MIPS 2004

Tutorial: Data Stream Management Systems (DSMS)
- Applications, Concepts, and Systems -

Appendix: Literature References for DSMS

Vera Goebel & Thomas Plagemann
Department of Informatics
University of Oslo

DSMS - Surveys

- http://db.uwaterloo.ca/~ddbms/publications/stream/streamsurvey.pdf
- http://www.bell-labs.com/user/rastogi/tutorials.html
- http://www-db.stanford.edu/stream/
- http://www.research.att.com/conf/mpds2003/schedule/
- B. Babcock, S. Babu, M. Datar, R. Motawani, and J. Widom, "Models and issues in data stream systems", ACM PODS 2002 (tutorial).
- M. Garofalakis, J. Gehrke, and R. Rastogi, "Querying and mining data streams: You only get one look", ACM SIGMOD 2002 (tutorial).
- Golab & Ozsu, Issues in Data Stream Management. ACM SIGMOD Record, June 2003.

DSMS - Applications

- S. Babu, L. Subramanian, and J. Widom. A Data Stream Management System for Network Traffic Management, *Proc. of the Workshop on Network-Related Data Management (NRDM 2001), May 2001*
- Cranor, Johnson, Spatscheck & Shkapenyuk. Gigascope: A Stream Database for Network Applications. SIGMOD 2003.
- Joseph M. Hellerstein. **From Database to Dataflow: New Directions in IT**. *Medical Records Institute Health IT Advisory Report* 3(6) (2002).
- Lerner & Shasha. The Virtues and Challenges of Ad Hoc + Streams Querying in Finance. IEEE Data Engineering Bulletin, March 2003.
- Oracle 9i Stream
- Sistal, Wolfson, Chamberlain, Dao. Modeling and Querying Moving Objects. ICDE 1997.
- Yao & Gehrke. Query Processing for Sensor Networks. CIDR 2003.

DSMS - Systems

- **Aurora** (Brandeis, Brown, MIT)
  - H. Balakrishnan, M. Balazinska, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, E. Galvez, J. Salz, M. Stonebraker, N. Tatbul, R. Tibbetts, S. Zdonik,

Retrospective on Aurora, *VLDB Journal Special Issue on Data Stream Processing*, 2004 (to appear).

- J. Hwang, M. Balazinska, A. Rasin, U. Çetintemel, M. Stonebraker, S. Zdonik. **A Comparison of Stream-Oriented High-Availability Algorithms**, Brown CS TR-03-17 (submitted).
- D. Carney, U. Çetintemel, A. Rasin, S. Zdonik, M. Cherniack, M. Stonebraker. **Operator Scheduling in a Data Stream Manager**. In proceedings of the *29th International Conference on Very Large Data Bases* (*VLDB*'03), Berlin, Germany, September 2003.
- N. Tatbul, U. Çetintemel, S. Zdonik, M. Cherniack, M. Stonebraker. **Load Shedding in a Data Stream Manager**. In proceedings of the *29th International Conference on Very Large Data Bases* (*VLDB*'03), Berlin, Germany, September 2003, pp. 309-320.
- D. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, S. Zdonik. **Aurora: A New Model and Architecture for Data Stream Management**. In *VLDB Journal (12)2: 120-139*, August 2003.
- D. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, C. Erwin, E. Galvez, M. Hatoun, J. Hwang, A. Maskey, A. Rasin, A. Singer, M. Stonebraker, N. Tatbul, Y. Xing, R.Yan, S. Zdonik. **Aurora: A Data Stream Management System (Demonstration)**. In proceedings of the *ACM SIGMOD International Conference on Management of Data* (*SIGMOD*'03), San Diego, CA, June 2003.
- D. Carney, U. Çetintemel, A. Rasin, S. Zdonik, M. Cherniack, M. Stonebraker. **Reducing Execution Overhead in a Data Stream Manager**. In proceedings of the *ACM Workshop on Management and Processing of Data Streams* (*MPDS*'03), San Diego, CA, June 2003.
- N. Tatbul, U. Çetintemel, S. Zdonik, M. Cherniack, M. Stonebraker. **Load Shedding on Data Streams**. In proceedings of the *ACM Workshop on Management and Processing of Data Streams* (*MPDS*'03), San Diego, CA, June 2003.
- S. Zdonik, M. Stonebraker, M. Cherniack, U. Cetintemel, M. Balazinska, and H. Balakrishnan. **The Aurora and Medusa Projects**. IEEE Data Engineering Bulletin, Vol. 26, No. 1, March, 2003.
- M. Cherniack, H. Balakrishnan, M. Balazinska, D. Carney, U. Cetintemel, Y. Xing, S. Zdonik. **Scalable Distributed Stream Processing.** In proceedings of the *First Biennial Conference on Innovative Database Systems* (*CIDR*'03), Asilomar, CA, January 2003.
- D. Carney, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, G. Seidman, M. Stonebraker, N. Tatbul, and S. Zdonik. **Monitoring Streams: A New Class of Data Management Applications**. In proceedings of the *28th International Conference on Very Large Data Bases* (*VLDB'*02), Hong Kong, China, August 2002.

- **COUGAR** (Cornell)
  - Niki Trigoni, Yong Yao, Alan Demers, Johannes Gehrke, Rajmohan Rajaraman. WaveScheduling: Energy-Efficient Data Dissemination for Sensor Networks. *Internet Draft 2004*

- Yong Yao, J. E. Gehrke. Query Processing in Sensor Networks. In *Proceedings of the First Biennial Conference on Innovative Data Systems Research (CIDR 2003).* Asilomar, California, January 2003.
- Yong Yao, J. E. Gehrke. The Cougar Approach to In-Network Query Processing in Sensor Networks. Sigmod Record, Volume 31, Number 3. September 2002.
- Anton Faradjian, J. E. Gehrke, and Philippe Bonnet. GADT: A Probability Space ADT For Representing and Querying the Physical World. *Proceedings of the 18th International Conference on Data Engineering.* San Jose, California, February 2002.
- Philippe Bonnet, J. E. Gehrke, and Praveen Seshadri. Towards Sensor Database Systems. *Proceedings of the Second International Conference on Mobile Data Management.* Hong Kong, January 2001.
- Philippe Bonnet, J. E. Gehrke, and Praveen Seshadri. Querying the Physical World. *IEEE Personal Communications*, Vol. 7, No. 5, October 2000, pages 10-15. Special Issue on Smart Spaces and Environments.
- Philippe Bonnet, Praveen Seshadri. Device Database Systems. In *Proceedings of the 16th International Conference on Data Engineering*. San Diego, California, February 2000.

