Virt&I-Comm.3.2012.25

A VIRTUAL RESEARCH COMMUNITY FOR SYNERGISTIC COMPUTING IN CHEMISTRY, MOLECULAR AND MATERIALS SCIENCE AND TECHNOLOGIES

Antonio Laganà[1], Alessandro Costantini[2]

1. Department of Chemistry, University of Perugia, Perugia, Italy
2. INFN Perugia, Perugia, Italy

ABSTRACT

The present paper outlines the planning lines of the EGI CMMST Virtual Team (VT) project aimed at investigating services and organisational possibilites for the creation of a sustainable Virtual Research Community of the Chemistry, Molecular & Materials Science and Technology computational community. The paper examines the usual ways of accessing compute facilities and how the distributed computing model (as well as its evolution towards a strongly collaborative environment bridging high throughput and high performance computing) meets the needs of the CMMST community. Finally a road map for the implementation of Quality criteria and the establishing of a credit system allowing the rewarding of members' contributions to the progress of the community is proposed.

1 INTRODUCTION

Virtual Research Communities (VRCs) are groups of like-minded individuals and researchers' clusters organised by discipline or computational model. A VRC can establish a support relationship, formalised through a Memorandums of Understanding (MoU), with the European Grid Infrastructure collaboration (EGI, http://www.egi.eu/). EGI VRCs typically have an established presence in their field and represent well-defined scientific research communities. Multi-national scientific communities can draw many benefits from having a VRC partnership with EGI. For example, they can benefit from the resources and support that are available within the National Grid Initiatives (the main stakeholders of EGI.eu), they can benefit from the workshops and forums organised by EGI, they can receive support on resolving specific technical issues with EGI services, and they become involved in the user-focussed evolution of EGI's production infrastructure.
For this reason, the COMPCHEM (https://www3.compchem.unipg.it/compchem/) Virtual Organization (VO) together with the GAUSSIAN VO (both belonging to the Chemistry, Molecular & Materials Science and Technology (CMMST) community) by recognising the advantages that membership as a VRC within EGI will bring has assembled the EGI Virtual Team (VT) project [1] that has taken the first step towards the setup of the CMMST VRC. The VRC status will help the CMMST community to satisfy the requirements of its members concerning the access and use the computing resources that are federated in EGI. The VT will take care of documenting:

- the structure that the proposed VRC should have to represent the CMMST community in EGI;
- the technologies, resources and services that already exist within EGI and could be used to satisfy the requirements of the CMMST VRC;

ISSN: 2279-8773

- the tools that need to be developed or brought into EGI and then integrated with the production infrastructure so the VRC members can efficiently manage and use resources from EGI;
- the applications that need to be imported in order to assemble higher level of complexity CMMST simulations.

2 THE CMMST-VRC PROPOSAL

2.1 Current scenario of computing resources allocation and usage

The evolution of computing technologies has led both to the growth of remotely accessible high-performance platforms (High Performance Computing, HPC) and, on the other hand to the growth of local (departments, research groups, institutions, etc.) clustered computational platforms and networks. These two types of platforms adopt two different models of resources allocation:

- the first (grant model) assigns, as already mentioned, the resources as a result of an ex ante evaluation of a proposal following a specific call for user projects;

- the second (opportunistic model) meets the requirements of registered users as they come without a specific selection of the merits for their request.

The grant model is in general adopted by large scale supercomputing facilities (as well as by the medium-large computer centres) in which specialized staff is devoted to the maintenance of HW and SW. The opportunistic model, instead, is in general adopted either by research groups or departments to run the machines and update the software of their computational members (with the help of some permanent or temporary technical staff).

In large scale computing facilities (as are the computer centres networked in PRACE, http://www.prace-ri.eu/) the users can rely on computational resources, software and skills that evolve according to the current technological development and to time assignments (grants) born out of a centralized vision of computing needs and strategies. In such model the user develops a rather passive attitude (with respect to the hardware and to the evolution of technology, networking and policy) and is forced to either adapt his/her application to the new computing paradigms or to adopt an already adapted one.

In the case of local computing facilities, instead, the users (which are either responsible or co-responsible for managing the available resources and related environments) can customize the platform and tune it to fit their needs and those of the services they wish to use/offer. In this case, however, hardware, technology, networking and policy evolution is much slower and the user group takes care of updating (usually rather partially) skills and computational environment (with the risk of underutilizing the computing resources and applications due to a lack of expertise).

