**Problématique :**

  Dans un environnement commercial de plus en plus compétitif, la fidélisation de la clientèle est devenue un pilier fondamental pour la pérennité et la croissance d'une entreprise. Dans cette dynamique, les systèmes de fidélisation de la clientèle jouent un rôle crucial en permettant aux entreprises de comprendre les besoins et les préférences de leurs clients fidèles de manière approfondie. La fromagerie, consciente de l'importance stratégique de la fidélisation de sa clientèle, a mis en place un système sophistiqué de fidélisation basé sur un datawarehouse. Dans le cadre de ce système, elle s'efforce d'extraire des informations pertinentes et exploitables à partir des données de fidélisation afin de perfectionner ses programmes de fidélisation et d'enrichir sa relation avec sa clientèle.

Le LOT 1 du projet vise à une analyse minutieuse des données de fidélisation des clients de la fromagerie pour la période allant de 2006 à 2010, avec une attention particulière portée aux départements 53, 61 et 28. Cette analyse approfondie permettra à l'entreprise de comprendre en profondeur les comportements d'achat de ses clients fidèles dans ces régions spécifiques. Les statistiques extraites pourront inclure la fréquence d'achat, les produits préférés des clients fidèles, ou encore les récompenses les plus sollicitées par ces derniers. Grâce à ces informations précieuses, la fromagerie sera en mesure d'ajuster ses programmes de fidélisation pour mieux répondre aux attentes de ses clients fidèles, renforçant ainsi leur engagement envers la marque et favorisant une relation durable et mutuellement bénéfique.

Pour le LOT 2, l'analyse est étendue à la période de 2011 à 2016 et aux départements 22, 49 et 53. La fromagerie cherche à identifier les tendances de fidélisation spécifiques à ces régions en filtrant les données selon ces critères. Ensuite, elle sélectionnera de manière aléatoire 5% des 100 meilleures commandes par ville, en excluant celles sans "timbrecli" (le “timbrecli” non renseigné ou à 0), et en calculant la moyenne des quantités de chaque commande. Le résultat sera exporté dans un fichier Excel, accompagné d'un graphe (PIE) par ville pour visualiser la répartition des commandes sélectionnées.

Pour le LOT 3, l'objectif est de mettre en place une base de données NoSQL HBase pour stocker le contenu du fichier CSV extrait du data warehouse de la fromagerie. Cette base de données sera utilisée pour interroger le data warehouse à l'aide de scripts Python et extraire des informations pertinentes sur les commandes des clients. Les données extraites incluront notamment la meilleure commande de Nantes de l'année 2020, le nombre total de commandes effectuées entre 2010 et 2015 réparties par année, ainsi que les informations sur le client ayant eu le plus de frais de timbrecde.

Pour le LOT 4, il est prévu de mettre en œuvre un moteur de recherche avec Power BI pour interroger le data warehouse HBase mis en place dans le cadre du LOT 3. Ce moteur de recherche permettra d'explorer les données de manière interactive et de générer des visualisations pour répondre aux exigences du cahier des charges. Il inclura notamment des graphiques répondant aux lots 1 et 2, ainsi que d'autres types de requêtes et de visualisations comme la géolocalisation. De plus, un tableau de bord interactif sera mis en place pour permettre aux utilisateurs d'explorer les données de manière dynamique. En mettant en œuvre ces outils d'analyse avancée, la fromagerie pourra optimiser ses programmes de fidélisation de la clientèle et consolider sa relation avec ses clients fidèles, contribuant ainsi à sa croissance et sa pérennité dans un environnement commercial compétitif.

En conclusion, l'utilisation stratégique des données de fidélisation à travers des jobs de traitement de données offre à la fromagerie une opportunité unique d'optimiser ses programmes de fidélisation et de consolider sa relation avec ses clients fidèles. En comprenant plus profondément les besoins et les comportements d'achat de sa clientèle fidèle dans divers contextes régionaux, l'entreprise peut adapter ses stratégies pour offrir des expériences personnalisées et pertinentes à ses clients. Cette approche proactive devrait se traduire par une fidélité accrue, une augmentation des ventes et une croissance durable pour la fromagerie

**Données qualifiées**

Nous avons procédé à une qualification minutieuse des donnés car Il est crucial de qualifier les données avant de les utiliser dans toute analyse ou prise de décision stratégique. La qualification des données garantit leur fiabilité, leur pertinence, leur complétude et leur cohérence, tout en assurant la conformité réglementaire et éthique. Cette étape permet d'identifier et de corriger les erreurs, les incohérences et les duplications, assurant ainsi des résultats précis et exploitables.

