IL PROGETTO IBiSCo E LA TRANSIZIONE VERSO IL "*CENTRO NAZIONALE DI RICERCA IN HIGH PERFORMANCE COMPUTING, BIG DATA E QUANTUM COMPUTING (ICSC)*"

18-19 Aprile 2024
Complesso Universitario di Monte S. Angelo
Università di Napoli "Federico II"

BOOK OF ABSTRACTS

Editors: Gianpaolo Carlino, Giovanni Cantele, Luisa Carracciuolo, Alessandra Doria, Giorgio Maggi, Guido Russo,

UNIONE EUROPEA
Fondo Sociale Europeo
Fondo Europeo di Sviluppo Regionale

Ministero dell'Istruzione, dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE 2014 - 2020


18 aprile 2024

09:30–10:00 WELCOME COFFEE e REGISTRAZIONE
10:00–10:15 C1-0 Saluti istituzionali (proff. Gennaro Miele, prof. Luca Lista, prof. Fabio Miletto Granozio)

Chairperson: Giorgio Pietro Maggi
10:15–10:40 C1-1 *Carlino*, Il progetto IBiSCo
10:40–11:00 C1-2 *Doria et al.*, L'upgrade delle infrastrutture del DataCenter IBiSCo verso ICSC a INFN-NAPOLI
11:00–11:15 C1-3 *Vilucchi et al.*, Scientific Computing Facilities at LNF and INAF-OAR
11:15–11:35 C1-4 *Monforte et al.*, The IBiSCo infrastructure at INFN-CT
11:35–11:50 C1-5 *Russo et al.*, Le infrastrutture realizzate presso UNINA
11:50–12:05 C1-6 *di Bari et al.,* Le infrastrutture impiantistiche del Data Center IBiSCo-ReCaS-Bari dopo l'upgrade di IBiSCo
12:05–12:20 C1-7 *Donvito et al.*, Le infrastrutture di calcolo, storage e rete per il Data Center IBiSCo-ReCaS-Bari dopo l'upgrade di IBiSCo
12:20–12:35 C1-8 *Cantele et al.,* Le infrastrutture realizzate presso il CNR
12:35–13:00 C1-9 *Grandi*, Il Centro Nazionale ICSC
13:00–13:20 Questions & Answers
13:20–14:30 PAUSA PRANZO (buffet)

Chairperson: Alessandra Doria
14:30–14:45 C2-1 *Nicodemi et al.*, High-performance computing in biological physics
14:45–15:00 C2-2 *Pardi et al.*, IBiSCo usage for the Belle II experiment
15:00–15:15 C2-3 *Carracciuolo et al.*, Progettazione, Implementazione e Validazione di una risorsa HPC eterogenea
15:15–15:30 C2-4 *Vizzuso,* Quantum hybrid algorithms for combinatorial optimization problems
15:30–15:45 C2-5 *Spisso et al.*, The IBiSCo HPC seed for INFN experiments workflow-based cluster
15:45–16:00 Questions & Answers
16:00–16:30 COFFEE BREAK

Chairperson: Giovanni Cantele
16:30–16:45 C2-6 *Sabella et al.,* The User Experience in IBiSCo Data Center
16:45–16:55 C2-7 *Tortora et al.*, The Monitoring System in the IBiSCo Data Center
16:55–17:10 C2-8 *Delli Veneri et al.*, Cloud Infrastructure and storage for IBiSCo and ICSC at Napoli
17:10–17:20 C2-9 *Brescia et al.*, High-performance computing in astrophysics
17:20–17:30 C2-10 *Conte et al.*, Application of the IBiSCo Data Center for cultural heritage projects
17:30–17:50 Questions & Answers

19 aprile 2024

Chairperson: Silvio Pardi

09:20–09:30 C3-1 *Bonano et al.*, Methods and algorithms for the massive processing of synthetic aperture radar interferometric satellite data

09:30–09:40 C3-2 *Fontanarosa et al.*, Computational fluid dynamics of complex multiphase systems through electrohydrodynamic effect (EHD) for microfluidic applications

09:40–09:50 C3-3 *Vitale et al.,* Large scale molecular dynamics simulations of biomolecular systems in membrane environment

09:50–10:00 C3-4 *Massuoli et al.*, Development and application of parallel CPU/GPU tools to extend the characterization of direct and solvent-mediated H-bond interactions in large structural protein databases

10:00–10:10 C3-5 *di Bari et al.*, Contribution of the IBiSCo-ReCaS-Bari Data **C**enter to the Italian National Smart Specialization Strategy

10:10–10:20 C3-6 *Nicotri et al.*, The users of the IBiSCo-ReCaS-Bari data center and their support: organization and tools

10:20–10:30 C3-7 *Italiano et al.*, IBiSCo-ReCaS-Bari Cloud site evolution: Introducing the architecture and implementation of the upgraded IaaS infrastructure

10:30–10:40 C3-8 *Patano et al.*, Delivering integrated high-level services at Bari IBiSCo-ReCaS site through the PaaS dashboard

10:40–11:00 Questions & Answers

11:00–11:30 COFFEE BREAK

Chairperson: Melania Paturzo

11:30–11:40 C3-9 *Vino et al.*, GPU-based cluster at IBiSCo-ReCaS-Bari for containerized scientific applications

11:40–11:50 C3-10 *La Salandra et al.*, High-performance computing in UAV photogrammetry

11:50–12:00 C3-11 *Russomanno*, Many-body quantum dynamics simulations with the IBiSCo resource

12:00–12:10 C3-12 *Izzo et al.,* Unlocking AI Capabilities: Advanced Applications through High-Performance Computing

12:10–12:20 C3-13 *Piergentili et al.,* Open Access repositories for scientific literature and research data

12:20–12:40 Questions & Answers

12:40–13:40 PAUSA PRANZO (buffet)

Chairperson: Giacinto Donvito

13:40–13:55 C3-14 *Mutarelli et al.,* Setting up an automated framework for bioinformatic and multiomic analyses on IBiSCo

13:55–14:10 C3-15 *Lo Giudice et al.*, Development of a state of the art computational environment for handling human genetic data: the effort of ELIXIR-IT

14:10–14:25 C3-16 *Barreca et al.*, Material science problems on IBiSCo

14:25–14:40 C3-17 *Viola et al.*, Towards machine learning-based PDE identification for crowd dynamics

14:40–15:00 Conclusioni (*Carlino*)

Sommario

C1-1

Authors: G. Carlino

Title: The IBiSCo project

ABSTRACT:

IBISCO is a project involving the main Italian Research Institutions, INFN (leader), CNR, INAF and INGV and the Universities of Bari and Napoli funded by the Ministry of University and Research call in the Action II.1 of the National Operative Programme (PON) – Research and Innovation 2014-2020, aimed at enhancing the Research Infrastructures in the south of Italy.