2.2 The synergistic Grid model for the CMMST VRC

Although widely adopted both models are alone inadequate to meet the present requirements of the scientific communities. As an alternative a synergistic innovative model of allocating resources was developed within a series of European projects and collaborations (DATATAG (http://datatag.web.cern.ch/datatag/), EGEE-I-II-III (http://www.egee.eu), WLCG (http://wlcg.web.cern.ch/), etc.) mainly under the pressure of the research needs of High Energy Physics. Such synergistic model, often called High Throughput Computing (HTC) is based on the aggregation of a large number of geographically dispersed CPUs and users connected over the public network and is managed through the use of appropriate middleware and tools, has made it possible to execute distributed programs on a large number (over hundreds of thousands) of processors reaching very high performances for applications composed by decoupled or loosely coupled tasks.

ISSN: 2279-8773

After the first impulse given by the High Energy Physics community the model was extended to other disciplines. After all, for academic research groups (not belonging to a single Institution or tightly bound to the management of a shared experimental infrastructure) the synergistic model is the best approach to access the necessary volume of computing resources. Moreover, the synergistic model encompasses much more than the simple task of aggregating a large amount of computing resources and users for massive distributed computing. It can target, in fact, the more radical objective of changing the model of using and managing computing resources by allowing:

- an on demand allocation of the available computing time among registered users
- a synergy among different types of programs and expertise (including experiment and monitoring)
- a service oriented organization with a (Grid economy) stimulated proactive participation of the users
- a coupled distributed and parallel (HTC-HPC) computing methodology.

After all, within the EGI-Inspire European project, significant effort has been spent to make the Grid model undergo a significant increase in participating communities, networked resources, software and skills (otherwise confined in the local sites) availability as well as inter-disciplinary collaboration enhancement. Moreover, in order to meet the peculiarities of the different research areas and strengthen their role the users of the Grid have been gathered in VOs (http://www.egi.eu/community/vos/) which are now encouraged to gather together into VRCs (http://www.egi.eu/community/vrcs). For this reason the specific commitment of the COMPCHEM VO in the CMMST VRC project is to join its forces with those of other actors of the community (especially those of similar VOs like GAUSSIAN (http://accounting.egi.eu/vodis.php?vo=gaussian) and WeNMR (http://www.wenmr.eu/) and to encourage other VOs to join as well). In particular through DCC this opportunity will be offered to the researchers of the Division of the Computational Chemistry of EUCHEMS and to ECTNA. In order ~~to~~ better achieve this, quarterly teleconferences for planning and management will be held and one or two face-to-face meetings a year will be held during main Grid events like conferences or forums. Such meetings will also focus on technical problems of the VRC including operations, services, support to user, communication and dissemination.

In particular the Grid model to be implemented for CMMST should realize the dream of several computational scientists (also of other disciplines) of being able to carry out simulations of realistic systems and reproduce their observable properties starting from first principles (the so called virtual experiment). This is, indeed, what the Grid Empowered Molecular Simulator (GEMS) schema has been designed for [2]. In this way, by making use of appropriate grid tools (portals, workflows, frameworks, etc.) and blocks of applications and data, VRC scientists should be able to compose such simulations a la carte from components and applications made available to the members of the community by other members (or even from external providers) and run them by having flat access to the most suited computing resources (within the limits of availability).

This grounds also the VRC purpose of competing for acquiring, as a general policy, computing resources from resource providers for the community (to be added to the ones available to the VRC members from other sources). These resources will be used by the community for supporting fundamental activities (like basic research and algorithms development), sustaining collaborative projects, pursuing special innovative targets, etc.) and will be assigned to the members of the community in return for the gained credits.

2.3 The adoption of a credit system in the CMMST VRC

A key aspect of the synergistic nature of the proposed CMMST VRC is, in fact, the possibility of monitoring the activities of the members to the end of evaluating the work they do by them on behalf of the community and the services they are offered by the other members of the community. In other words all the activities and services offered to the community are be identified, recognised and rewarded.

The solid pillars on which such endeavour relies are

ISSN: 2279-8773

1. the commitment of the members of the community to offer the products of their work (new algorithms and applications, new validated data, etc.) as a service to the other members
2. the exploitation of the functionalities embodied in two tools being developed in COMCPHEM.

Such tools are aimed at evaluating the quality of both services (QoS) and users (QoU) by means of a framework performing active and passive filtering (GriF [3]) on Grid data and at awarding credits [4] based on the evaluation of monitored activities.

Because of this, the VRC can offer to its members the clear advantage of allowing them to carry out their computational campaigns (especially when the calculations are so complex to be unfeasible on the ground of only their own competences and their own limited platforms) in return for their commitment to carry out community related duties. Obviously, one can also choose to join the community at an entry level (passive user, first lower layer as discussed in Ref. [5]) member and either run the codes made available for demo purpose or his/her own code for personal use. The reason for having established this entry layer (usually meant to be temporary) is to offer an opportunity to the newcomers to check their real willingness to operate on a DCI platform. Already at this level, in fact, it becomes soon apparent that it is necessary to bear certain competences to exploit the advantages of the Grid. As a result, at this point one has choice of acquiring such competences and enter the second layer (active user) able to restructure the code so as to take advantage of distributing computing and usefully relate their codes to the ones of other users.

