**MongoDB -Aggregation Exercises**

Import the zips.json file into your MongoDB. Database name is "population" and collection name is "zipcodes".

mongoimport --db population --collection zipcodes --file zips.json

**Soltions: commands are in purple colour and results of the query is in green colour.**

**Atlas atlas-68i9f5-shard-0 [primary] myFirstDatabase> use population**

switched to db population

**Atlas atlas-68i9f5-shard-0 [primary] population> show collections**

zipcodes

**Atlanta Population**

1. use db.zipcodes.find() to filter results to only the results where city is ATLANTA and state is GA.

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.find({$and:[{city:"ATLANTA"},{state:"GA"}]})**

[

{

\_id: '30303',

city: 'ATLANTA',

loc: [ -84.388846, 33.752504 ],

pop: 1845,

state: 'GA'

},

{

\_id: '30305',

city: 'ATLANTA',

loc: [ -84.385145, 33.831963 ],

pop: 19122,

state: 'GA'

},

{

\_id: '30306',

city: 'ATLANTA',

loc: [ -84.351418, 33.786027 ],

pop: 20081,

state: 'GA'

},

{

\_id: '30307',

city: 'ATLANTA',

loc: [ -84.335957, 33.769138 ],

pop: 16330,

state: 'GA'

},

{

\_id: '30308',

city: 'ATLANTA',

loc: [ -84.375744, 33.771839 ],

pop: 8549,

state: 'GA'

},

{

\_id: '30309',

city: 'ATLANTA',

loc: [ -84.388338, 33.798407 ],

pop: 14766,

state: 'GA'

}

]

Type "it" for more

->showing few of the results from find query.

2. use db.zipcodes.aggregate with $match to do the same as above.

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$match:{$and:[{city:"ATLANTA"},{state:"GA"}]}})**

[

{

\_id: '30303',

city: 'ATLANTA',

loc: [ -84.388846, 33.752504 ],

pop: 1845,

state: 'GA'

},

{

\_id: '30305',

city: 'ATLANTA',

loc: [ -84.385145, 33.831963 ],

pop: 19122,

state: 'GA'

},

{

\_id: '30306',

city: 'ATLANTA',

loc: [ -84.351418, 33.786027 ],

pop: 20081,

state: 'GA'

},

{

\_id: '30307',

city: 'ATLANTA',

loc: [ -84.335957, 33.769138 ],

pop: 16330,

state: 'GA'

},

{

\_id: '30308',

city: 'ATLANTA',

loc: [ -84.375744, 33.771839 ],

pop: 8549,

state: 'GA'

},

{

\_id: '30309',

city: 'ATLANTA',

loc: [ -84.388338, 33.798407 ],

pop: 14766,

state: 'GA'

}

]

Type "it" for more

->showing few of the results from find query.

3. use $group to count the number of zip codes in Atlanta.

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate([ {$group: {\_id:{city:"$city"},total:{$sum:1}}},{$match: {"\_id.city": "ATLANTA"}}])**

[ { \_id: { city: 'ATLANTA' }, total: 41 } ]

4. use $group to find the total population in Atlanta.

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate([ {$group: {\_id:{city:"$city"},total:{$sum:"$pop"}}},{$match: {"\_id.city": "ATLANTA"}}])**

[ { \_id: { city: 'ATLANTA' }, total: 630046 } ]

