HEPiX Spring 2018 Workshop

Monday, 14 May 2018 - Friday, 18 May 2018

University of Wisconsin-Madison

Book of Abstracts

Contents

Site reports / **89**

DESY Site Report

Author: Timm Essigke[1]

[1] *DESY*

Corresponding Author: timm.essigke@desy.de

News about what happened at DESY during the last months

Desired length:

12

Computing and batch systems / **90**

Fast Distributed Image Reconstruction using CUDA/MPI

Authors: Eduardo Miqueles[1]; Gilberto Martinez Jr.[1]; Fernando Furusato[1]

[1] *LNLS/CNPEM*

Corresponding Authors: gilberto.martinez@lnls.br, fernando.furusato@lnls.br, eduardo.miqueles@lnls.br

In this work, we present a fast implementation for analytical image reconstruction from projections, using the so-called "backprojection-slice theorem" (BST). BST has the ability to reproduce reliable image reconstructions in a reasonable amount of time, before taking further decisions. The BST is easy to implement and can be used to take fast decisions about the quality of the measurement, i.e., sample environment, beam-line conditions, among others. A synchrotron facility able to measure a three-dimensional dataset Y within few seconds, needs a fast reconstruction algorithm able to provide a fast "preview" of the tomography within the same amount of time. If the experimental conditions are not satisfactory, the quality of the reconstruction will decrease, and the researcher can decide either to make another scan, or to process later the data using advanced reconstruction algorithms or even high quality segmentation methods. The difficulty here is that inversion algorithms depends on the backprojection operator, which is defined as an average through all the x-rays passing at a given pixel. Backprojection presents a high computational complexity of $O(N^3)$ for an image of N^2 pixels. The brute-force approach to compute the backprojection operator can be made extremely slow, even using a GPU implementation. Sophisticated ray-tracing strategies can also be used to make the running time faster and others analytical strategies reduce the backprojection complexity to $O(N^2 \log N)$. The BST approach have the same low complexity of $O(N^2 \log N)$ although easier to implement than his competitors, producing less numerical artifacts and following a more traditional "gridding strategy".

Desired length:

20

Networking and security / **91**

Cyberinfrastructure and China Science and Technology Cloud Plan in Chinese Academy of Sciences

Authors: YANG WANG[1]; Hong Xuehai[2]

Co-author: Li Jingjing [1]

[1] *Computer Network Information Center, Chinese Academy of Sciences*

[2] *Institute of Computing Technology, Chinese Academy of Sciences*

Corresponding Authors: wangyang@cnic.cn, hxh@ict.ac.cn, lijingjing@cstnet.cn

Chinese Academy of Sciences has 104 research institutes, 12 branch academies, three universities and 11 supporting organizations in 23 provincial-level areas throughout the country. These institutions are home to more than 100 national key labs and engineering centers as well as nearly 200 CAS key labs and engineering centers. Altogether, CAS comprises 1,000 sites and stations across the country.

As the science research methods develops, we are coming to the fourth paradigm of the science research—Data intensive science research. Data, compute and the link of them, network has played an more important role in science research. All the institutes have various demands in cyberinfrastructure.

China Science and Technology Cloud(CSTC) was constructed in order to meet the needs of the research institutes under the Chinese Academy of Sciences and even the whole scientific and technological community in China. It is an IT-based resources management and cloud service platform with smart resource dispatching and user self-service. It constructs a new-generation information infrastructure which is at high speed, dynamic and self-adaptive; speed up the state-level high-performance computing environment development, and achieve one-stop service for scientific computing. It integrates cloud computing and cloud storage facilities to enhance data recovery capabilities of the whole academy's scientific data assets and application systems. It also integrates and gather various scientific and technological information resources, and sets up a smart cloud service platform to provide scientific and technological resources and information services.

CSTC has maintained long-term partnerships with world-class research organizations such as U.S. National Center for Supercomputing Applications and Forschungszentrum Jülich. The predecessor of CSTC, China Science Technology Network (CSTNET) was one of the founding organizations of the Global Ring Network for Advanced Application Development (GLORIAD) which connect the North American with 10Gb/s bandwidth.
Based on CSTC, we build an expandable basic environment which can carry Big Data resources and support Big Data analysis and processing, realizing management and on-line processing of massive scientific data; targeting fields of relevant disciplines, as well as major research projects and special projects of the state and the academy, to deploy a batch of Big Data driven scientific research and application service in the fields of astronomy, biology, high-energy physics, etc..

The CSTC plans to build a test bed for network research and big science research, for example a Dynamic Virtual Dedicate Network for VLBI research, a DMZ for Advanced Light Source, an Open Network Environment for LHC. We are looking forward for further cooperation with global science research institutes.

Desired length:

20

End-user services and operating systems / **92**

Scientific Linux update

Author: Pat Riehecky[1]

Co-author: Bonnie King [1]

[1] *Fermilab*

Corresponding Authors: riehecky@fnal.gov, bonniek@fnal.gov

Updates on the status of Scientific Linux

Desired length:

20m

Networking and security / **93**

Computer Security Update

Author: Stefan Lueders[1]

[1] *CERN*

Corresponding Author: stefan.lueders@cern.ch

This presentation provides an update on the global security landscape since the last HEPiX meeting. It describes the main vectors of risks to and compromises in the academic community including lessons learnt, presents interesting recent attacks while providing recommendations on how to best protect ourselves. It also covers security risks management in general, as well as the security aspects of the current hot topics in computing and around computer security.

This talk is based on contributions and input from the CERN Computer Security Team.

Desired length:

20

Site reports / **94**

CERN Site Report

Author: Andrei Dumitru[1]

[1] *CERN*

Corresponding Author: andrei.dumitru@cern.ch

News from CERN since the HEPiX Fall 2017 workshop at KEK, Tsukuba, Japan.

Desired length:

12

Networking and security / **95**

SDN implementation plan in China Science and Technology Network

Author: JINGJING LI[None]

Co-author: YANG WANG [1]

[1] *Computer Network Information Center, Chinese Academy of Sciences*

Corresponding Authors: jimmy@cnic.cn, wangyang@cnic.cn

Scientific activities generate huge data and need to transfer them to some places to research. Traditional networking infrastructure has a defined architecture and can not satisfy such real-time and high-quality transferring requirements.

China Science and Technology Network(CSTNet) was constructed in order to meet the needs of the research institutes under the Chinese Academy of Sciences and even the whole scientific and technological community in China. CSTNet has planned to construct a new-generation infrastructure using some new technology such as SDN, NFV and etc.

CSTNet has started to build a new NOC running system to achieve real-time measurements, monitoring and management of the network flows. The new NOC running system can provide an user interface to enable user to submit networking requirements dynamic and self-adaptive, some network configure task can become effective on time no need to connect to console board.

Dynamic network management will accelerate the integrating of cloud computing and cloud storage facilities because of faster data transfer command delivering and implementing to service research in the fields of astronomy, biology, high-energy physics, etc.

Desired length:

15

Storage and file systems / **96**

WLCG Archival Storage group report

Authors: Oliver Keeble[1]; Vladimir Bahyl[1]

[1] *CERN*

Corresponding Authors: oliver.keeble@cern.ch, vladimir.bahyl@cern.ch

The group has been formed to tackle two main themes

- establish a knowledge-sharing community for those operating archival storage for WLCG
- understand how to monitor usage of archival systems and optimise their exploitation by experiments

I will report on the recent activities of this group.

