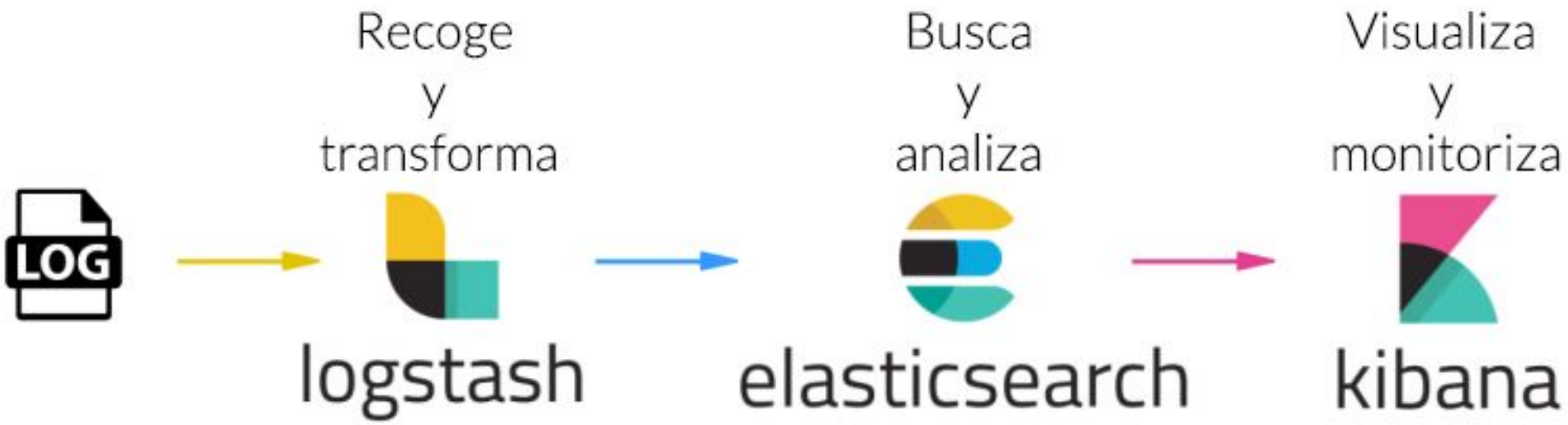
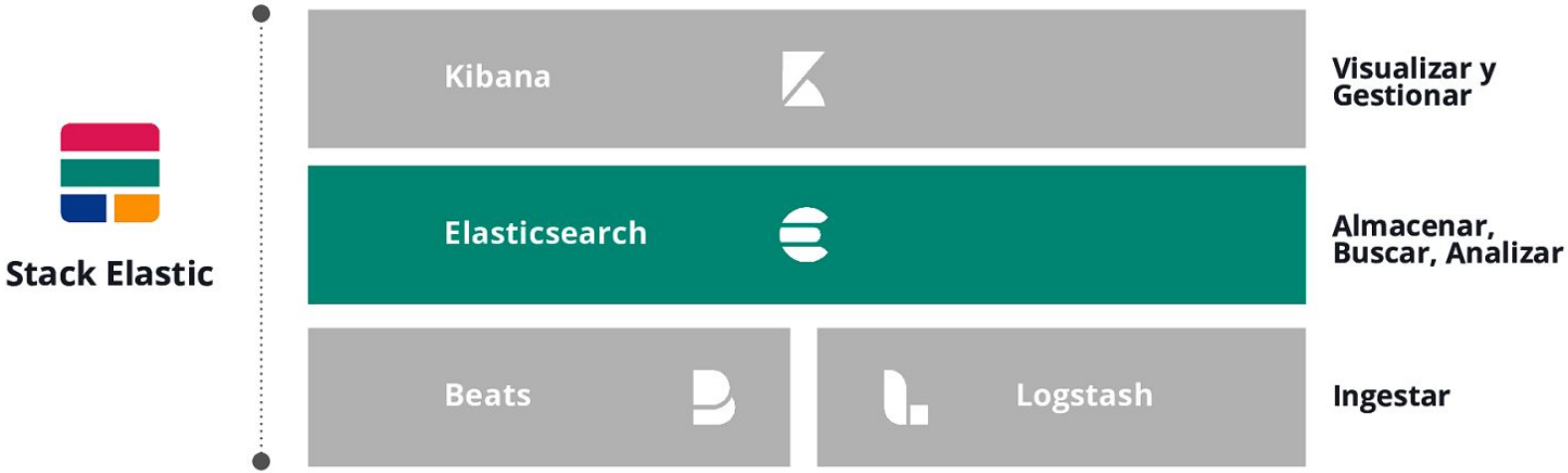

elasticsearch
logstash
kibana


LOG
Recoge
y
transforma
logstash
Busca
y
analiza
elasticsearch
Visualiza
y
monitoriza
kibana

Elasticsearch



elastic

INTRODUCCIÓN A ELASTICSEARCH

- Introducción y conceptos
- Preparación del entorno
- Arquitectura

Introducción y conceptos

https://db-engines.com/en/ranking/search+engine

Rank			DBMS	Database Model	Score		
Sep 2021	Aug 2021	Sep 2020			Sep 2021	Aug 2021	Sep 2020
1.	1.	1.	Elasticsearch	Search engine, Multi-model	160.24	+3.16	+9.74
2.	2.	2.	Splunk	Search engine	91.61	+1.01	+3.71
3.	3.	3.	Solr	Search engine, Multi-model	49.81	-1.26	-1.81
4.	4.	4.	MarkLogic	Multi-model	9.60	+0.45	-2.34
5.	5.	↑ 7.	Sphinx	Search engine	7.64	-0.11	+1.17
6.	6.	↓ 5.	Algolia	Search engine	7.30	+0.24	+0.53
7.	7.	↓ 6.	Microsoft Azure Search	Search engine	6.85	+0.71	+0.15
8.	8.	8.	ArangoDB	Multi-model	4.79	+0.53	-1.01
9.	9.	↑ 10.	Virtuoso	Multi-model	4.42	+0.19	+1.86
10.	10.	↓ 9.	Amazon CloudSearch	Search engine	2.20	+0.02	-0.41

Introducción y conceptos

Elasticsearch es un motor de búsqueda:

- Desarrollado en java.
- Open source.
- Distribuido.
- Escalable.
- Basado en **lucene.**

