



Expert
EXPO
EX

SQL Server 2014

Implémentation d'une solution
de **Business Intelligence**

(Sql Server, Analysis Services,
Power BI...)

Téléchargement
www.editions-eni.fr



Thomas GAUCHET

Les éléments à télécharger sont disponibles à l'adresse suivante :
<http://www.editions-eni.fr>
Saisissez la référence ENI du livre **EI14SQL** dans la zone de recherche
et validez. Cliquez sur le titre du livre puis sur le bouton de téléchargement.

Avant-propos

1. À qui s'adresse ce livre ?	15
2. Pré-requis	15
3. Objectifs du livre	16
4. Notations	17

Partie Business Intelligence

Chapitre 1

Introduction à la Business Intelligence

1. Du transactionnel au décisionnel	19
2. Business Intelligence	22

Chapitre 2

Théories de la Business Intelligence

1. Architectures des systèmes décisionnels	25
1.1 Variétés des systèmes décisionnels	25
1.2 Data Mart et Data Warehouse	28
2. Modélisation OLAP	30
2.1 Modèles en étoile et en flocon	30
2.2 Dimensions	33
2.2.1 Attributs	33
2.2.2 Clés	33
2.2.3 Variation des dimensions	34

2 --- SQL Server 2014

Implémentation d'une solution de Business Intelligence

2.3	Faits, agrégations et cubes	36
3.	Conduite de projets	38
3.1	La démarche itérative	38
3.2	Les étapes de conception du Data Mart	39
3.3	Implémentation	40
4.	Self-service.	41
4.1	Self-service, du personnel au groupe	42
4.2	Business Intelligence d'entreprise	44
5.	Big data	45
6.	Cloud.	46

Chapitre 3 **Business Intelligence Microsoft**

1.	Offre générale	47
1.1	Positionnement.	48
1.1.1	BI pour tous	48
1.1.2	BI en libre-service	48
1.1.3	Offre évolutive	49
1.1.4	Offre extensible	49
1.2	Modèle unifié	50
1.2.1	Toutes les sources de l'entreprise.	50
1.2.2	Un portail unique	50
2.	La suite SQL Server.	50
2.1	SQL Server.	53
2.2	Analysis Services.	53
2.2.1	BISM multidimensionnel.	54
2.2.2	BISM tabulaire	56
2.2.3	Power Pivot	58
2.2.4	Choix du modèle.	58
2.3	Reporting Services	63
2.4	Integration Services	64

- 2.5 Master Data Services 66
- 2.6 Data Quality Services..... 68
- 2.7 Éditions SQL Server 69
 - 2.7.1 Gamme évolutive 69
 - 2.7.2 Licence..... 70
 - 2.7.3 Configuration du système 72
- 3. Office..... 74
 - 3.1 Excel..... 74
 - 3.2 Power BI pour Excel 76
 - 3.3 SharePoint..... 78
 - 3.3.1 SharePoint Foundation..... 78
 - 3.3.2 Microsoft SharePoint Server 78
 - 3.3.3 Complément SSRS pour les technologies SharePoint ... 79
 - 3.3.4 Excel Services 79
 - 3.3.5 Performance Point Server..... 80
 - 3.3.6 Power Pivot pour SharePoint 80
 - 3.3.7 Power View pour SharePoint..... 81
- 4. Le nuage..... 81
 - 4.1 SQL Azure..... 82
 - 4.2 Machines virtuelles Azure 82
 - 4.3 Power BI 83
 - 4.4 SharePoint Online 83
 - 4.5 HDInsight..... 84
 - 4.6 Place de marché Azure 84
- 5. Premiers pas avec SQL Server 84
 - 5.1 SSDT 85
 - 5.2 SSMS..... 92
 - 5.3 L'exemple Adventure Works 94

Partie Data Mart et modèle sémantique

Chapitre 4

Modéliser le Data Mart SQL

1. Caractéristiques du DM	97
2. Dimensions	98
2.1 Utilisation	98
2.2 Colonnes	100
2.3 Clés de substitution	101
2.4 Index	104
2.5 Membre inconnu	104
2.6 Dimension calendrier	105
3. Table de faits	109
3.1 Colonnes	110
3.2 Index	111
3.3 Index columnstore	112
3.4 Partitions	114
4. Base de données	117
4.1 Intégrité référentielle	117
4.2 Bases de données contained	118
4.3 Mode de récupération, sauvegardes	119
4.4 Compression	119
5. Assistant de création du modèle	120
6. Requête décisionnelle	131
6.1 Vues indexées	131
6.2 Requêtes	132

