

## # WORK HISTORY & EXPERIENCE

### SOLE PROPRIETOR

Okome Studio

Sep. 2007 - Current

- Provided short-term consulting as a software developer/data scientist to startups.
- Translated technical/educational materials from English to Japanese for long-term clients including the International Baccalaureate.

### RESEARCH ENGINEER

Restless Bandit, Inc.

Sep. 2014 - Jul. 2019

As the first employee in the founding team, I made the initial and sustained contribution to the design, coding, and maintenance of data infrastructure built on AWS, Python, and Linux, integrating job-candidate match scoring models by machine learning scientists; this remained the pillar of the business value proposition before a major pivot. The services (REST API/Flask) for which I was responsible include:

- Data ingestion service integrating external applicant tracking systems (e.g., Taleo & iCIMS), ingesting canonical job and candidate entities (PostgreSQL, RDS) for the customer-facing service as well as for batch processing (S3), implementing replay-via-cache for efficient use of stringent API usage quota.
- Job scheduling/monitoring service enabling non-trivial chained mapreduce/Spark workflows (mrjob, pyspark, boto3, moto, EMR, S3).
- Autoscaler/load balancer of Daxtra resume parser (SOAP) cluster, increasing the throughput demanded by highly concurrent ingestion.
- Scraping service for > 20 million public resumes for ML model training (scrapy).

Judicious use of microservice architecture and inter-process communication with Redis, RabbitMQ, SQS, and Lambda enabled smooth operation with a small team of 2 - 3 engineers; I also strove for and maintained a penchant for better quality control by participating in the DevOps best practices:

- CI/CD (CircleCI, Drone CI, Docker, Ansible) to facilitate local development and integration tests in environment closer to production.
- Active participation to documentation, testing (pytest), and code review.
- Effortless full migration from Python 2.7 to 3.7 with broad test coverage.

Before shifting heavily to a developer role, I did some modeling as well:

- Job title parser based on hidden Markov model and Viterbi algorithm; a tiny demo at <https://okomestudio.net/#/portfolio>

### DATA SCIENTIST

Bright Media Corp.

Nov. 2012 - Mar. 2014

- Increased data processing capacity by setting up Hadoop pipeline around EMR.
- Reduced similar document contamination by a MinHash deduper with CouchDB/mapreduce.
- Improved internal categorization of data by running hierarchical clustering on logs.
- Supported marketing and internal data standardization by analyzing and visualizing quantitative and textual data using PyData stack (numpy, scipy, matplotlib).

## # EDUCATION HISTORY

### Ph.D. in ASTROPHYSICS

University of California, Santa Barbara

2007