

# Module 2 – Connexion aux données

**Objectifs du module** - Accéder à l'éditeur Power Query et comprendre son rôle - Connaître le catalogue des sources de données de Power BI - Se connecter à des fichiers plats (txt, csv, xls), un dossier, une base de données (Access, SQL Server, Oracle, SAP) et un site Web - Choisir entre mode **Import** et **DirectQuery**

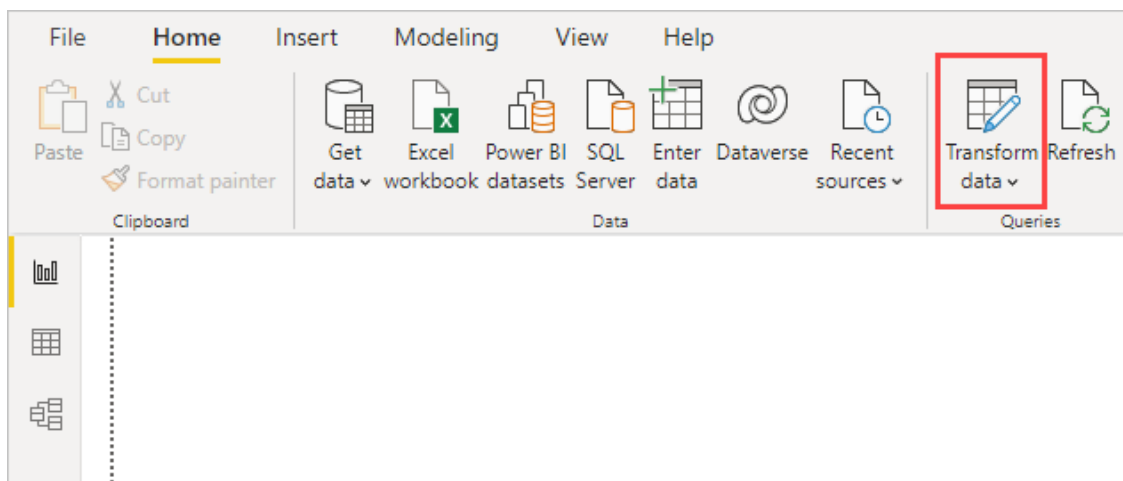
Durée estimée : 3 h | Niveau : Débutant Fil rouge : connexion aux 6 fichiers NovaRetail - -

## 2.1 Power Query : la porte d'entrée des données

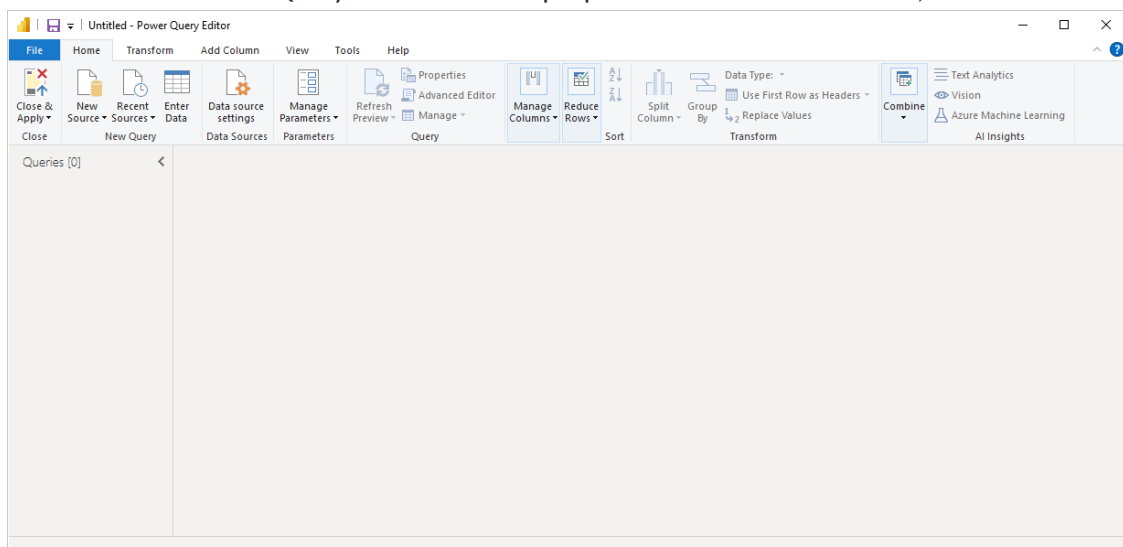
Avant d'alimenter un rapport, il faut **extraire** les données. C'est le rôle de **Power Query**, le moteur ETL (*Extract, Transform, Load*) intégré à Power BI Desktop (et aussi présent dans Excel).

