



CentraleSupélec

CENTRALESUPÉLEC  
2021-2022

---

Big Data, Techniques and Platforms

## Document Databases

---

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

- `artistsnestedsongs.json`

This file contains the same informations you used in the previous exercise of Spark/MapReduce. 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

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 that give you the following informations and the returned result.

1. **(1 point)** - Retrieve one element of the collection.
2. **(1 point)** - Retrieve the artist name (`Artist` attribute) and the main genre of any artist for which the `link` to the webpage is provided.
3. **(1 point)** - Show the list of artist (just their name) that contain `Stones` in the name.

4. **(1 point)** - Show the number of artists that are returned by the previous query.
5. **(1 point)** - Provide the name of the artists that have popularity greater than 4.
6. **(1 point)** - Show the artists with the lowest popularity.
7. **(1 point)** - Give the list of artists that have at list 10 songs listed.
8. **(1 point)** - Give the artists (if any) that do not have songs listed.
9. **(3 point)** - Give the list of artists that have at list 10 distinct songs listed and the criteria that you used for this distinct operation.
10. **(1 point)** - Show the complete set of genres for 40 artists.
11. **(1 point)** - Comment on how the list is specified and on how it is possible to query this attribute.
12. **(3 points)** - Give the number of artists having Folk among the genres.
13. **(3 points)** Give the number of song by main genre of the artist and language.

### 3 BONUS - ENJOY WITH DATA

Now you can give freedom to your self-initiative and fantasy.

Using Spark, Dataframes and/or MongoDB provide one analysis of your choice on the dataset.