



Comment les outils de
Datafactory peuvent améliorer
l'efficacité de la récupération et
du traitement de la donnée.

JULIENNE Maxence

Technomade 31/07/2025

Sommaire :

- I. [Introduction](#)
 - Problématique
 - Objectif du rapport
- II. [Les outils de Datafactory](#)
 - Dataflow Gen2
 - Lakehouse
 - Data Pipeline
 - Tache Apache Airflow
 - Copiez le travail
 - Miroiring
 - Bibliothèque de variables
- III. [Exemples complets](#)
 - Vigneron (Excel)
 - Vigneron (SQLEXPRESS)
- IV. [Sécurité](#)
 - Problèmes de sécurité de Datafactory
 - Solution avec Pureview
- V. [Tarification](#)
 - Fonctionnement global
 - Grilles
 - Exemple
- VI. [Bilan](#)
- VII. [Documentation / tutoriels](#)
 - Lier un fichier Excel à un Lakehouse via un Dataflow
 - Ajouter une table dans ce Dataflow

Introduction

Les entreprises déjà clientes ou non de Technomade disposent souvent de manières différentes pour stocker et gérer leurs données. Ces pratiques, parfois non professionnelles et mal optimisées rendent difficile l'exploitation efficace des informations, en particulier lors de l'intégration dans le CRM. Les problèmes fréquents sont :

- Multiplication des fichiers, sources
- Risque d'erreur manuelle dans les valeurs ou les formats
- Tâches répétitives et chronophage
- Difficulté à centraliser et exploiter les données

La problématique à l'origine de ce rapport est donc : Comment les outils de Datafactory peuvent améliorer l'efficacité de la récupération et du traitement de la donnée.

En deuxième partie du rapport des tutoriels sur les outils présentés les plus utiles sont disponibles.

Les outils de Datafactory et leur utilité

Dans cette section, chaque outil de Datafactory sera présenté à travers une approche structurée :

- Identification du besoin métier auquel il répond
- Présentation de l'outil approprié
- Illustration par un exemple rapide
 - Dataflow Gen2
 - **Besoin** : interfaçage et transformation des données sans coder
 - **Description** : Interface visuelle qui permet de filtrer, fusionner, corriger et transformer des données de manière intuitive, ETL
 - **Exemple** : Récupérer un fichier Excel, enlever les ligne vide, corriger les noms et charger ces données propres dans un Lakehouse
 - Lakehouse
 - **Besoin** : Stocker des données brutes ou transformées dans un format structuré, consultable et optimisé pour l'analyse
 - **Description** : conteneur qui combine la flexibilité du Datalake (stockage massif de données) avec la structure et les performances d'un data Warehouse (requête SQL, intégration avec Power BI)
 - **Exemple** : Garder les fichiers de stock mensuels d'une entreprise sur le Lakehouse, en faire des analyses avec Power BI qui y est intégré (historisation).
 - Data Pipeline
 - **Besoin** : Automatiser les taches de récupération et de transformation des données
 - **Description** : Suite d'étapes programmées qui s'exécutent automatiquement à intervalles réguliers selon les déclencheurs
 - **Exemple** : Tous les matins, le pipeline prend les ventes de la veille, les nettoie et les envoie dans un Lakehouse et met à jour le rapport Power BI.
 - Tache Apache Airflow
 - **Besoin** : Orchestrer des étapes de traitement complexe avec des logiques conditionnelles
 - **Description** : Outil d'orchestration qui exécute des étapes en suivant des règles
 - **Exemple** : Après le dépôt d'un fichier, Airflow vérifie qu'il est complet, lance le traitement et envoie un mail de confirmation
 - Copy Activity
 - **Besoin** : Transférer des donnée entre deux sources sans transformation
 - **Description** : Outil simple qui déplace les données entre deux sources
 - **Exemple** : Copier des fichiers d'un Warehouse vers un Lakehouse toutes les heures en ne prenant que les enregistrements modifiés

- Mirroring
 - **Besoin** : accéder aux données d'une base SQL en temps réel dans Fabric sans les dupliquer
 - **Description** Révolution dans le traitement de données, chaque modification de la source est visible instantanément dans Power BI, cela remplace le schéma Dataflow + Pipeline + Power BI par une connexion directe et en temps réel. ([description](#))
 - **Exemple** : Le client ajoute une commande dans sa base de données, quelques secondes plus tard elle apparaît dans le rapport power BI sans toucher à rien
- Bibliothèque de variables
 - **Besoin** : Réutiliser des paramètres dans plusieurs traitement sans tout reconfigurer
 - **Description** : Espace de stockage des variables qui permet d'adapter un pipeline à plusieurs environnements
 - **Exemple** : Le nom du dossier varie en fonction de si je suis en développement ou en production mais le pipeline reste le même grâce à la variable

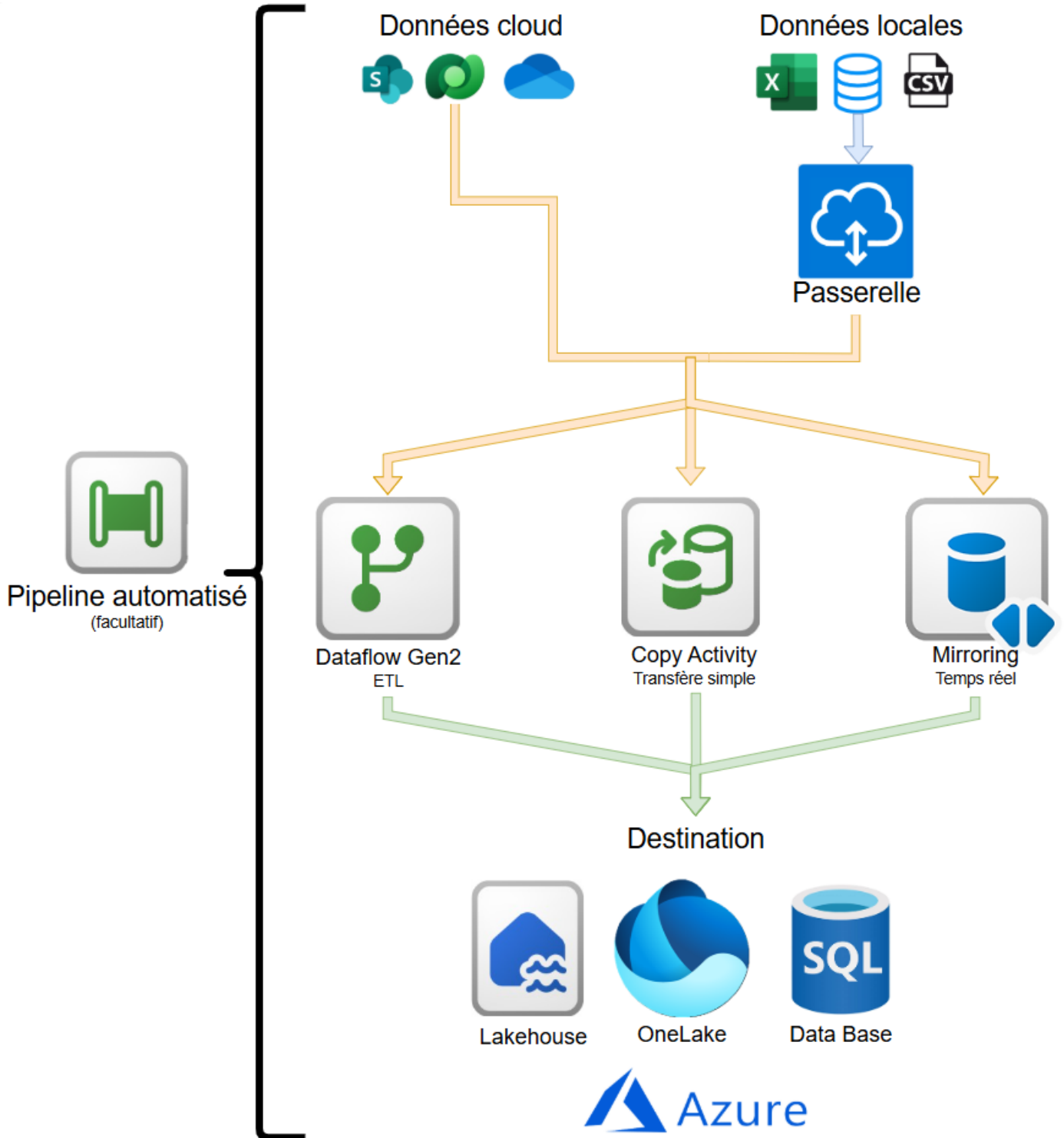


Schéma global des outils principaux de Data Factory

Exemples complets

Cas numéro 1 : fichiers Excel sur Onedrive

Contexte :

Pour illustrer l'apport des outils vu précédemment dans un cas concret, prenons l'exemple générique d'un propriétaire d'un domaine viticole. Il y fait de la production de vin et propose aussi la location de son domaine pour des événements. Depuis 15 ans, il utilise plusieurs tableaux Excel pour gérer toutes ses données (clients, réservations, production de vin...). Ces fichiers ne sont pas optimaux et trop compliqués à modifier pour lui.

Son CRM développé par Technomade correspond à ses besoins mais nécessite l'import manuel de ses fichiers Excel ce qui a plusieurs désavantages :

- Tache répétitive
- Risque d'erreur dans la saisie ou le choix du format
- Chronophage

Le propriétaire, peu familier avec les outils informatiques récents ne souhaite pas s'y former ni modifier ses habitudes avec Excel.

L'objectif est donc d'automatiser ce processus sans changer ses habitudes.

Ainsi, grâce à Datafactory, il deviendra possible de :

- Détecter automatiquement tout nouveau fichier ou la mise un jour des anciens fichiers sur OneDrive
- Intégrer, transformer, trier ses données afin qu'elles soient compatibles pour le CRM
- En bref, permettre au client de centraliser toutes ses données sur le CRM uniquement en rentrant ses informations dans son tableur.

