

22 Comments

The Jackal of Javascript

 Login ▾ Recommend Share

Sort by Best ▾



Join the discussion...

**acveer** · 7 months ago

Aravind,

Excellent tutorial. Well-written. It help me get started with mapReduce. I have tried your examples with mongoDB aggregate framework & thought these might help someone. My intention here is share what I have learned and to seek expertise on the aggregation framework.

Below queries can be run on mongo shell to get the same results with aggregate framework.

Example1 :

```
db.sourceData.aggregate( {$group: { _id: "$gender", total: {$sum: 1}}});
```

Example2 was a little complicated for me, still learning/trying. If you know how to achieve this, please help.

Example3:

The dummy-json template + custom helper logic above is returning hobbies like below:

```
"hobbies" : "[Acrobatics,Meditation,Music]"
```

which was not a valid array for using \$unwind in the aggregation framework. So I changed the template(schema.hbs) for hobbies from

```
"hobbies" : "{{hobbies}}"
```

to

[see more](#)

1 ^ | v · Reply · Share ›

**Arvind Ravulavaru** Mod → acveer · 7 months ago

Hello acveer. Thanks! And Thanks for sharing your solutions.

I guess you can tweak the dataGen.js to give you the array you want from inside the hobby(). Let me know if it works.

A few days ago I was planning to do a comparison between AF and MR, never got to it though. Did you try anything in that space?

Thanks.

^ | v · Reply · Share ›



acveer → Arvind Ravulavaru · 7 months ago

Arvind,

I have not done enough MP or AF for the detailed comparison, but here is what I understood so far.

1. AF is much faster than MP.
2. Both are targeted to be used in batch operations in background mode (not real-time). (so convenience/control).
3. Both have respective limitations in terms of document size, max documents in process (100MB) and sharding support.
4. MongoDB 2.6 has more handy features added for AF.

I could not figure out how to insert hobbies array by changing the template or helpers function. The array needs to be this way `""hobbies" : ["Painting", "Cooking", "Reading"]`. So I just inserted empty hobbies array from datagen.js and written another script to update hobbies array separately.

Will post solution for example 2, once I make progress.

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → acveer · 7 months ago

Thanks for sharing your findings. I will also take a look at hobbies array and get back.

^ | v · Reply · Share ›



Peter Boot · 2 months ago

What is the correct format for template.hbs ?

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → Peter Boot · 2 months ago

Correct format as in?

^ | v · Reply · Share ›



Logic Town · 3 months ago

Do you know if it's possible to invoke a function provided by an npm module inside the mapper or reducer functions?

For instance

```
var mymod = require('mymod');
```

```
var mapper = function(){
```

```
emit(this.key, mymod.apply(this));  
}
```

...

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → Logic Town · 3 months ago

I have not tried but it should be possible. Did you try it?

^ | v · Reply · Share ›



Logic Town · 3 months ago

I reckon the functions such as

```
db.example3_results.find(function (err, docs) {  
if(err) console.log(err);  
console.log(docs);  
});
```

should be passed as a callback to the mapReduce command, otherwise they will be invoked before the job completes.

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → Logic Town · 3 months ago

I am not sure if Mongojs API has a callback for `mapreduce()`, if it does then that is the place to put it. But I think `mapreduce()` in a blocking way here. Not sure though.

^ | v · Reply · Share ›



Harshavardhan Reddy · 5 months ago

Hey Aravind,

I want to do with multiple collections.

Can you please help with that?

Thank you.

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → Harshavardhan Reddy · 5 months ago

Hello Harsha,

Take a look at : <http://stackoverflow.com/a/383...>

Thanks,

Arvind.

^ | v · Reply · Share ›



rajkumar · 7 months ago

Hi i have json like this.

```
"marks":{
  "sem1" :{
    "mark1":10,
    "total":100
  },
  "sem2":{
    "mark2":20,
    "total":200
  },
  "sem3":{
    "mark2":30,
    "total":300
  }
}
```

I need result like

mark total sem

10 100 sem1

20 200 sem2

30 300 sem3

how can i achive above format using monogodb query.query is jaspersoft related means very useful

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → rajkumar · 7 months ago

I have not worked on jaspersoft so I am not sure how the query needs to written. If you want to return an object and compare a value refer example 2 in the post. -- Thanks.

^ | v · Reply · Share ›



syd · 8 months ago

and thanks for your quick response.

^ | v · Reply · Share ›



syd · 8 months ago

can you explain a little briefly please, as i am doing a project to calculate the time taken by mongoDB and hadoop in mapReduce algorithm when they store or retrieve different types of bulk

data.

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → syd · 7 months ago

I have not dug deeper into the algorithms, so I am not aware of it. You can look for more info here : <https://github.com/mongodb>

^ | v · Reply · Share ›



syd → Arvind Ravulavaru · 7 months ago

Thank you

1 ^ | v · Reply · Share ›



syd · 8 months ago

what are the different techniques to find the time taken by my query to fetch output in mapReduce

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → syd · 8 months ago

AFAIK, there is no official benchmarking tool for MongoDB. You can track the time taken from within the node application. Try this : <http://blog.nodejs.org/2012/04...>

Or you can try

```
console.time('myJob');  
// your job  
console.timeEnd('myJob');
```

This should give a fair idea.

^ | v · Reply · Share ›



sppericat · a year ago

This is nice, but why don't you use the the aggregation framework instead ?

^ | v · Reply · Share ›



Arvind Ravulavaru Mod → sppericat · a year ago

Thanks sppericat,

No argument there. You can use an aggregation framework to do the same. I wanted to throw some light on aggregating data using MapReduce.

IMO, MapReduce is a verbose version of the aggregation framework & is also an alternative. The only key difference I see between MapReduce & Aggregation Framework is the built in pipe operators like \$geoNear. \$sum, \$gte etc. In a MapReduce paradigm, you

end up writing these on your own.

Thanks,
Arvind.

^ | v • Reply • Share ›

ALSO ON THE JACKAL OF JAVASCRIPT

WHAT'S THIS?

Pushbots and Cordova – Easy Push Notifications for your app

24 comments • 3 months ago

Getting Started with Ionic Creator

8 comments • 9 months ago

M.E.A.N. Stack Development with Vagrant

10 comments • 7 months ago

Javascript 101

2 comments • 6 months ago

 [Subscribe](#)

 [Add Disqus to your site](#)

 [Privacy](#)