Accelerating the Design of Scientific Workflows with Simulation-Based Rapid Prototyping

Frédéric Suter
Oak Ridge National Laboratory, Oak Ridge, TN, USA
email: suterf@ornl.gov

***Abstract*—Extreme-scale science requires scientists to combine multiple heterogeneous computational tasks in complex workflows, efficiently managing large amounts of data, and fully exploiting the performance of the entire edge-to-HPC computing continuum. Going from science to workflows requires domain scientists to express their research ideas from a computer science perspective so that the most adapted and efficient tools and techniques can be selected. However, this is a long, error-prone, and tedious process, as there is no "turnkey solution" in such a diverse ecosystem. Therefore, we propose to develop a comprehensive simulation-based framework that will be used by domain scientists to easily prototype their scientific workflows, while expressing all the important information needed by computer scientists to provide them with the most efficient implementation.**

I. Introduction

The path from a scientific idea that requires the orchestrated execution of multiple software pieces on different data sets to its actual implementation as an actionable workflow running on the resources of leadership computing facilities is a complex workflow itself which requires the interaction of multiple actors with different backgrounds, knowledge, vocabulary, preferences, and constraints.

It begins with domain scientists who express their needs as a combination of computational tasks and data exchanges, based on their scientific problem and their experience of running the different components at small scale. This abstract view of the scientific workflow is then enriched to better specify the precise computing needs, data movements, and interactions between tasks. It may also integrate information about the individual performance and scalability of the different components. This step usually requires multiple exchanges between domain and computer scientists which may lead to imperfect translations of the expression of needs, because of different vocabularies for instance, or to incomprehension, e.g., about the importance of a specific constraint.

Then, particular workflow and data management systems have to be selected among a plethora of existing tools [1] to implement this enriched view of the workflow. This selection step usually requires some adaptations, as no "turnkey solution" does exist, and may suffer from the influence of some biases or preferences. More importantly, selecting tools leads to a *technological lock*. However, it precedes the necessary testing and optimization phase of the workflow implementation. If performance at large-scale is unsatisfactory, it can be very difficult, if even possible, and very time-consuming for all the actors to operate a *technological change*. In most of the cases, scientific results can nevertheless be obtained, but at an important human cost and without exploiting the computational resources at their full potential.

To reduce the effects of these identified drawbacks, we propose to design, implement, and test a *simulation-based rapid prototyping tool*. The main advantage of resorting to simulation is the capacity to explore multiple scenarios for the implementation of a scientific workflow, be they existing or exploratory, in a controlled environment, a reasonable amount of time, and without the need for a deployment at scale. This will allow scientists to report the technological lock after the testing and optimization phase of a simulated version of their workflow, and thus to take more informed decisions. Moreover, simulating the most prominent features of data and workflow management systems rather than specific implementations increases the capacity to easily switch from one feature to another in the evaluation process. Then, it reduces the impact of biases and preferences and limit the need for adaptation to that of a well defined version of the workflow. Finally, having domain and computer scientists work together on the writing of a simulator of the abstract workflow reduces the risk of incomprehension or caveats in the design of the actionable workflow. Scientist needs can be implemented and tested in a quick interaction cycle to converge faster on a full prototype implementation of the workflow.

This work builds on the SimGrid toolkit [2] which enables the simulation of large-scale distributed applications in a way that is accurate (via validated simulation models), scalable (ability to run large scale simulations on a single computer with low compute, memory, and energy footprints), and expressive (ability to simulate arbitrary platform, application, and execution scenarios). It also leverages the WRENCH project [3], which builds on SimGrid, to implement simulated core services commonly used by production workflow management systems. WRENCH makes it possible to implement simulators of complex workflow management scenarios in only a few hundred lines of code.

This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid up, irrevocable, world-wide license to publish or reproduce the published form of the manuscript, or allow others to do so, for U.S. Government purposes. The DOE will provide public access to these results in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

In this poster presentation, we will present how we address the different challenges related to the development of this simulation-based rapid prototyping tool, follwing a bottom-up approach by: (i) producing faithful descriptions of the resources of leadership class supercomputers; (ii) developing simulated core services for efficient data management, and (iii) extracting and enabling the simulation of the prominent features of workflow management systems.

II. Leadership Class Supercomputers and Beyond

To obtain sound and objective performance indicators for the execution of scientific workflows, we need to feed the simulation framework with faithful descriptions of large and complex computing and storage infrastructures. The new programmatic description interface of SimGrid eases the description of such infrastructures. It consists in describing the computing resources (i.e., number and characteristics of compute nodes), network interconnect (i.e., topology and nominal bandwidth and latency of the network links), and storage resources (i.e., number and types of disks and file system characteristics), and use this information to instantiate the underlying resource simulation models [4]. For instance, it is possible to describe the Summit pre-exascale machine in less than a hundred lines of code. Such a description is optimistic, as if an application could run alone and reach peak performance, but can be used to get coarse preliminary insights into the performance of a workflow execution. The accuracy of the simulation-based performance assessment can then progressively be improved by augmenting the initial descriptions with information extracted from benchmarks or traces (e.g., network switch saturation, compute kernel affinity, parallel file system throughput).

