



Resume

Van-Duyet LE

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

I am a **Data Engineer** with 3+ years of experience. Currently, helping Fossil Inc with a Terrabyte-scale Data Platform on Kubernetes.

Working on Data Engineering (*Data Pipeline, DWH, Data Processing and Big Data based on AWS and GCP*), Data Visualization (*Redash, Data Studio*), Machine Learning especially NLP (*Sentiment Analysis, Text Classification, Word Embedding, ...*).

Experiences

OCTOBER 2018 - CURRENT

Data Engineer - *Fossil Group Inc*

Working on Data Pipeline and Data Warehouse project

- Build Data Pipeline (based on Kubernetes, Apache Airflow) with ~10TB processed data in near-realtime per month, transform data to Data Warehouse based on AWS and GCP. Technologies: AWS (S3, EMR, Lambda), GCP (Firebase, Big Query, Data Studio), Kubernetes, Airflow.
- Design and build the internal tools for Data Pipeline (Management Dashboard, Configuration Dashboard, ...) using Typescript, React, Python.
- Data Visualization (Redash, Data Studio)

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, PySpark.
 - 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, Work2vec).
 - Visualization dashboard: hiring competitiveness, ideal candidates, etc (by 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 (<https://arxiv.org/pdf/1707.09751.pdf>).
- Advertising Network Company

- Predict the Demographics of users from tracking data (Movie ads network).
- Technologies: Apache Spark (PySpark), 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 by Nodejs.
- Web Spider, 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 (EC2, ECS, EMR, S3), MS Azure (Azure Function, Data Factory, Databricks), Python, Javascript, Apache Spark, Apache Airflow, Kubernetes, Helm, FluxCD, Data Studio.

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