**Hurtownie Danych - laboratorium**

**Lista 4**

***Podstawy Integration Services oraz Analysis Services***

**Wstęp teoretyczny**

W ramach niniejszej listy wykorzystane zostaną podstawowe skrypty DDL oraz DML.

Tworzenie schematu:

|  |
| --- |
| **CREATE SCHEMA** schema\_name [ <schema\_element> [ ...n ] ] |

Tworzenie tabeli:

|  |
| --- |
| **CREATE TABLE**  { database\_name.schema\_name.table\_name. |  schema\_name.table\_name | table\_name }  ( { <column\_definition> } [ ,...n ] ) |

Wypełnianie tabeli danymi:

|  |
| --- |
| [ WITH <common\_table\_expression> [ ,...n ] ]  **INSERT**  {  [ TOP ( expression ) [ PERCENT ] ]  [ **INTO** ]  { <object> | rowset\_function\_limited  [ WITH ( <Table\_Hint\_Limited> [ ...n ] ) ]  }  {  [ ( column\_list ) ]  [ <OUTPUT Clause> ]  { VALUES ( { DEFAULT | NULL | exp } [ ,...n ] )  | derived\_table  | execute\_statement  | <dml\_table\_source>  | DEFAULT VALUES  }  }  } |

Dodawanie więzów integralności:

|  |
| --- |
| **ALTER TABLE** table\_name  **ADD CONSTRAINT** constraint\_name  {  [ NULL | NOT NULL ]  { PRIMARY KEY | UNIQUE }  [ CLUSTERED | NONCLUSTERED ]  [ WITH FILLFACTOR = fillfactor ]  [ WITH ( index\_option [, ...n ] ) ]  [ ON { partition\_scheme\_name (partition\_column\_name)  | filegroup | "default" } ]  **FOREIGN KEY** attribute\_name  **REFERENCES** [ schema\_name . ] referenced\_table\_name  [ ( ref\_column ) ]  [ ON DELETE { NO ACTION | CASCADE | SET NULL | SET DEFAULT }]  [ NOT FOR REPLICATION ]  | CHECK [ NOT FOR REPLICATION ] ( logical\_expression )  } |

Źródła:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-schema-transact-sql?view=sql-server-ver15>

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-transact-sql?view=sql-server-ver15>

<https://docs.microsoft.com/en-us/sql/t-sql/statements/insert-transact-sql?view=sql-server-ver15>

<https://docs.microsoft.com/en-us/sql/t-sql/statements/alter-table-column-constraint-transact-sql?view=sql-server-ver15>

**Zadania do wykonania**

Baza danych: **AdventureWorks**

**Zad 1. Przygotowanie schematu**

W bazie danych należy utworzyć schemat, którego nazwa będzie odpowiadać nazwisku wykonującego ćwiczenie (zapisać zapytanie tworzące ten schemat).

**Zad 2. Tworzenie tabel wymiarów i tabeli faktów**

W nowo utworzonym schemacie utworzyć tabele wymiarów: klienta, produktu i sprzedawcy (zapisać skrypt CREATE TABLE), opisane w następujących schematach:

* DIM\_CUSTOMER (CustomerID, FirstName, LastName, Title, City, TerritoryName, CounrtyRegionCode, Group)
* DIM\_PRODUCT (ProductID, Name, ListPrice, Color, SubCategoryName, CategoryName, Weight, Size, IsPurchased)
* DIM\_SALESPERSON (SalesPersonID, FirstName, LastName, Title, Gender, CountryRegionCode, Group)

oraz tabelę faktów:

* FACT\_SALES (ProductID, CustomerID, SalesPersonID, OrderDate, ShipDate, OrderQty, UnitPrice, UnitPriceDiscount, LineTotal)

*Uwaga 1: Proszę sprawdzić, dlaczego niektóre atrybuty mają wartość NULL. Czy zawsze wynika to z braku danych?*

*Uwaga 2: Kolumny OrderDate oraz ShipDate przechowują dane typu całkowitego, gdzie cztery pierwsze cyfry oznaczają rok, dwie następne miesiąc, a dwie ostatnie dzień. Do pobrania poszczególnych części daty użyć funkcji datepart.*

*Uwaga 3: Atrybuty OrderQty oraz LineTotal będą miarami w naszej hurtowni.*

**Zad. 3. Wypełnianie danych – denormalizacja źródłowej bazy**

Wypełnić nowoutworzone tabele danymi znajdującymi się w tabelach źródłowych. Do wypełnienia użyć instrukcji INSERT INTO. Proszę sprawdzić liczbę skopiowanych rekordów.