elastic

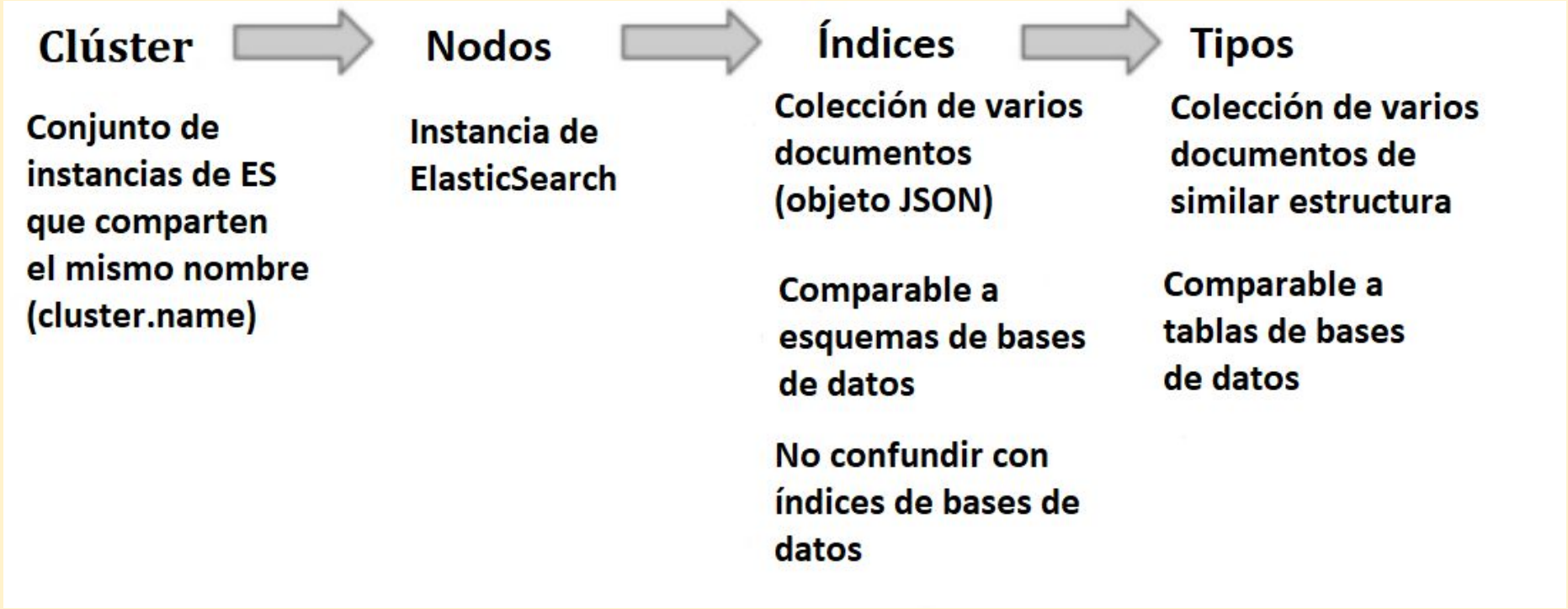
Introducción y conceptos

Lucene es una librería de búsqueda de texto:

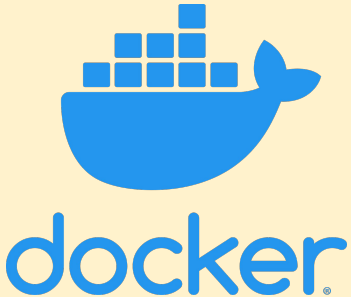
- Desarrollado en java.
- Open source.
- Escalable.
- Con alto rendimiento.
- Basada en **índices invertidos.**

Introducción y conceptos

Introducción y conceptos



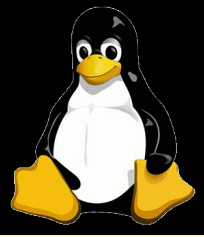
Preparación del entorno



https://github.com/jmortega/curso-elk

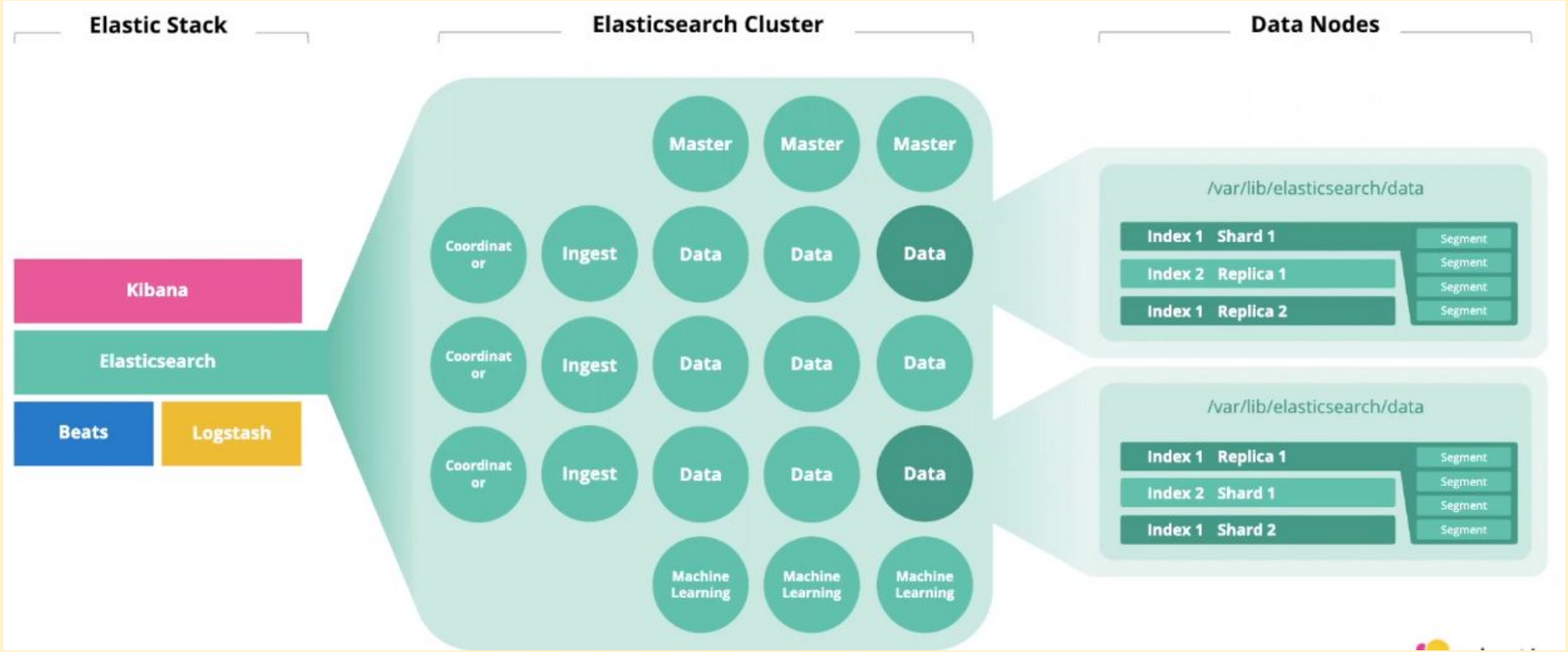
Preparación del entorno

```
$ sudo sysctl -w vm.max_map_count=262144
$ sudo sysctl -p
$ cat /proc/sys/vm/max_map_count
262144
```

