



# DANG NHAT

## DATA ENGINEER

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

## Profile



## Education

### Obninsk Institute for Nuclear Power Engineering

B.Sc in Nuclear Power Engineering and Thermal Physics (2014-2019).

GPA: 4.6/5.0

### IBM Skills Network

Data Engineering (2021-2022).

### DataCamp

Data Engineer (2021-2022).

### Google Career

IT Automation with Python (2021-2022).

### Google Career

Data Analytics (2021-2022).

## Certificate

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

## Working Experience

Oct 2022 – Present

### Data Engineer

PECC2

Contribute to Big Data Development Plan for Company. Writing 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

Web scraping and using APIs to extract data with BeautifulSoup, Selenium, Scrapy.

Creating, designing, & managing relational databases & applying database administration (DBA) concepts to RDBMSs such as MySQL, PostgreSQL, & IBM Db2. Composing more powerful queries with advanced SQL techniques like views, transactions, stored procedures and joins. Monitoring and optimizing important aspects of database performance. Backing up and restoring databases, managing user roles and permissions. Troubleshooting database issues.

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

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

Implementing ETL & Data Pipelines with Bash, Airflow & Kafka; architecting, populating, deploying Data Warehouses; creating 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.

### py-frame

Creating a Telegram Bot with python-telegram-bot that can do API call to coingecko and get real time price of coin.

Creating search function for searching specific keywords across multiple files combined with simple GUI using tkinter.

Web scraping with Scrapy, Splash, Selenium on real estate and e-commerce websites.

Core technologies: Python with frameworks.

# DANG NHAT

DATA ENGINEER

## Languages

Vietnamese

English

Russian

Chinese

## Project

### DataCamp

Ingesting data from external sources with many of different file types. SQL Database Design.  
Cleaning data, ensure data integrity, completeness, uniformity.  
Using record linkage technique to merge multiple datasets together, used when values have typos or different spellings.  
Writing efficient, timing and profiling code. Pandas optimizations.  
Streamlined data ingestion with pandas and data visualization.  
Git, Cloud. Software Engineering Principles in Python.  
Understanding Resilient 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. Manipulate, clean, query data with pyspark.  
Testing data pipeline. Manage and Orchestrate Workflows with Apache Airflow, DAG schedule.  
Core technologies: Python, Spark, ETL, Airflow, SQL, Git.

### Big Data with Spark

Using the Spark Python API, PySpark, leveraging parallel computation with large datasets, and getting ready for high-performance machine learning.  
From cleaning data to creating features and implementing machine learning models, execute end-to-end workflows with Spark.  
Getting data into Spark and then delving into the three fundamental Spark Machine Learning algorithms: Linear Regression, Logistic Regression/Classifiers, and creating pipelines.  
Data wrangling and feature engineering with PySpark.  
Building recommendation engines using Alternating Least Squares in PySpark.  
Core technologies: Big Data, ETL & Data Pipelines, Apache Spark, SparkSQL, SparkMLlib, Python Programming, Java Programming, SQL, Shell Scripting.

### Automation

Using 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.  
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.  
Writing a script to check the health status of the system.  
Core technologies: Bash, Linux, Python, Shell Scripting.

## Skills

Domain  
Programming  
Framework  
Scripting  
Database  
Big Data  
Cloud  
Visualization  
Soft Skills

Data Structures & Algorithms, Probability & Statistics.  
Python, R, Java, LaTeX.  
Flask, Django, Spring.  
Bash, Shell with Linux, Git.  
SQL(MySQL, PostgreSQL), NoSQL(MongoDB, Cassandra).  
Hadoop, Spark, Kafka.  
GCP, IBM.  
Tableau, Power BI.  
Detail-oriented, Structured & Analytical Thinking, Problem Solving, Teamwork, Diligent, Energetic, Resilient.