The main goals of the project are to strengthen the multi-regional scientific computing infrastructure realized in southern Italy with previous PON calls (PI2S2, SCOPE and RECAS projects) and with the ministry Ordinary Operating Fund (DHTCS project) and to evolve it towards a multi-disciplinary digital platform supporting all the Italian scientific communities.

The enhanced IBISCO research infrastructure consists of a geographically distributed data center, subdivided between Bari, Catania and Naples data centers in southern Italy and the Laboratories of Frascati data center, that is visible as a single distributed system thanks to the high physical connection and the logic Data Lake among the sites.

This presentation will summarize the whole story of the IBISCO project, the main achievements reached and the connection with the High-Performance Computing, Big Data e Quantum Computing National Center (ICSC) established by the National Recovery and Resilience Plan (NRRP).

C1-2

Authors: A. Doria, G. Carlino, S. Pardi, G. Russo, B. Spisso, A. Tortora, V. Artiano, M. Delli Veneri, F. Di Nucci, S. Stellacci, V. Rega.

Title: The upgrade of the Data Center infrastructures towards ICSC at INFN-NAPOLI

ABSTRACT:

INFN-NAPOLI hosts a Data Center (DC) since 2006, initially as Tier2 site in Worldwide LHC Computing GRID, dedicated to particle physics experiments like ATLAS and Belle II. With IBiSCo project the Data Center infrastructure has grown in size and opened to other scientific communities.

The IT resources are placed in closed racks, internally cooled with a water-air system, up to 30KW cooling capacity per rack. The infrastructure was enlarged with eight new racks and the power system was upgraded to supply the upgraded DC.

Each new rack hosts servers and storage, for a total amount of about 5000 CPUs and 10PB of storage, that are managed as a single cluster together with the pre-existing resources. The DC network has also been upgraded, making the DC ready for the latest challenges of scientific computing.

C1-3

Authors: S. Gallozzi, E. Vilucchi

Title: Scientific Computing Facilities at INFN-LNF and INAF-OAR

ABSTRACT:

The Frascati National Laboratories (LNF) actively contribute to the Worldwide LHC Computing Grid (WLCG) through their dedicated scientific computing data center (DC), serving as one of Italy's four ATLAS Tier2 facilities. The room that hosts the DC has a surface area of approximately 100m2 and a power of 160kW available, set up with a double power system, redundant with UPS and generator. The electrical power supply comes from the LNF electrical substation and is therefore extremely reliable. Moreover, the DC incorporates an eco-friendly air-cooling system, leveraging winter thermal waste to heat adjacent LNF facilities covering over 12,000m².

The geographical proximity of this computing centre, which currently hosts around 5PB of storage disks and 5000 computational cores, with the INAF-OAR in Monteporzio Catone has fostered a collaboration between the institutes for the development of the Astri, Miniarray and Cherenkov Telescope Array, CTA.

This collaboration was strengthened by the IBISCO PON, thanks to which it was possible to update the DC network infrastructure, increasing its speed by a factor of ten and allowing connection to the global Tier2 network, LHCONE, at 100Gbps. Furthermore, the project funds facilitated the acquisition of six service machines (for hosting virtualization systems and prototypes of the INFN Data-Cloud infrastructure), twenty-four computing servers and 1PB of storage disk that represent the first installation of the Italian data center of the CTA Observatory (one of four).

The Italian CTA data center, result of collaboration between INFN and INAF, provides a set of observatory services and tools together with an high throughput access to all data products generated by the CTA observatory runtime. Such services include distributed storage system, a computing orchestration, a dedicated workload management system and other higher level astronomers users services well described in the CTA Computing Model.

Finally general-purpose services are placed in the Observatory of Rome within the institution datacenter, which is 10Gbps connected point-to-point to the INFN-LNF and consists of 12 machines acquired with IBISCO founds to host general services for the shared ASTRI, Miniarray and CTA datacenter.

C1-4

Authors: S. Monforte, O. Conti, C. Rocca, R. Rotondo, P. Belluomo, G. Platania

Title: The IBiSCo infrastructure at INFN-CT

ABSTRACT:

This abstract succinctly encapsulates the strategic augmentation of IT infrastructure within the IBISCO project, aimed at supporting the escalating demands for computational resources essential for Big Data Analysis and High Throughput Computing. Through the empowering of the IT infrastructure, including a newly established data center, several keys topics were addressed such as increase processing capabilities, fortify security measures, and ensure the reliability of technological foundations. Aligned with contemporary standards, this enhancement stands as a pivotal stride in the evolutionary trajectory of the IBISCO project, specifically tailored to accommodate the intensifying requirements of computational resources for advanced data analytics. The abstract outlines achieved objectives, highlights technologies implemented, and anticipates the transformative impact on overall IT operations within the expansive domain of the IBISCO project.

C1-5

Authors: G. Russo, G.B. Barone, D. Bottalico, L. Carracciuolo, A. Izzo, D. Michelino, C. Piccolo, G. Sabella

Title: The IBiSCo infrastructure realized at UNINA

ABSTRACT:

The IBiSCo project has allowed us at UNINA to update the hardware infrastructure, both for power and cooling, and for servers, storage, networking. We describe with some detail what has been done, namely new UPS, chiller revamping, new servers for a joint cluster (UNINA/CNR/INFN) with 128 GPUs, new servers for a cloud (OpenStack) cluster, and 2 Pbytes of storage. For the networking, we added Ethernet switches at 25 Gb/s, and Infiniband switches, thus posing the basis for a light-HPC cluster organization of the computational power.

