Editor's Notes

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

The article opening this issue is an outline of research around "The Continuous Distributed Monitoring Model" by Cormode. The setting considered is one where multiple distributed sites collaborate in monitoring some (potentially distributed) phenomenon and must implement, together, a specific computation on their monitoring result, such as counting events, or detecting "unusual" activity according to a given (un)usual activity profile. The key is to accurately capture the interesting information, for instance, record each event exactly once, or detect each unusual pattern, with as little communication effort as possible among the participating sites. The survey outlines existing works on variants of the problem such as distributed countdown (the sites must sum up their observations to detect the arrival of a given number of events) and entropy monitoring (where the problem is to detect changes in the entropy of the distributed system under observation), and other extensions, while also pointing to the connection with related areas of work and outlining possible advances to be made in future research.

The survey by Schomm, Schall and Vossen focuses on the hot topic of data marketplaces, defined as places where anyone (or at least a large number of users) can connect to upload and/or obtain datasets, for free or for a fee. The authors identify twelve categories according to which existing data marketplaces can be classified, and present the distribution of 46 data marketplaces surveyed as of mid-2012 according to these categories. Depending on the core product (=data) they peddle, marketplaces can be split in several classes, such as crawlers that are services to be invoked to obtain data from a given target crawl site, search engines over specific existing databases, raw data vendors, tagging services (which enrich a dataset given as input with annotations based on an ontology (semantic model) etc. The analysis is also based on other dimensions such as the data pricing model, data formats, and the maturity of the marketplace platforms.

In the Distinguished Profiles column, we have the pleasure of reading Hank Korth' reflections on his long successful career as a researcher at Bell Labs and Panasonic, and a professor and deparment head at LeHigh University among others. The Korth-Silberschatz (later Korth-Silberschatz-Sudarshan) database book has reached its sixth edition in print, which is at least a reason why most if not all of our readers already know one of Hank's works very well! Hank also shares in the interview thoughts on his experience with transferring research results to the industry – a topic obviously non-trivial even for a large company such as Bell Labs, teaching computer science to non-CS majors, playing ultimate Frisbee with the undergraduate students, seizing career opportunities and much more.

The Research Centers column features two research center presentations. The first one, by Chakrabarti, Ramakrishnan, Ramamrithan, Sarawagi and Sudarshan, is from the IIT Bombay. The paper outlines the group's research areas, featuring an interesting mix of database, information retrieval and machine learning. The research topics comprise graph querying, keyword search, entity annotation, Web tables, working with imprecise data, continuous queries and streams. The second one, by Stonebraker, Madden and Dubey, is from the Intel "Big Data" science and technology center, created in 2012. The report outlines the center's vision of the current state of "Big Data" technology and the problems currently open, which in the authors' view include complex analytics on big-volume data, and integrating high-velocity data with a large persistent state.

Two workshop reports are included in this issue. Pedersen, Lehner and Hackebroich report on the works of the Energy Data Management (EnDM) 2012 workshop, held in conjunction with EDBT 2012. The workshop focused on the data management issues raised by intelligent distribution networks, e.g., electricity networks using smart meters, performance and energy reduction—aware file placement, smart

grids, database techniques for energy environmental impact assessment, and gaming techniques to encourage users to adopt energy saving behaviors. The Cloud Data Management (CloudDB) workshop, held next to ACM CIKM 2012, is described in the report by Meng, Wang and Silberstein. The workshop has covered topics such as OLTP benchmarking in the cloud, data analytics, scaling out social applications, privacy and security, energy-efficient clouds, as well as more specific performance-oriented techniques for large-scale data management in the cloud.

The Call for Participation to the ACM SIGMOD/PODS yearly conference closes this issue. Looking forward to see you all in New York!

Your contributions to the Record are welcome via the RECESS submission site (http://db.cs.pitt.edu/recess). 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

March 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)