Tuan Kiet Nguyen
Data Engineer

Tel. (*Zalo*): 0368516377 / Email: kiet.nguyentuan.1652000@gmail.com / Github / Linkedin

EDUCATION

ITMO University, Saint Petersburg, Russia   **2019 - 2024**
Major: Information Systems and Technologies   GPA: 3.95/5

WORK EXPERIENCE

Intern Data engineer   **05/2023 - 08/2023**
Faculty of Information Technology and Programming, ITMO University, Russia
Details: Complete the tasks given by the instructor in building ETL.

PROJECTS

Real Time-streaming-Vehicle-Speed-Pipeline
- **Objective**: Developed a real-time data pipeline to monitor and analyze vehicle speed data.
- **Technologies**: Apache Kafka, Apache Spark, PostgreSQL, HDFS, Python, Docker.
- **Details**: Streamed data from Kafka, processed it with Spark Streaming, stored raw data in HDFS, and loaded processed results into PostgreSQL for analytics.

Brazil-E-commerce-ELT-Pipeline
- **Objective:** Designed and implemented data warehouse and ELT pipeline for Brazilian e-commerce data.
- **Technologies:** Apache Spark, Apache Hive, HDFS, MySQL, Apache Airflow, Docker.
- **Details:** Extracted data from MySQL, stored it in a data lake on HDFS, processed it with Scala, and loaded it into Hive for reporting and analysis.

Youtube-data-pipeline-with-aws-and-airflow
- **Objective:** An automated pipeline that collects trending Youtube video data from multiple countries.
- **Technologies:** AWS S3, AWS Lambda, AWS Glue, AWS Quicksight, Apache Airflow, Docker.
- **Details:** An automated pipeline collects trending video data from multiple countries via the YouTube API, managed by Airflow. Data is stored in an S3 data lake, cleaned by Lambda, processed in AWS Glue, and made ready for analysis in QuickSight.

Football livescore.com crawling
- **Objective:** This project collects data like match details, scores, and statistics and stores them in JSON files for further analysis or use in other applications.
- **Technologies:** Python, Selenium.
- **Details:** A Selenium-based Python project to automate the extraction of match information and statistics from the Premier League section of the Livescore website.

Tiki product price ETL
- **Objective:** Developed a pipeline to extract and store daily product prices and reviews from Tiki.vn.
- **Technologies**: Apache Airflow, Apache Spark, PostgreSQL, Python, Docker.
- **Details:** Automated daily data extraction, transformation, and loading using Airflow and Spark to populate a PostgreSQL data warehouse.

Other projects:
- **[Crypto k-lines data_ETL pipeline_with_Spark and Airflow]()**
- **[Backend of job Searching web application with NestJS]()**
- **[Backend of football team management with java spring boot]()**

SKILLS

Technical:
- **Languages:** Python, Pyspark, Java, SQL.
- **Technologies:** Hadoop, Spark, Kafka, Airflow, AWS.
- **Others:** Data crawling, Data Streaming, Data Modeling, ETLs.

Foreigner languages:
- **English:** Communication, reading and understanding documents.
- **Russian:** Communication, reading and understanding documents.