POLITECHNIKA WROCŁAWSKA

WYDZIAŁ ELEKTRONIKI

KIERUNEK: INFORMATYKA (INF)

SPECJALNOŚĆ: SYSTEMY I SIECI KOMPUTEROWE (ISK)

PRACA DYPLOMOWA

MAGISTERSKA

Analiza porównawcza relacyjnych i nierelacyjnych baz danych na wybranych przykładach

Comparative analysis of relational and   
non-relational databases on selected examples

AUTOR:

Paweł Idziak

PROWADZĄCY PRACĘ:

Dr inż. Jacek Cichosz, W4\K2

OCENA PRACY:

WROCŁAW, 2019

**Spis treści**

[Spis rysunków 4](#_Toc4662652)

[Spis tabel 5](#_Toc4662653)

[Spis listingów 6](#_Toc4662654)

[Skróty 7](#_Toc4662655)

[1. Wstęp 8](#_Toc4662656)

[1.1. Wprowadzenie 8](#_Toc4662657)

[1.2. Cel pracy 8](#_Toc4662658)

[1.3. Uzasadnienie wyboru baz danych 8](#_Toc4662659)

[1.4. Przedstawienie struktury pracy 9](#_Toc4662660)

[2. Analiza wybranych baz danych 10](#_Toc4662661)

[2.1. MySQL 10](#_Toc4662662)

[2.2. PostgreSQL 10](#_Toc4662663)

[2.3. MongoDB 10](#_Toc4662664)

[2.4. Cassandra 10](#_Toc4662665)

[3. Użyte technologie 11](#_Toc4662666)

[3.1. NodeJS 11](#_Toc4662667)

[3.2. ExpressJS 11](#_Toc4662668)

[3.3. Angular 7 11](#_Toc4662669)

[3.4. TypeScript 11](#_Toc4662670)

[3.5. GIT 11](#_Toc4662671)

[4. Implementacja 13](#_Toc4662672)

[4.1. Założenia projektowe 13](#_Toc4662673)

[4.2. Wykorzystane technologie i narzędzia 13](#_Toc4662674)

[4.3. Opis działania systemu 13](#_Toc4662675)

[5. Analiza wydajnościowa 14](#_Toc4662676)

[6. Podsumowania 15](#_Toc4662677)

[Literatura 16](#_Toc4662678)

[Dodatek A 17](#_Toc4662679)

# Spis rysunków

[Rys. 1. Przykład podpisu rysunku 4](#_Toc465685478)

# Spis tabel

[Tab. 1. Przykład podpisu tabeli 4](#_Toc465685652)

# Spis listingów

[Listing. 1. Początkowe żadanie HTTP 4](#_Toc465685644)

# Skróty

**OGC** (ang. *Open Geospatial Consortium*)

**XML** (ang. *eXtensible Markup Language*)

**SOAP** (ang. *Simple Object Access Protocol*)

**WSDL** (ang. *Web Services Description Language*)

**UDDI** (ang. *Universal Description Discovery and Integration*)

**GIS** (ang. *Geographical Information System*)

**SDI** (ang. *Spatial Data Infrastructure*)

**ISO** (ang. *International Standards Organization*)

**WMS** (ang. *Web Map Service*)

**WFS** (ang. *Web Feature Service*)

**WPS** (ang. *Web Processing Service*)

**GML** (ang. *Geography Markup Language*)

**SRG** (ang. *Seeded Region Growing*)

**SOA** (ang. *Service Oriented Architecture*)

**IT** (ang. *Information Technology*)

1. Wstęp

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus. Aliquam erat volutpat. Integer maximus est turpis, ut bibendum ligula accumsan et. Ut eget vestibulum libero. Aliquam erat volutpat. Nullam placerat mauris a lectus tincidunt, et aliquet turpis aliquam. Etiam in malesuada lacus. Proin dignissim augue sit amet auctor elementum. Suspendisse potenti. Vivamus suscipit vulputate massa ac molestie. Suspendisse a justo porttitor, commodo mi at, placerat risus. Integer lobortis augue ac neque suscipit, vel sodales lacus fringilla.

