

**International Symposium on
Grids & Clouds (ISGC) 2022
Virtual Conference**

Report of Contributions

Contribution ID: 5

Type: **Oral Presentation**

Assurance 2.0 - The Evolution of the REFEDS Assurance Suite

Tuesday, March 22, 2022 4:00 PM (30 minutes)

With the start of the global COVID-19 pandemic in 2019 we all experienced an unexpected shift of our daily life and business to the virtual. With that, collaborating services, such as videoconferencing tools or wikis started to become an integral part of our life. In order to access such tools in the higher Research and Education (R&E) space, federated access and single sign on is commonly used. Federated access and the concept of identity federations, such as national federations operated in many cases by NRENs, heavily rely on trust; i.e. trust between Federation Operators (FOs), Identity Providers (IDPs), Service Providers (SP), and users. While trust is multifaceted and requires establishment in many areas, such as organizational trust by adhering to a common set of agreements, or technical trust with the use of signatures and certificates, another important trust dimension is the trust in the user and that the user, who is accessing the service, is indeed the person who (s)he claims to be. To communicate qualitative identity and authentication information of a user, assurance information is used, with the strength being expressed by different Levels of Assurance (LoA) or 'assurance profiles'. To address the assurance needs in R&E, the REFEDS Assurance Suite was released in 2018, which comprises orthogonal components on identity assurance (the REFEDS Assurance Framework (RAF)) and authentication assurance (Single Factor Authentication profile (SFA), Multi Factor Authentication Profile (MFA)). However, one of the drawbacks identified in the REFEDS RAF identity proofing section is the usage of links to external documents, such as eIDAS, Kantara or IGTF, which makes the framework hard to understand and use. This is why the REFEDS Assurance Working Group decided to evolve the current REFEDS RAF version 1. The version 2 which is in draft status at the time of writing this abstract, will define its own criteria on Identity Proofing (while maintaining backwards compatibility) but will also make other parts of the specification more clear by bringing informative text into the framework. In addition to that, with the National Institute of Health in the U.S. being one of the driving factors, the REFEDS Assurance Working Group formed an MFA Subgroup to also provide supplementary material, particularly implementation guidance, for the REFEDS MFA Profile.

This presentation addresses the evolution of REFEDS Assurance. The talk starts with an overview of the REFEDS Assurance Suite version 1 and its specifications REFEDS RAF, SFA and MFA. The focus of this talk lies in the enhancements provided to REFEDS RAF and MFA as well as the community consultation process.

Primary author: ZIEGLER, Jule A. (Leibniz Supercomputing Centre)

Co-authors: KREMERS, Maarten (SURF); GROEP, David (Nikhef); KELSEY, David (STFC-RAL)

Presenter: ZIEGLER, Jule A. (Leibniz Supercomputing Centre)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 6

Type: **Oral Presentation**

US Network for Advanced NMR

Monday, March 21, 2022 3:00 PM (30 minutes)

NMR spectroscopy is one of the most versatile methods available for investigating matter, able to probe the composition, structure, and dynamics of complex liquid solutions and solids. Recent advances, including the development of powerful superconducting magnets employing high-temperature superconductors, are enabling advances in structural biology, metabolomics, and material science. However the barriers to accessing and using state-of-the-art NMR instrumentation remain unacceptably high. The US government funds several national and regional facilities, yet there is no single source of information on availability and capabilities of instruments. Once a scientist has identified an instrument, there are no resources enumerating best practices for sample preparation, experiment design, or processing and analysis workflows. High-field NMR remains largely the domain of specialists who have apprenticed in one of the cathedrals of NMR. The US Network for Advanced NMR (NAN) is being developed to address these and other barriers to the wider application of NMR to problems that are manifestly important. NAN will deploy state-of-the-art 1.1 GHz NMR spectrometers, for solid-state applications in Madison, WI, liquid-state applications in Athens, GA, and will be connected via a portal developed in Farmington, CT. In addition to providing resource discovery, scheduling and remote access, and resources for data stewardship, NAN will provide knowledgebases of vetted protocols for solid-state applications in structural biology and material science, and liquid-state applications for structural biology and metabolomics. When fully operational, NAN will provide access to 26 networked NMR spectrometers. Data from networked spectrometers will be automatically uploaded to a secure archive, with opt-in tools for depositing data in publicly accessible data resources. The overall goal of NAN is to democratize NMR for scientific applications in the US.

Primary author: HOCH, Jeffrey (UConn Health)

Co-authors: Prof. HENZLER-WILDMAN, Katherine (University of Wisconsin); Prof. EDISON, Arthur (University of Georgia); Prof. RIENSTRA, Chad (University of Wisconsin)

Presenter: HOCH, Jeffrey (UConn Health)

Session Classification: Health & Life Science Applications

Track Classification: Track 5: Virtual Research Environment (including tools, services, workflows, portals, ... etc.)

Contribution ID: 7

Type: **Oral Presentation**

An overview of technologies behind the Russian Scientific Data Lake prototype

Friday, March 25, 2022 11:30 AM (30 minutes)

A substantial data volume growth is expected with the start of the HL-LHC era. Even taking into account the hardware evolution it will require substantial changes to the ways data is managed and processed. The WLCG DOMA project was established to address the relevant research, and along with the national Data Lake R&Ds it studied the possible technology solutions for the organization of intelligent distributed federated storage. In this talk we will present one such R&D: the Russian Scientific Data Lake prototype, which is ongoing for the last three years, with the focus on software and technologies used for its deployment and testing.

Primary authors: ALEKSEEV, Alexandr; KIRYANOV, Andrey (NRC "Kurchatov Institute"); KLIMENTOV, Alexei; ZAROCHEVSEV, Andrey; KORCHUGANOVA, Tatiana

Presenter: KIRYANOV, Andrey (NRC "Kurchatov Institute")

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 8

Type: **Oral Presentation**

Running Highly Available SOC services on ex-worker nodes

Wednesday, March 23, 2022 1:30 PM (30 minutes)

Setting up on premise Security Operations Center (SOC) services can carry a serious initial hardware investment. To make this important piece of security more accessible, at Nikhef we have been leveraging ex worker nodes to provide a platform for a reliable elasticsearch cluster and highly available SOC services.

Over the last 1,5 year, we have experimented with various ways of deploying and running software with the least amount of interruptions as possible. We will go into our current setup and lessons learned from previous attempts.

Primary author: ROORDA, Jouke (Nikhef)

Presenter: ROORDA, Jouke (Nikhef)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 9

Type: **Oral Presentation**

Caching for Analysis Workloads of the ATLAS LHC experiment

Since the start of the Large Hadron Collider (LHC) in 2008, the Worldwide LHC Computing Grid (WLCG) has been serving the needs of the largest LHC experiments' detectors - ATLAS, CMS, LHCb, and ALICE. The volume of the data coming from these experiments every year exceeds 90 PB per year, and the rate of the raw data reaches 100 GB/s, which requires the best approaches in computing and storage system architectures.

WLCG has a hierarchical three-tier structure. The tasks of data archiving and serving, computing, and running the infrastructure services are distributed amongst the grid sites depending on the Tier they belong to.

Optimizing WLCG is crucial for the success of the whole LHC program, especially in the light of the upgrade of the accelerator – the High Luminosity LHC (HL-LHC). This will make it possible to obtain higher energies and a much greater number of collisions, which may lead physicists to new scientific discoveries. The increase in data volume and complexity outpaces the expected hardware gains, and staying within the budget requires significant improvements in the efficiency.

ATLAS is the largest of the LHC detectors and is built and operated by a collaborative effort of physicists, engineers, and technicians from around the world. The typical data processing workflow of this experiment could be roughly divided into reconstruction and analysis activities [1]. The reconstruction campaigns are aimed at preparing the raw data coming from the LHC detectors for the physics users. They are run in a well-organized, scheduled manner, while the analysis tasks are relatively sporadic, as they are independently triggered by different users according to their needs. The analysis data are stored in a form of events, which are further united into files and datasets. This hierarchical data organization implies specific data access patterns.

Caches have proved to be an effective technique to hide the latency of the file accesses and reduce the network traffic. From the detailed log information coming from the CERN Data Center (Tier 0), we extract traces of user accesses and evaluate the potential benefit of adding a caching layer under the different cache eviction policies. In particular, we compared the effects of the widely used and state-of-the-art techniques (LRU, 2-LRU) and compare them to the optimal policies (we use the PFOO-L approximation [2]). To evaluate the cache performance, we use both the Object Miss Ratio and the Bytes Miss Ratio. For the latter, we suggest a modification of the PFOO-L algorithm to estimate the optimal for the Bytes Miss Ratio. Additionally, we analyse the dependencies between the individual file accesses within the same dataset and evaluate the performance of dataset-specific policies, which are suitable for any workloads reading files in clusters (datasets).

We also consider the remote computations model, where we explore the possibility for some Tier 2 sites to move to the storageless mode and be solely used as a processing unit. In this case, the temporary storage to couple the CPU capacities will be mostly used as a cache for the processing data. In this model, we take into account the constraint on the connectivity between the storage and the processing units. This in turn may lead to the phenomenon called 'delayed hits', which affects the cache performance. In this work, we evaluate models with both file-based and dataset-based caching algorithms in order to estimate the influence on the cache performance and evaluate the connectivity bandwidth needed for the transition to the storageless sites.

[1] ATLAS, Collaboration, et al. "ATLAS Computing: technical design report." LHCC Reports; ATLAS Technical Design Reports (2005).

[2] Berger, Daniel S., Nathan Beckmann, and Mor Harchol-Balter. "Practical bounds on optimal caching with variable object sizes." Proceedings of the ACM on Measurement and Analysis of Computing Systems 2.2 (2018): 1-38.

Primary authors: CHUCHUK, Olga (CERN, University of Cote d'Azur); Dr SCHULZ, Markus (CERN); Dr NEGLIA, Giovanni (Inria)

Presenter: CHUCHUK, Olga (CERN, University of Cote d'Azur)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 10

Type: **Oral Presentation**

Software defect prediction: A study on software metrics using statistical and machine learning methods

Monday, March 21, 2022 3:30 PM (30 minutes)

Software defect prediction is aimed at identifying defect prone software modules in order to allocate testing resources [1, 2]. In the development software life cycle, software testing plays an essential role: its criticality is proved by the significant amount of spending that companies allocate to it [3]. In the last decades, furthermore, software systems are becoming more and more complex in order to meet functional or non-functional requirements [4], this complexity represents a suitable environment for defects. Several researchers have striven to develop models able to identify defective modules with the aim of reducing time and cost of software testing [5, 6]. Such models are typically trained on software measurements i.e. software metrics [7]. Software metrics are of paramount importance in the field of software engineering because they describe the characteristics of a software project such as size, complexity and code churn [8]. They reduce the subjectivity of software quality assessment and can be relied on for decision making e.g. to decide where to focus software tests [9, 10].

Our study is based on different kinds of software dataset metrics derived from different software projects [11, 12, 13]. The collected metrics belong to three main categories: size, complexity, object oriented [14, 15]; our research has highlighted the lack of consistency [16] among metrics' names: on the one hand some metrics have similar names but measure different software features, on the other hand different metrics' names measure similar software features. The involved datasets are both labelled and unlabelled i.e. they may (or may not) contain the information on the defectiveness of the software modules. Moreover, some datasets include metrics computed leveraging metrics' thresholds [17] - by default available in software application used to conduct metrics' measurement.

Software defect prediction models use both statistical and machine learning (ML) methods as described in previous literature [18, 19]. Due to the characteristics of the data, usually with a non Gaussian distribution, this work includes techniques, such as Decision Tree, Random Forest, Support Vector Machine, LASSO, Stepwise Regression. We have also employed statistical techniques that enable to compare all these algorithms by performance indicators such as precision, recall and accuracy as well as nonparametric tests [20].

To make our study available to research community, we have developed an open source and extensible R application that supports researchers to load the selected kinds of datasets, to filter them according to their features and to apply all the mentioned statistical and ML techniques.

[1] 1. Akimova EN, Bersenev AY, Deikov AA, et al. A survey on software defect prediction using deep learning. *Mathematics*. 2021;9(11):1180. doi: <http://dx.doi.org/10.3390/math9111180>.

[2] Peng He, Bing Li, Xiao Liu, Jun Chen, and Yutao Ma. 2015. An empirical study on software defect prediction with a simplified metric set. *Inf. Softw. Technol.* 59, C (March 2015), 170–190. DOI:<https://doi.org/10.1016/j.infsof.2014.11.006>

[3] S. Huda et al., "A Framework for Software Defect Prediction and Metric Selection," in *IEEE Access*, vol. 6, pp. 2844–2858, 2018, doi: 10.1109/ACCESS.2017.2785445.

[4] M. Cetiner and O. K. Sahingoz, "A Comparative Analysis for Machine Learning based Software Defect Prediction Systems," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225352.

- [4] L. Šikić, P. Afrić, A. S. Kurdija and M. Šilić, "Improving Software Defect Prediction by Aggregated Change Metrics," in *IEEE Access*, vol. 9, pp. 19391-19411, 2021, doi: 10.1109/ACCESS.2021.3054948.
- [5] Meiliana, S. Karim, H. L. H. S. Warnars, F. L. Gaol, E. Abdurachman and B. Soewito, "Software metrics for fault prediction using machine learning approaches: A literature review with PROMISE repository dataset," 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), 2017, pp. 19-23, doi: 10.1109/CYBERNETICSCOM.2017.8311708.
- [6] R. Jadhav, S. D. Joshi, U. Thorat and A. S. Joshi, "A Survey on Software Defect Prediction in Cross Project," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), 2019, pp. 1014-1019.
- [7] Wang, H., Khoshgoftaar, T.M., & Seliya, N. (2011). How Many Software Metrics Should be Selected for Defect Prediction? FLAIRS Conference.
- [8] T. Honglei, S. Wei and Z. Yanan, "The Research on Software Metrics and Software Complexity Metrics," 2009 International Forum on Computer Science-Technology and Applications, 2009, pp. 131-136, doi: 10.1109/IFCSTA.2009.39.
- [9] H. M. Olague, L. H. Etzkorn, S. Gholston and S. Quattlebaum, "Empirical Validation of Three Software Metrics Suites to Predict Fault-Proneness of Object-Oriented Classes Developed Using Highly Iterative or Agile Software Development Processes," in *IEEE Transactions on Software Engineering*, vol. 33, no. 6, pp. 402-419, June 2007, doi: 10.1109/TSE.2007.1015.
- [10] Ulan, M., Löwe, W., Ericsson, M. et al. Weighted software metrics aggregation and its application to defect prediction. *Empir Software Eng* 26, 86 (2021). <https://doi.org/10.1007/s10664-021-09984-2>
- [11] M. D'Ambros, M. Lanza and R. Robbes, "An extensive comparison of bug prediction approaches," 2010 7th IEEE Working Conference on Mining Software Repositories (MSR 2010), 2010, pp. 31-41, doi: 10.1109/MSR.2010.5463279.
- [12] Tóth Z., Gyimesi P., Ferenc R. (2016) A Public Bug Database of GitHub Projects and Its Application in Bug Prediction. In: Gervasi O. et al. (eds) *Computational Science and Its Applications – ICCSA 2016*. ICCSA 2016. Lecture Notes in Computer Science, vol 9789. Springer, Cham. https://doi.org/10.1007/978-3-319-42089-9_44
- [13] M. Shepperd, Q. Song, Z. Sun and C. Mair, "Data Quality: Some Comments on the NASA Software Defect Datasets," in *IEEE Transactions on Software Engineering*, vol. 39, no. 9, pp. 1208-1215, Sept. 2013, doi: 10.1109/TSE.2013.11
- [14] Malhotra, R., & Jain, A. (2012). Fault Prediction Using Statistical and Machine Learning Methods for Improving Software Quality. *Journal of Information Processing Systems*, 8(2), 241–262
- [15] S. R. Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," in *IEEE Transactions on Software Engineering*, vol. 20, no. 6, pp. 476-493, June 1994, doi: 10.1109/32.295895
- [16] Siket, I., Beszédes, Á., & Taylor, J. (2014). Differences in the Definition and Calculation of the LOC Metric in Free Tools.
- [17] A. Boucher, M. Badri, Software metrics thresholds calculation techniques to predict fault-proneness: An empirical comparison, *Information and Software Technology*, Volume 96, 2018, Pages 38-67, ISSN 0950-5849, <https://doi.org/10.1016/j.infsof.2017.11.005>.
- [18] Esteves, G., Figueiredo, E., Veloso, A. et al. Understanding machine learning software defect predictions. *Autom Softw Eng* 27, 369–392 (2020). <https://doi.org/10.1007/s10515-020-00277-4>
- [19] Atif, Farah & Rodriguez, Manuel & Araújo, Luiz & Amartiwi, Utih & Akinsanya, Barakat & Mazzara, Manuel. (2021). A Survey on Data Science Techniques for Predicting Software Defects. 10.1007/978-3-030-75078-7_31
- [20] Demšar, Janez Statistical Comparisons of Classifiers over Multiple Data Sets 2006 *Journal of Machine Learning Research*, Vol. 7, No. 1 p. 1-30

Primary authors: RONCHIERI, Elisabetta (INFN CNAF); Mr BERTACCINI, Gianluca (Department of Statistical Sciences, University of Bologna); Dr CANAPARO, Marco (INFN CNAF)

Presenter: Dr CANAPARO, Marco (INFN CNAF)

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)

Contribution ID: 11

Type: **Oral Presentation**

OpenForBC, GPU partitioning made easy

Thursday, March 24, 2022 3:50 PM (20 minutes)

In recent years, compute performances of GPUs (Graphics Processing Units) dramatically increased, especially in comparison to those of CPUs (Central Processing Units). Large computing infrastructures such as HPC (High Performance Computing) centers or cloud providers offer high performance GPUs to their users. GPUs are nowadays the hardware of choice for several applications involving massive parallel operations, such as deep learning (DL) and Artificial Intelligence (AI) workflows. However the programming paradigms for GPUs significantly vary according to the GPU model and vendor, often posing a barrier to their use in scientific applications. In addition, powerful GPUs such as those available in HPCs or data centers are hardly saturated by typical computing applications. The OpenForBC (Open For Better Computing) project was born in this context, and aims to ease the use of GPUs in an efficient manner. OpenForBC is an open source software framework that allows for effortless partitioning of GPUs from different vendors in Linux virtualized environments. OpenForBC supports dynamic partitioning for various configurations of the GPU, which can be used to optimise the utilisation of GPU kernels from different users or different applications. For example training complex DL models may require a full GPU, but inference may need only a fraction of it. In this contribution we describe the most common GPU partitioning options available on the market, discuss the implementation of the OpenForBC interface, and show results of benchmark tests in typical scenarios.