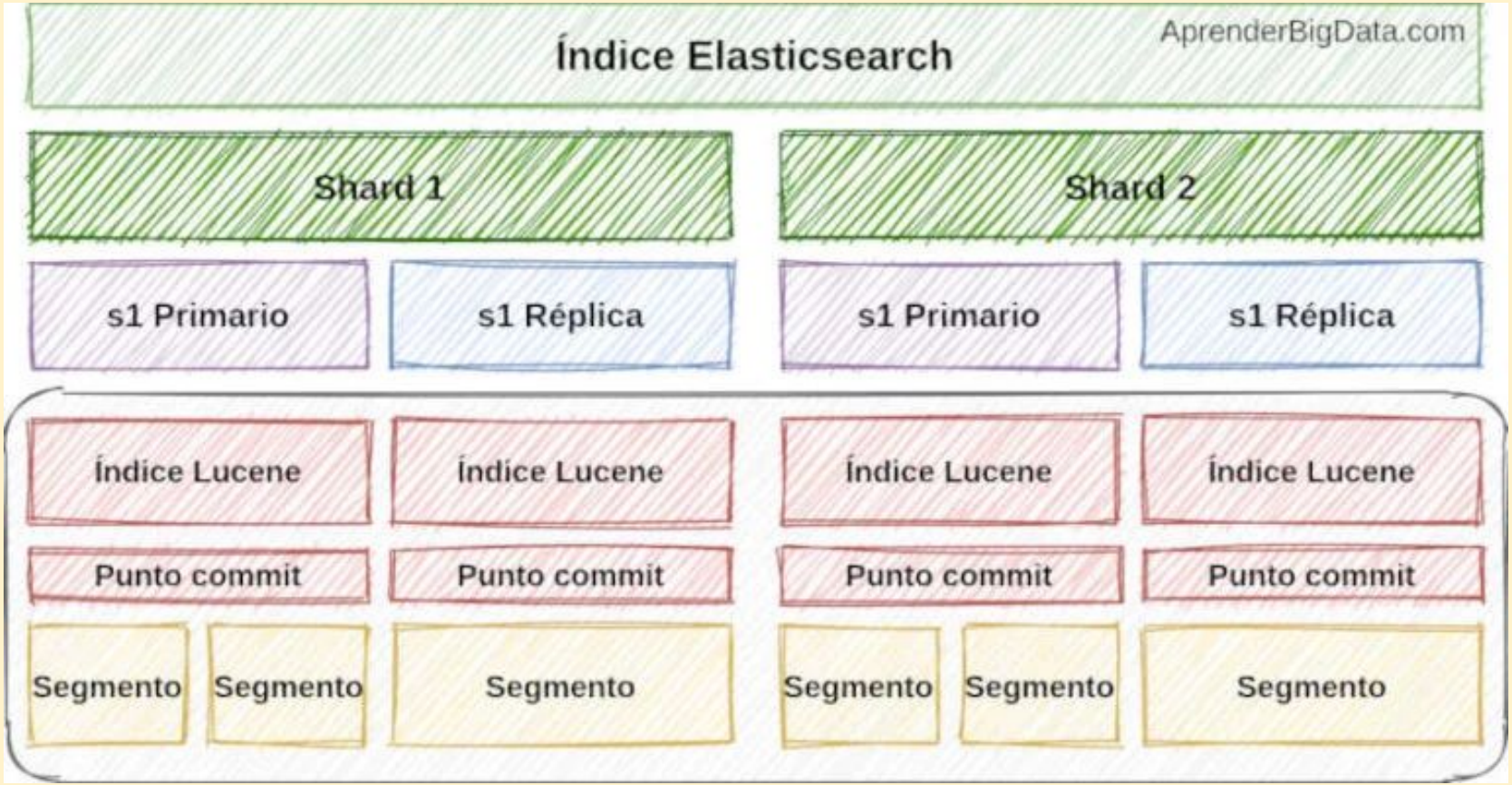


```
wsl -d docker-desktop
sysctl -w vm.max_map_count=262144
```

