

Documentation Urbanistique & Métier

SmartShop 360

23 février 2026

Table des matières

1 Stratégie MDM (Master Data Management)

1.1 Approche implémentée dans le projet

Le projet met en place une table pivot `product_mapping` qui joue le rôle de *golden record* produit entre la source ERP et la source avis. Le mapping repose sur :

- normalisation des libellés dans le pipeline ETL,
- rapprochement fuzzy TF-IDF (`char_wb`, n-grammes 2 à 4),
- seuil de similarité cosinus (0.35),
- fallback `rank` quand aucun match n'atteint le seuil,
- conservation du `MatchScore` et de `MatchStrategy` (traçabilité).

1.2 Approche recommandée en contexte réel

En production, la stratégie cible est hiérarchisée :

1. **EAN/GTIN** comme clé de référence quand disponible (matching déterministe),
2. **fuzzy matching** (Levenshtein + TF-IDF) quand le code universel est absent,
3. **matching sémantique** par embeddings pour cas difficiles,
4. revue humaine des cas ambigus,
5. traçabilité systématique (règle, score, date, validateur) et boucle d'amélioration.

2 Dictionnaire de données (tables clés)

2.1 Product _ Mapping (pivot MDM)

Table de référence de rapprochement :

- `ERP_StockCode` et `Review_ProductCode`,
- nom unifié produit,
- catégorie,
- score et stratégie de matching.

2.2 Products

Table de référence produits ERP (Top 50 dans le POC) :

- `ProductID`, `ProductName`, `Category`

2.3 Customers

Table clients ERP :

- `ClientID`, `Nom`, `Pays`

2.4 Sales_Facts

Table de faits des ventes :

- granularité transactionnelle (facture/ligne),
- quantités, chiffre d'affaires, marge,
- clés vers produits et clients.

2.5 Review_Facts

Table de faits des avis :

- identifiant produit côté avis,
- note, sentiment, texte d'avis,
- date d'avis.

2.6 Vues analytiques

- v_product_kpi (CA, note, avis, quantités)
- v_customer_kpi (CA client, panier moyen, commandes)
- v_alerts (statuts CRITIQUE/A_SURVEILLER/OK)
- v_data_quality (couverture et volumétrie)

3 Scénario de démonstration (vidéo 5 à 10 min)

3.1 Script recommandé

1. **Contexte (30s)** : présenter les deux silos et l'absence d'identifiant commun.
2. **Ingestion ETL (2-3 min)** : lancer `streamlit run app.py` (ETL auto si base vide).
3. **Validation (1 min)** : montrer `product_mapping + v_product_kpi`.
4. **Question complexe à l'agent (2-3 min) :**
“Quels produits vendus à plus de 50 exemplaires ont une note inférieure à 3 ?”
5. **Analyse de réponse (1-2 min)** : montrer le SQL généré, le résultat, et la décision métier.
6. **Robustesse (30s)** : démontrer le fallback hors-ligne (sans clé API) si nécessaire.

3.2 Critères de réussite

- cohérence entre les sources et les résultats,
- explicabilité des règles MDM,
- capacité de l'agent à répondre avec des données vérifiables,
- traçabilité : SQL généré + vues/colonnes utilisées.