

# Van-Duyet LE | Resume

Ho Chi Minh City | Year of birth: 1995 | e: [me@duyet.net](mailto:me@duyet.net)

<https://duyet.net> | <https://github.com/duyet> | <https://linkedin.com/in/duyet> | <https://t.me/duyet>

I'm a **Data Engineer** with 4+ years of experience. Currently, I'm helping Fossil Group with a Terabyte-scale Data Platform based on Kubernetes.

## Experiences

OCTOBER 2018 - CURRENT

**Data Engineer** - *Fossil Group Inc*

Lead the Data Platform Team of 5 members

- Maintained, saved a couple of thousand dollars each month by refactoring the Data Platform (Kubernetes-based, AWS, Kafka, Airflow, Spark) with ~10TB/month processed data in near-real time. Transforming data into Data Warehouses, visualization dashboard with Google Data Studio.
- Develop the services for Data Monitoring, Data Catalog, and Self-service Analytics.
- Technologies: Python, Typescript, AWS (S3, Redshift, SQS), GCP (Firebase, BigQuery, Data Studio), Kubernetes, Helm, Apache Airflow, Apache Spark, Apache Kafka.

JUNE 2017 - OCTOBER 2018

**Senior Software Engineer** - *FPT Software*

Working on the Recommendation System (RS) and Data Lake project

- RS project: develop a Data Pipeline to process 2TB/day of data by AWS services (EC2, S3, EMR, Lambda, Spark), Using Elasticsearch to serving RS models with CI/CD by CodePipeline and CodeBuild.
- Data Lake project: develop a Data Pipeline on MS 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
  - Develop the data pipeline, clean data, and implement Machine Learning models for Candidate matching with JDs (Python, Elasticsearch, Cloudera, Word2vec).
  - Visualization: hiring competitiveness, ideal candidates, etc., using RShiny, D3js, and 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 users' Demographics from tracking data (Movie ads network) using Apache Spark on self-managed Hortonworks Data Platform.

OCTOBER 2014 - SEPTEMBER 2016

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

Working on many projects: contribute to Ad Platform project (RTB core using Nodejs), Web Crawler for Opinion Mining, Research and Develop websites using Nodejs, D3js, Kafka, Socket, etc.

## Skills

**Frontend + Backend Engineering:** Javascript, TypeScript, Reactjs (basic), Python, SQL/NoSQL.

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

**DataOps:** Kubernetes, Apache Airflow, Helm/FluxCD, Prometheus/Grafana.

## 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**

John von Neumann Institute (VNU) | e: [an.dang@jvn.edu.vn](mailto:an.dang@jvn.edu.vn)

Ph.D. **Thinh PHAM**

John von Neumann Institute (VNU) | e: [thinh.pham@jvn.edu.vn](mailto:thinh.pham@jvn.edu.vn)