

# PROIECT ADMINISTRAREA BAZELOR DE DATE

Student: Barbu Manuela Florentina  
Master TI, anul I  
Coordonator: BOCAN VALER

Security | Cloud: MongoDB Cl... | Curs: Administrarea bazelor de... | cv.upt.ro/course/view.php?id=439

Command Prompt

```
C:\Users>cd..
C:\>mongoimport --db us_zips --collection us_zips --type csv --headerline --file C:\zips\uszips.csv
2024-05-25T19:32:31.010+0300    connected to: mongodb://localhost/
2024-05-25T19:32:31.924+0300    33787 document(s) imported successfully. 0 document(s) failed to import.
```

Prerequisites:

- Download the latest US Zips dataset from <https://simplemaps.com/data/us-zips> (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:

- Get the states with a total population of over 10 million.
- Get the average city population by state.
- Get the largest and the smallest city in each state.
- Get the largest and the smallest counties in each state.
- Get the nearest 10 zips from one of Chicago's landmarks, the Willis Tower situated at coordinates 41.878876, -87.635918.
- 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.

Notes:

- Create the indexes you deem relevant for your collection. You will be asked on the performance of your indexes so be prepared to defend your choice, preferably by analyzing the execution statistics.
- For requirements e) and f), you may add a geo field to your collection in order to leverage geospatial query operators.
- Your solution must be original so please don't rely on cheating.

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

```
us_zips> db.us_zips.createIndex({ population: 1 })
us_zips> db.us_zips.aggregate([
  {
    $group:{_id:"$state_name",total_population:{$sum:"$population"}}
  },
  {
    $match:{total_population:{$gt: 10000000}}
  }
])
```

Security | Cloud: MongoDB Cl... | Curs: Administrarea bazelor de... | Curs: Administrarea bazelor de...

cv.upt.ro/course/view.php?id=439

Campus Virtual™ | UPT | Cel... | Meniul meu | ZOOM | Cursurile mele

Română (ro) | Barbu Manuela-Florentina

## 2. Project description

```
mongosh mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000
us_zips> db.us_zips.createIndex({ population: 1 })
us_zips> db.us_zips.aggregate([{$group: { _id: "$state_name", total_population: { $sum: "$population" } } }, {$match: {total_population:
{ $gt: 10000000 } } } ] )
[
  { _id: 'Ohio', total_population: 11774683 },
  { _id: 'California', total_population: 39364820 },
  { _id: 'Texas', total_population: 29242096 },
  { _id: 'New York', total_population: 19994379 },
  { _id: 'Florida', total_population: 21632200 },
  { _id: 'Pennsylvania', total_population: 12989208 },
  { _id: 'North Carolina', total_population: 10470214 },
  { _id: 'Illinois', total_population: 12757583 },
  { _id: 'Michigan', total_population: 10057902 },
  { _id: 'Georgia', total_population: 10722352 }
]
us_zips>
```

oDB as the database platform.

the free tier). The dataset has approximately 33k

se a local instance. For local instances Docker

OS.

ates 41.878876, -87.635918.

Statue of Liberty at coordinates 40.689247,

ance of your indexes so be prepared to

defend your choice, preferably by analyzing the execution statistics.

- For requirements e) and f), you may add a geo field to your collection in order to leverage geospatial query operators.
- Your solution must be original so please don't rely on cheating.

## 3. Scoring:

Type here to search

70°F Mostly sunny 3:08 PM 5/26/2024

## b) Get the average city population by state

```
us_zips> db.us_zips.createIndex({ state_name: 1, city: 1 })
```

```
us_zips> db.us_zips.aggregate([{$group: { _id: {state: "$state_name", city: "$city"},
population: { $sum: "$population" } } }, {$group: { _id: "$_id.state", avg_city_population:
{ $avg: "$population" } } } ] )
```

Security | Cloud: MongoDB Cl... | Curs: Administrarea bazelor de... | Curs: Administrarea bazelor de...