Desired length:

10

Site reports / **97**

INFN-T1 Site report

Authors: Andrea Chierici[1]; Stefano Dal Pra[2]

[1] *INFN-CNAF*

[2] *INFN*

Corresponding Authors: stefano.dalpra@cnaf.infn.it, chierici@cnaf.infn.it

A brief update on INFN-T1 site, what is our current status and what is still to be done to reach 100% functionality

Desired length:

12

Computing and batch systems / **98**

Batch on EOS Extra Resources moving towards production

Authors: David Smith[1]; Markus Schulz[1]; Andrey kiryanov[None]; Ben Jones[1]; Gavin McCance[1]; Massimo Lamanna[1]; Herve Rousseau[1]

[1] *CERN*

Corresponding Authors: gavin.mccance@cern.ch, massimo.lamanna@cern.ch, ben.dylan.jones@cern.ch, david.smith@cern.ch, markus.schulz@cern.ch, andrey.kiryanov@cern.ch, herve.rousseau@cern.ch

At the last HEPix meeting we described the results of a proof of concept study to run batch jobs on EOS disc server nodes. By now we have moved forward towards a production level configuration and the first pre-production nodes have been setup. Beside the relevance for CERN this is also a more general step towards a hyper-converged infrastructure.

Desired length:

15

Computing and batch systems / **99**

HSF-WLCG Cost and Performance Modeling Working Group

Authors: Andrea Sciaba[1]; Markus Schulz[1]; Jose Flix Molina[2]

[1] *CERN*

[2] *Centro de Investigaciones Energéti cas Medioambientales y Tecno*

Corresponding Authors: jose.flix.molina@cern.ch, andrea.sciaba@cern.ch, markus.schulz@cern.ch

The working group has been established and is now working towards a cost and performance model that allows to quantitatively estimate the computing resources needed for HL-LHC and map them towards the cost at specific sites.

The group has defined a short and medium term plan and identified the main tasks. Around the tasks teams with members from experiments and sites have formed and started concrete work. We will report on the goals and status of the working group.

Desired length:

15

End-user services and operating systems / 100

TRIDENT Tool for collecting and understanding performance hardware counters

Authors: Servesh Muralidharan[1]; David Smith[1]

[1] *CERN*

Corresponding Authors: servesh.muralidharan@cern.ch, david.smith@cern.ch

Trident, a tool to use low level metrics derived from hardware counters to understand Core, Memory and I/O utilisation and bottlenecks. The collection of time series of these low level counters does not induce significant overhead to the execution of the application.

The Understanding Performance team is investigating on a new node characterisation tool, ¹Trident¹, that can look at various low level metrics with respect to the Core, Memory and I/O. Trident uses a three pronged approach to analysing node¹s utilisation and understand the stress on different parts of the node based on the given job. Currently core metrics such as memory bandwidth, core utilization, active processor cycles, etc., are being collected. Interpretation of this data is often non intuitive. The tool preprocesses the data to make the data usable by developers and site managers without the need of in-depths expertise of CPU and systems architecture details.

Desired length:

15

IT facilities / 101

ExDeMon: a new scalable monitoring tool for the growing CERN infrastructure

Author: Daniel Lanza Garcia[1]

[1] *CERN*

Corresponding Author: daniel.lanza@cern.ch

When monitoring an increasing number of machines, infrastructure and tools need to be rethinked. A new tool, ExDeMon, for detecting anomalies and raising actions, has been developed to perform well on this growing infrastructure. Considerations of the development and implementation will be shared.

Daniel has been working at CERN for more than 3 years as Big Data developer, he has being implementing different tools for monitoring the computing infrastructure in the organisation.

Desired length:

20

Clouds, virtualisation, grids / **102**

HNSciCloud Status Report

Author: Andreas Petzold[1]

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

Corresponding Author: andreas.petzold@cern.ch

The Helix Nebula Science Cloud (HNSciCloud) Horizon 2020 Pre-Commercial Procurement project (http://www.hnscicloud.eu/) brings together a group of 10 research organisations to procure innovative cloud services from commercial providers to establish a cloud platform for the European research community.
This 3 year project has recently entered its final phase which will deploy two pilots with a combined capacity of 20,000 cores and 2 PB of storage integrated with the GEANT network at 40Gbps.
This presentation will provide an overview of the project, the pilots, the applications being deployed and lessons learned to-date.

Desired length:

20

Storage and file systems / **103**

The Software Defined Online Storage System at the GridKa WLCG Tier-1 Center

Author: Jan Erik Sundermann[1]

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

Corresponding Author: jan.sundermann@kit.edu

The computing center GridKa is serving the ALICE, ATLAS, CMS and LHCb experiments as one of the biggest WLCG Tier-1 centers world wide with compute and storage resources. It is operated by the Steinbuch Centre for Computing at Karlsruhe Institute of Technology in Germany. In April 2017 a new online storage system was put into operation. In its current stage of expansion it offers the HEP experiments a capacity of 23 Petabytes of online storage distributed over 16 redundant storage servers with 3900 disks and 50TB SSDs. The storage is connected via two redundant infiniband fabrics to 44 file servers which in turn are connected each via 40Gbit/s and several 100Gbit/s ethernet uplinks to the GridKa backbone network. The whole storage is partitioned into few large file systems, one for each experiment, using IBM Spectrum Scale as software-defined-storage base layer. The system offers a combined read-write performance of 70Gbyte/s. It can be scaled transparently both in size and performance allowing to fulfill the growing needs especially of the LHC experiments for online storage in the coming years.

In this presentation we discuss the general architecture of the storage system and present first experiences with the performance of the system in production use. In addition we present the current plans for expansion of the system.

Desired length:

20

End-user services and operating systems / **104**

Run the latest software on a stable enviroment - A simpler way

Author: Troy Dawson[None]

Corresponding Author: tdawson@redhat.com

What do our users want?
One group wants the latest version of foo, but the stable version of bar.
The other group wants the latest version of bar, but the old version of foo.

What have we tried?
SCL
SCL's are great in theory. But in practice they are hard for the packagers. They also make the developers have to jump through several hoops. If something was developed in an SCL enviroment, it often wouldn't translate into a non-SCL enviroment.
Containers
Containers are also great in theory. They are especially great for allowing people to run their code in the exact enviroment on different machines. But for developers, they still have to jump through many hoops to develop on SCL's. And often we restrict what version of foo and bar we have in those containers.
Tarballs and Zip files
We admit it, that's what our developers are really doing. They are pulling down what they need, from who knows where, writting their code around it, and then asking us (or you) to support it. This is very bad for security, as well as for administrators trying to duplicate the enviroment on another machine.

Modules - The simpler way
The easiest way to explain modules is they are like yum groups, done right.
Using dnf admins are able to install nodejs 6, or nodejs 9. They aren't installed over in /opt/, they are installed in their usual place in /usr/.
Are you ready to move from python36 to python37? Just change the version of the python module and dnf with change it all.
Users will only be allowed to have one version of python, or nodejs, just like normal python or nodejs. But the developers (or their code) won't have to do anything special to use them.

This presentation will go through our up and downs as we've worked on getting a new technology created to help our users.

https://docs.pagure.org/modularity/

Desired length:

20

Clouds, virtualisation, grids / **105**

Automatic for the People: Containers for LIGO software development on the Open Science Grid and other diverse computing resources

Author: Thomas Downes[1]

[1] *University of Wisconsin-Milwaukee*

Corresponding Author: downes@uwm.edu

Distributed research organizations are faced with wide variation in computing environments to support. LIGO has historically resolved this problem by providing RPM/DEB packages for (pre-)production software and coordination between clusters operated by LIGO-affiliated facilities and research groups. This has been largely successful although it leaves a gap in operating system support and in the development process prior to formal point releases.

