Editorial

Suraj Pandey

IBM Research,
Lvl 5 / 204 Lygon Street,
Carlton, VIC 3053, Australia
E-mail: surpande@au1.ibm.com

Surya Nepal*

CSIRO ICT Centre,
Crn Vimiera and Pembroke Roads,
Marsfield NSW 2122, Australia
E-mail: surya.nepal@csiro.au
*Corresponding author

Biographical notes: Suraj Pandey is a Research Scientist at IBM Research Australia. He joined the research lab in May 2012 as a research staff member to focus on industrial research and produce high-quality and high-impact results. Before joining IBM, he worked as a Research Fellow at the ICT Centre, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), in Marsfield, Sydney. At CSIRO, he published top research articles in journals and conferences, designed and implemented external client projects, and supervised PhD students.

Surya Nepal is a Principal Research Scientist at CSIRO ICT Centre, Australia. His main research interest is in the development and implementation of technologies in the area of service-oriented architectures, web services and cloud computing. He received his PhD from RMIT University, Australia and MSc from AIT, Thailand. He has published several journal and conference papers in the areas of multimedia databases, web services and service-oriented architectures, and security, privacy and trust in collaborative environment, cloud computing and social networks. He is also a programme committee member in many international conferences.

National Institute of Standards and Technology (NIST) defines cloud computing as "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell and Grance, 2009). As it is clear from this definition, cloud computing promises reliable services delivered through next-generation data centres that are built on compute and storage virtualisation technologies. Users will be able to access applications and data from a cloud anywhere in the world on-demand. In other words, the cloud appears to be a single point of access for all the computing needs of users. The users are assured that the cloud infrastructure is robust and will always be available at any time. Especially, application and research scientists are excited to use the new and improved model for executing their ever-increasing computing problems. Previously, scientists faced the problem of limited resources, authorisation and access issues with existing local resources, resource contentions, performance degradation due to shared resource model, collaboration across networks, and so forth. In contrast to the past cluster and/or grid computing model, cloud offers a virtualised data centre, which theoretically, is infinitely scalable and always available for anyone who is ready to pay. Also in practice, market leaders, such as Amazon Inc., are providing cloud services at massive scale through its web portals.

Cloud computing has started delivering impressive results. However, like its immediate predecessor, grid computing, it also has many challenges and unresolved research issues. Many of these issues have been simply transferred from previous paradigms to cloud in general. This makes the problem even more difficult to solve given that cloud computing model brings along new dimensions for management and optimisation. In order to study, communicate, and propose applicable solutions to these existing problems, the event on cloud computing and scientific applications (CCSAs) series was established.

The CCSA event series targets to promote research and development activities focused on enabling and scaling scientific applications using distributed computing paradigms, such as cluster, grid, and cloud computing. With the rapid emergence of software systems and their applicability, the volume of users is growing exponentially. User requirements are getting more and more complex. Existing computing infrastructure, software system designs, and use cases will have to take into account the huge volume of requests, size of data, computing load, locality and type of users, and so forth.

CCSA 2011 was organised in conjunction with UCC 2011 at Melbourne Australia and received innovative work on enabling and scaling computing systems to support the execution of scientific applications. CCSA brought together researchers and practitioners from around the world to share their experiences on modelling, executing, and monitoring scientific applications on clouds. The audience included researchers and industry practitioners who were interested in distributed systems, particularly focusing on scaling of applications using cloud computing.

The CCSA 2011 event was a huge success. The paper titled 'Improving data transfer performance of web service workflows in the cloud environment' by Donglai Zhang, Paul Coddington and Andrew L. Wendelborn, was given the best paper award. CCSA 2011 also had two excellent keynote speeches. The first keynote, titled: 'Resiliency principles for emergency management systems', was given by Jürg von Känel, Senior Research Manager, IBM Research and Development Australia. An extended abstract of this keynote is presented as a position paper in this special issue. The second keynote was titled: 'Issues and challenges in cloud storage services' by Surya Nepal, principal research scientist at CSIRO ICT Centre, Australia. This special issue comprises a position paper based on the presented keynote by Jürg von Känel and six selected research papers presented at CCSA 2011. We briefly present the summary of those papers in the following paragraph.

One of the basic requirements for the adaptation of cloud infrastructure in scientific applications is to be able to execute scientific workflows. Execution of scientific workflows in cloud is not straightforward. Though the management of control flow can be implemented, implementation of a data flow in a heterogeneous cloud environment is a challenging problem, as different cloud infrastructures are owned by different providers where data privacy may need to be maintained. The paper 'Improving data transfer performance of web service workflows in the cloud environment' aims to address this issue through a web service framework. This paper was given the best paper award at the CCSA 2011 event. The paper presents the performance analysis of their web service data forwarding (WSDF) framework using the ScienceCloud provided by the nimbus cloud infrastructure. In this framework, intermediate result from a previous service is treated as a resource of the composite service and can be directly used by its subsequent service, without sending it back to the centralised controller. The authors also show that the WSDF framework has significant performance advantage over normal web service framework for workflows consisting of large data transfers. This claim is substantiated with theoretical models and is presented in the paper.

