DANG NHAT

DATA ENGINEER

👤 12/04/1995

✉ Go Vap District, HCMC

📞 +84 0354883673

@ dangnhatsimon@gmail.com

Profile

in LinkedIn

○ Github

Skills

Basic:

Data Structures	●●●●○
Algorithms	●●●●○
Prob & Stats	●●●●○

Programming:

Python, R, LaTeX	●●●●○
JavaScript, Scala	●●●○○

Database:

MySQL, SQL Server	●●●●○
MongoDB	●●●○○

Big Data:

Hadoop, Spark	●●●●○

Machine Learning:

numpy, pandas, polars	●●●●○
polars, tensorflow	●●●○○
scikit-learn	●●●○○

Data Visualization:

Tableau, Power BI	●●●●○

Detail-oriented   Structured thinking

Analytical thinking   Hard working

Teamwork   Data Storytelling

Working Experience

Oct 2022 – Present

Data Engineer   PECC2

Contribute to Big Data Development Plan for Company. Hadoop, Spark for Data Analytics. Running automation with Python, Bash scripting on Ubuntu. Version control system like Git. Monitoring health of server computers.

Getting data like power output, weather, current and voltage from site, through pipelines to database server at company, after processing and manipulating, data will be analyzed, visualized in apps. All reports and dynamic dashboard will be sended streaming to the end users to help them improve experiences, making decisions.

Core technologies: Big Data Analytics, ETL, Data Engineering, Python,Bash, Git, Ubuntu.

Oct 2019 – Oct 2022

Mechanical Piping Engineer   PECC2

BIM Manager and Mechanical Lead.

Thermal-Mechanical and piping engineering.

Project

Aug 2023 – Aug 2023

Automation   **Data Engineer**

Scale and convert images using PIL.

Process Text Files with Python Dictionaries and Upload to Running Web Service.

Automatically Generate a PDF and send it by Email.

Automate updating catalog information.

Core technologies: Bash, Linux, Python.

Aug 2023 – Aug 2023

Data Engineering   **Data Engineer**

Importing data from external sources (SAS,STATA,HDF5,MAT,..) to python and using numpy, pandas, matplotlib, seaborn to process data.

Database Design.

Use PySpark to Create a Data Transformation Pipeline.Understand Resilent Distributed Datasets (RDDs) of Spark, using pyspark 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.

Jul 2023 – Jul 2023

freeCodeCamp   **Data Analyst**

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: visualize a dataset containing the number of page views each day on the freeCodeCamp.org forum from 2016-05-09 to 2019-12-03, understand the patterns in visits and identify yearly and monthly growth.

Sea Level Predictor: analyze a dataset of the global average sea level change since 1880, use the data to predict the sea level change through year 2050.

Core technologies: Data Analysis with Python by using library Numpy, Pandas, Seaborn, Matplotlib.

DANG NHAT

DATA ENGINEER

About Me ——————

I am a results-driven Data Engineer with a passion for creating efficient data pipelines and optimizing data workflows. My expertise lies in designing and implementing robust data warehouses, modeling data structures for optimal performance, and integrating diverse data sources. With hands-on experience in data transformation and real-time processing, I am adept at turning raw data into actionable insights.

Languages ——————

Language	Level
Vietnamese	● ● ● ● ●
English	● ● ● ○ ○
Russian	● ● ● ● ○
Chinese	● ● ○ ○ ○

Project

Oct 2022 –
Oct 2022

Cyclistic Bike Share   **Data Analyst**

Using MySQL, R and Python to 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.

Core technologies: MySQL, Python, R, Tableau.

Oct 2022 –
Oct 2022

FitBit Fitness Tracker   **Data Analyst**

The user data from FitBit Fitness Tracker has been merged and categorised into different sections; daily activity, daily calories daily intensities, daily steps, etc. Using MySQL, R and Python to combine, cleaning, manipulate, analyze FitBit fitness tracker data to gain insights into how consumers are using the FitBit app and discover trends for Bellabeat marketing strategy.

Core technologies: MySQL, Python, R, Tableau.

Education

Undergraduate Study

2014 – 2019

B.Sc. in Nuclear Power Engineering and Thermal Physics   Russia

Title: Analysis of the transient mode in the VVER-1000 core when two main circulation pumps are switched off at rated power.

Supervisors: Prof. Leskin Sergei Terent'evich.

Knowlegde: Nuclear Power system and Nuclear Reactor, especially WWER-1000, Steam Generator and Heat Exchangers, Steam Turbine, Industrial Pump, Nuclear reactor safety systems, Valves

GPA: 4.6/5.0

Certificate

Date	Certificate	Provider
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	Python (Basic) Certificate	HackerRank
Jul 2023	SQL (Intermediate) Certificate	HackerRank
Jan 2023	Google Data Analytics	Google
Jun 2023	Google IT Automation with Python	Google

References

Ref. 1

Huynh The Minh, Data Engineer
Parcel Perform
htminh2307@gmail.com

Ref. 2

Nguyen Nhat Nam, Data Scientist
CEL
nathan.nguyennhat@gmail.com