

# Documentation – Préparation des données

Ce guide technique, conçu pour les utilisateurs sous environnement **Windows**, détaille la procédure d'ingestion de données via API et les étapes de transformation (ETL) nécessaires à l'élaboration d'un tableau de bord de veille technologique.

**Prérequis technique** : Installation préalable du logiciel **Microsoft Power BI Desktop**.

## 1. Introduction et Objectifs

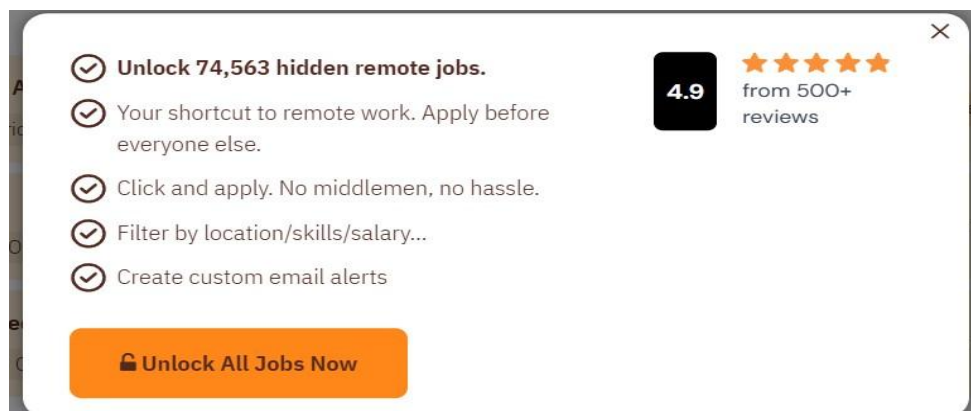
### La Veille Technologique par l'Analyse du Marché de l'Emploi

L'objectif de cette démarche n'est pas de suivre l'actualité médiatique ("le buzz"), mais d'identifier les **tendances structurelles** ("ce qui est acheté"). L'analyse quantitative des offres d'emploi constitue un indicateur avancé pertinent : si une technologie est explicitement demandée dans les fiches de poste, cela atteste de son utilisation effective en production au sein des entreprises.

### Source de Données : L'API Remoteive.com

- **Spécialisation Tech** : Contrairement aux agrégateurs généralistes (Indeed, Monster), Remoteive cible spécifiquement les métiers de la Tech (Développement, Data, DevOps, Design), garantissant un ratio signal/bruit élevé.
- **Accessibilité** : Il s'agit d'une "Open API" ne requérant aucune clé d'authentification pour l'accès en lecture.

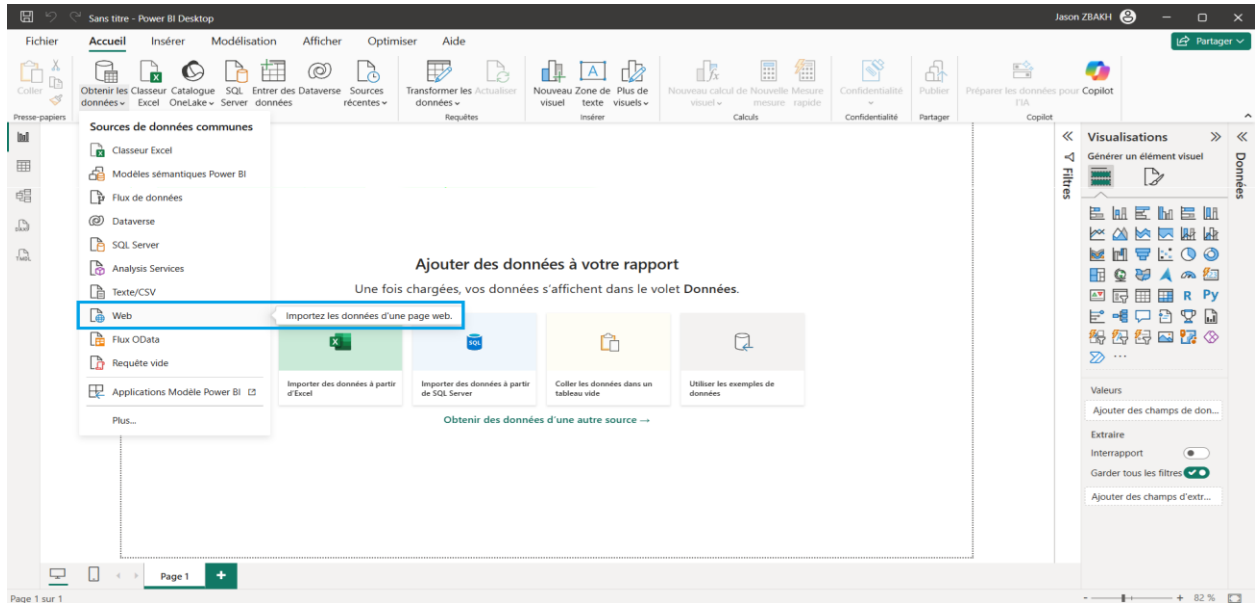
**Note sur les limitations (Version Gratuite)** : L'API publique limite le nombre d'offres retournées et la profondeur de l'historique. Pour un déploiement en production critique, l'acquisition d'un accès complet à la base de données historique est recommandée.



