**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ**

**КАФЕДРА КОМП’ЮТЕРНИХ НАУК**

**СЕКЦІЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ПРОЕКТУВАННЯ**

ОДЗ з дисципліни

«Сховища даних»

|  |  |  |
| --- | --- | --- |
| Викладач: |  | Марченко А.В. |
| Студент: |  | Тиченко Я.М. |
| Група: |  | ІТ.мз-62с |
|  |  |  |
|  |  |  |
|  |  |  |

Суми 2017

**Зміст**

[Практикум 1: Вивчення вимірювань з бази даних AdventureWorksDW2012 2](#_Toc502103228)

[Завдання 1. Вивчення схеми бази даних AdventureWorksDW2012 4](#_Toc502103227)

[Практикум 2: Вивчення вимірювань з бази даних AdventureWorksDW2012 7](#_Toc502103228)

[Завдання 1. Дослідження вимірювань AdventureWorksDW2012 7](#_Toc502103229)

[Практикум 3: Вивчення таблиць фактів з бази даних AdventureWorksDW2012 10](#_Toc502103230)

[Завдання 1. Створення діаграми для таблиці фактів з AdventureWorksDW2012 10](#_Toc502103231)

[Завдання 2. Аналіз стовпців таблиці фактів 11](#_Toc502103232)

[Вправа 1. Короткостроковий пробний проект 12](#_Toc502103233)

[Вправа 2. Розширення короткострокового пробного проекту 13](#_Toc502103234)

[Практикум 4: Реалізація таблиць вимірів 14](#_Toc502103235)

[Завдання 1. Створення бази даних сховища даних і послідовності 14](#_Toc502103236)

[Завдання 2. Створення вимірювань 15](#_Toc502103237)

[Завдання 3. Створення таблиці фактів 17](#_Toc502103238)

[Практикум 5: Завантаження даних, застосування стиснення даних і індекси columnstore 20](#_Toc502103239)

[Завдання 1. Загрузка вашего хранилища данных 20](#_Toc502103240)

[Завдання 2. Застосування стиснення даних і створення індексу columnstore 21](#_Toc502103241)

[Висновок 22](#_Toc502103239)

**Вступ**

Вміння працювати та взаемодії із різними сховищами даних є важливою та необзхідною навичкою для будь-якого спеціаліста, пов'язаного із інформаційними технологіями. В ході виконання роботи буде виконано практични завдання для закріплення вивченого матеріалу.

Всього буде виконано п'ять практикумів з програмою AdventureWorksDW2012, а саме: створення схеми інтернет-продажів з бази даних, вивчення вимірювань з бази даних, вивчення таблиць фактів з бази даних, реалізація таблиць вимірів, завантаження даних, застосування стиснення даних і індекси columnstore.

**Практикум 1: Схема інтернет-продаж з бази даних AdventureWorksDW2012**

База даних AdventureWorksDW2012 – хороший приклад сховища даних. В ньому є всі елементи, які дозволять побачити приклади проектування змінних різних типів.

**Завдання 1. Вивчення схеми бази даних AdventureWorksDW2012**

У цьому завданні створюємо заново схему бази даних.

1. Запустіть середу SSMS і підключіть ваш екземпляр SQL Server. Відкрийте в браузері об'єктів папку Databases (Бази даних), а потім базу даних AdventureWorksDW2012
2. Клацніть правою кнопкою миші на папці Database Diagrams (Діаграми баз даних) і виберіть команду New Database Diagram (Створити діаграму бази даних). Якщо в цій базі даних діаграми ще не створювалися, ви побачите інформаційне повідомлення про те, що необхідно налаштувати створення схем баз даних. Якщо це повідомлення з'явиться на екрані, натисніть кнопку Yes (Так).
3. Зі списку Add Table (Додати таблицю) виберіть наступні таблиці (спочатку клацайте кнопкою миші кожну таблицю, а потім натискайте кнопку Add (Додати)):
   * DimCustomer;
   * DimDate;
   * DimGeography;
   * DimProduct;
   * DimProductCategory;
   * DimProductSubcategory;
   * FactlnternetSales.

