

Atelier N°4

Extraction de données à partir d'un fichier plat

Objectif :

Dans cet atelier, vous allez créer un package ETL qui extrait des données à partir d'un fichier source, transforme les données en utilisant deux composants de transformation de recherche, et charge les données dans la table de faits FactCurrencyRate dans AdventureWorksDW2012.

Source de données

La source de données est un ensemble de données de change historiques contenues dans le fichier **SampleCurrencyData.txt**. La source de données possède les quatre colonnes suivantes : le taux moyen de la monnaie, une clé de monnaie, une clé de la date, et le taux de fin de journée.

Voici un exemple des données contenues dans le fichier **SampleCurrencyData.txt** :

```
1.00010001    ARS    9/3/2001    0:00    0.99960016
1.00010001    ARS    9/4/2001    0:00    1.001001001
1.00020004    ARS    9/5/2001    0:00    0.99990001
1.00020004    ARS    9/6/2001    0:00    1.00040016
1.00050025    ARS    9/7/2001    0:00    0.99990001
1.00050025    ARS    9/8/2001    0:00    1.001001001
1.00050025    ARS    9/9/2001    0:00    1
1.00010001    ARS    9/10/2001   0:00    1.00040016
1.00020004    ARS    9/11/2001   0:00    0.99990001
1.00020004    ARS    9/12/2001   0:00    1.001101211
```

Destination

La destination des données est la table de faits **FactCurrencyRate** dans **AdventureWorksDW2012**. La table de faits **FactCurrencyRate** possède quatre colonnes, et est reliée à deux tables de dimension, comme indiqué dans le tableau suivant.

Column Name	Data Type	Lookup Table	Lookup Column
AverageRate	float	None	None

CurrencyKey	int (FK)	DimCurrency	CurrencyKey (PK)
DateKey	Int (FK)	DimDate	DateKey (PK)
EndOfDayRate	float	None	None

Étape 1 : Création d'un nouveau projet Integration Services

La première étape dans la création d'un package dans **Integration Services** est de créer un projet **Integration Services**.

1. Ouvrir **SQL Server Data Tools**
2. Dans le menu **Fichier**, pointer sur **Nouveau** et cliquer sur **Projet** pour créer un nouveau projet **Integration Services**.
3. Dans la boîte de dialogue Nouveau projet, sélectionner **Projet Integration Services** dans le volet Modèles.
4. Dans la zone Nom, remplacer le nom par défaut pour **SimpleETLSSIS**.
5. Cliquer sur **OK**.
6. Par défaut, un package vide, intitulé **Package.dtsx**, est créé et ajouté à votre projet.
7. Dans la barre d'outils Explorateur de solutions, cliquer droit sur **Package.dtsx**, cliquer sur **Renommer** et renommer le package par défaut en **SimpleETL.dtsx**.

Étape 2 : Ajout et configuration d'un gestionnaire de connexions de fichiers

Dans cette étape, vous ajoutez un **gestionnaire de connexions de fichiers** au package que vous venez de créer. Un gestionnaire de connexions de fichiers plats permet à un package d'extraire des données à partir d'un fichier plat. En utilisant le gestionnaire de connexions de fichiers, vous pouvez spécifier le nom de fichier et l'emplacement, le code de page et les paramètres et le format de fichier, y compris les séparateurs de colonnes. En outre, vous pouvez spécifier manuellement le type de données pour les colonnes individuelles, ou utiliser la boîte de dialogue **Proposer des types de données** pour mapper automatiquement les colonnes de données extraites à des types de données Integration Services.

Pour ajouter un gestionnaire de connexions de fichiers au package SSIS

1. Cliquer-droit n'importe où dans la zone **Gestionnaires de connexion**, puis cliquer sur **New Flat File Connexion**.
2. Dans la boîte de dialogue Éditeur du gestionnaire de connexions de fichiers, pour le nom du gestionnaire, taper **Currency Flat File Data Source**.
3. Cliquez sur **Parcourir**.
4. Dans la boîte de dialogue **Ouvrir**, rechercher le fichier **SampleCurrencyData.txt** sur votre machine.
5. Effacer les noms de colonnes dans la première case de la ligne de données.

