

FraudLens - Healthcare Fraud Detection Platform

Description

FraudLens est une plateforme de detection de fraudes dans le secteur de la sante. Elle ingere et centralise plus de 10 datasets publics CMS/OIG pour detecter des anomalies dans les paiements Medicare/Medicaid et identifier les prestataires a risque.

Architecture

Data Flow: CMS Medicare → OIG LEIE → Open Payments → NPPES → S3 → Snowflake (Bronze → Silver → Gold) → Dashboard



Stack Technique

Composant	Technologie	Role
Orchestration	Apache Airflow 3.x	Scheduling et execution des pipelines
Data Lake	Amazon S3	Stockage des fichiers Parquet
Data Warehouse	Snowflake	Stockage et traitement analytique
Transformations	dbt Core	Modelisation et tests de donnees
CI/CD	GitHub Actions	Tests et deploiement automatises
Dashboard	Streamlit + Plotly	Visualisation et analyse
Conteneurisation	Docker Compose	Environnement Airflow
Traitement	Python 3.12, Pandas	ETL et manipulation de donnees

Frequence de mise a jour des datasets

Dataset	Frequence	Periode
NPPES	Hebdomadaire	Tous les lundis (Snowflake Marketplace)
LEIE	Mensuelle	1ere semaine du mois
Medicare Hospital Spending	Trimestrielle	Jan, Avr, Juil, Oct
Provider Info / Nursing Homes	Trimestrielle	Jan, Avr, Juil, Oct
Long-term Care Hospital	Trimestrielle	Dec, Mar, Juin, Sep
Hospice	Trimestrielle	Fev, Mai, Aout, Nov
Home Health Care	Trimestrielle	Jan, Avr, Juil, Oct
Open Payments	Annuelle	Fin juin (+refresh janvier)
Medicare Part D	Annuelle	Printemps/ete

Outputs

Dashboard Streamlit

- **Overview** - KPIs et metriques cles
- **Fraud Alerts** - Alertes actionnables par priorite
- **Provider 360** - Vue complete d'un provider (recherche NPI) + ML Analysis
- **Analytics** - Analyses geographiques et tendances

Tables Gold

Table	Description
provider_360	Vue 360° complete du provider
fraud_risk_score	Score de risque 0-100
high_risk_alerts	Alertes priorisees pour investigation
payments_summary	Aggregations paiements par NPI
prescriptions_summary	Aggregations prescriptions par NPI
provider_ml_features	Features ML pour detection d'anomalies (z-scores, HHI, percentiles)

Documentation Complementaire

[Snowflake_doc.pdf](#) - Infrastructure, schemas, tables

[DBT_doc.pdf](#) - Transformations, tests, architecture Medallion

[CICD_doc.pdf](#) - GitHub Actions, deploiement, secrets

[DataModeling_doc.pdf](#) - Modele de donnees, scoring fraude

[ML_Features_doc.pdf](#) - Detection d'anomalies ML (z-scores, HHI, diversite pharma)