

# 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

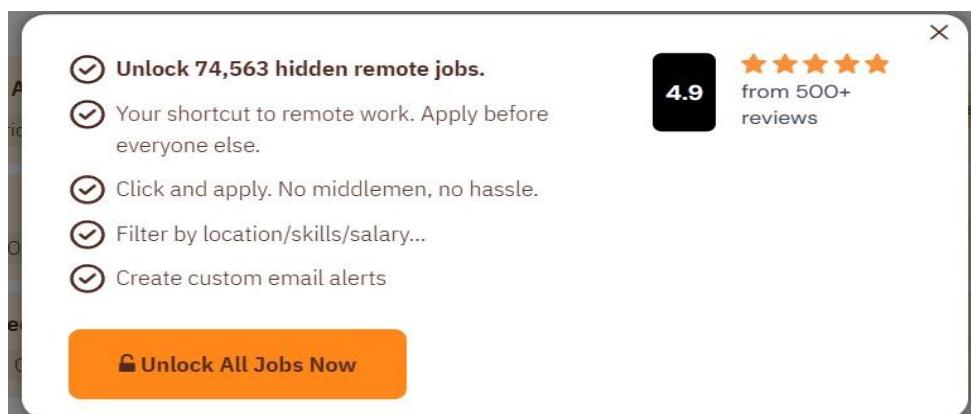
### La Veille Technologique par l'Analyse du Marché de l'emploi

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 Remotive.com

- **Spécialisation Tech** : Contrairement aux agrégateurs généralistes (Indeed, Monster), Remotive 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.

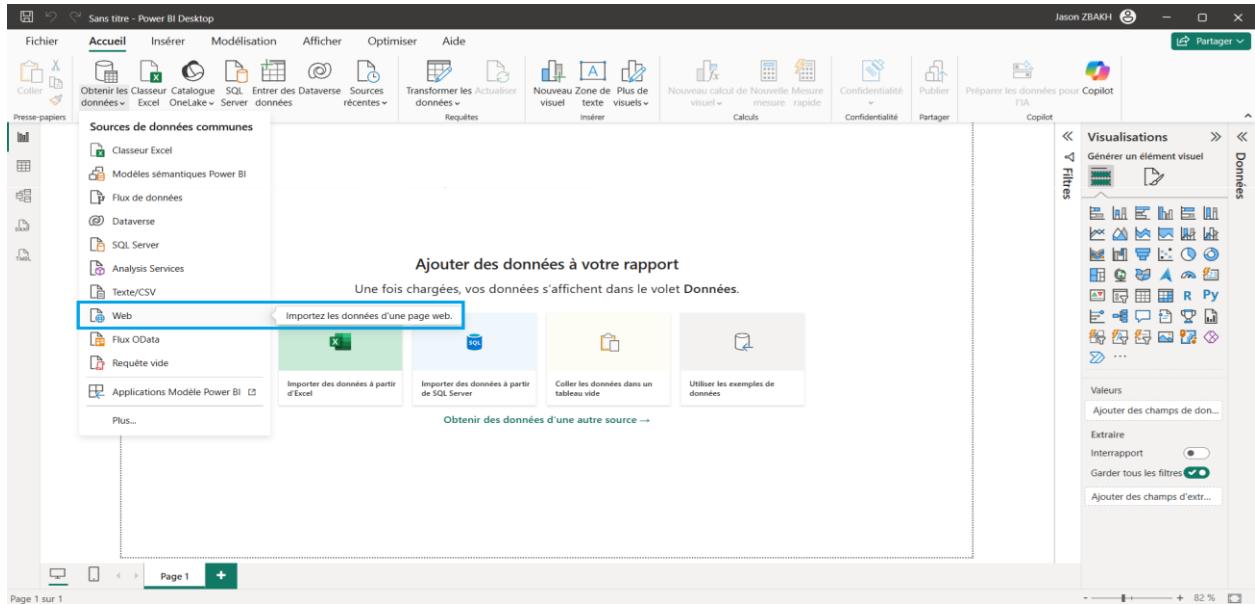
**Note sur les limitations (Version Gratuite)** : L'API publique limite le nombre d'offres retournées et la profondeur de l'historique. Pour un déploiement en production critique, l'acquisition d'un accès complet à la base de données historique est recommandée.