Pour définir les propriétés des paramètres régionaux

1. Dans la boîte de dialogue Éditeur du gestionnaire de connexions de fichiers, cliquez sur **Général**.

2. Définir les paramètres régionaux en anglais (États-Unis) et la CodePage en 1252.
3. Vérifiez que la case à cocher « colonne name in the first row » n'est pas cochée dans columns vérifiez que le délimiteur est une tabulation au lieu d'une virgule

Pour renommer des colonnes dans le gestionnaire de connexions de fichiers plats

1. Dans la boîte de dialogue Éditeur du gestionnaire de connexions de fichiers, cliquer sur **Avancé**.
2. Dans le volet des propriétés, effectuez les modifications suivantes :
 - Changer le nom de la **colonne 0** à **AverageRate**.
 - Changer le nom de la **colonne 1** à **currencyID**.
 - Changer le nom de la **colonne 2** à **CurrencyDate**.
 - Changer le nom de la **colonne 3** à **EndOfDayRate**.

Pour remapper les types de données de colonne

1. Dans la boîte de dialogue Éditeur du gestionnaire de connexions de fichiers, cliquez sur **Suggérer les types**.
2. Dans ce TP, Integration Services suggère les types de données figurant dans la deuxième colonne du tableau ci-dessous pour les données du fichier SampleCurrencyData.txt. Cependant, les types de données qui sont nécessaires pour les colonnes de la destination, et qui seront définis lors d'une étape ultérieure, sont présentés dans la dernière colonne du tableau.

Flat File Column	Suggested Type	Destination Column	Destination Type
AverageRate	float [DT_R4]	FactCurrency.AverageRate	float
CurrencyID	string [DT_STR]	DimCurrency.CurrencyAlternateKey	nchar(3)
CurrencyDate	date [DT_DATE]	DimDate.FullDateAlternateKey	date
EndOfDayRate	float [DT_R4]	FactCurrency.EndOfDayRate	float

Le type de données suggéré pour la colonne **currencyID** est incompatible avec le type de données du champ dans la table de destination. Parce que le type de données de **DimCurrency.CurrencyAlternateKey** est **nchar(3)**, **currencyID** doit être changé de **string[DT_STR]** en **unicode string[DT_WSTR]**. En outre, le champ **DimDate.FullDateAlternateKey** est défini comme un type de données **Date** ; Par conséquent, **CurrencyDate** doit être changé de **date[DT_DATE]** en **database date[DT_DBDATE]**.

3. Dans la liste, sélectionner la colonne **currencyID** et dans le volet propriétés, modifier le type de données de la colonne **currencyID** de **string[DT_STR]** à **unicode string[DT_WSTR]**.
4. Dans le volet propriétés, modifier le type de données de la colonne **CurrencyDate** de **date[DT_DATE]** à **database date[DT_DBDATE]**.
5. Cliquez sur **OK**.

Étape 3 : Ajout et configuration d'un gestionnaire de connexions OLE DB

Après avoir ajouté un gestionnaire de connexion de fichier pour se connecter à la source de données, la tâche suivante consiste à ajouter un **gestionnaire de connexions OLE DB** pour se connecter à la destination. Un gestionnaire de connexions OLE DB permet à un package d'extraire ou de charger des données à partir d'une source de données OLE DB conforme. En utilisant le gestionnaire de connexions OLE DB, vous pouvez spécifier le serveur, la méthode d'authentification, et la base de données par défaut de la connexion.

Dans cette étape, vous allez créer un gestionnaire de connexions OLE DB qui utilise l'authentification Windows pour se connecter à l'instance locale d'**AdventureWorksDB2012**. Le gestionnaire de connexions OLE DB que vous créez sera également référencé par d'autres composants que vous allez créer plus tard dans ce TP.

Pour ajouter et configurer un gestionnaire de connexions OLE DB

1. Cliquer-droit n'importe où dans la zone **Gestionnaires de connexion**, puis cliquer sur **Nouvelle connexion OLE DB**.
2. Dans la boîte de dialogue **Configurer le gestionnaire de connexion OLE DB**, cliquer sur **Nouveau**.
3. Pour le nom du serveur, entrer **localhost**.
4. Vérifier que l'option **Utiliser l'authentification Windows** est sélectionnée.
5. Dans le connecter à un groupe de base de données, dans la zone Sélectionner ou entrer un nom de base de données, taper ou sélectionner **AdventureWorksDW2012**.
6. Cliquer sur **Tester la connexion** pour vérifier que les paramètres de connexion que vous avez spécifiées sont valides.
7. Cliquez sur **OK**.
8. Cliquez sur **OK**.
9. Dans le volet Connexions de données de la boîte de dialogue **Configurer le gestionnaire de connexion OLE DB**, vérifier que **localhost.AdventureWorksDW2012** est sélectionné.
10. Cliquez sur **OK**

Étape 4 : Ajout d'une tâche de flux de données au package

Après avoir créé les gestionnaires de connexions pour les données source et destination, la tâche suivante consiste à ajouter une tâche de flux de données à votre package. La tâche de flux de données encapsule le moteur de flux de données qui se déplace entre les données sources et destinations, et fournit la fonctionnalité pour la **transformation**, le **nettoyage** et la **modification** des données pendant leur déplacement. La tâche de flux de données est où la plupart des tâches d'extraction, de transformation et de chargement (ETL) se produisent.

