Jimmy Nguyen
Data Engineer

PROFILE

Data Engineer with 4+ years of experience building robust and scalable data systems. **Looking for a technically focused role where I can design, build, and optimize impactful data solutions within collaborative teams.** Passionate about building high-impact data systems and guiding teams toward best practices in engineering and delivery.

CONTACT ME

- +33 6 95 04 21 09
- jimmy-94@hotmail.fr
- www.linkedin.com/in/ji-nguyen/
- https://github.com/Jimmmyy

EDUCATION

ESIEA - Paris
Engineer School
2015 - 2020

LANGUAGES

- French - Native
- English - Professional

HOBBIES

- Outdoor activities
  - Hiking
  - Climbing
  - Calisthenics
- Traveling

SKILLS

Infrastructure
- AWS
- GCP
- Terraform
- Airflow
- Spark
- Docker
- Monitoring
- Grafana
- Linux / Bash

Developement
- Python
- JavaScript
- React
- CI/CD
- MySQL/ Postgres

Collaboration Tools
- Jira, Confluence, Bitbucket (Git)
- Notion
- Slack

WORK EXPERIENCE

Lead Data Engineer
Quantcube Technology l Paris
2023 - Feb. 2025

Technical leadership
- Guided data engineers and software developers (squad of 6 persons) in delivering data and infrastructure projects.
- Led the creation of scalable data systems, making sure they met performance, security, and business needs.
- Established Agile practices (Atlassian tools like Jira), leading to continuous improvements in project delivery and team collaboration.
- Created an effective collaborative environment, encouraging knowledge sharing and skill development for everyone.
- Coordinated with Data Science and business teams to establish project roadmaps, aligning deliverables with strategic goals.

Infrastructure
- Managed, optimized, and automated cloud infrastructure using AWS services such as EC2 for compute, S3 for storage, IAM for access control, and RDS for database management, ensuring efficient resource management and cost optimization.
- Deployed a small infrastructure using Terraform, automating the setup and management of cloud resources. (VPC, Load Balancer, EC2, etc).
- Used Airflow to orchestrate workflows and Docker for consistent and reproducible deployment environments.

Developement
- Developed Python applications, machine learning models, and scripts for various environments, with a focus on performance and scalability.
- Implemented CI/CD pipelines , automating the deployment process in production (Docker, Bitbucket pipeline, Unit testing, Python, SonarQube, Jenkins).
- Monitoring : Set up Slack-based alerts for monitoring systems, facilitating immediate issue tracking and team communication (Python).

Data Engineer
Quantcube Technology l Paris
2021 - 2023

- Utilized AWS SQS and S3 for managing and storing satellite data in a deforestation analysis project, collaborating with data scientists.
- Worked on deploying various company-specific indicators and led multiple project refactorings to enhance process efficiency and performance.
- Used Python for web scraping to automate data collection from diverse web sources, integrating the data into data engineering workflows for Data Scientist models.
- Executed a data collection project using AWS S3 for storage and AWS Athena for querying.

Apprenticeship Data Scientist
Groupe BPCE l Paris
2019 - 2020

- Developed and implemented models and technologies for automatic data classification, focusing on Natural Language Processing (NLP) techniques and various neural network architectures, including Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), and BERT.
- Utilized Neo4j to represent and visualize the resulting data relationship.

PERSONAL PROJECTS

Creating Essential Cloud Services in GCP
First version completed

Stack: GCP, Terraform

Provisioned a fully functional GCP infrastructure using Terraform, with reusable and modular code.
- Designed and deployed a custom Virtual Private Cloud (VPC) with private and public subnets.
- Configured a Bastion Host to securely access internal instances through SSH.
- Set up a NAT Gateway to allow private instances to access the internet securely.
- Deployed multiple Compute Engine VMs simulating development environments.
- Implemented an HTTP Load Balancer to distribute traffic across backend instances.
- Followed best practices for IAM, network segmentation, and infrastructure-as-code versioning.

Creating Portfolio on React
In progress

Stack : Ansible, Terraform, Kubernetes, GCP, Apache Spark, BigQuery

- Built a modern and responsive React portfolio using JSX and component-based architecture.
- Designed a FastAPI backend in Python to expose a REST API serving dynamic project data.
- Integrated frontend with backend to display real-time data such as projects, skills, and contact info.
- Followed best practices for code structure, reusability, and separation of concerns.
- Deployed the full-stack app to a custom domain using a cloud provider (e.g. GCP, Vercel, or AWS).
- Configured CI/CD pipeline and ensured seamless frontend-backend communication in production.

