

Curriculum Vitae
CHAITANYA K. BARU
 San Diego Supercomputer Center
 University of California San Diego
baru@sdsc.edu

Education

- 1983 – 85 Doctor of Philosophy (*Ph.D.*) in Electrical Engineering, University of Florida, Gainesville, FL, 1985.
- 1981 – 83 Master's in Engineering (*M.E.*) in Electrical Engineering, University of Florida, Gainesville, FL, 1983
- 1974 – 79 Bachelor of Technology (*B.Tech.*) in Electrical Engineering, Indian Institute of Technology, Madras, India, 1979.

Awards

- Outstanding Master's Thesis, Dept. of Electrical Engineering, University of Florida, 1983
- *Recognition Award*, Center for Advanced Studies, IBM Toronto Labs for efforts in bridging research and development activities at IBM, 1996.

Employment

San Diego Supercomputer Center, UC San Diego

9500 Gilman Drive, La Jolla, CA 92093-0505; Dates: June 1996-current; Salary: \$185,000/year.

2013 – *Associate Director, Data Initiatives*

Responsibilities: Coordinate and expand SDSC's data-centric initiatives, while leveraging capabilities in high-performance computing and data analytics. Develop a research and education strategy for data science, targeted towards academic researchers as well as industry.

2011 – *Director, Center for Large-scale Data Systems Research (CLDS), San Diego Supercomputer Center*

Responsibilities: Create a industry-university collaboration in the area of "big data".

Set the technical agenda for the Center; recruit researchers; invite researcher/staff participation from other parts of SDSC and UCSD; plan and coordinate activities of the Center.

Since its initiation in October 2011, CLDS has garnered support from six industry sponsors and also established an open, community-based activity in Big Data Benchmarking, now called the *BigData Top100 List* (bigdatatop100.org). Conducted three workshops on Big Data Benchmarking in May 2012 in San Jose, CA, December 2012 in Pune, India, and July 2013 in Xi'an, China. The fourth workshop will be in October 2013 in San Jose. Workshops have been supported by grants from the NSF (CISE and ACI) as well as by industry sponsorships. Baru also co-chairs the NIST Working Group on Big Data.

2008 – *Distinguished Scientist, SDSC*

Responsibilities: Organize and run an internal grants competition (Collaborative Research Opportunities, CRO) to foster collaboration between SDSC researchers and other UCSD faculty with the objective of generating full-scale proposals to external funding agencies or creating useful computational and/or data services at SDSC. Created a *SDSC Project Advisory Council* consisting of UCSD faculty members representing a broad cross-section of disciplines across campus, to serve as the review committee for the CRO competitions.

2008 – *Director, Advanced Cyberinfrastructure Development (ACID) Lab*

Responsibilities: The ACID Lab is my research and development group at SDSC. As Lab Director, I am responsible for ensuring that projects that the lab is associated with are making appropriate and good progress, and continuity in funding and projects for the 15 members of the lab (5 PhD, 10 MS in Computer Science or related disciplines). About nine staff members are currently supported on projects for which I am the principal investigator (PI). The rest are supported on projects for which either they are the PIs themselves, or the PI is at UCSD or some other institution in the United States.

The lab was formally established in 2008. Over the years, our lab has participated in numerous cyberinfrastructure projects including the Geosciences Network (GEON), National Ecological Observatory Network Cyberinfrastructure Testbed (NEON CI), Network for Earthquake Engineering Simulations Cyberinfrastructure Center (NEESit), OpenTopography, Tropical Ecology Assessment and Monitoring Network (TEAM), the Moores Cancer Center Bioinformatics and Biostatistics Share Resource, the UCSD/Salk Center for Academic Training and Research in Anthropogeny (CARTA), CyberGIS, and others.

2004 – 2008 *Division Director, Science R&D Division and Member of the SDSC Senior Management Team.*

Responsibilities: Manage a division of 100+ staff, including senior as well as junior PIs. Manage a division budget of about \$2M/year.

Created the *SDSC/Calit2 Synthesis Center*, in collaboration with Prof. Ramesh Rao, Director of the UCSD Division of Calit2 and Prof. Larry Smarr, Director of Calit2. A large meeting space was allocated in the newly constructed Calit2 Building (Atkinson Hall), which was outfitted with a large format screen with dual high-resolution 4K projectors, and a 3D screen with a 4K projector.

