Mark Cahill

hello@thinkjson.com
www.thinkjson.com
github.com/thinkjson

Experience

YouVersion, Data Analyst and Developer
2011-2013
- Developed and maintained a scalable analytics infrastructure that allowed us to analyze 100 billion requests per year
- Wrote a number of data processing applications in both Python and Java and deployed them to Hadoop on Google Compute Engine and Elastic MapReduce
- Developed a self-service business intelligence application that interfaced with BigQuery and enforced row-level multitenancy, allowing Bible societies to access usage data on their content
- Successfully deployed BigQuery as an enterprise data warehouse solution
- Learned a new database technology each of the first ten months I was on the job
- Extended the telemetry system I developed at Analytical Labs to capture mobile telemetry data, and wrote backends for a variety of data storage mechanisms

Dynamo BI, Contractor
2011
- Implemented HTML5 analysis frontend that utilizes precalculated dimensions served over JSONP

Analytical Labs, Contributor
2010 – 2011
- Developed a distributed telemetry system for capturing exceptions and crash information
- Worked with the user interface team in creating a rich web interface to the Saiku open source analysis tool on Backbone.js and HTML5 that loaded in 300ms

Stetson University, Development Operations Coordinator
2006 – 2011
- Developed a self-service business intelligence and workflow management application that streamlined data access and made the division more agile
- Maintained a MySQL data warehouse
- Coordinated data mining, statistical analysis, and multivariate regression predictive
- Managed a team of 30 student callers in a fundraising call center, as well as ran the direct mail program

Education

B.S. Computer Science
Stetson University
Partially complete

B.A. Communication
Palm Beach Atlantic University
Fall 2004

Skills

Programming languages
Python *(MapReduce, shell programs, APIs, web application development)*
Java *(MapReduce, streaming JSON processing)*
C/C++ *(data processing, log conversion)*
Node.js *(distributed telemetry system)*
PHP *(business intelligence platform)*

Database technologies
MySQL, PostgreSQL, LucidDB, BigQuery *(database administration, data analysis)*
Hadoop and Hadoop Streaming *(distributed data analysis)*
CouchDB *(data collection and aggregation)*
Redis *(atomic queue for data processing applications)*

Projects

haproxy2tab (C++)
Converts HAProxy logs into tab-delimited format, ready to be loaded into a database for analysis

bqstreamload (Python)
A wrapper around the bq command line tool that reads data from stdin and streams it to a BigQuery table (used at YouVersion as part of our operational intelligence system)

django-maintenance (Python)
Django middleware that puts a Django application in maintenance mode (displays a notice and blocks use of the application) without altering the Django settings

django-issue-tracker (Python)
Coordinates workflow management, providing an easy-to-use interface for non-technical users