Chapitre 5

Créer le modèle tabulaire

1. Présentation	137
1.1 Bases du modèle tabulaire	139
1.2 Développement et gestion	141
1.3 Gérer les bases déployées	147
1.4 Métadonnées	148
1.5 Architecture	149
1.6 Méthodologie tabulaire	150
2. Données	150
2.1 Connexions	151
2.2 Tables	152
2.3 Relations	159
2.4 Exemple	161
3. Modèle sémantique	174
3.1 Colonnes calculées	174
3.2 Mesures	177
3.3 KPI (IPC)	180
3.4 Hiérarchies	183
3.5 Perspectives	186
4. Expressions DAX	188
4.1 Références	188
4.2 Opérateurs	189
4.3 Types	189
4.4 Fonctions logiques	190
4.5 Agrégats	192
4.6 Relations	193
4.7 Contextes et filtres	195
4.8 Hiérarchies parent-enfants	200

6 SQL Server 2014

Implémentation d'une solution de Business Intelligence

5. Table de dates	203
5.1 Déclarer la table de dates	203
5.2 Fonctions de dates	206
5.3 Dimensions jouant un rôle	208
6. Sécurité	210
7. Déploiement de la base SSAS	212
8. Alimentation de la base SSAS	213
8.1 Options de traitement	213
8.2 Traitement depuis SSMS	215
8.3 Traitement depuis SSIS	219
9. DirectQuery et partitions	219
9.1 Partitions	220
9.2 DirectQuery	224
10. Requêtes DAX	228

Chapitre 6

Créer le modèle multidimensionnel

1. Présentation de SSAS multidimensionnel	231
1.1 Modèle dimensionnel unifié	233
1.2 Métadonnées	234
1.3 Propriétés du projet	235
1.4 Méthodologie pour créer l'UDM	235
2. Données	236
2.1 Sources de données	236
2.2 Vue de sources de données	237
3. Dimensions	239
3.1 Terminologie	239
3.2 Assistant Dimension	243
3.3 L'éditeur de dimension	250

3.4	Propriétés de la dimension	252
3.5	Structure : attributs et hiérarchies	253
3.6	Relations d'attributs	256
3.7	Test	265
4.	Cubes	267
4.1	Terminologie	267
4.2	Assistant Cube	269
4.3	L'éditeur de cube	270
4.4	Structure de cube	272
4.4.1	Mesures	272
4.4.2	Dimensions	276
4.4.3	Cube	277
4.5	Utilisation des dimensions	277
4.6	Traitement et test du cube	282
5.	Déploiement de la base SSAS	285
5.1	Options de conception du studio	285
5.2	Déploiement en développement	289
5.3	Assistant de déploiement	290
6.	Alimentation de la base OLAP	291
6.1	Options de traitement	292
6.2	Traitement depuis SSMS	294
6.3	Traitement depuis SSIS	296
7.	Expressions MDX	297
7.1	Éléments de MDX	298
7.2	Manipulation	301
7.3	Mesure calculée et membre calculé	303
7.4	Membre par défaut	308
7.5	Jeu nommé (Named Set)	313

Chapitre 7**Finaliser le modèle multidimensionnel**

1.	Enrichir fonctionnellement la base SSAS	317
1.1	Commande de script	317
1.2	KPI	319
1.3	Actions	320
1.4	Traductions	321
1.5	Perspectives	321
1.6	Assistant Business Intelligence	323
1.6.1	Exécuter l'assistant Time Intelligence	324
1.6.2	Définir l'intelligence comptable	330
1.6.3	Spécifier un opérateur unaire	331
1.6.4	Créer une formule de membre personnalisée	332
1.6.5	Définir la conversion monétaire	333
1.7	Objets liés	335
1.8	Écriture (Writeback)	336
1.8.1	Dimension en écriture	336
1.8.2	Groupe de mesures en écriture	337
2.	Rôles de sécurité	340
3.	Stockage et partitions	343
3.1	Partitions	344
3.2	Modes de stockage	345
3.3	Agrégations	347
3.3.1	Conception d'agrégations	347
3.3.2	Optimisation basée sur l'utilisation	352
3.4	Traitement	356
3.5	Mise en cache proactive	357