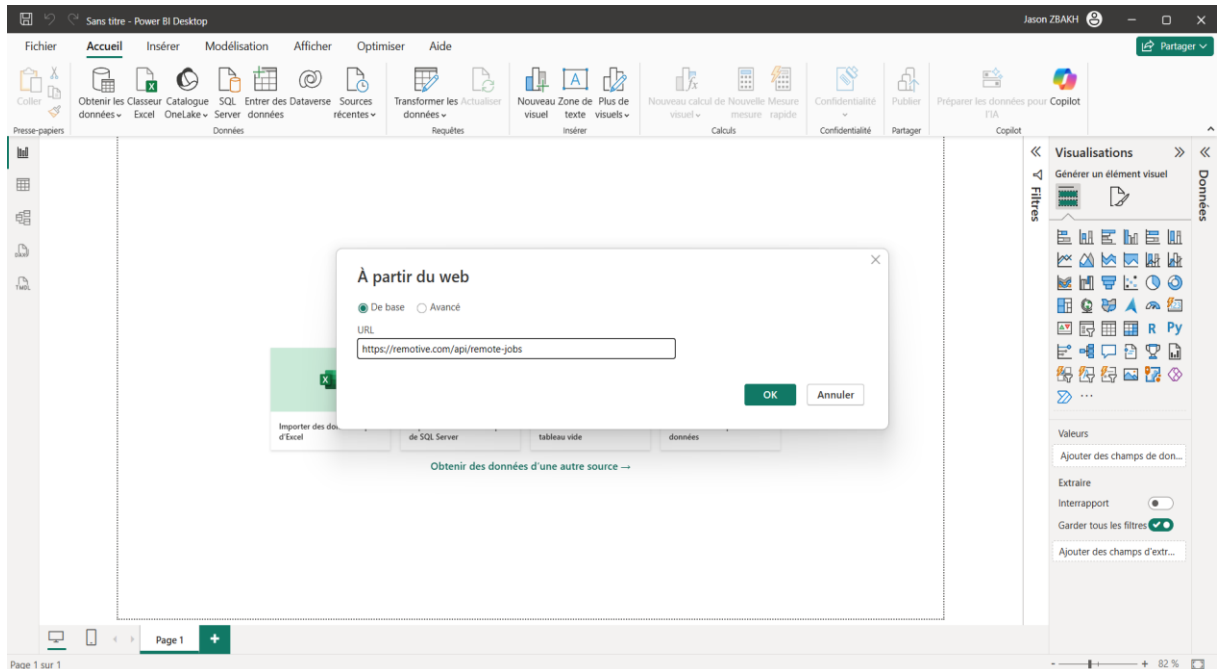
## 2. Procédure d'Importation et de Traitement (ETL)

### Étape 2.1 : Connexion à la Source Web

1. Lancez **Power BI Desktop**.
2. Depuis l'écran d'accueil, créez un **nouveau rapport vierge**.
3. Dans le ruban "Accueil", sélectionnez **Obtenir les données** puis choisissez **Web**.



4. Dans la boîte de dialogue, saisissez l'URL du point de terminaison de l'API :  
<https://remotive.com/api/remote-jobs>



5. Validez par **OK**. Une fois la connexion établie, l'éditeur **Power Query** s'ouvre automatiquement.

## Étape 2.2 : Structuration des Requêtes

Sur le panneau latéral gauche (volet "Requêtes"), une requête nommée remote-jobs (ou similaire) apparaît. Elle contient l'intégralité des données brutes JSON.

**Bonne pratique :** Nous allons utiliser cette table comme **source de référence** immuable. Cela permet de conserver l'état original des données pour audit et de dériver plusieurs tables de travail sans recharger l'API plusieurs fois.

### Création de la Table de Faits "Emploi"

1. Dans le volet de gauche, effectuez un **clic droit** sur la requête remote-jobs.
2. Sélectionnez l'option **Référence**. Une nouvelle requête liée est créée.

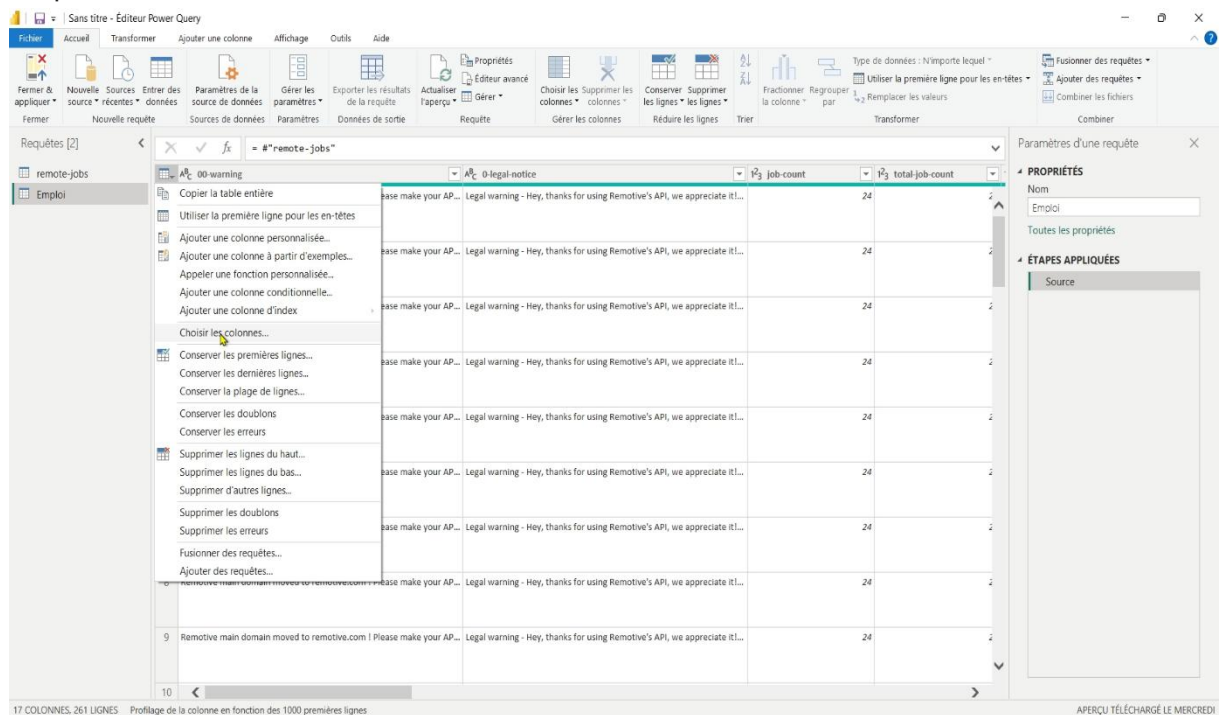
The screenshot shows the Power Query Editor interface. In the left-hand 'Requêtes' (Queries) pane, the 'remote-jobs' query is selected. A right-click context menu is open over this query, and the 'Référence' (Reference) option is highlighted. The main workspace displays a table with the following columns: '0-legal-notice', '0-legal-notice', 'job-count', and 'total-job-count'. The table contains 10 rows of data. The right-hand pane shows the 'Propriétés' (Properties) and 'Étapes appliquées' (Applied Steps) for the selected query. The 'Propriétés' section shows the name 'remote-jobs' and the source 'Source'. The 'Étapes appliquées' section shows the steps 'Converti en table', 'jobs développés', 'jobs développés1', and 'jobs tags développés'.