We describe early developments in LIGO's use of GitLab, GitHub, and DockerHub to continuously deploy researcher-maintained containers for immediate use on all LIGO clusters, the Open Science Grid, and user workstations. Typical latencies are below an hour, dominated by the build-time of the software itself and the client refresh rate of the CernVM File System

Desired length:

15 minutes

Basic IT services / **106**

Evolution of the Hadoop and Spark platform for HEP

Author: Zbigniew Baranowski[1]

[1] *CERN*

Corresponding Author: zbigniew.baranowski@cern.ch

The interest in using Big Data solutions based on Hadoop ecosystem is constantly growing in HEP community. This drives the need for increased reliability and availability of the central Hadoop service and underlying infrastructure provided to the community by the CERN IT department.
This contribution will report on the overall status of the Hadoop platform and the recent enhancements and features introduced in many areas including the service configuration, availability, alerting, monitoring and data protection, in order to meet the new requirements posed by the users community.

Desired length:

20

Basic IT services / **107**

Monitoring Infrastructure for the CERN Data Centre

Authors: Asier Aguado Corman[1]; Alberto Aimar[2]; Pedro Andrade[2]; Simone Brundu[2]; Javier Delgado Fernandez[2]; Luis Fernandez Alvarez[2]; Borja Garrido Bear[2]; Edward Karavakis[2]; Dominik Marek Kulikowski[3]; Luca Magnoni[2]

[1] *Universidad de Oviedo (ES)*

[2] *CERN*

[3] *Wroclaw University of Science and Technology (PL)*

Corresponding Authors: alberto.aimar@cern.ch, simone.brundu@cern.ch, asier.aguado@cern.ch, luis.fernandez.alvarez@cern.ch, luca.magnoni@cern.ch, pedro.andrade@cern.ch, borja.garrido.bear@cern.ch, dominik.marek.kulikowski@cern.ch, javier.delgado.fernandez@cern.ch, edward.karavakis@cern.ch

Since early 2017, the MONIT infrastructure provides services for monitoring the CERN data centre, together with the WLCG grid resources, and progressively replaces in-house technologies, such as LEMON and SLS, using consolidated open source solutions for monitoring and alarms.

The infrastructure collects data from more than 30k data centre hosts in Meyrin and Wigner sites, with a total volume of 3 TB/day and a rate of 65k documents/sec. It includes OS and hardware metrics, as well as specific IT service metrics. Logs and metrics collection is deployed by default in every machine of the data centre, together with alert reporting. Each machine has a default configuration that can be extended for service-specific data (e.g. for specifically monitoring a database server). Service managers can send custom metrics and logs from their applications to the infrastructure through generic endpoints, and they are provided with an out-of-the-box discovery and visualization interface, data analysis tools and integrated notifications.

The infrastructure stack relies on open source technologies, developed and widely used by the industry and research leaders. Our architecture uses collectd for metric collection, Flume and Kafka for transport, Spark for stream and batch processing, Elasticsearch, HDFS and InfluxDB for search and storage, Kibana and Grafana for visualization, and Zeppelin for analytics. The modularity of collectd provides flexibility to the infrastructure users to configure default and service-specific monitoring, and allows to develop and deploy custom plugins.

This contribution is an updated overview of the monitoring service for CERN data centre. We present our main use cases for collection of metrics and logs. Given that the proposed stack of technologies is widely used, and the MONIT architecture is well consolidated, a main objective is to share the lessons learned and find common monitoring solutions within the community.

Desired length:

20

Computing and batch systems / **109**

Benchmarking Working Group. An update

Authors: Michele Michelotto[1]; Manfred Alef[2]; Domenico Giordano[3]

[1] *Università e INFN, Padova (IT)*

[2] *Karlsruhe Institute of Technology (KIT)*

[3] *CERN*

Corresponding Authors: michele.michelotto@cern.ch, domenico.giordano@cern.ch, manfred.alef@kit.edu

The benchmarking working group holds biweekly meeting. we are focusing on the health of HS06, fast benchmark and study of a new benchmark to replace HS06 since SPEC has moved to a new family of benchmark

Desired length:

20

Computing and batch systems / 111

Swiss HPC Tier-2 @ CSCS

Authors: Dino Conciatore[1]; Miguel Gila[1]; Dario Petrusic[1]; Pablo Fernandez[1]; Gianni Mario Ricciardi[1]

[1] *CSCS (Swiss National Supercomputing Centre)*

Corresponding Authors: dino.conciatore@cscs.ch, dario.petrusic@cern.ch, pablo.fernandez.fernandez@cern.ch, ricciardi@cscs.ch, miguel.angel.gila@cern.ch

For the past 10 years, CSCS has been providing computational resources for the ATLAS, CMS, and LHCb experiments on a standard commodity cluster.
The High Luminosity LHC upgrade (HL-LHC) presents new challenges and demands with a predicted 50x increase in computing needs over the next 8 to 10 years. High Performance Computing capabilities could help to equalize the computing demands due to there ability to provide specialized hardwre and economies of scale. For the past year, CSCS has been running the Tier-2 workload for these experiments on the flagship system Piz Daint, a Cray XC system.

Desired length:

20

Basic IT services / 112

A fully High-availability logs/metrics collector @ CSCS

Author: Dino Conciatore[1]

[1] *CSCS (Swiss National Supercomputing Centre)*

Corresponding Author: dino.conciatore@cscs.ch

As the complexity of systems increases and the scale of these systems increases, the amount of system level data recorded increases.
Managing the vast amounts of log data is a challenge that CSCS solved with the introduction of a centralized log and metrics infrastructure based on Elasticsearch, Graylog, Kibana, and Grafana.
This is a fundamental service at CSCS that provides easy correlation of events bridging the gap from the computation workload to nodes enabling failure diagnosis.
Currently, the Elasticsearch cluster at CSCS is handling more than 22'000'000'000 online documents (one year) and another 20'000'000'000 archived. The integrated environment from logging to graphical representation enables powerful dashboards and monitoring displays.

Desired length:

20

Site reports / 113

Purdue University CMS T2 site report

Authors: Stefan Piperov[1]; Erik Gough[1]; Majid Arabgol[1]; Thomas Hacker[1]; Norbert Neumeister[1]

[1] *Purdue University (US)*

Corresponding Authors: erik.gough@cern.ch, majid.arabgol@cern.ch, thomas.joe.hacker@cern.ch, norbert.neumeister@cern.ch, piperov@fnal.gov

Through participation in the Community Cluster Program of Purdue University, our Tier-2 center has for many years been one of the most productive and reliable sites for CMS computing, providing both dedicated and opportunistic resources to the collaboration. In this report we will present an overview of the site, review the successes and challenges of the last year of operation, and outline the perspectives and plans for future developments.

