

Symposium on HPC and Data-Intensive Applications in Earth Sciences - Day2

Keynotes Session - Chairman: V. Balaji (Princeton University & NOAA, USA)

Speaker: B. Lawrence (NCAS & University of Reading, U.K.)

Title: "Infrastructure for Environmental Supercomputing: beyond the HPC!"

09:00 - 09:45 **Short-Abstract:** We begin by motivating the problems facing us in environmental simulations across scales: complex community interactions, and complex infrastructure. Looking forward we see the drive to increased resolution and complexity leading not only to compute issues, but even more severe data storage and handling issues. We worry about the software consequences before moving to the only possible solution, more and better collaboration, with shared infrastructure. To make progress requires moving past consideration of software interfaces alone to consider also the "collaboration" interfaces. We spend considerable time describing the JASMIN HPC data collaboration environment in the UK, before reaching the final conclusion: Getting our models to run on (new) supercomputers is hard. Getting them to run performantly is hard. Analysing, exploiting and archiving the data is (probably) now even harder!

Speaker: Kwo-Sen Kuo (N.A.S.A. Goddard Space Flight Center, USA)

Title: "Big Data Infrastructure for Earth Science Data-Intensively Analysis"

09:45 - 10:30 **Short-Abstract:** What is "Big Data" exactly? How is Big Data infrastructure different from "small data" infrastructure? What infrastructures are NASA investigating? What are the "components" or "ingredients" of an "optimal" Big Data infrastructure for Earth Science data-intensive analysis? How do we get there? These are some of the questions I will be exploring in this presentation.

10:30 - 11:00

Coffee Break

Big Data - Convener: R. Farneti (ICTP)

Speaker: A. Rietbrock (University of Liverpool, U.K.)

Title: "From VERCE to EPOS: How to build an efficient and comprehensive multidisciplinary research platform for the solid Earth sciences in Europe"

11:00 - 11:25 **Short-Abstract:** The mission of EPOS is to build an efficient and comprehensive multidisciplinary research platform for the solid Earth sciences in Europe. In particular, EPOS is a long-term plan to facilitate integrated use of data, models and facilities from mainly distributed existing, but also new, research infrastructures for Earth Science. EPOS will enable innovative multidisciplinary research for a better understanding of the physical processes controlling earthquakes, volcanic eruptions, unrest episodes, ground stability, and tsunamis as well as those processes driving tectonics and Earth surface dynamics. EPOS will allow the Earth Science community to make a significant step forward by developing new concepts and tools for accurate, durable, and sustainable answers to societal questions concerning geo-hazards and those geodynamic phenomena relevant to the environment and human welfare. We have successfully concluded the PP (Preparatory Project) phase and expect to continue with the implementation phase in 2015. The key aspects of EPOS concern providing services to allow homogeneous access by end-users over heterogeneous data, software, facilities, equipment and services. The e-infrastructure of EPOS is the heart of the project since it integrates the work on organisational, legal, economic and scientific aspects. Following the creation of an inventory of relevant organisations, persons, facilities, equipment, services, datasets and software (RIDE) the scale of integration required became apparent.

The EU-funded project VERCE (Virtual Earthquake and seismology Research Community in Europe) aims to deploy technologies which satisfy the HPC and data-intensive requirements of modern seismology. As such VERCE can be seen as a pilot study for EPOS centered around the seismological community. As a result of VERCE's official collaboration with the EU project SCI-BUS, access to computational resources, like local clusters and international infrastructures (EGI and PRACE), is made homogeneous and integrated within a dedicated science gateway. In this presentation we give an overview of EPOS challenges and also on the progress achieved with the developments of the VERCE Science Gateway, according to a use-case driven implementation strategy.

Speaker: S. Denvil (IPSL-UPMC, Paris, France)

Title: "IS-ENES Project"

11:25 - 11:50 **Short-Abstract:** Global climate models are at the basis of climate change science and of the provision of information to decision-makers and a large range of users. Within Europe, the European Network for Earth System Modelling (ENES) gathers together the European climate/Earth system modelling community, which is working on understanding and prediction of future climate change. Climate Earth system models are key tools to understanding climate change and its effects on society and are at the basis of the International Panel on Climate Change projections (IPCC). ENES, through IS-ENES, promotes the development of a common distributed modelling research infrastructure in Europe in order to facilitate the development and exploitation of climate models and better fulfill the societal needs with regards to climate change issues. IS-ENES gathers 18 partners from 10 European countries and includes the 6 main European Global Climate Models. IS-ENES combines expertise in climate Earth system modelling (ESM), in computational science, and in studies of climate change impacts.

Speaker: G. Fiameni (CINECA, Bologna, Italy)

Title: "Towards a data science infrastructure"

11:50 - 12:15 **Short-Abstract:** The goal of this talk is to present the transition of a classic HPC infrastructure towards a more integrated, data driven, and flexible system to respond to the users' demand of using heterogeneous services together. Starting from the processing of new requirements to the design of a new infrastructure, the presentation will illustrate the various steps that have led Cineca progressively expand its offering with new components and services and how this transition will evolve in the near future.

Speaker: A. Fiore (CMCC, Lecce, Italy)

Title: "The Ophidia framework: toward big data analytics for climate change"

12:15 - 12:40 **Short-Abstract:** the Ophidia project is a research effort on big data analytics facing scientific data analysis challenges in the climate change domain. It provides parallel (server-side) data analysis, an internal storage model and a hierarchical data organization to efficiently manage large amount of multidimensional scientific data (data cubes). The Ophidia analytics framework provides several MPI-based parallel operators to manipulate (as a whole) the entire set of chunks associated to a data cube. Some key examples include: (i) data sub-setting (slicing and dicing), (ii) data aggregation, (iii) array-based primitives, (iv) data cube duplication, (v) NetCDF-import and export. The most relevant data analytics use cases implemented in national and international projects target fire danger prevention, sea situational awareness, interactions between climate change and biodiversity, remote data analysis, large scale data analytics on CMIP5 data.

Speaker: Kwo-Sen Kuo (N.A.S.A. Goddard Space Flight Center, USA)

Title: "Collaborative Infrastructure to Integrate Local and Cloud Resources and Ease Scientific Collaboration"

12:40 - 13:05 **Short-Abstract:** Modern researchers are busy, busy with scientific investigations and all the chores associated with the activities. We know we need to collaborate in order to better understand our Earth, a system of systems. But, collaborations have been difficult. I will first identify the causes of such difficulties and then describe a prototype collaboration infrastructure that addresses them.

13:05 - 14:00

Lunch Break