SMBUD 2021 - Project work – part 3

Prof. Marco Brambilla

Suppose you want to build an information system for managing pandemic information for a given country.

DELIVERY #3

You need 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.

We suggest you get and import the dataset available as open data at:

https://raw.githubusercontent.com/italia/covid19-opendata-vaccini/master/dati/somministrazioni-vaccini-latest.csv

(Pick a time interval of at least 3 months. The main website with the data description is available at https://github.com/italia/covid19-opendata-vaccini/. In the following page we will report the translated version)

By using an ElasticSearch installation: import the data of the dataset, apply the appropriate schema design choices, and implement some queries aiming at exploring the data statistics (evolution in time, multiple perspectives on the data for understanding the vaccine uptake by genre, region, etc.), and design a basic visualization dashboard of the results.

Tasks to perform:

- Write the specification and hypotheses of the problem and solution
- Report the schema of the data, including the types of the different fields. Make sure that the format/schema is correct and motivate it (even if you use the automatic mapping)
- Store the dataset in ElasticSearch
- Write basic Queries (minimum 8) and data update commands (minimum 2) useful for typical usage scenarios
- Implement a simple visualization dashboard using Kibana. Exploration, navigation and dynamicity of the dashboard will be considered a valuable contribution too
- Prepare a short report describing the above aspects

- Optional 1: if you want you can integrate other datasets

- Optional 2: implement some features in another nosql platform (different from Neo4J and MongoDB)

Deliverables to submit:

- Short report as above
- ZIP with the implementation

Deadline:

- January 10th 2022

Field Name	Data Type*	Description
Index	Integer	The index of the record
Area	String	Acronyms of the region of delivery
Supplier	String	Complete name of the supplier of the vaccine
Administration Date	Datetime	Administration date of the vaccines
Age Group	String	Age group of the people administered with the vaccines
Male Count	Integer	Number of vaccinations administered to males
Female Count	Integer	Number of vaccinations administered to females
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 people administered with a dose after they have been infected
Booster Doses	Integer	Number of people administered with an additional dose/recall
NUTS1 Code	String	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
NUTS2 Code	String	https://en.wikipedia.org/wiki/NUTS_statistical_regions_of_Italy
Region ISTAT Code	Integer	ISTAT code of a region
Region Name	String	Name of the region (bilingual, when necessary)

* It is not necessary to follow the data type when creating your Elasticsearch implementation. If you believe that some fields would be better suited for another type of data, create your own mapping