3. **Clic droit** sur cette nouvelle requête > **Renommer**. Nommez-la Emploi.

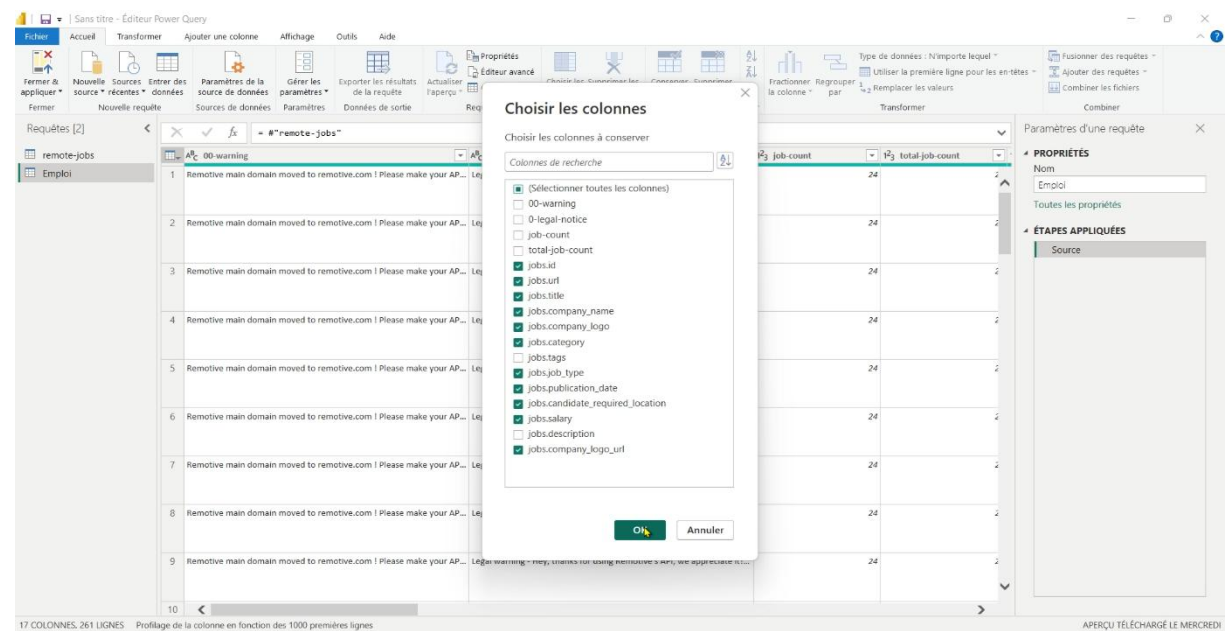
The screenshot shows the Power Query Editor interface after the query has been renamed. In the left-hand 'Requêtes' (Queries) pane, the query is now named 'Emploi'. The main workspace still displays the same table with columns '0-legal-notice', '0-legal-notice', 'job-count', and 'total-job-count'. The right-hand pane shows the 'Propriétés' (Properties) and 'Étapes appliquées' (Applied Steps) for the renamed query. The 'Propriétés' section now shows the name 'remote-jobs (2)' and the source 'Source'. The 'Étapes appliquées' section shows the steps 'Source'.

