Future Generation Computer Systems 43–44 (2015) 38–39

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Editorial

Software Tools and Techniques for Big Data Computing in Healthcare Clouds

As we delve deeper into the 'Digital Age', we witness an explosive growth in the volume, velocity, and variety of the data available on the Internet. For example, in 2012 about 2.5 quintillion bytes of data was created on a daily basis. The data originated from multiple types of sources including mobile devices, sensors, individual archives, social networks, Internet of Things, enterprises, cameras, software logs, health data etc. Such 'Data Explosions' has led to one of the most challenging research issues of the current Information and Communication Technology (ICT) era: how to effectively and optimally manage such large amount of data and identify new ways to analyze large amounts of data for unlocking information. The issue is also known as the 'Big Data' problem, which is defined as the practice of collecting complex data sets so large that it becomes difficult to analyze and interpret manually or using on-hand data management applications. From the perspective of real-world applications, the Big Data problem has also become a common phenomenon in domain of science, medicine, engineering, and commerce. Representative applications include clinical decision support systems, digital agriculture, social media analytics, high energy physics, earth observation, genomics, automobile simulations, medical imaging, body area networks, translational medicine, and the like.

An important class of Big Data application exists in the healthcare domain. There are wide varieties of health related datasets that play a critical role in the health information systems (HIS) and clinical decision support systems (CDSS). These datasets differ widely in their volume, variety, and velocity, from patient focused sets such as electronic medical records to population focused sets such as public health data, and knowledge focused sets such as drug-to-drug, drug-to-disease, disease to disease interaction registries. While decision makers' (healthcare practitioner, government decision makers) ability to understand and process the health data dictates the accuracy of the final decision, the exponential growth in the size of the aforementioned health related raw data sets has widened this integration gap. This further makes the timely information aggregation, retrieval, and analysis a challenge. This is severely limiting the potential benefits of having large datasets and HIS/CDSS for medical decision-making processes.

Another important class of Big Data application in the healthcare domain includes the Medical Body Area Networks (MBANs). According to the market intelligence company ABI research (http://www.abiresearch.com/), over the next five years, close to five million disposable wireless MBAN sensors will be shipped. MBANs enable a continuous monitoring of patient's condition by sensing and transmitting measurements such as heart rate, electrocardiogram (ECG), body temperature, respiratory rate, chest sounds, and blood pressure etc. MBANs will allow: (i) real-time and historical monitoring of patient's health; (ii) infection control; (iii) patient identification and tracking; and (iv) geo-fencing and vertical alarming. However, to manage and analyze such massive MBAN data from millions of patients in real-time, healthcare providers will need access to an intelligent and highly secure ICT infrastructure.

In all of the aforementioned health application scenarios, hundreds of petabytes of heterogeneous data (images, text, video, raw sensor data, and the like) will be generated and required to be efficiently processed (stored, distributed, and indexed with an ontology and semantics) in a way that does not compromise end-users' Quality of Service (QoS) in terms of data availability, data search delay, data analysis delay, and the like. Many of the existing ICT systems that store, process, distribute, and index hundreds of petabytes of heterogeneous data fall shortly of this challenge or do not exist. We need to develop new techniques that aims to optimize all these in less than 10 ms and to achieve this without any cloud configuration knowledge (i.e., by automatically detecting cloud storage proximity and the QoS of network links between storage alternatives).

We believe that Cloud computing infrastructures (e.g., Amazon, Microsoft Azure, etc.) in conjunction with fast communication networks, data-intensive programming paradigms (MapReduce, distributed storage system, etc.), semantic web, and machine learning algorithms will form the basis of designing and developing Big Data Analytics based innovation framework in health domain. We need to develop software tools and techniques that allow for fast query processing and speeds-up data analytics in a global cloud computing based Big Data network that exploits such data provide awareness and knowledge in real-time.

