

# Analyse des besoins métier

Groupe 14

18 novembre 2025

Vinted



ALJIBBAOUI Tala  
AMMARI Reda  
SOUFYANI Anouar

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Analyse des besoins métier</b>	<b>2</b>
2.1	Analyse du cas Vinted . . . . .	2
2.2	Informations utiles pour la prise de décision . . . . .	2
2.3	Actions importantes à tracer . . . . .	2
2.4	Traitements analytiques possibles . . . . .	2
2.5	Classement des actions par importance . . . . .	3
2.6	Actions retenues pour la modélisation . . . . .	3
<b>3</b>	<b>Conception du premier modèle détaillé</b>	<b>3</b>
3.1	Conception du Data-Mart des Ventes . . . . .	3
3.2	Dimensions du modèle . . . . .	4
3.3	Mesures de la table des faits . . . . .	5
3.4	Validation du modèle par rapport au besoin métier . . . . .	5
3.5	Instance de l'entrepôt et estimation de la volumétrie . . . . .	6
3.5.1	Exemple d'instance . . . . .	6
3.5.2	Estimation de la taille du datamart sur 12 mois (France uniquement) . . .	6
3.5.3	Justification de l'utilisation d'un entrepôt de données . . . . .	7
3.5.4	Vérification des mesures du datamart . . . . .	7
<b>4</b>	<b>Conception des data-marts secondaires</b>	<b>7</b>
4.1	Data-mart Snapshot : suivi du catalogue d'annonces . . . . .	7
4.2	Data-mart Updated-Records : suivi des livraisons . . . . .	9
4.3	Réponse au besoin métier . . . . .	11
<b>5</b>	<b>Conception – Techniques avancées de modélisation</b>	<b>11</b>
5.1	Table-pont pour la hiérarchie des catégories . . . . .	11
5.2	Analyse de la dimension la plus volumineuse : D.article . . . . .	12
5.3	Partitionnement hybride de la dimension Article . . . . .	13

# 1 Introduction

Dans ce projet, nous avons décidé de travailler sur Vinted. C'est une plateforme que nous utilisons tous et qui génère beaucoup de données chaque jour. Nous voulions comprendre comment ces informations pouvaient être organisées dans un entrepôt de données et comment elles pouvaient être utilisées pour analyser le fonctionnement de la plateforme.

Le projet suit les cinq grandes étapes demandées dans le module : l'analyse du besoin métier, la conception du data-mart principal, la construction de deux modèles secondaires, l'application de techniques avancées vues en cours et enfin l'implémentation et les requêtes. Nous avons essayé de rester proches du fonctionnement réel de Vinted tout en respectant les contraintes de modélisation d'un entrepôt de données.

## 2 Analyse des besoins métier

### 2.1 Analyse du cas Vinted

Vinted est aujourd'hui l'une des plateformes de revente d'articles d'occasion les plus utilisées en Europe. Elle fonctionne entièrement en C2C, ce qui veut dire que les utilisateurs publient eux-mêmes leurs articles et gèrent leurs ventes. La plateforme sert d'intermédiaire pour les paiements et la livraison. Elle ne possède donc aucun stock, mais elle doit suivre un volume important d'utilisateurs, d'annonces et de transactions.

Le modèle économique repose surtout sur deux éléments : les frais de protection acheteur, ajoutés à chaque achat, qui couvrent la sécurité du paiement et la gestion des litiges et les options de visibilité payantes pour les vendeurs. Par exemple, un vendeur peut utiliser un Boost pour remonter un article en haut des résultats pendant quelques heures, ou un Wardrobe Spotlight pour mettre en avant plusieurs articles de sa penderie pendant plusieurs jours.

En plus de ça, Vinted s'appuie sur des partenaires logistiques comme Mondial Relay, Colissimo ou Relais Colis pour l'envoi des colis. La qualité de ces livraisons a un impact direct sur la satisfaction des utilisateurs et sur la réputation de la plateforme.

### 2.2 Informations utiles pour la prise de décision

Pour améliorer son fonctionnement, Vinted doit pouvoir répondre à plusieurs questions simples :  
Quels types d'articles se vendent rapidement ?

Quels vendeurs sont les plus fiables ?

Quels services de visibilité fonctionnent le mieux ?

Quels transporteurs livrent dans les délais ?

etc...

La plateforme doit aussi comprendre quelles zones géographiques sont les plus actives et où les livraisons posent le plus de problèmes.

Ces informations ne sont pas centralisées dans les bases opérationnelles, car elles sont réparties entre les annonces, les profils utilisateurs, les transactions et les systèmes de livraison. Un entrepôt de données permettrait d'obtenir une vision globale et d'analyser l'activité sur plusieurs mois, ce qui est difficile à faire directement dans les systèmes internes.

### 2.3 Actions importantes à tracer

Nous avons retenu trois actions essentielles pour le projet. La mise en vente d'un article permet de suivre le stock visible sur la plateforme, ainsi que la façon dont les utilisateurs fixent leurs prix ou utilisent les options de visibilité. La vente d'un article est l'action principale, puisqu'elle génère les revenus de la plateforme et reflète le comportement des acheteurs et vendeurs. La livraison associée à la vente est aussi importante, car elle regroupe plusieurs étapes logistiques et fait intervenir des transporteurs externes avec des performances différentes.

### 2.4 Traitements analytiques possibles

Chaque action peut être analysée de plusieurs manières. Pour la vente, il est possible d'observer la rapidité de vente selon les catégories, les marques ou les pays, de mesurer l'impact réel d'un Boost sur la durée avant la vente, ou de suivre le chiffre d'affaires et la commission générée par période. Pour la livraison, il est intéressant de comparer les délais réels entre transporteurs comme

Mondial Relay et Colissimo, d'identifier les zones où les retards sont les plus fréquents, ou de repérer les types de litiges les plus courants. Pour la mise en vente, il est possible de suivre le nombre d'annonces actives, les vues et favoris accumulés, ou la variation des prix affichés dans le temps. Ces analyses peuvent aider la plateforme à améliorer ses recommandations, à ajuster les tarifs des services payants, et à comparer la performance des transporteurs dans différents pays.

## 2.5 Classement des actions par importance

Parmi ces actions, la vente d'un article est la plus importante, car elle représente directement la source de revenus de Vinted. C'est aussi une action qui dépend fortement du comportement des utilisateurs et qui détermine la santé générale du système. La livraison arrive ensuite, car une mauvaise expérience à ce niveau peut entraîner des litiges et faire perdre des utilisateurs. La mise en vente est utile pour comprendre l'évolution du stock et le comportement des vendeurs, mais son impact est moins direct que celui des deux autres actions.

