



# Mémoire de stage de fin d'études Pour l'obtention du Diplôme de Master

## Informatique Décisionnelle Et Vision Intelligente Spécialité : Informatique

### Analyse Collaborative dans le décisionnel pour tous Projet ANR BI4people



Réalisé par :

TAJMOUATI Moad

Sous la direction de :

M. Ali YAHYAOUY

Professeur à la FSDM-Université Sidi Mohamed Ben Abdellah  
Fès – Maroc

Mme. Sabine Loudcher

Professeure des universités, université Lumière Lyon 2 – France.

Mme. Cécile Favre

Maîtresse de conférences, université Lumière Lyon 2 – France.

M. Jérôme Darmont

Professeur d'informatique, université Lumière Lyon 2 – France  
et coordinateur du projet BI4people.

Soutenu le XX/07/2022, devant le jury composé de :

M. Ali YAHYAOUY

Professeur à la FSDM

Président du jury

M. Xxx XXX

Mme. Xxx XXX

---

## Dédicaces

*Au nom de Dieu, le Tout Miséricordieux, le Très Miséricordieux*

*Je dédie ce travail*

*A ma chère mère,*

*A mon cher père,*

*Qui n'ont jamais cessé de faire des prières à mon égard de me soutenir  
Et de leurs encouragements pour que je puisse atteindre mes objectifs.*

*A mes frères, Hamza, Omar et Hatim*

*Pour leur soutien moral et leurs conseils précieux tout au long de mes études.*

*A mon cher ami, Amine*

*Qui m'a supporté durant ces deux dernières années et  
Chez qui j'ai trouvé l'entente dont j'avais besoin.*

*A mes chers ami, Yazid, Mohammed*

*Qui m'ont aidé plusieurs moments tout au long de ces années.*

*A mes chères ami(e)s, Imane, Souraya*

*Pour leurs aides et support durant les moments difficiles.*

*A toute ma famille*

*A tous ce que j'aime et ceux qui m'aiment.*

---

## Remerciements

*Au terme de ce projet, Je tiens à exprimer ma gratitude envers tous ceux qui ont assisté de près ou de loin à sa réalisation.*

*Je remercie ardemment mon encadrant à la faculté des sciences Dhar El Mahraz de Fès (Maroc) Monsieur le professeur **Ali YAHYAOUY** pour la qualité inédite de son suivi et la pertinence de ses conseils.*

*Je tiens à remercier chaleureusement Monsieur **Jérôme Darmont** chef du projet BI4people, Madames **Cécile Favre** et **Sabine Loudcher** pour m'avoir impliqué dans ce projet et pour le savoir-faire qu'elles m'ont transférés entant qu'encadrantes.*

*Je tiens à remercier toute personne ayant contribué à la réussite de ce projet, **M. Fahad Muhammad** Post-Doctorant ANR au laboratoire ERIC de son aide toutes au long de cette période de stage.*

*Mes remerciements s'adressent également à tous mes enseignants et aux cadres pédagogiques de l'Université Sidi Mohammed Ben Abdellah surtout de la **Faculté des Sciences Dhar el Mahraz de Fès, Maroc** pour les efforts déployés dans le but de développer toutes nos compétences.*

***J'adresse mon vif remerciement à :***

*Monsieur le Professeur **Jaouad BOUMHIDI**, Professeur à la Faculté Des Sciences Dhar El Mahraz de Fès (Maroc) et coordonnateur du master **INFORMATIQUE DECISIONNELLE ET VISION INTELLIGENTE** pour son support durant toutes ces années d'études entant que professeur et coordonnateur.*

*Monsieur le professeur **Julien Jacques**, Professeur à l'université Lumière Lyon 2 et directeur du laboratoire ERIC.*

*Madame **Habiba Osman**, Responsable Administrative et Financière du LABORATOIRE ERIC de l'université Lyon 2 pour son soutien durant cette période. A toutes les membres du laboratoire ERIC surtout celles de L'équipe SID (Systèmes d'Information Décisionnels) pour tous leurs aides.*

---

## Résumé

*L'informatique décisionnelle (BI) est un processus de gestion des données qui combine la collecte et le stockage des données, la gestion des connaissances et l'analyse afin d'alimenter le processus décisionnel.*

*La BI met l'accent sur l'analyse de grands volumes de données sur l'entreprise et ses opérations. Elle comprend la veille concurrentielle (surveillance des concurrents) en tant que sous-ensemble. Dans les environnements informatiques, la veille stratégique utilise une grande base de données, généralement stockée dans un entrepôt de données (Data Warehouse/Data Mart), comme source d'information et comme base d'une analyse sophistiquée. Le principal but de ce projet est de rendre ces techniques d'analyses accessible par toutes et tous, d'où vient le nom du projet BI4people et qui veut dire business intelligence pour tous, qui va permettre une prise en main simplifiée du processus décisionnel, en masquant les phases d'intégration de données et de conception d'un entrepôt à travers des techniques de CBI (Collaborative Business Intelligence). C'est sur cela que portera ce projet, en étalant toutes les étapes de conception du système à travers un choix des outils et une mise en place ainsi que des tests et des résultats.*

**Mots clés :** *Business Intelligence, Données, Entrepôt de données, Intégration des données, Extraction des données, Chargement des données, Data Mart, Collaborative Business Intelligence.*

---

## Abstract

*Business Intelligence (BI) is a data management process that combines data collection and storage, knowledge management, and analysis to feed the decision-making process.*

*BI focuses on the analysis of large volumes of data about the company and its operations. It includes competitive intelligence (competitor monitoring) as a subset.*

*In IT environments, BI uses a large database, usually stored in a (Data Warehouse/Data Marts), as a source of information and as a basis for sophisticated analysis.*

*The main goal of this project is to make these analysis techniques accessible to everyone, hence the name of the BI4people project, which means business intelligence for all, which will allow a simplified handling of the decision-making process, by masking the phases of data integration and design of a warehouse through CBI techniques (Collaborative Business Intelligence). This is what this project will focus on, spreading out all the stages of system design through a choice of tools and implementation as well as tests and results.*

**Keywords:** Business Intelligence, Data, Data warehouse, Data integration, Data mining, Data loading, Data Mart, Collaborative Business Intelligence

## Table des matières

|   |    |
|---|----|
| Dédicaces .....   | 2  |
| Remerciements .....   | 3  |
| Résumé .....  | 4  |
| Abstract .....  | 5  |
| Tables des figures.....                                     | 8  |
| Introduction .....  | 10 |
| 1. Environnement et contexte de stage.....                  | 11 |
| 1.1. L'unité de recherche ERIC : .....                      | 11 |
| 1.2. Projet ANR BI4people « Le décisionnel pour tous »..... | 12 |
| 1.2.1. Mode de travail : .....                              | 12 |
| 1.2.2. Descriptif et objectif : .....                       | 12 |
| 1.2.3. Partenaires :.....                                   | 13 |
| ❖ Partenaires académiques : .....                           | 13 |
| ❖ Partenaires Industriel : .....                            | 13 |
| ❖ Partenaires associés : .....                              | 14 |
| 1.2.4. Plateforme du projet « BI en libre-service » : ..... | 14 |
| 1.2.5. Missions du stage : .....                            | 17 |
| 2. L'informatique décisionnelle : .....                     | 18 |
| 2.1. Introduction : .....                                   | 18 |
| 2.2. Concepts généraux du BI : .....                        | 18 |
| 2.2.1. La Business Intelligence .....                       | 18 |
| 2.2.2. Avantages du BI : .....                              | 18 |
| 2.2.3. Les limites du BI : .....                            | 18 |
| 2.3. Les principes des systèmes décisionnels : .....        | 19 |
| 2.3.1. Sources de données : .....                           | 19 |
| 2.3.2. Entrepôt de données : .....                          | 20 |
| 2.3.3. Extract-Transform-Load : .....                       | 20 |
| 2.3.4. La différence entre OLTP et OLAP : .....             | 20 |
| 2.3.5. Conclusion : .....                                   | 21 |
| 3. Etat de l'art : .....                                    | 22 |

|        |  |    |
|--------|--|----|
| 3.1.   | Concepts de base.....  | 22 |
| 3.2.   | Principes de conception de La CBI :.....                         | 23 |
| 3.3.   | Phases du processus de prise de décision en collaboration :..... | 23 |
| 3.4.   | Différents aspects de la recherche en BI collaborative :.....    | 24 |
| 3.5.   | Conclusion : .....   | 25 |
| 4.     | Différents cas pour l'analyse collaborative des données :.....   | 25 |
| 4.1.   | Types de données disponibles :.....                              | 25 |
| 4.2.   | Objectif : .....   | 26 |
| 4.3.   | Conclusion : .....   | 27 |
| 5.     | Réalisation du stage :.....                                      | 27 |
| 5.1.   | Etude de cas sur la scolarisation en France :.....               | 27 |
| 5.1.1. | Implémentation d'ETL :.....                                      | 27 |
| 5.1.2. | Base de données utilisée : .....                                 | 27 |
| 5.1.3. | Etude des sources de données et type de chargement :.....        | 29 |
| 5.1.4. | Visualisations des données : .....                               | 37 |
| 5.2.   | Dashboard bilingues (Fr, En) :.....                              | 44 |
| 5.2.1. | L'enjeux de multilingues : .....                                 | 44 |
| 5.2.2. | Choix de l'outil : .....   | 44 |
| 5.2.3. | Réagir avec React-i18next :.....                                 | 44 |
| 5.2.4. | Résultat : .....   | 47 |
| 5.2.5. | Conclusion :.....  | 47 |
|        | Conclusion.....  | 48 |
|        | Bibliographie .....  | 49 |

# Tables des figures

|   |    |
|---|----|
| Figure 1: Axes de recherche d'ERIC .....  | 11 |
| Figure 2: Schéma global de fonctionnement de BI4people .....                        | 12 |
| Figure 3 : les outputs des Workpackage .....  | 15 |
| Figure 4 : Architecture générale d'un système décisionnel.....                      | 19 |
| Figure 5 : La différence entre OLTP et OLAP .....                                   | 21 |
| Figure 6: Différentes formes de la Business Intelligence Collaborative .....        | 22 |
| Figure 7: Phases du processus de prise de décision en collaboration.....            | 24 |
| Figure 8: Vue générale de la base source (Excel).....                               | 28 |
| Figure 9 : Dictionnaire des variables.....  | 28 |
| Figure 10 : Schéma de notre source des données .....                                | 29 |
| Figure 11 : Diagramme UML du processus d'alimentation.....                          | 30 |
| Figure 12 : Connexion à Mariadb et création de la base datawarehouse_BI4people..... | 31 |
| Figure 13 : Structure après la division de chaque colonne par ses racines .....     | 32 |
| Figure 14 : Création des tables de dimensions .....                                 | 33 |
| Figure 15 : Création de la table de fait.....                                       | 33 |
| Figure 16 : Insertion dans la table de fait .....                                   | 34 |
| Figure 17 : Quelques visualisations pour le test avec Power BI.....                 | 37 |
| Figure 18 : Architecture fonctionnelle de Cube Js .....                             | 38 |
| Figure 19 : Contenu du fichier .env de notre projet .....                           | 39 |
| Figure 20 : Génération du Schéma.....   | 39 |
| Figure 21 : Démarage du dashboard.....  | 41 |
| Figure 22 : Interface Dashboard avec React .....                                    | 41 |
| Figure 23 : Visualisation sous forme d'une Table .....                              | 41 |
| Figure 24 : Visualisation sous forme de Bar .....                                   | 42 |
| Figure 25 : Visualisation sous forme de Pie chart .....                             | 42 |
| Figure 26 : Enregistrement du Dashboard .....                                       | 42 |
| Figure 27 : Vue des visualisations enregistrer .....                                | 43 |
| Figure 28 : Modification des visualisations après la création .....                 | 43 |
| Figure 29 : Dictionnaire des données en JSON .....                                  | 45 |

## Tableau d'abréviation

|                  |  |
|------------------|--|
| <b><i>BI</i></b> | Business Intelligence  |
| <b>CBI</b>       | Collaborative business intelligence                          |
| <b>LLSHS</b>     | Domaines des lettres, langues, sciences humaines et sociales |
| <b>OLAP</b>      | Online Analytical Processing                                 |
| <b>CI</b>        | Communication interne  |
| <b>SBI</b>       | Social Business Intelligence                                 |
| <b>GDM</b>       | Group Decision Making  |
| <b>ETL</b>       | Extract-Transform-Load                                       |
| <b>ORM</b>       | Object Relational Mapping                                    |
| <b>OLTP</b>      | Online Transaction Processing                                |
| <b>UML</b>       | Unified Modeling Language                                    |
| <b>SSAS</b>      | Software-as-a-service  |
| <b>CDM</b>       | Collaborative Decision Making                                |
|                  |  |
|                  |  |
|                  |  |

# Introduction

Etudiant en deuxième année master INFORMATIQUE DECISIONNELLES ET VISION INTELLIGENTE au sein de la faculté des sciences Dhar El Mahraz de Fès, Maroc.

Je suis en période de stage de fin d'étude auprès du laboratoire ERIC,

Il s'agit d'une unité de recherche des Universités Lyon 2 et Lyon 1, c'est un stage pour une durée 18 semaines, du 23 mars au 29 juillet 2022.

Ce stage a pour but de confirmer mon diplôme de master qui est orienté vers la recherche.

La mission de stage que l'organisme d'accueil m'a proposé, en plus d'être cohérent avec le programme d'étude suivi tout au long de ma formation correspond justement à cet objectif, c'est pourquoi j'ai choisi cette unité de recherche.

De plus, le secteur de recherche vers lequel il est tournée, à savoir l'informatique décisionnelle, la science des données, m'a semblé très intéressant.

Nous allons commencer ce rapport de stage par une présentation de l'organisme d'accueil qui m'a accueillie : ERIC, en décrivant le secteur d'activité de recherche.

## Plan du mémoire

Ce mémoire s'articule de la manière suivante.

- La section 2 introduit les concepts généraux (les avantages et les limites, etc...) de la Business intelligence
- La section 3 présente un peu d'histoire sur la littérature de la CBI (Collaborative Business intelligence)
- La section 4 présente les différents cas de l'analyse collaboratives des données, les types des données possible en indiquant notre choix de type des données ainsi
- La section 5 présente les réalisations du stage et enfin une conclusion s'impose.

## 1. Environnement et contexte de stage

Nous allons débiter cette première partie en présentant l'unité de recherche qui m'a accueilli pour effectuer ce stage ainsi que son secteur d'activité. Nous décrivons également le projet ANR BI4people « Le décisionnel pour tous » et les partenaires associés ainsi que ma mission de stage.

### 1.1. L'unité de recherche ERIC :

Le laboratoire ERIC (Unité de recherche des Universités Lyon 2 et Lyon 1) développe des recherches théoriques et appliquées dans les domaines de la science des données et de l'informatique décisionnelle. Elles visent à valoriser les grandes bases de données complexes, notamment dans les domaines des lettres, langues, sciences humaines et sociales (LLSHS) et se situent dans les domaines suivants :

- Science des données : machine Learning, prévision, décision, fouille de données complexes, textuelles ou fonctionnelles, agrégation multicritère.
- Informatique décisionnelle : entrepôts de données, intégration intelligente de données complexes, modélisation multidimensionnelle d'objets complexes, analyse en ligne (OLAP) personnalisée, lacs de données, gestion de métadonnées, sécurité des données.
- Humanités numériques, gestion et analyse de données variées (au sens des mégadonnées), métadonnées, interprétabilité des modèles de fouille de données, informatique et genre.

