# Advances in Data Warehousing and OLAP in the Big

# Data Era

Ladjel Bellatreche, LIAS/ISAE-ENSMA – Poitiers University, France. E-mail: bellatreche@ensma.fr

Alfredo Cuzzocrea, DIA Department, University of Trieste and ICAR-CNR, Italy. E-mail: alfredo.cuzzocrea@dia.units.it

Il-Yeol Song, College of Computing & Informatics, *Drexel University, PA, USA.* E-mail: *songiy@drexel.edu*

We welcome you to the special issue featuring the best papers presented at the 16th ACM Data Warehousing and OLAP Workshop 2013 (ACM DOLAP) in Burlingame, CA, USA. DOLAP continues its tradition of being a high quality forum where both researchers and practitioners in data warehousing and On-Line Analytical Processing (OLAP) share their findings on theoretical foundations, current database design methods, new trends and practical experiences. In recent years, research in these areas has addressed many topics, covering all phases of the life cycle of data warehousing applications (user requirements, conceptual phase, ETL, logical phase, deployment phase, physical design, exploitation). However, the successful use of data warehousing and OLAP technologies within organizations has brought up new requirements and research issues, in particular to cope with non-traditional application domains, such as text, blogs, tweets, etc.

The DOLAP 2013 call for papers attracted 26 submissions. After careful review and discussion, the program committee accepted 7 full papers and 6 short papers. Out of the 7 full papers, we selected 5 papers to be invited for the special issue in the Information Systems journal. After a second round of reviews we finally accepted all these papers, which attest to the quality of DOLAP papers.

In general, research papers presented at DOLAP 2013 show data warehousing is now a mature and well-established computer science field, but still with many important research issues. ETL and modelling aspects remain challenging given the ever-changing nature of databases, both in structure and in size. New applications, faster hardware and the Internet are making data warehousing technology evolve even further. Thus we expect research on data warehousing and OLAP to remain active in the future.

The paper titled "Clustering binary Cube Dimensions to Compute Relaxed Group by Aggregations" by Carlos Garcia-Alvarado and Carlos Ordonez, proposes to replace a GROUP BY with several cube dimensions with another GROUP BY clause, by grouping by a cluster identifier, corresponding to groups of similar fact table records. Such aggregation is useful when the cardinality of the output table is large or when the cube has high dimensionality. A key innovation is that a traditional SQL grouping is based on the equality of tuples, whereas a cluster-based grouping is based on a similarity measure. The authors propose a fast incremental window-based k-means algorithm that allows the end user to provide the number of cluster k as an input parameter, effectively controlling the output table size. The paper provides an experimental evaluation showing that their clustering approach produces a more succinct summary than traditional cube queries. Moreover, since clustering is done fast, queries remain interactive.

The paper titled “CineCubes: aiding data workers gain insights from OLAP queries,” by Dimitrios Gkesoulis, Panos Vassiliadis and Petros Manousis proposes a Human–Computer Interaction extension to cube processing based on the metaphor of movies given in acts to provide further insights over the results of a traditional OLAP query. The proposed system, called CineCubes, receives an OLAP query as input and enriches it with auxiliary, explanatory queries that give context and extra details to the original one. Then, the results, along with important highlights that are extracted from them, are presented to the user as PowerPoint slides that are further accompanied with commenting text, which is transformed, into speech using text-to-speech synthesis. The findings of a user study assessing the system reveal novel usage patterns along with a pressing need for concise results.

The paper “Advanced Topic Modeling for Social Business Intelligence,” by Stefano Rizzi, Enrico Gallinucci and Matteo Golfarelli proposes to combine data with user generated contents to allow decision makers to improve their business activities. To do so, the authors propose the meta-stars approach to model topic hierarchies in ROLAP systems.   The basic idea of the approach is to use meta modelling coupled with navigation tables and with dimension tables.  The paper gives detailed formalizations of the meta-stars model as well as the querying expressiveness of the approach. The paper also provides an execution plan strategy and a set of tests to validate the approach.

