

Introducción al Stack Elastic

Introducción al Stack Elastic	1
Demo 1	2
Setup del entorno	2
CRUD	5
Queries	6
Agregaciones	8
Gist	8
Demo 2 - Ingesta de documentos	9

Para ejecutar las demos se debe clonar el proyecto <https://github.com/immavalls/viu-elk-ml-talk>

```
git clone git@github.com:immavalls/viu-elk-ml-talk.git
```

E instalar los prerequisites documentados. Estas demos requieren **docker** y **docker-compose**.

Demo 1

Setup del entorno

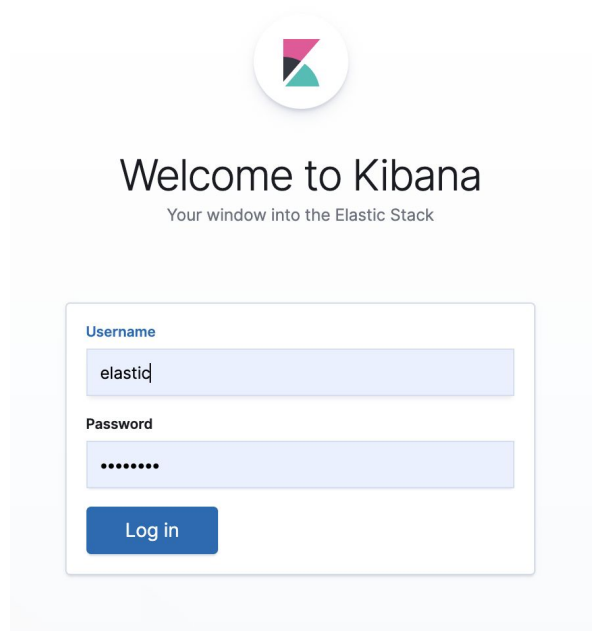
Para ello, situados en el raíz del proyecto **viu-elk-mi-talk**, ejecutar:

```
docker-compose up -d
```

Comprobar que ha arrancado correctamente:

<http://localhost:5601/>

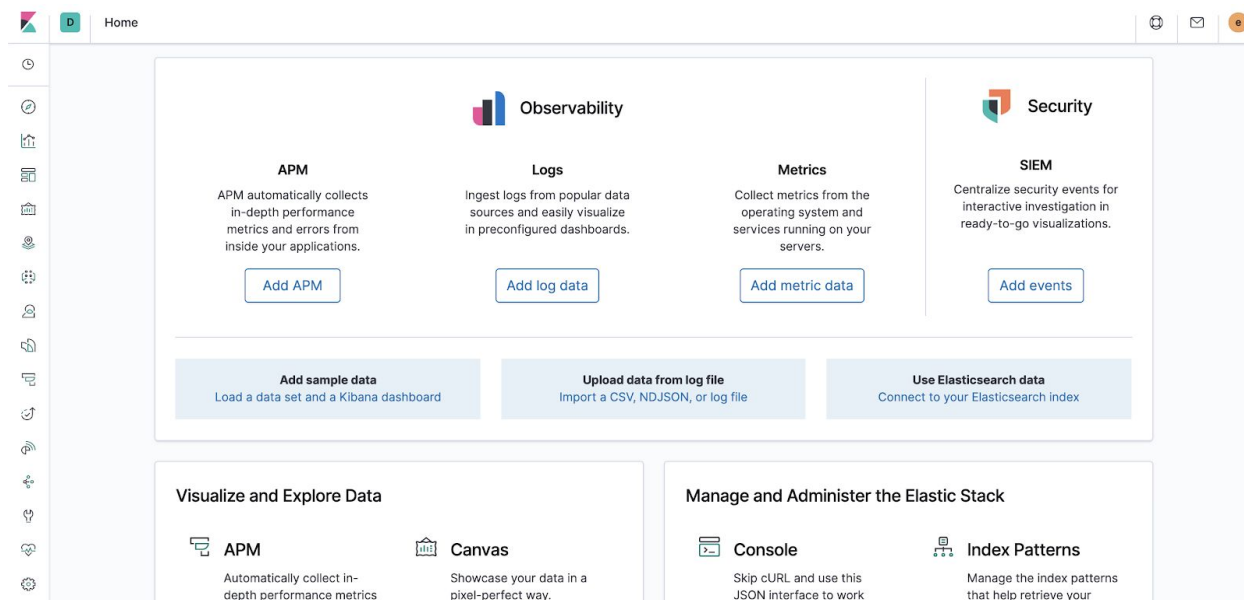
Si Elasticsearch y Kibana han arrancado correctamente, veremos el pantalla de Login de Kibana.



