

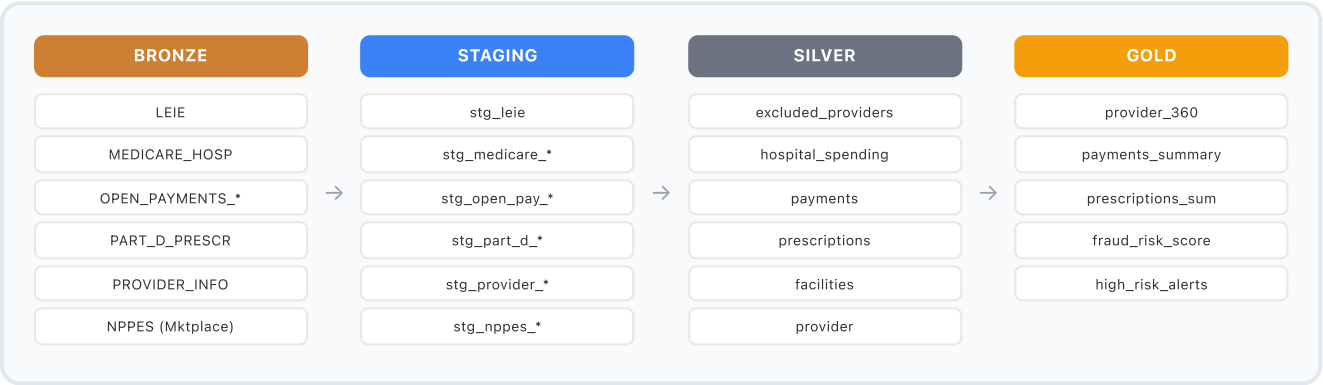
FraudLens - Modélisation

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



Couches de données

Bronze Tables brutes

Données brutes chargées depuis S3 (format Parquet). Aucune transformation, structure identique aux fichiers source.

Staging Vues dbt

Standardisation des noms de colonnes, cast des types. Matérialisées en vues (pas de stockage additionnel).

Silver Tables nettoyées

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

Gold Tables analytiques

Table	Description	Usage
provider_360	Vue 360° complète du provider	Dashboard, investigation
payments_summary	Agrégations paiements par NPI	Analyse pharma
prescriptions_summary	Agrégations prescriptions par NPI	Analyse Part D
fraud_risk_score	Score de risque (0-100)	Priorisation alertes
high_risk_alerts	Alertes actionnables	Investigation

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 utilisée par le dashboard et le scoring.

provider_360 (Gold - Table centrale)

Identité Provider

NPI (clé primaire), FULL_NAME, ORGANIZATION_NAME
ENTITY_TYPE (Individual/Organization), CREDENTIAL

Localisation

STATE, CITY, ZIP_CODE, PHONE

Spécialité

SPECIALTY, SPECIALTY_CLASSIFICATION, PROVIDER_TYPE

Statut Exclusion

IS_EXCLUDED, EXCLUSION_REASON, EXCLUSION_DATE
IS_EXCLUDED_HIGH_RISK (exclu mais encore actif)

Métriques Paiements Pharma

TOTAL_PAYMENT_AMOUNT, TOTAL_PAYMENTS, UNIQUE_PAYERS
RECIPIENT_TIER (MEGA, MAJOR, MODERATE, MINOR)
PCT_HIGH_RISK_PAYMENTS

Métriques Prescriptions

TOTAL_PRESCRIPTION_COST, TOTAL_PRESCRIPTION_CLAIMS
PCT_BRAND_CLAIMS, TOTAL_HIGH_RISK_DRUGS
PRESCRIBER_VOLUME_TIER

Flags Activité

HAS_PHARMA_PAYMENTS, HAS_PRESCRIPTIONS, IS_NPI_ACTIVE
TOTAL_FINANCIAL_EXPOSURE (paiements + prescriptions)

