Guest Editors Introduction: Special Issue on Scientific Cloud Computing

Kate Keahey, Ioan Raicu, Kyle Chard, and Bogdan Nicolae

COMPUTATIONAL and Data-Driven Sciences have become the third and fourth pillar of scientific discovery in addition to experimental and theoretical sciences. Scientific Computing has transformed scientific discovery, enabling scientific breakthroughs through new kinds of experiments and simulations that would have been impossible only a decade ago. It is the key to solving grand challenges in many domains and providing breakthroughs in new knowledge by combining the lessons and approaches from multiple large-scale distributed computing areas, including high performance computing (HPC), high throughput computing (HTC), many-task computing, and data-intensive computing. Todays Big Data problems are generating datasets that are increasing exponentially in both complexity and volume, making their analysis, archival, and sharing one of the grand challenges of the 21st century.

Cloud computing has shown a great deal of promise as a scalable and cost-effective computing model for supporting scientific applications. Indeed, over the last several years adoption has been swift. Cloud computing offers elastic computing capacity, virtualized resources, and pay-as-you-go billing models, these capabilities enable scientists to outsource analyses, scale to large problem sizes, and to do so by paying only for the resources they used–rather than requiring large upfront investments. However, there are many inherent challenges on how to adapt the mixed techniques used by modern scientific computing to make best use of cloud computing infrastructures and vice-versa.

This journal Special Issue on Scientific Cloud Computing in the IEEE Transaction on Cloud Computing provides an opportune forum for presenting new research, development, and deployment efforts to address conducting scientific analyses on Cloud Computing infrastructures. This is a timely special issue, as we are seeing rapid growth of scientific computing using clouds, and in many cases, a lack of technological advancements are limiting the efficiency of executing these applications on the cloud. The importance of this area is reflected by the strong participation in this special issue: receiving 41 submissions of which 8 have been selected.

The selected papers contribute important advances towards leveraging clouds for scientific applications. The contributions focus on a broad range of topics, including: performance modeling and optimization, data management, resource allocation and scheduling, elasticity, reconfiguration, cost prediction and optimization. Most papers revolve around general techniques and approaches that are agnostic of the applications, while two contributions demonstrate how domain-specific scientific applications can be migrated to the cloud.

While clouds provide access to enormous on-demand computing resources, challenges arise when attempting to scale analyses automatically and efficiently, both from a performance and a cost perspective. Righi et al. propose AutoElastic, a transparent, Platform-as-a-Service level approach for elastically scaling HPC applications on clouds. This approach enables applications to scale without requiring user intervention or source code modification. J. Chen et al. present a complementary system, Ensembe, that is able to construct performance models for applications running in clouds. The resulting performance models can be used by systems like AutoElastic to optimizing provisioning and allocation.

Scientific workflows are one of the most common methods for running scientific applications on clouds. As such, several of the papers in this special issue address challenges faced by workflow systems. Zhou et al. investigate the economic landscape of running scientific workflows on clouds and present a scheduling system and associated cost optimizations that are able to minimize expected cost given user-specified probabilistic deadlines. In response to the challenges of running workflows in the presence of failure. W. Chen et al. present a theoretical analysis of the impact of failures on the runtime performance of scientific workflows. They apply a general task failure modeling approach to estimate performance under failure and present three fault-tolerant task clustering strategies and a dynamic clustering strategy to improve performance and adjust the granularity of clusters based on failure rate.

New data management and transfer approaches are needed to move data efficiently and reliable to and from the cloud and between cloud instances. In this space, Yildirim et al. focus on optimizing large data transfers composed of heterogenous file sizes in heterogeneous environments. Tudoran et. al propose OverFlow, a data management system that runs across geographically distributed sites and empowers large-scale scientific applications with a set of tools to monitor and perform low level manipulations of data (e.g. compression, de-duplication, geo-replication) that enable achieving a desired performance-cost trade-off.

Finally, two articles address the challenges of migrating scientific applications to the Cloud. First, Frattini et al. study migration from multiple perspectives: performance, resource

K. Keahey was with the Mathematics and Computer Science Division, Argonne National Laboratory e-mail: keahey@mcs.anl.gov.

I. Raicu was with the Department of Computer Science, Illinois Institute of Technology e-mail: iraicu@cs.iit.edu

K. Chard was with the Computation Institute, University of Chicago and Argonne National Laboratory e-mail: chard@uchicago.edu.

