

Robert Grossman

Curriculum Vita

Summary

Robert Grossman is a faculty member at the University of Chicago, where he is the Director of Informatics at the Institute for Genomics and Systems Biology, a Senior Fellow at the Computation Institute, and a Professor of Medicine in the Section of Genetic Medicine. His research group focuses on bioinformatics, data mining, cloud computing, data intensive computing, and related areas. He is also the Chief Research Informatics Officer of the Biological Sciences Division.

From 1998 to 2010, he was the Director of the National Center for Data Mining at the University of Illinois at Chicago (UIC). From 1984 to 1988 he was a faculty member at the University of California at Berkeley. He received a Ph.D. from Princeton in 1985 and a B.A. from Harvard in 1980.

He is also the Founder and a Partner of Open Data Group. Open Data provides management consulting and outsourced analytic services for businesses and organizations. At Open Data, he has led the development of analytic systems that are used by millions of people daily all over the world.

He has published over 150 papers in refereed journals and proceedings and edited seven books on data intensive computing, bioinformatics, cloud computing, data mining, high performance computing and networking, and Internet technologies.

Prior to founding the Open Data Group, he founded Magnify, Inc. in 1996. Magnify provides data mining solutions to the insurance industry. Grossman was Magnify's CEO until 2001 and its Chairman until it was sold to ChoicePoint in 2005. ChoicePoint was acquired by LexisNexis in 2008. Also, in 1996 he founded Magnify Research, which provides data mining solutions to the federal government. Magnify Research was sold to Baesch Computer Consulting in 2002, and is now part of Unisys.

He is a Member of the Board of Directors of the ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) for the term 2009–2011. He also served as a Member of the Board of Directors for the terms: 2005–2007 and 2007–2009.

Since 2008, he has been Chair of the Open Cloud Consortium, which develops standards and interoperability frameworks for cloud computing and supports the development of open source software for cloud computing.

From 1998 to 2010, he was the Chair for the Data Mining Group (DMG), an industry consortium responsible for the Predictive Model Markup Language (PMML), an XML language for data mining and predictive modeling.

Education and Work Experience

University of Chicago. Professor of Medicine in the Section of Genetic Medicine, 2010 – present; Core Faculty and Director of Informatics at the Institute for Genomics and Systems Biology; 2010 – present; Core Faculty and Senior Fellow at the Computation Institute, 2010 – present. Chief Research Informatics Officer, Biological Sciences Division, 2011 – present.

University of Illinois at Chicago. Professor, Department of Mathematics, Statistics, & Computer Science, 1995 – 2010; Associate Professor 1991 – 1995. Assistant Professor 1988 – 1991. Professor, Department of Computer Science, 1995 – 2010. Professor, Department of Bioengineering, 2009 – 2010. Director, Laboratory for Advanced Computing, 1990 – 2010. Director, National Center for Data Mining, 1998 – 2010. Part time appointment 1994 – 2010.

Institute for Genomics and System Biology, University of Chicago. Associate Senior Fellow, 2008 – 2010.

Open Data Group. Founder and Managing Partner, 2001 – present.

Magnify, Inc. Founder and Chairman, 1994 – 2005. CEO, 1994 – 2001.

Cornell University. Visiting Associate Professor, Department of Computer Science 1994 – 1995. Visiting Scientist, Mathematical Sciences Institute, 1989.

University of California at Berkeley. NSF Postdoctoral Research Fellow. 1984 – 1988.

Princeton University. 1980 – 1984. Ph.D. Mathematics, 1985.

Harvard University. 1976 – 1980. A.B., Mathematics, 1980.

Selected Awards

- Pritzker Scholar, University of Chicago, 2011
- Overall Winner, SC 09 Bandwidth Challenge. Maximizing Bandwidth Utilization in Distributed Data Intensive Applications, was the Overall Winner in the Bandwidth Challenge at the ACM/IEEE International Conference for High Performance Computing and Communications 2009 (SC09).
- First Place, SC 08 Bandwidth Challenge. Towards Global Scale Cloud Computing: Using Sector and Sphere on the Open Cloud Testbed won First Place in the Bandwidth Challenge at the ACM/IEEE International Conference for High Performance Computing and Communications 2008 (SC08).
- First Place, 2007 Analytics Challenge, Angle: Detecting Anomalies and Emergent Behavior from Distributed Data in Near Real Time, ACM/IEEE International Conference for High Performance Computing and Communications 2007 (SC07).
- ACM SIGKDD 2007 Service Award for “... the development of open and scalable architectures and standards for the SIGKDD and Global KDD Communities.” SIGKDD is the ACM’s professional group on Knowledge Discovery and Data Mining.
- First Place, 2007 Data Mining Practice Prize awarded at the ACM Conference on Knowledge Discovery and Data Mining (KDD 2007).
- First Place, 2006 Bandwidth Challenge, Distributing the Sloan Digital Sky Survey Using Sector, ACM/IEEE International Conference for High Performance Computing and Communications 2006 (SC06).
- First Place, 2005 Analytics Challenge, Real Time Change Detection of Highway Sensor Data, ACM/IEEE International Conference for High Performance Computing and Communications 2005 (SC05).

Research Interests

Data intensive computing, bioinformatics, cloud computing, data mining, high performance computing and networking, and Internet technologies.

Publications

Articles in Refereed Journals

1. R. Grossman, A note on the application of Morse theory to the study of the potential extrema of body surface potential maps, Volume 11, 1978, pages 201-202.
2. M. Beals, C. Fefferman, R. Grossman, Strictly pseudoconvex domains in \mathbb{C}^n , Bulletin American Mathematical Society, Volume, 8, 1983, pages 125-322.

3. R. Grossman and R. Larson, Hopf algebraic structures of families of trees, *Journal Algebra*, Volume 26, 1989, pages 184-210.
4. R. Grossman and R. Larson, Hopf-algebraic structure of combinatorial objects and differential operators, *Israel Journal Mathematics*, Volume 72, 1990, pages 109-117.
5. Matt Grayson and Robert Grossman, Models for free, nilpotent Lie algebras, *Journal Algebra*, Vol. 35, 1990, pages 177-191.
6. Robert Grossman and Richard Larson, Solving nonlinear equations from higher order derivations in linear stages, *Advances in Mathematics*, Vol. 82, 1990, pages 180-202.
7. Robert Grossman, The evaluation of expressions involving higher order derivations, *Journal of Mathematical Systems, Estimation, and Control*, Volume 1, 1991, pages 91-106.
8. Robert L. Grossman and Richard G. Larson, The realization of input-output maps using bialgebras, *Forum Mathematicum*, Volume 4, 1992, pages 109-121.
9. M. W. Bern, D. P. Dobkin, D. Eppstein, and R. Grossman, Visibility with a moving point of view, *Algorithmica*, Volume 11, pages 360-378, 1994.
10. Robert L. Grossman and Richard G. Larson, The symbolic computation of derivations using labeled trees, *Journal of Symbolic Computation*, Volume 13, pages 511-523, 1992.
11. Peter Crouch and Robert L. Grossman, Numerical integration of ordinary differential equations on manifolds, *Journal of Nonlinear Science*, Volume 3, pages 1-33, 1993.
12. Robert Grossman and H. Vincent Poor, Wavelet transforms associated with finite cyclic groups, *IEEE Transactions on Information Theory*, Volume 39, 1993, pp. 1157-1166.
13. Peter E. Crouch, Robert Grossman, Y. Yan, On the numerical integration of the rolling ball equations using geometrically exact algorithms, *Mechanics of Structures of Machines*, Volume 23, Issue 2, 1995, pages 257-272.
14. Robert L. Grossman and Robert G. Larson, An algebraic approach to hybrid systems, *Journal of Theoretical Computer Science*, Volume 138, pages 101-112, 1995.
15. R. L. Grossman, Data Mining Challenges for Digital Libraries, *ACM Computing Surveys*, Volume 28A (electronic), December, 1996.
16. R. L. Grossman, S. Bailey, A. Ramu and B. Malhi, P. Hallstrom, I. Pulley and X. Qin, The Management and Mining of Multiple Predictive Models Using the Predictive Model Markup Language (PMML), *Information and Software Technology*, Volume 41, 1999, pages 589-595.
17. Robert Grossman, and Marco Mazzucco, DataSpace - A Web Infrastructure for the Exploratory Analysis and Mining of Data, *IEEE Computing in Science and Engineering*, July/August, 2002, pages 44-51.
18. Robert Grossman, Mark Hornick, and Gregor Meyer, Data Mining Standards Initiatives, *Communications of the ACM*, Volume 45-8, 2002, pages 59-61.
19. H. Sivakumar, R. L. Grossman, M. Mazzucco, Y. Pan, Q. Zhang, Simple Available Bandwidth Utilization Library for High-Speed Wide Area Networks, *Journal of Supercomputing*, Volume 34, Number 3, pages 231-242, 2005.
20. Robert L. Grossman, Yunhong Gu, Dave Hanley, Xinwei Hong, Dave Lillethun, Jorge Levera, Joe Mambretti, Marco Mazzucco, and Jeremy Weinberger, Experimental Studies Using Photonic Data Services at IGrid 2002, *Journal of Future Computer Systems*, 2003, Volume 19, Number 6, pages 945-955.

21. Andrei L. Turinsky and Robert L. Grossman, Intermediate Strategies: A Framework for Balancing Cost and Accuracy in Distributed Data Mining, Knowledge and Information Systems, 2004, to appear.
22. Asvin Ananthanarayan, Rajiv Balachandran, Yunhong Gu, Robert Grossman, Xinwei Hong, Jorge Levera, Marco Mazzucco, Data Webs for Earth Science Data, Parallel Computing, Volume 29, 2003, pages 1363-1379.
23. J. Mambretti, J. Weinberger, J. Chen, E. Bacon, F. Yeh, D. Lillethun, R. Grossman, Y. Gu, M. Mazzucco, The Photonic TeraStream: Enabling Next Generation Applications Through Intelligent Optical Networking at iGrid 2002, Journal of Future Computer Systems, Elsevier Press, Volume 19, Number 6, pages 897-908.
24. Chong Zhang, Jason Leigh, Thomas A. DeFanti, Marco Mazzucco and Robert Grossman, TeraScope: Distributed Visual Data Mining of Terascale Data Sets Over Photonic Networks, Journal of Future Computer Systems, 2003, Volume 19, Number 6, pages 935-943
25. Ian Foster and Robert L. Grossman, Data Integration in a Bandwidth Rich World, Communications ACM, Volume 46, Issue 11, November, 2003, pages 50-57.
26. A. Chien, T. Faber, A. Falk, J. Bannister, R. Grossman, J. Leigh, Transport Protocols for High Performance: Whither TCP?, Communications ACM, Volume 46, Issue 11, November, 2003, pages 42-49.
27. Robert L. Grossman, Pavan Kasturi, Donald Hamelberg, Bing Liu, An Empirical Study of the Universal Chemical Key Algorithm for Assigning Unique Keys to Chemical Compounds, Journal of Bioinformatics and Computational Biology, 2004, Volume 2, Number 1, 2004, pages 155-171.
28. Yunhong Gu and Robert L. Grossman, SABUL: A Transport Protocol for Grid Computing, Journal of Grid Computing, Volume 1, pages 377-386, 2004.
29. Bing Liu, Robert L. Grossman and Yanhong Zhai, Mining Web Pages for Data Records, IEEE Intelligent Systems, November/December, 2004, pages 49-55.
30. Robert L. Grossman, Yunhong Gu, Xinwei Hong, Antony Antony, Johan Blom, Freek Dijkstra, and Cees de Laat, Teraflows over Gigabit WANs with UDT, Journal of Future Computer Systems, Elsevier Press, Volume 21, Number 4, 2005, pages 501-513.
31. Robert L. Grossman, Yunhong Gu, Dave Hanley, Xinwei Hong and Parthasarathy Krishnaswamy, Experimental Studies of Data Transport and Data Access of Earth Science Data over Networks with High Bandwidth Delay Products, Computer Networks, Volume 46, 2004, pages 411-421.
32. Robert L. Grossman and Richard G. Larson, Differential Algebra Structures on Families of Trees, Advances in Applied Mathematics, Volume 35, pages 97-119, 2005. Also arXiv:math/0409006v1 [math.QA].
33. Leland Wilkinson, Anushka Anand and Robert L Grossman, High-dimensional Visual Analytics: Interactive Exploration Guided by Pairwise Views of Point Distribution, IEEE Transactions on Visualization and Computer Graphics, Volume 12, Number 6, pages 1363-1372, 2006.
34. Robert L. Grossman, Yunhong Gu, David Handley, and Michal Sabala Joe Mambretti, Alex Szalay and Ani Thakar, Kazumi Kumazoe and Oie Yuji, Minsun Lee, Yoonjoo Kwon, and Woojin Seok, Data Mining Middleware for Wide Area High Performance Networks, Journal of Future Generation Computer Systems (FGCS), Volume 22, Number 8, pages 940-948, 2006.
35. Yunhong Gu and Robert L. Grossman, UDT: UDP-based Data Transfer for High-Speed Wide Area Networks, Computer Networks, Volume 51, Number 7, pages 1777-1799, 2007.

