

# Formation Microsoft Fabric :



## Table des matières

|   |    |
|---|----|
| Implémenter un entrepôt de données avec Microsoft Fabric .....                        | 8  |
| Présentation .....  | 8  |
| Comprendre les principes de base d'un entrepôt de données .....                       | 8  |
| Comprendre les entrepôts de données dans Fabric.....                                  | 10 |
| Interroger et transformer des données .....   | 11 |
| Préparer des données pour l'analyse et la création de rapports .....                  | 13 |
| Sécuriser et monitorer votre entrepôt de données .....                                | 15 |
| Exercice - Analyser des données dans un entrepôt de données.....                      | 16 |
| Résumé.....   | 16 |
| Présentation de l'analytique de bout en bout à l'aide de Microsoft Fabric .....       | 17 |
| Introduction .....  | 18 |
| Explorer l'analytique de bout en bout avec Microsoft Fabric .....                     | 18 |
| Équipes de données et Microsoft Fabric .....  | 19 |
| Activer et utiliser Microsoft Fabric.....   | 20 |
| Résumé.....   | 21 |
| Bien démarrer avec les lakehouses dans Microsoft Fabric.....                          | 23 |
| Introduction .....  | 23 |
| Explorer le lakehouse Microsoft Fabric .....  | 23 |
| Travailler avec des lacs de données Microsoft Fabric.....                             | 23 |
| Explorer et transformer des données dans un lakehouse .....                           | 24 |
| Exercice – Créer et ingérer des données avec un lac de données Microsoft Fabric ..... | 25 |
| Exercice .....  | 26 |
| Résumé.....   | 26 |
| Utiliser Apache Spark dans Microsoft Fabric .....                                     | 27 |
| Introduction .....  | 27 |
| Se préparer à l'utilisation d'Apache Spark.....                                       | 27 |
| Exécuter du code Spark .....  | 28 |
| Utiliser des données dans un dataframe Spark.....                                     | 29 |
| Utiliser des données à l'aide de Spark SQL .....                                      | 30 |
| Visualiser des données dans un notebook Spark.....                                    | 31 |
| Exercice .....  | 34 |
| Utiliser des tables Delta Lake dans Microsoft Fabric .....                            | 34 |
| Introduction .....  | 34 |

|  |    |
|--|----|
| Comprendre Delta Lake .....  | 35 |
| Créer des tables delta .....   | 36 |
| Utiliser des tables delta dans Spark .....                                     | 38 |
| Utiliser des tables delta avec des données de streaming .....                  | 38 |
| Exercice .....   | 39 |
| Résumé .....   | 40 |
| Utiliser des pipelines Data Factory dans Microsoft Fabric .....                | 40 |
| Introduction .....   | 40 |
| Comprendre les pipelines .....   | 41 |
| Utiliser l'activité copier des données .....                                   | 42 |
| Utiliser des modèles de pipeline .....   | 44 |
| Exécuter et superviser des pipelines .....                                     | 45 |
| Exercice .....   | 47 |
| Ingérer des données avec des flux de données Gen2 dans Microsoft Fabric .....  | 48 |
| Présentation .....   | 48 |
| Comprendre les flux de données Gen2 dans Microsoft Fabric .....                | 48 |
| Explorer un flux de données Gen2 dans Microsoft Fabric .....                   | 49 |
| Intégrer des flux de données Gen2 et des pipelines dans Microsoft Fabric ..... | 51 |
| Exercice .....   | 52 |
| Résumé .....   | 53 |
| Bien démarrer avec Real-Time Analytics dans Microsoft Fabric .....             | 53 |
| Introduction .....   | 53 |
| Qu'est-ce que Synapse Real-Time Analytics ? .....                              | 54 |
| Comprendre la base de données et les tables KQL .....                          | 54 |
| Ecrire des requêtes en KQL .....   | 56 |
| Exercice : Explorer Synapse Real-Time Analytics dans Fabric .....              | 57 |
| Résumé .....   | 57 |
| Utiliser des flux d'événements en temps réel dans Microsoft Fabric .....       | 58 |
| Introduction .....   | 58 |
| Principaux composants d'Eventstream .....                                      | 59 |
| Configurer Eventstream .....   | 61 |
| Routage des données de flux d'événements .....                                 | 62 |
| Exercice .....   | 65 |
| Résumé .....   | 65 |
| .....  | 65 |

|  |    |
|--|----|
| Interroger des données dans une base de données KQL dans Microsoft Fabric .....                    | 66 |
| Présentation .....   | 66 |
| Prise en main des requêtes KQL.....  | 66 |
| Bonnes pratiques relatives à KQL .....   | 68 |
| Fonctionnalités avancées .....   | 69 |
| Exercice : Interroger une base de données KQL dans Microsoft Fabric .....                          | 71 |
| Résumé .....   | 71 |
| Bien démarrer avec la science des données dans Microsoft Fabric.....                               | 72 |
| Présentation .....   | 72 |
| Comprendre le processus de science des données.....  | 72 |
| Explorer et traiter des données avec Microsoft Fabric .....  | 74 |
| Entraîner et évaluer des modèles avec Microsoft Fabric.....  | 74 |
| Exercice .....   | 75 |
| Résumé .....   | 75 |
| Bien démarrer avec Data Activator dans Microsoft Fabric.....                                       | 76 |
| Présentation .....   | 76 |
| Comprendre Data Activator .....  | 77 |
| Démarrage avec Data Activator .....  | 78 |
| Comprendre les déclencheurs, les conditions et les actions dans Data Activator .                   | 78 |
| Obtenir des données à partir de rapports et de flux d'événements Power BI avec Data Activator..... | 79 |
| Affecter des données dans Data Activator .....   | 83 |
| Créer des déclencheurs dans DATA Activator .....   | 85 |
| Exercice .....   | 89 |
| Résumé.....  | 89 |
| Administrer Microsoft Fabric .....   | 90 |
| Introduction .....   | 90 |
| Comprendre l'architecture Fabric .....   | 90 |
| Comprendre le rôle Administrateur Fabric.....  | 91 |
| Gérer la sécurité Fabric.....  | 92 |
| Gouverner des données dans Fabric .....  | 93 |
| Résumé.....  | 94 |
| Ingérer des données avec des notebooks Spark et Microsoft Fabric .....                             | 94 |
| Présentation .....   | 94 |
| Se connecter aux données avec Spark.....   | 95 |

|   |     |
|---|-----|
| Envisager des usages adaptés aux données ingérées .....   | 96  |
| Exercice .....  | 96  |
| Implémenter un lakehouse avec Microsoft Fabric .....  | 97  |
| Introduction .....  | 97  |
| Décrire l'architecture en médaillon .....   | 97  |
| Implémenter une architecture en médaillon dans Fabric.....  | 98  |
| Interroger les données, puis générer des rapports les concernant dans votre lakehouse Fabric..... | 99  |
| Facteurs à prendre en compte dans la gestion de votre lakehouse .....                             | 100 |
| Exercice .....  | 101 |
| Résumé.....   | 102 |
| Obtenir des données dans Power BI.....  | 102 |
| Introduction .....  | 102 |
| Obtenir des données de fichiers.....  | 103 |
| Obtenir des données de sources de données relationnelles.....                                     | 103 |
| Créer des rapports dynamiques avec des paramètres.....  | 105 |
| Obtenir des données d'une base de données NoSQL .....   | 106 |
| Obtenir des données de services en ligne .....  | 107 |
| Sélectionner un mode de stockage .....  | 108 |
| Obtenir des données d'Azure Analysis Services.....  | 110 |
| Résoudre les problèmes de performances.....   | 111 |
| Résoudre les erreurs d'importation de données.....  | 113 |
| Exercice .....  | 114 |
| Résumé.....   | 114 |

1. **Data Lake** : Un Data Lake est un référentiel de données qui stocke une grande quantité de données brutes et non structurées dans leur format natif, offrant ainsi une flexibilité maximale pour l'analyse ultérieure.
2. **Data Warehouse** : Un Data Warehouse est une base de données centralisée et organisée qui stocke des données structurées et optimisées pour l'analyse et la génération de rapports, facilitant ainsi la prise de décision basée sur les données.
3. **Table de fait** : Une Table de fait est une table dans un entrepôt de données qui contient des données de transaction ou d'événement, généralement à un niveau granulaire, fournissant ainsi les détails nécessaires pour l'analyse et le reporting.
4. **Table de dimension** : Une Table de dimension est une table dans un entrepôt de données qui contient des informations descriptives sur les entités ou les événements enregistrés dans la table de fait, permettant ainsi une analyse plus approfondie et contextuelle.
5. **Contrôle d'accès en fonction du rôle (RBAC)** : Mécanisme permettant de contrôler l'accès aux ressources en fonction des rôles des utilisateurs. Chaque utilisateur se voit attribuer un ou plusieurs rôles, déterminant ainsi les actions qu'il peut effectuer dans l'entrepôt de données.
6. **Chiffrement SSL** : Protocole de sécurité qui utilise un algorithme de chiffrement pour sécuriser les communications entre un client et un serveur, garantissant ainsi la confidentialité et l'intégrité des données échangées.
7. **Azure Storage Service Encryption** : Service de chiffrement fourni par Microsoft Azure pour protéger les données stockées dans Azure Storage, assurant ainsi leur sécurité tant au repos que pendant leur transmission.
8. **Azure Monitor et Azure Log Analytics** : Outils de surveillance fournis par Microsoft Azure pour collecter, analyser et visualiser les données de télémétrie des ressources Azure, permettant ainsi de surveiller les performances, la disponibilité et la sécurité de l'entrepôt de données.
9. **Authentification multifacteur (MFA)** : Méthode d'authentification qui nécessite la présentation de plusieurs formes d'identification distinctes pour vérifier l'identité d'un utilisateur, renforçant ainsi la sécurité des comptes utilisateur.
10. **Intégration de Microsoft Entra ID** : Service d'identité et d'accès fourni par Microsoft pour gérer les identités utilisateur et l'accès aux ressources Microsoft, offrant ainsi une authentification sécurisée et une gestion centralisée des utilisateurs et des autorisations.
11. **Autorisations d'espace de travail** : Permissions définies au niveau de l'espace de travail dans Fabric, permettant de contrôler l'accès et les actions des utilisateurs sur les données et les services dans l'ensemble de l'espace de travail.

12. **Autorisations d'élément** : Permissions définies au niveau de l'élément dans Fabric, permettant de contrôler l'accès des utilisateurs à des entrepôts de données individuels, offrant ainsi une granularité supplémentaire dans la gestion des autorisations.
13. **Vues de gestion dynamique (DMV)** : Vues système dans une base de données qui fournissent des informations sur l'état et les performances du système, telles que les connexions, les sessions et les demandes, permettant ainsi de surveiller et de diagnostiquer les problèmes de performance en temps réel.
14. **Monitoring des requêtes** : Processus de surveillance et d'analyse des requêtes SQL exécutées dans l'entrepôt de données, permettant d'identifier les requêtes longues ou inefficaces et de prendre des mesures correctives pour optimiser les performances et l'utilisation des ressources.

# Implémenter un entrepôt de données avec Microsoft Fabric

## Présentation

L'entrepôt de données relationnels est importante pour les solutions décisionnelles en entreprise, basé sur un schéma multidimensionnel et dénormalisé. Microsoft Fabric offre une version moderne de cet entrepôt en centralisant les données de divers services et systèmes dans une vue unifiée pour l'analyse et les rapports. Unique car basé sur le lakehouse en format Delta, il offre une sémantique SQL complète pour manipuler les données. Fabric est conçu pour toute l'équipe de données, pas seulement les ingénieurs.

Fabric permet aux ingénieurs, analystes et scientifiques des données de collaborer pour créer et interroger un entrepôt de données adapté à leurs besoins spécifiques. Ce module vous initie aux entrepôts de données dans Fabric, couvrant leur création, chargement, interrogation et visualisation des données.

## Comprendre les principes de base d'un entrepôt de données

Le processus de création d'un entrepôt de données moderne comprend généralement plusieurs étapes :

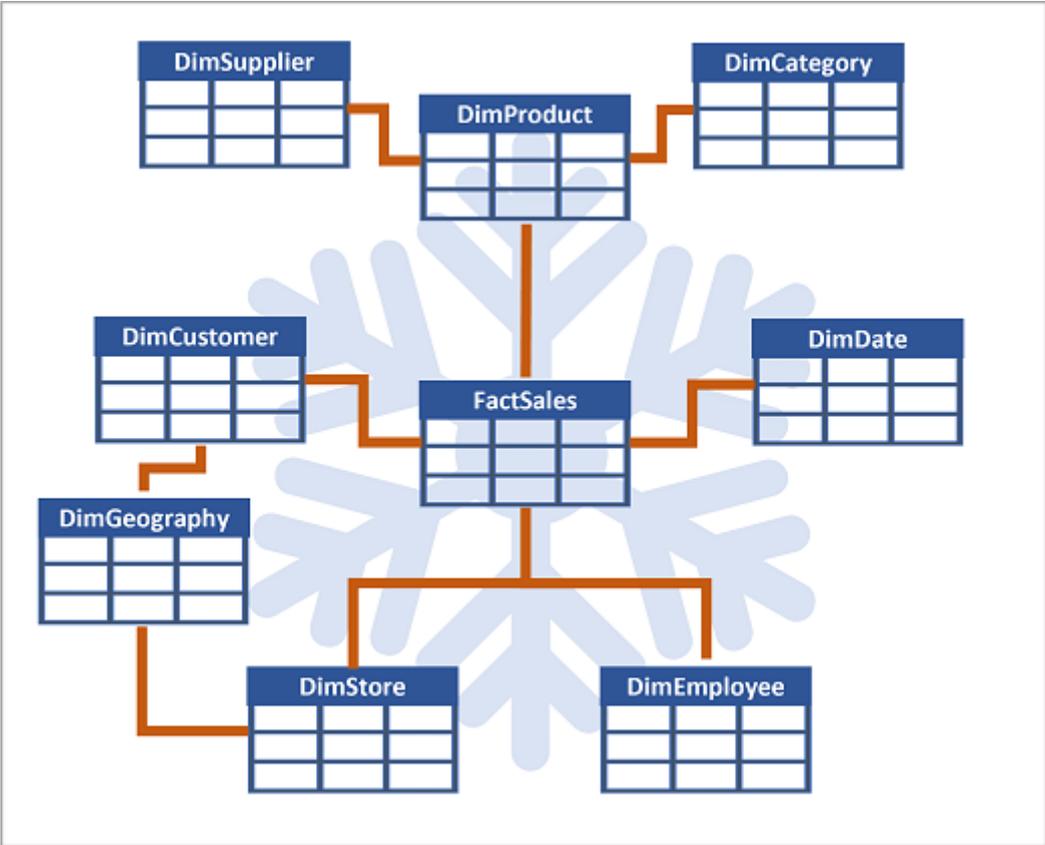
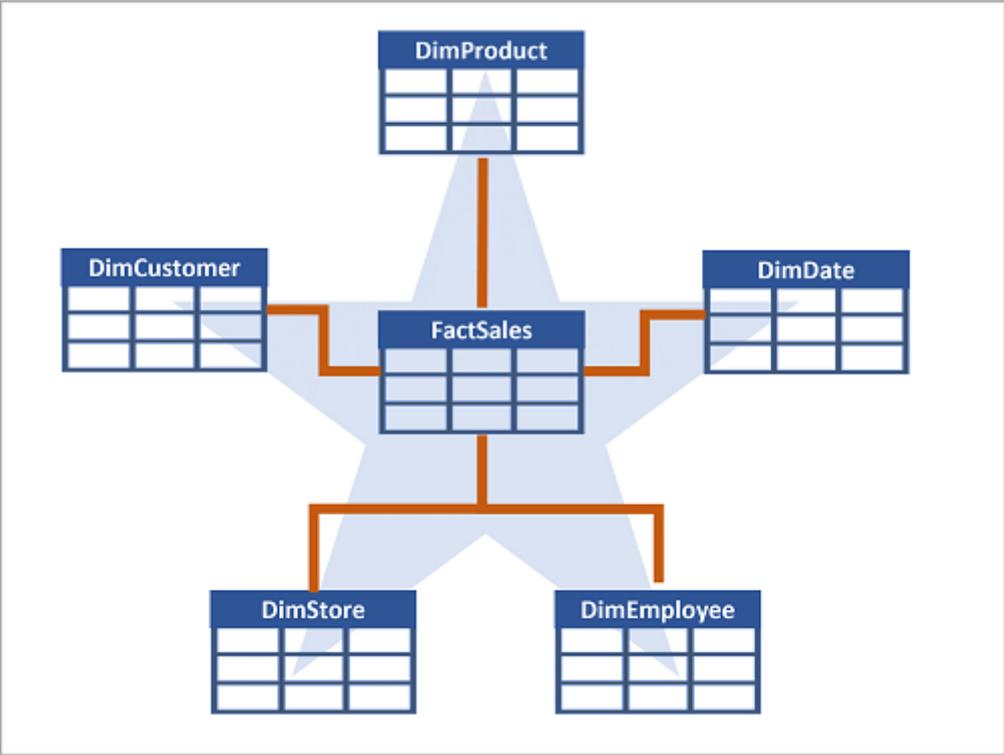
1. **Ingestion des données** : Déplacement des données des systèmes sources vers l'entrepôt de données.
2. **Stockage des données** : Stockage des données dans un format optimal pour l'analyse.
3. **Traitement des données** : Transformation des données pour les rendre utilisables par les outils analytiques.
4. **Analyse et restitution des données** : Analyse des données pour obtenir des insights et leur restitution à l'entreprise.

Microsoft Fabric simplifie ce processus en combinant l'expérience traditionnelle avec des fonctionnalités "low-code". Son entrepôt de données relationnel supporte toutes les fonctionnalités T-SQL attendues, géré, scalable et hautement disponible, stockant et interrogeant les données dans le lakehouse.

Dans Fabric, ingénieurs et analystes peuvent collaborer, partageant la même expérience. Les ingénieurs créent une couche relationnelle sur les données du lakehouse, permettant aux analystes d'utiliser T-SQL et Power BI pour explorer les données.

Pour concevoir un entrepôt de données, les tables sont organisées dans un schéma optimisé pour la modélisation multidimensionnelle. Les tables de faits contiennent les données numériques à analyser, tandis que les tables de dimension fournissent le contexte. Les dimensions spéciales comme celles du temps et à variation lente enrichissent l'analyse des données.

Les schémas d'entrepôts de données, souvent en étoile ou en flocon, déterminent la structure des relations entre les tables, optimisant la performance des requêtes.

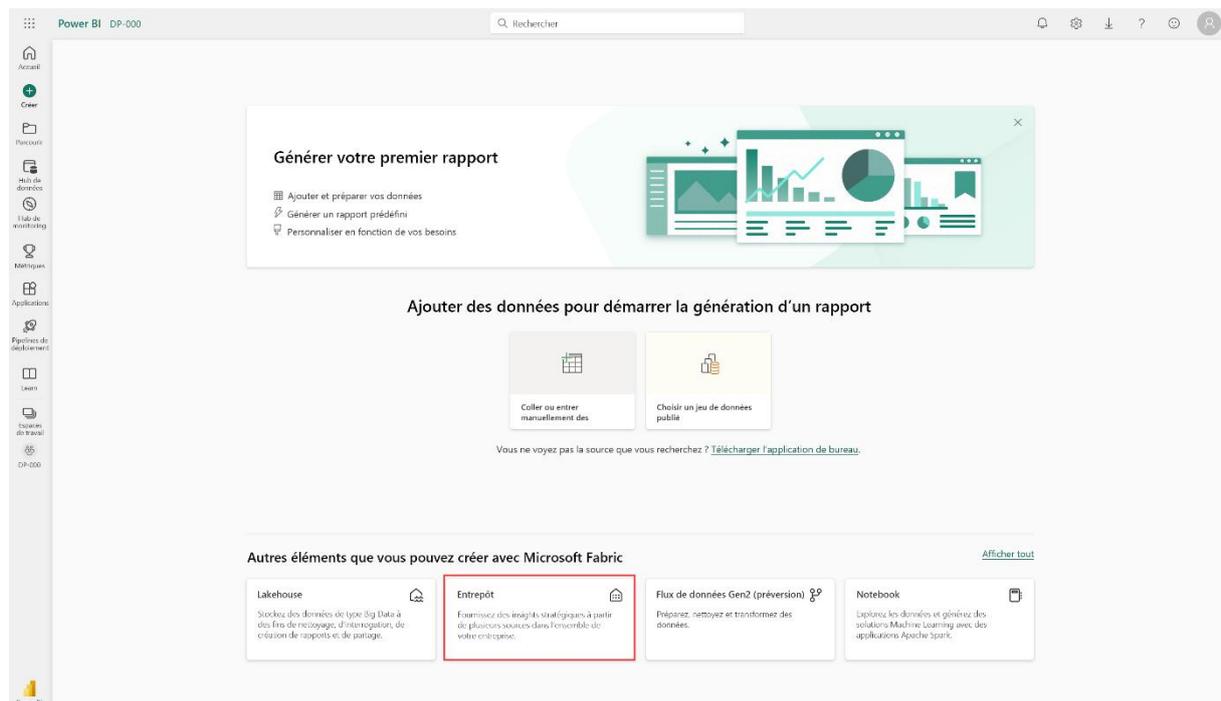


## Comprendre les entrepôts de données dans Fabric

Le lakehouse de Microsoft Fabric constitue une base de données sur un lac de données, utilisé pour stocker des données dans un format optimisé pour le traitement du Big Data par les moteurs Spark et SQL. Il offre des fonctionnalités ACID via des tables au format Delta open source. Cependant, bien que le lakehouse soit idéal pour le stockage des données, il n'est pas optimal pour la restitution des données aux utilisateurs professionnels, pour cela, un entrepôt de données est nécessaire.

L'expérience d'entrepôt de données de Fabric permet de passer de la vue du lac du lakehouse aux expériences SQL d'un entrepôt de données traditionnel. Dans cette expérience, les utilisateurs peuvent modéliser les données à l'aide de tables et de vues, exécuter des commandes T-SQL pour interroger les données, utiliser T-SQL pour effectuer des opérations DML sur les données, et restituer des données à des couches de création de rapports comme Power BI.

Pour créer un entrepôt de données dans Fabric, vous pouvez créer une couche relationnelle au-dessus des données physiques dans le lakehouse et l'exposer à des outils d'analyse et de création de rapports. Vous pouvez ensuite créer des tables directement dans l'interface de Fabric, en utilisant T-SQL.



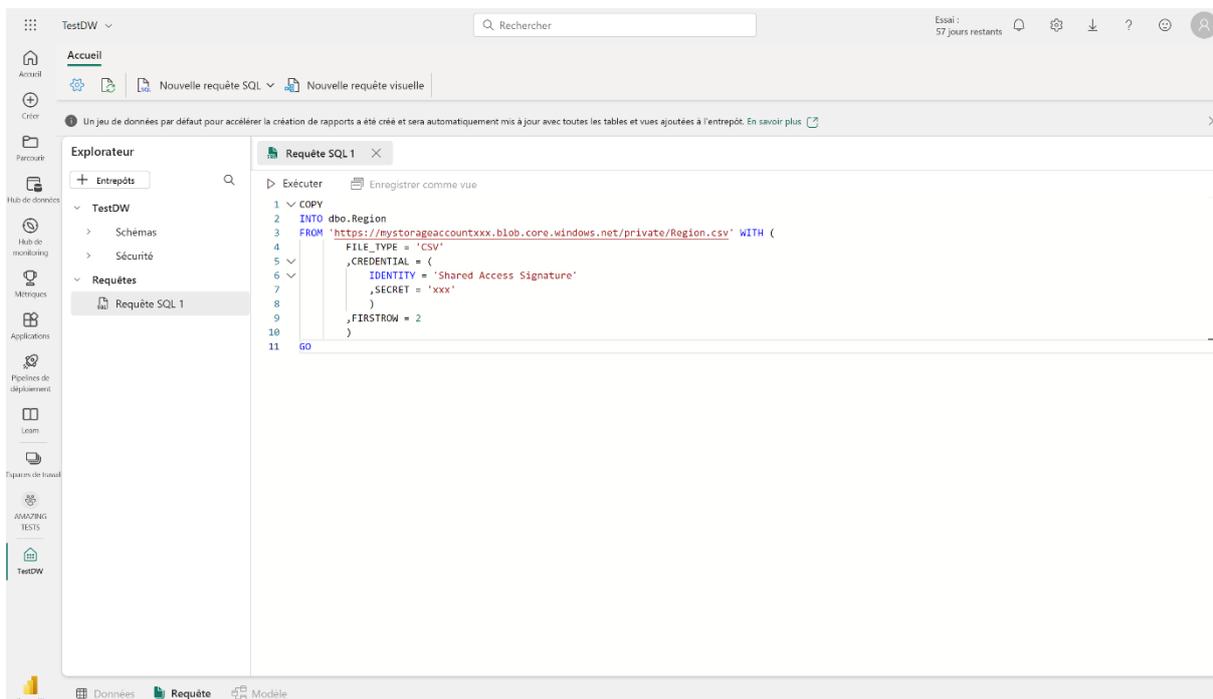
Pour ingérer des données dans un entrepôt de données Fabric, vous disposez de plusieurs méthodes telles que l'utilisation de pipelines, des flux de données, ou encore des requêtes d'interrogation entre bases de données. Vous pouvez également copier des données à partir d'un emplacement externe dans une table de l'entrepôt de données en utilisant la syntaxe COPY INTO.

Après la création des tables dans l'entrepôt de données, vous pouvez implémenter un processus de chargement périodique, souvent réalisé par lots, qui consiste à ingérer

de nouvelles données dans un lac de données, les charger dans des tables de mise en lots dans

```
SQL Copier  
  
COPY INTO dbo.Region  
FROM 'https://mystorageaccountxxx.blob.core.windows.net/private/Region.csv' WITH (  
    FILE_TYPE = 'CSV'  
    , CREDENTIAL = (  
        IDENTITY = 'Shared Access Signature'  
        , SECRET = 'xxx'  
    )  
    , FIRSTROW = 2  
)
```

GO



l'entrepôt de données, charger les tables de dimension à partir des données de dimension, et enfin, charger les tables de faits à partir des données de faits. Il est également possible d'interroger les données dans le lakehouse directement à partir de l'entrepôt de données en utilisant l'interrogation entre bases de données.

## Interroger et transformer des données

Pour interroger et transformer les données dans votre entrepôt de données Fabric, vous avez deux options : l'éditeur de requête visuelle et l'éditeur de requête SQL.

L'éditeur de requête visuelle vous permet de créer des requêtes par "glisser-déposer", idéal si vous préférez éviter le code. Si vous êtes à l'aise avec T-SQL, vous pouvez utiliser l'éditeur de requête SQL pour écrire vos requêtes. Dans les deux cas, vous pouvez créer des tables, des vues et des procédures stockées pour interroger les données dans l'entrepôt de données et le lakehouse.

Pour utiliser l'éditeur de requête SQL, vous pouvez créer une requête en utilisant le bouton "Nouvelle requête SQL" dans le menu, puis écrire et exécuter vos requêtes T-

SQL. Par exemple, vous pouvez créer une vue pour les analystes afin qu'ils puissent créer des rapports dans Power BI.

The screenshot shows the SQL editor interface in Power BI. On the left, the 'Explorateur' (Explorer) pane shows the database structure, including the 'dbo' schema and the 'OrderSummaryView' view. The main editor area displays the following SQL code:

```

1 CREATE VIEW OrderSummaryView AS
2 SELECT
3     d.Month,
4     c.Email,
5     SUM(s.Quantity) AS TotalOrders
6 FROM
7     factsales_gold s
8 JOIN
9     dimdate_gold d ON s.OrderDate = d.OrderDate
10 JOIN
11     dimcustomer_gold c ON s.CustomerID = c.CustomerID
12 GROUP BY
13     d.Month, c.Email;
14

```

Si vous préférez l'éditeur de requête visuelle, utilisez le bouton "Nouvelle requête visuelle" pour commencer. Vous pouvez faire glisser une table de votre entrepôt de données sur le canevas, puis utiliser le menu Transformer pour ajouter des colonnes, des filtres et d'autres transformations à votre requête. C'est similaire à l'affichage des diagrammes dans Power Query en ligne.

The screenshot shows the Power BI interface with the visual query editor. The 'Explorateur' pane on the left shows the database structure. The main editor area displays a visual query diagram with a table 'dimproduct' and a table 'factcustomer'. The 'Transformer' menu is open, showing options like 'Ajouter des colonnes', 'Ajouter des filtres', 'Ajouter des mesures', etc. The bottom pane shows the results of the query, which is a table with columns: 'Month', 'Email', 'TotalOrders', and 'ProductLine'. The data is as follows:

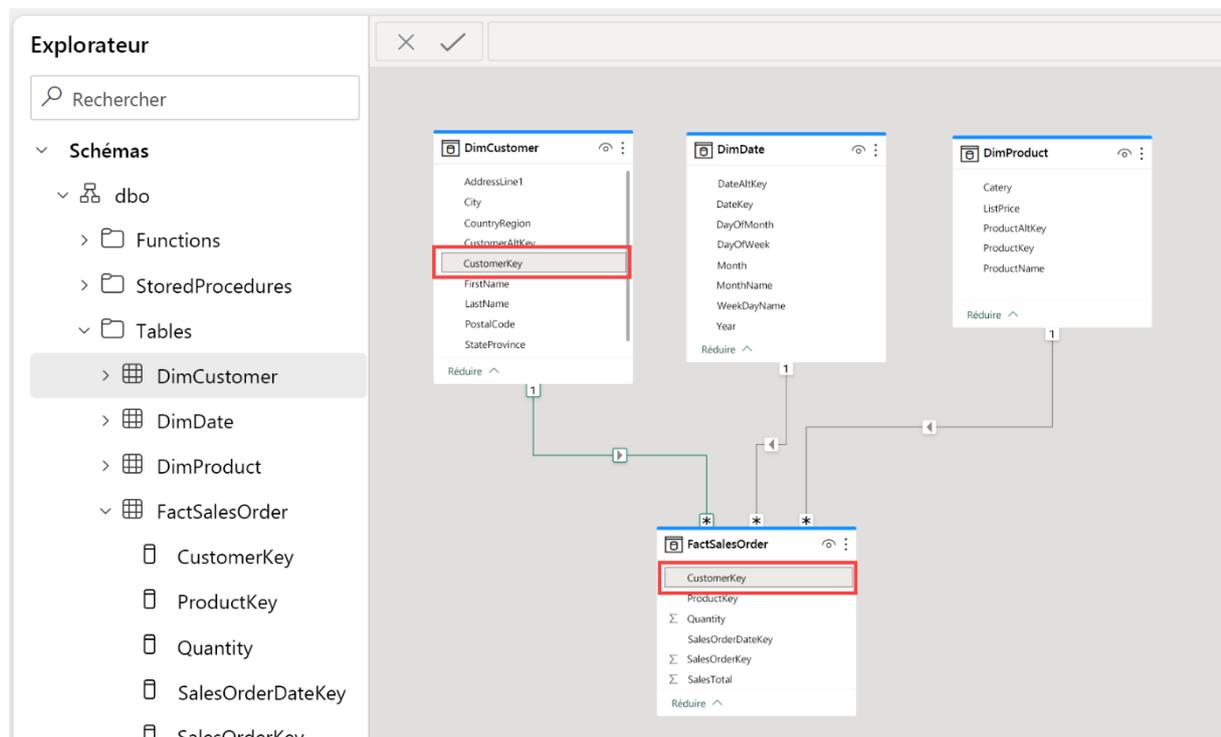
| Month | Email | TotalOrders | ProductLine |
|-------|-------|-------------|-------------|
| 1     | 1528  | 2021102     | 486         |
| 2     | 4605  | 2021021     | 42          |
| 3     | 4029  | 2020021     | 258         |
| 4     | 2029  | 2020101     | 2           |
| 5     | 2524  | 2020011     | 486         |
| 6     | 2917  | 2020011     | 2           |
| 7     | 2625  | 2019011     | 2           |
| 8     | 2627  | 2017011     | 2           |
| 9     | 2078  | 2021102     | 2           |
| 10    | 2108  | 2021021     | 2           |
| 11    | 1419  | 2019118     | 2           |
| 12    | 1206  | 2019011     | 2           |
| 13    | 1528  | 2019011     | 2           |
| 14    | 1528  | 2019011     | 2           |

Quelle option préférez-vous utiliser pour interroger et transformer vos données ?

## Préparer des données pour l'analyse et la création de rapports

Dans Fabric, la création d'un modèle de données sémantique est essentielle pour structurer les relations entre les tables, définir les règles d'agrégation et de synthèse des données, ainsi que pour créer des mesures permettant de générer des insights à partir des données. Ce modèle sémantique est ensuite utilisé pour créer des rapports dans Power BI.

Pour créer un modèle de données sémantique dans Fabric, vous pouvez facilement basculer entre les vues Données, Requête et Modèle en utilisant le menu situé en bas à gauche de l'écran. La vue Données présente les tables du modèle sémantique, la vue Requête les requêtes SQL utilisées pour créer le modèle sémantique, et la vue Modèle le modèle sémantique lui-même.



Dans la vue Modèle, vous pouvez créer des relations entre les tables en utilisant l'interface "cliquer-glisser", ce qui permet de connecter les différentes tables du modèle sémantique. De plus, vous pouvez créer des mesures en utilisant le bouton "Nouvelle mesure", en écrivant des formules en langage de formule DAX (Data Analysis Expressions).

Home Reporting Table tools Measure tools

Data category Uncategorized

New measure

Explorer

Schemas

- dbo
  - Functions
  - StoredProcedures
  - Tables
    - DimCustomer
    - DimDate
    - DimProduct
    - FactSalesOrder
      - CustomerKey
      - Measure
      - ProductKey
      - Quantity
      - SalesOrderDateKey
      - SalesOrderKey
      - SalesTotal
    - Views
    - Queries

```

1 Month over Month Sales =
2 VAR _CurrentMonthSales =
3   CALCULATE ( SUM ( FactSalesOrder[SalesAmount] ), DATESYTD ( DimDate[Date] ) )
4 VAR _PreviousMonthSales =
5   CALCULATE (
6     SUM ( FactSalesOrder[SalesAmount] ),
7     DATESYTD ( DATEADD ( DimDate[Date], -1, MONTH ) )
8   )
9 RETURN
10  DIVIDE ( _CurrentMonthSales - _PreviousMonthSales, _PreviousMonthSales, 0 )

```

DimCustomer: AddressLine1, City, CountryRegion, CustomerAltKey, CustomerKey, FirstName, LastName, ...

DimDate: DateAltKey, DateKey, DayOfMonth, DayOfWeek, Month, MonthName, WeekDayName, ...

DimProduct: Category, ListPrice, ProductAltKey, ProductKey, ProductName, ...

FactSalesOrder: CustomerKey, ProductKey, Quantity, SalesOrderDateKey, SalesOrderKey, SalesTotal, ...

Pour simplifier l'utilisation du modèle sémantique par les utilisateurs, vous pouvez masquer des éléments tels que des tables ou des colonnes en les cachant dans la vue. Cela ne les supprime pas, mais les rend simplement invisibles dans la vue du modèle.

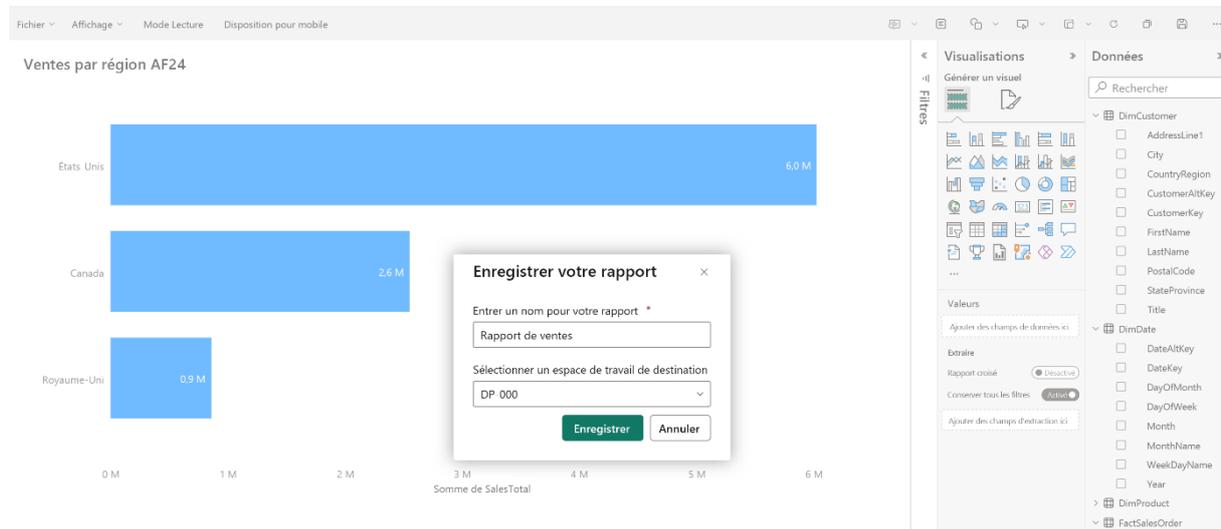
FactSalesOrder

- CustomerKey
- ProductKey
- Σ Quantity
- SalesOrderDateKey
- Σ SalesOrderKey

Chaque fois qu'un entrepôt de données est créé dans Fabric, un modèle sémantique est automatiquement généré. Ce modèle comprend des métriques utilisées pour créer des rapports. Il est synchronisé avec l'entrepôt de données, ce qui évite toute

maintenance manuelle. Vous pouvez également créer des modèles sémantiques personnalisés pour répondre à des besoins spécifiques.

Enfin, Fabric permet de visualiser les données directement dans l'interface, sans avoir à quitter l'expérience d'entrepôt de données. Vous pouvez créer des rapports Power BI basés sur le contenu de votre entrepôt de données, ce qui facilite l'analyse et la prise de décision.



## Sécuriser et monitorer votre entrepôt de données

La sécurité et le monitoring sont des aspects cruciaux de la gestion d'un entrepôt de données dans Fabric.

### Sécurité

**Contrôle d'accès en fonction du rôle (RBAC) :** Vous pouvez définir des rôles pour contrôler l'accès à l'entrepôt et à ses données, garantissant ainsi que seules les personnes autorisées peuvent accéder à certaines fonctionnalités ou données.

**Chiffrement SSL et Azure Storage Service Encryption :** Ces fonctionnalités assurent la sécurité des données en transit et au repos, garantissant que les données sont protégées contre tout accès non autorisé.

**Azure Monitor et Azure Log Analytics :** Ces outils permettent de surveiller l'activité de l'entrepôt et d'auditer l'accès aux données, aidant ainsi à détecter et à répondre rapidement à toute anomalie ou activité suspecte.

**Authentification multifacteur (MFA) :** L'ajout d'une couche de sécurité supplémentaire aux comptes utilisateur grâce à l'authentification multifacteur renforce la sécurité de l'entrepôt.

**Intégration de Microsoft Entra ID :** Cette fonctionnalité permet de gérer les identités utilisateur et l'accès à l'entrepôt, assurant ainsi que seules les personnes autorisées peuvent accéder aux données.

### Autorisations d'espace de travail et d'élément

Les autorisations d'espace de travail permettent de contrôler l'accès et la gestion des données et des services au niveau de l'espace de travail. Les autorisations d'élément permettent quant à elles d'accorder l'accès à des entrepôts de données individuels, offrant ainsi une granularité supplémentaire dans la gestion des autorisations.

## Monitoring

Le monitoring des activités dans l'entrepôt de données est crucial pour garantir des performances optimales, une utilisation efficace des ressources et la sécurité des données. Les vues de gestion dynamique (DMV) permettent de monitorer l'état des connexions, des sessions et des demandes, fournissant ainsi des insights sur le cycle de vie des requêtes SQL en direct.

Les DMV disponibles dans Fabric, tels que

- **sys.dm\_exec\_connections**
- **sys.dm\_exec\_sessions**
- **sys.dm\_exec\_requests**

Permettent d'obtenir des détails sur les connexions, les sessions et les demandes, facilitant ainsi la détection des problèmes de performance et la prise de mesures correctives.

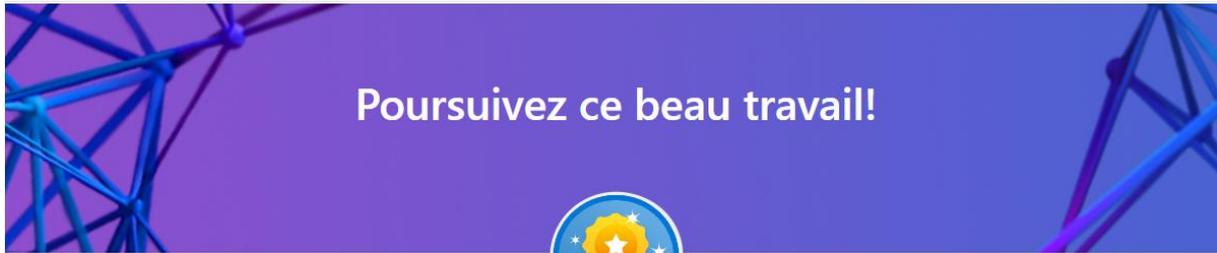
En résumé, la combinaison de fonctionnalités de sécurité robustes, d'une gestion efficace des autorisations et d'un monitoring continu garantit que l'entrepôt de données dans Fabric est sécurisé, bien géré et offre des performances optimales.

## Exercice - Analyser des données dans un entrepôt de données

- Créer un espace de travail
- Créer un entrepôt de données
- Créer des tableaux et insérer des données
- Définir un modèle de données
- Interroger les tables de l'entrepôt de données
- Créer une vue
- Nettoyer les ressources

## Résumé

Dans ce module, les participants ont été initiés aux entrepôts de données et à la modélisation dimensionnelle. Ils ont ensuite procédé à la création d'un entrepôt de données, ainsi qu'au chargement, à l'interrogation et à la visualisation des données. Enfin, une étude des modèles sémantiques et de leur utilisation dans la génération de rapports en aval a été menée.



Bien démarrer avec les  
entrepôts de données dans  
Microsoft Fabric

Vous avez gagné un succès!

Félicitations, mais que devez-vous faire ensuite?

## Présentation de l'analytique de bout en bout à l'aide de Microsoft Fabric

1. **Microsoft Fabric** : Une plateforme d'analytique de bout en bout qui offre un environnement intégré pour les professionnels des données et les entreprises afin de collaborer sur des projets de données.
2. **OneLake** : L'architecture centrée sur le lac de Microsoft Fabric qui fournit un environnement unique et intégré pour stocker et accéder aux données de manière sécurisée et conforme.
3. **Delta** : Un format de stockage par défaut pour OneLake de Fabric, conçu pour fournir une gestion des données efficace et une traçabilité complète des modifications.
4. **Data Factory** : Une charge de travail dans Microsoft Fabric utilisée pour déplacer et transformer des données entre différents systèmes et services.
5. **Ingénierie des données** : Processus de préparation et de transformation des données brutes en un format utilisable pour l'analyse et la génération de rapports.
6. **Science des données** : Domaine qui utilise des algorithmes et des techniques statistiques pour extraire des informations et des connaissances à partir de données structurées et non structurées.
7. **Analytique en temps réel** : Processus d'analyse des données en temps réel pour obtenir des insights instantanés et prendre des décisions en temps opportun.
8. **Capacité Premium** : Un niveau de capacité dans Power BI nécessaire pour utiliser certaines fonctionnalités avancées telles que Microsoft Fabric.

9. **DirectLake** : Un mode dans Microsoft Fabric qui permet aux analystes de données de se connecter directement aux données sans duplication ni déplacement.
10. **Administration Microsoft 365** : Les autorisations requises pour activer Microsoft Fabric au niveau du locataire ou de la capacité dans une organisation.

## Introduction

Microsoft Fabric est une plateforme d'analytique complète qui favorise la collaboration entre les professionnels des données et l'entreprise. Elle offre un ensemble intégré de services pour ingérer, stocker, traiter et analyser les données dans un environnement unifié.

Destinée aux acteurs des données citoyens et aux professionnels, Fabric s'intègre aux outils nécessaires à la prise de décisions. Ses services incluent l'ingénierie des données, l'intégration des données, les entrepôts de données, l'analyse en temps réel, la science des données et le décisionnel.

Ce module présente la plateforme Fabric, son public cible et explore ses différents services pour mieux comprendre son fonctionnement.

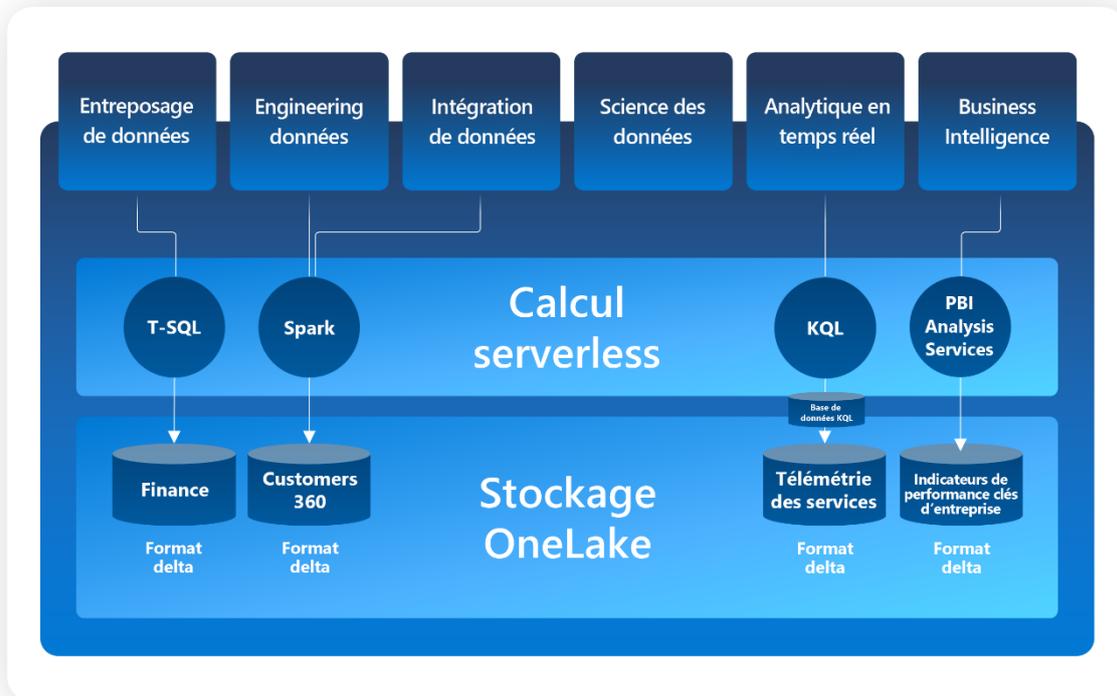
## Explorer l'analytique de bout en bout avec Microsoft Fabric

Microsoft Fabric offre une solution complète d'analytique de bout en bout, simplifiant la complexité et les coûts associés à l'analytique scalable. Il centralise les données dans OneLake, offrant un environnement intégré pour la collaboration entre les professionnels des données et l'entreprise.

OneLake repose sur Azure Data Lake Storage et prend en charge divers formats de données. Il permet aux moteurs de calcul dans Fabric d'accéder aux données sans les déplacer. Les raccourcis facilitent l'accès aux données cloud existantes sans duplication.



Fabric propose plusieurs expériences d'analytique, telles que l'ingénierie des données, l'entrepôt de données, la science des données, l'analytique en temps réel, Data Factory et Power BI, toutes unifiées sur une plateforme unique.



La sécurité et la gouvernance sont centralisées dans le centre d'administration de Fabric, offrant une gestion des groupes et des autorisations, la configuration des sources de données, la supervision de l'utilisation et des performances, ainsi que l'accès aux API d'administration et aux Kits de développement logiciel (SDK). Fabric est nativement intégré à Microsoft Purview Information Protection pour la classification et la protection des données sensibles.

## Équipes de données et Microsoft Fabric

Microsoft Fabric facilite la collaboration entre les professionnels des données en supprimant les silos et en permettant une gestion et une gouvernance unifiées des projets de données. Les équipes de données travaillent ensemble dans un même produit SaaS, améliorant la compréhension mutuelle des besoins et renforçant la capacité à transformer les données en insights.

Avec Fabric, les ingénieurs données peuvent simplifier l'organisation de leurs modèles sémantiques et explorer des techniques de science des données. Les analystes de données bénéficient d'une meilleure traçabilité et peuvent se connecter aux données plus directement grâce au mode DirectLake. Les scientifiques des données peuvent intégrer des techniques de science des données natives et utiliser Power BI pour fournir des insights de manière interactive.

Fabric, en tant que plateforme SaaS, offre une provision rapide et facile de différents types de charges de travail, permettant une adaptation agile aux besoins changeants de l'entreprise. Il intègre également des fonctionnalités à faible niveau de code pour rendre la création et la représentation visuelle du code plus accessibles.

## Activer et utiliser Microsoft Fabric

Pour activer Microsoft Fabric pour votre organisation, vous aurez besoin des autorisations appropriées telles que Administrateur Fabric, Administrateur Power Platform ou Administration Microsoft 365. Cela peut être effectué au niveau du locataire ou de la capacité, selon vos besoins. Si vous ne voyez pas l'icône Power BI dans votre interface, cela signifie que Fabric n'est pas disponible pour vous.

### Portail d'administration

Métriques d'utilisation

Utilisateurs

Premium par utilisateur

Journaux d'audit

Domaines (préversion) **Nouveau**

**Paramètres du locataire**

Paramètres de capacité

Actualiser le résumé

Codes incorporés

Visuels organisationnels

Azure connections

Espaces de travail

Marques personnalisées

Mesures de protection

Contenu proposé

#### Microsoft Fabric (préversion)

▾ Les utilisateurs peuvent créer des éléments Fabric (préversion publique)  
*Activé pour toute l'organisation*

Les utilisateurs peuvent créer des éléments Fabric avec de nouvelles fonctionnalités dans Microsoft Fabric. Ce paramètre peut être géré à la fois au niveau du locataire et au niveau de la capacité. [En savoir plus](#)

Activé

Appliquer à :

Toute l'organisation

Groupes de sécurité spécifiques

À l'exception de groupes de sécurité spécifiques

Déléguer le paramètre à d'autres administrateurs

Les administrateurs de capacité peuvent activer/désactiver cette option

Appliquer Annuler

Les administrateurs peuvent activer Fabric via le Centre d'administration Power BI et le mettre à disposition de l'ensemble de l'organisation ou de groupes d'utilisateurs spécifiques. Une fois activé, vous pouvez créer des espaces de travail Fabric où toutes les ressources, telles que lakehouses, notebooks et pipelines, sont stockées dans OneLake.

Assurez-vous d'avoir accès à une capacité Premium pour utiliser Fabric et sélectionnez "Essai" dans les paramètres de capacité Premium de votre espace de travail pour l'activer. Ensuite, vous pouvez créer des ressources dans Fabric à l'aide du menu "Créer" dans Power BI.

**Workspace settings**

- About
- Premium**
- Azure connections
- System Storage
- Git integration
- Other

Power BI

Data Engineering/Science

### Choose a license

**License mode**

- Pro  
Select Pro to use basic Power BI features and collaborate on reports, dashboards, and scorecards. To access a Pro workspace, users need Pro per-user licenses. [Learn more](#)
- Premium per-user  
Select Premium per-user to collaborate using Power BI Premium features, including paginated reports, dataflows, and datamarts. To collaborate and share content in a Premium per-user workspace, users need Premium per-user licenses. [Learn more](#)
- Premium capacity  
Select premium capacity if the workspace will be hosted in a premium capacity. When you share, collaborate on, and distribute Power BI and Microsoft Fabric content, users in the viewer role can access this content without needing a Pro or Premium per-user license. [Learn more](#)
- Embedded  
Select embedded if the workspace will be hosted in an Azure embedded capacity. ISVs and developers use Power BI Embedded to embed visuals and analytics in their applications. [Learn more](#)
- Fabric capacity  
Select Fabric capacity if the workspace will be hosted in a Microsoft Fabric capacity. With Fabric capacities, users can create Microsoft Fabric items and collaborate with others using Fabric features and experiences. Explore new capabilities in Power BI, Data Factory, Data Engineering, and Real-time Analytics, among others. [Learn more](#)
- Trial  
Select the free trial per-user license to try all the new features and experiences in Microsoft Fabric for 60 days. A Microsoft Fabric trial license allows users to create Microsoft Fabric items and collaborate with others in a Microsoft Fabric trial capacity. Explore new capabilities in Power BI, Data Factory, Data Engineering, and Real-time Analytics, among others. [Learn more](#)

**Default storage format**

- Small dataset storage format
- Large dataset storage format

[Learn more about dataset storage formats](#)

Apply Cancel

Les expériences Fabric regroupent différentes fonctionnalités telles que l'ingénierie des données, la science des données et l'analytique en temps réel. Fabric s'appuie sur Power BI et Azure Data Lake Storage, tout en intégrant des fonctionnalités d'Azure Synapse Analytics, Azure Data Factory, Azure Databricks et Azure Machine Learning dans une expérience SaaS et intégrée.

## Résumé

Dans ce module, les apprenants explorent Microsoft Fabric, une collection d'outils et de services conçus pour répondre aux besoins croissants des professionnels des données et des entreprises. Ils découvrent le stockage OneLake de Fabric, les différentes charges de travail incluses dans Fabric, ainsi que les étapes pour activer et utiliser Fabric dans leur organisation. Avec Fabric, les professionnels des données peuvent travailler de manière sécurisée, conforme et économique avec des données à grande échelle, tandis que l'entreprise peut les utiliser de manière plus efficace et rapide pour prendre de meilleures décisions.

**Poursuivez ce beau travail!**



Présentation de l'analytique de  
bout en bout à l'aide de  
Microsoft Fabric

**Vous avez gagné un succès!**

Félicitations, mais que devez-vous faire ensuite?

