# PROJECT IN BIG DATA 2 REPORT

Managing Databases with Sql and NoSql:
*A Study of the French Administration*

Fiona Gérent  Axelle Renaud  Yaelle Coudray  Souleymane Faye  Max Michel Koffi

**Magistère Ingénieur Economiste 1**

# Summary

**Abstract**

In this report, we provide a model for the French Administration. First, we study it graphically by making CDM, LMD and data dictionaries. Then, we take our data in an SQL interface to provide queries. We thus make json files of our data with python in order to query it with MongoDB.

# 1 Introduction

We have chosen to study the organization of the French administration. This project will allow us to grasp the relationships between political figures, between the different organizations, to draw up a representation of the political diversity of the major bodies of the administration during a given period. The database contains the central administration, the territorial administration of the state and includes the institutions, organizations as well as the political and administrative members attached. We'll present it with bSQL and NoSQL programming.

*For ease of reading, the code will, in this report, be provided in purple.*

# 2 Model

## 2.1 Data Dictionary

**Fig. 1: Data Dictionary for the *French Administration Model*[1]**

| Field Name | Table(s) | Data Format | Data Type | Field Size |
|---|---|---|---|---|
| Administrative center | regions | | varchar(20) | 10 |
| Age | head of state; minister; president of the departmental council; president of the region; prime minister; deputies; senators | NN | int(20) | 70 |
| Area | regions | NNNNN | int(20) | 10 |
| Budget | ministries | NNNNNNNN | int(50) | 10 |
| Capital | country | | varchar(20) | 10 |
| City | departmental council; regional council | | varchar(20) | 20 |
| Date (*primary*) | In all tables | NNNN | int(20) | |

[1] A primary key variable is not the primary key for all tables where it is included, it may be a *foreign key* for some. Thus, we marked in red tables for which the variable is actually a primary key.