Primary authors: Mr BORRIERO, Alessio (INFN); LEGGER, Federica (Istituto Nazionale di Fisica Nucleare); Mr MONTELEONE, Daniele (INFN); Dr FRONZÉ, Gabriele Gaetano (INFN); Dr VALLERO, Sara (INFN); Dr BAGNASCO, Stefano (INFN); Dr LUSSO, Stefano (INFN)

Presenter: Dr FRONZÉ, Gabriele Gaetano (INFN)

Session Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators

Track Classification: Track 9: Converging High Performance Computing Infrastructures: Supercomputers, clouds, accelerators

Contribution ID: 13

Type: **Oral Presentation**

INFN Cloud: an open, extensible, federated Cloud infrastructure and service portfolio targeted to scientific communities

Friday, March 25, 2022 12:00 PM (20 minutes)

Starting at the end of 2019, INFN has been restructuring its internal organization and solutions for what regards support to computing and storage resources and services. One of the first tangible results of this effort has been the creation of INFN Cloud, a data lake-centric, heterogeneous federated Cloud infrastructure. Building on the INFN experience gained in the past 20 years with the development and operation of Grid and Cloud computing solutions for scientific communities, associated with a state-of-the-art country-wide infrastructure spanning multiple sites across Italy, INFN Cloud provides an extensible portfolio tailored to multi-disciplinary scientific communities. Based on community requirements, the INFN Cloud portfolio spans from traditional IaaS offerings to more elaborate PaaS and SaaS solutions. This contribution will discuss the INFN Cloud architecture, its governance, the security policies, the operational implementation, the main elements of the INFN Cloud portfolio, and will describe how private or public Cloud infrastructures can be federated with it.

Primary authors: COSTANTINI, Alessandro (INFN-CNAF); SALOMONI, Davide (INFN); ITALIANO, Alessandro (INFN-Bari); MARTELLI, Barbara (INFN-CNAF); SPIGA, Daniele (INFN-Perugia); CIANGOTTINI, Diego (INFN-Perugia); MICHELOTTO, Diego (INFN-CNAF); DUMA, Doina Cristina (INFN-CNAF); FANZAGO, Federica (INFN-Padova); SINISI, Francesco (INFN-CNAF); DONVITO, Giacinto (INFN-Bari); PECO, Gianluca (INFN-Bologna); GAIDO, Luciano (INFN-Torino); VERLATO, Marco (INFN-Padova); ANTONACCI, Marica (INFN-Bari); SGARAVATTO, Massimo (INFN-Padova); TRACCOLLI, Mirco (INFN-Perugia); FOGGETTI, Nadina (INFN-Bari); VERALDI, Riccardo (INFN-CNAF); VALENTINI, Roberto (INFN-Bari); TRALDI, Sergio (INFN-Padova); STALIO, Stefano (INFN-LNGS); CIASCHINI, Vincenzo (INFN-CNAF); SPINOSO, Vincenzo (INFN-Bari)

Presenter: SALOMONI, Davide (INFN)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 5: Virtual Research Environment (including tools, services, workflows, portals, ... etc.)

Contribution ID: 14

Type: **Oral Presentation**

Multiple Scenarios Oriented HTC Computing System Based on HTCondor at IHEP

Tuesday, March 22, 2022 2:00 PM (30 minutes)

IHEP is a multi-disciplinary comprehensive research institution which is hosting or attending about 15 experiments around high energy physics, including LHAASO, BES, JUNO, HEPS, DYW, ALI, ATLAS, CMS, LHCb etc.

Corresponding to the multiple experiments, in the computing system, the multiple scenarios have to be considered and the proper technologies should be suitable for the different requirements from the multiple experiments and applications. The scenarios and the corresponding tech-architectures are: 1. A big sharing pool for massive offline data processing; 2. A unified sharing pool with a special resource policy for WLCG and JUNO distributed computing grid; 3. dHTC pool for resources sharing between HTC and HPC and between sites; 4. Dedicated pools for on-site data pre-processing; 5. Realtime computing pool for astrophysics streaming data processing; 6. Customized clusters for the cooperated institutions or universities.

HTCondor is a relative popular distributed batch system in HEP computing. This paper will make discussions on the multi-scenarios HTC computing system based on HTCondor at IHEP including how to consider and design the architecture, how to take advantage of HTCondor's functions and what we have done on research and development with HTCondor. Currently, the whole computing system is managing about 50,000 CPU cores (including x86-arch CPU and arm-arch CPU) and some GPU cards.

Primary authors: JIANG, Xiaowei (Institute of High Energy Physics, Chinese Academy of Sciences); Prof. SHI, Jingyan (IHEP)

Co-authors: DU, Ran (Institute of High Energy Physics, Chinese Academy of Sciences); HU, Qingbao (IHEP); Prof. SUN, Gongxing (IHEP)

Presenter: JIANG, Xiaowei (Institute of High Energy Physics, Chinese Academy of Sciences)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 15

Type: **Oral Presentation**

Open maintenance analysis platform at IHEP

Wednesday, March 23, 2022 2:30 PM (30 minutes)

As the scale of equipment continues to grow and the computing environment becomes more and more complex, the difficulty of operation and maintenance of large-scale computing clusters is also increasing. Operation and maintenance methods based on automation technology cannot quickly and effectively solve various service failures in computing clusters. It is urgent to adopt emerging technologies to obtain all-around cluster operation and maintenance information, integrate monitoring data from multiple dimensions, and comprehensively analyze abnormal monitoring data. Based on the results of data analysis, locate the root cause of service failures and help computing clusters quickly restore services.

In order to provide a more stable cluster operating environment, IHEPCC combined big data technology and data analysis index tools to design and implement a set of open operation and maintenance analysis tool sets (OMAT), which include data collection, correlation analysis, strategy of monitoring data, abnormal alarm and other functions. This report will introduce the architecture, processing capabilities, and some key functions of OMAT. Combined with the processing flow of monitoring data, introduce the specific implementation of the system in data collection, data processing, data storage and data visualization.

The current OMAT platform has been applied to multiple cross-regional computing clusters including IHEP, covering about 5k nodes. The collected information includes node status, storage performance, network traffic, user operations, account security, power environment and other operation and maintenance indicators to ensure the computing cluster's performance Stable operation.

Primary authors: HU, Qingbao (IHEP); ZHENG, Wei (Institute of High Energy Physics, CAS); JIANG, Xiaowei (Institute of High Energy Physics, Chinese Academy of Sciences)

Presenter: HU, Qingbao (IHEP)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 16

Type: **Oral Presentation**

Research and implementation of IHEP Network performance analysis platform

Friday, March 25, 2022 2:30 PM (30 minutes)

The business application often suffers from poor performance, when it happens, business always suspect network problems, while network administrators feel innocent because the network seems fine from both the device running states and the network traffic monitoring. We do the research on IHEP network performance analysis combining the business performance monitoring. By connecting the network traffic and 7-layer network connection packets, analyzing these data to create the baseline of the normal connection of the business and alert the abnormal connections or traffic. we also use this platform to find the real bottleneck of the business.

Primary author: ZENG, SHAN (IHEP)

Presenter: ZENG, SHAN (IHEP)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 17

Type: **Oral Presentation**

Effective open-science practices for organizing a scientific software repository

Monday, March 21, 2022 2:00 PM (30 minutes)

The way we develop scientific software has drastically evolved, especially in recent years, from “scripts in a folder” to open source projects deposited in worldwide distributed community platforms. This paradigm change occurred because we, developers of scientific software, felt an increasing need to unite efforts and resources among the community. The surge of new online platforms, such as GitHub or GitLab, made this leap also possible (or was it the other way around?). Software developers and users can now interact in ways never seen before, boosting the development of projects and sparking discussions to the “commit” resolution. However, with great power comes greater responsibility. Having our code open to the wild facilitates usability and promotes collaboration and progress. But, despite most of the research projects being now open source, there is still a huge leap between a source/project that is open and a source/project that is usable and that others can build upon. In other words, we need our house clean for guests to feel comfortable and make the magic happen. Users and other developers (and our future selves) expect our project to be readable, understandable, operable, modifiable, testable. Therefore, embracing an open-source community means adopting best organizing practices if we wish our repository to shine, our community to grow, and our project to thrive. But, what are “best practices”? We refer to “best practices” as any behavior we do today that naturally solves or avoids problems in the future. Repository organizing practices are agnostic to the project scope, hence can be adopted by anyone, and englobe source organization, documentation, versioning, contributing guidelines, traceability (issues and pull requests), testing, and CI/CD. As we will discuss: source organization communicates where implementations reside and where new ones should be placed; documentation tells you everything and should be versioned as well; traceability is crucial to maintain a cohesive community and register history; testing makes you sleep well at night (aim beyond 100% coverage); CI/CD “saves” you countless time and from forgot-the-keys situations; versioning defines API expectations and grant full-reproducibility.

Maintaining such practices takes effort (at the beginning way beyond coding); yet, we shouldn't see them as a pain in the neck but as a relief, for us to sleep well at night, knowing everything is perfect. We should rewire our brains to dislike chaotic practices naturally. To share our experience in effective good practices, we selected two of our repositories HADDOCK3 and pdb-tools projects, because we believe learning by example is an excellent practice in the field of open science. Utrecht University recently awarded our pdb-tools package the AWESOME SOFTWARE badge, meaning it considered it among top packages regarding openness, reusability, and transparency, following FAIR principles and Open Science spirit. You can find HADDOCK3 and pdb-tools in the links below:

<https://github.com/haddock/haddock3> • <https://github.com/haddock/pdb-tools>

Primary authors: CORREIA TEIXEIRA, Joao Miguel (Utrecht University); BONVIN, Alexandre M.J.J. (Utrecht University)

Presenter: CORREIA TEIXEIRA, Joao Miguel (Utrecht University)

Session Classification: Health & Life Science Applications

Track Classification: Track 2: Health & Life Sciences (including Pandemic Preparedness Applications)

Contribution ID: 18

Type: **Oral Presentation**

The automatic reconstruction of the particles with machine learning at e+e- collider

Thursday, March 24, 2022 11:20 AM (20 minutes)

We have studied the automatic reconstruction of particles at e+e- collider, especially B meson, using machine learning algorithm. The main purpose of the research with e+e- collider is to precisely measure the Standard Model and search for the evidence of New Physics beyond the Standard Model. The analyses of B meson are the main objects for e+e- collider. A pair of B meson are created from e+e- collider and one of them are regarded as signal and reconstructed. Automatic reconstruction of other B makes it possible to improve the quality of events by utilizing information provided by other B, even in a situation where complete reconstruction of signal B is impossible due to non-searchable particles, such as neutrino, are included in the decay mode. In order to take advantage of automatic reconstruction, a 'tagging' method has been developed and used in B meson analyses with e+e- collider. We introduce the tagging method of 'other' B meson using machine learning algorithm. We have applied the method of automatic reconstruction and checked the effect of it when studying one of lepton-flavor-violating decay modes with simulated samples.

Primary authors: Dr KIM, Kyungho (KISTI); Prof. CHO, Kihyeon (KISTI/UST)

Presenters: Dr KIM, Kyungho (KISTI); Prof. CHO, Kihyeon (KISTI/UST)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 19

Type: **Oral Presentation**

A study of dark photon at e+ e- colliders using KISTI-5 supercomputer

Thursday, March 24, 2022 11:00 AM (20 minutes)

Dark matter is one of the most important challenges of particle physics. One of the ways to research dark matter is to identify dark matter signals in particle collision experiments. However, a large amount of simulation is required due to the very small cross-section and broad parameter space of dark matter. Therefore, it is important to reduce the Central Processing Unit (CPU) time. In order to reduce the CPU time, we have studied CPU time of simulation used for dark matter research using machines.

We have studied a few signal channels of dark photon including $A' \rightarrow \mu^+ \mu^-$ at electron-positron colliders using High Energy Physics (HEP) simulations. We have conducted this study in the stand-alone frame using KISTI-5 supercomputer (Nurion) and a local Linux machine. We have included the electron-positron collision experimental energies of Belle II, FCC-ee, CEPC, and ILC. We have generated signal events using MadGraph5 based on the Simplified Model. And the detector simulation of each experiment has been performed through Delphes. Next, physics analysis has been performed using MadAnalysis5. As a result, we have compared the physical quantities of dark photon at the generation level and the reconstruction level. Finally, the detector acceptance for each experiments have been calculated. The CEPC have showed the highest detector acceptance for the signal channel.

We have compared CPU time of simulation using the Nurion and the local machine. Nurion is composed of the KNL, a computational node, and the SKL, a CPU-only node, and both the KNL and the SKL were used in this study. We have compared how much CPU time is reduced when using one node (multiple cores) compared to when using one core by machines. When using multiple cores, CPU time decreased in the order of the KNL, the SKL, and the local Linux machine. In addition, we have compared the CPU time for each machines as the number of jobs increase on one node. As a result, the parallel processing efficiency according to the number of jobs of the Nurion have shown much better performance than that of the local Linux machine.

Primary authors: PARK, Kihong (KISTI/UST); Prof. CHO, Kihyeon (KISTI/UST)

Presenters: PARK, Kihong (KISTI/UST); Prof. CHO, Kihyeon (KISTI/UST)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 20

Type: **Oral Presentation**

Study for jet flavor tagging by using machine learning

Thursday, March 24, 2022 11:40 AM (20 minutes)

In collisions like the Large Hadron Collider (LHC), a large number of physical objects, called jets, are created. They are originated from hadrons such as gluons and quarks, and it is important to identify their origin. For example, a b-jet produced from a bottom quark has features, which can be used for its identification called a “b-tagging” algorithm, enabling precise measurement of the Higgs boson and search for other new particles from the beyond standard model.

Machine learning models have been proposed by various groups to identify jet flavors, but only for specific flavor classification, e.g., classification of the bottom quark and other quarks/gluons (b-tagging), or classification of quarks and gluons (quark and gluon separation). In this study, we propose a method and show results, where we extends the classification to all flavors: b/c/s/d/u/g at once using a modern method based on recent training methods for image recognition models.

Primary authors: TANAKA, Junichi (University of Tokyo); MORINAGA, Masahiro; SAITO, Masahiko (ICEPP, The University of Tokyo); Dr GANGULY, Sanmay; KISHIMOTO, Tomoe (University of Tokyo)

Presenter: MORINAGA, Masahiro

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 21

Type: **Oral Presentation**

Development of a Scout Job Framework for Improving Efficiency of Analysis Jobs in Belle II Distributed Computing System

Thursday, March 24, 2022 1:30 PM (30 minutes)

The Belle II experiment is a next generation B-factory experiment using an electron-positron collider, SuperKEKB. Our goal is to broadly advance our understanding of particle physics, e.g. search for physics beyond the Standard Model, precise measurements of electroweak interaction parameters, and exploring properties of the strong interaction. We started collecting collision data in 2019, aiming to acquire 50 times more data than the predecessor experiment, Belle. At the end of the data taking, we need several hundred PB disk storage and tens of thousands of CPU cores are required. To store and process the massive data with the resources, a worldwide distributed computing system is utilized. Its interconnection among end-users and heterogeneous computing resources is done by the DIRAC interware [1] and its extension Belle DIRAC to meet our requirements.

For the physics analysis, an end-user prepares analysis scripts and can submit a cluster of jobs to the system with a single command including input data sets. The system submits jobs to computing sites where input data is hosted. For more efficient analysis, it is important to use computer resources efficiently. However, about ten percent of executed jobs failed in 2019. The reason is not problems of the system but can be broadly categorized as problems in the analysis script or improper settings of the job parameters specified by the end-user. They reduce the job execution efficiency in two points. First, in our system, any jobs spend a few minutes on a worker node in downloading input files, authentication, and so on. Therefore, worker nodes are unnecessarily occupied for failed jobs. Second, when many jobs are submitted at once and they fail quickly, access to the central system is concentrated for a short time. It often triggers system trouble and solving the trouble becomes a load on the operation side.

Therefore, we have developed two features to suppress the failed jobs. For problems originating from analysis scripts, we add a syntax checker in the job submission command. This detects syntax errors at the language level of analysis scripts and notifies the end-user before the job submission. However, this is not enough to detect complicated syntax errors or the improper settings of job parameters. Therefore, we also develop a scout job framework, which submits a small number of test jobs (henceforth referred to as “scout jobs”) with the same analysis script to process a small number of events prior to massive job submissions. Then, only if the scout job succeeds, main jobs are submitted. Otherwise, all submissions of the main jobs are canceled. With these two features, we could reduce the operation load and waste of computational resources. Furthermore, this is beneficial also for end-users because the pre-test is streamlined. In the future, we aim to improve further by adding a function to automatically correct the problematic job parameters and by implementing this framework into the system for automatically generating simulation samples.

In this presentation, an overview of physical analysis in distributed computing systems, the design and operational results of the developed features, and prospects will be given.

[1] DIRAC, <https://github.com/DIRACGrid/DIRAC>

Primary authors: HIRATA, Hikari (Nagoya University); THE BELLE II COMPUTING GROUP

Presenter: HIRATA, Hikari (Nagoya University)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 22

Type: **Oral Presentation**

Running Fermi-LAT analysis on Cloud: the experience with DODAS with EGI-ACE Project

Wednesday, March 23, 2022 11:20 AM (20 minutes)

The Fermi-LAT long-term Transient (FLT) monitoring aim is the routine search of gamma-ray sources on monthly time intervals of Fermi-LAT data.

The FLT analysis consists of two steps: first the monthly data sets were analyzed using a wavelet-based source detection algorithm that provided the candidate new transient sources; finally these transient candidates were analyzed using the standard Fermi-LAT maximum likelihood analysis method. Only sources with a statistical significance above 4σ in at least one monthly bin were listed in a catalog.

The strategy adopted to implement the maximum likelihood analysis pipeline has been based on cloud solutions adopting the Dynamic On Demand Analysis Service (DODAS) service as technology enabler. DODAS represents a solution to transparently exploit cloud computing with almost zero effort for a user community.

This contribute will detail the technical implementation providing the point of view of the user community.

Primary authors: DUMA, Cristina (INFN-CNAF); SPIGA, Daniele (INFN Perugia); CIANGOTTINI, Diego (INFN Perugia); Dr MEREU, Isabella (INFN Perugia); GAIDO, Luciano (INFN Torino); ANTONACCI, Marica (INFN Bari); Dr CUTINI, Sara (INFN Perugia); NICOTRI, Stefano (INFN Bari)

Presenter: Dr MEREU, Isabella (INFN Perugia)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 23

Type: **Oral Presentation**

Distributed Computing at Belle II

Thursday, March 24, 2022 3:50 PM (20 minutes)

The Belle II experiment, an asymmetric energy electron-positron collider experiment, has a targeted integrated luminosity of 50 ab^{-1} . Data taking has already started with more than 250 fb^{-1} recorded thus far. Due to the very high data volume and computing requirements, a distributed “Grid” computing model has been adopted. Belle II recently integrated Rucio, a distributed data management software, into its workflow in order to improve scalability, automation etc. Since then, client tools have been taking advantage of Rucio to make the Belle II distributed computing system more robust. Grid job submission time is vastly improved by using the container concept of Rucio, where a single path corresponds to a collection of files, resulting in single call resolution of file paths. Replication and deletion of files are done asynchronously to improve usability. Including the completed and on-going development in this area, we will report the operation status of the Belle II distributed computing.

