Nguyen Thanh Hung

0355726538 · nguyent.hung.pntv9@gmail.com · https://github.com/hnguyen021
District 4, Ho Chi Minh City

INTERN/FRESHER DATA ENGINEER

Driven by the incandescent zeal of youth and an unwavering commitment to personal growth, I always wanted to learn and constantly evolve. Therefore, I hope to have the opportunity to join a new working environment gain experience in a real-world environment, and become a proficient data engineer and data analyst.

PROFESSIONAL EXPERIENCE

DXC Technology Service Vietnam **Jan 2022 - Feb 2023**

. Net Developer

- Microsoft Power platform (Power app, Power automate, Share point online)
- .Net legacy systems development(.Net core,SQL Server, Oracle, MS Report Builder)
- Back-end processing for .Net framework and .Net core

PROJECTS

Building Data Warehouse and Sagemaker Recommendation Engine **Aug 2023 - Oct 2023**
https://github.com/hnguyen021/Sagemaker-Recommendation-Engine

- Crawling data from AWS resources, conducting ETL process by Pentaho data integration, storing data into PostgreSQL and AWS Redshift as a basic Data Warehouse which follows star schema, and visualizing data by Power BI
- Using Python and AWS SageMaker, AWS S3 bucket, and Factorization Machine algorithm to build a personalized product recommendation engine

Building ETL Data Pipeline and Stock Prices Prediction Engine **Oct 2023 - Nov 2023**
https://github.com/hnguyen021/Apache-Airflow-Docker

- Utilizing Apache Airflow and Python, Crawling stock data to construct a robust DAG serving as an efficient ETL data pipeline. Within the encapsulated Docker environment, build an LSTM network to develop an engine for predicting stock prices.

Building Realtime Data Streaming Pipeline **Nov 2023 - Jan 2024**
https://github.com/hnguyen021/kafka-spark-streaming-pipeline

- Within a Docker environment, we leverage Apache Airflow and Python to extract data from APIs, constructing a resilient DAG tailored for efficient streaming data processing in Kafka. The PySpark framework is then applied to handle, transform, and seamlessly write the streaming data into a Cassandra database.

Building Realtime Election Voting System **Jan 2024 - Mar 2024**
https://github.com/hnguyen021/Realtime-Election-Voting-Engine

- Within a Docker environment, utilizing Python to extract data from APIs and store it in the PostgreSQL database, and duplicate the votes data into Kafka. The PySpark framework is then employed to manage, transform, aggregate, and seamlessly write the streaming data into specific topics on Kafka. Finally, using Streamlit for displaying real-time voting data by consuming aggregated data from Kafka and PostgreSQL.

EDUCATION & CERTIFICATIONS

Ton Duc Thang University

- Bachelor's Degree in Computer Science
- 2018 – 2023
- GPA 7.5/10

English Certification

- **IELTS Academic Certification 5.5**

SKILLS/TECHNOLOGIES

- **AWS Services: S3 Bucket, SageMaker, Redshift, EMR**
- **Database: SQL Server, PostgreSQL, MongoDB**
- **Apache PySpark, Apache Airflow, Apache Kafka**
- **Power BI**
- **Docker, Docker Compose**
- **Pentaho Data Integration**
- **Agile Scrum, Github**