Voici donc la mise en œuvre technique : le client continue à modifier ses fichiers Excel comme à son habitude, stockés sur OneDrive. Il en utilise 3 :

- clients pour ses clients classiques
- clients_speciaux pour ses clients VIP
- vin pour sa production viticole

Pour simplifier l'analyse, Technomade décide de centraliser les données dans deux tables seulement : client et vin. Une colonne supplémentaire client_special est ajoutée à la table clients pour fusionner les deux sources.

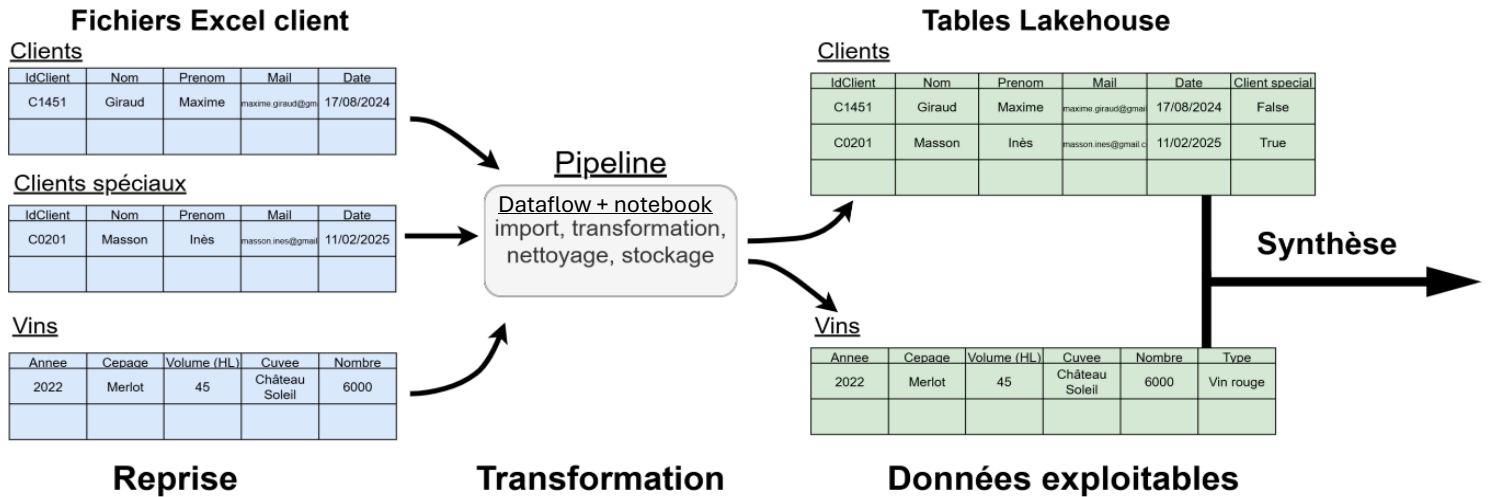
Un Lakehouse est donc créé avec ses deux tables dont les colonnes sont précisément adaptées aux besoins du client.

Chaque jour, à heure fixe, un Pipeline se lance automatiquement. Il orchestre un flux de travail : lance un Dataflow puis un notebook.

- Le Dataflow récupère, transforme les tables fusionne client et client_special et envoie toutes les données dans les tables correspondantes dans le Lakehouse
- Le notebook lance une suite d'opération (ex : supprimer les doublons de la table finale)

Résultat : Le client obtient chaque jour deux tables (clients et vin) parfaitement à jour, nettoyées et prêtes à être utilisées dans son CRM sans avoir dû modifier ses habitudes.

L'automatisation est invisible pour lui mais fait gagner un temps précieux et améliore la fiabilité de ses données.



Cas numéro 2 : Base de données SQL en local

Le client utilise déjà une base de données local, hébergée sur SQLEXPRESS, il l'administre grâce à SSMS (Visual studio). Etant stockée en local, elle n'est donc pas directement accessible via un Dataflow. Il y a alors plusieurs solutions (plus ou moins simples et efficaces) :

- Power BI desktop pour créer des rapports en local et pouvoir ensuite les envoyer sur le CRM (action manuelle)
 - L'outil Mirroring : récupère les données et actualise instantanément les rapports power BI
- La passerelle

Passerelle :

Le client ne change rien à ces habitudes, il continue de remplir sa base de données. Une passerelle installée sur son serveur SQL peut être appelée par le Dataflow. Il est ensuite possible de récupérer et traiter les données comme vu précédemment.

Mirroring :

Mirroring fonctionne aussi à l'aide d'une passerelle. Une fois configuré, il permet l'actualisation automatique des données sur le CRM dès la modification de la base de données. Le client modifie ou ajoute des valeurs dans ces tables, quelques secondes plus tard, ses rapports sont à jour.

Point faible : la passerelle doit rester ouverte, sinon le programme s'arrête et ne redémarre pas.

id_vente	id_client	id_vin	date_commande	quantité	prix_unitaire	prix_total	date_livraison	mode_paiement	notes
1	1	9	2024-01-20	10	480.00	4800.00	2024-01-25	carte	Livraison rapide
2	2	10	2024-01-22	1	420.00	420.00	2024-01-27	virement	NULL
3	3	6	2024-02-06	3	2500.00	7500.00	2024-02-10	virement	Commande exceptionnelle
4	4	15	2024-02-13	6	180.00	1080.00	2024-02-18	carte	NULL
5	5	28	2024-02-20	4	180.00	720.00	2024-02-25	espèces	Client fidèle
6	6	12	2024-03-02	2	350.00	700.00	2024-03-05	carte	NULL
7	7	13	2024-03-09	5	320.00	1600.00	2024-03-14	virement	Commande pro
8	8	4	2024-03-16	1	950.00	950.00	2024-03-21	chèque	NULL
9	9	14	2024-03-23	2	300.00	600.00	2024-03-28	carte	Livraison express
10	10	33	2024-04-03	10	75.00	750.00	2024-04-08	virement	NULL
11	11	26	2024-04-11	3	350.00	1050.00	2024-04-14	carte	NULL
12	12	11	2024-04-19	4	380.00	1520.00	2024-04-24	virement	Client entreprise
13	13	8	2024-05-02	2	1150.00	2300.00	2024-05-07	chèque	NULL
14	14	3	2024-05-09	1	1100.00	1100.00	2024-05-14	espèces	NULL
15	15	24	2024-05-16	6	85.00	510.00	2024-05-20	carte	NULL
16	16	35	2024-05-23	10	55.00	550.00	2024-05-28	virement	NULL
17	17	1	2024-06-02	1	850.00	850.00	2024-06-07	virement	Marriage
18	18	7	2024-06-09	3	1300.00	3900.00	2024-06-14	chèque	NULL
19	19	2	2024-06-16	1	1200.00	1200.00	2024-06-21	virement	NULL
20	20	25	2024-06-23	5	75.00	375.00	2024-06-28	carte	NULL
21	21	5	2024-07-02	2	1050.00	2100.00	2024-07-07	virement	NULL
22	22	38	2024-07-09	4	110.00	440.00	2024-07-13	chèque	NULL
23	23	23	2024-07-16	6	110.00	660.00	2024-07-21	carte	NULL

Données initiales stockées dans une base SQL



Redu final en quasi-temps-réel sur le CRM

Sécurité

Les outils de Datafactory sont des outils complets, cependant, ils demandent une certaine vigilance au niveau de la sécurité, en particulier sur :

- L'interaction avec des systèmes externes (bases SQL locales)
- L'usage de passerelles
- La gestion des identifiants pour les connexions

Datafactory bénéficie de la sécurité native de L'active directory de Microsoft, il est donc déjà bien protégé.

Toutefois, il est possible d'améliorer la sécurité en utilisant l'outil complémentaire Pureview.

(Toutes les informations suivantes sont tirées de Microsoft Learn sans véritable retour sur leur utilité, n'ayant pas de compte Azure, je n'ai pas eu l'occasion de les tester)

Microsoft Pureview est une plateforme puissante pour gérer et sécuriser des données. Il peut être intégré dans l'écosystème de Fabric. Il est possible de :

- Obtenir une visibilité sur les données dans l'ensemble de l'organisation
- Protéger et gérer les données sensibles tout au long de leur cycle de vie, où qu'elles se trouvent
- Gérer les données de manière transparente selon des méthodes nouvelles et complètes
- Gérer les risques liés aux données critiques et les exigences réglementaires

Voici la liste de toutes les fonctionnalités de Pureview :

- **Data Loss Prevention (DLP)**
 - **Besoin** : Empêcher le partage inapproprié de données sensibles avec des tiers non autorisés
 - **Description** : Identification et blocage automatique de contenus sensibles sur Teams, SharePoint, OneDrive...
 - **Exemple** : Un fichier Excel listant des numéros de carte bancaire est intercepté, chiffré et l'envoi est bloqué.
- **Data investigation**
 - **Besoin** : Comprendre rapidement ce qu'il s'est passé lors d'un incident de données
 - **Description** : Reconstitue la chronologie des actions à partir des logs
 - **Exemple** : On retrace les actions d'un utilisateur qui a copié des données sensibles sur un disque externe
- **Data Security Posture Management (DSPM)**
 - **Besoin** : Vérifier que le réseau cloud reste sécurisé
 - **Description** : Scanne les environnement cloud et repère les failles
 - **Exemple** : Alerte qu'un stockage Azure est publique

- Information Barriers
 - **Besoin** : restreindre la communication et la collaboration entre certains groupes spécifiques d'utilisateurs pour protéger des informations internes
 - **Description** : Empêche les communications entre différents groupes d'utilisateurs, sur Teams, SharePoint, OneDrive...
 - **Exemple** : Une équipe interne disposant de documents de secrets commerciaux ne doit pas appeler ou discuter en ligne avec les utilisateurs de certains groupes au sein de son organisation.