Primary author: PANTA, Anil (University of Mississippi)

Presenter: PANTA, Anil (University of Mississippi)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 24

Type: **Oral Presentation**

Exploiting INFN-Cloud to implement a Cloud solution to support the CYGNO computing model

Thursday, March 24, 2022 2:30 PM (30 minutes)

The aim of the CYGNO project is to demonstrate the capability of a high resolution gaseous TPC based on sCMOS (scientific CMOS) optical readout for present and future directional Dark Matter searches at low WIMP masses (1-10 GeV) down to and beyond the Neutrino Floor.

CYGNO is a typical medium size astro-particle experiment that requires a relatively small amount of computing resources and for this reason can be subjected to a fragmentation and low utilisation rate. A typical use case that could exploit and benefit from all the features of a Cloud infrastructure. In the context of the INFN-Cloud project, a container based system has been developed in order to provide a seamless integration between storage and compute system. The latter is based on JupyterHub to provide a multi user environment to access the experiment environment (ROOT, GEANT, GARFIELD++, libraries, ecc). The token based authentication and authorization system allows a seamless integration with S3 Cloud Storage where a remote DAQ system continuously uploads acquired files. The result is a Software as a Service layer for data analysis and simulation with common tools of our community.

The talk will detail the overall project and preliminary user experiences.

Primary authors: AMARO, Fernando Domingues (IBPhys, Department of Physics, University of Coimbra); ANTONACCI, Marica (Istituto Nazionale di Fisica Nucleare, Sezione di Bari); BARACCHINI, Elisabetta (Gran Sasso Science Institute); BENUSSI, Luigi (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); BIANCO, Stefano (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); CAPOCCIA, Cesidio (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); CAPONERO, Michele (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); CARDOSO, Danilo Santos (Centro Brasileiro de Pesquisas Físicas); CAVOTO, Gianluca (Dipartimento di Fisica, Università La Sapienza di Roma); CIANGOTTINI, Diego (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia); CORTEZ, André (Gran Sasso Science Institute); COSTA, Igor Abritta (Università degli Studi Roma Tre); DANÉ, Emiliano (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); DHO, Giorgio (Gran Sasso Science Institute); DI GIAMBATTISTA, Flaminia (Gran Sasso Science Institute); DI MARCO, Emanuele (Istituto Nazionale di Fisica Nucleare, Sezione di Roma); D'IMPERIO, Giulia (Istituto Nazionale di Fisica Nucleare, Sezione di Roma); DUMA, Cristina (Istituto Nazionale di Fisica Nucleare, CNAF); IACOANGELI, Francesco (Istituto Nazionale di Fisica Nucleare, Sezione di Roma); LIMA JÚNIOR, Herman Pessoa (Centro Brasileiro de Pesquisas Físicas); LOPES, Guilherme Sebastiao Pinheiro (Universidade Federal de Juiz de Fora, Faculdade de Engenharia); MACCARRONE, Giovanni (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); MANO, Rui Daniel Passos (LIBPhys, Department of Physics, University of Coimbra); MARAFINI, Michela (Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"); MARCELO GREGORIO, Robert Renz (Department of Physics and Astronomy, University of Sheffield); MARQUES, David José Gaspar (Gran Sasso Science Institute); MAZZITELLI, Giovanni (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); MCLEAN, Alasdair Gregor (Department of Physics and Astronomy, University of Sheffield); MESSINA, Andrea (Dipartimento di Fisica, Università La Sapienza di Roma); MONTEIRO, Cristina Maria Bernardes (LIBPhys, Department of Physics, University of Coimbra); NOBREGA, Rafael Antunes (Universidade Federal de Juiz de Fora, Faculdade de Engenharia); PAINS, Igor Fonseca (Uni-

versidade Federal de Juiz de Fora, Faculdade de Engenharia); PAOLETTI, Emiliano (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); PASSAMONTI, Luciano (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); PELOSI, Sandro (Istituto Nazionale di Fisica Nucleare, Sezione di Roma); PETRUCCI, Fabrizio (Dipartimento di Matematica e Fisica, Università Roma TRE); PIACENTINI, Stefano (Dipartimento di Fisica, Università La Sapienza di Roma); PICCOLO, Davide (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); PIERLUIGI, Daniele (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); PINCI, Davide (Istituto Nazionale di Fisica Nucleare, Sezione di Roma); PRAJAPATI, Atul (Gran Sasso Science Institute); RENGA, Francesco (Istituto Nazionale di Fisica Nucleare, Sezione di Roma); ROQUE, Rita Cruz (LIBPhys, Department of Physics, University of Coimbra); ROSATELLI, Filippo (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); RUSSO, Andrea (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); SANTOS, Joaquim Marques Ferreira dos (LIBPhys, Department of Physics, University of Coimbra); SAVIANO, Giovanna (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); SPIGA, Daniele (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia); SPOONER, Neil John Curwen (Department of Physics and Astronomy, University of Sheffield); STANLIO, Stefano (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Gran Sasso); TESAURO, Roberto (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); TOMASSINI, Sandro (Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati); TORELLI, Samuele (Gran Sasso Science Institute); TRACOLLI, Mirco (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia)

Presenter: COSTA, Igor Abritta (Università degli Studi Roma Tre)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 25

Type: **Oral Presentation**

Data I/O at edge sites for traditional experiments in a distributed system

Friday, March 25, 2022 2:10 PM (20 minutes)

“One platform, multi Centers” is a distributed computing platform in China managed by manpower of computing center IHEP. It consists of 10 distributed computing centers which belongs to HEP related institutes and departments. The computing center of IHEP at Beijing and the big data center of IHEP-CSNS-branch at Guangdong Province contribute to 90% of its computing and storage resources, while the other small and medium scale sites contribute the rest 10%. The platform is also capable of adding opportunistic computing resources rented from public cloud for the peak of data processing requirement.

In such a system, the small and medium sites and public cloud are designed as edge sites which have only computing resources and volatile disk storage. Therefore, remote data I/O with the two data centers is a necessity. For modern HEP experiments which access data via XRootD protocol and have mature data management systems such as RUCIO, technology stack adapted from WLCG can be reused directly. However, for traditional experiment which relies on POSIX I/O protocol and organizes dataset with namespaces of distributed file systems, data I/O at edge sites is a new and tricky task. Many HEP experiments at China belongs to this category, and have the intention to join the distributed computing platform.

We propose a solution for data I/O at edge sites of traditional experiments. This solution re-use members of XROOTD family as communication protocol over WAN, data proxy, read only cache manager and FUSE file system. Therefore, all the sites of the platform can see a same set of file systems of the two central data centers. The cache resources managed by XCache can be shared by clients of multiple file systems and experiments at the same time. A user-mapping plugin adopted from OSG is used on proxy nodes. With this plugin user identities embedded in XRootD requests can be read out and translated to local POSIX usernames. Currently, small size output is transferred back by HT-Condor while big size output has to be uploaded manually by using xrdcp. To overcome the inconvenience, a writable data cache and a data synchronization mechanism is imperative and under development.

Results of small scale tests with BESIII simulation and reconstruction at CSNS sites will be presented at the end of the presentation. It validates the feasibility and performance acceleration effect of our design.

Primary authors: LI, Haibo (Chinese); WANG, Lu (Institute of High Energy Physics, Chinese Academy of Sciece); HU, Qingbao (IHEP); CHENG, Yaodong (IHEP, CAS); BI, Yujiang (IHEP)

Presenter: WANG, Lu (Institute of High Energy Physics, Chinese Academy of Sciece)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 27

Type: **Oral Presentation**

Open-source and cloud-native solutions for managing and analyzing heterogeneous and sensitive clinical Data

Friday, March 25, 2022 1:50 PM (20 minutes)

The requirement for an effective handling and management of heterogeneous and possibly confidential data continuously increases within multiple scientific domains.

PLANET (Pollution Lake ANalysis for Effective Therapy) is a INFN-funded research initiative aiming to implement an observational study to assess a possible statistical association between environmental pollution and Covid-19 infection, symptoms and course. PLANET builds on a “data-centric” based approach that takes into account clinical components, environmental and pollution conditions, complementing primary data and many eventual confounding factors such as population density, commuter density, socio-economic metrics and more . Besides the scientific one, the main technical challenge of the project is about collecting, indexing, storing and managing many types of datasets guaranteeing FAIRness as well as adherence to the prescribed regulatory frameworks, such as the GDPR.

In this contribution we describe the developed open-source DataLake platform, detailing its key features: the event-based storage system centered on MinIO, which automates metadata processing; the data pipeline, implemented via Argo Workflows; the GraphQL-based mechanisms to query object metadata; finally, the seamless integration of the platform with a compute multi-user environment, showing how all these frameworks are integrated in the Enhanced PrIvacy and Compliance (EPIC) Cloud partition of the INFN-Cloud federation.

Primary authors: COSTANTINI, Alessandro (INFN-CNAF); MARTELLI, Barbara (INFN - CNAF); DUMA, Cristina (Istituto Nazionale di Fisica Nucleare, CNAF); SALOMONI, Davide (INFN); Dr SPIGA, daniele (INFN-PG); CIANGOTTINI, Diego (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia); RONCHIERI, Elisabetta (INFN CNAF); Mrs SERGI, Giusy (INFN-CNAF); Mr GASPARETTO, Jacopo (INFN-CNAF); Prof. STORCHI, Lorian (Dipartimento di Farmacia Universita' G. d'Annunzio); TRACOLLI, Mirco (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia); Dr LUBRANO, Pasquale (INFN -Sezione Perugia); CUITINI, Sara (INFN Perugia)

Presenter: Dr SPIGA, daniele (INFN-PG)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 28

Type: **Oral Presentation**

The Next Generation University Campus Network : Design Challenges & Opportunities

Friday, March 25, 2022 1:30 PM (30 minutes)

The University of Jammu is a very big regional University having 7 Campuses which are separated by 400 kilometers. Managing network connectivity is a big challenge. The existing network was setup in 2004 and now the user requirements especially due to the present scenario of pandemic as well as ONLINE classes as well as work from home scenario makes it a challenge.

With the emergence of New Campus Network Paradigm and advancement in Technology augmented with Big Data as well as Artificial Intelligence, it is of at most importance realizing an infrastructure suitable for emerging Campus workloads.

The Next Generation Campus Network must be designed focused on key parameters such as Automation, Scalability, open standard, security. It must have a balance between LAN and WI-FI in such a way that the availability and usage of the internet bandwidth is optimized. Also, the seven geographically diverse campuses must be seamlessly integrated. Jammu being in one of the most sensitive areas of India, security of the network, its resources and users has also to be taken care of.

The paper will discuss the key performance indicator for deciding the architecture as per the required workload in campus network. A qualitative analysis of existing technology and roadmap will be discussed with use cases. Network Virtualization techniques such as VxLAN-EVPN for Campus Network will be evaluated with use cases over legacy campus Network design.

The paper proposes to evolve a framework that shall support selection of best suited architecture and associated protocols from lowest level to highest level but also try to optimize the operation in such a way so as to strike a balance between performance key indicators for a campus network. The proposed framework would in addition, attempt to meet the end user SLAs, which take care of Quality of Service, availability and other parameter of significance. Consequently, this would involve achieving the desired degree of cohesion between involved entities.

Primary authors: GUPTA, Anik (University of Jammu, Jammu, India); Mr RANJAN, Pranjal (IIT Jammu)

Presenter: GUPTA, Anik (University of Jammu, Jammu, India)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 29

Type: **Oral Presentation**

Adapting HTCondor fairshare for mixed workloads

Friday, March 25, 2022 11:20 AM (20 minutes)

The INFN Tier-1 data centre is the main Italian computing site for scientific communities on High Energy Physics and astroparticle research. Access to the resources is arbitrated by a HTCondor batch system which is in charge of balancing the overall usage by several competing user groups according to their agreed quotas. The set of different workloads submitted to the computing cluster is highly heterogeneous and a very rich set of different requirements is to be considered by the batch system in order to provide user groups with a satisfactory fair share over the available resources. To prevent or reduce usage disparities a system to self-adjust imbalances has been developed and it is being used with satisfactory results. This work explains how and when fair share implementations can miss optimal performances and describes a general method to improve them. Results of the current solution are presented and possible further developments are discussed.

Primary authors: Mr PELLEGRINO, Carmelo (INFN); DAL PRA, Stefano (Istituto Nazionale di Fisica Nucleare)

Presenter: DAL PRA, Stefano (Istituto Nazionale di Fisica Nucleare)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 30

Type: **Oral Presentation**

Providing secure Interactive access to HTCondor batch resources with Jupyterlab

Friday, March 25, 2022 11:40 AM (20 minutes)

The Italian WLCG Tier-1 located in Bologna and managed by INFN provides batch computing resources to several research communities in the fields of High-Energy Physics, Astroparticle Physics, Gravitational Waves, Nuclear Physics and others. The capability of manually executing jobs in the Computing Nodes managed by the Batch System, as they normally would when using owned or cloud resources, is frequently requested by members of our user communities. In fact, the setup-tuning and troubleshooting of a computation campaign on a Batch System can be very difficult and time consuming as on the final Compute Node is never accessible by the end user and because job execution is generally delayed in time with respect to submission. In order to help speeding up the process of setting up the proper execution environment, most batch systems typically support interactive job submission modes, which allow users to execute arbitrary commands in the very same environment where a batch job would run. This is however usually forbidden by administrators for security and other reasons. To overcome these problems and enable users to better and more easily set up their computing tasks, while not impacting on the security policies of the Centre, a new way to deliver interactive on demand access to batch resources has been designed and implemented. This method, based on JupyterHub and JupyterLab technologies, allows users to gain interactive access in a HTCondor job slot executed on the same Compute Nodes running normal batch payloads. This approach allows user groups to address their computing problems in a timely and more effective manner, as the interactive resource being used is the very same where a normal batch job would run. Moreover, the accounting for the used batch resource is the same of a normal job and only based on the lifetime of the user session, which is then accounted as a normal job by the batch system. This work describes the project, the technical details of its implementations, also focusing on the security aspects, the early results and its possible evolution.

Primary authors: PELLEGRINO, Carmelo (INFN-CNAF); DAL PRA, Stefano (Istituto Nazionale di Fisica Nucleare)

Presenter: PELLEGRINO, Carmelo (INFN-CNAF)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 31

Type: **Oral Presentation**

Customizable Integrative Modelling Workflows with HADDOCK3

Monday, March 21, 2022 2:30 PM (30 minutes)

HADDOCK3 is a ground-up rework of our well-established flagship software for integrative modelling with as main goals improved extensibility, modularization, testability and documentation in order to allow a high degree of customization. The core routines of HADDOCK have been modularized and the underlying python interface adapted so that these core modules, as well as new ones, including third party software, can be organized in different manners to generate custom workflows. This modularity is the most significant contrast with the previous HADDOCK2.x version which is running a fixed workflow. In HADDOCK3, users and developers can create their own project-specific modules that can be easily integrated into the HADDOCK machinery to best suit their research goals. This redesign directly benefits our efforts towards exascale computing in the context of the BioExcel Center of Excellence (<https://www.bioexcel.eu>). We divided the new Python-shell architecture managing HADDOCK3 functionalities into independent building blocks, such as: non-variable physical parameters, command-line interfaces, plug-ins, the CNS code for the actual simulations, and most importantly, the functional modules configuring for each calculation stage. We can manage each of these blocks and blocks' pieces independently, adding or removing them without breaking the remaining of the software. To achieve this pipelining capacity, we have developed a dedicated input/output system that homogenizes the I/O interfaces of the modules. Thus, the only requirement to assemble two modules is that the output of the first conceptually (scientifically) matches the input of the next module. This new modular version of HADDOCK is available as an open-source repository at <https://github.com/haddock/haddock3>

Primary authors: HONORATO, Rodrigo (Utrecht University); CORREIA TEIXEIRA, Joao Miguel (Utrecht University); BONVIN, Alexandre M.J.J. (Utrecht University)

Presenter: HONORATO, Rodrigo (Utrecht University)

Session Classification: Health & Life Science Applications

Track Classification: Track 2: Health & Life Sciences (including Pandemic Preparedness Applications)

Contribution ID: 32

Type: **Oral Presentation**

A weakly-supervised method for encrypted malicious traffic detection

Monday, March 21, 2022 4:00 PM (30 minutes)

In order to defend against complex threats and attacks in cyberspace, IHEPSOC has been developed and deployed at the Institute of High Energy Physics Chinese Academy of Sciences (IHEP), which is considered as an integrated security operation platform to ensure a secure state of the network and scientific researches in IHEP. Nowadays the integration of the state-of-the-art cyber-attack detection algorithms for IHEPSOC has become an important task for enhancing the attack discovery capability of IHEPSOC. Meanwhile, malicious traffic detection comes to be increasingly challenging recently. With the extensive application of data encryption techniques to protect communication security and privacy, malicious software could also hide their attack information, making most of malicious traffic identification methods such as port-based methods and DPI-based methods ineffective.

Machine learning based detection methods have been proposed to address the encrypted malicious traffic detection problem, which usually construct statistical features of internet traffic flow and train classification models for detecting encrypted malicious traffic. There have been some drawbacks with above mentioned detection methods. On the one hand, feature selection is a time-consuming procedure that depends on expert experience. On the other hand, most traffic classification schemes employ supervised learning methods, while the acquisition of the large fine-grained labeled datasets is a tedious task. In this paper, we propose a weakly-supervised method for encrypted malicious traffic detection, which combines the generative adversarial network (GAN) and the multiple instance learning detector to achieve the fine-grained classification of encrypted traffic with a small number of coarse-grained labeled samples and a large number of unlabeled samples. Thus, we could focus on the accuracy of the malware detector instead of spending efforts on dataset annotation with the weakly-supervised learning approach. First of all, we convert the traffic data into single-channel grayscale images in the proposed method, and then input them into the GAN, so that the original traffic features can be learned without manual effort. Secondly, the improved semi-supervised learning generative adversarial network, which based upon convolution neural network (CNN) architecture, generates more synthetic samples for data augmentation, and addresses the insufficiency of labeled samples. In addition, a multi-instance detector with an attention mechanism is used to identify encrypted malicious traffic from coarse-grained labeled data. We validate the proposed approach on two datasets, a real-world dataset and a public dataset. Compared with other malicious traffic detection methods, the experimental results show that our proposed framework can effectively perform fine-grained detection of encrypted malicious traffic.