- **GigaScope** and **Hancock** (At&T)
  - C. D. Cranor, T. Johnson, O. Spatscheck, V. Shkapenyuk: Gigascope: A Stream Database for Network Applications. SIGMOD Conference 2003: 647-651
  - C. D. Cranor, T. Johnson, O. Spatscheck, V. Shkapenyuk: The Gigascope Stream Database. IEEE Data Engineering Bulletin 26(1): 27-32 (2003)

- **Medusa** (Brown University)
  - High-Availability Algorithms for Distributed Stream Processing
    Jeong-Hyon Hwang, Magdalena Balazinska, Alexander Rasin, Ugur Cetintemel, Michael Stonebraker, and Stan Zdonik
    The 21st International Conference on Data Engineering. ICDE 2005 (to appear)

  - Contract-Based Load Management in Federated Distributed Systems
    Magdalena Balazinska, Hari Balakrishnan, and Mike Stonebraker
    1st Symposium on Networked Systems Design and Implementation (NSDI)
    San Francisco, CA, March 2004.

  - A Comparison of Stream-Oriented High-Availability Algorithms
    Jeong-Hyon Hwang, Magdalena Balazinska, Alexander Rasin, Ugur Cetintemel, Michael Stonebraker, and Stan Zdonik
    Technical Report CS-03-17. Brown University, September 2003.

  - The Aurora and Medusa Projects
    Stan Zdonik, Michael Stonebraker, Mitch Cherniack, Ugur Cetintemel, Magdalena Balazinska, and Hari Balakrishnan
    Bulletin of the Technical Committe on Data Engineering
    IEEE Computer Society. March 2003. p.3-10. (Invited Paper).

- Scalable Distributed Stream Processing
  Mitch Cherniack, Hari Balakrishnan, Magdalena Balazinska,
  Don Carney, Ugur Cetintemel, Ying Xing, and Stan Zdonik
  CIDR 2003 - First Biennial Conference on Innovative Data Systems Research,
  Asilomar, California, January 2003.
- Richard Tibbetts
  Linear Road: Benchmarking Stream-Based Data Management Systems
  M. Eng. Thesis, Massachusetts Institute of Technology, August 2003.
  [Postscript (1.3 MB)] [.ps.gz (563 KB)] [PDF (631 KB)](61 pages)

- Hiroyoshi Iwashima
  Differential Bandwidth Allocation with Multiplexed TCP Connections
  M. Eng. Thesis, Massachusetts Institute of Technology, August 2003.
  [Postscript (1 MB)] [.ps.gz (250 KB)] [PDF (345 KB)](66 pages)

- **NiagaraCQ** (University of Wisconsin-Madison)
  - J. Chen, D. J. DeWitt, F. Tian, Y. Wang: NiagaraCQ: A Scalable Continuous Query System for Internet Databases. SIGMOD Conference 2000: 379-390.
  - **NiagaraCQ: A Scalable Continuous Query System for Internet Databases**. Jianjun Chen, David J. DeWitt, Feng Tian, Yuan Wang. *SIGMOD 2000* , p379-390.
  - **Active Query Caching for Database Web Servers**. Qiong Luo, Jeffrey F. Naughton, Rajasekar Krishnamurthy, Pei Cao, and Yunrui Li. *WebDB'2000.*
  - **The Niagara Internet Query System**. Jeffrey Naughton, David DeWitt, David Maier, and many others.
  - **Architecting a Network Query Engine for Producing Partial Results**. Jayavel Shanmugasundaram, Kristin Tufte, David J. DeWitt, Jeffrey F. Naughton, and David Maier. *WebDB'2000*
  - **Rate-Based Query Optimization for Streaming Information Sources** Stratis D. Viglas, Jeffrey F. Naughton *SIGMOD 2002*
  - **Design and Evaluation of Alternative Selection Placement Strategies in Optimizing Continuous Queries**. Jianjun Chen, David J. DeWitt, and Jeffrey F. Naughton. *ICDE02*
  - **Building XML Statistics for the Hidden Web**. Ashraf Aboulnaga and Jeffrey F. Naughton.
  - **Evaluating Window Joins Over Unbounded Streams**. Jaewoo Kang, Jeffrey F. Naughton and Stratis D. Viglas.
  - **Maximizing the Output Rate of Multi-Join Queries over Streaming Information Sources**. Stratis D. Viglas, Jeffrey F. Naughton, Josef Burger.
  - **STEP: Extending Relational Query Engines for Efficient XML Query Processing**. Feng Tian, David J. DeWitt.
  - **Optimizing Fixed-Schema XML to SQL Query Translation**. Rajasekar Krishnamurthy, Raghav Kaushik, Jeffrey F. Naughton.
  - **Dynamic Re-grouping of Continuous Queries**. Jianjun Chen, David J. DeWitt.
  - **Merge as a Lattice-Join of XML Documents**. Kristin Tufte, David Maier.
  - **Putting XML Query Algebras into Context**. Stratis D. Viglas, Leonidas Galanis, David J. DeWitt, David Maier, Jeffrey F. Naughton.

