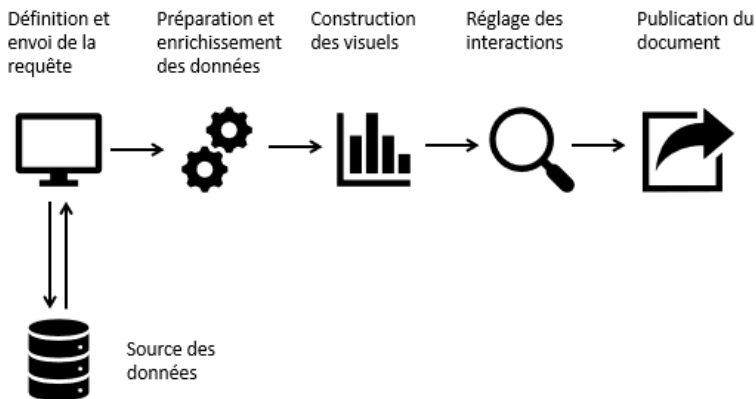


A. Les concepts clés

1. Circulation des données

Comprendre la façon dont circulent et sont stockées les données lorsque vous utilisez Power BI Desktop peut s'avérer utile, notamment lors du choix de la connexion, lors de l'actualisation, de la création des colonnes ou des mesures, et de l'utilisation même du rapport.

Reprenons le schéma du chapitre précédent :

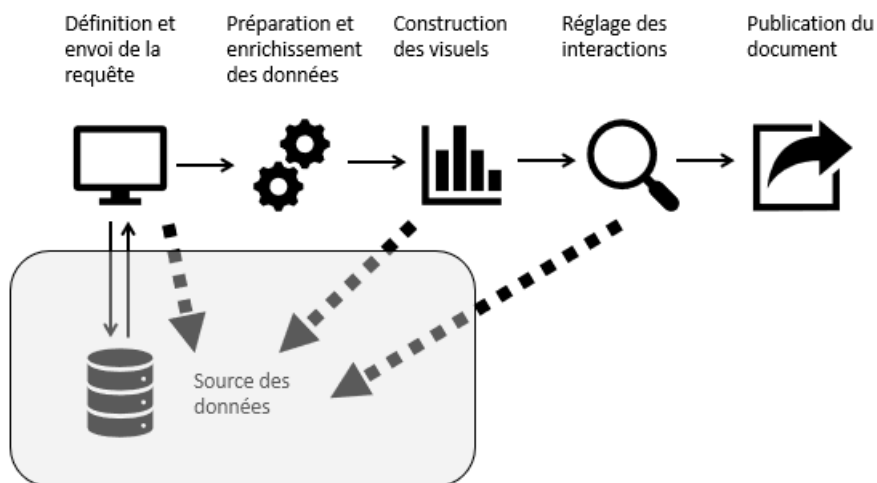


Le premier point clé concerne le type de connexion : la volumétrie des données, ou le besoin de temps (quasi) réel, peuvent vous amener à utiliser une connexion en direct sur votre base de données (mode Direct) – dans ce cas en effet, les données restent dans la base, qui est interrogée à chaque manipulation dans le fichier au prix d'un effort important sur la base.

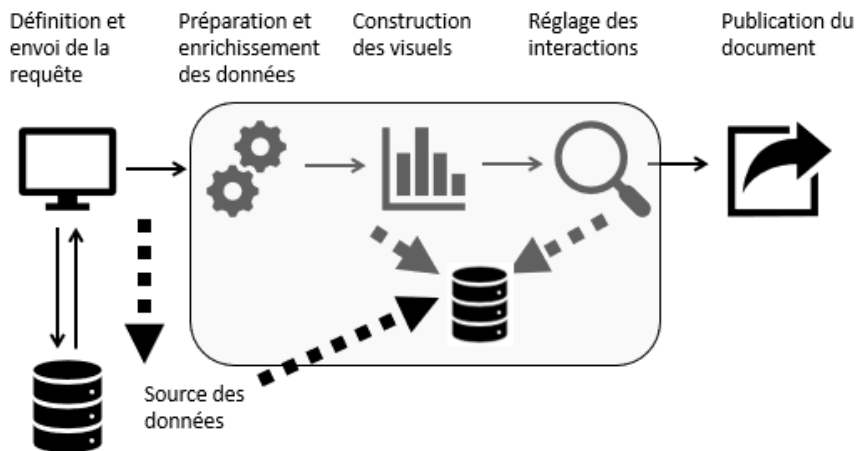


Les types de connexion sont précisés dans la section suivante.