Desired length:

12

Storage and file systems / 115

Data To Network: building balanced throughput storage in a world of increasing disk sizes

Author: Tristan Suerink[1]

[1] *Nikhef National institute for subatomic physics (NL)*

Corresponding Author: t.suerink@cern.ch

The ever-decreasing cost of high capacity spinning media has resulted in a trend towards very large capacity storage 'building blocks'. Large numbers of disks - with up to 60 drives per enclosure being more-or-less standard – indeed allow for dense solutions, maximizing storage capacity in terms of floor space, and can in theory be packed almost exclusively with disks. The result are building blocks with a theoretical gross capacity of about 180 TByte per unit height when employing 12 TByte disks. This density comes at a cost, though: getting the data to and from the disks, via front-end storage management software and through a network, has different scaling characteristics than the gross storage density, and as a result maintaining performance in terms of throughput per storage capacity is an ever more complex challenge. At Nikhef, and for the NL-T1 service, we aim to maintain 12MiB/s/TiB combined throughput, supporting at least 2 read- and 1 write stream per 100TiB netto storage, from any external network source down to the physical disks. Especially this combined read-write operational pattern poses challenges not usually found in commercial deployments. Yet this is the pattern most commonly seen for our scientific applications in the Dutch National e-Infrastructure.

In this study we looked at each of the potential bottlenecks in such a mixed-load storage system: network throughput, limitations in the system bus between CPU, network card, and disk subsystem, at different disk configuration models (JBOD, erasure-encodings, hardware, and software RAID) and the effect on processor load in different CPU architectures. We present the results of different disk configurations and show the limitations of commodity redundancy technologies and how they affect processor load in both x86-64 and PowerPC systems, and how the corresponding system bus design impacts overall throughput.

Combining network and disk performance optimizations we show how high-density commodity components can be combined to build a cost-cutting system without bottlenecks – offering constant-throughput multi-stream performance with over 700TiB netto in just 10U and able to keep a 100Gbps network link full – as a reference architecture for everything from a single Data Transfer Node down to a real distributed storage cluster.

Desired length:

20 minutes

Site reports / 116

PIC site report

Author: Jose Flix Molina[1]

[1] *Centro de Investigaciones Energéti cas Medioambientales y Tecno*

Corresponding Author: jose.flix.molina@cern.ch

News from PIC since the HEPiX Fall 2017 workshop at KEK, Tsukuba, Japan.

Desired length:

12

IT facilities / **117**

Evolution of technology and markets

Authors: Bernd Panzer-Steindel[1]; Helge Meinhard[1]

[1] *CERN*

Corresponding Authors: helge.meinhard@cern.ch, bernd.panzer-steindel@cern.ch

A short review of how technology and markets have evolved in areas relevant for HEP computing

Desired length:

20 minutes (see comment)

IT facilities / **118**

Proposal for a technology watch WG

Authors: Helge Meinhard[1]; Bernd Panzer-Steindel[1]

[1] *CERN*

Corresponding Authors: bernd.panzer-steindel@cern.ch, helge.meinhard@cern.ch

Following up from abstract #117, a proposal to form a working group dedicated to technology watch

Desired length:

20 minutes (see comment)

Storage and file systems / **119**

Next generation of large-scale storage services at CERN

Author: Jakub Moscicki[1]

Co-authors: Enrico Bocchi [1]; Cristian Contescu [1]; Hugo Gonzalez Labrador [1]; Massimo Lamanna [1]; Luca Mascetti [1]; Andreas Joachim Peters [1]; Giuseppe Lo Presti [1]; Herve Rousseau [1]; Roberto Valverde Cameselle [2]

[1] *CERN*

[2] *Universidad de Oviedo (ES)*

Corresponding Authors: jakub.moscicki@cern.ch, massimo.lamanna@cern.ch, luca.mascetti@cern.ch, giuseppe.lopresti@cern.ch, andreas.joachim.peters@cern.ch, cristian.contescu@cern.ch, hugo.gonzalez.labrador@cern.ch, enrico.bocchi@cern.ch, roberto.valverde.cameselle@cern.ch, herve.rousseau@cern.ch

CERN IT Storage (IT/ST) group leads the development and operation of large-scale services based on EOS for the full spectrum of use-cases at CERN and in the HEP community. IT/ST group also provides storage for other internal services, such as Open Stack, using a solution based on Ceph. In this talk we present current operational status, ongoing development work and future architecture outlook for next generation storage services for the users based on EOS — a technology developed and integrated at CERN.

EOS is the home for all physics data-stores for LHC and non-LHC experiments (at present 250PB storage capacity). It is designed to operate at high data rates for experiment data-taking while running concurrent complex production work-loads. EOS also provides a flexible distributed storage back-end and architecture with plugins for tape archival (CTA - evolution and replacement for CASTOR), synchronization&sharing services (CERNBox) and general-purpose filesystem access for home directories (FUSE for Linux and SMB Gateways for Windows and Mac).

CERNBox is the cloud storage front-end for desktop,mobile and web access focused on personal user files, general-purpose project spaces and smaller physics datasets (at present 12K user accounts and 500M files). CERNBox provides simple and uniform access to storage on all modern devices and operating systems. CERNBox is also hub for integration with other services: Collaborative editing — MS Office365 and alternatives: Collabora and OnlyOffice; Web-based analysis — SWAN Jupyter Notebooks with access to computational resources via Spark and Batch; and software distribution via CVMFS.

This storage service ecosystem is designed to provide “total data access”: from end-user devices to geo-aware data lakes for WLCG and beyond. It also provides a foundation for strategic parternships (AARNet, JRC, ...), new communities such as CS3 (Cloud Storage and Synchronization Services) and new application projects such as Up2University (cloud storage ecosystem for education). CERN Storage technology has been showcased to work with commercial cloud providers such as Amazon, T-Systems (Helix Nebula) or COMTRADE (Openlab) and there is an increasing number of external sites testing the CERN storage service stack in their local computing centers.

This strategy proves very successful with the users and as a result storage services at CERN see exponential growth: CERNBox alone has grown by 450% in 2017. Growing overall demand drive the evolution of the service design and implementation of the full ecosystem: EOS core storage as well as CERNBox and SWAN. Recent EOS improvements include new distributed namespace to provide scaling and high-availability; new robust FUSE module providing client-side caching, lower latency and more IOPs; new workflow engine and many more. CERNBox is moving to micro-service oriented architecture and SWAN is tested with Kubernetes container orchestration.

New developments come together with a constant effort to streamline QA, testing and documentation as well as reduce manual configuration and operational effort for managing large-scale storage services.

Desired length:

20

Storage and file systems / **120**

Operating a large scale distributed XRootd cache across Caltech and UCSD

Authors: Edgar Fajardo Hernandez[1]; Matevz Tadel[1]; Alja Mrak Tadel[1]; Terrence Martin[None]; Frank Wuerthwein[2]

[1] *Univ. of California San Diego (US)*

[2] *UCSD*

Corresponding Authors: edgar.mauricio.fajardo.hernandez@cern.ch, matevz.tadel@cern.ch, fkw@fnal.gov, alja.mrak.tadel@cern.ch, tmartin@physics.ucsd.edu

After the successful adoption of the CMS Federation an opportunity arose to cache xrootd requests in Southern California. We present the operational challenges and the lessons learned from scaling a federated cache (a cache composed of several independent nodes) first at UCSD and the scaling and network challenges to augment it to include the Caltech Tier 2 Site. In which would be a first of a kind multisite Xrootd cache which could potentially ease the data management of CMS.

Desired length:

20 minutes

Clouds, virtualisation, grids / **121**

Status update of the CERN private cloud

Author: Spyridon Trigazis[1]

[1] *CERN*

Corresponding Author: spyridon.trigazis@cern.ch

CERN runs a private OpenStack Cloud with ~300K cores, ~3000 users and a number of OpenStack services. CERN users can built services using a pool of compute and storage resources using the OpenStack APIs like Ironic, Nova, Magnum, Cinder and Manila, on the other hand CERN cloud operators face some operational challenges at scale in order to offer them. In this talk, you will learn about the status of the CERN cloud, new services and plans for expansion.

