

VO DUC DAN

Binh Thanh district , Ho Chi Minh city

0972184325 voducdand99@gmail.com <https://voducdan.github.io/cv/>

SUMMARY

As a Data Engineer, I am passionate and dedicated to my work. I am committed to continuously improving my skills and expertise. Leveraging a solid foundation in the Data Engineering sector, I am confident in contributing to project success by delivering robust and efficient data solutions. I find joy in every aspect of the design process, from initial discussions and collaboration to concept development and execution. My greatest satisfaction comes from seeing the final product achieve its intended goals.

EDUCATION

HCM University of Science

8/2017 - 10/2021

Bachelor Information Technology

GPA: 7.3/10

WORK EXPERIENCE

MoMo

7/2022 - now

Data Engineer

Main responsibilities:

- Operate and scale Airflow on Kubernetes to serve over 3,000 DAGs, ensuring tasks meet their SLA by running on time. Guarantee DAG files are updated and reflected on the Airflow UI with a P90 latency of less than 15 seconds. Leverage advanced Airflow features such as Pools, Queues, Priority Weights, and Cluster Policies to optimize resource utilization and isolate workloads across multiple teams. Set up CI pipelines using Gitlab CI to test and sync DAG code from GitLab to Airflow, and build a development environment for users to validate their code before deployment.
- Developed a tool to support batch ingestion from various data sources using Spark on K8S. Took part in all stages of the development lifecycle, including ideation, coding, infrastructure setup, and UI development. Implemented solutions such as Kubernetes Admission Control and secret management to monitor and secure data ingestion jobs, ensuring data security and access control.
- Proposed and implemented a new solution to increase the report update frequency for critical requests from once per day to every 5 minutes by setting up ClickHouse Kafka Connect to stream data into ClickHouse. Leveraged ClickHouse features such as the ReplacingMergeTree engine, materialized views, materialized columns, and function combinators to deduplicate data and perform pre-aggregation, significantly reducing the load on transformation processes and dashboard queries.
- Designed and developed a lakehouse system based on the medallion architecture using Spark, Iceberg, Hive Metastore, and Google Cloud Storage (GCS).
- Managed infrastructure on Kubernetes in an Infrastructure-as-Code (IaC) approach using Pulumi.

Technologies: GCP, Kubernetes, Spark, Airflow, Clickhouse

Amanotes

6/2021 - 7/2022

Data Engineer

Main responsibilities:

- Designed data models for event data from games
- Crawled/pulled data from partners' APIs/Dashboards
- Applied DBT for transforming raw data from firebase

Technologies: Python, GCP, Airflow, DBT, Docker

FPT Software

1/2021 - 7/2021

Data Engineer

Main responsibilities:

- Created data pipelines to extract, transform and load large datasets from csv files into the data warehouse
- Designed and developed data models
- Participated in building the infrastructure required for optimal extraction, transformation, and loading processes

Technologies: Python, GCP, Airflow

Asia Commercial Bank (ACB)

7/2020 - 1/2021

Data Engineer Collaborator

Main responsibilities:

- Maintained data infrastructure
- Built data lineage to quickly identify the source tables of reports
- Crawled and enriched data for analytical purposes

Technologies: Spark, Hadoop, Python, SQL, Oracle

ADDITIONAL**Technical:** Spark, SQL, Data Modeling, Clickhouse, GCP, K8S, Docker, Airflow, Linux, Git**Programming Skills:** Python, Javascript, OOP, Data Structures and Algorithms**Language:** Fluent in English**Certification & Training:** [ClickHouse Certified Developer](#), [HackerRank SQL \(Advanced\) Certificate](#), [GCP Professional Data Engineer](#)**References:** [Github](#)