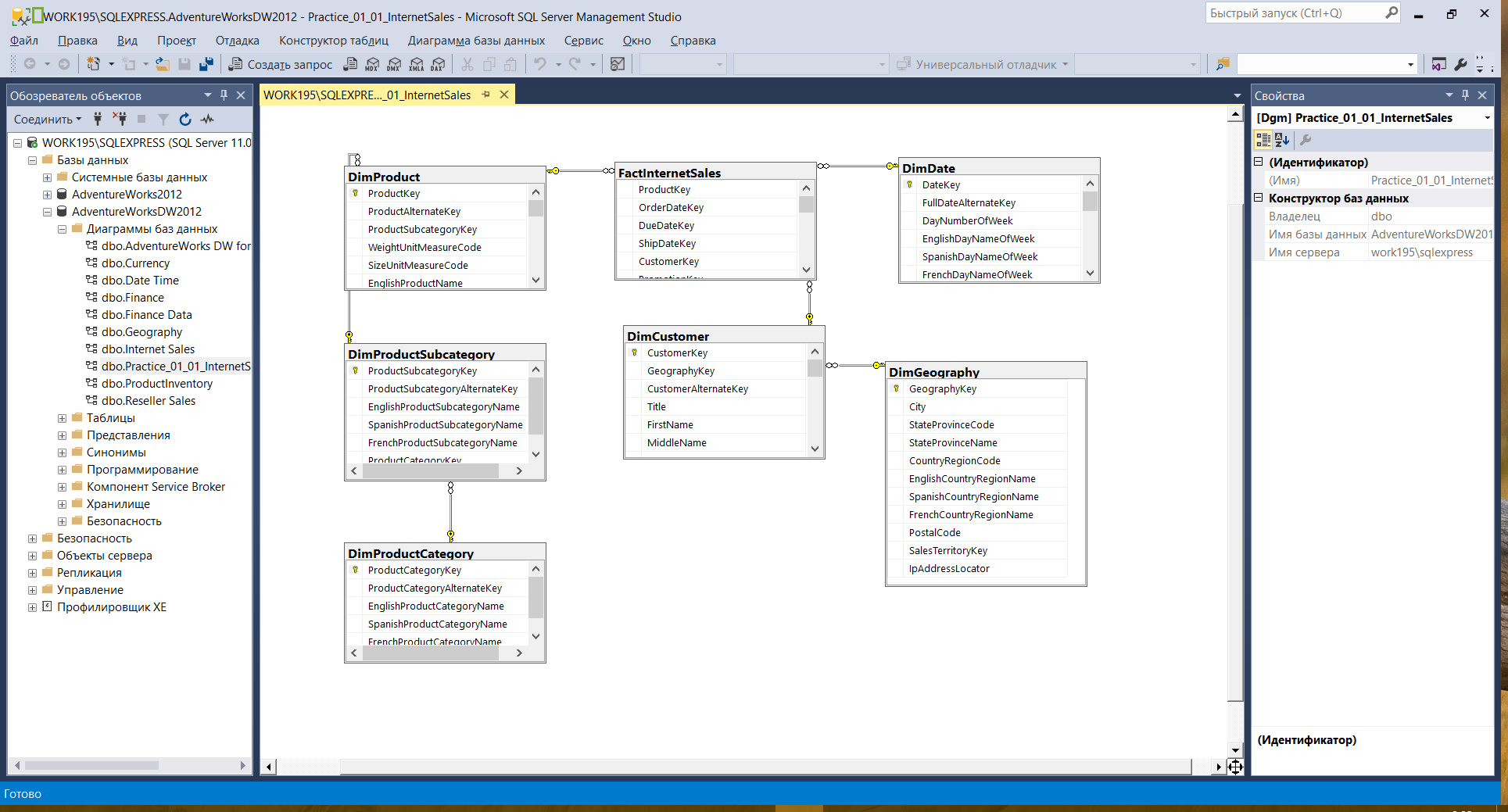


Рис 1.1 — схема бази даних

**Завдання 2. Аналіз діаграми**

Проаналізуйте схему AdventureWorksDW2012, щоб відзначити наступні факти.

* У вимірювання DimDate немає пов'язаних з ним додаткових таблиць уточнюючих запитів, і, отже, воно застосовується в схемі "зірка".
* У таблиці DimProduct зв'язку як в схемі "сніжинка"; вона використовує таблицю уточнюючих запитів DimProductSubcategory, яка в свою чергу використовує таблицю уточнюючих запитів DimProductCategory.
* Вимірювання DimCustomer застосовується в гібридної схемою - з першим рівнем схеми "сніжинка", які використовують тільки таблицю уточнюючих запитів DimGeography. Таблиця DimGeography денормалізована: у неї немає зв'язків з будь-якою іншою таблицею уточнюючих запитів.
* Ні в одній таблиці немає спеціальних стовпців для інформації про життєвий шлях даних.

**Резюме заняття**

* У сховищі даних схема "зірка" застосовується найчастіше.
* Схема "сніжинка" більше підходить для пробних проектів.
* Завжди слід визначати гранулярность таблиць фактів, а також необхідність аудиту і відомостей про життєвий шлях даних.

**Закріплення матеріалу**

1. Формувати звіти в схемі "зірка" простіше, ніж в нормалізованої схемі для оперативної обробки транзакцій (online transactional processing, OLTP). Чому виникає бажання спростити формування звітів? (Виберіть всі, що підходить).

* У схемі "зірка", як правило, менше таблиць, ніж в нормалізованої схемі. Тому запити простіше, т. К. Вимагають менше з'єднань.
* Схема "зірка" стандартизована і інформативна; ви можете швидко знайти відомості, необхідні для формування звіту.

1. Ви створюєте короткостроковий пробний проект. Яка схема найбільше підходить для такого проекту?

* Схема "сніжинка" більше підходить для короткострокових проб¬них проектів, оскільки вимірювання в ній нормалізовані, і, отже, вона більше відповідає вихідної нормализованной схемою.

1. У схемі "зірка" застосовуються таблиці двох типів. Які це типи? (Виберіть всі, що підходить.)

* вимірювання - частина схеми "зірка, таблиця фактів - центральна таблиця схеми" зірка ".

**Практикум 2: Вивчення вимірювань з бази даних AdventureWorksDW2012**

У зразку бази даних AdventureWorksDW2012 є багато вимірів. На цьому практичному занятті ми досліджуємо деякі з них.

**Завдання 1. Дослідження вимірювань AdventureWorksDW2012**

У цьому завданні ви створите діаграму для вимірювань.

1. Якщо ви закрили середу SSMS, запустіть її і підключіть ваш екземпляр SQL Server. У браузері об'єктів розкрийте папку Databases (Бази даних), а потім базу даних AdventureWorksDW2012.
2. Клацніть правою кнопкою миші на папці Database Diagrams (Діаграми баз даних) і виберіть команду New Database Diagram (Створити діаграму бази даних).
3. Зі списку Add Table (Додати таблицю) виберіть наступні таблиці (спочатку клацайте кнопкою миші на кожній таблиці, а потім натисніть кнопку Add (Додати)):

* DimProduct;
* DimProductCategory;
* DimProductSubcategory.

