

Introduction à SSIS

1. Présentation de SSIS	13
1.1 Concepts de l'ETL	13
1.2 Management Studio et SQL Server Data Tools	14
1.3 Architecture	14
2. SSIS 2012	17
2.1 Les nouveautés de la version 2012	17
2.2 Les éditions SQL Server 2012	18
3. Installation de SSIS	19
3.1 Les composants à installer	19
3.2 Installation répartie	20
3.3 Installation étape par étape	21
4. Premier lot SSIS	24
4.1 Export de données	24
4.2 SQL Server Data Tools	31

Flux de contrôle, principes de fonctionnement

1. Présentation du flux de contrôle	35
1.1 Objectifs	35
1.2 Aperçu du concepteur	36
2. Gestionnaires de connexions	39
2.1 Concepts	39
2.2 Création	40
2.3 Types de connexions	41
2.3.1 Bases de données relationnelles	41
2.3.2 Fichiers	47

2.3.3	Protocoles réseau	54
2.3.4	Accès à l'API Windows	58
2.3.5	Autres gestionnaires de connexions spécifiques	60
2.3.6	Gestionnaires supplémentaires à télécharger	63
3.	Orchestration avec les contraintes de précedence	63
3.1	Mode de fonctionnement	64
3.2	Contraintes basées sur des expressions	69
3.3	Combinaisons logiques de contraintes	71
4.	Conteneurs et boucles	72
4.1	Principe de conteneur	72
4.2	Paramétrage des boucles	74
4.2.1	La boucle For	74
4.2.2	La boucle ForEach	75
5.	L'appel de packages enfants	83
5.1	Tâche d'exécution de package	84
5.1.1	Référence de projet (mode SQL Server 2012)	84
5.1.2	Référence de package (mode SQL Server 2008 R2 et précédents)	86
5.2	Configuration de Package Parent	87
6.	Les autres tâches de flux de contrôle	90
6.1	Tâches sur les bases de données relationnelles	90
6.1.1	Tâche d'exécution de requêtes SQL	90
6.1.2	Tâche de contrôle du Change Data Capture	95
6.1.3	Tâches de plan de maintenance SQL Server	98
6.1.4	Tâches de transferts d'objets SQL Server	99
6.2	Tâches sur les bases Analysis Services	99
6.2.1	Tâche de traitement Analysis Services	100
6.2.2	Tâche d'exécution de DDL Analysis Services	102
6.2.3	Tâche de requête de Data Mining	103
6.3	Les autres tâches	105
6.3.1	Tâche d'envoi de mail	105
6.3.2	Tâche de script	105

6.3.3 Tâche de service Web.....	109
6.3.4 Tâche XML.....	111
6.3.5 Tâche de système de fichiers.....	112
6.3.6 Tâche FTP.....	113
6.3.7 Tâche d'expression.....	113
6.3.8 Tâche d'exécution de processus.....	114
6.3.9 Tâche MSMQ.....	115
6.3.10 Tâches WMI.....	116
6.3.11 Tâche de profilage de données.....	118

Variables, paramètres et expressions

1. Les variables.....	121
1.1 Présentation des variables.....	121
1.2 La fenêtre Variables.....	122
1.3 Les variables système.....	125
1.3.1 Présentation des variables système.....	125
1.3.2 Utilisation des variables système.....	131
1.4 Les variables utilisateur.....	133
1.4.1 Propriétés principales.....	134
1.4.2 Création d'une variable utilisateur.....	138
1.4.3 Utilisation des variables utilisateur.....	139
2. Les paramètres.....	141
2.1 Présentation des paramètres.....	141
2.2 Propriétés des paramètres.....	141
2.3 Les types des paramètres.....	142
2.3.1 Les paramètres de package.....	142
2.3.2 Les paramètres de projet.....	143
2.3.3 Utilisation des paramètres.....	144
2.4 Création d'un paramètre.....	145
3. Les expressions.....	148
3.1 Le générateur d'expressions.....	148

