



**Awa NDIAYE**  
**Lorynda LOUFOUA**

# **Rapport de projet**

## Intégration de données

**10 janvier 2024**



**IUT de Paris - Rives de Seine**  
Université de Paris

# TABLE DES MATIÈRES

- 01** Contextualisation
- 02** Modélisation relationnelle
- 03** Table de Faits
- 04** Mesures
- 05** Table de Dimensions
- 06** Schéma Relationnel
- 07** Dashboard

# Contextualisation

Salut Nil.fr,

Alors, on a plongé dans l'analyse du comportement des utilisateurs sur votre plateforme. C'est comme creuser dans une mine d'or pleine de données, mais à quoi bon si on ne sait pas ce qu'il y a dedans, non ? 🤔

Bon, revenons au sérieux. Votre plateforme est cool, mais on veut la rendre encore plus cool en comprenant comment les utilisateurs interagissent avec elle. On sait que chaque livre a sa propre histoire, mais maintenant on veut découvrir celle de vos utilisateurs.

Chaque clic, chaque consultation de page détaillée, chaque ajout au panier et chaque commande sont autant d'indices qui peuvent nous dire beaucoup de choses. On veut dénicher les pépites d'information dans cette jungle de données pour vous aider à prendre des décisions plus intelligentes.

Pour cela, on va plonger dans la modélisation relationnelle. On va cartographier tout ça, des livres aux éditeurs, des maisons d'édition aux lois régissant les prix des livres en France, et même les sociétés de livraison qui font le sale boulot de porter les livres jusque chez les clients. C'est un peu comme faire un puzzle géant, mais avec des données.

On va vous montrer comment tout cela peut être rangé dans une base de données, un peu comme organiser votre collection de livres sur une étagère. On va même ajouter quelques fausses données pour que vous puissiez voir comment ça marche.

Mais ce n'est pas tout. On va identifier deux mesures géniales à surveiller pour chaque niveau d'additivité. Genre, combien de fois un livre particulier a été consulté, ou combien de commandes ont été livrées par La Poste. Vous voyez le truc.

Enfin, on va créer une esquisse d'un outil de reporting ou d'un dashboard, juste pour montrer à quel point ces données peuvent être sexy. Parce que parfois, il vaut mieux voir les données danser plutôt que de les regarder statiquement, non ?

Voilà, c'est le plan. On va rendre votre plateforme encore plus brillante.

Vous êtes prêts ? 🚀

# Modélisation relationnelle

## Tables de faits

Maintenant, parlons du cœur du système : les tables de faits. Ces mauvais garçons enregistrent les événements essentiels qui se produisent sur votre plateforme. Nous en avons identifié deux qui sont cruciaux pour comprendre ce qui se passe : Commande et Détails.

Pour cela nous initialisons :

### Les tables de faits

- Commande
- Détails
- Regarder

Commande :

La table Commande est notre indicateur principal de l'activité transactionnelle. Chaque fois qu'un utilisateur décide de passer à la caisse, c'est là que ça se passe. Donc, pourquoi la granularité "Instantané récapitulatif" ? Pour saisir la quantité totale vendue sur une période indéterminée. Ainsi, la table des commandes reflète une vue récapitulative des activités au fil du temps.

Détails :

Maintenant, plongeons dans les détails. La table Détails conserve, eh bien, les détails de chaque visite. De la date de visite aux page visitée en passant par les petits détails croustillants. Encore une fois, "Instantané récapitulatif" règne en maître ici.

Regarder :

Avec l'Instantané Récapitulatif, nous enregistrons les moments où un produit est regardé sans forcément être commander, offrant une vue récapitulative de l'engagement des utilisateurs au fil du temps.

Dans ce monde complexe de données, la granularité de l'Instantané Récapitulatif s'accorde avec la nécessité de récapitulatifs sur des périodes variables et indéterminées.

