



UNIVERSITÉ
DE MONTPELLIER

UberRides Data Warehouse

Projet EDBD

2024-2025

Auteurs

GRARI Youssef
BENOMAR Fadel
AL AYOUBI Ibrahim
SAOUMA Christen

Encadrants

FRIGON Anne-Murel
FEDERICO Ulliana

Table des matières

Introduction	2
1 Analyse	3
1.1 Présentation de l'entreprise	3
1.1.1 Position dans le marché	4
1.1.2 Services proposés par Uber Rides	4
1.1.3 Formes de revenu	5
1.2 Actions et opérations à tracer	6
1.3 Traitements	6
1.4 Actions par ordre de rentabilité et d'importance	8
1.5 Actions Stratégiques Classées par Rentabilité et Importance	8
2 Conception de l'entrepot de données	10
2.1 Choix de deux actions les plus importantes	11
2.2 Modélisation des Data Mart	11
2.2.1 Data Mart performances des courses	11
2.2.2 Data Mart Gestion de ressources humaines	14
2.3 Définition des dimensions	16
2.3.1 Dimensions analyse des performances des courses	16
2.3.2 Dimensions gestion ressources humaines	21
2.4 Traitements indiqués avec le modèle choisit	25
2.4.1 Action 1 : Performances des courses 1.3	25
2.5 Instance du modèle	27
2.5.1 Instance du Data Mart Courses	27
2.5.2 Instance du Data Mart Ressources Humaines	29
2.6 Estimation de la taille des tables	31
2.6.1 Fait_course	31
2.6.2 Fait_RH	31
3 Implémentation de l'entrepôt de données	32
3.1 Action 1 : Performance de course	32
3.1.1 Implémentation des dimensions et de la table de faits	32
3.1.2 Requêtes analytiques	32
3.2 Action 2 : Suivi des employés et événements	33
3.2.1 Implémentation des dimensions et de la table de faits	33
3.2.2 Requêtes analytiques	34
3.3 Vues Matérialisées	35
3.3.1 Selection de Vues	35

Introduction

Dans un contexte où la demande pour les services de transport ne cesse de croître, l'optimisation des opérations d'une entreprise de taxis comme Uber s'avère essentielle pour garantir sa compétitivité et sa pérennité. Ce projet, réalisé sous la supervision de Mme Anne-Muriel Frigon et M. Ulliana Federico, a pour objectif de concevoir un entrepôt de données performant permettant d'explorer et d'analyser efficacement les données relatives aux courses effectuées, à la performance des employés, à la satisfaction des clients, et à la gestion des ressources.

Notre démarche vise à identifier des leviers d'amélioration susceptibles d'augmenter les revenus tout en renforçant l'efficacité opérationnelle et en tenant compte des enjeux écologiques croissants. Pour ce faire, nous adopterons une approche analytique rigoureuse, en nous concentrant sur des indicateurs tels que le gain par course, la durée des trajets, et la disponibilité des véhicules. En parallèle, nous analyserons les préférences des clients et leur niveau de satisfaction, des éléments cruciaux pour orienter les stratégies commerciales et améliorer l'expérience utilisateur.

À travers une série de traitements de données, notre objectif est de fournir des recommandations concrètes qui permettront à l'entreprise d'optimiser ses ressources humaines et ses services, tout en adoptant une approche durable de la mobilité urbaine.

Le rapport se structure de la manière suivante : dans le Chapitre 1, nous proposons une analyse approfondie des besoins de l'entreprise ainsi que des actions clés sur lesquelles nous nous concentrerons. Le Chapitre 2 détaillera ensuite la conception de la solution, suivie par son implémentation dans le Chapitre 3.

Chapitre 1

Analyse

Dans ce chapitre, nous réaliserons une analyse approfondie du cas considéré, en nous concentrant sur les objectifs de l'entreprise, sa position sur le marché, ainsi que les services et produits qu'elle propose. Nous examinerons également ses sources de revenus et identifierons les informations clés nécessaires à la prise de décision stratégique. Pour ce faire, nous identifierons les actions et opérations à tracer, telles que les ventes et les livraisons, qui permettront de récupérer ces données. Pour chaque action, nous proposerons au moins trois traitements analytiques visant à éclairer les décisions et à optimiser les performances de l'entreprise. Enfin, nous classerons ces actions par ordre d'importance et de rentabilité potentielle, afin de fournir une vue d'ensemble stratégique sur les leviers d'amélioration à envisager.

1.1 Présentation de l'entreprise

Uber Rides a été fondée en 2009 et a commencé ses opérations en 2010 à San Francisco. Son modèle d'affaires repose sur la mise en relation de chauffeurs et de passagers via une application mobile, rendant les trajets urbains plus pratiques. En 2023, Uber a réalisé un chiffre d'affaires annuel de 31,8 milliards de dollars, avec plus de **7,6 milliards de trajets** effectués dans le monde entier. Uber compte environ 131 millions d'utilisateurs actifs par mois, répartis dans plus de 900 villes à travers le globe. En termes d'infrastructure humaine, Uber emploie environ 32 600 personnes dans le monde, en plus de **plusieurs millions de chauffeurs** indépendants utilisant la plateforme. La société s'appuie fortement sur des technologies avancées comme l'intelligence artificielle et l'analyse de données pour optimiser ses opérations, notamment via des algorithmes de gestion de la demande et de tarification dynamique. L'analyse des données joue un rôle crucial dans les décisions stratégiques d'Uber, telles que l'optimisation des trajets, l'amélioration de l'expérience client et la gestion de la flotte de chauffeurs. Dans un marché de plus en plus compétitif, où des acteurs comme Lyft et Bolt tentent de capter une part de la clientèle, un entrepôt de données permettrait à Uber de centraliser et de structurer ses informations, facilitant les prises de décision basées sur des analyses précises et de se maintenir à l'avant-garde de l'innovation.



FIGURE 1.1 – Logo de l'entreprise

1.1.1 Position dans le marché

Uber s'est rapidement imposé comme un acteur incontournable dans l'industrie des transports urbains, en révolutionnant le marché avec un modèle basé sur l'économie collaborative. Son principal atout réside dans sa capacité à proposer un service de mise en relation directe entre les utilisateurs et les chauffeurs, simplifiant le transport en ville. Toutefois, cette domination ne va pas sans défis, notamment avec l'émergence de concurrents locaux et internationaux tels que Bolt, Didi Chuxing en Chine, ou Grab en Asie du Sud-Est. Ces rivaux mettent en avant des offres adaptées à des marchés spécifiques, s'attaquant à des segments de clientèle que Uber peine parfois à fidéliser dans certaines régions du globe.

Pour rester compétitif, Uber cherche à se diversifier en élargissant ses services au-delà du transport de personnes. Par exemple, Uber Eats est devenu une part essentielle de l'activité de l'entreprise, lui permettant de capter une part significative du marché en pleine expansion de la livraison de repas. Par ailleurs, Uber explore les services de freight et les mobilités électriques, renforçant ainsi sa position sur des secteurs en pleine croissance. Malgré cette diversification, Uber fait face à une pression accrue des régulateurs dans des marchés clés, ce qui complique son expansion internationale et son modèle basé sur les chauffeurs indépendants.

La concurrence croissante sur le marché des plateformes de mobilité pousse Uber à innover constamment, notamment en investissant dans la recherche sur les véhicules autonomes et les solutions de transport durable. Ces initiatives visent à réduire les coûts et améliorer la satisfaction client, mais elles mettent également en lumière les enjeux de long terme de l'entreprise face à un environnement réglementaire et concurrentiel de plus en plus complexe.

1.1.2 Services proposés par Uber Rides

UberX

Le service standard d'Uber, parfait pour vos trajets quotidiens. Avec UberX, commandez facilement une voiture privée avec chauffeur via l'application mobile. C'est une option pratique et abordable pour vous déplacer en ville.

Uber Comfort

Si vous recherchez un peu plus de confort, Uber Comfort est fait pour vous. Ce service propose des véhicules modernes et spacieux, ainsi que des chauffeurs bien notés, vous permettant de profiter d'un trajet agréable, avec plus d'espace pour vous détendre.

Uber Black

Pour les occasions spéciales ou simplement pour vous faire plaisir, Uber Black offre une expérience haut de gamme. Avec des voitures de luxe et des chauffeurs professionnels, chaque trajet devient un moment d'exception.

Uber Pool

Ce service de covoiturage est idéal pour économiser de l'argent tout en faisant un geste pour la planète. En partageant votre trajet avec d'autres passagers se rendant dans la même direction, vous réduisez les coûts et l'empreinte carbone.

Uber Pet

Voyagez avec votre compagnon à quatre pattes ? Pas de souci ! Avec Uber Pet, vous pouvez

emmener votre animal de compagnie avec vous, rendant les déplacements plus simples.

Uber Reserve

Pour ceux qui aiment planifier à l'avance, Uber Reserve permet de réserver un trajet à l'avance. C'est parfait pour les rendez-vous importants ou les déplacements vers des événements.

Uber Hourly

Si vous avez besoin d'un chauffeur pour plusieurs heures, Uber Hourly est la solution. Ce service vous permet de louer un chauffeur pour la durée de votre choix, ce qui est très pratique lors d'événements ou de journées bien remplies.

Chaque service est conçu pour s'adapter à des besoins spécifiques, et Uber ne cesse d'ajuster son offre pour mieux correspondre aux attentes de ses utilisateurs. Que vous ayez besoin d'un simple trajet ou d'une expérience plus raffinée, Uber est là pour rendre vos déplacements aussi agréables que possible.

1.1.3 Formes de revenu

Uber Rides génère des revenus à travers divers modèles économiques qui lui permettent de capitaliser sur sa vaste plateforme de mise en relation entre chauffeurs et passagers. Les principales sources de revenus d'Uber Rides sont les suivantes :

Frais de service

Uber prélève des frais de service sur chaque course effectuée via sa plateforme. Ces frais, généralement calculés comme un pourcentage du tarif total de la course, varient selon la ville et le type de service choisi. Par exemple, pour des services premium comme Uber Black, les frais peuvent être plus élevés que pour des options standard telles qu'UberX.

Tarification dynamique

La tarification dynamique, ou "tarification à la demande", est un mécanisme qui ajuste les tarifs en temps réel en fonction de la demande des passagers et de l'offre de chauffeurs. Pendant les périodes de forte demande, comme lors d'événements spéciaux ou pendant les heures de pointe, les tarifs peuvent augmenter, ce qui permet à Uber de maximiser ses revenus.

Partenariats commerciaux

Uber a établi des partenariats avec divers établissements, tels que des restaurants, des hôtels et des entreprises, pour offrir des services de transport à leurs clients. Ces collaborations peuvent inclure des réductions ou des commissions sur les courses générées à partir de ces partenariats, contribuant ainsi aux revenus d'Uber.

