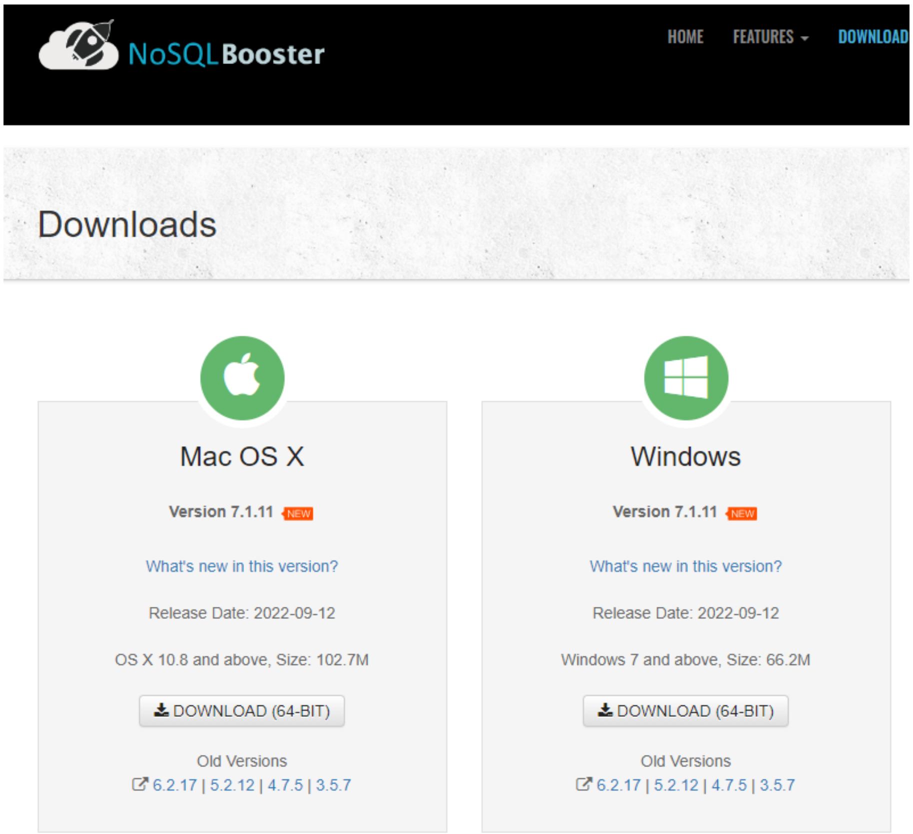
# Tutorial - Week 11

## Objectives:

- Explore NoSQL database (MongoDB)
- Install and use a browser for MongoDB - NoSQLBooster
- Get understanding of JSON
- Revise theoretical concepts of NoSQL Databases

## A. Download & Install

1. Go to: https://nosqlbooster.com/downloads (version 7.1.x – whatever is the latest version).



Downloads

Mac OS X

Version 7.1.11 NEW

What's new in this version?

Release Date: 2022-09-12

OS X 10.8 and above, Size: 102.7M

DOWNLOAD (64-BIT)

Old Versions
6.2.17 | 5.2.12 | 4.7.5 | 3.5.7

Windows

Version 7.1.11 NEW

What's new in this version?

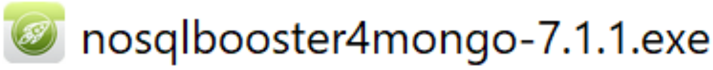
Release Date: 2022-09-12

Windows 7 and above, Size: 66.2M

DOWNLOAD (64-BIT)

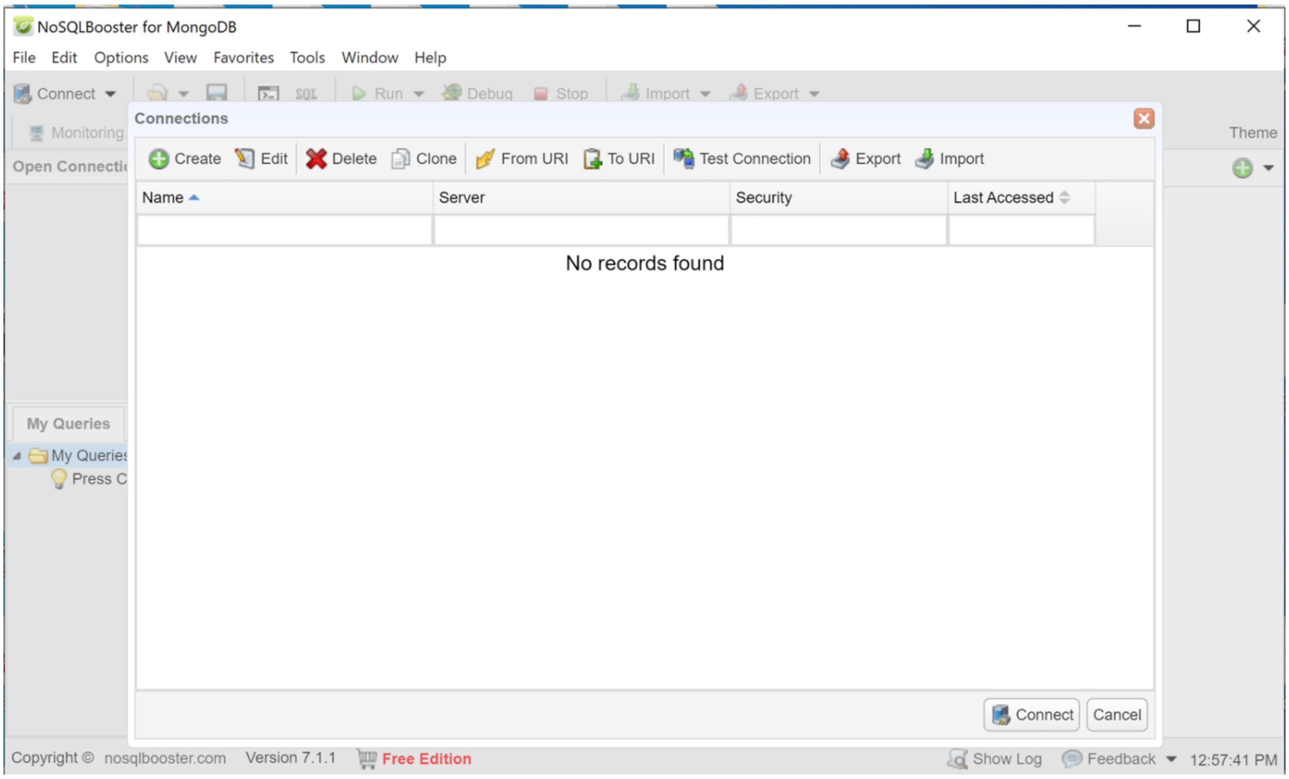
Old Versions
6.2.17 | 5.2.12 | 4.7.5 | 3.5.7

2. Download the version relevant for your OS, run it to install (takes a few minutes). The screenshots below are for Windows OS. This is the downloaded software NoSQLBooster:



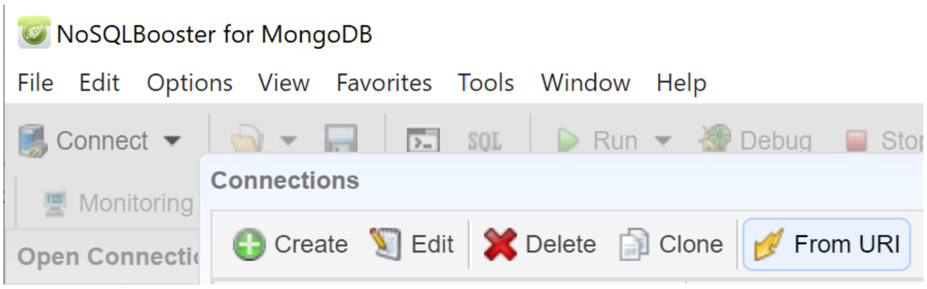
nosqlbooster4mongo-7.1.1.exe

3. Start/run NoSQLBooster. It will take a few minutes then you will see something similar to the screenshot below.



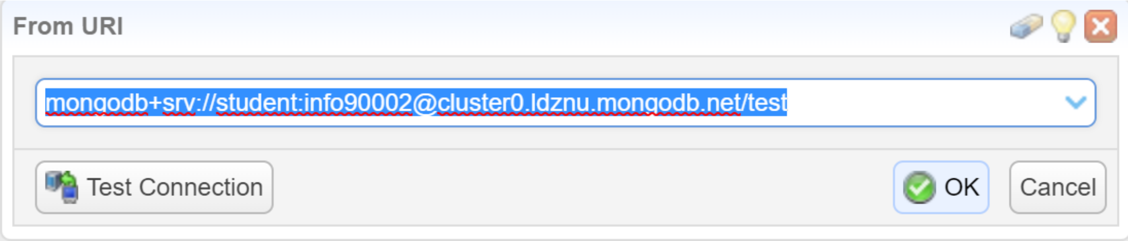
## B. Connect to MongoDB

1. Click on the button ‘From URI’ (see screenshot below right)



2. Copy and paste this connection string into the pop-up (**don’t type it**):

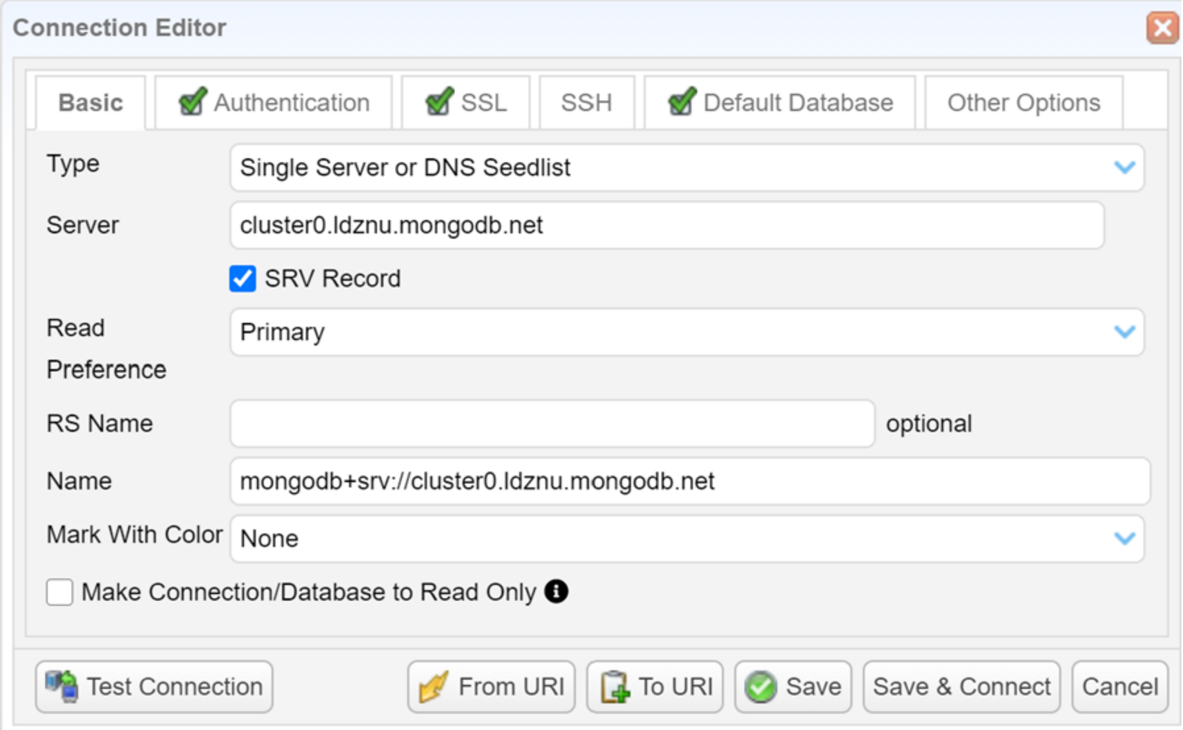
mongodb+srv://student:info90002@cluster0.ldznu.mongodb.net/test



