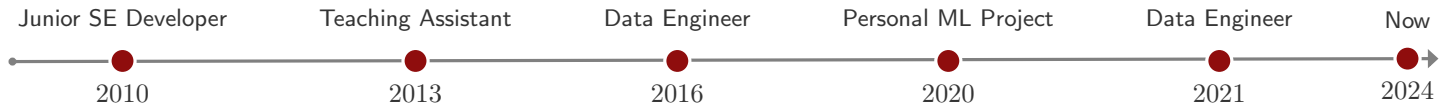


DUY LAM VO

Data Engineer (7+ years)

@ duy-lam.vo1102@outlook.com
in duylamvo

📍 Berlin, Germany
📱 duylamvo



EXPERIENCE

Data Engineer

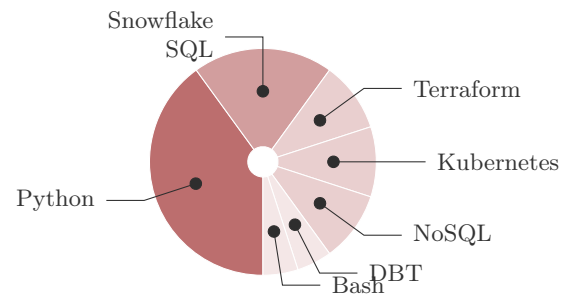
Wefox

📅 June 2021 – Now

📍 Berlin, Germany

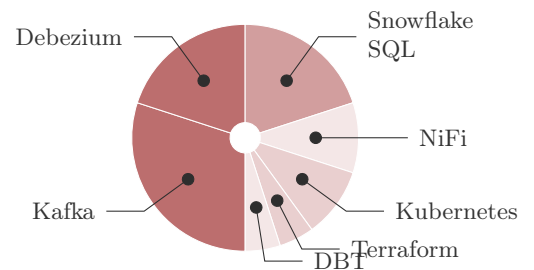
◀ Build and maintain Batch Processing platform

- Profoundly, design and develop generic Airflow-based ingestion framework to standardize the process of ingestion to data lake using yaml file.
- Write sql-based stored procedures to load data from S3 to Snowflake via Snowflake Task and Snowpipe automatically
- Create automation pipeline via Github Actions to trigger Terraform and deployment
- Design and implement data modelling for stakeholders via DBT.
- Ingest different data sources including BigQuery, Google Analytics, MySQL, Postgresql, Oracle, DocumentDB (MongoDB), Excel, Google Spreadsheet, HTTP Restful, APIs using our own framework.
- Monitor logs and events via Prometheus and Elastic Search



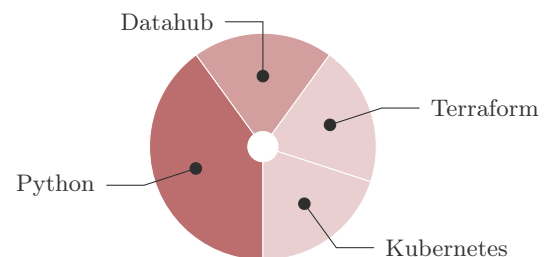
◀ Build and maintaining Streaming Processing platform

- Build streaming platform to capture CDC data via Debezium to Kafka (MSK)
- Design and deploy unified ingestion Kafka-based framework
- Design workflow in NiFi to consume messages from Kafka and backup data to S3, then load to Snowflake
- Design Streams and Tasks in Snowflake to upsert data to Snowflake and convert CDC data to relational tables.
- Ensure the schema consistency between source and target
- Automate creation and updating schema in Snowflake table when new column added
- Monitor events produced and consumed via Prometheus and error logs via Elastic Search



◀ Deploy Data Discovery and Lineage platform

- Build data dictionary and governance platform based on [Datahub](#) to support data discovery in the company
- Write python-based custom pipeline for Datahub to load missing data lineage from source to be available in the platform.
- Ensure the visualization of data lineage from source to target not be broken
- Write Terraform and K8S to deploy it on AWS EKS