Obviously, one can still remain passive user after the expiration date of the trial period though it would have to contribute to the VRC sustainability using different means like for example being a paying user. On the contrary, the involvement reached by being active use may call the VRC member to an upper layer of involvement. As a matter of fact, opening the code implemented on the Grid to a shared usage by the other members of the VRC is the next membership level (passive software provider, third layer) that implies the validation of a stable version of the code and the assemblage of all the necessary (Grid User Interfaces) GUIs and the adoption of proper (even only de facto standard) data formats for use by other researchers possibly in conjunction with other codes. When software providers guarantee also additional basic services (like software upgrade, maintenance as well as user support (active software provider)) they become active software providers (fourth layer).

The commitment to confer to the Grid additional hardware (especially for those suites of codes which need special devices) after a negotiation with the Management Committee (MC) of the VRC about the relevance of such a commitment to the strategic choices is also welcome. This is, indeed, the fifth layer (passive resource provider). It applies also to hardware providers the possibility of providing basic services and user support. In this case they become active resource provider (sixth layer).

Obviously, the going through all these layer by conferring both software and hardware together with related services will take place gradually along a certain period of time (and not all the member of the VRC will be willing to do so). On the contrary some members or group of users or resource providers would like to take this as a mission (for example devote to development work on behalf of the VRC also unshared resources) and reach the status of “VRC stakeholder” (seventh layer).

The belonging to a given layer is periodically revised on the ground of credits acquired and multiple layer status for different products and services may coexist within a given member. For example, the status of stakeholder does not exclude that of paying customer or paid supplier for certain items. This becomes particularly appropriate when the VRC wins a bid or becomes a funded project holder.

Crucial to such dynamical geometry of the VRC is the adoption of “terms of exchange credits” (toecs or simply credits [6]) through which all activities of the members can be rewarded. The mechanism through which credits are assigned and redeemed are established and regulated by the governing body of the VRC. In general credits are assigned to reward the work that the VRC declares as useful to its sustainability and can be then redeemed through a privileged allocation of the resources of the VRC,

ISSN: 2279-8773

discounts in the use of commercial tools and/or other services. They may be even redeemed, in some clearly regulated cases, as cash to be used for research activities.

3 CONCLUSIONS

In the present document the advantages that a VRC status would offer to the members of the CMMST community in terms of access and use of the computing resources federated in EGI are listed. The document depicts the present scenario of computational resources and their usage by analysing the characteristics of the grant and the opportunistic models as opposed to the synergistic one proposed for the CMMST community. The document enumerates also the technical and non technical aspects of such cooperative model based on the selection of the resources (from personal systems to supercomputers) and services (from number crunching to massive data handling on heterogeneous platforms).

ACKNOWLEDGEMENTS

The activities presented in this document has been possible thanks to the Grid resources and services provided by the European Grid Infrastructure (EGI), the National Grid Initiatives the VOs and the Instituitions that support the Chemistry, Molecular & Materials Science and Technology computational community.

REFERENCES

[1] https://wiki.egi.eu/w/images/5/5f/VT_CMMST_Proposal_v1.pdf

[2] A. Costantini, O. Gervasi, C. Manuali, N. Faginas Lago, S. Rampino, A. Laganà, COMPCHEM: progress towards GEMS a Grid Empowered Molecular Simulator and beyond, Journal of Grid Computing, 8(4), 571-586 (2010)

[3] C. Manuali, S. Rampino, A. Lagan\`a, GRIF: A Grid Framework for a Web Service Approach to Reactive Scattering, Comp. Phys. Comm. 181, 1179-1185 (2010); C. Manuali, A. Laganà GRIF: A New Collaborative Framework for a Web Service Approach to Grid Empowered Calculations Future Generation of Computer Systems, 27(3), 315-318 (2011)

[4] C. Manuali, A. Lagan\`a, A Grid Credit System Empowering Virtual Research Communities Sustainability, Lecture Notes Computer Science, 6784, 397-411 (2011)

[5] A. Lagana', A. Riganelli, O. Gervasi, On the structuring of the computational chemistry virtual organization COMPCHEM, Lecture Notes in Computer Science (ISSN:0302-9743) 3980, 665-674 (2006).

[6] A. Laganà, S. Crocchianti, N. Faginas Lago, A. Riganelli, C. Manuali, S. Schanze, From Computer Assisted to Grid Empowered Teaching and Learning Activities in Higher Chemistry Education in Innovative Methods in Teaching and Learning Chemistry in Higher Education, I. Eilks and B. Byers Eds, RSC Publishing (2009) p. 153-190 ; ISBN 978-1-84755-958-6

ISSN: 2279-8773