

2009 SIGMOD/PODS Award Winners

SIGMOD Edgar F. Codd Innovations Award

For innovative and highly significant contributions of enduring value to the development, understanding, or use of database systems and databases. Until 2003, this award was known as the "SIGMOD Innovations Award." In 2004, SIGMOD, with the unanimous approval of ACM Council, decided to rename the award to honor Dr. E.F. (Ted) Codd (1923 - 2003) who invented the relational data model and was responsible for the significant development of the database field as a scientific discipline.



2009 SIGMOD Edgar F. Codd Innovations Award Winner: Masaru Kitsuregawa

Masaru Kitsuregawa is the recipient of the 2009 SIGMOD Edgar F. Codd Innovations Award for *contributions to high-performance database technology*. Kitsuregawa made major contributions to the development of hash-join algorithms, which significantly improved the performance of join operations in relational database systems. That work has influenced related research in areas such as query execution, plan optimization and dynamic query-workload balancing, as well as the development of commercial database products. He implemented the hash-based approach on a variety of platforms, including the Functional Disk System and multi-node PC clusters, demonstrating its substantial advantages through detailed evaluations. He has also applied hash-based strategies to parallel association mining and showed its effectiveness there. His contributions in the hardware area include a high-speed sorting system with a sophisticated memory management algorithm. That work was eventually commercialized in collaboration with colleagues, and won the Datamation sort benchmark in 2000.

Details

Professor Kitsuregawa has contributed extensively to the area of high-performance database systems, particularly involving hash-based methods. The work began in the early 1980's in the context of the GRACE relational-database machine. He is particularly known for his work on hash-based join algorithms, which is still widely cited. By the late 1980's and early 1990's, others had built on that work to develop various hybrid versions of hash join, and most database conference of that time had sessions devoted to the topic. His own refinements include dynamic destaging and bucket tuning. At the time, most commercial relational-database products used only looping and sort-based joins. Nearly all current system include hash-based join implementations. He also contributed hash-based approaches to aggregation operations.

He went on to implement the Functional Disk System (FDS), a parallel, hash-based relational system with a shared-memory architecture. He demonstrated that efficient parallel execution of relational operations was possible with hash-based methods, showing substantial performance improvement on the Wisconsin Benchmark. He also developed database-engine software for a shared-nothing architecture on a 100-node PC cluster, which was evaluated against other systems on the TPC-D Benchmark in the late 1990s. He is among the first to apply hash-based approaches to parallel data mining.

SIGMOD Contributions Award

For significant contributions to the field of database systems through research funding, education, and professional services.



2009 SIGMOD Contributions Award Winner:

Beng Chin Ooi

Beng Chin Ooi is the recipient of the 2009 SIGMOD Contributions Award for *his sustained and selfless contributions to the database community in promoting and pursuing high standards of database research at both the international and regional level*. Beng Chin Ooi has been consistently involved in activities that push the frontier of database research. He has taken leadership roles on technical programs for ICDE, SSD, SIGMOD and VLDB. He has helped ensure a high standard for database journal publications as an editor for The VLDB Journal, IEEE Transactions on Knowledge and Data Engineering, International Journal of Geographical Information Science, and Geoinformatica. Currently, he is the editor-in-chief of IEEE Transactions on Knowledge and Data Engineering, and an editor for the Distributed and Parallel Databases journal. He also serves on the SIGMOD Doctoral Dissertation Award committee and on the board of the VLDB Endowment. Ooi has contributed greatly to promoting database research, especially in the Asia-Pacific region, by building an eminent database group at the National University of Singapore that supports high quality research throughout the region through direct collaborations and participation in regional conferences.

Details

Beng Chin Ooi has been continuously involved in leadership roles that help create strong technical programs for the community. He served as the co-PC chair for the Third International Symposium on Large Spatial Databases (SSD) in 1993. He also served as the vice PC Chair for the International Conference on Data Engineering (ICDE) in 2000, 2004 and 2006. More recently, Ooi was involved in running two major database conferences, serving as PC chair for the ACM SIGMOD International Conference on Management of Data in Beijing in 2007 and as the Core DB PC chair in the Very Large Database Conference (VLDB) in Auckland in 2008. Ooi has consistently worked to maintain the high quality of these conferences.

