Advances in cloud and big data computing - Foreward to the special issue

Ladjel Bellatreche, Carson Leung, Yinglong Xia, Didier El Baz

Foreward to the Special Issue of Advances in Cloud and Big Data Computing

Ladjel Bellatreche 1, Carson Leung 2 Yinglong Xia 3 Didier Elbaz 4
1 LIAS/ISAE-ENSMA – Poitiers University, Futuroscope, Poitiers, France
2 University of Manitoba, Canada
3 Huawei Research America, USA
4 LAAS CNRS, Toulouse, France

1. BACKGROUND

We welcome you to this special issue initially dedicated to the best papers presented at the IEEE International Conference on Cloud and Big Data Computing (CBDCom 2016) that was held in Toulouse, France 18-21 July, 2016. Similar to CBDCom 2015, CBDCom 2016 is once again an event in the 2016 IEEE Smart World Congress, in conjunction to five other events, i.e., UIC 2016, ATC 2016, ScalCom 2016, IoP 2016 and SmartWorld 2016. CBDCom is a premier forum for researchers, practitioners, developers, and users who are interested in Cloud computing and Big Data and would like to explore new ideas, techniques and tools, as well as to exchange experience. CBDCom 2016 has been organized into 11 tracks: Big Data Algorithms, Applications and Services (co-chaired by Boualem Benatallah, University of New SouthWales, Australia Yassine Ouhammou, ISAE-ENSMA, France), Big Data Mining and Analytics (co-chaired by Sanjay Madria, Missouri University of Science and Technology, USA and Praveen Rao, University of Missouri-Kansas City, USA), Big Data Processing and Querying (chaired by Mayank Daga, AMD, USA), Cloud and Big Data for Internet of Things (IoT) (chaired by: Jian Tang, Syracuse U, USA), Government and Industrial Experiences for Cloud and Big Data (co-chaired by Toyotaro Suzumura, IBM Research, USA), Green Computing and Networking Technologies for Cloud and Big Data (co-chaired by: Kalyana Chadalavada, Intel, USA and Samee Khan, NDSU, USA), Software Engineering for Cloud Computing and Big Data (chaired by Annie T.T. Ying, IBM T.J. Watson Research Center, USA), Cloud Computing Solutions and Platforms (Cochaired by: Manisha Gajbe, Intel, USA and Carlos Garcia-Alvarado, Amazon, USA), Privacy and Security for Cloud and Big Data (co-chaired by Alfredo Cuzzocrea, CNR & University of Trieste, Italy and Nuyun Zhang, Clemson University, USA), Big Data Visualization (chaired by Nan Cao, New York University Shanghai, China & New York University, USA) and Big Data Education (co-chaired by Dickson K.W. Chiu, The University of Hong Kong, China and Wenhua Yu, Jiangsu Big Data Education Lab, China). CBDCom 2016 received 35 submissions from 28 countries or districts, covering 12 topics related to Cloud and Big Data, more particularly Big Data graph algorithms, virtualization, networking, Big data mining, privacy, parallel system technology (Spark, MapReduce, HDFS), emerging hardware trends in large-scale data processing, service composition in Cloud, and industrial experiences.All submissions were peer-reviewed by at least three reviewers from our international program committee consisting of professors and industrial researchers in relevant fields from 16 countries. Through many days of hard work and persistence, we were able to complete the review process for the papers appearing in this proceeding. Finally, we accepted 12 full papers, at the acceptance rate of 35%. To attract good papers, we manage our special issue for Journal Concurrency and Computation: Practice and Experience, Wiley as follows: out of the 12 full papers accepted in CDBCom 2016, five of them were invited to extend their paper by at least 50% new content. Also, an open call for papers has been organized and attracted seven papers covering the different topics of CBDCom 2016. In total, our special issue got 12 papers. After a second round of reviews, we finally accepted six papers. Thus, the relative acceptance rate for the papers included in this special issue is competitive. We congratulate the authors who submitted articles to CBDCom 2016 and our special issue. The six selected papers are summarized as follows:

The first paper titled, ”Learning the way to the Cloud: Big Data Park”, by Marchiori Massimo [1] , proposes a multidisciplinary topic combining the learning by doing teaching/learning model with the map-reduce programming model. It comes up with an interesting idea which can be put in practice and might represent a way to help children to get used with cloud computing. To doso, the paper shows the interest in making the new generations familiar with at least the basics of these new technologies. As in the field of technology, big data computing creates a total stir, among the young generations must intervene a readjustment to the new trends, too. In order to reach its objective, the article puts on the first line the Big Data Park one-line educational tool, which represents an innovative idea for making kids be interested and