Desired length:

20

Clouds, virtualisation, grids / **122**

Baremetal provisioning in the CERN cloud

Author: Spyridon Trigazis[1]

[1] *CERN*

Corresponding Author: spyridon.trigazis@cern.ch

Virtual machines is the technology that formed the modern clouds - private and public - however the physical machine are back in a more cloudy way. Cloud providers are offering APIs for baremetal server provisioning on demand and users are leveraging containers for isolation and reproducible deployments. In this talk, I will be presenting one of the newest services at the CERN cloud, Ironic, the Baremetal service of OpenStack. You will learn how the Cloud team improves its operational workflow and accounting and how users can use the same tooling they are used to when working with virtual machines. Finally, you will hear about the recent integration efforts between the container and baremetal services.

Desired length:

20

Site reports / **123**

Nikhef Site Report

Author: Bart van der Wal[1]

Co-author: Paul Kuipers [2]

[1] *NIkhef*

[2] *Nikhef*

Corresponding Authors: bwal@nikhef.nl, paulks@nikhef.nl

Site report from Nikhef

Desired length:

12

Computing and batch systems / **124**

Changing Compute Landscape at Brookhaven

Author: William Edward Strecker-Kellogg[1]

[1] *Brookhaven National Laboratory (US)*

Corresponding Author: william.edward.strecker-kellogg@cern.ch

Computing is changing at BNL, we will discuss how we are restructuring our Condor pools, integrating them with new tools like Jupyter notebooks, and other resources like HPC systems run with Slurm.

Desired length:

15-20

Basic IT services / **125**

A smorgasbord of tools around Linux at DESY

Author: Yves Kemp[1]

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

Corresponding Author: yves.kemp@cern.ch

In the past, we have developed lots of smaller and larger tools to help in various aspects of Linux administration at DESY.
We present (some) of them in this talk.
An incomplete list is:
- Two-Factor-Authentication
- Timeline repositories
- Making Kernel upgrade notifications (more) audit safe
- Fail2ban

Desired length:

20

Computing and batch systems / **126**

Techlab benchmarking web portal

Author: Maxime Reis[1]

[1] *CERN*

Corresponding Author: maxime.reis@cern.ch

Techlab, a CERN IT project, is a hardware lab providing experimental systems and benchmarking data for the HEP community.

Techlab is constantly on the lookout for new trends in HPC, cutting-edge technologies and alternative architectures, in terms of CPUs and accelerators.
We believe that in the long run, a diverse offer and a healthy competition in the HPC market will serve science in particular, computing in general, and everyone in the end.
For this reason, we encourage the use of not-quite-there-yet alternatives to the standard x86 quasi-monopoly, in the hope that in the near future, such alternative architectures can proudly compete, on an equal footing.
We buy hardware, set it up, test and benchmark it, then make it available to members of the HEP community for porting and testing their scientific applications and algorithms. On a best-effort basis, we try and help users make the best out of the hardware we provide.

To serve as basis for hardware choice, we run extensive benchmarks on all the systems we can get our hands on, and share the results to help others make fully informed choices when buying hardware that will fit their computing needs. As a means to achieve this, we developed a benchmarking web portal, open to everyone in the HEP community, to upload and publish data about all kinds of hardware. It was built with security in mind, and provides fine-grained access control to encourage even people working on yet-unreleased hardware to contribute.
As Techlab cannot possibly buy and test everything, it is our hope that this portal gets used by other HEP labs, and the database we build together becomes the 'one-stop shop' for benchmarking.

This presentation both gives an overview of Techlab's benchmarking web portal — what and whom it is designed for, what we hope to achieve with it — and delves into the technology choices of the implementation.

Desired length:

20

Clouds, virtualisation, grids / **127**

Integration of OpenStack and Amazon Web Service into local batch job system

Author: Wataru Takase[1]

Co-authors: Tomoaki Nakamura [2]; Koichi Murakami [2]; Takashi Sasaki [2]

[1] *High Energy Accelerator Research Organization (JP)*

[2] *KEK*

Corresponding Author: wataru.takase@kek.jp

Cloud computing enables flexible resource provisioning on demand. Through the collaboration with National Institute of Informatics (NII) Japan, we have been integrating our local batch job system with clouds for expanding its computing resource and providing heterogeneous clusters dynamically. In this talk, we will introduce our hybrid batch job system which can dispatch jobs to provisioned instances on on-premise OpenStack and Amazon Web Service as well as local servers. We will also report some performance test results conducted for investigation of the scalability.

Desired length:

20

Computing and batch systems / **128**

News from the DESY batch-clusters

Authors: Thomas Hartmann[1]; Thomas Finnern[1]; Sven Sternberger[1]; Christoph Beyer[None]

[1] *DESY*

Corresponding Authors: thomas.hartmann@desy.de, thomas.finnern@desy.de, sven.sternberger@desy.de, christoph@treibsand.net

The batch facilities at DESY are currently enlarged significantly while at the same time partly migrated from SGE to HTCondor.
This is a short overview of what is going on on site in terms of GRID-, local- and HPC cluster development.

Desired length:

15

Networking and security / **129**

Deployment of IPv6 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

For several years the HEPiX IPv6 Working Group has been testing WLCG services to ensure their IPv6 compliance. The transition of WLCG central and storage services to dual-stack IPv4/IPv6 is progressing well, thus enabling the use of IPv6-only CPU resources as agreed by the WLCG Management Board and presented by us at previous HEPiX meetings.

By April 2018, all WLCG Tier 1 data centres have provided access to their services over IPv6. The LHC experiments have requested all WLCG Tier 2 centres to provide dual-stack access to their storage by the end of LHC Run 2.The working group, driven by the requirements of the LHC VOs to be able to use IPv6-only opportunistic resources, continues to encourage wider deployment of dual-stack services and has been monitoring the transition. We will present the progress of the transition to IPv6.

Desired length:

20 minutes

Basic IT services / **130**

First Impressions of Saltstack and Reclass as our new Configuration Management System

Author: Dennis Van Dok[None]

Co-author: Andrew Pickford [1]

[1] *Nikhef*

Corresponding Authors: dennisvd@nikhef.nl, andrewp@nikhef.nl

In the Autumn of 2016 the Nikhef data processing facility (NDPF) found itself at a junction on the road of configuration management. The NDPF was one of the early adopters of Quattor, which served us well since the early days of the Grid. But where grid deployments were uniquely complex to require the likes of Quattor then, nowadays a plethora of configuration systems have cropped up to fulfill the needs of the booming industry of cloud orchestration.

Faced with the choice of an overhaul of our Quattor installation to bring our site up-to-date, or an investment to adopt a brand new system, we opted for the latter. And among the many candidates like Chef, Puppet, and Ansible, we chose Saltstack and partnered it with Reclass.

This led to hours of discussions designing the flows and processes of the new system, as well figuring out how to do the transition from the old to the new. And many more were spent trying out these ideas, pioneering as it were to find a way forward that would work for us.

We like to present the insights we've gained about Saltstack and configuration management in general, and our design choices in particular. This is very much a work in progress, as we are getting to know our new system while we are implementing more and more of the moving parts. We are gaining invaluable experience with tasks that are otherwise rarely among the day-to-day business of the system administrator, and we like to share it with our peers.

