



CONTACTS

✉ thaibatuan.amu@gmail.com
☎ +33 6 25 33 26 07
🌐 github.com/batuan
🌐 linkedin.com/in/batuan97
🏠 Ivry Sur Seine
🌐 thaibatuan.com/about

COMPÉTENCES

PROGRAMMATION

Python, Java, C/C++, Swift, Scala,
JavaScript

DATA ENGINEER OUTILS

- Hadoop Ecosystem: HDFS, Hive
- Apache Cassandra, Hudi, Delta-Lake
- Airflow, Kafka, Apache Spark, Nifi
- Linux (Ubuntu, Mac Os), Cloud Server, GCP, AWS, Docker, Gitlab CD/CI.
- Bash/Scripts, Terraform, Ansible
- Looker Studio, Apache Zeppelin.

PROGRAMTION OUTILS

- Visual Studio Code, Sublime Text
- Terminal: Iterm2, Zsh, Tmux
- Gestion de projet: JIRA, Notion
- Méthodologie: Agile-Scrum

BIBLIOTHÈQUES

Numpy, Spark, Airflow, Kafka-Python,
Pandas, Tensorflow, Wandb, Scikit-learn, Selenium, MLflow.

LANGUES

Français: Intermédiaire
Anglais: Intermédiaire
Vietnamien: Maternelle

CENTRES D'INTÉRÊT

Réparation ordinateur, téléphone
League of legends
Football: MU et PSG
Cinéma

DIVERS

2017-2018: Collaborateur du programme "Good Puppet Show" pour guider les enfants à se protéger des problèmes de violence

Ba-Tuan THAI

Data Engineer, Data Scientist

Diplômé au Master 2 IASD à l'Université Paris Dauphine-PSL et Cloud Data Engineer à la recherche contrat CDI, CDD dans le domaine informatique à partir de 01/07/2023

EXPÉRIENCES PROFESSIONNELLES

CLOUD DATA ENGINEER - STAGE

3/2023 - 6/2023

Jems-Group, DataScientest

- Participer le formation POEI sur Cloud Data Engineer
- Projet: Data pipeline pour surveiller, stocker les données automobiles en temps-réel
- Google Cloud Platform: Pub/Sub, DataFlow, BigQuery, Looker Studio, BigLake, DataPlex
- Open source: Apache Kafka, NiFi, Spark, Cassandra, Zeppelin, Delta-lake, MQTT Broker
- Source code: [GitHub](#) (batuan/project_iot_automobile), Demo: [Youtube](#) (bit.ly/tuan_iot)

MACHINE LEARNING RESEARCH ENGINEER

5/2023 - 12/2023

Institut national de recherche en informatique et en automatique (INRIA)

- Construire Machine Learning model dans la reconnaissance des activités humaines
- Installer, collecter et traitement des données des capteurs
- Utiliser Raspberry pi 4 pour connecter fonctionner des capteurs

RECHERCHE TRAITEMENT AUTOMATIQUE DES LANGUES

4/2021 - 6/2021

Aix-Marseille Université

- Participer à l'inférence de traductions à partir de dictionnaires (TIAD-2021)
- Créer une nouvelle dictionnaire basée sur une dictionnaire connue
- Python, Selenium, Word2vec, MUSE. [Source code GitHub](#) (batuan/TER_TIAD_2021)

DATA SCIENTIST - STAGE FIN D'ÉTUDE

2/2020 - 6/2020

Cenhomes.vn (Viet Nam)

- Collecter, traiter des données immobilières; appliquer des méthodes d'apprentissage automatique pour prédire les prix de l'immobilier
- Améliorer 30% de la précision du Crawler. [Source code GitHub](#) (batuan/vips_datn)

FORMATIONS

Cloud Data Engineer

3/2023 - 6/2023

DataScientest

Master 2 IASD Intelligence Artificielle, Systèmes, Données

2021 - 2022

Université Paris Dauphine-PSL

- Bourse d'Excellence par l'institut PRAIRIE (Paris Artificial Intelligence Research Institute)
- Bases de données avancées, Big Data, Machine Learning, Deep Learning

Master 1 - Informatique

2020 - 2021

Aix-Marseille Université (AMU)

- Réseaux, Science des données, Génie logiciel, Programmation objet concurrente

Diplôme d'Ingénieur en Systèmes d'Information - PFIEV

2015 - 2020

Institut polytechnique de Hanoï (IPH)

- Big Data Analyse et stockage, Réseaux, Projet de développement du système d'information

CERTIFICATES

GOOGLE PROFESSIONAL DATA ENGINEER

GOOGLE

- Design et construit des Systèmes de pipeline de donnée sur Google Cloud Platform
- BigQuery, Cloud Storage, DataFlow, GCP Pub/Sub, GCP Dataplex

CLOUD DATA ENGINEER

DATASCIENTEST

- Data Engineer, Google Cloud Platform, Databricks, Agilité
- Bash, docker, Hadoop, OOP, Spark, SQL, Python

TENSORFLOW DEEP LEARNING DEVELOPER CERTIFICATE (100/100)

GOOGLE

- Créer et ajuster des modèles ML pour le traitement d'images et du langage naturel

DOCKER POUR DÉBUTANTS À AVANCÉS

UDEMY, KODEKLOUD

- Docker, Docker Swarm, Kubernetes