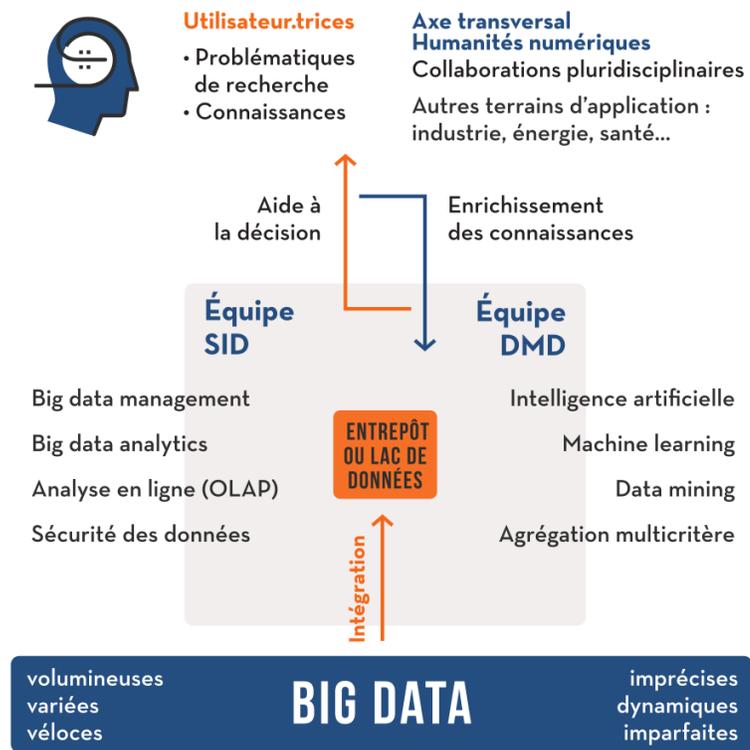


Figure 1: Axes de recherche d'ERIC

## 1.2. Projet ANR BI4people « Le décisionnel pour tous »

### 1.2.1. Mode de travail :

Suite à des problèmes administratifs je suis arrivé plus tard, j'ai commencé le stage en télétravail, j'ai participé à une réunion plénière qui regroupe tous les partenaires en distanciel le 24 mars et une autre en présentiel le 23,24 juin à la MSH de Lyon, ça été l'occasion de voir comment se déroule le projet et de se mettre d'accord avec les différents workpackage sur l'avancement du projet ainsi y'avais des réunions hebdomadaires qui m'ont données l'occasion de poser des questions et savoir pas mal des choses.

### 1.2.2. Descriptif et objectif :

Le descriptif et l'objectif présenté ci-dessous sont reprise du site web du projet dans la référence est la suivante : [Projet ANR BI4people - Le décisionnel pour tou.les \(univ-lyon2.fr\)](http://Projet ANR BI4people - Le décisionnel pour tou.les (univ-lyon2.fr))

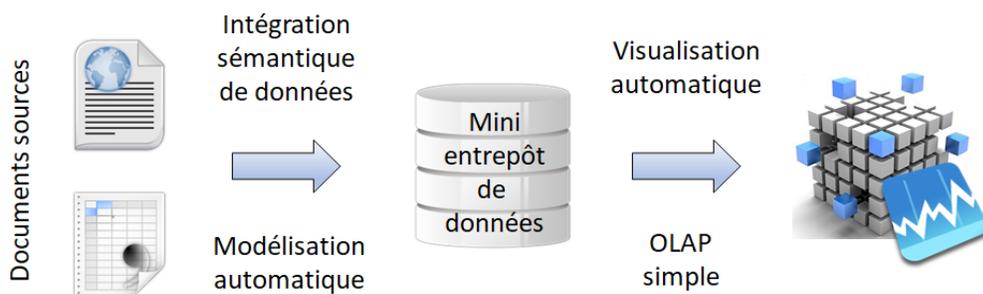
Les technologies de l'informatique décisionnelle (business intelligence ou BI), telles que les entrepôts de données OLAP, sont des outils primordiaux dans l'aide à la décision qui ont longtemps nécessité un investissement financier et humain très lourd.

Toutefois, il existe désormais de nombreuses solutions de BI gratuites, qu'elles soient propriétaires, libres ou infonuagiques. Les logiciels propriétaires se focalisent cependant sur les tableaux de bord et la visualisation, et ont tous des fonctionnalités limitées, comme l'absence d'une intégration efficace de données depuis des sources disparates.

De plus, bien que quelques logiciels libres proposent des explorations OLAP, ils demeurent techniquement hors de portée des petites entreprises, des chercheurs et des indépendants comme des journalistes, que nous ciblons particulièrement dans ce projet.

Finalement, la tendance à déporter la BI dans le nuage rejoint la demande grandissante d'outils collaboratifs qui permettent aux utilisateurs de croiser des données privées, publiques, ainsi que des self data, d'effectuer des analyses conjointes, d'annoter des figures ou des rapports et de communiquer via les réseaux sociaux. Les réponses actuelles à cette demande globale restent en deçà des attentes en se limitant au partage en ligne de résultats d'analyse.

L'objectif de BI4people est de rendre accessible la puissance de l'analyse interactive OLAP à la plus large audience possible, en mettant en œuvre le processus d'entreposage de données en mode SSAS, de l'intégration de données multi source, hétérogènes (typiquement sous la forme de tableaux issus de tableurs, de documents textuels ou semi-structurés, ou encore du Web) à une analyse OLAP et une visualisation très simple. Pour atteindre ce but, le service de BI doit inclure la privacy by design, être autonome, extrêmement simple, ergonomique et intelligible (jargon informatique ou BI interdit !). Dans ce contexte, les étapes classiques de l'entreposage de données s'appliquent mais doivent être complètement automatisées.



Relevant actuellement plutôt de l'apprentissage automatique. De plus, le prototype logiciel que nous proposons comme le livrable principal du projet prendra en charge la confidentialité des données dans toutes les étapes, permettra des analyses collaboratives et sera réellement intelligible par ses utilisateurs. Nous insistons en effet sur l'importance de l'appropriation des visualisations fournies par l'outil par les utilisateurs, ce qui implique une collaboration interdisciplinaire entre l'informatique et les sciences de l'information et de la communication.

### 1.2.3. Partenaires :

Le projet regroupe des partenaires reconnus dans les domaines mentionnés : Mon organisme d'accueil (**ERIC**) pour ses compétences en analyse OLAP, personnalisation de la BI et cryptographie ; l'IRIT pour ses compétences en modèles d'entrepôts des données, intégration de données et processus ETL, Le laboratoire **LIFAT** de Tours pour leurs compétences de visualisations des données et enfin **ELICO** pour ses compétences en études d'usage. S'ajoutent à ce consortium, l'entreprise **TRIMANE** qui possède un département R&D en Big Data et deux associations citoyennes qui militent dans le domaine des « données personnelles » : le **TUBA** à Lyon et la **FING (Fondation Internet Nouvelle Génération)**.

#### ❖ Partenaires académiques :



#### ❖ Partenaires Industriel :



❖ Partenaires associés :



#### 1.2.4. Plateforme du projet « BI en libre-service » :

##### 1.2.4.1. Objectif principal :

La business intelligence en self-service consiste à permettre aux utilisateurs non experts de prendre des décisions bien informées en enrichissant le processus de décision avec des données situationnelles, c'est-à-dire des données qui ont une focalisation étroite sur un problème métier spécifique sur une durée de vie courte pour un petit groupe d'utilisateurs.

- Le principal résultat de ce projet sera un prototype de logiciel open source supportant toutes les fonctionnalités mentionnées dans différents lots de travail
- Nous incluons BI4people en tant que plateforme BI en libre-service

##### 1.2.4.2. Environnement de travail :

Pour faciliter ce processus, BI4people sera conçu de manière modulaire et orientée service dès le départ.

La technologie utilisée est NodeJS, les scripts NodeJS relient divers modules écrits dans différents langages et assurent l'interopérabilité entre les différents WP (Work Packages).

##### 1.2.4.3. C'est quoi Node.JS ?



Node.js est un environnement d'exécution JavaScript open-source et multiplateforme :

- Node.js exécute le moteur JavaScript V8, le cœur de Google Chrome, en dehors du navigateur.
- Une application Node.js s'exécute dans un seul processus, sans créer un nouveau thread pour chaque requête.
- Node.js fournit un ensemble de primitives d'E/S asynchrones dans sa bibliothèque standard.

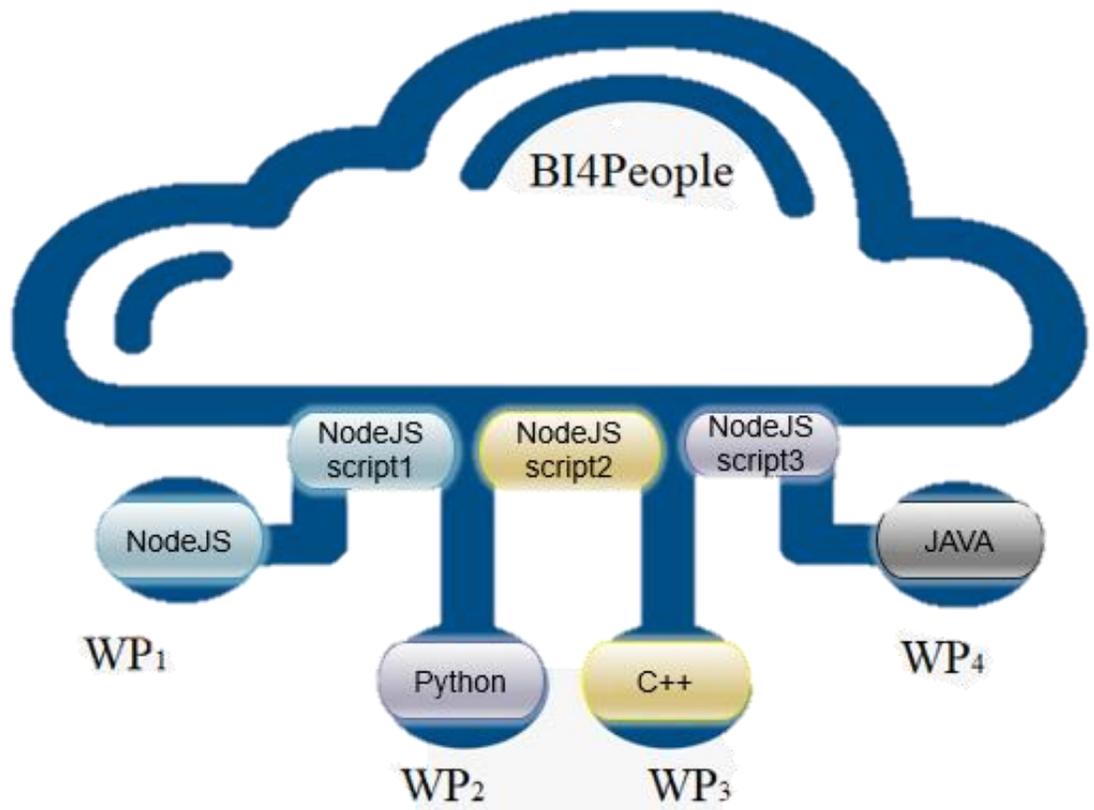
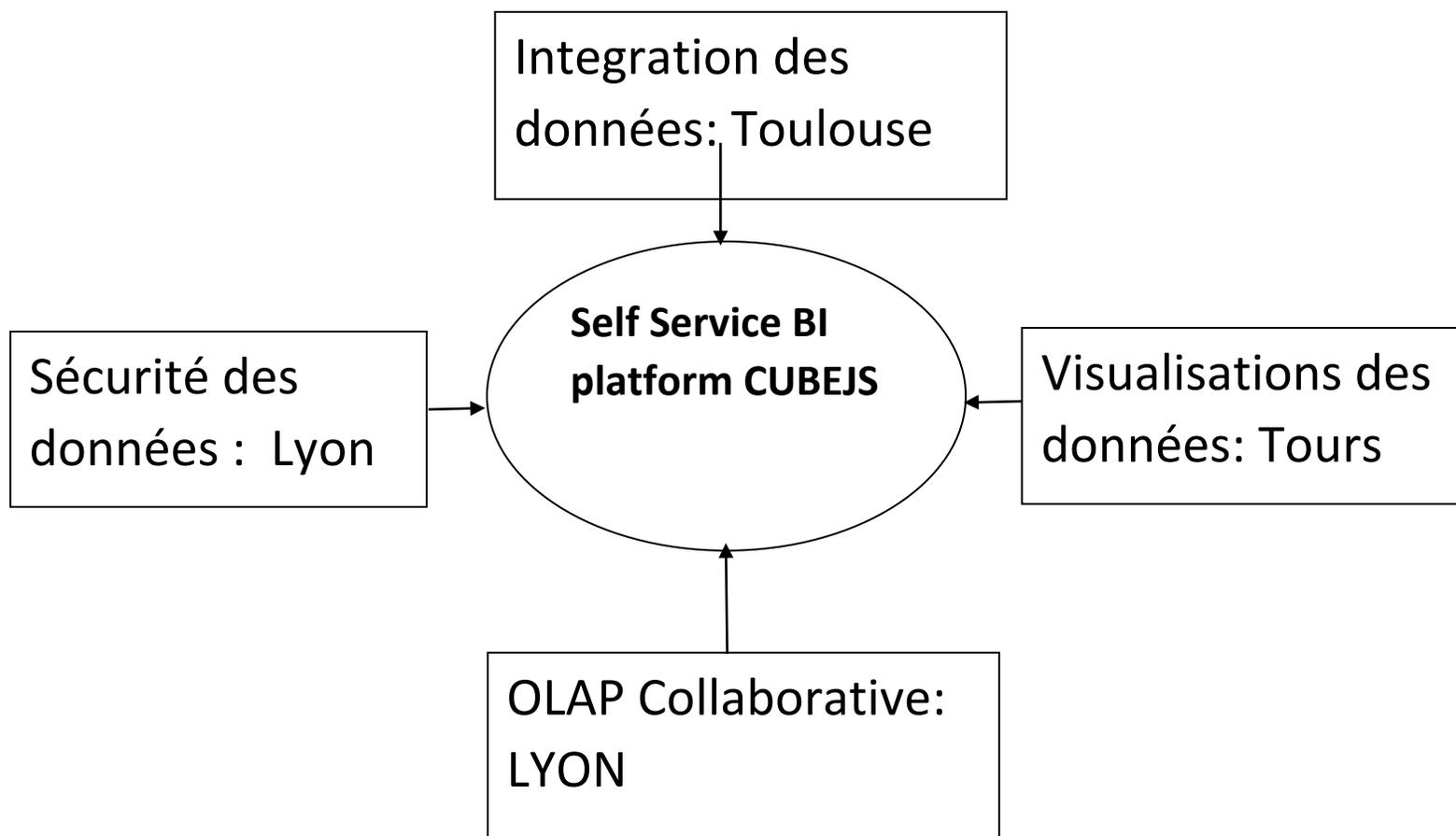


Figure 3 : les outputs des Workpackage



#### 1.2.4.4. *Front End des équipes de travail (WPs) :*

**NodeJS** peut se connecter avec n'importe quelle application frontale basée sur html5

**React - bibliothèque JavaScript** : Il s'agit d'une bibliothèque JavaScript open-source, front-end, pour la construction d'interfaces utilisateur ou de composants UI. Elle est maintenue par Facebook ...

Développer un module de la plateforme BI revient à développer une application JS frontale.

Tout ce dont vous avez besoin est NodeJS 10.

Chaque application fonctionnant dans la plateforme BI est une application HTML5 côté client qui interagit avec vos données par le biais de l'API du serveur.

#### 1.2.4.5. OLAP collaboratif : Extension de CubeJS

Cube est une plateforme de business intelligence basée sur des API, destinée aux ingénieurs de données et aux développeurs d'applications, qui permet de rendre les données accessibles et cohérentes dans toutes les applications.