### Accéder à l'éditeur Power Query

Depuis Power BI Desktop : ruban **Accueil** > **Transformer les données**.



L'éditeur Power Query s'ouvre dans sa propre fenêtre. Sans connexion, il est vide :

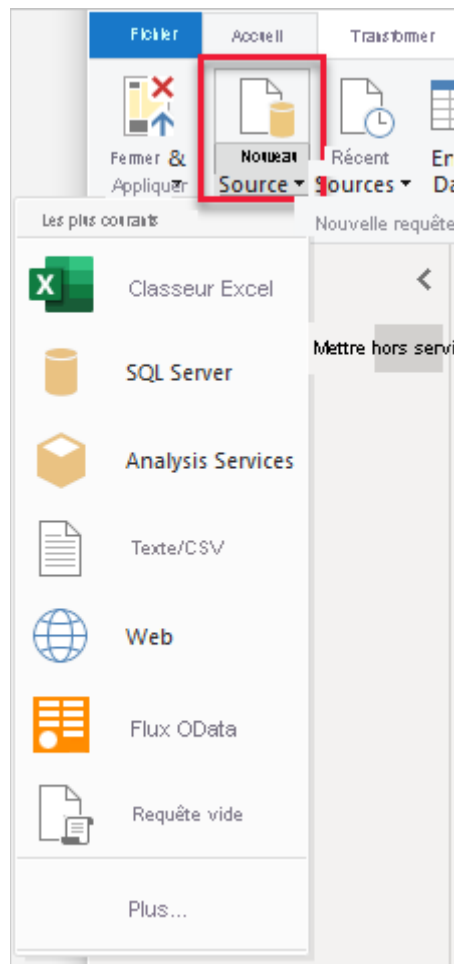


**À retenir :** chaque opération effectuée dans Power Query est enregistrée comme une *étape* reproductible. Les données sources ne sont **jamais modifiées** : on applique une “recette” sur leur copie.

Référence : [Vue d'ensemble des requêtes — Microsoft Learn](#) - -

## 2.2 Le catalogue des sources de données

Ruban Accueil > Nouvelle source fait apparaître le menu des sources courantes :



### Panorama professionnel (extrait)

| Catégorie              | Sources                                                                                                           | Connecteur |
|------------------------|-------------------------------------------------------------------------------------------------------------------|------------|
| Fichiers               | Excel (.xls/.xlsx), texte/CSV, XML, JSON, dossier complet                                                         | natif      |
| Bases de données       | SQL Server, Oracle, MySQL, PostgreSQL, IBM Db2, Sybase, Teradata, SAP HANA, SAP Business Warehouse, <b>Access</b> | natif      |
| Power Platform / Azure | Dataaverse, Azure SQL, Synapse, Blob Storage...                                                                   | natif      |
| Services en ligne      | SharePoint, Salesforce, Google Analytics, Dynamics 365...                                                         | natif      |

| Catégorie | Sources                                                         | Connecteur |
|-----------|-----------------------------------------------------------------|------------|
| Web       | Page Web (scraping de tableaux),<br>API REST via connecteur Web | natif      |
| Autres    | ODBC, OLE DB, script R/Python,<br>requête vierge                | avancé     |

Liste exhaustive et officielle : [Sources de données Power BI Desktop](#)

Plus de **200 connecteurs** sont disponibles. Si aucun ne convient : ODBC ou développement d'un connecteur personnalisé. - -