**Populations By State**

1. use aggregate to calculate the total population for each state

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}})**

[

{ \_id: { state: 'DC' }, ItsTotalPopulation: 606900 },

{ \_id: { state: 'SD' }, ItsTotalPopulation: 695397 },

{ \_id: { state: 'LA' }, ItsTotalPopulation: 4217595 },

{ \_id: { state: 'NV' }, ItsTotalPopulation: 1201833 },

{ \_id: { state: 'OR' }, ItsTotalPopulation: 2842321 },

{ \_id: { state: 'VA' }, ItsTotalPopulation: 6181479 },

{ \_id: { state: 'IA' }, ItsTotalPopulation: 2776420 },

{ \_id: { state: 'TX' }, ItsTotalPopulation: 16984601 },

{ \_id: { state: 'VT' }, ItsTotalPopulation: 562758 },

{ \_id: { state: 'DE' }, ItsTotalPopulation: 666168 },

{ \_id: { state: 'MS' }, ItsTotalPopulation: 2573216 },

{ \_id: { state: 'ME' }, ItsTotalPopulation: 1226648 },

{ \_id: { state: 'FL' }, ItsTotalPopulation: 12686644 },

{ \_id: { state: 'ID' }, ItsTotalPopulation: 1006749 },

{ \_id: { state: 'MA' }, ItsTotalPopulation: 6016425 },

{ \_id: { state: 'MN' }, ItsTotalPopulation: 4372982 },

{ \_id: { state: 'AR' }, ItsTotalPopulation: 2350725 },

{ \_id: { state: 'IL' }, ItsTotalPopulation: 11427576 },

{ \_id: { state: 'WV' }, ItsTotalPopulation: 1793146 },

{ \_id: { state: 'ND' }, ItsTotalPopulation: 638272 }

]

Type "it" for more

2. sort the results by population, highest first

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}},{$sort:{ItsTotalPopulation:-1}})**

[

{ \_id: { state: 'CA' }, ItsTotalPopulation: 29754890 },

{ \_id: { state: 'NY' }, ItsTotalPopulation: 17990402 },

{ \_id: { state: 'TX' }, ItsTotalPopulation: 16984601 },

{ \_id: { state: 'FL' }, ItsTotalPopulation: 12686644 },

{ \_id: { state: 'PA' }, ItsTotalPopulation: 11881643 },

{ \_id: { state: 'IL' }, ItsTotalPopulation: 11427576 },

{ \_id: { state: 'OH' }, ItsTotalPopulation: 10846517 },

{ \_id: { state: 'MI' }, ItsTotalPopulation: 9295297 },

{ \_id: { state: 'NJ' }, ItsTotalPopulation: 7730188 },

{ \_id: { state: 'NC' }, ItsTotalPopulation: 6628637 },

{ \_id: { state: 'GA' }, ItsTotalPopulation: 6478216 },

{ \_id: { state: 'VA' }, ItsTotalPopulation: 6181479 },

{ \_id: { state: 'MA' }, ItsTotalPopulation: 6016425 },

{ \_id: { state: 'IN' }, ItsTotalPopulation: 5544136 },

{ \_id: { state: 'MO' }, ItsTotalPopulation: 5110648 },

{ \_id: { state: 'WI' }, ItsTotalPopulation: 4891769 },

{ \_id: { state: 'TN' }, ItsTotalPopulation: 4876457 },

{ \_id: { state: 'WA' }, ItsTotalPopulation: 4866692 },

{ \_id: { state: 'MD' }, ItsTotalPopulation: 4781379 },

{ \_id: { state: 'MN' }, ItsTotalPopulation: 4372982 }

]

Type "it" for more

3. limit the results to just the first 3 results. What are the top 3 states in population?

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}},{$sort:{ItsTotalPopulation:-1}},{$limit:3})**

[

{ \_id: { state: 'CA' }, ItsTotalPopulation: 29754890 },

{ \_id: { state: 'NY' }, ItsTotalPopulation: 17990402 },

{ \_id: { state: 'TX' }, ItsTotalPopulation: 16984601 }

]

**Populations by City**

1. use aggregate to calculate the total population for each city (you have to use city/state combination). You can use a combination for the \_id of the $group: { city: '$city', state: '$state' }

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{city:"$city",state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}})**

