Long Nguyen Hoang

Blockchain Data Engineer

longcnttbkhn.github.io | hoanglong180695@gmail.com | 0963207903

Experience

ARCHITECT AND FOUNDER | Chainslake.com

July 2024 – Now | Hanoi, Vietnam

- ➔ Start-up Blockchains Data Platform provides high-performance, cost-effective data processing technology solutions for customers in the blockchain field, customers have full control over the data warehouse, ensuring privacy and security.
  - Team size: 1 members.
  - Designed Data Lakehouse optimized for blockchain data analytics, supporting multiple chains (5 chains) including EVM and non-EVM.
  - Built a high-performance data processing system, ensuring high availability, scaleability, high performance, high reliability, can be deployed on both on-premise and on-cloud.
  - Optimized to update data daily, hourly or near realtime at a controllable cost.
  - Provided SDK allowing integration of analytics into other websites or sharing within our community.
  - Skills: **Hadoop**, **Spark**, **Hive**, **Delta lake**, **Trino**, **Airflow**, **SQL**, **Scala**
- ➔ To do plan: Developing an AI administrator for Lakehouse can automatically administer and operate the system with minimal developer intervention, supporting regular users so they can easily and effectively exploit data.
  - Develop AI Agent that acts as a Data Analyst, understands tables and data structures in the Data Lake, automatically writes SQL queries, and builds data analysis charts.
  - Develop AI Agent to act as Developer, can develop ETL jobs to build new data tables on Lakehouse
  - Develop AI Agents that act as Business Analysts with complex analytical and inferential capabilities, can analyze difficult user requests, plan the development of new data tables, and build insights from data.
  - Skills: **Prompt**, **RAG**, **MCP**

HEAD OF ENGINEER | DareNFT

Mar 2022 – Jun 2024 (2 year) | Hanoi, Vietnam

- ➔ Business Intelligence for Blockchain and Crypto data.
  - Team size: 4 members.
  - Built data lake infrastructure, data warehouse on-premise using **Hadoop**, **Hive**, **Delta lake**.
  - Designed and built a parallel, distributed data collection job using **Spark** to retrieve Blockchain data many times faster.
  - Designed and implemented data processing pipeline using **DBT**.
  - Deployed data analysis system on cloud environment: **AWS**, **Databricks**
  - Trained a virtual assistant (chatbot) to automatically generate SQL queries from natural language using **GPT**

Skills

BIGDATA

Technology:

Spark • Hadoop • Delta • Hive • Trino • Airflow • DBT • Kafka

Language:

Scala • Python • SQL

AWS

EMR • S3 • Glue • Databricks • EC2 • EBS • RDS

BACKEND

Java Spring MVC • Python Django • Elasticsearch • Graphql

DATABASE

Mariadb • Posgress

Education

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

Master in Data Science and Artificial Intelligence

2020 - 2022 | Hanoi, Vietnam

Engineer in Information Technology

2013 - 2018 | Hanoi, Vietnam

GPA: 3.2 / 4.0

References

Tran Viet Trung, Lecturer at School of Information and Communication Technology - Hanoi University of Science and Technology

trungtv@soict.hust.edu.vn

098 506 2706

SENIOR DATA ENGINEER | CENGROUP

Oct 2018 – Mar 2022 (3 years) | Hanoi, Vietnam

- ➔ Built AI-power Data Crawler allow crawling data from any website without predefined data structure.
  - – Team size: 3 members
  - – Trained the page classifier model using **Fasttext** to determine the target page with 90% accuracy
  - – Used **Web2text** to remove Boilerplate from the page achieve 85% accuracy.
  - – Built Information Extraction Model by using **Weak Supervised Learning** achieve 93% accuracy.
- ➔ Build real estate data analyst system
  - – Team size: 2 members
  - – 21 million records raw data were crawled from many big real estate website, were stored in **HDFS**
  - – Built ETL pipeline to pre-process raw data, filter and transform to built data analyst table.
  - – Calculated average price of real estates by area and time using **Spark**

DATA ENGINEER | GIASAN.VN

Jan 2018 – Oct 2018 (1 year) | Hanoi, Vietnam

- ➔ Designed and built a real-time real estate market monitoring system
  - – Team size: 1 members
  - – Built real time real estate data stream base on **Kafka** with throughput about several thousand records per seconds.
  - – Process a lot of user query concurrency on data stream using **Spark streaming** and sent the result to the corresponding channel for each user.