Facultatea de Automatică și Calculatoare
Master Tehnologii Informatice
Anul universitar 2023-2024

ADMINISTRAREA BAZELOR DE DATE
-MongoDB-

Student: Denisa-Valentina MANTA

Coordonator științific: Valer BOCAN

Sesiunea: Iunie 2024

1. Overview

The goal of the project is to assess the MongoDB database skills of the student. The project has a medium difficulty level and is relevant to industry employers of today.

2. Project description

You are required to make several statistical computations on some US Zips dataset using MongoDB as the database platform.

Prerequisites:

- Download the latest US Zips dataset from (choose the free tier). The dataset has approximately 33k entries.
- Create a MongoDB instance. You may use your own MongoDB Atlas instance in cloud or use a local instance. For local instances Docker is preferred, but you may also choose to install MongoDB as a standalone server on your OS.
- Import the dataset into the MongoDB instance.

Requirements:

a) Get the states with a total population of over 10 million.

b) Get the average city population by state.

c) Get the largest and the smallest city in each state.

d) Get the largest and the smallest counties in each state.

e) Get the nearest 10 zips from one of Chicago's landmarks, the Willis Tower situated at coordinates 41.878876, -87.635918.

f) Get the total population situated between 50 and 200 kms around New York's landmark, the Statue of Liberty at coordinates 40.689247, -74.044502.

Pentru baza de date am folosit MongoDB Compass. Mi-am denumit baza de date "abd_denisamanta" si am importat datele de aici https://simplemaps.com/data/us-zips.

CERINTA A:

Am folosit comanda:

db.abd_denisamanta.aggregate([{ "$group": {"_id": "$state_id","state_population": {"$sum": "$population"}} },{ "$match":{"state_population": {"$gt": 10000000}} }])

CERINTA B:

Am folosit comanda:

db.abd_denisamanta.aggregate([{ $group: { _id: { state_id: "$state_id", city:"$city" }, population: { $sum: "$population" } } }, { $group: { _id:"$_id.state_id", average_city_population: { $avg: "$population" } } }])

```
abd_denisamanta> db.abd_denisamanta.aggregate([{ $group: { _id: { state_id: "$state_id", city:"$city" }, population: { $sum: "$population" } } }, { $group: { _id:"$_id.stat
e_id", average_city_population: { $avg: "$population" } } }])
[
  { _id: 'VI', average_city_population: 0 },
  { _id: 'DC', average_city_population: 223529 },
  { _id: 'IA', average_city_population: 3444.1868250539956 },
  { _id: 'MN', average_city_population: 7349.458064516129 },
  { _id: 'DE', average_city_population: 18408.962962962964 },
  { _id: 'WV', average_city_population: 2514.178120617111 },
  { _id: 'CA', average_city_population: 32988.113998323555 },
  { _id: 'TX', average_city_population: 20887.64 },
  { _id: 'MO', average_city_population: 6846.150166852058 },
  { _id: 'NM', average_city_population: 6320.586826347306 },
  { _id: 'NJ', average_city_population: 17096.234750462107 },
  { _id: 'PA', average_city_population: 7915.422303473491 },
  { _id: 'OK', average_city_population: 6965.724561403509 },
  { _id: 'MA', average_city_population: 15083.846652267819 },
  { _id: 'NC', average_city_population: 14582.470752089137 },
  { _id: 'SD', average_city_population: 2449.1184573002756 },
  { _id: 'SC', average_city_population: 14735.67335243553 },
  { _id: 'MI', average_city_population: 11973.692857142858 },
  { _id: 'MD', average_city_population: 15480.547738693467 },
  { _id: 'WY', average_city_population: 3360.3313953488373 }
]
Type "it" for more
abd_denisamanta> _
```

CERINTA C:

Am folosit comanda:

db.abd_denisamanta.aggregate([{ $group: { _id: { state_id: "$state_id", city: "$city",

state_name: "$state_name" }, population: { $sum: "$population" } } }, { $sort: {

"_id.state_id": 1, "population": -1 } }, { $group: { _id: "$_id.state_id", largest_city:

{ $first: "$_id.city" }, largest_city_population: { $first: "$population" },

smallest_city: { $last: "$_id.city" }, smallest_city_population: { $last: "$population" }