## Nettoyage de la Table "Emploi"

1. Sélectionnez la requête Emploi.
2. Cliquez sur **Choisir les colonnes**.



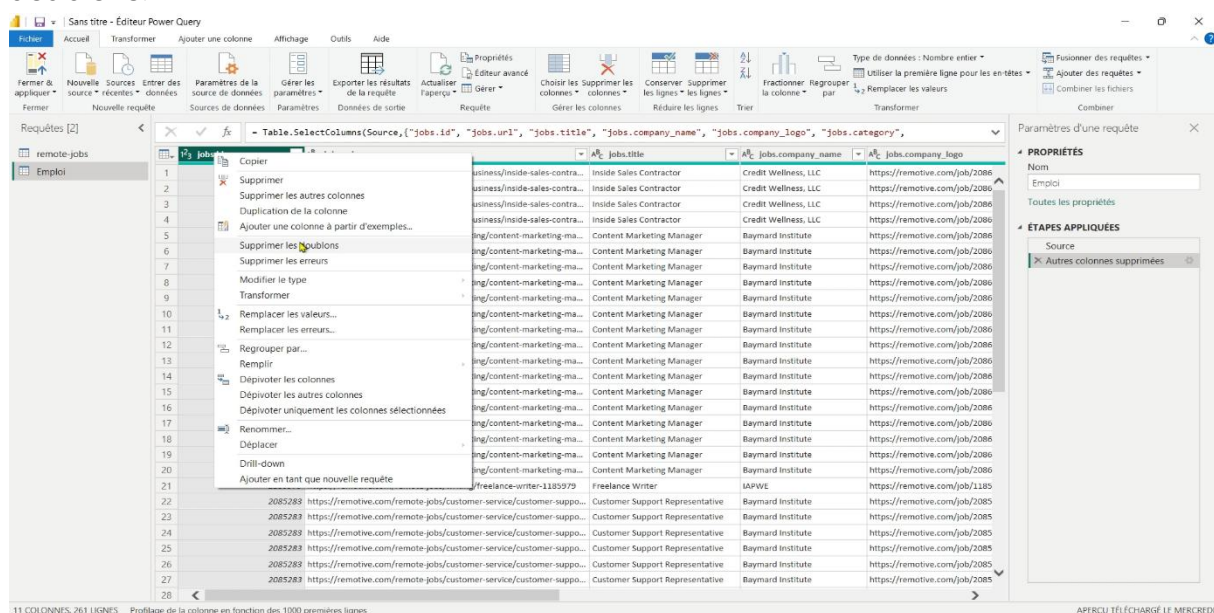
3. Cochez uniquement les champs nécessaires à l'analyse pour optimiser le modèle :  
**id, title, url, company\_name, company\_logo, category, job\_type, publication\_date, candidate\_required\_location, salary et company\_logo\_url.**



- Décochez impérativement la colonne tags (nous la traiterons séparément).

4. Validez par **OK**.

## 5. Garantissez l'unicité des offres : **Clic droit** sur l'en-tête de la colonne id > **Supprimer les doublons.**

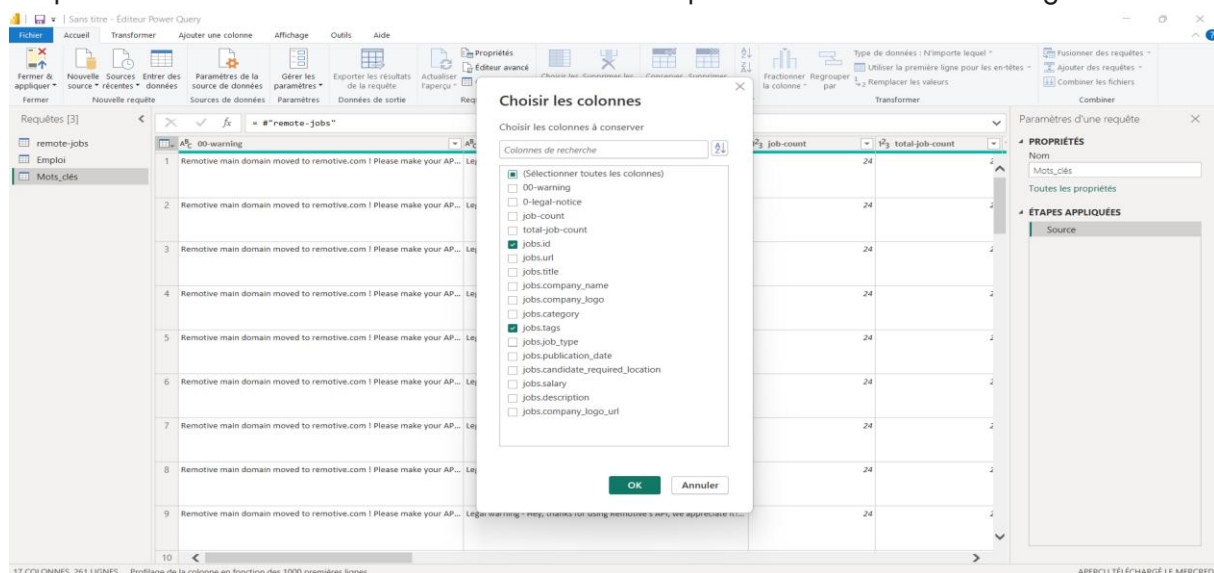


- *Résultat : La table Emploi contient désormais une ligne unique par offre.*

## Création de la Table de Dimension "Mots-Clés"

Nous allons isoler les compétences (tags) dans une table dédiée pour faciliter l'analyse croisée.

1. Dans le volet de gauche, effectuez de nouveau un **clic droit** sur la requête source **remote-jobs**.
2. Sélectionnez **Référence**.
3. Renommez cette nouvelle requête **Mots\_clés**.
4. Cliquez sur **Choisir les colonnes** et ne conservez que deux colonnes : **id** et **tags**.





### 3. Classification Sémantique (Taxonomie)

Les tags bruts sont hétérogènes (mélange de langages, d'outils, de concepts). Une classification manuelle est nécessaire pour structurer l'analyse.

#### Étape 3.1 : Extraction et Qualification

1. Copiez la liste des tags uniques depuis Power Query (ou exportez-les).
2. Collez-les dans un classeur **Excel** pour attribution manuelle d'une catégorie.

