Future Generation Computer Systems 63 (2016) 96–99

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Editorial

Modeling and Management of Big Data: Challenges and opportunities

David Gil [a,*], Il-Yeol Song [b]

[a] Lucentia Research Group, Computing Technology and Data Processing, University of Alicante, Spain
[b] College of Computing and Informatics, Drexel University, Philadelphia, PA 19104, USA

HIGHLIGHTS

- Objectives of the third International Workshop on Modeling and Management of Big Data (MoBiD'14).
- Summary of the selected papers.
- Conceptual modeling in the big data era.
- Expectation in these topics for this and the next editions of this workshop.

ARTICLE INFO

Article history:
Received 14 June 2015
Received in revised form
28 July 2015
Accepted 31 July 2015
Available online 24 August 2015

Keywords:
Conceptual modeling Big Data
Ecosystem
Integrate & analyze & visualize

ABSTRACT

The term Big Data denotes huge-volume, complex, rapid growing datasets with numerous, autonomous and independent sources. In these new circumstances Big Data bring many attractive opportunities; however, good opportunities are always followed by challenges, such as modelling, new paradigms, novel architectures that require original approaches to address data complexities. The purpose of this special issue on Modeling and Management of Big Data is to discuss research and experience in modelling and to develop as well as deploy systems and techniques to deal with Big Data. A summary of the selected papers is presented, followed by a conceptual modelling proposal for Big Data. Big Data creates new requirements based on complexities in data capture, data storage, data analysis and data visualization. These concerns are discussed in detail in this study and proposals are recommended for specific areas of future research.

1. Introduction

The aim of the International Workshop on Modeling and Management of Big Data is to bring together researchers, developers and practitioners to discuss research issues and experience in modeling, developing and deploying systems and techniques to deal with Big Data. The third International Workshop on Modeling and Management of Big Data (MoBiD'14), held in Atlanta, October, 27–30, 2014 was a continuation/evolution of the previous workshops, the International Workshop on Modeling for Data-Intensive Computing (MoDIC'12), held in Florence, Italy, October 15–18, 2012 and MoBiD'13 held in Hong Kong, November 11–13, 2013. MoBiD'14 was presented with the aim to attract papers on the latest and best proposals for modeling and managing Big Data in this new era of the data-drive paradigm. This new conceptualization of big data applications incorporating both internal and external Big Data requires new models and methods to accomplish their conceptual modeling phase. Papers focusing on the application and the use of conceptual modeling approaches for Big Data, MapReduce, Hadoop and its ecosystems, big data analytics, social networking, cloud computing, security and privacy, data science, etc. were highly encouraged. Therefore, the workshop has been an international forum for researchers and practitioners who are interested in the different facets related to the use of the conceptual modeling approaches for the development of next generation of applications based on Big Data. We view that several key themes with the Big Data trend include (i) using a cloud for large-scale external and internal data; (ii) providing an easy-to-use but powerful services to access/manage/analyze the big data in the cloud; (iii) defining a problem-solving space and developing an architecture for a big data environment to conceptualize goals, tasks, and problem-solving methods to apply to domains; and (iv) managing big data and analyzing them to discover business values.

2. Papers

MoBiD'14 attracted papers from 9 different countries distributed all over the world: France, Greece, India, Japan, Kenya, Korea, Spain, United Kingdom and USA. We have finally received 14

* Corresponding editor.
E-mail address: dgil@dtic.ua.es (D. Gil).

http://dx.doi.org/10.1016/j.future.2015.07.019
0167-739X/

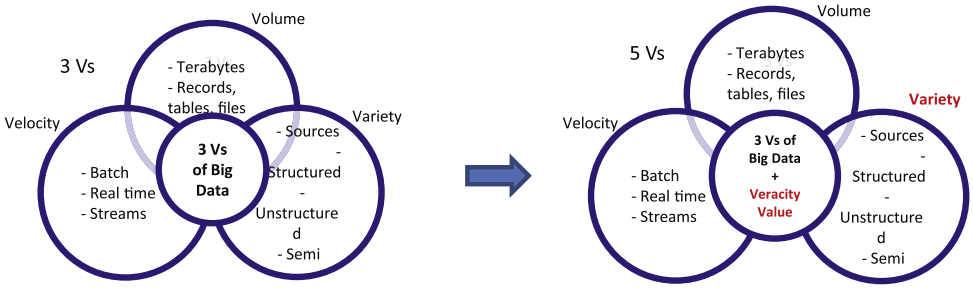

Fig. 1. Defining big data with 3 V's and moving towards 5 V's.

papers and the Program Committee has selected 5 papers, making an acceptance rate of 35%. In the following, we summarize these selected papers:

The first paper, "From Business Intelligence to Semantic Data Stream Management" by Marie-Aude Aufaure, Raja Chiky, Olivier Curé, Houda Khrouf and Gabriel Kepeklian [1], introduces recent work on Real-Time Business Intelligence that utilizes semantic data stream management. This paper addresses the new tendencies of real-time systems that are continuously generating data to be analyzed, processed, and stored. They also present underlying approaches to continuous queries and data summarization.

