

International Symposium on Grids & Clouds 2017 (ISGC 2017)

Sunday 05 March 2017 - Friday 10 March 2017

BHSS, Academia Sinica

Book of Abstracts

Contents

Explore multi-core virtualization on the ATLAS@home project	1
Factors Affecting Air Surface Temperature Variations in the Philippines	1
The DisVis and PowerFit web servers: Explorative and Integrative Modeling of Biomolecular Complexes harvesting EGI GPGPU resources	2
Building CNGrid As a HPC Application Cloud Service Provider	3
EGI federated platforms supporting accelerated computing	4
The 'Cloud Area Padovana': lessons learned after two years of a production OpenStack-based IaaS for the local INFN user community	5
Synergy, a new approach for optimizing the resource usage in OpenStack	6
The SDN application on data transfer and cyber security in IHEP	6
Convective cloud model calibration for short and medium term precipitation forecast in the Philippines	7
Exploiting clouds for smart cities applications - The Cagliari 2020 project	7
A Novel Architecture towards Exascale Computing	8
The EGI CernVM-FS infrastructure - latest developments and evolution towards a global facility	9
VCondor - an implementation of dynamic virtual computing cluster	9
WLCG Security Operations Centres Working Group	10
Using Advanced e-Systems for Community-Engaged Research	11
Revealing Philippine Climate Type Using Remotely-sensed Rainfall Estimates	11
IPv6 Deployment and Migration of WLCG TEIR-2 site resources on Private Cloud	12
The High Throughput Strategy of IHEP	12
Examination of dynamic partitioning for multi-core jobs in the Tokyo Tier-2 center	13
Collaborating for WISER Information Security	14
Status of Network Security Operations at IHEP	14

Design and Implementation of Unified Authentication Management System of IHEP . . .	15
The Billing System of IHEP Data Center	16
NeIC EISCAT_3D support project: Nordic computing challenge	17
Towards a cloud-based computing and analysis framework to process environmental science big data	18
GUOCCI – The Entryway to Federated Cloud for Small-scale Users	19
Platform for Humanities Open Data	19
Monitoring Virtual Devices in Mass Storage Environments	20
Data storage accounting at RAL	21
Running WLCG and HPC Workload on Single Openstack Cloud	21
TRANSITION FROM IPv4 TO IPv6 IN MONGOLIA: KEY ISSUES AND RECOMMENDATIONS	22
SDN implementation in IHEP	23
dCache, managing Quality of Service in Cloud Storage Infrastructures	24
2D and 3D Medical Images for Anatomy Education using a cloud computing platform . .	25
Data Provenance Tracking as the Basis for a Biomedical Virtual Research Environment .	25
Molecular dynamics of proteins in the cloud	26
EGI-CSIRT: Coordinating Operational Security in evolving distributed IT-Infrastructures.	27
A Proposal: ePortfolio for enhancing active learning for the future generation	27
Can R&E federations trust Research Infrastructures?	28
ICT Enhanced Interactive Remedial Mathematics Program for Science and Engineering Students	29
THE VIRTUAL REALITY (VR) TRAINING SYSTEM FOR DISASTER PREPAREDNESS . .	29
Investigating community detection algorithms and their capacity as markers of brain diseases	30
Automated quality control for data from ASTI automatic weather stations	31
A solution for secure use of Kibana and Elasticsearch in multi-user environment	32
Using virtualized computing resources with the DIRAC Interware	32
Towards Environmental Computing Compendium	33
Identifying Suspicious Network Activities in Grid Network Traffic: Finding the needle in a stack of needles	34

Design and Implementation of Portal System for Subscribed Cloud Services in Identity Federation	35
Machine Learning analysis of CMS data transfers	35
A Method for Remote Initial Vetting of Identity with PKI Credential	36
Modern Monitoring Systems	37
dCache, towards Federated Identities and Anonymized Delegation	37
When and Why Are Citizens OK to Publicize Their Data?	38
Occupation recommendation with major programs for adolescents	38
Framework for distributing Radio Astronomy processing across Clusters and Clouds	39
GridPP wider community support and SKA synergies	40
Virtualized Web Portals in EGI Federated Cloud	40
Data Science as a Foundation Toward Open Data and Open Science: The Case of Taiwan Indigenous Peoples open research Data (TIPD)	41
NMRBox and VCell: Common Flexible Infrastructure Supporting Two Very Different Virtual Research Communities	42
Framework for Developing Cloud enabled Applications for IoT	43
CloudIPStore: A Cloud SaaS Repository for your Intellectual Properties	43
Security Incident Response Procedure for Inter-Federation	44
e-Sciences Activities in MAS	44
A Preliminary Study on Reconstructing Faded Color by Spectral Estimation Method for Heritage Object	45
e-Science Activities in Thailand	45
GSDC activities for scientific computing	45
Ten years of operations at the PIC Tier-1	46
Application of non-uniform sampling method in NMR spectroscopy	46
General introduction to biological cryoEM	47
Image Processing in cryoEM: Open problems and current perspectives	47
EMAN 2 (Part 1)	47
EMAN 2 (Part 2)	47
Leggion and Appion	47
Image Processing in cryoEM: Open problems and current perspectives	47

Three key breakthroughs that enable cryo-EM to become a routine density map generator	47
Relion (Part 1)	47
Relion (Part 2)	48
Three key breakthroughs that enable cryo-EM to become a routine density map generator	48
Applications of cryo-electron microscopy to understand complex structures	48
e-Science Activities in Japan - Building Academic Inter-Cloud Infrastructure	48
eScience Activities in China	48
eScience Activities in Korea	48
e-Science Activities in Taiwan	48
Q&A	48
e-Science Activities in Indonesia	49
eScience Activities in Malaysia	49
eScience Activities in Vietnam	49
eScience Activities in Philippine	49
Q&A	49
eScience Activities in Singapore	49
e-Science Activities in India	49
eScience Activities in Australia	49
eScience Activities in Nepal	50
eScience Activities in Pakistan	50
Q&A	50
Opening Remarks	50
On-the-fly Capacity Planning in Support of High Throughput Workloads	50
Big Data-Driven Drug Discovery	50
High Performance Computing Environment and Applications in CAS	50
Application of non-uniform sampling method in NMR spectroscopy	51
Introduction	51
Security intrusions and their detection	51
How to perform forensic analysis?	51
How have Federated Cloud security incidents been detected?	51

Security Incident handling in Federated Clouds	51
Introduction to the hands-on exercises	51
Hands on session I	52
Hands on session II	52
Wrap-up and conclusions	52
Introduction: Goal Setting	52
PBL through TBL in Active Learning Way	52
Writing Program (ESL, the mirror of the actively learning mind)	52
Writing Program (JSL, the mirror of the learning mind)	53
Fundamental Math Programs (symbol manipulation and higher order thinking)	53
ePortfolio for Learning Goal Settings & Artifacts of Learning Processes	53
ePortfolio for career development (Scenario Planning for the future)	53
Hands-on session for the participants	53
summary and wrap-up session	53
Efficiency Improvement on Distributed Cloud System	54
Future warming scenario and impacts study over Taiwan: Results from ECHAM5/MPIOM- WRF dynamical downscaling	54
Supporting Open Science with the EGI Federated Cloud - Experiences, success stories, lessons learnt	54
Q&A	54
Q&A	54
Q&A	54
Q&A	55
Image processing in cryo Electron Microscopy (cryo EM): Analyzing reliability and quality	55
High-resolution Integrative Modelling of Biomolecular Complexes from Fuzzy Data	55
POWERFIT and DISVIS (part 1)	55
POWERFIT and DISVIS (part 2)	55
Scipion (part 1)	56
Scipion (part 2)	56
Closing remark/round table discussion	56

Docker and dCache	56
Container Technology and Software Delivery	56
Running LHC jobs using Kubernetes	56
Docker for tests and software preservation in LHCb	56
Citizen Earthquake Science in Taiwan	57
The Emergence of Computational Archival Science	57
Caches all the way down: Infrastructure for Data Science	57
Listening to the ecosystem: the integration of machine learning and a long-term sound- scape monitoring network	58
Exploring the Bujang Valley - A Pre-production Survey	58
Atlas of Maritime Buddhism Reports on the Initial 3D Filming Projects in Myanmar, India, and Sri Lanka and Plans for Data Collection	58
Maritime Religious Networks: Aesthetic Sources of Flora and Fauna Motifs from India and China on Tomb Elaborations in Taiwan	58
Orphan Works and Collective Memories: On Setting Up The Sunflower Movement Archive	59
Metadata Development and Documentation for a Research Data Repository	59
Regional Religious System and Social Changes in China	59
Automatic Collective Commentaries with Corpus Positioning System	59
Mining Textual Image Data with Cloud Technologies: Challenges and Opportunities	59
Challenges and Possibilities: Translating and Building Buddhist Lexicography in the Mod- ern Era	59
Dictionary Translation Project	59
Report on Technical Developments	60
Stopping the flow - The Yellow River and China's Grand Canal after 1855	60
Volunteered Geographic Information and Arches	60
Digital Economy and Asian Production Network – A Reality Check for Humanities	60
Endangered Languages and Flow of Identities: State Policies and Ethnic Boundaries of the Thao	60
History and Projections of Tombs Research in the Taiwan Area	60
Earth Deity Mapping and Community Networks in Taiwan	60
Affiliate Updates: Short reports on work in progress	61

Translation of Buddhist Texts: Building of Translation Memories, Terminology Databases and Corpus Analysis Tools	61
Introduction	61
Numerical Analysis on Mesoscale Dynamics of the Extreme Rainfall and Flood Event (May 2016) over Sri Lanka	61
Q&A	61
Parameterization Study of Chemically Reactive Pollutant Dispersion using Large-Eddy Simulation	61
Parallel taxonomic classification algorithm for metagenomic sequences	61
Detecting Incidents with Limited Linguistic Knowledge for Low Resource Languages . .	61
Malaysia	62
Indonesia	62
Philippines	62
Vietnam	62
Taiwan	62
Thailand	62
Panel Discussion: State of the Art and new opportunities/challenges	62
Introduction	63
EUGridPMA Update	63
TAGPMA Update	63
APGridPMA Update	63
New NAREGI CA Software System	63
KISTI CA Report and Review	63
Remote Vetting	63
Member Report	63
Collaboration on monitoring Asian soundscape and the challenges	64
GDB Introduction	64
ASGC Report	64
Asian Tier Forum report	64
Asian Network Status	64
Traceability & Isolation Working Group update	64

IPv6 Rollout	65
Sirtfi - Incident response for for Identity Federations	65
WLCG Workshop Supdate	65
Tier 1 Configuration Evolution & Options	65
Workload Management Trends in WLCG	65
Wrap Up & WLCG Workshop update	65
In Silico Drug Discovery of Potential HCV Helicase Inhibitors	65
GROMACS in the Clouds: A Portative User-friendly Interface Bridged to Grid-computing Resources	66
Automated Quality Control for Data from ASTI Automatic Weather Stations	66
ICT Enhanced Interactive Remedial Mathematics Program for Science and Engineering Stu- dents	66
The Virtual Reality (VR) Training System for Disaster Preparedness	66
Virtualized Web Portals in EGI Federated Cloud	66
A Switching Mechanism of Visualization Middleware and Application Using Docker	66
Design and Implementation of Unified Authentication Management System of IHEP	67
Security Incident Response Procedure for Inter-Federation	67
Ten Years of Operations at the PIC Tier-1	67
Monitoring Virtual Devices in Mass Storage Environments	67
Metadata as Linked Data for Research Data Repositories	67
A Single Rack Cloud Center with Unprecedented Power and Thermal Efficiency	67
Toward Construction of Resilient Software-Defined IT Infrastructure for Supporting Disas- ter Management Applications	68

Massively Distributed Computing and Citizen Sciences / 1**Explore multi-core virtualization on the ATLAS@home project****Author:** wenjing wu¹**Co-authors:** Andrej Filipcic²; David Cameron³¹ *IHEP*² *Department of Experimental High Energy Physics, Jozef Stefan Institute*³ *Department of Physics, University of Oslo***Corresponding Author:** wuwj@ihep.ac.cn

The exploitation of volunteer computing resources has become a popular practice in the HEP (High Energy Physics) computing community as the huge amount of potential computing power it provides. ATLAS@home, as a pioneer project in this practice, uses the BOINC middleware to harness thousands of worldwide volunteer computers, has been harvesting a very considerable number of computing resources since its official launch in 2013. With more and more resource diverse volunteer computers participating in this project, it is very necessary to explore ways to optimize the usage of volunteer computing resources in terms of memory, storage and CPU to qualify more computers of running ATLAS@home jobs. The ATLAS software Athena has already supported multi-core processing, and been running stably and efficiently on its Grid sites for over 1 year. Statistics from Grid sites show that using multi-core processing can significantly reduce the total memory usage upon utilizing the same amount of CPU cores. Based on this practice, we explore the usage of multi-core virtualization on the ATLAS@home project, which is to spawn one virtual machine with all the available CPU cores from the volunteer computer, and then a multi-core Athena job is launched inside this virtual machine. This Athena job uses all the available CPU cores from the virtual machine. The ATLAS@home multi-core application was officially launched in July 2016 and the statistics from a few months' full load running confirmed the reduction of total memory usage, however the performance of this practice is different from the Grid sites due to the virtualization. Through the historical statistics and testing, we find out that factors including the allocation of CPU cores, different versions of hypervisors and the scheduling policies of the BOINC middleware can significantly impact the performance of the multi-core application.

In this paper, we will cover the following aspects of this practice: 1) the implementation of the multi-core virtualization; 2) experiences gained through a few months' full load running; 3) tuning and optimization of its performance.

2

Factors Affecting Air Surface Temperature Variations in the Philippines**Author:** John Christian Lequiron¹**Co-author:** Gerry Bagtasa¹¹ *Institute of Environmental Science and Meteorology, University of the Philippines, Diliman, Quezon City***Corresponding Author:** jonglequiron@gmail.com

Changes in air surface temperature plays an important role in the Philippine's economy, industry, health, and food production. While increasing global mean temperature in the recent several decades has prompted a number of climate change and variability studies in the Philippines, most studies still focus on rainfall and tropical cyclones. This study aims to investigate the trend and variability of observed air surface temperature and determine its major influencing factor/s in the Philippines. A non-parametric Mann-Kendall trend test was applied to monthly mean temperature of 17 synoptic stations covering 56 years from 1960 to 2015 and a mean change of 0.58 °C or a positive trend of 0.0105 °C/year ($p < 0.05$) was found. In addition, wavelet decomposition was used to determine the

frequency of temperature variability show a 12-month, 30-80-month and more than 120-month cycles. This indicates a strong annual variations, an interannual variations that coincides with ENSO events, and an interdecadal variations that is attributed to PDO and CO₂ concentrations. Air surface temperature was also correlated with smoothed sunspot number and galactic cosmic rays, the results show a low to no effect. The influence of ENSO teleconnection on temperature, wind pattern, cloud cover, and outgoing longwave radiation on different ENSO phases had significant effects on regional temperature variability. Particularly, an anomalous anticyclonic (cyclonic) flow east of the Philippines during the peak and decay phase of El Niño (La Niña) events leads to the advection of warm southeasterly (cold northeasterly) air mass over the country. Furthermore, an apparent increasing cloud cover trend is observed over the West Philippine Sea including portions of the Philippines, and this is believed to lessen the effect of the increasing air surface temperature. However, relative humidity was also found to be increasing especially on the central part of the country, which results in a high positive trend of heat index, exacerbating the effects on human discomfort. Finally, an assessment of gridded temperature datasets were done to look at the viability of using three high-resolution datasets in future climate analysis and model calibration and verification. Several error statistics (i.e. Pearson correlation, Bias, MAE, and RMSE) were used for this validation. Results show that gridded temperature datasets generally follows the observed surface temperature change and anomalies. In addition, it is more representative of regional temperature rather than a substitute to station-observed air temperature.

Summary:

Changes in air surface temperature plays an important role in the Philippine's economy, industry, health, and food production. While increasing global mean temperature in the recent several decades has prompted a number of climate change and variability studies in the Philippines, most studies still focus on rainfall and tropical cyclones. This study aims to investigate the trend and variability of observed air surface temperature and determine its major influencing factor/s in the Philippines. A non-parametric Mann-Kendall trend test was applied to monthly mean temperature of 17 synoptic stations covering 56 years from 1960 to 2015 and a mean change of 0.58 °C or a positive trend of 0.0105 °C/year ($p < 0.05$) was found. In addition, wavelet decomposition was used to determine the frequency of temperature variability show a 12-month, 30-80-month and more than 120-month cycles. This indicates a strong annual variations, an interannual variations that coincides with ENSO events, and an interdecadal variations that is attributed to PDO and CO₂ concentrations. Air surface temperature was also correlated with smoothed sunspot number and galactic cosmic rays, the results show a low to no effect. The influence of ENSO teleconnection on temperature, wind pattern, cloud cover, and outgoing longwave radiation on different ENSO phases had significant effects on regional temperature variability. Particularly, an anomalous anticyclonic (cyclonic) flow east of the Philippines during the peak and decay phase of El Niño (La Niña) events leads to the advection of warm southeasterly (cold northeasterly) air mass over the country. Furthermore, an apparent increasing cloud cover trend is observed over the West Philippine Sea including portions of the Philippines, and this is believed to lessen the effect of the increasing air surface temperature. However, relative humidity was also found to be increasing especially on the central part of the country, which results in a high positive trend of heat index, exacerbating the effects on human discomfort. Finally, an assessment of gridded temperature datasets were done to look at the viability of using three high-resolution datasets in future climate analysis and model calibration and verification. Several error statistics (i.e. Pearson correlation, Bias, MAE, and RMSE) were used for this validation. Results show that gridded temperature datasets generally follows the observed surface temperature change and anomalies. In addition, it is more representative of regional temperature rather than a substitute to station-observed air temperature.

Biomedicine & Life Science I / 4

The DisVis and PowerFit web servers: Explorative and Integrative Modeling of Biomolecular Complexes harvesting EGI GPGPU resources

Author: Alexandre Bonvin¹

Co-authors: Antonio Rosato²; Gydo van Zundert³; Joerg Schaarschmidt³; Marco Verlato⁴; Mario David⁵; Mikael Trellet³; Seynep Kurkcuglu³

¹ *Utrecht University*

² *University of Florence, Italy*

³ *Utrecht University, the Netherlands*

⁴ *INFN, Padova, Italy*

⁵ *LIP, Lisbon, Portugal*

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

Structure determination of complex molecular machines requires combination of an increasing number of experimental methods with highly specialized software geared towards each data source to properly handle the gathered data. Recently we introduced the two software packages PowerFit [1,2] and DisVis [3]. These combine high-resolution structures of atomic subunits with density maps from cryo-electron microscopy or distance restraints, typically acquired by chemical cross-linking coupled with mass spectrometry, respectively. To facilitate their use by a broad community, they have been implemented as web portals harvesting both local CPU resources and GPGPU-accelerated EGI HTC resources [4], making use of GPGPU-enabled Docker containers developed under the MoBrain competence center of EGI-Engage[5] and the INDIGO-Datacloud EU project [6]. The web portals offer user-friendly interfaces, while minimizing computational requirements, and provide a first interactive view of the results. The portals can be accessed freely after registration via <http://milou.science.uu.nl/services/DISVIS> and <http://milou.science.uu.nl/services/POWERFIT>.

