Long Hoang Tran

Linkedin | +17746969820| maxtran1997@gmail.com| Github

A data enthusiast with 6 years of industry experience, passionate about offering technical solutions, building data systems and applying data analytics, data engineering, and NLP solutions for companies in need. On the side, I also train junior data professionals and beginners.

Technical Knowledges

Relational Databases : MySQL , SQL Server, Postgres, Big Query, Oracle, AzureSQL , AWS RDS, AWS Redshift

NoSQL Databases : MongoDB, Cassandra , Elastic Search.

Big Data : Databricks, Hadoop Distributed File System (HDFS) , Apache Spark , Apache Flink

DataLakehouse : Azure Synapse, AWS Athena , Oracle , Presto

Programming : Python , Scala

Business Intelligence : PowerBI , Tableau , DataStudio , Cluvio , Oracle BI , Grafana, Superset , Kibana

Cloud Infrastructure : AWS , Google , Azure , Oracle

Event Streaming : Kafka , ksqlDB, Apache Druid

ETL tools : Pentaho , Talend , SSIS , Azure Data Factory, Oracle Data Integration.

Jobs Scheduler , Automation : Airflow , Celery , Cronjobs , N8N

Spreadsheets : Excel and Google Spreadsheets

Technical Skills

Structure Query Language : ETL & ELT, transforming OLTP data from Data Lakes to OLAP data in Data Warehouse using joins , window function ,pivot , optimize query, dynamic SQL.

ETL data with batch processing and realtime streaming using Spark and Flink with Python / Scala.

Developing data infrastructure , real time analytics data pipelines from data lakes to data warehouses and Analytics tools from stratch.

Crawling data using Selenium , Beautiful Soup and processing with Regex.

Performing data mining techniques and analyzing data for diagnostic analytics using SQL, Python , and predictive analytics using Machine Learning libraries and advance statistics such as Pycaret , Profet and Statsmodels.

Developing Visualization , analyzing data using calculated fields and metrics , developing embeded analytics dashboards using PowerBI and Tableau and other cloud services.

Developing Time Series Analysis , Supervised / Unsupervised ML Models and ML pipelines with ML Ops for regression and classification purposes, quantitative analysis and quantitative analysis using Python.

Quick adhoc reports with excel / google sheets applying data transforming functions , pivot table and charts.

WORK EXPERIENCE

FPT Telecom   **HCMC, Vietnam**

Analytics Engineer   *1st March 2020 – 31ˢᵗ March 2023*

- Assisting other Business Departments by analyzing customer behavior and customer interactions data Customer Segmentation Models that help maintain 30 percent increase in Revenue annually.
- Conducting diagnostic analytics such as Customer Taste & Behavior Analysis as well as Trending Analysis for weekly and monthly reports and making ad hoc reports for managers.
- Crawling data from other sources and performing comparison with internal data to identifying key trends in Customer's Trending and Sentiment Analysis using Python , Beautiful Soup and Selenium.
- Using data mining techniques to explore insights from data and developing new metrics using Power Query and making Business Intelligence Dashboards / realtime & streaming dataset with PowerBI Service.
- Developing and maintaining automation data pipelines for DA & BI team by using PySpark to ETL logs data from Kafka to HDFS and Data warehouses, or from noSQL databases like ElasticSearch / MongoDB.
- Developing ELT flows between data warehouse and data marts (SQL Server , AzureSQL , PostgreSQL , MySQL) using Transact-SQL and event scheduler using Airflow / Shell / crontab.
- R&D new data analytics models and predictive models with Big data for Churn Rate Prediction / Recommendation Systems / Related Search and Digital Rights Managements.

FPT Software   **Full time Remote**

Machine Learning Engineer & Data Engineer   *1ˢᵗ March 2022 – 28ᵗʰ February 2023*

- Working in Machine Learning, Deep Learning and AI projects for US and Japan clients by building , optimizing and preprocessing Supervised / Unsupervised models using Python , SKLearn and AutoMLs.
- Improving & Optimizing current AI models performance by applying LSTM for TimeSeries Analysis, Natural Language Processing models with Tensorflow and Keras using BERT and SimCSE.
- Working in data project as an analytics engineer for US client by building an entire real time data pipeline from scratch using Kafka , PySpark , MySQL and Cluvio.

Topica Edtech Group   **Hanoi, Vietnam**

Data Analyst   *1ˢᵗ June 2019 – 30 October 2019*

- Building data warehouse for the company using SQL Server , performing data cleansing , data mapping and data standardizing from company current data from Google Sheets to Data Warehouse.
- Building BI Dashboards and embeded analytics system for HR performance tracking purposes (such as Internal orders , recruitment and KPIs) using Power BI and PowerBI Service.
- Making ad-hoc reports and data queries using SQL and Google Sheets for other departments.

Nielsen Vietnam   **HCMC, Vietnam**

Market Research Analyst   *Septemmber 2018 – March 2019*

- Performing Data Processing, Quantitative Analytics and Qualitative Analytics for Market Research needs using Excel and SPSS.

PREVIOUS PROJECTS

1.Customer 360 Project for Customer Analytics

June 2021 – March 2023

Developing Customer Segmentation models by using RFM models and other statistical methods (Inter Quartile Range / Histogram / Distribution) to analyze customer behavior and payments analytics to enhance customer care, decrease customer turnover and up sale revenue boost that help maintain a 30 percent increase in revenue annually using Python , Spark , SQL and PowerBI.

