

Analyse des besoins métier

Groupe 14

3 novembre 2025

— ALJIBBAOUI Tala
— AMMARI Reda
— SOUFYANI Anouar

Table des matières

1	Introduction	2
2	Analyse	2
2.1	Définition et enjeux du domaine	2
2.2	Les pratiques classiques de gestion et d'analyse des données	3
2.2.1	Outils de reporting et d'analyse	3
2.2.2	Limites de ces approches	3
2.3	Application au Projet d'Entrepôt de Données "Type Vinted"	4
2.3.1	Objectif du Modèle d'Entrepôt	4
2.3.2	Ordre d'Importance des Actions (Rentabilité Potentielle)	4
2.3.3	Actions Analytiques Prioritaires pour la Conception	4
2.3.4	Modèle d'Entrepôt Envisagé : Schéma en Étoile	4
2.3.5	Indicateurs Clés et Exploitation (Exemples)	6
3	Conception (traitement, schema et estimation)	7
3.1	Conception du Data-Mart (Action la plus importante)	7
3.1.1	Modèle d'Entrepôt Envisagé : Schéma en Étoile	7
3.2	Les 6 Dimensions et leurs Attributs	7
3.3	Réponse au Besoin Métier	9
3.4	Instance de l'Entrepôt et Estimation de Taille	10
3.4.1	Exemple d'Instance (2 lignes de faits)	10
3.4.2	Estimation de Taille sur 12 Mois (Basée sur les chiffres France)	10
3.4.3	Justification Entrepôt de Données vs. Excel	11
3.5	Vérification des Mesures	11
4	Conception des Data-Marts Secondaires (Snapshot et Updated-Records)	11
4.1	Data-mart Snapshot (Inventaire des Annonces)	11
4.1.1	Dimensions du Modèle Snapshot	12
4.1.2	Table des Faits et Typologie des Mesures (Snapshot)	13
4.1.3	Réponse au Besoin Métier (Snapshot)	13
4.1.4	Exemple d'Instance (Snapshot)	13
4.1.5	Estimation de Taille (Snapshot)	14
4.2	Data-mart Updated-Records (Suivi des Livraisons)	15
4.2.1	Dimensions du Modèle Updated-Records	15
4.2.2	Table des Faits et Typologie des Mesures (Updated-Records)	16
4.2.3	Réponse au Besoin Métier (Updated-Records)	17
4.2.4	Exemple d'Instance (Updated-Records)	17
4.2.5	Estimation de Taille (Updated-Records)	18
4.3	Typologie des Mesures par Table de Faits	18
4.4	Réponse aux Besoins Métier	20

1 Introduction

2 Analyse

2.1 Définition et enjeux du domaine

Le marché de la revente en ligne entre particuliers fait partie de ce qu'on appelle l'économie circulaire. Son but est de donner une seconde vie aux produits au lieu de les jeter. Ce secteur connaît une forte croissance ces dernières années, car de plus en plus de consommateurs cherchent à acheter moins cher et à consommer de façon plus responsable.

Dans ce contexte, Vinted est aujourd'hui l'un des acteurs principaux. Créée en 2008 en Lituanie, la plateforme permet à des particuliers de vendre, acheter ou échanger des vêtements et accessoires d'occasion. En quelques années, elle est devenue le leader européen de la mode de seconde main, avec plusieurs dizaines de millions d'utilisateurs. Vinted se distingue par sa simplicité d'utilisation, la mise en ligne gratuite des articles et un système sécurisé pour les paiements et les livraisons.

Présente dans plusieurs pays européens comme la France, l'Allemagne, l'Espagne ou encore les Pays-Bas, Vinted s'est imposée grâce à une interface simple, une communauté très large et une communication centrée sur la durabilité et l'économie. Ses principaux concurrents sont des plateformes comme Depop, Vestaire Collective ou eBay, mais Vinted reste la plus populaire pour les vêtements du quotidien et les articles à petits prix.

Cependant, malgré son succès, Vinted doit relever plusieurs défis :

- **Gérer un grand nombre d'utilisateurs et d'annonces** : chaque jour, des millions d'articles sont publiés et vendus. La plateforme doit pouvoir suivre toutes ces activités sans problème et rester stable.
- **Assurer la sécurité des transactions** : il est important que les paiements se passent correctement et que les échanges entre vendeurs et acheteurs soient fiables.
- **Faciliter la recherche et l'accès aux articles** : les utilisateurs doivent pouvoir trouver facilement ce qu'ils cherchent, que ce soit par le moteur de recherche ou les listes de suggestions.
- **Étudier le comportement des utilisateurs** : savoir quels articles sont consultés ou achetés permet à Vinted de mieux comprendre les attentes des utilisateurs et d'améliorer ses services.

Pour relever ces défis, la donnée joue un rôle essentiel. Les informations collectées sur les ventes, les recherches ou les habitudes des utilisateurs peuvent aider l'entreprise à mieux comprendre son fonctionnement et à prendre de meilleures décisions. C'est pourquoi la création d'un entrepôt de données est importante : il permet de rassembler, organiser et analyser toutes ces informations pour soutenir la stratégie et le développement de Vinted.

Modèle économique et exploitation des données chez Vinted

Vinted repose sur un modèle de vente entre particuliers (Consumer To Consumer - C2C), c'est-à-dire que les utilisateurs eux-mêmes publient, vendent et achètent des articles sans passer par des professionnels. L'entreprise ne possède donc pas de stock : elle met simplement en relation les acheteurs et les vendeurs, en gérant les paiements et la logistique comme intermédiaire sécurisé.

Sa principale source de revenus provient des **frais de protection acheteur**, ajoutés à chaque transaction. Ces frais incluent une commission fixe et un pourcentage du prix de vente, servant à couvrir les coûts liés à la sécurité des paiements, à la gestion des litiges et à la maintenance du service. Cette approche assure à Vinted un flux financier régulier proportionnel au volume d'échanges réalisés sur la plateforme.

En complément, Vinted propose des options payantes de mise en avant pour les vendeurs souhaitant accroître la visibilité de leurs articles. Par exemple, l'option *Boost* permet de mettre un article en haut des résultats pendant 4 heures, et coûte environ 1 à 3 euros selon le prix de l'article. L'option *Wardrobe Spotlight* met en avant plusieurs articles simultanément pendant 7 jours et peut coûter jusqu'à 10 euros. Ces services génèrent des revenus supplémentaires et fournissent également

des informations précieuses sur le comportement des vendeurs et l'efficacité des promotions.

À cela s'ajoutent des partenariats logistiques et financiers avec des transporteurs (par exemple Mondial Relay ou Colissimo) et des prestataires de paiement, qui génèrent des commissions indirectes sur chaque transaction ou étiquette d'envoi produite. Certaines annonces ou marques peuvent également être mises en avant de manière sponsorisée sur la plateforme via des services publicitaires.

Bien que la société ne commercialise pas directement les données personnelles de ses utilisateurs, celles-ci constituent un actif stratégique essentiel. L'exploitation intelligente des données alimente les algorithmes de recommandation personnalisée, permettant à chaque utilisateur de se voir proposer des articles adaptés à ses goûts et à son historique d'achat. Ces analyses contribuent à augmenter le taux de conversion et à renforcer la fidélisation des membres.

Les données transactionnelles et comportementales sont également mobilisées pour analyser la performance des catégories de produits, optimiser les campagnes marketing et identifier les zones géographiques les plus actives. Vinted s'appuie sur ces analyses pour ajuster sa stratégie commerciale, orienter ses partenariats logistiques et planifier ses investissements publicitaires.

Un autre usage majeur des données concerne la détection de fraudes et de comportements suspects. À l'aide de modèles prédictifs, la plateforme repère les transactions anormales, les annonces dupliquées ou les activités massives non conformes aux conditions d'utilisation. Cela contribue à sécuriser l'écosystème et à préserver la confiance entre utilisateurs.