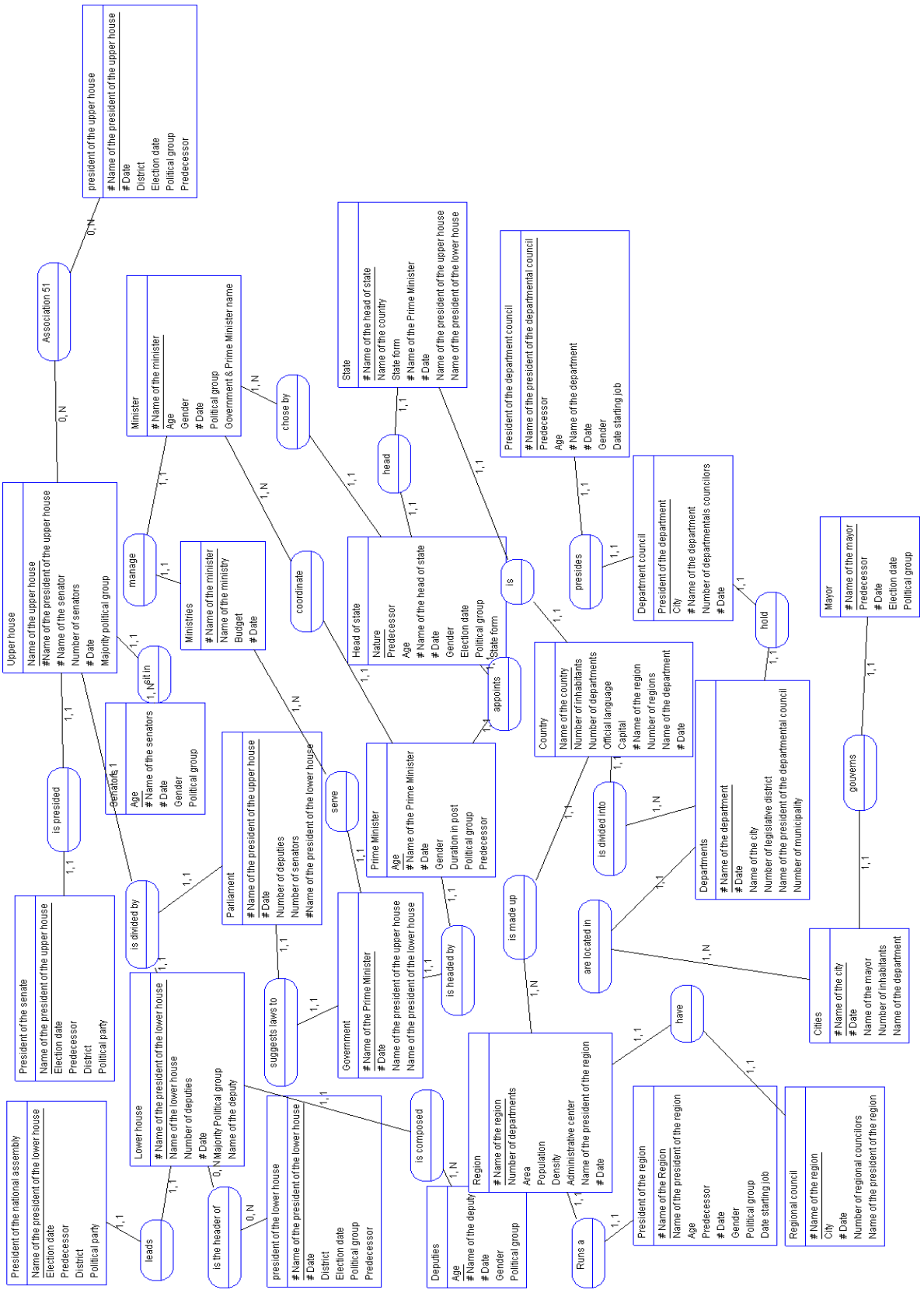
| Field Name | Table(s) | Data Format | Data Type | Field Size |
|---|---|---|---|---|
| Date starting job | president of the departmental council ; president of the region | YYYY/MM/DD | date | 20 |
| Density | regions | NN | int(20) | 10 |
| District | president of the lower house ; president of the upper house[2] | | varchar(40), varchar(20) | 20 |
| Duration in post | prime minister | | varchar(30) | 10 |
| Election date | head of state; mayor; president of the lower house; president of the upper house | YYYY/MM/DD | date | 40 |
| Gender | head of State; minister; president of the departmental council ; president of the region ; prime minister; deputies; senators | | varchar(20) | 60 |
| Government & Prime Minister name | minister | | varchar(20) | 10 |
| Majority political group | lower house ; upper house | | varchar(20) | 20 |
| Name of the city (*Primary*) | cities | | varchar(20) | 10 |
| Name of the country | country; state | | varchar(20) | 20 |
| Name of the department (*Primary)* | country; president of the departmental council; departmental council | | varchar(20); varchar(50) (for departmental council) | 30 |
| Name of the deputy | lower house; deputies | | varchar(20) | 20 |

[2] For some variables, their types are not the same for each table. Hence, we marked in green the table concerned by the other type.

| Field Name | Table(s) | Data Format | Data Type | Field Size |
|---|---|---|---|---|
| Name of the head of state (*Primary*) | head of state; state | | varchar(20); varchar(50) (for state) | 20 |
| Name of the lower house | lower house; state | | varchar(20); varchar(50) | 20 |
| Name of the mayor (*Primary*) | mayor, cities | | varchar(50) | 20 |
| Name of the minister (*Primary*) | minister; ministries | | varchar(50) | 20 |
| Name of the ministry | ministries | | varchar(60) | 10 |
| Name of the president of the departmental council | president of the departmental council | | Text | 10 |
| Name of the president of the lower house (*primary*) | government ; lower house; parliament; president of the lower house | | varchar(40); varchar(50) (for parliament) | 40 |
| Name of the president of the region | president of the region; regional council ; regions | | varchar(20) | 30 |
| Name of the president of the upper house (*Primary*) | government ; parliament; president of the upper house; upper house | | varchar(20); varchar(50) (for president of the upper house) | 40 |
| Name of the Prime Minister (*Primary*) | government; prime minister; state | | varchar(20); varchar(50) ( for state) | 30 |
| Name of the region (*Primary*) | country; president of the region; regional council; regions | | varchar(40); varchar(20) (for president of the region and regions) | 40 |
| Name of the senator (*Primary*) | upper house; senators | | varchar(50) | 20 |
| Name of the upper house (*Primary*) | state; upper house | | varchar(50) | 20 |
| Nature | head of state | | varchar(50) | 10 |

