

Vigneshwar GURUNATHAN

☎ +33-(0) 605502403 — ✉ vicky.april02@gmail.com — 🌐 Portfolio — 🔗 LinkedIn — 🐙 Github — 📍 Mérignac

Ingénieur Data innovant avec plus de 4 ans d'expérience dans la conception et la mise en œuvre d'architectures de données hautes performances. Motivé à appliquer mes compétences en analyses avancées et en processus ETL pour relever des défis complexes dans des environnements dynamiques.

Compétences

Langues: Python, SQL, NoSQL,R

Technologies et outils: Pandas, Airflow, VTK (DICOM),Paraview, Tableau, Kibana, Google Cloud, Docker,Kubernetes, Gitlab, Jenkins, PyFlink, PySpark, SSIS

Expérience Professionnelle

IHU LIRYC - Université de Bordeaux, Pessac, France

Data Engineer

May 2021 – Present

- **Contribution au projet de l'Union Européenne:** Rôle clé dans le projet Beat-AF, une initiative financée par l'Union Européenne pour la gestion avancée des données cliniques en fibrillation auriculaire.
- **Développement de serveurs Web:** Conception d'un serveur Web robuste utilisant Girder, Python et MongoDB, déployé sur le serveur DMZ pour gérer divers ensembles de données, y compris DICOM et les signaux EP.
- **Outils de visualisation de données:** Développement de visualisations en utilisant VTK et des plugins personnalisés de Girder pour améliorer l'interprétation des données et la prise de décision clinique.
- **Pipeline de données conforme au RGPD:** Conception d'un pipeline de preuve de concept utilisant Apache Airflow pour anonymiser les données cliniques en conformité avec les réglementations RGPD.
- **Pipeline ETL pour les données ECG:** Pipeline ETL pour les données ECG : Développement d'un pipeline ETL pour les données ECG anonymisées via l'API Kardia Pro avec stockage centralisé, accompagné de tableaux de bord interactifs en temps réel pour surveiller et visualiser les ensembles de données cliniques.
- **Technologies utilisées:** Girder, Python, MongoDB, VTK, Apache Airflow, Kardia Pro API, Trame.

Vectorive, Paris, France

Data Engineer

Jan 2020 – May 2021

- **Développement de pipelines de données:** Intégration de Dynamics 365 avec Azure AD pour une authentification fluide, création de pipelines pour l'extraction de données CRM via l'API WEB et OData, et prétraitement des données.
- **Solutions de données en streaming:** Développement d'un pipeline de données en streaming (POC) utilisant PyFlink avec Kafka comme source de données et PostgreSQL comme récepteur.
- **Conception et optimisation ETL:** Création et maintenance d'ETL pour garantir une haute qualité et fiabilité des données dans des environnements de traitement en temps réel et par lot.
- **Technologies utilisées:** Dynamics 365, Azure Active Directory, WEB API, OData, Pandas, PyFlink, Kafka, PostgreSQL.

Formations

M.Sc. in Data Engineering for AI, Data Science Tech Institute, Paris, France	2019 – 2020
MBA in International Trade Management, IPAG Business School, Paris, France	2017 – 2018
M.Tech in Digital Communication Network, Kalasalingam University, India	2011 – 2013
B.E in Electronics Communication,Mailam Engineering College ,India.	2007 – 2011

Projets Sélectionnés

Airflow-2019: Création d'un pipeline de données en temps réel utilisant Airflow dans un environnement Docker, avec des tâches incluant la vérification des fichiers sources, le nettoyage des données (MySQL), l'agrégation parallèle, l'exportation en CSV et la notification par e-mail. ([Git](#))

Object-detection-API - 2020: Utilisation de modèles pré-entraînés de détection d'objets TensorFlow pour détecter des objets sur des pages web, avec Selenium pour l'interaction web, Docker pour la configuration de l'environnement et Jupyter pour le développement et les tests. ([Git](#))

ELK-2020: Chargement du jeu de données MovieLens dans Elasticsearch à l'aide de Python et visualisation avec Kibana. ([Git](#))

SSIS-2019: Création d'un modèle de Data Warehouse avec SSIS pour les données de crimes en France (1996-2010), intégration du staging et du data store opérationnel pour un Data pipeline et un import des fichiers excel. ([Git](#))

Certifications

Airflow (Dag-Authoring) - Astronomer
Neo4J Certified Professional - Neo4J
Data Analysis with Python - Cognitive Class
Machine Learning with Python - Cognitive Class
GCP - Professional Data Engineer - Prévu pour mars 2025

Langues et Centres d'Intérêt

Anglais - Courant
Français - Intermédiaire (B1)
Tamoul - Langue maternelle
Données - LLM,Computer Vision, Snowflake,MLOps.
Sports - Rugby, Cricket, Kabaddi, Natation.

