

Aaron Beppu

Veteran software engineer seeks to apply data engineering and ML skills to problems with clear social impact.

Experience

2013 - Now, Sr to Principal Engineer, Sift (San Francisco, CA)

I have been a key contributor within engineering during a period of continued and transformative growth. During that period I have driven many of the projects underpinning that transformation. Here's a sample:

Product:

- A flexible workflow automation product. Customers can describe “if X then Y else Z” conditions and automations which have access to our ML features and predictions.
- A real-time reporting product. Customers can see up-to-the-moment reports and aggregates describing their use of our product.

Data hacking:

- A model calibration system which supports migrating between models while preserving marginal score distributions. This lets us change/improve models without disrupting downstream automations.
- Designed and built jobs to track prediction accuracy over time. Designed to consider widely differing customers, and labels which may arrive months after predictions.

Platform/Infra:

- Migrated between incompatible versions of HBase, with no downtime. Involved dual-writing through all tables, backfill jobs to transform between HFile versions, and tooling to check values at a byte-level.
- Encrypted all data at rest. Involved replacing every instance in our fleet, with new images with different FS initialization.
- Migrated from SQS to Kafka as our primary queue. Included developing a suite of libraries to handle multi-threaded consumption, retries, monitoring, etc.

Org Hacking:

- Led a rotation-based system to pay down technical debt.
- Organized and launched a trial of “project-based teams” – temporary, inter-disciplinary teams scoped to key projects.
- Ethics Committee Member

Feb - July 2013, Software Engineer, Prismatic (San Francisco, CA)

Built and improved a range of backend services, including topic modeling, document life-cycle, and social media integrations.

Jan 2011 - Jan 2013, Software Engineer, Etsy (New York, NY)

Data-mining system to improve search ranking. (e.g. see my [Hadoop World 2011 presentation](#))

Big data tools and infrastructure:

- Migrating from EMR to a Hadoop cluster on our own hardware, including a large codebase of legacy jobs
- Breaking performance bottlenecks in our ongoing processing
- Tooling for creating, scheduling and running workflows of jobs
- Tooling for monitoring and error-recovery

Jun 2008 - Dec 2010, Software Development Engineer, A9.com (Palo Alto, CA)

Wrote and improved jobs for large-scale click-stream analysis. Aggregating and reporting data about product search and search quality.

Education

2005 - 2008 **BA, Cognitive Science**; UC Berkeley, Departmental Citation

Public communications

Patent **US10339472B2**

Technical blog **Models in Disguise: How Sift Science Ships Non-Disruptive Model Changes**

Public speaking **Non-disruptive Model Changes**

Academic publication Beppu, Aaron & Griffiths, Thomas L. (2009). **Iterated learning and the cultural ratchet**. In N. A. Taatgen & H. van Rijn (eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. pp. 2089–2094.

Technologies

Languages I have used in production: java, scala, python, clojure, javascript, php

libraries/frameworks/services: hadoop, spark, kafka, hbase, elastic search, solr, mysql, postgres, memcached, airflow, oozie, protobuf, avro, thrift, AWS services (many), GCP services (several)