



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**Université des Sciences et de la Technologie Houari Boumediene**

**Faculté d'Informatique**

**Département d'Intelligence Artificielle et Sciences des données  
(DIASD)**

**Mémoire de Licence**

**Filière: Informatique**

**Spécialité: ACAD**

## **Thème**

---

# **VISUALISATION ET ANALYSE DE DONNÉES NPS EN UTILISANT LES TECHNIQUES DE BIGDATA**

---

**Sujet Proposé par :**

**M. BENTALEB ABDELFTTAH**

**Encadré par :**

**M. BENTALEB ABDELFTTAH**

**Dr. TAMEN ZAHIA**

**Soutenu le: 2023-06-05**

**Présenté par : SAHARAOUI HANANE**

**SAADOUNE YASMINE**

**Devant le jury composé de :**

**M(e) HEDJAZI BADIAA Présidente**

**M(e) BELHADI HIBA Membre**

# Dédicace

À mon cher papa, Abd El Kader, ma chère maman Bahia pour leur amour inconditionnel, leur soutien indéfectible et leurs encouragements constants tout au long de mon parcours académique. Votre présence et votre confiance ont été une source d'inspiration et de motivation pour moi.

À ma tante Fatma Zohra et mes sœurs Amira, Meriem et Zina, qui ont toujours été là pour moi, me soutenant et me poussant à donner le meilleur de moi-même. Votre soutien inébranlable et votre amitié précieuse ont fait de ce parcours un voyage mémorable.

À mon binôme Yasmine, avec qui j'ai partagé des moments de travail acharné, de collaboration et d'apprentissage. Votre dévouement, votre engagement et votre esprit d'équipe ont été essentiels dans la réussite de notre projet.

Et enfin, à ma meilleure amie MOKEDDEM Imène, qui a toujours été là pour moi, m'apportant son soutien inconditionnel, sa positivité et son amitié sincère. Tu as été mon roc et ma source d'inspiration, et je suis reconnaissante de t'avoir à mes côtés.

Je dédie ce projet à toutes ces personnes exceptionnelles, qui ont joué un rôle essentiel dans ma vie et m'ont soutenue tout au long de ce parcours. Votre encouragement indéfectible restera gravé dans mon cœur.

*"SAHARAOUI Hanane"*

# Dédicace

À mes chers parents, qui ont été ma source de force et d'inspiration tout au long de ce projet. Votre amour, votre soutien constant et vos encouragements infatigables ont été les piliers de ma réussite. Je suis profondément reconnaissante envers-vous pour tous vos efforts louables afin de m'encourager et de me permettre de prospérer jusqu'à ce stade.

À mes frères Amine et Youcef et ma sœur Sara, leur camaraderie et leur inspiration. Vous êtes mes complices de toujours, et grâce à vous, j'ai pu surmonter tous les défis qui se sont présentés sur mon chemin.

À mes cousins et aux membres de ma famille, qui ont toujours cru en moi et m'ont soutenu(e) de près ou de loin. Votre soutien moral a été une source de motivation constante. Je suis reconnaissante de pouvoir compter sur vous.

À mon binôme Hanane, pour notre collaboration harmonieuse, notre travail d'équipe et notre persévérance. Ton dévouement, ton professionnalisme et ton amitié ont été un moteur essentiel dans la réalisation de notre projet.

Et à mes chers amis de la section C, Maya, Abdelhak, Amina, Riad, Ikram, Aya, Fella, Maroua, Scintia et Asma, qui ont partagé avec moi tant de rires, de moments de joie et de défis. Votre amitié précieuse, votre soutien mutuel et votre esprit de camaraderie ont rendu cette expérience encore plus enrichissante et mémorable.

Je dédie ce projet à toutes ces personnes merveilleuses, qui ont contribué à ma croissance personnelle et professionnelle. Votre présence, vos encouragements et votre amitié ont fait toute la différence. Je suis honorée de vous avoir dans ma vie.

*"SAADOUNE Yasmine"*

# Remerciement

Nous rendons grâce à Dieu, le miséricordieux, de nous avoir donnée la force, la volonté et la patience pour pouvoir accomplir ce modeste travail.

Nous tenons à exprimer notre profonde gratitude envers Monsieur **BENTALEB Abdelfettah** pour sa guidance et ses conseils précieux, ainsi que pour son accompagnement et son encadrement tout au long de notre stage chez Djezzy. Sa disponibilité et sa patience ont été d'une valeur inestimable pour nous.

Nous souhaitons également remercier chaleureusement Madame **TAMEN Zahia** pour ses commentaires éclairés, ses orientations académiques, sa gentillesse et son soutien continu. Ses efforts ont été essentiels dans la réussite de notre projet. Nous avons grandement bénéficié de sa présence et de ses conseils.

Nous tenons également à exprimer un grand merci à Monsieur **AMOR Azedine** pour nous avoir obtenu ce stage, et nous remercions également Mademoiselle **BOUNOUH Nesrine** et toute l'équipe de Djezzy pour leur aide et leur encouragement.

Nous remercions les membres du jury, pour l'intérêt qu'ils portent à notre travail en acceptant de l'examiner.

Enfin, Nous adressons notre reconnaissance à nos proches respectifs, pour leur soutien moral et pour leur aide précieuse.

# Table des matières

<b>Table des figures .....</b>	<b>8</b>
<b>Les abréviations .....</b>	<b>9</b>
<b>Introduction générale .....</b>	<b>10</b>
<b>1 Etat de l'art .....</b>	<b>11</b>
I. Introduction .....	11
II. Net Promotor score .....	11
1. Définition .....	11
2. Marketing ciblé .....	11
2.1. Définition .....	11
2.2. Les méthodes de collecte de données dans le marché .....	12
3. L'utilité du NPS dans le marketing ciblé .....	12
4. Business Intelligence (BI) .....	12
III. Extract-Transform-Load (ETL) .....	13
1. Définition.....	13
2. Les bénéfices d'ETL .....	13
IV. Big Data .....	14
1. Historique .....	14
2. Caractéristiques du Big Data .....	14
3. Initiation au Big data .....	15
4. L'état actuel du Big Data et son potentiel pour l'avenir .....	15
5. L'importance du Big Data dans le monde des affaires .....	16
6. Exemples d'application du Big Data dans les entreprises .....	16
V. Conclusion .....	16
<b>2 Étude de l'existence .....</b>	<b>17</b>
I. Introduction .....	17
II. Présentation de l'entreprise .....	17
1. A propos de Djezzy .....	17
2. Optimum Telecom Algérie.....	18
3. Les buts de l'Enterprise .....	18
4. Organigramme.....	18
5. Les départements .....	19

5.1. Les départements spéciaux.....	19
5.2. Les départements principaux.....	19
5.3. Service d'accueil (Big Data) .....	19
III. L'application de NPS chez Djezzy .....	19
IV. Conclusion .....	21
<b>3 Conception .....</b>	<b>22</b>
I. Introduction .....	22
II. Intégration des données .....	22
1. Présentation des données .....	22
1.1. Les fichiers .....	22
2. Les étapes de l'intégration .....	23
2.1. La création des tables .....	23
2.2. Le traitement et l'importation des données .....	23
2.2.1. Apache NiFi .....	23
2.2.2. Le choix d'utiliser NiFi.....	24
2.2.3. L'application de NiFi.....	24
2.3. Le traitement et l'insertion des données.....	25
III. Extraction des données .....	35
IV. Conclusion .....	37
<b>4 Réalisation, tests et analyse des résultats .....</b>	<b>38</b>
I. Introduction .....	38
II. L'environnement de développement .....	38
1. Environnement de travail .....	38
1.1. Environnement Matériel .....	38
2. Langages de programmation .....	38
3. Outils et logiciels de développement .....	39
4. Librairies et bibliothèques .....	40
III. Visualisation des données.....	40
1. Définition de la visualisation.....	40
2. Définition des tableaux de bords .....	40
IV. Présentation de l'application .....	41
1. Page d'authentification.....	41
2. Page d'inscription.....	41
3. Page d'accueil .....	42
4. Page de profil d'utilisateur .....	42
5. Page des tableaux de bords.....	43

V. Analyse des statiques et résultat .....	44
VI. Conclusion.....	45
<b>Conclusion Générale.....</b>	<b>46</b>
<b>Références Bibliographique .....</b>	<b>48</b>
<b>Annexe .....</b>	<b>50</b>

# Table des figures

Figure 1.1	Indicateur de fidélité .....	11
Figure 1.2	Le processus d'ETL .....	13
Figure 1.3	Les 3Vs du Big Data .....	14
Figure 1.4	Taille annuelle de la sphère mondiale des données .....	15
Figure 2.1	L'organigramme de l'entreprise OTA .....	18
Figure 2.2	Schéma explicatif de processeur "La solution NPS" .....	21
Figure 3.1	Configuration du processeur GetFile.....	25
Figure 3.2	Configuration du processeur ListFile.....	26
Figure 3.3	Configuration du processeur FetchFile.....	27
Figure 3.4	Gestion des connexions à une base de données via le Service de mise en pool .....	27
Figure 3.5	Configuration du processeur PutDatabaseRecord.....	28
Figure 3.6	Configuration du processeur ReplaceText.....	29
Figure 3.7	Schéma du groupe de processeur de remplacement .....	30
Figure 3.8	Schéma de flux de données pour le traitement du fichier question .....	31
Figure 3.9	Schéma de flux de données pour le traitement du fichier responses .....	32
Figure 3.10	Configuration du Processeur QueryRecord.....	33
Figure 3.11	Configuration du Processeur ConvertRecord.....	34
Figure 3.12	Schéma de flux de données pour le traitement du fichier NPS base .....	35
Figure 4.1	La page authentification .....	41
Figure 4.2	La page Inscription .....	41
Figure 4.3	La page d'accueil (en tant qu'administrateur et en tant qu'utilisateur) .....	42
Figure 4.4	Les tableaux d'actualité .....	42
Figure 4.5	La page de profile .....	42
Figure 4.6	Tableau des nombre de réponses globales .....	43
Figure 4.7	Exemple de graphe linéaire filtré .....	43
Figure 4.8	Exemple de graphe à barres filtré.....	44
Figure 4.9	Exemple de graphe en secteurs .....	44
Figure A.1	Création d'une nouvelle connexion .....	50
Figure A.2	La sélection d'une base de données .....	50
Figure A.3	La configuration de connexion .....	50
Figure A.4	Sélection de schéma de fichier .jar .....	51
Figure A.5	Un exemple de requête de création pour l'une des tables .....	51



# Abréviations

**2G** désignent la 2ème Génération de réseaux mobiles.

**3G** désignent la 3ème Génération de réseaux mobiles.

**4G** désignent la 4ème Génération de réseaux mobiles.

**SMS** Short Message Service

**SMSC** Short Message Service Center

**SMPP** Short Message Peer-to-Peer

**GSM** Global System for Mobile Communication

**VIP** Very Important Person

**IT** Information Technology

**IA** Intelligence Artificielle

**NPS** Net Promoter Score

**ETL** Extract Transform Load

**ADSL** Asymmetric Digital Subscriber Line

**SQL** Structured query language

**SFTP** Secure File Transfer Protocol

**JDBC** Java Database Connectivity

**URL** Uniform Resource Locator

**JVM** Java Virtual Machine

**HTTP** HyperText Transfer Protocol

**IOT** Internet Of Things

# Introduction Générale

Le marché de la téléphonie en Algérie continue de connaître une croissance soutenue du nombre d'abonnés. Dans ce contexte, les enjeux et la concurrence dans le domaine des télécommunications imposent aux opérateurs téléphoniques d'exploiter efficacement les données massives dont ils disposent. En optimisant les coûts et en maximisant les profits, ces opérateurs peuvent répondre aux exigences du marché tout en assurant une meilleure expérience utilisateur. Djezzy, en tant qu'opérateur de référence en Algérie pour l'innovation, s'investit en mettant l'accent sur ces aspects qui jouent désormais un rôle clé dans la satisfaction des clients.

Djezzy a mis en place une solution qui lui permet de mieux comprendre son public cible. Cette solution utilise le NPS (Net Promoter Score), un indicateur de croissance qui aide à étudier plus en détail l'expérience utilisateur. En recueillant des données sur le NPS, Djezzy peut prendre des décisions éclairées pour améliorer ses services et garantir la satisfaction de ses clients. Cette approche favorise une meilleure compréhension des besoins et des attentes des utilisateurs, permettant ainsi à l'entreprise de mieux répondre à leurs exigences.

Notre projet vise à concevoir des workflows qui facilitent l'extraction, le traitement et l'insertion des données du NPS dans une base de données, dans le but ultime de les représenter sous forme d'une interface web. Cette approche s'appuie sur les techniques de Big data, permettant ainsi de manipuler de grandes quantités de données de manière efficace et d'exploiter pleinement leur potentiel analytique. L'interface web résultante offrira une visualisation claire et conviviale des informations extraites, favorisant ainsi une compréhension approfondie des évaluations du NPS et facilitant la prise de décisions informées pour améliorer l'expérience utilisateur.

Notre document sera divisé en quatre chapitres, présentés selon le plan suivant :

- Chapitre 1 : Ce chapitre présente les concepts du Big Data, en explorant son histoire et son importance dans le monde des affaires. Nous abordons également la notion de marketing ciblé et la satisfaction des clients, mettant en évidence l'utilisation du Net Promoter Score dans cette approche.
- Chapitre 2 : Ce chapitre a pour objectif de fournir des informations relatives à l'entreprise où notre projet a été réalisé, et aborder l'utilisation du Net Promoter Score (NPS) en tant qu'outil stratégique au sein de cette organisation.
- Chapitre 3 : Nous allons présenter dans ce chapitre les différentes étapes de la conception de notre projet à savoir l'intégration et l'extraction des données.
- Chapitre 4 : Ce chapitre sera consacré à la partie réalisation de notre application. Nous commencerons par l'environnement de développement ainsi que les différentes bibliothèques utilisées. Nous détaillerons ensuite les différentes interfaces liées à l'implémentation de notre site web, puis nous finirons ce chapitre par donner les résultats obtenus ainsi que l'analyse des différents tableaux de bord.

Une conclusion générale terminera notre document incluant quelques perspectives à ce travail.

# Chapitre 1

## État de l'art

### I. Introduction

Dans ce chapitre nous allons présenter des généralités sur le Bigdata, le marketing ciblé et le Net Promoter Score (NPS). Comprendre le bonheur des clients, le domaine en développement de l'analyse des données repose fortement sur ces idées. Nous examinerons leurs définitions, leurs utilisations et leurs avantages.

### II. Net Promoter Score

#### 1. Définition

Le Net Promoter Score est l'indicateur qui mesure la fidélisation et la satisfaction client, développé par Fred Reichheld via le cabinet de consulting Bain & Company en 2003. Il évalue la qualité de la relation entre une marque et ses clients, en d'autres termes, il permet d'estimer la probabilité qu'une marque, qu'un produit ou qu'un service soit recommandé par ses clients.[1]

On le calcule en posant des questions aux clients. Ils donnent une note entre 0 et 10 (voir figure 1.1) qui permettent de les regrouper en trois catégories :

- Détracteurs : les clients dont la note est entre 0 et 6.
- Passifs : les clients avec une note entre 7 et 8.
- Promoteurs : les clients ayant une note entre 9 et 10.

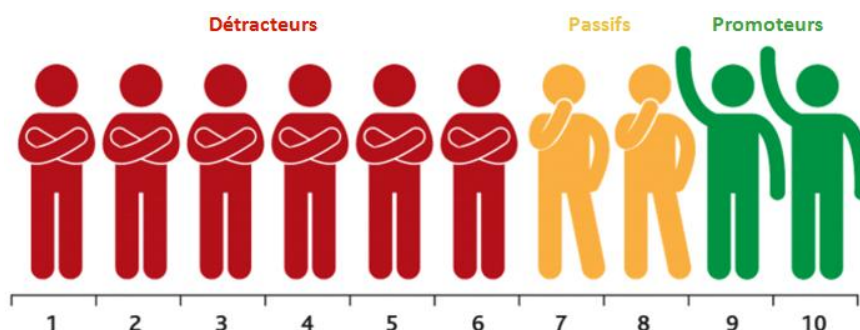


Figure 1.1 – Indicateur de fidélité

#### 2. Marketing ciblé

##### 2.1. Définition

Le marketing ciblé est conçu pour transmettre le bon message aux bonnes personnes au bon moment, ce qui le rend très efficace. Il permet de se démarquer dans un marché surchargé en ciblant les clients potentiels qui ont déjà manifesté leur intérêt pour ce que vous offrez.

Il est bien plus efficace et rentable que les techniques de marketing de masse traditionnelles. Au lieu de dépenser de l'argent pour faire de la publicité à tout le monde, on peut concentrer les ressources sur les clients qui ont déjà manifesté leur intérêt pour nos produits ou services. Cela nous permet de maximiser le retour sur investissement tout en nous assurant que notre message atteint un public susceptible de s'y intéresser.