Et voilà, nos tables de faits sont prêtes à enregistrer l'action. On se prépare à plonger dans ces données comme des détectives des temps modernes. 🕵️🔍

# Mesures par niveaux d'additivités !

FAITS	TEMPS	LIEU	PRODUITS	CARACTERISTIQUES
TOTAL DES QUANTITÉS VENDUES PAR LIVRE	✓	✓	✓	ADDITIF
CHIFFRE D'AFFAIRES TOTAL PAR MAISON D'ÉDITION	✓	✓	✓	ADDITIF
NOMBRE DE CLIENTS PAR REGION	×	✓	✓	SEMI ADDITIF
NOMBRE DE LIVRES COMMANDÉS	×	✓	✓	SEMI ADDITIF
POURCENTAGE DE PERSONNES ENTRANT SUR LE SITE SANS COMMANDER	×	×	×	NON ADDITIF
DELAI MOYEN DE LIVRAISON	×	×	×	NON ADDITIF

# Justifications !

## Additifs

Pour les mesures additives, on a choisi deux pépites : le chiffre d'affaires et le nombre total de consultations de pages détaillées par livre. Ces mesures peuvent être agrégées de manière significative, que ce soit par livre ou par maison d'édition, sans perte de sens dans l'analyse. Le chiffre d'affaires cumule toutes les transactions, quel que soit le livre ou la maison d'édition, donnant une vue globale de la performance financière. De même, le nombre total de consultations de pages détaillées par livre nous offre une vision agrégée de l'engagement des utilisateurs pour chaque livre, quel que soit l'auteur ou la catégorie.

## Semi additif

Du côté semi-additif, nous avons jeté notre dévolu sur le nombre de clients et le nombre de livres commandés. Le nombre de clients est semi-additif sur la dimension produit, car l'addition du nombre de clients n'a pas de sens lorsqu'il est analysé du point de vue du produit. Agréger le nombre total de clients pour différents produits ne fournit pas une information pertinente sur les habitudes d'achat par produit. Pour le nombre de livres commandés, cette mesure serait semi-additive, car on additionne simplement le nombre de livres commandés. Cependant, lors de l'agrégation par client ou par date, on garde la quantité individuelle pour chaque commande au lieu de l'additionner. Cela nous permet de conserver une vue précise du comportement d'achat au niveau individuel tout en permettant des agrégations pertinentes pour les analyses globales.

## Non-additif

Enfin, pour les mesures non-additives, on a choisi le pourcentage de personnes qui visitent le site mais ne commandent aucun livre, ainsi que le délai moyen de livraison. Le pourcentage de visiteurs non-commandants ne peut pas être agrégé de manière significative sur différentes dimensions, car il représente une proportion fixe du total des visiteurs, indépendamment des autres dimensions. Quant au délai moyen de livraison, additionner ces délais moyens individuels ne donne pas une représentation correcte du délai moyen global de livraison pour l'ensemble des commandes. Ces mesures conservent leur signification uniquement lorsqu'elles sont analysées individuellement, offrant un aperçu précieux du comportement des utilisateurs et de la performance logistique sans perte de contexte.

# Requêtes utilisées

## Additifs

- Total des quantités vendues par livre

```
SELECT id_livre, SUM(quantite) AS total_quantite_vendue
FROM Commande
GROUP BY id_livre;
```

- Chiffre d'affaires total par maison d'édition

```
SELECT maison_edition, SUM(prix * quantite) AS chiffre_affaires_total
FROM Commande c, details
JOIN Livre l ON c.id_livre = l.id_livre
GROUP BY maison_edition;
```

## Semi additif

- Nombre de Transactions (semi-additif sur la dimension produit)

```
SELECT COUNT(id_commande) AS NombreTransactions, id_livre
FROM Commande
GROUP BY id_livre
ORDER BY NombreTransactions DESC;
```

"

- Nombre de clients par région

```
SELECT pays, COUNT(DISTINCT id_client) AS nombre_clients_par_region
FROM Visiteurs
GROUP BY pays;
```

## Non-additif

- Pourcentage de personnes entrant sur le site sans commander

```
SELECT COUNT(DISTINCT id_visiteur) AS total_visiteurs,
COUNT(DISTINCT CASE WHEN id_commande IS NOT NULL THEN id_visiteur END)
visiteur_ayant_commande
FROM Clients;
```

- Delai moyen de livraison

```
SELECT AVG(Societe_livraison.date_livraison - Commande.date_commande) AS
DelaiMoyenLivraison
FROM Commande
JOIN Societe_livraison ON Commande.id_societe = Societe_livraison.id_societe;
```

# Modélisation relationnelle

## Tables de dimensions

Le type de gestion des changements, souvent appelé Slowly Changing Dimension (SCD), définit comment les modifications apportées aux données sont traitées dans les tables de dimensions au fil du temps. Il existe 3 types :

- Type 1 : Ecrasement de l'ancienne valeur par la nouvelle
- Type 2 : Ajout d'une ligne pour intégrer la modification
- Type 3 : Ajout d'une colonne de dimension

### 1. Client

On a tous nos clients préférés, et on veut garder une trace de leurs changements au fil du temps. Pour le type de gestion des changements (SCD), on a opté pour le Type 1. Pourquoi? Parce que si un client change son nom ou son adresse, on écrase simplement l'ancienne valeur par la nouvelle. On ne veut pas surcharger la base de données avec des historiques sans fin.

