

SIGMOD/PODS 2010 Awards

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.



Dr. Umeshwar Dayal is the recipient of the 2010 SIGMOD *Edgar F. Codd Innovation Award* for a succession of pioneering, influential contributions in distributed heterogeneous databases, high-performance active databases, generalized transitive closure, transaction models for long-running activities, and business process discovery, among other topics.

Umeshwar Dayal has a track record of 30 years of research accomplishments in data management and has made a succession of fundamental contributions to the field. His research in the mid-1980's on Multibase, the world's first large-scale heterogeneous distributed database system, paved the way for research over the past two decades at universities and industrial research labs (HP Labs, IBM Research, and elsewhere) in data integration and federated databases. PROBE, one of the first extensible database management systems, made significant contributions to the field in several areas, including object and semantic data modeling, spatial, temporal and recursive query processing, and system architecture. HiPAC was one of the leading active database systems of its time, and the only one to focus on the needs of real-time applications, resulting in an innovative transaction model. The event-condition-action (ECA) rule model introduced in HiPAC has now been widely adopted in reactive computing systems, complex-event-processing systems, and distributed middleware.

Dayal also has significant results in query-processing research, with particularly strong contributions to the processing of multi-database queries, spatial queries, and recursive queries. The approach described in his VLDB 1987 paper on unnesting SQL queries was later adopted in at least five commercial products. In addition, Dayal performed pioneering work in long-duration transactions, business-process management, and database design. In particular, he pioneered the field of business-process intelligence, which combines data warehousing, data mining, analytics and optimization techniques to monitor, control, analyze, and optimize business processes. Over 160 research papers and over 25 patents testify to Dayal's innovation and productivity.

In 2001, he received the prestigious 10-year best paper award from VLDB for his paper on a transactional model for long-running activities. He is an HP Fellow, recognized for career contributions that “caused substantial change” in the state of the art while also improving HP products. In addition to his many innovative technical contributions, he has a distinguished record of service to our community, as a member of the editorial board of major journals (including ACM TODS and VLDB), chairing conferences (including SIGMOD, VLDB and ICDE), serving on boards and steering committees (VLDB, IEEE TC on Electronic Commerce, SIAM International Conference on Data Mining among them), and mentoring junior colleagues and young researchers.

SIGMOD Contributions Award for outstanding and sustained services to and promotion of the database field through activities such as education, conference organization, journals, standards, and research funding.



Dr. David Lomet is the recipient of the 2010 SIGMOD Contribution Award for his outstanding leadership as the Editor-in-Chief of the IEEE Data Engineering Bulletin, a key forum for dissemination of emerging ideas in academia and industry. Lomet has been a key figure in our field, holding many additional leadership roles and demonstrating in each his dedication to service and to our community.

By awarding David Lomet the ACM Contributions Award, we recognize his outstanding contributions to our community in leading the IEEE Data Engineering Bulletin for nearly 20 years and thereby creating a collection of timely articles of great value. He has almost single-handedly driven the IEEE Data Engineering Bulletin, providing our community with a constant stream of special issues, assembled by world-class invited guest editors. This service has been a wonderful benefit to the field, as each issue has provided a "root node" into key projects, both academic and industrial, and into the research literature related to the topic of the issue. These articles and issues have thus provided a "fast path" to see what's happening in an area as well as a way to make sure industrial highlights are brought to the attention of academics and vice versa. Not only has Lomet run the Bulletin, but he also negotiated with IEEE Computer Society and authors to make issues of the Bulletin available on CDROM via the SIGMOD DiSC, and later, to digitize the entire set of issues from 1977 on so that they can be available online to all.

Lomet has made significant contributions to our field through service in other roles as well. He was a Co-PC Chair of VLDB and he is currently on the VLDB Board of Trustees. He has been both a Co-PC Chair and a Co-General Chair of IEEE Data Engineering Conferences, and served as a member of the Steering Committee of the IEEE Technical Committee on Data Engineering from 2004-2009. He sets a high standard of service for our community.

SIGMOD Test-of-Time Award for the paper from the 2000 SIGMOD Conference that has had the most impact (research, products, methodology) over the intervening decade.

