

Adam DUSZEŃKO  
Politechnika Śląska, Instytut Informatyki

## SYSTEMY ZARZĄDZAJĄCE BAZAMI DANYCH W ARCHITEKTURZE CHMURY OBLICZENIOWEJ

**Streszczenie.** Struktura systemów informatycznych o charakterze chmury obliczeniowej posiada specyfikę, która może zostać odpowiednio wykorzystana w budowie systemów bazodanowych. To, w jaki sposób budować bazy danych opierając się na chmurach obliczeniowych, które elementy chmur, w jaki sposób wspierają mechanizmy baz danych, które są dla nich zagrożeniem i w jakim zakresie jest przedmiotem tego artykułu. Na bazie przeprowadzonych analiz zostanie również zaproponowany rozszerzony model systemu zarządzania bazą danych przeznaczony w pełni do pracy w środowisku chmury obliczeniowej.

**Słowa kluczowe:** system zarządzający bazami danych, chmura obliczeniowa, wirtualizacja

## DATABASE MANAGEMENT SYSTEM IN CLOUD COMPUTING

**Summary.** The structure of information systems in cloud computing has the specificity that can suitably be used in the construction of database systems. This is how to build a database based on cloud computing, which means how the cloud database support, which is a threat to them and to what extent is the subject of this chapter. On the basis of the analyzes will also be proposed extended model of a database management system fully dedicated to work in a cloud computing environment.

**Keywords:** database management system, cloud computing, virtualization

### 1. Wstęp

Niezależnie od przyjętego modelu chmury obliczeniowej ukierunkowanego na infrastrukturę, platformę, aplikację czy pochodne, osadzenie usług bazodanowych w chmurze musi być uzasadnione. Elementami uzasadniającymi mogą być korzyści płynące z wykorzystania ele-

mentów chmury obliczeniowej dla zwiększenia wydajności, bezpieczeństwa lub uniwersalności systemu zarządzania bazą danych. Podstawowe mechanizmy chmur obliczeniowych, czyli wirtualizacja i automatyzacja konfiguracji są same w sobie elementami posiadającymi radykalny wpływ na wydajność systemów uruchomionych na chmurach. Zarządzanie wszystkimi zasobami włącznie z przestrzenią składowania na poziomie mechanizmów wirtualizacji jest codziennością i dobrą praktyką prowadzącą do radykalnie większej użycia całego systemu. Choć głównym celem wirtualizacji i automatyzacji zarządzania nią jest zwiększenie użycia zasobów na potrzeby wielu systemów działających wspólnie, to korzyści mogą być również zauważalne dla poszczególnych systemów, jak zostanie to przedstawione poniżej.

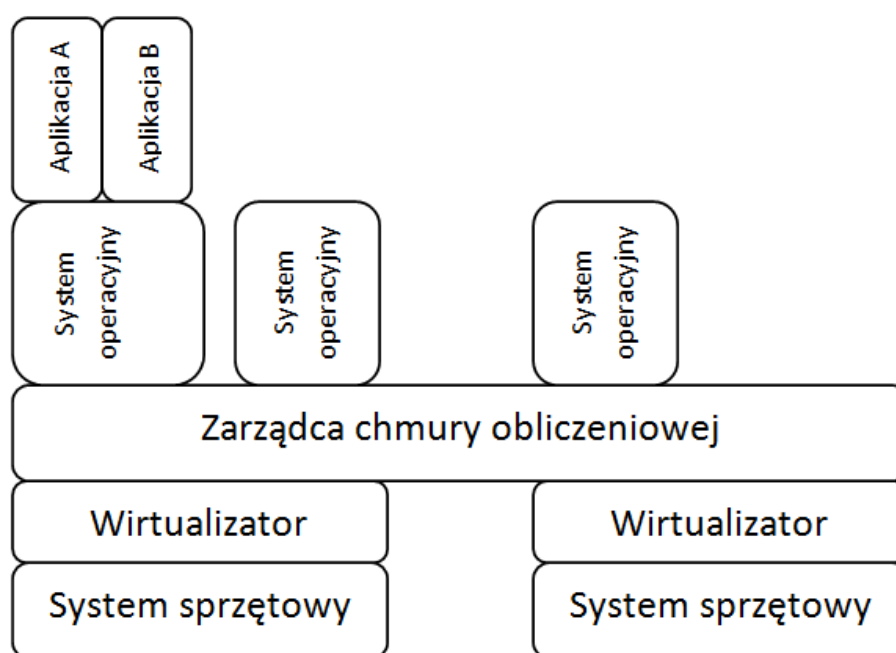
## 2. Architektura systemu IT o charakterze chmury obliczeniowej

