

Analiza porównawcza wydajności baz danych pracujących pod kontrolą systemu Windows

Serhii Stets*, Grzegorz Kozieł

Politechnika Lubelska, Instytut Informatyki, Nadbystrzycka 36B, 20-618 Lublin, Polska

Streszczenie. Celem niniejszego artykułu jest przeprowadzenie badań najpopularniejszych systemów zarządzania bazami danych takich jak MySQL, PostgreSQL oraz Firebird pod względem szybkości wykonania zapytań, obciążenia procesora oraz pamięci na dysku. Na pierwszym etapie przeprowadzono analizę popularności baz danych. W drugim etapie została napisana aplikacja umożliwiająca komunikację z wybranymi bazami danych. Na trzecim etapie zostały przeprowadzone badania za pomocą opracowanej aplikacji oraz przeanalizowano uzyskane wyniki.

Słowa kluczowe: analiza; MySQL; postgresql; firebird.

*Autor do korespondencji.

Adres e-mail: kompanybbc@gmail.com

Comparative analysis of databases working under the control of Windows system

Serhii Stets*, Grzegorz Kozieł

Institute of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract. The aim of this paper is to conduct research on the most popular database management systems such as MySQL, PostgreSQL and Firebird in terms of query performance, CPU load and disk usage. At the first stage, the analysis of the popularity of databases was carried out. In the second stage an application was created that allows for communication with selected databases. At the third stage, tests were carried out using a written application and analysis of results.

Keywords: analysis; MySQL; postgresql; firebird.

*Corresponding author.

E-mail address: kompanybbc@gmail.com

1. Wstęp

W dzisiejszych czasach niemożliwe jest sprawne działanie bez skutecznego zarządzania danymi. Ważną kategorią są systemy przetwarzania oraz gromadzenia informacji, od których zależy wydajność każdego przedsiębiorstwa lub instytucji [1]. Taki system powinien zapewniać możliwość otrzymywania ogólnych i/lub szczegółowych raportów wyników pracy, umożliwiać łatwiejszą identyfikację trendów kluczowych wskaźników, zapewniać możliwość odbioru ważnej informacji bez znacznych opóźnień oraz wykonywać dokładną i kompletną analizę danych [2].

DBMS (ang. Database Management System) – zorganizowany zestaw oprogramowania i danych językowych, które obejmują wykresy, tabele, zapytania, raporty, widoki i inne. Zasadniczo DBMS to program, który pozwala na interakcję z bazami danych jako obiektem abstrakcyjnym (bez konieczności pisania zapytań) [3].

Obecnie znaczna część aplikacji korzysta z baz danych. Zakres stosowań baz danych jest dość szeroki: od aplikacji dla pojedynczego użytkownika, takich jak notatniki, po duże rozproszone systemy informacyjne, takie jak wyszukiwarki, sklepy internetowe, systemy bankowości elektronicznej [4].

W 1968 r. uruchomiono pierwszy przemysłowy DBMS - system IBM IMS (Information Management System). Mimo że IMS jest pierwszym ze wszystkich komercyjnych systemów zarządzania bazami danych, nadal pozostaje głównym hierarchicznym systemem zarządzania bazami danych używanym w większości komputerów głównego szeregu [7]. Pierwszy etap rozwoju DBMS wiąże się z organizacją bazy danych na dużych maszynach, takich jak IBM360/370. Bazy danych są zapisane w pamięci zewnętrznej komputera centralnego. Programy dostępu do baz danych zostały napisane w różnych językach. Interaktywny dostęp zapewniono za pomocą terminali, które nie miały własnych zasobów obliczeniowych i służyły jedynie jako urządzenia wejścia/wyjścia dla komputera centralnego.

W drugim etapie, wraz z pojawieniem się komputerów osobistych, zaczęły się rozwijać desktopowe wersje DBMS. Większość z nich miała przyjazny interfejs użytkownika, oferowała interaktywny tryb pracy z bazą danych, zarówno do tworzenia bazy danych, jak i do projektowania zapytań [5].

