

International Symposium on Grids & Clouds 2016 (ISGC 2016)

Sunday 13 March 2016 - Friday 18 March 2016

Academia Sinica

Book of Abstracts

Contents

Elastic extension of a CMS Computing Centre resources on external Clouds (remote presentation)	1
Kipper – a Grid bridge to Identity Federation	1
West-Life: A VRE for Structural Biology	2
Synergy: a service for optimising resource allocation in cloud based environments (remote presentation)	3
Enhanced Utilization and Allocation of Grid Resources	4
Considerations on using CernVM-FS for dataset sharing within various research communities	5
Visualization of dCache billing files with modern BigData tools.	5
Using containers to manage dCache	6
dCache on steroids - delegated storage solutions	6
Quality of Service in storage and the INDIGO-DataCloud project.	7
Opportunistic usage of the CMS online cluster using a cloud overlay	8
Scaling the Geneva Library Collection to large HPC clusters	8
Involving the public into HEP through IT challenges and projects	9
Agent-Based Modelling And Simulation For The Geospatial Network Model Of The Roman World	10
Experience of IPv6 and Cloud Deployment at NCP	11
High Performance Computing System for Scientific Applications	11
Beauty@LHC: The WMSecureGW service to interface untrusted volunteers machines to the DIRAC System.	12
Authentication and Authorisation for Research and Collaboration	12
Who doesn't need to be WISE?	13
Finding the Optimum Resolution, and Microphysics and Cumulus Parameterization Scheme Combinations for Numerical Weather Prediction Models in Northern Thailand: A First	

Step towards Aerosol and Chemical Weather Forecasting for Northern Thailand	14
SuMegha Cloud Kit: Create Your Own Private Scientific Cloud	15
Managing Virtual Appliance Lifecycle in IaaS and PaaS Clouds	16
A Ubiquitous Urgent Computing Framework for Ensembles of Flash Flood Forecasts . . .	16
EU-India cooperation on e-Infrastructures: from EU-IndiaGrid to e-INIT project	17
A Scientific Paper Reproducible Environment with Overlay Cloud Architecture	18
C++ and the delivery of portable program binaries in the Linux ecosystem	19
ICT-Enhanced Interactive Writing Program for International Students - A Plagiarism-free Writing Program -	20
Educational Informatics: A Paradigm Shift for IT-enhanced Education	21
Networking for WLCG: LHCOPN and LHCONE	21
IPv6 deployment at CERN	22
The Impact of the UNIX Scheduler on IO Dominated Applications	23
The EUDAT Service Suite and the Data Lifecycle	23
A Survey of Chinese Spallation Neutron Source Cloud Computing System	24
IPv6 Security	24
A Study of Certification Authority Integration Model in a PKI Trust Federation on Dis- tributed Infrastructures for Academic Research	24
Raising Security and Trust in our Inter-Federated World	25
DE-KIT(GridKa) – 100G @ LHCONE	26
LHCb experience during the LHC 2015 run	26
Macarons in dCache: sharing data made easy	27
INDIGO-DataCloud: enabling collaboration in an identity-rich world	28
Installation of Public Opinion Sensors: How a Community Based Web Survey Platform Facilitates the Collection of Qualitative Opinion Data	28
Options for the evolution of the LHCb Computing Model for LHC run 3	29
High-Throughput Processing of Space Debris Data	30
The Cluster Monitoring System of IHEP	30
Context-aware cloud computing for HEP applications	31
Data Processing and Visualization for Global High-Resolution Climate Simulations . . .	31
KEK Central Computer System	32

Mesos in a WLCG Tier 1 Grid Site	32
The Networking of the IHEP Data Center	33
Improvement of Scalability in Sharing Visualization Contents for Heterogeneous Display Environments	33
Elastic CNAF DataCenter extension via opportunistic resources	34
Seasonal Ensemble Forecasting Application On SuMegha Scientific Cloud Infrastructure	35
Review and evaluation of the outcome of Pre-University Program	36
DIRAC Data Management Framework	37
Advanced visualisation for environmental computing	37
Automatic dynamic stack management in large scientific applications: A case study using a global spectral model	38
Computer Security Landscape	39
A blueprint for Environmental Computing Applications	39
Evolving the RAL Tier 1 - Beyond High Energy Physics	40
The Dutch National e-Infrastructure	40
Generic Policy Engine for Virtual Infrastructure	41
Recommending Majors of Students by using Fuzzy Modeling	41
EGI-CSIRT: Organising Operational Security in evolving distributed IT-Infrastructures .	42
Modeling the Past and Future of Identity Management for Scientific Collaborations . . .	42
Open Platform for Academic Humanities Data	43
State Space Models in Analyzing Big Election Data	44
Serious Game Experimentation for Measuring Trust	45
The Implementation of Combining Problem Based Learning and Hands-on Learning . .	46
Importance of User Deprovisioning from Services	46
4th system upgrade of Tokyo Tier2 center	47
iRODS 4.2: Increasing Data Management Programming Productivity and Capability with the Pluggable Rule Engine Architecture	47
The Inevitable End of Moore's Law beyond Exascale will Result in Data and HPC Conver- gence and More	48
Building Smart Regions Through Multi-Sector Collaborations	49
eScience Activities in Japan	49

eScience Activities in China	49
eScience Activities in Korea	49
eScience Activities in Taiwan	49
eScience Activities in Mongolia	50
Panel Discussion	50
eScience Activities in Thailand	50
eScience Activities in Indonesia	50
eScience Activities in Malaysia	50
eScience Activities in Vietnam	50
eScience Activities in Philippine	50
Panel Discussion	50
eScience Activities in India	50
eScience Activities in Pakistan	51
eScience Activities in Singapore (remote presentation)	51
eScience Activities in Australia	51
Panel Discussion	51
Introduction	51
Mechanics of Interfederation	51
Requirements for assurance	51
Common attack vectors & threats	51
Proper communications mechanisms with reference to IR	52
Introduction to Exercises	52
Hands on session I	52
Hands on session II	52
General introduction to Cryo-EM single particles technique	52
Introduction to Scipion framework	52
Movies alignment (Xmipp)	52
CTF estimation and screening (Ctffind4)	52
Particle picking (Xmipp and Eman)	53
2D classification (Spider)	53

More picking with templates (Relion)	53
2D classification (Relion and Xmipp, maybe show precomputed results)	53
Initial volume methods (Several packages)	53
3D classification and refinement (Relion, maybe show precomputed results)	53
TransPAC – Pragmatic Application-Driven International Networking	53
European Research Infrastructures for Heritage Science	54
Introduction	54
Mechanisms of Interfederation	54
Trust Assurance and Policy for Federated Identity	54
Common attack vectors & threats	54
Proper communications mechanisms with reference to IR	54
Introduction to Exercises	55
Hands on session I	55
Hands on session II	55
The Cutting-Edge Technologies	55
The Cutting-Edge Technologies	55
Hands-on Exercise	55
Hands-on Exercise	55
Introduction	55
Hands-on Exercise	56
Opening Remarks	56
Complex Big Data: The road to bad policy?	56
Kai Nan	56
NMRbox: Toward Reproducible Computation for Bio-NMR	56
Closing Ceremony	56
Building Smart Regions Through Multi-Sector Collaborations	56
Prediction of drug targets by curated pharmacophore database	56
Meeting Introduction	57
LHCOPN status update	57
IPv6 deployment update	57

WLCG Workshop 2016 report	57
LHCONE operations update	57
LHCONE VRFs status update	57
LHCONE reachability measurement	57
perfSONAR introduction and update	57
TEIN activities and VRF implementation update	58
ASGC activities and VRF implementation update	58
Belle II update	58
Update on SINET5 implementation	58
Commercial Cloud services for WLCG	58
GEANT approach to Cloud Providers' R&E traffic	58
NORDUnet views on cloud services	58
LHCONE P2P service	58
CAS Clouds: a Case Study of Community Cloud	59
Time Series Analysis of Protein Dynamics	59
Efficient Large Scale Biomedical Data Analysis	59
A Novel Approach of Dimensionality Reduction for Revealing Protein Functional Dynamics	59
Facing Computing and Data Management Challenges in Electron Microscopy: The Scipion Software Framework	59
Introduction	59
CAS Clouds: A Case Study of Community Cloud	60
Microsoft Intelligent Cloud	61
The Development of Storm Surge Forecasting System and the Case Study of 2013 Typhoon Haiyan	61
Numerical study of Typhoon Haiyan (2013) with WRF model	61
Distributed Cloud for e-Science	61
Welcome speech and ASGC update	61
Discussion time on LHCONE Asia future development	61
Clouds discussion time	61
LHCONE P2P activities introduction	61

Next meeting and wrap-up	62
Installation of Public Opinion Sensors: How a Community Based Web Survey Platform Facilitates the Collection of Qualitative Opinion Data	62
e-Science Activities in Mongolian Academy of Sciences	62
GIS of Dr Sun Yat-sen, a life history	62
Spatial Humanities Going Dutch: Historical Mapping of 17th-Century Formosa Manuscripts	63
The Taiwan Missions Project - Late 19th Century (1865-1895)	63
Status of Austronesian Voyaging and Religions	63
CBETA-RP – an Integrated Digital Research Platform of Chinese Buddhism Materials	63
Offline Full Text Search on iOS and Android for Chinese Tripitaka and Tibetan Kangyur	63
The Fo Guang Dictionary of Buddhism Project	63
TBA	63
Digital Transformations of the Cultural Imaginary	64
Digital and Physical Preservation: Comparing and Bridging Records of the Xindian First Graveyard	64
Maritime Buddhism:Current state of research	64
ECAI Project and Content Collection Updates	64
ECAI Data Portal and SinicaView 4D GIS Platform	64
ECAI Community Discussion - Project Updates	64
Arches - to register heritage, Getty conservation	64
Mapping Chinese Taoist religious networks	64
Update on GEANT projects and plans	64
Russian VRF introduction and update	65
APAN update	65
Introduction to Environmental Computing Workshop	65
Development and Implementation of a Global-to-Urban Climate Model Suite	65
Discussion	65
The big picture of environmental computing	65
Mekong Delta	65
Environmental Exascale Computing	66

Land Use Development Simulation Systems	66
Application of numerical model on extreme weather and environmental studies	66
The Applications of Advanced Numerical Simulation on the Tsunami and Flooding Hazard Mitigation	66
Panel Discussion	66
Panel Discussion	66
Discussion time on LHCONE Asia future development	66
ESnet activities and plans	66
NORDUnet view on cloud services	67
Cloud discussion time	67
Monitoring at a Multi-Site Tier 1	67
Visualization of dCache accounting information with state-of-the-art Data Analysis Tools	67
CERN Network and WLCG Network Monitoring	67
Introduction of IHEP Monitoring System	67
eScience Activity in Nepal	67
BGP Route Server Proof of Concept (remote presentation)	68
Caltech on SDN activities	68
Getting SDN to DevOps in ATLAS	68
Questions and Next steps	68
Nova experiment request to join LHCONE	68
KISTI activities and plans	68

Infrastructure Clouds and Virtualisation Session I / 1**Elastic extension of a CMS Computing Centre resources on external Clouds (remote presentation)****Author:** Giuseppe Codispoti¹**Co-authors:** Alessandro Costantini ²; Claudio GRANDI ³; Cristina Aiftimiei ⁴; Daniele Bonacorsi ⁵; Diego Michelotto ²; Donato DeGirolamo ²; Gianluca Peco ⁶; Giovanni Zizzi ²; Massimo Sgaravatto ⁷; Matteo Panella ²; Patrizia Calligola ⁶; Riccardo Di Maria ⁸; Sonia TTaneja ²; Stefano Dal Pra ²; Vincenzo Ciaschini ²; Vladimir Sapunenko ²¹ *INFN & Bologna University*² *CNAF Bologna*³ *INFN - Bologna*⁴ *CNAF Bologna and IFIN-HH, Magurele*⁵ *University of Bologna*⁶ *INFN Bologna*⁷ *Universita e INFN, Padova (IT)*⁸ *Universita e INFN, Bologna (IT)***Corresponding Author:** giuseppe.codispoti@bo.infn.it

After the successful LHC data taking in Run-I and in view of the future runs, the LHC experiments are facing new challenges in the design and operation of the computing facilities. The computing infrastructure for Run-II is dimensioned to cope at most with the average amount of data recorded. The usage peaks, as already observed in Run-I, may however originate large backlogs, thus delaying the completion of the data reconstruction and ultimately the data availability for physics analysis. In order to cope with the production peaks, CMS - along the lines followed by other LHC experiments as well - is exploring the opportunity to access Cloud resources provided by external partners or commercial providers. Specific use cases have already been explored and successfully exploited during Long Shutdown 1 (LS1).

In this work we present the proof of concept of the elastic extension of a CMS site, specifically the Bologna Tier-3, on an external OpenStack infrastructure. We focus on the “Cloud Bursting” of a CMS Grid site using a newly designed LSF configuration that allows the dynamic registration of new worker nodes to LSF. In this approach, the dynamically added worker nodes instantiated on the OpenStack infrastructure are transparently accessed by the LHC Grid tools and at the same time they serve as an extension of the farm for the local usage. The amount of resources allocated thus can be elastically modeled to cope up with the needs of CMS experiment and local users. Moreover, a direct access/integration of OpenStack resources to the CMS workload management system is explored.

In this paper we present this approach, we report on the performances of the on-demand allocated resources, and we discuss the lessons learned and the next steps.

Networking, Security, Infrastructure & Operations Session III / 2**Kipper – a Grid bridge to Identity Federation****Author:** Andrey Kiryanov¹¹ *CERN*

Identity Federation (IdF, aka Federated Identity) is the means of interlinking people’s electronic identities stored across multiple distinct identity management systems. This technology has gained momentum in the last several years and is becoming popular in academic organisations involved in

international collaborations.

One example of such federation is eduGAIN, which interconnects European educational and research organisations, and enables trustworthy exchange of identity-related information.

In this work we will show an integrated Web-oriented solution code-named Kipper with a goal of providing access to WLCG resources using a user's IdF credentials from their home institute with no need for user-acquired X.509 certificates.

Kipper achieves "X.509-free" access to Grid resources with the help of two additional services: STS and IOTA CA. STS allows credential translation from SAML2 format used by Identity Federation to the VOMS-enabled X.509 used by most of the Grid, and the IOTA CA is responsible for automatic issuing of short-lived X.509 certificates.

Kipper comes with a JavaScript API considerably simplifying development of rich and convenient "X.509-free" Web-interfaces to Grid resources, and also advocating adoption of IOTA-class CAs among WLCG sites.

We will describe a working prototype of IdF support in the WebFTS interface to the FTS3 data transfer engine, enabled by integration of multiple services: WebFTS, CERN SSO (member of eduGAIN), CERN IOTA CA, STS, and VOMS.

Biomedicine & Life Science Session II / 3

West-Life: A VRE for Structural Biology

Author: Alexandre Bonvin¹

Co-authors: Anastassis Perrakis²; Antonio Rosato³; Chris Morris⁴; Jose Maria Carazo⁵; Ludek Matyska⁶; Marco Verlato⁷; Martyn Winn⁴; Ruty Francois⁸; Susan Daenke⁹; Victor Lamzin¹⁰

¹ *Utrecht University*

² *Netherlands Cancer Institute (NKI), Amsterdam*

³ *CIRMMP, Florence, Italy*

⁴ *STFC, Daresbury Laboratory, UK*

⁵ *CSIC Madrid, Spain*

⁶ *CESNET*

⁷ *INFN*

⁸ *Luna Technology, Paris, France*

⁹ *INSTRUCT, Oxford, UK*

¹⁰ *EMBL Hamburg*

Corresponding Author: a.m.j.j.bonvin@uu.nl

The focus of structural biology is shifting from single macromolecules produced by simpler prokaryotic organisms, to the macromolecular machinery of higher organisms, including systems of central relevance for human health. Structural biologists are expert in one or more techniques. They now often need to use complementary techniques in which they are less expert. INSTRUCT supports them in using multiple experimental techniques, and visiting multiple experimental facilities, within a single project.

The Protein Data Bank is a public repository where final structures and (some) of the data leading to them are deposited. Nowadays journals require such a deposition as a precondition of publication. However, metadata are often incomplete.

West-Life will pilot an infrastructure for storing and processing data that supports the growing use of combined techniques. There are some technique-specific pipelines for data analysis and structure determination but little is available in terms of automated pipelines to handle integrated datasets. Integrated management of structural biology data from different techniques is lacking altogether.

West-Life will integrate the data management facilities and services (e.g. from WeNMR) that already exist, and enable the provision of new ones. The resulting integration will provide users with an overview of the experiments performed at the different research infrastructures visited, and links to the different data stores. It will extend existing facilities for processing this data. As processing is performed, it will automatically capture metadata reflecting the history of the project. The

effort will use existing metadata standards, and integrate with them new domain-specific metadata terms.

West-Life will provide the application level services specific to uses cases in structural biology, enabling structural biologists to get the benefit of the generic services developed by EUDAT and the EGI.

Summary:

West-Life is a H2020 Virtual Research Environment project that will provide the application level services specific to uses cases in structural biology, covering all experimental techniques (e.g. Xray, cryo-EM, NMR, SAXS), enabling structural biologists to get the benefit of the generic services developed by EUDAT and the EGI.

Infrastructure Clouds and Virtualisation Session I / 4

Synergy: a service for optimising resource allocation in cloud based environments (remote presentation)

Author: L. Zangrando¹

Co-authors: Federica Fanzago²; Marco Verlatto²; Massimo Sgaravatto²

¹ INFN - Sez. Padova

² INFN

Corresponding Authors: zangrando@pd.infn.it, marco.verlatto@pd.infn.it

Computing activities performed by user groups in Public Research and Administration are usually not uniformly distributed over periods in the order a year. The amount of computing resources effectively used by such groups may vary significantly. In those well defined environments customers generally stipulate contracts with their Data Centers to guarantee the provisioning of an average computing level during a long period of time rather than of an agreed amount of computing resources that should be available at any given point in time.

In the current OpenStack scheduling model only fixed quotas can be granted to user groups. Those resources cannot be exceeded by one group even if there are unused resources allocated to other groups. Therefore, in a scenario of full resource usage for a specific project, new requests are simply rejected.

When resources are statically allocated amongst user groups, the global efficiency in the Data Center's resource usage may become quite low. The recently started European INDIGO-DataCloud project will address this issue through 'Synergy', a new advanced scheduling service to be integrated in the OpenStack Cloud Management Framework. 'Synergy' will adopt a resource provisioning model based on a fair-share algorithm to maximize resources usage. The INDIGO team is considering to use the SLURM multifactor fair-share algorithm for its first release. Beside the improved usage of resources, the algorithm guarantees that those resources are equally distributed among users and groups by accounting for the portion of the resources allocated to them and the resources already consumed in previous usage periods. Moreover, the mechanism will provide a persistent priority-queuing for handling user requests that can't be immediately fulfilled. Those requests will be processed as resources will become available again.

As our improvements are well structured and only have to be applied once to the upstream code repository, we don't expect significant maintenance efforts.

Starting from a list of selected requirements, 'Synergy' has to satisfy, this paper will provide a high level architecture of services, focusing on integration and interoperability aspects of existing OpenStack components, preferably those, which don't need to be extended. Along with preliminary results, we'll elaborate on the status of the current 'Synergy' implementation.

Summary:

The presentation will describe 'Synergy', a new advanced scheduling service developed by the INDIGO-DataCloud European project, to be integrated in the OpenStack Cloud Management Framework.

