Editor's Notes

Welcome to the December 2013 issue of the ACM SIGMOD Record!

First, congratulations to the newly recognized ACM fellows which are also prominent members of our scientific community: Charu C. Aggarwal, Stefano Ceri, Peter Haas, Matthias Jarke, Timos Sellis, Dennis Shasha, Kyuseok Shim, Val Tannen, and Limsoon Wong!

The issue opens with a Database Principles article by Ngo, Ré and Ruda, on new developments in the theory of join algorithms. While the general perception may be that the important things about join processing are already well-known, the authors make a very solid case against this perception. Specifically, they present a novel unifying framework of two recent results on the algorithmic complexity of join query evaluation; these optimal algorithms are not the ones taught in textbooks over the last twenty to thirty years or so, although a single commercial system already implemented one of them! Two main ideas lie behind the novel optimal algorithms: first, exploit together the query shape and statistics on the data (whereas most previously proposed optimization frameworks used either one or the other); second, process all the joins simultaneously, as opposed to ordering them and performing them one by one. I am thrilled to have this paper published in the Record, as it provides a great introduction to a very recent family of contributions on which many future works will likely be based.

The Surveys column features three articles. First, Bordawekar, Blainey and Apte consider the currently hot area of research and development generally termed "data analytics", outline the main algorithmic techniques involved, and study how each can be mapped effectively for execution on a parallel infrastructure. The authors identify a set of popular analytical models and identify for each suitable parallelization strategies for large-scale data management workloads.
The second survey, by Li, Wang, Li and Gao, focuses on the problem of approximate (similarity) joins on XML data sets. Such joins are useful, for instance, when integrating heterogeneous data sets of tree-structured data. In such contexts, approximate joins are needed, and to avoid performing too many comparisons, lower bounds on the distance between two trees are necessary. The survey outlines three families of edit distances, namely string-based, histogram-based and binary branch distance. The authors show how such distances can be combined, and also compare their associated computational costs.
The survey by Felix Naumann focuses on the topic of profiling data, seen broadly as deriving or extracting metadata under the form of statistics or other information, out of a dataset. Data profiling is recognized as an important area given the abundance of new data sets, which must be integrated in existing or novel applications. The survey identifies data profiling application areas, such as data integration, analytics, or scientific data management; it outlines previous works in the area and also discusses novel trends such as incremental data profiling, profiling on novel architectures etc.

The Distinguished Profiles in Databases column features an interview with Anand Rajaraman from WallMart Labs. Read it to find out how one co-authors the same year a SIGMOD and a VLDB paper that would 10 years later each get the respective Test-of-Time/10 years best paper award, leave the PhD program to create the Junglee start-up, come back to finish the PhD without a scholarship but paying the registration costs out of the pocket, be close to buying an obscure start-up named Google (without buying it in the end), co-head @WalmartLabs, and now invest in and advise Sillicon Valley startups!

In the Research Centers column, Christodoulakis, Garofalakis, Petrakis, Deligiannakis, Samoladas, Ioannou, Papapetrou and Sotiriadis describe ongoing data management research at the Technical University of Crete. Within the Software Technology and Network Applications (SoftNet) lab, research areas include stream monitoring, data as a service in a cloud context, and uncertain/probabilistic data management. The Intelligent Systems lab studies architectures and tools for cloud-based deployment of

complex software applications, in particular for healthcare services. Finally, the Distributed Multimedia Information Systems and Applications (MUSIC) laboratory investigates semantic-based integration and interoperability of digital contents, for application contexts such as digital libraries, natural history, or cultural heritage.

The issue features three event reports. Darmont and Pedersen report on the First International Workshop on Cloud Intelligence (Cloud-I 2012), held in conjunction with the VLDB 2012 Conference in Istanbul, Turkey. The report sessions focused on topics such as data analytics, security and privacy in the cloud.
The report on the Second International Worskhop on Energy Data Management (EnDM 2013) is by Pedersen and Lehner. The workshop was held in conjunction with EDBT 2013 in Genova; it featured works on representing, extracting and visualizing energy data, energy forecasts in local distribution networks etc.
The last workshop report is from the 2nd workshop on Scalable Workflow Enactment Engines and Technology (SWEET) held in 2013 next to the SIGMOD conference. Sroka, Hidders and Missier outline the keynotes (by prof. Paul Watson from Newcastle University and Jelena Pjesivac-Grbovic from Google), and the papers on topics such as simulating execution on heterogeneous hardware, workflow scheduling, or user steering within e.g. scientific workflows.

The issue closes with two call for contributions for the ACM SIGIR conference, respectively, for the new ACM e-Energy conference, both to be held in 2014.

On behalf of the SIGMOD Record Editorial board, I wish all our community a happy and prosperous 2014, full of success and joy!

Your submissions to the Record are welcome via the submission site:

http://sigmod.hosting.acm.org/record

Prior to submitting, be sure to peruse the Editorial Policy on the SIGMOD Record's Web site (http://www.sigmod.org/publications/sigmod-record/sigmod-record-editorial-policy).

Ioana Manolescu

December 2013

Past SIGMOD Record Editors:

Harrison R. Morse (1969)
Daniel O'Connell (1971 – 1973)
Randall Rustin (1974-1975)
Douglas S. Kerr (1976-1978)
Thomas J. Cook (1981 – 1983)
Jon D. Clark (1984 – 1985)
Margaret H. Dunham (1986 – 1988)
Arie Segev (1989 – 1995)
Jennifer Widom (1995 – 1996)
Michael Franklin (1996 – 2000)
Ling Liu (2000 – 2004)
Mario Nascimento (2005 – 2007)
Alexandros Labrinidis (2007 – 2009)