- Protection des données
 - **Besoin** : Classifier et chiffrer les données selon leur sensibilité
 - **Description** : Applique des étiquettes et chiffre les fichiers selon leur sensibilité
 - **Exemple** : Un rapport RH confidentiel est chiffré et nécessite une clé pour l'ouvrir

- Insider Risk Management
 - **Besoin** : Détecter et traiter les fuites de données par les employés
 - **Description** : Surveille les comportements à risque (export massif, accès hors horaire)
 - **Exemple** : Alerte si un employé télécharge 100 fichiers en 3 min.

- Privileged Access Management
 - **Besoin** : Limiter l'accès à des comptes à haut privilèges
 - **Description** : Fourni un accès temporaire, surveillé et soumis à des approbations sur des fichiers sensibles
 - **Exemple** : Un administrateur obtient 30 min d'accès pour une tâche critique, toutes ses actions sont surveillées

Tarification

La tarification des tous module de Fabric se fait uniformément de la même manière : avec des CU (Capacity unit). Leur prix change en fonction de la zone géographique.

Pour Data Factory, on paye uniquement les ressources de stockage et de calcul réellement utilisés, chaque outil de Data Factory a des grilles de facturation différentes. Les voici pour les principaux outils. (ctrl+clic sur l'image pour accéder à la page)

Data Pipeline :

Type de moteur de pipelines de données	Compteurs de charge et unités de métriques	Taux de consommation des unités de capacité d'infrastructure (CU)
Déplacement des données	En fonction de la durée d'exécution de l'activité Copy en heures et des ressources de débit d'optimisation intelligente utilisées	1,5 heures CU
Orchestration de données	Incorpore les exécutions d'activités d'orchestration et les frais d'exécution d'intégration d'activité	0,0056 heures CU pour chaque exécution d'activité non-Copy

Dataflow Gen2 :

Type de moteur Dataflow Gen2	Compteurs de consommation	Taux de consommation en CU Fabric	Granularité des rapports de consommation
Compute Standard	En fonction de chaque durée d'exécution de requête du moteur mashup en secondes.	16 CU	Par élément Dataflow Gen2
Calcul des flux de données à grande échelle	En fonction de la durée d'exécution (sans préproduction activée) du moteur SQL lakehouse/entrepôt en secondes.	6 CU	Par espace de travail
Déplacement des données	En fonction de la durée d'exécution de la copie rapide en heures et des ressources de débit d'optimisation intelligente utilisées.	1.5 unité(s) CU	Par élément Dataflow Gen2

Copy Data :

Modèles de copie	Compteurs de consommation	Taux de consommation des unités de capacité d'infrastructure (CU)	Granularité des rapports de consommation
Copie complète	Déplacement des données	1,5 heures CU	Élément par tâche de copie
Copie incrémentielle	Déplacement des données : copie incrémentielle	3 crédits heures	Élément par tâche de copie

Tâche Apache Airflow :

Taille du travail Apache Airflow (base)	Compteurs de consommation	Taux de consommation en CU Fabric	Granularité des rapports de consommation
Petite	DataWorkflow Petit	5 unités de requête	Par élément de travail Apache Airflow
Grande	DataWorkflow volumineux	10 unités de requête	Par élément de travail Apache Airflow

Exemple :

Charger 1 To de données CSV dans un Lakehouse via un pipeline qui utilise l'outil : copy data.

Durée : 290 secondes

Capacités demandées : 95 760 secondes = 26.6 CU heure

Prix CU : 0.18\$

Prix total = nombre CU heure * prix CU = 0.18 * 26.6 = 4.79 \$

Sur [cette page](#) un tutoriel pour estimer les couts d'infrastructure de Data Factory :

Voici une [calculatrice de tarification Azure](#)

Bilan

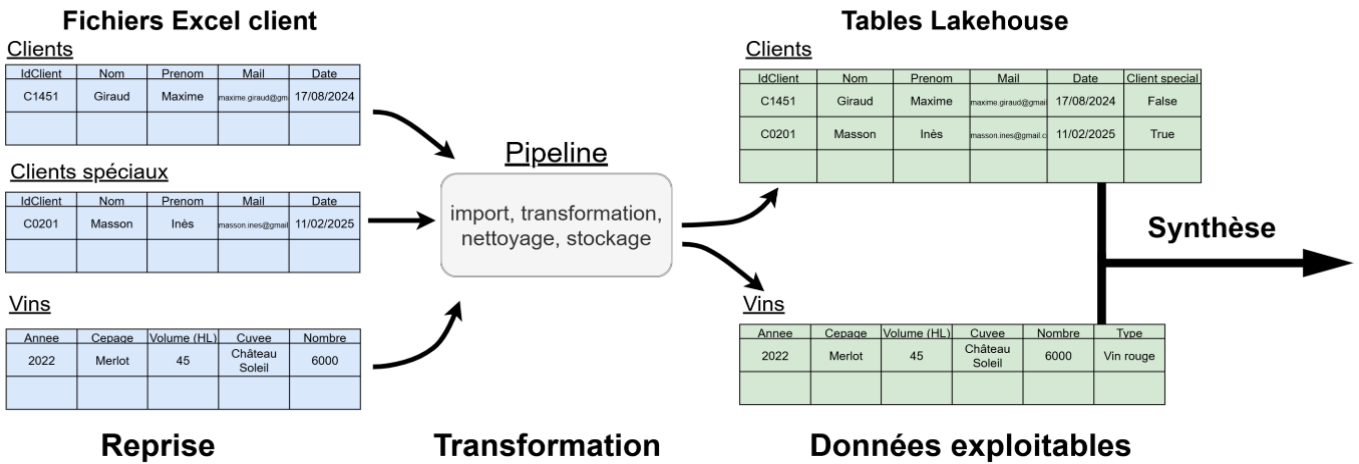
Les outils de Datafactory peuvent permettre de simplifier grandement la reprise, l'interfaçage et la transformation de données dans des cas simple comme dans des cas complexe. Ils offrent une flexibilité importante et une prise en main rapide et simple.

Pour conclure, Datafactory est un excellent outil d'intégration mais il demande une configuration rigoureuse pour un tirer le meilleur tout en garantissant la sécurité des données.

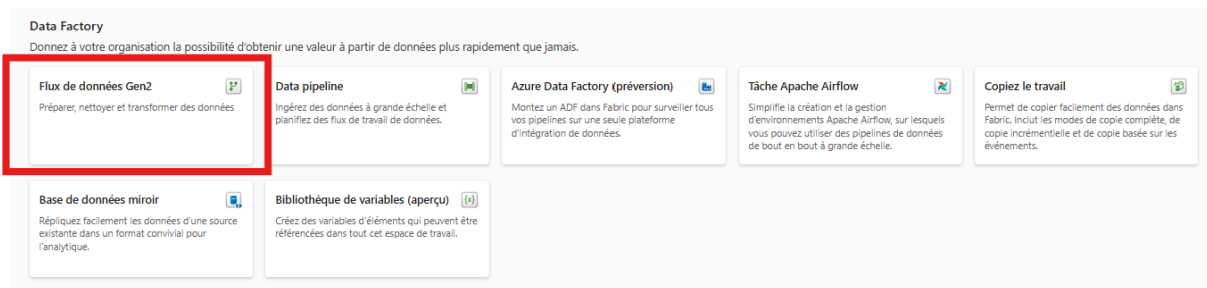
Documentation / tutoriels

Relier un fichier Excel à un Lakehouse via un pipeline

Nous allons voir comment mettre en place le processus de cette image avec cet exemple précis.



1. Créer la table nécessaire dans le Dataflow (voir documentation Dataflow)
2. Le client modifie son fichier Excel (A demander : est ce qu'il est possible d'avoir accès à ses fichiers OneDrive)
3. Créer un nouveau Dataflow Gen2



4. Choisir la source de donnée, ici Excel, choisir le fichier et configurer la connexion en faisant attention aux accès donnés.

Obtenir les données

Connexion à une source de données

Classeur Excel
Fichier
[En savoir plus](#)

Paramètres de connexion

Lien vers le fichier Charger le fichier

Chemin ou URL du fichier *

Exemple : Parcourir OneDrive...

Informations d'identification de connexion

Connexion

Nom de la connexion

Passerelle de données

Type d'authentification

Niveau de confidentialité

Cette connexion peut être utilisée avec des passerelles de données locales et des passerelles de données de réseau virtuel.

5. Sélectionner la table à importer

6. Composants clés de l'interface :

The screenshot shows the Power Query interface with the following components highlighted:


- 1:** The 'Requêtes' (Queries) pane on the left, showing the 'Feuil1' query.
- 2:** The 'Navigation 1' step in the 'Requêtes' pane, which is selected.
- 3:** The 'Type de colonne' (Column types) step in the 'Requêtes' pane, which is selected.
- 4:** The data table displayed in the main workspace, with columns like 'ID_Client', 'Nom', 'Prénom', 'Email', 'Date_Réservation', 'Type_Service', 'Nb_Personnes', and 'Montant (€)'. The table contains 17 rows of data.
- 5:** The 'Enregistrer et lancer' (Save and Run) button in the top-left corner of the interface.

- 1) Liste des requêtes.
- 2) Requête, on y voit les différentes étapes (source, navigation...). En cliquant sur une étape, on peut y voir son effet sur le tableau.
- 3) Sélection de la destination.
- 4) Tableau des données chargées, il est possible de cliquer sur une colonne afin de la cibler pour une étape supplémentaire.
- 5) Enregistrer et lancer.
7. Il est conseillé de directement changer le type des colonnes qui ne sont pas en raccord avec la table de destination avant d'ajouter d'autres étapes à la requête.