Enhanced Utilization and Allocation of Grid Resources

Author: Nam Beng Tan¹

¹ *Nanyang Polytechnic*

Corresponding Author: tan_nam_beng@nyp.edu.sg

Grid computing systems comprise of several resource nodes in a networked system. Each resource node may further comprise of a single machine or several machines forming another network. By utilizing the computing power of all these machines in this grid computing system, calculations that usually require super computers may be realized by the combined computing capability of all the machines in the system.

Presently, grid computing has become one of the effective ways for the sharing of resources and computing power over the internet. Several resource nodes of a grid computing system may be able to utilize each other's spare or idle computing power. An example of such sharing is when a resource node is located in a differing time zone, say at night where the computing power of the machines in that node are idle. Another resource node currently requiring extra computing power may be able to send jobs to this idle node to speed up the calculations or even to free up computing power in its own node for more crucial applications. Furthermore, present methods for assigning jobs to nodes are mainly based on a round robin queuing system that is disadvantageously arbitrary and takes no consideration of the characteristics of the idle or available resource nodes.

Our proposed solution thus provides a method and system for monitoring and managing grid resources in a grid computing environment with a plurality of resource nodes and at least one resource broker/supervisor node. The resource broker serves to retrieve information stored in a attribute repository about the resource nodes. All the resource nodes would send and update this information on the attributes and status of the resource nodes to the attribute repository periodically. The step of job brokering starts with the step of checking the attribute repository for attributes and status for all resource nodes in the grid computing environment. The information retrieved by the resource broker from the attribute repository may be attributes such as CPU utilization, Random Access memory (RAM) capacity or Virtual Memory (VM) Capacity. The status of the resource nodes may also be retrieved.

Next, the step of receiving a first priority attribute and a second priority attribute for the matching of resource nodes that may be best suitable for performing the job. These first and second priority attributes may be decided by a user based on the requirements of the job. For example, the job submitted may require very large RAM capacity for handling large amounts of variable data, the first priority attribute will then be RAM. The second priority may then be CPU utilization, depending on the user's requirements for the job. The step of matching available resource nodes corresponding to the first and second priority attributes is then performed. Following which the step of recommending the matching resource node to the job submitter is then determined.

Summary:

This paper describes an effective mean of determining the optimum resource for job execution in a Grid environment that will further improve the performance of job computation which Grid Computing offers. In summary, we proposed a solution with the following features :

- An improved method for Grid job submission, monitoring and management
- Monitoring Grid processes and switching between Grid resources in the event of a process failure
- Job brokering to determine the "Best match" grid resource to perform the job request
- Distributing jobs to available resources whereby the number of jobs exceeded the number of resources
- Notification in the event of a process failure through SMS, Email and messaging.

Physics & Engineering Session / 6**Considerations on using CernVM-FS for dataset sharing within various research communities**

Author: Catalin Condurache¹

¹ *STFC Rutherford Appleton Laboratory*

Corresponding Author: catalin.condurache@stfc.ac.uk

Firmly established as a method of software distribution for the Large Hadron Collider (LHC) experiments and many other research communities at Grid sites, the CernVM File System (CernVM-FS) is reaching now a new stage, its advantages starting to be acknowledged by communities activating within and making use of various Cloud computing environments.

As the manipulation of research data within various grid and cloud infrastructures (EGI FedCloud, TWGrid, OpenScience Grid) becomes more important for many communities, their members started to look into the CernVM-FS as a technology that could bring expected benefits.

Also the developers have optimized the technology for access to conditions data and other auxiliary data, and because of the use of standard technologies (http, squid) the CernVM-FS can now be used everywhere (local clusters, grid and cloud environments) and for more than software distribution.

The presentation will give an overview on the status of the EGI CernVM-FS infrastructure developed for the benefit of the non-LHC communities and its integration with other similar infrastructures for consolidated and better software and data access across the globe.

We will explain when CernVM-FS can be used for dataset sharing without losing the main benefits of the technology and then information on how to properly use it will be given.

Pros and cons will be discussed and available use cases from different High Energy Physics (HEP) and non-HEP research communities (i.e. Space, Natural and Life Sciences) will be analysed.

Data Management Session II / 7**Visualization of dCache billing files with modern BigData tools.**

Author: Tigran Mkrtchyan¹

Co-authors: Birgit Lewendel¹; Lusine Yakovleva¹; Patrick Fuhrmann²

¹ *DESY*

² *DESY/dCache.org*

Corresponding Author: tigran.mkrtchyan@desy.de

Running a data center is never a trivial job. In addition to daily routine tasks, service provider teams have to provide a meaningful information to upper management, end users and operators. The dCache production instances at DESY, produce gigabytes of billing files per day. Crunching the millions of numbers into a useful and handy information is unpleasant and boring task. Nevertheless, we can analyse and visualize this huge amount of log information with a modern BigData processing tools.

This presentation will show how we built a real-time monitoring system which visualizes dCache billing files and provides intuitive, simple to use web interface using Elasticsearch, Logstash and Kibana.

Infrastructure Clouds and Virtualisation Session II / 8**Using containers to manage dCache****Author:** Tigran Mkrtchyan¹**Co-authors:** Christian Bernardt¹; Karsten Schwank¹; Patrick Fuhrmann²; Paul Millar¹¹ DESY² DESY/dCache.org**Corresponding Author:** tigran.mkrtchyan@desy.de

For over a decade, dCache.ORG has provided robust software that is used at more than 80 Universities and research institutes around the world, allowing these sites to provide reliable storage services for the WLCG experiments and many other scientific communities. The flexible architecture of dCache allows its component services to be deployed in a wide variety of configurations and platforms, from a single Raspberry Pi up to hundreds of nodes in multi-petabyte infrastructures.

Even on multi node setups, it's common to run groups of services on the single node. This is motivated by the desire to optimise the performance (e.g., to reducing communication overhead), or simply to minimise the cost of running dCache. However, hosting dCache services on the same node implies the services are locked to the same dCache version: they must run the same dCache version and all services are upgraded at the same time.

Operating-System (OS) virtualization, often called containers, allows multiple isolated instances of a user space environment to run, while sharing the same kernel and OS. Unlike hardware virtualization, an application running in a container incurs little or no overhead making such deployments suitable for IO intensive applications such as dCache. There are a wide variety of container solutions, with almost all platforms providing at least one solution.

This presentation will show how we run dCache inside Docker, a popular Linux-based container solution. In addition to all benefits of the container technology, docker provides docker-hub - a place to store and share docker recipes. We will introduce two dCache containers - a full dCache installation, useful for dCache testing and evaluation, and a per-service container, allowing to execute and manage each dCache component in an independent container.

Data Management Session II / 9**dCache on steroids - delegated storage solutions****Author:** Tigran Mkrtchyan¹**Co-authors:** Gerd Behrmann²; Patrick Fuhrmann³; Paul Millar¹¹ DESY² NDGF³ DESY/dCache.org**Corresponding Author:** tigran.mkrtchyan@desy.de

For over a decade, dCache.ORG has provided robust software that is used at more than 80 Universities and research institutes around the world, allowing these sites to provide reliable storage services for the WLCG experiments and many other scientific communities. The flexible architecture of dCache allows running it in a wide variety of configurations and platforms - from all-in-one Raspberry-Pi up to hundreds of nodes in multi-petabyte infrastructures.

Due to lack of managed storage at the time, dCache implemented data placement, replication and data integrity directly. Today, many alternatives are available: S3, GlusterFS, CEPH and others. While such systems position themselves as scalable storage systems, they can not be used by many

scientific communities out of the box. The absence of specific authentication and authorization mechanisms, the use of product specific protocols and the lack of namespace are some of reasons that prevent wide-scale adoption of these alternatives.

Most of these limitations are already solved by dCache. By delegating low level storage management functionality to the above mentioned new systems and providing the missing layer through dCache, we provide a system which combines the benefits of both worlds - industry standard storage building blocks with the access protocols and authentication required by scientific communities.

In this presentation, we focus on CEPH, a popular software for clustered storage that supports file, block and object interfaces. CEPH is often used in modern computing centres, for example as a backend to OpenStack services. We will show prototypes of dCache running with a CEPH backend and discuss the benefits and limitations of such an approach. We will also outline the roadmap for supporting 'delegated storage' within the dCache releases.

Data Management Session I / 10

Quality of Service in storage and the INDIGO-DataCloud project.

Authors: Benjamin Ertl¹; Giacinto Donvito²; Marcus Hardt¹; Patrick Fuhrmann³; Paul Millar⁴

¹ *KIT*

² *INFN/BARI*

³ *DESY/dCache.org*

⁴ *DESY*

Corresponding Author: patrick.fuhrmann@desy.de

The pressure to provide cheap, reliable and unlimited cloud storage space in the commercial area has provided science with affordable storage hardware and open source storage solutions offering low maintenance costs and tuneable performance as well as durability properties, resulting in different cost models per storage unit. Those models, already introduced by WLCG a decade ago (disk vs tape) are now offered by large companies like Amazon (Bock Storage, S3 and Glacier) or Google (Standard, Durable Reduced, Cloud Storage Nearline). Industry appliances or software stacks (e.g. HPSS) offer similar storage properties for your locally installed computer centre storage space.

However, other than with SRM for WLCG, those offered storage quality properties don't follow a common description or specification and are hard to compare programatically. Moreover they aren't even close to a common way of been negotiated between the requesting client and the providing storage technology, which would be a prerequisite for federating different public and private storage services as planned by EGI and EUDAT.

To fill this gap, the INDIGO-DataCloud project is proposing a process to agree on common semantics in describing QoS attributes in storage in a consistent way, independently of the used API or protocol.

The process involves gathering uses-cases from scientific communities and creating working groups within international organisations, like RDA, OGF and SNIA to further discuss possible solutions with other interested parties like EUDAT and EGI. The activity has already been presented at the Paris RDA meeting and an RDA interest group is being prepared.

In a second step, based on feedback received, INDIGO will propose an implementation of the defined semantics to steer quality of service in storage. As a proof of concept INDIGO will implement the proposed solution in storage systems used within the INDIGO project, like dCache, StoRM, some typical industry products and a selected public storage cloud.

For within INDIGO-DataCloud a consistent QoS specification to uniquely describe the different distributed storage infrastructure components is essential to allow the INDIGO platform layer to broker the most appropriate endpoint for each individual data store, replication or access request.

We are presenting our work at ISGC, in order to receive feedback from interested communities in the Asian and Australian area as we believe that the approach is extremely useful beyond the activities within INDIGO-DataCloud.

This presentation will report on the goals achieved so far and on our next steps.

Infrastructure Clouds and Virtualisation Session II / 11

Opportunistic usage of the CMS online cluster using a cloud overlay

Author: Olivier chaze¹

¹ CERN

Corresponding Author: olivier.chaze@cern.ch

After two years of maintenance and upgrade, the LHC (Large Hadron Collider), the largest and most powerful particle accelerator in the world, has started its second three year run. Around 1500 computers make up the CMS (Compact Muon Solenoid) online cluster. This cluster is used for Data Acquisition of the CMS experiment at CERN, selecting and sending to storage around 20 TBytes of data per day that are then analysed by the WLCG (Worldwide LHC Computing Grid) infrastructure that links hundreds of data centres worldwide. 3000 CMS physicists can access and process data, and are always seeking more computing power and data. The backbone of the CMS online cluster is composed of 16000 cores which provide as much computing power as all CMS WLCG Tier 1 sites (352K HEP-SPECHS06 score in the CMS cluster versus 300K across Tier 1 sites). The computing power available in the CMS cluster can significantly speedup the processing of data, so an effort has been made to allocate the resources of the CMS online cluster to the GRID when it isn't used to its full capacity for data acquisition. This occurs during the maintenance periods when the LHC is non-operational, which happens one week out of every six. During 2016, the aim is to increase the availability of the CMS online cluster for data processing by making the cluster accessible during the time between two physics collision while the LHC and beams are being prepared. This is usually the case for a few hours every day, which would vastly increase the computing power available for data processing. Work has already been undertaken to provide this functionality, as an Openstack cloud layer has been deployed as a minimal overlay that leaves the primary role of the cluster untouched. This overlay also abstracts the different hardware and networks that the cluster is composed of. The operation of the cloud (starting and stopping the virtual machines) is another challenge that has been overcome as the cluster has only a few hours spare during the aforementioned beam preparation. By improving the virtual image deployment and integrating the Openstack services with the core services of the Data Acquisition on the CMS Online cluster it is now possible to start a thousand virtual machines within 10 minutes and to turn them off within seconds. This presentation will explain the architectural choices that were made to reach a fully redundant and scalable cloud, with a minimal impact on the running cluster configuration while giving a maximal segregation between the services. It will also present how to speed up of 25 the cold starting of 1000 virtual machines, using tools commonly utilised in all data centres.

Summary:

The Compact Muon Solenoid (CMS) experiment at CERN is maintaining a large infrastructure to readout and filter the data from the detector. The High Level Trigger (HLT) online cluster is composed of 16000 cores dedicated to online data and event filtering. However, this resource is used only for about 30% of the time, due to the Large Hadron Collider (LHC) duty cycle and the various maintenance periods thus the rest of time it is free. Only during these unused times an OpenStack cloud is started on top of the cluster allowing it to join the Worldwide LHC Computing Grid (WLCG) for offline data analysis.

Massively Distributed Computing and Citizen Sciences Session / 12

Scaling the Geneva Library Collection to large HPC clusters

Authors: Jan Knedlik¹; Kilian Schwarz¹; Lutz Matthias¹; Ruediger Berlich²

Co-authors: Ariel Garcia²; Sven Gabriel²

¹ *GSI Darmstadt*

² *Gemfony scientific UG (haftungsbeschaenkt)*

Corresponding Authors: r.berlich@gemfony.eu, k.schwarz@gsi.de

The Geneva Library Collection was designed to allow parametric optimization of demanding scientific and engineering problems in distributed and parallel computing environments. It has been successfully tested with a code relevant for hadron physics and is in production use in the automotive industry. Scalability of Genevas distributed execution model may be of particular importance for deployment scenarios involving complex models and simulations. Where these feature a non-trivial quality surface and many parameters to be optimized, as well as long running evaluation functions, computing demands may rise to hundreds or even thousands of cores, used over days or weeks. Scalability and stability in this case depend on many factors, some of which must be tuned in highly distributed environments. GSI Darmstadt has teamed up with the KIT spin-off Gemfony scientific to modify the Geneva library in such a way that it allows distributed execution on the new GSI Kronos Cluster, potentially involving more than thousand cores. The presentation introduces the steps that had to be taken both on the code- and the cluster-side to make Geneva scale to the desired level, and involves a practical demonstration on the target cluster.

Massively Distributed Computing and Citizen Sciences Session / 14

Involving the public into HEP through IT challenges and projects

Author: Claire Adam-Bourdarios¹

¹ *LAL*

Corresponding Author: claire.bourdarios@cern.ch

The ATLAS collaboration has recently setup three outreach projects and global challenges which have a strong IT component and could not have been envisaged without the growth of general public computing resources and network connectivity.

HEP has exciting and difficult problems like the extraction of the Higgs boson signal, and at the same time data scientists have advanced algorithms. The goal of the Higgs Machine Learning (HiggsML) project was to bring the two together by a “challenge”: machine learning experts could compete online to obtain the best Higgs $\rightarrow\tau\tau$ signal significance on a set of ATLAS fully simulated Monte Carlo signal and background events. The first challenge of this kind ran from May to September 2014, drawing considerable attention, and new projects followed in the context of the CERN open data initiative.

Higgs Hunters is the only physics-related project hosted on a web-based citizen science platform called Zooniverse. Volunteers usually contributing to space, natural world and humanities projects are asked to scan ATLAS events, looking for secondary vertices. Their results are compared to the ATLAS secondary vertex finding algorithm in the context of the search for long-lived particles in supersymmetric models.

ATLAS@home belongs to the well established family of BOINC projects: volunteers run simulations of collisions in the ATLAS detector on their highly distributed and heterogeneous personal computers, thanks to the Virtual Machine and ARC technologies. So far many thousands of members of the public signed up and already provide a significant fraction of ATLAS computing resources.

Each of these three axes of interaction with very specific communities are in development. In this talk, the setup, current success and future of such projects will be reviewed.

Humanities, Arts, and Social Sciences Session II / 15

Agent-Based Modelling And Simulation For The Geospatial Network Model Of The Roman World

Author: Jan Fousek¹

Co-authors: Adam Mertel²; Aleš Chalupa³; Eva Hladká¹; Eva Výtvarová¹

¹ Faculty of Informatics, Masaryk University, Brno, Czech Republic

² Department of Geography, Faculty of Science, Masaryk University, Brno, Czech Republic

³ Department for Study of Religions, Faculty of Arts, Masaryk University, Brno, Czech Republic

Corresponding Authors: izaak@mail.muni.cz, eva.vytvarova@gmail.com, eva@fi.muni.cz

Computational models have large potential for enhancing our understanding of human-environment interaction as a factor in various social and historical phenomena 1. One such an approach are agent-based models that provide useful model paradigm for human behavior 3. When coupled with geospatial data, such models can be spatially explicit, and have variety of applications in computational social science 2.

The ORBIS project [4, 5] provides a geospatial network model of travel in the Roman Empire. It combines the road network with maritime transport model derived from historical data, and provides cost and time expense prediction for given routes. The time and cost prediction take into account various influential factors, such as seasonal changes, distinguishes coastal and open sea routes, and onshore means of transport. Currently the online interface enables the researchers to examine routes between given locations, analyze distance from one location to all others, and visualize the importance of paths connecting a given location to the rest of the network.

In our case, we use agents traveling between the cities of the Roman Empire on routes defined by the ORBIS transportation model. As the preferred routes can change depending on season and other external factors, advancing the model from current average estimates to probabilistic distributions derived from the simulation will provide with better understanding and robustness to subsequent analysis. The agent-based approach allows for such probabilistic simulations, and it is a direct extension of the current model.

In this paper we present a computational environment for agent-based modelling on the ORBIS geospatial transport model. We provide web-based interface to specify the agent-based model and visualize the results of the simulation. We also enable the user to specify the parameters of the transportation model to create and visualize a static network, with the possibility to download the network for further analysis. The functionality of the environment is demonstrated on a model of diffusion process on the transport network.

1 Epstein, J. M. Why model? *Journal of Artificial Societies and Social Simulation*, 2008, 11, 12

- 2 Crooks, A. Brunson, C. & Singleton, A. (Eds.) Agent-Based Modeling And Geographical Information Systems Geocomputation: A Practical Primer, SAGE, 2015
- 3 Conte, R. & Paolucci, M. On agent-based modeling and computational social science *Frontiers in psychology*, Frontiers Media SA, 2014, 5
- 4 Scheidel, W.; Meeks, E. & Weiland, J. ORBIS: The Stanford Geospatial Network Model of the Roman World Stanford University Libraries, 2012
- 5 <http://orbis.stanford.edu/>

16

Experience of IPv6 and Cloud Deployment at NCP

Authors: Muhammad Tanzeel Murtaza Khan¹; Saqib Haleem²

Co-author: Fawad Saeed³

¹ *National Centre for Physics*

² *National Centre for Physics, Islamabad, Pakistan*

³ *National Centre of Physics*

Corresponding Author: saqib.haleem@ncp.edu.pk

National Centre for Physics (NCP) in Pakistan, maintains a large computing infrastructure for scientific community, including a Teir-2 site of Worldwide LHC Computing Grid (WLCG). Need for IP address space has been increased, due to expansion of infrastructure, and adoption of Cloud technology for hosting virtual machines. On the other side, IPv4 address space is almost near to depletion, and hence migration to IPv6 is inevitable. NCP is among the few organizations in the country, which is actively involved in promoting IPv6 and has deployed next generation IPv6 protocol in its campus network. NCP network is configured to provide IPv6 support by ensuring high availability of services, security, and optimized routing. Most of the corporate services are running in dual stack mode. WLCG site is also being tested and evaluated on IPv6. This paper discusses the detail about experience with IPv6 and current status of computing services over IPv6. With the complete deployment of IPv6 for WLCG services, storage pool nodes and computing nodes will need not to use NAT to communicate with the global WLCG nodes at different sites. Foreman and Puppet is being used for automated installation configuration of computing nodes at NCP. Physical resources are being monitored through Zabbix. OpenStack based cloud has also been deployed for efficient utilization of computing resources for different projects. Till now, we have tested the WLCG computing nodes deployment over private cloud using OpenStack. This paper also presents the plan, for availability of GRID resources through private cloud which will enable on demand creation of compute nodes.

