Bazy danych - NoSQL MongoDB - zadania

Imię i nazwisko: Błażej Kustra

Tydzień B, czwartek 12:50

Zadania:

- 1. Wykorzystując bazę danych **yelp dataset** wykonaj zapytanie i komendy MongoDB, aby uzyskać następujące rezultaty:
- a. Zwróć bez powtórzeń wszystkie nazwy miast w których znajdują się firmy (business). Wynik posortuj na podstawie nazwy miasta alfabetycznie.

db.yelpbusiness.distinct("city").sort()

```
/* 1 */
[
    "Ahwatukee",
    "Anthem",
    "Apache Junction",
    "Arcadia",
    "Atlanta",
    "Avondale",
    "Black Canyon City",
    "Bonnyrigg",
    "Boulder City",
    "Buckeye",
    "C Las Vegas",
    "Cambridge",
    "Carefree",
```

b. Zwróć liczbę wszystkich recenzji, które pojawiły się po 2011 roku (włącznie).

```
db.yelpreview.countDocuments({"date": { $gte: '2011-01-01'}})

① 1.55 sec.

880318
```

c. Zwróć dane wszystkich zamkniętych (*open*) firm (*business*) z pól: nazwa, adres, gwiazdki (*stars*).

d. Zwróć dane wszystkich użytkowników (*user*), którzy nie uzyskali ani jednego pozytywnego głosu z kategorii (*funny lub useful*), wynik posortuj alfabetycznie według imienia użytkownika.

e. Określ, ile każde przedsiębiorstwo otrzymało wskazówek/napiwków (tip) w 2012. Wynik posortuj alfabetycznie według liczby (tip).

f. Wyznacz, jaka średnia ocen (stars) uzyskała każda firma (business) na podstawie wszystkich recenzji. Wynik ogranicz do recenzji, które uzyskały min 4.0 gwiazdki.

g. Usuń wszystkie firmy (business), które posiadają ocenę (stars) równą 2.0.

```
db.yelpbusiness.deleteMany({stars:2.0})

@ 0.126 sec.

/* 1 */
{
    "acknowledged": true,
    "deletedCount": 1576.0
}
```

2. Zdefiniuj funkcje (*MongoDB*) umożliwiająca dodanie nowej recenzji (*review*). Wykonaj przykładowe wywołanie.

```
function add_review(user_id, review_id, text, business_id) {
    db.yelpreview.insert({
        votes: {
            funny: 0,
            useful : 0,
            cool : 0
        },
        user_id: user_id,
        review_id: review_id,
        stars: 0,
        date: new Date(),
        text: text,
        type: "review",
        business_id: business_id
    })
                                                                                           ( 0.008 sec.
Inserted 1 record(s) in 8ms
```

3. Zdefiniuj funkcje (MongoDB), która zwróci wszystkie biznesy (business), w których w kategorii znajduje się podana przez użytkownika cechę. Wartość kategorii należy przekazać do funkcji jako parametr. Wykonaj przykładowe wywołanie zdefiniowanej funkcji.

4. Zdefiniuj funkcje (*MongoDB*), która umożliwi modyfikacje nazwy użytkownika (*user*) na podstawie podanego id. Id oraz nazwa maja być przekazywane jako parametry.

```
function modify_user(user_id, new_name) {
   db.yelpuser.findOneAndUpdate({_id: new ObjectId(user_id) }, {$set: {name: new_name}})
}
```

5. Zwróć średnia ilość wszystkich wskazówek/napiwków dla każdego z biznesów, wykorzystaj map reduce.

```
var map = function() { emit(this.business_id, { count: 1 }) }
var reduce = function(key, values){
    var counter = 0
    values.forEach(function(val){ counter+=val.count })
    return counter
}

db.yelptip.mapReduce(map,reduce, {out: "number_of_tips" })
db.count.find()

**Tid": "--lemggGHgOGGIpd_RMb-g",
    "value": 6.0
}

**Jav'

*"_id": "--5jkZ3-nUPZxUvtcbr8Uw",
    "value": 16.0
}

**Jav'

*"_id": "--BlvYOO_RGZyElKu9XAl_g",
    "value": 21.0
}

**Jav'

*"_id": "--DlzrM_XO8GuYBomlg9zw",
    "value": 2.0
}
```

