

SMBUD

Vaccinations Analysis

*Specification, Entity-Relationship model and
Elasticsearch analysis*



POLITECNICO
MILANO 1863

Filippo Lazzati (10629918) - Martina Magliani
(10682333) - Christian Grasso (10652464) - Sofia
Martellozzo (10623060) - Giacomo Lombardo
(10674987)
Year: 2021/2022

Contents

1	Problem specification	2
1.1	Dataset	2
2	Conceptual model	3
3	Sample dataset	4
3.1	Import the dataset	4
4	Queries and Commands	5
4.1	Queries	5
4.2	Commands	10
5	Kibana Dashboard	11
6	References and sources	12
7	Image gallery	12

1 Problem specification

The purpose of this project is to design, store and query data on a NoSQL database in order to analyze data about COVID-19 vaccination statistics.

Starting from an open dataset containing raw data about the distribution and administering of COVID-19 vaccines in Italian regions, the main goal of the project is to obtain relevant information through queries and graphs.

To perform this task, the dataset has been loaded into **ElasticSearch**, a search engine with an internal database focused on high performances that interacts with external components through RESTful APIs. Once loaded into the search engine, information can be retrieved from the dataset not only through queries but also through Kibana, a powerful user interface that allows to visualize ElasticSearch data. More specifically, through **Kibana** it is possible to create Dashboards that generate and display graphs and other significant information in an intuitive way.

1.1 Dataset

The used dataset is provided by the Italian Government and contains the official data about vaccines distribution and administration. More specifically it provides regional overview of the amount of doses administered daily in the time period between 27/12/2020 and today. The daily administration are divided by:

- type of dose (first, second, booster and post-infection)
- supplier (Pfizer, Moderna,...)
- age group
- gender
- area

A secondary dataset, containing additional information about the amount of daily delivered doses in the same time period, has also been implemented in order to better contextualise the primary dataset.

2 Conceptual model

The following is the Entity-Relationship (E-R) model of our database:

-

3 Sample dataset

3.1 Import the dataset

4 Queries and Commands

4.1 Queries

We have identified the following queries:

1. list of brands with number of vaccinations in last 30 days and brand with the highest number.

```
1 GET /somministrazioni-vaccini-latest/_search
2 {
3   "size": 0,
4   "query": {
5     "range": {
6       "@timestamp": {
7         "gte": "now-30d/d",
8         "lt": "now/d"
9       }
10    }
11  },
12  "runtime_mappings": {
13    "vaccini": {
14      "type": "long",
15      "script": ""
16      long total = doc['sesso_maschile'].value + doc['sesso_femminile'].value;
17      emit(total);
18    }
19  },
20  },
21  "aggs": {
22    "vaccinations_per_brand": {
23      "terms": {
24        "field": "fornitore"
25      },
26      "aggs": {
27        "total_vaccinations": {
28          "sum": {
29            "field": "vaccini"
30          }
31        }
32      }
33    }
34  },
35  "most-used_brand": {
36    "max_bucket": {
37      "buckets_path": "vaccinations_per_brand > total_vaccinations"
38    }
39  }
40 }
41 }
```

2. region with highest number of 1st vaccinations in last 3 months.

```
1 GET /somministrazioni-vaccini-latest/_search
2 {
3   "size": 0,
4   "query": {
5     "range": {
6       "@timestamp": {
7         "gte": "now-90d/d",
8         "lt": "now/d"
9       }
10    }
11  },
12  "aggs": {
13    "first_vaccinations_per_region": {
14      "terms": {
15        "field": "area"
16      },
17      "aggs": {
18        "first_vaccinations": {
19          "sum": {
20            "field": "prima_dose"
21          }
22        }
23      }
24    },
25    "most-vaccinated_region": {
26      "max_bucket": {
27        "buckets_path": "first_vaccinations_per_region >
28        first_vaccinations"
29      }
30    }
31  }
```

3. total number of boosters in italy in last 30 days.

```
1 GET /somministrazioni-vaccini-latest/_search
2 {
3   "size": 0,
4   "query": {
5     "range": {
6       "@timestamp": {
7         "gte": "now-30d/d",
8         "lt": "now/d"
9       }
10    }
11  },
12  "aggs": {
13    "total_number_of_boosters": {
14      "sum": {
15        "field": "dose_addizionale_booster"
16      }
17    }
18  }
19 }
```

4. age range with highest number of vaccinations in last 30 days

```
1 GET /somministrazioni-vaccini-latest/_search
2 {
3   "size": 0,
4   "query": {
5     "range": {
6       "@timestamp": {
7         "gte": "now-30d/d",
8         "lt": "now/d"
9       }
10    }
11  },
12  "runtime_mappings": {
13    "vaccini": {
14      "type": "long",
15      "script": ""
16        long total = doc['sesso_maschile'].value + doc['
17        sesso_femminile'].value;
18        emit(total);
19      ""
20    }
21  },
22  "aggs": {
23    "vaccinations_per_age_range": {
24      "terms": {
25        "field": "fascia_anagrafica"
26      },
27      "aggs": {
28        "total_vaccinations": {
29          "sum": {
30            "field": "vaccini"
31          }
32        }
33      }
34    },
35    "most-used_brand": {
36      "max_bucket": {
37        "buckets_path": "vaccinations_per_age_range >
38        total_vaccinations"
39      }
40    }
41  }
```

