777 KB/s	9.163 MB/s	263.3 MB/s	276.1 TB
5.	storage04.spacescience.ro	16.1 MB/s	0.152 KB/s	9.63 MB/s	295.3 MB/s	290.1 TB
6.	storage05.spacescience.ro	15.08 MB/s	0.152 KB/s	8.789 MB/s	280.6 MB/s	264.8 TB
7.	storage06.spacescience.ro	14.74 MB/s	0.193 KB/s	10.84 MB/s	716.8 MB/s	326.6 TB
	Total	**72.27 MB/s**		**45.5 MB/s**		**1.331 PB**

PROBLEMS ENCOUNTERED

- Because of the aging hardware we encounter many stuck nodes due to insufficient memory relative to the actual needs of current software for analysis and reconstruction.

- The Information Systems and monitoring tools are fragmented and poorly documented and as such is difficult to decouple and interface with the said modules for providing complete and exact hardware and jobs status information. (E.g. GGUS ticket 123392)

- Due to wear there is a constant stream of failed hard-drives per year that require constant funding for replacement.

Thank you for your attention!