Ooi has also been involved in ensuring high standards of database journal publications. He was an editor for The VLDB Journal, IEEE Transactions on Knowledge and Data Engineering (TKDE), International Journal of Geographical Information Science and Geoinformatica. He has recently been appointed editor-in-chief of TKDE and continues to serve as an editor for the Distributed and Parallel Databases journal.

Ooi is currently a board member of the VLDB Endowment and is involved in the liaison and fostering of cooperation between the Endowment and database researchers in the Asia-Pacific region. Since 2005 he has served as a member the ACM SIGMOD Dissertation Award committee. Since 2008 he has served as co-chair of this committee.

Ooi is an active proponent for database research, especially in the Asia-Pacific region. He has been engaging researchers from other countries in the region in collaborative projects with a view to enhancing the standard of work being done. This outreach has resulted in fruitful collaborations with leading universities in the region, including Fudan University, Tsinghua

University and Renmin University of China. Ooi has served on the program committees of many regional conferences. In 2005, he was co-PC chair of the International Conference on Database Systems for Advanced Applications (DASFAA) and he acted as the general chair for the same conference in 2006. He successfully advocated Singapore as the venue for VLDB 2010. His research group has taken an active part in regional conferences such as DASFAA, COMAD, ADC, APWeb, WAIM and NDBC, both as organizers and participants.

SIGMOD Test of Time Award

The ACM SIGMOD Test of Time Award recognizes the best paper from the SIGMOD proceedings 10 years prior (i.e., for 2009 the 1999 proceedings were consulted), based on the criterion of identifying the paper that has had the most impact (research, products, methodology) over the intervening decade. This paper is chosen by the SIGMOD Awards Committee.

2009 SIGMOD Test of Time Award

Approximate Computation of Multidimensional Aggregates of Sparse Data Using Wavelets
Jeffrey Scott Vitter (Texas A&M University) and Min Wang (IBM Research)

This influential paper showed that aggregates over sparse, high-dimensional arrays can be approximated with wavelets to give a compact data-cube representation that supports queries at interactive speeds. Prior histogram-based methods had prohibitive I/O costs for massive data sets of high dimensionality. The method is more accurate than random sampling and supports progressive refinement of answers when additional accuracy is desired. The paper inspired significant follow-on work by others, in the OLAP domain and also more broadly in approximate query processing, selectivity estimation, indexing of images and time series, and data-stream processing.

Abstract of the 1999 SIGMOD paper:

Computing multidimensional aggregates in high dimensions is a performance bottleneck for many OLAP applications. Obtaining the exact answer to an aggregation query can be prohibitively expensive in terms of time and/or storage space in a data warehouse environment. It is advantageous to have fast, approximate answers to OLAP aggregation queries. In this paper, we present a novel method that provides approximate answers to high-dimensional OLAP aggregation queries in massive sparse data sets in a time-efficient and space-efficient manner. We construct a compact data cube, which is an approximate and space-efficient representation of the underlying multidimensional array, based upon a multiresolution wavelet decomposition. In the on-line phase, each aggregation query can generally be answered using the compact data cube in one I/O or a small number of I/Os, depending upon the desired accuracy. We present two I/O-efficient algorithms to construct the compact data cube for the important case of sparse high-dimensional arrays, which often arise in practice. The traditional histogram methods are infeasible for the massive high-dimensional data sets in OLAP applications. Previously developed wavelet techniques are efficient only for dense data. Our on-line query processing algorithm is very fast and capable of refining answers as the user demands more accuracy. Experiments on real data show that our method provides significantly more accurate results for typical OLAP aggregation queries than other efficient approximation techniques such as random sampling.

