

 + QUIZ

Version en ligne

OFFERTE !

pendant 1 an

Power Query et M

Extraire et préparer les données en vue
de leur exploitation dans Excel ou Power BI

André MEYER-ROUSSILHON

En téléchargement



fichiers sources
et exemples



Solutions Business



Chapitre 1

Introduction

A. Power Query, M et les produits Microsoft	11
B. À quoi sert Power Query ?	12
C. À qui s'adresse ce livre et comment est-il structuré ?	13
D. Aperçu général de l'interface de Power Query	15
1. Les rubans	15
2. Les requêtes	16
3. La zone centrale	16
4. La barre de formule	17
5. Les étapes	17
E. Power Query et le modèle de données	18
F. Le modèle en étoile	20
G. Power Query et le modèle composite	21
H. Les ressources	22
I. Note importante à propos des versions du logiciel Power BI Desktop	23

Chapitre 2

Acquérir les données

A. Introduction	27
1. Organisation des fichiers exemples et des sources	27
2. Les principes généraux	27
3. L'accès à la collection de connecteurs	28
B. Les sources de type fichier	30
1. Fichier Excel	30
a. Une feuille dans un fichier	30
b. Plusieurs feuilles dans un fichier	32
c. Plusieurs feuilles de structure identique dans un seul fichier	33
d. Plusieurs fichiers avec des feuilles de structure identique	34
e. Plusieurs fichiers avec des feuilles de noms différents	36
f. Obtenir une liste de répertoires ou une liste de fichiers	37
g. Cas particulier du tableau croisé	39

2. Fichiers plats	40
a. Fichier TXT ou CSV	40
b. Le cas des données empilées dans un fichier TXT ou CSV	42
c. Fichier XML	45
d. Fichier JSON	46
e. Fichier PDF	47
C. Les base de données	48
1. Base Access	48
2. SQL Server	50
a. Accéder aux tables	50
b. Effectuer une recherche sur la base	52
c. Le mode Import et le mode DirectConnect (ou LiveConnect)	53
d. Utiliser les options avancées	53
D. Les sources web	55
1. Extraire les fichiers d'un dossier SharePoint	55
2. Extraire un tableau d'une page web	56
3. Extraire une liste structurée implicite d'une page web	58
4. Autre sources web	62
a. API	62
b. Flux Odata	63
c. Microsoft 365	64
d. Google Analytics	64
E. Les sources de données de la Power Platform	65
1. Les jeux de données	65
2. Les flux de données (Dataflows)	68
3. Le Common Data Service ou Dataverse	68
a. Connexion par le CDS	69
b. Connexion par le connecteur Dataverse	69
F. Changer la source de données d'une requête	70
G. Exporter et partager une source de données	72
H. Épingler une source de données	73

Chapitre 3

Nettoyer et transformer les données

A. Introduction	77
B. Fonctionnement général	78
C. Lire et comprendre le code M	79
1. Aperçu de la structure d'un code M	79
2. Afficher et éditer le code M	81
D. Transformer les colonnes	82
1. Conserver ou supprimer les colonnes	82
a. Conserver ou supprimer les colonnes à l'aide de l'interface	83
b. Supprimer ou conserver les colonnes ?	85
c. Maîtriser la suppression de colonnes à l'aide du code M	86
2. Ajouter les en-têtes	87
3. Renommer les colonnes	88
a. Avec l'interface	88
b. Avec M	88
4. Changer le type de la colonne	90
a. Changer le type d'une colonne	91
b. Changer le type de plusieurs colonnes d'un coup	92
5. Conserver ou supprimer les lignes	93
6. Développer les colonnes	96
E. Transformer le texte	100
1. Remplacer des portions de texte	103
2. Fractionner une colonne, extraire, recombinaison	104
3. Ajouter un préfixe ou un suffixe	109
4. Nettoyer le texte à l'aide du code M	109
F. Transformer les numériques	113
1. Remplacer le séparateur de décimales	116
2. La fonction Modulo et List.Split	116
G. Transformer les dates	120
1. Utiliser l'interface pour générer des champs liés à la date	121
2. Utiliser l'interface pour créer une table du temps	122
3. Générer une table du temps avec le code M	122
4. Quelques éléments de M	124
a. Obtenir la date d'aujourd'hui	124
b. Convertir du texte en date	125