Trzeci etap rozwoju DBMS wiąże się z rozległym rozwojem sieci lokalnych. Praca na izolowanym komputerze z małą bazą danych staje się obecnie nietypowa dla większości aplikacji. Komputery są połączone w sieć a baza danych jest jednocześnie dostępna dla wielu użytkowników [5].

Jednym z podstawowych standardów w dziedzinie oceny DBMS są standardy opracowane przez zarząd TPC.

TPC (ang. Transaction Processing Performance Council) jest korporacją założoną w celu przeprowadzenia testów porównawczych baz danych oraz rozpowszechniania obiektywnych, weryfikowalnych danych dotyczących wydajności DBMS w branży. TPC został stworzony w celu obserwacji nie tylko poprawności testów, ale także sposobu interpretacji wyników tych testów. TPC została założona w 1988 roku przez wiodących producentów w celu obserwacji nie tylko poprawności testów, ale także sposobu interpretacji wyników tych testów, wśród których byli IBM, HP, Control Data. Każda firma może zostać członkiem TPC [8].

Testy TPC opierają się na koncepcji „transakcji”, która tradycyjnie jest powiązana z relacyjnymi bazami danych, ale ma bardziej ogólne znaczenie dla zastosowań komercyjnych. Transakcja w aplikacjach biznesowych oznacza wymianę informacji o towarach, przekazywanie zleceń płatniczych czy wymianę różnego rodzaju usług. TPC definiuje i zarządza kilkoma formatami testowymi do oceny wydajności OLTP (On-Line Transaction Processing), w tym testów TPC-A, TPC-B, TPC-C, TPC-D, TPC-E i TPC-H. Jak już wspomniano, za stworzenie testu oceny odpowiedzialna jest organizacja przeprowadzająca ten test. Zespół TPC wymaga tylko spełnienia pewnych warunków. Chociaż wspomniane testy TPC nie są testami do bezpośredniej oceny wydajności bazy danych, relacyjne systemy baz danych są kluczowymi składnikami każdego systemu przetwarzania transakcji [8].

Uruchomiony w listopadzie 1989 r. Test TCP-A został zaprojektowany w celu oceny wydajności systemów działających w środowisku intensywnie aktualizowanych baz danych, typowych dla aplikacji do interaktywnego przetwarzania danych. Test TPC-A określa przepustowość systemu, mierzoną liczbą transakcji na sekundę, które system może wykonać podczas pracy z wieloma terminalami.

Pakiet testowy TPC-C pod względem rzeczywistych potrzeb użytkowników zrobił ogromny krok do przodu w odniesieniu do testów TPC-A i TPC-B. Chociaż w rzeczywistości modeluje również operacyjne przetwarzanie transakcji, jego złożoność jest przynajmniej o rząd wielkości większa niż złożoność testów A i B: wykorzystuje kilka typów transakcji, bardziej złożoną bazę danych i ogólną strukturę wykonawczą. Test TPC-C modeluje zadanie przetwarzania aplikacji. Testuje wszystkie główne elementy systemu: terminale, linie komunikacyjne, procesor, dyskowe operacje wejścia/wyjścia oraz bazę danych.

Konieczne jest również ponowne zwrócenie uwagi na fakt, że porady TPC dostarczają jedynie opisu wymagań dla testów. Sama realizacja testów jest zależna od tych, którzy bezpośrednio przeprowadzają porównanie.

W ramach niniejszej pracy zostaną wygenerowane bazy danych o różnych parametrach oraz stworzona zostanie aplikacja dla wysyłania zapytań. Za pomocą różnych narzędzi zostanie przeprowadzone badanie porównywanych baz danych oraz analiza otrzymanych wyników.

Zgodnie ze stanem na początek 2019 roku bazy danych takie jak MySQL oraz PostgreSQL znajdują się w pierwszej dziesiątce popularności DBMS [6]. Baza danych Firebird zajmuje nieco gorszą pozycję, co nie oznacza, że ten system zarządzania bazami danych nie jest warty uwagi (rys. 1).