Programmes de fidélité et abonnements

Uber propose également des programmes de fidélité, comme Uber Rewards, qui incitent les utilisateurs à choisir leur plateforme plutôt que celle de la concurrence. Les abonnements, tels qu'Uber Pass, offrent aux clients des réductions sur les courses et d'autres avantages, tout en assurant un flux de revenus régulier pour l'entreprise.

Conclusion

En résumé, Uber s'est solidement établi comme un leader sur le marché des transports urbains grâce à son modèle économique innovant et ses divers services. Dans ce qui suit, nous nous focaliserons sur l'option standard, **UberX**, qui incarne l'essence même de l'expérience Uber. Ce service, alliant praticité et coût abordable, répond parfaitement aux besoins quotidiens des usagers.

1.2 Actions et opérations à tracer

Performances des courses : Les données de chaque course, incluant le gain généré, la durée de la course, et le trajet parcouru (kilométrage), sont enregistrées. La rentabilité des courses est analysée en fonction de la distance et la période pour optimiser les investissements.

Gestion des ressources humaines : Une entreprise prospère est avant tout une entreprise qui sait gérer efficacement ses équipes. Il est donc essentiel de disposer d'un modèle capable de soutenir la prise de décision en matière de gestion des employés, en suivant et en analysant les événements qui leur sont associés.

Gestion de la flotte et des véhicules : La disponibilité des véhicules est analysée par heure de la journée pour déterminer combien de véhicules sont disponibles (vides) ou occupés (pleins). Les périodes de sous-utilisation ou de forte utilisation des véhicules sont identifiées pour améliorer la gestion des ressources humaines et planifier les shifts des conducteurs.

Satisfaction des clients et typologie des courses : La satisfaction des clients est analysée en fonction des types de courses choisis (grandes distances vs petites distances). Les types de clients qui préfèrent certains types de courses sont identifiés afin de mieux cibler les offres et promotions.

Optimisation écologique des trajets : Les trajets des taxis avant l'arrivée chez les clients sont étudiés pour optimiser les parcours et réduire l'impact écologique. Des moyens d'améliorer l'efficacité énergétique des trajets à vide sont identifiés. Les économies réalisées grâce à l'optimisation écologique des trajets, notamment la réduction des coûts de carburant et de maintenance, sont calculées.

Appels et gestion des clients : Le nombre d'appels entrants par heure est suivi et le taux de satisfaction après l'appel est analysé. Des efforts sont faits pour réduire le temps d'attente des clients en améliorant le traitement des appels et la rapidité des réponses.

1.3 Traitements

Nous présentons ci-dessous, pour chaque action, plusieurs traitements possibles qui contribueront à la prise de décision visant à améliorer et optimiser les services proposés, ainsi qu'à gérer efficacement les ressources humaines.

Action 1 : Performances des courses

- **Traitement 1 :** Analyser le nombre de courses effectuées par heure, jour et mois pour identifier les périodes de forte demande et en mesurer l'impact sur le chiffre d'affaires.

- **Traitement 2 :** Identifier les facteurs externes (événements, météo) influençant les pics de demande.
- **Traitement 3 :** Évaluer la satisfaction des clients par région, en distinguant les courses longues et courtes, pour mieux orienter les offres et optimiser les investissements.
- **Traitement 4 :** Identifier les clients qui utilisent le moins fréquemment les services UberX afin d'améliorer leur fidélité.
- **Traitement 5 :** Analyser les performances des chauffeurs afin d'attribuer des primes et des incitations aux chauffeurs les plus performants, en se basant sur des critères tels que le nombre de courses effectuées, le chiffre d'affaires généré, la distance parcourue et la satisfaction des clients afin d'attribuer des primes aux plus engagés.

Action 2 : Gestion des ressources humaines

- **Traitement 1 :** Analyser les scores de satisfaction des employés par équipe et par région pour identifier les zones de satisfaction basse. Mettre en évidence les équipes avec les employés les moins satisfaits pour mieux cibler les actions de remédiation.
- **Traitement 2 :** Identifier les événements les plus coûteux et évaluer leur impact (positif ou négatif) sur les équipes et la société. Cela permet de réévaluer les investissements futurs en fonction de leur efficacité.
- **Traitement 3 :** Suivre la durée moyenne des affectations des employés par type d'événement et identifier les éventuels déséquilibres. Optimiser les processus pour réduire les interruptions ou rotations excessives qui pourraient nuire à la stabilité des équipes.
- **Traitement 4 :** Analyser la relation entre la durée des affectations et les niveaux de satisfaction des employés, afin d'identifier les affectations excessivement longues ou fréquentes qui peuvent nuire à la satisfaction. Proposer des ajustements pour améliorer l'expérience des employés.

Action 3 : Gestion de la flotte et des véhicules

- **Traitement 1 :** Identifier le type de véhicules pour lequel les courses sont les moins notées afin de connaître le type de véhicules que les clients apprécient le moins pour les remplacer et améliorer la satisfaction des clients.
- **Traitement 2 :** Identifier les véhicules dont le nombre de réparations est au-dessus de la moyenne afin d'éviter d'en acquérir davantage.
- **Traitement 3 :** Analyser la consommation des véhicules pour identifier ceux ayant les coûts d'exploitation les plus élevés.
- **Traitement 4 :** Suivre l'âge et le kilométrage des véhicules pour prévoir leur remplacement en fonction de leur ancienneté et leur usure.

Action 4 : Satisfaction des clients et typologie des courses

- **Traitement 1 :** Analyser les scores de satisfaction des clients pour les courses longues et courtes afin d'adapter le service aux attentes spécifiques de chaque type de trajet.
- **Traitement 2 :** Analyser la satisfaction des clients selon les quartiers de départ des courses pour proposer des services mieux adaptés aux besoins locaux.
- **Traitement 3 :** Identifier les clients ayant vécu les pires expériences pour leur proposer un dédommagement et améliorer la fidélisation.
- **Traitement 4 :** Étudier la satisfaction en fonction des horaires (heures creuses ou de pointe) pour ajuster la qualité de service et mieux gérer les périodes de forte demande.

Action 5 : Optimisation écologique des trajets

- **Traitement 1** : Optimiser les parcours en étudiant les itinéraires et en identifiant les trajets les plus écologiques.
- **Traitement 2** : Analyser les économies de carburant réalisées grâce à l'optimisation des trajets.
- **Traitement 3** : Suivre l'impact des pratiques écologiques sur les coûts de maintenance des véhicules.
- **Traitement 4** : Analyser les bénéfices financiers générés par les initiatives de réduction d'impact écologique.

Action 6 : Appels et gestion des clients

- **Traitement 1** : Suivre le nombre d'appels entrants par heure pour identifier les pics.
- **Traitement 2** : Évaluer la satisfaction client post-appel pour mesurer l'efficacité du service.
- **Traitement 3** : Analyser les raisons d'insatisfaction des clients lors des appels pour optimiser le service.
- **Traitement 4** : Mesurer les temps d'attente pour repérer les goulots d'étranglement et en évaluer l'impact sur la satisfaction client.

1.4 Actions par ordre de rentabilité et d'importance

Dans le cadre de la gestion de notre entrepôt de données, il est essentiel de définir et prioriser les actions permettant d'optimiser les revenus et les performances de l'activité de l'entreprise. Voici les principaux axes classés par leur impact sur la rentabilité et l'importance stratégique.

1.5 Actions Stratégiques Classées par Rentabilité et Importance

Notre mission est d'identifier et de prioriser les actions permettant à uber d'optimiser ses revenus pour son option UberX, améliorer ses performances, et favoriser sa croissance. Voici les principaux axes de développement proposés, classés selon leur impact potentiel sur la rentabilité et leur importance stratégique.

1. Analyse des Performances des Courses

L'analyse approfondie des données de courses, incluant les revenus, la durée, et les trajets, est essentielle pour permettre à uberX de maximiser ses profits. En identifiant les périodes de forte demande et les trajets rentables, l'entrepôt de données facilitera des décisions éclairées sur l'allocation des ressources. De plus, cette analyse contribuera à fidéliser les clients en leur proposant des offres adaptées (par exemple, des codes promotionnels pour les utilisateurs les moins réguliers).

2. Gestion des Ressources Humaines

L'impact d'une gestion efficace des ressources humaines sur la satisfaction des clients et les revenus de l'entreprise est certes indirect, mais essentiel. Une telle gestion renforce les piliers de l'entreprise, en optimisant la performance et l'engagement des équipes. Grâce à ce modèle, l'entreprise pourra prendre des décisions ciblées concernant ses ressources humaines, ce qui améliorera directement leur rendement et, à terme, la performance globale des équipes.

3. Fidélisation Client et Typologie des Courses

L'entrepôt de données permettra à uber d'identifier les préférences de ses clients et d'adapter ses offres aux comportements observés. Cette approche, orientée vers l'expérience utilisateur, favorisera la fidélisation à long terme en optimisant l'offre de services et en répondant aux attentes des clients.

4. Optimisation Écologique des Trajets

En facilitant une analyse précise des trajets à vide et des émissions de CO₂, l'entrepôt de données contribuera à réduire l'impact environnemental des activités d'uber. Cette approche permet de limiter les coûts opérationnels tout en répondant aux exigences d'une clientèle de plus en plus soucieuse de durabilité.

5. Gestion de la Flotte et Disponibilité des Véhicules

L'entrepôt de données permettra une gestion efficace de la flotte destinée au service uberX, réduisant les coûts d'exploitation et augmentant la réactivité de l'entreprise face aux demandes de course. La mise à disposition de données précises sur l'entretien et la disponibilité des véhicules contribuera directement à la satisfaction client et à la rentabilité.

6. Optimisation de la Gestion des Appels Clients

Même si la gestion des appels n'impacte pas directement les revenus, elle reste cruciale pour la qualité du service uberX. En intégrant des données sur les temps d'attente et le traitement des appels, notre solution renforcera la relation de confiance avec les clients, ce qui soutient l'image de marque d'uber.

Cette approche stratégique vise à structurer efficacement l'entrepôt de données pour soutenir uber dans sa quête d'optimisation des revenus, d'efficacité opérationnelle, et de satisfaction client, contribuant ainsi à sa croissance et à sa compétitivité.

Chapitre 2

Conception de l'entrepot de données

Dans le cadre de notre mission d'optimisation pour Uber, l'analyse approfondie des besoins réalisée précédemment nous a permis d'identifier les axes stratégiques prioritaires pour l'entreprise. Fort de cette compréhension des enjeux, il est maintenant essentiel d'aborder la phase de conception de notre solution de datawarehouse, une étape cruciale qui transformera ces besoins en une architecture data robuste et performante. Cette phase de conception représente le pont entre la théorie et la pratique, où chaque décision technique doit être méticuleusement pensée pour garantir une solution qui répond précisément aux objectifs d'amélioration des gains, d'optimisation des ressources humaines et de réduction de l'impact environnemental.

