CrossMark

Guest Editorial: A Special Issue in Physical Design for Big Data Warehousing and Mining

**Ladjel Bellatreche[1] · Pedro Furtado[2] ·
Mukesh K. Mohania[3]**

Big data is data that exceeds the processing capacity of conventional systems, typically because it is too big, moves too fast (high incoming rates), or is too unstructured for the traditional database approaches to handle. Research involving big data primarily concerns itself with how to discover and make sense of such high amounts of data. Researchers devise solutions for future systems to deal with these challenges more effectively and efficiently. Issues investigated include capturing, searching, storing, analyzing, sharing, and visualizing big data.

Meanwhile, data warehousing and mining technologies have evolved to scale and to analyze large volumes of data. There were major advances concerning architectures, physical design, indexing, query processing, optimization and parallel processing. But the volume, variety, and velocity of big data require us to rethink mechanisms to be able to cope with new requirements.

There was great response to the call for papers; we received 13 papers from 9 countries (Australia, China, France, India, Italy, Korea, Morocco, Tunisia, and the USA). Due to the limited space, only five papers were accepted and selected for this special issue.

✉ Ladjel Bellatreche
   bellatreche@ensma.fr

   Pedro Furtado
   pnf@dei.uc.pt

   Mukesh K. Mohania
   mkmukesh@in.ibm.com

[1] LIAS/ISAE-ENSMA – Poitiers University, 86961 Chasseneuil, France

[2] Coimbra University, 3030-290 Coimbra, Portugal

[3] IBM Research, Delhi 110001, India

🍂 Springer

This special issue discusses advances in architecture and design for the warehousing and mining of big data. The works presented provide interesting answers to questions on how to deal efficiently with high-rate industrial process data, the visualization of big data, the handling of huge scientific datasets, and graph big data and spatial data. These topics cover domains such as astronomy, engineering, energy, etc. The particularity of these papers is that most of them used case studies derived from academic and industrial projects [e.g. PetaSky: http://com.isima.fr/Petasky and Electricity of France (EDF)].

The five selected papers are summarized as follows:

The paper titled *Chronos: a NoSQL System on Flash Memory for Industrial Process Data* by Brice Chardin, Jean-Marc Lacombe, and Jean-Marc Petit presents the design of a system to handle the archiving of industrial process data. Given the lack of optimizations of traditional database management systems concerning the use of flash memories, especially in scenarios with write-intensive workloads, they propose Chronos, an open-source NoSQL system that supports acquisition rate improvements in the range of 20–54 when compared with other existing solutions. Their solution is based on an append-only approach for insertions and index management techniques optimized for process data management over flash memories.

The paper titled *On visualizing large multidimensional datasets with a multi-threaded radial approach* by Tianyang Liu, Fatma Bouali, and Gilles Venturini deals with the issue of the visualization of large amounts of multidimensional data. The authors propose POIViz, a radial visualization approach that uses points of interest to determine the layout of a large dataset. They also study the efficiency of the approach using parallelization on CPUs and GPUs, concluding that it is possible to visualize, in less than one second, millions of pieces of data with tens of dimensions, and to support "real-time" interactions even for large datasets.

The paper titled *Benchmarking SQL On MapReduce systems using large astronomy databases* by Amin Mesmoudi, Mohand-Saïd Hacid, and Farouk Toumani discusses the ability of SQL on MapReduce systems to handle large astronomy databases where the data size can reach many petabytes. In this paper, Mesmoudi et al. focus on the problem of evaluating the performance of existing SQL on MapReduce data management systems using astronomy data and queries. They experiment on the ability of such systems to support large-scale declarative queries. They mainly investigated the impact of data partitioning, indexing, and compression on query execution performances in that context. In practice, this work compares mostly the performances of Hadoop-based data management systems while dealing with a number of diverse configurations related to the queries, the data, and the machines that reside within the clusters.