Vigneshwar GURUNATHAN

☎ +33-(0) 605502403 — ✉ vicky.april02@gmail.com — 🌐 Portfolio — 🔗 LinkedIn — 🐙 Github — 📍 Mérignac

Innovative Data Engineer with over 4 years of experience in designing and implementing high-performance data architectures. Excited to apply my skills in advanced analytics and ETL processes to solve complex challenges in fast-paced environments.

Skills

Languages: Python, SQL, NoSQL,R

Technologies & Tools: Pandas, Airflow, VTK (DICOM),Paraview, Tableau, Kibana, Google Cloud, Docker,Kubernetes, Gitlab, Jenkins, PyFlink, PySpark, SSIS

Work Experience

IHU LIRYC - Université de Bordeaux, Pessac, France

Data Engineer

May 2021 – Present

- **European Union Project Contribution:** Played a key role in Beat-AF, an European Union-funded initiative for advanced clinical data management in atrial fibrillation.
- **Web Server Development:** Built a robust Web server using Girder, Python, and MongoDB and deployed on the DMZ server to manage diverse datasets, including DICOM and EP signals.
- **Data Visualization Tools:** Developed visualizations leveraging VTK and custom Girder plugins to enhance data interpretation and clinical decision-making.
- **GDPR-Compliant Data Pipeline:** Designed a proof-of-concept pipeline using Apache Airflow to anonymize clinical data in compliance with GDPR regulations.
- **ECG Data ETL Pipeline:** Developed an ETL pipeline for anonymized ECG data from Kardia Pro API with centralized storage, alongside real-time interactive dashboards for monitoring and visualizing clinical datasets.
- **Technologies used:**Girder, Python, MongoDB, VTK, Apache Airflow, Kardia Pro API,Trame.

Vectorive, Paris, France

Data Engineer

Jan 2020 – May 2021

- **Data Pipeline Development:** Integrated Dynamics 365 with Azure AD for seamless authentication, built pipelines for CRM data extraction via WEB API and OData, and preprocessed data using Pandas.
- **Streaming Data Solutions:** Developed a proof-of-concept (POC) data streaming pipeline using PyFlink with Kafka as the data source and PostgreSQL as the sink.
- **ETL Design and Optimization:** Built and maintained ETLs to ensure high data quality and reliability in real-time and batch processing environments.
- **Technologies used:** Dynamics 365, Azure Active Directory, WEB API, OData, Pandas, PyFlink, Kafka, PostgreSQL

Education

M.Sc. in Data Engineering for AI, Data Science Tech Institute, Paris, France	2019 – 2020
MBA in International Trade Management, IPAG Business School, Paris, France	2017 – 2018
M.Tech in Digital Communication Network, Kalasalingam University, India	2011 – 2013
B.E in Electronics Communication,Mailam Engineering College ,India.	2007 – 2011

Selected Projects

Airflow-2019: Built a real-time data pipeline using Airflow in a Docker environment with tasks including source file check, data cleaning (MySQL), parallel aggregation, CSV export, and email notification. ([Git](#))

Object-detection-API - 2020: Leveraged pre-trained TensorFlow Object Detection models to detect objects on web pages, using Selenium for web interaction, Docker for environment setup, and Jupyter for development and testing. ([Git](#))

ELK-2020: Load Movie Lens dataset into Elasticsearch using Python and visualize with Kibana.([Git](#))

SSIS-2019: Created a Data Warehouse model using SSIS for Crimes data in France (1996-2010), integrating staging (STA) and operational data store (ODS) for efficient data flow and import from excel files. ([Git](#))

Certifications

Airflow (Dag-Authoring) - Astronomer
Neo4J Certified Professional - Neo4J
Data Analysis with Python - Cognitive Class
Machine Learning with Python - Cognitive Class
GCP Professional Data Engineer - Expected by March 2025

Languages and Interests

English - Fluent
French - Intermediate (B1)
Tamil - Native
Data - LLM,Computer Vision,Snowflake,MLOps.
Sports - Rugby, Cricket, Kabaddi, Swimming.