

Chester “Chet” Weger

chester.weger@gmail.com || 1(818)484-5557 || chet-weger.herokuapp.com

Summary Software Engineer passionate about delivering *simple, maintainable, and correct* solutions.

Experience

Senior Data Engineer – *Walt Disney Studio Technologies* via *Insight Global* – Burbank, CA Sept 2018 – April 2020

- Developed a suite of data pipelines processing marketing data from sources including Facebook Marketing API, Display & Video 360, Google Campaign Manager, and AdWords. This data provided reporting and analytics for a 40+ person Media and Analytics team spending millions on ads. Built these pipelines using *Python, Spark, and Airflow*.
- Collaborated with stakeholders to understand and apply data transformations, and troubleshoot and analyze data quality issues. Added unit tests to ensure code readability & maintainability. Automated reporting processes to save ~40 hours per week of manual excel-centric reporting.
- Proved the usefulness of the new data platform “Route 66” using tools such as *Apache Airflow, Apache Spark, DataDog, Docker, Sops, & GitLab*. Helped establish best practices around writing new data pipelines and engaged in code review and design discussions.

Software Development Engineer – *The Rubicon Project* – Playa Vista, CA

July 2014 – Sept 2018

- Wrote new Spark job for reporting on Rubicon’s “Prebid” traffic. Used Amazon EMR for rapid deployment and scaling.
- Designed & Developed Spark job for “First-Party” dataset persisted in Druid to sunset an older reporting product and add new functionality.
- Migrated several legacy Hadoop jobs processing *billions* of transaction logs per hour to Spark using Spark SQL and Dataset API, increasing performance and maintainability.
- Contributed to data-pipeline, API, and application layer for next-gen performance reporting product based on Druid capable of ingesting over a billion aggregated rows per month and allowing an arbitrary selection of over 40 dimensions and over 30 base and derived metrics.
 - Designed new tables, built REST API, and developed application for managing *saved, standard, and scheduled* reports.
 - Developed new dimension columns such as *Site Groups* that represent many-to-many relationships to deprecate legacy reporting.
- Contributed to *Blocked Advertiser Revenue* and *Optimized Price Floors* which produce recommendations based on historical data.

Student – *HMC Clinic Program* with *The Rubicon Project* – *Respectful Cross Device Attribution* Sept 2013 – May 2014

- Worked with three other students, an academic advisor, and a liaison from *The Rubicon Project* to gain industry experience and explore novel approaches to tackling a big-data problem.
- Project's goal is to be able to match devices – to be able to say these two devices might belong to the same person.
- Played a key role in creating the design for the project, pushed for similarity hashing as a technique to group device profiles.
- Researched, designed, and implemented similarity measures that operate on pairs of device profiles.

Software Engineering Intern – [Pagewoo/Nearwoo](#) (now defunct)

June 2013 – Aug 2013

- Developed a module to automatically generate information about a user’s business. Performed integration with the yelp business search API as well as building a scraping component to find, among other things, relevant pictures that would be useful in banner ad creation.
- Performed other miscellaneous scraping and web development tasks.

Technical Assistant – *Pomona College*; CS052: Foundations of Computer Science

Aug 2012 – Dec 2012

- Responsibilities included grading homework using an automated script and holding office hours to answer questions on homework problems.

Skills

- Proficient writing readable and idiomatic code in a variety of languages including *Python, Scala, JavaScript, Java, Bash, and SQL*.
- Pragmatic ability to balance the [benefits of functional programming](#), without [harming code readability and comprehension](#).
- Experience using Gitlab and Jenkins for testing and deployment. Used build tools including *setuptools, sbt, and maven*. Proficient using Docker.
- Experience applying [Data Engineering best practices](#) such as making tasks idempotent, and separating ingestion from processing.
- Experience using *Druid* for OLAP reporting on extremely high-volume transaction logs. Experience with *data-modeling* for OLAP & OLTP apps. Experience applying *data warehouse* design and concepts.
- Experience using *Django, vert.x*, and other frameworks to develop *RESTful* web services following best practices.

Career Highlights

- Developed pipelines for a 40+ person Media and Analytics team spending millions of dollars on programmatic ads on a variety of channels.
- Druid Performance Tuning – worked with [Imply.io](#) to implement optimizations for Druid querying and indexing. Benchmarked potential performance optimizations to evaluate cost effectiveness. Achieved over 40% reduction in average report query time.
- Won Internal Hackathon “Automation Prize” – used a Sankey chart to visualize the flow of money among marketplace entities.

Education

Bachelor of Arts, *Computer Science, Minor in Mathematics, 2014, Pomona College*, Claremont, California

Additional Courses

- Databricks training – [completed 3-day onsite training](#) course and discussed best practices and approaches to solving business needs.
- [Functional Programming in Scala](#), – completed five part Coursera specialization on functional programming and Spark.

Personal Projects

- [Interactive Tic-Tac-Toe](#): can you beat the AI?
- [Self Learning Meta Tic-Tac-Toe](#): watch the AI improve its state-evaluation constants using temporal difference learning.
- [Interactive Meta Tic-Tac-Toe](#): test your abilities against the AI. Adjust its search depth parameter to change its skill level. Adjust its evaluation constants to change its behavior.
- [Using the Copy Model for Edge Attachment to Evaluate the Stability of Graph Clustering](#): reproducing and extending results from a paper in graph/social-network analysis.