

Andrzej Niesler

METAJĘZYK XML I WBUDOWANE BAZY DANYCH W PROCESIE INTEGROWANIA SYSTEMÓW INFORMATYCZNYCH

1. Wstęp

Powstanie i rozwój technologii bazodanowych zapoczątkowały aplikacje, w których wprowadzono strukturalizację i ujednolicony sposób składowania przetwarzanych danych. Stosowane wówczas rozwiązania w zakresie gromadzenia i operowania danymi stanowiły integralną część kodu programu, statycznie wkomponowaną w przebieg jego wykonywania. Brak jakichkolwiek możliwości wykorzystania tych mechanizmów przez inne aplikacje powodował, że musiały być zaimplementowane w każdej z nich. Korzystanie z przechowywanych danych odbywało się poprzez fizyczny dostęp do plików systemu operacyjnego.

Wraz z rozwojem technologii informacyjnych ten pierwotny model przetwarzania danych podlegał licznym przemianom, związanym z ewoluującymi potrzebami rynku. Z kodu programu wyodrębniono części odpowiedzialne za składowanie i manipulowanie danymi. Początkowo występowały one w postaci przystosowanych do wielokrotnego wykorzystania bibliotek procedur, następnie niezależnych aplikacji funkcjonujących jako motor bazodanowy, aż po rozbudowane, kompleksowe systemy zarządzania bazą danych (SZBD). Potrzeba integracji źródeł danych, reguł przetwarzania i protokołów komunikacyjnych między różnymi systemami informatycznymi znalazła odzwierciedlenie w uporządkowaniu od strony metodologicznej stosowanych modeli danych oraz w powstaniu koncepcji języka zapytań jako uniwersalnego narzędzia komunikacji z SZBD. Upowszechnienie się sieci komputerowych umożliwiło dodatkowo realizację idei scentralizowanego gromadzenia danych i administrowania nimi, niezależnie od miejsca ich fizycznego powstawania lub dalszego wykorzystywania.

Do momentu rewolucji komunikacyjnej, którą zapoczątkowało rozpowszechnienie się technologii internetowych i mobilnych, podstawową architekturą przetwarzania danych był centralny, niezależny SZBD, oparty na relacyjnym modelu danych,

standardzie języka zapytań SQL oraz wspierający przetwarzanie w trybie klient-serwer. W efekcie połączenia setek tysięcy lokalnych sieci komputerowych w jedną, globalną sieć teleinformatyczną powstała „nowa rzeczywistość”, w której funkcjonują obecne przedsiębiorstwa i ich systemy informatyczne. Z perspektywy technologii baz danych środowisko to charakteryzuje się dużym rozproszeniem i heterogenicznością źródeł danych, modeli danych i metadanych. Brak jednorodności widoczny jest również w przypadku procedur przetwarzania i administrowania zasobami. Przekłada się to w naturalny sposób na tendencje integracyjne w wymiarze zarówno technologicznym, jak i organizacyjnym funkcjonujących w tym środowisku podmiotów.

Nowy kierunek rozwoju technologii bazodanowych w świetle dzisiejszych realiów rynkowych wyznacza problematyka rozproszonych baz danych. Podejmuje się w niej próbę połączenia systemów bazodanowych i źródeł danych – często odmiennych, odległych fizycznie i technologicznie – w jedną całość, która od strony funkcjonalnej i administracyjnej mogłaby być postrzegana jako pojedynczy SZBD. Takie podejście wynika m.in. z przyczyn technologicznych – przyrost informacji w sieciach teleinformatycznych jest nieporównywalnie szybszy niż wzrost możliwości technologii związanych z przechowywaniem i przetwarzaniem danych. Z drugiej strony istotny wpływ na to zagadnienie mają też czynniki pozatechniczne. W wielu przypadkach przechowywanie i przetwarzanie informacji w miejscu ich powstawania i wykorzystywania jest bardziej efektywne i mniej kosztowne niż w klasycznym wariantcie scentralizowanym. Odpowiada to również dzisiejszym tendencjom rynkowym, zgodnie z którymi centralne bazy przedsiębiorstw są zasilane informacjami lub świadczą dostęp do nich coraz większej liczbie użytkowników. Źródłem danych lub klientem dla takiej bazy może być w dobie komunikacji bezprzewodowej zarówno przemysłowy system informatyczny, domowy komputer podłączony do Internetu, notebook handlowca, jak i cała gama miniaturowych urządzeń przenośnych: od elektronicznych notatników po telefony komórkowe.