Enfin, la consolidation de l'ensemble de ces informations dans un entrepôt de données permet à Vinted de suivre ses indicateurs de performance (ventes, délais de livraison, taux de litige, satisfaction client) et d'améliorer continuellement son fonctionnement. Cet entrepôt devient ainsi un outil stratégique de pilotage et de décision, soutenant la croissance de l'entreprise tout en garantissant une meilleure expérience utilisateur.

2.2 Les pratiques classiques de gestion et d'analyse des données

Décrire comment les plateformes géraient historiquement leurs données : bases relationnelles pour les transactions et utilisateurs, outils d'analyse limités à des tableaux de bord simples, données dispersées entre plusieurs services.

Montrer les limites : difficile de croiser ventes / utilisateurs / zones géographiques / temps, pas d'historique exploitable pour l'analyse prédictive, ralentissement des bases opérationnelles lors d'analyses lourdes.

Les plateformes comme **Vinted** géraient historiquement leurs données avec des bases de données relationnelles. Ces bases de données étaient utilisées pour stocker les transactions et informations utilisateurs telles que les profils, avis, publications, etc. Souvent, chaque service disposait de sa propre base de données (ventes, utilisateurs, articles, etc.). Ainsi, les données étaient souvent dispersées, ce qui rendait leur exploitation complexe.

2.2.1 Outils de reporting et d'analyse

Vinted utilisait des outils de reporting classiques tels que :

- des tableaux **Excel**,
- des requêtes **SQL**,
- des tableaux de bord basiques.

Ces outils servaient à produire des indicateurs simples comme le nombre de ventes quotidiennes, les articles populaires, etc. Les analyses étaient **descriptives et ponctuelles**, se basant sur des exports manuels de données provenant des bases opérationnelles. Cette approche limitait la compréhension du comportement des utilisateurs ou de l'évolution du marché sur le long terme.

2.2.2 Limites de ces approches

Difficulté à croiser les données : Les informations sur les ventes, les utilisateurs, les périodes de temps et les zones géographiques étaient stockées dans des systèmes différents. Cela rendait difficile la production d'analyses croisées, par exemple :

- identifier les régions les plus actives ;
- déterminer les articles populaires selon la saison.

Absence d'historique exploitable : L'absence d'un historique complet empêchait toute **analyse prédictive** ou étude d'évolution dans le temps.

Ralentissement des systèmes opérationnels : Lors d'analyses lourdes exécutées sur les bases de données, le fonctionnement de la plateforme pouvait ralentir. Certaines opérations subissaient alors des **latences**, notamment en période de forte activité.

Manque de cohérence et de vision unifiée : Comme chaque service disposait de sa propre base, il devenait difficile d'obtenir une **vision globale et fiable** de l'activité de la plateforme. Les chiffres pouvaient différer d'un service à l'autre à cause de décalages de mise à jour, compliquant ainsi la prise de décision stratégique.

2.3 Application au Projet d'Entrepôt de Données "Type Vinted"

Le projet consiste à concevoir un modèle d'**Entrepôt de Données** (Data Warehouse) capable de répondre aux besoins analytiques stratégiques d'une plateforme de revente en ligne de type Vinted. L'objectif principal de cet entrepôt est de transformer le volume massif de données transactionnelles et comportementales en informations cohérentes et exploitables pour le pilotage de l'entreprise.

2.3.1 Objectif du Modèle d'Entrepôt

L'entrepôt de données simulé visera à fournir une **vision historique et agrégée** des performances de la plateforme. Cela permettra aux équipes de direction, marketing et logistique de :

1. **Optimiser la rentabilité :** En analysant l'efficacité des options payantes et les revenus générés par catégorie.
2. **Améliorer l'expérience utilisateur :** En réduisant les délais de livraison et en gérant mieux les litiges.
3. **Soutenir la croissance :** En identifiant les tendances de marché et les zones géographiques à fort potentiel.

2.3.2 Ordre d'Importance des Actions (Rentabilité Potentielle)

Afin de concentrer les efforts de conception sur les axes les plus critiques pour la pérennité et la croissance de la plateforme, les actions / opérations sont ordonnées comme suit :

1. **Vente d'un article :** C'est le cœur du modèle économique (C2C) et la source principale des revenus via les **frais de protection acheteur** et les **options payantes**. Sans vente, l'entreprise ne génère pas de revenu.
2. **Livraison :** Elle affecte directement les **partenariats logistiques** et la **satisfaction client**. Une logistique optimisée réduit les coûts et les litiges.
3. **Utilisateur (Comportement et Fidélité) :** L'analyse de l'activité (panier moyen, taux de fidélité) est essentielle pour l'**acquisition** et la **rétenion** des membres à long terme.
4. **Réclamation / Litige :** Bien que vitale pour la confiance, cette action est principalement un coût opérationnel. Son optimisation vient soutenir la **fidélisation** générée par les actions précédentes.

2.3.3 Actions Analytiques Prioritaires pour la Conception

Conformément à l'analyse de rentabilité, les deux actions les plus importantes à analyser, et qui détermineront la structure de l'entrepôt, sont :

1. **La Vente d'un article** (pour les revenus).
2. **La Livraison** (pour les coûts et la satisfaction client).

Ces deux actions serviront de base à la création des **faits** (tables de faits) dans notre modèle de données.

2.3.4 Modèle d'Entrepôt Envisagé : Schéma en Étoile

Pour répondre efficacement aux requêtes analytiques complexes (section 3.2) tout en assurant une performance optimale, le modèle de l'entrepôt de données sera basé sur un **schéma en étoile** (*star schema*), reconnu pour sa simplicité et son efficacité en matière de requêtes.

Table de faits VENTE La table **VENTE** centralise les informations relatives à chaque transaction finalisée. Elle relie l'ensemble des dimensions contextuelles et contient les mesures quantitatives utiles à l'analyse des performances.

TABLE 1 – Structure de la table de faits **VENTE**

Attribut	Type	Description
IDVente	INT	Identifiant unique de la transaction.
FK_Vendeur	INT	Référence vers la dimension Utilisateur (vendeur).
FK_Acheteur	INT	Référence vers la dimension Utilisateur (acheteur).
FK_Article	INT	Référence vers la dimension Article .
FK_DatePublication	INT	Référence vers la dimension Date (date de mise en ligne).
FK_DateVente	INT	Référence vers la dimension Date (date de vente).
FK_LocalisationVendeur	INT	Référence vers la dimension Localisation (vendeur).
FK_LocalisationAcheteur	INT	Référence vers la dimension Localisation (acheteur).
FK_Livraison	INT	Référence vers la dimension Livraison .
FK_Visibilite	INT	Référence vers la dimension Visibilite .
MontantVente	DECIMAL(8,2)	Montant total payé par l'acheteur (mesure additive).
FraisProtection	DECIMAL(6,2)	Commission perçue par la plateforme (mesure additive).
CoutOptionVisibilite	DECIMAL(6,2)	Coût payé pour une mise en avant (mesure additive).
CoutLivraison	DECIMAL(6,2)	Coût du transport (mesure additive).
DureeAvantVenteJours	INT	Délai entre publication et vente (mesure non-additive par la dimension Date).
NombreVuesArticle	INT	Nombre total de vues avant la vente (mesure additive).
NombreFavoris	INT	Nombre d'ajouts aux favoris avant la vente (mesure additive).

Dimensions du modèle Le modèle comporte six dimensions principales, chacune décrivant un axe d'analyse complémentaire.

- **Utilisateur** : décrit les profils des vendeurs et acheteurs (ancienneté, réputation, nombre de ventes et d'achats, localisation, statut du compte).
- **Article** : caractérise l'article (catégorie, marque, état, matière, couleur, prix initial, taille).
- **Date** : fournit les niveaux d'agrégation temporels (jour, mois, trimestre, année, saison, jour de la semaine).
- **Livraison** : détaille les options de transport (transporteur, délai estimé, coût, zone géographique, format de colis, assurance, option suivi).
- **Localisation** : précise la zone géographique (pays, région, ville, type de zone, code postal) du vendeur et de l'acheteur.
- **Visibilité** : décrit les services de mise en avant (type de service, durée, niveau de priorité, coût, canal d'affichage, description du service).

