Ankur Dave

2703 Ridge Rd Apt. 203   Berkeley, CA 94709

ankurdave@gmail.com   http://ankurdave.com

mobile: 650-701-7705

RESEARCH EXPERIENCE

UC Berkeley AMPLab   *Berkeley, CA*
Undergraduate Research Assistant   September 2010 to May 2013

Contributed to Spark, an open source cluster computing framework written in Scala. Advised by Matei Zaharia and Scott Shenker.

- Led the design and development of Arthur, a distributed replay debugger for Spark programs; gave a talk at the AMPLab Winter 2012 retreat; and wrote a technical report [3]
- Used Spark to build Bagel, a Pregel-like graph processing framework, and presented a poster at the AMPLab Summer 2011 retreat
- Coauthored an NSDI 2012 paper on Spark [1] that won Best Paper Award, contributing Bagel and an evaluation of Spark's user-controllable partitioning

Microsoft Research   *Redmond, WA*
Research Intern   June to August 2010

Designed and implemented CloudClustering, a scalable clustering algorithm on the Windows Azure cloud, for Microsoft Research's eXtreme Computing Group using C#. Generalized the design into a set of architectural patterns for data processing using cloud services. Published a workshop paper [2] and gave a talk at DataCloud 2011.

INDUSTRY EXPERIENCE

Google   *Mountain View, CA*
Software Engineer Intern   May to August 2012

Contributed to Google's workflow execution system, which handles collections of processes with dependencies.

- Designed and implemented an algorithm to improve efficiency at allocating cluster resources for complex workflows
- Productionized a tool for static analysis of user workflows to help design and prioritize new features
- Contributed cache-related bugfixes and features to Google's next-generation data processing substrate

Facebook   *Palo Alto, CA*
Software Engineer Intern   May to August 2011

Added network usage tracking and limit enforcement to Facebook's cluster manager using C++ and Linux cgroups.

PUBLICATIONS

Conference and Workshop Papers

[1] Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauley, Michael Franklin, Scott Shenker, and Ion Stoica. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing, *NSDI 2012,* April 2012. *Best Paper Award* and *Honorable Mention for Community Award.*

[2] Ankur Dave, Wei Lu, Jared Jackson, and Roger Barga. CloudClustering: Toward an Iterative Data Processing Pattern on the Cloud. *DataCloud 2011*, May 2011.

Technical Reports

[3] Ankur Dave, Matei Zaharia, Scott Shenker, Ion Stoica. Arthur: Rich Post-Facto Debugging for Production Analytics Applications. *Submitted to USENIX ATC 2013*, January 2013.

[4] Ankur Dave. Optimizing Boggle Boards: An Evaluation of Parallelizable Techniques. *IB Extended Essay*, January 2009.

EDUCATION

University of California, Berkeley   Starting August 2013
Ph.D. candidate, Computer Science

University of California, Berkeley   August 2010 to May 2013
B.S., Electrical Engineering and Computer Science   GPA: 3.79/4.0

SKILLS

- Programming languages: Scala, Python, C, C++, Scheme, C#, Java, PHP, JavaScript