Rank			DBMS
Jun 2019	May 2019	Jun 2018	
1.	1.	1.	Oracle +
2.	2.	2.	MySQL +
3.	3.	3.	Microsoft SQL Server +
4.	4.	4.	PostgreSQL +
5.	5.	5.	MongoDB +
29.	↓ 28.	↓ 27.	Vertica +
30.	30.	↓ 28.	Firebird

Rys. 1. Ranking popularności baz danych [6]

MySQL – jest to darmowy system zarządzania relacyjnymi bazami danych. MySQL jest rozwijany i wspierany przez firmę Oracle Corporation. Wcześniej MySQL był rozwijany przez szwedzką firmę „MySQL AB”, która została kupiona w 2008 roku przez Sun Microsystems, a ta przez Oracle w 2010 roku. Produkt jest rozpowszechniany na licencji GNU General Public License oraz na własnej licencji komercyjnej. MySQL posiada API i wtyczki dla Delphi, C, C ++, Javy, PHP, Pythona, Ruby, Smalltalk, Component Pascal i Tcl, .NET, a także posiada obsługę ODBC za pomocą sterownika MyODBC [11].

PostgreSQL – jest jednym z najpopularniejszych otwartych systemów zarządzania bazami danych. Rozwój Postgresa, który rozpoczął się w 1986 roku, był bezpośrednio związany z Michałem Stonebrakerem, szefem wcześniejszego projektu Ingres, w tym czasie już zakupionym przez Computer Associates. Nazwa została odszyfrowana jako „Post Ingres”. Stonebraker i jego uczniowie opracowywali nowy DBMS przez osiem lat (od 1986 do 1994 roku). W tym okresie zostały wprowadzone procedury, reguły, typy zdefiniowane przez użytkownika oraz inne komponenty. W 1995 r. rozwój został ponownie podzielony: Stonebraker wykorzystując swoje doświadczenie stworzył komercyjny DBMS „Illustra”, a jego uczniowie opracowali nową wersję Postgres -Postgres95, w której zastąpiono język Ingres na SQL. Następnie rozwój Postgres95 został przeprowadzony poza uniwersyteckim zespołem entuzjastów. Nowy DBMS otrzymał nazwę PostgreSQL, pod którą jest obecnie rozwijany. [10]

Firebird – jest systemem zarządzania relacyjnymi bazami danych. Działa w środowiskach systemów operacyjnych: Linux, Windows, Mac OS X i wielu innych [9].

2. Aplikacja testowa

W celu przeprowadzenia zaplanowanych badań systemów zarządzania bazami danych została napisana prosta aplikacja w języku Java w środowisku Intelij IDEA przeznaczona do

pracy na platformie Windows, pozwalająca na przesyłanie zapytań do baz danych oraz mierzenie czasu odpowiedzi na zapytanie.

3. Metodyka badań oraz stanowisko badawcze

Badania przeprowadzono na systemie Windows 10, za każdym razem uruchamiano jedynie system oraz silnik bazy danych. Każde badanie zostało powtórzone 10 razy, a jako ostateczny wynik przyjęto średnią wartość z pomiarów.

Parametry komputerów, na których zostały przeprowadzone badania.

Komputer 1:

- system operacyjny: Windows 10 Pro 64-bit,
- processor: Intel(R) Core(TM) i5-3230M,
- pamięć RAM: Kingston 8 GB,
- dysk SSD: Kingston 128 GB.

Komputer 2:

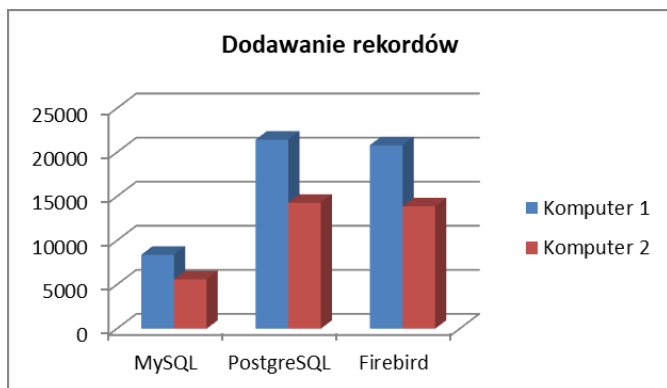
- system operacyjny: Windows 10 Pro 64-bit,
- processor : Intel(R) Core(TM) i5-8305G,
- pamięć RAM: Kingston 12 GB,
- dysk SSD: Kingston HyperX 512 GB.