Dans ce premier cas, les données circulent « à la demande » et ne sont pas stockées localement. Une opération aussi simple qu'afficher un total au bas d'un tableau ou filtrer un graphique amène à réinterroger la source. Les performances dépendent des performances du système de base de données :



Dans les autres cas, l'application sera plus performante si les données sont rapatriées localement (mode Import) : la source de données n'est interrogée qu'une fois, lors de l'exécution ou de l'actualisation de la requête, puis les données sont stockées localement, à l'intérieur du fichier.



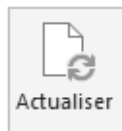
Dans ce cadre, Power BI repose sur un outil très performant : le moteur de stockage tabulaire, VertiPaq, accélère considérablement la mise en place des visuels par la sélection des champs, les interrogations du modèle à chaque utilisation et à chaque calcul d'une mesure, la mise à jour des visuels lors de l'utilisation des filtres et des interactions en général.

Il est également possible de mettre en place des modèles composites, dont une partie des tables (ou des sources), sont exploitées en mode Import, et les autres en mode Direct. Par exemple, des tables Produits ou Sites géographiques, qui évoluent peu, peuvent être importées, mais les tables des transactions, ou des chaînes de production, qui ont parfois besoin d'être interrogées en quasi-temps réel, peuvent rester en mode Direct.

2. Actualisation

Dans le cadre d'une connexion avec importation des données, l'actualisation dans Power BI Desktop se fait exclusivement à l'aide de l'outil **Actualiser**.

Dans le cadre d'une connexion directe à la source de données, nous l'avons vu, l'actualisation dans Power BI Desktop se fait à chaque action sur un visuel. Elle peut aussi être déclenchée manuellement à l'aide de l'outil **Actualiser** dans l'onglet Accueil :



Enfin, c'est dans le cadre de la publication sur Power BI Service (le serveur Power BI) que les possibilités d'actualisation sont les plus complètes.

Lorsque vous publiez un rapport sur Power BI Service, un jeu de données est également envoyé sur le serveur, et disponible aux côtés des rapports et des tableaux de bord (composés par l'utilisateur à partir des visuels présents dans les rapports).

Ce jeu de données peut faire l'objet d'une actualisation immédiate (manuelle), automatique (notamment si le fichier source est stocké dans SharePoint ou OneDrive) ou programmée (planifiée).



Le plus souvent, il faut avoir mis en place une passerelle entre le serveur Power BI et le serveur de base de données, pour que la mise à jour puisse se faire.

Si l'utilisateur a créé un tableau de bord, en épinglant différents visuels disponibles dans les rapports auxquels il a accès, une fonctionnalité **Actualiser les vignettes du tableau de bord** est disponible, pour forcer l'actualisation (en dehors de cette actualisation manuelle, une actualisation automatique a lieu toutes les 15 ou 30 minutes).



Pour certains types de sources – fichiers stockés sur OneDrive, SharePoint, bases de données en ligne Salesforce, base de données Azure, l'actualisation est le plus souvent automatique (toutes les heures par exemple pour des fichiers stockés sur OneDrive).



Il est également possible d'aller chercher uniquement les données ajoutées ou modifiées depuis la précédente actualisation : l'actualisation incrémentielle repose sur l'existence d'une date témoin, et accélère considérablement le temps d'actualisation d'un rapport. Pour plus d'information sur ce point, je vous invite à lire l'article que je lui ai consacré : <https://daxone.fr/un-choix-parmi-les-nouveautes-de-fevrier-2020/>.