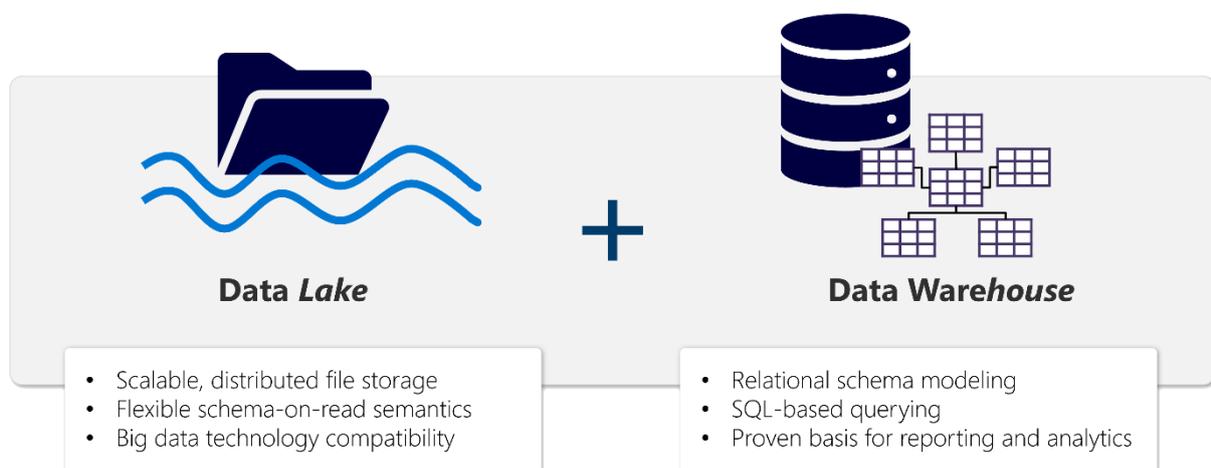
# Bien démarrer avec les lakehouses dans Microsoft Fabric

## Introduction

Le Microsoft Fabric repose sur un lakehouse, combinant stockage scalable, Apache Spark et SQL pour le Big Data. Il intègre le stockage de données flexible et l'analyse de l'entrepôt de données. Le lakehouse répond aux besoins de gestion des données structurées et non structurées pour améliorer la prise de décision.

## Explorer le lakehouse Microsoft Fabric

Un lakehouse repose sur un lac de données utilisant des tables au format Delta et combine les fonctionnalités analytiques SQL d'un entrepôt de données avec la flexibilité d'un lac de données. Les avantages incluent le traitement à grande échelle avec Spark et SQL, la prise en charge des transactions ACID et l'accès centralisé pour les ingénieurs, scientifiques et analystes de données. Dans Microsoft Fabric, vous pouvez créer un lakehouse dans un espace de travail Premium, charger des données à partir de différentes sources et les explorer à l'aide de notebooks ou de flux de données. Des pipelines Data Factory permettent de gérer les processus de transformation de données, et une fois transformées, les données peuvent être interrogées avec SQL, utilisées pour le Machine Learning, l'analyse en temps réel ou les rapports dans Power BI, tout en appliquant des stratégies de gouvernance des données.



## Travailler avec des lacs de données Microsoft Fabric

Pour utiliser les lakehouses dans Microsoft Fabric, commencez par créer un espace de travail et configurez-le pour autoriser les ressources Fabric. Ensuite, créez et configurez un nouveau lakehouse dans la charge de travail Engineering données, ce qui produit trois éléments dans l'espace de travail Fabric : le Lakehouse, un modèle sémantique (par défaut) et un point de terminaison SQL en lecture seule. Vous pouvez explorer les données du lakehouse de deux manières : directement dans le Lakehouse ou via le point de terminaison d'analytique SQL. Les raccourcis vous permettent

d'intégrer des données externes dans votre lakehouse tout en conservant ces données dans un stockage externe, avec des autorisations gérées par OneLake. Plusieurs méthodes, telles que le chargement, les flux de données (Gen2), les notebooks et les pipelines Data Factory, permettent d'ingérer des données dans le lakehouse. Enfin, vous pouvez accorder l'accès au lakehouse au niveau de l'espace de travail ou de l'élément, en contrôlant les autorisations à l'aide de différents rôles et en permettant l'accès direct ou via le point de terminaison d'analytique SQL. La sécurité est essentielle, donc référez-vous à la documentation pour comprendre les différentes options disponibles.

## Explorer et transformer des données dans un lakehouse

Après avoir chargé des données dans le lakehouse, vous pouvez les explorer et les transformer de différentes manières :

- Utilisation d'Apache Spark : vous pouvez utiliser des pools Spark via des notebooks ou des définitions de travaux Spark pour traiter les données dans des fichiers et des tables du lakehouse avec Scala, PySpark ou Spark SQL.
- Notebooks : vous pouvez utiliser des interfaces interactives de programmation pour lire, transformer et écrire des données directement dans le lakehouse sous forme de tables et/ou de fichiers.
- Définitions de travaux Spark : vous pouvez utiliser des scripts à la demande ou planifiés pour traiter des données dans le lakehouse à l'aide du moteur Spark.
- Point de terminaison analytique SQL : chaque lakehouse dispose d'un point de terminaison analytique SQL pour exécuter des instructions Transact-SQL et explorer les données dans des tables de lakehouse.
- Flux de données (Gen2) : en plus de l'ingestion de données, vous pouvez créer des flux de données pour effectuer des transformations via Power Query et renvoyer les données transformées vers le lakehouse.
- Pipelines de données : orchestrez une logique de transformation complexe des données dans le lakehouse via une séquence d'activités telles que des flux de données et des travaux Spark.

Pour analyser et visualiser les données, les tables de lakehouse sont incluses dans un modèle sémantique définissant un modèle relationnel. Vous pouvez modifier ou créer des modèles sémantiques, puis utiliser Power BI pour visualiser et analyser les données, combinant ainsi les fonctionnalités de visualisation avec le stockage et le schéma tabulaire d'un lakehouse pour une solution d'analytique complète sur une plateforme unique.

## Exercice – Créer et ingérer des données avec un lac de données Microsoft Fabric

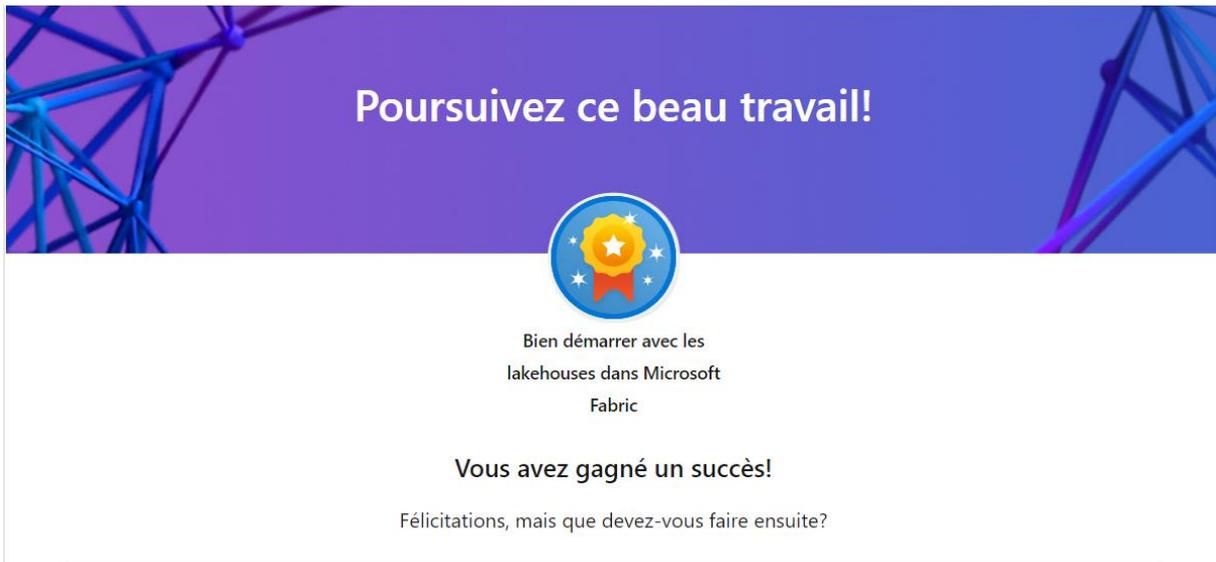
Dans cet exercice, explorez les tâches de lakehouse Microsoft Fabric telles que la création d'un lakehouse, l'importation de données, l'interrogation de tables avec SQL et la génération de rapports. L'exercice met l'accent sur l'importance du lakehouse en tant que composant central dans l'engineering données, l'entreposage et l'analytique, permettant aux utilisateurs de gérer et d'analyser efficacement leurs données dans l'environnement de lakehouse.

## Exercice

- Créer un espace de travail
- Créer une maison au bord du lac
- Télécharger un fichier
- Explorer les raccourcis
- Charger les données du fichier dans une table
- Utiliser SQL pour interroger des tables
- Créer une requête visuelle
- Créer un rapport
- Nettoyer les ressources

## Résumé

Les lakehouses Microsoft Fabric fournissent aux ingénieurs et aux analystes de données les avantages combinés du stockage de lac de données et d'un entrepôt de données relationnelles. Vous pouvez utiliser un lakehouse comme base d'une solution d'analytique données de bout en bout qui comprend l'ingestion, la transformation, la modélisation et la visualisation des données.



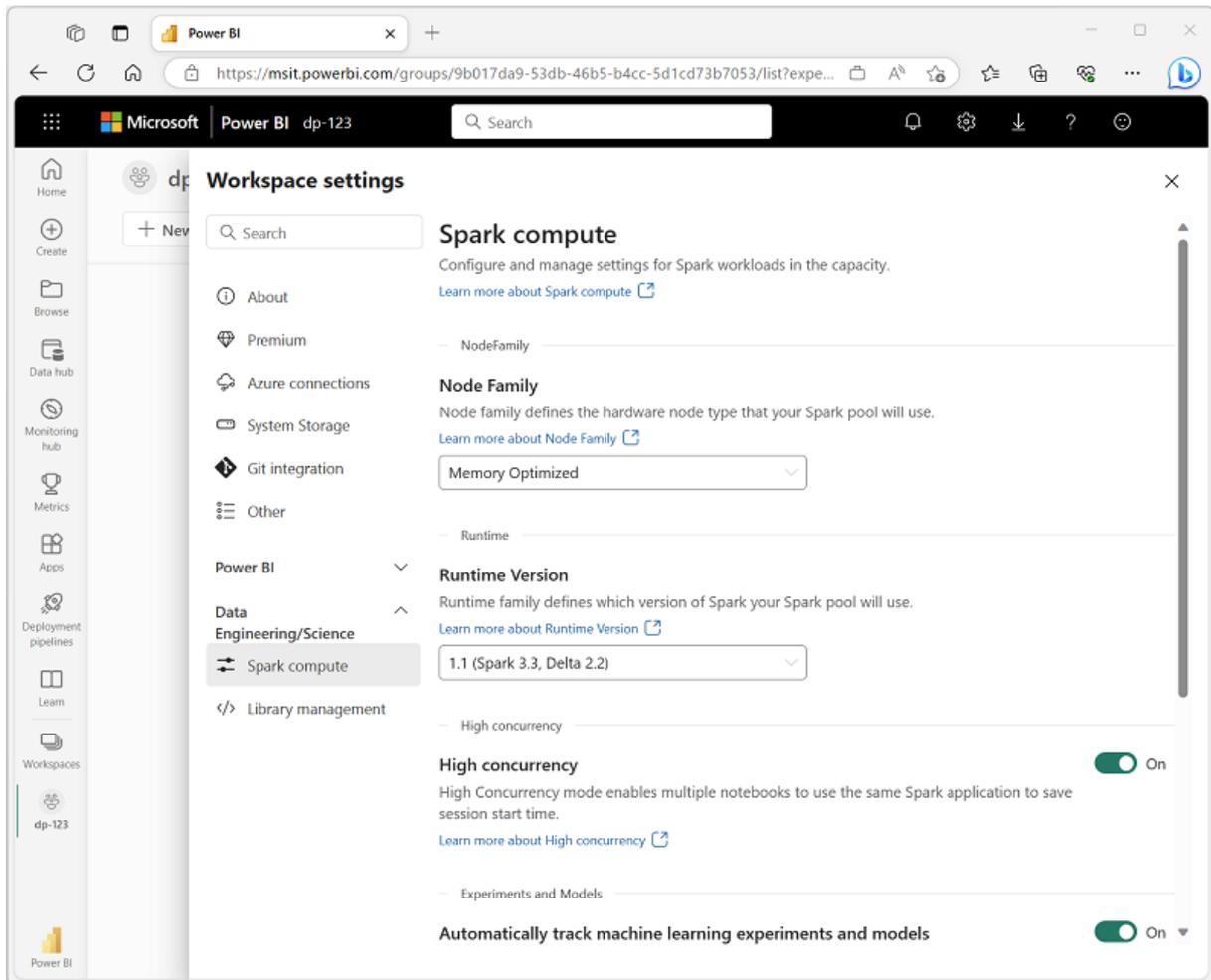
# Utiliser Apache Spark dans Microsoft Fabric

## Introduction

Apache Spark est un framework open source pour le traitement parallèle des données à grande échelle, populaire dans le Big Data et disponible sur plusieurs plateformes, dont Microsoft Fabric. Ce module explore son utilisation dans Fabric pour le traitement des données d'un lakehouse, mettant en avant les avantages d'intégration avec les autres services de données de la plateforme.

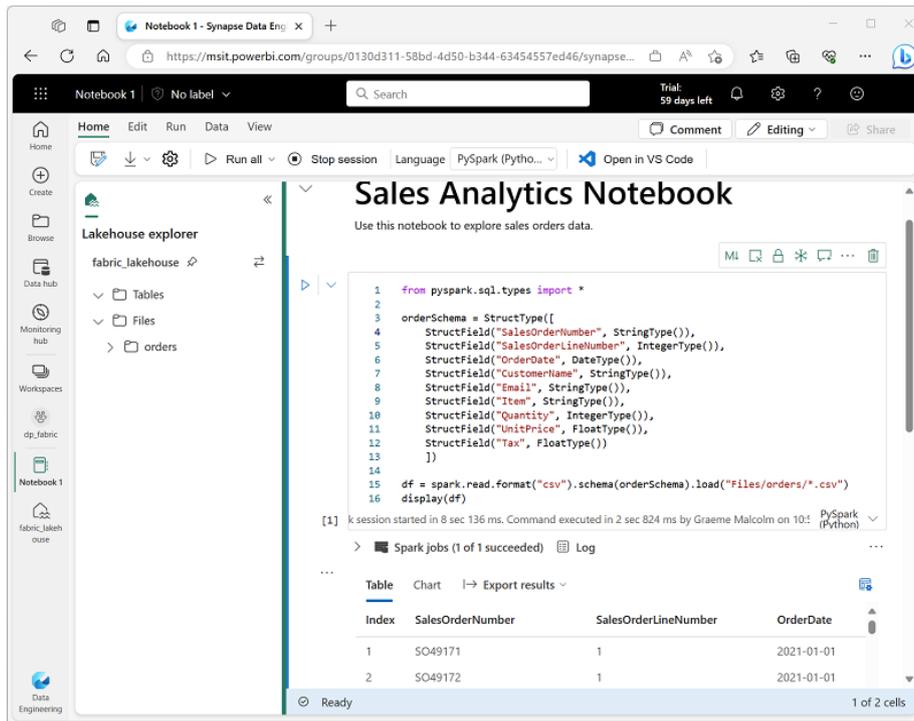
## Se préparer à l'utilisation d'Apache Spark

Apache Spark est un framework de traitement distribué qui permet l'analyse de données à grande échelle en répartissant le travail sur plusieurs nœuds dans un cluster. Spark utilise une approche "diviser et conquérir" pour traiter rapidement de grands volumes de données, gérant la distribution des tâches et le regroupement des résultats. Il prend en charge plusieurs langages de programmation, notamment Java, Scala, Spark R, Spark SQL et PySpark. Dans Microsoft Fabric, chaque espace de travail dispose d'un cluster Spark, avec des paramètres configurables tels que la famille de nœuds et la version du runtime. L'écosystème Spark comprend une variété de bibliothèques pour différentes tâches, avec la possibilité d'ajouter des bibliothèques supplémentaires dans Fabric avec les autorisations d'administrateur.



## Exécuter du code Spark

Pour exécuter du code Spark dans Microsoft Fabric, vous avez deux options : les notebooks et la définition de travaux Spark. Les notebooks permettent une exploration interactive des données en combinant texte, images et code, tandis que les travaux Spark sont utilisés pour l'ingestion et la transformation automatisées des données. Les notebooks se composent de cellules pouvant contenir du texte ou du code exécutable, offrant une interactivité immédiate. Les travaux Spark sont configurés en définissant un script à exécuter sur demande ou selon une planification, avec la possibilité de spécifier un fichier de référence et un lakehouse contenant les données à traiter.



The screenshot shows a Databricks Notebook interface. The notebook is titled "Sales Analytics Notebook" and contains a PySpark code cell. The code defines a schema for sales orders and reads a CSV file. The output is a table with 2 rows and 4 columns.

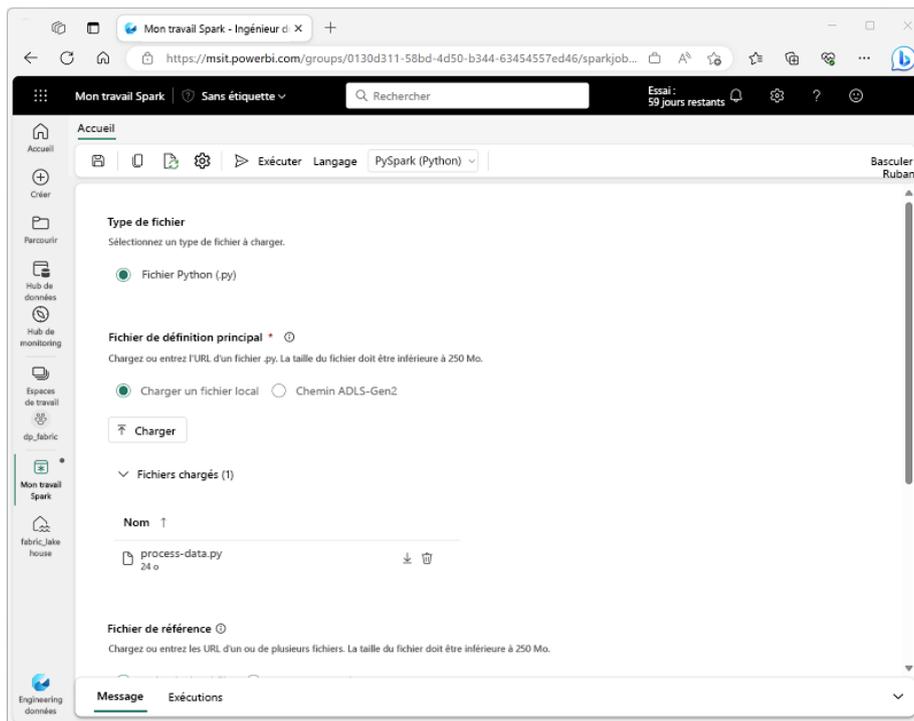
```

1 from pyspark.sql.types import *
2
3 orderSchema = StructType([
4     StructField("SalesOrderNumber", StringType()),
5     StructField("SalesOrderLineNumber", IntegerType()),
6     StructField("OrderDate", DateType()),
7     StructField("CustomerName", StringType()),
8     StructField("Email", StringType()),
9     StructField("Item", StringType()),
10    StructField("Quantity", IntegerType()),
11    StructField("UnitPrice", FloatType()),
12    StructField("Tax", FloatType())
13 ])
14
15 df = spark.read.format("csv").schema(orderSchema).load("Files/orders/*.csv")
16 display(df)

```

Output: k session started in 8 sec 136 ms. Command executed in 2 sec 824 ms by Graeme Malcolm on 105. PySpark (Python)

| Index | SalesOrderNumber | SalesOrderLineNumber | OrderDate  |
|-------|------------------|----------------------|------------|
| 1     | SO49171          | 1                    | 2021-01-01 |
| 2     | SO49172          | 1                    | 2021-01-01 |



The screenshot shows the Databricks workspace interface for uploading a file. The page is titled "Mon travail Spark - Ingénieur d..." and displays the "Fichiers chargés (1)" section. A file named "process-data.py" (24 o) is listed under "Fichiers chargés (1)".

File upload options:

- Fichier Python (.py)
- Fichier de définition principal
- Chemin ADLS-Gen2

Buttons: Charger

Message: Message Exécutions

## Utiliser des données dans un dataframe Spark

Dans Apache Spark, les dataframes sont utilisés principalement pour manipuler des données structurées, en remplacement des RDDs (Resilient Distributed Datasets). Ces dataframes sont intégrés à Spark SQL et fonctionnent de manière similaire aux dataframes de Pandas, mais sont optimisés pour le traitement distribué.

Pour charger des données dans un dataframe, on utilise la méthode **spark.read.load** avec les paramètres nécessaires pour spécifier le format de fichier et l'option

d'inclusion des entêtes de colonne. On peut améliorer les performances en définissant explicitement un schéma, évitant ainsi à Spark de devoir inférer le type de données.

Les dataframes permettent des manipulations variées comme le filtrage, la sélection de colonnes, ou l'agrégation par groupes. Par exemple, on peut sélectionner et afficher les premières lignes avec des commandes simples en PySpark ou utiliser des méthodes comme **select** et **where** pour filtrer et transformer les données. De plus, enregistrer les résultats transformés, notamment dans des formats optimisés comme Parquet, facilite les analyses ultérieures et peut inclure des techniques comme le partitionnement pour améliorer encore les performances.

```
csv Copier

ProductID,ProductName,Category,ListPrice
771,"Mountain-100 Silver, 38",Mountain Bikes,3399.9900
772,"Mountain-100 Silver, 42",Mountain Bikes,3399.9900
773,"Mountain-100 Silver, 44",Mountain Bikes,3399.9900
...
```

## Inférence d'un schéma

Dans un notebook Spark, vous pouvez utiliser le code PySpark suivant pour charger les données du fichier dans un dataframe et afficher les 10 premières lignes :

```
Python Copier

%%pyspark
df = spark.read.load('Files/data/products.csv',
    format='csv',
    header=True
)
display(df.limit(10))
```

La ligne `%%pyspark` au début est appelée *magic* et indique à Spark que le langage utilisé dans cette cellule est PySpark. Vous pouvez sélectionner le langage que vous souhaitez utiliser comme valeur par défaut dans la barre d'outils de l'interface Notebook, puis utiliser une commande magic pour remplacer ce choix pour une cellule spécifique. Par exemple, voici le code Scala équivalent pour l'exemple de données des produits :

```
Scala Copier

%%spark
val df = spark.read.format("csv").option("header", "true").load("Files/data/products.csv")
display(df.limit(10))
```

La commande magic `%%spark` est utilisée pour spécifier Scala.

Ces deux exemples de code produisent une sortie comme suit :

## Utiliser des données à l'aide de Spark SQL

Le texte explique comment l'API Dataframe de Spark, intégrée à Spark SQL, est utilisée pour manipuler des données et comment le catalogue Spark sert de metastore pour les objets de données relationnelles comme les tables et les vues. Spark SQL permet aux analystes de données d'utiliser des expressions SQL pour interroger et manipuler des données, rendant le processus plus naturel pour certains utilisateurs. Le catalogue Spark permet une intégration fluide du code écrit dans les langages supportés par Spark avec SQL.

Les vues temporaires sont créées pour rendre les données d'un dataframe interrogeables dans le catalogue Spark et sont supprimées à la fin de la session. Les tables persistantes peuvent être définies dans le catalogue pour une interrogation récurrente, et les données de ces tables managées sont stockées dans l'emplacement de stockage spécifié par Microsoft Fabric.

Les tables Delta sont privilégiées dans Microsoft Fabric en raison de leur support des transactions, du versioning et de la prise en charge des données de streaming. Les tables externes, qui utilisent des métadonnées stockées dans le catalogue mais obtiennent leurs données d'un emplacement externe, peuvent également être définies.

L'utilisation de l'API Spark SQL permet d'interroger les données à partir du catalogue directement en utilisant SQL, rendant le code plus accessible pour certains développeurs. Par exemple, on peut exécuter des requêtes SQL directement dans un notebook Spark via la commande magique `%%sql`, facilitant l'interrogation et la visualisation des données.

En résumé, Spark SQL et le catalogue Spark offrent des outils puissants pour la gestion de données relationnelles dans un environnement distribué, permettant une manipulation efficace des données à travers des approches déclaratives et programmatiques.

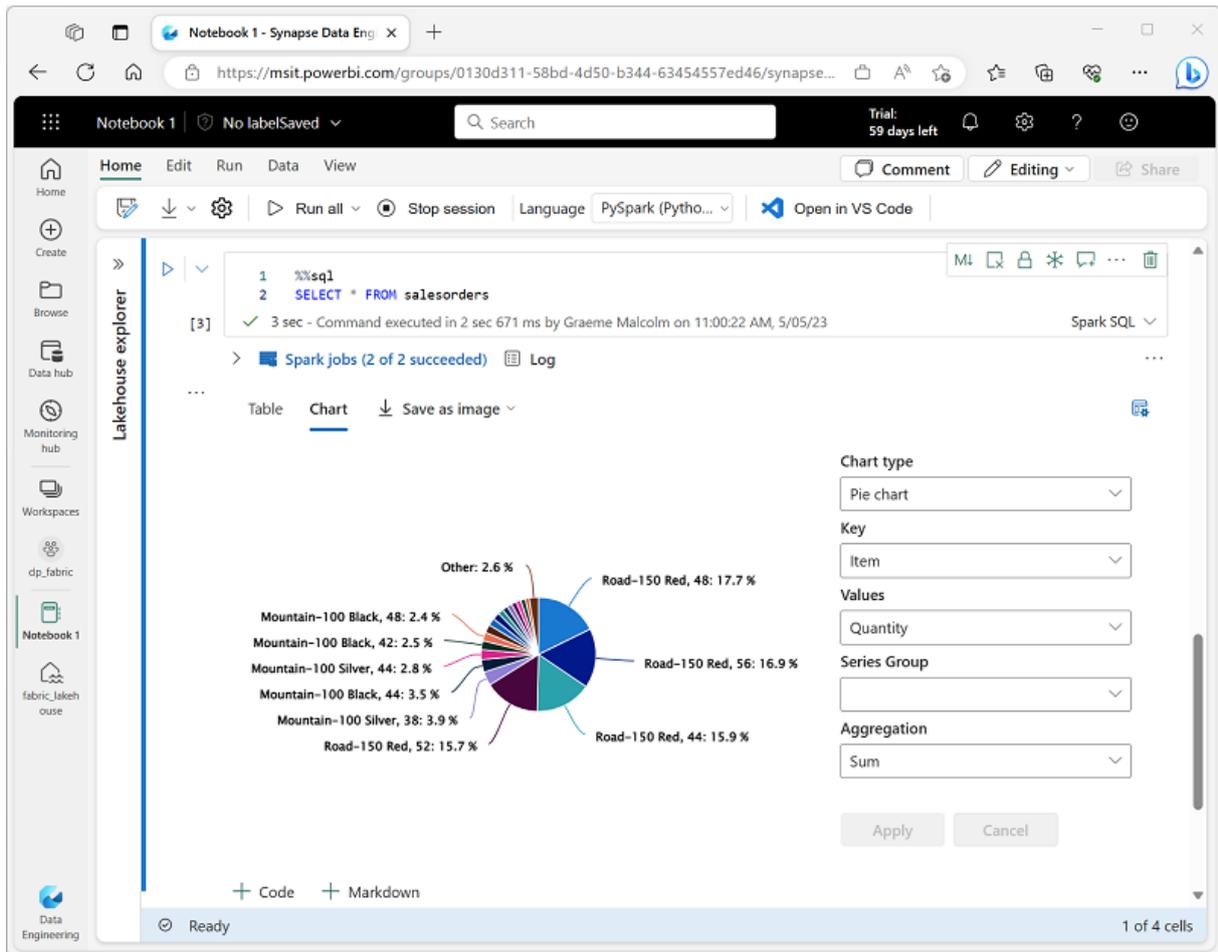
```
SQL Copier  
  
%%sql  
  
SELECT Category, COUNT(ProductID) AS ProductCount  
FROM products  
GROUP BY Category  
ORDER BY Category
```

L'exemple de code SQL retourne un jeu de résultats qui s'affiche automatiquement dans le notebook sous forme de table :

## Visualiser des données dans un notebook Spark

Pour analyser visuellement les résultats des requêtes de données dans Microsoft Fabric, les notebooks offrent des fonctionnalités graphiques intégrées ainsi que la possibilité d'utiliser des bibliothèques graphiques Python. Les résultats d'un dataframe ou d'une requête SQL dans un notebook Spark peuvent être visualisés sous forme de graphiques, permettant une représentation visuelle directe des données.

Les graphiques intégrés dans les notebooks sont pratiques pour une visualisation rapide et simple. Cependant, pour plus de contrôle sur la présentation, des bibliothèques telles que Matplotlib peuvent être utilisées. Ces bibliothèques permettent une personnalisation approfondie des visualisations et sont intégrables directement dans les notebooks, combinant ainsi code, visualisation et documentation en un seul endroit.



Pour créer des graphiques personnalisés, les données Spark peuvent être converties en dataframes Pandas via la méthode **toPandas**. Un exemple concret serait d'utiliser Matplotlib pour générer un graphique à barres des comptages de produits par catégorie à partir de données agrégées. Ce processus implique la création d'une figure, l'ajout d'un graphique à barres et la personnalisation de divers éléments du graphique comme les titres, les axes, et la grille, avant de finalement afficher le résultat.

Python

Copier

```
from matplotlib import pyplot as plt

# Get the data as a Pandas dataframe
data = spark.sql("SELECT Category, COUNT(ProductID) AS ProductCount \
FROM products \
GROUP BY Category \
ORDER BY Category").toPandas()

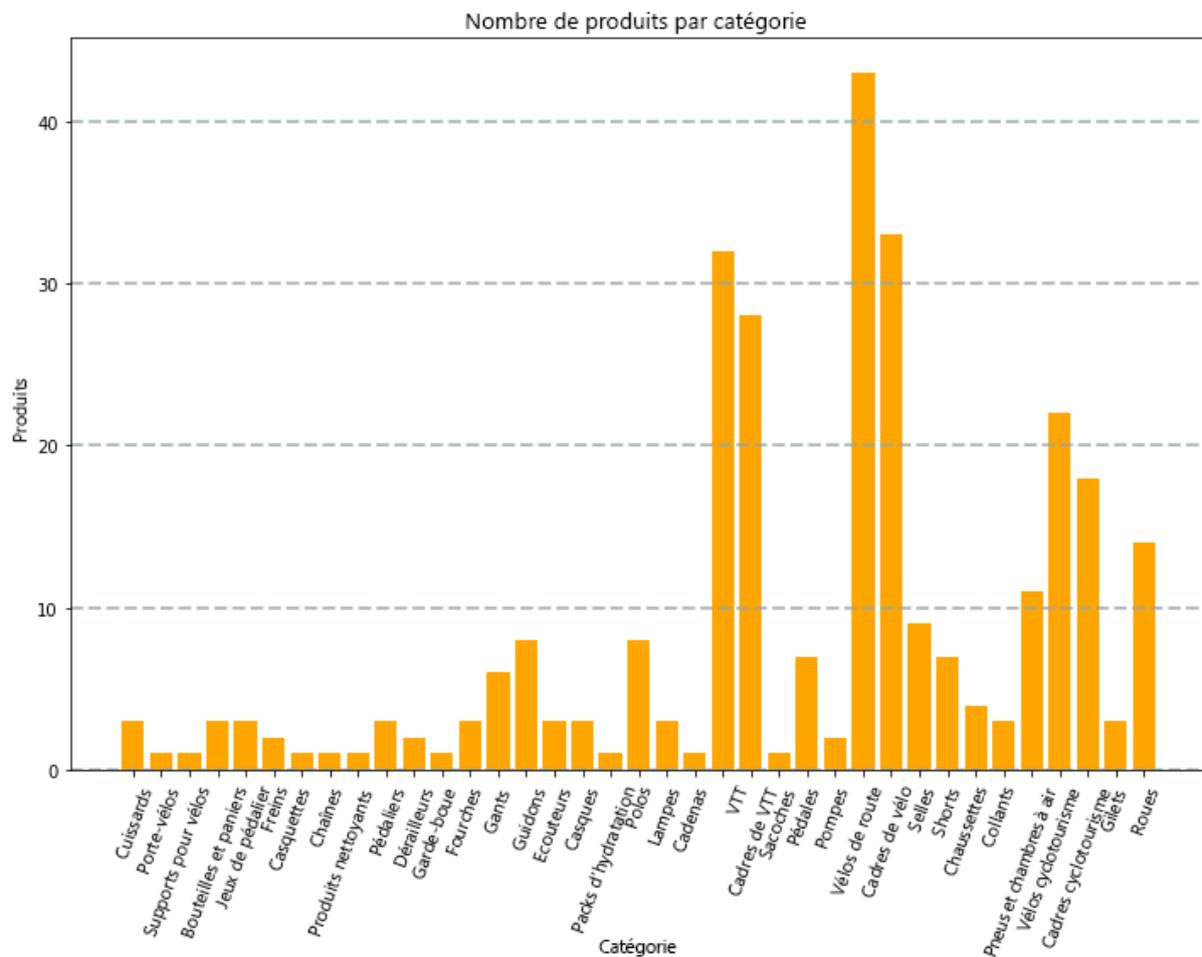
# Clear the plot area
plt.clf()

# Create a Figure
fig = plt.figure(figsize=(12,8))

# Create a bar plot of product counts by category
plt.bar(x=data['Category'], height=data['ProductCount'], color='orange')

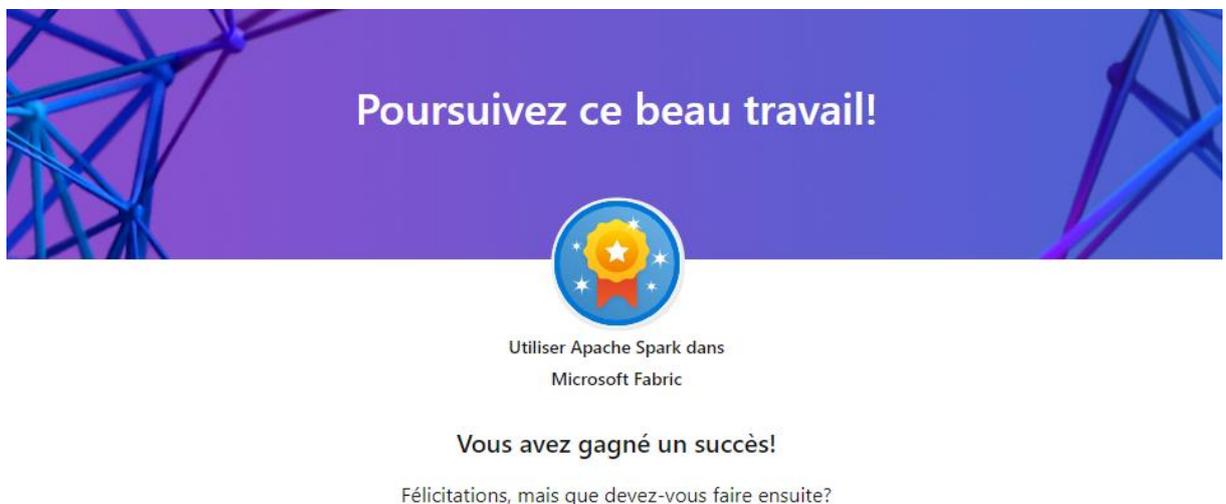
# Customize the chart
plt.title('Product Counts by Category')
plt.xlabel('Category')
plt.ylabel('Products')
plt.grid(color='#95a5a6', linestyle='--', linewidth=2, axis='y', alpha=0.7)
plt.xticks(rotation=70)

# Show the plot area
plt.show()
```



## Exercice

- Créer un espace de travail
- Créez une maison au bord du lac et téléchargez des fichiers
- Créer un cahier
- Charger des données dans un dataframe
- Explorer les données dans un dataframe
- Utilisez Spark pour transformer des fichiers de données
- Travailler avec des tables et SQL
- Visualisez les données avec Spark
- Enregistrez le bloc-notes et terminez la session Spark
- Nettoyer les ressources



## Utiliser des tables Delta Lake dans Microsoft Fabric

### Introduction

Dans Microsoft Fabric, les tables d'un lakehouse sont basées sur le format de table Delta Lake de la Linux Foundation, largement utilisé dans Apache Spark. Delta Lake est une couche de stockage open source pour Spark qui offre des fonctionnalités similaires à celles des bases de données relationnelles pour les données traitées par lots et en streaming. Cette technologie permet de mettre en œuvre une architecture de lakehouse qui intègre une manipulation des données basée sur SQL, avec support des transactions et de la gestion rigoureuse des schémas. Le résultat est un magasin de données analytique qui combine les avantages d'un système de base de données relationnelles avec la flexibilité des systèmes de fichiers dans un lac de données.

Bien que l'utilisation directe des API Delta Lake ne soit pas nécessaire pour travailler avec des tables dans un lakehouse Fabric, comprendre l'architecture de Delta Lake et

connaître les opérations spécialisées sur les tables Delta peut grandement améliorer la capacité à développer des solutions analytiques avancées sur Microsoft Fabric.

## Comprendre Delta Lake

Delta Lake est une couche de stockage open source qui enrichit Spark avec la sémantique d'une base de données relationnelle pour le traitement des données dans un lac de données. Les tables Delta, utilisées dans les lakehouses Microsoft Fabric, sont reconnaissables à une icône triangulaire Delta sur les interfaces utilisateur des lakehouses. Chaque table Delta est une abstraction de schéma représentant des fichiers de données stockés au format Delta, comprenant un dossier contenant des fichiers Parquet et un dossier `_delta_log` pour les journaux de transaction au format JSON.

| ABC | NumComma... | 123 | SalesOrderL... | DateComma... | ABC                 | NomClient...       | ABC | AdresseEmail | ABC | Élément               | 123 | Quantité | 1,2 | PrixUnité | 1,2 | Montant |
|-----|-------------|-----|----------------|--------------|---------------------|--------------------|-----|--------------|-----|-----------------------|-----|----------|-----|-----------|-----|---------|
| 1   | SO51555     | 7   |                | 23/06/2021   | Chloe Garcia        | chloe27@advent...  |     |              |     | Kit de patches/8 p... | 1   | 2,29     |     | 0,1832    |     |         |
| 2   | SO54042     | 7   |                | 09/08/2021   | Logan Collins       | logan29@advent...  |     |              |     | Mitaines...           | 1   | 24,49    |     | 1,9592    |     |         |
| 3   | SO54784     | 7   |                | 22/08/2021   | Autumn Li           | autumn3@advent...  |     |              |     | Vélo à usage génè...  | 1   | 159      |     | 12,72     |     |         |
| 4   | SO58572     | 7   |                | 25/10/2021   | Cesar Sara          | cesar9@adventu...  |     |              |     | Manches courtes...    | 1   | 53,99    |     | 4,3192    |     |         |
| 5   | SO58845     | 7   |                | 30/10/2021   | Peter She           | peter8@adventu...  |     |              |     | Casque Sport-100...   | 1   | 34,99    |     | 2,7992    |     |         |
| 6   | SO58845     | 8   |                | 30/10/2021   | Peter She           | peter8@adventu...  |     |              |     | Manches longues...    | 1   | 49,99    |     | 3,9992    |     |         |
| 7   | SO60233     | 7   |                | 16/11/2021   | Jason Mitchell      | jason40@advent...  |     |              |     | Casque Sport-100...   | 1   | 34,99    |     | 2,7992    |     |         |
| 8   | SO61412     | 7   |                | 03/12/2021   | Nathaniel Cooper    | nathaniel9@adv...  |     |              |     | Manches courtes...    | 1   | 53,99    |     | 4,3192    |     |         |
| 9   | SO62984     | 7   |                | 29/12/2021   | Miguel Sanchez      | miguel72@adve...   |     |              |     | Chaussettes de co...  | 1   | 8,99     |     | 0,7192    |     |         |
| 10  | SO51555     | 6   |                | 23/06/2021   | Chloe Garcia        | chloe27@advent...  |     |              |     | Bouteille de mont...  | 1   | 9,99     |     | 0,7992    |     |         |
| 11  | SO52058     | 6   |                | 04/07/2021   | Elijah Ross         | elijah7@adventu... |     |              |     | Manches courtes...    | 1   | 53,99    |     | 4,3192    |     |         |
| 12  | SO53255     | 6   |                | 28/07/2021   | Edward Taylor       | edward31@adve...   |     |              |     | Manches courtes...    | 1   | 53,99    |     | 4,3192    |     |         |
| 13  | SO53852     | 6   |                | 05/08/2021   | Maria Reed          | maria4@adventu...  |     |              |     | Pack d'hydratatio...  | 1   | 54,99    |     | 4,3992    |     |         |
| 14  | SO54042     | 6   |                | 09/08/2021   | Logan Collins       | logan29@advent...  |     |              |     | Manches courtes...    | 1   | 53,99    |     | 4,3192    |     |         |
| 15  | SO54377     | 6   |                | 15/08/2021   | Ashlee Xu           | ashlee12@adven...  |     |              |     | Bouteille d'eau - ... | 1   | 4,99     |     | 0,3992    |     |         |
| 16  | SO54784     | 6   |                | 22/08/2021   | Autumn Li           | autumn3@adven...   |     |              |     | Kit de patches/8 p... | 1   | 2,29     |     | 0,1832    |     |         |
| 17  | SO55957     | 6   |                | 10/09/2021   | Melissa Richards... | melissa31@adve...  |     |              |     | Bouteille d'eau - ... | 1   | 4,99     |     | 0,3992    |     |         |
| 18  | SO56533     | 6   |                | 21/09/2021   | Max Alvarez         | max5@adventur...   |     |              |     | Nettoyant pour v...   | 1   | 7,95     |     | 0,636     |     |         |
| 19  | SO58845     | 6   |                | 30/10/2021   | Peter She           | peter8@adventu...  |     |              |     | Bouteille d'eau - ... | 1   | 4,99     |     | 0,3992    |     |         |
| 20  | SO59161     | 6   |                | 01/11/2021   | Isabella Long       | isabella20@adve... |     |              |     | Pack d'hydratatio...  | 1   | 54,99    |     | 4,3992    |     |         |
| 21  | SO59384     | 6   |                | 04/11/2021   | Miranda Ross        | miranda4@adve...   |     |              |     | Pack d'hydratatio...  | 1   | 54,99    |     | 4,3992    |     |         |
| 22  | SO60232     | 6   |                | 16/11/2021   | Kristi Malhotra     | kristi21@advent... |     |              |     | Porte-vélos - 4-Bi... | 1   | 120      |     | 9,6       |     |         |

| Nom   | Date de modification | Type    | Taille     |
|---|----------------------|---------|------------|
| _delta_log  | 14/12/2023 12:39...  | Dossier | 2 éléments |
| part-00000-e7d47785-e211-43f9-816d-06c702bbf065-c000... | 14/12/2023 12:39...  | PARQUET | 494 Ko     |

Les tables Delta offrent plusieurs avantages significatifs :

1. **Fonctionnalités relationnelles pour la modification des données** : Avec Apache Spark, les données peuvent être stockées dans des tables Delta qui supportent les opérations CRUD—créer, lire, mettre à jour, et supprimer—similaires à celles des bases de données relationnelles.
2. **Prise en charge des transactions ACID** : Delta Lake apporte à Spark une prise en charge transactionnelle robuste, avec des transactions qui assurent atomicité, cohérence, isolation (isolation sérialisable), et durabilité.
3. **Contrôle de version des données et voyage dans le temps** : Le journal des transactions permet de suivre plusieurs versions de chaque ligne de table et d'accéder à des versions antérieures des données via la fonctionnalité de voyage dans le temps.
4. **Prise en charge des données de traitement par lots et de diffusion en continu** : Alors que les bases de données traditionnelles gèrent des données statiques, les tables Delta dans Spark supportent les données en streaming grâce à l'API Spark Structured Streaming, permettant aux tables Delta d'agir comme cibles et sources pour la diffusion en continu.
5. **Formats standard et interopérabilité** : Les données dans les tables Delta sont stockées au format Parquet, favorisant l'interopérabilité dans les pipelines d'ingestion de lac de données. En outre, les tables Delta peuvent être interrogées en SQL via le point d'analytique SQL du lakehouse Microsoft Fabric.

Ces caractéristiques font des tables Delta une solution puissante pour la gestion de données relationnelles dans un environnement distribué, combinant la flexibilité du stockage de fichiers avec les fonctionnalités avancées des systèmes de bases de données relationnelles.

## Créer des tables delta

Lorsque vous créez une table dans un lakehouse Microsoft Fabric, elle est définie dans le metastore du lakehouse et les données sont stockées dans des fichiers Parquet. Ces détails sont souvent abstraits dans les outils interactifs de Microsoft Fabric, mais l'utilisation d'Apache Spark offre un contrôle accru pour créer et gérer des tables delta.

Pour créer une table delta à partir d'un dataframe dans Spark, le dataframe est d'abord chargé à partir d'un fichier et ensuite enregistré au format delta, enregistrant les données dans des fichiers Parquet et un dossier de journaux de transactions. Voici un exemple de code pour enregistrer un dataframe en tant que table delta :

```
python Copy code  
  
df = spark.read.load('Files/mydata.csv', format='csv', header=True)  
df.write.format("delta").saveAsTable("mytable")
```

Il existe deux types de tables dans ce contexte: les tables gérées, où Spark gère à la fois la définition de la table dans le metastore et les fichiers de données, et les tables

externes, où les fichiers de données sont stockés dans un emplacement spécifié, mais la définition de la table reste dans le metastore. Par exemple, pour créer une table externe :

```
python Copy code  
  
df.write.format("delta").saveAsTable("myexternaltable", path="Files/myexternaltable")
```

En plus de créer des tables à partir de dataframes, vous pouvez également définir des métadonnées de table sans données associées en utilisant l'API DeltaTableBuilder ou des commandes Spark SQL. Par exemple, pour créer une table avec DeltaTableBuilder :

```
python Copy code  
  
from delta.tables import *  
  
DeltaTable.create(spark)  
    .tableName("products")  
    .addColumn("Productid", "INT")  
    .addColumn("ProductName", "STRING")  
    .addColumn("Category", "STRING")  
    .addColumn("Price", "FLOAT")  
    .execute()
```

Ou en utilisant Spark SQL pour créer une table gérée ou externe :

```
sql Copy code  
  
CREATE TABLE salesorders  
(  
    Orderid INT NOT NULL,  
    OrderDate TIMESTAMP NOT NULL,  
    CustomerName STRING,  
    SalesTotal FLOAT NOT NULL  
)  
USING DELTA
```

Pour une table externe spécifiant un emplacement :

```
sql Copy code  
  
CREATE TABLE MyExternalTable  
USING DELTA  
LOCATION 'Files/mydata'
```

Enfin, il est possible d'enregistrer des données au format delta sans créer de définition de table dans le metastore, ce qui permet de conserver les résultats des transformations de données dans un format de fichier modifiable ultérieurement avec

Delta Lake. Voici un exemple d'enregistrement d'un dataframe dans un nouveau dossier au format delta :

```
python Copy code  
  
delta_path = "Files/mydatatable"  
df.write.format("delta").save(delta_path)
```

Ces techniques offrent une flexibilité pour gérer et manipuler des données dans un lakehouse Microsoft Fabric, alliant les avantages de la gestion relationnelle et du stockage de données à grande échelle.

## Utiliser des tables delta dans Spark

Dans Spark, les tables delta et les fichiers au format delta offrent des méthodes flexibles pour récupérer et modifier des données. La manière la plus commune d'interagir avec ces données est l'utilisation de Spark SQL, qui permet d'incorporer des instructions SQL directement dans d'autres langages tels que PySpark ou Scala. Par exemple, vous pouvez insérer ou mettre à jour des données dans une table delta en utilisant des commandes SQL simples, comme montré dans les exemples de code pour insérer une ligne ou mettre à jour le prix d'un produit.

Au-delà de l'utilisation directe de SQL, l'API Delta Lake offre une alternative pour travailler directement avec des fichiers delta sans passer par les tables de catalogue. Cette API permet de créer une instance d'une DeltaTable à partir d'un dossier et d'utiliser diverses méthodes pour modifier les données, telles que la mise à jour conditionnelle des prix.

Un aspect unique des tables delta est leur gestion des versions grâce au journal des transactions, permettant un "voyage dans le temps". Ce mécanisme enregistre chaque modification apportée à la table, vous permettant de consulter l'historique des opérations et de récupérer des versions antérieures des données. Vous pouvez afficher cet historique avec la commande SQL **DESCRIBE HISTORY** ou accéder à une version spécifique des données en utilisant les options **versionAsOf** ou **timestampAsOf** lors de la lecture des fichiers delta.

Ces fonctionnalités rendent les tables delta particulièrement puissantes pour les scénarios où la traçabilité des modifications et la capacité de revenir à des états antérieurs sont cruciales, améliorant ainsi la robustesse des solutions analytiques dans Spark.

## Utiliser des tables delta avec des données de streaming

Les données explorées jusqu'à présent dans Microsoft Fabric étaient statiques, stockées dans des fichiers. Cependant, de nombreux scénarios analytiques nécessitent le traitement de données de diffusion en continu, traitées en temps quasi réel, comme les données issues des appareils IoT.

**Spark Structured Streaming** est une API de Spark qui prend en charge la diffusion en continu. Elle permet de lire en continu les données depuis diverses sources, comme les ports réseau, les services de messages en temps réel, ou les systèmes de fichiers, et de les traiter à l'aide de dataframes dynamiques.

### Utilisation des tables Delta avec Spark Structured Streaming :

- **Comme source** : Les tables Delta peuvent être utilisées pour lire des données en continu, par exemple, des commandes commerciales en ligne ajoutées à la table. Les opérations de lecture doivent se limiter à l'ajout de données, les modifications pouvant entraîner des erreurs sauf si des options spécifiques sont configurées.
- **Comme récepteur** : Les tables Delta peuvent aussi servir à stocker des données en continu, telles que les états d'appareils IoT, en écrivant ces données dans un format Delta sur un emplacement de stockage.

Voici un exemple de création et d'utilisation d'une table Delta comme récepteur de données de streaming :

```
python Copy code  
  
from pyspark.sql.types import StructType, StructField, StringType  
from pyspark.sql.functions import *  
  
inputPath = 'Files/streamingdata/'  
jsonSchema = StructType([  
    StructField("device", StringType(), False),  
    StructField("status", StringType(), False)  
])  
  
stream_df = spark.readStream.schema(jsonSchema).option("maxFilesPerTrigger", 1).json(  
table_path = 'Files/delta/devicetable'  
checkpoint_path = 'Files/delta/checkpoint'  
  
delta_stream = stream_df.writeStream.format("delta").option("checkpointLocation", che
```

Ensuite, pour interroger les données les plus récentes et gérer le flux de données, vous pouvez utiliser des commandes SQL pour créer une table de catalogue et l'interroger, et la méthode **stop** pour arrêter le flux.

Cette intégration des tables Delta dans le traitement de données en continu offre une grande flexibilité et robustesse pour des applications analytiques avancées dans un environnement de lakehouse, combinant les avantages des bases de données relationnelles avec la gestion de données à grande échelle.

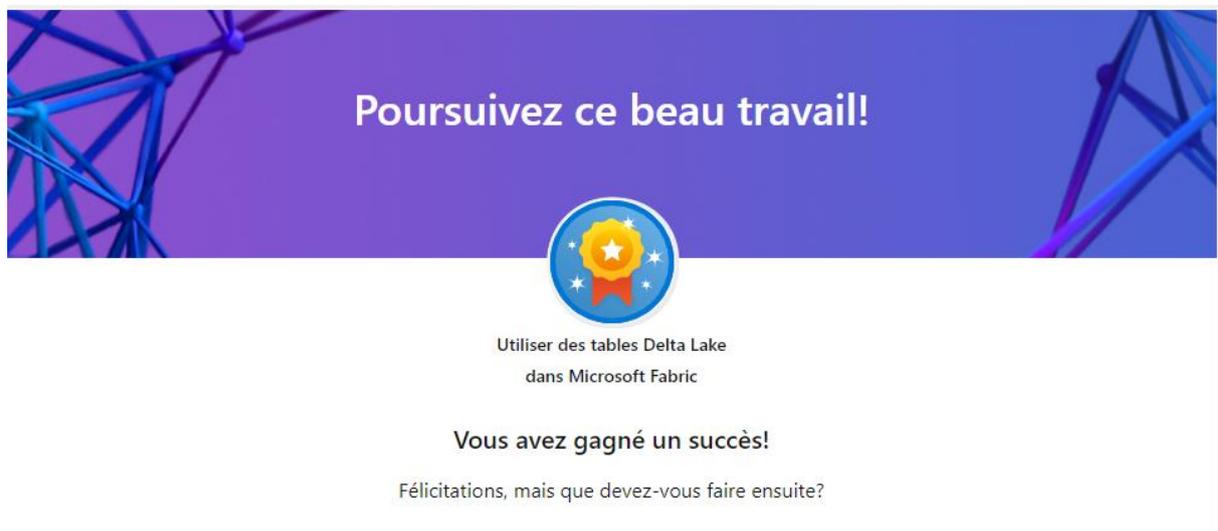
### Exercice

- Créer un espace de travail

- [Créer une maison au bord d'un lac et télécharger des données](#)
- [Explorer les données dans un dataframe](#)
- [Créer des tables delta](#)
- [Explorer la gestion des versions de table](#)
- [Utiliser des tables delta pour diffuser des données](#)
- [Nettoyer les ressources](#)

## Résumé

Delta Lake est une technologie qui ajoute une sémantique de base de données relationnelle à Apache Spark. Les tables d'un lakehouse Microsoft Fabric sont basées sur Delta Lake, ce qui vous permet de tirer parti de nombreuses fonctionnalités et techniques avancées par le biais de l'API Delta Lake.