1. G.C.P. van Zundert and A.M.J.J. Bonvin. Fast and sensitive rigid-body fitting into cryo-EM density maps with PowerFit. *AIMS Biophysics*. **2**, 73-87 (2015).
2. G.C.P. van Zundert and A.M.J.J. Bonvin. Defining the limits and reliability of rigid-body fitting in cryo-EM maps using multi-scale image pyramids. *J. Struct. Biol.*, **195**, 252-258 (2016).
3. G.C.P. van Zundert and A.M.J.J. Bonvin. DisVis: Quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics*. **31**, 3222-3224 (2015).
4. <http://www.egi.eu>
5. <https://mobrain.egi.eu>
6. <http://www.indigo-datacloud.eu>

Supercomputing, High Throughput, Accelerator Technologies and Integration / 5

Building CNGrid As a HPC Application Cloud Service Provider

Author: Haili XIAO¹

¹ *Supercomputing Center, Chinese Academy of Sciences*

Corresponding Author: haili@sccas.cn

Public grids (WLCG, OSG, XSEDE2) provide huge computing resources to scientific users all around the world, those services are continuously and constantly evolving for the last 20 years. In the mean while, recently public clouds show great interest in HPC besides its traditional IaaS/PaaS/SaaS market. Companies like Amazon, Microsoft, Google and Ali all have their own HPC cloud solutions. As one of the biggest computing grid in the world, China National Grid (CNGrid) integrates top HPC resources all over China, including Tianhe, Tianhe2 and Shenwei. It has been providing grid computing services for scientific user for more than 15 years. CNGrid has developed its own grid software SCE, as the kernel middleware to link all the HPCs and schedule job requests. Based upon SCE, there is a RESTful wrapper called SCEAPI, enabling rapid development of a dozen of science gateways—specialized domain communities or easy-to-use mobile apps. A very successful example is that CERN ATLAS simulation jobs are running in CNGrid, by connecting ARC-CE middleware and SCEAPI.

Summary:

In conclusion, CNGrid becomes a HPC application cloud service provider, it's a very good practice of fusion between grid and cloud technologies.

Supercomputing, High Throughput, Accelerator Technologies and Integration / 6**EGI federated platforms supporting accelerated computing**

Authors: David Rebatto¹; Jan Astalos²; Lisa Zangrando³; Marco Verlato³; Miroslav Dobrucky²; Paolo Andreetto³; Viet Tran²

¹ *Istituto Nazionale di Fisica Nucleare - Sez. di Milano, Italy*

² *Institute of Informatics Slovak Academy of Sciences, Slovakia*

³ *Istituto Nazionale di Fisica Nucleare - Sez. di Padova, Italy*

Corresponding Author: marco.verlato@pd.infn.it

While accelerated computing instances providing access to NVIDIA GPUs are already available since a couple of years in commercial public clouds like Amazon EC2, the EGI Federated Cloud has put in production its first OpenStack-based site providing GPU-equipped instances at the end of 2015. However, many EGI sites which are providing GPUs or MIC co-processors to enable high performance processing are not directly supported yet in a federated manner by the EGI HTC and Cloud platforms. In fact, to use the accelerator cards capabilities available at resource centre level, users must directly interact with the local provider to get information about the type of resources and software libraries available, and which submission queues must be used to submit accelerated computing workloads.

EU-funded project EGI-Engage since March 2015 has worked to implement the support to accelerated computing on both its HTC and Cloud platforms addressing two levels: the information system, based on the OGF GLUE standard, and the middleware. By developing a common extension of the information system structure, it was possible to expose the correct information about the accelerated computing technologies available, both software and hardware, at site level. Accelerator capabilities can now be published uniformly, so that users can extract all the information directly from the information system without interacting with the sites, and easily use resources provided by multiple sites. On the other hand, HTC and Cloud middleware support for accelerator cards has been extended, where needed, in order to provide a transparent and uniform way to allocate these resources together with CPU cores efficiently to the users.

In this paper we describe the solution developed for enabling accelerated computing support in the CREAM Computing Element for the most popular batch systems and, for what concerns the information system, the new objects and attributed proposed for implementation in the version 2.1 of the GLUE schema.

For what concerns the Cloud platform, we describe the solutions implemented to enable GPU virtualization on KVM hypervisor via PCI passthrough technology on both OpenStack and OpenNebula based IaaS cloud sites, which are now part of the EGI Federated Cloud offer, and the latest developments about GPU direct access through LXD container technology as a replacement of KVM hypervisor.

Moreover, we showcase a number of applications and best practices implemented by the structural biology and biodiversity scientific user communities that already started to use the first accelerated computing resources made available through the EGI HTC and Cloud platforms.

Summary:

In this paper we describe the solution developed for enabling accelerated computing support in the CREAM Computing Element for the most popular batch systems and, for what concerns the information system, the new objects and attributed proposed for implementation in the version 2.1 of the GLUE

schema. We describe also the solutions implemented to enable GPU virtualization on KVM hypervisor via PCI passthrough technology on both OpenStack and OpenNebula based IaaS cloud sites, which are now part of the EGI Federated Cloud offer, and the latest developments about GPU direct access through LXD container technology as a replacement of KVM hypervisor.

Moreover, we showcase a number of applications and best practices implemented by the structural biology and biodiversity scientific user communities that already started to use the first accelerated computing resources made available through the EGI HTC and Cloud platforms.

Infrastructure Clouds and Virtualisation I / 7

The 'Cloud Area Padovana': lessons learned after two years of a production OpenStack-based IaaS for the local INFN user community

Authors: Alberto Crescente¹; Alvise Dorigo¹; Ervin Konomi¹; Fabrizio Chiarello¹; Federica Fanzago¹; Fulvia Costa¹; Lisa Zangrando¹; Marco Verlati¹; Massimo Sgaravatto¹; Matteo Segatta¹; Nicola Tritto¹; Paolo Andreetto¹; Sergio Fantinel²; Sergio Traldi¹

¹ *INFN-Padova*

² *INFN-LNL*

The Cloud Area Padovana is an OpenStack-based scientific cloud, spread across two different sites - the INFN Padova Unit and the INFN Legnaro National Labs - located 10 km away but connected with a dedicated 10 Gbps optical link. In the last two years its hardware resources have been scaled horizontally by adding new ones: currently it provides about 1100 logical cores and 50 TB of storage.

Special in-house developments were also integrated in the OpenStack dashboard, such as a tool for user and project registrations with direct support for Single Sign-On via the INFN-AAI Identity Provider as a new option for the user authentication. The collaboration with the EU-funded INDIGO-DataCloud project, started one year ago, allowed to experiment the integration of Docker-based containers and the fair-share scheduling: a new resource allocation mechanism analogous to the ones available in the batch system schedulers for maximizing the usage of shared resources among concurrent users and projects. Both solutions are expected to be available in production soon.

The entire computing facility now satisfies the computational and storage demands of more than 100 users afferent to about 25 research projects. In this paper we'll present the architecture of the Cloud infrastructure, the tools and procedures used to operate it ensuring reliability and fault-tolerance. We'll especially focus on the lessons learned in these two years, describing the challenges identified and the subsequent corrective actions applied. From the perspective of scientific applications, we'll show some concrete use cases on how this Cloud infrastructure is being used. In particular we'll focus on two big physics experiments which are intensively exploiting this computing facility: CMS and SPES. CMS deployed on the cloud a complex computational infrastructure, composed of several user interfaces for job submission in the Grid environment/local batch queues or for interactive processes; this is fully integrated with the local Tier-2 facility. To avoid a static allocation of the resources, an elastic cluster, based on cernVM, has been configured: it allows to automatically create and delete virtual machines according to the user needs. SPES is using a client-server system called TraceWin to exploit INFN's virtual resources performing a very large number of simulations on about a thousand nodes elastically managed.

Summary:

In this paper we'll present the architecture of the Cloud infrastructure, the tools and procedures used to operate it ensuring reliability and fault-tolerance. We'll especially focus on the lessons learned in these two years, describing the challenges identified and the subsequent corrective actions applied. From the perspective of scientific applications, we'll show some concrete use cases on how this Cloud infrastructure is being used. In particular we'll focus on two big physics experiments which are intensively exploiting this computing facility: CMS and SPES.

Infrastructure Clouds and Virtualisation I / 8**Synergy, a new approach for optimizing the resource usage in OpenStack****Author:** Lisa Zangrando¹**Co-authors:** Ervin Konomi²; Marco Verlato²; Massimo Sgaravatto²; Vincent Llorens³¹ *INFN - Sez. Padova*² *INFN*³ *in2p3***Corresponding Author:** lisa.zangrando@pd.infn.it

Managing resource allocation in a Cloud based data center serving multiple virtual organizations is a challenging issue. In fact, while the LRMS (Local Resource Management Systems) are able to maximize the resource usage by fairly distributing computing resources among different user groups according to specific policies imposed by the data centre administrator, this is not so straightforward in the most common Cloud management frameworks (e.g. OpenStack, OpenNebula).

For example, the current OpenStack implementation provides a too simplistic scheduling model based on an immediate First Come First Served paradigm. Therefore, a user request will be rejected if no resources are immediately available: it is then up to the user to later re-issue the same request. Moreover the resource provisioning is only limited to the static partitioning strategy. In particular each project is assigned an agreed and fixed quota of resources that cannot be exceeded by one group even if there are unused resources allocated to other groups.

☒The EU-funded INDIGO-DataCloud project is addressing this issue through 'Synergy', a new advanced scheduling and resource provisioning service targeted at OpenStack. With Synergy it is possible to maximize the resource utilization by allowing OpenStack projects to consume extra shared resources in addition to those statically allocated. Therefore such projects can now access two different kind of quotas: the private quota and the shared one. The first one is the OpenStack quota operated in a standard OpenStack way. The shared quota is instead handled by Synergy and is composed of resources non statically allocated. Such shared resources are fairly distributed among users following the fair-share policies defined by the administrator. In case the user request cannot be immediately satisfied, it is not rejected but instead inserted in a persistent priority queue and scheduled later.

We present the architecture of Synergy, the status of its implementation, some results demonstrating its functionalities and the foreseen evolution of the service.

Network, Security, Infrastructure & Operations III / 9**The SDN application on data transfer and cyber security in IHEP****Author:** Fazhi QI¹**Co-authors:** Shan Zeng²; Zhihui Sun²¹ *Institute of High Energy Physics, CAS*² *IHEP***Corresponding Author:** qfz@ihep.ac.cn

Network performance and security are important for HEP experiment data and information exchange. the Software Defined Network(SDN)

gives us a new idea and choice to design and implement a new network model to guarantee or improve the network performance and security. There are two SDN application projects in IHEP that have been launched, SDN@WAN is focusing on the quality of HEP data transfer among different computing sites around China with the current available network infrastructure and SDN@SEC is focusing on the network intrusion-detection and protection with SDN architecture.

10

Convective cloud model calibration for short and medium term precipitation forecast in the Philippines

Author: Gerry Bagtasa¹

¹ *Institute of Environmental Science & Meteorology, University of the Philippines*

Corresponding Author: gbagtasa@iesm.upd.edu.ph

In the Philippines, rain is the most important weather parameter. In particular, convective rain or cumulonimbus clouds make up majority of precipitation during the Asian southwest monsoon season. Aside from the hazards associated with intense rainfall, crop productivity of rain-fed agricultural farms relies on the onset and amount of seasonal precipitation. This study explores the calibration of several parameters in the Kain-Fritsch cumulus parameterization included in the WRF weather model. Convection trigger, CAPE consumption time and auto-conversion rate are found to be most sensitive in rainfall amount quantification. However, parameter calibration varies with different global models (i.e. GFS, CFS, etc.) that drive the WRF model. In addition, significant improvement is seen with adjusted cumulus mode while its impact on overall computing time is negligible. Case studies for (1) a successive thunderstorm events for short-range forecast and (2) five month-long rain and wind for medium-term seasonal prediction will be presented.

Physics & Engineering II / 11

Exploiting clouds for smart cities applications - The Cagliari 2020 project

Author: Alberto Masoni¹

Co-author: Daniele Mura²

¹ *INFN National Institute of Nuclear Physics*

² *INFN - National Institute of Nuclear Physics - Italy*

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

CAGLIARI 2020 is a 25 million euro project funded within the framework of the National Operational Program for Research and competitiveness of the Italian Ministry of Education University and Research.

The project starts end 2016 with a duration of three years. The partnership includes public and private organisms of the South Sardinia for the development of ICT technologies aimed at optimising the usage of the "city system" and improving the quality of life for those who work and/or live in the city.

The main goal of CAGLIARI 2020 is the development of innovative and environmentally friendly solutions for urban mobility (and possibly metropolitan area mobility), so to boost energy and environmental performances.

The project idea originates from the ever increasing need of innovative tools and technological solutions for the optimisation of urban mobility, for lowering travel times and improving air quality. Cagliari represents the ideal case study for the development and testing of the project, mainly because of its centralised public transport management system ranking among the most advanced in

Europe.

CAGLIARI 2020 is based on the study and testing of a sensors network comprised of:

1. Fixed sensors for the tracking of vehicles entering/exiting the urban area. These sensors allow real-time and/or historical analysis, especially helpful in gathering the information required to manage traffic lights systems and sending routing optimisation information to interested users;
2. mobile sensors for the collection of environmental data. Such data will be used to feed decision-making models for the reduction of carbon emissions and the consequent improvement of air quality in the urban area.
3. Mobile devices for the acquisition of the motion habits of people.

The integration of environmental models and smart systems for the management of urban mobility will allow to optimise public and private traffic flows as well as to reduce carbon emissions.

The main innovation brought by CAGLIARI 2020 is related to the application of the “netcentric” paradigm by means of a dynamic and pervasive net (the urban information grid) whose nodes can be both fixed and mobile. This feature allows the sensorial integration of the devices distributed in the urban area and turns public transport buses into “mobile platforms” for the urban road system monitoring thanks to the continuous gathering of traffic, carbon emissions and noise pollution data. It is therefore possible to develop models for the analysis of environmental parameters and to provide support tools to policies aimed at curbing traffic flows, energy consumptions, and carbon emissions within urban areas.

The integration between the aforementioned information and the people’s travelling habits (by means of the anonymous tracking of their mobile phones) allows for the creation of people’s mobility maps.

Moreover, the project intends to spur and help the growth of new multi-sectorial entrepreneurial realities operating in the fields of mobility management and energy consumption. Cloud services will play a key role within the project in supporting the applications dedicated to data traffic monitoring and analysis.

Summary:

CAGLIARI 2020 is a 25 million euro project funded within the framework of the National Operational Program for Research and competitiveness of the Italian Ministry of Education University and Research. Its main goal is the development of innovative and environmentally friendly solutions for urban mobility (and possibly metropolitan area mobility), so to boost energy and environmental performances. The project idea originates from the ever increasing need of innovative tools and technological solutions for the optimisation of urban mobility, for lowering travel times and improving air quality. Cloud services will play a key role within the project in supporting the applications dedicated to data traffic monitoring and analysis.

Supercomputing, High Throughput, Accelerator Technologies and Integration / 12

A Novel Architecture towards Exascale Computing

Author: Shaun de Witt¹

Co-authors: David Bond²; Keeran Brabazon³; Oliver Perks³; Sai Narasimhamurthy⁴; Salem El Sayed Mohamed⁵; Sergio Rivas Gómez⁶; Stefano Markidis⁶

¹ *Culham Centre for Fusion Energy*

² *Diamond Light Source*

³ *Allinea Software Ltd.*

⁴ *Seagate Technology*

⁵ *Jülich Supercomputing Centre*

⁶ *KTH Royal Institute of Technology*

Corresponding Author: shaun.de-witt@ukaea.uk

As data volumes grow rapidly in the science domain, the ability to process this data efficiently is becoming increasingly of interest. While in many applications the processing of very large volumes can be accomplished efficiently with map/reduce algorithms (e.g., using frameworks such as Hadoop),

this does not cover a large class of problems which are best run in an HPC environment. This class of problem requires a new paradigm commonly referred to as Big Data and Extreme Computing (BDEC).

This talk explains why current HPC systems are not fully suited to BDEC class problems, and how some upcoming technologies such as advanced non-volatile memories could help to resolve some of these issues. We introduce a novel architecture that aims to address this problem, currently being developed within the SAGE project (sagestorage.eu) as part of the Horizon2020 program. We explain how this architecture has been co-designed between leading industrial partners and research organisations covering a wide spectrum of scientific disciplines, and outline new programming models that are being developed to make use of this new technology, including tools to optimise its use. Finally, we present details of testing of this new architecture and explain where such a system can overcome the limitations of traditional HPC systems.

Network, Security, Infrastructure & Operations IV / 14

The EGI CernVM-FS infrastructure - latest developments and evolution towards a global facility

Author: Catalin Condurache¹

¹ *STFC Rutherford Appleton Laboratory*

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

The CernVM-FS is firmly established as a method of software and condition data distribution for the LHC experiments at WLCG sites. Use of CernVM-FS outside WLCG has been growing steadily and an increasing number of Virtual Organizations (VOs), both within the High Energy Physics (HEP) and in other communities (i.e. Space, Natural and Life Sciences), have identified this technology as a more efficient way of maintaining and accessing software across Grid and Cloud computing environments.

Following initial success of a CernVM-FS service offered to small VOs in the UK, RAL Tier-1 enlarged it and an EGI CernVM-FS infrastructure has been developed since September 2013. In this paper we describe the work carried out at RAL to expand the infrastructure to provide a resilient, distributed CernVM-FS service to non-LHC VOs across Europe and replicated around the world. We focus on the current status of its main elements: the Master Repository (Stratum-0), the Replica/Mirror (Stratum-1) and the customised mechanism to upload and maintain the master repositories by the VO Software Grid Managers.

The latest developments to widen and consolidate the CernVM-FS infrastructure as a global facility (with main contributors in Europe, North America, Asia) are reviewed, such as the mechanism implemented to publish external repositories hosted by emerging regional infrastructures (eg. South Africa Grid). Progress on enabling the 'squid auto discovery' mechanism at CernVM-FS client level (which is a specific demand from the communities using the EGI Federated Cloud resources) is described alongside the implementation of protected CernVM-FS repositories, a requirement for academic communities willing to use CernVM-FS technology.

Summary:

The EGI CernVM-FS infrastructure - latest developments and evolution towards a global facility

Infrastructure Clouds and Virtualisation II / 15

VCondor - an implementation of dynamic virtual computing cluster

Author: Yaodong CHENG¹

¹ *IHEP, CAS*

Corresponding Author: chyd@ihep.ac.cn

As a new approach to manage resource, virtualization technology is more and more widely applied in high energy physics field. We have built virtual computing cluster at IHEP based on Openstack, with HTCondor as the job management system. In traditional computing cluster, fixed number of slots are pre-allocated to the job queue of different experiments. However, this kind of policy has gradually become dissatisfy with the peak requirements of different experiments, and also leads to a low CPU utilization. To solve the problem, we designed and implemented a dynamic virtual computing cluster system - VCondor based on HTCondor and Openstack. This system performs unified management of virtual machines according with queue status in HTCondor. One or more VMs will be created automatically when some jobs are waiting to run. VM will be destroyed when job is finished and there is no more job in HTCondor queue. Job queue status is checked in a period of time such as 10 minutes, so a VM will continue to run if there are new jobs in the period of time. VCondor also support resource provision and reservation for different experiments. VCondor has to request and get the available number of VM from a VM resource scheduling system called VMQuota before it create VMs. VMQuota tells how many VMs VCondor can create and how long these VMs will be reserved before they are created.

