

Model systemu klasy Business Intelligence w chmurze obliczeniowej

Patryk Dzięgielewski*, Dariusz Pałka**

Warszawska Wyższa Szkoła Informatyki

Abstrakt

W procesach zarządzania organizacjami gospodarczymi wymagane jest podejmowanie jak najlepszych decyzji, które pozwolą na zwiększenie konkurencyjności firmy oraz optymalizację jej działalności. Pomocne są w tym systemy informatyczne takie jak Business Intelligence. Jeszcze więcej można osiągnąć, gdy te systemy oparte są o technologie chmurowe. Artykuł przybliży tematykę systemów BI, technologii chmurowych oraz wskazuje różnice wydajnościowe między systemem opartym o serwer fizyczny, a środowiskiem usług analitycznych opartych o technologie chmurowe.

Słowa kluczowe – BI, Business Intelligence, technologia chmurowa, Azure, ERP

* E-mail: p_dziegielewski@wwsi365.edu.pl

** E-mail: dpalka@poczta.wwsi.edu.pl

1 Wprowadzenie

Termin „Business Intelligence” pierwszy raz został użyty w roku 1958 [1], jednak dopiero w roku 1989 określono, że BI to szeroka dziedzina, która jest ukierunkowana na realne wspieranie podejmowania decyzji biznesowych. Same systemy BI oferują przeprowadzenie wielopoziomowych analiz, monitorowanie bieżącej sytuacji danej organizacji czy odkrywanie nowej wiedzy.

Systemy informatyczne BI często połączone są z systemami ERP (*Enterprise Resource Planning*). System ERP to rozwiązanie, które pozwala na integrację wszystkich modułów firmy, a więc produkcji, księgowości czy magazynów. Integracja polega na wykorzystywaniu jednej bazy danych dzięki czemu unika się trudności w realizacji pewnych procesów czy minimalizuje ryzyko utraty kontroli nad tym, co dzieje się w organizacji. Takie systemy pozwalają na udoskonalanie działań prowadzonych w firmie czy wyciąganie wniosków.

Samą Genezę systemów BI można określić przez ich generację:

- BI 1.0
- BI 2.0
- BI 3.0

BI 1.0 odnosi się ery, która istniała pod koniec lat 90. zeszłego wieku. Wraz z pojawieniem się i rozwojem hurtowni danych, SQL (*Structured Query Language*) czy OLAP (*Online Analytical Processing*), dane zostały zunifikowane i skonsolidowane w jeden system, co pozwalało na ich wyodrębnienie z wielu tabel jednocześnie, ostatecznie pomagając przedsiębiorstwom uzyskać bardziej efektywny dostęp do danych i ich przechowywania. W BI 1.0 można wyróżnić dwa podstawowe składniki jakimi są dane i raporty.

BI 2.0 można określić na koniec pierwszego dziesięciolecia XXI wieku. Wtedy nastąpił znaczący krok dla BI, ponieważ zaczęła ona wykraczać poza proste dane i raportowanie, integrując przetwarzanie w czasie rzeczywistym, współpracę, samoobsługę, a także dostęp do danych offline i online. Podczas gdy BI 1.0 skupiał się na udoskonalaniu różnych narzędzi (takich jak hurtownie danych czy technologii OLAP), BI 2.0 skupiał się na wykorzystaniu łączności internetowej do stworzenia środowiska.

BI 3.0 rozwinął się głównie pod względem prezentacji danych dla analityków. Wizualizacje, interaktywne pulpity czy animowane wykresy słupkowe mają pomóc skutecznie analizować informacje pochodzące z zewnątrz i z wewnątrz. Również rozwój urządzeń mobilnych takich jak smartfony czy tablety stał się nowym wyzwaniem: duży nacisk kładzie się na dostęp do danych z każdego urządzenia [2].

2 Analiza dostępnych rozwiązań

Cloud computing (ang.) czyli chmura obliczeniowa, to usługa oparta na przeniesieniu wszystkich usług IT, takich jak dane, moc obliczeniowa czy samo oprogramowanie na serwer i świadczenie możliwości uzyskania do nich dostępu w każdej chwili przez komputer kliencki [3].

