Special issue on best papers of VLDB 2014

H. V. Jagadish¹ · Aoying Zhou²

Published online: 28 August 2015
© Springer-Verlag Berlin Heidelberg 2015

The VLDB conference is a premier venue for presenting advances in the research and practice of data management. The VLDB 2014 conference took place in Hangzhou, China, in September 2014.

Papers for the VLDB conference are chosen through a year-round reviewing process with cutoff dates for conference inclusion. The VLDB 2014 conference roughly coincided with Volume 7 of PVLDB, which received 695 submissions. Of these, 139 were accepted, but 21 were accepted too late for inclusion in VLDB 2014. The remaining 118 were presented at VLDB 2014. In addition, there were 47 papers from PVLDB Vol. 6 presented at VLDB 2014, for a total of 165 papers.

A best paper committee, comprising Dimitris Papadias (Chair), Jayant Haritsa, and Kian-Lee Tan, chose seven papers out of these 165 to invite for inclusion in VLDB Journal. Of these seven, five papers were finally accepted in extended form for publication in this special issue, after two additional rounds of review. These papers provide a nice sampling of the rich frontier of database research today, touching upon five high points, each in a very distinct sub-area.

In today's era of Big Data, Hadoop-style map-reduce implementations are everywhere. However, these implementations make strong assumptions about the structure of the data and the computation, rendering them unsuitable for many practical situations. The epiC system, described in *epiC: an Extensible and Scalable System for Processing Big Data* by Dawei Jiang, Sai Wu, Gang Chen, Beng Chin Ooi and Kian-Lee Tan, overcomes these limitations through an

innovative modular design and an actor-like programming model. The result is a system that can be used effectively for complex analyses and also for multiple distinct applications in parallel.

Indexes are central to database performance. However, index design is an art, and index construction is a heavy-weight task. Database cracking has emerged in recent years as a technique to construct indexes incrementally and adaptively as a side effect of query processing. Many algorithms have been proposed for this purpose. In *An Experimental Evaluation and Analysis of Database Cracking* by Felix Martin Schuhknecht, Alekh Jindal, and Jens Dittrich, the authors present a comprehensive study and comparative evaluation of the current state of the art.

Data visualization is an important output of database systems and is increasingly used with data analytics today. As data volumes increase, visualization capabilities are stressed, as are other parts of the system. *VDDA: Automatic Visualization-Driven Data Aggregation in Relational Databases* by Uwe Jügel, Zbigniew Jerzak, Gregor Hackenbroich, and Volker Markl takes a creative approach to dealing with data volume by recognizing that the number of pixels on the screen limits the granularity of data that can be shown. Working backwards from here, data can be aggregated to obtain much smaller databases, without hurting the quality of displayed results. In contrast, naïve aggregation, without taking visualization effects into account, can lose information of value.

There is great interest today in seamless querying across multiple types of media. Traditional indexing techniques rely on notions of distance that make sense only if data objects can be mapped to points in some meaningful feature space. How to take objects in different media types and map them to a uniform meaningful feature space is a challenge. This challenge is addressed in *Effective Deep Learning-Based Multi-Modal*

✉ H. V. Jagadish
jag@umich.edu

¹ University of Michigan, Ann Arbor, MI, USA

² East China Normal University, Shanghai, China

Retrieval by Wei Wang Xiaoyan Yang, Ph.D. Beng Chin Ooi, Ph.D. Dongxiang Zhang, and Ph.D. Yueting Zhuang. The paper proposes technique to learn a mapping from heterogeneous data source types to a common metric space over which an index can be constructed.

On k-Path Covers and their Applications by Stefan Funke, Andre Nusser, and Sabine Storandt is directed at very large graphs that we find so often today, such as in social networks or road network data. A common need in such large graphs is to find a small subset of “important” nodes, to which further analysis can be restricted. While there are many ways in which this importance could be defined, and the preferred definition could be different for different applications, one definition that is frequently of interest is the k-path cover. Informally, we seek a subset of nodes such that there is at least one node on every long-enough path in the graph. This paper presents a comprehensive analysis of this problem and effective solutions for large graphs.

We thank the authors of the invited papers for investing substantial time and effort into extending the original conference version of the papers. We would also like to express our sincere thanks to the diligent reviewers who provided perspective, advice, and keen assessment of the submissions. Finally, we are grateful to the best paper committee for their work in selecting the best among a large set of very nice papers at the VLDB conference. We hope you enjoy this issue.

H. V. Jagadish and Aoying Zhou
Guest Editors of the Special Issue
Program Chairs of VLDB 2014, and
Editors in Chief of Vol. 7 of PVLDB.