



Querying MongoDB

Practice Works

ESSEC

nicolas.travers (at) cnam.fr

1 MongoDB	3
1.1 Installation	3
1.1.1 Windows	3
1.1.2 Linux and MacOS	3
1.2 Import datafile	3
1.3 Robomongo	4
2 MongoDB	5
2.1 Practice Work's data set	5
2.2 Querying MongoDB collections	6
2.2.1 MongoDB simple queries	6
2.2.2 Distinct queries	6
2.2.3 Aggregates	7
2.3 MongoDB Updates	8
2.3.1 Examples	8
2.3.2 Update queries	8
2.4 Indexing in MongoDB	9
2.4.1 Indexing single attributes	9
2.4.2 2Dsphere indexes	9

The MongoDB¹ database server can be installed in a local or distributed environment. You can download it here: <http://www.mongodb.org/downloads/> (Find some tips on Moodle or online)

1.1 Installation

Download and install MongoDB² ;

1.1.1 Windows

Open a first shell cmd

Extract archive on « C:\ » (for example)

```
> cd /
```

```
> md data (create the data folder)
```

```
> md data\db
```

Launch server: (Do not close it afterwards!!!!)

```
> C:\mongodb\bin\mongod.exe
```

Or

```
> C:\mongodb\bin\mongod.exe --dbpath "your own data folder"
```

Client shell:

```
> C:\mongodb\bin\mongo.exe
```

1.1.2 Linux and MacOS

Open a first shell cmd:

```
> tar zxvf mongodb-xxx.tgz (xxx:version)
```

```
> mv mongodb-xxx /your-own-folder
```

```
> ln -s mongodb-xxx mongodb (create a symbolic link)
```

```
> sudo mkdir /data/db (create the data folder)
```

launch server:

```
> bin/mongod (with /data/db)
```

Or

```
> bin/mongod --dbpath /your-own-data-folder
```

Launch a second shell cmd:

```
> bin/mongo
```

1.2 Import datafile

Parameters:

¹<http://www.mongodb.org/>

²Documentation : <http://docs.mongodb.org/manual/contents/>

- localhost:27017 is the mongodb server address
- DBLP: is the database name
- publis: is the collection name
- dblp.json: is the file name to import from

```
mongoimport --host localhost:27017 --db DBLP --collection publis dblp.json
```

where:

1.3 Robomongo

Robomongo³ is UI client for Mongo. You can create a connection to the default port 27017 for MongoDB.

Then, you can query the database directly by selecting the DBLP.publis collection.

³<https://robomongo.org/>

To use MongoDB, the server has to be launched in a first command tool. The MongoDB console has to be launched in a second command tool. All the related instructions (installation, launcher, console, querying) are described in the guide available on the Moodle.

2.1 Practice Work's data set

2.1.1 Download and unzip DBLP.zip (available at chewbii.com, section “MongoDB” → “Data from DBLP”),

2.1.2 Import the data file in the “DBLP” database and “publis” collection (cf. Guide),

2.1.3 In the MongoDB console, connect to the DBLP database and count the number of stored documents.

This dataset is an automatic extraction from DBLP, a website which stores all the research publications in computer science. You have here a sample of 118k documents.

Note that some keys can be available or not (“booktitle”, “cites”...), some others can change of data types(“pages”).

```
{
  "_id" : "books/daglib/0025185",
  "type" : "Book",
  "title" : "MongoDB - The Definitive Guide: Powerful and Scalable Data Storage.",
  "pages" : "I",
  "publisher" : "O'Reilly",
  "year" : 2010,
  "authors" : [ "Kristina Chodorow", "Michael Dirolf" ],
  "isbn" : [ "978-1-449-38156-1" ]
}
```

```
{
  "_id" : "conf/icod/BouzeghoubG83",
  "type" : "Article",
  "title" : "The Design of an Expert System for Database Design.",
  "booktitle" : "ICOD-2 Workshop on New Applications of Data Bases",
  "pages" : { "start" : 203, "end" : 223 },
  "year" : 1983,
  "url" : "db/conf/icod/icod83w.html#BouzeghoubG83",
  "authors" : [ "Mokrane Bouzeghoub", "Georges Gardarin" ],
  "cites" : [ "journals/tods/AstrahanBCEGGKLMPTWW76", "books/bc/AtzeniA93",
    "journals/tcs/AtzeniABM82", "journals/jcss/AbiteboulB86" ]
}
```

You can find the following keys:

- Conference and scientific journal are of “type” “Article”, with a “title” and sometimes “booktitle”,
- Book chapters are of “type” “Book” with a title and booktitle,
- PhD thesis are of “type” : “Phd”,
- Every publications are linked to a list of authors and sometimes publisher (also called editor),
- Some of them can embed a “cites” key with a list of references

2.2 Querying MongoDB collections

2.2.1 MongoDB simple queries

For each of the following sentence, give the corresponding query in the MongoDB API:

2.2.1 All the book titles (type 'Book'),

2.2.2 All the publications since 2011,

2.2.3 All the books since 2014,

2.2.4 Titles of publications with an existing publisher,

2.2.5 Titles of *Jeffrey D. Ullman*'s publications,

2.2.6 Give those in which he is the first author,

2.2.7 And when he is the only author,

2.2.8 List of all publications which title contains the word "database",

2.2.2 Distinct queries

2.2.1 List of distinct publishers,

2.2.2 List of distinct authors,

2.2.3 Aggregates

2.2.1 All *Jeffrey D. Ullman*'s publications, projected the title and sorted by the starting page,

2.2.2 Count how much publications he published per year,

2.2.3 Count how much publications he published,

2.2.4 For each author, count the number of its publications. Sort the result in descending order,

2.2.5 For each publisher (if exists) and year, give the number of publications

2.2.6 For each publisher (if exists), give the average rate of publications per year. Sort the result decreasingly.

2.3 MongoDB Updates

Store documents can be updated in the database. You have to use the MongoDB API to query the required documents.

update function: 1st param is for mapping, 2nd param is for setting (\$set, \$unset).

delete function: 1st param is for mapping.

2.3.1 Examples

Give a sentence corresponding to the following update queries:

- `db.publis.update ({ "authors" : "Jeffrey D. Ullman" }, { $set : { "label" : "Ishida" } });`
- `db.publis.remove({ "Title" : "Parallel, Distributed and Multiagent Production Systems" })`

2.3.2 Update queries

For each of the following sentence, query the updating query. Verify the targeted information before the update.

2.3.1 Update all books containing “database” by adding the label attribute “database”,

2.3.2 Update all publications with a publisher containing “ACM” by removing the “number” attribute,

2.3.3 Remove all publications without any authors,

2.3.4 Update all publications to add a “pp” field which corresponding to the number of pages of the publication.

2.4 Indexing in MongoDB

Indexing attributes enables fast querying on this attribute.

2.4.1 Indexing single attributes

2.4.1 For the ”Jeffrey D. Ullman” query, generates the execution plan by adding “.explain()”,

2.4.2 Add an index on authors: “db.publis.createIndex({ "authors" :1 });”,

2.4.3 Repeat the query with the explain plan. You will obtain a new one which uses the created index (more efficient).