WSTEP – co jest w pracy, co jest celem i jej zakres ale bez wchodzenia w szczególy.

* To, że w dzisiejszych czasach obserwuje sie gwałtowny wzrost gromadzonych danych (pokazad statystki z cisco)
* To, że wzrost danych wiąże się z zarządzaniem tymi danymi w odpowiedni sposób.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

* 1. Cel pracy

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

* 1. Zastosowanie baz danych

Coś o zastosowaniu, później definicja, DBMS, architektura, rodzaje

Internetowy słownik Merriam–Webster [1] definiuję bazę danych jako „dużą kolekcję danych, zorganizowaną w celu szybkiego przeszukiwania i dostępu do informacji”. Rozwijając tą definicję, baza danych jest zorganizowanym zbiorem danych, zapisanym z określonymi regułami oraz przechowywanym w określony sposób. W znaczeniu technologii komputerowych, definicja bazy danych obejmuje dane cyfrowe gromadzone według ogólnych zasad przyjętych dla danego systemu, przeznaczonego do wykonywania określonych operacji na danych.

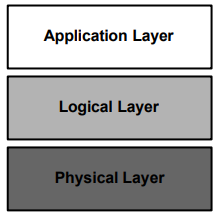
Zarządzanie bazą danych realizowane jest poprzez *DBMS* [2](ang. *Database Management System*), czyli *system zarządzania bazą danych*. Jest to najczęściej oprogramowanie, składające się z zestawu bibliotek oraz narzędzi, które umożliwiają użytkownikowi między innymi zdefiniowanie oraz utworzenie bazy. DBMS posiada interfejs dostępu do danych oraz funkcjonalności służące do zarządzania, analizowania czy przechwytywania danych. Głównym założeniem DBMS jest współdziałanie z użytkownikami końcowymi, aplikacjami, a samą bazą danych. Funkcjonalności, jakie spełnia system zarządzania bazą danych obejmuje:

* **Tworzenie.** Każda baza danych może być tworzona na różne sposoby. DBMS posiada określoną procedurę tworzenia bazy danych, dzięki czemu użytkownik nie ma potrzeby interakcji z niskopoziomowymi strukturami plików. System zarządzania bazą udostępnia dostarcza odpowiedni dostęp potrzebny użytkownikom poprzez zdefiniowany interfejs.
* **Selekcja i uaktualnienie.** Systemy DBMS powiadają funkcjonalności tworzenia zapytań o dane spełniające określone kryteria (np. wszystkie zamówienia pojedynczego użytkownika), oraz zapytań realizujących funkcje usuwania czy aktualizacji określonych danych.
* **Wielozadaniowość.** W sytuacji, gdy kilka procesów (np. aplikacji) korzysta z bazy danych jednocześnie, system zapewnia dla każdego z połączenia dostęp do zasobów / funkcjonalności bazy. Wielozadaniowość zapewnia odebranie oraz przetworzenie każdego żądania bez negatywnego wpływu na inne elementy systemu.
* **Bezpieczeństwo.** System zarządzania bazą danych zapewnia mechanizm kontroli dostępu do poszczególnych elementów. Dzięki takiemu rozwiązaniu, operacje manipulowania danymi oraz ich strukturą są dozwolone jedynie dla uprawnionych użytkowników. System DBMS zawiera mechanizmy zarządzania użytkownikami, a zdefiniowanie odpowiedniej hierarchii użytkowników pozwala na dostęp do określonych zasobów dla danej grupy i / lub operacji wykonywanych na bazie danych.
* **Historia operacji.** Kolejną ważną funkcjonalnością DBMS jest tworzenie oraz zapisywanie historii operacji wykonanych na przestrzeni określonego czasu. Taką funkcjonalność można wykorzystać do śledzenia zmian, wykrywania błędów czy rekonstrukcji danych w sytuacji awarii systemu. Kolejną zaletą jest wykorzystanie funkcjonalności historii do tworzenia rejestru operacji oraz kopi zapasowej danych.