| Field Name | Table(s) | Data Format | Data Type | Field Size |
|---|---|---|---|---|
| Number of departments | country; regions | N | int(20) | 20 |
| Number of departmentals councilors | departmental council | | int(20) | 10 |
| Number of deputies | lower house; parliament | NNN | int(20) | 20 |
| Number of inhabitants | cities; country | NNNNNNNN | int(20); int(40) | 20 |
| Number of legislative district | departments | NN | int(20) | 10 |
| Number of municipality | departments | NNN | int(20) | 10 |
| Number of regional councilors | regional council | NNN | int(20) | 10 |
| Number of senators | parliament; upper house | NNN | int(20); varchar(20) | 20 |
| Number of regions | country | NN | int(20) | 10 |
| Official language | country | | varchar(20) | 10 |
| Political group | head of state; minister ; president of the lower house; president of the region ; President of the upper house ; Prime minister; deputies; mayor; senators | | varchar(20); varchar(50); varchar(40) | 90 |
| Population | regions | NNNNNN | int(20) | 10 |
| Predecessor | head of state; mayor; president of the lower house; president of the region; president of the upper house; prime minister | | varchar(20); varchar(50) | 60 |
| President of the department | departmental council | | varchar(50) | 10 |
| State form | head of state; state | | varchar(40); varchar(50) | 20 |

## 2.2 Conceptual Data Model



## 2.3 Logical Data Model

Here is the Logical Data Model of our study:

- ➔ president of the lower house (Name of the president of the lower house, Date,Election date, Predecessor, District, Politic group)
- ➔ lower house (Name of the lower house, Date, Number of deputy, Majority political group, #Name of the president of the lower house,Date, #Name of the deputy,Date)
- ➔ -deputies (Name of deputy, Date,Age, Gender, Political group)
- ➔ president of the upper house (Name of the president of the upper house,Date ,Election date, Predecessor, District, Political group)
- ➔ upper house (Name of the upper house,Date,Majority political group, #Name of the president of the upper house,Date, #Name of the senator,Date, Number of senators)
- ➔ senators (Name of the senator,Date,Gender, Political group, Age)
- ➔ parliament (Name of the president of the upper house,Date,Number of deputies, Number of senators,#Name of the president of the lower house,Date)
- ➔ government (Name of the Prime Minister, Date,#Name of the president of the lower house,Date, #Name of the president of the upper house,Date)
- ➔ ministries (Name of the ministry,Date, #Name of the minister,Date, Budget)
- ➔ minister (Name of the minister, Date, Gender, Age, Political group, Government and Prime Minister name)
- ➔ prime minister (Name of the Prime Minister, Date ,Political group, Gender, Age, Duration in post, Predecessor)
- ➔ head of state (Name of the head of state,Date, State form, Nature, Political group, Election Date, Predecessor, Gender, Age)
- ➔ state (Name of country,Date, State form, #Name of the head of state,Date #Name of the Prime Minister,Date, #Name of the upper house,Date, Name of the lower house)
- ➔ regions (Name of the region, Date, Number of department, Administrative center, Area, Population, Density, #Name of the president of the region,Date)
- ➔ president of the region (Name of the president of the region, Date, Date starting job, Gender, Political group, Age, Predecessor, #Name of the region,Date)
- ➔ regional council (Name of the president of the region, Date, Number of regional councillors, City, Name of the region)
- ➔ country (Name of the country, Date, Number of inhabitant, Number of regions, Number of departments, official language, Capital, #Name of the region,Date, # Name of the department,Date)
- ➔ departments (Name of the department, Date,Number of municipality, Number of legislative district, #Name of the president of the department council,Date,#Name of the city,Date)
- ➔ cities (Name of the city, Date, #Name of the mayor,Date, Number of inhabitants, #Name of the department,Date)
- ➔ department council (President of the department, Date, Number of departmentals councillors, City, #Name of the department,Date)
- ➔ mayor (Name of the mayor, Date,Election date, Political group, Predecessor)
- ➔ president of the department council (Name of the president of the department council, Date,Date Starting job, Predecessor, Gender, Age, #Name of the department,Date)

