

Atelier2

Mise en place d'un Datawarehouse

Objectif :

Mettre en place un entrepôt de données implique la création du schéma de l'entrepôt de données (DW) et les objets de base de données. Les principaux objets de base de données sont des dimensions et des tables de faits. Pour accélérer l'extraction, la transformation et le chargement (ETL), vous pouvez avoir besoin des objets supplémentaires dans votre DW, comme des séquences, des procédures stockées et des tables de transfert.

Après cet atelier, vous serez en mesure de:

- **Créer une base de données de l'entrepôt de données ;**
- **Créer des séquences ;**
- **Mettre en place des dimensions ;**
- **Mettre en place vos des tables de fait.**

Création de LightAdventureWorksDW

1. Lancer **SSMS** et se connecter à votre instance de SQL Server. Ouvrir une nouvelle fenêtre de requête en cliquant sur le bouton Nouvelle requête.
2. D'après le contexte de la base de données **master**, créer une nouvelle base de données appelée **LightAdventureWorksDW**.
3. La base de données doit avoir les propriétés suivantes :
 - Il devrait y avoir un fichier de données unique et un seul fichier journal dans le dossier de **LightAdventureWorksDW**. Vous pouvez créer ce dossier dans n'importe quel lecteur.
 - Le fichier de données doit avoir une taille initiale de 300 Mo et avoir l'**autogrowth** permis en morceaux de 10 Mo.
 - La taille du fichier journal doit être de 50 Mo, avec 10% des morceaux de croissance automatique.
4. Après avoir créé la base de données, modifier le modèle de récupération à simple. Voici le code de création de base de données complète.

```

USE master;
IF DB_ID('LIGHTADVENTUREWORKSDW') IS NOT NULL
    DROP DATABASE LIGHTADVENTUREWORKSDW;
GO
CREATE DATABASE LIGHTADVENTUREWORKSDW
    ON PRIMARY
    (NAME = N'LIGHTADVENTUREWORKSDW', FILENAME = N'C:\DW\LIGHTADVENTUREWORKSDW.mdf',
    SIZE = 307200KB , FILEGROWTH = 10240KB )
    LOG ON
    (NAME = N'LIGHTADVENTUREWORKSDW_log', FILENAME = N'C:\DW\LIGHTADVENTUREWORKSDW_log.ldf',
    SIZE = 51200KB , FILEGROWTH = 10%);
GO
ALTER DATABASE LIGHTADVENTUREWORKSDW SET RECOVERY SIMPLE WITH NO_WAIT;
GO

```

5. Dans votre nouvel entrepôt de données, créer un objet séquence. Nommer le **SeqCustomerDwKey**. Commencer la numérotation à 1, et utiliser un incrément de 1 pour les autres options de séquence, utiliser les paramètres par défaut de SQL Server. Vous pouvez utiliser le code suivant.

```

USE LIGHTADVENTUREWORKSDW;
GO
IF OBJECT_ID('dbo.SeqCustomerDwKey','SO') IS NOT NULL
    DROP SEQUENCE dbo.SeqCustomerDwKey;
GO
CREATE SEQUENCE dbo.SeqCustomerDwKey AS INT
START WITH 1
INCREMENT BY 1;
GO

```

Création des dimensions

- Créer la dimension « clients ». La source de cette dimension est la dimension **DimCustomer** de la base de données exemple **AdventureWorksDW2012** (c'est juste pour simplifier l'atelier, normalement il faut extraire les données depuis **AdventureWorks2012**). Ajouter une colonne clé de substitution appelé **CustomerDwKey**, et de créer une contrainte de clé primaire sur cette colonne. Utilisez le tableau suivant pour les informations nécessaires pour définir les colonnes de la table et remplir la table.

Nom de la colonne	Type de donnée	Nullability	Remarques
CustomerDwKey	INT	NOT NULL	Clé de substitution ; attribuer des valeurs avec une séquence
CustomerKey	INT	NOT NULL	
FullName	NVARCHAR(150)	NULL	Concaténer <i>FirstName</i> et <i>LastName</i> de <i>DimCustomer</i>
EmailAddress	NVARCHAR(50)	NULL	
BirthDate	DATE	NULL	

MaritalStatus	NCHAR(1)	NULL	
Gender	NCHAR(1)	NULL	
Education	NVARCHAR(40)	NULL	<i>EnglishEducation</i> de <i>DimCustomer</i>
Occupation	NVARCHAR(100)	NULL	<i>EnglishOccupation</i> de <i>DimCustomer</i>
City	NVARCHAR(30)	NULL	<i>City</i> de <i>DimGeography</i>
StateProvince	NVARCHAR(50)	NULL	<i>StateProvinceName</i> de <i>DimGeography</i>
CountryRegion	NVARCHAR(50)	NULL	<i>EnglishCountryRegionName</i> de <i>DimGeography</i>
Age	Inherited	Inherited	Colonne calculée. Calculer la différence entre <i>BirthDate</i> et la date du jour, et la classer en trois groupes : <ul style="list-style-type: none"> • Lorsque la différence <= 40, étiqueter "Younger" • Lorsque la différence > 50, étiqueter "Older" • Sinon étiqueter "Middle Age"
CurrentFlag	BIT	NOT NULL	1 par défaut

