Peter Lee

My research interest includes Data Infa, Data Science, Backend and IoT. I focus on big data and data infra since 2015. Contributor of TensorFlow, Rocket.Chat and Nvidia Jetson community and other open-source projects.

in ○

Taipei, Taiwan
peterlee0127@gmail.com   Github Website

Experience

Vpon Taiwan - Senior Data Engineer (Architect),

Oct 2018 - Present

- In data team, I am responsible for building up data infrastructures and designing pipelines. I have huge passion for learning new knowledge and sharing with members. In addition, I am expert in optimizing the performance of a system by introducing new technology. Like containers and DevOps.
- Data Collection and Exchange:

  1. Design different algorithms for automatic data exchange with business partners.
  2. Design the backend system for data collection of Data SDK. The system reduces the time to update data from a day to an hour and achieves a cost reduction of 40% for ETL machines.
  3. Deploy Elastic Stack to validate and visualize data from each new client in real time.

- Data Architecture

  1. Introduce the Kubernetes Engine to automate the process of deployment and monitoring, significantly reducing the cost by 75% and the personnel expenses by 20 hours.
  2. Provide advices about architecture design as well as deployment of Kubernetes with Istio, Cloud Function, Airflow, and CloudRun.
  3. Develop internal bot notification system to provide instant error notifications of ETL states and Health Check Monitor.

- Data Application

  1. Build up a platform for highly customized tagging services. The platform gives each individual ID different kinds of tags such as personal interests, language, age, and gender and reports the tagging rate of each tag.
  2. 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.
  3. Replace previous VM/EMR environment for ETL with Apache Airflow and Kubernetes CronJob to make ETL pipeline more flexible and cost-effective.

- Continuous Integration and Deployment

  1. Design the complete CI/CD pipeline including building Docker images, deploying to Container Registry, 2-hours deploying images on the Kubernetes cluster, and deploying images on application environment after conducting unit tests.
  2. Build up Jenkins on the GKE and design related architectures to automatically run ETL pipelines each day.

- Data Migration:
  1. Migrate existing projects and the ETL pipeline from AWS to GCP. Transfer data from Hive and S3 to GCP. Set up a Gitlab server in GCP. Build up ETL pipelines for effective and efficient data migration, which is very important for team members to access the latest data on GCP.

Public Digital Innovation Space (PDIS) 唐鳳政委辦公室- Software Engineer,

Sep 2017 - Aug 2018

PDIS is a goveronment origanization in Taiwan, it also has another name: Executive Yuan - Ministry of

Digital - Audrey Tang's Office. We use new technology to improve public participation and open

government.

1: As a software developer: I use Line BOT API to connect our internal systems to make an easy-to-use app to help our colleagues saving their time. I also designed API interfaces to allow other internal systems to hook up with our bot system.Up to now, it already bridged across our meeting reserve system and electronic bulletin board system.

2: About productive and quality: I designed a software that can display real-time subtitle during streaming. To meet the schedule, I took only one week to make it from idea to prototype. This system now works perfectly in all of our video conferences.

3: As a quick learner and contributor: I involved in many open source projects. Like Sandstorm(A open source private cloud with the container), Rocket.Chat (A open source chat app which like Slack). We have fixed many issues and add lots of features, and giving back to the community.

4: Eager to learn everything: when I have free time, I like to think about how to make system more efficient. I still find more possibility to improve the current system. Like the subtitle display system, I would like to let some part of the system can automatically work. Let the computer can label the people with their name in real-time and display the information on the subtitle. On the road of development, I fixed various issues and bugs of Tensorflow. NVIDIA give me an award as the outstanding community developer in this April.

Until now, I still help PDIS to maintain our cloud services. Like our internal systems, website and translator.
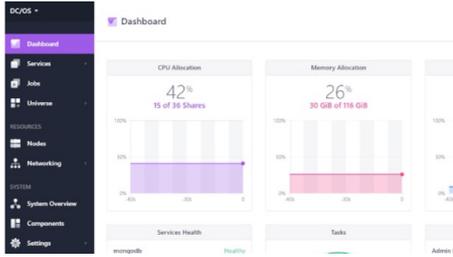https://pdis.nat.gov.tw/en/   https://github.com/pdis

Nexdoor (91 App) - iOS Engineer, July 2013 - Sep 2014

Nexdoor is a famous App company in 2013, we build iOS/Android/Web for the enterprise. I was an iOS Engineer Intern when I was an undergraduate student. I developed some interesting project, like Geofenece App, LBS Service framework and customize UI. In 2016, most of employees join new company 91 app.
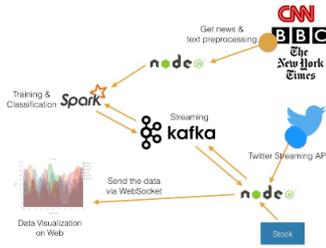
91APP

Publication

Implementation of Lambda Architecture: A Restaurant Recommender System over Apache Mesos

Advanced Information Networking and Applications (AINA) 2017 IEEE 31st International Conference
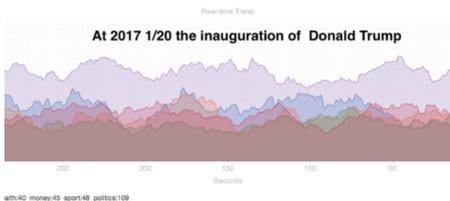A Lambda Architecture at DC/OS, using Spark, Spark Streaming, Kafka, Hadoop HDFS. Full Text

Stock market analysis from Twitter and news based on streaming big data infrastructure

Awareness Science and Technology (iCAST) 2017 IEEE 8th International Conference
Real-time trend analysis by twitter, using Spark, Spark Streaming, Kafka, Classification Data and Visualize on Web. Full Text

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) Full Text

Selected Awards

Nvidia Jetson developer community contributor

Nvidia Jetson developer community contributor
Porting TensorFlow for Nvidia Jetson.

TensorFlow Contributor

Porting TensorFlow to Nvidia Jetson. Repository tensorflow-nvJetson and tensorflow #26985 #20025, #19075, #17394.

Rocket.Chat Contributor

Update and patch Rocket.Chat for Sandstorm. Due to there is no official build since 2017. I started to fix issues and help new official build.

Selected Skills

Data

Kubernetes with Istio
Elasticsearch
Spark
Airflow
Kafka

Backend/Web

Node.js
Python
MongoDB
MySQL

General

DevOps
GCP/AWS
Gitlab CI/CD
Jenkins
Docker

Education

University Of Aizu, Fukushima, Japan.

2016 March - 2017 March Master Degree of Computer Science. Double Degree Program.

Tamkang University, Taipei, Taiwan.

2015 - 2017 Master Degree of Computer Science. Cloud Computing Lab.
2011 - 2014 Bachelor of Engineering (B.E.) Computer Science. Embedded System Lab.