B. Se connecter

Power BI propose une vaste bibliothèque de connecteurs – et encore celle-ci évolue avec chaque nouvelle livraison du logiciel : les décrire tous serait fastidieux, mais il est possible de dégager les grandes lignes et le fonctionnement des principaux connecteurs.



Vous trouverez sur le site de l'éditeur une description très complète de la plupart des connecteurs : <https://docs.microsoft.com/fr-fr/power-bi/desktop-data-sources>

1. Les trois types de connexion

Un premier choix essentiel consiste à définir le type de connexion : import ou direct.

- ▶ L'option la plus fréquente consiste à **importer** les données : une fois la requête lancée, Power BI récupère et stocke en local l'ensemble des données.
 - ▶ L'intérêt majeur de ce type de connexion repose sur les performances de Power BI lors de la création de rapports, les possibilités de modélisation et la réactivité aux actions de filtrage de l'utilisateur.
 - ▶ L'inconvénient tient à ce que le poste sur lequel est conçu ou utilisé le rapport doit avoir une puissance suffisante, en termes de RAM (pour le stockage des données) et de CPU (pour les calculs).
 - ▶ Un autre inconvénient possible est la limite de taille du fichier : 1 Go pour la version standard, 10 Go pour les versions Pro et Premium.
 - ▶ Enfin, ce type de connexion implique une actualisation manuelle ou programmée des données – par opposition à la visualisation de données en temps réel.

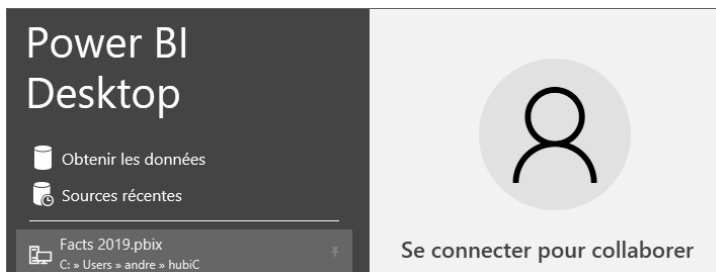
Malgré ces quelques inconvénients, l'import des données est le mode à privilégier.

- ▶ Il est également possible de se connecter en **direct** à certaines sources (bases de données).
 - ▶ L'intérêt ici est d'avoir en permanence des rapports à jour, sans avoir besoin d'actualiser les données. Une simple action (filtre ou sélection dans le visuel) suffit à mettre à jour les données. Par ailleurs, la puissance du poste n'est plus un critère limitant.
 - ▶ En revanche, un inconvénient tient à la capacité de la source : il faut qu'elle soit suffisamment puissante pour accepter une connexion directe. Les performances de l'application peuvent s'en ressentir.
 - ▶ Par ailleurs, toutes les sources ne sont pas disponibles.
 - ▶ Les possibilités de modélisation des données sont limitées (puisqu'elles ont déjà été définies dans la source).
 - ▶ Enfin d'autres limitations incluent la restriction de certaines fonctionnalités de transformation et de création de formule (notamment les fonctions Time Intelligence).

Il existe deux types de connexion en direct : la connexion DirectQuery et la connexion directe. La première est plus générale, mais aussi moins performante, la seconde concerne exclusivement SSAS, Azure ou Power BI Service – elle est plus performante dans la mesure où les données sont déjà stockées sur le modèle qu'utilise Power BI.

