

A. Objectif

Ce premier chapitre permet tout d'abord de vous familiariser et d'apporter des connaissances concrètes en matière de manipulation de données avec Power Query.

Tout au long de ce chapitre, des exemples d'application simples seront proposés afin d'illustrer les notions abordées.

Enfin, un cas pratique de mise en application inspiré d'une expérience professionnelle réelle, sera proposé.

Les fichiers nécessaires à la réalisation des exemples et du cas pratique sont disponibles en téléchargement.

B. Power Query, un outil pour nettoyer et manipuler les données

Power Query est un outil de Microsoft spécialisé dans la manipulation et la transformation de données.

Pour l'analyste, cette étape est probablement la plus importante car sans données structurées, rien n'est possible.

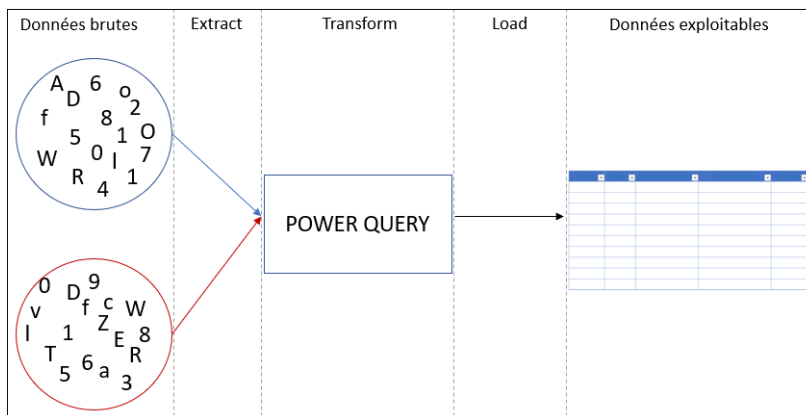
Cet outil est disponible dans Excel à partir sa version 2010 mais aussi dans Power BI Desktop, Microsoft Flow, Common Data Service et Analysis Service.

Lorsque Microsoft a commencé à intégrer des outils de business intelligence dans Excel, Power Query a été quelque peu négligé par les utilisateurs, ces derniers préférant la plupart du temps les possibilités de calcul spectaculaires du langage DAX ou bien les nouvelles fonctionnalités en matière de graphique.

C'est à mon avis une erreur, Power Query est probablement l'outil le plus utile depuis l'arrivée des tableaux croisés dynamiques en 1993.

Power Query est ce que l'on appelle un ETL (*Extract, Transform, Load*).

- ▶ **Extraction** : se connecter à une ou plusieurs sources pour en extraire les données souhaitées.
- ▶ **Transformer** : nettoyer, mettre en forme, agréger les données.
- ▶ **Load** : importer les données, transformées ou non dans une location spécifique afin de les analyser.



De fait, depuis des décennies, les utilisateurs d'Excel manipulent des données à la manière d'un ETL, cependant jusque-là ces tâches étaient réalisées soit :

- ▶ manuellement, ce qui engendre à la fois une perte de temps et un risque d'erreur,
- ▶ grâce aux formules d'Excel et/ou au code VBA. Excel est un outil puissant offrant une grande liberté à l'utilisateur. Toutefois, cette liberté peut aussi être un piège. Les formules complexes et le code demandent une certaine expertise qui ne s'improvise pas forcément.

Power Query propose à la fois de simplifier et d'automatiser l'importation et la manipulation des données.

Concrètement, Power Query permet :

- ▶ de se connecter à un grand nombre de sources de données,
- ▶ de filtrer, adapter et transformer les données brutes pour permettre une analyse et une modélisation poussées (conjointement avec Power Pivot qui sera étudié au chapitre suivant),
- ▶ de créer des procédures stockées pour automatiser la préparation des données.

C. Première prise en main

L'objectif de cette section est de réaliser un premier tour d'horizon du fonctionnement de Power Query à travers un exemple simple. L'ensemble des fonctionnalités sera étudié plus en détail dans les sections suivantes.

Présentation des données

Un refuge pour animaux dispose d'un petit tableau de données comportant le nom, l'espèce, la date d'accueil et l'âge de plusieurs animaux.

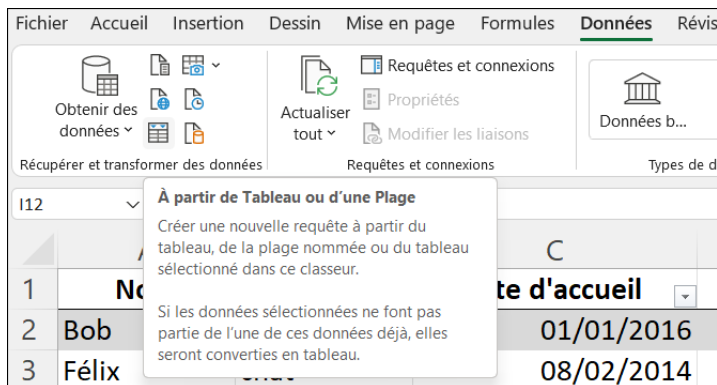
Les données se présentent sous cette forme :