Primary authors: LIU, Junyi; LI, Zhenyu; WANG, Jiarong (Institute of High Energy Physics); YAN, Tian (IHEP); QI, Fazhi (Institute of High Energy Physics,CAS); CHEN, Gang (Institute Of High Energy Physics)

Presenter: LIU, Junyi

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)

Contribution ID: 33

Type: **Oral Presentation**

Imbalanced Malicious Traffic Detection Based on Coarse-grained Data Labels

Wednesday, March 23, 2022 2:00 PM (30 minutes)

In order to resist complex cyber-attacks, IHEPSOC is developed and deployed in the Institute of High Energy Physics of the Chinese Academy of Sciences (IHEP), which provides a reliable network and scientific research environment for IHEP. It has become a major task to integrate cutting-edge cyber-attacks detection methods for IHEPSOC to improve the ability of threat detection. Malicious traffic detection based on machine learning is an emerging security paradigm, which can effectively detect known and unknown cyber-attacks. However, the existing studies usually adopt traditional supervised learning, which often encounters some problems contrary to the implicit hypothesis in the real-world service. For example, most studies are often based on data sets that already have accurate data labels, but these labeled data sets take a lot of manual effort to carry out such accurate data labels according to the requirements. In addition, in the real-world service, the benign traffic data is much more than the malicious traffic data, and the imbalance between benign and malicious categories also makes many machine learning detection models difficult to apply in the production environment. Based on these problems, we propose an imbalanced malicious traffic detection method based on coarse-grained data labels. First of all, malicious traffic detection is modeled as a weakly supervised learning problem of multi-instance and multi-classification learning, which only needs to use coarse-grained data labels for traffic detection. Specifically, experts only need to confirm whether there is malicious traffic in the original data stream for a period of time when labeling data manually. There is no need to find out the accurate data of malicious traffic, which greatly reduces the difficulty of data labeling. In addition, in weakly supervised learning with only coarse-grained labels, the above class imbalance of benign and malicious is more serious, and the solutions to class imbalance in traditional machine learning, such as sampling, cost-sensitive functions, etc., will destroy the premise of coarse-grained labeling in weakly supervised learning. In view of this, we design a corresponding scheme to deal with the imbalance under weak supervision. The possibility of malicious traffic is pre-estimated by integrating the results of multiple clustering models and updated during the training process to shield the negative impact of the majority benign traffic in a group of coarse-grained labeled data. We change the fine-grained data labels in Android Malware dataset to coarse-grained data labels, and show that the proposed method with coarse-grained data labels outperforms traditional supervised learning method. In addition, we carried out an ablation study to verify the effectiveness of each module in our method.

Primary authors: LI, Zhenyu; LIU, Junyi; WANG, Jiarong (Institute of High Energy Physics); QI, Fazhi (Institute of High Energy Physics,CAS); YAN, Tian (IHEP)

Presenter: LI, Zhenyu

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 34

Type: **Oral Presentation**

Quantum Computing and Simulation Platform for HEP at IHEP

Thursday, March 24, 2022 2:00 PM (30 minutes)

With the dramatic growth of data in high-energy physics and the increasing demand for accuracy of physical analysis, the current classical computing model and computing power will become the bottleneck in the near future. There is an urgent need to explore new computational approaches. As a new computing device with great potential, quantum computers have become one of the possible solutions.

A quantum computer is a computing device which explicitly leverage the principles of quantum mechanics. The computing unit - qubit, can represent more information than classical bit in classical computers widely used by scientific computing nowadays. Because of the unique quantum superposition and quantum entanglement properties of quantum systems, quantum computers can naturally “parallelize” a large number of different situations and have a computing power far beyond that of classical computers.

Current quantum computers are very immature in terms of hardware manufacturing and software ecosystem. Unlike classical program development, quantum programs currently have great difficulties in problem abstraction, data processing, algorithm design and implementation due to the difference in information representation and computing model. A large amount of infrastructure needs to be researched and developed to help the development of quantum algorithm.

Motivated by the interests of LHAASO, BESIII and other HEP experiments and theory research, we have developed and deployed a quantum computing simulation platform for high-energy physics analysis combined with existing data computing clusters to provide a friendly and efficient development environment for the development of high-energy physics quantum algorithms, simplify the quantum algorithm development process, and improve the efficiency of quantum program development. Simulating large-scale quantum bits on a classical computing cluster, thus prototyping quantum algorithms and quantum software, is a more realistic approach at this stage.

Specifically, based on jupyter, we have developed and deployed a visualized interactive quantum algorithm development environment combined with the unified authentication of IHEP to help researchers in quantum algorithm development. By combining with high performance computing clusters, researchers can simulate large-scale quantum circuits and quantum annealing algorithms on our quantum simulation platform to make proof of concept of their quantum algorithms.

Together with various physical experimental and theoretical researchers, we will explore the applications of quantum computing in physical analysis and multi-body simulation, etc. We will abstract generic algorithms to form a library of generic quantum algorithm templates, and provide corresponding algorithm application examples. Users can use these examples to conduct research quickly.

Based on this platform, we will also conduct training related to quantum computing, and together promote the research and application practice of quantum computing.

Primary authors: SHI, Jingyan (IHEP); WANG, Lu (IHEP,CAS); SUN, Wei (Institute of High Energy Physics, Chinese Academy of Science); CHENG, Yaodong (IHEP, CAS); BI, Yujiang (IHEP)

Presenter: BI, Yujiang (IHEP)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 35

Type: **Oral Presentation**

Sparse-view CT reconstruction based on fusion learning in hybrid domain

Wednesday, March 23, 2022 3:30 PM (30 minutes)

In the synchrotron radiation tomography experiment, sparse-view sampling is capable of reducing severe radiation damages of samples from X-ray, accelerating sampling rate and decreasing total volume of the experimental dataset. Consequently, the sparse-view CT reconstruction has been a hot topic nowadays. Generally, there are two types of traditional algorithms for CT reconstruction, i.e., the analytic and iterative algorithms. However, the widely used analytic CT reconstruction algorithms usually lead to severe stripe artifacts in the sparse-view reconstructed images, due to the Nyquist rule is not satisfied. While the more accurate iterative algorithms often result in prohibitively high computational costs and difficulty in selecting production parameters.

Recently, using machine learning to improve the image quality of analytic algorithms is proposed as an alternative, for which multiple promising results have been successively shown. Generally, the machine learning approach of CT sparse reconstruction involves two domains, i.e., image and sinogram domain. The image domain method aims at solving the reconstructed mapping problem from the perspective of computer vision, while the sinogram domain one from the perspective of statistics and physics. For the image domain method, the performance of denoising the stripe artifacts is distinguished, while the generalization ability is relatively poor, due to the image-to-image process procedure. That method mostly employs convolution neural networks to extract features, which is lack of consideration of the global correlation of the extracted features. For the sinogram domain method, the generalization ability is rather good, due to a direct estimate of unmeasured views on the sinogram by interpolations. Nevertheless, imperfect interpolations could introduce extra artifacts. Recently, some attempts on the hybrid method, which combines image and sinogram domain methods, have been reported. Up to now, the reported hybrid method merely employed those two methods in series, which could lead to uncertain interference in the reconstruction results due to the neglect of the asymmetry of information processing during the mapping process.

In this paper, we propose a new hybrid domain method based on fusion learning. In the image domain, we employ a UNet-like network which contains the Transformer module to consider the global correlation of the extracted features. In the sinogram domain, we employ a Laplacian Pyramid network to recover unmeasured data in the sinogram, which progressively reconstructs the sub-band residuals and can reduce the quantity of network parameters. Subsequently, we employ a deep fusion network to fuse the two reconstruction results at a feature-level, which can merge the useful information of the two reconstructed images. We also compared the performances of those single-domain methods and the hybrid domain method. Experimental results indicate that the proposed method is practical and effective for reducing the artifacts and preserving the quality of the reconstructed image.

Primary authors: Dr TIAN , Haolai (Institute of High Energy Physics, CAS); Ms LI, Ling (Institute of High Energy Physics, CAS); HU, Yu (IHEP); Ms QIU, XiaoMeng (Zhengzhou university)

Co-authors: Dr DENG, Tijian (Institute of High Energy Physics, CAS); Ms HUANG, Mei (Institute of High Energy Physics, CAS); Prof. LI, Gang (Institute of High Energy Physics, CAS); Prof. QI, Fazhi (Institute of High Energy Physics, CAS)

Presenter: HU, Yu (IHEP)

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)

Contribution ID: 36

Type: **Oral Presentation**

Application of transfer learning to event classification in collider physics

Monday, March 21, 2022 4:30 PM (30 minutes)

Transfer Learning technique has been successfully applied to many scientific fields such as computer vision, natural language processing, and so on. This presentation reports an enhancement of data analysis in collider physics experiments based on this Transfer Learning technique.

Experimental particle physics aims to understand the fundamental laws of nature using a huge amount of data. In collider physics experiments, each event of data is produced from particle collisions using a high energy accelerator, such as the Large Hadron Collider. The classification of events is quite important in data analysis, where interesting signal events are separated from background events as much as possible.

Deep Learning (DL) technique has been widely used to enhance the performance of event classification by utilizing the huge parameter space of model. However, DL with such huge parameter space requires a large amount of data to maximize performance. In the field of collider physics, training data are typically generated by Monte Carlo simulations based on theories for signal and background processes. However, the simulations cost high computational power to generate a large number of events. Therefore, applying DL technique with a limited amount of data is a key concept for the collider physics experiment.

DL model consists of a stack of layers with non-linear functions. It is considered that their initial part of layers learns local features of data, and then the subsequent layers learn global features. This indicates that knowledge gained while solving one problem, such as extracting local features, can be transferred to different problems which involve common tasks. In the collider physics experiments, there are many data analysis channels for targeting different signal events. If DL can learn common knowledge or features for different data analysis channels, Transfer Learning (TL) technique should work effectively. This technique allows us to avoid training DL models from scratch by re-using weight parameters. If these weight parameters can be re-used for many data analysis channels, we can save a lot of computational power for generating the simulation data.

In this presentation, we report that the event classification can be performed with a high accuracy for different signal events by applying the TL technique. The event classification is typically based on the information of final state particles (objects) from collisions. A re-training with a small amount of data (fine-tuning) is performed to absorb differences of object topologies. For example, the number of objects in the final state is different depending on the signal process. Thus, the DL model needs to work with a variable number of objects and be insensitive to the ordering of objects to apply TL effectively. We propose a DL model to overcome these problems. The proposed model is compared to a simple multilayer perceptron model, which has a similar number of trainable parameters. Technical details of the DL model and limitations of this study are also discussed.

Primary authors: TANAKA, Junichi (University of Tokyo); SAITO, Masahiko (ICEPP, The University of Tokyo); MORINAGA, Masahiro; KISHIMOTO, Tomoe (High Energy Accelerator Research Organization)

Presenter: KISHIMOTO, Tomoe (High Energy Accelerator Research Organization)

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)

Contribution ID: 37

Type: **Oral Presentation**

Long-term data archive of IHEP: From CASTOR to EOSCTA

Friday, March 25, 2022 12:00 PM (30 minutes)

CASTOR is the primary tape storage system of CERN and has been used for over fifteen years at IHEP. By 2021, the data volume from experiments has reached 12PB. Two replicas are saved in tape for most raw data, as a result, the capacity of CASTOR has exceeded 20PB. However, numerous factors hinder the performance of CASTOR. For example, new experiments such as JUNO and HEPS, ask for long-term storage and quick increasement of data. To satisfy these requirements, we plan to replace tape storage system from CASTOR to EOSCTA. New data of LHAASO has been saved gradually in EOSCTA since late 2021. Moreover, BESIII online data and JUNO raw data will be saved directly in EOSCTA from 2022.

In this paper, we describe the current infrastructure of EOSCTA at IHEP, whose components comprise a CTA tape system as back-end, an EOS filesystem as disk buffer, a Ceph as queue manager and a PostgreSQL database. We set up two EOSCTA instances which serve for four experiments, and the front end is consisted with multiple disk file systems (LUSTRE and EOS). According to the difference of data generation, different workflows are designed to receive data from remote experimental stations or local disk arrays to EOSCTA. In order to ensure the effectiveness and reliability of EOSCTA, we have adopted several tools.

CASTOR will be replaced by EOSCTA and all the existed data of CASTOR will be migrated to EOSCTA. To achieve this target, we also have to upgrade tape drives from LTO4 to LTO7, which includes five CASTOR instance and two types of tape library. It is planned to complete most of the migration by 2023. The paper reports the migration plan, the steps and methods of data migration, and the inspection to ensure the data integrity. To elucidate the process of data migration, this paper takes the data of BESIII as an example.

Based on the previous experience of EOSCTA, we present the outlook in the requirements of experiments at IHEP, discuss a possible way to use EOSCTA as massive tape data storage for multiple data sources, and also design a unified workflow which is more suitable for local data.

Primary author: YAO, qiuling (IHEP)

Presenter: YAO, qiuling (IHEP)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 39

Type: **Oral Presentation**

The Integration of Computational Power: From CNGrid to IoSC

Thursday, March 24, 2022 4:30 PM (20 minutes)

CNGrid, the national high-performance computing environment in China, is consist of high-performance computing resources contributed by many supercomputing centers, universities and research institutes. It aims to provide high quality computing services to scientific researchers and industrial users. In last few years, CNGrid focused on upgrading its service and operation to higher levels, and better supports scientific discoveries and industrial innovation. Compared with traditional HPC clusters, the HPC environments shows its advantages in robust services and abundant software, hence the better service quality for users. CNGrid provides computing resources through unified access entry, and uses programming APIs to support the construction of specialized communities and platforms. It also helps applications of large scientific equipment by providing high throughput data processing power.

With the appearance of new concepts such as Metaverse, Computational Power Network and Wide-Area Resource Scheduling, CNGrid also needs the upgrading to the new era. Internet of Super-Computing (IoSC), the large structure composed of HPC infrastructures, storage, software and applications that are highly connected by the network, has been seen as the next generation of HPC environments. Being as the national level heterogeneous computing architecture and the national strategic reserve of computing resources in China, IoSC will be constructed to found the basement of high-performance computing research, development, application, construction, service, operation and education. The computing services provided by the IoSC will be turned from resource-oriented to task-oriented, which provides multi-form cross-domain computational power. IoSC aims to promote the connectivity of computing resources, the coordination among subjects and the merging of research work and industrial demands.

Similar to the OSI model, IoSC is designed to have multiple layers. The bottom level is the Physical Layer, which includes computing infrastructures such as clusters, storage and bandwidth. The next level is the Network Layer that is founded by the software-defined network, which greatly increases the data transmission performance across the HPC environment. The Service Layer contains a series of core HPC functionalities, including the task scheduling and the resource management for computational power, software and data in IoSC. The Presentation Layer bases on the functionalities provided by the Service Layer, and establish programming APIs, models and gateways. The top level is the Application Layer, which provides diversified and workflow-fit computing services to users through programming APIs and models.

Being as the high-end national computing facilities, IoSC will keep exploring and practicing the operation and management mechanism in the context of large scale cross-domain and cross-organization HPC environment. It will also continuously improve the evaluation system on aspects such as service level and data security.

Primary authors: Prof. CHI, Xuebin (Computer Network Information Center, Chinese Academy of Sciences); Mr XIAO, Haili (Computer Network Information Center, Chinese Academy of Sciences); Dr WANG, Xiaoning (Computer Network Information Center, Chinese Academy of Sciences); Dr ZHAO, Yining (Computer Network Information Center, Chinese Academy of Sciences)

Presenter: Prof. CHI, Xuebin (Computer Network Information Center, Chinese Academy of Sciences)

Session Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators

Track Classification: Track 9: Converging High Performance Computing Infrastructures: Supercomputers, clouds, accelerators

Contribution ID: 40

Type: **Oral Presentation**

Large scale ARM computing cluster and its application in the field of high energy physics

Thursday, March 24, 2022 3:30 PM (20 minutes)

In the field of high energy physics, with the advancement of several large-scale scientific experiments, the speed of data production is constantly increasing, and it will reach EB level in the next decade. This puts forward higher request to data storage and computation processing. In order to reduce the dependence on single type chip architecture and provide a more cost-effective storage and computing solution, we built a super computer cluster based on ARM architecture with 9600 computing cores in Dongguan, China. Based on this cluster, we have done a lot of work.

In terms of storage, we ported a large distributed storage system EOS to the ARM cluster, and incorporated the 4.7.7-aarch64 version into the official code base of EOS. In the process of software migration, the biggest challenge is that many software dependent programs do not have an officially released version of aarch64 architecture. And, due to the difference of underlying architecture, some original codes also need to be adjusted. Despite these challenges, we successfully ported it to the aarch64 architecture. The performance test is carried out.

In terms of HPC, we carried out the related work of lattice quantum chromodynamics (LQCD) based on ARM cluster, including the algorithm library transplantation and numerical simulation. QCD is the basic theory to describe strong interaction and hadron structure. Because it involves strong coupling, its strict solution is a very challenging theoretical problem. LQCD is a numerical gauge field theory method to study the properties of quarks and gluons based on the first principle of QCD. It is also a numerical simulation method to study elementary particles. Because of its huge amount of data and computing scale, it has become one of the main scientific research applications of supercomputers in the world.

In terms of HTC, we cooperated with LHAASO experiment and ported several high-energy physics data analysis softwares, such as GEANT4, Corsika, km2a, g4wcda, etc. We tested them on different architectures, like X86 and ARM, and found similar results.

In order to facilitate unified management and make better use of this ARM cluster and our existing X86 cluster, we standardized the directory structure of the system and formulated a set of scheduling strategies for heterogeneous computing clusters at remote sites. At the same time, we also deployed a monitoring system to observe the operation of ARM computing cluster and find abnormalities in time.

Primary authors: CHENG, Yaosong (Institute of High Energy Physics Chinese Academy of Sciences); BI, Yujiang (IHEP); CHENG, Yaodong (IHEP, CAS); WANG, Lu (IHEP,CAS); YAO, qiuling (IHEP); SHI, Jingyan (IHEP); HU, Qingbao (IHEP); ZHANG, minxing (IHEP)

Presenter: CHENG, Yaosong (Institute of High Energy Physics Chinese Academy of Sciences)

Session Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators

Track Classification: Track 9: Converging High Performance Computing Infrastructures:

Supercomputers, clouds, accelerators

Contribution ID: 41

Type: **Oral Presentation**

Slurm workbench: a cluster visualized research system based on Slurm REST API

Tuesday, March 22, 2022 1:30 PM (30 minutes)

As many MPI and GPU computing requirements are raised from experiments, the computing center of IHEP founded the Slurm cluster in 2017 and started the parallel computing services afterwards. Since then, users and applications of the Slurm cluster are increased from time to time. By the end of 2021, there are 16 applications and 160 users served by more than 6200 CPU cores and 200 GPU cards from the Slurm cluster.