SIGMOD Best Paper Award

2009 SIGMOD Best Paper Award

Generating Example Data for Dataflow Programs

Christopher Olston, Shubham Chopra, Utkarsh Srivastava (Yahoo! Research)

While developing data-centric programs, users often run (portions of) their programs over real data, to see how they behave and what the output looks like. Doing so makes it easier to formulate, understand and compose programs correctly, compared with examination of program logic alone. For large input data sets, these experimental runs can be time-consuming and inefficient. Unfortunately, sampling the input data does not always work well, because selective operations such as filter and join can lead to empty results over sampled inputs, and unless certain indexes are present there is no way to generate biased samples efficiently. Consequently new methods are needed for generating example input data for data-centric programs.

We focus on an important category of data-centric programs, *dataflow programs*, which are best illustrated by displaying the series of intermediate data tables that occur between each pair of operations. We introduce and study the problem of generating example intermediate data for dataflow programs, in a manner that illustrates the semantics of the operators while keeping the example data small. We identify two major obstacles that impede naive approaches, namely (1) highly selective operators and (2) noninvertible operators, and offer techniques for dealing with these obstacles. Our techniques perform well on real dataflow programs used at Yahoo! for web analytics.

<http://doi.acm.org/10.1145/1559845.1559873>

2009 SIGMOD Best Paper Award Runner-up

An Architecture for Recycling Intermediates in a Column-store

Milena G. Ivanova, Martin L. Kersten, Niels J. Nes, Romulo A.P. Goncalves (CWI)

Automatically recycling (intermediate) results is a grand challenge for state-of-the-art databases to improve both query response time and throughput. Tuples are loaded and streamed through a tuple-at-a-time processing pipeline avoiding materialization of intermediates as much as possible. This limits the opportunities for reuse of overlapping computations to DBA-defined materialized views and function/result cache tuning. In contrast, the operator-at-a-time execution paradigm produces fully materialized results in each step of the query plan. To avoid resource contention, these intermediates are evicted as soon as possible.

In this paper we study an architecture that harvests the by-products of the operator-at-a-time paradigm in a column store system using a lightweight mechanism, the *recycler*. The key challenge then becomes selection of the policies to admit intermediates to the resource pool, their retention period, and the eviction strategy when facing resource limitations. The proposed recycling architecture has been implemented in an open-source system. An experimental analysis against the TPC-H ad-hoc decision support benchmark and a complex, real-world application (SkyServer) demonstrates its effectiveness in terms of self-organizing behavior and its significant performance gains. The results indicate the potentials of recycling intermediates and charters a route for further development of database kernels.

<http://doi.acm.org/10.1145/1559845.1559879>

SIGMOD Jim Gray Doctoral Dissertation Award

SIGMOD has established the annual SIGMOD Jim Gray Doctoral Dissertation Award to recognize excellent research by doctoral candidates in the database field. This award, which was previously known as the SIGMOD Doctoral Dissertation Award, was renamed in 2008 with the unanimous approval of ACM Council in honor of Dr. Jim Gray.

2009 SIGMOD Jim Gray Doctoral Dissertation Award - Winner

- Daniel Abadi (advisor: Samuel Madden), MIT
Dissertation: *Query Execution in Column-Oriented Database Systems*.

This dissertation addresses all the key aspects of implementing an efficient database system based on a column-store architecture including storage management, compression strategies, query execution techniques, and alternative strategies for materializing result rows. The thesis is remarkable for its breadth, depth, system implementation and impact, and is an excellent treatise on the design and implementation of database systems employing a column-oriented storage model. The work is significant in bringing sound database design principles into practice and offering efficient solutions to urgent real problems.

2009 SIGMOD Jim Gray Doctoral Dissertation Award - Honorable Mentions:

- Bee-Chung Chen (advisor: Raghu Ramakrishnan), University of Wisconsin at Madison
Dissertation: *Cube-space Data Mining*
- Ashwin Machanavajjhala (advisor: Johannes Gehrke), Cornell University.
Dissertation: *Defining and Enforcing Privacy in Data Sharing*

The annual SIGMOD Jim Gray Dissertation Award recognizes excellent research by doctoral candidates in the database field. For the 2009 award a dissertation needed to be completed and accepted by the candidate's department between September 1, 2007 and December 15, 2008. This year 14 outstanding dissertations were nominated by their departments. The dissertations were evaluated by the SIGMOD Jim Gray Dissertation Awards Committee (Beng Chin Ooi (co-chair), Johannes Gehrke (co-chair---excused due to COI), Alfons Kemper, Hank Korth, Alberto Laender, Gerome Miklau, Timos Sellis, Kyu-Young Whang) for technical depth and significance of the research contribution, potential impact on theory and practice, and quality of the presentation. All submissions were of extremely high quality and we congratulate all the nominated students and their departments for their excellent work and contributions to our field.