	A	B	C	D
1	Nom	Espèce	Date d'accueil	Age
2	Bob	chien	01/01/2016	5
3	Félix	chat	08/02/2014	8
4	Yvan	écureuil	16/08/2015	3

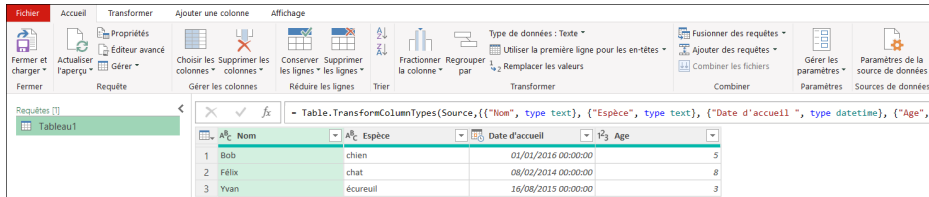
Vous retrouverez les données de cet exemple dans le fichier 1_animaux.xlsx.

1. Importation des données dans Power Query

- 🔗 Ouvrez le fichier 1_animaux.xlsx.
- 🔗 Sélectionnez une cellule contenue dans le tableau de données, par exemple la cellule A2.
- 🔗 Dans l'onglet **Données** du ruban, dans le groupe **Récupérer et transformer des données**, cliquez sur le bouton **À partir de Tableau ou d'une Plage**.



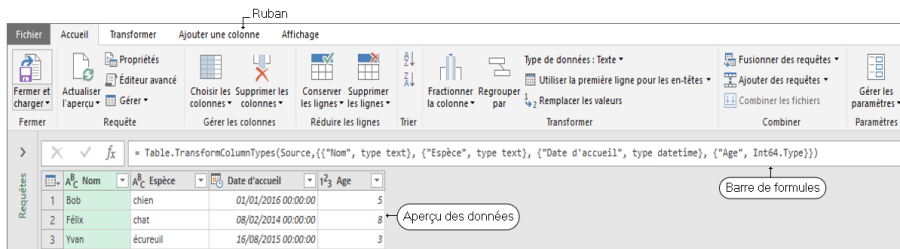
Une nouvelle fenêtre s'ouvre, il s'agit de l'éditeur Power Query.



2. Présentation de l'interface

a. Premier aperçu de l'éditeur Power Query

L'éditeur de requête se présente de la manière suivante :



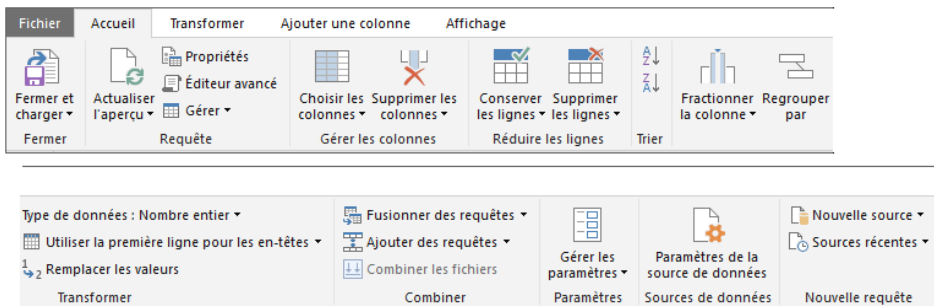
Le ruban

En haut de l'écran se trouve un ruban ressemblant à celui qui se trouve dans Excel. Il est composé de cinq onglets :

Accueil

Cet onglet contient les outils de transformation les plus courants.

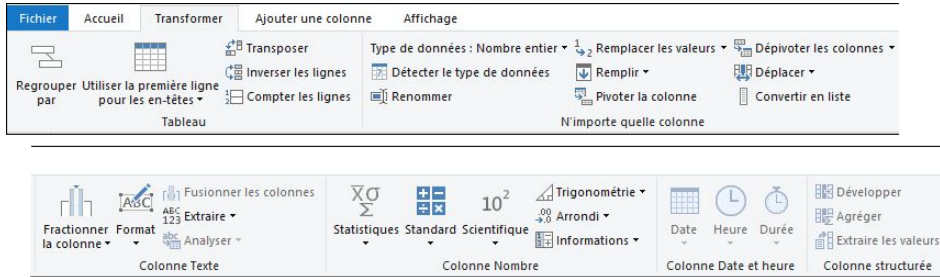
Il est possible entre autres de supprimer des colonnes, supprimer des lignes, trier, grouper, combiner les données, séparer des colonnes.