2001 – 2004 *Co-Director, Data and Knowledge Systems Program. Member of SDSC Senior Management Team.*

Responsibilities: I was Co-Director responsible for the R&D activities while my colleague Dr. Phil Andrews was Co-Director for production storage services. My responsibility was to provide a common strategy across all data R&D groups and to develop a full portfolio of data programs. Created data groups for Data Mining, Advanced Database Projects, and Geoinformatics. A total of 9 labs provided a full portfolio of data activities.

San Diego Supercomputer Center, General Atomics, La Jolla, CA

1996–2001 *Senior Principal Scientist, Data Intensive Computing Environments (DICE) group. Technical Project Manager for the Distributed Object Computation Testbed (DOCT) project.*

Responsibilities: SDSC was still a division of General Atomics. As the senior technical person in the group, my responsibility was to provide technical direction, assist with technical design, and oversee technical activities in the group, and do the same for the DARPA-funded Distributed Object Computation Testbed (DOCT) project in the group, which was one of the largest projects at SDSC at that time.

Developed the Storage Resource Broker (SRB) concept for managing storage in a distributed, heterogeneous environment (with Arcot Rajasekar and Michael Wan). Assisted General Atomics (GA) to spin out the software into a separate division of GA called, NirvanaStorage.

IBM Corporation

1992 – 95 *Advisory Programmer, Database Technology Institute, IBM Almaden Research Labs, San Jose, CA (1995).
Advisory Development Analyst and Group Lead, Database Technology Group, IBM Toronto Labs (1992-95).*

Responsibilities: One of the three group leads for development of a parallel version of IBM's DB2 database system for UNIX-based, shared-nothing clusters.

Participated in the design and development of the TPC-D decision support benchmark standard. Led the team at IBM that published the first results for TPC-D on a 100GB database on 100 nodes.

University of Michigan, Ann Arbor, MI

1985 – 92 *Assistant Professor*, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122.

Responsibilities: Conduct research and teach courses at the undergraduate and graduate levels in Computer Science.

Received NSF grants for my research. Member of the Software Systems Research Lab (SSRL) and the Advanced Computer Architecture Lab (ACAL). Advised three PhD students.

Tata Engineering and Locomotive Company, Ltd., Pune, India

1979 – 81 *Programmer Analyst*, Management Services Division.

Programming costing and inventory control applications in Autocoder on IBM 1401 and COBOL on Burroughs systems.

Cyberinfrastructure Leadership Activities

1. *XLDB2013, Member, Organizing Committee.*
2. *Co-Chair, National Institute of Standards and Technologies (NIST) Big Data Working Group, 2013-.*
This activity is just beginning.
3. *Chair, Organizing Committee, Workshops on Big Data Benchmarking (WBDB), 2012-.*
Have conducted 2 WBDB's, first in San Jose in May 2012 and second in Pune, India in December 2012, with 60 and 40 attendees each, respectively. Third WBDB will be held in July 2013 in Xi'an, China, and fourth will be held in October 2013 in San Jose. Currently, we are considering collocating the 5th WBDB with the Standard Performance Evaluation Corporation (SPEC) conference in March 2014, in Dublin, Ireland. Selected papers from the 2012 and 2013 workshops will be published in Springer-Verlag Lecture Notes in Computer Science.
4. *Chair, Board of Directors, BigData Top100 List (www.bigdatatop100.org), 2012-.*
Helping to drive forward the creation of industry standards.
5. *PI/Project Director of NSF-funded EarthCube Summer Institute, 2013.*
This will be the first ever summer institute for EarthCube, and builds upon 8 years of our experience in running the Cyberinfrastructure Summer Institute for Geoscientists (CSIG).
6. *Co-Chair, NCI/NSF Invitational Workshop Health Cyberinfrastructure: A Seeded Cloud Approach, January 10-11, 2011, San Diego Supercomputer Center, UCSD.*
The workshop advocated creation of a "seeded cloud" infrastructure where representative datasets are already loaded into the system to facilitate rapid investigations in interdisciplinary and transdisciplinary topics.
7. *PI/Project Director of NSF-funded Cyberinfrastructure Summer Institute for Geoscientists, 2004-2011.*
Conducted a week-long seminar for 8 years, with about 40 students/attendees per year.
8. *Coordinator, SDSC Data Science Institute.*
I am working with a team of about 6 researchers/Pis at SDSC to develop curricula for short format courses in data science to be offered to industry attendees and also scientists.
9. *PI, OpenTopography: A Cyberinfrastructure-enabled Facility for High-Resolution Topography Data and Tools, 2009-2012 and renewed 2012-2015.*
OpenTopography is funded as an Earth Science Facility. The project is a spinoff from the GEON LiDAR Workflow activity in the GEON project. (total funding \$3.2M).
10. *PI/Lead, Cyberinfrastructure for the Tropical Ecology, Assessment and Monitoring Network (TEAM), 2007-present.*