## 2.6 Actions retenues pour la modélisation

Pour la suite du projet, nous avons décidé de nous concentrer sur deux actions principales : la vente et la livraison. La vente servira de base au modèle détaillé en étoile, tandis que la livraison sera utilisée pour construire un modèle secondaire. La mise en vente sera étudiée avec un modèle snapshot pour suivre l'évolution du stock visible au fil du temps.

1. **Vente d'un article** : principale source de revenu (commission + services payants).
2. **Livraison** : impact majeur sur satisfaction et coûts.

# 3 Conception du premier modèle détaillé

## 3.1 Conception du Data-Mart des Ventes

L'action la plus importante identifiée dans l'analyse métier est la vente d'un article. C'est cette opération qui génère les revenus de la plateforme et qui dépend directement du comportement des utilisateurs. Nous avons donc construit un data-mart basé sur un schéma en étoile, centré sur une table de faits appelée F\_Ventes.

Ce data-mart a pour objectif de décrire chaque vente à l'aide d'un ensemble de dimensions (article, utilisateur, date, livraison, visibilité) et d'un groupe de mesures permettant d'analyser la performance commerciale, la rapidité des ventes et l'impact des services payants.

La table de faits regroupe toutes les informations propres à l'événement « vente », tandis que les dimensions fournissent le contexte nécessaire pour effectuer des analyses par type d'article, par période, par vendeur ou encore par transporteur.

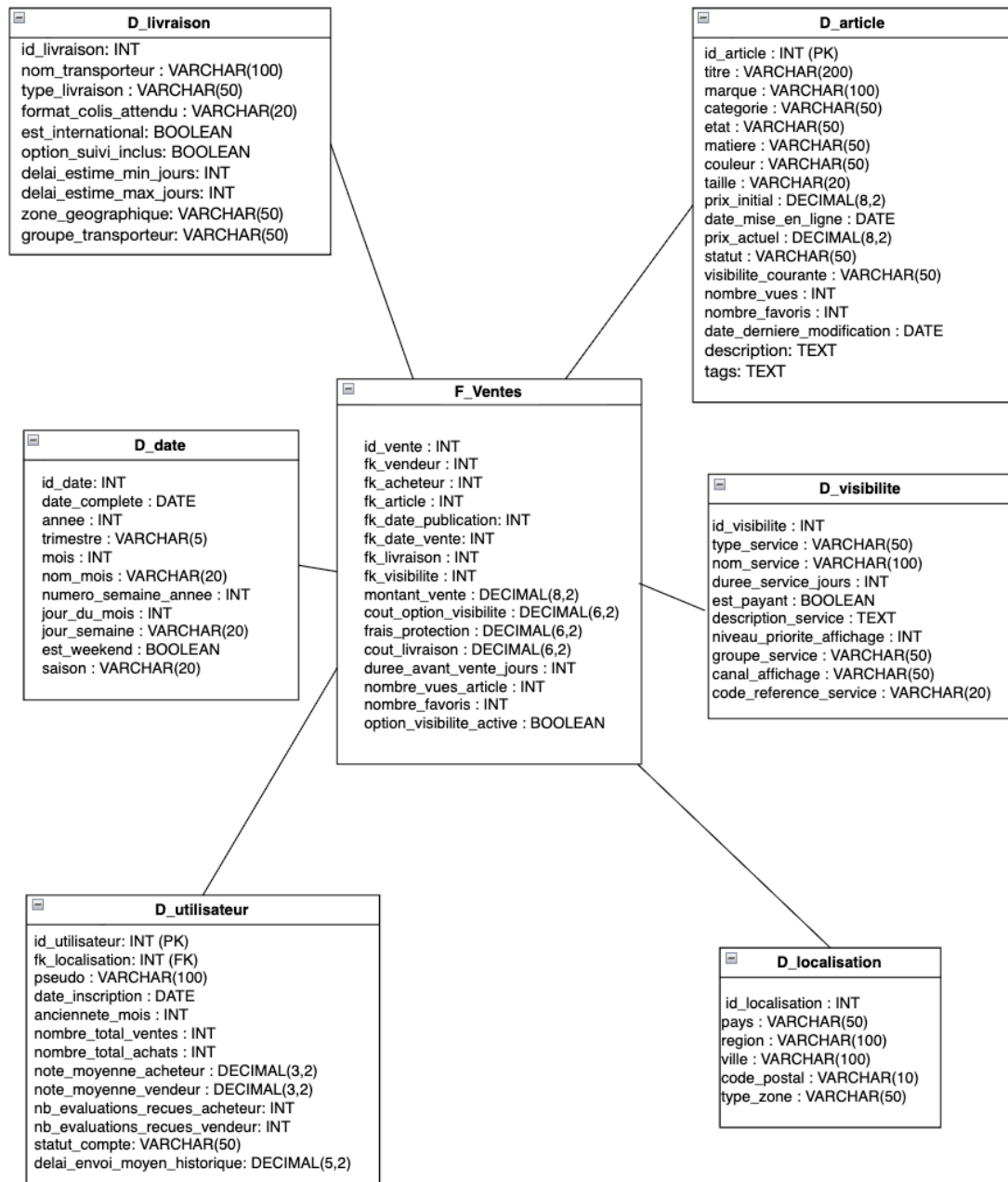


FIGURE 1 – Schéma en étoile du data-mart F\_Ventes.

### 3.2 Dimensions du modèle

Pour décrire chaque vente de manière complète, nous avons retenu cinq dimensions principales. Elles couvrent les informations sur l'utilisateur, l'article vendu, la date de la transaction, le mode de livraison et l'éventuel service de mise en avant utilisé. Ces dimensions apportent suffisamment de détails pour analyser les ventes sous différents angles tout en restant simples à exploiter.

**Dimension Utilisateur :** Cette dimension rassemble les informations sur les vendeurs et les acheteurs. On y retrouve par exemple le pseudo, la date d'inscription, l'ancienneté, le statut du compte, le nombre de ventes réalisées, le pays, la ville et la note moyenne reçue. Ces attributs permettent d'étudier le comportement des utilisateurs et de repérer les profils les plus actifs.

**Dimension Article :** L'article est décrit à partir d'une vue virtuelle construite sur les trois tables article de notre modèle (statiques, dynamiques et textes). Elle regroupe les attributs essentiels : catégorie, marque, couleur, matière, état, taille, prix initial, description courte et date de mise en ligne. Cette dimension permet d'analyser les ventes selon la typologie des produits.

**Dimension Date :** Elle contient les différentes informations temporelles nécessaires aux analyses : jour, mois, année, trimestre, saison, semaine, jour de la semaine et indicateur férié. Cette dimension sert aux analyses de saisonnalité et aux comparaisons entre périodes.

**Dimension Livraison :** Elle décrit la méthode d'envoi choisie lors de la vente : transporteur (Colissimo, Mondial Relay...), type de livraison (domicile ou point relais), zone couverte, délai estimé et format du colis. Cette dimension permet de comparer la performance des transporteurs et d'identifier les zones où les délais sont plus importants.