Data Engineer

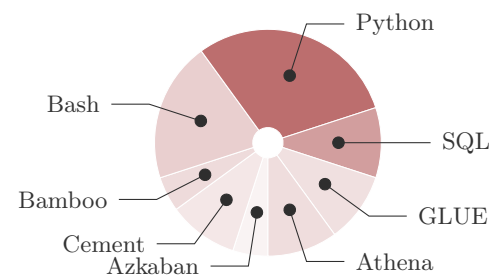
Avira (Data Intelligence)

◀ Ingest Data from Advertisement Platforms to S3

- Fetch data from Google Ads, Facebook, Bing and Criteo Ads
- Fetch data from Google Playstore and Apple Appstore
- Pipelines from Ad Engines → S3 → AWS Athena → Tableau
- Automation via Azkaban, Lambda, Python, and Bash scripting

◀ Write command line interfaces (CLI) for Python

- Develop interfaces using Cement Framework
- Support data analysts fast query and upload data.



◀ Develop a Marketing Alert System

- Monitor sales and send emails about drops in sales via R-language.
- Supports marketing, managerial team for detecting drops.

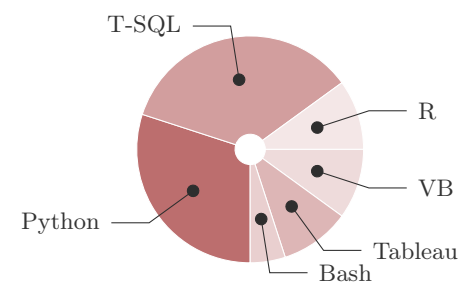
◀ Develop Auto Report Generator

- Automation scripts to generate PowerPoint from Excel via VB.
- Helps to reduce time in reporting.

◀ Ingest Data from Google Analytics to AWS S3

- Pipelines from → Big Query → AWS S3
- Automation via Google Data Flow, Apache Beam, and Bamboo

◀ Design dashboards and data sources in Tableau



TECHNICAL SKILLS

Programming Language

Python, Bash, SQL, NoSQL, VB, Terraform

Backend Frameworks

Cement, AWS Boto3, Restful, OAuth2, Pytest, Unittest, Kubernetes

Database

DBT, Snowflake, MSSQL, Athena, BigQuery, InfluxDB, Postgresql, MySQL, MongoDB, Grafana, Prometheus, Elastic-Search

Data Engineer Frameworks

NiFi, Kafka, Airflow, Apache Beam, Data Flow, Dask, Lambda, PySpark

CI/CD Tools

Bamboo, Docker, Github, Bitbucket, Github Actions

Data Science

Scipy, Scikit-learn, Numpy, Pandas, Tensorflow, Pytorch, Dask

Data Visualization

Matplotlib, Plotly, Dash, Bokeh, Tableau

EDUCATION

M.S. in Computer and Data Science

📅 Apr 2017 – Mar 2021

📍 Konstanz University

B.S. in Business Informatics

📅 Oct 2011 – Oct 2016

📍 Südwestfalen University

B.S. in Software Engineering

📅 Oct 2003 – Oct 2009

📍 HUTECH University

PUBLICATIONS

◀ **TS-MULE**: Local Interpretable Model-Agnostic Explanations for Time Series Forecast Models

ACHIEVEMENTS

🔗 **Increase Productivity Time**
data is mine, focus in your good job

⚙️ **Get rid of human error-prone excels**
and move to auto dashboards

📈 **Support colleagues**
with auto alert on unexpectedly low revenue

STRENGTHS

⌚ **I can work hard**
for team's objectives.

💬 **I would collaborate**
with my colleges and learn from them.

💓 **I am ready**
to take risk re-sharpening skills for new objectives.

🏆 **I train myself**
to keep positive thinking in making decisions.

LANGUAGE

Vietnamese ● ● ● ● ●

English ● ● ● ● ●

German ● ● ● ● ●

INTEREST

📖
Reading

💡
Science

</>
Open Source

⚙️
Swimming

⚽
Running

📺
Movies