TEAM collects field ecological data from 16 sites in tropical forests across Central America, South America, Africa, and Asia. TEAM is a partnership funded by the Moore Foundation, led by Conservation International and includes the Missouri Botanical Garden, the Smithsonian Institution, and the Wildlife Conservation Society. SDSC provides all of the necessary cyberinfrastructure for this globally distributed data collection and data curation effort.

11. *Chair, SDSC Data Services Design Team, 2010-2011.*

This team developed a data services stack for all of SDSC—from production-level storage services to advanced data modeling and data analytics services. It recommended the creation of a stable storage service for users. A 5.5PB SDSC Cloud has been built, deployed, and is in use.

12. *Project Director, The Geosciences Network, GEON, 2002-2012.*

PI of the original large NSF ITR grant (\$11.5M total funding) involving 12 PI institutions. The project was renewed as GEON 2.0 (\$1.26M) and also spun off OpenTopography. My task was to coordinate a research agenda across the 12 PI institutions.

13. *Director of Cyberinfrastructure for NSF National Earthquake Engineering Simulations (NEESit), 2007-2010.*

I was first affiliated with NEESit in 2006, as a Cyberinfrastructure Advisor. Agreed to oversee the project during its transition period.

14. *Member of Senior Management Team, NSF National Ecological Observatory Network (NEON), 2005-2007.*

I was member of the original team that won the grant to conceive of and plan NEON. As member of the Senior Management Team, I was the Cyberinfrastructure Lead and worked in collaboration with Prof. Deborah Estrin, who was the lead for the Embedded Cyberinfrastructure, to develop the high-level concepts for NEON cyberinfrastructure. I was also co-PI (with PI David Schimel) of the NEON Cyberinfrastructure Diagnostic Testbed.

15. *Lead, KatrinaSafe Project.*

During Hurricane Katrina in 2005, I worked with the American Red Cross to develop KatrinaSafe, a “live” database of victims of Hurricane Katrina, with the objective of connecting hurricane victims with family and friends. Subsequently, the American Red Cross deployed this as a permanent service called disastersafe.redcross.org, hosted at SDSC. This is now called <https://safeandwell.communityos.org> and his hosted outside of SDSC and Red Cross. We processed information on about 200,000 individuals during the 2-month period of Hurricanes Katrina and Rita.

16. *Member, IRIS Data Management System Standing Committee, 2007-2009.*

Provide strategic input and advice on cyberinfrastructure-related issues to the IRIS data management team.

17. *Member, Cyberinfrastructure Advisory Committee, Long-Term Ecological Research Network (LTER, www.lternet.org), 2006.*

Served on a site visit committee that reviewed the information needs and requirements of the \ LTER network at large

18. *Member, Advisory Board, CLEANER Project Office, 2005-2006.*

Provided input to the project based on my experiences at SDSC and with GEON.

19. *Co-Convener, NSF Earth Science CyberInfrastructure (ES-CI) Task Force (with Lee Allison and Tom Jordan), 2004.*

20. *SDSC PI, CUAHSI Hydrologic Information System (<http://his.cuahsi.org>), 2004-2008.*

21. *Member and co-Investigator, Leadership Team, Biomedical Informatics Research Network-Coordinating Center (BIRN-CC), 2001-2004.*

Other Significant Collaborative Activities

1. **DELPHI**: – Data E-platform Leveraged for Patient Empowerment and Population Health Improvement. NSF-funded project on Smart Health and Well-being, led by Prof. Kevin Patrick (PI). I am a project co-investigator.