36. Robert L Grossman and Richard G. Larson, Hopf Algebras of Heap Ordered Trees and Permutations, *Communications in Algebra*, Volume 37, Issue 2, 2009, pages 453-459. Also arxiv.org/abs/0706.1327.
37. Robert L. Grossman, Yunhong Gu, Michael Sabala and Wanzhi Zhang, Compute and Storage Clouds Using Wide Area High Performance Networks, *Journal of Future Generation Computer Systems (FGCS)*, Volume 25, Issue 2, 2009, pages 179-183.
38. Yunhong Gu and Robert L Grossman, Sector and Sphere: Towards Simplified Storage and Processing of Large Scale Distributed Data, *Philosophical Transactions of the Royal Society A*, Volume 367, Number 1897, pages 2429–2445, 2009.
39. Robert L. Grossman, The Case for Cloud Computing, *IT Professional*, volume 11, number 2, pages 23-27, March/April 2009.
40. Feng Tian, Parantu K Shah, Xiangjun Liu, Nicolas Negre, Jia Chen, Oleksiy Karpenko, Kevin P White, Robert L Grossman, Flynet: a genomic resource for *Drosophila melanogaster* transcriptional regulatory networks, *Bioinformatics*, Volume 25, Number 22, pages 3001-3004, 2009.
41. Feng Tian, Jia Chen, Suying Bao, Lin Shia, Xiangjun Liua, and Robert Grossman, A graph model based study on regulatory impacts of transcription factors of *Drosophila melanogaster* and comparison across species, *Biochemical and Biophysical Research Communications*, Volume 386, Issue 4, 2009, Pages 559-562.
42. Yunhong Gu, Robert Grossman, Towards Efficient and Simplified Distributed Data Intensive Computing, *IEEE Transactions on Parallel and Distributed Systems*, Volume 22, Issue 6, 2011, pages 974-984, [[doi.ieeecomputersociety.org/10.1109/TPDS.2011.67](https://doi.org/10.1109/TPDS.2011.67)].
43. The modENCODE Consortium, Sushmita Roy, Jason Ernst, Peter V. Kharchenko, et. al., Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE, *Science*, Volume 330 (6012), pages 1787-1797, 2010, [[doi:10.1126/science.1198374](https://doi.org/10.1126/science.1198374)], PMID: 21177974.
44. Nicolas Negre, Christopher D. Brown, Lijia Ma, et. al., Cis-Regulatory Map of the *Drosophila* Genome, *Nature*, Volume 471, pages 527-531, 2011, [[doi:10.1038/nature09990](https://doi.org/10.1038/nature09990)], PMID: 21430782.
45. Xin Feng , Robert Grossman and Lincoln Stein, PeakRanger: A cloud-enabled peak caller for CHIP-seq data, *BMC Bioinformatics* 2011, 12:139, [doi:10.1186/1471-2105-12-139](https://doi.org/10.1186/1471-2105-12-139), PMCID: PMC3103446.
46. Robert L. Grossman and Kevin P. White, A vision for a Biomedical Cloud, *Journal of Internal Medicine*, Volume 271, Number 2, pages 122-130, 2012. PMID: 22142244.

Articles in Refereed Proceedings

1. R. Grossman, Quantum Controllability, *Proceedings of the 23rd Conference on Decision and Control*, IEEE, 1984 pages 1466-7.
2. R. Grossman and C. Martin, The approximation and control of symmetric systems over the circle, *Mathematical Theory of Network and Systems*, Lecture Notes in Computer Science Voume 58, P. A. Fuhrmann, editor, Springer-Verlag, New York, 1984, pages 376-388.
3. R. Grossman and C. Martin, The approximation and control of symmetric systems over compact groups, *Proceedings of the Berkeley-Ames Conference on Nonlinear Problems in Control and Fluid Dynamics*, L. R. Hunt and C. F. Martin, editors, Math Sci Press, Brookline, 1984, pages 145-170.

4. R. Grossman and R. Larson, The symbolic computation of higher order derivations: symmetries of expressions and actions of group algebras, *Differential Geometry: The Interface Between Pure and Applied Mathematics, Contemporary Mathematics*, 68, M. Luksic, C. Martin and W. Shadwick, editors, American Mathematical Society, Providence, 1987, pages 121-131.
5. R. Grossman, P. S. Krishnaprasad, and J. E. Marsden, The dynamics of two coupled rigid bodies, *Dynamical Systems Approaches to Nonlinear Problems in Systems and Circuits*, F. M. A. Salam and M. L. Levi, editors, SIAM, Philadelphia, 1988, pages 373-378.
6. R. Grossman and R. Larson, Labeled trees and the algebra of differential operators, *Graphs and Algorithms, Contemporary Mathematics*, Volume 89, B. Richter, editor, American Mathematical Society, Providence, 1989, pages 81-87.
7. R. Fateman and R. Grossman, Computer algebra and operators, *Symbolic Computation: Applications to Scientific Computing*, R. Grossman, editor, SIAM, Philadelphia, 1989, pp. 1-14.
8. M. Grayson and R. Grossman, Nilpotent Lie algebras and vector fields, *Symbolic Computation: Applications to Scientific Computing*, R. Grossman, editor, SIAM, Philadelphia, 1989, pages 77-96.
9. R. Grossman and R. G. Larson, Labeled trees and the efficient computation of derivations, *Proceedings of 1989 International Symposium on Symbolic and Algebraic Computation*, ACM, 1989, pages 74-80.
10. R. Grossman, Querying databases of trajectories of differential equations I: data structures for trajectories, *Proceedings of the 23rd Hawaii International Conference on Systems Sciences*, IEEE, 1990, pages 18-23.
11. M. W. Bern, D. P. Dobkin, D. Eppstein, and R. Grossman, Visibility with a moving point of view (extended abstract), *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, SIAM, 1990, pp. 107-117.
12. R. Grossman, Querying databases of trajectories of differential equations II: index functions, *Fourth NASA Workshop on Computational Control of Flexible Aerospace Systems*, NASA Conference Proceedings, Number 10065, Part 1, L. W. Taylor, Jr., editor, NASA Langley Research Center, 1991, pp. 35-39.
13. R. Grossman, Using trees to compute approximate solutions of ordinary differential equations exactly, *Differential Equations and Computer Algebra* M. Singer, editor, Academic Press, New York, 1991, pp. 29-59.
14. P. Crouch, R. Grossman, and R. G. Larson, Computations involving differential operators and their actions on functions, *Proceedings of 1991 International Symposium on Symbolic and Algebraic Computation*, ACM, 1991, pp. 301-307.
15. A. Baden and R. Grossman, Database computing and high energy physics, *Computing in High-Energy Physics 1991*, edited by Y. Watase and F. Abe, Universal Academy Press, Inc., Tokyo, 1991, pp. 59-66.
16. R. Grossman and R. G. Larson, The symbolic computation of vector field expressions, *Algebraic Computing in Control 1991*, edited by G. Jacob and F. Lamnabhi-Lagarrigue, Springer-Verlag, Berlin, 1991, pp. 1-10.
17. R. Grossman and D. Radford, A simple construction of bialgebra deformations, *Contemporary Mathematics: Quantum Groups and Deformations*, AMS, pp. 115-117, 1992.

18. P.E. Crouch and R. L. Grossman, The Explicit Computation of Integration Algorithms and First Integrals for Ordinary Differential Equations With Polynomial Coefficients Using Trees, Proceedings of the 1992 International Symposium on Symbolic and Algebraic Computation, ACM Press, pp. 89-94.
19. R. L. Grossman and R. G. Larson, Viewing hybrid systems as products of control systems and automata, Proceedings of the 31st IEEE Conference on Decision and Control, IEEE Press, 1992, pages 2953-2955.
20. P.E. Crouch, R. Grossman, and Y. Yan, On the numeric integration of dynamic attitude equations, Proceedings of the 31st IEEE Conference on Decision and Control, IEEE Press, 1992, pages 1497-1501.
21. R. L. Grossman, D. Lifka, and X. Qin, A proof-of-concept implementation interfacing an object manager to a hierarchical storage system, Twelfth IEEE Symposium on Mass Storage Systems, IEEE Press, Los Alamites, 1993, pp. 209-214.
22. E. May, D. Lifka, E. Lusk, L. E. Price, C. T. Day, S. Loken, J. F. MacFarlane, A. Baden R. Grossman, X. Qin, L. Cormell, A. Gauthier, P. Leibold, J. Marstaller, U. Nixdorf, B. Scipioni Requirements for a System to Analyze High Energy Physics Events Using Database Computing Twelfth IEEE Symposium on Mass Storage Systems, IEEE Press, Los Alamites, 1993, pp. 31-36.
23. C. T. Day, S. Loken, J. F. MacFarlane, E. May, D. Lifka, E. Lusk, L. E. Price, A. Baden, R. Grossman, X. Qin, L. Cormell, P. Leibold, D. Liu, U. Nixdorf, B. Scipioni, T. Song, Database Computing in HEP – Progress Report, Proceedings of the International Conference on Computing in High Energy Physics '92, C. Verkerk and W. Wojcik, editors, CERN-Service d'Information Scientifique, 1992, ISSN 0007-8328, pp. 557-560.
24. R. L. Grossman and R. L. Larson, Some Remarks About Flows in Hybrid Systems, in R. L. Grossman, A. Nerode, A. P. Ravn, and H. Rischel, editors, Hybrid Systems, Lecture Notes in Computer Science, Volume 736, Springer-Verlag, New York, 1993, pp. 357-365.
25. R. L. Grossman, D. Valsamis and X. Qin, Persistent stores and hybrid systems, Proceedings of the 32st IEEE Conference on Decision and Control, IEEE Press, 1993, pp. 2298-2302.
26. R. L. Grossman, Working With Object Stores of Events Using PTool, 1993 Cern Summer School in Computing, C. E. Vandoni and C. Verkerk, editors, CERN-Service d'Information Scientifique 94-06, pages 66-97, 1994.
27. R. L. Grossman and X. Qin, Ptool: a scalable persistent object manager, Proceedings of SIGMOD 94, ACM, 1994, page 510.
28. R. L. Grossman, A. Sundaram, H. Ramamoorthy, M. Wu, S. Hogan, J. Shuler and O. Wolfson, Viewing the U.S. Government Budget as a Digital Library, Proceedings of Digital Libraries 1994: Conference on the Theory and Practice of Digital Libraries, ACM, 1994.
29. R. L. Grossman, W. Sluis, and W. Shadwick, On Nonlinear Normal Forms, Proceedings of the 33st IEEE Conference on Decision and Control, IEEE Press, 1994.
30. D. R. Quarrie, C. T. Day, S. Loken, J. F. Macfarlane, D. Lifka, E. Lusk, D. Malon, E. May, L. E. Price, L. Cormell, A. Gauthier, P. Liebold, J. Hilgart, D. Liu, J. Marstaller, U. Nixdorf, T. Song, R. Grossman, X. Qin, D. Valsamis, M. Wu, W. Xu, A. Baden, The PASS Project: A Progress Report, Proceedings of the Conference on Computing in High Energy Physics 1994, edited by S. C. Loken, pages 229-232, 1995.