This talk will present several use cases of LHAASO and JUNO experiments. The results show virtual computing cluster can dynamically expanded or shrunk while computing requirements changed. Additionally, CPU utilization of overall computing resource is significantly improved compared with traditional resource management system. The system also has good performance when there are multiple condor schedulers and multiple job queues. It is stable and easy to maintain as well.

Network, Security, Infrastructure & Operations I / 16

WLCG Security Operations Centres Working Group

Authors: David Crooks¹; Liviu Valsan²

¹ *University of Glasgow*

² *CERN*

Corresponding Author: liviu.valsan@cern.ch

Security monitoring is an area of considerable interest for sites in the Worldwide LHC Computing Grid (WLCG), particularly as we move as a community towards the use of a growing range of computing models and facilities. There is an increasingly large set of tools available for these purposes, many of which work in concert and use concepts drawn from the use of analytics for Big Data. The integration of these tools into what is commonly called a Security Operations Centre (SOC), however, can be a complex task - the open source project Apache Metron (which at the time of writing is in incubator stage and is an evolution of the earlier OpenSOC project) is a popular example of one such integration. At the same time, the necessary scope and rollout of such tools can vary widely for sites of different sizes and topologies. Nevertheless, the use of such platforms could be critical for security in modern Grid and Cloud sites across all scientific disciplines.

In parallel, the use and need for threat intelligence sharing is at a key stage. Grid and Cloud security is a global endeavour - modern threats can affect the entire community, and trust between sites is of utmost importance. Threat intelligence sharing platforms are a vital component to building this trust as well as propagating useful threat data. The MISP software (Malware Information Sharing Platform) is a very popular and flexible tool for this purpose, in use at a wide range of facilities in different domains across the world.

In this context we present the work of the WLCG Security Operations Centres Working Group, which was created to coordinate activities in these areas across the WLCG. The mandate of this group includes the development of a scalable SOC reference design applicable for a range of sites by examining current and prospective SOC projects & tools. In particular we report on the first

work on the deployment of MISP and the Bro Intrusion Detection System at a number of WLCG sites, including areas of integration between these tools. We also report on our future roadmap and framework, which includes the Apache Metron project.

Summary:

We present the work of the WLCG Security Operations Centres Working Group, which was created to coordinate work on analytic security platforms across the WLCG (a more detailed description being in the abstract). The mandate of this group includes the development of a scalable SOC reference design applicable for a range of sites by examining current and prospective SOC projects & tools. In particular we report on the first work on the deployment of MISP and the Bro Intrusion Detection System at a number of WLCG sites, including areas of integration between these tools. We also report on our future roadmap and framework, which includes the Apache Metron project.

Humanities, Arts & Social Sciences II / 17

Using Advanced e-Systems for Community-Engaged Research

Author: David J. Bodenhamer¹

¹ *Indiana University-Purdue University*

Corresponding Author: intu100@iupui.edu

Community Information Systems is a generic term that describes a wide range of methods to organize, manage, and disseminate community data for the purpose of increasing the capacity of citizens and organizations to participate effectively in decision-making. CIS is a growing phenomenon in the U.S., beginning in 2005 with a Brookings Institution report, National Infrastructure for Community Statistics. A 2011 U.S. General Accountability Office study revealed that over 35 American cities has a form of CIS, with many of them part of an emerging consortium, the National neighborhood Indicators Partnership.

The SAVI Community Information System for Central Indiana (SAVI), developed and managed by the Polis Center at Indiana University Purdue University, Indianapolis, is the nation's largest and most comprehensive CIS. In existence since 1995, SAVI brings together over 35 data providers into a rich spatially enabled, web-based environment that allows citizens and researchers to understand a wide range of social issues at more than a dozen geographic scales, from census blocks and neighborhoods to counties and metropolitan areas. Now it is transitioning to a community intelligence system, with information integrated at both individual and geographic units and with enhanced reporting capabilities and embedded predictive analytics. This system also has been selected to serve an innovative new Indiana University initiative to link scientific and social science data for the purpose of helping the state's citizens to respond more effectively to 21st Century changes.

This presentation will discuss both the technical infrastructure and research potential of SAVI, especially in the rapidly emerging area of community-engaged research. It also will outline how SAVI is part of an emerging system of systems that uses advanced computing to link a variety of administrative and human service databases for community advancement.

Earth, Environmental Science & Biodiversity I / 18

Revealing Philippine Climate Type Using Remotely-sensed Rainfall Estimates

Author: Jay Samuel Combinido¹

Co-author: John Robert Mendoza¹

¹ *Advanced Science and Technology Institute*

Corresponding Author: jaysamuel@asti.dost.gov.ph

Rainfall variability is a key feature of climate. Understanding rainfall distribution over space and time is particularly of interest because it impacts several aspects of human activities. It has been known that rainfall variations in the Philippines is influenced by an interplay of various synoptic systems affecting the country – southwest and northeast monsoons, easterlies and tropical cyclone activity. In a 1920 report, Coronas was able to identify and map four distinct climate types based on monthly rainfall time series obtained from synoptic stations over the Philippines. Using the same source of data, subsequent studies by various authors demonstrated similar results.

A common challenge in utilizing ground observations for climate analysis, especially when dealing with rainfall, is the spatial resolution of the ground station network. Taking into consideration the sparsity of the stations, the geographical delineation between climate types must have been determined by the archipelago's topographical features such as mountain ranges blocking passages of rain-producing systems (rain shadow) and other factors known to the authors. Satellite-based technology overcomes this limitation.

In this work, we demonstrate the viability of satellite-based rainfall estimates in capturing rainfall variations in the Philippines. Performing K-means clustering algorithm onto Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis (TRMM/TMPA) 3B43 rainfall data revealed four climate types with unique characteristics: (Type 1) two pronounced season with peak rainfall occurring during JJA; (Type 2) no dry seasons with extreme rainfall during DJF season; (Type 3) relatively dry from January to April with rainfall amount gradually increasing until December; and (Type 4) evenly distributed rainfall throughout the year. These rainfall patterns can be explained by southwest monsoon, easterlies, tropical cyclone visits, position of the inter-tropical convergence zone and topography. It is also interesting that the spatial extent of each climate type naturally manifested during the classification process and that the effect of topography was inferred from the results.

Network, Security, Infrastructure & Operations III / 19

IPv6 Deployment and Migration of WLCG TEIR-2 site resources on Private Cloud

Author: Saqib Haleem¹

¹ *National Centre for Physics, Islamabad, Pakistan*

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) and local scientific cluster. 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 in this region, 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 TEIR-2 site is also being tested on IPV6. In order to optimize the utilization of computing resources, open stack based private Cloud has also been deployed. All the computing resources are now being managed through that cloud. This paper discusses the details of IPv6 deployment status and migration status of TEIR-2 site on private cloud.

Physics & Engineering I / 20

The High Throughput Strategy of IHEP

Author: Jiaheng Zou¹

Co-authors: Jingyan Shi²; Ran DU²; Xiaowei JIANG²; Zhenyu SUN²

¹ *IHEP, Chinese Academy of Sciences*

² *IHEP*

Corresponding Author: zoujh@ihep.ac.cn

IHEP computing center serves for many high energy physics experiments. We have more than 13,000 CPU cores and hundreds of active users. There are tens of thousands jobs per day. We divide users into many groups classically according to which experiment they belong to. And each computing node is privately owned by one group. The peak requirements of different groups are not coincident in general. Then there can be a great waste without resource sharing between groups.

It makes good sense to improve the resource utilization and the jobs throughput. We managed to deploy a high throughput system, which considers both the sharing of resources and the fairness between groups. The system is based on HTCondor. However, it is necessary to customize a strategy to manage user groups in a single cluster pool. We keep a number of their own nodes for each group considering fairness. This ensures that there are resources available for each group forever. Meanwhile, a ratio of resources have to be shared between all groups. So it is possible for busy groups to take benefits from free groups. This is important to increase the entire jobs throughput. We provide real time statistics and try to ask free groups for more sharing resources. The sharing ratio of each group can be tuned automatically with owners' approval. We are developing an accounting system, which will provide statistical details of free groups' contribution and busy groups' occupation. An error recovery mechanism is provided and integrated with the cluster monitoring system. Nodes with fatal problems are removed from the pool automatically. We also developed a set of toolkits to users. The toolkits add a series of attributes to users' jobs, which are necessary in our approach.

The entire strategy needs an enhanced central control system of HTCondor. We have implemented the essential components for central control, and deployed the customized strategy. The result shows great effects to our high throughput computing management.

Physics & Engineering I / 21

Examination of dynamic partitioning for multi-core jobs in the Tokyo Tier-2 center

Author: Tomoe Kishimoto¹

Co-authors: Hiroshi Sakamoto¹; Nagataka Matsui¹; Tetsuro Mashimo¹; Tomoaki Nakamura²

¹ *The University of Tokyo*

² *KEK*

Corresponding Author: tomoe@icepp.s.u-tokyo.ac.jp

The Tokyo Tier-2 site, which is located in International Center for Elementary Particle Physics (ICEPP) at the University of Tokyo, is providing computer resources for the ATLAS experiment in the Worldwide LHC Computing Grid (WLCG). The official site operation in the WLCG was started in 2007 after the several years development since 2002, and the site has been achieving a stable operation since then. In the current system, which was upgraded in 2016, 6144 CPU cores have been deployed as worker nodes for the WLCG, where each worker node consists of 24 CPU cores.

The ATLAS experiment developed the multi-core implementation of their software framework for reconstruction and simulation jobs, which provides an efficient memory sharing. In 2014, the experiment started to submit eight-core jobs using the above software framework to the Grid. The Tokyo Tier2 site has been processing this multi-core job and normal single-core job (e.g. user analysis job) separately using dedicated worker nodes and computing elements (static partitioning). However, we have often observed idle CPUs of the worker nodes due to this static partitioning when either multi-core or single-core jobs are not assigned to the site. Therefore, we started to evaluate an implementation of the dynamic partitioning of the worker nodes using HTCondor batch scheduler to

reduce this idle CPU time, and have deployed a small cluster (1536 CPU cores) in the production. For the dynamic partitioning, draining of single-core jobs is necessary in order to dispatch a new multi-core job into a worker node when the worker node is filled by only single-core jobs. This draining should be performed until the number of running multi-core jobs reaches a target share. In order to perform an efficient draining, we need to consider several parameters, such as the number of draining machines at the same time, based on properties of the jobs. In this presentation, improvement of CPU efficiencies by introducing the dynamic partitioning and optimization of the drain parameters in the Tokyo Tier-2 center will be reported.

Network, Security, Infrastructure & Operations I / 22

Collaborating for WISer Information Security

Authors: Adam Slagell¹; Alf Moens²; David Kelsey³; Hannah Short⁴; Jim Marsteller⁵; Ralph Niederberger⁶; Urpo Kaila⁷

¹ *NCSA/XSEDE*

² *SURF*

³ *STFC-RAL*

⁴ *CERN*

⁵ *PSC*

⁶ *HBP/PRACE/EUDAT*

⁷ *CSC/EUDAT*

Corresponding Author: hannah.short@cern.ch

As most are fully aware, cybersecurity attacks are an ever-growing problem as larger parts of our lives take place on-line. Distributed digital infrastructures are no exception and action has to be taken to both reduce the security risk and to handle security incidents when they inevitably happen. These activities are carried out by the various e-Infrastructures and it has become very clear in recent years that collaboration with others both helps to improve the security and to work more efficiently.

The WISE (Wise Information Security for collaborating E-infrastructures) community was born as the result of a workshop in October 2015, which was jointly organised by the GÉANT group SIG-ISM (Special Interest Group on Information Security Management) and SCI, the ‘Security for Collaboration among Infrastructures’ group of staff from several large-scale distributed computing infrastructures. All agreed at the workshop that collaboration and trust is the key to successful information security in the world of federated digital infrastructures for research.

WISE provides a trusted global framework where security experts can share information on topics such as risk management, experiences about certification processes and threat intelligence. With participants from e-Infrastructures such as EGI, EUDAT, PRACE, XSEDE, NRENs and more, WISE focuses on standards, guidelines and practices, and promotes the protection of critical infrastructure.

This talk will focus on ongoing work in the WISE Working Groups, each tackling different aspects of collaborative security and trust: Security Training and Awareness; Risk Assessment; Review and Audit; Security in Big and Open Data; and SCI version 2. This final Working Group aims to produce a whitepaper detailing the community’s requirements. We will report on progress made and highlight the expected challenges to be faced by participating E-infrastructures in the coming years.

Details on the WISE community and our Working Groups can be found at <https://wise-community.org>

Network, Security, Infrastructure & Operations II / 23

Status of Network Security Operations at IHEP

Authors: Fazhi QI¹; Tian Yan²

Co-authors: Dehai An²; Lanxin Ma²

¹ *Institute of High Energy Physics, CAS*

² *Institute of High Energy Physics, CAS, China*

Corresponding Author: yant@ihep.ac.cn

Institute of High Energy Physics (IHEP) is an institute of Chinese Academy of Sciences which explore elementary particle physics. The network of IHEP campus and data center connects about 1200 servers and 3000 PC clients. It supports IPv4 and IPv6 protocols, both of them has 10 Gbps internet access.

This report gives a brief introduction to the status of network security operations at IHEP. Firstly, an architecture overview is presented. Then four subsystems are described in detail, including a monitoring and early warning system, an intranet proactive defense system based on user behavior, a network security self-service platform and a cyber-security federation for HEP community in China.

A network security monitoring and early warning system was developed and deployed based on a 10 Gbps network probe. By monitoring the network traffic data and analyzing the log of network devices and servers, it can find out the malicious IP addresses and then block the attack by associated firewall. The affected hosts in local network can be located too. A warning system is integrated and it will send short messages to administrators when high risk attack appears.

An intranet proactive defense system was designed as a supplement of traditional strategy of network boundary protection. It's based on user behavior analysis with help of artificial neural network. Suspicious user behavior can be detected and visualized, as well as early warning and even proactive attack block.

In order to help users getting rid of their host vulnerability, we developed and deployed a network security self-service platform. This system can present straightforward result of quantized fuzzy evaluation of host security risk. The result is obtained by analytic hierarchy process and cloud model theory. We improved the multi-level index system of the analytic process by dynamic weight method, which increased the adaptability and objectivity of the index system.

In June 2016, we proposed a China cyber-security federation for high energy physics (CSFHPEP), about 10 universities and institutes have joined in this federation. The constitution of CSFHPEP is completed and the federation is now in test run. In framework of CSFHPEP, a cooperative security response center and a cyber-security study group are founded. All the members will benefit from this federation on security incident response support, threat information sharing, training on secure operations and other related services.

24

Design and Implementation of Unified Authentication Management System of IHEP

Authors: Fazhi QI¹; Li Wang²; Zhen Xu²

¹ *Institute of High Energy Physics, CAS*

² *Institute of High Energy Physics, CAS, China*

Corresponding Author: wangli320@ihep.ac.cn

High energy physics network computing environment has the characteristics of complex and heterogeneous, and the projects researched by high energy physics are mostly carried out in the way

of international cooperation. These factors increase the complexity of the management infrastructure and application services system. Therefore, reasonable and efficient management of resources, taking into account the scientific application service flexibility requirements, has become the core issue of protecting information and resources. The traditional application and service authentication method is to save the user information in each application system and establishes an independent authentication mechanisms with independent authentication in their respective identity authentication modules. This traditional method of identity authentication makes the user's security identity fragmentation when using the application. These fragmented identity fragments consist of one to one customer service relationship in all the authentication modules, which implies a huge security risk. From the perspective of the user, each access to a system to verify the login once when the user needs frequent access to different systems, which will lead to the user needs to remember a large number of account information, while causing unnecessary time consumption.

The Institute of high energy physics, has all kinds of information system and service resources, such as personnel information management system, data sharing service, performance management system, conference system, and each user has an account in these systems. The traditional management mode is configured different permissions for each account on each system. It is easy to be wrong because the system is too much, for example users are assigned permissions without modifying high timely which eventually lead to the information divulgence and the system safety risk increasing. At the same time each user needs to remember their own user name and password on each system that is not only more troublesome but also easy to appear lost password.

It then proposes the unified authentication solution under the IHEP network environment. This solution uses LDAP directory services for storing user information, builds a unified authentication mechanism by using OAuth2.0 protocol. This approach eventually combines a variety of applications and enables the secure access to these applications.

Physics & Engineering I / 25

The Billing System of IHEP Data Center

Authors: Ge Ou¹; Hongmei Zhang¹; Li Wang¹; Wenshuai Wang¹

¹ *Institute of High Energy Physics, CAS, China*

IHEP manages a number of China's major scientific facilities, including BEPC, BES, BSRF, HXMT, ADS, JUNO, CSNS, CEPC, the International Cosmic-Ray Observatory at Yangbajing in Tibet, the Daya Bay Neutrino Experiment etc. Data generated by these facilities is processed in IHEP data center.

There are many computing resources in IHEP data center, including cloud computing resources, local cluster resources, distributed computing resources, GPU resources, etc. But sometimes the utilization rate is imbalance because some resources belong to and can only be used in an experiment. So we develop the billing system to promoting resource sharing.

The billing system is a system for renting some kinds of computing resources of the data center and implementing integrated billing for these resources. Users can rent any computing resources by this system, and they can check real-time bill too. The managers of experiment resources can also check the resource usage and the income.

The billing system is designed and implemented based on B/S structure and MVC pattern, using JAVA programming technology. It mainly covers three function modules: product & service, user center and system management.

(1) It lists all the computing resources and the accounting rules in product & service module. Users can login the billing system use the account of IHEP unified authentication, they can rent the resources according to their needs, and they can apply to use free resources too.

(2) In user center, users can view details of resources they had rented, and they can check the real-time bill too.

(3) Management platform provides services for managers. They can manage and publish computing resources product, they can redeploy resources pool, and they can manage billing information too.

Now the billing system mainly manages computing resources, in future we will extend modules for private cloud storage, database resources, etc.

Earth, Environmental Science & Biodiversity II / 26

NeIC EISCAT_3D support project: Nordic computing challenge

Author: John White¹

¹ *NeIC*

EISCAT_3D will establish a system of distributed phased array radars that will enable comprehensive three-dimensional vector observations of the atmosphere and ionosphere above Northern Fennoscandinavia.

The use of new radar technology, combined with the latest digital signal processing, will achieve ten times higher temporal and spatial resolution than obtained by present radars while simultaneously offering, for the first time, continuous measurement capabilities.

The flexibility of the

EISCAT_3D system will allow the study of atmospheric phenomena at both large and small scales unreachable by the present systems.

The EISCAT_3D system will, in its first stage, consist of three radar sites: one with both transmitting (TX) and receiving (RX) capabilities and two with only RX capabilities.

The sites will be located in remote locations in three different countries (Finland, Norway and Sweden)

and will be separated geographically by approximately 130 km.

Two additional receive sites, at distances 200-250 km from the transmit site, are planned for the full EISCAT_3D system.

In addition to the radar sites, EISCAT_3D will also have an operations centre and one or more data centres.

The NeIC EISCAT_3D support (E3DS) project aids the future EISCAT_3D project in planning and tendering their required e-infrastructure.

This includes the gathering of the EISCAT_3D use-cases and transforming these into a set of standard

requirements for the various components of the overall EISCAT_3D computing e-infrastructure.