2.Develop data pipeline for ML & Analytics purposes for an American Hiring Platform

June 2022 – March 2023

Migrating data from AWS to Oracle Cloud MySQL database using Python

ELT data from Data Warehouse to Data Mart (MySQL) using Routines and Triggers

ETL data from Cassandra DB to Oracle Cloud MySQL using Python & Spark, making pipelines that process data in near real time interval (2 minutes).

Develop realtime Analytics Streams using Kafka and Cluvio for realtime visualizations.

3.Optimize and Improve performance of search engine for an American rental company

April 2022 – June 2022

R&D solutions to improve performance of current Natural Language Processing model using SimCSE and AutoEncoder , Tensorflow and Pytorch, implementing the solutions and boosting 200% of efficiency for their related search system.

4.R&D Embedded AI Algorithms for Electrical Devices for a Japanese Electrical Company

February 2022 – April 2022

Developing Supervised and Unsupervised Machine Learning models for classification and regression tasks, optimizing models and turning models for better prediction results on their devices using Python, implementing the solutions and achieving 8x% in accuracy for the prediction.

5.Data Integration between IPTV system and OTT system of a Telecommunication Company

October 2021 – January 2022

Integrating between 2 large data systems, synchronize data between data warehouses (PostgreSQL – MySQL – SQL Server) , remodifying ETL data pipelines that combines data from different HDFS using PySpark.

Governance and checking data accuracy between analytics systems (Redash , Kibana and PowerBI)

6. HRIS System & Data Warehouse build for a South East Asia Unicorn Edtech Startup

June 2019 – December 2019

Designing and Building a Data Warehouse and Business Intelligence system for the company using SQL Server and Power BI. Mapping data from current Google Sheets to fact and dim tables in Data Warehouse for company HR operations.

Building PowerBI dashboards and develop metrics as well as embedded analytics system for the company to allow departments to monitor internal operations and recruitment process.

7. FMCG & AUTOMOBILE MARKET RESEARCH

December 2016 – December 2018

Performing data processing and data analysis about content strategies for company incoming TVCs and marketing purposes using SPSS and Excel.

8. Data Science Capstone Project at HCMUS

Applying Data Analytics and Data Engineering to realistic issue : Where to open a restaurant in HCMC. Develope a program that can crawl venues data and geospatial data from government sources , Foursquare and Google API.

Performing data cleasing from unstructured text data to structured data using Python and Regex. Using Pandas and Matplotlib and Folium as well as SK Learn to analyze the crawled data to answer the research question.

9. Natural Language Processing Final Project at Clark University

Developing a program that can crawl economics articles from US newspapers and analyze them for financial investment purpose using Python and Selenium.

Perform sentiment analysis to evaluate market sentiment using RoBERTa , implement gemma-3-27b-it as LLM and optimize the appropipate prompt engineering to identify the mentioned industry as well as suggesting the stocks within the industry for user considerations.

Developing front end as interactive dashboard that is user friendly using streamlit , DASH and Plotly.

TEACHING EXPERIENCE

Mindx Technology School Vietnam

Data Science Partner   *June 2021 – June 2022*

Making lectures and developing Data Science training programs for a Vietnamese edtech unicorn and teaching their Data Analytics classes with an average 30 – 40 students per class.

Academic Manager for Data Science Program   *June 2022 – January 2023*

Responsible for the training quality of the center's Data Science program. Developing metrics and benchmarks to monitoring students' performance and hiring new teachers. Ensuring the quality of the lectures and the context are sufficient and up to date through students' and professionals feedback and student's results.

Regularly conducting enterprise and university training as requested.

Assisting students and other teaching staff to solve conflicts and working with the board of directors to further develop the program.

Freelance Data Science Instructor   *October 2022 - Present*

Entrepreneur for a small data analyst and data engineer training businesses online that have successfully supported more than 200 students from zero to their first Data Analyst / Data Engineer job offer in the past 4 years.

Making Data Science training contents and Data Science related Facebook posts for students and other data enthusiasts in need , with an average 3000 followers and 500-1000 interactions per post.

Offering technical solutions consults and trainings for businesses in need.

Clark University

Teaching Assistant   *August 2024 – December 2024*

TA for Data Mining Class , being professor's first line of defense , responsible for replying emails , clarifying technical issues and grading 43 students works, making remedial sessions for students in needs and giving lectures about using Splunk and other data mining tools.

EDUCATION

Clark University

Master of Science in Data Analytics with 30 percent of scholarship   *August 2023 – May 2025*

Royal Melbourne Institute of Technology   *October 2016 – December 2019*

Bachelor of Business in Economics and Finance with Minor in Market Research

Ho Chi Minh City University of Natural Sciences   *June 2020 – June 2021*

Applied Data Science Professional Certificate

CERTIFICATES

Oracle Certified Data Science Professional	bit.ly/oracleds
IBM Data Science Professional Certificates	bit.ly/3ypVIJ5
IBM Advanced Data Science Certificates	bit.ly/3YAcnEq
IBM Data Engineering Professional Certificates	bit.ly/3FfjCe6
Google Data Analytics Certificates	bit.ly/3ZTN2WP
FPT Group Big Data Engineer Professional Certificates	bit.ly/3YCpZi8

<u>REFERENCES</u>

FPT Telecom	**Link**
FPT SOFTWARE	**Link**