# 3 Queries

## 3.1 SQL queries

### 3.1.1 Basic queries

-Display the cities names
-Display ,without duplicate,the different political groups of ministers
-Display ,without duplicate,the different state form
- Display the different names of department
- Display the different names of ministry

### 3.1.2 Where clause queries

- Display the name of the president of the department in Lille.
- Display the name of the president of the lower house, who has more deputies in his assembly
- Display the name of the prime minister who stayed the shortest time in office
- Display the name of the senator of the political party: UMP
- Display the cities where the number of inhabitants is more than 150000

### 3.1.3 Order by queries

-Display the name of ministers in alphabetical order.
-Display the number of deputies by date in descending order.
-Display from the youngest to the oldest the name of the head of government when they started their term
-Display the Name of the Region by area in ascending order
-Display the Name of the ministry by budget in descending order

### 3.1.4 Multi-Table queries

-Display for each city,the political group of his actual mayor
- Display the name of the head of the government, the name of the president of the lower house and the upper house for 2020
-Display the differents names of predecessors of president of upper house for which the number of senators in the upper house is more than 325
-Display the differents names of the head of state for which the number of inhabitant is more than 55000000
-Display for each city,the predecessor of his actual mayor

### 3.1.5 Queries with numeric expressions and functions

-Display the average budget of the ministries
-Display the region with the highest density
-Display the population growth rate between 2020 and 1880
-Display the number of times where Gerard Larcher has been president of the upper house
-Display the older minister

### 3.1.6 Group by queries

-Display the average age of  ministers by political group
-Display the people who have been president of the upper house more than once
-Display the average number of senators by president of the upper house
-Display for each political group,the number of times where the head of state has been in this political group
- Display the average age of  deputies by gender

### 3.1.7 Nested queries

-Display the number of inhabitants in the cities where the political group of the mayor is 'RPR'
-Display the number of regional councillors in the regions where there are more than 6 departments
-Display the budget of ministries where the age of minister is more than 58
-Display the area of regions where the number of regional councillors is more than 100
- Display the name of the president of the upper house where the majority political group in the upper house is the same from the political group of the president of the upper house

## 3.2 SQL solutions

Here, we provide the solutions of  queries and recalling the type of query.

### **3.2.1** Solutions for the 5 ORDER BY queries

- Display the name of ministers  in alphabetical order.

```
SELECT `Name of the minister` FROM `minister` ORDER BY `Name of the minister`;
```

-Display the number of deputies by date in descending order.

```
SELECT `Number of deputy`, `Date` FROM `lower house` ORDER BY `Number of deputy`DESC;
```

- Display from the youngest to the oldest the name of the head of government when they started their term

```
SELECT `Name of the head of government`, `Age` FROM `head of government` ORDER BY Age;
```

- Display the Name of the Region by area in ascending order

```
SELECT `Name of the region` FROM `regions` ORDER BY Area;
```

- Display the Name of the ministry by budget in descending order

```
SELECT `Name of the ministry`, `Budget` FROM `ministries` ORDER BY Budget;
```

### **3.2.2** Solutions for the 5 Multi-Table queries

- Display for each city,the political group of his  actual mayor

```
SELECT `Name of the city`, `Political group` FROM cities, mayor WHERE cities.`Name of the mayor`=mayor.`Name of the mayor`;
```