2. CARTA: Cyberinfrastructure and bioinformatics lead for the UCSD / Salk Institute-led Center for Academic Research and Training in Anthropogeny led by Profs. Ajit Varki, Margaret Schoeninger, and Rusty Gage (Salk). Funded by the Mathers Foundation. Duration: 2007—ongoing.
3. CYCORE: Co-PI of Cyberinfrastructure for Comparative Effectiveness Research project funded by NIH. Project is led by Dr. Kevin Patrick (SOM & Calit2) in collaboration with M.D. Anderson Cancer Center, Houston. Duration: October 2009—September 2010.
4. CISA3: Co-PI with Profs. Tom Levy and Falko Kuester of the *Mediterranean Archaeology Network (MedArchNet)*. Funded by the UCSD Chancellor's Collaboratory initiative, for 2009-2010 academic year.
5. 911: PI of NSF-funded project on Spatiotemporal Analysis of 911 Call Stream Data, with Prof. William Hodgkiss, SIO, as co-PI. Duration: 2004-2008.
6. WIISARD: Co-investigator (with Prof. Leslie Lenert as PI), Wireless Internet Information System for Medical Response in Disasters project. Funded by NIH, 2005-2007. Was responsible for the data management component.
7. I2T: PI of the NSF-funded Information Integration Testbed (I2T) project with Prof. Yannis Papakonstantinou (CSE) as co-PI. Duration: 2002-2004.

Software Development

1. One of the group leaders and developers of IBM's DB2 Parallel Edition Version 1.0, released commercially in December 1995.
2. One of the designers of the SDSC Storage Resource Broker (SRB). Version 1 was released in September 1997.
3. One of the designers of the Data Integration Cart™ technology for ontology-based data integration (invention disclosure filed: 2007).

U.S. Patents

1. *Persistent Archives*, R. Moore, A. Rajasekar, **C. Baru**, B. Ludaescher, A. Gupta, R. Marciano, US Patent 7,349,915, March 25, 2008. Assigned to Nirvana, Inc.
2. *Persistent Archives*, R. Moore, A. Rajasekar, **C. Baru**, B. Ludaescher, A. Gupta, R. Marciano, US Patent 6,963,875, November 8, 2005. Assigned to Nirvana, Inc.
3. *System and method for construction, storage, and transport of presentation-independent multimedia content*, **C. Baru**, J. Chase, T. Elvins, R. Fassett, E. Nebel, Patent No. 7,028,252, March 22, 2001. Assigned to Oracle, Corp.
4. *Method and apparatus for achieving uniform data distribution in a parallel database system*, **C. Baru** and F. Koo, Patent No. US5970495, IBM, Oct. 19, 1999.
5. *Method and apparatus for implementing partial declustering in a parallel database system*, **C. Baru**, G. Fecteau, J. Kirton, L. Kollar, F. Koo, Patent No. US5878409, IBM, March 2, 1999.

Ph.D. Committees Chaired

- Ophir Frieder. Dissertation title: "*Database processing on a cube-connected multicomputer system*," EECS Dept., University of Michigan, Dec. 1987. Recipient of IBM Graduate Fellowship Award. Currently, chaired professor and Chair of Computer Science Department, Georgetown University.
- Piyush Goel. Dissertation title: "*Dataflow query processing and optimization*," EECS Dept., University of Michigan, May 1992. Was co-Founder of Everypath.com, San Jose, CA.
- Sriram Padmanabhan. Dissertation title: "*Data placement in shared-nothing parallel database systems*," EECS Dept., University of Michigan, July 1992. Recipient of IBM Graduate Fellowship Award. Currently, Distinguished Engineer, IBM Infosphere, IBM Silicon Valley Labs.

Key Committee Memberships

- Chair, SDSC Project Advisory Council, 2009-2012.

- Lead, TeraGrid Data Working Group, 2001–2002.
- Member of Review Committee, Canada Research Chairs program, Natural Sciences and Engineering Research Council (NSERC) of Canada, 2000–2002.
- Member of the Architecture Working Group, California Digital Library, University of California, Office of the President, Oakland, CA, 1998–2000.
- Member of the *Grants Selection Committee (GSC)* for Computer and Information Sciences, Natural Sciences and Engineering Research Council (NSERC) of Canada, 1994-97. (The GSC is responsible for annually reviewing grant proposals from computer science faculty in Canada and making funding decisions).
- IBM representative on the Transaction Processing Council's TPC-D Benchmark Standard Subcommittee, 1993-95. Participated in drafting the original TPC-D specification.

Funded Research

- Principal Investigator for over 25 grants, total ~\$18M in funding.
- Co-principal Investigator for about 10 grants, ~\$8M funding.
- Senior Personnel in over 25 projects.