Partie Reporting et Power BI

Chapitre 8

Créer des rapports avec SSRS

1. Présentation de SSRS	365
1.1 Langage de définition de rapports	366
1.2 Services	366
1.2.1 Serveur de rapports.	366
1.2.2 Gestionnaire de rapports	368
1.3 Développement.	368
1.3.1 Éléments à déployer	368
1.3.2 Outils de développement	370
1.3.3 Développement de rapport MDX	370
1.4 Solution SSRS	371
2. Extraction des données.	373
2.1 Sources de données.	373
2.1.1 Source de données partagée	373
2.1.2 Source de données du rapport	373
2.1.3 Propriétés de la source de données partagée	374
2.1.4 Propriétés de la source de données du rapport	376
2.2 Dataset	378
2.3 Concepteur de requêtes MDX graphique	379
2.4 Requêtes MDX pour SSRS.	380
2.4.1 Clause FROM	381
2.4.2 Clause ON ROWS	381
2.4.3 Clauses ON COLUMNS et WITH MEMBER	383
2.4.4 Filtres.	385
2.4.5 Paramètres.	388
3. Mise en forme d'un rapport avec SSDT	395
3.1 Boîte à outils	395
3.1.1 Régions de données	395
3.1.2 Visualisation des données	400
3.1.3 Éléments de rapport	407

3.1.4	Sous-rapport	408
3.1.5	Parties de rapport	408
3.2	Expressions	409
3.3	Format des valeurs	411
3.4	Actions	411
3.5	Exemple des commandes par canal	412
3.6	Exemple Carte du CA par département	431
4.	Utilisation de Report Builder	442
4.1	Choix entre SSDT et Report Buidler	443
4.2	Choix entre Report Builder 1.0 et 3.0	444
5.	Déploiement	445
5.1	Déploiement depuis SSDT	445
5.2	Déploiement avec le gestionnaire de rapports	447
6.	Impression	448
6.1	Mise en page	448
6.2	Formats de rendu	449
6.3	Impression client	450
7.	Administration	450
7.1	Configuration du serveur	451
7.2	Sécurité	453
7.2.1	Rôles sur les éléments	453
7.2.2	Rôles système	457
7.2.3	Utilisateurs extérieurs	458

Chapitre 9

Exploiter SSRS

1.	Introduction	459
2.	Serveur de rapports	460
2.1	Déploiement	460
2.2	Abonnements	461
2.2.1	Introduction	461

2.2.2	Sécurité	461
2.2.3	Planification	462
2.2.4	Abonnement standard	464
2.2.5	Abonnement piloté par les données	468
2.3	Le rapport lié (Linked Report)	469
2.4	L'instantané (Snapshot)	470
2.5	Cache	472

Chapitre 10 Power Pivot et Power View

1.	BI libre-service Microsoft	477
2.	Power Pivot	479
2.1	Déploiement Power Pivot	480
2.2	Principes du Power Pivot	481
2.2.1	Préparation des données	481
2.2.2	Mise en forme des rapports	485
2.2.3	Fonctionnalités Excel	486
2.2.4	Fonctionnalités Power Pivot	487
2.3	Exemple DADS	489
2.3.1	Présentation de l'exemple	489
2.3.2	Création du jeu de données Power Pivot	491
2.3.3	Mise en forme du rapport	507
2.4	Création d'une hiérarchie	512
2.5	Publication dans SharePoint	514
3.	Power View	515
3.1	Power View pour SharePoint	516
3.2	Power View pour Excel	518
3.3	Visualisations	519
3.4	Champs	525
3.5	Images	526
3.6	Tri	526
3.7	Mosaïque	527

3.8 Segment (Slicer)	529
3.9 Focus sur sélection	529
3.10 Multiples	531
3.11 Axe de lecture	532
3.12 Filtres	533
3.13 Dimensionnement d'une visualisation	536
3.14 Mode Lecture	537
3.15 Export PowerPoint	538