6. Odwzoruj wszystkie zadania z punktu 1 w Pythonie z pomocą API do MongoDB. Wykorzystaj dla każdego zadania odrębn<u>a</u> metodę.

```
import pymongo

myclient = pymongo.MongoClient("mongodb://localhost:27017/")
mydb = myclient["bazydanych"]

reviews = mydb["yelpreview"]
business = mydb["yelpbusiness"]
tips = mydb["yelptip"]
users = mydb["yelpuser"]
```

Połączenie pythona z mongodb.

a. Zwróć bez powtórzeń wszystkie nazwy miast w których znajdują się firmy (business). Wynik posortuj na podstawie nazwy miasta alfabetycznie.

```
def get_all_cities():
    return sorted(business.distinct("city"))

['Ahwatukee', 'Anthem', 'Apache Junction', 'Arcadia', 'Atlanta', 'Avondale', 'Black Canyon City', 'Bonnyrigg', 'Boulder City', 'Buckeye', 'C Las Vegas', 'Cambridge', 'Carefree', 'Casa Grande', 'Cave Creek', 'Centennial Hills', 'Central City Village', 'Central Henderson', 'Chandler', 'Chandler-Gilbert', 'City of Edinburgh', 'Clark County', 'Columbus', 'Coolidge', 'Cottage Grove', 'Cramond', 'Dalkeith', 'Dane', 'De Forest', 'DeForest', 'Deforest', 'Eagan', 'Edinburgh', 'El Mirage', 'Enterprise', 'Fitchburg', 'Florence', 'Fort Kinnaird', 'Fort McDowell', 'Fort Mcdowell', 'Fountain Hills', 'Fountain Hls', 'Gila Bend', 'Gilbert', 'Glendale', 'Glendale Az', 'Gold Canyon', 'Goldfield', 'Goodyear', 'Green Valley', 'Henderson (Green Valley)', 'Henderson of Green Valley', 'Henderson and Las vegas', 'Henderston', 'Higley', 'Inverkeithing', 'Juniper Green', 'Kitchener', 'Lake Las Vegas', 'Las Vegas', 'Las Vegas East', 'Las Vegas, NV 89147', 'Las Vegass', 'Las Vegas', 'Las Vegas', 'Las Vegas, NV 89147', 'Las Vegass', 'Las Vegas', 'Las Vegas', 'Las Vegas', 'Las Vegas', 'Norristown', 'Maricopa', 'Mc Farland', 'McFarland', 'McFarland', 'Mesa', 'Mesa', 'Midleton', 'Midlothian', 'Monona', 'Morristown', 'Muss elburgh', 'N E Las Vegas', 'N Las Vegas', 'N Las Vegas', 'Nellis Afb', 'North
```

b. Zwróć liczbę wszystkich recenzji, które pojawiły się po 2011 roku (włącznie).

```
def count_reviews_after_2011():
    return reviews.count_documents({"date": { "$gte": '2011-01-01'}})
blazejkustra@bk-3 untitled % python3 program.py
880318
```

c. Zwróć dane wszystkich zamkniętych (*open*) firm (*business*) z pól: nazwa, adres, gwiazdki (*stars*).

