**HPC/HTC Workflows for the Computational Chemistry Community**

**(XSEDE-EGI Collaborative Use Example)**

|  |  |
| --- | --- |
| Date: | **13/2/2013** |
| Document Status: | **V2.0** |
| Dissemination Level: | **Final** |
| Document Link: | **N.A.** |

1. **Overview**

In the recent years, following the requests from some EGI VOs, the need to bridge HTC and HPC infrastructures making them interoperable to run complex computational workflows has dramatically grown.

The Chemistry and Molecular and Materials Science and Technology (CMMST) community, at present operatively supported in EGI by the COMPCHEM VO [1], is developing a set of activities aimed at equipping their Computational Chemistry applications, largely based on the use of the Grid Empowered Molecular Simulator (GEMS) {2}, with a set of tools facilitating interoperability between HPC and HPC platforms. Unfortunately, the adoption of different middlewares on the different platforms made, as matter of facts, such interoperability (and the related activities) difficult to be finalized. Another crucial aspect limiting the interoperability between the two platforms is the lack of a unique point of ingress.

Starting from this interest, the community has developed prototypical workflows (hereafter called HTPC skeletons) grouped in two main categories:

* HTPC1, right hand side of Fig. 1: where a large quantity of independent tasks run on a HTC platform and their outcomes are each passed as input to a strongly coupled treatment implemented on a HPC platform
* HTPC2, see left hand side of Fig. 1: where a strongly coupled task runs on a HPC platform and its outcomes are passed as input to a large quantity of independent tasks running on a HTC platform

Within the EGI-XSEDE collaboration we would like to perform a feasibility study on running such HTPC skeletons by exploiting the two infrastructures in an automated way so as to minimize the human intervention and possibly develop prototypical tools able to handle the skeletons (in both the computing and data management parts). For this purpose we propose a set of use-cases that provide real life requirements to this activity and can be used to operatively proof the work performed.

[1] compchem website: www3.compchem.unipg.it/

[2] Laganà et al, J. of Grid Computing Vol. 8, No. 4, 571-586, (2010)



Fig. 1 – HTPC Skeletons: HTC computation following a HPC ones (left hand

Side HTPC1) and HPC computation following a HTC one (right hand side HTPC2).

1. **Collaborative teams**

To implement its activities the CMMST Community is collaborating with several actors both in Europe and US. On the European side the Community is interacting with the already mentioned COMPCHEM VO with the support of many NGIs (in particular the Italian NGI which is providing support and resources). On the US side the Community is interacting with the Chemical Dynamics group at the Department of Chemistry and Biochemistry (specialized in classical, semi-classical, and quantum mechanical methods applied to chemical reactions at an atomistic level) which have developed the VENUS package and the High Performance Computing Center both at Texas Tech University.

Appendix A lists the people involved in this collaboration.

1. **Proposed usecases**

The selected use-cases currently used by some members of the Community are presented in this section and a summary table (see Tab. 1) highlighting the application used and their computational requirements is given.

* 1. Exact quantum dynamics calculations for Li + FH (HTPC2)

Exact quantum dynamics calculations on the reaction Li + FH -> LiF + H require the nuclei wavefunction to be expanded in a suitable basis set parametrically dependent on a molecular geometry coordinate expressing the advance of the reactive process. Within the ABC program, firstly very large matrices are calculated (HPC section), then independent fixed E (scattering energy) and J (total angular momentum quantum number) propagations are distributed (HTC section).

* 1. Exact dynamics of nuclear and electonic systems within cavities (HTPC2)

The dynamics of closely coupled confined systems are going to be studied through the use of the MCTDH scheme, in particular the exact dynamics of electronic and/or nuclear systems within cavities. The high number of coupled degrees of freedom necessitates first HPC, while the possibility of running many jobs with different initial conditions points then towards the HTC requirement.

* 1. Ab initio calculations for N2 + N2 system (HTPC1)

The General Atomic and Molecular Electronic Structure System, is a general ab initio quantum chemistry package. Our work is focused on the calculation of a large number of electronic energy values for different nuclear geometries of the N2 + N2 system, using the CCSD(T) and MRPT2 methods implemented in GAMESS-US. In this scheme preliminary calculations of the electronic structure can be run on HTC nodes while the CCSDT(T) and the MRPT2 codes have to be run on HPC nodes. Ab initio electronic energy values will be used for building a global Potential Energy Surface to be used in quasi-classical or classical dynamics calculations.