**Dimension Visibilité :** Elle regroupe les informations liées aux services de mise en avant disponibles sur la plateforme : type de service (standard, boost, spotlight), durée d'affichage, niveau de priorité, est-payant, canal d'affichage et code interne du service. Cette dimension est utile pour mesurer l'impact réel des options payantes sur la rapidité des ventes.

### 3.3 Mesures de la table des faits

Mesure	Type	Description	Non-additive sur
montant_vente	Additive	Montant payé par l'acheteur pour l'article.	–
frais_protection	Additive	Commission perçue par la plateforme sur la vente.	–
cout_option_visibilite	Additive	Montant payé pour une option payante (Boost, Spotlight).	–
cout_livraison	Additive	Coût du transport associé à la transaction.	–
duree_avant_vente_jours	Semi-additive	Nombre de jours entre la mise en ligne et la vente.	Date
nombre_vues_article	Additive	Nombre total de vues accumulées avant la vente.	–
nombre_favoris	Additive	Nombre d'ajouts aux favoris avant la vente.	–
option_visibilite_active	Non-additive	Indique si une option de visibilité était active (0/1).	–

TABLE 1 – Mesures de la table des faits F\_Ventes.

### 3.4 Validation du modèle par rapport au besoin métier

Le besoin métier principal était de comprendre ce qui influence la performance des ventes : la rapidité avec laquelle les articles se vendent, les revenus générés, l'intérêt des utilisateurs et l'efficacité des options de visibilité. Notre modèle en étoile permet de répondre directement à ces questions, car il rassemble dans une même table toutes les mesures liées à la transaction et les relie aux dimensions utiles pour l'analyse.

Grâce à la dimension Article, nous pouvons savoir quelles catégories, marques ou états se vendent le plus vite. La dimension Utilisateur permet de comparer les vendeurs entre eux ou d'identifier les profils les plus actifs. La dimension Visibilité nous aide à mesurer l'impact réel des options payantes sur la durée avant la vente ou sur le montant final. Avec la dimension Livraison, on peut aussi croiser les ventes avec le mode d'envoi choisi pour repérer les transporteurs qui offrent les meilleures performances. Enfin, la dimension Date simplifie les analyses temporelles comme la saisonnalité ou les pics d'activité.

L'ensemble du modèle couvre donc toutes les informations nécessaires pour analyser les ventes de façon fiable et complète. Il permet de suivre les revenus, de comprendre le comportement des utilisateurs et d'évaluer les services proposés par la plateforme. C'est exactement ce qui est attendu pour répondre au besoin métier identifié dans la première partie.

### 3.5 Instance de l’entrepôt et estimation de la volumétrie

#### 3.5.1 Exemple d’instance

Pour vérifier la cohérence du modèle, nous présentons un petit exemple contenant quelques lignes dans les dimensions et dans la table de faits. Nous ne montrons que les attributs nécessaires au lien avec la table **F\_VENTES**.

TABLE 2 – Exemple – D\_date

id_date	date_complete	annee	mois	jour_du_mois
20240110	10/01/2024	2024	1	10
20240115	15/01/2024	2024	1	15

TABLE 3 – Exemple – D\_utilisateur

id_utilisateur	pseudo	statut_compte	date_inscription
101	VendeurA	particulier	12/03/2021
205	AcheteurB	particulier	05/07/2022

TABLE 4 – Exemple – D\_article

id_article	categorie	marque	prix_initial	etat
5012	Robe	Zara	18.00	tres_bon
8841	Jean	Levi's	25.00	bon

TABLE 5 – Exemple – D\_livraison

id_livraison	nom_transporteur	type_livraison
1	Mondial Relay	relais
2	Colissimo	domicile

TABLE 6 – Exemple – D\_visibilite

id_visibilite	type_service	est_payant
1	standard	non
2	boost	oui

TABLE 7 – Exemple – D\_localisation

id_localisation	pays	region	ville
10	France	Île-de-France	Paris
11	Belgique	Bruxelles-Capitale	Bruxelles

TABLE 8 – Exemple – F\_Ventes

id_vente	fk_vendeur	fk_acheteur	fk_article	fk_date_vente	fk_livraison	fk_visibilite	montant_vente
1	101	205	5012	20240115	1	2	18.00
2	101	205	8841	20240115	2	1	25.00

#### 3.5.2 Estimation de la taille du datamart sur 12 mois (France uniquement)

Le grain de la table **F\_VENTES** est une vente finalisée. En France, Vinted compte environ **27 millions d'utilisateurs** (*source : Les Échos Publishing, 2024*).

Une étude *Joko* (basée sur 700 000 transactions bancaires) indique qu’un utilisateur français réalise en moyenne **11 achats par an**.

Ainsi, le volume annuel de transactions est estimé à :

$$27\,000\,000 \times 11 = 297\,000\,000 \text{ ventes/an}$$

La table **F\_Ventes** contiendra donc environ **297 millions de lignes** sur 12 mois (France).

Pour les dimensions principales :

- **D\_utilisateur** : environ 27 millions de lignes (nombre d'inscrits en France).
- **D\_article** : selon *BFM TV*, environ **400 000 nouvelles annonces sont publiées chaque jour** (France, 2019).

$$400\,000 \times 365 \approx 146\,000\,000$$

- **D\_date** : 365 lignes/an.
- **D\_livraison** et **D\_visibilite** : moins de 50 lignes chacune.

### 3.5.3 Justification de l'utilisation d'un entrepôt de données

Excel est limité à **1 048 576 lignes**. Avec **297 millions de lignes** rien que pour les ventes annuelles en France, cette limite serait dépassée par :

$$\frac{297\,000\,000}{1\,048\,576} \approx 283$$

Même une seule journée d'activité (environ 800 000 ventes estimées) remplirait presque entièrement une feuille Excel.

Les volumes, la nécessité d'historiser, et les jointures complexes entre dimensions justifient pleinement l'utilisation d'un entrepôt de données dédié.

### 3.5.4 Vérification des mesures du datamart

Nous avons vérifié que toutes les mesures présentes dans la table F\_Ventes correspondent bien au grain du modèle, qui est une vente finalisée. Elles décrivent toutes l'événement lui-même et non une dimension, donc elles sont bien à leur place dans la table des faits.

Les mesures financières et les compteurs (*montant\_vente*, *frais\_protection*, *cout\_livraison*, *cout\_option\_visibilite*, *nombre\_vues\_article*, *nombre\_favoris*) sont additives, ce qui permet de les additionner sur n'importe quelle dimension.

La mesure *duree\_avant\_vente\_jours* est semi-additive, car on peut la moyenner mais pas l'additionner sur la dimension Date. L'indicateur *option\_visibilite\_active* est non-additif puisqu'il s'agit d'un booléen.

