Molecular Dynamics with HyperFlow and Scalarm on the PaaSage Platform

Maciej Malawski[1(✉)], Bartosz Balis[1], Kamil Figiela[1], Maciej Pawlik[1], Marian Bubak[1], Dariusz Król[1], Renata Słota[1], Michał Orzechowski[1], Jacek Kitowski[1], and Dennis Hoppe[2]

[1] Department of Computer Science, Faculty of Computer Science, Electronics, and Telecommunications, AGH University of Science and Technology, Mickiewicza 30, 30-059 Krakow, Poland
{malawski,dkrol}@agh.edu.pl
[2] High Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Nobelstr. 19, 70569 Stuttgart, Germany
dennis.hoppe@hlrs.de

This paper demonstrates how scientific workflow applications executed by HyperFlow engine [1] and data farming experiments managed by the Scalarm platform [2] can benefit from the capabilities offered by PaaSage platform[1], on the example of molecular dynamics simulation. While HyperFlow engine enables users to execute tasks of scientific workflows and Scalarm supports parameter studies oriented on data farming on available computing resources (e.g. virtual machines in a cloud), the role of PaaSage platform is to provision these resources, deploy an application, and automatically scale them according to the application demands. PaaSage uses a model-based approach, which means that the cloud application together with its requirements needs to be described using Cloud Application Modeling and Execution Language (CAMEL) [4].

The main new capabilities resulting from combining HyperFlow and Scalarm with PaaSage are to deploy the entire runtime environment as part of the application. Consequently, the whole platform such as HyperFlow or Scalarm, supporting a class of applications, can be seen as a single cloud application from the perspective of PaaSage. In particular, the workflow enactment engine, HyperFlow [1], acts as a component of the PaaSage application. Similarly, in the case of the Scalarm platform [2], we modelled the whole platform as an application, which is capable to execute different simulation models and scale itself in a cross-cloud environment. Consequently, we avoid tight coupling to a particular cloud infrastructure and middleware in both cases.

According to the PaaSage methodology, we modelled both HyperFlow and Scalarm using CAMEL. HyperFlow deployment model includes the master node which contains the workflow engine components, while the worker (which is scalable from 1 to n nodes) includes the executor together with application dependencies such as MPI or POV-Ray. Additionally, we deploy Redis and RabbitMQ components, as well as InfluxDB and Grafana as monitoring dashboard. Scalarm's architecture also follows the master-worker design pattern, where the

[1] http://www.paasage.eu.

A. Lazovik and S. Schulte (Eds.): ESOCC 2016 Workshops, CCIS 707, pp. 299–300, 2018.
https://doi.org/10.1007/978-3-319-72125-5

master part consists of loosely coupled services and is responsible for coordinating experiments, while the worker part handles actual computations. In addition, Scalarm can be run on different clouds as described in [3], hence its actual runtime deployment can be divided vertically and horizontally, where the amount of resources for each part can be dynamically adjusted to the actual load.

We demonstrate the benefits of using HyperFlow and Scalarm with PaaSage by deploying a molecular dynamics (MD) simulation. MD simulations are highly representive for e-science applications, because they comprise resource– and compute–intensive calculations that are massively parallelizable via MPI or OpenMP. These kinds of simulations provide information about how a given substance behaves under a given set of physical conditions including temperature and pressure. It allows us to predict material behaviour, for example for industrial purposes, such as the crack distribution across bridges. Simulations usually extend over multiple iterations starting with a coarse simulation over selected data points, and ending in a fine-granular simulation around a point of interest such as the origin of a crack. That's why accurate simulations require to be executed up to several hundred times with different parameter sets to yield accurate results. Our simulation includes a post-processing step that generates a video out of the raw data received from the simulation.

HPC is usually first-choice when it comes to executing e-science applications. However, the trend is towards a hybrid HPC/Cloud model, where high-performance resources are combined with the advantages of the cloud: flexibility, high availability, and disaster recovery to name but a few. Applications that will benefit from such a hybrid model are, in particular, e-science applications that usually include pre- and post-processing steps such as generating a video, which are not compute intensive. Those tasks can then be moved into the cloud, whereas compute-intensive tasks will continue to run on HPC infrastructure. PaaSage enables us to model our e-science applications once using CAMEL, and then deploy individual tasks, managed by HyperFlow or Scalarm, on different HPC/cloud infrastructures with ease.

Acknowledgments. We thankfully acknowledge the support of the EU FP7-ICT project PaaSage (317715), Polish grant 3033/7PR/2014/2 and AGH grant 11.11.230.124. Access to EC2 Cloud was provided by AWS in Education Grant.

References

1. Balis, B.: HyperFlow: a model of computation, programming approach and enactment engine for complex distributed workows. Future Gener. Comput. Syst. **55**, 147–162 (2016)
2. Król, D., Kitowski, J.: Self-scalable services in service oriented software for cost-effective data farming. Future Gener. Comput. Syst. **54**, 1–15 (2016)
3. Krol, D., Slota, R., Kitowski, J., Dutka, L., Liput, J.: Data farming on heterogeneous clouds. In: 2014 IEEE 7th International Conference on Cloud Computing, pp. 873–880. IEEE, June 2014
4. Nikolov, N., Rossini, A., Kritikos, K.: Integration of DSLs and migration of models: a case study in the cloud computing domain. Procedia Comput. Sci. **68**, 53–66 (2015)