**■ README.md** 

## ServerLogsNoSQL

A REST API project in Flask & MongoDB for post-graduate class Database Management Systems

### **Authors**

- Papadopoulos Christos-Charalampos
- Papamichalopoulos Marios

### **Tools Used**

- Python 3.6.9
- Flask 1.1.1
- MongoDB Community Server 4.2.2
- Faker
- Postman

## How to run

• populate the database by running the scripts:

populateAdmins.py

and

populateLogs.py

inside logs directory

- traverse inside the directory
- type in your terminal:
- . flask-mongodb/bin/activate

or set up a virtual environment downloading:

flash flask-pymongo

• start the REST API:

flask run

• app is up and running on

# Schema Design

Our database consists of two primary collections:

- log
- admin

The first one contains all types of logs. We felt it was not correct design principal to furtherly normalize our log collection, since the point of NoSQL is to keep normalization at a minimum. Another reason is that lookups are really costly and one wants to avoid it at all costs.

As a result we have merged all the types of logs into one collection, to avoid using joins for the queries.

The second collection, contains all the admin related data as mentioned in the requirements of the project. In more detail, each admin owns the following properties:

- username
- email
- telephone
- an array of the upvotes casted

All the admin data have been generated using Faker.

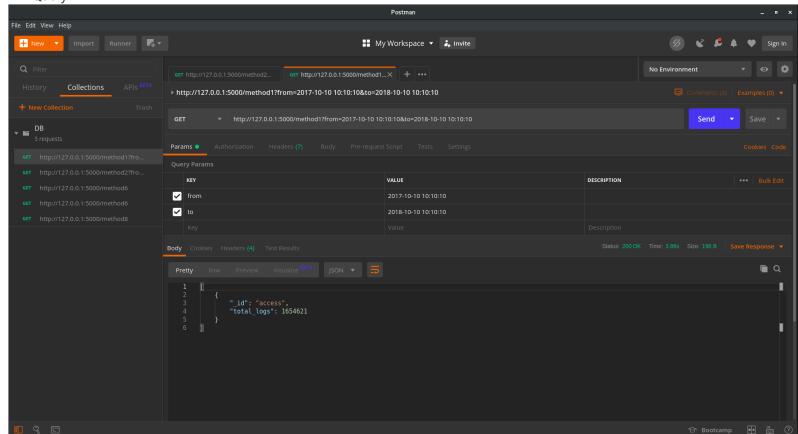
### Queries

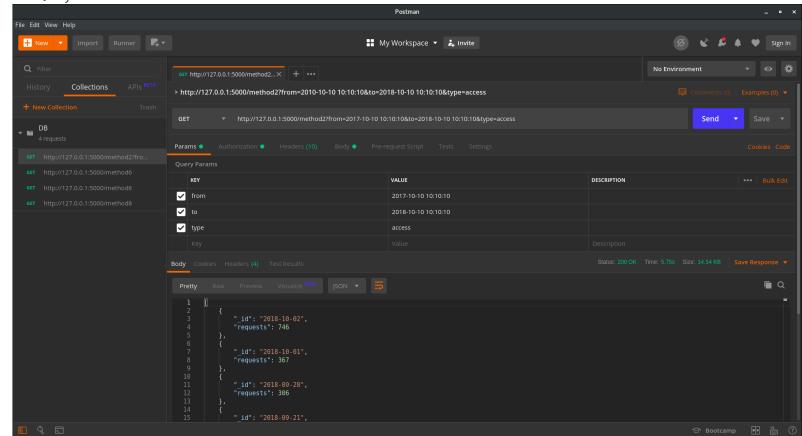
All the queries can be found at:

- ./ServerLogsFlask/ServerLogsNoSQL/methods.py
- ./ServerLogsFlask/ServerLogsNoSQL/insert.py

