IHEP Tier-2 computing center: status and operation

V. Kotliar[a], V. Gusev[b], V. Kukhtenkov[c], N. Savin[d]

National Research Center "Kurchatov Institute" State Research Center of Russian Federation Institute for High Energy Physics, Protvino, Russia

E-mail: [a] Viktor.Kotliar@ihep.ru, [b] Victor.Gusev@ihep.ru, [c] kvi@ihep.ru, [d] Nikolay.Savin@ihep.ru

RU-Protvino-IHEP site is the one of three biggest WLCG Tier-2 centers in Russia. The computing infrastructure serves for "big four" LHC high energy physics experiments such as Atlas, Alice, CMS, LHCb and local experiments at IHEP such as OKA, BEC, radio biology stands and others. In this work the current status of the computing capacities, networking and engineering infrastructure is shown as well as the contribution of the grid site to the collaboration experiments.

Keywords: WLCG, GRID-computing, Tier-2

Introduction

RU-Protvino-IHEP site participates in the Worldwide LHC Computing Grid from the very beginning since 2003. At that time there were installed and configured the first grid infrastructure services like CE, SE, WNs, UI on 16 two-core Pentium III 900MHz. LCG (LHC Computing Grid) and EDG (European Data Grid) grid middleware was used [Kotlyar, Gusev, …, 2012]. IHEP participated in the EGEE I – EGEE III projects in NA2, NA3, NA4, SA1 [EGEE homepage] activities. After increasing network bandwidth to 100Mb/s, then to 1Gb/s and in the end to 10Gb/s grid site in Protvino became one of the biggest Tier-2 site in Russia after JINR with ~3k CPU (25k HEP-SPEC06) and 2PB disks space.

At the present time our site serves for big four LHC experiments (Atlas, Alice, CMS, LHCb) and many physical experiments inside the Institute. We implement shared CPU schema that allows achieving 24x7 CPU resource usage and it makes resources usage more effective.

Site overview

IHEP cluster is a cluster with shared CPU resources among Grid and Internal experiments. In the core of the cluster is the Kerberos5 “Single Sign on” system. It is used by local users to get an access to the resources of the cluster directly without any grid services. It means that a batch system on the cluster supports Kerberos tickets and these tickets are forwarded to all Working Nodes too. As soon as we also use Andrew File System for user home directories cluster must provides support for AFS tokens which are based on Kerberos tickets.

The next major service is a Lustre parallel cluster file system which is shared across all working nodes to be able to allow local users to perform data analysis as faster as possible.

Both for the long data store and for the store of the RAW data we use CASTOR [CASTOR home page] (the CERN Advanced STORage manager). It is a HSM system that has been using in IHEP for ten years. We store data on LTO4 and LTO5 tapes with tape libraries and stand alone tape drives.

The main administration system on the cluster is Puppet with FAI boot installation.

Almost all grid-services including site BDII, APEL, VO BOX CMS, Puppet and some others are ran under Xen hypervisors on the high availability system as virtual machines. This technique allows to use resources as much as needed for the current setup and dynamically adjust them if it will be necessary in the future.

There are several internal servers that are used for the site infrastructure. They are: network gateway servers which provide NAT (network address translation or masquerade) for the site internal network; DNS (Domain Name Servers) for this internal network and for caching DNS queries on the cluster; squid http proxy servers for caching CVMFS [CVMFS home page] (CERN Virtual Machine File System) requests and special frontier proxies for CMS and ATLAS to cache Oracle DB (data base) queries of experiments to the central repositories; also there are GPU Nvidia Tesla computing systems.

Grid middleware is EMI3 and base OS (operating system) on the cluster is Scientific Linux 6 64bit. For GPU’s we use CUDA 7.5. There are Ansys 15.0 and Mathematica 8.0 installed on several nodes on the cluster and Intel® Fortran Composer XE 64bit is installed on UI (User Interface). All this software allows to use more flexibly computing farm at IHEP as multi-purpose computing cluster.

Grid resources of the site

Grid center at IHEP has storage and computing resources which it declares for usage to the WLCG infrastructure. At the moment there are:

- 2828 CPU, 24390 HEP-SPEC06;
- 1942 TB: Atlas 1185, CMS 395, Alice 297, LHCb 65;
- 2x10Gb/s Internet channels (LHCONE shared with RDIG 10Gb/s);
- manpower is six system administrators;

All four LHC experiments have their own disk resources and CPU resources shared among all of them on fair share basis. The main storage technologies is used on site are dCache and xrootd which allow to build distributed storage systems for grid-computing [Kotliar, Latyshev, …, 2013]. An access to outside world is done by two 10Gb/s links where IHEP is connected to the LHCONE computing network for scientific research. Manpower for site support is six system administrators. And site counts as a third big Grid site in Russia (figure 1).