- ***2001***
- **Estimating the Selectivity of XML Path Expressions for Internet Scale Applications**. Ashraf Aboulnaga, Alaa R. Alameldeen, and Jeffrey F. Naughton. *VLDB 2001.*
- **Generating Synthetic Complex-structured XML Data** . Ashraf Aboulnaga, Jeffrey F. Naughton, and Chun Zhang. *WebDB 2001.* [software download].
- **Form-Based Proxy Caching for Database-Backed Web Sites**. Qiong Luo and Jeffrey F. Naughton. *VLDB 2001.* [Full version].
- **The Design and Performance Evaluation of Various XML Storage Strategies**. Feng Tian, David J. DeWitt, Jianjun Chen, and Chun Zhang.
- **X-Diff: An Effective Change Detection Algorithm for XML Documents**. Yuan Wang, David J. DeWitt, and Jin-Yi Cai. [The XML diff tool].
- **On Supporting Containment Queries in Relational Database Management Systems**. Chun Zhang, Jeffrey F. Naughton, David J. DeWitt, Qiong Luo, and Guy Lohman. *SIGMOD 2001.* [Short version]. [Full version].

- **OpenCQ** (Georgia Tech)
  - L. Liu, C. Pu, W. Tang: Continual Queries for Internet Scale Event-Driven Information Delivery. TKDE 11(4): 610-628 (1999)

- **StatStream** (New York University)
  - Yunyue Zhu, Dennis Shasha "StatStream: Statistical Monitoring of Thousands of Data Streams in Real Time" VLDB, August, 2002. pp. 358-369
  - http://cs.nyu.edu/cs/faculty/shasha/papers/statstream.html

- **Stream** (Stanford)
  - The STREAM Group. STREAM: The Stanford Stream Data Manager (short overview paper), *IEEE Data Engineering Bulletin, Vol. 26 No. 1, March 2003*
  - Motwani et al. (Stanford group), Query Processing, Resource Management, and Approximation and in a Data Stream Management System. CIDR 2003.

- **Streaminer** (UIUC)
  - C. Aggarwal, J. Han, J. Wang, and P. S. Yu, "A Framework for Clustering Evolving Data Streams", Proc. 2003 Int. Conf. on Very Large Data Bases (VLDB'03), Berlin, Germany, Sept. 2003
  - http://www-faculty.cs.uiuc.edu/~hanj/projs/streamine.htm

- **Tapestry** (Xerox)
  - D. B. Terry, D. Goldberg, D. Nichols, B. M. Oki: Continuous Queries over Append-Only Databases. SIGMOD Conference 1992: 321-330

- **Telegraph(CQ)** (Berkeley)
  - Sirish Chandrasekaran, Owen Cooper, Amol Deshpande, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, Sailesh Krishnamurthy, Samuel R. Madden, Vijayshankar Raman, Fred Reiss, Mehul A. Shah. **TelegraphCQ: Continuous Dataflow Processing for an Uncertain World**, *CIDR 2003.*
  - Sailesh Krishnamurthy, Sirish Chandrasekaran, Owen Cooper, Amol Deshpande, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, Samuel R. Madden, Vijayshankar Raman, Fred Reiss, and Mehul A. Shah. **TelegraphCQ: An Architectural Status Report**. *IEEE Data Engineering*

Bulletin, Vol 26(1), March 2003.

- **TinyDB** (Berkeley)
  - Samuel R. Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. The Design of an Acquisitional Query Processor for Sensor Networks. To Appear, SIGMOD, June 2003.
  - Joseph M. Hellerstein, Wei Hong, Samuel Madden, and Kyle Stanek. Beyond Average: Towards Sophisticated Sensing with Queries. To Appear, 2nd International Workshop on Information Processing in Sensor Networks (IPSN '03), March 2003.
  - Samuel R. Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. TAG: a Tiny AGgregation Service for Ad-Hoc Sensor Networks. OSDI, December 2002.
  - Samuel R. Madden, Robert Szewczyk, Michael J. Franklin and David Culler. Supporting Aggregate Queries Over Ad-Hoc Wireless Sensor Networks. Workshop on Mobile Computing and Systems Applications, 2002 (to appear).
  - Samuel R. Madden. Query Processing For Streaming Sensor Data. Ph.D. Qualifying Exam Proposal, May 2002.
  - Samuel R. Madden and Michael J. Franklin. Fjording the Stream: An Architecture for Queries over Streaming Sensor Data. ICDE Conference, 2002.
  - Samuel R. Madden, Mehul A. Shah, Joseph M. Hellerstein, and Vijayshankar Raman. Continuously Adaptive Continuous Queries over Streams. SIGMOD Conference, 2002.

- **Tradebot** (www.tradebot.com)

- **Tribeca** (Bellcore)
  - Mark Sullivan: Tribeca: A Stream Database Manager for Network Traffic Analysis Source Proceedings of the 22th International Conference on Very Large Data Bases 1996
  - Mark Sullivan: Tribeca: A System for Managing Large Databases of Network Traffic, USENIX Annual Technical Conference (NO 98), 1998

DSMS – Query Languages

- Seshadri, P., Livny, M., Ramakrishnan, R., Sequence Query Processing. in *ACM Conference on the Management of Data (SIGMOD)*, (1994), 430-441.
- Arasu, A., Babu, S., Widom. J., The CQL Continuous Query Language: Semantic Foundations and Query Execution, *Technical Report, Oct. 2003*
- Wang, H., Zaniolo, C., ATLaS: A Native Extension of SQL for Data Mining and Stream Computations, UCLA CS Dept., 2002.
- Arasu, A., Babu, S., Widom, J., An Abstract Semantics and Concrete Language for Continuous Queries over Streams and Relations. Technical Report, 2003 (Abbreviated version in Proc. International Conference on Data Base Programming Languages, 2003).
- Chandrasekaran et al., Windows Explained, Windows Expressed. Technical Report, 2003.