NiagaraCQ: A Scalable Continuous Query System for Internet Databases
(<http://portal.acm.org/citation.cfm?doid=342009.335432>)

Jianjun Chen, David J. DeWitt (University of Wisconsin, now Microsoft), Feng Tian (University of Wisconsin, now VMware), Yuan Wang (University of Wisconsin, now Microsoft)

This paper from the SIGMOD 2000 Conference bridged from the world of continuous, or standing, queries against a changing stored database, to stream processing systems. NiagaraCQ was a pioneering system, the first to address the problem of the millions of overlapping queries that would need to be supported in a truly internet-scale system. It used relational-style operators to optimize a given set of continuous queries. Similar frameworks appeared in subsequent studies of stream databases, sensor databases, information delivery systems, and complex-event-processing (CEP) systems. The idea of dynamic optimization of continuous queries leveraging database operators (including dynamic query grouping and split) became a baseline for modern streaming data platforms. In summary, this paper helped open the new field of high-performance systems for continuous query processing, and was a strong force in shaping the following generations of stream processing systems.

Abstract: Continuous queries are persistent queries that allow users to receive new results when they become available. While continuous query systems can transform a passive web into an active environment, they need to be able to support millions of queries due to the scale of the Internet. No existing systems have achieved this level of scalability. NiagaraCQ addresses this problem by grouping continuous queries based on the observation that many

web queries share similar structures. Grouped queries can share the common computation, tend to fit in memory and can reduce the I/O cost significantly. Furthermore, grouping on selection predicates can eliminate a large number of unnecessary query invocations. Our grouping technique is distinguished from previous group optimization approaches in the following ways. First, we use an incremental group optimization strategy with dynamic regrouping. New queries are added to existing query groups, without having to regroup already installed queries. Second, we use a query-split scheme that requires minimal changes to a general-purpose query engine. Third, NiagaraCQ groups both change-based and timer-based queries in a uniform way. To insure that NiagaraCQ is scalable, we have also employed other techniques including incremental evaluation of continuous queries, use of both pull and push models for detecting heterogeneous data source changes, and memory caching. This paper presents the design of NiagaraCQ system and gives some experimental results on the system's performance and scalability.

SIGMOD Best Paper Award

FAST: Fast Architecture Sensitive Tree Search on Modern CPUs and GPUs
(<http://portal.acm.org/citation.cfm?doid=1807167.1807206>)

Changkyu Kim, Jatin Chhugani, Nadathur Satish (Intel), Eric Sedlar (Oracle), Anthony Nguyen (Intel), Tim Kaldewey (Oracle), Victor Lee (Intel), Scott Brandt (University of California, Santa Cruz), Pradeep Dubey (Intel)

This paper presents FAST, a layout for an in-memory binary tree index that is well-suited for state-of-the-art CPU and GPU architectures. The layout and associated search methods take advantage of SIMD instructions and thread-level parallelism (TLP). FAST also accounts for cache-line sizes and hides cache-miss and TLB-miss latency by processing many outstanding queries simultaneously (with software pipelining and TLP). The paper shows that with all these optimizations, search on GPU is compute bound and search on a CPU is bandwidth bound. To optimize the latter further, the paper presents a key-compression scheme, which also takes advantage of SIMD instructions, to alleviate bandwidth limits and handle larger keys. Experiments show how CPU and GPU perform on trees with different sizes, how many concurrent queries are needed to achieve their peak throughput, and how compression can improve search performance. This paper is an excellent research contribution that provides an end-to-end system design and associated algorithms and techniques to develop a complete solution that leverages the underlying hardware architecture. Given the modular structure of the overall design, the solution can easily be adapted to future architectures.

PODS Best Paper Award

An Optimal Algorithm for the Distinct Elements Problem
(<http://portal.acm.org/citation.cfm?doid=1807085.1807094>)

Daniel M. Kane, Jelani Nelson and David P. Woodruff

Abstract: We give the first optimal algorithm for estimating the number of distinct elements in a data stream, closing a long line of theoretical research on this problem begun by Flajolet and Martin in their seminal paper in FOCS 1983. This problem has applications to query optimization, Internet routing, network topology, and data mining. For a stream of indices in $\{1, \dots, n\}$, our algorithm computes a $(1 \pm \epsilon)$ -approximation using an optimal $O(\epsilon^{-2} + \log(n))$ bits of space with $2/3$ success probability, where $0 < \epsilon < 1$ is given. This probability can be amplified by independent repetition. Furthermore, our algorithm processes each stream update in $O(1)$ worst-case time, and can report an estimate at any point midstream in $O(1)$ worst-case time, thus settling both the space and time complexities

simultaneously. We also give an algorithm to estimate the Hamming norm of a stream, a generalization of the number of distinct elements, which is useful in data cleaning, packet tracing, and database auditing. Our algorithm uses nearly optimal space, and has optimal $O(I)$ update and reporting times.

