

Dokumentowe bazy danych – MongoDB

Materialy:

Książki

Np.

- Shannon Bradshaw, Eoin Brazil, Kristina Chodorow, MongoDB: The Definitive Guide. Powerful and Scalable Data Storage, O'Reilly 2019
- Alex Giamas, Mastering MongoDB 4.x., Pact 2019

Dokumentacja

- <https://www.mongodb.com/docs/manual/reference/program/mongo/>

MongoDB University Courses

- <https://university.mongodb.com/courses/catalog>
- MongoDB Basics
 - <https://university.mongodb.com/courses/M001/about>
- The MongoDB Aggregation Framework
 - <https://university.mongodb.com/courses/M121/about>
- Data Modeling
 - <https://university.mongodb.com/courses/M320/about>

Zbiory danych

The MongoDB Atlas Sample Datasets

<https://www.mongodb.com/developer/article/atlas-sample-datasets/>

Yelp Dataset

www.yelp.com - serwis społecznościowy – informacje o miejscach/lokalach

- restauracje, kluby, hotele itd. (*businesses*),
- użytkownicy piszą recenzje (*reviews*) o miejscach i wystawiają oceny,
- użytkownicy odwiedzają te miejsca - "meldują się" (*check-in*)
- Przykładowy zbiór danych zawiera dane z 5 miast: Phoenix, Las Vegas, Madison, Waterloo i Edinburgh.

Kolekcje:

- **42,153** businesses
- **320,002** business attributes
- **31,617** check-in sets
- **252,898** users
- **955,999** edge social graph
- **403,210** tips
- **1,125,458** reviews

business

```
{
  'type': 'business',
  'business_id': (encrypted business id),
  'name': (business name),
  'neighborhoods': [(hood names)],
  'full_address': (localized address),
  'city': (city),
  'state': (state),
  'latitude': latitude,
  'longitude': longitude,
  'stars': (star rating, rounded to half-stars),
  'review_count': review count,
  'categories': [(localized category names)]
  'open': True / False (corresponds to closed, not business hours),
  'hours': {
    (day_of_week): {
      'open': (HH:MM),
      'close': (HH:MM)
    },
    ...
  },
  'attributes': {
    (attribute_name): (attribute_value),
    ...
  },
}
```

review

```
{
  'type': 'review',
  'business_id': (encrypted business id),
  'user_id': (encrypted user id),
  'stars': (star rating, rounded to half-stars),
  'text': (review text),
  'date': (date, formatted like '2012-03-14'),
  'votes': {(vote type): (count)},
}
```

user

```
{
  'type': 'user',
  'user_id': (encrypted user id),
  'name': (first name),
  'review_count': (review count),
  'average_stars': (floating point average, like 4.31),
  'votes': {(vote type): (count)},
  'friends': [(friend user_ids)],
  'elite': [(years_elite)],
  'yelping_since': (date, formatted like '2012-03'),
  'compliments': {
    (compliment_type): (num_compliments_of_this_type),
    ...
  },
  'fans': (num_fans),
}
```

check-in

```
{
  'type': 'checkin',
  'business_id': (encrypted business id),
  'checkin_info': {
    '0-0': (number of checkins from 00:00 to 01:00 on all Sundays),
    '1-0': (number of checkins from 01:00 to 02:00 on all Sundays),
    ...
    '14-4': (number of checkins from 14:00 to 15:00 on all Thursdays),
    ...
    '23-6': (number of checkins from 23:00 to 00:00 on all Saturdays)
  }, # if there was no checkin for a hour-day block it will not be in the dict
}
```

tip

```
{
  'type': 'tip',
  'text': (tip text),
  'business_id': (encrypted business id),
  'user_id': (encrypted user id),
  'date': (date, formatted like '2012-03-14'),
  'likes': (count),
}
```

Narzędzia

- Konsola mongo/mongosh
- MongoDB Compass
- Robo3T
- VS Code
- DataGrip
- Studio3T