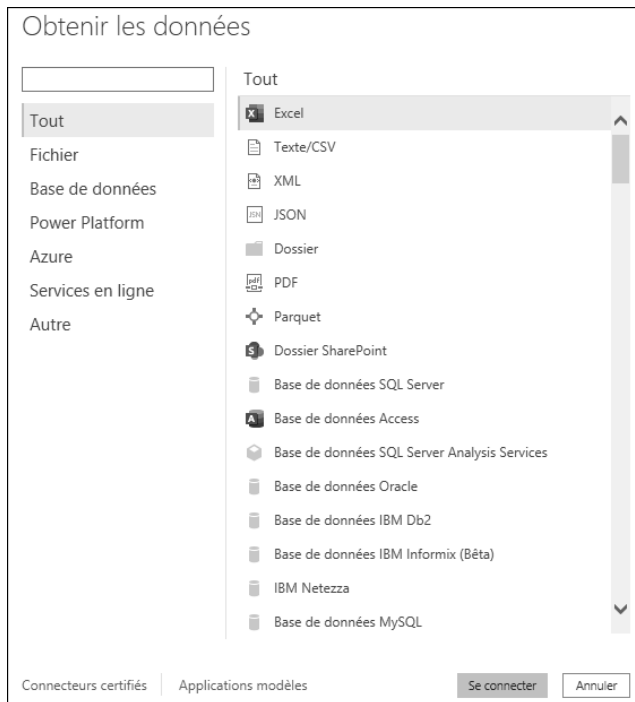
Découvrir les types de connexion

👉 Lancez Power BI Desktop et cliquez sur **Obtenir les données**.



*Vous trouvez aussi la fonctionnalité **Obtenir les données** sur l'onglet **Accueil**.*

👉 Pour accéder à une source, sélectionnez le connecteur dans la liste, puis cliquez sur **Se connecter**.



En fonction du connecteur choisi, des options apparaissent, le nom de la base, le mode de connectivité.

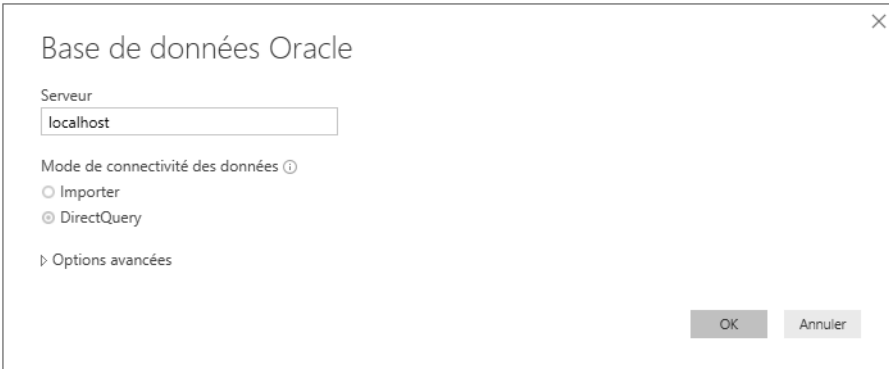
Voici quelques exemples de connecteurs à des bases de données Oracle et SQL Server :



The screenshot shows a dialog box titled "Base de données Oracle". It contains the following elements:

- A "Serveur" field with a dropdown menu showing "A B C" and an empty text input box.
- A "Mode de connectivité des données" section with two radio buttons: "Importer" (selected) and "DirectQuery".
- An "Options avancées" section with a collapsed arrow.
- A "Délai de commande en minutes (facultatif)" field with an empty text input box.
- An "Instruction SQL (facultative)" field with a large empty text area.
- Two checked checkboxes: "Inclure des colonnes de relation" and "Naviguer avec la hiérarchie complète".
- "OK" and "Annuler" buttons at the bottom right.

Connexion avec import sur une base Oracle : dans les options avancées, remarquez la possibilité de saisir directement du code SQL.



The screenshot shows a dialog box titled "Base de données Oracle". It contains the following elements:

- A "Serveur" field with a text input box containing "localhost".
- A "Mode de connectivité des données" section with two radio buttons: "Importer" and "DirectQuery" (selected).
- An "Options avancées" section with an expanded arrow.
- "OK" and "Annuler" buttons at the bottom right.

Connexion DirectQuery sur une base Oracle

A. Introduction

Le rôle central du langage DAX dans l'analyse des données n'est plus à démontrer : aussi bien pensée que soit votre source de données, aussi bien construit que soit votre modèle, il ne s'agit que des pierres sur lesquelles vous allez pouvoir fonder vos analyses, et celles-ci seront toujours basées sur des formules DAX.