Przedstawione tendencje paradoksalnie doprowadziły do sytuacji, w której zaczyna się powracać do pierwotnego modelu, znanego z pierwszych aplikacji bazodanowych. Koncepcja wbudowanych baz danych (*embedded databases*) zakłada zuniifikowanie funkcjonalności znanej z dzisiejszych SZBD z kodem aplikacji.

W artykule prezentowana jest idea wykorzystania wbudowanych baz danych obsługujących standard XML na potrzeby integracji systemów informatycznych. Zastosowanie tych nowych technologii bazodanowych jest analizowane z uwzględnieniem specyfiki środowiska rozproszonego. Takie podejście uwzględnia w szczególności istotny obecnie problem integrowania źródeł danych semistrukturalnych w środowiskach heterogenicznych.

2. Semistrukturalność danych i metadanych

Rola współczesnych systemów informatycznych coraz częściej zaczyna wykroczać poza proste wspomaganie tradycyjnych obszarów funkcjonowania przedsię-

biorstwa, takich jak: produkcja, księgowość czy gospodarka magazynowa. Dzisiejsze uwarunkowania rynkowe, ale także wewnętrzne potrzeby organizacji w ogólności i jej kadry zarządczej w szczególności, wymuszają rozwój funkcjonalny systemów w kierunku coraz bardziej zaawansowanego wspomagania analitycznego. Konkurencyjność przedsiębiorstwa jest w coraz większym stopniu uzależniona od precyzyjnych, zweryfikowanych jakościowo informacji, których dostarczanie wykracza poza możliwości typowych systemów poziomu operacyjnego. W wielu koncepcjach zarządzania pojawiają się postulaty technologicznego wspomagania procesów pozyskiwania i wykorzystywania wiedzy jako jednego z najważniejszych zasobów przedsiębiorstwa.

Systemy informatyczne zatem są rozwijane w kierunku podniesienia ich możliwości analitycznych oraz operowania informacją i wiedzą na potrzeby zarządzania. Aby aparat analityczny spełniał stawiane przed nim zadania, konieczne jest stworzenie warunków do gromadzenia odpowiednio dużej ilości szczegółowych danych oraz zapewnienie mechanizmów efektywnego pozyskiwania informacji w nich zawartych. Co więcej, istotna zmiana następuje również w strukturze gromadzonych i przetwarzanych danych. Charakter informacji, którymi operują systemy analityczne, jest bardziej zróżnicowany od tych znanych z systemów poziomu operacyjnego (na przykład rejestrujących powtarzalne transakcje). Dane pochodzą z wielu heterogenicznych źródeł. Część z nich jest kontrolowana przez przedsiębiorstwo, inne – zewnętrzne, pozostają zwykle poza jego wpływem. Proces integracji takich systemów informatycznych, na samym tylko poziomie wymiany danych, jest w związku z tym złożony.