## **2.2. Les méthodes de collecte de données dans le marché**

Pour créer un plan marketing efficace, il est préférable d'utiliser une variété de méthodes pour recueillir les données, notamment :

- **L'analyse**

Les outils d'analyse fournissent des informations précieuses sur le comportement des clients. Regarder les tendances du trafic sur le site web, les types de clients qui interagissent avec la présence de notre marque sur les réseaux sociaux, et d'autres tendances comportementales numériques pour mieux comprendre les préférences et les intérêts de notre public cible.

- **La recherche**

La recherche nous aide à comprendre le contexte plus large de notre marché cible - des choses comme les conditions économiques, les tendances de l'industrie et les paysages concurrentiels. Toutes ces informations peuvent être utilisées pour élaborer une stratégie marketing efficace pour cibler des publics spécifiques.

- **Les enquêtes**

Les enquêtes sont un excellent moyen d'en savoir plus sur les attentes du marché cible et sur la manière dont il réagit à différentes techniques de marketing. Poser des questions qui nous donneront des informations utiles sur leurs préférences et comportements, telles que les canaux publicitaires qu'ils préfèrent, leurs habitudes de dépenses en ligne, etc. Et parmi les enquêtes les plus utiles, il y a le NPS.

## **3. L'utilité du NPS dans le marketing ciblé**

Le NPS permet de cibler les clients les plus satisfaits de l'entreprise, qui sont plus susceptibles de devenir des promoteurs de la marque. Cela contribue à accroître la notoriété de la marque et à attirer de nouveaux clients, tout en identifiant les domaines à améliorer en matière de service client. En recueillant les commentaires des clients, les entreprises peuvent identifier les problèmes et les lacunes dans leur service client, et prendre des mesures pour les résoudre, améliorant ainsi la satisfaction des clients et augmentant le NPS à long terme.

## **4. Business Intelligence (BI)**

La Business Intelligence désigne le processus de collecte, d'analyse et d'interprétation de données pour inférer les décisions commerciales. Cela implique l'utilisation de divers outils et techniques pour collecter des données provenant de différentes sources, notamment des commentaires de clients, des données de vente et des tendances du marché.[2]

Ce processus est crucial dans le marketing ciblé car il fournit des informations précieuses sur le comportement et les préférences des clients, ce qui aide les entreprises à créer des campagnes de marketing plus efficaces. Il existe différentes technologies de BI, notamment le data Warehousing<sup>1</sup>, le Online Analytical Processing(OLAP) et le data mining<sup>2</sup>, ainsi que le Big data.

---

<sup>1</sup>Collecte et organisation de données pour analyse et rapports.

<sup>2</sup>Découverte de modèles et relations dans de grandes quantités de données, utilisant l'apprentissage automatique et l'analyse statistique.

### III. Extract-Transform-Load (ETL)

Il est courant que dans la plupart des entreprises, les données qui pourraient être potentiellement utiles restent inaccessibles. Selon une étude, environ deux tiers des entreprises ne tirent que peu ou pas d'avantages concrets de leurs données, et parfois même aucun. Les données sont souvent enfermées dans des silos cloisonnés, des systèmes obsolètes ou des applications peu utilisées.

Pour remédier à cela, le processus ETL permet de collecter des données à partir de sources multiples, de les nettoyer et de les transformer, afin de les rendre disponibles pour des analyses métier ultérieures.

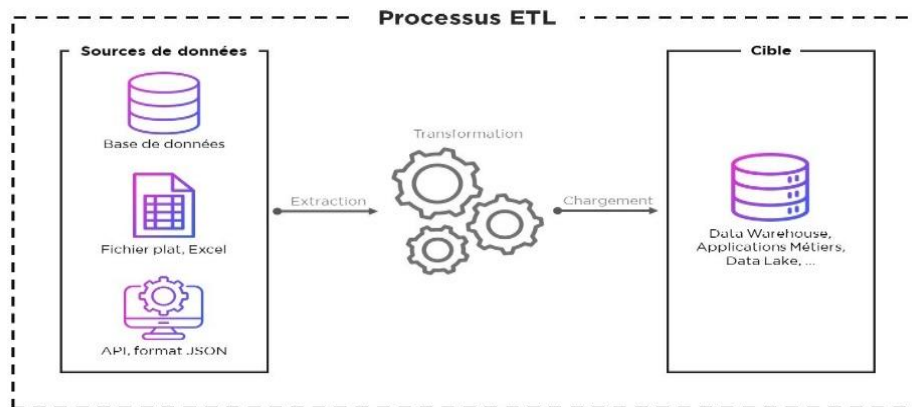


Figure 1.2 – Le processus d'ETL

#### 1. Définition

Il s'agit d'une méthode de collecte de données à partir de différentes sources, de leur transformation en un format adapté à une analyse ultérieure et de leur insertion dans une base de données ou un entrepôt de données pour un accès ultérieur.

C'est un processus qui permet aux entreprises de collecter rapidement et avec précision des données à partir de sources multiples et de les centraliser en un seul endroit. Par exemple, le processus ETL peut rassembler des données-clients stockées dans une base de données Sales et des informations de commande stockées dans une base de données, dans un référentiel centralisé pour une analyse ultérieure. De cette façon, nous pouvons nous assurer que les données sont précises, à jour et facilement accessibles lorsque l'on en a besoin.[3]

#### 2. Les bénéfices d'ETL

Ce processus est très important lorsqu'il s'agit de transformer des données brutes en informations utilisables. Il peut rapidement collecter des données à partir de sources multiples et de les normaliser dans un système plus automatisé, offrant toutes sortes d'avantages à ceux qui les utilisent, comme :

- **Augmentation de l'efficacité :** En automatisant la transformation des données, les erreurs manuelles et la duplication peuvent être considérablement réduites. Cela améliore l'efficacité en termes de temps et de coûts.
- **Amélioration de la qualité des données :** Étant donné que les processus ETL standardisent et valident les données, cela améliore la qualité du produit final.

- Analyse rationalisée : Les données normalisées facilitent l'analyse des tendances et l'extraction automatique des informations commerciales par les équipes.

#### IV. Big Data

##### 1. Historique

Avec la prolifération d'Internet, les données générées ont connu une croissance exponentielle. Cette nouvelle réalité numérique a entraîné un ajustement nécessaire des outils de traitement et d'exploitation des données volumineuses pour répondre à ces besoins croissants.

##### 2. Caractéristiques du Big Data

Comme tout objet, le Big data peut avoir différentes caractéristiques qui affectent sa qualité, qu'il peut varier selon de nombreuses dimensions. En 2001, Doug Laney qui est ancien scientifique en chef à Silicon Graphics<sup>3</sup> a détaillé que les Big Data étaient caractérisées par trois traits [4] :

- **Volume** : La quantité de données générées par les utilisateurs peut être considérée comme Big Data si elle dépasse notre capacité de stockage actuelle, par exemple, en 2016 la capacité d'un disque dur est arrivée jusqu'à 4 Téraoctet qui sont 4000 fois plus grand qu'un disque dur en 1990[5]
- **Variété** : Ce terme englobe tous les types de données, structurés ou non structurés, et les méthodes traditionnelles d'analyse sont incapables de gérer cette variété. [5]
  - **Données structurées** : Il s'agit de données organisées sous forme de tableaux appelés tables de données. Elles peuvent être facilement traitées, consultées et modifiées. Exemples : les noms, les prénoms, etc.
  - **Données non structurées** : Il s'agit de données dont la forme ou la structure est inconnue, exemples : les images, les vidéos, les fichiers audio, etc.
- **Vélocité** : Ce terme fait référence à la vitesse à laquelle les données sont générées, capturées et partagées, tel que le volume est une conséquence d'incompatibilité entre la vitesse d'arrivée et la vitesse de traitement des données avec le temps. [5]

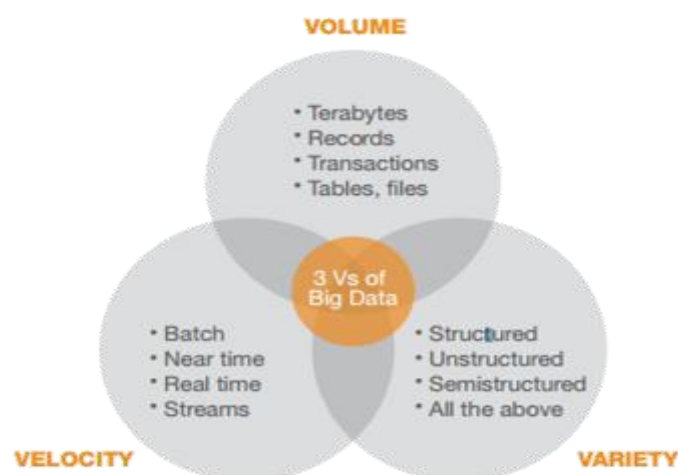


Figure 1.3 – Les 3Vs du Big Data

<sup>3</sup>Une ancienne entreprise américaine spécialisée dans les solutions informatiques et graphiques.

### 3. Initiation au Big data

Avec l'évolution technologique, la quantité de données a rapidement augmenté dans tous les domaines, ce qui a donné naissance au concept de Big Data. Les progrès technologiques et l'augmentation de la production et de l'utilisation des données ont rendu difficile le traitement en temps réel de ces volumes massifs d'informations à l'aide des méthodes traditionnelles.

En réponse à ce défi, les informaticiens ont cherché de nouvelles solutions pour traiter, analyser et extraire les informations pertinentes. Ainsi, le terme "Big Data" est utilisé pour décrire des données de taille supérieure au petabyte. Les entreprises se tournent vers le Big Data pour obtenir un avantage concurrentiel en prenant des décisions basées sur les données, en améliorant leurs produits et services, et en augmentant la satisfaction des clients. Ce besoin croissant a stimulé l'émergence de nouvelles technologies adaptées à la gestion du Big Data.

### 4. L'état actuel du Big Data et son potentiel pour l'avenir

Les objets connectés à l'Internet jouent un rôle essentiel dans un futur incertain où les Big Data sont omniprésentes. Les données accumulées grâce à ces objets connectés permettent d'entraîner des systèmes d'intelligence artificielle qui structurent des enjeux planétaires. La majorité des experts du Big Data s'accordent à dire que la quantité de données générées augmentera de façon exponentielle à l'avenir. Selon le rapport "Data Age 2025" de Seagate réalisé par IDC (International Data Corporation), la sphère de données mondiale devrait atteindre 175 ZéttaOctets d'ici 2025[6].

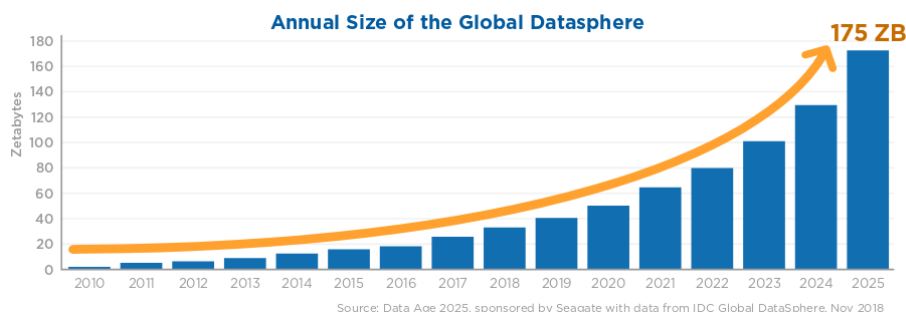


Figure 1.4 – Taille annuelle de la sphère mondiale des données.

La croissance exponentielle des données générées s'accélère avec la popularité d'Internet. En 2000, les chercheurs de l'Université de Californie à Berkeley ont estimé, dans la revue "How Much Information?", que les données numériques générées par l'humanité équivalaient à 11,5 exabits, ce qui était déjà considérable. Depuis 2010, le volume des données générées a explosé avec l'avènement de l'Internet des objets, l'accessibilité des appareils mobiles et la popularité des réseaux sociaux tels que Facebook<sup>4</sup>, Instagram<sup>5</sup>, etc. Cette réalité numérique en constante évolution nécessite donc une adaptation des outils de traitement et d'exploitation des données volumineuses.[7]

<sup>4</sup>Réseau social en ligne, il permet aux utilisateurs de créer des profils, de se connecter avec des amis, de partager du contenu.

## 5. L'importance du Big Data dans le monde des affaires

Le Big Data permet aux entreprises de révolutionner leur fonctionnement. En exploitant de grandes quantités de données, les entreprises peuvent obtenir des informations sur leurs activités et prendre de meilleures décisions qui conduisent à une efficacité accrue et à un meilleur service client. Avec l'analyse de données volumineuses, les entreprises peuvent identifier les motifs et les patterns de comportement des clients, prédire les évolutions futures, affiner leurs stratégies de marketing et optimiser les modèles de tarification.

Les sociétés peuvent économiser de l'argent grâce au Big Data en optimisant leurs opérations, en améliorant la chaîne d'approvisionnement, et en identifiant des domaines pour réduire les coûts et il est devenu possible aussi d'anticiper les risques potentiels pour prévenir les coûts élevés ou les dommages à leur réputation.

## 6. Exemples d'application du Big Data dans les entreprises

- **Améliorer le service client**

L'analyse fournie par Le Big data est un outil puissant pour comprendre les clients, leurs besoins et leurs préférences. Les entreprises peuvent tirer parti des informations obtenues grâce à cette analyse pour développer de meilleures stratégies de service client et offrir de meilleures expériences à leur public cible.

- **Optimiser les opérations**

En analysant de grands ensembles de données provenant de différentes sources (telles que les enquêtes de satisfaction client, les rapports opérationnels, les enregistrements de ventes, etc.), les entreprises peuvent repérer rapidement les tendances, acquérir de nouvelles perspectives et optimiser leur fonctionnement.

- **Automatiser les décisions**

Le Big data peut également aider les entreprises à automatiser les processus de prise de décision en intégrant des capacités d'IA dans les solutions d'analyse qui analysent de grandes quantités de données structurées et non structurées. Cela permet aux entreprises de prendre des décisions plus rapidement, plus précisément et plus régulièrement, ce qui se traduit par une amélioration de l'efficacité et des économies de coûts.

## V. Conclusion

Dans ce chapitre, nous avons exploré les concepts clés qui sous-tendent notre projet. Nous avons découvert le Net Promoter Score (NPS) et sa capacité à mesurer la fidélité et la satisfaction des clients. Nous avons également découvert l'importance d'un marketing ciblé pour atteindre des segments de clientèle spécifiques. En outre, nous avons discuté du rôle de l'informatique décisionnelle dans la compréhension du comportement et des préférences des clients, ainsi que de l'importance du Big Data dans l'environnement numérique d'aujourd'hui.

Dans le chapitre suivant, nous allons explorer la société Djazzy, mettre en évidence son architecture interne, et étudier son besoin pour l'utilisation du NPS.

---

<sup>5</sup>Plateforme de partage de photos et vidéos, Elle permet aux utilisateurs de prendre des photos, d'appliquer des filtres, de les partager sur leur profil et de les suivre par d'autres utilisateurs



# Chapitre 2

## Étude de l'existence

### I. Introduction

Notre stage s'est déroulé chez Djazzy, une entreprise pionnière dans le domaine des télécommunications en Algérie. Dans ce chapitre, nous allons présenter cette entreprise, mettre en évidence ses différents services, expliquer son architecture interne, et aborder l'utilisation du Net Promoter Score (NPS) en tant qu'outil stratégique au sein de la société.

### II. Présentation de l'entreprise

#### 1. A propos de Djazzy

Djazzy, un opérateur de télécommunications algérien, a été fondé en juillet 2001. Il est considéré comme un leader dans le domaine de la téléphonie mobile, avec plus de 14 millions d'abonnés en août 2022. L'entreprise propose une large gamme de services, notamment des forfaits prépayés et post-payés, des services de données, des services à valeur ajoutée.

Depuis juillet 2022, Djazzy est devenue une entreprise nationale après que Veon<sup>6</sup> ait signé l'acte de cession de la totalité de ses actions dans l'entreprise au profit du Fonds National d'Investissement. Ainsi, la propriété de Djazzy est partagée entre le FNI<sup>7</sup> qui possède 96,57% des parts et CEVITAL<sup>8</sup> qui en détient 3,43%. [8]

Elle fournit des services de télécommunications dans tout le pays et couvre 95 % de la population. Ses services 3G sont disponibles dans les 58 wilayas et elle a commencé à offrir la 4G dans 20 wilayas le 1<sup>er</sup> octobre 2016, avec l'objectif de couvrir tout le territoire national. En ce qui concerne les licences, Djazzy a obtenu la licence 2G le 30 juillet 2001, la licence 3G le 2 décembre 2013 et la licence 4G le 4 septembre 2016. [8]

Cette entreprise a remporté plusieurs prix pour la qualité de son service à la clientèle, y compris le prix du meilleur service client de l'année en 2020 en Algérie. Voici une liste de quelques-uns des services offerts par cette société [8] :

- Services de téléphonie mobile : Djazzy propose des forfaits de téléphonie mobile prépayés et post payés avec des appels illimités, des SMS et des données mobiles.
- Services Internet : Djazzy propose également des offres d'accès à internet haut débit pour les particuliers et les entreprises, avec des offres ADSL et des offres 4G.
- Services pour les entreprises : Djazzy propose des solutions de communication pour les entreprises, telles que la communication par SMS en masse, l'hébergement de sites web et la mise en place de systèmes de communication interne.