8. Ajouter les étapes nécessaires dans la requête en sélectionnant la/les colonnes à affecter. Ici ajouter une colonne personnalisée.

Colonne personnalisée ?

Ajoutez une colonne calculée en fonction d'autres colonnes ou valeurs.

 Dites-nous ce que vous souhaitez effectuer ⓘ ▶

Nouveau nom de colonne *

Type de données

Formule de colonne personnalisée * ⓘ

```
= let
    jour = Date.DayOfWeekName([Date_Réservation],
    "fr-FR"),
    groupe =
        if [Nb_Personnes] = 1 then "Solo"
        else if [Nb_Personnes] <= 4 then "Petit
groupe"
        else "Gros groupe"
    in
    jour & " - " & groupe
```

Colonne(s) disponible(s)

- ID_Client
- Nom
- Prénom
- Email
- Date_Réservation
- Type_Service
- Nb_Personnes

[En savoir plus sur les formules Power Query](#)

OK

Annuler


9. Choisissez ensuite la destination



10. Sélectionnez votre destination, ici Lakehouse

11. Entrez ensuite les paramètres de connexion en faisant attention au niveau de confidentialité.

Se connecter à la destination des données

 Lakehouse
Microsoft Fabric

Informations d'identification de connexion


Connexion

Nom de la connexion

Passerelle de données
aucun

Type d'authentification

Vous êtes actuellement connecté en tant que :

 **Maxence Julienne**
maxence.julienne@technomade.com
[Changer de compte](#)

Niveau de confidentialité

Cette connexion peut être utilisée avec des passerelles de données locales et des passerelles de données de réseau virtuel.

12. Sélectionner les données à exporter et où les envoyer.

Choisir la cible de destination

Nouvelle table Table existante

Rechercher

Options d'affichage

- Lakehouse [2]
 - My workspace
 - stage [2]
 - StagingLakehouseForDat...
 - Vigneron [4]
 - Fichiers
 - clients**
 - commandes
 - vins

La table clients dans Vigneron sera utilisée comme destination.

clients

ID_Client	Nom	Prenom	Email	Date_Reservation	Type_Service	Nb_Personnes	Montant	client_special
C001	bongo	Claire	claire.martin@email.fr	15/07/2024	Nuitée + Dégust.	2	180	TRUE
C002	Dubois	Julien	julien.dubois@email.fr	03/08/2024	Visite guidée	4	60	FALSE
C003	Lefèvre	Sophie	sophie.lefevre@email.fr	10/08/2024	Mariage	50	2500	FALSE
C004	Bernard	Paul	paul.bernard@email.fr	20/07/2024	Nuitée	2	160	FALSE
C005	Moreau	Elise	elise.moreau@email.fr	22/07/2024	Dégustation	3	90	FALSE
C006	Petit	Marc	marc.petit@email.fr	25/07/2024	Visite guidée	5	75	FALSE
C007	Garcia	Laura	laura.garcia@email.fr	28/07/2024	Nuitée + Spa	2	220	TRUE
C008	Lambert	Jean	jean.lambert@email.fr	01/08/2024	Mariage	80	4000	FALSE
C009	Faure	Alice	alice.faure@email.fr	05/08/2024	Nuitée	1	90	FALSE
C010	Mireau	Pierrooooo	Mireau.pierre@email.fr	05/07/2024	Nuitée + Spa	2	550	TRUE
C011	Colin	Emma	emma.colin@email.fr	09/08/2024	Visite guidée	3	45	FALSE
C012	Marchand Lucie	Lucie	lucie.marchand@email.fr	11/08/2024	Nuitée + Dégust	2	180	FALSE
C013	Noël	Antoine	antoine.noel@email.fr	13/08/2024	Nuitée	2	160	FALSE
C014	Henry	Clara	clara.henry@email.fr	15/08/2024	Mariage	60	3000	FALSE
C015	Giraud	Maxime	maxime.giraud@email.fr	17/08/2024	Visite guidée	6	90	FALSE
C016	Blin	Julie	julie.blin@email.fr	19/08/2024	Nuitée	1	80	FALSE
C017	Vidal	Thomas	thomas.vidal@email.fr	21/08/2024	Séminaire	15	950	FALSE
C018	Perrot	Chloé	chloe.perrot@email.fr	23/08/2024	Dégustation	4	120	FALSE
C019	Julienne	Yann	yann.lefort@email.fr	25/08/2024	Nuitée + Spa	2	210	FALSE
C020	Masson	Inès	ines.masson@email.fr	27/08/2024	Nuitée	2	160	FALSE
C021	Gabillé	Liam	liam.gabille@cacamail.com	28/08/2024	Nuitée+massage	1	10	FALSE

13. Faire le mappage adéquat.

Destination des données

Choisir les paramètres de destination

Pour améliorer les performances de la charge de données dans la destination, nous allons désactiver la copie intermédiaire de la requête source.

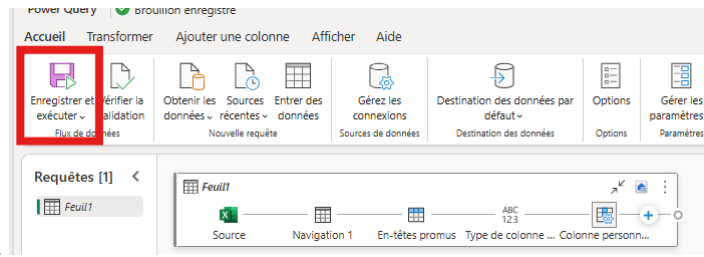
Méthode de mise à jour



Mappage de colonnes

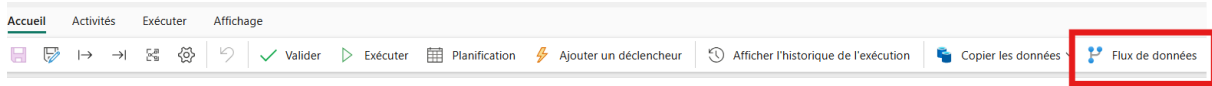
Source	Destination	Type
ID_Client	ID_Client	Texte
Nom	Nom	Texte
Prénom	Prenom	Texte
Email	Email	Texte
Date_Réservation	Date_Reservation	Date
Type_Service	Type_Service	Texte
Nb_Personnes	Nb_Personnes	Nombre entier
Montant (€)	Montant	Nombre décimal
client_spécial	client_special	Vrai/Faux

14. Possibilité de s'arrêter là. Appuyer sur enregistrer et exécuter enverra les données



transformées dans la destination.

15. Possibilité de créer un pipeline pour automatiser le processus, ajouter un flux de données.



Créer un pipeline de données pour organiser et déplacer vos données

Démarrer avec un canevas vierge

Activité de pipeline

Automatisez les orchestrations de données en tirant parti d'activités enrichies sans code.

Démarrer avec des conseils

Assistant Copier des données

Suivez les étapes guidées pour copier des données dans Microsoft Fabric, ainsi que d'autres magasins de données. Avez-vous besoin d'aide ? [Voir une démonstration](#)

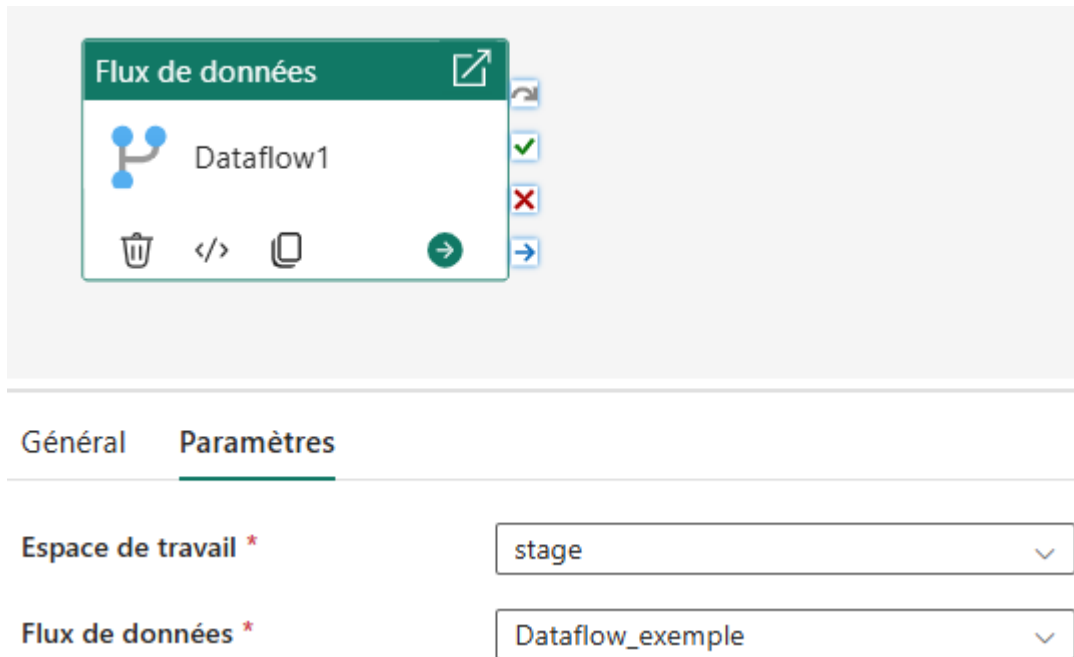
S'exercer avec des exemples de données

Générez rapidement un pipeline de données avec un modèle prédéfini pour charger des données dans Lakehouse.

Modèles

Générez rapidement un pipeline de données en utilisant un scénario de données prédéfini.