The NeIC EISCAT_3D support project interacts with EISCAT_3D and Grid and Cloud e-infrastructure projects.

This interaction is needed to match the expertise in EISCAT_3D with corresponding expertise in the existing e-infrastructure projects in the various fields.

In effect, the E3DS project builds the collaborations among EISCAT_3D, national e-infrastructure providers

and network providers.

The NeIC EISCAT_3D support project contributes by making it possible to expand and enhance the usage of existing Nordic e-infrastructures.

By introducing a new field of research to the existing Nordic e-infrastructure the overall capabilities will be extended.

Adding another large field of research into the existing e-infrastructures can spread the load of resource

and cost-sharing.

In return the EISCAT_3D project will benefit from already supported and maintained e-infrastructures for

computing and storage.

Ideally, the EISCAT_3D project will be able to use the e-infrastructure transparently and focus on atmospheric science.

Data Management & Big Data / 27

Towards a cloud-based computing and analysis framework to process environmental science big data

Authors: Eleonora Luppi¹; Luca Tomassetti¹; Sebastiano Fabio Schifano¹

¹ *University of Ferrara and INFN*

Corresponding Authors: luca.tomassetti@unife.it, eleonora.luppi@unife.it, sebastiano.fabio.schifano@unife.it

Environmental sciences are quickly and increasingly adopting research methodologies based on data coming from satellites, large network of sensors installed on the ground or sea-floating stations, as well as from devices installed on balloons or aircraft. These networks produce a big amount of data that needs to be appropriately processed and analyzed to extract information useful for scientists to investigate natural phenomena.

This requires for example, the capacity to collect and store huge amount of data together with space and time information and the availability of large and powerful computing resources to run analysis and visualization codes. However, for environmental scientific community to develop or to have in house the infrastructures to accomplish and support efficiently all these steps can not be convenient, because they require a non negligible effort for maintenance.

On the other hand, in the last decades, scientific communities of other scientific fields, such as high-energy physics, have developed and acquired a large experience in computer grid infrastructures to store and analyze big amount of data coming from particle accelerator laboratories experiments. Computer grids have recently evolved into clouds which are more user-friendly allowing access to new and more heterogeneous communities with less computing expertise [1,2].

In this contribution we discuss how to bridge the experience acquired in using the grids for high energy physics experiments and the needs of environmental sciences. In particular, we apply this strategies in the context of the interdisciplinary EU-ERASMUS+ TORUS project which includes Europe's and south east Asia's partners with a strong expertise in distributed and cloud computing and earth and environmental sciences.

The TORUS project aims to make soon available to environmental scientists a cloud based computing and analysis framework to manage and process big-data.

This includes the ability to access clouds to virtualize the computing resources, and knowledges to use software tools to process and analyze data coming from the different data sources. We also describe how to store data together with meta-data information related to time and space, and how to present data at high-level that can be easily used and interpreted by user scientists.

Finally, we also discuss how to integrate into this framework high-performance computing to boost for example satellite image processing, that for the intrinsic computational complexity may require the use of recently developed accelerators like GP-GPUs or many-core processors [3,4].

References

[1] Fella, A., Luppi, E., Manzali, M., Tomassetti, L., A general purpose suite for Grid resources exploitation (2012) IEEE Nuclear Science Symposium Conference Record, art. no. 6154459, pp. 99-103.

[2] Roiser, S., et al., The LHCb Distributed computing model and operations during LHC runs 1, 2 and 3 (2015) Proceedings of Science, art. no. 005.

[3] Calore, E., et al., Massively parallel lattice Boltzmann codes on large GPU clusters, Parallel Computing 58 (2016), pp. 1-24.

[4] Adinetz, A. V., et al., Performance evaluation of scientific applications on POWER8, Lecture Notes in Computer Science 8966 (2015), pp. 24–45.

Infrastructure Clouds and Virtualisation II / 28

GUOCCI – The Entryway to Federated Cloud for Small-scale Users

Author: Radim Janča¹

Co-authors: Dusan Baran¹; Zdeněk Šustr¹

¹ CESNET

Corresponding Author: janca@ics.muni.cz

The Open Cloud Computing Interface (OCCI) – a standard released by the Open Grid Forum – has found wide adoption. First came, quite naturally, server-side components. Later, with a variety of attractive computing resources made available over the standardized protocol, user-side submission tools or science gateways followed. These are usually tailored with a specific use case in mind, and – as such – they are generally capable of setting up heterogeneous data processing platforms and orchestrating complex workflows specific to their given area of science. They put all this functionality at user's disposal, often “at a single click”. On the other hand, small-scale users, backed by no software development teams, are usually relegated to the use of the simplest OCCI client in the command line. That is freely available and amply documented, but still, it by no means lowers the threshold of entry for potential federated cloud users, who often come from non-technical areas in the long tail of science. Therefore, the GUOCCI (GUI for OCCI) has been conceived as a rudimentary graphical user interface to OCCI-compliant cloud services. It is by design kept as simple as possible to address the needs of small-scale or one-off users who typically require little dynamism in their virtual resources, and who are perfectly happy to set up their virtual resources by hand, even one-by-one, especially if they are offered

a comprehensive graphical interface to do it. GUOCCI integrates not only with OCCI-compliant cloud sites, but also with the EGI Application Database and with authentication technologies used in academic federated clouds, namely with Virtual Organization Membership Services (VOMS). With that, the considerable resources available for instance in the EGI Federated Cloud are open up to all such small-scale or beginning users. This article introduces in greater depth the reasoning behind developing GUOCCI, and details the architecture of the product, making it into an example OCCI client implementation.

Humanities, Arts & Social Sciences I / 29

Platform for Humanities Open Data

Author: HARA Shoichiro¹

Co-author: KAMEDA Akihiro¹

¹ Center for Integrated Area Studies, Kyoto University

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

Construction of humanities databases have difficulties due to some reasons. First, construction process requires expert knowledge and techniques of database systems, which impedes database construction by humanities researchers. Second reason is diversity of resource media, which enriches the humanities researches but is an obstacle to metadata standardization, and brings about heterogeneous databases. Third reason is this heterogeneity of metadata, which makes sharing data difficult.

Center for Integrated Area Studies, Kyoto University (CIAS) has developed two information tools, named MyDatabase (MyDB) and Resource Sharing System (RSS), to solve these difficulties. The main component of MyDB is a database builder, allowing humanities researchers to construct and revise databases without expert knowledge. MyDB stores metadata and accepts any vocabulary of metadata, including nonstandard one. This enables humanities researchers to use their own metadata vocabulary according to their own purpose. On the other hand, those metadata varieties make the integration processes difficult. RSS is developed to integrate heterogeneous databases on the Internet and to provide users with a uniform interface to retrieve databases seamlessly in one operation. Thus, MyDB and RSS have contributed to accelerate humanities open data, but there are still two problems to solve, especially of RSS: small coverage of databases and initial costs of integration. First, for example, Kyoto University releases KULINE (OPAC), KURENAI (repository), KURRA (archive), Open Course Ware and various databases developed by each research institute in the university, but RSS does not integrate these databases. Second, it is time consuming to integrate new databases into RSS and impossible to trace links automatically, that is, for now, RSS is not the appropriate tool to discover hints and/or create new knowledge.

To overcome these drawbacks, a new project has been launched to develop an innovative information platform for open humanities data. This platform comprises three sublayers. The first layer is “Open Data Layer” which accumulates heterogeneous metadata. This layer uses RDF to describe data of different structures. The second layer is “Data Link Layer.” This layer uses ontology techniques such as RDFS and OWL to link ambiguous (uncontrolled) vocabularies and emerge “humanities big data.” The third layer is “Application Layer.” As humanities big data is too huge and complicated to retrieve, categorize, and analyze by hands. This layer provides utilities to process big data. This platform will prepare for APIs to help mashup applications. We expect the platform to reconstruct knowledgebase from heterogeneous databases, which is used to construct meaningful chunks from scattered data.

As a pilot study to determine the validity of the platform, we prepared a dataset of “Japanese Journal of Southeast Asian Studies” as a core dataset (the first layer) for trying to link words or documents in the core dataset to external resources such as KURENAI or DBpedia (the first layer also). Then, the relationship between heterogeneous internal and external databases will be described in the second layer, so that whole data is structured for clean API and exploits that data in an annotated paper viewer (the third layer).

30

Monitoring Virtual Devices in Mass Storage Environments

Author: Tim Chou¹

¹ *Brookhaven National Laboratory*

Corresponding Author: tchou@bnl.gov

The Scientific Data and Computing Center (SDCC) at Brookhaven National Laboratory (BNL) is a major scientific data storage site with more than 90 PB of archived experimental data, which amount is supposed to grow exponentially during the next several years in connection with the newly emerging enormous storage needs of ATLAS – one of the four particle physics experiments based at the Large Hadron Collider (LHC), CERN, Geneva. The Large Hadron Collider (LHC) is the biggest particle accelerator/collider in the world and BNL is a Tier 1 data center for the ATLAS collaboration.

BNL’s data storage model is based on the use of the High Performance Storage System (HPSS) which is a flexible, scalable, policy-based Hierarchical Storage Manager, developed as the result of a collaboration among five Department of Energy laboratories and IBM.

The challenge is to meet the data storage and processing requirements of the Relativistic Heavy Ion Collider (RHIC) experiments (at BNL) and especially the newly emerging enormous storage needs of the ATLAS experiment.

In the above context the presentation will review BNL's data storage complex design and architecture with a special emphasis and focus on proprietary optimization and monitoring software.

The developed tools use gathered statistical information to provide adaptive methods and techniques for device management and rigorous performance tuning and optimization of the entire storage environment. The software suite provides an intuitive user-friendly web-based interface enabling dynamic parameter visualization based on a priority-driven device-monitoring paradigm and addresses the following major technical issues:

- maintain the health of the deep storage archive
- identify which data volumes are degrading and move the valuable data to new volumes
- predict device failures before they actually happen and service the devices
- identify the backlogs and congestions to optimize the bandwidth of deep storage systems
- balance the usage of deep storage devices and media
- detect and protect storage devices and media from unauthorized access
- enable encryption feature on encryption capable devices

Summary:

Monitoring Virtual Devices in Deep Storage Environments

Data Management & Big Data / 31

Data storage accounting at RAL

Author: Rob Appleyard¹

Co-authors: Alison Packer¹; Bruno Canning¹; Cheney Ketley¹; Jens Jensen¹; John Gordon¹

¹ STFC

Corresponding Author: rob.appleyard@stfc.ac.uk

Accounting for storage available and used is important to many communities using grids (and clouds). Several large grids came together to define the GLUE schema (Grid Laboratory Uniform Environment) in order to promote the interoperation and cross-grid use of the infrastructures. Ensuring consistency (across implementations and grids) and usefulness of the information published by these grids was a significant task for every storage system, and the consistency exercise had to be repeated for the second version of GLUE, GLUE2. Creating the schema alone is not sufficient, so collaborations defined their "interpretations" of the schemas and guidelines for publishing metadata.

The present paper describes the work involved in developing and deploying such an information system for the CASTOR and CEPH systems at the RAL Tier 1. In part 1 of the paper, we outline some of the important design decisions taken throughout the development process and focus on how we obtain the required information from several disparate parts of the systems, and describe the difficulties associated with accounting when files can be compressed on tape, have ephemeral copies on disk, filesystem vs object store, etc. We finish part 1 with an outlook towards dynamic ("cloudy") use of storage resources. In part 2, we look at how to format the information to become standards-compliant and future-proof.

Running WLCG and HPC Workload on Single Openstack Cloud

Author: Manzoor Ahmad¹

¹ *National Centre for Physics, Islamabad*

Corresponding Author: manzoor.ahmad@ncp.edu.pk

Our world is changing so fast. We have to adopt this change and act according to the new technologies quickly. Many of the biggest world recognized companies are no longer in the business just because they did not act quickly to adopt the new changing in technologies. Nokia is the example of this. Today, cloud computing has revolutionized Information Technology field. It is the emerging concept and change that every Data Center will have to adopt in near future or even today.

HPC and WLCG applications are widely used in scientific and industrial research, data analytic and visualization, social behavioral studies etc. Most HPC applications and WLCG setup require dedicated, available and highly customized resources and environments for computation since they exhibit intense resource utilization. As stated above, Cloud computing is emerging as a latest computing technology. The on-demand nature of cloud has provoked interest to explore if cloud properties can be useful for HPC and WLCG setups. This paper is a work around in that direction.

Currently, user community of our 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. Whereas, our WLCG setup has 524 CPU Cores and 441 TB of raw Storage. DPM based WLCG Tier-2 site has currently 59 worker nodes dedicated machines. The main innovation towards running HPC and WLCG workload on cloud was to better utilization of IT resources by running WLCG worker nodes on cloud, as well as to test the behavior of jobs being run on cluster nodes and WLCG Worker nodes simultaneously. Private cloud was setup by using Openstack latest release Mitaka. Initially, three Octa Core Sun Servers, each having 16GB of RAM and 160 GB of local storage were installed on CentOS 7. One node was made as controller node while other two nodes were made as compute node. Different subnets were used for cluster and WLCG instances to avoid network latency. Finally, Rocks Cluster and WLCG worker nodes instances were created and added in our production environment and real jobs were submitted, ran and monitored. We observed that jobs are being run smoothly on cluster instances as well as on our WLCG worker nodes. We also observed that running HPC and WLCG workload on single cloud was more easy, scalable and flexible than running these workload on physical hosts because we could add new nodes in both HPC and WLCG very easily and allocate hardware resources according to the job requirements.

Network, Security, Infrastructure & Operations III / 33

TRANSITION FROM IPv4 TO IPv6 IN MONGOLIA: KEY ISSUES AND RECOMMENDATIONS

Author: TumeUlzii Naranmandakh¹

¹ *Communications Regulatory Commission, Mongolia*

Corresponding Author: naran@crc.gov.mn

- List item

- 1.Introduction
 - 1.1.ICT Sector policy and Regulations
 - 1.2.Results of IPv6 Survey
 - 2.IPv6 deployment,challenges and current initiatives
 - 2.1.1.Policy and regulation

- 2.1.2.IPv6 deployment for Industry and Business
- 2.1.3.Security Considerations in IPv4 to IPv6 Migration
3. Mongolia Industry Case Studies
4. IPv6 Deployment and Infrastructure Security Training for Mongolia
5. The way forward–Summary of Recommendations

Summary:

The steady growth of broadband and mobile networks, leading to universal network access, requires an adequate supply of IP addresses. “Transition from IPv4 to IPv6” is a top issue for the Governments, regulators, the internet community and users. While the field of IPv6 transition technology is still evolving, operators need to carefully consider which mechanism is most suitable to their networks. Broadband and mobile network deployment planning checklists must include IPv6 readiness at the service provider level. Governments and Regulators also need to make sure their e-services support IPv6 for equal accessibility across their population.

The objective of this paper is to introduce and share the current ICT sector developments, policy objectives and regulatory frameworks on implementation of the National Broadband Program and IPv6 deployment initiatives, Mongolia-IPv6 survey results, some achievement and key challenges in Mongolia compared with global trends. In its National Broadband program, Communications and Information Technology Authority (CITA, former IPTTA) set out 7 strategic objectives for the next midterm period (2015-2018). To implement the strategic objectives, IP network growth and scalability is critical. It is, therefore, necessary that Mongolia facilitates a timely transition from IPv4 to IPv6. The need for IP addresses would escalate as the country rolls out 4G-LTE, IPTV, broadband wireless access and Internet of Things (IoT) devices.

Developing countries like Mongolia recognize the need to harness ICT applications for E-Government and public service delivery. Policy and regulatory environment is one of the most critical factors affecting the development of ICT infrastructure development including transition from IPv4 to IPv6, basis for e-applications, e-sciences and research. Emphasis is needed from policy makers and regulators with the cooperation from all other stakeholders in improving of utilization the state owned fiber optic backbone network (more than 36.5 thous.km long) and its transition to IPv6 based network. Technological options are available for new installations as well as for upgrading existing networks.

The paper attempts to define key challenges Mongolia is now facing for example legal frameworks, policy and regulatory issues on IPv6 deployment and key challenges including all stakeholders, HRD and capacity building activities, data and network security issues.

Mongolia is currently examining to institute a framework, roadmap and task force on IPv6 deployment in keeping with approaches in other countries. The policy maker (CITA) and the regulator (CRC) have set up a joint working group-IPv6 to assess the status of IPv6 readiness and to prepare draft paper on IPv6 transition strategy in Mongolia.

Finally, paper proposed the way to forward-set of recommendations on Roadmap and IPv6 deployment at national level, recommendation for telecom industry and business, economics of transition to IPv6, ssecurity policy in IPv4 to IPv6 migration for the Government organizations, all level of networks and stakeholders for starting transition from IPv4 to IPv6 in Mongolia.

34

SDN implementation in IHEP

Author: SHAN ZENG¹

Co-author: Fazhi QI²

¹ IHEP

² Institute of High Energy Physics, CAS

Corresponding Author: zengshan@ihep.ac.cn

High energy physics experiments produce huge amounts of raw data, while because of the sharing characteristics of the network resources, there is no guarantee of the available bandwidth for each experiment which may cause link competition problems. On the other side, with the development of cloud computing technologies, IHEP have established a cloud platform based on OpenStack which can ensure the flexibility of the computing and storage resources, and more and more computing applications have been moved to this platform, however, under the traditional network architecture, network capability become the bottleneck of restricting the flexible application of cloud computing. This report introduces the SDN implementation in IHEP to solve the above problems, we built a dedicated and elastic network platform based on the data center SDN technologies and network virtualization technologies, meanwhile the SDN@WAN solution in IHEP will also be introduced. In the end, the test results and future works will be shared and analyzed.

Data Management & Big Data / 35

dCache, managing Quality of Service in Cloud Storage Infrastructures

Authors: Albert Rossi¹; Anupam Ashish²; Dmitry Litvintsev¹; Gerd Behrmann³; Juergen Starek²; Marina Sahakyan²; Olufemi Adeyemi²; Patrick Fuhrmann⁴; Paul Millar²; Sibel Yasar²; Tigran Mkrtchyan²

¹ FNAL

² DESY

³ NDGF

⁴ DESY/dCache.org

Corresponding Author: patrick.fuhrmann@desy.de

For the previous decade, high performance, high capacity Open Source storage systems have been designed and implemented, accommodating the demanding needs of the LHC experiments. However, with the general move away from the concept of local computer centers, supporting their associated communities, towards large infrastructures, providing Cloud-like solutions to a large variety of different scientific groups, storage systems needed to adjust their capabilities in many areas, as there are federated identities, non authenticated delegation to portals or platforms, modern sharing and user defined Quality of Storage.

This presentation will give an overview on how dCache is keeping up with modern Cloud storage requirements by partnering with EU projects, which provide the necessary contact to a large set of Scientific Communities.

Regarding authentication, there is no strict relationship anymore between the individual scientist, the scientific community and the infrastructure, providing resources. Federated identity systems like SAML or "OpenID Connect" are growing into the method-of-choice for new scientific groups and are even sneaking their way into HEP. Therefore, under the umbrella of the INDIGO-DataCloud project, dCache is implementing those authentication mechanisms in addition to the already established ones, like username/password, Kerberos and X509 Certificates.

To simplify the use of dCache as back-end of scientific portals, dCache is experimenting with new anonymous delegation methods, like "Macarons", which the dCache team would like to introduce in order to start a discussion, targeting their broader acceptance in portals and at the level of service providers.