We also plan to go beyond the case of a single monolithic supercomputer anticipate the future evolution of leadership class computing facilities towards a more flexible ecosystem along the Edge-to-HPC computing continuum.

III. Versatile Data Transport Layer

Many scientific workflow are data-intensive and move large amounts of data from one computational task to another [5], which can easily become a performance bottleneck on leadership class supercomputers and over the edge-to-HPC continuum if not handled properly. However, this aspect of scientific workflows related to data management is often underestimated or even disregarded by domain scientists, despite its importance in achieving performance at scale. Allowing scientists to easily measure the effect of parallel I/O on the overall performance of their workflow is thus paramount. Our simulation framework thus includes a simulated *data transport layer* core service. This versatile service is inspired by the ADIOS high-performance I/O framework [6] that exposes simple concepts to its users and hides from them all the complex techniques that are needed to achieve the best performance at scale. It relies on the same high-level concepts (i.e., self-describing data and publish-subscribe paradigm) and allows users to seamlessly switch from file-based I/O to data streaming.

IV. Workflow Management System Features

To design and implement their scientific workflows, domain scientists are faced with a plethora of workflow management systems with different levels of maturity, specific features and/or technical constraints, and many configuration parameters [1]. The performance assessment of different such systems is a time- and resource-consuming task. As the consequence, scientists tend to either select tools that they have been told were interesting, but may not be adapted to their needs; spend a significant amount of time to evaluate candidate tools, which delay the actual execution of the workflows; or design yet another tool that suits their needs, but requires efforts and may fell in many avoidable pitfalls in the process.

Our approach is to focus on prominent features of actively-developed workflow systems rather than simulating a specific tool. The rationale is that using high level concepts will facilitate the interactions with domain scientists and simplify the design of a prototype simulator corresponding to their needs. Indeed, the objective is to be able to translate simple questions such as: “Is this workflow component a set of high throughput sequential tasks or a tightly-coupled MPI parallel job?”, “Do you store the results in a database that you later query or as files in a specific directory?”, or “Does this numerical simulation need to be coupled to in-situ analyses or visualization?” into corresponding simulated building blocks that can be combined at will. When the answers to such questions are not definitive, scientists will obtain preliminary objective performance results, allowing them to refine the expression of their needs and make more educated decisions to select a specific workflow management system and decide on how to configure it.

Acknowledgements

This research is partially supported by Laboratory Directed Research and Development Strategic Hire funding No. 11134 from Oak Ridge National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy.

References

[1] P. Amstutz, M. Mikheev, M. R. Crusoe, N. Tijanić, S. Lampa *et al.*, “Existing Workflow systems,” [Online] https://s.apache.org/existing-workflow-systems, updated 2022-09-13, accessed 2023-01-23.

[2] H. Casanova, A. Giersch, A. Legrand, M. Quinson, and F. Suter, “Versatile, Scalable, and Accurate Simulation of Distributed Applications and Platforms,” *JPDC*, vol. 74, no. 10, pp. 2899 – 2917, 2014.

[3] H. Casanova, R. Ferreira da Silva, R. Tanaka, S. Pandey, G. Jethwani, S. Albrecht, J. Oeth, and F. Suter, “Developing Accurate and Scalable Simulators of Production Workflow Management Systems with WRENCH,” *FGCS*, vol. 112, pp. 162–175, 2020.

[4] A. Degomme, A. Legrand, G. Markomanolis, M. Quinson, M. Stillwell, and F. Suter, “Simulating MPI applications: the SMPI approach,” *IEEE TPDS*, vol. 18, no. 8, pp. 2387–2400, 2017.

[5] J. Liu, E. Pacitti, P. Valduriez, and M. Mattoso, “A Survey of Data-Intensive Scientific Workflow Management,” *Journal of Grid Computing*, vol. 13, no. 4, pp. 457–493, 2015.

[6] W. F. Godoy, N. Podhorszki, R. Wang, C. Atkins, G. Eisenhauer, J. Gu, P. Davis, J. Choi, K. Germaschewski, K. Huck, A. Huebl, M. Kim, J. Kress, T. Kurc, Q. Liu, J. Logan, K. Mehta, G. Ostrouchov, M. Parashar, F. Poeschel, D. Pugmire, E. Suchyta, K. Takahashi, N. Thompson, S. Tsutsumi, L. Wan, M. Wolf, K. Wu, and S. Klasky, “ADIOS 2: The Adaptable Input Output System. A framework for high-performance data management,” *SoftwareX*, vol. 12, p. 100561, 2020.