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 1.5K stars.

Work Experience

17 LIVE - Data Engineer   Nov. 2019 - Now

- Refactor ETL, create a airflow project by **Cloud Composer** to transfer ETL tools from digdag to airflow and transfer ETL develop method from shell script to python.

- Maintenance **BigQuery** more than 100 tables.

- Create pipelines from **mysql** and **mongo** to bigquery.

- Create a good development culture, including the introduction of CICD, dev-stage-master and unit tests.

- Assist in interview new data engineer.

SinoPac Holdings - Software Engineer   Nov. 2019 - Now

- 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.

Projects

FinMind python project   **1.5K stars on github.** Sep. 2017 - Now

- 3 - 5 team member.

- Create a payment flow by **ECPay**, and getting some order from user.

- More than **1,000 people registered**.

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

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

- **Deploy 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/).

- API has about **70,000** 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)   **Tamkang University**
Master of Science, Sep. 2017.   Bachelor of Science, Sep. 2015.
Major : Mathematics and Statistics.   Major : Mathematics.