- Display the name of the head of the government, the name of the president of the lower house and the upper house for 2020

SELECT `Name of the head of state`, `Name of the president of the lower house`,`Name of the president of the upper house` FROM `head of state`, `government` WHERE `head of state`.`Date`=`government`.`Date` AND `government`.`Date`=2020;

- Display the differents names of predecessors of president of upper house for which the number of senators in the upper house is more than 325

SELECT DISTINCT `Predecessor`, `Number of senators`FROM `president of the upper house`, `upper house` WHERE `president of the upper house`.`Name of the president of the upper house`=`upper house`.`Name of the president of the upper house`and `upper house`.`Number of senators`>325

- Display the differents names of the head of state for which the number of inhabitant is more than 55000000

SELECT `Name of the head of state`,`Number of inhabitant` FROM `head of state`,`country` WHERE `head of state`.`Date`=`country`.`Date`and `country`.`Number of inhabitant`>55000000;

- Display for each city,the predecessor of his actual mayor

SELECT `Name of the city`,`Predecessor` FROM `cities`,`mayor` WHERE `cities`.`Name of the mayor`=`mayor`.`Name of the mayor`;

### **3.2.3** Solutions for the 5 Queries with Numeric expressions and functions

- Display the average budget of the ministries

SELECT AVG(Budget) FROM `ministries`;

- Display the region with the highest density

SELECT `Name of the region`, max(Population/Area) as Max_Density FROM `regions`;

- Display the population growth rate between 2020 and 1880

SELECT (POWER(MAX(`Number of inhabitant`)/ MIN(`Number of inhabitant`),(1/140)) - 1)*100 as growth_rate FROM `country`;

-Display the number of times where Gerard Larcher has been president of the upper house

SELECT COUNT(`Name of the president of the upper house`) as Count_Gérard_Larcher FROM `president of the upper house` WHERE `Name of the president of the upper house`='Gérard Larcher'

- Display the older minister

SELECT `Name of the minister`,MAX(`Age`) FROM `minister`;

### **3.2.4** Solutions for the 5 GROUP BY queries

- Display the average age of ministers by political group

SELECT `Political group`, AVG(`Age`) as Average_Age_Minister FROM `minister` GROUP BY `Political group`;

- Display the people who have been president of the upper house more than once

SELECT `Name of the president of the upper house`,COUNT(*) FROM `president of the upper house` GROUP BY `Name of the president of the upper house` HAVING COUNT(`Name of the president of the upper house`)>1;

- Display the average number of senators by president of the upper house

SELECT `Name of the president of the upper house`,AVG(`Number of senators`) FROM `upper house` GROUP BY `Name of the president of the upper house`;

- Display for each political group,the number of times where the head of state has been in this political group

SELECT COUNT(*), `Political group` FROM `head of state` GROUP BY `Political group`;

- Display the average age of deputies by gender

```
SELECT `gender`,AVG(`Age`) as Average_Age_Deputies FROM `deputies` GROUP BY
`Gender`;
```

### **3.2.5** Solutions for the 5 nested queries

-Display the number of inhabitants in the cities where the political group of the mayor is 'RPR'

```
SELECT `Name of the city`, `Number of inhabitants`FROM `cities` WHERE `Name of the
mayor`IN (SELECT `Name of the mayor`FROM `mayor`WHERE `Political group`= 'RPR');
```

-Display the number of regional councillors in the regions where there are more than 6 departments

```
SELECT `Name of the region`, `Number of regional councillors` FROM `regional council`
WHERE `Name of the president of the region` IN (SELECT `Name of the president of the region`
FROM `regions` WHERE `Number of department`>6);
```

-Display the budget of ministries where the age of minister is more than 58

```
SELECT `Name of the ministry`,`budget`FROM `ministries` WHERE `Name of the minister`IN
(SELECT `Name of the minister`FROM `minister`WHERE `Age`>58);
```