### 2. Livre

Chaque livre a sa propre histoire, mais il peut aussi connaître des changements. On a choisi le Type 2 pour les livres. Imaginez qu'un livre change d'auteur ou de maison d'édition. On veut garder la trace de ces modifications, alors on ajoute une nouvelle ligne pour intégrer le changement sans toucher à l'ancienne version du livre.

### 3. Visiteur

Les visiteurs peuvent évoluer au fil du temps, mais pour les nôtres, on a opté pour le Type 1. Si un visiteur change son nom ou son pays, on écrase simplement l'ancienne info. Pourquoi ? Parce que parfois, c'est plus simple de ne pas compliquer les choses.

### 4. Société de Livraison

La logistique est cruciale, mais les sociétés de livraison peuvent aussi changer. On a choisi le Type 1 ici aussi. Si une société change son nom, on écrase l'ancien nom par le nouveau. Simplicité avant tout.

### 5. Catégorie

Les catégories peuvent évoluer, mais là, on a opté pour le Type 2. On veut garder une trace de l'évolution des catégories, donc à chaque modification, on ajoute une nouvelle ligne. Simple, non ?

### Les Dimensions Douteuses

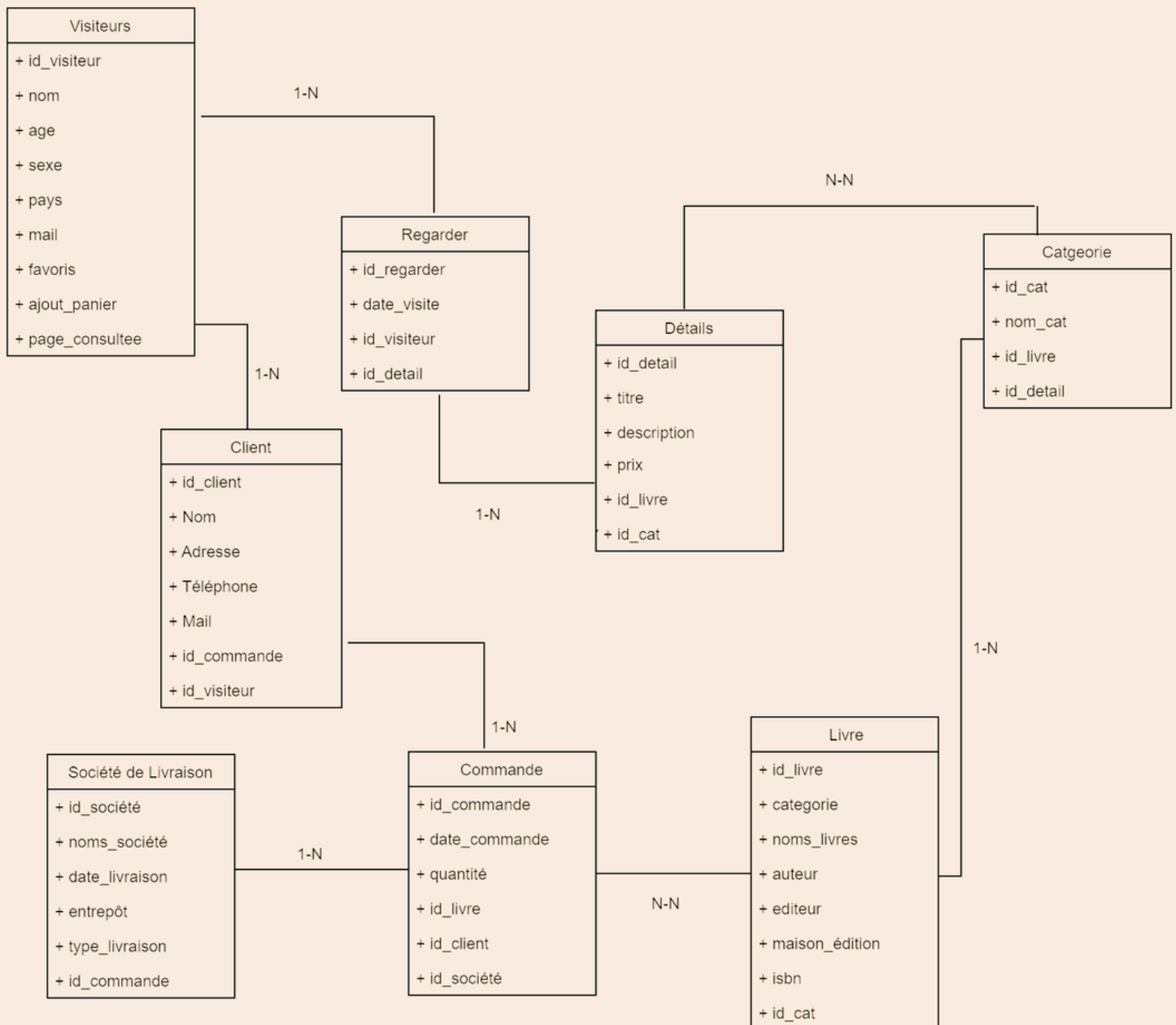
On a quelques tables avec des dimensions un peu floues. Pour les tables Client, Société de Livraison, Livre, et Catégorie, on peut se demander si elles sont vraiment des dimensions ou si elles jouent plusieurs rôles. Elles sont un peu comme des caméléons dans le monde des dimensions, d'où le doute. C'est pour ça qu'on les qualifie de dimensions douteuses.