Cubejs : <https://cube.dev/>

#### 1.2.5. Missions du stage :

La mission de stage que l'organisme d'accueil m'a proposé, en plus d'être cohérent avec le programme d'études suivies tout au long de ma formation, correspond justement à cet objectif, c'est pourquoi j'ai choisi cette unité de recherche. De plus, le secteur de recherche vers lequel il est tournée, à savoir l'informatique Décisionnelle, la science des données, m'a semblé très intéressant. J'ai donc pour mission principale, premièrement de voir l'état de l'art sur la CBI (Collaborative business intelligence) et après une étude de cas sur la scolarisation en France et la création d'un script qui permet l'importation des données depuis une base de données Excel sous Maria DB prenant en compte les relations et les jointures entre toutes les tables de dimensions afin de faire des visualisations des données par la création d'un Dashboard avec React et après une phase de traduction du Dashboard afin de rendre notre Dashboard bilingues.

## 2. L'informatique décisionnelle :

### 2.1. Introduction :

Ce chapitre sera réservé pour définir l'informatique décisionnelle. Nous présentons dans un premier temps ses avantages et ses limites. Nous abordons, ensuite, ses termes et les concepts clés en détaillant la notion d'ETL et d'entrepôt de données.

### 2.2. Concepts généraux du BI :

#### 2.2.1. La Business Intelligence

L'informatique décisionnelle, également Business Intelligence ou BI en anglais, désigne les moyens, les méthodes et les outils qui apportent des solutions en vue d'offrir une aide à la décision aux professionnels afin de leur permettre d'avoir une vue d'ensemble sur l'activité de l'entreprise et de leur permettre de prendre des décisions plus avisées à travers des tableaux de bord de suivi et des analyses.

#### 2.2.2. Avantages du BI :

Déployer une solution BI apporte de nombreux avantages :

- Améliorer la visibilité sur les chiffres, les écarts et les anomalies.
- La combinaison de plusieurs sources de données (ERP, systèmes comptable, feuilles de calcul, des budgets ...).
- La présentation uniforme d'informations fiables.
- L'automatisation permettant l'accélération de la collecte et de la diffusion de l'information.
- La performance dans le calcul d'agrégats sur de gros volume de données.
- La prise de décision grâce à des indicateurs pertinents et à une structure cohérente des informations.
- La prise de décision grâce à des indicateurs pertinents et à une structure cohérente des informations.

#### 2.2.3. Les limites du BI :

Parmi les limites de la Business Intelligence :

- La mise en place d'une solution de BI prend beaucoup de temps : de nombreuses entreprises dans le scénario industriel rapide ne sont pas assez patientes pour attendre la mise en place du système décisionnel dans leur organisation.
- Complexité : un autre inconvénient de BI pourrait être sa complexité dans la mise en œuvre des données.

- Erreur : les résultats produits par les systèmes décisionnels sont le résultat de conceptions informatiques et mathématiques complexes, qui peuvent révéler des erreurs, par ailleurs les résultats sont souvent statistiques, donc non déterministes. La possibilité d'une erreur ou une approximation inadaptée devra toujours être prise en compte dans les décisions.

### 2.3. Les principes des systèmes décisionnels :

Le système décisionnel est architecturé de la façon suivante :

- Plusieurs sources de données en lecture
- Un entrepôt de données fusionnant les données requises.
- Un ETL permettant d'alimenter l'entrepôt de données à partir des données existantes.
- Des magasins de données permettant de simplifier l'entrepôt de données
- Des applications d'exploitation de données pour présenter l'étude aux utilisateurs finaux et décideurs.

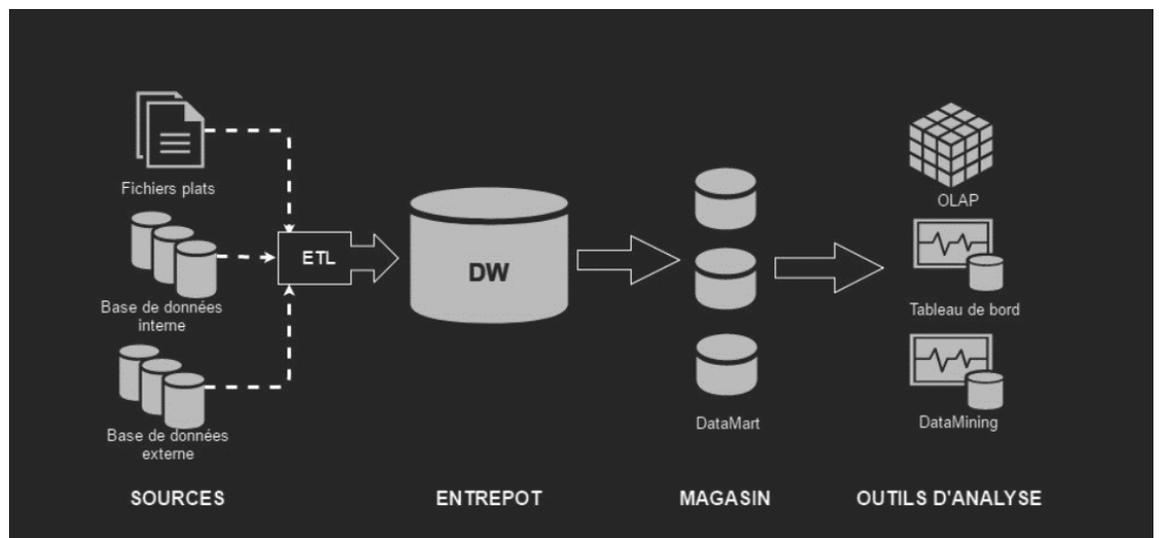


Figure 4 : Architecture générale d'un système décisionnel

#### 2.3.1. Sources de données :

Afin d'alimenter l'entrepôt, les informations doivent être identifiées et extraites de leurs emplacements originels. Il s'agit des sources de données hétérogènes qui peuvent comporter des données internes à l'entreprise, stockées dans les bases de données de production des différents services.

Elles peuvent être aussi des sources externes, récupérées via des services distants et des web services ou des sources qui peuvent être sous format de fichier plats comme le cas de notre projet.

### 2.3.2. Entrepôt de données :

D'après **BILL Inmon** : "Un entrepôt de données est une collection de données thématiques, intégrées, non volatiles et historiques, organisées pour la prise de décision"

D'après cette définition nous distinguons les caractéristiques suivantes :

- Données Orientées sujet : les données des entrepôts sont organisées par sujet et donc triées par thème.
- Données intégrées : les données provenant des différentes sources doivent être intégrées avant leur stockage dans l'entrepôt de données. Un nettoyage préalable des données est nécessaire afin d'avoir une cohérence et une normalisation de l'information.
- Données non-volatiles : à la différence des données opérationnelles, celles de l'entrepôt sont permanentes et ne peuvent pas être modifiées. Le rafraîchissement de l'entrepôt, consiste à ajouter de nouvelles données sans perdre celles qui existent.
- Historiées : les données doivent être datées.

### 2.3.3. Extract-Transform-Load :

ETL acronyme d'extraction transformation loading est un processus d'intégration des données. Il permet de transférer des données brutes d'un système source, de les préparer pour une utilisation et de les envoyer vers l'entrepôt de données. Ce système doit faire passer les données par un tas de processus pour les dénormaliser, les nettoyer, les contextualiser, puis de les charger de la façon adéquate. Cependant, la réalisation de l'ETL est une étape très importante et très complexe parce qu'il constitue 70% d'un projet décisionnel en moyenne.

### 2.3.4. La différence entre OLTP et OLAP :

Les systèmes informatiques peuvent se subdiviser en deux catégories : **Les systèmes transactionnels OLTP** (Online Transaction Processing) et **les systèmes analytiques OLAP** (Online Analytical Processing) :

- Les systèmes OLTP sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes et donc directement opérationnels, le mode de travail est transactionnel, l'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base. Ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs simultanément. Il est proposé essentiellement pour les applications gérant des opérations commerciales comme les opérations bancaires, ou l'achat de bien divers.
- Les systèmes OLAP sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité. C'est un outil de reporting dont la couche d'analyse permet de générer des résultats en fonction du contenu d'un entrepôt de données. Les

programmes consultent une quantité importante de données pour procéder à des analyses. Les objectifs principaux sont : regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour permettre à l'utilisateur de retrouver et analyser l'information facilement et rapidement.

Bien que les systèmes d'informations OLTP et OLAP aient le point commun de regrouper les données de l'entreprise dans un SGBD et d'en fournir l'accès aux utilisateurs, ils présentent de profondes différences :

| Caractéristiques                  | OLTP  | OLAP   |
|-----------------------------------|---|--|
| Utilisation                       | SGBD base de production                     | Entrepôt de données                                      |
| Opération typique                 | Mise à jour                                 | Analyse  |
| Type d'accès                      | Lecture écriture                            | Lecture  |
| Taille BD                         | Faible (max quelque GB)                     | Importante (pouvant aller à plusieurs TB)                |
| Requête                           | Simple, régulières, répétitives, nombreuses | Complexes, irrégulières, peu nombreuses, non prévisibles |
| Quantité d'informations échangées | Faible                                      | Important  |
| Orientation                       | Ligne                                       | Multi dimension  |
| Structure de données              | Beaucoup de tables                          | Peu de table mais de grande taille                       |
| Ancienneté des données            | Récente                                     | Historique   |

Figure 5 : La différence entre OLTP et OLAP

### 2.3.5. Conclusion :

Dans ce chapitre, nous avons détaillé toutes les notions relatives aux systèmes décisionnels, pour les maîtriser afin de favoriser le bon déroulement du projet et puis nous revenons sur nos recherches à propos de la littérature de la BI collaborative qui sera représenté dans le chapitre suivant.

### 3. Etat de l'art :

#### 3.1. Concepts de base

Voici quelques concepts de base sur la Business Intelligence collaborative (CBI en anglais) qui sont très utiles et favorisent la compréhension sur la problématique.

Selon (Kaufmann et Chamoni, 2014), Les CBI peuvent prendre plusieurs formes.

Il peut s'agir d'une discussion générale où les personnes au sein des entreprises génèrent plusieurs rapports basés sur certaines discussions où des sujets généraux où il peut s'agir d'une discussion plus centrée sur le rapport qui se concentre sur les commentaires et les commentaires sur un rapport particulier.

D'autres formes de CBI peuvent être considérées comme l'ajout d'annotations à des éléments spécifiques dans un rapport où de nombreuses personnes ajoutent leur analyse ou leurs commentaires en fonction de leurs domaines d'expertise. Il peut s'agir d'analyses de rapports et de visualisation du comportement des clients, d'un support de partage d'informations et d'un support de coordination des tâches, etc. La figure 1 illustre diverses formes de CBI.

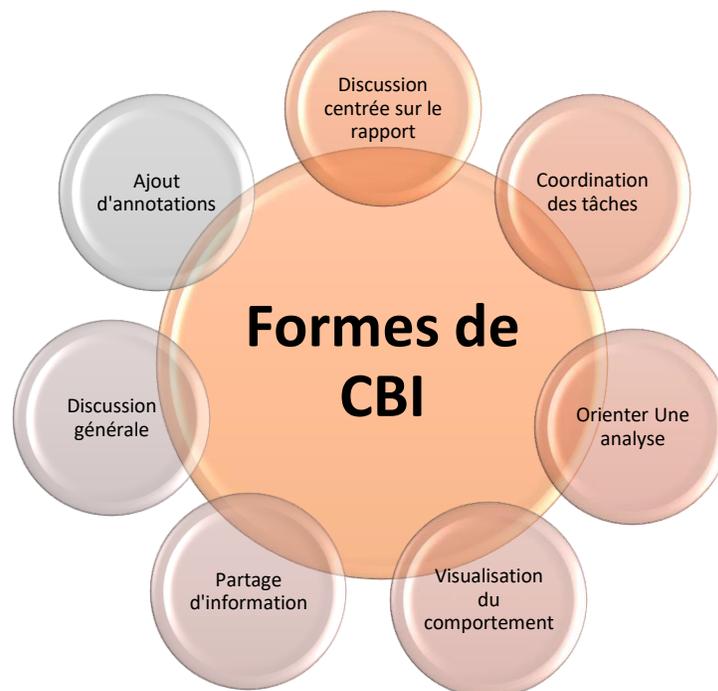


Figure 6: Différentes formes de la Business Intelligence Collaborative

Une approche collaborative de l'analytique rassemble les compétences, l'expérience et les connaissances de personnes occupant des rôles variés, ce qui apporte des avantages à l'organisation et aux membres individuels de l'équipe.

### **3.2. Principes de conception de La CBI :**

Kretzer et al. (2015) ont conçu des principes pour la diffusion des rapports et l'utilisation innovante des plateformes de Business Intelligence. Pour la diffusion des rapports, ils suggèrent que la plate-forme de BI devrait proposer que le rapport à un nouvel utilisateur potentiel soit basé sur Ses préférences

- La contagion des utilisateurs précédents de ce rapport.
- L'équivalence des rôles entre l'utilisateur et les utilisateurs précédents de ce rapport.
- La cohésion sociale entre l'utilisateur et les utilisateurs précédents de ce rapport.

En plus de l'utilisation innovante, la plate-forme de BI doit offrir un bac à sable permanent aux utilisateurs finaux dans lequel ils peuvent charger les données de la plate-forme de BI, les modifier et les enrichir avec des données provenant de sources de données externes. Leurs principes de conception basés sur des études de cas et des recherches existantes sont précieux pour les gestionnaires BI ainsi que pour les chercheurs BI.

### **3.3. Phases du processus de prise de décision en collaboration :**

Adla et al. (2010) ont proposé un modèle en trois phases du processus de prise de décision en collaboration (c'est-à-dire la phase de pré-décision, la phase de décision et la phase de post-décision).

Dans la phase de pré-décision, tous les participants doivent partager une compréhension commune du problème et des objectifs à atteindre. Cela permet d'obtenir une représentation commune du problème selon différents points de vue et de définir les limites et les frontières du problème.

La phase de décision se déroule en quatre étapes :

1. Les idées sont générées par tous les participants et leur permettent de comparer leurs idées.
2. Ces idées sont organisées afin d'accroître leur visibilité et leur compréhension.
3. Les idées et solutions proposées sont comparées selon les différents points de vue dans le référentiel commun de la décision.
4. On identifie et publie les accords des participants et on choisit parmi toutes les solutions proposées.

Enfin, la phase post décisionnelle est très importante car elle permet le suivi d'une décision, qui consiste essentiellement à réaliser un plan afin de mettre en œuvre la décision prise.

La figure 3 illustre ces phases du processus de prise de décision en collaboration.

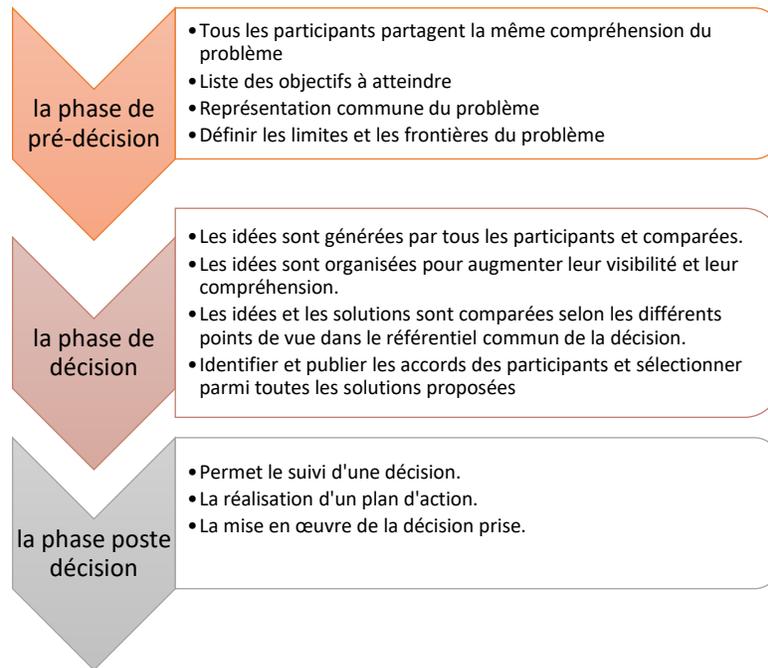


Figure 7: Phases du processus de prise de décision en collaboration

### 3.4. Différents aspects de la recherche en BI collaborative :

Il existe de nombreux aspects pour lesquels nous avons besoin de l'informatique décisionnelle collaborative tels que la formulation des requêtes, la recherche des sources, l'acquisition, l'intégration et la présentation des données comme le présentent Abelló et al. (2013).

