

# Ying Wang

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

## Experience

---

### Caribou (formerly MotoRefi)

Founding Data Engineer

Contract-to-hire

Arlington, VA

November 2020 - Present

July 2020 - November 2020

- Leading data engineering team evolution to a data platform engineering division of ~3 engineering teams, addressing emergent data enablement, analytics engineering, and ML engineering needs.
- Leading implementation of data quality vision for the entire company, including but not limited to: iterating on data engineering platform architecture, data warehouse dimensional modeling, data observability/governance, steering marketing / contact center / upstream engineering data quality, etc.
- Founded a cross-department data interest group, including AutoML discovery & transforming / ingesting / normalizing Equifax loan performance data for ML-based credit modeling w/ 2 others, etc.
- As acting tech lead, 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 for CTO/VP reporting, Airflow 1.x to 2.x migration, CI for Airflow on GCP GKE, data quality testing/alerting, dbt analytics layer creation, BI platform migrations, etc.
- Founded and scaled the data engineering team from myself to 10 people, including writing and publishing job descriptions, evaluating technical screening tools, developing and validating multiple bespoke coding practicums, interviewing >20 job candidates for {PM, EM, IC} roles, onboarding new team hires, mentoring/training junior ICs, executing a \$250k+ contract engagement, etc.
- Addressed BigQuery internal Sev 1 outage by migrating all pipelines based on GCP DataFlow / Python to GCP Composer / BQ Standard SQL in 1 week, eliminating identical future Sev 1s, increasing IC SRE capacity by cutting back core data platforms from 3 to 2, and reducing data warehouse refresh latency from 6 hours to 10 minutes.
- 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”, moderating mental health awareness week, 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

- Product Owner, Kinetica I/O (graph-based ETL tool for highly-structured and semi-structured data)
  - 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