

FraudLens Documentation DBT

Couche de Transformation des Données - Snowflake + dbt

3

Couches

20+

Modeles

80+

Tests

5

Sources

Architecture Medallion

Couche Bronze (Staging)

Ingestion brute des données. Sources : NPPES, LEIE, Medicare Part D, Open Payments, CMS.

Couche Silver (Nettoyée)

Données nettoyées, typees, enrichies. Jointures avec listes d'exclusion, métriques calculées.

PAYMENTS

PRESCRIPTIONS

FACILITIES

Couche Gold (Analytique)

Aggregations métier et KPIs.

PROVIDER_360

FRAUD_RISK_SCORE

HIGH_RISK_ALERTS

PROVIDER_ML_FEATURES

Materialisation Incrementale

| Modele | Cle | Stratégie |
|-------------------|-----------------|------------------------|
| prescriptions | PRESCRIPTION_ID | Filtre _LOAD_TIMESTAMP |
| hospital_spending | SPENDING_ID | Filtre _LOAD_TIMESTAMP |
| facilities | FACILITY_ID | Filtre _LOAD_TIMESTAMP |
| payments | PAYMENT_ID | Rebuild complet |

Clustering Snowflake

| Modele | Cles de Cluster |
|---------------|------------------------------|
| prescriptions | STATE, NPI |
| payments | STATE, NPI, PAYMENT_CATEGORY |

Note : Les sous-requêtes correlees avec MAX() ne sont pas supportées. Utiliser la macro Jinja `run_query`.

Scoring de Risque Fraude (0-100)

| Categorie | Max | Facteurs |
|---------------|-----|--|
| Exclusion | 40 | Liste LEIE (+25), Actif malgré exclusion (+15) |
| Paiements | 20 | Tier Mega/Major (+5-10), Paiements risques (+10) |
| Prescriptions | 20 | % marques élevées (+10), Médicaments risques (+10) |
| Activité | 20 | Paiements+Rx (+10), Inactif avec activité (+10) |

CRITIQUE (70-100)

Investigation immédiate

ELEVE (50-69)

Revue prioritaire

MOYEN (30-49)

Préoccupation modérée

FAIBLE (0-29)

Surveillance standard

Modeles Gold

| Modele | PK | Description |
|-----------------------|----------|--|
| provider_360 | NPI | Profil complet du provider avec paiements, prescriptions, statut d'exclusion |
| fraud_risk_score | NPI | Score de risque fraude composite (echelle 0-100) |
| high_risk_alerts | ALERT_ID | Alertes actionnables pour investigation |
| payments_summary | NPI | Metriques agregees des paiements pharma par provider |
| prescriptions_summary | NPI | Metriques agregees des prescriptions par provider |
| provider_ml_features | NPI | Features ML pour detection d'anomalies (voir ML_Features_doc.pdf) |

Types d'Alertes

| Type d'Alerte | Priorite | Declencheur |
|--------------------------|----------|--|
| EXCLUDED_STILL_ACTIVE | 1 | Provider sur LEIE mais toujours actif dans NPPES |
| PRESCRIPTION_BY_EXCLUDED | 2 | Provider exclu a des prescriptions Medicare |
| PAYMENT_TO_EXCLUDED | 3 | Provider exclu recoit des paiements pharma |
| HIGH_RISK_SCORE | 4 | Tier de risque CRITIQUE ou ELEVE |
| HIGH_BRAND_PRESCRIBER | 5 | >80% prescriptions de marque (indicateur kickback) |

Tests de Qualite des Donnees

Types de Tests

Tests de Schema

```
unique, not_null, accepted_values, relationships
```

Tests de Plage

`dbt_utils.accepted_range` pour les bornes

SQL Personnalise

Logique métier dans `tests/`

Tests Personnalises

Test

```
assert_critical_risk_providers_are_excluded
```

```
assert_no_negative_financial_amounts
```

```
assert_all_alerts_have_valid_npi
```

```
assert_financial_exposure_calculation
```

Commandes Courantes

```
dbt run                      # Executer tous les modeles
dbt run --select gold         # Executer une couche spécifique
dbt run --select prescriptions --full-refresh # Refresh complet incremental
dbt test --select gold        # Executer les tests
dbt docs generate && dbt docs serve      # Documentation
```