2009 SIGMOD Undergraduate Awards

- Daniel Scott Brotherston (University of Waterloo) -- *Comparing and visualizing query optimizer search spaces*
- David M. Lewis (University of Maryland, Baltimore County) -- *An evaluative comparison of similarity coefficients for binary valued data*
- Sang-Phil Lim (Sung Kyun Kwan University) -- *FASTER FTL for enterprise-class flash memory SSDs*
- Manasi Vartak (WPI) -- *Recommendation based query relaxation via mapping functions and space partitioning*
- Shashank Yaduvanshi (IIT Bombay) -- *An architecture for regulatory compliant database management*

A complete listing of all SIGMOD Awards is available at: <http://www.sigmod.org/awards/>

2009 PODS Best Paper Award

Size and Treewidth Bounds for Conjunctive Queries

Georg Gottlob, Stephanie Tien Lee (Oxford University), and Gregory J. Valiant (University of California, Berkeley)

This paper provides new worst-case bounds for the size and treewidth of the result $Q(D)$ of a conjunctive query Q to a database D . We derive bounds for the result size $|Q(D)|$ in terms of structural properties of Q , both in the absence and in the presence of keys and functional dependencies. These bounds are based on a novel "coloring" of the query variables that associates a *coloring number* $C(Q)$ to each query Q . Using this coloring number, we derive tight bounds for the size of $Q(D)$ in case (i) no functional dependencies or keys are specified, and (ii) simple (one-attribute) keys are given. These results generalize recent size-bounds for join queries obtained by Atserias, Grohe, and Marx (FOCS 2008). An extension of our coloring technique also gives a lower bound for $|Q(D)|$ in the general setting of a query with arbitrary functional dependencies. Our new coloring scheme also allows us to precisely characterize (both in the absence of keys and with simple keys) the treewidth-preserving queries--the queries for which the output treewidth is bounded by a function of the input treewidth. Finally we characterize the queries that preserve the sparsity of the input in the general setting with arbitrary functional dependencies.

<http://doi.acm.org/10.1145/1559795.1559804>

2009 PODS Best Student Paper Award

XPath Evaluation in Linear Time with Polynomial Combined Complexity

Pawel Parys (Warsaw University)

We consider a fragment of XPath 1.0, where attribute and text values may be compared. We show that for any unary query in this fragment, the set of nodes that satisfy the query can be calculated in time linear in the document size and polynomial in the size of the query. The previous algorithm for this fragment also had linear data complexity but exponential complexity in the query size.

<http://doi.acm.org/10.1145/1559795.1559805>

ACM PODS Alberto O. Mendelzon Test-of-Time Award

In 2007, the PODS Executive Committee decided to establish a Test-of-Time Award, named after the late Alberto O. Mendelzon, in recognition of his scientific legacy, and his service and dedication to the database community. Mendelzon was an international leader in database theory, whose pioneering and fundamental work has inspired and influenced both database theoreticians and practitioners, and continues to be applied in a variety of advanced settings. He served the database community in many ways; in particular, he served as the General Chair of the PODS conference, and was instrumental in bringing together the PODS and SIGMOD conferences. He also was an outstanding educator, who guided the research of numerous doctoral students and postdoctoral fellows. The Award is to be awarded each year to a paper or a small number of papers published in the PODS proceedings ten years prior, that had the most impact (in terms of research, methodology, or transfer to practice) over the intervening decade. The decision was approved by SIGMOD and the ACM. The funds for the Award were contributed by IBM Toronto.