# Utiliser des pipelines Data Factory dans Microsoft Fabric

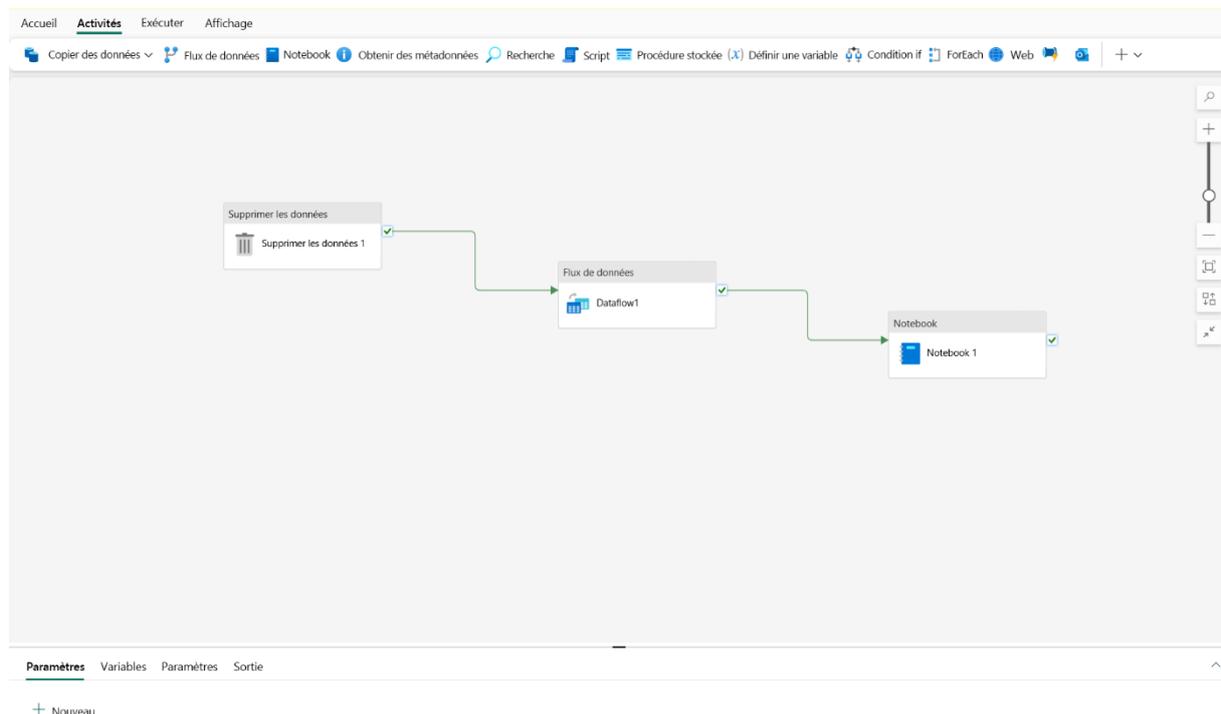
## Introduction

Le texte discute des pipelines de données utilisés pour orchestrer des processus d'extraction, de transformation et de chargement (ETL). Ces pipelines sont essentiels pour automatiser les flux de données depuis des sources transactionnelles vers des magasins analytiques comme les lakehouses ou les entrepôts de données. En utilisant

une architecture similaire à celle d'Azure Data Factory, Microsoft Fabric permet de définir et d'exécuter ces processus via une séquence d'activités connectées. Ces activités peuvent varier en nature, allant du traitement des données à la logique de contrôle de flux, et peuvent être exécutées de manière interactive ou programmée dans l'interface de Microsoft Fabric.

## Comprendre les pipelines

Les pipelines dans Microsoft Fabric permettent de définir et d'orchestrer des activités de traitement de données, y compris leur transfert et leur transformation. Ils utilisent un canevas graphique qui simplifie la création de processus complexes. Voici les concepts clés à comprendre pour travailler avec ces pipelines :



## Concepts principaux des pipelines

- **Activités:** Ce sont des tâches spécifiques au sein du pipeline, divisées en deux catégories :
  - **Activités de transformation des données :** Elles comprennent les opérations de copie simple, les transformations complexes au cours du transfert, l'exécution de notebooks Spark, de procédures stockées SQL, et des activités de suppression de données.
  - **Activités de flux de contrôle :** Ces activités gèrent la logique de flux tels que les boucles, les branches conditionnelles, et la gestion de variables et paramètres, permettant de construire une logique de pipeline complexe.

- **Paramètres:** Les pipelines peuvent être paramétrisés pour améliorer la réutilisabilité, permettant aux utilisateurs de spécifier des variables telles que les noms de dossiers lors de l'exécution du pipeline.
- **Exécutions de pipeline:** Chaque exécution d'un pipeline est suivie par un ID unique qui aide à examiner les détails et le succès de l'exécution. Les exécutions peuvent être déclenchées à la demande ou programmées régulièrement.

En résumé, Microsoft Fabric offre une plateforme robuste pour automatiser et gérer les processus ETL en utilisant des pipelines de données, facilitant ainsi l'ingestion et la transformation des données à grande échelle. Pour une utilisation efficace, il est essentiel de maîtriser les activités de base, la paramétrisation des pipelines, et le suivi des exécutions.

## Utiliser l'activité copier des données

L'activité "Copier des données" est cruciale dans les pipelines de données de Microsoft Fabric, souvent utilisée pour l'ingestion de données simples. Voici un aperçu des éléments clés liés à cette activité :

### L'activité Copier des données

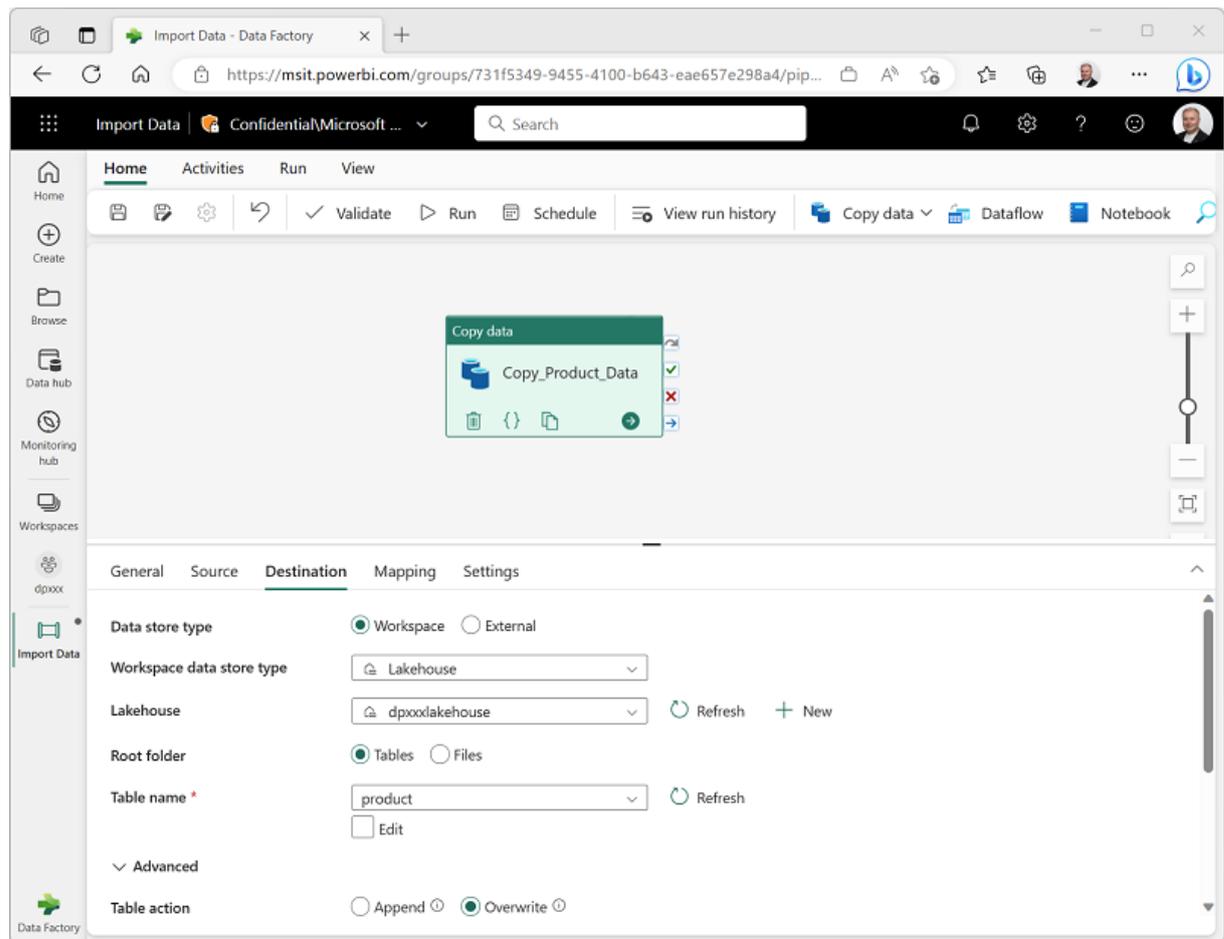
- **Utilisation principale :** Elle permet de copier des données directement d'une source externe à un fichier ou une table dans un lakehouse, souvent utilisée comme première étape dans un pipeline pour l'ingestion initiale de données.
- **Combinaison avec d'autres activités :** L'activité Copier des données peut être combinée avec des activités comme Supprimer des données (pour nettoyer les données existantes avant la copie) et des activités de transformation (comme exécuter un notebook Spark pour transformer les données après leur copie).

### Outil Copier des données

- **Interface graphique :** Un outil graphique assiste les utilisateurs dans la configuration de l'activité, depuis la sélection de la source de données jusqu'à la destination de l'opération de copie.
- **Support de sources variées :** Un large éventail de connexions sources est supporté, facilitant l'ingestion de données à partir de divers systèmes.

### Paramètres de l'activité Copier des données

- **Configuration :** Après avoir ajouté une activité Copier des données au pipeline, elle peut être sélectionnée et ses paramètres modifiés via un volet dédié dans l'interface.



### Quand utiliser l'activité Copier des données

- **Sans transformations** : Idéale pour les cas où les données doivent être copiées sans modification entre la source et la destination.
- **Avec transformations ultérieures** : Utile pour importer des données brutes qui seront transformées par des activités subséquentes dans le pipeline.

### Avancé

- **Activité de flux de données (Gen2)** : Pour des cas plus complexes nécessitant des transformations pendant l'ingestion ou la fusion de données de plusieurs sources, l'utilisation de l'activité de flux de données (Gen2) avec Power Query peut être envisagée. Cette approche permet d'intégrer plusieurs étapes de transformation dans un pipeline.

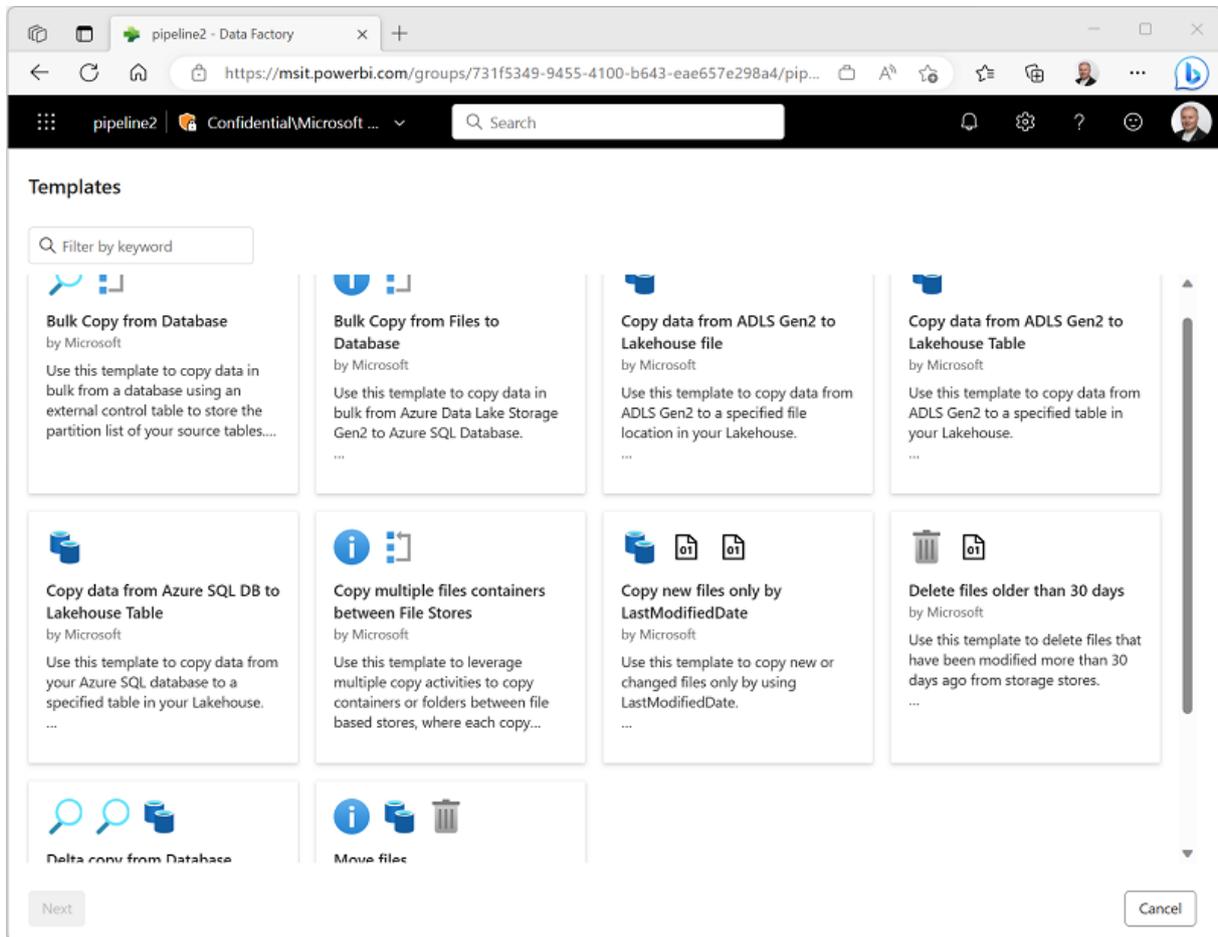
### Conseil

- **Formation recommandée** : Pour ceux qui souhaitent approfondir l'utilisation des flux de données (Gen2) pour l'ingestion de données, Microsoft Fabric propose un module spécifique sur ce sujet.

Ce résumé couvre l'utilisation et la configuration de l'activité Copier des données dans Microsoft Fabric, ainsi que des conseils pour son utilisation efficace dans des scénarios d'ingestion de données.

## Utiliser des modèles de pipeline

Dans Microsoft Fabric, l'utilisation de modèles de pipeline préconçus permet une création rapide et personnalisable de pipelines pour divers scénarios d'ingestion et de transformation de données. Voici comment cela fonctionne :



### Utilisation des modèles de pipeline

- **Flexibilité** : Vous pouvez démarrer avec n'importe quelle combinaison d'activités pour créer des pipelines adaptés à vos besoins spécifiques.
- **Modèles prédéfinis** : Pour les scénarios courants, Microsoft Fabric offre des modèles de pipeline qui facilitent le démarrage rapide. Ces modèles couvrent une gamme de cas d'utilisation typiques et sont entièrement personnalisables.

### Création d'un pipeline à partir d'un modèle

- **Sélection du modèle** : Dans l'interface utilisateur de Microsoft Fabric, commencez par sélectionner la vignette 'Choisir une tâche pour démarrer' dans un nouveau pipeline. Cela vous mènera à une sélection de modèles de pipeline.
- **Personnalisation** : Après avoir choisi un modèle approprié, vous pouvez le personnaliser selon vos besoins spécifiques en utilisant le canevas de pipeline. Ce processus vous permet d'ajuster ou d'ajouter des activités pour répondre précisément à vos exigences.

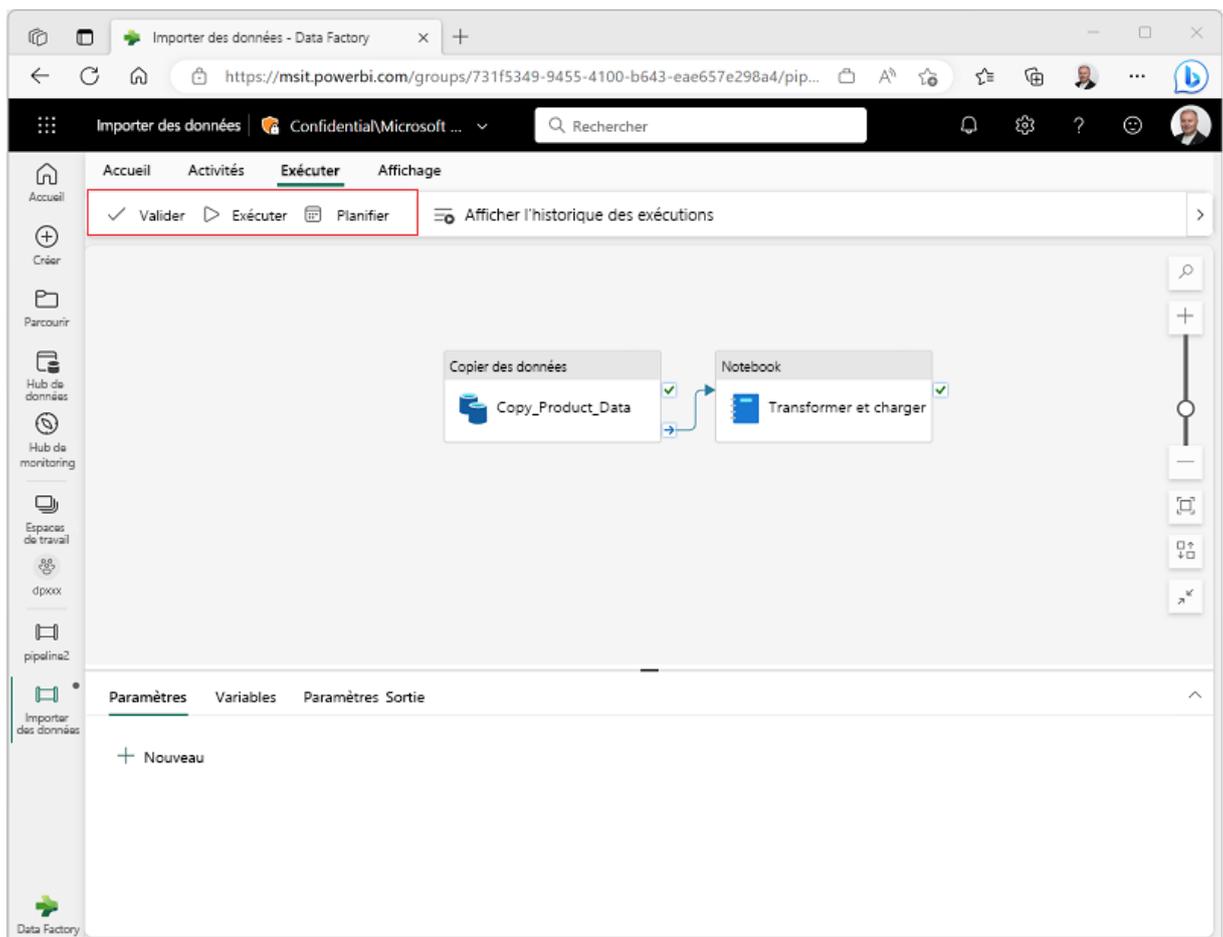
Ces modèles facilitent non seulement la mise en place rapide de solutions d'ingestion et de transformation de données mais offrent également une base solide pour des adaptations plus détaillées et complexes.

## Exécuter et superviser des pipelines

L'exécution et la supervision des pipelines dans Microsoft Fabric sont des étapes clés pour gérer les processus d'ingestion et de transformation de données. Voici comment vous pouvez procéder :

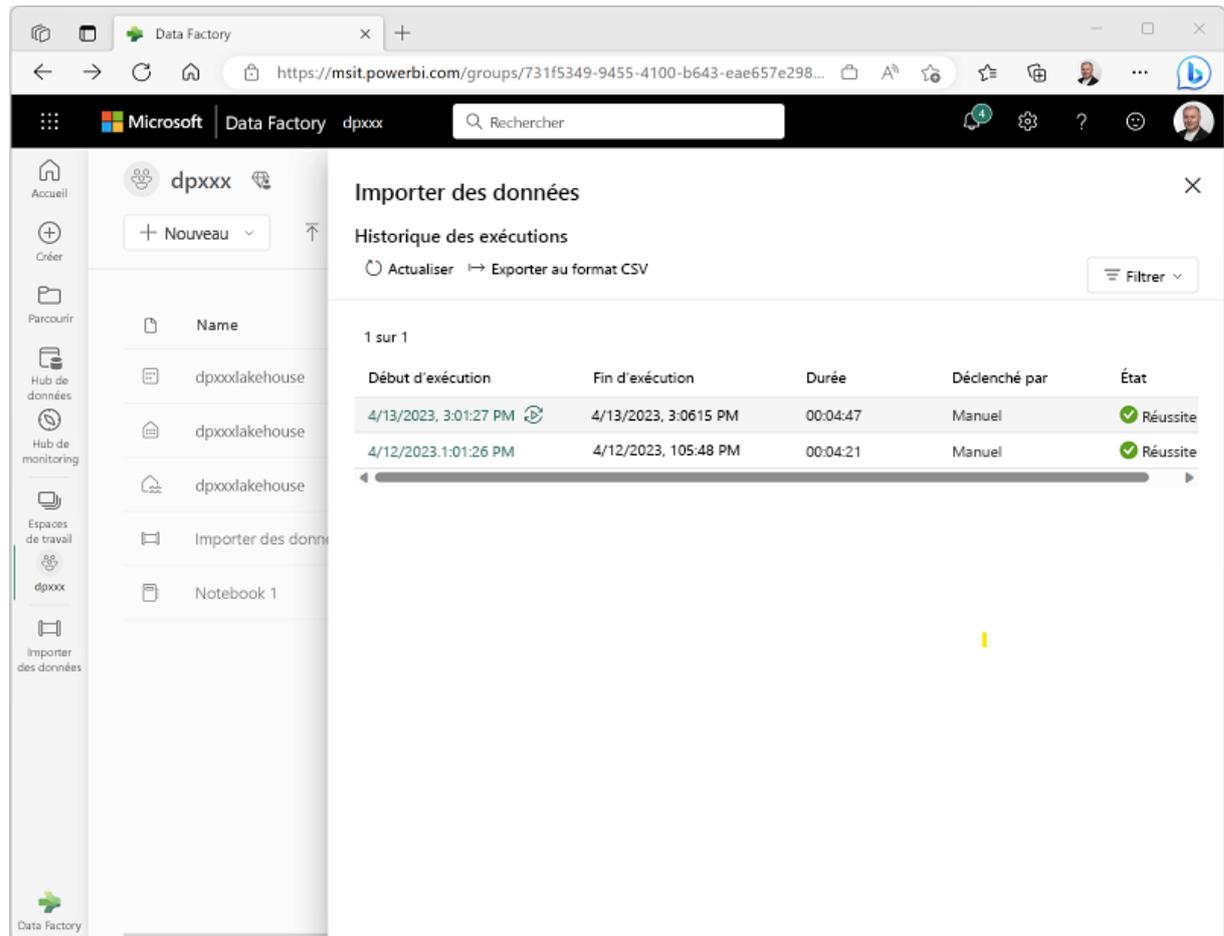
### Exécuter et superviser des pipelines

- **Validation du pipeline** : Avant de lancer un pipeline, utilisez l'option **Valider** pour vérifier que sa configuration est correcte. Cette étape est essentielle pour s'assurer que tous les éléments du pipeline sont configurés correctement et prêts à fonctionner sans erreurs.
- **Exécution du pipeline** : Une fois validé, le pipeline peut être exécuté de manière interactive ou selon une planification prédéfinie. Cela permet de lancer des processus d'ETL à des moments spécifiques ou en réponse à certaines conditions.



### Supervision des exécutions

- **Historique des exécutions** : Microsoft Fabric permet de visualiser l'historique des exécutions de chaque pipeline. Cela se fait depuis le canevas du pipeline ou via la liste des pipelines dans la page de l'espace de travail.
- **Détails des exécutions** : Pour une exécution spécifique, vous pouvez consulter des détails plus approfondis, y compris le temps d'exécution de chaque activité. Ces informations sont souvent présentées sous forme de graphique de Gantt, ce qui aide à comprendre le déroulement et la performance de chaque partie du pipeline.



The screenshot displays the Microsoft Fabric Data Factory interface. The main content area is titled "Importer des données" and shows the "Historique des exécutions" (Execution History) for a pipeline. The history is presented as a table with the following columns: "Début d'exécution" (Start of execution), "Fin d'exécution" (End of execution), "Durée" (Duration), "Déclenché par" (Triggered by), and "État" (Status). Two execution records are visible, both marked as "Réussite" (Success).

| Début d'exécution     | Fin d'exécution       | Durée    | Déclenché par | État       |
|-----------------------|-----------------------|----------|---------------|------------|
| 4/13/2023, 3:01:27 PM | 4/13/2023, 3:06:15 PM | 00:04:47 | Manuel        | ✓ Réussite |
| 4/12/2023, 1:01:26 PM | 4/12/2023, 10:5:48 PM | 00:04:21 | Manuel        | ✓ Réussite |

### Importance de la supervision

La capacité à superviser et à analyser chaque exécution permet non seulement de s'assurer du bon fonctionnement des pipelines, mais aussi de détecter et de corriger les éventuelles inefficacités ou erreurs dans le processus.

Ces fonctionnalités de Microsoft Fabric sont essentielles pour maintenir l'intégrité et l'efficacité des opérations de données à grande échelle, facilitant une gestion robuste des processus ETL dans des environnements complexes.

Tous les artefacts > Importer des données

Réexécuter Annuler Actualiser Mettre à jour le pipeline Liste Gantt

Copier des données ✓ Notebook ✓

Copy\_Product\_Data Transformer et charger

Détails de l'exécution du pipeline

Propriétaire  
Graeme Malcolm

Exécuter par  
Déclencheur manuel

Heure de début  
13/4/2023, 15:01:27

Heure de fin  
13/4/2023, 15:06:15

État  
✓ Réussite

Paramètres  
0

ID d'exécution du pipeline  
d7066dd6-499d-4d93-93b4-19ed7c296538

Exécutions de l'activité

Tous les états Exporter au format CSV

13 avril 03:02:00 13 avril 03:03:00 13 avril 03:04:00 13 avr 03:06:00

Transformer et...  
Copy\_Product...

## Exercice

- Créer un espace de travail
- Créer une maison au bord du lac
- Créer un pipeline
- Créer un cahier
- Modifier le pipeline
- Nettoyer les ressources

Poursuivez ce beau travail!

Utiliser des pipelines Data  
Factory dans Microsoft Fabric

Vous avez gagné un succès!

Félicitations, mais que devez-vous faire ensuite?

# Ingérer des données avec des flux de données Gen2 dans Microsoft Fabric

## Présentation

Microsoft Fabric propose une solution intégrée pour l'ingénierie, l'intégration et l'analytique des données. Un élément clé de cette solution est l'utilisation des flux de données Gen2, qui permettent l'ingestion et la transformation de données provenant de diverses sources, facilitant ainsi leur nettoyage et leur transfert vers des destinations spécifiques. Ces flux peuvent être intégrés dans des pipelines de données pour un traitement ultérieur et utilisés comme sources dans Power BI.

Un scénario d'utilisation typique impliquerait un ingénieur données dans une entreprise de vente au détail mondiale, préparant et transformant des données de différents magasins pour l'analyse et le reporting. Les flux de données Gen2 simplifient ce processus en unifiant les sources disparates et en garantissant la cohérence des données, tout en offrant des possibilités de réutilisation et de mise à jour facile des données. Ce module détaille comment exploiter les flux de données Gen2 dans Microsoft Fabric pour optimiser l'ingestion de données, remplaçant les méthodes manuelles fastidieuses et sujettes aux erreurs.

## Comprendre les flux de données Gen2 dans Microsoft Fabric

Les flux de données Gen2 de Microsoft Fabric représentent une avancée majeure pour l'ingénierie, l'intégration et l'analytique des données dans des environnements cloud. Ils offrent une solution ETL (Extract, Transform, Load) qui facilite la création et l'exécution de processus de transformation de données à grande échelle, grâce à Power Query Online et une interface visuelle intuitive.

### Utilisation des flux de données Gen2

Ces flux permettent de connecter différentes sources de données, les transformer et les préparer pour l'analyse ou d'autres opérations. Ils peuvent ensuite être intégrés dans des pipelines de données pour un traitement supplémentaire ou utilisés directement comme sources dans des outils comme Power BI.

### Scénario d'application

Dans le cas d'une entreprise de vente au détail mondiale, un ingénieur des données utiliserait les flux de données Gen2 pour normaliser les données provenant de divers magasins et les rassembler dans un modèle sémantique unifié. Cela facilite l'analyse et la génération de rapports à l'échelle de l'entreprise.

### Options de gestion des données

Les flux de données peuvent être stockés directement dans un Lakehouse ou traités via un pipeline de données vers d'autres destinations. Ils supportent également le processus ELT (Extract, Load, Transform), où les données sont extraites et chargées avant d'être transformées.

## Avantages

- **Cohérence des données:** Ils permettent de standardiser et d'uniformiser les données, comme les tables de dimensions de dates, et d'assurer leur qualité.
- **Accès en libre-service:** Facilitent l'accès aux données pour les utilisateurs finaux sans exposer directement l'entrepôt de données.
- **Performances optimisées:** Les flux de données permettent une réutilisation efficace des données, réduisant le temps de traitement pour les sources lentes.
- **Simplicité:** Offre une interface low-code pour l'intégration de données de sources diverses.

## Limitations

- **Non substituable à un entrepôt de données complet:** Ils ne remplacent pas les fonctionnalités complètes d'un entrepôt de données.
- **Sécurité:** Ne prennent pas en charge la sécurité au niveau des lignes.
- **Exigences de capacité:** Nécessitent un espace de travail avec une capacité suffisante pour gérer les opérations.

En résumé, les flux de données Gen2 dans Microsoft Fabric simplifient l'ingestion, la transformation, et l'intégration des données à grande échelle, tout en offrant des options flexibles pour le traitement et l'analyse des données. Toutefois, il est essentiel de considérer leurs limites lors de la planification de l'architecture de données d'une entreprise

## Explorer un flux de données Gen2 dans Microsoft Fabric

Explorer un flux de données Gen2 dans Microsoft Fabric vous permet d'ingérer, de visualiser et de transformer des données efficacement via la charge de travail Data Factory ou l'espace de travail Power BI. L'utilisation de Power Query Online enrichit cette expérience en offrant une interface visuelle pour gérer les transformations de données.

The screenshot displays the Power Query Online interface. At the top, there is a search bar and a ribbon with tabs: Home, Transform, Add column, View, and Help. The Home tab is active, showing various options like 'Get data', 'Enter data', 'Options', 'Manage parameters', 'Refresh', 'Advanced editor', 'Add data destination', 'Choose columns', 'Remove columns', 'Keep rows', 'Remove rows', 'Filter rows', 'Sort', 'Transform', 'Combine', 'Map to entity', and 'Export temp'. Below the ribbon is a 'Queries' pane on the left, showing a list of queries including 'orders' and 'orders-2'. The main area shows a data preview for the 'orders' query, with columns: SalesOrderID, OrderDate, MonthNo, CustomerID, Lineltem, ProductID, and OrderID. The data is displayed in a table format. On the right, the 'Query settings' pane is open, showing the 'Properties' section with the name 'orders', 'Entity type' set to 'Custom', and a list of 'Applied steps' including 'Source', 'Promoted h...', 'Changed co...', 'Appended ...', 'Filtered rows', 'Added cust...', 'Reordered c...', and 'Changed co...'. The 'Data destination' section shows 'Lakehouse' as the selected destination. A 'Publish' button is visible at the bottom right.

## Interface Power Query Online

### 1. Ruban Power Query:

- **Connecteurs de source:** Supporte une variété de sources comme des bases de données cloud et locales, des fichiers Excel, SharePoint, et Salesforce.
- **Transformations possibles:** Incluent filtrer, trier, fusionner des requêtes, manipuler des colonnes, calculer des rangs et des pourcentages, et plus encore.

### 2. Volet Requête:

- Affiche les sources de données, appelées requêtes, avec des options pour renommer, dupliquer, et gérer les requêtes.

### 3. Affichage Diagramme:

- Permet de visualiser les connexions entre les sources de données et les transformations appliquées.

### 4. Volet Aperçu des données:

- Montre un aperçu des données permettant d'évaluer l'impact des transformations. Offre des fonctionnalités interactives pour réorganiser ou modifier les données directement.

### 5. Volet Paramètres de la requête:

- Contient les détails des étapes de transformation appliquées, avec une option pour visualiser et modifier le code M à travers l'éditeur avancé.

## Advanced editor



```
1 let
2   Source = Lakehouse.Contents([]),
3   Navigation = Source[{workspaceId = "91dd9a9b-e5b1-4d41-88f1-3e315bdd9f7e"}][Data],
4   #"Navigation 1" = Navigation[{lakehouseId = "5c04bd01-14eb-4393-8a2e-b15277568c26"}][Data],
5   #"Navigation 2" = #"Navigation 1"#{Id = "COVID_data", ItemKind = "Table"}[Data],
6   #"Added rank column" = Table.AddRankColumn(#"Navigation 2", "Rank", {{"positive", Order.Descending}}),
7   #"Added index" = Table.AddIndexColumn(#"Added rank column", "Index", 0, 1, Int64.Type),
8   #"Marked key columns" = Table.AddKey(#"Added index", {"Index"}, false),
9   #"Split column by delimiter" = Table.SplitColumn(#"Marked key columns", "load_time", Splitter.SplitTextByEachDelimiter
10  ({" "}, QuoteStyle.Csv), {"load_time.1", "load_time.2"}),
11  #"Changed column type" = Table.TransformColumnTypes(#"Split column by delimiter", {{"load_time.1", type date},
12  {"load_time.2", type time}}),
13  #"Filtered rows" = Table.SelectRows(#"Changed column type", each ([hospitalized] <> null)),
14  #"Changed column type 1" = Table.TransformColumnTypes(#"Filtered rows", {{"hospitalized", Int64.Type}}),
15  #"Duplicated Column" = Table.AddColumn(#"Changed column type 1", "iso_subdivision - Copy", each [iso_subdivision],
16  type text)
17 in
18  #"Duplicated Column"
```

## Intégration avec Power BI

Les données transformées peuvent être facilement intégrées et consommées via Power BI Desktop, permettant aux analystes de se connecter directement aux flux de données et de les utiliser dans leurs modèles sémantiques et rapports.

### Conseil

Pour optimiser l'utilisation des données, il est conseillé de connecter directement les analystes de données au flux via Power BI Desktop, facilitant l'accès et l'exploitation des données transformées dans des analyses plus complexes.

Ces outils et interfaces de Microsoft Fabric offrent une plateforme puissante pour l'ingestion et la transformation de données, permettant une intégration fluide des données à travers différentes sources et destinations, tout en simplifiant les processus ETL pour les ingénieurs et les analystes de données.

## Intégrer des flux de données Gen2 et des pipelines dans Microsoft Fabric

L'intégration des flux de données Gen2 et des pipelines dans Microsoft Fabric permet une gestion efficace et flexible des opérations sur les données. Voici comment ces outils fonctionnent ensemble pour améliorer le traitement des données :

### Utilisation des flux de données Gen2

Les flux de données Gen2 sont spécialement conçus pour l'ingestion et la transformation des données. Ils permettent de préparer les données pour des analyses ultérieures ou pour être chargées dans des destinations comme un Lakehouse. Ils offrent un contrôle précis sur les transformations de données et peuvent être ajustés visuellement via Power Query Online.

### Création et utilisation des pipelines de données

Les pipelines dans Microsoft Fabric facilitent l'orchestration de diverses activités d'ingénierie de données, tels que :

- **Copier des données:** Transférer des données entre sources et destinations.

- **Incorporer un flux de données:** Utiliser un flux de données préexistant pour transformer les données.
- **Ajouter un notebook:** Exécuter des scripts de notebook pour des analyses complexes.
- **Obtenir les métadonnées:** Extraire des métadonnées pour une gestion des données améliorée.
- **Exécuter un script ou une procédure stockée:** Automatiser l'exécution de scripts pour le traitement des données.

Ces pipelines offrent une interface visuelle où vous pouvez organiser ces activités dans un ordre spécifique, permettant un flux de travail continu et automatisé pour les données transformées par les flux de données Gen2.

### Processus typique

1. **Ingestion et transformation initiale:** Utilisez un flux de données Gen2 pour ingérer et transformer les données, les préparant pour l'analyse ou d'autres traitements.
2. **Intégration dans un pipeline:** Une fois le flux de données complété, incorporez-le dans un pipeline de données pour orchestrer d'autres activités nécessaires, comme l'exécution de scripts supplémentaires ou l'analyse des données transformées.

### Exemple d'intégration

Vous pouvez démarrer avec un flux de données qui ingère et transforme les informations. Une fois ce processus terminé, le flux de données peut être intégré dans un pipeline où d'autres opérations sont orchestrées, comme la mise à jour d'un Lakehouse ou l'exécution de procédures qui dépendent des données transformées.

L'ensemble offre une solution robuste et flexible pour le traitement des données dans des environnements complexes, rendant le processus de gestion des données à grande échelle plus efficace et moins sujet à des erreurs manuelles.

### Exercice

Dans ce module, nous avons décrit un scénario dans lequel les ingénieurs Données et les analystes de données ont besoin de préparer des données en vue de leur consommation et d'étendre des modèles sémantiques. Nous avons identifié que les flux de données Gen2 constituent la meilleure solution pour les étapes de transformation des données. Dans cet exercice, vous créez et utilisez un flux de données Gen2 pour ingérer des données transformées dans un Lakehouse.

- Créer un espace de travail
- Créer une maison au bord du lac
- Créer un Dataflow (Gen2) pour ingérer des données
- Ajouter une destination de données pour Dataflow
- Ajouter un flux de données à un pipeline

- Nettoyer les ressources

## Résumé

Avec Microsoft Fabric, vous pouvez créer des flux de données Gen2 pour effectuer l'intégration des données pour votre Lakehouse, et éventuellement inclure le flux de données dans un pipeline de données. Dans ce module, vous avez découvert un flux de données Gen2 et comment les utiliser dans le cadre de votre processus d'intégration des données. Power Query Online offre une interface visuelle permettant d'effectuer des transformations de données complexes sans écrire de code.



## Bien démarrer avec Real-Time Analytics dans Microsoft Fabric

### Introduction

Microsoft Fabric est une plateforme complète pour les solutions de données, intégrant l'analytique en temps réel via Synapse Real-Time Analytics. Cette fonction utilise une base de données Kusto (KQL) pour gérer le stockage et l'analyse des données, avec un accent particulier sur les données structurées ou textuelles et les séries chronologiques issues de fichiers journaux ou de services de streaming. Le langage de requête Kusto (KQL) est spécialement conçu pour analyser efficacement ces types de données, facilitant ainsi la création d'insights et de modèles.

Microsoft Fabric permet également le streaming de données et l'exécution de requêtes en KQL en utilisant simplement une source et une base de données Kusto. La plateforme offre des fonctionnalités de partage de jeux de requêtes KQL au sein d'un

même locataire et avec Power BI, grâce à un outil de « création rapide », optimisant ainsi la diffusion et l'utilisation des données au sein de l'organisation.

## Qu'est-ce que Synapse Real-Time Analytics ?

Le service Synapse Real-Time Analytics de Microsoft Fabric est une solution de streaming complète pour l'analyse de données haut débit, adaptée aux données de séries chronologiques et supportant le partitionnement et l'indexation automatiques pour tous formats de données.

Real-Time Analytics gère efficacement des volumes de données variant de quelques gigaoctets à plusieurs pétaoctets, provenant de diverses sources et formats, et est utilisable dans des domaines variés tels que l'IoT ou l'analytique des journaux pour des secteurs comme l'industrie ou l'automobile.

Ce service est entièrement géré, optimisé pour les données de séries chronologiques en streaming, et intégré à la suite de fonctionnalités de Fabric, ce qui facilite les flux de travail depuis le chargement jusqu'à la visualisation des données. Il permet :

- L'ingestion de données de toute source et format.
- L'exécution de requêtes analytiques sur des données brutes sans nécessiter de modèles complexes ni de scripts de transformation.
- Une analyse des données performante, à faible latence et à forte actualisation grâce au streaming par défaut.
- Un partitionnement et une indexation automatiques des données importées.
- La gestion et l'interrogation de données structurées, semi-structurées ou libres avec des performances élevées.

Le langage de requête Kusto (KQL) permet d'analyser et d'extraire des insights de données structurées à grande échelle. Il supporte les analyses reproductibles, améliore les processus DevOps, et s'intègre facilement aux flux de développement grâce à la possibilité de lier des fichiers KQL aux dépôts Git et pipelines CI/CD. KQL simplifie également la création et la manipulation de données avec une gamme variée de fonctions et d'opérateurs.

## Comprendre la base de données et les tables KQL

Explorons les composants de base de Synapse Real-Time Analytics (RTA) dans Microsoft Fabric, une plateforme qui intègre divers éléments essentiels pour l'analyse de données en temps réel.

### Composants de base de Real-Time Analytics

#### 1. Base de données KQL (Kusto)

- Elle représente l'entité de niveau supérieur qui héberge :
  - Une collection de tables
  - Des fonctions stockées

- Des vues matérialisées
- Des raccourcis
- Des flux de données

## 2. Jeu de requêtes KQL

- Permet d'exécuter des requêtes, de visualiser et de manipuler les résultats sur les données provenant de la base de données KQL.
- Les requêtes peuvent être enregistrées pour une utilisation ultérieure, exportées et partagées.
- Supporte le langage de requête Kusto, T-SQL, et certaines fonctions T-SQL.

## 3. Fonctionnalité Eventstream

- Intègre les données de streaming provenant de diverses sources, telles que les hubs d'événements ou des applications personnalisées.
- Les données peuvent être envoyées vers diverses destinations comme un lakehouse, une base de données KQL, ou une application personnalisée.

## Objets de base de la base de données KQL

### 1. Table

- Contient des données dans un ensemble structuré de colonnes et de lignes.
- Possède un schéma bien défini (liste ordonnée de paires nom de colonne/type de données).
- Commandes associées :
  - **.create table** pour créer une table.
  - **.show table** pour afficher le schéma de la table.
  - **.ingest** pour ingérer des données dans une table.

### 2. Fonction

- Encapsule une expression de sous-requête pouvant être appelée à partir d'autres requêtes KQL.
- Peut avoir des paramètres facultatifs et un corps contenant l'expression de sous-requête.
- Commandes associées :
  - **.create function** pour créer une fonction stockée.
  - **.show functions** pour afficher les fonctions stockées.

### 3. Vue matérialisée

- Stocke les résultats précalculés d'une requête pour accélérer la récupération.

- Peut avoir des paramètres facultatifs et un corps contenant l'expression de requête.
- Commandes associées :
  - **.create materialized-view** pour créer une vue matérialisée.
  - **.show materialized-views** pour afficher les vues matérialisées.

#### 4. Flux de données

- Représente tous les flux d'événements KQL attachés qui sont connectés à la base de données KQL.

Ces composants sont essentiels pour la gestion efficace de l'analyse des données en temps réel dans Microsoft Fabric, facilitant l'interrogation, la visualisation et le partage des données au sein de l'organisation.

### Écrire des requêtes en KQL

Pour interroger des données dans une base de données KQL (Kusto), on utilise le Kusto Query Language (KQL), compatible avec plusieurs services Azure et Microsoft Fabric. Une requête KQL est constituée de commandes de traitement de données en lecture seule.

#### Instructions de requête KQL :

- Une requête se base sur un nom de table suivi d'opérateurs tels que **take**, **filter**, **transform**, **aggregate**, ou **join**. Les opérateurs relationnels utilisés sont similaires à SQL, mais KQL inclut des extensions pour la correspondance de texte avancée, l'analyse statistique, les projections de séries chronologiques, les données géospatiales, et le machine learning.

#### Création et chargement de tables :

- Les tables sont généralement créées via des outils graphiques dans Microsoft Fabric, mais peuvent aussi être définies directement avec KQL.
- Pour créer une table, utilisez **.create table** avec un schéma spécifié. Par exemple, pour créer une table **sales** avec divers types de données pour les colonnes.
- Les données sont chargées dans les tables via la commande **.ingest into table** en spécifiant le chemin du fichier source.

```
Kusto Copier  
  
.create table sales (  
    SalesOrderNumber: string,  
    SalesOrderLineItem: int,  
    OrderDate: datetime,  
    CustomerName: string,  
    EmailAddress: string,  
    Item: string,  
    Quantity: int,  
    UnitPrice: real,  
    TaxAmount: real)
```

### Exemples de requêtes de récupération :

- Une requête de base pourrait simplement interroger toutes les données d'une table avec des filtres et des transformations appliqués pour affiner les résultats.
- L'instruction **take** permet de retourner un nombre spécifié d'enregistrements, utile pour l'échantillonnage des données.
- La clause **where** filtre les données selon des conditions spécifiques, comme filtrer les enregistrements où une colonne contient une chaîne de caractères précise.
- Pour les données de séries chronologiques, on peut utiliser **where** avec des fonctions comme **now()** pour filtrer les données sur des intervalles de temps.
- L'opérateur **sort** permet de trier les résultats selon un ou plusieurs attributs.
- L'opérateur **summarize** regroupe les données par colonne et permet d'effectuer des agrégations, comme calculer la somme des quantités vendues par article.

Ces fonctions rendent KQL particulièrement puissant pour interroger et analyser des volumes de données significatifs de manière flexible et efficace.

### Exercice : Explorer Synapse Real-Time Analytics dans Fabric

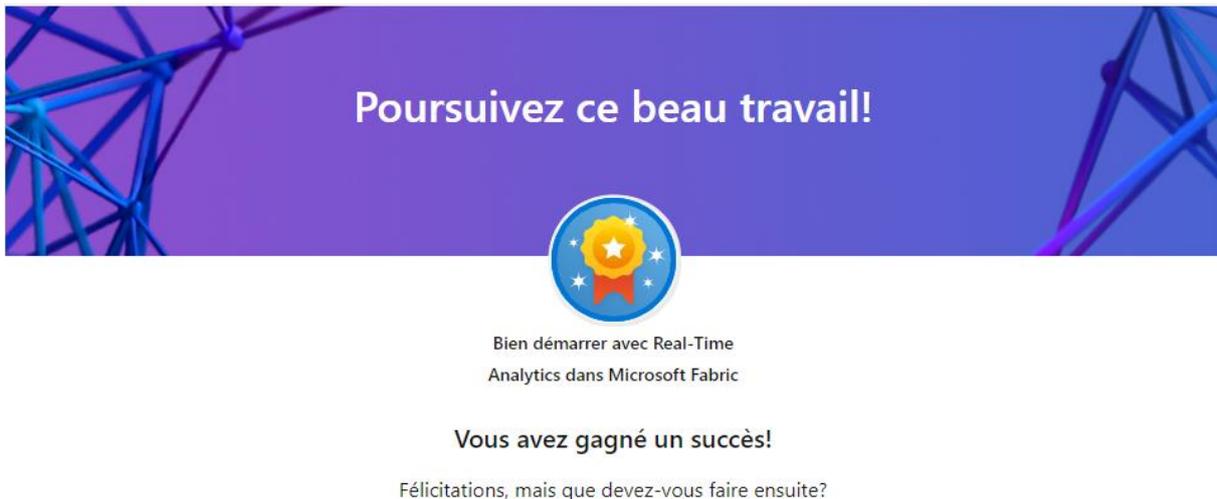
- Créer un espace de travail
- Télécharger le fichier pour la base de données KQL
- Créer une base de données KQL
- Utilisez KQL pour interroger la table des ventes
- Créer un rapport Power BI à partir d'un ensemble de requêtes KQL
- Nettoyer les ressources

### Résumé

Real-Time Analytics est une plateforme d'analytique de Big Data complètement managée et optimisée pour les données de séries chronologiques en streaming. Real-Time Analytics contient un moteur et un langage de requête dédiés qui permettent d'effectuer des recherches dans des données structurées, semi-structurées et non structurées avec un haut niveau de performance. Real-Time Analytics est entièrement

intégré à la suite de produits Fabric, à la fois pour le chargement de données et la visualisation avancée.

La base de données KQL et le langage de requête Kusto (KQL) sont des fonctionnalités puissantes qui vous aident à explorer vos données de diffusion en continu dans Fabric. Kusto est similaire à SQL en termes de structure puisqu'il utilise des bases de données, des tables et des colonnes. Il propose également des fonctionnalités étendues qui permettent d'équilibrer les données volumineuses afin d'améliorer les performances des requêtes.



## Utiliser des flux d'événements en temps réel dans Microsoft Fabric

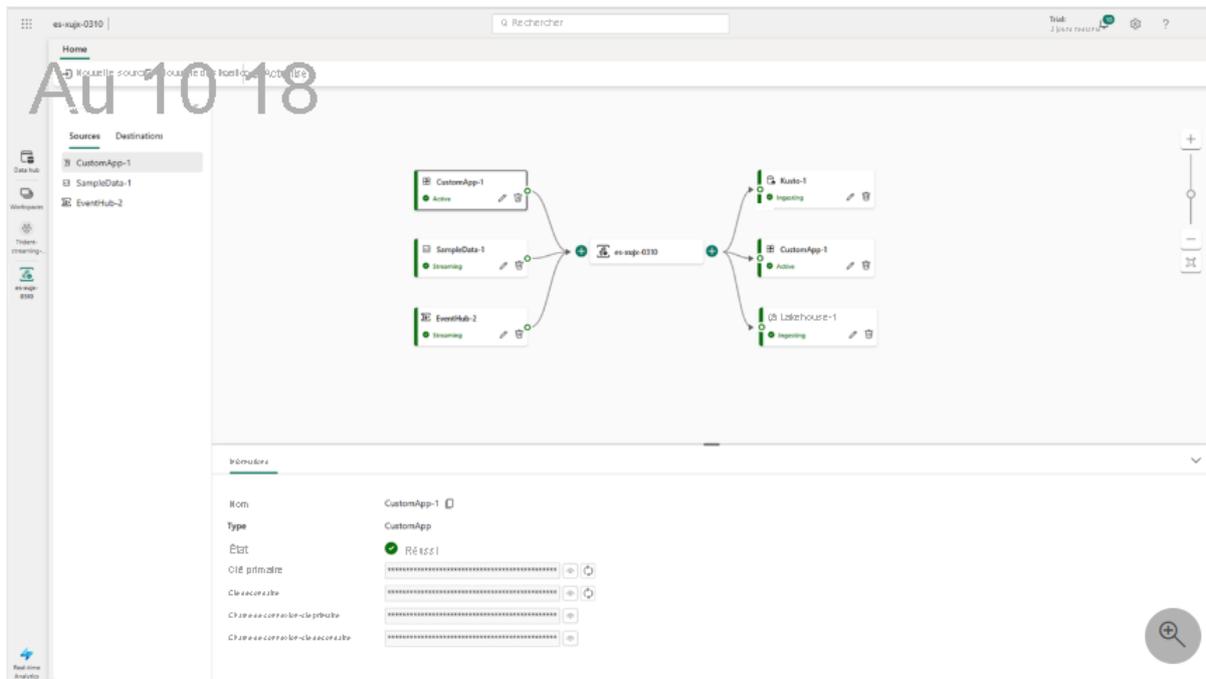
### Introduction

Quand une entreprise opte pour Microsoft Fabric pour l'analyse en temps réel (RTA), elle commence par intégrer les données dans son lakehouse Fabric. Une des fonctions clés de Microsoft Fabric est Eventstreams, qui facilite la gestion des événements en temps réel sans nécessiter de programmation. Cette fonctionnalité permet de configurer les sources, les destinations et les processeurs d'événements dans le flux d'événements, en offrant diverses options pour le traitement des événements telles que la collecte, l'agrégation, le filtrage, le regroupement, la transformation, le développement et la jonction.

L'interface principale pour configurer ces flux est l'éditeur Eventstream. Il permet d'établir les sources et les destinations, de visualiser les données en temps réel, et de capturer, transformer et acheminer les données. À travers ce module, l'utilisateur apprendra à utiliser les flux d'événements Fabric pour gérer et router les données événementielles vers les destinations requises par ses charges de travail.

## Principaux composants d'Eventstream

Microsoft Fabric Eventstream est une solution no-code permettant de capturer et d'acheminer des événements en temps réel vers diverses destinations. Idéale pour les applications telles que l'IoT, le streaming et les tableaux de bord, cette fonctionnalité facilite la connexion à diverses sources et destinations, et offre une visualisation en temps réel des données événementielles.



## Fonctionnement d'Eventstream

Eventstream opère comme un pipeline qui transfère les données de diverses sources vers différentes destinations. Ce processus est simplifié grâce à un éditeur principal qui permet de configurer le flux d'événements via une interface de glisser-déposer, sans nécessité de codage. Les utilisateurs peuvent observer les données transitant dans le pipeline en temps réel, tandis que Microsoft Fabric s'occupe de l'évolutivité, de la fiabilité et de la sécurité du flux.

## Composants d'Eventstream

- **Flux d'événements** : Configurés via le portail pour créer et gérer les flux de données événementielles.
- **Source** : Origine des données, pouvant inclure des options comme Azure Event Hubs ou des applications personnalisées. Les formats de données et les groupes de consommateurs sont configurables.
- **Destination** : Emplacement final des données, pouvant être une base de données KQL ou un lakehouse, avec possibilité de définir le nom de la table et le mappage de schéma.

- **Éditeur principal** : Interface principale pour ajouter et gérer les sources et destinations, offrant une visualisation des données événementielles et des insights.