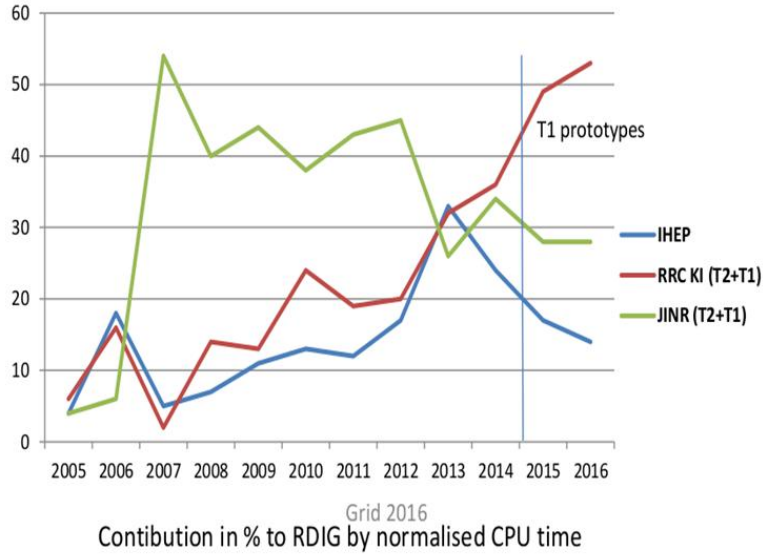

Figure 1. A history of the contribution for big Russian Grid-sites

IHEP Grid VO usage

As soon as computing cluster at IHEP is shared among four Grid experiments following fair-share schema implemented for CPU resources: ATLAS 52%, CMS 30%, ALICE 12%, LHCb 6%. It needs to be taken into account that CPU also shared with local IHEP experiments like BEC, OKA, etc. So the real usage for the last year represents as: ATLAS 44%, CMS 17%, Alice 32%, LHCb 7%. The main consumer for the site is the ATLAS experiment where RU-Protvino-IHEP site has a status of Tier-2D alpha which means Tier2 site with availability and accessibility more than 95%. The site contributes with 30% of all ATLAS jobs in RU-cloud, 40% of all RU-cloud users data and 20% of all RU-cloud data stored at IHEP. For CMS the percent of jobs computed on site is around 10% inside Russian segment. For Alice it is a 12% and for LHCb 9%. For LHCb IHEP has also a status of the site Tier-2D (with data) and we only Tier2 site with such status in Russia.

IHEP site shows really good stability and has cpu usage efficiency near 90% (figure 2) which is above average cpu efficiency for Grid sites in WLCG.

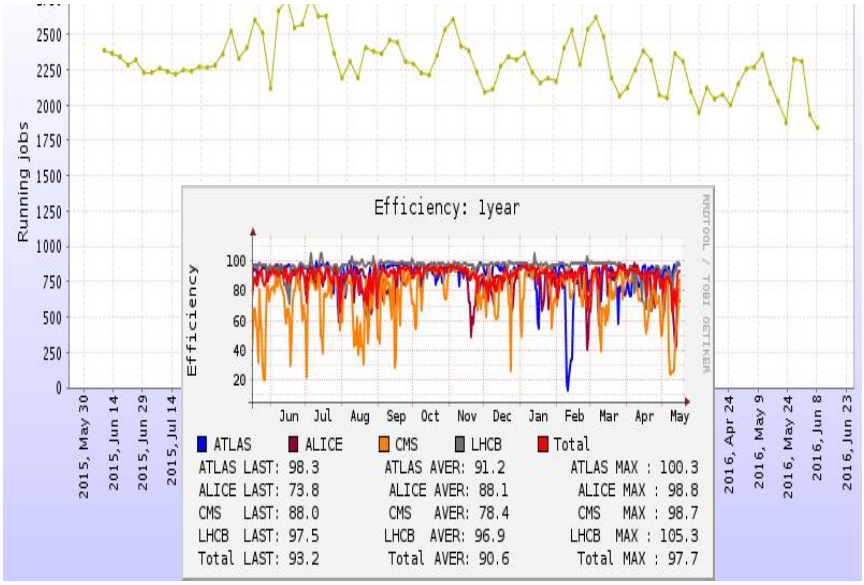

Figure 2. Running jobs and efficiency per year

As benefit of the following works performed on the site it was possible to achieve such parameters and smooth run in 2016:

- two new cooling systems were installed at IHEP data center;
- additional manpower was added for site monitoring activities;
- many upgrades for grid middleware and system software was performed without site outages;
- more works were focused in supporting current infrastructure availability, reliability and efficiency;
- external network connectivity was completely changed by using another Internet service provider and by connecting to LHCONE network for scientific research.

Future plans

Future plans more focused on works available without increasing resources of the computer center. As it is possible easy to add resources to the current infrastructure without any additional modifications it is only matter of money to increase site's computing or storage resources. For the future works it is a better integration to the LHCONE network. This depends on modernization of the current network infrastructure at IHEP and splitting cluster and campus networks at IHEP. The next major activity is a system software and Grid middleware upgrades with minimum outage on the site. All storage systems has to be upgraded to the new version (gold release) of dCache to be able to run in LHC run 3. More works focused on supporting current level of availability, reliability, efficiency and leveling it up when possible. For such activity more “smart” monitoring tools is going to be deployed for better understanding the complex infrastructure.

Conclusion

The WLCG Grid site at IHEP has a big potential of development for the future and it has established reputation of the site with high availability and reliability in Russian segment of the Grid infrastructure. We use leading technologies in the computing science for computing clusters with batch systems to run jobs. There are many investigations and approbations inside the site infrastructure like

cloud technology and GPU computing which are going to be a future in data centers. The generic goal is to be the best Tier-2 in Russia.

References

Kotlyar V., Gusev V., Kukhtenkov V., Popova E., Savin N., Soldatov A. WLCG Tier-2 computing infrustructure at IHEP. // Distributed computing and grid-technologies in science and education (Grid`2012). — 2012. — P. 150-157.

Kotlyar V., Latyshev G., Popova E., Yutalova A. IHEP Data Storage Systems for Experiments. // Proceedings of XXIV International Symposium on Nuclear Electronics & Computing (NEC`2013). — 2013. — P. 166-172.

EGEE homepage [Electronic resource]: http://public.eu-egee.org/.

CASTOR homepage [Electronic resource]: http://castor.web.cern.ch/.

CVMFS homepage [Electronic resource]: http://cernvm.cern.ch/portal/cvmfs/release-2.0.