17

High Performance Computing System for Scientific Applications

Authors: Muhammad Deen Jafar¹; Muhammad Tanzeel Murtaza Khan¹

Co-authors: Adeel Ur-Rehman¹; Saeed Fawad¹; Saqib Haleem²

¹ *National Centre for Physics*

² *National Centre for Physics, Islamabad, Pakistan*

Corresponding Author: jafar@ncp.edu.pk

Modern research institutes strive to offer smart computing systems to entertain robust applications modeled to address and solve eminent research problems being faced by the current scientific community. The central research institute in Pakistan dealing with various areas of Physics – NCP, emphasizes to compete with its small-scaled facility with other related research setups over the globe by deploying a computational cluster setup comprising of computing nodes having 56 physical as well as 24 virtual cores of CPUs serving under both serial and parallel fashioned batch queues. The virtual nodes are provisioned by private cloud using manual instance creation which will be automated in near future. Data persistence of the infrastructure running is governed by an IP SAN over iSCSI setup responsible to share its storage across the compute nodes involved. The state-of-the-art facility is dedicated to carry out the corresponding workflow simulations for the local research community spread across the various provinces of the country and is meant to operate round the clock without any foreseen interruption. The user community of the cluster comprises but not limited to researchers working in the fields of Computational Condensed Matter Physics, Climate Modeling, Density Functional Theory, Computational Plasma Physics, and Semi-conductor Nanostructures. Software packages/compiler support is available for Intel Fortran, gcc, Quantum Espresso, Gaussian09, Mopac2012, Turbo mole, Wien2k, and Geant4 etc. The results produced making use of this facility by submitting user jobs over it looks comparable to the those obtain in the past using similar computational clusters hosted by other international institutes. NCP anticipates enhancing the performance and capacity of its deployed facility by scaling it further to incorporate GPUs as well in the near future, thereby widening its scope for other impending research problems/areas.

Poster Session / 18

Beauty@LHC: The WMSecureGW service to interface untrusted volunteers machines to the DIRAC System.

Author: Cinzia Luzzi¹

Co-authors: Christophe Haen¹; Federico Stagni¹; Philippe Charpentier¹

¹ CERN

Corresponding Author: cinzia.luzzi@cern.ch

Considering the growing need of computing power, in addition to the experiment resources, the LHCb community aspires to profit also from volunteer computing.

Beauty@LHC is the LHCb volunteer computing project that aims to exploit opportunistic resources to run simulation jobs. The project uses the CERNVM Virtual Software Appliance, the Berkeley Open Infrastructure for Network Computing (BOINC) framework, and the DIRAC system for distributed computing.

A first prototype of Beauty@LHC was developed in 2013 and was used by volunteer users belonging to the LHCb Virtual Organisation. However, the architecture did not provide a secure technique to authenticate volunteers, a trusted host certificate was contained in the machine dispatched to the user. A secure authorization and authentication process was a mandatory requirement to open the project to the outside world and triggered the development of a gateway service called WMSecureGW (Workload Management System Secure Gateway).

The objective of the WMSecureGW service is to authenticate the BOINC users against the DIRAC framework and to authorize them to execute LHCb jobs.

This new service enables the execution of LHCb jobs by untrusted VMs bypassing the necessity of having a valid grid certificate to talk with the DIRAC services and thus allows the transition from the insecure volunteer computing world to the secure Grid computing environment.

The WMSecureGW runs on a trusted machine and accepts a dummy grid certificate signed by a dummy CA. The service is responsible for receiving all calls to different DIRAC services and to properly dispatch them. Before the real storage phase, the output data produced by the volunteer machines are uploaded on the gateway machine. Here a check has to be performed in order to avoid the storage of wrong data on the LHCb storage resources.

This paper describes the new architecture of the Beauty@LHC project and the implementation of the WMSecureGW service.

Networking, Security, Infrastructure & Operations I / 19**Authentication and Authorisation for Research and Collaboration**

Author: Alessandra Scicchitano¹

¹ GEANT

Corresponding Author: alessandra.scicchitano@geant.org

National Identity Federations for Research and Education have been globally emerging during the last decade together with a growing interest from different research communities in using federated access. Some challenges though still prevent the wide usage of this approach. The lack of seamless integration for example among the different AAI's operated by the various research collaborations and e-Infrastructures and the non-ubiquity of federated credentials and policy challenges are hindering the sharing of knowledge across the European and global research collaborations.

The AARC project addresses these challenges of interoperability and is driven by the principle to support the collaboration model across institutional and sectorial borders and to advance mechanisms that will improve the experience for users and guarantee their privacy and security.

The concept behind AARC is to build on existing AAI's used in the R&E sector, to analyse them in light of user requirements, integration aspects and security, and to design, test and pilot missing components and integrate them with existing working flows.

More over AARC aims at creating the conditions to allow identity to become a core element of e-infrastructure, as has been the case for network connectivity for many years already and at harmonising policies among e-Infrastructures to make it easy for resources and services providers to offer their services on a cross-border and cross-organisation basis.

The first part of the work has therefore been focus on collecting the community's requirements based on which develop a training package to support the uptake of federated access and a possible architecture to integrate existing AAI used in the R&E sector to enable access to services across different R&E infrastructures;

This presentation will provide an overview of the project and the results achieved so far in the training development and the design of the architecture as well as the adoption of common best practices and the deliver of related pilots.

Networking, Security, Infrastructure & Operations Session II / 20**Who doesn't need to be WISE?**

Author: Alessandra Scicchitano¹

¹ GEANT

Corresponding Author: alessandra.scicchitano@geant.org

Nowadays internet enables opportunities for research and development on a global scale. Researchers use the network to run their experiments in clouds and grid infrastructures and use the internet to exchange results and data. In order to do that, they have to have confidence that they can use the Internet for secure and reliable communication all across the world.

Real security on the Internet can only be realised within a broader context of trust and respect. Collaboration is the key to successful information security.

WISE stands for Wise Information Security for collaborating E-infrastructure and was born as a workshop from the joint effort of SIG-ISM (Special Interest Group on Information Security Management) and SCI (Security for Collaboration among Infrastructures).

The goal of the workshop was bringing together 4 big e-infrastructure EGI, EUDAT, GEANT and

PRACE in the same room in order to facilitate the exchange of experience and knowledge on information security management and other topics of relevance.

During the three days spent together in Barcelona in October 2015 where not only the e-infrastructures but also NRENs, XSEDE, NCSA, CTSC and communities like HEP/CERN, HBP and many others were present, a more profound need for such a collaboration together with the benefits that it could bring became evident.

The audience engaged in lively discussions on how to collaborate and help each other giving life to what today can be called the WISE community.

WISE aims at providing a trusted global framework where security experts can share information on different topics like risk management, experiences about certification process and threat intelligence.

This presentation will detail the outcome of the workshop and the work being undertaken by WISE to bring into reality this global collaboration, with a summary of the benefits to be gained by e-infrastructures and not only in joining the community.

Earth & Environmental Sciences & Biodiversity Session II / 21

Finding the Optimum Resolution, and Microphysics and Cumulus Parameterization Scheme Combinations for Numerical Weather Prediction Models in Northern Thailand: A First Step towards Aerosol and Chemical Weather Forecasting for Northern Thailand

Author: Ronald Macatangay¹

Co-authors: Boonrucksar Soonthornthum¹; Gerry Bagtasa²; Suparerk Aukkaravittayapun¹; Thiranan Sonkaew³; Utane Sawangwit¹

¹ *National Astronomical Research Institute of Thailand, Chiang Mai, Thailand*

² *Institute of Environmental Science and Meteorology, University of the Philippines, Diliman, Quezon City, Philippines*

³ *Lampang Rajabhat University, Lampang, Thailand*

Corresponding Author: ronmcd@gmail.com

Weather forecasts dictate our daily activities and allow us to respond properly during extreme weather events. However, weather forecasts are never perfect, but differences with model output and with observations can be minimized. Discrepancies between meteorological observations and weather model outputs are often caused by resolution differences (point vs. grid comparisons) and by the parameterizations used in the model. Atmospheric model parameterization refers to substituting small-scale and complicated atmospheric processes by simplified ones. In order to make weather forecasts more accurate, one can either increase the model resolution or improve the parameterizations used. Increasing model resolution can simulate small-scale atmospheric processes better, but takes a longer simulation time. On the other hand, improving model parameterization schemes involve in-depth measurements, analysis and research on numerous atmospheric processes. However, one can find a combination of existing parameterization schemes that would minimize observation-model differences. It is therefore essential to ask the question, "What model resolution and parameterization scheme combinations at a particular location and at particular seasons produce model output that has the smallest difference with observations simulated at a reasonable amount of time?"

Northern Thailand is a meteorologically active and unstable region especially during the summer and monsoon months (e.g. intense thunderstorms, hail storms, etc). It is also where high concentrations of air pollutants occur during the dry months (e.g. haze). It is therefore essential to have model forecasts close to observations for this region to reduce the risk from weather and from air quality degradation. This study aims to find the optimum model resolution and parameterization scheme

combinations at particular provinces in northern Thailand with available data during the wet and dry seasons that produces minimum differences with observations.

Nested model simulations are performed using the Weather Research and Forecasting (WRF) model (v. 3.6) ran in the High-Performance Computer (HPC) cluster of the National Astronomical Research Institute of Thailand (NARIT) for northern Thailand (2 km spatial resolution and hourly output), for the whole of Thailand (10 km spatial resolution and hourly output), and for the entire Southeast Asia (50 km spatial resolution and 3-hourly output). Combinations of the Lin et al., the WRF Single-Moment 5-class, the WRF Single-Moment 6-class and the WRF Double-Moment 6-class microphysics parameterization schemes, as well as the Kain-Fritsch, the Grell-Devenyi (GD) ensemble and the Grell 3D cumulus parameterization schemes would also be utilized to determine the optimum resolution and parameterization of the model when compared to observations. Meteorological data would come from selected weather stations in Chiang Mai, Chiang Rai and Lampang in northern Thailand from December 1-15, 2014 (cool dry season), from April 1-15, 2015 (warm dry season) and from August 1-15, 2015 (wet season).

Infrastructure Clouds and Virtualisation Session I / 22

SuMegha Cloud Kit: Create Your Own Private Scientific Cloud

Author: Vineeth Arackal¹

Co-authors: Arunachalam B¹; Kalasagar B¹; Mangala N¹; Prahlada Rao B B¹; Sarat Chandra Babu¹; Sukeshini R¹; Sumit Kumar¹

¹ *Centre for Development of Advanced Computing*

Corresponding Authors: kalasagarb@cdac.in, vineeth@cdac.in, barunachalam@cdac.in

Abstract— In this paper, we discuss the implementation details of SuMegha Scientific Cloud Lab Kit, which enables users to setup their own private scientific cloud. The scientific cloud offers infrastructure, platform, and/or software services for the scientific community working on modeling and simulation of problems in domains of bioinformatics, climate modeling, etc. It enables the creation of virtual clusters (High Performance Computing as a Service). The lab kit is very useful to the researchers and the student community, who want to set up a cloud, but lack the expertise to do so. It not only helps to effectively utilize the idle computers in the organization, but also provides MPI or Hadoop based clusters on demand to its users. Storage as a service is provided by the in-house developed CloudVault solution integrated into the lab kit. Supplemented with Software as a Service such as Seasonal Forecast Model (SFM), Next Generation Sequencing (NGS), etc. SuMegha Cloud Lab kit offers a comprehensive environment for scientific computing.

The cloud lab kit is designed in a modular fashion using open source components; Nimbus is used as the cloud middleware as it supports virtual cluster creation or contextualization, and Xen is the hypervisor. The SuMegha Portal interface enables users to request for virtual machines/clusters and easily submit jobs to cloud for execution. The SuMegha cloud lab kit is a useful invention to the cloud community which can setup a cloud on a single desktop or on a group of servers. The fact that the Cloud can be set up in a single desktop can greatly aid in adoption of Cloud Computing, since almost anyone can now establish a cloud in their premises. SuMegha will also promote the use of parallel paradigms like MPI and Hadoop to solve compute-data intensive problems. We believe that this lab kit is very useful to the vast number of academic institutions, to easily set up the required Cloud Lab.

Summary:

Abstract— In this paper, we discuss the implementation details of SuMegha Scientific Cloud Lab Kit, which enables users to setup their own private scientific cloud. The scientific cloud offers infrastructure, platform, and/or software services for the scientific community working on modeling and simulation of problems in domains of bioinformatics, climate modeling, etc. It enables the creation of virtual clusters (High Performance Computing as a Service). The lab kit is very useful to the researchers and the student community, who want to set up a cloud, but lack the expertise to do so. It not only helps to effectively utilize the idle computers in the organization, but also provides MPI or Hadoop based clusters on demand to its users. Storage as a service is provided by the in-house developed CloudVault

solution integrated into the lab kit. Supplemented with Software as a Service such as Seasonal Forecast Model (SFM), Next Generation Sequencing (NGS), etc. SuMegha Cloud Lab kit offers a comprehensive environment for scientific computing.

The cloud lab kit is designed in a modular fashion using open source components; Nimbus is used as the cloud middleware as it supports virtual cluster creation or contextualization, and Xen is the hypervisor. The SuMegha Portal interface enables users to request for virtual machines/clusters and easily submit jobs to cloud for execution. The SuMegha cloud lab kit is a useful invention to the cloud community which can setup a cloud on a single desktop or on a group of servers. The fact that the Cloud can be set up in a single desktop can greatly aid in adoption of Cloud Computing, since almost anyone can now establish a cloud in their premises. SuMegha will also promote the use of parallel paradigms like MPI and Hadoop to solve compute-data intensive problems. We believe that this lab kit is very useful to the vast number of academic institutions, to easily set up the required Cloud Lab.

Infrastructure Clouds and Virtualisation Session III / 23

Managing Virtual Appliance Lifecycle in IaaS and PaaS Clouds

Authors: Boris Parák¹; Michal Kimle¹; Zdeněk Šustr¹

¹ CESNET

Corresponding Author: sustr4@cesnet.cz

Both the IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) models of providing cloud services rely on virtual appliances. In popular terms, they are “images” of either bare operating systems, typically entailing popular Linux distributions, which can be further contextualized once users instantiate their own virtual resources, or operating systems with applications pre-installed for use in the given platform, which may often consist of a number of complimentary appliances. Such appliances must be offered to users of any cloud service – they are the basic units the users see and select from when they decide to procure resources in the cloud. Understandably, cloud service providers are often expected to offer a variety of appliances. Even in a simple IaaS scenario, users expect to see a range of OS distributions and flavours. With PaaS, the variety is even greater. Obviously a range of appliances can be obtained from cloud marketplaces, but that only offsets rather than solves the problem since the challenges of maintaining their appliances are the same for local cloud site administrators and marketplace maintainers alike. This, inevitably, means that cloud site or marketplace administrators must not only offer a selection of appliances, but also manage them throughout their life cycle, keep them secured and updated, and eventually discontinue them when the time comes. It is not only cumbersome but also inherently insecure to leave updates to the user instantiating the given appliance. On top of that, the ability to always offer “fresh” appliances to its users is a competitive advantage a cloud site may wish to exploit. This paper introduces a concept of automated periodic appliance updates in a federated cloud environment, alongside actual tools developed to perform that task. It also sums up up-to-date experience with operating such tools in the European Grid Initiative’s Federated Cloud.

Earth & Environmental Sciences & Biodiversity Session II / 24

A Ubiquitous Urgent Computing Framework for Ensembles of Flash Flood Forecasts

Author: Siew Hoon Leong¹

Co-author: Dieter Kranzlmüller²

¹ Leibniz Supercomputing Centre

² LMU Munich

Corresponding Authors: leong@lrz.de, kranzlmuller@ifi.lmu.de

Flash flood is a common weather abnormality that plagues many countries including Germany. It is arguably the most dangerous type of floods as it can form swiftly due to high or extremely high rainfall rates with little or no prior warning. According to the findings of a flood risk assessment study 1 carried out by the European Commission (EC) Joint Research Centre (JRC), the frequency and damage of floods in Europe are expected to increase sharply over the next decades. In Germany, it is expected that 170,000 to 323,700 of the population to be affected by the floods with a damage of 5.3 to 33.9 billion EUR in 2080.

Flash flood numerical forecasts can generate information and provide predictions to facilitate the process of making timely decisions for managing affected areas and reduce casualties. If the forecasts have to be computed within a (short) required timeframe, this class of computing is referred to as urgent computing 2. Urgent computing enables responsible authorities to make educated decisions by providing simulated predictions of disasters, the impact and required evacuation zones, etc., within a small lead time.

The need to access the framework from anyway and at anytime, leads to the construction and implementation of a task-based ubiquitous (TbU) approach 3. This approach facilitates access to the underlying distributed resource sets from ubiquitous end user device, which is particularly crucial in the chaotic environment that entails a disaster. The inherent unpredictability of disasters can render any best made plans to prepare resources in advance futile.

Additionally, due to the inherent uncertainties in most forecast models, stochastic methods based on an ensemble of forecasts are recommended. Enabling multiple forecasts with varying execution time to complete within a deadline, require a number of different resources. It is thus also essential to acquire the ability to swiftly organise a set of resources, while facing uncertainty in computation requirements and dynamism of computing environments on heterogeneous distributed resources. Consequently, an adaptable framework that can swiftly and effectively manage and allocate multiple resources for such computations is designed.

Thus, a ubiquitous urgent computing framework is designed to efficiently allocate ensembles of forecasts on a set of heterogeneous distributed resources robustly and reliably in an energy aware manner 4. E-Infrastructures 5 like PRACE and EGI, are good candidates to integrate this framework.

1. EM-DAT The International Disaster Database, Centre for Research on Epidemiology of Disasters – CRED, Belgium, viewed 5 November 2015, < <http://www.emdat.be/>>
2. S. H. Leong and D. Kranzlmüller. Towards a General Definition of Urgent Computing. In ICCS, volume 51 of *Procedia Computer Science*, pages 2337 – 2346. Elsevier, 2015.
3. S. H. Leong and D. Kranzlmüller. A Task-based Ubiquitous Approach to Urgent Computing for Disaster Management. In *ICT-DM*, 2015. To be Published.
4. S. H. Leong and D. Kranzlmüller, “Robust Reliable Energy-Aware Resource Allocation Heuristics on Distributed HPC Resources for an Urgent Computing Ensemble Forecast,” Submitted.
5. S. H. Leong, A. Frank, and D. Kranzlmüller. Leveraging e-Infrastructures for Urgent Computing. In ICCS, volume 18 of *Procedia Computer Science*, pages 2177–2186. Elsevier, 2013.

Summary:

A ubiquitous urgent computing framework for ensemble flash flood forecasts that can potentially leverage on distributed e-Infrastructures, e.g. PRACE and EGI, is shared.

e-Science Activities in Asia Pacific III / 25

EU-India cooperation on e-Infrastructures: from EU-IndiaGrid to e-INIT project

Author: Alberto Masoni¹

¹ *INFN National Institute of Nuclear Physics*

Corresponding Author: alberto.masoni@ca.infn.it

Since 2006 INFN set up a cooperation with India on Grid technology and e-Infrastructures, further developed along the years with the EU-IndiaGrid/EU-IndiaGrid2 and CHAIN/CHAIN-REDS projects. These projects, co-funded by the European Commission within the Research Infrastructures Workprogramme, supported the exploiting of e-Infrastructures across Europe and India for the benefit of a variety of scientific disciplines, including biology, earth science, material science and the Indian collaboration for the Large Hadron Collider (LHC).

