

Chia-Yin Wu

DATA ENGINEER · SOFTWARE ENGINEER

Kaohsiung City, Taiwan

☎ (+886) 975-193693 | ✉ apexkaws911@gmail.com | 📱 apexkaws911 | 🌐 apexkaws911

Summary

Currently as data engineer of data analytics & AI team at Foxconn tech group. With 3+ years experience specializing in the data engineering, data architecture, and backend development. Super geek who loves Python, data engineering and macOS and enjoys to customize all of the development environment. Interested in devising a better problem-solving method for challenging tasks, and learning new technologies and tools if the need arises.

Skills

| | |
|----------------------------------|--|
| Programming | Python, SQL |
| Backend | FastAPI, Flask, PostgreSQL, MongoDB, Redis, Elasticsearch, RabbitMQ, Celery, Nginx, Memcached |
| Big Data/Machine Learning | Kafka, Hadoop, Spark(PySpark), Prefect, Impala, Sqoop, Hive, Kudu, Scikit-learn, Gensim, Pandas, Numpy |
| DevOps | Docker, GCP, Linux(Ubuntu, CentOS), Drone CI, Gitlab, Supervisor, Shell Scripting |
| Web crawler | Scrapy, Scrapy-splash, BeautifulSoup, Selenium |
| Languages | Chinese, English (Toeic 790) |

Experience

Foxconn Technology Group

Kaohsiung, Taiwan

DATA ENGINEER

Aug. 2019 - Now

- Designed, developed and deployed the overall asynchronous architecture in pcb printer parameter system via Python, Celery, RabbitMQ, Docker, Supervisor, PostgreSQL.
- Design and developed the overall backend analytical job architecture and RESTful API services with Docker-Compose, RabbitMQ, Flask, PostgreSQL in panel data analysis system.
- Refactored the code of Blind-Search service to improve reusability and flexibility in panel data analysis system.
- Built fully automated CI/CD pipelines via Drone CI, Gitlab for data platform project with docker-compose.
- Surveyed massive big data tech stacks, implemented and deployed CDH infrastructure including HDFS, Spark, Kafka, Impala, Kudu, Zookeeper.
- Designed the overall tech architecture and leveraged suitable tech stack in data platform project.
- Implemented core pipeline running sort algorithm with Python, MongoDB in data platform project.
- Designed and developed many RESTful APIs via FastAPI, Uvicorn, Gunicorn, also designed schema for MongoDB in data platform project.
- Actively attended internal tech sharing to give talks, talked about clean code guide, virtual environment and package management experience in Python.
- Actively involved to set up Python coding style guide reference Google style guide.
- Wrote the SQL commands for web team to integrate the API.

Pi Software Corp.

Kaohsiung, Taiwan

DATA ENGINEER

Mar. 2017 - July. 2019

- Built and developed streaming and batch data pipeline of <https://zi.media> (Alexa Taiwan top 50 site) via Python, Kafka, MongoDB, Redis, Memcached, Elasticsearch.
- Developed and deployed RESTful APIs to handle article data(daily request around 1 million times) in zi.media with Nginx, UWSGI, Flask, Redis, MongoDB.
- Designed and developed NLP services and web api including article keyword extraction(tf-idf), content based recommendations(cosine-similarity), text classification(fasttext), topic model (LDA).
- Developed and maintained more than 50 web crawlers and optimized the company's internal distributed crawler system with Kafka, Scrapy, Redis and OrangePi.
- Developed and deployed ETL system to process company and interview data with calculating ranking score for website (<https://interview.tw>), and use the RFM model to improve the popular ranking algorithm with Elasticsearch and MongoDB.
- Developed and deployed salary data analysis report with ETL job for website (<https://salary.tw>) with Pandas, MongoDB, Numpy.
- Developed and deployed finance data ETL system and stock news robot for website (<https://www.moneybar.com.tw/>).

Taiwan Hopax Chemicals. Mfg. Co., Ltd.

Kaohsiung, Taiwan

PROCUREMENT SPECIALIST

July. 2014 - Mar. 2017

- Responsible for the company's major chemical raw materials sourcing.
- Responsible for cost reduction and supplier management.
- Developed the price prediction model of styrene and acrylic acid, using R language.

Education

NCHU(National Chung Hsing University)

MASTER OF BUSINESS ADMINISTRATION

- Field of research: Game theory
- Class Rank: 6th/28.
- Master thesis - Product Line Design of Conspicuous Goods in a Distribution Channel.

*Taichung, TAIWAN**Sep. 2011 - June. 2013***National Kaohsiung University of Science and Technology**

B.A. IN INFORMATION MANAGEMENT

*Kaohsiung, TAIWAN**Sep. 2006 - June. 2011*