Books

1. *Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences*. Edited by Randy Keller and Chaitanya Baru. Cambridge University Press. June 2011, ISBN-10: 0521897157 | ISBN-13: 978-0521897150 | Edition: 1.

Book Chapters

1. “Science Gateways: Harnessing clouds and software services for science,” N. Wilkins-Diehr, C. Baru, D. Gannon, K. Keahey, J. McGee, M. Pierce, R. Wolski, W. Wu. To appear in *Cloud Computing and Software Services: Theory and Techniques*, under preparation. CRC Press.
2. “Data Intensive Computing,” (with Moore, R., Marciano, R., Rajasekar, A., and Wan, M.), in *The Grid: Blueprint for a New Computing Infrastructure*, edited by Ian Foster and Carl Kesselman, January 1999, Morgan Kaufmann.
3. “Virtualization Services for Data Grids” (with Moore, R.), to appear in, *Grid Computing: Making the Global Infrastructure a Reality*, edited by Fran Berman, Geoffrey Fox, and Tony Hey, to be published February 2003, Wiley.

Journal Articles

1. “Benchmarking Big Data Systems and the BigData Top100 List,” C. Baru, M. Bhandarkar, R. Nambiar, M. Poess, T. Rabl, *Big Data Journal*, Vol.1, No.1, March 2013.
2. “Setting the Direction for Big Data Benchmark Standards”, C. Baru, M. Bhandarkar, R. Nambiar, M. Poess, and T. Rabl, *Selected Topics in Performance Evaluation and Benchmarking*, Lecture Notes in Computer Science, Volume 7755, pp 197-208, 2013, DOI: 10.1007/978-3-642-36727-4_14.
3. “Cyberinfrastructure for Observatory/Monitoring Networks: A Case Study from the TEAM Network,” Baru, C., Fegraus, E., Andelman, S., Chandra, S., Kaya, K., Lin, K., Youn, C., *BioScience*, July 2012, Vol.62, No.7: 667–675. ISSN 0006-3568, electronic ISSN 1525-3244.
4. “Database Design for High-Resolution LIDAR Topography Data,” V. Nandigam, C. Baru, C. J. Crosby, in *SSDBM 2010*, Lecture Notes in Computer Science 6187, edited by M. Gertz and B. Ludascher, pp. 151–159, June 2010.
5. “Mediating among GeoSciML resources,” C. Baru and K. Lin. Volume 2, Supplement 1, 2009. *International Journal of Digital Earth*. Publishers: Taylor & Francis.
6. “The GEON service-oriented architecture,” C. Baru, S. Chandra, K. Lin, A. Memon, C. Youn. Volume 2, Supplement 1, 2009. *International Journal of Digital Earth*. Publishers: Taylor & Francis.

7. "Sharing and caring of eScience data," C. Baru. International Journal on Digital Libraries, Springer, ISSN: 1432-5012. Vol. 7, Nos. 1-2, Oct 2007. DOI: 10.1007/s00799-007-0029-2.
8. "An extensible information model for shared scientific data collections," (with A. Gupta), *Journal on Future Generation Computer Systems*, 1999.
9. "DB2 Parallel Edition," (with Fecteau et al), *IBM Systems Journal*, April 1995.
10. "Site selection and query scheduling policies for a multicomputer database system," (with O. Frieder), *IEEE Transactions on Knowledge and Data Engineering*, August 1994.
11. "Inter-subcube routing in hypercubes," (with S. Padmanabhan), accepted for publication in the *IEEE Transactions on Parallel and Distributed Systems*, 1993.
12. "Join and data redistribution algorithms for hypercubes," (with S. Padmanabhan), *IEEE Transactions on Knowledge and Data Engineering*, June 1993.
13. "Squashed embedding of E-R schemas in hypercubes," (with P. Goel), *Journal of Parallel and Distributed Computing*, Vol.8, No.4, April 1990, pp.340-348, Academic Press.
14. "Database operations in a cube-connected multicomputer system," (with O. Frieder), *IEEE Trans. on Computers*, Vol. 38, No. 6, June 1989, pp.920-927.
15. "The Architecture of SM3: A dynamically partitionable multicomputer system," (with S.Y.W. Su), *IEEE Trans. on Computers*, Vol. 35, No. 9, Sept. 1986, pp.790-802. Also appeared in the IEEE Tutorial on Parallel Architecture for Database Systems, IEEE Computer Society Press, 1989.
16. "Dynamically partitionable multicomputers with switchable memories," (with S.Y.W. Su), *Journal of Parallel and Distributed Computing*, Vol. 1, No. 2, Academic Press, Nov. 1984, pp.152-184.