In this special issue, the progress has been made by applying and extending well-founded formal models and techniques from multiple domains of computer science. Xu et al. develops the Knowle [1], a semantics-rich self-organized network, which reflects various semantic relations of concepts, news, and events. Furthermore, Knowle can be used for organizing and mining health news, which shows the potential on forming the basis of designing and developing big data analytics based innovation framework in health domain. Tang et al. [2] proposes an optimal reduce scheduling policy called SARS (Self Adaptive Reduce Scheduling) for reduce tasks' start times in the Hadoop platform for Big Data processing. Yang et al. [3] develops a Medical Image File Accessing System (MIFAS) based on HDFS of Hadoop in cloud. The proposed system can improve medical imaging storage,

http://dx.doi.org/10.1016/j.future.2014.11.001
0167-739X/

transmission stability, and reliability while providing an easy-to-operate management interface. Yang et al. [4] proposes a practical solution for privacy preserving medical record sharing for cloud computing. Based on the classification of the attributes of medical records, they use vertical partition of medical dataset to achieve the consideration of different parts of medical data with different privacy concerns. Chang et al. [5] build an inference model based on ontology and Bayesian Network for inferring the possibility of getting depressed and implement a prototype using mobile agent platform to show the proof-of-the-concept using mobile cloud. Abbas et al. [6] use the Multi-attribute Utility Theory (MAUT) to help users compare different health insurance plans based on coverage and cost criteria, such as: (a) premium, (b) co-pay, (c) deductibles, (d) co-insurance, and (e) maximum benefit offered by a plan. Chen et al. [7] proposes an approach that measures the synchronization strength of bivariate non-stationary nonlinear data against phase differences. Castiglione et al. [8] propose an engine for lossless dynamic and adaptive compression of 3D medical images, which also allows the embedding of security watermarks within them. Jrad [9] designed an ontological model for a semantic description of the problem and developed a novel utility-based genetic matching algorithm for selecting the Cloud services with respect to the user requirements and the properties of the Clouds. Zhang [10] propose a task-level adaptive MapReduce framework. This framework extends the generic MapReduce architecture by designing each Map and Reduce task as a consistent running loop daemon.

References

[1] Zheng Xu, Xiangfeng Luo, Yunhuai Liu, Lin Mei, Chuanping Hu, Lan Chen, Knowle: a semantic link network based system for organizing large scale online news events, Future Gener. Comput. Syst. 43–44 (2015) 40–50.

[2] Zhuo Tang, Lingang Jiang, Junqing Zhou, Kenli Li, Keqin Li, A self-adaptive scheduling algorithm for reduce start time, Future Gener. Comput. Syst. 43–44 (2015) 51–60.

[3] Chao-Tung Yang, Wen-Chung Shih, Lung-Teng Chen, Cheng-Ta Kuo, Fuu-Cheng Jiang, Fang-Yie Leu, Accessing medical image file with co-allocation HDFS in cloud, Future Gener. Comput. Syst. 43–44 (2015) 61–73.

[4] Ji-Jiang Yang, Jian-Qiang Li, Yu Niu, A hybrid solution for privacy preserving medical data sharing in the cloud environment, Future Gener. Comput. Syst. 43–44 (2015) 74–86.

[5] Yue-Shan Chang, Chih-Tien Fan, Win-Tsung Lo, Wan-Chun Hung, Shyan-Ming Yuan, Mobile cloud based depression diagnosis using ontology and Bayesian network, Future Gener. Comput. Syst. 43–44 (2015) 87–98.

[6] Assad Abbas, Kashif Bilal, Limin Zhang, Samee U. Khan, A cloud based health insurance plan recommendation system: A user centered approach, Future Gener. Comput. Syst. 43–44 (2015) 99–109.

[7] Chang Cai, Ke Zeng, Lin Tang, Dan Chen, Weizhou Peng, Jiaqing Yan, Xiaoli Li, Towards adaptive synchronization measurement of large-scale non-stationary non-linear data, Future Gener. Comput. Syst. 43–44 (2015) 110–119.

[8] Arcangelo Castiglione, Raaele Pizzolante, Alfredo De Santis, Bruno Carpentieri, Aniello Castiglione, Francesco Palmieri, Cloud-based adaptive compression and secure management services for 3d healthcare data, Future Gener. Comput. Syst. 43–44 (2015) 120–134.

[9] Foued Jrad, Jie Tao, Ivon Brandic, Achim Streit, SLA enactment for large-scale healthcare workflows on multi-cloud, Future Gener. Comput. Syst. 43–44 (2015) 135–148.

[10] Fan Zhang, Junwei Cao, Samee Khan, Keqin Li, Kai Hwang, A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications, Future Gener. Comput. Syst. 43–44 (2015) 149–160.

Lizhe Wang*
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, PR China
E-mail address: Lizhe.wang@gmail.com.

Rajiv Ranjan
CSIRO Computational Informatics, Australia

Joanna Kołodziej
Department of Computer Science, Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Cracow, Poland

Albert Zomaya
School of Information Technologies, The University of Sydney, Australia

Leila Alem
CSIRO Computational Informatics, Australia

* Corresponding editor.