mongodb+srv://student:info90002@cluster0.ldznu.mongodb.net/test

3. Click ‘OK’

4. You will be returned to the original window where you need to click on ‘Save & Connect’ (below right)



Connection Editor

Type: Single Server or DNS Seedlist

Server: cluster0.ldznu.mongodb.net

☑ SRV Record

Read Preference: Primary

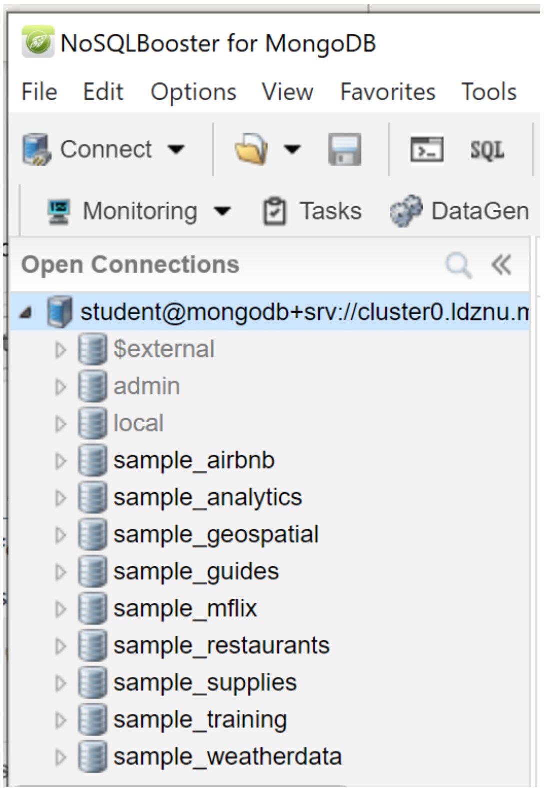
RS Name: optional

Name: mongodb+srv://cluster0.ldznu.mongodb.net

Mark With Color: None

☐ Make Connection/Database to Read Only

5. In the left pane you will see sample databases:



NoSQLBooster for MongoDB

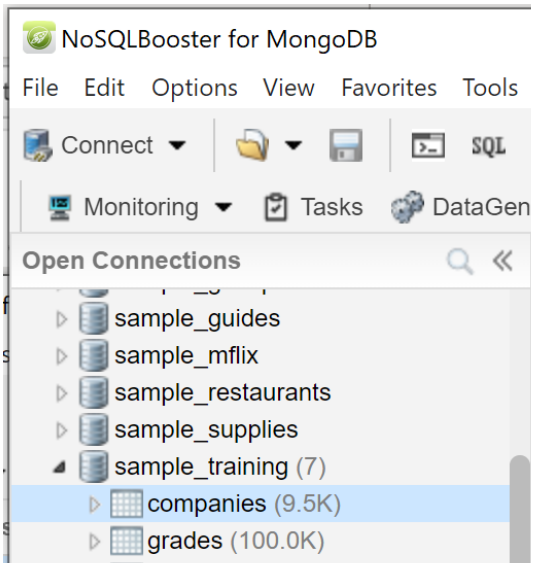
Open Connections

student@mongodb+srv://cluster0.ldznu.m

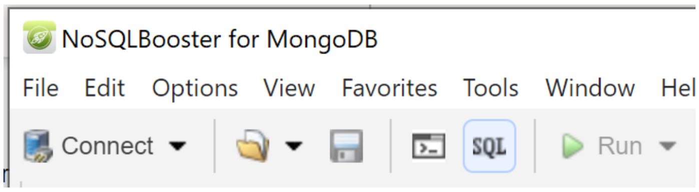
- $external
- admin
- local
- sample_airbnb
- sample_analytics
- sample_geospatial
- sample_guides
- sample_mflix
- sample_restaurants
- sample_supplies
- sample_training
- sample_weatherdata

## C. Explore MongoDB sample data using NoSQLbooster

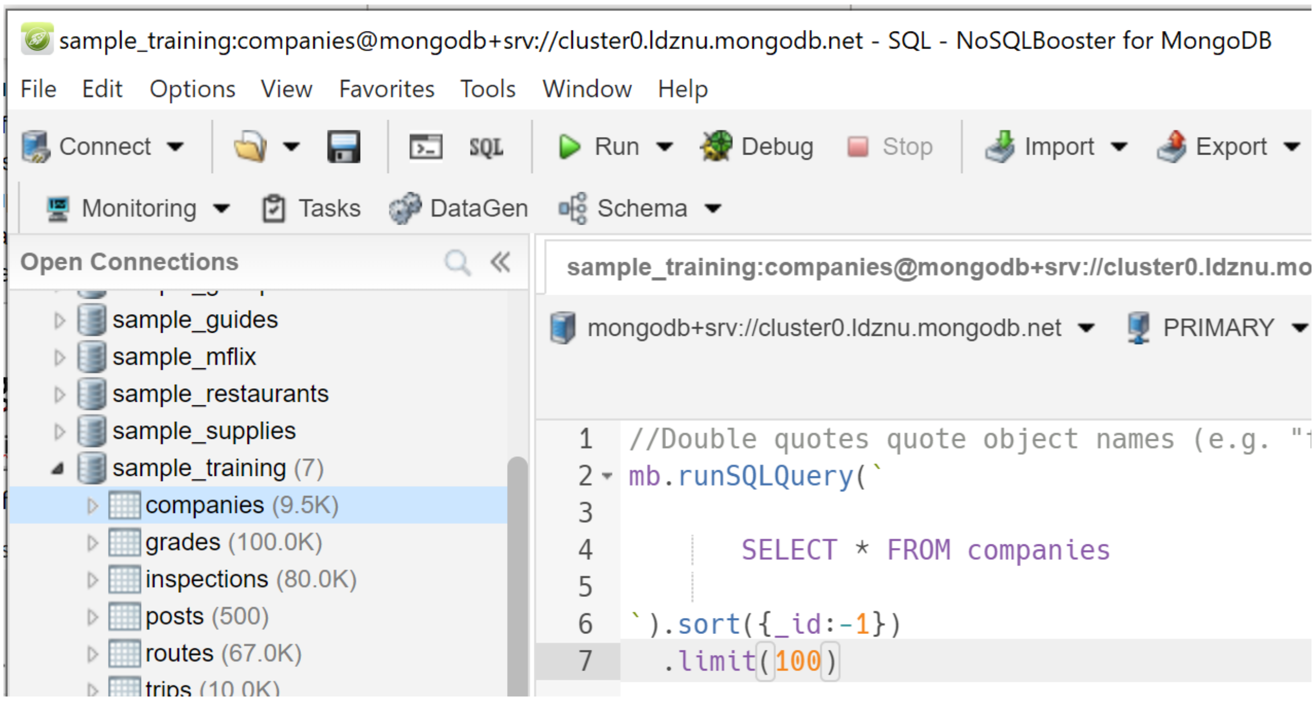
1. Select the database ‘sample_training’, then the collection ‘companies’



2. Click on the ‘**SQL**’ button (as shown below)



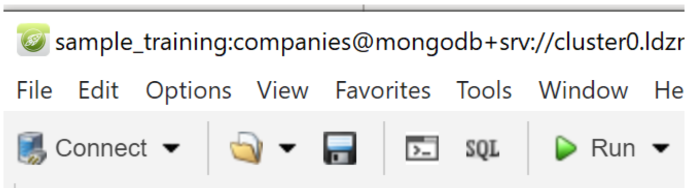
3. The browser will show a default SQL query.



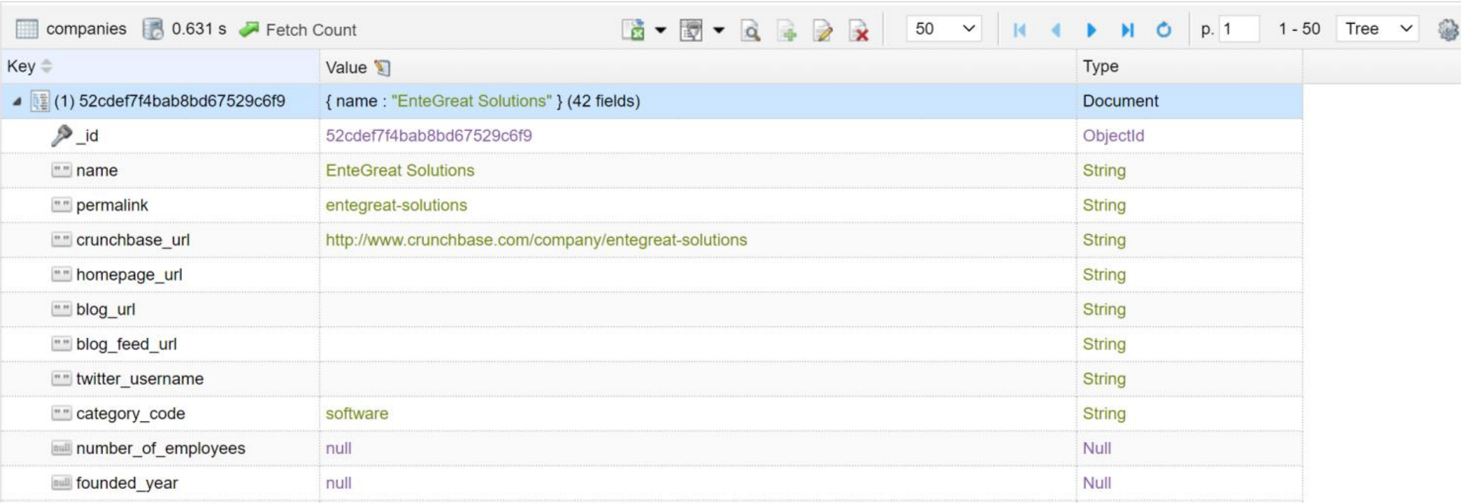
```
//Double quotes quote object names (e.g. "t
mb.runSQLQuery(`

      SELECT * FROM companies

`).sort({_id:-1})
  .limit(100)
```