Chmura obliczeniowa [1] jest dodatkową warstwą zarządzającą, umożliwiającą między innymi automatyzację procesów konfiguracyjnych, kontrolę wykorzystania poszczególnych zasobów, przenaszalność środowisk niezależnie od uwarunkowań sprzętowych. Zarządzanie jest w tym momencie możliwe dzięki wykorzystaniu mechanizmów wirtualizacyjnych, które niezależnie od charakteru zasobu czynią je podzielnymi i zarządzalnymi. W systemie komputerowym bez wirtualizacji wszystkie zasoby są zarządzane przez jedną jedyną instancję systemu operacyjnego zainstalowanego natywnie na systemie sprzętowym. Wszelkie zmiany zasobów, na jakich pracuje system, muszą być realizowane jako fizyczna zmiana podzespołów komputerowych, takich jak pamięć, procesor itp. Co więcej, wszystkie te zmiany muszą być realizowane dopiero po wstrzymaniu pracy systemu, a więc powodując przestoje.

Wprowadzenie warstwy wirtualizacji daje na poziomie danego systemu znacznie większą elastyczność, i co najważniejsze możliwość współbieżnego uruchomienia wielu niezależnych instancji systemów operacyjnych. W większości przypadków realokacja zasobów takich systemów wymaga wstrzymania ich pracy, choć są wyjątki w postaci systemów klasy i5 OS, AIX, PowerLinux które pozwalają na dynamiczną realokację zasobów, takich jak pamięć, procesor w trakcie pracy bez wyłączenia systemu.

Dodanie warstwy zarządczej realizującej chmurę obliczeniową [6], jak to przedstawiono na rysunku 1, rozszerza funkcjonalność płynącą z wirtualizacji pojedynczego systemu na wszystkie systemy wchodzące w skład chmury. W tym przypadku nie tylko realokacja zasobów, ale tworzenie klonów systemów w celu tworzenia na przykład klastrów wydajnościowych lub niezawodnościowych sprowadza się do wykonania kilku operacji administracyjnych. W tej konfiguracji systemy mogą być dynamicznie przenoszone pomiędzy fizycznymi systemami goszczącymi w miarę potrzeb. Ta funkcjonalność jest już bardzo ciekawa z punk-

tu widzenia baz danych. O ile sama wirtualizacja serwerów nie wnosi dla baz danych wiele korzyści, o tyle możliwość łatwego tworzenia duplikatów środowiska daje już szerokie możliwości budowania baz rozproszonych na poziomie chmury. Takie klastry bazodanowe mogą być tworzone dynamicznie dla potrzeb uzyskania większej wydajności lub dla zapewnienia większego poziomu bezpieczeństwa.



Rys. 1. Struktura chmury obliczeniowej  
Fig. 1. Cloud Computing structure

Jak wspomniano, wirtualizacja systemów komputerowych jest istotna, choć z punktu widzenia baz danych nie jest krytyczna. Inna sytuacja dotyczy krytycznego dla baz danych zasobu, jakim jest przestrzeń trwałego składowania, czyli przestrzeń dyskowa. W tym przypadku wprowadzenie wirtualizacji na poziomie zarządzania przestrzenią dyskową radykalnie wpłynęła na wydajność systemów bazodanowych. Wirtualizacja macierzy dyskowych, bo o tym mowa, zwiększając użycie dysków, znacząco poprawiła wydajność całego systemu składowania. Budowanie logicznych wolumenów dyskowych w poprzek dysków, a nie jak to było do tej pory w pionie, spowodowało znaczący wzrost wydajności operacji wejścia, wyjścia, co dla baz danych jest krytyczne.