16. Modifier ensuite les paramètres du Dataflow.



17. Possibilité choisir des planifications pour son exécution, y rajouter des blocs d'action etc...

Ajouter une table au Dataflow depuis Excel

Le client souhaite rajouter une table et l'intégrer au fonctionnement déjà présent.

1. Le client créé sa table, l'enregistre sur le même OneDrive que les autres et prévient Technomade.

	A	B	C	D	E	F	G	H	I
1	ID_Commande	Date_Commande	Client	Cuvée	Année	Cépage	Quantité_Cor	Prix_Unitaire	Montant_Total (€
2	CMD001	01/06/2024	Dupont SA	Château Soleil	2022	Merlot	120	125,5	15060
3	CMD002	03/06/2024	Vins & Co	Nuit d'Or	2023	Cabernet	300	14	4200
4	CMD003	05/06/2024	Le Sommelier	Rouge Royal	2022	Syrah	500	13,3	6650
5	CMD004	06/06/2024	Domaine Lenoir	Blanc Prestige	2023	Sauvignon	250	11,1	2775
6	CMD005	07/06/2024	Caviste 33	Château Soleil	2021	Merlot	200	12	2400
7	CMD006	08/06/2024	Vins & Co	Nuit d'Or	2020	Cabernet	150	13,5	2025
8	CMD007	09/06/2024	Le Sommelier	Rouge Royal	2022	Syrah	100	1300	130000
9	CMD008	10/06/2024	Dupont SA	Blanc Prestige	2020	Sauvignon	180	10,8	1944
10	CMD009	11/06/2024	Domaine Lenoir	Château Soleil	2023	Merlot	220	12,8	2816
11	CMD010	12/06/2024	Caviste 33	Nuit d'Or	2021	Cabernet	300	13,7	4110

2. Créer la nouvelle table dans le Lakehouse à l'aide d'un notebook. Définir le Lakehouse comme

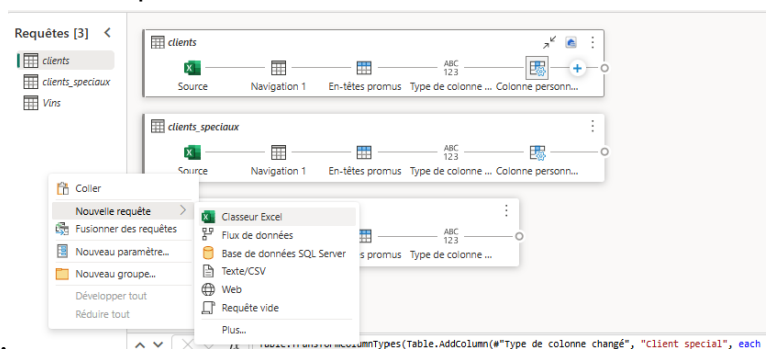
```

1 %%sql
2 -- Supprimer la table si elle existe déjà
3 DROP TABLE IF EXISTS commandes_vin;
4
5 -- Créer la table vide avec les bons types de données
6 CREATE TABLE commandes_vin (
7   ID_Commande      STRING,
8   Date_Commande   DATE,
9   Client           STRING,
10  Cuvée            STRING,
11  Année            INT,
12  Cépage           STRING,
13  Quantité_Commande INT,
14  Prix_Unitaire_EUR DECIMAL(10,2),
15  Montant_Total_EUR DECIMAL(12,2)
16 )
17 USING DELTA;

```

source de données.

3. Créer une nouvelle requête dans le Dataflow qui servira à connecter le fichier Excel du client à



la nouvelle table dans le Lakehouse.

- Choisir la source en faisant attention à la connexion utilisée et aux autorisations. Sélectionner ensuite la table que l'on veut importer.

Connexion à une source de données

Classeur Excel
Fichier
[En savoir plus](#)

Paramètres de connexion

Lien vers le fichier Charger le fichier (?)

Chemin ou URL du fichier *

 Parcourir OneDrive...

Informations d'identification de connexion

Connexion
 (?)

Nom de la connexion

Passerelle de données
 aucun

Type d'authentification

Vous êtes actuellement connecté en tant que :

MJ

Maxence Julienne
 maxence.julienne@technomade.com
[Changer de compte](#)

Niveau de confidentialité

Cette connexion peut être utilisée avec des passerelles de données locales et des passerelles de données de réseau virtuel.

- Ajouter les étapes de transformation dans la nouvelle requête que vous souhaitez.
- Renseigner la destination, ici dans le Lakehouse.

Requêtes [4] <

- clients
- clients_speciaux
- Vins
- Commandes

Source
 Navigation 1
 En-têtes promus
 Type de colonne ...
 Colonne personn...

Source
 Navigation 1
 En-têtes promus
 Type de colonne ...
 Colonne personn...

Source
 Navigation 1
 En-têtes promus
 Type de colonne ...

Source
 Navigation 1
 En-têtes promus
 Type de colonne ...

Destination par défaut

Aucune destination par défaut disponible

Nouvelle destination

- Lakehouse
- Warehouse
- Base de données SQL
- SharePoint (préversion)
- Plus...

Table.TransformColumnTypes(#"En-têtes pr

ID_Commande	Date_Commande	Client
1	01/06/2024	Dupont SA
2	03/06/2024	Vins & Co
3	05/06/2024	Le Sommelier
4	06/06/2024	Domaine Lenoir
5	07/06/2024	Caviste 33
6	08/06/2024	Vins & Co

7. Renseigner les options de connexion, puis choisir où exporter les données.

Destination des données
Choisir la cible de destination

Nouvelle table Table existante

Rechercher

Options d'affichage

- Lakehouse [2]
 - My workspace
 - stage [2]
 - StagingLakehouseForDat...
 - Vigneron [4]
 - Fichiers
 - clients
 - commandes
 - vins

① La table commandes dans Vigneron sera utilisée comme destination.

commandes

ID_Commande	Date_Commande	Client	Cuvee	Annee	Cepage	Quantite_Commandee	Prix_Unitaire_EUR	Montant_Total_EUR
? Cette table ne contient aucune ligne.								

8. Faire le mappage des colonnes et le choix d'ajouter ou de remplacer les données (si « table existante » sélectionnée précédemment).

Destination des données
Choisir les paramètres de destination

① Pour améliorer les performances de la charge de données dans la destination, nous allons désactiver la copie intermédiaire de la requête source.

Méthode de mise à jour

Données existantes → Nouvelles données → Ajouter / Remplacer

Mappage de colonnes

Source	Destination	Type
ID_Commande	ID_Commande	Texte
Date_Commande	Date_Commande	Date
Client	Client	Texte
Cuvée	Cuvee	Texte
Année	Annee	Nombre entier
aucun.	Cepage	Texte
Quantité_Commandée	Quantite_Commandee	Nombre entier
Prix_Unitaire (€)	Prix_Unitaire_EUR	Nombre décimal
Montant_Total (€)	Montant_Total_EUR	Nombre décimal

9. Exécuter ensuite le pipeline ou directement votre Dataflow.

Explorateur

tableaux de recherche

Vigneron

- table
- clients
- commandes
- vins
- Fichiers

commandes

Affichage Table

ID_Commande	Date_Commande	Client	Cuvee	Annee	Cepage	Quantite_Commandee	Prix_Unitaire_EUR	Montant_Total_EUR
1	6/10/2024 12:00:00 AM	Duport SA	Chateau Sirey	2023	Mérid	120	12.50	1500.00
2	6/9/2024 12:00:00 AM	Vins & Co	Nuit d'Or	2023	Cabernet	300	14.00	4200.00
3	6/9/2024 12:00:00 AM	Le Sommelier	Rouge Royal	2022	Sirah	500	13.30	6650.00
4	6/8/2024 12:00:00 AM	Domaine Lensez	Blanc Prestige	2023	Sauvignon	250	11.10	2775.00
5	6/7/2024 12:00:00 AM	Carole D	Chateau Sirey	2021	Mérid	200	12.00	2400.00
6	6/6/2024 12:00:00 AM	Vins & Co	Nuit d'Or	2020	Cabernet	150	13.50	2025.00
7	6/6/2024 12:00:00 AM	Le Sommelier	Rouge Royal	2022	Sirah	100	13000.00	13000.00
8	6/10/2024 12:00:00 AM	Duport SA	Blanc Prestige	2020	Sauvignon	180	13.00	2340.00
9	6/11/2024 12:00:00 AM	Domaine Lensez	Chateau Sirey	2023	Mérid	220	12.80	2816.00
10	6/12/2024 12:00:00 AM	Cavite 333333	Nuit d'Or	2021	Cabernet	300	13.70	4110.00

Affichage des 10 lignes

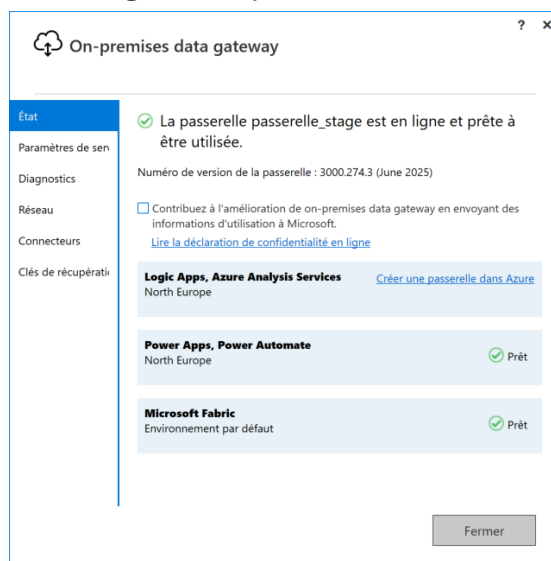
10. Constaté le résultat.

Mettre en place un mirroring (avec passerelle):

Pour ce tutoriel, les outils utilisés sont : SSMS, SQL Server Configuration Manager, On-premises data gateway.