Colonne	Type de Donnée	Description
ID_Vente_FK	Clé étrangère	Liaison avec la transaction associée.
ID_Transporteur_FK	Clé étrangère	Lien vers la dimension Transporteur .
ID_Date_Envoi_FK	Clé étrangère	Lien vers la dimension Temps (date d'expédition).
ID_Date_Reception_FK	Clé étrangère	Lien vers la dimension Temps (date de réception).
Cout_Livraison	Mesure	Coût total pour Vinted ou le client.
Delai_Effectif_Jours	Mesure	Nombre de jours réels pour la livraison.
Litige_Logistique	Mesure/Flag	Indicateur s'il y a eu un problème de livraison.

TABLE 2 – Table de Faits pour la Livraison

Table de Faits (pour la Logistique) : Fact_Livraison :

2.3.5 Indicateurs Clés et Exploitation (Exemples)

L'entrepôt permettra d'extraire directement les **Indicateurs Clés de Performance (KPI)** nécessaires au pilotage, en s'appuyant sur les mesures des tables de faits :

- **Taux de Conversion des articles Boostés** : $\frac{\sum \text{Ventes_Boostées}}{\sum \text{Articles_Boostés}}$. (Aide à optimiser le prix et l'efficacité de l'option payante).
- **Délai Moyen de Vente par Catégorie** : $\text{AVG}(\text{Duree_Avant_Vente_Jours})$ agrégé par $\text{Dim_Article}(\text{Catégorie})$. (Aide à ajuster les recommandations et à cibler les promotions).
- **Performance Logistique** : $\text{AVG}(\text{Delai_Effectif_Jours})$ agrégé par Dim_Transporteur et $\text{Dim_Utilisateur}(\text{Pays})$. (Aide à renégocier les contrats et à choisir les meilleurs partenaires par zone).

Ces requêtes seront matérialisées par la **génération de rapports analytiques** et des tableaux de bord dynamiques (par exemple dans **Power BI** ou Tableau), transformant l'entrepôt en un outil stratégique d'aide à la prise de décision.

3 Conception (traitement, schema et estimation)

3.1 Conception du Data-Mart (Action la plus importante)

L'analyse métier réalisée a permis d'identifier la **vente d'articles** comme l'action la plus stratégique pour la plateforme. En effet, la majorité des revenus de l'entreprise provient des commissions prélevées sur les ventes finalisées. Il est donc essentiel de comprendre les facteurs qui influencent la transformation d'une annonce en vente effective, ainsi que les leviers permettant d'augmenter la valeur moyenne des transactions.

L'objectif de ce Data-Mart est d'offrir une vision multidimensionnelle des ventes afin de répondre aux questions suivantes :

- Quels types d'articles se vendent le plus rapidement ?
- Quelles catégories ou marques génèrent le plus de chiffre d'affaires ?
- Quels profils de vendeurs réalisent le plus de ventes ?
- Quel est l'impact des options de visibilité et des modes de livraison sur la performance des ventes ?

3.1.1 Modèle d'Entrepôt Envisagé : Schéma en Étoile

Le Data-Mart proposé repose sur un **schéma en étoile** centré sur la table de faits **VENTE**. Ce modèle permet d'analyser chaque transaction selon plusieurs axes : article, vendeur, acheteur, temps, localisation, livraison et visibilité. Les mesures quantitatives contenues dans la table de faits seront agrégées selon ces dimensions afin de répondre aux besoins décisionnels identifiés.

Modélisation détaillée de la table des faits : La table **F_Ventes** contiendra des clés étrangères (FK) pointant vers chaque table de dimension et des mesures numériques décrivant la transaction. Les dimensions fourniront le contexte (**Qui, Quoi, Où, Quand, Comment**) de la vente.

- **FK_Vendeur** et **FK_Acheteur** → **D_Utilisateur**
- **FK_Article** → **D_Article**
- **FK_DatePublication** et **FK_DateVente** → **D_Temps**
- **FK_Livraison** → **D_Livraison**
- **FK_Visibilite** → **D_Visibilite**
- **FK_LocalisationVendeur** et **FK_LocalisationAcheteur** → **D_Localisation**

Les mesures principales associées à chaque ligne de la table **F_Ventes** sont :

- **MontantVente** — montant total payé par l'acheteur (mesure additive).
- **FraisProtection** — commission prélevée par la plateforme (mesure additive).
- **CoutOptionVisibilite** — coût payé pour une option de mise en avant (mesure additive).
- **CoutLivraison** — coût total du transport (mesure additive).
- **DureeAvantVenteJours** — délai entre publication et vente (mesure non additive).
- **NombreVuesArticle** — nombre total de vues (mesure additive).
- **NombreFavoris** — nombre d'ajouts aux favoris (mesure additive).

3.2 Les 6 Dimensions et leurs Attributs

Nous avons mis en place 6 dimensions principales qui décrivent le contexte de chaque vente. Elles apportent la richesse analytique nécessaire pour comprendre les comportements des utilisateurs, la typologie des articles, la temporalité des ventes, les modes de livraison et l'effet de la visibilité sur la performance commerciale.

