SMBUD 2021 - Project work 3

Aman Gabba - 10793117

Andrea Cerasani - 10680486

Giovanni Demasi - 10656704

Pasquale Dazzeo - 10562130

Vlad Marian Cimpeanu - 10606922

POLITECNICO
MILANO 1863

Contents

1 Introduction

1.1 Problem Specification

The aim of this project was to design, store and query data on a NoSQL DB supporting a data analysis scenario over data about COVID-19 vaccination statistics. The purpose is that of building a comprehensive database of vaccinations.

A given vaccinations dataset has been assigned, with the purpose to pick a time interval of at least 3 months from it and, by using an ElasticSearch installation, import the data, apply the appropriate schema design choices, implement some queries aiming at exploring the data statistics and design a basic visualization dashboard of the results.

1.2 Hypothesis

The assumptions taken into account are the following:

- ...

2   ... diagram

...

3 Dataset description

3.1 Vaccine summary dataset

The Dataset of the project has been downloaded from The official Italian Government Github repository at the following link :
https://raw.githubusercontent.com/italia/covid19-opendata-vaccini/master/dati/somministrazioni-vaccini-latest.csv.

It contains information about administered vaccines in Italy and it is made by the following fields:

Field	Data type	Description
index	integer	Record identification code
area	string	Code of the delivery region
supplier	string	Complete name of the supplier of the vaccine
administration date	datetime	Administration date of the vaccines
age group	string	Age group to which the subjects to whom the vaccine were administered belong
male count	integer	Number of vaccinations administered to males per day, region and age group
female count	integer	Number of vaccinations administered to females per day, region and age group
first doses	integer	Number of people administered with the first dose
second doses	integer	Number of people administered with the second dose

post infection doses	integer	Number of administrations given to subjects with previous covid-19 infection in the 3-6 month period and who, therefore, conclude the vaccination cycle with a single dose
booster doses	integer	Number of people administered with an additional dose/recall
NUTS1 code	string	European classification of NUTS territorial units: NUTS level 1
NUTS2 code	string	European classification of NUTS territorial units: NUTS level 2
ISTAT region code	integer	ISTAT code of the Region
region name	string	Standard denomination of the area (where necessary bilingual denomination)

The data types written in the table are the 'original' ones, so the ones used by the dataset creator.

The same data types have been used to implement and use the dataset in ElasticSearch because they well represent the different parameters, so no changes were needed. (CHECK??)

The dataset period taken into consideration for statistical purposes is the one that goes from the first day of the year 2021 to the last day of September 2021. (CHECK??)

3.2   ... dataset

We have also, as optional point of this porject, integrated our analysis with another dataset ...

4 Queries and Commands

In the following chapter all the queries and commands parameters (part of the code to substitute with desired values) will be highlighted with **magenta** bold text.

Some parameters information can be useful for different queries or commands so they are written here to avoid writing them multiple times:

- ...

4.1 Queries

4.1.1 Query 1

The following query...

```
CODE
```

4.1.2 Query 2

The following query...

```
CODE
```

4.1.3 Query 3

The following query...

```
CODE
```

4.1.4 Query 4

The following query...

```
CODE
```

4.1.5 Query 5

The following query...

```
CODE
```

4.1.6 Query 6

The following query...

```
CODE
```

4.1.7 Query 7

The following query...

```
CODE
```

4.1.8 Query 8

The following query...

```
CODE
```

4.2   Commands

4.2.1   Command 1

The following query...

```
CODE
```

4.2.2   Command 2

The following query...

```
CODE
```

5   Dashboard description

The Kibana Dashboard has been made by...

6 User guide

...

7 Conclusion

Some interesting conclusions can be drawn from the development of this project:

...

8 References and Sources

- Elastic Guide: https://www.elastic.co/guide/index.html

- Italian Government repository: https://github.com/italia/covid19-opendata-vaccini