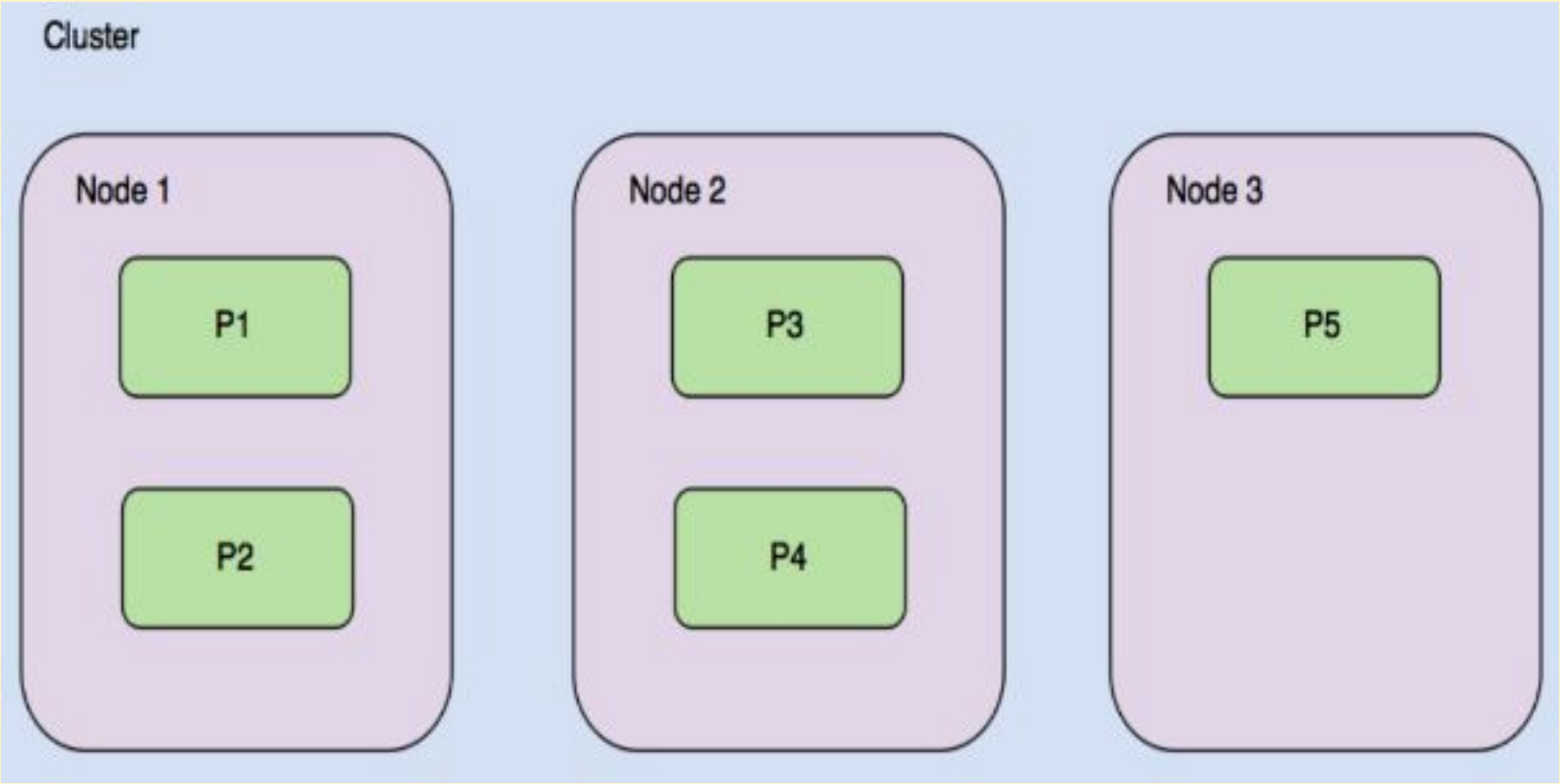
Arquitectura



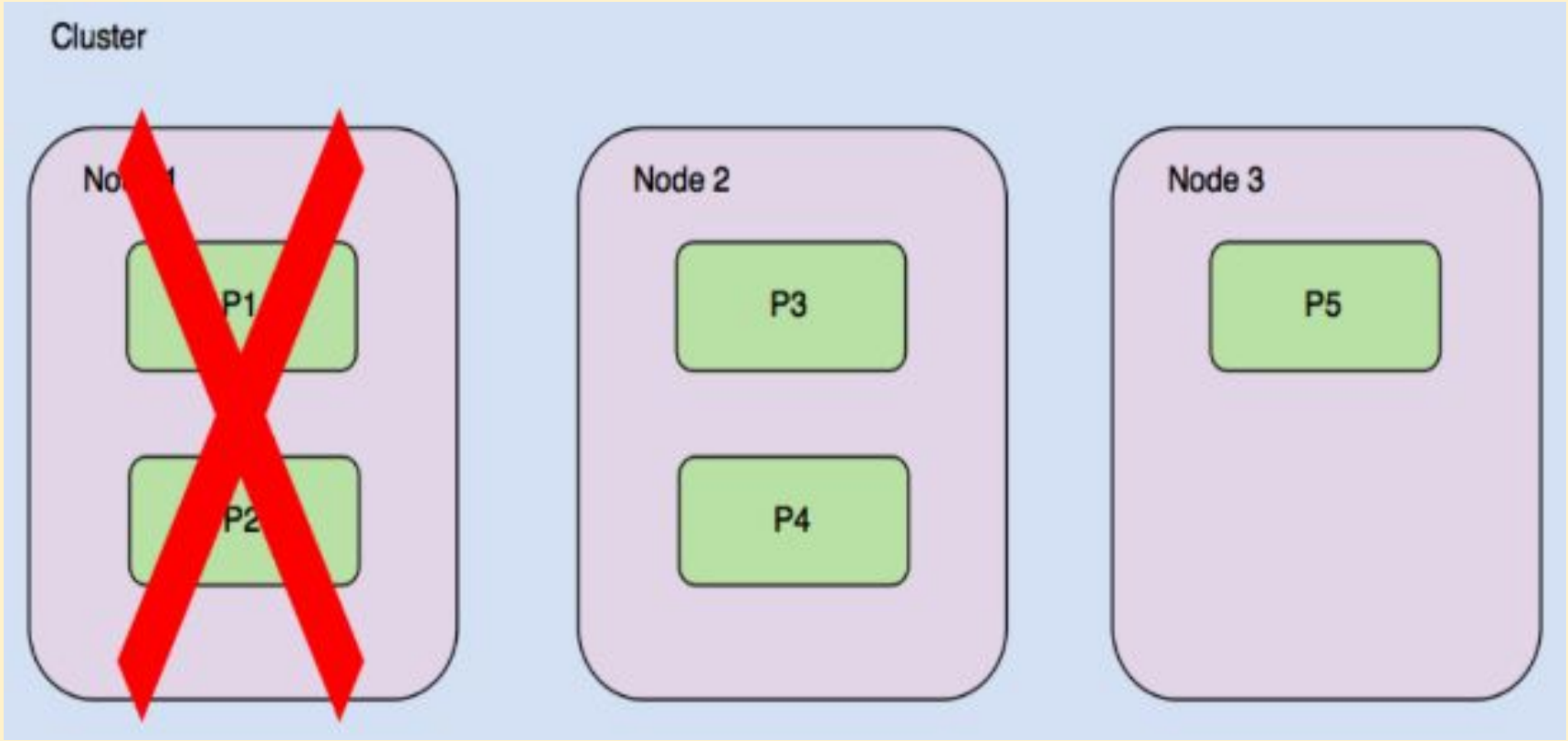
Shards


Índice Elasticsearch
AprenderBigData.com
Shard 1
Shard 2
s1 Primario
s1 Réplica
s1 Primario
s1 Réplica
Índice Lucene
Índice Lucene
Índice Lucene
Índice Lucene
Punto commit
Punto commit
Punto commit
Punto commit
Segmento
Segmento
Segmento
Segmento
Segmento
Segmento

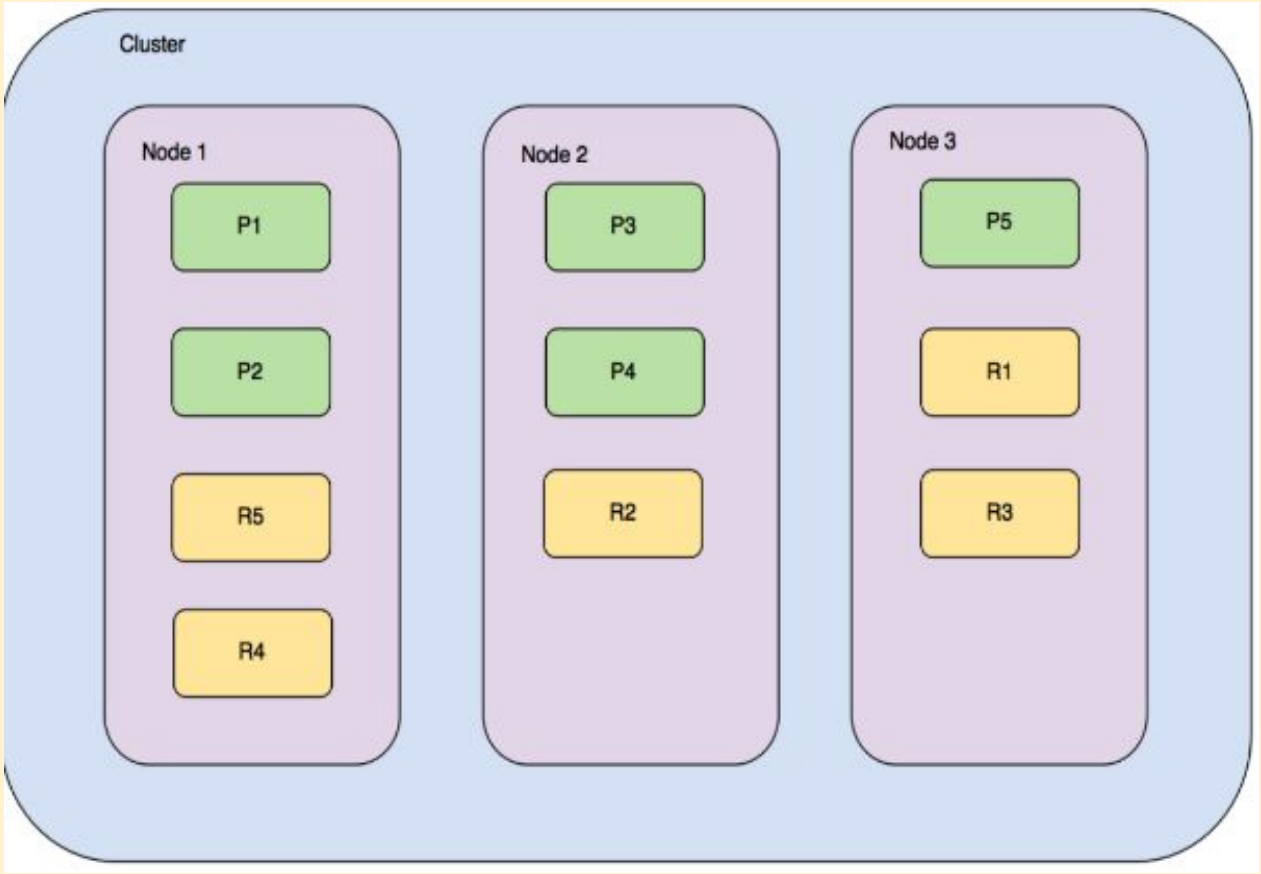
Shards y réplicas



Shards y réplicas



Shards y réplicas



Estados del cluster

Verde: Todo OK. Todos los shards y sus réplicas se han asignado a los nodos del cluster.

