

Bazy NoSQL: przeгляд rozwiązań



Your personal e-commerce solutions



Agenda

1. Czym są bazy NoSQL?
2. Bazy dokumentowe
3. Bazy klucz-wartość
4. Bazy kolumnowe
5. Bazy grafowe
6. Bazy hybrydowe
7. Zalety NoSQL
8. Wady NoSQL
9. NoSQL vs Relacyjne bazy danych
10. Jaką bazę NoSQL wybrać?
11. NoSQL - najlepsze praktyki
12. Podsumowanie



Czym są bazy NoSQL?

“Baza danych zapewniająca mechanizm do przechowywania i wyszukiwania danych modelowanych w inny sposób niż relacje tabelaryczne używane w relacjach baz danych SQL.”

Źródło: Wikipedia (<https://pl.wikipedia.org/wiki/NoSQL>)

Bazy dokumentowe

- Przechowują dane jako dokumenty, czyli strukturach składających się z pól klucz-wartość (np. pliki JSON)
- Nie ma wymogu co do schematu przechowywanych danych
- Zoptymalizowane na rozpraszanie i odczyty
- Dobre do przechowywania danych przychodzących z API (JSONy mogą lecieć bezpośrednio do bazy bez żadnych transformacji)
- Przykłady: MongoDB, CouchDB, Orient DB

Document 1

```
{
  "id": "1",
  "name": "John Smith",
  "isActive": true,
  "dob": "1964-30-08"
}
```

Document 2

```
{
  "id": "2",
  "fullName": "Sarah Jones",
  "isActive": false,
  "dob": "2002-02-18"
}
```

Document 3

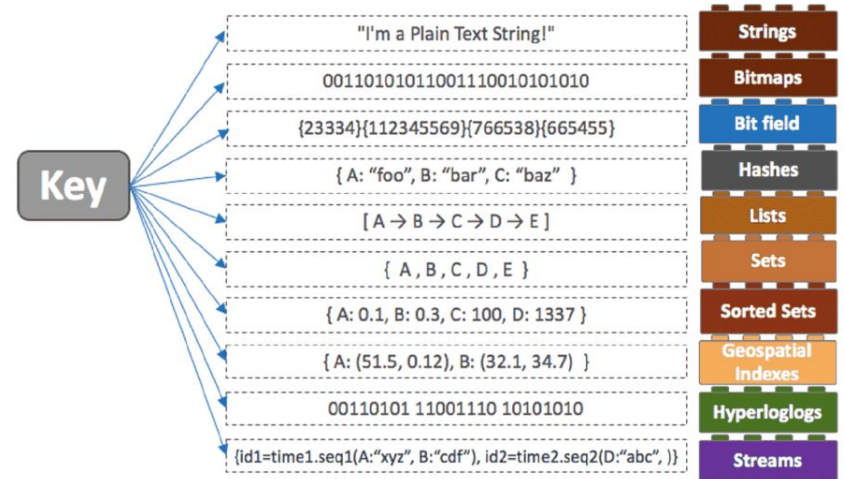
```
{
  "id": "3",
  "fullName": {
    "first": "Adam",
    "last": "Stark"
  },
  "isActive": true,
  "dob": "2015-04-19"
}
```

Źródło:

<https://lennilobel.wordpress.com/2015/06/01/relational-databases-vs-nosql-document-databases/>

Bazy klucz-wartość

- Przechowują dane na zasadzie klucz-wartość
- Wartość może być typem prostym bądź bardziej złożonym (tablica-obiekt)
- Im prostsza wartość tym lepiej
- Nie ma wymogu co do schematu przechowywanych danych
- Zoptymalizowane na rozpraszanie i odczyty
- Często realizowane jako baza rezydująca w pamięci
- Mniej zalecana do wykonywania skomplikowanych agregacji
- Dobra jako pamięć cache bądź baza pośrednia
- Przykłady: BerkeleyDB, Memcached, Redis



Źródło:

<https://redislabs.com/redis-enterprise/data-structures/>

Bazy kolumnowe

- Przechowują dane w kolumnach zamiast w wierszach jak w relacyjnych bazach danych
- Dzięki temu możliwe są efektywne agregacje po kolumnach
- Im więcej danych tym większy zysk widać
- Często wykorzystywane jako baza analityczna bądź do hurtowni danych
- Przykłady: Cassandra, HBase, Apache Druid