3.2 Les opérateurs	150
3.3 Les fonctions	151
3.4 Marqueurs d'icônes	153
4. Exemple d'utilisation	154

Manipulation de données simples

1. Le flux de données	165
1.1 Principaux objectifs	165
1.2 Aperçu	166
1.3 Entrée-sortie et flux de données	168
2. Composants source et destination	171
2.1 Source et destination OLE DB	171
2.2 Fichier plat	176
2.3 Fichier Excel	180
2.4 Fichier brut	181
2.5 ADO.NET	181
2.6 Source XML	182
2.7 CDC Source	183
2.8 Source ODBC	183
2.9 Destination de l'ensemble d'enregistrements	184
2.10 Destination DataReader	184
2.11 Destination pour SQL Server	184
2.12 Destination pour SSAS	185
3. Fenêtres avancées	186
3.1 Éditeur avancé des composants	186
3.2 Fenêtre de mise à jour des mappages	187
4. Flux de données monosource, monodestination	189
4.1 Stockage de données temporaires SQL	189

4.2 Utilisation des transformations	190
---	-----

Transformation des données

1. Transformations simples	193
1.1 Colonnes dérivées	193
1.2 Multidiffusion	195
1.3 Fractionnement conditionnel	196
1.4 Conversion de données	198
1.5 Compteur de lignes	198
1.6 Tri	199
1.7 Agrégation	200
1.8 Commande OLE DB	202
1.9 Composant Script	202
2. Transformations moins communes	206
2.1 Copie de colonne	206
2.2 Transformation du cache	206
2.3 Audit	207
2.4 Échantillonnage des données	207
2.5 Table de caractères	207
2.6 Tableau croisé dynamique	208
2.7 Extraction de terme	211
2.8 Recherche de terme	212
2.9 Dimension à variation lente	212
2.10 Requête d'exploration de données	218
2.11 Export et Import de colonnes	219
2.12 Regroupement probable	220
3. Change Data Capture	221
3.1 Considérations sur le CDC	221
3.2 Mise en place	221
4. Data Quality Services	226

4.1 Considérations sur DQS	226
4.2 Nettoyage DQS	227

Flux de données multisource et jointures

1. Introduction	233
2. Union de flux de structure identique	233
2.1 Le composant Union All	234
2.2 Le composant Fusion	235
2.3 Comparaison des deux transformations et bonnes pratiques	237
3. Jointure de flux sur une clé : le composant Fusion de jointure	237
3.1 Jointure interne	238
3.2 Jointures externes à gauche et à droite	238
3.3 Jointures externes totales	239
3.4 Paramétrage de la tâche	240
4. Ajout d'informations externes dans un flux	241
4.1 Le composant Recherche	241
4.1.1 Cas de base du cache total sur connexion OLE DB	242
4.1.2 Que se passe-t-il si la valeur est introuvable ?	244
4.1.3 Avantages/inconvénients de la mise en cache et alternatives	246
4.1.4 Utilité de modifier la requête de recherche en cache partiel ou nul	247
4.1.5 Utilisation du composant Cache en cache total	248
4.2 Le composant Recherche Floue	251
4.2.1 Cas d'utilisation classique	251
4.2.2 Confiance et similarité	253
4.2.3 Paramétrage avancé	254
4.2.4 Les index de la Recherche Floue	255
4.2.5 Paramètres avancés	257

Événements et suivi d'exécution

SQL Server Integration Services 2012

Mise en oeuvre d'un projet ETL avec SSIS

1. Le gestionnaire d'événements	259
1.1 Principes de fonctionnement	259
1.2 Mise en place	264
2. Journalisation	266
2.1 Principes de fonctionnement	266
2.2 Mise en place	268
2.3 Journalisation de production avec SSIS 2012	273
3. Débogage	275
3.1 Les points d'arrêt	276
3.2 Résultats d'exécution et journal d'événements	279
3.3 Le flux de données	280
4. Monitoring	283
5. Data Taps	285

Administration SSIS

1. Introduction	289
1.1 Présentation des modèles de déploiement	289
1.2 Comparaison des modèles de déploiement	290
2. Le catalogue SSISDB	292
2.1 Présentation du catalogue SSISDB	292
2.2 Création de la base SSISDB	294
2.3 Administration avec SSISDB	297
2.3.1 Création de dossiers dans SSISDB	297
2.3.2 Utilisation d'environnements serveur	299
2.3.3 Gestion des versions	302