Amarillo: Todos los shards primarios se han asignado a nodos del cluster. Una o más réplicas no se han podido asignar a ningún nodo.

Rojo: Uno o más shards primarios no se han podido asignar a ningún nodo.

Estados del cluster

GET /_cluster/health


```
{
  "cluster_name" : "docker-cluster",
  "status" : "yellow",
  "timed_out" : false,
  "number_of_nodes" : 1,
  "number_of_data_nodes" : 1,
  "active_primary_shards" : 38,
  "active_shards" : 38,
  "relocating_shards" : 0,
  "initializing_shards" : 0,
  "unassigned_shards" : 6,
  "delayed_unassigned_shards" : 0,
  "number_of_pending_tasks" : 0,
  "number_of_in_flight_fetch" : 0,
  "task_max_waiting_in_queue_millis" : 0,
  "active_shards_percent_as_number" : 86.36363636363636
}
```

CRUD

- Crear índices
- Añadir documentos
- Leyendo documentos
- Actualizando documentos
- Actualización con script
- Upserts
- Borrar documentos
- Borrar índices
- Importando datos con cURL
- Actualizaciones en lote

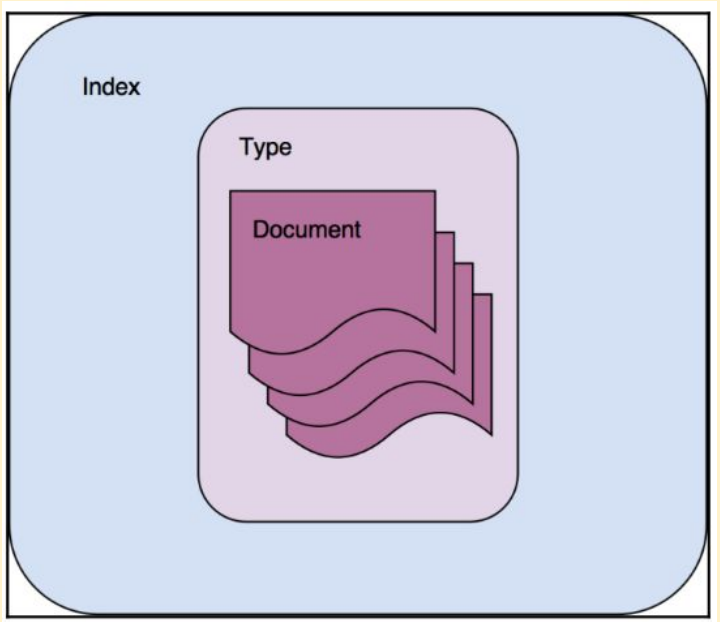
Crear índices

```
$ curl -XPUT 'http://elastic:elastic@localhost:9200/mi_indice/'
```

```
$ curl -XGET
'http://elastic:elastic@localhost:9200/mi_indice?pretty'
```

Crear índices

```
PUT /my_index
{
"settings": { ... settings ... },
"mappings": {
"type": { ... mappings ... }
}
}
```



Crear índices

```
PUT /catalog
{
  "settings": {
    "index": {
      "number_of_shards": 1,
      "number_of_replicas": 1,
      "codec": "best_compression"
    }
  }
}
```

Crear índices

```
PUT /catalog
{
  "mappings": {
     "properties": {
       "texto": {
         "type": "text"
       },
       "keyword": {
         "type": "keyword"
       }
     }
  }
}
```

Elasticsearch

¿Cómo almacenar los datos?

- Datos en origen en distintos formatos: BBDD, CSV, etc.

FlightNum	Origin	Dest	timestamp	FlightDelayMin	Cancelled
652J76O	Amsterdam Airport Schiphol	Jorge Chavez International Airport	Mar 8, 2020 @ 21:08:25.000	225	false
O9N0WNY	Shanghai Hongqiao International Airport	Ukrainka Air Base	Mar 8, 2020 @ 21:03:31.000	45	false
7ZS3Z50	Guangzhou Baiyun International Airport	Munich Airport	Mar 8, 2020 @ 20:21:16.000	285	true

- Se deben convertir a **objetos JSON** para enviar al **API REST** de Elasticsearch. Los documentos se almacenan en **índices**, agrupaciones lógicas de **Lucene shards**.

```
{
  "FlightNum": "652J76O",
  "Origin": "Amsterdam Airport Schiphol",
  "Dest": "Jorge Chavez International Airport",
  "timestamp": "Mar 8, 2020 @ 21:08:25.000",
  "FlightDelayMin": 225,
  "Cancelled": false
}
```

Añadir documentos

PUT <nombre_indice>/_doc/<identificador>

```
PUT 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"
}
```

Añadir documentos

POST <nombre_indice>/_doc/

POST kibana_sample_data_flights/_doc/

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

Añadir documentos


```
{
"_index": "kibana_sample_data_flights",
"_type": "_doc",
"_id": "AVrASKqgaBGmnAMj1SBe",
"_version": 1,
"result": "created",
"_shards": {
"total": 2,
"successful": 1,
"failed": 0
},
"created": true
}
```

Obtener documentos

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

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

```
GET kibana_sample_data_flights/_doc/1?_source=Origin
```

Obtener documentos

```
GET /kibana_sample_data_flights/_doc/_mget
{
"ids" : [ "1", "AVbA4WNg7uqRWQFJiJSn" ]
}
```

```
GET /kibana_sample_data_flights/_mget
{
  "docs": [
    {
      "_id": "1"
    },
    {
      "_id": "AVbA4WNg7uqRWQFJiJSn"
    }
  ]
}
```

Obtener documentos de diferentes índices

```
GET /_mget
{
  "docs": [
    {
      "_index":"kibana_sample_data_flights",
      "_id": "1"
    },
    {
     "_index":"kibana_sample_data_log",
      "_id": "1"
    }
  ]
}
```

Actualizando documentos

```
PUT kibana_sample_data_flights/_doc/1
{
  "FlightNum": "652J760",
  "Origin": "Amsterdam Airport Schiphol",
  "Dest": "Stockholm-Arlanda Airport",
  "FlightDelayMin": 100,
  "Cancelled": false,
  "timestamp": "2020-05-01T05:21:34"
}
```

Actualizando documentos de forma parcial

```
POST kibana_sample_data_flights/_update/1
{
  "doc": {
    "FlightDelayMin": 260,
    "tags":["flight1","delayed"]
  }
}
```

Actualización con script

```
POST kibana_sample_data_flights/_doc/1/_update
{
  "script": "ctx._source.FlightDelayMin = 200"
 }
```

Actualización con script

```
POST kibana_sample_data_flights/_update/1
{
  "script": {
    "source": "if (ctx._source.Origin.contains(params.origin)) {
ctx._source.FlightDelayMin = 100 } else {
ctx._source.FlightDelayMin = 300 }",
    "lang": "painless",
    "params": {
      "origin": "Amsterdam Airport Schipholl"
    }
  }
}
```