Pour ajouter une tâche de flux de données

1. Cliquer sur l'onglet **Flux de contrôle**.
2. Dans la boîte à outils, glisser une **Tâche de flux de données** sur la surface de dessin de l'onglet Flux de contrôle.
3. Cliquer-droit sur la tâche de flux de données nouvellement ajouté, cliquez sur **Renommer** et modifier le nom à **Extract Sample Currency Data**.

4. Cliquer-droit sur la tâche de flux de données, cliquer sur **Propriétés**, et dans la fenêtre Propriétés, vérifier que la propriété **LocaleID** est l'**anglais (États-Unis)**.

Étape 5 : Ajout et configuration du fichier source

Dans cette étape, vous allez ajouter et configurer le fichier source de votre package. Un fichier source est un composant de flux de données qui utilise les métadonnées définies par un gestionnaire de connexions de fichiers pour spécifier le format et la structure des données à extraire du fichier par un processus de transformation.

Pour ce TP, vous allez configurer le fichier source pour utiliser le gestionnaire de connexions de fichiers que vous avez créé précédemment.

Pour ajouter un composant de fichier source

1. Ouvrir le **concepteur de flux de données**, soit en double-cliquant sur la tâche de flux de données **Extract Sample Currency Data** ou en cliquant sur l'onglet **Flux de données**.
2. Dans la boîte à outils, glisser une **source de fichier plat** sur la surface de dessin de l'onglet Flux de données.
3. Cliquer-droit sur la source de fichier plat nouvellement ajoutée, cliquer sur **Renommer** et modifier le nom en **Extract Sample Currency Data**.
4. Double-cliquer sur la source de fichier plat pour ouvrir la boîte de dialogue **Editeur de source de fichier plat**.
5. Dans le **gestionnaire de connexions de fichiers plats**, taper ou sélectionner **Sample Flat File Source Data**.
6. Cliquer sur **Colonnes** et vérifier que les noms des colonnes sont corrects.
7. Cliquer sur **OK**.
8. Cliquer-droit sur la source de fichier plat et cliquer sur **Propriétés**.
9. Dans la fenêtre Propriétés, vérifier que la propriété **LocaleID** est l'**anglais (États-Unis)**.

Étape 6 : Ajout et configuration des transformations de recherche

Après avoir configuré la source de fichier plat pour extraire les données du fichier source, la tâche suivante consiste à définir les transformations de recherche nécessaires pour obtenir les valeurs de la **CurrencyKey** et **DateKey**. Une transformation de recherche effectue une recherche en joignant les données de la colonne d'entrée spécifiée avec une colonne dans un ensemble de données de référence. L'ensemble de données de référence peut être une table ou une vue existante, une nouvelle table, ou le résultat d'une instruction SQL. Dans ce TP, la transformation de recherche utilise un gestionnaire de connexions OLE DB pour se connecter à la base de données qui est la source de l'ensemble de données de référence.

Pour ce TP, vous allez ajouter et configurer les deux composants de transformation de recherche suivants pour le package :

- Une transformation pour effectuer une recherche de valeurs de la colonne de **CurrencyKey** de la table de dimension **DimCurrency** basée sur l'adéquation entre ces valeurs et la colonne **currencyID** du fichier.

- Une transformation pour effectuer une recherche de valeurs de la colonne de **DateKey** de la table de dimension **DimDate** basée sur l'adéquation entre ces valeurs et la colonne **CurrencyDate** du fichier.

Dans les deux cas, les transformations de recherche utiliseront le gestionnaire de connexions OLE DB que vous avez créé précédemment.

