Bill Howe

University of Washington, Department of Computer Science and Engineering
Box 351202, Seattle, WA 98195-2350

Professional Preparation

Georgia Institute of Technology	Atlanta, GA	BS, Honors, Industrial and Systems Engineering	1999
Portland State University	Portland, OR	PhD, Commendation, Computer Science	2006

Appointments

2014-present	Associate Director, eScience Institute, University of Washington
2014-present	Affiliate Associate Professor, Computer Science and Engineering, University of Washington
2012-2014	Director of Research, Scalable Data Analytics, eScience Institute, University of Washington
2009-2014	Affiliate Assistant Professor, Computer Science and Engineering, University of Washington
2009-2012	Senior Scientist, eScience Institute, University of Washington
2008-2009	Staff Scientist, NSF Science and Technology Center for Coastal Margin Observation and Prediction, Oregon Health & Science University
2006-2008	Senior Research Associate, NSF Science and Technology Center for Coastal Margin Observation and Prediction, Oregon Health & Science University
2001-2006	Graduate Research Assistant, Portland State University
1999-2001	Consultant, Deloitte Consulting, Microsoft, Schlumberger Inc., Siebel Systems.

Awards and Honors

Best of VLDB 2010, selected for publication in VLDB Journal special issue, for "Haloop: Efficient iterative data processing on large clusters." (currently the most cited paper from VLDB 2010)

Jim Gray Seed Award, Microsoft Research, $40,000, April 2010. Project: Client+Cloud: Bridging the Gap Between Spreadsheets and Databases for eScience

Jim Gray Seed Award, Microsoft Research, $25,000, April 2008. Project: The eScience Appliance

Department Commendation Award, Maseeh College of Engineering and Computer Science, Portland State University, May 2007

Best of VLDB 2004, one of 4 best papers (of 81) selected for publication in VLDB Journal special issue, for "Algebraic manipulation of scientific datasets"

Publications most closely related to the proposal

[1] K. R. Y. Kwon, M. Balazinska, and B. Howe. Hadoops adolescence: An analysis of hadoop usage in scientific workloads. In *VLDB*, 2013.

[2] M. Shaw, B. Howe, P. Koutris, and D. Suciu. Optimizing large-scale semi-naive datalog evaluation in hadoop. In *Datalog 2.0*, 2012.

[3] Y. Kwon, M. Balazinska, B. Howe, and J. A. Rolia. Skewtune: mitigating skew in mapreduce applications. In *SIGMOD Conference*, pages 25–36, 2012.

[4] H. Vo, B. Summa, J. Comba, J. Freire, B. Howe, C. Silva, and V. Pascucci. Parallel visualization on large clusters using mapreduce. In *Proceedings of Large-Scale Data Analysis and Visualization (LDAV '11)*, 2011.

[5] Y. Bu, B. Howe, M. Balazinska, and M. Ernst. Haloop: Efficient iterative data processing on large clusters. In *Proc. of International Conf. on Very Large Databases (VLDB)*, 2010.

Other publications

[6] B. Howe, G. Cole, E. Souroush, P. Koutris, A. Key, N. Khoussainova, and L. Battle. Database-as-a-service for long tail science. In *Proceedings of the 23rd Scientific and Statistical Database Management Conference (SSDBM '11)*. Springer, 2011.

[7] Y. Kwon, M. Balazinska, B. Howe, and J. Rolia. Skew-resistant parallel processing of feature-extracting scientific user-defined functions. In *Proc. of the ACM Symposium on Cloud Computing (SOCC 2010)*, June 2010.

[8] K. Grochow, B. Howe, M. Stoermer, and E. Lazowska. Client + cloud: Seamless architectures for visual data analytics in the ocean sciences. In *Proceedings of the 22nd International Conference on Scientific and Statistical Database Management (SSDBM '10)*. IEEE Computer Society, 2010.

[9] B. Howe, H. Green-Fishback, and D. Maier. Scientific Mashups: Runtime-Configurable Data Product Ensembles. In *Proceedings of the 21st International Conference on Scientific and Statistical Database Management (SSDBM '09)*. Springer, 2009.

[10] B. Howe, D. Maier, and L. Bright. Smoothing the ROI curve for scientific data management applications. In *Proceedings of the Third Biennial Conference on Innovative Data Systems Research (CIDR '07)*, Ansilomar, CA, 2007.

Synergistic Activities

eScience Institute. Director of Research, Scalable Data Analytics; outreach and consulting for big data in the sciences (http://escience.washington.edu)

Course development. Designed a coursera MOOC "Introduction to Data Science" with over 100,000 registrants and 9,000 earned certificates Spring 2013; Developed a new Introductory "Data Programming" Course, Summer 2012; New course "Data-Intensive Computing in the Cloud," Spring 2012; Advisory Board, Data Science Certificate, UW Educational Outreach; Advisory Board, Cloud Computing Certificate, UW Educational Outreach; "Scientific Data Management" (2010), University of Washington (with Magdalena Balazinska).

Science Advisory Board Member. SciDB project, http://scidb.org/

Organizing Committee Co-Chair, DataMASS 2012; Demo Co-chair, SSDBM 2013; Chair, HPCDB 2011/2012; XLDB Organizing Committee, 2011; Co-Chair, Workshop on Array Databases, 2011 (with Peter Baumann); Registration Chair, SSDBM 2011.

Reviewer. Reviewer, PVLDB, 2012-2013; Program Committee, LDAV 2013; Program Committee, Science-Cloud 2012; Reviewer, VLDB Journal, 2011; Program Committee, EDBT 2011; Demo Program Committee, SIGMOD 2011; Registration Chair, SSDBM 2011. Program Committee Program Committee, SSDBM 2010. Reviewer, Journal of Parallel and Distributed Computing, May 2010. Reviewer, VLDB Journal, 2007

Collaborators and other affiliations

Collaborators (outside the University of Washington):
Matthew Arrott (UCSD), Roger Barga (Microsoft Research), Mike Cafarella (University of Michigan), Ian Foster (University of Chicago), Christine Borgman (UCLA), Bryan Heidorn (Arizona), Carl Kesselman (USC), Lois Delcambre (Portland State University), Juliana Freire (NYU Poly), David Maier (Portland State University), Jerome Rolia (HP), Rich Signell (USGS), Claudio Silva (NYU Poly)

Advisors: David Maier, Portland State University, Thesis Advisor; Antonio Baptista, Oregon Health & Science University, Postdoctoral Advisor