

## Introduction à SSIS

<b>1. Présentation de SSIS</b>	<b>11</b>
1.1 Concepts ETL et ELT	11
1.2 Architecture	13
1.3 Les éditions SQL Server	15
<b>2. Installation et outils</b>	<b>17</b>
2.1 Installation de SSIS	17
2.1.1 Installation étape par étape de SSIS	18
2.1.2 Installation en ligne de commande de SSIS	21
2.1.3 Finalisation de l'installation	24
2.2 Installation des outils de développement	27

## Nouveautés SSIS

<b>1. Introduction</b>	<b>29</b>
<b>2. Installation et administration</b>	<b>29</b>
2.1 Installation sur Linux	29
2.2 Installation distribuée (Scale Out)	33

2.2.1 Architecture générale	33
2.2.2 Installation d'un nœud Master	35
2.2.3 Installation d'un Worker	38
2.2.4 Configuration et surveillance du cluster	41
2.3 Support Always-On de la base SSISDB	45
2.4 Nouveau rôle SSIS_logreader	47
2.5 Personnalisation du niveau de journalisation	47
2.6 Support d'Always Encrypted	50
2.7 Déploiement incrémental des packages	52
2.8 Choix de la version de déploiement	54
<b>3. Composants et tâches</b>	<b>55</b>
3.1 Les tâches Cloud	55
3.1.1 Tâche Azure Blob Download Task	57
3.1.2 Tâche Azure Blob Upload Task	61
3.1.3 Tâche Azure HDInsight Create Cluster	62
3.1.4 Tâche de suppression d'un cluster Azure HDInsight	67
3.1.5 Tâche Hive Azure HDInsight	68
3.1.6 Tâche Azure HDInsight Pig	72
3.1.7 Tâche de chargement Azure SQL Data Warehouse (DW)	74
3.1.8 Tâche de système de fichiers Azure Data Lake Store	80

3.2 Nouveaux énumérateurs Foreach	82
3.2.1 Énumérateur HDFS	82
3.2.2 Énumérateur ADLS	83
3.2.3 Énumérateur Azure Blob	84
3.3 Composants Cloud de flux de données	86
3.3.1 Source Azure Blob	86
3.3.2 Destination Azure Blob	90
3.3.3 Source Azure Data Lake Store	91
3.3.4 Destination Azure Data Lake Store	92
3.4 Les composants Big Data	93
3.4.1 Tâche du système de fichiers Hadoop	95
3.4.2 Tâche Hadoop Hive	97
3.4.3 Tâche Hadoop Pig	98
3.4.4 Composant source de fichier HDFS	99
3.4.5 Composant destination de fichier HDFS	100
3.5 Les autres composants	101
3.5.1 Composant de publication de flux de données	101
3.5.2 Distributeur de données équilibrées	104
3.5.3 Les parties de package	105
3.5.4 Déploiement SSIS dans Azure	108

## Les bases de l'ETL

<b>1. Présentation et installation de la solution de développement de ce livre</b>	<b>113</b>
<b>2. Les bases avant une première implémentation</b>	<b>114</b>
2.1 Généralités	114
2.2 Le flux de contrôle : généralités	114
2.2.1 Définition	114
2.2.2 Le concepteur de flux de contrôle	115
2.3 Le flux de données : généralités	117
2.3.1 Objectifs	117
2.3.2 Le concepteur de flux de données	118
2.4 Contraintes de précédence versus Pipeline	120
2.4.1 Contraintes de précédence	120
2.4.2 Pipeline : les bases	125
2.5 Gestionnaires de connexions	127
2.5.1 Définition	127
2.5.2 Configuration d'un gestionnaire de connexions pour base de données	127
2.5.3 Configuration d'un gestionnaire de connexions vers des fichiers plats	132
2.6 Les variables	138
<b>3. Première implémentation</b>	<b>140</b>