warm-blooded of cloud computing. Also, the system comes up as a response against the classical way of teaching in schools which limits the thinking flexibility of the kids. Instead of watching and struggling with learning problems from outside of the scenario, the system brings the user in the middle of the action, starting from a low level and trying to reach the top of the game. In addition to this aspect, challenges, impacts and results on the young class of people have been reported after they passed through both the instructive/educational and funny/attractive techniques that the system offers. The second paper, titled ”Fuzzy ACID Properties for Self-Adaptive Composite Cloud Services Execution”, by Yudith Cardinale, Joyce El Haddad , Maude Manouvrier, and Marta Rukoz [2] discusses a self-adaptive model that aims at relaxing atomicity of composite service execution by introducing the notion of fuzzy atomicity. The fuzzy atomicity is built upon a set of transactional properties and is relaxed using either compensation or checkpointing. According to the user requirements, the (acceptable fuzzy atomicity), and the state of the composite service execution, the system provides an all-something-or-(almost)nothing model: users can obtain either all results, or partial results (but still significant), or nothing. The authors provide also a comparison of different transactional models relaxing ACID properties versus their model. The third paper titled, ”Accelerating Apache Spark with FPGAs”, by Ghasemi Ehsan and Chow Paul [3] presents a heterogeneous CPU-FPGA distributed platform that leverages the capabilities of the Spark data analytic framework and combines it with the power of FPGAs (field-programmable gate array) to achieve a high-performance, low-power, cluster-computing engine. As the case study, the focus is on demonstrating the feasibility of incorporating FPGA acceleration into Spark, and uses a MapReduce implementation of the k-means clustering algorithm to show that acceleration is possible even when using a hardware platform that is not well-optimized for performance. The fourth paper titled, ”Classifying very high-dimensional and large-scale multi-class datasetswith Latent-lSVM”, by Do Nghi and Poulet Franc¸ois [4] presents an interesting topic with a significant novelty aspect, which lies in the development of a local version of SVMs based on pre-extracted clusters with the help of Latent Diriclet Allocation (LDA) algorithm. Based on this development, it is possible to reduce large-scale problems containing several thousands of classes significantly to partially low-scale problems (as in each cluster typically a small subset of classes are present). The numerical results show significantly improved performance compared to classical SVMs and to other well-known and widely used classifiers on various real-world data sets, and this with a remarkably reduced computation time. The fifth paper titled, ”Parallel and distributed core label propagation with graph coloring”, by Attal Jean-Philippe, Malek Maria and Zolghadri Marc [5] studies an important problem related to the community detection problem via label propagation with graph coloring. This paper develops a parallel and distributed algorithm on Hadoop using the MapReduce framework to ensure the scalability of this algorithm. The algorithm is simple and intuitive, and its proposed implementation is sophisticated. The experiment results are convincing. The sixth paper titled, ”Deep Bayesian network architecture for Big Data Mining”, by Njah Hasna, Jamoussi Salma and Mahdi Walid [6] proposes numerous heuristics to improve 3 key learning steps: (i) clustering of variables: The algorithm implements overlapped clustering which yields to a loose hierarchy. This means that a given variable can have multiple parents. (ii) Latent variable learning that implements numerous heuristics such as Equilibrium Criterion for latent variable cardinality. It also implements a way to classify instances based on instance frequency and hamming distance, avoiding the computational burden of classical Expectation-Maximization algorithm. (iii) Hidden layer number: a stopping criterion for limiting the number of hidden layers is introduced. The iterative process for learning the Deep-Bayesian Network stops when the information loss of a hidden layer given its subsequent layer is greater than a given threshold. The authors evaluate the efficiency and the effectiveness of the proposed learning steps and compare them against the state-of-art methods.

2. ACKNOWLEDGEMENTS

We hope readers will find the content of this special issue interesting and will inspire them to look further into the challenges that are still ahead before designing, deploying and exploiting Big Data and Cloud applications. The guest editors of this special issue wish to express their sincere gratitude to all the authors who submitted their papers to this special issue. We are also grateful to the Reviewing Committee for the hard work and the feedback provided to the authors. As guest editors of this special issue, we also wish to express our gratitude to the Editor-in-Chief Geoffrey C. Fox for the opportunity to edit this special issue related to the Advances in Cloud and Big Data Computing, his assistance during the special issue preparation, and for giving the authors the opportunity to present their work in the international journal of Concurrency and Computation: Practice and Experience. Last but not least, we wish to thank the Journal’s staff for their assistance and suggestions. We acknowledge the following Reviewing Committee members: Amin Beheshti (Australia), Karim Benouaret (France), Yousef Elmehdwi (USA), Noura Faci (France),

Catarina Ferreira da Silva (France), Peter Graham (Canada), I-Ching Hsu (Taiwan), Wendy Hui-Wang (USA), Fan Jiang (Canada), Selma Khouri (Algeria), Hung-Yi Lin (Taiwan), Edwin Lughofer (Austria), Samir Ouchani (France), Amin Mesmoudi (France), Mourad Raphael (France), Syed Rizvi (USA), Soror Sahri (France), George Suciu (Roumania), and Muhammad Younas (UK).

REFERENCES

1. Marchiori Massimo, Learning the way to the Cloud: Big Data Park, Concurrency and Computation: Practice and
Experience [This issue].
2. Cardinale Yudith, El Haddad Joyce, Manouvrier Maude, Rukoz, MartaFuzzy ACID Properties for Self-Adaptive
Composite Cloud Services Execution, Concurrency and Computation: Practice and Experience [This issue].
3. Ghasemi Ehsan and Chow Paul, Accelerating Apache Spark with FPGAs, Concurrency and Computation: Practice
and Experience [This issue].
4. Do Nghi and Poulet Franois, Classifying very high-dimensional and large-scale multi-class datasets with LatentlSVM,
Concurrency and Computation: Practice and Experience [This issue].
5. Attal Jean-Philippe, Malek Maria and Zolghadri Marc, Parallel and distributed core label propagation with graph
coloring, Concurrency and Computation: Practice and Experience [This issue].
6. Njah Hasna, Jamoussi Salma and Mahdi Walid, Deep Bayesian network architecture for Big Data Mining,
Concurrency and Computation: Practice and Experience [This issue].