5.	Travailler avec les durées.	125
a.	Utiliser les fonctionnalités de calcul de durée de Power Query.	125
b.	Générer un identifiant pour travailler avec une table des heures.	127
H.	Transformer les requêtes.	129
1.	Ajouter des colonnes.	129
a.	Ajouter des colonnes à l'aide d'une formule (Colonne personnalisée). ...	129
b.	Ajouter des colonnes à partir d'exemples.	132
c.	Ajouter des colonnes conditionnelles.	135
d.	Ajouter des colonnes à l'aide d'une fonction personnalisée.	137
e.	Ajouter une colonne d'index.	137
2.	Dépivoter un tableau croisé.	138
a.	En utilisant l'interface.	138
b.	En utilisant le code M.	141
c.	Cas particulier d'un tableau avec des sous-totaux.	143
3.	Combiner des requêtes.	146
a.	Empiler les lignes (ajouter).	146
b.	Remarques sur la combinaison de fichiers avec le connecteur Dossier. ...	149
c.	Assembler les colonnes (fusionner).	149
4.	Importer plusieurs fichiers avec des feuilles de noms différents.	153

Chapitre 4

Exploiter la puissance de Power Query

A.	Introduction.	159
B.	Organiser les requêtes et les étapes.	160
1.	Gérer les requêtes.	160
a.	Créer des répertoires pour regrouper les requêtes.	160
b.	Dupliquer ou référencer une requête.	162
2.	Travailler avec les étapes.	163
a.	Renommer une étape.	164
b.	Gérer et modifier une étape.	166
C.	Récupérer le contexte de la requête.	167
D.	Dynamiser la requête et travailler avec les paramètres.	171
1.	Travailler sur des périodes glissantes.	171
2.	Travailler avec les paramètres de requête.	174
a.	Créer et utiliser un paramètre pour filtrer la requête.	174
b.	Créer un paramètre de type texte avec une liste de valeurs dynamiques. .	181
c.	Créer et utiliser un paramètre pour changer la source de la requête.	182
d.	Paramètres et modèle de document.	187

E. Entrer des données	188
1. Entrer des données par copier-coller.....	188
a. Où trouver la fonctionnalité ?	189
b. Copier-coller à partir d'Excel.....	189
c. Copier-coller à partir d'un fichier texte.....	190
d. Copier-coller à partir d'un tableau PDF.....	191
e. Copier-coller à partir d'un tableau Word	192
2. Ajouter une table par saisie manuelle	193
3. Ajouter une table avec du code M	194
a. Ajouter une table du temps.....	194
b. Ajouter une table et son contenu manuellement	194
F. Analyser le texte avec un produit cartésien	196
1. Avec un nombre de termes de recherche limité.....	197
2. Avec un produit cartésien	199
G. Agréger les tables	203
1. Comprendre l'agrégation.....	203
a. Définition standard de l'agrégation.....	203
b. Définition étendue de l'agrégation	204
c. Pourquoi et quand recourir à un agrégat dans Power BI	204
d. Comment s'utilise un agrégat.....	205
2. Mettre en place les données agrégées	206
a. Agréger un indicateur dans une table dimension	207
b. Créer une table agrégée	212
c. Ramener un modèle en flocon à un modèle en étoile.....	216
d. Utiliser la fonction d'agrégation Toutes les lignes.....	217

Chapitre 5

Gérer les erreurs au chargement

A. Introduction	223
B. Comment repérer les erreurs	223
1. Dans Power Query.....	223
2. Au moment du chargement complet des données	226
C. Supprimer les erreurs sur une colonne	227
D. Construire un rapport d'erreur lors du chargement	229
E. Erreurs liées aux paramètres régionaux	238
F. Utiliser la fonction TRY localement.....	240

G. Les messages d'erreurs courants	243
1. Expression.Error : La clé ne correspondait à aucune ligne dans la table	243
2. Expression.Error : Désolé... Nous n'avons pas trouvé la colonne « nom » de la table	243
3. DataSource.error : Le fichier <chemin/nom du fichier> est introuvable.	244
4. Expression.Error : Le nom « Table.promoteHeaders » n'a pas été reconnu ...	245
5. DataFormat.Error : Désolé... Nous ne pouvons pas procéder à la conversion en un nombre.....	245