d. Zwróć dane wszystkich użytkowników (user), którzy nie uzyskali ani jednego pozytywnego głosu z kategorii (funny lub useful), wynik posortuj alfabetycznie według imienia użytkownika.

```
def get_all_users_with_no_votes():
    return users.find({"$or": [{"votes.funny": 0}, {"votes.useful": 0}]})

{'_id': ObjectId('5e859d6d1319bd6bb9796417'), 'yelping_since': '2013-10', 'votes': {'funny': 1, 'useful': 0, 'cool': 0},
    'review_count': 1, 'name': 'Anamika', 'user_id': 'CYKWsSZPMJYE6uXzydKGUA', 'friends': [], 'fans': 0, 'average_stars': 5.0
    ', 'type': 'user', 'compliments': {}, 'elite': []}

{'_id': ObjectId('5e859d6d1319bd6bb979641a'), 'yelping_since': '2013-07', 'votes': {'funny': 0, 'useful': 3, 'cool': 0},
    'review_count': 3, 'name': 'Jeanine', 'user_id': 'UEOCuejDwMmK4KyFx8Begw', 'friends': [], 'fans': 0, 'average_stars': 4.3
    3, 'type': 'user', 'compliments': {}, 'elite': []}

{'_id': ObjectId('5e859d6d1319bd6bb9796421b), 'yelping_since': '2009-12', 'votes': {'funny': 0, 'useful': 0, 'cool': 0},
    'review_count': 1, 'name': 'Ali', 'user_id': 'Sn_bI9-2uz2zAsSSbdfgXw', 'friends': [], 'fans': 0, 'average_stars': 5.0, 't
    ype': 'user', 'compliments': {}, 'elite': []}

{'_id': ObjectId('5e859d6d1319bd6bb9796421'), 'yelping_since': '2010-11', 'votes': {'funny': 0, 'useful': 1, 'cool': 0},
    'review_count': 3, 'name': 'Allister', 'user_id': 'ZRzhH3_33EkbimjsWYHRNw', 'friends': [], 'fans': 0, 'average_stars': 4.0, 'type': 'user', 'compliments': {}, 'elite': []}

{'_id': ObjectId('5e859d6d1319bd6bb9796423'), 'yelping_since': '2011-02', 'votes': {'funny': 0, 'useful': 2, 'cool': 1},
    'review_count': 6, 'name': 'Holly', 'user_id': 'dTRxrg[cYztvbbb_P_rv7g', 'friends': ['32_905]RUayEINgoRARaTQ'], 'fans': 1, 'average_stars': 4.17, 'type': 'user', 'compliments': {'more': 1}, 'elite': []}

{'_id': ObjectId('5e859d6d31319bd6bb9796424'), 'yelping_since': '2012-02', 'votes': {'funny': 0, 'useful': 0, 'cool': 0},
    'review_count': 1, 'name': 'Derik', 'user_id': 'CNUBWiJlmIhjN_2ipIauBg', 'friends': [], 'fans': 0, 'average_stars': 5.0,
    'type': 'user', 'compliments': {}, 'elite': []}
```

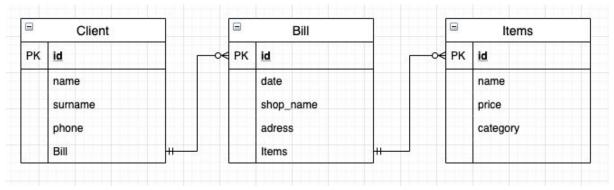
e. Określ, ile każde przedsiębiorstwo otrzymało wskazówek/napiwków (tip) w 2012. Wynik posortuj alfabetycznie według liczby (tip).

```
def list_business_with_their_tips():
      return tips.aggregate([
               {"$match": {"date": { "$regex": "2012" }}},
               {"$group": {"_id": "$business_id", "count": {"$sum": 1}}},
               {"$sort": {"count": 1}}
      1)
              '5EeGGEFbc2KlrBYhL80ciA',
'MlRBLmH9JwUMnJ8yvTpcOQ',
      id':
                                                         'count': 1}
                                                         'count': 1}
             'dHXoP7bZYJAv005-uAQadg',
'gxuVySgACHDqJlwme1FHLA',
    _id': 'dHXOP7bZYJAVOUS-LAQAGG',
_id': 'gxuVySgACHDQJlwmelFHLA',
_id': 'PtJdFzFitBa6rXQWx3v7_A',
_id': '0HDkCZtCvua-Wahio1-pBw',
_id': '2BGLuqsrqpT6t09M1Q19Q',
_id': 'U-tiHSh0P38H-6-yDUHrZA',
_id': '274bgOpS524g_tzcfaKYog',
_id': 'aPu5WZKYYro4Jhw0JJYAQQ',
_id': 'M7AMdpRPlrOol5JMYeg3dg',
_id': 'ZVWZaTHG5bB1A9ihtBD1ww',
                                                         'count':
                                                         'count':
                                                         'count':
                                                         'count':
              'ZVmZaTHC5bBlA9ihtBDlww',
             '4XuQ-oAsPQo5xwh-DPZovg',
'JtmzJ18FXNTBbWcXwroZuw',
                                                        'count':
       id': 'apGVTRZRCQ9-89hu2qW-vw',
id': 'RkZIVN-Wyii8he3o07GdgQ',
                                                        'count':
                                                                       1}
                                                        'count':
               'bNmsfLXaKKcDigEMidEcGg'
```