---

<sup>6</sup>C'est une entreprise néerlandaise de télécommunication, un des plus importants opérateurs de réseau mobile au monde, dont le premier marché est la Russie.

<sup>7</sup>(Fonds National d'Investissement) : anciennement Banque Algérienne de Développement

<sup>8</sup>C'est un conglomérat algérien de l'industrie agroalimentaire, la grande distribution, l'industrie et les services.

## 2. Optimum Telecom Algérie

Optimum Telecom Algérie (OTA), propriétaire de la marque commerciale « Djezzy », situé à **Dar El Beida** est autorisée par le gouvernement Algérien à exploiter les services de téléphonie mobile de type GSM, 3G et 4G sur le territoire Algérien. Conformément aux licences octroyées à cet effet, OTA a mis en place un site web [www.djezzy.dz](http://www.djezzy.dz) afin de permettre à tout utilisateur de s'informer sur l'actualité et les évolutions concernant la société[9]

## 3. Les buts de l'Enterprise

Pour réaliser sa vision, Djezzy s'engage à :

- Déployer des infrastructures à la pointe de la technologie.
- Créer pour ses employés le meilleur environnement de travail et d'épanouissement.
- Contribuer activement au bien-être des Algériens.
- Optimiser la création de valeurs pour ses actionnaires, à travers un contrôle strict des coûts.
- Appliquer rigoureusement sa politique environnementale.
- Améliorer sans cesse ses processus internes dans le respect de sa politique-qualité.

## 4. Organigramme

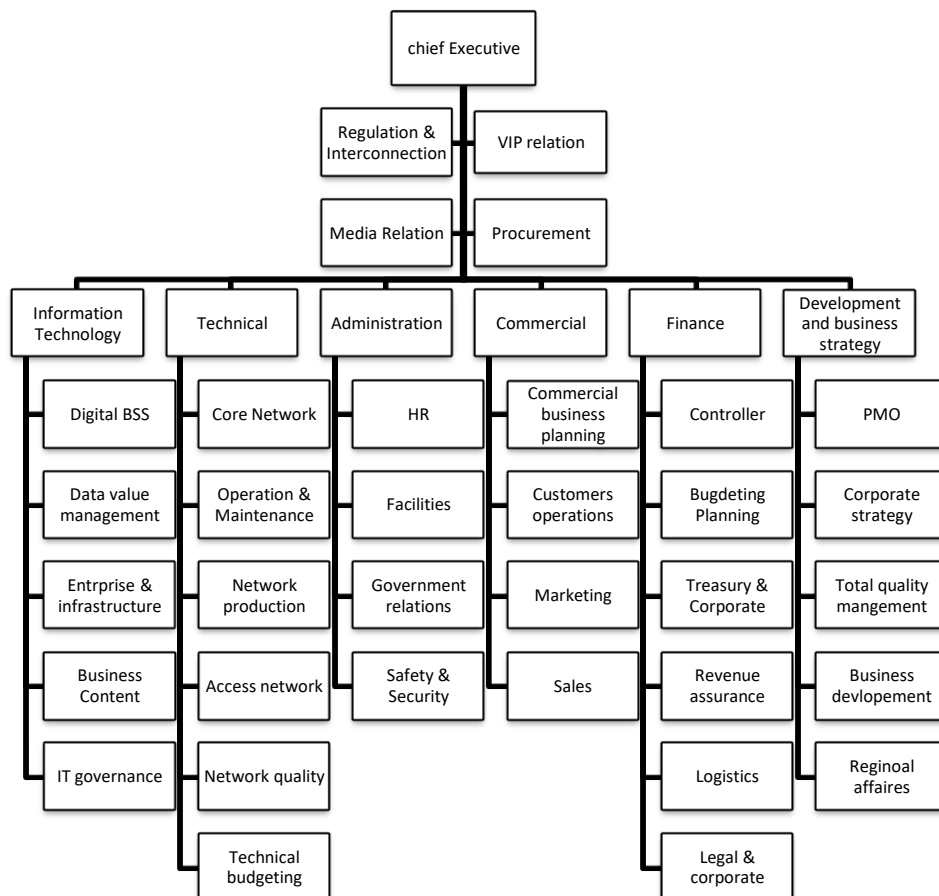


Figure 2.1 – L'organigramme de l'entreprise OTA

## 5. Les départements

### 5.1. Les départements spéciaux

- Régulations et interconnexions : Responsable de la régulation des activités télécoms et de la communication avec les régulateurs.
- Relation VIP : Gère les clients VIP et leurs services spécifiques.
- Approvisionnement : Gère les finances et les budgets de l'entreprise.
- Media relation : S'occupe des relations avec les médias et de la communication externe.

### 5.2. Les départements principaux

- Département digital : Gère la stratégie et le marketing numérique.
- Département Technique : Gère l'expansion du réseau et la maintenance technique.
- Département administration : Gère les tâches administratives et les ressources humaines.
- Département commercial : Responsable du marketing, du service client et des ventes.
- Département finances : Gère les revenus et les finances de l'entreprise.
- Département stratégie de l'entreprise : Prépare et évalue la stratégie à long terme.
- Département IT (technologies et systèmes d'informations): Développe et maintient les outils informatiques.

### 5.3. Service d'accueil (Data Management) :

Nous effectuons notre projet dans le département IT précisément au service Data management qui a comme principales missions :

- **Comprendre et cibler les abonnés** : Ce service a pour objectif d'aider Djezzy dans la compréhension de ses abonnés en collectant les données de l'entreprise, puis en les analysant pour obtenir des informations exploitables. Il est responsable de l'extraction et de l'agrégation des données pour une utilisation dans la prise de décision et la direction de l'entreprise.
- **Aider les décideurs à prendre des décisions stratégiques** : Fournir un ensemble de données et d'analyses aux décideurs afin de mettre à leur disposition toutes les informations relatives à l'entreprise, au marché, aux clients, aux télécoms et à la concurrence et ce pour les aider à prendre des décisions data-driven.
- **Optimisation des coûts** : utiliser les technologies des Big data pour optimiser les coûts et les dépenses de l'entreprise.
- **Améliorer l'expérience client** : permettant à Djezzy de se rapprocher de sa clientèle, en surveillant de près son comportement et en étant à l'écoute de ses besoins.
- **Advanced data analysis** : Rajouter de l'intelligence à la donnée afin d'automatiser certains processus et services de l'entreprise.

## III. L'application de NPS chez Djezzy

Dans le cadre des activités de Djazzy afin d'évaluer les fidélités de ses clients, celle-ci a recours à la méthodologie NPS (Net Promoter Score). Pour ce faire, un processus nommé « La solution NPS » a été mis en place où un groupe spécifique d'abonnés est sélectionné chaque jour et se voit envoyer des sondages d'opinion par SMS, en fonction de critères précis définis.

Djazzy a utilisé sept types de sondages différents, à savoir :

- **Join** (Rejoindre) : cette enquête est disponible pour les nouveaux clients une fois qu'ils commencent à utiliser l'offre. Ils reçoivent ensuite un sondage.
- **CC** (Centre calls) : Cette enquête s'adresse aux clients qui ont des problèmes et appellent le centre d'appels pour exprimer leurs réclamations. Une fois leur conversation avec l'agent terminée, ils reçoivent un sondage.
- **Product**: Ce type d'enquête est destiné aux clients qui utilisent les produits Djazzy tels que les clés ou les modems.
- **Retail**: Cette enquête est envoyée aux clients qui envisagent de mettre fin à leur relation avec Djazzy.
- **Network** (Réseau) : cette enquête recherche des clients rencontrant des problèmes de réseau et de connectivité.
- **B2B** (Business to Business) : Ce type d'enquête s'adresse aux entreprises qui concluent des contrats commerciaux avec Djazzy.
- **Djazzy App** : Cette enquête concerne les utilisateurs d'une application mobile développée par Djazzy.

#### ▪ Implémentation du processeur "La solution NPS"

La solution NPS repose sur un système modulaire qui met en œuvre une architecture événementielle utilisant un bus de messagerie appelé "Kafka" pour échanger le trafic sortant et entrant des SMS vers/depuis les abonnés.

Il utilise un module en amont appelé NIFI. Ce module publie toute l'activité SMS sur un sujet Kafka dédié. Cela permet de transmettre ces informations hors plateforme au SMSC (Short Message Service Center) pour la livraison de SMS. Dans le même temps, le module NIFI lit également les SMS entrants et déclenche le logique métier approprié.

L'intégration du SMSC dans la solution NPS joue un rôle crucial en tant qu'interface avale. Il effectue les actions suivantes :

- Il lit les SMS sortants provenant de Kafka et les envoie au SMSC via une connexion SMPP (Short Message Peer-to-Peer).
- Il reçoit les SMS entrants et les publie sur les topics de Kafka.

Ce processus permet à Djazzy de gérer efficacement les échanges de SMS avec les abonnés, assurant la bonne diffusion des sondages et la réception des réponses.

Le schéma explicatif ci-dessous illustre visuellement ce processus :

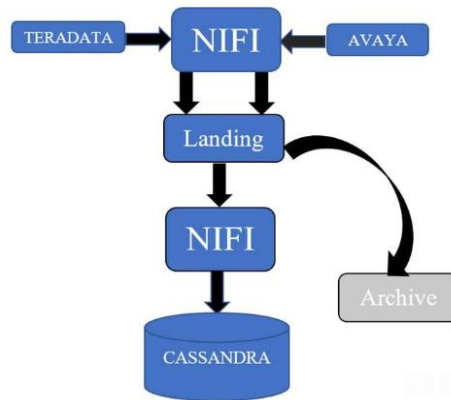


Figure 2.2 – Schéma explicatif de processeur "La solution NPS"

- ⇒ **Avaya** est une entreprise qu'elle génère et collecte des données relatives aux interactions téléphoniques, ce qui lui permet d'optimiser ses solutions et de répondre aux besoins de ses clients en matière de communication.[10]
- ⇒ **Teradata** est une plateforme d'analyse de données et un système de gestion de base de données qui est couramment utilisé par les entreprises pour collecter, stocker, gérer et analyser de grandes quantités de données.[11]
- ⇒ **Cassandra** est une base de données distribuée hautement évolutive conçue pour gérer de grandes quantités de données non structurées tout en assurant une haute disponibilité. [12]
- ⇒ **Apache Kafka** est une plateforme de diffusion d'événements distribuée en open source utilisée par de nombreuses entreprises pour des pipelines de données haute performance, l'analyse en continu, l'intégration de données et les applications critiques.[13]

#### IV. Conclusion

Dans ce chapitre, nous avons cherché à donner un aperçu de l'environnement de travail au sein de Djezzy, afin de mieux comprendre les besoins et les exigences liés à la réalisation de notre projet.

Dans le chapitre qui suit, nous aborderons la partie de conception du projet, qui se concentre sur le traitement et l'intégration des données.

# Chapitre 3

## Conception

### I. Introduction

L'objectif principal de notre travail consiste à valoriser les données dont nous disposons en les rendant opérationnelles pour l'analyse, la prise de décision et la mise en œuvre de stratégies de marketing ciblé. Pour atteindre cet objectif, deux étapes essentielles sont nécessaires. Dans ce chapitre, nous nous commencerons par la partie intégration des données et nous terminerons par l'extraction et analyse de ces dernières.

### II. Intégration des données

De manière générale, l'intégration de données consiste à rassembler des données de différentes sources, comme des bases de données, des fichiers, des réseaux sociaux, etc. On nettoie et transforme ces données pour qu'elles soient exploitables. De cette manière, tout le monde peut voir les mêmes informations et ce processus aide à prendre les bonnes décisions et améliorer les processus de travail. Dans le contexte de notre projet, les résultats des sondages revêtent une importance primordiale.

#### 1. Présentation des données

Comme mentionné précédemment, notre principal intérêt se porte sur les résultats de NPS, incluant les réponses fournies par les participants, les questions qui leur ont été posées, ainsi que les informations démographiques et comportementales des clients. Ces données sont fournies par l'entreprise sous la forme de fichiers CSV (comma-separated-values).

**1.1. Les fichiers :** Ils consistent en quatre fichiers CSV où les colonnes sont séparées par un pipe (|) :

**1.1.1. Questions :** Ce tableau contient les questions posées dans les enquêtes NPS menées par Djezzy en trois langues (français, arabe et arabe algérien). Il s'agit de taille de 80KB, Les colonnes incluent un identifiant de type d'enquête, un identifiant de langue qui indique dans quelle langue la question est posée, un numéro de question (question\_number), le texte de la question (question), le nom de la question (question\_name) et le type de question.

**1.1.2. Réponses :** Ce tableau est de taille 296MB, il contient les réponses des clients aux enquêtes NPS menées par Djezzy. Les colonnes incluent la date et l'heure de l'enquête (survey\_dttm), le numéro de téléphone du client crypté (msisdn), la date et l'heure de la réponse (response\_dttm), le numéro de la question à laquelle la réponse a été donnée (question\_number), le nom de la question (question\_name), le type de question (question\_type), la réponse brute du client (response\_raw) et la réponse analysée (response\_parsed).

**1.1.3. Type :** un tableau de taille 1Kb, qui contient deux colonnes une pour les identifiants et une autre pour le nom de sondage de la manière suivante : 1 pour

le sondage join, 2 pour contact center, 3 : Product, 4 : Djezzy App, 5 : retail, 6 : network et 8 pour B2B.

- 1.1.4. Survey** : Le tableau de données est très volumineux, dépassant 20GB, avec plus de 60 millions de lignes. Il contient des informations complètes sur les enquêtes envoyées aux clients, y compris les données des clients et les données de traitement de l'enquête. Les informations stockées incluent la date de l'enquête, le numéro de téléphone du client, le type d'enquête, le score Net Promoter Score (NPS), le statut de l'enquête, le groupe d'âge du client, la ville de résidence, la marque et le modèle du téléphone, l'identifiant du point de vente, etc. De plus, des informations sur le traitement de l'enquête, telles que le rôle de l'utilisateur, les dates et heures des réponses, sont également présentes dans le tableau.

## **2. Les étapes de l'intégration**

### **2.1. La création des tables**

Après l'installation de PostgreSQL, nous avons créé une base de données appelée "nps-Survey". Ensuite, nous avons établi une connexion entre DBeaver, un logiciel gratuit et open source utilisé pour la gestion et l'administration de bases de données, et PostgreSQL, qui prend en charge les requêtes SQL et JSON<sup>9</sup>. Pour plus de détails sur PostgreSQL et DBeaver (voir chapitre 4). Les étapes de création de cette connexion sont décrites en détail dans l'annexe 1.

L'objectif de cette procédure est d'effectuer diverses tâches liées à la base de données, telles que l'exécution de requêtes SQL, la gestion des tables et des données, l'importation et l'exportation de données, la création de vues, et l'administration des utilisateurs de manière plus facile et efficace.

Dans notre projet, nous avons utilisé DBeaver comme éditeur SQL convivial. Il nous a permis de créer facilement des tables dans notre base de données PostgreSQL. De plus, DBeaver a automatiquement converti les données du format CSV vers le format approprié pour le serveur PostgreSQL, simplifiant ainsi l'importation des données.

Une fois la connexion établie, nous avons procédé à l'exécution du script de création des tables (voir l'annexe 1 pour plus de détails).

### **2.2. Le traitement et l'importation des données**

Chaque fichier doit être importé dans la table correspondante de la base de données. Bien que cette tâche puisse être effectuée manuellement à l'aide de requêtes SQL, cela peut devenir fastidieux et prendre beaucoup de temps, surtout lorsqu'on travaille avec de grandes quantités de données. Il est donc souvent nécessaire d'utiliser des outils pour automatiser ces tâches et gérer les flux de données de manière efficace et rapide. C'est ici que les outils de gestion de flux de données, tels que Apache NiFi, entrent en jeu.

#### **2.2.1. Apache NiFi**

Apache NiFi est une plateforme open source qui automatise les flux de données entre divers systèmes, y compris les bases de données, les services Cloud et les appareils IoT. Elle

---

<sup>9</sup>(JavaScript Object Notation) est un format de données léger et facile à utiliser pour les humains et les machines. Il est largement utilisé pour l'échange de données entre différentes langues de programmation.

simplifie l'intégration de données, en particulier pour les organisations traitant de grandes quantités de données. Pour plus de détails sur Apache NiFi, veuillez-vous référer au chapitre 4 pour la définition complète.

### 2.2.2. Le choix d'utiliser NiFi

Il y a plusieurs raisons pour lesquelles NiFi peut être préféré par rapport à d'autres outils de traitement de flux de données:

- **L'interface utilisateur graphique** : l'interface de NiFi est caractérisée par sa clarté et sa simplicité d'utilisation, ce qui favorise la réalisation des opérations liées au flux et au traitement des données.
- **Facilité de déploiement et d'administration** : Cette fonctionnalité facilite le traitement des données, notamment pour les grandes quantités, pour les entreprises qui utilisent la plate-forme NiFi grâce à sa prise en charge du système distribué. Ainsi, elle rend le traitement des données plus flexible et plus aisé pour les utilisateurs.
- **Évolutivité** : NiFi est très évolutif, automobile il peut gérer des flux de données de grande envergure avec des volumes de données importants et des fréquences de mise à jour élevées.
- **Processeurs prédéfinis** : NiFi contient un grand nombre de processeurs qui répondent aux besoins de gestion de flux de données. Il existe plusieurs types de ces processeurs selon le cas d'utilisation, tels que les processeurs d'entrée et de sortie, la transformation, et autres.