Actualización con script

```
POST kibana_sample_data_flights/_doc/1/_update
{
  "script": {
    "source": "ctx._source.remove('tags')",
    "lang": "painless"
  }
}
```

Upserts

```
POST kibana_sample_data_flights/_update/1
{
  "script": {
    "source": "ctx._source.FlightDelayMin += params.count",
    "lang": "painless",
    "params": {
      "count": 4
    }
  },
  "upsert": {
    "origin": "new airport"
  }
}
```

Upserts

```
POST kibana_sample_data_flights/_update/1
{
  "scripted_upsert": true,
  "script": {
    "source": "ctx._source.FlightDelayMin += params.count",
    "lang": "painless",
    "params": {
      "count": 4
    }
  },
  "upsert": {
    "origin": "new airport",
    "FlightDelayMin":200
  }
}
```

Upserts

```
POST kibana_sample_data_flights/_update/1
{
  "doc": {
    "FlightDelayMin": 260,
    "tags": [
      "flight1",
      "delayed"
    ]
  },
  "doc_as_upsert": true
}
```

Borrar documentos

```
DELETE kibana_sample_data_flights/_doc/<id>
```

```
POST kibana_sample_data_flights/_delete_by_query
{
"query":{"match":{"_id":<id>}}
}
```

Borrar índices

```
curl -XDELETE 'http://localhost:9200/<indice>'
```

```
curl -XPOST 'http://localhost:9200/<indice>/_open'
```

```
curl -XPOST 'http://localhost:9200/<indice>/_close'
```

Importando datos con cURL

$ cat data.json

```
{ "create": { "_index": "my_index", "_id": "1" }}
{ "name": "my name", "date": "20021-01-01","country": "Spain" }
{ "create": { "_index": "my_index", "_id": "2" }}
{ "name": "my name1", "date": "20021-01-01","country": "Italy" }
{ "create": { "_index": "my_index", "_id": "3" }}
{ "name": "my name2", "date": "20021-01-02","country": "France" }
```

$ curl -s -H "Content-Type: application/json" -XPOST http://elastic:elastic@localhost:9200/_bulk --data-binary "@data.json"; echo


```
{"took":2009,"errors":false,"items":[{"create":{"_index":"my_index","_type":"_doc","_id":"1","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":0,"_primary_term":1,"status":201}},{"create":{"_index":"my_index","_type":"_doc","_id":"2","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":1,"_primary_term":1,"status":201}},{"create":{"_index":"my_index","_type":"_doc","_id":"3","_version":1,"result":"created","_shards":{"total":2,"successful":1,"failed":0},"_seq_no":2,"_primary_term":1,"status":201}}]}
```

Actualizaciones en lote

```
POST /_bulk?pretty
{"delete":{"_index":"kibana_sample_data_flights","_type":"_doc","_id":"1"}}
{"create":{"_index":"kibana_sample_data_flights","_type":"_doc","_id":"1"}}
{"FlightNum":"652J760","Origin":"Amsterdam Airport
Schiphol","Dest":"Stockholm-Arlanda
Airport","FlightDelayMin":100,"Cancelled":false,"timestamp":"2020-05-01T05
:21:34"}
{"update":{"_index":"kibana_sample_data_flights","_type":"_doc","_id":"1"}
}
{"doc":{"FlightDelayMin":200}}
```

Estadísticas de un índice

GET http://localhost:9200/<nombre_indice>/_stats


```
},
"indexing" : {
  "index_total" : 18,
  "index_time_in_millis" : 21,
  "index_current" : 0,
  "index_failed" : 0,
  "delete_total" : 2,
  "delete_time_in_millis" : 1,
  "delete_current" : 0,
  "noop_update_total" : 1,
  "is_throttled" : false,
  "throttle_time_in_millis" : 0
},
```

Elasticsearch

¿Cómo almacenar los datos?

- Operaciones **CRUD** - **API REST**

```
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"
}
```

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

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

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

MAPPING

- Tipos de datos
- Mapping dinámico
- Meta fields
- Añadiendo mappings a índices existentes
- Parámetros de Mapping
- Formatos personalizados para fechas

Tipos de datos

https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-types.html

- **text**
- **keyword**
- **byte, short, integer, and long**
- **bool**
- **date**
- **Tipo de datos array**
- **Tipo de datos de objeto:** permite objetos dentro de documentos JSON.
- **Tipo de datos anidado:** útil para admitir arrays de objetos

Tipos de datos

- **text:** Se utiliza para indexar campos con valores de texto completo (FullText), como por ejemplo, el contenido de un correo electrónico o la descripción de un producto. Estos campos se analizan, es decir, se pasan a través de un analizador para convertir la cadena en una lista de términos individuales antes de indexarse.
- **keyword:** Se utiliza para los campos con valores que no pueden ser analizados para descomponerse en una serie de términos individuales, es decir, que solamente pueden ser buscados por su valor exacto.

Tipos de datos

- **Tipo de datos de puntos geográficos (Geo-point):** permite almacenar puntos geográficos con longitud y latitud. El tipo de datos de puntos geográficos permite realizar consultas, como buscar en todos los documentos a una distancia de 2 km de un punto.
- **Tipo de datos de forma geográfica (Geo-shape):** permite almacenar formas geométricas como polígonos y mapas. Geo-shape permite consultas como buscar todos los elementos dentro de una forma.
- **Tipo de datos IP:** permite almacenar direcciones IPv4 e IPv6.

Tipos de datos geo point

```
PUT my-index-geo
{
  "mappings": {
    "properties": {
      "location": {
        "type": "geo_point"
      }
    }
  }
}
```

```
PUT my-index-geo/_doc/1
{
  "text": "Geo-point as an object",
  "location": {
    "lat": 41.12,
    "lon": -71.34
  }
}
```

Tipos de datos objeto

```
POST my_index_object_type/_doc/1
{
  "person": {
    "name": {
      "firstname": "Martin",
      "lastname": "Fowler"
    }
  }
}
```

Tipos de datos nested

```
PUT /library/_mapping
{
    "properties": {
      "title": {
        "type": "text"
      },
      "review": {
        "type": "nested",
        "properties": {
          "nickname": {
            "type": "text"
          }
        }
      }
    }
}
```

Mapping dinámico

```
PUT /catalog/_doc/1
{
  "sku": "SP000001",
  "title": "Elasticsearch for Hadoop",
  "description": "Elasticsearch for Hadoop",
  "author": "Author",
  "ISBN": "1785288997",
  "price": 26.99
}
```

Mapping dinámico

```
PUT /catalog/_doc/2
{
"sku": "SP000002",
"title": "Google Pixel Phone 32GB - 5 inch
display",
"description": "Google Pixel Phone 32GB",
"price": 400.0,
"long":100,
"resolution": "1440 x 2560 pixels",
"os": "Android 7.1"
}
```

Meta fields

- **_id:** este es el identificador único del documento dentro del tipo, al igual que una clave principal en una tabla de base de datos. Puede ser autogenerado o especificado por el usuario.
- **_type:** este campo contiene el tipo de documento.
- **_index**: este campo contiene el nombre de índice del documento.