Wyróżnia się różne rodzaje chmur:

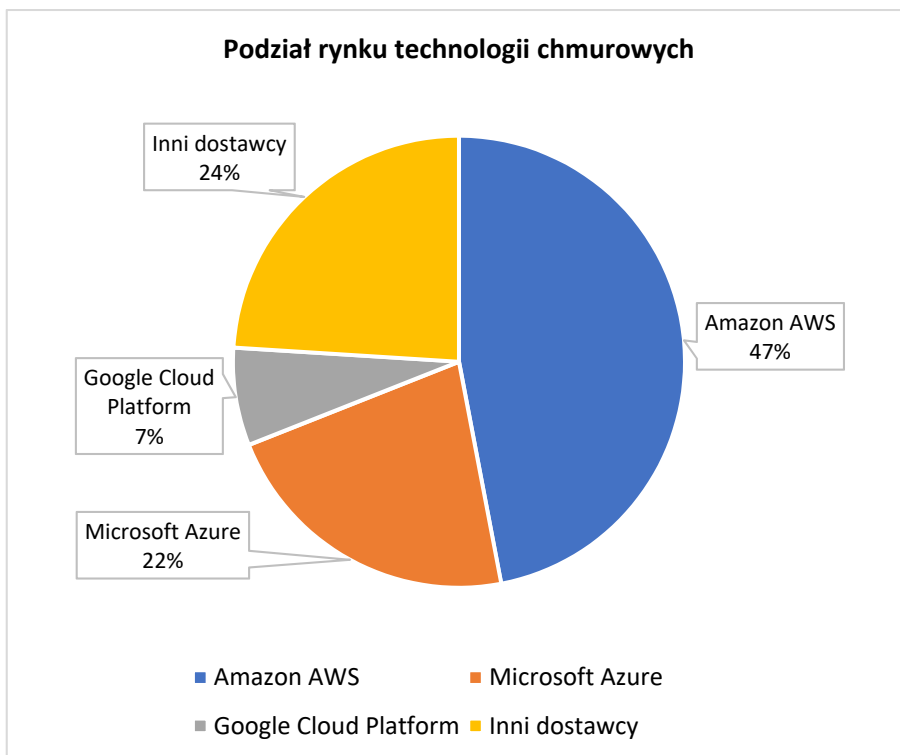
- Prywatne – są to wewnętrzne serwery danego przedsiębiorstwa, które jest dla siebie jednocześnie dostawcą usługi chmurowej.
- Publiczne – najbardziej popularne rozwiązanie, gdzie usługa jest zamawiana w zewnętrznej firmie np. Amazon Web Services, Google Cloud Platform czy Microsoft Azure.
- Hybrydowe – są to połączone rozwiązania chmury prywatnej i publicznej, gdzie pewna część aplikacji pracuje na usłudze wewnętrznej, a część umiejscowiona jest w chmurze publicznej.

Aktualnie coraz częściej przedsiębiorstwa decydują się na umiejscowienie swoich danych w modelu chmury obliczeniowych.

Do porównania wybrane zostały trzy najbardziej popularne rozwiązania chmurowe:

- Amazon AWS
- Microsoft Azure
- Google Cloud

Poniższy wykres pokazuje aktualny podział rynku technologii chmurowych na świecie.



Rysunek 1. Podział rynku technologii chmurowych [4].

Jak widać, Amazon ze swoją technologią AWS kontroluje ponad 40% światowego rynku chmur publicznych, co stawia go na pozycji lidera. Znacznie ustępuje mu Microsoft ze swoimi 22%, a dosyć daleko w tyle jest Google (na podstawie danych od firmy Gartner).

Swoją przewagę Amazon wypracował dzięki wcześniejszemu pojawieniu się na rynku oraz stabilności swoich usług, które zależne są jednak od światłowodowych połączeń internetowych. W krajach mniej rozwiniętych, gdzie częściej spotyka się prywatne rozwiązania dostępu do Internetu, wydajność Amazona wyraźnie spada, przez co chętniej wykorzystywane są rozwiązania od pozostałych dostawców. Każda z tych platform posiada swoje wady i zalety, co uwidoczniło w tabeli 1.

