IGI training and user support activities

Facilitated by the technological advances in networking and the reduction of costs for hardware manufacturing, in the last years distributed computing infrastructures have established themselves as a key enabler for international scientific collaborations. High Energy Physics has been a pioneer in this sense. A shared infrastructure was the most practical solution in order to analyze the huge amount of data generated by particle accelerators, like the Large Hadron Collider, and allow the data share across the members of the international experiments. This approach was eventually extended to other scientific areas, such as bioinformatics, earth sciences, social simulations, all of which have in common the usage of large and shared data sets, that naturally enable and foster distributed collaborations. This essential role of distributed infrastructures was acknowledged by funding agencies, that supported creation of national and international projects evolving the infrastructures and promoting their usage. With regards to Europe, are worth mentioning, EDG (European DataGrid, 2001-2004), and the series of EGEE (Enabling Grid for E-sciencE) projects, 2004-2010) which evolved in the European Grid Infrastructure (EGI - http://www.egi.eu). EGI relies on the National Grid Infrastructures (NGIs), among which the Italian Grid Infrastructure (IGI). IGI operates the Italian Grid ensuring its full compatibility with the European Infrastructure, based on the compliance to the agreed shared policies, and taking also care of the support to Italian users. The support activity is intended in a very broad sense, being comprehensive of the training and assistance to new users, the consultancy on adapting their applications to the distributed environment, as well as the development of instruments facilitating their approach to the grid infrastructure. In particular, those user communities without a strong IT background have usually difficulties approaching the default grid interfaces, consisting in a rather complex UNIX-like command line. In order to overcome this difficulties, high level web interfaces are developed, with the two-folded purpose of simplifying via a graphical interface the beginners' approach to the grid providing at the same time a seamless, location independent access.

The rest of this paper provides examples of the interaction happened during the last year between the IGI training and user support team (IGI-FUS) and some selected new grid user communities which led to the porting to the national infrastructure of applications belonging to various scientific research fields. The applications are selected to exemplify various aspect and capabilities of the distributed computing infrastructures: CPU power, storage capacity, parallel jobs support, licensed applications support and the federating power for human and technical resources.

One of the main collaboration was established with the Mario Negri institute (http://www.marionegri.it/mn/en/), which is an italian institute for Pharmacological Research, interested, among other things, in the role of haemodynamics in the development of vascular diseases. These studies are supported simulating clinical scenarios with OPENFOAM (http://www.openfoam.com/), a Computational Fluid-Dynamics software. Therefore, the support action performed by IGI-FUS consisted mainly in the installation and configuration of OPENFOAM on the Italian Grid Infrastructure, that allowed the execution of CFD simulations for large arteries.

The initial performances have been quite encouraging for the user community: compared to the resources available to the group, the grid runs allowed for a speed-up factor close to 4 for large CFD simulations, which makes conceivable a more intensive usage, for instance with numerical simulations including settings of fluid-structure interaction (FSI) for blood vessels,

a new challenge for the cardiovascular biomechanics research. Moreover, OPENFOAM is rather generic, supporting an extensive range of partially derivate equations (PDE) that can be applied to chemical reactions, turbulence, electromagnetism, and so on, extending the range of scientific communities potentially supported.

If Mario's Negri researchers are essentially demanding more computing power, scientists from the EMSO (the European Multidisciplinary Seafloor Observatory - http://www.emso-eu.org/) collaboration need to exploit the grid capabilities for the distribution of large quantity of data to scientific collaboration in a federated way. EMSO is an EU ESFRI project aiming at deploying a large set of deep-seafloor observatories for the long-term monitoring of the environmental processes related to the interaction between the geosphere, biosphere, and hydrosphere, including natural hazards. The huge quantity of data detected by the observatories need to be made available to a globally distributed geophysical community, in a scenario remembering very closely the LHC one, and it was therefore natural to design a similar architecture. Data collected from the detector are first saved in the observatories, and copied to a tier0 site, which is as close as possible to the observatory. Data are eventually moved to tier1 sites, where they can be analyzed. This activity has started with two pilot sites, in Catania and Porto Palo, Sicily, and it's rather promising as it's promoting a joint effort both at EU and national levels between infrastructure EU funded projects (EGI and ESFRI).

Again on the geophysical field, IGI-FUS supported a community from the Italian Institute of Geophysics and Volcanology (INGV, bologna section) to exploit the grid resources for the NEMO (http://www.nemo-ocean.eu/) ocean modelling framework which is developed by an european consortium. NEMO is a parallel application that requires the exploitation of the low and mid parallelism capabilities offered by the grid which were recently improved by the middleware releases. To easy the submission process of the application's jobs to the grid a high level web interface was developed based on the IGI-Portal and addressing the user community requirements (https://portal.italiangrid.it/).

Other applications that exploit the parallel jobs support of the infrastructure were ported, in particular collaborating with the computation chemistry community. Among them we quote NAMD, Gaussian, DL_POLY and Gromacs (for a short descrition of these application please refer to https://wiki.italiangrid.it/twiki/bin/view/UserSupport/UserSupportCompchem).

In collaboration with the CNR-ISOF institute (Bologna section - http://www.isof.cnr.it/), using NAMD, IGI-FUS also tested the support of the infrastructure for very long non embarrassingly parallel simulation. The use case implied the simulation of huge molecules (40 thousands atoms) and required months of computational time on a 32 cores machines. The porting was possible developing a grid-wise checkpointing of the calculation.

IGI-FUS collaborated also with an INFN community (the SPES experiment - Selective Production of Exotic Species - http://web.infn.it/spes/) in order to port the ANSYS (http://www.ansys.com/) engineering simulation software (computer-aided engineering, or CAE) that offers a comprehensive range of engineering simulation solution sets providing access to virtually any field of engineering that a design process requires. ANSYS is a commercial suite and in installing it on the infrastructure required particular technical solutions in order to fulfill the licence restrictions. An easy to use web interface to run the simulation was provided through a dedicated portlet of the IGI Portal. The web GUI was built in strong partnership with the user community interested in porting the application to gather their requirements and particular focus

was given to the evolution's audit of the calculations at runtime to check the consistency of the output before the end of the calculation, evaluating possible strategies aimed at saving time, computing resources and at avoiding waste of license usage.
All these support and training activities are done in strong partnership not only with user communities interested in exploiting the grid capabilities, but also in collaboration with other support teams disseminated around Europe. The interaction are mainly mediated by EGI through the so called NGI International Liaisons and Virtual Team frameworks (https://wiki.egi.eu/wiki/Virtual_team) created at the European level in order to improve the efficiency and flexibility of the interaction between the NGIs, EGI.eu and foster engagment of new user communities facilitate the achievement of common research goals. Specifically, IGI-FUS is involved in several virtual team projects such as: collaboration between EGI/NGIs and large ESFRI project ELIXIR, Science Gateway primer, MPI within EGI, Environmental & Biodiversity, Federated Identity Providers within the EGI Community, ESFRI Contact List.
As it can be seen, the range of scientific areas, and related research institutes supported is quite broad and, though the support and consultancy process consolidated by IGI-FUS in the last year could be further extended making distributed infrastructures a general purpose commodity.