Entrar con:

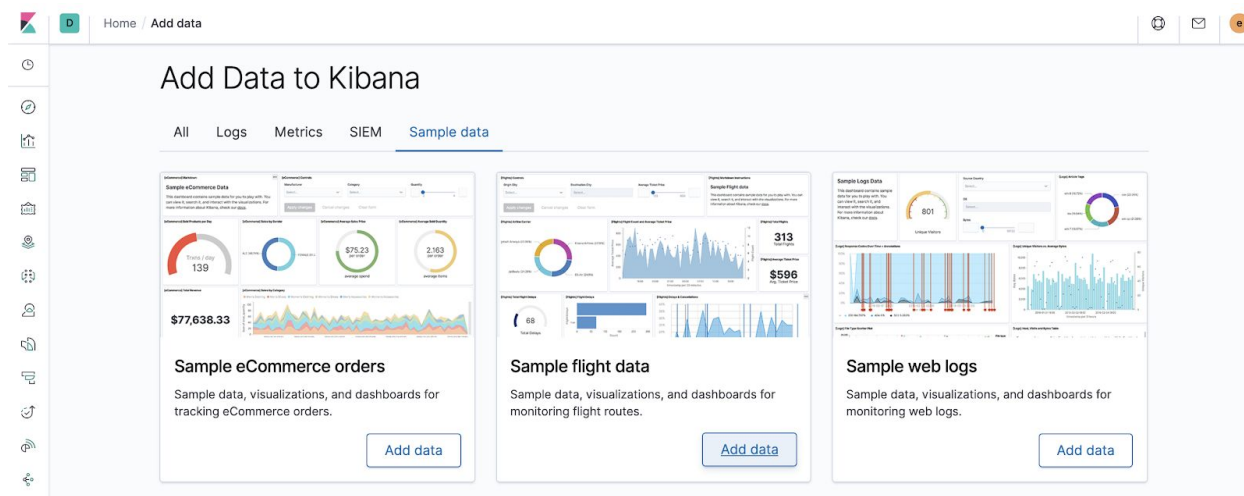
Usuario: elastic

Contraseña: changeme



Vamos a cargar datos de prueba que proporciona Kibana, para facilitar la exploración del Stack.

En la parte central de la pantalla Home, Pulsar en “Add sample data!”, “Load a data set and a Kibana dashboard”.



Pulsar “Add data” en **Sample flight data** y en **Sample web logs**.

En la parte inferior izquierda, en el menú, pulsar en la flecha  “Expand”. Aquí podemos

ver todas las opciones de menú de Kibana. Escogemos , “Dev Tools”.

Vamos a preparar los datos de vuelos para poder hacer búsquedas de texto libre. Para ello, ejecutar lo siguiente.

