DANG NHAT

DATA ENGINEER

- 12/04/1995
- Go Vap District, HCMC
- +84 0354883673
- dangnhatsimon@gmail.com

Profile

- LinkedIn
- Github
- LeetCode

Education

Obninsk Institute for Nuclear Power Engineering
B.Sc in Nuclear Power Engineering and Thermal Physics. (2014-2019)
GPA: 4.6/5.0

MongoDB University
MongoDB Python Developer. (2024)

IBM Skills Network
Data Engineering. (2022)

DataCamp
Professional Data Engineer. (2024)
Big Data with PySpark. (2024)

Google Career
IT Automation with Python. (2021)
Data Analytics. (2021)

Working Experience

May 2024 – Present

Data Engineer Specialist
GREENFEED Vietnam Corporation - Fulltime
- GREENFEED is a fully integrated clean food chain FEED – FARM – FOOD with the mission to build quality and trusted brands that relentlessly innovate and create sustainable values for customers and society.
- ELT & Data Integration: designed, built, and tested end-to-end ELT workflows using Docker, integrating data from diverse sources such as SAP, Solomon, MISA, Porcitec, and FileSystem.
- Streaming Data Pipelines: developed real-time data pipelines using Apache Kafka and Spark Streaming, enabling low-latency processing of streaming data from IoT devices and external sources.
- Batch Data Processing: designed and executed scheduled data transformation workflows using PySpark on Apache Airflow & Spark, efficiently processing large-scale batch data.
- Big Data Platform: built from scratch a Docker-based big data platform using Dockerfile and Docker Compose, ensuring scalable and efficient deployment of Apache Spark, Kafka, and Airflow.
- Data Pipelines & Orchestration: developed and monitored robust batch and streaming pipelines using Airbyte, Airflow, Apache Beam (Dataflow), Spark Streaming, and Structured Streaming.
- CI/CD: Implemented GitLab CI/CD pipelines to automate testing, build, and deployment processes.
- Metadata Management & Data Lineage: implemented OpenLineage to track and manage data lineage, dependencies, and metadata across Airflow and Spark pipelines.
- Data Migration: managed migration and synchronization of data from internal systems to AWS services such as S3 and Redshift.
- Data Lakehouse: designed, built from scratch, and validated an AWS-based Lakehouse architecture for domains including Farm, Food, Social Media, Supply Chain, and Financial & Accounting.
- Data Transformation: leveraged DBT, Python, and Java to deliver clean, structured datasets for analytics and business intelligence.
- Social Media Data Integration: built and managed data pipelines from social media platforms and APIs to the Lakehouse, enabling deeper insights from external data sources.
- Documentation & Collaboration: maintained through documentation and collaborated closely with stakeholders to ensure timely and aligned data solutions across FEED, FARM, and FOOD.
- Skills: Python, Java, AWS, GCP, Spark, Beam, Kafka, Data Architecture, Docker, Kubernetes, GitLab, CI/CD, Integrated Businesses in Closed Loop.

DANG NHAT

DATA ENGINEER

Certificate ——————

IBM - Big Data 101.
IBM - Hadoop 101.
DataCamp - Building Data Pipelines.
freeCodeCamp - Data Analysis with Python.
Hackerrank - Problem Solving Basic.
Hackerrank - Python Basic.
Hackerrank - SQL Advanced.
Sololearn - Python Developer.
Sololearn - Python Intermediate.
Sololearn - SQL Intermediate.

Languages ——————

Vietnamese   English   Russian
Chinese

Oct 2022 –   **Data Engineer**
May 2024   PECC2 - Fulltime
             Definer - Remote Part-time
             - Definer specializes in building DevOps, automating cloud IT infrastructure, and developing analytical platforms for big data. PECC2 is a leading consultancy in the electricity energy industry, delivering power generation, transmission, and distribution solutions while leveraging data for efficiency and innovation.
             - Solution Design & DevOps: architected and implemented comprehensive solutions with a focus on automating cloud IT infrastructure and workflows.
             - AWS & CI/CD: developed task definitions and configuration files (CircleCI, Docker Compose, Dockerfile) to automate build, test, and deployment processes on AWS.
             - Containerization & Orchestration: built Docker-based environments for scalable deployment of applications and data processing workflows.
             - Data Ingestion & Processing: managed ingestion pipelines processing 1GB/day from power plant monitoring systems, ensuring data availability for Data Scientist and ML teams.
             - ETL & Data Warehousing: designed and deployed data warehouses, built ETL pipelines, and developed interactive reports to enhance customer insights.
             - Data Migration & Cloud Storage: automated data synchronization between internal servers and cloud platforms like AWS (S3, Redshift) and Google BigQuery.
             - Web Scraping & API Development: extracted and processed data from dynamic JavaScript websites with pagination, utilizing APIs for automation and real-time insights.
             - Database Management: designed and optimized relational databases with advanced SQL techniques (views, transactions, stored procedures, joins), ensuring high performance and reliability.
             - Scripting & Workflow Automation: developed Linux shell scripts, Python scripts, and cron jobs to automate catalog updates, supplier data processing, and document generation.
             - NoSQL Databases: designed and optimized large-scale NoSQL databases using MongoDB and Cassandra to handle high-throughput workloads and unstructured data efficiently.
             - Custom Tools & Monitoring: created internal tools such as a Telegram bot for real-time data retrieval, keyword search GUIs, and system health monitoring scripts.
             - Skills: Data Engineering, ETL & Data Pipelines, RDBMS, NoSQL and Big Data, Spark, Kafka, Beam, CI/CD, Python, SQL, Scripting, Docker, Git, Cloud.

Skills

Domain	Electricity, Agriculture, F&B, Inventory.
Foundation	Data Structures & Algorithms, Probability & Statistics.
Programming	Python, Java, dbt, R, LaTeX.
Framework	Flask, Django, Spring, Data Engineering.
Scripting	Bash, Shell on Linux, Git.
Database	SQL(MySQL, PostgreSQL, MSSQL), NoSQL(MongoDB, Cassandra), SAP.
Big Data	Spark, Kafka, Beam, Hadoop.
Cloud	AWS, GCP, IBM, Azure.
DevOps	Docker, Kubernetes.
Visualization	Tableau, Power BI.
Soft Skills	Detail-oriented, Structured & Analytical Thinking, Problem Solving, Teamwork, Diligent, Energentic, Resilient.