### 3. Bazy danych a chmury obliczeniowe

Aktualnie rozwijane systemy zarządzające bazami danych nie posiadają funkcjonalności umożliwiającej bezpośrednią współpracę z chmurą obliczeniową. SZBD są jedynie aplikacjami pracującymi na zwirtualizowanych platformach opierając się na klasycznych syste-

mach operacyjnych. Chmura ma zdolność zarządzania takim systemem od góry, ale brak jest wbudowanych mechanizmów komunikacji od dołu, czyli od SZBD do zarządcy chmury. Taka komunikacja od dołu pozwoliłaby faktycznie SZBD współpracować z chmurą obliczeniową, przekazując jej polecenia wykonania zadań rekonfiguracji dla uzyskania założonych celów wydajnościowych lub bezpieczeństwa.

Dla zilustrowania możemy wyobrazić sobie przypadek, w którym klaster wydajnościowy bazy danych wykrywa, iż średni czas wykonania zapytań interaktywnych przekroczył zadany limit maksymalny. W tym momencie SZBD mógłby wysłać do zarządcy chmury komunikat z żądaniem uruchomienia kolejnego kłona węzła systemu bazodanowego w celu włączenia go w strukturę klastra. Taka funkcjonalność byłaby faktycznie bazą danych w chmurze.

### 3.1. Bezpieczeństwo informacji

Podstawowym poziomem zabezpieczenia systemu informatycznego jest poziom dostępu fizycznego. W większości przypadków bezpośredni dostęp do systemu daje możliwość uzyskania większości uprawnień do zawartości systemu. Dodatkowo, bezpośredni dostęp daje możliwość wykonania kopii nośników pamięci i późniejsze uzyskanie dostępu do informacji. W przypadku chmur obliczeniowych, jeśli nie jesteśmy zarządcą chmury, nie posiadamy fizycznego dostępu do systemu sprzętowego, co więcej posiadają ten dostęp osoby trzecie. Oczywiście dostawca chmury działa w trybie instytucji zaufania publicznego, gdzie jego staranność o zapewnienie poufności danych klientów jest kluczowym zadaniem. Mimo wszystko dla zapewnienia bezpieczeństwa danych, zwłaszcza w kontekście ich poufności, należy podjąć odpowiednie kroki zabezpieczające na wypadek nadużycia przez operatorów chmury lub innych użytkowników uprawnień wynikających z pełnego fizycznego dostępu do danych. Krokami tymi może być wprowadzenie szyfrowania, dywersyfikacja dostawcy lub ukrywanie informacji.

Szyfrowanie danych daje możliwość zabezpieczenia danych, pod warunkiem że proces odzyskiwania danych realizowany jest w innym środowisku niż system bazy danych. Realizacja szyfrowania na poziomie SZBD, w sytuacji gdy cała jego logika i konfiguracja jest poza naszą bezpośrednią kontrolą, nie daje wystarczającego poziomu zabezpieczeń. Dlatego szyfrowanie całości tabel nie jest w tym przypadku właściwym rozwiązaniem. Lepiej w tym momencie zastosować szyfrowanie pojedynczych wartości lub zwiększenie poziomu granulacji danych przez ich agregację na poziomie tworzenia modelu danych. W tym przypadku realizacja szyfrowania i deszyfrowania danych poza SZBD daje większe bezpieczeństwo, nawet w przypadku naruszenia bezpieczeństwa środowiska bazodanowego.

Dywersyfikacja dostawcy sprowadza się do podziału repozytorium danych na części umieszczane w kilku środowiskach chmury. Rzecz jasna rozwiązanie to posiada oczywiste

problemy wydajnościowe i nie może być używane w zastosowaniach o dużych wymaganiach wydajnościowych. Podział danych na części, z których żadna z osobna nie pozwala na odtworzenie pełnej informacji, umożliwia zabezpieczenie danych na wypadek naruszenia bezpieczeństwa u jednego z dostawców chmury.

Ukrywanie danych takimi technikami jak steganografia [2] podobnie jak szyfrowanie pozwala zabezpieczyć krytyczne informacje, nie ujawniając jednak miejsca ich zapisu. Realizowane musi być również poza SZBD, gdyż tylko wtedy daje odpowiedni poziom zabezpieczenia. Przykładem zastosowania tego typu technik może być ukrycie hasła użytkownika w zdjęciu profilowym. W tym przypadku aplikacja dokonująca autoryzacji użytkownika pobiera z bazy danych tylko identyfikator użytkownika i jego zdjęcie. Uzyskanie informacji o hasle jest operacją bazującą na danych zawartych w obrazie.

