# Van-Duyet LE  |  Resume

Ho Chi Minh City | Year of birth: 1995 | e: me@duyet.net
https://duyet.net  |  https://github.com/duyet  |  https://linkedin.com/in/duyet

I am a **Data Engineer** with 4+ years of experience. Currently, helping Fossil Inc with a Terabyte-scale Data Platform based on Kubernetes.
Working on Data Engineering *(Data Pipeline, DWH, Data Processing and Big Data based on AWS and GCP)*, Data Analytics (*Data Studio*), Machine Learning especially NLP *(Text Classification, Word Embedding, ...)*.

# Experiences

OCTOBER 2018 - CURRENT

**Data Engineer –** *Fossil Group Inc*

Lead the Data Platform team

- Build and maintain the Data Platform (based on Kubernetes, Apache Kafka, Apache Airflow) with ~10TB/month processed data in near-realtime, transforming data into Data Warehouse, visualization dashboard with Google Data Studio. Technologies: AWS (S3, Redshift, SQS), GCP (Firebase, BigQuery, Data Studio), Kubernetes, Apache Airflow, Apache Kafka.
- Design and build the internal tools for Data Platform (Management Dashboard, Data Catalog, Parser, ...) using Typescript, React, Python.

JUNE 2017 - OCTOBER 2018

**Senior Developer –** *FPT Software*

Working on Recommendation System and Data Lake project

- Build a Data Pipeline to process 2TB of data per day in AWS.
    - Technology: AWS (EC2, S3, EMR, Lambda), ElasticSearch, Apache Spark.
    - CI/CD for automatic deployment (CodePipeline, CodeBuild, Docker, ECS).
- Data Pipeline on Microsoft Azure: build Databricks workflows by using Data Factory and Azure Functions, ETL batch processing the data to Data Lake with 1TB+/day.

SEPTEMBER 2016 - JUNE 2017

**Data Engineer –** *John von Neumann Institute - Vietnam National University*

Main contribution: Team member of two projects:

- Headhunter Company
    - Cleaning, building data pipeline, and ML model for Candidate matching with JDs (Python, ElasticSearch, Cloudera, Word2vec).
    - Visualization dashboard: hiring competitiveness, ideal candidates, etc (RShiny, D3js, Python).
    - Project: Skill2vec - Relevant Skills from Job Description: It transforms skills to new vector space, which has the characteristics of calculation and presents skills relationships.
- Advertising Network Company
    - Predict the Demographics of users from tracking data (Movie ads network) using Apache Spark, Hortonworks.

OCTOBER 2014 - SEPTEMBER 2016

**Full-stack Developer** – *Information System Lab, UIT*

Main contribution: Working on many projects such as:

- Self-serve ad platform for marketing, Real-time Bidding server using Nodejs.
- Web crawler
- Research and Develop websites with Node.js/Javascript. Real-time dashboard with D3.js, Node.js, Kafka, and Socket.io.

## Skills

**Web:** Javascript & Node.js (D3js, Reactjs), Python (Flask), SQL/NoSQL.

**Data Engineering:** AWS (EMR, S3, Redshift), MS Azure (Azure Function, Data Factory, Databricks), GCP (Firebase, BigQuery, Data Studio), Python, Javascript, Apache Spark, Apache Airflow, Kubernetes, Helm/FluxCD.

## Education

JULY 2018 - CURRENT

Master of Computer Science | **University of Information Technology - HCMC**

JUNE 2013 - 2017

Bachelor of Information System | **University of Information Technology - HCMC**
Thesis: *Build a network of career skills and support an optimal job search.*

## Awards

**Third Prize "Entropy 2016"** – *John von Neumann Institute (VNU).*

The first competition of Data Science in Vietnam ⬈

**Third Prize "FOSS 2015"** – *Hanoi University of Business and Technology.*
Free and open-source software (FOSS) competition.

**Final round "Creative IT Camp 2013"** – *University of Information Technology (VNU).*
Innovation and Programming hackathon competition.

## References

MEng. **An DANG** – *Tech Lead of Glandore System Project*
John von Neumann Institute (VNU) | email: an.dang@jvn.edu.vn

Ph.D. **Thinh PHAM**
John von Neumann Institute (VNU) | email: thinh.pham@jvn.edu.vn