HEPiX Fall 2014 Workshop

Monday, 13 October 2014 - Friday, 17 October 2014

University of Nebraska - Lincoln

Book of Abstracts

Contents

Site Reports / **1**

INFN-T1 site report

Author: Andrea Chierici[1]

[1] *INFN-CNAF*

Corresponding Author: chierici@cnaf.infn.it

Updates on INFN Tier1 site

Summary:

Updates on INFN Tier1 site

Storage and Filesystems / **2**

Experience in running relational databases on clustered storage

Author: Ruben Domingo Gaspar Aparicio[1]

[1] *CERN*

Corresponding Author: ruben.gaspar.aparicio@cern.ch

CERN IT-DB group is migrating its storage platform, mainly NetApp NAS's running on 7-mode but also SAN arrays, to a set of NetApp C-mode clusters. The largest one is made of 14 controllers and it will hold a range of critical databases from administration to accelerators control or experiment control databases. This talk shows our setup: network, monitoring, use of features like transparent movement of file systems, flash pools (SSD + HDD storage pools), snapshots, etc. It will also show how these features are used on our infrastructure to support backup & recovery solutions with different database solutions: Oracle (11g and 12c multi tenancy), MySQL or PostgreSQL. Performance benchmarks and experience collected while running services on this platform will be also shared. It will be also covered the use of the cluster to provide iSCSI (block device) access for OpenStack windows virtual machines.

Basic IT Services / **3**

DataBase on Demand : insight how to build your on DBaaS

Author: Ruben Domingo Gaspar Aparicio[1]

[1] *CERN*

Corresponding Author: ruben.gaspar.aparicio@cern.ch

Inspired on different database as a service, DBaas, providers, the database group at CERN has developed a platform to allow CERN user community to run a database instance with database administrator privileges providing a full toolkit that allows the instance owner to perform backup/point in time recoveries, monitoring specific database metrics, start/stop of the instance and uploading/downloading specific logging or configuration files. With about 150 instances Oracle (11g and 12c), MySQL and PostgreSQL the platform has been designed and proofed to be flexible to run different RDBMS vendors and to scale up.

Initially running on virtual machines, OracleVM, the instances are represented as objects in the management database toolset, making it independent of its physical representation. Nowadays instances run on physical servers together with virtual machines. A high availability solution has been implemented using Oracle cluster ware.
This talk explains how we have built this platform, different technologies involved, actual user interface, command execution based on a database queue, backups based on snapshots, and possible future evolution (Linux containers, storage replication, OpenStack, Puppet,...).

Site Reports / **4**

Fermilab Site Report - Fall 2014 HEPiX

Author: Keith Chadwick[1]

[1] *Fermilab*

Corresponding Author: chadwick@fnal.gov

Fermilab Site Report - Fall 2014 HEPiX

Summary:

Fermilab Site Report - Fall 2014 HEPiX

Site Reports / **5**

CERN Site Report

Author: Arne Wiebalck[1]

[1] *CERN*

Corresponding Author: arne.wiebalck@cern.ch

News from CERN since the Annecy meeting.

Grids, Clouds, Virtualisation / **6**

Addressing the VM IO bottleneck

Author: Arne Wiebalck[1]

[1] *CERN*

Corresponding Author: arne.wiebalck@cern.ch

This is summary of our efforts to address the issue of providing sufficient IO capacity to VMs running in our OpenStack cloud.

Grids, Clouds, Virtualisation / **7**

CERN Cloud Report

Author: Arne Wiebalck[1]

[1] *CERN*

Corresponding Author: arne.wiebalck@cern.ch

This is a report on the current status of CERN's OpenStack-based Cloud Infrastructure.

IT Facilities and Business Continuity / **8**

First Experience with the Wigner Data Centre

Author: Wayne Salter[1]

[1] *CERN*

Corresponding Author: wayne.salter@cern.ch

After a tender for a CERN remote Tier0 centre issued at the end of 2011, and awarded to the Wigner Data Centre in May 2012, operations commenced at the beginning of 2013. This talk will give a brief introduction to the history of this project and it scope. It will then summarise the initial experience that has been gained to-date and highlight a number of issues that have been encountered; some maybe expected but others not.

Basic IT Services / **9**

Ermis service for DNS Load Balancer configuration

Authors: Aris Angelogiannopoulos[1]; Ignacio Reguero[2]

[1] *Ministere des affaires etrangeres et europeennes (FR)*

[2] *CERN*

Corresponding Author: aris.angelogiannopoulos@cern.ch

This presentation describes the implementation and use cases of the Ermis Service. Ermis is a RESTful service to manage the configuration of DNS load balancers. It enables direct creation and deletion of DNS delegated zones using a SOAP interface provided by the Network group thus simplifying the procedure needed for supporting new services. It is written in Python as a Django Application. This is quite generic and can be easily adapted to other types of Load Balancers. Ermis is being integrated with Openstack. It uses the Openstack Keystone API as a means of authentication and authorization as well as Kerberos and e-groups. The ultimate aim of the project is to provide Load Balancing as a Service (LBaaS) to the end users of the cloud.

