

Jacek BARTMAN

Uniwersytet Rzeszowski, Polska

Dariusz SOBCZYŃSKI

Politechnika Rzeszowska, Polska

Dane przestrzenne w relacyjnych bazach danych

Wstęp

Nowoczesne społeczeństwo to społeczeństwo informacyjne, w którym informacja jest najbardziej wartościowym i pożądanym towarem. Dlatego konieczne jest gromadzenie i szybkie oraz efektywne przetwarzanie danych w celu uzyskania z nich niezbędnych informacji. Aktualnie gromadzimy coraz to większe ilości danych, ale co bardzo charakterystyczne, dane te pochodzą z bardzo różnych źródeł. Do klasycznych źródeł danych, takich jak dokumenty czy pomiary, w ostatnich latach dołączyły również zdjęcia, mapy, dźwięk, strony internetowe czy filmy. Dane te mają zupełnie inny charakter i właściwe ich gromadzenie oraz przetwarzanie stanowi poważne wyzwanie. W niniejszej publikacji przedstawimy konsekwencje tych zmian – korzyści, jakie uzyskujemy z poszerzenia zakresu przechowywanych i przetwarzanych danych, oraz problemy z tym związane.

Bazy danych jako źródło informacji

Informacja może zostać należycie wykorzystana jedynie wtedy, gdy jest pełna, wiarygodna, uporządkowana i posiada odpowiednią formę.

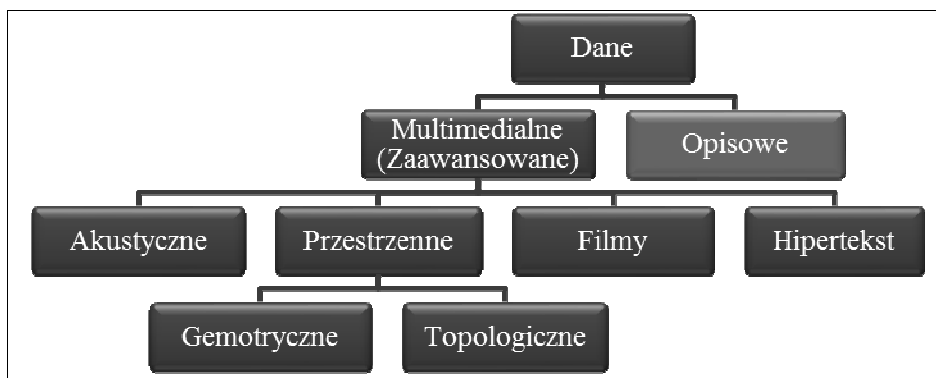
Pełną informację możemy uzyskać tylko wtedy, gdy gromadzone dane w pełni opisują wszystkie interesujące nas cechy rzeczywistości, gdy stanowią wierne odzwierciedlenie swego obszaru analizy. Dane niekompletne mogą być przyczyną wprowadzenia w błąd, jeżeli użytkownik nie jest świadom ich niekompletności. Natomiast gdy użytkownik ma świadomość niekompletności danych, musi szukać dodatkowych informacji w innych źródłach, jeżeli pragnie zminimalizować swoją niepewność (swój stopień niewiedzy) w interesującym go obszarze rzeczywistości. Z punktu widzenia komplementarności informacji możliwość zapisu w bazach danych danych w nowych formatach (obrazy, dźwięk, filmy) jest bardzo pożądana.

Informacja niewiarygodna to informacja bezwartościowa. Trudno ją nazwać nawet informacją. Przyczyną niewiarygodności informacji jest niewiarygodność danych i tu mamy kolejny atut nowych rodzajów formatów danych, które pozwalają zapisać w bazie materiał nieprzetworzony, np. mapę zamiast jej opisu.

Posortowanie, czyli uporządkowanie danych, ułatwia ich analizę oraz przeszukiwanie, dzięki czemu dostęp do informacji jest łatwiejszy.

W celu gromadzenia, przetwarzania i udostępniania informacji stosuje się bazy danych, które winny być odpowiednio zaprojektowane – zgodnie z wymogami modelu danych, który chcemy zastosować [Bartman 2013: 1]. Odpowiednia forma danych służy przede wszystkim zachowaniu zawartych w nich informacji. Jednocześnie forma danych winna ułatwiać proces ich gromadzenia oraz przetwarzania. Poszczególne właściwości modelowanej rzeczywistości zapisuje się w danych w postaci atrybutów, które ze względu na rodzaj przechowywanej w nich informacji dzielimy na (rys. 1):

- atrybuty opisowe – mające charakter zapisów w postaci słów, liczb, dat itp. symbolicznie wyrażających określoną cechę,
- atrybuty multimedialne – mające charakter złożony, wyrażające często w sposób niepodzielny wiele cech modelowanej rzeczywistości.