cv.upt.ro/course/view.php?id=439

Campus Virtual™ | UPT | Cel... | Meniul meu | ZOOM | Cursurile mele

Română (ro) | Barbu Manuela-Florentina

## 2. Project description

```
mongosh mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000
us_zips> db.us_zips.createIndex({ state_name: 1, city: 1 })
state_name_1.city_1
us_zips> db.us_zips.aggregate([{$group: { _id: {state: "$state_name", city: "$city"}, population: { $sum: "$population" } } }, {$group: { _id: "$_id.state",
avg_city_population: { $avg: "$population" } } } ] )
[
  { _id: 'Colorado', avg_city_population: 14835.881748071979 },
  { _id: 'Missouri', avg_city_population: 6846.150166852058 },
  { _id: 'Guam', avg_city_population: 0 },
  { _id: 'Maryland', avg_city_population: 15480.547738693467 },
  { _id: 'New Hampshire', avg_city_population: 5821.426160337553 },
  { _id: 'Maine', avg_city_population: 3239.054502369668 },
  { _id: 'Nebraska', avg_city_population: 3725.74144486692 },
  { _id: 'Minnesota', avg_city_population: 7349.458064516129 },
  { _id: 'Montana', avg_city_population: 3085.864406779661 },
  { _id: 'Ohio', avg_city_population: 11635.062252964426 },
  { _id: 'South Dakota', avg_city_population: 2449.1134572002756 },
  { _id: 'Alabama', avg_city_population: 8468.883239171375 },
  { _id: 'Indiana', avg_city_population: 9789.903318903318 },
  { _id: 'Massachusetts', avg_city_population: 15083.846652267819 },
  { _id: 'Louisiana', avg_city_population: 10357.174107142857 },
  { _id: 'Idaho', avg_city_population: 7270.113725490196 },
  { _id: 'South Carolina', avg_city_population: 14735.67335243553 },
  { _id: 'District of Columbia', avg_city_population: 223529 },
  { _id: 'Virgin Islands', avg_city_population: 0 },
  { _id: 'Northern Mariana Islands', avg_city_population: 0 }
]
Type "it" for more
```

oDB as the database platform.

the free tier). The dataset has approximately 33k

se a local instance. For local instances Docker

OS.

ates 41.878876, -87.635918.

Statue of Liberty at coordinates 40.689247,

ance of your indexes so be prepared to

defend your choice, preferably by analyzing the execution statistics.

- For requirements e) and f), you may add a geo field to your collection in order to leverage geospatial query operators.
- Your solution must be original so please don't rely on cheating.

## 3. Scoring:

Type here to search

Video highlight 3:14 PM 5/26/2024

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

- Se sortează înregistrările aferente fiecărui stat după populație în ordine descrescătoare
- Se grupează înregistrările după stat și se folosește \$first și \$last ca să se obțină cel mai mare și cel mai mic oraș

```
us_zips> db.us_zips.aggregate([
{
  $sort:{state_name:1,population:-1}
},
{
  $group:{
    _id: "$state_name",
    largest_city:{$first:"$city" },
    largest_city_population:{$first:"$population"},
    smallest_city:{$last:"$city"},
    smallest_city_population:{$last:"$population"}
  }
}
])
```

The screenshot shows a web browser window with a URL bar displaying 'cv.upt.ro/course/view.php?id=439'. The page content includes a '2. Project description' section with a terminal window showing a MongoDB aggregation query and its results. The query is:

```
us_zips> db.us_zips.createIndex({ state_name: 1, county: 1 })
us_zips>
us_zips> db.us_zips.aggregate([{$sort:{state_name:1,population:-1}},{$group:{_id:'$state_name',largest_city:{$first:'$city'},largest_city_population:{$first:'$population'},smallest_city:{$last:'$city'},smallest_city_population:{$last:'$population'}}]])
```

The results are:

```
[
  {
    _id: 'District of Columbia',
    largest_city: 'Washington',
    largest_city_population: 70043,
    smallest_city: 'Washington',
    smallest_city_population: 0
  },
  {
    _id: 'Missouri',
    largest_city: 'Saint Peters',
    largest_city_population: 74967,
    smallest_city: 'Springfield',
    smallest_city_population: 0
  },
  {
    _id: 'Guam',
    largest_city: 'Barrigada',
    largest_city_population: '',
    smallest_city: 'Vigo',
    smallest_city_population: ''
  },
  {
    _id: 'Nebraska',
    largest_city: 'Lincoln',
    largest_city_population: 46714,
    smallest_city: 'Beatrice',
    smallest_city_population: 0
  }
]
```

The page also includes a '3. Scoring' section with the following text:

defend your choice, preferably by analyzing the execution statistics.