Site Reports / **10**

Australia Site Report

Author: Sean Crosby[1]

[1] *University of Melbourne (AU)*

Corresponding Author: sean.christopher.crosby@cern.ch

An update on the ATLAS Tier 2 and distributed Tier 3 of HEP groups in Australia. Will talk about our integration of Cloud resources, Ceph filesystems and integration of 3rd party storage into our setup

IT End User and Operating Systems / **11**

Next Linux Version at CERN.

Author: Thomas Oulevey[1]

Co-author: Jarek Polok [1]

[1] *CERN*

Corresponding Author: thomas.oulevey@cern.ch

CERN is maintaining and deploying Scientific Linux CERN since 2004.

In January 2014 CentOS and Red Hat announced joining forces in order to provide common platform for open source community project needs.

CERN decided to see how CentOS 7 fits his needs and evaluate CentOS release 7 as their next version.

An updated report will be provided, as agreed at HEPiX Spring 2014.

Storage and Filesystems / **12**

DPM performance tuning hints for HTTP/WebDAV and Xrootd

Author: Fabrizio Furano[1]

Co-authors: Adrien Devresse [1]; Alejandro Alvarez Ayllon [1]; Andrea Manzi [1]; David Smith [1]; Ivan Calvet [1]; Martin Philipp Hellmich [1]; Oliver Keeble [1]

[1] *CERN*

Corresponding Author: andrea.manzi@cern.ch

In this contribution we give a set of hints for the performance tuning of the upcoming DPM releases, and we show what one can achieve by looking at different graphs taken from the DPM nightly performance tests.
Our focus is on the HTTP/WebDAV and Xrootd protocols and the newer “dmlite” software framework, and some of these hints may give some benefit also to older, legacy protocol implementations.
Our goal is to make sure that single-VO and multi-VO DPM sites can join HEP and non-HEP computing models and HTTP and Xrootd federations, while giving the needed level of performance and the best system administration experience.

Grids, Clouds, Virtualisation / **13**

Evolution of WLCG monitoring

Authors: Alexandre Beche[1]; Cristovao Jose Domingues Cordeiro[1]; David Tuckett[1]; Edward Karavakis[1]; Hassen Riahi[1]; Hector Martin De Los Rios Saiz[1]; Ivan Antoniev Dzhunov[1]; Ivan Kadochnikov[2]; Julia Andreeva[1]; Lionel Cons[1]; Luca Magnoni[1]; Marian Babik[1]; Pablo Saiz[1]; Sergey Belov[2]; Uthayanath Suthakar[3]

[1] *CERN*

[2] *Joint Inst. for Nuclear Research (RU)*

[3] *Brunel University*

Corresponding Author: edward.karavakis@cern.ch

The WLCG monitoring system provides a solid and reliable solution that has supported LHC computing activities and WLCG operations during the first years of LHC data-taking. The current challenge consists of ensuring that the WLCG monitoring infrastructure copes with the constant increase of monitoring data volume and complexity (new data-transfer protocols, new dynamic types of resource providers - cloud computing). At the same time, simplification of the monitoring system is desirable in order to reduce maintenance and operational costs.
The current evolution of the system aims to achieve these two goals: decrease the complexity of the system and ensure its scalability and performance with the steady increase of monitoring information. The presentation will describe the new WLCG monitoring platform including the new technology stack for large-scale data analytics.

IT End User and Operating Systems / **14**

FTS3, large scale file transfer service with simplicity and performance at its best

Author: Michail Salichos[1]

Co-authors: Alejandro Alvarez Ayllon [1]; Andrea Manzi [1]; Michal Kamil Simon [1]; Oliver Keeble [1]

[1] *CERN*

Corresponding Author: michail.salichos@cern.ch

FTS3 is the service responsible for globally distributing the majority of the LHC data across the WLCG infrastructure. It is a file transfer scheduler which scales horizontally and it's easy to install and configure. In this talk we would like to bring the attention to the FTS3 features that could attract wider communities and administrators with several new friendly features. We will present both the new tools for the management of the FTS3 transfer parameters, for instance bandwidth-limits, max active transfers per endpoint and VO, banning users and endpoints, plus new Data Management operations (deletions and staging files from archive) easily accessed via REST-API. In addition we will also showcase the new captivating FTS3 Graphical Interface for end-users to manage their transfers (WebFTS) together with the new activities to extend the FTS3 transfers capabilities outside the grid boundaries (Dropbox, S3, etc.) In this manner we demonstrate that FTS3 can cover the needs from casual users to high load services.

Summary:

The evolution of FTS3 is addressing the technical and performance requirements and challenges for LHC RUN2, moreover, its simplicity, generic design, web portal and REST interface makes it an ideal file transfer scheduler both inside and outside HEP

Grids, Clouds, Virtualisation / **15**

The Adoption of Cloud Technologies within the LHC Experiments

Author: Laurence Field[1]

[1] *CERN*

Corresponding Author: laurence.field@cern.ch

The adoption of cloud technologies by the LHC experiments is currently focused on IaaS, more specifically the ability to dynamically create virtual machines on demand.
This talk provides an overview of how this alternative approach for resource provision fits into the existing workflows used by the experiments.
It shows that in order to fully exploit this approach, solutions are required in the areas of image management, capacity management, monitoring, accounting, pilot job frameworks and supporting services.
Each of those areas is covered in more detail to explain the various architectural choices and the rational for the decisions made.
Finally, a general overview of the state of adoption within each experiment is given that describes the ongoing integration with their individual frameworks.

