Dai Vo - Data Engineer

+84 944 433 047 | 2000 | daivt1805@gmail.com | LinkedIn | GitHub | HoChiMinh, VietNam

Objective

Detail-oriented data engineering with a strong focus on cost optimization, seeking a challenging role to design and implement efficient data pipeline. Committed to leveraging cloud-based technologies, automation, and data modeling expertise to deliver cost-effective solutions and optimize resource utilization.

Work Experience

Mobile World JSC HoChiMinh, Vietnam

DBA System Fresher *Oct 2023 – Present*

- Collaborated with team members and mentors to find effective solutions to problems.
- Built a change data capture (CDC) pipeline to capture changes from Postgres using Debezium.
- Pushed messages into Kafka for subsequent processing.
- Processed streaming messages from Kafka using Apache Flink and stored them in TimescaleDB.
- Built a Hazelcast cluster to cache distributed memories for saving precomputed values using Java.
- Developed an API for querying data from Hazelcast using Java Spring Boot.
- Utilized Apache Spark for end-of-day batch processing to ensure data accuracy using Python.
- Visualized insights of data using Superset.

Education

University of Information Technology, VNU-HCM HoChiMinh, VietNam

B.Sc. in Data Science; ***GPA: 8.51/10 Very Good*** *Sep 2019 – Jun 2023*

Skills

Programming: Python, Java, SQL, Git.

Data Engineering: Spark, Hadoop, Airflow, Kafka, Docker, AWS.

BI tools: Superset.

Soft skill: Life-long learning, critical thinking, logical reasoning, research, problem-solving.

Languages: Vietnamese (Native), English (Limited working).

Certificates

MWG IT Fresher *MWG*

Accenture North America Data Analytics and Visualization Virtual Experience *Forage*

Lyft Back-End Engineering Virtual Experience Program *Forage*

IBM Data Science Specialization *Coursera*

Machine Learning Specialization *Coursera*

Projects

Vehicle Speed Estimation on the Vietnam's Street Real-time

Video from data sources are pushed into model by Kafka. The vehicles are detected and tracked to estimate speed real-time.

- Train models on multiple GPU.
- Use Kafka to send the frames of video from data source into model.
- Write speed estimation of the vehicles based on the distance of the crossing-lines.
- **Skills:** Python, Kafka, OpenCV, Flask, Models (Yolov7, DeepSORT).

Hate speech detection on Vietnamese Social Media Text on Facebook Real-time

Comments from select public Facebook pages are crawled and pushed into Spark Structured Streaming to process them. Deep learning models trained using distributed methods will predict sentiment of comments. A real-time web app dashboard will display aggregated predictions.

- Use BigDL to improve performance by combining the power of distributed clusters when training models.

- Crawl data from Facebook by Selenium.
- Use Kafka to push data into Spark.
- Handle data real-time with Spark Structured Streaming.
- Build a web-app dashboard to display statistics of data and visualize prediction.
- **Skills:** Python, Selenium, BigDL, Keras, Spark, Flask, Plotly, Models (LSTM, GRU, Text-CNN, NaiveBayes, Logistic Regression, Decision Tree).