Phan Thanh Hai

Thu Duc City, HCMC | 0912738824 | phanthanhhai2411@gmail.com
linkedin.com/in/haiphan24112000 | github.com/haiphan2000

Summary

Potential graduate student with more than 1 year of hands-on experience in Python, SQL, data visualization tools and working with both technical and nontechnical business stakeholders, seeking a data engineer at **VNG** to leverage my experience in building robust data pipelines, and enhance effectiveness in the data on-boarding process

Education

University of Information Technology Thu Duc City, HCMC
Bachelor of Computer Science *2018 – 2023*
GPA: 8.52/10.0
Relevant Courses: Design and analysis of algorithms (**9.5**/10.0), Object-oriented Programming (**9.2**/10.0), Data Structures and Algorithms (**9.1**/10.0), Databases(**8.7**/10.0)
Activities: Teacher assistant for Introduction to Programming and Information Literacy class

Skills & Additional

Languages: Conversational Proficiency in English (TOEIC LR: 775); Basic Communication in Korean
Programming Languages: Python
Database Management: Microsoft SQL Server, Data Modeling, ETL, Data Warehousing
Tools and Software: PySpark, Azure Data Factory, Azure Databricks, Azure Synapse, Apache Superset, Apache Druid, Apache Airflow, Power BI, Visual Studio Code, Jupyter Notebooks, PowerPoint, Github
Soft Skills: Problem Solving, Teamwork, Communication, Adaptability, Attention to detail, Work Ethic
Courses & Certifications: End-To-End Data Engineering Project - LinkedIn Learning, Problem Solving (Intermediate) - HackerRank, SQL (Advanced) - HackerRank, Python (Basic) - HackerRank, English For Career Development - University of Pennsylvania (Coursera)

Projects

Real Estate Prices 10/2023 – 11/2023

Description: The project aimed to create a solution for web scraping, PySpark processing, and real-time analytics, providing dynamic insights into house prices
Project Type: Personal project
Technologies: *Python, PySpark, Apache Airflow, Apache Druid, Apache Superset, Jupyter Notebooks*
Key Contributions:

- Developed a web scraper for Immo Scout 24, extracting real estate data and transforming it into a structured format for analysis
- Utilized **PySpark** in **Jupyter Notebooks** to perform data processing, cleaning, and modeling, resulting in the creation of a predictive Machine Learning (ML) model for house prices
- Orchestrated a scalable data pipeline using **Apache Airflow**, managing tasks such as web scraping, ETL processes, and machine learning model training for efficient data processing
- Implemented data versioning with **Delta** tables, and integrated ML model predictions into **Apache Druid** for real-time analytics

GitHub Link: github.com/haiphan2000/real-estate-prices

Simple End-To-End Azure Data Engineering 04/2023 – 05/2023

Description: Implemented a comprehensive Azure data engineering solution, initiating data ingestion from a local SQL database and concluding with dynamic reporting in Power BI. The project focused on ETL pipeline techniques and served as a practical learning experience for businesses migrating local data to the cloud

Project Type: Personal project
Technologies: *PySpark, Azure Data Factory, Azure Databricks, Azure Synapse, MS SQL Server*
Key Contributions:

- Implemented secure data transfer using **Azure Data Factory**, leveraging the installation of self-hosted integration runtime for enhanced control
- Established seamless connections between **Azure Data Factory** and the local **MS SQL Server**
- Configured a copy pipeline within **Azure Data Factory**, transferring tables to **Azure Data Lake**'s
- Implemented a medallion data lake architecture using **Azure Databricks** and **PySpark** for efficient data processing

GitHub Link: github.com/haiphan2000/simple-end-to-end-azure-data-engineering

Global Superstore Sales Analyst 11/2022 – 12/2022

Description: Designed and implemented a dashboard tailored for retailers, offering a comprehensive analysis of product performance. Key features include showcasing yearly, quarterly, and monthly growth rates, which enhancing decision-making through an understanding of product performance trends
Project Type: Personal project
Technologies: *Power BI*
Key Contributions:

- Built **Power BI** dashboard using data from global superstore sales data to visualize core business KPIs (total sales, total profit, and total return), saving time of manual reporting work
- Provided valuable insights to businesses about the effectiveness of their sales strategies through charts and visualizations
- Applied time series analysis to generate sales forecasts for the next 6 months

GitHub Link: https://github.com/haiphan2000/superstore-sales-dashboard

A Comprehensive Analysis Of Zara's Supply Chain 03/2022 – 05/2022

Description: The focus of the project was the analysis of Zara's "fast fashion" model, aiming to evaluate both its successes and limitations within the supply chain
Project Type: Group project (5 members)
Technologies: *PowerPoint*
Key Contributions:

- **Led** a team of 4 students with **e-commerce backgrounds** in facilitating the presentation of insightful analysis on Zara clothing company's supply chain to the lecturer, resulting in our team scoring **25% higher** than the class average.
- Drafted a report for a broad analysis of Zara's fashion brand, covering aspects such as its vision, mission, sales channels, business model, etc.
- Utilized non-coding visualization tools, specifically **PowerPoint**, to generate comprehensive analysis chart and graphs pertaining to the supply chain model, pricing strategy, etc.

GitHub Link:
github.com/haiphan2000/EC214-Introduction-to-Supply-Chain-Management/tree/main/Zara