*Uwaga 1. Do tabeli DIM\_PRODUCT należy także skopiować produkty, które nie mają przypisanej podkategorii.*

*Uwaga 2. Do tabeli FACT\_SALES należy skopiować również transakcje, które nie mają sprzedawcy.*

Przykładowe rozwiązania:



*Przykładowe rozwiązanie – tabela DIM\_CUSTOMER.*



*Przykładowe rozwiązanie – tabela DIM\_PRODUCT.*



*Przykładowe rozwiązanie – tabela DIM\_SALESPERSON.*



*Przykładowe rozwiązanie – tabela FACT\_SALES.*

**Zad. 4. Więzy integralności**

1. Dodać integralność referencyjną i klucze główne do tabel już zdefiniowanych.
2. Przygotować instrukcję INSERT INTO, która sprawdzi poprawność integralności referencyjnej oraz klucze główne.

**Zad. 5. Tworzenie kostki**

Należy utworzyć projekt Analysis Services, w którym zostanie przygotowana kostka zawierająca utworzone wymiary (klienta, produktu i sprzedawcy) oraz tabelę faktów.

Używając Visual Studio utworzyć projekt typu Analysis Services Project (Menu File -> New Project)

1. Dodać źródło danych (Solution Explorer -> Data Sources -> New Data Source), które będzie wskazywało na bazę danych, która przechowuje tabele faktów i wymiarów. Przeanalizować opcje związane z bezpieczeństwem dostępu do danych.
2. Utworzyć nowy widok źródła danych (Solution Explorer -> Data Source Views -> New Data Source View). Dodać wcześniej utworzone tabele.
3. Utworzyć nową kostkę za pomocą asystenta (Solution Explorer -> Cubes -> New Cube):

* Wybrać utworzenie kostki na podstawie istniejących tabel (Use existing tables)
* Wybrać, utworzony w poprzednim punkcie, widok źródła danych
* Jako tabelę faktów (Measure group tables) wybrać FACT\_SALES
* Na stronie dotyczącej miar wybrać OrderQty, UnitPriceDiscount, Line Total. Zastanowić się nad użytecznością wybranych miar. Dlaczego nie wszystkie atrybuty tabeli FACT\_SALES mogą być użyte jako miary?
* Na stronie dotyczącej wymiarów wybrać wszystkie tabele z przedrostkiem DIM.

Po utworzeniu kostki dokonać edycji wymiarów (Solution Explorer -> Dimensions lub zakładka Cube Structure -> Dimensions).

Dla każdego z wymiarów zdefiniować potrzebne atrybuty. Przykładowo wymiar produkt powinien zawierać: Nazwę, Cenę, Kolor, Podkategorię i Kategorię. W przypadku tabeli Produkt należy zmienić jej definicję i sposób ładowania, tak aby zawierała ona zarówno kategorię jak i podkategorię (jeśli nie zrobiło się tego już podczas pracy z kreatorem dodawania nowej kostki).

**Zad. 6. Uruchomienie kostki**

1. Sprawdzić ustawienia projektu (Solution Explorer -> Nazwa projektu -> Properties).
2. Sprawdzić ustawienia konfiguracji wdrożenia (Configuration properties -> Deployment -> Target) projektu – Server: localhost, Database: PierwszyProjektKostek
3. Wdrożyć projekt (Menu Build -> Deploy). W przypadku błędu:
   * w konfiguracji wdrożenia (Configuration properties -> Deployment -> Target) przekleić nazwę serwera (z ekranu wyboru serwera w MSSQL Management Studio);
   * klikając dwukrotnie na utworzony Data Source, sprawdzić działanie innych opcji „Impersonation information” (VS powinien być uruchamiany z konta o uprawnieniach administratora);
4. Przetworzyć kostkę (Przycisk Process).

**Zad. 7. Proste raporty**

Połączyć się z programem MS Excel i przygotować raporty (tabele i wykresy przestawne), w których zostaną ujęte ciekawe zależności pomiędzy danymi.

**Rozwiązania:**

**Wnioski:**

***Uwaga!***

* Sprawozdanie, bez wniosków podsumowujących aspekt zagadnień analizowanych na zajęciach laboratoryjnych i zawartych w sprawozdaniu, jest automatycznie oceniane negatywnie!