Conference Presentations

1. TEAM Network: Building Web-based Data Access and Analysis Environments for Ecosystem Services, Youn, C., S. Chandra, E.H. Fegaus, K. Lin, and C. Baru, Proceedings of the International Conference on Computational Science. 4:146-155, 2011.
2. Data acquisition and management software for camera trap data: A case study from the TEAM Network, Fegaus, E., Lin, K., Ahumada, J., Baru, C., Chandra, S., Youn, C., *Ecological Informatics*, 6 (2011) 345-353.
3. NMR Cyberinfrastructure: Web-based virtual file system for managing distributed NMR data, Youn, C, Baru, C., Mrse, A., and O'Connor, J., Gateway Computing Environments Workshop (GCE), 2010, DOI: 10.1109/GCE.2010.5676122.
4. Evaluation of MapReduce for Gridding LIDAR data, Krishnan, S., Baru, C., and Crosby, C.J. *Proceedings of the 2nd IEEE International Conference on Cloud Computing Technology and Science*, Dec 2010.
5. M. Smeekens, C. Baru, G.R. Keller, R. Arrowsmith, C. Crosby, "Teaching and Training in Geoinformatics: Experiences from the Cyberinfrastructure Summer Institute for Geoscientists (CSIG)," to be presented at the Fall AGU, Dec. 2009, San Francisco.
6. C. Crosby, V. Nandigam, J. R. Arrowsmith, S. Balakrishnan, N. Alex, C. Baru (2008), "A Cyberinfrastructure Platform for Distribution of GeoEarthScope LiDAR Topography Data," *Eos Trans. AGU*, 89(53), Fall Meet. Suppl., Abstract IN51A-1146.
7. C. Baru and K. Lin, "Federating metadata catalogs," (invited), European Geosciences Union, General Assembly, Vienna, Austria, 19-24, 2009, EGU2009-11848.
8. S. Andelman, C. Baru, S. Chandra, E. Fergaus, K. Lin, R. Unwin (2009), "A data scheduling and management infrastructure for the TEAM Network," European Geosciences Union, General Assembly, Vienna, Austria, 19-24, 2009, EGU2009-11095.
9. C. Crosby, V. Nandigam, C. Baru (2009), "Recommendation systems for Geoscience data portals – Harnessing usage patterns," European Geosciences Union, General Assembly, Vienna, Austria, 19-24, 2009, EGU2009-6620.
10. A. Agarwal, C. Baru, C. Crosby, R. Keller, K. V. Subbarao, V. Nandigam, "iGEON-India: An international collaborative activity of the GEON Project," European Geosciences Union, General Assembly, Vienna, Austria, 19-24, 2009, EGU2009-11792.