### 2.2.3. L'application de NiFi

Nous traitons chaque table en utilisant des processeurs NiFi pour effectuer des tâches spécifiques selon nos besoins.

- **Les processors utilisés**

**ListFile**: C'est un processor de type source (input processors) qui liste les fichiers d'un répertoire spécifié et crée des FlowFiles contenant les métadonnées des fichiers correspondants. Il ne traite pas directement les fichiers, mais fournit des informations sur ces derniers. [14]

**FetchFile** : C'est un processor de type traitement (processing processors) qui récupère et achemine le contenu des fichiers listés par ListFile pour un traitement ultérieur. Il peut renommer ou supprimer les fichiers après récupération. Il est utilisé en combinaison avec d'autres processeurs pour effectuer des actions telles que le traitement de fichiers, la conversion de formats ou l'insertion de données dans une base de données. [14]

**PutDatabaseRecord** : C'est un processor de type destination (output processor) qui permet d'insérer, mettre à jour ou supprimer des données dans une base de données à partir des FlowFile de manière rapide et efficace. [14]

**ConvertRecord** : C'est un processeur de transformation qui permet d'utiliser un intermédiaire pour convertir les données entre différents formats. Il peut également ajouter, supprimer, renommer les champs et convertir les types de données avant d'envoyer les données à un système de stockage ou à un tiers. [14]



**QueryRecord** : est un processeur de transformation flexible qui permet d'exécuter des requêtes SQL sur des données et de les transformer de manière personnalisée avant de les envoyer à d'autres systèmes ou de les stocker dans des sources de données. [14]

**ReplaceText** : C'est un processeur de transformation qui remplace du texte dans les FlowFiles. Il supprime des caractères indésirables, remplace des valeurs obsolètes et ajoute des informations supplémentaires. Il propose des options de configuration et des fonctionnalités avancées avec des expressions régulières. Les FlowFiles modifiés sont ensuite transmis pour un traitement ultérieur. [14]

## 2.3. Le traitement et l'insertion des données

Dans cette partie nous fournirons des informations sur la manière dont nous avons utilisé NiFi dans notre travail pour traiter et nettoyer les données.

### 2.3.1. La table type

Comme on l'a expliqué plus tôt dans ce document, NiFi propose une gamme étendue de processeurs ayant diverses fonctionnalités, notamment ceux qui permettent de récupérer des fichiers. Dans notre première tentative d'intégration de fichiers CSV dans une table de notre base de données, nous avons utilisé l'un des processeurs d'entrée les plus populaires, à savoir GetFile.

- ❖ La configuration du processeur GetFile implique de sélectionner le type de processeur, d'accéder à ses propriétés et de les personnaliser selon nos besoins. Nous avons modifié les propriétés suivantes : Input Directory (répertoire source) et File Filter (filtre de fichiers). Cependant, nous avons rencontré un problème de récupération multiple du fichier. Pour résoudre ce problème, nous avons examiné le processeur ListSFTP, qui fonctionne avec des serveurs SFTP externes, mais nous l'avons exclu car nous opérons sur un serveur local. Finalement, nous avons trouvé une alternative avec le processeur ListFile, qui permet de récupérer une liste de fichiers à partir du répertoire d'entrée spécifié.

**Configure Processor** | GetFile 1.19.1

Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

Property	Value
Input Directory	E:\Noun
File Filter	nps_survey_types_202303120925.csv
Path Filter	No value set
Batch Size	10
Keep Source File	true
Recurse Subdirectories	true
Polling Interval	0 sec
Ignore Hidden Files	true
Minimum File Age	0 sec
Maximum File Age	No value set
Minimum File Size	0 B
Maximum File Size	No value set

Figure 3.1 – Configuration du processeur GetFile

- ❖ Lors de la configuration du processeur ListFile, nous définissons le répertoire source à scanner en spécifiant le chemin complet et l'expression régulière pour filtrer les fichiers requis. Nous paramétrons également la recherche pour qu'elle soit limitée au répertoire spécifié, sans inclure les sous-répertoires. Les autres propriétés sont conservées telles quelles. Pour chaque fichier listé, un FlowFile est créé avec les métadonnées du fichier, telles que le nom et le chemin absolu.

En utilisant le processeur FetchFile, nous pouvons ensuite récupérer ces fichiers en utilisant les métadonnées, ce qui nous permet d'importer les fichiers dans la base de données une seule fois sans duplication indésirable.

**Configure Processor** | ListFile 1.20.0

Stopped

SETTINGS SCHEDULING **PROPERTIES** RELATIONSHIPS COMMENTS

Required field

Property	Value
Input Directory	E:\Noun\USTHB
Listing Strategy	Tracking Timestamps
Recurse Subdirectories	false
Record Writer	No value set
<b>Input Directory Location</b>	<b>Local</b>
File Filter	nps_responses_d{12}-\d+\.csv
Path Filter	No value set
Include File Attributes	true
Minimum File Age	0 sec
Maximum File Age	No value set
Minimum File Size	0 B
Maximum File Size	No value set

Figure 3.2 – Configuration du processeur ListFile

- ❖ Dans le processeur FetchFile, nous configurons simplement la propriété "file to fetch" en spécifiant le chemin complet et le nom du fichier à récupérer. Les autres propriétés sont laissées avec leurs valeurs précédemment définies, qui conviennent à notre cas d'utilisation. Le FlowFile généré par le processeur FetchFile est ensuite transmis au processeur PutDatabaseRecord, qui convertit les données du FlowFile en un format compatible avec la structure de la base de données en utilisant des champs et des types de données prédéfinis. Une fois la conversion effectuée, le processeur insère les données dans la base de données en utilisant une requête SQL préalablement définie.

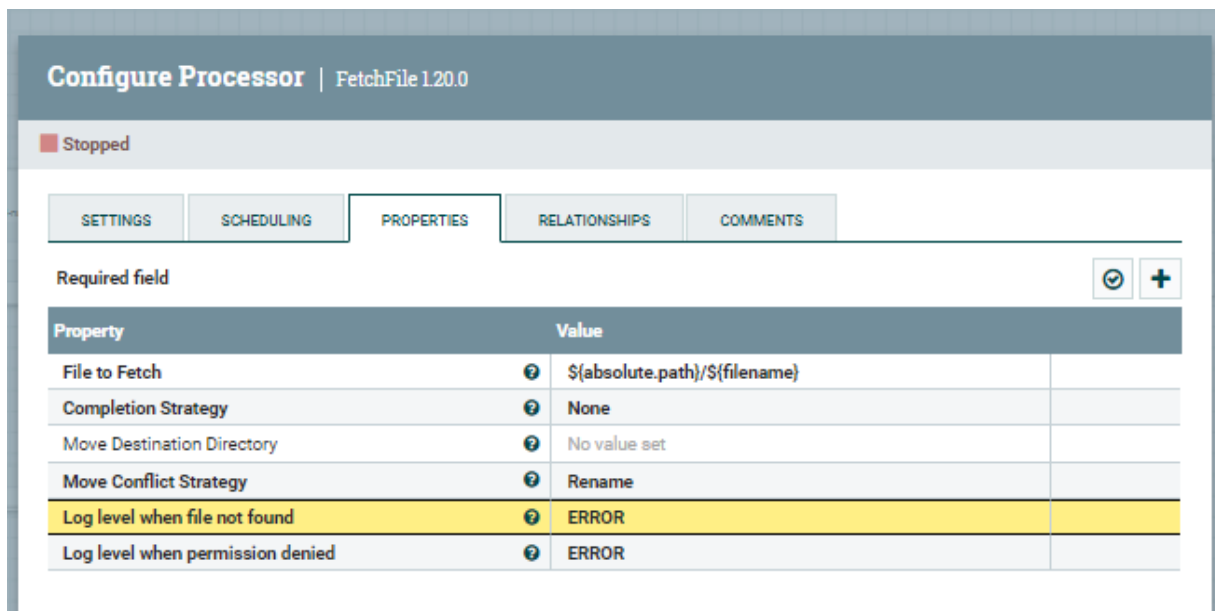


Figure 3.3 – Configuration du processeur FetchFile

- ❖ La configuration du processeur PutDatabaseRecord a impliqué plusieurs étapes. Tout d'abord, nous avons configuré le "Database Connection Pooling Service" pour établir la connexion entre NiFi et la base de données PostgreSQL. Voici les étapes que nous avons suivies :
  - Nous nous sommes assurés d'avoir installé le pilote JDBC PostgreSQL.
  - Nous avons ajouté un contrôleur de service "Database Connection Pooling Service" dans NiFi.
  - Nous avons configuré le contrôleur de service en spécifiant les informations de connexion à la base de données, telles que l'URL de connexion, la classe du pilote JDBC, l'emplacement du pilote et les identifiants d'utilisateur.
  - Nous avons cliqué sur "Apply" pour valider les changements.

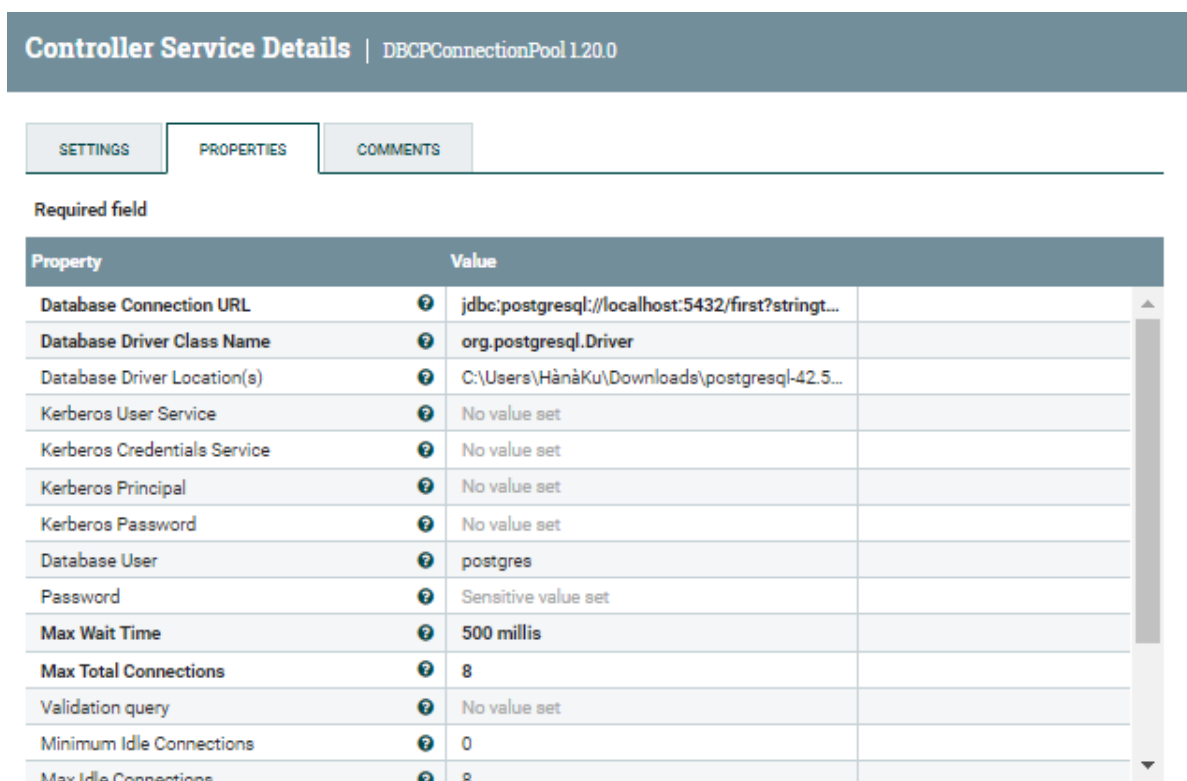


Figure 3.4 – Gestion des connexions à une base de données via le Service de mise en pool

Une fois le "Database Connection Pooling Service" configuré, nous sommes passés à la configuration du processeur PutDatabaseRecord lui-même. Voici les principales propriétés que nous avons configurées :

- "Record Reader" : Nous avons sélectionné la stratégie d'accès au schéma des données (dans notre cas, "infer schema") et nous avons spécifié l'analyseur CSV à utiliser (comme "Apache Commons CSV").
- Nous avons configuré les paramètres spécifiques au format CSV, tels que le format du fichier CSV, le séparateur de valeurs, le séparateur d'enregistrement, etc.
- Nous avons spécifié si la première ligne du fichier CSV devait être traitée comme l'en-tête des colonnes.
- Nous avons configuré les caractères de guillemets et d'échappement utilisés dans le fichier CSV.
- Nous avons configuré les propriétés de base du processeur, notamment le type de base de données (dans notre cas, PostgreSQL), le type de requête SQL (par exemple, "Insert" pour insérer de nouvelles lignes), le nom du schéma et le nom de la table cible.

En configurant correctement le processeur PutDatabaseRecord, nous avons pu exécuter des requêtes SQL sur la base de données PostgreSQL en utilisant les données extraites du flux de données.

**Configure Processor** | PutDatabaseRecord 1.20.0

Stopped

SETTINGS | SCHEDULING | **PROPERTIES** | RELATIONSHIPS | COMMENTS

Required field

Property	Value
Record Reader	CSVReaderResponses
Database Type	PostgreSQL
Statement Type	INSERT
Data Record Path	No value set
Database Connection Pooling Service	DBCPConnectionPool_Nps
Catalog Name	No value set
Schema Name	public
Table Name	responses
Translate Field Names	true
Unmatched Field Behavior	Ignore Unmatched Fields
Unmatched Column Behavior	Fail on Unmatched Columns
Quote Column Identifiers	false

CANCEL APPLY

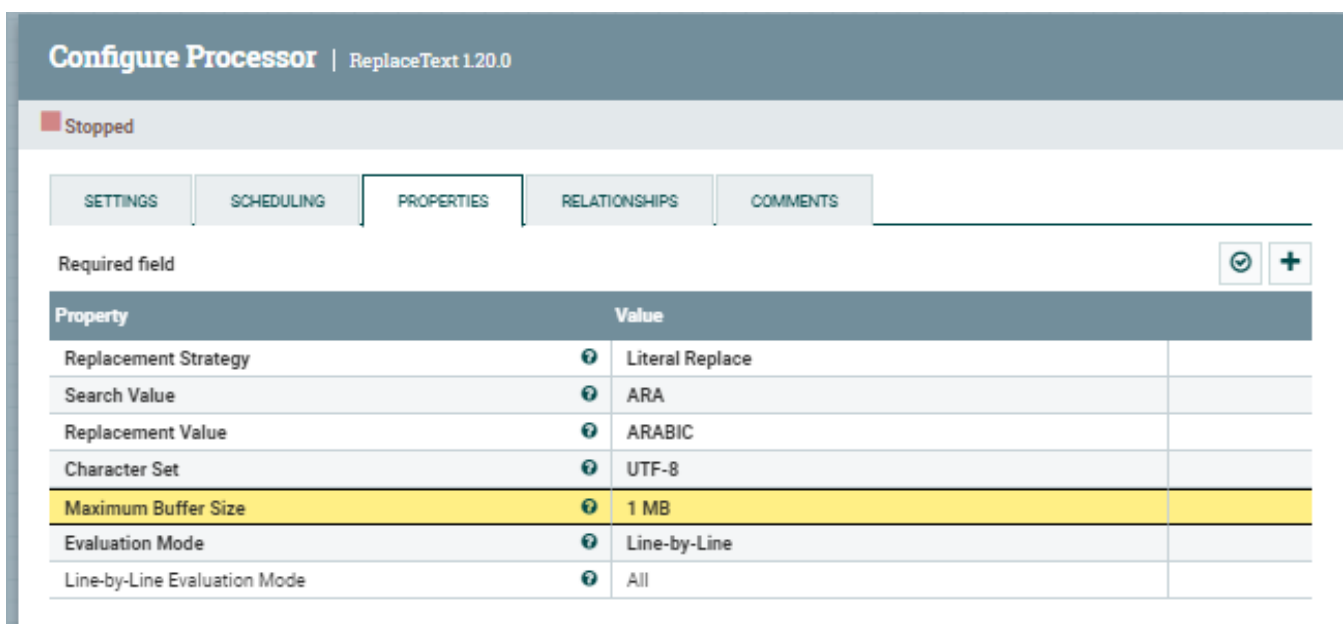
Figure 3.5 – Configuration du processeur PutDatabaseRecord

### 2.3.3. La table question

Dans le cadre de notre travail sur la table de questions, nous devons effectuer quelques ajustements mineurs avant d'insérer les données dans la base de données. Plus précisément, nous souhaitons remplacer les identificateurs de langues dans le fichier CSV (ARA, FRE, ARD) par leurs équivalents complets (ARABIC, FRENCH, ARABIC\_DZ).

