INP BSU grid site

V. Mossolov Dz. Yermak V. Yermolchyk

Research Institute for Nuclear Problems of Belarusian State University
Minsk, Belarus

14 September 2018

Laboratory of Fundamental Interactions

- Participation in developement of physical program and data analysis in modern and future particle accelerator experiments: CMS, ATLAS, CLIC and others.
- Calculation and analysis of observables in particle scattering processes including higher order effects.

ATLAS

INP BSU ATLAS group participates in a wide variety of physics studies[1]:

- Activities within the hadronic Tile Calorimeter working group
- Jet Energy Measurements
- Jet Cross Sections Measurements
- Parton Distribution Functions (PDF) constraints
- Physics Beyond Standard Model (BSM)

[1] https://indico.cern.ch/event/598242/contributions/2428265/attachments/1396942/2130005/INPATLASOverviewPPT.pdf

CMS

- Calculation of Radiative Corrections
- Monte-Carlo generator LPPG development
- Take part in Standart Model Physics group analyses: **Drell-Yan double differential contribution** and **Afb/Angular coefficients measurement**[1]

[1] https://twiki.cern.ch/twiki/bin/viewauth/CMS/AfbAngularWG13TeV

WLCG GRID site

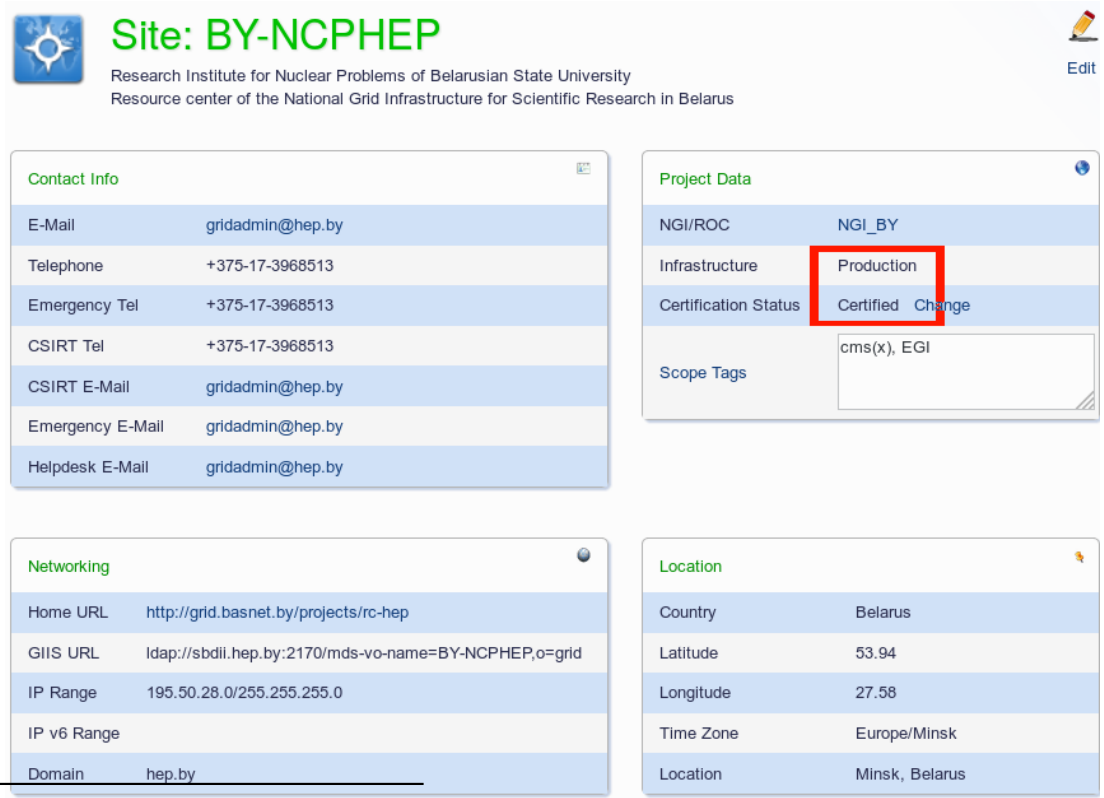
GRID site had been created for the needs of INP ATLAS and CMS groups.