Podstawowym obszarem pozyskiwania i wymiany informacji dla nowej klasy systemów stają się (poza lokalnymi systemami operacyjnymi przedsiębiorstwa) heterogeniczne środowiska rozproszone, których najlepszym przykładem jest Internet. Natura informacji przechowywanych w tego typu sieciach powoduje, że dane z nich pozyskiwane mają charakter semistukturalny. Semistukturalność oznacza w tym wypadku istnienie określonej struktury danych, ale struktura ta może być nieregularna lub niekompletna. Do tego może ona podlegać częstym zmianom w czasie, i to w sposób nieprzewidywalny [COBE04, s. 500]. Nieregularność i niekompletność jest prostą konsekwencją heterogeniczności samego środowiska. Różnorodność systemów i modeli danych powoduje, że informacje tego samego typu przechowywane są w różny sposób i z uwzględnieniem różnego poziomu szczegółowości. Zmienność wynika natomiast z rozproszenia zasobów i charakteru poszczególnych struktur danych. Typowym tego przykładem mogą być złożone struktury, które współtworzy kilka wzajemnie powiązanych zbiorów danych, pozostających w różnych fizycznie miejscach sieci. Nieobecność mechanizmów wymuszających utrzymywanie spójności lub kontrolę zmiany zawartości informacyjnej poszczególnych składowych stwarza możliwość zaistnienia sytuacji, w której nastąpi nieoczekiwana zmiana treści lub utrata kompletności w ramach całej struktury. Pomimo tych własności (a czasem wręcz dzięki nim) wykorzystanie źródeł danych semistukturalnych jest często niezwykle przydatne.

W odniesieniu do tego typu danych nie jest z góry wymagana zgodność z żadnym ustalonym schematem, jak ma to miejsce np. w modelu relacyjnym czy obiektowym. Informacje, które taki schemat zwykle zawiera, umieszczane są w tym wypadku razem z danymi, przez co dane semistrukturalne określa się często mianem samoopisujących lub bezschematowych [PGMW95, s. 4]. W pewnym sensie schemat jest tworem wtórnym w stosunku do danych, co jest podejściem odwrotnym do powszechnie stosowanego definiowania schematu *a priori*. Istnienie zewnętrznego schematu umożliwia weryfikowanie danych zgodnie z zadanymi regułami. Nie warunkuje to jednak w żaden sposób możliwości odczytywania ich lub operowania nimi w podstawowym zakresie (por. [COBE04, s. 499 i n.]).

```

Organizacja (&1)
  Biuro (&2)
    ulica (&7) "Floriańska 6"
    Dyrektor &3
  Personel (&3)
    imięNazwisko (&8)
      imię (&13) "Marek"
      nazwisko (&14) "Kowalski"
    pensja (&9) "13 500"
    DyrektorBiura &2
  Personel (&4)
    imięNazwisko (&10) "Jan Ryng"
    pensja (&11) 12000
    Nadzoruje &5
  Magazyn (&5)
    adres (&12) "Wiślana 27"
    NadzorowanaPrzez &4

```

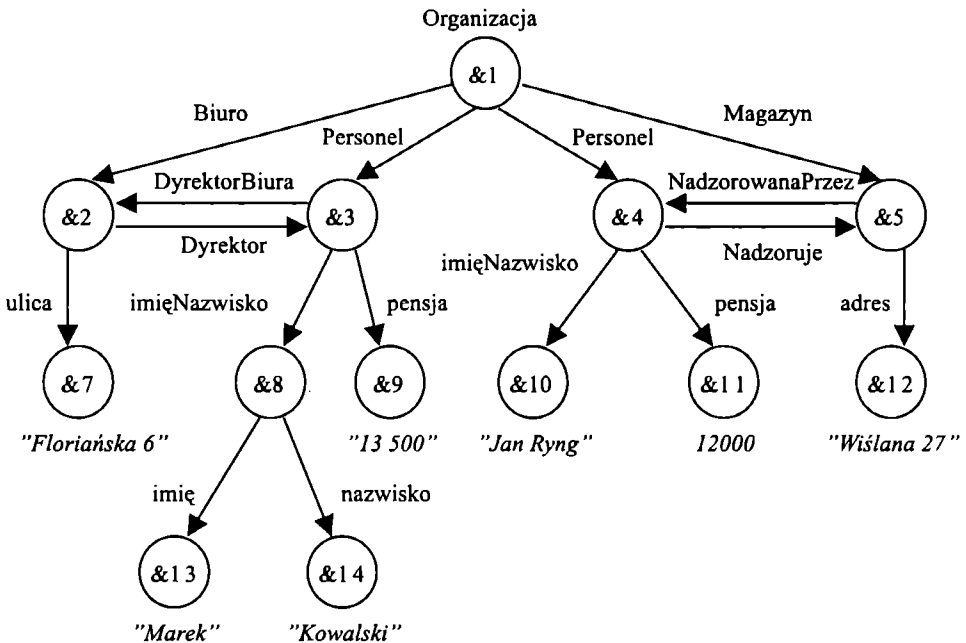
Rys. 1. Przykładowa reprezentacja danych semistrukturalnych

Źródło: opracowanie własne na podstawie [COBE04, s. 501].