5. total number of Moderna doses received in the last 3 days in Italy.

```
1 GET /consegne-vaccini-latest/_search
2 {
3   "size": 0,
4   "query": {
5     "bool": {
6       "filter": [
7         {"term": {
```



```

8         "fornitore": "Moderna"
9     },
10 },
11     {"range": {
12         "@timestamp": {
13             "gte": "now-3d/d",
14             "lt": "now-1d/d"
15         }
16     }
17 }
18 ]
19 }},
20 "aggs": {
21     "vaccine_doses_received": {
22         "sum": {
23             "field": "numero_dosi"
24         }
25     }
26 }
27 }

```

6. region with highest disparity of vaccinations between males and females.

```

1 GET /somministrazioni-vaccini-latest/_search
2 {
3     "size": 0,
4     "runtime_mappings": {
5         "vaccini": {
6             "type": "long",
7             "script": ""
8             long total = doc['sesso_maschile'].value - doc['
9 sesso_femminile'].value;
10             if(total < 0){
11                 total = total * - 1;
12             }
13             emit(total);
14         }
15     },
16     "aggs": {
17         "vaccinations_per_region":{
18             "terms": {
19                 "field": "nome_area"
20             },
21             "aggs": {
22                 "max_dif_of_vaccinations_between_male_female": {
23                     "max": {
24                         "field": "vaccini"
25                     }
26                 }
27             }
28         },
29         "highest_disparity_of_vaccinations": {
30             "max_bucket": {
31                 "buckets_path": "vaccinations_per_region >
max_dif_of_vaccinations_between_male_female"

```

```

31     }
32   }
33 }
34 }

```

7. number of children between 5-11 vaccinated this week in Basilicata.

```

1  GET /somministrazioni-vaccini-latest/_count
2  {
3    "query": {
4      "bool": {
5        "must": [
6          { "match" :
7            "fascia_anagrafica": "5-11"
8          },
9          { "range" : {
10             "@timestamp": {
11               "gte": "now-7d/d",
12               "lte": "now/d"
13             }
14           },
15           { "match" : {
16             "nome_area": "Basilicata"
17           }
18         }
19       }
20     }
21 }

```

8. to do.

9. brand of the less-used vaccine with young people (<30 years old).

```

1  GET /somministrazioni-vaccini-latest/_search
2  {
3    "query": {
4      "bool": {
5        "should": [
6          { "match" :
7            "fascia_anagrafica": "5-11"
8          },
9          { "match" :
10             "fascia_anagrafica": "12-19"
11           },
12           { "match" :
13             "fascia_anagrafica": "20-29"
14           }
15         ]
16       },
17       "runtime_mappings": {
18         "vaccini": {

```

```

19     "type": "long",
20     "script": ""
21     long total = doc['sesso_maschile'].value + doc['
sesso_femminile'].value;
22     emit(total);
23     ""
24 }
25 },
26 "aggs": {
27   "vaccinations_per_region": {
28     "terms": {
29       "field": "nome_area"
30     },
31     "aggs": {
32       "total_vaccinations": {
33         "sum": {
34           "field": "vaccini"
35         }
36       },
37       "less-used_brand": {
38         "min_bucket": {
39           "buckets_path": "vaccinations_per_region >
total_vaccinations"
40         }
41       }
42 }

```

4.2 Commands

We have identified the following commands to show how the system works:

1. Update one row.

```

1      POST /somministrazioni/_update_by_query
2  {
3    "query": {
4      "bool": {
5        "filter": [
6          {"term": {"area": "ABR"}},
7          {"term": {"fascia_anagrafica": "40-49"}},
8          {"term": {"fornitore": "Moderna"}},
9          {
10           "range": {
11             "@timestamp": {
12               "time_zone": "+02:00",
13               "gte": "2020-03-30T00:00:00",
14               "lt": "2020-03-31T00:00:00"
15             }
16           }
17         }
18       ]
19     }
20   },
21   "script": {

```

```

22     "source": "ctx.sesso_maschile = params.sesso_maschile; ctx
    .seconda_dose = params.seconda_dose",
23     "lang": "painless",
24     "params": {
25         "sesso_maschile": 100,
26         "seconda_dose": 2012
27     }
28 }
29 }

```

2. Create one row.

```

1      POST /somministrazioni/_doc
2  {
3      "data_somministrazione": "2021-12-28",
4      "fornitore": "Moderna",
5      "area": "BAS",
6      "fascia_anagrafica": "30-39",
7      "sesso_maschile": 910,
8      "sesso_femminile": 1211,
9      "prima_dose": 93,
10     "seconda_dose": 563,
11     "pregressa_infezione": 33,
12     "dose_addizionale_booster": 1465,
13     "codice_NUTS1": "ITF",
14     "codice_NUTS2": "ITF5",
15     "codice_regione_ISTAT": 17
16 }

```

5 Kibana Dashboard

By using the Kibana tool we realised a simple dashboard containing some relevant graphs to visualised the data provided by the dataset. The dashboard displays the following information and graphs:

- total number of administered vaccines
- administered doses of vaccines per region
- administered doses per supplier
- administered doses per week, grouped by:
 - supplier
 - age range
 - gender
- total number of delivered doses
- delivered doses per region
- delivered doses per week

6 References and sources

- [Elasticsearch](#)
- [Kibana](#)
- [Dataset](#)
- [draw.io](#)

7 Image gallery

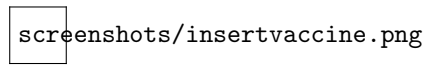


Figure 1: Page for the insertion of a new vaccine dose to a person.