Editorial

Bahman Javadi*

School of Computing, Engineering and Mathematics,
University of Western Sydney,
Locked Bag 1797,
Penrith, NSW 2751, Australia
Email: b.javadi@uws.edu.au
*Corresponding author

Saurabh Kumar Garg

Department of Computing and Information Systems,
|School of Engineering and ICT,
University of Tasmania,
Sandy Bay, TAS 7005, Australia
Email: Saurabh.Garg@utas.edu.au

Biographical notes: Bahman Javadi is a Senior Lecturer in Networking and Cloud Computing at the University of Western Sydney, Australia. Prior to this appointment, he was a Research Fellow at the University of Melbourne, Australia. From 2008 to 2010, he was a Postdoctoral Fellow at the INRIA Rhone-Alpes, France. He is co-founder of the Failure Trace Archive, which serves as a public repository of failure traces and algorithms for distributed systems. He has received numerous best paper awards at IEEE/ACM conferences for his research papers. His research interests include cloud and grid computing and reliability and fault tolerance.

Saurabh Kumar Garg is currently working as a Lecturer at the University of Tasmania, Australia. He is one of the few PhD students who completed in less than three years from the University of Melbourne in 2010. He has published more than 35 papers in highly cited journals and conferences with H-index 20 and more than thousand citations. His doctoral thesis focused on devising novel and innovative market-oriented meta-scheduling mechanisms for distributed systems under conditions of concurrent and conflicting resource demand. His current research interests are cloud computing, green computing, and big data analytics in healthcare, education and disaster management.

Recent exponential growth in data which is called as big data has created a revolution in the way our businesses and government organisations use to make decisions. Whether it is real-time fraud detection, web-based competitive analysis, detection of disease outbreak, disaster management or intelligent traffic management, the extraction and interpretation of information from big data is now a fundamental part of all decision making processes. In layman terms, big data can be defined as datasets/data streams that become so large that they become awkward to work with using on-hand computer data and computation management tools.

Such datasets are often from various sources (variety) in unstructured form, such as those from social media, sensors, scientific applications, surveillance, video and image archives, internet texts and documents, internet search indexing, medical records, business transactions and web logs. In general, these data streams are not only of a large size (volume) that cannot be stored but also have fast data in/out (velocity) and hide valuable knowledge. Big data size is beyond the capacity of commonly used system storage or computing capabilities within a reasonable time frame, hence, demanding new innovative solutions. Considering the complexity and scale of big data, using traditional techniques and models is not enough, and we need to propose new methodologies and frameworks for big data.

This opens several research questions and challenges, including investigating computing systems which can handle the storage, processing and networking requirements of big data. Using cloud resources for the storage and processing of big data applications is currently under investigation by many researchers.

This special issue features some of the most recent and emerging advances in the distributed computing systems to handle and process big data. This special issue covers state-of-art research in the area of hardware technologies and databases that are designed to support the processing of big data. Nunna et al. presented the current trends in hardware infrastructure that are utilised to process big data. They also presented the architectural features of future hardware systems that can combine CPUs along with GPUs to enable big data processing. Playne and Hawick benchmarked several modern GPU architectures and multi-GPU systems using the shallow water model to

understand their performance in relation to big data processing.

Another important advancement that is driven by big data is in the area of databases which aims at fast retrieval of the big data without much overhead. Lomotey et al. presented a data analytics tool that can discover knowledge from NoSQL databases. Zhang et al. presented a highly efficient caching and prefetching mechanism to retrieve efficiently the data from heterogeneous storage systems that is essential for high speed analysis of the big data. Finally, Sharma et al. presented classification and comparison of NoSQL databases to enable selection of data models for reliable and efficient management of big data.

We believe this special issue will not only increase the awareness of advancements in the area of large-scale distributed systems for big data but also inspire leading researchers and developers to investigate the problems in this area and thus benefit our society and businesses.