Face à la complexité et à la diversité des données à traiter, notre approche de conception repose sur une méthodologie éprouvée de modélisation en étoile (*Data Mart*). Cette phase de conception, loin d'être un simple exercice technique, constitue la colonne vertébrale de notre solution, où chaque table de faits, chaque dimension et chaque métrique est pensée pour transformer les données brutes en informations stratégiques directement exploitables par les décideurs. la Figure 2.1 ci-dessous représente une modélisation simplifiée de la conception de notre entrepot de données.

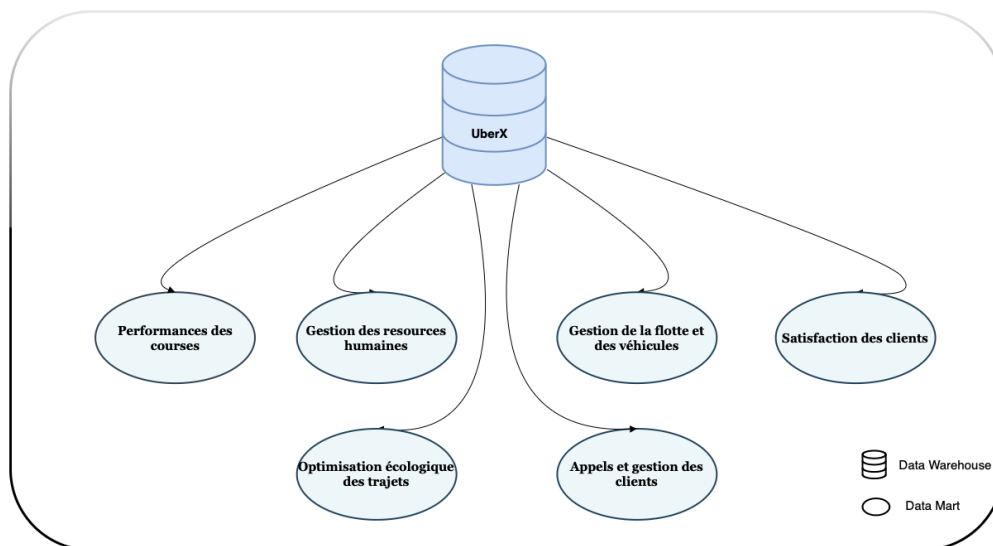


FIGURE 2.1 – UberX Data-warehouse

2.1 Choix de deux actions les plus importantes

Suite à une analyse approfondie des besoins d'Uber, nous avons identifié deux axes stratégiques d'analyse, essentiels pour maximiser la performance de l'entreprise :

A. Analyse des performances des courses

Ce domaine central de l'activité d'Uber regroupe les indicateurs clés tels que les gains, les durées et les distances de trajet. Une compréhension fine de ces données permettra de maximiser la rentabilité en ciblant les types de courses les plus profitables et en affinant les stratégies tarifaires.

B. Gestion des ressources humaines

En lien direct avec la qualité de service et l'efficacité opérationnelle, cet axe se focalise sur l'optimisation des affectations, le suivi des sanctions et des formations des employés. L'objectif est de répondre de manière proactive aux besoins de l'entreprise en contribuant à la prise de décisions stratégiques pour une gestion plus efficace de ses ressources humaines.

Ces deux axes ont été choisis pour leur contribution directe aux objectifs stratégiques d'Uber : rentabilité accrue, optimisation des ressources et satisfaction client.

2.2 Modélisation des Data Mart

2.2.1 Data Mart performances des courses

La table de faits **Fait_Course**, présentée dans la Figure 2.2, enregistre les mesures clés de chaque course, permettant de relier les informations provenant de plusieurs dimensions : **Dimension_Dates**, **Dimension_Heure_Arrivee**, **Dimension_Heure_Depart**, **Dimension_Client**, **Dimension_Véhicule**, **Dimension_Meteo**, **Dimension_Depart**, **Dimension_Arrivé**, **Dimension_Chauffeur**. Elle constitue le cœur de l'entrepôt de données, facilitant l'analyse des coûts, des profits et de la productivité.

La Table 2.1 présentée ci-après détaille la table fait :

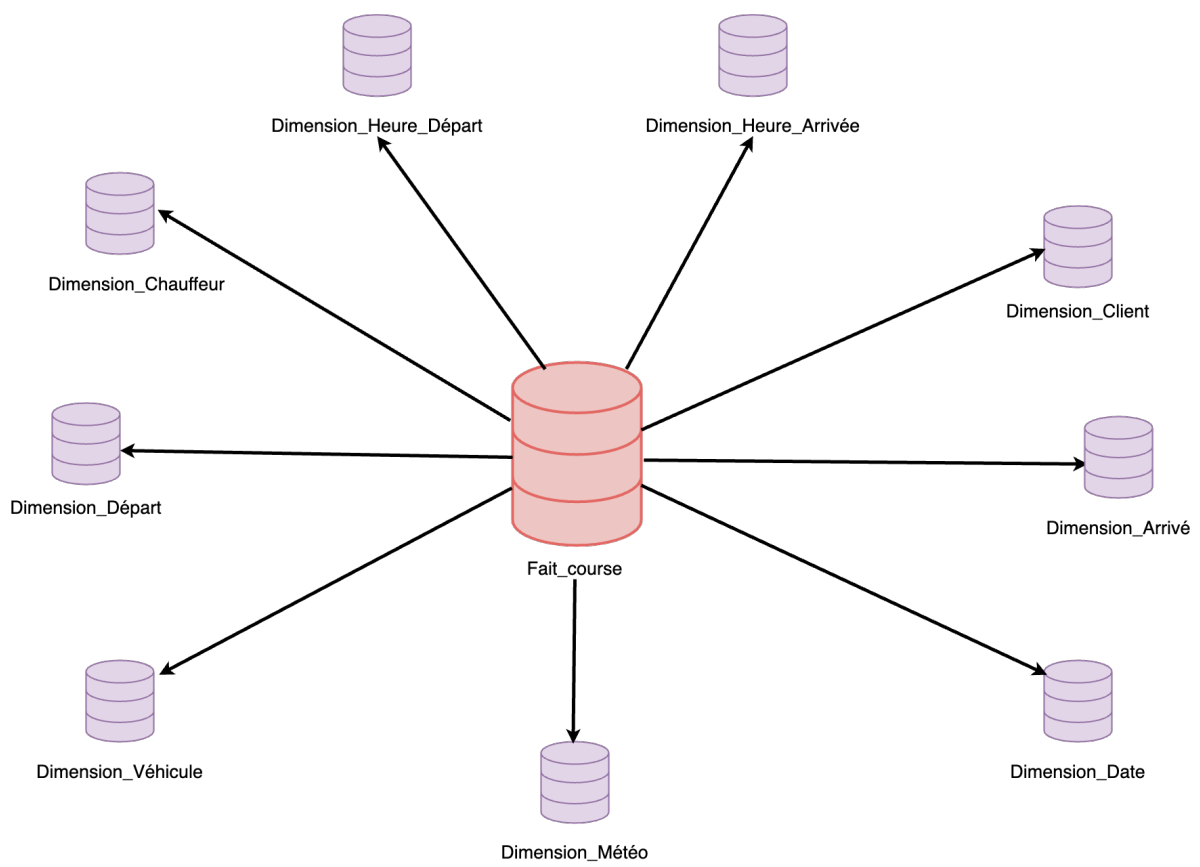


FIGURE 2.2 – Data Mart Course

Attribut	Description
id_Dates (Clé étrangère)	Référence vers la dimension Date
id_Arrivée (Clé étrangère)	Référence vers la dimension Arrivée
id_Départ (Clé étrangère)	Référence vers la dimension Départ
id_Client (Clé étrangère)	Référence vers la dimension Client
id_Heure_Départ (Clé étrangère)	Référence vers la dimension Heure Départ
id_Heure_Arrivée (Clé étrangère)	Référence vers la dimension Heure Arrivé
id_Chauffeur (Clé étrangère)	Référence vers la dimension Chauffeur
id_Véhicule (Clé étrangère)	Référence vers la dimension Véhicule
id_Météo (Clé étrangère)	Référence vers la dimension Météo
id_Course (Dimension dégénérée)	Identifiant unique de la course
Prix	Revenus générés par la course
Distance parcourue	Distance totale effectuée pendant la course
Durée du trajet	Temps total entre le début et la fin de la course
Nombre_de_passagers	Nombre de personnes transportées durant la course.
Temps d'attente du client (en minutes)	Délai entre la réservation et l'arrivée du véhicule
Revenu (net)	le gain net après la déduction du pourcentage du chauffeur
Indice de satisfaction client	Note moyenne des courses selon les avis clients (échelle de 1 à 5, par exemple)

TABLE 2.1 – Attributs de la table de faits **fait_COURSE**

Analyse des Mesures

Les mesures ci-dessous sont classées selon leur type d'additivité. Une mesure est dite **additive** si elle peut être sommée sur plusieurs dimensions (comme le temps ou le chauffeur) sans perdre de sens. Les **semi-additives** peuvent être additionnées dans des cas spécifiques mais nécessitent souvent d'être moyennées sur de longues périodes. Enfin, les **non-additives** ne peuvent pas être agrégées par addition.

Attribut	Type	Type d'additivité
Prix	Float	Additif
Distance parcourue	Float	Additif
Durée du trajet	Float	Additif
Nombre de passagers	Int	Additif
Temps d'attente du client (en minutes)	Float	Semi-additif
Revenu (net)	Décimal	Semi-additif
Indice de satisfaction client	Décimal	Semi-additif

TABLE 2.2 – Tableau des mesures et de leur type d'additivité

Justification des Types d'Additivité

- **Mesures additives** : sont celles qui peuvent être additionnées sur plusieurs trajets ou périodes sans perdre leur pertinence. Elles offrent une vision claire des revenus et de la productivité. Parmi elles, on retrouve **Prix**, la **distance parcourue**, la **durée du trajet**, le **nombre de passagers** et le **nombre total de courses**. Ces mesures permettent d'évaluer efficacement les performances des véhicules, des chauffeurs et la fréquentation générale.
- **Mesures semi-additives** : en revanche, nécessitent une interprétation plus nuancée, car elles ne sont pertinentes que dans certains contextes. Elles incluent le **temps d'attente du client**, le **revenu net** et l'**indice de satisfaction client**.

2.2.2 Data Mart Gestion de ressources humaines

La table de faits **Fait_suivi_employé** présentée dans la Figure 2.3 est conçue pour collecter les informations clés permettant une analyse approfondie des événements affectant les ressources humaines. Cette table a pour objectif principal de centraliser les données nécessaires pour les traitements visant le suivi des ressources humaines.

