

Utworzenie relacyjnej bazy danych i wypełnienie jej danymi wydobytymi z pliku XML w procesie ETL

A. Utworzenie bazy danych

1. Otwórz narzędzie *Microsoft SQL Server 2008* → *SQL Server Management Studio*
2. W oknie połączenia wybierz *Server Type: Database Engine*, *Server Name: localhost*
3. Wybierz *Connect*
4. Otwórz dokument *tworzenie_bazy.sql* (*File* → *Open* → *File*)
5. Wykonaj komendę *Execute*

B. Proces ETL

1. Otwórz narzędzie *Microsoft SQL Server 2008* → *SQL Server Business Intelligence Development Studio*
2. Utwórz nowy projekt dla procesu integracji danych (*File* → *New* → *Project*). Typ tworzonego projektu to *Integration Services Project*

Odczytanie danych z pliku *filmy.xml*

3. Na zakładkę *Control Flow* przeciągnij z *Toolbox*'a element *Data Flow Task*.
4. Zmień nazwę tego zadania na "Informacje o filmach" klikając na niego i wybierając F2.
5. Kliknij dwukrotnie na to zadanie aby przejść do zakładki *Data Flow*.
6. Na zakładkę *Data Flow* przeciągnij z *Toolbox*'a element *XML Source*.
7. Zmień jego nazwę na "Filmy".
8. Kliknij dwukrotnie na to zadanie - otworzy się okienko z jego właściwościami.
9. W polu *XML location*: wskaż adres pliku *filmy.xml*.
10. W polu *XSD location*: wybierz plik *filmy.xsd*.
11. Kliknij OK.

Pliki XML Schema są potrzebne w celu zdefiniowania typów zmiennych reprezentujących dane zawarte w plikach XML. Program *SQL Server Business Intelligence Development Studio* pozwala na wygenerowanie plików Schema, jednak typy zdefiniowane w pliku docelowym mogą być różne od oczekiwanych. W naszym przypadku stworzenie i poprawna modyfikacja pliku XSD sprowadzałyby się do następujących kroków:

- a) Otwarcie pliku *filmy.xml* w Visual Studio (*File* → *Open* → *File*).
- b) Wygenerowaniu pliku XML Schema (*XML* → *Create Schema*).
- c) Zapisaniu wygenerowanego pliku na dysku i dokonaniu w nim następujących zmian:

I. W wierszu:

```
<xs:element name="cena" type="xs:unsignedByte" />
```

należy zmienić typ elementu z *xs:unsignedByte* na *xs:decimal*

II. Element:

```
<xs:element minOccurs="0" name="tytulFilmu" type="xs:string" />
```

należy rozwinąć w celu ograniczenia dopuszczalnej liczby znaków do sześćdziesięciu poprzez zastąpienie go następującym wpisem:

```

<xs:element minOccurs="0" name="tytulFilmu" >
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:maxLength value="60"/>
    </xs:restriction>
  </xs:simpleType>
</xs:element>

```

III. Z elementami o nazwach *gatunek* oraz *rodzaj* należy postąpić analogicznie do punktu II, ustawiając odpowiednio wartość atrybutu *maxLength* - maksymalna liczba znaków przeznaczona na zapis atrybutu "gatunek" powinna wynosić 20, natomiast dla atrybutu "rodzaj" powinna ona wynosić 30.

12. Na zakładkę *Data Flow* przeciągnij z *Toolbox*'a element *OLE DB Destination*.
13. Zmień nazwę elementu *OLE DB Destination* na "film".
14. Wyjście poprzedniego elementu (*XML Source*) połącz z wejściem tego elementu.
15. Kliknij dwukrotnie na to zadanie - otworzy się okienko z jego właściwościami. W polu *Name of the table or the view* zakładki *Connection Manager* wybierz [dbo].[Film].
16. W zakładce *Mappings* ustaw odwzorowania względem nazw (*ID_Filmu* nie jest mapowane). Następnie wybierz *OK*.
17. Wykonaj zadanie "Informacje o filmach" z zakładki *Control Flow* poprzez kliknięcie na nie prawym przyciskiem myszki i wybranie opcji *Execute Task*.
18. Sprawdź zawartość docelowej bazy danych - kliknij dwukrotnie na zadanie "film" i wybierz opcję *Preview...*, a następnie sprawdź, czy zawartość pierwszych czternastu wierszy zgadza się z poniższym rysunkiem. Tabela "FILM" powinna zawierać 100 różnych wpisów i wszystkie atrybuty powinny być poprawnie wypełnione.

| ID_FILMU | TYTUL | GATUNEK | RODZAJ | CENA |
|----------|-------------------------|---------------------|--------------|------|
| 1 | Blok | sciene-fiction | czarno-biały | 4 |
| 2 | Kula | sciene-fiction | 3D | 2 |
| 3 | Cena honoru | horror | czarno-biały | 5 |
| 4 | Historie kuchenne | sciene-fiction | 3D | 5 |
| 5 | Casino Royale | sciene-fiction | kolorowy | 2 |
| 6 | Zagubieni w kosmosie | thriller | czarno-biały | 6 |
| 7 | Mission: Impossible III | komedia romantyczna | kolorowy | 9 |
| 8 | Raport mniejszości | thriller | 3D | 4 |
| 9 | Silent Hill | wojenny | czarno-biały | 8 |
| 10 | Matrix: Rewolucje | romans | kolorowy | 8 |
| 11 | Dirty Dancing 2 | komedia | 3D | 4 |
| 12 | Ajlawju | wojenny | 3D | 7 |
| 13 | Czas, który pozostał | komedia | 3D | 2 |
| 14 | Lara Croft: Tomb R... | thriller | kolorowy | 4 |

Rys. 1: Zawartość wypełnionej tabeli "FILM"