31. D. R. Quarrie, C. T. Day, S. Loken, J. F. Macfarlane, D. Lifka, E. Lusk, D. Malon, E. May, L. E. Price, L. Cornell, A. Gauthier, P. Liebold, J. Hilgart, D. Liu, J. Marsteller, U. Nixdorf, T. Song, R. Grossman, X. Qin, D. Valsamis, M. Wu, W. Xu, A. Baden, The PASS Project Architectural Model, Proceedings of the Conference on Computing in High Energy Physics 1994, edited by S. C. Loken, pages 233-235, 1995.
32. E. N. May, D. Lifka, D. Malon, L. E. Price L. Cornell, A. Gauthier, J. Marsteller, S. Mestad, U. Nixdorf R. Grossman, X. Qin, D. Valsamis, M. Wu, W. Xu A Demonstration of a Multi-level Object Store and its Application to the Analysis of High Energy Physics Data, Proceedings of the Conference on Computing in High Energy Physics 1994, edited by S. C. Loken, pages 236-238, 1995.
33. D. Malon, D. Lifka, E. May R. Grossman, X. Qin, W. Xu Parallel Query Processing for Event Store Data, Proceedings of the Conference on Computing in High Energy Physics 1994, edited by S. C. Loken, pp. 239-240, 1995.
34. R. L. Grossman, N. Araujo, X. Qin, and W. Xu, Managing physical folios of objects between nodes, Persistent Object Systems (Proceedings of the Sixth International Workshop on Persistent Object Systems), M. P. Atkinson, V. Benzaken and D. Maier, editors, Springer-Verlag and British Computer Society, 1995, pages 217-231.
35. R. L. Grossman, X. Qin, D. Valsamis, W. Xu, C. T. Day, S. Loken, J. F. MacFarlane, D. Quarrie, E. May, D. Lifka, D. Malon, L. Price, Analyzing High Energy Physics Data Using Databases: A Case Study, Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management, IEEE Press, 1994, pages 283-286.
36. R. L. Grossman, A. Nerode, and W. Kohn, Nonlinear Systems, Automata, and Agents: Managing their Symbolic Data Using Light Weight Persistent Object Managers, International Symposium on Fifth Generation Computer Systems, 1994: Workshop on Heterogeneous Cooperative Knowledge-Bases, Kazumasa Yokota, editor, ICOT, pages 65-74.
37. J. Leigh, C. A. Vasilakis, T. A. DeFanti, R. Grossman, C. Assad, B. Rasnow, A. Protopappas, E. DeSchutter, J. M. Bower, Virtual Reality in Computational Neuroscience, Virtual Reality Applications, edited by R. Earnshaw, J. A. Vince and H. Jones, Academic Press, London, 1995, pages 293-306.
38. R. L. Grossman D. Hanley, and X. Qin Caching and migration for physical collections of objects: Interfacing persistent object stores and hierarchical storage systems, in Proceedings of the 14th IEEE Computer Society Mass Storage Systems Symposium, S. Coleman, editor, IEEE, 1995, pages 127-135.
39. R. L. Grossman, D. Hanley, and X. Qin, PTool: A Light Weight Persistent Object Manager, Proceedings of SIGMOD 95, ACM, 1995, p. 488.
40. R. L. Grossman and M. Sweedler Hybrid Systems and Quantum Automata: Preliminary Announcement, Hybrid Systems II, P. Antsaklis, W. Kohn, A. Nerode, S. Sastry, editors, Springer Lecture Notes in Computer Science, Volume 999, pages 191-201, 1995.
41. M. J. Doffou and R. L. Grossman, The Symbolic Computation of Differential Invariants of Polynomial Vector Field Systems Using Trees, Proceedings of the 1995 International Symposium on Symbolic and Algebraic Computation, A. H. M. Levelt, editor, ACM, 1995, pages 26-31.
42. N. Araujo, R. Grossman, D. Hanley, W. Xu, S. Ahn, K. Denisenko, M. Fischler, M. Galli D. Malon and E. May, Some Remarks on Parallel Data Mining Using a Persistent Object Manager, Proceedings of the Conference on Computing in High Energy Physics 1995.

43. S. Bailey, R. Grossman, and D. Hanley, D. Benton and B. Hollebeek, Scalable Digital Libraries of Event Data and the NSCP Meta-Cluster, Proceedings of the Conference on Computing in High Energy Physics 1995.
44. R. L. Grossman and H. V. Poor, Optimization Driven Data Mining and Credit Scoring, in Proceedings of the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFEr), IEEE, Piscataway, 1996, pages 104-110.
45. S. Bailey, R. L. Grossman, L. Gu, and D. Hanley, A Data Intensive Approach to Path Planning and Mode Management for Hybrid Systems, in R. Alur, T. A. Henzinger, and E. Sontag, Hybrid Systems III, Proceedings of the DIMACS Workshop on Verification and Control of Hybrid Systems, Springer-Verlag, LNCS 1066, 1996.
46. R. L. Grossman, H. Bodek, D. Northcutt, and H. V. Poor, Data Mining and Tree-based Optimization, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD 1996), E. Simoudis, J. Han and U. Fayyad, editors, AAAI Press, Menlo Park, California, 1996, pp 323-326.
47. R. L. Grossman, The Terabyte Challenge: An Open, Distributed Testbed for Managing and Mining Massive Data Sets, Proceedings of the 1996 Conference on Supercomputing, IEEE, 1996.
48. R. L. Grossman, S. Bailey and D. Hanley, Data Mining Using Light Weight Object Management in Clustered Computing Environments, Proceedings of the Seventh International Workshop on Persistent Object Stores, Morgan-Kaufmann, San Mateo, 1997, pages 237-249.
49. S. Bailey, A. Goldstein, R. L. Grossman, and D. Hanley, Accessing Warehoused Collections of Objects Through Java, Proceedings of the First International Workshop on Persistence and Java, Sun Microsystems, 1996.
50. S. Bailey and R. L. Grossman, JTool: Accessing Warehoused Collections of Objects with Java, Proceedings of the Second Workshop on Persistence and Java, Sun Microsystems, 1998.
51. R. L. Grossman and S. Bailey, An Overview of Dynamic Classification: Mining Collections of Trajectories (invited paper), 1998 Proceedings of the Section on Physical and Engineering Sciences, American Statistical Association, Alexandria, Virginia, pages 24-28.
52. R. L. Grossman, S. Bailey, A. Ramu, B. Malhi and A. Turinsky, The Preliminary Design of Papyrus: A System for High Performance, Distributed Data Mining over Clusters, in Advances in Distributed and Parallel Knowledge Discovery, H. Kargupta and P. Chan, editors, AAAI Press/The MIT Press, Menlo Park, California, 2000, pages 259-275.
53. R. L. Grossman, S. Bailey, A. Ramu, B. Malhi and H. Sivakumar, A. Turinsky, Papyrus: A System for Data Mining over Local and Wide Area Clusters and Super-Clusters, Proceedings of Supercomputing 1999, IEEE.
54. R. L. Grossman, The Role of QoS in Wide Area Data Mining, Proceedings of the First Internet 2 Joint Applications Engineering QoS Workshop: Enabling Advanced Applications Through QoS, UCAID, 1999, pages 19-21.
55. J. Leigh, A. Johnson, T. DeFanti, S. Bailey, R. L. Grossman, A Methodology for Supporting Collaborative Exploratory Analysis of Massive Data Sets in Tele-Immersive Environments, 8th IEEE International Symposium on High Performance and Distributed Computing, Redondo Beach, California, Aug 3-6, 1999.
56. J. Leigh, A. Johnson, T. DeFanti, S. Bailey, R. L. Grossman, A Tele-Immersive Environment for Collaborative Exploratory Analysis of Massive Data Sets, ASCI 99, pages 3-9, Heijen, the Netherlands, 1999.

57. S. Bailey, E. Creel, R. Grossman, S. Gutti, and H. Sivakumar, A High Performance Implementation of the Data Space Transfer Protocol (DSTP), Large-Scale Parallel Data Mining, M. J. Zaki and C.-T. Ho, editors, Springer-Verlag, Berlin, 2000, pages 55-64.
58. R. L. Grossman and Yike Guo, Parallel Methods for Scaling Data Mining Algorithms to Large Data Sets, Handbook on Data Mining and Knowledge Discovery, Jan M Zytkow, editor, Oxford University Press, 2002, pages 433 - 442.
59. H. Sivakumar, R. Grossman, B. Schiefer, X. Xue, and S. Syed, Performance of DB2 UDB EEE on NT with Virtual Interface Architecture, Lecture Notes in Computer Science: Advances in Database Technology EDBT 2000, 7th International Conference on Extending Database Technology Konstanz, Germany March 2000.
60. H. Sivakumar, S. Bailey, R. L. Grossman, Pockets: The Case for Application-level Network Striping for Data Intensive Applications using High Speed Wide Area Networks, Proceedings of the 2000 ACM/IEEE Conference on Supercomputing (CDROM), IEEE Computer Society, Washington, DC, USA, 2000, page 38.
61. N. Sawant, C. Scharver, J. Leigh, A Johnson, G. Reinhart, E. Creel, S. Batchu, S. Bailey, R. L. Grossman, The Tele-Immersive Data Explorer: A Distributed Architecture for Collaborative Interactive Visualization of Large Data-sets, 4th International Immersive Projection Technology Workshop, Ames, Iowa, June 19-20, 2000.
62. R. L. Grossman and R. Hollebeck, The National Scalable Cluster Project: Three Lessons about High Performance Data Mining and Data Intensive Computing, in Handbook of Massive Data Sets, J. Abello, P. M. Pardalos, and M. G. C. Resende, editors, Kluwer Academic Publishers, 2002.
63. A DataSpace Infrastructure for Astronomical Data, Robert Grossman, Emory Creel, Marco Mazzucco, Roy Williams in R. L. Grossman, C. Kamath, W. Philip Kegelmeye, V. Kumar, and R. Namburu, Data Mining for Scientific and Engineering Applications, Kluwer Academic Publishers, 2001, pages 115-123.
64. Robert Grossman, Mark Hornick, and Gregor Meyer, Emerging Standards and Interfaces in Data Mining, Handbook of Data Mining, Nong Ye, editor, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 2003, pages 453-459.
65. M. Cornelson, E. Greengrass, R. L. Grossman, R. Karidi, and D. Shnidman, Combining Information Retrieval Algorithms Using Machine Learning, Survey of Text Mining: Clustering, Classification, and Retrieval Michael W. Berry, editor, Springer-Verlag, 2003, pages 159-169.
66. Marco Mazzucco, Asvin Ananthanarayan, Robert L. Grossman, Jorge Levera, and Gokulnath Bhagavantha Rao, Merging Multiple Data Streams on Common Keys over High Performance Networks, Proceedings of the IEEE/ACM SC2002 Conference, 2002, IEEE Computer Society, page 67.
67. R. L. Grossman and R. G. Larson, An Algebraic Approach to Data Mining: Some Examples, Proceedings of the 2002 IEEE International Conference on Data Mining, IEEE Computer Society, Los Alamitos, California, 2002, pages 613-616.
68. Robert L. Grossman, Yunhong Gu, Dave Hanley, Xinwei Hong, Dave Lillethun, Jorge Levera, Joe Mambretti, Marco Mazzucco, and Jeremy Weinberger, Photonic Data Services: Integrating Path, Network and Data Services to Support Next Generation Data Mining Applications, Data Mining: Next Generation Challenges and Future Directions, H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha, editors, AAAI Press, 2004.

69. Robert L. Grossman, Yunhong Gu, Dave Hanley, Xinwei Hong, Dave Lillethun, Jorge Levera, Joe Mambretti, Marco Mazzucco, and Jeremy Weinberger, Global Access to Large Distributed Data Sets using Photonic Data Services, Proceedings of the 20th IEEE/11th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2003), IEEE Computer Society, Los Alamitos, California, pages 62-66.
70. Thomas A. DeFanti, Jason Leigh, Maxine D. Brown, Daniel J. Sandin, Oliver Yu, Chong Zhang, Rajvikram Singh, Eric He, Javid Alimohideen, Naveen K. Krishnaprasad, Robert Grossman, Marco Mazzucco, Larry Smarr, Mark Ellisman, Phil Papadopoulos, Andrew Chien, John Orcutt, Teleimmersion and Visualization with the OptIPuter, Proceedings of the 12th International Conference on Artificial Reality and Telexistence (ICAT 2002), Ohmsha/IOS Press.
71. R. Grossman, X. Qin, W. Xu, H. Hulen, and T. Tyler, An Architecture for a Scalable, High-Performance Digital Library, 14th IEEE Symposium on Mass Storage Systems, IEEE Press, pages 89-98, 1995.
72. Robert L. Grossman and Dave Northcutt, A Note on Interfacing Object Warehouses and Mass Storage Systems for Data Mining Applications, Proceedings of the Goddard Conference on Mass Storage Systems, 1996.
73. Shitij Mutreja, Stuart Bailey, Robert Grossman, and Dave Hanley, Lightweight Video Service for Multi-Media Digital Libraries, Proceedings of CASCON '95, 1995.
74. Robert Grossman, Donald Hamelberg, Pavan Kasturi, and Bing Liu, Experimental Studies of the Universal Chemical Key (UCK) Algorithm on the NCI Database of Chemical Compounds, Proceedings of the 2003 IEEE Computer Society Bioinformatics Conference (CSB 2003), IEEE Computer Society, Los Alamitos, California, pages 244-250.
75. Bing Liu, Robert L. Grossman and Yanhong Zhai, Mining Data Records in Web Pages, Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2003), pages 601-606.
76. Robert L. Grossman, Alert Management Systems: A Quick Introduction, in Managing Cyber Threats: Issues, Approaches and Challenges, edited by Vipin Kumar, Jaideep Srivastava and Aleksandar Lazarevic, Springer Science+Business Media, Inc., New York, 2005, pages 281-291, ISBN 0-387-24226-0.
77. R.L. Grossman, Y. Gu, D. Hanley, X. Hong, and G. Rao, Open DMIX - Data Integration and Exploration Services for Data Grids, Data Web and Knowledge Grid Applications, Proceedings of the First International Workshop on Knowledge Grid and Grid Intelligence (KGGI 2003), W. K. Cheung and Y. Ye, editors, pages 16-28.
78. S. Bailey, R. L. Grossman and D. Hanley, Clusters, meta-clusters, and digital libraries: digital libraries for scientific, engineering and medical applications, ACM SIGWEB Newsletter, Volume 4, Number 2, ACM Press, New York, NY, 1995, pages 8-10.
79. Chetan Gupta and Robert L. Grossman, GenIc: A Single Pass Generalized Incremental Algorithm for Clustering, 2004 SIAM International Conference on Data Mining (SDM 04), to appear.
80. Robert L. Grossman and Richard G. Larson, Bialgebras and Realizations, in Hopf Algebras, Jeffrey Bergen, Stefan Catoiu, and William Chin, editors, Marcel Dekker, Inc., New York, 2004, pages 157-166.
81. Robert L. Grossman, Yunhong Gu, Chetan Gupta, David Hanley, Xinwei Hong, and Parthasarathy Krishnaswamy, Open DMIX: High Performance Web Services for Distributed Data Mining, 7th International Workshop on High Performance and Distributed Mining, in association with the Fourth International SIAM Conference on Data Mining, 2004.

