

# TRINH DINH PHUC

## Data Scientist

🌐 [thebigdatascientists.com](https://thebigdatascientists.com)

@ [phuctrinh@soongsil.ac.kr](mailto:phuctrinh@soongsil.ac.kr)

☎ +82.10.9934.9876

📍 Seoul, South Korea



## EXPERIENCE

### Data Scientist



#### POPS WORLDWIDE

📅 Aug 18 – Feb 19 (7 months) 📍 HCMC, Vietnam

- Led POPS's data team, including one business analyst, one data analyst, and one system administrator.
- Designed and built a complete data pipeline for a finance department with  $\approx 35\text{M}$  records per month (2TB in total) on Google Cloud Platform (GCP), including ETL, data warehouses, analytics, and BI.
- Designed and built various reporting dashboards using Power BI, retrieving data from a data warehouse on GCP.
- Created a tool that automatically downloads monthly reports through YouTube Reporting API using Python (saved \$2800 per month on buying reports).
- Migrated on-premises data to GCP and built internal tools: web crawlers, parsing tools using Python.

### Machine Learning Engineer



#### CBD Robotics

📅 Feb 18 – Feb 19 (1 year) 📍 HCMC, Vietnam

Built various ML-based products with a team:

- Built a deep-learning-based [lung cancer detection solution](#) using X-ray images collaborated with Cho Ray hospital.
- Built a [CBD diagnosis website](#) integrated with a trained detection model using Flask.
- Taught various data science problems to trainees: NLP classification, clustering, and sentiment analysis at [CBD bootcamp](#).

### Web Developer Intern



#### DevNet [start-up company]

📅 Apr 17 – Aug 18 (5 months) 📍 Nha Trang, Vietnam

- Developed modules and maintained company's front-end website using HTML, CSS, and PHP.

## TECHNOLOGY AND SKILLS

- Programming Languages: Python, Scala, Java, R, DAX, C++.
- Databases: Oracle SQL, MySQL, SQL Server, PostgreSQL.
- Big data: AWS Services (EC2, S3, EFS, RDS, DynamoDB, Aurora, CloudWatch, SageMaker, EMR, Kinesis, ElasticSearch, Glue, S3DistCP, Athena, Redshift), BigQuery, Google Cloud Engine, Apache Kafka, Apache Spark, Apache Hadoop, Apache Parquet, Zookeeper, Cloudera.
- ML framework: Scikit learn, Keras, Tensorflow.
- BI tools: Power BI, Tableau.
- Other: Git, Jupyter, Stack.
- Language: English (Advanced), Korean (Basic), Vietnamese.

## EDUCATION

### Master of Engineering



[Soongsil University](#) | GPA: 4.13/4.5

📅 Mar '19 – Feb '21 📍 Seoul, South Korea

### B.Eng in Computer Science



[Telecommunications University](#) | GPA: 3.34/4

📅 Sep '14 – Aug '18 📍 Nha Trang, Vietnam

## NOTABLE PROJECTS

#### Covid-19 Research project - [\[nature.com\]](#)

- Developed and implemented a full data-science workflow (excluding data collection).
- Developed and evaluated ML models with ICE, PDP, SHAP, LIME, H-statistic.
- Deploying proposed scheme on the AWS cloud platform (ongoing).

#### Design a big data framework for SDN-NFV paradigm [\[pub:#1\]](#)

- Designed and developed a big data framework for a large-scale SDN-NFV paradigm.

#### Compromised switch & Economic DoS in SDN applied deep learning projects in collaboration with KAIST [\[pub:#2-#6\]](#)

- Proposed various intelligent-based schemes using deep learning techniques to detect abnormal behaviors in multivariate time-series.

## PUBLICATIONS

- T. D. Phuc, Minho Park, "BDF-SDN: A Big Data Framework for DDoS Attack Detection in Large-Scale SDN-Based Cloud," IEEE DSC, Japan, Jan 2021.
- T. D. Phuc, Minho Park, "ECSD: Enhanced Compromised Switch Detection in an SDN-Based Cloud Through Multivariate Time-Series Analysis," IEEE ACCESS - [Q1 - SCIE].
- T. D. Phuc, Minho Park, "Economic EDoS Detection using GANs in SDN-based Cloud," IEEE ICCE, Phu Quoc - Vietnam, Jan 2021.
- T. D. Phuc, Minho Park, "Dynamic EDoS Detection in SDN-based Cloud" IEEE FMEC, Paris, July, 2020.
- T. D. Phuc, Lee Tae-Hee, Nguyen Canh Thang, Pham Dang Sa, Minho Park, "Abnormal SDN switches detection based on chaotic analysis of network traffic", 38th IEEE APCC.
- Kim D., T. D. Phuc, Noh Sinchul, Minho Park, "An Effective Defense Against SYN Flooding Attack in SDN" IEEE ICTC, Jeju, Nov, 2019.