Transformer

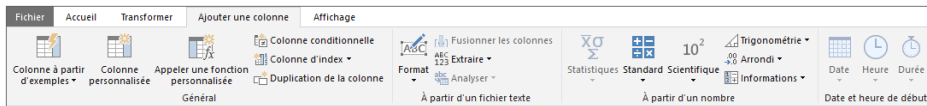
Cet onglet comprend des outils qui permettent de modifier les colonnes existantes.

Ainsi, il est possible notamment de scinder ou grouper des colonnes, transposer du texte, d'utiliser des fonctions de calcul ou de manipulation des dates et heures.



Ajouter des colonnes

Dans cet onglet se trouvent des outils permettant d'ajouter différents types de colonnes : colonnes conditionnelles, duplication, date et heure, etc.



Certains outils sont communs aux onglets **Transformer** et **Ajouter des colonnes** comme par exemple les fonctions de calcul ou de manipulation des dates ou heures.

La barre de formule

La barre de formule contient du code M.

☞ Si la barre de formule n'est pas affichée, allez dans l'onglet **Affichage**, dans le groupe **Disposition**, cochez **Barre de Formule**.

M est un langage de programmation développé par Microsoft et spécialement conçu pour la manipulation et la modélisation des données.

Après l'importation du tableau de données dans l'éditeur Power Query, la barre de formule contient le code suivant :

```
= Table.TransformColumnTypes(Source,{{"Nom", type text}, {"Espèce", type text}, {"Date d'accueil", type datetime}, {"Age", Int64.Type}})
```



Notez que cette ligne de code a été créée automatiquement, un peu de la même manière que l'enregistreur de macro VBA d'Excel.

Power Query traduit automatiquement chaque action que vous effectuez sur une table en code M.

L'aperçu des données

Les données sont affichées sous forme tabulaire au centre de l'interface, de manière similaire à Excel :

	A ^B _C Nom	A ^B _C Espèce	Date d'accueil	1 ² ₃ Age
1	Bob	chien	01/01/2016 00:00:00	5
2	Félix	chat	08/02/2014 00:00:00	8
3	Yvan	écureuil	16/08/2015 00:00:00	3

À gauche de l'intitulé de chacune des colonnes se trouve un symbole renseignant sur le type de donnée identifié et utilisé par Power Query.

Ainsi, les colonnes **Nom** et **Espèce** ont été paramétrées en format **Texte**, la colonne **Date d'accueil** a été paramétrée en format **Date/Heure** et la colonne **Age** en format **Nombre décimal**.

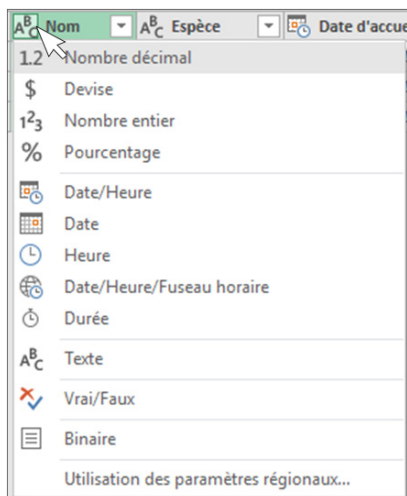
Ce paramétrage se retrouve dans le code M au niveau de la barre de formule :


```
= Table.TransformColumnTypes(Source,{{"Nom", type text}, {"Espèce", type text}, {"Date d'accueil ", type datetime}, {"Age", Int64.Type}}).
```

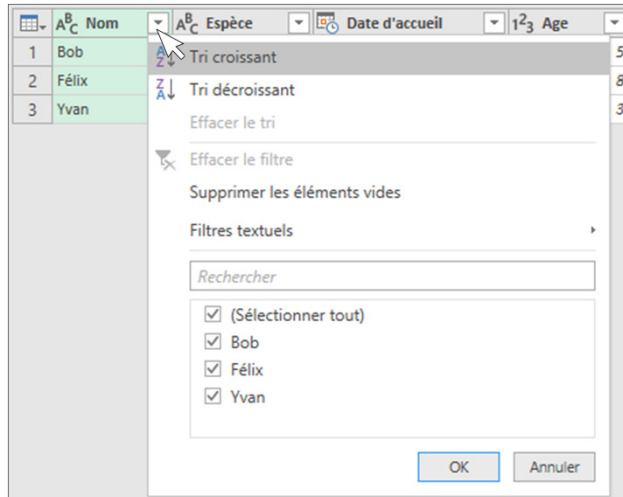
Il est possible de modifier le format octroyé par Power Query aux données de la colonne.

🖱️ Au niveau de la colonne **Nom**, cliquez sur le symbole .

Plusieurs choix sont alors possibles :



- ☞ Cliquez ensuite sur le bouton  sur le côté droit de chacune des colonnes. Il permet de réaliser des **Tri** et des **Filtres**, de la même manière que les filtres existant dans les tableaux de données d'Excel.



b. Le volet Paramètres d'une requête

Le volet **Paramètres d'une requête** se situe sur le côté droit de l'éditeur Power Query.