Tabela 1. Porównanie mocnych i słabych stron platform [5]

Dostawca	Mocne strony	Słabe strony
Amazon AWS	Bezsprzeczny lider na rynku Bardzo dużo usług Bardzo duże doświadczenie Wsparcie dla dużych firm Obszerne materiały szkoleniowe Globalny zasięg	Trudny w użytkowaniu Zawile zarządzanie kosztami Przytłaczająca ilość opcji
Microsoft Azure	Drugi największy dostawca na rynku Integracja ze wszystkimi dotychczasowymi rozwiązaniami Microsoftu Szeroki zestaw funkcji Chmura hybrydowa Wsparcie dla rozwiązań Open Source	Problemy z dokumentacją Niekompletne narzędzia do zarządzania
Google Cloud Platform	Zaprojektowana dla firm działających w chmurze Przyjazna dla rozwiązań Open Source i mobilności Duże zniżki i elastyczne umowy Ekspertyza dla developerów	Jako ostatni wszedł na rynek IaaS Mniej funkcji i usług niż konkurencja Początkowo rozwiązanie nie kierowane do zewnętrznych odbiorców

Na podstawie przeprowadzonego porównania, do badania wybrane zostało rozwiązanie Microsoft Azure.

System ERP do realizacji badania wybrany został z pomiędzy dwóch najpopularniejszych w Polsce rozwiązań:

- Comarch ERP XL
- SAP

Pierwszy z nich to system stworzony przez polską firmę Comarch, zawiera w sobie kompletne moduły do zarządzania finansami, księgowością, sprzedażą, dystrybucją, kadrami, płacami, produkcją, procesami, obiegami dokumentów, analizą dokumentów oraz BI. Jest to zdecydowanie jeden z najchętniej wybieranych systemów ERP w Polsce.

Drugi system, który rozważano pochodzi od niemieckiej firmy SAP SE, działającej w Polsce od roku 1995. Jest to czołowy dostawca aplikacji na rynku biznesowym. Liczba klientów od rozpoczęcia działalności na terenie Polski wyniosła już

ponad 1.5 tysiąca. Zdecydowano, że do badania zostanie wybrana zintegrowana aplikacja BI w systemie Comarch.

3 Weryfikacja założeń

Do przeprowadzenia założonych badań utworzona została maszyna wirtualna modelu wykorzystującego technologie chmurową w specyfikacji jak najbardziej zbliżonej do tej występującej w modelu klasycznym (tabela 2).

Tabela 2. Specyfikacja serwerów

Specyfikacja	Podzespoły serwera fizycznego	Podzespoły serwera Azure
Procesor	1x Intel Xeon C4 2.6 GHz	1x Intel Xeon E5 v4 2.6 GHz
RAM	16 GB	16 GB
Dysk	2 x SAS 15k rpm	1
Pojemność dysków	300 GB	300 GB

Wykorzystana do badań baza danych oparta na Microsoft SQL Server, jest bazą prezentacyjną Comarch, zawierającą przykładowych kontrahentów oraz towary. Do bazy zostały zaimportowane losowo wygenerowane zamówienia przez program napisany dla celów niniejszych badań. Zamówienia zostały oparte na danych znajdujących się w bazie prezentacyjnej czyli towarach i kontrahentach.

Następnie wybrane zostały trzy raporty dostarczone przez producenta oprogramowania Comarch XL Business Intelligence. Są to:

- Trend sprzedaży – porównanie marży i przychodu – jako raport nr 1.
- Liczba sprzedanych produktów w zależności od województwa – jako raport nr 2.
- Ranking klientów ze względu na wielkość oraz wartość złożonych zamówień – jako raport nr 3.

Każdy z tych raportów został generowany dziesięć razy, po czym czasy długości wywołania raportu zostały uśrednione, co pozwoliło najlepiej zobrazować czas potrzebny na wykonanie danego raportu na każdym z serwerów. Czasy wykonywania

raportów mierzone były za pomocą wbudowanego w Comarch XL BI narzędzia pozwalającego określić, jak długo trwała realizacja polecenia, aż do wyświetlenia raportu.

3.1 Badanie nr 1.

Tabela 3 pokazuje czasy potrzebne na wygenerowanie raportu na każdym z serwerów.

Tabela 3. Czasy uzyskane na serwerach przy wywołaniu raportu nr 1.