Main usage:

- Long term single job calculations.
- Testing grid jobs before main submission.
- Priority job execution.
- Storage area for analysis output data.

INP BSU GRID site

- Grid site **BY-NCPHEP**[1] is part of **NGI-BY**.
- It is the only production and certified site of the Worldwide LHC Computing Grid in Belarus.



Site: BY-NCPHEP

Research Institute for Nuclear Problems of Belarusian State University
Resource center of the National Grid Infrastructure for Scientific Research in Belarus

Edit

Contact Info

E-Mail	gridadmin@hep.by
Telephone	+375-17-3968513
Emergency Tel	+375-17-3968513
CSIRT Tel	+375-17-3968513
CSIRT E-Mail	gridadmin@hep.by
Emergency E-Mail	gridadmin@hep.by
Helpdesk E-Mail	gridadmin@hep.by

Project Data

NGI/ROC	NGI_BY
Infrastructure	Production
Certification Status	Certified Change
Scope Tags	cms(x), EGI

Networking

Home URL	http://grid.basnet.by/projects/rc-hep
GIIS URL	ldap://sbdli.hep.by:2170/mds-vo-name=BY-NCPHEP,o=grid
IP Range	195.50.28.0/255.255.255.0
IP v6 Range	
Domain	hep.by

Location

Country	Belarus
Latitude	53.94
Longitude	27.58
Time Zone	Europe/Minsk
Location	Minsk, Belarus

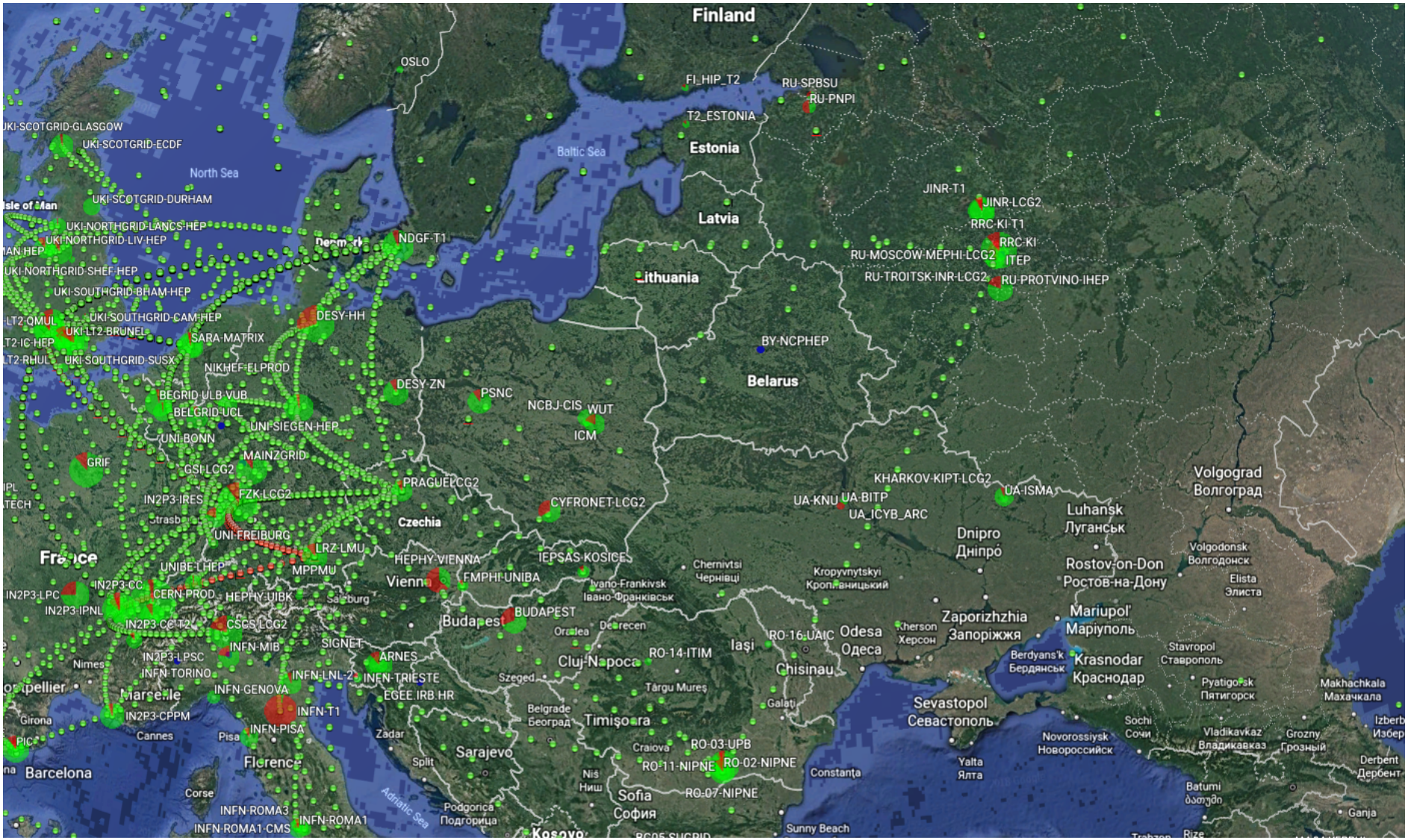
[1] https://goc.egi.eu/portal/index.php?Page_Type=Site&id=1225