Zadania

1. Uruchom lokalny serwer MongoDB
2. Połącz się lokalnym serwerem MongoDB
 - za pomocą konsoli mongo/mongosh
 - ewentualnie zdefiniuj połączenie w wybranym przez siebie narzędziu
3. Stwórz własny klaster/bazę danych w serwisie MongoDB Atlas
 - <https://www.mongodb.com/atlas/database>
 - Serwis umożliwia założenie darmowego konta
 - Połącz się z bazą danych w serwisie MongoDB Atlas
 - za pomocą konsoli mongo/mongosh
 - ewentualnie zdefiniuj połączenie w wybranym przez siebie narzędziu
4. Stwórz bazę danych/kolekcję/dokument
 - na lokalnym serwerze bazy danych
 - w serwisie MongoDB Atlas
 - może to być dowolny dokument – o dowolnej strukturze
5. Zaimportuj przykładowe zbory danych
 - MongoDB Atlas Sample Dataset
 - <https://docs.atlas.mongodb.com/sample-data>
 - w przypadku importu do lokalnej bazy danych wykorzystaj komendę mongorestore
 - `mongorestore <data-dump-folder>`
 - np.
`mongorestore samples`
 - Yelp Dataset
 - wykorzystaj komendę mongoimport
 - `mongoimport --db <db-name> --collection <coll-name> --type json --file <file>`
 - np.
`mongoimport --db yelp --collection business --type json --file ./yelp_academic_dataset_business.json`
 - ewentualnie do importu danych można wykorzystać narzędzie MongoDB Compass
6. Zapoznaj się ze strukturą przykładowych zbiorów danych/kolekcji
7. Operacje CRUD
 - <https://www.mongodb.com/docs/manual/crud/>
 - stwórz nową bazę danych, jako nazwy bazy danych użyj swoich inicjałów
 - stwórz kolekcję "student"
 - informacje o studentach, przedmiotach ocenach z przedmiotów itp.
 - zaproponuj strukturę dokumentu

- wykorzystaj typy proste/podstawowe, dokumenty zagnieżdżone, tablice
 - wprowadź kilka przykładowych dokumentów
 - przetestuj operacje wstawiania, modyfikacji/usuwania dokumentów
 - przetestuj operacje wyszukiwania dokumentów
8. Operacje wyszukiwania danych

Dla zbioru Yelp wykonaj następujące zapytania

W niektórych przypadkach może być potrzebne wykorzystanie mechanizmu Aggregation Pipeline
<https://www.mongodb.com/docs/manual/core/aggregation-pipeline/>

Przykładowe zadania:

- a) Zwróć dane wszystkich zamkniętych (*open*) firm (*business*). Zapytanie powinno zwracać dane z pól: nazwa, adres, gwiazdki (*stars*)
- b) Ile miejsc ocenianych na 5 gwiazdek (pole *stars*, kolekcja *business*)
- c) Ile restauracji znajduje się w każdym mieście. (pole *categories* w dokumencie *business* musi zawierać wartość *Restaurants*).
- d) Zwróć bez powtórzeń wszystkie nazwy miast w których znajdują się firmy (*business*)
- e) Ile każda firma otrzymało wskazówek (*tip*) w 2012. Wynik posortuj według liczby wskazówek (*tip*).
- f) Recenzje mogą być oceniane przez innych użytkowników jako *cool*, *funny* lub *useful* (jedna recenzja może mieć kilka głosów w każdej kategorii). Napisz zapytanie, które zwraca dla każdej z tych kategorii, ile sumarycznie recenzji zostało oznaczonych przez te kategorie (np. recenzja ma kategorię *funny* jeśli co najmniej jedna osoba zagłosowała w ten sposób na daną recenzję)
- g) Zwróć dane wszystkich użytkowników (*user*), którzy nie uzyskali ani jednego pozytywnego głosu z kategorii (*funny* lub *useful*), wynik posortuj alfabetycznie według imienia użytkownika.
- h) Wyznacz, jaką średnią ocenę (*stars*) uzyskała każda firma (*business*) na podstawie wszystkich recenzji. Wynik ogranicz do recenzji, które uzyskały min 4.0 gwiazdki.

Należy wykonać 4 wybrane zadania

W sprawozdaniu należy umieścić kod rozwiązań oraz zrzuty ekranów z uzyskanymi wynikami

9. Modelowanie danych

- Zaproponuj strukturę bazy danych dla wybranego/przykładowego zagadnienia/problemu
- Należy wybrać jedno zagadnienie/problem (A lub B)

Przykład A

- Wykładowcy, przedmioty, studenci, oceny
- Wykładowcy prowadzą zajęcia z poszczególnych przedmiotów
- Studenci uczęszczają na zajęcia
- Wykładowcy wystawiają oceny studentom
- Studenci oceniają zajęcia

Przykład B

- Firmy, wycieczki, osoby

- Firmy organizują wycieczki
- Osoby rezerwują miejsca/wykupują bilety
- Osoby oceniają wycieczki

Warto zaproponować/rozważyć różne warianty struktury bazy danych i dokumentów w poszczególnych kolekcjach oraz przeprowadzić dyskusję każdego wariantu (wskazać wady i zalety każdego z wariantów)

W sprawozdaniu należy zamieścić przykładowe dokumenty w formacie JSON (wraz z odpowiednim komentarzem opisującym strukturę dokumentów) oraz polecenia ilustrujące wykonanie przykładowych operacji na danych

Do sprawozdania należy kompletny zrzut wykonanych/przygotowanych baz danych (taki zrzut można wykonać np. za pomocą poleceń mongoexport, bsondump ...). Do sprawozdania należy dołączyć plik zip.

Punktacja za zadanie (razem 2pkt)

Pkt 7 – 20% - 0,4pkt

Pkt 8 – 40% - 0,8pkt

Pkt 9 – 40% - 0,8pkt