Rys. 1. Podział danych gromadzonych we współczesnych bazach danych

Specyfika relacyjnego modelu danych

Relacyjny model danych został opublikowany w 1970 r. przez E. Codda. Jego siłą są mocne podstawy teoretyczne gwarantujące przewidywalność przetwarzania danych [Bartman 2013: 1]. Model zbudowany został na bazie teorii zbiorów oraz teorii predykatów i stanowi podstawę relacyjnych baz danych. Zasadą, na której został zbudowany model relacyjny, jest założenie, iż typowa baza danych składa się z szeregu nieuporządkowanych relacji, którymi można manipulować, używając nieproceduralnych operacji zwracających całe relacje. Zgodnie z założeniami modelu relacja służy do reprezentowania „różnych rzeczy” ze świata rzeczywistego i każda powinna reprezentować tylko jedną taką rzecz – relacja powinna opisywać tylko jeden rodzaj rzeczy. Rzeczy mogą być rzeczywistymi podmiotami bądź zdarzeniami. Na przykład rzeczywistym obiektem może być klient, przedmiot z inwentarza lub faktura. Przykładami zdarzeń

mogą być wizyty pacjentów, zlecenia czy rozmowy telefoniczne. Relacje stanowią zbiór jednakowych rekordów zapisanych w formie tablic spełniających następujące warunki:

- każda relacja w bazie danych ma jednoznaczną nazwę,
- każda kolumna ma jednoznaczną nazwę w ramach relacji,
- wszystkie wartości w kolumnie są tego samego rodzaju,
- porządek kolumn w relacji nie jest istotny,
- kolejność wierszy nie jest istotna,
- wiersze mają unikalne wartości (są różne),
- każde pole relacji winno zawierać elementarne wartości.

Większość z wymienionych wyżej cech wydaje się oczywista, jedynie dwie ostatnie stanowią wyróżnik relacji wobec zwykłych tabel.

W celu zapewnienia unikalności wierszy stosuje się w modelu relacyjnym klucze pierwotne (klucze główne), które umożliwiają jednoznaczną identyfikację krotek i pozwalają niezależnie od struktury rekordu i charakteru gromadzonych w jego polach danych uniknąć niejednoznaczności. Wypełnienie więc tego warunku jest możliwe również w sytuacji, gdy w polach rekordu wpisywane są wartości o charakterze nierelacyjnym.

Wymóg, aby każde pole zawierało wartości elementarne, jest bardzo ważny, gdyż spełnienie go ułatwia proces wyszukiwania danych w tabelach. Podczas projektowania baz danych wymóg ten jest wymuszany poprzez pierwszą postać normalną mówiącą o funkcyjnej zależności atrybutów niekluczowych od klucza głównego. Oczywiście elementarność danych należy traktować w sensie semantycznym, co sprowadza się do tego, iż pewne dane mogą w jednej bazie być elementarne, a w innej już nie. Tym niemniej wymóg ten jest praktycznie nie do spełnienia przez dane zaawansowane, a więc również dane przestrzenne.

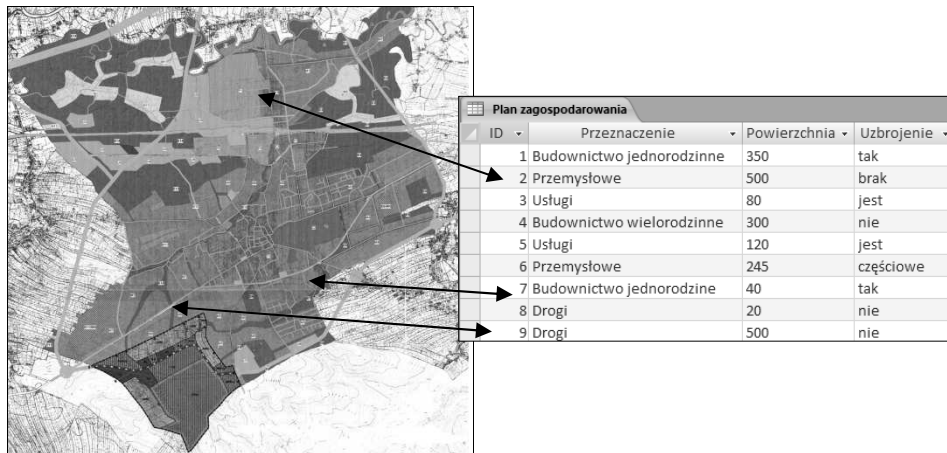
Obowiązujący standard najpopularniejszego języka dedykowanego do obsługi relacyjnych baz danych – języka SQL – nie przewidywał konstrukcji, które pozwalałyby na obsługę danych przestrzennych czy jakichkolwiek danych multimedialnych. Podczas prac nad uzupełnieniem standardu o obsługę zaawansowanych typów danych zwyciężyła koncepcja opracowania nowego standardu zawierającego specyfikacje bibliotek opartych na typach obiektowych SQL99, które byłyby dedykowane do obsługi specjalistycznych typów danych i aplikacji [Jankiewicz 2014: 2]. Nowy standard funkcjonuje pod nazwą SQL/MM. Wykorzystanie w specyfikacji standardu SQL/MM obiektowych typów SQL daje możliwość wykorzystania bibliotek z poziomu języka SQL, co zwiększa funkcjonalność systemu.