Desired length:

Miscellaneous / **132**

Logistics and announcements

Miscellaneous / **133**

Welcome to University of Wisconsin-Madison

Networking and security / **134**

News from the world of Federated Identity Management and AAI

Authors: Dave Kelsey[1]; Hannah Short[2]

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

[2] *CERN*

Corresponding Authors: david.kelsey@stfc.ac.uk, hannah.short@cern.ch

There are many ongoing activities related to the development and deployment of Federated Identities and AAI (Authentication and Authorisation Infrastructures) in research communities and cyber Infrastructures including WLCG and others. This talk will give a high-level overview of the status of at least some of the current activities in FIM4R, AARC, WLCG and elsewhere.

Desired length:

20 minutes

Computing and batch systems / **135**

What's new in HTCondor? What is upcoming?

Author: Todd Tannenbaum[1]

[1] *University of Wisconsin Madison (US)*

Corresponding Author: tannenba@cs.wisc.edu

he 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 talk will present recent changes and enhancements to HTCondor, including details on some of the enhancements created for the forthcoming HTCondor v8.8.0 release, as well as changes created on behalf of the HEP community. It will also discuss the upcoming HTCondor development roadmap, and seek to solicit feedback on the roadmap from HEPiX attendees.

Desired length:

15 min, but I am flexible shorter or longer

Site reports / **136**

BNL Site Report

Author: David Yu[1]

[1] *BNL*

Corresponding Author: david.yu@bnl.gov

Updates from BNL since KEK meeting

Desired length:

15

Site reports / **137**

FZU site report

Author: Jiri Chudoba[1]

[1] *Acad. of Sciences of the Czech Rep. (CZ)*

Corresponding Author: jiri.chudoba@cern.ch

Recently we deployed new cluster with worker nodes with 10 Gbps network connection
and new disk servers for DPM and xrootd. I will also discuss migration from Torque/Maui to HTCondor batch system.

Desired length:

12

IT facilities / **138**

BNL New Data Center - Status and Plans

Authors: Tony Wong[1]; Alexandr Zaytsev[2]; Shigeki Misawa[1]; Eric Christian Lancon[3]

[1] *Brookhaven National Laboratory*

[2] *Brookhaven National Laboratory (US)*

[3] *BNL*

Corresponding Authors: elancon@bnl.gov, alezayt@bnl.gov, misawa@bnl.gov, tony@bnl.gov

BNL is planning a new on-site data center for its growing portfolio of programs in need of scientific computing support. This presentation will provide an update on the status and plans for this new data center.

Desired length:

20

Computing and batch systems / **139**

PDSF - Current Status and Migration to Cori

Author: Tony Quan[1]

Co-authors: Jan Balewski [2]; Rath Georg [2]; James Botts [3]; Glenn Lockwood [2]; Damian Hazen [2]

[1] *LBL*

[2] *NERSC*

[3] *LBNL*

Corresponding Authors: gbrath@lbl.gov, balewski@lbl.gov, twquan@lbl.gov, jfbotts@lbl.gov, glock@lbl.gov, dhazen@lbl.gov

PDSF, the Parallel Distributed Systems Facility, has been in continuous operation since 1996 serving high energy physics research. It is currently a tier-1 site for Star, a tier-2 site for Alice and a tier-3 site for Atlas. We are in the process of migrating PDSF workload from commodity cluster to the Cori a Cray XC40 system. The process will involve preparing containers that will allow PDSF community to effectively move workloads between systems and other HPC centers. We are in the discovering process of optimizing serial jobs in a parallel environment. The goal is to minimize service interruptions as we shut down the existing PDSF by the second half of 2019.

Desired length:

15

Site reports / **140**

RAL Site Report

Author: Martin Bly[1]

[1] *STFC-RAL*

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

Update on activities at RAL

Desired length:

12

Networking and security / **141**

WLCG/OSG Networking Update

Authors: Shawn Mc Kee[1]; Marian Babik[2]

[1] *University of Michigan (US)*

[2] *CERN*

Corresponding Authors: marian.babik@cern.ch, shawn.mckee@cern.ch

WLCG relies on the network as a critical part of its infrastructure and therefore needs to guarantee effective network usage and prompt detection and
resolution of any network issues, including connection failures, congestion and traffic routing. The OSG Networking Area is a partner of the WLCG effort
and is focused on being the primary source of networking information for its partners and constituents. We will report on the changes and updates that have
occurred since the last HEPiX meeting.

The WLCG Network Throughput working group was established to ensure sites and experiments can better understand and fix networking issues. In addition,
it aims to integrate and combine all network-related monitoring data collected by the OSG/WLCG infrastructure from both network and transfer systems.
This has been facilitated by the already existing network of the perfSONAR instances that is being commissioned to operate in full production.

We will provide a status update on the LHCOPN/LHCONE perfSONAR infrastructure as well as cover recent changes in the higher level services due to reorganisation of the OSG. This will include details on the central service migrations, updates to the dashboards; updates and changes to the Web-based mesh configuration system and details on the newly established pipeline for processing perfSONAR results.

In addition, we will provide an overview of the recent major network incidents that were investigated with the help of perfSONAR infrastructure and provide
information on changes that will be included in the next perfSONAR Toolkit version 4.1. We will also cover the status of our WLCG/OSG deployment
and provide some information on our future plans.

Desired length:

20

Networking and security / **142**

Network Functions Virtualisation Working Group Update

Authors: Shawn Mc Kee[1]; Marian Babik[2]

[1] *University of Michigan (US)*

[2] *CERN*

Corresponding Authors: marian.babik@cern.ch, shawn.mckee@cern.ch

High Energy Physics (HEP) experiments have greatly benefited from a strong relationship with Research and Education (R&E) network providers and thanks to the projects such as LHCOPN/LHCONE and REN contributions, have enjoyed significant capacities and high performance networks for some time. RENs have been able to continually expand their capacities to over-provision the networks relative to the experiments needs and were thus able to cope with the recent rapid growth of the traffic between sites, both in terms of achievable peak transfer rates as well as in total amount of data transferred. For some HEP experiments this has lead to designs that favour remote data access where network is considered an appliance with almost infinite capacity. There are reasons to believe that the network situation will change due to both technological and non-technological reasons starting already in the next few years. Various non-technological factors that are in play are for example anticipated growth of the non-HEP network usage with other large data volume sciences coming online; introduction of the cloud and commercial networking and their respective impact on usage policies and securities as well as technological limitations of the optical interfaces and switching equipment.

As the scale and complexity of the current HEP network grows rapidly, new technologies and platforms are being introduced, collectively called Network Functions Virtualisation (NFV), ranging from software-based switches such as OpenVSwitch, Software Defined Network (SDN) controllers such as OpenDaylight up to full platform based open solutions such as Cumulus Linux. With many of these technologies becoming available, it's important to understand how we can design, test and develop systems that could enter existing production workflows while at the same time changing something as fundamental as the network that all sites and experiments rely upon. In this talk we'll give an update on the Network Functions Virtualisation (NFV) WG that was established at the last HEPiX meeting. We'll provide details on its mandate, objectives, organisation of work as well as areas of interest that were already discussed and plans for the near-term future.

Desired length:

15

Clouds, virtualisation, grids / **143**

Data analysis as a service

Author: James Adams[1]

[1] *STFC RAL*

Corresponding Author: james.adams@stfc.ac.uk

We are seeing an increasingly wide variety of uses being made of Hybrid Cloud (and Grid!) computing technologies at STFC, this talk will focus on the services being delivered to end users and novel integrations with existing local compute and data infrastructure.