SQL Server Integration Services 2012

Mise en oeuvre d'un projet ETL avec SSIS

3. Service Integration Services	305
3.1 Présentation du service	305
3.2 Administration des packages	307
4. Déploiement	308
4.1 Utilisation du modèle de déploiement de projet	308
4.1.1 Les fichiers de déploiement ISPAC	308
4.1.2 L'assistant de déploiement du modèle de projet	310
4.2 Utilisation du modèle de déploiement de packages	317
4.2.1 Conversion vers un modèle de déploiement de package	318
4.2.2 Ajout d'une configuration de package	322
4.2.3 Création de l'utilitaire de déploiement	327
4.2.4 Exécution de l'utilitaire de déploiement	329
5. Exécution de package	334
5.1 Exécution dans SSISDB	334
5.2 DTEXEC	335
5.3 DTEXECUI	336
5.4 Utilisation de l'agent SQL Server	337
6. Sécurité	342
6.1 Sécurité du modèle de déploiement de projet	342
6.2 Sécurité du modèle de déploiement de package	347
6.2.1 Sécurité du service Integration Services	347
6.2.2 Sécurité du contenu du serveur SSIS	347
7. Reporting opérationnel dans SSIS	348
7.1 Les rapports prédéfinis	348
7.2 Les rapports personnalisés	351

Checkpoints et transactions

1. Introduction	353
2. Checkpoints	354
2.1 Principe général	354
2.1.1 Un cas d'usage classique	354
2.1.2 Mise en place	355
2.1.3 Lien entre échec et création de checkpoint	356
2.1.4 Fonctionnement	358
2.2 Limitations	361
2.2.1 Conteneurs mal gérés	361
2.2.2 Mauvaise gestion du parallélisme	362
2.2.3 Gestionnaires d'évènements	364
2.3 Conclusion	364
3. Transactions	365
3.1 Principe des transactions	365
3.1.1 A.C.I.D. ?	365
3.1.2 Transactions distribuées	366
3.1.3 Cas du DTC en réseau	367
3.2 Mise en place d'une transaction MSDTC dans SSIS	368
3.2.1 TransactionOption pour activer les transactions	369
3.2.2 IsolationLevel pour gérer l'interaction avec les autres transactions	370
3.2.3 Comportement à l'exécution	372
3.2.4 Troubleshooting de base	374
3.2.5 Conseils de design de flux de contrôle dans un cas transactionnel	375
3.3 Utilisation de transactions natives	375
3.3.1 BEGIN / ROLLBACK / COMMIT	375
3.3.2 Implémentation avec des Execute SQL Task et RetainSameConnection	376
3.3.3 Bonnes pratiques	378
3.4 Conclusion	379

Notions avancées et bonnes pratiques

1. Les fondements du flux de données	381
1.1 Les buffers	381
1.2 Les types de composants	383
1.3 Incidences sur les performances	386
1.4 Execution Tree et threading	389
2. Des packages performants	393
2.1 Utilisation des moteurs relationnels	393
2.2 Particularités sur les Recherches	396

Programmation de composants SSIS

1. Introduction	401
1.1 Pourquoi développer un nouveau composant ?	401
1.2 Possibilités offertes par les composants personnalisés	403
2. Pour bien démarrer	403
2.1 Quelques notions avant de commencer	403
2.2 Éléments à installer sur le poste de développement	408
2.3 Configurer un projet de base	409
3. Créer des composants	412
3.1 Choisir ce que l'on veut développer	412
3.1.1 Un exemple concret : une tâche d'envoi de mail personnalisée	412
3.1.2 Rappels sur les composants disponibles	414
3.2 Étendre la bonne classe de base et renseigner les métadonnées	416
3.3 Vérifier que le déploiement s'est bien déroulé	417
3.4 Ajouter des propriétés à l'élément personnalisé	418
3.5 Implémenter les méthodes importantes du runtime SSIS	419
3.5.1 La validation, une étape capitale	419
3.5.2 Les méthodes Initialize pour récupérer les connexions, les variables... ..	420
3.5.3 ConnectionManager : acquisition et relâchement des connexions	421

3.5.4 Réaliser un traitement dans une Task avec Execute()	422
3.5.5 LogProvider : ouverture, logging et fermeture	423
3.5.6 Énumérer sur une collection avec un ForEachEnumerator	426
3.5.7 PipelineComponent : sources, destinations et transformations	427
3.6 Comment déboguer tout cela ?	434
4. Ajouter une interface graphique	435
4.1 Principes généraux	435
4.2 Une Form pour une Task, un PipelineComponent ou un ConnectionManager	437
4.2.1 Développer la Form	437
4.2.2 Implémenter la bonne interface de base	439
4.2.3 Référencer l'assembly dans le composant source	440
4.3 Un UserControl pour un énumérateur ForEach	441
4.4 Une simple ComboBox pour un LogProvider	442
4.5 Débogage de l'interface graphique	443
5. Conclusion	444
Index	445