Le schéma en étoile illustré dans la Figure 2.3 relie cette table de faits à six tables de dimensions : **Dimension_Région**, **Dimension_Date_Début_Evénement**, **Dimension_Date_Fin_Evénement**, **Dimension_Employé**, **Dimension_Evénement** et **Dimension_Equipe**.

Nous choisissons ici une approche **transactionnelle**. Une telle approche est plus adaptée pour ce datamart car il suit des événements discrets (transferts, sanctions, formations) avec des dates précises de début et fin. Cette nature événementielle, combinée

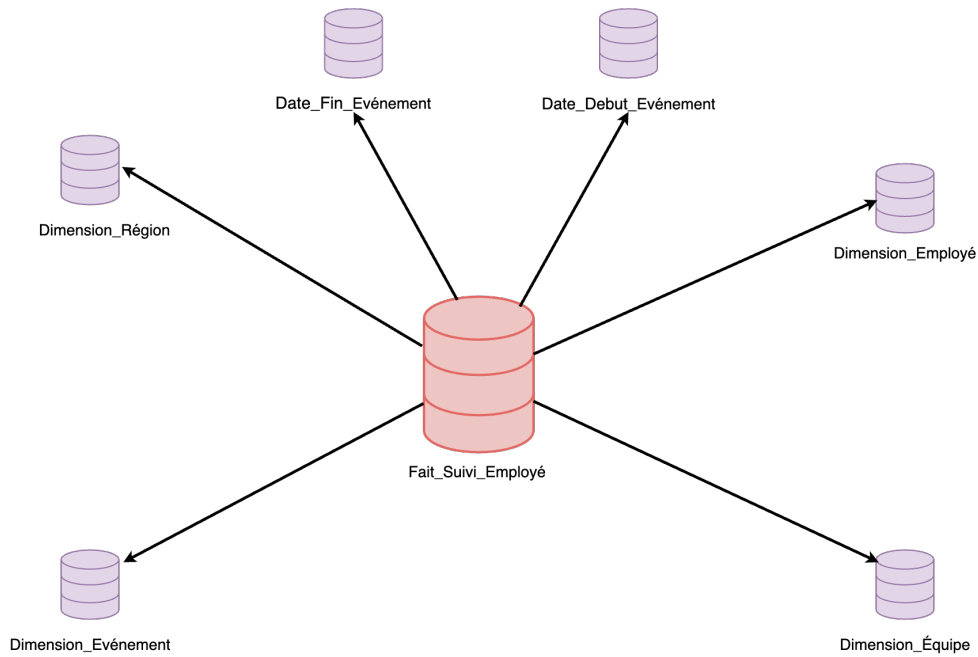


FIGURE 2.3 – Data Mart Suivi Employé

aux besoins d’analyse des coûts individuels et des durées exactes des affectations, rend la capture transactionnelle plus pertinente qu’une approche par snapshots qui générerait des données redondantes et pourrait manquer des événements courts entre deux captures d’état.

De plus, une approche par snapshots serait moins efficace car elle nécessiterait une fréquence très élevée de captures pour maintenir la précision temporelle requise dans les traitements (comme l’analyse des durées moyennes d’affectation), ce qui augmenterait significativement le volume de données stockées sans apporter de valeur ajoutée pour les analyses demandées de satisfaction des employés et d’impact des événements sur les équipes.

La Table 2.3 présentée ci-après détaille les mesures spécifiques utilisées dans la table des faits pour capturer ces indicateurs clés et leur évolution.

Mesure	Description	Caractère
Coût _Evenement	Somme dépensée par l’entreprise pour un employé.	Additive
Satisfaction _Employe	Score de satisfaction de l’employé (échelle de 1 à 10).	Non-additive
Durée	Durée passée par l’employé dans l’événement.	Additive

TABLE 2.3 – Table des mesures pour Fait _Suivi _Employe

Justification des types d’additivité

- **Additives** (Coût _Evenement, Durée) : Ces métriques sont additives car elles représentent des valeurs absolues. Par exemple, le coût total dépensé par l’entreprise pour un employé ou la durée totale passée par un employé dans un événement

peuvent être additionnés sur différentes dimensions (par exemple, sur différentes périodes ou événements) de manière cohérente, car chaque valeur est indépendante des autres.

- **Non-additive** (Satisfaction_Employe) : Un score de satisfaction ne peut pas être additionné directement sur les dimensions, car cela donnerait des résultats incohérents. Par exemple, additionner les scores de satisfaction des employés dans différents événements ou périodes ne permettrait pas de calculer un score global correct, car chaque score doit être traité indépendamment et ne peut pas être cumulé de manière simple.

2.3 Définition des dimensions

2.3.1 Dimensions analyse des performances des courses

Attribut	Type	Description
ID_Client (Clé primaire)	Int	Identifiant unique du client
Nom	Varchar(50)	Nom du client
Prénom	Varchar(50)	Prénom du client
Âge	Int	Âge du client
Sexe	Enum('H', 'F')	Genre du client
Type_Client	Varchar(20)	Type de client (régulier ou occasionnel)
Langue_Préférée	Varchar(20)	Langue préférée du client
Email	Varchar(50)	Adresse email du client
Téléphone	Varchar(15)	Numéro de téléphone du client
Point_gagnee	Int	Nombre de points gagnés suite aux courses

TABLE 2.4 – Attributs de la dimension **Client**

Attribut	Type	Description
ID_Chauffeur (Clé primaire)	Int	Identifiant du chauffeur
Nom	Varchar(26)	Nom du chauffeur
Prenom	Varchar(26)	Prénom du chauffeur
Date_Embauche	Date	Date d'embauche du chauffeur
Numéro_Permis	Varchar(20)	Numéro de permis de conduire
Type_Permis	Varchar(10)	Catégorie de permis (ex : B, C)
Téléphone	Varchar(15)	Numéro de téléphone du chauffeur
Email	Varchar(50)	Adresse email du chauffeur
Statut	Enum('actif', 'inactif')	Statut du chauffeur
Zone_Principale	Varchar(50)	Zone géographique où le chauffeur opère

TABLE 2.5 – Attributs de la dimension **Chauffeur**

Attribut	Type	Description
ID_Départ (Clé primaire)	Int	Identifiant du point de départ
Adresse	Varchar(100)	Adresse de départ du trajet
Ville	Varchar(50)	Ville de départ du trajet
Code_Postal	Varchar(20)	Code postal de départ du trajet
Distance_Centre	Int	Distance entre le point de départ et le centre-ville (en km)
Zone_Urbaine	Booléen	Indique si le point de départ est en zone urbaine.
Prise_en_Charge	Énumération (Pris_en_charge, Annulé, En_attente)	État de la course, sera utile pour déterminer si la course n'a pas été annulée
Description_Depart	Varchar(255)	Description supplémentaires concernant le départ

TABLE 2.6 – Attributs de la dimension **Départ**

Attribut	Type	Description
ID_Véhicule (Clé primaire)	Int	Identifiant unique du véhicule
Marque	Varchar(30)	Marque du véhicule
Modèle	Varchar(30)	Modèle du véhicule
Année	Int	Année de fabrication du véhicule
Km_Parcourus	Int	Kilométrage total du véhicule
Dernière_Révision	Date	Date de la dernière révision
Consommation	Float	Consommation moyenne en L/100 km
Assurance_Valide	Boolean	L'assurance est-elle valide ?
Type_véhicule	Énum('essence', 'électrique', 'diesel', 'hybrid')	Motorisation du véhicule
État_véhicule	Énum('bon', 'excellent', 'accidenté')	État du véhicule

TABLE 2.7 – Attributs de la dimension **Véhicule**

Attribut	Type	Description
ID_Arrivée (Clé primaire)	Int	Identifiant du point d'arrivée
Adresse	Varchar(100)	Adresse d'arrivée du trajet
Ville	Varchar(50)	Ville d'arrivée du trajet
Code_Postal	Varchar(10)	Code postal d'arrivée du trajet
Distance_Centre	Int	Distance entre le point d'arrivée et le centre-ville (en km)
Zone_Urbaine	Boolean	Le point d'arrivée est-il une zone urbaine ?
Description_arrivée	Varchar(255)	Description supplémentaire concernant le point d'arrivée

TABLE 2.8 – Attributs de la dimension **Arrivée**

Attribut	Type	Description
ID_Heure_Départ (Clé primaire)	Int	Identifiant de l'heure
Heure_Complète	Time	Heure au format HH :MM :SS
Période_Journée	Enum('matin', 'après-midi', 'soir', 'nuit')	Partie de la journée
AM_PM	Enum('AM', 'PM')	Avant midi , apres midi
Heure_Pointe	Boolean	Indique si c'est une heure de pointe (oui/non)
Heure_Format24h	Time	Heure au format 24 heures (HH :MM)
Fuseau_Horaire	Varchar(10)	Fuseau horaire associé
Heure_UTC	Time	Heure au format UTC (Universal Time Coordinated)

TABLE 2.9 – Attributs de la dimension **Heure_Départ**

Attribut	Type	Description
ID_Heure_Arrivée (Clé primaire)	Int	Identifiant de l'heure
Heure_Complète	Time	Heure au format HH :MM :SS
Période_Journée	Enum('matin', 'après-midi', 'soir', 'nuit')	Partie de la journée
AM_PM	Enum('AM', 'PM')	Avant midi , apres midi
Heure_Pointe	Boolean	Indique si c'est une heure de pointe (oui/non)
Heure_Format24h	Time	Heure au format 24 heures (HH :MM)
Fuseau_Horaire	Varchar(10)	Fuseau horaire associé
Heure_UTC	Time	Heure au format UTC (Universal Time Coordinated)

TABLE 2.10 – Attributs de la dimension **Heure_Arrivée**

Attribut	Type	Description
ID_Meteo (Clé primaire)	Int	Identifiant unique de l'enregistrement météo
Date_Meteo	Date	Date de la mesure météorologique
Temperature	Float	Température mesurée (en °C)
Humidite	Int	Taux d'humidité (entre 0 et 100)
Precipitations	Float	Quantité de précipitations (en mm, >= 0)
Vent_Vitesse	Float	Vitesse du vent (en km/h, >= 0)
Vent_Direction	Varchar(10)	Direction du vent (ex : N, NE, S)
Condition_Meteo	Varchar(50)	Description des conditions météorologiques (ex : ensoleillé, nuageux)
Pression_Atmo	Float	Pression atmosphérique (en hPa)
Indice_UV	Int	Indice UV (entre 0 et 11)

TABLE 2.11 – Attributs de la table **Meteo**

Attribut	Type	Description
ID_Date (Clé primaire)	Int	Identifiant de la date
Date_Complete	Date	Date complète (AAAA-MM-JJ)
Jour	Int	Jour du mois
Mois	Int	Mois de l'année
Année	Int	Année
Type_Jour	Enum('semaine', 'week-end', 'jour férié')	Type de jour
Trimestre	Int	Trimestre de l'année
Jour_Semaine	Varchar(10)	Jour de la semaine
Semaine_Anee	Int	Numéro de la semaine de l'année
Est_Férié	Boolean	Jour férié ou pas
Est_Weekend	Boolean	Jour est en week-end ou pas
Commentaire	Varchar(255)	Commentaire supplémentaire (ex. Évènement spécial, fêtes etc.)