2. Le code pour créer la dimension « clients » devrait être similaire au code suivant.

```

USE LIGHTADVENTUREWORKSDW;
GO
CREATE TABLE dbo.Customers
(
    CustomerDwKey INT NOT NULL,
    CustomerKey INT NOT NULL,
    FullName NVARCHAR(150) NULL,
    EmailAddress NVARCHAR(50) NULL,
    BirthDate DATE NULL,
    MaritalStatus NCHAR(1) NULL,
    Gender NCHAR(1) NULL,
    Education NVARCHAR(40) NULL,
    Occupation NVARCHAR(100) NULL,
    City NVARCHAR(30) NULL,
    StateProvince NVARCHAR(50) NULL,
    CountryRegion NVARCHAR(50) NULL,
    Age AS
    CASE
        WHEN DATEDIFF(yy, BirthDate, CURRENT_TIMESTAMP) <= 40
        THEN 'Younger'
        WHEN DATEDIFF(yy, BirthDate, CURRENT_TIMESTAMP) > 50
        THEN 'Older'
        ELSE 'Middle Age'
    END,
    CurrentFlag BIT NOT NULL DEFAULT 1,
    CONSTRAINT PK_Customers PRIMARY KEY (CustomerDwKey)
);
GO

```

3. Créer la dimension « Produits ». La source de cette dimension est la dimension **DimProducts** de la base de données exemple **AdventureWorksDW2012**. Utilisez le tableau suivant pour les informations nécessaires pour définir les colonnes de la table et remplir la table.

Nom de la colonne	Type de donnée	Nullability	Remarques
ProductKey	INT	NOT NULL	
ProductName	NVARCHAR(50)	NULL	<i>EnglishProductName</i> de <i>DimProduct</i>
Color	NVARCHAR(15)	NULL	
Size	NVARCHAR(50)	NULL	
SubcategoryName	NVARCHAR(50)	NULL	<i>EnglishProductSubcategoryName</i> de <i>DimProductSubcategory</i>
CategoryName	NVARCHAR(50)	NULL	<i>EnglishProductCategoryName</i> de <i>DimProductCategory</i>

4. Le code pour créer la dimension « Produits » devrait être similaire au code suivant.

```

USE LIGHTADVENTUREWORKSDW;
GO
CREATE TABLE dbo.Products
(
    ProductKey          INT          NOT NULL,
    ProductName         NVARCHAR(50) NULL,
    Color               NVARCHAR(15) NULL,
    Size                NVARCHAR(50) NULL,
    SubcategoryName     NVARCHAR(50) NULL,
    CategoryName        NVARCHAR(50) NULL,
    CONSTRAINT PK_Products PRIMARY KEY (ProductKey)
);
GO

```

5. Créer la dimension « Dates ». La source de cette dimension est la dimension **DimDate** de la base de données exemple **AdventureWorksDW2012**. Utilisez le tableau suivant pour les informations nécessaires pour définir les colonnes de la table et remplir la table.

Nom de la colonne	Type de donnée	Nullability	Remarques
DateKey	INT	NOT NULL	
FullDate	DATE	NOT NULL	<i>FullDateAlternateKey</i> de <i>DimDate</i>
MonthNumberName	NVARCHAR(15)	NULL	Concaténer <i>MonthNumberOfYear</i> (avec des zéros à gauche lorsque le nombre est inférieur à 10) et <i>EnglishMonthName</i> de <i>DimDate</i>
CalendarQuarter	TINYINT	NULL	
CalendarYear	SMALLINT	NULL	

6. Le code pour créer la dimension « dates » devrait être similaire au code suivant.

```

CREATE TABLE dbo.Dates
(
    DateKey          INT          NOT NULL          IDENTITY(1,1),
    FullDate         DATE         NOT NULL,
    MonthNumberName  NVARCHAR(15) NULL,
    CalendarQuarter  TINYINT      NULL,
    CalendarYear     SMALLINT     NULL,
    CONSTRAINT PK_Dates PRIMARY KEY (DateKey)
);
GO

```

Création de la table de faits

Dans cet exemple simplifié d'un véritable entrepôt de données, vous allez créer une seule table de faits.

1. Créer la table de faits **InternetSales**. (si vous n'arrivez pas à extraire les données de AdventureWorks2012, vous pouvez considérer la table de faits **FactInternetSales** de la base de données exemple **AdventureWorksDW2012** comme source de cette table de faits) . Ajouter les clés étrangères des trois dimensions créées dans les étapes précédentes. Ajouter une colonne integer en utilisant la propriété **IDENTITY**, et l'utiliser comme clé primaire. Utilisez le tableau suivant pour les informations nécessaires pour définir les colonnes de la table et remplir la table.

