

# THE VINH HA

Data Engineer / Data Scientist

+61420235867



thevinhha21082003@gmail.com



[Portfolio](#), [GitHub](#), [LinkedIn](#)



Sydney, NSW



## WORK EXPERIENCE

### AI Development Intern

Access Grants | Jul 2024 - Oct 2024

Leveraged LLM-driven system utilizing LangChain framework:

- Extracted key patterns from R&D documents automatically.
- Reduced manual review time by 15 hours weekly.
- Elevated data retrieval precision to 88%.
- Led a team of 6 engineers in daily stand-ups and sprint planning.
- Delivered the first MVP increasing 40% in processing speed.
- Cut down 15% of input errors.

### Assistant Manager

Miniso Australia | Apr 2025 - Present

- Tracked daily sales and inventory trends, supported restocking.
- Audited weekly stock, reducing inventory discrepancies.
- Optimized staff schedules based on sales and store visit pattern.
- Prepared weekly performance reports on customer insights.

## PORTFOLIO PROJECTS

### E-Commerce Analytics Pipeline

Tech stack: Docker, Python, PostgreSQL, Snowflake, Apache Spark, dbt

Built an ELT pipeline mapping E-commerce data into data marts:

- Boosted processing time by 75% via Apache Spark's distributed computing architect.
- Reduce 85% manual workload by creating 3 analytical-ready data marts with dbt.

### Global Fashion Retail Analysis

Tech stack: Docker, PostgreSQL, PowerBI, DAX

Built an end-to-end analysis pipeline for over 6M retail sale records:

- Reduced report generation time by 30%.
- Identified 5 key personas that drove 80% of revenue, suggested a new marketing strategy.

### Diabetes Prediction Machine Learning Model

Tech stack: Python, Scikit-Learn, Numpy, Pandas, Matplotlib, Seaborn

Trained a machine learning model to predict diabetes in female:

- Utilized a dataset of over 800 patient records to train and evaluate the model.
- Achieved prediction accuracy of 80% after key feature selection and fine-tuning process.

## SKILLS

- **Programming:** Python, SQL, Java, JavaScript
- **Technologies:** Docker, Apache Spark, Apache Kafka, Apache Airflow, dbt, Excel, PowerBI
- **Platforms:** AWS, Azure, GCP, Snowflake, Databricks
- **Soft skills:** Data Analysis, Communication, Critical Thinking, Problem Solving, Collaboration

## CERTIFICATIONS

- AWS Certified Solutions Architect Associate
- SQL Associate - DataCamp

## ACTIVITIES

### OGTIP Data Science Training & Internship

Oeson | Feb 2024 - May 2024

- Completed hands-on training in Python, machine learning, and data visualization tools.
- Applied ML models to real-world projects like fraud detection and sentiment analysis.
- Built end-to-end pipelines from data cleaning to model evaluation and reporting.

### Data Analytics Job Simulation

KPMG AU | Nov 2023

- Assessed data quality and completeness and drafted recommendations to address issues.
- Analyzed customer and transaction datasets to identify high-value segments for targeted marketing.
- Built interactive dashboards to present key findings and customer insights visually.

### Representative / Collaborator

FPT University | Sep 2022

- Collected and organized partner university data using Excel; promoted programs internationally in Macquarie University Exchange Fair 2022.
- Built Power BI dashboards to support program presentations.

## EDUCATION

### Bachelor of Information Technology

Macquarie University | 2022 - 2025