#### Étape 3.2 : Méthodologie de Classification Tripartite

Pour garantir la cohérence de l'analyse, appliquez la taxonomie suivante :

Pilier	Critère de Distinction	Sous-Catégories (Exemples)
<b>1. Technologique</b> (L'Outillage)	<b>Critère : Instrumentalité.</b> Désigne un objet (logiciel, langage, infrastructure) que l'on <i>utilise, code</i> ou <i>configure</i> . C'est un moyen technique.	<ul style="list-style-type: none"><li>• <b>Langages</b> : Python, Java, C++</li><li>• <b>Frameworks</b> : React, Django</li><li>• <b>Infra/DevOps</b> : Docker, Kubernetes</li><li>• <b>SaaS</b> : Salesforce, Jira</li><li>• <b>Standards</b> : JSON, REST</li></ul>
<b>2. Compétence</b> (Le Savoir-Faire)	<b>Critère : Intrinsèque.</b> Désigne une capacité intellectuelle ou un métier qui "voyage" avec l'individu, indépendamment des outils.	<ul style="list-style-type: none"><li>• <b>Méthodologies</b> : Agile, Scrum</li><li>• <b>Disciplines</b> : SEO, Copywriting, Data Science</li><li>• <b>Soft Skills</b> : Leadership, Communication</li><li>• <b>Processus</b> : Lead Gen, QA Testing</li></ul>
<b>3. Contexte</b> (L'Environnement)	<b>Critère : Situationnel.</b> Ne décrit ni le candidat ni ses outils.	<ul style="list-style-type: none"><li>• <b>Industrie</b> : Fintech, Santé, Gaming</li><li>• <b>Modèle éco</b> : SaaS, B2B, Marketplace</li><li>• <b>Stade</b> : Startup, Scale-up</li></ul>

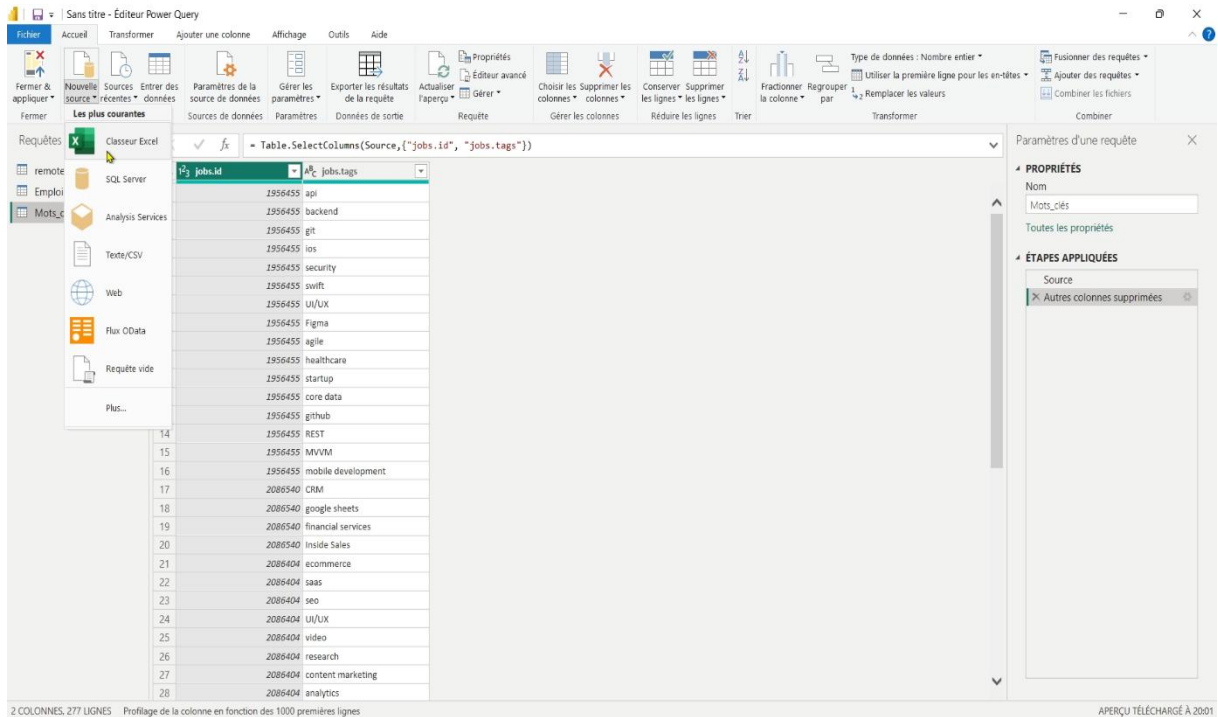
1	jobs.tags	Type
2	CRM	Technologie
3	google sheets	Technologie
4	financial services	Context
5	Inside Sales	Compétence
6	ecommerce	Context

Une fois le fichier Excel complété, enregistrez-le localement.