Voici [la vidéo](#) à l'origine de ces informations.

- 1) Installer et configurer la passerelle avec le même compte que celui utilisé sur Fabric



- 2) Sur Fabric, se rendre dans Paramètres -> Gérer des connexions et des passerelles.
- 3) Créer et configurer sa connexion sur la passerelle avec ces paramètres
Conseil : pour éviter tout problème lié à la sécurité, créez un utilisateur dans la base de données auquel vous donnez juste les autorisations nécessaires, ici 'passerelle'



Nom du cluster de passerelle *

passerelle_stage

Nom de connexion *

connexion_exemple

Type de connexion *

SQL Server

Serveur * i

DELL-3F3FJL3\MSSQLSERVER01

Base de données * i

Maxence

Authentification i

Méthode d'authentification *

Base

Nom d'utilisateur *

passerelle

Mot de passe *

.....

Ignorer la connexion de test

Authentification unique

Utiliser SSO via Kerberos pour les requêtes DirectQuery i

Utiliser SSO via Kerberos pour les requêtes DirectQuery et d'importation i

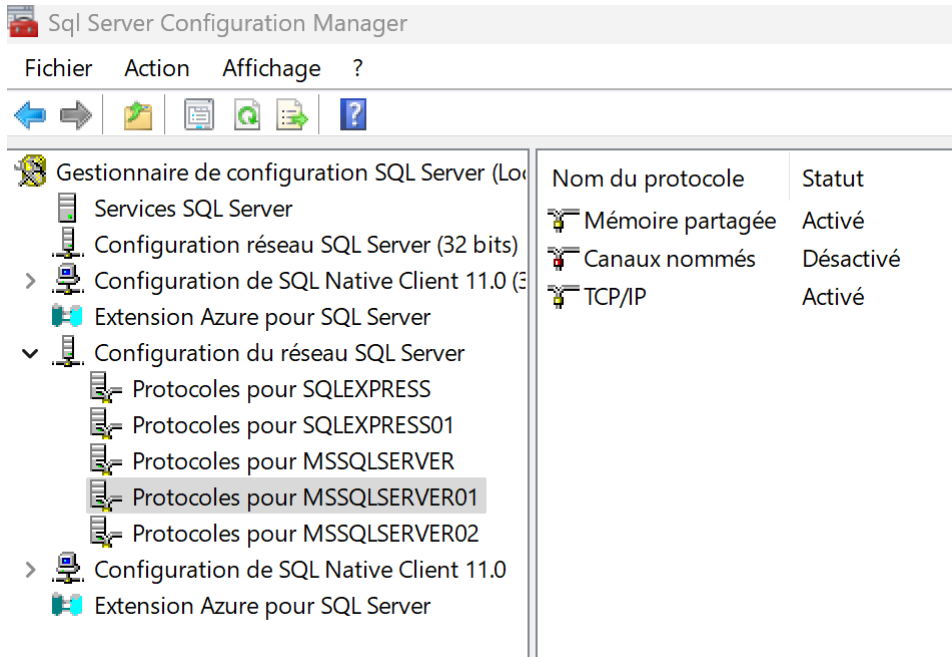
Utiliser SSO via Azure AD pour les requêtes DirectQuery i

Général

Créer

Fermer

- 4) Activer le TCP/IP sur votre serveur SQL puis le redémarrer, ne pas oublier de démarrer aussi l'agent



- 5) Activer le CDC sur la base en lançant les requêtes suivantes.
 CDC, (Change Data Capture) : méthode qui permet de détecter et transmettre uniquement les modifications apportées aux données d'une base : ajouts, suppression ou mise à jour.

```
USE Maxence --db
EXEC sys.sp_cdc_enable_db
```

- 6) Activer le CDC sur chaque table à miroiter

```
USE Maxence
EXEC sys.sp_cdc_enable_table
    @source_schema = N'dbo',
    @source_name = N'table_exemple', --nom table
    @role_name = NULL;
```

7) Vérifier si le CDC a bien été activé :

```
SELECT is_cdc_enabled, name FROM
sys.databases
WHERE name = 'Maxence'
```

Résultat attendu : is_cdc_enabled : 1
name : Maxence

8) Créer la base de données miroir sur Fabric.

Mirrored SQL Server (preview)



Répliquez facilement les données d'une source existante dans un format convivial pour l'analytique.

9) Repérer et sélectionner la connexion configurée plus tôt.

Choisir une connexion de base de données pour commencer

Rechercher

Nouvelles sources

- Base de données SQL Server

Catalogue OneLake

Tout | Recommandé

Nom	Propriétaire	Actualisé	Emplacement	Approbation	Confidentialité
connexion_exemple	Maxence Julienne	—	DELL-3F3JL3\MSSQ...	—	—
dell-3f3fj3\mssqlserver01,Maxenc...	Maxence Julienne	—	dell-3f3fj3\mssqlser...	—	—
dell-3f3fj3\mssqlserver01,Maxenc...	Maxence Julienne	—	dell-3f3fj3\mssqlser...	—	—
dell-3f3fj3\mssqlserver01,Maxenc...	Maxence Julienne	—	dell-3f3fj3\mssqlser...	—	—
dell-3f3fj3\sqlexpress01,Maxence	Maxence Julienne	—	dell-3f3fj3\sqllexpre...	—	—
Passerell	Maxence Julienne	—	dell-3f3fj3\mssqlser...	—	—

10) Choisir les données à importer et la destination.

Nouvelle mise en miroir SQL Server
Choisir des données

❗ Réplication des données Seules les colonnes valides sont répliquées vers Fabric OneLake. Vous pouvez définir vos contrôles et stratégies de sécurité à partir du point de terminaison d'analyse SQL. En savoir plus

Choisir des données

Rechercher

6 tables sur 6 sélectionnées

- Sélectionner tout
- dbo.clients
- dbo.locations_chateau
- dbo.sysdiagrams
- dbo.systranschemas
- dbo.ventes_vin
- dbo.vins

Aperçu des données: dbo.clients

id_client	nom	prenom	email	telephone	adresse	ville	code postal	pays	date_inscription	type_client
1	Dubois	Marie	marie.dubois@email.com	0145678901	12 rue de la Paix	Paris	75001	France	01/15/2024	particulier
2	Martin	Pierre	pierre.martin@email.com	0298765432	5 avenue Victor Hugo	Bordeaux	33000	France	01/20/2024	particulier
3	Rousseau	Sophie	sophie.rousseau@email.com	0387654321	8 place de la République	Lyon	69000	France	02/05/2024	professionnel
4	Moreau	Jean	jean.moreau@email.com	0476543210	15 boulevard Saint-Germain	Toulouse	31000	France	02/12/2024	particulier
5	Petit	Anne	anne.petit@email.com	0565432109	3 rue du Commerce	Marseille	13000	France	02/18/2024	cavisse
6	Durand	Paul	paul.durand@email.com	0654321098	7 avenue de la Liberté	Nice	06000	France	03/01/2024	particulier
7	Lefebvre	Catherine	catherine.lefebvre@email.com	0743210987	22 rue Nationale	Lille	59000	France	03/08/2024	professionnel
8	Roux	Michel	michel.roux@email.com	0832109876	11 place Bellecour	Lyon	69002	France	03/15/2024	particulier
9	Fournier	Isabelle	isabelle.fournier@email.com	0921098765	14 avenue Montaigne	Paris	75008	France	03/22/2024	particulier
10	Girard	Philippe	philippe.girard@email.com	0110987654	6 rue de la Gare	Nantes	44000	France	04/02/2024	cavisse

Miroir automatiquement les tables futures Toutes les nouvelles tables ajoutées à la source seront répliquées vers la destination.

Précédent Connecter

11) Une fois la base de données miroir créée, les données chargées doivent apparaitre.

Détails pour Maxence_ex

ID de base de données mise en miroir
6e1a7bd4-7bf8-4986-abc1-31c5d8d401b2

Source
Serveur langage SQL

Connexion source
connexion_exemple

Superviser la réplication

Voici les tables de la source. Une fois la réplication terminée, les tables sont ensuite synchronisées avec le point de terminaison d'analyse SQL de l'élément en miroir.

Vous pouvez également utiliser la surveillance de l'espace de travail avec les journaux d'activité des opérations de base de données en miroir pour obtenir des insights plus approfondis et activer la surveillance proactive.

État

● En cours d'exécution

Actualiser

Nom	État	Lignes répliquées	Dernière exécution terminée
[dbo].[sysdiagrams]	▲ Exécution avec avertissements	0	--
[dbo].[clients]	● En cours d'exécution	53	7/22/2025, 9:50 AM
[dbo].[locations_chateau]	● En cours d'exécution	8	7/22/2025, 9:50 AM
[dbo].[ventes_vin]	● En cours d'exécution	50	7/22/2025, 9:50 AM
[dbo].[vins]	● En cours d'exécution	52	7/22/2025, 9:50 AM

12) Les données sont ainsi miroitées en quasi-temps réel dans la base de données miroir. Il est ensuite possible d'y appliquer des requêtes SQL ou de faire une analyse Power BI.