The paper titled *Scalable Graph-based OLAP Analytics over Process Execution Data* by Seyed-Mehdi-Reza Beheshti, Boualem Benatallah, and Hamid Reza Motahari-Nezhad proposes a framework and a set of methods to support scalable graph-based OLAP analytics over process execution data. In this paper, the authors note that graph data has some fundamentally different characteristics from traditional analytics solutions. They propose a new framework and approach to deal with analytics over huge graph data and study scalable graph-based OLAP analytics over process execution data. The approach is able to summarize big process graphs and to provide multiple views at different granularities using OLAP specific abstractions in process

context, such as process cubes, dimensions, and cells. A MapReduce-based graph-processing engine is defined to support big data analytics over those process graphs.

Finally, the paper titled *Spatial Data Warehouses and Spatial OLAP come towards the Cloud: design and performance* by Rodrigo Costa Mateus, Thiago Luís Lopes Siqueira, Valéria Cesário Times, Ricardo Rodrigues Ciferri, and Cristina Dutra de Aguiar Ciferri studies how to bring spatial OLAP to the cloud. In this paper, Mateus et al. discuss the issues raised when hosting a spatial data warehouse in the cloud and processing spatial OLAP queries over such data. They introduce novel concepts such as a cloud spatial data warehouse and spatial OLAP as a service. Then, they detail the design of a novel schemata and approaches for handling spatial OLAP in that environment. They also introduce a CSB-index to speed up the performance of spatial OLAP queries over cloud spatial data warehouses. Finally, they evaluate the performance of the proposals.

We hope the readers of DAPD will find the content of this special issue timely and that it will inspire them to look further into the challenges that are still ahead before designing advanced information systems using Computational Intelligence. We would like to thank all the authors who submitted their papers to this special issue. In addition, we are grateful for the support of various reviewers that ensured the high quality of this special issue. Last but not least, we would like to thank Professors Amit Sheth and Divyakant Agrawal, Editors-in-Chief of this journal for accepting our proposal for this special issue focused on Advances in Physical Design for Big Data Warehousing and Mining, and for assisting us whenever required. We would like to thank very much Springer's editorial and publication support teams for their endless help and support. The complete International Program Committee of this special issue is listed next.

International Program Committee

Reza Akbarinia, INRIA, Montpellier, France
Mohammed Al-Kateb, Teradata, USA
Ladjel Bellatreche, LIAS/ISAE-ENSMA, Poitiers, France
Luc Bouganim, INRIA Paris-Rocquencourt, France
Jalil Boukhobza, University of Occidental Brittany, Brest, France
Sebastian Breß, TU Dortmund, Germany
Brice Chardin, LIAS/ISAE-ENSMA, Poitiers, France
Alain Crolotte, Teradata, USA.
Pedro Furtado, Coimbra University, Portugal
Helena Galhardas, University of Lisbon, Portugal
Carlos Garcia Alvarado, Pivotal Software Inc., USA
Allel Hadj Ali, LIAS/ISAE-ENSMA, Poitiers, France
Omar Hussain, UNSW Canberra, Australia
Stéphane Jean, LIAS/ISAE-ENSMA, Poitiers, France
Carson K. Leung, The University of Manitoba, Canada
Samee Khan, North Dakota State University, USA
Selma Khouri, LIAS/ISAE-ENSMA, France
Sanjay Kumar Madria, Missouri University of Science and Technology, USA

Jens Lechtenboerger, University of Munster, Germany
Sofian Maabout, Labri, Bordeaux, France
Yannis Manolopoulos, Aristotle University of Thessaloniki, Greece
Sameep Mehta, IBM Research, India
Anirban Mondal, Xerox Research, India
Rim Moussa, INRIA, Montpellier, France
Carlos Ordonez, Houston University, USA
Jorge R Bernardino, Instituto Superior de Engenharia de Coimbra, Portugal
Srinath Srinivasa, IIIT, Bangalore, India
Nambiar Ullas, EMC, India
Panos Vassiliadis, University of Ioannina, Greece
Robert Wrembel, Poznan University of Technology, Poland