Grid Interoperability for e-Research

Morris Riedel · Gabor Terstyanszky

Published online: 25 September 2009
© Springer Science + Business Media B.V. 2009

Computational simulations and thus scientific computing is the third pillar alongside theory and experiment in science and engineering today. The term e-science evolved as a new research field that focus on collaboration in key areas of science using next generation computing infrastructures such as Grids to extend the potential of scientific computing. More recently, increasing complexity of e-science applications that embrace multiple physical models (i.e. multi-physics) and consider a larger range of scales (i.e. multi-scale) is creating a steadily growing demand for world-wide interoperable Grid infrastructures that allow for new innovative types of e-science by jointly using a broader variety of computational resources. Since such interoperable Grid infrastructures are still not seamlessly provided today, the topic ‘Grid interoperability’ emerged as a broader research field in the last couple of years.

The lack of Grid interoperability is a hindrance since we observe a growing interest in the coordinated use of more than one Grid with a single

client that controls interoperable components deployed in different Grid infrastructures. In fact, we have shown in a recent classification [1] that among simple scripts with limited control functionality (i.e. loops), scientific application client plug-ins, complex workflows, and interactive access, there is also Grid interoperability mentioned as one approach to perform e-science today. Such interoperable federated Grids have the potential to facilitate e-research and thus scientific advances, which would not be possible using only a single Grid infrastructure. These advances arise from the advantages that federated Grid resources provide, such as access to a wide variety of heterogeneous resources, aggregated higher throughput, and lower time-to-solution.

In more detail, we observe that more and more Grid end-users raise the demand to access both High Throughput Computing (HTC)-driven Grids (EGEE, OSG, etc.) and High Performance Computing (HPC)-driven infrastructures (DEISA, TeraGrid, etc.) from a single client or science portal. In this context, the fundamental difference between HPC and HTC is that HPC resources (i.e. supercomputers) provide a good interconnection of cpus/cores while HTC resources (i.e. pc pools) do not. This joint use is typically motivated by the theory and concept that tackle the scientific problem and is modeled within the corresponding codes that in turn lead to some codes that are ‘nicely parallel’ (i.e. HTC) and others

M. Riedel (✉)
Institute for Advanced Simulation,
Forschungszentrum Jülich, Jülich, Germany
e-mail: m.riedel@fz-juelich.de

G. Terstyanszky
University of Westminster, London, UK
e-mail: terstyg@westminster.ac.uk

that are better suited to be computed as ‘massively parallel’ (i.e. HPC) simulations. In addition, the joint use of HTC- and HPC-Grids is often motivated by the fact that often end-users perform smaller evaluation runs with their codes on HTC resources before performing full-blown production runs on large-scale HPC resources. This saves rare computational time on costly HPC resources within HPC-driven Grids.

At the time of writing, it is an interesting time period for European Grids in the sense of the upcoming transition process from the project-based EGEE project to a longer sustainable European Grid Initiative (EGI) while DEISA and the Partnership for Advanced Computing in Europe (PRACE) are jointly creating an HPC infrastructure for emerging peta-scale applications. In the US, we see an upcoming third phase of the TeraGrid in the context of the extreme digital (XD) resources for science and engineering transition. Nevertheless, what we learned from the past and what we can expect from the future is that the underlying computing paradigms will remain in the sense that requirements for HTC and HPC will be still present. That’s still valid even in times where, in principle, HTC and HPC codes could be executed on one large-scale cluster such as the IBM BlueGene/P while having thus much more focus on the computed data itself instead on the computational paradigms that are being used.

Since the difference between these underlying computational paradigms (i.e. HTC and HPC) will still exist in the future, interoperability between Grid infrastructures that offer seamless access to both types of computational resources will be further needed in future. Since, we observe a rather slow adoption of emerging open standards in deployed Grid middleware systems on these infrastructures in the past; the Grid communities developed many different approaches to Grid interoperability that are classified by Riedel et al. in [3]. Nevertheless, common open standards are

the one and only way to enable a long-term seamless cross-Grid access and thus we have worked on the understanding of how such standards can be further improved to increase their adoption in production Grid middleware. In fact, we have performed many interoperability tests and worked with a wide variety of interoperability use cases [2] in the context of the Grid Interoperation Now (GIN) community group of OGF. The lessons learned from all these activities have been given as an input to the OGF Production Grid Infrastructure (PGI) working group in order to improve existing emerging open standards towards an improved production usage following a well defined reference model [3].

This journal special issue highlights selected contributions to the greater research field of Grid interoperability in general and provides an interesting set of information about world-wide projects that work in this particular research field. It thus represents a good supplement to the proceedings of the International Grid Interoperability and Interoperation Workshops (IGIIW) that we have organized in the past.

References

1. Riedel, M., Streit, A., Wolf, F., Lippert, T.H., Kranzlmüller, D.: Classification of different approaches for e-science applications in next generation computing infrastructures. In: Proceedings of the 4th IEEE Conference on e-Science (e-Science), pp. 198–205, Indianapolis, Indiana, USA (2008)
2. Riedel, M., Laure, E., et al.: Interoperation of world-wide production e-science infrastructures. *Journal of Concurrency and Computation: Practice and Experience* **21**, 961–990 (2009)
3. Riedel, M., Wolf, F., Kranzlmüller, D., Streit, A., Lippert, T.: Research advances by using interoperable e-science infrastructures—the infrastructure interoperability reference model applied in e-science. *Journal of Cluster Computing, Special Issue Recent Advances in e-Science*. doi:[10.1002/cpe.1402](https://doi.org/10.1002/cpe.1402)