Chapitre 6

Guide pratique du code M

A. Introduction	249
B. Présentation générale du code M.....	249
1. DAX ou M ?	250
2. Afficher la librairie des fonctions.....	251
3. Les étapes	255
4. Les types de valeurs primitives	256
5. Les types complexes (ou abstraits)	257
C. Réutiliser le code M entre plusieurs requêtes ou documents	259
D. Créer et utiliser une fonction personnalisée	261
1. Créer la fonction personnalisée	261
2. Utiliser la fonction personnalisée	263
3. Créer une fonction complexe	264
E. Travailler avec les tables	265
1. Manipulations courantes avec les tables	265
2. Fonctions de table courantes	265
3. Exemples de manipulations de tables	266
a. Extraire des colonnes dynamiquement.....	266
b. Ajouter un suffixe aux noms des colonnes.....	266
c. Nettoyer le nom des colonnes	266
F. Travailler avec les listes	267
1. Manipulations courantes avec les listes	267
2. Fonctions de liste courantes.....	267
G. Travailler avec les enregistrements (records).....	268
1. Manipulations courantes avec les enregistrements	268
2. Fonctions d'enregistrement courantes	269
H. Les fonctions Texte, Date et Nombre	269
I. L'expression each	270

J. Comprendre l'actualisation incrémentielle.....	271
1. Comment la mettre en place ?	272
2. Limitations et réserves	276
K. Extraire les données sur plusieurs pages d'un site avec le code M	276
1. Créer la requête sur la première page	277
2. Transformer cette requête en fonction personnalisée	278
3. Créer une table à une colonne contenant les numéros de page	278
4. Créer une colonne personnalisée appelant la fonction.....	280
5. Opérer des transformations.....	282
L. Où trouver des ressources sur M ?	282

Chapitre 7

R et Python dans Power Query : premier pas

A. Utiliser R dans Power Query.....	285
1. Installation (R et R Studio)	285
2. R pour importer les données	287
3. R pour transformer les données.....	291
B. Utiliser Python dans Power Query	294
1. Installation	294
2. Python pour importer les données	297
3. Python pour transformer les données	299

Index	303
-------------	-----

A. Introduction

Avec ce nouveau chapitre, nous entrons de plain-pied dans Power Query et ses capacités à nettoyer et à transformer les données issues de la source.

Nous allons voir le fonctionnement général de Power Query, et l'articulation avec les lignes de code M.

Nous aborderons le travail au niveau des requêtes (ou tables, les termes sont synonymes), au niveau des colonnes (modifier, supprimer, ou ajouter des colonnes), la transformation des champs de type texte, numérique ou date et le traitement des tableaux croisés.

La plupart du temps, nous verrons que les opérations peuvent être réalisées à l'aide de l'interface graphique, mais aussi à l'aide du code M, et pourquoi comprendre ce code et en connaître certaines syntaxes simples permet d'accélérer le travail, et souvent de dépasser les limites de l'interface.

Les opérations que nous illustrons dans ce chapitre sont courantes, des opérations que tout utilisateur de Power Query doit connaître.

Le chapitre Exploiter la puissance de Power Query vous permettra de pousser plus loin le travail à l'aide de l'interface, avec entre autres la mise en place de tables d'agrégation, ou encore de paramètres.

Le chapitre Guide pratique du code M propose, lui, une approche plus globale de ce code, avec des éclairages sur sa structure, ses entités et ses concepts, et illustrera ces différents points à l'aide d'un ensemble de cas de mises en œuvre fréquents.

B. Fonctionnement général

L'accès à Power Query est possible à tout moment du travail dans Power BI : soit à partir du navigateur que nous avons vu au chapitre précédent (options **Transformer les données**), en d'autres termes immédiatement après la phase de connexion aux données, soit à partir de Power BI (bouton **Transformer les données** dans le ruban **Accueil**), en cours d'élaboration du rapport.