## 2.3 Import ou DirectQuery : le choix stratégique

Au moment de se connecter (surtout aux bases de données), deux modes principaux :

| Critère       | Import (défaut)                             | DirectQuery                                             |
|---------------|---------------------------------------------|---------------------------------------------------------|
| Données       | Copie en mémoire (moteur VertiPaq)          | Restent dans la source ; requêtes envoyées à la volée   |
| Vitesse       | Très rapide                                 | Dépend de la source et du réseau                        |
| Volume        | Jusqu'à ~1 Go+ par modèle (plus en Premium) | Illimité côté source                                    |
| Actualisation | Manuelle ou planifiée                       | Temps quasi réel                                        |
| Langage M     | Toutes transformations autorisées           | Transformations limitées ("query folding")              |
| DAX           | Complet                                     | Sous-ensemble + restrictions                            |
| Cas typique   | 90 % des projets (dont NovaRetail)          | Très gros volumes, données temps réel, contraintes RGPD |

**Réflexe senior:** par défaut → **Import**. On ne passe en DirectQuery que pour une raison précise (volume colossal, fraîcheur seconde, interdiction de copier les données). Le mode **composite** permet même de mélanger les deux. > Approfondir : [Mode DirectQuery](#) - -

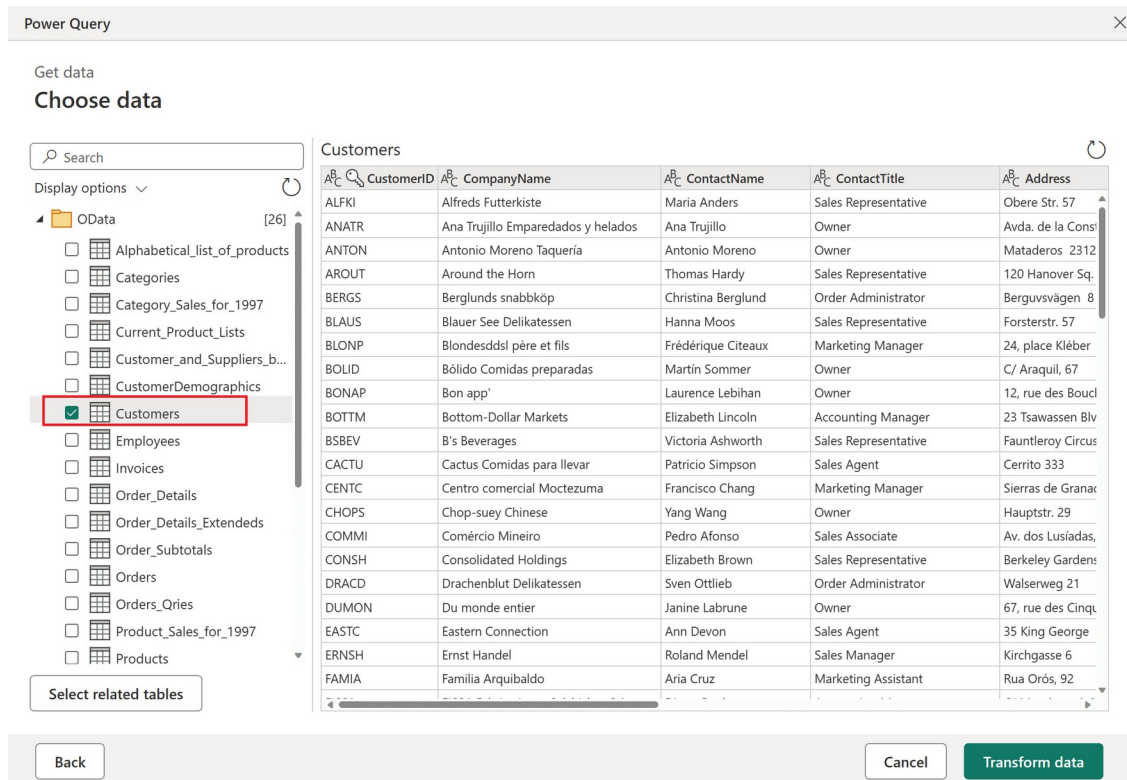
## 2.4 Connexion pas à pas : le flux commun à toutes les sources