DSMS – Query Processing

- Query Processing for Streams
  - Tucker, P., Maier, D., Sheard, T. and Fegaras, L. Punctuating Continuous Streams, Oregon Graduate Institute, Portland, OR, 2002.
  - U. Srivastava, S. Babu, and J. Widom. Monitoring Stream Properties for Continuous Query Processing (short paper), *Proc. of the 2003 Workshop on Management and Processing of Data Streams (MPDS 2003), June 2003*
  - A. Arasu, B. Babcock, S. Babu, J. McAlister, and J. Widom. Characterizing Memory Requirements for Queries over Continuous Data Streams , *ACM Transactions on Database Systems, March 2004.*
  - S. Babu, U. Srivastava, and J. Widom. Exploiting k-Constraints to Reduce Memory Overhead in Continuous Queries over Data Streams, *Technical Report, November 2002*
  - Vijayshankar Raman, Amol Deshpande, and Joseph M. Hellerstein. Using State Modules for Adaptive Query Processing. ICDE 2003.

- Adaptive Query Processing
  - Madden, Shah, Hellerstein and Raman. Continuously Adaptive Continuous Queries over Streams. ACM SIGMOD 2002.
  - Avnur & Hellerstein. Eddies: Continuously Adaptive Query Processing. ACM SIGMOD 2000.
  - Chandrasekaran and Franklin. Streaming Queries over Streaming Data. IEEE VLDB 2002.

- Query Processing with Windows
  - Kang, J., Naughton, J.F., Viglas, S.D., Evaluating Window Joins Over Unbounded Streams. Evaluating Window Joins over Unbounded Streams, ICDE 2003.
  - Golab & Ozsu. Processing Sliding Window Multi-Joins in Continuous Queries over Data Streams. Technical Report, February 2003.

- Memory Requirements for Query Processing
  - Arasu et al. Characterizing Memory Requirements for Queries over Continuous Data Streams. PODS 2002.
  - Babu, Srivastava, Widom. Exploiting k-Constraints to Reduce Memory Overhead in Continuous Queries over Data Streams. Technical Report, November 2002.
  - Tucker, Maier, Sheard, Fegaras. Exploiting Punctuation Semantics in Continuous Data Streams. Transactions on Knowledge and Data Engineering, May 2003. *(http://www.cse.ogi.edu/~ptucker/PStream/)*

- Approximation Techniques
  - Gibbons & Tirthapura. Distributed Streams Algorithms for Sliding Windows, ACM TOCS (earlier version appeared in ACM Symp. on Parallel Algorithms & Architectures, 2002).
  - Gilbert, Kotidis, Muthukrishnan & Strauss. Surfing Wavelets on Streams: One-pass Summaries for Approximate Aggregate Queries. IEEE VLDB 2001.

- Dobra, Garofalakis, Gehrke and Rastogi. Processing Complex Aggregate Queries over Data Streams. ACM SIGMOD 2002.

- Query Optimization
  - [Viglas02] Viglas and Naughton. Rate-based Query Optimization for Streaming Information Sources. SIGMOD 2002.
  - Chen, DeWitt and Naughton. Design and Evaluation of Alternative Selection Placement Strategies in Optimizing Continuous Queries. ICDE 2002.

- QoS, Scheduling, Load Shedding, Sampling
  - B. Babcock, S. Babu, M. Datar, R. Motwani, and D. Thomas. Operator Scheduling in Data Stream Systems*, Technical Report, Oct. 2003*
  - B. Babcock, M. Datar, and R. Motwani. Load Shedding Techniques for Data Stream Systems (short paper), *Proc. of the 2003 Workshop on Management and Processing of Data Streams (MPDS 2003), June 2003*
  - Babcock, Babu, Datar and Motwani. Chain: Operator Scheduling for Memory Minimization in Data Stream Systems. ACM SIGMOD 2003.
  - Babcock, Datar & Motwani. Sampling from a Moving Window Over Streaming Data (extended abstract). Proc. Symposium on Discrete Algorithms (SODA), 2002. *(http://www.stanford.edu/~babcock/papers/soda02.ppt)*
  - Don Carney, Ugur Cetintemel, Alex Rasin, Stan Zdonik, Mitch Cherniack and Michael Stonebraker, Operator Scheduling in a Data Stream Environment, Proceedings of the 29th International Conference on Very Large Data Bases (VLDB), September, 2003.
  - Das, Gehrke & Riedewald. Approximate Join Processing Over Data Streams. SIGMOD 2003.
  - Hammad, Franklin, Aref and Elmagarmid. Scheduling for Shared Window Joins over Data Streams. VLDB 2003.
  - *R. Motwani, J. Widom, A. Arasu, B. Babcock, S. Babu, M. Datar, G. Manku, C. Olston, J. Rosenstein, and R. Varma. Query Processing, Resource Management, and Approximation in a Data Stream Management System, Proc. of the 2003 Conf. on Innovative Data Systems Research (CIDR), January 2003.*
  - Nesime Tatbul, Ugur Cetintemel, Stan Zdonik, Mitch Cherniack and Michael Stonebraker, Load Shedding in a Data Stream Manager, Proceedings of the 29th International Conference on Very Large Data Bases (VLDB), September, 2003.
  - *Vijayshankar Raman and Joseph M. Hellerstein.* ***Partial Results for Online Query Processing****. ACM SIGMOD Conference, Madison, WI, June 2002.*

- Distributed Processing / Streams
  - Mitch Cherniack, Hari Balakrishnan, Don Carney, Ugur Cetintemel, Ying Xing and Stan Zdonik, **Scalable Distributed Stream Processing**, Proceedings of the Conference for Innovative Database Research (CIDR), January, 2003, Asilomar, CA.
  - *C. Olston, J. Jiang, and J. Widom. Adaptive Filters for Continuous Queries over Distributed Data Streams*
    In Proc. of the ACM Intl Conf. on Management of Data (SIGMOD 2003), June, 2003

- *B. Babcock and C. Olston. Distributed Top-K Monitoring, Proc. of the ACM Intl Conf. on Management of Data (SIGMOD 2003), June 2003*
- *Shah, M.A., Hellerstein, J.M., Chandrasekaran, S. and Franklin, M.J., Flux: An Adaptive Partitioning Operator for Continuous Query Systems. International Conference on Data Engineering (ICDE), (Bangalore, India, 2003).*
- Srivastava & Widom. Flexible Time Management in Data Stream Systems. Technical Report, July 2003.
- Olston, Jiang & Widom. Adaptive Filters for Continuous Queries over Distributed Data Streams. ACM SIGMOD 2003.