Computing and Batch Systems / **16**

Compute node benchmarks for Compact Muon Solenoid workflows

Author: Samir Cury Siqueira[1]

Co-author: Dorian Kcira [1]

[1] *California Institute of Technology (US)*

Corresponding Author: samir.cury.siqueira@cern.ch

Hardware benchmarks are often relative to the target application. In CMS sites, new technologies, mostly processors, need to be evaluated on an yearly basis. A framework was developed at the Caltech CMS Tier-2 to benchmark compute nodes with one of the most CPU-intensive CMS workflows - The Tier-0 Reconstruction.

The benchmark is a CMS job that reports the results to a central database based on CPU model and makes them available to real-time monitoring web interfaces. The goal is to provide to the collaboration a reference for CPU performance, which can also be used in automated systems through an API. The jobs run in parallel to normal Grid activity and could have their submission and reporting automated.

Networking and Security / **17**

Situational Awareness: Computer Security

Author: Stefan Lueders[1]

[1] *CERN*

Corresponding Author: stefan.lueders@cern.ch

Computer security is important as ever outside the HEP community, but also within. This presentation will give the usual overview on recent issues being reported or made public since the last HEPix workshop (like the ripples of "Heartbleed"). It will discuss trends (identity federation and virtualisation) and potential mitigations to new security threats.

Networking and Security / **18**

EEX: ESnet Extension to Europe and ESnet support for the LHC Community

Author: Joe Metzger[1]

[1] *LBL*

Corresponding Author: metzger@es.net

The ESnet Extension to Europe (EEX) project is building out the ESnet backbone in to Europe. The goal of the project is to provide dedicated transatlantic network services that support U.S. DOE funded science.

The EEX physical infrastructure build will be substantially completed before the end of December. Initial services will be provided to BNL, FERMI and CERN while the infrastructure is being built out and tested. EEX services will be expanded, following the build, to serve all current ESnet sites and the U.S. LHC community including some computing centers at U.S. Universities.

This talk will cover the EEX architecture, schedule, and services, and include initial thoughts about what universities will need to do to take advantage of this opportunity.

Summary:

This talk will cover the EEX architecture, schedule, and services, and include initial thoughts about what universities will need to do to take advantage of this opportunity.

Site Reports / **19**

Site report: NDGF-T1

Author: Ulf Tigerstedt[1]

[1] *CSC Oy*

Corresponding Author: mattias.wadenstein@cern.ch

Site report for NDGF-T1, mainly focusing on dCache.

Site Reports / **20**

Oxford Particle Physics Computing update

Author: Peter Gronbech[1]

[1] *University of Oxford (GB)*

Corresponding Author: p.gronbech1@physics.ox.ac.uk

Site report from the University of Oxford Physics department.

Summary:

Site report from the University of Oxford Physics department.

Basic IT Services / **21**

Puppet at USCMS-T1 and FermiLab - Year 2

Author: Timothy Michael Skirvin[1]

[1] *Fermi National Accelerator Lab. (US)*

Corresponding Author: tskirvin@fnal.gov

USCMS-T1's work to globally deploy Puppet as our configuration management tool is well into the "long tail" phase, and has changed in fairly significant ways since its inception. This talk will discuss what has worked, how the Puppet tool itself has changed over the project, and our first thoughts as to what we expect to be doing in the next year (hint: starting again is rather likely!).

Computing and Batch Systems / **22**

Future of Batch Processing at CERN: a Condor Pilot Service

Author: Jerome Belleman[1]

[1] *CERN*

Corresponding Author: jerome.belleman@cern.ch

The CERN Batch System comprises 4000 worker nodes, 60 queues and offers a service for various types of large user communities. In light of the developments driven by the Agile Infrastructure and the more demanding processing requirements, it will be faced with increasingly challenging scalability and flexibility needs.

Last HEPiX, we presented the results of our evaluation of SLURM, Grid Engine derivatives and HTCondor. The latter being the most promising one, we started setting up an HTCondor pilot service. This talk will present the initial functions provided by this service. We will then discuss the development steps we took to implement them. Finally, we will name the next features which will ultimately be leading to the new production CERN Batch Service.

Basic IT Services / **23**

CFEngine Application at AGLT2

Author: Ben Meekhof[1]

[1] *University of Michigan*

CFEngine is a highly flexible configuration management framework. It also has a very high learning curve which can sometimes make decisions about how to deploy and use it difficult. At AGLT2 we manage a variety of different systems with CFEngine. We also have an effective version-controlled workflow for developing, testing, and deploying changes to our configuration. The talk will demonstrate what a practical and useful CFEngine infrastructure might look like. It will also include examples of how we organize our policy and effectively use the CFEngine policy language.

IT Facilities and Business Continuity / **24**

UPS Monitoring with Sensaphone-A cost-effective solution

Authors: Alexandr Zaytsev[1]; Christopher Hollowell[2]; Christopher Lee[2]; Tony Wong[2]; William Strecker-Kellogg[3]

[1] *Brookhaven National Laboratory (US)*

[2] *Brookhaven National Laboratory*

[3] *Brookhaven National Lab*

We describe a cost-effective indirect UPS monitoring system that was implemented recently in parts of its RACF complex. This solution was needed to address a lack of centralized monitoring solution, and it is integrated with an event notification mechanism and overall facility management.