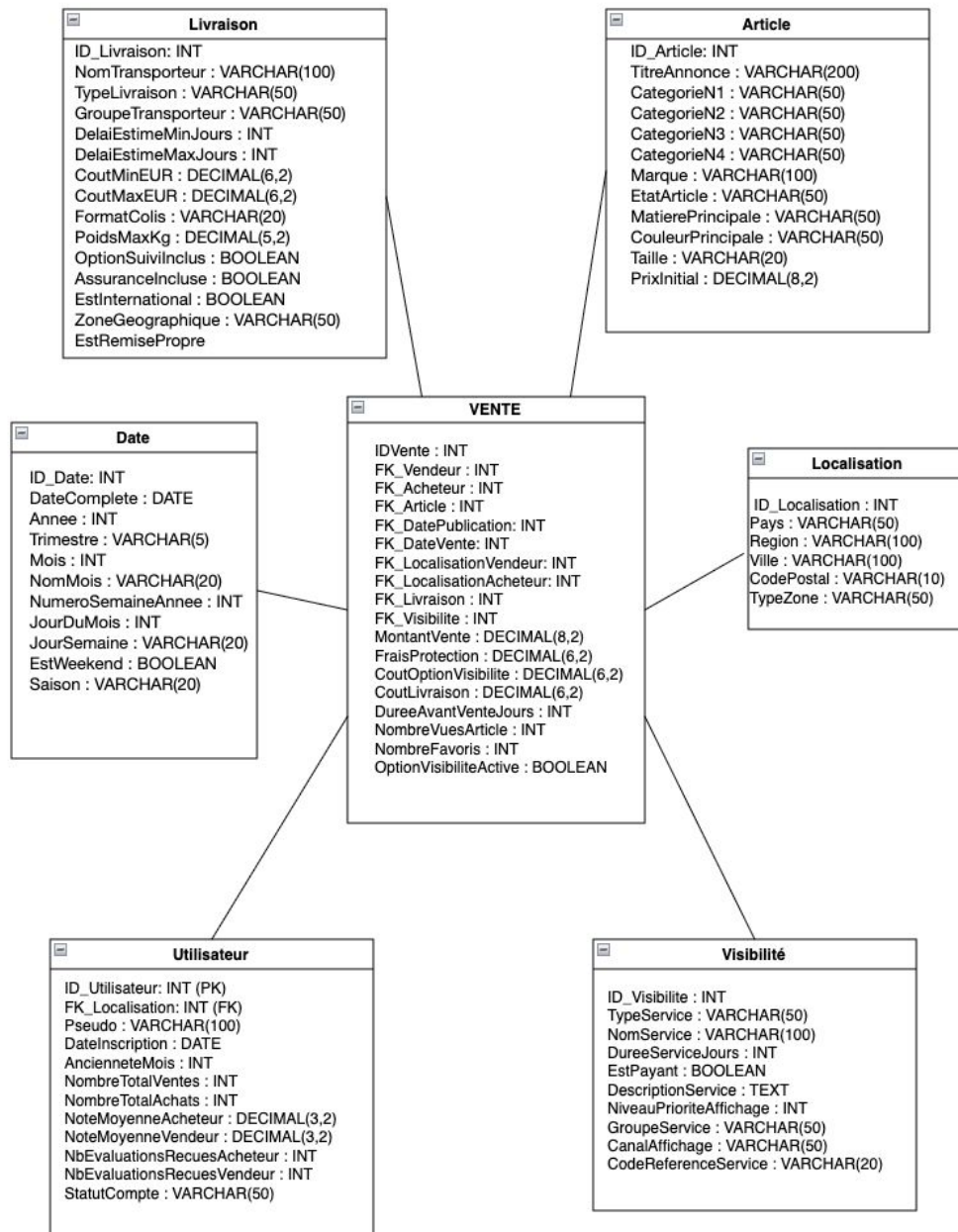


FIGURE 1 – Schéma UML du modèle d'entrepôt de données pour la plateforme Vinted.

Mesures (les indicateurs numériques et booléens) : La table **VENTE** contient également un ensemble de mesures quantitatives et qualitatives permettant d'évaluer la performance commerciale, la rapidité des ventes, et l'efficacité des services payants.

- **MontantVente : DECIMAL(8,2)**
 - **Type :** Additive
 - **Description :** Montant payé par l'acheteur pour l'article.
 - **Justification :** Se somme sur toutes les dimensions (produit, date, localisation, vendeur...).
- **FraisProtection : DECIMAL(6,2)**
 - **Type :** Additive
 - **Description :** Commission prélevée par la plateforme sur la vente (revenu brut).
 - **Justification :** Totalement additive sur toutes les dimensions.
- **CoutOptionVisibilite : DECIMAL(6,2)**
 - **Type :** Additive
 - **Description :** Montant payé par le vendeur pour une option de visibilité (boost, vitrine,

- etc.).
- **Justification** : Additive, agrégée par période, catégorie ou profil vendeur.
- **CoutLivraison** : **DECIMAL(6,2)**
 - **Type** : Additive
 - **Description** : Coût du transport associé à la vente.
 - **Justification** : Se somme logiquement à travers les dimensions (date, région, transporteur...).
- **DureeAvantVenteJours** : **INT**
 - **Type** : Semi-additive
 - **Description** : Nombre de jours entre la mise en ligne et la vente.
 - **Dimension non-additive** : **DATE**.
 - **Justification** : Ne peut être sommée dans le temps, mais se moyenne pour mesurer la rapidité des ventes.
- **NombreVuesArticle** : **INT**
 - **Type** : Additive
 - **Description** : Nombre total de vues de l'article avant sa vente.
 - **Justification** : Somme agrégée utile pour mesurer la visibilité moyenne par produit ou catégorie.
- **NombreFavoris** : **INT**
 - **Type** : Additive
 - **Description** : Nombre d'ajouts aux favoris reçus avant la vente.
 - **Justification** : Additive sur les agrégations, utile pour mesurer l'attractivité des produits.
- **OptionVisibiliteActive** : **BOOLEAN**
 - **Type** : Non-additive
 - **Description** : Indique si une option de visibilité était active au moment de la vente (TRUE / FALSE).
 - **Justification** : Sert à distinguer les ventes avec ou sans mise en avant dans les analyses comparatives.

Synthèse : La table de faits **VENTE** agrège ainsi les indicateurs clés de la plateforme : des mesures financières (**MontantVente**, **FraisProtection**, **CoutOptionVisibilite**, **CoutLivraison**), des mesures d'activité (**DureeAvantVenteJours**, **NombreVuesArticle**, **NombreFavoris**) et une mesure booléenne stratégique (**OptionVisibiliteActive**) permettant de mesurer l'efficacité des options payantes.

Ces données combinées offrent une base solide pour les analyses de performance, de rentabilité et de comportement utilisateur.

3.3 Réponse au Besoin Métier

Principal Besoin Métier : "Augmenter les revenus de la plateforme (commissions et services payants) en optimisant la liquidité du marché (vendre plus, plus vite) et en améliorant l'expérience utilisateur."

Oui, ce modèle permet d'y répondre parfaitement.

Comment ?

1. **Analyse des Revenus (Commissions)** : On peut analyser **SUM(Commission_Plateforme)** en la croisant avec n'importe quelle dimension.
 - *Exemple* : **SUM(Commission_Plateforme) GROUP BY D_Article.Categorie_N1** montre quelles catégories génèrent le plus de commissions.
2. **Mesure de l'Efficacité des Services (Boosts)** : C'est la requête clé permise par la dimension **D_Visibilite**.
 - *Exemple* : **AVG(Jours_Avant_Vente) GROUP BY D_Visibilite.Type_Service**.
 - *Pourquoi ?* Cela compare directement le temps de vente moyen d'un article "Standard" contre un article "Article Boosté". Si le temps est plus bas pour les articles boostés, le service est efficace et justifie son prix (ce qui répond au besoin "d'augmenter les revenus des services").
3. **Optimisation de la Liquidité (Vitesse)** : On peut analyser la vitesse de vente via **AVG(Jours_Avant_Vente)**.

- *Exemple* : `AVG(Jours_Avant_Vente) GROUP BY D_Article.Marque, D_Article.Etat_Article` montre quelles marques en "très bon état" se vendent le plus vite.
4. **Amélioration de l'Expérience (Transport & Notes) :**
- *Exemple* : `COUNT(*) GROUP BY D_Livraison.Nom_Transporteur` montre quels transporteurs sont les plus plébiscités.
 - *Exemple* : `AVG(Note_Transaction_Vendeur) GROUP BY D_Utilisateur.Statut_Compte` (du vendeur) permet de voir si les vendeurs "Pro" offrent une meilleure expérience.

3.4 Instance de l'Entrepôt et Estimation de Taille

3.4.1 Exemple d'Instance (2 lignes de faits)

TABLE 3 – Instance de D_Temps

ID_Date	Date_Complete	Annee	Mois
20251020	20/10/2025	2025	10
20251025	25/10/2025	2025	10
20251026	26/10/2025	2025	10

TABLE 4 – Instance de Utilisateur

ID_Utilisateur	Pseudo	Pays	Statut_Compte
101	VendeurStyle	France	Particulier
102	AcheteurEco	Belgique	Particulier
103	ProVintage	France	Pro

TABLE 5 – Instance de D_Article

ID_Article	Categorie_N3	Marque	Prix_Initial_Demande
5001	Robes	Zara	20.00
5002	Jeans	Levi's	30.00

TABLE 6 – Instance de D_Livraison

ID_Livraison	Nom_Transporteur	Type_Livraison
1	Mondial Relay	Point Relais
2	Colissimo	Domicile

TABLE 7 – Instance de D_Visibilite

ID_Visibilite	Type_Service	Est_Payant
1	Standard	Non
2	Article Boosté	Oui

TABLE 8 – Instance de F_Ventes (Le Cœur du modèle)

FK_Vendeur	FK_Client	FK_Article	FK_Date_Vente	FK_Date_Mise_En_Ligne	FK_Livraison	FK_Visibilite	Montant_Vente	Commission...	Jours_Avant_Vente
101	102	5001	20251025	20251020	1	2	18.00	1.80	5
103	101	5002	20251026	20251010	2	1	30.00	3.00	16

3.4.2 Estimation de Taille sur 12 Mois (Basée sur les chiffres France)

Pour estimer la taille de la table des faits **F_Ventes**, nous devons estimer le nombre total de transactions (ventes) par an, en nous concentrant sur le marché français qui est le plus grand.