TABLE 2.12 – Attributs de la dimension **Dates**

2.3.2 Dimensions gestion ressources humaines

Attribut	Type	Description
ID _Date _Debut _Evenement	Int	Identifiant de la Date du début de l'événement
Date _Debut _Evenement	Date	Date exacte du début de l'événement
Jour	Varchar(15)	Jour de la semaine (Lundi, Mardi, etc.)
Mois	Varchar(20)	Mois de l'année (Janvier, Février, etc.)
Trimestre	Varchar(2)	Le trimestre dans lequel l'événement commence (Q1, Q2, etc.)
Année	Int	L'année de début de l'événement
Saison	Varchar(20)	Saison correspondant à la date (Printemps, Été, Automne, Hiver)
Férié	Enumeration (Y , N)	Indication si la date est un jour férié ou non
Semaine	Numérique	Numéro de la semaine dans l'année
Jour _ou _Nuit	Booléen	Indication de l'événement débutant le jour ou la nuit
Mois/Année	Texte	Combinaison mois/année pour des rapports périodiques

TABLE 2.13 – Attributs de la dimension **Date _Debut _Evenement**

Attribut	Type	Description
ID_Date_Fin_Evénement	Int	Identifiant de la Date de fin de l'événement
Date_Fin_Evenement	Date	Date exacte de fin de l'événement
Jour	Varchar (15)	Jour de la semaine (Lundi, Mardi, etc.)
Mois	Varchar (20)	Mois de l'année (Janvier, Février, etc.)
Trimestre	Varchar (2)	Le trimestre dans lequel l'événement finit (Q1, Q2, etc.)
Année	Int	L'année de fin de l'événement
Saison	Varchar (20)	Saison correspondant à la fin de l'événement (Printemps, Été, Automne, Hiver)
Férié	Enumeration (Y, N)	Indication si la date est un jour férié ou non
Semaine	Numérique	Numéro de la semaine dans l'année
Jour_ou_Nuit	Booléen	Indication de la fin de l'événement en journée ou la nuit
Mois/Année	Texte	Combinaison mois/année pour des rapports périodiques

TABLE 2.14 – Attributs de la dimension **Date_Fin_Evenement**

Attribut	Type	Description
id_Région	Numérique	Code unique pour chaque région géographique
Nom_Région	Varchar(100)	Nom de la région
Pays	Varchar(50)	Pays auquel la région appartient
Zone_Géographique	Varchar(20)	Type de zone (urbaine, rurale, etc.)
Code_Postal	Int	Code postal de la région
Capital	Booléen	Indication si la région est une capitale ou non
Superficie	Int	Superficie de la région (en km ²)
Population	Int	Population approximative de la région
Statut_Régional	Texte	Indication si la région est en développement, stable, ou en déclin
Indice_de_développement	Int	Indice de développement de la région

TABLE 2.15 – Attributs de la dimension **Région**

Attribut	Type	Description
Id_Événement	Numérique	Identifiant unique pour chaque événement
Type_Événement	Varchar(50)	Type de l'événement (transfert, promotion, formation, etc.)
Description	Varchar(255)	Brève description de l'événement
Coût_Événement	Float	Coût total de l'événement pour l'entreprise
Priorité	Varchar(20)	Priorité de l'événement (haute, moyenne, basse)
Support	Enumeration(Y , N)	Indication si l'événement nécessite des ressources externes ou internes (ex : formation -> oui)
Type_Impact	Varchar(20)	Type d'impact (opérationnel ou financier)
Impact_Événement_Société	Varchar(20)	Impact de l'événement sur la société (qualifié de *positif* ou *négatif*)

TABLE 2.16 – Attributs de la dimension **Événement**

Attribut	Type	Description
Id_Équipe	Numérique	Code unique pour chaque équipe ou département
Nom_Équipe	Varchar(100)	Nom de l'équipe
Chef_d'Équipe	Varchar(100)	Nom du responsable de l'équipe
Taille_Équipe	Numérique	Nombre d'employés dans l'équipe
Type_Équipe	Varchar(50)	Type d'équipe
Budget_Équipe	Numérique	Budget alloué à l'équipe
Lieu_Équipe	Varchar(100)	Ville où se situe l'équipe
Mission_Principale	Varchar(255)	Mission ou objectif principal de l'équipe
Moyenne_Performance	Numérique	Indicateur de performance global de l'équipe

TABLE 2.17 – Attributs de la dimension **Équipe**

Attribut	Type	Description
ID_Employe	Numérique	Identifiant unique pour chaque employé
Nom_Employe	Varchar(100)	Nom de l'employé
Prenom_Employe	Varchar(100)	Prénom de l'employé
Date_Naissance	Date	Date de naissance de l'employé
Poste_Occupe	Varchar(50)	Poste occupé par l'employé
Date_Embauche	Date	Date d'embauche de l'employé
Salaire	Numérique	Salaire de l'employé
Email_Employe	Varchar(100)	Adresse email de l'employé
Telephone_Employe	Varchar(15)	Numéro de téléphone de l'employé
Adresse_Employe	Varchar(255)	Adresse de l'employé

TABLE 2.18 – Attributs de la dimension **Employé**

2.4 Traitements indiqués avec le modèle choisit

Dans cette section, nous évaluons la capacité de notre modèle à répondre aux besoins d'analyse énoncés dans la section 1.2. Pour cela, nous examinerons chaque traitement en vérifiant si notre modèle contient les données et les relations nécessaires entre les tables. Cette analyse nous permettra de déterminer si le modèle peut fournir les informations suffisantes pour effectuer les analyses requises.

2.4.1 Action 1 : Performances des courses 1.3

Traitement 1 : Analyser le nombre de courses effectuées par heure, jour et mois pour identifier les périodes de forte demande et en mesurer l'impact sur le chiffre d'affaires.

Le traitement repose sur la table `Fait_Course` et les dimensions `Heure_Arrive`, `Heure_Depart`, `Date`, `Client`, et `Chauffeur`. En agrégeant les données des courses par heure, jour et mois, il est possible de calculer la demande sur ces périodes. Cela permet d'identifier les heures de pointe, les jours ou les mois de forte demande et d'analyser l'impact de ces pics sur le chiffre d'affaires en croisant ces informations avec le revenu généré par chaque course. Une telle analyse aide à optimiser la gestion de la flotte et l'allocation des chauffeurs pour maximiser les revenus durant ces périodes.

Traitement 2 : Identifier les facteurs externes (événements, météo) influençant les pics de demande.

Ce traitement utilise la table `Fait_Course` et la dimension `Météo` pour analyser l'impact des conditions météorologiques sur la demande. En associant chaque course aux données météorologiques (telles que la température ou les précipitations) pour l'heure et la date correspondantes, on peut observer l'influence de la météo sur la fréquence des courses. Par exemple, la pluie ou la neige peut entraîner une augmentation de la demande. Cette analyse permet d'identifier des tendances saisonnières et météorologiques pour ajuster la gestion de la flotte en fonction des prévisions.

Traitement 3 : Évaluer la satisfaction des clients par région, en distinguant les courses longues et courtes, pour mieux orienter les offres et optimiser les investissements.

La table des faits capture les mesures clés, comme la distance parcourue, l'indice de satisfaction, et le prix, tandis que la dimension `Depart` fournit des informations géographiques sur la région de départ. En regroupant et en filtrant les données de la table des faits selon la distance (longue ou courte) et en croisant ces informations avec les dimensions, nous pouvons calculer rapidement des métriques agrégées comme la moyenne de satisfaction et le volume de courses. Ce design favorise une navigation intuitive et performante pour produire des analyses ciblées sur les besoins d'investissement et d'amélioration des courses selon leurs types.

Traitement 4 : Identifier les clients qui utilisent le moins fréquemment les services UberX afin d'améliorer leur fidélité.

En croisant les données de la table `Fait_Course` avec la dimension `Client`, ce traitement permet d'identifier les clients ayant une faible fréquence d'utilisation du service

UberX. L'analyse repose sur le nombre de courses effectuées par chaque client, ce qui permet de repérer ceux qui utilisent rarement le service. Une fois ces clients identifiés, des stratégies de fidélisation peuvent être mises en place, telles que des promotions ciblées, des récompenses ou des offres spéciales, afin d'inciter ces clients à utiliser plus fréquemment les services UberX.

Traitement 5 : Analyse des performances des chauffeurs

La table des faits **Fait_Course** capture les informations clés telles que le nombre de courses, le chiffre d'affaires, la distance parcourue et l'indice de satisfaction client. Les **dimensions** associées, telles que **Chauffeur**, **Client** et la dimension dégénérée **ID_Course**, permettent d'analyser les performances des chauffeurs. Grâce à cette structure, nous pouvons facilement agréger les données à différents niveaux pour évaluer la performance des chauffeurs en termes de volumes de courses, de revenus générés et de satisfaction des clients.

Action 2 : Gestion des ressources humaines 1.3

Traitement 1 : Analyser les scores de satisfaction des employés par équipe et par région

Ce traitement peut être réalisé en utilisant la dimension **Employé** pour accéder aux scores de satisfaction de chaque employé, qui sont stockés dans la mesure **Satisfaction_Employé** de la table de faits. Les dimensions **Équipe** et **Région** permettent d'agréger ces scores de satisfaction par équipe ou par région, facilitant ainsi l'identification des zones de faible satisfaction. Une fois ces zones identifiées, il est possible de mettre en place des actions ciblées pour améliorer la satisfaction des employés, comme des formations ou des réorganisations d'équipes.

Traitement 2 : Identifier les événements les plus coûteux et évaluer leur impact

Le coût des événements est stocké dans la mesure **Coût_Evenement** de la table de faits. En associant cette mesure aux dimensions **Événement** et **Équipe**, il est possible d'identifier les événements les plus coûteux et de déterminer leur impact sur les équipes et l'entreprise. L'analyse de l'impact (positif ou négatif) permet de déterminer si l'investissement dans ces événements est justifié ou s'il faut réévaluer les priorités.

Traitement 3 : Suivre la durée moyenne des affectations des employés par type d'événement

La durée des affectations des employés est mesurée par la mesure **Durée** dans la table de faits, associée aux dimensions **Événement** et **Employé**. En analysant cette donnée, il est possible de suivre la durée moyenne des affectations des employés pour chaque type d'événement, comme les formations, les transferts ou les sanctions. En comparant ces durées par type d'événement, on peut identifier des déséquilibres, tels que des rotations excessives ou des interruptions dans les affectations des employés. Cela permettra d'optimiser les processus en ajustant les affectations pour améliorer la stabilité des équipes et réduire les périodes d'interruption.