Aucune dimension ne contient de mesure ou de valeur calculée : chaque dimension ne stocke que des attributs descriptifs. L'ensemble des mesures est donc cohérent avec le grain du datamart et respecte les règles du schéma en étoile.

## 4 Conception des data-marts secondaires

Dans cette partie, nous construisons deux data-marts complémentaires à celui des ventes : un modèle *snapshot* et un modèle *updated-records*. Le snapshot sert à enregistrer régulièrement l'état d'un ensemble d'objets (comme un inventaire), tandis que le modèle updated-records suit un processus qui avance en plusieurs étapes.

Même si la vente et la livraison sont les actions principales de la plateforme, aucune des deux ne correspond vraiment à la logique snapshot : la vente est un événement ponctuel, et la livraison suit un déroulement étape par étape. Nous avons donc choisi une action secondaire mieux adaptée : le suivi quotidien des annonces visibles. La livraison, en revanche, se prête parfaitement au modèle updated-records.

### 4.1 Data-mart Snapshot : suivi du catalogue d'annonces

#### Objectif métier

Même si la vente reste l'action centrale sur Vinted, il est important de suivre l'évolution du stock d'articles encore visibles. Le data-mart snapshot permet d'enregistrer, chaque jour, l'état global des annonces selon plusieurs axes : la catégorie, la visibilité, l'ancienneté ou encore le statut. Cela permet par exemple de connaître le nombre d'annonces actives, les prix moyens, ou les types d'articles qui restent en ligne le plus longtemps.

## Grain du modèle

Le grain retenu est un état agrégé des annonces par date et par type d'article. Nous ne stockons donc pas une ligne par article, mais une ligne par "groupe" d'annonces (exemple : toutes les robes actives le 20 octobre). Cela réduit fortement la volumétrie tout en gardant les informations utiles.

## Dimensions du modèle

Le snapshot utilise cinq dimensions :

- Date (jour, mois, année, saison...)
- Type d'article (catégorie, état, marque...)
- Visibilité (standard, boost...)
- Ancienneté (tranches de jours depuis la publication)
- Statut (actif, réservé, vendu...)

Chaque dimension contient une dizaine d'attributs descriptifs, ce qui permet d'obtenir des analyses détaillées sans complexifier le modèle.

## Table de faits et mesures

La table des faits contient les mesures suivantes afin de représenter l'état du catalogue à un instant donné :

- nombre\_annonces : additive
- prix\_moyen, prix\_min, prix\_max : semi-additives (non additionnables sur la dimension Date)
- nombre\_vues, nombre\_favoris : semi-additives

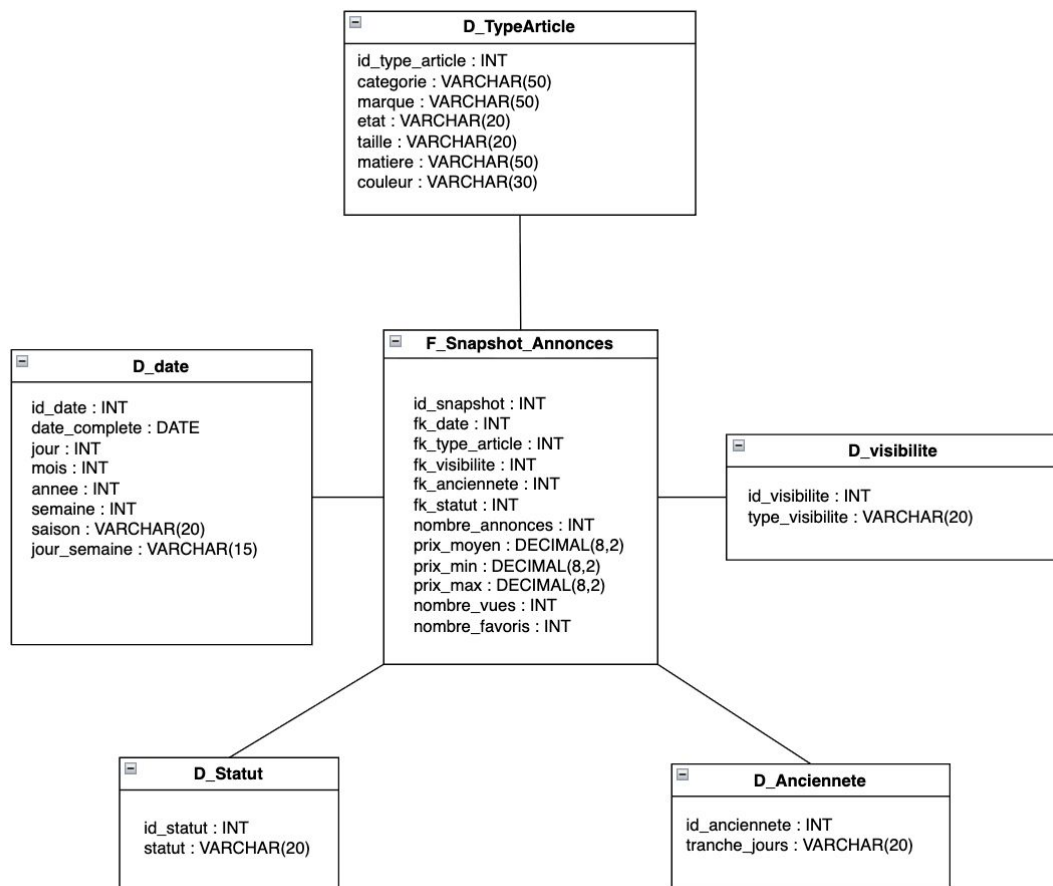


FIGURE 2 – Schéma du modèle snapshot F\_Snapshot\_Annonces.



## Exemple d'instance

date	type_article	visibilite	anciennete	statut	prix_moyen	vues
20/10	Robes	Standard	0-7 j	Actif	18.5	920
21/10	Robes	Boost	0-7 j	Actif	19.3	1240

## Estimation de taille

En supposant environ 5 000 types d'articles actifs par jour :

$$5000 \times 365 = 1\,825\,000 \text{ lignes par an}$$

Excel étant limité à 1 048 576 lignes par feuille, un snapshot annuel complet ne tiendrait pas dans un fichier. Un entrepôt de données est donc nécessaire pour stocker et analyser ces données correctement.

## 4.2 Data-mart Updated-Records : suivi des livraisons

### Objectif métier

La livraison est la deuxième action la plus importante sur Vinted après la vente. Elle suit un processus structuré comprenant plusieurs étapes : validation de la vente, préparation par le vendeur, envoi, transit et réception. Le modèle *updated-records* est parfaitement adapté à ce type de processus, car il permet de mettre à jour la même ligne de faits à chaque nouvelle étape. Comme expliqué dans le cours, ce modèle repose sur le principe : « *One fact table row contains all the movements of one product, and this row is updated repeatedly until the process is complete.* »

Ce data-mart permet ainsi de mesurer les délais de préparation, les temps de transit, les retards récurrents selon les zones géographiques, et la qualité globale de la livraison à travers la note donnée par l'acheteur.