3.1 Conteneurs	140
3.1.1 Principes de fonctionnement	140
3.1.2 Conteneur de boucles ForEach	142
3.1.3 Conteneur de boucles For	145
3.1.4 Portée et cloisonnement	146
3.2 Les expressions	147
3.2.1 Générateur d'expression	148
3.2.2 Les bases de la syntaxe	149
3.2.3 Expressions et gestionnaire de connexions	150
3.3 Extraire les données : composants source	153
3.3.1 Composant Source du fichier plat (Flat File Source)	154
3.3.2 Composant Source OLE DB	156
3.3.3 Éditeur avancé des composants source	158
3.4 Transformer les données	161
3.4.1 Colonne dérivée (Derived column)	161
3.4.2 Composant Recherche (Lookup)	162
3.4.3 Composant Transformation du cache	169
3.5 Charger les données : composants destination	170
3.6 Entrée et sortie des composants	174
3.6.1 Définition	174
3.6.2 Fractionnement conditionnel	174

3.6.3 Éditeur du chemin d'accès au flux de données	177
3.7 Les types de données sur SSIS	180
<b>4. Patterns de développement</b>	<b>182</b>
4.1 Éviter le composant Tri	182
4.2 Éviter le composant OLE DB Command	187
4.2.1 Configuration du composant OLE DB Command	187
4.2.2 Configuration de la Tâche d'exécution de requêtes SQL	190
4.3 Jointure SQL versus composant Jointure de fusion	194
4.3.1 Composant Jointure de fusion (Merge Join)	194
4.3.2 Composant Fusionner	198
4.4 Stocker temporairement des données	199
4.4.1 En base de données	199
4.4.2 Destination de fichier brut	199
4.5 Pattern Ignorer l'échec pour le composant Recherche	201
4.5.1 Composant Unir tout	202
4.5.2 Ignorer l'échec et colonne dérivée	204
4.6 Lot SSIS Parent-Enfant	204
4.6.1 Référence du projet	205
4.6.2 Référence externe	207

## Concepts de développement avancés

<b>1. Gestion de la qualité et du changement dans les données</b>	<b>211</b>
1.1 Data Quality Services (DQS)	211
1.1.1 Considérations sur DQS	211
1.1.2 Nettoyage DQS	213
1.2 CDC	215
1.2.1 Qu'est-ce que le CDC ?	215
1.2.2 Mise en place du CDC	216
<b>2. Interactions avec SQL Server et SSAS</b>	<b>221</b>
2.1 Insertion en bloc	221
2.2 Tâches de maintenance et de transfert d'objets de la base de données	224
2.3 Tâches relatives à SQL Server Analysis Services (SSAS)	225
2.3.1 Tâche de traitement Analysis Services	225
2.3.2 Tâche d'exécution de DDL Analysis Services	227
2.3.3 Tâche de requête de Data Mining	228
<b>3. Checkpoints</b>	<b>230</b>
3.1 Principe général	230
3.1.1 Cas d'usage classique	230
3.1.2 Mise en place	232

3.1.3 Lien entre échec et création de checkpoint	232
3.1.4 Fonctionnement	234
3.2 Limitations	236
3.2.1 Conteneurs mal gérés	236
3.2.2 Mauvaise gestion du parallélisme	237
3.2.3 Gestionnaires d'évènements	239
3.3 Conclusion	239
<b>4. Transactions</b>	<b>239</b>
4.1 Principe des transactions	239
4.1.1 A.C.I.D.	239
4.1.2 Transactions distribuées	240
4.1.3 Cas du DTC en réseau	242
4.2 Mise en place d'une transaction MSDTC dans SSIS	243
4.2.1 TransactionOption pour activer les transactions	243
4.2.2 IsolationLevel pour gérer l'interaction avec les autres transactions	243
4.2.3 Comportement à l'exécution	245
4.2.4 Analyse des problèmes	247
4.2.5 Conseils de design de Control Flow dans un cas transactionnel	248
4.3 Utilisation de transactions natives	249
4.3.1 BEGIN/ROLLBACK/COMMIT	249