**What is the default SQL query?**

4. Execute the query by clicking the 'Run' button:
   (next to the SQL button as in the screenshot below right)



5. To see some company data (e.g. 'EnteGreat Solutions' below):



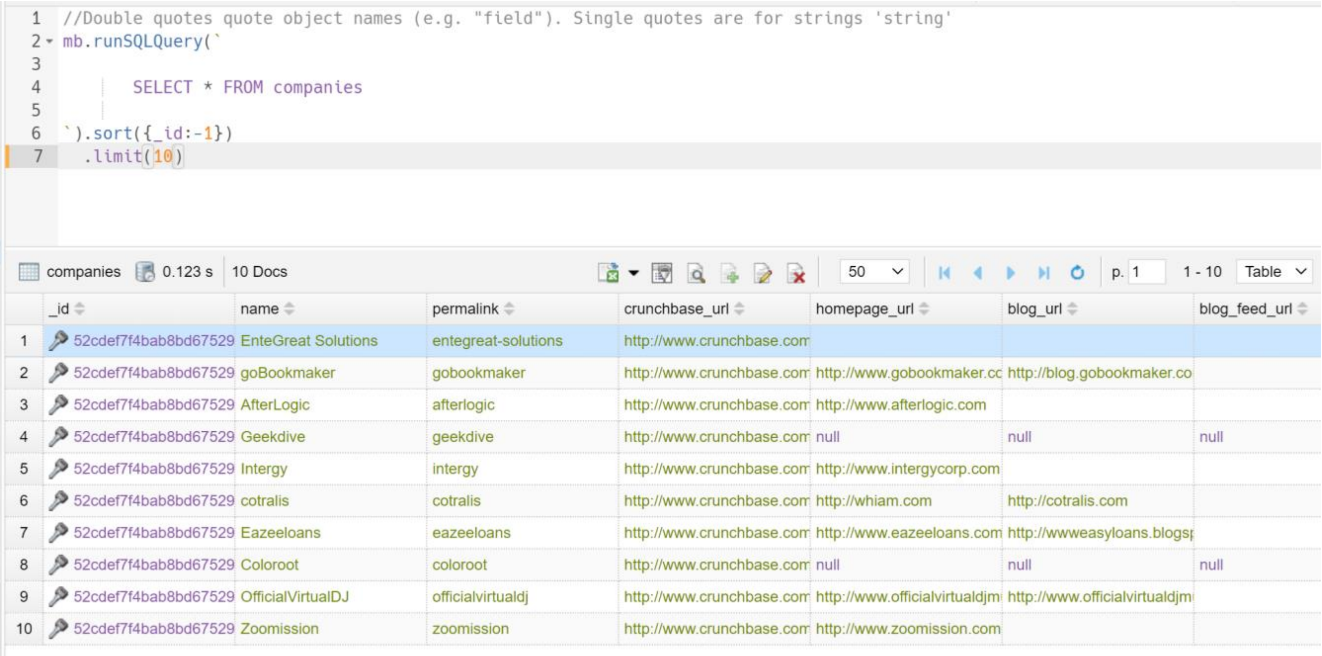
| Key | Value | Type |
|---|---|---|
| (1) 52cdef7f4bab8bd67529c6f9 | { name : "EnteGreat Solutions" } (42 fields) | Document |
| _id | 52cdef7f4bab8bd67529c6f9 | ObjectId |
| name | EnteGreat Solutions | String |
| permalink | entegreat-solutions | String |
| crunchbase_url | http://www.crunchbase.com/company/entegreat-solutions | String |
| homepage_url | | String |
| blog_url | | String |
| blog_feed_url | | String |
| twitter_username | | String |
| category_code | software | String |
| number_of_employees | null | Null |
| founded_year | null | Null |

509 -

6. **This is 'Tree' view look at data types, how many are there? What types are they and what is the highest level (or parent)? Why is it called 'Tree'?** Document. It has branches or levels.

7. On the far right change 'Tree' to 'Table'

8. Now change limit to 10 and run again:

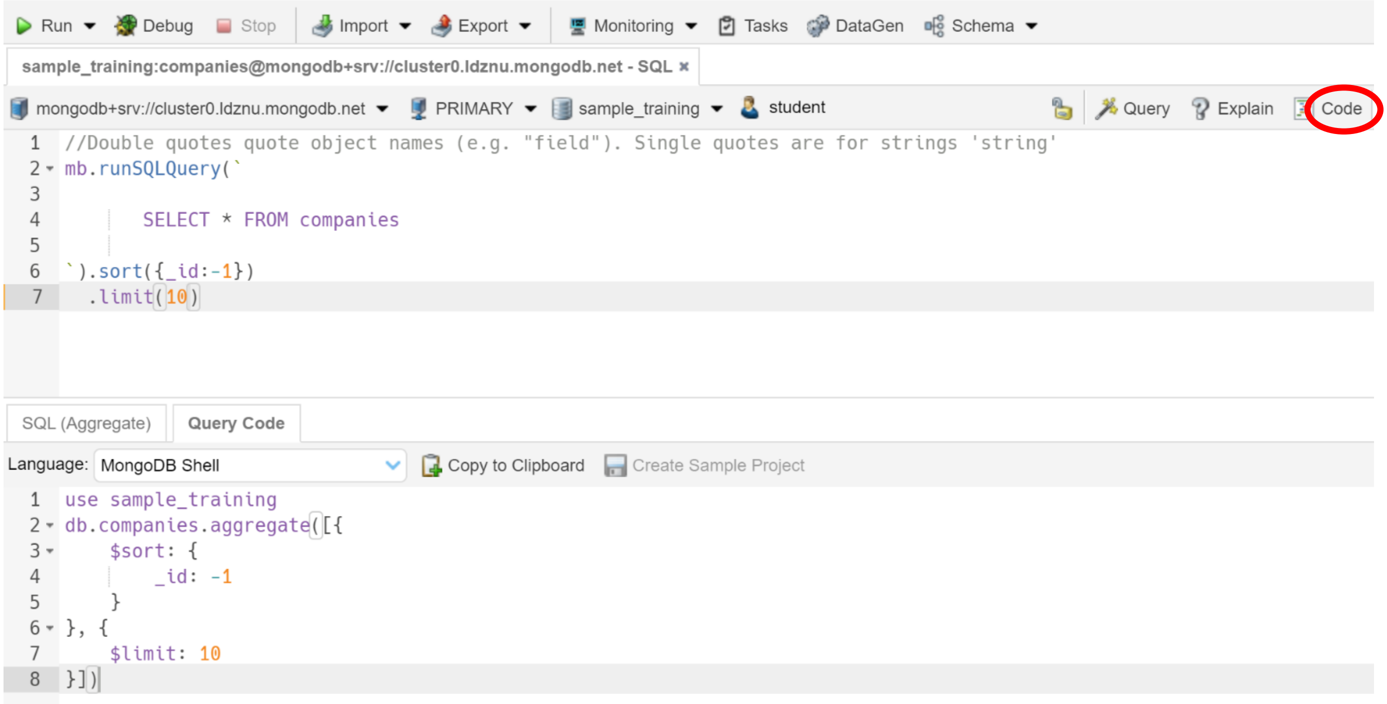


```
//Double quotes quote object names (e.g. "field"). Single quotes are for strings 'string'
mb.runSQLQuery(`

    SELECT * FROM companies

`).sort({_id:-1})
  .limit(10)
```