C1-6

Authors: D. Di Bari, A. Italiano, A. Leserri, D. Grittani, E. Renna, E. Serra, F. De Biase, G. Donvito, G. Lamacchia, G. Vino, G. Barbaro, G. Savarese, L. Cristella, M. Panetta, M. Antonacci, M. Patano, M. Perniola, N. Foggetti, S. Nicotri, R. Gervasoni, R. Valentini, R. Sguera, V. Digravina, V. Spinoso, G. Miniello, M. Tangaro

Title: The CED Infrastructure of the IBiSCo-ReCaS-Bari data center after the IBiSCo upgrade

ABSTRACT:

To allocate and keep running the new computing resources acquired with the IBiSCo Project (PIR01_00011 - PON R&I 2014-2020 call 424/2018 Action II. 1), it was necessary to improve the basic infrastructure of the ReCaS-Bari data center

In particular, the rack islands in the computing room were completed adding an island with 20 new racks. A second island was assembled using the 18 racks coming from the dismantling of the Bc2S Computer Room.

Moreover, the fire extinguishing system had to be expanded, the number of refrigeration units increased and the control network extended to newly installed racks.

The activity also included the upgrade to the MAN network and the acquisition and commissioning of a three-way-generator to produce electricity and cold water.

It is important to note that the IBiSCo underlying infrastructure is also used to operate the computational resources acquired by the CNR-ITB with funds of PON project CNRBIOmics (PIR01_00017).

C1-7

Authors: D. Di Bari, A. Italiano, A. Leserri, D. Grittani, E. Renna, E. Serra, F. De Biase, G. Donvito, G. Lamacchia, G. Vino, G. Barbaro, G. Savarese, L. Cristella, M. Panetta, M. Antonacci, M.Patano, M. Perniola, N. Foggetti, S. Nicotri, R. Gervasoni, R. Valentini, R. Sguera, V. Digravina, V. Spinoso, G. Miniello, M. Tangaro

Title: The computing, storage and network Infrastructure of the IBiSCo-ReCaS-Bari data center after the IBiSCO upgrade

ABSTRACT:

The upgrade of the Local Area Network of the ReCaS-Bari data center implemented in the framework of the IBiSCo Project (PIR01_00011 - PON R&I 2014-2020 call 424/2018 Action II. 1) will be presented.

It will also be described the expansion of computational and storage resources implemented in the framework of the IBiSCo Project but with smaller contributions from the CNRBIOmics (PIR01_00017) and LifeWatchPlus (PIR01_00028) projects, also financed under the R&I 2014-2020 PON call 424 /2018 Action II. 1. In addition the Bari data center hosts the resources acquired by the CNR-IREA group within the IBiSCo Project.

At the end of the upgrade the data center in Bari can count on 78 racks, 321 computing server, for a total of 36979 cores, 69 GPU's, disk storage: 12.5 PB GPFS and 2.5 PB XrootD, 20 PB of tape storage, and the possibility to activate 426 point to point connections non blocking at 10 Gbps and 40 at 25 Gbps.

C1-8

Authors: G. Cantele, P. Lucignano

Title: Le infrastrutture realizzate presso il CNR

ABSTRACT:

One of the key objectives of the IBiSCo project for the National Research Council has been to build up an infrastructure for high performance computing, aimed at hosting computational problems with different nature and requirements.

To this purpose, the computing nodes have been equipped with a significant amount of RAM memory (1408 GB per node); 4 GPU NVIDIA V100 accelerators; Infiniband connection.

The versatility of this configuration has allowed it to run problems belonging to different fields, ranging from material science to statistical mechanics to molecular dynamics to others.

In this talk we shortly outline the main architecture of the infrastructure and its versatility with respect to the job kind/size/requirements.

C1-9

Authors: C. Grandi

Title: ICSC: The Italian National Research Centre on HPC, Big Data and Quantum computing and TeRABIT

ABSTRACT:

ICSC, the Italian National Research Centre on HPC, Big Data and Quantum computing is one of the five Italian National Centres funded by the NRRP. The aim of ICSC, started in September 2022, is to create the national digital infrastructure for research and innovation, leveraging existing HPC, HTC and Big Data infrastructures and evolving towards a cloud data-lake model. The ICSC IT infrastructure is built upon existing scientific digital infrastructures as provided by the major national players: GARR, the Italian NREN, provides the network infrastructure, whose capacity will be upgraded to multiples of Tbps; CINECA hosts Leonardo, one of the world largest HPC systems, with a power of over 250 Pflops, to be further increased, and complemented with a quantum computer; INFN contributes with its distributed Big Data cloud infrastructure, built in the last decades to respond to the needs of the HEP community. On top of the IT infrastructure, several thematic activities focus on the development of tools and applications in several research domains. TeRABIT is a project, also funded by the NRRP, started in January 2023, with the aim to upgrade the IT infrastructures on GARR, CINECA and INFN, in total synergy with ICSC.

C2-1

Authors: A. Abraham, S. Bianco, A. Chiariello, M. Conte, F. Di Pierno, A. Esposito, A. Fontana, S. Goha, F. Vercellone, M. Nicodemi

Title: High-performance computing in biological physics

ABSTRACT:

Chromosome spatial organization in the nucleus of cells is fundamental to the control of the activity of our genome as, for example, it defines the contacts between genes and their regulators. We review recent technological and conceptual developments that we introduced in the computer simulation of models of polymer physics of chromosome folding to overcome the challenges deriving from the huge size of those systems in terms of the atoms they are made of. By those advancements, and by exploiting the computational resources of the Università di Napoli, of INFN and CINECA, we showed that the use of Molecular Dynamics, Monte Carlo and Machine Learning approaches is key to understand the mechanisms underlying the 3D organization of the human genome and to predict how mutations, typically associated to human diseases, impact gene activity by rewiring the network of regulatory contacts. Taken together, those computational developments are deepening our understanding of how our DNA works and are ushering in novel biomedical approaches for genetic diseases such as congenital disorders and cancer.

C2-2

Authors: S. Pardi

Title: IBiSCo usage for the Belle II experiment

ABSTRACT:

Belle II is an international experiment conducted at the SuperKEKB e+e- collider at the KEK laboratory located in Tsukuba, Japan. The scientific program focuses on searching for new physics beyond the Standard Model by accumulating approximately 50 times more data (~50ab-1) than its predecessor, Belle, which operated at the KEKB accelerator. At maximum luminosity, the experiment will produce 40TB of data per day, which must be analyzed using a distributed computing infrastructure.