Quel que soit le connecteur, le schéma mental est toujours le même :

Nouvelle source → Paramètres de connexion → Aperçu (Navigateur)  
→ [Transformer les données] ou [Charger]

1. Accueil > Nouvelle source > type choisi
2. Renseigner les paramètres (chemin, serveur, identifiants...)
3. Valider les **identifiants** (Windows, base, anonyme...) — mémorisés par source
4. Dans le **Navigateur**, cocher les tables/vues souhaitées
5. Choisir :

- **Charger** → directement dans le modèle ;
  - **Transformer les données** → passer d'abord par Power Query (*le réflexe pro!*).



## 2.5 Les familles de sources en détail

### A. Fichier texte / CSV (.txt, .csv)

Le format d'échange le plus répandu (exports ERP, CRM...). Points de vigilance professionnels :

| Vigilance  | Explication                                                                |
|------------|----------------------------------------------------------------------------|
| Séparateur | , ; tabulation... dépend de la locale ! Un CSV français utilise souvent ;  |
| Encodage   | Préférer UTF-8 ; sinon accents cassés (Ã©...)                              |
| Origine    | Fichier > Origine = encodage ; Paramètres régionaux = format dates/nombres |
| En-têtes   | Vérifier que la 1 <sup>re</sup> ligne devient bien l'en-tête               |

**Démo:** Nouvelle source > Texte/CSV > ventes\_2024.csv → Vérifiez l'aperçu : séparateur , ;, encodage UTF-8, types détectés automatiquement. → Cliquez sur **Transformer les données**.

### B. Fichier Excel (.xls / .xlsx)

- Le navigateur liste **feuilles** et **tables nommées**.

- Bonne pratique : nommer explicitement ses plages en tant que *Tableaux* dans Excel (**Ctrl+T**) avant import.
- Éviter : cellules fusionnées, totaux insérés au milieu des données, lignes de titre décoratives — ce sont des bombes à retardement.

### C. Un dossier entier (combiner N fichiers)

Cas d'usage n°1 du monde réel : un export mensuel par fichier (*ventes\_2023.csv*, *ventes\_2024.csv*, demain *ventes\_2025.csv*...). 1. **Nouvelle source > Dossier** → pointer vers le dossier *donnees*. 2. Bouton **Combiner et transformer** → Power Query : - crée une fonction d'exemple, - concatène tous les fichiers compatibles en une seule table, - ajoute une colonne *Source.Name* (utile pour tracer l'origine). 3. Ajouter un nouveau fichier dans le dossier ⇒ il sera pris en compte à la prochaine actualisation. Magique.

Nous utiliserons cette technique au module 4 (fusion multi-tables) pour empiler 2023 + 2024.

### D. Bases de données : Access, SQL Server, Oracle, SAP

| SGBD            | Paramètres requis                                                        | Remarques pro                                                                             |
|-----------------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| Access (.accdb) | Chemin du fichier                                                        | Idéal en TP/pédagogie ; limité en production                                              |
| SQL Server      | Serveur (+ instance), base optionnelle ; authentification Windows ou SQL | La plus utilisée en entreprise. Préférer des <b>vues</b> SQL plutôt que des tables brutes |
| Oracle          | Nom de service TNS ou connecteur EZConnect + pilote installé             | Installer le pilote Oracle Client (64 bits) sur le poste                                  |
| SAP HANA        | Serveur, port, identifiants ; choix cube (BW) ou vue calculée            | Licence & pilotes requis ; souvent piloté par l'équipe IT                                 |
| SAP BW          | Serveur applicatif/message server, infosystème                           | Navigation par InfoCubes/BEx queries                                                      |

#### Démo guidée SQL Server:

Nouvelle source > SQL Server  
 Serveur : SRVSQL01\PROD (ou localhost pour un TP local)  
 Base : NOVARETAIL\_DWH  
 Mode : Import  
 Identifiants : Authentification Windows (ou SQL : login/mot de passe)  
 Navigateur : cocher vues *dbo.v\_ventes*, *dbo.dim\_produits*...  
 >Transformer les données

En entreprise, l'accès direct aux serveurs passe souvent par une **passerelle de données** (on-premises data gateway) quand le Service doit actualiser : sujet traité au module 7.