11. "Generating composite thematic maps from semantically-different collections of shapefiles and map services," November 2004, Memon, G., A. Memon, K. Lin, I. Zaslavsky, C. Baru, ESRI 2005
12. Geoinformatics: Coordination of Environmental Cyberinfrastructure for the Earth Sciences
13. November 2002, Allison, M.L., C. Baru, T.H. Jordan, Abstract, GSA Annual Meeting, November 2-5, 2003, Seattle, WA
14. "The GEON Grid Software Architecture," Twenty-fourth Annual ESRI International User Conference, August 9-13, 2004, San Diego, CA.
15. "Standards-based Secure Invocation of ArcWeb Services," (with Ashraf Memon, Ilya Zaslavsky, Steve Mock, and Amit Behere), Twenty-fourth Annual ESRI International User Conference, August 9-13, 2004, San Diego, CA
16. "Interoperability of Databases and Software Tools: Building Geoinformatics and the Cyberinfrastructure," poster (with Lee Allison), American Association of Petroleum Geologists Annual Meeting, April 18-21, 2004, Dallas, TX
17. "GEON: Cyberinfrastructure for the Geosciences," (with Dogan Seber, Randy Keller, and Krishna Sinha), American Geophysical Union, Fall Meeting, 2003, San Francisco, CA.
18. "Creating Grid Services to Enable Data Inter-operability: An Example from the GEON Project," (with Bhatia, K., Memon, A., Zaslavsky, I., Seber, D), GSA Annual Meeting, November 2-5, 2003, Seattle, WA.
19. "Semantic Mediation Services in Geologic Data Integration: A Case Study from the GEON Grid," (with Lin, Kai, Bertram Ludaescher, Boyan Brodaric, Dogan Seber, and Krishna Sinha), GSA Annual Meeting, November 2-5, 2003, Seattle, WA
20. "Building the Geoinformatics System: Coordination of the environmental Cyberinfrastructure for the Earth Sciences," (with Lee Allison and Tom Jordan), GSA Annual Meeting, November 2-5, 2003, Seattle, WA.
21. "GEON: The Geoinformatics Network," (with Krishna Sinha), Annual Meeting of the Geological Society of America, October 26-30, 2002, Denver, CO.
22. "Grid-Enabled Mediation Services for Geospatial Information," (with Ilya Zaslavsky), Workshop on Next-Generation Geospatial Information, October 20-21, 2003, Cambridge, MA.
23. "Data Access and Management Services on the Grid," (with Raman, V., Narang, I., Crone, C., Haas, L., Malaika, S., Mukai, T., Wolfson, D.), Global Grid Forum 5, July 21-24, 2002, Edinburgh, Scotland.
24. "Information integration of geospatial information", Demo, NSF Digital Government Conference, May 19-21, 2002, Redondo Beach, CA.
25. "Information Integration Technologies for Current Research Information Systems," *CRIS 2000 Conference*, May 25-27, 2000, Helsinki, Finland.
26. "XML-based information mediation for digital libraries," (with Papakonstantinou, Gupta, Marciano, Ludaescher, Velikhov), demo at *ACM Digital Libraries'99*, Berkeley, CA, August 1999.
27. "XML-based information mediation with MIX," (with Papakonstantinou, Gupta, Marciano, Ludaescher, Velikhov), demo at *ACM-SIGMOD'99*, Philadelphia, PA, June 1999.
28. "Querying Web Mediators," (with Chu, Gupta, Ludaescher, Marciano, Papakonstantinou, Velikhov), demo at the *Internet2* meeting, Washington, DC, April 1999.
29. "Xviews: XML views of relational schemas," in *Database and Expert Systems Applications (DEXA'99)*, Sept. 2, 1999, Florence, Italy.
30. "Integrating GIS and Imagery through XML-Based Information Mediation," (with Gupta, Marciano, Zaslavsky), NSF International Workshop on Integrated Spatial Databases: Digital Images and GIS, June 14-16, 1999, Portland, Maine.
31. "The SDSC Storage Resource Broker," (with R. Moore, A. Rajasekar, M. Wan), in the *Proceedings of CASCON'98*, Nov.30-Dec.3, 1998, Toronto, ON.
32. "A Hierarchical Access Control Scheme for Digital Libraries," (short paper, with A. Rajasekar), in the *Proceedings of the ACM Digital Libraries Conference 1998*, June 24-26, 1998, Pittsburgh, PA.
33. "Data Handling Architecture for a Prototype Federal Application," (with Choi et al), *Proceedings of the IEEE Mass Storage Systems Conference*, March 1998, College Park, MD.