INP BSU GRID site

- Provide support for **ATLAS** and **CMS** VO.
- Registered and certified by CMS as **T3_BY_NCPHEP**[1].

[1] https://cmsweb.cern.ch/sitedb/prod/sites/T3_BY_NCPHEP

Google Earth view



EGI nagios dashboard

https://argo-mon.egi.eu/nagios/

Current Network Status
Last Updated: Mon Feb 5 19:43:34 CET 2018
Updated every 90 seconds
Nagios® Core™ 4.3.1 - www.nagios.org
Logged in as */DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=vyermolc/CN=733165/CN=Vitaly Yermolchyk*

View History For This Host
View Notifications For This Host
View Service Status Detail For All Hosts

Host Status Totals

Up	Down	Unreachable	Pending
2	0	0	0

All Problems	All Types
0	2

Service Status Totals

Ok	Warning	Unknown	Critical	Pending
18	0	0	0	0

All Problems	All Types
0	18

Service Status Details For Host '*.by'

Limit Results: 100

Host	Service	Status	Last Check	Duration	Attempt	Status Information
grid02.hep.by	eu.egi.sec.StoRM-SL5	OK	02-05-2018 18:51:30	4d 6h 48m 14s	1/1	BDII_ENTRIES_NUM OK - 0 entries found
	eu.egi.sec.dCache-2.10	OK	02-05-2018 18:57:30	4d 6h 42m 14s	1/1	BDII_ENTRIES_NUM OK - 0 entries found
	eu.egi.sec.dCache-2.13	OK	02-05-2018 19:03:30	4d 6h 36m 14s	1/1	BDII_ENTRIES_NUM OK - 0 entries found
	hr.srce.SRM2-CertLifetime	OK	02-05-2018 18:39:30	3d 1h 4m 5s	1/2	CERT LIFETIME OK - Certificate will expire in 371.35 days (Feb 12 02:04:02 2019 GMT
	org.sam.SRM-All-ops	OK	02-05-2018 18:45:30	0d 0h 58m 2s	1/4	OK: success.
	org.sam.SRM-Del-ops	OK	02-05-2018 18:45:44	0d 0h 57m 48s	1/4	OK: File was deleted from SRM.
	org.sam.SRM-Get-ops	OK	02-05-2018 18:45:44	0d 0h 57m 48s	1/4	OK: File was copied from SRM. Diff successful.
	org.sam.SRM-GetSURLs-ops	OK	02-05-2018 18:45:31	3d 5h 47m 32s	1/4	OK: Got SRM endpoint(s) and Storage Path(s) from BDII
	org.sam.SRM-GetTURLs-ops	OK	02-05-2018 18:45:39	0d 0h 57m 53s	1/4	OK: protocols OK-[file, gsiftp]
	org.sam.SRM-Ls-ops	OK	02-05-2018 18:45:36	0d 0h 57m 56s	1/4	OK: listing [/ops/testfile-put-1517852731-42e2437b5ace.txt]-ok:
	org.sam.SRM-LsDir-ops	OK	02-05-2018 18:45:31	3d 5h 45m 34s	1/4	OK: Storage Path[/ops/]-ok:
	org.sam.SRM-Put-ops	OK	02-05-2018 18:45:35	0d 0h 57m 57s	1/4	OK: File was copied to SRM.
sbdii.hep.by	eu.egi.sec.Site-BDII-SL5	OK	02-05-2018 18:53:11	4d 6h 52m 33s	1/1	BDII_ENTRIES_NUM OK - 0 entries found
	org.bdii.Entries	OK	02-05-2018 18:59:11	4d 6h 46m 33s	1/4	OK: time=0.17s, entries=2
	org.bdii.Freshness	OK	02-05-2018 19:05:11	4d 6h 40m 33s	1/4	OK: freshness=116s, entries=1
	org.bdii.GLUE2-Validate	OK	02-05-2018 18:41:11	4d 9h 29m 19s	1/2	OK - errors 0, warnings 0, info 2
	org.nagios.BDII-Check	OK	02-05-2018 19:38:04	4d 6h 58m 33s	1/4	LDAP OK - 0.172 seconds response time
	org.nagios.GLUE2-Check	OK	02-05-2018 18:53:11	4d 6h 52m 33s	1/3	BDII_ENTRIES_NUM OK - 1 entry found

accounting.egi.eu statistic

Countries — Total number of jobs by Country and Month (All VOs)

Country	Jun 2018	Jul 2018	Aug 2018	Total
Algeria	2,418	4,182	5,253	**11,853**
Armenia	5	0	201	**206**
Australia	273,073	374,578	247,122	**894,773**
Austria	46,829	19,482	7,515	**73,826**
Azerbaijan	10,945	14,789	14,780	**40,514**
Belarus	4,744	16,218	13,336	**34,298**
Belgium	80,028	67,645	59,852	**207,525**
Brazil	127,355	116,174	112,117	**355,646**
Canada	1,008,416	1,253,521	1,169,938	**3,431,875**
Chile	44,697	63,963	151,853	**260,513**
China	167,490	197,710	186,237	**551,437**
Czech Republic	471,541	478,935	536,625	**1,487,101**
Denmark	1,196	1,885	3,345	**6,426**
Estonia	52,773	41,601	47,185	**141,559**
Finland	71,424	49,807	67,870	**189,101**

