



CentraleSupélec

CENTRALESUPÉLEC  
2020-2021

---

Big Data, Techniques and Platforms

## Document Databases

---

For this assignment you are required to use MongoDB. The only necessary file is:

- `tourpedia_paris.json`

The file comes from the <http://wafi.iit.cnr.it/openervm/about/index.html> project, the Wikipedia of Tourism. It contains information about accommodations, restaurants, points of interest and attractions of different places in Europe. We will focus on Paris.

Download the files and store them in a directory of your computer. I suggest you to create again a directory called `data`.

### 1 IMPORT DATA

Import data in your MongoDB database, and take advantage of the Studio3T graphical interface.

### 2 QUERYING DATA

1. **[1 point]** Before starting with the queries look at data and provide a short description of them: the most common structure of the documents (the most present attributes, nested documents, etc.).

Then **provide the queries** (in the form you must put the queries) that give you the following informations:

2. **[1 point]** Give the operation that retrieves one element of the collection and provide a short description of the element: the attributes, and the types that are present.

3. **[1 point]** Retrieve the name and the contact phone of any location for which the phone number is provided.
4. **[1 point]** Show the list of structures that appear to be McDonald's. **Hint:** Looking at the names can help.
5. **[1 point]** Show the number of structures that are returned by the previous query.
6. **[1 point]** Provide the names of the structures that offer the chauffage among their services.
7. **[1 point]** Give the list of structures that provide less than 3 services.
8. **[1 point]** Give the list of structures that provide more than 15 services.
9. **[2 points]** Show the structure (or the list of structures if many) with the greatest number of services.
10. **[2 points]** Provide the query that gives the list of structures that do not provide services.
11. **[1 point]** Comment about how these last structures are represented in the database: is the fact that a structure does not provide services matching the best practices of MongoDB data modeling? Why?
12. **[2 points]** Write the query that shows the list of the names of the structures having an average rate greater than 4.
13. **[3 points]** Give the query that returns the number of locations having accomodation as category and that provide, among their services bar.
14. **[2 points]** Give the query that returns the number of reviews by category and language.
15. **[Bonus]** Show the date (if any) in which the system received the most number of reviews.