

CUNY MSDA - IS607 : DATA ACQUISITION AND MANAGEMENT

Week 10 Quiz

A. Leopold HILLAH

i. Import zip data

```
C:\MongoDB\bin>mongoimport --collection zips --file
C:\MongoDB\input\zips.json
connected to: 127.0.0.1
2014-11-02T11:02:36.789+0100 check 9 29353
2014-11-02T11:02:36.807+0100 imported 29353 objects
```

ii. Exploring loaded data

```
> show dbs
admin          (empty)
employment     0.078GB
local          0.078GB
test           0.078GB
>
> use test
switched to db test
>
> show collections
system.indexes
zips
>
```

```
> db.zips.find()

{ "_id" : "01001", "city" : "AGAWAM", "loc" : [ -72.622739, 42.070206 ], "pop" : 15338, "state" : "MA" }
{ "_id" : "01002", "city" : "CUSHMAN", "loc" : [ -72.51565, 42.377017 ], "pop" : 36963, "state" : "MA" }
{ "_id" : "01005", "city" : "BARRE", "loc" : [ -72.108354, 42.409698 ], "pop" : 4546, "state" : "MA" }
...
Type "it" for more
>
```

1. States with populations less than eight million

```
db.zips.aggregate ( [{ $group :  
  { _id : "$state",  
    TotalPop : { $sum : "$pop" } } },  
  { $match : { TotalPop : { $lt : 8*1000*1000 } } },  
  { $sort : { TotalPop : -1 } } ] )
```

```
> db.zips.aggregate ( [{ $group :  
... { _id : "$state",  
... TotalPop : { $sum : "$pop" } } },  
... { $match : { TotalPop : { $lt : 8*1000*1000 } } },  
... { $sort : { TotalPop : -1 } } ] )  
{ "_id" : "NJ", "TotalPop" : 7730188 }  
{ "_id" : "NC", "TotalPop" : 6628637 }  
{ "_id" : "GA", "TotalPop" : 6478216 }  
{ "_id" : "VA", "TotalPop" : 6181479 }  
{ "_id" : "MA", "TotalPop" : 6016425 }  
{ "_id" : "IN", "TotalPop" : 5544136 }  
{ "_id" : "MO", "TotalPop" : 5110648 }  
{ "_id" : "WI", "TotalPop" : 4891769 }  
{ "_id" : "TN", "TotalPop" : 4876457 }  
{ "_id" : "WA", "TotalPop" : 4866692 }  
{ "_id" : "MD", "TotalPop" : 4781379 }  
{ "_id" : "MN", "TotalPop" : 4372982 }  
{ "_id" : "LA", "TotalPop" : 4217595 }  
{ "_id" : "AL", "TotalPop" : 4040587 }  
{ "_id" : "KY", "TotalPop" : 3675484 }  
{ "_id" : "AZ", "TotalPop" : 3665228 }  
{ "_id" : "SC", "TotalPop" : 3486703 }  
{ "_id" : "CO", "TotalPop" : 3293755 }  
{ "_id" : "CT", "TotalPop" : 3287116 }  
{ "_id" : "OK", "TotalPop" : 3145585 }  
Type "it" for more  
>
```

2. Fifth largest city in New York

```
db.zips.aggregate( [  
  { $match : { state : "NY"} },  
  { $group : { _id : {city : "$city"} ,  
    TotalPop : { $sum : "$pop" } } },  
  { $sort : { TotalPop : -1 } },  
  { $limit : 5 },  
  { $skip : 4} ] )
```

```
> db.zips.aggregate( [  
... { $match : { state : "NY"} },  
... { $group : { _id : {city : "$city"} ,  
... TotalPop : { $sum : "$pop" } } },  
... { $sort : { TotalPop : -1 } },  
... { $limit : 5 },  
... { $skip : 4}  
... ] )  
{ "_id" : { "city" : "STATEN ISLAND" }, "TotalPop" : 378977 }  
>
```

3. Total number of cities in each state

```
db.zips.aggregate ( [  
  { $group : { _id : {state : "$state", city : "$city"} } },  
  { $group : { _id : "$_id.state" ,  
    CityCount : { $sum : 1 } } },  
  { $sort : { _id : 1 } }  
)
```

```
> db.zips.aggregate( [  
... { $group : { _id : {state : "$state", city : "$city"} } },  
... { $group : { _id : "$_id.state" ,  
... CityCount : { $sum : 1 } } },  
... { $sort : { _id : 1 } }  
... ] )  
{ "_id" : "AK", "CityCount" : 183 }  
{ "_id" : "AL", "CityCount" : 511 }  
{ "_id" : "AR", "CityCount" : 563 }  
{ "_id" : "AZ", "CityCount" : 178 }  
{ "_id" : "CA", "CityCount" : 1072 }  
{ "_id" : "CO", "CityCount" : 330 }  
{ "_id" : "CT", "CityCount" : 224 }  
{ "_id" : "DC", "CityCount" : 2 }  
{ "_id" : "DE", "CityCount" : 46 }  
{ "_id" : "FL", "CityCount" : 463 }  
{ "_id" : "GA", "CityCount" : 561 }  
{ "_id" : "HI", "CityCount" : 70 }  
{ "_id" : "IA", "CityCount" : 889 }  
{ "_id" : "ID", "CityCount" : 233 }  
{ "_id" : "IL", "CityCount" : 1148 }  
{ "_id" : "IN", "CityCount" : 598 }  
{ "_id" : "KS", "CityCount" : 648 }  
{ "_id" : "KY", "CityCount" : 771 }  
{ "_id" : "LA", "CityCount" : 403 }  
{ "_id" : "MA", "CityCount" : 405 }  
Type "it" for more  
>
```