NoSQL - DIA2

MANNAI Hasna (≈ 33%)
CHENIK Yassine (≈ 33%)
BOUCHIBA Emine (≈ 33%)

RAPPORT
TP ElasticSearch

TABLE DES MATIERES

ElasticSearch :

In ElasticSearch, we need to ad additionnal rows :

{"index":{"_index":"companies2","_id":k}}
With $k \in [\![1, N]\!]$ having N as the original number of rows of the JSON File.

To do that, we decided to create a python code to add these rows into our original JSON File.

Here is the code we compiled :

```
with
open("C:\\Users\\emine\\Documents\\Ecole\\A4\\S8\\Advanced_topics_in_NoSql_database
s\\companies2.txt", "r", encoding="utf-8") as f:
    lines = f.readlines()

with
open("C:\\Users\\emine\\Documents\\Ecole\\A4\\S8\\Advanced_topics_in_NoSql_database
s\\companies2.txt", "w", encoding="utf-8") as f:
    for i, line in enumerate(lines):
        f.write('{"index":{"_index":"companies2","_id":%d}}\n' % (i+1))
        f.write(line)
```

Importation du json dans MongoDB

Lancement de Docker :

Lancer Docker et le container ElasticSearch.

Vérification du statut de Kibana et ElasticSearch :

- ElasticSearch :
  - ✓ http://localhost:9200
- Kibana :
  - ✓ http://localhost:5601

Importation du fichier JSON :

- Menu/Kibana/Index Patterns
- « Upload File »
- And we drag and drop our file « companies.json ».

Simple Queries

1. Get all the companies that have a category_code=”nanotech” :

✓ Code :

```
GET /companies/_search
{
 "query": {
  "match": {
   "category_code": "nanotech"
  }
 }
}
```

✓ *Ici, on veut afficher les entreprises qui sont dans le domaine de la nanotechnologie en affichant le* ***"name"****, le* ***"category_code"*** *ainsi que l’****id****.*

✓ *Output :*


```
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "k9cevIYBlYorPAo2BDpe",
  "_score" : 7.970709,
  "_source" : {
    "category_code" : "nanotech",
    "name" : "Nanosolar"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "s9cevIYBlYorPAo2J4VW",
  "_score" : 7.970709,
  "_source" : {
    "category_code" : "nanotech",
    "name" : "Nantero"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "59cevIYBlYorPAo2MrTC",
  "_score" : 7.970709,
  "_source" : {
    "category_code" : "nanotech",
    "name" : "QD Vision"
  }
},
```

2. Get all the Companies that founded in 2008 :

✓ Code :

```
GET /companies/_search
{
  "_source": ["name", "founded_year", "founded_month", "founded_day"],
  "query": {
    "bool": {
      "must": [
        {"range": {
          "founded_year": {"gte": 2008, "lte": 2008}
        }}
      ]
    }
  }
}
```

✓ *Ici, on veut afficher les entreprises qui ont été fondés en 2008. en affichant le* ***"name"****, le* ***"founded_year"****, le* ***"founded_month"****, le* ***"founded_day"*** *ainsi que l'****id****.*

✓ *Output :*


```
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "MdcevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "founded_month" : null,
    "founded_year" : 2008,
    "name" : "BeliefNet",
    "founded_day" : null
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "8dcevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "founded_month" : 1,
    "founded_year" : 2008,
    "name" : "Fancast",
    "founded_day" : 1
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "O9cevIYBlYorPAo2CTtR",
  "_score" : 1.0,
  "_source" : {
    "founded_month" : null,
    "founded_year" : 2008,
    "name" : "Newspepper",
    "founded_day" : null
  }
},
```

3. Get the number of Companies whith more than 100 employees :

✓ Code :

```
GET /companies/_search
{
  "_source": ["name", "number_of_employees"],
  "query": {
    "bool": {
      "must": [
        {"range": {
          "number_of_employees": {"gte": 100}
        }}
      ]
    }
  }
}
```

✓ *Ici, on veut afficher les entreprises qui ont plus de 100 employés en affichant le **"name"**, le **"number_of_employees"** ainsi que l'**id**.*

✓ *Output :*


```
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "b9cevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "number_of_employees" : 400,
    "name" : "Deem"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "cdcevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "number_of_employees" : 350,
    "name" : "CDNetworks"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "ddcevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "number_of_employees" : 101,
    "name" : "KAYAK"
  }
},
```

4. Get the total amount raised per companies where category_code= "security" :

✓ Code :

```
GET /companies/_search
{
  "query": {
    "match": {
      "category_code": "security"
    }
  },
  "aggs": {
    "companies": {
      "terms": {
        "field": "name"
      },
      "aggs": {
        "total_raised": {
          "sum": {
            "field": "funding_rounds.raised_amount"
          }
        }
      }
    }
  }
  , "size": 1
  , "_source": ["name"]
}
```

Simple Queries

- ✓ *Ici nous cherchons à afficher le **montant total amassé** par les entreprises qui travaillent dans le **domaine de la sécurité**.*
- ✓ *Output :*


```
        "key" : "Lockheed Martin",
        "doc_count" : 2,
        "total_raised" : {
          "value" : 5.48E7
        }
      },
      {
        "key" : "Nevis Networks",
        "doc_count" : 2,
        "total_raised" : {
          "value" : 3249600.0
        }
      },
      {
        "key" : "NitroSecurity",
        "doc_count" : 2,
        "total_raised" : {
          "value" : 5.0E7
        }
      },
      {
        "key" : "PGP TrustCenter",
        "doc_count" : 2,
        "total_raised" : {
          "value" : 1.7E7
        }
      },
```

