

Ying Wang

me@yingw787.com | [linkedin.com/in/yingw787](https://www.linkedin.com/in/yingw787) | github.com/yingw787

Experience

Caribou (formerly MotoRefi)

Arlington, VA

Founding Data Engineer, Technical Lead (Data Engineering)

November 2020 - Present

Contract-to-hire

July 2020 - November 2020

- Redefined the entire company general ledger in SQL in one month for integration into an event-driven pipeline. In a cross-collaborative team, integrated Oracle NetSuite to internal monolith in migrating accounting operations from manual data entry contractors to automated pipeline under emergency circumstances. Turned project from red to green for direct reporting to board of directors, query runs <1s for operational purposes, ledger fully reconciled after 8 iterations and approved by finance.
- Addressed BigQuery + DataFlow internal Sev 1 outage by migrating all pipelines based on GCP DataFlow / Python to GCP Composer / BQ Standard SQL in 1 week, saved enterprise relationship with Credit Karma (80% of top-of-funnel leads), cut back core data platforms from 3 to 2 to increase IC capacity + remove future Sev1s, reduced data warehouse refresh latency from 6 hours to 10 minutes.
- Founded and help scale the data engineering team from myself to 10 (7 FTEs including EM, PM, 3 mids, 1 junior + 3 contractors) and product/engineering from 7 to 50+ FTEs, including writing and publishing job descriptions, evaluating technical screening tools, developing and validating multiple bespoke coding practicums, interviewing >25 job candidates for {PM, EM, IC} roles, onboarding new team hires, mentoring/training junior ICs, drafting a \$250k+ contract expansion engagement, etc.
- Reviewed IC design and implementation of single-touch marketing attribution, offline ads conversion platform hydration (enabling 3x marketing spend w/ only 10% increase in CAC), data system health metrics (99.99% general pipeline uptime) for CTO/VP reporting, Airflow 1.x to 2.x migration, CI for Airflow using k8s + GitOps, data quality testing/alerting with bespoke + Monte Carlo, dbt layer creation, BI platform migration from Google Data Studio to Metabase, Kafka-based data pipelines, etc.
- Securely processed and sent SPII data to TransUnion for Lost Sales Analysis of loan applications and validation of general ledger lending (G/L lending), company-wide Series C investor commitment.
- Wrote, tested, and published bespoke Python coding practicum for interviewing software generalists for onsite interviews. Popularity tied at #1 w.r.t. candidate selection for all engineering.
- Set up a custom BigQuery reverse ETL pipeline, hydrating our then-new Braze (email/SMS marketing tool) datastore with 20+ customer attributes, and flattening marketing spend on Segment.
- Developed Equifax and TransUnion credit report ingest and normalization pipeline to BigQuery, expanding industry-differentiating customer propensity modeling capabilities.
- Steered industry-differentiating corporate culture via Culture Pod during >6x headcount growth, including leading development of “Humans of Caribou”, leading employee resource groups, etc.

Sabbatical (Tooling / Infrastructure Exploration)

Arlington, VA

October 2019 - July 2020

- Created personal dotfiles to formalize development processes: <https://github.com/yingw787/dotfiles>
- Achieved beginner-level proficiency with Haskell: <https://github.com/yingw787/thehaskellbook>
- Built TinyDevCRM, a hybrid OLTP/OLAP streaming BI framework composed of PostgreSQL, Apache Parquet, and HTTP/2 Server-Sent Events: <https://docs.tinydevcrm.com>
- Started Paint Creek Software (software contracting), >\$3k revenue in 2020 (built out a logo comparison PoC using React.js + AWS Mechanical Turk)

Bytes by Ying (<https://bytes.yingw787.com>)

Author

Arlington, VA

October 2018 - Present

- 70,000 page views and 40,000 visitors from ~150 countries with ~10% return rate

Kinetica DB

Software Engineer | Developer Tools, Data Engineering

Arlington, VA

February 2018 - October 2019

- Expanded Kinetica I/O (graph-based ETL tool for highly-structured / semi-structured data) functions
 - Integrated Apache Spark as a data acceleration layer to boost data ingestion speeds by >2000x (1K records in 4 seconds → 20M records in 35 seconds)
 - Implemented engineering processes to shorten sustainable feature development lifecycle from 3 weeks to 1 day, enabling multiple \$1M - \$10M+ enterprise B2B commitments.
 - Implemented support for ESRI Shapefile ingestion supporting GIS use cases within Kinetica, part of the core product offering's market differentiation and value add.
 - Designed and implemented data-driven integration test harness in support of rapid test-driven development with support for test auto-generation, improving mean-time-to-repair by >3x.

Xometry

Software Engineer | Full-Stack

Bethesda, MD

December 2016 - February 2018

- Refactored machine learning pipeline for cost and win rate calculations, brought testing coverage to ~80%, and integrated individual testing coverage and linting reports with gated commits.
- Led a team of three SWEs to refactor cost-plus pricing algorithms, contributing to a ~80% speedup.
- Prototyped front-end expediting shipping feature, contributing to primary expedite shipping functionality generating \$10K in first week of production.
- Migrated parts of internal ERP software from Jinja / Angular / jQuery to React.js, boosting developer productivity 5x and improving page performance 3x.
- Migrated parts of landing page from WordPress to bespoke solution by developing Craft CMS templates, allowing sales/marketing to add site content without developer input while scaling to 4000+ requests per day.

Education

Duke University (Durham, NC)

B.S.E, Electrical and Computer Engineering (3.23/4.00)

2012 - 2016