

(+84) 382034536

Hanoi, Vietnam

anhluanhoang2001@gmail.com

github.com/tuanx18/Data-Engineer

linkedin.com/in/tuanx18

Anh Tuan Hoang

A recent graduate holding a Bachelor's degree in **Software Application and Technology**, my skill set spans *data manipulation, big data handling, and programming*. Proficient in *SQL, PowerBI, Python, Pandas, Apache Spark, Hadoop, and MongoDB* allows me to be the best at extracting meaningful insights and creating scalable solutions. Equipped a keen programming attitude developed during my university time learning about Software and a natural tendency for leadership, desired to contribute technical expertise and a collaborative spirit to drive impactful data-driven initiatives in a dynamic **Data Engineering** role.

EDUCATION

Bachelor of Information and Communication Technology (Software Technology)

Jan 2020 – Nov 2023

SWINBURNE UNIVERSITY OF TECHNOLOGY

Courses: Database Analysis and Design, User-Centred Design, Introduction to Data Science, Introduction to Artificial Intelligence, Big Data Management, Cloud Computing Architecture, Professional Issues in Information Technology

Computer Science

Aug 2019 – Jan 2020

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

TECHNICAL EXPERIENCE

Data Specialist, Intern

Aug 2022 – Dec 2022

FPT Software

- Worked as a team of seven, built samples and useful databases data specialists learned and researched database management and data insights.
- Mainly working on SQL, Python, Oracle, AWS and Microsoft Database Access

Data Engineer, Scrum Member

May 2023 – Aug 2023

Novaland Data Gathering and Consolidated Report Automation Project

- Developed a chatbot using NLP and an AI API, enabling user interaction, inquiries, and integration with Novaland's system.
- Designed a script using Python and Google Apps Script to detect keywords from Client Phone Calls to extract useful data for the Daily Report
- Extracted the provided datasets to the local machine, cleaned and modified the data, then established an automated procedure for generating Consolidated Reports.

SKILLS AND CERTIFICATES

Programming Languages: Python, C++, JavaScript, HTML / CSS, Scala, Kotlin

Technologies Pandas, Matplotlib, Database (Access, Excel, Azure), NoSQL Database (MongoDB), Batch Processing (Apache Spark, Hadoop), Machine Learning Models, ETL Data Pipeline, Amazon Web Services

Others Git, AI path-finding Algorithms, Agile Scrum Methodology, Object-Oriented Programming, Testing (PyTest), **English IELTS Overall Score 7.5**

PROJECTS [GitHub](#)

Billionaire Research and Exploration

Oct 2023

- Conducted comprehensive analysis leveraging Scala Spark and PySpark and the Billionaires Dataset to explore the realm of extreme wealth.
- Utilized PySpark's Machine Learning Models to implement machine learning algorithms, including K-means clustering and logistic regression.

Technologies: Python, Pandas, Matplotlib, Apache Spark, Confusion Matrix, Machine Learning Classifier Models (MLlib)

IT Department Project, Swinburne University

May 2022 – August 2022

Database Developer

- Organized and Set up Database for the Student Council within the university using Microsoft Database Access
- Managed Lower-level Data Specialists to ensure the dataflow run smoothly

Technologies: Microsoft Access, SQL, Python, Pandas, Machine Learning

Stellar Classification

Mar 2021

- Equipped with a dataset containing 18 curated feature columns, our project holds the promise of advancing astronomical science and discovery in several remarkable ways.
- Aimed to uncover the mysteries of the universe by classifying celestial objects such as stars, quasars, and galaxies.

Technologies: Python, Pandas, Matplotlib, Correlation Heatmap, Confusion Matrix, Classification Report, ROC Curve, Machine Learning Classifier Models.