34. "Managing Very Large Scientific Data Collections," (poster), *5th International Conference on High Performance Computing (HiPC'98)*, Dec.17-20, 1998, Chennai, India.
35. "Features and requirements for an XML view definition language: Lessons from XML information mediation," (with Ludaescher, Papakonstantinou, Vianu), position paper in *QL'98-The Query Languages Workshop*, World Wide Web Consortium, Dec.3-4 1998, Boston, MA.
36. "Integrating DB2 with HPSS," (poster, with M. Lo, S. Padmanabhan, and V. Gottemukalla) to be presented at *Supercomputing'98*, Nov.10-12, Orlando, FL.
37. "Archiving Metadata," (poster), *European Conference on Digital Libraries, EuroDL'98*, September 21-23, 1998, Crete, Greece.
38. "Information Based Computing," (with Moore, Karin, Rajasekar) *Proceedings of the Workshop on Research and Development Opportunities in Federal Information Services*, March 1997.
39. "Metadata to Support Information-Based Computing Environments," (with Moore et al), *IEEE Conference on Metadata, Metadata-97*, September 10-11, 1997, Silver Springs, MD.
40. "An Overview of DB2 Parallel Edition," (with G. Fecteau et al), *ACM SIGMOD-95, International Conference on the Management of Data*, San Jose, CA, May 1995.
41. "TPC-D: Analysis of an industry standard decision support systems benchmark," (with P. Larson, T.S. Liu, B. Schiefer, and S. Retnasena), *Proceedings of DB/EXPO-94*, San Francisco, May 1994.
42. "Data reorganization in parallel database systems," (with Daniele C. Zilio, *IEEE Workshop on Advances in Parallel and Distributed Systems*, October 6, 1993, Princeton, N.J.
43. "Data placement in shared-nothing parallel database systems," (with S. Padmanabhan), *Proceedings of the Conference on Information and Knowledge Management*, November 1992, Baltimore, MD.
44. "Routing between subcubes in a hypercube," (with S. Padmanabhan), *Proceedings of the 6th Distributed Memory Computing Conference*, Portland, OR, April 1991.
45. "A comparison of join algorithms for hypercubes," (with S. Padmanabhan, Q. Stout, and B. Wagar), *Proceedings of the 4th Hypercube Conference*, Monterey, CA., March 1989.
46. "Mapping E-R schema graphs onto hypercubes," (with P. Goel), *Proceedings of the 4th Hypercube Conference*, Monterey, CA., March 1989.
47. "Query scheduling and site selection algorithms for a cube-connected multicomputer system," (with O. Frieder), *Proceedings of the International Conference On Distributed Computer Systems*, June 1988, San Jose, CA.
48. "Study of dynamic data redistribution for database processing," (with O. Frieder), *Proceedings of the 3rd Hypercube Conference*, Pasadena, CA., Jan. 1988.
49. "Join on a cube: Analysis, simulation, and implementation," (with O. Frieder, D. Kandlur, and M. Segal), *Proceedings of the 5th International Workshop on Database Machines*, Oct. 1987, Karuizawa, Japan, pp.74-87.
50. "Data distribution and query scheduling policies for a cube-connected multicomputer system," (with O. Frieder), *Proceedings of the 2nd International Conference on Supercomputing Systems*, San Francisco, CA. May 1987.
51. "Issues in query processing on a cube-connected multicomputer system," (with O. Frieder), *Proceedings of the 6th IEEE Pheonix Conference on Computers and Communications*, Scottsdale, AZ., Feb. 25-27, 1987, pp.315-319.
52. "Implementing relational database operations in a cube-connected multicomputer system," (with O. Frieder), *Proceedings of the 3rd IEEE International Conference on Data Engineering*, Los Angeles, CA., Feb.2-6, 1987.
53. "Matrix multiplication on a multicomputer system with switchable main memory modules," (with A. Thakore and S.Y.W. Su), *Proceedings of the 1st International Supercomputing Systems Conference*, Tarpon Springs, FL., Dec.17-20, 1985, pp.650-659.
54. "Performance evaluation of the statistical aggregation by categorization in the SM3 system," (with S.Y.W. Su), *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, Boston, MA., June 18-21, 1984, pp.77-89.

55. "SM3: A dynamically partitionable multicomputer system with switchable main memory modules," (with T. Fei and S.Y.W. Su), *Proceedings of the IEEE International Conference on Data Engineering*, Los Angeles, CA., April 24-27, 1984, pp.42-49.

Conference and Workshop Organization

1. Program Chair, 4th Workshop on Big Data Benchmarking, IEEE BigData Conference, October 2013, San Jose, CA (with Shane Canon, LBL; Ian Foster, U.Chicago/ANL).
2. Program Chair, Workshop on Big Data in Science, IEEE BigData Conference, October 2013, San Jose, CA (with Tilmann Rabl, U.Toronto; Milind Bhandarkar, Pivotal).
3. Member, Organizing Committee, 7th Extremely Large Databases Conference (XLDB), September 9-12, 2013, Palo Alto, CA.
4. Chair, Program / Organizing Committee, Third Workshop on Big Data Benchmarking, July 16-17, 2013, Xi'an, China.
5. Chair, Program / Organizing Committee, Second Workshop on Big Data Benchmarking, December 17-18, 2012, Pune, India.
6. Chair, Program / Organizing Committee, First Workshop on Big Data Benchmarking, May 8-9, 2012, San Jose, CA. Meeting sponsored by NSF/CISE.

Involved as organizer, co-organizer, Program Committee member, or session chair for more than 35 different conferences / workshops.