



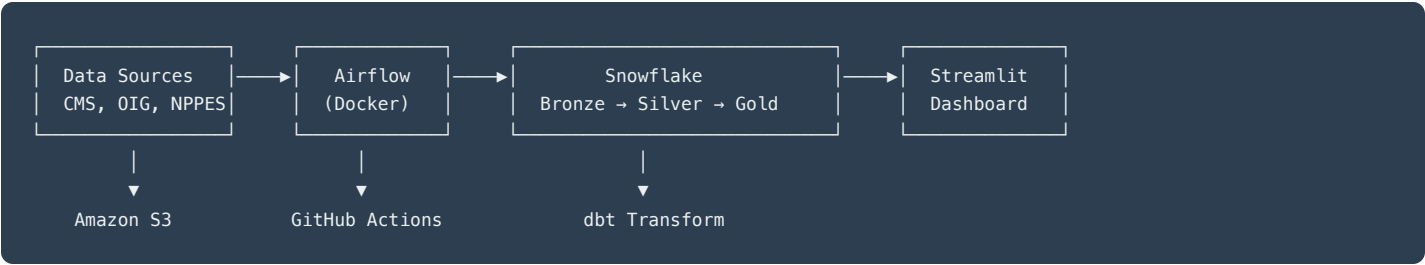
FraudLens - Healthcare Fraud Detection Platform

Description

FraudLens est une plateforme de détection de fraudes dans le secteur de la santé. Elle ingère et centralise plus de 10 datasets publics CMS/OIG pour détecter des anomalies dans les paiements Medicare/Medicaid et identifier les prestataires à risque.

Architecture

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



Stack Technique

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

Sources de Données

Source	Description	Fréquence
OIG LEIE	Liste des providers exclus Medicare/Medicaid	Mensuelle
CMS Open Payments	Paielements pharma aux médecins (General, Research, Ownership)	Annuelle
Medicare Part D	Prescriptions médicaments par prescripteur	Annuelle
Medicare Hospital Spending	Dépenses hospitalières par établissement	Annuelle
NPES	Registre national des providers (7M+ NPIs)	Continue
Provider Information	Informations nursing homes	Trimestrielle
Hospice/Home Health/LTCH	Établissements de santé spécialisés	Annuelle

Outputs

Dashboard Streamlit

- **Overview** - KPIs et métriques clés
- **Fraud Alerts** - Alertes actionnables par priorité
- **Provider 360** - Vue complète d'un provider (recherche NPI)
- **Analytics** - Analyses géographiques et tendances

Tables Gold

Table	Description
provider_360	Vue 360° complète du provider
fraud_risk_score	Score de risque 0-100
high_risk_alerts	Alertes priorisées pour investigation
payments_summary	Agrégations paiements par NPI
prescriptions_summary	Agrégations prescriptions par NPI

Documentation Complémentaire

- **Snowflake_Documentation.pdf** - Infrastructure, schemas, tables
- **DBT_Documentation.pdf** - Transformations, tests, architecture Medallion
- **CICD_Documentation.pdf** - GitHub Actions, déploiement, secrets

- **DataModeling_Documentation.pdf** - Modèle de données, scoring fraude