As the separation between managing scientific mass data and scientific semi-private data, like publications, is no longer strict, large data management systems are supposed to provide a simple interface to easily share data among individuals or groups. While some systems are offering that feature through web portals only, dCache will show that this can be provided uniquely for all protocols the system supports, including NFS and GridFTP.

Furthermore, in modern storage infrastructures, storage media, and consequently the quality and price of the request storage space are no longer negotiated with the responsible system administrators but dynamically selected by the end user or by automated computing platforms. The same is true for data migration between different qualities of storage. To accommodate this conceptual change, dCache is exposing its entire data management interface through a RESTful service and a graphical user interface. The implemented mechanisms are following the recommendation of the corresponding working groups in RDA and SNIA and are agreed-upon with the INDIGO-DataCloud project to be compatible with similar functionalities of other INDIGO provided storage systems.

Summary:

This presentation will give an overview on how dCache is keeping up with modern Cloud storage requirements by partnering with EU projects, especially INDIGO-DataCloud, providing the necessary contacts to a large set of Scientific Communities. We will elaborate on the integration of modern AAI mechanisms, like OpenID Connect and Macarons, the close integration of dCache services using REST, an improved graphical interface for data access and manipulation and the first implementation of the new Quality of Service in storage interface, defined by the IaaS team of INDIGO.

Biomedicine & Life Science II / 36**2D and 3D Medical Images for Anatomy Education using a cloud computing platform**

Author: Lihshyang Chen¹

¹ *National Cheng Kung University*

Human anatomy is the basic scientific study for students in medical schools to learn the shape, position, size, and various relationships of the organ structures in the human body. In this paper, we make use of the images produced by the Visible Human Project (VHP) that are free for access through the internet to develop a 2D and 3D anatomy learning system for the students in medical schools to learn the anatomy more effective and efficient. Since the system needs to use a huge amount of storage to store the data and to use a huge amount of computation power to generate 3D images of various organs interactively, the system is installed in a cloud platform. The students can access the system through a mobile device, such as a pad computer so that they can easily learn anytime anywhere with affordable devices.

VRE / 37**Data Provenance Tracking as the Basis for a Biomedical Virtual Research Environment**

Author: Richard McClatchey¹

¹ *University of the West of England, Bristol UK*

Corresponding Author: richard.mcclatchey@uwe.ac.uk

In complex data analyses it is increasingly important to capture information about the usage of data sets in addition to their preservation over time in order to ensure reproducibility of results, to verify the work of others and to ensure appropriate conditions data have been used for specific analyses. Scientific workflow based studies are beginning to realize the benefit of capturing this provenance 1 of their data and the activities used to process, transform and carry out studies on that data. This is especially true in biomedicine where the collection of data through experiment is costly and/or difficult to reproduce and where that data needs to be preserved over time. There is a clear requirement for systems to handle data over extended timescales with an emphasis on preserving the analysis procedures themselves and the environment in which the analyses were conducted alongside the processed data.

One way to support the development of workflows and their use in (collaborative) biomedical analyses is through the use of a Virtual Research Environment. However, the dynamic and geographically distributed nature of Grid/Cloud computing makes the capturing and processing of provenance information a major research challenge. In addition most workflow provenance management services are designed only for data-flow oriented workflows but researchers are now realising that tracking

data alone is insufficient to support the scientific process [2]. What is required for collaborative research is traceable and reproducible provenance support in a Virtual Research Environment (VRE) that enables researchers to define their analyses in terms of the datasets and processes used, to monitor and visualize the outcome of their analyses and to log their results so that others users can call upon that acquired knowledge to support subsequent analyses.

We have extended the work carried out in the neuGRID and N4U projects in providing a Virtual laboratory [3] to provide the foundation for a generic VRE in which sets of biomedical data (images, laboratory test results, patient records, epidemiological analyses etc.) and the workflows (pipelines) used to process those data, together with their provenance data and results sets are captured in the CRISTAL software [4]. This paper outlines the functionality provided for a VRE by the Open Source CRISTAL software and examines how that can provide the foundations for a practice-based knowledge base for biomedicine and, potentially, for a wider research community.

References:

- 1 Y. Simmhan et al., "A Survey of Data Provenance in e-Science". In SIGMOD RECORD, Vol 34, P. 31-36. ACM, 2005.
- [2] S. Bechhofer et al., "Why Linked Data is not Enough for Scientists". Future Generation Computer Systems Vol 9 No. 2 pp 599-611, Elsevier Publishers, 2013.
- [3] R. McClatchey et al, "Traceability and Provenance in Big Data Medical Systems". Proc of CBMS 2015, Sao Carlos, Brazil.
- [4] A. Branson et al., "CRISTAL : A Practical Study in Designing Systems to Cope with Change". Information Systems 42, pp 139-152. Elsevier publishers.

Biomedicine & Life Science II / 38

Molecular dynamics of proteins in the cloud

Author: Antonio Rosato¹

¹ *University of Florence*

Corresponding Author: rosato@cerm.unifi.it

A key computational technique in Structural Biology is Molecular Dynamics (MD), a computer simulation of the motion of atoms and molecules as a function of time. MD simulations capture the behavior of biological macromolecules in full atomic detail using statistical thermodynamics laws. Such simulations serve as a computational microscope, revealing biomolecular mechanisms at various spatial and temporal scales.

In our work, we have implemented and tested a variety of approaches to the use of grid and cloud computational infrastructures for MD simulations. Previously, we implemented a web interface to setup MD simulations that were then executed on the European Grid Infrastructure 1. We then exploited the technologies developed by the INDIGO-Datacloud project to expand both the types of computational infrastructures that can be used and the portfolio of services. For this we implemented the AmberTools suite [2] as a container for Docker. Next, we implemented the use of cloud storage resources to save trajectories and perform their analysis and comparison with experimental data, such as NMR order parameters. Cloud storage uses the Onedata solution, which can work with large files such as those typically output by MD simulations. Cloud computing can also be exploited for specific applications. The setup of simulations and analysis are still performed via web interfaces, so that the user does not need to know which kind of infrastructure is used. For this use the Future-Gateway, a programmable interface of a RESTful API Server developed within INDIGO-Datacloud. The availability of these new solutions to support MD simulations allows non-expert users to get access to standardized protocols for state-of-the-art calculations and analysis that enable the successful application of MD with a low learning barrier. The solutions are available also through the West-Life Virtual Research Environment for Structural Biology [3].

1 A Grid-enabled web portal for NMR structure refinement with AMBER. Bertini I, Case DA, Ferella L, Giachetti A, Rosato A.

[2] <http://ambermd.org/>

[3] <http://about.west-life.eu/>

Network, Security, Infrastructure & Operations I / 39

EGI-CSIRT: Coordinating Operational Security in evolving distributed IT-Infrastructures.

Authors: Daniel Kouril¹; Sven Gabriel²; Vincent Brillault³

¹ CESNET

² Nikhef/EGI

³ CERN

Corresponding Authors: sveng@nikhef.nl, vincent.brillault@cern.ch, kouril@ics.muni.cz

Operational Security in Scientific distributed IT-Infrastructures like EGI is challenging. Existing computation frameworks get further extended, and new technologies implemented.

In this evolving environment new policies have to be developed, and existing policies and procedures have constantly to be extended to meet new requirements.

To efficiently enforce new policies, the security monitoring infrastructure has to be further developed to cover all elements of the evolving infrastructure. Finally the incident response (ir) tool set has to be extended to be able to efficiently handle security incidents affecting 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 already involved in incidents handled by EGI-CSIRTs Incident Response Task Force (IRTF). Here it got apparent that the existing procedures and tools, which were successfully applied in IR in EGI, exposed deficiencies when applied to the FedCloud services.

This triggered the development of a central User- and Virtual Machine-Management for frameworks deployed in EGI-FedCloud.

The status of these tools will be demonstrated and the integration with the existing IR tools discussed.

In EGI the policies and procedures are put to a test in so called Security Service Challenges (SSCs) to check if they indeed help in security operations to prevent and respond to incidents.

An SSC addressing EGI-FedCloud interfaces and IR procedures will be described.

Humanities, Arts & Social Sciences II / 40

A Proposal: ePortfolio for enhancing active learning for the future generation

Author: Tosh Yamamoto¹

Co-authors: Maki Okunuki ¹; Masaki Watanabe ²; Yuri Kite ¹

¹ *Kansai University*

² *iGroup Japan*

Corresponding Authors: soetosh@gmail.com, ykite@kansai-u.ac.jp, okufield@yahoo.co.jp, masaki@igroupjapan.com

It is proposed that, the students' meta-cognitive reflection being the key to learning, the development of the students' expressive skills in writing literacy will, indeed, bring about a success in life after graduation. The four years of college education must fill up the "fuel tank of knowledge, wisdom, competencies and skills for lifelong learning", which must not be depleted for over 40 years until retirement. By nurturing the students' writing skills as the artifact or evidence for academic learning through meta-cognitive activities, is it possible to conduct the real education, producing graduates of high quality and caliber.

The basic assumption of incorporating the development of writing skills in the college curriculum or program is for the propose of making our culture richer and elevating the value of our heritage for the benefit of the better future in terms of constructive and humanistic communication. It is seen that ePortfolio has potential to grow into such robust ICT enhanced system for the education in a new paradigm fulfilling the need for transdisciplinarity.

This proposal is to put ePortfolio into a bigger picture in higher education, namely, in the realm of ePortfolio for academia, in which the process of learning leads to the benefit of career design and life-long learning. It cannot be denied that the proficiency in academic writing will bring students to a success in career as well as in the life long learning. In other words, the artifact in writing is the mirror of the learning mind. As the IFTF (Institute for the Future) claims that the 2020 skills include Global Awareness and Rich ICT, Media Literacy, as well as Digital Communication/ Presentation skills as the essential future skills, the mirror of the reflective learning mind incorporates not only the written information but also the rich media. Thus, the future education fortified with ePortfolio must also incorporate artifact of learning in rich media. While in the past, paper and a pencil were the optimal technologies to reflect the evidence for learning, digital media literacy has been becoming dominant due to the advancement of ICT. It is believed that the digital media have been providing us with richer ways of communication and presentation of the learning mind.

Network, Security, Infrastructure & Operations I / 41

Can R&E federations trust Research Infrastructures?

Author: David Kelsey¹

Co-author: David Groep²

¹ *STFC-RAL*

² *Nikhef*

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

Research Infrastructures increasingly use national and global "Research and Education" (R&E) authentication federations to provide access to their services. Collaborators log on using their home organization credentials, the Research Infrastructure (RI) enriches it with community information, and access decisions are made on its combined assertions. Studies in the AARC project have shown that research communities connect to the R&E federation using an 'SP-IdP proxy' design pattern: a single logical component makes the RI appear as a single entity towards the R&E federations. The RI can also augment the 'R&E identity' of their users with membership information. Thus the RI shields itself from heterogeneity in the global R&E federations and it eases service deployment by 'hiding' all services behind a single proxy identity provider (IdP) that itself needs to be registered only once in R&E federations. The AARC Blueprint Architecture identifies this model as a recommendation for engaging research collaborations with R&E federations.

The use of a proxy in itself poses policy challenges: services 'internal' to the community see only a single IdP that they have to trust ultimately. Generic service providers that span multiple research communities will have to trust many of these proxies – there are over a hundred multi-national RIs

in the world but many multi-purpose e-Infrastructures as well. And towards the R&E federations, the SP-IdP proxy hides all of the research services: home organisations and R&E federations see just a single service provider, even if the services behind it are provided in hundreds of different administrative domains.

Building on the Security for Collaboration among Infrastructures (SCI) framework, the “Security Networked-Community Trust-framework for Federated Identity” (Sntcfti) proposes a policy framework that allows determination of the ‘quality’ of such SP-IdP proxies and the research services behind them. For example, a SP-IdP-proxy for EGI – proxying for all its compute and storage services – would be able to express to the R&E federation space that it has an internally-consistent policy set, that it can make collective statements about all its constituent services and resource providers, and that it will abide by best practices in the R&E community, such as adherence to the Data Protection Code of Conduct (DPCoCo), REFEDS Research and Scholarship (R&S) entity category, Sirtfi – the security incident response trust framework that is in itself a development from the SCI structure.

By addressing the structure of the security policy set that binds services ‘hiding’ behind the SP-IdP proxy, Sntcfti allows comparison between proxies, assign trust marks for meeting requirements, and it allows a scalable way to negotiate and filter based on such policies. It eases authentication and attribute release by R&E federations as well as service providers (by easier enrolment in federations and because R&E IdPs may be more willing to release attributes if the proxy can convincingly assert DPCoCo and R&S), but also aids assessment by generic e-Infrastructures providers that know the RI proxy meets their trust requirements.

We will describe the requirements on the Sntcfti framework and show how it applies to research and generic e-Infrastructures.

42

ICT Enhanced Interactive Remedial Mathematics Program for Science and Engineering Students

Author: Tosh Yamamoto¹

Co-authors: Kunio Hamamoto¹; Masaki Watanabe²

¹ *Kansai University*

² *iGroup Japan*

Corresponding Authors: soetosh@gmail.com, kunio.hamamoto@gmail.com, masaki@igroupjapan.com

The purpose of this presentation is to share with the academic audience, the KU’s experience of implementing GradeAnything® by TurnItIn® to the basic math remedial courses in the departments of Science and Engineering.

The students were issued a review assignment in the form of paper in every class. After the class, the students solved the math problems on the paper. Then they took a snap shot of their paper using their smartphones and submitted their answers to TurnItIn®.

In order to assure the same high quality of advice for corrections given to all the students and the efficiency of correctional advice by the team of instructors or tutors, the grade mark feature of TurnItIn® was employed. In this way, even in a large size class, the frequent interactions were provided. The students were able to refer to the rubric for each assignment to review the important points and to check their mistakes from individualized comments by the instructor. All the processes were conducted in terms of the smartphone on the students’ side and GradeAnything® on the cloud.

The presentation includes the result of the experiment; the students’ achievements before and after the implementation of GradeAnything® as well as the reflections by the presenters.

43

THE VIRTUAL REALITY (VR) TRAINING SYSTEM FOR DISASTER PREPAREDNESS

Author: Deep Ayadi¹

¹ *The Thin Page*

Corresponding Author: deep.ayadi@thethinpage.com

In last 30 years, Nepal is hit by a wide-ranging variety of natural disasters that shook the nation. That has caused loss of thousands of life and million worth property. Floods occur with the greatest frequency all over the Nepal. Flooding is followed by landslides, forest fire, thunderstorm etc. Occurrence of these disaster is not known spatially and temporally and need to be prepared when they do. Emergency and response teams can minimize casualties if they get the right kinds of training. Virtual reality is an advanced technology that may be used for this purpose. It provides users with artificial worlds that they can interact with and explore in virtual environment in same phenomenon as disaster occurs. Emergency and response team can spend time in these digital environments while they enhance their skills in dealing with disasters.

This paper aims to evaluate the potential uses of Virtual Reality (VR) technologies for Disaster Risk Reduction (DRR) education. It presents an analysis of the state of virtual reality training systems, its prerequisite and outline the need of National Emergency Operation Center (NEOC) for disaster response and relief with an explicit focus on disaster preparedness. This paper discusses the application of VR technology in addressing disaster preparedness challenges as effective training in preparing emergency and response team more effective and efficient for work in disaster conditions in real environments in Nepal. Virtual reality could be the excellent way to help emergency and response team become more aware of a perilous risks of a disaster.

Summary:

Keywords: Disaster, Virtual Reality, Disaster Risk Reduction (DRR), Training and Education, Nepal

Biomedicine & Life Science II / 44

Investigating community detection algorithms and their capacity as markers of brain diseases

Author: Eva Vytvarova¹

Co-authors: Eva Hladka¹; Irena Rektorova²; Jan Fousek¹; Michal Mikl²

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

² *Central European Institute of Technology, Masaryk University, Brno, Czech Republic*

Corresponding Author: eva.vytvarova@gmail.com

In this paper, we present the workflow for evaluation of brain functional connectivity with different community detection algorithms, and their strengths to discriminate between health and brain disease. We further analyze the computational complexity of particular pipeline steps aiming to provide guidelines for both its execution on computing infrastructure and further optimization efforts.

The human brain consists of 10^{11} neurons interconnected with approximately 10^{14} connections creating large complex structure. This anatomical connectivity is a substrate for brain function which we can both measure and model on various scales. The goal of current research is understanding the brain as a system and its behavior as whole, i.e. identifying the spatiotemporal structure of brain function. This large-scale functional connectivity of neuronal populations is defined as statistical relations between neural activities of distinct brain cortical regions and is

usually measured by fMRI (functional magnetic resonance imaging), EEG or MEG (electro/magneto-encephalography). The very promising tool is to model the system as a network using network science methodology, drawing inspiration from previous applications in research areas such as economics, transportation, communication or immunology.

The network analysis models complex system as a set of discrete units and describes interactions between them, and provides ways to measure the importance of nodes and detects communities, by diverse metrics (properties) describes topology and whole-system events or can be used to model the growth of the system in size. Primarily, network analysis is a data-driven technique and is an excellent tool to study real systems and to understand them. It requires interdisciplinary and big data processing approaches.

Neuroscience community puts great effort into finding a potential marker of disease that would be noninvasive, easy to establish and stable. In this paper, we evaluate strengths of community detection methods to differentiate between networks of healthy and diseased subjects on the example of MCI-AD and AD (Mild Cognitive Impairment and Alzheimer's Disease) as illnesses known to influence network topology. We apply discriminant analysis including parameter sweeps of methods for data preprocessing to quantify the ability of community detection algorithms to distinguish between patient and control and thus evaluate each one of them as a promising or inconclusive marker of a disease. The pipeline consists of a relatively large number of subsequent steps. We provide the analysis of the computational complexity which will guide further optimization efforts. Since we're dealing with large amounts of data we use computing infrastructure of MetaCentrum, CESNET, Czech Republic, to be able to analyze the data.

45

Automated quality control for data from ASTI automatic weather stations

Author: Jay Samuel Combinido¹

Co-author: John Robert Mendoza¹

¹ *Advanced Science and Technology Institute*

Corresponding Author: jaysamuel@asti.dost.gov.ph

Meteorological observations are vital for scientific research, for weather monitoring and weather-related decision making. Such data sources act as support for several applications relating to agriculture, aviation, early warning, weather prediction, and disaster mitigation. In the last four years, the Advanced Science and Technology Institute (ASTI) has deployed over 1600 ground stations (automatic weather stations and rain gauges) over the Philippines in response to the government's effort in disaster risks management. The growing demand for these data requires that the quality control be performed to ensure their reliability and integrity. Notably, this tightly associates with the validation of high-quality data against *\emph{ground truth}* coming from manned synoptic stations of PAGASA. The ultimate goal is not only to provide guidance to the end-users but also aid operations in identifying and potentially predicting problems that may require immediate attention.

This work presents the development of quality control procedures applied onto data obtained from ASTI ground stations. These procedures include verification of geolocation information, timestamp, range, step, persistence and internal consistency checks. In the proposed system, data are not altered; rather, assigned a flag according to the result of each check. In the initial tests performed on data (air pressure, air temperature, relative humidity, rainfall and wind speed) collected between 2012 and 2015 from 15 stations, 7.97% of the records were detected to have invalid timestamps, 0.11% were flagged in the range check, 0.26% in the step check, 6.35% in the persistence check and 0.87% in the internal consistency check.