4.3.2 Implémentation avec des Execute SQL Task et RetainSameConnection	249
4.3.3 Bonnes pratiques	252
4.4 Conclusion	252

## Déploiement et administration

<b>1. Introduction</b>	<b>253</b>
<b>2. Le catalogue SSISDB</b>	<b>253</b>
2.1 Configuration du catalogue	255
2.2 Rapports de supervision	256
<b>3. Les objets du catalogue SSIS</b>	<b>258</b>
3.1 Dossier	258
3.2 Environnement	260
3.3 Projet	262
3.3.1 Les fichiers de déploiement ISPAC	262
3.3.2 Déployer à l'aide de l'assistant	264
3.3.3 Déployer en T-SQL	268
3.3.4 Déployer en PowerShell	270
3.3.5 Configuration de projet	271
3.4 Package	275

3.4.1 Exécution de package	275
3.4.2 Planification de package	279
<b>4. La sécurité</b>	<b>283</b>
4.1 Sécurité niveau SSISDB	284
4.2 Sécurité niveau dossier	285
4.3 Sécurité niveau Projet	286
4.4 Sécurité niveau Environnement	286
Notions avancées et optimisations	
<b>1. Introduction</b>	<b>289</b>
<b>2. Les fondements du flux de données</b>	<b>289</b>
2.1 La mémoire tampon (buffer)	290
2.2 Execution Tree et threading	292
2.3 Les types de composants	296
2.4 Incidences sur les performances	300
<b>3. Mesure des performances et diagnostic</b>	<b>302</b>
3.1 Utilisation des journaux SSISDB	302
3.2 Utilisation de rapport SSIS	307

3.3 Journalisation personnalisée	308
3.4 Utilisation du moniteur de performance	310
<b>4. Des packages performants</b>	<b>312</b>
4.1 Utilisation des moteurs relationnels	312
4.2 Particularités sur les Lookup	315
4.3 Particularités de Fusion et Jointure de fusion	318
Bases de la programmation de composants	
<b>1. Introduction</b>	<b>323</b>
1.1 Pourquoi développer un nouveau composant ?	323
1.2 Possibilités offertes par les composants personnalisés	324
<b>2. Pour bien démarrer</b>	<b>325</b>
2.1 Quelques notions avant de commencer	325
2.2 Éléments à installer sur le poste de développement	328
2.3 Configurer un projet de bibliothèque	329
<b>3. Exemple d'une tâche de flux de contrôle</b>	<b>333</b>
3.1 Tâche personnalisée d'envoi d'e-mail	333
3.2 Rappels sur les composants disponibles	334

3.3 Héritage et métadonnées	336
3.4 Vérification de déploiement	337
3.5 Ajouter des propriétés de configuration	338
3.6 Implémenter la validation	339
3.7 La méthode Initialize	340
3.8 Réaliser un traitement	341
3.9 Gestionnaire de connexions personnalisées	342
3.10 Interface utilisateur du gestionnaire de connexions	345
<b>4. Log Provider, ouverture, logging et fermeture du log</b>	<b>348</b>
4.1 Création d'un Log Provider	348
4.2 Interface utilisateur du Log Provider	351
<b>5. Énumérer sur toute collection avec un ForEachEnumerator</b>	<b>352</b>
5.1 Création d'un ForEachEnumerator	352
5.2 Interface utilisateur du ForeachEnumerator	353
<b>6. Sources, destinations et transformations personnalisées</b>	<b>355</b>
6.1 Cas d'usage	356
6.2 Préparation de la solution	357
6.3 Développement de composant source	358

<b>7. Débogage</b>	<b>366</b>
7.1 Débogage des composants personnalisés	366
7.2 Débogage de l'interface graphique	368
Index	371