La formulation de requête collaborative où l'utilisateur désigne une requête étendue avec une recherche par mot clé en se basant sur les actions précédentes des autres utilisateurs dans les mêmes cas ou des cas similaires.

Un autre aspect peut être la découverte collaborative de sources ou la recherche de données pertinentes de haute qualité sur la base des balises fournies par d'autres utilisateurs démontrant le contenu exact des sources. Lorsque les données souhaitées ne sont pas disponibles en ligne, il est possible de demander des données à la communauté, où d'autres personnes peuvent télécharger les données souhaitées qui seront utilisées par l'ensemble de la communauté, ce que l'on appelle également l'acquisition collaborative de données.

Après cette phase, il peut arriver qu'un utilisateur souhaite ajouter une dimension géographique au cube de fusion. Cela peut se faire en réutilisant des parties existantes disponibles en ligne ou en collaborant avec d'autres utilisateurs, ce que l'on appelle également l'intégration collaborative des données.

Les applications de Business Intelligence utilisent des données recueillies à partir d'un entrepôt de données (DataWarehouse) ou d'un data Mart et les concepts de BI et de DW se combinent où l'entrepôt de données contient une copie des données analytiques qui facilitent l'aide à la décision.

Par conséquent, nous étudions les applications existantes de la BI pour le DW dans notre revue de la littérature.

Il existe déjà un document d'étude sur des sujets similaires à l'IBC, à savoir la prise de décision collaborative (CDM) et la Social Business Intelligence (SBI) (Kaufmann et Chamoni, 2014).

Leur étude porte davantage sur les approches CDM, qui comprennent le CBI et le Group Decision Making (GDM), et le SBI, également appelé Social Media Analytics ou Social Analytics.

Le SBI traite de la demande d'intelligence sociale, suit l'impact des activités sociales sur l'organisation et favorise la compréhension de la façon dont les gens pensent et agissent. Dans leur document de recherche, ils identifient les principaux concepts et domaines de l'IBC et répondent à deux questions majeures.

**Quelles sont les définitions des IBC et dans quelle mesure sont-elles populaires ? Quels aspects sont principalement abordés par les chercheurs et quels domaines sont prometteurs pour les recherches futures ?**

Ils décrivent l'IBC comme " une extension à dessein du processus d'analyse en BI en ajoutant des partenaires internes et externes et en ouvrant nécessairement l'accès aux données ".

Sur la base de cette définition, ils classent les travaux existants en trois grandes catégories. Premièrement, la plupart des systèmes existants abordent les défis de recherche concernant la communication interne (CI). Deuxièmement, les travaux de recherche abordent les défis relatifs au partenariat en matière de données (PD), où les partenaires externes sont impliqués dans le processus de fourniture de données. Troisièmement, les travaux de recherche abordent les questions relatives au partenariat dans l'analyse (PA), où les partenaires travaillent ensemble dans le processus d'analyse des données. Dans notre enquête, nous ignorons complètement les travaux de recherche qui se concentrent sur la mesure des stratégies de BI sociale après les activités sociales dans les organisations. Il existe donc un besoin pour un nouvel état de l'art sur les approches et les techniques de CBI excluant SBI et CDM. Cela nous motive à combler cette lacune et à analyser les approches et les outils de CBI. Par conséquent, nous étudions en détail la littérature de recherche sur l'IBC en ce qui concerne non seulement les défis et les objectifs qu'ils atteignent, mais aussi sous divers aspects techniques, tels que l'évolutivité, la flexibilité, la configuration de nouveaux modèles, etc.

### **3.5. Conclusion :**

La Business Intelligence collaborative (CBI) est un concept qui implique l'intégration de la Business Intelligence (BI) et des outils technologiques collaboratifs afin d'aider une organisation à prendre des décisions commerciales nouvelles et améliorées, dans la section suivante on verra les types des données disponibles pour notre analyse ,ainsi nous faisons une étude pour voir le type des données qui sont les plus appropriées pour l'analyse collaborative.

## **4. Différents cas pour l'analyse collaborative des données :**

### **4.1. Types de données disponibles :**

- Données scientifiques
- Virus, maladie, gène ...
- Données générales collectées par les pouvoirs publics
- Consommation, Chômage, Emploi, (ex. données INSEE)
- Bases de données ouvertes comme Skyscraper
- Journaux de données et de requêtes



- Les données 2007, 2012, et 2017 sont diffusées selon la géographie en vigueur au 1er janvier 2020.
- En complément des données, la base comprend plusieurs onglets : la liste des variables, les formules de calcul et la documentation générale.
- Les bases en téléchargement sont au format : xls et csv.
- Niveau géographique : commune, arrondissement municipal.
- Résultats : les variables sont disponibles pour 2007, 2012 et 2017

### 4.3. Conclusion :

Un post Doctorant qui travaille dans le projet à pouvoir analyser différents types de données disponibles (scientifiques, gouvernementales, données d'enregistrement) et en conclusion il a pu voir que les données gouvernementales sont les plus appropriées pour l'analyse collaborative.

D'où nous avons conclu que les données gouvernementales sont plus appropriées à prendre en compte car :

- Elles présentent un niveau de complexité minimal
- Elles sont très compréhensibles et nécessitent très peu d'expertise pour être manipulées.
- Bien que certains ensembles de données soient plus ou moins sensibles, ils le sont beaucoup moins que d'autres (par exemple, les données médicales).
- Il a la capacité et suffisant pour l'utilisabilité et la représentation de notre étude de cas.

## 5. Réalisation du stage :

### 5.1. Etude de cas sur la scolarisation en France :

#### 5.1.1. Implémentation d'ETL :

L'alimentation de l'entrepôt de données est une phase cruciale dans un projet décisionnel, elle a pour objectif l'extraction, la transformation et le chargement des données depuis les systèmes sources vers l'entrepôt de données.

L'implémentation d'un processus ETL requies les étapes suivantes :

- **Etude des sources de données et type de chargement.**
- **Choix de l'architecture technique.**
- **Conception de processus de chargement.**

#### 5.1.2. Base de données utilisée :

La base utilisée c'est une base INSEE sur le Recensement de la population en France, l'INSEE c'est l'Institut national de la statistique et des études économiques, le recensement de la population permet de connaître la diversité et l'évolution de la

population de la France. L'Insee fournit ainsi des statistiques sur les habitants, leur nombre et leurs caractéristiques : répartition par sexe et âge...

Le recensement repose désormais sur une collecte d'informations annuelles, concernant successivement tous les territoires communaux au cours d'une période de cinq ans. Les communes de moins de 10 000 habitants réalisent une enquête de recensement portant sur toute la population, à raison d'une commune sur cinq chaque année.

Figure 8: Vue générale de la base source (Excel)

Comme entrée un fichier Excel sous le nom : base-cc-diplomes-formation-2017-2012-2007.xlsx qui contient plusieurs colonnes et lignes, nous disposons d'un dictionnaire pour bien comprendre les significations des colonnes

Figure 9 : Dictionnaire des variables

### 5.1.3. Etude des sources de données et type de chargement :

#### 5.1.3.1. Modèle flocon de neige :

C'est une manière de relier une dimension à un fait dans un entrepôt de données. C'est le modèle en étoile avec une normalisation des dimensions. Il peut exister des hiérarchies des dimensions pour diviser les tables de dimensions lorsqu'elles sont trop volumineuses.

#### 5.1.3.2. L'étude des sources des données :

L'étude des sources de données représente une phase primordiale pour le processus de l'ETL, elle consiste à l'identification des données sources pour l'alimentation de notre entrepôt.

En ce qui concerne notre projet, La source de données pour notre entrepôt est une base de données sous Mariadb : « datawarehouse\_BI4people » qui contient une table de faits « fact\_population » et 8 dimensions : « dim\_annee », « dim\_catDiplome », « dim\_sexe », « dim\_scolarisation », « dim\_trancheage », « dim\_commune », « dim\_departement », « dim\_region ».

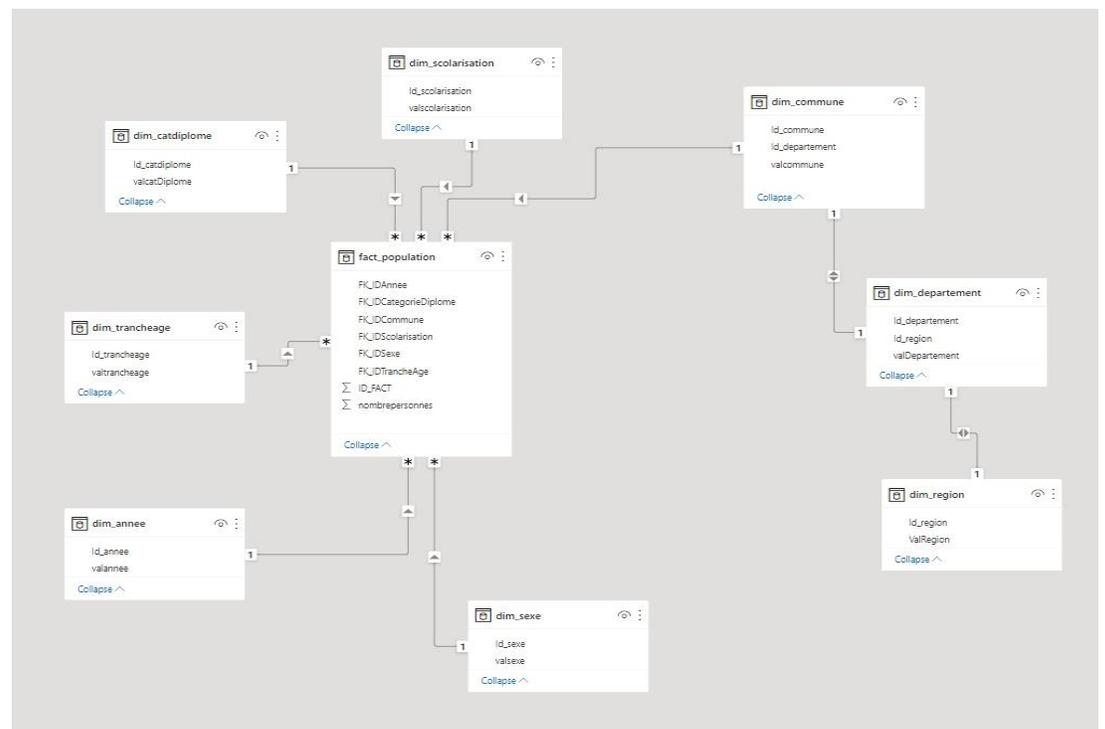


Figure 10 : Schéma de notre source des données

### 5.1.3.3. Stratégie du chargement des données :

Les différents datamarts de notre entrepôt de données engendrent des chargements considérables, en termes de volume.

Les données liées à l'activité du service sont stockées au niveau du système transactionnel. Afin de capter les nouvelles données et les données modifiées, nous allons choisir un ETL incrémental, pour cela, dans notre entrepôt de données nous aurons :

- **Des mises à jour des dimensions.**
- **Des nouveaux faits.**
- **Le chargement de notre entrepôt de données est conditionné par l'inactivité du système sources, en effet nous allons fixer le temps de l'alimentation à minuit où l'activité du système est minimale.**

### 5.1.3.4. Processus d'alimentation :

Pour décrire le processus de l'alimentation de notre entrepôt de données nous allons employer le diagramme en utilisant le langage UML :

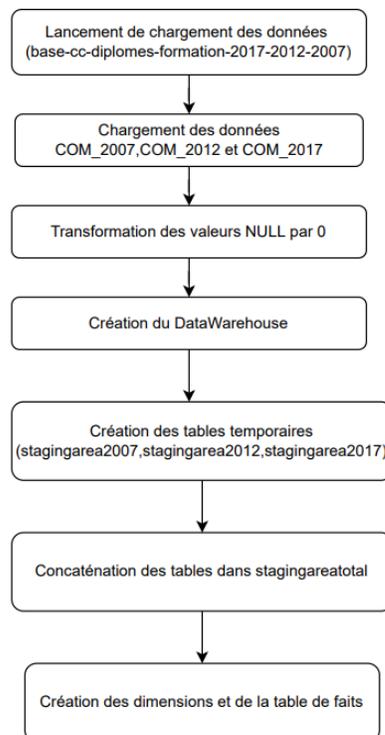


Figure 11 : Diagramme UML du processus d'alimentation

```

46 import mysql.connector as mariadb
47 import sys
48 use_pure = True
49
50
51 # Connect to MariaDB
52 try:
53     conn = mariadb.connect(
54         user="root",
55         password="123456789",
56         host="localhost",
57         port=3306
58     )
59 except mariadb.Error as e:
60     print(f"Error connecting to MariaDB Platform: {e}")
61     sys.exit(1)
62
63
64
65 cur = conn.cursor(buffered=True)
66
67 cur.execute("CREATE OR REPLACE DATABASE datawarehouse_BI4people")
68 cur.execute("use datawarehouse_BI4people")
69

