

# Brian F. Cooper

## Office

1600 Amphitheatre Parkway  
Mountain View, CA 94043

## Internet

www.brianfrankcooper.net  
cooperb@google.com  
brianfrankcooper@gmail.com

## INTERESTS

Massive scale information systems: search, databases, storage systems, middleware systems, event systems and data stream processing

## EDUCATION

### **Ph.D. in Computer Science, 2003**

Stanford University, Stanford, California

*Dissertation: Information Preservation in Networks of Autonomous Archives*

Advisor: Professor Hector Garcia-Molina

### **M.S. in Computer Science, 2000**

Stanford University, Stanford, California

### **B.S. in Computer Science (highest honors), 1998**

### **B.A. in Chemistry, 1998**

University of Colorado, Boulder, Colorado

## EXPERIENCE

### **Google**

*Search Team*

*Software Engineer, June 2010-present*

Working in the search quality team on core ranking algorithms.

### **Yahoo! Research**

*Web Information Management Group*

*Principal Research Scientist, September 2006-June 2010*

Built massive scale distributed storage and cloud database systems for Yahoo!'s web serving and OLTP workloads.

- Served as key technical architect for PNUTS/Sherpa, a distributed, replicated database system in production at Yahoo! for web workloads. Sherpa provides low latency, high reliability and automated operations. Sherpa is designed to scale to thousands of servers in globally distributed datacenters.
- Managed a research group consisting of research scientists and engineers, and served as technical lead for the design and implementation of several Sherpa components.
- Worked wherever needed to move the project forward, including gathering requirements from customers, presenting to executives, writing and running integration and performance tests, and prototyping next generation features.

### **Georgia Institute of Technology**

*Division of Computing Sciences and Systems, College of Computing*

*Assistant Professor, August 2003-August 2006*

Conducted an active research program in distributed data and information systems involving students and faculty collaborators. Projects include:

- *Overlay Dynamic Information Networks (ODIN)* – Techniques for automatically adapting the structure of information network overlays to changing workloads, environments and requirements to result in higher performance
- *IFLOW/inTransit* – Techniques for distributed data stream processing, including

resource-aware deployment of processing tasks, automatic adaptation of deployment to respond to resource changes or failures, and responsiveness of the system to business concerns and workloads

- *InfoBeacons* – Techniques for routing user text searches through large networks to information sources, even when the sources are dynamic and uncooperative

### **Stanford University**

*Department of Computer Science*

*Research Assistant, September 1998- August 2003*

Conducted research in the archival repository project. Contributions include:

- *Stanford Archival Vault (SAV)* – Designed and implemented a reliable archival database system based on remote replication and repeated content checking. Implemented an ingest tool, *InfoMonitor*, which automatically scanned a filesystem for updated content
- *Data Trading* – Designed and tested algorithms for trading content between distributed, networked data archives; trading efficiently allocates resources to replicas to ensure fault tolerance
- *SIL model* – Developed a model of search overlays to allow resilient and efficient searching of content in scattered repositories

### **IBM Almaden Research Center**

*Database Technology Institute*

*Research Intern, July-September 2002*

Examined integration of SOAP request capability into DB2 database engine:

- Implemented features and enhanced stability for SOAP requestor user defined function
- Implemented SOAP requestor as a built in function of the DB2 engine
- Conducted performance studies of SOAP request/response and identified optimizations for latency reduction

### **RightOrder, Inc.**

*Enterprise software corporation, San Jose, California*

*Research Engineer, September 2000-June 2002*

Developed algorithms for efficient indexing and query processing over XML data, using trie-based path indexing

### **IBM T.J. Watson Research Center**

*Programming Technologies Department*

*Research Intern, May-August 1998*

Developed program profiling techniques for an adaptive Java compiler

- Designed and implemented run-time program profiler for an adaptive Java compiler
- Implemented program profile data structure based on calling context tree

### **University of Colorado**

*Department of Computer Science*

*Undergraduate Researcher, September 1997-May 1998*

Developed Java runtime profiling tools for the Compiler Optimization Group:

- Developed ProfBuilder, a package for rapid construction of runtime Java program profilers
- Designed and implemented program analysis tools based on dynamic construction of calling context tree and control flow graph structures

## SKILLS

I have been working primarily in C++ and Java, but have done some Perl as well. I am also familiar with SQL. I have written plugins to Apache, installed and run MySQL, Oracle, and Cassandra, implemented a P2P framework using sockets, and worked with a variety of other systems.

## SELECTED PUBLICATIONS

*Full list available at [www.brianfrankcooper.net/pubs](http://www.brianfrankcooper.net/pubs)*

1. Adam Silberstein, Jeffrey Terrace, Brian F. Cooper and Raghu Ramakrishnan. *Feeding Frenzy: Selectively Materializing Users' Event Feeds*. ACM SIGMOD Conference, Indianapolis, IN, USA, 2010.
2. Brian F. Cooper, Adam Silberstein, Erwin Tam, Raghu Ramakrishnan and Russell Sears. *Benchmarking Cloud Serving Systems with YCSB*. ACM Symposium on Cloud Computing (SoCC), Indianapolis, IN, USA, 2010.
3. Brian F. Cooper, Raghu Ramakrishnan and Utkarsh Srivastava. *Cloud Storage Design in a PNUTShell*. In "Beautiful Data," edited by Toby Segaran and Jeff Hammerbacher. O'Reilly, 2009.
4. Ymir Vigfusson, Adam Silberstein, Brian F. Cooper and Rodrigo Fonseca. *Adaptively Parallelizing Distributed Range Queries*. PVLDB 2(1-2), 2009 (VLDB Conference, Lyon, France, August 2009).
5. Parag Agrawal, Adam Silberstein, Brian F. Cooper, Utkarsh Srivastava and Raghu Ramakrishnan. *Asynchronous View Maintenance for VLSD Databases*. ACM SIGMOD Conference, Providence, RI, USA, 2009.
6. Brian F. Cooper, Raghu Ramakrishnan, Utkarsh Srivastava, Adam Silberstein, Philip Bohannon, Hans-Arno Jacobsen, Nick Puz, Daniel Weaver and Ramana Yerneni. *PNUTS: Yahoo!'s Hosted Data Serving Platform*. VLDB Conference (industry track), Auckland, New Zealand, 2008.
7. Adam Silberstein, Brian F. Cooper, Utkarsh Srivastava, Erik Vee, Raghu Ramakrishnan and Ramana Yerneni. *Efficient Bulk Insertion into a Distributed Ordered Table*. ACM SIGMOD Conference, Vancouver, BC, Canada, 2008.
8. Brian F. Cooper. *Trading off resources between overlapping overlays*. ACM/IFIP/USENIX 7th International Middleware Conference, Melbourne, Australia, 2006.
9. Yong Yang, Rocky Dunlap, Michael Rexroad and Brian F. Cooper. *Performance of Full Text Search in Structured and Unstructured Peer-to-Peer Systems*. IEEE INFOCOM, Barcelona, 2006.
10. Brian F. Cooper, Neal Sample, Michael J. Franklin, Gisli R. Hjaltason and Moshe Shadmon. *A fast index for semistructured data*. VLDB 2001.

## SERVICE

- Program committee member of ICDE 2011, VLDB 2010, ICDE 2010, SIGMOD 2009, ICDE 2009, SIGMOD 2006, ICDE 2006, ICDCS 2006, ICDE 2008, IPDPS 2008 and at least 20 other conferences and workshops.
- Area co-chair for ICDE 2011 (Cloud Computing and Web Applications track)
- Co-program chair of Tenth ACM/IFIP/USENIX International Middleware Conference (Middleware '09), December 2009, Champaign, IL.
- Co-program chair of Third International Workshop on Networking Meets Databases (NetDB '07), April 10, 2007, Cambridge, MA
- Co-program chair of the IEEE Workshop on Workflow and Data Flow for Scientific Applications (SciFlow 2006) in conjunction with ICDE
- Treasurer, ACM 1<sup>st</sup> Symposium on Cloud Computing (SoCC 2010)

- Local arrangements chair, ICDE 2006
- Reviewer for ACM Computing Surveys, ACM/IEEE Transactions on Networking, IEEE Transactions on Parallel and Distributed Systems, Software Practice and Experience, Information Systems, and at least 15 other journals
- Associate Editor, ACM SIGMOD Record (2004-present)
- Editor, ACM SIGMOD DiSC Electronic Literature Collection (2003-2006)
- Associate Editor, IEEE Data Engineering Bulletin