The second paper, entitled "Design Science Research Contribution to Business Intelligence in the Cloud — A Systematic Literature Review" by Odette Sangupamba Mwilu, Nicolas Prat and Isabelle Comyn-Wattiau [2] deals with the new opportunities for business intelligence (BI) and analytics offered by Cloud computing and big data. They propose a typology of artifacts potentially produced by researchers in design science. Then, after analyzing the state of the art through that typology, they use it to sketch opportunities of new research to improve BI and analytics capabilities in the cloud and from big data.

The third paper, "A Data Quality in Use Model for Big Data" by Jorge Merino, Ismael Caballero, Bibiano Rivas, Manuel Serrano and Mario Piattini [3] is a position paper that proposes the 3Cs model, which is composed of three data quality dimensions for assessing the quality-in-use of big datasets: Contextual Consistency, Operational Consistency and Temporal Consistency. The aim is that the quality of data lacks a quality-in-use model adapted for big data.

The fourth paper, "A Hybrid Integrated Architecture for Energy Consumption Prediction" by Alejandro Maté, Jesus Peral, Antonio Ferrández, David Gil, Juan Trujillo [4] explores the opportunities of using ICT (Information and Communication Technologies) as an enabling technology to reduce energy consumption in cities. It proposes a multidimensional hybrid architecture that makes use of current energy data and external information (with unstructured data sources) to improve knowledge acquisition and allow managers to make better decisions.

The last paper of this Special Issue, entitled "Benchmarking Performance for Migrating a Relational Application to a Parallel Implementation" by Krishna Karthik Gadiraju, Manik Verma, Karen C. Davis, and Paul G. Talaga [5] investigates the impact of scaling up the data sizes for several benchmarking queries. They illustrate what kind of performance results an organization could expect when they migrate current applications to big data environments. The authors measure the speedup for query execution for all dataset sizes resulting from the scale up. They conclude that Hive loads the large datasets faster than MySQL, while it is marginally slower than MySQL when loading the smaller datasets.

3. Conceptual modeling in the big data era

The experience in conjunction with the novelty and the new trends developed during the last three years lead us to summarize our thoughts and expectation in these topics for this and the next editions of this workshop.

Big data is a very broad term which is often easily understood by means of a graphical representation (Fig. 1) in order to not only pay attention to the "Big" word, but especially to understand that "big data" express the difficulty into dealing with data in different dimensions.

Currently there are too many scenarios where the term Big Data appears. Scientists, business executives, practitioners of media and advertising and governments alike regularly meet difficulties with large data sets in areas including Internet search, finance and business informatics [6,7].

There are many domains whose data management needs have exploded. For example, we can discuss data management challenges of E-commerce along the three dimensions: volume, velocity and variety.

- On Volume: "The lower cost of e-channels enables an enterprise to offer its goods or services to more individuals or trading partners. The explosion of the data to be collected in e-commerce are even up to 10x of the quantity of data about an individual transaction, thereby significantly increasing the overall volume of data to be managed."
- On Velocity: "E-commerce has also increased point-of-interaction (POI) speed, and consequently the pace data used to support interactions and generated by interactions."
- On Variety: "No greater barrier to effective data management will exist than the variety of incompatible data formats, non-aligned data structures, and inconsistent data semantics."

Where does big data come from? (i) "data exhaust" from customers; (ii) new and pervasive sensors; (iii) the ability to "keep everything" [8,9].

In [10] it is indicated that with the significant advances in Information and Communications Technology (ICT) over the last half century, there is an increasingly perceived vision that computing will one day be the 5th utility (after water, electricity, gas, and telephony). It is defined Cloud computing and provide the architecture for creating Clouds with market-oriented resource allocation by leveraging technologies such as Virtual Machines (VMs). The proliferation of the devices in a communicating actuating network creates the Internet of Things (IoT) [11]. In [12] it is analysed the challenges and requirements for next-generation Big Data services and presented a solution designed to support next-generation Big Data applications.

Regarding the difficulty of managing Big Data, it has been stated that Big Data is any data that is expensive to manage and hard to extract value from [13]. Among the Vs shown in Fig. 1, in this brief summary, we will focus on The V of Variety. This will lead us to the main topic of the workshop which is Conceptual modeling of Big Data.