Dans un premier temps, nous avons tenté de réaliser les trois modifications simultanément en utilisant un unique processeur ReplaceText.

- ❖ Le processeur ReplaceText permet de rechercher et remplacer une valeur spécifique dans un FlowFile. Pour le configurer, il faut définir la valeur à rechercher et la valeur de remplacement.



Configure Processor | ReplaceText 1.20.0

Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

Property	Value
Replacement Strategy	Literal Replace
Search Value	ARA
Replacement Value	ARABIC
Character Set	UTF-8
Maximum Buffer Size	1 MB
Evaluation Mode	Line-by-Line
Line-by-Line Evaluation Mode	All

Figure 3.6 – Configuration du processeur ReplaceText

On peut choisir la stratégie de remplacement, comme la première occurrence ou toutes les occurrences. Dans notre cas, nous utilisons la stratégie "Literal" pour remplacer les occurrences exactes sans interprétation de caractères spéciaux. Il est recommandé de laisser le jeu de caractères par défaut et d'évaluer le FlowFile ligne par ligne. Cependant, ce processeur ne peut remplacer qu'une seule valeur à la fois. Pour effectuer plusieurs remplacements, nous devons utiliser plusieurs processeurs ReplaceText. Après chaque modification, les données sont envoyées à l'étape suivante via un FlowFile.

Pour regrouper les processeurs ReplaceText, nous pouvons créer un Process group avec ListFile, FetchFile et PutDatabaseRecord. Une fois toutes les modifications effectuées, le résultat final peut être inséré dans la base de données.

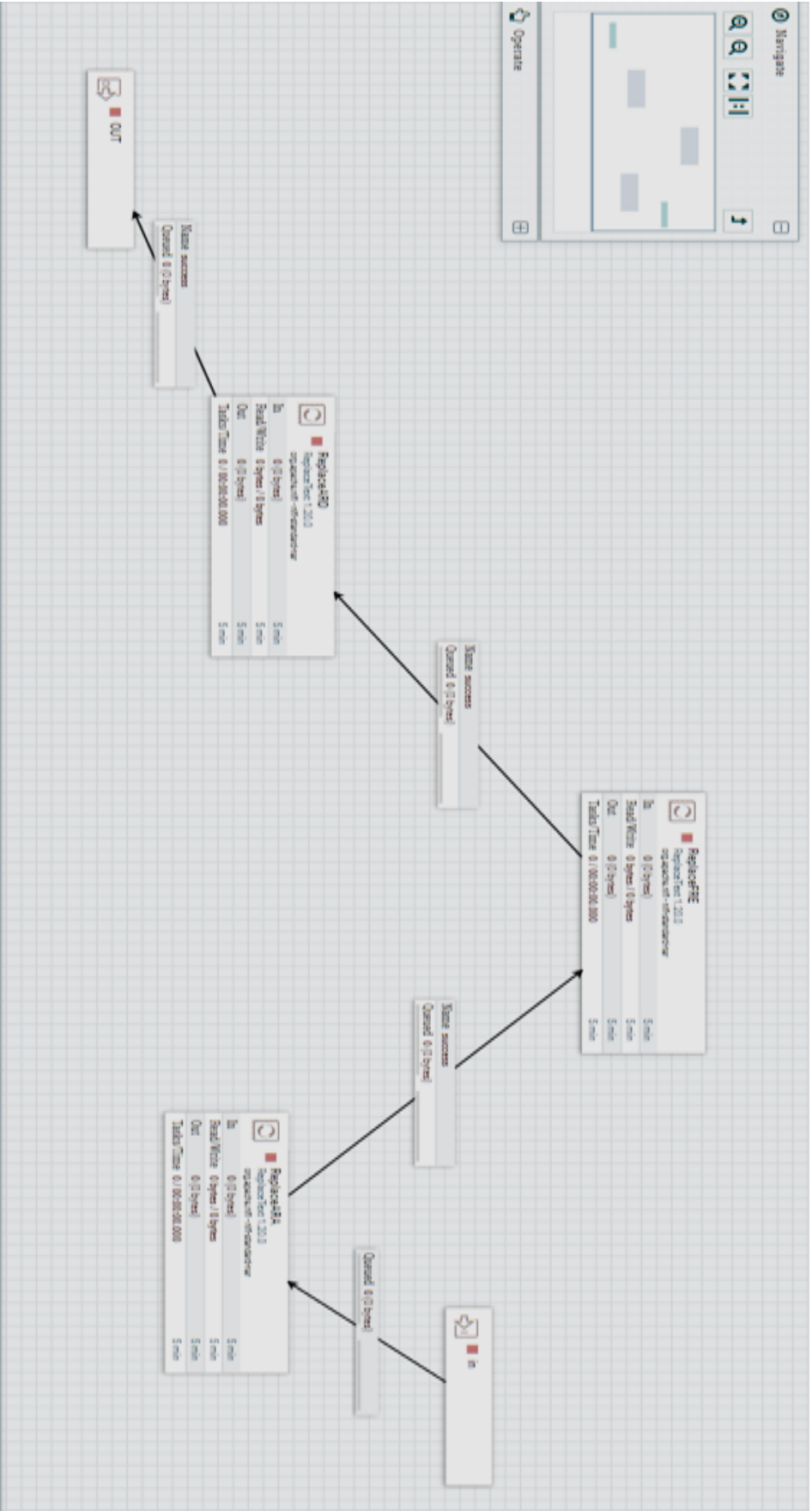


Figure 3.7 –Schéma du groupe de processeur de remplacement

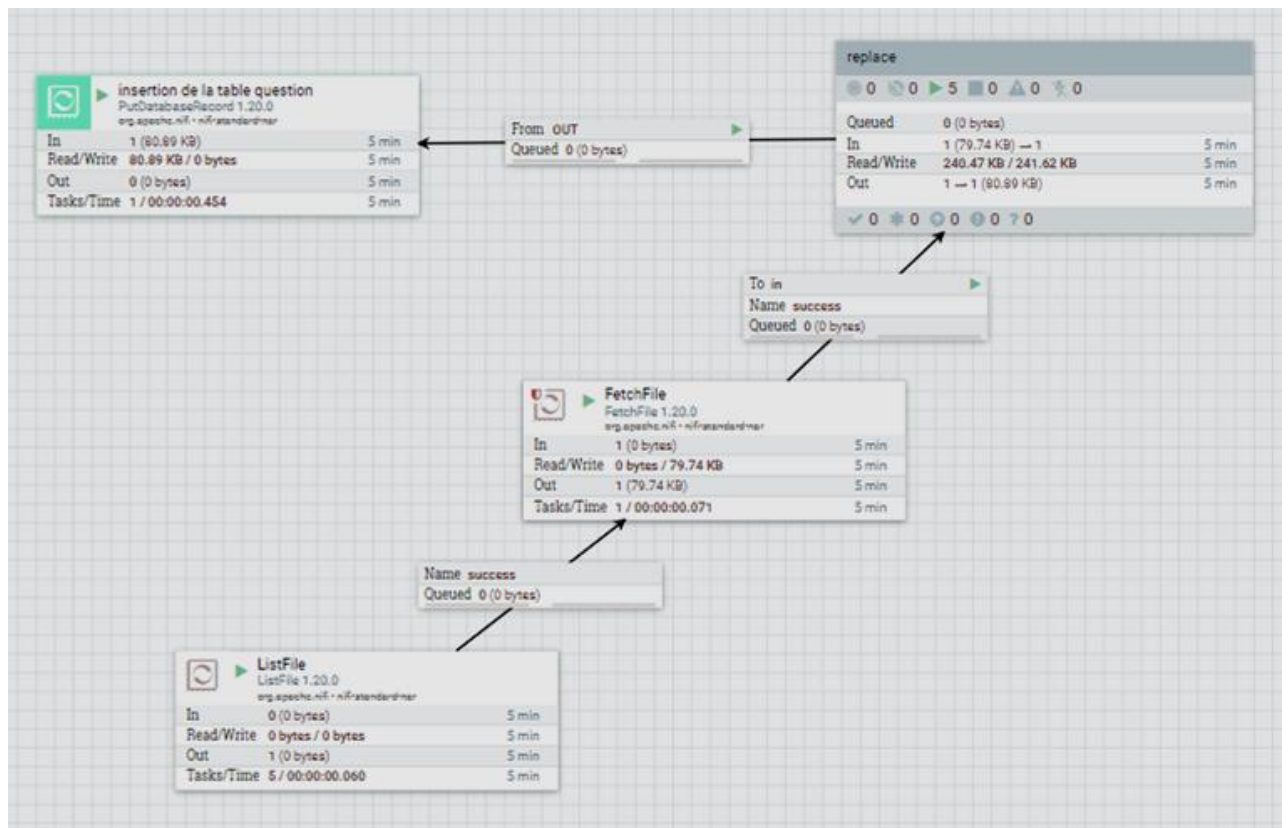


Figure 3.8 – Schéma de flux de données pour le traitement du fichier question

### 2.3.4. Table Response

Nous avons rencontré plusieurs problèmes et avons dû effectuer différentes configurations pour les résoudre. Tout d'abord, une erreur est survenue lors de la lecture d'un type de données, ce qui nous a amenés à modifier la configuration du processeur PutDatabaseRecord, en particulier le CsvReader. Nous avons changé la stratégie d'accès au schéma en utilisant les champs de chaîne de l'en-tête du fichier CSV. Malgré cela, nous avons rencontré des plantages fréquents, que nous avons tenté de résoudre en exécutant NiFi en tant qu'administrateur et en libérant les ports utilisés. Cependant, ces mesures n'ont pas résolu le problème, ce qui nous a finalement conduits à réinstaller NiFi.

Nous avons également découvert que le problème était lié à la taille du fichier CSV, qui dépassait la mémoire allouée à la JVM de NiFi. Nous avons essayé d'augmenter la mémoire, mais cela n'a pas été une solution permanente. Par conséquent, nous avons décidé de diviser le fichier en plusieurs fichiers plus petits et de les télécharger un par un. Après avoir divisé le fichier en 20 fichiers de taille 14 Mo chacun, nous avons rencontré des problèmes de configuration de la liste de fichiers lors de l'importation de tous les fichiers en même temps. Nous avons résolu ce problème en supprimant un écrivain et en réinitialisant les configurations de FetchFile aux paramètres par défaut.

Cependant, lors de l'insertion des données via le processeur PutDatabaseRecord, nous avons rencontré une erreur indiquant que la colonne "msisdn" était vide, ce qui était incorrect car elle était censée être une clé primaire. Nous avons identifié que chaque ligne était entourée de guillemets, ce qui empêchait la lecture correcte des données. Après avoir



supprimé les guillemets, nous avons redémarré le processus depuis le début et avons réussi à effectuer l'insertion sans problème.

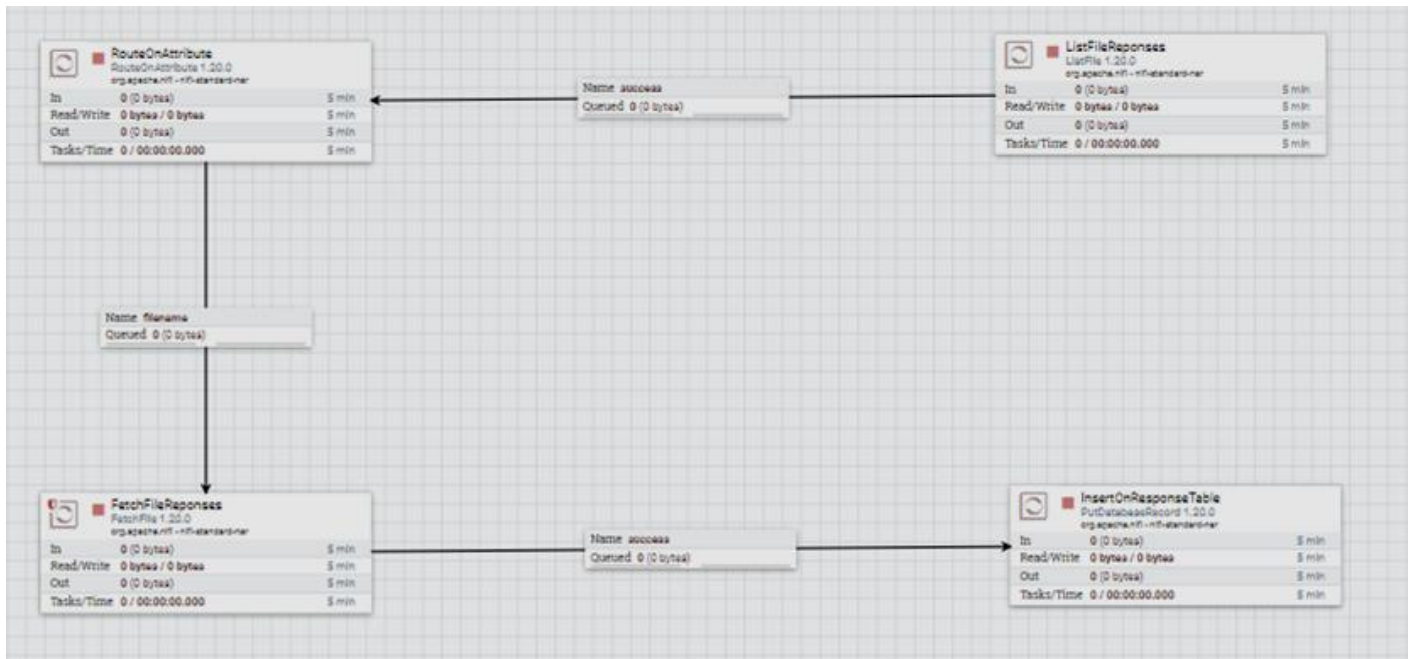


Figure 3.9 – Schéma de flux de données pour le traitement du fichier responses

### 2.3.5. Table Survey base

Notre objectif était d'effectuer un traitement sur une table en n'insérant que les colonnes nécessaires pour la visualisation. La table d'origine contenait 70 colonnes, mais nous devions en garder seulement 27, ce qui nécessitait un nettoyage des données.

Le principal défi auquel nous avons été confrontés était la taille du fichier initial, qui était trop volumineux pour être importé dans nos machines, avec une taille de 20 Go. Malgré nos tentatives pour le diviser en fichiers de 1 Go, nous avons continué à diviser le fichier jusqu'à obtenir 124 fichiers, chacun ayant une taille comprise entre 176 Mo et 173 Mo.

Après l'importation des fichiers à l'aide des processeurs ListFile et FetchFile, nous avons ajouté un intervalle de temps de 3 minutes entre chaque flux de fichiers. Chaque flux de fichiers a ensuite été traité par un processeur Query Record pour ne sélectionner que les colonnes nécessaires à l'insertion. Cependant, nous avons rencontré une erreur dans le deuxième fichier et les suivants, indiquant que les colonnes que nous essayions de récupérer n'existaient pas. Après avoir examiné les fichiers, nous avons constaté que les en-têtes manquaient en raison de la division, à l'exception du premier fichier. Nous avons donc dû ajouter manuellement les en-têtes à tous les fichiers, puis recommencer le processus depuis le début.

Lorsque nous avons enfin atteint le processeur PutDatabaseRecord, nous avons rencontré un autre problème lié au type de données d'une colonne spécifique. Le contenu était de type



texte, mais la valeur était de type Int, ce qui empêchait l'écriture dans la base de données. Pour résoudre ce problème, nous avons ajouté un processeur ConvertRecord pour convertir les types de données et garantir la compatibilité avec notre base de données. Après cela, nous avons recommencé le processus depuis le début, et cette fois-ci, tout s'est déroulé parfaitement.

- ❖ **La configuration de QueryRecord** : Consiste à spécifier les paramètres de lecture et d'écriture des lignes. Pour ce faire, nous devons configurer les composants suivants :

**Configure Processor** | QueryRecord 1.20.0

Stopped

Required field

Property	Value
Record Reader	CSVReader-query
Record Writer	CSVRecordSetWriter
Include Zero Record FlowFiles	true
Cache Schema	true
Default Decimal Precision	10
Default Decimal Scale	0
flowfile	SELECT "day", "month", "year", ingestion_dttm, msisdn, sur...

Figure 3.10 – Configuration du processeur QueryRecord

- ⇒ **Record Reader (Lecteur de données)** : Comme mentionné précédemment, nous choisissons le lecteur CSVReader et le configurons en conséquence. Il s'agit du composant chargé de lire les données à partir de la source, en l'occurrence, un fichier CSV.
- ⇒ **Record Writer (Écrivain de données)** : Nous devons évaluer cette propriété afin de pouvoir écrire le contenu d'un flow file reçu. Dans notre cas, nous choisissons le composant "CSVRecordSetWriter". Ce composant est responsable de l'écriture des données dans le format CSV.