Basic IT Services / **25**

Configuration Services at CERN: update

Author: Ben Jones[1]

[1] *CERN*

Corresponding Author: ben.dylan.jones@cern.ch

A status of the Puppet-based Configuration Service at CERN will presented giving a general update and discussing our current plans for the next 6 months.
The presentation will also highlight the work being done to secure the Puppet infrastructure making it appropriate for use by a large number of administratively distinct user-groups.

Networking and Security / **26**

Do You Need to Know Your Users?

Author: Bob Cowles[1]

[1] *BrightLite Information Security*

Corresponding Author: bob.cowles@gmail.com

After several years investigation of trends in Identity Management (IdM), the eXtreme Scale Identity Management (XSIM) project has concluded there is little reason for resource providers to provide IdM functions for research collaborations or even for many groups within the institution. An improved user experience and decreased cost can be achieved with "a small amount of programming."

Networking and Security / **27**

Plans for Dual Stack IPv4/IPv6 services on WLCG - an update from the HEPiX IPv6 Working Group

Author: Dave Kelsey[1]

[1] *STFC - Rutherford Appleton Lab. (GB)*

Corresponding Author: david.kelsey@stfc.ac.uk

This talk will present an update on the recent activities of the HEPiX IPv6 Working Group including our plans for moving to dual-stack services on WLCG.

IT End User and Operating Systems / **28**

Issue Tracking and Version Control Services status update

Author: Borja Aparicio Cotarelo[1]

Co-authors: Alvaro Gonzalez Alvarez [1]; Anna Trzcinska [2]; David Asbury [1]; Georgios Koloventzos [1]; Nils Hoimyr [1]; Terje Andersen [1]

[1] *CERN*

[2] *Warsaw University of Technology (PL)*

Corresponding Author: borja.aparicio.cotarelo@cern.ch

The current efforts around the Issue Tracking and Version Control services at CERN will be presented. Their main design and structure will be shown giving special attention to the new requirements from the community of users in terms of collaboration and integration tools and how we address this challenge in the definition of new services based on GitLab for collaboration and Code Review and Jenkins for Continuous Integration.

The presentation will also address infrastructure issues for these services, such as the back-end storage, administration portal "Cernforge" for power-users and experience using Puppet for managing the server software configuration.

IT End User and Operating Systems / **29**

Monviso: a portal for metering and reporting CNAF resources usage

Author: Giuseppe Misurelli[1]

Co-author: Andrea Chierici [2]

[1] *INFN*

[2] *INFN-CNAF*

Corresponding Author: chierici@cnaf.infn.it

CNAF T1 monitoring and alarming systems produce tons of data describing state, performance and usage of our resources. Collecting this kind of information centrally would benefit both resource administrators and our user community in processing information and generating reporting graphs. We built the "Monviso reporting portal" that consumes a set of key metrics, graphing them based on two main viewpoints: resource administration and experiments support.
The resulting portal is a lightweight charts gallery used also by our operator-on-call.

Computing and Batch Systems / **30**

Benchmarking on System on Chip Archtecture and fast benchmarking

Author: Michele Michelotto[1]

[1] *Universita e INFN (IT)*

Corresponding Author: michele.michelotto@cern.ch

The traditional architecture for High Energy Physics is x86-64 but in the community there is interest in processor more efficient in term of computing power per Watt. I'll show my measurement on ARM and Avoton processor.
I'll conclude with some measurements on candidate for fast benchmark that are requested by the physics community, mostyl to measure the performance of machine in cloud.

Storage and Filesystems / **31**

New High Availability Storage at PDSF

Author: Tony Quan[1]

Co-authors: Iwona Sakrejda ; James Botts [2]; Larry Pezzaglia [2]

[1] *LBL*

[2] *LBNL*

Corresponding Author: twquan@lbl.gov

The PDSF Cluster at NERSC has been providing a data-intensive computing resource for experimental high energy particle and nuclear physics experiments (currently Alice, ATLAS, STAR, ICECUBE, MAJORANA) since 1996. Storage is implemented as a GPFS cluster built out of a variety of commodity hardware (Dell, Raidinc, Supermicro storage and servers). Recently we increased its capacity by 500TB by adding two file systems using NetApp E5600 series storage directly SAS attached to a pair of servers in a high availability configuration. Data IO routes to the cluster through dual 10Gb Ethernet. A 1Gb private network is used for monitoring and management.

We will describe the configuration, share observations from the deployment process and provide the initial performance results. One of the new file systems was used to replace the back-end of the Tier3 ATLAS Storage Element (Bestman in gateway mode). We will share our experiences related to that move.

Basic IT Services / **32**

Developing Nagios code to suspend checks during planned outages

Author: Ray Spence[1]

[1] *u*

Corresponding Author: respence@lbl.gov

Lawrence Berkeley National Laboratory/NERSC Division

Developing Nagios code to suspend checks during planned outages.
Raymond E. Spence

NERSC currently supports more than 13,000 computation nodes spread over six supercomputing or clustered systems. These systems access cumulatively more than 13.5PB of disk space via thousands of network interfaces. This environment enables scientists from anywhere on the planet to login, run code and thereby to conduct science at elite levels. Scientists depend on NERSC for 24x7 availability and NERSC personnel in turn depend on industrial-strength system administration tools for our support efforts. Since monitoring everything from our largest system to the last network uplink is a chief concern at NERSC we chose several years ago to employ Nagios for our monitoring solution. Nagios is a mature product with a great degree of flexibility. Although NERSC has found the free, open source Nagios version sufficient in many ways we had eventually tired of one specific hole in this tool's arsenal. The hole NERSC found in Nagios' configuration involves planned downtime.

