

WORK HISTORY & EXPERIENCE

SOLE PROPRIETOR Okome Studio Sep. 2007 - Current

- Short-term consulting as a software developer/data scientist
- Translation & localization of technical/educational materials (English-Japanese) for 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 preparing job/candidate/resume entities for the source-of-truth database (PostgreSQL, RDS) as well as the data lake for batch processing (S3); implementing cache/replay for efficient use of stringent API usage quota; integrating external applicant tracking systems (e.g., Taleo & iCIMS)
- Job scheduling/monitoring service enabling DAG workflows with mapreduce/Spark (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 collecting >20 million resumes for ML model training (scrapy)

Judicious use of microservice architecture and inter-process communication (Redis, RabbitMQ, SQS, Lambda) enabled smooth operation with a small team of 2 - 3 engineers with 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
- Broad test coverage, which made full migration from Python 2.7 to 3.7 effortless

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