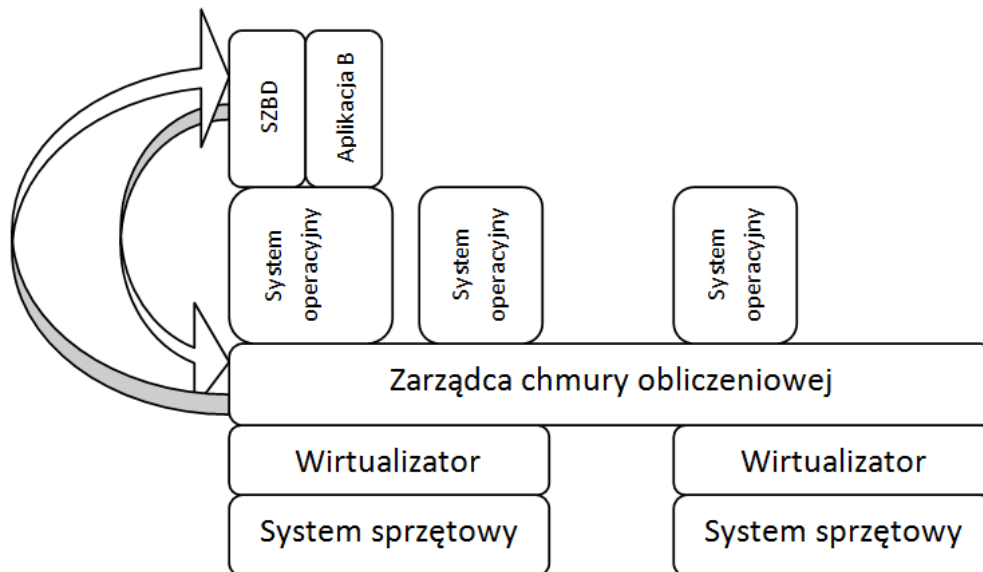
Oczywiście aspekt bezpieczeństwa należy tu widzieć zarówno w kryteriach zagrożenia, jak i zabezpieczenia. W sytuacji gdy większość przypadków wykradzenia danych związana jest z działaniami pracowników wewnętrznych, posiadanie systemów baz danych poza siedzibą organizacji zwiększa bezpieczeństwo danych. Dodatkowo dostawcy chmur obliczeniowych, dbając o duże bezpieczeństwo, oferują wysokiej jakości usługi wykrywania intruzów i aktywności niepożądanych. Sumarycznie zastosowanie chmur obliczeniowych oferuje więcej korzyści niż zagrożeń.

Bezpieczeństwo informacji w sensie trwałości danych jest bezspornie większe w przypadku zastosowania chmur obliczeniowych. Na stan taki wpływają przede wszystkim zaawansowane mechanizmy zarządzania przestrzenią składowania niezbędne dla chmur obliczeniowych oraz wspomniana wyżej łatwość tworzenia struktur redundantnych o charakterze klastrów bezpieczeństwa.

#### **4. Rozszerzony model systemu zarządzającego bazą danych dedykowany środowisku chmur obliczeniowych**

Podstawowym założeniem wirtualizacji jest jej przeźroczystość dla systemów goszczonych. Innymi słowy, system goszczony nie musi posiadać informacji, iż działa na podstawie zwirtualizowanych zasobów. Dla zwiększenia wydajności można jednak dodać do systemu aplikacje, które bazując na specyfice wirtualizatora mogą przyspieszać działanie całego systemu. Aplikacje tym bardziej nie muszą dostosowywać swojego działania do faktu uruchomienia ich w środowisku o charakterze chmury obliczeniowej. Systemy zarządzania bazami danych są aplikacjami na tyle angażującymi zasoby, że łatwo wyobrazić sobie sytuację, w której elastyczność środowiska chmury może zostać zaangażowana do dostosowania zwirtualizowanego środowiska do potrzeb baz danych. Sytuacja taka prowadzi do stworzenia

kanału informacyjnego od zarządcy wirtualizacji do SZBD i kanału sterującego od SZBD do zarządcy chmury, jak to przedstawiono na rysunku 2.



