

Steven Dao

📍 Houston, TX | ✉ steven@stevendao.com | 🌐 github.com/sndao | 🔗 [linkedin.com/in/stevendao](https://www.linkedin.com/in/stevendao)

EXPERIENCE

Ori Technologies, LLC

December 2020 - May 2022

Founder / Creator / Software + Data Engineer

- Built my own data infrastructure to ingest financial data near real-time (<50ms (P50) latency, w/ in-house servers) for all major US markets (stocks, futures, bonds, equity/index options, indices, commodities, FX, cryptos)
 - Designed system w/ async-first/streams-based approach, where all I/O data is: 1) cached to Redis for quick lookups, and 2) persisted to SQL DB (for indexing and fast/complex JOINS, and for storing historic/bigger data)
 - Created pipelines to 1) consume raw data (e.g., time series, company data, text corpuses) via 3rd-party APIs, websockets, web-scraping (curl, headless, site-internal APIs), and 2) transform/re-ingest/store output data
 - Prototypes w/ idea that markets are members of a single, chaotic (/deterministic), complex adaptive system
 - GUI prototypes (static): <https://archive.ph/BfdHJ> (V0), <https://traderui-static.onrender.com> (V1)
 - Core thesis: CS/SWE toolkit is more effective at early identification of non-normal economic events (compared to traditional methods, e.g., stat/ML models), esp. for non-stationary data and outlier/anomalous events
- Technologies:** Python (asyncio/Flask/Django), Redis/Streams, Sqlite, Azure/AWS (VMs, DataLake, etc.)

LinkedIn

March 2017 - November 2020

Senior Software Engineer

Jobs Data Ingestion (Jobs Scraping Infrastructure), J2C (Job-to-Company ML Standardization)

- Built, maintained, and extended LinkedIn's end-to-end job posting data pipelines, optimizing the data ingestion of over 50M+ listed job postings, with scrapers (concurrently) crawling 30K+ different external job sites
- Lead data engineer on the J2C Project, a joint effort with the Company AI team. Worked cross-functionally to build and productionalize a ML model with the objective of accurately assigning a scraped job's raw company name to its correct ID in LinkedIn's data taxonomy (ex: "IBM, Inc" → "urn:li:company:1009")
- Wrote copious amounts of ETL/ELT code to clean/transform/normalize raw job data, and created task dependency flows (DAGs) to systemize the preparation of the data (top of mind, always: 'garbage in, garbage out')
- Submitted a US patent for one of the first innovations to take advantage of the partnership between LinkedIn and Microsoft, by using the Bing API to discover new jobs data sources using LinkedIn's existing metadata

Trust, Anti-Abuse Infrastructure (Anti-Scraping, Anti-Abuse Request Scoring Platform)

- Core infra engineer building and maintaining LinkedIn's high-performance, ultra low-latency (>300K requests/sec) UserRequest scoring platform, which analyzes and scores all member/guest traffic throughout all of LinkedIn's services to prevent, deter, and detect any potential abuse to LinkedIn network or its members
- Shipped super critical C++ code to improve performance across every single UserRequest Kafka event emitted from the Apache Traffic Servers (ATS) deployed in all of LinkedIn's globally co-located data centers
- Worked daily with ML engineers to integrate and productionalize various Anti-Abuse ML models (e.g., Registration scoring, UserRequest scoring), which was critical as LinkedIn's first line of defense against bad actors, e.g., malicious bots, web scraping, fake accounts, and the never-ending deluge of cyberattacks on the site

Technologies: Java, Python, Oracle DB, Kafka, Samza, HDFS, Redis, Spark, Presto, SQL, Scala
(NIHS): LinkedIn inGraphs, Espresso, Venice/Voldemort, Azkaban, DataHub, Nuage

JPMorgan Chase & Co.

July 2015 - September 2016

Software Engineer, PYXIS Team

- Core Java engineer on a distributed RESTful service that managed credit risk for large institutional clients

Software Engineering Intern, BASEL Securitization Team

June 2013 - September 2015

PERSONAL PROJECTS

🏠 **Steven's Online Garage Sale** (Python, Flask, Sqlite, Bootstrap) – steven.sale **2022 - present**

- Necessity is the mother of invention

JobMate.io (Python, Flask, MongoDB, Bootstrap) – stevendao.com/jobmateio.png **2016 - 2017**

- Aggregated companies data, then targeted best ones to work for in Silicon Valley, and their open positions
- Featured in my viral article: **I quit my job, took one-way flight to SF, and hacked the job search:**
<https://linkedin.com/pulse/how-i-quit-my-job-booked-one-way-flight-sf-hacked-search-steven-dao>

EDUCATION **University of Houston, BS in Computer Science. Extracurriculars: Lambda Phi Epsilon 2014**