Stonebraker Receives IEEE John von Neumann Medal

David DeWitt, Michael Carey, and Joseph M. Hellerstein

In December 2004, Michael Stonebraker was selected to receive the 2005 IEEE John von Neumann Medal for his "*contributions to the design, implementation, and commercialization of relational and object-relational database systems."* Mike is the first person from the database field selected to receive this award. He joins an illustrious group of former recipients, including Barbara Liskov (2004), Alfred Aho (2003), Ole-Johan Dahl and Kristen Nygaard (2002), Butler Lampson (2001), John Hennessy and David Patterson (2000), Douglas Engelbart (1999), Ivan Sutherland (1998), Maurice Wilkes (1997), Carver Mead (1996), Donald Knuth (1995), John Cocke (1994), Fred Brooks (1993), and Gordon Bell (1992).

Mike has had, and continues to have, a profound impact on the database field. The relational data model and its associated benefits of "data independence" and "non-procedural access" were first invented by Tedd Codd. However, more than any other individual, Mike is responsible for making Codd's vision of independence a reality through the architectures and algorithms embodied in the series of open-source prototypes and commercial systems that he has initiated and led. While many others have certainly made important contributions to the field, no one else comes close to his continuous sequence of landmark innovations over a period of almost 30 years. In 1992, Mike received the SIGMOD Innovations Award the very first time it was given.

Mike has been the primary driving force behind major shifts in the research agenda of the database community, including two occasions where he launched entire new directions for the field to follow (the implementation of relational systems and object-relational systems). His *modus operandi* has been fairly uniform across his career: he declares a research target; he rallies the research community in that direction via workshops, position papers and talks; he builds an open-source prototype with his students and publishes papers in the open literature; he then transfers the technology directly to practice via a startup company. Some people catch on to his ideas early on in this process, some catch on after the research begins to appear, and some are only convinced by the commercial success of the resulting company – and/or of the competitors that emerge based on his vision. Stonebraker's approach is a compelling model for end-to-end innovation, involving community-building, mentoring, technical publication, open-source development, and the "last mile" of direct technology transfer via commercialization.

As is well known, Stonebraker's earliest contributions occurred as the leader of the INGRES project at Berkeley, which he later transferred to industry via his startup company, RTI. The INGRES project gambled on the possibility that the paper proposals of relational theory could be realized in a high-performance software system for data management. At the time, serious database systems were based on the "network" or "hierarchical" models, which required users to programmatically access data via pointers and custom logic. The challenge of realizing Codd's vision in INGRES was truly grand: it required automatic techniques that could compete with the commercial database systems and IT programmers of the day. The impact has been enormous. Every database system today implicitly works within extensions of the relational system frameworks realized by INGRES and its competitor, IBM's System R. Stonebraker's personal dual thrust on both systems engineering and model/language design in INGRES set the standard for a holistic view of database research, moving the community beyond its prior dichotomy of systems engineers and modeling conceptualizers. As we are all aware, every large organization today depends upon relational databases to manage mission-critical data, and the relational database industry accounts for multiple billions of dollars of business each year. The INGRES research project and its open-source software led directly to the development of a number of commercial products that remain at the core of modern systems, including Sybase SQL Server, Microsoft SQL Server, and the commercial version of INGRES. INGRES and System R shared the 6th ACM Software Systems Award in 1988. While RTI's marketing turned out to not be a match for Oracle's, the competition between RTI and Oracle drove the entire industry forward at a rapid pace.

Subsequent to his work on INGRES, Stonebraker continued to lead the database field with the development of object-relational databases, exemplified by the POSTGRES research project. This was an effort to marry data independence to a rich, extensible data model, an idea first espoused by Mike as part the of the ADT-INGRES project. When Mike started the POSTGRES project, most of the database field was exploring object-oriented data

models based on the addition of persistence and transactions to object-oriented programming languages such as C++. These models lacked a declarative query language and the attendant benefits of data independence. In contrast, Mike realized that many of the benefits that an object-oriented data model might provide could be achieved without giving up on the key notions of declarative query languages and data independence. In the object-relational case, POSTGRES led the charge both in conceptual terms (data model and query language) and in terms of the system architecture to enable declarative queries to be automatically optimized and efficiently executed in this semantically rich environment. POSTGRES extended the relational model so that users could define and store rich objects with methods and rules in the database system and invoke them from declarative queries. Its extensible type architecture enabled POSTGRES to optimize these queries automatically, and it supported indexes for efficiently retrieving these data types. The object-relational extensions pioneered by POSTGRES allowed database systems to provide significantly enhanced intelligence and efficiency to both business-oriented and scientific applications.