Mining Data Streams

- D. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, C. Erwin, E. Galvez, M. Hatoun, J. Hwang, A. Maskey, A. Rasin, A. Singer, M. Stonebraker, N. Tatbul, Y. Xing, R.Yan, S. Zdonik. Aurora: A Data Stream Management System (Demonstration). In proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'03), San Diego, CA, June 2003.
- G. Abowd, Ubiquitous Computing: Research themes and open issues from an applications perspective. Technical Report GIT-GVU 96-23, GVU Center, Georgia Institute of Technology, October 1996.
- A. Arasu, B. Babcock. S. Babu, M. Datar, K. Ito, I. Nishizawa, J. Rosenstein, and J. Widom. STREAM: The Stanford Stream Data Manager Demonstration description - short overview of system status and plans; in Proc. of the ACM Intl Conf. on Management of Data (SIGMOD 2003), June 2003.
- C. Aggarwal, J. Han, J. Wang, P. S. Yu, A Framework for Clustering Evolving Data Streams, Proc. 2003 Int. Conf. on Very Large Data Bases (VLDB'03), Berlin, Germany, Sept. 2003.
- C. Aggarwal, J. Han, J. Wang, and P. S. Yu, A Framework for Projected Clustering of High Dimensional Data Streams, Proc. 2004 Int. Conf. on Very Large Data Bases (VLDB'04), Toronto, Canada, Aug. 2004.
- C. Aggarwal, J. Han, J. Wang, and P. S. Yu, On Demand Classification of Data Streams, Proc. 2004 Int. Conf. on Knowledge Discovery and Data Mining (KDD'04), Seattle, WA, Aug. 2004.
- M. Ajtai, T.S. Jayram, R. Kumar, and D. Sivakumar. Approximate counting of inversions in a data stream. In 34th ACM Symposium on Theory of Computing, Montral, Qubec, Canada, 2002.
- B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom. Models and issues in data stream systems. In Proceedings of PODS, 2002.
- B. Babcock, M. Datar, and R. Motwani. Load Shedding Techniques for Data Stream Systems (short paper) In Proc. of the 2003 Workshop on Management and Processing of Data Streams (MPDS 2003), June 2003
- B. Babcock, M. Datar, R. Motwani, L. O'Callaghan: Maintaining Variance and k-Medians over Data Stream Windows, to appear in Proceedings of the 22nd Symposium on Principles of Database Systems (PODS 2003)
- M. Burl, Ch. Fowlkes, J. Roden, A. Stechert, and S. Mukhtar, Diamond Eye: A distributed architecture for image data mining, in SPIE DMKD, Orlando, April 1999.

- Ben-David, S, Johannes Gehrke and Daniel Kifer, Detecting Change in Data Streams, Proceedings of VLDB 2004.
- R. Bhargava, H. Kargupta, and M. Powers, Energy Consumption in Data Analysis for On-board and Distributed Applications, Proceedings of the ICML'03 workshop on Machine Learning Technologies for Autonomous Space Applications, 2003.
- M. Charikar, K. Chen and M. Farach-Colton. Finding Frequent Items in Data Streams. International Colloquium on Automata,Languages, and Programming (ICALP '02) 508--515.
- M. Charikar, L. O'Callaghan, and R. Panigrahy. Better streaming algorithms for clustering problems In Proc. of 35th ACM Symposium on Theory of Computing (STOC), 2003.
- Y. D. Cai, D. Clutter, G. Pape, J. Han, M. Welge, and L. Auvil, MAIDS: Mining Alarming Incidents from Data Streams, (system demonstration), Proc. 2004 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'04), Paris, France, June 2004.
- Y. Chen, G. Dong, J. Han, B. W. Wah, and J. Wang. Multi-Dimensional Regression Analysis of Time-Series Data Streams In VLDB Conference, 2002.
- F.Chu and C.Zaniolo, Fast and light boosting for adaptive mining of data streams, in Proc. of the 5th Pacific-Asic Conference on Knowledge Discovery and Data Mining (PAKDD), Sydney, May 2004.
- Liadan O'Callaghan, Nina Mishra, Adam Meyerson, Sudipto Guha, and Rajeev Motwani. Streaming-data algorithms for high-quality clustering. Proceedings of IEEE International Conference on Data Engineering, March 2002.
- G. Cormode, S. Muthukrishnan What's hot and what's not: tracking most frequent items dynamically. PODS 2003: 296-306
- G. Cormode and S. Muthukrishnan., What is new: Finding significant differences in network data streams, INFOCOM 2004.
- L. Chen, K. Reddy, and G. Agrawal, GATES: A Grid-Based Middleware for Processing Distributed Data Streams, in proceedings of Conference on High Performance Distributed Computing (HPDC), 2004.
- Y. Dora Cai, D. Clutter, G. Pape, J. Han, M. Welge, L. Auvil. MAIDS: Mining Alarming Incidents from Data Streams. Proceedings of the 23rd ACM SIGMOD (International Conference on Management of Data), June 13-18, 2004, Paris, France.
- Q. Ding, Q. Ding, and W. Perrizo, Decision Tree Classification of Spatial Data Streams Using Peano Count Trees," Proceedings of the ACM 124 Symposium on Applied Computing, Madrid, Spain, March 2002, pp. 413417.
- Mayur Datar, Aristides Gionis, Piotr Indyk, Rajeev Motwani: Maintaining Stream Statistics Over Sliding Windows (Extended Abstract) in Proceedings of 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2002).
- Abhinandan Das, Johannes Gehrke and Mirek Riedewald, Approximate Join Processing Over Data Streams,
- Proc. of the 2003 ACM SIGMOD International Conference on Management of Data, 2003
- P. Domingos and G. Hulten. Mining High-Speed Data Streams. In Proceedings of the Association for Computing Machinery Sixth International Conference on Knowledge Discovery and Data Mining, pages 71--80, 2000.
- P. Domingos and G. Hulten, A General Method for Scaling Up Machine Learning Algorithms and its Application to Clustering, Proceedings of the