Pour ajouter et configurer la transformation de recherche Currency Key

1. Dans la boîte à outils, glisser **Recherche** sur la surface de dessin de l'onglet **Flux de données**. Placer **Recherche** directement en dessous d'**Extract Sample Currency Data**.
2. Cliquer sur **Extract Sample Currency Data** et glisser la flèche bleu sur la transformation de recherche nouvellement ajoutée pour relier les deux composants.
3. Renommer le composant Recherche en **Lookup Currency Key**.
4. Double-cliquer sur la transformation **Lookup Currency Key**.
5. Sur la page **Général**, effectuer les sélections suivantes :
 - Sélectionner **cache complet**.
 - Dans la zone **Type de connexion**, sélectionnez gestionnaire de connexions OLE DB.
6. Sur la page de **connexion**, effectuer les sélections suivantes :
 - Dans la boîte de dialogue de gestionnaire de connexions OLE DB, veiller à ce que **localhost.AdventureWorksDW2012** est affiché.
 - Sélectionner **Utiliser les résultats d'une requête SQL**,

```

select * from (select * from [dbo].[DimCurrency]) as refTable
where [refTable].[CurrencyAlternateKey] = 'ARS'
OR
[refTable].[CurrencyAlternateKey] = 'AUD'
OR
[refTable].[CurrencyAlternateKey] = 'BRL'
OR
[refTable].[CurrencyAlternateKey] = 'CAD'
OR
[refTable].[CurrencyAlternateKey] = 'CNY'
OR
[refTable].[CurrencyAlternateKey] = 'DEM'
OR
[refTable].[CurrencyAlternateKey] = 'EUR'
OR
[refTable].[CurrencyAlternateKey] = 'FRF'
OR
[refTable].[CurrencyAlternateKey] = 'GBP'
OR
[refTable].[CurrencyAlternateKey] = 'JPY'
OR
[refTable].[CurrencyAlternateKey] = 'MXN'
OR
[refTable].[CurrencyAlternateKey] = 'SAR'
OR
[refTable].[CurrencyAlternateKey] = 'USD'
OR
[refTable].[CurrencyAlternateKey] = 'VEB'

```

7. Sur la page **Colonnes**, effectuer les sélections suivantes :
 - Dans le panneau **Colonnes d'entrée disponibles**, glisser **currencyID** dans le panneau **Colonnes de recherche disponibles** et le déposer sur **CurrencyAlternateKey**.
 - Dans la liste **Colonnes recherche disponibles**, activer la case à cocher à droite de **CurrencyKey**.
8. Cliquer sur **OK** pour revenir à l'espace de conception de flux de données.
9. Cliquer-droit sur la transformation **Lookup Currency Key**, cliquer sur **Propriétés**.
10. Dans la fenêtre **Propriétés**, vérifier que la propriété **LocaleID** est l'**anglais (États-Unis)** et la propriété **DefaultCodePage** est fixée à **1252**.

Pour ajouter et configurer la transformation de recherche DateKey

1. Dans la boîte à outils, glisser **Recherche** sur la surface de dessin de l'onglet **Flux de données**. Placer **Recherche** directement en dessous de la transformation **Lookup CurrencyKey**.
2. Cliquer sur **Lookup CurrencyKey** et glisser la flèche bleu sur la transformation de recherche nouvellement ajoutée pour relier les deux composants.
3. Renommer le composant Recherche en **Lookup Date Key**.
4. Double-cliquer sur la transformation **Lookup Date Key**.
5. Sur la page **Général**, sélectionner **cache partiel**.
6. Sur la page de **connexion**, effectuer les sélections suivantes :

7. Dans la boîte de dialogue gestionnaire de connexions OLE DB, veiller à ce que **localhost.AdventureWorksDW2012** est affiché.
8. Dans la zone **Utiliser une table ou une vue**, taper ou sélectionner **[dbo].[DimDate]**.
9. Sur la page **Colonnes**, effectuez les sélections suivantes :
 - a. Dans le panneau **Colonnes d'entrée disponibles**, glisser **CurrencyDate** dans le panneau **Colonnes de recherche disponibles** et le déposer sur **FullDateAlternateKey**.
 - b. Dans la liste **Colonnes recherche disponibles**, activer la case à cocher à droite de **DateKey**.
11. Cliquer sur **OK** pour revenir à l'espace de conception de flux de données.
12. Cliquer-droit sur la transformation **Lookup Date Key**, cliquer sur **Propriétés**.
13. Dans la fenêtre **Propriétés**, vérifier que la propriété **LocaleID** est l'**anglais (États-Unis)** et la propriété **DefaultCodePage** est fixée à **1252**.

Étape 7 : Ajout et configuration la destination OLE DB

Votre package peut maintenant extraire des données à partir du fichier source et transformer ces données dans un format qui est compatible avec la destination. L'étape suivante consiste à charger effectivement les données transformées dans la destination. Pour charger les données, vous devez ajouter une destination OLE DB pour le flux de données. La destination OLE DB peut utiliser une table de base de données, une vue ou une commande SQL pour charger des données les bases de données.

Dans ce TP, vous ajoutez et configurez une destination OLE DB à utiliser le gestionnaire de connexions OLE DB que vous avez créé précédemment.