Na rysunku 1 przedstawiona została przykładowa reprezentacja danych semistrukturalnych. Dane obejmują informacje dotyczące takich obiektów organizacji, jak: biuro, dwóch zatrudnionych pracowników i magazyn. Imię i nazwisko jednego pracownika zostały wprowadzone osobno, drugiego – w postaci pojedynczego ciągu. Nieregularność przejawia się również w odmiennym traktowaniu wartości pensji – raz zapisana jest jako wartość liczbowa, innym razem jako ciąg znaków.

Na rysunku 2 dane te zostały przedstawione za pomocą grafu skierowanego. W obydwu przypadkach zastosowano podejście zgodne z rozszerzonym OEM (Object Exchange Model) [PGMW95], według którego pojedynczy element danych jest obiektem opisywanym przez czwórkę: identyfikator obiektu (np. &14), opisującą

etykietę tekstową (np. nazwisko), typ (np. ciąg znaków) i wartość (np. „Kowalski”). Obiekty dzielą się na atomowe i złożone. Identyfikacja poszczególnych obiektów następuje przez określenie ich miejsca w kolekcji i etykiety tekstowej. Identyfikatory pozwalają na swobodne definiowanie występujących między nimi wzajemnych relacji, w ramach danej kolekcji.



Rys. 2. Graficzna reprezentacja danych przedstawionych na rys. 1

Źródło: opracowanie własne na podstawie [COBE04, s. 502].

Zaakceptowanie semistrukturnalnego charakteru danych wydaje się być racjonalne w odniesieniu do integracji heterogenicznych środowisk rozproszonych. Pozwala ono na tworzenie rozwiązań wprowadzających oczekiwaną funkcjonalność, a jednocześnie umożliwia zachowanie niskiego poziomu inwazyjności w stosunku do integrowanych elementów. Osiągnięta w ten sposób elastyczność i łatwość dalszej rozbudowy systemu jest jednak uzyskiwana kosztem zwiększonych wymagań w zakresie mocy obliczeniowej i złożoności samej procedury przetwarzania danych. System, poza podstawowym dotąd poziomem weryfikacji danych – poprawnością syntaktyczną, musi teraz rozpatrywać kolejne, jak np. kompletność strukturalna, i w odpowiedni sposób reagować na związane z tym możliwe stany szczególne.

Wykorzystanie danych semistrukturnalnych w procesie integrowania systemów informatycznych na dużą skalę wymaga jednak istnienia powszechnie akceptowanego standardu, wyznaczającego podstawowy zbiór zasad formatowania takich danych. Role tę obecnie zaczyna spełniać metajęzyk XML (eXtensible Markup Language).

3. Metajęzyk XML jako uniwersalny standard wymiany danych

XML jest metajęzykiem opisu danych opracowanym na podstawie standardu SGML (Standard Generalized Markup Language). Pierwotnym, przewidywanym przez jego twórców zastosowaniem było formatowanie dokumentów udostępnianych w sieci WWW, jednak dzięki uniwersalności i łatwej rozszerzalności obszar potencjalnych zastosowań jest znacznie większy. Od momentu powstania standardu XML, zatwierdzonego w wersji 1.0 przez W3C (World Wide Web Consortium) w 1998 r., minęło już kilka lat i wszystko wskazuje na to, że został on w pełni zaakceptowany przez rynek, jako uniwersalny format wymiany danych.