Row-oriented

ID	Name	Grade	GPA
001	John	Senior	4.00
002	Karen	Freshman	3.67
003	Bill	Junior	3.33

Column-oriented

Name	ID
John	001
Karen	002
Bill	003

Grade	ID
Senior	001
Freshman	002
Junior	003

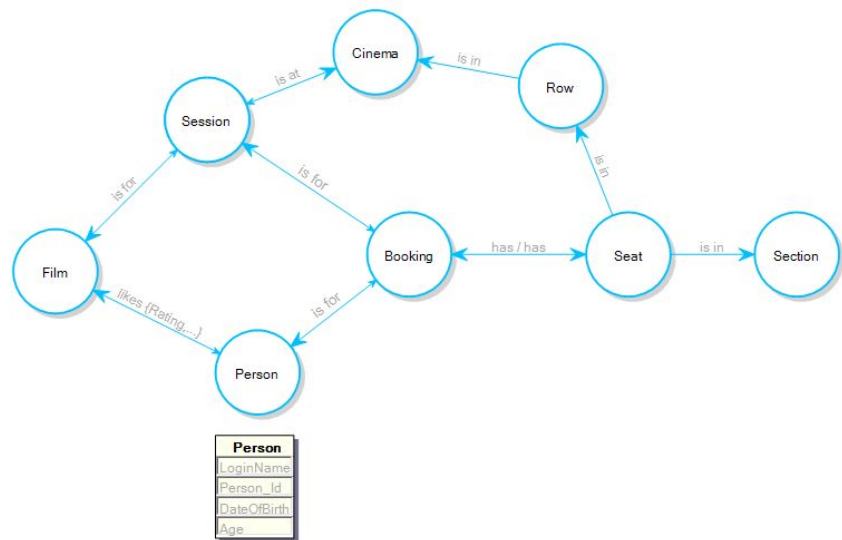
GPA	ID
4.00	001
3.67	002
3.33	003

Źródło:

<https://searchdatamanagement.techtarget.com/definition/columnar-database>

Bazy grafowe

- Przechowują dane jako wierzchołki, powiązania między nimi oraz cechy na węzłach i gałęziach
- Dobra analityka dla danych dobrze reprezentowanych jako sieć (najkrótsza ścieżka w grafie, sąsiedztwo, stopnie wierzchołków)
- Ciężkie do rozproszenia
- Zalecane jako baza specjalistyczna, raczej nie jako podstawowe źródło danych
- Przykłady: FlockDB, Infinite Graph, Neo4J

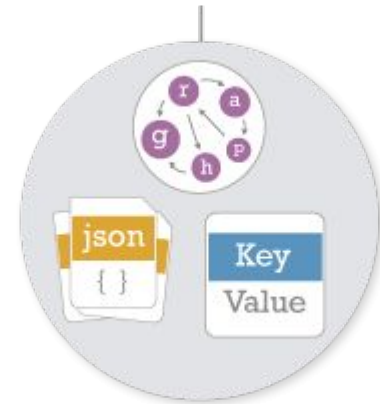


Źródło:

<https://towardsdatascience.com/what-is-a-graph-database-249cd7fdf24d>

Bazy hybrydowe

- Bazy, które posiadają reprezentowania danych w wielu modelach, np. klucz-wartość i grafowe, lub dokumentowo-grafowe.
- Język zapytań dostosowanych do wielu modeli danych
- Wygodne gdy charakter danych jest różnorodny
- Mniej zoptymalizowane pod konkretny model danych
- Przykłady: ArangoDB, OrientDB



Źródło: <https://spyurk.am/tags/multi-model>

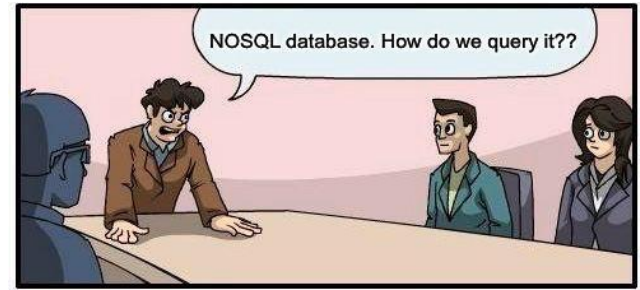
Zalety NoSQL

- Zaprojektowane z myślą o dużych zbiorach danych
- Zaprojektowane z myślą o rozproszeniu
- Pozwalają na dynamiczny schemat danych
- Brak join-ów z relacyjnych baz danych :-)