```

Figure 12 : Connexion à Mariadb et création de la base datawarehouse\_BI4people

L'alimentation des différentes tables à savoir les dimensions et la table de faits est faite par un outil ETL développé afin de garantir une homogénéité technologique au niveau du SI du service.

```

if __name__ == "__main__":

    path = "base-cc-diplomes-formation-2017-2012-2007.xlsx"
    sheet_name_2007 = "COM_2007"
    df_com_2007 = read_sheet(path, sheet_name_2007)
    filtredData2007 = df_com_2007.filter(regex='^INTEGRER')
    filtredData2007 = filtredData2007.drop(0)
    filtredData2007.columns = filtredData2007.iloc[0]
    filtredData2007 = filtredData2007.drop(1)

    sheet_name_2012 = "COM_2012"
    df_com_2012 = read_sheet(path, sheet_name_2012)
    filtredData2012 = df_com_2012.filter(regex='^INTEGRER')
    filtredData2012 = filtredData2012.drop(0)
    filtredData2012.columns = filtredData2012.iloc[0]
    filtredData2012 = filtredData2012.drop(1)

    sheet_name_2017 = "COM_2017"
    df_com_2017 = read_sheet(path, sheet_name_2017)
    filtredData2017 = df_com_2017.filter(regex='^INTEGRER')
    filtredData2017 = filtredData2017.drop(0)
    filtredData2017.columns = filtredData2017.iloc[0]
    filtredData2017 = filtredData2017.drop(1)

    filtredData2007 = filtredData2007.fillna(0)
    filtredData2012 = filtredData2012.fillna(0)
    filtredData2017 = filtredData2017.fillna(0)

    print("dataframes created succesfully ...")

columnDict = {
    "P07_HSCOL0205": {"annee": "2007", "age": "02-05 ans", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL0610": {"annee": "2007", "age": "06-10 ans", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL1114": {"annee": "2007", "age": "11-14 ans", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL1517": {"annee": "2007", "age": "15-17 ans", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL1824": {"annee": "2007", "age": "18-24 ans", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL2529": {"annee": "2007", "age": "25-29 ans", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL30P": {"annee": "2007", "age": "30 ans ou plus", "sexe": "Homme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HSCOL0205": {"annee": "2007", "age": "02-05 ans", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HSCOL0610": {"annee": "2007", "age": "06-10 ans", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HSCOL1114": {"annee": "2007", "age": "11-14 ans", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HSCOL1517": {"annee": "2007", "age": "15-17 ans", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HSCOL1824": {"annee": "2007", "age": "18-24 ans", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HSCOL2529": {"annee": "2007", "age": "25-29 ans", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HSCOL30P": {"annee": "2007", "age": "30 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "non specifie"},
    "P07_FSCOL0205": {"annee": "2007", "age": "02-05 ans", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_FSCOL0610": {"annee": "2007", "age": "06-10 ans", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_FSCOL1114": {"annee": "2007", "age": "11-14 ans", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_FSCOL1517": {"annee": "2007", "age": "15-17 ans", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_FSCOL1824": {"annee": "2007", "age": "18-24 ans", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_FSCOL2529": {"annee": "2007", "age": "25-29 ans", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_FSCOL30P": {"annee": "2007", "age": "30 ans ou plus", "sexe": "Femme", "scolarite": "oui", "diplome": "non specifie"},
    "P07_HNSCOL0205": {"annee": "2007", "age": "02-05 ans", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL0610": {"annee": "2007", "age": "06-10 ans", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL1114": {"annee": "2007", "age": "11-14 ans", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL1517": {"annee": "2007", "age": "15-17 ans", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL1824": {"annee": "2007", "age": "18-24 ans", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL2529": {"annee": "2007", "age": "25-29 ans", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL30P": {"annee": "2007", "age": "30 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "non specifie"},
    "P07_HNSCOL15P_DIPLMIN": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "aucun diplome"},
    "P07_HNSCOL15P_BEPC": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "BEP, brevet des collèges"},
    "P07_HNSCOL15P_CAPBEP": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "CAP ou d'un BEP"},
    "P07_HNSCOL15P_BAC": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "un baccalauréat ou d'un brevet professionnel"},
    "P07_HNSCOL15P_BAC2": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "diplôme de l'enseignement supérieur court"},
    "P07_HNSCOL15P_SUP": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Homme", "scolarite": "non", "diplome": "diplôme de l'enseignement supérieur long"},
    "P07_HNSCOL15P_DIPLMIN": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "aucun diplome"},
    "P07_HNSCOL15P_BEPC": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "BEP, brevet des collèges"},
    "P07_HNSCOL15P_CAPBEP": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "CAP ou d'un BEP"},
    "P07_HNSCOL15P_BAC": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "un baccalauréat ou d'un brevet professionnel"},
    "P07_HNSCOL15P_BAC2": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "diplôme de l'enseignement supérieur court"},
    "P07_HNSCOL15P_SUP": {"annee": "2007", "age": "15 ans ou plus", "sexe": "Femme", "scolarite": "non", "diplome": "diplôme de l'enseignement supérieur long"}
}

```

## Extraction :

On a utilisé "base-cc-diplômes-formation-2007-2012-2017.xlsx" comme notre source données en utilisant 3 Sheets "COM\_2007", "COM 2012" et "COM 2017" et en filtrant juste les colonnes qui commencent par INTEGRER :

```
alldata = []
for k in range(len(desiredColumns)):
    cur.execute("""select CODGEO,REG,DEP,LIBGEO,"""+desiredColumns[k][0]+"""" from stagingarea2012 where reg=84""")
    currentRowtemp = cur.fetchall()
    alldata.append(currentRowtemp)

for i,field in enumerate(zip(currentRow)):

    print(i)
    for j,secondfield in enumerate(zip(columnDict.items())):
        print(j)

        CODEGEO = alldata[j][i][0]
        REG = alldata[j][i][1]
        DEP = alldata[j][i][2]
        LIBGEO = alldata[j][i][3]
        NOMBREPERSONNES = round(alldata[j][i][4])

        sexe = columnDict[desiredColumns[j][0]]['sexe']
        scolarite = columnDict[desiredColumns[j][0]]['scolarite']
        age = columnDict[desiredColumns[j][0]]['age']
        annee = columnDict[desiredColumns[j][0]]['annee']
        diplome = columnDict[desiredColumns[j][0]]['diplome']

    cur.execute("""
        INSERT INTO stagingareaTotal values(%s,%s,%s,%s,%s,%s,%s,%s,%s,%s)
        """, (CODEGEO,REG,DEP,LIBGEO,sexe,scolarite,age,annee,diplome,NOMBREPERSONNES))
    conn.commit()
    conn.start_transaction()
```

## Transformations:

- ✓ La modification de valeurs NULL par 0.
- ✓ Remplissage de tables "stagingarea2007","stagingarea2012" et "stagingarea2017" par les données désirées (via la source).
- ✓ La création d'un dictionnaire afin de transformer le modèle initial de la source vers le modèle de notre cas d'étude.
- ✓ On prend pour chaque ligne les colonnes suivants: (P07\_HSCOL0205,P07\_HSCOL0610, P07\_HSCOL1114,P07\_HSCOL1517,P07\_HSCOL1824,P07\_HSCOL2529,P07\_HSCOL30P,P07\_HNSCOL0205,P07\_HNSCOL 0610, P07\_HNSCOL1114,P07\_HNSCOL1517,P07\_HNSCOL1824,P07\_HNSCOL2529,P07\_HNSCOL30P,P07\_FSCOL0205,P07\_FSCOL0610, P07\_FSCOL1114,P07\_FSCOL1517,P07\_FSCOL1824,P07\_FSCOL2529,P07\_FSCOL30P,P07\_FNSCOL0205,P07\_FNSCOL06 10, P07\_FNSCOL1114,P07\_FNSCOL1517,P07\_FNSCOL1824,P07\_FNSCOL2529,P07\_FNSCOL30P,P07\_HNSCOL15P\_DIPLMI N,P07\_HNSCOL15P\_BEPC,P07\_HNSCOL15P\_CAPBEP,P07\_HNSCOL15P\_BAC,P07\_HNSCOL15P\_BACP2,P07\_HNSCO L15P\_SUP,P07\_FNSCOL15P\_DIPLMIN,P07\_FNSCOL15P\_BEPC,P07\_FNSCOL15P\_CAPBEP, P07\_FNSCOL15P\_BAC,P07\_FNSCOL15P\_BACP2,P07\_FNSCOL15P\_SUP)

Et on divise chaque colonne par ses racines, par exemple : la colonne P07\_HSCOL0205 on la transforme par le script ci-dessus vers une autre structure comme la suivante :

| CODGEO | REG | DEP | LIBGEO                  | SEXE  | SCOLARITE | AGE       | ANNEE | DIPLOME      | NOMBREPERSONNES |
|--------|-----|-----|-------------------------|-------|-----------|-----------|-------|--------------|-----------------|
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 02-05 ans | 2007  | non specifie | 20              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 06-10 ans | 2007  | non specifie | 34              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 11-14 ans | 2007  | non specifie | 22              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 15-17 ans | 2007  | non specifie | 19              |

Figure 13 : Structure après la division de chaque colonne par ses racines

## Chargement (Loading):

```

cur.execute("""
CREATE OR REPLACE TABLE `dim_region` (
  `id_region` int(11) NOT NULL AUTO_INCREMENT,
  `nomsregion` varchar(100) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci DEFAULT NULL,
  PRIMARY KEY (`id_region`),
  KEY `id_region` (`id_region`)
) ENGINE=InnoDB;
""")

cur.execute("""select DISTINCT REG from stagingareatotal""")
rows=cur.fetchall()
rowsreg=[0]
j=len(rows)
for i in range(len(rows)):
  cur.execute("""
INSERT INTO dim_region(valregion) values(%)
""", (rows[i][0],))
conn.commit()

cur.execute("""
CREATE OR REPLACE TABLE `dim_departement` (
  `id_departement` int(11) NOT NULL AUTO_INCREMENT,
  `nomsdepartement` varchar(100) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci DEFAULT NULL,
  `id_region` int(11) NOT NULL DEFAULT 0,
  foreign key(id_region) references dim_region(id_region),
  PRIMARY KEY (`id_departement`),
  KEY `id_departement` (`id_departement`)
) ENGINE=InnoDB;
""")

cur.execute("""select dep,dr_id_region from stagingareatotal st inner join dim_region dr on dr.valregion = st.reg
group by dep""")
rowsreg = cur.fetchall()
rowsreg[0]
j=len(rowsreg)
for i in range(len(rowsreg)):
  cur.execute("""
INSERT INTO dim_departement(valdepartement, id_region) values(%,%)
""", (rowsreg[i][0],rowsreg[i][1]))
conn.commit()

cur.execute("""
CREATE OR REPLACE TABLE `dim_sexe` (
  `id_sexe` int(11) NOT NULL AUTO_INCREMENT,
  `nomssexe` varchar(100) CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci DEFAULT NULL,
  PRIMARY KEY (`id_sexe`),
  KEY `id_sexe` (`id_sexe`)
) ENGINE=InnoDB;
""")

sexe = ["homme", "femme"]
for i in range(2):
  cur.execute("""
INSERT INTO dim_sexe(valsexe) values(%)
""", (sexe[i],))
conn.commit()

```

Figure 14 : Création des tables de dimensions

On remplit les dimensions et notre table de faits avec les données transformées en utilisant les jointures entre les différentes tables.

Ensuite en utilisant nos dimensions et stagingareaTotal on va merger les données pour extraire les identifiant pour remplir la table de faits 'fact\_population' :

```

print("create fact table population ...")

cur.execute("""
CREATE OR REPLACE TABLE `fact_population` (
  `ID_FACT` int(11) NOT NULL AUTO_INCREMENT,
  `FK_IDScolarisation` int(11) NOT NULL DEFAULT '0',
  constraint fk_scolarisation foreign key(FK_IDScolarisation)
references dim_scolarisation(Id_scolarisation),
  `FK_IDCommune` int(11) NOT NULL DEFAULT '0',
  constraint fk_commune foreign key(FK_IDcommune)
references dim_commune(Id_commune),
  `FK_IDAnnee` int(11) NOT NULL DEFAULT '0',
  constraint fk_annee foreign key(FK_IDannee)
references dim_annee(Id_annee),
  `FK_IDSexe` int(11) NOT NULL DEFAULT '0',
  constraint fk_sexe foreign key(FK_IDSexe)
references dim_sexe(Id_sexe),
  `FK_IDCategorieDiplome` int(11) NOT NULL DEFAULT '0',
  constraint fk_catdiplome foreign key(FK_IDCategorieDiplome)
references dim_catdiplome(Id_catdiplome),
  `FK_IDTrancheAge` int(11) NOT NULL DEFAULT '0',
  constraint fk_trancheage foreign key(FK_IDtrancheage)
references dim_trancheage(Id_trancheage),
  `nombrepersonnes` int(11) NOT NULL DEFAULT '0',
  PRIMARY KEY (`ID_FACT`)
)ENGINE=InnoDB;
""")

conn.commit()

```

Figure 15 : Création de la table de fait

```

cur.execute("set SESSION optimizer_search_depth=0;")

stagingareatotal = pd.read_sql("select * from stagingareatotal", conn)
dim_commune = pd.read_sql("select * from dim_commune", conn)
dim_departement = pd.read_sql("select * from dim_departement", conn)
dim_region = pd.read_sql("select * from dim_region", conn)
dim_scolarisation = pd.read_sql("select * from dim_scolarisation", conn)
dim_annee = pd.read_sql("select * from dim_annee", conn)
dim_sexe = pd.read_sql("select * from dim_sexe", conn)
dim_trancheage = pd.read_sql("select * from dim_trancheage", conn)
dim_catdiplome = pd.read_sql("select * from dim_catdiplome", conn)

dfest1 = pd.merge(stagingareatotal, dim_scolarisation,
                  how='inner', left_on=['SCOLARITE'],
                  right_on = ['valscolarisation'])

dfest2 = pd.merge(dfest1,dim_commune,
                  how='inner', left_on=['LIBGEO'],
                  right_on = ['valcommune'])

dfest3 = pd.merge(dfest2,dim_annee,
                  how='inner', left_on=['ANNEE'],
                  right_on = ['valannee'])
dfest4 = pd.merge(dfest3,dim_sexe,
                  how='inner', left_on=['SEXE'],
                  right_on = ['valsexe'])

dfest5 = pd.merge(dfest4,dim_catdiplome,
                  how='inner', left_on=['DIPLOME'],
                  right_on = ['valcatDiplome'])

dfest6 = pd.merge(dfest5,dim_trancheage,
                  how='left', left_on=['AGE'],
                  right_on = ['valtrancheage'])

```

```

for i in range(len(dfest6)):
    print(i)
    cur.execute("""
        INSERT INTO fact_population(FK_IDscolarisation,FK_IDCommune,FK_IDAnnee,FK_IDCategorieDiplome,
                                   FK_IDSexe,FK_IDTrancheAge,nombrepersonnes
                                   ) values(%s,%s,%s,%s,%s,%s,%s)
        """,(int(dfest6["Id_scolarisation"][i]),int(dfest6["Id_commune"][i]),int(dfest6["Id_annee"][i]),
            int(dfest6["Id_catdiplome"][i]),int(dfest6["Id_sexe"][i]),int(dfest6["Id_trancheage"][i]),
            int(dfest6["NOMBREPERSONNES"][i])))
    conn.commit()
    conn.start_transaction()

conn.close()

```

Figure 16 : Insertion dans la table de fait

### 5.1.3.5. Le chargement des tables de dimensions et la table de fait après le lancement du script :

Database: datawarehouse\_b4people | Table: stagingaretotal | 460,388 rows total (approximately), limited to 1000

| COGNEC | REC | DEP | LIBSEO                  | SEXE  | SCOLARITE | AGE            | ANNEE | DIPLOME      | NOMBREPERSONNES |
|--------|-----|-----|-------------------------|-------|-----------|----------------|-------|--------------|-----------------|
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 02-05 ans      | 2007  | non specific | 30              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 06-10 ans      | 2007  | non specific | 34              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 11-14 ans      | 2007  | non specific | 22              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 15-17 ans      | 2007  | non specific | 19              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 18-24 ans      | 2007  | non specific | 13              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 25-29 ans      | 2007  | non specific | 0               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | oui       | 30 ans ou plus | 2007  | non specific | 3               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 02-05 ans      | 2007  | non specific | 15              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 06-10 ans      | 2007  | non specific | 0               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 11-14 ans      | 2007  | non specific | 0               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 15-17 ans      | 2007  | non specific | 1               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 18-24 ans      | 2007  | non specific | 17              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 25-29 ans      | 2007  | non specific | 18              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Homme | non       | 30 ans ou plus | 2007  | non specific | 248             |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 02-05 ans      | 2007  | non specific | 12              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 06-10 ans      | 2007  | non specific | 28              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 11-14 ans      | 2007  | non specific | 23              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 15-17 ans      | 2007  | non specific | 14              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 18-24 ans      | 2007  | non specific | 11              |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 25-29 ans      | 2007  | non specific | 7               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | oui       | 30 ans ou plus | 2007  | non specific | 7               |
| 01001  | 84  | 01  | L'Abergement-Clémenciat | Femme | non       | 02-05 ans      | 2007  | non specific | 3               |

```

64 SELECT * FROM information_schema.KEY_COLUMN_USAGE WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingare2017' AND REFERENCED_TABLE_NAME IS NOT NULL;
65 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingare2017`;
66 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingare2017';
67 SELECT * FROM `information_schema`.`COLUMNS` WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingare2012' ORDER BY ORIGINAL_POSITION;
68 SHOW INDEXES FROM `stagingare2012` FROM `datawarehouse_b4people`;
69 SELECT * FROM `information_schema`.`REFERENTIAL_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingare2012' AND REFERENCED_TABLE_NAME IS NOT NULL;
70 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingare2012`;
71 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingare2012';
72 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingaretotal`;
73 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingaretotal`;
74 SELECT * FROM `datawarehouse_b4people`.`stagingaretotal` LIMIT 1000;
75 SHOW TABLE STATUS LIKE `stagingaretotal`;
  
