MSO4SC

Mathematical Modelling, Simulation and Optimization for Societal Challenges with Scientific Computing

Using HPC and Cloud Resources for Running your Simulations

European Commission

Horizon 2020
European Union funding
for Research & Innovation

Executive Summary

Running applications, which perform complex mathematical simulations, is difficult and requires a lot of resources, also complex to use. This white paper analyzes some of the challenges related to these applications and describes the solution proposed by MSO4SC for executing them, based on a combination of HPC and Cloud resources to optimize running time, cost, efficiency and ease of usage. It describes the three main features that enable such combination: the orchestration, the monitoring and the interfaces. Finally, it provides information about the benefits of the proposed approach, based on the experimentation we have carried out.

Introduction

Mathematical modelling, simulation and optimization (MSO) algorithms and computational techniques are increasingly used in all fields of science, technology, industry and public service. Most of these techniques are based on tools and libraries which are very demanding in terms of computational resources. Therefore, they are typically executed in high performing hardware, although not all the operations involved may require it.

Applications based on complex mathematical frameworks usually carry out a process with four main kinds of tasks: i) data pre-processing, ii) simulation, iii) data post-processing and iv) visualization. Due to their complexity, parallelizing the tasks is important, and this is where High Performance Computing (HPC) and Cloud enter the scene.

These technologies have been evolving independently during years, one mainly focused on performance (HPC) and the other mainly focused on scalability (Cloud). Nevertheless, in order to run the MSO software efficiently, it is possible to get the best of each world by combining them in a smart way.

Challenges to be Faced

When we aim at running MSO software in HPC systems, there are several aspects that must be faced by the users: the complexity of the systems (that needs to be learnt) and the access to the resources (which are not always available).

In the first aspect, end users need to learn about the particularities of these systems, how to access them, how to compile for them, how to launch applications there, etc. The learning curve of these technologies uses to be steep.

With respect to the second aspect, when we want to run something in an HPC environment, it goes to a queue of tasks where it waits for its turn to run. Depending on the center load and the resources requested, it may take a long time until the application can run.

Moreover, it is not easy to get access to HPC resources, as the biggest centers are usually managed by public entities and are more dedicated to research. They are scarce resources and they are not especially cheap, in general.

In the case of HPC, the main challenge is to make easier the usage of the resources, while we reduce waiting time and we optimize the cost of a simulation.

On the other hand, Cloud resources are easily accessible (i.e. see Amazon EC2, Microsoft Azure, etc.) and they scale pretty well. Also, there is an important market and the cost of Cloud resources is getting lower and lower. The problem is that they do not provide good performance when running parallelized software, due to network virtualization and the way to manage the physical resources.

In this case, the main challenge is to make an adequate usage of Cloud resources for computing tasks with good enough performance, while we minimize the effects of moving data to the remote resources.

When looking at the whole picture, the main challenge to solve is how to run the MSO applications in such a way they will get the best from availability and scalability of Cloud systems, while they exploit the performance of HPC

systems and the cost of running the application is kept as low as possible.

What is MSO4SC Proposing?

MSO4SC is proposing a solution in which the tasks to be run in MSO applications are split between HPC and Cloud resources, considering aspects like the kind of tasks to run and the availability of the resources.

MSO4SC vision is a layered approach in which MSO applications and mathematical frameworks run on top of some computational services, which can be seen as other Cloud Platform as a Service (PaaS). Such PaaS provides access to computational resources, abstracting the complexity to the end users, who do not need to know about the underlying hardware and do not need to prepare the applications (compilation, deployment, etc.).

We use containers technology in the background to facilitate the combination of HPC and Cloud resources, since the adequate containers with the MSO applications are generated automatically. Such containers are deployed at the adequate location where the computational resources will be used.

Users just need to select the application to run, parametrize it and then launch it. The MSO4SC platform will select the appropriate resources providers (HPC + Cloud) and will run the tasks according to the application workflow defined by developers (pre-processing / simulation / post-processing / visualization). Once the process is finished, the resulting data will be ready for its visualization.

In the meantime, MSO4SC will be monitoring the usage of resources, their status, the application status and its internal metrics (if these are available). This information is available from the graphical user interface, so end users may access to information they are interested in.