En primer lugar, añadiremos un campo de tipo text, necesarios para realizar las búsquedas.¹

```
POST kibana_sample_data_flights/_mappings
{
  "properties": {
    "origin": {
      "type": "keyword",
      "fields": {
        "text": {
          "type": "text"
        }
      }
    }
  }
}
```

Forzaremos un update de todos los documentos del índice para que este campos se actualice²

```
POST kibana_sample_data_flights/_update_by_query
{
  "script": {
    "source": ""
    ctx._source.origin = ctx._source.Origin;
    "",
    "lang": "painless"
  }
}
```

¹ <https://www.elastic.co/blog/strings-are-dead-long-live-strings>

² <https://www.elastic.co/guide/en/elasticsearch/reference/current/docs-update-by-query.html>

CRUD

- Crear un documento con identificador 1.

```
POST kibana_sample_data_flights/_doc/1
{
  "FlightNum": "652J760",
  "Origin": "Amsterdam Airport Schiphol",
  "Dest": "Stockholm-Arlanda Airport",
  "FlightDelayMin": 0,
  "Cancelled" : false,
  "timestamp" : "2020-05-01T05:21:34"
}
```

- Leer un documento

```
GET kibana_sample_data_flights/_doc/1
```

- Actualizar un documento

```
POST kibana_sample_data_flights/_update/1
{
  "doc": {
    "FlightDelayMin": 260
  }
}
```

- Borrar un documento

```
DELETE kibana_sample_data_flights/_doc/1
```

Queries

Podemos realizar distintas queries, usando Query DSL³.

- ¿Vuelos con origen en Amsterdam?

```
GET kibana_sample_data_flights/_search
{
  "query": {
    "match": {
      "origin.text": "amsterdam"
    }
  }
}
```

- ¿Vuelos retrasados 60 minutos o más?

```
GET kibana_sample_data_flights/_search?filter_path=hits.hits._source
{
  "size": 10,
  "_source": ["FlightDelayMin", "FlightNum"],
  "query": {
    "range": {
      "FlightDelayMin": {
        "gte": 60
      }
    }
  }
}
```

³ <https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl.html>

- ¿Vuelos con destino Amsterdam retrasados 60 minutos o más?

```
GET kibana_sample_data_flights/_search
{
  "size": 5,
  "_source": [
    "FlightDelayMin",
    "FlightNum"
  ],
  "query": {
    "bool": {
      "must": [
        {
          "match": {
            "origin.text": "amsterdam"
          }
        }
      ],
      "filter": [
        {
          "range": {
            "FlightDelayMin": {
              "gte": 60
            }
          }
        }
      ]
    }
  }
}
```

Agregaciones

- ¿Cuáles son los top 3 aeropuertos origen?

```
GET kibana_sample_data_flights/_search
{
  "size": 0,
  "aggregations": {
    "top_aeropuertos_origen": {
      "terms": {
        "field": "origin",
        "size": 3
      }
    }
  }
}
```

- ¿Qué retraso tienen en media los vuelos en los top 2 aeropuertos origen?

```
GET kibana_sample_data_flights/_search
{
  "size": 0,
  "aggregations": {
    "top_aeropuertos_origen": {
      "terms": {
        "field": "origin",
        "size": 3
      },
      "aggregations": {
        "media_retraso": {
          "avg": {
            "field": "FlightDelayMin"
          }
        }
      }
    }
  }
}
```

Gist

<https://gist.github.com/immavalls/b3d9b1b3985e75f86f64836632250baf>

Demo 2 - Ingesta de documentos

En esta demo, vamos a ingestar en Elasticsearch un documento en formato JSON. Para ello, usaremos Filebeat.

El documento en JSON que usaremos como ejemplo es el listado de bibliotecas de la Comunitat Valenciana en 2020: <https://dadesobertes.gva.es/es/dataset/cul-dir-bibliotecas-2020>

Dentro del proyecto encontraréis el documento en formato CSV (original) y JSON. El fichero que vamos a usar es el ubicado en `doc/datasets/bibliotecas-comunidad-valenciana-2020.ndjson`. Hemos convertido el CSV a JSON, para facilitar la ingesta.