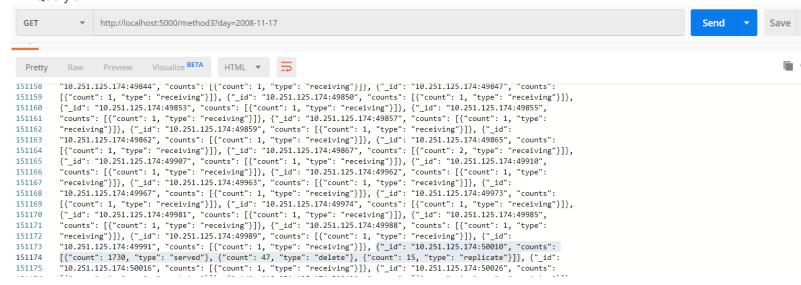
# Sample Snapshots

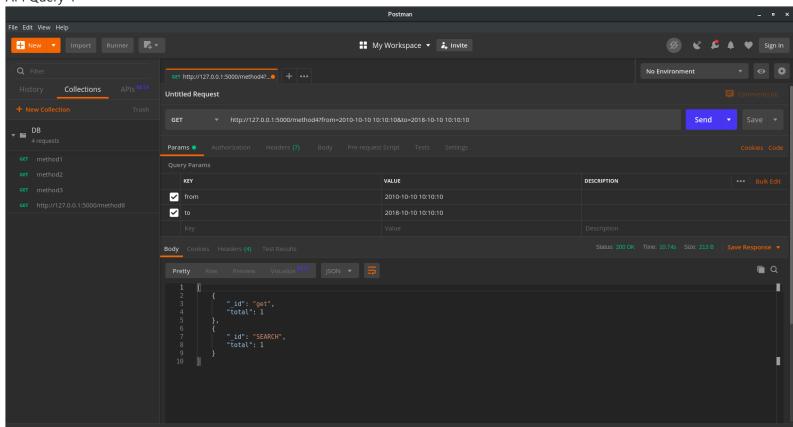
We present some sample snapshots without the use of indeces:

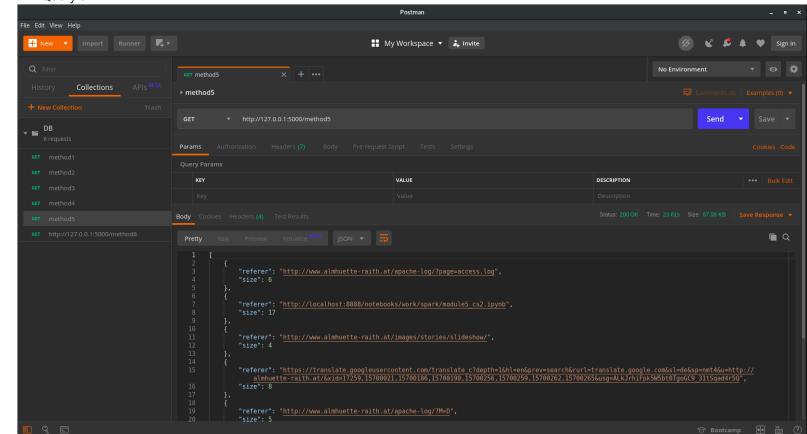


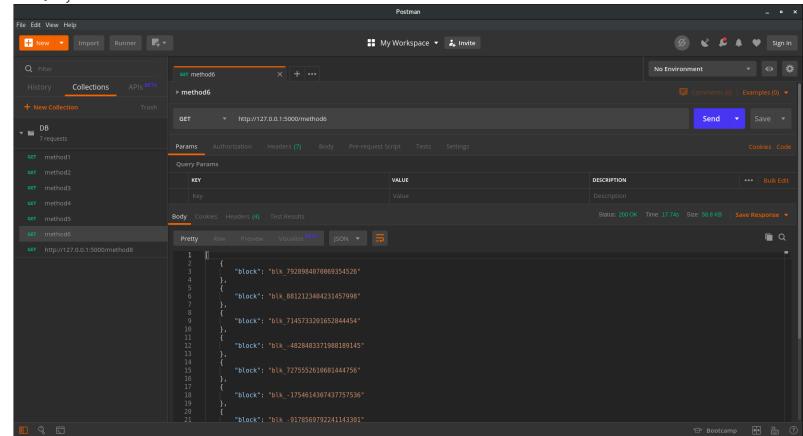


#### • API Query 3

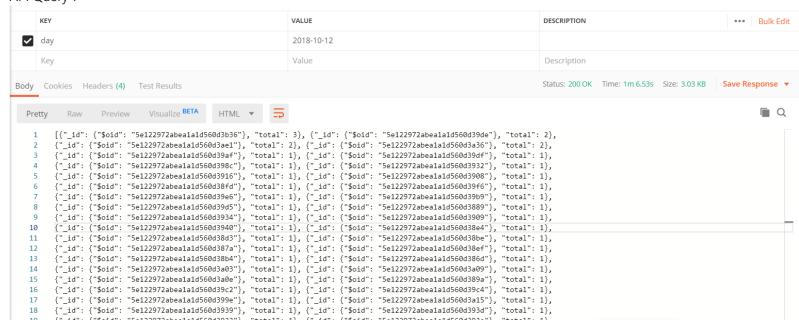


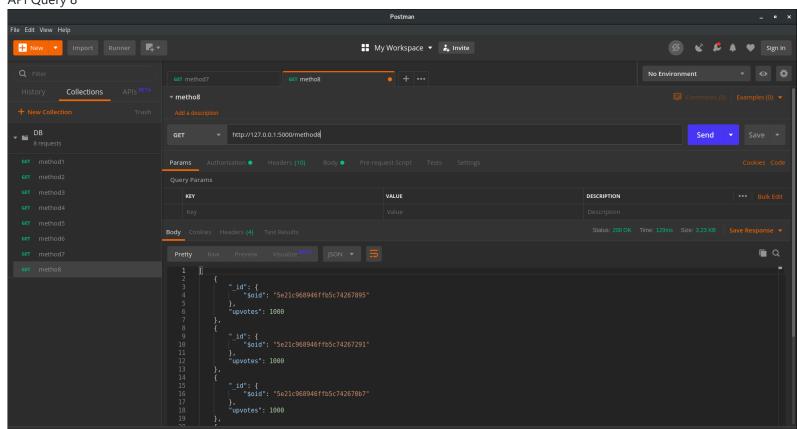


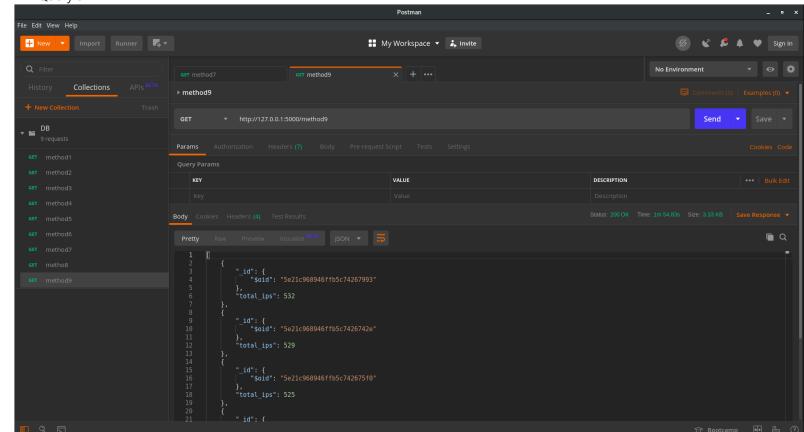




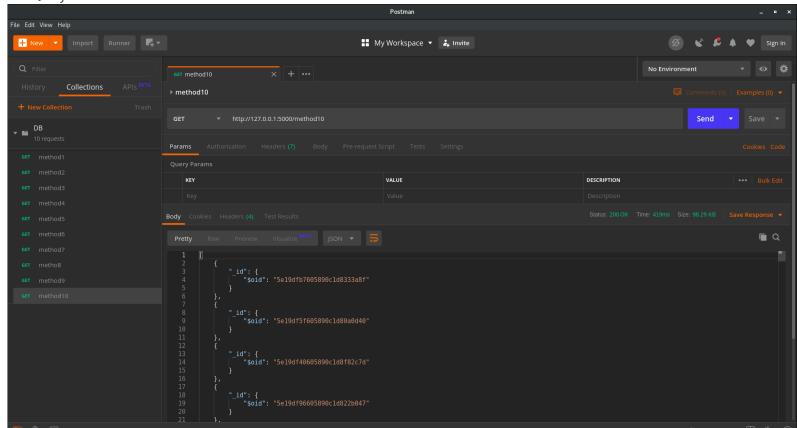
#### • API Query 7

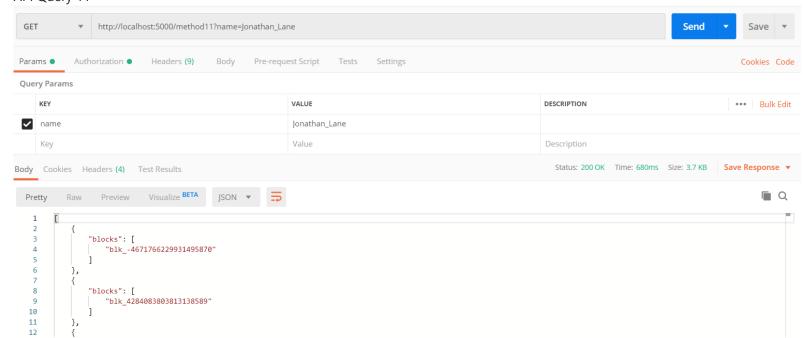






• API Query 10





• Insert Log Examples 
 + New
