

International Symposium on Grids & Clouds 2020 (ISGC 2020)

Sunday, 23 August 2020 - Friday, 28 August 2020

BHSS, Academia Sinica

Scientific Programme

Theme: Challenges in High Performance Data Analytics: Combining Approaches in HPC, HTC, Big Data and AI

While the research data are becoming a real asset nowadays, it is an information and knowledge gained through thorough analysis that makes them so valuable. To process vast amounts of data collected, novel high performance data analytics methods and tools are needed, combining classical simulation oriented approaches, big data processing and advanced AI methods. Such a combination is not straightforward and needs novel insights at all levels of the computing environment – from the network and hardware fabrics through the operating systems and middleware to the platforms and software, not forgetting the security – to support data oriented research. Challenging use cases that apply difficult scientific problems are necessary to properly drive the evolution and also to validate such high performance data analytics environments.

The goal of ISGC 2020 is to create a face-to-face venue where individual communities and national representatives can present and share their contributions to the global puzzle and contribute thus to the solution of global challenges. We cordially invite and welcome your participation!

Physics (including HEP) and Engineering Applications

Submissions should report on experience with physics and engineering applications that exploit grid and cloud computing services, applications that are planned or under development, or application tools and methodologies. Topics of interest include: (1) End-user data analysis including ML/DL based one; (2) Management of distributed data; (3) Applications level monitoring; (4) Performance analysis and system tuning; (5) Workload scheduling; (6) Management of an experimental collaboration as a virtual organization; (7) Comparison between grid and other distributed computing paradigms as enablers of physics data handling and analysis; (8) Expectations for the evolution of computing models drawn from recent experience handling extremely large and geographically diverse datasets; and (9) Application software development, optimization and benchmarking.

Biomedicine & Life Sciences Applications

During the last decade, research in Biomedicine and Life Sciences has dramatically changed thanks to the continuous developments in High Performance Computing and highly Distributed Computing Infrastructures such as grids and clouds, but also in big-data solutions to deal with the explosion in genomic data. This track aims at discussing problems, solutions and application examples related to this area of research, with a particular focus on non-technical end users. Submissions should concentrate on practical applications and solutions in the fields of Biomedicine and Life Sciences, such as Drug discovery, Structural biology, Bioinformatics, Medical imaging, Public health applications / infrastructures, High throughput (grid and cloud-based) data processing/analysis, Distributed data computing and services, and Big data management issues. Submissions should ideally highlight how the availability and use of Big Data has enabled new processes for or dramatically evolved the scope of their research.

Earth/Environmental Sciences & Biodiversity Applications

Natural and Environmental sciences are placing an increasing emphasis on the understanding of the Earth as a single, highly complex, coupled system with living and dead organisms. It is well accepted, for example, that the feedbacks involving oceanic and atmospheric processes can have major consequences for the long-term development of the climate system, which in turn affects biodiversity, natural hazards and can control the development of the cryosphere and lithosphere. Natural disaster mitigation is one of the most critical regional issues in Asia. Despite the diversity of environmental sciences, many projects share the same significant challenges. These include the collection of data from multiple distributed sensors (potentially in very remote locations), the management of large low-level data sets, the requirement for metadata fully specifying how, when and where the data were collected, and the post-processing of those low-level data into higher-level data products which need to be presented to scientific users in a concise and intuitive form. This session would in particular address how these challenges are being handled with the aids of e-Science paradigm.

Virtual Research Environment (including Middleware, tools, services, workflow, ... etc.)

Virtual Research Environments (VRE) provide an intuitive, easy-to-use and secure access to (federated) computing resources for solving scientific problems, trying to hide the complexity of the underlying infrastructure, the heterogeneity of the resources, and the interconnecting middleware. Behind the scenes, VREs comprise tools, middleware and portal technologies, workflow automation as well as security solutions for layered and multifaceted applications. Topics of interest include but are not limited to: (1) Real-world experiences building and/or using VREs to gain new scientific knowledge; (2) Middleware technologies, tools, services beyond the state-of-the-art for VREs; (3) Science gateways as specific VRE environments, (4) Innovative technologies to enable VREs on arbitrary devices, including Internet-of-Things; and (5) One-step-ahead workflow integration and automation in VREs.