Calcul du Score de Fraude

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

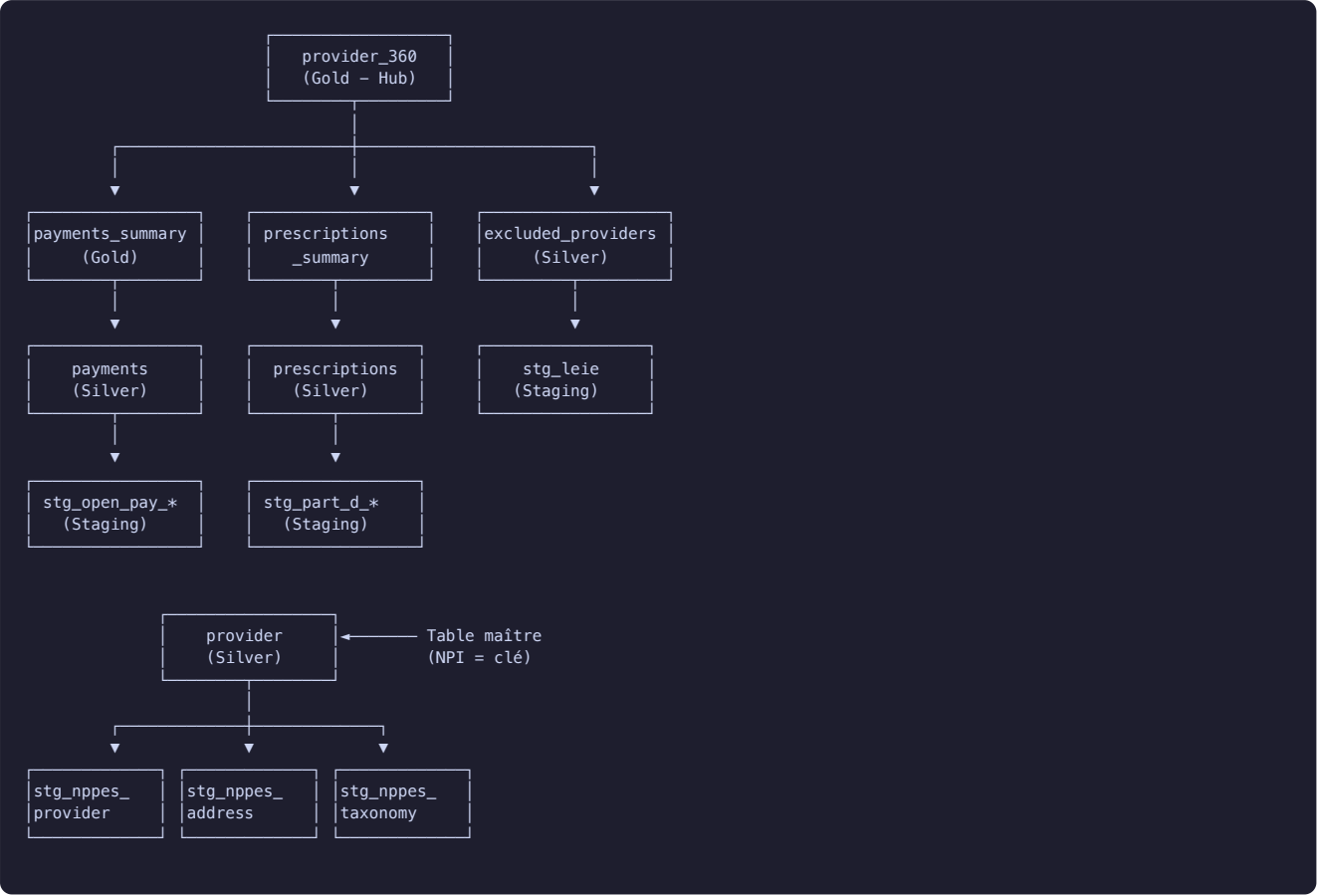
Facteur de risque	Points	Condition
Exclusion OIG	+25	Provider sur liste LEIE
Exclusion + Actif	+15	Exclu mais toujours activité
MEGA_RECIPIENT	+10	Top tier paiements pharma
MAJOR_RECIPIENT	+5	Paiements pharma élevés
Paiements haut risque	+10 max	PCT_HIGH_RISK_PAYMENTS ≥50%
Prescriptions marque	+10 max	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
---------	-----	-----------------------

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_PRESCRIBERS	Annuelle	Sur-prescription, pill mills
CMS Hospital Spending	MEDICARE_HOSPITAL_SPENDING	Annuelle	Dépenses anormales
CMS Provider Data	PROVIDER_INFO, HOSPICE, etc.	Trimestrielle	Établissements à risque
NPPES (Marketplace)	DIM_PROVIDER, DIM_ADDRESS	Temps réel	Référentiel NPI maître