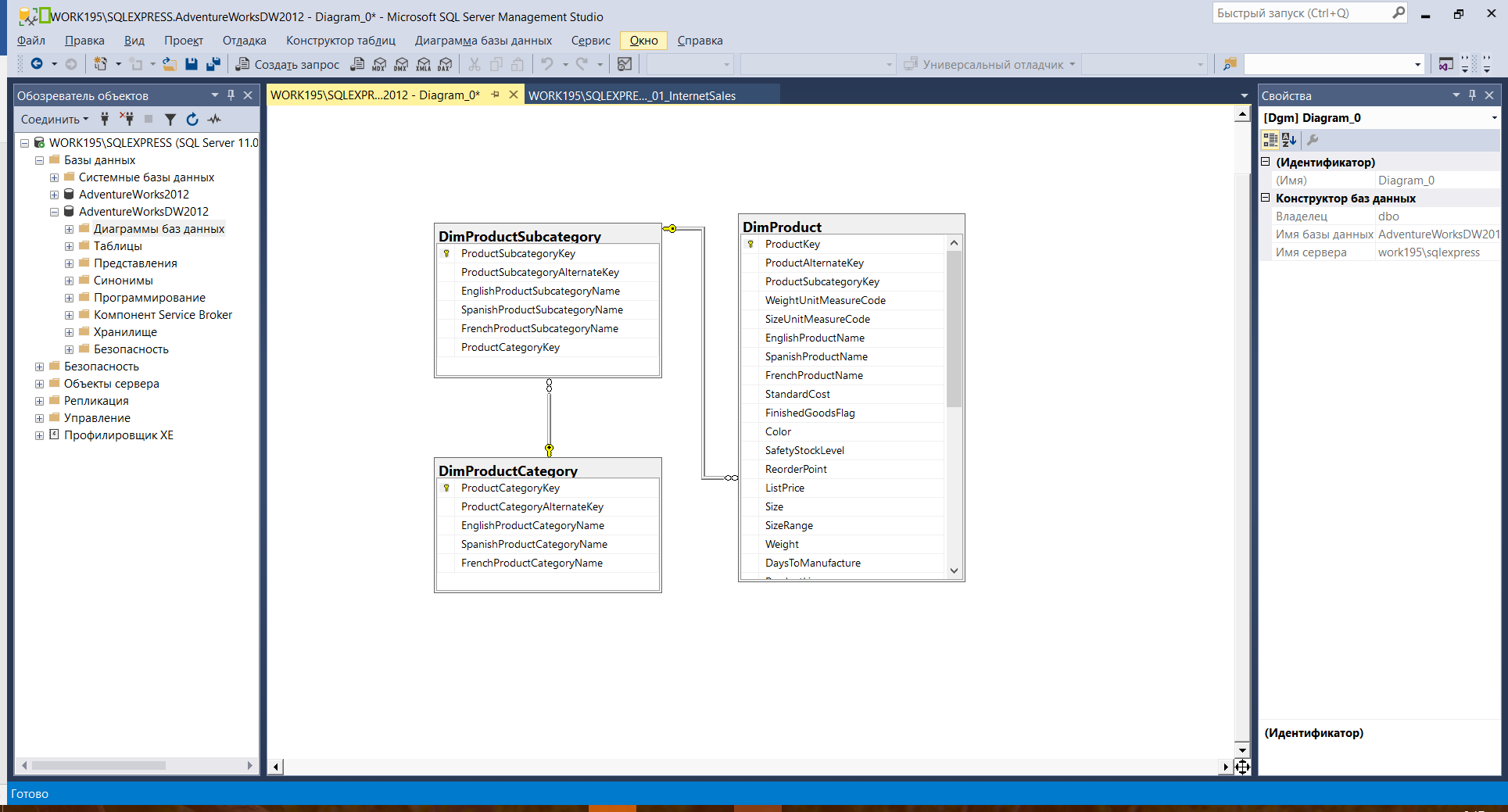


Рис 2.1 — схема бази даних

1. Спробуйте визначити, які стовпці відповідають наступним цільовим призначенням:

* ключи - ProductSubcategoryKey, ProductKey, ProductCategoryKey;
* імена - ModelName;
* переклад - EnglishProductName, SpanishProductName;
* У вимірі DimProduct є природна ієрархія: ProductCategory —» ProductSubcategory *—>* Product.
* У вимірі DimProduct є багато додаткових атрибутів, корисних для складання зведень, але які не є частиною будь-якої природної іерархії. До них відносяться, наприклад, атрибути Color (Колір) і size (Розмір).
* Деякі стовпці в вимірі DimProduct, наприклад LargePhoto (Велика фотографія) і Description (Опис), - це властивості елементів.
* Вимірювання DimSalesReason використовується в схемі "зірка". У цьому виді схеми важче знаходити природні ієрархії. У схемі "сніжинка" ви просто про¬сматріваете таблиці уточнюючих запитів і знаходите рівні ієрархій, а в схемі "зірка" вам доведеться знаходити ієрархії по іменах атрибутів. Якщо не вдається виділити ієрархії виходячи з імен стовпців, можна перевірити дані. У вимірі DimSalesReason (Причина продажу), здається, є природна ієрархія SalesReasonType (Тип причин продажу) - "SalesReasonName (Найменування причини продажу).

1. Спробуйте встановити, чи готові таблиці до опрацювання змін Туре 2 SCD. Таблиця DimProduct готова до опрацювання змін Туре 2 SCD, тому що в ній присутні поля StartDate і EndDate.
2. Увімкніть в діаграму таблицю DimSalesReason.

