

# Pham Hoang Tuan

## Senior Data Engineer

Last update: June 26, 2023

The online version is available at  
<https://phamhoangtuan.github.io/cv>

### Residence

Vietnam

### LinkedIn

<https://www.linkedin.com/in/phamhoangtuan>

### Email

[phamhoangtuan.it@gmail.com](mailto:phamhoangtuan.it@gmail.com)

### Languages

Vietnamese, English

### Phone

+84 909 770 891

## Professional Experience

### Aitomatic, Inc., Vietnam

Product Engineer

Mar 2021 - Present

- Developed product based on Large Language Model (LLM) to answer questions based on documents and many agents' reasoning process.
- Used Terraform and Terragrunt to create a Kubernetes cluster that is scalable and cost saving by automatically use suitable spot instances based on cluster usage.
- Developed service API with FastAPI to serve requests from UI to build ML models and infer data by that saved model.
- Developed framework libraries with OOP philosophy to support product development.
- Developed weekly/daily ETL pipeline to ingest raw coldchain data from customers, transform to parquet format and identify anomaly stores based on Data Scientist models with some positive feedbacks.
- Developed AWS Lambda jobs based on docker images to replace old ETL pipeline and reduce 90% cost of running infrastructure in AWS.

Python AWS Kubernetes Terraform FastAPI LLM

### SP Digital, Singapore

Senior Data Engineer

Feb 2019 - Feb 2021

- Upgraded data infrastructure to newer version in production with no downtime in 3 months.
- Took full ownership of a new system to collect, transform and feed useful information to a data science model, which can predict when the electric meter need to be replaced and reduce false alarm for operation team. As a result, time to train model is decreased by 30% and the accuracy is increased by 45%.
- Created coding standards for SQL that helps team members easier to read and review. Moreover, a platform to manage SQL effectively is suggested as an effort to improve code performance and readability.
- Developed python scripts to monitor data automatically and reduce MTTR (Mean Time To Repair) of data lake.
- Developed hourly jobs in Kubernetes to ingest electric price data as an input for the next 48-hours price prediction model with more than 80% accuracy.

Python Java Talend Hadoop Hive Spark HBase Kubernetes

### Honestbee, Singapore

Data Engineer

Sep 2016 - Jan 2019

- Developed a data pipeline to collect users' orders information, users' behaviors and other third parties data, process them and store in Datawarehouse. As a result, BI team always have clean and accuracy data to show in their reports.
- Optimized queries to increase overall performance by 25% for product ranking and recommendation system.
- Maintained legacy data infrastructure and ensure its high availability for other dependent data services.

Python AWS Airflow Postgres Docker Mesos Looker

Review: [Nick Young, Head of Data](#)

## Crayon Data, Singapore

Data Scientist Intern

Aug 2015 - Mar 2016

- Developed web crawler with Python libraries to collect data automatically from e-commerce website. As a result, data team can have many raw data to process and train data models for product improvement.
- Implemented a novel extractor to retrieve websites' information based on HTML structure in under 5 seconds per website.

Python Java MySQL

Review: [Vivek Rensol, Product Director](#)

## Certifications

[Professional Scrum Developer I - May 2020](#)

## Education

Master degree in Knowledge Engineering, NUS, Singapore [2015 - 2016]

Bachelor degree in Software Engineering, FPT University, Vietnam [2009 - 2013]