Bazy danych mają szerokie zastosowane w różnych dziedzinach. ………..

Cos o przeznaczeniach i przejście do architektury

Architekturę bazy danych można przedstawić za pomocą abstrakcyjnego standardu *ANSI-SPARC* (ang. *American National Standards Institute, Standards Planning And Requirements Committee*), na którym opiera się większość współczesnych baz danych. Schemat *ANSI-SPARC* został zdefiniowany w 1975 roku i przedstawia sposób projektowania systemów zarządzani bazami danych (*DBMS*). Koncepcja standardu opiera się na wyróżnieniu trzech głównych warstw systemu (ODNOŚNIK DO RYS), dlatego często spotykaną nazwą jest *Architektura trójwarstwowa*. Architektura trójwarstwowa posiada trzy główne komponenty, każdy rozumiany jako inny poziom abstrakcji.



Warstwa aplikacji (ang. *Application Layer*)

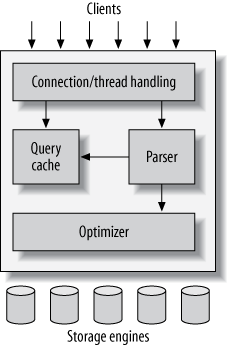
Warstwa aplikacji przedstawia pewien interfejs dostępny dla wszystkich użytkowników systemu, dzięki któremu użytkownicy mogą nawiązać interakcję z serwerem bazy danych. Wśród użytkowników można wyróżnić między innymi dwie główne grupy:

1. **Standardowi użytkownicy** – korzystają z systemu pośrednio, przez wcześniej zdefiniowany interfejs użytkownika.
2. **Zaawansowani użytkownicy** – współpracują z systemem bezpośrednio, bez użycia zdefiniowanej aplikacji. Użytkownicy tej grupy formułują żądania bezpośrednio za pomocą języka zapytań konkretnej bazy danych. Do tej grupy można zaliczyć administratorów lub programistów, którzy zajmują się implementacją aplikacji użytkowych.

Warstwa logiczna (ang. *Logical Layer*)

Warstwa logiczna zawiera podstawowe funkcjonalności zapewniające komunikację z fizyczną bazą danych. Warstwa logiczna przetwarza żądania użytkowników, definiuje reguły logiczne oraz wykonuje zdefiniowane obliczenia. Warstwa ta również pełni funkcję pośrednią pomiędzy otaczającymi ją warstwami (aplikacji oraz fizycznej) zapewniając przepływ danych. W tej części systemu istnieje wiele różnych implementacji komunikacji specyficznych dla danego systemu.

Warstwa fizyczna (ang. *Physical Layer*)



W zależności od stosowanej struktury organizacji danych w konkretnym systemie, można wyróżnić między innymi takie rodzaje baz danych jak:

* Relacyjne (RDBMS – ang. *Relational Database Management System*)
* Nierelacyjne (NoSQL – ang. *Non SQL / Non relational*)
* Obiektowe (OODBMS – ang. *Object-Oriented Database Management System*)
* Obiektowe – relacyjne (ORDBMS – ang. *Object-Relational Database Management System*)

