

# DIEGO NEGRIER

Data Engineer

[contact@data-worlds.com](mailto:contact@data-worlds.com)

+33 X XX XX XX XX

Paris / Bordeaux

[data-worlds.com](http://data-worlds.com)

DATA ENGINEER

PIPELINE ARCHITECT

CLOUD DATA

PYTHON EXPERT

BIG DATA

---

### Pilotage de la Performance Industrielle

Raisonnement en processus, flux et causalités pour traiter la performance comme un système global et non par indicateurs isolés. Identification des leviers à impact réel ...

### Lien Production — Maintenance — Énergie

Capacité à relier les pannes aux usages, la consommation énergétique au fonctionnement réel des machines, et les décisions d'exploitation aux résultats observés. C'est ce lien ...

### Data & Aide à la Décision Terrain

Transformation des données industrielles (capteurs, SCADA, IoT, GMAO) en constats actionnables. Pas seulement produire des analyses : transformer les constats en règles d'exploitation, participer à ...

### Maintenance Prédicative & Fiabilité

Développement de modèles de détection d'anomalies et de prévision de défaillances sur équipements industriels. Analyse de séries temporelles, corrélation pannes/usages, réduction des arrêts non planifiés. ...

### Posture Hybride & Conduite du Changement

Échange avec la direction sur les enjeux globaux, travail avec la maintenance et la production au quotidien, structuration de méthodes simples et durables. Formation des ...

## >> MISSIONS & EXPÉRIENCES

### Ministère de l'Éducation Nationale

2024 - Présent

Conception et déploiement d'un catalogue de données national centralisant les métadonnées de +200 flux de données métier. Architecture data : pipelines ETL Python pour la collecte, la normalisation et la transformation des données issues de multiples SI. Mise en place de contrôles qualité automatisés (Great Expectations, SQL). Dashboards Power BI de monitoring de la qualité et de la complétude des données. Stack : Python, SQL, PostgreSQL, Power BI, dbt, Git.

[STACK] [Data Governance & Engineering](#)

### SUD OUEST — Média

2023

Architecture data et pipelines ETL sur Google Cloud Platform. Ingestion de données multi-sources (API, fichiers, bases) dans BigQuery. Transformation SQL et Python/Spark pour alimenter les modèles analytiques. Développement de modèles de machine learning (Scikit-learn) pour l'optimisation des processus opérationnels. Automatisation du reporting, mise en place CI/CD (GitHub Actions). Stack : Python, SQL, Spark, GCP, BigQuery, Scikit-learn, Git.

[STACK] [Data Engineering GCP](#)

### AGIRC ARRCO

2022 - 2023

Mise en place de la gouvernance des données sur cluster Cloudera/Hadoop. Développement d'outils de monitoring de la qualité des données avec Flask et Python (contrôles de complétude, cohérence, fraîcheur). Création de pipelines de reporting automatisé. Documentation du dictionnaire de données, standardisation des flux ETL et accompagnement des équipes data. Stack : Python, Flask, SQL, Hadoop, Cloudera, Hive, Git.

[STACK] [Data Governance Cloudera](#)

### VEOLIA — Eau & Assainissement

2020 - 2022

Construction d'une plateforme data pour la collecte, le stockage et l'analyse de données IoT et SCADA en temps réel. Pipelines Spark Streaming pour l'ingestion de flux capteurs. Data lake structuré (Bronze/Silver/Gold) pour l'historisation. Tableaux de bord Power BI pour le suivi opérationnel. API REST (FastAPI) pour exposer les données aux applications métier. Stack : Python, Spark, FastAPI, PostgreSQL, Power BI, Docker, Git.

[STACK] [Data Platform & BI](#)

### LAPISCINE — Formation

2020

Formateur Python et Data Science auprès de publics en reconversion professionnelle. Conception de programmes pédagogiques couvrant : Python (bases, POO, librairies), SQL, Pandas, NumPy, machine learning (Scikit-learn) et visualisation (Matplotlib, Seaborn). Animation d'ateliers pratiques sur des jeux de données réels (Kaggle, open data). Évaluation continue et projets fil rouge.

[STACK] [Formation Python & Data](#)

### PRIMAGAZ — Énergie

2019 - 2020

Développement end-to-end d'un système de détection d'anomalies sur les séries temporelles de consommation de gaz. Exploration des données (Pandas, Jupyter), feature engineering temporel, entraînement et validation croisée de modèles (Isolation Forest, LSTM). Pipeline de scoring automatisé en production (Python, FastAPI, Docker). Dashboards de suivi des alertes et des performances du modèle. Résultat : identification anticipée des compteurs défaillants. Stack : Python, Pandas, Scikit-learn, FastAPI, Docker, PostgreSQL, Power BI.

[STACK] [Data Science & ML](#)

### SUEZ — Environnement

2015 - 2019

Architecture data cloud complète sur Azure pour le traitement de données industrielles à grande échelle. Pipelines ETL avec Azure Data Factory et Databricks (Spark/Python) pour l'intégration de données capteurs IoT, SCADA et systèmes métier. Modélisation d'un data warehouse (Azure SQL) optimisé pour le reporting. Développement de modèles ML (Scikit-learn, XGBoost) pour la détection d'anomalies sur les processus de production. CI/CD Azure DevOps, tests de qualité de données, documentation technique. Stack : Python, Spark, Azure, SQL, Scikit-learn, Power BI, Docker.

[STACK] [Data Engineering Azure](#)

### # DATA ENGINEERING

- Python
- SQL
- Spark
- Airflow
- Kafka
- dbt
- PostgreSQL

### # CLOUD & INFRA

- Azure
- AWS
- GCP
- Docker
- Kubernetes
- CI/CD

### # BASES DE DONNÉES

- PostgreSQL
- MongoDB
- Redis
- Elasticsearch
- Data Lake

### # OUTILS & DEVOPS

- Git
- Linux
- Terraform
- FastAPI
- Django

## >> LANGUES

- Français [Maternel](#)
- Anglais [Professionnel](#)
- Espagnol [Courant](#)

## >> FORMATIONS

### Licence Pro SARI

Python, C++, Java  
CERGY (95) - 2017

### BTS Assitance Technique d'ingénieur

Electrotechnique  
Passy Buzenval (92) - 2014

### Bac Pro Electrotechnique

Electrotechnique  
Passy Buzenval (92) - 2012

### CNAM

MASTER MEDAS — Méthodes statistiques et data science. Machine learning, ...  
Paris - 2019

### Université de Cergy-Pontoise

Licence Professionnelle SARI — Systèmes, Administration et Réseaux Informatiques. Infrastructure, ...  
Cergy - 2017

### Lycée Technique

BTS ATI — Assistance Technique d'ingénieur. Systèmes industriels, gestion de ...  
France - 2014

### Lycée Professionnel

Bac Pro Électrotechnique — Électricité industrielle et tertiaire. Installations électriques, ...  
France - 2012