Lp.	Microsoft Azure [s]	Serwer fizyczny [s]
1	8,5	8,9
2	8,2	8,2
3	7,9	6,7
4	7,9	6,8
5	7,9	6,5
6	7,9	7,0
7	7,8	7,1
8	8,0	7,0
9	8,4	6,7
10	8,0	6,9

Model klasyczny znajdujący się na serwerze fizycznym radzi sobie zdecydowanie lepiej. Jak widać, pierwszy czas jest najdłuższy, ponieważ raport wywoływany był pierwszy raz. Kolejne czasy skracały się, osiągając swój optymalny wymiar. Nie są to bardzo duże różnice, jednak jak widać po wynikach zamieszczonych w tabeli 3, praca serwera fizycznego jest bardziej stabilna, niż serwera w technologii chmurowej dla tego raportu.

Podsumowując, z raportem nr 1 lepiej poradził sobie model oparty na serwerze fizycznym. Czasy uzyskiwane w modelu klasycznym są średnio niższe o prawie sekundę (0,87 sekundy). Nie jest to duża różnica, ale warta odnotowania, jeśli skala częstości wywoływania takiego raportu jest duża.

3.2 Badanie nr 2.

Badanie nr 2 zostało przeprowadzone w ten sam sposób co badanie poprzednie. Tym razem raport zawiera trochę więcej danych dlatego jego przetwarzanie powinno być dłuższe (tabela 4).

Tabela 4. Czasy uzyskane na serwerach przy wywołaniu raportu nr 2.

Lp.	Microsoft Azure [s]	Serwer fizyczny [s]
1	8,7	12,4
2	7,5	11,7
3	8,3	9,4
4	11,9	9,6
5	7,7	9,5
6	7,7	10,3
7	7,8	11,2
8	8,1	10,5
9	7,8	10,1
10	17,4	9,9

Czasy serwera Azure mimo bardziej rozbudowanego raportu nie były bardzo wysokie. System dosyć szybko rysował potrzebny raport, tylko jedno wywołanie trwało zdecydowanie dłużej. Suma czasu potrzebna na wykonanie wszystkich raportów wyniosła 92,90 sekundy, co daje wynik prawie 10 sekund dłuższy niż raportu nr 1.

Wyniki na serwerze fizycznym oscylowały w podobnych zakresach jak na serwerze Azure, jednak tym razem były nieznacznie wyższe. Początkowy czas wywołania raportu był dosyć wysoki, jednak potem kolejne wywołania pokazywały się już w podobnych odstępach.

Podsumowując, z raportem nr 2 lepiej poradził sobie model oparty na serwerze Azure, jednak tutaj różnica jest większa niż w poprzednim przypadku. Czasy uzyskiwane w modelu klasycznym są średnio wyższe o około 1,17 sekundy.

3.3 Badanie nr 3.

Trzeci raport również zawierał dużo informacji (tabela 5), jednak wymagał dużo mniej obliczeń niż raport nr 1.

Tabela 5. Czasy uzyskane na serwerach przy wywołaniu raportu nr 3.

Lp.	Microsoft Azure [s]	Serwer fizyczny [s]
1	15,6	16,2
2	7,8	9,1
3	7,7	8,7
4	8,0	8,2
5	7,9	12,2
6	7,9	9,5
7	7,9	8,4
8	7,9	8,6
9	8,0	8,3
10	7,8	8,1

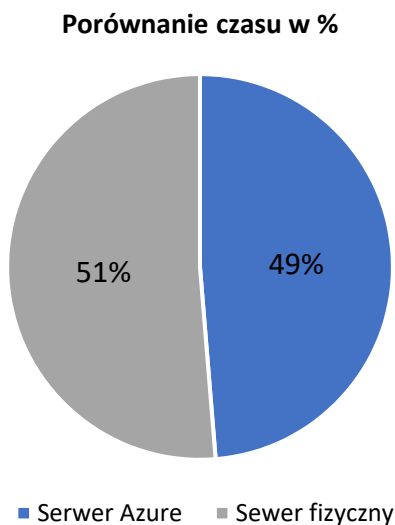
Czasy uzyskane na serwerze Azure dla raportu nr 3. były zadowalające. Nie były one przesadnie długie, dzięki czemu użytkownik nie powinien czuć dyskomfortu przy użytkowaniu całego systemu. Czasy średnio oscylowały w zakresie 8 sekund, z zaznaczeniem wyjątku przy pierwszym wywołaniu, gdzie czas był znacznie dłuższy. Suma czasu potrzebna na wykonanie wszystkich raportów wyniosła 86 sekund, co plasuje ten wynik pośrodku stawki w czasie potrzebnym na wykonanie.