* 1. Molecular Dynamics simulations of CTN (HTPC1)

Molecular Dynamics simulations using DLPOLY software to study the hydrogen storage in single and bundles walled carbon nanotubes (CNT) of various diameters and chiralities (whose determination can run on HTC nodes) run subsequently a force field dynamical calculation that takes into account all possible interaction between H-CNT and CNT-CNT ones and has HPC reuqirements. The considered use case is concerned with the force field for the H2 (fully describing the interaction with a three body potential) requires high CPU power and memory when investigating the storage of hydrogen in CNT.

* 1. Ab initio and classical MD simulations in micropourus materials (HTPC2)

This project aims at developing reliable force fields for classical Molecular Dynamics (MD) simulations in microporous materials. These force fields will be derived starting from high quality ab-initio computations (CP2K), requiring a large amount of resources available in HPC environment. The following classical MD simulations (NAMD), instead, are suitable for HTC environment: requiring less density of computational resources but a larger number of independent runs.

* 1. Interactions between doxorubicine and γ-cyclodextrin (HTPC1)

Doxorubicin is a quite powerful antitumoral agent, of the class of antracyclines. Selection of an appropriate carrier is extremely important to improve bio-availability of the drug and to reduce side effects, especially its cardiotoxicity. This research project is aimed to study interactions between doxorubicine and γ-cyclodextrin, which has been successfully used as a suitable carrier. Computational molecular modeling of these systems will be carried on by molecular mechanics and molecular dynamics methods, using the NAMD package. The four major doxorubicin dimers and the corresponding intermolecular interactions with γ-cyclodextrin in water solutions will be studied by molecular dynamics runs encompassing at least 100 ns of simulation time (HPC part). Estimates of the binding energies for the different arrangements will be obtained by the Potential of Mean Force (PMF), extracted from a series of umbrella sampling simulations (HTC part)

* 1. Semiclassical simulations of spectroscopy (HTPC2 inverse)

This project aims at reproducing a spectroscopy structure by running first a large number of trajectories (HTC section) whose outcomes are collected and used as input for the SC-IVR code that works out the spectrum using semiclassical techniques (HPC section) applying so far the HTPC2 scheme in an inverted order.

 Tab.1: Summary table highlighting the application used and their computational requirements for the presented case-studies.

|  |  |  |
| --- | --- | --- |
| Application | HPC requirements | HTC requirements |
| Cores /job | Memory(GB/core) | Total allocated time (h) | Cores/job | Memory(GB/core) | Total allocated time (h) |
| ABC (use case a) | 32 | 2 | 25K | 1 | 2 | 50K |
| MCTDH (use case b) | 1 | 5 | 50K | 5 | 1 | 40K |
| GAMESS-US (use case c) | 32 | 2 | 300K |  |  |  |
| DL\_POLY (use case d) | 64 | 2 | 100K | 8 | 1 | 500K |
| CP2K (use case e – HPC) | 512 | 2 | 512K |  |  |  |
| NAMD (use case e – HTC) |  |  |  | 8 | 1 | 40K |
| NAMD (use case f) | 256 | 1 | 50k | 32 | 1 | 60k |
| VENUS | 1 | 2 | 50K | 1 | 2 | 50K |
| SC-IVR (use case g) | 16 | 2 | 50 | 1 | 2 | 50 |

**Appendix A: People involved and affiliation**

**William L Hase**

**Alan Sill**

Dept. of Chemistry and Byochemistry

HPC Center

Texas Tech University

**Sergio Rampino**

 Institute of Molecular Science and Technologies of the Italian National Research Council, (CNR-ISTM) – Perugia Dept.

Italy

**Marco Sant**

**Andera Gabrieli**

University of Sassari

Dept. of Chemistry and pharmacy

Italy

**Noelia Faginas Lago**

**Leonardo Pacifici**

**Dimitrios SKOUTERIS**

University of Perugia

Dept. of Chemistry

Italy

**Stefano Ottani**

Institute of Organic Synthesis and Photoreactivity of the Italian National Research Council (CNR-ISOF) – Bologna Dept.

Italy

**Michele Ceotto**

University of Milano

Department of Chemistry

Italy

**Daniele Cesini**

**Alessandro Costantini**

INFN-IGI – Italian NGI user support team