Data Management & Big Data

The rapid growth of the data available to scientists and scholars – in terms of Velocity and Variety as well as sheer Volume – is transforming research across disciplines. Increasingly these data sets are generated not just through experiments, but as a byproduct of our day-to-day digital lives. This track explores the consequences of this growth, and encourages submissions relating to two aspects in particular - firstly, the conceptual models and analytical techniques required to process data at scale; secondly, approaches and tools for managing and creating these digital assets throughout their lifecycle.

Humanities, Arts, and Social Sciences (HASS) Applications

Disciplines across the Humanities, Arts and Social Sciences (HASS) have critically engaged with technological innovations such as grid- and cloud computing, and, most recently, various data analytic technologies. The increasing availability of data, ranging from social media text data to consumer big data has led to an increasing interest in analysis methods such as natural language processing, social network analysis, machine learning and text mining. These developments pose challenges as well as opening up opportunities and members of the HASS community have been at the forefront of discussions about the impact that novel forms of data, novel computational infrastructures and novel analytical methods have for the pursuit of science endeavours and our understanding of what science is and can be. The ISGC 2020 HASS track invites papers and

presentations covering applications demonstrating the opportunities of new technologies or critically engaging with their methodological implications in the Humanities, Arts and Social Sciences. Innovative application of analytical tools for survey data, social media data, and government (open) data are welcomed. We also invite contributions that critically reflect on the following subjects: (1) the impact that ubiquitous and mobile access to information and communication technologies have for society more generally, especially around topics such as smart cities, civic engagement, and digital journalism; (2) philosophical and methodological reflections on the development of the techniques and the approaches by which data scientists use to pursue knowledge.

Network, Security, Infrastructure & Operations

Networking and the connected e-Infrastructures are becoming ubiquitous. Ensuring the smooth operation and integrity of the services for research communities in a rapidly changing environment are key challenges. This track focuses on the current state of the art and recent advances in these areas: networking, infrastructure, operations, and security. The scope of this track includes advances in high-performance networking (software defined networks, community private networks, the IPv4 to IPv6 transition, cross-domain provisioning), the connected data and compute infrastructures (storage and compute systems architectures, improving service and site reliability, interoperability between infrastructures, data centre models), monitoring tools and metrics, service management (ITIL and SLAs), and infrastructure/systems operations and management. Also included here are issues related to the integrity, reliability, and security of services and data: developments in security middleware, operational security, security policy, federated identity management, and community management. Submissions should address solutions in at least one of these areas.

Infrastructure Clouds and Virtualisation

This track will focus on the development of cloud infrastructures and on the use of cloud computing and virtualization technologies in large-scale (distributed) computing environments in science and technology. We solicit papers describing underlying virtualization and "cloud" technology including integration of accelerators and support for specific needs of AI/ML and DNN, scientific applications and case studies related to using such technology in large scale infrastructure as well as solutions overcoming challenges and leveraging opportunities in this setting. Of particular interest are results exploring the usability of virtualization and infrastructure clouds from the perspective of machine learning and other scientific applications, the performance, reliability and fault-tolerance of solutions used, and data management issues. Papers dealing with the cost, price, and cloud markets, with security and privacy, as well as portability and standards, are also most welcome.

Converging High Performance infrastructures: Supercomputers, clouds, accelerators

The classical simulation-oriented computing is nowadays complemented by the novel general machine learning and specifically deep neural networks based approaches. This requires novel approaches to build high performance infrastructures, combining supercomputers, high performance clouds, specialized DNN hardware and other accelerators. An additional challenge lies in the individual components being provided by different owners, usually in a federated distributed way. This track solicits recent research and development achievements and best practices in building and exploiting these converging high performance infrastructures or their components. The topics of interest include, but are not limited to the followings: (1) Building and use of modern high performance computing systems, including special support for AI and DNN in particular; (2) Use of virtualization techniques and containers to support access to and portability

across different heterogeneous systems; (3) Experiences, use cases and best practices on the development and operation of large-scale heterogeneous applications; (4) Integration and interoperability to support coordinated federated use of different e-infrastructures (supercomputers, accelerated clouds, ...) and their building blocks; (5) Performance of different applications on these integrated high performance infrastructures.

Keynote Session

eScience in Asia Pacific Session