Rys. 2. Sprzężenie aplikacji i zarządcy chmury  
Fig. 2. Application feedback to cloud manager

Proponowane sprzężenie celowo realizowane jest z pominięciem systemu operacyjnego, gdyż to właśnie zarządca chmury ma możliwość interakcji z systemem operacyjnym, choćby na poziomie generowania przerw i wysyłania do systemu odpowiednich sygnałów umożliwiających jego restart, hibernację i inne czynności zarządcze. Kanał sterowania może w tym momencie służyć do kontroli pracy całego środowiska, dynamicznej zmiany jego konfiguracji itp.

Zdolność zamawiania od zarządcy chmury zasobów na żądanie, zgodnie z możliwościami środowiska i zapotrzebowaniem SZBD, jest jednym bardzo ważnym aspektem posadowienia środowiska bazy danych w chmurze obliczeniowej. Drugim bardzo ważnym, jeśli nie ważniejszym, elementem jest możliwość stworzenia na bazie chmury obliczeniowej heterogenicznego środowiska bazodanowego o zróżnicowanej wewnętrznej organizacji opartej na różnych modelach danych. Hybrydowe środowiska bazodanowe, takie jak na przykład IBM DB2 [3] łączące implementacje modelu relacyjnego i hierarchicznego cieszą się dużą popularnością dzięki swojej uniwersalności. Budując środowisko bazy danych w chmurze, można połączyć w jeden system wiele odrębnie działających systemów baz danych. Celem takiego działania byłoby przygotowanie przestrzeni przetwarzania danych, oferującej mechanizmy dedykowane różnym kategoriom danych. Dla przykładu, integrując środowisko baz danych relacyjnych, NoSQL i hierarchicznych dedykowanych językowi XML możemy stworzyć uniwersalne rozwiązanie oferujące wydajne mechanizmy zarządzania zarówno danymi transakcyjnymi, dużymi danymi nieustrukturyzowanymi, jak i dokumentami XML. Odpowiednia warstwa zarządzająca odpowiedzialna byłaby za umieszczanie danych we właściwych im

modułach środowiska. Modułami środowiska byłyby znane i aktualnie rozwijane środowiska baz danych. Elementem integrującym wszystkie moduły byłby element typu IBM DB2 Connect [3] oferujący możliwość tworzenia sfederowanych środowisk bazodanowych [7]. Środowiska sfederowane mimo swojej heterogenicznej struktury logicznie prezentują się jako jednolite i zintegrowane środowisko. Zastosowanie mechanizmów chmury może zapewnić odpowiedni poziom wydajności komunikacji pomiędzy poszczególnymi modułami, aby mechanizmy federacji działały wystarczająco wydajnie. Dodatkowo dzięki mechanizmom chmury obliczeniowej poszczególne moduły lub ich klony mogłyby być uruchamiane i zatrzymywane na żądanie wedle aktualnego zapotrzebowania środowiska bazodanowego. W tym podejściu tworzymy uniwersalną usługę zarządzania danymi realizującą wewnętrzne procesy oparte na dedykowanym typom danych implementującym systemy bazodanowe. Podobnie działającym w aspekcie dynamiczności zmian struktury jest środowisko IBM PureScale [4]. Znajdujące się w tym rozwiązaniu węzły zarządzające pozwalają na dynamiczną zmianę węzłów roboczych, usuwanie ich w przypadku awarii i zastępowanie lub uzupełnianie nowymi. W tym rozwiązaniu wszystkie węzły są instancjami bazy relacyjnej, w proponowanej strukturze poszczególne węzły mogą być instancjami różnych środowisk w różnych modelach danych.

