Guest Editorial——Cloud and Big Data

Niklas Carlsson, Xiaolin (Andy) Li, Mukesh Singhal, and Mea Wang

In the current digital age, massive amounts of data are generated in many different ways and forms. The data may be collected from everything from personal web logs to purposefully placed sensors. Today, companies and researchers use this data for everything from targeted personalized ads based on social data to solving important scientific problems that may help future generations of word citizens. Regardless if measured in monetary profit or other measures, the value of this data has proven valuable for many purposes and has led us into the Big Data era. Due to the large volume of data, Big Data requires significant storage, processing, and bandwidth resources. To date, the Cloud provides the largest collection of disk storage, CPU power, and network bandwidth, which makes it a natural choice for housing the Big Data.

However, the Cloud is more than just storage and computing facilities. It provides services in many forms: Infrastructure as a service (IaaS), Platform as a service (PaaS), Storage as a service (SaaS), Data as a service (DaaS), etc. The power of the Cloud has been demonstrated by the fast emergence of cloud services, such as iCloud and Google Drive. The Cloud provides great flexibility in storage, data processing, data management, new service provisioning, and more. In addition, the Cloud makes the data accessible anywhere at any time, which can further facilitate data generation, making Big Data even bigger.

When moving Big Data into the Cloud, it is important to note that large storage, processing, and bandwidth requirements offer three basic technical challenges. First, we need efficient and scalable storage solutions to store the data. The storage system should not only house the data, but also provide efficient access to the data. Second, we need efficient data processing solutions of heterogeneous structures. Third, we need efficient and scalable network solutions to move the data, particularly as data backup and data migration can be costly due to the high demand for network bandwidth. However, these are not the only challenges. The rapid growth of Big Data pose many other challenges, including challenges related to data mining, programming models, security, and many more.

In this special issue, we present four very interesting research papers that have gone through a thorough peer-review process. The papers present interesting results on various aspects of cloud management and big data, from failure detection to data placement, from application platform to workflow scheduling.

The first paper, "Cwmwl, a LINDA-based PaaS fabric for the Cloud", by Joerg Fritsch and Coral Walker, introduces Cwmwl, a cloud Platform-as-a-Service (PaaS) for efficient execution of tenant plugins and applications. The platform integrates the LINDA coordination language, Redis in-memory key-value store, and functional programming. The LINDA tuple space and coordination model provides mechanism for thread synchronization, while Redis makes parallel constructs of business application an elastic distributed application. Overall, the functional programming provides an environment for parallel execution. The system is implemented and tested in the SETTLE framework, a tool for measuring tuple space performance. The evaluation results indicate that Cwmwl offers effective system throughput (operations/second). The proposed system opens up new research opportunities. For example, it can be used as the basis for dynamic memory virtualization.

In the second paper, "Thermal Modeling and Analysis of Cloud Data Storage Systems", the authors, Xunfei Jiang, Maen M. Al Assaf, Xiaojun Ruan, Meikang Qiu, and Xiao Qin, investigate the power efficiency in cloud data centers, especially with regards to the power usage due to I/O accesses in the storage system. The study monitors the thermal behavior of a data storage node and proposed a thermal model to predict the outlet temperature of data nodes. The study also focuses on the impact of data placement strategy on the cooling cost and the thermal performance, based on which a thermal-aware energy-efficient data placement strategy was proposed. The prediction algorithm and the data placement strategy direct us to a new research direction on addressing the timely issues on energy efficiency and cooling cost in cloud data centers.

The final paper, "High-throughput Scientific Workflow Scheduling under Deadline Constraint in Clouds" by Michelle M. Zhu, Fei Cao, and Chase Q. Wu, proposes a job deadline-aware scheduling algorithm. The proposal is based on optimization for maximizing resource utilization for high system throughput, subject to delay constraints. The

doi: 10.12720/jcm.9.4.i-iii

scheduling algorithm first minimizes the end-to-end delay to meet the deadlines and then reduces the VM overhead to be efficient in resource utilization. Workflow scheduling is essential in handling big data and allocating cloud resources. The proposed algorithm not only effectively manages the workflow, but also significantly reduces the total execution time.