Nous conservons les autres paramètres par défaut, qui sont les valeurs prédéfinies pour les configurations supplémentaires du processeur. De plus, nous avons ajouté une propriété spécifique qui concerne les colonnes que nous souhaitons lire à partir du flow file. Voici la requête SQL utilisée pour sélectionner les colonnes spécifiées :

```
SELECT "day", "month", "year", ingestion_dttm, msisdn, survey_type, lang_id,
question_number, first_resp, last_resp, alert, alert_status, assigned_role, assigned_user,
nps_score, status, age_group, channel, city_id, city_name, curr_price_plan_desc,
```

curr\_price\_plan\_key, handset\_brand, handset\_type, retail\_pos\_id, retail\_region\_name, segment\_type

**FROM FLOWFILE**

- ❖ **La configuration de ConvertRecord** : la configuration de ce processeur consiste à évaluer les trois propriétés suivantes :

Configure Processor | ConvertRecord 1.20.0

Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

Property	Value
Record Reader	CSVReader
Record Writer	CSVRecordSetWriter
Include Zero Record FlowFiles	true

Figure 3.11 – Configuration du processeur ConvertRecord

- ⇒ Record Reader (Lecteur de données) : Dans notre cas, nous utilisons le lecteur CSVReader, qui est responsable de la lecture des données à partir de la source au format CSV.
- ⇒ Record Writer (Écrivain de données) : Nous choisissons le composant CSVRecordSetWriter comme écrivain de données. Il est chargé d'écrire les données dans le format CSV.
- ⇒ IncludeZero Record Flowfiles (Inclure les flow files sans enregistrements) : Cette propriété est évaluée avec la valeur par défaut de "true". Lorsque cette option est activée, les flow files qui ne contiennent aucun enregistrement sont inclus dans le flux de données de sortie. Cela peut être utile pour des raisons de traçabilité ou de gestion des flux, même si aucun enregistrement n'est présent dans le flow file.

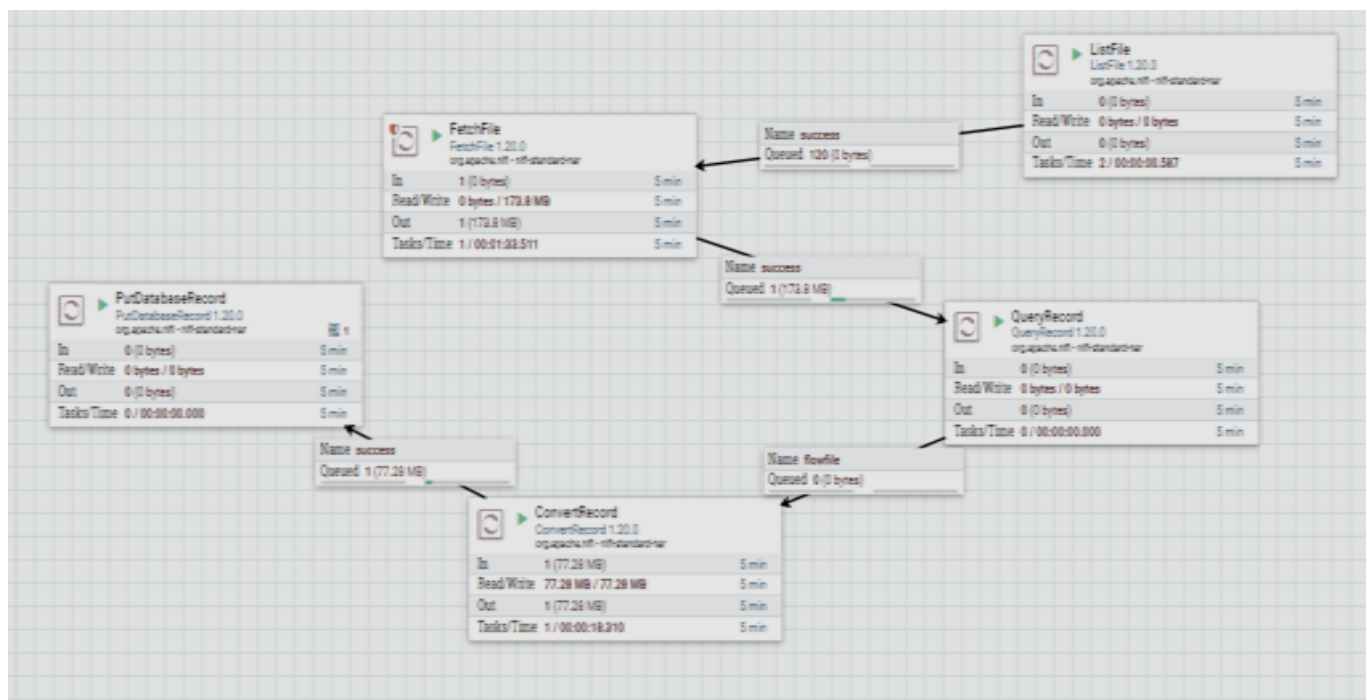


Figure 3.12 – Schéma de flux de données pour le traitement du fichier NPS base

### III. Extraction des données

Dans L'étape d'extraction des données, nous avons utilisé des requêtes SQL exécutées à l'aide de DBEaver pour extraire les informations pertinentes à partir des bases de données. Cette revue revêt une importance cruciale, car elle nous permet de collecter les données nécessaires pour alimenter nos analyses et nos réalisations.

#### 1. Les statistiques globales : Pour obtenir les nombres totaux :

- De réponses, nous devons simplement compter toutes les lignes de la table 'responses'. Cette requête nous permettra de déterminer le nombre total de réponses en prenant en compte toutes les colonnes de la table.
- De réponses valides, nous devons compter les colonnes de la table 'responses' qui ont une valeur non nulle. Cette requête nous permettra de compter le nombre de réponses qui ont été complétées sans valeurs manquantes.
- De réponses non valides, nous devons compter les colonnes de la table 'responses' qui ont une valeur nulle. Cette requête nous permettra de compter le nombre de réponses qui sont incomplètes ou qui contiennent des valeurs manquantes.
- De questions, nous devons compter les colonnes de la table 'nps\_base'. Cette requête nous permettra de déterminer le nombre total de questions dans la table.

#### 2. Les statistiques spécifiques : On s'intéresse au nombre total de réponses :

- **Selon le type de sondage** : On sélectionne et compte le nombre total de lignes à partir de la table 'nps\_base'. On effectue une jointure interne entre les tables 'nps\_base' et 'responses' en utilisant la colonne 'msisdn' qui correspond au numéro des clients crypté. On filtre les résultats pour ne conserver que les enregistrements où la colonne 'survey\_type' de la table 'nps\_base' correspond au type spécifique de sondage qu'on souhaite évaluer. Par exemple, pour le type "I Join", la valeur de 'survey\_type' serait 1. Cette requête permettra de compter le nombre total de réponses pour chaque type de sondage.
- **Selon le groupe d'âge** : On sélectionne et compte le nombre total de clients à partir de la table 'nps\_base'. On effectue une jointure interne entre les tables 'nps\_base' et 'responses' en

utilisant la colonne 'msisdn'. On filtre les résultats pour ne conserver que les enregistrements où la colonne 'age\_group' de la table 'nps\_base' correspond au groupe d'âge spécifique qu'on souhaite évaluer. Par exemple, pour le groupe d'âge '18-25', la valeur de 'age\_group' serait '18-25'. Cette requête permettra de compter le nombre total de réponses pour chaque groupe d'âge spécifique.

- **Pour chaque jour** : On sélectionne et compte le nombre total d'enregistrements et les dates correspondantes dans la table 'responses'. On regroupe les résultats par date en ignorant les informations horaires. On trie les résultats par ordre chronologique croissant de date. Pour l'extraction ci-dessous, on utilise la même jointure entre les tables 'responses' et 'nps\_base' sur la colonne 'msisdn'. Cette jointure permet de lier les données pertinentes pour chaque tableau de bord.
- **Selon la satisfaction par jour** : Nous sélectionnons et comptons le nombre total d'enregistrements et les dates correspondantes dans les tables 'responses' et 'nps\_base'. Après la jointure, nous filtrons les résultats pour ne conserver que les enregistrements où la colonne 'response\_parsed' de la table 'responses' correspond aux valeurs '7', '8', '9', '10' ou 'YES' dans le cas de la satisfaction et '0', '1', '2', '3', '4' ou 'NO' dans le cas de l'insatisfaction. Ensuite, nous regroupons les résultats par date en ignorant les informations horaires et les trions par ordre chronologique croissant de date.
- **Selon le groupe d'âge par jour** : Nous comptons le nombre total d'enregistrements et les dates correspondantes à partir de la table 'responses'. Une fois que la jointure est effectuée, nous filtrons les résultats pour ne conserver que les enregistrements où la colonne 'age\_group' de la table 'nps\_base' correspond à l'un des groupes d'âge mentionnés précédemment. Ensuite, nous regroupons les résultats par date en ignorant les informations horaires et les trions par ordre chronologique croissant de date.
- **La satisfaction de chaque type par jour** : Nous sélectionnons le nombre total d'enregistrements et les dates correspondantes à partir de la table 'responses'. Ensuite, après avoir effectué la jointure en utilisant la colonne 'survey\_type' de la table 'nps\_base', nous filtrons les résultats pour ne conserver que les enregistrements associés à l'identifiant de type de sondage spécifique. Nous appliquons également un filtre en fonction des valeurs de 'response\_parsed' dans la table 'responses', indiquant la satisfaction ou la non-satisfaction. Les résultats sont ensuite regroupés par date, en ignorant les informations horaires, et triés par ordre chronologique croissant de date.
- **La validité des réponses selon la langue** : Nous sélectionnons le nombre total d'enregistrements de 'msisdn' et les dates correspondantes à partir de la table 'responses'. En utilisant la colonne 'msisdn', nous effectuons une jointure interne entre les tables 'responses' et 'nps\_base'.

### 3. Résolution du problème de lenteur d'exécution

Nous avons constaté que l'exécution des requêtes sur la base de données prenait un temps énorme, rendant leur utilisation dans la partie réalisation quasiment impossible. Afin de remédier à cela, nous avons adopté une méthode d'agrégation couramment utilisée dans les domaines du business intelligence et du Big Data.

Cette approche consiste à créer des petites tables contenant uniquement les informations nécessaires, en se basant sur les requêtes précédentes. Cette optimisation a permis de réduire le nombre de lignes des tables de plusieurs millions à quelques centaines,

ainsi que de compresser les données de 11 Go à seulement 1 Mo. Grâce à cette méthode, l'accès aux données a été grandement facilité.

Exemple d'agrégation :

Pour extraire le nombre de réponses satisfaites par jour, une table appelée "reponse\_satisfait" a été créée, comprenant les colonnes "reponse" et "date". Cette table a été générée en utilisant la sélection mentionnée précédemment, et elle ne dépasse pas 800 lignes à la fin du processus.

#### **IV. Conclusion**

Ce chapitre a mis en lumière l'importance cruciale de l'intégration des données dans le processus de valorisation, en soulignant les étapes clés et les considérations pertinentes pour assurer la réussite de cette première étape.

Dans le chapitre suivant, nous aborderons l'environnement de développement de notre projet et son implémentation. Nous présenterons les différents outils et bibliothèques utilisés, ainsi que les éléments nécessaires à la réalisation de notre interface.

# Chapitre 4

## Réalisation, tests et analyse des résultats

### I. Introduction

Dans ce chapitre, nous débuterons par une présentation de l'environnement de développement utilisé pour l'implémentation de notre système d'analyse, incluant les outils et bibliothèques utilisés. Ensuite, nous examinerons les différentes composantes liées à l'implémentation de notre système, en fournissant une analyse des résultats obtenus.

### II. L'environnement de développement

#### 1. Environnement de travail

##### 1.1. Environnement Matériel

L'environnement matériel qui a été utilisé pour développer cette application consiste en :

- Ordinateur Portable: ACER
- Processeur : Intel(R) Core™ i5 CPU-5200U @ 2.20GHz
- Mémoire (RAM) 4 Go.
- Ordinateur Portable: DELL
- Processeur : Intel(R) Core™ i5 CPU-5300U @ 2.30GHz
- Mémoire (RAM) 8 Go.

#### 2. Langages de programmation

- **HTML** (HyperText MarkupLanguage) ou "langage de balisage d'hypertexte", est un langage informatique utilisé sur Internet pour la création d'hypertexte basé sur une structure de balisage. Cette langue est considérée comme le "squelette" des pages web. En effet, il donne une vue complète des éléments de la page, tels que des images, des textes et autres [15]
- **CSS** (Cascading Style Sheets), ou feuilles de style en cascade, est un langage de programmation web qui permet de gérer le style d'une page HTML. On peut dire que CSS concerne l'aspect visuel d'une page web, tel que les couleurs, les animations, les dimensions, etc. [16]
- **JavaScript** : Un langage de programmation qui s'exécute côté client, c'est-à-dire sur le navigateur web, et qui offre aux utilisateurs la possibilité d'interagir avec le contenu. Il est largement utilisé pour rendre les applications web plus interactives en ajoutant des fonctionnalités dynamiques. JavaScript peut être considéré comme un élément essentiel d'une page web, car il permet de manipuler le contenu et de répondre aux actions de l'utilisateur [17]
- **PHP** est un langage de script impératif orienté objet utilisé pour créer des pages Web dynamiques via un serveur HTTP. Il s'agit d'un langage libre et il peut également être utilisé en tant que langage interprété localement. [18]



- **SQL** (StructuredQueryLanguage) est un langage de programmation utilisé pour stocker et traiter des informations dans des bases de données relationnelles. Les bases de données relationnelles organisent les données sous forme de tableaux avec des lignes et des colonnes, représentant les attributs et les relations entre les valeurs. Avec SQL, vous pouvez effectuer des opérations telles que le stockage, la mise à jour, la suppression, la recherche et l'extraction de données dans la base de données. Il permet également de maintenir et d'optimiser les performances de la base de données. [19]



### 3. Outils et logiciels de développement

- **Postgresql** : C'est un système de gestion de base de données relationnelle open source fiable et flexible. Il offre des fonctionnalités avancées pour le stockage et la gestion sécurisée des données complexes. Avec plus de 35 ans de développement, PostgreSQL dispose d'une architecture solide et d'une intégrité des données éprouvée. Il est compatible avec différents systèmes d'exploitation et propose des extensions telles que PostGIS pour les données géo spatiales. PostgreSQL est largement utilisé et bénéficie d'une communauté open source active. [20]
- **pgAdmin** : C'est un logiciel open source de gestion et d'administration des bases de données PostgreSQL. Il simplifie la création, la gestion et l'utilisation des objets de base de données. PgAdmin est livré avec un assistant de connexion, un éditeur SQL intégré pour importer des scripts SQL, et un système pour construire automatiquement des scripts SQL qui permet d'exécuter des scripts SQL sur le shell de ligne de commande de la base de données. [21]
- **Dbeaver** : C'est un logiciel gratuit et open source multiplateforme de gestion et d'administration de bases de données, ainsi qu'un client SQL, destiné aux administrateurs de bases de données, développeurs et analystes. Souvent considéré comme un outil universel de bases de données, DBeaver prend en charge plus de 80 types de bases de données, qu'elles soient relationnelles ou non, telles que PostgreSQL, MySQL, SQLite, SQL Server, DB2, Sybase, Phoenix, MS Access, Teradata, Apache Hive, etc. [22]
- **Apache Nifi** : C'est une plateforme open source qui automatise le flux de données entre les systèmes tels que bases de données, services Cloud et appareils IoT (Internet of things). Elle est reconnue comme une technologie clé pour l'intégration des modèles de systèmes et d'applications, et est particulièrement utile pour les organisations traitant de grandes quantités de données, Nifi offre des fonctionnalités avancées pour gérer efficacement les flux de données entre les systèmes. [23]
- **XAMPP** : C'est une distribution Apache gratuite et conviviale incluant MariaDB, PHP et Perl. Il offre un environnement de développement local facile à installer et à utiliser, permettant d'exécuter des applications web et des bases de données sur son propre ordinateur. Les développeurs l'utilisent largement pour simplifier le processus de développement, de test et de débogage avant de déployer leurs applications sur un serveur en ligne. [24]
- **Notepad++** : C'est un éditeur de code source gratuit et open source pour Windows. Il offre une interface conviviale et de nombreuses fonctionnalités avancées pour l'édition de texte et le développement de logiciels. Il est largement utilisé par les programmeurs pour sa légèreté, sa rapidité et sa polyvalence [25]



PostgreSQL



pgAdmin  
Management Tools



**Visual Studio Code :** C'est un éditeur de code léger et puissant disponible sur Windows, macOS et Linux. Il prend en charge JavaScript, TypeScript et Node.js, et propose de nombreuses extensions pour d'autres langages et environnements d'exécution. [26]



#### 4. Librairies et bibliothèques

Dans ce qui suit, nous allons décrire les bibliothèques et les librairies qui ont été les plus utilisées dans notre projet.