W praktyce do tej pory zrealizowane zostały próby tworzenia środowiska rozproszonej bazy danych na podstawie zwirtualizowanych zasobów sprzętowych. Integracja wykonana była na poziomie mechanizmów federacji systemu IBM DB2. Element rekonfiguracji środowiska realizowany był za pomocą skryptów, które w przyszłości zastąpione zostaną narzędziem samodzielnie modyfikującym konfigurację modułu federacji. W zakresie rozwiązań wysokiej wydajności tworzone były klastry niezawodności na poziomie mechanizmów IBM DB2 HADR [8] wykorzystując maszyny wirtualne jako odpowiedniki klastrowanych serwerów. Wspomniane próby zakończyły się pomyślnie i pozwoliły wyciągnąć pozytywne wnioski co do możliwości automatycznej realizacji mechanizmów zarządzania środowiskiem systemu bazy danych w środowisku chmury obliczeniowej. Planowane są jako kontynuacja stworzenia instalacji bazującej na środowisku IBM PureScale do sprawdzenia możliwości dynamicznego rekonfigurowania wielowęzłowej struktury systemu bazy danych rozproszonej oraz stworzenia własnego oprogramowania zarządzającego chmurą systemów baz danych w zakresie ich konfiguracji początkowej i rekonfiguracji w trakcie działania.

## 5. Podsumowanie

Zastosowanie chmur obliczeniowych do budowy środowisk baz danych stwarza pewne wyzwania związane głównie z zapewnieniem poufności danych, jednak w głównej mierze

znacząco upraszcza tworzenie rozbudowanych wielowęzłowych systemów. Pozwala automatyzować proces zarządzania klastrami bazodanowymi, upraszcza tworzenie środowisk zapasowych, zmniejsza czas reakcji na sytuacje awaryjne. Rozbudowanie klasycznego klastra bazodanowego o warstwę zarządzającą daje możliwość tworzenia systemów heterogenicznych wielomodułowych. Jest to nowy rodzaj rozproszonej bazy danych zdolnej do realokacji, rekonfiguracji, rozbudowy lub redukcji struktury na żądanie, zgodnie ze zmieniającymi się wymaganiami wydajnościowymi. Umieszczenie baz danych w chmurze daje szerokie korzyści, wymaga jednak rozbudowy, aby w pełni skorzystać z funkcjonalności charakterystycznej dla chmur obliczeniowych.

## BIBLIOGRAFIA

1. Mell P., Grance T.: The NIST Definition of Cloud Computing. Special Publication 800-145 National Institute of Standards and Technology, 2011.
2. Umiński W.: Steganografia – nowoczesny sposób ukrywania informacji. Wydawnictwo Politechniki Częstochowskiej, 2005.
3. Dokumentacja produktu IBM DB2, <http://www-01.ibm.com/software/data/db2/linux-unix-windows/>.
4. Ebbers M., Schulz M.: Highly Available and Scalable Systems with IBM eX5 and DB2 pureScale. IBM Redbooks, 2011.
5. Bienko C. D., Greenstein M., Holt S. E., Phillips R. T.: IBM Cloudant Database as a Service Advanced Topics. IBM Redbooks, 2015.
6. Rosenberg J., Mateos A.: Chmura obliczeniowa. Rozwiązania dla biznesu, Gliwice 2011.
7. Zhu W.-D., Bacalzo R., Edeen E., Jun Y., Ouimet D., Schmitt J. D., Wang B., Wersin D., Werts D. B., Willingham M.: Federated Content Management Accessing Content from Disparate Repositories with IBM Content Federation Services and IBM Content Integrator. IBM Redbooks, 2010.
8. Bartkowski S., De Buitlear C., Kalicki A., Loster M., Marczewski M., Mosaad A., Nelken J., Soliman M., Subtil K., Vrhovnik M., Zimmol K.: High Availability and Disaster Recovery Options for DB2 for Linux, UNIX, and Windows. IBM Redbooks, 2012.



**Abstract**

The usage of cloud computing to build a database environments pose some challenges related mainly to confidentiality of the data, but also significantly simplifies the creation of complex multi-node systems. The automation of the process of database-cluster management is also simplified just like creation of backup environment, reduction of the response time to emergencies. Expanding the database of the classic cluster management layer provides the ability to create heterogeneous multi-module systems. It is a new type of distributed database capable of reallocation, reconfiguration, expansion or reduction of on-demand structure according to the changing performance requirements. Running databases in the cloud provides broad benefits, however, requires expansion to take advantage of the functionality characteristic of cloud computing.

**Adres**

Adam DUSZEŃKO: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, adam.duszenko@polsl.pl.