Para ingestar este fichero, usaremos Filebeat. La pipeline de ingestar sería la siguiente, en la que Filebeat va a leer de un fichero en formato JSON (un documento por línea), y lo enviará a Elasticsearch:



Para arrancar Filebeat, editaremos el fichero `docker-compose.yml` que está en el raíz del proyecto, y descomentamos la parte donde se define el contenedor de Filebeat.

```
docker-compose.yml ×
docker-compose.yml > {} services > {} filebeat > abc container_name

29     networks:
30     - elk
31
32     # Filebeat container
33     filebeat:
34     container_name: filebeat
35     hostname: filebeat
36     image: "docker.elastic.co/beats/filebeat:${ELK_VERSION}"
37     volumes:
38     - # Mount filebeat configuration
39     - ./filebeat/config/filebeat.yml:/usr/share/filebeat/filebeat.yml
40     - # Mount file directory into container /var/log
41     - ./doc/dataset/:/var/log/
42     - # Mount data registry in host
43     - filebeatdata:/usr/share/filebeat/data/
44     networks:
45     - elk
46     restart: on-failure
47     depends_on:
48     - elasticsearch: {condition: service_healthy}
49
50     # Kibana container
51     kibana:
52     container_name: kibana
```

Ese contenedor está configurado para leer el fichero

`doc/datasets/bibliotecas-comunidad-valenciana-2020.ndjson` y enviar los documentos JSON a Elasticsearch.

Arrancamos Filebeat con:

```
docker-compose up -d
```

Y podemos visualizar los logs para comprobar que está funcionando correctamente:

```
docker-compose logs -f filebeat
```

Para comprobar que la ingesta ha sido correcta ejecutando en *Kibana Dev Tools* la siguiente query:

```
GET bibliotecas-valencia/_search?filter_path=hits.total,hits.hits._source.NOMBRE
```

Que nos devolverá un total de 685 resultados, y al no definir size en la query, 10 de esos documentos:

```
{
  "hits" : {
    "total" : {
      "value" : 685,
      "relation" : "eq"
    },
    "hits" : [
      {
        "_source" : {
          "NOMBRE" : "AGENCIA DE LECTURA SANTA CECILIA MISLATA"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "AGENCIA DE LECTURA MOIXENT"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "BIBLIOTECA PUBLICA MUNICIPAL MONTSERRAT"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "OTRAS PÚBLICAS MONTÁN"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "AGENCIA DE LECTURA MONTAVERNER"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "AGENCIA DE LECTURA MONTROI"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "BIBLIOTECA PUBLICA MUNICIPAL JUAN BAUTISTA MUÑOZ FERRANDIS MUSEROS"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "OTRAS PÚBLICAS NOVETLÈ"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "AGENCIA DE LECTURA SANT RAFEL ONTINYENT"
        }
      },
      {
        "_source" : {
          "NOMBRE" : "AGENCIA DE LECTURA ORBA"
        }
      }
    ]
  }
}
```

Finalmente, si queremos visualizar estos datos en Kibana, crearíamos un **Index Pattern**⁴. Para ello, pulsamos en el menú de Kibana,  Management “Management”, y creamos un index pattern para “*bibliotecas-valencia*”, seleccionando en “*I don’t want to use the Time Filter*” (estos documentos no tienen ningún timestamp).



Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

☐ Include system indices

Step 1 of 2: Define index pattern

Index pattern

bibliotecas-valencia

You can use a * as a wildcard in your index pattern.
You can't use spaces or the characters \, /, ?, *, <, >, |.

✓ **Success!** Your index pattern matches **1 index**.

bibliotecas-valencia-2020.05.04-000001

Rows per page: 10

> Next step

⁴ <https://www.elastic.co/guide/en/kibana/current/index-patterns.html>

Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

☐ Include system indices

Step 2 of 2: Configure settings

You've defined **bibliotecas-valencia** as your index pattern. Now you can specify some settings before we create it.

Time Filter field name Refresh

I don't want to use the Time Filter

The Time Filter will use this field to filter your data by time.
You can choose not to have a time field, but you will not be able to narrow down your data by a time range.

Hide advanced options

Custom index pattern ID

custom-index-pattern-id

Kibana will provide a unique identifier for each index pattern. If you do not want to use this unique ID, enter a custom one.