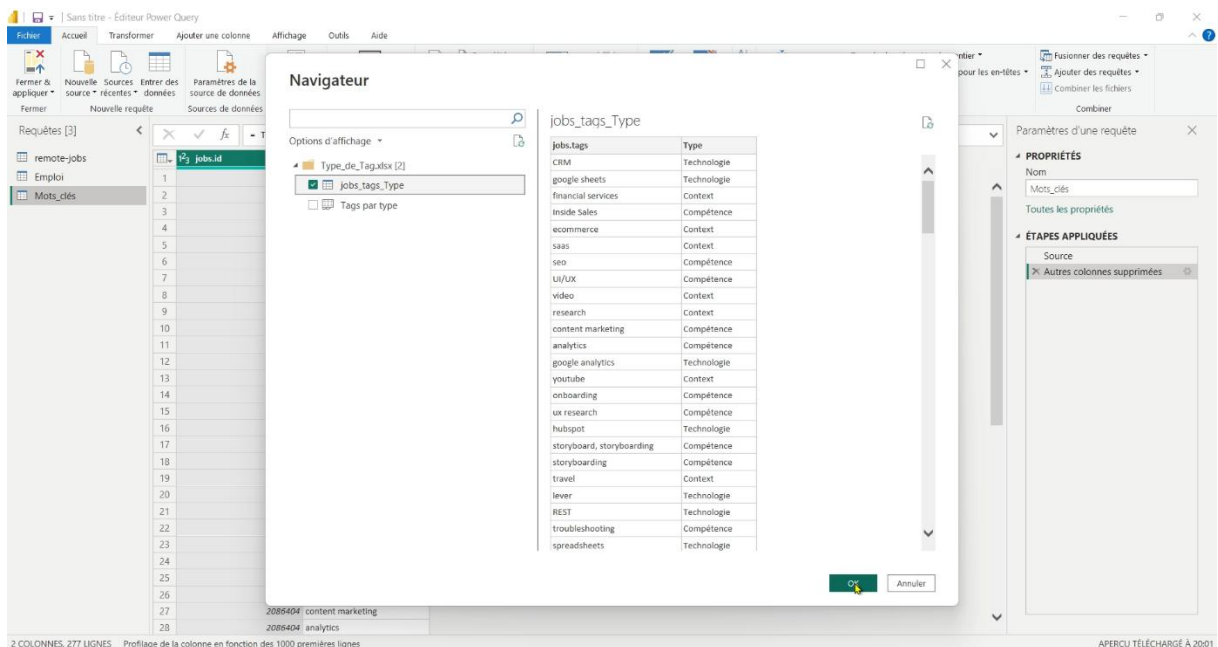
## 4. Enrichissement du Modèle de Données

### Étape 4.1 : Import de la Classification

1. Dans Power Query, cliquez sur **Nouvelle source** > **Classeur Excel**.



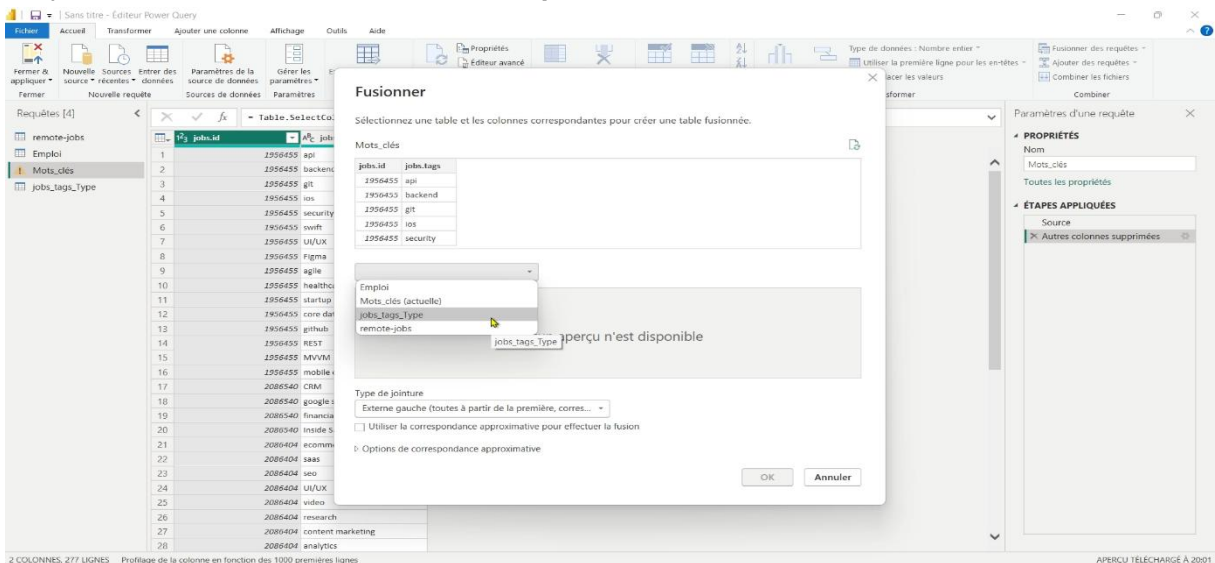
2. Sélectionnez votre fichier de classification.
3. Importez la table contenant vos données qualifiées. Nommez cette requête **job\_tag\_Type**.



## Étape 4.2 : Fusion des Requêtes (Merge)

Nous allons rapatrier la classification (Type) vers notre table de mots-clés issue de l'API.