Partie Préparation des données

Chapitre 11

Alimenter le Data Mart avec SSIS

1. Présentation de SSIS	541
1.1 Comparaison des modèles de déploiement	543
1.1.1 Modèle de déploiement de package	543
1.1.2 Modèle de déploiement de projet	545
1.2 Changement de modèle	548
1.3 Paquets DTSX	550
1.4 Le studio de développement	552
1.4.1 Aperçu du studio	552
1.4.2 Projets SSIS	556
1.4.3 Exécution	558
2. Gestion des connexions	559
2.1 Centraliser les connexions du paquet	559
2.2 Types de connexions	560
2.3 Connexions aux bases de données	561
2.4 Sécurité du paquet	566
3. Flux de contrôle (Control Flow)	568
3.1 Contraintes de précedence	568
3.2 Boîte à outils du flux de contrôle	570
3.2.1 Tâches de flux de contrôle	572

- 3.2.2 Conteneurs de flux de contrôle 578
- 4. Flux de données (Data Flow) 586
 - 4.1 Boîte à outils du flux de données. 587
 - 4.1.1 Sources de flux de données 588
 - 4.1.2 Destinations de flux de données 592
 - 4.1.3 Transformations de flux de données. 594
 - 4.2 Alimentation d'une dimension 602
 - 4.3 Alimentation d'un groupe de mesures 618
 - 4.4 Développements 625
 - 4.4.1 Variables 626
 - 4.4.2 Expressions 627
 - 4.4.3 Paramètres. 630
 - 4.4.4 Gestionnaires de connexions 636

Chapitre 12
Exploiter et maîtriser SSIS

- 1. Développement avancé 637
 - 1.1 Ordre et tri des données. 637
 - 1.2 Audit 642
 - 1.3 Reprise sur erreur 643
 - 1.3.1 Transaction. 643
 - 1.3.2 Checkpoint 644
 - 1.4 Événements. 645
 - 1.5 Formats internationaux 646
- 2. Production. 647
 - 2.1 Architecture 647
 - 2.2 Catalogue SSIS : SSISDB 648
 - 2.3 Projets du catalogue 651
 - 2.4 Environnements du catalogue 653
 - 2.5 Liaisons entre projets et environnements 657
 - 2.6 Exécution depuis le catalogue 662
 - 2.7 Exécution depuis l'Agent 671

2.8 Exécution depuis l'utilitaire	674
2.9 Déploiement	674
2.10 Journaux	676
2.11 Drainage de données	678

Chapitre 13

Tirer parti de SQL pour l'ETL

1. Présentation	681
2. Chargement des dimensions et des faits	681
2.1 Contexte	681
2.2 Dimensions	682
2.3 Faits	688
2.4 Déploiement	691
3. Détection des modifications de la base source	696
3.1 Principe du CDC	696
3.2 Exemple CDC	697
4. Choix entre SQL Server et SSIS	701
4.1 Deux moteurs de données	701
4.2 Extraction avec Transact SQL	702
4.3 Transformation avec Transact SQL	704
4.4 Chargement avec Transact SQL	705
4.5 Le meilleur des deux mondes	706

Index	707
-----------------	-----

Chapitre 6

Créer le modèle multidimensionnel

1. Présentation de SSAS multidimensionnel

SSAS (*SQL Server Analysis Services*) multidimensionnel est un serveur de bases de données OLAP et un serveur de *Data Mining*. Cet ouvrage ne traite que de la partie OLAP : le *Data Mining* ne peut être laissé à des informaticiens seuls. Ces derniers doivent collaborer avec des statisticiens pour créer une solution viable de *Data Mining*.

■ Remarque

Pour rappel, SSAS peut être installé en mode tabulaire ou multidimensionnel. Si ces deux modes ont des points communs, ils restent très différents. Ce chapitre concerne le mode multidimensionnel uniquement.

Dans une solution de BI, le moteur de base de données OLAP apporte une couche métier entre l'utilisateur final et la donnée OLTP, la possibilité d'effectuer des calculs complexes sur les données et de les partager, une couche centralisée de sécurité liée à la donnée et surtout des requêtes d'extraction très performantes.

SSAS fournit au travers des dimensions et des mesures des objets métier simples à manipuler dans une interface graphique pour construire des tableaux. Au-delà des dimensions et des mesures, SSAS permet d'implémenter d'autres objets comme les KPI (*Key Performance Indicator*, indicateurs de performance clés) qui enrichissent l'interface métier.