Pour ajouter et configurer la destination OLE DB exemple

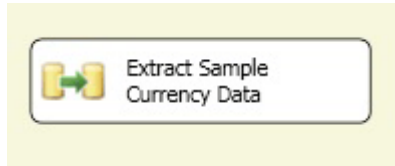
1. Dans la **boîte à outils SSIS**, glisser **destination OLE DB** sur la surface de dessin de l'onglet **Flux de données**. Placer la destination OLE DB directement en dessous de la transformation **Lookup Date Key**.
2. Cliquer sur **Lookup Date Key** et glisser la flèche bleu sur la transformation de recherche nouvellement ajoutée pour relier les deux composants.
3. Renommer le composant de destination en **Sample OLE DB Destination**.
4. Double-cliquer sur **Sample OLE DB Destination**.
5. Dans la boîte de dialogue **Éditeur de destination OLE DB**, veiller à ce que **localhost.AdventureWorksDW2012** est sélectionné dans la boîte de gestionnaire de connexions OLE DB.
6. Dans le nom de la table ou de la vue, taper ou sélectionner **[dbo].[FactCurrencyRate]**.
7. Cliquer sur le bouton **Nouveau** pour créer une nouvelle table. Changer le nom de la table en **NewFactCurrencyRate**. Cliquer sur **OK**.
8. Après avoir cliqué sur **OK**, la boîte de dialogue se ferme et le nom de la table ou de la vue passe automatiquement à **NewFactCurrencyRate**.
9. Cliquez sur **Mappings**.
10. Vérifier que les colonnes d'entrée **AverageRate**, **CurrencyKey**, **EndOfDayRate**, et **DateKey** sont mappés correctement aux colonnes de destination.
11. Cliquer sur **OK**.
12. Cliquer-droit sur **Sample OLE DB Destination** et cliquer sur **Propriétés**.

13. Dans la fenêtre **Propriétés**, vérifier que la propriété **LocaleID** est l'**anglais (États-Unis)** et la propriété **DefaultCodePage** est fixée à **1252**.

Étape 8 : Tester la transformation

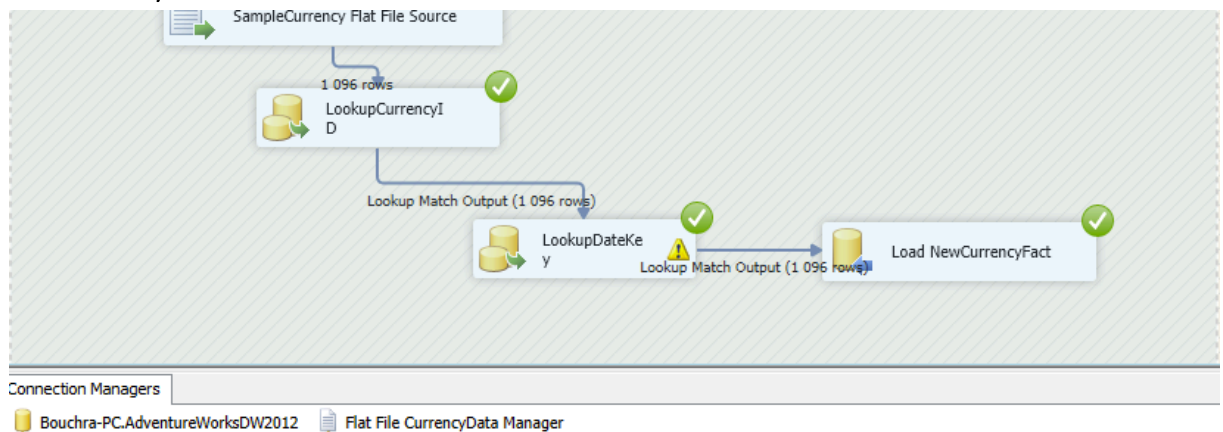
Avant de tester le package, vous devez vérifier que le contrôle et les flux de données dans le package **SimpleETL** contient les objets présentés dans les schémas suivants.

Flux de contrôle



Flux de données

1. Dans le menu **Déboguer**, cliquez sur **Démarrer le débogage**.
2. Le package doit s'exécuter, entraînant 1097 lignes ajoutées avec succès dans la table de faits **FactCurrencyRate** dans **AdventureWorksDW2012**.



Package execution completed with success. Click here to switch to design mode, or select Stop Debugging from the Debug menu.

- 3.
4. Après la fin de l'exécution, dans le menu **Déboguer**, cliquez sur **Arrêter le débogage**.