Relacyjne bazy danych

Nierelacyjne bazy danych

Obiektowe bazy danych

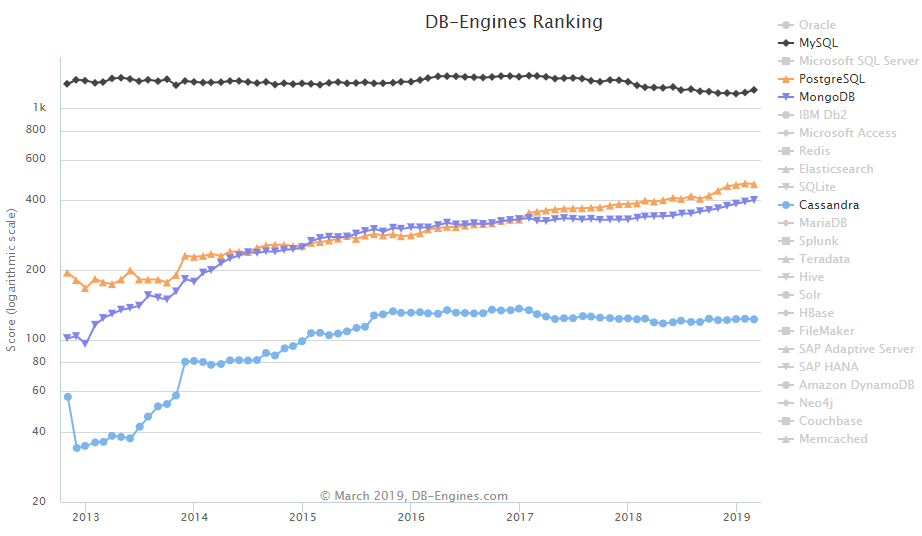
Obiektowo – relacyjne bazy danych

* 1. Uzasadnienie wyboru baz danych

W ramach niniejszej pracy magisterskiej zostały wybrane dwie relacyjne oraz dwie nierelacyjne bazy danych. Głównym kryterium wyboru konkretnych baz danych jest ich popularność datująca na marzec 2019r. Ranking popularności został zaczerpnięty ze strony internetowej *DB-ENGINES.COM* [3]. W momencie pisania pracy metoda użyta do zmierzenia popularności baz danych używa takich parametrów jak:

* **Liczba „wzmianek” o danym systemie na stronach internetowych**, mierzona liczbą wyników w popularnych internetowych wyszukiwarkach. Używane wyszukiwarki to Google, Bing oraz Yandex.
* **Ogólne zainteresowanie systemem bazy danych** mierzona jako częstotliwość wyszukiwania w Google Trends.
* **Częstotliwość dyskusji technicznych na temat systemu.** Dane są pobierane ze znanych witryn internetowych związanych z technologią IT takich jak Stack Overflow i DBA Stack Exchange.
* **Liczba ofert pracy, w których wspomniany jest dany system.** Oferty prace są przeglądane w serwisach takich jak Indeed oraz Simply Hired.

Jako relacyjne systemy zostały wybrane takie bazy jak *MySQL* oraz *PostgreSQL*, natomiast jako nierelacyjne *MongoDB* oraz *Cassandra*. Wykres poniżej przedstawia tendencję popularności wybranych systemów, na podstawie której zostały wybrane wyżej wymienione bazy.



Rys. 1. Tendencja popularności baz danych - marzec 2019. Źródło: <https://db-engines.com/en/ranking_trend>

* 1. Przedstawienie struktury pracy

Niniejsza praca magisterska została podzielona na …

1. Analiza wybranych baz danych

W tym rozdziale zostanie przedstawiona analiza wybranych relacyjnych i nierelacyjnych baz danych. Wybór baz danych został uzasadniony w rozdziale 1.3.

Jednym z analizowanych aspektów jest model bazy danych. Relacyjne bazy danych posiadają zdefiniowane modele, lecz nierelacyjne bazy danych nie posiadają spójnego modelu danych. Wśród nierelacyjnych modeli baz danych możemy wyróżnić między innymi TUTAJ JAKIŚ MODEL. Porównanie będzie zawierać przedstawienie cech wspólnych lub istotnych różnic.

MODEL DANYCH CZYLI JAK TE DANE SA PRZECHOWYWANE

Kolejnym analizowanym aspektem jest architektura bazy danych rozumiana jako elementy, z których składa się dany system oraz funkcjonalności takie jak między innymi NIE WIEM CZY TEN AKAPIT BĘDZIE OK. WYJDZIE W PRANIU.

Kolejnym analizowanym aspektem jest dostępność bazy danych. Każda baza danych potrzebuje pewnego sterownika (biblioteki), który służy do komunikacji pomiędzy aplikacją kliencką, a bazą danych. Aspekt dostępności uwzględnia takie założenia jak poziom trudności implementacji połączenia z bazą danych, rozumiany jako jakość instrukcji dostępnych w oficjalnej dokumentacji, oraz rodzaj języków programowania, poprzez które można ustanowić połączenie. Im większa liczba dostępnych języków programowania, tym mniejsze ograniczenie wyboru bazy danych.