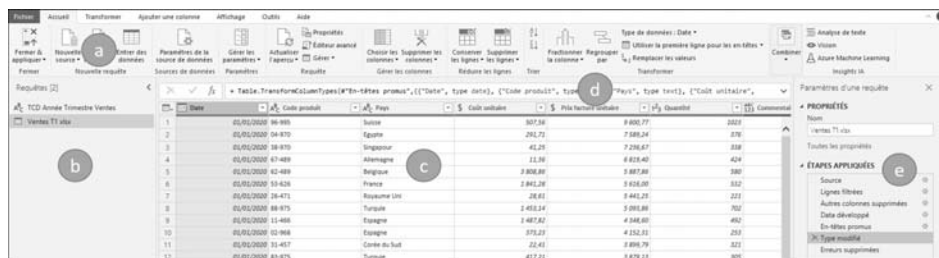
Il faut ici distinguer les sources locales (ou personnelles) et les sources d'entreprise (base de données notamment, mais aussi toute source partagée) : les premières nécessitent le plus souvent un travail de nettoyage et de préparation, ce qui est moins vrai des secondes. Mais il n'existe aucune règle définitive en la matière.

Il est utile de rappeler que le travail dans Power Query est effectué sur un échantillon de données (1000 lignes) : sur la base de cet échantillon, vous allez pouvoir définir un ensemble d'opérations (dont le nom précis est « étapes »). C'est seulement lorsque vous serez satisfait par cet ensemble d'étapes que vous allez envoyer la requête à la source, qui ramènera alors l'intégralité des lignes de la source.

Autrement dit, tout travail dans Power Query se conclut par l'activation de la fonctionnalité **Fermer & appliquer** : c'est à ce moment que vous envoyez la requête.

Il faut souligner un autre aspect de Power Query : toute étape de transformation correspond à une ligne de code M, visible dans la barre de formule. Nous allons y revenir dans la section suivante.

Pour rappel, un aperçu de l'interface graphique de Power Query :



a : l'ensemble de rubans, et dans le ruban **Accueil** la fonctionnalité **Fermer & appliquer** ;

b : les requêtes, ou tables (mais aussi, nous le verrons plus loin, les fonctions personnalisées, les paramètres, etc.).

c : la table active.

d : la barre de formule ; si vous ne voyez pas la barre de formule, rendez-vous dans le ruban **Affichage** et cochez la case **Barre de formule**.

e : les étapes.

C. Lire et comprendre le code M

1. Aperçu de la structure d'un code M

Dans ce chapitre, nous allons souvent alterner entre l'interface graphique et le code M : dans la présente section, je vous propose une petite introduction à ce code, sa structure, et comment l'afficher, mais uniquement dans le but de vous proposer par la suite des petites modifications souvent très simples, en général à partir de la barre de formule – comme vous pourriez le faire dans Excel.

Voici à quoi ressemble un code M résultant de transformations courantes. Les étapes apparaissent en gras dans le code ci-dessous (chacune des étapes correspond à une transformation). Les retours à la ligne ont été ajoutés dans un souci de clarté. Ce code a été généré automatiquement par l'interface graphique.

En deux mots : il s'agit d'abord d'aller chercher les données (étapes **Source** et **Feuill Sheet**), puis d'opérer une série de transformations (de l'étape **En-têtes promus** à **Lignes filtrées**), et enfin de retourner le résultat (après le **IN**) :

```
let
    Source =
    Excel.Workbook(File.Contents(ChoixSource), null, true),
    Feuill_Sheet =
    Source[[Item="chine", Kind="Sheet"]][Data],
    "En-têtes promus" =
    Table.PromoteHeaders(
        Feuill_Sheet,
        [PromoteAllScalars=true]),
    "Type modifié" =
    Table.TransformColumnTypes(
        "En-têtes promus",
        {{"date", type date},
        {"heure", type time},
        {"confirmés", Int64.Type}},
    "Index ajouté" =
    Table.AddIndexColumn("Type modifié", "Index", 1, 1),
    "Personnalisée ajoutée" =
    Table.AddColumn(
        "Index ajouté",
        "Date heure",
        each [date] & [heure]),
    "Lignes filtrées" =
    Table.SelectRows(
        "Personnalisée ajoutée",
        each ([date] <> null))
in
    "Lignes filtrées"
```

Extraire et préparer les données en vue de leur exploitation dans Excel ou Power BI

