

Nguyen Hoang Long

Senior Data Engineer



longcnttbkhn.github.io



hoanglong180695@gmail.com



0963207903

EXPERIENCE

TECHNICAL MANAGER | DARENFT

Mar 2022 – Current (9 months) | Hanoi, Vietnam

- Business Intelligence for Blockchain and Crypto data. (3 members)
 - Collected all BSC and ETH blockchain data and store in Data Lake base on **HDFS**.
 - Built Data Warehouse using **DBT**, **Spark**, **Hive** and **Airflow** for data analysis.

SENIOR DATA ENGINEER | CENGROUP

Jan 2020 – Mar 2022 (2 years) | Hanoi, Vietnam

- Built AI-power Data Crawler allow crawling data from any website without predefined data structure. (3 members)
 - Trained the page classifier model using **Fasttext** to determine the target page
 - Used **Web2text** to remove Boilerplate from the page
 - Built Information Extraction Model by using **Weak Supervised Learning**.
- Lead development team to build a new Real Estate Online Appraisal System. (4 members)
- Research the Customer Data Platform. (2 members)

DATA ENGINEER | CENGROUP

Oct 2018 – Dec 2019 (1 year) | Hanoi, Vietnam

- Used Big Data processing technology including **Spark**, **Hadoop** to build ETL pipeline for real estate listing data. (2 members)
- Built Address Normalization Engine uses the **Weak Supervised Learning** and **Elasticsearch** allows to handle strange and incomplete address strings. (1 member)
- Built Google Map Cache System uses **Leaflet** and Tile Coordinates System, thereby reduces the cost of using Google Map API. (1 member)
- Design database, participate in building backends for online valuation system using **Java Spring MVC**. (2 members)

INTERNS | GIASAN.VN

Sep 2016 – Jun 2018 (1.5 years) | Hanoi, Vietnam

- Designed and built a real-time real estate market monitoring system using **Spark Streaming**.
- Developed chatbot to asking price, look up real estate information on Facebook Messenger using **RasaNLU** and **Chatterbot**.
- Used **CRF** algorithm to train model to extract real estate information from the description in the listing data, serving the function of posting information on the website Giasan.vn.
- Built a property price heatmap using **Leaflet** and the Tile Coordinate System combined with **Spark** calculations.
- Contributed to the development of website Giasan.vn using **Java Spring MVC** and **Angular**.

SKILLS

MACHINE LEARNING - NLP

Framework:

Sklearn • Pytorch • Rasa •
Spacy •
Fasttext • Tensorflow •
Fondue • Snorkel • Deepdive

AI/ML Model:

CRF • LSTM

BIGDATA

Spark • Hadoop

BACKEND

Java Spring MVC • Python
• Elasticsearch • GraphQL

DATABASE

Mariadb • Posgress

FRONTEND

Angular

EDUCATION

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

MASTER IN DATA SCIENCE AND
ARTIFICIAL INTELLIGENCE
2020 - 2022 | Hanoi, Vietnam

ENGINEER IN INFORMATION
TECHNOLOGY
2013 - 2018 | Hanoi, Vietnam
GPA: 3.2 / 4.0

REFERENCES

Tran Viet Trung, Lecturer at
School of Information and
Communication Technology -
Hanoi University of Science
and Technology

✉ trungtv@soict.hust.edu.vn
☎ 098 506 2706

AWARDS

2017 3rd Nhân tài đất Việt