The ACM PODS Alberto O. Mendelzon Test-of-Time Award 2010

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.

After careful consideration, the Award Committee for 2010 has decided to select the following papers as the award winners for 2010:

- **Typechecking for XML Transformers**, Tova Milo, Dan Suciu, and Victor Vianu (<http://doi.acm.org/10.1145/335168.335171>)

Abstract: We study the typechecking problem for XML transformers: given an XML transformation program and a DTD for the input XML documents, check whether every result of the program conforms to a specified output DTD. We model XML transformers using a novel device called a k-pebble transducer, that can express most queries without data-value joins in XML-QL, XSLT, and other XML query languages. Types are modeled by regular tree languages, a robust extension of DTDs. The main result of the paper is that typechecking for k-pebble transducers is decidable. Consequently, typechecking can be performed for a broad range of XML transformation languages, including XML-QL and a fragment of XSLT.

- **Integrity Constraints for XML**, Wenfei Fan and Jérôme Siméon (<http://doi.acm.org/10.1145/335168.335172>)

Abstract: Integrity constraints are useful for semantic specification, query optimization and data integration. The ID/IDREF mechanism provided by XML DTDs relies on a simple form of constraint to describe references. Yet, this mechanism is not sufficient to express semantic constraints, such as keys or inverse relationships, or stronger, object-style references. In this paper, we investigate integrity constraints for XML, both for semantic purposes and to improve its current reference mechanism. We extend DTDs with several families of constraints, including key, foreign key, inverse constraints and constraints specifying the semantics of object identities. These constraints are useful both for native XML documents and to preserve the semantics of data originating in relational or object databases. Complexity and axiomatization results are established for the (finite) implication problems associated with these constraints. These results also extend relational dependency theory on the interaction between (primary) keys and foreign keys. In addition, we investigate implication of more general constraints, such as functional, inclusion and inverse constraints defined in terms of navigation paths.

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.

The 2010 recipient is **Chris Re** (University of Washington, advised by Dan Suciu) for his thesis titled “*Managing Large-scale Probabilistic Databases*”.

The two runners-up are Soumyadeb Mitra, (University of Indiana at Urbana Champaign, advised by Marianne Winslett), for his thesis titled “*Trustworthy and Cost Effective Management of Compliance Records*”, and Fabian Suchanek (Max-Planck Institut für Informatik, advised by Gerhard Weikum) for his thesis titled “*Automated Construction and Growth of a Large Ontology*”

SIGMOD undergraduate posters

SIGMOD conference has started a new initiative that provides undergraduate students an opportunity to showcase their research accomplishments in a poster competition. The recipients for 2010 are:

1. **Zhenqiang Gong**, University of Science and Technology of China (USTC) *Parallel Algorithms for Top-k Query Processing*
2. **Yael Amsterdamer**, Tel Aviv University *Top-k Algorithms for Interactive Processes*
3. **Ian Charlesworth**, University of Waterloo *Analyzing Plan Spaces of Query Optimizers*
4. **Bharath Vissapragada**, International Institute of Information Technology, Hyderabad (IIIT-H) *Query Optimization on Cloud*
5. **Thodoris Georgiou**, National and Kapodistrian University of Athens *Extracting Topics of Debate between Users on Web Discussion Boards*
6. **Manos Karvounis**, National and Kapodistrian University of Athens *Utilizing the Quoting System of Online Web Forums to Estimate User Agreement*

SIGMOD best undergraduate posters: Thodoris Georgiou and Manos Karvounis, National and Kapodistrian University of Athens

The **2010 SIGMOD programming contest**, as well as and the **best SIGMOD 2010 demonstration award**, are detailed in the respective events report of this SIGMOD Record issue.