Zgodnie z oryginalnym przeznaczeniem, dokumenty sformatowane w XML są dobrze przystosowane do zastosowań w heterogenicznym środowisku rozproszonym. Oznacza to m.in. zgodność z internetowymi protokołami komunikacyjnymi i niezależność od platformy sprzętowej lub programowej, możliwość swobodnego definiowania związków (odnośników) względem innych zasobów oraz oddzielenie zawartości informacyjnej od formy jej prezentacji. Dowolna struktura danych, oparta na każdym z obecnie stosowanych schematów, może być przedstawiona w reprezentacji XML. Dotyczy to oczywiście również dowolnych danych semistrukturalnych. Rozszerzalność języka jest gwarancją, że wszystkie obecne oraz przyszłe schematy danych będzie można rzutować na odpowiedni, kompletny zestaw wymaganych znaczników. Dzięki standardom pochodnym względem XML, dane za jego pomocą sformatowane można powiązać z zewnętrznymi dokumentami opisującymi m.in.: reguły poprawności składniowej (schemat), dyrektywy dotyczące formy prezentacji danych czy też reguły zautomatyzowanej transformacji do innych formatów. Umożliwia to funkcjonowanie XML jako bazowego standardu przechowywania dowolnych danych, które na potrzeby składowania konwertowane są do postaci XML, a w razie potrzeby użycia bezproblemowo i automatycznie przekształcane ponownie do formatu źródłowego. Baza danych przechowująca dane w źródłowej postaci XML może funkcjonować równolegle ze stosowanymi do tej pory SZBD.

Podejście to jest szczególnie przydatne w procesie integracji systemów informatycznych. Dodanie tak działającego modułu repozytorium wspierającego standard XML w niewielkim stopniu ingeruje w istniejące rozwiązania, a jednocześnie umożliwia uniwersalną wymianę danych z każdym innym systemem obsługującym standard XML. Zastosowanie interfejsów typu SAX (Simple API for XML) lub DOM (Document Object Model) pozwala na reprezentację obiektową dokumentu XML w pamięci aplikacji. Ułatwia to w dużym stopniu operowanie danymi z poziomu aplikacji pisanych w językach obiektowych. Przetwarzanie dokumentu odbywa się poprzez odwzorowanie danych w reprezentację zbioru obiektów i metod udostępniających mechanizmy manipulowania nimi.

Na rysunku 3 została przedstawiona reprezentacja omawianych wcześniej przykładowych danych semistrukturalnych z wykorzystaniem notacji standardu XML.

```

<?xml version="1.0"?>
<!DOCTYPE ORG "c:\org.dtd"
<Organizacja id="1" >
  <Biuro id="2" >
    <ulica id="7" >Floriańska 6</ulica>
    <Dyrektor hRef="3" />
  </Biuro>
  <Personel id="3" >
    <imięNazwisko id="8" >
      <imię id="13" >Marek</imię>
      <nazwisko id="14" >Kowalski</nazwisko>
    </imięNazwisko>
    <pensja id="9" typ="ciąg" >13 500</pensja>
    <DyrektorBiura hRef="2" />
  </Personel>
  <Personel id="4" >
    <imięNazwisko id="10" >Jan Ryng</imięNazwisko>
    <pensja id="11" typ="liczba" >12000</pensja>
    <Nadzoruje hRef="5" />
  </Personel>
  <Magazyn id="5" >
    <adres id="12" >Wiślana 27</adres>
    <NadzorowanaPrzez hRef="4" />
  </Magazyn>
</Organizacja>

```

Rys. 3. Przykładowa reprezentacja danych semistrukturalnych w notacji XML

Źródło: opracowanie własne.

Zastosowanie standardu XML w procesie integracji systemów informatycznych nie ogranicza się jedynie do najniższego poziomu reprezentacji danych. Uniwersalność tego podejścia pozwala na wykorzystanie metajęzyka XML do opisu informacji na wyższym poziomie złożoności i abstrakcji. Branżowe organizacje standaryzacyjne opracowują zestawy własnych znaczników, które odpowiadają potrzebom i specyfice zastosowań charakterystycznych dla konkretnej dziedziny. Następuje również migracja środowisk elektronicznej wymiany danych (Electronic Data Interchangei – EDI) do rozwiązań bazujących na technologiach internetowych i XML. Wszechobecność Internetu pozwala na odejście od zestawiania trwałych połączeń typu punkt–punkt na rzecz komunikacji *ad hoc* [CUMM02, s. 256].

