

Dokumentowe bazy danych – MongoDB

Ćwiczenie 1 - do wykonania podczas zajęć

Imię i nazwisko:

Jakub Szarekko

Materiały:

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: mongosh, MongoDB Database Tools
- Studio3T
- MongoDB Compass
- DataGrip
- VS Code

Zadania

1. Połącz się serwerem MongoDB
 - Można skorzystać z własnego/lokanego serwera MongoDB
 - Można skorzystać własny klaster/bazę danych w serwisie MongoDB Atlas
 - <https://www.mongodb.com/atlas/database>
 - Serwis umożliwia założenie darmowego konta
 - Połącz za pomocą konsoli mongosh
 - Ewentualnie zdefiniuj połączenie w wybranym przez siebie narzędziu

Pobrałem serwer bazodanowy MongoDB, a następnie powłokę Mongosh. Połączyłem się z bazą danych za pomocą powłoki.

```
PS C:\Users\Jakub> mongosh

Current Mongosh Log ID: 647665ed217395d1efc81162
Connecting to:
mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=
mongosh+1.9.1
Using MongoDB:      6.0.6
Using Mongosh:      1.9.1

For mongosh info see: https://docs.mongodb.com/mongodb-shell/

-----
The server generated these startup warnings when booting
  2023-05-30T22:03:34.331+02:00: Access control is not enabled for the database. Read and write
access to data and configuration is unrestricted
-----

test>
```

2. Stwórz bazę danych/kolekcję/dokument

- może to być dowolna kolekcja, dowolny dokument – o dowolnej strukturze
- chodzi o przetestowanie działania połączenia

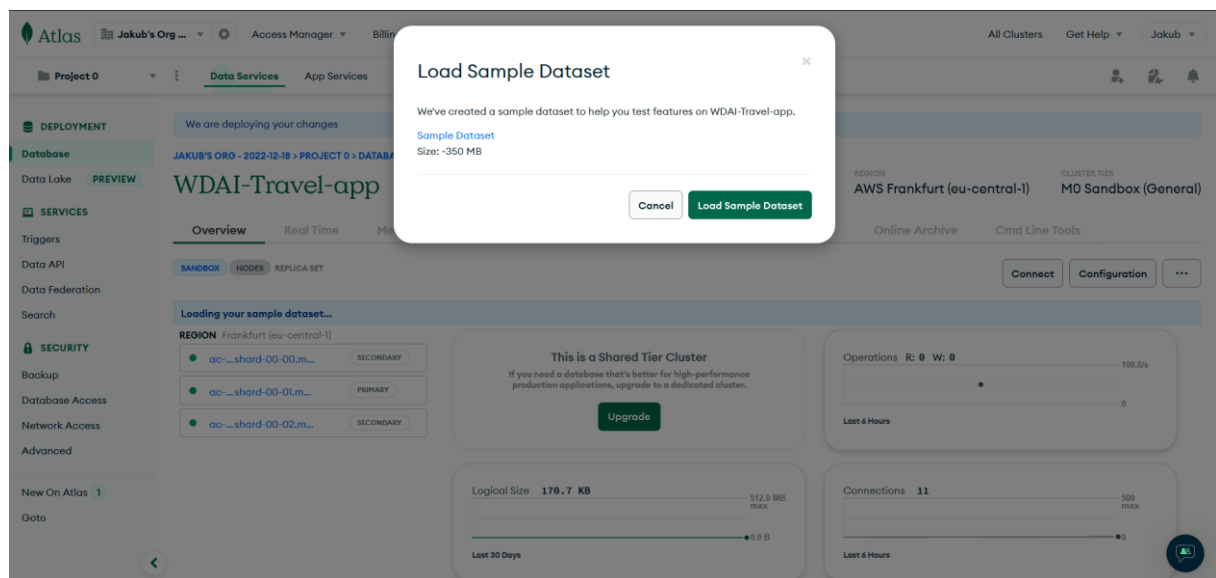
Połączenie działa, stworzyłem nową kolekcję BazyTestCollection.

```
test> db.createCollection('BazyTestCollection')
{ ok: 1 }
test> db.getCollectionNames()
[ 'BazyTestCollection' ]
```