Since the very beginning international cooperation of Indian leading actors with relevant European and Asia-Pacific Institutes represented one of the main asset for the development of the Indian e-Infrastructure and the benefit of the concerned scientific applications.

The e-INIT project: India-ITaly cooperation on e-Infrastructures support for High Energy Physics applications, is funded by Italian Ministry for Foreign Affairs and International Cooperation within the frame of the Executive Programme of Scientific and Technological Cooperation between Italy and India, where was selected as one of the six Significant Research Projects approved by the Programme. The project leading partners are INFN, the Italian National Institute of Nuclear Physics and the Office of Principal Scientific Adviser to Government of India (PSA). The project capitalized on the experience and the achievements of almost 10 years of projects in the area of e-Infrastructures addressing India and the Asia-Pacific region. These projects were led by INFN on the European side and premium research Government Institutions on the Indian side.

The e-INIT project worked for four years (2012-2015) in close synergy with the CHAIN and CHAIN-REDS project which set the activity within a worldwide context. Following the way paved by previous EU-IndiaGrid and EU-IndiaGrid2 projects it addressed the support of cooperation of major European and India e-Infrastructures to the advantage of several scientific domains with particular focus on the LHC experiments, where India and Italy collaborates in the context of the ALICE and CMS experiments.

A major step in Indian e-Infrastructures development was the set-up of the National Knowledge Network (NKN). The PSA office to Government of India conceived and proposed the National Knowledge Network plan, approved by the Indian Government in 2009 with a budget exceeding 1 billion euro for a period of over 10 years. In these years NKN developed as “the e-Infrastructure of India” connecting with a high speed network backbone for all knowledge related institutions in the country. At present over 1000 Institutions are connected. The e-INIT project actively supported the development of connectivity between NKN and GEANT, the Pan-European Research network, in particular within the context of the Trans-Eurasia Information Network Programme (TEIN3 and TEIN4).

In cooperation with the CHAIN and CHAIN-REDS projects e-INIT also supported the interoperability and interoperability between Indian and European grid/cloud infrastructures. Project partners as CDAC, spearhead of GARUDA India National Grid Initiative, played a key role, together with the strong cooperation of INFN with ALICE and CMS Indian Institutes, managing the Indian component of the Worldwide LHC Computing Grid Infrastructure.

Summary:

In a period of about 10 years India e-Infrastructures experienced a most impressive evolution. A policy plan to support LHC experiments and strategic scientific sectors, as well as the international cooperation of Indian leading Institutions with key players in Europe Asia-Pacific gave a fundamental contribution. The presentation describes the major developments of 10 years of India e-Infrastructure within the perspective of a series of S&T cooperation projects whose Indian partners where at the core of this developments.

VRE Session / 26

A Scientific Paper Reproducible Environment with Overlay Cloud Architecture

Author: Shigetoshi Yokoyama¹

¹ *National Institute of Informatics*

Corresponding Author: yoko@nii.ac.jp

The Inter-Cloud is a promising approach for the distributed application demands in some of HPC applications, like Next-Generation-Sequencing Data analytic. However, building the Inter-Cloud environments requires IT expert knowledge. This paper introduces an architecture called Overlay Cloud and Virtual Cloud Provider (VCP), which is a middle-ware to automatically build a set of virtual resources on the Inter-Cloud and ease the knowledge requirements. That aims to help to realize ubiquitous scientific paper reproducible environments.

In a data-centric science fields such as bioinformatics, demands for reproducibility to ensure experiments in papers is strong. For example, we share genomic analysis programs as open source software and have the public databases of the DNA sequences in the bioinformatics field. However, the following issues related to program execution environments still exist.

1. Data processing software is complex and diverse.
2. Massive data from many data sources, such as the next generation sequencers.
3. Amount of data analysis processing is increasing.

In my study, I utilize Overlay Cloud architecture in order to solve these problems. Overlay Cloud is an architecture that is Overlay a container environment over existing cloud environments (private / community / public) and overlay clouds do not depend on cloud environments in order to separate the interface between the application user and the cloud infrastructure. Therefore, the user can freely select the container execution platform. Paper readers can simply press a button which is linked to the paper in order to obtain the paper reproduction verification environment on their selected cloud.

As a prototype, I constructed a bioinformatics workflow reproduction environment on inter-cloud in Overlay Cloud architecture. This is aiming to show that the paper reproduction environment has following properties.

1. Data analysis software portability across the clouds by two level containerization.
2. Network delay reduction between the data and the data analysis program by container distribution.
3. Processing performance improvement of data analysis by distributed processing infrastructure to deploy containers.

In this paper, we report in particular a result of our efforts to ensure portability, which is the first problem.

Summary:

The Inter-Cloud is a promising approach for the distributed application demands in some of HPC applications, like Next-Generation-Sequencing Data analytics. However, building the Inter-Cloud environments requires IT expert knowledge. This paper presents an architecture called Overlay Cloud and Virtual Cloud Provider (VCP), which is a middle-ware to automatically build a set of virtual resources on the Inter-Cloud and ease the knowledge requirements. That aims to help to realize ubiquitous scientific paper reproducible environments.

Poster Session / 27

C++ and the delivery of portable program binaries in the Linux ecosystem

Authors: Ralf Mikulla¹; Ruediger Berlich²; Stefan Hacker¹; Stefan Hug¹

Co-authors: Ariel Garcia³; Sven Gabriel³

¹ *PTV Group*

² *Gemfony scientific UG (haftungsbeschraenkt)*

³ *Gemfony scientific*

Corresponding Author: r.berlich@gemfony.eu

Linux excels as the dominant platform particularly in High Performance Computing, as well as as a general server operating system in less HPC-centered environments. Both on the server- and the desktop side, a huge variety of programs is available, readily delivered in pre-compiled form through Linux distributions. Very rarely will typical users have to compile source code themselves. Organizations wishing to supply independent applications in binary format will however face many difficulties, particularly where C/C++ was used as the main programming language. Technical differences between Linux distributions, while being subtle, often prevent the „compile once, run anywhere“ philosophy being one of the hallmarks of the Windows programming environment. Particular difficulties exist with respect to dynamic libraries which, when delivered with the operating system rather than the application, represent a direct dependency on the underlying Linux OS. While static linking may help, some libraries (in particular C and C++ standard libraries) may not easily be linked statically into an application. Shipping dynamic libraries along with the binaries of applications or custom libraries, while being heavy-weight, may be an option, but will present users with restrictions regarding the libraries they may use in their own code. Issues exist in particular, where users need to load pre-compiled binary objects or access the API of a supplied library. They would then be forced to link their code with other standard libraries than those found on their home system, which is both error-prone and represents a huge restriction. The presentation discusses possible approaches for dealing with these issues, both in the context of Grid- and Cloud-Computing and for application development at the PTV group, Germany.

Humanities, Arts, and Social Sciences Session I / 28

ICT-Enhanced Interactive Writing Program for International Students - A Plagiarism-free Writing Program -

Author: Tomoki Furukawa¹

Co-author: Tosh Yamamoto²

¹ *Kansai University, International School*

² *Kansai University, CTL*

Corresponding Authors: tomokifurukawa@hotmail.com, soetosh@gmail.com

With the formulation of the “300,000 International Students Plan” in 2008, Japan has made efforts to increase the number of international students, and acceptance of these students is moving forward at Japan’s institutions of higher learning. However, with the acceptance of many international students, the acquisition of advanced Japanese-language skills, particularly improvement of academic writing abilities in Japanese, has become an urgent issue. Among these problems, one of the largest is knowledge and understanding of plagiarism. The widespread use of the Internet has made it extremely easy to plagiarize (Hinchliffe, 1998), and the low level of awareness of plagiarism among international students has made it a major issue in the context of writing academic papers and reports as well as raising awareness about it.

Thus, this paper provides a report on results of a class called “Academic Writing” for international students at Kansai University, and the use of TurnItIn® for preventing plagiarism. TurnItIn® is an online tool that compares sentences composed by students with a vast amount of information culled from sources, such as webpages and academic databases. TurnItIn® quickly checks for similarities, and matches submissions of students with the database. It then displays the level of similarity, enabling a quick confirmation of whether a student has used quotes appropriately or is plagiarizing. The results output by TurnItIn® can then be used in providing instruction to students. In addition, the system has features that enable teaching staff to make direct corrections to and evaluations of submitted materials online, allowing teachers to use it as an online educational tool for developing writing abilities.

As a result of using TurnItIn®, there has been an increase in the appropriate use of quotations. In addition, the results of questionnaires given after these classes have shown an improvement in awareness of plagiarism among international students, as compared with classes prior to using this system.

This paper presentation will share the features of TurnItIn® and its functionality for enhancing

the quality of the writing program and reports on the implementation of an academic writing curriculum with a particular focus on a shift in plagiarism awareness that occurred among international students.

Summary:

Topics of Interest:

1.Applications and results from the Virtual Research Communities and Industry

(4) Humanities, Arts, and Social Sciences (HASS) Applications

Key Words: Japanese as Second Language (JSL),College Writing, International Students, Plagiarism

Humanities, Arts, and Social Sciences Session I / 29

Educational Informatics: A Paradigm Shift for IT-enhanced Education

Author: Tosh Yamamoto¹

Co-author: Ti-Chuang Chiang²

¹ *Kansai University*

² *the Division of Medical Informatics, College of Medicine, NTU*

Corresponding Authors: soetosh@gmail.com, tcc@ntu.edu.tw

This paper proposes a new realm of research field for ICT-enhanced learning, called educational informatics. While arguing the pros and cons of the current issues surrounding the ICT-enhanced education, it is proposed that there should be a need for a research field to measure and visualize the learning effectiveness in the scientific way. In order to achieve such goal, there should be an immediate need for achieving consensus on SOP for educational informatics.

Key Words: Constructivism, Educational Paradigm Shift, Educational Informatics, ICT, SNS, SOP.

Summary:

This paper proposes a new realm of research field for ICT-enhanced learning, called educational informatics. While arguing the pros and cons of the current issues surrounding the ICT-enhanced education, it is proposed that there should be a need for a research field to measure and visualize the learning effectiveness in the scientific way. In order to achieve such goal, there should be an immediate need for achieving consensus on SOP for educational informatics.

Observing the current trend in the development of ICT surrounding education, there are fundamental concepts and the guideline for driving us into the future design for education. This paper will walk the reader through the process of thinking the educational trend critically and then propose an optimal Standard Operating Procedure (SOP) for the future design for education enhanced with ICT.

Networking, Security, Infrastructure & Operations Session IV / 30

Networking for WLCG: LHCOPN and LHCONE

Author: Edoardo Martelli¹

¹ *CERN*

Corresponding Author: edoardo.martelli@cern.ch

The Worldwide LHC Computing Grid (WLCG) is a global collaboration of almost 200 interconnected computing centres that provide global computing resources to store, distribute and analyse the massive volume of physics data generated by the Large Hadron Collider (LHC) experiments at CERN, Alice, ATLAS, CMS and LHCb.

The LHCOPN (LHC Optical Private Network) connects the Tier 0 and Tier 1 sites. It is reserved for LHC data transfers and analysis, has a highly resilient architecture and relies on dedicated long distance links.

The LHC Open Network Environment (LHCONE) is the network deployed to meet the requirements of the new computing model of the LHC experiments, which demands to transfer data among any pair of Tier1, Tier2 and Tier3 sites.

The LHCOPN and LHCONE both successfully supported the data transfer needs of the LHC community during Run 1 and have now evolved to serve the networking requirements of the new computing models for Run 2.

The presentation will explain how the two networks are designed and operate. The LHCONE will be described in more details because it's still open to all the Tier2 and Tier3 sites to connect. For this, the concept of Science DMZ (De Militarized Zone) and how it must be used to connect to LHCONE will be explained in details.

LHCONE consists of three main services: L3VPN, P2P, perfSONAR. L3VPN is the production service giving high throughput connectivity to more than 50 sites around the world. P2P is a prototype service still being designed which aims to provide on demand point to point dynamic circuits between any pairs of LHCONE sites. perfSONAR is the monitoring service used by the WLCG community, built using the perfSONAR tool suite.

Focus will be given to the current status and the key changes, notably the delivered and planned bandwidth increases, the ongoing work to better address the needs of the Asia-Pacific region, developments to improve redundancy and progress made for provisioning point-to-point links.

Summary:

LHCOPN and LHCONE are the network that contributed to the LHC data distribution. The presentation describes the architectures of the two networks and the advantages they brought

Networking, Security, Infrastructure & Operations Session III / 31

IPv6 deployment at CERN

Author: Edoardo Martelli¹

¹ CERN

Corresponding Author: edoardo.martelli@cern.ch

In 2013 CERN completed the deployment of IPv6 in its Campus and Datacentre network. The full project lasted 2 years and involved several engineers to address all the aspects of a production level deployment, notably the management of addresses, the design and then the automatic provisioning of services. The presentation will explain how IPv6 has been deployed at CERN, the main design principles, the issues encountered and how they were solved. Special focus will be given to the addressing plan designed; to the development done on CSDB (the CERN network database and its network interface) and CFMGR (the CERN Network Management software); to the DNS configuration and how the users can modify it; to the DHCPv6 service and why and how it has been preferred to SLAAC (State Less Address Auto Configuration); to security aspects, especially on the tools given to the users to modify the CERN main firewall; to all the challenges encountered during the deployment; to the lessons learnt. The IPv6 deployment has required a large investment but will pay off when IPv6 will finally take over.

Summary:

The talk describes how IPv6 has been implemented in the CERN network and management tools

High Throughput & Supercomputing Systems and their Integration Session / 32**The Impact of the UNIX Scheduler on IO Dominated Applications****Author:** Shaun de Witt¹**Co-authors:** Andrew Lahiff²; George Vasilakakos²; Robert Appleyard²¹ *Science and Technology Facilities Council*² *STFC*

The kernel scheduler of a LINUX system is responsible for reordering and optimising requests for access to storage, be it spinning disks or SSDs. Since disk seek and read times are one of the slowest part of computing operations, this scheduling is essential to maintain performance on any modern computing system. In this paper, we look at the performance of different schedulers under a range of conditions, including heavy I/O, single stream performance and multi stream sequential and random access.

Summary:

In this paper we look at the optimisation of IO Performance on a LINUX based RAIDed filesystem by tuning the kernel scheduler and associated parameters. We demonstrate that the default scheduler does not perform well on these systems when under heavy I/O load and recommend settings for RAID6 systems based on the configuration used at the RAL Tier 1.

Data Management Session I / 33**The EUDAT Service Suite and the Data Lifecycle****Author:** Shaun de Witt¹**Co-author:** Jens Jensen²¹ *Science and Technology Facilities Council*² *STFC*

EUDAT, along with EGI, PRACE and GEANT represent four major 'transverse' projects across Europe, providing services, infrastructure and resources to support projects and communities on a European and International Scale in collaboration with international partners such as the RDA. . In this talk we introduce the services provided by EUDAT, and show how they support projects in their data management and data lifecycle. In addition, we show how EUDAT is evolving both to lower the barrier to usage, and provide a more integrated toolset - thus evolving from a set of standalone services to a suite of tools through which data can easily move dependent on the use case. One of the keys to this, and any other project, is the need for a federated AAI and in this we detail the work that has been done on allowing users to use their own local credentials to access the service

Summary:

The EUDAT Project has developed a suite of tools to assist users and communities in their data management needs. In this talk, we explain the origins and purpose of the project, the services currently in production, planned new services and how these services will evolve in the future.

34

A Survey of Chinese Spallation Neutron Source Cloud Computing System

Author: Yakang Li¹

¹ *Institute of High Energy Physics of the Chinese Academy of Sciences*

Corresponding Author: liyk@ihep.ac.cn

The requirements of computing and storage for High Energy physics experiments are growing rapidly with the continually expansion of experiments, and the forthcoming completion of Chinese Spallation Neutron Source (CSNS) has also higher needs for computing and storage system. The new computing mode—cloud computing—can make IT resources configuration flexible and management centralized. Under this computing environment, the cost of hardware can be reduced and the resource utilization has also been a lot more improved. So the application of cloud computing in High Energy Physics Experiments will be showed in this paper.

From the research and practice aspects, firstly, the application status of cloud computing science in High Energy Physics Experiments are introduced in this paper. Secondly, the special requirements of CSNS, which is constructing by the insitute of High Energy Physics of Chinese Academy of Sciences, are discussed further. Thirdly, the design and practice of cloud computing platform based on OpenStack will be mainly demonstrated from the aspects of deployment architecture, optimization, cloud storage system, maintenance system and monitoring system. And finally, the present status of CSNS computing environment are summarized and some future prospects are put forward in the ending of this paper.

Networking, Security, Infrastructure & Operations Session III / 35

IPv6 Security

Author: David Kelsey¹

¹ *STFC-RAL*

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

IPv4 network addresses are running out and the deployment of IPv6 networking in many places is now well underway. Some large IT distributed infrastructures, such as the Worldwide Large Hadron Collider Computing Grid, are starting to deploy dual-stack IPv6/IPv4 services to support IPv6-only clients. The IPv6 protocols involve new challenges for operational IT security. We have spent many decades understanding and fixing security problems and concerns in the IPv4 world. Many IT support teams have only just started to consider IPv6 security! The lack of maturity of IPv6 implementations together with all the additional complexities, particularly in a dual-stack environment, brings many new challenges.

The HEPiX IPv6 Working Group is producing guidance on best practices in this area. This talk will consider some of the security concerns in an IPv6 world and present the HEPiX IPv6 working group guidance both for the system administrators who manage IT services on distributed infrastructures and also for their related security teams.

Networking, Security, Infrastructure & Operations I / 36

A Study of Certification Authority Integration Model in a PKI Trust Federation on Distributed Infrastructures for Academic Research

Author: Eisaku Sakane¹

Co-authors: Kento Aida¹; Takeshi Nishimura¹

¹ *National Institute of Informatics*

Corresponding Author: sakane@nii.ac.jp

Among certification authorities (CAs) in an academic PKI trust federation such as IGTF (Interoperable Global Trust Federation), most of academic organizations that operate CA install by themselves the CA equipment in their building. To keep CA trustworthy, it is necessary to maintain such CA equipment and to obtain the special operators. Consequently, the high cost of CA operation weighs heavily on the CA organization. For research institutes whose essential duties are not the CA operation, the burden on the high cost of CA operation is an earnest problem, and cost reduction with increasing the efficiency of the operation is an important issue.

In this paper, we discuss not further operational optimization of a single CA itself but cost reductions with an integration of more than one CA in a PKI federation. However, it is not easy to straightforwardly integrate CAs as follows: current CAs are turned into intermediate CAs, and a new root CA is built. This integration model would remain CA operation duties such as issuing, revoking, and identity vetting. Also, building the new root CA that covers different academic research communities would raise the total costs of CA operations. From the point of view of feasibility, the issues of CA integration are provided. As a solution to the issues, this paper takes notice of issuing and registration authorities constituting CA, and proposes the following integration model: it integrates the issuing duties, and each organization carries out the registration duties as before. In the proposed model, integrating the issuing duties means that one issuing authority (IA) takes over the duty of the other IA. Since each registration authority (RA) performs the registration duty as usual, most of procedures such as for application for certificate usage remain unchanged, so that it does not confuse the users.

