

PHU D NGUYEN

nphu01@vt.edu | (540) 988-6514 | Blacksburg, VA 24060 | [LinkedIn](#) | [GitHub](#)

EDUCATION

B.S in Computer Science

Expected Graduation: May 2024

- **Virginia Tech** (Blacksburg, VA)
- **Programming Languages:** Java, Python, Golang, C Programming
- **Technologies:** Natural Language Processing, Apache Hadoop, Spark, Kafka, Airflow, Docker, Kubernetes, SQL (PostgreSQL, MySQL), Git Version Control
- **Certifications:** [HackerRank Java \(Basic\) Certificate](#), [HackerRank Problem Solving \(Basic\) Certificate](#)
- **Relevant Courses:** Data Structure & Algorithm, DBMS & SQL, Machine Learning, Computer Systems, Mobile Software Development, Applied Combinatorics, Statistics for Engineer

WORK EXPERIENCE

Data Engineer | ISTT Co, Ltd

May 2023 – Oct 2023

- Developed a robust, end-to-end Natural Language Processing SMS anti-spam system using Python, PySpark, BERT and Ensemble Learning, achieving **90% accuracy on test data**.
- Architected a multi-stage pipeline system using Python, MySQL, Airflow, visualizing data using Superset and Matplotlib. The system accelerated project performance, achieving **a 5-fold increase in processing speed**.
- Deployed a highly scalable ETL pipeline using Python, Kafka, and Hadoop, capable of extracting data from diverse external sources and efficiently transforming and loading it into Amazon S3.

Software Engineer Intern | ISTT Co, Ltd

August 2021 – December 2021

- Implemented a scalable API using Golang that enhanced Spoofing's detection capabilities, analyzing and processing over 1 million call records, led to **a 95% detection rate improvement** in identifying fraudulent calls.
- Improved the micro-service system architecture by deploying and managing **over 100 distinct tasks** using the JHipster platform.
- Assisted a team of 6 senior developers in implementing REST API with the Spring Boot framework and optimized data accessibility by creating API endpoints to read data from SQL databases.

PROJECTS

[JobScraper](#)

July 2023 – Present

- **Technologies:** Python, Flask, MySQL, JavaScript, HTML/CSS
- Developed an aggregator platform for job listings from LinkedIn, indeed, and ZipRecruiter using Python libraries.
- Stored and managed aggregated data by using MySQL as the database management system.
- Features a user-friendly UI using Flask that allows real-time queries, customizable search criteria, and a dynamic result display.

[YourFlight](#)

April 2023 – Jun 2023

- **Technologies:** Python, PySpark, Airflow, PostgreSQL, Elasticsearch, Logstash, Kibana
- Programmed a flight price tracking application that scrapes real-time data from Kayak and Google Flights.
- Processed data using PySpark, orchestrated data collection and transformation workflows using Airflow to streamline the process.
- Used PostgreSQL as the DBMS, then stored on AWS S3. Analytics and logs are handled using ELK stack (Elasticsearch, Logstash, and Kibana).

[Subscription Streaming Service Website](#)

Aug 2022 – Oct 2022

- **Technologies:** Python, Tailwind CSS, Django
- Designed a user-friendly streaming service website using Python Django, with robust user authentication, email validation, and password security features.
- Improved website aesthetics and responsiveness using Tailwind CSS, enhancing the overall user experience.
- Integrated SQLite database with Django for dynamic data querying and display.