| | _id | name | permalink | crunchbase_url | homepage_url | blog_url | blog_feed_url |
|---|---|---|---|---|---|---|---|
| 1 | 52cdef7f4bab8bd67529 | EnteGreat Solutions | entegreat-solutions | http://www.crunchbase.com | | | |
| 2 | 52cdef7f4bab8bd67529 | goBookmaker | gobookmaker | http://www.crunchbase.com | http://www.gobookmaker.cc | http://blog.gobookmaker.co | |
| 3 | 52cdef7f4bab8bd67529 | AfterLogic | afterlogic | http://www.crunchbase.com | http://www.afterlogic.com | | |
| 4 | 52cdef7f4bab8bd67529 | Geekdive | geekdive | http://www.crunchbase.com | null | null | null |
| 5 | 52cdef7f4bab8bd67529 | Intergy | intergy | http://www.crunchbase.com | http://www.intergycorp.com | | |
| 6 | 52cdef7f4bab8bd67529 | cotralis | cotralis | http://www.crunchbase.com | http://whiam.com | http://cotralis.com | |
| 7 | 52cdef7f4bab8bd67529 | Eazeeloans | eazeeloans | http://www.crunchbase.com | http://www.eazeeloans.com | http://wwweasyloans.blogs | |
| 8 | 52cdef7f4bab8bd67529 | Coloroot | coloroot | http://www.crunchbase.com | null | null | null |
| 9 | 52cdef7f4bab8bd67529 | OfficialVirtualDJ | officialvirtualdj | http://www.crunchbase.com | http://www.officialvirtualdjm | http://www.officialvirtualdjm | |
| 10 | 52cdef7f4bab8bd67529 | Zoomission | zoomission | http://www.crunchbase.com | http://www.zoomission.com | | |

**9.** Now try ‘JSON’ view (far right).
**Explore**, note that data is in key:value pairs e.g. "name" : "EnteGreat Solutions" (and separated by commas). **Find ‘Acquisitions’, what data type is it?** array

10. Now try ‘Code’ (button circled at the far right) to see:

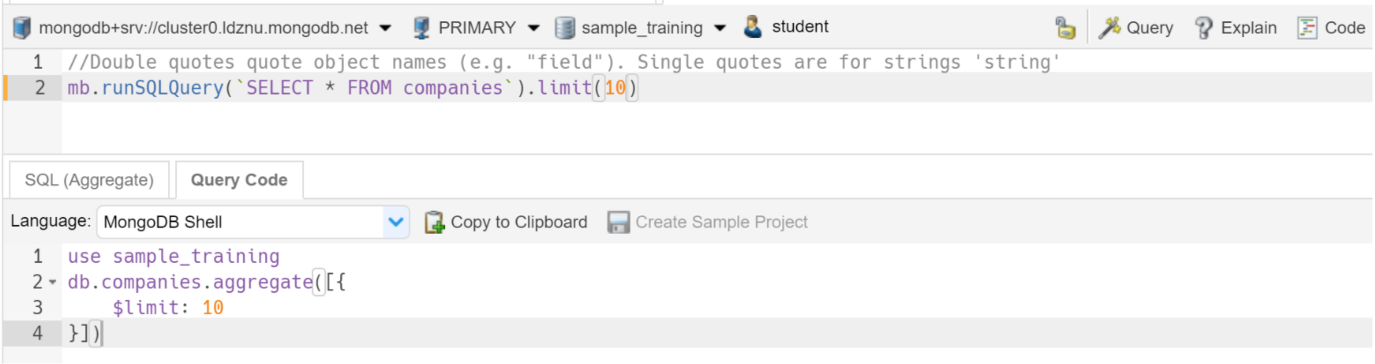


```
//Double quotes quote object names (e.g. "field"). Single quotes are for strings 'string'
mb.runSQLQuery(`

      SELECT * FROM companies

`).sort({_id:-1})
  .limit(10)
```

```
use sample_training
db.companies.aggregate([{
    $sort: {
        _id: -1
    }
}, {
    $limit: 10
}])
```

**Obviously we have some sort of SQL above but what is the ‘Query Code’?** db. companies. aggregated

11. Simplify this query so that there is no sort, just limit (you can use the ‘Table’ view to sort columns). Run to test if it’s still working. Examine ‘Query Code’ again.

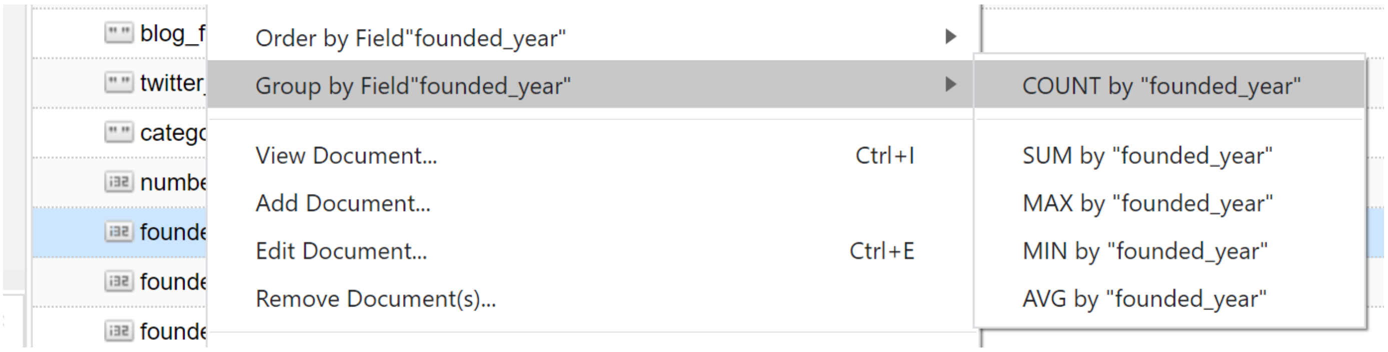


```
//Double quotes quote object names (e.g. "field"). Single quotes are for strings 'string'
mb.runSQLQuery(`SELECT * FROM companies`).limit(10)
```

```
use sample_training
db.companies.aggregate([{
    $limit: 10
}])
```

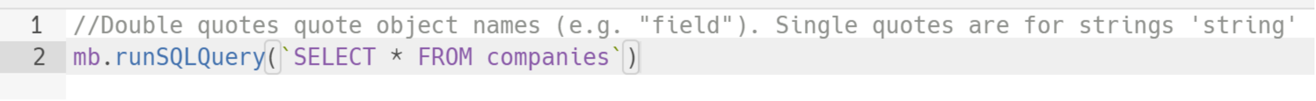
**12. Now search for companies founded after 2010 using SQL. How many are there?**