The success of scientific applications in cloud environment depends on the sharing of analytic tools and data among scientists. In recent time, online social networks (OSNs) have been used as a platform for sharing tools and results of scientific experiments such as 'my experiment' (http://www.myexperiment.org/). The paper 'Extending HTTP models to Web 2.0 applications: the case of online social networks' identifies the failure of classical HTTP models and proposes an extension to the standard HTTP behavioural models by explicitly considering new elements characterising OSN applications. They study a traffic analysis of a popular social network for proving the correctness of the proposed approach.

One of the characteristics of scientific applications is the volume of data that needs to be processed (i.e., big data). Scientific applications ranging from astrophysics to climate change have to deal with an enormous amount of data. Traditional algorithms do not scale well for such a large dataset. There have been a number of programming model that have been developed for dealing with a large dataset. MapReduce is one of such programming models developed by Google. However, this programming model is not a solution for all scientific applications, as the transformation of scientific algorithms to MapReduce may not be straightforward and difficult to ingest for many scientists. Hence, a good programming model for scientific applications in a cloud environment is a challenging problem. In recent time, there have been attempts to tweak the MapReduce to fit it for certain scientific applications. The paper 'Simplifying MapReduce data processing' is one of such attempts. Although MapReduce is useful for decreasing complexity of processing large datasets in a distributed manner, the composition of an application into map and reduce function is a daunting task for common users. This paper emphasises this and aims to reduce this complexity. It has developed a web-based graphical user interface for end-users to use MapReduce without the need to programme the functions at the lowest level. Users specify their processing requirements in the form of target-value-action tuples, which then gets converted to a map-reduce task by their system. The paper provides real example scenarios as demonstrators.

It is envisaged that cloud computing technologies will become the driving force for realising the vision of 'computing as a utility' and driving the cost of computing significantly downward. The advancement of technologies has not yet matured to realise this dream. Therefore, scientists have to rely on volunteer contributions of resources such as Einstein@Home. The paper 'Managing volunteer resources in the cloud' aims to develop a similar concept, called Cloud@Home. In this paper, authors have developed a logical organisation of all the resources into a hierarchical cluster, implementing autonomic, distributed and self adapting algorithms to manage the hierarchical infrastructure according to the resource availability variations.

Many scientists rely on computational methods and simulations to solve the problems. The paper 'Applications of heterogeneous computing in computational and simulation science' reports some practical case studies in this area. This paper presents some heterogeneous algorithms and applications that have been developed by Commonwealth

Scientific and Industrial Research Organisation (CSIRO) for solving practical and challenging science problems. These algorithms use GPUs and multi-core CPUs on a scale ranging from single desktop workstations through to large GPU clusters.

Due to the private nature of many scientific data such as health data, it is not always possible to use public cloud. It is envisaged that the hybrid cloud model is going to be a solution for such scenarios where both public and private clouds are used for a single application. The paper 'How to exploit grid infrastructures for federated cloud purposes with CLEVER' aims to develop a middleware solution for hybrid cloud, called CLEVER. It also explores the potential benefits of using such middleware for GRID.

Finally, the position paper discusses the importance of resilience in IT infrastructures and systems as a quality property. This is equally applicable while developing scientific applications in clouds. The position paper 'Global technology trends: perspectives from IBM Research Australia on resilient systems' describes this property by presenting a case study of developing IT solution for the management of emergency scenarios in Australia.

We are confident this special issue will shed more light on the role of cloud computing in enabling and enhancing many aspects of scientific computing. The next series of the event, CCSA 2012, was held in conjunction with the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2012), on 13–16 May, 2012, in Ottawa, Canada.

Acknowledgements

The bringing out of this special issue is credited to all the researchers who submitted their excellent work to the CCSA series, presented the work at the event in person, and extended their work further to be included as part of the *IJCSE* journal. We thank the editors of *IJCSE* for including the CCSA series in *IJCSE*. We also thank all the programme committee members for reviewing the articles and helping improve their quality: Albert Zomaya, University of Sydney, Australia; Alberto Sanchez, Universidad Rey Juan Carlos, Spain; Craig Lee, Open Grid Forum, USA; Pavan Balaji, Argonne National Laboratory, USA; Ivona Brandic, Vienna University of Technology, Austria; Klaus-Dieter Schewe, Software Competence Center, Austria; Bahman Javadi, University of Western Sydney, Australia; Bart Pindor, University of Melbourne, Australia; Richard O. Sinnott, e-Research at University of Melbourne, Australia; James Sankar, AARNet Pty Ltd, Australia; Christian Simone Vecchiola, IBM Research and Development Australia; Bruno Schulze, National Laboratory for Scientific Computing, Brazil; Manish Parashar, Rutgers, The State University of New Jersey, USA; Anton Beloglazov, University of Melbourne, Australia; William Voorsluys, University of Melbourne, Australia; Abedelaziz Mohaisen, University of Minnesota, USA; and Nathalie Colineau, CSIRO, Australia.

References

Mell, P. and Grance, T. (2009) *The NIST Definition of Cloud Computing*, National Institute of Standards and Technology, Vol. 53, p.50.