MDX, le langage de requête, est plus adapté que le SQL pour écrire des requêtes décisionnelles et des formules de calcul. Obtenir le CA du mois précédent nécessite une requête SQL complexe et peu performante. En MDX, une simple expression permet de retourner instantanément la valeur. La puissance de calcul s'en trouve démultipliée. Les formules métier les plus utilisées sont créées dans la base de données OLAP, et disponibles pour tous les utilisateurs. Ces derniers bénéficient donc d'une bibliothèque de formules déjà créées et disponibles pour leurs propres besoins. Les chiffres diffusés dans l'entreprise sont théoriquement les mêmes pour tout le monde.

Les rôles de sécurité de la base de données SSAS permettent de sécuriser l'accès à une dimension ou un cube, mais aussi à certains membres d'une dimension, par utilisateur. Il est ainsi possible de ne rendre accessible à un commercial que les chiffres de sa zone géographique.

Le moteur OLAP construit des agrégats qui permettent une exécution très rapide des requêtes. En SQL, le calcul du CA du mois précédent implique une lecture de toutes les lignes des ventes du mois, alors que dans le cube le moteur n'a qu'à lire les bons agrégats précalculés.

■ Remarque

À l'inverse, si les utilisateurs finaux font des calculs ou des lectures sur les données détaillées, les performances de SSAS sont désastreuses. Dans ce cas, soit SSAS a été choisi à tort, soit des améliorations sont à apporter dans la modélisation de la base SSAS. Calculer un CA à partir des prix unitaires et des quantités vendues n'est pas une requête OLAP. Dans la conception du cube, vous devez remplacer le prix unitaire par le montant, qui est une donnée agrégable. Le CA sera alors l'agrégation des montants dans une bonne modélisation OLAP.

La version actuelle du moteur SSAS multidimensionnel est quasiment identique à celle de SQL Server 2008 R2 à deux exceptions près, du point de vue de la conception. D'une part, les fonctionnalités de l'édition Entreprise sont désormais accessibles dans la nouvelle édition dite "Business Intelligence". D'autre part, un nouveau mode de gestion des chaînes de caractères est disponible, mais pas utilisé par défaut. Il permet la limite des 4 Go, existante dans les précédentes versions. Pour l'utiliser, il suffit de modifier la propriété "STRING STORE COMPATIBILITY LEVEL" avec la valeur "1100" des dimensions et des groupes de mesures (pour le "DistinctCount").

1.1 Modèle dimensionnel unifié

L'UDM (*Unified Dimensional Model*) est le cœur du moteur OLAP de SSAS. Le modèle est unifié car il unit un modèle dimensionnel à un modèle relationnel. En plus des qualités décrites ci-dessus, SSAS permet de modéliser des relations entre les attributs d'une dimension, différents types de relations entre les dimensions et les mesures. Par exemple, une fonctionnalité comme les tables de liens NxN est typique du moteur relationnel. Elle est présente dans le moteur OLAP SSAS.

Avec l'UDM, le cube n'est plus limité à une seule table de faits, comme avec SSAS 2000. C'est un ensemble de tables de faits et de dimensions. Pour bien comprendre, il convient d'énumérer les différents composants de l'UDM (hors *Data Mining*).

Base de données

La base de données SSAS est l'équivalent de la base de données SQL. À vous de choisir son périmètre. Elle est le plus haut niveau de l'UDM. Elle contient tous les autres objets de l'UDM.

■ Remarque

La bonne pratique est de ne créer qu'un cube par base de données. Mettre plusieurs cubes qui, eux-mêmes, contiennent plusieurs tables de faits dans la même base est difficile à gérer.

Source de données

La source de données définit l'emplacement des données source et les informations de sécurité qui seront utilisées pour les lire. Une base peut contenir plusieurs sources.

Vue de source de données (DSV)

La vue de source de données (*Data Source View* - DSV) est une couche logique relationnelle entre le moteur OLAP et les données source. Elle peut être utilisée pour décrire le DM, ou pour créer un DM logique sur une source OLTP. Une base peut contenir plusieurs DSV.

Remarque

Il est possible d'avoir des objets provenant de plusieurs sources de données dans une DSV.