- **Contexte (Trafic) :** La plateforme enregistre environ **5,7 millions de visiteurs uniques par jour** en France (Médiamétrie, Nov. 2024, cité par TF1Info). Ce chiffre mesure le trafic, pas les ventes.
- **Base de calcul (Utilisateurs et Fréquence) :**
 - Vinted compte **27 millions de membres** en France (Les Echos Publishing, 2024).

- Une étude Joko (analysant 700 000 transactions bancaires) rapporte que les utilisateurs français de Vinted effectuent en moyenne **11 achats par an** (Les Echos Publishing, 2024).
- **Calcul (Table des Faits F_Ventes) :**
 - Taille par an (France) = 27,000,000 (membres) \times 11 (achats/membre/an)
 - **Taille F_Ventes \approx 297,000,000 lignes par an (pour la France).**
 - (Cela correspond à une moyenne de $\sim 813,700$ ventes par jour en France).
- **Calcul (Dimensions) :**
 - Utilisateur : ~ 27 millions de lignes (pour la France) + les utilisateurs étrangers.
 - D.Article : 400 000 articles ajoutés par jour en 2019 (BFM TV, 2019), soit ~ 146 millions par an. Ce chiffre ayant augmenté, la dimension Article contiendra plusieurs **centaines de millions** de lignes.
 - D.Temps : 365 lignes pour un an (très petit).
 - D.Livraison / D.Visibilite : < 100 lignes chacune (très petit).

3.4.3 Justification Entrepôt de Données vs. Excel

Un fichier Excel est totalement et absolument inadapté.

1. **Limite de Lignes :** Microsoft Excel a une limite maximale de **1 048 576 lignes** par feuille.
2. **Comparaison :**
 - Taille estimée des ventes *par an* (France) : **297 000 000** lignes.
 - Limite Excel : **1 048 576** lignes.

Conclusion : Notre estimation (basée sur des chiffres français sourcés) montre que la table des faits F_Ventes atteindra près de **300 millions de lignes par an**.

Un fichier Excel serait saturé avec les données de ventes de moins de deux jours ($\frac{1048576}{813700} \approx 1.3$ jours).

La volumétrie (des centaines de millions de lignes de faits par an, des centaines de millions de lignes pour les dimensions Utilisateur et Article) et la vitesse ($\sim 800\,000$ nouvelles ventes chaque jour) relèvent purement du **Big Data**. La mise en place d'un entrepôt de données (Data Warehouse) sur une plateforme cloud dédiée (type BigQuery, Snowflake, Redshift) ou un cluster Spark est **fondamentale et indispensable**.

3.5 Vérification des Mesures

Toutes les mesures de la table F_Ventes sont désormais alignées sur le grain (la transaction de vente) et la clarification concernant la dimension D_Visibilite :

- **Mesures Additives** (Montant_Vente, Frais_Port, Commission_Plateforme, Montant_Total_Transaction, Quantite_Vendue, Jours_Avant_Vente) : Elles sont correctes. Elles peuvent être sommées (SUM) sur toutes les dimensions.
- **Mesure Semi-additive** (Nb_Articles_En_Vente_Vendeur) : Elle est correcte. C'est un snapshot du stock au moment de la vente. Son addition n'a pas de sens sur la dimension D_Temps.
- **Mesure Non-additive** (Note_Transaction_Vendeur) : Elle est correcte. C'est un ratio (une note) qui ne peut être que moyenné (AVG).
- **Vérification clé :** La mesure Cout_Visibilite_Paye (le prix payé par le vendeur pour le boost) a été **correctement retirée** de la table F_Ventes. Elle n'appartient pas au processus de "Vente" mais à un processus "Achat de Service". Sa présence aurait pollué le modèle. La dimension D_Visibilite décrit maintenant uniquement l'état de l'article au moment de la vente, ce qui est parfait pour l'analyse d'efficacité.

4 Conception des Data-Marts Secondaires (Snapshot et Updated-Records)

4.1 Data-mart Snapshot (Inventaire des Annonces)

Objectif Métier : L'objectif de ce data-mart est de suivre l'évolution du stock d'annonces disponibles sur la plateforme dans le temps. Cette analyse se fait selon différents axes : la catégorie d'article, le niveau de visibilité, l'ancienneté de l'annonce, le vendeur, et l'état de l'article.

Ce modèle permet de répondre à des questions sur l'inventaire, telles que "Combien d'articles sont actifs chaque jour?" ou "Quelle catégorie d'articles reste la plus longtemps visible sans être vendue?".

Modèle en étoile et Table de Faits : Le datamart est un modèle en étoile centré sur la table de faits `Fait_Annonce_Snapshot`.

Grain (Granularité) : Chaque ligne correspond à l'état d'une annonce à une date donnée (snapshot journalier). Si un article reste en vente 10 jours, il générera 10 lignes dans cette table, une pour chaque jour.

TABLE 9 – Structure de la table des faits `Fait_Annonce_Snapshot`

Attribut	Description	Type	Nature
<code>id_snapshot</code>	Identifiant du snapshot	int	PK
<code>id_date</code>	Date du snapshot	int	FK
<code>id_article</code>	Référence vers l'article	int	FK
<code>id_utilisateur</code>	Référence vers le vendeur	int	FK
<code>id_visibilite</code>	Référence vers la visibilité	int	FK
<code>id_anciennete</code>	Réf. vers la tranche d'ancienneté	int	FK
<code>statut_annonce</code>	Actif / Réservé / Vendu / etc.	string	Attribut descriptif
<code>prix_actuel</code>	Prix affiché à la date du snapshot	decimal	Mesure (Semi-additive)
<code>nombre_vues</code>	Nombre de vues cumulées	int	Mesure (Semi-additive)
<code>jours_depuis_creation</code>	Nb. de jours depuis la mise en ligne	int	Mesure (Non-additive)
<code>favoris_cumulés</code>	Nb. total d'ajouts aux favoris	int	Mesure (Semi-additive)

4.1.1 Dimensions du Modèle Snapshot

Les 5 dimensions associées à cette table de faits sont les suivantes :

Date (Date du snapshot)

- `id_date` (PK)
- `date_complete`
- `jour`
- `mois`
- `nom_mois`
- `trimestre`
- `année`
- `jour_semaine`
- `est_weekend` (Oui/Non)
- `saison`

Article

- `id_Article` (PK)
- `titre`
- `marque`
- `categorie_n1` (ex : "Femme")
- `categorie_n2` (ex : "Vêtements")
- `categorie_n3` (ex : "Robes")
- `état`
- `taille`
- `couleur`
- `prix_initial`

Utilisateur (Vendeur)

- `id_utilisateur` (PK)
- `pseudo`
- `pays`
- `region`
- `date_inscription`

- `ancienneté_compte` (ex : "0-1 an", "1-3 ans", "+3 ans")
- `note_moyenne`
- `nb_evaluations_recues`
- `type_utilisateur` (Occasionnel / Régulier / Pro)

Visibilite (Statut de l'article ce jour-là)

- `id_visibilite` (PK)
- `type_annonce` (Standard / Boostée / Premium)
- `est_payant` (Oui / Non)
- `duree_boost` (ex : 0, 3, 7 jours)
- `cout_promotion` (ex : 1.95, 6.95)
- `groupe_service` (ex : "Boost", "Vitrine", "Standard")
- `niveau_priorite_affichage` (ex : 1, 2, 3)
- `code_reference_service`

Anciennete_Annonce (5ème dimension) Cette dimension permet de "ranger" (bucketize) l'ancienneté d'une annonce pour faciliter l'analyse.