} }, { $project: { _id: 0, state_id: "$_id", state_name: "$_id.state_name", largest_city:

1, largest_city_population: 1, smallest_city: 1, smallest_city_population: 1 } }])

```
Type "it" for more
... state_name: "$state_name" }, population: { $sum: "$population" } } }, { $sort: {
... "_id.state_id": 1, "population": -1 } }, { $group: { _id: "$_id.state_id", largest_city:
... { $first: "$_id.city" }, largest_city_population: { $first: "$population" },
... smallest_city: { $last: "$_id.city" }, smallest_city_population: { $last: "$population" }
... } }, { $project: { _id: 0, state_id: "$_id", state_name: "$_id.state_name", largest_city:
... 1, largest_city_population: 1, smallest_city: 1, smallest_city_population: 1 } }])
[
  {
    largest_city: 'Phoenix',
    largest_city_population: 1447391,
    smallest_city: 'Eden',
    smallest_city_population: 0,
    state_id: 'AZ'
  },
```

```
{
  largest_city: 'Fargo',
  largest_city_population: 129531,
  smallest_city: 'Absaraka',
  smallest_city_population: 0,
  state_id: 'ND'
},
{
  largest_city: 'Brooklyn',
  largest_city_population: 2679609,
  smallest_city: 'Rose',
  smallest_city_population: 0,
  state_id: 'NY'
},
{
  largest_city: 'Columbia',
  largest_city_population: 342092,
  smallest_city: 'Coosawhatchie',
  smallest_city_population: 0,
  state_id: 'SC'
},
```

```
{
  largest_city: 'Des Moines',
  largest_city_population: 221145,
  smallest_city: 'Ira',
  smallest_city_population: 0,
  state_id: 'IA'
},
{
  largest_city: 'Portland',
  largest_city_population: 934313,
  smallest_city: 'Post',
  smallest_city_population: 0,
  state_id: 'OR'
},
{
  largest_city: 'Sioux Falls',
  largest_city_population: 199595,
  smallest_city: 'Redig',
  smallest_city_population: 0,
  state_id: 'SD'
},
{
  largest_city: 'Honolulu',
  largest_city_population: 416884,
  smallest_city: 'Tripler Army Medical Center',
  smallest_city_population: 0,
  state_id: 'HI'
},
{
  largest_city: 'Columbus',
  largest_city_population: 871112,
  smallest_city: 'Hockingport',
  smallest_city_population: 0,
  state_id: 'OH'
},
```

```
},
{
  largest_city: 'Los Angeles',
  largest_city_population: 2406666,
  smallest_city: 'Mill Creek',
  smallest_city_population: 0,
  state_id: 'CA'
},
{
  largest_city: 'Indianapolis',
  largest_city_population: 978291,
  smallest_city: 'Smithville',
  smallest_city_population: 0,
  state_id: 'IN'
},
{
  largest_city: 'Burlington',
  largest_city_population: 44661,
  smallest_city: 'Northfield Falls',
  smallest_city_population: 0,
  state_id: 'VT'
},
{
  largest_city: 'Seattle',
  largest_city_population: 988217,
  smallest_city: 'Carrolls',
  smallest_city_population: 0,
  state_id: 'WA'
},
{
  largest_city: 'Miami',
  largest_city_population: 1855275,
  smallest_city: 'Mango',
  smallest_city_population: 0,
  state_id: 'FL'
},
```

```
{
  largest_city: 'Louisville',
  largest_city_population: 767460,
  smallest_city: 'Marshes Siding',
  smallest_city_population: 0,
  state_id: 'KY'
},
{
  largest_city: 'Baltimore',
  largest_city_population: 598831,
  smallest_city: 'Mount Victoria',
  smallest_city_population: 0,
  state_id: 'MD'
},
```

```
{
  largest_city: 'Billings',
  largest_city_population: 143280,
  smallest_city: 'Sumatra',
  smallest_city_population: 3,
  state_id: 'MT'
},
{
  largest_city: 'Wichita',
  largest_city_population: 423979,
  smallest_city: 'Treece',
  smallest_city_population: 0,
  state_id: 'KS'
},
{
  largest_city: 'Anchorage',
  largest_city_population: 242190,
  smallest_city: 'Meyers Chuck',
  smallest_city_population: 1,
  state_id: 'AK'
},
```

```
},
{
  largest_city: 'Newark',
  largest_city_population: 307483,
  smallest_city: 'Fort Monmouth',
  smallest_city_population: 0,
  state_id: 'NJ'
}
]
Type "it" for more
abd_denisamanta>
```