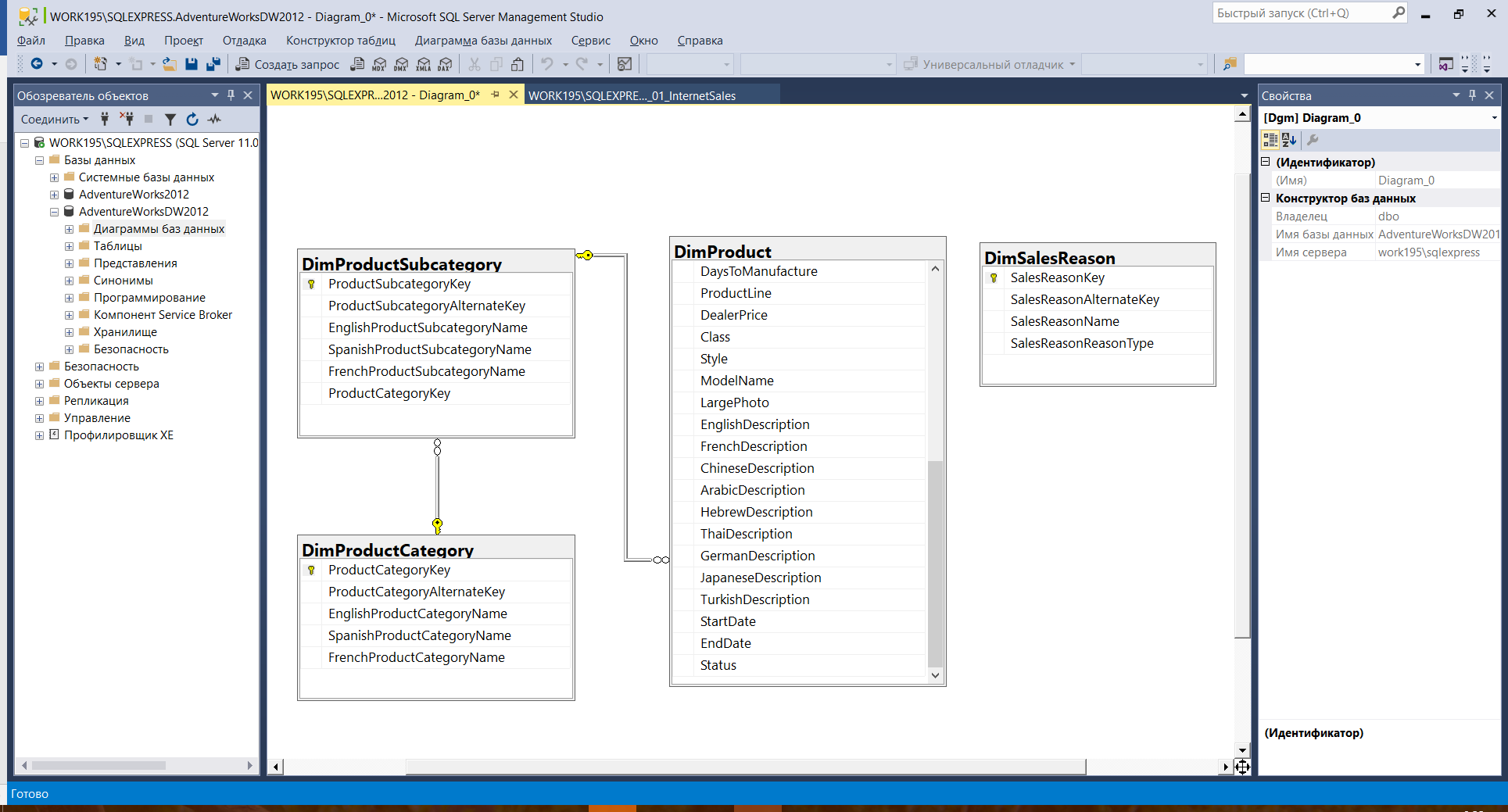


Рис 2.2 — схема бази даних

**Резюме заняття**

У вимірі можуть бути стовпчики наступних типів: ключі, імена, атрибути, властивості елементів, переклади та відомості про життєвий шлях даних.

Деякі атрибути формують природні ієрархії.

У проблеми повільно мінливих вимірювань (Slowly Changing Dimensions, SCD) є стандартні рішення.

**Закріплення матеріалу**

1. Ви реалізуєте для конкретного стовпця варіант Туре 2 рішення проблеми SCD. Що слід зробити при отриманні з вихідної системи зміненого значення для даного стовпця?

 Вставити новий рядок для того ж самого елемента вимірювання з новим значенням в обновлюваному стовпці. Застосувати сурогатний ключ, оскільки бізнес-ключ тепер не унікальний. Додати прапор, який вказує на те, яка рядок елемента актуальна або дійсна.

2. Який тип стовпчика не є частиною вимірювання?

Заходи - частина таблиць фактів.

3. Як знайти природні ієрархії в схемі "сніжинка"?

Природні ієрархії зберігаються в таблицях уточнюючих запитів, таблиці уточнюючих запитів для вимірювань вказують на природні ієрархії в схемі "сніжинка".

**Практикум 3: Вивчення таблиць фактів з бази даних AdventureWorksDW2012**

У зразку бази даних AdventureWorksDW2012 є безліч типів таблиць фактів, щоб продемонструвати всі можливі заходи. На цьому практичному занятті розглянемо одну з них.

**Завдання 1. Створення діаграми для таблиці фактів з AdventureWorksDW2012**

У цьому завданні ви створите діаграму бази даних для таблиці фактів і двох пов'язаних з нею вимірювань.

1. Якщо ви закрили середу SSMS, запустіть її і підключіть ваш екземпляр SQL Server. У браузері об'єктів розкрийте папку **Databases** (Бази даних), а потім базу даних **AdventureWorksDW2012**.

2. Клацніть правою кнопкою миші на папці **Database Diagrams** (Діаграми баз даних) і виберіть команду **New Database Diagram** (Створити діаграму бази даних).

3. Зі списку **Add Table** (Додати таблицю) виберіть наступні таблиці (спочатку клацайте кнопкою миші на кожній таблиці, а потім натисніть кнопку **Add** (Додати)):

* DimProduct;
* DimDate;
* FactProductlnventory.