Obszar zastosowań XML stale się powiększa, a w nowoczesnych technologiach integracyjnych stał się już wręcz normą. Wystarczy wymienić tutaj tylko standardy opisu komunikatów na platformach *middleware* czy charakterystyki (Web Services Description Language – WSDL) i rejestry (Universal Description, Discovery and Integration – UDDI) usług sieciowych (*web services*). W praktyce stwarza to sytuację, w której XML staje się kluczową technologią integracyjno-komunikacyjną. Zaimplementowanie jej w systemie informatycznym jest podstawą wymiany informacji i kooperowania z innymi systemami w przyszłości.

4. Koncepcja wbudowanych baz danych

Większość współczesnych SZBD oferuje na podstawowym poziomie określoną minimalną funkcjonalność, która najczęściej obejmuje takie cechy, jak: równoczesny dostęp wielu użytkowników, obsługa transakcji, wprowadzanie i udostępnianie danych poprzez zestandaryzowany język zapytań czy weryfikację danych zgodnie z ustalonym schematem. Cechy te występują zwykle we wszystkich dużych, scentralizowanych SZBD. Wobec rozwoju środowisk rozproszonych i związanej z tym decentralizacji przetwarzania danych coraz częściej funkcjonalność dostępna w klasycznych systemach bazodanowych znajduje zastosowanie w pojedynczych aplikacjach, które przetwarzają dane w odpowiednio mniejszej skali lub które nie mają stałej możliwości komunikowania się z centralną bazą danych. W takich sytuacjach nie ma potrzeby, a często nawet możliwości, instalowania rozbudowanych systemów, które poza dużymi wymaganiami sprzętowymi wprowadzają dodatkowo konieczność administrowania lub dostrajania przez wykwalifikowanego pracownika.

W przypadku systemów OLTP (On-Line Transaction Processing) obserwowany jest coraz częściej problem dużego spadku wydajności w sytuacjach narastającego obciążenia. Koszt przeprowadzenia prostej operacji pozyskiwania danych, obejmującej takie etapy, jak: sformułowanie zapytania w języku SQL, wysłanie go przez sieć do bazy danych, obsłużenie zapytania przez bazę, przesłanie wygenerowanej odpowiedzi do systemu i zinterpretowanie wyniku, staje się zbyt duży w kontekście milionów, kolejno wykonywanych, analogicznych operacji.

Próbą rozwiązania wymienionych i podobnych problemów stają się obecnie wbudowane bazy danych (*embedded databases* – EmDB). Idea EmDB zakłada wbudowanie podstawowych mechanizmów SZBD bezpośrednio w kod aplikacji w takim zakresie, jaki jest niezbędny do realizowanych przez nią zadań. Aplikacja ma pełną kontrolę oraz bezpośredni dostęp do danych i mechanizmów bazodanowych, ponieważ znajdują się one wewnątrz jej przestrzeni adresowej. Oznacza to skrócenie praktycznie do minimum czasu potrzebnego na operowanie danymi, przy jednoczesnym zachowaniu użyteczności typowej dla wydzielonych, scentralizowanych SZBD. Nie ma potrzeby dokonywania przeformatowywania danych, formułowania zapytań SQL czy przesyłania czegokolwiek przez sieć. Model danych w przypadku EmDB jest zwykle zgodny z językiem programowania, w którym napisano aplikację. Poza takim trybem dostępu, określanym zwykle jako „natywny”, wbudowana baza danych może być wyposażona również w interfejsy obsługujące powszechnie wykorzystywane standardy relacyjne, obiektowe czy oparte na języku XML [CHRZ03, s. 69 i n.].