- `id_anciennete` (PK)
- `nb_jours_en_ligne` (ex : 0, 1, ... 365)
- `tranche_jours` (ex : "0-7 jours", "8-30 jours", "31-90 jours", "+ 90 jours")
- `tranche_semaines` (ex : "Semaine 1", "Semaine 2-4", "+ 4 semaines")
- `tranche_mois` (ex : "Mois 1", "Mois 2-3", "+ 3 mois")
- `est_nouveaute` (Oui/Non, si \leq 7 jours)
- `est_vieux_stock` (Oui/Non, si \geq 90 jours)
- `saison_publication` (ex : "Printemps 2024")

4.1.2 Table des Faits et Typologie des Mesures (Snapshot)

La liste des mesures est incluse dans la table `Fait_Annonce_Snapshot`. Leur typologie est la suivante :

- `prix_actuel`
 - **Type : Semi-additive.** On peut le sommer à travers les articles, vendeurs... (pour un jour J), mais son addition n'a pas de sens à travers le temps (`D.Date`).
- `nombre_vues_et_favoris_cumulés`
 - **Type : Semi-additive.** Ce sont des soldes cumulés. On peut les sommer à travers toutes les dimensions (articles, vendeurs...) pour un jour donné, mais leur addition n'a pas de sens à travers la dimension `D.Date`.
- `jours_depuis_creation`
 - **Type : Non-additive.** On ne peut pas sommer cette mesure. L'agrégation pertinente est la moyenne (AVG) pour analyser l'ancienneté du stock.

4.1.3 Réponse au Besoin Métier (Snapshot)

Besoin Métier : Analyser l'évolution temporelle du catalogue et du stock disponible.

Oui, ce modèle répond parfaitement à ce besoin.

Comment ? Ce datamart permet de constituer une vision historique du stock d'articles disponibles, ce qui est impossible avec un modèle transactionnel pur. Il permet de :

- Suivre le nombre d'annonces actives et leur évolution dans le temps.
- Calculer des taux de conversion, par exemple en comparant le statut "Actif" un jour J au statut "Vendu" à J+7.
- Analyser le temps moyen avant vente selon le type d'article ou la visibilité.
- Évaluer l'efficacité des promotions (boosts, annonces premium) en analysant la performance des articles ayant ces statuts de visibilité.

4.1.4 Exemple d'Instance (Snapshot)

Voici un exemple d'instance illustrant le fonctionnement du snapshot journalier.

TABLE 10 – Instance de D_Date

id_date	date_complete	mois	année
1	2025-03-01	Mars	2025
2	2025-03-02	Mars	2025

TABLE 11 – Instance de D_Article

id_Article	titre	categorie_n3	état
101	Jean Levi's	Jeans	Bon
102	Basket Nike	Baskets	Très bon

TABLE 12 – Instance de D_Utilisateur

id_utilisateur	pays	type_utilisateur
1	France	Régulier
2	Espagne	Occasionnel

TABLE 13 – Instance de D_Anciennete_Annonce

id_anciennete	nb_jours_en_ligne	tranche_jours
3	3	"0-7 jours"
4	4	"0-7 jours"
5	5	"0-7 jours"
6	6	"0-7 jours"

TABLE 14 – Instance de D_Visibilite

id_visibilite	type_annonce	est_payant
1	Standard	Non
2	Boostée	Oui

TABLE 15 – Instance de Fait_Annonce_Snapshot

id_snapshot	id_date	id_article	id_utilisateur	id_visibilite	id_anciennete	statut_annonce	prix_actuel	nombre_vues	jours_depuis_creation
1	1	101	1	1	5	Actif	25	120	5
2	2	101	1	1	6	Vendu	23	140	6
3	1	102	2	2	3	Actif	45	90	3
4	2	102	2	2	4	Actif	45	95	4

4.1.5 Estimation de Taille (Snapshot)

Estimation de Taille sur 12 Mois Le grain est "une ligne par article en vente, par jour".

- **Base de calcul (Stock)** : En 2019, 400 000 articles étaient ajoutés *par jour* en France (Source : BFM TV). En supposant une durée de vie moyenne d'un article à 30 jours (avant vente ou suppression), le stock "actif" moyen en France serait de : 400,000 articles/jour \times 30 jours = 12,000,000 articles actifs. (Ce chiffre est conservateur, le stock réel est probablement plus élevé).
- **Calcul (Table des Faits Fait_Annonce_Snapshot)** :
 - Taille par jour (France) = 12,000,000 (articles actifs) \times 1 (snapshot)
 - Taille par an (France) = 12,000,000 (lignes/jour) \times 365 (jours/an)
 - **Taille Fait_Annonce_Snapshot \approx 4,380,000,000 lignes par an.**

Justification Entrepôt de Données vs. Excel Un fichier Excel est radicalement inadapté.

1. **Limite de Lignes** : Microsoft Excel est plafonné à **1 048 576 lignes**.
2. **Comparaison** : Notre estimation du stock pour **un seul jour** est de **12 000 000** de lignes.

Conclusion : Excel ne peut même pas stocker 10% des données d'inventaire d'une seule journée. Le volume (plus de **4 milliards de lignes par an**) et la vélocité (12 millions de nouvelles lignes à insérer chaque jour) exigent impérativement une architecture **Big Data** (Spark, BigQuery, Snowflake, etc.) pour le stockage et l'analyse.

4.2 Data-mart Updated-Records (Suivi des Livraisons)

Pour ce modèle, nous choisissons la deuxième action la plus importante : **le processus de Livraison**.

Un modèle transactionnel (comme **F_Ventes**) enregistre un événement ponctuel. Un modèle snapshot (comme **F_Inventaire**) enregistre un état à un instant T.

Un modèle **Updated-Records** (aussi appelé *Accumulating Snapshot*) est conçu pour suivre un processus métier avec un début, des étapes intermédiaires claires, et une fin. La livraison est le cas d'usage parfait.

Objectif Métier : Suivre le cycle de vie complet d'une livraison, de la vente à la réception, pour mesurer et optimiser les délais à chaque étape (préparation par le vendeur, transit par le transporteur).

Table de Faits : F_Livraisons

Grain (Granularité) : Une ligne par livraison / expédition.

Fonctionnement du Modèle :

1. À la vente (étape 1), une nouvelle ligne est insérée dans **F_Livraisons**. Les clés de dimension et la **FK_Date_Vente** sont renseignées. Toutes les autres dates et mesures de délai (lags) sont NULL.
2. Lorsque le vendeur dépose le colis (étape 2), le processus ETL **met à jour** (*updates*) cette même ligne pour renseigner la **FK_Date_Envoi** et la **FK_Date_MAJ**. Il calcule **Delai_Preparation_Jours**.
3. Lorsque l'acheteur confirme la réception (étape 3), le processus ETL **met à jour** la ligne une dernière fois pour renseigner **FK_Date_Reception** et **FK_Date_MAJ**. Il calcule les mesures finales : **Delai_Transit_Jours** et **Delai_Total_Livraison_Jours**.

Cette table est donc "vivante", les enregistrements sont complétés au fur et à mesure.

4.2.1 Dimensions du Modèle Updated-Records

Ce modèle utilise 5 dimensions, dont deux "à rôle multiple" (Utilisateur et Temps).

D_Utilisateur (Rôle : Vendeur et Acheteur) Contient les informations sur le **vendeur** et l'**acheteur** (y compris la **destination** via l'adresse de l'acheteur).

- id_utilisateur (PK)
- pseudo
- pays (contient l'information de "destination")
- region
- code_postal
- note_moyenne_vendeur
- nb_evaluations_recues
- statut_compte (Particulier, Pro)
- delai_envoi_moyen_historique (SCD)
- date_inscription

D_Temps (Rôle : Jalon de date) La table de faits pointerait 4 fois vers cette dimension : **FK_Date_Vente**, **FK_Date_Envoi**, **FK_Date_Reception** et **FK_Date_MAJ**.