- **Bulma**

C'est une bibliothèque CSS qui propose des classes CSS pour vous aider à styliser le code HTML. Il offre deux options d'utilisation : vous pouvez soit utiliser le fichier .css précompilé, soit installer les fichiers .sass, qui permettent une personnalisation en fonction de vos besoins. [27]

- **Animate**

C'est une bibliothèque d'animations prédéfinies conçue pour assurer une compatibilité entre les navigateurs. Elle permet de mettre l'accent sur des éléments, d'améliorer les pages d'accueil, de créer des sliders et de guider l'attention de manière efficace. [28]

- **Chart.js**

C'est une bibliothèque graphique JavaScript populaire utilisée pour créer des tableaux et des graphiques interactifs et personnalisables. Il est largement adopté et activement maintenu, offrant des fonctionnalités adaptées aux débutants et une communauté de soutien. [29]



Chart.js

- **Font Awesome**

C'est une bibliothèque d'icônes très répandue, utilisée par les concepteurs, les développeurs et les créateurs de contenu. Dans notre projet, nous utilisons spécifiquement la version 5.15.4. [30]

## 2. Visualisation des données

Comme mentionné précédemment dans le chapitre précédent, notre projet vise principalement à exploiter les données pour les rendre analysables. Pour ce faire, nous avons opté pour l'utilisation de la méthode de la visualisation des données en créant des tableaux de bord. Ces tableaux de bord jouent un rôle essentiel dans la facilitation de la compréhension des informations clés et répondent aux besoins spécifiques de l'entreprise.

### 1. Définition de la visualisation

Grâce à des représentations visuelles telles que des tableaux de bord, des rapports interactifs et des graphiques, la visualisation des données permet de développer rapidement des perspectives d'affaires et d'optimiser la prise de décision en fournissant une vue d'ensemble claire des performances, en identifiant l'importance des données et en suivant les tendances au fil du temps. [31]

### 2. Définition des tableaux de bords

Ce sont des outils permettant de surveiller, analyser et afficher les données de manière conviviale. Ils facilitent la connexion de différentes sources de données et permettent d'extraire des informations pertinentes. Les tableaux de bord sont associés à des indicateurs de performance clés pour une analyse décisionnelle et une exploration approfondie des



informations. Les tableaux de bord fournissent des informations actualisées, éclairent les décisions commerciales et permettent une action proactive. [32]

#### IV. Présentation de l'application

Nous fournissons une description pour chaque fenêtre correspondant aux différentes interfaces qui composent notre application web.

##### 1. Page d'authentification

Dans cette page, l'employé s'authentifie en saisissant son email et son mot de passe. Si les informations sont incorrectes, le système affiche un message d'erreur sur la même page, permettant à l'utilisateur de réessayer.



Figure 4.1 – La page authentification

##### 2. Page d'inscription

Si l'administrateur souhaite créer un compte pour un employé, il doit remplir le formulaire qui apparaît sur cette page. Le formulaire contient les informations de l'employé, ainsi que son rôle (administrateur ou utilisateur simple). Des messages d'avertissement s'afficheront en cas de saisie d'un e-mail déjà existant ou en cas de différence entre le mot de passe et sa confirmation.

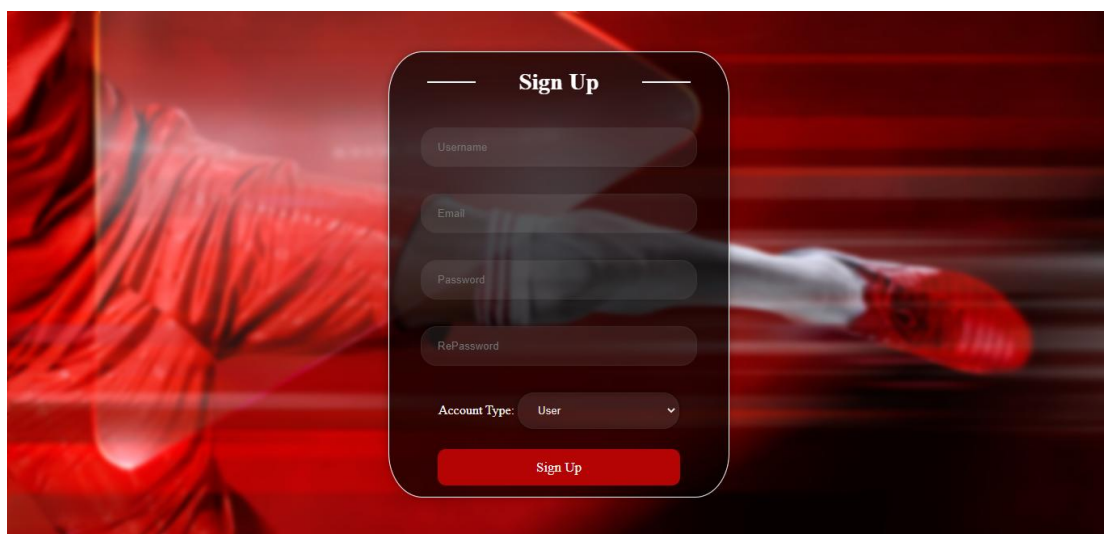


Figure 4.2 – La page Inscription

### 3. Page d'accueil

La page présente les contenus clés de l'application, y compris les mises à jour récentes des scores NPS. Elle permet aux utilisateurs de se déconnecter, de modifier leur profil et d'accéder à des tableaux de bord statistiques plus spécifiques. Elle met en évidence la différence entre les types de comptes utilisateurs, en particulier le compte administrateur qui a l'autorisation de créer de nouveaux comptes pour les employés. Cette fonctionnalité est exclusive aux administrateurs.

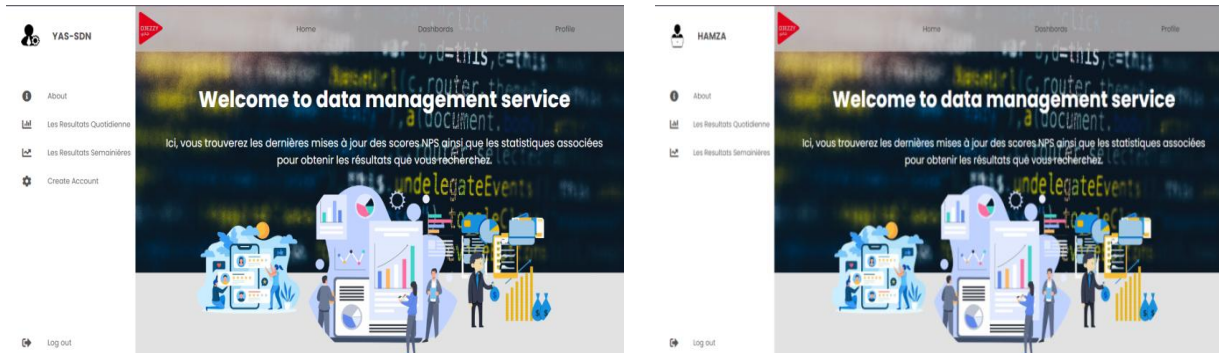


Figure 4.3 - La page d'accueil (en tant qu'administrateur et en tant qu'utilisateur)

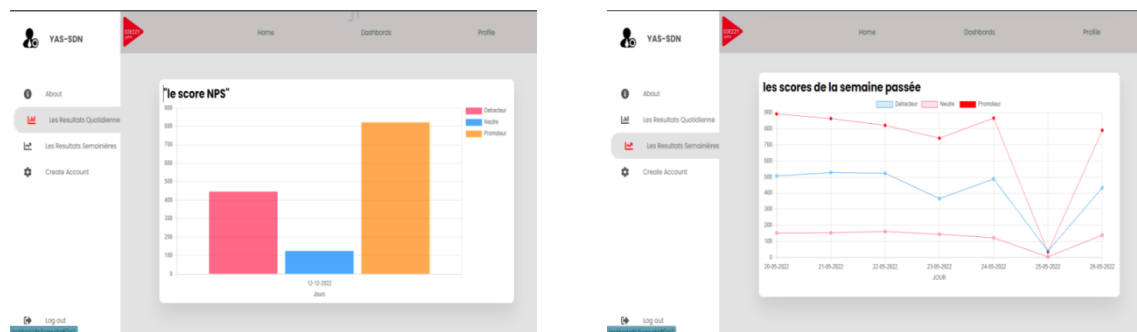


Figure 4.4 - Les tableaux d'actualité

### 4. Page de profil d'utilisateur

Cette page permet à l'employé de modifier son nom d'utilisateur et son mot de passe. Après avoir modifié l'une de ces informations, l'employé devra se reconnecter pour que les changements prennent effet.

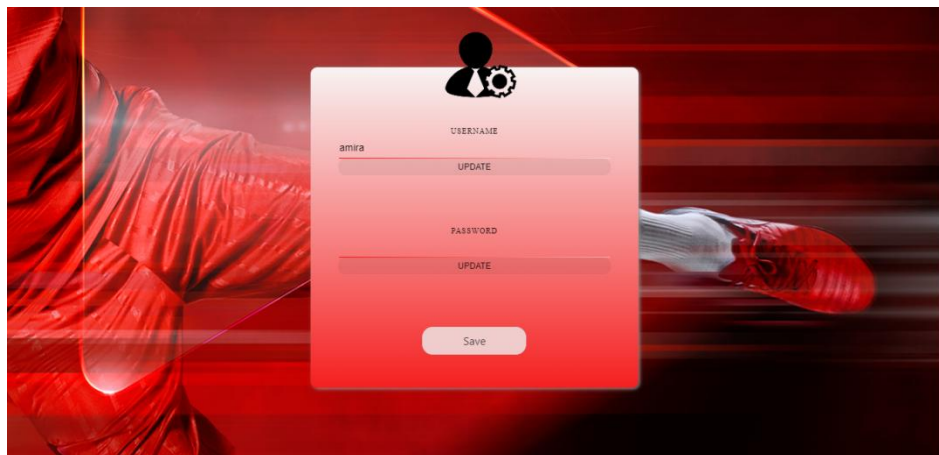


Figure 4.5 – La page de profile

## 5. Page des tableaux de bords

Cette page présente les tableaux de bord que nous avons réalisés, ainsi que les options permettant de se déconnecter et de modifier le profil de l'utilisateur. La page propose des tableaux de bord personnalisés avec des fonctionnalités de déconnexion, de modification de profil et de retour à la page d'accueil. Les tableaux sont générés à partir de données extraites dans la partie d'agrégation et visualisées à l'aide de PHP et Chart.js. Un filtre de date permet d'ajuster la présentation des graphiques en fonction de la plage de dates choisie.

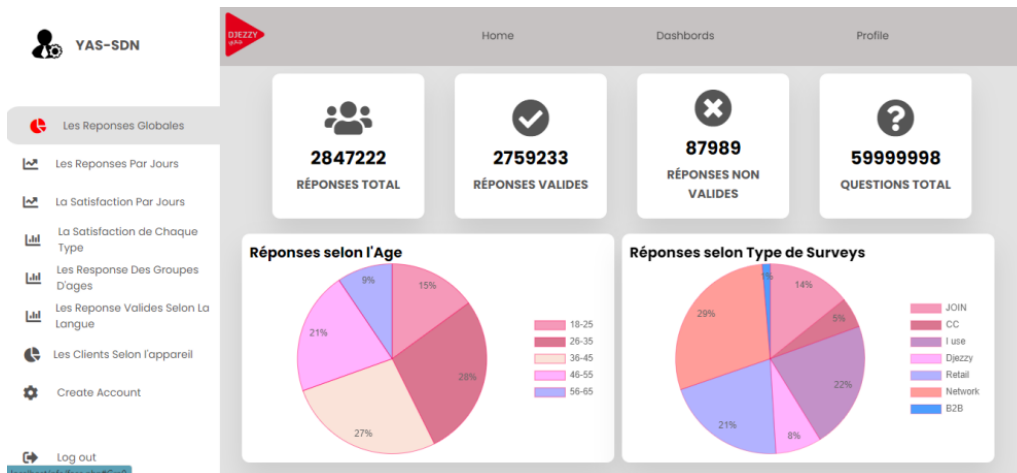


Figure 4.6 – Tableau des nombre de réponses globales

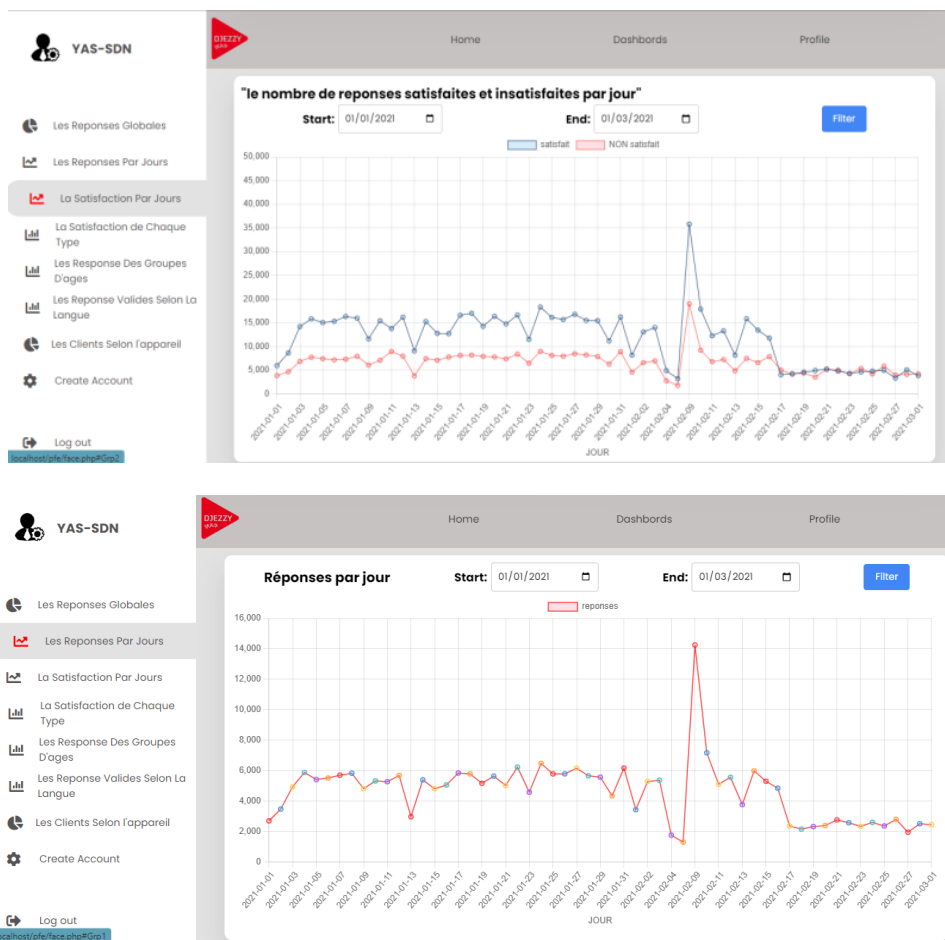


Figure 4.7 - Exemple de graphe linéaire filtré

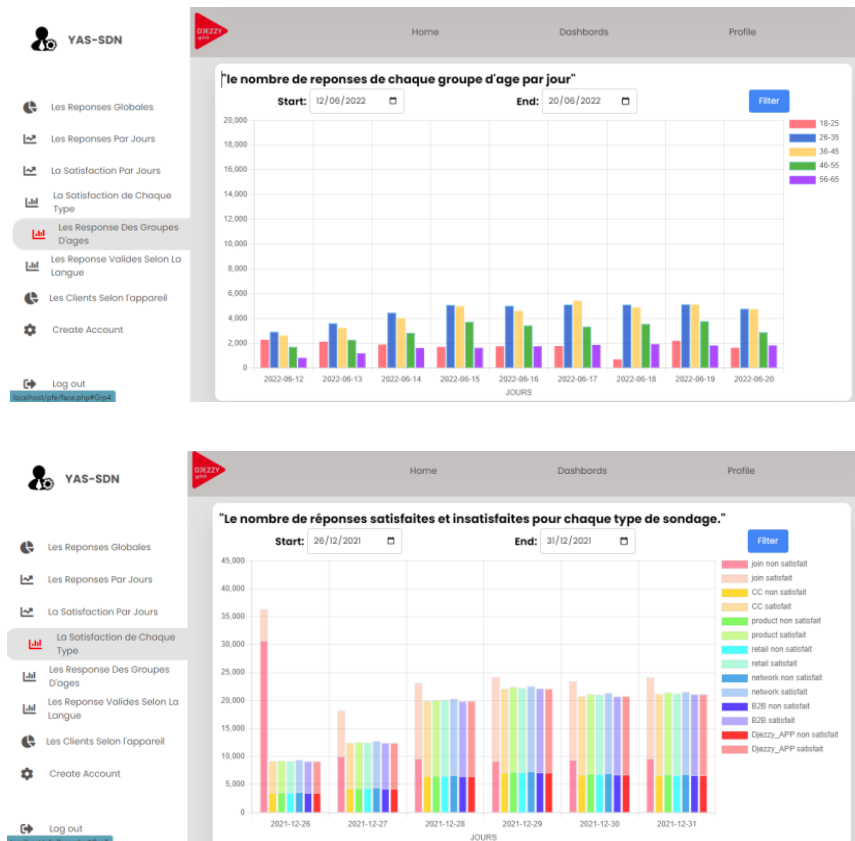


Figure 4.8 –Exemple de graphe à barres filtré

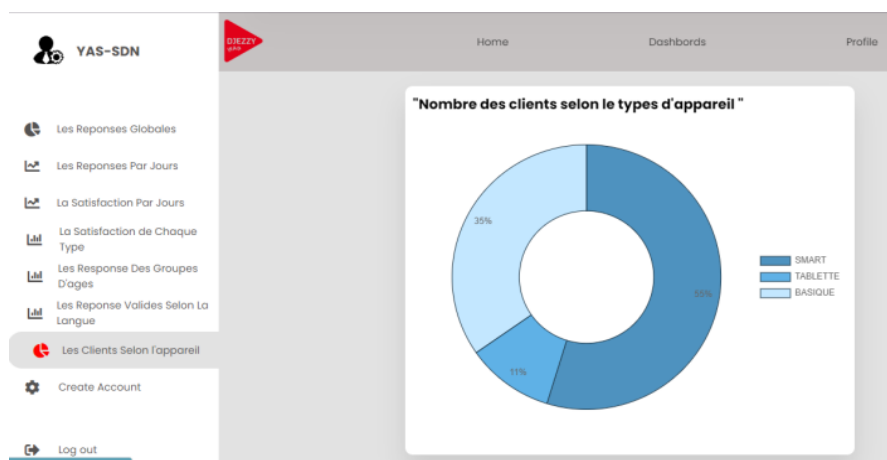


Figure 4.9 –Exemple de graphe en secteurs

## V. Analyse des statiques et résultat

L'objectif principal de cette plateforme est d'assister l'entreprise dans la prise de décisions visant à améliorer ses performances et sa rentabilité en analysant les résultats obtenus à partir des graphes. En se basant sur notre expérience limitée dans le domaine du marketing, il est possible de tirer certaines conclusions à partir de ces résultats. Par exemple, à travers la première section qui présente des statistiques numériques nous avons constaté qu'un pourcentage significatif des clients auxquels l'enquête a été envoyée n'a pas répondu. Cependant, il est important de souligner que le taux de réponses invalides ne représente que

3% de l'ensemble des réponses reçues, ce qui renforce la crédibilité des résultats analysés. Les deux diagrammes circulaires illustrent également la répartition des réponses en fonction de l'âge et du type de vote.