Voilà comment on a joué avec les dimensions et leurs changements. Vous savez, la vie, c'est compliqué, mais au moins, nos données sont organisées. 😎



# Schéma Relationnel

Pour modéliser cette situation, voici notre Modèle Conceptuel de Données (MCD) final :



# Maquette

**Chiffre d'Affaire**  
**1,6 Billion**



Nombres visiteurs

**126**

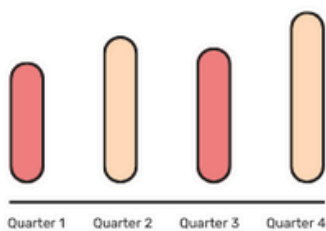
Nombres Clients

**96**

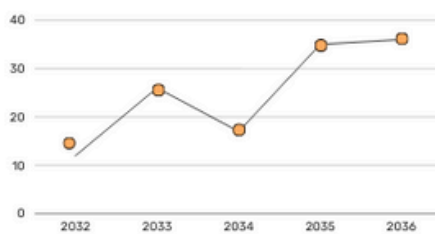
Pourcentage cliques

**92%**

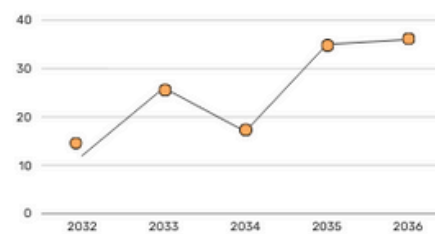
Quarterly Sale Development



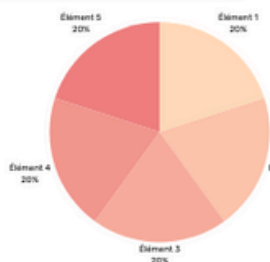
Total des quantités vendues par livre



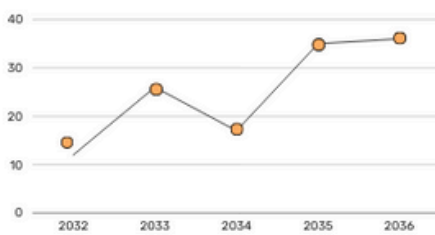
Chiffre d'affaires total par maison d'édition



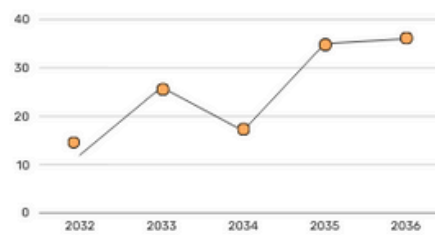
Quarterly Sale Development



Niveau de stock par livre



Nombre de clients par région



**Merci pour votre  
attention**

---

