

Diagramme Technique – DFD Pipeline ETL

SmartShop 360

23 février 2026

Objectif

Détailler le flux de données de bout en bout :

Source → Nettoyage → Mapping/Unification → Base finale

Étapes du pipeline

1. Détection incrémentale : `should_run_etl()` compare les hashes des sources.
2. Ingestion/Nettoyage transactions : suppression annulations, quantités/prix invalides, calcul **Revenue** et **Margin**.
3. Ingestion avis : normalisation du sentiment et génération des notes.
4. Qualité : exécution des règles `run_all_validations()` sur transactions, avis et produits.
5. Mapping/Unification MDM : `build_product_mapping()` avec TF-IDF (seuil 0.35) et fallback rank.
6. Enrichissement avis : `attach_product_to_reviews()` + `ReviewDate`.
7. Chargement PostgreSQL : `products`, `customers`, `product_mapping`, `sales_facts`, `review_facts`.
8. Publication des vues analytiques : `v_product_kpi`, `v_customer_kpi`, `v_alerts`, `v_data_quality`.

Sorties principales

- Base finale relationnelle : PostgreSQL 16
- Vues KPI prêtes pour UI et Agent IA
- Historique incrémental des sources : `.etl_hashes.json`

Diagramme