Skalowalność wbudowanych baz danych jest stosunkowo duża. Firmy opracowujące tego typu technologie zwykle mają w ofercie co najmniej kilka wariantów sprzedawanego oprogramowania. Od podstawowych wersji z pojedynczym procesem zapisującym i wieloużytkownikowym odczytem danych, poprzez pełny, równoległy wielodostęp i transakcje, aż po specjalizowane warianty na potrzeby replikacji i równoważenia obciążenia (*load balancing*). W związku z rozwojem urządzeń

przenośnych wyposażonych w łączność bezprzewodową i minisystemy operacyjne stosowane są również wersje mikro, które zajmują poniżej 100kB przestrzeni adresowej systemu. Osadzenie motoru bazodanowego wewnątrz aplikacji powoduje marginalizację lub wręcz wyeliminowanie roli administratora SZBD. Większość prac, które są związane z tym stanowiskiem, przejmuje programista i w dużej mierze wykonywane są one tylko raz, na etapie tworzenia i wdrażania systemu.

5. Wykorzystanie wbudowanych baz danych w procesie integrowania systemów informatycznych

Jednym z podstawowych zagadnień w procesie integrowania systemów informatycznych jest wymiana danych. W przypadku środowiska rozproszonego występuje zwykle duża różnorodność lokalizacji i procedur przetwarzania danych. Integrowane systemy czasem pełnią skrajnie różne funkcje – od ukierunkowanych na szybkość przetwarzania rozwiązań e-biznesowych, po ultraoszczędne i kompaktowe systemy urządzeń mobilnych. Przekłada się to na takie oczekiwane charakterystyki technologii integracyjnej, jak:

- zdolność dopasowywania się do istniejących rozwiązań,
- skalowalność w zakresie wydajności przetwarzania,
- skalowalność w zakresie minimalizacji wykorzystania zasobów,
- otwartość, bazująca na wspieraniu powszechnie akceptowanych standardów.

Koncepcja wbudowanych baz danych obsługujących natywnie standard XML dobrze wpisuje się w przedstawione wymagania. Oprogramowanie tego typu dostępne jest w wersjach dla wszystkich podstawowych platform programowych, w których tworzone są współczesne rozwiązania biznesowe (C/C++, Java). Istnieje zatem możliwość zaimplementowania wybranego zakresu funkcji bazodanowych w większości aplikacji i środowisk. Zastosowanie dowolnego wariantu EmDB, spełniającego lokalne wymagania, wprowadza automatycznie zdolność kooperowania z pozostałymi, zastosowanymi w pozostałych elementach.

Wykonywanie funkcji motoru bazodanowego w przestrzeni adresowej aplikacji, stanowiące podstawę do tworzenia maksymalnie szybkich i wydajnych systemów przetwarzania danych, znajduje zastosowanie w wielu rozwiązaniach *middleware*. Tradycyjne podejście ewolucyjne w procesie integracji systemów informatycznych wymaga zwykle funkcjonalnego połączenia istniejących, nie rozwijanych już systemów (*legacy systems*) z nowymi elementami infrastruktury informatycznej. Oznacza to konieczność konstruowania interfejsów umożliwiających prostą wymianę danych z tymi systemami lub poszerzania istniejącej architektury przetwarzania danych o zaawansowane systemy pełniące rolę brokera czy multipleksa [BRITT01, s. 143]. Wkomponowywanie takich mechanizmów w procesy przetwarzania danych wymaga rozwiązań na tyle transparentnych, żeby wydajność całego układu pozostawała na poziomie porównywalnym z wyjściowym. Różnorodność interfejsów dostępnych w technologii EmDB sprawia, że bardzo dobrze nadaje się ona do zastosowania

w realizacji takich projektów. Tryb natywny gwarantuje wysoką wydajność przetwarzania, interfejsy obiektowe i relacyjne zapewniają kompatybilność z dominującymi obecnie na rynku technologiami, a wsparcie dla języka XML gwarantuje uniwersalność i wymianę danych w przyszłości.

