

# *Captura y almacenamiento de información*

Especialización en Inteligencia de Datos orientada a Big Data

2021

# *Captura y almacenamiento de información*

En esta sección se presentan las principales características del motor Elasticsearch

## *ElasticSearch*



Facultad de  
**INFORMÁTICA**



UNIVERSIDAD  
NACIONAL  
DE LA PLATA

# ElasticSearch

---

## *Características*

- Open-source, basado en Apache Lucene. Provee interface RESTful .
- Utiliza JSON.
- Cada índice puede tener sus propias configuraciones.
- Puede utilizarse en conjunto con soluciones más tradicionales como base de datos relacionales.
- Soporta búsquedas con querystring al estilo de Lucene.

# ElasticSearch

## Índices

- Cada vez que se guarda información en ElasticSearch se lo almacena en un índice que posee un tipo.
- Convención:

<server>:<port>/{índice}/{tipo}

ElasticSearch	MongoDB	RDBMS
índice	Base de datos	Base de datos
tipo	colección	tabla

# ElasticSearch

## Mapeo

- El mapeo define como se deberían almacenar los documentos y como indexarlos para la búsqueda.
- Si no se define un mapeo ElasticSearch asignará automáticamente uno, pero no es recomendable ya que es muy genérico.

```
1 {  
2   "first_name": "bam",  
3   "last_name": "margera",  
4   "gender": "male",  
5   "age": 36  
6 }
```

# ElasticSearch

## Mapeo

```
1 PUT localhost:9200/test/
2
3 {
4   "mappings": {
5     "users": {
6       "properties": {
7         "age": {
8           "type": "long"
9         },
10        "first_name": {
11          "type": "string"
12        },
13        "gender": {
14          "type": "string"
15        },
16        "level": {
17          "type": "string"
18        },
19        "last_name": {
20          "type": "string"
21        }
22      }
23    }
24  }
25 }
```

# ElasticSearch

## Indexación

- Es el proceso de insertar datos para que sea posible realizar búsquedas sobre éstos.

```
1 POST localhost:9200/test/users/
2 {
3   "first_name": "Bam",
4   "last_name": "Margerá",
5   "gender": "male",
6   "level": "super awesome",
7   "age": 36
8 }
9
10 POST localhost:9200/test/users/
11 {
12   "first_name": "Stephanie",
13   "last_name": "Hodge",
14   "gender": "female",
15   "level": "awesome",
16   "age": 34
17 }
18
19
20 POST localhost:9200/test/users/
21 {
22   "first_name": "Johnny",
23   "last_name": "Knoxville",
24   "gender": "male",
25   "level": "awesome",
26   "age": 45
27 }
28 }
```

# ElasticSearch

## Búsquedas

- Para realizar una búsqueda en un índice por un tipo en particular se utiliza la convención:
- POST localhost:9200/test/users/\_search

```
1 {
2   "took": 4,
3   "timed_out": false,
4   "_shards": {
5     "total": 5,
6     "successful": 5,
7     "failed": 0
8   },
9   "hits": {
10    "total": 3,
11    "max_score": 1,
12    "hits": [
13      {
14        "_index": "test",
15        "_type": "users",
16        "_id": "AVRQQ1CE0YBBUjDwpzQZ",
17        "_score": 1,
18        "_source": {
19          "first_name": "Bam",
20          "last_name": "Marger",
21          "gender": "male",
22          "level": "super awesome",
23          "age": 36
24        }
25      },
26      **... the other 2 docs go here**
27    ]
28  }
29 }
```



# ElasticSearch

## Consultas

- Se pueden utilizar criterios de relevancia
- Por ejemplo, utilizando “match” para obtener aquellos usuarios con cierto nivel...

```
1 POST localhost:9200/test/users/_search
2
3 {
4   "query": {
5     "match": {
6       "level": "super awesome"
7     }
8   }
9 }
```

# ElasticSearch

## Filtros

- Los filtros no se basan en la relevancia de la información

```
1 POST localhost:9200/test/users/_search
2
3 {
4   "filter": {
5     "match": {
6       "gender": "male"
7     }
8   }
9 }
```

# ElasticSearch

## *Consultas sobre varios campos*

- Se puede consultar por varios campos buscando un determinado valor

```
POST /{index}/{type}/_search
```

```
{
  "query": {
    "multi_match" : {
      "query" : "{search term}",
      "fields" : ["{field 1}", "{field 2}", "{field 2}", "{field 3}"]
    }
  }
}
```

# ElasticSearch

---

## ***Herramientas***

- Cerebro: consola de administración y monitoreo para ElasticSearch.
- Logstash: herramienta para recuperación, transformación e inserción masiva de datos.
- Kibana: visualización y análisis de la información.