XUAN-QUANG NGUYEN
Data Engineer

Date of birth: 17 August 1994

Address: S1002 Vinhome Grand Park, Thu Duc, Ho Chi Minh city

Phone: (+84) 938 643 547

Email: quangnx.engineer@gmail.com

Skype: nxquang178

Linkedin: Xuan-Quang Nguyen

Profile

A skillful software engineer working in data criteria including ETL, Big data and Backend

Build environment and infrastructure to handle data flow, analyze data and visualize on automatic report

Working hard in pressure, a striver always moving forwards to become an expert in big data, data analysis

Core Competencies

- **Programming language:** Python, JavaScript, PHP, Golang, Java
- **Libraries**: Numpy, Matplotlib, Pandas, Scikit-Learn, Request, Json, Re, TensorFlow
- **Data Science & Machine learning**: Data Preprocessing, Data Visualization, Logistic Regression, ANN, CNN, Clusterings
- **Data stacks**: DataLake, Data WareHouse, ETL, Data Pipeline, HDFS, MapReduce, Spark, Airflow, Scala, Kafka, Dataiku, ChatGPT Prompt Genius, Tableau, DBT
- **Database**: MySQL, PostgreSQL, MongoDB, T-SQL
- **Cloud**: Google Cloud (BigQuery, Firebase), AWS (RedShift, Glue, S3, ECS, ECR, Lambda, Sagemaker), Azure (Landing Zone, Machine Learning, storage Gen 2, Synapse Analytic)
- **CICD stacks**: docker, docker-compose, Azure DevOps, CircleCI

Work experience

03/2023 - Present: **Data Engineer** *at FPT Software Hanoi*

- **Description:** This is a traveling platform focused on the user experience. It needs to improve the usage performance, detect fraud, product risk, product & user ranking
- **Technologies:** Python, Lambda function, S3, Airflow, Glue, Spark, Redshift, Google Big Query
  - Implement complex Airflow DAGs based on several operators: SensorOperator, PythonOperator, BranchOperator, EcsOperator. Also, define custom operators to adapt variety of processing logic
  - Implement the data processing jobs via Spark running in AWS Glue. Also, leverage dynamic frame to familiarize the Glue philosophy. On besides, several Spark transformations, actions and cache are applied intending optimize performance
  - Evaluate fraud detection model efficiency based on determining metrics precision, recall, f1 score, false positive rate, true positive rate

- ○ Compose queries in PostgreSQL and define the store procedures of Google Big Query to extract valuable data from database for automatic reports

- ○ Write dockerfile, set up ECS, define task definition to deploy Airflow

- ○ Define several ChatGPT prompts to populate the web page's contents, collect destinations, POI and set up mechanism to validate ChatGPT returned data

- ○ Design data warehouse applying both Star Schema and Snowflake for each context. Mostly use Slow Change Dimension Type 2 to track data history.

- ○ Define DBT model to populate data for metric tables in Data Warehouse

07/2021 - 02/2023 **Data Engineer** *at Vingroup JSC*

- ● **Description:** Develop the data platform for Vinfast electrical cars: VF8, VF9, VFe34. This platform supports ADAS validation by collecting, transform huge CAN, LIN, Rtag data from road test

- ● **Technologies**: Python(Numpy, Matplotlib, Pytorch, Django), Golang, T-SQL, MongoDB, MySQL, Docker Kubernetes, Airflow, Azure

- ● **Responsibilities:**

- ○ Build ETL and data pipeline by defining Airflow DAG with KubernetesPodOperator to extract, transform a huge amount of semi-structured and unstructured data from cars. Integrate Azure Kubernetes Services with Virtual Machine Scale Set to achieve the scalable infrastructure where Airflow KubernetesPodOperator performs well on it.