The paper titled, “Slowly Changing Measures,” by Mathias Goller and Stefan Berger gives a systematic view of the possible techniques for dealing with changes in the way measures are computed. The authors use ideas for handling dimension changes as a foundation to handle changing measures. One of the merits of the paper is the fact that it borrows techniques, which are already normally adopted by practitioners who have to face that problem. After discussing the pros and cons of these techniques, the authors present their method through a representative set of examples.  Each slowly changing measure type is tested in a proof-of-concept prototype.

The paper titled “INDREX: In-Database Relation Extraction,” by Torsten Kilias, Alexander Loser and Periklis Andritsos, proposes INDREX, a complex system that allows users to describe so-called relation extraction tasks across documents and relational data stored in a RDBMS, allowing these tasks to be processed via SQL statements. INDREX provides built-in indexing structures and query optimization techniques that enable fast response times. Semantically, relation extraction tasks are enriched by white-box user-defined functions to adapt extraction rules with existing domain specific data from RDBMS. The authors complement their analytical contributions by means of a comprehensive experimental evaluation performed in both a single pipelined DBMS (i.e., PostgreSQL) and a columnar-based massively parallel processing query engine (i.e., Cloudera Impala), which clearly show the effectiveness and efficiency of INDREX in supporting in-database relation extraction.

BIO

Ladjel Bellatreche is a Professor at National Engineering School for Mechanics and Aerotechnics ([ENSMA](http://www.ensma.fr" \t "_blank)), Poitiers, where he joined as a faculty member since Sept 2010. He leads the Data and Model Engineering Team of Laboratory of Computer Science and Automatic Control for Systems (LIAS) ([LIAS](http://www.lias-lab.fr" \t "_blank)). Prior to that, he spent eight years as Assistant and then Associate Professor at [Poitiers University, France](http://www.univ-poitiers.fr" \t "_blank). He was a Visiting Professor of the Québec en [Outaouais, Canada](http://www.uqo.ca" \t "_blank), a Visiting Researcher at [Department of Computer Science, Purdue University, USA](http://www.cs.purdue.edu" \t "_blank) and [Department of Computer Science of Hong Kong University of Science and Technology, China](http://www.cse.ust.hk/%22%20%5Ct%20%22_blank). He is also involved in Research Postgraduate Programmes in Computer Science of several Universities and Schools in Algeria. Prof. Ladjel Bellatreche has been actively involved in the research community by serving as reviewer for technical journals (IEEE TKDE, DKE, Distributed and Parallel Database Journal, JoDS, etc.) and Editorial Board Member, International Journal of Reasoning-based Intelligent Systems, Inderscience, subject area editor of the Scalable Computing Journal, Springer and as an organizer/co-organizer of numerous international and National Conferences and Workshops (ADBIS, DAWAK, DASFAA, DOLAP, MEDI, WISE, EDA, JFO). Some recent conferences in which he is playing or has played major roles include DAWAK, DOLAP, MEDI, and WISE Workshops. In addition, he served as a program committee member for over forty international conferences and Workshops. Ladjel Bellatreche actively contributes in promoting research in Africa and Asia, where he co-supervises several students and organizes conferences and workshops (ICT-EurAsia, MEDI, CIIA, etc.).