The screenshot shows the main editor interface. On the left, there's a 'Données' sidebar with 'Sources' and 'Destinations' tabs, and a 'Stock' source selected. The central area displays a data flow diagram with three components: 'Stock' (Diffusion en continu), 'StockDS', and 'ReflexQueryDB' (Ingestion). On the right, a 'Modifier ReflexQueryDB' panel is visible, showing fields for 'Nom de destination', 'Espace de travail', and 'Base de données KQL'. Below the diagram, there's an 'Aperçu des données' section with a table of data points.

| heure                    | symbole | secteur            | securityType | bidPrice | bidSize |
|--------------------------|---------|--------------------|--------------|----------|---------|
| 2023-08-23T14:01:38.356Z | HOJI    | mediaentertainment | commonstock  | 1330.28  | 66      |
| 2023-08-23T14:01:38.356Z | NSFT    | softwareservices   | commonstock  | 290.63   | 14      |
| 2023-08-23T14:01:38.356Z | BMZM    | vente              | commonstock  | 2336.84  | 50      |
| 2023-08-23T14:01:38.356Z | HOJI    | mediaentertainment | commonstock  | 1350.28  | 106     |
| 2023-08-23T14:01:38.356Z | NSFT    | softwareservices   | commonstock  | 320.63   | 120     |
| 2023-08-23T14:01:38.356Z | BMZM    | vente              | commonstock  | 2306.84  | 107     |

## Éditeur principal

L'éditeur principal est l'espace de travail où les utilisateurs peuvent développer et concevoir des flux d'événements:

The screenshot shows the main editor interface with several numbered callouts highlighting key features:

- 1. Ruban** : Menu offering options for sources and destinations.
- 2. Volet de navigation des données** : Allows navigating among sources and destinations.
- 3. Vue 'Canevas et diagramme'** : Graphical representation of the flow topology, with icons representing each source or destination.
- 4. Modifier Kusto** : Panel for configuring the destination, workspace, and database.
- 5. Aperçu des données** : Data preview section showing a table of data points.

| entityTime                  | messageId | deviceId | température | EventProcessedUtcTime        | PartitionId |
|-----------------------------|-----------|----------|-------------|------------------------------|-------------|
| 2023-03-15T13:03:58.0000000 | 390       | 653      | 35          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 430       | 273      | 19          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 154       | 733      | 17          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 462       | 226      | 17          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 888       | 602      | 0           | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 940       | 144      | 14          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 547       | 466      | 20          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 826       | 644      | 13          | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 600       | 931      | 9           | 2023-03-15T13:07:59.8185848Z | 1           |
| 2023-03-15T13:03:58.0000000 | 506       | 46       | 20          | 2023-03-15T13:07:59.8185848Z | 1           |

1. **Ruban** : Menu offering options for the sources and destinations.
2. **Volet de navigation des données** : Permet de naviguer parmi les sources et destinations.
3. **Vue 'Canevas et diagramme'** : Représentation graphique de la topologie des flux, avec des vignettes représentant chaque source ou destination.

4. **Volet 'Configuration et modification'** : Permet de configurer ou de modifier les sources et destinations.
5. **Volet inférieur** : Affiche des informations supplémentaires telles que l'aperçu des données, des insights et des journaux d'exécution.

### Avantages d'Eventstream

Eventstream facilite la capture, la transformation, et le routage des données en temps réel vers diverses destinations, tout en permettant une intégration avec les services Azure et une évolutivité automatisée. Cette infrastructure simplifie les opérations ETL en temps réel et offre une flexibilité accrue pour exploiter les données événementielles.

## Configurer Eventstream

L'ajout et la gestion des sources d'événement dans Microsoft Fabric Eventstream permettent une intégration transparente des données d'événements à partir de diverses sources. Voici un résumé des étapes clés et des configurations possibles :

### Ajout de Sources d'Événement

1. **Sélection de la Source** : Les sources peuvent inclure Azure Event Hubs, des exemples de données, ou des applications personnalisées. Vous sélectionnez le type de source dans l'éditeur principal d'Eventstream après avoir initié un flux d'événements dans votre espace de travail.
2. **Configuration de la Source** : Après avoir choisi le type de source, vous nommez la source, sélectionnez la connexion cloud (ou créez-en une nouvelle), et définissez le format des données (JSON, Avro, CSV) ainsi que le groupe de consommateurs.

### Utilisation de Sources Spécifiques

- **Azure Event Hubs** : Créez un espace de noms et un hub d'événements, configurez une stratégie d'accès partagé et une chaîne de connexion, et choisissez un format de données et un groupe de consommateurs.
- **Exemples de Données** : Sélectionnez "Exemples de données" comme source, choisissez les jeux de données désirés (IoT, Détail, Finances, etc.), et ajustez la fréquence et la durée de génération des données.
- **Application Personnalisée** : Ajoutez une source d'application personnalisée en récupérant et utilisant la chaîne de connexion pour envoyer des événements à votre flux. Développez une application qui utilise cette chaîne pour envoyer des données via des interfaces comme les clients Kafka ou les hubs d'événements.

### Gestion et Visualisation des Données

- Configurez et ajoutez des destinations au flux d'événements pour diriger les données vers des récepteurs appropriés (lakehouse, base de données KQL, Power BI, etc.).
- Utilisez l'éditeur principal pour visualiser les données en temps réel, vérifier les insights de données, et assurer que les données sont correctement routées vers les destinations configurées.

Ce processus permet une intégration flexible des données d'événements dans des applications en temps réel, facilitant la gestion et l'analyse des données sans nécessiter de compétences de programmation avancées

## Routage des données de flux d'événements

Pour ajouter et gérer des destinations d'événement dans Microsoft Fabric Eventstream, suivez ces étapes clés :

### Ajout de Destinations d'Événement

1. **Sélection de la Destination:** Utilisez le ruban pour choisir « Nouvelle destination » et sélectionnez le type de destination souhaitée, comme une base de données KQL, un lakehouse, une application personnalisée ou le système Reflex.
2. **Configuration de la Destination:** Entrez un nom pour la destination, choisissez un espace de travail et configurez la destination spécifique. Confirmez en cliquant sur « Ajouter et configurer ».

### Configuration Détaillée

- **Destination:** Choisissez d'utiliser une table existante ou de créer une nouvelle pour recevoir les données.
- **Source:** Configurez la source des données pour établir une connexion d'ingestion.
- **Schéma:** Sélectionnez le format de données et le type de compression, et vérifiez que le schéma des données correspond à vos attentes pour prévenir des erreurs d'ingestion.
- **Résumé:** Examinez le résumé de l'ingestion pour vous assurer que tout est configuré correctement avant de finaliser.

Une fois terminé, une vignette représentant la destination configurée apparaît dans l'éditeur principal, indiquant que les données peuvent maintenant être routées vers cette destination.

### Fonctions de Traitement et Fenêtrage

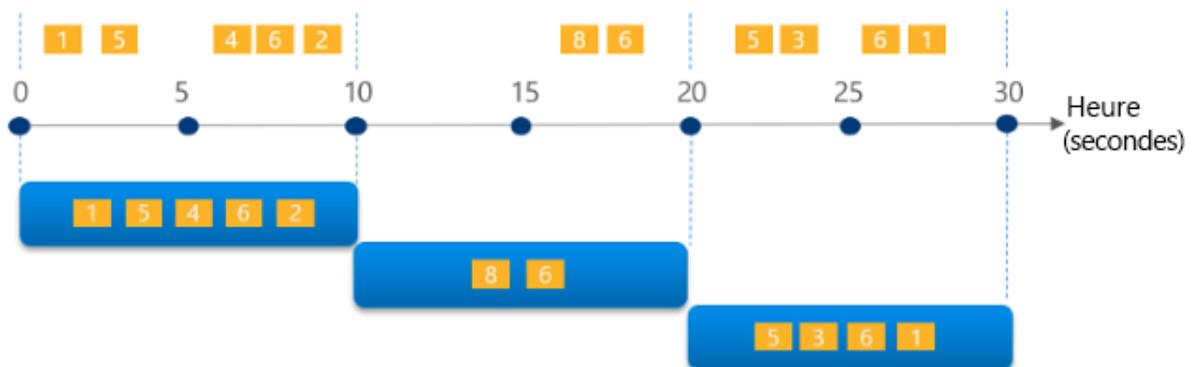
Les destinations peuvent intégrer diverses opérations de traitement pour affiner les données avant leur stockage :

- **Agrégation:** Regroupe les événements pour appliquer des fonctions comme la somme ou la moyenne.

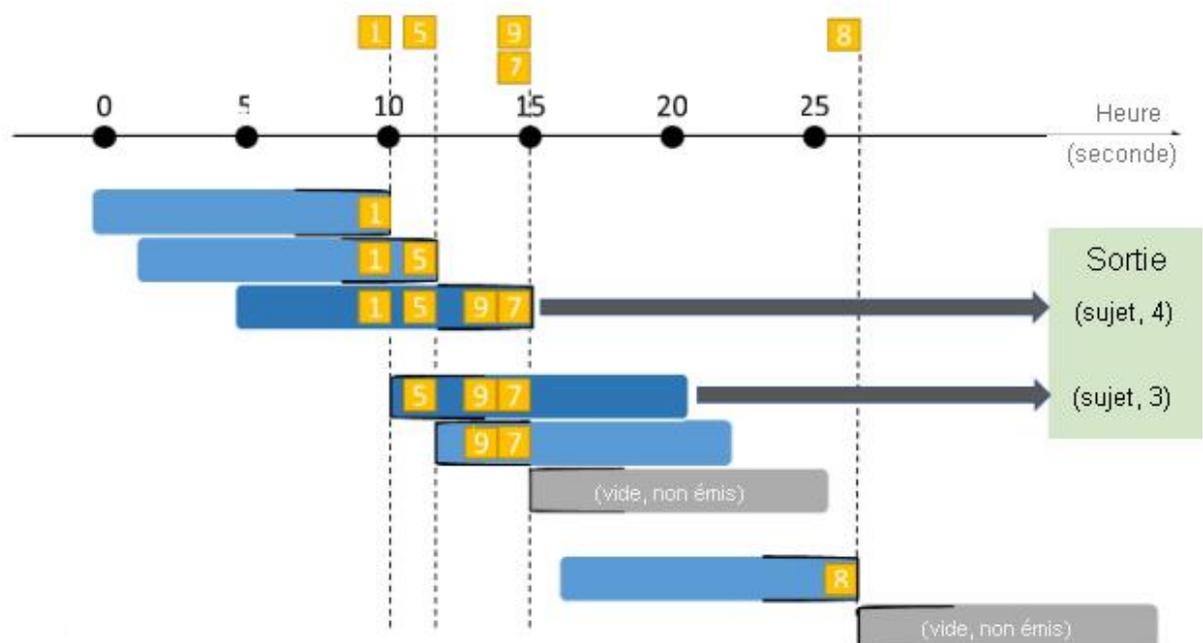
- **Développer:** Crée une ligne pour chaque valeur d'un tableau.
- **Filtrer:** Élimine les événements qui ne correspondent pas aux critères spécifiés.
- **Regrouper par:** Calcul des agrégations sur les événements d'une période donnée.
- **Unir:** Combine les événements de plusieurs sources ayant des champs en commun.

Les fonctions de fenêtrage dans Eventstream permettent d'analyser les données en fonction de périodes spécifiques :

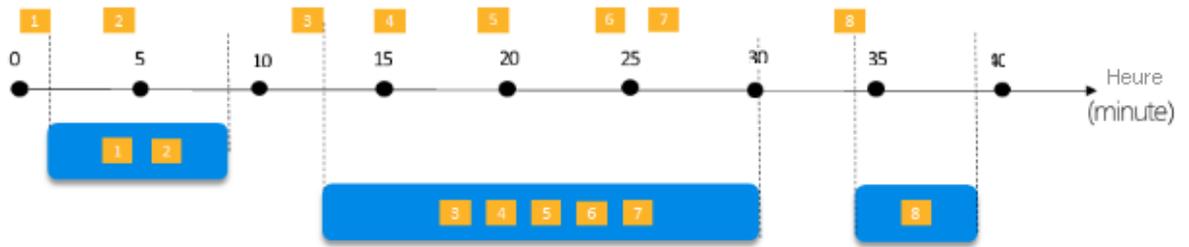
- **Fenêtres bascules:** Divisent les événements en intervalles fixes selon leur heure d'arrivée.



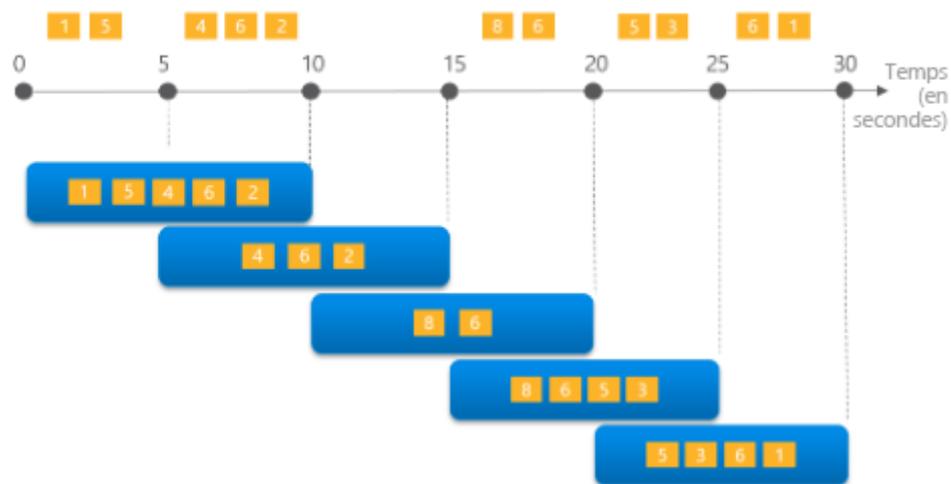
- **Fenêtres glissantes:** Gèrent les événements dans des intervalles qui se chevauchent.



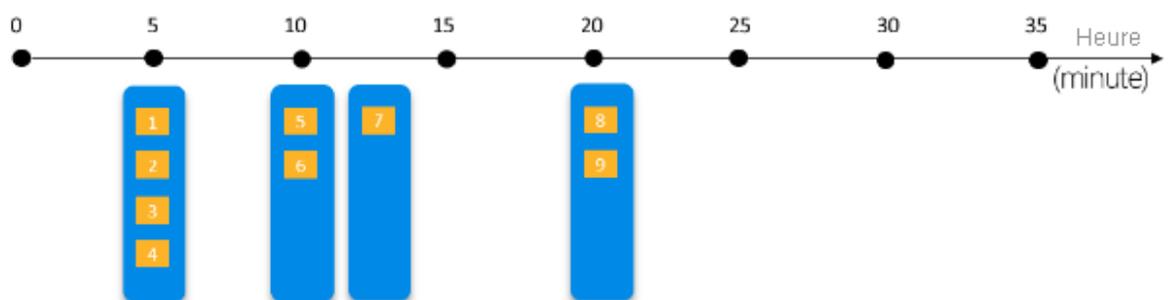
- **Fenêtres de session:** Groupent les événements en fonction de périodes d'activité et d'inactivité.



- **Fenêtres récurrentes:** Fenêtres programmées qui se répètent à intervalles réguliers.



- **Fenêtres d'instantés:** Regroupent les événements ayant le même horodatage.



Ces outils permettent de personnaliser l'analyse et le traitement des données en temps réel, optimisant ainsi leur utilité pour des applications spécifiques comme le monitoring de capteurs ou l'analyse comportementale en ligne.

## Exercice

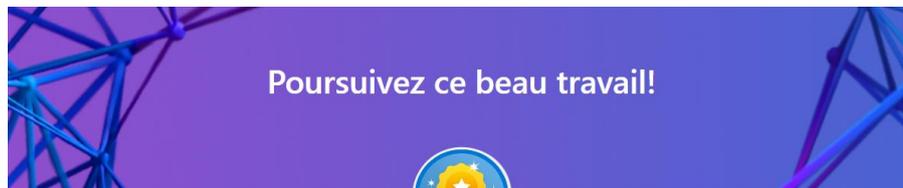
- Créer un espace de travail
- Scénario
- Créer une base de données KQL
- Créer un flux d'événements
- Établir une source et une destination de flux d'événements
- Requêtes KQL
- Exemples de requêtes SQL
- Nettoyer les ressources

## Résumé

Ce module de formation Microsoft aborde l'utilisation d'Eventstream dans Microsoft Fabric. Il décrit les notions de base des flux d'événements, leur création et leur utilisation, mais aussi la procédure de gestion des erreurs. Le module comprend également un labo pratique où les apprenants peuvent s'entraîner à créer et à utiliser des flux d'événements dans un exemple d'application.

Points essentiels :

- Eventstream est un service permettant de publier et de consommer des événements dans un système distribué.
- Azure Service Fabric fournit une prise en charge intégrée d'Eventstreams.
- Les apprenants découvrent comment créer et utiliser des composants Eventstream dans des applications Microsoft Fabric.
- Le labo pratique permet aux apprenants de s'entraîner à la création et l'utilisation d'Eventstream avec une base de données KQL.



Utiliser des flux d'événements  
en temps réel dans Microsoft  
Fabric

**Vous avez gagné un succès!**

Félicitations, mais que devez-vous faire ensuite?

# Interroger des données dans une base de données KQL dans Microsoft Fabric

## Présentation

Ce didacticiel sur Microsoft Fabric introduit les principes de base du KQL (Langage de requête Kusto), un langage de requête puissant utilisé dans Azure Data Explorer, Azure Monitor, et Microsoft Fabric pour l'analyse de grands volumes de données. Le tutoriel aborde l'écriture de requêtes KQL et T-SQL sur une base de données KQL, incluant la sélection, le filtrage et l'agrégation des données. À la fin, les utilisateurs apprendront à utiliser efficacement les requêtes KQL pour analyser leurs données dans Microsoft Fabric.

La formation se poursuit en mettant l'accent sur les meilleures pratiques pour optimiser l'utilisation de KQL, telles que le choix des types de données adéquats, l'application de l'indexation, et l'évitement des fonctions superflues pour améliorer la performance et l'efficacité des requêtes.

Enfin, la session explore des fonctionnalités avancées pour augmenter la productivité, y compris des raccourcis, des outils d'automatisation et des options de collaboration. Ces outils permettent une exécution plus efficace des tâches, automatisent les opérations répétitives et facilitent la collaboration via l'édition en temps réel et le partage de fichiers, améliorant ainsi la productivité globale.

## Prise en main des requêtes KQL

**Configuration Initiale** : Pour débiter avec KQL dans Microsoft Fabric, il est essentiel de posséder un compte Azure et de configurer Azure Data Explorer.

**Création d'un Ensemble de Requêtes KQL** : Un ensemble de requêtes KQL vous permet d'exécuter, modifier, et afficher les résultats des requêtes à partir d'une base de données KQL. Cela est particulièrement utile pour l'analyse en temps réel dans Microsoft Fabric.

### Fonctionnalités de l'Ensemble de Requêtes KQL :

- Chaque onglet de l'ensemble peut être lié à une base de données KQL différente.
- Permet de sauvegarder et partager des requêtes.

### Limitations :

- L'utilisation de SQL dans un ensemble de requêtes KQL empêche la création de rapports Power BI directement à partir de l'ensemble.

### Requêtes de Base :

- Pour récupérer tous les enregistrements d'une table : **TableName**

- Pour récupérer des colonnes spécifiques : **TableName | project Column1, Column2**
- Pour filtrer les enregistrements : **TableName | where Column1 == "value"**
- Pour trier les enregistrements : **TableName | sort by Column1 asc/desc**

#### **Avantages de KQL par rapport à T-SQL :**

- **Simplicité** : KQL est plus facile à apprendre et à utiliser.
- **Performances** : Optimisé pour gérer de grandes quantités de données.
- **Flexibilité** : Permet des requêtes complexes plus facilement.
- **Intégration** : Bien intégré avec d'autres produits Microsoft comme Azure Monitor et Azure Sentinel.

#### **Avantages de T-SQL par rapport à KQL :**

- **Compatibilité** : Large utilisation et compatibilité avec divers systèmes de base de données.
- **Fonctionnalités** : Offre plus de fonctions et fonctionnalités.
- **Programmation procédurale** : Supporte des scripts et procédures stockées complexes.

**Transition de T-SQL à KQL** : Si vous êtes familiarisé avec SQL, vous pouvez apprendre KQL en utilisant la syntaxe **-- explain** pour traduire des requêtes SQL en KQL, ce qui peut aider à comprendre et utiliser KQL efficacement.

Par exemple :

```

Transact-SQL Copier
--
Explain
SELECT CASE
    WHEN pickup_boroughname IS NULL OR pickup_boroughname = '' THEN 'Unidentified'
    ELSE pickup_boroughname
END AS Borough,
SUM(trip_distance) AS [Total Trip Distance]
FROM Trips
GROUP BY CASE
    WHEN pickup_boroughname IS NULL OR pickup_boroughname = '' THEN 'Unidentified'
    ELSE pickup_boroughname
END;

```

génère la syntaxe KQL suivante. Il s'agit d'une fonctionnalité puissante dans le moteur KQL

```

Kusto Copier
Trips
| summarize ['Total Trip Distance']=sum(trip_distance) by groupByExpression1=iff((isnull(pickup_boroughname)
| project Borough=groupByExpression1, ['Total Trip Distance']

```

## Ressources Complémentaires

Pour une pratique plus approfondie, Microsoft Learn propose des exercices pratiques pour s'habituer aux requêtes KQL dans le cadre de l'analyse en temps réel avec Microsoft Fabric.

## Bonnes pratiques relatives à KQL

### Bonnes Pratiques pour l'Utilisation de KQL dans Microsoft Fabric

**Optimisation des Performances de Requêtes :** La performance d'une requête KQL est influencée par la quantité de données traitées. Réduire cette quantité peut accélérer la requête et réduire la consommation de ressources.

#### Exemples de Requêtes Optimisées :

##### 1. Filtrage par Dates et Projection de Colonnes Nécessaires :

kusto

```
Trips | where ((pickup_datetime >= todatetime("2014-10-01")) && (pickup_datetime < todatetime("2014-11-01"))) | project passenger_count, trip_distance, fare_amount
```

##### 2. Application de Prédicats sur Types de Paiement et Dates :

kusto

```
Trips | where (payment_type == 'CRD' && (pickup_datetime >= todatetime("2014-10-01")) && (pickup_datetime < todatetime("2014-11-01"))) | project passenger_count, trip_distance, fare_amount
```

### 3. Filtrage par Comptage de Passagers et Dates :

kusto

```
Trips | where (passenger_count > 2 && (pickup_datetime >= todatetime("2014-10-01")) && (pickup_datetime < todatetime("2014-11-01"))) | project payment_type, passenger_count, trip_distance, fare_amount
```

**Référencement des Entités :** Évitez l'utilisation de noms qualifiés redondants, car cela simplifie la lecture des requêtes et peut accélérer leur exécution.

**Sélection des Types de Données Appropriés :** Utiliser le bon type de données pour chaque champ améliore les performances des requêtes.

#### Utilisation Stratégique des Opérateurs KQL :

- **Project :** Limitez les colonnes dans vos résultats pour minimiser la quantité de données traitées.
- **Where :** Filtrez les données avant de les traiter pour réduire la charge.
- **Summarize :** Utilisez pour les agrégations, réduisant ainsi la quantité de données traitées.

**Évitez les Caractères de Remplacement :** L'utilisation de caractères génériques peut ralentir les performances des requêtes. Exemple « \* » et « ? ». Si possible, évitez-les.

**Optimisation Générale :** Assurez-vous de filtrer les données avec **where** avant d'effectuer des opérations qui pourraient être coûteuses en termes de performances. Évitez l'utilisation de l'opérateur \* pour sélectionner toutes les colonnes si toutes ne sont pas nécessaires.

**Ressources Supplémentaires :** Pour approfondir vos connaissances sur les requêtes KQL et leur optimisation, vous pouvez consulter des ressources éducatives sur Microsoft Learn, spécialement celles dédiées à Microsoft Fabric.

## Fonctionnalités avancées

### Fonctionnalités Avancées des Requêtes KQL dans Microsoft Fabric

**Vues Matérialisées :** Les vues matérialisées permettent de synthétiser et d'accéder rapidement à des données depuis une table source ou une autre vue. Il y a deux types principaux :

- **Vue matérialisée vide :** Utilise **.set-or-append** pour ajouter des enregistrements après la création de la vue. Disponible immédiatement pour la requête.
- **Vue matérialisée basée sur des enregistrements existants :** Sa création peut être longue en fonction du volume de données de la table source.

## Exemple de Création de Vue Matérialisée :

kusto

```
.create materialized-view MyMV on table Trips { Trips | summarize count() by vendor_id }
```



**Fonctions Personnalisées :** Permet de créer des fonctions personnalisées dans KQL pour des tâches d'analyse de données spécifiques.

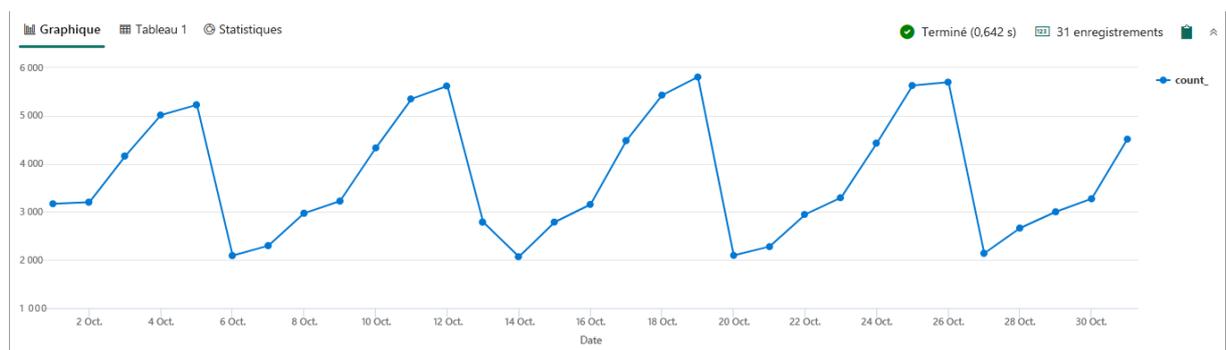
**Visualisation des Données :** KQL supporte diverses techniques de visualisation pour afficher les données :

- **Render timechart :** Utilisé pour visualiser des données chronologiques.
- **Render columnchart :** Utilisé pour visualiser des données en diagramme à colonnes.
- **Render anomalychart :** Utilisé pour la détection d'anomalies et la prédiction de tendances.

## Exemple de Visualisation Temporelle :

kusto

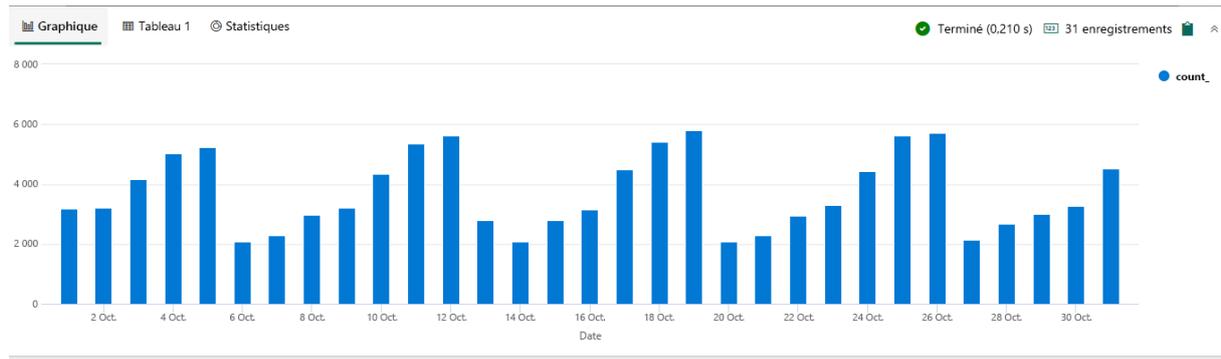
```
Trips | where pickup_datetime >= datetime("2014-10-01") and pickup_datetime < datetime("2014-11-01") | summarize count() by bin(pickup_datetime, 1d) | render timechart with (ytitle="Number of Trips", xtitle="Date")
```



## Exemple de Visualisation en Diagramme à Colonnes :

kusto

```
Trips | where pickup_datetime >= datetime("2014-10-01") and pickup_datetime < datetime("2014-11-01") | summarize count() by bin(pickup_datetime, 1d) | render columnchart with (ytitle="Number of Trips", xtitle="Date")
```



**Ingestion des Données :** KQL facilite l'ingestion de données depuis des sources comme Azure Event Hubs, Azure IoT Hub, et le Stockage Azure.

**Exportation de Données :** Les résultats des requêtes KQL peuvent être exportés vers diverses destinations comme Azure Blob Storage, Azure Data Lake Storage, et Azure Event Hubs.

**Ressources Supplémentaires :** Pour approfondir vos connaissances sur KQL et ses applications dans Microsoft Fabric, des ressources telles que la terminologie Microsoft Fabric et des guides pour interroger des données dans un ensemble de requêtes KQL sont disponibles sur Microsoft Learn.

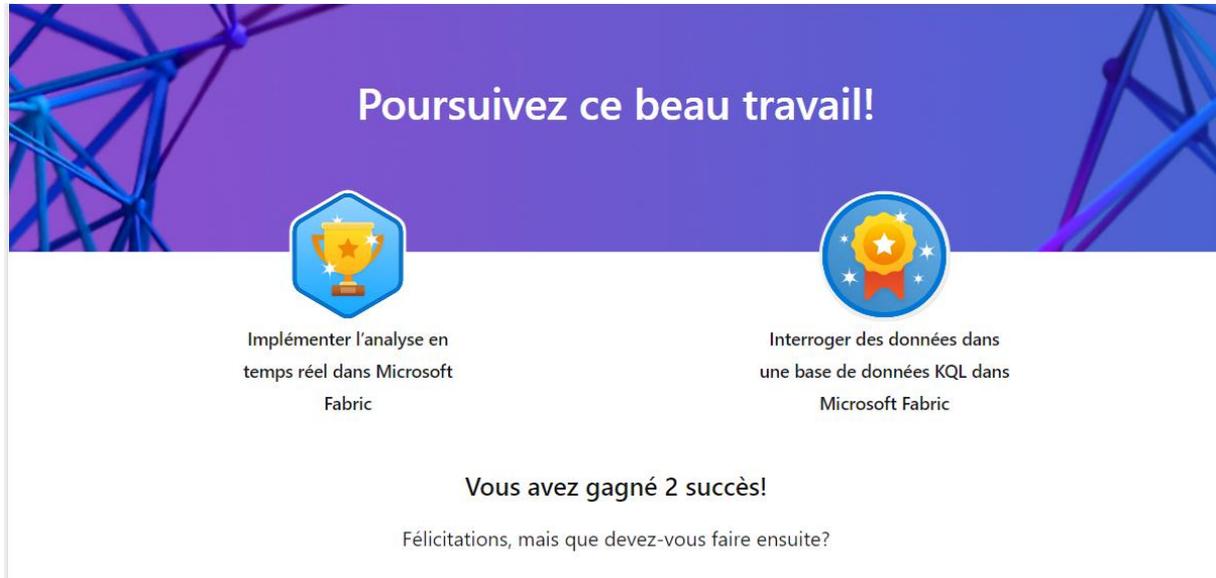
## Exercice : Interroger une base de données KQL dans Microsoft Fabric

- [Créer un espace de travail](#)
- [Créer une base de données KQL](#)
- [Introduction au langage de requête Kusto \(KQL\) et à sa syntaxe](#)
- [SÉLECTIONNER les données de notre exemple d'ensemble de données à l'aide de KQL](#)
- [Données GROUP BY de notre exemple d'ensemble de données utilisant KQL](#)
- [Données ORDER BY de notre exemple d'ensemble de données utilisant KQL](#)
- [Clause WHERE pour filtrer les données dans notre exemple de requête KQL](#)
- [Utiliser T-SQL pour interroger des informations récapitulatives](#)
- [SÉLECTIONNER les données de notre exemple d'ensemble de données à l'aide de T-SQL](#)
- [Données GROUP BY de notre exemple d'ensemble de données à l'aide de T-SQL](#)
- [Données ORDER BY de notre exemple d'ensemble de données utilisant T-SQL](#)
- [Nettoyer les ressources](#)

## Résumé

Dans ce module, vous avez exploré comment initier des requêtes KQL dans Microsoft Fabric. Il est nécessaire de configurer un compte Azure et Azure Data Explorer pour démarrer. Les requêtes KQL sont structurées autour d'un modèle de base qui inclut la sélection de tables, l'application de conditions, et la projection de colonnes spécifiques. Vous avez également appris à utiliser diverses fonctions et opérateurs pour réaliser des manipulations arithmétiques, de comparaison, logiques, et sur des chaînes de

caractères. Les techniques avancées comme les agrégations, les jointures et les sous-requêtes ont été couvertes. Les exercices pratiques de Microsoft Learn offrent une expérience concrète avec KQL. Enfin, KQL offre des avantages significatifs sur T-SQL en termes de simplicité, performance, flexibilité, et intégration avec d'autres produits Microsoft.



## Bien démarrer avec la science des données dans Microsoft Fabric

### Présentation

La science des données permet d'optimiser les opérations et la prise de décisions dans un supermarché, par exemple en ajustant les stocks pour éviter le gaspillage ou en ciblant les clients avec des offres personnalisées. Elle combine mathématiques, statistiques et ingénierie informatique pour extraire des insights des données, développer des modèles d'intelligence artificielle et appliquer le machine learning avec des outils comme scikit-learn.

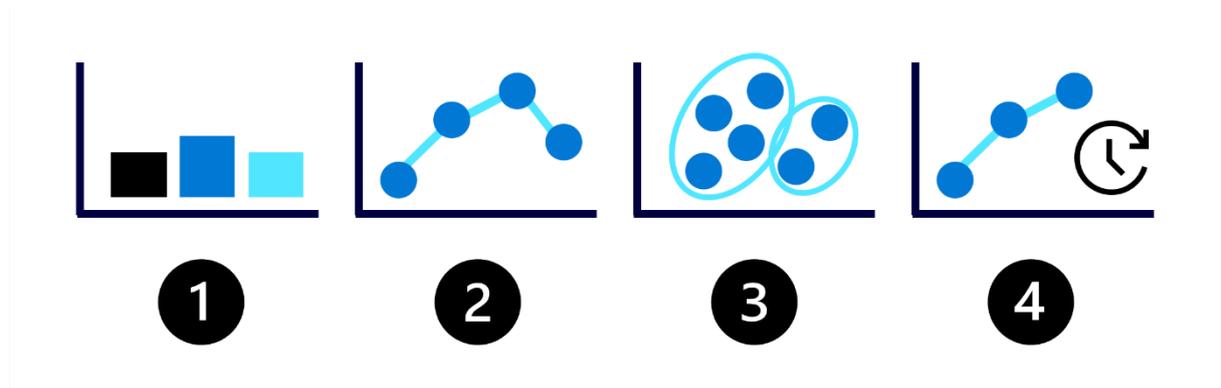
Microsoft Fabric propose une plateforme pour gérer intégralement les projets de science des données, facilitant chaque étape du processus, de la collecte à l'analyse des données. Ce module explore ces applications pratiques au sein d'un projet typique.

### Comprendre le processus de science des données

La science des données implique la visualisation et l'analyse de jeux de données complexes pour détecter des modèles et générer des insights. Le Machine Learning (ML) joue un rôle clé dans ce processus, permettant de prédire des résultats tels que les ventes futures en entraînant des modèles sur des données historiques.

Les quatre types courants de modèles ML sont :

1. **Classification** : Prédit une catégorie, comme identifier les clients à risque.
2. **Régression** : Prédit une valeur numérique, comme le prix d'un produit.
3. **Clustering** : Regroupe des données similaires pour identifier des segments distincts.
4. **Prévision** : Prédit des valeurs futures basées sur des séries chronologiques, telles que les ventes mensuelles.



Le processus de science des données comprend plusieurs étapes :

1. **Définir le problème** : Clarifier les objectifs avec les utilisateurs et déterminer les succès du modèle.
2. **Obtenir les données** : Accéder à des sources pertinentes, souvent stockées dans un lakehouse.
3. **Préparer les données** : Explorer, nettoyer et transformer les données via des outils comme Pandas et Numpy.
4. **Entraîner le modèle** : Sélectionner des algorithmes et ajuster les hyperparamètres, souvent avec l'aide de bibliothèques comme Scikit-Learn, PyTorch, ou SynapseML.
5. **Générer des insights** : Utiliser le modèle pour produire les prédictions nécessaires.



Les scientifiques des données passent beaucoup de temps à préparer les données et à ajuster les modèles. Le suivi des expériences via des outils comme MLflow dans Microsoft Fabric aide à gérer et optimiser les modèles développés.

## Explorer et traiter des données avec Microsoft Fabric

Microsoft Fabric offre des outils robustes pour ingérer, explorer, et transformer les données, essentiels pour entraîner des modèles de Machine Learning. Les performances des modèles s'améliorent avec des jeux de données plus grands et de meilleure qualité.

**Ingestion de données:** Microsoft Fabric permet l'ingestion de données depuis diverses sources, y compris locales et cloud, stockant ces données dans un lakehouse centralisé pour un accès et une gestion faciles.

**Exploration et transformation de données:** Utilisant des notebooks optimisés par Apache Spark, Microsoft Fabric facilite l'exécution et la gestion du code. Les utilisateurs peuvent choisir le langage de programmation de leur choix, comme PySpark ou SparkR, pour travailler sur les données. Les notebooks permettent de visualiser, d'explorer, et de transformer les données directement.

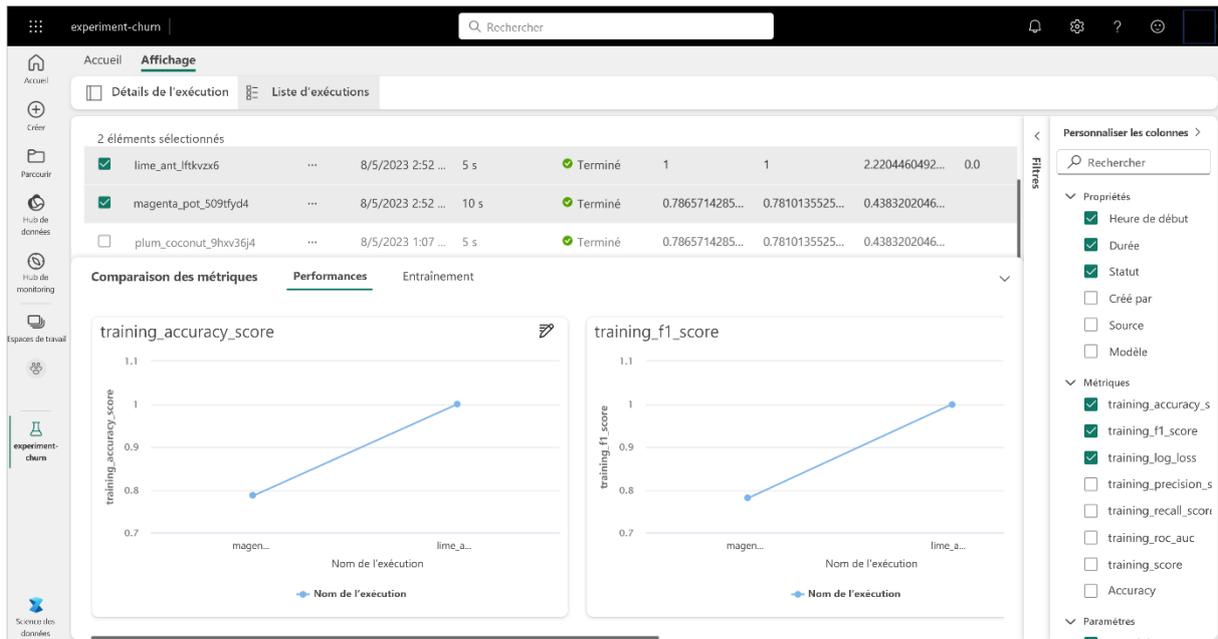
**Data Wrangler:** Cet outil intégré à Microsoft Fabric aide à nettoyer et transformer les données rapidement. Il fournit une vue d'ensemble des données avec des statistiques récapitulatives, permettant de détecter et de résoudre les problèmes comme les valeurs manquantes. Les transformations choisies peuvent être exportées en code pour être appliquées aux données, simplifiant le processus de préparation des données.

## Entraîner et évaluer des modèles avec Microsoft Fabric

Microsoft Fabric intègre MLflow pour suivre et journaliser les processus d'entraînement des modèles de Machine Learning, rendant les étapes reproductibles et permettant d'évaluer les meilleures méthodes pour entraîner le modèle final.

**Création et suivi des expériences:** Dans Microsoft Fabric, chaque session d'entraînement d'un modèle est appelée une "expérience", qui peut inclure plusieurs "exécutions". Ces exécutions représentent différents essais ou configurations lors de l'entraînement d'un modèle. Les paramètres, métriques, et artefacts de chaque exécution sont suivis pour faciliter la comparaison et l'optimisation des modèles.

**Utilisation et gestion des modèles:** Une fois un modèle entraîné, il peut être utilisé pour le scoring, c'est-à-dire pour faire des prédictions sur de nouvelles données. Les modèles et leurs métadonnées sont enregistrés comme artefacts dans Microsoft Fabric, permettant une gestion facile et le suivi des versions.



**Prédiction et intégration:** La fonction PREDICT de Microsoft Fabric permet d'intégrer facilement les modèles à des applications pour générer des prédictions. Par exemple, un modèle prédictif peut être utilisé pour estimer les ventes hebdomadaires à partir des données historiques, et les résultats peuvent être visualisés dans un rapport Power BI.

En résumé, Microsoft Fabric offre des outils avancés pour l'entraînement, le suivi, la gestion et l'utilisation des modèles de Machine Learning, facilitant ainsi la création de solutions d'intelligence artificielle performantes et adaptées aux besoins spécifiques des entreprises.

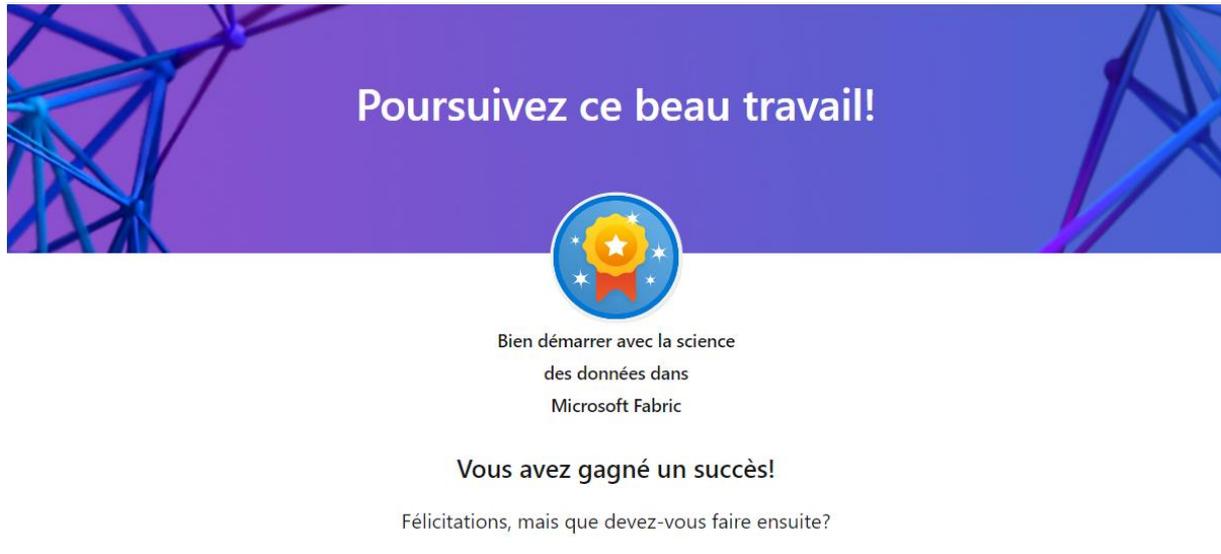
## Exercice

- Créer un espace de travail
- Créer un cahier
- Obtenez les données
- Préparer les données
- Former des modèles d'apprentissage automatique
- Explorez vos expériences
- Enregistrez le modèle
- Enregistrez le bloc-notes et terminez la session Spark
- Nettoyer les ressources

## Résumé

Microsoft Fabric offre un espace de travail centralisé pour gérer l'intégralité d'un projet de science des données, de la définition du problème à la génération d'insights. Le processus commence par la définition du problème, suivi de l'identification et de l'ingestion des données nécessaires. Après l'ingestion, les données peuvent être explorées et préparées via des notebooks ou Data Wrangler. Pour l'entraînement des modèles de Machine Learning, Microsoft Fabric permet de suivre le progrès à travers

des expériences. Enfin, pour générer des insights à partir des modèles entraînés, la fonction PREDICT intégrée peut être utilisée.



## Bien démarrer avec Data Activator dans Microsoft Fabric

### Présentation

Data Activator dans Microsoft Fabric est conçu pour agir en réponse à des événements spécifiques dans les flux de données, facilitant l'automatisation et la réactivité dans divers scénarios. Trois exemples illustratifs incluent :

1. **Gestion de l'inventaire** : Un responsable d'entrepôt utilise Data Activator pour détecter les baisses de stock et déclencher automatiquement le réapprovisionnement afin d'éviter les ruptures.
2. **Engagement client** : Lorsqu'un client potentiel visite un site web, Data Activator alerte un vendeur pour initier une conversation en direct, améliorant l'interaction client.
3. **Sécurité réseau** : Un administrateur réseau utilise l'outil pour détecter des violations de sécurité en temps réel et prendre des mesures de protection immédiates.

Data Activator permet d'ingérer des données, de déclencher des flux Power Automate et de visualiser des données en temps réel avec Power BI, rendant l'intégration et l'action sur les données fluides et efficaces. Ce module fournit une compréhension approfondie de l'architecture de Data Activator et de son application pratique pour surveiller et répondre à des événements de données en temps réel, offrant une base pour son utilisation efficace.

## Comprendre Data Activator

Data Activator dans Microsoft Fabric est optimisé pour une réponse rapide et une analyse en temps réel des données. C'est une solution adaptée à divers scénarios nécessitant des actions immédiates basées sur des analyses en temps réel, comme la gestion dynamique de l'inventaire, l'engagement client en temps réel, et la gestion efficace des ressources dans les environnements cloud.

### Cas d'usage pour Data Activator :

- **Marketing** : Déclenchez des actions marketing en réaction à une baisse des ventes.
- **Gestion de produits périssables** : Envoyez des notifications liées à des changements de température qui pourraient affecter les produits.
- **Surveillance des performances** : Signalez les problèmes affectant l'expérience utilisateur sur les applications et les sites web en temps réel.
- **Logistique** : Déclenchez des alertes pour les expéditions non mises à jour dans les délais prévus.
- **Finance** : Alertez lorsqu'un compte dépasse un seuil financier spécifique.
- **Surveillance des processus** : Réagissez immédiatement aux anomalies ou aux défaillances dans les flux de traitement des données.

### Principes clés de Data Activator :

- **Événements** : Traite les sources de données comme des séquences d'événements qui reflètent des observations sur l'état de différents objets.
- **Objets** : Peuvent être tangibles (comme des véhicules) ou abstraits (comme des campagnes publicitaires), avec des propriétés spécifiques définies par des flux d'événements.
- **Déclencheurs** : Surveillent les événements et déclenchent des actions lorsque certaines conditions sont remplies.
- **Propriétés** : Utilisées pour réutiliser la logique entre différents déclencheurs et améliorer l'efficacité des actions prises.

### Interopérabilité avec d'autres fonctionnalités de Fabric :

Data Activator s'intègre bien avec d'autres outils de Microsoft Fabric, tels que EventStreams et Power BI, pour une ingénierie et une visualisation des données plus robustes.

Ce module vous donne une base pour comprendre et utiliser Data Activator efficacement dans le cadre de la réponse aux données en temps réel.

## Démarrage avec Data Activator

la configuration initiale de Data Activator dans Microsoft Fabric, un outil conçu pour réagir en temps réel à des données spécifiques. Voici les étapes essentielles pour commencer :

### Vérification de disponibilité :

- Avant de commencer, assurez-vous que Data Activator est disponible dans votre région et que votre organisation dispose des capacités nécessaires dans Fabric, surtout si vous utilisez la préversion publique qui peut être limitée géographiquement.

### Sélection de Data Activator :

- Pour utiliser Data Activator, sélectionnez-le dans le menu de Microsoft Fabric pour personnaliser votre espace de travail autour de cette fonctionnalité.

### Création d'un Reflex :

- Dans Fabric, les fonctionnalités de Data Activator sont gérées à travers des éléments appelés Reflexes. Chaque Reflex est configuré pour se connecter à des sources de données, surveiller des conditions spécifiques, et déclencher des actions en réponse. Vous pouvez créer un Reflex pour chaque processus ou segment d'affaires que vous souhaitez surveiller.

### Navigation entre les modes :

- Lorsque vous travaillez avec un Reflex, vous naviguerez entre deux modes principaux : le mode Données et le mode Conception.
  - En mode Données, vous pouvez visualiser et associer les nouvelles données aux objets spécifiques.
  - En mode Conception, vous développez des déclencheurs basés sur ces objets. Initialement vide, ce mode vous permettra de configurer et peupler vos données dans les éléments Reflex après leur création.

Ces instructions détaillées vous aideront à configurer et à utiliser Data Activator pour connecter des sources de données, créer des déclencheurs, et lancer des actions en surveillant les conditions en temps réel.

## Comprendre les déclencheurs, les conditions et les actions dans Data Activator

Data Activator dans Microsoft Fabric exploite des déclencheurs, des conditions, et des actions pour répondre efficacement à des événements en temps réel, transformant la gestion de données d'une manière proactive et réactive. Voici un résumé des concepts clés :

### Déclencheurs :

- Les déclencheurs sont essentiellement des instructions "si-alors" appliquées à vos données. Ils surveillent les événements et déclenchent des actions spécifiques lorsque certaines conditions sont remplies. Cela permet de répondre instantanément aux modifications importantes ou aux anomalies détectées dans les flux de données.

**Conditions :**

- Les conditions définissent les circonstances ou les états spécifiques qui activent un déclencheur. Elles agissent comme des règles qui déterminent quand une action doit être entreprise en réponse aux données et événements observés en temps réel.

**Actions :**

- Les actions sont les interventions exécutées lorsque les conditions d'un déclencheur sont remplies. Ces interventions peuvent varier de l'envoi d'alertes à l'initiation de workflows complexes, en fonction des besoins spécifiques identifiés par les déclencheurs.

La synergie entre déclencheurs, conditions, et actions permet une gestion des données dynamique et adaptative, facilitant une réponse immédiate aux insights générés en temps réel. Cette configuration crée un système de gestion des données qui n'est pas seulement réactif, mais aussi proactif, capable d'agir sur des données en constante évolution. La prochaine section approfondira la mise en œuvre et la gestion pratique de ces éléments dans le cadre de Data Activator.

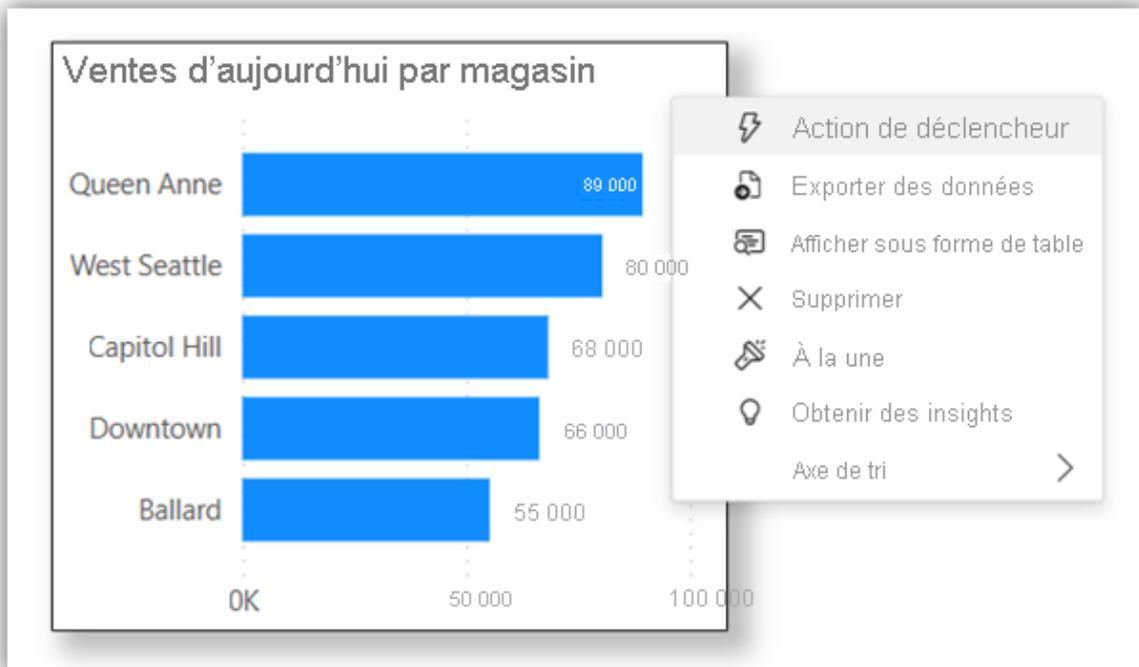
## Obtenir des données à partir de rapports et de flux d'événements Power BI avec Data Activator

Data Activator de Microsoft Fabric permet de déclencher des actions automatiques en réaction à des conditions prédéfinies observées dans les données. Ce module guide à travers la configuration de notifications pour des rapports Power BI et des flux d'événements.

**Rapports Power BI:**

- Utilisez Data Activator pour configurer des notifications basées sur des conditions spécifiques dans les rapports Power BI. Par exemple, si le trafic d'un site web suivi par un rapport Power BI tombe en dessous d'un seuil prédéfini, Data Activator peut automatiquement envoyer une alerte pour une action rapide.
- Avant de configurer, assurez-vous que votre rapport Power BI est publié en ligne dans un espace de travail Fabric avec une capacité Premium. Data Activator supporte les visuels avec des axes X et Y, mais pas ceux avec un axe de temps sur l'axe X.
- Pour configurer une alerte, sélectionnez l'option de déclenchement d'action sur un visuel de votre rapport, définissez les mesures à surveiller, la fréquence de

vérification, et spécifiez les conditions qui déclenchent l'alerte. Enregistrez ensuite le déclencheur dans un Reflex existant ou créez-en un nouveau.



### Flux d'événements:

- Connectez Data Activator à des données en temps réel via les flux de données de Fabric pour configurer des notifications basées sur des conditions spécifiques observées dans ces flux. Par exemple, si le trafic d'un site web suivi par un flux d'événements tombe en dessous d'un seuil prédéfini, une alerte peut être envoyée.
- Pour utiliser les flux d'événements, vous avez besoin d'un élément de flux d'événements et d'une connexion à une source. Les données de l'événement doivent être structurées en JSON avec une clé indiquant un ID d'objet unique.
- Pour connecter votre flux d'événements à Data Activator, ajoutez un Reflex comme nouvelle destination sur l'élément de flux d'événements.