Any Nagios user eventually comes to know where to point and click to acknowledge alerts and twiddle other Nagios switches. However, when it comes to running large systems with multiple monitored services per node, point and click solutions do not scale. Like any supercomputing center NERSC has many planned downtimes of varying size throughout the year. Unfortunately we found no obvious path to configure Nagios to temporarily turn off checks on a to-be downed resource. NERSC then began writing code to communicate directly with Nagios to suspend these checks. Over the past year NERSC has produced scripts which configure Nagios to respectively obey a planned downtime, remove a planned downtime and to list scheduled downtimes. Further, each downtime can cover any number of services running on any number of nodes. We used our dedicated Physics cluster, PDSF, as our test bed and first production system for the scripts. Managing planned outages on PDSF aided debugging the code and how to avoid misuse of its various configuration options.

Today NERSC system managers can use our Nagios downtime scripts to quickly and easily accommodate downtime for anything Nagios monitors. Our downtime tool has saved a mountain of both point and click tasks and avoided the risky last resort of manually disabling Nagios checks.

NERSC wishes to present these Nagios downtime scripts and describe more fully how this code has aided our support efforts.

Summary:

NERSC has created and implemented original code to directly suspend Nagios monitors to accommodate planned outages.

Site Reports / **33**

IRFU site report

Author: Frederic Schaer[1]

[1] *CEA*

Corresponding Author: frederic.schaer@cea.fr

In this site report, we will speak about what changed at CEA/IRFU and what has been interesting since Hepix@Annecy, 6 months ago.

IT End User and Operating Systems / 34

Experiences with EL 7 at T2_US_Nebraska

Author: Garhan Attebury[1]

[1] *University of Nebraska (US)*

Corresponding Author: garhan.attebury@cern.ch

Seven years have passed since the initial EL 5 release and yet it's still found in active use at many sites. The successor EL 6 is also showing age with its 4th birthday just around the corner. While both are still under support from RedHat for many years to come, it never hurts to prepare for the future.

This talk will detail the experiences at T2_US_Nebraska in transitioning towards EL 7 using CentOS 7 as a base. Highlights will include the major differences between EL6 and EL7 and how they relate to our daily operations as a Tier2 CMS site.

Summary:

This talk will detail the experiences at T2_US_Nebraska in transitioning towards EL 7 using CentOS 7 as a base.

Site Reports / 35

KIT Site Report

Author: Andreas Petzold[1]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

Corresponding Author: andreas.petzold@cern.ch

News about GridKa Tier-1 and other KIT IT projects and infrastructure.

Site Reports / 36

Updates from Jefferson Lab HPC and Scientific Computing

Author: Sandy Philpott[1]

[1] *JLAB*

Corresponding Author: sandy.philpott@jlab.org

An overview since our spring meeting on JLab's latest developments for 12 GeV physics computing and storage, Lustre update, openZFS plan, load balancing between HPC and data analysis, Facilities changes in the Data Center, ...

Grids, Clouds, Virtualisation / **37**

RAL Tier 1 Cloud Computing Developments

Author: Ian Peter Collier[1]

Co-author: Andrew David Lahiff [1]

[1] *STFC - Rutherford Appleton Lab. (GB)*

Corresponding Author: ian.peter.collier@cern.ch

Update on the RAL Tier 1 cloud deployment and cloud computing activities.

Grids, Clouds, Virtualisation / **38**

LHC@home status - Outlook for wider use of volunteer computing at CERN

Author: Nils Hoimyr[1]

Co-authors: Alvaro Gonzalez Alvarez [1]; Helge Meinhard [1]; Miguel Marquina [1]; Pete Jones [1]; Tomi Juhani Asp [2]

[1] *CERN*

[2] *University of Jyvaskyla (FI)*

Corresponding Author: nils.hoimyr@cern.ch

LHC@home was brought back to CERN-IT in 2011, with 2 projects; Sixtrack and Test4Theory, the latter using virtualization with CernVM. Thanks to this development, there is increased interest in volunteer computing at CERN, notably since native virtualization support has been added to the BOINC middleware. Pilot projects with applications from the LHC experiment collaborations running on CernVM have also been deployed, opening the perspective for wider use of BOINC also for High Energy Physics software. The presentation will address the current status of LHC@home and the evolution of the CERN BOINC service to address the needs of a wider range of applications and users.

Summary:

Use of BOINC at CERN for LHC@home, Virtualization support with BOINC and CernVM allows for running HEP software under BOINC, and LHC@home will be extended to include applications from ATLAS, CMS and LHCb. A Description of BOINC application and server infrastructure at CERN is given.