Desired length:

20

Networking and security / **144**

Recent status of KEK network

Authors: Soh Suzuki[None]; Tadashi Murakami[1]; Tomoaki Nakamura[1]; Fukuko YUASA[1]; Teiji Nakamura[1]; Kiyoharu Hashimoto[1]; Atsushi Manabe[2]

[1] *KEK*

[2] *High Energy Accelerator Research Organization (KEK)*

Corresponding Authors: fukuko.yuasa@kek.jp, atsushi.manabe@kek.jp, teiji.nakamura@kek.jp, tadashi.murakami@kek.jp, kiyoharu.hashimoto@kek.jp, soh.suzuki@kek.jp

The Belle II detector is already taking data by cosmic ray test and is about to record data by beam. The importance of the network connectivity becomes higher than all other experiments in KEK. It is not only for the data transfer but also for researchers who are watching the condition of detectors from off sites.
We will report the present status of the campus network and the upgrade plan in this summer.

Desired length:

15minutes

Storage and file systems / 145

AFS and Linux Containers

Authors: Jeffrey Altman[1]; Marc Dionne[1]; David Howells[2]

[1] *AuriStor, Inc*

[2] *Red Hat*

Corresponding Authors: marc@auristor.com, dhowells@redhat.com, jaltman@auristor.com

One future model of software deployment and configuration is containerization.

AFS has been used for software distribution for many decades. Its global file namespace, the @sys path component substitution macro which permits file paths to be platform-agnostic, and the atomic publication model ("vos release") have proven to be critical components of successful software distribution systems that scale to hundreds of thousands of systems and have survived multiple OS and processor architecture changes.

The AuriStorFS security model consisting of combined-identity authentication, multi-factor authorization, and mandatory security policies permits a global name space to be shared between internal, dmz and cloud; and to store a mix of open and restricted data.

The combination of Linux Containers, the global AFS namespace, and the AuriStorFS security model is powerful permitting the development of container based software deployments that can safely bridge internal, dmz and cloud with reduced risk of data leaks.

This session will discuss the most recent updates to AuriStorFS and the Linux kernel implementation of AF_RXRPC socket family and (k)AFS filesystem. A demonstration will be included consisting of:

- Containers with binary executable files stored in /afs
- Containers mounting private AFS volume for scratch space
- AuriStorFS and (k)AFS file system implementations running side-by-side
- Linux namespaces for /afs

AuriStorFS milestones since HEPiX Spring 2017 include:

1. Successful migration and replication of volumes exceeding 5.5TB. The largest production volume so far is 50TB with a 250TB volume

2. Deployment of a single AuriStorFS cell spanning an internal data center, AWS and GCP with more than 25,000 nodes for distribution of software and configuration data.
3. Meltdown and Spectre remediation. In response to nearly 30% performance hit from Meltdown and Spectre the AuriStor team optimized the Rx stack, Ubik database and fileserver to reduce the number of syscalls by more than 50%
4. AES-NI, SSSE3, AVX and AVX2 Intel processor optimization of AES256-CTS-HMAC-SHA1-96 cryptographic operations for kernel cache managers reduces computation time by 64%

AF_RXRPC and kAFS highlights:

- IPv6 support for AuriStorFS
- dynamic root mount -o dyn
- @sys and @cell support
- multipage read and write support
- local hero directory caching
- per file acls (for AuriStorFS)
- server failover and busy volume retries

Desired length:

20 minutes

Clouds, virtualisation, grids / **146**

RAL Cloud update

Authors: Ian Collier[1]; Alexander Dibbo[2]

[1] *Science and Technology Facilities Council STFC (GB)*

[2] *STFC RAL*

Corresponding Authors: ian.peter.collier@cern.ch, alexander.dibbo@stfc.ac.uk

As our OpenStack cloud enters full production, we give an overview of the design and how it leverages the RAL Tier 1 infrastructure & support. We also present some of teh new use cases and science being enabled by the cloud platform.

Desired length:

20 minutes

End-user services and operating systems / **147**

CC-IN2P3 User Portal

Authors: Gino Marchetti[1]; Cyril L'Orphelin[2]; Renaud Vernet[3]

[1] *CNRS*

[2] *CNRS/IN2P3*

[3] *CC-IN2P3 - Centre de Calcul (FR)*

Corresponding Authors: cyril.lorphelin@cc.in2p3.fr, gino.marchetti@cc.in2p3.fr, renaud.vernet@cern.ch

CC-IN2P3 is one of the largest academic data centers in France. Its main mission is to provide the particle, astroparticle and nuclear physics community with IT services, including largescale compute and storage capacities. We are a partner for dozens of scientific experiments and hundreds of researchers that make a daily use of these resources. The CC-User Portal project's goal is to develop a web portal providing the users with a single-entry point to monitor their activity and incidents, to receive vital information and have the necessary links in order to access and use our services efficiently.

During HEPiX Fall 2017, we presented our first developments and exchanged with the community our thoughts on how to display the information to the users. With this presentation we would like to show what is now deployed in production and which features we are already developing to complete the web portal offer and meet the users' needs.

Desired length:

20 minutes

Site reports / **148**

AGLT2 Site Update

Author: Shawn Mc Kee[1]

Co-authors: Robert Ball [1]; Philippe Laurens [2]

[1] *University of Michigan (US)*

[2] *Michigan State University (US)*

Corresponding Authors: ball@umich.edu, shawn.mckee@cern.ch, philippe.laurens@cern.ch

We will present an update on our site since the Fall 2017 report, covering our changes in software, tools and operations.

Some of the details to cover include the enabling of IPv6 for all of our AGLT2 nodes, our migration to SL7, exploration of the use of Bro/MISP at the UM site, the use of Open vSwitch on our dCache storage and information about our newest hardware purchases and deployed middleware.

We conclude with a summary of what has worked and what problems we encountered and indicate directions for future work.

Desired length:

12

Basic IT services / **149**

Teraflops of Jupyter: A Notebook Based Analysis Portal at BNL

Authors: Ofer Rind[None]; Doug Benjamin[1]; Thomas Throwe[2]; William Strecker-Kellogg[3]

[1] *Duke University (US)*

[2] *BNL*

[3] *Brookhaven National Lab*

Corresponding Authors: rind@bnl.gov, douglas.benjamin@cern.ch, throwe@bnl.gov, willsk@bnl.gov

The BNL Scientific Data and Computing Center (SDCC) has begun to deploy a user analysis portal based on Jupyterhub. The Jupyter interfaces have back-end access to the Atlas compute farm via Condor for data analysis, and to the GP-GPU resources on the Institutional Cluster via Slurm, for machine learning applications. We will present the developing architecture of this system, current use cases and results, and discuss future plans.

Desired length:

20 minutes

Networking and security / **150**

Network status at IHEP and LHCONE progress in China

Author: Shan Zeng[1]

Co-author: Fazhi QI

[1] *Chinese Academy of Sciences (CN)*

Corresponding Authors: fazhi.qi@ihep.ac.cn, zengshan@ihep.ac.cn

Present the Network status at IHEP and LHCONE progress in China

Desired length:

15

Miscellaneous / **151**

Workshop wrap-up

Storage and file systems / **152**

AFS Update: Spring 2018

Author: Jeffrey Altman[1]

[1] *AuriStor, Inc.*

Corresponding Author: jaltman@auristor.com

Last May it was announced “AFS” was awarded the 2016 ACM System Software Award. .This presentation will discuss the current state of the AFS file system family including:

- IBM AFS 3.6
- OpenAFS
- kAFS
- AuriStor File System

IBM AFS 3.6 is a commercial product no longer publicly available.