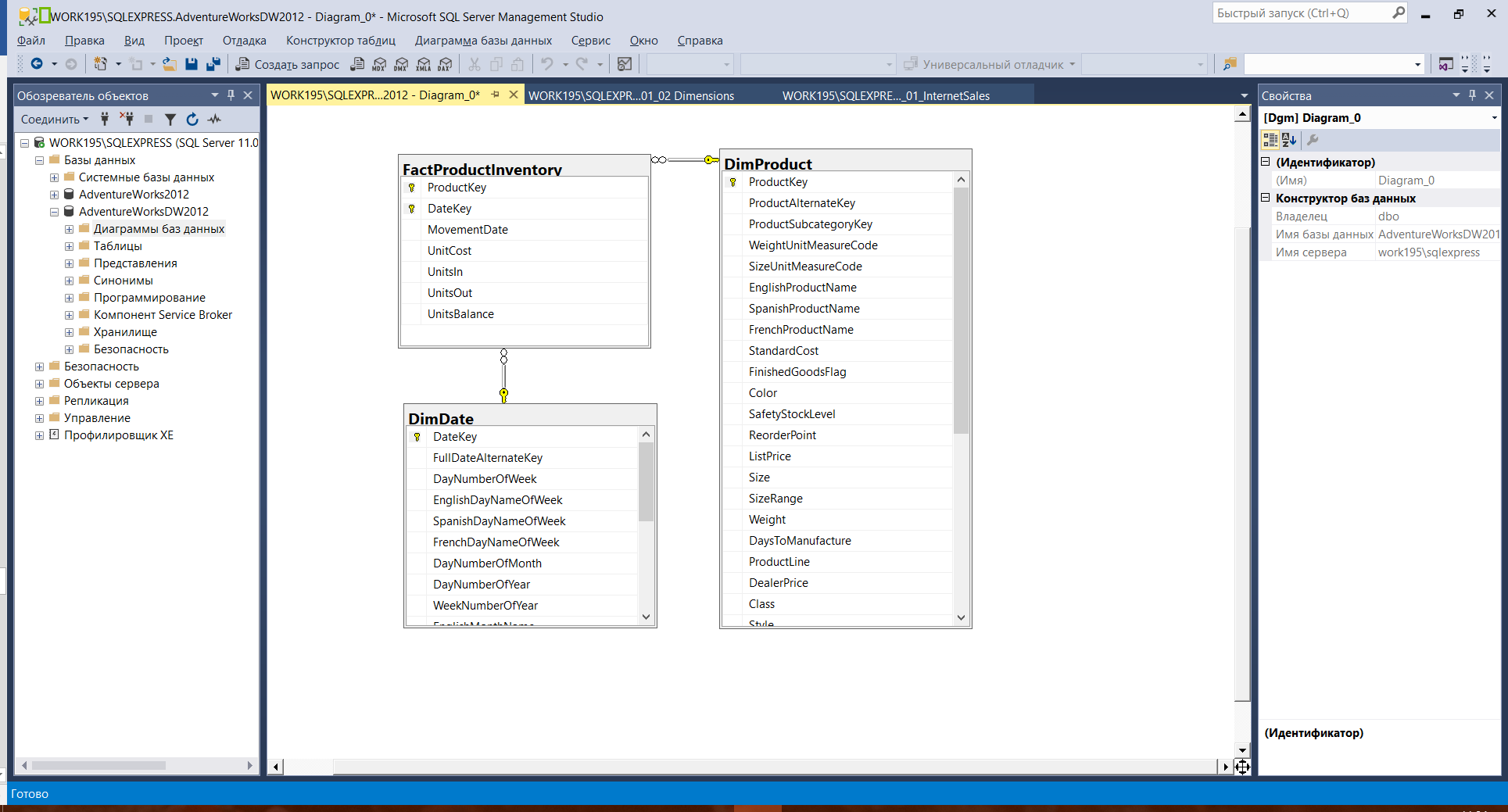


Рис 3.1 — схема бази даних

**Завдання 2. Аналіз стовпців таблиці фактів**

У цьому завданні ви дізнаєтеся нові подробиці про таблиці фактів зі схеми, яка була створена в попередньому завданні. Зверніть увагу на те, що ви з'ясуєте всі ці подробиці, аналізуючи імена стовпців з заходами; в реальному проекті також слід перевіряти вміст стовпців.

* Знаючи, як ведеться інвентарний облік, можна зробити висновок про те, що стовпчики Unitsin і unitsOut - адитивні заходи. Застосування в цих двох шпальтах агрегатної функції SUM цілком обгрунтовано для отримання сумарних значень або агрегатів з будь-якого виміру.
* Міра Unitcost - неаддитивну міра. Підсумувати її з будь-якого виміру не має сенсу.
* Міра unitsBalance- полуаддітівная міра. Ви можете застосувати агрегатную функцію SUM з будь-якого виміру, крім вимірювання часу.
* Збережіть діаграму з ім'ям Practice\_01\_03\_ProductInventory. Закрийте діаграму і вийдіть з-поміж SSMS

**Резюме заняття**

Таблиці фактів включають в себе заходи, зовнішні ключі і іноді додатковий первинний ключ і стовпці з відомостями про життєвий шлях даних.

* Заходи можуть бути адитивними, неаддитивну і підлозі аддитивними.
* Для відносин "багато-до-багатьох" можна вводити проміжне вимір.

**Закріплення матеріалу**

1. За яким виміром не слід застосовувати агрегатну функцію SUM для полуаддітівних заходів?

Неможливо застосовувати для полуаддитивних заходів агрегатні функції sum по вимірюванню Date..

1. Які міри по-вашому неадитивні? (Виберіть всі можливі варіанти.)

Ціни і відсотки знижок.

1. Стовпець якого типу не є частиною таблиці фактів?

Властивість елемента - тип стовпця в вимірі.

**Вправа**

**Вправа 1. Короткостроковий пробний проект**

Вас взяли на роботу для реалізації короткострокового пробного проекту сховища даних. Ви повинні підготувати схему для даних про продажі. Фахівці відділу продажів вашого замовника хотіли б аналізувати дані по клієнтах, товарам і періодами часу. Перш ніж створювати ХД і таблиці, ви повинні прийняти пару рішень і відповісти на кілька запитань.

* Яку схему ви стали б використовувати?

У короткостроковому пробному проекті слід застосовувати схему "сніжинка".