### Grain du modèle

Le grain est une ligne par livraison, créée à la vente puis complétée au fur et à mesure.

### Dimensions du modèle

Les dimensions retenues sont les suivantes :

- **Utilisateur** : informations sur le vendeur et l'acheteur.
- **Article** : caractéristiques descriptives du produit expédié (catégorie, marque, poids, matière...).
- **Livraison** : informations stables sur le mode d'envoi (transporteur, type de livraison, zone, délai estimé...).
- **Localisation** : localisation géographique associée à l'utilisateur.
- **Transaction** : identifiant de la vente et informations associées.
- **Date** : dimension temporelle partagée dans l'ensemble de l'entrepôt de données.

À noter que les dates propres aux étapes du processus (vente, envoi, réception) sont stockées directement dans la table de faits et non via plusieurs clés étrangères vers la dimension Date.

### Table de faits et mesures

La table contient les mesures suivantes :

- `cout_livraison` (additive)
- `delai_preparation_jours` (additive)
- `delai_transit_jours` (additive)
- `delai_total_jours` (additive)
- `note_livraison` (non-additive)
- `quantite_expediee` (additive)

Les délais sont calculés lorsque les dates manquantes sont renseignées.

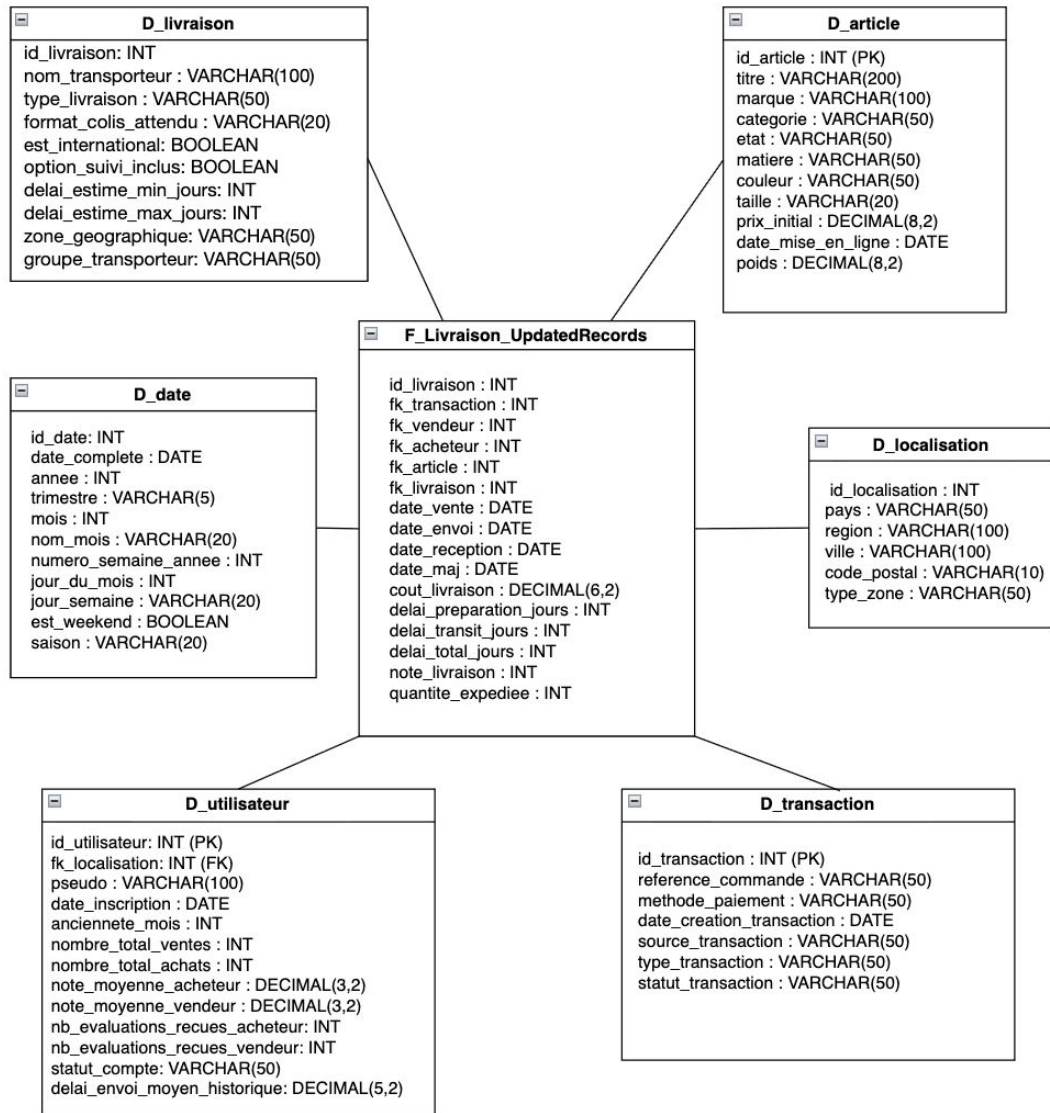


FIGURE 3 – Schéma du modèle Updated-Records F\_Livraison\_UpdatedRecords.

### Exemple d'instance

vendeur	acheteur	transporteur	vente	envoi	reception	delai_prepa	delai_transit
101	205	Mondial Relay	20/10	22/10	27/10	2	5
102	330	Colissimo	23/10	27/10	—	4	—

### Estimation de taille

Comme chaque vente déclenche une livraison, la taille du data-mart est comparable à celle des ventes :

≈ 255 000 000 lignes par an

Un tel volume dépasse largement les capacités d'un fichier Excel et nécessite une architecture d'entrepôt de données capable de gérer des mises à jour fréquentes.

### 4.3 Réponse au besoin métier

Les deux modèles secondaires complètent le data-mart principal. Le snapshot fournit une vision claire de l'évolution du catalogue, tandis que le modèle updated-records permet de suivre précisément le processus de livraison. Ensemble, ils couvrent les besoins liés au stock, à la logistique, à la visibilité des articles et à la qualité de service.

## 5 Conception – Techniques avancées de modélisation

Dans cette partie, nous appliquons trois techniques vues en cours : l'utilisation d'une table-pont pour gérer une relation hiérarchique, la prise en compte de l'évolution des dimensions, et enfin un schéma de partitionnement pour optimiser les lectures.

### 5.1 Table-pont pour la hiérarchie des catégories

Sur Vinted, les catégories sont organisées de manière hiérarchique et peuvent atteindre plusieurs niveaux (par exemple : *Women* → *Clothing* → *Dresses* → *Special-occasion Dresses* → *Prom dresses*). Comme la profondeur n'est pas la même pour tous les articles, il n'est pas adapté de créer plusieurs colonnes fixes (catégorie 1, catégorie 2, etc.). Pour modéliser proprement cette hiérarchie sans modifier le schéma en étoile, nous utilisons une table-pont.

