

FraudLens - Data Modeling Documentation

Architecture des données et modèles analytiques

Vue d'ensemble

FraudLens utilise une architecture Medallion (Bronze → Staging → Silver → Gold) pour transformer les données brutes en insights actionnables. Le modèle est centré sur le **Provider (NPI)** qui agrège toutes les sources pour la détection de fraude.

Architecture Medallion

| BRONZE | STAGING | SILVER | GOLD |
|-----------------|------------------|----------------------|---------------------|
| LEIE | → stg_leie | → excluded_providers | |
| MEDICARE_HOSP | → stg_medicare_* | → hospital_spending | provider_360 |
| OPEN_PAYMENTS_* | → stg_open_pay_* | → payments | → payments_summary |
| PART_D_PRESCR | → stg_part_d_* | → prescriptions | → prescriptions_sum |
| PROVIDER_INFO | → stg_provider_* | → facilities | fraud_risk_score |
| NPPES | → stg_nppes_* | → provider | → high_risk_alerts |

Couches de données

| Couche | Description | Matérialisation |
|---------|--------------------------------------|-----------------|
| Bronze | Données brutes depuis S3 (Parquet) | Tables |
| Staging | Standardisation colonnes, cast types | Vues |
| Silver | Tables nettoyées, enrichies, jointes | Tables |
| Gold | Tables analytiques agrégées | Tables |

Tables Silver

| Table | Description | Clé | Jointures |
|--------------------|-------------------------------|-------------|-----------------------------------|
| provider | Table maître des providers | NPI | NPPES + adresse + taxonomy + LEIE |
| excluded_providers | Providers exclus OIG | NPI | LEIE enrichi avec provider |
| payments | Paiements pharma consolidés | Record_ID | Open Payments (3 tables) |
| prescriptions | Prescriptions Medicare Part D | NPI + Drug | Medicare Part D + provider |
| facilities | Établissements de santé | CCN | Hospice, Home Health, LTCH |
| hospital_spending | Dépenses hospitalières | Facility_ID | Medicare Hospital Spending |

Modèle Provider 360

Le modèle **provider_360** est le cœur du système. Il agrège toutes les données autour du provider (NPI) pour créer une vue unifiée.

Identité: NPI, FULL_NAME, ORGANIZATION_NAME, ENTITY_TYPE, CREDENTIAL

Localisation: STATE, CITY, ZIP_CODE, PHONE

Spécialité: SPECIALTY, SPECIALTY_CLASSIFICATION, PROVIDER_TYPE

Exclusion: IS_EXCLUDED, EXCLUSION_REASON, EXCLUSION_DATE, IS_EXCLUDED_HIGH_RISK

Paiements: TOTAL_PAYMENT_AMOUNT, TOTAL_PAYMENTS, UNIQUE_PAYERS, RECIPIENT_TIER

Prescriptions: TOTAL_PRESCRIPTION_COST, TOTAL_PRESCRIPTION CLAIMS, PCT_BRAND CLAIMS

Flags: HAS_PHARMA_PAYMENTS, HAS_PRESCRIPTIONS, IS_NPI_ACTIVE, TOTAL_FINANCIAL_EXPOSURE

Calcul du Score de Fraude

Le modèle **fraud_risk_score** calcule un score composite de 0 à 100 basé sur plusieurs facteurs de risque:

| Facteur de risque | Points | Condition |
|------------------------|---------|---------------------------------|
| Exclusion OIG | +25 | Provider sur liste LEIE |
| Exclusion + Actif | +15 | Exclu mais toujours en activité |
| MEGA_RECIPIENT | +10 | Top tier paiements pharma |
| MAJOR_RECIPIENT | +5 | Paiements pharma élevés |
| Paiements haut risque | +10 max | PCT_HIGH_RISK_PAYMENTS ≥50% |
| Prescriptions marque | +10 max | PCT_BRAND CLAIMS ≥80% |
| Drugs high-risk | +10 max | TOTAL_HIGH_RISK_DRUGS ≥10 |
| Double activité | +10 | Paiements ET prescriptions |
| NPI inactif + activité | +10 | NPI désactivé mais transactions |