The IBiSCo data center in Napoli plays a significant role in these computing activities, representing one of the largest Italian sites within the collaboration. It serves as a testing site for several innovations to be incorporated into the computing model. Additionally, the Napoli site plays a crucial role in the collaboration by coordinating the sites and network of the experiment.

In this paper, we will summarize the activities conducted using the IBiSCo infrastructure, highlighting its key role in supporting the experiment.

C2-3

Authors: G.B. Barone, D. Bottalico, L. Carracciuolo, D. Michelino, G. Sabella

Title: Design, implementation, and validation of a heterogeneous resource for high performance computing

ABSTRACT:

The work describes the strategies and the solutions on which are based the implementation and the validation of an HPC heterogeneous resource installed at the data center of the University of Naples "Federico II" thanks to the funds of the Italian National Project IBiSCo (Infrastructure for Big Data and Scientific COmputing).

C2-4

Authors: M. Vizzuso

Title: Quantum hybrid algorithms for combinatorial optimization problems

ABSTRACT:

The Quantum Approximate Optimization Algorithm (QAOA) is a promising hybrid quantum-classical algorithm that can solve combinatorial optimization problems. The quantum part of the algorithm involves using parametric unitary operations on a quantum computer to prepare a trial solution state. The parametric QAOA angles are variationally optimized minimizing a cost function using classical methods. For classical optimization, we used a parallel optimization (optimalparallel) with method L-BFGS-B. We study a generalized QAOA ansatz that includes corrections to the Trotter expansion at the first and second order based on the Baker-Campbell-Hausdorff (BCH) expansion. By utilizing terms in the BCH expansion as additional control unitaries, each with its own angle, we can improve convergence compared to standard QAOA. The additional angles are treated as independent free parameters, rather than keeping them fixed to the prescription of the BCH expansion, resulting in a cost function simpler to deal with. In our work all quantum calculations (Pauli matrices and unitaries) were performed on classic simulators.

C2-5

Authors: B. Spisso, G. Carlino, F. Cirotto, A. D'Onofrio, A. Doria, G. Sabella

Title: The HPC IBiSCo seeds for INFN experiments workflow-based cluster

ABSTRACT:

The deployment of High-Performance Computing (HPC) clusters in the field of high-energy physics experiments necessitates specialized software capable of managing and analyzing the colossal datasets produced by these scientific endeavors. The software infrastructure of such HPC clusters is a critical component, comprising a suite of tools for job scheduling, resource management, data processing, and simulation.

The work illustrates the IBiSCo HPC cluster of the Naples INFN section focusing on the different interaction workflow considering as principal use case the ATLAS experiment. Beside the base software workflows are treated additional software tools (Jupyter Notebook, Miniconda, DASK, etc...) chosen to facilitate the exploitation the computing resources to the experiment's

communities. As a possible bridge to the HTC paradigm, was developed a prototype HPC/HTC mixed workflow, based on the storage manager dCache. Eventually, the future extension of the cluster due the ICSC program, is described.

C2-6

Authors: P. Hegde, L. Carracciuolo, G. Sabella, A. Tortora

Title: The User Experience in IBiSCo Data Center

ABSTRACT:

The Data Center of the Department of Physics at the University of Naples Federico II hosts diverse research activities spanning physics, engineering, and humanities. All these activities share a common thread: the utilization of HPC (High Performance Computing) and HTC (High Throughput Computing) resources by users. Within the framework of the IBiSCo project, services have been implemented to provide users with the best possible experience. Among all the services provided, relevance is given to the realization of a Wiki page and a collaborative computing platform. Equally crucial aspect is security management, ensuring the protection of user activities from both internal threats and potential attacks or infiltrations from external. This paper aims to illustrate the implementation of various user experience services along with their corresponding support services.

C2-7

Authors: A. Tortora, D. Bottalico, A. Izzo, D. Michelino, S. Pardi, G. Sabella

Title: The Monitoring System in the IBiSCo Data Center

ABSTRACT:

Data Centers require uninterrupted operation, making it extremely important to monitor key metrics related to power supply and cooling systems. Monitoring IT resources is crucial for job allocation within computing clusters and for regulating individual machine temperatures. This also indirectly assesses the effectiveness of cooling systems at the rack level. These aspects inspired the design and implementation of the monitoring system developed at the Data Centers of the Monte Sant'Angelo complex at the University of Naples and of the Naples section of INFN, during the IBiSCo project. The Monitoring system utilizes the Round Robin Database for representing metrics related to power and cooling as well as those obtained from IT technologies. This article begins with a description of the requirements and proceeds to present the measurements taken by the instrumentation at the Data Centers. After that, the focus is on a description of the monitoring system implementation.

C2-8

Authors: G. Sabella, M. Delli Veneri, S. Pardi, B. Spisso, D. Michelino, F. Di Nucci, G. Carlino, A. Doria

Title: Cloud Infrastructure and storage for IBiSCo and ICSC at the INFN Section of Naples

ABSTRACT:

Cloud computing has revolutionized the way computational resources and storage are managed in research institutions. The National Institute for Nuclear Physics of Naples has embraced this paradigm shift by establishing a robust cloud infrastructure leveraging OpenStack and dCache technologies. OpenStack provides the foundation for managing compute, storage, and networking resources in a highly scalable and flexible manner, while dCache serves as the distributed storage solution for handling large volumes of data. In this article, we delve into the architecture and implementation of the cloud infrastructure deployed at the INFN of Naples. We discuss the role of OpenStack in orchestrating virtualized resources, enabling researchers to provision and manage computational instances seamlessly. Moreover, we explore how dCache integrates with OpenStack to provide reliable and efficient storage solutions for the diverse data requirements of scientific experiments and analyses. Through the integration of OpenStack and dCache, the INFN of Naples has established a versatile and agile cloud environment capable of meeting the demanding computational and storage needs of various research projects. This article sheds light on the challenges encountered and the strategies employed in harnessing the potential of cloud technologies to advance scientific research and innovation at the INFN of Naples.