Traitement 4 : Analyser la relation entre la durée des affectations et les niveaux de satisfaction des employés

Ce traitement peut être réalisé en utilisant les mesures **Durée** et **Satisfaction** **Employé** de la table de faits **Fait_Suivi_Employe**. Les dimensions **Événement** et **Employé** permettent d’analyser les données en regroupant par type d’événement.

En calculant des statistiques telles que la **durée moyenne**, la **durée maximale**, la **durée minimale**, ainsi que la **satisfaction moyenne** pour chaque type d’événement, il devient possible d’identifier les affectations problématiques. Par exemple, des affectations excessivement longues ou fréquentes qui nuisent à la satisfaction des employés.

Cette analyse permet de mettre en place des ajustements tels que :

- Réduire la durée ou la fréquence des affectations problématiques.
- Réorganiser les plannings ou les types d’événements pour améliorer l’expérience des employés.

Ces ajustements contribuent à optimiser la gestion des affectations et à augmenter la satisfaction globale des employés.

2.5 Instance du modèle

Dans cette section, nous illustrons notre modèle en présentant un échantillon de données pour chaque table afin d’évaluer sa pertinence et sa capacité à répondre aux besoins d’analyse. Les tables ont été créées en SQL, et les Figures 2.4 à 2.20 montrent le résultat de requêtes `SELECT * FROM table` exécutées sur chacune d’elles. Nous débutons par le modèle Course, décrit dans la section 2.2.1, en présentant d’abord sa table de faits puis ses dimensions. La même démarche est appliquée au modèle Ressources Humaines, abordé dans la section 2.2.2.

2.5.1 Instance du Data Mart Courses

Table des Faits

ID_COURSE	ID_DATE	ID_ARRIVEE	ID_DEPART	ID_CLIENT	ID_HEURE_ARRIVEE	ID_HEURE_DEPART	ID_CHAUFFEUR	ID_VEHICULE	ID_METEO	PRIX	DISTANCE_PARCOURUE	DUREE_DU_TRAJET	NOMBRE_PASSAGERS
TEMPS_D'ATTENTE_CLIENT	REVENUE_MOYEN_PAR_COURSE	INDICE_SATISFACTION_CLIENT											
19	9 4	2	3	19	38 4,2	39	3	4	5	30	15	50	3
20	2 4	1	4 28,5	20	36 4,2	37	9	2	1	31,5	16,2	70	2
21	1 6	3	4 23,5	21	32 4,5	33	1	1	2	26,5	13,2	35	3

FIGURE 2.4 – Table des Faits Course

Dimensions

ID_CLIENT NOM TELEPHONE	POINT_GAGNEE	PRENOM	AGE S	TYPE_CLIENT	LANGUE_PREFEREE	EMAIL
78 Meier 0123456866	55	Anna	31	F Standard	Allemand	anna.meier@example.com
79 Sharma 0123456867	45	Priya	27	F Standard	Hindi	priya.sharma@example.com
80 Rodriguez 0123456868	90	Maria	34	F Premium	Espagnol	maria.rodriguez@example.com
81 Kim 0123456869	125	Jiwoo	39	F VIP	Coréen	jiwoo.kim@example.com
82 Davies 0123456870	40	Oliver	28	H Standard	Anglais	oliver.davies@example.com
83 Singh 0123456871	85	Rohit	32	H Premium	Hindi	rohit.singh@example.com

FIGURE 2.5 – Table Client

ID_CHAUFFEUR NOM EMAIL	PRENOM	STATUT	DATE_EMBAU	NUMERO_PERMIS	TYPE_PERMIS	TELEPHONE
ZONE_PRINCIPALE						
28 Gauthier laetitia.gauthier@example.org Bordeaux	Laetitia	Actif	12/05/2020	78901234532	D	+33 5 33 44 55 66
29 Poulain theo.poulain@example.com Strasbourg	Th??o	Inactif	20/10/2021	89012345643	B	+33 4 44 55 66 77

FIGURE 2.6 – Table Chauffeur

ID_VEHICULE	MARQUE	MODELE	ANNEE	TYPE_VEHICULE	ETAT_VEHICULE	KM_PARCOURS	DERNIERE_R	CONSOMMATION	ASSURANCE_VALIDE
1	Toyota	Corolla	2019	Essence	Excellent	30000	15/05/2023	6,5	1
2	Ford	Focus	2020	Diesel	Bon	45000	22/11/2022	5,2	1
3	Tesla	Model 3	2021	Electrique	Excellent	12000	05/01/2024	0	1
4	BMW	320i	2018	Essence	Mauvais	75000	30/06/2022	8	1
5	Audi	A3	2017	Diesel	Bon	60000	18/07/2023	7,2	1
6	Volkswagen	Golf	2019	Essence	Excellent	25000	10/04/2023	6	1
7	Honda	Civic	2020	Essence	Bon	35000	25/03/2023	7,5	1
8	Nissan	Altima	2021	Diesel	Excellent	15000	20/02/2024	5,8	1
9	Chevrolet	Malibu	2022	Hybride	Bon	5000	15/03/2024	4,5	1
10	Hyundai	Elantra	2020	Essence	Mauvais	80000	11/09/2022	8,3	1
11	Mercedes	A-Class	2021	Essence	Excellent	10000	05/12/2023	6,2	1

FIGURE 2.7 – Table Vehicule

ID_DEPART	ADRESSE	VILLE	CODE_POSTA	DISTANCE_CENTRE	ZONE_URBAINE	P
DESCRIPTION_DEPART						
94	330, rue J??r??me Dupr??	Delorme	26774	4364	1	A
Ignorer rouler convenir neuf ciel l une trace.						
95	315, avenue Carlier	Denis	62642	544	0	E
n Attente Passage derri??re large danser.						

FIGURE 2.8 – Table Depart

ID_ARRIVEE	ADRESSE	VILLE	CODE_POSTA	DISTANCE_CENTRE	ZONE_URBAINE
DESCRIPTION_ARRIVEE					
91	95, rue Anouk Normand	Saint Paulette-sur-Mer	07226	1588	0
Apprendre type d??fendre libert?? si croire.					
92	92, boulevard de Martel	Jacob	52676	3525	1
Puis fleur si peuple image naturellement.					
93	4, boulevard Valentine Peron	GuyotBourg	19172	1572	0
Mille papier moyen v??ritable ??lever.					

FIGURE 2.9 – Table Arrivee

ID_HEURE_DEPART	HEURE_CO	PERIODE_JOURNEE	AM_PM	HEURE_POINTE	HEURE_FO	FUSEAU_HOR	HEURE_UT
67	22:00:00	nuit	PM	0	22:00	CET	21:00
68	00:30:00	nuit	AM	0	00:30	CET	23:30
69	03:00:00	matin	AM	1	03:00	CET	02:00
70	06:15:00	matin	AM	1	06:15	CET	05:15
71	08:30:00	matin	AM	1	08:30	CET	07:30
72	11:00:00	matin	AM	1	11:00	CET	10:00
73	13:15:00	apres-midi	PM	0	13:15	CET	12:15
74	16:30:00	apres-midi	PM	0	16:30	CET	15:30
75	19:45:00	soir	PM	1	19:45	CET	18:45
76	22:00:00	nuit	PM	0	22:00	CET	21:00
77	00:15:00	nuit	AM	0	00:15	CET	23:15

FIGURE 2.10 – Table Heure_Depart

Dimensions

ID_Evenement	Type_Evenement	Description_Evenement	Cout_Evenement	Priorite	Support	Type_Impact	Impact_Evenement_Societe
1	Promotion	Augmentation de poste	5000	Haute	Y	Financier	Positif
2	Formation	Formation en leadership	3000	Moyenne	N	Opérationnel	Positif
3	Réorganisation	Changement d'équipe	2000	Moyenne	Y	Opérationnel	Neutre

FIGURE 2.15 – Table Evenement

ID_Date_Deb...	Date_Debut_Ev...	Jour	Mois	Trimestre	Annee	Saison	Ferie	Semaine	Jour_ou_Nuit	Mois_Annee
1	2024-01-01	Lundi	Janvier	Q1	2024	Hiver	N	1	J	Janvier/2024
2	2024-02-14	Mercredi	Février	Q1	2024	Hiver	N	7	N	Février/2024
3	2024-03-10	Dimanche	Mars	Q1	2024	Printemps	N	10	J	Mars/2024

FIGURE 2.16 – Table Date_Debut_Evenement

ID_Date_Fin...	Date_Fin_Even...	Jour	Mois	Trimestre	Annee	Saison	Ferie	Semaine	Jour_ou_Nuit	Mois_Annee
1	2024-01-02	Mardi	Janvier	Q1	2024	Hiver	N	1	J	Janvier/2024
2	2024-02-15	Jeudi	Février	Q1	2024	Hiver	N	7	N	Février/2024
3	2024-03-11	Lundi	Mars	Q1	2024	Printemps	N	10	J	Mars/2024

FIGURE 2.17 – Table Date_Fin_Evenement

ID_Employe	Nom_Employe	Prenom_Employe	Date_Naissance	Poste_Occupe	Date_Embauche	Salaire	Email_Employe	Telephone_Empl...	Adresse_Employe
1	Martin	Alice	1990-01-01	Développeuse	2015-06-01	45000	alice.martin@exam...	0612345678	10 Rue de Paris, Paris
2	Dupont	Bob	1985-03-15	Manager	2010-09-01	55000	bob.dupont@exam...	0623456789	20 Avenue de Lyon, Lyon
3	Dubois	Caroline	1988-07-12	Commerciale	2012-04-10	48000	caroline.dubois@ex...	0634567890	15 Avenue des Champs, ...

FIGURE 2.18 – Table Employe

ID_Equipe	Nom_Equipe	Chef_d_Equipe	Taille_Equipe	Type_Equipe	Budget_Equipe	Lieu_Equipe	Mission_Principale	Moyenne_Performance
1	Développement	Alice Martin	10	Projet	50000	Paris	Créer un nouveau pro...	8.5
2	Marketing	Bob Dupont	8	Support	40000	Lyon	Augmenter la visibilité	7.8
3	Ventes	Caroline Dubois	12	Projet	45000	Marseille	Accroître les ventes	8.2

FIGURE 2.19 – Table Equipe

ID_Region	Nom_Region	Pays	Zone_Geographi...	Code_Postal	Capital	Superficie	Population	Statut_Regional	Indice_de_Developpem...
1	Auvergne-Rhône-Al...	France	Urbaine	34000	N	72787	2950000	Stable	0.89
2	Bourgogne-Franch...	France	Rurale	21000	N	47784	2800000	En déclin	0.86
3	Bretagne	France	Rurale	35000	N	27288	3340000	Stable	0.88

FIGURE 2.20 – Table Region

2.6 Estimation de la taille des tables

Dans cette section, nous estimons la taille de nos tables après **12 mois**. Il est toujours important d’avoir une idée préalable de la taille des données à laquelle notre entrepôt de données fera affaire afin de pouvoir évaluer sa performance dessus.