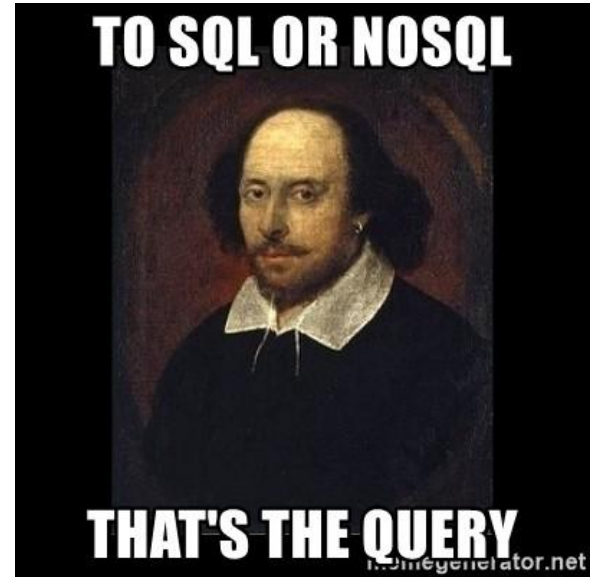
Wady NoSQL

- W większości przypadków nie da się zapewnić transakcyjności po stronie bazy danych
- Nie wybaczą błędów projektowych :-)
- Wykonywanie agregacji może być utrudnione z uwagi na rozproszenie danych
- Preferowanie INSERT-ów i SELECT-ów zamiast UPDATE-ów i DELETE-ów



NoSQL vs Relacyjne bazy danych

- Relacyjna baza danych wybacza wiele błędów - NoSQL już nie
- Transakcyjność i ACID w relacyjnych bazach danych normą, w NoSQL problem
- NoSQL pozwala względnie łatwo rozproszyć dane i organizować je w klastry - relacyjne bazy danych mają to utrudnione
- NoSQL przetwarza ogromne zbiory danych względnie łatwo - relacyjna baza danych nie wytrzyma takiego wolumenu danych



Jaką bazę NoSQL wybrać?

Heurystyki:

- Pamięci cache i bazy z szybkim dostępem do danych - **baza klucz-wartość**
- Analizy i raporty z dużą liczbą agregacji - **baza kolumnowa**
- Podstawowe źródło danych - **baza dokumentowa**
- Dane łatwo reprezentowane jako sieć powiązań i analizy/raporty na tych danych - **baza grafowa**
- Baza ogólnego przeznaczenia i chęć minimalizacji używanych technologii - **baza hybrydowa**



NoSQL - najlepsze praktyki

- Dobierać bazę danych do problemu - nie próbować rozwiązywać wszystkiego jednym młotkiem
- Każdą wybraną bazę sprawdzić pod kątem potencjalnego jej skalowania (bez doktoryzowania się z niej) - jak zachowa się przy naszych danych i jak dużą skalę problemu jest w stanie obsłużyć
- Zawsze Proof-of-concept przed produkcyjnym użyciem!
- Zawsze zaczynać od bazy relacyjnej i dopiero jak nie daje rady myśleć o NoSQL. **NoSQL nie wybacza błędów!**



O firmie Endora

Zajmujemy się implementacją sklepów i aplikacji internetowych oraz informatyzacją procesów biznesowych w branży e-commerce. Od zrozumienia procesów biznesowych do wdrożenia skalowalnego rozwiązania – tak właśnie przekuwamy pomysły w realizację.

- ponad 10 lat na rynku
- od początku związani z platformą Magento (od wersji 1.6)
- ponad 90 ukończonych projektów
- współpraca z produktami e-commerce różnej skali: od mikrosklepów po platformy działające na kilkunastu rynkach zagranicznych
- projektujemy i wykorzystujemy modele sztucznej inteligencji i uczenia maszynowego w e-commerce

W przypadku pytań zapraszam do kontaktu



Michał Zabielski

Właściciel / Owner

michal.zabielski@endora.pl

<https://www.linkedin.com/in/mzabielskipl/>

ENDORA - Your personal e-commerce solutions

<https://endora.pl>

Aleja Rzeczypospolitej 18/93

02-972 Warszawa, Polska

+48 512 707 629 (Kom.)

biuro@endora.pl

NIP 522 289 11 66 | REGON 147047424



Mendora

Your personal e-commerce solutions