

# Mai Nguyen

Tel: +84383891252 | [LinkedIn](#) | [mai.nvphuong@gmail.com](mailto:mai.nvphuong@gmail.com) | [Github](#)

## EMPLOYMENT

---

<b>Data Engineer</b>	<a href="#">Fram^</a>	11/2022 - Now
<ul style="list-style-type: none"><li>Join a Singaporean logistics company's regional team to build, optimize and automate ETL pipeline and data validation using Airflow, Jupyterhub, Kubernetes and Google Cloud services</li><li>Work with BI team to conduct data governance, data validation and implement new logic for metrics</li></ul>		
<b>Data Engineer   Staff Consultant</b>	<a href="#">Rackspace Technology</a>	07/2020 - 10/2022
<ul style="list-style-type: none"><li>Implemented end-to-end modern data platforms/data lake/data warehouse to support analytics using Microsoft's data and analytics services</li><li>Owned the process from tables design, ETL implementation, SIT (system integration testing), UAT (user acceptance testing) to sign-off and hypercare for two workstreams in a migration project for a Southeast Asia's super-app company:<ul style="list-style-type: none"><li>Created a script to analyze queries of +200 tables and +10 reports from a data governance tool and generated text lineage, from which we could estimate the workload and timeline of the migration process</li><li>Created a script that did basic data validation (row duplicate detection, calculating percentage mismatch for metrics, finding missing primary keys), which saved time when doing data validation for a large number of tables</li><li>Implemented CI/CD for ETL pipelines with source code management tool</li><li>Did root-cause analysis to tune and optimize data pipelines and queries</li><li>Hosted sessions to handover ETL processes, reports to users</li></ul></li><li>Implemented ETL pipelines and data validation stage for a Norwegian gas shipping company. Created a script to quickly updated pipelines when change requests occurred, which saved time and efforts and generated no error compared to manual update</li><li>Owned the implementation, SIT and supported UAT of ETL pipelines for a trading company from the Maldives. Created STTM (source to target mapping) document, functional specifications and technical design document</li><li><b>Leveraged knowledge</b> in Azure Data Factory, Azure SQL Database, Databricks, Apache Spark, Azure DevOps, Dimensional Modeling, Change Data Capture, ETL Process, Power BI, Python, SQL</li></ul>		
<b>Data Engineer</b>	<a href="#">Deliveree</a>	10/2018 - 06/2020
<ul style="list-style-type: none"><li>Led a 3-member team to conduct some ETL projects and trained new members.</li><li>Built in-house tools that schedules ETL processes, from many sources (PostgreSQL, Twilio, Crashlytics, App Store,...) to data lake (PostgreSQL) or export to reports (csv, google sheets) with Celery and Airflow.</li><li>Optimized an in-house tool that scrapes data from analytics web apps. It can open and switch multiple tabs within a Firefox browser for data extraction.</li><li>Built an in-house tool that uses asyncio streams to do query load balancing.</li><li><b>Leveraged knowledge</b> in ETL Process, Python's concurrency programming, Apache Airflow, Celery, Producer-consumer model, RabbitMQ, Selenium, Google Data Studio, Google Apps Script, TCP/IP socket, Project Lead</li></ul>		

- Built frontend and backend for a powerful self-service business intelligence platform that aims to enable self-service data access for the entire organization.
- Managed product backlogs and updated new product's features, worked with product managers to plan and execute product roadmap.
- Performed data analytics for insights about features usage.
- **Leveraged knowledge** in BI Tools, Ruby on Rails, VueJS, PostgreSQL, ETL Process, Product Development, SQL

## EDUCATION

---

RMIT University Vietnam (GPA: 3.3 / 4) - Bachelor of Information Technology

2014 - 2017

## SKILLS

---

**ETL Tools:** Azure Data Factory, Databricks, Apache Airflow

**Databases:** Azure SQL Database, PostgreSQL, Azure Synapse Analytics, Azure Storage, BigQuery

**Languages:** SQL, Python

**BI Tools:** Power BI, Holistics, Redash, Superset, Looker Studio

**CI/CD:** Azure DevOps, GitHub, Bitbucket

**Database Design:** Star Schema, Change Data Capture

## CERTIFICATES

---

- [Azure Data Engineer Associate](#)
- [Azure Administrator Associate](#)
- [Databricks Certified Data Engineer Associate](#)
- [Power BI Data Analyst Associate](#)
- [Google Cloud Certified - Associate Cloud Engineer](#)