The Combination of HPC and Cloud Resources

Orchestrating Applications

Those applications which integrate complex mathematical simulations can be split in several tasks that need to be executed. The orchestration is the process which manages the execution by identifying the tasks, assigning the adequate resources and, later on, running them. It is responsible to execute the different operations that compose a simulation, while optimizing the resources usage. MSO4SC is based on a well-known orchestrator called Cloudify, hiding the computing infrastructure complexities to users and applications, and automating its management.

Simulation developers can define their applications using blueprints (YAML files, similar to XML) that describe the multiple smaller tasks involved, their constrains and the dependencies between them. The resulting information is a directed graph that represents the simulation from a hardware agnostic perspective. As an added value, the developer can define loops (group of tasks that repeats for a static/dynamic number of iterations), and scalability (group of tasks that scale in parallel). In most cases, the application developer also provides the application binaries inside containers to maximize the computing infrastructures interoperability.

At the background, the orchestrator works with containers technology both in HPC and Cloud infrastructures, by means of Singularity and Docker, so it is easier to perform different deployments without the need to compile and deploy manually. The Continuous Integration and Continuous Deployment process allows just to provide the source code and obtain automatically the containers with the compiled components ready to be used.

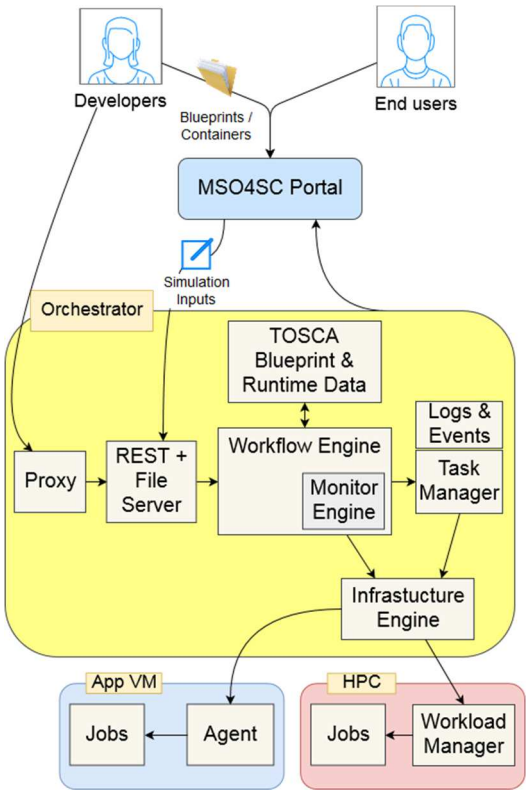

Figure 1 – Orchestration System

The orchestrator is provided with a set of inputs that configure the concrete simulation (e.g.: datasets to be used) when the user wants to start it. Then, it queries the application features and the infrastructures available, select the most suitable ones, and start the simulation. In this selection it is decided in which infrastructure it is going to be executed each task, looking for the best resources usage in terms of core/hour used and taking into consideration potential issues because of moving large datasets.

Monitoring all the Elements Involved

Due to the complexity and time to run the MSO applications, it is important that end users and resource providers

have access to information about the applications execution. Therefore, MSO4SC provides a mean to monitor the status and features of the available computing infrastructures, and the current performance of the ongoing simulations.

The monitoring feature keeps historic data about the infrastructure's general performance and load, as well as the specifics of each tasks being executed by the orchestrator. It is able to collect information from different sources, allowing MSO4SC to combine together HPC and Cloud metrics, such as availability, queues status, current load or time to run the applications. It is also able to retrieve information from the application logs, since important messages could be printed for end users capable of interpreting the results.

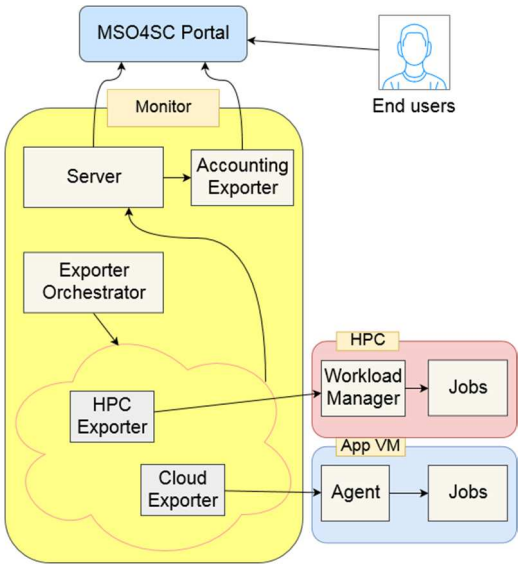

Figure 2 – Monitoring System

The system is extensible by means of new "exporters", that is, small programs that can send new metrics and information to the collector.