82. Jorge Levera, Benjamin Barin, and Robert Grossman, Experimental Studies Using Median Polish Procedures to Reduce Alarm Rates in Data Cubes of Intrusion Data, *Intelligence and Security Informatics for National and Homeland Security*, Hsinchun Chen, Reagan Moore, Daniel Zeng, John Jeavitt, editors, LNCS 3073, Springer Verlag, New York, 2004, pages 482-491.
83. Robert L. Grossman, Dave Hanley, Xinwei Hong and Parthasarathy Krishnaswamy, Using DataSpace to Support Long-Term Stewardship of Remote and Distributed Data, *NASA/IEEE MSST 2004, 12th NASA Goddard/21st IEEE Conference on Mass Storage Systems and Technologies*, 2004, pages 239-244.
84. Yunhong Gu, Xinwei Hong, and Robert Grossman, Experiences in Design and Implementation of a High Performance Transport Protocol, *ACM/IEEE International Conference for High Performance Computing and Communications (SC '04)*, page 22.
85. Andrei L. Turinsky and Robert L. Grossman, A Greedy Algorithm for Selecting Models in Ensembles, *Proceedings 4th IEEE International Conference Data Mining (ICDM 2004)*, Brighton, UK, pages 547-550, IEEE Computer Society Press, 2004.
86. Yunhong Gu, Xinwei Hong and Robert Grossman, An Analysis of AIMD Algorithms with Decreasing Increases, *Proceedings of GridNets 2004*, IEEE Press, 2004.
87. Parthasarathy Krishnaswamy, Stephen G. Eick, Robert L Grossman, Visual Browsing of Remote and Distributed Data, *IEEE Symposium on Information Visualization (INFOVIS'04)*, 2004, page 12.
88. Yunhong Gu and Robert L. Grossman, Optimizing UDP-Based Protocol Implementations, *Proceedings of the Third International Workshop on Protocols for Fast Long-Distance Networks PFLDnet 2005*, 2005.
89. Greeshma Neglur and Robert L. Grossman, Assigning Unique Keys to Chemical Compounds for Data Integration: Some Interesting Counter Examples, *2nd International Workshop on Data Integration in the Life Sciences (DILS 2005)*, La Jolla, July 20-22, 2005.
90. Joseph Bugajski, Robert L. Grossman, Eric Sumner and Tao Zhang, An Event Based Framework for Improving Information Quality That Integrates Baseline Models, Causal Models and Formal Reference Models, *Second International ACM SIGMOD Workshop on Information Quality in Information Systems (IQIS 2005)*, June 17th, Baltimore, Maryland, co-located with ACM SIGMOD/PODS 2005.
91. Robert L. Grossman, Michal Sabala, Javid Alimohideen, Anushka Aanand, John Chaves, John Dillenburg, Steve Eick, Jason Leigh, Peter Nelson, Mike Papka, Doug Rorem, Rick Stevens, Steve Vejckik, Leland Wilkinson, and Pei Zhang, Real Time Change Detection and Alerts from Highway Traffic Data, *ACM/IEEE International Conference for High Performance Computing and Communications (SC '05)*.
92. Yunhong Gu and Robert Grossman, Supporting Configurable Congestion Control in Data Transport Services, *ACM/IEEE International Conference for High Performance Computing and Communications (SC '05)*.
93. Joseph Bugajski, Robert Grossman, Eric Sumner, Tao Zhang, A Methodology for Establishing Information Quality Baselines for Complex, Distributed Systems, *10th International Conference on Information Quality (ICIQ)*, 2005.
94. L. Wilkinson, A. Anand and R. Grossman, Graph-theoretic scagnostics, *Proceedings of the IEEE Information Visualization 2005 (INFOVIS'05)*, pages 157-164.

95. Rajmonda Sulo, Stephen Eick, Robert Grossman, DaVis: A tool for Visualizing Data Quality, Proceedings of the IEEE Information Visualization 2005 (INFOVIS'05).
96. Yong Mao, Yunhong Gu, Jia Chen and Robert L. Grossman, SDCS: Simplified Data Communications in Parallel/Distributed Applications, IEEE International Symposium on Cluster Computing and the Grid (CCGrid06), pages 292-295, 2006.
97. Greeshma Neglur, Robert L. Grossman, Natalia Maltsev, and Clement Yu, Using Term Lists and Inverted Files to Improve Search Speed for Metabolic Pathway Databases, 3rd International Workshop on Data Integration in the Life Sciences 2006 (DILS'06), Lecture Notes in Bioinformatics, Volume 4075, Springer-Verlag, Berlin, 2006, pages 168-184.
98. Yunhong Gu, Robert L. Grossman, Alex Szalay and Ani Thakar, Distributing the Sloan Digital Sky Survey Using UDT and Sector, Proceedings of e-Science 2006.
99. Joseph Bugajski, Robert L. Grossman, Eric Sumner and Steve Vejcik, Monitoring Data Quality for Very High Volume Transaction Systems, Proceedings of the 11th International Conference on Information Quality, 2006.
100. Rajmonda Sulo, Anushka Anand, Leland Wilkinson, Robert Grossman, Stephen Eick, Topographically-Based Real-time Traffic Anomaly Detection in a Metropolitan Highway System, Proceedings of the IEEE Information Visualization 2006 (INFOVIS'06).
101. Joseph M. Bugajski, Robert L. Grossman, and Steve Vejcik, A Service Oriented Architecture Supporting Data Interoperability for Payments Card Processing Systems, Proceedings of the International Conference on Service-Oriented Computing (ICSOC) 2006, Springer Lecture Notes in Computer Science, Volume 4296, 2006, pages 591-600.
102. Yong Mao, Yunhong Gu, Jia Chen, and Robert L. Grossman, FastPara: A High-Level Declarative Data-Parallel Programming Framework on Clusters, Parallel and Distributed Computing and Systems, 2006.
103. Joseph Bugajski, Chris Curry, Robert L. Grossman, David Locke, Steve Vejcik, Detecting Changes in Large Data Sets of Payment Card Data: A Case Study, Proceedings of The Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007), ACM, 2007.
104. G. Caire, Robert L. Grossman and H. Vincent Poor, Wavelet transforms associated with finite cyclic groups, The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers, Volume 1, pages 113 - 119, 1992.
105. Fang Fang, Robert L. Grossman and Ziangjun Liu, An Algorithm for Assigning Unique Keys To Metabolic Pathways, Proceedings of the 2007 IEEE International Conference on Bioinformatics and Biomedicine, pages 374-382, 2007.
106. Yunhong Gu, Robert L. Grossman and Joe Mambretti, A Peer-to-Peer Infrastructure for Distributing Large Scientific Data Sets over Wide Area High-Performance Networks: Experimental Studies Using Wide Area Layer 2 Services, Proceedings of the First International Conference on Networks for Grid Applications (GridNets 2007), ICST, ISBN: 978-963-9799-02-8, 2007.
107. Joseph Bugajski and Robert L. Grossman, An Alert Management Approach to Data Quality: Lessons Learned from the Visa Data Authority Program, Proceedings of the 12th International Conference on Information Quality, (ICIQ 2007).
108. Joseph Bugajski, Chris Curry, Robert L. Grossman, David Locke and Steve Vejcik, Data Quality Models for High Volume Transaction Streams: A Case Study, Proceedings of the Second Workshop on Data Mining Case Studies and Success Stories, ACM 2007.

109. Robert L. Grossman, A Review of Some Analytic Architectures for High Volume Transaction Systems, The 5th International Workshop on Data Mining Standards, Services and Platforms (DM-SSP '07), ACM, 2007, pages 23-28.
110. Chetan Gupta and Robert L. Grossman, Outlier Detection with Streaming Dyadic Decomposition, Proceedings of the 7th Industrial Conference on Data Mining, LNCS Volume 4597, Springer-Verlag, 2007, pages 77-91.
111. David Ferrucci, Robert L. Grossman, Anthony Levas, PMML and UIMA Based Frameworks For Deploying Analytic Applications and Services, Proceedings of the 4th International Workshop on Data Mining Standards, Services and Platforms (DM-SSP 06), ACM, New York, 2006, pages 14-26.
112. Robert L Grossman and Yunhong Gu, Data Mining Using High Performance Clouds: Experimental Studies Using Sector and Sphere, Proceedings of The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), ACM, 2008, pages 920-927.
113. Yunhong Gu and Robert Grossman, UDTv4: Improvements in Performance and Usability, Proceedings of GridNets 2008, Springer-Verlag 2008.
114. Yunhong Gu and Robert Grossman, Exploring Data Parallelism and Locality in Wide Area Networks, Proceedings of the Workshop on Many-task Computing on Grids and Supercomputers (MTAGS), IEEE, 2008, pages 1-10.
115. Robert L Grossman, Michal Sabala, Yunhong Gu, Anushka Anand, Matt Handley, Rajmonda Sulo and Lee Wilkinson, Discovering Emergent Behavior from Network Packet Data: Lessons From the Angle Project, in Next Generation Data Mining, edited by Hillol Kargupta, Jiawei Han, Philip S Yu, Rajeev Motwani and Vipin Kumar, CRC Press, Boca Raton, 2009, pages 243-260.
116. Robert Grossman, Yunhong Gu, Michal Sabala, Collin Bennett, Jonathan Seidman and Joe Mambretti, The Open Cloud Testbed: A Wide Area Testbed for Cloud Computing Utilizing High Performance Network Services, GridNets 2009, Springer-Verlag, 2009.
117. Yunhong Gu and Robert L Grossman, Lessons Learned From a Year's Worth of Benchmarks of Large Data Clouds, 2nd Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS 2009), ACM, 2009.
118. Wenxuan Gao, Robert Grossman, Philip Yu, and Yunhong Gu, Why Naive Ensembles Do Not Work in Cloud Computing, Proceedings of the The First Workshop on Large-scale Data Mining: Theory and Applications (LDMTA 2009), 2009.
119. Collin Bennett, Robert L. Grossman, David Locke, Jonathan Seidman and Steve Vejcek, Mal-Stone: Towards a Benchmark for Analytics on Large Data Clouds, The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010), ACM, 2010.
120. Robert L. Grossman, Yunhong Gu, Joe Mambretti, Michal Sabala, Alex Szalay, and Kevin White, An Overview of the Open Science Data Cloud, Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (HPDC '10), ACM, 2010.

Books

1. M. Beals, C. Fefferman, R. Grossman, Strictly Pseudoconvex Domains in C_n , Moscow, 1987, 286 pages, in Russian. Translation of Article 2.
2. R. Grossman, editor, Symbolic Computation: Applications to Scientific Computing, SIAM, Philadelphia, 1989, 185 pages.