Storage and Filesystems / **39**

Ceph Based Storage Systems for RACF

Authors: Alexandr Zaytsev[1]; Hironori Ito[1]

Co-authors: Antonio Wong [1]; Christopher Hollowell [1]; Tejas Rao [1]

[1] *Brookhaven National Laboratory (US)*

Corresponding Author: alezayt@bnl.gov

Ceph based storage solutions are becoming increasingly popular within the HEP/NP community over the last few years. With the current status of Ceph project, both object storage and block storage layers are production ready on a large scale, and the Ceph file system storage layer (CephFS) is rapidly getting to that state as well. This contribution contains a thorough review of various functionality, performance and stability tests performed with all three (object storage, block storage and file system) levels of Ceph by using the RACF computing resources in 2012-2014 on various hardware platforms (including HP Moonshot) and with different networking solutions (10/40 GbE and IPoIB/4X FDR Infiniband based). We also report the status of commissioning a large scale (1 PB of usable capacity, 4.0k HDDs behind the RAID arrays by design) Ceph based object storage system provided with AMZ/S3 complaint RadosGW interfaces which is currently being finalized within RACF, and the early performance results obtained with it.

IT End User and Operating Systems / **40**

Scientific Linux current status update

Authors: Pat Riehecky[1]; connie sieh[1]

[1] *Fermilab*

Corresponding Author: riehecky@fnal.gov

This presentation will provide an update on the current status of Scientific Linux, descriptions for some possible future goals, and allow a chance for users to provide feedback on its direction.

Summary:

This presentation will provide an update on the current status of Scientific Linux, descriptions for some possible future goals, and allow a chance for users to provide feedback on its direction.

Site Reports / **41**

University of Wisconsin Madison CMS T2 site report

Authors: Ajit mohapatra[1]; Carl Vuosalo[1]; Daniel Charles Bradley[1]; Sridhara Dasu[1]; Tapas Sarangi[1]

[1] *University of Wisconsin (US)*

Corresponding Author: tapas.sarangi@cern.ch

As a major WLCG/OSG T2 site, the University of Wisconsin Madison CMS T2 has provided very productive and reliable services for CMS MonteCarlo production/processing, and large scale global CMS physics analysis using high throughput computing (HT-Condor), highly available storage system (Hadoop), efficient data access using xrootd/AAA, and scalable distributed software systems

(CVMFS). An update on the current status of and activities (since the last report at Ann Arbor meeting) at the UW Madison Tier-2 will be presented that includes efforts on the 100Gb network upgrade and IPv6 etc., among other things.

Summary:

An update on the current status of and activities (since the last report at Ann Arbor meeting) at the UW Madison Tier-2 will be presented that includes efforts on the 100Gb network upgrade and IPv6 etc., among other things.

Storage and Filesystems / **42**

OpenZFS on Linux

Author: Brian Behlendorf[1]

[1] *LLNL*

Corresponding Author: behlendorf1@llnl.gov

OpenZFS is a storage platform that encompasses the functionality of a traditional filesystem and volume manager. It's highly scalable, provides robust data protection, supports advanced features like snapshots and clones, and is easy to administer. These features make it an appealing choice for HPC sites like LLNL which uses it for all production Lustre filesystems.

This contribution will discuss the state of OpenZFS on Linux including its goals and challenges. It will review the core features which make OpenZFS an excellent choice for managing large amounts of storage. Several new features will be discussed along our current plans for future improvements. I'll report on LLNL's use of OpenZFS on Linux over the last year to manage 100PB of production storage, including our experiences regarding stability, performance, and administration.

Storage and Filesystems / **43**

SSD benchmarking at CERN

Author: Liviu Valsan[1]

[1] *CERN*

Corresponding Author: liviu.valsan@cern.ch

Flash storage is slowly becoming more and more prevalent in the High Energy Physics community. When deploying Solid State Drives (SSDs) it's important to understand their capabilities and limitations, allowing to choose the best adapted product for the use case at hand. Benchmarking results from synthetic and real-world workloads on a wide array of Solid State Drives will be presented. The new NVM Express SSD interface specification will be detailed with a look towards the future of enterprise flash storage. The presentation will end by touching on endurance concerns often associated with the use of flash storage.

Site Reports / **44**

BNL RACF Site Report

Author: James Pryor[1]

[1] *B*

A summary of developments at BNL's RHIC/ATLAS Computing Facility since the last HEPiX meeting.

45

IHEP Site Report

Author: Jingyan Shi[1]

Co-author: Fazhi QI

[1] *IHEP*

It's the site report including what we have done with the storage, computing. Besides, it will discuss the serious error happened with our central switch and how we deal with. The progress of cloud computing based on openstack will be also discussed.

Networking and Security / **46**

Evaluating Infiniband Based Networking Solutions for HEP/NP Data Processing Applications

Author: Alexandr Zaytsev[1]

Co-authors: Christopher Hollowell [2]; Ofer Rind [2]; Tony Wong [2]; William Strecker-Kellogg [3]