By comparing AWS temperature and rainfall data with that of PAGASA synoptic stations, we found that eliminating dubious observations improved the agreement of AWS measurements with ground truth. Validation statistics for temperature and rainfall were (*\emph{r}*=0.94; bias=0.37°C; RMSE=1.01°C) and (*r_s*=0.80; bias=-0.79mm/d; RMSE= 12.04mm/d), respectively, as opposed to the baseline statistics

using raw data given by ($r=0.84$; bias= 0.47°C ; RMSE= 1.77°C) for temperature and ($r_s=0.80$; bias= -1.17mm/d ; RMSE= 14.36mm/d) for rainfall.

While the system was able to identify various errors in the measurements, continuous improvements to the algorithm are necessary to achieve more robust quality control. A few examples are spatial consistency checks, application of site-specific quality control parameters, long-term time series analytics, and data interpolation. Nevertheless, this development of quality control system provides considerable value to ASTI ground station data that is expected to promote further end-user utilization for research and various applications.

Network, Security, Infrastructure & Operations IV / 46

A solution for secure use of Kibana and Elasticsearch in multi-user environment

Author: Wataru Takase¹

Co-authors: Takashi Sasaki¹; Tomoaki Nakamura¹; Yoshiyuki Watase¹

¹ KEK

Corresponding Author: wataru.takase@kek.jp

In order to check health, activities, or resource usage of IT service, monitoring is indispensable. A combination of Kibana and Elasticsearch is used for monitoring in many places such as KEK, CC-IN2P3, CERN, and also non-HEP communities. Kibana provides a web interface for rich visualization and Elasticsearch is a scalable distributed search engine. However, these tools do not support authentication and authorization features by default. There is no problem in the case of single-user environment. On the other hand, in the case of single Kibana and Elasticsearch services shared among many users, any user who can access Kibana can retrieve other's information from Elasticsearch. In multi-user environment, in order to protect own data from others or share part of data among a group, fine-grained access control is necessary.

The CERN cloud service group provides cloud utilization dashboard to each user by Elasticsearch and Kibana. The group has been deployed a homemade Elasticsearch plugin to restrict data access based on a user authenticated by the CERN Single Sign On system. It enables each user to have a separated Kibana dashboard for cloud usage and cannot access to others. Based on the solution, we propose an alternative one which enables user/group based Elasticsearch access control and Kibana dashboards separation. It is more flexible and can be applied to not only the cloud service but also other various situations. We confirmed our solution works fine in CC-IN2P3. And a pre-production platform for CC-IN2P3 is under construction.

We will describe our solution for the secure use of Kibana and Elasticsearch including integration of Kerberos authentication, development of a Kibana plugin which allows Kibana dashboards to be separated based on user/group, and contribution to Search Guard which is an Elasticsearch plugin enabling user/group based access control. We will also describe the effect on performance from using Search Guard.

Massively Distributed Computing and Citizen Sciences / 47

Using virtualized computing resources with the DIRAC Interware

Author: Andrei Tsaregorodtsev¹

¹ CPPM-IN2P3-CNRS

Corresponding Author: atsareg@in2p3.fr

Multiple scientific communities are using computationally intensive applications in their research activities. They are largely relying on national and international distributed computing infrastructures, which are mostly based on well-known grid technologies. However, with a progressively wide adoption of the cloud technologies more and more computing power is available in a form of groups of dynamically created virtual machines. The machines can be of general purpose or specialized virtual appliances suitable for a particular application. This provides a very high level of flexibility in using computing resources but makes it quite difficult to manage large amount of cloud computing power by an average user. On the other hand, still most of the computing resources are available through grid infrastructures. Therefore, it is necessary to provide a transparent user access interface to both types of computing infrastructures in order to enlarge the overall available power and to ensure a smooth transition to the use of the new technology. The DIRAC Interware project offers software and multiple ready to use components to build distributed computing infrastructures. It provides tools to integrate various types of computing resources including grid and cloud systems. DIRAC users see cloud resources as logical entities in the same way as grid sites. The DIRAC Workload Management System allows combination of grid, cloud and other resources within the same complex workflow. In this contribution we will describe the recent progress in the development of the cloud management subsystem of the DIRAC Project, its architecture and main components. We will demonstrate how the combined configuration, usage and monitoring of the grid, cloud and other computational resources is performed. We will present how resources provided by cloud federation infrastructures are made available via the DIRAC services. We will give several examples of their usage by large high energy physics experiments and other scientific communities.

Earth, Environmental Science & Biodiversity II / 48

Towards Environmental Computing Compendium

Author: Matti Heikkurinen¹

Co-author: Dieter Kranzlmüller²

¹ LMU

² LMU Munich

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

Environmental computing is focusing on producing actionable knowledge by advanced environmental modelling on high performance computing platforms. The environmental computing community shares a tacit understanding of what are the initiatives, tools and approaches that belong to the core scope of this discipline. A website has been used as an interim community resource for collecting links to resources that are of relevance to this discipline. This link collection forms an interesting dataset for assessing different aspects of environmental computing as a discipline. Obviously it allows making an estimate of the size of the community – both the core group considering themselves as environmental computing specialists and the group of individuals and organisations involved in the related initiatives. While this measure is very reductionist, it already allows to make arguments e.g. with regards to inclusion of environmental computing aspects in a broad range of curricula (ranging from computer science to different specialties dealing with environmental modelling).

In terms of analysing the more fine-grained distribution of the entries in the above mentioned environmental computing dataset, the main value of the current data is that it is sufficiently large for us to test different categorisation approaches. The challenges encountered in this process are especially useful, as they will force us to reflect not only which (if any) of the categories are exclusive in nature and which of them are actually attributes (or “tags”) that represent specific aspects of them. This differentiation can be important when considering ways to identify and benefit from synergies between different environmental computing groups, initiatives and tools. As an example, a project and an organisation both represent a community, and in both cases you can identify both the formally “core” team (individuals with a contract of some kind, specifying their role in the community), and the surrounding, larger group of stakeholders. However, due to the typically limited lifetime of a project and different internal structure (a project usually consists of several legal entities), the engagement

strategy needs to be adjusted. A project collaboration usually needs to provide relatively short-term tangible benefits to be successful, whereas inter-organisational collaboration arrangements can be more loosely defined.

In addition to presenting the observations the current dataset allows us to make we will also discuss the limitations of the current dataset. The paper also presents plans on how to overcome them to increase the representativeness of the dataset and its ability to capture more comprehensive picture of the environmental computing landscape.

Network, Security, Infrastructure & Operations II / 49

Identifying Suspicious Network Activities in Grid Network Traffic: Finding the needle in a stack of needles

Author: Fyodor Yarochkin¹

¹ *Academia Sinica*

Corresponding Author: fy@iis.sinica.edu.tw

In this presentation we will share our experience with analysing a year of grid network flow data. The network flow data provides only limited information regarding the nature of network traffic that traveled through the network segments. Therefore researchers need to come up with additional methods of anomaly detection, data enrichment and cross-referencing in order to effectively identify 'true-positives': a subset network flows which could be of some interest to security officers: from denial of service attacks, to malware operations, network scanning and attacker's lateral movements. In this study having access to other network data feeds (such as honeypot networks) and full packet payload monitoring, so we demonstrate how such sources could be effectively leveraged in identifying and verifying suspicious network activities.

Summary:

A quick summary of the presentation:

Visibility

Data subsets by Protocol

Analyzing Traffic Direction (internal/external/incoming/outgoing flows)

Outgoing connections using high-risk protocols

Identifying Local Assets

Threats

Identifying Recon activities :Scans, Bruteforce sessions

Anomalies: TCP and UDP

High-risk flows

Identifying C2 calls

Lateral spreading and worming activities

Exfiltration

Infrastructure abuse

DDOS

DDOS and DNS traffic sessions

DDOS and SSDP

Exploring 'bad' SNMP

'bad' NTP traffic

Conclusions and Future research

VRE / 50

Design and Implementation of Portal System for Subscribed Cloud Services in Identity Federation

Author: Takeshi Nishimura¹

Co-authors: Eisaku Sakane²; Kento Aida²

¹ *Project Researcher*

² *National Institute of Informatics*

Corresponding Author: takeshi@nii.ac.jp

With growing number of online services, identity federation is rapidly spreading, especially in the academic world. Identity federations have been established in many countries. In Japan, an academic identity federation called "GakuNin" is operated since 2010.

In identity federation, IdP can provide information about a user as attributes in addition to the authentication-related information. An SP can request attributes about a user from an IdP and then utilize those attributes to authorize users and provide services. In general, an IdP releases attributes they derived from authoritative source systems within the organization. Although academic society affiliations or research community membership across universities are user data that would often be interesting for academic services, a typical campus identity management system is not designed to manage such data, and the IdP is thus unable to provide it to services. Some of the advanced academic federations are beginning to make available attributes like this, especially related to the membership of a group that spans multiple organizations, by means of the Attribute Provider (AP). The SP is typically responsible for aggregation of additional attributes from APs after the initial exchange with the IdP. A unique user identifier is supplied by the IdP to the SP and is shared with APs in order to look up additional attributes of that user.

An IdP usually joins one federation and is connected to some services such as organization-specific services provided by the organization. If users in the IdP could use all the services joining the federation, they could easily recognize usable services through their credentials of the IdP. But the reality is not the case, for example, because the organization does not subscribe a service, or the IdP is not configured to send attributes to the service correctly, and so on. Users must access each service that joins the federation in order to make sure that they can use the service. It is not realistic especially if the federation has a lot of services.

Furthermore, adding AP makes this problem more complex. The AP must know user identifiers in advance to identify themselves, usually by means of attributes from IdPs. Some SPs must also receive specific attribute values from an AP for authorization, which is uncertain for ordinary users. We propose a portal system for each stake holder to register some information. With this information it displays his / her usable services for any accessing user. For example, IdP operators register SPs configured on their IdPs, and SP operators register their strategies for authorization. APs provides membership information of the user.

We do an elementary evaluation about our system, which calculates availability of some complicated SPs by some conditioned users.

We will present a demo of this system working in our production federation.

Data Management & Big Data / 51

Machine Learning analysis of CMS data transfers

Authors: Daniele Bonacorsi¹; Nicolò Magini²; Valentin Kuznetsov³

Co-authors: Diotallevi Tommaso¹; Kančys Kipras⁴; Matonis Zygimantas⁵; Repečka Aurimas⁴

¹ *University of Bologna*

² *Fermilab (US)*

³ *Cornell University*

⁴ *Vilnius University*

⁵ *University of Vilnius*

Corresponding Authors: daniele.bonacorsi@unibo.it, vkuznet@gmail.com, nicolo.magini@cern.ch

Tens of Petabytes of collision and simulated data have been collected and distributed across WLCG sites in Run-1 and Run-2 at LHC. A low latency in transfers among dozens of computing centres is crucial to make an efficient use of the computing resources. Despite on average the desired level of throughput has been successfully achieved to serve the LHC physics programs, it is not uncommon to observe transfer latencies caused by a large variety of causes, from file corruptions to site issues, most of which require operator intervention. To improve on this front, in particular, the CMS experiment equipped the PhEDEx dataset replication system with a system to collect the latency data, and a mechanism to categorise and analyse them promptly, matching them to quick and focussed operators intervention. The transfer latencies data has also been the target of Machine Learning techniques - already used in CMS to study and predict the dataset popularity - and preliminary results on the predictability potential of this approach will be presented and discussed.

Network, Scurity, Infrastructure & Operations IV / 52

A Method for Remote Initial Vetting of Identity with PKI Credential

Author: Eisaku Sakane¹

Co-authors: Kento Aida¹; Takeshi Nishimura¹

¹ *National Institute of Informatics*

Corresponding Author: sakane@nii.ac.jp

With the growth of large-scale distributed computing infrastructures, a system that enables researchers – not only international collaborative research projects but also small research groups – to use high performance computing resources in such infrastructures is established. For the computing resource use system which invites researchers in the world to submit the research proposal, it is tough to carry out initial vetting of identity based on a face-to-face meeting at a window for the system if the researcher whose proposal is accepted lives in a foreign country.

The purpose of this paper is to propose a method to solve the difficulty of initial vetting of identity for a remote user.

An identity management (IdM) system vets the identity and reality of a user by checking the beforehand registered personal information against the identity documents. After the identity vetting, the user can obtain a credential used in the infrastructure. Suppose that the IdM system(A) needs to initially vet the identity of a user and that the user already possesses a credential issued by the other IdM system(B). The basic idea of this paper is that the IdM system(A) uses the credential issued by the IdM system(B) for the initial identity vetting if the level of assurance of the IdM system(B) is the same as or higher than the IdM system(A). However, the IdM system(A) cannot

always check the identity against the attribute information provided by the credential. In a trust federation, the IdM system will be able to finish vetting the identity by making reference to the other IdM system that issued the credential for the necessary and sufficient identity data.

As the credential handled in this paper, we focus on Public Key Infrastructure (PKI) credentials that often used in large-scale high performance computing environments. We discuss necessary condition and procedure for ensuring that the remote initial vetting of identity with a PKI credential is the same assurance as the one based on a face-to-face meeting. The proposed method can be introduced to an existing PKI without large changes. The basic idea of the proposed method can be also applied to an infrastructure based on another authentication technology. The applicability of the basic idea is also considered.

Network, Security, Infrastructure & Operations II / 53

Modern Monitoring Systems

Author: Aleksander Paravac¹

¹ *University of Wuerzburg*

Corresponding Author: aleksander.paravac@uni-wuerzburg.de

Monitoring your infrastructure is vital to ensure your services are running for your users in an Cloud or Grid environment.

Modern Monitoring Systems can not only check the health of your hardware but also the status of provided services or issues that are relevant to security.

In this talk an overview of Modern Monitoring Systems is given and how they interact, scale and integrate with your infrastructure.

Network, Security, Infrastructure & Operations IV / 54

dCache, towards Federated Identities and Anonymized Delegation

Authors: Albert Rossi¹; Anupam Ashish²; Dmitry Litvintsev¹; Gerd Behrmann³; Marina Sahakyan²; Olufemi Adeyemi²; Patrick Fuhrmann⁴; Paul Millar²; Tigran Mkrtchyan²

¹ *FNAL*

² *DESY*

³ *NDGF*

⁴ *DESY/dCache.org*

Corresponding Author: paul.millar@desy.de

For over a decade, X509 Proxy Certificates are used in High Energy Physics (HEP) to authenticate users and guarantee their membership in Virtual Organizations, on which subsequent authorization, e.g. for data access, is based upon. Although the established infrastructure worked well and provided sufficient security, the implementation of procedures and the underlying software is often seen as a burden,

especially by smaller communities trying to adopt existing HEP software stacks. In addition, it is more efficient to guarantee the identity of a scientist at his home institute, since the necessary identity validation has already been performed. Scientists also depend on service portals for data access and processing, on their behalf. As a result, it is imperative for the infrastructure providers to support delegation of access to these portals for their end-users without compromising data security and identity privacy.

The growing usage of distributed services for similar data sharing and processing have led to the development of novel solutions like OpenID Connect, SAML etc. OpenID Connect is a mechanism for establishing the identity of an end-user based on authentication performed by a trusted third-party identity provider, which thereof can be used by infrastructures to delegate the identity verification and establishment to the trusted entity. After a successful authentication, the portal is in possession of an authenticated token, which can be further used to operate on infrastructure services on behalf of the scientist. Furthermore, these authenticated tokens can be exchanged for more flexible authorized credentials, like Macaroons. Macaroons are bearer tokens and can be used by services to ascertain whether a request is originating from an authorized portal. They are cryptographically verifiable entities and can be embedded with caveats to attenuate their scope before delegation.

In this presentation, we describe how OpenID Connect is integrated with dCache and how it can be used by a service portal to obtain a token for an end-user, based on authentication performed with a trusted third-party identity-provider. We also propose how this token can be exchanged for a Macaroon by an end-user and we show how dCache can be enabled to accept requests bearing delegated Macaroons.

55

When and Why Are Citizens OK to Publicize Their Data?

Author: Frank Liu¹

¹ *National Sun Yat-Sen University*

Corresponding Author: frankesliu@gmail.com

Open Data is becoming a “fashion” and a trend in the field of data science is built upon the availability of big data provided by the public and private sectors. While data scientists have been urging and demanding the government to publicize data with the “value-adding” rationale, it has been little discussion and reflection about the data supply side, not the government but the citizen. This study is one of the first attempts in political science and data science to bring people’s voice and perspectives into the development of open data community. By utilizing smilepoll.tw, a web-survey open-data platform that was initiated in 2011 and matured in 2015, and by reflecting upon big data methodology, I designed a throughout questionnaire about citizens’ attitudes toward publicizing their data for a variety of usage purposes, with regard to financial, public service, research, and business usage. This study includes two parts: an Internet survey of over 1,000 citizens and a series of in-depth interviews of 20 selected citizens who are polarized in their attitudes toward publicizing their data. The report of the findings will communicate directly to the demand side of open data, to the governments that hesitate to move forward to facilitate data, and to the researchers who have been concerned about and criticizing the trend of open data.

Occupation recommendation with major programs for adolescents

Author: Ankhtuya Ochirbat¹

Co-author: Timothy K. Shih²

¹ *National Central University*

² *NCU*

Corresponding Author: ankhaa3@yahoo.com

Students make initial and critical decisions regarding what to study further and which career path to pursue. Many students enter post-secondary education without a clear idea of their major and future career plans. Mismatch of the major choice and lack of processing information through the professional study is one of the reasons to switch the major. Such changes are wasteful in time and resources, and also it is the cause of emotional and financial stresses of students. An approximated twenty to fifty percent of students enter college as undecided, and an approximated seventy-five percent of students change their major at least once before graduation.

Hence, to find an adolescent's major/occupation as earlier as possible can help them to choose correct learning direction. Due to the rapid development of society, adolescents need counselling session to enable them to choose a suitable major and occupation. The choice of the major and occupation has become increasingly complex due to the existence of multiple human abilities and interests which mean each person or human has their ability or skill at the certain area and can be applied to multiple occupations. The main problem of difficulty making the major/occupation selection among students is they do not know how to make decisions, lack of knowledge and information about majors/occupations and over information of common occupations. Therefore, it is essential to build the major/occupation recommendation system for the adolescents with a capacity to meet all the needs where it provides direction and guidance to adolescents in choosing the major/occupation that suits with their vocational interests and competencies.

In this regard, this study proposes Occupation Recommendation System (ORS) with major program of study using Collaborative Filtering (CF) methods, a framework to assist students in the major/occupation selection, to provide the information about contemporary occupations, and to suggest the suitable occupations with the major program of study based on their interests, skills, favorite courses, high score courses, earned certificates, and learned tools/technologies. To do this, firstly we normalize data and secondly compute similarities. Then the predictions of factors are calculated if the factors are missing for the active student. Finally, the system executes hybrid CF algorithms to recommend the occupations with major program to the active student. The study was carried out in accordance with following two questionnaires with Mongolian 26 adolescents in the spring semester of the 2015/ 2016 academic year. In addition, Holland vocational Interest and skill questionnaires were employed respectively during the experiment for collecting data.