3. R. L. Grossman, A. Nerode, A. P. Ravn, and H. Rischel, editors, Hybrid Systems, Lecture Notes in Computer Science, Volume 736, Springer-Verlag, New York, 1993.
4. Yike Guo and Robert Grossman, editors, High Performance Data Mining: Scaling Algorithms, Applications and Systems, Kluwer Academic Publishers, 1999.
5. R. L. Grossman, J. Han and V. Kumar, editors, Proceedings of the SIAM First International Conference on Data Mining (SDM-01), SIAM, 2001, ISBN 0-89871-495-8.
6. Robert L. Grossman, Chandrika Kamath, Philip Kegelmeyer, Vipin Kumar, and Raju R. Namburu, Data Mining for Scientific and Engineering Applications, Kluwer Academic Publishing, 2001. ISBN 1-4020-0033-2.
7. Robert Grossman, Jiawei Han, Vipin Kumar, Heikki Mannila, and Rajeev Motwani, editors, Proceedings of the Second SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2002.
8. Robert L. Grossman, Roberto Bayardo, Kristin Bennet, and Jaideep Vaidya, editors, Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2005), ACM Press, New York, 2005, ISBN 1-59593-135-X.

Other Publications

1. Robert Grossman, Simon Kasif, Reagan Moore, David Rocke, and Jeff Ullman, Data Mining Research: Opportunities and Challenges. A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data, <http://www.ncdm.uic.edu/m3d2.htm>, 1998.
2. Haim Bodek, Robert Lee Grossman and Ivan Pulleyn, Detecting Network Intrusions through the Data Mining of Network Packet Data Using the ACT Algorithm, 1997.
3. Robert L. Grossman, Symbolic Computation and Flows of Differential Equations, Second Workshop on Computer Algebra, July 21, 1993, Rio de Janeiro.
4. Andrew Baden and Robert L. Grossman, A Model for Computing at the SCC, SSC Technical Report, June 6, 1990.
5. KDD-2003 Workshop on Data Mining Standards, Services, and Platforms (DM-SSP 03), ACM SIGKDD Explorations, Volume 5, Issue 2, page 197, 2003.
6. Robert L. Grossman and David Radford, Bialgebra Deformations of Certain Universal Enveloping Algebras, Laboratory for Advanced Computing Technical Report, University of Illinois at Chicago, 1991.
7. This technical report is now archaic.
8. R. L. Grossman, S. Mehta and X. Qin, Path planning by querying persistent stores of trajectory segments, Laboratory for Advanced Computing Technical Report Number LAC 93-R3, September, 1992.
9. Jason Leigh, Eric He, and Robert Grossman, Grid Networks and UDT Services, Protocols, and Technologies, in Franco Travostino, Joe Mambretti, and Gigi Karmous-Edwards, editors, Grid Networks: Enabling Grids with Advance Communication Technology, Wiley, 2006, pages 171-184.
10. Robert L. Grossman, Yunhong Gu, Michal Sabala, and Joel J. Mambretti, Real Time, Distributed Detection of Anomalies and Emergent Behavior Using the Angle Algorithm, University of Illinois at Chicago, Laboratory for Advanced Computing, Technical Report, 2006.

11. Robert L. Grossman, Fifth International Workshop on Data Mining Standards, Services, and Platforms, Preface, Proceedings of The Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2007.
12. Bennett Bertenthal, Robert Grossman, David Hanley, Mark Hereld, Sarah Kenny, Gina Levow, Michael E. Papka, Steve Porges, Kavithaa Rajavenkateshwaran, Rick Stevens, Thomas Uram, and Wenjun Wu, Social Informatics Data Grid, Third International Conference on e-Social Science October 7-9, 2007, Ann Arbor, Michigan.
13. This technical report is now archaic.
14. Gregory Piatetsky-Shapiro, Robert Grossman, Chabane Djeraba Ronen Feldman, Lise Getoor, Mohammed Zaki Is There A Grand Challenge or X-Prize for Data Mining?, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2006, pages 954 - 956. See also SIGKDD Explorations, Volume 8, Number 2, 2006.
15. Robert L Grossman and Yunhong Gu, On the Varieties of Clouds for Data Intensive Computing, Bulletin of the Technical Committee on Data Engineering, March 2009, Volume 32, Number 1, pages 44-50.
16. John Chaves, Chris Curry, Robert L. Grossman, David Locke and Steve Vejcik, Augustus: the Design and Architecture of a PMML-based Scoring Engine, in Proceedings of the 4th International Workshop on Data Mining Standards, Services and Platforms (DMSSP '06), ACM, New York, NY, 38-46.
17. Robert L. Grossman, What is Analytic Infrastructure and Why Should You Care?, SIGKDD Explorations, July 2009, Volume 11, Issue 1, pages 5-9.

Funded Research

- PIRE: Training and Workshops in Data Intensive Computing Using The Open Science Data Cloud. Sponsor: NSF. Role: PI. Award: OISE 1129076. Period: March 7, 2011 - February 10, 2016. Amount: \$3,489,528.
- P50: Center for Transcriptional Network Dynamics and Evolution. Sponsor: NIH. Role: Co-Principal Investigator. Award: P50 081892. Period: September 1, 2008 - August 31, 2013. Amount: \$930,953.
- Chicago Biomedical Consortium Funds for Center for Transcriptional Network Dynamics and Evolution. Sponsor: Chicago Biomedical Consortium. Role: Principal Investigator. Period: July 1, 2008 - June 30, 2011. Amount: \$494,544.
- Visually-Motivated Characterization of Point Sets Embedded in High-Dimensional Geometric Spaces. Sponsor: National Science Foundation. Role: Co-Principal Investigator. Period: July 1, 2008 - June 30, 2011. Amount: \$475,000.
- The Teraflow Network: Instrumentation to Support Experimental Studies of High Volume Data Flows. Role: Principal Investigator. Sponsor: Army Research Office. Period: September 23, 2008 - September 22, 2009. Amount: \$150,000.
- The Teraflow Network: Instrumentation to Support Experimental Studies of High Volume Data Flows. Role: Principal Investigator. Sponsor: Office of Naval Research. Period: September 23, 2008 - September 22, 2009. Amount: \$479,509.

- SCI: II: The TeraFlow Project: High Performance Flows for Mining Large Distributed Data Archives. Sponsor: NSF. Role: Principal Investigator. Award: SCI-0430781. Period: October 1, 2004 - September 30, 2010. Amount: \$3,226,050.
- ITR: Collaborative Research: A Data Mining and Exploration Middleware for Grid and Distributed Computing. Sponsor: NSF. Role: Principal Investigator. Award Number: ACI-0325013. Period: November 1, 2003 - October 31, 2009. Amount: \$315,699.
- Cyberinfrastructure for Collaborative Research in the Social and Behavioral Sciences. Sponsor: NSF. Role: Co-Principal Investigator. Award Number: 0537849. Period: October 1, 2005 - September 30, 2007. Amount: \$1,011,442.
- MRI: International Data Mining Grid Testbed for Research in High Performance Data Transport, Data Integration, and Data Exploration – Instrument Development Proposal. Sponsor: NSF. Role: Principal Investigator. Award: CNS-0420847. Period: September 1, 2004 - August 31, 2007. Amount: \$237,000.
- Secure Agency Interoperation for Effective Data Mining in Border Control and Homeland Security Applications. Sponsor: National Science Foundation. Role: Co-Principal Investigator. Award Number: EIA-0306838. Amount: \$1,064,697. Period: September 1, 2003 - August 31, 2006
- Detecting Distributed Network Intrusions using Real-Time Data Mining. Sponsor: Department of Defense. Role: Principal Investigator. Award Number: MDA904-03-C-0620. Amount: \$130,098. Period: March 28, 2003 - March 27, 2004
- CISE Research Resources: Matching Advanced Visualization and Intelligent Data Mining to High-Performance Experimental Networks. Sponsor: National Science Foundation. Role: Co-Principal Investigator. Award Number: ANI-0224306. Amount: \$850,000. Period: October 1, 2001 - September 30, 2005.
- Tera Mining: A Testbed for Distributed Data Mining over High Performance SONET and Lambda Networks. Sponsor: National Science Foundation. Role: Principal Investigator. Award Number: ANI-0129609. Amount: \$660,000. Period: March 1, 2002 - February 28, 2005.
- ITR: The OptIPuter. Sponsor: National Science Foundation. Role: Senior Scientist. Award Number: ANI-0225642. Amount: \$13,500,000. Period: October 1, 2002 - September 30, 2007.
- Intelligent Computational Genomic Analysis. Sponsor: National Science Foundation. Role: Co-Principal Investigator. Award: MCD-9980088. Amount: \$1,700,000. Period: September 1, 2000 - February 28, 2003.
- REU Supplement for the Terabyte Challenge: Developing Software Tools and Network Services for Mining Remote and Distributed Data Over High Performance Networks. Sponsor: National Science Foundation. Role: Principal Investigator. Award Number: ANI-9977868. Amount: \$27,807. Period: June 1, 2000 - July 31, 2004
- The Terabyte Challenge: Developing Software Tools and Network Services for Mining Remote and Distributed Data over High Performance Networks. Sponsor: National Science Foundation. Role: Principal Investigator. Award Number: ANI-9977868. Amount: \$2,410,401. Period: September 1, 1999 - July 31, 2005.
- Cavern - The Cave Research Network. Sponsor: National Science Foundation. Role: Co-Principal Investigator. Award Number: EIA-9802090. Amount: \$2,000,000. Period: October 1, 1998 - September 30, 2003

- The High Performance and Wide Area Analysis and Mining of Scientific and Engineering Data. Sponsor: Department of Energy. Amount: \$825,000. Period: September 1, 1998 - September 14, 2001.
- CISE Research Infrastructure. Sponsor: National Science Foundation. Award Number: 9802090. Role: Co-Investigator. Amount: \$2,000,000. Period October 1, 1998-September 30, 2003.
- Data Intensive Computing of Hybrid Systems Data. Sponsor: NASA. Role: Principal Investigator. Amount: \$180,000. January 1, 1998 - December 31, 2000.
- Academic Infrastructure Grant. Sponsor: National Science Foundation. Role: Co-Investigator. Amount: \$4,000,000 Period: November 1, 1996 - October 31, 1999.
- Computing with Persistent Stores of Scientific Data. Sponsor: Department of Energy. Role: Principal Investigator. Amount: \$825,000. Period: April 1, 1995-March 31, 1998.
- Persistent Stores of Multi-media data. IBM Principal Investigator. \$75,000 January 1, 1995 - December 31, 1998.
- Academic Infrastructure Grant. Sponsor: National Science Foundation (CDA 9413948) Role: Co-Investigator. Amount: \$4,000,000. Period: November 15, 1994-October 30, 1996.
- CISE Research Infrastructure. Sponsor: National Science Foundation (CDA 9303433) Role: Co-Investigator. Amount; \$2,000,000. Period July 1, 1993-June 30, 1998.
- Computing with persistent stores of scientific objects Sponsor: National Science Foundation (IRI 9224605) Role: Principal Investigator. Amount: \$401,000 Period: July 1, 1993-June30, 1996.
- Database computing for the SSC. Sponsor: Department of Energy (ER-25133) Role; Principal Investigator. Amount: \$575,000. Period: September, 1992 - March 31, 1995.
- Symbolic computation and differential equations. Sponsor: National Science Foundation. Role: Principal Investigator. Amount: \$178,010. Period: November, 1991 - April, 1993.
- Design of data model for the analysis of HEP data Using Database Computing. Sponsor: Argonne National Laboratory. Role: Principal Investigator. Amount: \$110,171 Period: October, 1991 - June, 1993.
- The analysis of control trajectories using symbolic and database computing. Sponsor: NASA-Ames Research Center (NAG 2-513) Role: Principal Investigator. Amount: \$230,000. Period: January, 1991 - December, 1993.
- The symbolic computation of trajectories. Sponsor: National Science Foundation. Role: Principal Investigator. Amount: \$20,000. Period: August, 1989 - January, 1991.
- Symbolic computation and the automated analysis of trajectories of control systems. Sponsor: NASA-Ames Research Center. Role: Principal Investigator. Amount: \$124,899. Period: July, 1988 - December, 1990.
- A symposium on the use of symbolic methods to solve algebraic and geometric problems arising in engineering. Sponsor: NASA-Ames Research Center. Role: Project Director. Amount: \$5000. Period: January - June, 1987.
- A symposium on the use of symbolic methods to solve algebraic and geometric problems arising in engineering. Sponsor: University of California Coordinating Committee on Nonlinear Science. Role: Project Director. Amount: \$5000. Period: January - June, 1987.
- Postdoctoral Research Fellowship. Sponsor: National Science Foundation. Role: Principal Investigator. Amount: \$62,000. Period: September, 1985 - May, 1988.

Administrative Activities

National Center for Data Mining, 1998-present. Grossman founded the National Center for Data Mining at UIC in 1998. The Center has raised over \$20 million through external funding for projects. It has led the development of the Sector/Sphere large data cloud and the UDT high performance network protocol. It has also led the development of the Predictive Model Markup Language (PMML), the acknowledged standard in data mining and statistical modeling, operated a global data mining testbed called the Teraflow Testbed, and led the development of open source tools for data intensive computing, distributed computing, and high performance networking.

