Chen-Ming **Yang**

DATA ENGINEER · DATA SCIENTIST

Taipei, Taiwan

✉ afun@afun.tw | 🐙 afunTW | in chen-ming-yang | M @afun | ✈ a_fun

"A Foolish Consistency is the Hobgoblin of Little Minds"

Summary

Looking for an environment with open-sharing and enthusiasm. Team management is comfortable on scrum. Willing to challenge the data-intensive system both on-premise and cloud-native. Available to take on the data pipeline and data warehouse design. But still open to the position related to data science or software engineering.

I am the person who will keep pushing myself on learning not only the tech about jobs but any news. Believing in the concepts of connecting the dots, there is no useless information.

Work Experience

FULL-TIME

WeMo Scooter Corp. *Taipei, Taiwan*

DATA SCIENTIST *Sep. 2019 - Now*

- Redesign and Refactoring the data warehouse alone
  - Considering the BigQuery leverage the columnar store format, redesign the warehouse in denormalized makes better performance
  - Controlling the SSOT by well-document and restrict following the design principle
  Value: Self-formulate an SOP to follow and documentation
- Refactor the Airflow pipeline and improve workflow alone
  - Auto-generated the DAGs by self-defined jinja template and pre-commit trigger
  - Reduce errors by setting linter and code-formatter on pre-commit hook
  - 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
  - Provide near-real-time company KPI metrics Dashboard, data updated in 30 mins
  - 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 researcher to replace 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 colour 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 Tirumisu 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

PART-TIME

KYC Right *Remote*

SPECIALLY APPOINTED RESEARCH ASSISTANT *Nov. 2019 - Apr. 2020*

- PoC of community detection on Neo4j

Cupoy *Remote*

LECTURER *Apr. 2019 - Now*

- Cooperate and publish two courses on Python Crawling and Computer Vision

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

Familiar	**Airflow,** Integrated with GCP Composer to build the pipeline	*Frequently*
Familiar	**BigQuery,** As data warehouse platform and the source of BI tools	*Frequently*
Known	**Docker,** Used to deploy a service on a single machine	*Frequently*
Known	**Scrapy,** As side-project crawling framework	*Used*

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*