Ce chapitre est centré sur LA fonction DAX par excellence : **CALCULATE**.

Parce que c'est la fonction pivot, celle qui demande de maîtriser les deux notions clés du DAX, à savoir les fonctions de table et le contexte de filtre et, par conséquent de maîtriser toute une batterie d'autres fonctions considérées à juste titre comme centrales.

Et parce que derrière une apparente simplicité – après tout, **CALCULATE** se contente de modifier le contexte de filtre avant de calculer une expression -, se cache un mécanisme d'une richesse et parfois d'une complexité remarquables.

Après un préambule destiné à poser les bases du travail (rappel des notions clés, debugging, validation d'un calcul), nous aborderons **CALCULATE**, nous évoquerons des modèles de formule, nous entrerons dans les arcanes de la fonction, et nous verrons comment **CALCULATE** permet de réaliser des analyses courantes.



Depuis la mise à jour de mai 2020, les séparateurs DAX sont par défaut la virgule pour séparer les listes et les arguments au sein d'une fonction, et le point comme séparateur décimal. Un système « à l'américaine », donc, par opposition aux paramètres localisés, le point-virgule pour séparer les arguments et virgule comme séparateur décimal, pour ce qui concerne la France. Nous restons dans ce livre fidèle à ces derniers. Vous pouvez passer de l'un à l'autre à l'aide du menu **Fichier - Options et paramètres - Options**, puis dans la rubrique **Global - Paramètres régionaux**, faites votre choix des séparateurs DAX :

The screenshot shows the 'Options' dialog box in Power BI Desktop. The left sidebar is expanded to 'Paramètres régionaux' (Regional Settings) under the 'GLOBAL' category. The main content area shows the following settings:

- Langue du modèle**: Utiliser la langue d'affichage par défaut de Windows (dropdown menu).
- Langue utilisée lors de la comparaison de chaînes dans les données et pour créer des champs de date internes**: Utiliser la langue de l'application (dropdown menu).
- Étapes de la requête**: Spécifiez la langue à utiliser pour les noms d'étapes générés automatiquement:
 - Utiliser la langue de l'application
 - Toujours en anglais
- Séparateurs DAX**: Spécifiez la culture à utiliser pour les séparateurs de liste et de décimale dans les expressions DAX:
 - (Recommandé) Utiliser les séparateurs DAX standard : virgule (,) comme séparateur de liste et point (.) comme séparateur décimal
 - Utiliser des séparateurs DAX localisés : les séparateurs de liste et de décimale sont définis par les paramètres régionaux Windows

At the bottom of the 'Séparateurs DAX' section, there is a link: 'En savoir plus'.

Si vous préférez utiliser les (nouveaux) paramètres recommandés par Power BI, il vous suffira, dans tous les exemples de formules qui suivent, de remplacer le point-virgule par la virgule.

B. Préambule



Cette section traite d'un ensemble de points fondamentaux et généraux de DAX. Vous n'y trouverez donc pas d'exercices d'application, mais je vous encourage à appliquer ces concepts et règles dans toutes vos formules DAX. Il est essentiel que vous lisiez très attentivement les pages qui suivent. Et peut-être, que vous les relisiez par la suite !

1. Les mesures, leur format, leur nom

Pour l'essentiel, les formules doivent être utilisées dans le cadre de la création de mesures, et plus rarement pour la création de colonne.

Rappelons en effet que les mesures ne prennent pas de place dans votre modèle (n'utilise pas d'espace de stockage), et garantissent donc un document plus léger et plus performant. Rappelons que les mesures ne sont calculées qu'au moment de leur utilisation, selon le contexte dans lequel elles apparaissent, et pas pour chaque ligne de la table dans laquelle, à l'inverse, une colonne est « physiquement » créée.

Rappelons également que toute donnée numérique issue de la source et destinée à être analysée (montant des ventes, quantité, résultat de test, etc.), doit être masquée, et remplacée par une mesure équivalente (`[Montant] = SUM(Ventes[Montant des ventes])`). Nous verrons plus loin le lien avec CALCULATE.