Añadiendo mappings a índices existentes

```
PUT /<indice>/_mapping
{
"properties": {
"name": {
"type": "text"
}
}
}
```

Formatos personalizados para fechas

https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-date-format.html

```
PUT my-index-date
{
  "mappings": {
    "properties": {
      "date": {
        "type": "date"
      }
    }
  }
}
```

Formatos personalizados para fechas

```
PUT my-index-date/_doc/1
{
  "date": "2021-01-01"
}

PUT my-index-date/_doc/2
{
  "date": "2021-01-01T12:10:30Z"
}

GET my-index-date/_search
{
  "sort": { "date": "asc"}
}
```

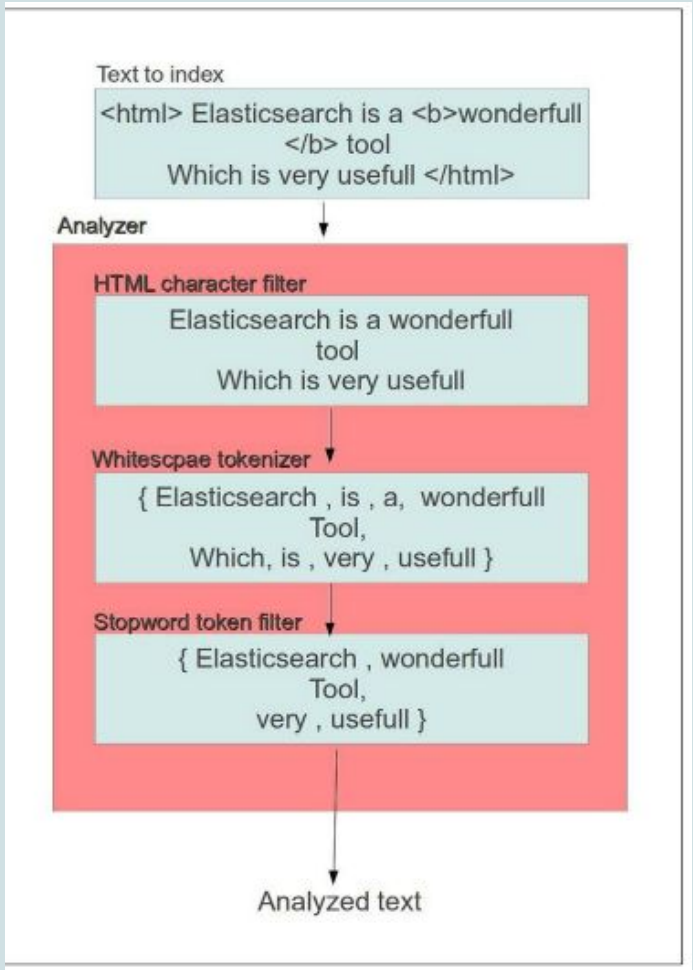
Formatos personalizados para fechas

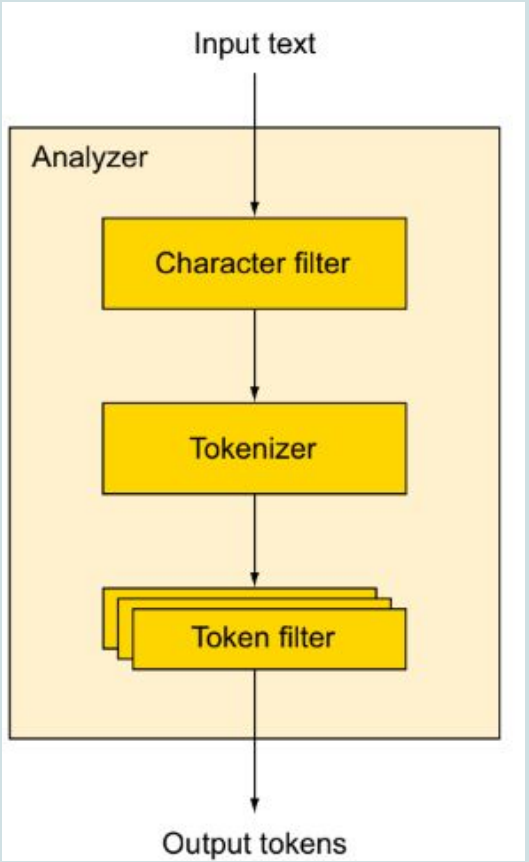
```
PUT my-index-date/_mapping
{
  "properties": {
    "date2": {
      "type": "date",
      "format": "yyyy-MM-dd HH:mm:ss||yyyy-MM-dd"
    }
  }
}
```

ANÁLISIS

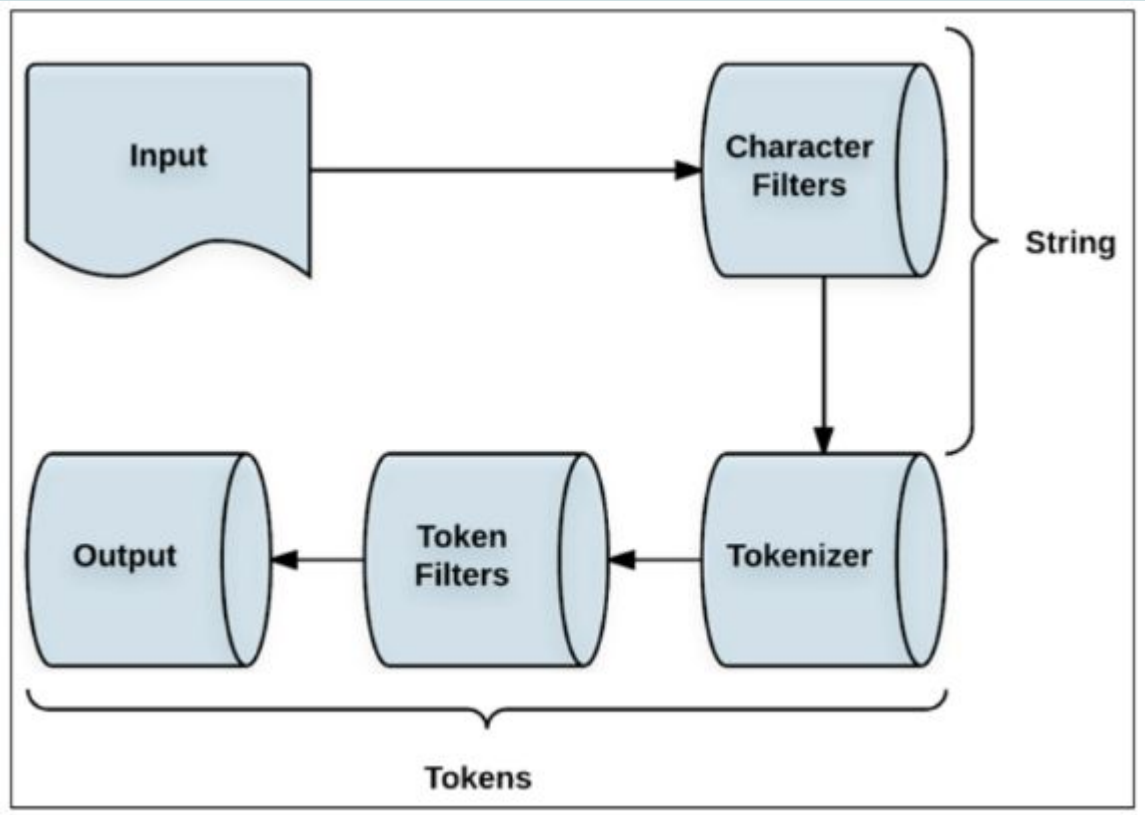
- Introducción
- Filtros de caracteres
- Tokenizers
- Filtros de Tokens
- API de Analyze
- Analizadores de sistema
- Analizadores personalizados

