

Chen-Ming Yang

DATA ENGINEER · DATA SCIENTIST

Taipei, Taiwan

✉ afun@afun.tw | 🐱 afunTW | in chen-ming-yang | 📧 @afun | 📧 a_fun

"A Foolish Consistency is the Hobgoblin of Little Minds"

Summary

Familiar with the progress of building a team from scratch, and cross-functionality from engineer to the business side. Rich experience in Python development and diverse data domain from geospatial, commerce, and finance.

Looking for the challenge of either data-intensive services, low-latency data transmission, or artificial intelligence production.

Work Experience

FULL-TIME

Kronos Research

Taipei, Taiwan

SOFTWARE ENGINEER II

Sep. 2021 - Now

- Design the distributed container-free data collator on demand
 - Cross-team cooperation on low-latency process development and migration
 - Develop the custom dispatcher, monitor, alert, and recovery flow.**Value: Catch the high-volatile data makes the post-development error-tolerant possible**
- Develop data-intensive pipelines on Argo
 - Profiling the resource consumption for each task based
 - Dynamically dispatch the task by the trade-off between accuracy and resource spec**Value: Report the statistics of considerable data on time**
- Initialize the market maker analysis, dashboard, and bot reporting

WeMo Scooter Corp.

Taipei, Taiwan

DATA SCIENTIST

Sep. 2019 - Sep. 2021

- Redesign the data warehouse
 - Leverage the columnar storage and distributed computing by BigQuery
 - Redesign with the data warehouse schema and SSOT**Value: Regularize the principle of revising the data warehouse**
- Refactor the Airflow pipeline to reduce the landing time
 - Design the template of the pipeline and auto-generate by config
 - Decouple the pipeline logic and business into DAGs and SQL files for better cooperation**Value: Reduce 90% pipeline lead time and less error**
- Serve the BI tool Metabase for cross-department data ingestion
 - Define and build the company-wise metrics dashboard
 - Setting Google SSO and permission control**Value: Centralized Dashboard and Report**
- Maintain the ELK Cluster by helm
- Geospatial data analysis on operation metrics and resource allocation
- Marketing analysis on user behavior

Academia Sinica, Institute of Information Science

Taipei, Taiwan

RESEARCH ASSISTANT & RDSS (COMPULSORY MILITARY SERVICE)

Aug. 2016 - Aug. 2019

- Build a beetle trajectory tracking pipeline for biology researchers to replace the human labeling process.
 - Apply tracking-by-detection methodology on multi-object tracking and enhanced the performance with Kalman filter and the Hungarian matching algorithm
 - Using OpenCV on pre-data cleaning and generation**Value: Phased success on tracking without light interference and high-resolution video**
- Cooperate the paper "Artificial intelligence reveals environmental constraints on color diversity in insects", accepted by the nature communication journal
 - Using PyQt5 develop the GUI tools to generate 3,000 high-resolution moth segmentation data
 - Experiment the Tiramisu and RCNN model on 30,000 raw images to generate the segmentation data**Value: Data collection and pre-processing the own dataset, narrow down the problem to model tuning**
- Recording the Python crawling tutorial for AI Academic schools lecture
- Maintain the laboratory 10+ Linux servers

Side Projects

IPI

Remote

IoT TELEMETRY DATA ANALYSIS PIPELINE

Sep. 2021 - Now

- PoC of automatically archive items to S3 using DynamoDB TTL with Lambda and Firehose
- PoC of building data analysis pipeline with Glue and Athena

Taiwan-Ebook-Lover

Remote

SEARCH THE E-BOOKS ON DIFFERENT PLATFORMS

Sep. 2021 - Now

- Using Prometheus to export metrics and monitoring on Grafana on 1 virtual machine

Education

National Chi Nan University (NCNU)

Nantou, Taiwan

BS & MASTER'S DEGREE IN COMPUTER SCIENCE AND INFORMATION ENGINEERING

Sep. 2011 - Jul. 2016

- Paper "Integrated Sleep Scheduling for Multiple Relays and UEs in LTE-TDD"
- The Director of Student Association of CSIE Department.

Memorial University of Newfoundland (MUN)

St. John, NL, Canada

BACHELOR EXCHANGE STUDENT IN COMPUTER SCIENCE

Sep. 2014 - Apr. 2015

- Two semesters with GPA 3.5, 3.6 respectively.

Skills & Interested

DEVELOPMENT

Experienced	Argo , Develop the pipeline rely on EKS and AWS ecosystem	Frequently
Experienced	Airflow , Develop the pipeline rely on Composer and GCP ecosystem	Frequently
Experienced	BigQuery , As data warehouse platform and the source of BI tools	Frequently
Known	Docker , Used to deploy a service on a single machine	Frequently

DOMAIN

Interested	Data Engineering , Data Governance on business insight and build the data pipeline on distributed system	Frequently
Known	Data Scientist , Using Keras on classification and tracking	Used
Known	Computer Vision , Using OpenCV for image and video pre-processing	Used