Spark Cluster on GCP with Kubernetes for BigQuery Analytics
In progress

Stack : Ansible, Terraform, Kubernetes, GCP, Apache Spark, BigQuery

- Automated provisioning of a Kubernetes cluster on GCP using Terraform and GKE.
- Deployed a scalable Apache Spark cluster on Kubernetes for distributed data processing.
- Used Ansible to configure Spark workers and manage cluster setup efficiently.
- Connected Spark to Google BigQuery to run large-scale analytics queries on cloud data.
- Designed infrastructure-as-code modules for reusability, security, and scalability.
- Monitored resource usage and optimized Spark performance in a Kubernetes environment.

CRUD FastAPI
First version completed

Stack : Python, FastAPI, SqlAlchemy, Postgres, Poetry, Pytest

- Developed a modular CRUD API using FastAPI, SQLAlchemy, and PostgreSQL, with dependency management via Poetry.
- Implemented database models and asynchronous endpoints for efficient and clean API structure.
- Used SQLAlchemy ORM for robust query handling and data validation.
- Designed the API with scalability and maintainability in mind, applying best practices for routing, schemas, and error handling.

PROFIL

Data Engineer avec plus de 4 ans d'expérience dans la conception de systèmes de données fiables et scalables. À la recherche d'un poste technique permettant de concevoir et de développer des solutions data dans une équipe collaboratives.

CONTACT

📞 +33 6 95 04 21 09

✉️ jimmy-94@hotmail.fr

💻 www.linkedin.com/in/ji-nguyen/

EDUCATION

ESIEA - Paris
Ecole d'inégnieur
2015 - 2020

LANGUES

- Français - Native
- Anglais - Professionel

HOBBIES

- Activités extérieurs
  - Radonnées
  - Escalade
  - Calisthenics
- Voyages

COMPETENCES

Infrastructure
- AWS
- GCP
- Terraform
- Airflow
- Spark
- Docker
- Monitoring
- Grafana
- Linux / Bash

Developement
- Python
- JavaScript
- React
- CI/CD
- MySQL/ Postgres

Communication
- Jira, Confluence, Bitbucket (Git)
- Notion
- Slack

Jimmy Nguyen
Data Engineer

EXPERIENCE PRO

Lead Data Engineer
Quantcube Technology l Paris
2023 - Fev. 2025

Leadership technique
- Coordination d'une équipe de data engineer et software developer pour livrer des projets de données et d'infrastructure.
- Création de projet de data répondant aux besoins en performance, sécurité et business.
- Adoption des pratiques Agiles (outils Atlassian comme Jira), améliorant la livraison des projets et la collaboration.
- Mise en place d'un environnement collaboratif efficace, favorisant le partage de connaissances, la bienveillance et le développement des compétences.
- Collaboration avec les équipes Data Science et business pour définir des roadmaps de projets alignées sur les objectifs stratégiques.

Infrastructure
- Géré, optimisé et automatisé l'infrastructure cloud avec AWS (EC2, S3, IAM, RDS), assurant une gestion efficace des ressources et une optimisation des coûts.
- Implémentation d'une petite infrastructure avec Terraform pour automatiser la gestion des ressources cloud (VPC, Load Balancer, EC2).
- Utilisation d'Airflow pour orchestrer les workflows et Docker pour des environnements de déploiement reproductibles.

Développement
- Développement d'applications Python, de modèles de machine learning et de scripts pour divers environnements, en mettant l'accent sur la performance et la scalabilité.
- Installation des pipelines CI/CD pour automatiser le déploiement en production (Docker, Bitbucket pipeline, tests unitaires, Python, SonarQube, Jenkins).
- Mis en place des alertes Slack pour surveiller les applications, facilitant le suivi des problèmes en temps réel (Python).

Data Engineer
Quantcube Technology l Paris
2021 - 2023

- Utilisation de AWS SQS et S3 pour la gestion et le stockage de données satellites dans un projet d'analyse de la déforestation, en collaboration avec des data scientists.
- Déploiement de divers indicateurs spécifiques à l'entreprise et encadrement de plusieurs refactorisations de projet pour améliorer l'efficacité des processus.
- Automatisation de la collecte de données via du web scraping en Python, puis intégration des données dans les modèles de data scientists.
- Réalisation d'un projet de collecte de données avec AWS S3 pour le stockage et AWS Athena pour les requêtes.

Apprenticeship Data Scientist
Groupe BPCE l Paris
2019 - 2020

- Développement et mise en œuvre de modèles pour la classification automatique des données, en mettant l'accent sur les techniques de Traitement du Langage Naturel (NLP) et diverses architectures de réseaux de neurones (DNN), les Réseaux Neurones Récurrents (RNN) et BERT.
- Utilisation de Neo4j pour la représentation et la visualisation des resultats sous formes de relations.