

HIROAKI OSHIMA

San Jose, CA 95128
kifa0422@gmail.com
2532758522

<https://www.linkedin.com/in/hiroaki-oshima-31246a15b/>

Education

Bachelor of Arts in Data Science
University of California Berkeley, Berkeley, CA | August 2019

Certifications

AWS Certified Machine Learning Engineer – Associate

Skills

- Python
- SQL,
- Spark
- AWS (Glue, ECS, EKS, etc)
- terraform
- React.js
- REST API
- Docker & Kubernetes
- Data Lake
- Data Warehouse
- Databricks
- Scalability Programming
- Machine Learning
- DevOps

Languages

- English - *Native or Bilingual*
- Japanese - *Native or Bilingual*

Work History

Machine Learning Software Engineer (Volunteer)

DataKind , Remote | September 2024 - Current

- Worked as volunteer and applied big data and ML skills for social cause. Created a ML software application commissioned to NPO and counties workers in CA and FL to help their policy making
- Created an application to visualize housing affordability and renters burden across all 3000 counties. Provided the interactivity on the map to improve the user experience. Deployed to AWS cloud environment
- Created a RAG service and LLM-based chat agent which generate data visualizations and understands natural language with REST API. Made data analysis work 10x faster as a result

Data Engineer

Blue River Technology, Santa Clara, CA | July 2021 - November 2023

- Worked as big data engineer and partnered with data scientists and robotics software engineers to develop autonomous tractor See & Spray
- Built Data Lakehouse with AWS and Databricks to solve scalability problem for data replication pipelines. Shortened the replication lag from 24 hours to 1 hour while keeping the cost largely same. Enabled more timely data analysis and dashboard creation
- Architected and built ETL pipelines between multiple organizations that processes large amount of machinery data (> 1TB) daily with AWS and Spark while ensuring data security and access control for stakeholders
- Act as a consultant and regularly held office hours to troubleshoot data pipelines built by data scientists and robotics software engineer. Optimized their pipelines and unblocked multiple workflows bottle-necked by slow data pipeline execution.
- Built data lake and cataloged metadata for imagery data on AWS environment. Offloaded query from the original database (Mongodb), resulted in 40-200x faster query executions

Software Engineer Intern

FriendlyRobots.co, Berkeley, CA | November 2019 - February 2020

- Built AWS cloud infrastructure and CI/CD pipeline for self-driving vacuuming robot application to automate the build-deploy-testing process, which made the development cycle 60% faster
- Containerized ROS and simulator applications to deploy the functional tests to AWS cloud.
- Created alarms in CloudWatch service for monitoring the robot application's performance, memory and CPU usage

Projects

Databricks Data Lakehouse

- Innovated existing AWS data lake and introduces data Lakehouse architecture with Databricks/delta lake. Significantly improved replication lag while maintaining low cost. Cataloged NoSQL ML imagery data, annotation, and machine log data and helped data scientists and data analysis query data easier and faster

Choropleth Map for Housing Affordability of US Counties

- Utilized data provided by Housing and Urban Development and Created a interactive map that visually represents how many people are paying more than 50% of their income on housing costs in all 3000+ US counties per capita. The map highlights areas with particularly high housing costs compared to county's medium income.