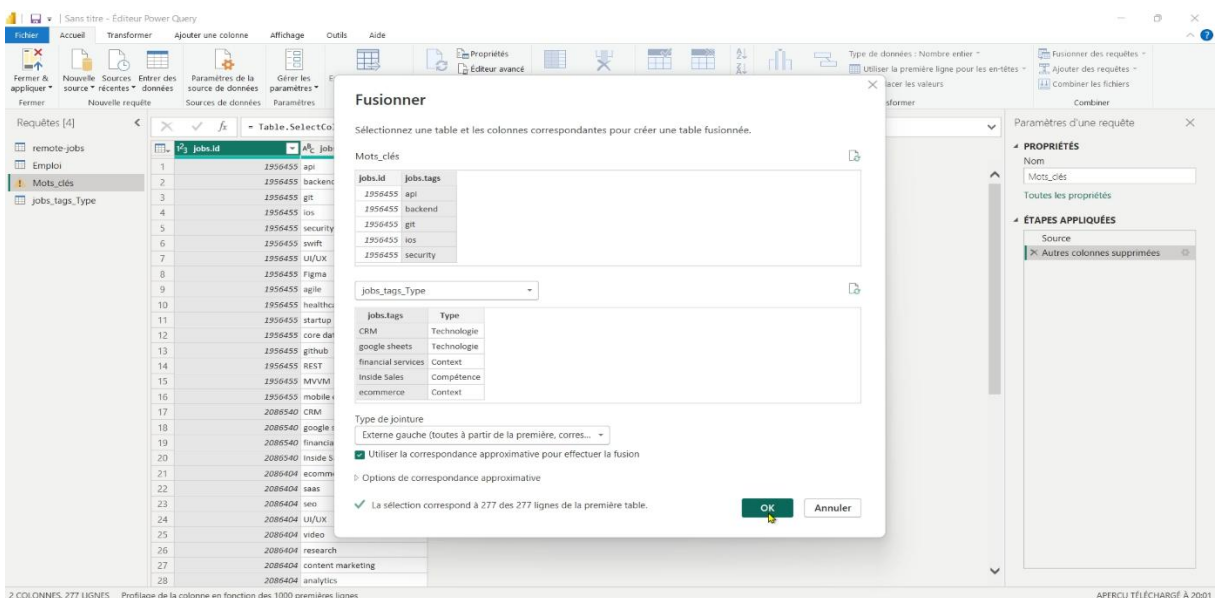
1. Sélectionnez la requête Mots\_clés dans le volet de gauche.
2. Cliquez sur le bouton **Fusionner des requêtes** (situé en haut à droite).



### 3. Configuration de la fusion :

- **Table 1 (Haut) :** Mots\_clés. Sélectionnez la colonne contenant le tag brut.
- **Table 2 (Bas) :** jobs\_tags\_Type. Sélectionnez la colonne correspondante (Tag).
- **Type de jointure :** Sélectionnez **Externe gauche** (Left Outer) afin que tous les tags sont conservés, même ceux qui n'ont pas encore été classifiés dans Excel.
- Cochez l'option **Utiliser la correspondance approximative pour effectuer la liaison** pour activé la **Gestion des variations orthographiques**.

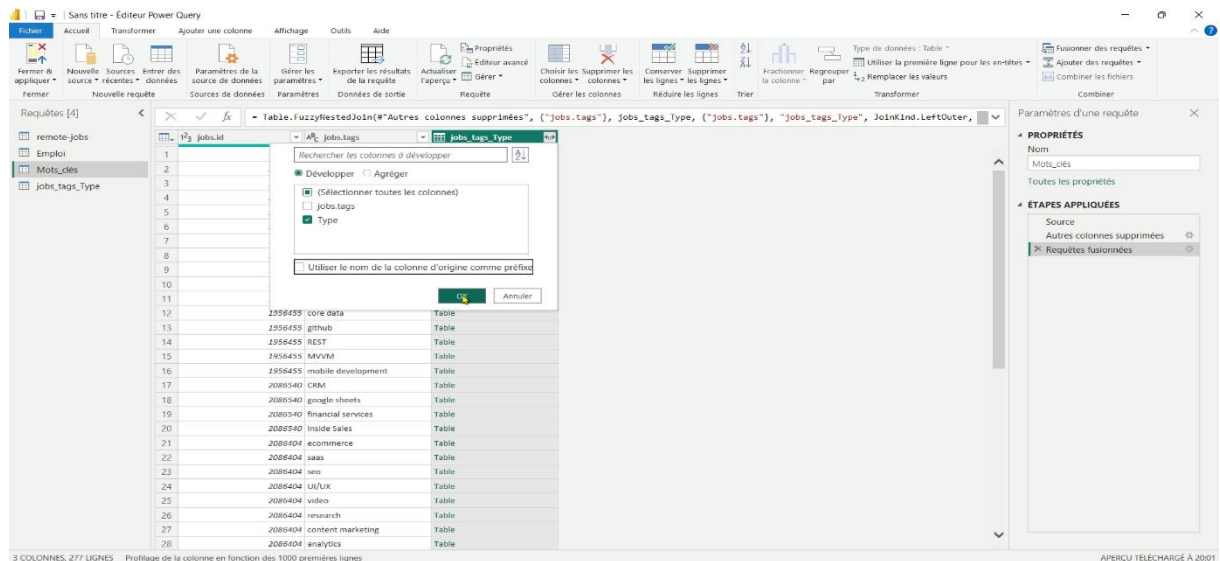
### 4. Validez par OK.