-Display the area of regions where the number of regional councillors is more than 100

```
SELECT `Name of the region`,`Area`FROM `regions` WHERE `Name of the region` IN
(SELECT `Name of the region`FROM `regional council`WHERE `Number of regional
councillors`>100);
```

-Display the name of the president of the upper house where the majority political group in the upper house is the same from the political group of the president of the upper house

```
SELECT `Name of the president of the upper house`, Date FROM `upper house` WHERE
`Majority political group` in (SELECT DISTINCT `Political group` FROM `president of the upper
house`)
```

### **3.2.6** Solutions for the 5 basic queries

- Display the cities names

```
SELECT `Name of the city` FROM `cities`;
```

- Display ,without duplicate,the different political groups of ministers

```
SELECT Distinct `Political group` FROM `minister`;
```

- Display ,without duplicate,the different state form

```
SELECT Distinct `state form` FROM `state`;
```

- Display the different names of department

```
SELECT `Name of the department` FROM `departments`;
```

- Display the different names of ministry

```
SELECT `Name of the ministry` FROM `ministries`;
```

### **3.2.7** Solutions for the 5 WHERE clause queries

- Display the name of the president of the department in Lille.

```
SELECT `President of the department` FROM `departmental council` WHERE city= 'Lille';
```

- Display the name of the president of the lower house, who has more deputies in his assembly

```
SELECT `Name of the president of the lower house`, `Number of deputy` FROM `lower house`
WHERE `Number of deputy` =(SELECT MAX(`Number of deputy`)FROM `lower house`)
```

- Display the name of the prime minister who stayed the shortest time in office

SELECT `Name of the Prime Minister` FROM `prime minister` WHERE `Duration in post`=(SELECT MIN(`Duration in post`) FROM `prime minister`);

- Display the name of the senator of the political party : UMP

SELECT `Name of the senator`, `Political group` FROM `senators` WHERE `Political group`= 'UMP';

- Display the cities where the number of inhabitant is more than 150000

SELECT `Name of the city`, `Number of inhabitants` FROM `cities` WHERE `Number of inhabitants` >150000;

# **4** MongoDB

In this part, we chose the *government*, *cities* and *state* tables.

## 4.1 MongoDB: Mongo import

### 4.1.1 Command prompt

We decided to work on a local host, using the Windows Command Prompt to import the 3 json files in MongoDB

> cd C:\Program Files\MongoDB\Server\5.0\bin

When we're located where our Mongoimport file tool is, we can use it to import the three json files.

- mongoimport C:\Users\soule\BD2_Project\government_table_data.json -d French_admin -c governments --drop
- > mongoimport C:\Users\soule\BD2_Project\state_table_data.json -d French_admin -c state --drop
- > mongoimport C:\Users\soule\BD2_Project\cities_table_data.json -d French_admin -c cities --drop

### 4.1.2 Pymongo

We then switched to Python to use PyMongo for queries. We provided 5 queries on the table "Government". Here is the code to connect to the french admin and thus the government collection.

```
# Install pymongo
pip install pymongo
# importing the required libraries
import pymongo
import pprint
import json
import warnings
warnings.filterwarnings('ignore')
```

```
# connect to the mongoclient
client = pymongo.MongoClient('mongodb://localhost:27017')

# get the database
french_admin_db = client['French_admin']

# we get then to the collection
gov_collection = french_admin_db.governments
```

## 4.2 MongoDB queries

### 4.2.1 Filter with a condition query

The first query provides the element that respects the condition. Here, we put a condition on the name of the prime Minister, it has to be Jules Ferry. Only one element matches the query.

```
# First query
gov_collection.find_one({'Name of the Prime Minister': 'Jules Ferry'})
```

## 4.3 MongoDB solutions

### 4.3.1 Filter with a condition query

The first query, filter with a condition, provides the following element:

{"Name of the Prime Minister": "Jules Ferry", "Name of the president of the lower house": "Léon Gambetta", "Name of the president of the upper house": "Léon Say", "Date": 1880}

### 4.3.2 Equal to query

The second query, an equal to one, provides the following element:

{"Name of the Prime Minister": "François Fillon", "Name of the president of the lower house": "Bernard Accoyer", "Name of the president of the upper house": "Gérard Larcher", "Date": 2010}

### 4.3.3 Less than query

The third query, a less than one, provides the following element:

{"Name of the Prime Minister": "Pierre Laval", "Name of the president of the lower house": "Lucien Romier", "Name of the president of the upper house": "Lucien Romier", "Date": 1942},
{"Name of the Prime Minister": "Léon Blum", "Name of the president of the lower house": "Vincent Auriol", "Name of the president of the upper house": "Pierre de Felice", "Date": 1946},
{"Name of the Prime Minister": "Pierre Messmer", "Name of the president of the lower house": "Achille Perreti", "Name of the president of the upper house": "Alain Poher", "Date": 1972},
{"Name of the Prime Minister": "Raymond Barre", "Name of the president of the lower house": "Jacques Chaban-Delmas", "Name of the president of the upper house": "Alain Poher", "Date":

1978},
{"Name of the Prime Minister": "Edith Cresson", "Name of the president of the lower house": "Henri Emmanuelli", "Name of the president of the upper house": "René Monory", "Date": 1992},
{"Name of the Prime Minister": "Pierre Bérégovoy", "Name of the president of the lower house": "Philippe Séguin", "Name of the president of the upper house": "René Monory", "Date": 1993}

### 4.3.4 Sort query

The fourth query, a sort one, provides the following element:

{"Name of the Prime Minister": "Edouard Philippe", "Name of the president of the lower house": "Richard Ferrand", "Name of the president of the upper house": "Gérard Larcher", "Date": 2020},
{"Name of the Prime Minister": "Manuel Valls", "Name of the president of the lower house": "Claude Bartolone", "Name of the president of the upper house": "Gérard Larcher", "Date": 2015},
{"Name of the Prime Minister": "François Fillon", "Name of the president of the lower house": "Bernard Accoyer", "Name of the president of the upper house": "Gérard Larcher", "Date": 2010},
{"Name of the Prime Minister": "Pierre Bérégovoy", "Name of the president of the lower house": "Philippe Séguin", "Name of the president of the upper house": "René Monory", "Date": 1993},
{"Name of the Prime Minister": "Edith Cresson", "Name of the president of the lower house": "Henri Emmanuelli", "Name of the president of the upper house": "René Monory", "Date": 1992},
{"Name of the Prime Minister": "Raymond Barre", "Name of the president of the lower house": "Jacques Chaban-Delmas", "Name of the president of the upper house": "Alain Poher", "Date": 1978},
{"Name of the Prime Minister": "Pierre Messmer", "Name of the president of the lower house": "Achille Perreti", "Name of the president of the upper house": "Alain Poher", "Date": 1972},
{"Name of the Prime Minister": "Léon Blum", "Name of the president of the lower house": "Vincent Auriol", "Name of the president of the upper house": "Pierre de Felice", "Date": 1946},
{"Name of the Prime Minister": "Jules Ferry", "Name of the president of the lower house": "Léon Gambetta", "Name of the president of the upper house": "L\u00e9on Say", "Date": 1880},
{"Name of the Prime Minister": "Pierre Laval", "Name of the president of the lower house": "Lucien Romier", "Name of the president of the upper house": "Lucien Romier", "Date": 1942},

### 4.3.1 Sort and find at once query

The fifth query, a sort and find one, provides the following element:

{"Name of the Prime Minister": "Edouard Philippe", "Name of the president of the lower house": "Richard Ferrand", "Name of the president of the upper house": "Gérard Larcher", "Date": 2020},

# **5** Conclusion

The “French Administration” database was built with the aim of getting as close as possible to the life of the structure of the administration pattern in France by characterizing the different interactions between the actors of this system. At the end, a complex and complete datable was established. The potential of the database could be illustrated by the SQL and NoSQL queries that were presented in this report. This database could therefore be used to centralize all information on the French administration.

# Table of contents