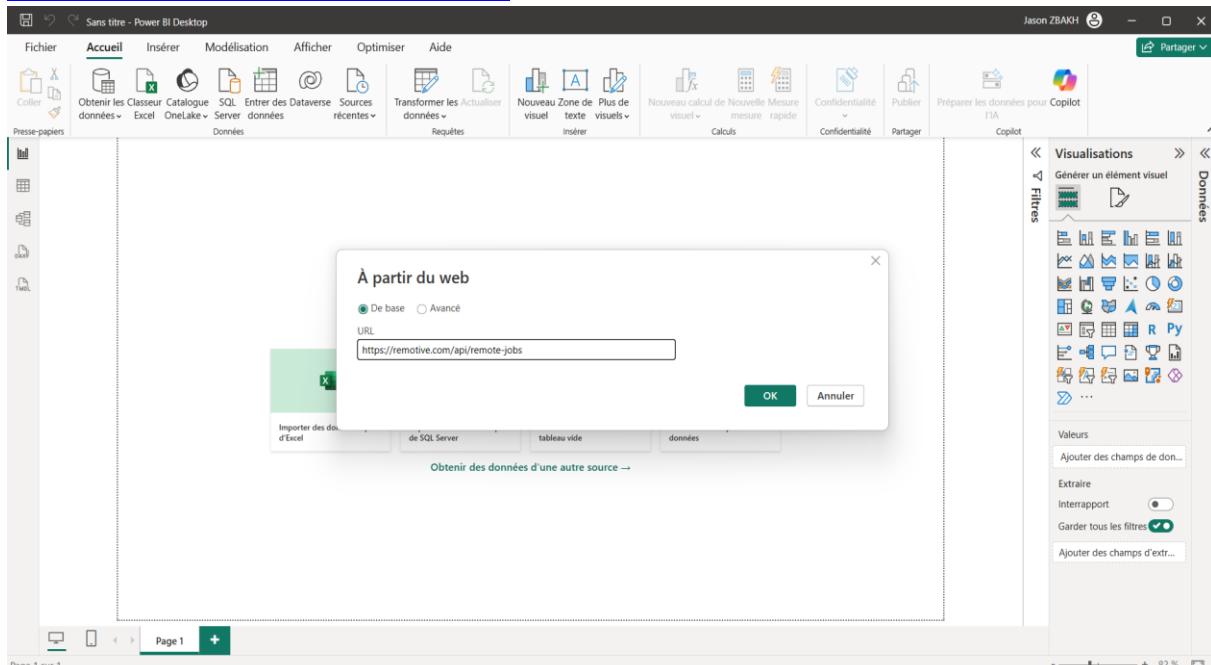
## 2. Procédure d'Importation et de Traitement (ETL)

### 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. On the left, the 'Requêtes' pane lists two items: 'remote-jobs' and 'remote-jobs (2)'. The 'remote-jobs' item has a context menu open with options like 'Copier', 'Coller', 'Supprimer', 'Renommer', etc. The 'remote-jobs (2)' item is selected. The main workspace displays a table with 17 columns and 261 rows, showing log entries from Remotive's API. The ribbon at the top has tabs like 'Fichier', 'Accueil', 'Transformer', etc. The 'Transformer' tab is active. The 'ÉTAPES APPLIQUÉES' pane on the right shows the steps applied to the source query, including 'Converti en tableau', 'jobs développé', 'jobs développé1', and 'jobs.tags développé'. The status bar at the bottom right says 'APERÇU TÉLÉCHARGÉ LE MERCREDI'.

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

This screenshot shows the same Power Query Editor interface after renaming the query. The 'remote-jobs (2)' entry in the 'Requêtes' pane has been renamed to 'Emploi'. The main workspace and ribbon remain the same, displaying the transformed table and log entries. The 'ÉTAPES APPLIQUÉES' pane still shows the applied steps. The status bar at the bottom right remains 'APERÇU TÉLÉCHARGÉ LE MERCREDI'.