CERINTA D:

Am folosit comanda:

db.abd_denisamanta.aggregate([{ $group: { _id: { state_id: "$state_id", county_name: "$county_name", state_name: "$state_name" }, population: { $sum: "$population" } } }, { $sort: { "_id.state_id": 1, "population": -1 } }, { $group: { _id: "$_id.state_id", largest_county: { $first: "$_id.county_name" }, largest_county_population: { $first: "$population" }, smallest_county: { $last: "$_id.county_name" }, smallest_county_population: { $last: "$population" } } }, { $project: { _id: 0, state_id: "$_id", state_name: "$_id.state_name", largest_county: 1, largest_county_population: 1, smallest_county: 1, smallest_county_population: 1 } }])

```
Type "it" for more
... "$county_name", state_name: "$state_name" }, population: { $sum: "$population" } } }, {
... $sort: { "_id.state_id": 1, "population": -1 } }, { $group: { _id: "$_id.state_id",
... largest_county: { $first: "$_id.county_name" }, largest_county_population: { $first:
... "$population" }, smallest_county: { $last: "$_id.county_name" },
... smallest_county_population: { $last: "$population" } } }, { $project: { _id: 0,
... state_id: "$_id", state_name: "$_id.state_name", largest_county: 1,
... largest_county_population: 1, smallest_county: 1, smallest_county_population: 1 } }])
[
  {
    largest_county: 'Hinds',
    largest_county_population: 227234,
    smallest_county: 'Issaquena',
    smallest_county_population: 550,
    state_id: 'MS'
  },
  {
    largest_county: 'Cass',
    largest_county_population: 186640,
    smallest_county: 'Oliver',
    smallest_county_population: 1113,
    state_id: 'ND'
  },
  {
    largest_county: 'Greenville',
    largest_county_population: 588461,
    smallest_county: 'Allendale',
    smallest_county_population: 8356,
    state_id: 'SC'
  },
```

```
  },
  {
    largest_county: 'Kings',
    largest_county_population: 2679609,
    smallest_county: 'Hamilton',
    smallest_county_population: 8887,
    state_id: 'NY'
  },
  {
    largest_county: 'Minnehaha',
    largest_county_population: 204609,
    smallest_county: 'Sully',
    smallest_county_population: 942,
    state_id: 'SD'
  },
  {
    largest_county: 'Polk',
    largest_county_population: 489044,
    smallest_county: 'Adams',
    smallest_county_population: 3422,
    state_id: 'IA'
  },
  {
    largest_county: 'Honolulu',
    largest_county_population: 1010100,
    smallest_county: 'Kalawao',
    smallest_county_population: 50,
    state_id: 'HI'
  },
```

```
},
{
  largest_county: 'Franklin',
  largest_county_population: 1311502,
  smallest_county: 'Monroe',
  smallest_county_population: 11778,
  state_id: 'OH'
},
{
  largest_county: 'Los Angeles',
  largest_county_population: 9934597,
  smallest_county: 'Sierra',
  smallest_county_population: 3290,
  state_id: 'CA'
},
{
  largest_county: 'Multnomah',
  largest_county_population: 801225,
  smallest_county: 'Modoc',
  smallest_county_population: 345,
  state_id: 'OR'
},
{
  largest_county: 'Chittenden',
  largest_county_population: 170634,
  smallest_county: 'Essex',
  smallest_county_population: 6481,
  state_id: 'VT'
},
{
  largest_county: 'Marion',
  largest_county_population: 972652,
  smallest_county: 'Ohio',
  smallest_county_population: 5367,
  state_id: 'IN'
},
```

```
{
  largest_county: 'Miami-Dade',
  largest_county_population: 2688237,
  smallest_county: 'Liberty',
  smallest_county_population: 7704,
  state_id: 'FL'
},
{
  largest_county: 'Jefferson',
  largest_county_population: 784949,
  smallest_county: 'Robertson',
  smallest_county_population: 2071,
  state_id: 'KY'
},
```

```
{
  largest_county: 'Baltimore',
  largest_county_population: 1424394,
  smallest_county: 'Somerset',
  smallest_county_population: 22980,
  state_id: 'MD'
},
{
  largest_county: 'Yellowstone',
  largest_county_population: 165098,
  smallest_county: 'Petroleum',
  smallest_county_population: 356,
  state_id: 'MT'
},
{
  largest_county: 'Anchorage',
  largest_county_population: 278071,
  smallest_county: 'Bristol Bay',
  smallest_county_population: 471,
  state_id: 'AK'
},
```

```
{
  largest_county: 'Johnson',
  largest_county_population: 602436,
  smallest_county: 'Greeley',
  smallest_county_population: 1276,
  state_id: 'KS'
},
{
  largest_county: 'Bergen',
  largest_county_population: 953243,
  smallest_county: 'Salem',
  smallest_county_population: 66114,
  state_id: 'NJ'
},
{
  largest_county: 'Bernalillo',
  largest_county_population: 671497,
  smallest_county: 'Harding',
  smallest_county_population: 1590,
  state_id: 'NM'
}
]
Type "it" for more
abd_denisamanta>
```