Physics & Engineering II / 57

Framework for distributing Radio Astronomy processing across Clusters and Clouds

Author: Alexandar Mechev¹

Co-authors: A. Danezi²; Raymond Oonk³; T. W. Shimwell³

¹ *Sterrewacht Leiden*

² *SURFsara*

³ *Leiden Observatory*

Corresponding Author: apmechev@strw.leidenuniv.nl

The Low Frequency Array (LOFAR) radio telescope stationed near Exloo, the Netherlands is an international aperture synthesis radio telescope developed to image the universe in the 20-200MHz frequency bands. Unlike telescopes using dishes or mirrors to focus light, aperture synthesis requires

large amounts of processing between data acquisition and creating science ready images. While data can be split by frequency and processed in parallel, creating a wrapper around the processing software is needed in order to standardize data reduction across multiple locations while keeping track of the overall processing status.

This work presents a framework to wrap the radio processing software, a global location to track pipeline steps as well as an implementation of CERN's VM Filesystem (CVMFS) used to standardize the software install. With a central location to track the pipeline progress and a standard software install, this software suite allows to easily process LOFAR data on any computer, node or cluster that is connected to the internet. This distribution of processing allows to tap more resources than are available at any single cluster as well as a way to track the execution at all locations through a common interface.

Installing the LOFAR software on a CVMFS server allows every location to use the same software install, by tracking the software on the main server. The set of scripts that define the processing pipeline are placed in a sandbox folder along with a shell executable tasked with setting up the processing. The shell script sets up the environment, downloads, processes and uploads the data. The progress of the job is logged in an Apache CouchDB database inside a 'job token' document which contains all relevant data required to run the job.

The CouchDB token defines the parameters of the processing job. A user-friendly python interface was built to read/update fields from the tokens, download configuration files attached to the token, create and delete tokens as well as create/delete views from the database. Using this interface, it is easy to create batches of jobs to process large amounts of data on multiple nodes data centres or cloud services. Additionally, the python package logs the progress of each job by tracking the current processing step executed and updating the job token.

Further work will include integrating a scheduler which can make decisions which locations to use based on the current workload. Additionally, decision boundaries can be inserted between execution steps that analyse the intermediate solutions and decide on re-processing or terminating based on current data quality.

58

GridPP wider community support and SKA synergies

Author: Jeremy Coles¹

¹ *University of Cambridge*

Corresponding Author: jeremy.coles@cern.ch

GridPP has been a core member of the WLCG since its inception. In recent years various pressures have led to an increased focus on sharing GridPP resources and supporting new communities outside of the LHC experiments and beyond HEP. The first part of this paper brings together a review of the current approaches within GridPP to harness traditional resources for these 'other' Virtual Organisations, and also to bring in new resources such as HPC and commercial cloud providers. The paper presents several case studies and highlights common difficulties being encountered. A community whose impacts will be strongly felt in the coming years is astrophysics. Within this domain, the Square Kilometre Array (SKA) project presents some of the biggest computing challenges. The second part of this paper presents the background to SKA computing work being taken forward by the Science Data Processor (SDP) Consortium and outlines studies informing the design (many of which parallel HEP challenges in areas such as exploiting parallelism). The paper ends with a summary of directions and current findings of studies examining the challenge for SKA Regional Science Centres. Potential synergies with HEP distributed computing are discussed.

59

Virtualized Web Portals in EGI Federated Cloud

Author: Aleš Křenek¹

¹ *Masaryk University*

Corresponding Author: ljocha@ics.muni.cz

Web portals became a popular interface to complicated workflows of scientific computation in many areas due to their ability to shield the user from the complexity of the workflow implementation and control of the used computing resources.

On the other hand, implementation of such web portal itself involves rather complicated technology, which becomes difficult to operate and maintain as the number of its users grows. Moreover, user requirements may become contradictory (e.g. stick with an old version of simulation software to retain full backward compatibility vs. use a new version to leverage new features).

Cloud infrastructures available nowadays, together with cloud orchestration engines allow addressing those issues in a scalable way. Rather than maintaining “one size fits all” instance of an application web portal, we provide templates for deployment multiple instances of a specific portal, dedicated to smaller user groups. Such instances may vary in size, and they can be easily customized to the needs of their users, while still reusing 99% of the common code and configuration.

We describe in detail the solution adopted in the EU H2020 West-life project. The current implementation uses Cloudify orchestrator, extended with OCCI plugin, to spawn and to keep alive the set of virtual machines forming the portal in the environment of EGI Federated Cloud. Deliberately, minimalistic VM images containing the bare operating system only are used; all required software is installed and configured with Puppet system management upon VM instantiation. In this way, tedious management of huge VM images, which would require frequent updates and which would be complicated to distribute otherwise, is avoided.

We demonstrate the approach with a prototype of web portal on top of the Gromacs molecular dynamics package. The original version of the portal was developed in the WeNMR EU project, it is based on Gromacs version 4.5, and it uses gLite workload management system to submit payload jobs to the EGI grid resources. Besides wrapping the portal to the virtualized cloud environment we adapted it to use the more recent version of Gromacs (5.1), to leverage GPU acceleration when the hardware is available. Submission to the grid was replaced with use of local batch system (Torque) which is sufficient to manage resources at the size expected for any instance of such virtualized portal, while being less complicated, self-contained, and more reliable.

We also outline plans for future work – adopting the more advanced cloud orchestrator developed in EU Indigo-Datacloud project; in particular, we will focus on flexible sizing of the portal instances according to their current load.

Humanities, Arts & Social Sciences I / 60

Data Science as a Foundation Toward Open Data and Open Science: The Case of Taiwan Indigenous Peoples open research Data (TIPD)

Author: Ji-Ping Lin¹

¹ *Academia Sinica*

The research aims are threefold: to (1) demonstrate the methodology of data science in constructing Taiwan Indigenous Peoples open research Data (TIPD, see enter link description here <http://TIPD.sinica.edu.tw>, and enter link description here <https://osf.io/e4rvz/>) based on Taiwan Household Registration (THR) administrative data; (2) to illustrate automated or semi-automated data processing as a determinant of constructing effective open data; and (3) to demonstrate appropriate utilization of “old-school” data format such as multi-dimensional tables as an effect means to overcome legal and ethic issues. The research extracts valuable information embedded in micro data of THR and enriches the extracted information through the processes of cleaning, cleansing, crunching, reorganizing, reshaping the source data. Major outputs of TIPD amount to 7,300 files in number and around 32 GB in size. TIPD now consist of three categories of open research data: (1) categorical data, (2) household structure and characteristics data, and (3) population dynamics data. Categorical data include two broad dimensions. The data enrichment processes produce a number of data sets that contain no individual information but retain most of source data information. The enriched data sets thus can be open to the public for open administrative data study. The open data are systematically constructed in an automated or semi-automated way through the integration of compiler programming language, software, and script languages.

Keywords: data science, administrative data, open data, open science, TIPD

Biomedicine & Life Science I / 61

NMRBox and VCell: Common Flexible Infrastructure Supporting Two Very Different Virtual Research Communities

Author: Ion I. Moraru¹

¹ *UCONN Health*

Corresponding Author: moraru@uchc.edu

NMRbox is a shared computational platform for NMR which aims to simplify and integrate the dissemination, maintenance, support, and application of NMR data processing and analysis software packages. From the NMRbox perspective, the collection of myriad tools has enabled (1) development of meta-packages that utilize multiple tools to accomplish complex tasks that are beyond the scope of individual packages and (2) fosters critical comparison of different software packages that accomplish the same or similar tasks (e.g. comparing 11 different non-Fourier methods of spectrum analysis). BioMagResBank (BMRB) is the international repository for biological NMR data, and depositions have already enabled meta-analyses yielding novel insights into biomolecular structure and function based on chemical shifts (nuclear resonance frequencies). However depositions of other types of bioNMR data remain insufficient for meta-analysis. NMRbox utilizes the data model underlying BMRB as the organizing principle for annotating NMR data processing and analysis workflows, and consequently transparently facilitates and enriches BMRB depositions. VCell is a unique computational environment for modeling and simulation of cell biology that is deployed as a distributed application freely available over the Internet. VCell provides for “one-stop simulation shopping” whereby deterministic (compartmental ODE or reaction-diffusion-advection PDE), stochastic (several SSA solvers), spatial stochastic (reaction-diffusion with Smoldyn), hybrid deterministic/stochastic, and network-free agent based simulations can be easily created to study the same biological system/hypothesis. Model geometries may be derived from idealized analytical expressions or from experimental 2D or 3D microscope images and support for membrane flux, lateral membrane diffusion and electrophysiology is included. Inexperienced modelers can enter reactions and pathways in a biology-based interface and VCell automatically creates the mathematical system of equations according to the all possible formulations listed above. Models and simulations can be accessed from anywhere and can be shared among collaborators or made publicly available through the VCell database. NMRBox and VCell are supported by a common HPC infrastructure that exploits modern virtualization, distributed, and cloud computing technologies to provide high availability as well as software persistence, using customized, flexible, and evolving solutions for the respective projects and research communities. Both NMRbox and VCell lower the barrier to entry by non-experts in a number of ways: first, by solving the problem of software discovery and installation; second, by providing GUI “wrappers” that hide the complicated syntax utilized by many

packages and solvers; and third, by seamlessly handling data conversions and import/export operations required for interoperability within the respective software platforms or with other external tools.

VRE / 62

Framework for Developing Cloud enabled Applications for IoT

Authors: Arunachalam Bala¹; Mangala N¹; Sarat Chandara Babu¹

¹ C-DAC

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

In this paper we are proposing an Application Programming Interface (API) framework that can integrate cloud computing functionalities with Sensors to provide the on-demand computational resources, dynamic storage allocation, and database for developing cloud enabled application for Internet of Things (IoT). The functions of IoT, ie internetworking of physical devices, vehicles, buildings and other items—embedded with electronics, software, sensors, actuators, and network connectivity that enable these objects to collect and exchange data will be taken care by this API. Since these are embedded devices have limited CPU, memory and power resources. Cloud computing technology will be used for design and development of this Architecture and its APIs. Since the IoT and Cloud computing domains are an emerging area there will undoubtedly be a surge in cloud centric applications with Sensors. This API framework is designed to interface with sensor networks to transfer the required data to centralized Cloud Storage. Then the Cloud Storage serves as Database as a Service to any Cloud Applications, Mobile Apps or any other web applications.

For example, the proposed API framework will interface with Ubi-Sense sensor board developed with Air pollution and humidity sensor that supports digital output display, light sensors with infrared reacts and approximates person eyes, smoke sensor and buzzer to identify any abnormality in the above parameters. The well-known drawbacks of Sensor data such as lower visibility, non-accessibility will be addressed through this work. The Cloud computing functionalities Elasticity, Scalability, Optimal usage of resources, Minimal maintenance cost will be readily available for the IoT applications.

The advantage of using Sensors and its data are reasonable cost, power capacity and small size. Some of the Applications of Sensors are climate monitoring, battle field surveillance, air pollution monitoring, habitat monitoring, tactical surveillance, distributed computing, vehicular network, spying, etc can be supported using this framework. The applications of this work are measuring the pollution level in the office premise, temperature monitoring in the data center. The information generated from the sensor will be stored in the cloud storage. The data of sensor information will be sent as input to the web application or third party applications. The sensor information will be archived in hierarchical manner for later usage and from API will provide as Database as a Service.

Infrastructure Clouds and Virtualisation II / 63

CloudIPStore: A Cloud SaaS Repository for your Intellectual Properties

Authors: Arunachalam B¹; Battepati Kalasagar¹; Mangala N¹; Vineeth Simon Arackal¹

Co-authors: Sarat Chandra Babu ¹; Sukeshini R ¹

¹ C-DAC

There are thousands of organizations and large corporates involved in research and product development, which generate a huge number of different varieties of intellectual properties (IPs). Once these organizations reach a certain level of maturity, it becomes very important to organize and

categorize different IPs. In this paper, we describe CDAC CloudIPStore, which is a SaaS (Software as a Service) repository for the intellectual properties generated by an organization. This SaaS can be accessed through the Internet or the intranet, through a web-based graphical user interface. In the current version, it allows to store and index the patents, publications, software, trademarks and copyrights generated in the organization. Cloud IP Store allows storage, tracking, modifying, retrieving and searching of IPs. Role-based privileges for different categories of users are supported, that can be enabled or disabled. Multiple versions of Software IPs can coexist in the CloudIPStore as the SaaS offers a versioning mechanism/ version control feature

We believe that this software will be of immense use to organizations to build on their abilities and strengths. The OpenStack Swift-based cloud storage is provided in the backend. The cloud technologies employed in this SaaS enables the scalability of storage, increased availability, fault tolerance, and ubiquitous access from anywhere in the world, all of which are critical to efficient utilization of this tool to improve organization-level productivity.

66

Security Incident Response Procedure for Inter-Federation

Author: Hannah Short¹

Co-author: Romain Wartel¹

¹ CERN

Corresponding Author: hannah.short@cern.ch

How can we coordinate a response to a security incident affecting multiple national identity federations, research communities, campuses and online service providers? Who has responsibility for resolving an incident? Can we maintain confidentiality? Are the relevant capabilities for incident recovery available within the inter-federation and are these capabilities available to every participant?

This is one of the challenges being analysed by the AARC Project, <https://aarc-project.eu>, and is an essential building block for a reliable authentication and authorisation infrastructure. This poster presents the AARC proposal for building a sustainable security incident response capability for eduGAIN, the global inter-federation for research and education, in line with the existing trust model.

The proposal builds upon the requirements of Sirtfi, the Security Incident Response Trust Framework for Federated Identity, which has been widely accepted by the global community of identity federations over the past year. This framework is used to hold all participants to a common set of behaviours and guarantee the trustworthy exchange of information.

e-Science Activities in Asia Pacific I / 68

e-Sciences Activities in MAS

Authors: Batzaya E.¹; Galbaatar T.¹; Nergui B.¹

¹ Mongolian Academy of Sciences

We mainly focus on our Cultural Heritages which are old Mongolian scripts, Birch bark, Paintings and many cultural, and agricultural treasures. Mongolian Academy of Sciences has been working on the projects about saving our national treasures. Also, we have experience and cooperation with foreign Institutes and local museums and libraries. In order to archive and digitize our cultural heritages we needed delicate technologies and research. After a long time of research and procedure, we digitized hundreds of sutras and many more objects are planned.

Humanities, Arts & Social Sciences I / 69**A Preliminary Study on Reconstructing Faded Color by Spectral Estimation Method for Heritage Object****Authors:** M. James Shyu¹; Simon C. Lin²**Co-authors:** Arthur Chen³; Batzaya Enkhjargal⁴; Eric Yen²; Galbaatar T.⁴; Nergui Baasan⁴¹ *Chinese Culture University*² *ASGC*³ *Tamkang University*⁴ *Mongolian Academy of Science*

The color appearance of heritage object may reveal not only its unique style but also its culture characteristics. However as time goes by, the color of the heritage object may be changed, for example, the exposure to sunlight may cause the color on the exterior of historical building to fade gradually. This study proposes a color estimation method to recover faded color based on spectral reflectance data gathered from Bogd Khan Palace Museum in Ulaanbaatar, Mongolia. Several series of colors were measured at different locations of the building, which are under different level of shading in its structure. Photos of a white ruler were taken at the same time as a reference to indicate the intensity of sunlight exposure to each position of the measurement. The results indicate a possible application of spectral color technique to the culture heritage work. This study relies on the spectral technology in order to achieve unprecedented color restoration. Therefore, the data generated from the spectral processing technique is enormous especially when novel data analytics algorithms are required.

e-Science Activities in Asia Pacific II / 70**e-Science Activities in Thailand****Author:** Chalee Vorakulpipat¹¹ *NECTEC***Corresponding Author:** chalee.vorakulpipat@nectec.or.th

The National e-Science Infrastructure Consortium was formed to collaboratively develop computing, data storage and fundamental datasets as a sustainable infrastructure to support research in Thailand. The computing resources are used to support research projects in the areas of Computational Science and Engineering, Computer Science and Engineering, Water Resource, Energy and Environment Management, Climate Change and High Energy Particle Physics. The consortium involves members from a number of universities, research institutions, and public sectors which have committed to provide computing resources to the consortium and have participated in relevant activities to make contributions of collaboration with the consortium and to promote e-Science in Thailand.

e-Science Activities in Asia Pacific I / 71**GSDC activities for scientific computing****Author:** Sang-Un Ahn¹¹ *KISTI*

Corresponding Author: sahn@kisti.re.kr

GSDC is a national funding project to promote fundamental research fields in Korea by providing computing and networking infrastructure at KISTI. Currently we support 6 global and/or domestic experiments. In this talk, we present the status of GSDC activities about experiments support, especially WLCG Tier-1 center for ALICE experiment and the underlying system architecture and its operations.

74

Ten years of operations at the PIC Tier-1

Author: Josep Flix¹

Co-authors: Andreu Pacheco²; Antonio Pérez Calero-Yzquierdo³; Aresh Vedae²; Bruno Rodriguez²; Carles Acosta²; Elena Planas²; Esther Acción²; Fernando López³; Jordi Casals³; Manuel Delfino⁴; Marc Caubet³; Mari Carmen Porto³; Ricard Cruz²; Vanessa Acin²

¹ PIC / CIEMAT

² PIC/IFAE

³ PIC/CIEMAT

⁴ PIC/UAB

Corresponding Author: jflix@pic.es

This contribution reviews ten years of operations at the Spanish WLCG Tier-1 centre at Port d'Informació Científica (PIC), which serves the ATLAS, CMS and LHCb experiments. The implementation and the integration of all the components at the site, the reliability levels, and the resource usage will be displayed, as well as the current challenges and the ones that are expected for the next years towards the High Luminosity LHC (HL-LHC) era.

Biomedicine & Life Science I / 76

Application of non-uniform sampling method in NMR spectroscopy

Author: TsyR-Yan Yu¹

¹ *Inst. of Atomic & Molecular Sciences, Academia Sinica*

To harness the resolving power of high-field NMR spectroscopy, non-uniform sampling (NUS) methods and efficient data reconstruction methods are absolutely required to obtain NMR spectra at otherwise unreachable spectral resolution in indirect dimensions. I will describe necessity of NUS methods in high-field NMR spectroscopy and my experience using the methods developed by Wagner laboratory at Harvard Medical School [Hyberts, et al., 2011; Hyberts, et al., 2012.]. These important methods include the Poisson-gap sampling method, the Forward Maximum entropy (FM) reconstruction method [Hyberts et al., 2012], and the iterative soft thresholding (IST) method [Hyberts et al., 2011]. These methods will be used to record and reconstruct all demanding 3D or 4D experiments.

References

Hyberts S. G., Milbradt A. B., Wagner A. B., Arthanari H., Wagner G. (2011). Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap Scheduling. *J. Biomol. NMR* 52, 315-327.

Hyberts S. G., Arthanari H. and Wagner G. (2012). Applications of non-uniform sampling and processing. *Top. Curr. Chem.* 316, 125-148.