## Nettoyage de la Table "Emploi"

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

The screenshot shows the Power Query Editor interface. The ribbon at the top has 'Transformer' selected. A context menu is open over the 'Choisir les colonnes...' option in the 'Transformer' tab. The main area displays a preview of a table with columns: '0-warning', '0-legal-notice', 'job-count', and 'total-job-count'. The preview shows 17 columns and 261 rows. The status bar at the bottom indicates '17 COLONNES, 261 LIGNES' and 'Profilage de la colonne en fonction des 1000 premières lignes'.

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.**

The screenshot shows the 'Choisir les colonnes' dialog box open in the Power Query Editor. It lists various columns from the source table, with checkboxes next to them. Most checkboxes are unchecked except for 'jobs.id', 'jobs.url', 'jobs.title', 'jobs.company\_name', 'jobs.company\_logo', 'jobs.category', 'jobs.job\_type', 'jobs.publication\_date', 'jobs.candidate\_required\_location', 'jobs.salary', and 'jobs.company\_logo\_url'. At the bottom of the dialog box are 'OK' and 'Annuler' buttons. The main Power Query Editor window shows the preview of the table with the selected columns.

- 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.

The screenshot shows the Power Query Editor interface with the 'Jobs' table selected. A context menu is open over the 'id' column header, listing various options like 'Copier', 'Supprimer', and 'Supprimer les doublons'. The 'Supprimer les doublons' option is highlighted with a blue selection bar. The main table view shows several columns including 'Jobs.id', 'Jobs.title', 'Jobs.company\_name', 'Jobs.company\_logo', and 'Jobs.category'. The status bar at the bottom indicates '11 COLONNES, 261 LIGNES' and 'APERÇU TÉLÉCHARGE LE MERCREDI'.

- 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.

The screenshot shows the Power Query Editor with the 'Mots\_clés' query selected. The 'Choisir les colonnes' dialog is open, displaying a list of columns from the 'remote-jobs' table. The 'jobs.tags' column is checked and highlighted with a green selection bar. The main table view shows the 'remote-jobs' table with 261 rows. The status bar at the bottom indicates '17 COLONNES, 261 LIGNES' and 'APERÇU TÉLÉCHARGE LE MERCREDI'.

### 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)
1. Technologique (L'Outillage)	<b>Critère : Instrumentalité.</b> Désigne un objet (logiciel, langage, infrastructure) que l'on <i>utilise, code ou 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>
2. Compétence (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>
3. Contexte (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

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

Sans titre - Éditeur Power Query

Ficher Accueil Transformer Ajouter une colonne Affichage Outils Aide

Nouvelle source... Entrer des données Sources Paramètres de la source de données Gérer les paramètres Exporter les résultats de la requête Actualiser l'aperçu Propriétés Éditeur avancé Choisir les Supprimer les colonnes Gérer les colonnes Conservé les lignes Supprimer les lignes Réduire les lignes Trier Transformer Type de données : Nombre entier Utiliser la première ligne pour les en-têtes 1 2 Remplacer les valeurs Fusionner des requêtes Ajouter des requêtes Combiner les fichiers Combiner

Requêtes

- remote
- Emploi
- Mots\_clés
- Analysis Services
- Texte/CSV
- Web
- Flux OData
- Requête vide
- Plus...

Les plus courantes

Sources de données

Paramètres d'une requête

Propriétés

ÉTAPES APPLIQUÉES

Source

Autres colonnes supprimées

jobs.id	jobs.tags
1	api
2	backend
3	git
4	ios
5	security
6	swift
7	UI/UX
8	Figma
9	agile
10	healthcare
11	startup
12	core data
13	github
14	REST
15	MVVM
16	mobile development
17	CRM
18	google sheets
19	financial services
20	Inside Sales
21	ecommerce
22	saaS
23	seo
24	UI/UX
25	video
26	research
27	content marketing
28	analytics

2 COLONNES, 277 LIGNES Profilage de la colonne en fonction des 1000 premières lignes APERÇU TÉLÉCHARGÉ À 2001

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

Sans titre - Éditeur Power Query

Ficher Accueil Transformer Ajouter une colonne Affichage Outils Aide

Nouvelle source... Entrer des données Sources de données

Requêtes [3]

