

Chester “Chet” Weger

chetweger@gmail.com || 909-706-2026 || chet-weger.herokuapp.com

Experience

Senior Software Engineer, Data Services – [Motional](#) – Santa Monica, CA July 2020 – Present

- Built dataset-service to manage the release of ground truth and auto-labeled datasets using python, CDK, Fargate, DynamoDB.
- Lead integrating data mining services into annotation data pipeline
 - Added functionality to import data into the Image pipeline from a "log-miner" service, allowing the user to find and annotate objects of interest e.g. active emergency vehicle lights.
 - Automated the import of "scenes" into the Fusion pipeline using MWAA (airflow), ECS, Docker, and python
 - Integrated the Fusion pipeline with log-miner, allowing researchers to annotate scenes with rare or generic objects.
 - Integrated the Fusion pipeline with defect tracking, allow importing scenes based on intervals where the perception system gave incorrect outputs.
- Lead enhancements for ground-truth data pipelines used to annotate lidar, radar, and image data using python, cv2, pyquaternion.
 - Converted image segmentation “masks” produced by [Scale](#) into internal “polygon” representation of image segmentation.
 - Created new “Traffic Light” pipeline to support annotating traffic lights in house.
 - Reworked the image pipeline to remove a step annotating 2D-boxes, and instead inferred 2D-boxes from instance-segmentation polygon groups.
 - Added support for instance aware semantic segmentation aka panoptic segmentation to reduce annotations costs.
 - Integrated with new annotation partner Annotell, for lidar box cuboid annotations. Sent pilot tasks to Annotell and parsed annotations in [OpenLabel format](#).
- Performed ad-hoc data-devops to correct systematic errors or fix operational issues.
 - Wrote and ran script to unrectify 133,434 2D-box annotations and 19,027 images to achieve data consistency across batches.
 - Answered questions on tooling and provided rapid hot-fixes and bug-fixes to unblock the data pipeline and annotations workflows.
 - Supported auditing and patching incorrectly annotated annotations, "closing the loop" on the continuous learning framework.
- Added integration tests for the annotations data pipelines. Created a *docker-compose* setup for running the annotations backend *REST APIs*. The integration tests, combined with many unit tests brought code coverage to over 60%, from under 5% when I joined.

Senior Data Engineer – [Pepsico Inc.](#) via [Pivotal Solutions](#) – Remote May 2020 – July 2020

- Ingested transactions from Amazon Seller Central API and loaded data in Snowflake using Python, SQL, & Airflow.
- Created data pipeline to load Walmart Online Grocery data into Snowflake.
- Created data pipeline to load Shipt out of stock records into Snowflake.

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. Automated reporting processes to save ~40 hours per week of manual excel-centric reporting.

Software Development Engineer – [The Rubicon Project](#) – Playa Vista, CA July 2014 – Sept 2018

- Designed & Developed Spark job for “First-Party” dataset persisted in Druid to sunset an older reporting product.
- 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.

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.

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.

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.

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.

Skills

- *Languages*: Python, C#, Java, SQL
- *AWS Services*: CDK, DynamoDB, Lambda, Fargate, SNS, SQS, ECS, MWAA, EC2, S3
- *Libraries & Frameworks*: numpy, cv2, pyquaternion, sqlalchemy, Spark, Airflow, .Net Core, Dagger

Education

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