< Back

Create index pattern

Management / Index patterns / bibliotecas-valencia

bibliotecas-valencia

This page lists every field in the **bibliotecas-valencia** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch [Mapping API](#).

Fields (2593) | Scripted fields (0) | Source filters (0)

Filter: All field types

Name	Type	Format	Searchable	Aggregatable	Excluded
@timestamp	date		•	•	
CATALOGO	string		•	•	
CENTRAL	string		•	•	
COD_CARACTER	string		•	•	
COD_MUNICIPIO	number		•	•	
COD_PROVINCIA	number		•	•	
CP	number		•	•	
DECRETO	string		•	•	
DESC_CARACTER	string		•	•	
DIRECCION	string		•	•	

Rows per page: 10

< 1 2 3 4 5 ... 260 >

Una vez creado el patrón para el índice, escogemos Kibana Discover en el menú de la izquierda. Seleccionaremos ese index pattern, y podremos ver los documentos creados.

Discover

New Save Open Share Inspect

Search

+ Add filter

kibana_sample_data_fli...

CHANGE INDEX PATTERN

Filter options

- bibliotecas-valencia
- ✓ kibana_sample_data_flights
- kibana_sample_data_logs

Cancelled

Discover

New Save Open Share Inspect

Search KQL Refresh

+ Add filter

bibliotecas-valencia 685 hits

Search field names

Filter by type 0

Selected fields

- _source

Available fields

- @timestamp
- CATALOGO
- CENTRAL
- COD_CHARACTER

```

@timestamp: May 4, 2020 @ 10:53:04.809 ecs.version: 1.4.0 DESC_CHARACTER: PUBLICA COD_MUNICIPIO: 169 TELEFONO: TELF.963794660 FAX:
agent.hostname: filebeat agent.id: 261f5285-1b4d-46b9-a650-4419242bf461 agent.version: 7.6.2 agent.type: filebeat agent.ephemeral_id: 1dc1bd52-e93c-
48af-b559-8e98c144750d NOM_PROVINCIA: VALENCIA TIPO: AGENCIA DE LECTURA DECRETO: S WEB: HTTP://WWW.BIBLIOTECASPUBLICAS.ES/MISLATA-VAL/INDEX.JSP
NOM_MUNICIPIO: MISLATA EMAIL: BIBLIOTECA@MISLATA.ES host.name: filebeat CP: 46,920 CENTRAL: N NOMBRE: AGENCIA DE LECTURA SANTA CECILIA MISLATA
DIRECCION: CALLE SANTA CECILIA N° 4 log.file.path: /var/log/bibliotecas-comunidad-valenciana-2020.ndjson log.offset: 23,985 input.type: log

@timestamp: May 4, 2020 @ 10:53:04.809 COD_CHARACTER: PU FAX: FAX.962261393 log.offset: 24,488 log.file.path: /var/log/bibliotecas-comunidad-valenciana-
2020.ndjson TIPO: AGENCIA DE LECTURA WEB: HTTP://XLPV.GVA.ES/CGINET-BIN/ABNETOP?SUBC=0137 CP: 46,640 DECRETO: S ecs.version: 1.4.0 host.name: filebeat
CENTRAL: N COD_PROVINCIA: 46 DIRECCION: CALLE STAS.RELIQUIAS N° 3 PISO 1 NOM_MUNICIPIO: MOIXENT NOM_PROVINCIA: VALENCIA input.type: log
agent.ephemeral_id: 1dc1bd52-e93c-48af-b559-8e98c144750d agent.hostname: filebeat agent.id: 261f5285-1b4d-46b9-a650-4419242bf461 agent.version: 7.6.2
agent.type: filebeat EMAIL: BIBLIOTECA@MOIXENT.ES DESC_CHARACTER: PÚBLICA id: 52 COD_MUNICIPIO: 170 CATALOGO: HTTP://XLPV.GVA.ES/CGINET-BIN/ABNETOP?

```