- id_date (PK)
- date_complete
- annee
- mois
- jour_semaine
- est_weekend (Oui/Non)
- est_ferie_france (Oui/Non)
- semaine_annee
- trimestre
- saison

D_Article (Article expédié) Contient les informations sur le **article** (y compris le **poids**).

- **id_article** (PK)
- **categorie_n1**
- **categorie_n2**
- **categorie_n3**
- **marque**
- **etat_article**
- **prix_vente** (Le prix auquel il a été vendu)
- **poids_estime_grammes** (contient l'information de "poids")
- **taille**
- **couleur**

D_Livraison (Transporteur & Méthode) Contient les informations sur le **mode** et le **service** de livraison (et implicitement la **distance** via 'est_international').

- **id_livraison** (PK)
- **nom_transporteur** (contient "service de livraison")
- **type_livraison** (contient "mode de livraison", ex : Point Relais, Domicile)
- **format_colis_attendu** (Petit, Moyen...)
- **est_international** (contient une notion de "distance")
- **option_suivi_inclus** (Oui/Non)
- **delai_estime_min_jours**
- **delai_estime_max_jours**
- **zone_geographique** (National, Europe...)
- **groupe_transporteur** (Réseau Relais, Postal...)

D_Transaction_Vente (Dimension dégénérée) L'identifiant de la transaction de vente (le **order_id**) est placé directement dans la table de faits.

- **id_transaction_vente** (Clé naturelle de la vente)

4.2.2 Table des Faits et Typologie des Mesures (Updated-Records)

La table **F_Livraisons** contient les clés étrangères, les clés de date pour les jalons, et les mesures de délai.

Clés étrangères (les liens) :

- **FK_Vendeur** (pointe vers **D_Utilisateur**)
- **FK_Acheteur** (pointe vers **D_Utilisateur**)
- **FK_Article** (pointe vers **D_Article**)
- **FK_Livraison_Methode** (pointe vers **D_Livraison**)
- **FK_Date_Vente** (Jalon 1, pointe vers **D_Temps**)
- **FK_Date_Envoi** (Jalon 2, pointe vers **D_Temps**, NULL au début)
- **FK_Date_Reception** (Jalon 3, pointe vers **D_Temps**, NULL au début)
- **FK_Date_MAJ** (Votre **date_maj**, pointe vers **D_Temps**)
- **id_transaction_vente** (Dimension dégénérée)

Mesures (les chiffres et délais) :

- **Quantite_Expediee**
 - **Type : Additive.** Vaut 1. Permet de compter les livraisons.
- **Cout_Livraison_Paye** (Votre **cout livraison**)
 - **Type : Additive.** Le montant payé par l'acheteur pour l'envoi.
- **Delai_Preparation_Jours** (Votre **delai**, décomposé)
 - **Type : Additive.** (Calculé : **Date_Envoi** - **Date_Vente**). On peut le moyenner (**AVG**) pour mesurer la réactivité du vendeur.
- **Delai_Transit_Jours** (Votre **delai**, décomposé)
 - **Type : Additive.** (Calculé : **Date_Reception** - **Date_Envoi**). On peut le moyenner (**AVG**) pour mesurer la performance pure du transporteur.
- **Delai_Total_Livraison_Jours** (Votre **delai**, décomposé)
 - **Type : Additive.** (Calculé : **Date_Reception** - **Date_Vente**). Mesure l'expérience globale de l'acheteur.
- **Note_Livraison_Recue** (Votre **note**)

- **Type : Non-additive.** Une note sur 5 donnée par l'acheteur. Ne peut être que moyennée (AVG).

4.2.3 Réponse au Besoin Métier (Updated-Records)

Besoin Métier : Améliorer l'expérience utilisateur et optimiser les partenariats logistiques en réduisant les délais de livraison et en identifiant les points de friction.

Oui, ce modèle répond parfaitement à ce besoin.

Comment ? Ce modèle est spécifiquement conçu pour disséquer le délai de livraison total en sous-parties exploitables :

1. **Analyser la performance des Vendeurs :**
 - *Exemple :* `AVG(Delai_Preparation_Jours) GROUP BY D_Utilisateur.Region_Vendeur.`
 - *Action :* Permet à Vinted d'identifier les vendeurs (ou régions) systématiquement lents à l'envoi et de leur envoyer des rappels.
2. **Analyser la performance des Transporteurs :**
 - *Exemple :* `AVG(Delai_Transit_Jours) GROUP BY D_Livraison.Nom_Transporteur.`
 - *Action :* Permet à Vinted de comparer objectivement la vitesse de "Mondial Relay" contre "Colissimo" et de renégocier les contrats.
3. **Suivre l'état des livraisons en cours :**
 - *Exemple :* `COUNT(*) WHERE FK_Date_Envoi IS NOT NULL AND FK_Date_Reception IS NULL.`
 - *Action :* Donne une vue en temps réel du nombre de colis "en transit".

4.2.4 Exemple d'Instance (Updated-Records)

TABLE 16 – Instance de D_Temps

id_date	date_complete	jour_semaine
20251020	20/10/2025	Lundi
20251022	22/10/2025	Mercredi
20251023	23/10/2025	Jeudi
20251027	27/10/2025	Lundi

TABLE 17 – Instance de D_Utilisateur

id_utilisateur	pseudo	pays
101	VendeurStyle	France
102	AcheteurEco	Belgique

TABLE 18 – Instance de D_Livraison

id_livraison	nom_transporteur	type_livraison
1	Mondial Relay	Point Relais
2	Colissimo	Domicile

Instance de F_Livraisons (Le Cœur du modèle) Cette table montre deux livraisons :

1. **Ligne 1 :** Une livraison **terminée**.
2. **Ligne 2 :** Une livraison **en cours** (envoyée mais non reçue). La FK_Date_MAJ est celle de l'envoi.

TABLE 19 – Instance de F_Livraisons

FK_Vendeur	FK_Acheteur	FK_Livraison...	FK_Date_Vente	FK_Date_Envoi	FK_Date_Reception	FK_Date_MAJ	Cout_Livraison	Delai_Prep	Delai_Transit
101	102	1	20251020	20251022	20251027	20251027	2.99	2	5
102	101	2	20251023	20251027	NULL	20251027	4.95	4	NULL

4.2.5 Estimation de Taille (Updated-Records)

Estimation de Taille sur 12 Mois Le processus de "Livraison" est déclenché par une "Vente". Il y a donc une relation 1 :1 entre ces deux tables de faits.

- **Taille de F.Ventes (estimée dans la section précédente) : $\approx 297,000,000$ lignes par an (France).**
- **Calcul (Table des Faits F_Livraisons) :**
 - Taille F_Livraisons = Taille F_Ventes
 - **Taille F_Livraisons $\approx 297,000,000$ lignes par an.**

Justification Entrepôt de Données vs. Excel La justification est identique à celle du data-mart des ventes : un fichier Excel est inutilisable.

1. **Limite de Lignes :** Microsoft Excel est plafonné à **1 048 576** lignes.
2. **Comparaison :** Notre estimation (basée sur les chiffres français) est de **297 millions** de livraisons par an.

Conclusion : Excel ne pourrait même pas stocker les données de livraison de deux jours. De plus, la nature même de ce data-mart (*updated-records*) nécessite un moteur de base de données capable de gérer des centaines de milliers de mises à jour (UPDATE) quotidiennes sur une table de faits massive, une opération qu'Excel ne peut absolument pas gérer.

Un **entrepôt de données Big Data** est donc indispensable.