Finally, all the information can be accessed through a dedicated API (for other applications) and through a graphical user interface, where end users can filter the metrics they want to focus on.

Providing Friendly Interfaces

The MSO4SC interface provides a web application to present to the user the possibility to run simulations in a few clicks (the Click&Go feature), without knowing the specifics about the computing infrastructures, the MSO4SC platform, or the application itself.

The user just needs to define, for the infrastructures it has access to, its own credentials (e.g.: An HPC credentials, a Cloud provider credentials). After buying an application in the marketplace, it will be able to create a new "instance" of it, by defining the inputs that are prompt in the web page. Such inputs are determined by application developers, according to the 'blueprints' we mentioned, defining, in most of the cases, default values which reduce the parametrization complexity.

Then, just clicking on the "run" button, it is possible to access to the orchestration and simulation logs. When finished, the outputs will appear in the Data Catalogue, from which can be downloaded, and/or visualized in the "visualization tool" online.

For a more experienced user, an advanced data movement tool is also provided (based on Globus Connect) to manually move very large datasets from origin to the computing infrastructures, as fast as possible. Also, for such users, it is possible to have a more complete parametrization.

Where is the Benefit?

Reducing the Time to Run your Simulations

When running some tasks at Cloud resources, it is possible to request less HPC resources and, additionally, some pre-processing tasks can be executed while the simulation task is already waiting in the HPC queue. Requesting less resources usage led to reduce the waiting time in HPC queues and, moreover, as some pre-processing tasks are being performed in parallel to that waiting time, it is possible to reduce the total time of execution.

According to the validation done in the context of MSO4SC, tests have shown to reduce their execution time in about 23%, with respect to a full execution in HPC (reaching 48% when running only in Cloud).

Optimizing Resources

Providers of HPC resources need to be careful in the management due to the scarcity of such resources. When some tasks are run at Cloud resources, instead of HPC ones, we are releasing some HPC resources that can be used by other scientists or users. Since MSO4SC splits the application in smaller chunks, it generates smaller tasks which run during less time in HPC and the amount of resources requested is closer to the real usage (i.e. we do not have 60 nodes blocked several minutes for running some simple pre-processing tasks that could be handled with a few cores). As a result, time per core ratio decreases at the HPC center, achieving optimizations up to 17%.

On the other hand, due to the fact that each resource which is used is translated to costs for the end user, decreasing the amount of resources also decreases the cost for end users, especially if we take into account that HPC resources are, usually, more expensive than Cloud ones.

Abstracting the Usage of Computing Resources

Thanks to the graphical user interfaces available in MSO4SC, end users do not need to deal with complex systems, accessed by text consoles, which require a deep knowledge about concrete commands, customized compilation, etc. Application developers just need to define the application workflow with a high-level language and, once it is available, the platform handles all the complexity (through continuous integration and deployment). This saves a lot of time for application developers and users, who will not spend a lot of hours learning about these technologies.

End users will just need to enter the web interface, which will show them the parameters that are required for running their favorite MSO application. After filling them in and launching the application, they will be also able to follow the execution process and, once it is finished, they will have tools at hand for performing online visualization of the generated results. Running simulations and visualizing the results will take just a few minutes to end users, reducing drastically the time spent to configure, prepare and launch the applications.

Summary

MSO4SC is a new e-Infrastructure for running MSO applications and complex mathematical frameworks, in general, which implements a solution for combining HPC and Cloud resources, in an environment in which only HPC resources were used. Its main features are a new orchestration mechanism for applications execution management, an advanced monitoring solution merging data from HPC, Cloud and applications, and a simple web-based graphical user interface which makes easy to run MSO applications. Thanks to these features, end users can benefit with lower times to run, lower costs and lower times to prepare and run their simulations, while resources providers can manage more efficiently their resources.

Call to Action

We invite all those people interested to check out and test the MSO4SC e-Infrastructure by accessing our public portal. We are open to include new MSO applications and mathematical frameworks, increasing the MSO4SC offer for our stakeholders. We encourage you to contact us and get informed about how to proceed. And remember to follow our last updates on our web site, Twitter and ResearchGate!

Mathematical Modelling, Simulation and Optimization for Societal Challenges with Scientific Computing

www.mso4sc.eu

European Commission

Horizon 2020
European Union funding
for Research & Innovation