- For requirements e) and f), you may add a geo field to your collection in order to leverage geospatial query operators.
- Your solution must be original so please don't rely on cheating.

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

- Se sortează înregistrările din fiecare stat după populație în ordine descrescătoare - Se grupează înregistrările după stat și utilizăm \$first și \$last pentru a obține cel mai mare și cel mai mic oraș
- Se grupează după stat și județ pentru a calcula totalul populației pe fiecare județ
- Se sortează rezultatele după stat și populație în ordine descrescătoare
- Se folosește \$first și \$last pentru a determina cel mai mare și cel mai mic județ

```
us_zips> db.us_zips.aggregate([
{
  $sort:{state_name:1,population:-1}
},
{
  $group:{_id:{state:"$state_name",county:"$county"},total_population:{$sum:
    "$population"}}
},
{
  $sort:{"_id.state":1,total_population:-1}
},
{
  $group: {_id: "$_id.state",
    largest_county:{$first:"$_id.county"},
    largest_county_population:{$first:"$total_population"},
    smallest_county:{$last:"$_id.county"},
    smallest_county_population:{$last:"$total_population"}
  }
}
])
```

Campus Virtual

UPT

CeL

Meniu meu

ZOOM

Cursurile mele

Română (ro)

Barbu Manuela-Florentina

## 2. Project description

```
mongosh mongo://127.0.0.1:27017?directConnection=true&serverSelectionTimeoutMS=2000
us_zips> db.us_zips.aggregate([{$sort: { state_name:1,population:-1}},{$group:{_id:{state:"$state_name",county:"$county"},total_population:{$sum: "$population"}}},{$sort: {"_id.state":1,total_population:-1}},{$group:{_id:"$_id.state",largest_county:{$first:"$_id.county"},largest_county_population:{$first:"$total_population"},smallest_county:{$last:"$_id.county"},smallest_county_population:{$last:"$total_population"}}])
[
  {
    _id: 'Montana',
    largest_county: null,
    largest_county_population: 1092396,
    smallest_county: null,
    smallest_county_population: 1092396
  },
  {
    _id: 'Pennsylvania',
    largest_county: null,
    largest_county_population: 12989208,
    smallest_county: null,
    smallest_county_population: 12989208
  },
  {
    _id: 'Oklahoma',
    largest_county: null,
    largest_county_population: 3970463,
    smallest_county: null,
    smallest_county_population: 3970463
  },
  {
    _id: 'South Dakota',
    largest_county: null,
    largest_county_population: 889030,
    smallest_county: null,
    smallest_county_population: 889030
  }
]
```