C2-9

Authors: M. Delli Veneri, G. Longo, M. Brescia, G. Russo

Title: High performance computing for radio-interferometric applications in Astrophysics

ABSTRACT:

Radio Interferometrical instruments like the Atacama Large ALMA and SKA are transitioning in the Big Data regime, and thus the processing and analysis of interferometric data cubes requires vast amounts of computational resources. The employment of Deep Learning (DL) models for the cleaning and analysis of interferometric data has been proved successful on simple mock datasets, but their application on real data has still to be proven successful due to the lack of realistic simulations which are paramount for training and testing these models. In this article, we show how the IBISCO HPC Cluster has been used to generate large datasets of highly realistic data cubes thought he ALMASim framework (an MPI-based simulation pipeline), and how this datasets have been used to train and test Deep Focus, a DL metalearner which leverages multiple GPUs distributed across the IBISCO HPC Cluster to test in parallel multiple DL models with the aim of finding the best performing model for the given task.

C2-10

Authors: S.Conte, G.Russo, M.Salvatore, A.Tortora

Title: Application of the IBISCO Data Center for cultural heritage projects

ABSTRACT:

The hardware architecture of the IBiSCo Data Center is characterized by computing and storage nodes. The storage equipment is also functional for interdisciplinary scientific cooperation, in particular for the MAGIC project, born from the collaboration between the "Ettore Pancini" Department of Physics and the Department of Humanistic Studies of the University of Naples "Federico II". The objective is to create a Service Center, designated for the digitization and archiving of ancient manuscripts and books, preserved in national and international libraries, in line with the digital transformation, which is affecting cultural heritage. The project, in its various phases, involves a corpus of illuminated manuscript manuscripts of the Divine Comedy, incunabula and sixteenth-century manuscripts from the Pontaniana Academy and the Society of Sciences, Letters and Arts, selected manuscripts preserved at the Girolamini Library in Naples.

The archiving and analysis process requires the use of a Data Center for the management of large amounts of data, which guarantees data security, as well as a monitoring system, which allows an operator to insert the file names into a database and all the corresponding metadata, check the file formats (the original very high resolution images and the images created from them), the digitization progress and the percentage of storage occupancy.

C3-1

Authors: M. Bonano, S. Buonanno, F. Casu, C. De Luca, G. Donvito, R. Lanari, M. Manunta, G. Onorato, M. Perniola, G. Vino, I. Zinno

Title: Methods and algorithms for the massive processing of synthetic aperture radar interferometric satellite data for the analysis of deformation phenomena of the soil and the built environment

ABSTRACT:

We will describe the Differential SAR Interferometry (DInSAR) processing chain, for the extensive mapping of surface displacements, based on the Parallel Small BAseline Subset (P-SBAS) technique and its implementation at the IBiSCo-ReCaS-Bari data center. The exploited infrastructure component consists in 6 bare-metal nodes accessible via a front-end and equipped with 2 AMD EPYC 7H12 64-Core processors, 3 NVIDIA A100-PCIE-40GB GPUs and 2TBytes RAM and 21 TByte of local storage. Moreover, a shared storage of about 280 TByte, accessible via NFS, is used to save the DInSAR results for later usage. We have deployed, on every node, the Sentinel-1 P-SBAS processing chain operating at medium spatial resolution (about 80 m x 80 m), made of several steps executed through an IDL code performing in a multi-core parallel way. The presented analysis will be focused on the whole Italian territory and will be related to the products obtained by processing the SAR data acquired during the last seven years from both the ascending and descending orbit directions of the Sentinel-1 constellation. Moreover, we will show the results of a further experimental analysis

carried out by processing, through the Full Resolution P-SBAS processing chain, the datasets collected with high spatial resolution (about 3 m x 3 m), over some of the main Italian cities, by the sensors of the first (CSK) and second (CSG) generation of the COSMO-SkyMed constellation. In this case, the focus will be on the high computing capabilities offered by the GPUs of the infrastructure to perform the P-SBAS analysis at full spatial resolution and in relatively short time frames, to investigate possible displacements affecting the built-up environment and critical infrastructures.

C3-2

Authors: G. Fontanarosa, G. D'Avino, P. L. Maffettone.

Title: Computational fluid dynamics of complex multiphase systems through electrohydrodynamic effect (EHD) for microfluidic applications

ABSTRACT:

In this work, the behavior of a liquid droplet with low electrical conductivity under the influence of a constant electric field is investigated by numerical simulations. Both leaky-dielectric (air-distilled water) and pure dielectric model (air-PDMS) systems are considered. The electrohydrodynamic governing equations are solved by a volume-of-fluid finite volume method implemented in the open-source software Basilisk.

The deformation of the droplet is mainly influenced by a critical threshold of the applied electric field. The droplet is deformed without break-up below this threshold, reaching a new equilibrium state through damped oscillations (subcritical regime). Beyond this limit, the droplet gets an elongated shape and then ruptures, releasing microdroplets or creating a single filament (supercritical regime). To understand the impact of the relevant parameters on the droplet dynamics, simulations are carried out by varying the fluid properties (viscosity and electrical conductivity) and the electric field. The results show that the viscosity influences the transient behavior of the droplet in the subcritical domain, causing the system to respond with overdamped or underdamped oscillations. Concerning the supercritical regime, an inverse relationship between electrical conductivity and the number and size of droplets formed post-breakup is observed. Moreover, as the electric field or viscosity increase, a transition from an emission regime to a single-stranded regime is found. The electric field affects the steady-state attained in the subcritical regime, and the number of droplets produced in the supercritical one.

C3-3

Authors: R.M. Vitale, P. Amodeo

Title: Large scale molecular dynamics simulations of biomolecular systems in membrane environment

ABSTRACT:

Membrane proteins are large and usually multimeric molecular systems, which represent relevant pharmacological targets for a wide range of diseases. For this reason, continuous efforts from both the scientific community and pharmaceutical companies have been devoted so far either to understand their mechanism of action at molecular level, and to the discovery of ligands able to modulate their activity. Computational methods represent an unvaluable tool in the drug discovery process. In particular, in absence of experimental complexes, molecular dynamics (MD) simulations allow to assess the stability of docking binding poses and/or to evaluate the conformational transitions of biological systems occurring upon binding. However, MD simulations of such large biological systems are very challenging due to the huge number of atoms involved and simulated time scale required, ranging from hundreds of nanoseconds to microseconds, which renders such kind of studies highly demanding in terms of computational CPU and/or GPU resources. Here we discuss some examples of large-scale MD simulations performed on the IBiSCo HPC cluster, with a special attention to results that can be only obtained with the computational power provided by this kind of facilities.

