

Peter Lee

Taipei, Taiwan. Email: peterlee0127@gmail.com

Web: resume.peterlee.app Github: [peterlee0127](https://github.com/peterlee0127)

As a seasoned data engineer and DevOps engineer, I have a strong background in data infrastructure, data science, and backend development. I have practical experience working with public clouds (GCP and AWS), and have a track record of successfully designing and deploying pipelines in on-premises environments. I am skilled at maximizing system performance and am always eager to learn about new technologies.

SKILLS

Python, Node.js, Kubernetes, Docker, CI/CD, Airflow, GCP, AWS, Git, RKE, Kafka, Spark.

WORK EXPERIENCE

Data Engineer & DevOps Engineer, Snapask; Taipei, Taiwan. 2021 Dec - Present

- Designed new data infrastructure on AWS, resulting in a 2x reduction in cost and a 3x increase in query speed. For example, I migrated Airflow 1.x to Airflow 2.0 with AWS MWAA and optimized the allocation of machines to reduce cost on the cloud.
- Designed a log service for collecting custom log events, improving data collection reliability and reducing reliance on third-party analytics platforms.
- Maintained and developed terraform modules for AWS infrastructure. (IaC)

Senior Data Engineer, Vpon; Taipei, Taiwan. 2018 Oct - 2021 Jul

- Designed the backend system for data collection for the Data SDK, reducing data update time from one day to one hour and achieving a 40% cost reduction for ETL machines.
- Introduced the Kubernetes Engine to automate the process of deployment and monitoring, significantly reducing the cost by 75% and the personnel expenses by 20 hours. Optimize the pipeline of data update to directly generate files and reports and achieve a personnel expense reduction from 2 weeks to less than 1 hours.
- Replaced previous VM/EMR environment for ETL with Apache Airflow and Kubernetes to make the ETL pipeline more flexible and cost-effective. Also designed new tools to show the cost (size) of every query.
- Designed a complete CI/CD pipeline including building Docker images, deploying to Container Registry, deploying images on the Kubernetes cluster, and deploying images on application environments after conducting unit tests.

PDIS (Public Digital Innovation Space), Taipei, Taiwan. 2017 Sep - 2022 Aug

- PDIS is a government organization in Taiwan. Executive Yuan - Ministry of Digital - Audrey Tang's 唐鳳 Office.

RECENT PERSONAL PROJECTS

- Two local Ubuntu servers for building a Kubernetes 1.24 cluster using Rancher Kubernetes Engine. Airflow and a Line bot notification service with Flux CD.
- Cloudflare to restrict unauthorized web access, Caddy as a load balancer, and ingress-nginx to access internal Kubernetes services.
- Gitlab was deployed locally, along with gitlab-runner, which can build a Docker image and push it to a my private image registry.

PUBLICATION

- Implementation of Lambda Architecture: A Restaurant Recommender System over Apache Mesos. 2017 IEEE 31st International Conference: Advanced Information Networking and Applications (AINA). A Lambda Architecture at DC/OS, using Spark, Spark Streaming, Kafka, Hadoop HDFS.
- Stock market analysis from Twitter and news based on streaming big data infrastructure. 2017 IEEE 8th International Conference: Awareness Science and Technology (iCAST). Real-time trend analysis by twitter, using Spark, Spark Streaming, Kafka, Classification Data and Visualize on Web.
- Real-time Trend Analysis of Streaming Twitter and News Based on Big Data Infrastructure. 電子情報通信学会-講演論文. 信学技報, vol. 117, no. 184, SC2017-13, pp. 1-6, 2017年8月. SC2017-13 2017-08-18 (SWIM, SC).

EDUCATION

- University Of Aizu. Fukushima, Japan. Master Degree of Computer Science. 2016-2017
- Tamkang University. Taipei, Taiwan. Master Degree of Computer Science. 2014-2017
- Tamkang University. Taipei, Taiwan. Bachelor of Engineering Computer Science. 2011-2014