Based on the idea of the proposed model, we discuss how to connect between the taken-over IA and the RA2 operated by the organization that closes its IA. Among possible connections, we examine not only direct connection between the taken-over IA and the RA2 but also connection putting the RA1 operated so far by the organization that operates the IA between them. Furthermore, we improve the certificate policy of the taken-over IA so that it is compatible with the policy of the RA2. We also discuss an applicability of existing CA profiles such as MICS (Member Integrated Credential Service) profile and its extension.

Networking, Security, Infrastructure & Operations I / 37

Raising Security and Trust in our Inter-Federated World

Author: Hannah Short¹

Co-author: Romain Wartel¹

¹ *CERN*

Corresponding Author: hannah.short@cern.ch

The expanding network of Higher Education and Research facilities through inter-federation, whilst extremely valuable for collaboration and online security at large, exposes an inviting new vector of attack for malicious actors. A single compromised account may provide an entry point to this global network of resources linking thousands of organisations. How can we coordinate a response spanning countries and continents? How can we build trust between organisations in our communities? What lessons can we learn from existing architectures, such as WLCG?

REFEDS (the Research and Education FEDERations group), in conjunction with the European Commission funded AARC Project (Authentication and Authorisation for Research and Collaboration), is spearheading the Security Incident Response Trust Framework for Federated Identity (Sirtfi) as

a method for mitigating the impact of security incidents in our federated world. This framework provides a list of statements which an organisation must self-assert to be deemed Sirtfi compliant, such as “[OS4] A user’s access rights can be suspended, modified or terminated in a timely manner”. We are reaching out to members of academic communities to provide support and pilot the initiative.

Organic global trust groups already provide a platform for informal alliances within academia, research and industry, however there is a need for heightened transparency, inclusivity and structure to facilitate this process. The lack of centralised governance within this space, in contrast to individual organisations or even national federations, calls for a standard procedure that can be adopted by all participants. What role will individuals play as this network grows in magnitude? In this talk we will focus on the requirements for this trust framework and its implications on trust and collaboration. Join us as we explore the practicalities of closing the loop on federated security and discuss real-world scenarios.

This talk falls under “Networking, Security, Infrastructure & Operations”

Summary:

How can we raise security and trust in our increasingly complex, inter-federated world?

Networking, Security, Infrastructure & Operations Session IV / 38

DE-KIT(GridKa) – 100G @ LHCONE

Author: Bruno Hoefft¹

¹ *Karlsruhe Institute of Technology*

The Steinbuch Centre for Computing (SCC) at Karlsruhe Institute of Technology (KIT) is running the German LHC Tier-1 site and therefore involved in design and development of LHCOPN and LHCONE from the very beginning. KIT had previously established for the LHCOPN, the vpn network connecting tier-1 sites to tier-0 (CERN), 10Gbps links to multiple tier-1 sites in Europe. These links connected DE-KIT(GridKa), not only to tier-0, but also directed tier-1 to tier-1 traffic and provided the backup facility in case of a tier-1 to tier-0 link failure. The move towards 100Gbit technology was already highlighted at a talk at Chep 2012 “Status and Trends in Networking at LHC Tier1 Facilities”. However, it was only affordable with the latest range of emerging 100G technology products. With the last development of LHCONE, the dedicated LHCOPN links could be merged as part of the current 100G LHCONE uplink of DE-KIT.

Beside discussion the current LHCONE deployment at DE-KIT the talk will contain a view of the historic involvement of KIT of the fairly early development of 100G technology environment, e.g.:

- The first 100GE wide area network testbed over a distance of approx. 450 km - initiated by DFN - which was deployed between the national research organizations KIT and FZ-Jülich in 2010.
- In 2013, KIT joined the Caltech SuperComputing 2013 (SC13) 100GE “show floor” initiative using the transatlantic ANA-100GE link to transfer LHC data from a storage at DE-KIT (GridKa) in Europe to hard disks at the show floor of SC13 in Denver (USA).

The talk will cover the connection to LHCONE, based on DE-KIT(GridKa) as an example of a tier-1 center and will further discuss different possibilities of LHCONE connections in a more abstract manner. The requirements and restrictions for a LHCONE connected site as described in the LHCONE-AUP will be debated as well.

Physics & Engineering Session / 39

LHCb experience during the LHC 2015 run

Author: Christophe HAEN¹

Co-authors: Marco CATTANEO ¹; Philippe CHARPENTIER ¹; Stefan ROISER ¹

¹ CERN

Corresponding Author: christophe.haen@cern.ch

LHCb is one of the four high energy physics experiments currently in operation at the Large Hadron Collider at CERN, Switzerland. After a successful first running period (Run1 from 2011 to 2012), the LHC just entered the second exploitation phase (Run2, 2015-2017).

The technical break between these two running periods, known as Long Shutdown 1 (LS1), was the opportunity for LHCb to adapt, among other area of development, its data acquisition and computing models.

The operational changes on the data acquisition aspect include a clear split of the High Level Trigger (HLT) software in two distinct entities, running in parallel and in an asynchronous mode on the filtering farm, allowing a higher output rate to the final offline storage for further physics processing. A very challenging and innovative system performing full calibration and reconstruction in real time has been put in place. Thanks to this system, a fraction of the output of the HLT can be used directly for physics, without any intermediate step: this output is named “Turbo stream”.

Many changes were operated on the offline computing side as well. Besides the use of more modern and/or more scalable tools for the pure data management aspect, the computing model itself and the processing workflow were revisited in order to cope with the increased load and amount of data. The new Turbo stream requires new operational management compared to the other “standard” streams. The clear separation between the different levels of Tier (0, 1 and 2) has been abandoned for a more flexible, dynamic and efficient “Mesh” processing model, in which any site can process data stored at any other site. Validation and probing procedures were established and automatized before the start of massive Monte Carlo Simulation.

This paper presents the changes that were operated, and gives some feedback on their usage during the running period of 2015.

Data Management Session II / 40

Macaroons in dCache: sharing data made easy

Authors: Gerd Behrmann¹; Patrick Fuhrmann²; Paul Millar³; Tigran Mkrtchyan³

¹ NeIC

² DESY/dCache.org

³ DESY

Corresponding Authors: patrick.fuhrmann@desy.de, paul.millar@desy.de

For over a decade, dCache.ORG has provided robust software that is used at more than 80 Universities and research institutes around the world, allowing these sites to provide reliable storage services for the WLCG experiments and many other scientific communities. The flexible architecture of dCache allows its component services to be deployed in a wide variety of configurations and platforms, from a single Raspberry Pi up to hundreds of nodes in multi-petabyte infrastructures.

One problem that storage services, like dCache, share with other computer-related services is how to allow a user to share data with people the system does not know without making that data public: delegated access. In dCache, a user can shared data with other users by specifying POSIX group-ownership and ACLs; however, it does not allow a user to share data with people who are not known to dCache. While some services support delegated access by first requiring unknown recipients to register themselves, users often find this awkward and unnecessary.

Providing true delegated access is that it facilitates building aggregate services: services that depend on dCache for storage. A web portal that provides an enriched view on the data stored in dCache (by including additional metadata) is an example of such an aggregate service. Without a means for delegating access, either the users of the aggregate service must be known to the dCache instance or the portal must proxy all data transfers.

We present macaroons as a mechanism to support sharing data with people dCache does not know. Macaroons are a new cryptographic authorisation token that allows safe delegation. We describe various scenarios in which delegated access is useful, how macaroons are going to be supported in dCache, and the timeline including this support in future versions of dCache.

Networking, Security, Infrastructure & Operations Session II / 41

INDIGO-DataCloud: enabling collaboration in an identity-rich world

Authors: Andrea Ceccanti¹; Bas Wegh²; Patrick Fuhrmann³; Paul Millar⁴

¹ *CNAF-INFN*

² *KIT*

³ *DESY/dCache.org*

⁴ *DESY*

Corresponding Authors: andrea.ceccanti@cnafl.infn.it, paul.millar@desy.de

INDIGO-DataCloud is an €11m project funded by the EU's Horizon 2020 programme that harness 23 collaborator institutes from 11 countries. Over a 30 month period, it will develop a data/computing platform targeting scientific communities, deployable on multiple hardware and provisioned over hybrid (private or public) e-infrastructures.

It is now commonplace for collaborations within scientific communities to span organisational boundaries, which introduces the possibility for users within a collaboration authenticating using different technologies. In WLCG, a single technology was adopted: X.509. However, due to its overhead, X.509 has seen little use by scientific communities outside of particle physics. Instead most communities use either SAML or OpenID Connect. While the former is more mature and widely available within scientific communities, the latter has the backing of industry. As a result, INDIGO-DataCloud must allow scientists to authenticate with different mechanisms, supporting at least X.509, SAML and OpenID Connect.

Once users are authenticated they can interact with many INDIGO-DataCloud services directly. However, some services cannot easily be modified to support direct use of the INDIGO-DataCloud login session. Instead, the agent (a user or an application running on behalf of the user) must obtain the credentials necessary for interacting with the service; for example, an Amazon-S3-like service may require a username and password.

We present the INDIGO-DataCloud AAI infrastructure and describe how users can authenticate with X.509, SAML and OpenID Connect, along with how group membership and identity harmonisation are solved. We also describe how delegation between different agents is achieved and how these agents can obtain additional credentials when necessary for interacting with a service.

42

Installation of Public Opinion Sensors: How a Community Based Web Survey Platform Facilitates the Collection of Qualitative Opinion Data

Authors: Alex Yahja¹; Frank Liu²

¹ *National Center for Supercomputing Applications*

² *National Sun Yat-Sen University*

Corresponding Author: frankesliu@gmail.com

The political polarization in democracies, such as the U.S. and Taiwan, has emerging as a critical challenge to the legitimacy of governance. Not only political elites are likely to become divided along their ideology or stance on salient policy issues, but the general public are also subject to today's partisan media and self-selected news sources and are expected to become polarized in their partisanship. Employing the agent-based modeling (ABM) approach, we construct an artificial society where political parties, elites, news media, social networks, and ordinary citizens are considered as an ecosystem of preference formation. We pay close attention to the role of "independent voters," who are less partisan in their partisanship and policy stances, and their willingness of voting participation in the environment of political polarization. The theory of spiral of silence and theories of network heterogeneity suggest that the independents are the group of people who are less likely to turn out to vote. Therefore, that the independents becoming partisan, or least becoming stronger in their partisan orientation, suggests that society-wide participation in voting will increase. Will this theory stand true in all possible situations in the lab? In what situations will the independents become partisan? Does the turnout of the independents suggest a change in the preference redistribution at the aggregate level?

As these questions cannot be directly and systematically observed or empirically tested in the field, the lab-experiment approach like ABM employers researchers to explore possible answers to these critical questions. We use Netlogo and John Zaller's (1992) Response-Acceptance-Sample theoretical framework for the construction the model of voter turnout, use grid computing for system calibration and model validation, and then use linear models for testing hypotheses based on simulation data. Implications of the findings for the development of democratic governance are discussed.

Summary:

This study employs the agent-based modeling approach (ABM), use Netlogo and employ John Zaller's (1992) Response-Acceptance-Sample theoretical framework to construction the model of voter turnout, by which we examine the extent to which a political polarized environment influences the independent voters' political participation in voting.

Poster Session / 44

Options for the evolution of the LHCb Computing Model for LHC run 3

Author: Christophe HAEN¹

Co-authors: Marco CATTANEO¹; Philippe CHARPENTIER¹; Stefan ROISER¹

¹ CERN

Corresponding Author: christophe.haen@cern.ch

LHCb is one of the four high energy physics experiments currently in operation at the Large Hadron Collider at CERN, Switzerland. During the second long technical break of the LHC (LS2) to take place from 2018 to 2020, LHCb will undergo major upgrades. These upgrades concern not only the actual detector, but also the computing model driving the physics analysis.

The main incentive and driving constraint for the new computing model or Run 3 will be the increased amount of data recorded by the experiment. The current estimations lead to a data taking rate of 100 KHz, which is an order of magnitude higher than the Run 2 numbers. Such rate forces us to review the bare management of the data in terms of files and bookkeeping, as well as the way we are processing it on the grid in order to extract relevant physics figures.

Because of storage space restrictions, having a fixed and large number of replicas for each file is not a solution anymore. In particular, only files interesting for current analysis need fast access, while others can suffer reasonable delays. The data popularity system introduced for Run2 can help reaching such a dynamic data placement strategy. Storage technologies and transfer protocols will have to evolve and follow the latest trends encouraged by the various grid sites. It includes CEPH

based storages, S3 or HTTP. Processing model relying on stripping/skimming are not sustainable anymore, and the use of new concepts such as the LHCb TURBO stream must be extended. Opportunistic computing such as clouds and volunteer-computing are also part of the resources that should be accounted for and efficiently used.

This paper presents the challenges the LHCb computing model will have to face, as well are some solutions considered to address them.

High Throughput & Supercomputing Systems and their Integration Session / 45

High-Throughput Processing of Space Debris Data

Author: Andreas Schreiber¹

¹ *German Aerospace Center*

Corresponding Author: andreas.schreiber@dlr.de

Space Debris are defunct objects in space, including old space vehicles (such as satellites or rocket stages) or fragments from collisions. Space debris can cause great damage to functional space ships and satellites. Thus detection of space debris and prediction of their orbital paths are essential for today's operation of space missions. To detect the space debris, sensor networks of optical telescopes, radars, and laser systems are distributed at various places over the earth. The goal is a catalogue containing as many of the small debris items as possible. Based on the catalogue, a mission support system is built that gives real-time orbit information, collision detection, and re-entry prediction for each debris item.

The talk shows the software infrastructures BACARDI (Backbone Catalogue of Relational Debris Information) and Skynet (Network for surveillance of the sky). BACARDI is a high-level, domain-specific system for gathering space debris data from the various data sources, such as sensor networks or existing databases, for storing the data in databases, and for performing the data processing, such as object identification or orbital collision detection. BACARDI sits on top of Skynet. Skynet is a distributed computing infrastructure for high-throughput data processing. The architecture of Skynet was designed as a scalable, self-organizing network of nodes. Nodes communicate via decentralized message queues with minimal network overhead. The implementation is based on standard technologies. For the network infrastructure and for messaging, Skynet uses ZeroMQ. Computations on each node are performed by highly optimized and well-evaluated codes implemented in Fortran. The serialization of data that goes through the distributed network is done using Google Protocol Buffers, which guarantees a minimal protocol overhead. Skynet records Provenance information for traceability and provability of all processed data. This allows backtracking of each produced data product (e.g., ephemerides, state vectors, correlated objects, ...) and guarantees reproducibility of all generated data. Provenance is recorded of all activities during runtime and stored in a graph database.

Summary:

Space Debris are defunct objects in space, including old space vehicles or fragments from collisions. Space debris can cause great damage to functional space ships and satellites. Thus detection of space debris and prediction of their orbital paths are essential. The talk shows a Python based infrastructure for storing space debris data from sensors and high-throughput processing of that data.

Physics & Engineering Session / 46

The Cluster Monitoring System of IHEP

Author: Qingbao Hu¹

Co-authors: Xiaowei Jiang¹; jingyan Shi¹

¹ *IHEP*

Corresponding Author: huqb@ihep.ac.cn

With the rapid increase of the high-energy physics experimental requirements, the IHEP cluster scale is in rapid growth. More services running at the different devices, and more software and hardware status need to be monitored in real-time. A fine grained Monitor system can guarantee a device runs well, and solve the error happened to it. A Monitoring system ensures the stability of the whole platform.

In IHEP monitor system, Ganglia is used to record the status of the cluster machines, such as CPU load averages and network utilization; Nagios detects the service status, and sends the alarm actively based on NRPE remote plugin. A real-time logger-analyze the monitoring tool we developed, collects services log and machines log and provides an overview of the whole cluster health status. The aim of this tool is to give a summary of the cluster stability in real time.

Infrastructure Clouds and Virtualisation Session I / 47

Context-aware cloud computing for HEP applications

Author: Randall Sobie¹

¹ *University of Victoria*

Corresponding Author: rsobie@uvic.ca

Context-aware computing is a topical area for the multimedia or content-delivery industry. In this model, one uses situational and environmental information to anticipate needs and proactively offer situation-aware content, functions and experiences. We believe scientific computing can significantly benefit from a context-aware design. Currently we operate a distributed cloud computing system that is using thousands of cores on private and commercial clouds for HEP applications as well as applications in astronomy. We will describe this system, which has been running successfully for a number of years, highlighting the reliability and scalability, but also noting the challenges. To address these challenges, we have realized that we need systems with the ability to self-configure, retrieve software, locate data, self-diagnose faults and automatically recover from well-known errors situations. We are adding intelligence to the VM instances so that they can dynamically configure themselves and locate repositories. We will describe some of the steps we have made toward a context-aware cloud system and described the new features and services that will be deployed over the next year. The new system will make more efficient use of our existing resources and enable us to run all applications (e.g. compute, data and memory-intensive) on both private and opportunistic clouds. We also recognize the importance of developing software and system for a broader community. We have used and will continue to use and develop open-source packages with common standard protocols so that our context-aware cloud computing system can be used by any research discipline.

Earth & Environmental Sciences & Biodiversity Session I / 49

Data Processing and Visualization for Global High-Resolution Climate Simulations

Author: Chiaying Tu¹

Co-authors: Ching-Yao Lin²; Huang-Hsiung Hsu¹; Whey-Fone Tsai²

¹ *RCEC, Academia Sinica*

² *NCHC*

Corresponding Author: cytu@gate.sinica.edu.tw

Global high-resolution climate models have demonstrated the added value of enhanced resolution. They showed significant improvement in the simulation of large-scale circulation. In addition, the increased resolution enables more realistic simulation of small-scale phenomena. The improved simulations of climate also result in better representation of extreme events. Nevertheless, computing demands are rapidly increased for enhanced-resolution simulations. Other than computing resources, the storage and distribution of the high-resolution model data is another challenging issue. In this study we try to develop a big data experimental platform for environmental monitoring and simulation. The platform links computing environment for global climate simulation and big data analysis system that improves data management and processing power. Master and slave data nodes are designed and coupled with distributed database. The data node is not only for data storage, but also for in-place computing that enables processing further data operation and analysis effectively. The platform has applied to the global high-resolution climate simulations and demonstrates temporal and spatial visual analysis of the huge amount data.

Networking, Security, Infrastructure & Operations Session V / 50

KEK Central Computer System

Author: Go Iwai¹

¹ KEK

Corresponding Author: go.iwai@kek.jp

High Energy Accelerator Research Organization (KEK) plays a key role in particle physics experiments, as well as supporting the communities in Japanese universities. In order to ensure those important missions, KEK has two large-scale computer systems: the Supercomputer System (KEKSC) and the Central Computer System (KEKCC).

The KEKSC is mainly used by collaborative researches in theoretical elementary particle and nuclear physics, condensed matter physics, as well as for accelerator simulations. The system is composed of two different systems: Hitachi SR16000 model M1 (System A) and IBM Blue Gene/Q (System B).

The KEKCC caters to the research demands of particle physics, nuclear physics, the photon factory, neutron science, accelerator development, theory computation, etc. In addition, this system provides information infrastructure environment such as Web, e-mail, and Grid (EMI/iRODS) services and supports the research activities and collaborations of KEK.

As mentioned above, the EMI Grid middleware is deployed in the KEKCC for analysing and sharing experimental data over the distributed systems. The system is operated under the Worldwide LHC Computing Grid (WLCG) project. The Belle II, T2K, and ILC experiments do their data analysis using the Grid infrastructure to manage large amount of experimental data.

We would like to share our experiences and challenges in the security, the operation, and experiment-specific applications, as well as requirements for storage resources and computing resources particularly focusing on Grid through nearly 4 years operation of the current KEKCC. Also we discuss the outlook for the next KEKCC system, which will be newly introduced in September 2016.