 ▼

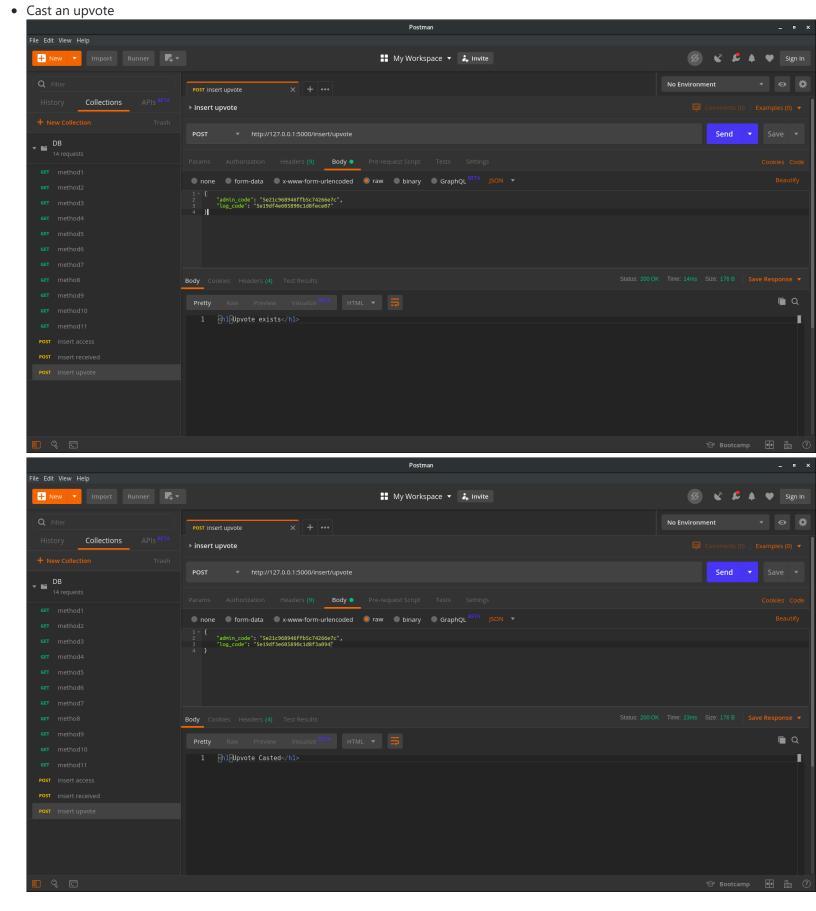
 Import
 Runner

 FL ▼
 My Workspace ▼ ♣ Invite POST Insert access ▶ insert access **■** Q Т 
 + New
 ✓
 Import
 Runner
 F. ▼
 ■ My Workspace ▼ ♣ Invite Sign In o 🌣 POST Insert received ▶ insert received Send ▼ Params Authorization Headers (9) Body Pre-request Script Tests Settings "source\_ip": "1.1.1.1",
"log\_timestamp": "2018-10-10 10:10:10",