Dimension

Les dimensions sont les axes d'analyse. Elles peuvent être partagées par plusieurs cubes.

Cube

Le cube est la structure multidimensionnelle, l'objet le plus complexe de l'UDM. Il est formé de groupes de mesures et de dimensions de cubes (des liens vers les dimensions de la base). Un cube est lié à une DSV.

Rôles

Les rôles permettent de sécuriser les accès aux données multidimensionnelles par les utilisateurs ou les groupes Windows.

Assemblés

Des bibliothèques .NET permettent d'augmenter les fonctionnalités de SSAS.

Remarque

Certaines fonctionnalités des cubes ou des dimensions reposent en partie sur le client, l'application de BI. C'est le cas par exemple des KPI. Vous pouvez ajouter des KPI dans un cube, mais l'outil utilisé pour lire le cube les ignorera s'il n'est pas conçu pour. De même, certaines propriétés des dimensions sont destinées au client.

1.2 Métadonnées

Tous les objets sont décrits dans un format XML dédié : ASSL. Les métadonnées sont enregistrées dans des fichiers XML : un par composant de l'UDM. Lors de la génération de la solution dans le studio de développement, l'ensemble des fichiers est agrégé en un fichier unique dans le répertoire `\bin` du projet portant l'extension `.asdatabase`.

1.3 Propriétés du projet

À la création d'un nouveau projet, commencez par éditer les propriétés de celui-ci, par un clic droit sur le projet dans l'Explorateur de solutions.

- Dans la page **Générer**, pensez à bien sélectionner l'**édition du serveur** SSAS cible en production. Conservez le choix de l'édition Développeur (vous ne pourrez pas déployer sur une édition Standard si vous implémentez une fonctionnalité non supportée par cette dernière).
- Dans la page **Déploiement**, choisissez la **cible** pour votre développement : **serveur** et nom de la **base de données** SSAS à créer.

■ Remarque

Comme pour la partie SQL, pour développer en SSAS vous aurez besoin d'un serveur, installé par exemple sur votre poste de travail.

1.4 Méthodologie pour créer l'UDM

L'**Explorateur de solutions** semble indiquer un ordre de création des objets de l'UDM. Ne le suivez pas : créez les dimensions avant le cube.

La méthodologie conseillée est la suivante :

- Créez les sources de données et la DSV.
- Créez les dimensions et testez-les jusqu'à recette complète de chaque dimension.
- Créez le cube : groupes de mesures et dimensions.
- Ajoutez, au-dessus des groupes de mesures, des expressions calculées et autres actions.
- Configurez les rôles.

Remarque

Commencer par lancer l'Assistant du cube est un piège. Il n'est pas possible d'arriver à un bon résultat sans passer le temps nécessaire sur chaque dimension. Dans SSAS 2005, l'objectif de création du cube « en un clic » avait amené Microsoft à désactiver la gestion des erreurs d'intégrité dans les dimensions pour donner aux développeurs l'illusion que l'assistant fonctionnait... L'éditeur est revenu en arrière, mais dans l'Explorateur de solutions, le cube est toujours présenté avant la dimension.

2. Données

Deux classes d'objets réalisent l'interface avec les données des systèmes source, qu'ils soient opérationnels ou décisionnels : les sources de données et les vues de sources de données.

2.1 Sources de données

Les sources de données sont des sources relationnelles accédées via un fournisseur, soit *.NET*, soit *OLE DB natif*. Le gestionnaire de connexions propose des connecteurs SSAS, mais SSAS ne peut pas les exploiter comme sources de données. Par contre, il est possible par la suite de créer des liaisons avec des composants d'une autre base SSAS.

Remarque

Préférez les connecteurs OLE DB. En effet, le service SSAS n'étant pas développé en .NET, l'utilisation d'un fournisseur .NET provoque des conversions de types qui ralentissent la lecture.

Pour la connexion *OLE DB* à SQL Server, préférez **SQL Server Native Client**.

Lorsque la source de données utilise une authentification Windows, l'onglet **Informations d'emprunt d'identité** permet de spécifier quel utilisateur sera utilisé par le serveur pour accéder à la source. SSAS propose quatre configurations :

- **Utiliser un nom d'utilisateur et un mot de passe Windows spécifique**
Cette option permet de saisir l'utilisateur Windows sélectionné.