[

{

\_id: { city: 'WEST HURLEY', state: 'NY' },

ItsTotalPopulation: 2365

},

{ \_id: { city: 'IOLA', state: 'WI' }, ItsTotalPopulation: 2589 },

{ \_id: { city: 'ELKHART', state: 'IA' }, ItsTotalPopulation: 1374 },

{ \_id: { city: 'ESBON', state: 'KS' }, ItsTotalPopulation: 342 },

{

\_id: { city: 'CASTLE SHANNON', state: 'PA' },

ItsTotalPopulation: 14952

},

{ \_id: { city: 'COLVILLE', state: 'WA' }, ItsTotalPopulation: 8805 },

{ \_id: { city: 'PETERSON', state: 'IA' }, ItsTotalPopulation: 640 },

{ \_id: { city: 'LINEVILLE', state: 'AL' }, ItsTotalPopulation: 4345 },

{

\_id: { city: 'POWERSVILLE', state: 'MO' },

ItsTotalPopulation: 284

},

{

\_id: { city: 'EAST ELMHURST', state: 'NY' },

ItsTotalPopulation: 52016

},

{ \_id: { city: 'KIMPER', state: 'KY' }, ItsTotalPopulation: 2317 },

{ \_id: { city: 'NEWHOPE', state: 'AR' }, ItsTotalPopulation: 400 },

{

\_id: { city: 'SEXTONS CREEK', state: 'KY' },

ItsTotalPopulation: 1776

},

{ \_id: { city: 'ALBURNETT', state: 'IA' }, ItsTotalPopulation: 1038 },

{

\_id: { city: 'EAST LIBERTY', state: 'OH' },

ItsTotalPopulation: 1351

},

{ \_id: { city: 'BIRCHLEAF', state: 'VA' }, ItsTotalPopulation: 997 },

{

\_id: { city: 'BEAR BRANCH', state: 'KY' },

ItsTotalPopulation: 400

},

{

\_id: { city: 'NORTH LEWISBURG', state: 'OH' },

ItsTotalPopulation: 1407

},

{ \_id: { city: 'SHERWOOD', state: 'MD' }, ItsTotalPopulation: 262 },

{

\_id: { city: 'JACKSONS GAP', state: 'AL' },

ItsTotalPopulation: 2236

}

]

Type "it" for more

2. sort the results by population, highest first

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{city:"$city",state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}},{$sort:{ItsTotalPopulation:-1}})**

[

{

\_id: { city: 'CHICAGO', state: 'IL' },

ItsTotalPopulation: 2452177

},

{

\_id: { city: 'BROOKLYN', state: 'NY' },

ItsTotalPopulation: 2300504

},

{

\_id: { city: 'LOS ANGELES', state: 'CA' },

ItsTotalPopulation: 2102295

},

{

\_id: { city: 'HOUSTON', state: 'TX' },

ItsTotalPopulation: 2095918

},

{

\_id: { city: 'PHILADELPHIA', state: 'PA' },

ItsTotalPopulation: 1610956

},

{

\_id: { city: 'NEW YORK', state: 'NY' },

ItsTotalPopulation: 1476790

},

{ \_id: { city: 'BRONX', state: 'NY' }, ItsTotalPopulation: 1209548 },

{

\_id: { city: 'SAN DIEGO', state: 'CA' },

ItsTotalPopulation: 1049298

},

{ \_id: { city: 'DETROIT', state: 'MI' }, ItsTotalPopulation: 963243 },

{ \_id: { city: 'DALLAS', state: 'TX' }, ItsTotalPopulation: 940191 },

{ \_id: { city: 'PHOENIX', state: 'AZ' }, ItsTotalPopulation: 890853 },

{ \_id: { city: 'MIAMI', state: 'FL' }, ItsTotalPopulation: 825232 },

{

\_id: { city: 'SAN JOSE', state: 'CA' },

ItsTotalPopulation: 816653

},

{

\_id: { city: 'SAN ANTONIO', state: 'TX' },

ItsTotalPopulation: 811792

},

{

\_id: { city: 'BALTIMORE', state: 'MD' },

ItsTotalPopulation: 733081

},

{

\_id: { city: 'SAN FRANCISCO', state: 'CA' },

ItsTotalPopulation: 723993

},

{ \_id: { city: 'MEMPHIS', state: 'TN' }, ItsTotalPopulation: 632837 },

{

\_id: { city: 'SACRAMENTO', state: 'CA' },

ItsTotalPopulation: 628279

},

{

\_id: { city: 'JACKSONVILLE', state: 'FL' },

ItsTotalPopulation: 610160

},

{ \_id: { city: 'ATLANTA', state: 'GA' }, ItsTotalPopulation: 609591 }

]

Type "it" for more

3. limit the results to just the first 3 results. What are the top 3 cities in population?