Eighteenth International Conference on Machine Learning, 2001, 106--113, Williamstown, MA, Morgan Kaufmann.
- G. Dong, J. Han, L.V.S. Lakshmanan, J. Pei, H. Wang and P.S. Yu. Online mining of changes from data streams: Research problems and preliminary results, In Proceedings of the 2003 ACM SIGMOD Workshop on Management and Processing of Data Streams. In cooperation with the 2003 ACM-SIGMOD International Conference on Management of Data (SIGMOD'03), San Diego, CA, June 8, 2003.
- W Fan, StreamMiner: A Classifier Ensemble-based Engine to Mine Concept Drifting Data Streams, VLDB'2004
- Wei Fan, Yi-an Huang, Haixun Wang, and Philip S. Yu, Active Mining of Data Streams, Proceedings of SIAM Int. Conference on Data Mining 2004.
- V. Ganti, Johannes Gehrke, Raghu Ramakrishnan: Mining Data Streams under Block Evolution. SIGKDD Explorations 3(2): 1-10 (2002).
- M. Garofalakis, Johannes Gehrke, Rajeev Rastogi: Querying and mining data streams: you only get one look a tutorial. SIGMOD Conference 2002: 635
- C. Giannella, J. Han, J. Pei, X. Yan, and P.S. Yu, Mining Frequent Patterns in Data Streams at Multiple Time Granularities, in H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (eds.), Next Generation Data Mining, AAAI/MIT, 2003.
- Amol Ghoting and Srinivasan Parthasarathy, Facilitating Interactive Distributed Data Stream Processing and Mining, In Proceedings of the IEEE International Symposium on Parallel and Distributed Processing Systems (IPDPS), April 2004.
- Anna C. Gilbert, Yannis Kotidis, S. Muthukrishnan, Martin Strauss: One-Pass Wavelet Decompositions of Data Streams. TKDE 15(3): 541-554 (2003)
- Gaber, M, M., Krishnaswamy, S., and Zaslavsky, A., Adaptive Mining Techniques for Data Streams Using Algorithm Output Granularity, The Australasian Data Mining Workshop (AusDM 2003), Held in conjunction with the 2003 Congress on Evolutionary Computation (CEC 2003), December, Canberra, Australia, Springer Verlag, Lecture Notes in Computer Science (LNCS).
- Gaber, M. M., Krishnaswamy, S. and Zaslavsky, A. (2004). Cost-Efficient Mining Techniques for Data Streams. In Proc. Australasian Workshop on Data Mining and Web Intelligence (DMWI2004), Dunedin, New Zealand. CRPIT, 32. Purvis, M., Ed. ACS.
- Gaber, M, M., Krishnaswamy, S., and Zaslavsky, A., (2004), A Wireless Data Stream Mining Model, Accepted for publication in the Third International Workshop on Wireless Information Systems (WIS 2004), Held in conjunction with the Sixth International Conference on Enterprise Information Systems (ICEIS 2004), Porto, Portugal, April 13-14, ICEIS Press, ISBN.
- Gaber, M, M., Krishnaswamy, S., and Zaslavsky, A., Ubiquitous Data Stream Mining, Current Research and Future Directions Workshop Proceedings held in conjunction with The Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining, Sydney, Australia May 26 2004.
- S. Guha, N. Mishra, R. Motwani, and L. O'Callaghan. Clustering data streams. In Proceedings of the Annual Symposium on Foundations of Computer Science. IEEE, November 2000.

- Sudipto Guha, Adam Meyerson, Nina Mishra, Rajeev Motwani, and Liadan O'Callaghan, Clustering Data Streams: Theory and Practice TKDE special issue on clustering, vol. 15, 2003.
- Lukasz Golab and M. Tamer Ozsu. Issues in Data Stream Management. In SIGMOD Record, Volume 32, Number 2, June 2003, pp. 5--14.
- Gaber, M, M., Zaslavsky, A., and Krishnaswamy, S., (2004), A Cost-Efficient Model for Ubiquitous Data Stream Mining, Accepted for publication in the Tenth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004), Perugia Italy, July 4-9.
- [GZK04b] Gaber, M, M., Zaslavsky, A., and Krishnaswamy, S., Resource-Aware Knowledge Discovery in Data Streams, Accepted for publication in the Proceedings of First International Workshop on Knowledge Discovery in Data Streams, to be held in conjunction with the 15th European Conference on Machine Learning (ECML 2004) and the 8th European Conference on the Principals and Practice of Knowledge Discovery in Databases (PKDD 2004), Pisa, Italy, 20-24 September 2004.
- M. Henzinger, P. Raghavan and S. Rajagopalan, Computing on data streams , Technical Note 1998-011, Digital Systems Research Center, Palo Alto, CA, May 1998
- J. Hsu, Data Mining Trends and Developments: The Key Data Mining Technologies and Applications for the 21st Century. In D Colton, M J Payne, N Bhatnagar, and C R Woratschek (Eds.), The Proceedings of ISECON 2002, v 19 (San Antonio): 224b. AITP Foundation for Information Technology Education. ISSN: 1542-7382.
- G. Hulten, L. Spencer, and P. Domingos. Mining Time-Changing Data Streams. ACM SIGKDD 2001.
- Z. He, X. Xu, S. Deng and J. Z. Huang. Clustering Categorical Data Streams, Journal of Computational Methods in Science and Engineering (JCMSE), 2004, to appear.
- Carsten Isert and Karsten Schwan. ACDS: Adapting Computational Data Streams for high performance. In International Parallel and Distributed Processing Symposium 2000
- R. Jin and G. Agrawal, Efficient Decision Tree Construction on Streaming Data, in proceedings of ACM SIGKDD 2003.
- Cheqing Jin, Weining Qian, Chaofeng Sha, Jeffrey X. Yu, and Aoying Zhou. Dynamically Maintaining Frequent Items over a Data Stream. In Proceedings of the 12th ACM Conference on Information and Knowledge Management (CIKM'2003).
- Hillol Kargupta, Ruchita Bhargava, Kun Liu, Michael Powers, Patrick Blair, Samuel Bushra, James Dull, Kakali Sarkar, Martin Klein, Mitesh Vasa, and David Handy, VEDAS: A Mobile and Distributed Data Stream Mining System for Real-Time Vehicle Monitoring, Proceedings of SIAM International Conference on Data Mining 2004.
- S. Krishnamurthy, S. Chandrasekaran, O. Cooper, A. Deshpande, M. Franklin, J. Hellerstein, W. Hong, S. Madden, V. Raman, F. Reiss, and M. Shah. TelegraphCQ: An Architectural Status Report. IEEE Data Engineering Bulletin, Vol 26(1), March 2003.