```

Database: datawarehouse\_b4people | Table: dim\_annee | 3 rows total (approximately)

| id_annee | valannee |
|----------|----------|
| 1        | 2007     |
| 2        | 2012     |
| 3        | 2017     |

```

71 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingare2012`;
72 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingare2012';
73 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingaretotal`;
74 SELECT * FROM `information_schema`.`COLUMNS` WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='stagingaretotal' ORDER BY ORIGINAL_POSITION;
75 SHOW INDEXES FROM `stagingaretotal` FROM `datawarehouse_b4people`;
76 SELECT * FROM `information_schema`.`COLUMNS` WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_annee' ORDER BY ORIGINAL_POSITION;
77 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_annee`;
78 SELECT * FROM `information_schema`.`REFERENTIAL_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_annee' AND REFERENCED_TABLE_NAME IS NOT NULL;
79 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_annee`;
80 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_annee';
81 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_annee`;
82 SELECT * FROM `datawarehouse_b4people`.`dim_annee` LIMIT 1000;
  
```

Database: datawarehouse\_b4people | Table: dim\_catiplome | 14 rows total (approximately)

| id_catiplome | valcatDiplome  |
|--------------|--|
| 1            | non specific   |
| 2            | aucun diplome  |
| 3            | BEPC, brevet des collèges                            |
| 4            | CAP ou d'un BEP                                      |
| 5            | un baccalauréat ou d'un brevet professionnel         |
| 6            | diplôme de l'enseignement supérieur court            |
| 7            | diplôme de l'enseignement supérieur long             |
| 8            | aucun diplôme ou au plus un CEP                      |
| 9            | BEPC, brevet des collèges, DNB                       |
| 10           | Baccalauréat, brevet professionnel ou équivalent     |
| 11           | un diplôme de l'enseignement supérieur de niveau ... |
| 12           | un diplôme de l'enseignement supérieur de niveau ... |
| 13           | un diplôme de l'enseignement supérieur de niveau ... |
| 14           | CAP, d'un BEP ou équivalent                          |

```

78 SELECT * FROM `information_schema`.`REFERENTIAL_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_annee' AND REFERENCED_TABLE_NAME IS NOT NULL;
79 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_annee`;
80 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_annee';
81 SELECT * FROM `information_schema`.`COLUMNS` WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_annee' ORDER BY ORIGINAL_POSITION;
82 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_catiplome`;
83 SELECT * FROM `information_schema`.`COLUMNS` WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_catiplome' ORDER BY ORIGINAL_POSITION;
84 SHOW INDEXES FROM `dim_catiplome` FROM `datawarehouse_b4people`;
85 SELECT * FROM `information_schema`.`REFERENTIAL_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_catiplome' AND REFERENCED_TABLE_NAME IS NOT NULL;
86 SELECT * FROM `information_schema`.`KEY_COLUMN_USAGE` WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_catiplome';
87 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_catiplome`;
88 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM `information_schema`.`CHECK_CONSTRAINTS` WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_catiplome';
89 SELECT * FROM `datawarehouse_b4people`.`dim_catiplome` LIMIT 1000;
  
```

Profils\datawarehouse\_b4people\dim\_commune - HeidiSQL 11.3.0.6295

Database: datawarehouse\_b4people | Table: dim\_commune | 3,926 rows total (approximately), limited to 1,000

| dim_commune | valcommune         | id_departement |
|-------------|--------------------|----------------|
| 1           | Abouin             | 7              |
| 2           | Abondance          | 12             |
| 3           | Abrest             | 2              |
| 4           | Accors             | 3              |
| 5           | Affoux             | 10             |
| 6           | Agnat              | 8              |
| 7           | Agnin              | 6              |
| 8           | Agnayes            | 2              |
| 9           | Aiguillette-le-Lac | 11             |
| 10          | Aigueperse         | 9              |
| 11          | Aiguilhe           | 8              |
| 12          | Ailhon             | 3              |
| 13          | Ailhon             | 7              |
| 14          | Aillon-le-Vieux    | 11             |
| 15          | Aillon-le-Vieux    | 11             |
| 16          | Aime-la-Plagne     | 11             |
| 17          | Amay-le-Château    | 2              |
| 18          | Aton               | 11             |
| 19          | Aux-la-Fayette     | 9              |
| 20          | Aux-les-Bains      | 11             |
| 21          | Azac               | 3              |
| 22          | Ajoux              | 3              |

```

93 SELECT * FROM information_schema.KEY_COLUMN_USAGE WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_departement' AND REFERENCED_TABLE_NAME IS NOT NULL;
94 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_departement`;
95 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_departement';
96 SELECT * FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_departement' LIMIT 1000;
97 SELECT * FROM information_schema.COLUMNS WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune' ORDER BY ORIGINAL_POSITION;
98 SHOW INDEXES FROM `dim_commune` FROM `datawarehouse_b4people`;
99 SELECT * FROM information_schema.REFERENTIAL_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune' AND REFERENCED_TABLE_NAME IS NOT NULL;
100 SELECT * FROM information_schema.REFERENTIAL_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune' AND REFERENCED_TABLE_NAME IS NOT NULL;
101 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_commune`;
102 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune';
103 SELECT * FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune' LIMIT 1000;
104 SHOW TABLE STATUS LIKE `dim_commune`;

```

Profils\datawarehouse\_b4people\dim\_trancheage - HeidiSQL 11.3.0.6295

Database: datawarehouse\_b4people | Table: dim\_trancheage | 8 rows total (approximately)

| id_trancheage | valtrancheage  |
|---------------|----------------|
| 1             | 00-05 ans      |
| 2             | 06-10 ans      |
| 3             | 11-14 ans      |
| 4             | 15-17 ans      |
| 5             | 18-24 ans      |
| 6             | 25-29 ans      |
| 7             | 30 ans ou plus |
| 8             | 15 ans ou plus |

```

100 SELECT * FROM information_schema.KEY_COLUMN_USAGE WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune' AND REFERENCED_TABLE_NAME IS NOT NULL;
101 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_commune`;
102 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune';
103 SELECT * FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_commune' LIMIT 1000;
104 SHOW TABLE STATUS LIKE `dim_commune`;
105 SELECT * FROM information_schema.COLUMNS WHERE TABLE_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_trancheage' ORDER BY ORIGINAL_POSITION;
106 SHOW INDEXES FROM `dim_trancheage` FROM `datawarehouse_b4people`;
107 SELECT * FROM information_schema.REFERENTIAL_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_trancheage' AND REFERENCED_TABLE_NAME IS NOT NULL;
108 SELECT * FROM information_schema.REFERENTIAL_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_trancheage' AND REFERENCED_TABLE_NAME IS NOT NULL;
109 SHOW CREATE TABLE `datawarehouse_b4people`.`dim_trancheage`;
110 SELECT CONSTRAINT_NAME, CHECK_CLAUSE FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_trancheage';
111 SELECT * FROM information_schema.CHECK_CONSTRAINTS WHERE CONSTRAINT_SCHEMA='datawarehouse_b4people' AND TABLE_NAME='dim_trancheage' LIMIT 1000;

```

\*

Profils\datawarehouse\_b4people\fact\_population - HeidiSQL 11.3.0.6295

Database: datawarehouse\_b4people | Table: fact\_population | 474,612 rows total (approximately), limited to 1,000

| ID_FACT | FK_IDScolarisation | FK_IDCommune | FK_IDAnnee | FK_IDSexe | FK_IDCategorieDiplome | FK_IDTrancheage | nombrespersonnes |
|---------|--------------------|--------------|------------|-----------|-----------------------|-----------------|------------------|
| 84      | 1                  | 63           | 1          | 1         | 1                     | 1               | 657              |
| 85      | 1                  | 63           | 1          | 1         | 1                     | 1               | 12               |
| 86      | 1                  | 63           | 1          | 1         | 1                     | 1               | 2                |
| 87      | 1                  | 63           | 1          | 1         | 1                     | 1               | 20               |
| 88      | 1                  | 63           | 1          | 1         | 1                     | 1               | 4                |
| 89      | 1                  | 63           | 1          | 1         | 1                     | 1               | 12               |
| 90      | 1                  | 63           | 1          | 1         | 1                     | 1               | 6                |
| 91      | 1                  | 63           | 1          | 1         | 1                     | 1               | 1                |
| 92      | 2                  | 63           | 1          | 1         | 1                     | 1               | 6                |
| 93      | 2                  | 63           | 1          | 1         | 1                     | 1               | 0                |
| 94      | 2                  | 63           | 1          | 1         | 1                     | 1               | 3                |
| 95      | 2                  | 63           | 1          | 1         | 1                     | 1               | 4                |
| 96      | 2                  | 63           | 1          | 1         | 1                     | 1               | 5                |
| 97      | 2                  | 63           | 1          | 1         | 1                     | 1               | 19               |
| 98      | 2                  | 63           | 1          | 1         | 1                     | 1               | 203              |
| 99      | 1                  | 72           | 1          | 1         | 1                     | 1               | 7                |
| 100     | 1                  | 72           | 1          | 1         | 1                     | 1               | 10               |
| 101     | 1                  | 72           | 1          | 1         | 1                     | 1               | 6                |
| 102     | 1                  | 72           | 1          | 1         | 1                     | 1               | 7                |
| 103     | 1                  | 72           | 1          | 1         | 1                     | 1               | 5                |
| 104     | 1                  | 72           | 1          | 1         | 1                     | 1               | 0                |
| 105     | 1                  | 72           | 1          | 1         | 1                     | 1               | 0                |
| 106     | 1                  | 72           | 1          | 1         | 1                     | 1               | 1                |

```

128 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingarea2012` LIMIT 1000;
129 SELECT * FROM `datawarehouse_b4people`.`stagingarea2012` LIMIT 1000;
130 SHOW TABLE STATUS LIKE `stagingarea2012`;
131 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingarea2017` LIMIT 1000;
132 SELECT * FROM `datawarehouse_b4people`.`stagingarea2017` LIMIT 1000;
133 SHOW TABLE STATUS LIKE `stagingarea2017`;
134 SHOW CREATE TABLE `datawarehouse_b4people`.`stagingareatotal` LIMIT 1000;
135 SELECT * FROM `datawarehouse_b4people`.`stagingareatotal` LIMIT 1000;
136 SHOW TABLE STATUS LIKE `stagingareatotal`;
137 SHOW CREATE TABLE `datawarehouse_b4people`.`fact_population` LIMIT 1000;
138 SELECT * FROM `datawarehouse_b4people`.`fact_population` LIMIT 1000;
139 SHOW TABLE STATUS LIKE `fact_population`;

```

### 5.1.4. Visualisations des données :

Exemple de test de visualisation après l'implémentation des tables avec Power BI et après nous abordons notre outil de visualisation implémenté :

#### Navigator

Display Options ▾

- localhost:3306;datawarehouse: datawarehouse...
- dim\_annee
- dim\_catdiplome
- dim\_commune
- dim\_departement
- dim\_region
- dim\_scolarisation
- dim\_sexe
- dim\_trancheage
- fact\_population
- stagingarea2007
- stagingarea2012
- stagingarea2017

dim\_scolarisation  
Preview downloaded on Thursday

| Id_scolarisation | valscolarisation | fact_population |
|------------------|------------------|-----------------|
| 1                | oui              | Table           |
| 2                | non              | Table           |

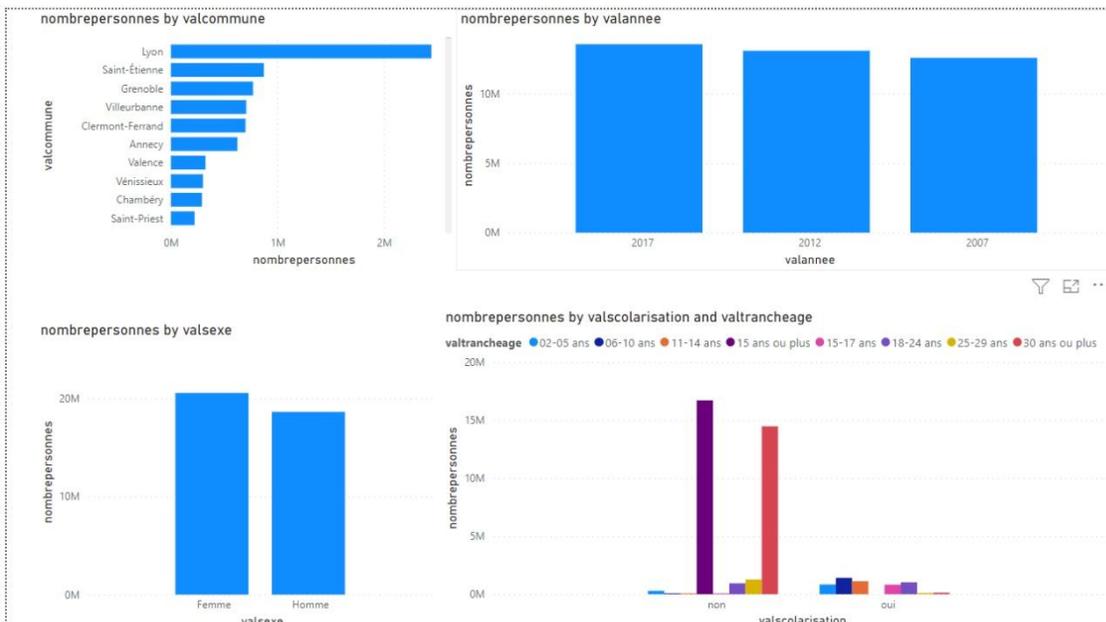
  


Figure 17 : Quelques visualisations pour le test avec Power BI

#### 5.1.4.1. Visualisations et notion avec cube.js :

Tout d'abord Cube Js est un Framework Open Source permettant de créer des applications web analytiques, Il est principalement utilisé pour créer des outils de la BI ou pour ajouter des analyses orientées client à une application existante.

Dans la plupart des cas, la première étape de la création d'une telle application est un tableau de bord analytique.

Classiquement, vous exécutez l'API Cube.js en tant que service il gère la connexion à votre base de données, y compris la file d'attente des requêtes, la mise en cache, la pré-agrégation, etc. Il expose également une API pour votre application frontale afin de créer des tableaux de bord et d'autres fonctionnalités d'analyse comme la montre la figure ci-dessous :



Figure 18 : Architecture fonctionnelle de Cube Js

Vous pouvez démarrer avec Cube localement ou l'auto-héberger avec Docker ou sur le Cloud (cube nuage)

Il existe également une option héritée pour démarrer avec Node.js, celle utilisée durant notre Projet.

- La première chose c'est d'exécutez la commande suivante pour démarrer avec Cube.js, en spécifiant le nom du projet et éventuellement votre base de données à l'aide de l'indicateur -d :

```
npx cubejs-cli create ProjectBI -d MySQL
```

Une fois exécutée, la création d'un nouveau répertoire de projet contenant l'échafaudage de votre nouveau projet Cube.js. Cela inclut tous les fichiers nécessaires pour lancer le backend Cube.js

Le fichier .env de ce répertoire de projet contient des espaces réservés pour les informations d'identification de base de données pertinentes comme vous voyez dans la figure :

```
CUBEJS_DB_HOST=127.0.0.1
CUBEJS_DB_PORT=3306
CUBEJS_DB_NAME=datawarehouse_bi4people
CUBEJS_DB_USER=root
CUBEJS_DB_PASS=123456789
CUBEJS_DB_TYPE=mysql
CUBEJS_API_SECRET=d461457c3947350f80ef36aaf71a270bf8ca2ebde496fa30235e39766fcde45b6b590d018ade408cb40616fec733a7015391e9404adeab571
CUBEJS_EXTERNAL_DEFAULT=true
CUBEJS_SCHEDULED_REFRESH_DEFAULT=true
CUBEJS_DEV_MODE=true
```

Figure 19 : Contenu du fichier .env de notre projet

Cube.js utilise **Data Schema** pour générer et exécuter SQL.