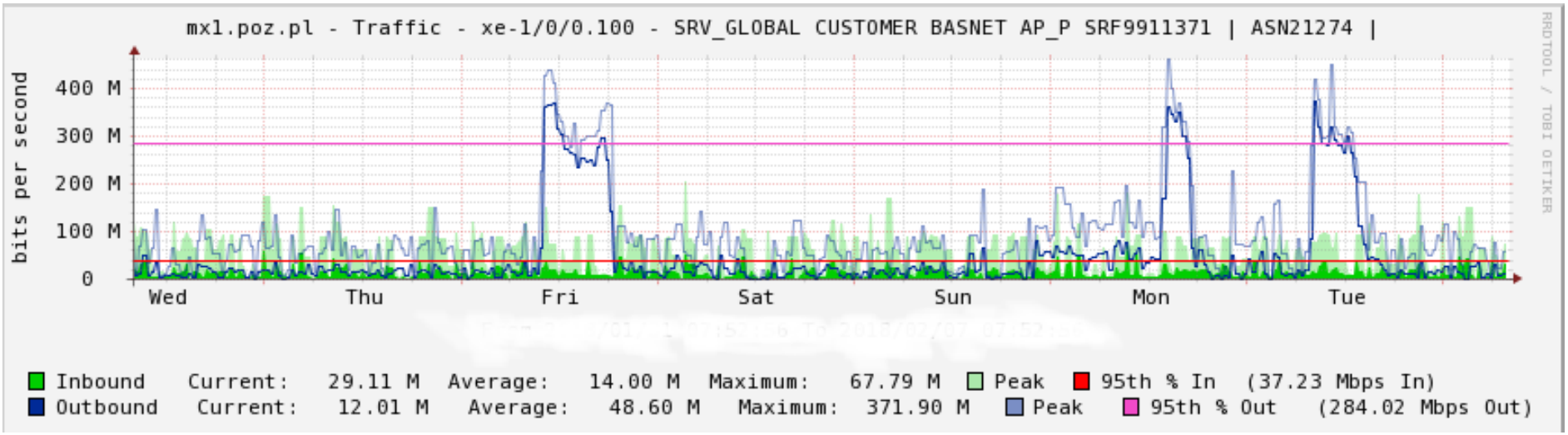
Previous limitations

- 50Mbit WAN

Network speed is too slow for work with CERN fts3 server. More than 50% of jobs failed with timeout error after waiting in queue.

Testing with GEANT network

Graphically	TaskName	NJobTotal	Pending	Running	Unknown	Cancelled	Success	Failed
	180206_073843:uyevarou_crab_crab_DataRun2016D-03Feb2017_ver2_v2	4860	0	0	0	0	4848	12
	180205_133220:uyevarou_crab_crab_DataRun2016C-03Feb2017_ver2_v2	2897	0	0	0	0	2897	0
	180202_093829:uyevarou_crab_crab_DataRun2016B-03Feb2017_ver2_v2	8715	0	0	0	0	8715	0