Infrastructure Clouds and Virtualisation Session III / 51

Mesos in a WLCG Tier 1 Grid Site

Author: Andrew Lahiff¹

¹ STFC

Corresponding Author: andrew.lahiff@stfc.ac.uk

Container orchestration is rapidly emerging as a means of gaining many potential benefits compared to a traditional static infrastructure, such as increased resource utilisation through multi-tenancy, the ability to handle changing loads due to elasticity, and improved availability as a result of self-healing. Whilst many large organisations are using this technology, in some cases for many years, it is not yet common in the scientific community.

At the RAL Tier-1 we have been investigating migration of services to an Apache Mesos cluster running on bare metal. In this architecture the whole concept of individual machines is abstracted away and services are run on the cluster in ephemeral Docker containers. Instead of the standard approach of manually placing long-running services on specific hosts, services are managed by a scheduler. This means that any host or application failures, as well as procedures such as rolling starts or upgrades, can be handled automatically and no longer require human intervention. Similarly, the number of instances of applications can be scaled automatically in response to changes in load. Even though there are these clear benefits, a number of new challenges arise, such as how monitoring, logging and in particular service discovery are dealt with in such a dynamic environment where services are no longer tied to specific hosts. In addition, an important question is whether it is even possible to run traditional grid middleware in this type of environment.

This talk will describe the Mesos infrastructure which has been deployed at RAL, the testing we have done, our progress towards migrating production services and discuss our future plans.

Networking, Security, Infrastructure & Operations Session IV / 52

The Networking of the IHEP Data Center

Author: Mengyao Qi¹

¹ IHEP

Corresponding Author: qmy@ihep.ac.cn

With constantly increasing volume of data from years of institutional research programs and sharply increasing use of server storage, the Data Center of Institute of High Energy Physics is facing heavy pressure of space layout, system wiring, and power consumption and thus needs further improvements and network architecture expansion. As the artery of Data Center business, the basic network architecture is in particular facing constantly and strictly challenges.

This paper elaborated on the network current condition of Data Center of IHEP, including the operation aspect of the wan network, campus network and Data Center network and related topological structure. Moreover, it showed new network update plans and wiring plans, focusing on solving the Data Center's current issues. These plans divided the Functional Network Area into five major regions, namely Management Network, Local Computing Environment, UI-DMZ1, Deposit -DMZ2 and Virtualization and helped to complete the transfer of servers and computing resources of research groups like juno and dyb. The serial test of the Data Center's whole network after these plans' implements proves that the performance and stability of the network is much better than the previous state and equipment management is getting smoother. More importantly, this paper explained how the real-time network monitoring of personar make contribution to overseeing and solving operation problems immediately and how independently developed public application service, such as IHEP unified authentication, IHEP IT Services Desk, Vidyo and IHEPBox, bring benefits to scientific research users.

VRE Session / 53

Improvement of Scalability in Sharing Visualization Contents for Heterogeneous Display Environments

Authors: Arata Endo¹; Haruo Takemura¹; Kiyoshi Kiyokawa¹; Shinji Shimojo¹; Susumu Date¹; Yasuhiro Watashiba²; Yoshiyuki Kido¹

¹ *Osaka University*

² *Nara Institute of Science and Technology*

Corresponding Author: endou.arata@ais.cmc.osaka-u.ac.jp

e-Science receives a lot of attention as the infrastructure that supports collaborative, computationally- or data-intensive researches through networks.

It is important in e-Science to build an infrastructure that streamlines the discussion between researchers that are located in different sites.

For efficient remote discussion, it is needed to visualize numerical data and to share visualized data because it is difficult to understand and analyze numerical data intuitively and there are many opportunities for researchers to change their research institute and university.

In recent researches that deal with large scale data, the visualization content that is the result of the visualization of the data is usually large and high-resolution image and video.

Tiled Display Wall (TDW) can display visualization contents on a large display in high-resolution. TDW is a technique to build a large and high-resolution display by arranging multiple displays in matrix.

TDW also facilitates people to see the visualization contents simultaneously and to share their insights easily.

These features of TDW promote the efficient discussion for the visualized results of experiments and simulations.

For realizing the e-Science infrastructure, sharing visualization contents between multiple TDWs in remote sites is required.

This requirement needs visualization contents to be distributed in different resolution because each TDW system has a unique configuration about display environment.

Visualcasting is a technique to distribute visualization contents to multiple TDWs that are heterogeneous display environments.

In Visualcasting, one or more relay servers are deployed on WAN and duplicate the images that is sent from applications and distribute them to TDW systems.

Visualcasting doesn't have the high scalability for the number of TDWs because the number of relay servers is static from start-up.

When much of TDWs share visualization contents, the lack of bandwidth is caused since each relay server sent received packets of a visualization content to all TDWs that share the visualization content.

We propose the dynamic rearrangement mechanism for the problem of Visualcasting.

In the mechanism, nodes that are the relay servers are launched dynamically with the increase of TDWs and are connected to existing nodes on tree structure.

This mechanism solves the problem of the scalability for the number of TDWs in Visualcasting.

For implementing the mechanism, we used Scalable Adaptive Graphics Environment (SAGE).

SAGE is a middleware to realize TDW and has a component called to SAGE Bridge that is developed as Visualcasting.

We implemented the dynamic rearrangement mechanism by remodeling SAGE Bridge.

In the evaluation, it is shown that our proposed method improves the scalability for the number of TDWs.

Infrastructure Clouds and Virtualisation Session II / 54

Elastic CNAF DataCenter extension via opportunistic resources

Authors: Stefano Dal Pra¹; Vincenzo Ciaschini¹; tommaso boccali²

¹ *INFN-CNAF*

² *INFN*

Corresponding Authors: tommaso.boccali@cern.ch, vincenzo.ciaschini@cnaf.infn.it, stefano.dalpra@cnaf.infn.it

CNAF/ Bologna, the biggest WLCG Computing Center in Italy, serves all WLCG Experiments plus more than 20 non WLCG Virtual Organizations, and currently deploys more than 180 kHS06 of Computing Power and more than 20 PB of Disk and 40 PB of tape via a GPFS SAN.

The Center has started a program to evaluate the possibility to extend its resources on external entities, either commercial or opportunistic or simply remote, in order to be prepared for future upgrades or temporary burst in the activity from experiments. The approach followed is meant to be completely transparent to users, with additional external resources directly added to the CNAF LSF batch system; several variants are possible, like the use of VPN tunnels in order to establish LSF communications between hosts, a multi-master LSF approach, or in the longer term the use of HTCondor. Concerning the storage, the simplest approach is to use Xrootd fallback to CNAF storage, unfortunately viable only for some experiments; a more transparent approach involves the use of GPFS/AFM module in order to cache files directly on the remote facilities. In this presentation we focus on the technical aspects of the integration, and assess the difficulties using different remote virtualisation technologies, as made available at different sites. A set of benchmark is provided in order to allow for an evaluation of the solution for CPU and Data intensive workflows.

Summary:

CNAF/ Bologna, the biggest WLCG Computing Center in Italy, serves all WLCG Experiments plus more than other 20 non WLCG Virtual Organizations, and currently deploys more than 180 kHS06 of Computing Power and more than 20 PB of Disk and 40 PB of tape via a GPFS SAN.

The Center has started a program to evaluate the possibility to extend its resources on external entities, either commercial or opportunistic or simply remote, in order to be prepared for future upgrades or temporary burst in the activity from experiments. The approach followed is meant to be completely transparent to users, with additional external resources directly added to the CNAF LSF batch system; several variants are possible, like the use of VPN tunnels in order to establish LSF communications between hosts, a multi-master LSF approach, or in the longer term the use of HTCondor. Concerning the storage, the simplest approach is to use Xrootd fallback to CNAF storage, unfortunately viable only for some experiments; a more transparent approach involves the use of GPFS/AFM module in order to cache files directly on the remote facilities. In this presentation we focus on the technical aspects of the integration, and assess the difficulties using different remote virtualisation technologies, as made available at different sites. A set of benchmark is provided in order to allow for an evaluation of the solution for CPU and Data intensive workflows.

Earth & Environmental Sciences & Biodiversity Session I / 55

Seasonal Ensemble Forecasting Application On SuMegha Scientific Cloud Infrastructure

Author: Ramesh Naidu LAVETI¹

Co-author: Prahlada Rao B B¹

¹ *C-DAC*

Corresponding Author: rameshl@cdac.in

Despite several advances in understanding the behavior of monsoon variability, innovations in the numerical modeling and the availability of higher computational capabilities, accurate prediction of Indian summer monsoon still remains a serious challenge. Seasonal Forecast Model (SFM), developed for seasonal forecast and climate research, is used for forecasting the Indian summer monsoon in advance of a season. Ensemble forecasting method helps us in finding and minimizing the uncertainty inherent in seasonal forecast. The inherent parallel nature and the bursty computational

demands of the ensemble forecasting method allows it to effectively utilize the Infrastructure-as-a-Service (IaaS) model on the cloud platform. However, realizing huge scientific experiments is still a challenge to the cloud service providers as well as to the climate modeling community.

To start with prototype experiments using SFM model were conducted at T-62 resolution (~ 200 km x 200 km grid). The experience gained from the prototype runs were used by the SuMegha operational community to fine tune the configuration of SuMegha Cloud resources to improve the quality of service. High resolution SFM at T-320 (~ 37 km x 37 km grid) was also configured and experiments were conducted to understand the scalability, computational performance of the application and the reliability of SuMegha Cloud.

In this paper, we use SFM as a case study to present the key problems found by climate applications, and propose a framework to run the same on SuMegha Cloud infrastructure to allow a climate model to take advantage of these cloud resources in a seamless and reliable way. The framework uses classification and outlier detection techniques to classify the resources and also to identify the faulty resources. It addresses the challenges such as unexpected hardware failures, power outages, failed porting and software bugs. We share our experience in conducting the ensemble forecasting experiments on SuMegha Cloud using the proposed framework. We also attempt to provide a perspective on the desirable features of a scientific cloud infrastructure, for easier adaptation of the same by the climate modeling community to conduct large scientific experiments.

Summary:

A framework for Seasonal Ensemble Forecasting Application On SuMegha Scientific Cloud Infrastructure is implemented to address the various challenges faced by climate modeling community while migrating to cloud infrastructure. We also presented our experience in conducting large ensemble forecast experiments on SuMegha scientific cloud infrastructure in a seamless and reliable way.

Poster Session / 56

Review and evaluation of the outcome of Pre-University Program

Author: Tomohiko SASAKI¹

Co-authors: Sadao HAYASHIDA²; Tosh Yamamoto¹

¹ *Kansai University*

² *Kobe City College of Technology*

Corresponding Author: ssk.tom@gmail.com

Kansai University had implemented pre-university program that had been provided through outsourcing for years. And since last academic year, as the plan changed, our team has developed the online program based on MOOC concept internally, which includes English for communication, Mathematics and Language arts. As regards managing Language arts program, we developed an effective educational model for academic writing in cooperation with iParadigms LLC. The aim of this presentation is to report the outcome of the Language arts program by outlining and analyzing the emerged problems and the results of questionnaire carried out for students after the course so that it makes an opportunity to exchange views with the audience about a more effectual model to cultivate learner's competence of writing, reading, thinking and communicating.

To begin with, this poster will provide an overview of Kansai University's pre-university program briefly before moving on to the main task. Then, the problems revealed during the course and the results of the questionnaire will be introduced. On the one hand, there certainly are many achievements in this program, on the other hand, not a few challenges we have to cope with in both curriculum and system are clarified. The poster will describe them, and especially it is the issue of student's attitude toward learning that became clear and made us recognize as a serious problem. Plagiarism, in particular, is the one of the most momentous matters to tackle: some of the student's assignments appeared that they had appropriated articles without indicating sources, and what was worse, they had showed to each other, which took staffs a lot of time to check them out. Besides solving the

problem of these reluctant and passive students, conducting follow-up survey for those who took this program is required in order to evaluate it more closely. Such topics will be examined through this presentation, and it is expected to show a clue of what model of academic writing, art of reading and logical thinking is effective and successful.

Summary:

The aim of this presentation is to exchange views with the audience about a more effectual model to cultivate learner's competence of writing, reading, thinking and communicating. Kansai University has developed pre-university program based on MOOC concept internally, which includes English for communication, Mathematics and Language arts. As regards managing Language arts program, we developed an effective online educational model for academic writing in cooperation with iParadigms LLC. This presentation will report the outcome of the Language arts program by outlining and analyzing the emerged problems and the results of questionnaire carried out for students after the course.

To begin with, this poster will provide an overview of Kansai University's pre-university program briefly before moving on to the main task. Then, the problems revealed during the course, such as plagiarism, and the results of the questionnaire will be introduced, which includes both achievements in this program and challenges we have to cope with in curriculum and system.

Reviewing and evaluating the program through such topics, it is expected to show a clue of what model of academic writing, art of reading and logical thinking is effective and successful.

Data Management Session I / 57**DIRAC Data Management Framework**

Author: Andrei Tsaregorodtsev¹

¹ *CPPM-IN2P3-CNRS*

Corresponding Author: atsareg@in2p3.fr

DIRAC Project is developing software for building distributed computing systems for the needs of research communities. It provides a complete solution covering both Workload Management and Data Management tasks of accessing computing and storage resources. The Data Management subsystem of DIRAC includes all the necessary components to organize data in distributed storage systems. It has a versatile File Catalog (DFC) service to keep track of data file physical replicas. This service is a central component to build a logical File System of DIRAC presenting all the distributed storage elements as a single entity for the users with transparent access to the data. The DFC service provides also a Metadata Catalog functionality to classify data with user defined tags. This can be used for an efficient search of the data necessary for a particular analysis. The Data Management system provides also support for usual data management tasks of uploading/downloading, replication, removal files with a special emphasis on the bulk data operations involving large numbers of files. Automation of data operations driven by new data registrations is also possible. In this contribution we will make an overview of the DIRAC Data Management System and will give examples of its usage by several research communities.

Earth & Environmental Sciences & Biodiversity Session I / 58**Advanced visualisation for environmental computing**

Author: Matti Heikkurinen¹

Co-authors: Christoph Anthes²; Dieter Kranzlmuller³

¹ *LMU*

² *Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities*

³ *LMU Munich*

Corresponding Author: heikku@nm.ifi.lmu.de

Environmental computing – supporting production of actionable knowledge from different environmental data sources and models – tends to produce results that describe the combined effects of several contributing phenomena. Efficient analysis of such amalgams, for example to understand how the event would be developing or to identify relative importance of different factors, requires different visualisation tools and techniques. Comparing different scenarios (and the effect of human responses to them) and developing action plans and policy recommendations requires seamless interdisciplinary (and often inter-sectoral) collaboration, which cannot be efficiently supported by the discipline-specific data sets alone.

In this paper we will present different visualisation approaches and their relative strengths and weaknesses in environmental computing approaches through using on-going initiatives as case studies. At the most straightforward level, simple graphs can already present information in a way that supports decision-making. Perhaps the best example of the power of such a simple visualisation is the impact of the “hockey stick” chart describing the global warming scenarios. However, the controversy surrounding the hockey stick graph also illustrates its limitations: finding the connection between data, models and the output requires some effort and human interpretation. This disconnect between the data and tools made it quite easy to misinterpret and misrepresent the data – accidentally or intentionally. Simply adding more dimensions (2D maps or videos) or resolution cannot overcome this problem on its own. In addition to presenting the results in an intuitive way, visualisation solutions for environmental computing need to cover equally efficiently the data sources as well as the models and workflows used.

As an approach to address this challenge, we present a framework model that supports efficient and intuitive visualisation of all the aspects of the environmental computing, ranging from the characteristics of the individual models and data sources to 3D presentations of the results. The framework components and their characteristics are discussed, ranging from model metadata frameworks to advanced virtual reality (VR) technologies used in environmental computing applications.

Since environmental computing data often consists of simulated or reassured spatial structures describing natural phenomena, VR tends to be an ideal tool visualising these structures. The observer is able to understand the spatial relations in a better way with the help of stereoscopic display and the intuitive change of geometrical perspective through head tracking technology.

High Throughput & Supercomputing Systems and their Integration Session / 59

Automatic dynamic stack management in large scientific applications: A case study using a global spectral model

Author: Ramesh Naidu Laveti¹

¹ *C-DAC*

Corresponding Author: rameshl@cdac.in

Compute and data intensive scientific applications demand compilers to allocate more temporaries on the stack. For example, the change resolution component of a global spectral model changes the resolution of the input files using Nearest Neighbor Interpolation which requires large temporaries on stack. Temporaries include sub-arrays, automatic arrays and, sub-sections corresponding to actual arguments of a subroutine. If the infrastructure cannot provide adequate stack space at runtime relative to the total size of the temporaries, then the application program runs out of stack and aborts. Using Heap memory allocation as a solution to this introduced around 5% additional runtime to allocate and free the memory. This is observed in various components of a global spectral model.

We propose an automatic dynamic stack management framework which uses application profile information and the information of required stack memory at compile time. It does not mandate any hardware configuration changes. This technique manages stack frames on RAM by compiler-inserted code into the application binary. Our experiments with a global spectral model show that the average runtime savings of 21% along with a compile time overhead of 4%. The actual gain depends on the size of the temporaries in an application and the size of the RAM. Currently, it supports sequential and OpenMP applications, we further enhance our framework to deal with the complex MPI and GPU programming paradigms.

Networking, Security, Infrastructure & Operations Session II / 60

Computer Security Landscape

Author: Romain Wartel¹

¹ CERN

Corresponding Author: romain.wartel@cern.ch

This presentation offers a short overview of the current security landscape, including threats, tool, techniques and procedures followed by attackers. The academic community is at a crucial time and needs to proactively manage the resulting risks, by collaborating further internally, as well as with other research collaborations, but also with the private sector and law enforcement.

VRE Session / 61

A blueprint for Environmental Computing Applications

Authors: Dieter Kranzlmuller¹; Matti Heikkurinen²; Nils gentschen Felde³

¹ LMU Munich

² LMU

³ LMU München

Corresponding Authors: kranzlmuller@ifi.lmu.de, felde@nm.ifi.lmu.de

Environmental Computing can be defined as the collaborative and multi-disciplinary approach to using computational sciences and technologies to support integration of environmental models capturing different aspects of the phenomena being studied. The analysis of the modelling results is often revealing previously unknown dependencies between the models, which is one of the ways it can provide a solid basis for further research and decision making. As such, environmental computing is the bleeding edge research area that bridges the gap between computer sciences and a multitude of application domains. The key technical challenge is to integrate heterogeneous models into flexible, cooperating wholes and providing application scientists with an easy access to (and use of) infrastructure components.

This paper focuses on the development of an e-Science infrastructure to provide end-to-end services (models, data, and both an easy to use workflow manager and graphical user interface) for environmental computing by exploiting HPC, Grid and Cloud resources. Best practices in adopting environmental computing applications to the newly formed e-Science infrastructure will be presented, as well as a blueprint of a flexible, extensible, and interoperable ICT infrastructure that supports composition of heterogeneous environmental computing workflows, while hiding low level complexity at the same time. This particular work will focus on the underlying ICT architecture and its core components, for example a central data store and (binary) repository, that will be described in more detail. Additionally, the generic approaches that allow running legacy codes on resources of different types (i.e. varying CPU architectures or different types of services, such as Clouds, Grids,

HPC clusters, etc.) presented and illustrated by concrete case studies based on the experiences and lessons learned related to the DRIHM project.