Il agit comme un ORM pour votre base de données et il est suffisamment flexible pour tout modéliser.

**(ORM signifie Object-Relational Mapping) : Il est un ensemble de classes permettant de manipuler les tables d'une base de données relationnelles comme s'il s'agissait d'objets.**

Un ORM est une couche d'abstraction d'accès à la base de données qui donne l'illusion de ne plus travailler avec des requêtes mais de manipuler des objets.

Accédez ensuite à <http://localhost:4000> et utilisez le pour générer des fichiers de schéma :

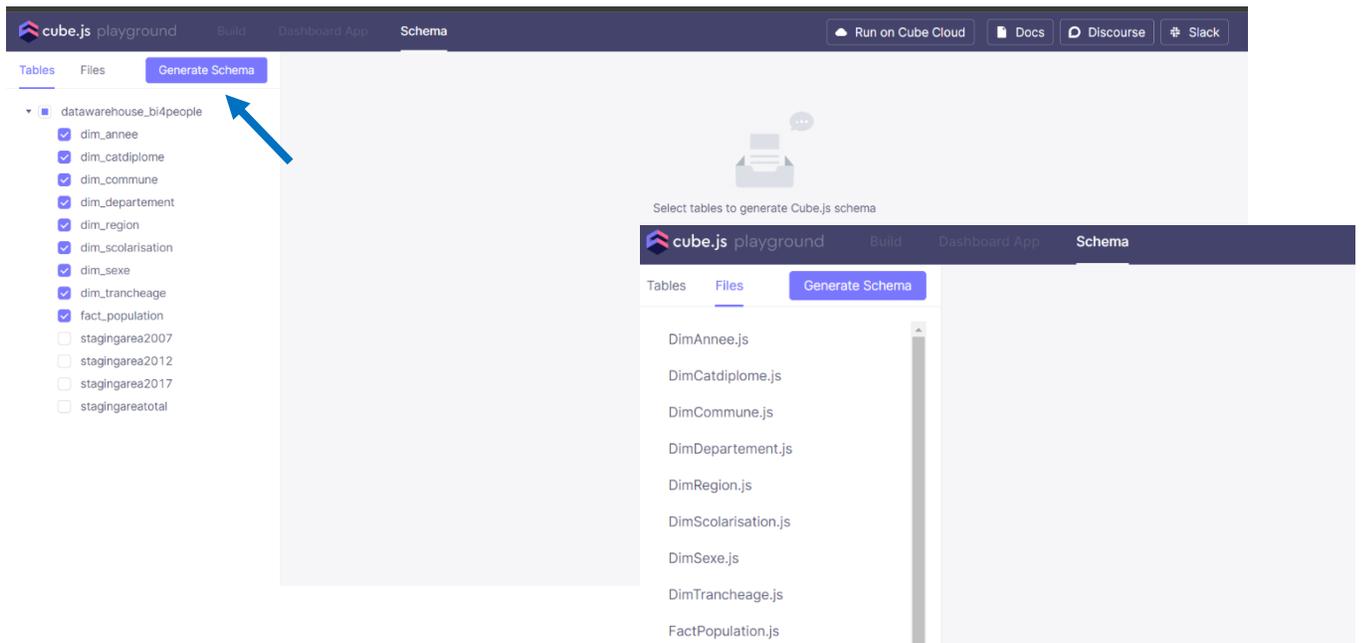


Figure 20 : Génération du Schéma

Les éléments de base du schéma de données sont les mesures et les dimensions :

- La mesure est appelée données quantitatives, telles que le nombre d'unités vendues, le nombre de profit, etc.
- La dimension est appelée données catégorielles, telles que l'état, le sexe, le nom du produit ou les unités de temps (par exemple, jour, semaine, mois).

➤ Installation du client Cube.js (Nous avons utilisé React)

```
npm i --save @cubejs-client/core  
npm i --save @cubejs-client/react
```

➤ Pour accélérer les performances des requêtes, envisagez d'utiliser des pré-agrégations, les pré-agrégations sont une couche de données agrégées créées et actualisées par Cube.js. Cela peut considérablement améliorer les performances des requêtes et fournir une simultanéité plus élevée.

```
This file can be edited at C:\Users\moadt\ProjectBI\schema\DimCommune.js  
  
cube(`DimCommune`, {  
  sql: `SELECT * FROM datawarehouse_bi4people.dim_commune`,  
  
  preAggregations: {  
    // Pre-Aggregations definitions go here  
    // Learn more here: https://cube.dev/docs/caching/pre-aggregations/getting-started  
  },  
  
  joins: {  
  },  
  
  measures: {  
    count: {  
      type: `count`,  
      drillMembers: []  
    },  
  },  
  
  dimensions: {  
    valcommune: {  
      sql: `valcommune`,  
      type: `string`  
    },  
  },  
  
  dataSource: `default`  
});
```

### 5.1.4.2. Tableau de bord React :

Nous allons construire notre interface et notre tableau de bord avec React, en utilisant le client Cube.js React.

Mais nous pouvons utiliser n'importe quel framework pour créer une interface avec Cube.js.

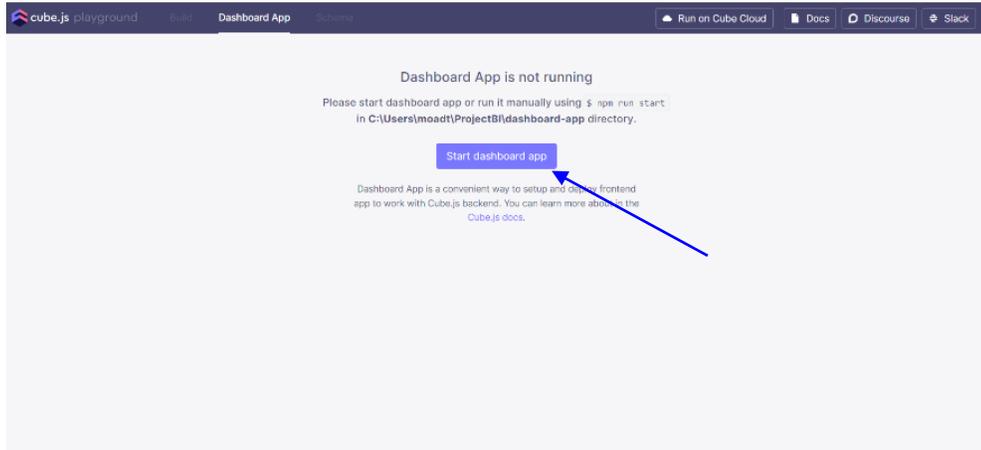


Figure 21 : Démarrage du dashboard

Un serveur sur le port 3000 s'ouvre sur votre navigateur Web sur localhost :3000

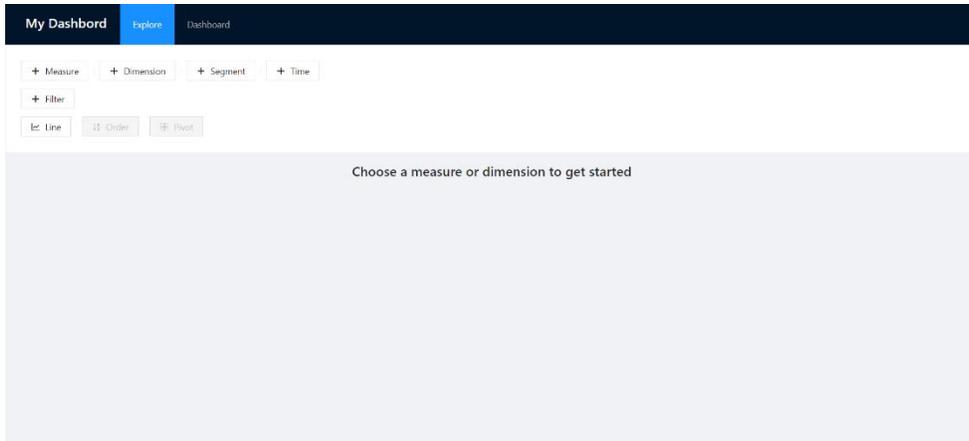


Figure 22 : Interface Dashboard avec React

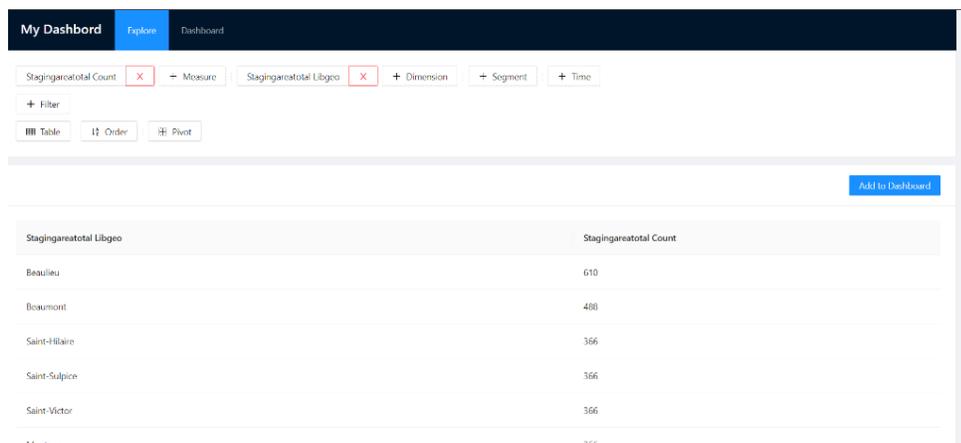


Figure 23 : Visualisation sous forme d'une Table

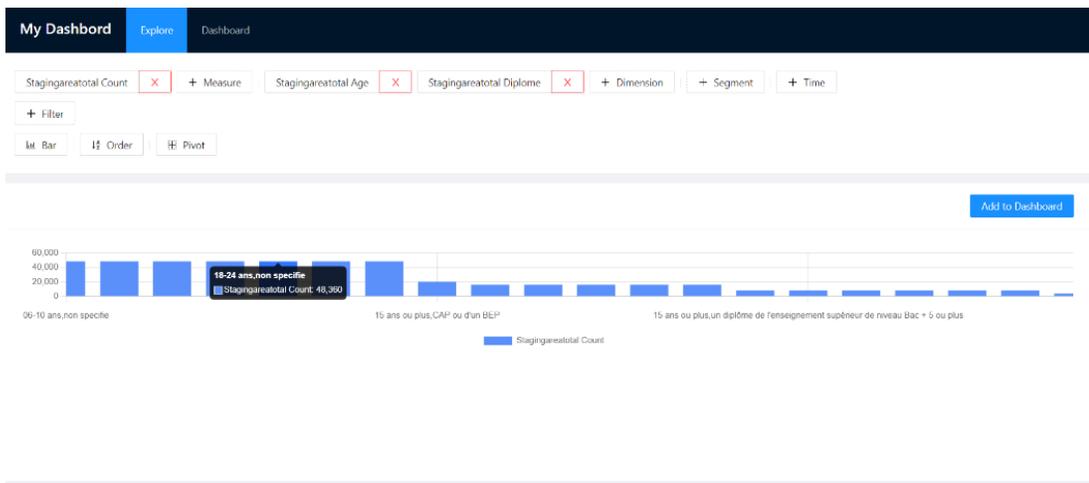


Figure 24 : Visualisation sous forme de Bar

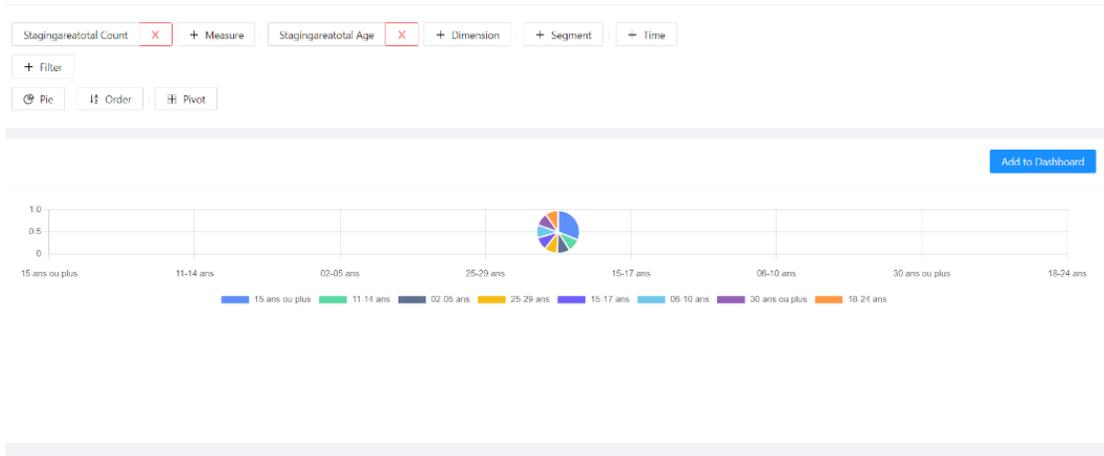


Figure 25 : Visualisation sous forme de Pie chart

Les deux figures ci-dessus montrent la population par tranches d'âge et catégorie de diplôme.

Si nous voulons enregistrer nos Dashboard pour des prochaines utilisations nous faisons **Add to Dashboard** :

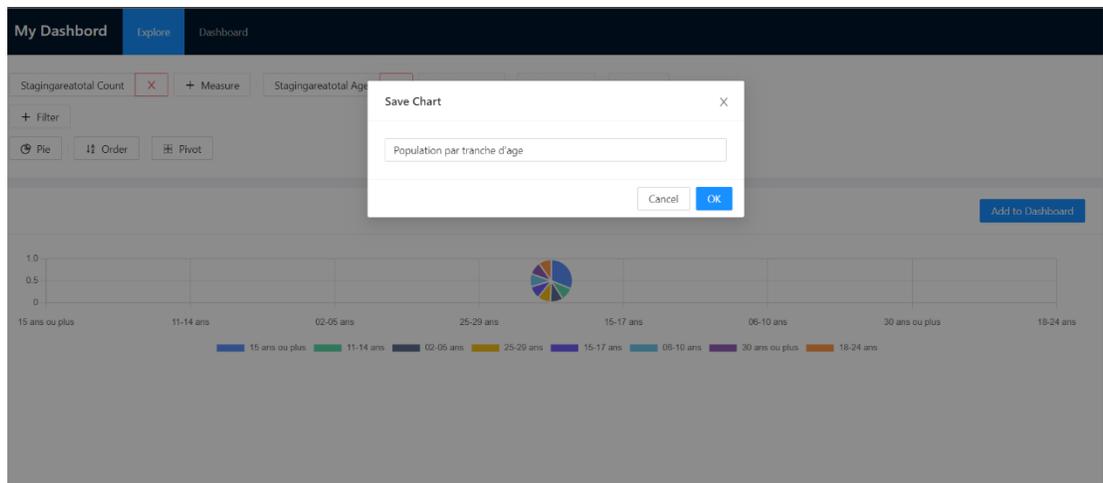


Figure 26 : Enregistrement du Dashboard

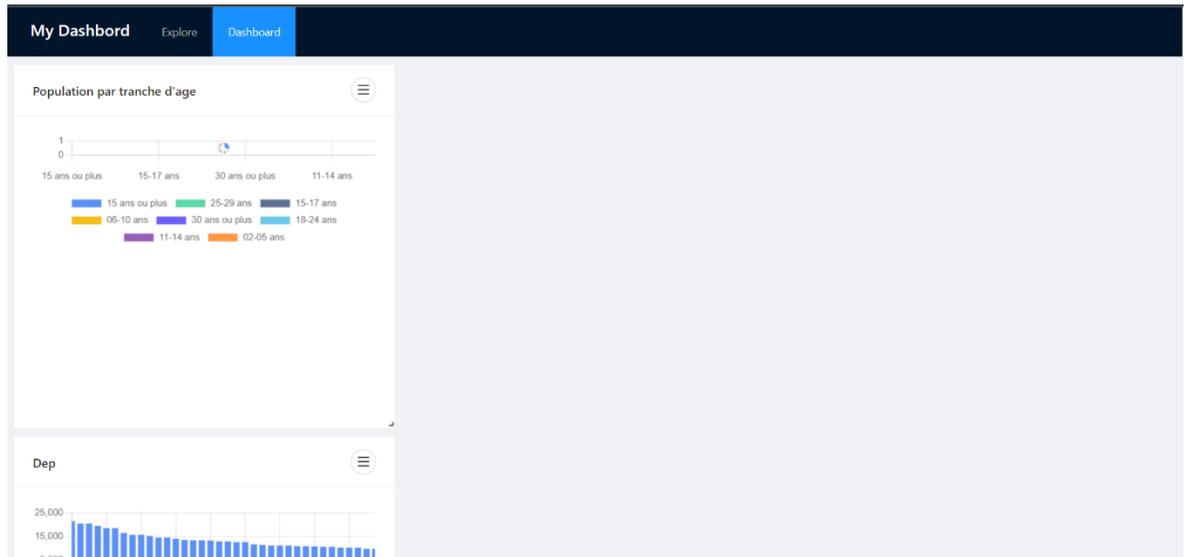


Figure 27 : Vue des visualisations enregistrer

Une fois le graphique enregistré on peut le modifier ou le supprimer ultérieurement comme vous voyez