ID	ID_client	nom	prenom	email	telephone	adresse	ville	code_postal	pays	date_inscription	type_client
1	1	Dubois	Marie	marie.dubois@email.com	0145575901	12 rue de la Paix	Paris	75001	France	2024-01-15	particulier
2	2	Marin	Pierre	pierre.marin@email.com	0326795432	5 avenue Victor Hugo	Bordeaux	33000	France	2024-01-20	particulier
3	3	Rousseau	Sophie	sophie.rousseau@email.com	031765321	8 place de la République	Lyon	69000	France	2024-03-05	professionnel
4	4	Moroux	Jean	jean.moroux@email.com	0479543210	15 boulevard Saint-Germain	Toulouse	31000	France	2024-02-12	particulier
5	5	Petit	Anne	anne.petit@email.com	056542109	3 rue du Commerce	Nantes	13000	France	2024-02-18	caisse
6	6	Durand	Paul	paul.durand@email.com	065421098	7 avenue de la Liberté	Nice	06000	France	2024-03-01	particulier
7	7	Leblond	Catherine	catherine.leblond@email.com	0743210987	22 rue Nationale	Lille	59000	France	2024-03-08	professionnel
8	8	Tan	Michel	michel.tan@email.com	083109876	11 place Belair	Lyon	69002	France	2024-03-15	particulier
9	9	Fournier	Isabelle	isabelle.fournier@email.com	0931098765	4 avenue Montaigne	Paris	75008	France	2024-03-22	particulier
10	10	Grand	Philippe	philippe.grand@email.com	011087654	6 rue de la Gare	Nantes	44000	France	2024-04-02	caisse
11	11	Faure	Sylvie	sylvie.faure@email.com	021987654	9 place Wilson	Dijon	21000	France	2024-04-10	particulier
12	12	Blanc	Laurent	laurent.blanc@email.com	0312876543	14 rue Gambetta	Strasbourg	67000	France	2024-04-18	professionnel
13	13	Lamy	Nathalie	nathalie.lamy@email.com	0401765432	18 avenue Jean Jaurès	Montpellier	34000	France	2024-05-01	particulier
14	14	Garnier	Alan	alan.garnier@email.com	051454321	21 rue de Ricci	Paris	75004	France	2024-05-08	particulier
15	15	Chevaller	Françoise	francoise.chevaller@email.com	061543210	23 boulevard Voltaire	Rennes	35000	France	2024-05-15	caisse
16	16	Pinson	Daniel	daniel.pinson@email.com	0715432109	30 rue Lafayette	Le Havre	76000	France	2024-05-22	particulier
17	17	Bourgeois	Monique	monique.bourgeois@email.com	0817321098	35 place Kléber	Strasbourg	67100	France	2024-06-01	professionnel
18	18	Vidal	François	francois.vidal@email.com	091210987	40 rue de la Pompe	Paris	75016	France	2024-06-08	particulier
19	19	Müller	Christine	christine.muller@email.com	018910876	45 avenue Foch	Rennes	35100	France	2024-06-15	particulier
20	20	Mercier	Bernard	bernard.mercier@email.com	029098765	50 rue Saint-Moroni	Paris	75001	France	2024-06-22	caisse
21	21	Guerin	Martine	martine.guerin@email.com	0391987654	55 place Massena	Nice	06000	France	2024-07-01	particulier
22	22	Lemoine	Gérard	gerard.lemoine@email.com	040276543	60 rue de la Paix	Caen	14000	France	2024-07-08	professionnel
23	23	Bonnet	Sylviane	sylviane.bonnet@email.com	0501765432	65 avenue Wilson	Tours	37000	France	2024-07-15	particulier
24	24	François	Thierry	thierry.francois@email.com	061454321	70 rue Victor Hugo	Clermont-Ferrand	63000	France	2024-07-22	particulier
25	25	Dupont	Véronique	veronique.dupont@email.com	071543210	75 place de la Comédie	Montpellier	34000	France	2024-08-01	caisse
26	26	Lemaire	Christophe	christophe.lemaire@email.com	081642109	80 rue Nationale	Arennes	80000	France	2024-08-08	particulier
27	27	Picard	Sarah	sarah.picard@email.com	091731098	85 avenue de la République	Beauvais	23000	France	2024-08-15	professionnel
28	28	Fleury	Patrice	patrice.fleury@email.com	019210987	90 rue du Faubourg Saint-Hippolyte	Paris	75008	France	2024-08-22	particulier
29	29	Rouere	Corinne	corinne.rouere@email.com	029109876	95 place Stanislas	Nancy	54000	France	2024-09-01	particulier
30	30	Lefevre	Olivier	olivier.lefevre@email.com	039098765	100 rue de la Liberté	Limoges	87000	France	2024-09-08	caisse
31	31	Brun	Dominique	dominique.brun@email.com	0491987654	105 avenue Gambetta	Perpignan	66000	France	2024-09-15	particulier
32	32	Caron	Sébastien	sebastien.caron@email.com	050276543	110 rue Saint-Hippolyte	Paris	75003	France	2024-09-22	professionnel
33	33	Barber	Valérie	valerie.barber@email.com	0601765432	115 place de Belleville	Nice	06100	France	2024-10-01	particulier
34	34	Renard	Hervé	herve.renard@email.com	071454321	120 rue de Belleville	Paris	75020	France	2024-10-08	particulier
35	35	Schmitt	Brigitte	brigitte.schmitt@email.com	081543210	125 avenue Jean Médecin	Nice	06000	France	2024-10-15	caisse

Présentation de Data Factory

Comment les outils de Data Factory peuvent améliorer l'efficacité de la récupération et du traitement de la donnée.

Sommaire

- I. Introduction
- II. Présentation de outils
- III. Exemples complets
- IV. Sécurité
- V. Bilan
- VI. Questions

I. Introduction

Problèmes :

- ▶ Multiplication des sources
- ▶ Risque d'erreur manuelle
- ▶ Difficulté à centraliser et exploiter les données

Problématique :

Comment les outils de Data Factory peuvent améliorer l'efficacité de la récupération et du traitement de la donnée.

II. Les outils de Data Factory

Présentations des outils de Data Factory en suivant ce schéma :

- ▶ Besoin
- ▶ Description
- ▶ Exemple

Lakehouse



- ▶ **Besoin** : Stocker des données brutes ou transformées dans un format structuré, consultable et optimisé pour l'analyse
- ▶ **Description** : Conteneur qui permet le stockage massif de données avec une structure propre et des bonnes performances. Les requêtes SQL et l'intégration avec Power BI y sont possibles.
- ▶ **Exemple** : Faire de l'historisation sur les stocks mensuels d'une entreprise

Dataflow Gen2



- ▶ **Besoin** : interfaçage et transformation des données sans coder
- ▶ **Description** : interface visuelle qui permet de filtrer, fusionner, corriger et transformer les données de manière intuitive. C'est un ETL.
- ▶ **Exemple** : Récupérer un fichier Excel, enlever les lignes vides, corriger les noms et charger ces données propres dans un Lakehouse

Data Pipeline



- ▶ **Besoin** : Automatiser des tâches successives de récupération et de transformation des données
- ▶ **Description** : Suite d'étapes programmées qui s'exécutent automatiquement selon les déclencheurs. Permet de faire des ETL plus complexes.
- ▶ **Exemple** : Tous les matins, le pipeline prend les ventes de la veille, nettoie les données et les envoie dans un Lakehouse et met à jour le rapport Power BI

Tâche Apache Airflow



- ▶ **Besoin** : Orchestrer des étapes de traitement avec des logiques conditionnelles
- ▶ **Description** : Outil d'orchestration qui exécute des étapes en suivant des règles
- ▶ **Exemple** : Après le dépôt d'un fichier, Airflow vérifie qu'il est complet, lance le traitement et envoie un mail de confirmation

Copy Activity



- ▶ **Besoin** : Transférer des données entre deux sources sans transformation
- ▶ **Description** : Outil simple qui déplace les données entre deux sources.
- ▶ **Exemple** : Copier des fichiers d'un Warehouse vers un Lakehouse toutes les heures en ne prenant que les enregistrements modifiés

Mirroring



- ▶ **Besoin** : Accéder aux données d'une base SQL en temps réel dans Fabric sans les dupliquer
- ▶ **Description** : Mise en miroir d'une source, chaque modification de cette source est visible instantanément dans Power BI. Remplace le schéma Dataflow + Pipeline + Power BI par une connexion directe en temps réel.
- ▶ **Exemple** : Le client ajoute des commandes dans sa base de données, quelques secondes plus tard, elle apparaît dans le rapport Power BI sans aucune intervention.

Bibliothèque de variables



- ▶ **Besoin** : Réutiliser des paramètres dans plusieurs traitements sans tout reconfigurer
- ▶ **Description** : Espace de stockage des variables qui permet d'adapter un pipeline à plusieurs environnements
- ▶ **Exemple** : Le nom du dossier varie en fonction de mon rôle, mais le pipeline reste le même grâce à une variable.

III. Exemple : Excel

- ▶ Propriétaire d'un domaine viticole : production + location.
- ▶ Stocke ses données de manière peu optimisée dans trois fichiers Excel :
 - ▶ Clients
 - ▶ Clients Spéciaux
 - ▶ Vins

Exemple : Excel

Problème de l'import manuel :