- Keogh, E., Chu, S., Hart, D. & Pazzani, M. (2001). An Online Algorithm for Segmenting Time Series. In Proceedings of IEEE International Conference on Data Mining. pp 289-296
- M. Khan, Q. Ding, and W. Perrizo, K-nearest Neighbor Classification on Spatial Data Stream Using P-trees, Proceedings of the PAKDD, Taipei, Taiwan, May 2002, pp. 517-528. 128
- J. Kleinberg, Bursty and Hierarchical Structure in Streams, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Edmonton, Alberta, Canada, 2002.
- E. Keogh, J. Lin, and W. Truppel. Clustering of Time Series Subsequences is Meaningless: Implications for Past and Future Research. In proceedings of the 3rd IEEE International Conference on Data Mining. Melbourne, FL. Nov 19-22, 2003.
- S. Krishnaswamy, S. Loke, and A. Zaslavsky, Towards Anytime Anywhere Data Mining E-Services, Proceedings of the Australian Data Mining Workshop (ADM'02) at the 15th Australian Joint Conference on Artificial Intelligence, (eds) S.J. Simoff, G.J. Williams, and M. Hegland. Canberra, Australia, December 2002, pp. 47 - 56, Published by the University of Technology Sydney, 2002. ISBN 0-9750075-0-5
- Nick Koudas and Divesh Srivastava. Data Stream Query Processing: A Tutorial. Presented at International Conference on Very Large Databases (VLDB), 2003.
- Christoph Koch, Stefanie Scherzinger: Attribute Grammars for Scalable Query Processing on XML Streams, Database Programming Languages, 9th International Workshop, DBPL 2003, Potsdam, Germany, September 6-8, 2003.
- Kargupta, H., Park, B., Pittie, S., Liu, L., Kushraj, D. and Sarkar, K. (2002). MobiMine: Monitoring the Stock Market from a PDA. ACM SIGKDD Explorations. January 2002. Volume 3, Issue 2. Pages 37--46. ACM Press.
- Christoph Koch, Stefanie Scherzinger, Nicole Schweikardt, Bernhard Stegmaier: Schema-based Scheduling of Event Processors and Buffer Minimization for Queries on Structured Data Streams. Proceedings of VLDB 2004.
- Christoph Koch, Stefanie Scherzinger, Nicole Schweikardt, Bernhard Stegmaier: FluXQuery: An Optimizing XQuery Processor for Streaming XML Data. Proceedings of VLDB 2004.
- Laerhoven, K. Van, Aidoo K., Lowette S., 2001. Real-time Analysis of Data from Many Sensors with Neural Networks. Proceedings of the fourth International Symposium on Wearable Computers (ISWC) Zurich, 7-9 October 2001. IEEE Press.
- M. Last, Online Classification of Nonstationary Data Streams, Intelligent Data Analysis, Vol. 6, No. 2, pp. 129-147, 2002.
- J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA. June 13, 2003.
- Hua-Fu Li, Suh-Yin Lee, and Man-Kwan Shan. An Efficient Algorithm for Mining Frequent Itemsets over the Entire History of Data Streams, Accepted for publication in the Proceedings of First International Workshop on Knowledge Discovery in Data Streams, to be held in conjunction with the 15th

European Conference on Machine Learning (ECML 2004) and the 8th European Conference on the Principals and Practice of Knowledge Discovery in Databases (PKDD 2004), Pisa, Italy, 20-24 September 2004.

- G. S. Manku and R. Motwani. Approximate frequency counts over data streams. In Proceedings of the 28th International Conference on Very Large Data Bases, Hong Kong, China, August 2002.