Slurm provides command lines to submit jobs, query and control cluster status. Those commands are powerful and comprehensive. However, from the view of administrator, a well-functional cluster not only asks for command lines, but also visualized support systems. Visualized systems can help administrators to monitor real-time cluster status, generate statistics from historic job records, and submit specific pattern of jobs out of research purpose. Such visualized systems are formed to be the Slurm ecosystem on top of the cluster itself.

Slurm started to release REST APIs since version 20.02. Slurm REST APIs can be used to interact with slurmctld and slurmdbd daemons, so that job submission and cluster management can be achieved on a web interface directly. In addition, responses from slurmctld and slurmdbd in JSON could be organized properly in a favored way by cluster administrators.

This paper presents the slurm workbench system, which is developed with Python Flask based on Slurm REST APIs. Slurm workbench is consisted with three subsystems, which are dashboard, jasmine and cosmos. Dashboard is developed to display current cluster status including jobs and nodes. Jasmine is used to generate and submit specific pattern of jobs according to job parameters, which is convenient to study resource allocation and job scheduling. Cosmos is a job accounting and analysis system, with which job statistical charts are generated based on history job records. With jasmine, cosmos and dashboard working together, Slurm workbench provides a visualized way to study application and Slurm cluster.

Primary author: DU, Ran (Institute of High Energy Physics, Chinese Academy of Sciences)

Co-authors: SHI, Jingyan (IHEP); JIANG, Xiaowei (Institute of High Energy Physics, Chinese Academy of Sciences)

Presenter: DU, Ran (Institute of High Energy Physics, Chinese Academy of Sciences)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 42

Type: **Oral Presentation**

Data Analysis Integrated Software System in IHEP, design and implementation

Tuesday, March 22, 2022 3:30 PM (30 minutes)

Large scale research facilities are becoming prevalent in the modern scientific landscape. One of these facilities' primary responsibilities is to make sure that users can process and analyse measurement data for publication. In order to allow for barrier-less access to those highly complex experiments, almost all beamlines require fast feedback capable of manipulating and visualising data online to offer convenience for the decision process of the experimental strategy. And recently, the advent of beamlines at fourth-generation synchrotron sources and high resolution with high sample rate detector has made significant progress that pushes the demand for computing resources to the edge of current workstation capabilities. On top of this, most synchrotron light sources have shifted to prolonged remote operation because of the outbreak of a global pandemic, with the need for remote access to the online instrumental system during the operation. Another issue is the vast data volume produced by specific experiments makes it difficult for users to create local data copies. In this case, on-site data analysis services are necessary both during and after experiments.

Some state-of-the-art experimental techniques, such as phase-contrast tomography and ptychography approaches, will be deployed. However, it poses a critical problem of integrating this algorithmic development into a novel computing environment used in the experimental workflow. The solution requires collaboration with the user research groups, instrument scientists and computational scientists. A unified software platform that provides an integrated working environment with generic functional modules and services is necessary to meet these requirements. Scientists can work on their ideas, implement the prototype and check the results following some conventions without dealing with the technical details and the migration between different HPC environments. Thus, one of the vital considerations is integrating extensions into the software in a flexible and configurable way. Another challenge resides in the interactions between instrumental sub-systems, such as control system, data acquisition system, computing infrastructures, data management system, data storage system and so on, which can be quite complicated.

In this paper, we propose a platform for integration and automation across services and tools, which ties together existing computing infrastructure and state-of-the-art algorithms. With modular architecture, it comprises loosely coupled algorithm components that communicate over the heterogeneous in-memory data store, and scales horizontally to deliver automation at scale based on kubernetes. To produce and integrate into applications high-performance products embodying data analysis and visualization methods, the platform also has native PyQt GUIs, Web UIs based on JupyterLab, ipython CLI clients, and APIs over ZeroMQ.

Primary authors: HU, Yu (IHEP); Dr TIAN, Haolai (Institute of High Energy Physics, CAS); LI, Ling (Institute of High Energy Physics, CAS); QIU, XiaoMeng (Zhengzhou university)

Co-authors: Mr LIU, Zhibing (Institute of High Energy Physics, CAS); HU, Qingbao (IHEP); QI, Fazhi (Institute of High Energy Physics, CAS)

Presenter: Dr TIAN, Haolai (Institute of High Energy Physics, CAS)

Session Classification: VRE

Track Classification: Track 5: Virtual Research Environment (including tools, services, workflows, portals, ... etc.)

Contribution ID: 44

Type: **Oral Presentation**

Data Lake as a Service for Open Science

Friday, March 25, 2022 11:00 AM (30 minutes)

Experiments and scientists, whether in the process of designing and building up a data management system or managing multi-petabyte data historically, gather in the European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE) project to address computing challenges by developing common solutions in the context of the EOSC. A modular ecosystem of services and tools constitutes the ESCAPE Data Lake, which is exploited by flagship ESFRIs in Astroparticle Physics, Electromagnetic and Gravitational-Wave Astronomy, Particle Physics, and Nuclear Physics to pursue together the FAIR and open-access data principles.

The aforementioned infrastructure fulfils the needs of the ESCAPE community in terms of data organisation, management, and access, and addresses the required functionalities and experiment-specific use cases. As a matter of fact, dedicated assessment exercises during specific testing-focused time windows - the 2020 Full Dress Rehearsal (FDR) and the 2021 Data and Analysis Challenge (DAC) exercises - demonstrated the robustness of the pilot and prototype phases of the various Data Lake components, tools, and services, providing scientists with know-how on their management and utilisation, and evening out differences in knowledge among ESCAPE partners. A variety of challenges and specific use cases boosted ESCAPE to carefully take into account both user and infrastructure perspectives, and contributed to successfully concluding the pilot and prototype phases beyond expectations, embarking on the last stage of the project. As a result, collaborating sciences are choosing their reference implementations of the various technologies among the proposed solutions.

The prototype phase of the project aimed at consolidating the functionalities of the services, e.g. integrating token-based AuthN/Z or deploying a tailored content delivery and caching layer, and at simplifying the user experience. In this respect, a considerable effort has been devoted towards a product named DataLake-as-a-Service (DLaaS). The focus was on further integration of the Data Lake, data access, and the related data management capabilities with the activities ongoing in the area of Science Platforms, with the goal to provide the end-user with a Notebook ready-to-be-used and fully Data Lake aware. The project was framed within the CERN-HSF Google Summer of Code (GSoC) programme in 2020, and under the CERN IT-OpenLab programme in 2021. The development of a state-of-the-art “data and analysis portal” as an interface to the Data Lake offers a wide range of possibilities, from very simple I/O access to more complex workflows such as enabling content delivery technologies and integration local storage facilities at the facility hosting the Notebook. The DLaaS project allows end-users to interact with the Data Lake in an easily-understandable and user-friendly way, and it is based on the JupyterLab and JupyterHub software packages. The Rucio-JupyterLab software package that was developed during GSoC2020 is used to integrate the service with the ESCAPE Rucio instance, and the DLaaS is deployed on the same cluster hosting the other Data Lake services and tools. Examples of the features of the DLaaS include token-based OpenID Connect authentication to ESCAPE IAM, data browser, data download and upload, local storage backend access to enlarge scratch Notebook space, multiple environment options, and a content delivery low latency-hiding data access layer based on XRootD-XCache.

ESCAPE milestones achieved during the length of the project represent a fundamental accomplishment under both sociological and computing model aspects for different scientific communities that should address upcoming data management and computing challenges in the next decade.

Primary authors: HILMY, Muhammad Aditya (Institut Teknologi Bandung); DI MARIA, Riccardo (CERN)

Presenter: DI MARIA, Riccardo (CERN)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 45

Type: **Oral Presentation**

The 2021 WLCG Data Challenges

Thursday, March 24, 2022 3:30 PM (20 minutes)

The gradual approach of the High-Luminosity LHC (HL-LHC) era poses interest in the status and in the expected behaviour of the WLCG infrastructure. In particular, efficient networks and tape storage are expected pillars of success in the upcoming LHC phase. Considering the current computing models and data volume of the LHC experiments, the requirements are mainly driven by custodial storage of data on tape, export of RAW data from CERN to the Tier-1 centres, as well as data reprocessing. For this reason, an activity under the WLCG Data Organisation, Management, and Access (DOMA) forum has been started to assess the readiness of the WLCG infrastructure for HL-LHC.

A hierarchical model referred to as “Minimal”, which considers a T0-T1-T2 traffic flow only, and a realistic model referred to as “Flexible” are taken into account during the assessment exercise. They represent the range within which future planning should occur. A capability to fill around 50% of the full bandwidth for the Minimal scenario with production-like storage-to-storage traffic should be demonstrated by the HL-LHC start. Consequently, increasingly larger challenges are scheduled throughout the years anticipating HL-LHC start in 2027, aiming at upsizing targets and goals for both scenarios.

The 2021 challenge was the first of these series of challenges, pivotal to set a baseline, and represented the cornerstone of the LHC Run-3 preparation. Moreover, it provided the playground for other activities, such as the commissioning of HTTPS as a Third Party Copy (TPC) protocol instead of gsiftp at pledged sites. During the challenge, which was split into a Network Challenge and a Tape Challenge, production traffic has been backfilled by additional dedicated, centrally-managed, Data Challenge traffic to reach the 2021 target. The usage of the production infrastructure of the experiments has the benefit of testing the presence of bottlenecks across production services, focusing on the currently deployed infrastructure. Despite production services of the WLCG sites and experiments being used, innovative and original work has been produced with respect to a centralised infrastructure handling the injection of data to transfer and a common monitoring solution to provide the WLCG community with a unified picture of the four LHC experiments. Finally, this challenge boosted ongoing activities and created new task forces on specific topics that have been identified to be crucial for future challenges in order to successfully reach the HL-LHC target.

Primary authors: FORTI, Alessandra (ATLAS); DI MARIA, Riccardo (CERN); MAIER, Benedikt (CERN); GARRIDO, Borja (CERN); HAEN, Christophe (CERN); CHRISTIDIS, Dimitrios (ATLAS); GOMEZ--CORTEZ, Felipe (CMS); GARZON, Fernando (CERN); PASPALAKI, Garyfalia (CMS); ELLIS, Katy (CMS); BETEV, Latchezar (CERN); MASCETTI, Luca (CERN); RIOS, Maria Arsuaga (CERN); LASS-NIG, Mario (CERN); PATRASCOIU, Mihai (CERN); DONA, Rizart (CERN); CAMPANA, Simone (CERN); ZHAO, Xin

Presenter: DI MARIA, Riccardo (CERN)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 46

Type: **Oral Presentation**

The verification of the Science Data and Computing System for HEPS

The Science Data and Computing System is designed for High Energy Photon Source (HEPS), which is responsible for automating the organization, processing, persistent store, analysis and distribution of the data produced from experiments. It consist of several sub-systems, including storage system, network, computing system, data analysis integrated software system (Daisy) and data management system (DMS). Since this July, the design schemes of all these sub-systems have passed the peer review and the development work has almost been finished. In order to verify the functions and interfaces of the system, we set up a testbed at beamline 3W1 of Beijing Synchrotron Radiation Facility (BSRF), which is a running synchrotron radiation facility and provides the technology R&D and test platforms for HEPS.

The 3W1 beamline, which is dedicated to test high-throughput instruments for HEPS, uses an Andor detector and generates data at a relatively high rate, making it is an ideal candidate to set up the testbed where we can deploy the system and verify the functions and the whole process of data acquisition, data processing, data transfer, data storage and data access.

In this paper, we start with a brief introduction to the design of the IT services that will be provided for experimental users before, during and after the experiment at HEPS. Next, the architecture of the system will be given, that illustrates the correlation and the data flow among sub-systems. Then the implementation detail of the integration and the verification of the Science data and Computing system is described. Finally, some deficiencies are listed which exposed via the integration test at the testbed and the improving plans are put forward.

Primary authors: QI, Fazhi (Institute of High Energy Physics,CAS); HU, Hao (Institute of High Energy Physics); TIAN, Haolai; WANG, Lu (IHEP,CAS); HU, Qingbao (IHEP)

Presenter: HU, Hao (Institute of High Energy Physics)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 47

Type: **Oral Presentation**

INDIGO IAM and its Application within Research Communities

Tuesday, March 22, 2022 3:30 PM (30 minutes)

The INDIGO IAM is an Identity and Access Management service first developed in the context of the INDIGO-Datacloud Horizon 2020 project and currently maintained and developed by INFN. IAM supports user authentication via local username and password credentials as well as SAML Identity Federations, OpenID Connect providers and X.509.

The INDIGO IAM has seen adoption within a number of research communities, notably including:

- WLCG, where IAM is the AAI solution chosen to power the next generation of WLCG Authentication and Authorization infrastructure.
- the IRIS collaboration in the UK, where it is used in production as the IRIS IAM service.

This presentation will present and update on the deployment and development progress made in both communities, providing a comprehensive overview of how the INDIGO IAM provides a powerful Identity and Access tool for research communities.

Primary author: DACK, Thomas (STFC - UKRI)

Presenter: DACK, Thomas (STFC - UKRI)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 49

Type: **Oral Presentation**

Cloud native approach for Machine Learning as a Service for High Energy Physics

Wednesday, March 23, 2022 11:00 AM (20 minutes)

Nowadays Machine Learning (ML) techniques are widely adopted in many areas of High-Energy Physics (HEP) and certainly will play a significant role also in the upcoming High-Luminosity LHC (HL-LHC) upgrade foreseen at CERN. A huge amount of data will be produced by LHC and collected by the experiments, facing challenges at the exascale.

Here, we present Machine Learning as a Service solution for HEP (MLaaS4HEP) to perform an entire ML pipeline (in terms of reading data, processing data, training ML models, serving predictions) in a completely model-agnostic fashion, directly using ROOT files of arbitrary size from local or distributed data sources.

With the new version of MLaaS4HEP code based on uproot4, we provide new features to improve user's experience with framework and their workflows, e.g. user can provide some preprocessing functions to be applied on ROOT data before starting the ML pipeline.

Then our approach is extended to use local and cloud resources via HTTP proxy which allows physicists to submit their workflows using HTTP protocol. We discuss how we enabled this pipeline on INFN Cloud provider and supplement it with real use-case examples used in HEP analysis.

Primary author: GIOMMI, Luca (INFN and University of Bologna)

Co-authors: Mr KUZNETSOV, Valentin (Cornell University, Ithaca, USA); SPIGA, Daniele (INFN Perugia); Mr PALADINO, Mattia (University of Bologna); BONACORSI, Daniele (University of Bologna)

Presenter: GIOMMI, Luca (INFN and University of Bologna)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 50

Type: **Oral Presentation**

Integration of network-restricted resources at the Barcelona Supercomputer Center into CMS computing

Thursday, March 24, 2022 4:10 PM (20 minutes)

Given the growing computing needs of the LHC experiments facing the start of the Run 3 and the High-Luminosity era, it is crucial to gain access to and ensure the efficient exploitation of new resources. The CMS computing ecosystem presents a number of standardized methods for authentication and authorization, access to remotely stored data and software deployment, enabling access to WLCG resources seamlessly. However, incorporating Cloud and HPC resources presents a number of challenges, which generally need to be solved on a case by case basis. The Barcelona Supercomputing Center (BSC), the main Spanish HPC site, represents a particularly difficult case, as severe network restrictions impact the feasibility of many of the aforementioned standardized solutions. This contribution describes a number of actions and novel solutions introduced by the Spanish CMS community in order to facilitate the inclusion of BSC resources into the CMS computing infrastructure for their use by the collaboration. This includes adapting the resource allocation and workload management tools, access to CMS data processing and simulation software, and to remote experimental conditions databases, as well as setting up a transfer service for output data to storage at a nearby WLCG facility. This work summarizes the current status of the integration efforts and also reports on the first experiences with the implemented solutions.

Primary authors: Dr DELGADO PERIS, Antonio (CIEMAT); Dr PÉREZ-CALERO YZQUIERDO, Antonio (CIEMAT - PIC); Mr PINEDA, Esteve (PIC); Dr HERNÁNDEZ CALAMA, Jose Maria (CIEMAT); FLIX, Josep (PIC / CIEMAT)

Presenter: Dr PÉREZ-CALERO YZQUIERDO, Antonio (CIEMAT - PIC)

Session Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators

Track Classification: Track 9: Converging High Performance Computing Infrastructures: Supercomputers, clouds, accelerators

Contribution ID: 51

Type: **Oral Presentation**

Intelligent audio monitoring for mother and child care innovation

Tuesday, March 22, 2022 2:00 PM (30 minutes)

As the global population ages, economic conditions improve and the childcare environment improves, the views of young families on childcare and consumption have changed dramatically, with “family participation”, “scientific childcare” and “personalised education” becoming the core concepts of family childcare. “Family participation”, “scientific parenting” and “personalised education” have become the core concepts of family parenting. Some studies have shown that the sound of peristalsis, frequency of peristalsis and length of peristalsis in the gastrointestinal tract can be a very accurate predictor of a baby’s needs. Therefore, this study has designed a new parenting intelligence device based on gut sound detection to help parents take care of their babies with less time and effort.

The research uses a combination of micro-sensors and a thin and flexible smart textile flex circuit printing technology. The micro-sensors are miniature microphone modules made using MEMS (MicroElectrical-Mechanical System) technology that provide comfortable, non-invasive and accurate access to the tummy to sense the physiological signals emitted by the gases and fluids flowing in the gut. These vital physiological signals are transmitted by the Smart Textiles flexible circuitry to the soft electronic substrate of the main board, where the digital signal processor receives the signals, processes them and then sends them wirelessly via Bluetooth to the mobile device app. This allows parents of infants with the mobile device to be aware of their baby’s physical condition at the first time.

This study evaluated five smart parenting apps on the market and combined them with innovative ideas to form the functions of this study app: cloud album, vaccination, developmental assessment, growth record, nutrition recipes and parenting encyclopedia, children’s songs, sleepy songs, stories and other contents, covering various sections such as child feeding, growth and development, child care, diseases, parent-child early education and lifestyle habits.

The Kano’s Model was used to analyse the importance and attributes of the functions, combined with the QFD (Quality Function Deployment) model to assess the relationship between the technical feasibility of photos, videos, voice, text, algorithms, microphones and speakers and the needs of users, and to summarise the sequence of function development for the innovative parenting smart device. This study is based on a ‘family-centred’ approach.