**codcli** : Le code client, une référence unique attribuée à chaque client dans un système ou une base de données.

**genrecli** : Le genre du client, indiquant s'il est de sexe masculin, féminin ou autre.

**nomcli** : Le nom du client.

**prenomcli** : Le prénom du client.

**cpcli** : Le code postal du client, une série de chiffres utilisée pour identifier une zone géographique spécifique dans un pays.

**ville** : Le nom de la ville où réside le client.

**codecde** : Le code de la commande, une référence unique attribuée à chaque commande passée par un client.

**date\_str** : La date de la commande sous forme de chaîne de caractères, indiquant quand la commande a été passée.

**timbrecli** : Le timbre du client, une indication de l'heure à laquelle la commande a été passée par le client.

**timbrecde** : Le timbre de la commande, une indication de l'heure à laquelle la commande a été enregistrée dans le système.

**Nbcolis** : Le nombre de colis inclus dans la commande.

**cheqcli** : La référence du chèque utilisé pour le paiement, le cas échéant.

**barchive** : Un indicateur booléen indiquant si la commande a été archivée.

**bstock** : Un indicateur booléen indiquant si l'article associé à la commande est en stock.

**codobj** : Le code de l'objet commandé.

**qte** : La quantité commandée de l'objet.

**Colis** : Le numéro de colis associé à l'objet dans la commande.

**libobj** : La description de l'objet.

**Tailleobj** : La taille de l'objet.

**Poidsobj** : Le poids de l'objet.

**points** : Le nombre de points attribués pour l'achat de cet objet, le cas échéant.

**indispobj** : Un indicateur booléen indiquant si l'objet est indisponible.

**libcondit** : Le libellé du conditionnement de l'objet, c'est-à-dire la description détaillée du mode de conditionnement de l'objet.

**prixcond** : Le prix du conditionnement de l'objet. Cette donnée indique le coût associé au conditionnement spécifique de l'objet décrit dans le champ "libcondit".

**puobj** : Le prix unitaire de l'objet.

**Procédures d’import des données**

Les mappers sont des composants clés dans le processus de traitement de données MapReduce. Leur rôle principal est de lire et de traiter les lignes d'un fichier d'entrée, puis de générer des paires clé-valeur qui seront traitées par les reducers. Voici une description plus détaillée du processus décrit :

1. Lecture et traitement des lignes CSV :

Les mappers sont chargés de lire et de traiter les lignes du fichier CSV d'entrée, dataw\_fro03.csv, une par une. Pour chaque ligne, la méthode split(",") est utilisée pour séparer les différentes valeurs de la ligne en fonction du délimiteur de champ, qui semble être la virgule dans ce cas. Cela crée une liste contenant les différentes colonnes de la ligne CSV.

1. Extraction des colonnes spécifiques et conversion :

Ensuite, les mappers accèdent aux éléments de cette liste pour extraire les colonnes spécifiques nécessaires. Par exemple, une opération de division peut être appliquée à une colonne comme "cpcli". Ces colonnes peuvent être traitées de différentes manières en fonction des besoins du traitement.

1. Filtrage des données :

Les données extraites et traitées sont ensuite soumises à des filtres basés sur des conditions spécifiques concernant la date et le département. Seules les lignes qui satisfont ces conditions sont conservées pour le traitement ultérieur.

1. Génération des paires clé-valeur :

Lorsque les conditions de filtrage sont satisfaites, les mappers génèrent des paires clé-valeur. La clé est représentée par le codecde, qui semble être un identifiant unique pour chaque enregistrement. La valeur est formatée avec des informations telles que la ville (ville), le timbre de commande (timbrecde), la quantité (qte), la date (date), et le département (departement).