La table-pont permet de gérer proprement la profondeur variable de la hiérarchie des catégories, en évitant les colonnes fixes (catégorie1, catégorie2...), et en permettant d'exploiter toutes les catégories dans les requêtes analytiques, quel que soit leur niveau.

#### Structure de la table-pont

Attribut	Description
id_article	Identifiant de l'article (FK vers D_article)
id_categorie	Identifiant de la catégorie (FK vers D_categories)
niveau_hierarchie	Niveau hiérarchique (1 = plus haut, jusqu'à 5)

TABLE 9 – Table-pont `bridge_article_categorie`

Chaque article peut donc apparaître plusieurs fois dans cette table (une ligne par niveau de catégorie), ce qui permet de gérer à la fois les catégories principales et les sous-catégories sans modifier la dimension des articles.

#### Dimension D\_categories

Attribut	Description
id_categorie	Identifiant unique de la catégorie
nom_categorie	Nom de la catégorie (Dresses, Shoes, etc.)
categorie_parent	Identifiant de la catégorie parente (ou NULL)
type_categorie	Niveau (racine, parent, enfant)

TABLE 10 – Dimension D\_categories

#### Schéma simplifié de la table-pont

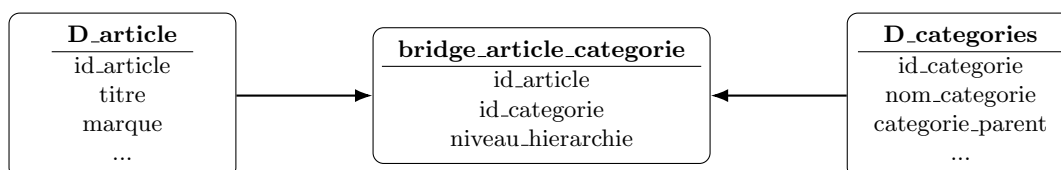


FIGURE 4 – Table-pont entre articles et catégories

#### Exemple de requête analytique

Cette requête calcule le chiffre d'affaires par catégorie, en utilisant la table-pont.

```

1 SELECT
2     c.nom_categorie,
3     SUM(v.montant_vente) AS total_ventes
4 FROM f_ventes v
5 JOIN bridge_article_categorie b
6     ON v.fk_article = b.id_article
7 JOIN d_categories c
8     ON b.id_categorie = c.id_categorie
9 GROUP BY c.nom_categorie
10 ORDER BY total_ventes DESC;

```

Cette requête fonctionne pour n'importe quel niveau de la hiérarchie, sans modifier le modèle principal.

## 5.2 Analyse de la dimension la plus volumineuse : D\_article

Dans notre entrepôt de données, la dimension la plus volumineuse est la dimension D\_article, car la plateforme enregistre plusieurs centaines de milliers de nouvelles annonces chaque jour. Comme expliqué dans le cours, lorsqu'une dimension contient beaucoup d'attributs, il est important de distinguer ceux qui ne changent jamais (statiques) de ceux qui peuvent évoluer au cours du temps (dynamiques). Cela permet ensuite d'appliquer la bonne stratégie de gestion des mises à jour (type 1, type 2 ou type 3).

**Attributs statiques** Les attributs suivants décrivent l'article lui-même et ne sont jamais modifiés par l'utilisateur ou par le système :

- id\_article
- titre
- marque
- categorie
- etat
- matiere
- couleur
- taille
- prix\_initial
- date\_mise\_en\_ligne
- description
- tags

**Attributs dynamiques** Ces attributs évoluent dans le temps et reflètent l'activité autour de l'annonce :

- prix\_actuel
- statut
- visibilite\_courante
- nombre\_vues
- nombre\_favoris
- date\_derniere\_modification

Les attributs dynamiques d'une dimension volumineuse doivent être gérés à l'aide des techniques SCD (Slowly Changing Dimensions). Ces stratégies permettent de conserver l'historique uniquement là où cela a un intérêt analytique, tout en évitant d'alourdir la dimension inutilement.

Pour chaque attribut dynamique de la dimension *D\_article*, nous indiquons ci-dessous un exemple de mise à jour et la stratégie SCD la plus adaptée.

### prix\_actuel

- **Exemple de mise à jour** : un vendeur baisse un article de 20€ à 15€.
- **Stratégie : Type 2**. Conserver l'historique des changements de prix est important pour analyser la durée avant vente, les comportements de négociation ou l'efficacité des baisses de prix.

### statut

- **Exemple de mise à jour** : une annonce passe de "Actif" à "Réservé", puis à "Vendu".

- **Stratégie : Type 2.** Le statut reflète le cycle de vie de l'annonce et il est essentiel de garder une trace de ces transitions pour analyser le temps passé dans chaque état.

#### visibilite\_courante

- **Exemple de mise à jour :** un vendeur active un boost pendant 3 heures.
- **Stratégie : Type 2.** Les périodes où l'annonce est boostée doivent être historisées afin d'évaluer l'impact des options payantes sur la vitesse de vente.

#### nombre\_vues

- **Exemple de mise à jour :** l'annonce passe de 120 à 135 vues.
- **Stratégie : Type 1.** Les vues sont des compteurs qui n'ont pas besoin d'historique complet. Seule la valeur la plus récente est utile.

#### nombre\_favoris

- **Exemple de mise à jour :** un nouvel utilisateur ajoute l'article à ses favoris.
- **Stratégie : Type 1.** Comme pour les vues, l'historique n'apporte pas de valeur analytique : seule la valeur actuelle compte.

#### date\_derniere\_modification

- **Exemple de mise à jour :** l'utilisateur modifie le prix ou change le statut.
- **Stratégie : Type 1.** Ce champ sert uniquement d'information technique pour suivre la synchronisation des données.

### 5.3 Partitionnement hybride de la dimension Article

La dimension *Article* est la plus volumineuse de l'entrepôt, car elle contient toutes les annonces publiées par les utilisateurs. Pour améliorer les performances de lecture et limiter la quantité de données scannées lors des requêtes analytiques, nous appliquons un schéma de partitionnement hybride, combinant un partitionnement horizontal et un partitionnement vertical.

#### Partitionnement horizontal

Le partitionnement horizontal consiste à diviser les lignes de la dimension selon un critère temporel. Nous avons choisi l'attribut *date\_mise\_en\_ligne*, car la majorité des analyses portent sur des périodes récentes (par exemple : articles mis en ligne en 2024).

Ainsi, la dimension est séparée en plusieurs partitions :

- **Article\_2022**
- **Article\_2023**
- **Article\_2024**
- **Article\_Archive**

Ce découpage permet de ne lire que la ou les partitions utiles. Par exemple, une requête portant sur l'année 2024 n'accède qu'à la partition **Article\_2024**, ce qui réduit fortement le volume de données lu.