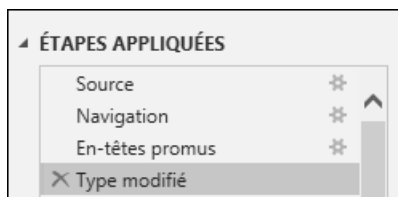
De manière plus générale (et peut-être plus lisible), la structure du code est le plus souvent la suivante :

```
let
    Source = <nom de la source> ,
    // cette ligne est un commentaire
    #"Etape 1" = FonctionM (Source, paramètres) ,
    #"Etape 2" = FonctionM ("Etape 1", paramètres) ,
    #"Etape finale" = FonctionM ("Etape 2", paramètres)
in
    #"Etape finale"
```

Le mot-clé LET (en minuscules) permet d'énoncer une série d'opérations tandis que le mot-clé IN (en minuscules) indique ce qui ressortira. C'est le plus souvent la dernière étape, mais pas nécessairement : il est possible d'invoquer une étape antérieure, par exemple pour corriger le code.

Comme vous pouvez le constater, l'élément clé est l'**étape**.

Elle porte un nom et appelle une fonction à laquelle elle passe des paramètres. Dans l'interface de Power Query, vous retrouvez le nom des étapes dans la rubrique **ÉTAPES APPLIQUÉES** du volet **Paramètres d'une requête** :



Dans cet exemple, quatre étapes ont été appliquées

Vous retrouvez le code dans la barre de formule :

Si le nom de l'étape est un seul mot, il apparaît tel quel : *Source*, *Feuil1_Sheet*. S'il est composé de plusieurs termes, il apparaît encadré de guillemets et précédé d'un dièse : *"Etape 2"*.



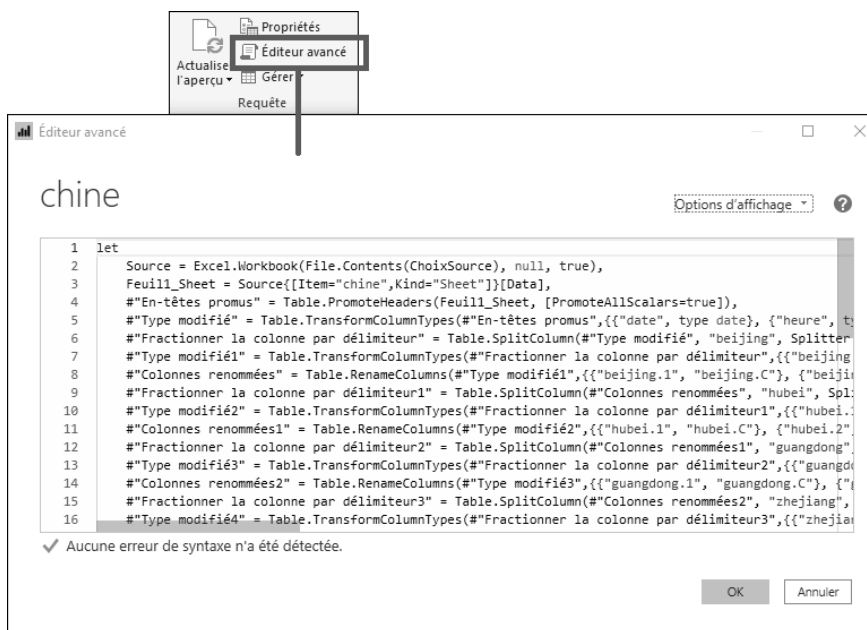
Attention, M est sensible à la casse : il faut impérativement respecter les minuscules et les majuscules (le plus souvent au début du nom de la fonction).

Autres remarques sur le code visible ci-dessus :

- ▶ Parmi les paramètres d'une fonction, vous retrouvez quasi systématiquement une référence à une autre étape, le plus souvent celle qui précède.
- ▶ Pour insérer un commentaire, faite précéder la ligne d'un double slash //.
- ▶ Pour insérer un commentaire sur plusieurs lignes, vous pouvez aussi utiliser la syntaxe /* ... */.

2. Afficher et éditer le code M

Bien que vous puissiez éditer le code directement dans la barre de formule, la façon la plus efficace d'afficher et de modifier le code M est d'utiliser l'éditeur avancé (accessible depuis l'onglet **Accueil**), fenêtre dans laquelle l'intégralité du code (c'est-à-dire toutes les étapes) est visible :



Plutôt que d'un éditeur avancé, il faudrait plutôt parler d'un éditeur rudimentaire pour utilisateur avancé ! En effet, vous n'y trouverez que peu de coloration syntaxique (mais tout de même bien utile), et une aide à l'écriture (M IntelliSense).

