Preface

**Eric Aubanel · Virendra C. Bhavsar ·
Michael Alex Frumkin**

Received: 3 August 2012 / Accepted: 4 August 2012 / Published online: 31 August 2012
© Springer Science+Business Media B.V. 2012

This special issue was inspired by the many contributions to the High Performance Grid and Cloud Computing workshop, which has been held annually at the International Parallel and Distributed Processing Symposium (IPDPS) since 2004. Our workshop initially only covered Grid computing, but has embraced cloud computing starting in 2011. Topics have ranged from Grid middleware, management and use of Grid resources, scheduling, and fault tolerance of distributed resources. Past participants of the workshop were invited to submit original contributions for this issue, but the invitation was extended to the entire community. It has turned out that the contributions address many of the same topics we have seen in the workshop.

We have selected the best ten papers out of many submissions received. We have categorized the papers in this issue into three broad groups: resource management, scheduling, and fault tolerant storage. These topics have been ongoing concerns for years because the rapidly evolving

distributed computing architectures keep generating new challenges. Many massive data centers are coming online, giving users an opportunity to access resources previously only available to national laboratories. Each improvement in compute, storage and network resources can have a huge impact on the productivity of high performance computing clients.

There are many ongoing issues in Grid and cloud management. In “Riding out the storm: How to deal with the complexity of Grid and cloud management,” J. Montes, A. Sanchez, and M.S. Perez outline a theoretical framework for the management of a Grid or a cloud infrastructure as single entity. This framework can be used for automatic management of such infrastructures. The Scientific Workflow Automation and Management Platform (SWAMP) is proposed by Q. Wu, M. Zhu, T. Gu, P. Brown, X. Lu, W. Lin and Y. Liu, in “A Distributed Workflow Management System with Case Study of Real-life Scientific Applications on Grids.” The SWAMP is designed for management of real-life scientific complex workflows in heterogeneous network environments. The SWAMP also includes efficient workflow mapping schemes for achieving superior performance. E. Yildirim and T. Kosar propose “End-to-end Data-flow Parallelism for Throughput Optimization in High-speed Networks.” They optimize parallel data streams over a network of disks and CPUs resulting in a

E. Aubanel (✉) · V. C. Bhavsar
University of New Brunswick,
Fredericton, NB, Canada
e-mail: aubanel@unb.ca

M. A. Frumkin
NVIDIA, Santa Clara, CA, USA

high increase in throughput. In “Using Broadcast Networks to Create On-demand Extremely Large Scale High throughput Computing Infrastructure” B. Costa, F. Brasileiro, G.L.Filho and D. Sousa present the On-Demand Distributed Computing Infrastructure (OddCI) architecture. Besides standard computing systems, OddCI can include mobile phones and digital television receivers. In “Energy-efficient Thermal-aware Autonomic Management of Virtualized HPC Cloud Infrastructure” I. Rodero, H. Viswanathan, E.K. Lee, M. Gamell, D. Pompili and M. Parashar present a reactive cross-layer thermal management scheme to avoid hotspots in cloud infrastructures. They also propose a virtual machine allocation strategy for significantly improving energy efficiency.

The first two papers in the scheduling category approach Grid scheduling from two angles: rescheduling and game-theoretic analysis. In “On the Improvement of Grid Resource Utilization: Preventive and Reactive Rescheduling Approaches”, Luis Tomas et al. address the fragmentation that results from scheduling Grid jobs well before they are executed. They employ a combination of reactive and preventive algorithms, and validate their approach on a Grid test-bed. Scheduling HPC workflows in clouds is complicated by the need to consider the dynamic nature

and cost of resources. Maria Grazia Buscemi, Ugo Montanari and Sonia Taneja analyze the impact of selfish and cooperative organizations in “A Game-Theoretic Analysis of Grid Job Scheduling”. Their model allows them to reason about different job scheduling mechanisms. The next two papers address cloud scheduling. Daniel de Oliveira propose in “A Provenance-based Adaptive Scheduling Heuristic for Parallel Scientific Workflows in Clouds” an algorithm with a cost model based on the objectives of makespan, reliability and financial cost, and validate it with a bioinformatics workflow on Amazon’s EC2. Oliver Niehorster et al. in “Cost-Aware and SLO-Fulfilling Software as a Service” have designed a self-optimizing multi-agent system to reduce human intervention in cloud management. Agents adapt resources during runtime using cost-aware scheduling to ensure quality of service.

Finally, Grids and clouds use redundancy to provide fault-tolerant storage. Zhen Huang, Yisong Lin and Yuxing Peng in “Robust redundancy scheme for the repair process: Hierarchical Codes in the bandwidth-limited systems” address the problem of reintegrating nodes which fail and are repaired, which can result in considerable network traffic. They find that hierarchical codes, that extend erasure codes, perform the best when dealing with all kinds of failures.