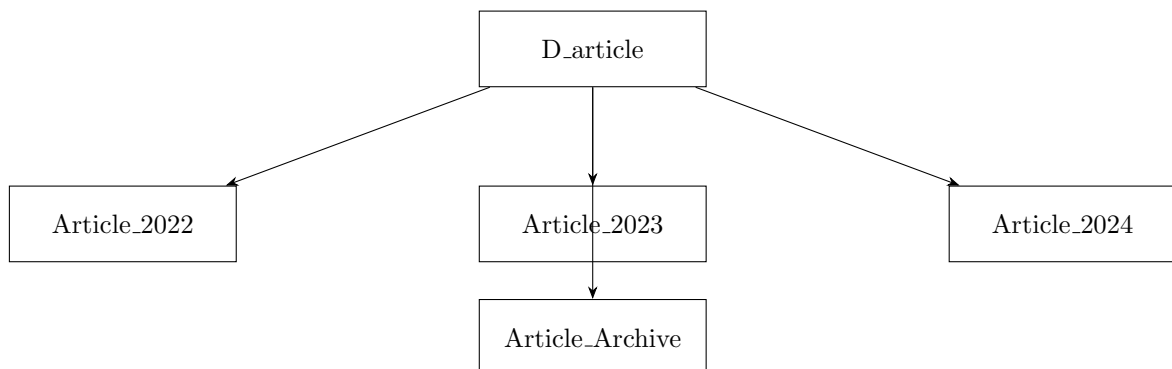


FIGURE 5 – Partitionnement horizontal de la dimension Article.

## Partitionnement vertical

Le partitionnement vertical consiste à séparer les colonnes selon leur évolutivité et leur rôle dans les requêtes analytiques. Dans notre cas, les attributs de la dimension se répartissent naturellement en deux groupes :

### Attributs statiques (peu ou pas modifiés)

- id\_article, marque, matiere, couleur, taille
- description, tags

### Attributs dynamiques (susceptibles de changer)

- prix\_actuel, statut, visibilite\_courante
- nombre\_vues, nombre\_favoris
- date\_derniere\_modification

Les attributs statiques sont regroupés dans une table **Article\_Static**, tandis que les attributs dynamiques sont stockés dans **Article\_Dynamic**. Cette séparation évite de lire des colonnes volumineuses (par exemple **description**) lorsque les requêtes portent uniquement sur des valeurs dynamiques comme les prix ou la visibilité.

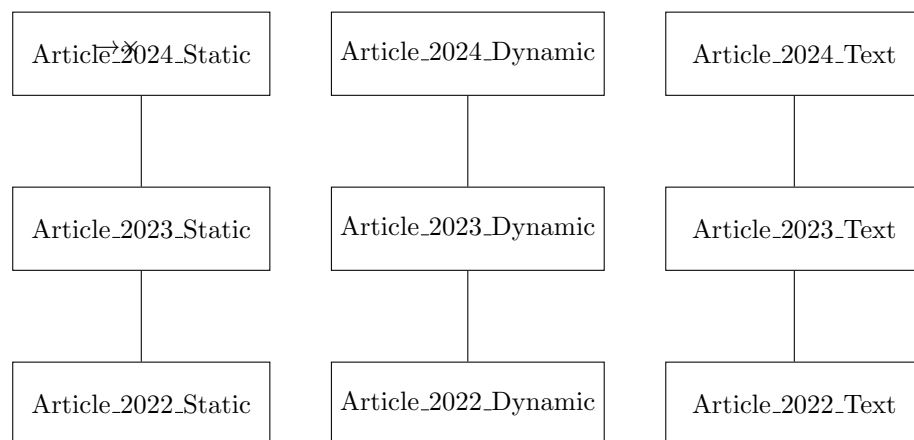
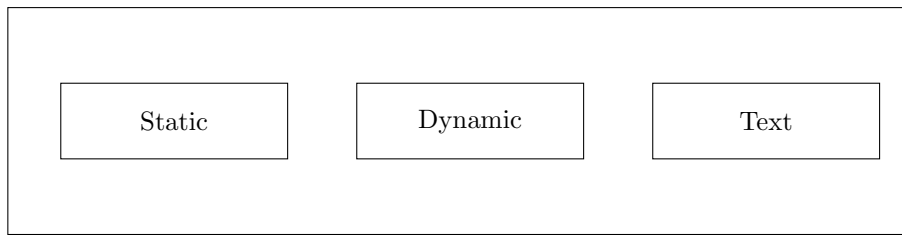


FIGURE 6 – Partitionnement hybride de la dimension Article.

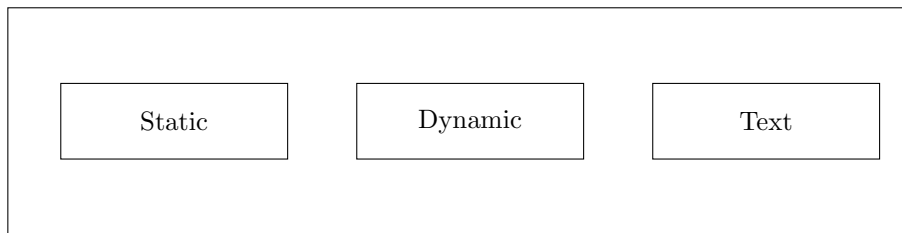
## Structure finale du partitionnement hybride

Le modèle hybride combine les deux approches. Chaque année possède une partition horizontale, et chaque partition est elle-même divisée verticalement :

### Partition horizontale : Articles 2024



### Partition horizontale : Articles 2023



### Partition horizontale : Articles 2022

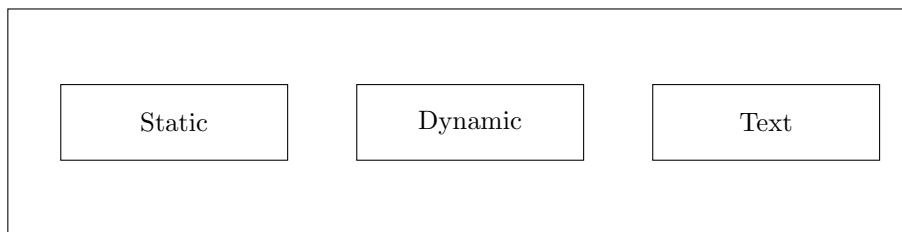


FIGURE 7 – Partitionnement hybride de la dimension Article.

#### Impact sur les lectures de données

Ce schéma réduit le volume lu à deux niveaux :

- seules les partitions temporelles pertinentes sont lues (réduction du nombre de lignes) ;
- seules les colonnes nécessaires sont scannées (réduction du nombre de colonnes).