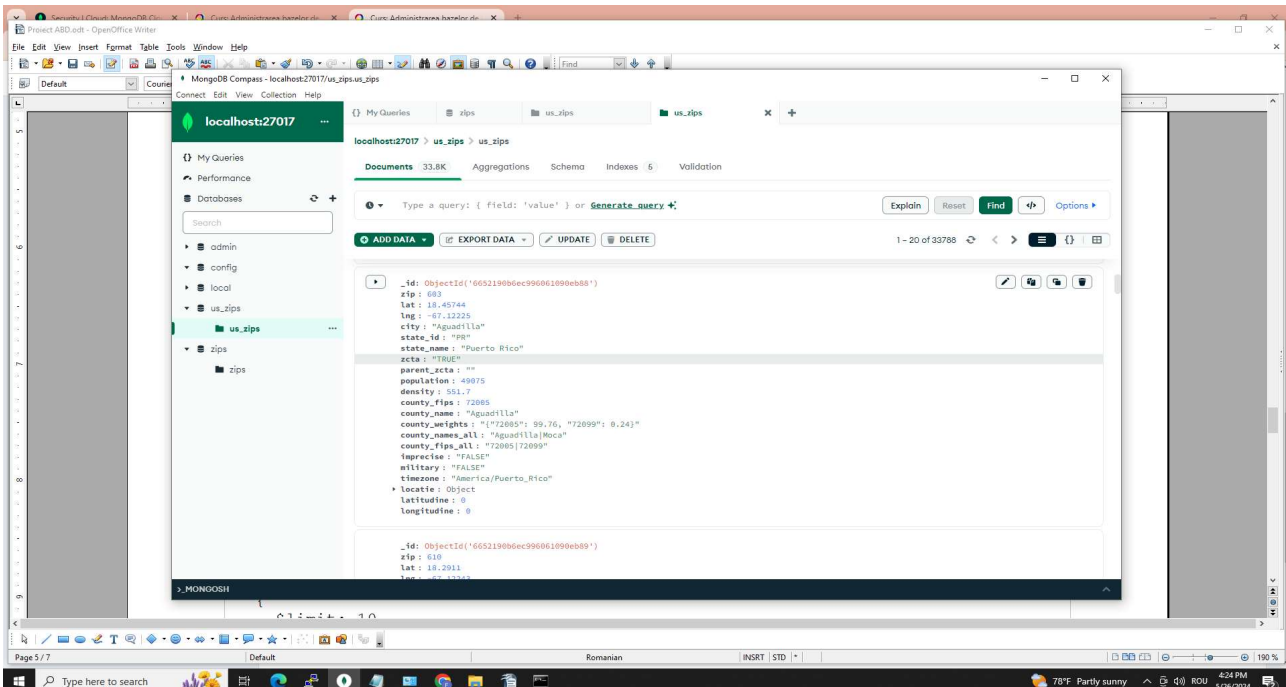
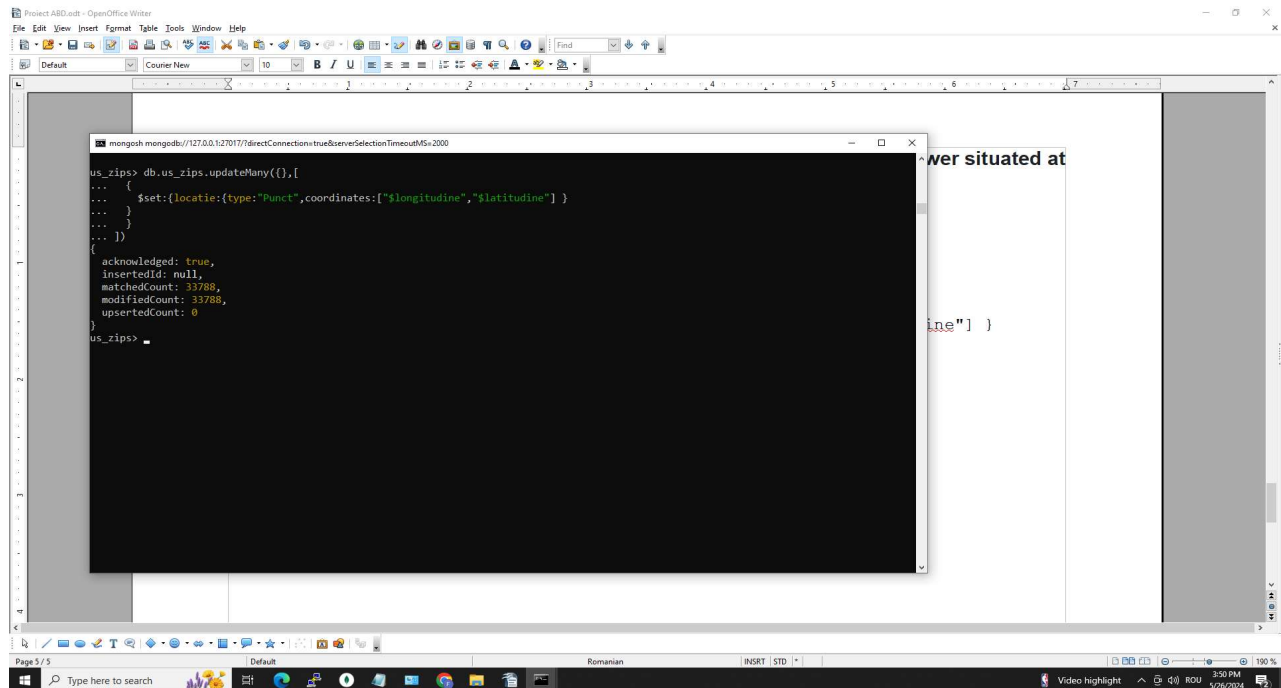
dened your choice, preferably by analyzing the execution statistics.

