

Otmane NKAIRA

Senior Data & Analytics Engineer GCP

PROFIL

Senior Data & Analytics Engineer GCP avec plus de 10 ans d'expérience dans la conception et l'industrialisation de plateformes data complexes. J'excelle dans la construction de **pipelines de données scalables** (batch et temps réel) sur des écosystèmes Big Data, avec une expertise approfondie de **Google Cloud Platform (GCP)**. Ma mission : bâtir des fondations data robustes, optimisées et automatisées (**Terraform, Airflow, CI/CD**) pour alimenter les stratégies BI et IA de l'entreprise.

CONTACT

📍 Plaisir (78)
📞 07 61 64 93 44
✉ otmane@passionatedata.fr
🌐 passionatedata.fr

COMPÉTENCES CLÉS

Cloud & Data Engineering

Google Cloud Platform (GCP) BigQuery Airflow
Cloud Composer PySpark DBT Dataform
Dataproc Pub/Sub Cloud Storage App Engine
Compute Engine ETL MS Azure

Bases de données

SQL (Expert) PostgreSQL MySQL MS SQL Server
MongoDB Elasticsearch Firestore SAP HANA
TERADATA

Programmation & DevOps

Python (Expert) Java JEE C# PHP R C/C++
HTML/CSS Terraform Git / Gitlab Kubernetes
CI/CD DevOps FastAPI Linux APIs

Data Science & IA

Machine Learning Scikit-learn Deep Learning
PyTorch MLOps Vertex AI LLM Application
Prompt Engineering NLP Gensim Power BI
Looker Studio

Agilité & Outils

Méthodologie Agile Scrum JIRA Confluence
Kibana

FORMATIONS

MASTER DONNÉES & LANGAGE NATUREL

Univ. de Montpellier | 2012-2014

MASTER EXPERT EN INFORMATIQUE

SUPINFO, Montpellier | 2010-2012

LICENCE PRO. DÉVELOPPEUR JAVA / C++

Univ. Hassan II, Casablanca | 2009-2010

PRIX ET DISTINCTIONS

Prix SILVER au Cristal Festival

Opération ACCORHOTELS #FeelWelcomeFromThe1stTweet

Émis par Cristal Festival | Déc. 2016

Associé à Publicis Conseil

Cat. Marketing Innovation - Data - IoT / Best use of social media

LANGUES

Anglais - Courant (C1)

Arabe - Bilingue

Français - Bilingue

RÉSULTATS & KPIs

- ROI de 5€ pour 1€ investi
- Taux de conversion doublé (+100%)
- Traitement réduit de 90% (30h à 3h)
- 90% de précision (maint. prédictive)
- Accès données 2x plus rapide

EXPÉRIENCES PROFESSIONNELLES

Depuis Septembre 2022 (4 ans)

Consultant Data & IA Indépendant

PASSIONATE DATA

De Septembre 2025 à Aujourd'hui (1 an)

Analytics Engineer & Développeur Data App (GCP / Python) @ SERVIER

Contexte : Création d'un **Data Product** de A à Z (**Data Engineering** & Application) pour le pilotage, la visualisation et la traçabilité des flux de la Supply Chain pharmaceutique.

Missions & Réalisations :

- Analytics Engineering** : Modélisation des données avec **dbt** sur **BigQuery**. Transformation des flux logistiques bruts en indicateurs métiers précalculés pour optimiser les performances.
- Développement Application** : Création from scratch d'une interface interactive en **Python (Streamlit)** offrant à la Supply Chain une visualisation sur mesure des transferts de produits.
- Cloud & DevOps** : Conteneurisation et déploiement **serverless** sur GCP via **Cloud Run. Infrastructure as Code** avec **Terraform** et automatisations des pipelines **CI/CD** via **GitLab**.
- Delivery & Produit** : Collaboration directe avec les utilisateurs finaux en méthodologie **Agile** (cycles courts) pour garantir l'adoption et le respect des normes pharmaceutiques.

🏆 Impact & Résultats

🕒 Performance technique : Déploiement automatisé permettant des mises en production quotidiennes sans interruption de service.