Magnify, 1994-2005. Grossman founded Magnify in 1994 and was its CEO from 1994 to 2001 and its Chairman until it was sold to ChoicePoint in 2005. Magnify develops data mining systems and provides outsourced analytics to the financial services and marketing industries. During this time, he raised three rounds of venture capital totaling over \$10 million, recruited a management team, signed key contracts with Visa, Allstate, Metropolitan Life, DOD, yesmail.com and coolsavings.com, and oversaw offices in Chicago, Maryland, and California. In addition, he founded Magnify Research, which provides data mining solutions to the federal government. Magnify Research was sold to Baesch Computer Consulting in 2002, and is now part of Unisys.

Software Projects

Sector/Sphere. Sector/Sphere is an open source cloud computing platform designed to manage and compute with large data that was first released in 2008. It was used to build the application that won the SC 08 Bandwidth Challenge Application. Unlike most other cloud computing platforms, Sector/Sphere is not only designed to operate within a data center but also across multiple geographically distributed data centers.

Augustus. Augustus is an open source infrastructure for building and deploying data mining and statistical models for large data sets and high volume data streams. The first release of Augustus was in 2005 on Source Forge. Augustus is compliant with the Predictive Model Markup Language (PMML). Augustus supports vectorized operations. Augustus also includes a Python based infrastructure for data integration and data preparation, which is usually one of the most time consuming components of building and deploying new statistical and data mining models.

UDT. During the period 1999-2003, Grossman led the development of a high performance network protocol called SABUL. SABUL used a UDP based data channel and a TCP based control channel. SABUL set a number of milestones for high performance data transport on wide area OC-3 and OC-12 networks. During the period, 2003 through the present, Grossman led the development of a successor to SABUL called UDT. UDT is entirely implemented in UDP and provides reliable, fair, and friendly data transport for high volume data flows. Both SABUL and UDT are open source and are available on source forge.

DataSpace. From 1998 to 2004, Grossman has led the development of open source data web clients and servers, including the data web Server Mercury designed for commodity networks, the data web Server Jupiter designed for high performance networks, and the data transport libraries Pockets and SABUL. See www.dataspaceweb.net

PMML. From 1998 to the present, Grossman has been the spokesperson and chairman of the Data Mining Group's working group on the Predictive Model Markup Language (PMML). PMML has now been adopted by most vendors of data mining and statistical software including IBM, Oracle, Microsoft, SAS, SPSS, and over ten others vendors. During this time, Grossman led the development of several reference implementations of PMML producers and consumers. See www.dmg.org.

PATTERN. During 1996-2000, Grossman led the development of Magnify's PATTERN data mining system. PATTERN was a scalable data mining system sold and marketed by Magnify and used by

a variety of financial institutions, as well as in-house by Magnify. PATTERN employed a layered architecture, consisting of a scalable column oriented data warehouse, a scalable data mining system, and an XML based infrastructure for quickly deploying predictive models. PATTERN was the first commercial data mining system to use ensemble based modeling techniques. It was also the first to use taxonomy-based modeling techniques. See www.magnify.com.

PTool. During 1992-1996, Grossman led the development of PTool, a scalable high performance persistent object manager designed for warehousing large data sets. Variants of PTool were later adopted by various scientific collaborations, including high energy physicists at Fermi Lab. PTool was used to create some of the earliest terabyte size data warehouses of scientific and engineering data. PTool was also used as an infrastructure for the data analysis and data mining of very large data sets.

Previous Projects

The Teraflow Project, 2004-2008. The goal of the Teraflow Project is to develop data mining middleware for transporting, exploring and mining high volume data flows. The Teraflow Project is supporting the development of several tools and applications, including UDT for high volume data transport, SOAP* for high performance web services, and applications in several domains including astronomy, bioinformatics, and sensor networks, built over UDT, SOAP*, and related tools. Part of the project includes the operation of a testbed called the Teraflow Testbed, which is an international application testbed for exploring, analyzing, integrating and detecting changes in massive and distributed data over wide area high performance networks. The Teraflow Testbed has nodes in Chicago, Kingston, Amsterdam, Geneva, Daejeon, and Tokyo that are connected by 1 Gbps and 10 Gbps wide area networks. The Teraflow Testbed is currently used to help distribute the Sloan Digital Sky Survey data to researchers world wide. It is also used for experiments for detecting changes in high volume data flows. For more information, please visit the Teraflow Testbed web site.

Project DataSpace, 1999-2004. DataSpace was middleware infrastructure allowing remote and distributed data to be explored, visualized, accessed, analyzed and mined with a point and click interface. The five year project began in 1999 by developing an internet protocol called the data space transfer protocol of DSTP specifically designed to access remote and distributed data broadly analogous to the way that http accesses remote documents. Later, the middleware was redesigned to use standard web services, as well as high performance web services specifically designed for large distributed data sets. By the end of the project in 2004, DataSpace applications had been developed to work with a variety of different data types, including business, e-business, scientific, engineering, and health care data.

The Terabyte Challenge, 1996-2004. The Terabyte Challenge was an international testbed for mining massive and distributed data using high performance networks. During its operation and the operation of its successor The Terra Wide Data Mining Testbed, from 1996-2004, it set several records milestones in data mining, high performance computing, and high performance networking. The Terra Wide Data Mining Testbed was one of the first application testbeds to demonstrate how applications could be built effectively over networks in which applications can set up, tear down, and monitor their optical paths (sometimes these are called lambda grids).

National Scalable Cluster Project (NSCP), 1994-1999. Robert Grossman was a co-founder and co-director of the National Scalable Cluster Project (NSCP). The project began in 1994 and ended in 1999. The NSCP was a collaboration between research groups at the University of Illinois at Chicago, the University of Pennsylvania, and the University of Maryland at College Park, together with partners from other universities, national laboratories, and companies. It connected high performance work station clusters at each of these locations using a high performance ATM network to create one of the first national grid based computing infrastructures. The support for this project was provided by two \$4M NSF Academic Research Infrastructure (ARI) consortia awards to the three core universities.

PASS Project, 2001-1995. Robert Grossman was a co-founder and co-director of the PASS Project. The goal of the PASS project was to develop an open, modular and scalable persistent object store for scientific computing applications requiring the ability to store, access and make numerically intensive queries on petabytes of data distributed across hierarchical storage systems. Twenty scientists and five institutions participated in this five year HPCC project which ended in 1995.

Professional Organizations

- I am a Member of the ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) Board of Directors for the term 2009-2011. I was also a Member of the Board for the terms 2005-2007 and 2007-2009.
- I am a member of the Association for Computing Machinery, the IEEE Computer Society, the American Statistical Association, the American Mathematical Society, the Society for Industrial and Applied Mathematics, the the American Association for the Advancement of Science, and Sigma Xi.
- I was the editor of the ACM SIGSAM Bulletin from 1991-1995.

Standards

- Since 2008, I have been the Chair of the Open Cloud Consortium, which develops standards and interoperability frameworks for cloud computing.
- Since 1998, I have been the Chair of the Data Mining Group, which is a vendor led organization that develops the Predictive Model Markup Language (PMML), an XML language for statistical and data mining models.

Professional Activities

- Member, Scientific Advisory Board, Chicago Biomedical Consortium, 2006-2008.
- Industrial/Government Applications Program Committee, Co-Chair, The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), Philadelphia, August 20-23, 2006.
- General Chair, The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2005), Chicago, August 21-24, 2005.
- Chair, KDD-2004 Workshop on Data Mining Standards, Services and Platforms, part of the The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, August 22, 2004.
- Chair, KDD-2003 Workshop on Data Mining Standards, Services and Platforms, part of the The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 27, 2003.
- Member, DOE Committee of Visitors (COV) subcommittee for Advanced Scientific Computing, DOE Office of Science, 2004.
- Reviewer, National Research Council's Report, "A Review of the FBI's Trilogy Information Technology Program," The National Academies Press, Washington, DC, 2004.

- Chair, CCR-P/DIMACS Workshop Mining Massive Data Sets and Streams: Mathematical Methods and Algorithms for Homeland Defense, June 17-22, 2002, Princeton, New Jersey.
- Co-Chair, Second SIAM International Conference on Data Mining 2002 (SDM-02), April 11-13 2002, Arlington.
- Co-Chair, First SIAM International Conference on Data Mining 2001 (SDM-01), April 5-6, 2001, Chicago.
- Co-Chair, Second Workshop on Mining Scientific Datasets, Army High Performance Computing Research Center University of Minnesota, Minneapolis, July 20-21, 2000.
- Sponsorship Chair, ACM Knowledge Discovery and Data Mining (KDD) Conference, 2001.
- Sponsorship Chair, ACM Knowledge Discovery and Data Mining (KDD) Conference, 2000.
- Chair, SIAM Mathematics in Industry Workshop, October 2-3, 1998, University of Illinois, Chicago.
- Chair, Mining and Managing Massive Data 1998 (M3D-98), San Diego Supercomputing Center, February 5-6, 1998.
- Chair, Mining and Managing Massive Data 1997 (M3D-97), July 12-15, 1997, Chicago.
- Chair, Mining and Managing Massive Data 1996 (M3D-96), San Diego Supercomputing Center, March 14-15, 1996.
- Co-Chair, Mathematics and Data Mining, Center for Communications Research, La Jolla, California, February 13-14, 1996.
- Chair, UIC-NASA Ames Workshop on Application-Specific Symbolic Techniques in High Performance Computing Environments, Fields Institute, Waterloo, Ontario, October 18-20, 1993.
- Co-Chair, Workshop on Hybrid Systems, Mathematical Sciences Institute, Ithaca, June 10 - 12, 1991.
- Co-Chair, The UIC-NASA Ames Symposium: New Directions in Computing: Analytic Computing, Intelligent Computing, and Database Computing, University of Illinois at Chicago, Chicago, April 7-9, 1991.
- Co-Chair, Workshop on Geometric and Algebraic Integration Algorithms, Mathematical Sciences Institute, Ithaca, November 4-8, 1989.
- Co-Chair, The NASA-Ames Symposium AI and Discrete Event Control Systems, NASA-Ames Research Center, Moffett Field, California, July 7-8, 1988.
- Co-Chair, The NASA-Ames Symposium on The Use of Symbolic Methods to Solve Algebraic and Geometric Problems Arising in Engineering, NASA-Ames Research Center, Moffett Field, California, January 15-16, 1987.

Selected Technical Talks

2011

- The Emergence of Genomics as a Data Intensive Science (invited talk), The 1st Data Intensive Science Workshop, Tokyo, March 9, 2011.

- A Quick Introduction to the Open Cloud Consortium's Open Science Data Cloud and Open Cloud Testbed, Japan Grid Consortium Workshop, March 10, 2011, Tokyo.
- A Quick Introduction to the Open Science Data Cloud, Data Intensive Research Technology Workshop, Baltimore, March 17, 2011.
- Bionimbus: A Cloud-Based Infrastructure for Managing, Analyzing and Sharing Genomics Data, EUIndiaGrid2 Biology Grid and Cloud Workshop, Cambridge, United Kingdom, March 29, 2011.
- An Overview of the Open Science Data Cloud, Cloud Computing and Its Applications 2011 (CCA-11), Chicago, April 13, 2011.
- The Open Science Data Cloud, part of the Science of Cloud Computing Plenary Panel, 2011 IEEE 4th International Conference on Cloud Computing (Cloud 2011), Washington, D.C., July 5, 2011 (plenary talk).

2010

- Building 10,000 Predictive Models: Scaling Health and Status Models to Large, Complex Systems, Predictive Analytics World, San Francisco, California, February 16, 2010.
- Bionimbus: An Open Community Cloud for Biological Data, NHGRI Cloud Computing Meeting, Bethesda, Maryland, March 31, 2010.
- Running R over Clouds, R/Finance 2010: Applied Finance with R, Chicago, Illinois, April 17, 2010.
- Project Matsu, Open Grid Forum (OGF) 29, Chicago, June 21, 2010.
- An Overview of the Open Science Data Cloud, Science Cloud 2010 (HPDC Workshop), Chicago, June 21, 2010.
- My Other Computer is a Data Center: The Sector Perspective on Big Data (Keynote Talk), 2nd Workshop on Large-scale Data Mining: Theory and Applications (LDMTA 2010), Washington, D.C, July 25, 2010.
- MalStone: Towards a Benchmark for Analytics on Large Data Clouds, 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010), Washington, DC, July 28, 2010.
- Distributed Data-Parallel Computing Using Sector/Sphere, Big Data for Science Workshop (Virtual Workshop that was broadcast to 12 universities), July 30, 2010.