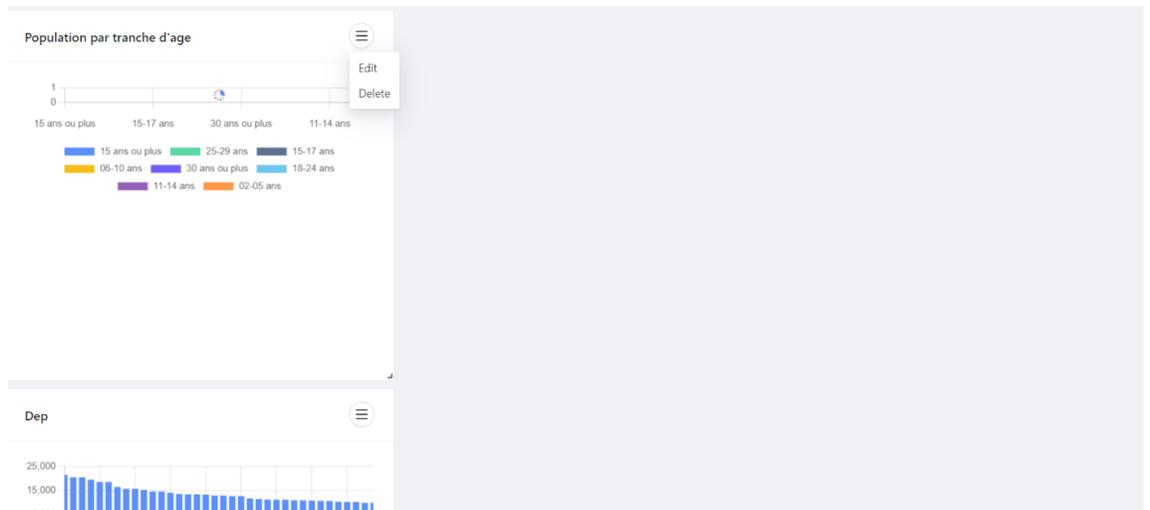


Figure 28 : Modification des visualisations après la création

## 5.2. Dashboard bilingues (Fr, En) :

### 5.2.1. L'enjeux de multilingues :

Surmonter la barrière de la langue pour les utilisateurs qui utilisent votre logiciel est un sujet important, l'anglais n'est plus la langue universelle d'Internet.

Dans cette partie, on s'intéresse à une phase d'internationalisation puisque dans ce projet y ont à des gens francophones et anglophones donc l'idée c'est de rendre notre Dashboard utilisateurs multilingues (Français, Anglais) ou bilingues.

C'est à partir de ce dictionnaire qui est construit au format JSON (fig.29) en prenons en compte y auront plusieurs perspectives, pour la traduction y avaient deux options : faire une version anglaise et une autre française mais ça va poser des problèmes de maintenabilités parce que si en évolue la version anglaise il faut aussi faire une mise à jour pour la version française d'où le faite de prendre en compte les deux langues pour qu'e ne sera pas un impact lors de l'évolution de l'application.

### 5.2.2. Choix de l'outil :

Dans l'écosystème JavaScript, il existe de nombreux frameworks d'internationalisation.

Dans ce projet, nous utiliserons le Framework **i18next** pour internationaliser notre application React.js .

#### 5.2.2.1. *Les avantages de react-i18next :*

- Taux d'adoption rapide en ce qui concerne les nouvelles fonctionnalités de React.
- API très efficace
- Écosystème i18n qui n'est pas lié à React

### 5.2.3. Réagir avec React-i18next :

Avant de pouvoir commencer à traduire notre application React, nous devons installer ses bibliothèques :

```
npm install react-i18next i18next --save  
npm install i18next-browser-languagedetector --save
```

Nous utiliserons la bibliothèque principale i18next pour la configuration et la bibliothèque react-i18next pour connecter ses capacités d'internationalisation à React.

Après nous créons des fichiers JSON pour les langues désirées, un identifiant unique est à chaque chaîne traduite, il est utilisé pour rechercher le JSON et localiser la valeur de la chaîne dans différentes langues.

```
dashboard-app > src > locale > {} FR.json > {} translation > chart
2  "translation": {
3    "Explore": "Explorer",
4    "Measure": "Mesure",
5    "measure": "mesure",
6    "Dashbord": "Tableau de bord",
7    "Dimension": "une dimension",
8    "Segment": "Segment",
9    "Time": "Temps",
10   "Line": "Ligne",
11   "Area": "Zone",
12   "Bar": "Bar",
13   "Pie": "Tarte",
14   "Settings": "configurations",
15   "Edit": "Editer",
16   "Number": "Nombre",
17   "chart": "Graphique",
18   "Query": "Requete",
19   "Security": "sécurité",
20   "schema": "schéma",
21   "Filter": "Filtre",
22   "Choose" : "Choisissez",
23   "get" : "pour",
24   "or" : "ou",
25   "started": "commencer",
26   "My": "Mon",
27   "Order": "Ordre",
28   "Pivot": "Pivot",
29   "members" : "Membre",
30   "Found" : "trouver",
31   "No" : "Pas"
32 }
33 }
```

Figure 29 : Dictionnaire des données en JSON

Nous créons un fichier de configuration `i18n` dans `src/i18n.js` qui ressemble à ceci :

```
PROJECTBI dashboard-app > src > JS i18n.js > react
1 import i18n from "i18next"; 56.2k (gzipped: 14.9k)
2 import { initReactI18next } from "react-i18next"; 2.9k (gzipped: 1.3k)
3 import LanguageDetector from "i18next-browser-languagedetector" 6.6k (gzipped: 2.3k)
4 import translationFR from "./locale/FR.json"
5 const resources = {
6
7   fr: {
8     translation: translationFR
9   }
10 };
11
12 i18n
13   .use(LanguageDetector)
14   .use(initReactI18next)
15   .init({
16     resources,
17     lng: "en",
18     interpolation: {
19       escapeValue: false
20     },
21     react: {
22       useSuspense: false
23     }
24   });
25
26 export default i18n;
```

Cette importation permet de détecter la langue de navigateur d'utilisateur ou le visiteur de la plateforme (si c'est le français il va afficher l'interface en français sinon en anglais) :

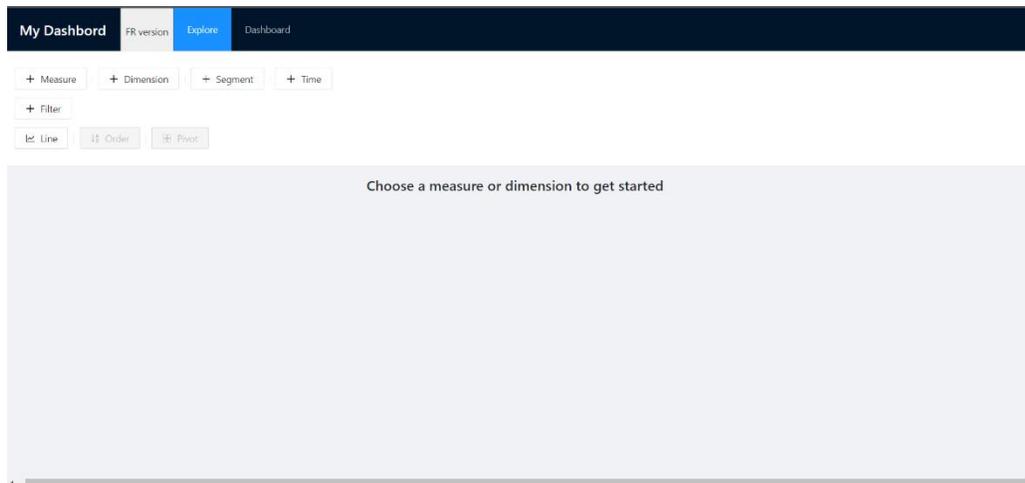
```
import LanguageDetector from "i18next-browser-languagedetector";
```

Modifications dans le fichier `Header.js` :

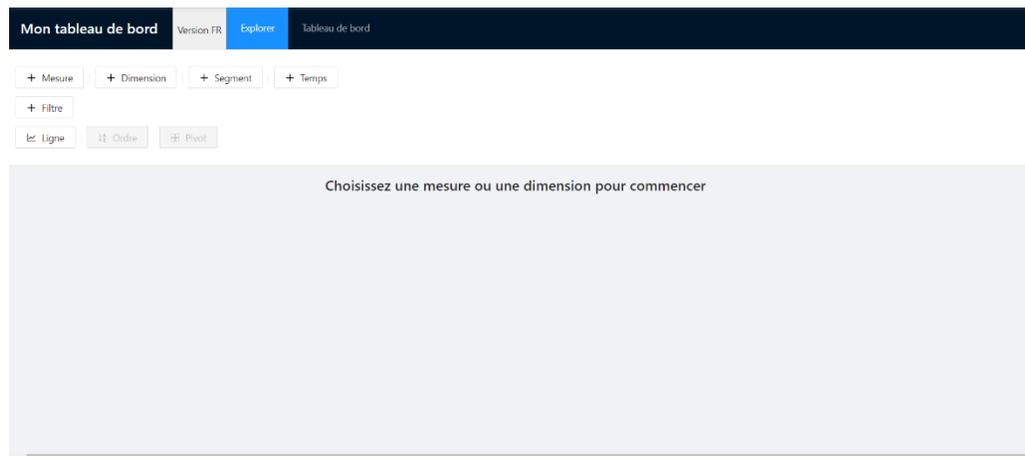
```
components 42
> QueryBuilder 43
JS ChartRenderer.js 44
JS Dashboard.js 45
JS DashboardItem.js 46
JS Header.js 1 47
JS TitleModal.js 48
graphql 49
i18n 50
{} en.json 51
{} fr.json 52
JS i18n.js 53
≡ index 54
i18n 55
locale 56
pages 57
services 58
App.css 59
App.js 60
App.test.js 61
body.css 62
i18n.js 63
index.css 64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

## 5.2.4. Résultat :

Version initiale du Tableau de bord :



Version Traduite à base du vocabulaire de la version originale :



## 5.2.5. Conclusion :

Cette tâche d'internationalisation est très importante pour notre projet, c'est une tâche qui n'est pas encore terminée, en cours pour une version finale.

## Conclusion

*Ce stage de fin d'études a été une bonne occasion pour s'intégrer dans le milieu de la recherche qui m'a permis de découvrir l'axe de la BI collaboratives.*

*Notre travail explore le concept d'organisation ou équipe virtuelle dont les membres sont issus de plusieurs organisations où unités de recherche et réunis à l'occasion de se rendre d'accord dans des réunions plénières : la résolution collaborative d'un problème décisionnel, la durée de vie de l'équipe est la durée de vie du projet qui permet la bonne conduite et à une satisfaction plus élevée pour la génération d'idées pour la prise de décision collective.*

*La facilitation semble également être une variable conduisant à une décision de meilleure qualité et à la satisfaction vis-à-vis du processus.*

*Cependant ceci n'était pas facile pour moi à cause de plusieurs contraintes, mon arrivée tardive en France à cause de quelques contraintes administratives(vers la dernière semaine du mois d'avril) ca était difficile à distance depuis le Maroc, ainsi pour l'intégration dans un grand projet comme celui-ci vraiment ce n'est pas facile parce qu'il faut bien comprendre le projet, savoir les buts désirés, communiquer avec les autres membres régulièrement, prendre en compte des nouvelles idées qui apparaissent.*

# Bibliographie

**Evelson B., March 2015**, an approach to converge the worlds of Big Data and BI,

[https://www.forrester.com/blogs/15-03-27 -  
an approach to converge the worlds of big data and bi/](https://www.forrester.com/blogs/15-03-27-an-approach-to-converge-the-worlds-of-big-data-and-bi/)

**Delaney Rebernik, (2015)** Collaborative Business Intelligence, TechTarget,

[https://searchbusinessanalytics.techtarget.com/definition/collaborative-BI-collaborative-business-intelligence \(2015\)](https://searchbusinessanalytics.techtarget.com/definition/collaborative-BI-collaborative-business-intelligence-(2015))

**Devon Tackels**, 6 benefits of a collaborative approach to analytics and bi,

(2015)<https://www.sigmacomputing.com/blog/6-benefits-of-a-collaborative-approach-to-analytics-and-bi/>

**Abelló, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazón, J., Naumann, F., Pedersen, T., Rizzi, S. B., Trujillo, J., Vassiliadis, P., & Vossen, G. (2013)**. Fusion Cubes: Towards Self-Service Business Intelligence. International Journal of Data Warehousing and Mining (IJDWM), 9(2), 66-88.

**Kaufmann J. and Chamoni P.**, Structuring Collaborative Business Intelligence: A Literature Review, 2014 47th Hawaii International Conference on System Science, pp. 3738-3747, IEEE (2014).

**Adla A, Nachet B, Ould-Mahraz A**, Multi-Agents Model for Web-based Collaborative Decision Support Systems. ICWIT, 294-299

**Eirinaki M., Abraham S., Polyzotis N., Shaikh N.. (2014)**. QueRIE: Collaborative database exploration. IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 26 (7), pp. 1778-1790. IEEE (2014).

**Cao L., Zhang C., and Liu J. (2006)** Ontology-based integration of business intelligence. Web Intelligence and Agent Systems, vol. 4(3), pp. 313-325 (2006).

**Fahad, Darmont, Favre, Loudcher** Collaborative Business Intelligence: A Survey French National Research Agency (ANR), project ANR-19-CE23-0005 BI4people (Business Intelligence for the people).

**Passlick, J.; Lebek, B.; Breitner, M. H. (2017)** A Self-Service Supporting Business Intelligence and Big Data Analytics Architecture, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. International en Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, Switzerland. 1126-1140

**Jerbi H., Ravat F., Teste O. and Zurfluh G. (2009)**., Preference-Based Recommendations for OLAP Analysis, DaWaK 2009, LNCS, vol. 5691, pp. 467-478. Springer.

**Liu T., Shen A., Hu X., Tong G., Gu W.**, The Application of Collaborative Business Intelligence Technology in the Hospital SPD Logistics Management Model, Iran J Public Health, Vol. 46, No.6, Jun 2017, pp.744-754adl