The PODS Executive Chair has appointed us to serve as the Award Committee for 2009. After careful consideration, we have decided to select the following paper as the award winner for 2009:

Georg Gottlob, Nicola Leone and Francesco Scarcello.
Hypertree Decompositions and Tractable Queries

The paper deals with a central problem in database research, namely finding classes of conjunctive queries for which problems, such as the evaluation of Boolean queries and query containment, are in polynomial time. This problem has attracted a lot of attention since the pioneering work of Yanakakis on acyclic queries. The paper shows that the earlier notion of bounded query width (introduced by Chekuri and Rajaraman in ICDT 97) is NP-hard, introduces the notion of bounded hypertree width, then shows that this notion properly generalizes earlier notions of acyclicity, that constant hypertree width is efficiently recognizable, and that Boolean queries with constant hypertree width can be efficiently evaluated. The results of the paper are applicable to both conjunctive query evaluation and to constraint satisfaction. The paper is extensively cited in the literature, and had an impact on subsequent research on these two problems. Hence, the committee has found it to be worthy of the Award.

Catriel Beeri (Chair), Phokion G. Kolaitis, Christos H. Papadimitriou
The Alberto O. Mendelzon Test-of-Time Award Committee for 2009



Georg Gottlob is a Professor of Computing Science at Oxford University, UK, and an Adjunct Professor of Computer Science at the Vienna University of Technology (TU Wien). His research interests are in database theory (in particular, query languages), Web information processing, AI, and computational logic. Gottlob got his Ph.D. degree in Computer Science from TU Vienna, Austria in 1979 and 1981, respectively. Before he moved to Oxford in 2006, he was a Professor of Computer Science at TU (since 1988). Before that, he was affiliated with the Italian National Research Council in Genoa, Italy. He also was a Research Associate and lecturer at the Politecnico di Milano, Stanford University and held visiting positions at Paris VII and at Berkeley. Georg has received the Wittgenstein Award from the Austrian National Science Fund. He is an ACM and an ECCAI Fellow, and a member of the Austrian Academy of Sciences, the German National Academy of Sciences Leopoldina, and the European Academy of Sciences Academia Europaea in London. He chaired the Program Committees of IJCAI 2003 and ACM PODS 2000. He has co-founded the Lixto software company (www.lixto.com) which offers software and services for Web data extraction.



Nicola Leone is full professor of Computer Science, Head of the Department of Mathematics, and Director of the PHD School ISIMR at University of Calabria. He has previously held the professor position for Database Systems at Vienna University of Technology in Austria from 1995 to 2000. Before that, he was with the Italian National Research Council in Cosenza, Italy. His research interests include Database Theory, Artificial Intelligence, and Computational Logics. He is the leader of the team which developed DLV -- the state-of-the-art Disjunctive-Datalog system. He is the author of over 200 scientific papers. He was the program chair of a number of conferences including LPNMR-99, JELIA-02, AGP-03, and LPNMR-05. He was the invited speaker of LPAR'06, RR'07, LPNMR'07, and ICLP'08, and he is the President of the Steering Committee of LPNMR. He has recently co-founded three spin-off companies, namely, Exeura, Artemat, and DLVSYSTEM, working on Databases, AI applications, and Knowledge Management.



Francesco Scarcello is a professor of computer science at the University of Calabria. His research interests include computational complexity, database theory, constraint satisfaction, graph theory, game theory, knowledge representation, and non-monotonic reasoning. He has extensively published in all these areas in leading conferences and journals, including the Journal of the ACM, Journal of Computer and System Sciences, Information and Computation, and Artificial Intelligence. He holds a Ph.D. in Computer Science from the University of Calabria. He was recipient of two grants from the Italian National Research Council (CNR) until 1999, when he became an assistant professor at the University of Calabria. He holds his current position of associate professor since 2001. Also, during these years, he has been visiting the Department of Information Systems at the Vienna University of Technology. He serves on program committees and as a reviewer for many international conferences and journals. He is co-recipient of the 2008 IJCAI-JAIR Best Paper Prize, awarded to an outstanding paper published in the Journal of Artificial Intelligence Research in the preceding five calendar years.

A complete listing of all PODS Awards is available at: <http://www.sigmod.org/pods/>