Traditional Database Management Systems (DBMS) software is built on the concept of persistent data sets, that are stored reliably in stable storage and queried several times throughout their lifetime. For several emerging application domains, however, data arrives and needs to be processed continuously, without the benefit of several passes over a static, persistent data image. Potential applications for data stream management range from network monitoring, sensor networks, financial data, to the real-time enterprise.

In January 2003, a group of researchers met at Stanford University for an informal workshop to discuss ongoing and open research questions in query processing for streaming data, and to ponder about the killer application for a data stream management system. While this special issue is not an official report from the workshop, at least one of the authors of each invited papers were present at the workshop, and the selection of papers mirrors some of the exciting work that was presented at the meeting. I encourage the interested reader to visit `http://telegraph.cs.berkeley.edu/swim/` for more information.

Research in data stream processing spans many communities with different research foci. This issue concentrates on some of the ongoing work in the database community. We start this issue with overview articles from four groups that are currently building prototype data stream processing systems. Stan Zdonik, Michael Stonebraker, Mitch Cherniack, Ugur C etintemel, Magdalena Balazinska, and Hari Balakrishnan survey the Aurora and Medusa Projects at MIT, Brandeis University, and Brown University. Sailesh Krishnamurthy, Sirish Chandrasekaran, Owen Cooper, Amol Deshpande, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, Samuel R. Madden, Fred Reiss, and Mehul Shah give us an update of the latest status of the TelegraphCQ Project at Berkeley. The Stanford STREAM Group — at the writing of this article consisting of Arvind Arasu, Brian Babcock, Shivnath Babu, Mayur Datar, Keith Ito, Rajeev Motwani, Itaru Nishizawa, Utkarsh Srivastava, Dilys Thomas, Rohit Varma, and Jennifer Widom — describes the status of their system and ongoing research. Chuck Cranor, Theodore Johnson, Oliver Spatscheck, and Vladislav Shkapenyuk from AT&T Research give an industry perspective; they survey challenges in the design and implementation of the Gigascope stream database system specialized for network monitoring.

The issue continues with two articles that cover algorithmic issue in data stream processing. Peter Tucker, David Maier, and Tim Sheard describe how punctuations — predicates that hold about a data stream — can be used to improve data stream queries. Yanlei Diao and Michael Franklin describe an XML publish-subscribe query processor that matches streaming XML data with a large number of continuous queries. We conclude this issue with an article by Alberto Lerner and Dennis Shasha which discusses a novel architecture of a data stream system based for data from the financial industry on integrating online stream processing with support for historical queries over a data archive.

I hope that you find this special issue as enjoying to read as I did, and I thank the authors for their contributions.