- ○ Define retention policies on Azure Storage Gen 2 to optimize the storage cost. Define data governance policies to control data access through services

- ○ Design and construct the Azure Landing Zone where multiple data platforms throughout many regions works smoothly throughout many regions

- ○ Implement the backend for the data platform with Golang, NodeJS, Python, MongoDB, MySQL, T-SQL. Also apply authentication with Azure Active Directory, Azure API management

- ○ Write Python script attaching a long complex MongoDB query to extract data from MongoDB for reporting and analyze car's features performance

- ○ Design the CICD workflow and compose yml files to define the full flow of app building and deployment via Azure Devops

08/2020 - 06/2021: **Data Engineer** *at FPT software HCM, Vietnam*

- ● **Description:** Build ETL to provide data for a visualization tool to optimize the supply chain operations of grid factories in China, Europe, North America

- ● **Technologies:** AWS Redshift, Greenplum, PostgreSQL, NodeJS, Python(numpy, pandas), Shell script

- ● **Responsibilities:**

- ○ Set up batch-based data ingestion flow to transport data from source such as CSV files, AWS Redshift, SQL server to inner Data Warehouse by dblink, Python pandas.

- ○ Implement multiple view layers for data processing, each view has its own business logic which is implemented by SQL. The data goes through multi views as processing in pipeline and finally stores in materialized views
- ○ Implement the API with Nodejs, Redix caching to provide insight from materialized views for UI visualization with minimum latency
- ○ Apply Dataiku to make recipes for ETL, visualize data following customer's business logic

07/2018 - 04/2020: **Software engineer** *at Vitalify Asia, Vietnam*

- **Description:** This is a Japanese software offshore company. Therefore, I joined multiple projects like bitcoin exchange service, real estate growth prediction (by Tensorflow) and stock management tool
- **Technologies:** Python(Pandas, Numpy), Typescript, PHP(Laravel), Docker, Google Firebase
- **Responsibilities:**
  - ○ Develop new features in both backend & frontend for platforms: food orders management and bitcoin transferring service via PHP Laravel, MySQL, JQuery, HTML, CSS, Angular
  - ○ Apply GCP Firebase, Compute Engine intending tuning model to tweak the hyper parameter on weekly training data

01/2018 - 06/2018: **Data engineer** *at Sendo.vn*

- **Description:** This is an ecommerce company, working in BI team in order to build the platform generating the automatic reports for the departments. Therefore, business leaders can give well informed decision in marketing, stock, customer feedbacks, orders from data insight
- **Technologies:** Python(Numpy, Pandas, Matplotlib), Google Cloud services (Compute Engine, BigTable, BigQuery, Data Studio)
- **Responsibilities:**
  - ○ Analyst department's requirements, determine the necessary metrics for visualization
  - ○ Implement Python scripts to collect and extract data from multi sources(internal DB, APIs, Google sheet, raw data in blob storage) and then use bash scripts to run the batch processing of data
  - ○ Define Google Big Query views to extract insight, optimize performance by partitioning data
  - ○ Improve query performance by partitioning data
  - ○ Build automatic report by Google App Script, Google BigQuery for headcount planning
  - ○ Use Google Data Studio to set up report containing charts visualizing data from final tables

08/2016 - 12/2017: **Full stack developer** *at Not A Square*

- **Description:** Build applications for an Austrian gene laboratory. These applications help them to manage their resources information, support parsing publications data and control whole operations
- **Technologies:** Python(Numpy, Pandas, Django, nltk), PHP(Symfony), Javascript, Docker
- **Responsibilities:**
  - ○ Implement both backend and frontend based on PHP - Symfony, Python, Javascript

- ○ Build NLP model by N-gram, base regex to detect language and extract data from text
- ○ Discuss directly with customers for new features, write detail design, implement, test and deploy

Education

10/2012 - 06/2018:   **Bachelor Information Technology** *at Vietnam National University - Ho Chi Minh city, University of Science*

Language

- **English**: Well communicate, read, write technical document
- **French, German**: Basic communicating and reading

Certificate *(click icons to view detail)*

Volunteer

10/2012 - 10/2015: **Chairman** *at Youth Union in the Faculty of Information, University of Science.*

- Leader in organizer of academic competitions, entertainment activities for student