C3-4

Authors: M. Massuoli, R.M. Vitale, M. Paturzo, P. Amodeo

Title: Development and application of parallel CPU/GPU tools to extend the characterization of direct and solvent-mediated H-bond interactions in large structural protein databases

ABSTRACT:

Since hydrogen bonds (H-bonds) are fundamental interatomic interactions for 3D structure and function of biological macromolecules, many software tools exist for their study. However, they are mostly interactive tools, operating on one/few structures. Even batch or interactive procedures supporting structure ensembles are typically focused on multiple structures of a single system, or, at most, on comparisons among few different structures/ensembles. Furthermore, many tools use hardwired definitions for H-bond donors, acceptors and geometrical parameters and are often unable to cope with some subtle features of structural formats included in databases, such as the Protein Data Bank (PDB), typically to manage either local structural variability, or the presence of solvent and other “small” (but often very important) molecules, and thus requiring heavy pre-editing and cleaning of the structure, with possible loss of relevant information.

We have developed a set of routines, heavily relying on combined CPU+GPU resources available at IBiSCo HPC cluster, performing fully customizable batch processing of the whole PDB dataset to search for H-bonds also involving solvent and other “small molecules”, and trying to provide reasonable assignment of possible interactions even for flexible groups. The final aim is a Web service, endowed with a user-friendly interface, possibly hosted on IBiSCo-related servers, to flexibly explore H-bonds within the PDB dataset, subsequently expandable to perform complex statistical

studies and/or interact with computational approaches, by testing their reliability and/or improving the description of these important but complex interactions.

C3-5

Authors: D. di Bari, A. Italiano, A. Leserri, D. Grittani, E. Renna, E. Serra, F. De Biase, G. Donvito, G. Lamacchia, G. Vino, G.Barbaro, G. Savarese, L. Cristella, M. Panetta, M. Antonacci, M. Patano, M. Perniola, N. Foggetti, S. Nicotri, R. Gervasoni, R. Valentini, R. Sguera, V. Digravina, V. Spinoso, G. Miniello, M. Tangaro

Title: Contribution of the IBiSCo-ReCaS-Bari data center to the Italian National Smart Specialization Strategy

ABSTRACT:

The Bari data center, in addition to serving High Energy Physics community where it plays the role of TIER2 for the ALICE and CMS experiments running at the LHC at CERN in Geneva, has allocated computational resources to communities operating in the sector of Earth & Space Science Informatics, territory monitoring, health, advanced diagnostics, nutrition, quality of life.

To achieve the objectives, the data center has signed agreements with ARPA Puglia, INGV, Planetek, has provided resources and services to the EOSC community using the pay per use mechanism and hosts the CNR computational resources acquired with funds assigned to the PON CNRBIomis (PIR01_00017) project and the GARR computational resources acquired with funds assigned to GARR-X project financed by MIUR on PAC funds (Notice D.D. n. 27 4 of 15 February 2013)

C3-6

Authors: S. Nicotri, M. Antonacci, D. Di Bari, A. Italiano, L. Cristella, F. Debiase, G. Donvito, D. Grittani, A. Italiano, A. Leserri, G. Miniello, M.P. Panetta, M. Patano, M. Perniola, L. E. Renna, R. Sguera, R. Valentini

Title: The users of the IBiSCo-ReCaS-Bari data center and their support: organization and tools

ABSTRACT:

This paper is aimed at describing the main use-cases currently supported by the ReCaS-Bari data center. Providing timely support to a heterogeneous and wide plethora of use-cases is a complex process, which requires dedicated effort and procedures, management, organization, and tools. Such a process will be described in detail.

C3-7

Authors: M. Antonacci, A. Italiano, R. Valentini, G. Donvito, S. Nicotri, R. Sguera, M. Perniola, L. E. Renna

Title: IBiSCo-ReCaS-Bari Cloud site evolution: Introducing the architecture and implementation of the upgraded IaaS infrastructure

ABSTRACT:

The IBiSCo Project has provided a boost, in terms of available resources, to one of the assets of the IBiSCo-ReCaS-Bari, the cloud. This talk aims to describe the architecture and the implementation of the upgraded cloud instance. Sharing the details about the main components deployment [compute, storage and network] will highlight how the resource is exploited and used by the end users. For instance, the storage has been dramatically reorganized. The current deployment can provide at the same time fast disk access to specific workloads and reduce disk space consumption for block storage use cases, the erasure-code features allow to reduce the underlay storage usage per GiB of user space.

C3-8

Authors: M. Antonacci, M. Patano, E. Serra, M. Tangaro, G. Savarese

Title: Delivering integrated high-level services at Bari IBiSCo-ReCaS site through the PaaS dashboard and the Orchestration system

ABSTRACT:

In the realm of IT services provisioning, the paradigm of cloud computing, facilitated by advancements in network and virtualization technologies, has emerged as a dominant force over the past decade.

This model encompasses various service delivery modalities such as Software as a Service (SAAS), Data as a Service (DAAS), Hardware as a Service (HAAS), Platform as a Service (PAAS), and Infrastructure as a Service (IAAS), aiming to provide a heterogeneous and distributed array of resources to end-users, albeit with varying degrees of transparency regarding technical specifications.

Key players in the commercial cloud computing sector, including Amazon Web Services, Microsoft Azure, Google Cloud Platform, Alibaba Cloud, offer paid services spanning computing, networking, content distribution, database management, application deployment and analytics. These services are readily accessible to customers.

This model of service supply and access is becoming increasingly important, even in general purpose Data Centers with a significant focus on research, such as ReCaS Bari, or in data center federations like INFN-Cloud, of which ReCaS is a member.

The first part of this contribution briefly describes the existing physical characteristics of the RecaS Bari Data Center as well as the distributed and federated architecture, built on heterogeneous and geographically distant cloud sites, to which ReCaS belongs.

The second part describes the current state of the PaaS Orchestration system which allows end-users to deploy virtual infrastructures and services via a simple and intuitive graphical web interface, the PaaS Dashboard. This web application enables users toAuthenticate, choose a service to deploy from a catalog of predefined templates, configure and customize the deployment using a simple form, monitor and manage deployments through dedicated menus and views, and finally receive a notification as soon as the deployment is complete.

In the final part, future developments of the platform are described.