- remote-jobs
- Emploi
- Mots\_clés

Navigateur

Options d'affichage

jobs\_tags\_Type

jobs.tags	Type
CRM	Technologie
google sheets	Technologie
financial services	Context
Inside Sales	Compétence
ecommerce	Context
saaS	Context
seo	Compétence
UI/UX	Compétence
video	Context
research	Context
content marketing	Compétence
analytics	Compétence
google analytics	Technologie
youtube	Context
onboarding	Compétence
ux research	Compétence
hubspot	Technologie
storyboard, storyboard	Compétence
storyboarding	Compétence
travel	Context
lever	Technologie
REST	Technologie
troubleshooting	Compétence
spreadsheets	Technologie

Paramètres d'une requête

ÉTAPES APPLIQUÉES

Autres colonnes supprimées

OK Annuler

2 COLONNES, 277 LIGNES Profilage de la colonne en fonction des 1000 premières lignes APERÇU TÉLÉCHARGÉ À 2001

## 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).

The screenshot shows the Power Query Editor interface with the 'Fusionner' (Merge) dialog open. The 'Mots\_clés' table is selected as the top table. The 'jobs\_tags\_Type' column from the 'jobs\_tags\_Type' table is chosen as the corresponding column for the merge. The 'Type de jointure' dropdown shows 'Externe gauche'. The 'Utiliser la correspondance approximative pour effectuer la liaison' checkbox is checked. The 'OK' button is visible at the bottom right.

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

The screenshot shows the Power Query Editor interface with the 'Fusionner' (Merge) dialog open. The 'jobs\_tags\_Type' table is selected as the bottom table. The 'jobs.tags' column is chosen as the corresponding column. The 'Type de jointure' dropdown shows 'Externe gauche'. The 'Utiliser la correspondance approximative pour effectuer la liaison' checkbox is checked. A message at the bottom states 'La sélection correspond à 277 des 277 lignes de la première table.' The 'OK' button is highlighted.

## Étape 4.3 : Expansion et Finalisation

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

The screenshot shows the Power Query Editor interface. A context menu is open over the 'Type' column header, with the 'Type' option selected. Below it, there is a checkbox labeled 'Utiliser le nom de la colonne d'origine comme préfixe'. The ribbon at the top has tabs for 'Fichier', 'Accueil', 'Transformer', 'Ajouter une colonne', 'Affichage', 'Outils', and 'Aide'. On the right side, there are several toolbars and a sidebar titled 'Paramètres d'une requête' which includes sections for 'PROPRIÉTÉS' (Nom: Mots\_clés), 'ÉTAPES APPLIQUÉES' (Source: Autres colonnes supprimées), and 'APERÇU TÉLÉCHARGÉ À 2001'.

## Étape 4.4 : Optimisation du Chargement

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

- Faites un **clic droit** sur la requête job\_tag\_Type.
- Décochez l'option **Activer le chargement**. Le nom de la requête passe en italique.

The screenshot shows the Power Query Editor interface. A context menu is open over the 'job\_tags\_Type' query in the list of queries. The 'Activer le chargement' checkbox is checked. The 'job\_tags\_Type' query is now italicized. The ribbon at the top has tabs for 'Fichier', 'Accueil', 'Transformer', 'Ajouter une colonne', 'Affichage', 'Outils', and 'Aide'. On the right side, there are several toolbars and a sidebar titled 'Paramètres d'une requête' which includes sections for 'PROPRIÉTÉS' (Nom: Mots\_clés), 'ÉTAPES APPLIQUÉES' (Source: Autres colonnes supprimées), and 'APERÇU TÉLÉCHARGÉ À 2001'.

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).

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table with three columns: 'jobs.Id', 'jobs.tags', and 'Type'. The 'Type' column contains values like 'Technologie', 'Compétence', and 'Context'. The ribbon at the top has 'Fermer & appliquer' selected. On the right, the 'ÉTAPES APPLIQUÉES' pane shows a single step named 'jobs\_tags\_Type développé'. The status bar at the bottom right indicates 'APERÇU TÉLÉCHARGÉ À 2001'.

## 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**