Nom de la colonne	Type de donnée	Nullability	Remarques
InternetSalesKey	INT	NOT NULL	IDENTITY(1,1)
CustomerDwKey	INT	NOT NULL	<i>FullDateAlternateKey de DimDate</i>
ProductKey	INT	NOT NULL	
DateKey	INT	NOT NULL	<i>OrderDateKey de FactInternetSales</i>
OrderQuantity	SMALLINT	NOT NULL	Default 0
SalesAmount	MONEY	NOT NULL	Default 0
UnitPrice	MONEY	NOT NULL	Default 0
DiscountAmount	FLOAT	NOT NULL	Default 0

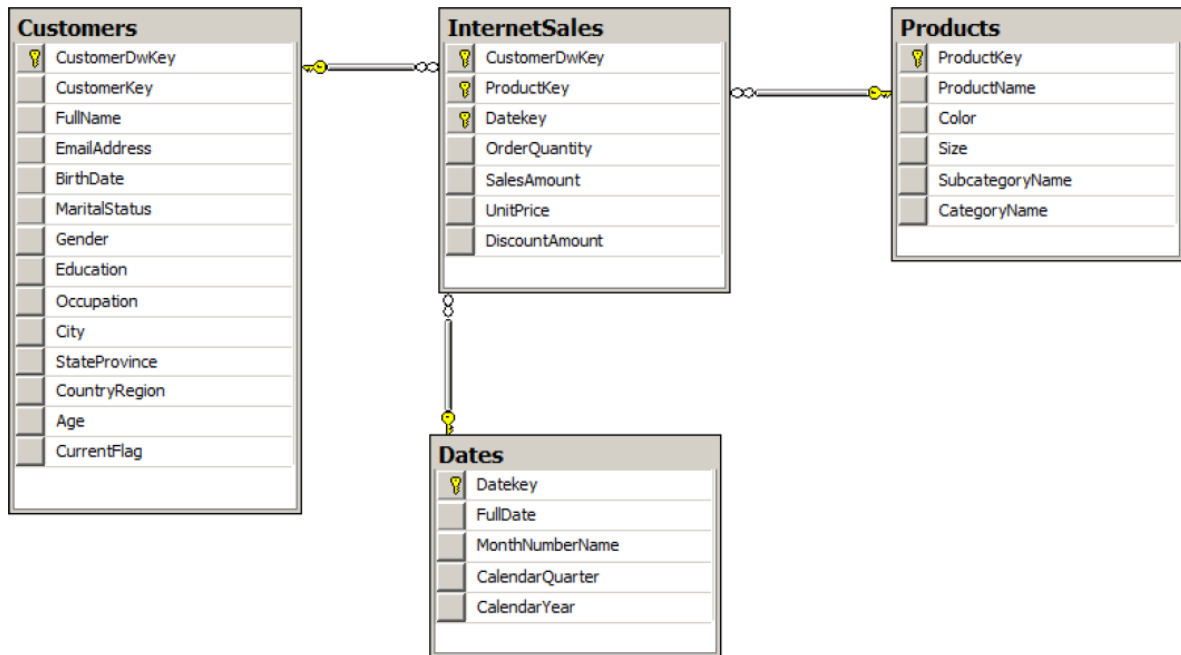
2. Le code pour créer la table de faits « InternetSales» devrait être similaire au code suivant.

```
USE LIGHTADVENTUREWORKSDW;
GO
CREATE TABLE dbo.InternetSales
(
    InternetSalesKey INT NOT NULL IDENTITY(1,1),
    CustomerDwKey INT NOT NULL,
    ProductKey INT NOT NULL,
    DateKey INT NOT NULL,
    OrderQuantity SMALLINT NOT NULL DEFAULT 0,
    SalesAmount MONEY NOT NULL DEFAULT 0,
    UnitPrice MONEY NOT NULL DEFAULT 0,
    DiscountAmount FLOAT NOT NULL DEFAULT 0,
    CONSTRAINT PK_InternetSales
    PRIMARY KEY (InternetSalesKey)
);
GO
```

3. Modifier la table de faits **InternetSales** pour ajouter des contraintes de clés étrangères des relations avec les trois dimensions.

```
USE LIGHTADVENTUREWORKSDW;  
GO  
ALTER TABLE dbo.InternetSales ADD CONSTRAINT  
    FK_InternetSales_Customers FOREIGN KEY(CustomerDwKey)  
    REFERENCES dbo.Customers (CustomerDwKey);  
ALTER TABLE dbo.InternetSales ADD CONSTRAINT  
    FK_InternetSales_Products FOREIGN KEY(ProductKey)  
    REFERENCES dbo.Products (ProductKey);  
ALTER TABLE dbo.InternetSales ADD CONSTRAINT  
    FK_InternetSales_Dates FOREIGN KEY(DateKey)  
    REFERENCES dbo.Dates (DateKey);  
GO
```

4. Créer un schéma de base de données, comme le montre la figure suivante. Le nommer **InternetSalesDW** et l'enregistrer.



5. Enregistrer le fichier avec le code T-SQL.