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{city:"$city",state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}},{$sort:{ItsTotalPopulation:-1}},{$limit:3})**

[

{

\_id: { city: 'CHICAGO', state: 'IL' },

ItsTotalPopulation: 2452177

},

{

\_id: { city: 'BROOKLYN', state: 'NY' },

ItsTotalPopulation: 2300504

},

{

\_id: { city: 'LOS ANGELES', state: 'CA' },

ItsTotalPopulation: 2102295

}

]

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{city:"$city"},ItsTotalPopulation:{$sum:"$pop"}}},{$sort:{ItsTotalPopulation:-1}},{$limit:3})**

[

{ \_id: { city: 'CHICAGO' }, ItsTotalPopulation: 2452177 },

{ \_id: { city: 'BROOKLYN' }, ItsTotalPopulation: 2341387 },

{ \_id: { city: 'HOUSTON' }, ItsTotalPopulation: 2123053 }

]

4. What are the top 3 cities in population in Texas?

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$match:{state:"TX"}},{$group: {\_id:{city:"$city",state:"$state"},ItsTotalPopulation:{$sum:"$pop"}}},{$sort:{ItsTotalPopulation:-1}},{$limit:3})**

[

{

\_id: { city: 'HOUSTON', state: 'TX' },

ItsTotalPopulation: 2095918

},

{ \_id: { city: 'DALLAS', state: 'TX' }, ItsTotalPopulation: 940191 },

{

\_id: { city: 'SAN ANTONIO', state: 'TX' },

ItsTotalPopulation: 811792

}

]

**Bonus**

1. Write a query to get the average city population for each state.

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{state:"$state"},AvgPopulation:{$avg:"$pop"}}})**

[

{ \_id: { state: 'ND' }, AvgPopulation: 1632.4092071611253 },

{ \_id: { state: 'CT' }, AvgPopulation: 12498.539923954373 },

{ \_id: { state: 'WI' }, AvgPopulation: 6832.079608938548 },

{ \_id: { state: 'NY' }, AvgPopulation: 11279.248902821317 },

{ \_id: { state: 'IL' }, AvgPopulation: 9238.137429264349 },

{ \_id: { state: 'NJ' }, AvgPopulation: 14315.162962962962 },

{ \_id: { state: 'KS' }, AvgPopulation: 3461.937062937063 },

{ \_id: { state: 'MD' }, AvgPopulation: 11384.235714285714 },

{ \_id: { state: 'RI' }, AvgPopulation: 14539.391304347826 },

{ \_id: { state: 'WY' }, AvgPopulation: 3239.4857142857145 },

{ \_id: { state: 'HI' }, AvgPopulation: 13852.8625 },

{ \_id: { state: 'GA' }, AvgPopulation: 10201.914960629922 },

{ \_id: { state: 'AL' }, AvgPopulation: 7126.255731922399 },

{ \_id: { state: 'MT' }, AvgPopulation: 2544.420382165605 },

{ \_id: { state: 'CO' }, AvgPopulation: 7955.929951690821 },

{ \_id: { state: 'MO' }, AvgPopulation: 5141.496981891348 },

{ \_id: { state: 'WA' }, AvgPopulation: 10055.148760330578 },

{ \_id: { state: 'AK' }, AvgPopulation: 2793.3230769230768 },

{ \_id: { state: 'AZ' }, AvgPopulation: 13574.918518518518 },

{ \_id: { state: 'OK' }, AvgPopulation: 5367.892491467577 }

]

Type "it" for more

2. What are the top 3 states in terms of average city population?

**Atlas atlas-68i9f5-shard-0 [primary] population> db.zipcodes.aggregate({$group: {\_id:{state:"$state"},AvgPopulation:{$avg:"$pop"}}},{$sort:{AvgPopulation:-1}},{$limit:3})**

[

{ \_id: { state: 'DC' }, AvgPopulation: 25287.5 },

{ \_id: { state: 'CA' }, AvgPopulation: 19627.236147757256 },

{ \_id: { state: 'FL' }, AvgPopulation: 15779.407960199005 }

]