62

Evolving the RAL Tier 1 - Beyond High Energy Physics

Author: Ian Collier¹

¹ *STFC-RAL*

Corresponding Author: ian.collier@stfc.ac.uk

In the course of GridPP 5 - the next four year project funding computing for high energy physics in the UK - the RAL Tier 1 plans to reduce the effort required to provide the Tier 1 service. We are investing now in deploying a new grid data storage platform based on Ceph, and in the deployment of a cloud platform with the Tier 1 as one tenant backed up by our existing configuration management, monitoring and on call systems which will result in a more scalable and low maintenance infrastructure. This will allow us to more easily meet the substantial computing needs of new communities inside and outside STFC. We discuss both the technical roadmap and the more social problem of enabling new user communities on our infrastructure.

Summary:

A discussion of plans at the RAL Tier 1 Centre of both the technical development and outreach activities involved in meeting the computing needs of new projects and scientific communities beyond HEP.

Networking, Security, Infrastructure & Operations Session V / 63

The Dutch National e-Infrastructure

Author: Jeffrey Templon¹

Co-author: Jan Bot²

¹ *Nikhef*

² *SURFsara*

Corresponding Author: templon@nikhef.nl

The Dutch National e-Infrastructure (DNI) was reorganised in 2013. The SURF foundation was charged to build upon and in some cases replace the e-Infrastructure built during the 5-year “BiG Grid” project. This reorganisation turned the DNI into a sustainable resource supported by earmarked funds from the Dutch government.

Since then, there has been much progress and activity towards increased utility, ubiquity, and ease-of-use of the DNI. Significant effort has also been expended in the area of user support and outreach, national alignment between several data repositories, and innovation of the infrastructure. eScience research and engineering, best practices and policies regarding data stewardship and software sustainability have been addressed by the Netherlands eScience Center which is closely associated with the DNI. The talk will cover all these topics in detail, as well as outline the current challenges faced in improving the reach and quality of the infrastructure.

Summary:

Explanation of the scope, governance, implementation, support, and evolution of the Dutch National e-Infrastructure.

64

Generic Policy Engine for Virtual Infrastructure

Author: Bukhary Ikhwan ismail¹

Co-authors: Hong Hoe Ong²; Mohammad Fairus Khalid²; Mohd Nizam Mohd Mydin²; Rama Tulasi Raju Morampudi²

¹ *Mimos Berhad*

² *MIMOS Berhad*

Corresponding Author: ikhwan.ismail@mimos.my

Service management in form of ITIL or SLAs refers to the activities which are directed by policies. The main goal is to ensure the service delivery i.e. performance, QoS, availability, reliability of service to end user. To implement these policies requires two important stages; service as well policy offering and policy enforcement.

In a typical deployment, policy implementation is embedded onto the service code. Modification, customization or manipulation of policy offering and enforcement would be nearly impossible without major changes to the product. To separate the logic and implementation between service and policy would be difficult.

In this paper we proposed a generic policy engine that decouples both implementations. Our framework would be able to cater the policy management onto multiple services. We present three parts of work.

First, we present an analysis towards decoupling both service and policy. It requires the generalization of policy implementation and analysis of service life cycle at which point of time the enforcement of policy would execute. We identified two distinct phases; non-run time; prior to service provisioning or termination, and run-time; during the service lifetime.

Secondly, we proposed a framework that could be adapted for any services targeted for virtual infrastructure i.e. VM, SDN, or container technology as part of our proof of concept work.

Third, we present our prototype output that covers the policy offering; customization, packaging and enforcement strategies for both service run-time and non-run-time.

From this work, we are able to improve the service offering to the user by providing the flexibility to the service provider to create or customize the policy independently from the service itself.

Summary:

This paper is to share the idea of Generic Policy Engine implemented on Virtual Infrastructure. It is part of our work in progress prototype solution.

Poster Session / 65

Recommending Majors of Students by using Fuzzy Modeling

Author: Ankhtuya Ochirbat¹

Co-author: Timothy K. Shih¹

¹ *National Central University*

Corresponding Author: ankhaa3@yahoo.com

Students in high school or undergraduate stage make critical decisions regarding what to study and which career path to pursue. For various reasons, many of them end up switching to other majors. This may potentially cause mismatch between personality, interest and abilities of the students and characteristics of majors. Such changes are wasteful in time and resources and they produce emotional and economical stresses. Due to the rapid development of society, students need counseling session to enable them to choose a suitable major. The choice of major has become increasingly complex due to the existence of multiple human skills which mean each person or human has their ability at the certain area and can be applied to multiple majors. The main problem of difficulty making a major choice among students is they don't know how make decisions and lack of knowledge and information about majors and occupations. Therefore, to recommend suitable majors to students, it is essential to build a recommendation system that provides direction and guidance to students for choosing their major. Hence, this study proposes the major recommendation system using the fuzzy model based on student's profile, personality, learning style and vocational interest. The study was carried out in accordance with following three questionnaires with a pool of Mongolian 107 eleventh grade students in the fall semester of the 2014/ 2015 academic year. In addition, Holland vocational Interest, big five inventory and index of learning styles questionnaires were employed respectively during the experiment for collecting data.

Networking, Security, Infrastructure & Operations Session V / 67

EGI-CSIRT: Organising Operational Security in evolving distributed IT-Infrastructures

Author: Sven Gabriel¹

¹ *Nikhef*

Corresponding Author: sveng@nikhef.nl

Operational Security in Scientific distributed IT-Infrastructures like EGI are challenging. Existing computation frameworks are further extended, and new technologies implemented. In this evolving environment new policies have to be developed, and existing policies and procedures have to be extended to meet the new requirements. These policies and procedures are then put to a test in so called Security Service Challenges (SSCs).

To efficiently enforce new policies, the security monitoring infrastructure has to be developed to cover all elements of the infrastructure. Finally the incident response tool set has to be extended to be able to efficiently handle security incidents involving new technologies.

In this presentation we will discuss EGI-CSIRTs way towards extending its portfolio to also provide all aspects of operational security in a Cloud environment. This covers the developments around the Virtual Machine Endorsement policy and the related technical aspects towards a trustworthy set of Virtual Machine Images (VMI) offered to the user community through an Application-DataBase.

VMIs with vulnerable configurations were involved in incidents handled by the EGI-CSIRTs Incident Response Task Force (IRTF). This triggered the extension of the existing incident response tool set to allow for central User- and Virtual Machine-Management frameworks, needed to efficiently respond to threads exposed by compromised systems in EGI-FedCloud.

Networking, Security, Infrastructure & Operations Session II / 68

Modeling the Past and Future of Identity Management for Scientific Collaborations

Author: Robert Cowles¹

Co-authors: Craig Jackson²; Von Welch²

¹ *Indiana Univ. CACR*

² *Indiana University CACR*

Corresponding Author: bob.cowles@gmail.com

Over its 3 year funding period, the eXtreme Science Identity Management (XSIM) research project collected and analyzed real world data on identity management (IdM) implementations in virtual organizations (VOs) representing the last 15+ years of collaborative DOE science. Based on that data, we constructed a descriptive VO IdM model. We used the model and existing trends to project the direction for IdM in the 2020 timeframe; and provided guidance to scientific collaborations and resource providers that are implementing or seeking to improve IdM functionality.

XSIM conducted over 20 semi-structured interviews of representatives from scientific collaborations and resource providers, both in the US and Europe; the interviewees supported diverse set of scientific collaborations and disciplines. We developed a definition of “trust,” a key concept in IdM, to understand how varying trust models affect where IdM functions are performed.

We searched for a descriptive IdM model sufficiently complex to produce accurate, useful descriptions of the observed trust relationships and technical implementations, but still simple enough to explain and use in novel situations. It was important that the model be comprehensible to both scientists and IT/Cyber security experts to support a dialog between stakeholder groups with different lexicons. The resulting model identifies how key IdM data elements are utilized in collaborative scientific workflows, and it has the flexibility to describe past, present and future trust relationships and IdM implementations.

In this talk, we will discuss the VO IdM model in depth, including the barriers, motivations, and enablers to IdM delegation and trust we uncovered in our interviews, as well as lessons learned in the process of conducting socio-technical research in this interdisciplinary space and utilizing the model to provide guidance to specific communities. Finally, we describe areas of needed or potentially fruitful research that would enhance the adoption of advanced IdM technologies in future scientific collaborations.

Summary:

We discuss a Virtual Organization (VO) Identity Management(IdM) model in depth, including the barriers, motivations, and enablers to IdM delegation and trust we uncovered in our interviews, as well as lessons learned in the process of conducting socio-technical research in this interdisciplinary space and utilizing the model to provide guidance to specific communities. Finally, we describe areas of needed or potentially fruitful research that would enhance the adoption of advanced IdM technologies in future scientific collaborations.

Humanities, Arts, and Social Sciences Session I / 69

Open Platform for Academic Humanities Data

Author: HARA Shoichiro¹

¹ *Center for Integrated Area Studies, Kyoto University*

Corresponding Author: shara@cias.kyoto-u.ac.jp

Universities are major stakeholders of academic data. Kyoto University, since its foundation in 1897, has collected, created and accumulated numerous and various materials, data and knowledge as its academic resources, and it has developed databases for researchers to access these resources, i.e., KULINE (the university OPAC operated by the library), KURENAI (the university repository developed by the library), KURRA (the university research archives developed by the museum), University Open Course Ware operated by the Academic Center for Computing and Media Studies, and various

databases developed by research institutes/centers. These databases include resources of research processes from data collections (original materials, observation data and experimental data etc.) to publications (papers, books etc.) and archives. However, as each database is independent and heterogeneous system, it is difficult to carry out even simple searches such as finding original experimental data related to the paper. Obviously, we cannot use such isolated databases for advanced research usages to discover hints and/or create new knowledge.

Kyoto University has just launched a new project to develop an innovative database platform adapted to Cloud and Big Data environments to accumulate and link its academic resources, and offered the platform as advanced research utilities. This platform will comprise three sublayers. The first layer is “Open Data Layer” to accumulate and open heterogeneous data. This layer uses RDF (Resource Description Framework) that can describe data of different structures by uniform way. For example, this layer can accumulate thesauri (tree structure), bibliographic catalogues (table structure) and documents (XML) simultaneously. The second layer is “Data Link Layer.” Academic data, especially humanities’ data, are ambiguous (i.e., a term “book” in a database may be expressed as “書” in the other database, “purple” may be the same notion of “紫”, and “cat” and “dog” may be the same category because of they are subordinate concept of “mammal”). This layer uses ontology techniques such as RDFS and OWL to link ambiguous notions and/or vocabularies and to create “Academic Big Data.” Academic Big Data comprise small fragments of heterogeneous data, which will form complex structures. This is the different feature from ordinary big data comprising simple structure data from sensors and IoT devices. The third layer is “Application Layer.” As Academic Big Data is too huge and complicate for researchers to retrieve, categorize and analyze by hands, and then applications to support these processes are necessary. The new project will develop some utilities, i.e., to estimate subjects of contents by natural language techniques, to categorize huge data sets by deep learning techniques, and to organize data according to topological spatiotemporal expressions. This platform will also provide APIs to create mashup applications easily.

This presentation will describe overview and state of progress about our project to promote advanced usages of academic resources of Kyoto University as “linked open data” on Cloud environment.

Humanities, Arts, and Social Sciences Session II / 70

State Space Models in Analyzing Big Election Data

Authors: Dennis Weng¹; Harold D. Clarke²; Karl Ho³

¹ *SUNY Cortland*

² *University of Texas at Dallas*

³ *University of Texas at Dallas*

Corresponding Author: kyho@utdallas.edu

This is a proposal to present a paper at the 2016 ISGC Meeting in Academia Sinica, Taiwan ROC. This study employs new methods in analyzing large amount of poll data from all survey firms to detect so called house and mode effects in election surveys. We collect all polling results and identify house (survey firm) and mode (telephone, internet, in-person) effects on British parties’ vote shares in nearly 2000 public opinion polls conducted between June 2010 and May 6, 2015. The 2015 British general election resulted in the Conservatives winning a majority of seats in parliament. This outcome was surprising because all major polling houses failed to forecast the parties’ vote shares accurately. The widely publicized failure to get the election polling right has reinvigorated debates concerning the accuracy of various survey modes and the British Polling Council is conducting an inquiry concerning what went wrong.

We will estimate house and mode effects using a Bayesian dynamic factor model. The model, pioneered by Jackman (2005), is a state space specification whereby a party’s vote intention share on any day t is a latent variable measured by observed polls conducted by one or more survey houses. A Bayesian approach to this state space model is attractive because it easily accommodates polls conducted by various houses at irregular intervals. We extend the model in three ways:

- estimate mode effects as well as house effects. This will provide evidence relevant to the debate about the accuracy of various survey modes—internet, telephone, in-person—currently used for political polling;
- reconceptualize house effects as dynamic entities, the magnitudes of which can vary over time. This will enable us to estimate the magnitude of possible ‘herding’ effects whereby the vote share estimates of various polling firms converge in the run-up to election day;
- use the dynamic factor model to forecast parties’ vote shares in the 2015 election using various time horizons (one week, one month, three months, six months). In addition unconditional forecasts, we will condition on a second factor that measures the public economic mood over the June 2010 - May 2015 period. Survey data indicate that economic evaluations improved markedly in the year preceding the election and individual-level analyses of national election survey data (Clarke et al., 2015, forthcoming) indicate that economic judgments had strong effects on the choices voters made. These results suggest that a time series forecasting model that conditions on the dynamics of the public economic mood may perform relatively well. We will extend this technique to other polling data including the case in Taiwan.

Humanities, Arts, and Social Sciences Session II / 71

Serious Game Experimentation for Measuring Trust

Author: Alex Yahja¹

Co-author: Marshall Scott Poole²

¹ *National Center for Supercomputing Applications*

² *University of Illinois at Urbana-Champaign*

Corresponding Author: alexy@uiuc.edu

With a deluge of data in this Big Data era, finding out a signal among the background of noise represents a challenge. It is almost impossible to be an expert in many subject matters needed to make rational judgments on an appealing pattern, so we rely on trust. While statistical methods help in finding correlations among data, the question is whether the data itself and/or the persons behind the data could be trusted. This kind of conflict manifests itself in the public debate of Climate Change, for example. In social systems, trust forms a foundation for social interactions. Resultant social network and interaction patterns could reveal the level of trust among social actors. In economics and business, trust forms a basis for economic exchanges.

In this talk, we present our work on an experiment with human participants using a 3D High-Fidelity Graphics serious game in which we measured trust and the effect of an intervention on the trust level. The game has good enough visuals and gameplays to induce players to immerse themselves in the game world. The players in the game are divided into two teams and are tasked to perform a mission. The teams could choose to be competitive and achieve their own team goal or they could choose to collaborate and share truthful information to achieve a greater multi-team goal. Not sharing information could be advantageous for competitive team, as well as intentionally sharing misleading information. Sharing truthful information in a collaborative teamwork depends on the level of trust about the other team. We ran the game for 60 sessions and observed that the subsequent actions and communication of players after a distrust judgment diverge from those after a trust judgment. This shows the implicit trust or distrust have observable phenomena in the form of subsequent actions and communication instances. The network of trust relationships manifests itself as a social network. Trust was shown to have contexts or dimension, e.g., a player may trust another for a particular expertise or task but not for another.

With this result, an interesting avenue opens up to extend the trust measurement using 3D serious game. Future work will use this trust measurement framework to evaluate disaster response scenarios and public policy design. Our trust measurement approach could also be useful to evaluate the analytics and the data in the Big Data deluge.

Poster Session / 73

The Implementation of Combining Problem Based Learning and Hands-on Learning

Author: Tomoya Ikezawa¹

Co-author: Tosh Yamamoto¹

¹ *Kansai University*

Corresponding Author: tomoya.9255@gmail.com

Abstract: The goal of this poster to share the experience of class management for problem based learning and hands-on learning. The purpose of class is to make good effectiveness for learning students by problem based learning and hands-on learning. A variety of problem based learning lecture-styles have been existed, yet, most problem based learning-styles classes have been limited space and under setting. With advances in information and communication technology, it's possible to do a variety of things. Space of learning doesn't need limited space by information and communication technology since various information and communication technologies have been developed for the past few years. For instance, collaborative working system, internet video call system advance and development of file share system. Also, those are able to do bidirectional learning. There are three main points in the implementation of this class. The first one Team based Learning, the second, extracurricular activities of farming, the third, exchanging and building good relationship to no connection to a school people.

The target readers for this poster are teachers, school administrator, government officers, stakeholders on campus, officers and people interested in education. The experiences of problem based learning class on the poster hope that those help your classes.

Networking, Security, Infrastructure & Operations I / 74

Importance of User Deprovisioning from Services

Authors: Michal Prochazka¹; Slavek Licehammer¹

Co-author: Ludek Matyska²

¹ *Masaryk University*

² *CESNET*

Corresponding Author: slavek@ics.muni.cz

Every service uses an authorization process to determine the access rights of individuals. Lots of services do authorization decisions only during the authentication process and by that the information about access rights is valid for the whole session. The other common approach is to run authorization process for each request from the user.

Both of the described approaches are commonly used and they are sufficient for most services. However there are services which enable users to work with persistent resources. An example of such services are cloud infrastructures which enable users to start virtual machines or use data storages for storing large amounts of data. Apart from the normal authorization done whilst user is interacting with the service, there is a need to know that the user is still authorized to use the resources, even though the user is not interacting with the service. Such knowledge enables services to free the persistent resources which were occupied by the user who is no longer authorized.

Deprovisioning is the process which enables service to know about users who are no longer authorized. It is the opposite of the well-known provisioning process, which is used in cases where the services need to know the users in advance of their first usage of the service.

In this paper we will describe the importance of deprovisioning process on real use-cases and services. Moreover we will focus on possible options to implement deprovisioning in existing infrastructures. That requires a well-defined user life-cycle process in the identity management system or a proper connection to the primary sources of user identities, in order to detect if the user is no longer authorized to use the service.

Last but not least, we will describe similarities between standard deprovisioning process and suspension of the users on the services due to security incidents. Based on those similarities, we will demonstrate on a real system how to utilize the deprovisioning process to automate mitigation of security incidents.

Physics & Engineering Session / 75

4th system upgrade of Tokyo Tier2 center

Author: Tomoaki Nakamura¹

Co-authors: Hiroshi Sakamoto²; Nagataka Matsui³; Tetsuro Mashimo³

¹ KEK

² The University of Tokyo

³ University of Tokyo

Corresponding Author: tomoaki.nakamura@kek.jp

The Tokyo Tier2 center, which is located at International Center for Elementary Particle Physics (ICEPP) in the University of Tokyo in Japan, was established as a regional analysis center for the ATLAS experiment. The official operation with Worldwide LHC Computing Grid (WLCG) was started in 2007 after the several years development since 2002. In December 2015, we have replaced a lot of hardware as the fourth system upgrade to cover the requirement of ATLAS experiment in LHC Run2. The total number of CPU cores is not increased from the previous system (9984 cores) including the CPUs for service instance, but the performance of individual CPU core is improved by 5% according to the HEPSPC06 benchmark test (Intel Xeon E5-2680 v3 2.50GHz). Since all worker nodes are made by 24 physical CPU cores configuration, we deploy 416 blade servers in total. They are connected to 10.56PB of disk storage system with 10Gbps internal network backbone by using two center network switch (NetIr

on MLXe-32, Brocade Communication Systems, Inc). The disk storage system is made by 80 of RAID6 disk arrays (Infotrend DS 3024G000F8C16D00) and served by equivalent number of 1U file servers (DELL PowerEdge R630) with 8G-FC connection. Among of the total computing resource in the fourth system, 3840 CPU cores and 7.392PB of storage capacity are reserved for the WLCG worker nodes and ATLAS disk storage area in upcoming three years, respectively. The remaining resources are dedicated to the Japanese collaborators. Since most of the data analysis jobs are I/O bound type jobs, we assigned 10Gbps of internal network bandwidth per two worker node for the effective use of such number of CPU cores. GPFS have been introduced for the non-grid resource, while Disk pool manager (DPM) are continued to be used for WLCG from the third system. In the third system, we had already have 3.168PB of ATLAS data in the DPM storage. All of those data is once migrated to the temporal storage so tha

t Grid jobs can use the data stored in Tokyo Tier2 with reduced number of worker nodes during the migration period. In this talk, we would like to introduce a procedure of the whole scale system upgrade, the improvement of the performance of the system and future perspectives based on the experience at the Tokyo Tier2 center.

iRODS 4.2: Increasing Data Management Programming Productivity and Capability with the Pluggable Rule Engine Architecture

Author: Daniel Bedard¹

Co-authors: Benjamin Keller²; Hao Xu³; Jason Coposky¹; Terrell Russell²

¹ *iRODS Consortium, RENCI at the University of North Carolina*

² *The iRODS Consortium, RENCI at the University of North Carolina*

³ *DICE at the University of North Carolina*

Corresponding Author: danb@renci.org

iRODS, the Integrated Rule-Oriented Data System, is open source data management software for finding, annotating, sharing, and protecting files throughout an organization's entire storage infrastructure. In this talk, we describe a new feature that makes iRODS both easier to use and more versatile: the pluggable rule engine architecture.