👥 Impact Business : Adoption par +50 utilisateurs métiers de la Supply Chain.

GCP BigQuery Cloud Run Python Streamlit dbt Terraform GitLab CI/CD Git

De Septembre 2022 à Juin 2025 (2 ans et 9 mois)

Senior Data Engineer GCP - Marketing & Personnalisation @ CARREFOUR

Contexte : Conception et industrialisation d'une **Data Platform** Marketing (promotions ciblées) à fort volume, traitant plusieurs millions de transactions quotidiennes.

Missions & Réalisations :

- Data Engineering & Orchestration** : Refonte complète des **pipelines de batch processing** et de traitement massif avec **PySpark** sur **Dataproc, BigQuery**, et **Dataform**. Orchestration via **Cloud Composer (Airflow)**.
- MLOps & Industrialisation** : Déploiement et industrialisation des algorithmes de **Machine Learning** (recommandation et ciblage) pour automatiser les campagnes marketing.
- DevOps Data & FinOps** : Automatisation de l'infrastructure en **Terraform** et création des pipelines **CI/CD** sous **GitLab**. Pilotage de l'optimisation des coûts **GCP** (règles de cycle de vie **Cloud Storage**, archivage, optimisation des requêtes **SQL**).
- Delivery & Produit** : Interface technique directe avec les équipes Produit et Marketing (**Scrum**) pour garantir l'alignement entre l'architecture Data et le time-to-market.

🏆 Impact & Résultats

🕒 Performance technique : Réduction de 90 % du temps de traitement data (passage de 30h à 3h).

👥 Impact Business : Taux de conversion des campagnes marketing doublé (+100 %) grâce à la fiabilité du ciblage.

🔥 Optimisation : Réduction de la facture Cloud GCP via l'implémentation de règles **FinOps** strictes.

GCP BigQuery Dataproc Cloud Storage Pub/Sub Cloud Composer Python PySpark SQL Airflow Dataform Terraform GitLab CI/CD Agile/Scrum

De Mars 2020 à Août 2022 (2 ans et 6 mois)

Consultant Data Senior

KEYRUS

De Septembre 2021 à Août 2022 (1 an)

Data Product Owner & Analytics Engineer GCP @ L'ORÉAL

Contexte : Pilotage à l'international de produits Data analytiques et modernisation complète de la **Data Platform** (migration des systèmes legacy SAP et Teradata vers **GCP**).

Missions & Réalisations :

- Product Management & Gouvernance** : Pilotage de la roadmap et du backlog (**Agile/Scrum**). Animation des comités de gouvernance et alignement stratégique entre les directions métiers et techniques à l'international.
- Data Engineering & Migration Cloud** : Conduite de la migration complexe de centaines de tables critiques depuis **SAP HANA** et **Teradata** vers **GCP (BigQuery)**.
- Analytics Engineering & BI** : Modélisation de la donnée avec **dbt** sur **BigQuery** pour simplifier les calculs intermédiaires. Gestion du cycle de vie complet jusqu'à la publication des dashboards **Power BI**.
- Data Quality & DevOps Data** : Garantie de la fiabilité des flux analytiques (standards de qualité, supervision UAT). Automatisation des déploiements via **GitLab CI** et **Infra-as-Code (Terraform)**.

🏆 Impact & Résultats

🔥 Performance : Accès aux données 2x plus rapide pour le reporting stratégique de la direction.

✓ Qualité : Fiabilisation totale des flux à l'international grâce aux standards de tests et de contrôle.

GCP BigQuery Cloud Composer Cloud Storage dbt Power BI Elasticsearch SAP HANA Teradata Terraform GitLab CI/CD Scrum / Agile

De Septembre 2020 à Août 2021 (1 an)

Senior Data Engineer GCP @ EIFFAGE

Contexte : Industrialisation (**MLOps**) de la chaîne prédictive de rentabilité des chantiers : passage d'un modele R&D à une solution de production robuste sur **GCP**.