Każdy aspekt analizy dla kolejnej bazy danych zostanie uwzględniony w poszczególnych podsumowaniach z uwzględnieniem cech nietypowych lub szczególnych dla danego rozwiązania.

* 1. MySQL

MySQL jest to system zarządzania relacyjnymi bazami danych (RDBMS – ang. *Relational Database Management System*), udostępniony na zasadzie *wolnego oprogramowania* (ang. *free software*). Termin ten określa oprogramowanie, które może być uruchamiane, powielane, zmieniane oraz rozpowszechniane przez użytkowników. Więcej informacji na temat *wolnego oprogramowania* można znaleźć na oficjalnej stronie FSF (ang. *Free Software Foundation)* (TUTAJ ODNOŚNIK).

Obecnie (marzec 2019) MySQL rozwijany jest przez firmę *Oracle*, lecz pierwsze wydanie systemu datowane jest na 23 maja 1995 roku. Pierwsza wersja MySQL została stworzona przez szwedzką firmę *MySQL AB*, która 16 stycznia 2008 roku została wykupiona przez *Sun Microsystems*, natomiast *Sun Microsystems* został wykupiony 27 stycznia 2019 roku przez *Oracle.*

* + 1. Model danych
  1. PostgreSQL
  2. MongoDB
  3. Cassandra

1. Użyte technologie
   1. NodeJS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

* 1. ExpressJS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

* 1. Angular 7

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

* 1. TypeScript

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

* 1. GIT

Tab. 1. Przykład podpisu tabeli

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

Listing. 1. Początkowe żądanie HTTP

GET /script/Articles/Latest.aspx HTTP/1.1  
Host: www.codeproject.com  
Connection: keep –alive  
Cache -Control: max-age=0  
Accept: text/html ,application/xhtml+xml,application/xml|  
User -Agent: Mozilla/5.0 ...  
Accept -Encoding: gzip ,deflate ,sdch  
Accept -Language: en-US...  
Accept -Charset: windows -1251,utf -8...

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus.

Przykład listy numerowanej:

1. Implementacja

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Mauris id dapibus enim. Etiam lobortis pulvinar enim in maximus. Aliquam erat volutpat. Integer maximus est turpis, ut bibendum ligula accumsan et. Ut eget vestibulum libero. Aliquam erat volutpat. Nullam placerat mauris a lectus tincidunt, et aliquet turpis aliquam. Etiam in malesuada lacus. Proin dignissim augue sit amet auctor elementum. Suspendisse potenti. Vivamus suscipit vulputate massa ac molestie. Suspendisse a justo porttitor, commodo mi at, placerat risus. Integer lobortis augue ac neque suscipit, vel sodales lacus fringilla.

3 i 4 polaczyc: opis środowiska testowego/platformy badawczej, rozwiazania jakie uzylem, technologie, architektura

* 1. Założenia projektowe
  2. Wykorzystane technologie i narzędzia
  3. Opis działania systemu

1. Analiza wydajnościowa
2. Podsumowania

# Literatura

[1] M. Bickley, C. Slominski. A MySQL-based data archiver: preliminary results. Proceedings of ICALEPCS07, Paz. 2007. http://www.osti.gov/scitech/servlets/purl/922267 [dostęp dnia 20 czerwca 2015].

[2] J. Jędrzejczyk, B. Sródka. Segmentacja obrazów metodą drzew decyzyjnych. Raport instytutowy, Politechnika Wrocławska, Wydział Elektroniki, 2007.

[1] „Dictionary by Merriam-Webster: America’s most-trusted online dictionary”. [Online]. Dostępne na: https://www.merriam-webster.com/. [Udostępniono: 29-mar-2019].

[2] C. Coronel i S. Morris, *Database Systems*. 2004.

[3] „DB-Engines - Knowledge Base of Relational and NoSQL Database Management Systems”. [Online]. Dostępne na: https://db-engines.com/en/. [Udostępniono: 25-mar-2019].

# Dodatek A