Stonebraker commercialized these ideas in his next startup company, Illustra (subsequently purchased by Informix, and now owned by IBM). The result was that the relational database industry quickly introduced object-relational research ideas into their systems. Today, the database systems from all of the major commercial vendors (including Oracle, DB2, and SQL Server) support the kinds of functionality first pioneered in POSTGRES. And the PostgreSQL open source community continues to drive the POSTGRES architecture and code base forward.

In addition to these major thrusts, Mike has made numerous technical contributions to many areas of the database field, including parallel database systems, distributed database systems, query processing, indexing and query processing techniques for geographic data management, data visualization, and storage systems including tertiary storage mechanisms and history-preserving transactional storage. MUFFIN, his late 1970s parallel database system project, was the first system to exploit a “shared-nothing” architecture, a term coined subsequently by Mike along with “shared-memory” and “shared-disk” to characterize the major alternative ways of architecting a parallel database system.

Soon after completing the initial versions of INGRES, Mike turned his attention to the design and implementation of distributed database systems, initially via the distributed database project INGRES* and then later via Mariposa. While INGRES* was never a commercial success (for mostly non-technical reasons), the project contributed many key technologies to the field, including the concept of horizontal partitioning of tables (later adopted by the parallel database field) and a variety of distributed query processing and optimization techniques. The technical challenges posed by such a project in the early 1980s are easily forgotten, but they were significant. For example, while the TCP/IP protocol stack was deployed as part of the ARPANet at the time, it was not yet part of any Unix distribution. In addition, local area networks were just beginning to become commercially available, and they were unreliable and hard to use at that time. Thus, building a distributed database system in the early 1980s required the builders to first design and implement a complete networking stack (!).

Mariposa, 10-15 years later, was Mike’s other major foray into distributed database systems. Mariposa pioneered the use of economic computing paradigms in federated database systems. Stated in classical relational terms, the goal of Mariposa was to extend the benefits of data independence across both geographic and administrative boundaries, so that declarative queries could be specified without concern for the location or management of machines and data. This required a significant leap of faith into new design territory. The key components of Mariposa were query optimization and data placement schemes that used notions of bidding and contracting for work, decoupling global optimization decisions from local administrative policies enforced at the individual sites. Mike’s next startup company, Cohera, commercialized the Mariposa research. Cohera’s technology was purchased by PeopleSoft for their Catalog Management solution. Although PeopleSoft never aligned their core architecture around a federated database model as had been planned, Mariposa remains another groundbreaking Stonebraker effort that took a major new conceptual idea and realized it in the form of a full-featured, practical software architecture.

In the last five years, Stonebraker has established himself in New England as a research community builder, research team leader, and East Coast entrepreneur. Using MIT as his base, he helped unify the New England Database Society (NEDS), and coordinated a series of research projects on streaming databases spanning MIT, Brown and Brandeis University. These projects – including Aurora, Medusa and Borealis – are important pieces of this active research ecosystem. Moreover, Stonebraker and his team are busy with a new startup called StreamBase, which

prompted Forbes magazine recently to once again ask the age-old question about a new Stonebraker startup: "Should IBM and Oracle worry?"

Without the INGRES project, relational database systems might never have been a commercial success. The object-relational ideas of the POSTGRES project provided the next major shift in direction for the field. Mariposa, with its radical economic model for query processing in a federated database environment, may someday prove to be the right approach for doing distributed query processing on the Internet. Stream query processors like Aurora may well change the way people conceive of data processing. Each of Mike's many contributions to the database field, whether it proved to be a commercial or technical success or not, demonstrated a level of creativity that has served to truly inspire the rest of our field. We are very proud of Mike's accomplishments, and we are proud that he is the first recipient of the John von Neumann medal from the database field. Please join us in congratulating Mike on this well-deserved honor!