4.3 Typologie des Mesures par Table de Faits

Table de Faits F.Ventes et Typologie des Mesures

Clés étrangères (les liens) :

- FK_Vendeur (pointe vers D.Utilisateur.ID.Utilisateur)
- FK_Client (pointe vers D.Utilisateur.ID.Utilisateur)
- FK_Article (pointe vers D.Article.ID.Article)
- FK_Date_Vente (pointe vers D.Temps.ID.Date)
- FK_Date_Mise_En_Ligne (pointe vers D.Temps.ID.Date)
- FK_Livraison (pointe vers D.Livraison.ID.Livraison)
- FK_Visibilite (pointe vers D.Visibilite.ID.Visibilite)

TABLE 20 – Mesures de la table F_Ventes

Mesure	Type	Commentaires
Montant_Vente	Additive	Somme des montants vendus (par jour, pays, catégorie).
Frais_Port	Additive	Somme totale des frais de port payés par les acheteurs.
Commission_Plateforme	Additive	Revenu clé pour la plateforme ("Frais de protection acheteur").
Montant_Total_Transaction	Additive	Montant_Vente + Frais_Port (total payé).
Quantite_Vendue	Additive	Vaut 1. $SUM(Quantite_Vendue)$ = nombre total de ventes.
Jours_Avant_Vente	Non-additive	Calculé : Date_Vente - Date_Mise_En_Ligne . On prend la moyenne (AVG).
Note_Transaction_Vendeur	Non-additive	Moyenne (ex : 4/5), pas de somme possible.
Nb_Articles_En_Vente_Vendeur	Semi-additive	Snapshot du stock au moment de la vente. Dimension non-additive : D.Temps.

Mesures (les chiffres) :

Table de Faits F.AnnonceSnapshot et Typologie des Mesures

Clés étrangères :

- FK_Date (pointe vers D_Date.ID_Date)
- FK_Article (pointe vers D_Article.ID_Article)
- FK_Utilisateur (pointe vers D_Utilisateur.ID_Utilisateur)
- FK_Visibilite (pointe vers D_Visibilite.ID_Visibilite)
- FK_Anciennete (pointe vers D_Anciennete_Annonce.ID_Anciennete)

TABLE 21 – Mesures de la table Fait_Annonce_Snapshot

Mesure	Type	Commentaires
Prix_Actuel	Semi-additive	C'est un solde. On peut le sommer par catégorie/vendeur, mais pas dans le temps. Dimension non-additive : D_Temps.
Nombre_Vues	Semi-additive	C'est un solde cumulé. Dimension non-additive : D_Temps.
Favoris_Cumulés	Semi-additive	C'est un solde cumulé. Dimension non-additive : D_Temps.
Jours_De puis_Creation	Non-additive	Âge de l'annonce. On prend la moyenne (AVG).

Mesures (les chiffres) : *Note : Les mesures Prix_Actuel, Nombre_Vues et Favoris_Cumulés sont des soldes ("balances"), leur état à un instant T. Elles sont donc semi-additives et non additives sur la dimension D_Temps, ce qui corrige l'analyse du document initial.*

Table de Faits F.Livraison Updated et Typologie des Mesures

Clés étrangères :

- FK_Vendeur (pointe vers D_Utilisateur.ID_Utilisateur)
- FK_Acheteur (pointe vers D_Utilisateur.ID_Utilisateur)
- FK_Article (pointe vers D_Article.ID_Article)
- FK_Mode_Livraison (pointe vers D_Livraison.ID_Livraison)
- FK_Date_Vente (jalon 1, pointe vers D_Temps.ID_Date)
- FK_Date_Envoi (jalon 2, pointe vers D_Temps.ID_Date)
- FK_Date_Reception (jalon 3, pointe vers D_Temps.ID_Date)
- FK_Date_MAJ (jalon 4, pointe vers D_Temps.ID_Date)
- ID_Transaction_Vente (Dimension dégénérée)

TABLE 22 – Mesures de la table F_Livraisons

Mesure	Type	Commentaires
Quantite_Expediee	Additive	Compte des livraisons (SUM = COUNT(*)).
Cout_Livraison_Paye	Additive	Montant payé par l'acheteur pour l'envoi.
Delai_Preparation_Jours	Additive	Date_Envoi - Date_Vente. Mesure la réactivité du vendeur (analyse via AVG).
Delai_Transit_Jours	Additive	Date_Reception - Date_Envoi. Mesure la performance du transporteur (AVG).
Delai_Total_Livraison_Jours	Additive	Date_Reception - Date_Vente. Mesure le délai global client (AVG).
Note_Livraison_Recue	Non-additive	Note donnée par l'acheteur (0-5). Agrégation correcte = moyenne (AVG).
Taux_Avancement_Livraison	Semi-additive	Indicateur d'état (0-100%). Seule la dernière version est pertinente. Dimension non-additive : D_Temps.
Montant_Indemnisation	Additive	Total des remboursements suite à un problème de livraison.

Mesures (les chiffres et délais) :

4.4 Réponse aux Besoins Métier

Oui, le modèle d'entrepôt de données que nous avons conçu répond efficacement aux besoins métier identifiés pour une plateforme de type Vinted. Les modèles **F.Vente**, **F.Annonce_Snapshot** et **F.Livraison_Updated** couvrent les trois axes stratégiques principaux de la plateforme : la performance commerciale, la dynamique du catalogue et la qualité de service logistique.

Modèle F.Vente → Analyse des performances commerciales et comportement des utilisateurs Le modèle en étoile **Vente** centralise toutes les informations liées aux transactions : prix de vente, frais de service, date de transaction, satisfaction, article, vendeur, acheteur, etc. Ce modèle permet de :

- Mesurer les ventes globales et détaillées (par catégorie, par pays, par période, par type de vendeur).
- Analyser les revenus de la plateforme grâce à la mesure additive des commissions.
- Étudier la satisfaction client et la rapidité des ventes via les mesures de note et de durée avant vente.
- Identifier les articles ou vendeurs performants et comprendre les facteurs de succès (visibilité, prix, catégorie...).

Ainsi, il répond directement au besoin métier de pilotage des ventes et d'optimisation du chiffre d'affaires.

Modèle F.Annonce_Snapshot → Suivi du catalogue et de la visibilité des articles Le datamart **Snapshot** capture, chaque jour, l'état complet des annonces actives sur la plateforme. Grâce à ce modèle, on peut :

- Suivre l'évolution du stock d'annonces (nombre d'articles actifs, expirés, vendus...).
- Mesurer la durée moyenne avant vente (en croisant `jours_depuis_creation` et `statut_annonce`).
- Analyser la performance des annonces boostées ou sponsorisées (via `id_visibilite`).
- Identifier les périodes de forte activité (via la dimension `Date`) et les catégories à forte rotation.

Ce modèle répond au besoin métier de compréhension du cycle de vie des articles et d'optimisation de la visibilité sur la plateforme.

Modèle F.Livraison_Updated → Suivi du processus logistique et de la satisfaction post-vente Le modèle **Updated-records Livraison** suit une ligne de fait unique par livraison, qui est **mise à jour** à chaque étape clé (envoi, réception) pour refléter l'état d'avancement du processus. Il permet de :

- Mesurer la qualité du service logistique : délais de préparation, de transit et totaux.
- Comparer la performance des transporteurs.
- Suivre la satisfaction client après la réception du colis (`note_livraison_recue`).
- Détecter les anomalies ou retards grâce à l'évolution des statuts et des délais.

Ce modèle répond au besoin métier de suivi de la qualité de livraison, de réduction des délais et d'amélioration de la satisfaction client.

Conclusion Le modèle mis en place permet de :

- Couvrir l'ensemble du cycle de vie de l'activité : mise en ligne → vente → livraison.
- Offrir une vision à 360° sur les ventes, les utilisateurs, les articles et la logistique.
- Faciliter les analyses transversales grâce aux dimensions communes (`Date`, `Utilisateur`, `Article`).
- Soutenir la prise de décision par des indicateurs fiables et historisés (ventes, satisfaction, délais, visibilité).

Ainsi, le modèle répond parfaitement aux besoins métier de pilotage, d'optimisation et d'analyse stratégique d'une plateforme commerciale de type Vinted.