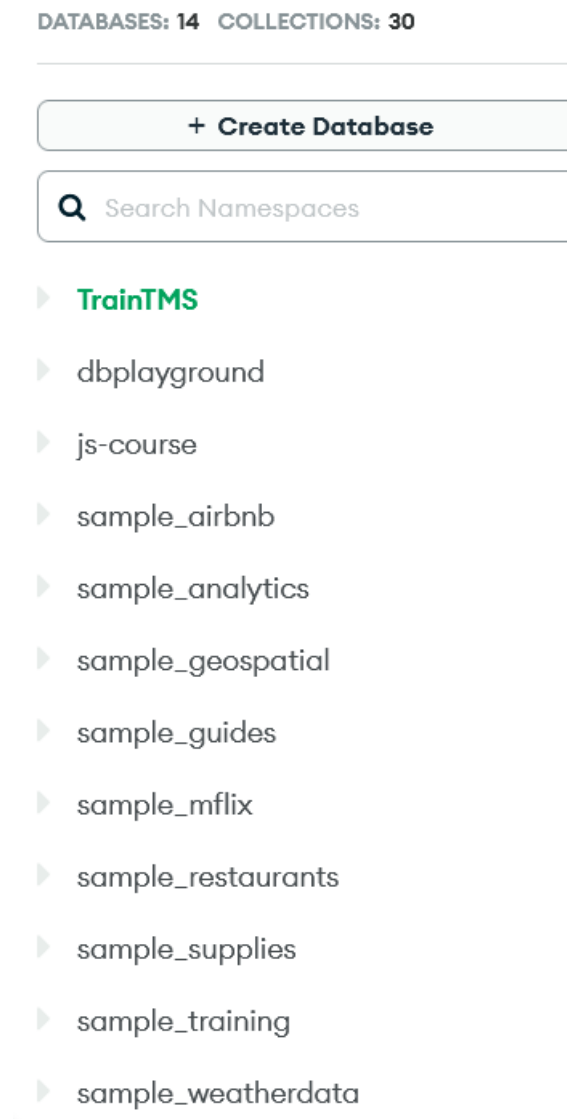
3. Zaimportuj przykładowe zbory danych

- MongoDB Atlas Sample Dataset
 - <https://docs.atlas.mongodb.com/sample-data>
 - w przypadku importu z lokalnych plików można wykorzystać polecenie mongorestore
 - mongorestore <data-dump-folder>
 - np.
mongorestore samples
 - Oczywiście, w przypadku łączenia się zdalnym serwerem należy podać parametry połączenia oraz dane logowania

Atlas Sample Dataset zaimportowałem bezpośrednio na MongoDB Atlas.



Import zbiorów danych się powiódł.



- 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
```
 - można też wykorzystać np. narzędzie MongoDB Compass

4. Zapoznaj się ze strukturą przykładowych zbiorów danych/kolekcji

- W bazach danych: MongoDB Atlas Sample Dataset

- Skomentuj struktury użyte w dokumentach dla dwóch wybranych zbiorów (takich które wydają ci się najciekawsze)
- Np. Sample Analytics Dataset i Sample Training Dataset
- W bazie Yelp
 - Skomentuj struktury użyte w dokumentach bazy Yelp

MongoDB Atlas Sample Dataset

Ze zestawu zbiorów danych od MongoDB Atlas wybrałem dwie:

- `sample_analytics`,
- `sample_guides`.

`sample_analytics`

Baza zawiera 3 kolekcje: `accounts`, `customers`, `transactions`.

`accounts`

Kolekcja zawierająca informacje dotyczące kont klientów. Zawiera 3 pola:

- `account_id` – id konta użytkownika;
- `limit` – limit (najprawdopodobniej pojedynczej transakcji);
- `products` – lista produktów zakupionych przez użytkownika.

`customers`

Kolekcja zawierająca szczegółowe informacje na temat samych klientów: Zawiera 7 pól:

- `username` – nazwa użytkownika;
- `name` – imię i nazwisko;
- `address` – pełny adres zamieszkania;
- `birthdate` – data urodzenia;
- `email` – adres email;
- `accounts` – lista id kont użytkowników;
- `tier_and_details` – obiekt zawierający wszystkie role lub rangi użytkownika wraz z ich szczegółami (każdy klucz obiektu to id, a wartość to kolejny obiekt).

`transactions`

Kolekcja zawierająca dane transakcji poszczególnych klientów: Zawiera 5 pól:

- `account_id` – id konta użytkownika;
- `transaction_count` – liczba dokonanych transakcji;
- `bucket_start_date` – najmłodsza data transakcji;
- `bucket_end_date` – najstarsza data transakcji;
- `transactions` – lista obiektów dokonanych transakcji:
 - `date` – data transakcji;
 - `amount` – liczba zakupionych produktów;
 - `transaction_code` – kod transakcji (`buy` | `sell`);
 - `symbol` – skrót witryny, na której została przeprowadzona transakcja;

- price – cena jednostkowa;
- total – łączny koszt.

sample_guides

Baza zawiera 1 kolekcję: planets.

planets

Kolekcja zawierająca informacje dotyczące planet Układu Słonecznego. Zawiera 5 pól:

- name – nazwa;
- orderFromSun – liczba porządkowa planety oddalonej od Słońca;
- hasRings – boolean czy posiada pierścienie;
- mainAtmosphere – lista pierwiastków tworzących atmosferę (string);
- surfaceTemperatureC – obiekt zawierający minimalną, maksymalną i średnią temperaturę.

Yelp

Zestaw danych posiada 5 kolekcji: business, review, user, check-in, tip.

business odpowiada za dany lokal, zawiera m. in. typ, pełen adres, koordynaty, ocenę (ilość gwiazdek zaokrąglonych do .5) obiekt z godzinami otwarcia.

review odpowiada za opinie.

user odpowiada za użytkowników 😊 Zawiera m. in. Typ, liczbę wystawionych ocen, listę przyjaciół, liczbę fanów, obiekt ze skargami.

check-in odpowiada za rezerwacje.

tip odpowiada za napiwki.

5. Operacje CRUD

- <https://www.mongodb.com/docs/manual/crud/>
- stwórz nową bazę danych, jako nazwy bazy danych użyj swoich inicjałów