### Créer une alerte ×

Pour que Reflex analyse ce graphique, sélectionnez la mesure, la condition et l'emplacement où enregistrer l'élément Reflex. En sélectionnant Continuer, vous serez redirigé vers Reflex où vous pourrez tester ou personnaliser votre alerte. [En savoir plus](#)

#### Éléments à surveiller

**Mesure**

**Pour chaque**

**Chaque**

#### Éléments à détecter

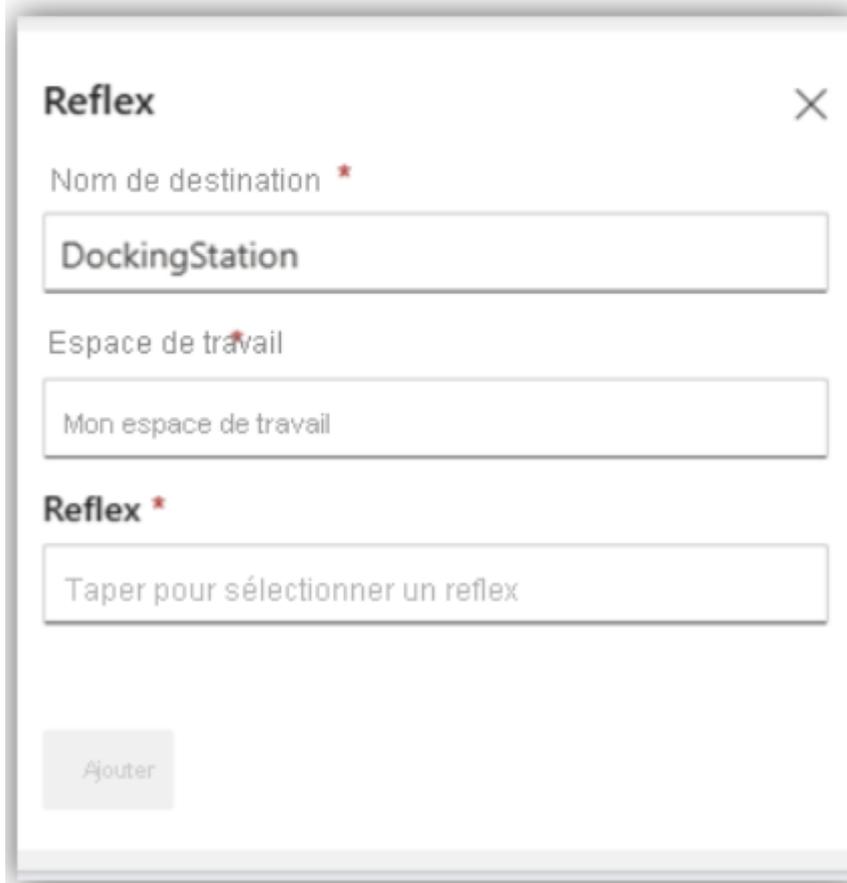
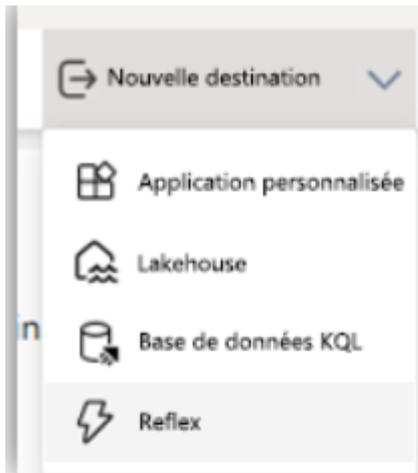
**Alerte lorsque la valeur est**

**Seuil**

#### Où enregistrer

**Espace de travail**

**Élément**

A screenshot of a configuration dialog box titled 'Reflex' with a close button (X) in the top right corner. The dialog contains three input fields: 'Nom de destination \*' with the value 'DockingStation', 'Espace de travail \*' with the value 'Mon espace de travail', and 'Reflex \*' with the placeholder text 'Taper pour sélectionner un reflex'. At the bottom left, there is a button labeled 'Ajouter'.

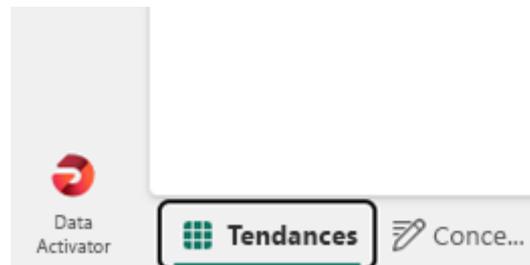
Ces étapes permettent d'intégrer et de réagir dynamiquement aux données surveillées, améliorant ainsi la réactivité aux conditions changeantes dans les rapports Power BI et les flux d'événements. La prochaine unité abordera plus en détail l'affectation des données dans Data Activator.

## Affecter des données dans Data Activator

Dans cette unité, nous explorons comment attribuer des données à des objets en utilisant le Mode Données de Data Activator, après avoir récupéré les données depuis Power BI et les EventStreams.

### Mode Données :

- **Sélection et Attribution** : Commencez par sélectionner le Mode Données dans l'interface de Data Activator. Ce mode est essentiel pour assigner les événements surveillés à des objets spécifiques. Si vous travaillez avec des données provenant de Power BI, Data Activator peut automatiquement affecter ces données à un nouvel objet créé.
- **Configuration de l'Objet** : Lors de l'attribution des données aux objets, vous définissez le nom de l'objet et la colonne clé qui identifie univoquement cet objet. Il est aussi possible d'attribuer des propriétés spécifiques à cet objet, ce qui sera expliqué plus en détail dans une unité ultérieure.



### Attribuer vos données

Attribuer à un nouvel objet    Attribuer à un objet existant

Connectez votre flux d'événements à un ou plusieurs objets pour l'utiliser dans Data Activator.

**Attribuer à un nouvel objet**

Nom de l'objet \*

Attribuer une colonne clé \*

Attribuer des propriétés

**Enregistrer**    Annuler

### Attribuer vos données

Attribuer à un nouvel objet    **Attribuer à un objet existant**

Connectez votre flux d'événements à un ou plusieurs objets préexistants pour l'utiliser dans Data Activator.

**Attribuer à un objet existant**

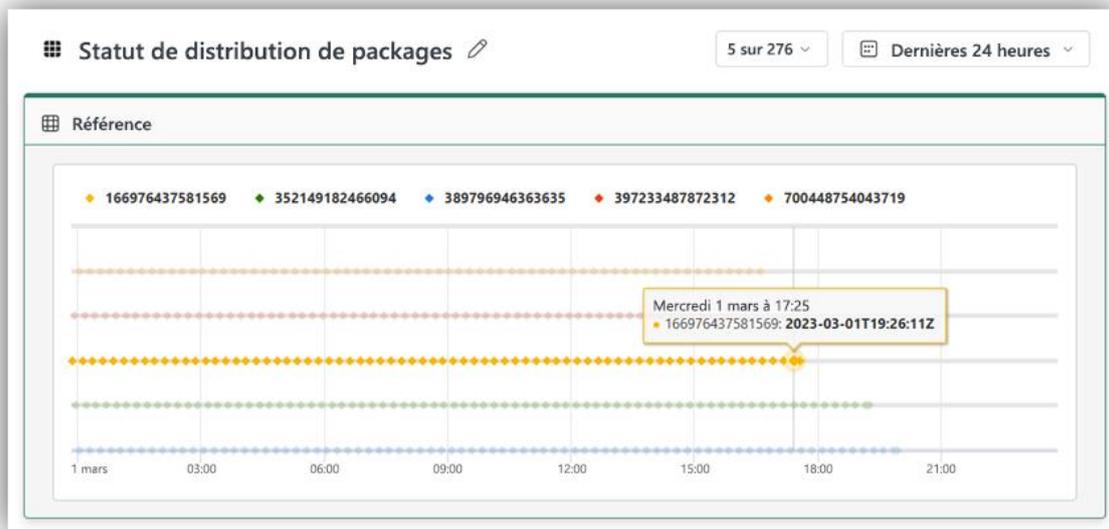
Sélectionner l'objet \*

Attribuer un ID pour le nom de l'objet \*

Attribuer des propriétés

## Visualisation et Personnalisation :

- Après la création d'un objet, vous passez au Mode Conception où vous pouvez visualiser les interactions et les données de l'objet au fil du temps. Vous pouvez voir un graphique qui montre, par exemple, l'état des livraisons d'une entreprise sur les dernières 24 heures. Ce graphique peut être personnalisé selon vos besoins pour mieux comprendre et suivre l'évolution des données.



## Flexibilité et Surveillance :

- Data Activator offre la flexibilité d'assigner plusieurs flux de données à un seul objet, ou de lier une source d'événement à différents objets. Cette capacité permet de surveiller divers aspects des données de manière efficace. Lors de l'ajout de données à un objet existant, il est crucial de bien utiliser la colonne clé pour éviter des anomalies dans les résultats, assurant ainsi la fiabilité de votre système de surveillance.

## Prochaines Étapes :

- La prochaine unité abordera la création de déclencheurs pour surveiller les conditions spécifiques et initier des actions basées sur ces conditions, renforçant ainsi la capacité de réaction de votre système aux événements en temps réel.

Ce processus d'attribution de données joue un rôle crucial dans la gestion et l'analyse des informations dans Data Activator, permettant une surveillance détaillée et réactive des différents éléments de votre entreprise.

## Créer des déclencheurs dans DATA Activator

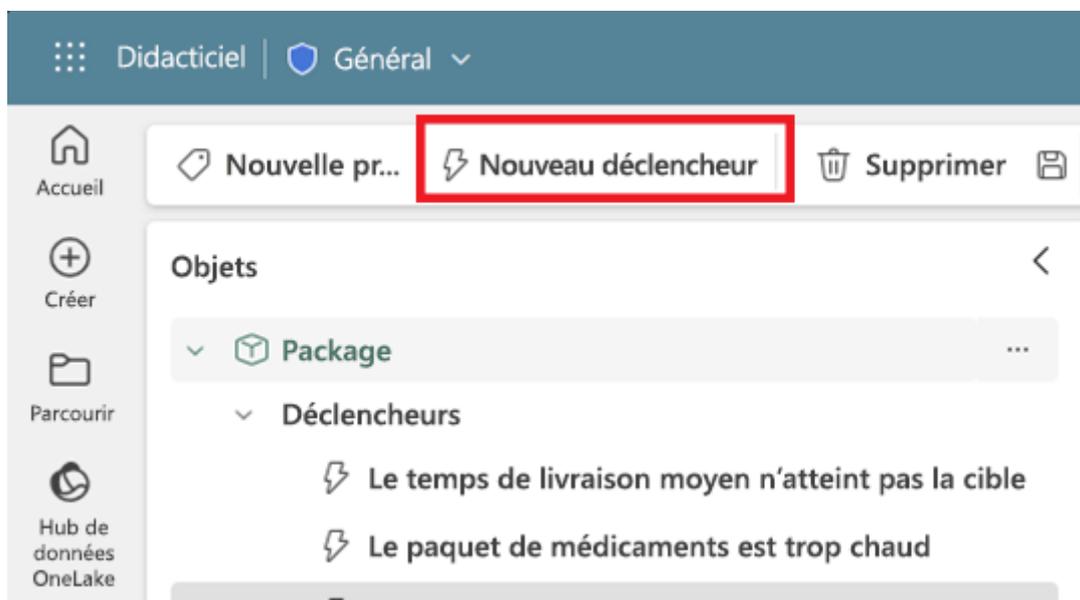
Dans cette unité, nous explorons la création de déclencheurs dans Data Activator pour agir sur les données collectées. Voici un résumé détaillé des étapes pour configurer ces déclencheurs :

### Mode Conception

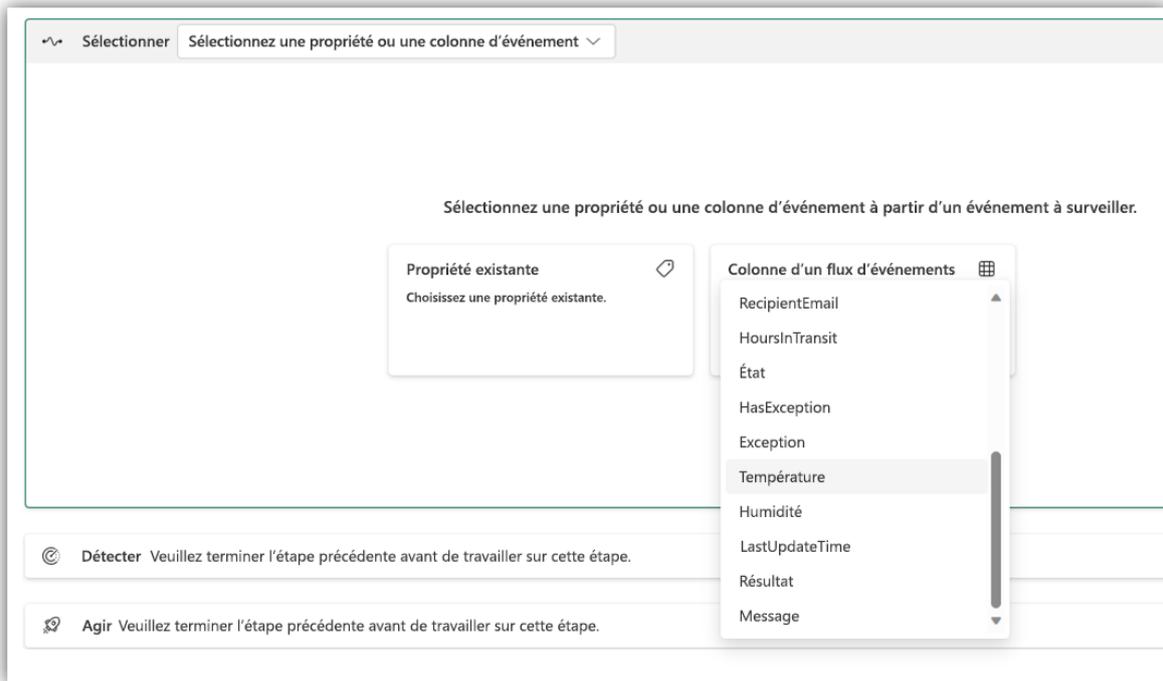
- **Sélectionner le Mode Conception:** Accédez à ce mode via l'onglet situé en bas de l'écran dans Data Activator. C'est ici que vous créerez et configurerez vos déclencheurs.

### Définir un Déclencheur et une Action

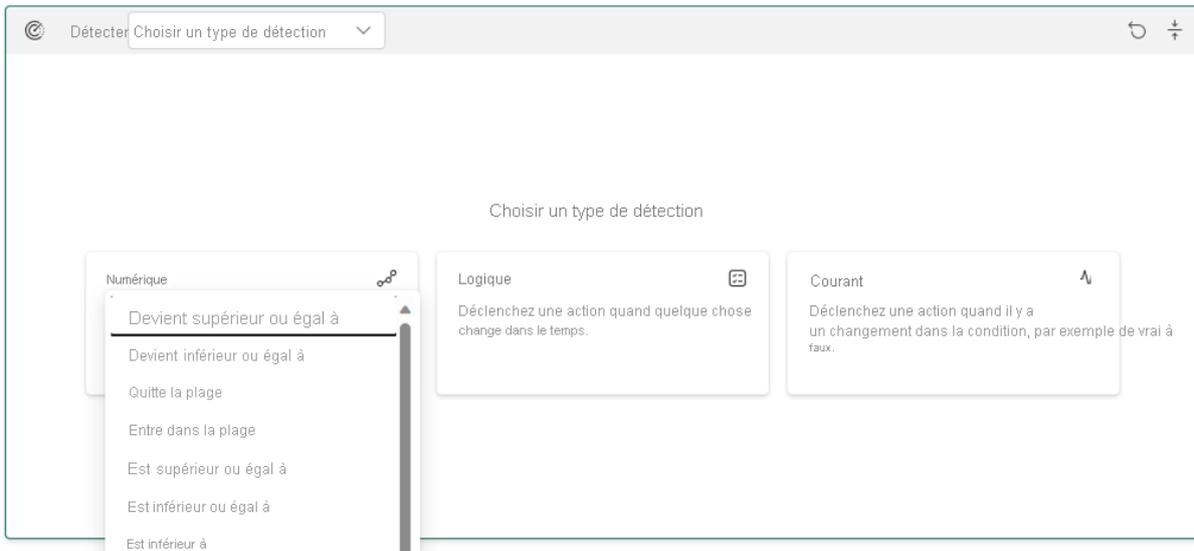
- **Création de Déclencheurs:** Les déclencheurs sont configurés pour réagir à des conditions spécifiques détectées dans vos données. Par exemple, pour une entreprise pharmaceutique, un déclencheur peut être créé pour alerter si la température des colis expédiés dépasse 48 degrés.



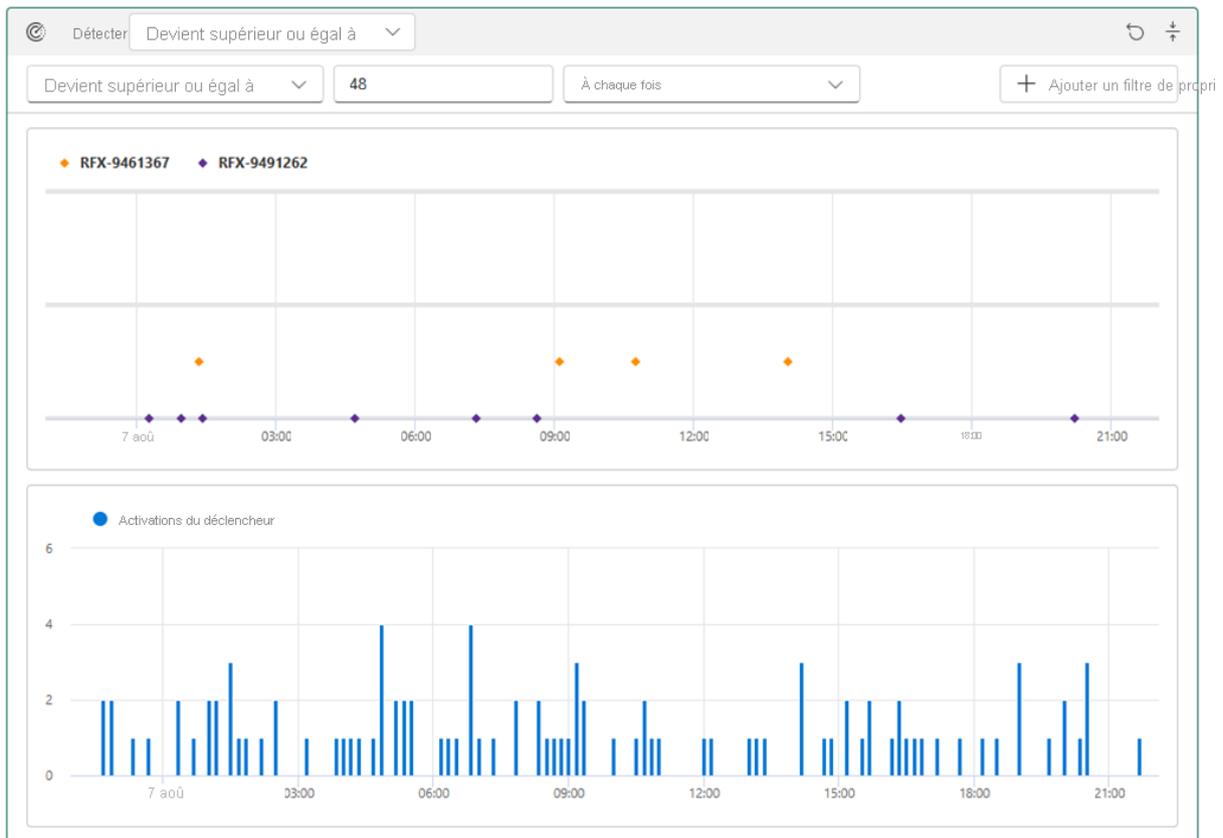
- **Sélection des Données:** Choisissez une colonne ou une propriété pour surveiller, comme la température dans un colis pharmaceutique.



- **Définir les Conditions:** Spécifiez les conditions sous lesquelles le déclencheur doit être activé, comme une température dépassant 48 degrés.



- **Aperçu des Données:** Visualisez un aperçu des valeurs de données pour confirmer le comportement du déclencheur.



## Configurer l'Action du Déclencheur

- **Choisir l'Action:** Une fois la condition déclenchée, choisissez une action appropriée, comme envoyer un message Teams au service d'expédition.
- **Personnalisation de l'Action:** Fournissez les détails nécessaires pour l'action, tels que les destinataires et le contenu du message.
- **Tester le Déclencheur:** Avant de finaliser, testez le déclencheur pour vous assurer qu'il fonctionne comme prévu en envoyant une alerte de test basée sur des données historiques.

The screenshot shows the configuration interface for the trigger action. The title is 'Envoyer un message Teams' (Send a Teams message). The form includes the following fields:

- Destinataires\*** (Recipients): A text input field.
- Titre\*** (Title): A text input field containing the text 'La condition du déclencheur pour l'objet suivant a été atteinte' (The trigger condition for the following object has been reached).
- Message facultatif** (Optional message): A text input field.
- Informations supplémentaires** (Additional information): A dropdown menu with the text 'Sélectionnez des propriétés à inclure dans votre notification' (Select properties to include in your notification).

## Création de Propriétés

- **Utiliser des Propriétés:** Pour appliquer la même logique de déclenchement à plusieurs déclencheurs, vous pouvez créer et réutiliser des propriétés.
- **Définir et Référencer des Propriétés:** Créez une propriété, comme "Maximum Temp last hour", et utilisez-la dans vos déclencheurs pour uniformiser et simplifier la gestion.

The screenshot shows the 'Temp maximale dernière heure' trigger configuration. The left sidebar lists 'Objets' (Objects) including 'Package', 'Déclencheurs' (Triggers), 'Propriétés' (Properties), and 'Événements' (Events). The main area displays the trigger configuration with the following settings:

- Sélectionner:** Température
- Package en transit:** Sélectionné
- Dépassement de temps maximum:** 1 heure

A line chart below the configuration shows temperature data for three packages: DFX-1311289 (blue), DFX-1023428 (orange), and DFX-1416806 (green). The x-axis represents time from 16:00 to 12:00 on August 11th. The y-axis represents temperature from 20 to 60. Data points are plotted for each package, with DFX-1023428 showing a peak temperature of approximately 55.

The screenshot shows the 'Le paquet est trop chaud' trigger configuration. The left sidebar lists 'Objets' (Objects) including 'Package', 'Déclencheurs' (Triggers), 'Propriétés' (Properties), and 'Événements' (Events). The main area displays the trigger configuration with the following settings:

- Sélectionner:** Sélectionnez une propriété ou une colonne d'événement
- Propriété existante:** Temp. maximale dernière heure
- Colonne d'un flux d'événements:** Sélectionné

The configuration page includes instructions: "Sélectionnez une propriété ou une colonne à partir d'un événement à surveiller." Below this, there are two options: "Propriété existante" (Choose an existing property) and "Colonne d'un flux d'événements" (Choose a column from the event flow). A button at the bottom reads "Détecter Veuillez terminer l'étape précédente avant de passer à cette étape".

## Gestion des Déclencheurs

- **Démarrage et Arrêt:** Une fois créé, le déclencheur peut être démarré, arrêté, ou mis à jour directement depuis Data Activator.

Cette structure permet une surveillance proactive et une réaction rapide aux changements significatifs dans les données, améliorant ainsi la réactivité opérationnelle de l'organisation. La prochaine étape implique la mise en pratique de ces configurations dans un exercice pratique.

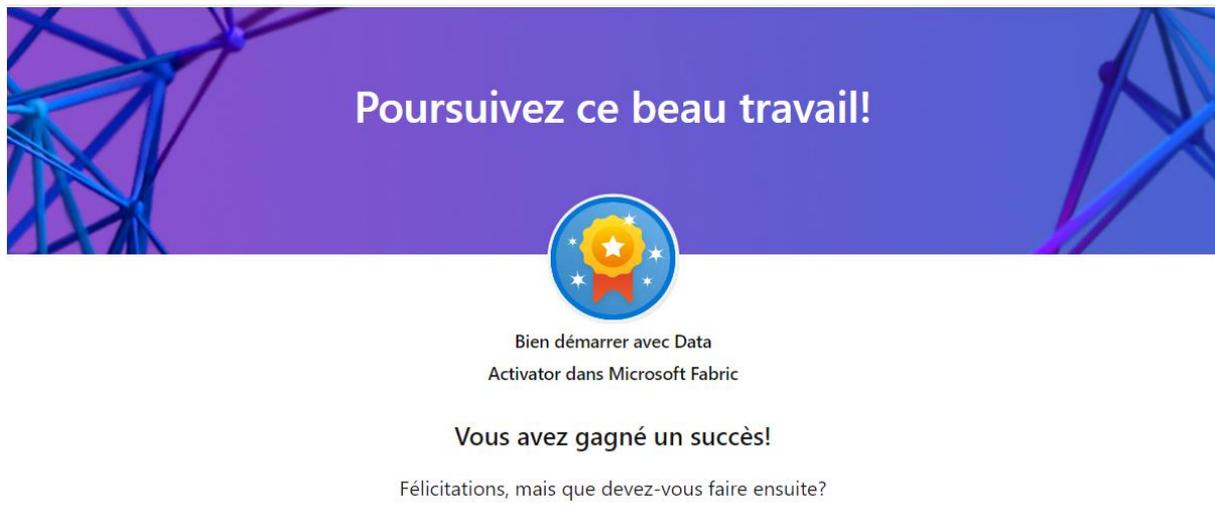
## Exercice

- Créer un espace de travail
- Scénario
- Créer un réflexe
- Familiarisez-vous avec l'écran d'accueil du reflex
- Créer un objet
- Créer un déclencheur
- Mettre à jour et arrêter un déclencheur
- Nettoyer les ressources

## Résumé

Ce module vous a introduit à Data Activator, un service de Microsoft Fabric conçu pour interagir avec les données en temps réel. Vous avez exploré les utilisations typiques de Data Activator, les aspects fondamentaux de son architecture, et sa capacité à s'intégrer avec d'autres fonctionnalités de Fabric. Data Activator traite des événements générés à partir de flux de données, offrant la possibilité de gérer, filtrer, et associer des actions spécifiques à ces événements.

Les points essentiels à retenir incluent la puissance de Data Activator pour gérer des situations nécessitant une réponse rapide et adaptée basée sur l'analyse en temps réel des données. Cet outil est idéal pour initier des actions marketing, envoyer des notifications, signaler des incidents en temps réel, activer des alertes, répondre à des anomalies ou à des défaillances dans les processus, et plus encore. La capacité de Data Activator à configurer des déclencheurs, conditions, et actions basées sur des données en temps réel permet de réagir de manière proactive et informée aux changements observés dans vos flux de données.



# Administrer Microsoft Fabric

## Introduction

L'administration de la plateforme Fabric nécessite une compréhension approfondie de son architecture, de ses fonctionnalités de sécurité et de gouvernance, ainsi que de ses capacités analytiques. Les administrateurs doivent aussi maîtriser les options de déploiement, de licence, et les outils d'administration comme le portail d'administration Fabric pour configurer et gérer l'environnement selon les besoins de l'organisation.

Les administrateurs de Fabric collaborent étroitement avec les utilisateurs métier, les analystes de données, et d'autres professionnels IT pour assurer que la plateforme est déployée et utilisée de manière à atteindre les objectifs métier tout en respectant les stratégies et normes de l'organisation. Ce module vise à fournir une compréhension claire du rôle de l'administrateur Fabric et des responsabilités associées à cette fonction.

## Comprendre l'architecture Fabric

Microsoft Fabric est une plateforme d'analytique tout-en-un pour les entreprises, englobant plusieurs services clés :

- **Entrepôt de données**
- **Engineering de données**
- **Intégration de données**
- **Science des données**
- **Analytique en temps réel**
- **Décisionnel**

La plateforme, fonctionnant comme un Logiciel en tant que service (SaaS), utilise OneLake comme fondation. OneLake sert de magasin natif centralisé sans configuration supplémentaire, organisant les données en conteneurs pour une gestion simplifiée. Il existe un seul OneLake par locataire, couvrant les utilisateurs, les régions, et même les clouds.



### Concepts Clés de Fabric :

- **Locataire Fabric** : Un espace unique par organisation aligné avec Microsoft Entra ID, situé à la racine de OneLake.
- **Capacité** : Les ressources dédiées disponibles, variées selon les besoins et utilisations.
- **Domaine** : Regroupe les espaces de travail logiquement (par exemple, ventes, marketing, finances).
- **Espace de travail** : Contient des éléments spécifiques à une fonction, utilisant la capacité disponible.
- **Éléments Fabric** : Composants créés et gérés sur la plateforme, tels que entrepôts de données et rapports.

La compréhension de ces concepts est cruciale pour l'administration efficace de l'environnement Fabric, permettant une gestion optimale et une organisation claire des ressources et des accès.

### Comprendre le rôle Administrateur Fabric

L'administration de Microsoft Fabric est assurée par plusieurs rôles spécifiques, y compris les administrateurs Microsoft 365, Power Platform, et les administrateurs de capacité Fabric. Ces rôles travaillent ensemble pour gérer efficacement la plateforme.

#### Rôles et Outils d'Administration:

- **Portail d'administration Fabric**: C'est le principal outil web pour gérer tous les aspects de la plateforme, où vous pouvez configurer les paramètres, gérer les utilisateurs et groupes, accéder aux journaux d'audit, et superviser l'utilisation et les performances de Fabric.
- **Applets de commande PowerShell**: Utilisées pour automatiser les tâches administratives courantes, telles que la création et gestion de groupes, la configuration de sources de données et de passerelles, et la surveillance de l'utilisation et des performances.

- **SDK et API d'administration:** Permettent d'interagir programmablement avec Fabric, facilitant l'automatisation des tâches administratives et l'intégration de Fabric avec d'autres systèmes.

### Tâches d'Administration Principales:

1. **Sécurité et contrôle d'accès:** Utiliser le contrôle d'accès basé sur les rôles (RBAC) et configurer des passerelles de données pour sécuriser l'accès aux données sensibles.
2. **Gouvernance des données:** Appliquer des stratégies de gouvernance pour sécuriser les connexions et superviser les métriques d'utilisation et de performances.
3. **Personnalisation et configuration:** Adapter la plateforme aux besoins spécifiques de l'organisation, y compris la configuration des liens privés et la définition de politiques de classification des données.
4. **Supervision et optimisation:** Gérer la capacité, surveiller les performances, et résoudre les problèmes de connectivité ou d'actualisation des données.

### Espace de travail de supervision de l'administration:

- Un espace dédié pour les administrateurs de locataires Fabric, offrant des rapports et des analyses pour aider à surveiller l'utilisation et les performances et à identifier les problèmes potentiels.

Ces outils et tâches garantissent que l'administration de Fabric est complète, permettant une gestion efficace de la plateforme dans le respect des objectifs et normes de l'organisation.

## Gérer la sécurité Fabric

La sécurité dans Microsoft Fabric, alignée sur les principes de sécurité de Power BI, est une priorité pour les administrateurs, qui sont responsables de la gestion des utilisateurs, des groupes, et de la manière dont le contenu est partagé et distribué au sein de l'environnement Fabric.

### Gestion des utilisateurs et des licences:

- **Attribution de licences:** Les licences déterminent le niveau d'accès et de fonctionnalité disponible pour les utilisateurs. Les administrateurs doivent attribuer les licences de manière à garantir un accès adéquat aux données et aux fonctionnalités analytiques tout en limitant l'accès aux données sensibles pour se conformer aux réglementations en matière de protection des données.
- **Optimisation des coûts:** Une gestion efficace des licences aide à contrôler les coûts, en s'assurant que les licences sont allouées uniquement aux utilisateurs qui en ont besoin, évitant ainsi des dépenses inutiles.

La gestion des licences est centralisée dans le Centre d'administration Microsoft 365, où les administrateurs peuvent attribuer et gérer les licences.

## Gestion des éléments et du partage:

- **Gestion du contenu:** En tant qu'administrateur, vous devez contrôler comment les utilisateurs partagent et distribuent le contenu, incluant les entrepôts de données, pipelines de données, modèles sémantiques, rapports et tableaux de bord.
- **Distributions via applications d'espace de travail:** Les éléments sont souvent mieux distribués à travers une application d'espace de travail, permettant un partage contrôlé. L'accès en lecture seule peut être accordé pour consulter les rapports, tandis que l'accès aux espaces de travail peut être octroyé pour la collaboration et le développement.
- **Principes de sécurité:** Appliquer les principes de sécurité comme l'octroi des droits les moins permissifs est essentiel pour protéger les données. Les bonnes pratiques de gestion et de distribution des éléments doivent être suivies pour renforcer la sécurité et l'efficacité.

Les administrateurs doivent aussi s'assurer que le partage et la distribution du contenu respectent les stratégies et procédures internes de l'organisation, tout en facilitant la collaboration et l'accès aux informations nécessaires pour les utilisateurs autorisés.

## Gouverner des données dans Fabric

Fabric offre des outils de gouvernance des données intégrés pour aider les administrateurs à contrôler et gérer les données au sein de leur organisation.

### Approuver et Certifier le Contenu dans Fabric

- **Approbaton de contenu:** Cette fonctionnalité permet aux administrateurs de marquer certains éléments Fabric comme fiables pour une utilisation organisationnelle. Cela inclut la promotion de contenus au sein d'un espace de travail ou de l'organisation entière, identifiés par un badge "Promu". Les administrateurs et les contributeurs de l'espace de travail peuvent promouvoir du contenu.
- **Certification de contenu:** C'est un processus formel d'examen du contenu par un réviseur désigné, après lequel le contenu approuvé reçoit un badge "Certifié". L'administrateur gère ce processus qui peut être adapté selon les besoins spécifiques de l'organisation.

### Recherche de Données Sensibles

- **Analyse des métadonnées:** L'API de scanner permet aux administrateurs d'analyser les éléments Fabric, tels que les entrepôts de données, les pipelines de données, les modèles sémantiques, et les rapports, pour identifier des données sensibles. Cette API peut analyser des données à la fois structurées et non structurées.

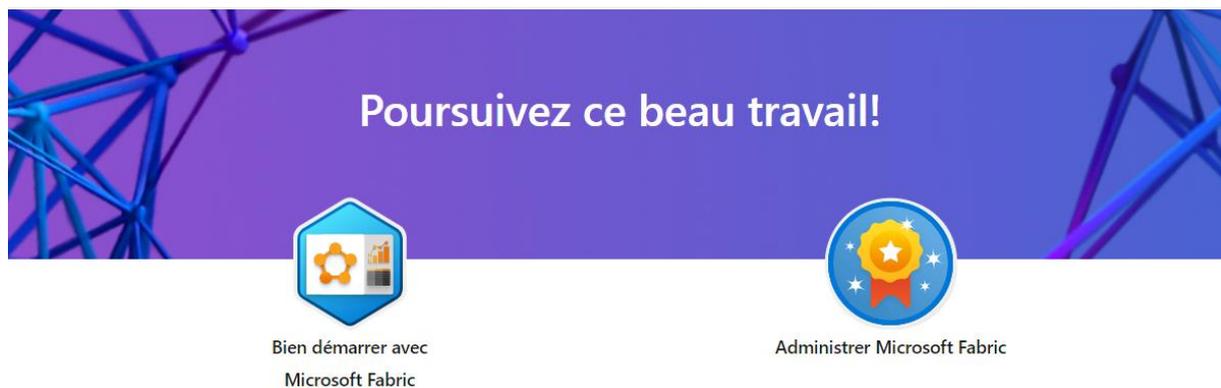
### Suivi de la Traçabilité des Données

- **Traçabilité des données:** Cette fonctionnalité permet de suivre le flux des données à travers Fabric, montrant l'origine, la transformation et la destination des données. Cela aide à comprendre comment les données sont utilisées et transformées au sein de l'environnement Fabric.

Ces outils de gouvernance de données fournissent des mécanismes robustes pour assurer la sécurité des données, leur conformité, et pour gérer la manière dont les informations sont partagées et traitées au sein de l'organisation. Ils permettent également une gestion stratégique des accès et de la visibilité des données pour maintenir des standards élevés de sécurité et de conformité.

## Résumé

Dans ce module, vous avez découvert l'architecture Fabric et le rôle d'un administrateur dans la gestion de la plateforme Fabric. Vous avez également exploré les différents outils disponibles pour la gestion de la sécurité et du partage, ainsi que les fonctionnalités de gouvernance qui peuvent être utilisées pour appliquer des normes et garantir la conformité. Bien comprendre comment gérer un environnement Fabric garantit qu'il est sécurisé, conforme et bien régi. Grâce à ces connaissances, vous êtes bien équipé pour permettre à votre organisation de tirer le meilleur parti de Fabric et de retirer de précieux insights de toutes vos données.



Vous avez gagné 2 succès!

Félicitations, mais que devez-vous faire ensuite?

## Ingérer des données avec des notebooks Spark et Microsoft Fabric

### Présentation

L'intégration de Microsoft Fabric dans une entreprise commence par l'ingestion de données dans le lakehouse Fabric, utilisant le processus ETL (extraction,

transformation, chargement). Selon les besoins spécifiques, vous pouvez commencer par l'ingestion de données ou opter pour leur nettoyage et transformation préalables. Les notebooks Fabric facilitent l'adaptation de ce processus à différents workflows, permettant l'extraction, le chargement et la transformation des données externes. Une connaissance préalable de Spark ou des notebooks peut être bénéfique mais n'est pas essentielle.

Les notebooks Fabric sont particulièrement utiles pour gérer de grandes quantités de données externes et effectuer des transformations complexes. À la fin de ce module, vous serez capable d'utiliser les notebooks Spark et Fabric pour ingérer des données externes dans votre lakehouse et vous apprendrez également des techniques de base de transformation et d'optimisation pour rendre le processus ETL plus efficace.

## Se connecter aux données avec Spark

Les notebooks Fabric offrent une automatisation pour l'ingestion et la transformation de données, contrastant avec les méthodes manuelles et les flux de données qui ne supportent pas aussi bien les grands modèles sémantiques. Ces notebooks permettent une gestion complète des processus ETL et sont stockés dans l'espace de travail où ils sont créés. Ils utilisent par défaut PySpark pour exploiter les capacités de Spark dans un environnement distribué, avec la possibilité d'utiliser également HTML, Spark SQL, et SparkR.

### Code PySpark pour se connecter à Azure Blob Storage :

Python

```
# Write DataFrame to Parquet file format
parquet_output_path = "your_folder/your_file_name"
df.write.mode("overwrite").parquet(parquet_output_path)
print(f"DataFrame has been written to Parquet file: {parquet_output_path}")

# Write DataFrame to Delta table
delta_table_name = "your_delta_table_name"
df.write.format("delta").mode("overwrite").saveAsTable(delta_table_name)
print(f"DataFrame has been written to Delta table: {delta_table_name}")
```

### Code PySpark pour se connecter à Azure SQL Database avec un principal de service :

Python

```
# Use format and save to load as a Delta table
table_name = "nyctaxi_raw"
filtered_df.write.mode("overwrite").format("delta").save(f"Tables/{table_name}")

# Confirm load as Delta table
print(f"Spark DataFrame saved to Delta table: {table_name}")
```

Ces extraits de code vous aideront à effectuer des tâches d'ingestion de données et à

vous connecter à différentes sources de données externes dans vos projets utilisant les notebooks Fabric.

## Envisager des usages adaptés aux données ingérées

Après avoir ingéré des données brutes dans le lakehouse Fabric, également connu sous le nom de couche bronze dans une architecture Medallion, il est important de considérer où et comment les transformations des données seront effectuées, en tenant compte des différents utilisateurs et de leurs interactions avec les données.

### Transformation pour différents utilisateurs

- **Nettoyage de base:** Avant de rendre les données disponibles, il est crucial de réaliser un nettoyage initial, qui peut inclure la suppression des doublons, la gestion des erreurs, la conversion des valeurs nulles, et l'élimination des entrées vides. Ces étapes garantissent la qualité et la cohérence des données.
- **Experts en science des données:** Ce groupe d'utilisateurs préfère souvent accéder aux données brutes pour pouvoir explorer de vastes ensembles de données sans modifications préalables. Le Data Wrangler de Fabric facilite cette exploration et permet de générer du code de transformation adapté à leurs besoins spécifiques.
- **Analystes de données Power BI:** Ces utilisateurs ont besoin de données plus transformées et modélisées avant de commencer à utiliser Power BI. Bien que Power BI puisse effectuer certaines transformations, travailler avec des données déjà bien préparées permet aux analystes de développer des rapports et des insights de manière plus efficace et rapide.

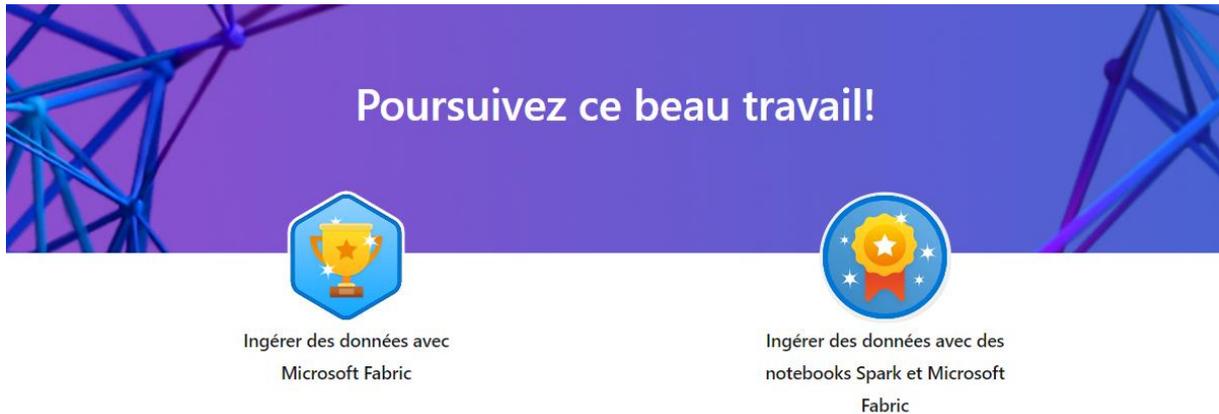
Le module "Utiliser Apache Spark dans Microsoft Fabric" approfondit comment utiliser les notebooks Fabric pour visualiser, agréger et transformer des données en utilisant Spark, offrant ainsi une flexibilité pour répondre aux besoins spécifiques des différents utilisateurs au sein de l'organisation.

## Exercice

- Créer un espace de travail
- Créer un espace de travail et une destination Lakehouse
- Créer un notebook Fabric et charger des données externes
- Transformer et charger des données dans une table Delta
- Optimiser les écritures de la table Delta
- Analyser les données de la table Delta avec des requêtes SQL
- Nettoyer les ressources

Dans ce module, vous avez appris à utiliser les notebooks Spark et Fabric pour automatiser l'ingestion et la transformation de données dans un lakehouse, utilisant le moteur Spark pour des opérations rapides et distribuées. Les lakehouses sont adaptés à divers formats de données, y compris les tables Delta, qu'il est bénéfique d'optimiser pour améliorer les performances. Il est aussi crucial de planifier soigneusement

l'emplacement des transformations et de considérer comment les utilisateurs interagissent avec les données pour assurer l'efficacité et la pertinence des processus.



Vous avez gagné 2 succès!

Félicitations, mais que devez-vous faire ensuite?

## Implémenter un lakehouse avec Microsoft Fabric

### Introduction

Dans le contexte des entreprises guidées par les données, l'importance d'exploiter des sources de données vastes et variées est cruciale. Le lakehouse Fabric, qui fusionne les lacs de données et les entrepôts de données, représente une solution optimale pour la gestion et l'analyse de ces données. L'architecture en médaillon, qui est devenue une norme pour l'analyse basée sur lakehouse, structure et optimise l'environnement lakehouse en trois couches : bronze, argent, et or.

Ce module est conçu pour les professionnels de tous niveaux, des ingénieurs données expérimentés aux débutants en analytique, et inclut des exemples pratiques ainsi qu'un exercice. Les participants doivent avoir une connaissance préalable du lac de données Fabric, ainsi qu'une compréhension de base de SQL et Power BI. Le contenu aborde l'organisation, le raffinement et la sécurisation des données, et propose de découvrir comment créer et gérer une architecture en médaillon dans les lakehouses Fabric, interroger les données, créer des rapports, et appliquer les meilleures pratiques de sécurité et de gouvernance.

### Décrire l'architecture en médaillon

Les entrepôts de données de Fabric utilisent le format Delta Lake, supportant les transactions ACID pour garantir la fiabilité des données. L'architecture en médaillon, recommandée pour les lakehouses, organise logiquement les données en trois couches principales pour améliorer leur qualité :

1. **Couche bronze** : Première étape où les données brutes sont stockées sans modification, incluant des données structurées, semi-structurées ou non structurées.
2. **Couche argent** : Les données sont validées et raffinées ici, avec des processus comme la fusion de données et l'application de règles de validation, servant de référentiel central pour consultation.
3. **Couche or** : Les données sont encore plus affinées et enrichies pour répondre aux besoins spécifiques de l'entreprise, prêtes à être utilisées par les équipes en aval.

## Medallion Architecture



L'architecture en médaillon est flexible, permettant des adaptations comme l'ajout d'une couche supplémentaire brute ou une couche "platine" pour des données spécifiques. La gestion des données à travers ces couches dans Fabric inclut le transfert (amélioration et préparation des données pour les analyses en aval) et l'orchestration des données (coordination des processus liés aux données).

Les outils de transformation dans Fabric incluent les flux de données pour des transformations simples et les notebooks pour des transformations complexes, permettant de sauvegarder les données transformées sous forme de tables Delta gérées. L'orchestration est réalisée principalement par les pipelines, qui peuvent être automatisés ou déclenchés par des événements, assurant le transfert efficace des données à travers les différentes couches de l'architecture en médaillon.

### Implémenter une architecture en médaillon dans Fabric

Pour mettre en pratique l'architecture en médaillon dans Fabric, suivez ces étapes :

1. **Configurez la base** : Créez votre lakehouse Fabric adapté à vos besoins. Vous pouvez opter pour un seul lakehouse pour diverses architectures en médaillon

ou pour plusieurs lakehouses dans différents espaces de travail selon votre cas d'usage.

## 2. Concevez votre architecture :

- **Couche bronze** : Ingérez des données brutes en utilisant des pipelines, des flux de données ou des notebooks.
  - **Couche argent** : Transformez et validez les données pour améliorer leur qualité et leur cohérence à l'aide de flux de données ou de notebooks. Cette étape ne se concentre pas sur la modélisation mais sur la préparation des données.
  - **Couche or** : Modélisez les données dans un schéma en étoile optimisé pour la création de rapports. Utilisez des points de terminaison analytiques SQL ou des modèles sémantiques pour transformer davantage les données et préparer des rapports. Vous pouvez créer plusieurs couches d'or pour différents publics ou fonctions au sein de l'entreprise.
3. **Ingérez des données dans le bronze** : Sélectionnez les méthodes appropriées (pipelines, flux de données, ou notebooks) pour ingérer des données dans la couche bronze.
  4. **Transformez les données, puis chargez-les en argent** : Utilisez les outils disponibles pour nettoyer et valider les données dans la couche argent. Assurez-vous que les données sont de haute qualité avant de passer à la couche suivante.
  5. **Générez une couche d'or** : Développez la couche or pour des besoins spécifiques tels que les rapports financiers, ventes, ou optimisations pour le Machine Learning. Modélisez les données pour répondre aux besoins spécifiques des utilisateurs finaux.
  6. **Activez la consommation en aval** : Configurez les autorisations nécessaires et connectez-vous au point de terminaison analytique SQL pour permettre l'accès aux données transformées. Utilisez des outils comme Power BI pour connecter, modéliser et visualiser les données pour la création de rapports.

Cette approche structurée garantit que les données circulent efficacement entre les couches, s'améliorent en qualité et en utilité tout au long du processus, et sont finalement prêtes pour une utilisation stratégique au sein de l'organisation.

## Interroger les données, puis générer des rapports les concernant dans votre lakehouse Fabric

Avec l'architecture en médaillon de Fabric en place, les équipes peuvent débiter l'interrogation et la création de rapports à partir des données stockées dans le lakehouse. Voici les outils et méthodes principaux à utiliser :

### Interrogation des Données

- **SQL et T-SQL** : Utilisez SQL pour interroger les données dans la couche or, et potentiellement dans toutes les couches de l'architecture en médaillon. Les fonctions peuvent être enregistrées, des vues générées, et la sécurité SQL appliquée pour contrôler l'accès et l'utilisation des données.
- **Point de terminaison d'analytique SQL** : Ce point de terminaison permet d'écrire des requêtes SQL, de gérer des modèles sémantiques, et de bénéficier d'une nouvelle expérience de requête visuelle dans Fabric. Il est important de noter que ce point fonctionne en mode lecture seule pour les tables delta.

### Création de Rapports

- **Power BI en mode Direct Lake** : Créez un modèle sémantique Power BI directement connecté à votre lakehouse, permettant d'interroger des données en temps réel. Le système approvisionne un modèle sémantique par défaut lors de la création d'un lakehouse, optimisant ainsi l'accès et la performance grâce à la mise en cache des données fréquemment demandées.

### Personnalisation des Couches de Médaillon

- **Adaptation aux Besoins Spécifiques** : Personnalisez les couches pour répondre aux besoins spécifiques des différentes équipes ou domaines fonctionnels de votre organisation, comme les finances, les ventes, ou la science des données. Chaque domaine peut avoir sa propre couche or optimisée, ce qui garantit la pertinence et la performance des données pour les utilisateurs finaux.
- **Formats de Données pour Applications et Outils Tiers** : Adaptez vos données pour répondre aux exigences spécifiques de certaines applications ou outils tiers, en utilisant les couches de votre architecture en médaillon pour générer des données nettoyées et correctement formatées.

En résumé, l'utilisation efficace de l'architecture en médaillon dans Fabric permet non seulement d'interroger et de rapporter les données de manière efficace mais aussi de personnaliser l'accès et le traitement des données pour divers cas d'utilisation et utilisateurs finaux, maximisant ainsi la valeur des données au sein de l'organisation.

## Facteurs à prendre en compte dans la gestion de votre lakehouse

Lors de la gestion d'un lakehouse dans Fabric, deux aspects principaux doivent être considérés : la sécurité et l'intégration continue/livraison continue (CI/CD). Voici comment ces facteurs influencent la gestion efficace de votre lakehouse :

### Sécurité du Lakehouse

- **Autorisations de Sécurité** : Assurez-vous que seulement les utilisateurs autorisés accèdent aux données en définissant des autorisations au niveau de l'espace de travail (pour tout l'espace) ou de l'élément (pour des éléments spécifiques). Cette stratégie permet de gérer l'accès de manière granulaire, ce qui est particulièrement utile lors de la collaboration avec des équipes diverses ou des membres hors de votre espace de travail habituel.

- **Gestion Stratégique des Couches** : Stockez les différentes couches (bronze, argent, or) dans des espaces de travail distincts pour améliorer la sécurité et la gestion des capacités. Ceci est crucial pour protéger les données sensibles et optimiser l'efficacité du système.
- **Contrôle d'Accès par Couches** : Restreignez l'accès à des niveaux spécifiques en fonction de la sensibilité des données. Par exemple, limitez l'accès à la couche or pour assurer que les données soient manipulées principalement en lecture seule et ajustez les permissions pour les couches argent et bronze en fonction des besoins de traitement et de sécurité des données.

### Intégration Continue et Livraison Continue (CI/CD)

- **Processus CI/CD** : Concevez un processus de CI/CD qui inclut des contrôles de qualité des données, gestion de versions, déploiements automatisés, surveillance et mesures de sécurité. Cela garantit que les pipelines de données sont déployés de manière fluide et efficace.
- **Scalabilité et Récupération d'Urgence** : Assurez-vous que le système peut gérer l'augmentation des charges de travail et qu'il existe des procédures de récupération en cas de pannes pour minimiser les interruptions.
- **Collaboration et Conformité** : Facilitez la collaboration entre les équipes grâce à des outils intégrés comme Git dans Fabric, permettant le contrôle de versions, la révision de code et le travail collaboratif ou individuel sur des branches séparées.
- **Intégration de Git** : Utilisez Git pour intégrer des pratiques de développement solides directement dans Fabric, aidant à sauvegarder les travaux, gérer les révisions, et collaborer efficacement.

### Application CI/CD dans la Couche Or

- **Qualité et Fiabilité des Données** : La CI/CD est cruciale pour la couche or car elle assure la disponibilité de données de haute qualité pour les analyses. Les processus automatisés réduisent les erreurs manuelles et garantissent que les données sont continuellement mises à jour et fiables, ce qui est essentiel pour les décisions basées sur des données précises.

En résumé, la gestion sécurisée de votre lakehouse et l'implémentation efficace de CI/CD sont vitales pour assurer l'intégrité, la sécurité, et l'utilité des données au sein de votre organisation. Ces stratégies permettent une gestion des données agile, sécurisée et conforme aux normes de l'industrie, tout en supportant des initiatives de données dynamiques et axées sur les résultats.

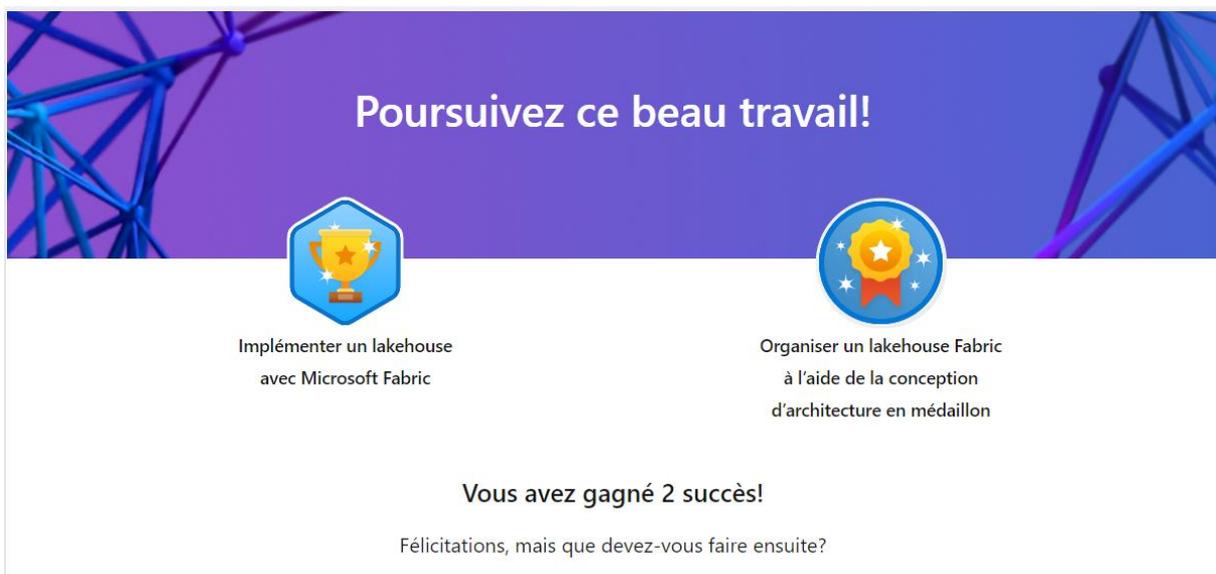
### Exercice

- Créer un espace de travail
- Créez une maison de lac et téléchargez les données sur la couche de bronze
- Transformer les données et les charger dans la table Silver Delta

- Explorez les données dans la couche Silver à l'aide du point de terminaison SQL
- Transformer les données pour la couche d'or
- Créer un modèle sémantique
- Nettoyer les ressources

## Résumé

L'architecture en médaillon structure et raffine efficacement les données dans un Lakehouse Microsoft Fabric, éliminant les silos de données et garantissant un accès facile et contrôlé pour les équipes et applications en aval. Cette configuration permet non seulement de stocker les données de manière organisée, mais aussi de les préparer pour une analyse rapide et efficace dès que nécessaire.