2.6.1 Fait_course

Nous estimons le nombre d’utilisateurs annuels d’Uber en France à 5 millions, avec une moyenne de 2 courses par jour pour chaque client.

Sur 12 mois, la table des faits retraçant les performances des courses contiendra environ :

$$12 \times 5M \times 2 = \mathbf{120} \text{ millions de lignes.}$$

Une requête analytique sur cette table pourrait, selon sa complexité, prendre plusieurs minutes. Cependant, nous pouvons réduire significativement le temps d’exécution en diminuant le nombre de lignes à parcourir grâce à la mise en place de *vues matérialisées*.

2.6.2 Fait_RH

Le nombre total d’employés UberX en France, incluant les chauffeurs, est estimé à **30 000**. Chaque employé fait l’objet, en moyenne, de **4 événements** par an, tels que des formations, réaffectations ou sanctions.

Ainsi, la table des faits dédiée au suivi des événements des employés contiendra :

$$30\,000 \times 4 = \mathbf{120\,000} \text{ lignes.}$$

Avec ce volume relativement modeste, il est possible d’optimiser les performances des requêtes analytiques sans nécessairement recourir aux *vues matérialisées*. Toutefois, leur mise en place reste une option judicieuse, notamment pour des analyses complexes ou fréquemment exécutées, afin d’améliorer davantage la réactivité du modèle.

Conclusion

Ce chapitre a permis de concevoir un entrepôt de données structuré et adapté aux besoins analytiques, en tenant compte des spécificités de chaque domaine, qu’il s’agisse de la gestion des courses ou des ressources humaines. La modélisation de nos tables de faits et dimensions a été orientée pour optimiser la pertinence et la précision des analyses, facilitant ainsi des prises de décisions éclairées.

En analysant les Data Marts, nous avons mis en évidence les relations essentielles entre les données, assurant une cohérence dans les traitements d’analyse, qu’il s’agisse de mesurer la performance des courses ou d’évaluer la productivité des chauffeurs. Cette structure de données garantit une vision globale et détaillée, permettant des analyses fines et ciblées, tant au niveau opérationnel que stratégique. Le modèle d’entrepôt ainsi conçu ouvre la voie à une exploitation flexible et évolutive des données, prête à répondre aux futurs besoins analytiques de l’organisation.

Dans le chapitre suivant, nous aborderons l’implémentation de ce modèle en utilisant Oracle, en détaillant les étapes de création des tables, d’intégration des données et de gestion des contraintes afin d’assurer une mise en œuvre robuste et performante de l’entrepôt de données.

Chapitre 3

Implémentation de l'entrepôt de données

Dans ce chapitre, nous détaillons l'implémentation de l'entrepôt de données, en réponse aux questions 11 et 12. Les fichiers associés aux dimensions, à la table des faits, et aux requêtes analytiques sont fournis en annexe.

3.1 Action 1 : Performance de course

3.1.1 Implémentation des dimensions et de la table de faits

Conformément à la question 11, nous avons créé les dimensions et la table des faits en SQL. Les dimensions ont été implémentées en tant que tables, ce qui permet une gestion plus claire et structurée des données.

Voici un exemple de la création de la table pour la dimension `Client` :

```
CREATE TABLE Client (  
    ID_Client INT PRIMARY KEY,  
    Nom VARCHAR2(50),  
    Prenom VARCHAR2(50),  
    Age INT,  
    Sexe CHAR(1) CHECK (Sexe IN ('H', 'F')),  
    Type_Client VARCHAR2(20),  
    Langue_Preferee VARCHAR2(20),  
    Email VARCHAR2(50),  
    Telephone VARCHAR2(15),  
    Point_gagnee INT  
);
```

Pour consulter toutes les dimensions, vous pouvez vous référer au fichier `dimension_course.sql`. En ce qui concerne la table des faits, elle a été créée selon les spécifications du modèle, et le fichier `table_fait_course.sql` contient la définition complète de cette table.

3.1.2 Requêtes analytiques

En réponse à la question 12, nous avons développé plusieurs requêtes analytiques correspondant aux traitements que nous avons identifiés. Ces requêtes sont détaillées dans le fichier `requetes_course.sql`.

Voici un exemple de requête analytique :

- **Nombre de courses par mois et chiffre d'affaires généré** : Cette requête met en évidence les mois les plus rentables, permettant une analyse saisonnière.

```
SELECT
    d.Mois,
    d.Annee,
    COUNT(c.ID_COURSE) AS Nombre_Courses,
    SUM(c.Prix) AS Chiffre_Affaires
FROM
    Course c
JOIN
    Dates d ON c.ID_DATE = d.ID_DATE
GROUP BY
    d.Mois, d.Annee
ORDER BY
    Nombre_Courses DESC;
```

Cette requête permet de comprendre la répartition des courses effectuées au cours de l'année, tout en calculant le chiffre d'affaires généré par mois. D'autres requêtes analytiques, comme celles mentionnées dans la section précédente, permettent d'explorer différents aspects de l'activité, notamment la satisfaction client, l'efficacité des chauffeurs et l'impact des promotions.

3.2 Action 2 : Suivi des employés et événements

3.2.1 Implémentation des dimensions et de la table de faits

Pour répondre à la question 11, nous avons créé les dimensions et la table des faits liées au suivi des employés et des événements en SQL. L'ensemble des scripts de création est disponible dans le fichier `creation_suivi_employe.sql`.

Exemple : Création de la table des faits `Fait_Suivi_Employe`

```
CREATE TABLE Fait_Suivi_Employe (
    -- Clés étrangères vers les dimensions
    ID_Date_Debut_Evenement INT NOT NULL,
    ID_Date_Fin_Evenement INT NOT NULL,
    ID_Employe NUMBER NOT NULL,
    ID_Region NUMBER NOT NULL,
    ID_Evenement NUMBER NOT NULL,
    ID_Equipe NUMBER NOT NULL,
    Cout_Evenement NUMBER(15, 2) NOT NULL,
    Satisfaction_Employe NUMBER(2, 1)
    CHECK (Satisfaction_Employe BETWEEN 1 AND 10),
    Duree NUMBER(5, 2) NOT NULL,
    PRIMARY KEY (ID_Date_Debut_Evenement, ID_Date_Fin_Evenement, ID_Employe,
    ID_Region, ID_Evenement, ID_Equipe),
    FOREIGN KEY (ID_Date_Debut_Evenement) REFERENCES
    Date_debut_evenement (ID_Date_Debut_Evenement),
```

```

    FOREIGN KEY (ID_Date_Fin_Evenement) REFERENCES
    Date_fin_evenement (ID_Date_Fin_Evenement),
    FOREIGN KEY (ID_Employe) REFERENCES
    Dimension_Employe (ID_Employe),
    FOREIGN KEY (ID_Region) REFERENCES
    Dimension_Region (ID_Region),
    FOREIGN KEY (ID_Evenement) REFERENCES
    Dimension_Evenement (ID_Evenement),
    FOREIGN KEY (ID_Equipe) REFERENCES
    Dimension_Equipe (ID_Equipe)
);

```

La table des faits utilise les clés étrangères pour relier les dimensions, tout en incluant des mesures comme le coût des événements, la satisfaction des employés et leur durée.

3.2.2 Requêtes analytiques

Pour répondre à la question 12, plusieurs requêtes analytiques ont été développées pour explorer et analyser les données relatives aux employés et aux événements. Ces requêtes sont regroupées dans le fichier `requetes_suivi_employe.sql`.

Exemple de requête : Analyse des coûts et de la participation par type d'événement

```

SELECT
    Ev.Type_Evenement AS Type_Ev,
    Ev.Description_Evenement AS Descr,
    Ev.Impact_Evenement_Societe AS Impact,
    Eq.Nom_Equipe AS Equipe,
    SUM(F.Cout_Evenement) AS Cout_Total,
    COUNT(F.ID_Employe) AS Nbre_Part,
    COUNT(DISTINCT F.ID_Equipe) AS Nombre_Equipes
FROM
    Fait_Suivi_Employe F
JOIN
    Dimension_Evenement Ev ON F.ID_Evenement = Ev.ID_Evenement
LEFT JOIN
    Dimension_Equipe Eq ON F.ID_Equipe = Eq.ID_Equipe
GROUP BY
    Ev.Type_Evenement, Ev.Description_Evenement,
    Ev.Impact_Evenement_Societe, Eq.Nom_Equipe
ORDER BY
    Cout_Total DESC
FETCH FIRST 10 ROWS ONLY;

```

Cette requête répond au **Traitement 2**, en identifiant :

- Les types d'événements ayant le coût total le plus élevé.
- Le nombre total de participants et d'équipes impliquées.
- Les impacts des événements sur la société.

Les résultats de cette requête peuvent être utilisés pour :

- Prioriser les types d'événements ayant un impact stratégique.
- Analyser les dépenses en fonction de la participation et de l'efficacité.

3.3 Vues Matérialisées

Dans notre projet, les vues matérialisées sont essentielles pour optimiser les performances des requêtes analytiques complexes, particulièrement face au volume élevé de données comme c'est le cas pour nos tables (cf. 2.6). Bien que le modèle en étoile simplifie la structure des requêtes, il reste nécessaire de les optimiser davantage pour répondre efficacement aux besoins décisionnels.

En stockant des **agrégats pré-calculés**, les vues matérialisées réduisent significativement le temps d'exécution des traitements et facilitent l'accès à des données consolidées.

3.3.1 Selection de Vues

Notre objectif dans cette section est de créer un ensemble minimal de vues qui permettra de répondre à toutes nos requêtes. pour cela nous utiliserons la méthode du **treillis d'agrégation**.

Employés

Le treillis d'agrégation présenté à la Figure 3.1 illustre la distribution des requêtes : deux sont liées au nœud **Evenement**, une à **Evenement**, **Equipe**, et une autre à **Equipe**, **Region**. Étant donné qu'une requête group-by pouvant être résolue par un nœud du treillis peut également être satisfaite par n'importe lequel de ses ancêtres, deux options s'offrent à nous : choisir la **racine** ou sélectionner les nœuds **rouge** et **vert**

Nous optons pour les nœuds rouge et vert pour construire nos vues matérialisées. En effet, choisir la racine entraînerait une surcharge importante en termes de rafraîchissements, chaque insertion dans l'une des trois tables nécessitant une mise à jour, ce qui n'est pas optimal pour nos besoins.