The study is based on a ‘family-centred’ design philosophy to promote continuity and holistic care. An app platform system is built with information technology to allow parents to keep track of their baby’s physiological status via their mobile phones, allowing them to understand their baby’s current situation and have contact with them. This study provides a new approach to parenting for the new generation of parents, helping them to take care of their babies with less time and effort, and contributing to the development of smart device manufacturers in the baby market.

Primary authors: Dr HAN, Chen (National Taipei University of Technology, Doctoral Program in Design, College of Design); CHEN, Chun Chieh; LIU, Hsin Lan

Presenters: Dr HAN, Chen (National Taipei University of Technology, Doctoral Program in Design, College of Design); CHEN, Chun Chieh

Session Classification: Humanities, Arts & Social Sciences Applications

Track Classification: Track 4: Humanities, Arts, and Social Sciences (HASS) Applications

Contribution ID: 52

Type: **Oral Presentation**

Exploring trust for Communities

Tuesday, March 22, 2022 11:00 AM (30 minutes)

Exploring trust for Communities

Building trust for research and collaboration

When exploring the (sometimes) intimidating world of Federated Identity, research communities can reap considerable benefit from using common best practices and adopting interoperable ways of working. EnCo, the Enabling Communities task of the GEANT 4-3 Trust and Identity Work Package, provides the link between those seeking to deploy Federated Identity Management and the significant body of knowledge accumulated within the wider community. Individuals from EnCo aim to ensure that outputs from projects (e.g. AARC) and groups (e.g. WISE, FIM4R, IGTF, REFEDS) are well known, available and kept up to date as technology changes. Since many of these groups are non-funded, it's vital for their survival that projects such as GN4-3 sponsor individuals to drive progress and maintain momentum. The ultimate aim is to enhance trust between identity providers and research communities/infrastructure, to enable researchers' safe and secure access to resources.

Although this activity has been ongoing for some years, 2022 is a highly appropriate time to expose the value that EnCo has brought to our community. The GN4-3 project is reaching a close and EOSC is being built. Ensuring that shared knowledge is maintained and updated in the future will be essential for interoperability, trust and security.

The Federated Identity Management for Research (FIM4R) community is a forum where Research Communities meet to establish common requirements, combining their voices to send a strong message to FIM stakeholders. For example, in 2020 people from EnCo were among those who led efforts to produce a position paper on the EOSC identity management strategy from the perspective of research communities.

The WISE community promotes best practice in information security for IT infrastructures for research. EnCo has been and is leading several activities within WISE. This includes the Security for Collaborating Infrastructures working group, which has produced a guidance document to encourage self-assessment against the SCI Trust Framework and is working towards updating the AARC Policy Development Kit (PDK). Also, since information security processes need periodic exercise, the community organises challenges for communications response and mitigation of incidents affecting collaborative communities, and at times even deep forensics - all to make sure communities are prepared, and the various tests complement each other.

REFEDS is the voice that articulates the mutual needs of research and education identity federations worldwide. EnCo has been leading and participating in several activities on both assurance (the REFEDS Assurance Suite) and security to increase the level of trust in federations. Trust in community AARC proxy services is also promoted with the IGTF guidance on secure attribute authority operations and exchanging assurance for the infrastructures.

Our target audience are the communities and the infrastructures providing their services.

Aims of the presentation:

- The audience will learn about essential trust, policies and guidance
- Raise awareness of the availability of common resources, including those owned by WISE, FIM4R, REFEDS, IGTF
- Promote participation in these bodies and groups

- Share news of progress, e.g. assessment of SCI, Sirtfi
- Inform about future activities, e.g. moving to OAuth2.0 and away from X.509 and SAML

Primary authors: GROEP, David (Nikhef); KELSEY, David (STFC-RAL); SHORT, Hannah (CERN); KREMERS, Maarten (SURF); Mr NEILSON, Ian; ZIEGLER, Jule A. (Leibniz Supercomputing Centre)

Presenter: KREMERS, Maarten (SURF)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 53

Type: **Oral Presentation**

Contextual Design of Emotional Intelligence Behavior Analysis with Immersive Experience Concept in New Normal Condition

Tuesday, March 22, 2022 4:00 PM (30 minutes)

The dissatisfaction record of the social distancing in new normal can still be seen upon unprepared rapid-shifting situations, which reflected the decreasing emotional intelligence factor towards people who adopted the current live virtual system for alternative interaction. This research proposes the contextual design analysis to enhance people's emotional intelligence with the immersive experience concept in the virtual environment. Ten participants from different residences around Asia and European countries were interviewed to apprehend the new normal adaptation, expectation, and demand to improve the immersive experience context in the pandemic and post-pandemic scenario. We also presented the Web Extended Reality (WebXR) prototype to visualize the research concept with the aim of developing preliminary feedback for further analysis. Adopting quality function deployment (QFD) evaluation study, our result reported that people denote the unprepared condition and knowledge to reconstruct daily habits following new normal conditions, which impacted their self-conflicts on emotional intelligence. At the same time, our participants mused some feedback concerning the prototype that revealed the need of optimizing the immersive experience in live virtual functionality system into three essential approaches: engagement, interaction, and user interface (UI) to encourage the natural simulated interaction in digital-virtual space adaptation. The findings contribute to promote an innovative service by integrating user-centered analysis and immersive experience concepts. The implications provide design assumptions to assemble the possible integration of users' emotional intelligence with indirect interaction in the live virtual system due to the new normal condition.

Primary author: AINUL YAQIN, Muhammad (National Taipei University of Technology)

Co-author: Prof. WANG, Sheng-Ming (National Taipei University of Technology)

Presenter: AINUL YAQIN, Muhammad (National Taipei University of Technology)

Session Classification: VRE

Track Classification: Track 5: Virtual Research Environment (including tools, services, workflows, portals, ... etc.)

Contribution ID: 54

Type: **Oral Presentation**

A Conditional Generative Adversarial Network for Rainfall Downscaling

Tuesday, March 22, 2022 11:00 AM (30 minutes)

Predicting extreme precipitation events is one of the main challenges of climate science in this decade. Despite the continuously increasing computing availability, Global Climate Models' (GCMs) spatial resolution is still too coarse to correctly represent and predict small-scale phenomena such as convection, so that precipitation prediction is still imprecise. Indeed, precipitation shows variability on both spatial and temporal scales (much) smaller than the current state-of-the-art GCMs resolution: the precipitation field can vary strongly on spatial scales of the order of the km or even less, whereas GCMs typical spatial resolution ranges from 50 to 200 km. Therefore, downscaling techniques play a crucial role, both for the understanding of the phenomenon itself and for applications like e.g. hydrologic studies, risk prediction and emergency management.

Seen in the context of image processing, a downscaling procedure has many similarities with super-resolution tasks, i.e. the process of enhancing the resolution of an image. In recent years this scope has taken advantage of the introduction of deep learning techniques, and in particular Convolutional Neural Networks (CNNs). In our work we exploit a conditional Generative Adversarial Network (cGAN) to train a generator model to perform precipitation downscaling. This generator, a deep CNN, takes as input the precipitation field at the scale resolved by GCMs, adds random noise, and outputs a possible realization of the precipitation field at higher resolution, preserving its statistical properties with respect to the coarse-scale field. Moreover, since the generator is appropriately conditioned by the coarse-scale precipitation field, also the spatial structure of the produced small-scale precipitation field is consistent with the structure prescribed by the GCMs prediction. The GAN is being trained and tested in a "perfect model" setup, in which we try to reproduce the ERA5 precipitation field starting from an upscaled version of it.

Compared to other downscaling techniques, our model has the advantage of being computationally inexpensive at run time, since the computational load is mostly concentrated in the training phase. Furthermore the approach is robust, as it does not depend on physical, statistical and empirical assumptions. For this reason it can be applied to any geographical area, including those whose morphology limits the performance of numerical models (such as areas with complex orography) or domains subject to local phenomena which are difficult to model properly (e.g. coastal territories). Lastly, the ability of the generator model to produce a whole set of possible realizations of the small-scale precipitation field compatible with the large-scale one provides us a straightforward way to model uncertainty of the prediction, a point of primary importance in the context of the extreme events prediction and disaster management.

Primary author: Mr IOTTI, Marcello (Dept. of Physics and Astronomy, University of Bologna, Bologna, Italy)

Co-authors: Dr DAVINI, Paolo (Institute of Atmospheric Sciences and Climate CNR-ISAC, Torino, Italy); Prof. VON HARDENBERG, Jost (Dept. of Environment, Land and Infrastructure Engineering, Polytechnic University of Turin, Turin, Italy); Dr ZAPPA, Giuseppe (Institute of Atmospheric Sciences and Climate CNR-ISAC, Bologna, Italy)

Presenter: Mr IOTTI, Marcello (Dept. of Physics and Astronomy, University of Bologna, Bologna, Italy)

Session Classification: Earth & Environmental Sciences & Biodiversity Applications

Track Classification: Track 3: Earth & Environmental Sciences & Biodiversity Applications

Contribution ID: 55

Type: **Oral Presentation**

IRIS Trust and Security

Tuesday, March 22, 2022 11:30 AM (30 minutes)

IRIS is STFC's federated national digital research infrastructure program, enabling science activities supported by STFC. We have seen the threat of a cybersecurity attack against digital research infrastructures grow in recent years. This is now acute, evidenced by high profile attacks against the research and education sector in the last year. It is timely, therefore, to reflect on the development of the cybersecurity capabilities of IRIS.

We discuss the development of the IRIS security policy framework and the lessons learned about "bootstrapping" a new policy set. We also present the evolution of the IRIS operational security capabilities, and look at developments for the future. The current status of the IRIS AAI development, focussed on the IRIS IAM, will be addressed in a separate talk. Finally, we also discuss collaboration between several different projects where we are involved with trust, identity and security.

Primary authors: CROOKS, David (UKRI STFC); KELSEY, David (STFC-RAL); COLLIER, Ian (STFC-RAL); NEILSON, Ian

Presenter: CROOKS, David (UKRI STFC)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 56

Type: **Oral Presentation**

dCache integration with CTA

Friday, March 25, 2022 2:30 PM (20 minutes)

The ever increasing amount of data that is produced by modern scientific facilities like EuXFEL or LHC puts a high pressure on the data management infrastructure at the laboratories. This includes poorly shareable resources of archival storage, typically, tape libraries. To achieve maximal efficiency of the available tape resources a deep integration between hardware and software components are required.

The CERN Tape Archive (CTA) is an open-source storage management system developed by CERN to manage LHC experiment data on tape. Although today CTA's primary target is CERN Tier-0, the data management group at DESY considers the CTA as a main alternative to commercial HSM systems.

dCache has an exible tape interface which allows connectivity to any tape system. There are two ways that a le can be migrated to tape. Ether dCache calls a tape system specific copy command or through interaction via an in-dCache tape system specific driver. The latter has been shown (by TRIUMF and KIT Tier-1s), to provide better resource utilization and efficiency. Together with the CERN Tape Archive team we are working on seamless integration of CTA into dCache.

This presentation will show the design of dCache-CTA integration, current status and first test results at DESY.

Primary authors: VOSS, Christian (DESY Hamburg); Mr CHODAK, Jacek (DESY); Ms MORSCHEL, Lea (DESY); Ms SAHAKYAN, Marina (DESY); Mr GASTHUBER, Martin (DESY); Mr KARIMI, Mwai (DESY); MILLAR, Paul (DESY); Mr SUCHOWSKI, Peter (DESY); Mr LUEKEN, Ralf (DESY); Ms MEYER, Svenja (DESY); MKRTCHYAN, Tigran (DESY)

Presenter: MKRTCHYAN, Tigran (DESY)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 57

Type: **Oral Presentation**

Collaborative threat intelligence and security operations centres

Tuesday, March 22, 2022 12:00 PM (30 minutes)

The threat faced by the research and education sector from determined and well-resourced cyber attackers has been growing in recent years and is now acute. A vital means of better protecting ourselves is to share threat intelligence - key Indicators of Compromise of an ongoing incidents including network observables and file hashes - with trusted partners. We must also deploy the technical means to actively use this intelligence in the defence of our facilities, including a robust, fine-grained source of network monitoring. The combination of these elements along with storage, visualisation and alerting is called a Security Operations Centre (SOCs).

We report on recent progress of the SOC WG, mandated to create reference designs for these SOC's, with particular attention to work being carried out at multiple 100Gb/s sites to deploy these technologies and a proposal to leverage passive DNS in order to further assist sites of various sizes to improve their security stance.

We discuss the plans for this group for the coming year and the importance of acting together as a community to defend against these attacks.

Primary authors: ARVANITIS, Christos (CERN); CROOKS, David (UKRI STFC); VALSAN, Liviu (CERN); WARTEL, Romain (CERN)

Presenters: CROOKS, David (UKRI STFC); VALSAN, Liviu (CERN)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 58

Type: **Oral Presentation**

Network bandwidth guarantee of data transmission in high energy physics experiment

Friday, March 25, 2022 2:00 PM (30 minutes)

High energy physics experiment has the characteristics of remote construction. It is often necessary to transfer a large number of experimental data from high energy physics experiment equipment to data center with the help of network link. The network links supporting the data transmission of high energy physics experiments often have the characteristics of sharing. The data from different experiments compete for link resources in the transmission. Therefore, this report mainly focuses on how to provide transmission guarantee for the data generated by a certain high-energy physics application in a certain period of time, that is, to ensure that it occupies enough available bandwidth so that it can transmit the experimental data to the data center of IHEP as soon as possible. On the high-energy physics experiment data transmission chain, a bandwidth estimation model is established for the experimental data generated by specific high-energy physics applications, and a supervised machine learning method is adopted to provide guidance and suggestions for specific applications to retain the specific amount of bandwidth when transmitting data, and verify it. The research on the network bandwidth guarantee of data transmission in high-energy physics experiment can improve the transmission efficiency of experimental data and promote the production of high-energy physics experiment results and have certain practical significance.

Primary author: Ms LIU, Yi (Institute of High Energy Physics, CAS)

Presenter: Ms LIU, Yi (Institute of High Energy Physics, CAS)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 59

Type: **Oral Presentation**

The ML_INFN Initiative

Wednesday, March 23, 2022 4:00 PM (30 minutes)

The ML_INFN initiative (“Machine Learning at INFN”) is an effort to foster Machine Learning activities at the Italian National Institute for Nuclear Physics (INFN). In recent years, AI inspired activities have flourished bottom-up in many efforts in Physics, both at the experimental and theoretical level. Many researchers have procured desktop-level devices, with consumer oriented GPUs, and have trained themselves in a variety of ways, from webinars, books, tutorials. ML_INFN aims to help and systematize such effort, in multiple ways: by offering state-of-the art hardware for Machine Learning, leveraging on the INFN-Cloud provisioning solutions and thus sharing more efficiently GPU-like resources and leveling the access to such resources to all INFN researchers, and by organizing and curating Knowledge Bases with production grade examples from successful activities already in production. On top of that, training events have been organized for beginners, based on existing INFN ML research and focussed on lowering the bar for beginners. We report on the status of the project after two years of intense activity and the plans for its conclusion and legacy.

Primary authors: Dr LUCIO, Anderlini (INFN Firenze); DUMA, Doina Cristina (INFN - CNAF); DAL PRA, Stefano (Istituto Nazionale di Fisica Nucleare); BOCCALI, tommaso (INFN)

Presenters: Dr LUCIO, Anderlini (INFN Firenze); DUMA, Doina Cristina (INFN - CNAF); DAL PRA, Stefano (Istituto Nazionale di Fisica Nucleare); BOCCALI, tommaso (INFN)

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)

Contribution ID: 60

Type: **Oral Presentation**

On-demand scheduling of GPUs for CI/CD with Kubernetes on Openstack

Friday, March 25, 2022 11:00 AM (20 minutes)

Machine Learning and Artificial Intelligence tools have become more and more commonplace in research over the last years and a growing need for organising models and source codes emerges. For the latter task, there are several version control tools, of which git is the de-facto standard. Together with Continuous Integration and Continuous Deployment (CI/CD) pipelines, however, git unfolds a lot more potential. We have established a GitLab platform with elastic CI/CD services for our researchers over the last years and now additionally provide GPUs for CI/CD workflows and pipelines for building, performance measurements and regression testing.

We are running an on-premise Openstack IaaS cloud for virtual machine provisioning, which we use to spawn manage Kubernetes clusters as a platform for deploying CI/CD runners. This allows easy scaling and straightforward integration of different hardware components under a common API.

This portability allows to interface with federated resources from large collaborations such as EOSC, EGI, and platforms within the Helmholtz Association (HIFIS, HIP).

In this talk, we will provide detail on our infrastructure setup, and elaborate on demo use cases for training networks with our platform.

Primary authors: SCHUH, Michael (DESY); WETZEL, Tim (Deutsches Elektronen-Synchrotron DESY); FUHRMANN, Patrick (DESY/dCache.org)

Co-authors: Dr RAJPUT, Humaira (DESY); Mr BUJACK, Stefan (DESY); Dr KEMP, Yves (DESY); Mr WENGERT, Markus (DESY); Dr REPPIN, Johannes (DESY)

Presenter: WETZEL, Tim (Deutsches Elektronen-Synchrotron DESY)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 61

Type: **Oral Presentation**

Implementation of CMSWEB Services Deployment Procedures using HELM

Wednesday, March 23, 2022 11:40 AM (20 minutes)

The Compact Muon Solenoid (CMS) experiment heavily relies on the CMSWEB cluster to host critical services for its operational needs. Recently, there has been migration of the CMSWEB cluster from the VM cluster to the Kubernetes (k8s) cluster. The new cluster of CMSWEB in Kubernetes enhances sustainability and reduces the operational cost. In this work, we added new features to the CMSWEB k8s cluster. The new features include the deployment of services using Helm's chart templates and the incorporation of canary releases using Nginx ingress weighted routing that is used to route traffic to multiple versions of the services simultaneously. The usage of Helm simplifies the deployment procedure and no expertise of Kubernetes is needed anymore for service deployment. Helm packages all dependencies and services are easily deployed, updated and rolled back. Helm enables us to deploy multiple versions of the services to run simultaneously. This feature is very useful for developers to test the new versions of the services by assigning some weight to the new service version and rollback immediately in case of issues. Using Helm, we can also deploy different application configurations at runtime.