## Obtenir des données dans Power BI

### Introduction

Dans le cadre de son travail pour Tailwind Traders, l'employé est chargé par un responsable senior de développer une série de rapports basés sur des données provenant de divers emplacements. Les données de transactions commerciales sont stockées dans SQL Server, une base de données relationnelle détaillant les achats des clients, les articles achetés, et les employés responsables des ventes. Cependant, les informations sur l'emploi des employés, telles que la date d'embauche et la fonction, sont conservées dans des fichiers Excel par le département Ressources Humaines.

Lorsqu'un article est expédié, cela est enregistré dans une nouvelle application de gestion des entrepôts qui utilise Cosmos DB pour stocker des données sous forme de documents JSON. En outre, les projections financières, cruciales pour prévoir les futures ventes, sont conservées dans Microsoft Azure Analysis Services.

L'employé doit initialement extraire ces données des différentes sources et les intégrer à l'aide de Power Query dans Power BI, en apprenant les spécificités de chaque système et en nettoyant les données, par exemple en renommant les colonnes ou en corrigeant les erreurs. Après avoir organisé les données, l'employé peut alors créer des rapports dans Power BI et publier les résultats, permettant à d'autres d'utiliser ou de créer des rapports à partir de ces modèles.

Ce processus est essentiel pour la création de rapports efficaces et précis chez Tailwind Traders, permettant à l'entreprise de mieux comprendre et prédire ses opérations commerciales.

## Obtenir des données de fichiers

Les organisations stockent fréquemment leurs données dans des fichiers plats, tels que les fichiers CSV, TXT et les fichiers de largeur fixe, ainsi que dans des classeurs Excel. Power BI Desktop permet d'importer des données de ces fichiers à travers la fonctionnalité "Obtenir des données".

Dans un scénario professionnel chez Tailwind Traders, l'équipe RH a préparé un fichier plat avec des données importantes des employés, stocké localement ou sur des plateformes cloud comme OneDrive Entreprise ou SharePoint. Pour les données statiques, un stockage local est adéquat, alors que pour les données nécessitant une synchronisation régulière, les options cloud sont recommandées.

Lors de l'utilisation de Power BI, après avoir choisi le fichier source via l'option "Obtenir les données" sur l'onglet Accueil, les utilisateurs doivent sélectionner le fichier (par exemple, un classeur Excel), puis choisir les données spécifiques à importer. Ils peuvent directement charger les données dans leur modèle Power BI ou les transformer préalablement via Power Query.

Si l'emplacement du fichier source change, il est crucial de mettre à jour les chemins de connexion dans Power BI pour maintenir l'exactitude des rapports. Cette mise à jour peut être réalisée dans les paramètres de la source de données ou via l'Éditeur avancé dans Power Query, en s'assurant de maintenir la structure du fichier pour éviter des interruptions dans le modèle de génération de rapports.

## Obtenir des données de sources de données relationnelles

Si une organisation utilise une base de données relationnelle pour la gestion de la vente, Power BI Desktop permet de se connecter directement à cette base, offrant ainsi une surveillance directe de la progression des activités et l'identification des tendances pour la prévision des ventes, la planification budgétaire et la définition des indicateurs de performance.

Dans le cas de Tailwind Traders, l'équipe commerciale a demandé de se connecter à la base de données SQL Server locale pour obtenir les données des ventes et créer des rapports dans Power BI Desktop.

Pour se connecter à une base de données relationnelle via Power BI, il faut utiliser la fonction "Obtenir des données", choisir l'option appropriée pour la base de données concernée (ici, SQL Server) et entrer les détails du serveur ainsi que les identifiants de connexion. Les modes de connexion disponibles sont "Importer", qui est généralement recommandé, et "DirectQuery".

Une fois connecté, la fenêtre Navigateur affiche les données disponibles. L'utilisateur peut alors sélectionner les tables ou les données spécifiques à importer, en ayant la possibilité de charger les données directement ou de les transformer préalablement via Power Query pour améliorer la qualité des données.

Il est également possible d'importer des données spécifiques en rédigeant une requête SQL directement, permettant de cibler des tables et des colonnes spécifiques et de filtrer les données selon les besoins.

SQL Server database

Server 

Database (optional)

Advanced options

Command timeout in minutes (optional)

SQL statement (optional, requires database)

```
SELECT
  ID
, NAME
, SALESAMOUNT
FROM SALES
```

Include relationship columns  
 Navigate using full hierarchy  
 Enable SQL Server Failover support

Des ajustements de la source de données peuvent être nécessaires, notamment pour des raisons de sécurité ou de mise à jour des identifiants, que l'on peut gérer via les "Paramètres de source de données" dans Power Query.

L'utilisation d'instructions SQL spécifiques pour filtrer et sélectionner les données nécessaires aide à optimiser la performance de Power BI en évitant l'importation de données superflues et en utilisant des vues pour faciliter la gestion des données.

## Créer des rapports dynamiques avec des paramètres

Pour créer des rapports dynamiques dans Power BI, qui permettent aux utilisateurs de filtrer les données selon leurs besoins spécifiques, vous devez suivre plusieurs étapes essentielles, à la fois pour configurer des rapports individuels et pour gérer des filtres multiples.

### Création de rapports dynamiques pour des valeurs individuelles

1. **Connexion à la base de données:** Utilisez la fonction "Obtenir des données" de Power BI Desktop pour vous connecter à la base de données SQL Server.
2. **Rédaction et intégration de la requête SQL:** Insérez votre requête SQL dans la zone "Instruction SQL" sous "Options avancées" et établissez la connexion.
3. **Modification des données via Power Query:** Après avoir chargé les données, ouvrez-les dans l'éditeur Power Query pour préparer et nettoyer les données selon les besoins.
4. **Création et gestion des paramètres:**
  - Créez un paramètre sous l'onglet "Accueil", en le nommant de manière descriptive (par exemple, "SalesPerson").
  - Configurez le type de paramètre et les valeurs attendues.
5. **Ajustement de la requête SQL:** Modifiez la requête SQL pour inclure le paramètre, permettant ainsi le filtrage dynamique des données.
6. **Application et vérification du paramètre:** Testez le paramètre en entrant une valeur spécifique pour s'assurer que la requête se met à jour correctement.

### Création de rapports dynamiques pour plusieurs valeurs

1. **Préparation des valeurs multiples:** Créez une feuille de calcul Excel avec une liste de valeurs dans une colonne, puis connectez-vous à cette feuille via Power BI.
2. **Importation et ajustement des données dans Power Query:**
  - Importez la table de valeurs et ajustez les noms et types des colonnes pour correspondre aux besoins de votre paramètre.
  - Transformez la requête originale (Query1) en une fonction dans Power BI pour pouvoir la réutiliser avec différentes valeurs.
3. **Utilisation de la fonction personnalisée:**
  - Appliquez la fonction aux valeurs de la colonne en utilisant "Appeler une fonction personnalisée" pour générer des résultats filtrés basés sur chaque ID de vendeur.
  - Sélectionnez les colonnes à charger et ajustez les paramètres d'affichage dans votre rapport.

#### 4. Mise à jour et application des changements:

- Ajoutez ou modifiez les valeurs dans la feuille Excel au besoin, puis actualisez les données dans Power Query.
- Finalisez en appliquant les changements pour revenir à l'éditeur de rapport de Power BI.

Ces étapes permettent de créer des rapports dynamiques efficaces qui peuvent être personnalisés par les utilisateurs finaux, améliorant ainsi l'interaction avec les données et la pertinence des insights générés.

### Obtenir des données d'une base de données NoSQL

Pour intégrer des données d'une base de données NoSQL comme Azure Cosmos DB dans Power BI, suivez ces étapes essentielles pour garantir une extraction et une transformation efficaces des données :

#### Connexion à Azure Cosmos DB

1. **Utilisation de Power BI Desktop:** Commencez par ouvrir Power BI Desktop et accédez à la fonction "Obtenir des données".
2. **Sélectionner Azure Cosmos DB:** Dans les options, sélectionnez "Plus...", puis choisissez Azure Cosmos DB sous la catégorie Azure.
3. **Configuration de la connexion:**
  - **URL du point de terminaison:** Entrez l'URL du point de terminaison de votre compte Azure Cosmos DB, que vous pouvez trouver dans le volet "Clés" de votre Portail Azure.
  - **Détails de la base de données:** Saisissez le nom de la base de données et de la collection, ou utilisez le navigateur pour les sélectionner directement.
  - **Clé du compte:** Si c'est votre première connexion à ce point de terminaison, entrez la clé principale, disponible dans le volet "Clés en lecture seule" du Portail Azure.

#### Importation et traitement des fichiers JSON

1. **Visualisation des données dans le Navigateur:** Une fois connecté, la fenêtre Navigateur affichera les bases de données disponibles. Sélectionnez la base ou la collection pertinente, comme "Product" pour les données produit.
2. **Ouverture des données dans Power Query:** Les données apparaîtront initialement comme des enregistrements. Utilisez le bouton "Modifier" pour passer à Power Query.
3. **Exploitation des enregistrements:**
  - Sélectionnez le bouton "Développer" à côté de la colonne contenant les enregistrements.

- Choisissez les champs spécifiques que vous souhaitez inclure dans votre rapport et décochez "Utiliser le nom de la colonne d'origine comme préfixe" pour simplifier les noms des colonnes.
4. **Vérification et application des données:** Examinez les données transformées pour vous assurer qu'elles répondent à vos attentes, puis cliquez sur "Fermer et appliquer" pour charger les données dans Power BI Desktop.

## Utilisation des données dans Power BI

Une fois les données chargées dans Power BI Desktop, elles sont présentées sous forme de table standard avec des lignes et des colonnes. Ces données peuvent maintenant être intégrées avec d'autres sources de données et utilisées pour créer des rapports dynamiques et informatifs dans Power BI. Cela permet une analyse approfondie et la visualisation de données complexes et non structurées issues de Cosmos DB, optimisant ainsi les insights d'affaires disponibles pour les décideurs et les analystes.

## Obtenir des données de services en ligne

Pour se connecter aux données d'applications en ligne via Power BI Desktop et les intégrer à des analyses, vous suivez un processus qui tire parti de la fonctionnalité "Obtenir des données". Voici les étapes détaillées à suivre, en utilisant SharePoint comme exemple pour illustrer le processus :

### Connexion à SharePoint pour accéder aux données

#### 1. Démarrage avec Power BI Desktop:

- Ouvrez Power BI Desktop et cliquez sur "Obtenir des données".
- Sélectionnez "Services en ligne" pour accéder aux différentes applications dont les données peuvent être importées.
- Choisissez "Liste SharePoint Online" parmi les options disponibles.

#### 2. Entrer l'URL de SharePoint:

- Copiez l'URL de votre site SharePoint depuis un navigateur web et collez-la dans la fenêtre de connexion de Power BI. Assurez-vous d'utiliser l'URL de base du site sans ajouter le chemin spécifique à un fichier ou une page.
- Après avoir entré l'URL, cliquez sur OK pour continuer.

#### 3. Authentification:

- Power BI vous demandera de vous authentifier. Connectez-vous avec votre compte Microsoft associé à SharePoint et autorisez Power BI à accéder à vos données SharePoint.

### Sélection des données à importer

#### 1. Navigation et sélection des données:

- Une fois la connexion établie, la fenêtre Navigateur s'ouvrira, affichant les listes et autres entités disponibles sur votre site SharePoint.
- Sélectionnez la liste spécifique que vous souhaitez importer. Dans le cadre de cet exemple, cela pourrait être une liste contenant les nouveaux objectifs de vente pour l'équipe commerciale.

## 2. Importation des données:

- Vous avez la possibilité de charger directement les données dans votre modèle Power BI ou d'utiliser l'éditeur Power Query pour modifier et transformer les données avant de les charger.
- Pour simplifier, vous pouvez choisir l'option "Charger" si aucune transformation supplémentaire n'est nécessaire. Cela ajoutera les données directement à votre modèle Power BI.

## Utilisation des données dans Power BI

- **Création de rapports:** Une fois les données chargées, utilisez les outils de visualisation et de rapport de Power BI pour analyser les données et générer des insights. Par exemple, vous pourriez croiser les objectifs de vente avec les performances actuelles pour évaluer l'état du pipeline des ventes.
- **Partage et collaboration:** Les rapports créés peuvent être partagés avec les responsables et les membres de l'équipe pour une analyse collaborative et des prises de décision basées sur les données actualisées.

En intégrant les données de SharePoint dans Power BI, Tailwind Traders peut efficacement suivre et analyser les objectifs de vente en temps réel, optimisant ainsi la gestion des performances commerciales tout au long de l'année fiscale.

## Sélectionner un mode de stockage

Pour gérer efficacement les données et répondre aux exigences de sécurité dans Power BI, surtout dans des cas comme celui du département commercial de Tailwind Traders, il est crucial de choisir le mode de stockage approprié. Voici un résumé des trois principaux modes de stockage disponibles dans Power BI, chacun avec ses propres avantages et utilisations spécifiques :

## Modes de Stockage dans Power BI

### 1. Mode Importation

- **Description :** Ce mode implique l'importation et le stockage des données directement dans Power BI. Une fois importées, les données sont disponibles pour l'analyse sans nécessiter de connexion continue à la source de données. Cela permet une interaction rapide et efficace avec les données.
- **Utilisation :** Recommandé pour les scénarios où les données ne changent pas fréquemment ou pour les ensembles de données de taille modérée qui ne nécessitent pas de mises à jour en temps réel. Cela

convient bien pour les analyses approfondies où la latence des requêtes n'est pas critique.

## 2. Mode DirectQuery

- **Description** : Au lieu de stocker des copies de données, DirectQuery maintient une connexion en temps réel avec la source de données. Chaque interaction ou requête dans Power BI déclenche une requête correspondante à la source de données, garantissant que les données affichées sont toujours à jour.
- **Utilisation** : Idéal pour les cas où les données doivent être actualisées en continu ou lorsque les réglementations de sécurité empêchent le stockage de données en dehors de la source originale. Cela est également bénéfique pour gérer de très grands volumes de données, car seule une fraction des données est interrogée à la fois, réduisant ainsi la charge sur Power BI.

## 3. Double (Mode Composite)

- **Description** : Combine les avantages des modes Importation et DirectQuery. Dans ce mode, certaines données peuvent être importées pour une récupération rapide, tandis que d'autres restent interrogées directement selon les besoins.
- **Utilisation** : Ce mode est utile lorsque certaines parties de vos données nécessitent des mises à jour fréquentes ou un accès en temps réel et d'autres moins. Cela permet une flexibilité maximale dans la gestion des données et l'optimisation des performances.

## Application Pratique

Pour le département commercial de Tailwind Traders, le mode DirectQuery semble être le plus approprié compte tenu des exigences de sécurité et de la nécessité d'accéder aux données les plus récentes sans les importer localement. Cela garantit que les données sensibles restent sécurisées et que les performances ne sont pas entravées par la gestion de grands volumes de données.

## Configuration dans Power BI

Pour configurer le mode de stockage :

- Accédez à la **Vue Modèle** dans Power BI.
- Sélectionnez la table de données souhaitée.
- Dans le **volet Propriétés**, choisissez le mode de stockage désiré dans la liste déroulante "Mode de stockage".

En sélectionnant judicieusement le mode de stockage, vous pouvez assurer que les rapports Power BI sont non seulement performants mais aussi conformes aux normes de sécurité et de gestion des données de l'organisation.

## Obtenir des données d'Azure Analysis Services

Pour connecter Power BI à Azure Analysis Services et intégrer des données complexes dans vos rapports, suivez les étapes ci-dessous. Cela vous permettra de tirer parti des modèles sémantiques avancés et des capacités de calcul de Azure Analysis Services, en comparant ces données, par exemple, avec les données de ventes réelles pour obtenir des insights approfondis sur les performances de Tailwind Traders.

### Étapes pour se connecter à Azure Analysis Services dans Power BI

#### 1. Ouvrir Power BI Desktop:

- Commencez par lancer Power BI Desktop et accédez à la fonctionnalité "Obtenir des données".

#### 2. Sélectionner Azure Analysis Services:

- Dans la liste des sources de données disponibles, choisissez "Azure Analysis Services". Cela peut nécessiter de naviguer à travers "Plus de sources" pour trouver cette option spécifique.

#### 3. Entrer les détails de connexion:

- Vous serez invité à fournir l'adresse du serveur Azure Analysis Services ainsi que le nom de la base de données. Ces informations sont cruciales pour établir la connexion appropriée.

#### 4. Choix du mode de connexion:

- **Importer:** Ce mode permet d'importer les données dans Power BI, où elles sont stockées et manipulées localement.
- **Connexion directe (DirectQuery):** Ce mode maintient les données sur le serveur Azure Analysis Services, permettant des mises à jour en temps réel et la conservation des calculs DAX existants.

#### 5. Authentification et accès aux données:

- Authentifiez-vous auprès du serveur Azure Analysis Services pour accéder aux modèles et tables disponibles.
- Sélectionnez les tables spécifiques ou les segments de données dont vous avez besoin pour vos analyses.

#### 6. Utilisation de DAX ou MDX pour des requêtes spécifiques:

- Si nécessaire, utilisez DAX (Data Analysis Expressions) ou MDX (Multidimensional Expressions) pour créer des requêtes personnalisées directement sur le modèle dans Azure Analysis Services, surtout si vous n'avez besoin que de parties spécifiques d'une table.

#### 7. Chargement des données dans Power BI:

- Après avoir sélectionné les données nécessaires, choisissez de les charger dans Power BI (si vous utilisez le mode Importer) ou de les connecter via DirectQuery pour une interaction directe avec les données stockées dans Azure Analysis Services.

### Avantages de l'utilisation d'Azure Analysis Services avec Power BI

- **Intégrité des données:** Les modèles sémantiques dans Azure Analysis Services sont hautement optimisés pour les performances et la sécurité, offrant une source fiable pour vos rapports.
- **Actualisations en temps réel:** Avec DirectQuery, les données dans vos rapports Power BI peuvent être actualisées en temps réel dès qu'elles sont mises à jour dans Azure Analysis Services.
- **Scalabilité:** Azure Analysis Services est conçu pour gérer de grands volumes de données et des calculs complexes, ce qui est idéal pour les entreprises ayant de grandes exigences en matière de données.

### Documentation et ressources supplémentaires

Pour plus d'informations sur la configuration spécifique ou pour résoudre des problèmes de connexion, consultez la documentation officielle de Power BI sur la connexion à Azure Analysis Services. Cette ressource offre des guides détaillés, des conseils de dépannage et des meilleures pratiques pour optimiser l'utilisation de vos données.

### Résoudre les problèmes de performances

L'optimisation des performances dans Power BI, surtout lors de l'utilisation de DirectQuery pour se connecter à des bases de données comme SQL Server, est cruciale pour maintenir une expérience utilisateur fluide et efficace. Voici un résumé détaillé des techniques principales à considérer pour optimiser les performances de vos rapports, en particulier dans le contexte de votre scénario avec l'équipe des ventes.

### Optimisation des Performances avec Power Query et DirectQuery

#### 1. Query Folding

Le Query Folding est une technique essentielle dans Power Query qui permet de déléguer le traitement des données à la source de données elle-même. Cette capacité est particulièrement utile car elle minimise la quantité de données transmises entre la source de données et Power BI, améliorant ainsi les performances générales des rapports.

- **Avantages:**
  - Réduction de la charge sur le réseau et les ressources informatiques locales.
  - Performances améliorées des actualisations des données.

- Meilleure intégration avec les modes de stockage DirectQuery et Composite, où les requêtes sont exécutées directement sur le serveur source.
- **Utilisation pratique:**
  - Dans l'éditeur Power Query, effectuez des transformations de données telles que renommer des colonnes ou appliquer des filtres. Si ces transformations peuvent se traduire en une requête SQL simple (ex. **SELECT**, **GROUP BY**, etc.), le Query Folding sera utilisé.

## 2. Diagnostics de Requête

Les diagnostics de requête permettent de surveiller et d'identifier les goulots d'étranglement dans les requêtes lors du chargement et de la transformation des données.

- **Procédure:**
  - Accédez à **Outils** dans l'éditeur Power Query et utilisez les options de diagnostics pour surveiller les performances de chaque étape de transformation.
  - Analysez le temps pris par chaque opération pour identifier les étapes qui nécessitent une optimisation.

## 3. Autres Techniques d'Optimisation

En plus du Query Folding et des diagnostics de requête, d'autres stratégies peuvent améliorer les performances de vos requêtes :

- **Exécution des transformations dans la source de données:**
  - Si possible, exécutez des transformations directement sur le serveur SQL pour réduire la charge sur Power BI.
- **Utilisation de requêtes SQL natives:**
  - Pour les sources de données SQL, l'utilisation de requêtes SQL natives au lieu de passer par des procédures stockées ou des fonctions peut réduire la complexité et améliorer les performances.
- **Séparation des colonnes de date et heure:**
  - Si les colonnes combinent la date et l'heure, séparez-les avant de charger les données dans Power BI pour améliorer la compression et réduire l'utilisation de la mémoire.

## Conclusion

En suivant ces techniques, vous pouvez améliorer significativement les performances de vos rapports Power BI, surtout dans un environnement d'entreprise où les décisions rapides et fondées sur des données actualisées sont cruciales. N'oubliez pas de

consulter la documentation de Power BI pour des détails supplémentaires sur le Query Folding et les diagnostics de requête pour maximiser l'efficacité de vos rapports.

## Résoudre les erreurs d'importation de données

Lors de l'importation de données dans Power BI, vous pouvez rencontrer diverses erreurs dues à la complexité des sources de données et à la variété des systèmes impliqués. Voici un aperçu des erreurs communes et des solutions pour les résoudre efficacement :

### 1. Dépassement du délai d'expiration de la requête

- **Cause:** Les serveurs de bases de données relationnelles limitent souvent le temps d'exécution des requêtes pour éviter la monopolisation des ressources.
- **Solution:**
  - Simplifiez les requêtes en réduisant le nombre de colonnes ou de lignes extraites.
  - Utilisez des agrégations et des regroupements pour minimiser la quantité de données traitées.
  - Divisez les requêtes complexes en requêtes plus petites et réassemblez les résultats dans Power Query.

### 2. Nous n'avons trouvé aucune donnée sous forme de tableau

- **Cause:** Erreur courante lors de l'importation de données depuis Excel lorsque les données ne sont pas formatées en tableau.
- **Solution:**
  - Ouvrez Excel, sélectionnez les données que vous souhaitez importer, et formatez-les en tableau en utilisant **Ctrl+T**.
  - Assurez-vous que les en-têtes de colonnes sont correctement définis.

### 3. Fichier introuvable

- **Cause:** Cette erreur se produit généralement lorsqu'un fichier a été déplacé ou que les autorisations ont changé.
- **Solution:**
  - Utilisez Power Query pour mettre à jour le chemin d'accès au fichier. Accédez à l'éditeur Power Query, sélectionnez la requête concernée, et ajustez les paramètres de la source pour pointer vers le nouvel emplacement du fichier.

### 4. Erreurs de type de données

- **Cause:** Les colonnes apparaissent vides dans Power BI lorsque le type de données n'est pas correctement interprété.

- **Solution:**
  - Convertissez explicitement les types de données dans votre requête SQL pour correspondre aux attentes de Power BI. Par exemple, convertissez les codes postaux en chaînes de caractères si Power BI ne parvient pas à interpréter le format d'origine.

```
sql
```

```
SELECT CAST(CustomerPostalCode AS varchar(10)) FROM Sales.Customers
```

### Conseils Généraux pour la Gestion des Erreurs

- **Documentation et Ressources** : Pour toute erreur spécifique non couverte ici, consultez la documentation Microsoft Power BI pour des conseils détaillés sur le dépannage.
- **Optimisation des Requêtes** : Améliorez la performance des requêtes en optimisant les requêtes SQL utilisées pour l'importation des données.
- **Validation des Données** : Vérifiez et validez les formats de données avant l'importation pour réduire les erreurs de type de données.

En suivant ces recommandations, vous pouvez améliorer la fiabilité de vos processus d'importation de données dans Power BI et réduire les interruptions causées par des erreurs communes.

## Exercice

Labo

## Résumé

Dans ce module sur Power BI, vous avez appris à intégrer des données provenant de diverses sources, chacune avec ses spécificités. Voici un résumé des points clés abordés et des meilleures pratiques pour utiliser Power BI de manière efficace :

### 1. Extraction des données de multiples sources

- **Fichiers** : Pour des fichiers comme Excel, les données doivent être formatées en tableaux pour que Power BI puisse les extraire efficacement.
- **Bases de données relationnelles** : Ces sources peuvent présenter des défis, tels que des délais d'expiration des requêtes, nécessitant souvent des optimisations de requête pour éviter les interruptions.
- **Azure Analysis Services** : La connexion peut être établie via Direct Query pour une actualisation en temps réel des données, permettant aux utilisateurs de voir les changements dès qu'ils se produisent.

### 2. Choix du mode de stockage

- **Importer** : Ce mode est idéal si vous avez besoin d'une interaction rapide avec les visuels et que les données n'ont pas besoin d'être actualisées fréquemment par rapport à la source originale.
- **Direct Query** : Choisissez ce mode si les mises à jour en temps réel des données sont cruciales, bien que cela puisse affecter les performances de l'interactivité.

### 3. Résolution des problèmes de performances et des erreurs d'importation

- **Outils de diagnostic** : Power BI fournit des outils pour identifier et résoudre les problèmes de performance, permettant une optimisation efficace des rapports.
- **Erreurs d'importation** : Bien que les erreurs d'importation puissent sembler intimidantes, elles sont souvent résolues par des ajustements simples, comme la reconfiguration des paramètres de la source de données ou la modification des requêtes pour respecter les limites de temps.

### Conclusion

Les compétences acquises dans ce module sont essentielles pour exploiter pleinement les capacités de Power BI, permettant de créer des rapports dynamiques et réactifs basés sur une grande variété de sources de données. En sélectionnant judicieusement le mode de stockage et en comprenant comment traiter et optimiser les données de différentes sources, vous pouvez améliorer significativement l'efficacité et la performance de vos solutions d'analyse de données.

# Poursuivez ce beau travail!



Obtenir des données dans  
Power BI

## Vous avez gagné un succès!

Félicitations, mais que devez-vous faire ensuite?

# Nettoyer, transformer et charger les données dans PowerBI

## Introduction

Dans un scénario d'importation de données dans Power BI, plusieurs problèmes peuvent rendre les données inadaptées à l'analyse. Les principales difficultés identifiées comprennent une colonne "Employment status" ne contenant que des chiffres, la présence d'erreurs dans plusieurs colonnes, des valeurs nulles, des doublons d'ID clients, et une colonne d'adresse combinant plusieurs informations. Ces problèmes conduisent à des erreurs dans les visuels et les rapports, notamment les totaux de ventes.

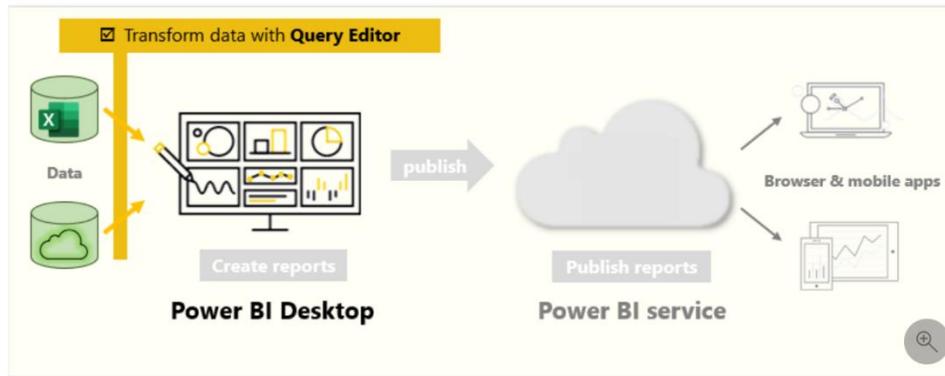
Pour résoudre ces problèmes, Power BI et Power Query offrent des outils pour nettoyer et préparer les données. Les bénéfices de travailler avec des données "propres" incluent des mesures et colonnes plus précises pour les calculs et agrégations, une meilleure organisation des tables, l'élimination des doublons, et la simplification des colonnes complexes. Ces améliorations permettent une meilleure navigation, des filtres efficaces, et une lisibilité accrue des données.

Le module traitera des méthodes pour résoudre les incohérences et les valeurs nulles, appliquer des remplacements de valeurs, profiler les colonnes pour mieux comprendre leur contenu, évaluer et transformer les types de données, ajuster les structures des tables, combiner des requêtes, appliquer des conventions de nommage aux colonnes et requêtes, et modifier le code M dans l'éditeur avancé.

## Mettre en forme les données initiales

L'éditeur Power Query de Power BI Desktop permet de transformer les données importées en effectuant des actions telles que renommer des colonnes ou des tables, convertir du texte en nombres, supprimer des lignes, et définir la première ligne comme en-tête. Cette mise en forme est cruciale pour garantir que les données sont adaptées à l'utilisation dans des rapports.

Les données de ventes brutes ont été chargées depuis deux sources distinctes : un fichier .csv créé manuellement dans Excel par l'équipe des ventes et une connexion au système ERP de l'organisation. Ces données apparaissent désordonnées dans Power BI Desktop, nécessitant une intervention via l'éditeur Power Query pour nettoyer et organiser les données correctement.



### Démarrer avec Power Query :

1. **Ouvrir l'éditeur** : Accessible via l'option 'Transformer des données' sous l'onglet 'Accueil'.
2. **Interface de Power Query** : Affichage central des données, volet de requêtes à gauche, et un historique des étapes de transformation à droite.

### Principales étapes de mise en forme :

- **Identifier et ajuster les en-têtes** : Utiliser la première ligne comme en-tête si nécessaire et corriger les noms des colonnes inappropriés.
- **Renommer des colonnes** : Modifier les en-têtes pour plus de clarté et de cohérence.
- **Supprimer les lignes et colonnes superflues** : Eliminer les données non nécessaires pour alléger le modèle et améliorer la performance.
- **Dépivoter des colonnes** : Transformer les colonnes pour simplifier la création de mesures DAX, organisant les données en colonnes telles que 'Month', 'Year', et 'SalesAmount'.
- **Pivoter des colonnes** : Réorganiser les données plates en une structure permettant de mieux identifier des modèles, en utilisant des fonctions d'agrégation.

**Finaliser les transformations** : Après avoir réalisé les modifications nécessaires, l'utilisateur doit sélectionner 'Fermer et appliquer' pour appliquer les changements au modèle sémantique de Power BI. Toutes les transformations sont enregistrées et appliquées automatiquement à chaque actualisation de la source de données, garantissant une cohérence continue sans altérer les données d'origine.

### Simplifier la structure des données

Dans Power BI Desktop, lors de l'importation de données depuis diverses sources, il est courant de rencontrer des noms de table et de colonne prédéfinis qui peuvent être peu clairs ou incohérents. L'éditeur Power Query permet de renommer ces éléments pour améliorer leur compréhensibilité et leur utilisation dans les rapports.

### Processus de simplification des données:

1. **Renommer les requêtes (tables):** Modifier les noms des requêtes pour les rendre plus explicites. Par exemple, renommer une table de faits de produits de "FactProductTable" en "Products" pour une meilleure clarté.
2. **Renommer une requête en fonction du contexte:** Ajouter des détails spécifiques, comme l'année à un nom de requête, pour éviter les confusions entre des données similaires mais distinctes d'une année sur l'autre.

#### Améliorer la qualité des données:

- **Remplacer des valeurs incorrectes:** Corriger les erreurs orthographiques ou les valeurs incorrectes dans les colonnes en utilisant la fonctionnalité "Remplacer les valeurs".
- **Gérer les valeurs Null:** Remplacer les valeurs Null par des valeurs significatives, telles que zéro, pour assurer des calculs précis dans les rapports.

#### Éliminer les redondances:

- **Supprimer les doublons:** Utiliser la fonctionnalité "Supprimer les doublons" pour ne conserver que des valeurs uniques dans une colonne, permettant de créer des tables plus efficaces pour le modèle sémantique.

#### Bonnes pratiques de nommage:

- Adopter des conventions de nommage qui utilisent un langage commun et des termes métier descriptifs.
- Préférer des termes simples et descriptifs pour les tables et les colonnes, éviter les préfixes non nécessaires et choisir des abréviations cohérentes.
- Remplacer les traits de soulignement par des espaces pour faciliter la lecture.

Ces modifications peuvent être effectuées facilement dans l'éditeur Power Query, où chaque action est enregistrée et appliquée systématiquement à chaque mise à jour des données, sans altérer les données sources originales. Une fois toutes les étapes nécessaires complétées, il suffit de sélectionner "Fermer et appliquer" pour intégrer les modifications dans le modèle de données de Power BI.

## Évaluer et changer les types de données des colonnes

Lors de l'importation de données dans Power BI Desktop, il est crucial d'évaluer et de corriger les types de données des colonnes pour assurer des performances optimales et éviter des erreurs dans les rapports. Power BI analyse initialement les premières lignes pour déduire les types de données, mais cette détection peut être incorrecte, particulièrement avec les fichiers plats comme les fichiers CSV ou les classeurs Excel.

#### Processus d'évaluation et de modification des types de données:

1. **Évaluation des types de données :** Examiner les colonnes pour s'assurer que les types attribués par Power BI correspondent à la nature des données. Par

exemple, vérifier qu'une colonne contenant des dates n'est pas incorrectement marquée comme texte.

2. **Changer le type de données** : Modifier le type de données dans l'éditeur Power Query de Power BI pour corriger les erreurs de détection. Ce changement est crucial car un type de données incorrect peut limiter les calculs, la création de hiérarchies temporelles, et les relations entre tables.

#### Implications d'un type de données incorrect:

- **Limitations dans les calculs** : Un type incorrect peut empêcher l'exécution de calculs spécifiques, comme les agrégations temporelles nécessaires pour des mesures telles que le cumul annuel des commandes.
- **Problèmes de hiérarchie** : Sans le bon type de données, il est impossible de créer des hiérarchies de dates, ce qui est essentiel pour l'analyse temporelle des données.

#### Méthodes pour changer le type de données dans Power Query:

- **Via le ruban Transformer** : Sélectionner la colonne, puis choisir le type de données approprié dans l'option Type de données du ruban.
- **Via l'icône de type de données** : Cliquer sur l'icône à côté de l'en-tête de la colonne et sélectionner le type correct dans la liste déroulante.

Changer le type de données dans Power Query est préférable avant le chargement des données dans le modèle sémantique pour assurer que les modifications soient appliquées de manière cohérente à chaque actualisation des données.

Après avoir effectué ces changements, il est important de sélectionner "Fermer et appliquer" pour intégrer les modifications dans le modèle de données de Power BI, préparant ainsi les données pour des analyses et des rapports précis et performants. Pour plus d'approfondissements, il est conseillé de consulter la documentation sur les types de données dans Power BI Desktop.

## Combiner plusieurs tables en une seule table

Dans Power BI, combiner des requêtes est une fonctionnalité puissante qui permet de simplifier et d'unifier les données en provenance de différentes sources, idéale pour créer un modèle sémantique plus compact et facile à naviguer. Cela est particulièrement utile lorsque vous gérez des tables multiples qui présentent des similarités fonctionnelles ou sont trop fragmentées.

#### Scénarios de combinaison de requêtes:

- **Fusion de requêtes** : Lorsque vous fusionnez des requêtes, vous combinez des colonnes de différentes tables. Cela nécessite une colonne clé commune pour lier les tables.
- **Ajout de requêtes** : Cette méthode consiste à ajouter des lignes de données d'une table à une autre, augmentant le nombre total de lignes sans modifier le nombre de colonnes.

**Exemple d'application:** Pour un rapport consolidé sur les informations de contact des employés, fournisseurs et clients, vous pourriez avoir besoin de combiner des informations provenant de différentes tables telles que HR.Employees, Production.Suppliers, et Sales.Customers. Chacune de ces tables devrait être préparée pour qu'elles aient des colonnes identiquement nommées (par exemple, ID, company, name, phone) pour permettre une vue consolidée.

### Processus de combinaison des requêtes:

1. **Préparation des tables** : Avant de combiner, formatez chaque table pour qu'elle contienne uniquement les colonnes pertinentes et que ces colonnes soient nommées de manière cohérente.
2. **Ajouter des requêtes** : Sous l'onglet Accueil de Power Query, utilisez l'option Ajouter des requêtes pour combiner les lignes des tables sélectionnées. Cela peut créer une nouvelle requête qui rassemble toutes les informations nécessaires.
3. **Fusionner des requêtes** : Pour les scénarios nécessitant la consolidation de détails reliés par une clé commune (comme des commandes et des détails de commande), utilisez l'option Fusionner des requêtes. Sélectionnez les tables à combiner et la colonne clé commune. Vous pouvez choisir le type de jointure (externe gauche, externe entière, interne) selon les besoins de vos données.

### Avantages de la combinaison des requêtes:

- **Réduction de la complexité** : Diminue le nombre de tables dans le modèle, simplifiant la navigation et la maintenance.
- **Vue unifiée** : Crée une source unique de vérité pour tous les rapports et analyses, évitant les incohérences et doublons.

En somme, combiner des requêtes dans Power BI permet d'optimiser la structure du modèle de données et de préparer le terrain pour des analyses plus efficaces et des rapports plus intégrés. Pour des instructions détaillées, il est conseillé de consulter la documentation officielle de Power BI sur la mise en forme et la combinaison des données.

## Profiler les données dans Power BI

Le profilage des données dans Power BI est une pratique essentielle pour comprendre en profondeur les caractéristiques et la qualité des données avant de les utiliser pour des analyses ou des rapports. Cela implique l'analyse des structures des données, la détection d'anomalies, et la visualisation de statistiques détaillées sur les données.

### Étapes clés du profilage des données dans Power BI:

#### 1. Examiner les structures des données

- Avant toute manipulation, il est crucial de comprendre comment les données sont organisées dans le modèle sémantique. Ceci peut être réalisé en visualisant le modèle sous l'onglet "Modèle" dans Power BI Desktop.

- Utilisez les fonctionnalités du ruban pour ajuster les propriétés des tables et des colonnes, et pour gérer les relations entre différentes tables.

## 2. Identifier les anomalies et les statistiques des données

- Dans l'éditeur Power Query, explorez les données pour identifier toute valeur aberrante ou anomalie qui pourrait affecter l'analyse.
- Utilisez la fonctionnalité "Distribution des colonnes" pour observer la répartition des valeurs dans chaque colonne, permettant de détecter des distributions anormales ou des valeurs extrêmes.

## 3. Analyse détaillée avec Profil de colonne

- Activez les options "Qualité de la colonne" et "Profil de colonne" pour obtenir des statistiques détaillées comme le nombre de lignes, les valeurs minimales et maximales, et les distributions des valeurs.
- Ce profilage aide à vérifier l'intégrité des données importées et à identifier les valeurs extrêmes qui pourraient nécessiter une investigation plus poussée.

## 4. Utilisation de la Distribution des valeurs

- Visualisez combien de fois certaines valeurs apparaissent dans une colonne, ce qui est crucial pour identifier les valeurs dominantes ou rares, qui peuvent être des indicateurs d'erreurs ou de particularités dans les processus d'affaires.

## 5. Statistiques de colonnes numériques

- Pour les colonnes numériques, examinez des statistiques supplémentaires telles que la moyenne, l'écart type, et le nombre de valeurs zéro ou nulles, fournissant une compréhension plus profonde de la distribution des données.

## 6. Gestion des erreurs et des valeurs atypiques

- L'identification d'erreurs potentielles ou de données atypiques peut conduire à des actions correctives telles que la mise à jour des données, garantissant l'exactitude des rapports et des analyses.

### Application pratique:

Supposez que vous préparez des rapports pour une équipe de ventes et que vous remarquez des anomalies dans les valeurs reportées par certains commerciaux. Le profilage des données vous permet de détecter ces anomalies et de prendre les mesures correctives nécessaires avant de finaliser les rapports.

### Conclusion:

Le profilage des données est une étape indispensable pour garantir l'exactitude et la fiabilité des analyses réalisées avec Power BI. Il vous aide à comprendre la structure, la qualité et la distribution de vos données, ce qui est crucial pour développer des rapports précis et fiables. En utilisant les outils intégrés dans Power BI, vous pouvez non seulement détecter et corriger les erreurs mais aussi optimiser l'utilisation des données pour des insights plus approfondis.

## Utiliser l'éditeur avancé pour modifier le code M

L'utilisation de l'éditeur avancé dans Power Query vous permet de manipuler directement le code M, ce qui est utile pour des ajustements fins ou des modifications qui ne sont pas facilement réalisables via l'interface graphique. Voici un guide succinct sur comment utiliser et modifier le code M dans l'éditeur avancé de Power Query dans Power BI.

### Accéder à l'Éditeur Avancé

1. **Ouvrir Power Query:** Lancez Power Query à partir de Power BI Desktop.
2. **Ruban Affichage:** Allez dans l'onglet Affichage sur le ruban.
3. **Éditeur Avancé:** Cliquez sur le bouton Éditeur Avancé pour ouvrir l'interface où le code M est visible.

### Comprendre le Code M

- **Structure du Code:** Le code M dans l'éditeur est structuré en séquences d'étapes, où chaque étape représente une transformation ou un nettoyage de données. Chaque étape est généralement une ou deux lignes de code.
- **Modification Simple:** Pour des modifications simples, comme changer le nom d'une base de données ou une valeur spécifique, localisez la ligne correspondante et apportez les modifications nécessaires.
- **Sauvegarde des Modifications:** Après avoir fait vos modifications, cliquez sur "Terminé" pour les appliquer.

### Bonnes Pratiques de Modification du Code M

- **Prudence avec les Dépendances:** Soyez vigilant lorsque vous réorganisez les étapes dans le code M. Chaque étape peut dépendre des résultats des étapes précédentes, donc réarranger les étapes peut entraîner des erreurs.
- **Expérimentation Contrôlée:** N'hésitez pas à expérimenter avec le code M pour apprendre son fonctionnement. Cependant, assurez-vous de tester les modifications dans un environnement contrôlé pour éviter de perturber les flux de travail existants.
- **Références par Noms:** Les étapes sont souvent référencées par le nom de la variable à gauche du signe égal. La dernière étape de la requête, utilisée dans l'instruction `in`, est généralement celle qui génère le résultat final du jeu de données.

### Exemple de Code M

Voici un exemple simplifié de ce que pourrait ressembler le code M dans l'éditeur avancé :

### Utilisations Avancées

## Sales Orders

Display Options ?

```
let
    Source = Sql.Database("localhost", "tsqlv4"),
    Sales_Orders = Source[[Schema="Sales",Item="Orders"]][Data],
    #"Split Column by Delimiter" = Table.SplitColumn(Sales_Orders, "shipaddress", Splitter.SplitTextByDelimiter(",", QuoteStyle.Csv), {"shipa
    #"Changed Type" = Table.TransformColumnTypes(#"Split Column by Delimiter",{{"shipaddress.1", type text}, {"shipaddress.2", type text}})
in
    #"Changed Type"
```

✓ No syntax errors have been detected.

Done Cancel

- **Intégration avec d'autres sources:** Vous pouvez ajouter des étapes pour intégrer des données provenant de différentes sources.
- **Fonctions personnalisées:** Le code M permet de créer des fonctions personnalisées pour réutiliser dans plusieurs transformations de données.
- **Optimisation des performances:** Modifier le code M peut aider à optimiser les performances en réduisant les étapes superflues ou en améliorant les méthodes de transformation.

En maîtrisant l'éditeur avancé et le code M, vous pouvez considérablement augmenter la puissance et l'efficacité de vos transformations de données dans Power BI, vous permettant de personnaliser profondément la manière dont les données sont préparées pour vos analyses.

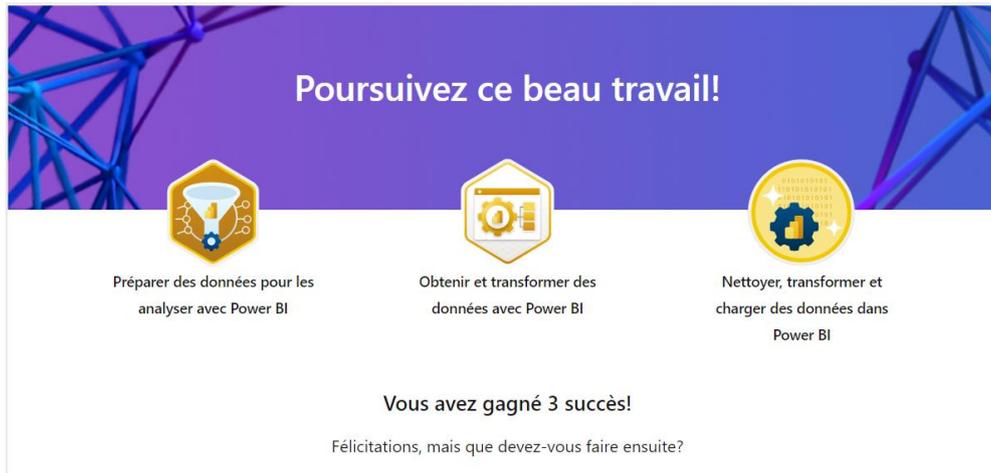
## Exercice

<https://learn.microsoft.com/fr-ch/training/modules/clean-data-power-bi/8-lab>

## Résumé

Ce module vous a montré comment transformer des données complexes en informations plus accessibles et faciles à analyser pour les utilisateurs et les créateurs de rapports. Vous avez appris à effectuer des calculs, à combiner des requêtes pour minimiser leur nombre et simplifier l'exploration des données, ainsi qu'à renommer les

colonnes de manière à les rendre compréhensibles. De plus, le module a couvert les meilleures pratiques de nommage pour les éléments dans Power BI.



## Concevoir un modèle sémantique dans Power BI

### Introduction

Dans ce scénario, les données importées dans Power BI depuis différentes sources ne sont pas préparées pour l'analyse en raison de plusieurs problèmes identifiés. Les principales difficultés incluent des colonnes contenant uniquement des nombres pour l'état d'emploi, des erreurs dans plusieurs colonnes, des valeurs Null, des duplications d'ID client, et des adresses combinées en une seule colonne. Ces problèmes entraînent des résultats incorrects et des rapports erronés lors de la création de visuels.

Pour remédier à ces défis, Power BI et Power Query offrent des outils pour nettoyer et préparer les données. Le nettoyage des données permet d'améliorer la précision des résultats lors des calculs, d'organiser les tables pour une navigation intuitive, de supprimer les doublons, de fractionner les colonnes complexes en colonnes plus simples, et de remplacer les codes et entiers par des valeurs lisibles.

Le module enseigne comment résoudre les incohérences, les valeurs inattendues ou Null, et les problèmes de qualité des données. Il couvre également le profilage des données pour mieux comprendre les colonnes, l'évaluation et la transformation des types de données, les transformations de la structure des tables, la combinaison des requêtes, l'application de conventions de nommage aux colonnes et requêtes, et la modification du code M dans l'éditeur avancé.

### Utiliser des tables

Quand un modèle sémantique contient de nombreuses tables importées de diverses sources, il devient complexe et difficile à utiliser. Pour simplifier ce modèle avant de créer des rapports, il est essentiel de structurer les tables de manière simple et intuitive. Une structure de table simple doit être facile à naviguer avec des propriétés de table et de colonne conviviales, comprendre des tables fusionnées pour simplifier

la structure des données, et présenter des relations logiques et de qualité entre les tables.

Pour configurer le modèle sémantique et établir des relations entre les tables dans Power BI, on utilise l'onglet Modèle pour visualiser et gérer les relations, par exemple entre les tables Order et Sales via la colonne OrderDate. La gestion des relations permet de créer, modifier, et supprimer des relations, et la fonctionnalité de détection automatique aide à identifier les relations entre les colonnes similaires.

En plus de configurer les relations, il est possible de personnaliser les propriétés des tables et des colonnes dans la vue Modèle. Ces propriétés incluent le changement de noms de colonnes, l'ajout de synonymes, l'organisation des colonnes dans des dossiers, et la mise en forme des données, comme le format des dates. La boîte de dialogue des propriétés des colonnes offre diverses options pour mettre à jour ces propriétés, facilitant ainsi l'organisation et la gestion des tables dans le modèle sémantique.

Ces étapes sont cruciales pour préparer le modèle de données avant de commencer la création de visuels, afin d'éviter des complications ultérieures et de garantir que les rapports soient basés sur une structure de données bien organisée et facile à comprendre.

## Créer une table de dates

Dans le cadre de la création de rapports dans Power BI, il est crucial de correctement mettre en forme les valeurs de dates et de temps, surtout quand les rapports se basent sur l'analyse temporelle comme les évolutions mensuelles, trimestrielles ou annuelles. Parfois, Power BI reconnaît automatiquement les colonnes de dates, mais des ajustements supplémentaires peuvent être nécessaires pour répondre aux besoins spécifiques de l'entreprise.

Par exemple, en travaillant sur des rapports pour l'équipe commerciale, on peut observer que les tables de ventes et de commandes possèdent leurs propres colonnes de dates. Pour créer un rapport qui compile les totaux de ventes et de commandes par année et par mois, il est essentiel de créer une table de dates commune utilisable par plusieurs tables.

### Créer une table de dates commune:

1. **Données sources:** Utiliser une table de dates préexistante dans la base de données si elle est disponible. Cela simplifie le processus car cette table est souvent déjà optimisée pour les besoins de l'entreprise, incluant la distinction entre années civiles et fiscales, et les jours de fermeture de l'entreprise.
2. **DAX:** Utiliser des fonctions comme **CALENDAR()** ou **CALENDARAUTO()** pour générer une table de dates. **CALENDAR()** permet de créer une table avec une plage de dates spécifiée, tandis que **CALENDARAUTO()** ajuste automatiquement la plage en fonction des données existantes dans le modèle.

DAX

```
Dates = CALENDAR(DATE(2011, 5, 31), DATE(2022, 12, 31))
```

```
1 Dates = CALENDAR(DATE(2011, 5, 31), DATE(2022, 12, 31))
```

3. **Power Query:** Créer une table de dates en utilisant le langage M, ce qui offre flexibilité pour définir la plage de dates et le niveau de granularité (jour, heure, etc.).

M

Copier

```
= List.Dates(#date(2011,05,31), 365*10, #duration(1,0,0,0))
```

fx

```
= List.Dates(  
#date(2011,05,31), //Start Date  
365*10, //Dates for every day for the next 10 years  
#duration(1,0,0,0) //Specifies duration of the period 1 = days, 0 = hours, 0 = minutes, 0 = seconds  
)
```

Une fois la table de dates commune créée, elle peut être intégrée dans le modèle sémantique de Power BI. Il est possible de marquer cette table comme la table de dates officielle, ce qui assure que Power BI l'utilise pour toutes les analyses temporelles. Cette table peut ensuite être liée aux autres tables de données, comme les ventes et les commandes, pour permettre des analyses cohérentes sur le temps.

Enfin, pour créer des visuels efficaces, il est important de bien configurer les relations entre la nouvelle table de dates et les autres tables concernées. Cela permet de réaliser des analyses complexes et de visualiser des données temporelles précises, telles que le total des ventes et des commandes par période.

## Utiliser des dimensions

La création d'un schéma en étoile dans Power BI utilise des tables de dimension et de faits, où les tables de faits contiennent des informations sur des événements tels que les commandes clients et les dates d'expédition, tandis que les tables de dimension stockent des détails sur des entités comme les produits ou l'heure. Ces tables sont connectées par des relations, facilitant l'analyse structurée et le reporting.

**Hiérarchies dans Power BI:** Les hiérarchies permettent de structurer et de naviguer dans les données plus facilement, en permettant une exploration détaillée au lieu d'une vue globale. Power BI peut générer automatiquement des hiérarchies pour les types de données comme les dates, ou vous pouvez créer des hiérarchies manuellement pour d'autres catégories, telles que les produits. Ces hiérarchies sont particulièrement utiles dans les visuels, permettant aux utilisateurs de percer les données de niveau supérieur jusqu'aux détails.

**Hiérarchie Parent-Enfant:** Une hiérarchie parent-enfant est utilisée pour montrer des relations dans des structures organisationnelles ou des séries de commandes. Par

exemple, dans une table d'employés, les responsables (parents) et les subordonnés (enfants) peuvent être représentés, permettant une visualisation des structures managériales. Power BI peut utiliser la fonction DAX **PATH()** pour tracer le chemin hiérarchique de chaque employé, avec **PATHITEM()** pour extraire les différents niveaux de cette hiérarchie, permettant une visualisation de chaque niveau de manière aplatie.

**Dimensions de Rôle Actif:** Les dimensions de rôle actif jouent un rôle crucial quand une table de dimension est connectée à plusieurs tables de faits. Par exemple, une table Calendar peut être connectée à la fois aux tables Sales et Order, permettant d'utiliser cette dimension pour regrouper et filtrer les données de différentes manières. Cela montre la polyvalence d'une table de dimension dans le filtrage et l'agrégation des données à travers plusieurs tables de faits.

En résumé, les tables de dimension et de faits, les hiérarchies, et les dimensions de rôle actif sont des composants essentiels du schéma en étoile dans Power BI, facilitant l'analyse complexe et rendant les données plus accessibles et compréhensibles pour les utilisateurs. Ces techniques avancées permettent une manipulation plus efficace des données et une meilleure intégration dans le processus décisionnel.

## Définir la précision des données

La précision des données dans Power BI se réfère à leur niveau de détail, et ajuster cette précision est crucial pour améliorer la performance et l'efficacité d'utilisation des rapports et des visuels. Dans un scénario où une entreprise gère 1 000 semi-remorques réfrigérés, avec des données de température recueillies à intervalles réguliers, la gestion de la précision des données peut aider à simplifier le modèle sémantique pour faciliter son exploitation.

**Optimisation de la précision des données:** Pour réduire la quantité de données manipulées, on peut ajuster la précision des données de diverses manières :

- **Agrégation quotidienne:** Calculer une moyenne quotidienne pour chaque camion réduit le nombre d'enregistrements et peut être suffisant pour le suivi des coûts et des erreurs.
- **Sélection de données spécifiques:** Importer uniquement les enregistrements qui excèdent ou sont inférieurs à certaines valeurs, ou choisir la dernière température enregistrée, peut aussi réduire la quantité de données tout en préservant les informations essentielles.

Dans d'autres contextes, il peut être approprié de définir la précision des données sur une base hebdomadaire, mensuelle ou trimestrielle, ce qui accélère les performances des rapports et permet des actualisations plus fréquentes du modèle sémantique.

**Conséquences de la réduction de la précision:** Une précision réduite peut limiter la capacité des utilisateurs à explorer les détails de chaque transaction, ce qui peut affecter négativement leur expérience. Il est donc crucial de discuter du niveau approprié de précision avec les utilisateurs pour s'assurer qu'ils comprennent les compromis impliqués.

**Ajustement de la précision pour les relations entre tables:** La précision des données influence également la création de relations entre les tables. Par exemple, pour créer des rapports pour une équipe commerciale en utilisant les tables Calendar, Sales et Budget, il est nécessaire de réconcilier les différents niveaux de précision des tables Sales (journalière) et Budget (mensuelle). Cela peut être accompli en transformant les colonnes Year et Month de chaque table en une nouvelle colonne qui standardise le format de la date, permettant ainsi d'établir une relation.

**Établissement de relations et création de visuels:** Une fois la relation établie, des mesures DAX peuvent être utilisées pour calculer des agrégats tels que le total des ventes et le montant du budget. Ces mesures peuvent ensuite être incorporées dans un visuel de type matrice pour comparer les ventes et les budgets sur le temps.

En somme, la gestion de la précision des données est une stratégie clé pour optimiser les performances et l'utilité des rapports dans Power BI, nécessitant un équilibre entre la granularité des données et les besoins de performance et d'analyse des utilisateurs.

## Relations et cardinalité

Power BI se distingue en intégrant la direction des relations, essentielle pour le filtrage des données entre plusieurs tables. Ce système cherche automatiquement les relations en correspondant les noms de colonnes lors du chargement des données, bien que ces relations puissent aussi être ajustées manuellement.

### Types de Relations dans Power BI:

#### 1. Relation Plusieurs-à-Un / Un-à-Plusieurs (:1/1:):

- C'est la relation la plus courante où plusieurs instances d'une valeur dans une colonne correspondent à une seule instance dans une autre.
- Utilisée principalement entre les tables de dimensions et de faits, cette relation permet un filtrage efficace où les flèches indiquent généralement de la dimension (plusieurs) vers le fait (un).

#### 2. Relation Un-à-Un (1:1):

- Chaque instance dans une table correspond exactement à une instance dans l'autre table.
- Souvent considérée comme redondante, cette relation peut indiquer une conception inefficace du modèle. Il est habituellement recommandé de fusionner les tables concernées.

#### 3. Relation Plusieurs-à-Plusieurs (.):

- Plusieurs valeurs dans une table peuvent correspondre à plusieurs valeurs dans une autre, sans exigence de valeurs uniques.
- Cette configuration est déconseillée en raison de son potentiel d'ambiguïté et de complexité accrue dans le filtrage et l'interprétation des données.

**Direction du Filtre Croisé:**

- **Direction Unique:**
  - Permet à une table de filtrer l'autre sans réciprocité. Cette configuration est typique des relations un-à-plusieurs où la table de dimension (le côté "un") filtre la table de faits (le côté "plusieurs").
- **Filtrage Croisé Bidirectionnel:**
  - Les deux tables peuvent se filtrer mutuellement, utile pour des analyses complexes mais peut réduire les performances et introduire des ambiguïtés si mal gérée.

**Cardinalité et Direction du Filtre:**

- Les relations un-à-un offrent toujours un filtrage croisé bidirectionnel, car elles connectent des valeurs distinctes et précises entre les tables.
- Les relations plusieurs-à-plusieurs permettent un choix entre le filtrage unidirectionnel et bidirectionnel, mais leur complexité nécessite une gestion prudente pour éviter des résultats inattendus ou incorrects.

**Création d'une Relation Plusieurs-à-Plusieurs:**

- En pratique, cette relation est utile dans des scénarios comme les analyses de budget où plusieurs clients peuvent être associés à plusieurs comptes.
- Lors de la création, il est crucial de ne pas s'attendre à des valeurs uniques, et le choix du filtrage croisé bidirectionnel doit être bien réfléchi pour éviter des résultats imprévisibles.

La gestion et la compréhension des relations dans Power BI est fondamentale pour construire des modèles de données robustes et performants, tout en évitant les pièges de configurations inappropriées qui pourraient nuire à la clarté et à l'efficacité des analyses.

## Résoudre les problèmes liés à la modélisation

La modélisation des données dans Power BI implique la structuration de relations entre les tables pour permettre une visualisation efficace des données selon les besoins de votre entreprise. Cependant, un défi majeur peut survenir lors de la création de ces relations : les relations circulaires. Ces relations peuvent complexifier la gestion du modèle et rendre la création de visuels plus difficile en obscurcissant la compréhension des relations existantes entre les tables.

**Comprendre les Relations Circulaires**

**Définition:** Une relation circulaire se produit lorsque trois tables (ou plus) sont reliées de manière que la dernière table se connecte à nouveau à la première, créant ainsi une boucle. Par exemple, si la Table 1 a une relation plusieurs-à-un avec la Table 2,

la Table 2 a une relation un-à-plusieurs avec la Table 3, et la Table 3 a ensuite une relation qui la relie de nouveau à la Table 1, nous avons une relation circulaire.

### Problèmes causés par les relations circulaires:

- **Ambiguïté dans la direction du filtrage** : Cela peut rendre le filtrage des données ambigu, car Power BI peut avoir du mal à déterminer dans quelle direction appliquer le filtre dans un modèle circulaire.
- **Performance** : Les relations circulaires peuvent également affecter négativement les performances du modèle en créant des chemins de calcul complexes et inutiles.
- **Complexité accrue** : Elles compliquent la gestion du modèle et la compréhension des dépendances entre les données, rendant ainsi les rapports plus susceptibles à des erreurs.

### Dépendances des Relations

Pour comprendre les dépendances dans les relations circulaires, il est utile de considérer comment les changements dans une table peuvent affecter d'autres tables dans le modèle. Par exemple, si une modification est apportée à la table **dCustomer**, cela peut entraîner une modification dans **fSales**, qui à son tour pourrait affecter **dSalesPerson**. Cela démontre une chaîne de dépendances où chaque table est influencée par les changements dans les autres.

**Exemple de dépendance de colonnes:** La colonne **CoûtTotal** dans la table **Ventes** est définie par une formule qui multiplie **Quantité** par **Prix**. Cela illustre une dépendance directe où le **CoûtTotal** est affecté par toute modification des valeurs de **Quantité** ou **Prix**.

### Solutions pour Gérer les Relations Circulaires

#### 1. Restructurer le Modèle :

- **Réexaminer les relations** : Assurez-vous que chaque relation est nécessaire et logique. Envisagez de supprimer ou de réarranger certaines relations pour éliminer les boucles.
- **Fusionner les tables** : Si possible, fusionnez les tables qui ont des relations un-à-un pour simplifier le modèle.

#### 2. Utilisation de tables intermédiaires :

- Parfois, l'introduction d'une table de jonction ou intermédiaire peut aider à simplifier les relations en décomposant une relation complexe en deux relations plus simples.

#### 3. Consultation et formation :

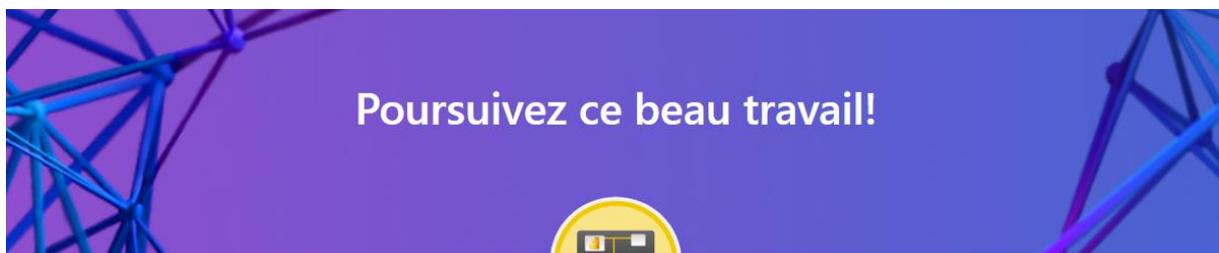
- Consulter des experts en modélisation de données ou suivre des formations spécifiques sur la conception de modèles de données peut fournir des stratégies supplémentaires pour éviter les pièges communs comme les relations circulaires.

## Exercice

<https://learn.microsoft.com/fr-ch/training/modules/design-model-power-bi/8-lab>

## Résumé

Dans Power BI, la modélisation de données englobe des aspects cruciaux tels que la création de tables de dates communes, la configuration de relations complexes comme les relations plusieurs-à-plusieurs, la résolution des problèmes de relations circulaires, et la conception de schémas en étoile. Maîtriser ces techniques est vital pour tout utilisateur de Power BI afin de faciliter la création de visuels et la communication efficace des rapports aux équipes concernées. Comprendre ces éléments enrichit votre approche de la modélisation sémantique et vous aide à naviguer dans les détails techniques avec plus d'aisance.



Concevoir un modèle  
sémantique dans Power BI

**Vous avez gagné un succès!**

Félicitations, mais que devez-vous faire ensuite?

## Ajouter des mesures aux modèles Power BI Desktop

### Introduction

Dans Power BI, les mesures sont classifiées en deux types : implicites et explicites. Les mesures implicites sont des résumés automatiques réalisés par Power BI lors de l'utilisation de colonnes numériques dans les visuels, identifiables par le symbole sigma ( $\Sigma$ ). Elles permettent des agrégations simples comme la somme, la moyenne, le minimum, etc., sans qu'il soit nécessaire de définir explicitement un calcul. Ces mesures offrent une grande flexibilité pour commencer l'analyse rapidement et sont simples à utiliser, mais elles sont limitées aux fonctions d'agrégation prédéfinies et ne conviennent pas pour des analyses plus complexes qui nécessitent des calculs personnalisés via DAX.

D'autre part, les mesures explicites, ou simplement les mesures, sont des calculs définis par l'utilisateur qui peuvent inclure des opérations complexes non disponibles dans les mesures implicites. Ces mesures sont essentielles pour des besoins

spécifiques d'analyse et doivent être utilisées pour des calculs avancés ou lorsque des fonctions d'agrégation personnalisées sont nécessaires.

Les colonnes non numériques supportent également des agrégations, mais sont limitées à des opérations telles que le comptage et l'identification du premier ou dernier élément par ordre alphabétique.

Un inconvénient notable des mesures implicites est que leur utilisation peut parfois conduire à des résultats trompeurs si l'agrégation n'est pas appropriée pour le contexte des données, comme sommer des prix unitaires qui devraient plutôt être moyennés. De plus, les mesures implicites ne sont pas reconnues lorsque le modèle est interrogé via des expressions multidimensionnelles (MDX), ce qui peut limiter leur utilité dans certaines applications avancées comme l'analyse dans Excel ou les rapports paginés de Power BI.

En résumé, bien que les mesures implicites facilitent l'analyse rapide et simple, les mesures explicites sont indispensables pour des analyses détaillées et personnalisées, nécessitant souvent l'usage de DAX pour des calculs précis et contextuels.

## Créer des mesures simples

Dans Power BI, la création de mesures, qu'elles soient implicites ou explicites, est essentielle pour le résumé et l'analyse des données dans vos modèles. Voici un résumé des points clés abordés sur l'utilisation des mesures dans Power BI :

### 1. Mesures Implicites vs. Explicites :

- Les mesures **implicites** sont automatiquement générées par Power BI pour les colonnes numériques marquées avec un symbole sigma ( $\Sigma$ ) et permettent des agrégations basiques telles que la somme, la moyenne, etc.
- Les mesures **explicites** sont définies manuellement par l'utilisateur via des formules DAX et sont utilisées pour des calculs plus complexes qui ne peuvent être réalisés automatiquement par Power BI.

### 2. Création de Mesures Explicites :

- Pour ajouter une mesure explicite, sélectionnez la table concernée, utilisez l'option "Nouvelle mesure" du ruban, et entrez une formule DAX pour définir le calcul, par exemple, **Revenue = SUM(Sales[Sales Amount])**.
- Il est possible de formater les mesures pour afficher les données de manière précise, en spécifiant le nombre de décimales ou d'autres paramètres de formatage.

### 3. Fonctionnalités des Formules DAX :

- DAX permet de créer des mesures pour sommer, calculer le coût total, le profit, la quantité, et les prix (minimum, maximum, et moyen) des produits.
- Des fonctions telles que **COUNT** et **DISTINCTCOUNT** permettent de compter les occurrences dans une colonne, ce qui est utile pour des données telles que les commandes ou les lignes de commande.

#### 4. Gestion et Contrôle des Mesures :

- Vous pouvez masquer des colonnes pour restreindre leur utilisation dans des résumés et forcer l'utilisation des mesures créées, renforçant ainsi l'intégrité des analyses effectuées.
- Il est recommandé de bien configurer les mesures dès leur création pour assurer une présentation cohérente des données dans tous les rapports.

#### 5. Avantages des Mesures Explicites :

- Offrent une flexibilité et une précision plus grandes dans le traitement et l'analyse des données.
- Nécessaires pour des calculs spécifiques qui ne sont pas pris en charge par les agrégations automatiques.

#### 6. Limitations des Mesures Implicites :

- Ne peuvent pas effectuer des calculs complexes nécessaires pour des analyses plus détaillées.
- Ne sont pas reconnues dans les requêtes MDX, ce qui limite leur utilisation dans certains contextes avancés comme les rapports paginés ou l'analyse dans Excel.

En résumé, bien que les mesures implicites soient utiles pour des analyses rapides et simples, les mesures explicites offrent une puissance et une flexibilité indispensables pour des analyses approfondies et personnalisées dans Power BI.

### Créer des mesures composées

Dans Power BI, utiliser des mesures composées est une technique efficace pour simplifier et optimiser les modèles sémantiques. Une mesure composée fait référence à d'autres mesures déjà définies dans le modèle pour effectuer des calculs complexes. Ce type de mesure est très utile pour réduire la redondance des calculs et améliorer la performance du modèle.

#### Exemple de Mesure Composée

##### Définition de la Mesure "Profit" :

DAXCopy code

Profit = [Revenue] - [Cost]

Dans cet exemple, la mesure "Profit" est calculée en soustrayant la mesure "Cost" de la mesure "Revenue". Chacune de ces mesures sous-jacentes est probablement déjà calculée ailleurs dans le modèle, ce qui permet de simplifier la gestion et la mise à jour des définitions de mesure.

## Avantages de l'Utilisation des Mesures Composées

### 1. Simplification des Modèles :

- Les mesures composées permettent de réduire le nombre de calculs stockés directement dans le modèle, ce qui simplifie la structure du modèle et facilite sa compréhension et sa gestion.

### 2. Optimisation des Performances :

- En réduisant le nombre de colonnes calculées, comme la suppression de "Profit Amount", on diminue la taille globale du modèle. Cela peut conduire à une réduction des temps de chargement et d'actualisation des données.

### 3. Flexibilité :

- Les mesures composées offrent une flexibilité accrue pour les analyses. Si les composantes sous-jacentes ("Revenue" ou "Cost") sont modifiées, la mesure "Profit" sera automatiquement mise à jour pour refléter ces changements, assurant ainsi que les analyses restent actuelles et précises.

## Implémentation dans un Visuel Matrice

Après avoir créé la mesure "Profit", vous pouvez l'ajouter à un visuel matrice dans votre rapport Power BI. Ce visuel montrera le profit pour différentes dimensions, telles que le temps, la géographie, ou les segments de clientèle, en fonction de la manière dont vous configurez le visuel.

## Conseils pour la Gestion des Mesures

- **Formatage** : Assurez-vous de formater les mesures correctement pour faciliter la lecture et la compréhension des rapports. Pour "Profit", vous pourriez choisir de formater avec deux décimales pour maintenir la cohérence avec les autres mesures financières.
- **Documentation** : Gardez une documentation claire sur la logique derrière chaque mesure, particulièrement pour les mesures composées qui dépendent d'autres mesures. Cela aide à éviter la confusion et les erreurs dans les rapports complexes.

En résumé, l'utilisation de mesures composées dans Power BI permet d'optimiser les modèles sémantiques tout en améliorant la performance et la flexibilité des analyses. Cette approche réduit la charge de travail nécessaire pour maintenir le modèle et augmente l'efficacité des processus d'analyse.

## Créer des mesures rapides

Power BI Desktop offre une fonction très utile pour simplifier la création de mesures complexes : les Mesures rapides. Cette fonctionnalité génère automatiquement l'expression DAX nécessaire pour des calculs courants, facilitant ainsi l'apprentissage de DAX et accélérant le développement de mesures pour les analyses.

### Utilisation des Mesures Rapides pour Créer une Mesure Composée

#### Exemple de Création de la Mesure de Marge Bénéficiaire :

##### 1. Sélection de la Mesure Rapide :

- Allez dans le volet Champs et sélectionnez la table *Sales*.
- Dans le ruban, sous l'onglet Outils de table, cliquez sur Mesure rapide, ce qui ouvre une fenêtre vous permettant de choisir le type de calcul.

##### 2. Configuration du Calcul :

- Dans la fenêtre Mesures rapides, parcourez les options disponibles jusqu'à trouver et sélectionner *Division* sous le groupe *Opérations mathématiques*.
- Définissez la mesure *Profit* comme numérateur et *Revenue* comme dénominateur pour calculer la marge bénéficiaire.

##### 3. Génération Automatique de DAX :

- Power BI génère automatiquement le DAX nécessaire :

DAXCopy code

Profit divided by Revenue = DIVIDE([Profit], [Revenue])

- Cette formule utilise la fonction **DIVIDE**, qui est plus sécurisée que la simple division, car elle gère les cas où le dénominateur pourrait être zéro.

##### 4. Finalisation de la Mesure :

- Renommez la mesure en *Profit Margin*.
- Formatez la mesure pour afficher les résultats en pourcentage avec deux décimales.

##### 5. Ajout de la Mesure au Visuel Matrice :

- Intégrez la nouvelle mesure de marge bénéficiaire dans un visuel matrice pour visualiser l'efficacité financière par différentes dimensions (comme le temps, la géographie, etc.).

### Avantages des Mesures Rapides

- **Efficacité** : Les mesures rapides accélèrent le processus de création de mesures en fournissant des templates de calculs courants, ce qui est particulièrement utile pour les utilisateurs moins familiers avec DAX.
- **Apprentissage** : Elles offrent une excellente opportunité pour les utilisateurs d'apprendre et de comprendre le DAX en observant comment les formules sont construites automatiquement.
- **Personnalisation** : Après la création, les mesures peuvent être personnalisées et étendues selon les besoins spécifiques.

### Limitations

- **Complexité Limitée** : Bien que les mesures rapides couvrent de nombreux scénarios de calcul courants, elles peuvent ne pas suffire pour des besoins très spécifiques ou extrêmement complexes qui nécessitent des ajustements manuels ou des développements DAX plus avancés.

En somme, les mesures rapides dans Power BI simplifient la création de mesures complexes et éduquent les utilisateurs sur l'utilisation efficace de DAX, rendant les analyses plus accessibles et personnalisables.

### Comparer des colonnes calculées à des mesures

Les débutants en DAX ressentent souvent une certaine confusion au sujet des colonnes calculées et des mesures. La section suivante passe en revue les similitudes et les différences entre les deux. En ce qui concerne les similarités entre les colonnes calculées et les mesures, les deux sont : Des calculs que vous pouvez ajouter à votre modèle sémantique. Définies à l'aide d'une formule DAX. Référencées dans des formules DAX en encadrant leurs noms entre crochets. Les domaines dans lesquels les colonnes calculées et les mesures diffèrent sont les suivants : Objectif : les colonnes calculées étendent une table avec une nouvelle colonne, tandis que les mesures définissent comment résumer les données du modèle. Évaluation : les colonnes calculées sont évaluées à l'aide du contexte de ligne au moment de l'actualisation des données, tandis que les mesures sont évaluées à l'aide du contexte de filtre au moment de la requête. Le contexte de filtre est présenté dans un module ultérieur ; il s'agit d'un sujet important à comprendre et à maîtriser pour vous permettre d'obtenir des résumés complexes. Stockage : les colonnes calculées (dans les tables du mode de stockage Importer) stockent une valeur pour chaque ligne de la table, mais une mesure ne stocke jamais de valeurs dans le modèle. Vous pouvez utiliser des colonnes calculées (comme n'importe quelle colonne) pour filtrer, regrouper ou résumer (comme mesure implicite), tandis que les mesures sont conçues pour résumer.

ChatGPT

### ChatGPT

Les concepts de colonnes calculées et de mesures en DAX sont essentiels pour une modélisation efficace des données dans Power BI. Bien qu'ils partagent certaines

similarités, ils servent des fonctions distinctes et ont des impacts différents sur les modèles de données. Voici un résumé clarifiant leurs similitudes et différences :

### Similitudes entre Colonnes Calculées et Mesures

1. **Définies par DAX** : Les deux utilisent des formules DAX pour leur définition.
2. **Calculs** : Les deux ajoutent des calculs personnalisés au modèle sémantique.
3. **Syntaxe** : Dans les formules DAX, les noms des colonnes calculées et des mesures sont encadrés par des crochets, par exemple **[NomColonne]** ou **[NomMesure]**.

### Différences entre Colonnes Calculées et Mesures

#### 1. Objectif :

- **Colonnes Calculées** : Elles ajoutent une nouvelle colonne à une table existante, ce qui étend les données de la table avec de nouvelles valeurs calculées.
- **Mesures** : Elles sont utilisées pour résumer ou calculer des données à travers une ou plusieurs tables, souvent à des fins d'analyse ou de reporting.

#### 2. Moment d'Évaluation :

- **Colonnes Calculées** : Elles sont calculées et stockées lors de l'actualisation des données, opérant dans un contexte de ligne où chaque calcul est effectué en référence aux autres valeurs de la même ligne.
- **Mesures** : Elles sont évaluées dynamiquement à la requête, utilisant un contexte de filtre qui peut varier selon la façon dont les données sont interrogées ou visualisées dans le rapport.

#### 3. Stockage :

- **Colonnes Calculées** : Elles sont stockées dans le modèle, avec une valeur calculée pour chaque ligne de la table.
- **Mesures** : Elles ne stockent aucune donnée; elles calculent les résultats à la volée en fonction des interactions de l'utilisateur avec le rapport.

#### 4. Utilisation :

- **Colonnes Calculées** : Peuvent être utilisées comme n'importe quelle autre colonne pour filtrer, trier ou comme base de calculs supplémentaires. Elles peuvent aussi être utilisées implicitement pour résumer les données.
- **Mesures** : Principalement utilisées pour créer des résumés dynamiques qui s'adaptent selon les filtres appliqués et sont souvent utilisées dans

les visualisations pour montrer des totaux, des moyennes, des minimums, des maximums, etc.

### Points Clés à Retenir

- Les **colonnes calculées** sont mieux adaptées lorsque vous avez besoin de valeurs qui doivent être recalculées avec chaque mise à jour des données et sont dépendantes du contexte de ligne.
- Les **mesures** sont préférables pour les analyses dynamiques où le besoin est de résumer ou d'analyser des données en fonction de différents filtres et segments sans nécessiter de stockage de valeurs calculées.

Comprendre ces différences aide à optimiser la performance du modèle Power BI et à utiliser les ressources de manière plus efficace, tout en évitant des recalculs inutiles et en assurant des réponses rapides aux requêtes des utilisateurs.

### Exercice

<https://learn.microsoft.com/fr-ch/training/modules/dax-power-bi-add-measures/5b-lab>

### Résumé

Dans ce module, vous avez appris que les mesures Power BI sont implicites ou explicites. Les mesures implicites sont des comportements automatiques qui sont pris en charge par les visuels, tandis que les mesures explicites utilisent des formules DAX qui résument les données du modèle.

Les mesures explicites sont importantes, car elles vous permettent de créer des formules DAX complexes pour obtenir des calculs précis dont vos visuels de rapport ont besoin. Bien que vous ayez appris à créer des mesures simples et composées dans ce module, dans les modules ultérieurs vous apprendrez à créer des mesures plus puissantes en utilisant des fonctions de modification de filtre et des fonctions d'itérateur.



Ajouter des mesures aux  
modèles Power BI Desktop

**Vous avez gagné un succès!**

Félicitations, mais que devez-vous faire ensuite?

# Ajouter des tables et des colonnes calculées aux modèles Power BI Desktop

## Introduction

Dans Power BI, il est possible d'ajouter une table calculée au modèle grâce à une formule Data Analysis Expressions (DAX). Cette table calculée peut dupliquer ou transformer des données existantes pour générer une nouvelle table, toutefois, elle ne peut se connecter à des sources de données externes. Une table calculée a pour contrainte d'augmenter la taille de stockage du modèle et peut rallonger les temps d'actualisation car elle nécessite un recalcul lors des mises à jour de données.

## Processus de Création d'une Table Calculée

- Duplication de la Table:** Pour illustrer un cas d'utilisation courant, le fichier Adventure Works DW 2020 M03.pbix est utilisé pour créer une table calculée. Dans ce modèle, la table "Sales" a trois relations avec la table "Date", mais seulement une est active. Les relations inactives entre ces tables doivent être supprimées pour éviter des conflits dans le filtrage des données.
- Création de la Table "Ship Date":**
  - Dans la vue Rapport, une nouvelle table nommée "Ship Date" est créée en utilisant la commande "Nouvelle table" sous l'onglet Modélisation.
  - La table "Ship Date" est une duplication de la table "Date", conservant les mêmes colonnes et lignes. Elle est synchronisée avec les mises à jour de la table "Date".
  - Une relation est ensuite établie entre "ShipDateKey" de la table "Sales" et "DateKey" de la table "Ship Date".

## Configuration Supplémentaire

- Renommage des Colonnes:** Les colonnes de la table "Ship Date" sont renommées pour refléter leur usage spécifique, par exemple "Fiscal Year" devient "Ship Fiscal Year".
- Tri et Masquage:** Les colonnes sont triées et certaines, comme "MonthKey", sont masquées pour optimiser l'affichage et l'usage.
- Création de Hiérarchies:** Une hiérarchie nommée "Fiscal" est créée pour organiser les données de manière plus intuitive.

## Création d'une Seconde Table Calculée: "Due Date"

- Utilisation de CALENDARAUTO:** Cette fonction DAX génère une table de dates couvrant toutes les dates pertinentes dans le modèle, basée sur les dates les plus anciennes et les plus récentes trouvées dans les données.

- **Configuration de la Table "Due Date":** Après sa création, des colonnes supplémentaires sont ajoutées pour soutenir des filtrages et groupements spécifiques, et la table est marquée comme table de dates.

## Importance des Tables Calculées

Les tables calculées facilitent la gestion de relations multiples et permettent d'ajouter des tables de dates essentielles pour utiliser la fonctionnalité "Time Intelligence" de Power BI, ce qui améliore considérablement la capacité d'analyse des données temporelles.

## Créer des colonnes calculées

Pour améliorer l'analyse des données dans Power BI, des colonnes calculées peuvent être ajoutées à une table du modèle. Celles-ci doivent retourner une valeur scalaire et peuvent augmenter la taille du modèle ainsi que les délais d'actualisation, car elles sont recalculées à chaque mise à jour des données dépendantes.

## Création de Colonnes Calculées dans la Table "Due Date"

1. **Renommage de la Colonne Date:** Avant de débiter, la colonne "Date" dans la table "Due Date" est renommée en "Due Date".
2. **Ajout de Colonnes Calculées:**
  - **Due Fiscal Year:** Une colonne calculée pour identifier l'année fiscale basée sur la date d'échéance. La formule concatène "FY" avec l'année, en ajoutant 1 si le mois est après juin pour aligner avec l'année fiscale qui commence en juillet.
  - **Due Fiscal Quarter:** Cette colonne détermine le trimestre fiscal en fonction du mois de la date d'échéance. Les trimestres sont définis avec des numéros qui reflètent le début de l'année fiscale en juillet.
  - **Due Month:** Formate la date d'échéance pour afficher l'année et le mois en format texte.
  - **Due Full Date:** Formate la date d'échéance pour afficher l'année, le mois et le jour.
  - **MonthKey:** Calcule une clé numérique en combinant l'année et le mois de la date d'échéance, utile pour le tri chronologique des dates.

## Vérification et Configuration Finale de la Table "Due Date"

- **Vérification des Colonnes:** S'assurer que toutes les nouvelles colonnes calculées sont correctement ajoutées et visibles dans la vue des données.
- **Configuration du Tri et du Masquage:**
  - Les colonnes "Due Full Date" et "Due Month" sont triées par la colonne "MonthKey" pour maintenir l'ordre chronologique.
  - La colonne "MonthKey" est masquée pour ne pas encombrer l'affichage.

- **Création de Hiérarchies:** Une hiérarchie nommée "Fiscal" est créée pour organiser les données fiscales par année, trimestre, mois et date complète.
- **Marquage de la Table Comme Table de Dates:** Essentiel pour utiliser la table dans des fonctions de Time Intelligence, la colonne "Due Date" est utilisée pour marquer la table comme table de dates.

### Utilité des Colonnes Calculées

Les colonnes calculées enrichissent les données en permettant des analyses plus fines, comme les comparaisons d'année fiscale ou le suivi des échéances par trimestre, mois, et date précise. Elles sont cruciales pour les rapports dynamiques et les analyses temporelles dans Power BI.

### En savoir plus sur le contexte de ligne

Le concept de contexte de ligne est essentiel dans l'évaluation des formules de colonnes calculées dans Power BI. Chaque formule est évaluée individuellement pour chaque ligne de la table, utilisant uniquement les données présentes dans cette ligne spécifique. Par exemple, la colonne calculée "Due Fiscal Year" utilise la date de chaque ligne pour déterminer l'année fiscale correspondante.

### Interaction entre Tables dans les Formules de Colonnes Calculées

#### 1. Utilisation de Fonctions Relatives aux Relations:

- **RELATED et RELATEDTABLE:** Ces fonctions sont utilisées pour accéder aux données d'autres tables liées. **RELATED** est utilisé pour récupérer des données d'une relation un-à-un ou un-à-plusieurs (de la partie "un" de la relation), tandis que **RELATEDTABLE** renvoie un ensemble de données de la partie "plusieurs" de la relation.
- **Performance:** Utiliser **RELATED** est préférable en termes de performance car elle exploite les indices des relations entre les tables.

#### 2. LOOKUPVALUE pour Tables Non Associées:

Si aucune relation directe n'existe, **LOOKUPVALUE** peut être employé pour chercher des valeurs à travers les tables, similaire à une recherche VLOOKUP dans Excel.

### Exemple de Contexte de Ligne dans la Pratique

Dans la table "Sales" (Ventes), la colonne calculée "Discount Amount" illustre l'utilisation du contexte de ligne. Pour chaque transaction, la quantité commandée et le montant des ventes sont multipliés et ajustés en fonction du prix listé du produit (récupéré via **RELATED**), montrant comment les valeurs dépendent des données de la ligne courante ainsi que des relations avec d'autres tables.

### Utilisation des Fonctions d'Itérateur

Les fonctions d'itérateur, qui seront explorées plus en détail dans un module ultérieur, permettent de créer des résumés avancés en traitant chaque ligne individuellement à travers le contexte de ligne, ce qui est crucial pour des calculs précis sur des ensembles de données complexes.

En résumé, le contexte de ligne dans Power BI est un pilier pour comprendre comment les données sont traitées dans les formules de colonnes calculées, permettant des analyses détaillées et précises au niveau de chaque enregistrement.

## Choisir une technique pour ajouter une colonne

Dans Power BI, il existe trois techniques principales pour ajouter des colonnes à un modèle de table, chacune adaptée à différents contextes et besoins :

### 1. Ajout de Colonnes à une Vue ou une Table (Colonne Persistante):

- **Utilisation:** Idéale quand la source de données est une base de données relationnelle et que l'utilisateur a les compétences et autorisations nécessaires.
- **Avantages:** Facilite la maintenance et permet la réutilisation de la logique de colonne dans d'autres modèles ou rapports.
- **Méthode:** La colonne est ajoutée directement dans la source de données (base de données), souvent via un outil de gestion de base de données ou un script SQL.

### 2. Ajout de Colonnes Personnalisées via Power Query (M):

- **Utilisation:** Recommandée pour transformer ou ajouter des données avant qu'elles ne soient chargées dans le modèle de données.
- **Avantages:** Permet une manipulation avancée et complexe des données avec le langage M. Les colonnes ajoutées via Power Query sont optimisées et compactes dans le modèle.
- **Méthode:** Utilisation de Power Query pour ajouter des transformations ou des colonnes personnalisées directement dans les requêtes avant le chargement des données.

### 3. Ajout de Colonnes Calculées avec DAX:

- **Utilisation:** Meilleure option quand il est nécessaire d'utiliser des fonctions de modélisation qui ne sont disponibles que dans DAX, telles que **RELATED** et **RELATEDTABLE**.
- **Avantages:** DAX est puissant pour des calculs dynamiques basés sur les données déjà chargées dans le modèle.
- **Méthode:** Les colonnes sont ajoutées directement dans le modèle de données de Power BI à l'aide de formules DAX.

## Quand Utiliser Chaque Technique ?

- **Colonnes Persistantes:** Quand vous pouvez modifier la structure de la base de données source. Cela est souvent géré par des administrateurs de bases de données ou des développeurs ayant un accès direct aux bases de données.

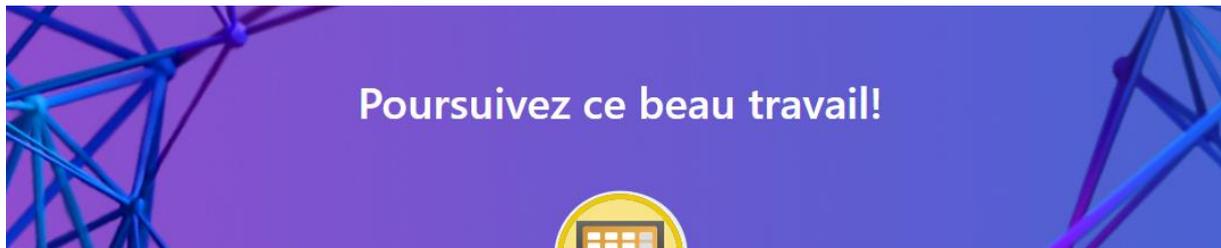
- **Colonnes via Power Query (M):** Privilégiez cette méthode pour les transformations de données avant le chargement, surtout si les transformations sont complexes ou si la performance du modèle est une priorité.
- **Colonnes Calculées (DAX):** Utilisez cette méthode pour les besoins spécifiques de calcul ou de relation qui ne peuvent être résolus qu'après que les données sont chargées dans le modèle.

### Recommandations Générales

- Les utilisateurs finaux des rapports ne distinguent généralement pas la méthode utilisée pour ajouter des colonnes, tant que les données sont correctes et utiles.
- Il est souvent conseillé d'utiliser Power Query pour l'ajout de colonnes lorsque cela est possible, en raison de l'optimisation de la performance du chargement des données.
- Réservez l'utilisation de colonnes calculées DAX pour des cas spécifiques où des fonctions avancées sont nécessaires, notamment pour des calculs qui dépendent d'autres données du modèle ou pour l'implémentation de hiérarchies complexes.

### Résumé

Dans ce module, vous avez exploré l'utilisation de DAX pour enrichir vos modèles avec des tables et des colonnes calculées. Vous avez compris que les colonnes calculées sont analysées selon le contexte de chaque ligne et appris à lier des colonnes de différentes tables. De plus, il a été souligné que, bien qu'il existe plusieurs méthodes pour ajouter des colonnes, l'ajout de colonnes personnalisées via Power Query est souvent la méthode la plus optimale.



Ajouter des tables et des  
colonnes calculées aux  
modèles Power BI Desktop

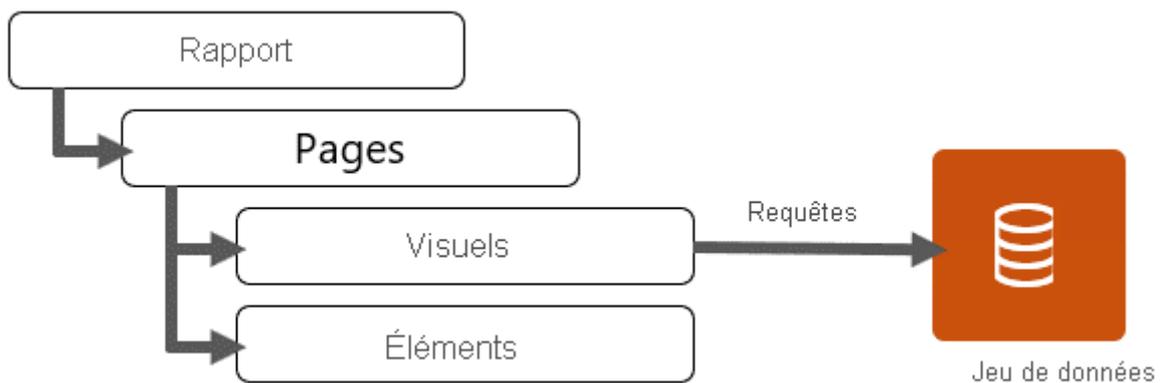
**Vous avez gagné un succès!**

Pour enregistrer vos succès et vos progrès, connectez-vous ou créez un profil.

# Concevoir des rapports Power BI

## Introduction

Microsoft Power BI permet la conception de rapports soit via Power BI Desktop (uniquement sur Windows) soit via le service Power BI en ligne, les applications mobiles étant réservées à la visualisation. Les rapports sont structurés autour d'un modèle sémantique unique avec au moins une page de rapport, bien que plusieurs pages soient courantes. Ces pages peuvent contenir des visualisations des données ou des éléments de design comme des zones de texte et des images.



La gestion des pages est flexible, permettant l'ajout, la duplication, ou la suppression de pages, similaire à la manipulation de feuilles dans Microsoft Excel. Dupliquer des pages facilite le développement, mais il est conseillé de ne pas surcharger les rapports avec des pages redondantes. La sécurité et la publication des rapports sont centralisées, sans possibilité de sécuriser ou publier des pages de manière indépendante. Les options de mise en forme des pages incluent des ajustements de taille et d'arrière-plan, et la navigation entre les pages peut être configurée avec des boutons ou des extraits interactifs.

## Concevoir la présentation des rapports analytiques

La conception de rapports analytiques doit adhérer aux principes fondamentaux de la conception graphique pour faciliter la compréhension automatique et inconsciente des données par les consommateurs. Un rapport peut comprendre plusieurs pages, chacune contenant des visuels de données et des éléments décoratifs. La conception débute par la définition du nombre, de la séquence et de l'objectif des pages, en évitant de mélanger des sujets contradictoires sur une même page.

Une présentation efficace doit respecter les principes de placement, d'équilibre, de contraste, de proximité et de répétition, en suivant la maxime "less is more" pour la simplicité et la clarté. Les objets doivent être placés de manière à créer une hiérarchie visuelle claire, souvent de gauche à droite pour les audiences occidentales, ou de droite à gauche pour les langues comme l'arabe ou l'hébreu. L'alignement et le

groupement logique des objets améliorent l'esthétique générale et l'organisation de la page.

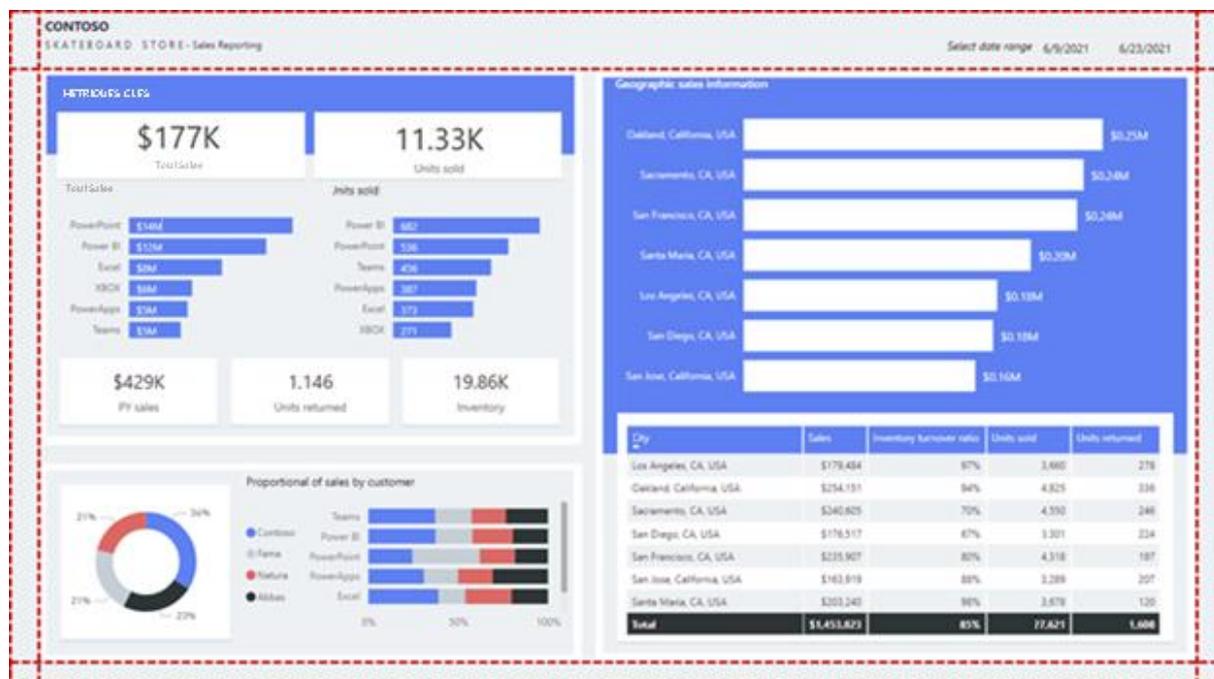
Le placement peut aussi suivre la règle des tiers, divisant l'espace en neuf parties égales pour une distribution harmonieuse des éléments. L'équilibre, qu'il soit symétrique ou asymétrique, contribue à la stabilité visuelle, souvent guidé par le nombre d'or pour un agencement harmonieux. La proximité entre les éléments aide à clarifier les relations entre les données, tandis que le contraste souligne les éléments clés. La répétition renforce la cohérence et aide à l'interprétation rapide des données, en liant visuellement les éléments associés.

## Concevoir des rapports agréables à l'œil

La conception de rapports agréables à l'œil est essentielle pour faciliter la compréhension rapide des données par les consommateurs. Un bon rapport combine un contenu bien organisé avec une esthétique soignée qui guide l'utilisateur à travers les informations importantes sans effort. Voici les principaux éléments à considérer lors de la conception de rapports :

### Espace

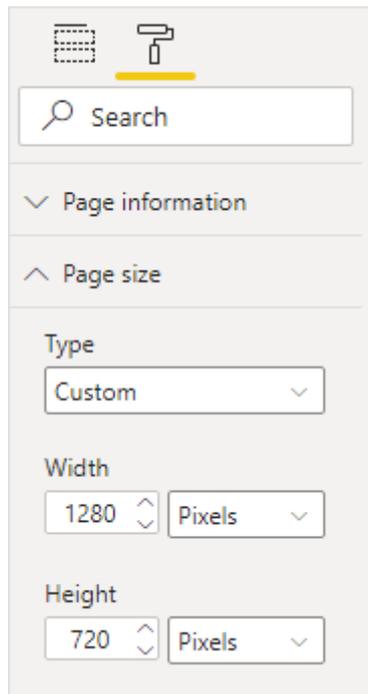
L'utilisation efficace de l'espace, y compris les marges et les intervalles entre les objets de rapport, est cruciale pour éviter un rapport encombré et améliorer la lisibilité. Les marges doivent être uniformes autour des pages pour encadrer les objets du rapport, et un espacement suffisant doit être maintenu entre les objets pour permettre une interaction facile et éviter le chevauchement.



### Taille

La taille des éléments, tant pour les visuels que pour les pages de rapport, doit être adaptée pour une visualisation optimale. Les visuels importants peuvent être agrandis

pour attirer l'attention, tandis que des formats de page personnalisés peuvent être nécessaires pour des visualisations complexes nécessitant des barres de défilement.

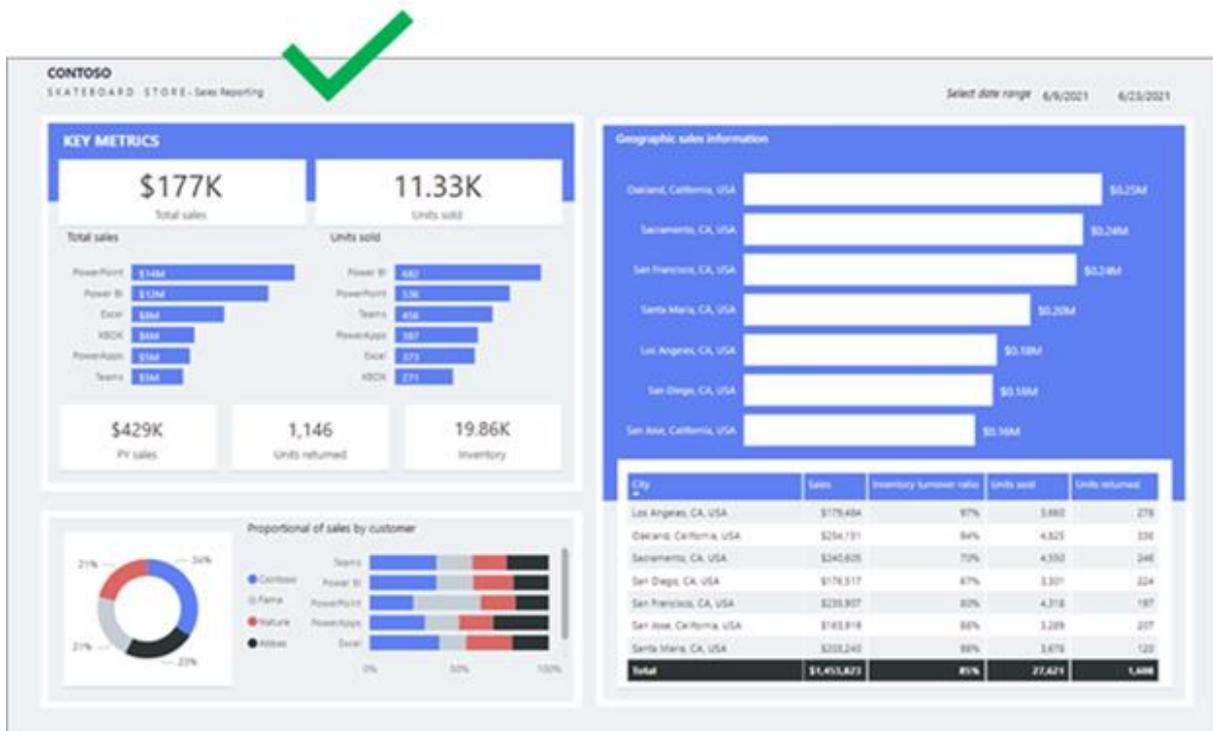
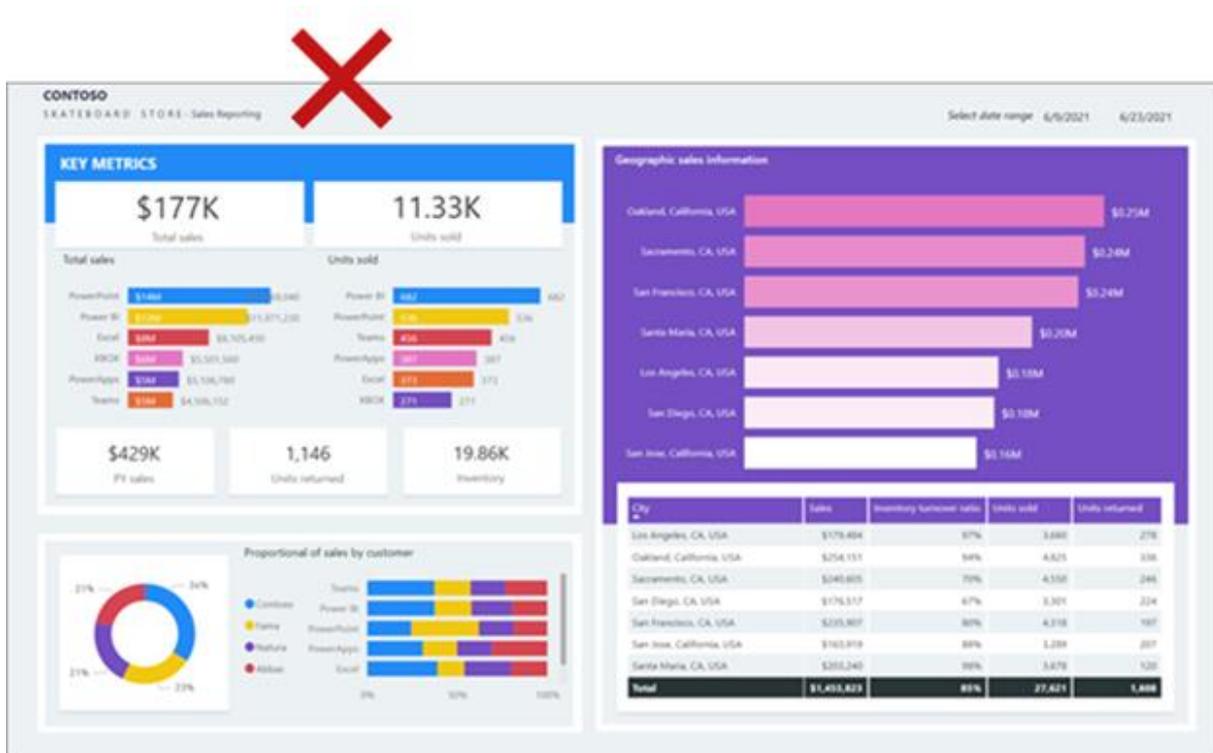


## Alignement

Un alignement précis des visuels améliore l'organisation visuelle et la perception de l'ordre. L'alignement aide à créer des sections implicites ou explicites sur la page de rapport, où les visuels liés sont regroupés visuellement pour une meilleure cohérence.

## Couleur

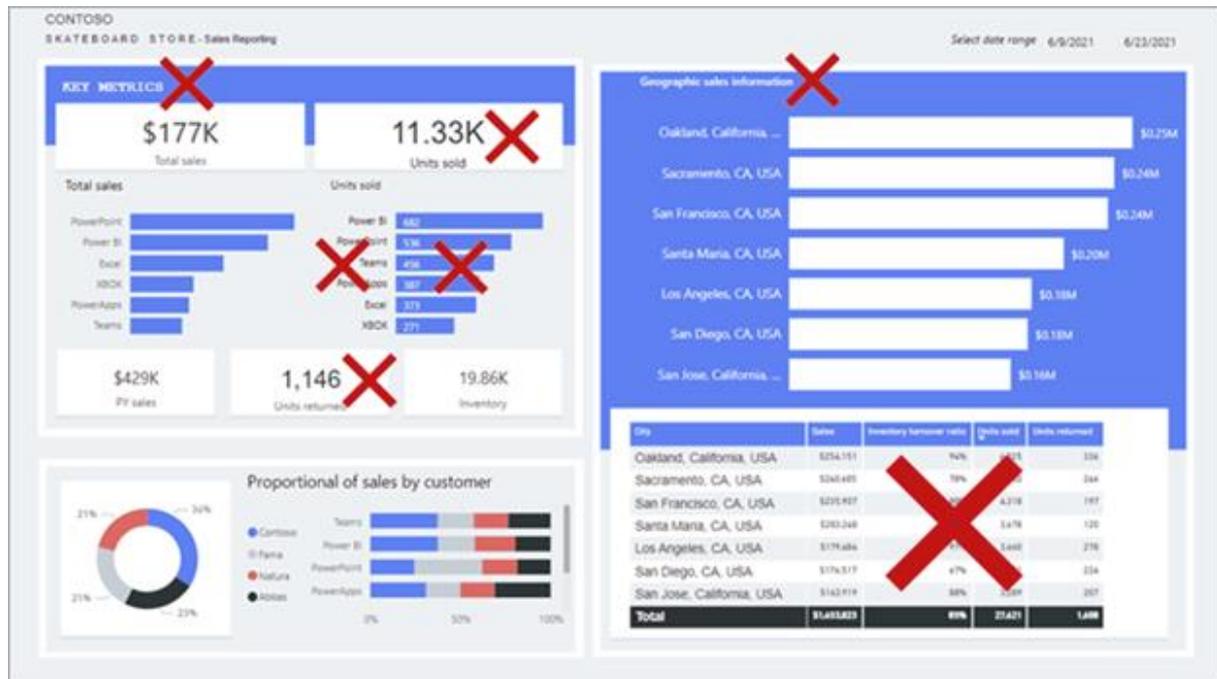
L'usage de la couleur doit être réfléchi et modéré, en utilisant des tons qui améliorent la lisibilité tout en évitant la distraction. Les couleurs doivent également offrir un contraste suffisant pour être accessibles aux personnes malvoyantes.



