DANG NHAT

DATA ENGINEER

👤 12/04/1995

✉ Go Vap District, HCMC

📞 +84 0354883673

@ dangnhatsimon@gmail.com

Profile ————————————

in LinkedIn

Education ————————————

Obninsk Institute for Nuclear Power Engineering
B.Sc in Nuclear Power Engineering and Thermal Physics.
GPA: 4.6/5.0

Working Experience

Oct 2022 – Present	**Data Engineer**	PECC2

Contribute to Big Data Development Plan for Company. Write automation script with Python, shell scripting on Linux to help company decrease working time.

Monitoring health of server computers, ingest data from power plants to database server at Company. Cleaning, manipulating, processing data with 1GB/day (estimated), ensure data integrity, ethics and available to Data Scientist and ML Teams.

Contributing to interactive reports and dynamic dashboard will be sent streaming to the end users to help them make decisions and improve customer experience.

Core technologies: Big Data Analytics, ETL, Data Engineering, Python, Shell, Git, Linux.

Oct 2019 – Oct 2022	**Mechanical Piping Engineer**	PECC2

BIM Manager and Mechanical Lead. Thermal-Mechanical and piping engineering.

Project

IBM Data Engineering

Implement webscraping and use APIs to extract data with Python.

Create, design, & manage relational databases & apply database administration (DBA) concepts to RDBMSs such as MySQL, PostgreSQL, & IBM Db2. Compose more powerful queries with advanced SQL techniques like views, transactions, stored procedures and joins. Monitor and optimize important aspects of database performance. Backing up and restoring databases, managing user roles and permissions. Troubleshoot database issues.

Develop shell scripts using Linux commands, environment variables, pipes, and filters. Schedule cron jobs in Linux with crontab.

Develop working knowledge of NoSQL & Big Data using MongoDB, Cassandra, Cloudant, Hadoop, Apache Spark, Spark SQL, Spark ML, and Spark Streaming.

Implement ETL & Data Pipelines with Bash, Airflow & Kafka; architect, populate, deploy Data Warehouses; create BI reports & interactive dashboards.

Core technologies: Data Science, ETL & Data Pipelines, RDBMS, NoSQL and Big Data, Apache Spark, Python Programming, Data Analysis, SQL, Shell Scripting

Automation

Use Python external libraries to create and modify documents, images, and messages.

Automate updating catalog information. Fetching and working with supplier data images. Process Text Files with Python Dictionaries.

Using APIs to interact with web services. Uploading images, descriptions to web server. Write a script that summarizes and processes sales data into different categories. Generate a PDF report and send it through email using Python.

Write a script to check the health status of the system.

Core technologies: Bash, Linux, Python.

DANG NHAT

DATA ENGINEER

Languages ──────────

Project

Data Engineering
Ingest data from external sources and process data. Database Design.
Understand Resilent Distributed Datasets, using pyspark to create a data transformation pipeline, to store data in distributed file system, transformation and actions datasets.
Spark SQL for Big Data Analytics. Create, import, inspecting, cleaning, querying and saving data with pyspark.
Testing data pipeline. Manage and Orchestrate Workflows with Apache Airflow, DAG schedule.
Core technologies: Python, Spark, ETL, Airflow, SQL.

Analytics
Mean-Variance-Standard Deviation Calculator: output the mean, variance, standard deviation, max, min, and sum of the rows, columns, and elements in matrix.
Demographic Data Analyzer: analyze, manipulate data from the 1994 Census database.
Medical Data Visualizer: converting, cleaning data, explore the relationship between cardiac disease, body measurements, blood markers, and lifestyle choices.
Page View Time Series Visualizer: analyze and visualize a dataset containing the number of page views each day on the webpage, understand the patterns in visits and identify yearly and monthly growth.
Sea Level Predictor: analyze the global average sea level change since 1880, predict the sea level change through year 2050.
Combine, cleaning, manipulate, analyze and visualize to get insights from 1 millions data casual and annual membership riders. Identify trend from Cyclistic's historical bike trip data and then planning digital media could affect marketing tactics.
Combine, cleaning, manipulate, analyze the user data from FitBit Fitness Tracker to gain insights into how consumers are using the FitBit app and discover trends for Bellabeat marketing strategy.
Core technologies: Data Analysis with Python by using library Numpy, Pandas, Seaborn, Matplotlib, MySQL, R, Tableau.

Skills

Domain	Data Structures & Algorithms, Probability & Statistics.
Programming	Python, R, Java, Scala, JavaScript, LaTeX.
Scripting	Bash, Shell.
Database	SQL(MySQL, PostgreSQL), NoSQL(MongoDB, Cassandra).
Big Data	Hadoop, Spark, Kafka.
Cloud	GCP, IBM
Visualization	Tableau, Power BI.

Certificate

Sep 2023	Google IT Automation with Python	Google
Sep 2023	Problem Solving (Basic) Certificate	HackerRank
Aug 2023	Building Data Engineering Pipelines in Python	DataCamp
Aug 2023	Data Analysis with Python	freeCodeCamp
Aug 2023	Python Intermediate	Sololearn
Aug 2023	SQL Intermediate	Sololearn
Jul 2023	Big Data 101	IBM
Jul 2023	Hadoop 101	IBM
Jul 2023	Python (Basic) Certificate	HackerRank
Jul 2023	SQL (Intermediate) Certificate	HackerRank
Jan 2023	Google Data Analytics	Google