"blocks" : ["block1", "block2"],
"destinations": ["dest1", "dest2"]

1 hlpReceived Log Inserted</hl>

**■** Q

Т



### **Indices**

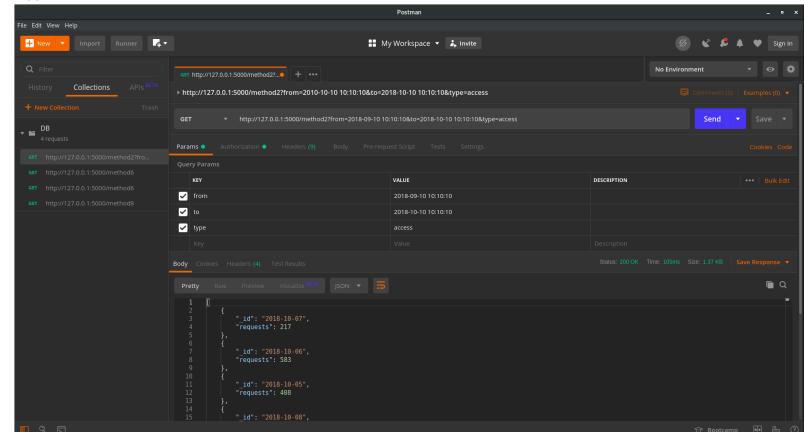
Log\_timestamp Index

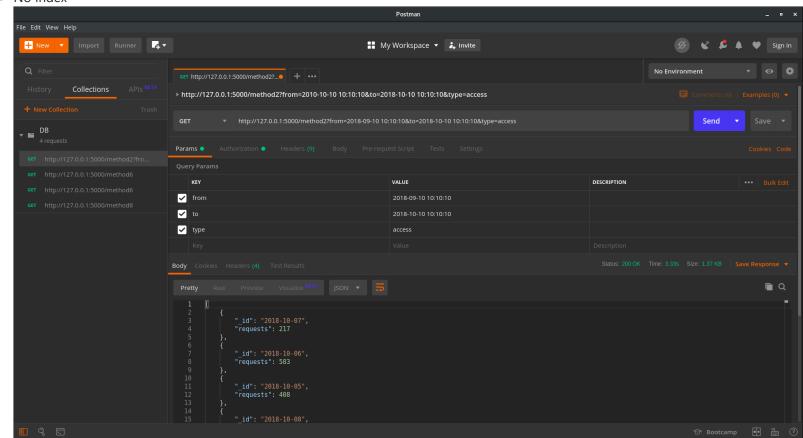
We found out that indices are situational since only two aggregation pipeline stages (sort, match) take indices into account. Group for example makes no use of it.