**Cohérence**

La cohérence dans la mise en forme des objets de rapport est essentielle pour maintenir une apparence professionnelle et organisée. L'utilisation d'un thème de rapport peut aider à appliquer de manière uniforme les styles à travers le rapport, incluant la typographie, les couleurs, et les options de mise en forme des visuels.

En respectant ces principes, les créateurs de rapports peuvent produire des visualisations qui ne sont pas seulement informatives, mais également attrayantes et faciles à naviguer, renforçant ainsi l'efficacité de la communication des données.



## Objets de rapport

Les objets de rapport dans Power BI sont essentiels pour créer des rapports visuellement attrayants et informatifs. Ils se classent en deux catégories principales : les visuels et les éléments.

### Visuels

Les visuels sont les composants principales des rapports Power BI, utilisés pour visualiser les données issues d'un modèle sémantique. Plus de 30 types de visuels sont disponibles directement dans Power BI, tels que les graphiques, les tableaux et les cartes. Les utilisateurs peuvent également ajouter des visuels personnalisés via Microsoft AppSource ou charger des fichiers .pbviz.

### Configuration des visuels

La configuration d'un visuel implique:

- Sélectionner et positionner le type de visuel dans le rapport.
- Mapper les champs du modèle sémantique aux "puits" du visuel, qui varient selon le type de visuel.
- Appliquer des filtres et ajuster les paramètres de tri et de totalisation.
- Personnaliser l'apparence à travers les options de mise en forme.

### Éléments

Les éléments ajoutent un intérêt visuel sans nécessairement utiliser de données du modèle sémantique. Les types d'éléments incluent:

- **Zones de texte:** Pour ajouter du texte enrichi, qui peut inclure des valeurs dynamiques du modèle sémantique.
- **Boutons:** Permettent une interaction similaire à celle d'une application, avec diverses actions comme naviguer entre les pages ou ouvrir des URL.
- **Formes:** Utilisées pour l'ornementation ou comme boutons interactifs.
- **Images:** Peuvent servir à charger des logos ou autres images pertinentes, avec la possibilité d'agir comme des boutons.

Les éléments et les visuels partagent des propriétés communes modifiables dans les options de Mise en forme, telles que l'emplacement, la taille, le titre, l'arrière-plan, les bordures et les ombres.

### Conseils pour la mise en forme

L'expérimentation est encouragée pour découvrir l'effet des différentes options de mise en forme. L'utilisation du raccourci Ctrl+Z permet de revenir aux configurations précédentes si un changement ne produit pas l'effet désiré.

En résumé, les objets de rapport dans Power BI offrent une flexibilité considérable pour la création de rapports personnalisés, engageants et interactifs, avec une forte capacité de narration visuelle grâce aux visuels et éléments dynamiques.

### Sélectionner des visuels de rapport

La sélection adéquate des visuels de rapport dans Power BI est cruciale pour communiquer efficacement des informations. Chaque type de visuel a ses propres avantages et contextes d'usage optimaux, facilitant ainsi la compréhension des données par les utilisateurs. Voici un guide pour choisir le visuel le plus approprié selon différents besoins de visualisation :

#### Visuels par Catégorie

- **Graphiques à barres ou histogrammes** : Idéaux pour afficher des données dans plusieurs catégories. Évitez les graphiques à barres empilées si les catégories sont nombreuses; privilégiez plutôt un graphique à barres simple avec des catégories sur l'axe horizontal.
- **Graphiques en courbes** : Non recommandés pour des données catégorielles non reliées, car ils impliquent une continuité ou une relation entre les points de données.

#### Visuels de Série Chronologique

- **Graphiques en courbes ou histogrammes** : Utilisés pour montrer des données temporelles, avec un axe des X représentant le temps de la plus ancienne à la plus récente.

- **Options analytiques** : Ajouter des prévisions ou d'autres indicateurs analytiques pour étendre la visualisation des données temporelles.

### Visuels Proportionnels

- **Graphiques à barres empilées ou histogrammes empilés 100 %** : Utiles pour montrer des parts d'un tout, permettant de visualiser des proportions sans afficher les valeurs exactes.

### Visuels Numériques

- **Cartes numériques** : Efficaces pour présenter des chiffres clés de manière très visible, utilisables seuls ou en groupes pour comparer plusieurs valeurs.

### Visuels de Grille

- **Tableaux et matrices** : Excellents pour détailler des informations avec la possibilité d'ajouter une mise en forme conditionnelle pour améliorer la visualisation des données et naviguer dans des structures hiérarchiques.

### Visuels de Performance

- **KPI et jauges** : Appropriés pour montrer les performances par rapport à des cibles, illustrant les variances avec des aides visuelles comme des couleurs ou des icônes.

### Visuels Géospatiaux

- **Cartes et cartes choroplèthes** : Choisissez en fonction du niveau de détail géographique des données. Les cartes sont bonnes pour des localisations précises, tandis que les cartes choroplèthes sont préférables pour visualiser des données par régions ou états.

La sélection de visuels doit toujours être guidée par le besoin de faire ressortir les informations les plus pertinentes de manière intuitive et accessible, en évitant de surcharger visuellement le rapport ou de présenter les données de manière trompeuse.

## Sélectionner des visuels de rapport en fonction de la présentation du rapport

Lors de la sélection des visuels pour un rapport Power BI, il est crucial de considérer la disposition spatiale de la page du rapport et de choisir des visuels qui non seulement répondent aux besoins de visualisation des données mais optimisent également l'utilisation de l'espace disponible. Voici quelques conseils pour bien choisir en fonction de l'espace et de la présentation du rapport :

1. **Analyser l'espace disponible** : Déterminez si l'espace sur la page est plus large que haut, ou l'inverse. Cela influencera directement le type de visuel le plus approprié.
2. **Choisir entre graphiques à barres et histogrammes** :

- **Espace large mais court** : Un graphique à barres empilées 100 % est préférable car il permet une meilleure lecture des longueurs de barre, ce qui facilite la comparaison des valeurs relatives entre différentes catégories.
  - **Espace étroit mais haut** : Un histogramme empilé 100 % est plus efficace, car les colonnes hautes permettent une meilleure visualisation des proportions dans un espace vertical restreint.
3. **Considérer l'agrément visuel** : Optez pour des visuels qui non seulement remplissent efficacement l'espace mais sont aussi agréables à regarder, ce qui peut améliorer l'engagement des utilisateurs et la compréhension des données.
  4. **Adaptabilité** : Assurez-vous que le visuel choisi reste lisible et efficace même lorsque les dimensions du rapport ou de la page sont modifiées, par exemple lors de la visualisation sur différents appareils ou résolutions d'écran.



Ce n'est pas toujours le cas. Quand le visuel doit remplir un espace étroit mais haut, l'historgramme efficace. La hauteur des *grandes* colonnes est plus facile à interpréter que la longueur des *petit*



## Mettre en forme et configurer des visualisations

Power BI Desktop offre une vaste gamme d'options de mise en forme pour personnaliser l'apparence des visualisations. Voici un guide étape par étape sur comment mettre en forme et configurer les visualisations pour optimiser l'efficacité de vos rapports :

## Sélection de la Visualisation et Accès au Volet de Mise en Forme

1. **Choisir la visualisation** sur le canevas.
2. **Ouvrir le volet de mise en forme** en cliquant sur l'icône de rouleau de peinture.

## Configuration des Options de Mise en Forme

- **Titre** : Ajoutez ou modifiez le titre pour décrire clairement les données. Personnalisez la police, la taille, la couleur, et l'alignement pour une meilleure visibilité et cohérence avec le thème du rapport.
- **Arrière-plan** : Choisissez une couleur ou une image. Un fond blanc ou neutre est souvent préférable pour une meilleure lisibilité des données.
- **Bordure** : Ajoutez une bordure pour distinguer clairement la visualisation sur le canevas. Personnalisez la couleur et l'épaisseur pour l'aligner avec le schéma de couleurs du rapport.
- **Positionnement et Taille** : Ajustez précisément la taille et l'emplacement de la visualisation pour un alignement parfait ou pour s'adapter à un espace spécifique sur le canevas.

## Personnalisation des Données

- **Couleurs des Données** : Définissez des couleurs spécifiques pour les différentes valeurs de données pour maintenir une cohérence visuelle à travers le rapport.
- **Étiquettes de Données** : Modifiez la police, la taille, et la couleur des étiquettes pour assurer qu'elles soient lisibles sur n'importe quel fond.

## Utilisation Avancée des Info-bulles

- **Info-bulles Standard** : Elles affichent par défaut la valeur et la catégorie du point de données. Vous pouvez enrichir les info-bulles avec plus de données ou des visualisations.
- **Info-bulles Personnalisées** : Créez une page d'info-bulle dans le rapport. Configurez cette page pour qu'elle serve d'info-bulle, en ajoutant des visualisations et des champs spécifiques.
- **Activation des Info-bulles** : Appliquez l'info-bulle personnalisée aux visualisations désirées pour qu'elle s'affiche lors du survol.

## Exemples Spécifiques de Configuration

- **Titre** : Changez le titre en "Total des ventes par pays" et augmentez la taille de la police à 16.
- **Arrière-plan** : Réglez la couleur de fond sur un gris clair pour une légère variation sans compromettre la clarté des données.

- **Info-bulle** : Ajoutez des informations contextuelles plus détaillées ou utilisez des visualisations dans les info-bulles pour une analyse plus profonde à la survol.

En suivant ces étapes, vous pouvez non seulement rendre vos visualisations plus attrayantes mais aussi plus informatives et utiles pour les utilisateurs de vos rapports. Chaque modification dans le volet de mise en forme est immédiatement reflétée dans la visualisation, permettant des ajustements rapides et précis. Si nécessaire, utilisez l'option pour rétablir les valeurs par défaut pour annuler des modifications spécifiques.

## Travailler avec les indicateurs de performance clés

Les indicateurs de performance clés (KPI) sont des outils essentiels dans Power BI pour suivre et visualiser la progression vers des objectifs spécifiques. Voici une explication détaillée de comment configurer et utiliser un KPI dans Power BI :

### Éléments nécessaires pour configurer un KPI

1. **Unité de mesure** : C'est la donnée principale que vous souhaitez suivre (exemples : ventes totales, nombre de recrutements, nombre de prêts attribués).
2. **Objectif** : La cible avec laquelle vous voulez comparer votre performance actuelle.
3. **Série chronologique** : La fréquence temporelle de la mesure (quotidienne, mensuelle, annuelle).

### Configuration d'un Indicateur de Performance Clé

1. **Ajout de l'objet visuel KPI** : Commencez par sélectionner l'icône de KPI dans le volet Visualisations de Power BI.
2. **Configuration des champs** :
  - **Indicateur** : Saisissez l'unité de mesure que vous suivez.
  - **Objectifs cibles** : Entrez la valeur de l'objectif que vous souhaitez atteindre.
  - **Axe de la tendance** : Sélectionnez la périodicité de la série chronologique (par exemple, quotidienne, mensuelle, annuelle).

### Visualisation d'un KPI

- Une fois configuré, le KPI affichera visuellement où vous vous situez par rapport à l'objectif fixé. Vous pourrez voir si vous êtes au-dessus ou en dessous de cet objectif et suivre l'évolution au fil du temps.

### Conseils pour l'utilisation efficace des KPI

- **Clarté** : Assurez-vous que les KPI sont faciles à comprendre pour tous les utilisateurs du rapport.
- **Visibilité** : Placez les KPI dans une section visible du rapport pour qu'ils attirent l'attention.
- **Comparaison** : Utilisez des séries de KPI pour comparer les performances à différentes échelles de temps ou entre différentes unités organisationnelles.

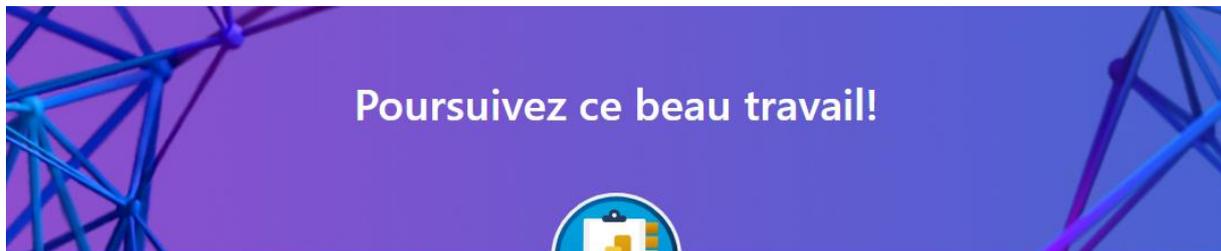
Les KPI sont particulièrement utiles dans les tableaux de bord de gestion où les décisions doivent être prises rapidement et sont basées sur des données actualisées et claires. Ils permettent non seulement de suivre les performances mais aussi de motiver les équipes à atteindre les objectifs fixés.

## Exercice

<https://learn.microsoft.com/fr-ch/training/modules/power-bi-effective-reports/9-lab>

## Résumé

La conception efficace d'un rapport Power BI repose sur l'agencement stratégique des objets sur une ou plusieurs pages pour communiquer clairement les données aux utilisateurs. Le succès de la conception dépend largement de la sélection des visuels appropriés, de l'utilisation judicieuse des options d'analytique, du classement intuitif des éléments visuels, et de la mise en œuvre de styles et de formats adaptés.



Concevoir des rapports  
Power BI

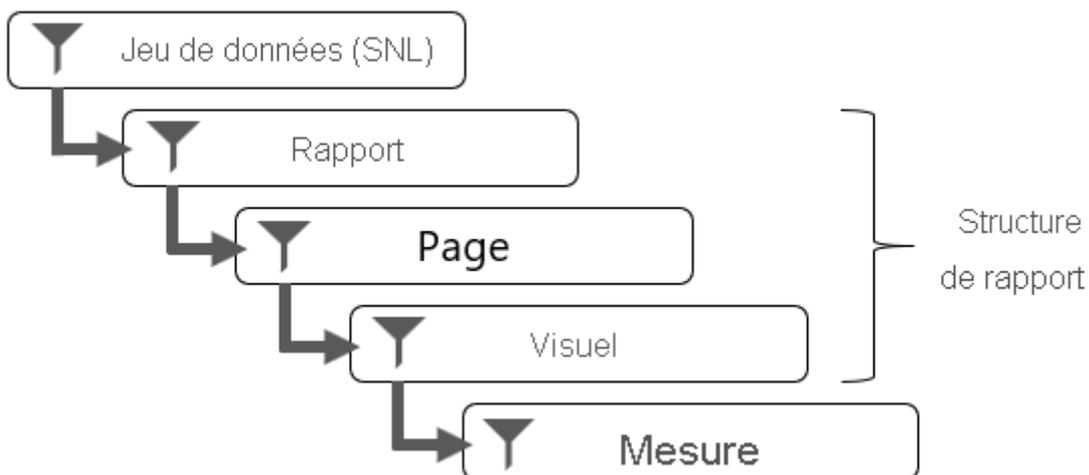
**Vous avez gagné un succès!**

Félicitations, mais que devez-vous faire ensuite?

# Configurer des filtres de rapport Power BI

## Présentation

Dans Microsoft Power BI, le filtrage peut être effectué à différents niveaux allant du modèle sémantique au visuel individuel. Au niveau du modèle sémantique, connu sous le nom de sécurité au niveau des lignes (RLS), le modèle sémantique gère l'accès aux données en restreignant la visibilité à des sous-ensembles spécifiques, basés sur le profil utilisateur. Ce niveau ne peut pas être modifié ou ignoré par les rapports. Au niveau du rapport, les filtres affectent l'ensemble du document, tandis que le filtrage au niveau de la page concerne tous les visuels d'une page spécifique. Les filtres visuels ciblent des éléments individuels dans une page, et enfin, les mesures, définies par des expressions DAX, permettent de calculer des données tout en modifiant le contexte de filtrage avec des fonctions telles que `CALCULATE` ou `CALCULATETABLE`. Ces dernières offrent la flexibilité nécessaire pour ajuster les filtres appliqués aux données. Il est crucial de gérer avec soin les filtres lorsqu'ils interagissent avec la SNL pour éviter que certains utilisateurs ne se retrouvent sans aucune donnée visible due à des restrictions de filtrage inadéquates.



## Appliquer des filtres à la structure du rapport

Lors de la conception de rapports dans Microsoft Power BI, les filtres sont appliqués via le volet Filtres qui comprend trois sections distinctes pour organiser les niveaux de filtrage : filtres dans toutes les pages, filtres dans cette page, et filtres sur ce visuel. Les filtres appliqués au niveau du rapport affectent toutes les pages et visuels, agissant comme des filtres globaux. Les filtres de page spécifient des restrictions pour tous les visuels d'une page donnée et se superposent aux filtres du rapport. Par exemple, un filtre de rapport peut limiter les données au pays États-Unis, et un filtre de page peut restreindre davantage les données à l'État de Washington.

Les filtres de visuel sont appliqués uniquement au visuel sélectionné et se cumulent avec les filtres de rapport et de page existants. Ces filtres peuvent également utiliser une mesure pour exclure des données spécifiques, comme éliminer des magasins d'un histogramme basé sur les ventes totales. Les filtres peuvent être de types divers, tels que basique, avancé, N premiers, ou date et heure relatives, chacun adapté à des besoins spécifiques de filtrage.

Il est également possible de verrouiller des filtres pour empêcher les modifications par les utilisateurs finaux, ou de les masquer pour que les consommateurs du rapport ne soient pas conscients de leur présence. En outre, le volet Filtres lui-même peut être masqué pour restreindre l'accès aux options de filtrage. Ces fonctionnalités permettent un contrôle précis et une personnalisation du filtrage dans les rapports Power BI, garantissant que les données sont présentées de manière appropriée et sécurisée.

## Appliquer des filtres avec des segments

Lors de la conception de rapports dans Microsoft Power BI, les filtres sont appliqués via le volet Filtres qui comprend trois sections distinctes pour organiser les niveaux de filtrage : filtres dans toutes les pages, filtres dans cette page, et filtres sur ce visuel. Les filtres appliqués au niveau du rapport affectent toutes les pages et visuels, agissant comme des filtres globaux. Les filtres de page spécifient des restrictions pour tous les visuels d'une page donnée et se superposent aux filtres du rapport. Par exemple, un filtre de rapport peut limiter les données au pays États-Unis, et un filtre de page peut restreindre davantage les données à l'État de Washington.

Les filtres de visuel sont appliqués uniquement au visuel sélectionné et se cumulent avec les filtres de rapport et de page existants. Ces filtres peuvent également utiliser une mesure pour exclure des données spécifiques, comme éliminer des magasins d'un histogramme basé sur les ventes totales. Les filtres peuvent être de types divers, tels que basique, avancé, N premiers, ou date et heure relatives, chacun adapté à des besoins spécifiques de filtrage.

Il est également possible de verrouiller des filtres pour empêcher les modifications par les utilisateurs finaux, ou de les masquer pour que les consommateurs du rapport ne soient pas conscients de leur présence. En outre, le volet Filtres lui-même peut être masqué pour restreindre l'accès aux options de filtrage. Ces fonctionnalités permettent un contrôle précis et une personnalisation du filtrage dans les rapports Power BI, garantissant que les données sont présentées de manière appropriée et sécurisée.

## Concevoir des rapports avec des techniques de filtrage avancées

En plus des filtres et des segments, les auteurs de rapports Power BI utilisent diverses techniques de filtrage pour affiner les interactions et la présentation des données. Les interactions entre visuels permettent aux consommateurs de rapports d'utiliser un visuel pour influencer les données affichées dans d'autres visuels sur la même page. Cette fonctionnalité, activable par la touche Ctrl pour ajouter des filtres croisés supplémentaires, peut être configurée pour permettre des interactions spécifiques entre paires de visuels, incluant la désactivation du filtrage croisé ou l'utilisation de la sélection croisée.

Les pages d'extraction offrent une possibilité d'exploration plus approfondie à partir d'un visuel, propageant tous les filtres appliqués au visuel original vers la page d'extraction. Les info-bulles de rapport, quant à elles, fournissent des informations contextuelles lorsqu'un utilisateur survole un visuel, héritant également des filtres du visuel.

Les signets captent une configuration spécifique d'un rapport, incluant les états des filtres et des visuels. Ils permettent une navigation rapide vers ces configurations prédéfinies, avec des options pour créer des signets personnels par les consommateurs. Les auteurs de rapports peuvent également configurer des signets pour réinitialiser les segments à un état par défaut.

Les options de rapport incluent la désactivation des filtres persistants, le masquage des en-têtes de visuel ou de l'icône de filtrage, et la limitation des modifications des types de filtre par les consommateurs de rapports. Ces options permettent de contrôler finement l'accès et la modification des filtres pour assurer la cohérence et la pertinence des données affichées.

Enfin, les options de réduction des requêtes visent à optimiser les performances en limitant le nombre de requêtes envoyées au modèle sémantique, particulièrement utile pour les configurations utilisant DirectQuery ou traitant de grands volumes de données. Ces paramètres incluent l'ajout de boutons d'application pour contrôler quand les filtres et les segments sont mis à jour, réduisant ainsi les rendus superflus et améliorant la réactivité des rapports.

## Filtrage au moment de la consommation

En mode Lecture, les consommateurs de rapports Power BI disposent de plusieurs options de filtrage pour interagir avec les données, telles que l'utilisation de segments, l'application de filtres, et la mise en œuvre d'actions de filtrage interactif. Chaque technique offre une dimension spécifique de contrôle sur les données affichées.

Les **segments** permettent aux consommateurs de modifier ou d'effacer des filtres facilement, avec une option de recherche pour localiser rapidement des éléments spécifiques à filtrer. Ils représentent une des interactions les plus intuitives et fréquemment utilisées.

L'**utilisation de filtres** est possible si le volet Filtres est accessible. Les consommateurs peuvent réviser, modifier ou effacer des filtres existants à moins qu'ils ne soient verrouillés. Les filtres peuvent varier de simples à avancés, permettant des ajustements précis sur les données à afficher. Cependant, les consommateurs ne peuvent pas ajouter de nouveaux filtres ou modifier le type de filtre en options plus spécifiques comme N premiers ou date/heure relative sans des ajustements par l'auteur du rapport ou via des API dans les scénarios d'incorporation.

Les **actions de filtrage interactif** incluent des capacités comme découper pour propager des filtres, inclure/exclure des filtres, et filtrer de manière croisée, qui permettent des interactions dynamiques et spécifiques avec les données au sein du même rapport. Les consommateurs peuvent aussi explorer des pages de rapport

supplémentaires à travers l'extraction, qui propage les filtres appliqués au visuel d'origine.

Les **signets** sont utiles pour les consommateurs souhaitant sauvegarder et revenir à des vues spécifiques du rapport, incluant des configurations de filtres et de visualisation.

La capacité de **déterminer les filtres appliqués** est cruciale pour comprendre comment les données sont filtrées dans un visuel particulier. En survolant l'icône Filtrer, une fenêtre contextuelle apparaît, montrant tous les filtres affectant le visuel. Si cette icône n'est pas visible, cela peut être dû à des paramètres désactivant les entêtes de visuel ou l'icône elle-même.

Enfin, les **filtres persistants** permettent de sauvegarder les paramètres de filtrage d'un consommateur pour qu'ils soient réappliqués lors de sessions futures. Ces paramètres peuvent être réinitialisés à l'état par défaut prévu par l'auteur du rapport grâce à la commande Rétablir les valeurs par défaut.

Chacune de ces techniques de filtrage enrichit l'expérience du consommateur en offrant divers degrés de personnalisation et de contrôle sur les données affichées, tout en maintenant une gestion appropriée de l'accès et de la visibilité des données selon les besoins de l'organisation et les préférences de l'utilisateur.

## Sélectionner les techniques de filtre des rapports

Les filtres et les segments dans Power BI sont deux techniques de manipulation de données qui, bien que similaires en fonction, diffèrent en termes de mise en œuvre et d'impact sur l'expérience utilisateur et les performances du rapport. Choisir entre l'utilisation d'un filtre ou d'un segment dépend souvent d'un compromis entre performance et conception.

### Filtres

Les **filtres** sont généralement plus performants car ils ne nécessitent pas de rendu de visuels supplémentaires et sont localisés dans le volet Filtres, qui est typiquement placé sur le côté droit du rapport. Ils permettent une configuration avancée et utilisent divers types d'expressions pour une manipulation précise des données affichées. Le volet offre également des fonctionnalités comme la recherche et le tri, améliorant l'efficacité lors de la gestion d'un grand nombre de filtres. Les filtres peuvent être verrouillés ou masqués, et il est possible de les appliquer globalement via un bouton d'application, réduisant ainsi le nombre de requêtes envoyées.

### Inconvénients des filtres :

- Moins de flexibilité visuelle par rapport aux segments.
- Difficulté à suivre les filtres appliqués, surtout au niveau des visuels.

### Segments

Les **segments** offrent une intégration visuelle plus harmonieuse et intuitive dans la page de rapport, ce qui peut améliorer l'expérience utilisateur. Ils sont hautement

configurables, supportent l'affichage d'images via des URL et peuvent être placés stratégiquement pour un accès facile. Les segments peuvent être configurés pour filtrer spécifiquement certains visuels ou pages, et leur état actuel est visible directement dans le rapport.

### **Inconvénients des segments :**

- Impact sur les performances dues au temps nécessaire pour le rendu des visuels.
- Occupation de l'espace précieux sur la page qui pourrait être utilisé pour d'autres données.

### **Utilisation Combinée et Conseils**

Il est généralement conseillé de choisir entre utiliser des filtres ou des segments pour éviter la confusion chez les utilisateurs. Voici quelques stratégies supplémentaires pour optimiser l'utilisation des deux techniques :

- **Verrouiller ou masquer les filtres au niveau des visuels** dans le volet Filtres pour éviter que les consommateurs de rapports ne modifient ou ne voient des filtres qui pourraient compliquer leur expérience.
- **Créer un signet pour réinitialiser tous les segments** aux valeurs par défaut et ajouter un bouton sur la page pour faciliter l'accès à ce signet.
- **Utiliser d'autres visuels comme segments** pour le filtrage croisé et bien expliquer aux consommateurs comment utiliser ces visuels.
- Envisager **la création d'une page dédiée aux segments** pour centraliser les options de filtrage, en synchronisant les segments avec d'autres pages et en les masquant sur les pages non dédiées.

Ces techniques peuvent aider à maintenir une interface claire et à optimiser les performances du rapport tout en offrant aux utilisateurs finaux les outils nécessaires pour interagir efficacement avec les données. Pour une mise en œuvre réussie, il est crucial de bien comprendre les besoins des utilisateurs et les objectifs de rapportage pour choisir la méthode de filtrage la plus appropriée.

## **Étude de cas - Configurer des filtres de rapport en fonction du feedback**

Chez Contoso Skateboard Company, le processus d'amélioration de la visibilité et de l'accessibilité des filtres pour le rapport des ventes illustre bien l'importance de l'expérience utilisateur dans la conception de rapports analytiques. Voici un récapitulatif de l'évolution de ce rapport, qui montre les adaptations nécessaires pour répondre aux besoins des utilisateurs :

### **1. Première tentative : Filtre global dans le volet Filtres**

- **Mise en œuvre** : Ajout du champ Catégorie au niveau du rapport dans le volet Filtres.
- **Problème** : Difficulté pour les utilisateurs à localiser et à comprendre la portée du filtre, car il n'était pas directement visible sur la page du rapport.
- **Résultat** : Feedback négatif, nécessitant une révision de l'approche.

## 2. Deuxième tentative : Segment de liste déroulante

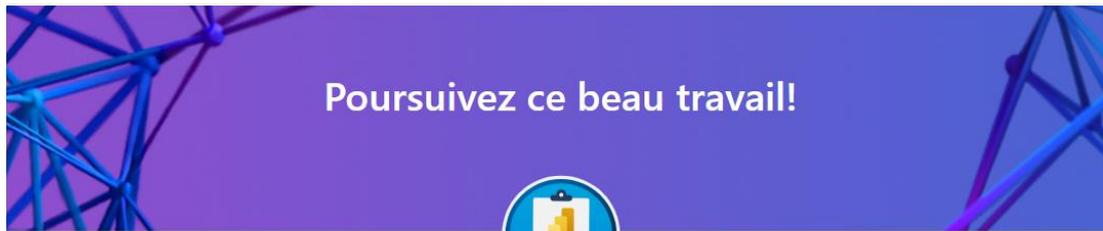
- **Mise en œuvre** : Remplacement du filtre global par un segment de liste déroulante en haut de la page de rapport.
- **Problème** : Lorsque plusieurs catégories étaient sélectionnées, le segment indiquait "Sélections multiples", obligeant les utilisateurs à ouvrir le segment pour voir les catégories spécifiques filtrées.
- **Résultat** : Persistance du mécontentement des utilisateurs, nécessitant une nouvelle approche.

## 3. Troisième tentative : Segment horizontal

- **Mise en œuvre** : Repositionnement et changement d'orientation du segment des catégories de produits en horizontal dans le coin supérieur droit de la page.
- **Avantage** : Permet l'affichage constant de trois catégories différentes ainsi que des sélections spécifiques, améliorant la visibilité directe et la compréhension des filtres appliqués.
- **Attention** : Il est conseillé d'éviter l'orientation horizontale lorsque de nombreuses valeurs distinctes sont présentes, car les utilisateurs pourraient ne pas remarquer les flèches de défilement et donc manquer certaines options.

## Résumé

Lors de la conception de rapports Power BI, il est essentiel de choisir entre le volet Filtres et les segments pour appliquer des filtres, afin de ne pas confondre les utilisateurs. Avant de partager un rapport, il est utile de faire une démonstration ou de fournir une vidéo explicative sur l'utilisation des filtres et des fonctionnalités interactives comme le filtrage croisé et les signets. Cela garantit que les consommateurs de rapports comprennent comment interagir avec le rapport de manière efficace. Pour approfondir vos connaissances en DAX, qui est crucial pour le filtrage avancé, consultez les ressources de Microsoft sur l'utilisation de DAX dans Power BI Desktop.



Configurer des filtres de  
rapport Power BI

**Vous avez gagné un succès!**

Félicitations, mais que devez-vous faire ensuite?

## Créer et gérer des espaces de travail dans Power BI

Après avoir chargé et transformé des données de diverses sources, créé des visualisations, formulé des équations DAX et publié des rapports sur Microsoft Power BI, l'étape suivante consiste à partager ces rapports au sein de votre organisation et avec un public plus large. Cela peut être réalisé grâce aux espaces de travail dans Power BI, qui servent de référentiels centralisés pour la collaboration entre collègues et équipes. Ces espaces permettent de regrouper des rapports et des tableaux de bord, facilitant ainsi les efforts de collaboration ciblés et la présentation de ces éléments dans un environnement sécurisé. Les espaces de travail assurent également un niveau de sécurité élevé en contrôlant l'accès aux modèles sémantiques, rapports et tableaux de bord. Ce module aborde les différentes tâches associées à la création et à la gestion d'un espace de travail, incluant l'importation et la mise à jour des ressources, la configuration de la protection des données et la résolution des problèmes liés aux données.

### Distribuer un rapport ou un tableau de bord

Après avoir créé des rapports pour l'équipe des ventes chez Tailwind Traders, le défi était de les rendre visibles et partageables. En utilisant le service Power BI, un espace de travail a été créé pour centraliser ces rapports, facilitant ainsi le partage et la collaboration. Pour créer un espace de travail, il faut se rendre dans le menu Espaces de travail du service Power BI, sélectionner le bouton pour créer un espace de travail, et fournir les informations nécessaires telles que le nom et la description. Il est aussi possible d'ajouter une image et de configurer des notifications pour des utilisateurs spécifiques en cas de problèmes.

Une fois l'espace créé, divers rôles peuvent être attribués aux membres de l'équipe pour gérer le contenu. Les rôles vont de l'administrateur, qui peut gérer entièrement l'espace, au simple observateur, qui peut seulement interagir avec le contenu. Il est crucial d'assigner le niveau d'accès minimal nécessaire pour chaque rôle afin d'optimiser la sécurité et l'efficacité.

Le contenu de l'espace de travail peut ensuite être organisé dans une application, qui est une collection de rapports, tableaux de bord, et autres ressources, destinée à être partagée en lecture seule. Lors de la création de l'application, il est possible de personnaliser son apparence et de définir le contenu accessible à différentes audiences. Une fois prête, l'application peut être publiée, nécessitant une licence Power BI Pro pour la publication et la visualisation.

Enfin, l'espace de travail peut être mis à jour aussi facilement que lors de la publication initiale de l'application, en sélectionnant simplement l'option de mise à jour dans l'espace de travail et en apportant les modifications nécessaires. La publication d'une application dans Power BI permet une distribution efficace et sécurisée du contenu analytique au sein de l'organisation.

## Superviser l'utilisation et le niveau de performance

Il est crucial de superviser l'utilisation et les performances des espaces de travail dans Power BI, car cela permet de concentrer les efforts d'amélioration là où les performances sont les plus faibles et de quantifier l'impact des rapports. Ces métriques, accessibles uniquement aux utilisateurs avec des rôles d'administrateur, membre ou contributeur dotés d'une licence Power BI Pro, offrent une visibilité sur qui utilise les rapports, les actions effectuées, et les problèmes de performance éventuels.

Pour voir ces métriques, il faut se rendre dans l'espace de travail concerné, sélectionner le rapport ou le tableau de bord voulu, et choisir l'option de visualisation des métriques d'utilisation. Cette action ouvre un tableau de bord qui inclut des graphiques et des cartes KPI indiquant le nombre d'observateurs par jour, le nombre total de vues, et les partages.

Le rapport de performance, accessible via un onglet dédié, fournit des données sur le temps d'ouverture des rapports et l'évolution de ce délai, aidant à comprendre comment les performances varient avec l'augmentation du nombre d'utilisateurs. Il montre également les performances quotidiennes et hebdomadaires pour divers percentiles, avec des filtres de date permettant d'analyser les tendances au fil du temps.

Ces outils sont essentiels pour assurer que les espaces de travail et les applications de Power BI restent performants et répondent aux besoins des utilisateurs, permettant ainsi de faire des ajustements basés sur des données concrètes et de maintenir une haute qualité de service. Pour plus de détails sur ces fonctionnalités, il est recommandé de consulter la documentation sur la surveillance de l'utilisation des métriques dans Power BI.

## Recommander une stratégie de cycle de vie de développement

Pour assurer l'intégrité des données et des rapports sans ralentir les chronologies de développement chez Tailwind Traders, la mise en place d'un pipeline de déploiement dans Power BI Premium est recommandée. Ce pipeline permet de gérer le contenu à travers différentes phases de développement, de test, et de production, en centralisant le développement et la gestion des tableaux de bord et des rapports.

## Avantages du Pipeline de Déploiement

- **Productivité améliorée** : Utiliser des pipelines existants minimise la duplication des efforts.
- **Livraison plus rapide** : Le développement des rapports est rationalisé, réduisant le temps de mise en production.
- **Réduction des erreurs humaines** : La réutilisation des pipelines diminue le risque d'erreurs lors du déplacement du contenu entre les environnements.

## Configuration des Pipelines de Déploiement

1. **Accès au service Power BI** : Allez dans les Pipelines de déploiement et créez un nouveau pipeline, par exemple, **SalesPipeline**.
2. **Attribuer des espaces de travail** : Définissez les espaces pour les phases de développement, test, et production. Chaque phase aura son propre espace de travail dédié.

## Gestion des Environnements

- **Développement** : Lieu initial où le contenu est créé et itéré avec d'autres développeurs.
- **Test** : Les utilisateurs cibles et les testeurs d'acceptation évaluent les rapports dans cet environnement pour s'assurer qu'ils sont exempts de défauts avant la production.
- **Production** : L'environnement final où les rapports validés sont rendus accessibles à un public plus large.

## Phases du Pipeline

- **Attribution des rôles** : Assurez-vous que les bonnes personnes ont accès aux bons outils et données dans chaque phase.
- **Modification de la source de données** : Adaptez les sources de données pour les tests afin de ne pas perturber les données opérationnelles tout en simulant les conditions de production.
- **Publication et mise à jour des applications** : Une fois les tests concluants, publiez l'application pour la rendre accessible aux utilisateurs finaux.

## Supervision et Maintenance

- Utilisez des outils de comparaison pour suivre les modifications entre les environnements et décider des éléments à déployer ou à rejeter.
- Soyez vigilant avec les dépendances entre les rapports et les modèles sémantiques pour éviter des incohérences.

En suivant ces recommandations, vous pouvez établir un pipeline de déploiement efficace qui soutient les processus itératifs de développement tout en maintenant l'intégrité et la performance des rapports et tableaux de bord dans Power BI. Pour plus

de détails, consultez les pratiques recommandées pour les pipelines de déploiement sur le site officiel de Power BI.

## Résoudre les problèmes liés aux données en consultant leur traçabilité

Dans Power BI, la vue Traçabilité est un outil essentiel pour gérer les modèles sémantiques et comprendre les interactions entre les différents artefacts d'un espace de travail, notamment les sources de données, les flux de données, les rapports et les tableaux de bord. Voici comment elle peut être utilisée efficacement pour résoudre les problèmes liés aux données, notamment pour Tailwind Traders.

### Avantages de la Vue Traçabilité

- **Simplification du dépannage:** Visualisez directement le chemin des données depuis la source jusqu'à la destination pour identifier rapidement les points sensibles.
- **Gestion efficace des modifications:** Observez comment un changement dans un modèle sémantique affecte d'autres rapports ou tableaux de bord, facilitant ainsi les mises à jour et les corrections.
- **Gain de temps:** Identifiez facilement les rapports et tableaux de bord nécessitant une mise à jour.

### Utilisation de la Vue Traçabilité

Pour accéder à cette vue, vous devez avoir un rôle d'administrateur, de membre ou de contributeur et disposer d'une licence Power BI Pro. Elle est uniquement disponible pour les espaces de travail d'application.

1. **Accéder à la Traçabilité:** Dans l'espace de travail, sélectionnez Traçabilité depuis le menu Vue pour ouvrir le canevas de la vue Traçabilité.
2. **Analyser le Flux de Données:** Examinez les artefacts inclus dans l'espace de travail et suivez les flèches indiquant le flux des données de la source jusqu'à la destination finale comme les tableaux de bord.

### Composants Clés de la Vue Traçabilité

- **Sources de données:** Chaque carte représente une source de données avec des détails comme le type et la passerelle utilisée. Une interaction avec la carte peut révéler des informations supplémentaires comme le chemin du fichier et l'état de la connexion.
- **Modèles sémantiques/Flux de données:** Connectés à des sources de données externes, ces cartes montrent quand le modèle a été mis à jour pour la dernière fois et permettent une actualisation directe via une icône dédiée.

### Exemples de Traçabilité pour Tailwind Traders

Supposons que des rapports montrent des données obsolètes. Vous pouvez utiliser la vue Traçabilité pour:

- **Identifier les modèles périmés:** Parcourez les modèles sémantiques et déterminez lesquels nécessitent une actualisation.
- **Mettre à jour les modèles:** Utilisez le bouton d'actualisation des données sur les cartes des modèles sémantiques pour les mettre à jour immédiatement.
- **Analyse d'impact:** Sélectionnez l'icône d'analyse d'impact pour voir comment un modèle sémantique spécifique affecte d'autres rapports et tableaux de bord dans l'espace de travail.

### Actions Spécifiques

- **Double-cliquer sur une carte:** Affichez des détails supplémentaires comme les métadonnées, la date de dernière actualisation, et les noms des tables.
- **Notifier des changements:** Utilisez l'option de notification pour informer les propriétaires des modèles sémantiques des modifications apportées.

En utilisant la vue Traçabilité, vous pouvez non seulement résoudre les problèmes de données de manière proactive mais aussi optimiser la gestion des ressources dans Power BI pour Tailwind Traders. Cette approche garantit que les données sont à jour et que les performances des rapports sont maintenues à un niveau optimal. Pour plus d'informations, vous pouvez consulter la documentation officielle sur la traçabilité des données dans Power BI.

### Configurer la protection des données

Pour assurer la sécurité des données au sein de l'espace de travail Tailwind Traders sur Power BI, il est crucial de mettre en œuvre des mesures de protection complètes, surtout avec l'augmentation du nombre de rapports et tableaux de bord ajoutés et le risque potentiel d'exportation non autorisée des données. Voici une approche recommandée pour configurer la protection des données en utilisant les fonctionnalités de sécurité de Power BI :

#### Utilisation des Étiquettes de Confidentialité

Les étiquettes de confidentialité de Microsoft permettent de classer et de protéger les données, les tableaux de bord, les rapports, les modèles sémantiques et les flux de données dans Power BI. Ces étiquettes assurent la continuité de la protection des données, même lorsqu'elles sont exportées hors de Power BI.

Procédure pour appliquer des étiquettes de confidentialité :

1. **Configuration des Étiquettes:** Avant de commencer, confirmez que vous disposez des licences nécessaires. Les étiquettes de confidentialité sont configurées dans le Microsoft 365 Security Center et peuvent être appliquées à Power BI rapidement.
2. **Attribution d'Étiquettes:** Accédez à l'espace de travail concerné dans Power BI, sélectionnez l'objet (rapport, tableau de bord, modèle sémantique) que vous souhaitez sécuriser, et sous les points de suspension (...), cliquez sur **Paramètres**. Là, vous pouvez attribuer une étiquette de confidentialité (par exemple, Confidentiel, Hautement confidentiel).

3. **Impact des Étiquettes:** Lors de l'exportation des données vers des formats tels que Excel, PowerPoint ou PDF, l'étiquette de confidentialité est automatiquement appliquée, assurant que les données exportées restent protégées.

#### Mesures de Protection Supplémentaires

- **Chiffrement et Filigranes:** Lors de l'exportation des données, des mesures comme le chiffrement et l'ajout de filigranes aident à sécuriser davantage les données et à dissuader l'utilisation inappropriée.
- **Utilisation de Microsoft Cloud App Security:** Cette solution de surveillance permet de suivre et d'examiner les activités dans Power BI, ce qui est crucial pour détecter et répondre aux utilisations anormales ou non autorisées des données.

#### Gestion des Accès et des Permissions

- **Restrictions sur l'Exportation des Données:** Configurez des politiques qui limitent les utilisateurs pouvant exporter des données et sous quelles conditions. Ces politiques peuvent être gérées par les étiquettes de confidentialité.
- **Surveillance des Activités:** Soyez vigilant sur les activités des utilisateurs dans l'espace de travail, surtout en ce qui concerne l'exportation des données pour prévenir toute fuite d'informations sensibles.

En mettant en place ces pratiques, vous pouvez non seulement sécuriser les données au sein de Power BI mais aussi lorsqu'elles sont partagées ou exportées. Cela est essentiel pour maintenir l'intégrité et la sécurité des données de Tailwind Traders. Pour plus de détails sur la mise en place et l'utilisation des étiquettes de confidentialité ainsi que d'autres mesures de sécurité dans Power BI, consultez la documentation officielle sur l'application des étiquettes de sensibilité des données dans Power BI.

## Résumé

Les espaces de travail dans Power BI sont cruciaux pour partager des rapports, créer des tableaux de bord, et collaborer avec des équipes. Ils améliorent la performance des rapports, garantissent la conformité aux normes de sécurité, et facilitent le partage de contenu. Savoir gérer ces espaces de travail est un atout majeur pour optimiser la création de tableaux de bord dans Power BI.



Créer et gérer des espaces de  
travail dans Power BI

**Vous avez gagné un exploit !**

Félicitations, mais que devez-vous faire ensuite ?

## Gérer des modèles sémantiques dans Power BI

### Introduction

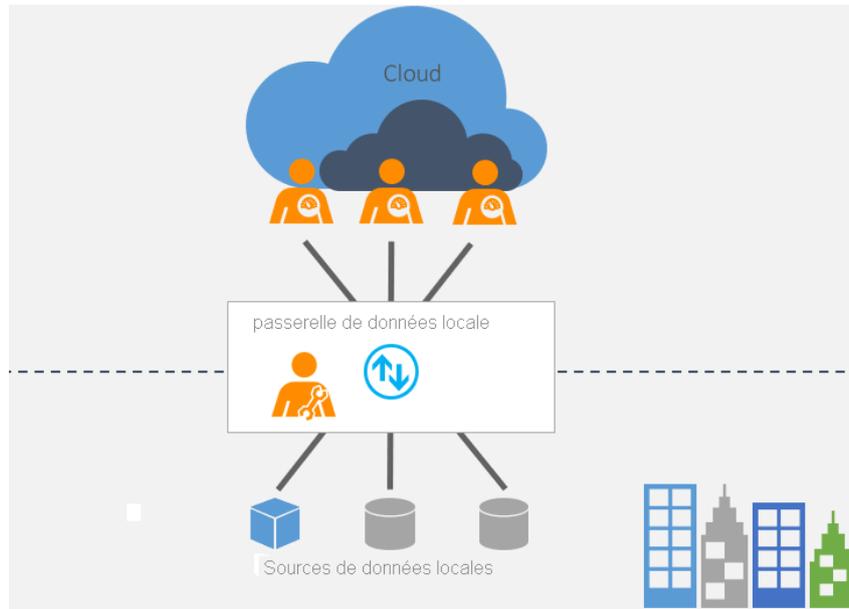
Lorsque les modèles sémantiques sont publiés dans l'espace de travail de Power BI d'une organisation, ils centralisent l'accès pour faciliter la collaboration et réduire les redondances, permettant ainsi une utilisation multiple pour divers besoins métier. Le partage efficace des modèles sémantiques accélère la productivité en minimisant le temps consacré à la préparation et au nettoyage des données. Pour optimiser les performances organisationnelles, il est crucial de gérer activement ces modèles en automatisant les actualisations pour garantir l'accès aux données les plus récentes et en promouvant les modèles les plus efficaces.

La gestion comprend également la configuration des paramètres des modèles pour adapter les sources de données ou améliorer les décisions métier, ainsi que la mise en place de passerelles pour accéder aux sources de données locales depuis le cloud. Il est essentiel de se préparer à résoudre tout problème de connectivité qui pourrait impacter l'accès aux modèles et, par conséquent, la productivité et la prise de décision.

Dans un scénario pratique, en tant que développeur Power BI chez Tailwind Traders, vous devez répondre aux demandes des utilisateurs pour plus de dynamisme dans les rapports et pour la mise en place de simulations. Vous êtes également chargé de maintenir la cohérence et l'intégrité des modèles sémantiques, de les centraliser pour une utilisation future et d'automatiser leur actualisation pour assurer des données toujours actualisées. Ce module vous enseignera à utiliser une passerelle Power BI pour des sources locales, configurer des actualisations planifiées et incrémentielles, promouvoir et gérer les modèles sémantiques, résoudre les problèmes de connectivité et améliorer les performances via la mise en cache des requêtes.

## Utiliser une passerelle Power BI pour se connecter à des sources de données locales

La passerelle Power BI est un logiciel qui sert de pont entre les bases de données locales et les services cloud tels que Power BI et Microsoft Azure Analysis Services, permettant l'accès sécurisé aux données locales depuis le cloud. Il existe deux types de passerelles :



1. **Mode Organisation** : Adapté à des scénarios complexes, ce mode permet à plusieurs utilisateurs de se connecter à plusieurs sources de données locales.
2. **Mode Personnel** : Utilisé par un seul utilisateur pour se connecter à des sources de données. Il est spécifique à Power BI et ne peut être partagé avec d'autres utilisateurs. Ce mode nécessite que l'ordinateur local sur lequel la passerelle est installée reste en ligne.

### Installation et Configuration de la Passerelle :

- L'installation et la configuration de la passerelle nécessitent généralement un administrateur. Une fois installée, la passerelle est démarrée et connectée via un compte Microsoft 365.

### Fonctionnement de la Passerelle :

- Les requêtes et les informations d'identification chiffrées sont envoyées de Power BI à la passerelle locale via Microsoft Azure Service Bus.
- La passerelle déchiffre ces informations et établit une connexion sécurisée avec les sources de données locales.
- Les requêtes sont exécutées localement et les résultats sont renvoyés via la passerelle au service cloud pour utilisation.

### Résolution des Problèmes :

- Les problèmes avec la passerelle peuvent nécessiter des ajustements dans les paramètres de communication, la configuration des informations de proxy, ou la vérification de la région du centre de données. Ces aspects sont cruciaux pour le bon fonctionnement de la passerelle et nécessitent des références à des documentations à jour pour une résolution efficace des problèmes.

## Configurer une actualisation planifiée d'un modèle sémantique

La fonctionnalité d'Actualisation planifiée de Power BI permet de programmer la fréquence et les horaires d'actualisation d'un jeu de données, rendant inutile l'actualisation manuelle et garantissant l'accès à des données à jour. Si des données ne sont pas récentes, comme observé dans un rapport de ventes actualisé il y a 10 jours, vous pouvez configurer une actualisation planifiée pour éviter d'oublier de mettre à jour manuellement le modèle sémantique. Cela implique d'installer une connexion de passerelle, d'activer l'actualisation planifiée via les paramètres du modèle sémantique, de choisir la fréquence et les horaires spécifiques, et d'appliquer ces paramètres.

En plus des actualisations planifiées, les actualisations à la demande sont disponibles pour rafraîchir les données immédiatement, sans affecter les horaires planifiés. Il est aussi crucial de vérifier régulièrement l'état et l'historique d'actualisation pour identifier tout problème, comme des erreurs d'actualisation ou des informations d'identification expirées, pouvant désactiver la planification après plusieurs échecs.

## Configurer les paramètres d'actualisation incrémentielle

La fonctionnalité d'actualisation incrémentielle de Power BI permet une mise à jour rapide et efficace des grands modèles sémantiques sans nécessiter de recharger les données historiques complètes à chaque fois. Cette méthode ne fonctionne que pour les sources de données qui supportent le Query Folding. Elle permet de rafraîchir uniquement les données modifiées récemment, réduisant ainsi le temps d'actualisation et la consommation de ressources.

Pour implémenter l'actualisation incrémentielle, vous devez:

1. Définir les paramètres de filtrage tels que RangeStart et RangeEnd dans Power Query pour limiter les données à la fenêtre de temps désirée.
2. Appliquer ces filtres à la colonne Date de vos données.
3. Configurer la stratégie d'actualisation incrémentielle dans Power BI Desktop pour spécifier la quantité de données historiques à conserver et la fréquence des actualisations.
4. Publier le modèle sur le service Power BI pour mettre en œuvre l'actualisation incrémentielle.

Cette stratégie assure que seul un sous-ensemble pertinent de données est actualisé, ce qui optimise les performances et maintient la pertinence des données sans surcharger le système.

## Gérer et promouvoir des modèles sémantiques

La gestion efficace des modèles sémantiques dans Power BI est cruciale pour maintenir des données cohérentes et promouvoir une culture saine des données au sein de l'organisation. Cela implique de diriger les utilisateurs vers des modèles de qualité et de limiter leur réutilisation inappropriée. Power BI propose deux méthodes pour valider officiellement vos modèles sémantiques :

1. **Promotion** : Cette option permet de marquer les modèles sémantiques prêts pour une utilisation plus large au sein de l'organisation.
2. **Certification** : Un processus plus sélectif par lequel un modèle promu peut être certifié pour garantir sa fiabilité. Seuls les modèles sémantiques répondant à des critères stricts obtiennent cette certification, assurant ainsi leur autorité et fiabilité.

L'utilisation de ces fonctionnalités aide à clarifier et à orienter les utilisateurs vers les modèles appropriés, évitant confusion et erreurs, particulièrement dans les grandes organisations avec de nombreux utilisateurs et jeux de données. Les modèles peuvent être promus ou certifiés via les paramètres dans l'espace de travail de Power BI, où les badges appropriés indiquent leur statut promu ou certifié.

## Résoudre les problèmes de connectivité du service

Pour connecter des services cloud tels que SharePoint Online à Power BI, aucune passerelle n'est nécessaire car les données sont déjà hébergées dans le cloud. Il suffit de fournir les informations d'identification d'autorisation correctes pour établir la connexion. Si l'actualisation d'un rapport échoue, il est important de vérifier et de s'assurer que les informations d'identification utilisées pour la source de données sont à jour. En cas de problèmes avec les informations d'identification, des étapes supplémentaires doivent être prises pour diagnostiquer et résoudre ces problèmes. Pour des détails supplémentaires sur la résolution de ces types de problèmes, il est recommandé de consulter les guides ou documents spécifiques aux scénarios d'actualisation.

## Améliorer les performances avec la mise en cache des requêtes (Premium)

La mise en cache des requêtes dans Power BI Premium permet d'améliorer les performances des rapports en stockant localement les résultats des requêtes, réduisant ainsi la charge sur les modèles sémantiques et les ressources de capacité dédiée. Cette fonctionnalité est particulièrement utile pour des modèles peu fréquemment actualisés mais régulièrement consultés, car elle diminue le temps de chargement et augmente la rapidité des requêtes.

Pour activer la mise en cache des requêtes, accédez aux paramètres du modèle sémantique dans votre espace de travail et activez cette fonctionnalité spécifiquement pour les modèles que vous choisissez. Cette option est disponible uniquement pour les utilisateurs de Power BI Premium ou Power BI Embedded et elle est configurable

pour chaque modèle sémantique. Choisir d'activer la mise en cache des requêtes signifie que les résultats sont conservés spécifiquement par utilisateur et page de rapport, ce qui améliore les performances sans compromettre la sécurité.

Cependant, il est important de noter que l'activation de la mise en cache pour plusieurs modèles sémantiques peut affecter les performances lors des actualisations simultanées. Si désactivée, elle peut entraîner un léger retard puisque les requêtes doivent être exécutées directement sur le modèle sémantique sans bénéficier des résultats mis en cache.

## Résumé

Ce module a exploré diverses fonctionnalités de Power BI pour améliorer la gestion des modèles sémantiques, notamment :

1. **Options d'actualisation des données** : Automatisation de l'actualisation pour rendre le processus plus efficace.
2. **Fonctionnalités d'approbation** : Aide les utilisateurs à identifier les modèles sémantiques fiables et appropriés pour leur utilisation.
3. **Utilisation de la passerelle locale** : Fournit des solutions aux problèmes de connectivité potentiels.

Ces techniques sont essentielles pour faciliter l'accès aux modèles sémantiques, en assurer l'actualisation régulière, et soutenir la création de rapports et de tableaux de bord précis qui aident les utilisateurs à prendre des décisions informées en temps réel.



Continuez à faire du bon travail !

 Développer des rapports dynamiques avec Microsoft Power BI

 Gérer des modèles sémantiques dans Power BI

Vous avez remporté 2 réalisations !

Félicitations, mais que devez-vous faire ensuite ?