Il est par ailleurs utile, lors de la création de formule, de prendre le réflexe de formater tout de suite la mesure nouvellement créée. Enfin pour ce qui concerne le nom de la mesure, évitez les lettres accentuées dans les quelques premiers caractères (afin qu'Intellisense retrouve rapidement votre mesure).

2. Le contexte de filtre, le contexte de ligne

Il ne s'agit pas pour nous ici de reprendre intégralement ces notions, mais simplement de rappeler que le contexte de ligne est généré automatiquement lors de la création d'une colonne par le biais d'une formule, ainsi que lors de l'utilisation d'une fonction de type itérateur (en particulier les fonctions `X – SUMX`, etc. – et la fonction `FILTER`).

Le contexte de filtre, lui, est défini par les visuels (tables, graphiques, segments) présents sur le rapport (et parfois sur les autres rapports). Mais il peut aussi être manipulé par formule, et c'est précisément le rôle de la fonction `CALCULATE`.

Le contexte de ligne filtre la table sur laquelle il s'applique, et n'en retient qu'une ligne. L'utilisation des fonctions `RELATED` ou `RELATEDTABLE` permet cependant au contexte de ligne de se propager vers d'autres tables.

Le contexte de filtre *filtre le modèle* dans son ensemble, selon le sens de propagation de 1 à N le long des relations. L'utilisation de la fonction `CROSSFILTER`, ou d'une relation bi-directionnelle, permettent toutefois au filtre de se propager dans le sens N à 1.

Enfin, dans certains cas, le contexte de ligne est transformé en contexte de filtre, selon le principe de la transition de contexte. Nous reviendrons sur cette notion capitale.

3. Mise en forme des formules

Afin d'assurer la lisibilité des formules, et notamment lorsqu'elles deviennent complexes, certaines règles de mises en forme peuvent être suivies :

- ▶ Les colonnes référencées le sont toujours en précisant le nom de la table d'abord, alors que les mesures sont simplement référencées entre crochets : cela aide en particulier à mieux « voir » le `CALCULATE` implicite.
- ▶ Ajouter autant de commentaires que nécessaire, précédés d'un `//`.
- ▶ Faciliter la vue des parenthèses ouvrantes et fermantes.
- ▶ Aller à la ligne autant de fois que vous le voulez.

```
[Montant moyen facturé] =  
    AVERAGEX (  
        Date[Date] ;  
        [Montant facturé]  
    )
```

Notez aussi qu'une variante, notamment dans le cadre d'un `CALCULATE`, peut consister à décaler le point-virgule, pour mieux être en mesure de mettre en commentaire une des lignes du code :

```
[Mesure] =
    CALCULATE (
        expression
        ; filtre1
        ; filtre2
        ; filtre3
    )
```

Écrite comme ça, la formule permet de mettre en commentaire facilement (ici, le filtre 3) :

```
[Mesure] =
    CALCULATE (
        expression
        ; filtre1
        ; filtre2
//      ; filtre3
    )
```

4. Raccourcis de l'éditeur de formule DAX

L'éditeur de formule fourmille de raccourcis clavier permettant d'accélérer la saisie ou de faciliter la correction des formules DAX.

Avec un minimum de pratique, vous gagnerez un temps précieux pour remettre en forme, corriger à plusieurs endroits en même temps, intervertir les lignes, etc.

Voici quelques-uns de ces raccourcis.

Les cinq raccourcis-clavier les plus utiles

- ▶ Sélectionner toutes les occurrences du terme (ou des caractères) actuellement sélectionné (exemple : une colonne, une table ou une fonction qu'il faut remplacer partout dans la formule).