[1] *Brookhaven National Laboratory (US)*

[2] *Brookhaven National Laboratory*

[3] *Brookhaven National Lab*

Corresponding Author: alezayt@bnl.gov

The Infiniband networking technology is a long established and rapidly developing technology which is currently dominating the field of low-latency, high-throughput interconnects for HPC systems in general and those included in the TOP-500 list in particular. Over the last 4 years a successful use of Infiniband networking technology combined with additional IP-over-IB protocol and Infiniband to Ethernet bridging layers was demonstrated well beyond the realm of HPC, covering various high throughput computing (HTC) systems, including data processing farms and private clouds devoted to HEP/NP data processing. With the recent advances of Mellanox VPI technology in 2013-2014 the 4X FDR IB now stands as the most versatile networking solution available for existing and future data centers that need to support both HTC and HPC oriented activities that can be seamlessly integrated into the existing Ethernet based infrastructure. Furthermore, it can be done completely transparently for the end users of these facilities, though certain modifications of the end user's activity-patterns are needed in order to utilize the full potential of the Infiniband based networking infrastructure. This contribution contains a detailed report on the series of tests and evaluation activities performed within the RACF over the last year in order to evaluate a Mellanox 4X FDR Infiniband based networking architecture (provided with an oversubscribed tree topology) as a potential alternative networking solution for both the RHIC and ATLAS data processing farms of the RACF, as well as the existing dCache and future Ceph based storage systems associated with them. Results of the price/performance comparison of such a networking system with a competing solution based on the 10 GbE technology (provided with non-blocking fabric topology) for a HEP/NP data processing farm consisting of 1500 compute nodes are presented. Job placement optimizations in Condor for

the offline data processing farm of the PHENIX experiment were implemented in order to demonstrate a sample user activity-pattern that more optimally utilizes the Infiniband-based networking solution. The results of using those optimizations in production for the PHENIX experiment over the last 9 months are presented.

Storage and Filesystems / **47**

EOS across 1000 km

Author: Luca Mascetti[1]

[1] *CERN*

Corresponding Author: luca.mascetti@cern.ch

In this contribution we report our experience in operating EOS, the CERN-IT high-performance disk-only solution, in multiple Computer Centres. EOS is one of the first production services exploiting the CERN's new facility located in Budapest, using his stochastic geo-location of data replicas.

Currently EOS holds more than 100PB of raw disk space for the four big experiments (ALICE, ATLAS, CMS, LHCb) and for our general purpose instance, of which 40PB are installed 1000 km away from Geneva.

Storage and Filesystems / **48**

The Lustre Filesystem for Petabyte Storage at the Florida HPC Center

Author: Dimitri Bourilkov[1]

Co-authors: Bockjoo Kim [1]; Craig Prescott [2]; Paul Ralph Avery [1]; Yu Fu [1]

[1] *University of Florida (US)*

[2] *UNIVERSITY OF FLORIDA*

Corresponding Author: bourilkov@phys.ufl.edu

Design, performance, scalability, operational experience, monitoring, different modes of access and expansion plans for the Lustre filesystems, deployed for high performance computing at the University of Florida, are described. Currently we are running storage systems of 1.7 petabytes for the CMS Tier2 center and 2.0 petabytes for the university-wide HPC center.

Summary:

Design, performance, scalability, operational experience, monitoring, different modes of access and expansion plans for the Lustre filesystems, deployed for high performance computing at the University of Florida, are described. Currently we are running storage systems of 1.7 petabytes for the CMS Tier2 center and 2.0 petabytes for the university-wide HPC center.

Site Reports / 49

DESY Site Report

Author: Andreas Haupt[1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: andreas.haupt@desy.de

News from DESY since the Annecy meeting.

Storage and Filesystems / 50

Using XRootD to Minimize Hadoop Replication

Author: Jeffrey Dost[1]

[1] *UCSD*

Corresponding Author: jdost@ucsd.edu

We have developed an XRootD extension to Hadoop at UCSD that allows a site to significantly free local storage space by taking advantage of the file redundancy already provided by the XRootD Federation. Rather than failing when a corrupt portion of a file is accessed, the hdfs-xrootd-fallback system retrieves the segment from another site using XRootD, thus serving the original file to the end user seamlessly. These XRootD-fetched blocks are then cached locally, so subsequent accesses to the same segment do not require wide area network access. A second process is responsible for comparing the fetched blocks with corrupt blocks in Hadoop, and injects the cached blocks back into the cluster. This on-demand healing allows a site admin to relax the file replication number, commonly required to ensure availability. The system has been put into production at the UCSDT2 since March of 2014, and we finished implementing the healing portion in September. The added resiliency of the hdfs-xrootd-fallback system has allowed us to free 236 TB in our local storage facility.

Basic IT Services / 51

Configuration Management, Change Management, and Culture Management

Author: James Pryor[1]

Co-authors: Jason Alexander Smith [2]; John Steven De Stefano Jr [2]

[1] *B*

[2] *Brookhaven National Laboratory (US)*

In 2010, the RACF at BNL began investigating Agile/DevOps practices and methodologies to be able to do more in less time or effort. We choose Puppet in 2010 and by Spring of 2011 we had converted about half our of configuration shell scripts into Puppet code on a handful of machines. Today we have scaled Puppet 3.x to support our entire facility and and host a common Puppet code base that is now shared and used upstream by the Physics and IT departments.