C3-9

Authors: G. Vino, M. Antonacci, G. Donvito, A. Italiano, M. Perniola

Title: GPU-based cluster at IBiSCo-ReCaS-Bari for containerized scientific applications

ABSTRACT:

The IBiSCo-ReCaS-Bari datacenter enriches its service portfolio providing a new GPU-based HPC cluster for Bari University and INFN users. This new service is ideal for complex applications that require a massively parallel processing architecture. The cluster is equipped with cutting edge Nvidia GPUs, like V100 and A100. Artificial intelligence, complex model simulation (weather and earthquake forecasts, molecular dynamics and galaxy formation) and all high precision floating-point based applications are possible candidates to be executed on the new service. The cluster is made up of 13 nodes and has a total computing resource of 2500 cpus, 20 TB RAM, 6 TB local SSD disk per node and 47 high performance GPUs (27 Nvidia A100 and 20 Nvidia V100). Each node can access the IBiSCo-ReCaS-Bari distributed storage based on GPFS amounting to 12 PB. Applications are executed only within Docker containers, conferring to the HPC/GPU cluster features like easy application configuration and execution, reliability, flexibility and security. Users can request web-based IDEs (JupyterLab and RStudio) and/or a job orchestrator for submitting complex workflows represented as DAG (Directed Acyclic Graphs). Apache Mesos, a distributed resource management system, is used to orchestrate the usage of a computer cluster equipped with GPUs in a scientific environment. Chronos, Marathon combined with Mesos, provide features like scalability, fault-tolerant, resource constraints and workflow orchestration. Users appreciated an impressive speed-up of their applications up to a factor of 10. The evolution of the service, where a performance evaluation of Kubernetes as a replacement for Apache Mesos, is in the pipeline

---

C3-10

Authors: M. La Salandra, S. Nicotri, G. Donvito, R. Colacicco, P. Dellino, D. Capolongo

Title: High-performance computing in UAV photogrammetry

ABSTRACT:

Mapping large areas of the Earth's surface through Unmanned Aerial Vehicle (UAV) imagery poses challenges in terms of data storage, computational resources and, above all, processing time. The capability to acquire large, high-resolution datasets, together with the complex photogrammetric processing (Structure from Motion) needed for accurate mapping, requires the usage of high-performance computing resources. The time-sensitive nature of applications, as e.g. disaster response and environmental monitoring, exacerbates the need for near-real-time processing. In this context, the present study introduces the implementation of a distributed photogrammetric workflow based on a divide-and-conquer approach and on the HTCondor software framework to exploit the ReCaS-Bari HTC/HPC cluster. Performance tests of the workflow, which leverages computing parallelism, hybrid bundle adjustment and GPU usage for image matching and depth map calculation, show a significant reduction in processing time for large UAV image datasets with up to 88% improvement over state-of-the-art approaches.

C3-11

Authors: A. Russomanno

Title: Many-Body Quantum Dynamics Simulations with the Ibisco resource

ABSTRACT:

My work is simulating the dynamics of many-body quantum systems in different situations. This amounts to diagonalize matrices, implement dynamical evolutions, store lots of data, and also massively parallelize, when averages over disorder or noise are in order. The Ibisco resource meets these needs of RAM, hard-disk space, and makes available many many processors to run parallelized codes. Without this resource my working life would be significantly harder. In this talk I will tell you some more details of this very important relationship in my life, focusing on the published works whose numerics I have performed on Ibisco, and the reason why such a powerful numerical resource was essential.

C3-12

Authors: S. Izzo, S. Amitrano, M. Savoia, E. Prezioso, F. Giampaolo, F. Piccialli

Title: Unlocking AI Capabilities: Advanced Applications through High-Performance Computing

ABSTRACT:

In a world eager to explore the full possibilities of Artificial Intelligence (AI), our journey is often limited by how much data we can process. High-Performance Computing (HPC) steps in as a game-changer, making it possible to push past these limits. This talk highlights how the IBiSco cluster is a key player in taking AI research and its applications to the next level.

Our exploration covers four case study: 1) The acceleration of the Deep Learning-based YOLO object detection algorithm through parallel processing, significantly reducing training times; 2) The advancement of Federated Learning, utilizing HPC's distributed computing to bolster model training across decentralized datasets; 3) The innovation in seismology through HPC-enabled simulation and dataset creation, offering groundbreaking insights into earthquake prediction and geological phenomena; 4) The application of deep learning in environmental conservation, specifically using satellite imagery analysis to detect illegal landfills. Each case study illustrates HPC's contribution to overcoming the computational barriers that hinder AI's progress, showcasing its ability to enable more sophisticated, efficient, and impactful AI solutions. This presentation aims to highlight HPC's transformative impact on AI capabilities, illustrating its essential role in driving forward scientific research and practical applications across diverse domains.

C3-13

Authors: I. Piergentili

Title: Open Access repositories for scientific literature and research data

ABSTRACT:

The Open Access Repository (OAR) project was created to implement Open Access policies and to preserve and share the scientific research results, including research data, of INFN Authors. With the approval of “Disciplinare per l’accesso aperto ai prodotti della ricerca dell’INFN”, in July 2023, OAR officially became the INFN's institutional repository.

In the past two years we studied the optimization of our institutional repository through the bulk upload of both digital and scanned documents, such as INFN Technical Notes and documents related to the ADONE project (1969-1993). Moreover, since 2023, a collaboration with INFN-CNAF has been established to migrate the repository to InvenioRDM (CERN).

To better support the OAR migration recent activities have been mainly focused on the record upload process, exploiting the following topics:Authentication, metadata customization,Author and entity names disambiguation and product approval flow management.

In addition to the study of the repository structure, we worked on communication as well, introducing the tool to users through a specific website, and user training activities about the use of OAR.

C3-14

Authors: L. Di Filippo, M. Mutarelli

Title: Setting up an automated framework for bioinformatic and multiomic analyses on IBiSCo

ABSTRACT:

Modern Biomedical Research has been revolutionized by the emergence of High Throughput Sequencing technologies and all their applications offering unprecedented insights into physiological mechanisms and diseases. Genomics, Transcriptomics, Epigenomics and their integration are now affordable and available to researchers, but at the cost of producing large amount of data that require high computational resources, tailored software and expertise to manage and analyze them.