Primary authors: Dr PFEIFFER, Andreas (CERN); Ms PERVAIZ, Aroosha (CERN); Dr IMRAN, Muhammad (CERN); Mr PAPARRIGOPOULOS, Panos (CERN); Mr TRIGAZIS, Spyridon (CERN); KUZNETSOV, Valentin (Cornell University, Ithaca, USA)

Presenter: Dr IMRAN, Muhammad (CERN)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 62

Type: **Oral Presentation**

Applying 5G Positioning Service for Digital Out-of-Home Advertising Innovation

Tuesday, March 22, 2022 2:30 PM (30 minutes)

This study focuses on developing the application of advertising innovation in digital out-of-home by using the 5G positioning service. 5G includes a new standard for services around the geographic position of objects, with significant improvements on accuracy and other performance parameters. The services are often called 'positioning services,' unlike 'location services' used for earlier generations. The new 3GPP REST API provides advanced positioning features, including better accuracy, address mapping, velocity, vertical positioning, and more. Digital Out-of-Home (DOOH) advertising is an interactive and eye-catching advertising strategy that empowers brands to digitize themselves and display content easily accessible to the general public. The real reason DOOH advertising is progressing way faster than anticipated is reaching the target audience in real-time. As we can see by the developing trend, with the coming of 5G, programmatic bidding and, AR, DOOH is rapidly evolving beyond its initial position as a video version of a static billboard.

Furthermore, some research results suggested that outdoor digital screens can be seen as an extension of social media. Mainly, the pairing of outdoor digital screens and social media on mobile devices has emerged as a distinct kind of user experience. Thus, the primary purpose of this study is to explore the integration of the 5G positioning service and DOOH for developing future advertising innovation. The study begins with using the service design method to explore how to create innovative service models for DOOH advertising by taking advantage of the characteristics of the 5G positioning service. Then, we propose a system architecture for developing an intelligent interactive advertisement service by using the high-density small cell feature of the 5G network to improve the inability of locating users in real-time. Thirdly, we build a DSP (Demand-side platform) and a DMP (Data management platform) to detect, collect and analyze users' "space" and "time" data. Finally, we set up the innovative advertising service according to the needs of different owners, the data filtering, potential audience discovery, interaction mechanism, and the flow of contents. The results of this study provide the solution to improve the problems of "the inability to deliver advertising content precisely" and "the uncertainty of advertising publishing expectation."

Primary authors: WANG, Ryan Sheng-Ming (Department of Interaction Design, National Taipei University of Technology); HSIAO, Yi Ting; HAN, chen

Presenter: HSIAO, Yi Ting

Session Classification: Humanities, Arts & Social Sciences Applications

Track Classification: Track 4: Humanities, Arts, and Social Sciences (HASS) Applications

Contribution ID: 63

Type: **Oral Presentation**

Design and Implementation of Token-based Authentication and Authorization System in High Performance Computing Infrastructure in Japan

Tuesday, March 22, 2022 4:30 PM (30 minutes)

Token-based technologies are attracting attention in order to realize authentication and authorization in distributed high-performance computing infrastructure for research. The purpose of this paper is to describe the design and implementation of the next authentication and authorization system in High Performance Computing Infrastructure (HPCI) in Japan.

Following the end of GSI (Grid Security Infrastructure) maintenance by the Globus Alliance, authentication and authorization technology to replace GSI is being considered worldwide for large-scale high performance computing environments. In Japan as well, HPCI uses currently GSI to realize single sign-on (SSO) among high performance computers and large-scale distributed file systems, therefore we have studied authentication technologies for the next authentication infrastructure that does not use GSI. As a result OAuth has been selected as the main authentication technology.

In order to use OAuth tokens for SSO among supercomputers and large-scale distributed file systems, it is necessary to skillfully delegate tokens. We must consider that in accesses to those resources there is not just Web user interface but command-line user interface (CUI), because end-users of HPCI typically log in to the front-end of a supercomputer with SSH and mount a distributed filesystem. We discuss token flows for typical CUI-based use cases in HPCI, which we consider beneficial for the other large-scale HPC infrastructure.

In this paper, we describe the details of token-based authentication and authorization system in HPCI such as access token details, token issuance and user information management, token processing in SSH and distributed file system, authorization management for services, and end-user environment.

Primary author: Dr SAKANE, Eisaku (National Institute of Informatics)

Co-author: AIDA, Kento (National Institute of Informatics)

Presenter: Dr SAKANE, Eisaku (National Institute of Informatics)

Session Classification: Network, Security, Infrastructure & Operations

Track Classification: Track 7: Network, Security, Infrastructure & Operations

Contribution ID: 64

Type: **Oral Presentation**

Simulating a network delivery content solution for the CMS experiment in the Spanish WLCG Tiers

Friday, March 25, 2022 1:30 PM (20 minutes)

The LHC experiments at CERN, the world's largest particle collider, have produced an unprecedented volume of data in the history of modern science since it started operations in 2009. Up to now, more than 1 Exabyte of simulated and real data have been produced, being stored on disk and magnetic support and processed in a worldwide distributed computing infrastructure, comprising 170 centers in 35 countries, known as WLCG (World-wide LHC Computing Grid). LHC operates in yearly periods, characterized by incremental steps in the number of particles that collide, gradually increasing the amount of experimental data to be stored and analyzed. By 2026, the experiment will face the High-Luminosity LHC (HL-LHC) era, where the produced data will increase a factor 10 as compared to today's values. The compute budget is not expected to increase substantially, so the LHC community is exploring novel ideas to integrate into experiment compute models in order to alleviate the expected compute and storage demands in that period. In terms of data management and access one of the strategic directions is to integrate storage caches as network delivery content solutions, and consolidating the main storage systems of WLCG into fewer sites. One of the benefits would be to be able to run the sites with less computing contribution without having its own storage system deployed. This technology would also have an impact over the latency hiding for remote reads and accelerate data delivery to opportunistic compute clusters and Cloud resources. In this contribution we simulate different behaviors and configurations of Least Recently Accessed (LRU) data caches for the CMS experiment in the Spanish CMS region, using real data accesses from both PIC Tier-1 and CIEMAT Tier-2. We expose and discuss the most efficient features and configurations to optimize caches and executed job performances in terms of the most relevant identified metrics.

Primary authors: Dr SIKORA, Anna (Universitat Autònoma de Barcelona (UAB)); PEREZ DENGRA, Carlos (CIEMAT); FLIX, Josep (PIC / CIEMAT)

Presenter: PEREZ DENGRA, Carlos (CIEMAT)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 65

Type: **Oral Presentation**

Integrating EGI Check-in in the Virtual Imaging Platform

Monday, March 21, 2022 1:30 PM (30 minutes)

The Virtual Imaging Platform (VIP) is a web portal (<https://vip.creatis.insa-lyon.fr>) allowing researchers from all over the world to have access to medical imaging applications as a service. They can thus execute scientific applications on distributed compute and storage resources (provided by the biomed EG VO) simply through their web browser. The platform currently counts more than 1300 registered users and approximately 20 applications available.

VIP is involved in the EGI-ACE project and registered in the EOSC market place (<https://marketplace.eosc-portal.eu/services/virtual-imaging-platform>). In this context, we are currently working on integrating The EGI Check-in authentication solution (<https://www.egi.eu/services/check-in/>). EGI Check-in is a proxy service that operates as a central hub to connect federated Identity Providers (IdPs) with EGI service providers. Check-in allows users to select their preferred IdP so that they can access and use EGI services in a uniform and easy way.

The presentation will give a general VIP overview, with a focus on the EGI Check-in integration and the two use-cases that are targeted: (i) facilitate users' authentication and (ii) access to external storage systems.

Primary authors: Mr CORNIER, Alexandre (Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294); Mr BONNET, Axel (Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294); POP, Sorina (CNRS - Creatis)

Presenter: Mr CORNIER, Alexandre (Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294)

Session Classification: Health & Life Science Applications

Track Classification: Track 2: Health & Life Sciences (including Pandemic Preparedness Applications)

Contribution ID: 66

Type: **Oral Presentation**

Exploiting cloud resources with DIRAC

Wednesday, March 23, 2022 12:00 PM (20 minutes)

The number of scientific communities needing access to large amounts of computing resources for their research work is growing. These demands are largely satisfied by grid computing infrastructures providing a unified access to resources distributed all over the world. A growing portion of those resources are provided as private clouds. These sites have different access protocols compared to traditional grid computing elements and can include more specialized capacities, for example, virtual machines with GPU accelerators, increased memory, etc. Therefore, there is a need to provide users with a uniform interface to access both grid and cloud resources. DIRAC is a project developing tools and services for the users of distributed computing resources. The DIRAC Workload Management Service allows to handle user jobs that can run on traditional computing clusters but also on cloud sites transparently. The service is provided by several large grid infrastructure projects, for example, EGI, GridPP, and others. In this contribution we will describe the DIRAC subsystem for managing cloud resources, mechanisms for creation and managing the life cycle of the virtual machines, secure access to clouds based on the OAuth2/OIDC tokens technology. We will give examples of cloud usage by research communities of users of GridPP and EGI grids.

Primary authors: TSAREGORODTSEV, Andrei (CPPM-IN2P3-CNRS); Mr LYTOVCHENKO, Andreii (Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France); Mrs BAUER, Daniela (Imperial College); Mr FAYER, Simon (Imperial College, UK)

Presenter: TSAREGORODTSEV, Andrei (CPPM-IN2P3-CNRS)

Session Classification: Infrastructure Clouds and Virtualisation

Track Classification: Track 8: Infrastructure Clouds and Virtualizations

Contribution ID: 67

Type: **Oral Presentation**

Machine Learning inference using PYNQ environment in an AWS EC2 F1 Instance

Thursday, March 24, 2022 12:00 PM (20 minutes)

In the past few years, using Machine and Deep Learning techniques has become more and more viable, thanks to the availability of tools which allow people without specific knowledge in the realm of data science and complex networks to build AIs for a variety of research fields. This process has encouraged the adoption of such techniques, e.g. in the context of High Energy Physics, new algorithms based on ML are being tested for event selection in trigger operations, end-user physics analysis, computing metadata based optimizations, and more. Time critical applications can benefit from implementing algorithms on low-latency hardware like specifically designed ASICs and programmable micro-electronics devices known as FPGAs. The latter offers a unique blend of the benefits of both hardware and software. Indeed, they implement circuits just like hardware, providing power, area and performance benefits over software, yet they can be reprogrammed cheaply and easily to implement a wide range of tasks, at the expense of performance with respect to ASICs.

In order to facilitate the translation of ML models to fit in the usual workflow for programming FPGAs, a variety of tools have been developed. One example is the HLS4ML toolkit, developed by the HEP community, which allows the translation of Neural Networks built using tools like TensorFlow to a High-Level Synthesis description (e.g. C++) in order to implement this kind of ML algorithms on FPGAs.

This paper presents and discusses the activity started at the Physics and Astronomy department of University of Bologna and INFN-Bologna devoted to preliminary studies for the trigger systems of the CMS experiment at the CERN LHC accelerator. A broader-purpose open-source project from Xilinx (a major FPGA producer) called PYNQ is being tested combined with the HLS4ML toolkit. The PYNQ purpose is to grant designers the possibility to exploit the benefits of programmable logic and microprocessors using the Python language. This software environment can be deployed on a variety of Xilinx platforms, from IOT devices like the ZYNQ-Z1 board, to the high performance ones, like Alveo accelerator cards and on the cloud EC2 F1 instances. The use of cloud computing in this work allows us to test the capabilities of this workflow, from the creation and training of a Neural Network and the creation of a HLS project using HLS4ML, to managing NN inference with custom Python drivers.

Hardware and software set-up, together with performance tests on various baseline models used as benchmarks, will be presented. The presence or not of some overhead causing an increase in latency will be investigated. Eventually, the consistency in the predictions of the NN, with respect to a more traditional way of interacting with the FPGA via the Vivado Software Development Kit using C++ code, will be verified.

As a next step for this study, Alveo accelerator cards are expected to be tested with the presented workflow as well, and a local server devoted to test NN in a fast, reliable and easy-to-use way will likely be assembled.

Primary author: Dr LORUSSO, Marco (Alma Mater Studiorum - University of Bologna)

Co-authors: Dr TRAVAGLINI, Riccardo (INFN - Bologna Division); Prof. BONACORSI, Daniele (University of Bologna)

Presenter: Dr LORUSSO, Marco (Alma Mater Studiorum - University of Bologna)

Session Classification: Physics & Engineering

Track Classification: Track 1: Physics (including HEP) and Engineering Applications

Contribution ID: 68

Type: **Oral Presentation**

Exploiting big data analytics for CMS computing operations cost-effectiveness

Thursday, March 24, 2022 4:10 PM (20 minutes)

Computing operations at the Large Hadron Collider (LHC) at CERN rely on the Worldwide LHC Computing Grid (WLCG) infrastructure, designed to efficiently allow storage, access, and processing of data at the pre-exascale level.

A close and detailed study of the exploited computing systems for the LHC physics mission represents an increasingly crucial aspect in the roadmap of High Energy Physics (HEP) towards the exascale regime. A deep knowledge of these systems will be essential to effectively express the discovery potential of LHC in the High Luminosity phase, which is estimated to collect up to 30 times more data than LHC in the next few years, and with novel part of the detectors which also adds large complexities to the overall picture.

In this context, the Compact Muon Solenoid (CMS) experiment has been collecting and storing over the last years a large set of heterogeneous non-collision-data (e.g. meta-data about data placement choices, transfer operations, and actual user access to physics datasets): all this data richness is currently residing on a distributed Hadoop cluster, and the data is organized so that running fast and arbitrary queries using the Spark analytics framework is a viable approach for focussed big data mining efforts.

CMS relies on its Monte Carlo Management (MCM) system, a tool to collect and monitor all Monte Carlo sample requests, namely a system that can be used to gain access to additional information about the simulated datasets produced for physics analysis.

Exploiting these sources, and using a data-driven approach oriented to the analysis of the aforementioned meta-data, we started to focus on data storage and data transfers over the WLCG infrastructure, and drafted an embrional software toolkit that can provide useful indicators about recurrent patterns and their correlations, for a deeper and more accurate exploration of the CMS computing beating heart in terms data movement and access. This aims – as a short-to-medium term goal – at exploring the effectiveness and adequateness of various choices in a data lake context, and – as a long term goal – at contributing to the overall design of a predictive/adaptive system that would eventually reduce cost and complexity of the CMS computing operations, while taking into account the stringent requests and the boundary conditions set by the physics analysts community.

Primary authors: BONACORSI, Daniele (University of Bologna); Dr GASPERINI, Simone (University of Bologna); ROSSI TISBENI, Simone (INFN-CNAF)

Presenters: Dr GASPERINI, Simone (University of Bologna); ROSSI TISBENI, Simone (INFN-CNAF)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data

Contribution ID: 69

Type: Oral Presentation

Reinforce-lib: A Reinforcement Learning Library for Scientific Research

Wednesday, March 23, 2022 4:30 PM (30 minutes)

Since the breakthrough achieved by the *DQN agent* in 2013 and 2015 on the *Atari learning environment* - a benchmark that was thought to be feasible only for humans - *Reinforcement Learning* (RL) and, especially, its combination with *Deep Learning* (DL), called *Deep Reinforcement Learning* (DRL), have both gained a major interest in the field since then, likewise *AlexNet* (thanks to the astounding improvement achieved on the ILSVRC challenge, compared to the best classical computer vision algorithm at that time) had started the deep learning era.

After few years, we have now powerful actor-critic distributed agents that are able to solve complex control and robotic tasks, and surprising even able to handle exponentially large search spaces like the ones found in the board games of Chess and Go, as well as high-dimensional continuous state-spaces of multi-agent environments. All these successes have common roots: powerful neural-networks function approximators borrowed from DL, and distributed training.

Although DRL seems to be incredibly powerful, in practice training successful agents is notoriously difficult, time consuming, resource intensive, costly, and error-prone: mainly due to very sensitive hyperparameters, beyond problem complexity itself. Such difficulties may arise from a still very limited understanding of the underlying mechanisms that power both RL and DRL, effectively preventing us to derive simpler (i.e. with way less moving parts, and hyperparameters) and thus more effective, sample-efficient, RL algorithms.

As happened in DL, having widely-used tools like *Keras* that simplify the building and training of neural networks is essential for speeding-up and improving research in that particular and related field(s). With such in mind, our aim is to provide a tool to ease the workflow of *defining, building, training, evaluating* and *debugging* DRL agents.

Our Python library **reinforce-lib** will provide simple code interfaces to a variety of implement agents and environments. We adopt a *modular design* that allows users to replace components like the agent's *networks*, its *policy*, and even *memory buffer* with other components made available by the library or new modules designed by the users themselves; this should enable the practitioner or researcher to easily prototype either novel research ideas, improvements to existing algorithms and components, or to adapt the agent for new, previously unsolved, research problems.

The *reinforce-lib* library other than being designed to be easy to use and extend, it will be also complete in the long-term (and, indeed, open-source), encompassing the three main paradigms found in reinforcement learning, namely: *model-free RL*, *model-based RL*, and *inverse RL*. This will allow users to solve a broader variety of problems according to the prior problem setting. For example, if the problem we want to solve naturally allow us to define a reward function, we will choose a model-free agent to solve it. If instead is easy for the researcher to provide a model of the environment (task or problem), model-based agents will do the job. Lastly, if we have many data coming from optimal sources (like domain experts, precise but expensive numerical simulations, exact algorithms, etc) we can leverage inverse RL algorithms to learn a reward function, and then use the learned reward to power the learning of model-free or hybrid agents.

The design principles of *reinforce-lib*, namely *usability*, *extensibility*, and *completeness*, should make the library distinguish itself from currently available RL libraries which are often based on *old* or legacy code, resulting in *unfriendly* code interfaces that are *hard to use* and even *harder to extend* or adapt. Mayor RL libraries like OpenAI's baselines, stable-baselines, Google's dopamine, and TensorFlow (to name a few) are also *very narrow*, often providing only few of the agents developed even by their own researchers.

We believe reinforcement learning to have many applications in scientific fields, thus further improving over classical or deep learning baselines, being even able to provide an answer to previously infeasible problems like the amazing breakthrough *AlphaFold* accomplished about *protein folding*. Today RL is mostly used in *games*, and *control systems* (often in simulation). Having the right tools, like the library we aim to develop with success, could help reinforcement learning to find applications in many more scientific scenarios.