Cryo-EM Workshop / 77

General introduction to biological cryoEM

78

Image Processing in cryoEM: Open problems and current perspectives

Cryo-EM Workshop / 79

EMAN 2 (Part 1)

Cryo-EM Workshop / 80

EMAN 2 (Part 2)

Cryo-EM Workshop / 81

Leggion and Appion

Cryo-EM Workshop / 82

Image Processing in cryoEM: Open problems and current perspectives

83

Three key breakthroughs that enable cryo-EM to become a routine density map generator

Cryo-EM Workshop / 84

Relion (Part 1)

Cryo-EM Workshop / 85

Relion (Part 2)

Cryo-EM Workshop / 86

Three key breakthroughs that enable cryo-EM to become a routine density map generator

Cryo-EM Workshop / 87

Applications of cryo-electron microscopy to understand complex structures

e-Science Activities in Asia Pacific I / 88

e-Science Activities in Japan - Building Academic Inter-Cloud Infrastructure

Corresponding Author: aida@nii.ac.jp

e-Science Activities in Asia Pacific I / 89

eScience Activities in China

Corresponding Author: gang.chen@ihep.ac.cn

90

eScience Activities in Korea

e-Science Activities in Asia Pacific I / 91

e-Science Activities in Taiwan

Corresponding Author: eric.yen@twgrid.org

e-Science Activities in Asia Pacific I / 92

Q&A

e-Science Activities in Asia Pacific II / 93

e-Science Activities in Indonesia

e-Science Activities in Asia Pacific II / 94

eScience Activities in Malaysia

Corresponding Author: suhaimi@upm.my

e-Science Activities in Asia Pacific II / 95

eScience Activities in Vietnam

e-Science Activities in Asia Pacific II / 96

eScience Activities in Philippine

e-Science Activities in Asia Pacific II / 97

Q&A

e-Science Activities in Asia Pacific III / 98

eScience Activities in Singapore

e-Science Activities in Asia Pacific III / 99

e-Science Activities in India

e-Science Activities in Asia Pacific III / 100

eScience Activities in Australia

Corresponding Author: glenn.moloney@nectar.org.au

101

eScience Activities in Nepal

Corresponding Author: deep.ayadi@thethinpage.com

e-Science Activities in Asia Pacific III / 102

eScience Activities in Pakistan

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

e-Science Activities in Asia Pacific III / 103

Q&A

Opening Ceremony & Keynote Session I / 104

Opening Remarks

Opening Ceremony & Keynote Session I / 105

On-the-fly Capacity Planning in Support of High Throughput Workloads

Keynote Session III / 106

Big Data-Driven Drug Discovery

Keynote Session III / 107

High Performance Computing Environment and Applications in CAS

108

Application of non-uniform sampling method in NMR spectroscopy

Security Workshop / 109

Introduction

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

Security Workshop / 110

Security intrusions and their detection

Corresponding Author: fy@iis.sinica.edu.tw

Security Workshop / 111

How to perform forensic analysis?

Corresponding Author: vincent.brillault@cern.ch

112

How have Federated Cloud security incidents been detected?

Corresponding Author: kouril@ics.muni.cz

Security Workshop / 113

Security Incident handling in Federated Clouds

Corresponding Author: sveng@nikhef.nl

Security Workshop / 114

Introduction to the hands-on exercises

Corresponding Author: sveng@nikhef.nl

Security Workshop / 115

Hands on session I

Corresponding Author: sveng@nikhef.nl

Security Workshop / 116

Hands on session II

Corresponding Author: sveng@nikhef.nl

Security Workshop / 117

Wrap-up and conclusions

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

ICT-Enhanced Educational Workshop / 118

Introduction: Goal Setting

Corresponding Author: soetosh@gmail.com

ICT-Enhanced Educational Workshop / 119

PBL through TBL in Active Learning Way

Corresponding Author: soetosh@gmail.com

ICT-Enhanced Educational Workshop / 120

Writing Program (ESL, the mirror of the actively learning mind)

Authors: Masaki Watababe¹; Yuri Kite²

¹ *iGroup*

² *Kansai Univeristy*

ICT-Enhanced Educational Workshop / 121

Writing Program (JSL, the mirror of the learning mind)

Corresponding Author: furukawa@kansai-u.ac.jp

ICT-Enhanced Educational Workshop / 122

Fundamental Math Programs (symbol manipulation and higher order thinking)

Corresponding Authors: soetosh@gmail.com, kunio.hamamoto@gmail.com

ICT-Enhanced Educational Workshop / 123

ePortfolio for Learning Goal Settings & Artifacts of Learning Processes

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

ICT-Enhanced Educational Workshop / 124

ePortfolio for career development (Scenario Planning for the future)

Corresponding Author: soetosh@gmail.com

ICT-Enhanced Educational Workshop / 125

Hands-on session for the participants

Corresponding Author: soetosh@gmail.com

ICT-Enhanced Educational Workshop / 126

summary and wrap-up session

Corresponding Author: soetosh@gmail.com

Infrastructure Clouds and Virtualisation I / 127**Efficiency Improvement on Distributed Cloud System**

Corresponding Authors: felix@twgrid.org, eric.yen@twgrid.org

Earth, Environmental Science & Biodiversity II / 128**Future warming scenario and impacts study over Taiwan: Results from ECHAM5/MPIOM-WRF dynamical downscaling****Infrastructure Clouds and Virtualisation II / 129****Supporting Open Science with the EGI Federated Cloud - Experiences, success stories, lessons learnt**

Author: Yin Chen¹

¹ EGI

The use of cloud computing for scientific research is on the rise: majority of scientific projects and communities are either already using clouds to store, share and process research data, or considering/implementing the transition to clouds. The EGI Federated Cloud offers a scalable, flexible and highly customisable platform to cater for researchers' needs. The EGI Federated Cloud is a multinational cloud system that integrates institutional Infrastructure as a Service (IaaS) clouds into a unified computing platform which can power data and/or compute intensive applications and services. Since its start in 2014, the EGI Federated Cloud has evolved into a hybrid system composed of public, community and private clouds. The participating clouds offer OpenStack-specific and Open Standard Interfaces to research users - depending on local capabilities and on the users' preferences. The federation is enabled by the EGI operational backbone, based on capabilities such as usage accounting, service registry, service availability monitor, Virtual Machine Image marketplace. This talk will present the current status of the EGI Federated Cloud, will demonstrate examples of scientific communities and use cases that already benefit from the system, will display some of the lessons learnt while we were serving research needs, and will provide an outlook into the future of the infrastructure services. The talk will highlight opportunities for cloud providers and for research communities of the Asia-Pacific region to participate or use the EGI Federated Cloud.

Infrastructure Clouds and Virtualisation II / 130**Q&A****Network, Security, Infrastructure & Operations IV / 131****Q&A**

Data Management & Big Data / 132**Q&A****Network, Security, Infrastructure & Operations I / 133****Q&A****Biomedicine & Life Science I / 134****Image processing in cryo Electron Microscopy (cryo EM): Analyzing reliability and quality**

Author: Jose Maria Carazo Garcia¹

¹ *National Center for Biotechnology - CNB - CSIC*

Corresponding Author: carazo@cnb.csic.es

Modern electron microscopes equipped with direct electron detectors are nowadays providing 2D images (“micrographs”) with quasi atomic resolution. However, the way from 2D image collection to the calculation of accurate structural maps showing the electrostatic potential of biological macromolecules demands complex image processing operations, involving many local optimizers. Naturally, the possibility always exists of getting trapped into local minima, representing wrong structural maps. In this context, I will present several approaches to analyze reliability and quality in cryo EM, including ways to analyze precision and accuracy in estimating the relative geometrical orientation among images (a key step in the 3D reconstruction process) and a novel approach to resolution estimation, totally automatic, allowing for local (pixel-based) analysis. All these new approaches are freely available as part of both XMIPP and Scipion image processing suites.

Cryo-EM Workshop / 135**High-resolution Integrative Modelling of Biomolecular Complexes from Fuzzy Data**

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

Cryo-EM Workshop / 136**POWERFIT and DISVIS (part 1)**

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

Cryo-EM Workshop / 137

POWERFIT and DISVIS (part 2)

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

Cryo-EM Workshop / 138

Scipion (part 1)

Corresponding Author: carazo@cnb.csic.es

Cryo-EM Workshop / 139

Scipion (part 2)

Corresponding Author: carazo@cnb.csic.es

Cryo-EM Workshop / 140

Closing remark/round table discussion

Workshop on Linux Containers in Grids & Clouds / 141

Docker and dCache

Corresponding Author: paul.millar@desy.de

Workshop on Linux Containers in Grids & Clouds / 142

Container Technology and Software Delivery

Workshop on Linux Containers in Grids & Clouds / 143

Running LHC jobs using Kubernetes

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

Workshop on Linux Containers in Grids & Clouds / 144

Docker for tests and software preservation in LHCb

Corresponding Author: ben.couturier@cern.ch

Massively Distributed Computing and Citizen Sciences / 145

Citizen Earthquake Science in Taiwan

Taiwan is located at seismically highly active area that geologists called the convergent plate boundary between the Eurasian plate and the Philippine Sea plate. To bring seismology in a simple way to citizens at school and home, we are incorporating the research-based Quake-Catcher Network (QCN) program into an educational seismic network that is maintained by teachers in tens of high schools in the whole island of Taiwan. We established a web-based educational platform so that users are encouraged to interact with these collected seismic waveform data and even to conduct further signal analysis on their own.

In addition, to collect field observations for any earthquake-induced ground damages, such as surface fault rupture, landslide, rock fall, liquefaction, and landslide-triggered dam or lake, etc., we are developing an earthquake damage reporting system for public but particularly relying on trained volunteers who have taken a series of workshops, organized by this project. This Taiwan Earthquake Scientific Report (TSER) system is based on the Ushahidi mapping platform, which has been widely used for crowdsourcing.

Some online games and materials for educational purposes on learning earthquakes will be ready in a near real-time manner for students and teachers. All These constructed products are now operated at the Taiwan Earthquake Research Center (TEC). With these newly developed platforms and materials, we are aiming not only to raise the earthquake awareness and preparedness, but also to encourage public participation in earthquake science in Taiwan.

Closing Keynote & Ceremony / 146

The Emergence of Computational Archival Science

The large-scale digitization of analog archives, the emerging diverse forms of born-digital archives, and the new ways in which researchers across disciplines (as well as the public) wish to engage with archival material, are resulting in disruptions to transitional archival theories and practices. Increasing quantities of 'big archival data' present challenges for the practitioners and researchers who work with archival material, but also offer enhanced possibilities for scholarship through the application of computational methods and tools to the archival problem space, and, more fundamentally, through the integration of 'computational thinking' with 'archival thinking'. The talk will discuss these paradigm shifts in the context of e-infrastructures.

Keynote Session II / 147

Caches all the way down: Infrastructure for Data Science

The rise of big data science has created new demands for modern computer systems. While floating performance has driven computer architecture and system design for the past few decades, there is renewed interest in the speed at which data can be ingested and processed. Early exemplars such as Gordon, the NSF funded system at the San Diego Supercomputing Centre, shifted the focus from pure floating point performance to memory and IO rates. At the University of Queensland we have

continued this trend with the design of FlashLite, a parallel cluster equipped with large amounts of main memory, Flash disk, and a distributed shared memory system (ScaleMP's vSMP). This allows applications to place data "close" to the processor, enhancing processing speeds. Further, we have built a geographically distributed multi-tier hierarchical data fabric called MeDiCI, which provides an abstraction very large data stores cross the metropolitan area. MeDiCI leverages industry solutions such as IBM's Spectrum Scale and SGI's DMF platforms.

Caching underpins both FlashLite and MeDiCI. In this talk I will describe the design decisions and illustrate some early application studies that benefit from the approach.

Earth, Environmental Science & Biodiversity I / 148

Listening to the ecosystem: the integration of machine learning and a long-term soundscape monitoring network

Information on the variability of environment and biodiversity is essential for conservation management. In recent years, soundscape monitoring has been proposed as a new approach to assess the dynamics of biodiversity. Soundscape is the collection of biological sound, environmental sound, and anthropogenic noise, which provide us the essential information regarding the nature environment, behavior of calling animals, and human activities. The recent developments of recording networks facilitate the field surveys in remote forests and deep marine environments. However, analysis of big acoustic data is still a challenging task due to the lack of sufficient database to recognize various animal vocalizations. Therefore, we have developed three tools for analyzing and visualizing soundscape data: (1) long-term spectrogram viewer, (2) biological chorus detector, (3) soundscape event classifier. The long-term spectrogram viewer helps users to visualize weeks or months of recordings and evaluate the dynamics of soundscape. The biological chorus detector can automatically recognize the biological chorus without any sound template. We can separate the biological chorus and non-biological noise from a long-term spectrogram and unsupervised identify various biological events by using the soundscape event classifier. We have applied these tools on terrestrial and marine recordings collected in Taiwan to investigate the variability of environment and biodiversity. In the future, we will integrate these tools with the Asian Soundscape monitoring network. Through the open data of soundscape, we hope to provide ecological researcher and citizens an interactive platform to study the dynamics of ecosystem and the interactions among acoustic environment, biodiversity, and human activities.

ECAI Workshop / 149

Exploring the Bujang Valley - A Pre-production Survey

ECAI Workshop / 150

Atlas of Maritime Buddhism Reports on the Initial 3D Filming Projects in Myanmar, India, and Sri Lanka and Plans for Data Collection

ECAI Workshop / 151

Maritime Religious Networks: Aesthetic Sources of Flora and Fauna Motifs from India and China on Tomb Elaborations in Taiwan

ECAI Workshop / 152

Orphan Works and Collective Memories: On Setting Up The Sunflower Movement Archive

ECAI Workshop / 153

Metadata Development and Documentation for a Research Data Repository

ECAI Workshop / 154

Regional Religious System and Social Changes in China

ECAI Workshop / 155

Automatic Collective Commentaries with Corpus Positioning System

ECAI Workshop / 156

Mining Textual Image Data with Cloud Technologies: Challenges and Opportunities

ECAI Workshop / 157

Challenges and Possibilities: Translating and Building Buddhist Lexicography in the Modern Era

ECAI Workshop / 158

Dictionary Translation Project

ECAI Workshop / 159

Report on Technical Developments

ECAI Workshop / 160

Stopping the flow - The Yellow River and China's Grand Canal after 1855

ECAI Workshop / 161

Volunteered Geographic Information and Arches

ECAI Workshop / 162

Digital Economy and Asian Production Network – A Reality Check for Humanities

ECAI Workshop / 163

Endangered Languages and Flow of Identities: State Policies and Ethnic Boundaries of the Thao

ECAI Workshop / 164

History and Projections of Tombs Research in the Taiwan Area

ECAI Workshop / 165

Earth Deity Mapping and Community Networks in Taiwan

ECAI Workshop / 166

Affiliate Updates: Short reports on work in progress

ECAI Workshop / 167

Translation of Buddhist Texts: Building of Translation Memories, Terminology Databases and Corpus Analysis Tools

Environmental Computing Workshop / 168

Introduction

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

Environmental Computing Workshop / 169

Numerical Analysis on Mesoscale Dynamics of the Extreme Rainfall and Flood Event (May 2016) over Sri Lanka

Environmental Computing Workshop / 170

Q&A

Environmental Computing Workshop / 171

Parameterization Study of Chemically Reactive Pollutant Dispersion using Large-Eddy Simulation

Environmental Computing Workshop / 172

Parallel taxonomic classification algorithm for metagenomic sequences

Environmental Computing Workshop / 173

Detecting Incidents with Limited Linguistic Knowledge for Low Resource Languages

Environmental Computing Workshop / 174

Malaysia

Corresponding Author: suhaimi@upm.my

Environmental Computing Workshop / 175

Indonesia

Environmental Computing Workshop / 176

Philippines

Environmental Computing Workshop / 177

Vietnam

Environmental Computing Workshop / 178

Taiwan

Corresponding Author: eric.yen@twgrid.org

Environmental Computing Workshop / 179

Thailand

Environmental Computing Workshop / 180

Panel Discussion: State of the Art and new opportunities/challenges

APGridPMA/IGTF Meeting / 181

Introduction

Corresponding Author: eric.yen@twgrid.org

APGridPMA/IGTF Meeting / 182

EUGridPMA Update

Corresponding Author: davidg@nikhef.nl

APGridPMA/IGTF Meeting / 183

TAGPMA Update

APGridPMA/IGTF Meeting / 184

APGridPMA Update

Corresponding Author: eric.yen@twgrid.org

APGridPMA/IGTF Meeting / 185

New NAREGI CA Software System

Corresponding Author: sakane@nii.ac.jp

APGridPMA/IGTF Meeting / 186

KISTI CA Report and Review

Corresponding Author: sahn@kisti.re.kr

APGridPMA/IGTF Meeting / 187

Remote Vetting

APGridPMA/IGTF Meeting / 188

Member Report

Earth, Environmental Science & Biodiversity I / 189

Collaboration on monitoring Asian soundscape and the challenges

GDB Meeting / 190

GDB Introduction

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

GDB Meeting / 191

ASGC Report

Corresponding Author: felix@twgrid.org

GDB Meeting / 192

Asian Tier Forum report

Author: Sang Un Ahn¹

¹ *Korea Institute of Science and Technology Information*

Corresponding Author: sahn@kisti.re.kr

GDB Meeting / 193

Asian Network Status

Corresponding Author: hychen@twgrid.org

GDB Meeting / 194

Traceability & Isolation Working Group update

Corresponding Author: vincent.brillault@cern.ch

GDB Meeting / 195

IPv6 Rollout

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

GDB Meeting / 196

Sirtfi - Incident response for for Identity Federations

Corresponding Author: hannah.short@cern.ch

197

WLCG Workshop Supdate

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

GDB Meeting / 198

Tier 1 Configuration Evolution & Options

Corresponding Author: jflix@pic.es

GDB Meeting / 199

Workload Management Trends in WLCG

Corresponding Author: maarten.litmaath@cern.ch

GDB Meeting / 200

Wrap Up & WLCG Workshop update

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

Poster Session / 201

In Silico Drug Discovery of Potential HCV Helicase Inhibitors

Corresponding Author: silverbot@um.edu.my

Poster Session / 202

GROMACS in the Clouds: A Portative User-friendly Interface Bridged to Grid-computing Resources

Corresponding Author: m.e.trellet@uu.nl

Poster Session / 203

Automated Quality Control for Data from ASTI Automatic Weather Stations

Corresponding Author: jaysamuel@asti.dost.gov.ph

Poster Session / 204

ICT Enhanced Interactive Remedial Mathematics Program for Science and Engineering Students

Corresponding Author: soetosh@gmail.com

Poster Session / 205

The Virtual Reality (VR) Training System for Disaster Preparedness

Corresponding Author: deep.ayadi@thethinpage.com

Poster Session / 206

Virtualized Web Portals in EGI Federated Cloud

Corresponding Author: ljocha@ics.muni.cz

Poster Session / 207

A Switching Mechanism of Visualization Middleware and Application Using Docker

Corresponding Author: ishida.kazuya@ais.cmc.osaka-u.ac.jp

Poster Session / 208

Design and Implementation of Unified Authentication Management System of IHEP

Corresponding Author: wangli320@ihep.ac.cn

Poster Session / 209

Security Incident Response Procedure for Inter-Federation

Corresponding Author: hannah.short@cern.ch

Poster Session / 210

Ten Years of Operations at the PIC Tier-1

Corresponding Author: jflix@pic.es

Poster Session / 211

Monitoring Virtual Devices in Mass Storage Environments

Corresponding Author: tchou@bnl.gov

Poster Session / 212

Metadata as Linked Data for Research Data Repositories

Corresponding Author: cjlee@iis.sinica.edu.tw

Poster Session / 213

A Single Rack Cloud Center with Unprecedented Power and Thermal Efficiency

Corresponding Author: chihhsun.lin@phys.sinica.edu.tw

Poster Session / 214

Toward Construction of Resilient Software-Defined IT Infrastructure for Supporting Disaster Management Applications