NGUYEN TRUNG NGHIA

DATA ENGINEER | DATAOPS | MLOPS

Thu Duc, Ho Chi Minh City, Vietnam

(+84) 334 349 789 trungnghia294@gmail.com in/tnghia294 Ren294 Site

SUMMARY

Highly motivated Data Engineer with a proven ability to design and build robust, scalable data pipelines. Skilled in leveraging cutting-edge big data technologies. Hands-on experience in data processing, system architecture, and cloud technologies through personal projects. Eager to collaborate with AI and BI teams, contribute to innovative solutions, and continuously expand knowledge in ML and AI to bridge the gap between data engineering and advanced analytics in a fast-paced, professional environment.

EDUCATION

Ho Chi Minh City University of Technology and Education Aug 2021 – 2025

Bachelor of Engineer Science, Information Technology *Ho Chi Minh City, Vietnam*

- Specialization in Information Systems, focusing on data engineering and big data technologies.
- Relevant coursework: Database Management, Data Analysis, System Architecture, Advanced SQL, Data Warehousing, and Big Data Technologies.

CERTIFICATIONS

Samsung Innovation Campus - Big Data | *Certificate* May 2024 – Aug 2024

Samsung Innovation Campus - SIC

- Key skills: Big data processing with Hadoop, Spark, and data visualization techniques.

IBM Data Engineering Specialization | *Certificate* Oct 2023 – Aug 2024

Including 13 certificates - Coursera

- Key skills: Data modeling, pipeline development, ETL processes, and data analysis using Python and SQL.

AWS Academy Graduate - AWS Academy Cloud Foundations | *Certificate* Aug 2023 – Dec 2023

AWS Academy

- Key skills: Understanding of AWS architecture, core services (EC2, S3, RDS) and cloud cost management.

EXPERIENCE

Data Engineering Mentor | *Portfolio* July 2024 – Dec 2024

TechData.AI *Ho Chi Minh City, Vietnam*

- Mentored students in Data Engineering, covering topics like Python, Linux, Docker, Apache NiFi, Apache Airflow, Apache Spark, Apache Kafka and GCP.
- Guided real-world projects in retail, finance, and healthcare, focusing on data pipelines, warehousing, and visualization using Power BI.
- Provided personalized mentoring, feedback on assignments, and improved course materials based on student needs.

PROJECTS

SmartTraffic Lakehouse for HCMC | *Personal Project* | *Github* Aug 2024 – Jan 2025

- **Overview:** A Smart Traffic Management System for Ho Chi Minh City, Vietnam leveraging batch and real-time data processing, intuitive dashboards, and monitoring tools to optimize traffic flow, enhance safety, and support sustainable urban mobility through advanced analytics and user-friendly applications.
- **Technologies:** Apache NiFi, Debezium, Apache Kafka, Apache Flink, Apache Spark, Redis, Apache Airflow, Clickhouse, Promotheus, Grafana, Trino, Streamlit, Metabase, Apache Superset, Minio, LakeFs, Apache Hudi, Apache Hive, Apache Parquet, Postgres, Docker.

Ecommerce Insights Lakehouse | *Personal Project* | *Github* June 2024 – Aug 2024

- **Overview:** Developed an end-to-end E-commerce Data Lakehouse to unify structured and semi-structured data from multiple sources, enabling high-performance analytics and business intelligence. The architecture supports real-time insights, trend analysis to optimize business operations.
- **Technologies:** Apache NiFi, Apache Sqoop, Apache Kafka, Redpanda, Apache Spark, Delta Lake, HDFS, Apache Hive, Cassandra, Grafana, Kibana, Apache Airflow, Power BI, Jupyter Notebook, Docker.

Log Analysis Project | *Personal Project* | *Github* May 2024 – June 2024

- **Overview:** Built a scalable, fault-tolerant log analytics pipeline based on the Lambda architecture, capable of processing and analyzing NASA web server logs in both real-time and batch modes. Provides comprehensive log insights, anomaly detection, and historical trend analysis to enhance system monitoring and security.
- **Technologies:** Utilized Ubuntu Server, Python, Apache NiFi, Apache Kafka, Apache Spark, Apache Hadoop HDFS, Apache Hive, Cassandra, Grafana, and Power BI.

COVID-19 Data Processing | *Personal Project* | *Github* Apr 2024 – May 2024

- **Overview:** Engineered a real-time COVID-19 data pipeline for efficient data ingestion, transformation, and visualization, enabling public health officials to track key metrics. Integrates streaming and batch processing to deliver up-to-date insights via interactive dashboards and automated reports.
- **Technologies:** Utilized Amazon EC2, Docker, Apache NiFi, Kafka, Redpanda, Apache Spark, Hadoop HDFS, Apache Hive, Apache Airflow, and Amazon QuickSight.

Employee Data Warehouse | *Team Project (4 members)* | *Github* Apr 2024

- **Overview:** Designed and implemented a data warehouse solution to centralize and analyze employee data for strategic decision-making in the HR department.
- **Technologies:** SQL Server, SSIS, SSAS, Power BI.
- **Role:** Data Warehouse Designer.

Crawling Tax Code and Company Information | *Personal Project* | *Github* Oct 2023

- **Overview:** Developed an automated web scraping pipeline to extract and process company information and tax codes in Vietnam.
- **Technologies:** Apache Airflow, Python, BeautifulSoup.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, SQL.

Data Engineering Tools: Apache NiFi, Apache Kafka, Apache Airflow, Apache Spark, Apache Flume, Apache Sqoop.

Databases: MySQL, PostgreSQL, SQL Server, Cassandra, HBase, Redis, ClickHouse, MongoDB.

Big Data Technologies: Hadoop, Spark, Hive, Delta Lake, Hudi, Trino, MinIO, AWS S3, LakeFS.

Data Visualization: Power BI, Grafana, Tableau, AWS QuickSight, Metabase, Apache Superset.

Cloud & Storage: MinIO, AWS S3, LakeFS

Monitoring & Logging: Prometheus, Grafana, Kibana.

Operating Systems: Linux/UNIX.

SOFT SKILLS

Analytical Problem-Solving: Skilled in breaking down complex data challenges, identifying efficient solutions, and optimizing workflows.

Growth Mindset & Curiosity: Committed to continuous learning and exploring emerging data technologies.

Independent Work: Demonstrated ability to work independently through successful completion of personal projects.

Adaptability & Quick Learning: Able to rapidly learn new concepts and adapt to evolving technologies.