A key focus of iRODS Consortium-led development has been making the software modular, which has fueled the growth of a broader developer community and further enhanced iRODS' versatility and reliability. The pluggable rule engine architecture, featured in the upcoming release of iRODS 4.2, will enable administrators to define data management rules in standard programming languages and to enact efficient, certifiable auditing of file activity, while maintaining compatibility with rules written for earlier versions of iRODS.

We will present a Python rule engine plugin to demonstrate how automated data management workflows can be implemented for high programming productivity. We will also present a high performance C++-based auditing rule engine to demonstrate low-level trapping of user and system activity. Each of these rule engine plugins is written with fewer than 100 lines of code. The plugins can run in parallel, and they interoperate with one another.

The pluggable rule engine architecture paves the way for iRODS rule-based automation in practically any programming language, and it will play an important role in the further development of a data management ecosystem and converged infrastructure powered by iRODS.

Opening Ceremony & Keynote Speech I / 77

The Inevitable End of Moore's Law beyond Exascale will Result in Data and HPC Convergence and More

Author: Satoshi Matsuoka¹

¹ *Tokyo Institute of Technology*

The so-called "Moore's Law", by which the performance of the processors will increase exponentially by factor of 2 every 2 years or so, is slated to be ending in 10-15 year timeframe due to the lithography of VLSIs reaching its limits around that time, and combined with other physical factors. This is largely due to the transistor power becoming largely constant, and as a result, means to sustain continuous performance increase must be sought otherwise than increasing the clock rate or the number of floating point units in the chips, i.e., increase in the FLOPS. The promising new parameter in place of the transistor count is the perceived increase in the capacity and bandwidth of storage, driven by device, architectural, as well as packaging innovations: DRAM-alternative Non-Volatile Memory (NVM) devices, 3-D memory and logic stacking evolving from VIAs to direct silicone stacking, as well as next-generation terabit optics and networks. The overall effect of this is that, the trend to increase the computational intensity as advocated today will no longer result in performance increase, but rather, exploiting the memory and bandwidth capacities will instead be the right methodology. However, such shift in compute-vs-data tradeoffs would not exactly be return to the old vector days, since other physical factors

such as latency will not change. As such, performance modeling to account for the evolution of such fundamental architectural change in the post-Moore era would become important, as it could lead to disruptive alterations on how the computing system, both hardware and software, would be evolving towards the future. We are now in the process of launching such innovative projects for the future of computing in Japan

78

Building Smart Regions Through Multi-Sector Collaborations

Author: Kevin Franklin¹

¹ *University of Illinois, Urbana-Champaign*

Corresponding Author: kdf@illinois.edu

The production and consumption of knowledge drives competitiveness in the global economy. As a result, nations and geographic regions that desire to succeed in a new interdependent and rapidly changing world must become “Smart”. They need to collect, process, analyze and interpret data to both develop knowledge and act with increasing efficiency to make decisions that will strengthen their government, business, and education sectors and improve the quality of life of their citizens. There are already a number of national data integration and analysis efforts underway - including the National Data Service, led by the University of Illinois National Center for Supercomputing Applications, and Europe’s EUDAT research data initiatives - that offer potential models for others to follow as they forge academic and industry partnerships (Harmon, 2014). The Organization of American States, Advanced Research and Technology Collaboratory for the Americas (OAS-ARTCA), is working to launch an effort to design and deploy a smart region big data plan for the Americas.

e-Science Activity in Asia Pacific I / 79

eScience Activities in Japan

Corresponding Author: tomoaki.nakamura@kek.jp

e-Science Activity in Asia Pacific I / 80

eScience Activities in China

e-Science Activity in Asia Pacific I / 81

eScience Activities in Korea

e-Science Activity in Asia Pacific I / 82

eScience Activities in Taiwan

83

eScience Activities in Mongolia

e-Science Activity in Asia Pacific I / 84

Panel Discussion

e-Science Activities in Asia Pacific II / 85

eScience Activities in Thailand

e-Science Activities in Asia Pacific II / 86

eScience Activities in Indonesia

e-Science Activities in Asia Pacific II / 87

eScience Activities in Malaysia

e-Science Activities in Asia Pacific II / 88

eScience Activities in Vietnam

e-Science Activities in Asia Pacific II / 89

eScience Activities in Philippine

e-Science Activities in Asia Pacific II / 90

Panel Discussion

e-Science Activities in Asia Pacific III / 91

eScience Activities in India

92

eScience Activities in Pakistan

e-Science Activities in Asia Pacific III / 93

eScience Activities in Singapore (remote presentation)

e-Science Activities in Asia Pacific III / 94

eScience Activities in Australia

e-Science Activities in Asia Pacific III / 95

Panel Discussion

96

Introduction

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

97

Mechanics of Interfederation

98

Requirements for assurance

Corresponding Author: davidg@nikhef.nl

99

Common attack vectors & threats

100

Proper communications mechanisms with reference to IR

101

Introduction to Exercises

102

Hands on session I

103

Hands on session II

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 104

General introduction to Cryo-EM single particles technique

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 105

Introduction to Scipion framework

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 106

Movies alignment (Xmipp)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 107

CTF estimation and screening (Ctffind4)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 108

Particle picking (Xmipp and Eman)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 109

2D classification (Spider)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 110

More picking with templates (Relion)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 111

2D classification (Relion and Xmipp, maybe show precomputed results)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 112

Initial volume methods (Several packages)

Cryo EM Solving the Structure of Macromolecular Complexes: A Hands-on Workshop / 113

3D classification and refinement (Relion, maybe show precomputed results)

Networking, Security, Infrastructure & Operations Session IV / 114

TransPAC – Pragmatic Application-Driven International Networking

Author: Andrew Lee^{None}

International collaboration is greatly improved by a robust set of network connectivity among research and education networks. Big Data requires resilient and abundant bandwidth between data sources and computational resources. National Research and Education networks often need to concentrate their resources building up the facilities that more directly impact the users they serve.

International connectivity is provided by collaborations created among nations such as the TEIN and APAN projects.

The TransPAC project works in close collaboration with other research networks in the Asia-Pacific region to provide resources for network researchers and science user alike. This talk will discuss the current state of the project, including the first trans-Pacific 100G research and education circuit, our ongoing work supporting end users of high-end networks and a brief sketch of future plans.

Keynote Speech II / 115

European Research Infrastructures for Heritage Science

Security Workshop / 116

Introduction

Security Workshop / 117

Mechanisms of Interfederation

Corresponding Author: alessandra.scicchitano@geant.org

Security Workshop / 118

Trust Assurance and Policy for Federated Identity

Corresponding Author: davidg@nikhef.nl

Summary:

Requirements for Assurance and Policy along the AAI chain

Security Workshop / 119

Common attack vectors & threats

Corresponding Author: hannah.short@cern.ch

Security Workshop / 120

Proper communications mechanisms with reference to IR

Corresponding Author: sveng@nikhef.nl

Security Workshop / 121

Introduction to Exercises

Corresponding Author: sveng@nikhef.nl

Security Workshop / 122

Hands on session I

Security Workshop / 123

Hands on session II

Educational Informatics Workshop / 124

The Cutting-Edge Technologies

Educational Informatics Workshop / 125

The Cutting-Edge Technologies

Educational Informatics Workshop / 126

Hands-on Exercise

Educational Informatics Workshop / 127

Hands-on Exercise

Introduction

EGI Federated Cloud for Open Science Tutorial / 129

Hands-on Exercise

Opening Ceremony & Keynote Speech I / 130

Opening Remarks

Keynote Speech II / 131

Complex Big Data: The road to bad policy?

132

Kai Nan

Closing Keynote & Ceremony / 133

NMRbox: Toward Reproducible Computation for Bio-NMR

Closing Keynote & Ceremony / 134

Closing Ceremony

Plenary / 135

Building Smart Regions Through Multi-Sector Collaborations

Biomedicine & Life Science Session II / 136

Prediction of drug targets by curated pharmacophore database

LHCONE Workshop / 137

Meeting Introduction

LHCONE Workshop / 138

LHCOPN status update

LHCONE Workshop / 139

IPv6 deployment update

LHCONE Workshop / 140

WLCG Workshop 2016 report

LHCONE Workshop / 141

LHCONE operations update

LHCONE Workshop / 142

LHCONE VRFs status update

LHCONE Workshop / 143

LHCONE reachability measurement

LHCONE Workshop / 144

perfSONAR introduction and update

LHCONE Workshop / 145

TEIN activities and VRF implementation update

LHCONE Workshop / 146

ASGC activities and VRF implementation update

LHCONE Workshop / 147

Belle II update

LHCONE Workshop / 148

Update on SINET5 implementation

LHCONE Workshop / 149

Commercial Cloud services for WLCG

LHCONE Workshop / 150

GEANT approach to Cloud Providers' R&E traffic

LHCONE Workshop / 151

NORUnet views on cloud services

152

LHCONE P2P service

153

CAS Clouds: a Case Study of Community Cloud

Biomedicine & Life Science Session I / 154

Time Series Analysis of Protein Dynamics

Biomedicine & Life Science Session I / 155

Efficient Large Scale Biomedical Data Analysis

Biomedicine & Life Science Session I / 156

A Novel Approach of Dimensionality Reduction for Revealing Protein Functional Dynamics

Biomedicine & Life Science Session II / 157

Facing Computing and Data Management Challenges in Electron Microscopy: The Scipion Software Framework

EGI Federated Cloud for Open Science Tutorial / 160

Introduction

Author: Gergely SIPOS¹

¹ *EGI.eu*

Corresponding Author: gergely.sipos@egi.eu

The EGI Federated Cloud (<http://go.egi.eu/cloud>) is a standards-based, open cloud system that federates institutional clouds to offer a scalable computing platform for data and/or compute driven applications and services. The infrastructure is already deployed on 20 academic institutes and offers access to approx. 6000 CPU cores, 8000 GB RAM and 430 TB storage. The infrastructure is available for free at the point of access through various interfaces and environments that are customised to the specific needs of users from research and education.

The technologies that enable the infrastructure are based on open solutions and are maintained by the EGI community. These technologies are available for institutes and research communities worldwide to federate cloud services and applications into large-scale infrastructures.

The tutorial will consist of short presentations and hands-on exercises that demonstrate the EGI

Federated Cloud from the user perspective. By covering the following topics the tutorial can be relevant to those who want to become users of the service, and/or want to design cloud applications and cloud infrastructures to support users from academia:

- List item

Introduction to compute clouds and to the distinguishing features of the EGI Federated Cloud.

- List item

Porting applications to the EGI Federated Cloud: Virtual Machines, Image Marketplace, Virtual Organisations.

- List item

Managing applications and data in EGI Federated Cloud: IaaS interfaces, PaaS and GUI environments.

- List item

Next steps to become active user of the EGI Federated Cloud, and/or member of the community.

Intended audience and prerequisites

- List item

Application developers, IT support teams from academia and industry who require access to cloud systems to develop, deploy and operate 'big data' applications and frameworks for researchers and research communities.

- List item

Researchers who would like to understand the basics of cloud computing and gain experience in using cloud resources and applications.

- List item

Representatives of scientific projects and collaborations who want to establish a cloud 'ecosystem' to support community applications and workloads.

Basic understanding and knowledge of cloud computing is a benefit, but not a prerequisite for this tutorial. Each attendee should have access to a PC with internet connection, an SSH client and a Web browser.

Instructor

Gergely Sipos (gergely.sipos@egi.eu) works as Technical Outreach Manager for EGI.eu, the coordinator institute of the EGI community. Gergely coordinates user engagement activities within the EGI community and supports communities exploit EGI services to push the boundaries of science. Since 2015 Gergely coordinates the 'Knowledge Commons' Work Package of the EGI-Engage project. This WP includes eight Competence Centres that support high-impact Research Infrastructures/communities with joint development of customised e-infrastructure services, training and consultancy. Since 2012 Gergely coordinates the training and technical user support of the EGI Federated Cloud infrastructure. Gergely holds PhD in computer science from the University of Miskolc, Hungary. Prior to EGI, Gergely worked in training and user support for the EGEE project from his base in Budapest, where he promoted grid technology and distributed computing practices to scientific communities.

CAS Clouds: A Case Study of Community Cloud

Opening Ceremony & Keynote Speech I / 162

Microsoft Intelligent Cloud

Earth & Environmental Sciences & Biodiversity Session II / 163

The Development of Storm Surge Forecasting System and the Case Study of 2013 Typhoon Haiyan

Earth & Environmental Sciences & Biodiversity Session II / 164

Numerical study of Typhoon Haiyan (2013) with WRF model

Infrastructure Clouds and Virtualisation Session III / 165

Distributed Cloud for e-Science

LHCONE Workshop / 166

Welcome speech and ASGC update

167

Discussion time on LHCONE Asia future development

168

Clouds discussion time

LHCONE Workshop / 169

LHCONE P2P activities introduction

LHCONE Workshop / 170

Next meeting and wrap-up

Humanities, Arts, and Social Sciences Session II / 171

Installation of Public Opinion Sensors: How a Community Based Web Survey Platform Facilitates the Collection of Qualitative Opinion Data

Author: Frank Liu¹

¹ *National Sun Yat-Sen University*

Corresponding Author: frankcsliu@gmail.com

In the emerging field of data science crawling social media for trends and patterns has been identified as important approach to understand the public. To social sciences, however, this approach may not be sufficient when it comes to obtain preferences and opinions about targeted subject or issues. Data collected from social media may not present expected qualitative data from which deep meanings can be drawn for further interpretation. I am arguing that conventional approaches of survey via telephone or face-to-face interviews have started to fail to reach this promise and achieve this goal of data “thickness,” either. Looking from a “thick data” perspective I created a web survey platform to serve as a community in which participants could feel comfortably to be interviewed and traced regarding sensitive and targeted political issues. In this paper I am demonstrating how this goal was achieved by sharing experiences of creating and maintaining smilepoll.tw. Although theoretically it is impossible to installing public opinion sensors to the voters and consumers, a community-based web survey platform could contribute (1) collection of sincerely answered survey responses and (2) the creation of multiple-wave panel data that have not been done in social sciences. The potential of this approach will be discussed. Scholars across disciplines, particularly data science and social sciences may find a new way to collaborate when the need of people and better governance in a democracy become a common interest.

e-Science Activity in Asia Pacific I / 172

e-Science Activities in Mongolian Academy of Sciences

Authors: Batzaya ENKHJARGAL¹; Nergui BAASAN¹

¹ *Mongolian Academy of Science*

This report is covering these issues :

- Current ICT development of Mongolian information technology,
- Current situation of information technology of Mongolian Academy of Sciences,
- e-Science project cooperating with Taiwan Academy of Science, research project cooperating with International research Institutes,
- Objectives and requirements of developing e-science in Mongolia.

ECAI Workshop / 173

GIS of Dr Sun Yat-sen, a life history

ECAI Workshop / 174

Spatial Humanities Going Dutch: Historical Mapping of 17th-Century Formosa Manuscripts

ECAI Workshop / 175

The Taiwan Missions Project - Late 19th Century (1865-1895)

ECAI Workshop / 176

Status of Austronesian Voyaging and Religions

ECAI Workshop / 177

CBETA-RP – an Integrated Digital Research Platform of Chinese Buddhism Materials

ECAI Workshop / 178

Offline Full Text Search on iOS and Android for Chinese Tripitaka and Tibetan Kangyur

ECAI Workshop / 179

The Fo Guang Dictionary of Buddhism Project

180

TBA

ECAI Workshop / 181

Digital Transformations of the Cultural Imaginary

ECAI Workshop / 182

Digital and Physical Preservation: Comparing and Bridging Records of the Xindian First Graveyard

ECAI Workshop / 183

Maritime Buddhism: Current state of research

184

ECAI Project and Content Collection Updates

ECAI Workshop / 185

ECAI Data Portal and SinicaView 4D GIS Platform

ECAI Workshop / 186

ECAI Community Discussion - Project Updates

ECAI Workshop / 187

Arches - to register heritage, Getty conservation

ECAI Workshop / 188

Mapping Chinese Taoist religious networks

LHCONE Workshop / 189

Update on GEANT projects and plans

LHCONE Workshop / 190

Russian VRF introduction and update

LHCONE Workshop / 191

APAN update

Corresponding Author: simon.lin@twgrid.org

Environmental Computing Workshop / 192

Introduction to Environmental Computing Workshop

Corresponding Author: kranzlmueeller@ifi.lmu.de

Environmental Computing Workshop / 193

Development and Implementation of a Global-to-Urban Climate Model Suite

Environmental Computing Workshop / 194

Discussion

Environmental Computing Workshop / 195

The big picture of environmental computing

Corresponding Author: heikku@nm.ifi.lmu.de

Environmental Computing Workshop / 196

Mekong Delta

Environmental Computing Workshop / 197

Environmental Exascale Computing

Corresponding Author: kranzlmue@ifi.lmu.de

Environmental Computing Workshop / 198

Land Use Development Simulation Systems

Environmental Computing Workshop / 199

Application of numerical model on extreme weather and environmental studies

Environmental Computing Workshop / 200

The Applications of Advanced Numerical Simulation on the Tsunami and Flooding Hazard Mitigation

Environmental Computing Workshop / 201

Panel Discussion

202

Panel Discussion

LHCONE Workshop / 203

Discussion time on LHCONE Asia future development

LHCONE Workshop / 204

ESnet activities and plans

LHCONE Workshop / 205

NORDUnet view on cloud services

LHCONE Workshop / 206

Cloud discussion time

Monitoring BoF / 207

Monitoring at a Multi-Site Tier 1

Monitoring BoF / 208

Visualization of dCache accounting information with state-of-the-art Data Analysis Tools

Corresponding Author: tigran.mkrtchyan@desy.de

Monitoring BoF / 209

CERN Network and WLCG Network Monitoring

Corresponding Author: edoardo.martelli@cern.ch

Monitoring BoF / 210

Introduction of IHEP Monitoring System

e-Science Activities in Asia Pacific III / 211

eScience Activity in Nepal

LHCONE Workshop / 212

BGP Route Server Proof of Concept (remote presentation)

LHCONE Workshop / 213

Caltech on SDN activities

LHCONE Workshop / 214

Getting SDN to DevOps in ATLAS

LHCONE Workshop / 215

Questions and Next steps

LHCONE Workshop / 216

Nova experiment request to join LHCONE

LHCONE Workshop / 217

KISTI activities and plans