Peter A. Alvaro

810 Grove St.
San Francisco, CA 94117
(415) 673-8931
palvaro (upon) eecs (point) berkeley (point) edu
http://eecs.berkeley.edu/~palvaro

Education

Bachelor of Arts, Literature, Philosophy minor, September 1997
Middlebury College, Middlebury, VT

PhD student, Computer Science, Philosophy minor (August 2008 - present)
University of California, Berkeley
Advisor: Joseph M. Hellerstein

Honors

Finalist, Qualcomm Innovation Fellowship 2009

Undergraduate Honors

Phi Beta Kappa
Magna Cum Laude
Winner, Reid L. Carr prize for achievement in English Literature
Highest Honors in English Literature

Research Interests

Databases and Declarative Systems
Distributed Systems
Programming Languages

Related Coursework

- **cs263 -** Design of Programming Languages (George Necula)
- **phil290-4 -** Gentzen and the Sequent Calculus (Paolo Mancosu)
- **cs261 -** Security in Computer Systems (David Wagner)
- **cs286 -** Implementation of Database Systems (Michael Frankin)
- **cs268 -** Graduate Computer Networking (Randy Katz)
- **cs262a -** Advanced Topics in Somputer Systems (Eric Brewer)

Related Experience

Senior Software Engineer, Ask.com, Oakland, CA
August 2003 – April 2008

- Designed and implemented a distributed SQL query processing and data aggregation engine, to solve business intelligence problems over data whose volume was too large to process with traditional RDBMS technology.

- Devised a SQL generation system to simplify the details of aggregation and intersection of VL datasets, and to minimize the data warehouse code base. The result was a reduction of several orders of magnitude in the number of lines of code needed to perform summarization and reporting, and the automatic generation of documents describing the summary business rules.
- Developed a scalable, highly parallel platform for performing ETL and other data transformations on a dynamic cluster of worker nodes. The system guaranteed atomicity of individual steps, and made forward progress even in cases of massive component and network failures.
- Designed and implemented a main-memory dimensional data aggregator for real-time reporting over multicast clickstream data. The application had to run persistently using constant memory, as traffic volumes and dimensions changed over time.
- Created a nomenclature to describe session-based clickstream event chains, and an algorithm to produce them from raw HTTP log data.

Database Engineer, Ask Jeeves, Inc., Emeryville, CA
September 2000 – August 2003

- Logical and physical design of data warehouses for VL clickstream datasets. A novel dimensional model was required to losslessly accommodate the volatile nature of the input data.
- Implemented a frequent pattern mining application for detecting significant token combinations in user queries. I used the FP-Tree structure and algorithm, but needed to optimize it to process hundreds of millions of queries per execution.
- Procedural programming within the Oracle, SQL Server, MySQL and postgres environments.
- Data and application integration following acquisitions of other internet companies.

Publications

- Peter Alvaro, William R. Marczak, Neil Conway, Joseph M. Hellerstein, David Maier, Russell Sears. *Dedalus: Datalog in Time and Space*. In submission.
- Peter Alvaro, Tyson Condie, Neil Conway, Khaled Elmeleegy, Joseph M. Hellerstein, Russell C. Sears. *BOOM Analytics: Exploring Data-Centric, Declarative Programming in the Cloud.* To appear in Eurosys 2010.
- Tyson Condie, Neil Conway, Peter Alvaro, Joseph M. Hellerstein, Khaled Elmeleegy, Russell Sears. *MapReduce Online*. To appear in NSDI 2010.
- Peter Alvaro, Tyson Condie, Neil Conway, Joseph M. Hellerstein, Russell C. Sears. *I Do Declare: Consensus in a Logic Language*. In Proceedings of the SOSP Workshop on Networking Meets Databases (NetDB), 2009. Best Paper.
- Peter Alvaro, Tyson Condie, Neil Conway, Khaled Elmeleegy, Joseph M. Hellerstein, Russell C. Sears. *BOOM: Data-Centric Programming in the Datacenter.* UC Berkeley Technical Report No. UCB/EECS-2009-98 2009.
- Peter Alvaro, Dmitriy Ryaboy, Divy Agrawal. *Towards Scaleable Architectures for Clickstream Data Warehousing.* Databases in Networked Information Systems: 5th Intl. Workshop Proceedings, Japan, October 2007.

Open Source Software Contributions

- Bloom::Faster: a perl wrapper and c library implementing high-performance bloom filters. (http://search.cpan.org/~palvaro/Bloom-Faster-1.4/lib/Bloom/Faster.pm)
- Underlayer: a symmetric, decentralized load-balancing and parallel computing middleware. (http://sourceforge.net/projects/underlayer/)

- Baobab: A high performance frequent pattern mining application. (http://sourceforge.net/projects/baobab-fp/)