1. Tri et transmission aux reducers :

Les paires clé-valeur générées par les mappers sont ensuite triées, généralement par clé, pour être transmises aux reducers. Ce tri est important car il facilite le regroupement des paires clé-valeur avec la même clé lors du traitement par les reducers. Les reducers, une fois qu'ils reçoivent ces paires clé-valeur triées, peuvent alors effectuer les opérations de traitement nécessaires, telles que l'agrégation, le calcul de statistiques, ou toute autre opération spécifiée dans le processus de traitement des données.

**Procédures de structuration :**

Nous avons réalisé une analyse approfondie des données fournies par la Fromagerie, telles que représentées dans le fichier CSV fourni.

Cette analyse nous a permis de comprendre les différentes entités, les attributs associés et les relations entre elles.

Sur la base de cette analyse, nous avons conçu un schéma de données adapté pour notre entrepôt de données.

Ce schéma inclut la création de tables pour représenter les entités pertinentes, ainsi que la définition des colonnes et des relations entre ces tables.

Nous avons également appliqué des techniques de normalisation pour réduire les redondances et les incohérences dans nos données, en séparant les données en différentes tables et en créant des clés primaires et étrangères pour maintenir l'intégrité des données.

Pour chaque attribut de nos tables, nous avons choisi des types de données appropriés, en tenant compte des valeurs stockées et des opérations à effectuer sur ces données, et nous avons défini des règles de gestion pour garantir l'intégrité et la qualité de nos données.

Enfin, nous avons documenté notre schéma de données en détaillant chaque table, chaque colonne et chaque relation, ainsi que les règles de gestion associées.

Cette documentation servira de référence pour les membres de l'équipe travaillant sur le projet.

En suivant cette approche, nous avons structuré nos données de manière à ce qu'elles soient bien organisées, cohérentes et faciles à gérer dans notre projet Big Data.

Cela facilitera l'analyse et l'exploitation des données pour répondre aux besoins spécifiques du client.

**Recommandations**

Dans le cadre de notre collaboration avec la fromagerie pour la réalisation du projet, nous formulons plusieurs recommandations clés pour assurer un déroulement fluide et réussi. Tout d'abord, nous préconisons une communication proactive et transparente entre notre équipe et la fromagerie. Il est essentiel d'établir un canal de communication régulier, favorisant ainsi l'échange d'informations pertinentes, la clarification des exigences et la résolution rapide des problèmes éventuels. Des réunions régulières seront organisées pour discuter des avancées du projet, des éventuels défis rencontrés et des décisions prises, permettant ainsi de maintenir un alignement constant sur les objectifs du projet.

En ce qui concerne la gestion des ressources, nous recommandons une planification minutieuse et une allocation judicieuse des compétences et des capacités nécessaires à chaque phase du projet. Cela permettra d'optimiser l'utilisation des ressources disponibles et d'assurer une progression efficace du projet dans les délais convenus.

Par ailleurs, nous insistons sur l'importance du contrôle qualité tout au long du processus de développement :

* Des processus et des procédures de contrôle qualité rigoureux devraient être mis en place pour garantir l'exactitude et la fiabilité des données produites par les jobs de traitement.
* Des tests systématiques et des validations croisées pourront être effectués pour détecter et corriger rapidement toute anomalie ou incohérence dans les résultats produits.

Enfin, nous recommandons un suivi post-implémentation attentif pour garantir la satisfaction continue de la fromagerie avec les solutions mises en œuvre. Nous resterons disponibles pour fournir un support continu, répondre à toutes les questions ou préoccupations et identifier les opportunités d'amélioration possible. En adoptant ces recommandations, nous sommes convaincus de notre capacité à garantir le succès du projet pour la fromagerie et à renforcer notre partenariat à long terme.

**Conclusion**

En conclusion, l'utilisation de HBase/Hadoop pour stocker et analyser les données de l'entreprise de fromagerie offre une infrastructure robuste, évolutive et flexible pour répondre aux besoins croissants de gestion et d'analyse des données.

Cette approche permet une meilleure gestion des gros volumes de données, un traitement parallèle efficace, une tolérance aux pannes élevée, et une intégration transparente avec d'autres outils Big Data pour des analyses avancées.

En fin de compte, l'utilisation de HBase/Hadoop pour l'analyse des données offre une base solide pour optimiser les programmes de fidélisation, favoriser une croissance durable et garantir le succès du projet grâce à une compréhension approfondie des clients et à des décisions stratégiques basées sur des données précises et fiables dans Power BI.

**MERCI**