- Martin H.C. Law, Nan Zhang, and Anil Jain, Nonlinear Manifold Learning for Data Stream, Proceedings of SIAM International Conference on Data Mining 2004.
- S. Muthukrishnan (2003), Data streams: algorithms and applications. Proceedings of the fourteenth annual ACM-SIAM symposium on discrete algorithms.
- Rajeev Motwani, Jennifer Widom, Arvind Arasu, Brian Babcock, Shivnath Babu, Mayur Datar, Gurmeet Manku, Chris Olston, Justin Rosenstein, and Rohit Varma, Query Processing, Approximation, and Resource Management in a Data Stream Management System, Proceedings of the 2003 Conference on Innovative Data Systems Research.
- O. Nasraoui, C. Cardona, C. Rojas, F. González, TECNO-STREAMS: Tracking Evolving Clusters in Noisy Data Streams with a Scalable Immune System Learning Model, in Proc. of Third IEEE International Conference on Data Mining (ICDM'03), Melbourne, FL, November 2003, pp. 235-242.
- Nasraoui O., Cardona C., Rojas C., and Gonzalez F., Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm, in Proc. of WebKDD 2003 – KDD Workshop on Web mining as a Premise to Effective and Intelligent Web Applications, Washington DC, August 2003, p. 71
- O. Nasraoui, C. Rojas, and C. Cardona. Single Pass Mining of Evolving Trends in Web Data with Explicit Retrieval Similarity Measures. In Proceedings of "International Web Dynamics Workshop", International World Wide Web Conference, New York, NY, May. 2004.
- O'Callaghan et al. Streaming-Data Algorithms For High-Quality Clustering. ICDE 2002.
- C. Olston, J. Jiang, and J. Widom. Adaptive Filters for Continuous Queries over Distributed Data Streams. ACM SIGMOD 2003 International Conference on Management of Data, San Diego, California, June 2003, pp. 563-574.
- Carlos Ordonez. Clustering Binary Data Streams with K-means ACM DMKD 2003.
- Kok-Leong Ong, Zili Zhang, Wee-Keong Ng, and Ee-Peng Lim. Agents and Stream Data Mining: A New Perspective. IEEE Intelligent Systems, to appear.
- B. Park and H. Kargupta. Distributed Data Mining: Algorithms, Systems, and Applications. To be published in the Data Mining Handbook. Editor: Nong Ye. 2002.
- S. Papadimitriou, C. Faloutsos, and A. Brockwell, Adaptive, Hands-Off Stream Mining, 29th International Conference on Very Large Data Bases VLDB, 2003.

- P. Domingos and G. Hulten. Catching Up with the Data: Research Issues in Mining Data Streams. Workshop on Research Issues in Data Mining and Knowledge Discovery, 2001. Santa Barbara, CA
- Byung-Hoon Park, George Ostrouchov, Nagiza F. Samatova, and Al Geist, Reservoir-Based Random Sampling with Replacement from Data Stream, Proceedings of SIAM International Conference on Data Mining 2004.
- S. Pirttikangas, J. Riekki, J. Kaartinen, J. Miettinen, S. Nissila, & J. Roning. Genie of The Net: A New Approach For A Context-Aware Health Club. In Proceedings of Joint 12th ECML'01 and 5th European Conference on PKDD'01. September 3-7, 2001, Freiburg, Germany.
- T. Palpanas, M. Vlachos, E. Keogh, D. Gunopulos, W. Truppel (2004). Online Amnesic Approximation of Streaming Time Series. In ICDE . Boston, MA, USA, March 2004.
- K. Rajaraman, Ah-Hwee Tan: Topic Detection, Tracking, and Trend Analysis Using Self-Organizing Neural Networks. PAKDD 2001: 102-107
- A. Srivastava and J. Stroeve, Onboard Detection of Snow, Ice, Clouds and Other Geophysical Processes Using Kernel Methods, Proceedings of the ICML'03 workshop on Machine Learning Technologies for Autonomous Space Applications
- R. Sadri, C. Zaniolo, A. Zarkesh, J. Adibi Optimization of pattern matching Queries on Database Sequences, PODS'2001, Twentieth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Santa Barbara, May 21-24 2001
- R. Sadri, C. Zaniolo, A. Zarkesh, J.Adibi A Sequential Pattern Query language for Supporting Instant Data Mining for e-Services, 27th International Conference on Very Large Data Bases (VLDB 2001), September 2001, Roma, Italy.
- S. Tanner, M. Alshayeb, E. Criswell, M. Iyer, A. McDowell, M. McEniry, K. Regner, EVE: On-Board Process Planning and Execution, Earth Science Technology Conference, Pasadena, CA, Jun. 11 - 14, 2002
- Wei-Guang Teng, Ming-Syan Chen, and Philip S. Yu, Resource-Aware Mining with Variable Granularities in Data Streams, Proceedings of SIAM International Conference on Data Mining 2004.
- Nesime Tatbul, Ugur Cetintemel, Stan Zdonik, Mitch Cherniack and Michael Stonebraker, Load Shedding in a Data Stream Manager Proceedings of the 29th International Conference on Very Large Data Bases (VLDB), September, 2003.
- N. Tatbul, U. Cetintemel, S. Zdonik, M. Cherniack, M. Stonebraker. Load Shedding on Data Streams, In Proceedings of the Workshop on Management and Processing of Data Streams (MPDS 03), San Diego, CA, USA, June 8, 2003.
- H. Wang, W. Fan, P. Yu and J. Han, Mining Concept-Drifting Data Streams using Ensemble Classifiers, in the 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Aug. 2003, Washington DC, USA.
- S. D. Viglas and Jeffrey F. Naughton Rate based query optimization for streaming information sources. In Proc. of SIGMOD, 2002
- A. Zhou, Z. Cai, L. Wei, and W. Qian. M-Kernel Merging: Towards Density Estimation over Data Streams. In Proceedings of the 8th International

Conference on Database Systems for Advanced Applications (DASFAA'2003), 2003.
- D. Zhang, D. Gunopulos, V. J. Tsotras, B. Seeger, Temporal and Spatio-Temporal Aggregations over Data Streams using Multiple Time Granularities, Journal of Information Systems, vol. 27, no. 8, 2002.
- D. Zhang, D. Gunopulos, V. J. Tsotras, B. Seeger, Temporal Aggregation over Data Streams using Multiple Granlarities, Proc. of 8th International Conference on Extending Database Technology (EDBT), Prague, Czech Republic, 2002.
- Y. Zhu, D. Shasha. StatStream: Statistical monitoring of thousands of data streams in real time. In VLDB 2002, pages 358--369.
- Y. Zhu, D. Shasha Efficient Elastic Burst Detection in Data Streams The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD-2003 24 August 2003 - 27 August 2003.