

## Summary

---

4-5 years experience with data engineer. (Distributed Queue System, Database, Web Crawling, RESTful API, ETL, Docker, CICD, GCP, Airflow ...etc.)

1-2 years experience with data science. (Data Analysis, Machine Learning and Deep Learning)

Github project get 1,600 stars.

## Work Experience

---

### 17 LIVE - Senior Data Engineer

May. 2021 - Now

- The main developer for refactor ETL. Creating a airflow project by **Cloud Composer** to transfer ETL tools from digdag to airflow and transfer ETL develop method from shell script to python. ( More than 20 large pipelines. )
- Maintenance **BigQuery** more than 300 tables and more than 10 TB data.
- Create pipelines from backend team, multiple mysql and mongodb, and appsflyer to bigquery.
- Create a good development culture, including the introduction of CICD, dev-stage-master, unit tests and test coverage.
- The finalists of **Break the Norm** awards on 2021-Q3 and 2021-Q4.
- Assist in interview more than 10 new data engineer.
- Mentor junior data engineers to be more effective individual contributors.

### SinoPac Holdings - Software Engineer

Nov. 2019 - May. 2021

- Develop **python, C#** api for stock/option/future trading.
- Setup **CI/CD** with github actions and gitlab-ci to publish package and unit testing.
- Deploy a test system to simulate transactions on **docker swarm**.
- Collecting distributed system log by **elk, grafana** and **prometheus**, 13 GB log data/day.
- System monitor of api latency, worker alive and user request count by **telegram bot**.
- Develop **transaction-by-trade and odd lot trading** API.

### Openup Speaker ( FinMind )

2019-12-01

### Tripresso - Data Engineer

Oct. 2018 - Oct. 2019

- Analysis travel data and build a machine learning model. Estimating increase **3%** orders (revenue).
- Maintain and develop an ETL **distributed queuing system** with 20 machines.
- Optimize the ETL system reduced more than **50%** execution time.
- Develop new product crawler let product volume increase **1.5%**.
- Making **BI charts** provide for other departments.

## Publications

---

- Published a book related to data engineering. Python 大數據專案 X 工程 X 產品資料工程師的升級攻略. On major platforms, like tenlong, books, momoshop ...etc, the weekly sales are the highest TOP 1, and the monthly sales are the highest TOP 2.

## Projects

---

### FinMind python project

1,600 Stars On Github. Sep. 2017 - Now

- 3 team member.
- Create a payment flow by **ECPay**, and getting some order from user.
- More than **1,500 people registered**.
- Open source of financial data, more than 50 dataset, more than **1 billion, 200GB**.
- **10 million** data per day.

- **Deploy restful api, backend, web, crawler, rabbitmq, mysql, dolphindb, redis and loading balance by docker swarm and traefik.**
- Transfer dataset from mysql to dolphindb, **reduce 80% (150GB)** data size.
- Using scheduler to automatic update data by **airflow, rabbitmq and celery** ( 8 cloud machines ).
- Automated deployment web, backend and api by **CI/CD**.
- Using **vue, python and django** to develop web pages. ( <https://finmindtrade.com/> ).
- The API has more than **100,000** requests per day. Upper bound of api is **8,000/minute** request by ApacheBench.
- Design data pipeline for crawler, backend and analysis by **airflow**.

**Bosch Production Line Performance ( Kaggle )**

**Post-competition analysis, top 6% rank.**

**Grupo Bimbo Inventory Demand ( Kaggle )**

**Post-competition analysis, top 8% rank.**

**Instacart Market Basket Analysis ( Kaggle )**

**Final top 25% rank.**

## Education

---

**National Dong Hwa University (GPA : 3.87)**

Master of Science, Sep. 2017.

Major : Mathematics and Statistics.

**Tamkang University**

Bachelor of Science, Sep. 2015.

Major : Mathematics.