### E. Site Web (Web connector)

Deux usages : 1. **Scraping de tableau HTML**: **Nouvelle source > Web** → URL → Power Query détecte les tableaux de la page → choisir celui à importer. Parfait pour des données publiques (taux de change, INSEE, bourse...). 2. **API/Web brut**: récupération du contenu brut (JSON/CSV) puis transformation M.

**Démo**: essayez avec une page contenant un tableau public, par exemple une page Wikipédia listant des

données statistiques : Nouvelle source > Web > [https://fr.wikipedia.org/wiki/D%C3%A9mographie\\_de\\_la\\_France](https://fr.wikipedia.org/wiki/D%C3%A9mographie_de_la_France)

**Cadre légal:** respecter les conditions d'utilisation des sites et le RGPD ; privilégier les API officielles quand elles existent. - -

## 2.6 Gérer les sources existantes

- **Modifier une source:** Power Query > Accueil > Paramètres de la source de données (chemins, identifiants).
- **Basculer Import ↔ DirectQuery:** Fichier > Options > Fichier actuel > Chargement des données.
- **Chemins relatifs (astuce pro):** éviter C:\Users\Moi\... codé en dur ; utiliser un paramètre de chemin ou OneDrive/SharePoint pour la portabilité. - -

## 2.7 Démo guidée – Fil rouge NovaRetail : brancher les données

**Mission 2 : connecter les 6 fichiers** 1. Ouvrez **NovaRetail.pbix** (module 1). 2. Accueil > Obtenir les données > Texte/CSV → **produits.csv** → **Transformer les données**. - Contrôlez l'aperçu : séparateur ; OK, en-têtes OK. - Renommez la requête **Produits** (volet Paramètres d'une requête, zone Nom). 3. Répétez pour **magasins.csv** (requête **Magasins**) et **clients.csv** (requête **Clients**), puis **budget\_2024.csv** (requête **Budget**). 4. Pour **ventes\_2023.csv** et **ventes\_2024.csv**: importez-les aussi (requêtes **Ventes2023**, **Ventes2024**). *Nous les empilerons au module 4*. 5. Dans le volet gauche de Power Query, vous avez maintenant **6 requêtes**. 6. Accueil > Fermer et appliquer.

**Résultat attendu:** retour dans la vue Rapport, le volet Données affiche 6 tables. Dans la vue Modèle, Power BI a peut-être créé quelques relations automatiquement — nous les auditerons au module 5.

Contrôle qualité express : ouvrez la vue Table et cliquez sur chaque table pour vérifier les lignes chargées. - -

## 2.8 Exercices

### Exercice 2.1 – QCM sources & modes

1. Quel bouton ouvre l'éditeur Power Query ?
2. Citez 2 différences majeures entre Import et DirectQuery.
3. Votre export mensuel arrive sous forme de 24 fichiers **ventes\_AAAAMM.csv**. Quelle stratégie ?
4. Pourquoi préférer une **vue** SQL à une table brute ?
5. Un CSV importé affiche **Ã©** à la place de **é**. Quelle est la cause ? La solution ?

### Exercice 2.2 – TP guidé : connexion complète

Reprenez la démo 2.7 et connectez les 6 fichiers NovaRetail. Puis : a) Renommez chaque requête sans accent ni espace (**Ventes2023...**). b) Dans Accueil > Paramètres de la source de données, vérifiez les chemins listés. c) Fermez et appliquez, notez le nombre de lignes par table (vue Table).