Cechą wynikającą z rozproszenia i heterogeniczności środowisk, w których funkcjonują dzisiejsze systemy informatyczne, jest różnorodność miejsc pozyskiwania, przetwarzania i przechowywania danych. Wbudowane bazy danych umożliwiają integrowanie takich składowych rozproszonego systemu poprzez dostarczenie funkcjonalności motoru bazodanowego dokładnie w takim minimalnym wymiarze, w jakim jest wymagana, oraz dokładnie w to miejsce, w którym jest potrzebna. Pozwala to na lokalne przetwarzanie danych zgodnie z najbardziej dogodnym modelem danych i na łatwą synchronizację zasobów z centralną bazą. Cecha EmDB, wynikająca z faktu osadzania bazy w kodzie aplikacji, przekłada się również na łatwość administrowania całą infrastrukturą.

Wbudowane bazy danych wyposażone w mechanizmy bezpośredniego przetwarzania dokumentów i komunikatów w języku XML można również z powodzeniem stosować w realizacji metod integracyjnych wykorzystujących właściwości danych semistrukturnalnych. Stanowią wówczas idealne uzupełnienie technologii scentralizowanych SZBD. Umiejscowione w pobliżu źródeł powstawania danych i miejsc cząstkowego przetwarzania umożliwiają automatyczną konwersję według różnych modeli danych. Odpowiadają również za ujednolicanie danych pozyskiwanych przez system ze źródeł zewnętrznych.

6. Podsumowanie

Dane semistrukturnalne wpisują się w specyfikę heterogenicznych środowisk rozproszonych. Wykorzystanie takiego modelu danych jest wskazane z integracyjnego punktu widzenia, pociąga za sobą jednak również określone wymagania w obszarze technologii. Metajęzyk XML można już obecnie uznać za standard spełniający rolę uniwersalnego formalizmu w obszarze wymiany danych na różnych poziomach złożoności. Nadaje się on do zastosowania w przypadku zarówno struktur danych związanych ze statycznie zdefiniowanym schematem, jak i danych semistrukturnalnych. Tym samym stanowi obecnie podstawowe narzędzie integracyjne oraz punkt wyjścia do rozwijania innych metod i narzędzi z tego obszaru.

Wbudowane bazy danych wyposażone w obsługę standardu XML stanowią użyteczne narzędzie integracyjne. Ich możliwości są szczególnie przydatne w środowiskach rozproszonych, przy integrowaniu źródeł danych semistrukturnalnych. Pozwalają na rozwinięcie architektury scentralizowanego systemu zarządzania bazą danych do postaci uwzględniającej rozproszenie i heterogeniczność zewnętrznych źródeł danych.

Omawiane koncepcje i technologie odzwierciedlają kierunek, w którym ewoluje rynek rozwiązań bazodanowych. Obok tradycyjnych aspektów skalowalności

i wydajności coraz istotniejszy, a przez to bardziej widoczny, staje się również wymiar integracyjny.

Literatura

- [BRIT01] Britton Ch., *IT Architectures and Middleware: Strategies for Building Large, Integrated Systems*, Addison-Wesley, Upper Saddle River 2001.
- [CHRZ03] Chaudhri A.B., Rashid A., Zicari R., *XML Data Management: Native XML and XML-enabled Database Systems*, Pearson Education Inc., Boston 2003.
- [COBE04] Connolly T., Begg C., *Systemy baz danych. Praktyczne metody projektowania, implementacji i zarządzania*, tom 2, Wydawnictwo RM, Warszawa 2004.
- [CUMM02] Cummins F.A., *Enterprise Integration: an Architecture for Enterprise Application and Systems Integration*, John Wiley & Sons Inc., New York 2002.
- [PGMW95] Papakonstantinou Y., Garcia-Molina H., Widom J., *Object Exchange Across Heterogeneous Information Sources*, IEEE International Conference on Data Engineering, s. 251-260, Taipei 1995.

XML AND EMBEDDED DATABASES IN COMPUTER SYSTEMS INTEGRATION

Summary

The article is about the usage of embedded databases with native XML support in the process of computer systems integration. It concerns storing, processing, and interchanging semistructural data with an XML representation. The author approaches the problem from the perspective of interoperability in a heterogeneous distributed environment. The main goal of this paper is to discuss the possibilities of using presented technology in integration undertakings.