

Vuk Ercegovic

address: 650 Harry Road, San Jose, CA 95120 e-mail: vercego@us.ibm.com
telephone: (W) 408-927-1146 web: <http://www.cs.wisc.edu/~vuk>

Education

- PhD: Computer Science, University of Wisconsin-Madison, 2002-2006. Advisors: Raghu Ramakrishnan and David J. DeWitt. Minor: Statistics. Thesis title: Integrating Text Search and Relational Databases: Functionality and Performance
- MS: Computer Science, University of Wisconsin-Madison, 1998-2000.
- BS: Electrical Engineering and Computer Science, University of California at Berkeley, 1993-1997.

Publications

Refereed Publications

1. Spyros Blanas, Jignesh M. Patel, Vuk Ercegovic, Jun Rao, Eugene J. Shekita, Yuanyuan Tian: A Comparison of Join Algorithms for Log Processing in MapReduce. SIGMOD Conference 2010: 975-986
2. Fei Xu, Kevin S. Beyer, Vuk Ercegovic, Peter J. Haas, Eugene J. Shekita: E = MC3: Managing Uncertain Enterprise Data in a Cluster-computing Environment. SIGMOD Conference 2009: 441-454
3. Kevin S. Beyer, Vuk Ercegovic, Rajasekar Krishnamurthy, Sriram Raghavan, Jun Rao, Frederick Reiss, Eugene J. Shekita, David E. Simmen, Sandeep Tata, Shivakumar Vaithyanathan, Huaiyu Zhu: Towards a Scalable Enterprise Content Analytics Platform. IEEE Data Eng. Bull. 32(1): 28-35 (2009)
4. Vuk Ercegovic, Vanja Josifovski, Ning Li, Mauricio R. Mediano, Eugene J. Shekita: Supporting Sub-Document Updates and Queries in an Inverted Index. Conference on Information and Knowledge Management (CIKM), pp. 659-668, Napa, CA, 2008.
5. Bishwaranjan Bhattacharjee, Vuk Ercegovic, Joseph Glider, Richard Golding, Guy Lohman, Volker Markl, Hamid Pirahesh, Jun Rao, Robert Rees, Frederick Reiss, Eugene Shekita, Garret Swart, "Impliance: A Next Generation Information Management Appliance", Procs. of the Third Biennial Conference on Innovative Data Systems Research (CIDR 2007), pp. 351-360, Monterey, CA, Jan. 2007
6. Vuk Ercegovic, David J. DeWitt and Raghu Ramakrishnan. *The TEXTURE Benchmark: Measuring Performance of Text Queries over a Relational Database*. Proceedings of the 31st International Conference on Very Large Data Bases (VLDB), pp. 313-324, Trondheim, Norway, 2005.
7. Kristen Lefevre, Rakesh Agrawal, Vuk Ercegovic, Raghu Ramakrishnan, Yirong Xu and David J. DeWitt. *Limiting Disclosure in Hippocratic Databases*. Proceedings of the 30th International Conference on Very Large Data Bases (VLDB), pp. 108-119, Toronto, Canada, 2004.
8. Raghu Ramakrishnan, Andrew Baptist, Vuk Ercegovic, Matt Hanselman, Navin Kabra, Amit Marathe and Uri Shaft. *Mass Collaboration: A Case Study*. International Database Engineering and Applications Symposium (IDEAS), pp. 133-146, Coimbra, Portugal, 2004.
9. Navin Kabra, Raghu Ramakrishnan and Vuk Ercegovic. *The QUIQ Engine: A Hybrid IR-DB System*. Proceedings of the 19th International Conference on Data Engineering (ICDE), pp. 741-743, Bangalore, India (short paper), 2003.

10. A. Woodruff, C. Olston, A. Aiken, M. Chu, V. Ercegovac, M. Lin, M. Spalding and M. Stonebraker. *DataSplash: A Direct Manipulation Environment for Programming Semantic Zoom Visualizations of Tabular Data*. Journal of Visual Languages and Computing, Special Issue on Visual Languages for End-user and Domain-specific Programming, 12(5), pp. 551-571, October 2001.

Other Publications

1. Navin Kabra, Raghu Ramakrishnan and Vuk Ercegovac. *The QUIQ Engine: A Hybrid IR-DB System*. University of Wisconsin-Madison Technical Report TR-1449, 2002.
2. Chris Olston, Allison Woodruff, Alexander Aiken, Michael Chu, Vuk Ercegovac, Mark Lin, Mybrid Spalding and Michael Stonebraker. *Datasplash*. Proceedings ACM SIGMOD International Conference on Management of Data, pp. 550-552, Seattle, Washington (demonstration), 1998.

Patents

1. Raghu Ramakrishnan, Navin Kabra, Uri Shaft, Vuk Ercegovac, "Unified Database and Text Retrieval System", US Patent 6681222, January 20th, 2004.

Experience

- 09/2006 - Present:** Research Staff Member, IBM Almaden Research Center, San Jose, CA
Architect and co-founder of the Jaql project (<http://code.google.com/p/jaql/>), a high-level, functional language for scalable data processing. Jaql achieves its scale through automatic exploitation of MapReduce (Hadoop). I have experience developing, using, and supporting Jaql through several research projects, commercial products, and client engagements.
- 02/2002 - 07/2006:** Research Assistant, Computer Science, University of Wisconsin - Madison
Advised by Raghu Ramakrishnan.
- 03/1999 - 06/2002:** Senior Software Engineer, Quiq Inc., Madison, WI and San Mateo, CA.
Among first four founding engineers. I worked on design, implementation, and evaluation of the application server as well as underlying data sources such as databases, file system, and text indexing.
- 08/1998 - 12/1999:** Research Assistant, Computer Science, University of Wisconsin - Madison
Advised by Miron Livny on the DEVISE project (data visualization system).
- 04/1997 - 04/1998:** Research Assistant, Electrical Engineering and Computer Science, UC Berkeley.
Worked on Datasplash, a visualization system built on top of Postgres. For more information, see <http://datasplash.cs.berkeley.edu>.

Internships

- 05/2005 - 09/2005:** Summer Internship, Microsoft Research, Mountain View, CA.
Mentored by Douglas Terry. Designed a simulator for WinFS replication in order to evaluate its performance.
- 06/2003 - 09/2003:** Summer Internship, Microsoft, SQL Server Full-text Group, Redmond, WA.
- 04/1998 - 08/1998:** Summer Internship, Ubilab (UBS), Zurich, Switzerland.
Worked on a visualization system. See <http://www.macrofocus.com> for a commercialization of the project.
- 05/1996 - 08/1996:** Summer Internship, Oracle, Redwood Shores, CA.
- 10/1995 - 10/1996:** Undergraduate Research, Nuclear Engineering, UC Berkeley.
Designed and developed a GUI that helps nuclear engineers specify simulation parameters for various nuclear power plant configurations.
- 05/1995 - 08/1995:** Database Administrator, UCLA, Tissue Typing Lab, Los Angeles, CA.

Skills

Java, C#, C/C++, Perl, Hadoop, HBase, SQL Server, DB2, Oracle, MySQL, Postgres, Linux, Unix, Windows, Apache (ant, lucene, hadoop, hbase, tomcat, server), XML processing, MatLab.