OpenAFS is fork from IBM AFS 3.6 available under the IBM Public License 1.0. The currently supported release branches are 1.6 and 1.8.

AuriStorFS is a commercial file system that is backward compatible with both IBM AFS 3.6 and OpenAFS clients and servers. AFS cells hosted on AuriStorFS servers benefit from substantial improvements in

kAFS is an AFS and AuriStorFS client distributed as part of the mainline Linux kernel distribution. kAFS shares no source code with IBM AFS or OpenAFS.

Desired length:

20 minutes

Site reports / **153**

IHEP Site Report

Author: Jingyan Shi[1]

[1] *IHEP*

Corresponding Author: shijy@ihep.ac.cn

The computing center of IHEP maintains a HTC cluster with 10,000 cpu cores and a site including about 15,000 CPU cores and more than 10PB storage. The presentation will talk about the its progress and next plan of IHEP Site.

Desired length:

12

IT facilities / **154**

INFN-T1 flooding report

Authors: Stefano Dal Pra[1]; Andrea Chierici[2]; Luca dell'Agnello[1]

[1] *INFN*

[2] *INFN-CNAF*

Corresponding Authors: chierici@cnaf.infn.it, luca.dellagnello@cern.ch, stefano.dalpra@cnaf.infn.it

On November 9 2017, a major flooding occurred in the computing rooms: this has turned into a down of all the services for a prolonged period of time.
In this talk we will go through all the issues we faced in order to recover the services in the quickest and most efficient way; we will analyze in detail the incident and all the steps made to recover the

computing rooms, electrical power, network, storage and farming.
Moreover, we will discuss the hidden dependencies among services discovered during the recovery of the systems and will detail how we solved them.

Desired length:

15

Site reports / **155**

University of Nebraska CMS Tier2 Site Report

Author: Garhan Attebury[1]

[1] *University of Nebraska Lincoln (US)*

Corresponding Author: garhan.attebury@cern.ch

Updates from T2_US_Nebraska covering our experiences operating CentOS 7 + Docker/Singularity, random dabbling with SDN to better HEP transfers, involvement with the Open Science Grid, and trying to live the IPv6 dream.

Desired length:

12

IT facilities / **156**

Planning new datacenter network architecture

Author: Szilvia Racz[1]

[1] *Wigner Datacenter*

Corresponding Author: racz.szilvia@wigner.mta.hu

In scope of the Wigner Datacenter cloud project we are consolidating our network equipment. According to our plans we would like to purchase 100 Gbps datacenter switches in order to anticipate our current and future needs. We need automated, vendor neutral and easily operable network. This presentation highlights our requirements and design goals, candidates we have tested in our lab. We take the opportunity here to introduce our knowledge lab initiative where we can expand the scope of testing solutions.

Desired length:

15

Storage and file systems / **157**

The OpenAFS Foundation

Authors: Todd DeSantis[1]; Margarete Ziemer[2]

[1] *IBM*

[2] *Sine Nomine Associates*

Corresponding Authors: emziemer@sinenomine.net, atd@us.ibm.com

We would like to have one of the Board members of The OpenAFS Foundation, Inc, speak about this 501(c)(3), US-based, non-profit organization dedicated to fostering the stability and growth of OpenAFS, an open source implementation of the AFS distributed network filesystem. The OpenAFS Foundation adopted a three-fold mission: to attract and increase the community of OpenAFS users, to foster the OpenAFS community of experts, and to nurture and evolve the OpenAFS technology; each will be explained briefly.

We would like to ask for help from the scientific community and ask its researchers to:

- Contribute to the OpenAFS code, as such contributions are critical for the survival and improvement of the OpenAFS technology. The Foundation has obtained insurance to protect all contributors from potential liability and infringement of IP lawsuits.
- Reviewing code, as their feedback is not only valuable and appreciated in several ways.
- Write documentation for already existing code, which is desperately needed.
  Communicate the changes to their computing needs, and how they would like OpenAFS to be even more useful to them in the future.
- Help craft code specifications and/or code design.
- Become a guardian and as such, an active, long-term champion shaping the future viability and well-being of OpenAFS technology.
- Donate and/or identify organizations possibly willing and able to contribute funds to sustain a lean operation and/or to fund specific development efforts to be assigned in an open bid process.

The presenter will be either Todd DeSantis or Margarete Ziemer, both Directors and Board members of the Foundation.

Desired length:

20 minutes

Computing and batch systems / **158**

HPL and HPCG Benchmark on BNL linux farm and SDCC

Author: Zhihua Dong[None]

Corresponding Author: dong10027@gmail.com

HPL and HPCG Benchmark on Brookhaven National Laboratory SDCC clusters and various generations of Linux Farm nodes has been conducted and compared with HS06 results. While HPL results are more aligned with CPU/GPU performance. HPCG results are impacted by memory performances as well.

Desired length:

15

Site reports / **159**

PDSF Site Report

Author: Georg Rath[1]

Co-authors: James Botts [2]; Jeff Porter [3]; Jan Balewski [4]; Douglas Jacobsen [4]; Lisa Gerhardt [2]; Tina Declerck [4]; Tony Quan [5]; Ershaad Basheer [4]

[1] *Lawrence Berkeley National Laboratory*

[2] *LBNL*

[3] *Lawrence Berkeley National Lab. (US)*

[4] *NERSC*

[5] *LBL*

Corresponding Authors: gbrath@lbl.gov, balewski@lbl.gov, dmjacobsen@lbl.gov, twquan@lbl.gov, ebasheer@lbl.gov, tmdeclerck@lbl.gov, rjporter@lbl.gov, lgerhardt@lbl.gov, jfbotts@lbl.gov

PDSF, the Parallel Distributed Systems Facility, was moved to Lawrence Berkeley National Lab from Oakland CA in 2016. The cluster has been in continuous operation since 1996 serving high energy physics research. The cluster is a tier-1 site for Star, a tier-2 site for Alice and a tier-3 site for Atlas.

This site report will describe lessons learned and challenges met, when migrating from Univa GridEngine to the Slurm scheduler, experiences running containerized software stacks using Shifter, as well as upcoming changes to systems management and the future of PDSF.

Desired length:

12

Storage and file systems / **160**

OpenAFS Release Team report

Authors: Michael Meffie[1]; Benjamin Kaduk[2]; Stephan Wiesand[3]

[1] *Sine Nomine Associates*

[2] *MIT (formerly)*

[3] *DESY*

Corresponding Authors: stephan.wiesand@desy.de, mmeffie@sinenomine.net, kaduk@mit.edu

A report from the OpenAFS Release Team on recent OpenAFS releases, including the OpenAFS 1.8.0 release, the first major release in several years. Topics include acknowlegement of contributors, descriptions of issues recently resolved, and a discussion of commits under review for post 1.8.0.

Desired length:

20 minutes

Networking and security / **161**

IPv6 Deployment Experience at the GridKa Tier-1 at KIT

Author: Andreas Petzold[1]

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

Corresponding Author: andreas.petzold@cern.ch

Recently, we've deployed IPv6 for the CMS dCache instance at KIT. We've run into a number of interesting problems with the IPv6 setup we had originally chosen. The presentation will detail the lessons we've learned and the resulting redesign of our IPv6 deployment strategy.

Desired length:

10