Dane przestrzenne i relacyjna baza danych

Klasyczne bazy danych zawierające tylko dane o charakterze opisowym nie są praktycznym narzędziem do obrazowania przestrzennego położenia obiektów.

Kłopotliwe jest bezpośrednie ich wykorzystanie w celu np. planowania zabudowy, zarządzania infrastrukturą itp.

Systemy uwzględniające geograficzną lokalizację opisywanych obiektów określa się mianem Systemów Informacji Geograficznej (GIS). Najprościej GIS można zdefiniować jako połączenie mapy z klasyczną bazą danych. Mapa przechowuje dane przestrzenne w postaci graficznej, zaś baza danych (tablice) przechowuje dane opisowe, przedstawia cechy atrybutowe obiektów i zjawisk. Oba te elementy są ze sobą połączone w taki sposób, aby dostarczać najpełniejszej informacji: zarówno w postaci graficznej, jak i opisowej (rys. 2).



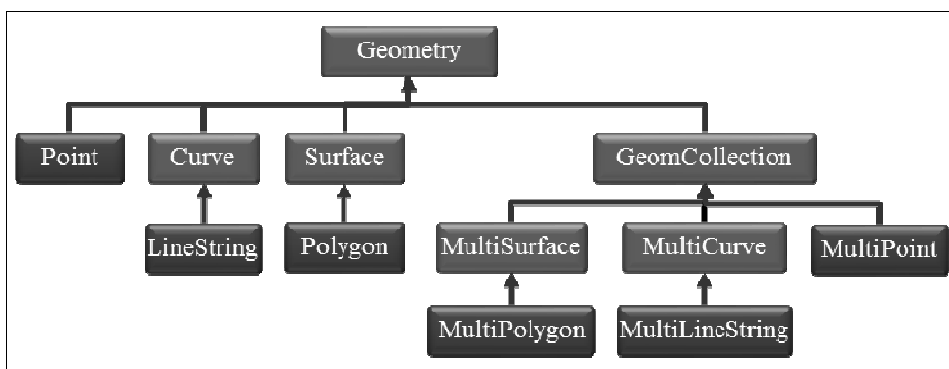
Rys. 2. Przykład powiązania danych opisowych i przestrzennych

W ostatnich latach większość z najpopularniejszych systemów zarządzania relacyjnymi bazami danych (np. MS SQL, Oracle) została wyposażona w możliwość zapisu danych przestrzennych pomimo ich nieelementarności i niemożności zadośćuczynienia formalnym wymogom podanym przez Codda.

Dane przestrzenne w MS SQL

MS SQL Server oferuje dwa typy danych przestrzennych: typ `GEOMETRY` oraz typ `GEOGRAPHY`. Są one dostarczane w bibliotekach `SqlServerSpatial.dll` oraz `Microsoft.SqlServer.Types.dll`. Oba typy są zgodne ze specyfikacją stworzoną przez konsorcjum OGC (Open Geospatial Consortium). Model OGC wprowadza (rys. 3):

- klasy abstrakcyjne: obiekt geometryczny, powierzchnia, krzywa, kolekcja, multipowierzchnia, multikrzywa,
- klasy zwykłe: punkt, łamana, odcinek, pierścień, wielobok, multipunkt, multikrzywa, multiwielobok,