Tiers de Risque

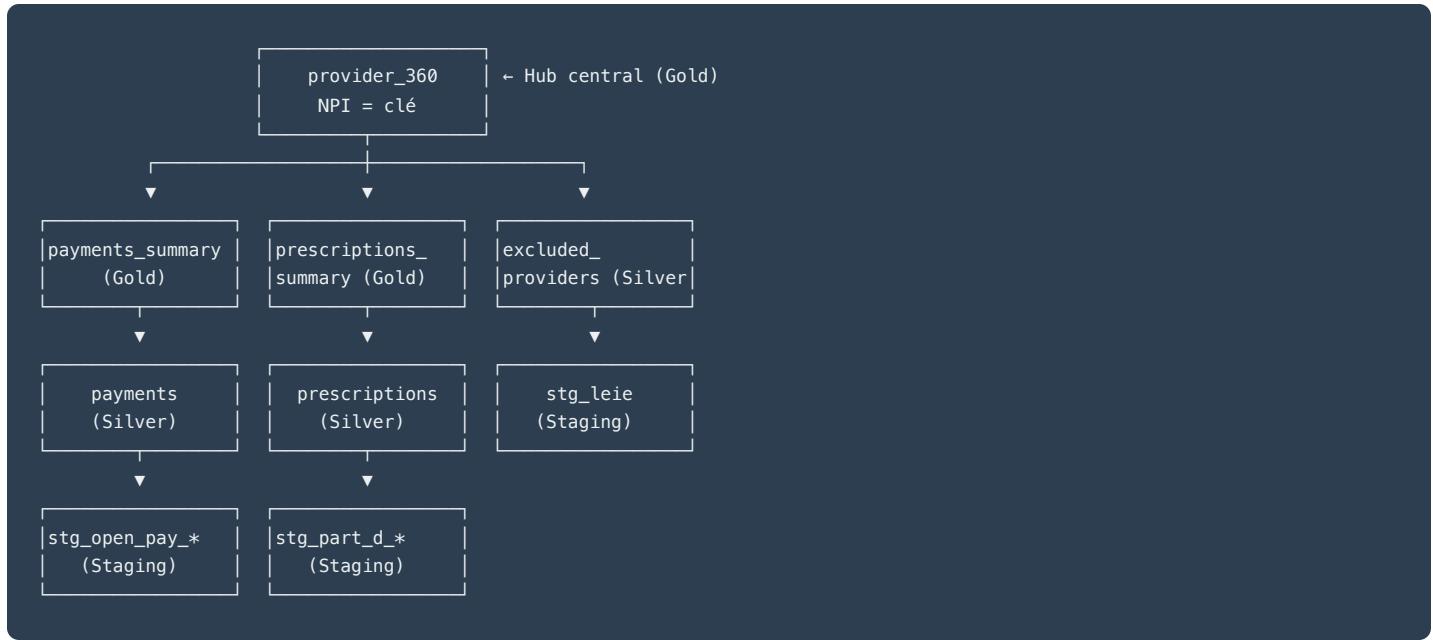
| Tier | Score | Action recommandée |
|----------|--------|-------------------------|
| CRITICAL | 70-100 | Investigation immédiate |
| HIGH | 50-69 | Revue prioritaire |
| MEDIUM | 30-49 | Surveillance accrue |
| LOW | 10-29 | Monitoring standard |
| MINIMAL | 0-9 | Aucune action requise |

High Risk Alerts

Le modèle `high_risk_alerts` génère des alertes actionnables par priorité:

| Priority | Alert Type | Description |
|----------|--------------------------|---|
| 1 | EXCLUDED_STILL_ACTIVE | Provider exclu mais encore actif dans NPPES |
| 2 | PRESCRIPTION_BY_EXCLUDED | Provider exclu a prescrit des médicaments |
| 3 | PAYMENT_TO_EXCLUDED | Provider exclu a reçu des paiements pharma |
| 4 | HIGH_RISK_SCORE | Provider avec score CRITICAL |
| 5 | HIGH_BRAND_PRESCRIBER | Taux de prescriptions marque >80% |

Relations entre Tables



Sources de Données

| Source | Tables Bronze | Fréquence | Usage Fraude |
|-------------------|--------------------|-----------|------------------------------------|
| OIG LEIE | LEIE | Mensuelle | Providers exclus Medicare/Medicaid |
| CMS Open Payments | OPEN_PAYMENTS_* | Annuelle | Conflits d'intérêts, kickbacks |
| Medicare Part D | MEDICARE_PART_D_* | Annuelle | Patterns prescriptions suspectes |
| NPPES | DIM_PROVIDER_* | Continue | Master data providers |
| CMS Facilities | HOSPICE, HHC, LTCH | Annuelle | Établissements à surveiller |