In addition to these, by running time range queries we observed that indices speed up the query when they search for a small time range. We can see that with the following snapshots based on method2:

1. For a small time range query using index makes it faster, to the point that is almost instant:

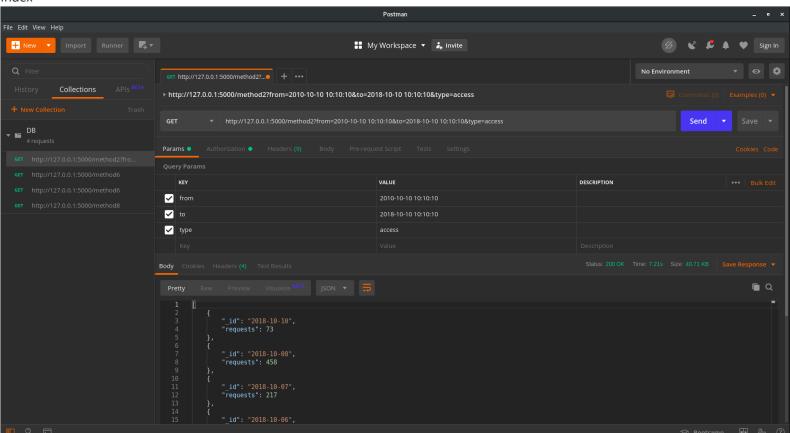
Index

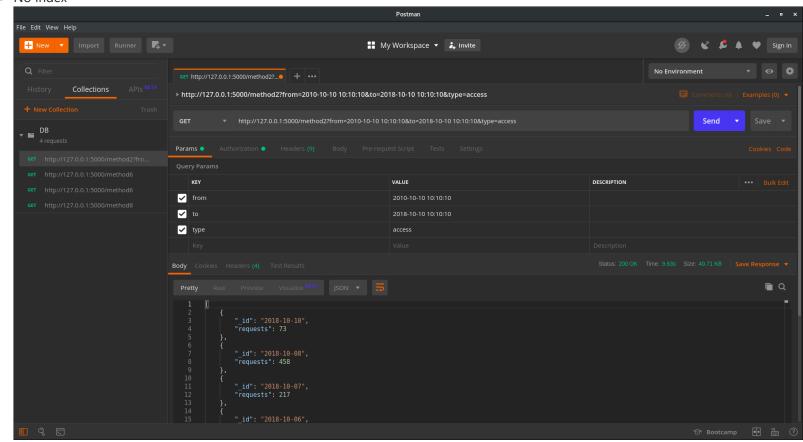




2. For a relatively large time range query we observe that the index gives only a slight boost in most cases:

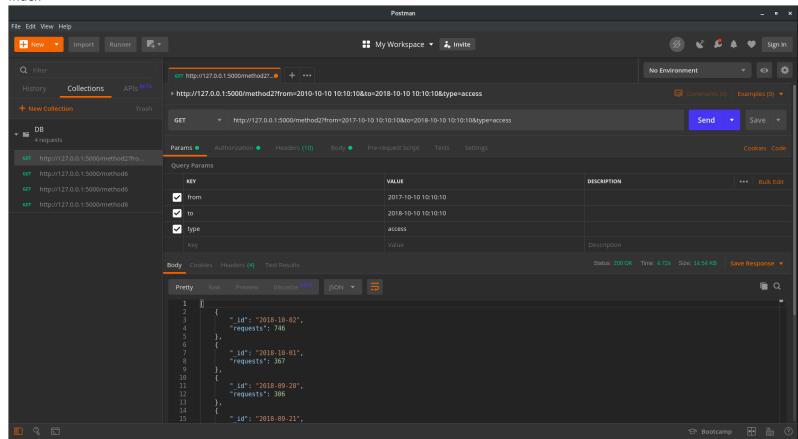


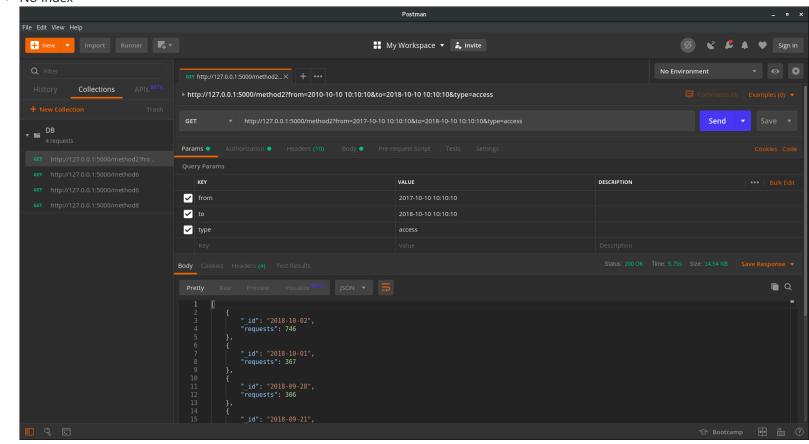




#### 3. Average case:

Index



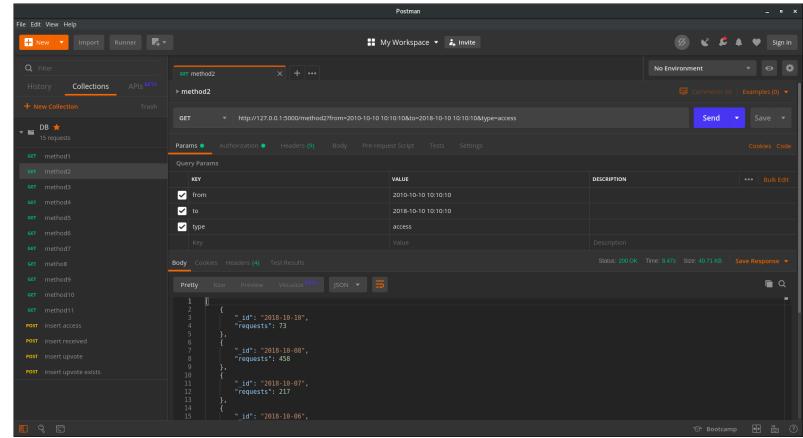


Despite the not so observable optimization, having a timestamp index is important since most of the queries involve a specific date or time range.

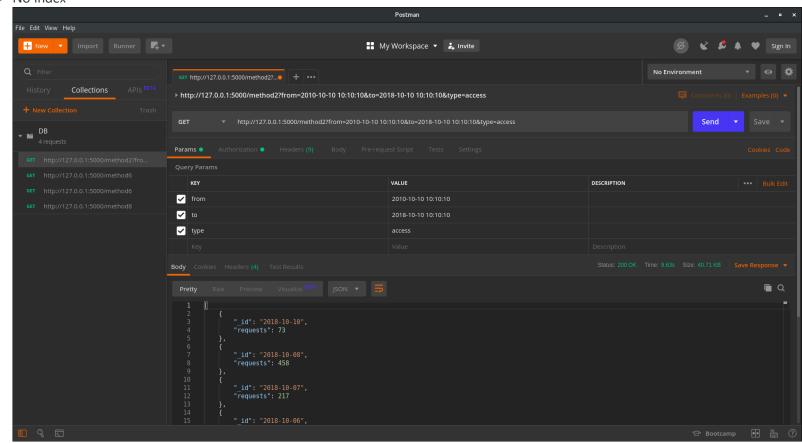
### Type Index

As we see from the snapshot following adding an index on type field gives a slight boot to the query:

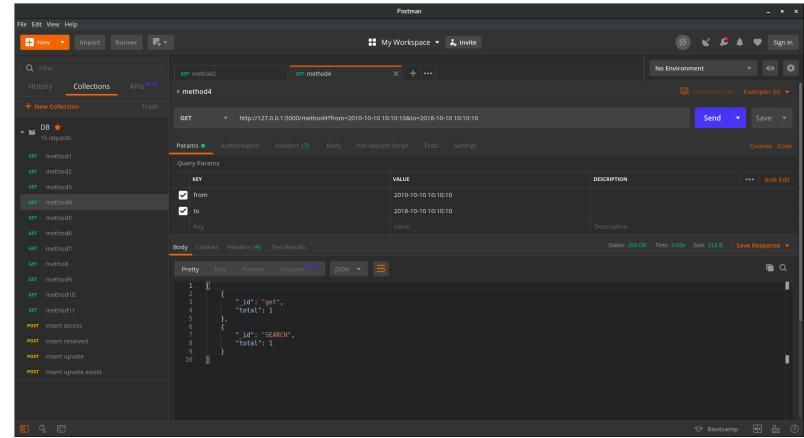
Index



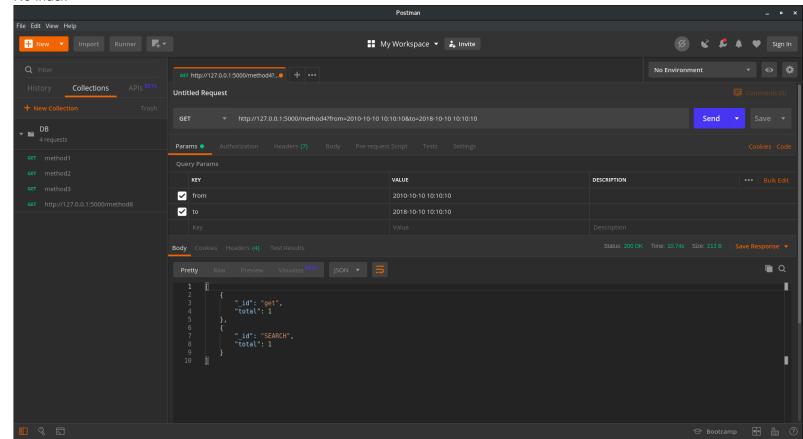
No Index



Index

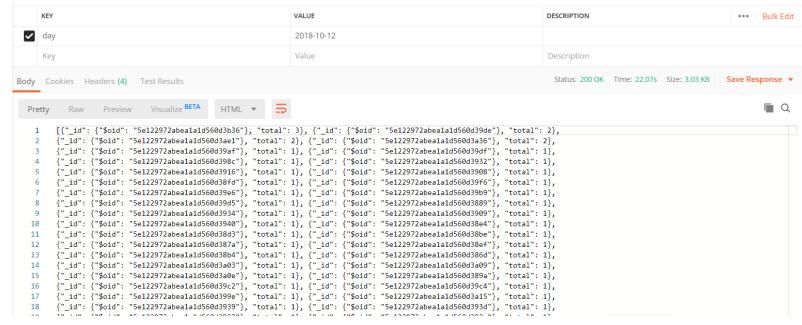


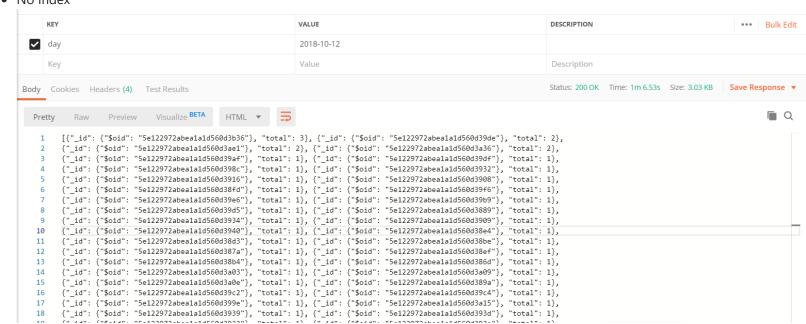
No Index



### **Upvote Index**







Indeces were tested on other fields as well but no noticeable optimizations was indicated. We also created a compound index on log\_timestamp and type but it actually made the query slower.

### Github Link

ServerLogsNoSQL