* Якими могли б бути вимірювання у вашій схемі?

У вас можуть бути вимірювання з замовниками, товарами і датами.

* Ви розраховуєте тільки на адитивні міри?

Ні, слід також розраховувати на кілька неаддитивних заходів. Скажімо, ціни і різні процентні ставки, наприклад, відсотки знижок, є неаддитивними.

**Вправа 2. Розширення короткострокового пробного проекту**

Після того як ви реалізували короткостроковий пробний проект сховища даних про продажі з вправи 1, ваш замовник був цілком задоволений. Його підприємству хотілося б розширити проект і довести його до реального довгострокового сховища даних. Але спілкуючись з аналітиками, ви також виявили кілька причин, що викликали незадоволеність.

**Результати опитування**

Нижче наведено перелік персоналу компанії, який висловив незадоволення під час бесіди, а також їхні зауваження.

**Спеціаліст відділу продажів: "**Я не бачу коректних агрегатів по регіонах для ретроспективних даних**".**

**Адміністратор бази даних, який створює звіти:** " Мої запити все одно ускладнені, з численними сполуками ".

**Питання**

1. Як ви б вирішили проблему фахівця відділу продажів?

Реалізувати для повільно мінливого вимірювання рішення Туре 2.

2. Схему якого типу ви б реалізували для довгострокового ХД?

Для довгострокового сховища даних потрібно вибрати схему "зірка".

3. Як ви б вирішили проблему адміністратора бази даних?

Застосувавши в проекті схему "зірка", ви автоматично вирішите проблему адміністратора бази даних.

**Практикум 4: Реалізація таблиць вимірів**

У цьому практикумі ви будете створювати сховище даних. Як джерело даних буде використовуватися зразок бази даних AdventureWorksDW2012. Вам не слід створювати явну проміжну область даних; в цій ролі буде використовуватися зразок бази даних AdventureWorksDW2012.

Якщо у вас виникнуть проблеми з виконанням завдань, можна встановити повністю завершені проекти з папки Solution (Рішення) супровідних матеріалів для цього розділу та заняття.

**Завдання 1. Створення бази даних сховища даних і послідовності**

У цьому завданні ви створите базу даних SQL Server для вашого сховища даних.

Запустіть середу SSMS і підключіть ваш екземпляр SQL Server. Відкрийте вікно створення запиту, натиснувши кнопку New Query (Створити запит).

Використовуючи контекст бази даних master, створіть нову базу даних з ім'ям TK463DW. Перед створенням бази даних перевірте, чи не існує вона, і якщо така база є, видаліть її. Завжди слід перед створенням перевіряти, чи існує об'єкт, і при необхідності видаляти його. У бази даних повинні бути наступні властивості.

Вона повинна мати один файл даних і один журнал транзакцій в папці ТК463. Створити цю папку можна на будь-якому дисковому пристрої.

У файлу даних повинен бути початковий розмір 300 Мбайт і повинен бути включений режим автоувеліченіе з приростом 10 Мбайт.

Файл журналу повинен бути об'ємом 50 Мбайт з 10-процентним збільшенням.

Після створення бази даних змініть модель відновлення на просту (simple). Нижче наведено повний код для створення бази даних.

USE master;

IF DB\_ID('TK463DW') IS NOT NULL DROP DATABASE TK463DW;

GO

CREATE DATABASE TK463DW ON PRIMARY

(NAME = N'TK463DW', FILENAME = N'C:\TK463\TK463DW.mdf',

SIZE = 307200KB, FILEGROWTH = 10240KB)

LOG ON

(NAME = N'TK463DW\_log', FILENAME = N'C:\TK463\TK463DW\_log.ldf’,

SIZE = 51200KB, FILEGROWTH = 10%);

GO

ALTER DATABASE TK463DW SET RECOVERY SIMPLE WITH NOJWAIT;

GO

У вашому новому сховищі даних створіть об'єкт-послідовність. Назвіть її SeqCustomerDwKey. Почніть з початкового значення 1 і задайте приріст 1. Для інших параметрів послідовності збережіть значення за замовчуванням, встановлені SQL Server. Можна використовувати наступний код:

USE TK463DW;

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

**Завдання 2. Створення вимірювань**

У цьому завданні ви створите вимір Customers (Замовники), для чого вам доведеться застосувати найрізноманітніші знання, отримані в цій та попередній главах. Вимірювання DimCustomer з бази даних AdventureWorksDW2012, яке послужить

джерелом для вашого вимірювання Customers, має на схемі додаткові лінії зв'язку. З ним пов'язана однорівнева таблиця для уточнюючих запитів DimGeography. Ви повністю денормалізуете це вимір. Крім того, ви вставите стовпці, необхідні для підтримки в вимірі способу SCD Туре 2, і пару обчислюваних стовпців. Крім вимірювання Customers ви також створите вимірювання Products (Товари) і Dates (Дати).

Створіть вимір Customers. Джерелом інформації для нього служить вимір DimCustomer з зразка бази даних AdventureWorksDW2012. Додайте стовпець сурогатного ключа customerDwKey і задайте обмеження первинного ключа для цього стовпця.

Ваш код інструкції створення вимірювання Customers повинен бути аналогічний наведеному далі.

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(l) NULL

Gender NCHAR(l) 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 BirthDate IS NULL THEN NULL

WHEN DATEDIFF(yy,BirthDate,CURRENT\_TIMESTAMP) > 50 THEN 'Older'

WHEN DATEDIFF(yy,BirthDate,CURRENT\_TIMESTAMP) > 40 THEN 'Middle Age'

ELSE 'Younger'

END

CurrentFlag BIT NOT NULL DEFAULT 1,

CONSTRAINT PK\_Customers PRIMARY KEY (CustomerDwKey)

) ;

GO

Створіть вимір Products. Джерело даних для цього виміру - таблиця DimProducts з зразка бази даних AdventureWorksDW2012. Для отримання інформації, необхідної для створення і заповнення цього виміру, використовуйте табл. 2.2.