5. Get total money raised for each domain :

✓ **Code :**

```
GET /companies/_search
{
  "aggs": {
    "category_code": {
      "terms": {
        "field": "category_code"
      },
      "aggs": {
        "total_raised": {
          "sum": {
            "field": "funding_rounds.raised_amount"
          }
        }
      }
    }
  }
  , "size": 1
  , "_source": ["category_code"]
}
```

Simple Queries

- ✓ *Ici, nous cherchons à afficher le **montant total amassé** par les entreprises du **même domaine**.*
- ✓ *Output :*


```
        "key" : "web",
        "doc_count" : 3769,
        "total_raised" : {
          "value" : 9.679404941E9
        }
      },
      {
        "key" : "software",
        "doc_count" : 2708,
        "total_raised" : {
          "value" : 1.6118613263E10
        }
      },
      {
        "key" : "games_video",
        "doc_count" : 1074,
        "total_raised" : {
          "value" : 8.661739405E9
        }
      },
      {
        "key" : "mobile",
        "doc_count" : 1007,
        "total_raised" : {
          "value" : 9.751133206E9
        }
      },
```

6. Get all companies that starts with an “F” :

➢ Code :

```
GET /companies/_search
{
  "query": {
    "prefix": {
      "name": "F"
    }
  },
  "_source": ["name"]
}
```

✓ *Ici, on cherche à afficher toutes les entreprises qui commencent par la lettre F. Pour ce faire, on utilise l’option* ***$regex*** *qui nous permet de reconnaître un motif dans une chaîne de caractères et on lui donne comme argument* ***‘^F’****. De et on affiche les noms des entreprises correspondantes.*

✓ *Output :*


```
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "8dcevIYBlYorPAo2BDle",
  "_score" : 1.0,
  "_source" : {
    "name" : "Fraud Sciences"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "FdcevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "name" : "FFWD Wheels"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "HdcevIYBlYorPAo2BDpe",
  "_score" : 1.0,
  "_source" : {
    "name" : "FastBooking"
  }
},
```

Complex Queries

1. Get the number of employees of each Company that have as category_code =”software” :

✓ Code :

```
GET /companies/_search
{
  "query": {
    "match": {
      "category_code": "software"
    }
  },
  "aggs": {
    "group_by_name": {
      "terms": {
        "field": "name.keyword",
        "size": 10
      },
      "aggs": {
        "employee_count": {
          "max": {
            "field": "number_of_employees"
          }
        }
      }
    }
  },
  "_source": ["name", "number_of_employees","category_code"],
  "size": 10
}
```

Complex Queries

- ✓ *Ici, on cherche à avoir le **nombre d'employés** travaillant au sein d'entreprises dans le domaine du **Software** et on affiche **le nom de l'entreprise**, sa **category_code** et le **nombre d'employés**.*
- ✓ *Output :*


```
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "L9cevIYBlYorPAo2BDpe",
  "_score" : 1.7713068,
  "_source" : {
    "number_of_employees" : 25,
    "category_code" : "software",
    "name" : "LiveJournal"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "bdcevIYBlYorPAo2BDpe",
  "_score" : 1.7713068,
  "_source" : {
    "number_of_employees" : 39,
    "category_code" : "software",
    "name" : "RipCode"
  }
},
{
  "_index" : "companies",
  "_type" : "_doc",
  "_id" : "qdcevIYBlYorPAo2BDpe",
  "_score" : 1.7713068,
  "_source" : {
    "number_of_employees" : 60,
    "category_code" : "software",
    "name" : "Bunchball"
  }
},
```

2. Get the number of products grouped by companies name :

- Code :

```
GET /companies/_search
{
  "aggs": {
    "companies": {
      "terms": {
        "field": "name"
      },
      "aggs": {
        "total_products": {
          "value_count": {
            "field": "products.name.keyword"
          }
        }
      }
    }
  }
  ,"size": 1
}
```

Complex Queries

- ✓ *Ici, on veut afficher le nombre de produits provenant de chacune des entreprises.*
- ✓ *Output :*


```
"key" : "2Big2Send",
"doc_count" : 2,
"total_products" : {
  "value" : 4
}
},
{
"key" : "2GeeksinaLab",
"doc_count" : 2,
"total_products" : {
  "value" : 0
}
},
{
"key" : "2pad",
"doc_count" : 2,
"total_products" : {
  "value" : 0
}
},
{
"key" : "360 Sports Software",
"doc_count" : 2,
"total_products" : {
  "value" : 2
}
},
```

Hard Queries

1. Get the total amount of money raised by companies founded in each year :

- Code :

```
GET /companies/_search
{
  "aggs": {
    "companies": {
      "terms": {
        "field": "founded_year"
      },
      "aggs": {
        "total_raised": {
          "sum": {
            "field": "funding_rounds.raised_amount"
          }
        }
      }
    }
  }
  , "size": 1
  , "_source": ["name"]
}
```

Hard Queries

- ✓ *Ici, on cherche à obtenir le montant total d'argent collecté par les entreprises fondées chaque année, trié par ordre décroissant du montant total collecté.*
- ✓ *Output :*


```
{
  "key" : 2007,
  "doc_count" : 2277,
  "total_raised" : {
    "value" : 1.4650692921E10
  }
},
{
  "key" : 2006,
  "doc_count" : 1418,
  "total_raised" : {
    "value" : 1.3606757327E10
  }
},
{
  "key" : 2005,
  "doc_count" : 957,
  "total_raised" : {
    "value" : 1.5032225448E10
  }
},
{
  "key" : 2004,
  "doc_count" : 749,
  "total_raised" : {
    "value" : 1.3269671942E10
  }
},
```