NGO Thi Thu Trang

Data Engineer | Ph.D. in Computer Science

Looking for a new opportunity

With 3+ years of experience in Data Engineering research and nearly 1 year of in industry experience in France, I specialize in designing and implementing efficient data pipelines in Kubernetes and cloud environments such as GCP, AWS, and OVH. I contributed to the international research community with two articles on big geospatial data integration, data processing, and data analysis.

PROFESSIONAL EXPERIENCES

Data Engineer
01/2024 – 08/2024
Paris, France

NW Group – Energy Transaction

Project: Designing and implementing scalable data pipelines for IoT data on Kubernetes infrastructure. Migrating NW Data Warehouse from GCP BigQuery to open-source solutions like Apache Iceberg and Trino.

Mission:
- Designed a data analytics platform from scratch using Kubernetes and cloud solutions.
- Implemented ETL processes by setting up Kubernetes jobs to ingest data from various static and streaming sources.
- Set up and deployed Apache Flink jobs for an IoT alerting system to improve energy power station monitoring.
- Developed Apache Flink jobs for ETL to Apache Iceberg data lakehouse, supporting decision-making and machine learning purposes.
- Deployed a Trino cluster analytics query engine on top of a Kubernetes cluster.
- Set up and managed the Influx Stack, using tools like Telegraf (with XPath parser) to retrieve data from multiple MQTT brokers, and Grafana for dashboard visualization and an alerting system.
- Configured Apache Airflow as the data orchestrator and created benchmark dashboards.

Main stack: Python, SQL, Flink, Airflow, Iceberg, Kafka, Kubernetes/GKE, Github Actions CI/CD, Docker, dbt, influxDB stack, Debezium CDC, MQTT, Agile.

PhD Researcher in Data Engineering
02/2020 – 10/2023
Clermont Ferrand, France

LIMOS CNRS – Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes

Project: Proposing new methods for the integration of IoT geospatial sensor data in the context of environment and agriculture, and for data-driven decision-making (data warehouse) and data processing optimizations.

Mission:
- Explored subjects (i) spatial data modeling and storage, (ii) big spatial data integration, and (iii) continuous querying in the context of streaming cloud environments.
- Proposed an analytical tool that enables real-time ETL analysis of geo-referenced data and fast searching on top of ELK stack [1], and proposed a mediation system belong with optimized techniques that enables the streaming analysis on top of Apache Spark [2] and cloud solutions.
- Implemented the proposals using Python, Docker, and the cloud-based tools (An internal cluster and GCP).
- Supervised a master internship student for his M2 internship projects in the domain data science.

Main contributions:
[1] NGO et al., March 2023, "A New Approach Based on ELK Stack for the Analysis and Visualization of Geo-Referenced Sensor Data" published by the Springer Nature Computer Science Journal (URL: https://link.springer.com/article/10.1007/s42979-022-01628-6)
[2] NGO et al., October 2023, "A Mediation System for Continuous Spatial Queries on a Unified Schema Using Apache Spark." published by the Big Earth Data Journal (URL: https://www.tandfonline.com/doi/full/10.1080/20964471.2023.2275854)

Main stack: Python, SQL, Elasticsearch stack, Apache Spark, Apache Kafka, Docker, Git

Research Trainee in Data Engineering
07/2018 - 04/2019
Bordeaux, France

LaBRI CNRS– Laboratoire Bordelais de Recherche en Informatique

Project: Implementing and optimizing algorithms of skylines queries

Mission:
- Explored the state-of-the-art of skyline queries and its algorithms.
- Implemented and optimized the algorithms of skyline queries on PostgreSQL.

Main stack: Python, PostgreSQL, GitHub.

PERSONAL INFORMATION

🏠 Île-de-France, France
📞 (+33) 7 66 08 99 32
✉ anna.ngo13@gmail.com
in https://www.linkedin.com/in/thi-thu-trang-ngo-1643a062/
GitHub https://github.com/AnnaNgo13

EDUCATION

02/2020 – 06/2023
Clermont Auvergne University | France
Ph.D. in Computer Science
Specialization: Data Engineering

11/2017 – 05/2019
Bordeaux University | France
Master in Software Engineering

09/2009 – 04/2014
FPT University | Vietnam
Bachelor of Business Administration

TECHNICAL SKILLS

Languages	Python, SQL
Libraries	Pandas, PySpark, PyFlink, NumPy, Matplotlib
Database	MySQL, PostgreSQL, InfluxDB, Elasticsearch, MongoDB, Big Query, Apache Iceberg, S3
Data Processing	Kafka, Flink, Spark, Airbyte, Airflow, dbt, Trino, Telegraf, logstash, Airbyte, MQTT, Debezium CDC
Visualization	Kibana, Grafana, Apache Superset
Cloud	GCP, AWS, Kubernetes, Docker, Terraform
Others	Git CI/CD, SharePoint, MS Visio, SOAP UI, Balsamiq, Postman

LANGUAGES

Vietnamese
Mother tongue

English
Fluent

French
Intermediate

CERTIFICATIONS

AWS DevOps
GCP Core Infrastructure
Advanced Photoshop (Ps)

INTERESTS

Swimming, action movies, volunteer

Senior IT Business Analyst

08/2016 – 06/2018
Aperia — US Fintech Company   Hochiminh, Vietnam

Project: Designing and maintaining functionalities for a Payment Card Industry product
Mission:

- Maintained the knowledge of the Payment Card Industry Data Security Standard (PCI DSS).
- Handled the client's requests, analyzed, and transferred the request to the development team.
- Designed and updated the Software Requirements Specification (e.g., workflows, and wireframes, activities diagrams).
- Reviewed the ticket packages before delivering to the US site.
- Supported the UAT User Acceptance Testing phase.

Main stack: SQL Server 2013, Balsamiq 3, MS Visio, SOAP UI 5, Agile.

IT Business Analyst

11/2014 – 06/2016
FPT Software — IT outsourcing company (ESN)   Vietnam - Japan

Project 1: Designing an education portal for a Japanese client in Tokyo, Japan

- Organized meetings with stakeholders to gather requirements in Tokyo.
- Designed wireframes and business requirements based on the client's requests then documented them in the Software Requirement Specification (SRS).
- Transferred confirmed requirements to the development team (dev and tester).
- Managed the Q&A between the development team and clients.
- Supported team members to review their SRS documents.

Project 2: Migrating 600 IBM Lotus Notes applications to SharePoint (SP) standard 2013
Mission:

- Estimated the complexity of the IBM Lotus Notes apps in three categories: complex, medium, simple.
- Reviewed the implemented apps in SharePoint version and analyzed functionalities of the assigned apps to correct the SharePoint version.
- Worked with the migration team on the historical data of the assigned app to update the SRS documents and worked with the tester team to transfer the application requirement and answered their questions regarding the application requirements.
- Reviewed the package before the delivery process and supported the UAT phase.

Main stack: IBM Domino Designer, SharePoint, SQL, MS Visio, Jira, Agile.
Reward: The best quality officer in 2014

Test Engineer

04/2014 - 10/2014
FPT Software — IT outsourcing company (ESN)   Hanoi, Vietnam

Project: Upgrading over 500 SharePoint apps version 2003 to version 2013
Mission:

- Reviewed the SharePoint version 2003 and wrote the test cases for SP v2013.
- Created the test cases for the data migrations process, back-end jobs, workflows, business functionalities, and views.
- Executed manually the test cases of the assigned apps and followed up the bugs via Jira.
- Created user manuals and supported the UAT phases for the assigned apps.

Main stack: SharePoint version 2003 and version 2013, MS Excel, SQL Server 2013, Waterfall.