Serwer fizyczny poradził sobie trochę gorzej. Pierwsze wywołanie, podobnie jak w przypadku serwera Microsoftu, trwało trochę dłużej, jednak potem czasy również się znormalizowały i wynosiły około 9,73 sekundy. Ciekawe jest to, że tym razem również niższe czasy miał serwer Azure, różnica w tym przypadku wynosiła 1,08 sekundy. Spowodowane może być to przez to, że Azure lepiej radził sobie podczas rysowania raportu w aplikacji Comarch BI.

Podsumowując, z raportem nr 3 należy stwierdzić, że model oparty na serwerze Azure działał efektywniej, jednakże różnica była tu jednak nieznacznie mniejsza niż w badaniu nr 2.

4 Podsumowanie badań

Jak widać na rysunku nr 2, oba serwery prezentują się prawie identycznie, jeśli chodzi o wydajność działania z systemem Comarch BI. Serwer Azure potrzebował jednak minimalnie mniej czasu na wykonanie postawionych mu zadań. Serwer Microsoftu na wykonanie wszystkich zadań potrzebował 259,9 sekund, zaś serwer w modelu klasycznym potrzebował ich aż 273,7. Z przeprowadzonych trzech badań tylko w jednym przypadku serwer fizyczny poradził sobie lepiej.



Rysunek 2. Wykres porównania czasu potrzebnego na wykonanie wszystkich raportów na obu serwerach

W trakcie badania zauważono, że z raportami, które wymagają rysowania bardziej skomplikowanych wykresów, lepiej radzi sobie serwer w modelu wykorzystującym technologie chmurową. Oznacza to, że wersja procesora znajdująca się na serwerze

fizycznym słabiej przetwarzała generowanie danych graficznych, co widać było na wykresach zużycia zasobów w badaniu nr 2 i 3, gdzie ich wykorzystanie było znacznie wyższe niż w przypadku badania nr 1.

5 Wnioski

Technologia chmurowa oraz oparty na niej system ERP z modułem BI to rozwiązanie, które zapewne coraz częściej będzie stosowane przez duże firmy. Sami producenci widzą w nim wielki potencjał, prezentując swoje systemy oparte na technologiach chmurowych z dostępem z każdego miejsca i na każdym urządzeniu.

Potrąfi to ułatwić pracę dla dużej liczbie osób, oszczędza czas potrzebny na konfigurację stanowisk i obniża w znaczny sposób koszty prowadzenia działalności. Jest bardzo możliwe, że w przyszłości rozwiązania oparte na technologiach chmurowych wyprą systemy budowane na indywidualnych serwerach fizycznych firmy.

Bibliografia

- [1] H.P. Luhn, *A Business Intelligence System*, “IBM Journal of Research and Development” Vol. 2, No. 4, pp. 314-319, October 1958.
- [2] C. Olszak, *Business Intelligence for informataion socjety. La société de l'information: perspective européenne et globale: la société de l'information et ses enjeux pour les entreprises*, Katowice, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2013.
- [3] A. Mateos, *Chmura obliczeniowa. Rozwiązania dla biznesu*, Warszawa, Helion, 2012.
- [4] K. Stalcup, *AWS vs Azure vs Google Cloud Market Share 2019: What the Latest Data Shows*, ParkMyCloud, <https://www.parkmycloud.com/blog/aws-vs-azure-vs-google-cloud-market-share/>, April 2019 [Online].
- [5] C. Harvey, A. Patrizio, *AWS vs. Azure vs. Google: Cloud Comparison [2019 Update]*, Datamation, <https://www.datamation.com/cloud-computing/aws-vs-azure-vs-google-cloud-comparison.html>, January 2019 [Online].

Model of the Business Intelligence Information System in the Cloud Computing Environment

Abstract

In the management processes of business organizations, it is necessary to make the best decisions that will increase the company's competitiveness and optimize its operations. IT systems such as Business Intelligence are helpful in this. Even more can be achieved when these systems are based on cloud technologies. The article introduces the topic of BI systems and cloud technologies and indicates differences in performance between a system based on a physical server and the environment of analytical services based on cloud technologies.

Keywords: *BI, Bussines Intelligence, Cloud Computing, Azure, ERP*