- For requirements e) and f), you may add a geo field to your collection in order to leverage geospatial query operators.
- Your solution must be original so please don't rely on cheating.

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

- Adăugare câmp geospațial

```
us_zips> db.us_zips.updateMany({}, [
  {
    $set: {locatie: {type: "Point", coordinates: ["$longitude", "$latitude"] }
  }
})
```



- Crearea unui index geospațial

```
us_zips> db.us_zips.createIndex({location:"2dsphere"})
```

- Identificarea celor mai apropiate 10 locații de Willis Tower

```
us_zips> db.us_zips.aggregate([
{
  $geoNear:{
    near:{type:"Point",coordinates:[-87.635918,41.878876]},
    distanceField: "dist.calculated",
    spherical:true,
    key:"locatie"
  }
},
{
  $limit:10
}
]).pretty();
```

The screenshot shows a web browser window with a URL bar containing 'cv.upt.ro/course/view.php?id=439'. The page header includes 'Campus Virtual', 'UPT', 'Cel', 'Meniul meu', 'ZOOM', and 'Cursurile mele'. The main content area is titled 'Prerequisites:' and contains a terminal window showing the following command and its output:

```
us_zips> db.us_zips.aggregate([
...
  $geoNear: {
    near: { type: "Point", coordinates: [-87.635918, 41.878876] },
    distanceField: "dist.calculated",
    spherical: true,
    key: "locatie" // Specifică numele câmpului indexului geospațial pe care vrei să-l folosești
  },
...
},
{
  $limit: 10
}
]).pretty();
```

The output shows a single document with the following fields:

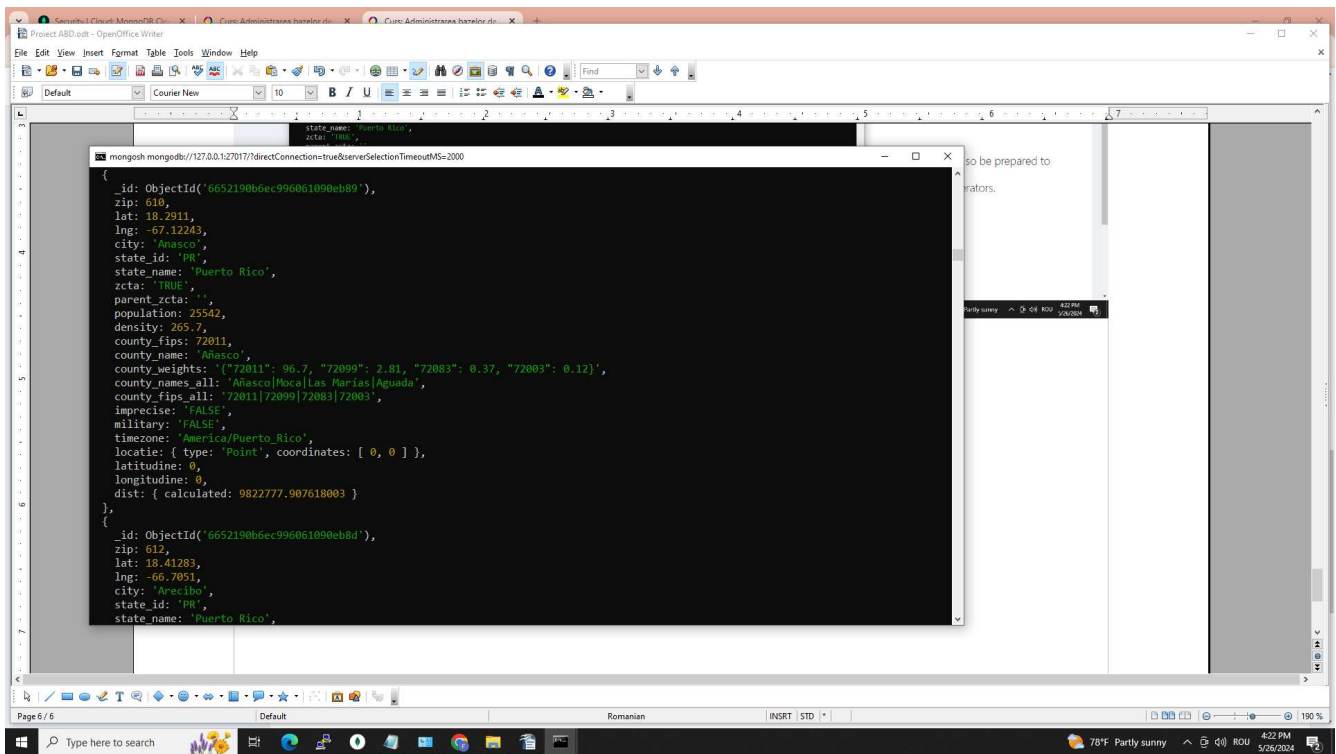
```
{
  _id: ObjectId('6652190b6ec996061090eb87'),
  zip: 602,
  lat: 18.36075,
  lng: -67.17541,
  city: 'Aguada',
  state_id: 'PR',
  state_name: 'Puerto Rico',
  zcta: 'TRUE',
  parent_zcta: '',
  population: 37642,
  density: 479.2,
  county_fips: 72003,
  county_name: 'Aguada',
  county_weights: '{"72003": 100}',
  county_names_all: 'Aguada',
  county_fips_all: 72003,
  imprecise: 'FALSE',
  military: 'FALSE',
}
```

Below the terminal window, there is a section titled '3. Scoring:' with the following requirements:

- Requirement a) - 1 point
- Requirement b) - 1.5 points
- Requirement c) - 1.5 points

The browser window also shows a sidebar with a search bar and a list of documents. The bottom of the browser window shows a Windows taskbar with the date and time '78°F Partly sunny 4:22 PM 5/26/2024'.





**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**

```
db.us_zip.aggregate([
  {
    $geoNear:{
      near:{type:"Point",coordinates:[-74.044502,40.689247]},
      distanceField:"distance",
      maxDistance:200000,
      minDistance:50000,
      spherical:true,
      query:{population:{$exists:true}},
      key:"locatie_2dsphere"
    }
  },
  {
    $group:{
      _id:null,
      totalPopulation:{$sum:"$population"}
    }
  }
]);
```

Curs: Administrarea bazelor de date

cv.upt.ro/course/view.php?id=439

Campus Virtual

UPT

Cel

Meniul meu

ZOOM

Cursurile mele

Română (ro)

Barbu Manuela-Florentina

D) Get the average city population by state.

```
us_zips>
{
  $geoNear: {
    near: { type: "Point", coordinates: [-74.044582, 40.689247] },
    distanceField: "distance",
    maxDistance: 200000,
    minDistance: 50000,
    spherical: true,
    query: { population: { $exists: true } }
  },
  $group: {
    _id: null,
    totalPopulation: { $sum: "$population" }
  }
}
us_zips>
```

es 41.878876, -87.635918.

atue of Liberty at coordinates 40.689247,

nce of your indexes so be prepared to

eospatial query operators.

4. Solution Delivery

Upload your final project solution to the designated area in the virtual campus.

Type here to search

Video highlight

ENG US 7:56 PM 5/31/2024