2009

- Extending Analytics to Clouds, 8th Annual ON*VECTOR International Photonics Workshop, La Jolla, California, February 24, 2009.
- Sector: An Open Source Cloud for Data Intensive Computing, CloudSlam '09, April 20, 2009.
- Panel on Emerging Trends in Open Standards and Cloud Computing for Data Mining, Paris, June 30, 2009, KDD 2009.
- IEEE New Technologies Conference, My Other Computer is A Data Center, Philadelphia, August 6, 2009.
- Cloud-based Services For Large Scale Analysis of Sequence and Expression Data: Lessons From Cistrack, Critical Assessment of Massive Data Analysis (CAMDA 2009), Chicago, October 6, 2009.

- Sector: An Open Source Cloud for Data Intensive Computing, Cloud Computing It's Applications 2009 (CCA 09), Chicago, October 20, 2009.
- Lessons from a Year's Worth of Benchmarks of Large Data Clouds, MTAGS 2nd Workshop on Many-Task Computing on Grids and Supercomputers, SC 09, Portland, November 16, 2009.
- Cloud-based Services For Large Scale Analysis of Sequence and Expression Data: Lessons From Cistrack, SC 09 Workshop On Using Clouds for Parallel Computations in Systems Biology, November 16, 2009.

2008

- Knowledge Discovery from Distributed Data: Lessons from the Teraflow Testbed, National Science Foundation, April 8, 2008, Arlington, Virginia.
- Data Mining Using High Performance Clouds: Experimental Studies Using Sector and Sphere, 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), August 27, 2008, Las Vegas, Nevada.
- Robert L Grossman, The Design and Implementation of a Data Cloud That Spans Data Centers, UK e-Science All Hands Meeting 2008, September 10, 2008, Edinburgh, UK.
- Sector and Sphere: Towards Large Scale, Distributed Data Storage, Data Sharing, and Simplified Data Processing, Cloud Computing and Its Applications 2008, Chicago, October 23, 2008.
- Towards Global-Scale Cloud Computing: Using Sector and Sphere on the Open Cloud Testbed, SC 08, November 18, 2008.

2007

- An Introduction to Data Mining on Grids, Midwest Grid Workshop, Chicago, March 25, 2007.
- Hopf Algebras of Labeled Trees and Some Associated Differential Algebra Structures, Second International Workshop on Differential Algebra and Related Topics, Rutgers University, Newark, New Jersey, April 13, 2007.
- Modeling Highly Large, Heterogeneous Data Sets: Towards a Billion Models, DIMACS Workshop on Recent Advances in Mathematics and Information Sciences for Analysis and Understanding of Massive and Diverse Sources of Data, Rutgers University, New Brunswick, May 15, 2007.
- Unique Keys for Chemical Compounds and Metabolic Pathways, Interface 2007, Philadelphia, May 25, 2007.
- Building Statistical Models on Large and Distributed Data, Analytical Computing Forum, June 28, 2007, Austin, Texas.
- Data Driven Discovery in E-Science, Interdisciplinary Strategic Issues in e-Science and Cyber-Infrastructure, Caltech, June 13, 2007.
- Detecting Changes in Large Data Sets of Payment Card Data: A Case Study, Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, August 14, 2007.
- Distributed Discovery in E-Science: Lessons from the Angle Project, National Science Foundation Workshop on Next Generation Data Mining (NGDM '07), Baltimore, Maryland, October 12, 2007.

- Sector: A Peer-to-Peer Infrastructure for Distributing Large Scientific Data Sets Over Wide Area High-Performance Networks, GridNets 2007, Lyon, France, October 17, 2007
- Angle: Detecting Anomalies and Emergent Behavior from Distributed Data in Near Real Time, SC 07, Reno, NV, November 13, 2007.
- Data Grids, Data Clouds and Data Webs: A Survey of High Performance and Distributed Data Mining , Workshop on Hardware and Software for Large-Scale Biological Computing in the Next Decade. Okinawa, Japan, December 12, 2007.

2006

- Other People's Petabytes: The Challenge of Distributed Data Mining and Distributed Data Integration, Salishan High Speed Computing Conference, April 26, 2006, Salishan, Oregon.
- Local Terabytes, Remote Terabytes, and Distributed Terabytes: Four Case Studies in Data Mining, Seminar, Computer Science Department, University of Chicago, May 3, 2006.
- Multiscale Analysis Of Data: Clusters, Outliers and Noise - Preliminary Results, Second NASA Data Mining Workshop: Issues and Applications in Earth Science, Pasadena, May 24, 2006.
- Change Detection using Cubes of Models (CDCM), Interface 2006, 38th Symposium on the interface of statistics, computing science, and applications, Pasadena, California, May 26, 2006.
- The Age of Data-Driven Discovery and Decision Support: The New Rules, The First Vyborny Memorial Lecture, University of Chicago, July 17, 2006.
- Using Term Lists and Inverted Files to Improve Search Speed for Metabolic Pathway Databases, 3rd International Workshop on Data Integration in the Life Sciences 2006 (DILS'06), July 21, 2006.
- Sector - An eScience Platform for Distributing Large Scientific Data Sets, eScience Workshop, October 13, 2006, Baltimore, Maryland.
- DataSpace - Data Integration Using Universal Keys, Workshop on Information Integration, Philadelphia, October 26, 2006.
- Transporting the Sloan Digital Sky Surey Using Sector, SC 06, November 14, 2006.
- Distributing the Sloan Digital Sky Survey Using UDT and Sector, Second IEEE International Conference on e-Science and Grid Computing, Amsterdam, December 4, 2006.

2005

- Yunhong Gu and Robert L. Grossman, Optimizing UDP-based Protocol Implementations, Third International Workshop on Protocols for Fast Long-Distance Networks Lyon, France, February 4, 2005 (presentation by Michal Sabala).
- The UDT Project and the Teraflow Testbed, 4th Annual ON*VECTOR International Photonics Workshop, La Jolla, March 1, 2005.
- Bioweb and Biogrids of Proteomics Data, Panel Presentation, Workshop on Proteomics and Informatics sponsored by the Chicago Biomedical Consortium, Northwestern University, April 22, 2005.
- An Event Based Framework for Improving Information Quality That Integrates Baseline Models, Causal Models and Formal Reference Models, Second International ACM SIGMOD Workshop on Information Quality in Information Systems, Baltimore, June 17, 2005.

- Assigning Unique Keys to Chemical Compounds for data integration: some interesting counterexamples, 2nd International Workshop on Data Integration in the Life Sciences, University of California, San Diego, July 22, 2005.
- High Performance Analytics: Why do Network Protocols and Light Paths Matter?, iGrid 2005, San Diego, California, September 26, 2005.
- A Tutorial Introduction to High Performance Analytics, SC 05, Seattle, November 14, 2005.
- Real Time Change Detection and Alerts from Highway Traffic Data, SC 05, Seattle, November 15, 2005.
- The Teraflow Challenge: High Performance Mining of Streaming Data, SC 05, Seattle, November 15, 2005.
- Master Works Talk: Data Mining Challenges: Technical, Pragmatic and Strategic, SC 05, Seattle, November 16, 2005.

2004

- UDT: An Application Level Transport Protocol for Grid Computing, Second International Workshop on Protocols for Fast Long-Distance Networks, PFLDnet 2004, Argonne National Laboratory, Argonne, Illinois, February 17, 2004.
- Tera Mining: A Testbed for Distributed Data Mining over High Performance SONET and Lambda Networks, NSF Shared Cyberinfrastructure (SCI) Meeting, Arlington, Virginia, February 19, 2004.
- Bioweb, UT-ORNL Bioinformatics Summit 2004, Fall Creek Falls State Park, Pikesville, TN, March 27, 2004.
- Using DataSpace Archives to Support Long Term Stewardship of Remote and Distributed Data, NASA/IEEE Conference on Mass Storage Systems and Technologies (MSST2004), College Park, Maryland, USA, April 14, 2004.
- Open DMIX: High Performance Web Services for Distributed Data Mining, 2004 SIAM International Conference on Data Mining (SDM 2004) Workshop on High Performance and Distributed Data Mining, Orlando, April 24, 2004.
- Distributed Alert Management Systems, Rutgers - CIMIC Workshop on Securing Critical Infrastructure and Resources Protection, Rutgers University, Newark, NJ, June 24, 2004.
- Some Hopf Algebras of Trees and their Applications, BIRS Workshop on Combinatorial Hopf Algebras, Banf, British Columbia, August 28 - September 2, 2004.
- Highly Scalable, UDT-Based Network Transport Protocols for Lambda and 10 GE Routed Network, DOE Office of Science High-Performance Network Research Workshop Ultranet 2004, Fermi National Laboratory, Batavia, Illinois, September 15, 2004.
- Unique chemical keys for biomolecules and integration of distributed data, Symposium on Computational Science of Biomolecules: Applications in Medicine and Therapeutics, University of Illinois at Chicago, October 8, 2004.
- Experiences in the Design and Implementation of a High Performance Transport Protocol, SC 04, Pittsburgh, November 9, 2004. (Presentation partly by Yunhong Gu.)

2003

- Bioweb, Data mining for the Americas, NSF AMPATH Workshop: Fostering Collaborations and Next Generation Infrastructure, Florida International University, Miami, January 29, 2003.
- The OptIPuter Data Stack, OptIPuter Project Meeting, San Diego, February 6, 2003.
- Data Grids and Beyond, Global Grid Forum/Internet Society Master Class, University of Amsterdam, March 25, 2003.
- Global Access to Large Distributed Data Sets using Photonic Data Services, 20th IEEE Symposium on Mass Storage Systems, San Diego, April 8, 2003.
- High Performance Data Transport Protocols employing UDP-based Data Channels and TCP-based Control Channels, DOE Workshop on Ultra High-Speed Transport Protocols and Network Provisioning for Large-Science Applications, Argonne National Laboratory, April 10, 2003.
- Some Case Studies For Alert Management Systems, DARPA Workshop, BBN, Cambridge, May 27, 2003.
- Beyond Data Grids: Photonic Data Services on Lambda Grids, Global Grid Forum Plenary Panel Presentation, Seattle, June 25, 2003.
- High Performance File Transfer Protocols and Congestion Control Mechanisms Using SABUL, Internet2 Techs Workshop, Lawrence, Kansas, August 5, 2003.
- Experimental Studies of the Universal Chemical Key (UCK) Algorithm on the NCI Database of Chemical Compounds Abstract, The Computational Systems Bioinformatics Conference (CSB), Stanford, August 12, 2003.
- Virtual Joins Using Universal Keys: Towards Data Integration Services in Data Mining Middleware, Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing, Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN, September 18, 2003.
- The SABUL Application Library for High Performance Data Transport: How to Move Very Large Data Sets Over Very Long Distances - Using Today's Network Infrastructure, Cern Computing Seminar, Geneva, Switzerland, October 1, 2003.
- Open DMIX - Data Integration and Exploration Services for Data Grids, Data Web and Knowledge Grid Applications, First International Workshop on Knowledge Grid and Grid Intelligence (KGGI 2003), Halifax, Canada, October 13, 2003.
- Beyond Data Grids: Data Webs, Lambda Grids, and All That, NASA Information Science and Technology Colloquium Series, NASA Goddard, November 12, 2003.
- A Tutorial Introduction to High Performance Data Transport, Bill Allcock (Argonne National Laboratory), Robert Grossman (University of Illinois at Chicago and Open Data Partners), and Steven Wallace (Indiana University), SC 03, Phoenix, November 16, 2003.
- Project DataSpace Bandwidth Challenge Presentation, SC 03, Phoenix, November 18, 2003.
- HPC Challenge Presentations, Using Virtual Joins in DataSpace to Mine and Visualize Distributed Data, SC 03, Phoenix, November 19, 2003.

2002

- Analyzing Remote Data and Mining Distributed Data Using Data Webs, Department of Computing Seminar, Imperial College, London, March 22, 2002.