f. Wyznacz, jaka średnia ocen (stars) uzyskała każda firma (business) na podstawie wszystkich recenzji. Wynik ogranicz do recenzji, które uzyskały min 4.0 gwiazdki.

7. Zaproponuj bazę danych składająca się z 3 kolekcji pozwalająca przechowywać dane dotyczące: klientów, zakupu oraz przedmiotu zakupu. W bazie wykorzystaj: pola proste, złożone i tablice. Zaprezentuj strukturę dokumentów w formie JSON dla przykładowych danych. Uzasadnij swoją propozycję

a. Schemat:



b. Uzasadnienie:

Klient posiada swoje dane osobowe oraz numer telefonu wraz z wszystkimi rachunkami, rachunki posiadają datę zakupów, nazwa sklepu, adres sklepu oraz listę przedmiotów. Każdy przedmiot posiada nazwę, cenę i kategorie. Struktura jest prosta, oraz brak w niej redundancji.

c. Client:

```
function add_client(name, surname, phone) {
    db.clients.insert({
        name: name,
        surname: surname,
        phone: phone,
        bill: []
    })
}

add_client("Blazej", "Kustra", 734467906);
add_client("Marcin", "Zielony", 123456789);
add_client("Mateusz", "Monada", 987654321);
```

d. Bill:

```
function add_bill(shop_name,shop_address){
   db.bill.insert({
        date: new Date(),
        shop_name: shop_name,
        shop_address: shop_address,
        items: []
   })
}

add_bill("Žabka","Myczkowskiego 21");
add_bill("Auchan","Merna 99");
```

e. Items:

```
function add_item(name,price,category){
    db.items.insert({
        name: name,
        price: price,
        category: category
    })
}

add_item("Woda",2.0,"Napoje");
add_item("Czekolada",4.99,"Sfodycze");
add_item("Sok",3.19,"Napoje");
add_item("Szpinak",5.99,"Zdrowe");
add_item("Chleb",4.0,"Pieczywo");
```

f. Dodanie przedmiotu do rachunku:

```
function add_to_cart(bill_id,item_id){
    db.bill.update(
        {_id: new ObjectId(bill_id)},
        {$addToSet: { items: {$ref: "item", $id: new ObjectId(item_id)}}}
    );
}

add_to_cart("5e8a3b617ab353eb806386e9","5e8a3c5e7ab353eb806386f0");
add_to_cart("5e8a3b617ab353eb806386ea","5e8a3c5e7ab353eb806386f2");
add_to_cart("5e8a3b617ab353eb806386ea","5e8a3c5e7ab353eb806386f4");
```

g. Dodanie rachunku do klienta:

```
function add_cart_to_client(client_id, bill_id){
    db.clients.update(
        {_id: new ObjectId(client_id)},
        {$addToSet: { bill: {$ref: "bill", $id: new ObjectId(bill_id)}}}
    );
}

add_cart_to_client("5e8a3a457ab353eb806386e6","5e8a3b617ab353eb806386e9");
add_cart_to_client("5e8a3a457ab353eb806386e7","5e8a3b617ab353eb806386ea");
```

h. Baza danych:

