

Mehul A. Shah

mashah@gmail.com
(510) 793-4391

12543 Palmtag Drive, Saratoga, CA 95070
http://www.hpl.hp.com/personal/Mehul_Shah/

Objective: Technical leadership position in areas related to efficient and scalable data storage and management.

Interests: Distributed computing, databases – analytics and transactional systems, energy-efficient systems, storage systems

Education

University of California, Berkeley

Ph.D., EECS Department, Database group
Advisor: Joseph M. Hellerstein

October 2004

Massachusetts Institute of Technology

MEng., Electrical Engineering and Computer Science
B.S., Computer Science, and B.S., Physics

June 1996

June 1996

Professional Experience

Hewlett-Packard Laboratories

Research Scientist

Palo Alto, CA

October 2004 to present

Main-Memory DB: Leading a team to develop a scalable, distributed transactional main-memory store for consistent low-latency operations and complex on-the-fly analytics. Applications include financial trading and social networking.

Scalable Storage: Founding member of HPKVS: a highly available, low-cost key-value service for the cloud. HPKVS is an eventually consistent, erasure-coded store designed for storing large objects and spanning multiple geographies.

Sinfonia: Highly scalable, distributed transactional system for building data-center infrastructure applications like clustered file systems, distributed B-Trees, and group membership. Best paper SOSP 2007.

Energy-efficient Systems: Characterized energy-efficiency of enterprise-grade, audit-class data analysis systems. Inventor of JouleSort, the first holistic, energy-efficiency benchmark for computer systems. Built 2007 winner.

University of California, Berkeley

Graduate Student Researcher

Berkeley, CA

September 1997 to October 2004

Thesis: "Flux: A Mechanism for Building Highly-Available, Fault-Tolerant, Scalable Dataflows": Founder of the TelegraphCQ project. Developed techniques for robustly scaling high-throughput, 24x7, data-streaming applications. Focused on methods that provide fault-tolerance and load-balancing for parallel, data-stream-processing systems.

Continuously Adaptive Continuous Queries (CACQ): Developed an adaptive query processing system that executes numerous long-running queries simultaneously over streaming data.

DB2/OSF Group / IBM Almaden

Intern

San Jose, CA

January to October 1999

Investigated alternative strategies for implementing collection types in IBM DB2/UDB. Designed language extensions for querying collection types. Gained experience with administration and software development in DB2/UDB.

AT&T Laboratories – Research

MEng., Thesis Student

Murray Hill, NJ

June 1996 to January 1997

Developed ReferralWeb, a system that automatically generates social networks by mining the public web. It identifies and recommends recognized experts on user-specified topics, and provides a social path by which to contact those experts.

Software Released: TelegraphCQ (<http://telegraph.cs.berkeley.edu/>), Amdb (<http://gist.cs.berkeley.edu/>)

Languages and Systems: C, C++, Java, Python, Perl, Windows, and Linux

Published in: SIGMOD, SOSP, VLDB, ICDE, NSDI, EuroSys, HotOS, CIDR, PODC (Visit my webpage for papers.)

Awards: Siebel Scholars Fellowship, 2003. U.C. Microelectronics Fellowship, 1997. Henry Ford II Scholar Award, 1996.

Membership: ACM SIGMOD, USENIX, SortBenchmark Committee (<http://www.sortbenchmark.org/>)

Hobbies: Running and photography

References available upon request.