

## TP 2 : ETL, premiers pas

L'objectif de ce TP2 est de faire les premiers pas vers l'utilisation de l'ETL : SSIS de Visual Studio Data Tool.

### Phase 1 : Création du premier projet SSIS.

Lancez **Visual Studio Data Tool**.

Créez un nouveau projet, de template Business Intelligence / Integration Services / Integration Services Project for Business Intelligence. Nommez-le **ssis\_premier\_prj**.

Dans la zone **Solution Explorer** à droite, vous avez les différents composants de votre projet :

- Project.params : qui définit les paramètres du projet.
- Connections Managers : définit l'ensemble des connections (sql server, excel, fichier plat, etc.) à utiliser dans ce projet.
- SSIS Packages : définit l'ensemble des tâches que ce projet va exécuter. Ces tâches sont organisées par packages.
- Miscellaneous : ou Divers, sert pour autres choses, à voir plus tard.

### Phase 2 : Extraction.

#### 1. Connexions aux bases de données

Créez deux **Connexion Managers** pour nos **deux bases de données**. Pour ce faire, cliquez **droit** sur **Connection Managers**, puis **New Connection Manager**. Avant de choisir le type de ce manager, défilez cette liste et découvrez tous les types de connexions que SSIS vous fournit (en particulier, Excel, Flat File, HTTP, SMTP, FTP). Pour vous connecter à SQL Server, choisissez le type **ADO.NET** et cliquez sur **Add**.

La fenêtre qui suit vous demande de configurer ce Connection Manager, pour le connecter à votre serveur et à votre base de données. Cliquez sur **New**, puis comme **server name**, mettez « . » et passez directement à la zone en bas **Connect to Database** pour choisir une des bases de données gérées par ce serveur. Choisissez notre base de données du TP précédent « **GestionDB** », cliquez sur **Test the connection** puis **Add**. Si tout se passe bien. Remarquez que cette connexion s'ajoute à la liste des connexions à gérer par ce manager.

Ce manager est maintenant disponible pour tout le projet. Ce qui veut dire que dans tout le projet nous pourrons nous connecter à cette base de données via ce manager.

Maintenant nous allons **nous connecter** à la **base source** pour en extraire les données. Double cliquez sur le package principal **Package.dtsx** dans la zone à droite. Cette action va ouvrir ce package dans **l'espace de travail** au milieu.

## 2. Création du premier Data Flow

Activez l'onglet **Control Flow**, depuis la zone **SSIS Toolbox** à gauche, glissez-déplacez vers l'espace de travail le composant **Data Flow Task**. Cliquez droit sur ce composant et cliquez sur **Edit**. Ceci vous mènera vers le deuxième onglet **Data Flow**.

Remarquez que la liste des composants du **SSIS Tool Box** à gauche a changé. **Glissez déplacez** le composant **ADO NET Source** (que vous trouvez sous **Other Sources**).

## 3. Extraction des données depuis la base de données.

**Cliquez droit** sur ce composant dans votre espace de travail, puis **edit**. Cliquez ensuite sur **New** pour choisir un **Connection Manager** à associer à cette **ADO NET Source**. Vous allez retrouver le **Connection Manager** que vous avez créé auparavant. Choisissez celui de la base de données **GestionDB**. Une fois le Manager ajouté, choisissez comme Data Access Mode **Table or View**, et sélectionnez ensuite la table **dbo.Produit**. Cliquez sur **Preview** pour visualiser les données de cette table.

## 4. Enregistrement des données vers une destination.

Maintenant que la source a bien été configurée. Nous allons en extraire les données et les rediriger vers **une destination**. Comme première destination, nous allons choisir un **fichier plat** (Flat File). Dans votre espace, glissez-déplacer le composant **Flat File Destination**. Cliquez sur le premier composant, deux flèches en sortent du bas, un bleu et un rouge, **glissez le bleu et connectez-le avec le composant de destination**.

Nous allons maintenant **configurer la destination**. Cliquez **droit** sur le composant **Flat File Destination** dans votre espace de travail, puis **edit**. Cliquez ensuite sur **New** pour configurer le manager de ce type. Choisissez le premier format **Delimited**. Cliquez ensuite sur **Browse** pour spécifier le nom du fichier, choisissez le nom **Produits.txt** dans le **bureau** de votre machine, puis cochez **Column names in the first row data**, puis **Ok**.

## 5. Exécution de notre package

Cliquez sur l'**icône triangle** pour **exécutez**. Si tout se passe bien, ouvrez le fichier **Produits.txt** et vérifiez bien que les produits ont bien été enregistrés.

Pour revenir en mode modification à votre package, arrêtez l'exécution en cliquant sur le **bouton carré**.

## Phase 3 : Transformation et Chargement.

Jusqu'à maintenant nous avons exploité uniquement les deux fonctionnalités **Extract** and **Load** de l'ETL, Essayons d'exploiter dans la suite une des fonctionnalités de transformation : **Transform**.

## 6. Le sort.

Comme premier exemple, essayons de **trier les produits** avant de les enregistrer dans le fichier plat.

Tout d'abord, **créez un nouveau package**, dans lequel vous **créez la source de données** comme précédemment fait, puis la **destination**.

Glissez déplacer le composant **Sort** (depuis la Zone SSIS ToolBox de votre environnement) dans votre espace de travail. **Connectez** le composant de **la source** à ce composant **Sort**, puis ce composant **Sort** au composant de **destination**.

Dans l'interface **Edit** de ce composant **Sort** (clic **droit** sur le composant, puis **edit**), **cochez** les colonnes par lesquelles vous souhaitez faire le **tri**. Essayez de lire toutes les options de ce composant (sur l'interface d'Edit), puis cliquez sur **Ok**. Nous souhaitons enregistrer les produits triés dans l'ordre **croissant de leurs catégories**, et **décroissant de leurs libellés**.

Exécutez et vérifiez que le résultat est bien comme attendu.

## 7. Le Conditionnal Split.

Nous souhaitons maintenant créer autant de fichiers plats que de catégories de produits. On va utiliser le composant **Conditionnal Split**, ce composant prend des données en entrée, fait des tests que vous lui donnez et, selon le résultat, redirige vers une sortie que vous précisez.

Ajoutez ce **composant à votre environnement**, connectez-la sortie de la source vers ce composant. Puis, sur l'interface **Edit** de ce composant, **configurez les différents tests nécessaires**. Cette fois-ci je vous laisse découvrir comment faire cette tâche. Une fois les tests faits, cliquez sur **Ok**.

Créez un composant de **destination Flat File Destination**, nommez-le **Flat File Destination Catégorie 1**. Connectez la sortie du composant de Split vers le composant Flat File Destination Catégorie 1, une interface vous demandant sous quelle condition rediriger les données vers cette destination, choisissez la bonne condition.

## 8. Le composant Agregate.

Nous souhaitons maintenant créer un fichier plat, contenant les différentes catégories et le nombre de produits pour chacune. Utilisez le composant **Agregate** pour accomplir cette tâche.

## 9. Le composant Merge Join.

Créez un nouveau package, réalisez le Data Flow qui,

- Charge les deux tables Vendeur, Region et Commande
- Fait les jointures entre elles, utilisez le composant Merge Join.
- Enregistre dans un fichier **excel**, un tableau contenant les régions et leurs employés, (changez de composant de destination vers **Excel Destination**)