The use of workflow managers like nextflow to design analysis pipelines, together with containerized software and open-source repositories, is the no-turning back direction of bioinformatic applications development. This approach allows to concatenate analysis steps performed by independent, even incompatible software, managing the correct definition of inputs and outputs between them, organizing job submission to the job manager and allowing to resume the analysis after failure for any reason or after partially changing the input or the analysis steps.

The presentation will show our experience in setting up several analysis pipelines and running on the IBiSCo resource the analysis of a number of projects that without the collaborative infrastructure would have been impossible with limited human resources.

C3-15

Authors: C. Lo Giudice, F. Licciulli, G. Miniello, M. Moscatelli, S.N. Cox, A. S. Varvara, B. Fosso, M.A. Tangaro, R. Cilli, D. Traversa, G. Donvito, E. Capriotti, M. Chiara, F. Zambelli, G. Pesole

Title: Development of a state-of-the-art computational environment for handling human genetic data: the effort of ELIXIR-IT

ABSTRACT:

Nucleic acid sequencing technologies are becoming more accessible by the day, paving the way for their use beyond research and the flourishing of sequencing-based services in many domains, including the health sector. Applications like personalized medicine and pharmacogenomics promise to redefine our approach to treating ailments like cancer and genetic diseases. However, several interwoven technical, legal and ethical hurdles must be addressed to harness this opportunity fully. Nucleic acid sequencing technologies are becoming more accessible by the day, paving the way for their use beyond research and the flourishing of sequencing-based services in many domains, including the health sector. Applications like personalised medicine and pharmacogenomics promise to redefine our approach to treating medical conditions like cancer and

genetic diseases. However, several interwoven technical, legal and ethical hurdles must be addressed to harness this opportunity fully.

From a scientific data infrastructure point of view, sequence data from human samples represent a unicum since they are sizable and, at the same time, must be transferred, stored, managed, and used in compliance with strict ethical and legal prescriptions, setting them apart from most other types of scientific data. Thus, the development of efficient solutions for secure human genetic data handling is in high demand, and their actionability as ready-to-use and cost-effective services represents a primary goal for a biological data public infrastructure such as ELIXIR-IT.

Hereby, we describe the architecture of a VM-based service integrated into a wider computational environment designed for managing human genetic data from production to their deposition in access-controlled repositories, to be based in the ReCaS datacenter in Bari, Italy.

First, data is transferred via SSH tunneling from the sequencing facility to a storage facility, named BioRepository (CNR at ReCaS-Bari), which provides data-at-rest encryption and geo-redundant replica and backup storage located at CNR-ITB (Milan) and CNR-ICAR (Naples), respectively.

A scalable virtualized computational environment will be provided over the cloud for the user to process their data. The VM-based environment will deploy state-of-the-art bioinformatics tools for data analysis steps including quality control, mapping to a reference, variant calling and prioritization, and VCF handling. The software tools will be self-contained by adopting containerization (e.g. Docker, Singularity) and package management (e.g. CONDA, MAMBA) solutions that will maximize their compatibility with the software environment, facilitate the release of updates, and improve the reproducibility of the analysis workflow. Workflow management systems such as Snakemake, Nextflow, or Galaxy will complete the analysis environment made available to the user.

The analysis software packages' installation on the VMs will be accomplished by IT automation engines (e.g., Ansible) to further improve the service's flexibility, customizability, and maintainability. Encryption at the file system level of the virtual volumes used by the VMs will provide secure storage for the data while they are analyzed. Furthermore, the virtual environments will also enjoy shared access to regularly updated bioinformatic reference databases stored in the BioRepository facility.

After the analysis process is completed, the user will be able to access relevant downstream services for data FAIRification, deposition, and discoverability offered by the same facility, thus easily integrated into the processing workflow to minimize off-site data transfers. First, a federated node of the EGA human genome-phenome archive (FEGA) is being developed and will be maintained by ELIXIR-IT in the same ReCaS facility. The FEGA node will make metadata browsable and searchable through the central EGA repository while allowing data owners to retain complete control over the data access authorization procedures. Finally, a service based on the Beacon protocol will complete the offer, allowing the discoverability of the datasets hosted by the FEGA node through its genomic variant-based query system from any other node of the Beacon network.

This integrated services approach represents a substantial leap forward in the infrastructure-level management of human genetic data in Italy, delivering a resource-wise, maintainable, and scalable solution to support their use for research and health applications.

C3-16

Authors: V. Barreca, D. Ninno, G. Cantele

Title: Material science problems on the IBiSCo infrastructure

ABSTRACT:

Material design and prediction is one of the most active activities in the field of material science and physics.
Atomistic simulations, based on first principles approaches such as density functional theory, can nowadays push the limit of the simulations in terms of systems size and number of atoms, thanks to the use of large-scale computational facilities for high-performance parallel computing. In the last years, some of the widely used packages (e.g. Quantum-ESPRESSO and VASP), have been upgraded to fit novel computational resources, as those offered by graphical accelerators, like the NVIDIA GPU installed on IBiSCo.

In this talk we will show some of the applications of these packages on the material study and design, with particular reference to two-dimensional materials and van der Waal heterostructures.

C3-17

Authors: G. Viola, A. Della Pia, L. Russo, C. Siettos

Title: Towards machine learning-based PDE identification for crowd dynamics

ABSTRACT:

A data-driven black-box macroscopic model for crowd (pedestrian) dynamics is here identified employing a state-of-the-art machine learning approach, namely Convolutional Neural Networks. Training data are generated by numerical simulation of the Hughes model, a seminal first-order model of crowd dynamics inspired by fluid dynamics conservation laws, which is employed to investigate pedestrian motions in a corridor.

The independent variables of the model are the crowd density ρ and the potential ϕ. The numerical data are employed to train a Convolutional Neural Network (CNN), which takes as inputs the variables ρ and ϕ and gives as output the time-derivative of the density, ∂ρ/∂t. The trained CNN model is then tested by generation of new data, namely starting from initial conditions different from those employed in the training step.

The performance of the learned black-box partial differential equation (PDE) model are thus compared with a classic computational fluid dynamics approach to solve the Hughes system of equations. The potential of the proposed methodology for future applications is finally highlighted, discussing the possibility of extending the CNN-based PDE learning workflow to experimental measurements and/or high-order numerical data (e.g. obtained by agent-based Monte Carlo simulations) of crowd dynamics.

END OF BOOK