Ваш код для інструкції створення вимірювання Products повинен бути аналогічний наведеному далі.

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

Створіть вимір Dates. Джерело даних для цього виміру - вимір DimDate з зразка бази даних AdventureWorksDW2012. Ваш код для інструкції створення вимірювання Dates повинен бути аналогічний наведеному далі.

CREATE TABLE dbo.Dates ( DateKey INT NOT NULL,

FullDate DATE NOT NULL,

MonthNumberName NVARCHAR(15) NULL, CalendarQuarter TINYINT NULL,

CalendarYear SMALLINT NULL,

CONSTRAINT PK\_Dates PRIMARY KEY (DateKey)

GO

**Завдання 3. Створення таблиці фактів**

В даному спрощеному прикладі реального сховища даних ви створите єдину таблицю фактів. У цьому прикладі можна застосовувати зовнішні ключі все разом як складової первинний ключ, т. К. У джерела даних для цієї таблиці - таблиці Factinternatsaies з бази даних AdventureWorksDW2012 - більш дрібна гранулярность (тобто велика ступінь деталізації), ніж у створюваній вами таблиці, і первинний ключ буде дублюватися. В якості первинного ключа ви могли б використовувати, як і в початковій таблиці, стовпці saiesOrderNumber і SaiesOrderLineNumber, але для того щоб показати, як можна автоматично пронумерувати стовпець з допомогою властивості identity, в цьому завданні ви вставите власний стовпець з цією властивістю. Він і буде вашим сурогатним ключем.

Створіть таблицю фактів internetsaies. Джерелом даних для цієї таблиці служить таблиця фактів Factinternetsaies з зразка бази даних AdventureWorksDW2012. Додайте зовнішні ключі з трьох вимірів, створених в завданні 2 даного заняття. Вставте цілочисельний стовпець, використовуючи властивість identity, і зробіть цей стовпець первинним ключем. Ваш код для інструкції створення таблиці фактів internetsaies повинен бути аналогічний наведеному далі.

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

Змініть таблицю фактів InternetSales, додавши обмеження зовнішнього ключа для зв'язків з усіма трьома вимірами. Код наведено в наступному лістингу:

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

**Резюме заняття**

У цьому занятті ви дізналися про реалізацію сховища даних.

Для бази даних сховища слід використовувати просту модель відновлення.

При створенні бази даних відводите достатньо місця на диску для файлів даних і файлів журналу транзакцій, щоб перешкодити автоувеліченіе файлів.

У вимірах з можливими змінами SCD Туре 2 застосовуйте сурогатні ключі.

Використовуйте обчислювані стовпці.

**Закріплення матеріалу**

1. Які об'єкти бази даних і які властивості об'єкта можна використовувати для формування автоматичної нумерації? (Виберіть всі підходящі варіанти.)

**Правильно:** властивість IDENTITY створює автоматичну нумерацію рядків.

**Правильно:** для автонумерації можна застосовувати новий, що з'явився в SQL Server 2012 об'єкт SEQUENCE.

1. Які стовпці вставляються в таблицю для обробки змін способом Туре 2 SCD? (Виберіть всі підходящі варіанти.)

**Правильно:** для реалізації збереження змін Туре 2 SCD необхідний прапор, позначати поточну або дійсну рядок.

**Правильно:** коли реалізується рішення Туре 2 SCD, потрібен новий сурогатний ключ.

1. Що таке виведений елемент?

**Правильно:** рядок, що додається в вимір під час завантаження таблиці фактів, називається виведеним елементом.

**Практикум 5: Завантаження даних, застосування стиснення даних і індекси columnstore**

У цьому практикумі ви завантажте в сховище дані, створені в практикумі заняття 1 цього розділу. Як джерело ваших даних буде використовуватися зразок бази даних AdventureWorksDW2012. Після завантаження даних ви застосуєте стиснення даних і створите індекс columnstore.

**Завдання 1. Загрузка вашего хранилища данных**

У першому завданні ви завантажте дані в ваше сховище.

Якщо ви закрили середу SSMS, запустіть її і підключіть ваш екземпляр SQL Server. Відкрийте вікно створення запиту, натиснувши кнопку New Query (Створити запит).

Підключіться до бази даних TK463DW. Завантажте вимір Customers, використовуючи інформацію.

Запит на завантаження наведено далі.

INSERT INTO dbo. Customers (CustomerDwKey, CustomerKey, FullName,

EmailAddress, Birthdate, MaritalStatus,

Gender, Education, Occupation,

City, StateProvince, CountryRegion)

SELECT

NEXT VALUE FOR dbo.SeqCustomerDwKey AS CustomerDwKey,

C.CustomerKey,

C.FirstName + ' ' + C.LastName AS FullName,

C.EmailAddress, C.BirthDate, C.MaritalStatus,

C.Gender, C.EnglishEducation, C.EnglishOccupation,

G.City, G.StateProvinceName, G.EnglishCountryRegionName FROM AdventureWorksDW2012.dbo.DimCustomer AS C INNER JOIN AdventureWorksDW2012.dbo.DimGeography AS G ON C.GeographyKey = G.GeographyKey;

GO

Завантажте вимір Products.

Запит на завантаження наведено далі.

INSERT INTO dbo.Products (ProductKey, ProductName, Color,

Size, SubcategoryName, CategoryName)

SELECT P.ProductKey, P.EnglishProductName, P.Color,

P.Size, S.EnglishProductSubcategoryName, C.EnglishProductCategoryName FROM AdventureWorksDW2012.dbo.DimProduct AS P INNER JOIN AdventureWorksDW2012.dbo.DimProductSubcategory AS S ON P.ProductSubcategoryKey = S.ProductSubcategoryKey INNER JOIN AdventureWorksDW2012.dbo.DimProductCategory AS C ON S.ProductCategoryKey = C.ProductCategoryKey;

GO

Завантажте вимір Dates.