Les requêtes analytiques qui portent sur des mesures dynamiques (par exemple : évolution du prix ou de la visibilité) n'accèdent plus aux colonnes statiques, ce qui diminue considérablement les accès disque et améliore les temps de réponse.

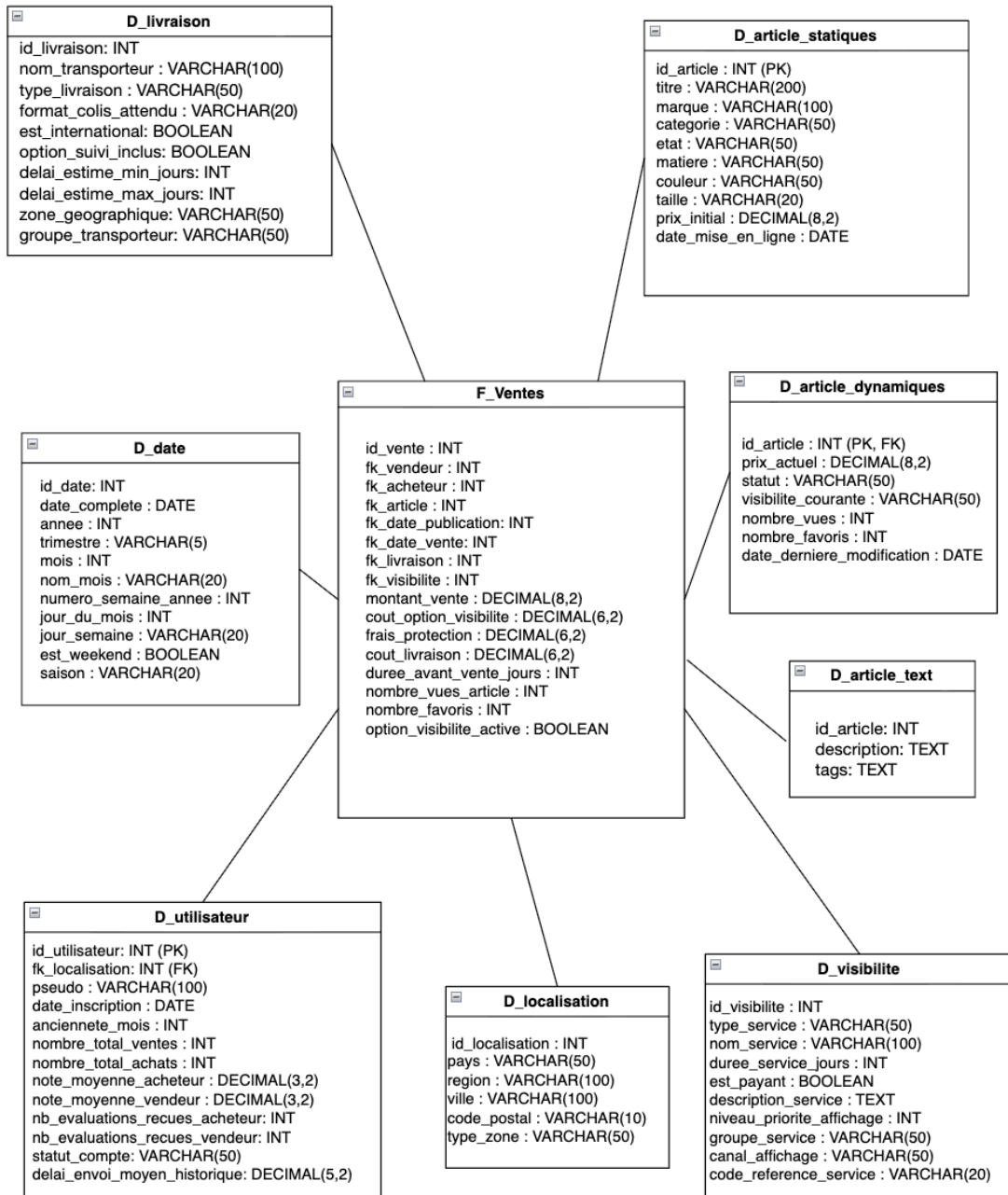


FIGURE 8 – Schéma UML final après application des techniques avancées.



# Annexe

## Références et Sources

Les principales sources utilisées pour les estimations et descriptions fonctionnelles sont indiquées ci-dessous. Elles sont regroupées ici pour ne pas alourdir le corps du rapport.

- **Le Monde du Tabac – Observatoire de la consommation des services colis** (2024). « Vinted compte 27 millions de membres inscrits en France ». <https://www.lemondedutabac.com/observatoire-de-la-consommation-services-colis-vinted-compte-27-millions-dinscrits-en-france>
- **Les Échos Publishing – Étude Joko : consommation mode** (2024). Analyse réalisée à partir d'un panel de 700 000 personnes ; fréquence moyenne d'achat : 11 fois par an. [https://demos.lesechos-publishing.fr/uploads/lep\\_content/lesessentiels\\_com\\_m\\_10.pdf](https://demos.lesechos-publishing.fr/uploads/lep_content/lesessentiels_com_m_10.pdf)
- **Product Mint – Vinted Statistics & Facts** (2025). Statistiques sur les visites mensuelles de la plateforme Vinted en France. <https://productmint.com/vinted-statistics/>
- **LSN Global – “Vinted tops France’s fashion market as second-hand sales surge”** (2025). Analyse du placement de Vinted comme leader du marché de la mode en France. <https://www.lsnglobal.com/news/article/32056/vinted-tops-france-s-fashion-market-as-second-hand-sales-surge>
- **Peaches.fr – Lifestyle** (2025). « En France, environ 400 000 nouveaux articles arrivent chaque jour sur Vinted ». <https://www.peaches.fr/lifestyle/vinted-la-phrase-magique-pour-vendre-plus-vite-et-le-mot-qui-fait-decoller-vos-vetements-ou-objets-en-lots-200429.html>
- **The Ethos – “How Vinted Became Europe’s Most Powerful Resale Platform”** (2025). Rapport sur le développement de Vinted et son volume d'utilisateurs. <https://the-ethos.co/how-vinted-became-europes-most-powerful-resale-platform/>
- **Vinted – Site officiel (Help)**. Informations sur les services de visibilité, les frais, et l'expédition. <https://www.vinted.fr/help>

## Utilisation d'une IA

Dans le cadre de ce projet, une assistance d'intelligence artificielle a été utilisée pour reformuler certaines parties du texte pour améliorer la clarté, la mise en forme LaTeX du rapport, vérifier la cohérence entre les différentes sections.

Les choix de modélisation, les analyses métier, la structuration des schémas, les estimations de volumétrie et l'ensemble du contenu technique ont été réalisés par les membres du groupe.

Nous garantissons que le travail fourni reflète notre compréhension personnelle du cours et des méthodes de modélisation enseignées dans l'UE.