```
dbplayground> use js
switched to db js
js>
```

```
dbplayground> use js
switched to db js
```

- stwórz kolekcję "student"
 - informacje o studentach, przedmiotach ocenach z przedmiotów itp.
 - zaproponuj strukturę dokumentu
 - wykorzystaj typy proste/podstawowe, dokumenty zagnieżdżone, tablice itp.


```
js> db.createCollection('student')
{ ok: 1 }

js> db.createCollection('student')
{ ok: 1 }
```

```
{
  "name": {
    "firstName":
    "lastName":
  },
  "address": {
    "street": "",
    "city": "",
    "zipCode": ""
  },
  attendedCourses: [Course],
  grades: [Grade] // grade będzie posiadało id przedmiotu
}
```

- wprowadź kilka przykładowych dokumentów

```
js> db.getCollection('student').insertOne({"name": {"firstName": "Jakub", "lastName": "Szaredko"}, "address": {
  "street": "Krakowska 33", "city": "Kraków", "zipCode": "00-000"}, "attendedCourses": ["Databases"], "grades":
  [2]})
{
  acknowledged: true,
  insertedId: ObjectId("64773a8680880f75fa426a31")
}

js> db.getCollection('student').insertOne({"name": {"firstName": "Anna", "lastName": "Kowalska"}, "address": {
  "street": "Wrocławska 1", "city": "Kraków", "zipCode": "00-000"}, "attendedCourses": ["Databases"], "grades":
  [5]})
{
  acknowledged: true,
  insertedId: ObjectId("64773b0c80880f75fa426a33")
}
```

- przetestuj operacje wstawiania, modyfikacji/usuwania dokumentów

```
js> db.getCollection('student').deleteOne({ "_id": ObjectId("64773ae580880f75fa426a32") })
{ acknowledged: true, deletedCount: 1 }
```

```
js> db.getCollection('student').updateOne({"name": {"firstName": "Jakub", "lastName": "Szaredko"}}, {$set: { "
address": {"street": "Wiejska 333", "city": "Białystok", "zipCode": "01-000"}}})
Uncaught:
SyntaxError: Unexpected token, expected "," (1:184)

> 1 | db.getCollection('student').updateOne({"name": {"firstName": "Jakub", "lastName": "Szaredko"}}, {$set: {
    "address": {"street": "Wiejska 333", "city": "Białystok", "zipCode": "01-000"}}})
```

^ Gdzieś jest niepoprawna składnia, za mało czasu zostało, żeby to zweryfikować.

- przetestuj operacje wyszukiwania dokumentów

```
js> db.getCollection('student').findOne({"name": {"firstName": "Jakub", "lastName": "Szaredko"}})
{
  _id: ObjectId("64773a8680880f75fa426a31"),
  name: { firstName: 'Jakub', lastName: 'Szaredko' },
  address: { street: 'Krakowska 33', city: 'Kraków', zipCode: '00-000' },
  attendedCourses: [ 'Databases' ],
  grades: [ 2 ]
}
```

W sprawozdaniu należy umieścić zrzuty ekranów (z kodem poleceń oraz z uzyskanymi wynikami).
Dodatkowo należy dołączyć plik tekstowy (najlepiej z rozszerzeniem .js) zawierający kod poleceń

Punktacja za zadanie (razem 1pkt)