13. **You can achieve a one-click grouping/filtering of data fields**.
Switch to tree view and right click on ‘founded_year’ then ‘Group by..’ then ‘COUNT..’ (see screenshot below)



**How many companies were founded after 2010? (and if the answer is different to the previous question.. review it).**

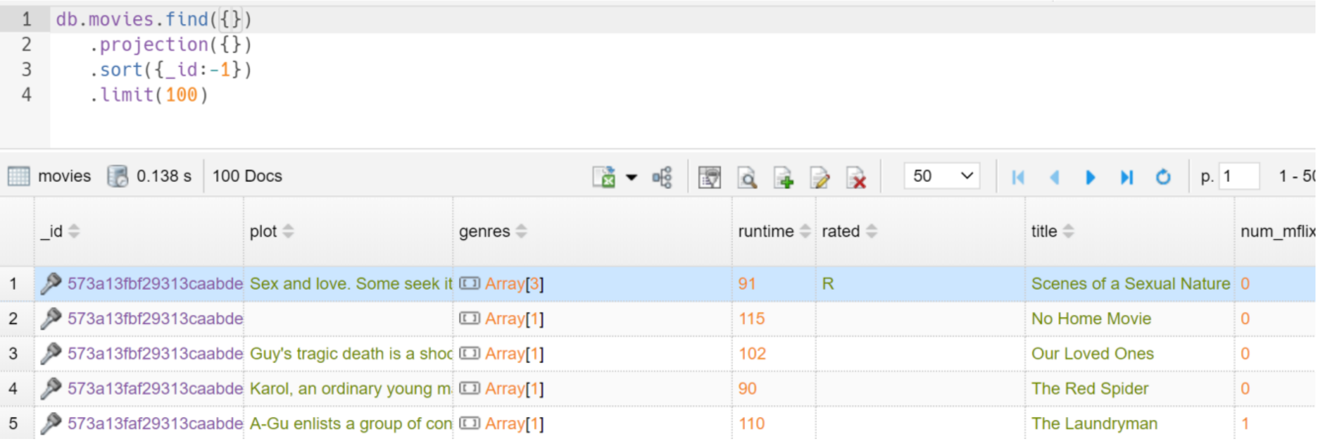
14. //Double quotes quote object names (e.g. "field"). Single quotes are for strings 'string'



```
1  //Double quotes quote object names (e.g. "field"). Single quotes are for strings 'string'
2  mb.runSQLQuery(`SELECT * FROM companies`)
```

**As above, on line 1, what does this mean?**
**Try a search to find companies that start with "E" (using SQL), how many are there?**

15. To the left of the SQL button is a code/shell button. Find and select the movies collection, then the shell button, then 'Run' to see (table view):

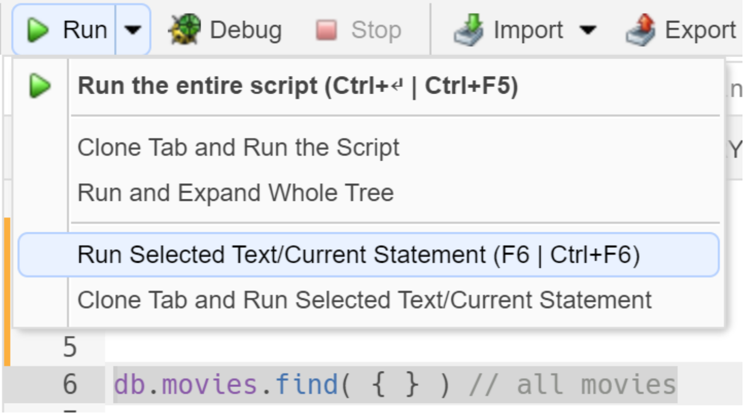


```
db.movies.find({})
   .projection({})
   .sort({_id:-1})
   .limit(100)
```

movies 0.138 s 100 Docs

| | _id | plot | genres | runtime | rated | title | num_mflix |
|---|---|---|---|---|---|---|---|
| 1 | 573a13fbf29313caabde | Sex and love. Some seek it | Array[3] | 91 | R | Scenes of a Sexual Nature | 0 |
| 2 | 573a13fbf29313caabde | | Array[1] | 115 | | No Home Movie | 0 |
| 3 | 573a13fbf29313caabde | Guy's tragic death is a shoc | Array[1] | 102 | | Our Loved Ones | 0 |
| 4 | 573a13faf29313caabde | Karol, an ordinary young m | Array[1] | 90 | | The Red Spider | 0 |
| 5 | 573a13faf29313caabde | A-Gu enlists a group of con | Array[1] | 110 | | The Laundryman | 1 |

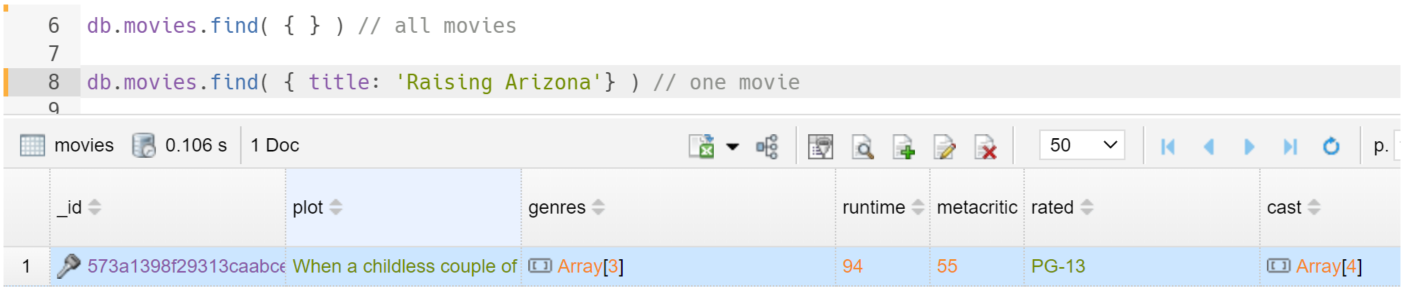
16. Similar to MySQL, you can 'Run' all, or a selection. Change the script to database.movies.find() (as shown below), then click on 'Run Selected..'


Run
Debug
Stop
Import
Export
Run the entire script (Ctrl+↲ | Ctrl+F5)
Clone Tab and Run the Script
Run and Expand Whole Tree
Run Selected Text/Current Statement (F6 | Ctrl+F6)
Clone Tab and Run Selected Text/Current Statement

```
db.movies.find( { } ) // all movies
```

Note the comment: // all movies - this is JavaScript (see the screenshot above).