Запит на завантаження наведено далі.

INSERT INTO dbo.Dates

(DateKey, FullDate, MonthNumberName,

CalendarQuarter, CalendarYear)

SELECT DateKey, FullDateAlternateKey,

SUBSTRING(CONVERT(CHAR(8), FullDateAlternateKey, 112), 5, 2)

+ ' ' + EnglishMonthName,

CalendarQuarter, CalendarYear FROM AdventureWorksDW2012.dbo.DimDate;

GO

Завантажте таблицю фактів internetSales.

Запит на завантаження наведено далі.

INSERT INTO dbo.InternetSales (CustomerDwKey, ProductKey, DateKey,

OrderQuantity, SalesAmount,

UnitPrice, DiscountAmount)

SELECT C.CustomerDwKey,

FIS.ProductKey, FIS.OrderDateKey,

FIS.OrderQuantity, FIS.SalesAmount,

FIS.UnitPrice, FIS.DiscountAmount FROM AdventureWorksDW2012.dbo.FactlnternetSales AS FIS INNER JOIN dbo. Customers AS C ON FIS.CustomerKey = C.CustomerKey;

GO

**Завдання 2. Застосування стиснення даних і створення індексу columnstore**

У цьому завданні ви застосуєте стиснення даних і створите індекс columnstore в таблиці фактів InternetSales.

Для обчислення простору, займаного таблицею InternetSales, скористайтеся системної збереженої процедурою sp spaceused. Використовуйте наступний код:

EXEC sp\_spaceused N'dbo.InternetSales', Supdateusage = N'TRUE';

GO

Таблиця повинна займати приблизно 3080 Кбайт замість виділеного простору. Тепер для стиснення таблиці застосуєте інструкцію ALTER TABLE. Використовуйте стиснення сторінок, як в наведеному далі коді.

ALTER TABLE dbo.InternetSales REBUILD WITH (DATA\_COMPRESSION = PAGE);

GO

Снова проверьте занимаемое пространство.

EXEC sp\_spaceused N'dbo.InternetSales', Supdateusage = N'TRUE';

GO

Тепер таблиця повинна займати приблизно тисяча дев'яносто шість Кбайт замість зарезерві¬рованного простору. Як бачите, за допомогою стиснення сторінок ви зекономили дві третини відведеного простору.

Створіть в таблиці InternetSales індекс columnstore. Скористайтеся наступним кодом:

CREATE COLUMNSTORE INDEX CSI\_InternetSales ON dbo.InternetSales

(InternetSalesKey, CustomerDwKey, ProductKey, DateKey,

OrderQuantity, SalesAmount,

UnitPrice, DiscountAmount);

GO

Насправді у вас недостатньо даних, щоб оцінити реальний виграш від застосування індексу columnstore і пакетної обробки запиту. Але ви все ж можете написати запит, який з'єднує таблиці і підраховує агрегати даних, і таким чином перевірити, чи застосовує SQL Server індекс columnstore. Нижче наведено приклад такого запиту.

SELECT С.CountryRegion, Р.CategoryName, D.CalendarYear,

SUM (I.SalesAmount) AS Sales FROM dbo.InternetSales AS I INNER JOIN dbo. Customers AS C ON I.CustomerDwKey = C.CustomerDwKey INNER JOIN dbo.Products AS P ON I.ProductKey = p.ProductKey

INNER JOIN dbo.Dates AS d ON I.DateKey = D.DateKey

GROUP BY C.CountryRegion, P.CategoryName, D.CalendarYear ORDER BY C.CountryRegion, P.CategoryName, D.CalendarYear;

Перевірте план виконання запиту, щоб дізнатися, застосовувався чи індекс columnstore. (В реальному тесті слід використовувати набагато більші набори даних.)

Цікаво підрахувати, скільки місця займає індекс columnstore. Скористайтеся знову системної процедурою sp\_spaceused.

EXEC sp\_spaceused N'dbo.InternetSales', @updateusage = N'TRUE';

GO

Цього разу відведений простір дорівнює приблизно 1560 Кбайт. Ви бачите, що незважаючи на застосоване стиснення сторінок, таблиця стиснута менше, ніж індекс columnstore. В даному випадку індекс columnstore займає приблизно половину простору, зайнятого таблицею.

**Резюме заняття**

У цьому занятті ви дізналися, як оптимізувати виконання запитів до сховища даних.

В ХД не слід застосовувати багато некластерізованних індексів.

Для первинних ключів кластеризації використовуйте маленькі цілочисельні сурогатні стовпці.

Користуйтеся індексованими уявленнями.

Застосовуйте індекси columnstore і виймайте вигоду з пакетної обробки.

**Закріплення матеріалу**

1. Які види стиснення даних підтримуються в SQL Server?

* SQL Server підтримує стиснення Unicode. Воно застосовується автоматично, коли використовується стиснення рядків або сторінок.
* SQL Server підтримує стиснення рядків.
* SQL Server підтримує стиснення сторінок.

1. Які оператори отримують виграш від застосування пакетної обробки?

* хеш-з'єднання можуть використовувати пакетну обробку.
* оператори Scan можуть отримати виграш від пакетної обробки.
* оператори фільтрації також використовують пакетну обробку.

1. Чому слід застосовувати індексовані уявлення?

* індексовані уявлення особливо корисні для прискорення запитів, що обчислюють агрегати даних.
* індексовані уявлення також можуть прискорити запити, які виконують множинні з'єднання.

**Висновок**

В ході виконання роботи, було набуто теоретичні та практичні навички роботи із сховищем даних AdventureWorksDW2012. В процесі, було виконано 5 практикумів, в яких ми навчились працювати із схемами баз даних, вимірюваннями, реалізувати таблиці вимірювань, завантажувати і стискати дані.

База даних AdventureWorksDW2012 – хороший приклад сховища даних. В ньому є всі елементи, які дозволять побачити приклади проектування змінних різних типів.