4. Analiza porównawcza

Pierwszym etapem było porównanie czasów odpowiedzi baz danych podczas dodawania rekordów (rys. 2).

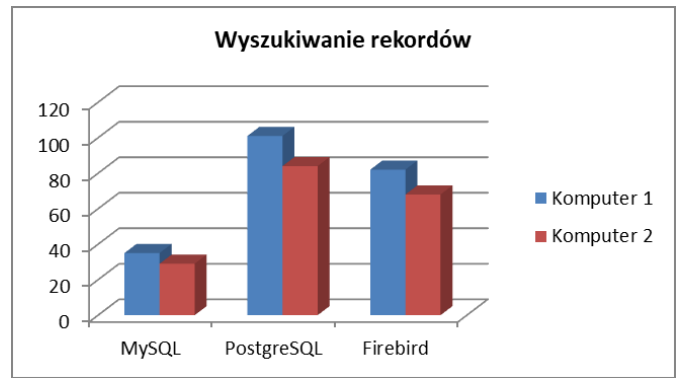
Tabela 1. Czas odpowiedzi baz danych

	Czas odpowiedzi [ms]			
	MySQL	PostgreSQL	Firebird	
Komputer 1	8400	21445	20825	Dodawanie rekordów
Komputer 2	5600	14295	13883	
Komputer 1	35	101	82	Wyszukiwanie rekordów
Komputer 2	29	84	68	
Komputer 1	20	84	47	Usuwanie rekordów
Komputer 2	30	41	39	

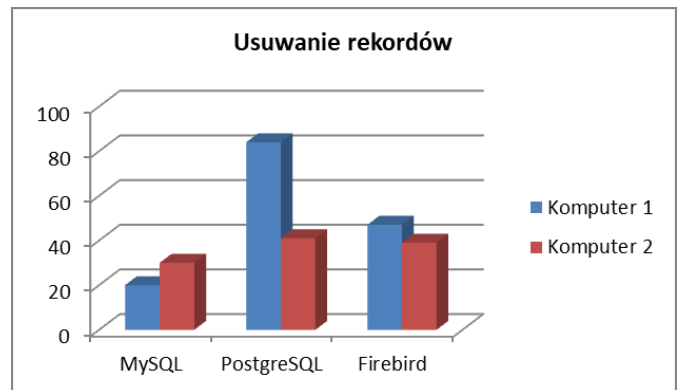


Rys. 2. Czas dodawania rekordów

Następnie analogicznie został zmierzony czas wyszukiwania rekordów oraz ich usuwania z baz danych (rys. 3, 4).



Rys. 3. Czas wyszukiwania rekordów

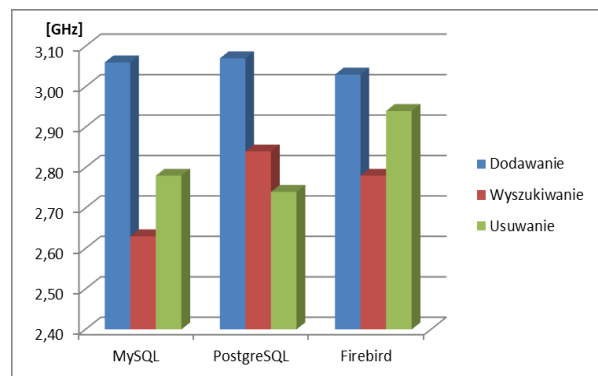


Rys. 4. Czas usuwania rekordów

Kolejnym zadaniem była analiza obciążenia procesora podczas wykonania powyższych zapytań. Wyniki są przedstawione na rysunku 5.