CERINTA E:

Am folosit comenzile de mai jos, pe care le-am rulat in terminalul din MongoDb Compass. (de aici diferenta de imagini)

db.abd_denisamanta.updateMany({}, [{ "$set": { "location2dsphere.type":"Point", location2dsphere.coordinates": ["$lng", "$lat"] } }])

db.abd_denisamanta.createIndex({ location2dsphere: "2dsphere" })

db.abd_denisamanta.find({ location2dsphere: { $near: { $geometry: {

type: "Point", coordinates: [-87.635918, 41.878876] } } } }).limit(10)

```
>_MONGOSH

  }
> db.abd_denisamanta.createIndex({ location2dsphere: "2dsphere" })
< location2dsphere_2dsphere
> db.abd_denisamanta.find({ location2dsphere: { $near: { $geometry: {
  type: "Point", coordinates: [-87.635918, 41.878876] } } } }).limit(10)
< {
    _id: ObjectId('6658d7e231c61111e55a9697'),
    zip: 60606,
    lat: 41.88178,
    lng: -87.63724,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: true,
    population: 3513,
    density: 5844.8,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {
      '17031': 100
    },
```

```
    county_names_all: 'Cook',

    county_fips_all: 17031,

    imprecise: false,

    military: false,

    timezone: 'America/Chicago',

    location2dsphere: {

      type: 'Point',

      coordinates: [

        -87.63724,

        41.88178

      ]

    }

}
```
```
  {

    _id: ObjectId('6658d7e231c61111e55a9695'),

    zip: 60604,

    lat: 41.87799,

    lng: -87.62928,

    city: 'Chicago',

    state_id: 'IL',

    state_name: 'Illinois',

    zcta: true,

    population: 737,

    density: 3684.3,

    county_fips: 17031,

    county_name: 'Cook',

    county_weights: {

      '17031': 100

    },
```

```
    county_names_all: 'Cook',
    county_fips_all: 17031,
    imprecise: false,
    military: false,
    timezone: 'America/Chicago',
    location2dsphere: {
      type: 'Point',
      coordinates: [
        -87.62928,
        41.87799
      ]
    }
  }
```

```
  }
  {
    _id: ObjectId('6658d7e231c61111e55a9694'),
    zip: 60603,
    lat: 41.88018,
    lng: -87.6274,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: true,
    population: 1110,
    density: 3612.5,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {
      '17031': 100
    },
```

```
    county_names_all: 'Cook',
    county_fips_all: 17031,
    imprecise: false,
    military: false,
    timezone: 'America/Chicago',
    location2dsphere: {
      type: 'Point',
      coordinates: [
        -87.6274,
        41.88018
      ]
    }
  }
```

```
  {
    _id: ObjectId('6658d7e231c61111e55a9693'),
    zip: 60602,
    lat: 41.88304,
    lng: -87.62906,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: true,
    population: 938,
    density: 4394.9,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {
      '17031': 100
    },
```

```
    county_names_all: 'Cook',
    county_fips_all: 17031,
    imprecise: false,
    military: false,
    timezone: 'America/Chicago',
    location2dsphere: {
      type: 'Point',
      coordinates: [
        -87.62906,
        41.88304
      ]
    }
}
```

```
    {
      _id: ObjectId('6658d7e231c61111e55a96c9'),
      zip: 60661,
      lat: 41.88291,
      lng: -87.64409,
      city: 'Chicago',
      state_id: 'IL',
      state_name: 'Illinois',
      zcta: true,
      population: 11513,
      density: 14546.6,
      county_fips: 17031,
      county_name: 'Cook',
      county_weights: {
        '17031': 100
      },
```

```
      county_names_all: 'Cook',
      county_fips_all: 17031,
      imprecise: false,
      military: false,
      timezone: 'America/Chicago',
      location2dsphere: {
        type: 'Point',
        coordinates: [