Moyennant quoi les codeurs **experts** utilisent souvent Notepad++ avec l'extension M de Lars Schreiber, ou encore Visual Studio (version 2017 et suivantes) avec le SDK Power Query.

Extraire et préparer les données en vue de leur exploitation dans Excel ou Power BI

Ce serait trop pour notre approche du M, aussi je me contenterai dans la suite de cet ouvrage de la barre de formule ou de l'éditeur avancé.



En termes de coloration syntaxique, le rouge indique une zone de texte (avec ses guillemets), le bleu turquoise un type de donnée, le bleu foncé un mot-clé. Tout le reste est noir.

D. Transformer les colonnes



*Dans chacune des sections de ce chapitre, nous allons reprendre les fichiers que nous avons commencé à créer au chapitre précédent. Vous pouvez repartir de vos fichiers ou les retrouver dans le répertoire **exemples\chapitre3** disponible en téléchargement.*

Transformer et nettoyer les colonnes sont les premières actions que vous réaliserez dans Power Query : cela inclut notamment supprimer les colonnes ou les lignes inutiles et vérifier ou changer le type de la colonne.

1. Conserver ou supprimer les colonnes



*Dans cette section, nous utilisons le fichier **source02.pbix**.*

Il existe une règle générale dans Power BI : ne ramenez que les données dont vous avez besoin. La quantité de données a un impact sur la performance de l'application, en lien avec les capacités de votre PC. Plus le fichier est léger, mieux Power BI réagira.

Conserver uniquement les colonnes nécessaires est donc la première action que vous devez faire. Cela suppose bien sûr que vous ayez une bonne idée de l'objectif que vous fixez à votre rapport. Mais rien n'est non plus définitif : vous pouvez à tout moment rajouter des colonnes ou annuler la suppression d'une colonne.

a. Conserver ou supprimer les colonnes à l'aide de l'interface

- 🔗 Ouvrez le fichier **source02.pbix**.
 - 🔗 Dans le ruban **Accueil** - groupe **Requêtes** cliquez sur **Transformer les données** pour lancer Power Query.
 - 🔗 Dans le volet **Requêtes**, cliquez sur **Ventes** pour faire apparaître la table.
- Sur la droite de la table, des colonnes superflues (**Column10** à **Column14**) ont été extraites : nous allons les supprimer.
- 🔗 En maintenant la touche **[Ctrl]** enfoncée, cliquez un à un sur les en-têtes des cinq colonnes concernées.
 - 🔗 Puis effectuez un clic droit pour ouvrir le menu contextuel et cliquez sur **Supprimer les colonnes**.

ABC 123	Column10	ABC 123	Column11	ABC 123	Column12	ABC 123	Column13	ABC 123	Column14	PROPRIÉTÉ
	null		null		null		null		null	
	null		null		null		null		null	
	null		null		null		null		null	
	null		null		null		null		null	
	null		null		null		null		null	

Les colonnes disparaissent, et une nouvelle étape (**Colonnes supprimées**) a été créée dans le volet **Paramètres d'une requête**. Le nom de l'étape est précédé d'une croix, qui permet la suppression et le retour à l'état avant transformation.

Il existe différentes manières de supprimer une ou plusieurs colonnes :

- ▶ par clic droit, ainsi que nous venons de le voir,
- ▶ à l'aide de la fonctionnalité **Choisir les colonnes**,
- ▶ en sélectionnant les colonnes à conserver, plutôt que celles à supprimer : lorsque vous supprimez plus de colonnes que vous n'en conservez, cette option peut être plus rapide. Cette opération peut être effectuée par clic-droit (option **Supprimer les autres colonnes** du menu contextuel), ou à l'aide de la fonctionnalité **Supprimer les autres colonnes** du menu déroulant **Supprimer les colonnes**.

Voyons par exemple la deuxième méthode :

- 🔗 Dans le volet **Paramètres d'une requête**, cliquez sur la croix à gauche de l'étape **Colonnes supprimées**. Ceci annule l'étape, et les colonnes réapparaissent donc.