Missions & Réalisations :

- Data Engineering & Ingestion** : Création de **pipelines d'extraction** des bases opérationnelles (**MS SQL Server**) vers **GCP (Cloud Storage, BigQuery)** pour consolider les données d'entraînement et d'inférence.
- MLOps & Batch Processing** : Refactorisation du code R&D (**Python**), conteneurisation (**Docker**, Artifact Registry) et déploiement du modèle sur **Cloud Run** pour l'exécution automatisée des prédictions en batch processing.
- Orchestration & DevOps Data** : Automatisation complète de la chaîne de préparation de données et d'inférence via **Cloud Composer (Airflow)**. Implémentation des pipelines **CI/CD** (Cloud Build, GitLab CI).
- Data Quality & Delivery** : Développement d'un framework de tests automatisés (**Pytest**) pour garantir la fiabilité des flux. Rédaction des spécifications d'architecture et transfert de compétences aux équipes RUN.

🏆 Impact & Résultats

🌟 Mise à l'échelle : Transformation réussie d'un code expérimental (R&D) en une architecture Cloud serverless, scalable et 100% automatisée.

✓ Fiabilité : Sécurisation des données injectées dans le modèle grâce aux contrôles de Data Quality continus.

GCP BigQuery Cloud Storage Cloud Run Cloud Composer Cloud Build Python Pandas Pytest MS SQL Server T-SQL Docker Airflow GitLab CI Scrum / Agile

De Mars 2020 à Août 2021 (6 mois)

Data Analyst & Data Quality Lead @ AGENCE DE BIOMÉDECINE

Contexte : Audit national de la qualité des données de santé et définition de la stratégie de **Data Governance** pour plus de 400 centres de greffe cardiaque.

Missions & Réalisations :

- Data Quality & Audit** : Conduite d'un audit d'envergure nationale sur les flux de données critiques de +400 centres hospitaliers afin de mesurer la fiabilité des registres de santé.
- Data Science & Scoring** : Modélisation et développement en **Python (Scikit-learn, Pandas)** d'un algorithme de scoring. Implémentation de modèles de détection d'anomalies pour évaluer et classifier la fiabilité des données.
- Data Governance & Stratégie** : Identification des biais structurels majeurs et élaboration de nouvelles directives officielles pour standardiser la collecte au niveau national.
- Advisory & Restitution** : Rédaction et présentation du rapport stratégique aux instances décisionnelles, posant les fondations de la nouvelle politique de contrôle qualité.

🏆 Impact & Résultats

👥 Échelle nationale : Cartographie et audit des données sur un réseau critique de +400 centres médicaux.

✓ Fiabilité : Création d'un référentiel de scoring automatisé permettant d'isoler et de corriger les incohérences de saisie.

🏛️ Gouvernance : Définition et adoption par les instances décisionnelles des nouvelles normes officielles de collecte de données.

Data Governance Data Quality Python Scikit-learn Pandas NumPy Statistiques descriptives Santé publique

De Mars 2017 à Février 2020 (3 ans)

Consultant Data Senior

PUBLICIS ETO

De Mars 2018 à Février 2020 (2 ans)

Data Engineer GCP - Marketing & Personalisation @ CARREFOUR

Contexte : Conception et industrialisation de la **Data Platform** d'une solution de ciblage marketing omnicanal à fort impact financier.

Missions & Réalisations :

- Data Engineering & Migration** : Création des **pipelines** fondateurs de **batch processing** sur **GCP** pour l'extraction massive de données depuis les bases legacy (**Teradata**) vers **BigQuery** et **Cloud Storage**.
- Architecture Event-Driven** : Mise en place d'une infrastructure **serverless** orientée événements (**Cloud Functions, App Engine, Firestore, Pub/Sub**) pour alimenter les moteurs de personnalisation à très grande échelle.
- MLOps & Data Science** : Développement de modèles prédictifs de propension d'achat (**Python, Scikit-learn**) et industrialisation de leur chaîne d'entraînement et de scoring.
- Delivery & Stratégie Business** : Traduction des enjeux marketing en spécifications techniques (Agile/Scrum) pour garantir un alignement total avec les objectifs de rentabilité.