<http://www.lias-lab.fr/members/bellatreche>

**Il-Yeol Song** is professor in the College of Computing and Informatics of Drexel University and Director of Ph.D. Program in Information Studies in his college. He served as Deputy Director of NSF-sponsored research center on Visual & Decision Informatics (CVDI) between 2012-2014. He is also an affiliated professor of Computer Science Department of KAIST, Korea. He is an ACM Distinguished Scientist and an ER Fellow. He is the recipient of 2015 Peter P. Chen Award in Conceptual Modeling. His research interests include conceptual modeling, data warehousing & OLAP, big data management & analytics, CRM, object-oriented analysis & design, healthcare informatics, and smart health. Dr. Song published over 190 peer-reviewed papers and co-edited 22 proceedings. He is a co-Editor-in-Chief of Journal of Computing Science and Engineering (JCSE) and is in an editorial board member of DKE, JDM, IJEBR, and JDFSL. He won the Best Paper Award in the IEEE CIBCB 2004. He won 14 research awards from competitions of annual Drexel Research Days. He also won four teaching awards from Drexel, including the most prestigious Lindback Distinguished Teaching Award. Dr. Song served as the Steering Committee chair of the ER conference between 2010-2012. He is also a steering committee member of DOLAP, BigComp, and ADFSL conferences. He served as a program/general chair of over 20 international conferences/workshops including DOLAP’98-14, CIKM’99, ER’03, FP-UML’06, DaWaK’07-‘08, , DESRIST’09, CIKM ‘09, MoDiC’12, and MoBiD’13-‘15.

http://www.cci.drexel.edu/faculty/**song**

**Alfredo Cuzzocrea** is currently Associate Professor at the DIA Department, University of Trieste, Italy. He is also habilitated as Full Professor in Computer Science Engineering by the French National Scientific Habilitation of the National Council of Universities (CUN) under the hegira of Ministry of Higher Education and Research (MESR). Previously, he has been Senior Researcher at the Institute of High Performance Computing and Networking of the Italian National Research Council, Italy, and Adjunct Professor at the University of Calabria, Italy. He holds 36 Visiting Professor positions worldwide (Europe, USA, Asia, Australia). He serves as Springer Fellow Editor. He serves as Elsevier Ambassador. He holds several roles in international scientific societies, steering committees for international conferences, and international panels, some of them having directional responsibility. He served as Panel Leader and Moderator in international conferences. He served as Invited Speaker in several international conferences worldwide (Europe, USA, Asia). He is member of scientific boards of several PhD programs worldwide (Europe, Asia, Australia). He serves as Editor for the Springer series “Communications in Computer and Information Science”. He covers a large number of roles in international journals, such as Editor-In-Chief, Associate Editor, Special Issue Editor (including JCSS, IS, KAIS, FGCS, DKE, INS, BigData Research). He edited more than 30 international books and conference proceedings. He is member of editorial advisory boards of several international books. He covers a large number of roles in international conferences, such as General Chair, Program Chair, Workshop Chair, Local Chair, Liaison Chair and Publicity Chair (including CSE, ODBASE, DaWaK, DOLAP, ICA3PP, ICEIS, APWeb, SSTDM, IDEAS, IDEAL). He served as Session Chair in a large number of international conferences (including EDBT, CIKM, DaWaK, DOLAP, ADBIS). He serves as Review Board Member in a large number of international journals (including TODS, TKDE, TKDD, TSC, TIST, TSMC, THMS, JCSS, IS, KAIS, FGCS, DKE, INS). He serves as Review Board Member in a large number of international books. He serves as Program Committee Member in a very large number of international conferences (including VLDB, ICDE, EDBT, CIKM, IJCAI, KDD, ICDM, PKDD, SDM). His current research interests include multidimensional data modeling and querying, data stream modeling and querying, data warehousing and OLAP, OLAM, XML data management, Web information systems modeling and engineering, knowledge representation and management models and techniques, Grid and P2P computing, privacy and security of very large databases and OLAP data cubes, models and algorithms for managing uncertain and imprecise information and knowledge, models and algorithms for managing complex data on the Web, models and algorithms for high-performance distributed computing and architectures. He is author or co-author of more than 350 papers in international conferences (including EDBT, CIKM, SSDBM, MDM, DaWaK, DOLAP), international journals (including JCSS, IS, KAIS, DKE, INS) and international books (mostly edited by Springer). He is also involved in several national and international research projects, where he also covers responsibility roles.

*http://si.deis.unical.it/~cuzzocrea*