Storage and Filesystems / **52**

Cernbox + EOS: Cloud Storage for Science

Authors: Andreas Joachim Peters[1]; Hugo Gonzalez Labrador[2]; Jakub Moscicki[1]; Luca Mascetti[1]; Massimo Lamanna[1]

[1] *CERN*

[2] *University of Vigo (ES)*

Corresponding Author: luca.mascetti@cern.ch

Cernbox is a cloud synchronization service for end-users: it allows to sync and share files on all major platforms (Linux, Windows, MacOSX, Android, iOS). The very successful beta phase of the service demonstrated high demand in the community for such easily accessible cloud storage solution. Integration of Cernbox service with the EOS storage backend is the next step towards providing sync and share capabilities for scientific and engineering use-cases. In this report we will present lessons learnt from the beta phase of the Cernbox service, key technical aspects of Cernbox/EOS integration and new, emerging usage possibilities. The latter include the ongoing integration of sync and share capabilities with the LHC data analysis tools and transfer services.

Site Reports / **53**

AGLT2 Site Report Fall 2014

Author: Shawn Mc Kee[1]

Co-authors: Ben Meekhof [2]; Philippe Alain Luc Laurens [3]; Robert Ball [1]

[1] *University of Michigan (US)*

[2] *University of Michigan*

[3] *Michigan State University (US)*

Corresponding Author: shawn.mckee@cern.ch

I will present an update on our site since the last report and cover our work with dCache, perfSONAR-PS and VMWare. I will also report on our recent hardware purchases for 2014 as well as the status of our new networking configuration and 100G connection to the WAN. I conclude with a summary of what has worked and what problems we encountered and indicate directions for future work.

Summary:

Update on AGLT2 including changes in software, hardware and site configurations and summary of status and future work.

Networking and Security / **54**

OSG IPv6 Software and Operations Preparations

Author: Robert Quick[1]

[1] *Indiana University*

Corresponding Author: rquick@iu.edu

OSG Operations and Software will soon be configuring our operational infrastructure and middleware components with an IPv6 network stack capabilities in addition to its existing IPv4 stack. For OSG services this means network interfaces will thus have at least one IPv6 address on which it listens, in addition to whatever IPv4 addresses it is already listening on. For middleware components we will test dual stacked servers versus dual-stacked clients and IPv4 only clients.

Site Reports / **55**

RAL Site Report

Author: Martin Bly[1]

[1] *STFC-RAL*

Corresponding Author: martin.bly@stfc.ac.uk

Latest from RAL Tier1

Computing and Batch Systems / **56**

HTCondor and HEP Partnership and Activities

Author: Todd Tannenbaum[1]

[1] *Univ of Wisconsin-Madison, Wisconsin, USA*

The goal of the HTCondor team is to to develop, implement, deploy, and evaluate mechanisms and policies that support High Throughput Computing (HTC) on large collections of distributively owned computing resources. Increasingly, the work performed by the HTCondor developers is being driven by its partnership with the High Energy Physics (HEP) community. This presentation will provide an overview of how the HTCondor developers and the HEP community currently work together to advance the capabilities of batch computing, and it will also expose some of the new HTCondor functionality currently under development.

Summary:

This presentation will provide an overview of how the HTCondor developers and the HEP community currently work together to advance the capabilities of batch computing, and it will also expose some of the new HTCondor functionality currently under development.

Computing and Batch Systems / **57**

Releasing the HTCondor-CE into the Wild

Author: Brian Paul Bockelman[1]

[1] *University of Nebraska (US)*

Corresponding Author: brian.bockelman@cern.ch

One of the most critical components delivered by the Open Science Grid (OSG) software team is the compute element, or the OSG-CE. At the core of the CE itself is the gatekeeper software for translating grid pilot jobs into local batch system jobs. OSG is in the process of migrating from the Globus gatekeeper to the HTCondor-CE, supported by the HTCondor team.

The HTCondor-CE provides an alternate view on how grid gatekeepers can offer provisioning services, with a model that significantly differs from other gatekeepers such as GRAM or CREAM. Further, the HTCondor-CE is not a standalone product in itself but a specialized configuration of the familiar HTCondor software. Basing the HTCondor-CE on the much larger HTCondor product will allow a rich set of new features in the future (with low development costs!).

In this presentation, I'll highlight some of the biggest technical similarities and differences between the HTCondor-CE and other gatekeepers. Further, I'll discuss some of the non-technical considerations (documentation, operations, rollout, etc) we had to take into account in managing such a large-scale software transition in the OSG Production Grid environment.

Computing and Batch Systems / **58**

HTCondor on the Grid and in the Cloud

Author: James Frey[None]

Corresponding Author: jfrey@cs.wisc.edu

An important use of HTCondor is as a scalable, reliable interface for jobs destined for other scheduling systems.
These include Grid intefaces to batch systems (Globus, CREAM, ARC) and Cloud services (EC2, OpenStack, GCE).
The High Energy Physics community has been a major user of this functionality and has driven its development.
This talk will provide an overview of HTCondor's Grid and Cloud capabilities, how they're being employed, and our plans for future functionality.

Summary:

This talk will give an overview of HTCondor's current and future capabilities in the Grid and Cloud realms.

IT Facilities and Business Continuity / **59**

Joint procurement of IT equipment and services

Author: Bob Jones[1]

Co-author: Wayne Salter [1]

[1] *CERN*

Corresponding Author: wayne.salter@cern.ch

The presentation describes options for joint activities around procurement of equipment and services by public labs, possibly with funding by the European Commission. The presentation is intended to inform the community and check whether there is interest.

HEPiX Business / 60

Local Organizer Info

Corresponding Author: brian.bockelman@cern.ch