En se référant aux délais de réponse en jours, une analyse des résultats sur deux ans consécutives et début de l'année 2023 permet de conclure que le premier trimestre de l'année 2021 présente un taux élevé de satisfaction client, qui diminue du deuxième trimestre jusqu'à la fin de l'année 2022, pour ensuite connaître une légère augmentation au début de l'année 2023. Nous pouvons observer que les clients âgés entre 26 et 35 ans, ainsi que ceux âgés de 36 à 45 ans, sont les plus enclins à répondre à l'enquête. De plus, le pourcentage le plus élevé concerne le vote du réseau.

Nous avons remarqué que la majorité des votes selon le type de service présentent un pourcentage élevé de satisfaction. Cependant, au cours de la seconde moitié de l'année 2022, nous avons constaté une augmentation du pourcentage d'insatisfaction dans tous les types de votes. Nous souhaitons mettre l'accent sur deux types importants, à savoir B2B et I Join. Il est important de prendre en charge rapidement les cas où les clients de type entreprise et ceux qui ont récemment rejoint nos services ne sont pas satisfaits des prestations. Par exemple, lorsqu'une entreprise n'est pas satisfaite, cela peut mettre en péril la continuité du contrat qui nous lie. De plus, il est pertinent de noter que le nombre de clients insatisfaits parmi les nouveaux venus a connu une augmentation significative à la fin de 2022 et au début de 2023 par rapport à la période précédente, où le pourcentage d'insatisfaction était pratiquement inexistant.

En ce qui concerne les réponses validées selon les langues, on constate une forte proportion dans la langue arabe par rapport au français. Cependant, il est à noter que la langue arabe algérienne est totalement absente.

En ce qui concerne les types d'appareils des clients, une analyse révèle que le taux de visionnage est de 55 % pour les utilisateurs de smart phones, ce qui est relativement faible par rapport aux attentes. En revanche, pour les détenteurs de téléphones basiques, ce taux s'élève à 35 %. Dans cette situation, l'entreprise a la possibilité d'ajuster ses offres en fonction du type de téléphone détenu par le client. Il est important de noter que la plupart des offres proposées par Djezzy sont axées sur les appels et l'accès à Internet, ce qui peut ne pas correspondre aux besoins des clients utilisant des téléphones basiques.

## **VI. Conclusion**

Ce chapitre est décomposé en deux parties distinctes. La première partie a été consacrée à la présentation des outils utilisés lors de la réalisation du projet, tandis que la deuxième partie a mis en évidence des exemples représentatifs d'interfaces concrètes. Tout au long de ce projet, nous avons fait de notre mieux pour couvrir un large éventail des besoins de l'entreprise et améliorer l'efficacité des fonctionnalités du site.

# Conclusion Générale

Dans cette étude, notre attention s'est portée sur la visualisation et l'analyse des données du Net Promotor Score (NPS) au profit de l'entreprise de téléphonie Djezzy. Nous avons développé une interface web dynamique composée de tableaux de bord afin de présenter les résultats recherchés de manière accessible et conviviale.

Pour mener à bien cette recherche, une étude bibliographique approfondie a été entreprise, portant sur les domaines d'étude liés à notre thème.

- Marketing ciblé, mettant en exergue son importance sur le marché des affaires et exposant les différentes notions qui lui sont associées, ainsi que l'utilité du Net Promoter Score (NPS) dans ce contexte.
- La gestion des données et l'analyse ont été traitées, en mettant en lumière la Business Intelligence (BI) et son rôle dans la fourniture d'informations pertinentes. L'étape essentielle de l'Extract-Transform-Load (ETL) a également été mentionnée en tant que processus clé dans la préparation des données pour une utilisation efficace dans la BI.
- Le domaine du Big Data a été exploré, définissant ce concept et soulignant son rôle essentiel dans la compréhension des clients, l'optimisation des opérations et la formulation de stratégies efficaces.

Lors de la conception, nous avons fait face à des problèmes liés à l'intégration des données en raison de leur volume important. Après avoir recherché les techniques appropriées, nous avons réussi à rendre ces données exploitables, mais cela a nécessité du temps pour accomplir cette intégration.

Un autre défi majeur était l'impossibilité d'utiliser ces données pendant la phase de développement en raison du temps d'exécution élevé requis pour accéder à la base de données et extraire les informations nécessaires. Pour remédier à cela, nous avons utilisé une approche alternative pour optimiser l'accès et l'extraction des données. Une fois cette étape terminée, nous avons pu passer à la phase de développement et créer les interfaces nécessaires.

Au cours de la réalisation de l'application, nous avons fait usage d'une variété de langages et d'outils de développement et de programmation. Nous avons utilisé HTML, CSS, PHP, JavaScript et Chart.js pour créer les tableaux de bord, en veillant à fournir une visualisation claire des données. Par ailleurs, nous avons opté pour PostgreSQL comme système de gestion de base de données, assurant ainsi une structure solide pour le stockage et l'accès aux informations nécessaires. Cette combinaison de langages et d'outils a permis de développer une application fonctionnelle et performante.

Pour l'avenir il serait intéressant développer des fonctionnalités de segmentation avancées pour analyser le NPS en fonction de critères tels que l'âge, le sexe, la localisation géographique, etc. la personnalisation des tableaux de bord en sélectionnant les indicateurs clés, les graphiques et les visualisations qui leur sont les plus pertinents l'analyse textuelle pour extraire des informations à partir des commentaires des clients et. De plus, l'intégration

de techniques de Machine Learning, comme la classification des commentaires et la prédiction des scores, pourrait améliorer notre analyse du NPS.

En conclusion, ce projet de fin d'études de Licence a représenté une opportunité précieuse pour mettre en pratique et consolider nos connaissances acquises tout au long de notre parcours universitaire. Nous espérons que ce travail aura un impact positif sur la rentabilité de l'entreprise ainsi que sur la relation avec ses clients, et que les résultats obtenus apporteront une valeur ajoutée à la société.

# Références Bibliographique

- [1] Outil 3. NPS (Net Promoter Score) Florence Gillet-Goinard, Bernard Seno Dans La boîte à outils de la qualité (2020), pages 20 à 21. (04/2023)
- [2] Kenneth\_C.Laudon,Jane\_P\_.Laudon\_-\_Management\_Information\_Sysrem\_13th\_Edition Chapter 2 Global E-business and Collaboration 77 (08/06/2023)
- [3] <https://aws.amazon.com/what-is/etl/> (04/2023)
- [4] An Introduction to Big Data and Hadoop - Student Guide (04/2023)
- [5] Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 2053951716631130.
- [6] <https://www.silicon.fr/data-age-2025-donnees-cloud-stockage-226317> (04/2023)
- [7] <https://www.praedictia.com/page/big-data/lhistoire-du-big-data> (04/2023)
- [8] <https://www.djezzy.dz/djezzy/nous-connaître/a-propos-de-djezzy/> (04/2023)
- [9] <https://www.djezzy.dz/mentions-legales/> (04/2023)
- [10] <https://www.avaya.com/> (04/2023)
- [11] <https://www.teradata.com/> (04/2023)
- [12] <https://cassandra.apache.org/> (04/2023)
- [13] <https://kafka.apache.org/> (04/2023)
- [14] [.https://nifi.apache.org/docs.html](https://nifi.apache.org/docs.html) (04/2023)
- [15] <https://developer.mozilla.org/en-US/docs/Web/HTML> (05/2023)
- [16] [.https://developer.mozilla.org/en-US/docs/Web/CSS](https://developer.mozilla.org/en-US/docs/Web/CSS) (05/2023)
- [17] <https://www.w3schools.com/js/> (05/2023)
- [18] <https://www.php.net/> (05/2023)
- [19] <https://aws.amazon.com/what-is/sql/> (05/2023)
- [20] <https://www.postgresql.org/about/> (05/2023)



- [21] <https://www.pgadmin.org/> (05/2023)
- [22] <https://dbeaver.io/about/> (05/2023)
- [23] <https://nifi.apache.org/>(05/2023)
- [24] <https://www.apachefriends.org/> (05/2023)
- [25] <https://notepad-plus-plus.org/> (05/2023)
- [26] <https://code.visualstudio.com/docs> (05/2023)
- [27] <https://bulma.io/documentation/overview/> (05/2023)
- [28] <https://animate.style/> (05/2023)
- [29] <https://www.chartjs.org/docs/> (05/2023)
- [30] <https://fontawesome.com/> (05/2023)
- [31] <https://aws.amazon.com/fr/what-is/data-visualization/> (05/2023)
- [32] <https://powerbi.microsoft.com/> (05/2023)
- [33] <https://kafka.apache.org/> (05/2023)
- [34] <https://storm.apache.org/> (05/2023)
- [35] <https://spark.apache.org/docs/latest/streaming-programming-guide.html> (05/2023)
- [36] <https://nightlies.apache.org/flink> (05/2023)



- On ajoute le fichier .jar du pilote JDBC PostgreSQL téléchargé lors de l'étape précédente. On localise le fichier postgresql.jar dans le répertoire d'installation.

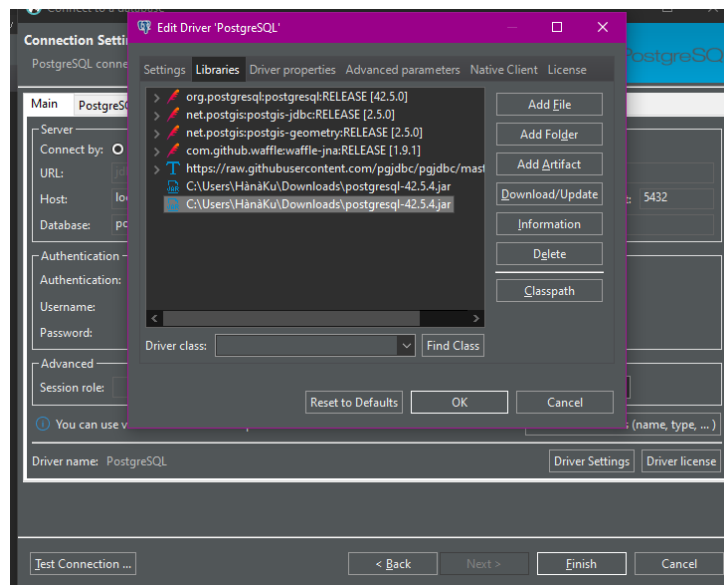


Figure A.4 – Sélection de schéma de fichier .jar

Une fois connectée, la base de données que nous avons créée dans PostgreSQL apparaîtra dans l'interface de DBeaver.

## - Création de tables

Pour créer une table dans DBeaver à l'aide d'un script SQL, les étapes sont comme suit : La création de tables dans DBeaver est facile. Il suffit de sélectionner la base de données appropriée dans laquelle on souhaite créer la table.

Dans l'éditeur SQL, sur "Nouveau script SQL" on rédige notre script SQL pour créer la table, en spécifiant les colonnes, les types de données et les contraintes nécessaires. Après avoir écrit le script, on exécute le script SQL.

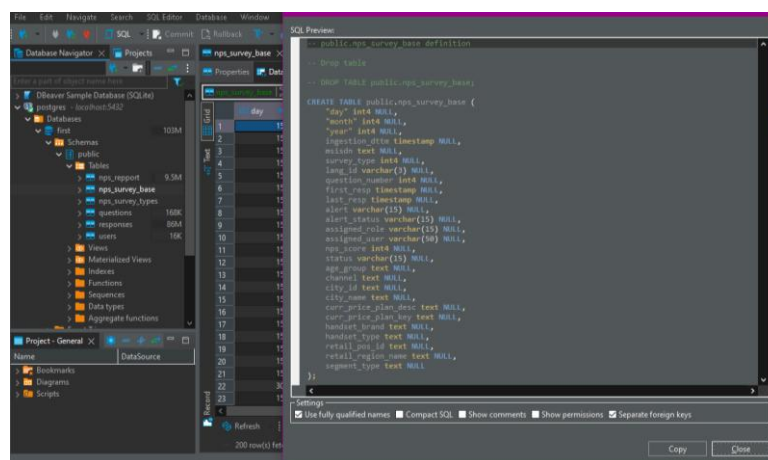


Figure A.5 - Un exemple de requête de création pour l'une des tables

## **II. Annexe B: Les outils qu'on peut utiliser au lieu d'utiliser NiFi**

Il existe plusieurs outils de traitement de flux de données disponibles sur le marché, chacun ayant ses propres avantages et inconvénients. Le choix de l'outil dépendra des besoins spécifiques du projet et des capacités stratégiques de l'équipe. Parmi eux :

- **Apache Kafka :**  
Apache Kafka est une plateforme de streaming d'événements distribués open source utilisée par des milliers d'entreprises pour les pipelines de données hautes performances, l'analyse de streaming, l'intégration de données et les applications critiques. [33]
- **Apache Storm :**  
Apache Storm est un système de calcul en temps réel distribué gratuit et open source. Apache Storm facilite le traitement fiable de flux de données illimités, faisant pour le traitement en temps réel ce que Hadoop a fait pour le traitement par lots. Apache Storm est simple, peut être utilisé avec n'importe quel langage de programmation et est très amusant à utiliser ! [34]
- **Apache Spark Streaming:**  
Spark Streaming est une extension de l'API principale de Spark, offrant un traitement en temps réel de flux de données évolutifs et à haut débit. Il permet l'ingestion de données provenant de diverses sources telles que Kafka, Kinesis et les sockets TCP, et permet l'utilisation d'algorithmes avancés à l'aide de fonctions de haut niveau comme map-reduce, join et window. [35]
- **Apache Flink :** C'est un Framework et un moteur de traitement distribué pour les calculs avec état sur des flux de données illimités et limités. Flink a été conçu pour fonctionner dans tous les environnements de clusters courants, effectuer des calculs à une vitesse en mémoire et à n'importe quelle échelle [36]

Quant aux autres outils de traitement de flux de données, ils ont également leurs propres avantages et inconvénients. Par exemple, Apache Kafka est souvent utilisé pour la messagerie, la collecte, la transformation et la distribution de données en continu, tandis qu'Apache Flink est souvent utilisé pour l'analyse de données en temps réel. Le choix de l'outil dépendra des besoins spécifiques du projet et des capacités techniques de l'équipe.

## Résumé

Le marché de la téléphonie en Algérie est en pleine croissance, ce qui pousse les opérateurs téléphoniques à maximiser leurs profits tout en améliorant l'expérience utilisateur. Djezzy, en tant qu'opérateur de référence, se concentre sur l'exploitation efficace des données massives pour répondre aux exigences du marché et garantir la satisfaction des clients. Dans ce contexte, Djezzy utilise le Net Promoter Score (NPS) pour mieux comprendre son public cible et améliorer ses services.

Notre projet vise à concevoir des workflows qui permettront de traiter et de nettoyer les données du NPS, de les insérer dans une base de données, puis de les extraire pour les représenter sous forme d'une interface web conviviale.

Cette interface facilitera l'analyse des données par les spécialistes de l'entreprise, leur permettant ainsi de prendre des décisions éclairées pour améliorer les services de la société.

**Mot-clés:** NPS, Marketing ciblé, big data, business intelligence, ETL, integration, extraction, visualization.