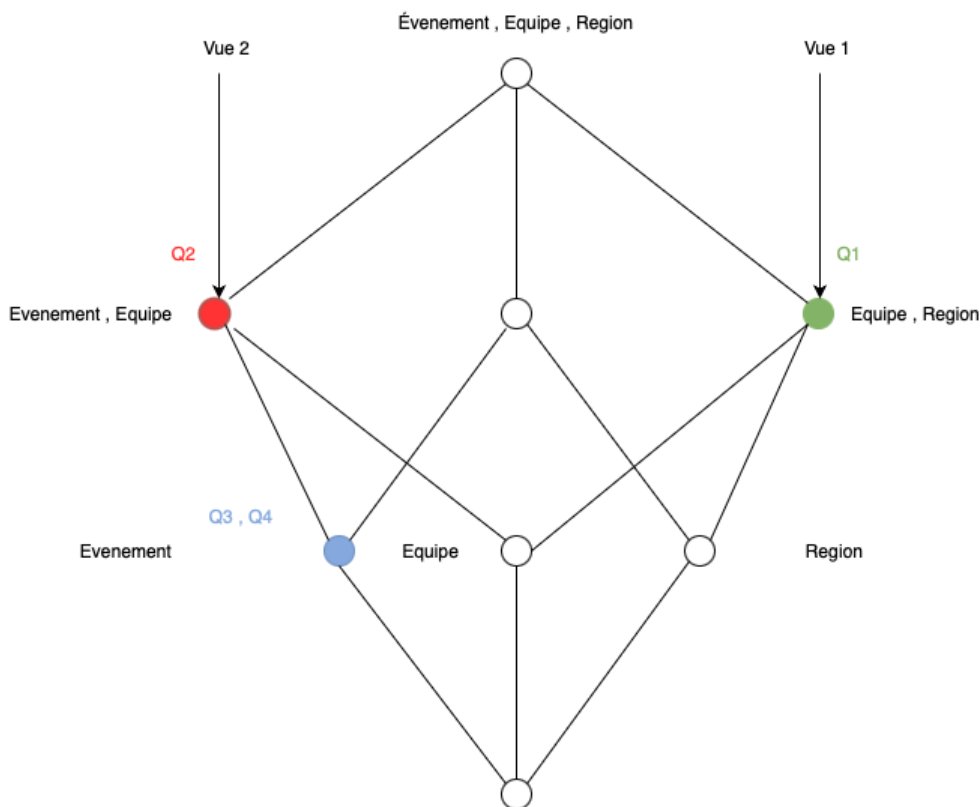


FIGURE 3.1 – Treillis d'agrégation

Ainsi, nous pouvons exploiter la deuxième vue (**Vue_Evenement_Equipe**) pour répondre aux requêtes 2, 3 et 4, et la première vue (**Vue_Satisfaction_Equipe_Region**) pour la requête 1. Cette approche simplifie considérablement les requêtes en réduisant le volume de calculs et le nombre de lignes à traiter, tout en garantissant une analyse rapide et efficace.

Cependant, cette solution présente un inconvénient : les vues matérialisées doivent être régulièrement rafraîchies pour intégrer les données nouvellement insérées. Le rafraîchissement peut être effectué simplement à l'aide de la commande suivante :

```
REFRESH MATERIALIZED VIEW <nom_de_la_vue>;
```

Course

Pour optimiser les performances et simplifier les analyses sur un volume conséquent de données environ **120 millions de lignes sur 12 mois**, nous avons adopté une approche basée sur **trois vues matérialisées**. Ces vues permettent de pré-calculer les agrégats et les jointures nécessaires, réduisant ainsi la charge sur le système et accélérant l'exécution des requêtes complexes ou répétitives.

La première vue, **Vue_Analyse_Traite1_2_3**, regroupe les données temporelles et contextuelles liées aux périodes de forte demande (heures, jours, mois) et aux facteurs externes comme la météo. Elle couvre les **Traitements 1(requetes 1 et 2), 2 et 3**, qui incluent l'analyse des pics de demande, leur impact sur le chiffre d'affaires, ainsi que l'étude des facteurs externes influençant ces variations.

La deuxième vue, `Vue_Analyse_Traite4`, est dédiée au suivi des **clients peu fréquents** dans le cadre du **Traitement 4**. En se concentrant sur les métriques essentielles comme le nombre de courses et le montant dépensé par client, cette vue simplifiera les requêtes liées à l'identification des utilisateurs nécessitant des efforts de fidélisation.

Enfin, la troisième vue, `Vue_Analyse_Traite5`, cible les **performances des chauffeurs** pour le **Traitement 5**. Elle agrège les données pertinentes, telles que le nombre de courses, le chiffre d'affaires, la distance parcourue et la satisfaction des clients, permettant une évaluation précise et rapide des performances.

En mettant ces 3 vues en place, nous évitons le traitement de données inutiles pour certains types d'analyses tout en répondant à l'ensemble des besoins analytiques. Cette approche limite les redondances, améliore les temps de traitement et garantit une utilisation efficace des ressources.

Conclusion

Ce chapitre a détaillé l'implémentation des tables, des requêtes, et des vues utilisées pour répondre aux besoins analytiques spécifiques du projet. Nous avons conçu une structure de base de données optimisée pour gérer un volume important de données, tout en assurant la flexibilité nécessaire pour des analyses fréquentes et variées.

Les tables relationnelles ont été définies de manière à capturer les informations essentielles, avec une normalisation adéquate pour éviter la redondance tout en garantissant l'intégrité des données. Les requêtes SQL ont été soigneusement élaborées pour répondre aux différents traitements, en s'appuyant sur des critères de performance, de lisibilité, et de pertinence.

L'intégration des vues matérialisées constitue un point clé de notre solution. Ces vues permettent d'agréger et de pré-calculer les données les plus demandées, réduisant ainsi la charge des calculs en temps réel tout en améliorant significativement les temps de réponse. En limitant le nombre de vues à trois, nous avons réussi à équilibrer efficacité et simplicité, tout en répondant à l'ensemble des traitements nécessaires, qu'il s'agisse d'analyser la demande, d'évaluer la satisfaction des clients, ou d'estimer les performances des chauffeurs.

Cette implémentation constitue une base robuste et performante pour des analyses futures. Elle garantit une gestion optimale des ressources tout en fournissant des outils puissants pour extraire des informations stratégiques. Le travail réalisé dans ce chapitre prépare également le terrain pour d'éventuelles extensions ou ajustements en fonction des évolutions des besoins métier.

Readme

Description

Ce projet contient deux modules principaux : **Course** et **Employes**, chacun organisé dans des dossiers distincts. Chaque dossier regroupe les fichiers nécessaires à la création des tables, au remplissage des données, ainsi qu'à l'exécution des requêtes et des vues associées.

Structure des dossiers

1. Course

- `dimensions_course.sql` : Création des dimensions.
- `table_fait_course.sql` : Création de la table de faits.
- `requetes_course.sql` : Requêtes associées analytiques.
- `vues_course.sql` : Création des vues et exécution des requêtes dépendant des vues.

2. Employes

- `creation_suivi_employe.sql` : Création des tables pour le suivi des employés (dimensions et table des faits).
- `remplissage_suivi_employe.sql` : Remplissage des données dans les tables créées.
- `requetes_suivi_employe.sql` : Requêtes analytiques.
- `vues_suivi_employe.sql` : Création des vues et requêtes supplémentaires basées sur ces vues.

Instructions d'exécution

1. Course

1. Création des tables :

- Exécuter `dimensions_course.sql` pour créer les tables dimensionnelles.
- Exécuter `table_fait_course.sql` pour créer la table de faits.

2. Exécution des requêtes :

- Exécuter `requetes_course.sql` pour tester les requêtes liées aux tables créées.

3. Création et exécution des vues :

- Exécuter `vues_course.sql` pour créer les vues et exécuter les nouvelles requêtes dépendantes des vues.

2. Employes

1. Création des tables :

- Exécuter `creation_suivi_employe.sql` pour créer les tables nécessaires.

2. Insertion des données :

- Exécuter `remplissage_suivi_employe.sql` pour remplir les tables avec des données pertinentes.

3. Exécution des requêtes :

- Exécuter `requetes_suivi_employe.sql` pour tester les requêtes sur les données insérées.

4. Création et exécution des vues :

- Exécuter `vues_suivi_employe.sql` pour créer les vues et exécuter les requêtes dépendantes des vues.

Remarque

Nous n'avons pas pu tester les vues et les nouvelles requêtes par manque de privilèges Oracle, mais leur logique a été soigneusement réfléchie et implémentée.

Conclusion

Dans ce rapport, nous avons montré qu'Uber, confronté à une concurrence accrue dans le secteur du transport, doit améliorer son service UberX pour maintenir sa position dominante. Pour soutenir cette démarche, nous avons mis en place un entrepôt de données permettant d'effectuer des requêtes analytiques, facilitant ainsi la collecte d'informations essentielles à la prise de décision. Nous avons aussi défini plusieurs objectifs stratégiques pour l'entreprise, notamment la fidélisation des clients, l'augmentation de la rentabilité, et l'optimisation de la gestion des ressources humaines.

Afin de répondre à ces besoins, deux schémas en étoile ont été conçus : l'un pour analyser les performances des courses, en tenant compte de la demande, de la satisfaction des clients et des revenus générés ; l'autre pour suivre l'évolution des employés, leur satisfaction et leur performance, tout en optimisant la gestion des ressources humaines et des événements qui les affectent. Ces schémas permettent à Uber de prendre des décisions plus informées grâce à des données claires et accessibles.

En conclusion, l'implémentation de ces outils analytiques, combinée à leur capacité à fournir des informations détaillées en temps réel, permettra à Uber d'atteindre ses objectifs stratégiques de manière efficace, tout en minimisant le temps d'exécution des requêtes. Cela contribuera à consolider sa position sur le marché, à optimiser la gestion de ses ressources humaines et à accroître sa rentabilité.

Nous tenons à vous exprimer notre gratitude pour l'attention portée à ce travail et l'intérêt manifesté. Votre engagement et vos réflexions sont grandement appréciés.

Table des figures

1.1	Logo de l'entreprise	3
2.1	UberX Data-warehouse	10
2.2	Data Mart Course	12
2.3	Data Mart Suivi Employé	15
2.4	Table des Faits Course	27
2.5	Table Client	27
2.6	Table Chauffeur	28
2.7	Table Vehicule	28
2.8	Table Depart	28
2.9	Table Arrivee	28
2.10	Table Heure_Depart	28
2.11	Table Heure_Arrivee	29
2.12	Table Meteo	29
2.13	Table Date	29
2.14	Table Fait_Suivi_Employe	29
2.15	Table Evenement	30
2.16	Table Date_Debut_Evenement	30
2.17	Table Date_Fin_Evenement	30
2.18	Table Employe	30
2.19	Table Equipe	30
2.20	Table Region	30
3.1	Treillis d'agrégation	36