B. Nicolae was with IBM Research, Ireland e-mail: bogdan.nicolae@acm.org.

utilization, and dependability aspects (i.e. resilience, availability). They propose a methodology to fine-tune the virtual machine configuration, consolidation strategy and fault tolerance strategies in order to meet the functional and nonfunctional requirements. Zinno et al. present a case study for migrating a traditional HPC earth surface deformation analysis to commercial clouds. They demonstrate that the cloud-based solution operates at reduced cost and improved performance.

The associate editors would like to thank all the researchers who responded to this call and submitted manuscripts for consideration. This special issue would not have been possible without their efforts. At the same time, they would like to acknowledge the efforts of all reviewers whose detailed reviews and suggestions were also instrumental in the publication of this special issue. They also thank the inaugural Editor-in-Chief Professor Rajkumar Buyya and current Editor-in-Chief Professor Irena Bojanova, for supporting this topic as a special issue. Finally, they thank the TCC administrator, Ms. Joyce Arnold, for her support in publishing this special issue in a timely fashion.

Bogdan Nicolae is a research scientist within the High Performance Systems group at IBM Ireland. He specializes in scalable storage and fault tolerance for large scale distributed systems, with a focus on cloud computing and high performance architectures. He holds a PhD from University of Rennes 1, France (2010) and a Dipl. Eng. degree from Politehnica University Bucharest, Romania (2007). He is interested by and authored numerous papers in the areas of scalable I/O, storage elasticity and virtualization, data and metadata decentralization and availability, multi-versioning, checkpoint-restart, live migration. He has developed and directed several R&D projects, including the BlobSeer large scale versioning storage system. He participates in the editing of several journals (e.g., IEEE TCC), the organization / program committees of several international conferences (e.g., IPDPS, HPDC, PPoPP, CCGrid, CLUSTER, CLOUD) and acts as an expert for broader initiatives (e.g. European Exascale Software Initiative) around the topics mentioned above.

Kate Keahey is one of the pioneers of infrastructure cloud computing. She created and leads the development of the Nimbus project, recognized as the first open source Infrastructure-as-a-Service implementation, and engages in many application projects popularizing the use of the cloud computing platform in science. She also leads the Chameleon project, a distributed experimental platform for cloud computing research. Kate is a Scientist at Argonne National Laboratory and a Senior Fellow at the Computation Institute at the University of Chicago.

Ioan Raicu is an assistant professor in the Department of Computer Science at Illinois Institute of Technology (IIT), as well as a guest research faculty in the Math and Computer Science Division at Argonne National Laboratory. He is also the founder (2011) and director of the Data-Intensive Distributed Systems Laboratory at IIT. He has received the prestigious NSF CAREER award (2011 - 2015) for his innovative work on distributed file systems for extreme-scales. He was a NSF/CRA Computation Innovation Fellow at Northwestern University in 2009–2010, and obtained his Ph.D. in Computer Science from University of Chicago under the guidance of Dr. Ian Foster in 2009. He is a 3-year award winner of the GSRP Fellowship from NASA Ames Research Center. His research work and interests are in the general area of distributed systems, with particular interests in resource management in large scale distributed systems with a focus on many-task computing, data intensive computing, cloud computing, grid computing, and many-core computing. He has founded and chaired several workshops, such as IEEE/ACM MTAGS, IEEE/ACM DataCloud, ACM ScienceCloud, and IEEE CASK. He is on the editorial board of IEEE TCC, Springer JoCCASA, and Springer Cluster. He has been leadership roles in several high profile conferences, such as HPDC, CCGrid, Grid, eScience, Cluster, ICAC, and BDC.

Kyle Chard is a Senior Researcher and Fellow in the Computation Institute at the University of Chicago and Argonne National Laboratory. He received his Ph.D. in Computer Science from Victoria University of Wellington. He also holds a BSc (Hons) in Computer Science from the same university. His research interests include distributed meta-scheduling, Grid and Cloud computing, economic resource allocation, social computing, and services computing. He has founded and chaired two workshops and contributes to the organization of several others. He has served as the local organizing chair for several international conferences, including CCGrid, Cluster, and IPDPS, and currently serves on the program committee for several other conferences, such as eScience, BDC, ICWS, and CLOUD.