This special issue presents a collection of works on timely issues in cloud and big data. We would like to thank all the authors who submitted their work for this issue, and the reviewers for their timely and constructive feedback. These papers also shed lights on future research directions in cloud and big data. We would also like to thank the staff at the JCM Academy Publisher for their help in handling the manuscripts. Lastly, we would like to extend our sincere appreciation to Dr. Haohong Wang, Editor-in-Chief of the Journal of Communications, for his great support and for providing us with the opportunity to organize this special issue.

Guest Editors (in alphabetic order of last name)
Niklas Carlsson, Linköping University, Sweden (niklas.carlsson at liu.se)
Xiaolin (Andy) Li, University of Florida, USA (andyli at ece.ufl.edu)
Mukesh Singhal (IEEE Fellow), University of California, Merced, USA (msinghal at ucmerced.edu)
Mea Wang, University of Calgary, Canada (meawang at ucalgary.ca)

Prof. Niklas Carlsson
Linköping University, Sweden
E-Mail: niklas.carlsson at liu.se
Niklas Carlsson is an associate professor at Linköping University, Sweden. He received his M.Sc. degree in engineering physics from Umeå University, Sweden, and his Ph.D. in computer science from the University of Saskatchewan, Canada. His research interests are in the areas of design, modeling, characterization, and performance evaluation of distributed systems and networks.

Prof. Xiaolin (Andy) Li
University of Florida, USA
E-Mail: andyli at ece.ufl.edu
Xiaolin (Andy) Li is an associate professor and area chair of Computer Engineering Division in the Department of Electrical and Computer Engineering at the University of Florida. His research interests include Cloud Computing, High Performance Computing, Big Data, Software-Defined Networking and Ecosystems, Cyber-Physical Systems, and Network Security. He is directing the Scalable Software Systems Laboratory (S3Lab). He is in the executive committee of IEEE Technical Committee of Scalable Computing (TCSC). He has been a TPC chair for over 10 international conferences and workshops and an associate editor for several journals. He received a PhD in Computer Engineering from Rutgers University. He is a recipient of the National Science Foundation CAREER Award in 2010 and Internet2 Innovative Applications Award in 2013.

Prof. Mukesh Singhal (IEEE Fellow)
University of California, Merced, USA
E-Mail: msinghal at ucmerced.edu
Mukesh Singhal is a Chancellor's Professor in the Computer Science and Engineering program at the University of California, Merced. He received a PhD degree in Computer Science from the University of Maryland, College Park, in May 1986. From 1986 to 2001, he was a faculty in the department of Computer and Information Science at The Ohio State University. From 1998 to 2001, he served as the program director of the Operating Systems and Compilers program at the National Science Foundation. From 2001 to 2012, he was a Professor and Gartner Group endowed chair in Network Engineering in the Department of Computer Science at The University of Kentucky. His current research interests include distributed and cloud computing, cyber-security, and computer networks.

He has published over 240 refereed articles in these areas. He is a Fellow of IEEE and he was a recipient of 2003 IEEE Technical Achievement Award. He has coauthored four books titled "Advanced Concepts in Operating Systems", McGraw-Hill, New York, 1994, "Distributed Computing Systems", Cambridge University Press 2007, "Data and Computer Communications: Networking and Internetworking", CRC Press, 2001, and "Readings in Distributed Computing Systems", IEEE Computer Society Press, 1993. He has served in the editorial board of "IEEE Trans. on Dependable and Secure Computing", "IEEE Trans. on Parallel and Distributed Systems", "IEEE Trans. on Data and Knowledge Engineering", and "IEEE Trans. on Computers".

Prof. Mea Wang
University of Calgary, Canada
E-Mail: meawang at ucalgary.ca

Dr. Mea Wang is currently an Assistant Professor in the Department of Computer Science at the University of Calgary. She received her Bachelor of Computer Science (Honours) degree from the Department of Computer Science, University of Manitoba, Canada, in 2002. She received her Master of Applied Science and Ph.D. degrees from the Department of Electrical and Computer Engineering at the University of Toronto, Canada, in 2004 and 2008, respectively. Her research interests include Peer-to-Peer networking, multimedia networking, cloud computing, as well as networking system design and development. She has been serving as an Associate Editor for the Journal of Communications since 2010, the Guest Editor of the Special Issue on Multimedia Streaming in 2010 – 2011.