`Ctrl` `F2` (ou `Ctrl` `⇧` `L`)

- ▶ Mettre en commentaire (ou à l'inverse, enlever la mise en commentaire)








Sélectionner la ou les lignes et `Ctrl` `/`

- ▶ Aller à la ligne avec indentation (retrait)








`⇧` `↵`

- ▶ Copier une ligne au-dessus/en-dessous
 ou 
- ▶ Valider une proposition de fonction d'Intellisense









Manipuler les lignes

- ▶ Déplacer une ligne vers le haut/le bas
 ou 
- ▶ Insérer la ligne ci-dessous

- ▶ Insérer la ligne ci-dessus

- ▶ Sélectionner la ligne actuelle

- ▶ Naviguer jusqu'à une ligne en indiquant son numéro
 G (taper ensuite  pour masquer la zone de saisie du numéro de ligne)



Saisir du texte

- ▶ Insérer plusieurs curseurs aux endroits choisis en cliquant (utile pour saisir un même texte à plusieurs endroits)

- ▶ Activer plusieurs curseurs au même endroit sur plusieurs lignes
 ou 
- ▶ Sélectionner toutes les occurrences de la sélection actuelle (exemple : une colonne, une fonction qu'il faut remplacer partout).
 L ou 
- ▶ Sélectionner le mot entier (là où se trouve le curseur) (Si une portion de texte a déjà été sélectionnée,  D sélectionne la prochaine occurrence de la même sélection, et ainsi de suite : cela permet de modifier l'ensemble des portions sélectionnées d'un coup)


Commentaire et indentation

- ▶ Mettre en commentaire (ou l'inverse)
Sélectionner la ou les lignes et  /
- ▶ Augmenter l'indentation / réduire l'indentation
 / 
- ▶ Aller à la ligne sans indenter
 
- ▶ Aller à la ligne avec indentation
 

Utiliser Intellisense

- ▶ Valider une proposition

- ▶ Pour rappeler la fenêtre Intellisense
 I

5. Correction de formule avec les variables

Il n'existe pas, dans Power BI, de moteur d'analyse des formules permettant d'en arrêter l'exécution au moment où une erreur est détectée (comme c'est le cas dans Excel notamment). Pour des formules complexes, longues, ceci peut être un problème.

C'est là qu'une utilisation astucieuse des variables peut vous aider.



Les variables sont déclarées, et exécutées, au début du script. Elles sont couramment utilisées pour n'exécuter un calcul qu'une fois, même si celui-ci est appelé plusieurs fois dans la formule. Leur première utilité est donc d'améliorer la performance de la formule.

Leur deuxième raison d'être est d'améliorer la lisibilité de la formule.

Leur troisième intérêt est d'effectuer un calcul sur l'état du contexte de filtre au début de l'exécution de la formule (en effet, le calcul n'étant effectué qu'une fois, la valeur de la variable est figée pendant toute la durée de la formule). Ce point, important, est développé à la fin de cette section.



Pour illustrer ce point, je me sers d'un exemple proposé par Marco Russo dans un tutoriel de Guy In A Cube : <https://www.youtube.com/watch?v=9SV2VnYbgg4>

Nous allons commencer par voir les trois états de mise en forme de la formule : d'abord sans mise en forme, puis avec, et enfin avec l'introduction des variables.

Cette formule retourne une liste de couleurs dont le nombre varie en fonction du choix fait par l'utilisateur à l'aide d'un segment *Couleurs_seg*.

Premier état :

```
Couleurs =  
IF (  
    COUNTROWS ( DISTINCT ( 'Produit'[Couleur] ) ) > SELECTEDVALUE  
( 'Couleurs_seg'[Nombre de Couleurs] );  
    CONCATENATEX ( TOPN ( SELECTEDVALUE ( 'Couleurs_seg'[Nombre de  
Couleurs] ); VALUES ( 'Produit'[Couleur] )); 'Produit'[Couleur]; " , "  
& " et plus...";  
    CONCATENATEX ( DISTINCT ( 'Produit'[Couleur] ); 'Produit'[Couleur];  
" , " )  
)
```

Vous remarquerez tout de suite que le code est très confus, donc difficile à lire, et par conséquent difficile à corriger.

Et en effet, ce code renvoie une erreur. Mais où se situe-t-elle ?

Le premier nettoyage consiste à formater le code selon les règles courantes, par exemple en le passant à la moulinette de www.daxformatter.com.