## Étape 4.3 : Expansion et Finalisation

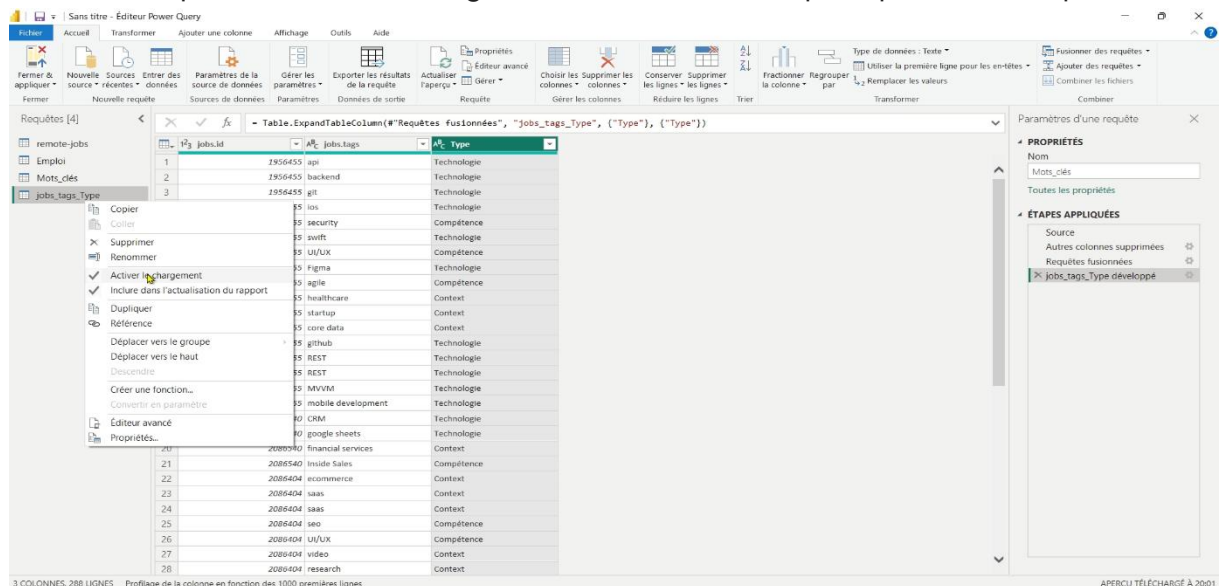
1. Une colonne `job_tag_Type` contenant des valeurs "Table" est apparue.
2. Cliquez sur l'icône d'expansion (double flèche) en haut à droite de cette colonne.
3. Sélectionnez uniquement le champ **Type** (la catégorie). Décochez "Utiliser le nom de la colonne d'origine comme préfixe".
4. Validez. La classification est maintenant intégrée.



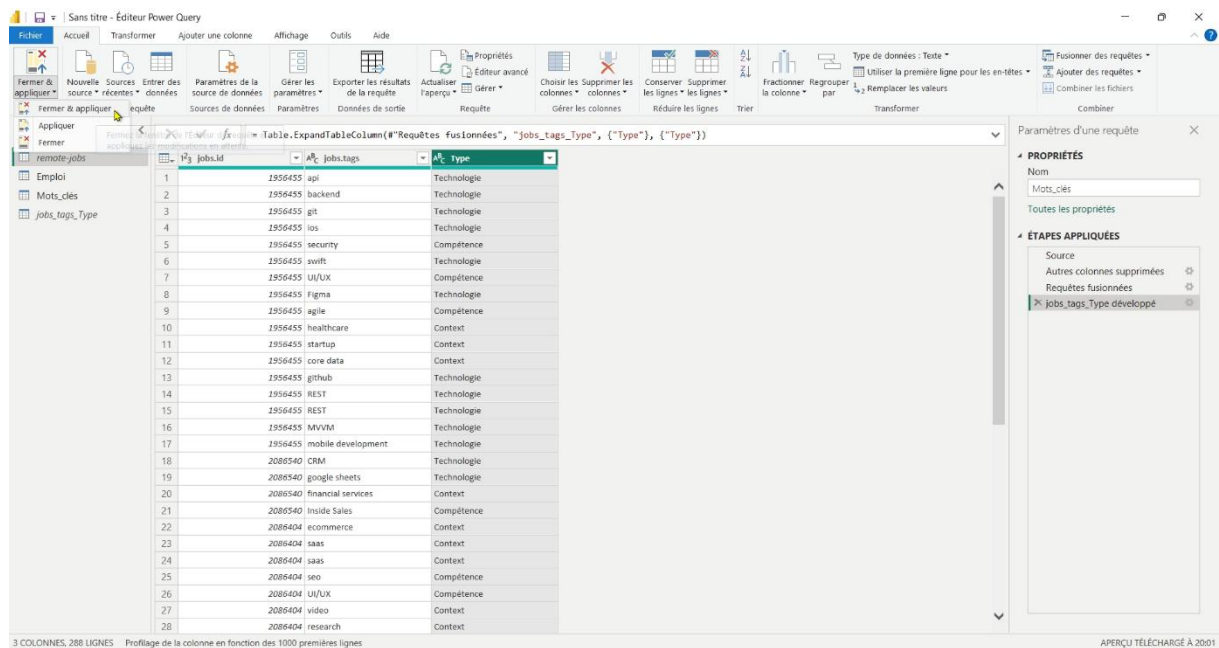
## Étape 4.4 : Optimisation du Chargement

Pour ne pas alourdir le modèle Power BI inutilement, nous ne chargerons que les tables finales.

1. Faites un **clic droit** sur la requête `job_tag_Type`.
2. Décochez l'option **Activer le chargement**. Le nom de la requête passe en italique.



3. Répétez l'opération pour la requête source remote-jobs.
4. Vérifiez que seules les tables Emploi et Mots\_clés sont activées (police normale).
5. Cliquez sur le bouton **Fermer et appliquer** (ruban Accueil, à gauche).



## Conclusion de la Procédure

Vous disposez désormais d'un modèle de données en étoile optimisé comprenant :

- Une **Table de Faits** (Emploi) centralisant les données des offres.
- Une **Table de Dimension** (Mots-Clés) segmentant les compétences par type.
- Une relation fonctionnelle permettant d'analyser le marché de l'emploi technologique.

**Jason ZBAKH** - Candidat Chef de Projet Data