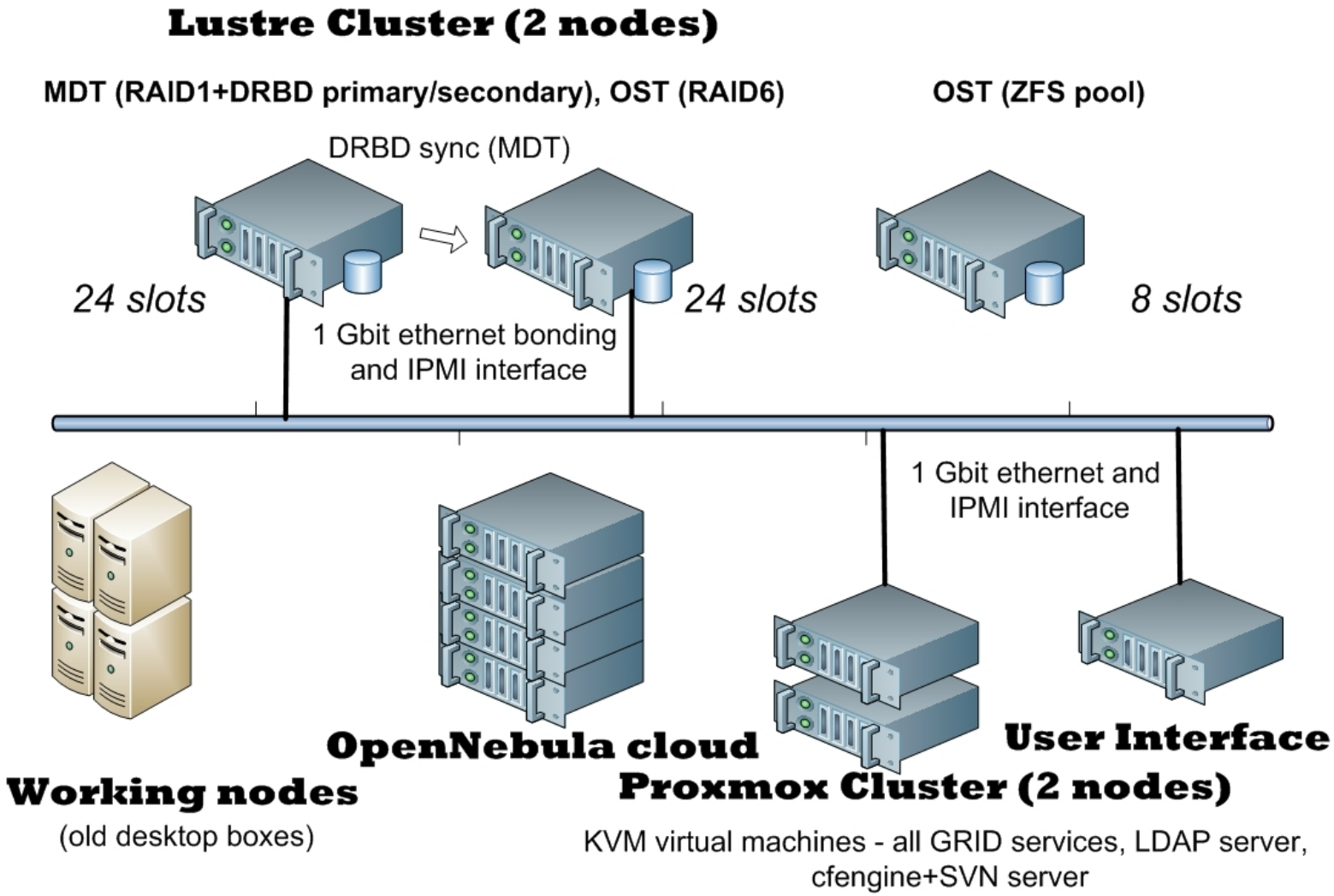


The part of successfully finished jobs has grown to 99.9% because no timeout error occurred!

Current state

- computational resources: 100 cores WN
- raid6 56TB storage, 55TB in use (+120TB in the next month)
- 1Gbit LAN
- 100Mbit WAN
- up to 1000Mbit WAN to GEANT network

Site scheme


Lustre Cluster (2 nodes)
MDT (RAID1+DRBD primary/secondary), OST (RAID6)
OST (ZFS pool)
DRBD sync (MDT)
24 slots
24 slots
8 slots
1 Gbit ethernet bonding
and IPMI interface
1 Gbit ethernet and
IPMI interface
OpenNebula cloud
User Interface
Working nodes
(old desktop boxes)
Proxmox Cluster (2 nodes)
KVM virtual machines - all GRID services, LDAP server,
cfengine+SVN server

Some Lustre on ZFS experience

Problem:

yum update was running for several hours in VM instance (kvm hypervisor in Opennebula cloud) if VM disk image is located on Lustre

More investigation:

opennebula_host# dd if=/dev/zero of=testfile bs=4K count=8000 oflag=sync
where testfile is placed on

Lustre (RAID6)	2.1 MB/s
local ZFS filesystem	460 kB/s
Lustre (ZFS)	26.9 kB/s

Symptoms of the problem disappeared:

when “unsafe” cache mode for VM drive enabled (= reduce sync operations)

Supposed problem source:

LU-4009 opened issue: "Add ZIL support to osd-zfs"

In order to improve sync performance on ZFS based OSDs Lustre must be updated to utilize a ZFS Intent log (ZIL)

Workaround:

for VM drive: use qcow2 base+derrived pair of images instead of single image.

Conclusions

- BY-NCPHEP is production and certified grid site of WLCG.
- It provide computational power for data analysis of LHC experiments.
- Also it can be used as testing site for new technologies.
- It is integrated to JINR cloud.
- Extension of storage system is in the first priority to provide area for analysis output data.