- Analyzing Remote Data and Mining Distributed Data Using Data Webs Departmental Colloquia, Computer Science Department, Indiana University, April 3, 2002.
- Analyzing Remote Data and Mining Distributed Data Using Data Webs, Indiana Pervasive Computing Research Initiative Colloquium, Indiana University Purdue University at Indiana (IUPUI), April 4, 2002.
- Combining Families of Information Retrieval Algorithms Using Meta-Learning, Second SIAM International Workshop on Text Mining, Arlington, VA, April 13, 2002.
- Finding Bad Guys in Distributed Streaming Data Sets, Panel Presentation on Resource and Location Aware Data Mining, Second SIAM International Workshop on High Performance Data Mining, Arlington, VA, April 13, 2002.
- Why Neural Networks Won't Catch Bad Guys, The 2002 Workshop on Information and Data Management IDM 2002 Arlington, Virginia, May 6, 2002.
- An Introduction to High Performance Data Mining for Homeland Defense, CCR-P/DIMACS Conference on Mining Massive Data Sets and Streams: Mathematical Methods and Algorithms for Homeland Defense, Princeton, New Jersey, June 17, 2002.
- The Freeing of Biological Data: From Biological Databases to Biogrids and Biowebs, Chicago Community Trust Symposium, Chicago, September 3, 2002.
- DataSpace (demonstration), IGrid 2002, Amsterdam, The Netherlands, September 25, 2002.
- Photonic Data Services: Integrating Path, Network, and Data Services, to Support Next Generation Data Mining Applications, NSF Workshop on Next Generation Data Mining Applications, Baltimore, Maryland, November 2, 2002.
- High Performance Data Webs (demonstration), SC 02, Baltimore, Maryland, November 18, 2002.
- High Performance Data Webs on the Terra Wide Data Mining Testbed (via video), CANARIE Advanced Networks Workshop, Montreal, Quebec, November 20, 2002.
- High Performance Computing Challenge: Data Exploration on the Terra Wide Data Mining Testbed, SC 02, Baltimore, Maryland, November 20, 2002.
- Merging Multiple Data Streams on Common Keys over High Performance Networks, SC 02, Baltimore, Maryland, November 21, 2002.
- Data Mining and Cyber Threat Analysis: Three Trends, Workshop on Data Mining for Cyber Threat Analysis, IEEE International Conference on Data Mining, December 9, 2002 Maebashi City, Japan.
- An Algebraic Approach to Data Mining: Some Examples, IEEE International Conference on Data Mining, December 10, 2002 Maebashi City, Japan.

2001

- Project DataSpace: An Infrastructure Supporting Real Time Analysis and Decision Making with Complex, Distributed Data, Multi-Sector Crisis Management Consortium (MSCMC), Alliance Center for Collaboration Education, Science and Software (ACCESS), Arlington, Virginia, March 14, 2001.
- Can Data Mining Ever be a Gigabit Application?, Lessons from DataSpace, Salishan Conference on High Speed Computing, Glen Eden, Oregon, April 25, 2001.

- Mining Distributed Exabytes of Data, Alliance All-Hands Meeting, National Center for Supercomputing Applications, Champaign-Urbana, Illinois, May 24, 2001.
- Tera-Mining, Star Tap Meeting, INET 2001, Stockholm, Sweden, June 5, 2001
- Steps Toward Real Time Data Mining, ICSA 2001, Applied Statistics Symposium, June 7-9, 2001, Chicago.
- An Introduction to PMML, Second Annual Workshop on the Predictive Model Markup Language, August 26, 2001, San Francisco, California.
- The Data Challenge, Testimony before the NSF Blue Ribbon Advisory Committee for Cyberinfrastructure, November 29, 2001, Arlington, Virginia.

2000

- Terabyte Challenge 2000: Project DataSpace, Asian Pacific Advanced Network Workshop, Tsukuba Science City, Japan, February 15, 2000.
- The Terabyte Challenge 2000/Project DataSpace, Workshop on Scientific Data Management, Minnesota High Performance Computing Center, July 20, 2000.
- The Terabyte Challenge 2000/Project DataSpace, NASA Ames, August 14, 2000.
- Distributed and Parallel Data Mining: Advances and Future Directions, Distributed and High Performance Knowledge Discovery 2000 (DPKD-2000), ACM Knowledge Discovery in Databases (KDD) 2000 Conference, Boston, August 20, 2000.
- A Framework for Distributed Data Mining Strategies that are Intermediate Between Centralized Strategies and In-Place Strategies, Distributed and High Performance Knowledge Discovery 2000 (DPKD-2000), ACM Knowledge Discovery in Databases (KDD) 2000 Conference, Boston, August 20, 2000.
- Introduction to PMML, PMML Workshop, ACM Knowledge Discovery in Databases (KDD) 2000 Conference, Boston, August 23, 2000.
- A Tutorial on High Performance Data Mining, Supercomputing 2000 (SC2000), Dallas, November 5, 2000.
- Pockets: The Case for Application-level Network Striping for Data Intensive Applications using High Speed Wide Area Networks, Supercomputing 2000 (SC2000), Dallas, November 8, 2000.

1999

- The Inevitable Emergence of Data Mining, Chicago Chapter of the American Statistical Association, East Bank Club, January 12, 1999.
- Terabyte Challenge 2000, Abilene Launch Event, Washington, D.C., February 24, 1999.
- The Terabyte Challenge: A Testbed for High Performance and Distributed Data Mining, Post-vBNS NSF-CRA Invitational Workshop, La Jolla, March 1, 1999.
- The Predictive Model Markup Language (PMML), R. L. Grossman and M. Cornelson (presentation by M. Cornelson) 1999 AFCEA Federal Data Mining Symposium and Exposition, Tysons Corner, Virginia, March 9-10, 1999.
- Mining Collection of Trajectories, SIAM 1999 International Conference on Parallel Processing (ICPP), San Antonio, March 23, 1999.

- Data Mining: Issues and Challenges in Wide Area Distributed Data Mining, KDD 99 Workshop on High Performance Data Mining, San Diego, August 15, 1999.
- A High Performance Implementation of the Data Space Transfer Protocol (DSTP), KDD 1999 Workshop on High Performance Data Mining, San Diego, August 15, 1999.
- Terabyte Challenge 2000: Project DataSpace, AHPRC Workshop, Minneapolis, September 9, 1999.
- Terabyte Challenge 2000: Project DataSpace, National Science Foundation, Arlington, VA, November 9, 1999.
- A Tutorial on High Performance Data Mining, Supercomputing 1999, Portland, November 15, 1999.
- Papyrus: A System for Data Mining over Local and Wide Area Clusters and Super-Clusters, Supercomputing 1999, Portland, November 16, 1999.
- Terabyte Challenge 2000: Project DataSpace, SuperComputing 99 Conference - High Performance Computing Challenge, Portland, November 17, 1999.

1998

- Combing Data Mining and Predictive Modeling, Advanced Information Processing and Analysis Steering Group (AIPA 98) Conference , Tysons Corner, Virginia, March 17-18, 1998.
- A Tutorial Introduction to High Performance Data Mining, Sixth NASA Goddard Space Flight Center Conference on Mass Storage and Technologies and Fifteenth IEEE Symposium on Mass Storage Systems, College Park, Maryland, March 23, 1998.
- Scaling Tree-based Classifiers, Second Pacific-Asia Conference on Knowledge Discovery and Data Mining, Melbourne, April 12-19, 1998.
- Scaling Tree-based Classifiers, DIMACS Workshop on High Performance Data Mining, Princeton, April 26-28, 1998.
- The Inevitable Emergence of Data Mining, UCAID/Internet 2 Quality of Service Workshop, Santa Clara, May 20-22, 1998.
- The Inevitable Emergence of Data Mining, Army Research Center (ARL) and US Army Test and Evaluation Command (TECOM) Workshop on Computing, Aberdeen, Maryland, August 19-21, 1998.
- A Tutorial Introduction to High Performance Data Mining, AAAI 1998 Conference on Knowledge Discovery and Data Mining (KDD-98), New York City, August 27-31, 1998.
- Data Mining on Clusters, Super-Clusters, and Meta-Clusters, 1998 Asian Conference on High Performance Computing, Singapore, September 22-27, 1998.
- Data Mining on Clusters, Super-Clusters, and Meta-Clusters, RCI Conference on High Performance Computing, Pentagon City, October 14, 1998.
- The Terabyte Challenge: A Testbed for High Performance and Distributed Data Mining, Research Demonstration, 1998 Supercomputing Conference (SC-98), Orlando, November 8 - November 12, 1998.
- The Terabyte Challenge: A Testbed for High Performance and Distributed Data Mining, Research Demonstration, IBM CASCON Conference, Toronto, November 30 - December 3, 1998.

- The Terabyte Challenge: A Global Testbed for High Performance and Distributed Data Mining, 2nd annual CA*net Workshop, Ottawa, December 15-16, 1998.

1997

- High Performance Data Mining, Tandem Computer, Austin, Texas, January 31, 1997.
- Four one hour talks on data mining and related topics: 1) An Overview to Data Mining, Data Warehousing and Intelligent Agents, 2) An Introduction to Data Mining, 3) An Introduction to High Performance Data Warehouses, 4) Integrated Architectures for Data Mining, Department of Defense, Fort Meade, Maryland, March 6, 1997.
- Detecting Network Intrusions Using Data Mining, Sixth Annual Symposium on Advanced Information Processing and Analysis, Tysons Corner, Virginia, March 26, 1997.
- The Old Order Changeth Yielding Place to the New: The Rise of High Performance Data Management and the Demise of High Performance Computing, Pittsburgh Supercomputing Center, March 29, 1997.
- The Data Mining and Analysis of Packet Data for Detecting Network Intrusion, Eleventh International Conference on Mathematical Modeling and Scientific Computing, Washington, DC, April 1, 1997.
- Data Mining in Financial Services, Financial Services Technology Consortium (FSTC) General Meeting, Orlando, FL, April 17, 1997.
- Using Data Mining to Detect Network Intrusion, Practical Applications of Data Mining and Knowledge Discovery PADD 97, London, England, April 25, 1997.
- An Introduction to Data Mining, Center for Communication Research, Princeton, NJ, May 1, 1997.
- High Performance Data Mining: A Tutorial Introduction, First European Symposium, Principles of Data Mining and Knowledge Discovery, PKDD 97, Trondheim, Norway, June 25-27, 1997.
- Dynamic Similarity: Mining Collections of Trajectory Segments, NSF Workshop on the Mathematics of Mining Massive Data, Chicago, Illinois July 12-15, 1997.
- JTool: Accessing Warehoused Collections of Objects with Java, Persistent Java Workshop 2, Half Moon Bay, California, August 13-15, 1997.
- An Introduction to Data Mining, NASA-CESDIS Workshop on Data Mining and Data Warehousing, Greenbelt, Maryland, August 19-21, 1997.
- Detecting Network Intrusions Using Data Mining, Seminar at Boeing Research, September 2, 1997.
- Data Mining Scientific and Engineering Data, NSF Mathematical and Physical Sciences Distinguished Lecturer, Arlington, Virginia, September 30, 1997.
- High Performance Data Mining, CASCON 97, Toronto, Ontario, November 11, 1997.
- High Performance Data Mining: A Tutorial Introduction, Supercomputing 97, San Jose, California, November 16, 1997.
- The Terabyte Challenge: An Open, Distributed Testbed for Managing and Mining Massive Data Sets, Supercomputing 96, Pittsburgh, November 19, 1997.

- The Terabyte Challenge: An Open, Distributed Testbed for Managing and Mining Massive Data Sets, Supercomputing 96, Pittsburgh, November 19, 1997.

1996

- The Symbolic Computation of Differential Invariants Using Trees, Closing session of the special year on Computational Differential Algebra and Algebraic Geometry, City College of New York, January 5, 1996.
- An Informal Introduction to Mathematics of Data Mining, La Jolla, February 14, 1996.
- The Old Order Changeth: The Rise of High Performance Data Management in Scientific Computing, colloquium at the Pittsburgh Supercomputing Center, March 29, 1996.
- Computing Differential Invariants with Trees, Special Session on Differential Algebra at the AMS Regional Meeting, New York City, New York, April 13, 1996.
- Data Mining, Object Warehouses, and Persistent Object Managers, Seventh International Conference on Persistent Object Systems, Cape May, New Jersey, May 30, 1996.
- Optimization Driven Data Mining and Object Warehouses, SIGMOD 96 Workshop on Data Mining, Monreal Quebec, June 2, 1996.
- Data Mining Challenges for Digital Libraries and Electronic Commerce, MIT-ACM 50th Anniversary of the ACM, Boston, June 14, 1996.
- Accessing Warehoused Collections of Objects Through Java, First International Conference on Persistent Java, Glasgow, Scotland, September 17, 1996.
- Discovering Critical Patterns in Large Data Sets, NSF Workshop on Data Mining, Arlington, Virginia, September 25, 1996.
- Mode Shifting, Mode Sharing, and Mode Superposition, Hybrid Systems IV (HSAC 96), Ithaca, New York, October 14, 1996.
- Data Mapping Support for Data Mining Applications, NASA-CESDIS Workshop on Data Mapping, CESDIS, Greenbelt, Maryland, November 7, 1996.
- Managing, Mining, Querying and Analyzing Very Large Data Sets, CASCON 96, Toronto, November 13, 1996.
- High Performance Data Mining, Australian National University, December 17, 1996.