          -87.64409,
          41.88291
        ]
      }
    }
```

```
{
    _id: ObjectId('6658d7e231c61111e55a9692'),
    zip: 60601,
    lat: 41.88527,
    lng: -87.62197,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: true,
    population: 16398,
    density: 16290.5,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {
      '17031': 100
    },
```

```
    county_names_all: 'Cook',
    county_fips_all: 17031,
    imprecise: false,
    military: false,
    timezone: 'America/Chicago',
    location2dsphere: {
      type: 'Point',
      coordinates: [
        -87.62197,
        41.88527
      ]
    }
  }
```

```
{
    _id: ObjectId('6658d7e231c61111e55a9698'),
    zip: 60607,
    lat: 41.87467,
    lng: -87.65175,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: true,
    population: 28797,
    density: 4806.3,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {
      '17031': 100
    },
```

```
  },
  county_names_all: 'Cook',
  county_fips_all: 17031,
  imprecise: false,
  military: false,
  timezone: 'America/Chicago',
  location2dsphere: {
    type: 'Point',
    coordinates: [
      -87.65175,
      41.87467
    ]
  }
}
```

```
{
  _id: ObjectId('6658d7e231c61111e55a96c3'),
  zip: 60654,
  lat: 41.89209,
  lng: -87.63673,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: true,
  population: 23475,
  density: 16729.1,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: {
    '17031': 100
  },
```

```
      county_names_all: 'Cook',
      county_fips_all: 17031,
      imprecise: false,
      military: false,
      timezone: 'America/Chicago',
      location2dsphere: {
        type: 'Point',
        coordinates: [
          -87.63673,
          41.89209
        ]
      }
    }
```

```
    }
    {
      _id: ObjectId('6658d7e231c61111e55a9696'),
      zip: 60605,
      lat: 41.86644,
      lng: -87.61956,
      city: 'Chicago',
      state_id: 'IL',
      state_name: 'Illinois',
      zcta: true,
      population: 34314,
      density: 9262.1,
      county_fips: 17031,
      county_name: 'Cook',
      county_weights: {
        '17031': 100
      },
```

```
    county_names_all: 'Cook',
    county_fips_all: 17031,
    imprecise: false,
    military: false,
    timezone: 'America/Chicago',
    location2dsphere: {
      type: 'Point',
      coordinates: [
        -87.61956,
        41.86644
      ]
    }
}
```

```
{
  _id: ObjectId('6658d7e231c61111e55a969c'),
  zip: 60611,
  lat: 41.89483,
  lng: -87.61944,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: true,
  population: 35247,
  density: 17163,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: {
    '17031': 100
  },
```

```
      county_names_all: 'Cook',
      county_fips_all: 17031,
      imprecise: false,
      military: false,
      timezone: 'America/Chicago',
      location2dsphere: {
        type: 'Point',
        coordinates: [
          -87.61944,
          41.89483
        ]
      }
    }
  abd_denisamanta >
```

CERINTA F:

Am folosit comanda de mai jos, pe care am rulat-o in terminalul din MongoDb Compass:

db.abd_denisamanta.aggregate([{ $geoNear: { near: { type: "Point", coordinates:

[-74.044502, 40.689247] }, key: "location2dsphere_2dsphere", distanceField:

"distance", maxDistance: 200000, minDistance: 50000, spherical: true } }, {

$group: { _id: null, total_population: { $sum: "$population" } } }])

```
> db.abd_denisamanta.aggregate([{ $geoNear: { near: { type: "Point", coordinates: [-74.044502, 40.689247] }, key: "location2dsphere", distanceField: "distance"
< {
    _id: null,
    total_population: 18089630
  }
abd_denisamanta >
```