## Exercice 2.3 – Exploration Web

Connectez-vous à une page Wikipédia contenant un tableau (ex. démographie de la France). Importez un tableau dans une nouvelle requête **DemoWeb**. Que remarque-t-on sur les types de colonnes ? \* (Bonus pro : créez un dossier **archives/**, déplacez-y une copie de **ventes\_2023.csv**, et testez le connecteur Dossier sur ce dossier.)\* - -

## Corrigés

### Corrigé 2.1

1. **Accueil > Transformer les données** (ruban Power BI Desktop).
2. Import = copie en mémoire, rapide mais figée jusqu'à actualisation, volume limité ; DirectQuery = données restant en source, quasi temps réel, volume illimité mais performances liées à la source et transformations/DAX restreints.
3. Le connecteur **Dossier + Combiner et transformer**: une seule requête absorbe les 24 fichiers, et les futurs fichiers seront inclus à l'actualisation.
4. Une vue encapsule la logique métier côté base (filtrage, jointures, renommages), réduit les volumes transférés et évite que des changements de structure de table cassent le rapport.
5. Cause : mauvais **encodage** (fichier ANSI lu comme UTF-8 ou inverse). Solution : dans l'étape Source de Power Query, régler **Origine** sur l'encodage correct (souvent **1252 : Europe de l'Ouest** pour du français legacy, ou UTF-8).

### Corrigé 2.2

- a) Sélectionner la requête > volet *Paramètres d'une requête* > champ *Nom* (ou double-clic sur la requête). Sans espaces/accents : **Ventes2023**, **Ventes2024**, **Produits**, **Magasins**, **Clients**, **Budget**.
- b) La boîte liste toutes les sources du fichier avec leurs chemins — c'est ici qu'on change un chemin si le dossier a bougé.
- c) Valeurs de référence attendues : **Produits** ≈ 21 lignes · **Magasins** = 13 · **Clients** = 80 · **Budget** = 4 · **Ventes2023** ≈ 907 (avec doublons volontaires!) · **Ventes2024** = 1050.

### Corrigé 2.3

Les colonnes arrivent majoritairement en type **texte** (« Tout objet » / Any) car le scraping HTML ne devine pas les types : il faut typer manuellement (années → Nombre entier, populations → Nombre...). On voit aussi parfois des notes de bas de tableau collées aux valeurs : nettoyage nécessaire (module 3!). -

## Les pièges du professionnel

1. **Chemins absolus personnels** (C:\Users\Pierre\...) : le rapport casse dès qu'on le transmet. → Utilisez un dossier partagé/SharePoint ou un paramètre.
2. **Cliquer « Charger » trop vite**: passez TOUJOURS par « Transformer les données » pour auditer avant chargement.
3. **Oublier les credentials expirés**: mot de passe SQL changé ⇒ actualisation KO. Centralisez dans *Paramètres de la source de données*.
4. **DirectQuery “au cas où”**: performances catastrophiques si mal maîtrisé. Import d'abord !

5. **Importer des tables entières géantes** alors que 5 colonnes suffisent : supprimez/filtrez tôt (bonne pratique *early filtering*). - -

## Auto-évaluation du module 2

| # | Question                             | A                     | B                   | C                            |
|---|--------------------------------------|-----------------------|---------------------|------------------------------|
| 1 | Power Query est...                   | un moteur ETL intégré | un site web         | un langage graphique fermé   |
| 2 | Le mode recommandé par défaut est... | DirectQuery           | Live                | Import                       |
| 3 | Pour empiler des exports mensuels... | 24 imports manuels    | connecteur Dossier  | copier-coller Excel          |
| 4 | Les étapes Power Query...            | modifient la source   | sont reproductibles | disparaissent à la fermeture |

- (Réponses : 1-A, 2-C, 3-B, 4-B)\* - -

## Ressources du module 2

- [Se connecter aux données dans Power BI Desktop](#)
- [Sources de données prises en charge](#)
- [Combiner des fichiers binaires \(dossier\)](#)
- [DirectQuery : présentation approfondie](#)
- Module Learn : [Nettoyer, transformer et charger des données dans Power BI](#)

**Module suivant : Préparation et nettoyage — on met les mains dans le cambouis NovaRetail!**