🏆 Impact & Résultats

🔥 Rentabilité (ROI) : 5 € générés pour 1 € investi grâce à la précision du ciblage prédictif.

👥 Business : Maximisation du taux de conversion des offres marketing via le scoring automatisé.

🌟 Scalabilité : Modernisation réussie du socle de données (passage de Teradata à une architecture GCP serverless et robuste).

GCP BigQuery Dataproc Cloud Storage Pub/Sub Cloud Composer Firestore App Engine Compute Engine Python PySpark SQL Airflow Dataform Terraform GitLab CI/CD Agile/Scrum

De Mars 2017 à Février 2018 (1 an)

Data Scientist - NLP & Automation @ RELAXNEWS

Contexte : Conception et industrialisation d'un moteur **NLP (Natural Language Processing)** de classification documentaire et de recherche full-text pour automatiser les processus métiers.

Missions & Réalisations :

- Data Engineering & Ingestion** : Automatisation de bout en bout des **pipelines d'extraction** des données brutes depuis **MS SQL Server**.
- Data Science & NLP** : Audit du corpus textuel, ingénierie des caractéristiques (**Text Mining**, vectorisation **TF-IDF/Embeddings**) et développement d'algorithmes de classification avec **Python (Scikit-learn)**.
- Search Engine & Architecture** : Intégration et configuration d'**Elasticsearch** pour l'indexation massive des documents et la restitution des prédictions.
- Delivery & Change Management** : Rédaction de la documentation d'architecture, restitution des métriques de performance et accompagnement des utilisateurs métiers vers le nouvel outil.

🏆 Impact & Résultats

🔥 Automatisé : Remplacement d'un processus de tri documentaire 100 % manuel par un pipeline automatisé de bout en bout.

✓ Précision IA : Atteinte d'un score de fiabilité de 90 % sur la catégorisation automatique des documents.

🕒 Productivité : Accélération drastique du temps de recherche et d'accès à l'information pour les équipes opérationnelles grâce au moteur full-text.

Python Scikit-learn Pandas NLP TF-IDF Embeddings Text Mining Elasticsearch MS SQL Server T-SQL Architecture Data

De Mars 2015 à Février 2017 (2 ans)

Data Scientist & Social Intelligence Analyst

PUBLICIS CONSEIL

Contexte : Traitement de données massives non structurées (**Social Listening**), modélisation **NLP** et génération d'insights stratégiques pour le compte d'acteurs majeurs (CAC 40).

Missions & Réalisations :

- Data Engineering & Ingestion** : Conception de **pipelines** pour l'extraction et la préparation de données massives non structurées issues des réseaux sociaux (**APIs, MongoDB**).
- Data Science & NLP** : Développement sous **Python (Scikit-learn, Gensim)** d'un moteur temps réel de détection d'incidents (**Topic Modeling, Text Mining**) pour analyser la connectivité réseau.
- Analytics & Modélisation** : Création de modèles comportementaux (**clustering**, analyse sémantique et géolocalisation) pour accompagner la transformation digitale des clients.
- Data to Business** : Traduction d'analyses de données complexes en recommandations opérationnelles directes pour les équipes marketing et créatives.

🏆 Impact & Résultats

🔥 Business Impact (Orange) : Extraction de l'insight stratégique ayant directement déclenché et nourri la campagne publicitaire nationale « La Fibre 100% Fibre » d'Orange.

🏆 Récompense (AccorHotels) : Lauréat du Prix SILVER au Cristal Festival pour la campagne data-driven #FeelWelcomeFromThe1stTweet.

🌟 Déploiement Grands Comptes : Adoption des modèles prédictifs et analytiques par des leaders de l'industrie (AXA, BNP Paribas, Sanofi, Nestlé, Renault, ENGIE).

Python Scikit-learn Gensim R NLP Topic Modeling Sentiment Analysis Text Mining MongoDB Social Media APIs Synthésio Netbase Data Marketing