17. Find your favourite movie e.g.



```
db.movies.find( { } ) // all movies

db.movies.find( { title: 'Raising Arizona'} ) // one movie
```

movies 0.106 s 1 Doc

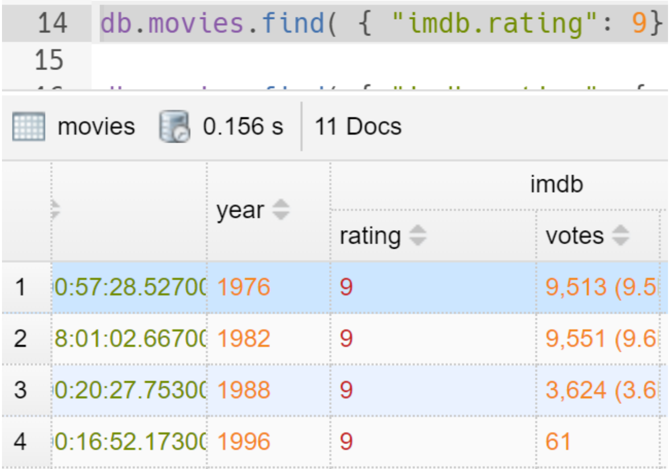
| | _id | plot | genres | runtime | metacritic | rated | cast |
|---|---|---|---|---|---|---|---|
| 1 | 573a1398f29313caabce | When a childless couple of | Array[3] | 94 | 55 | PG-13 | Array[4] |

Note the key:value syntax e.g. title: 'Raising Arizona'

**18. Find all 'G' rated movies. How many are there?**

19. Answer question 18 using coding. Hint: replace 'find' with something else.

20. Notice that some fields have complex structure, here's one way to use imdb rating:



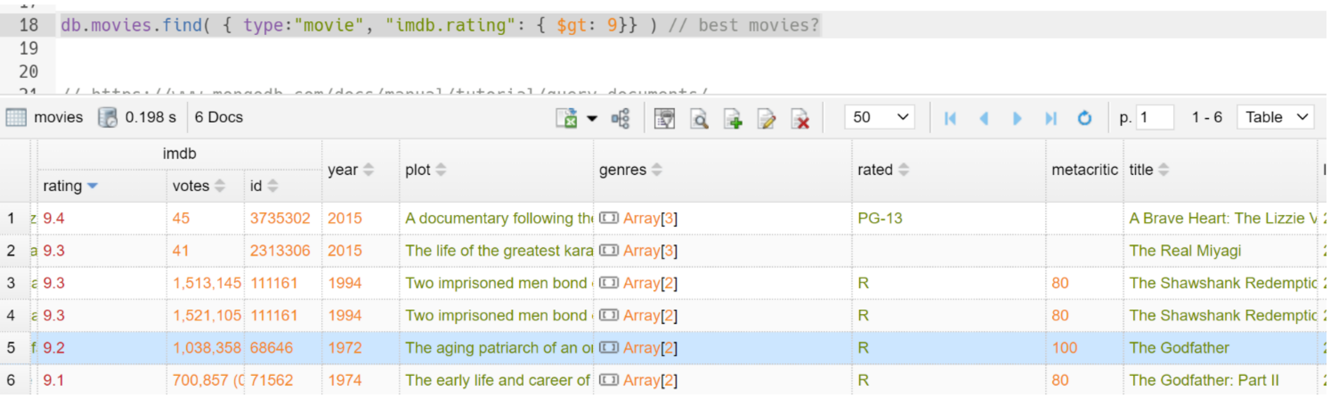
```
db.movies.find( { "imdb.rating": 9}
```

movies 0.156 s 11 Docs

| | | year | imdb | |
|---|---|---|---|---|
| | | | rating | votes |
| 1 | 0:57:28.5270( | 1976 | 9 | 9,513 (9.5 |
| 2 | 8:01:02.6670( | 1982 | 9 | 9,551 (9.6 |
| 3 | 0:20:27.7530( | 1988 | 9 | 3,624 (3.6 |
| 4 | 0:16:52.1730( | 1996 | 9 | 61 |

**21. How many shows have a rating of 2?**

22. How to find the movies with best 'rating'?



```
db.movies.find( { type:"movie", "imdb.rating": { $gt: 9}} ) // best movies?


// https://www.mongodb.com/docs/manual/tutorial/query-documents/
```

movies 0.198 s 6 Docs 50 p. 1 1 - 6 Table

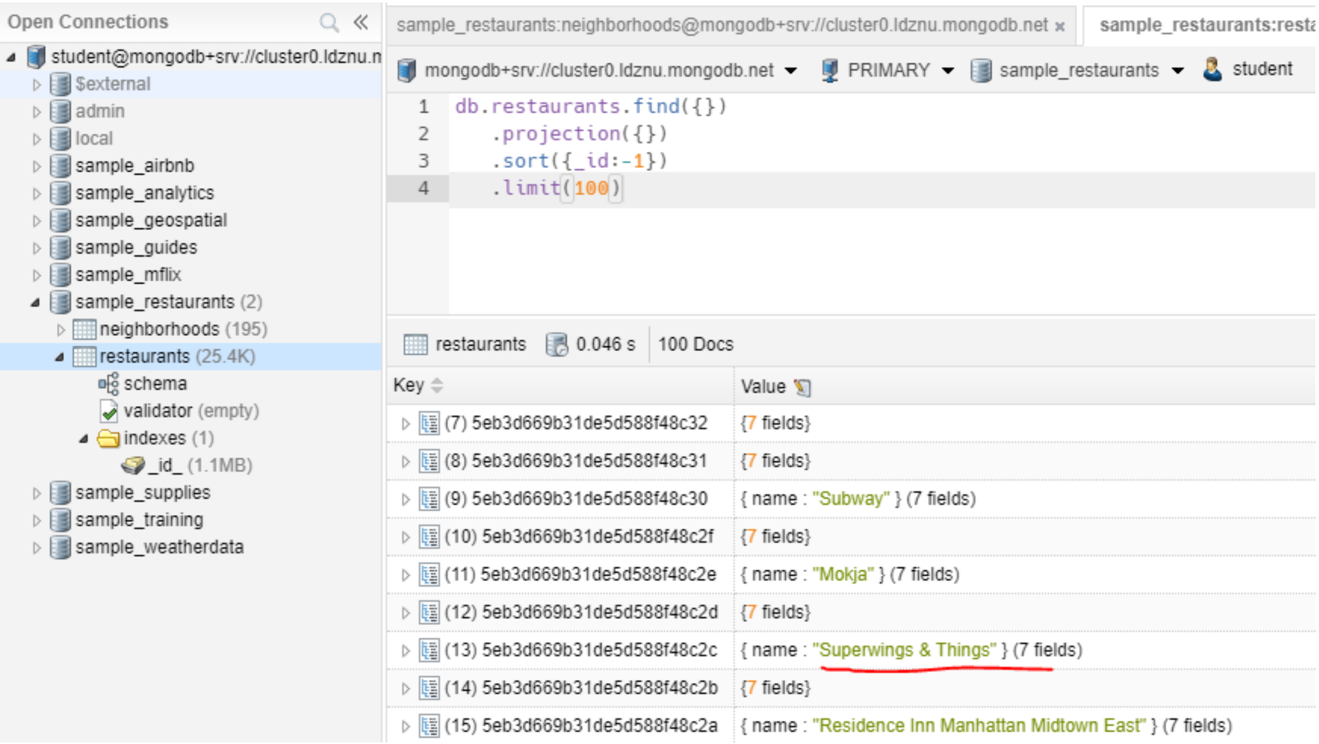
| | | imdb | | | year | plot | genres | rated | metacritic | title |
|---|---|---|---|---|---|---|---|---|---|---|
| | | rating | votes | id | | | | | | |
| 1 | z | 9.4 | 45 | 3735302 | 2015 | A documentary following th | Array[3] | PG-13 | | A Brave Heart: The Lizzie V |
| 2 | a | 9.3 | 41 | 2313306 | 2015 | The life of the greatest kara | Array[3] | | | The Real Miyagi |
| 3 | a | 9.3 | 1,513,145 | 111161 | 1994 | Two imprisoned men bond | Array[2] | R | 80 | The Shawshank Redemptic |
| 4 | a | 9.3 | 1,521,105 | 111161 | 1994 | Two imprisoned men bond | Array[2] | R | 80 | The Shawshank Redemptic |
| 5 | f | 9.2 | 1,038,358 | 68646 | 1972 | The aging patriarch of an o | Array[2] | R | 100 | The Godfather |
| 6 | | 9.1 | 700,857 (( | 71562 | 1974 | The early life and career of | Array[2] | R | 80 | The Godfather: Part II |