Tabela 2. Obciążenie procesora

	Obciążenie procesora [GHz]		
	MySQL	PostgreSQL	Firebird
Dodawanie rekordów	3,06	3,07	3,03
Wyszukiwanie rekordów	2,63	2,84	2,78
Usuwanie rekordów	2,78	2,74	2,94



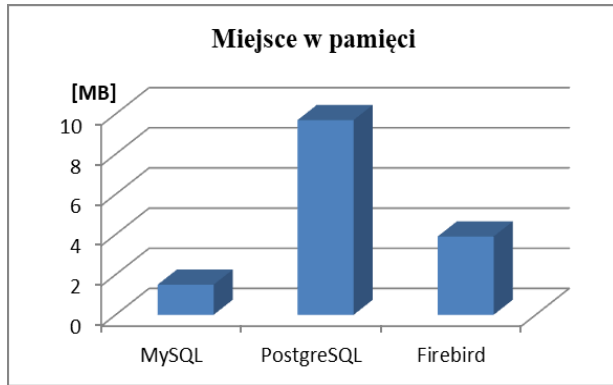
Rys. 5. Obciążenie procesora

W ostatnim etapie zostały porównane rozmiary baz danych pod względem miejsca zajmowanego na dysku

twardym. Wszystkie porównywane bazy zawierały ten sam zbiór danych.

Tabela 3. Ilość zajmowanego miejsca na dysku

Miejsce w pamięci [MB]		
MySQL	PostgreSQL	Firebird
1,5	9,7	3,9



Rys. 6. Rozmiar tabel baz danych

5. Wnioski

Podczas wszystkich przeprowadzonych badań baz danych pracujących pod kontrolą systemu Windows, pod względem szybkości wykonania zapytań pierwsze miejsce zajęła baza MySQL. Czas odpowiedzi na zapytania wysyłane do bazy MySQL był dużo krótszy od czasów uzyskanych podczas stosowania innych silników baz danych. Identycznie wygląda sytuacja pod względem ilości zajmowanego miejsca w pamięci – najlepsze wyniki uzyskała baza MySQL.

Wyniki badań obciążenia procesora wykazały zależność od rodzaju wykonywanej operacji. Podczas dodawania dużej liczby rekordów najmniejsze obciążenie procesora

zaobserwowano dla silnika bazy Firebird. Podczas wyszukiwania rekordów oraz usuwania zajął on ostatnie miejsce wśród badanych baz danych.

Na podstawie przeprowadzonych badań można stwierdzić, że system zarządzania bazami danych MySQL pod kontrolą systemu Windows jest najszybszy pod względem wykonania zapytań a utworzona w nim baza danych zajmuje mniej miejsca na dysku w porównaniu do analizowanych konkurencyjnych rozwiązań, ale odbywa się to kosztem większego obciążenia procesora. Ale ponieważ powyższe testy są przeprowadzone tylko na systemie Windows nie można stwierdzić, że zachowanie testowanych baz danych na innych systemach operacyjnych będzie podobne.

Literatura

- [1] Avi Silberschatz, Henry F. Korth, S. Sudarshan: Database System Concepts, McGraw-Hill Education, 2010.
- [2] Baron Schwartz, Peter Zaitsev, Vadim Tkachenko: High Performance MySQL: Optimization, Backups, and Replication, O'Reilly Media, 2008.
- [3] Rob Mattison: Understanding Database Management Systems, McGraw-Hill, 1997.
- [4] Date, C.J: Introduction to Database Systems, Pearson, 2000.
- [5] Toby J. Teorey: Database Modeling & Design: The Fundamental Principles, Morgan Kaufmann Pub, 1994.
- [6] DB-Engines Ranking, <https://db-engines.com/en/ranking> [31.03.2019 r.]
- [7] IBM Information Management System, https://en.wikipedia.org/wiki/IBM_Information_Management_System [23.06.2019]
- [8] TPC-Homepage V5, <http://www.tpc.org> [25.02.2019 r.]
- [9] Firebird, <https://pl.wikipedia.org/wiki/Firebird> [12.03.2019 r.]
- [10] PostgreSQL, <https://pl.wikipedia.org/wiki/PostgreSQL> [12.03.2019 r.]
- [11] MySQL, <https://pl.wikipedia.org/wiki/MySQL> [15.03.2019 r.]