

Guest Editorial: Special Section on the International Conference on Data Engineering

Johannes Gehrke, Beng Chin Ooi, *Fellow, IEEE*, and Evaggelia Pitoura

THE 28th International Conference on Data Engineering was held in Washington, DC, on April 1-5, 2012. ICDE 2012 attracted 413 submissions in the research track, 19 submissions in the industrial track, and 68 submissions in the demo track. After a thorough review from the corresponding committees, 100 research papers, six industrial contributions, and 28 demo proposals were selected for inclusion in the conference program.

This special section consists of journal versions of 10 outstanding papers selected among the 100 accepted research contributions. All papers were revised and substantially extended, over their conference versions and went through a rigorous review process to ensure the high quality standards of the *IEEE Transactions on Knowledge and Data Engineering*. They cover a broad range of topics highlighting the liveliness of the data engineering field.

The paper “Answering Why-Not Questions on Top-K Queries” by Zhian He and Eric Lo extends the problem, of explaining missing tuples in query results, known as the why-not problem, for the case of top-k queries. The authors develop algorithms for processing such why-not questions efficiently and present case studies and experimental results to demonstrate the efficiency of their approach in providing high quality explanations.

The paper “BestPeer++: A Peer-to-Peer Based Large-Scale Data Processing Platform” by Gang Chen, Tianlei Hu, Dawei Jiang, Peng Lu, Kian-Lee Tan, Hoang Tam Vo, and Sai Wu presents a peer-to-peer data management system that provides elastic data sharing services, for corporate data processing applications in the cloud. The system was evaluated on the Amazon EC2 platform and shown to outperform HadoopDB on enterprise workloads.

The paper “Discovering Conservation Rules” by Lukasz Golab, Howard Karloff, Flip Korn, Barna Saha, and Divesh Srivastava proposes a new class of constraints, termed Conservation Rules (CRs), for data quality analysis in monitoring infrastructure networks. The authors provide confidence metrics to quantify the strength of a CR and approximation algorithms for the problem of discovering a concise summary of the subset of the data that satisfies a given CR. The utility of CRs is demonstrated through a set of experiment with real data.

The paper “A Generic Framework for Top- k Pairs and Top- k Objects Queries over Sliding Windows” by Zhitao

Shen, Muhammad Aamir Cheema, Xuemin Lin, Wenjie Zhang, and Haixun Wang provides a general framework for processing multiple top-k object and top-k pair queries over sliding windows. Given a function that computes a score between two objects, a top-k pair query returns the k pairs with the largest score. A detailed complexity analysis is presented along with experimental results.

The paper “Incremental Detection of Inconsistencies in Distributed Data” by Wenfei Fan, Jianzhong Li, Nan Tang, and Wenyuan Yu focuses on incremental detection of functional dependency violations in distributed data. Given a distributed database partitioned either vertically or horizontally, a set of conditional functional dependencies (CFDs), a set of violations of the CFDs, and updates of the database, the problem studied is locating with minimum data shipment, changes to the set of violations in response to changes in the database. They show that the incremental detection problem is NP-complete and provide appropriate algorithms.

The paper “LARS*: An Efficient and Scalable Location-Aware Recommender System” by Mohamed Sarwat, Justin J. Levandoski, Ahmed Eldawy, and Mohamed F. Mokbel presents a recommender system that incorporates the notion of location. It combines spatial ratings for non-spatial objects, non-spatial ratings for spatial objects, and special ratings for spatial objects. The authors also exploit user rating locations to influence recommendations with ratings spatially close to the query. Experimental results are presented using real data from Foursquare and MovieLens.

The paper “Mapping XML to a Wide Sparse Table” by Liang Jeff Chen, Philip A. Bernstein, Peter Carlin, Dimitrije Filipovic, Michael Rys, Nikita Shamgunov, James F. Terwilliger, Milos Todoc, Sasa Tomasevic, and Dragan Tomic proposes a novel mapping of XML data into one wide table with sparsely populated columns that provides good performance for workloads observed in enterprise applications but not supported efficiently by existing work. Rewrite rules that minimize the number of joins are also presented.

The paper “On the Use of Side Information for Mining Text Data” by Charu C. Aggarwal, Yuchen Zhao, and Philip S. Yu presents a clustering algorithm for documents which combines classical partitioning algorithms with probabilistic models and then they show how to extend the approach to the classification problem. Experimental results on a number of real data sets are presented.

The paper “PrefDB: Supporting Preferences as First-Class Citizens in Relational Databases” by Anastasios Arvanitis and Georgia Koutrika makes the case for placing preference aware query processing closer to the DBMS. The authors propose a preference-aware relational data model and an associated algebra. They also present query

- J. Gehrke is with Cornell University and Microsoft.
- B.C. Ooi is with the National University of Singapore, Singapore.
- E. Pitoura is with the University of Ioannina, Greece.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2014.2314529

processing strategies that push preference evaluation inside the query plan and compare their approach with two plug-in variations using two real data sets.

Finally, the paper “Towards Multi-Tenant Performance SLOs” by Willis Lang, Srinath Shankar, Jignesh M. Patel, and Ajay Kalhan studies Service Level Objectives (SLOs) in the case of Database-as-a-Service (DaaS) cloud architectures. The authors present a framework that takes as input the tenant workloads, their performance SLOs, and the server hardware available to the DaaS provider and produces a cost effective specification of how much hardware to provision and how to schedule the tenants on each hardware resource.

We would like to thank all the paper authors and the reviewers of this special section for all their hard. We would also like to extend our thanks to the program committee, the organization committee, and the authors and participants of ICDE 2012 for making the conference a successful and memorable event. We hope that you will find the selected papers interesting and informative and that you will enjoy this special section.

Johannes Gehrke
Beng Chin Ooi
Evaggelia Pitoura
Guest Editors
ICDE 2012 Program Committee Chairs



Johannes Gehrke is the Tisch University Professor in the Department of Computer Science at Cornell University and a Distinguished Engineer at Microsoft. His research interests are in Big Data, Data Science, and Distributed Systems. He has received the 2011 IEEE Computer Society Technical Achievement Award, and he co-authored the undergraduate textbook. *Database Management Systems* (McGrawHill (2002), currently in its third edition), used at universities all over the world. He is also an adjunct professor at the University of Tromsø in Norway.



Beng Chin Ooi received the BSc (first class honors) and PhD degrees from Monash University, Australia, in 1985 and 1989, respectively. He is a distinguished professor of computer science at the National University of Singapore (NUS). His research interests include database system architectures, performance issues, indexing techniques, and query processing, in the context of multimedia, spatio-temporal, distributed, parallel, P2P, and Cloud database systems and applications. He is the recipient of the ACM SIGMOD 2009 Contributions Award, a co-winner of the 2011 Singapore President’s Science Award, the recipient of 2012 IEEE Computer Society Kanai Award, 2013 NUS Outstanding Researcher Award, and 2014 IEEE TCDE CSEE Impact Award. He is a fellow of the IEEE and the ACM.



Evaggelia Pitoura received the PhD degree from Purdue University. She currently serves as a professor and an associate chair in the Computer Science and Engineering Department at the University of Ioannina, Greece. Her research interests include the area of data management systems with a recent focus on social networks and graphs. For her research, she has received three Best Paper Awards (ICDE 1999, VLDB 2013, and DBSocial 2013), a Marie Curie Fellowship (2009) and two Recognition of Service Awards from the ACM.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.