Rys. 3. Model danych przestrzennych zaproponowany przez OGS

Dane przestrzenne o charakterze geometrycznym w MS SQL Server przechowywane są jako kolumny typu GEOMETRY zapisane w płaskim euklidesowym układzie współrzędnych, nieposiadające jednostki miary [Microsoft 2014: 4]. Dane przestrzenne o charakterze geograficznym w MS SQL Server przechowywane są jako dane typu GEOGRAPHY zapisane w układzie elipsoidalnym (w układzie „okrągłej ziemi”) [Microsoft 2014: 4]. Do ich zapisu wykorzystywane są współrzędne geograficzne – długość i szerokość. Dla obu typów danych przestrzennych podczas ich tworzenia konieczne jest podanie wartości identyfikatora SRID (*Spatial Reference System Identifier*). Dla typu GEOMETRY jego domyślna wartość wynosi 0 (identyfikator ten jest praktycznie bez znaczenia), zaś dla danych typu GEOGRAPHY domyślną wartością identyfikatora SRID jest 4326. Dostępne w MS SQL Server wartości identyfikatora SRID (oprócz wartości 0) można przeglądać, wykonując zapytania do widoku katalogowego `sys.spatial_reference_systems`:

```
SELECT * FROM sys.spatial_reference_systems
```

W kolumnie zawierającej wartość identyfikatora SRID mogą występować różne wartości, jednak w przypadku wykonywania operacji na dwóch instancjach o różnych identyfikatorach SRID metody typów przestrzennych MS SQL zwracają wartość NULL.

Typ GEOMETRY (podobnie jak typ GEOGRAPHY) jest typem abstrakcyjnym. Rodzaj danych przestrzennych (krzywa itd.) użytkownik określa dopiero podczas tworzenia obiektu. Przykładowo, jeżeli zdefiniowaliśmy tabelę:

```
CREATE TABLE Geo (id int primary key, g geometry);
```

to w celu wprowadzenia do niej danych należy zastosować zapis:

```
INSERT INTO Geo
VALUES (0, geometry::STPointFromText('POINT(1 2)',0));
```

Powyższy zapis korzysta z formatu tekstowego WKT (*Well Known Text*), zgodnie z którym polecenia zbudowane są według schematu: ST[rodzaj] FromText, gdzie w miejsce [rodzaj] wstawiana jest nazwa tworzonego obiektu zgodna z nazwami podanymi na rys. 3. Podczas definiowania danych oprócz systemu WKT możemy wykorzystać WKB (*Well Known Binary*) w przypadku danych binarnych oraz system GML (*Geography Markup Language*) w przypadku dokumentów XML [Microsoft 2014: 4]. Obiektowy charakter zmiennych czy kolumn implikuje konieczność korzystania z zapisów pozwalających na odwołanie się do metod dla nich dostępnych. Przykładowo wykonajmy zadanie polegające na zadeklarowaniu, utworzeniu oraz zwróceniu współrzędnych punktu na płaszczyźnie:

```
DECLARE @punkt geometry;  
SET @punkt = geometry::STGeomFromText('POINT(1 2)', 0);  
SELECT @punkt.STAsText() AS OpisPunktu
```

Podsumowanie

Wprowadzenie obsługi danych zaawansowanych (multimedialnych) poprzez systemy zarządzania oparte o relacyjne bazy danych było koniecznością i krok ten został wykonany. Jednak jego realizacja wymagała odstąpienia od pewnych wymogów stawianych relacjom. W konsekwencji uzyskaliśmy konstrukcję, która zapełnia powstałą lukę, jest ona jednak protezą i należy oczekiwać rozwiązań, które wprowadzą nowy architektoniczny model danych pozwalający w jednolity sposób korzystać z różnorodnych formatów zapisu.

Literatura

- Bartman J. (2013): *Bazy danych*, Rzeszów.
- Jankiewicz K., Wojciechowski M. (2004): *Standard SQL/MM: SQL Multimedia and Application Packages*, IX Seminarium PLUG „Przetwarzanie zaawansowanych struktur danych: Oracle interMedia, Spatial, Text i XML DB”.
- Litwin P. (1995): *Podstawy projektowania relacyjnych baz danych*, Software.
- Microsoft (2014): *Books Online for SQL Server 2014*, msdn.microsoft.com/en-us/library/ms130214.aspx.

Streszczenie

Do klasyczne danych zapisywanych w bazach danych w formie opisowej w ostatnim okresie dołączyły dane zaawansowane o zupełnie innym charakterze. Właściwe ich gromadzenie oraz przetwarzanie stanowi poważne wyzwanie. W niniejszej publikacji przedstawiamy korzyści, jakie uzyskujemy z poszerzenia typów przechowywanych i przetwarzanych danych, oraz problemy z tym związane. Jako przykład zaprezentowano implementację danych przestrzennych w systemie MS SQL.

Słowa kluczowe: relacyjne bazy danych, dane przestrzenne.

Spatial Data in Relational Databases

Summary

To the classic data stored in databases in a descriptive, recently joined Advanced data of a completely different nature. Competent their collection and processing is a major challenge. In this publication we present the benefits we get from the expansion of the types of data stored and processed, and problems associated with it. As an example of the implementation of spatial data presented MS SQL.

Keywords: relational database, spatial data.