Primary author: Mr ANZALONE, Luca (University of Bologna)

Co-author: Prof. BONACORSI, Daniele (University of Bologna)

Presenter: Mr ANZALONE, Luca (University of Bologna)

Session Classification: Artificial Intelligence

Track Classification: Track 10: Artificial Intelligence (AI)

Contribution ID: 70

Type: **Oral Presentation**

Proposing Authentic Assessment Strategies for New Education Normal Era Fortified with Academic Integrity Mindset and Culture

Tuesday, March 22, 2022 1:30 PM (30 minutes)

It is proposed in this presentation that a new authentic assessment approach to e-Portfolio may be realized in terms of such qualitative assessment method as Learning Analytics enhanced with text-mining technique in machine learning. Since the assessment in the educational paradigm of active learning cannot rely heavily on the way of the summative evaluation in the quantitative way, the key factor resides in the formative as well as qualitative assessment strategies along the learning processes evidenced with artifacts. This approach considers such key factors and proposes to make all stakeholders in education including learners as the major role player in the process of assessment. This presentation starts with the educational framework or philosophy in which the learning e-Portfolio is positioned in the curriculum. And then, the curriculum to support such educational paradigm is elaborated. With the supporting environment for the educational paradigm being elaborated, the theoretical background for the qualitative assessment is stated, making use of the text-mining strategies in terms of the Bloom's Taxonomy Matrix to visualize the state of the learning mind.

Primary authors: CHEN, Ru-Shan (Chihlee University of Technology); YAMAMOTO, Tosh (Kansai University); ZHANG , Zhi-hua (Kansai University of International Studies)

Presenter: YAMAMOTO, Tosh (Kansai University)

Session Classification: Humanities, Arts & Social Sciences Applications

Track Classification: Track 4: Humanities, Arts, and Social Sciences (HASS) Applications

Contribution ID: 71

Type: **Oral Presentation**

Remote Sensing Satellite Development and Earth Observation Applications from Taiwan Space Agency, NSPO

Tuesday, March 22, 2022 11:30 AM (30 minutes)

National Space Organization (NSPO) which is the Taiwan Space Agency. There are fifteen satellites have been developed and launched successfully since 1991. In 1999, FORMOSAT-1 was deployed with three scientific experimental missions: (1) Ionosphere Plasma and Electrodynamics Instrument for measuring the effects of ionospheric plasma and electrodynamics, (2) Ocean Color Imager for taking the images of visible and near-infrared light radiometric measurement on the ocean surface, and (3) Experimental Communication Payload for transmitting and receiving Ka-band signal from the ground stations. In 2004, FORMOSAT-2 was deployed as Taiwan's first optical Remote Sensing Instrument (RSI) satellite. Real-time images were taken daily with 2-m black & white and 8-m color images. The earth observation (EO) satellite images were applied on territorial planning, natural resource exploration, environmental protection, disaster prevention and relief, and other earth observation applications. In 2006, FORMOSAT-3 constellation (with six satellites) has deployed for observing global atmosphere and ionosphere. The global radio occultation (RO) data was applied on the weather forecast updates, long-term climate change research, dynamic monitoring of the ionosphere, earth gravity research and other related scientific research. These satellite programs were accomplished their missions and decommissioned.

In 2017, FORMOSAT-5 was deployed. It's the first self-developed optical RSI satellite by Taiwan, according to the experience of FORMOSAT-2. The optical RSI with 2-m black & white and 4-m color images continues to serve the global imagery users' community of FORMOSAT-2. In 2019, FORMOSAT-7 was deployed. It's a follow-on program to the successful FORMOSAT-3 with upgraded performance for spacecraft bus and mission payload. FORMOSAT-7 with six-satellites were forming a constellation to increase meteorological data collection in low-latitude regions (between 50 degrees north and south latitudes). An average of 4,000 points of RO data were received every day, and the accuracy of weather forecast, climate observation, and space weather monitoring are expected to be improved. Currently, FORMOSAT-5 and FORMOSAT-7 are on the orbit and serving the needs for Taiwan, as well as for the global society.

Nowadays, a couple of satellite programs are in the progress. Triton, also called as Wind-Hunter Satellite, is the self-derived meteorological satellite which carries Global Navigation Satellite System Reflectometry (GNSS-R) to collect GNSS signals reflected from the Earth surface in low Earth orbit and conduct researches on soil characteristics, air-sea interaction, and typhoon intensity prediction. FORMOSAT-8 is the self-derived optical RSI constellation which is planned to develop six satellites with 1-m black & white and 2-m color images continues the mission of FORMOSAT-5. Beyond 5G is planned to develop low earth orbit communication satellite to experiment the broadband internet service, video conference and internet of thing.

Presenters: Dr CHEN, Bo (NSPO); Dr HWANG, Feng-Tai (NSPO); Dr CHOU, Jill (NSPO)

Session Classification: Earth & Environmental Sciences & Biodiversity Applications

Contribution ID: 72

Type: **Oral Presentation**

Implementing the European Open Science Cloud

Wednesday, March 23, 2022 9:30 AM (1 hour)

The European Open Science Cloud (EOSC) aims to offer European researchers a virtual environment for open access to services to reuse scientific data. EOSC initiative started to be shaped by the end of 2015 and get a stable structure with creating the EOSC AISBL in 2020. EOSC addresses technical and organisational challenges, such as the federation of the research infrastructures, the fulfilment and development of the FAIR principles on data, the set up of services for data discovery, data access and processing integrated on a centralised access portal and the definition of a governance model with clear Rules of Participation. The implementation of EOSC is conducted through a tripartite model with a prominent role by the EOSC Association (EOSC-A), which gathers more than 230 members contributing to the implementation of these different dimensions. The operating bodies of the EOSC-A, such as the Task Forces, contribute to coordinating the activities of institutions and projects funded at the European and national levels on the development of the Strategic Research and Innovation Agenda (SRIA). The implementation of the European dimension of the EOSC strongly depends on the European Projects funded in the frame of the INFRA-EOSC calls, both in the H2020 and the HEU Framework Program. An overview of some of the main activities in the landscape will be given and discussed during the presentation.

Presenter: Prof. BLANQUER, Ignacio (Universitat Politècnica de València)

Session Classification: Keynote Speech

Contribution ID: 73

Type: **Oral Presentation**

Science, Computing and AI

Friday, March 25, 2022 9:30 AM (1 hour)

Presenter: Dr SEE, Simon (NVIDIA)

Session Classification: Keynote Speech

Contribution ID: 74

Type: **Oral Presentation**

e-Science Activities in Singapore

Thursday, March 24, 2022 9:10 AM (10 minutes)

Presenter: Dr TAN, Tinwee (NSCC)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 75

Type: **Oral Presentation**

e-Science Activities in Thailand

Tuesday, March 22, 2022 9:20 AM (10 minutes)

Presenter: Dr PIYOUNGKORN, Kajornsak (NECTEC)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 76

Type: **Oral Presentation**

e-Science Activities in Japan

Thursday, March 24, 2022 9:40 AM (10 minutes)

Presenter: AIDA, Kento (National Institute of Informatics)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 77

Type: **Oral Presentation**

e-Science Activities in China

Thursday, March 24, 2022 9:00 AM (10 minutes)

Presenter: CHEN, Gang (Institute Of High Energy Physics)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 78

Type: **Oral Presentation**

e-Science Activities in Taiwan

Tuesday, March 22, 2022 9:00 AM (10 minutes)

Presenter: LEE, Felix (ASGC)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 79

Type: **Oral Presentation**

EduceLab: Midscale Infrastructure for Heritage Science

Monday, March 21, 2022 3:30 PM (15 minutes)

The commissioning of the EduceLab infrastructure, funded through the National Science Foundation's "Mid-Scale Infrastructure" program, represents a tremendous opportunity in operational capacity for Heritage Science problems in the central region of the United States. This overview explains the infrastructure and its intended use and capabilities.

Presenter: Prof. CHAPMAN, Christy (University of Kentucky)

Session Classification: Humanities, Arts & Social Sciences Applications

Contribution ID: 80

Type: **Oral Presentation**

FLEX: A Flexible X-ray and Micro-CT Environment for Heritage Science

Monday, March 21, 2022 3:45 PM (15 minutes)

The use of x-ray analysis and imaging for Heritage Science problems has traditionally been limited to systems engineered for other uses. This talk explains the design and operational goals of the FLEX instrument cluster in EduceLab, which targets custom and flexible configurations for imaging using x-ray in order to accommodate the analysis of a wide variety of Heritage Science problems.

Presenter: Prof. PARKER, Seth (University of Kentucky)

Session Classification: Humanities, Arts & Social Sciences Applications

Contribution ID: **81**

Type: **Oral Presentation**

Ancient Lives, Modern Tech: Tools and Process for Digital Scholarship

Monday, March 21, 2022 4:00 PM (15 minutes)

The software tools and associated metadata that accompany the analysis of heritage material form a critical pathway for managing the analysis of data generated for projects that now depend on machine learning and other complex algorithmic approaches. This talk discusses the framework for doing scholarly work in a data-rich environment while maintaining high standards for visualization, peer review, and digital provenance.

Presenter: Prof. BRUSUELAS, James (University of Kentucky)

Session Classification: Humanities, Arts & Social Sciences Applications

Contribution ID: 82

Type: **Oral Presentation**

CyberInfrastructure for Heritage Science

Monday, March 21, 2022 4:15 PM (15 minutes)

This talk explains the computational environment to support the goals of the EduceLab Heritage Science ecosystem, including massive storage, internal and external access to data, mobile systems, instruments producing large streams of data, and computational cycles to support activities like machine learning algorithms running over massive datasets.

Presenter: Prof. PIKE, Lowell (University of Kentucky)

Session Classification: Humanities, Arts & Social Sciences Applications

Contribution ID: 83

Type: **Oral Presentation**

Segmentation and the M.910 From Tomography

Monday, March 21, 2022 4:30 PM (15 minutes)

This talk focuses on an early prototype project with the Morgan M.910 manuscript representing the kind of projects that Educelab will facilitate, showing the advances that await by applying x-ray imaging, micro-CT, and machine learning techniques to the analysis of ancient materials.

Presenter: Dr GESSEL, Kristina (University of Kentucky)

Session Classification: Humanities, Arts & Social Sciences Applications

Contribution ID: 84

Type: **not specified**

Machine Learning and CyberInfrastructure for inkID and trans-modal rendering

Monday, March 21, 2022 4:45 PM (15 minutes)

How can we turn back the clock on damaged, fragile manuscripts? X-ray CT allows us to see the detailed internal structure of scrolls and books, but alone does not reverse any of the damage that occurred. This talk discusses how we use machine learning to enhance the results of X-ray CT, restoring artifacts to their full original splendor by revealing otherwise invisible inks, virtually reversing physical damage, and generating full color images.

Presenter: Dr PARSONS, Stephen (University of Kentucky)

Session Classification: Humanities, Arts & Social Sciences Applications

Contribution ID: 85

Type: **Oral Presentation**

e-Science Activities in Korea

Tuesday, March 22, 2022 9:10 AM (10 minutes)

Presenter: CHO, Kihyeon (KISTI/UST)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 86

Type: **Oral Presentation**

"e-Science Activities in ARDC/ Australia

Tuesday, March 22, 2022 9:30 AM (10 minutes)

Presenter: Dr CODDINGTON , Paul (ARDC)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 87

Type: **Oral Presentation**

e-Science Activities in Indonesia

Thursday, March 24, 2022 9:20 AM (10 minutes)

Presenter: Dr SUHARDIMAN , Basuki (ITB)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 88

Type: **Oral Presentation**

e-Science Activities in Philippines

Thursday, March 24, 2022 9:30 AM (10 minutes)

Presenter: Dr TETANGCO, Jelina Tanya H. (DOST)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 89

Type: **Oral Presentation**

Integrating Quantum and High Performance Computing - Expectations and Challenges

Monday, March 21, 2022 9:40 AM (40 minutes)

Recent advances in Quantum Computing (QC) have generated high expectations from scientific and industrial communities. While first prototypes are already moving out of the physics labs into the data centers, it is still unclear how the expected benefits will support research and development. The Quantum Integration Center (QIC) at LRZ addresses these challenges with an approach, where QC is integrated with High Performance Computing (HPC). The idea is that the HPC delegates specific tasks, which are better suited to the QC, to a quantum accelerator inside the HPC system. As a consequence, the overall achieved performance of scientific simulations can be improved. To realize this integration, a number of gaps need to be addressed, mostly on the software engineering and computer science aspects of integration. In this talk, we will provide an overview of the current developments in the QIC@LRZ and the larger Munich Quantum Valley (MQV) initiative.

Presenter: KRANZLMULLER, Dieter (LMU Munich)

Session Classification: Opening Ceremony & Keynote Speech I

Contribution ID: **90**

Type: **Oral Presentation**

Opening Remarks

Monday, March 21, 2022 9:30 AM (10 minutes)

Presenters: MATYSKA, Ludek (CESNET); CHANG, Yuan-Hann (Institute of Physics, Academia Sinica)

Session Classification: Opening Ceremony & Keynote Speech I

Contribution ID: 91

Type: **not specified**

e-Science Activities in MAS/ Mongolia

Tuesday, March 22, 2022 9:40 AM (10 minutes)

Presenter: BADRAKH, Otgonsuvd (Researcher at Institute of Mathematics and Digital technology, Mongolian Academy of Sciences)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: **92**

Type: **not specified**

Panel Discussison

Tuesday, March 22, 2022 9:50 AM (40 minutes)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 93

Type: **not specified**

Panel Discussion

Thursday, March 24, 2022 9:50 AM (40 minutes)

Session Classification: e-Science Activities in Asia Pacific

Contribution ID: 94

Type: **Oral Presentation**

Hydro-Meteorological modelling using the H2020 LEXIS platform

Tuesday, March 22, 2022 12:00 PM (30 minutes)

Presenter: Dr PARODI, Antonio (CIMA)

Session Classification: Earth & Environmental Sciences & Biodiversity Applications

Contribution ID: 95

Type: **not specified**

Q&A

Thursday, March 24, 2022 4:30 PM (20 minutes)

Session Classification: Data Management & Big Data

Contribution ID: 96

Type: **Oral Presentation**

APOCAWLIPSEA - Prediction of Air Pollution from Wildfires in S-E Asia & the Open Call for LEXIS Cloud/HPC Platform Evaluation

Wednesday, March 23, 2022 3:30 PM (30 minutes)

Presenters: LUANGSODSAI, Arthorn; DITTRICH, Jirathana; NIINIMAKI, Marko; HAYEK, Mo-
hamad (Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities); HACHINGER,
Stephan (Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities); TAN-
PIPAT, Veerachai (Hydro-Informatics Institute and Faculty of Forestry, Kasetsart University)

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: 97

Type: **Oral Presentation**

Real time computation of tsunami inundation estimates with TsunAWI as part of the LEXIS project

Wednesday, March 23, 2022 4:00 PM (30 minutes)

Presenter: RAKOWSKI, Natalja

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: 98

Type: **Oral Presentation**

K2I - Tracing Unknown Water Pollutants with Artificial Intelligence

Wednesday, March 23, 2022 4:30 PM (30 minutes)

Presenter: PAUW, Viktoria

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: 99

Type: **Oral Presentation**

Introduction & Regional Collaborations

Wednesday, March 23, 2022 11:00 AM (10 minutes)

Presenter: YEN, Eric (ASGC)

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: **100**

Type: **Oral Presentation**

2022 Hunga Tonga eruption and tsunami event

Wednesday, March 23, 2022 11:40 AM (50 minutes)

Presenter: Prof. WU, Tso-Ren (National Central University)

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: **101**

Type: **Oral Presentation**

Floods

Wednesday, March 23, 2022 11:10 AM (30 minutes)

Presenter: LIEW, Ju Neng (Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia)

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: **102**

Type: **Oral Presentation**

Forest Fire/Haze/Maze and hydroinformatics

Wednesday, March 23, 2022 1:30 PM (20 minutes)

Presenter: TANPIPAT, Veerachai (Hydro-Informatics Institute and Faculty of Forestry, Kasetsart University)

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: 103

Type: **Oral Presentation**

Development and Knowledge transfer of early forest fire IoT detection system using AI

Wednesday, March 23, 2022 1:50 PM (20 minutes)

Presenter: HONGWITAYAKORN, Apisake

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: **104**

Type: **not specified**

Visual IoT for Disaster Mitigation

Wednesday, March 23, 2022 2:10 PM (20 minutes)

Presenter: MURATA, Ken T.

Session Classification: DMCC & Environmental Computing Workshop

Contribution ID: **105**

Type: **Oral Presentation**

Introduction

Thursday, March 24, 2022 11:00 AM (10 minutes)

Presenter: YEN, Eric (ASGC)

Session Classification: APGridPMA Meeting

Contribution ID: **106**

Type: **Oral Presentation**

EUGridPMA, IGTF, AARC and EOSC ISM

Thursday, March 24, 2022 11:10 AM (50 minutes)

Presenter: GROEP, David (Nikhef)

Session Classification: APGridPMA Meeting

Contribution ID: **107**

Type: **Oral Presentation**

Development of HPCI AAI in JP

Thursday, March 24, 2022 12:00 PM (30 minutes)

Presenter: SAKANE, Eisaku (National Institute of Informatics)

Session Classification: APGridPMA Meeting

Contribution ID: **108**

Type: **Oral Presentation**

SIFULAN and iFIRExMAN Project

Thursday, March 24, 2022 1:30 PM (30 minutes)

Presenter: SJAUGI, Muhammad Farhan (SIFULAN)

Session Classification: APGridPMA Meeting

Contribution ID: **109**

Type: **not specified**

Reports from CAs of APGridPMA

Thursday, March 24, 2022 2:00 PM (30 minutes)

Session Classification: APGridPMA Meeting

Contribution ID: **110**

Type: **not specified**

TAGPMA update

Thursday, March 24, 2022 2:30 PM (30 minutes)

Presenter: SIMMEL, Derek (Pittsburgh Supercomputing Center)

Session Classification: APGridPMA Meeting

Contribution ID: **111**

Type: **not specified**

Next Meeting and AOB

Thursday, March 24, 2022 3:00 PM (5 minutes)

Session Classification: APGridPMA Meeting

Contribution ID: 112

Type: **Oral Presentation**

A collaboration and compute platform from EGI for Open science in Asia

Wednesday, March 23, 2022 2:30 PM (30 minutes)

EGI-ACE is a 30-month European project (Jan 2021 - June 2023) with a mission to accelerate Open Science practices, by delivering online services for compute and data intensive research. EGI-ACE provides the Compute Platform of the European Open Science Cloud (EOSC): a federation of compute and storage facilities, complemented by diverse access, data management and analytical services. The EOSC Compute Platform is designed for a wide range of scientific data processing and analysis use cases, including the hosting of scientific portals and data spaces. The Compute Platform heavily builds on the 22 OpenStack providers of the EGI Federated Cloud. EGI-ACE recently allocated a resource pool from these cloud providers to environmental and agriculture researchers in Asia Pacific. The allocated cloud resources and the additional value-added services can be used for the hosting of scientific services and data, application delivery, interactive data analysis and visualisation, sharing of compute and storage cycles. The talk will present the opportunities that this new EGI-ACE infrastructure brings to open science in Asia.

Presenter: SIPOS, Gergely (EGI)

Session Classification: DMCC & Environmental Computing Workshop