

NAM VU HAI

DATA ENGINEER AND BACKEND DEVELOPER

<https://vhn-portfolio.vercel.app/>

An experienced, flexible data engineer and backend developer who has knowledge of big data, cloud computing and backend development, always eager to learn and adapt.

EDUCATION

MASTER OF DATA SCIENCE

RMIT

Mar 2026 – Ongoing

BACHELOR OF SOFTWARE ENGINEER -1ST CLASS HONOUR

RMIT Vietnam

Sep 2017 – Oct 2021

SKILLS

PROGRAMMING LANGUAGE

Python(Flask, FastAPI, Quart, OpenCV, Scikit-Learn, TensorFlow & Keras)
Java(Spring Framework, Quarkus)
Scala(Apache Spark)
JavaScript(ReactJs, NodeJs)
PowerShell script

DATABASE

SQL (PostgreSQL, MySQL)
NoSQL(MongoDB, Firebase, ElasticSearch, HBase, HDFS)

CLOUD

Microsoft Azure
Amazon Web Service

OTHER SKILLS

Docker
Design Database
Build ETL pipeline
LGTM full stack
Automation (Airflow, Power platform)
Data Analysis (MS App Insight, MS Monitor)
Data Visualization (Grafana, MS App Insight)
Monitoring system
Teamwork

CONTACT INFO

 +61431799079

 <https://github.com/Limbo2Forma>

 <https://www.linkedin.com/in/vhn6799>

 nam671999hai@gmail.com

EXPERIENCE

DATA ENGINEER – UDE CENTRAL MONITORING, NTTDATA VDS

NOV 2023 – MAR 2026

Design system of data ingestion with MS Azure products.

Building LGTM stacks, designing Grafana dashboard.

Build a Metric Anomalies Detection system using Airflow, GitHub Action, Azure Monitoring.

Create Databricks notebooks for ingestion logic, scheduled, automated availability checks in the customer environment.

Maintenance and migration of old pipeline with PowerShell script

Analysing customer applications and defining ways to integrate them into an E2E monitoring

Working directly with product manager, working peer to peer with Germany Developer

BACKEND DEVELOPER – LICENSIGHT, NTTDATA VDS

FEB 2023 – OCT 2023

Develop backend API using Quarkus.

Design database using MongoDB and PostgreSQL.

DATA ENGINEER – CUSTOMER DATA PLATFORM, VCCORPS

APR 2022 – DEC 2022

Design HBase and Elasticsearch database to store user information.

Using SpringBoot to create API for frontend and create data queue with Kafka.

Use Google Ads API and Facebook Ads API to sync daily statistics to the system.

Using Scala to create spark job to perform mass calculation from user information and merging data

REFERENCE

MR. STEFAN LEUTHOLF

Leading Technical Consultant | Application &
Cloud Wolfsburg, NTT DATA Deutschland SE
Stefan.Leutholf@nttdata.com

MR. QUANG VU

Bioinformatics Scientist at Genestory,
Research Assistant at RMIT
vthanhquang72881@gmail.com

DR. DANG PHAM THIEN DUY

Former Senior Program Manager
and Professor of School of Science
and Technology, RMIT VN
duy.dangphamthien@rmit.edu.vn

DATA ANALYST – DATA MINING AND ANALYSIS, VCCORPS

OCT 2021 – MAR 2022

Using PySpark to analyze logs of user notifications of
various news sites: Kenh14, Soha, Pega, ...

Handling 10TB of daily data traffic

Setup Airflow to create scheduled tasks for analysts

ACHIEVEMENT

TOP 1 KAGGLE COMPETITION

Used machine learning algorithm (Random Forest
Regression, XGBoost, Neural Network) to predict life
expectancy of countries in WHO dataset

[https://www.kaggle.com/c/2020-
cosc2753/leaderboard](https://www.kaggle.com/c/2020-cosc2753/leaderboard)

Used machine learning algorithm (Random Forest
Regression, XGBoost, Neural Network) to predict
book rating

[https://www.kaggle.com/competitions/cosc2789-
pds-sem3-2020/leaderboard](https://www.kaggle.com/competitions/cosc2789-pds-sem3-2020/leaderboard)