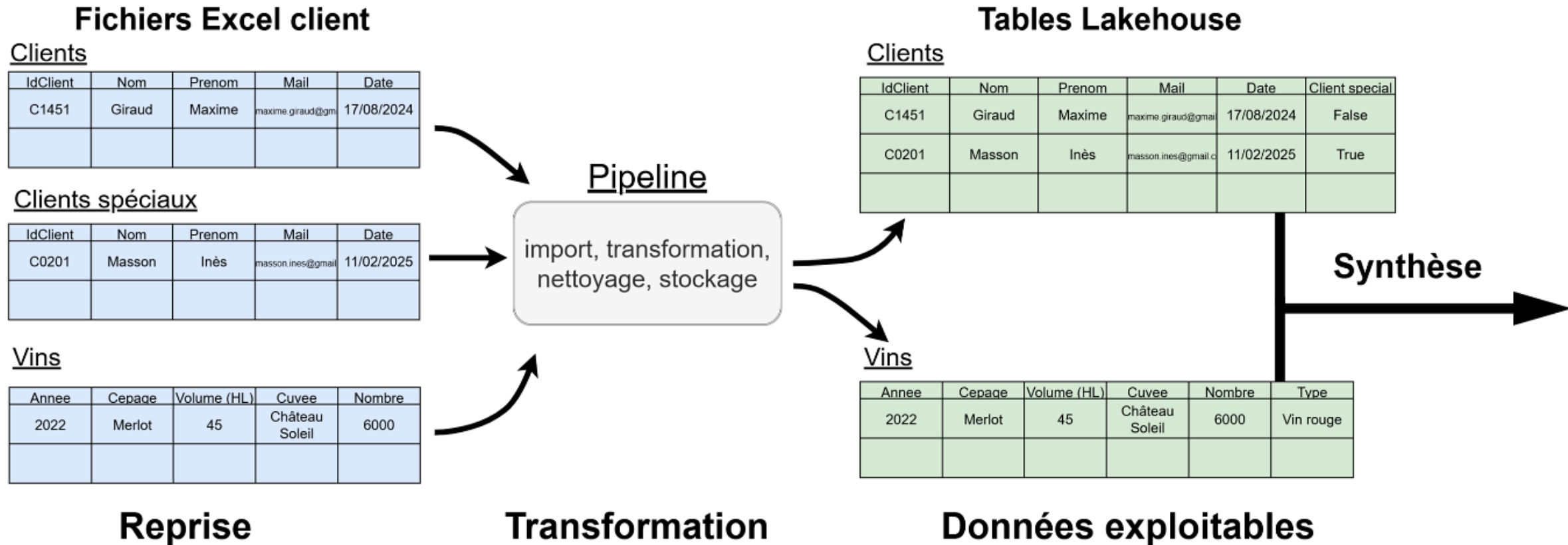
- ▶ Tache répétitive
- ▶ Risque d'erreur dans la saisie ou le choix du format
- ▶ Chronophage

Solution : Data Factory

- ▶ Détecter automatiquement la mise à jour d'un fichier
- ▶ Intégrer, transformer, trier ses données pour qu'elles soient compatibles pour le CRM
- ▶ Stocker les données et faire des rapports Power BI

Exemple : Excel

Mise en œuvre :



Exemple : Serveur SQL

- ▶ Problème : base de données stockée en locale
- ▶ Solutions :
 - ▶ Power BI desktop
 - ▶ Utilisation d'une passerelle
 - ▶ Mirroring

Exemple : Serveur SQL, passerelle

- ▶ Le client continue à remplir sa base SQL
- ▶ Une passerelle installée sur ce serveur accède aux données
- ▶ Un Dataflow est capable d'utiliser cette passerelle comme source

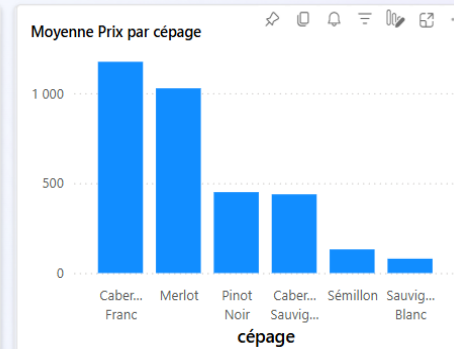
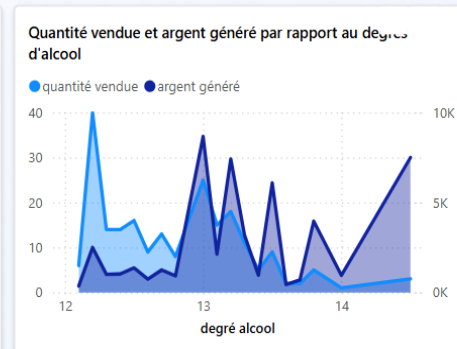
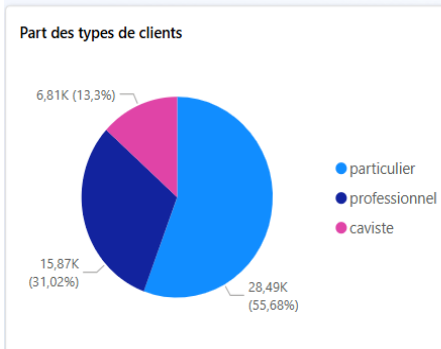
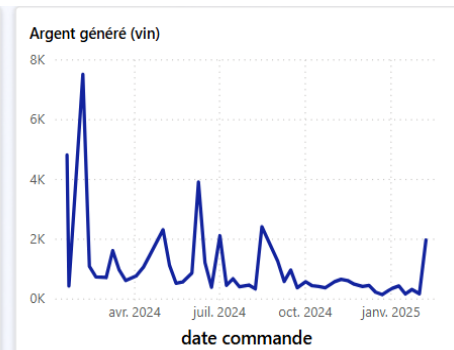
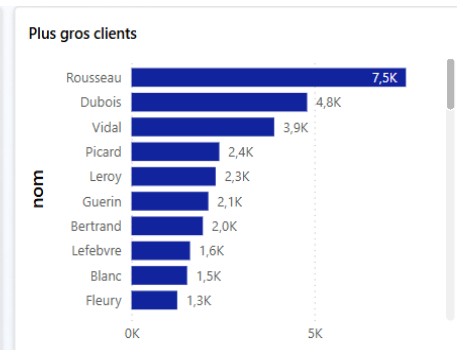
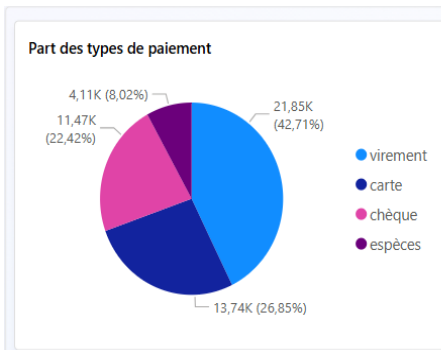
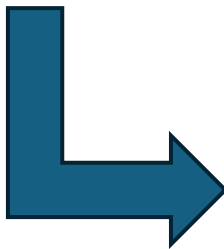
Exemple : Serveur SQL, Mirroring

- ▶ Le client remplit sa base SQL
- ▶ Mirroring, branché sur une passerelle détecte la modification
- ▶ Il charge les données dans Power BI
- ▶ Actualise le rapport en 1min

Résultat :

id_vente	id_client	id_vin	date_commande	quantité	prix_unitaire	prix_total	date_livraison	mode_paiement	notes
1	1	9	2024-01-20	10	480.00	4800.00	2024-01-25	carte	Livraison rapide
2	2	10	2024-01-22	1	420.00	420.00	2024-01-27	virement	NULL
3	3	6	2024-02-06	3	2500.00	7500.00	2024-02-10	virement	Commande exceptionnelle
4	4	15	2024-02-13	6	180.00	1080.00	2024-02-18	carte	NULL
5	5	28	2024-02-20	4	180.00	720.00	2024-02-25	espèces	Cient fidèle
6	6	12	2024-03-02	2	350.00	700.00	2024-03-05	carte	NULL
7	7	13	2024-03-09	5	320.00	1600.00	2024-03-14	virement	Commande pro
8	8	4	2024-03-16	1	950.00	950.00	2024-03-21	chèque	NULL
9	9	14	2024-03-23	2	300.00	600.00	2024-03-28	carte	Livraison express
10	10	33	2024-04-03	10	75.00	750.00	2024-04-08	virement	NULL
11	11	26	2024-04-11	3	350.00	1050.00	2024-04-14	carte	NULL
12	12	11	2024-04-19	4	380.00	1520.00	2024-04-24	virement	Cient entreprise
13	13	8	2024-05-02	2	1150.00	2300.00	2024-05-07	chèque	NULL
14	14	3	2024-05-09	1	1100.00	1100.00	2024-05-14	espèces	NULL
15	15	24	2024-05-16	6	85.00	510.00	2024-05-20	carte	NULL
16	16	35	2024-05-23	10	55.00	550.00	2024-05-28	virement	NULL
17	17	1	2024-06-02	1	850.00	850.00	2024-06-07	virement	Mariage
18	18	7	2024-06-09	3	1300.00	3900.00	2024-06-14	chèque	NULL
19	19	2	2024-06-16	1	1200.00	1200.00	2024-06-21	virement	NULL
20	20	25	2024-06-23	5	75.00	375.00	2024-06-28	carte	NULL
21	21	5	2024-07-02	2	1050.00	2100.00	2024-07-07	virement	NULL
22	22	38	2024-07-09	4	110.00	440.00	2024-07-13	chèque	NULL
23	23	23	2024-07-16	6	110.00	660.00	2024-07-21	carte	NULL

≈ 1 min



IV. Sécurité :

- ▶ Data Factory hérite de la sécurité native de l'active directory de Microsoft.
- ▶ Possibilité d'améliorer la sécurité en utilisant Pureview
- ▶ Capacités de Pureview :
 - ▶ Obtenir une visibilité sur les données dans l'ensemble de l'organisation
 - ▶ Protéger et gérer les données sensibles tout au long de leur cycle de vie, où qu'elles se trouvent
 - ▶ Gérer les données de manière transparente selon des méthodes nouvelles et complètes
 - ▶ Gérer les risques liés aux données critiques et les exigences réglementaires

Sécurité

Liste des outils :

- ▶ Data Loss Prevention
- ▶ Data investigation
- ▶ Data Security Posture Management
- ▶ Information Barriers
- ▶ Protection des données
- ▶ Insider Riks Management
- ▶ Privileger Access Management

V. Bilan

Les outils de Data Factory peuvent simplifier, la reprise, l'interfaçage et la transformation de données. Ils offrent une flexibilité importante et une prise en main rapide.

Data Factory propose d'excellents outils d'intégration mais demande une configuration rigoureuse pour en tirer le meilleur tout en garantissant la sécurité des données.

VI. Questions

Merci de m'avoir écouté, je suis disponible pour répondre à vos questions.