**What is $gt?**

**23. Why are there TWO Shawshank Redemptions?**

**24. Who features in two of the worst movies (by imdb rating)?**

25. In the left menu find the database sample_restaurants, expand it and click on restaurants. Change the script to database.restaurants.find() (as shown below), then click on 'Run Selected..'

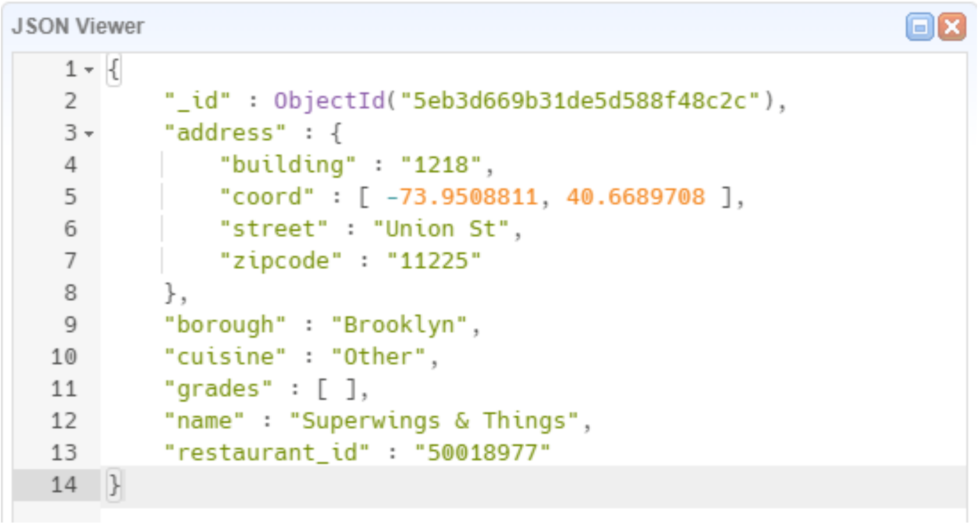


```
db.restaurants.find({})
   .projection({})
   .sort({_id:-1})
   .limit(100)
```

restaurants 0.046 s 100 Docs

| Key | Value |
|---|---|
| (7) 5eb3d669b31de5d588f48c32 | {7 fields} |
| (8) 5eb3d669b31de5d588f48c31 | {7 fields} |
| (9) 5eb3d669b31de5d588f48c30 | { name : "Subway" } (7 fields) |
| (10) 5eb3d669b31de5d588f48c2f | {7 fields} |
| (11) 5eb3d669b31de5d588f48c2e | { name : "Mokja" } (7 fields) |
| (12) 5eb3d669b31de5d588f48c2d | {7 fields} |
| (13) 5eb3d669b31de5d588f48c2c | { name : "Superwings & Things" } (7 fields) |
| (14) 5eb3d669b31de5d588f48c2b | {7 fields} |
| (15) 5eb3d669b31de5d588f48c2a | { name : "Residence Inn Manhattan Midtown East" } (7 fields) |

26. Find the restaurant 'Superwings and Things". Move your mouse pointer to the restaurant name, you will see a JSON pop-up describing the object. Press p to see the code in JSON viewer.

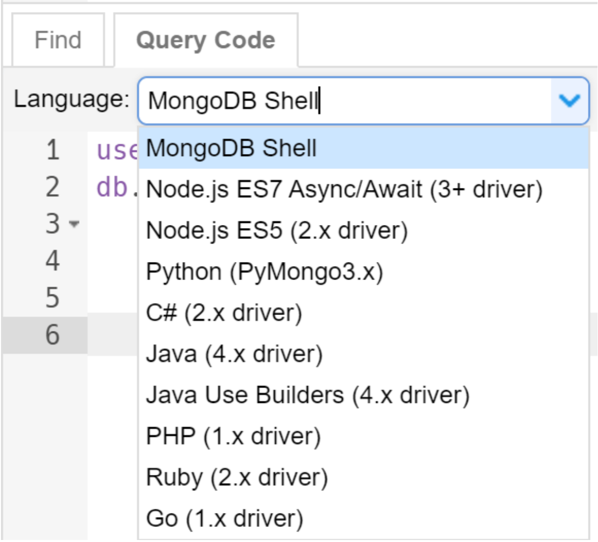


```
{
    "_id" : ObjectId("5eb3d669b31de5d588f48c2c"),
    "address" : {
        "building" : "1218",
        "coord" : [ -73.9508811, 40.6689708 ],
        "street" : "Union St",
        "zipcode" : "11225"
    },
    "borough" : "Brooklyn",
    "cuisine" : "Other",
    "grades" : [ ],
    "name" : "Superwings & Things",
    "restaurant_id" : "50018977"
}
```

27. **You need to write code adding more attributes to this record.** Note, your connection is read-only so you cannot add your changes to MongoDB.

**You need to add details of the signature dish:**
**name "Sticky BBQ wings"**
**diet: Gluten-free, Halal**

**For the programmers (and the curious), explore the various language options:**



Language: MongoDB Shell

You can save your script (as .js)

# D. Exercise. Choosing a NoSQL database

Libraries store information about their collections in their catalogue See example on the next page).

Match each of the following statements to the type of NoSQL database that would be best for storing that library's data. Select from the four types of NoSQL database discussed in lectures.

a. In one library, items are catalogued by author, title and publisher, as well as any number of other fields chosen by the cataloguer, such as physical description, subject codes and notes.

b. In another library, each catalogue record is stored in the MARC format (Figure 1), a coded text format that contains all the catalogue information for a particular item.

c. A public library wishes to store cover photos of all its items, which might be in JPEG, PNG or PDF format, or stored as a URL.

d. A university library wishes to keep track of which published academic papers reference each other in order to help researchers measure their metrics.

```
LEADER 00000nam  22000001  4500
008    730220s1955    ilu      b    00000 eng
019    55007351
050 0  QA276.5|b.R3
082    311.22
110 20 Rand Corporation.
245 12 A million random digits|bwith 100,000 normal deviates.
260 0  Glencoe, Ill.,|bFree Press|c[1955]
300    xxv, 400, 200 p.|c28 cm.
504    Bibliography: p. xxiv-xxv.
650  0 Numbers, Random.
984    |cMS T 519 R152
```

*Figure 1: An example of a MARC record. MARC is a very old format that predates NoSQL, JSON and even XML by several decades, yet it remains the industry standard in library data systems.*