林子軒 Lin Sam · ✉ samlin266118@gmail.com · 📞 0972724528 · **in** sam lin · ⬡ linsamtw ·

Summary

3-4 years experience with data engineer. (distributed queue system, database, web crawling, airflow, CI/CD, cloud)
1-2 years experience with data science. (data analysis, machine learning and deep learning)
Github project get 1K stars.

Work Experience

SinoPac Holdings - Software Engineer(Python)   Nov. 2019 - Now

- Develop **python, C#** api for stock/option/future trading. Setup **CI/CD** with GitHub actions.

- Deploy test system for simulating trading by **docker swarm**.

- Collecting distributed system log by **elk**, **grafana** and **prometheus**, 13 GB log data/day.

- System monitor by **telegram bot**.

- Develop **transaction-by-trade and odd lot trading** API.

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.

Projects

FinMind python package   **1.3K stars on github.**

- Open source of financial data, more than 50 dataset, more than **1 billion, 200GB**.

- **10 million, 1gb, streaming** data per day.

- Transfer dataset from mysql to dolphindb, **reduce 80% (150GB)** data size.

- Automatic update data by **airflow, rabbitmq and celery** (8 cloud machines).

- Automated deployment web, backend and api by **CI/CD**.

- Using **vue, python** to develop web pages. (https://finmindtrade.com/).

- Api request **20,000 ∼ 30,000** times/day. Up to 70,000 times.

- Design data pipeline for crawler, backend and analysis by **airflow**.

TaiwanTrainVerificationCode2text python package   **70 stars on github.**

- The verification code consists of 26 English and 10 number, which may be 5 or 6 yards.

- Simulate generating 100,000 images and building CNN models.

Bosch Production Line Performance (Kaggle)   **Post-competition analysis, top 6% rank.**

- Highly imbalance data, ratio is 1000 : 1, 10 GB dataset size. And the data is 50% missing value, more than 4000 variables.

Grupo Bimbo Inventory Demand (Kaggle)   **Post-competition analysis, top 8% rank.**

- Time series problem, 8 millions data size. Building models predict inventory demand after 2 weeks.

Instacart Market Basket Analysis (Kaggle)   **Final top 25% rank.**

- Which products will an instacart consumer purchase again?

Skills

- Experience with **distributed queue system**, rebbitmq and celery. (Linode Cloud)

- Experience with **database**. (MySQL, dolphindb, Redis)

- Experience with **restful api**. (Fastapi).

- Experience with **data pipeline**. (airflow).

- Experience with **virtualisation and containerisation**. (Docker Swarm) Using it manage services of crawler, backend api, finmind api, database and web.

- Experience with **traefik** to manage router and DNS. Include DNS of backend api, finmind api and web. And auto register SSL let's encrypt.

- Experience with **CI/CD** tools, gitlab-CI, github-action. Create automated tests and automated deploy.

- Experience with **monitoring systems** (Prometheus, Grafana, elk, telegram bot).

- Experience with **machine learning** by Python (**xgboost-gpu**) and R (**xgboost, svm, random forest, knn**).

- Experience with **deep learning** by Python (**kears-CNN, LSTM, GRU**).

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.