Introducción

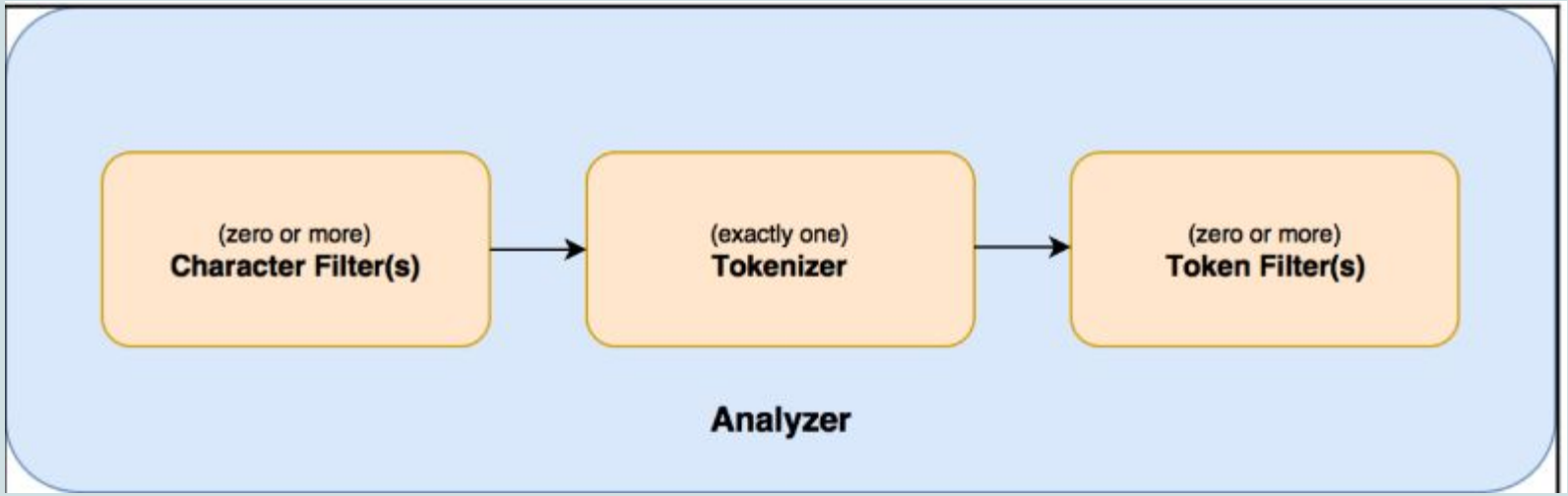




Analizadores



Composición de un analizador



Filtros de caracteres

```
GET /_analyze
{
  "tokenizer": "keyword",
  "char_filter": [
    {
      "type": "mapping",
      "mappings": [
        "&#35; => #",
        "&#37; => %",
        "&#38; => &"
      ]
    }
  ],
  "text": "HTML ASCII &#35; &#37; &#38;"
}
```

Filtros de caracteres

```
GET /_analyze
{
  "tokenizer": "keyword",
  "char_filter": [
    "html_strip",
    {
      "type": "mapping",
      "mappings": [
        "I'll => I will"
      ]
    }
  ],
  "text": "<p>I'll <b>learn elastic </b>!</p>"
}
```

Tokenizers

```
POST _analyze
{
"tokenizer": "standard",
"text": "Tokenizer breaks characters into tokens!"
}
```

Filtros de Tokens

- **Lowercase Token Filter:** reemplaza todos los tokens en la entrada con sus versiones en minúsculas.
- **Stop Token Filter**: Elimina las palabras vacías, es decir, las palabras que no añaden más significado al contexto.

Stop words

- https://www.ranks.nl/stopwords

Default English stopwords list

This list is used in our Page Analyzer and Article Analyzer for English text, when you let it use the default stopwords list.

a	ourselves
about	out
above	over
after	own
again	same
against	shan't
all	she
am	she'd
an	she'll

Analizadores de sistema

- **Standard Analyzer**: es el analizador predeterminado en Elasticsearch. Si no se reemplaza por ningún otro analizador de nivel de campo, nivel de tipo o nivel de índice, todos los campos se analizan utilizando este analizador.
- **Analizadores de idiomas:** los diferentes idiomas tienen diferentes reglas gramaticales.
- **Analizador de espacios en blanco:** el analizador de espacios en blanco divide la entrada en tokens siempre que encuentre un token de espacio en blanco, como un espacio, tabulación, nueva línea o retorno de carro.

Standard Analyzer

- **Standard Tokenizer**: un tokenizador que divide los tokens en caracteres de espacio en blanco.
- **Standard Token Filter**: el filtro de token estándar se utiliza como un filtro de token de marcador de posición dentro del analizador estándar.
- **Lowercase Token Filter:**: convierte todos los tokens en minúsculas.
- **Stop Token Filter:**: elimina las stop words especificadas.

Standard Analyzer

```
PUT index_standard_analyzer
{
"settings": {
"analysis": {
"analyzer": {
"std": {
"type": "standard"
}
}
}
},
"mappings": {
"properties": {
"my_text": {
"type": "text",
"analyzer": "std"
}
}
}
}
```

Standard Analyzer

```
POST index_standard_analyzer/_analyze
{
"field": "my_text",
"text": "The Standard Analyzer works this way."
}
```

Standard Analyzer stop words

```
PUT index_standard_analyzer_english_stopwords
{
"settings": {
"analysis": {
"analyzer": {
"std": {
"type": "standard",
"stopwords": "_english_"
}
}
}
},
"mappings": {
"properties": {
"my_text": {
"type": "text",
"analyzer": "std"
}
}
}
}
```

Standard Analyzer stop words

```
POST index_standard_analyzer_english_stopwords/_analyze
{
"field": "my_text",
"text": "The Standard Analyzer works this way."
}
```

Analizadores personalizados

```
PUT /my_index
{
"settings": {
"analysis": {
"char_filter": { ... custom character filters ... },
"tokenizer": { ... custom tokenizers ... },
"filter": { ... custom token filters ... },
"analyzer": { ... custom analyzers ... }
}
}
}
```

Analizadores personalizados

```
"keyword_tokenizer": {
"type": "custom",
"filter": [
"lowercase",
"asciifolding"
],
"tokenizer": "keyword"
}
```

Analizadores personalizados

PUT /custom_analyzer_index

```
{
"settings": {
"index": {
"analysis": {
"analyzer": {
"custom_analyzer": {
"type": "custom",
"tokenizer": "standard",
"filter": [
"lowercase",
"custom_edge_ngram"
]
}
},
}
```

```
"filter": {
"custom_edge_ngram": {
"type": "edge_ngram",
"min_gram": 2,
"max_gram": 10
}
}
}
}
},
"mappings": {
"my_type": {
"properties": {
"product": {
"type": "text",
"analyzer": "custom_analyzer",
"search_analyzer": "standard"
}
}
}
}
}
```

Analizadores personalizados

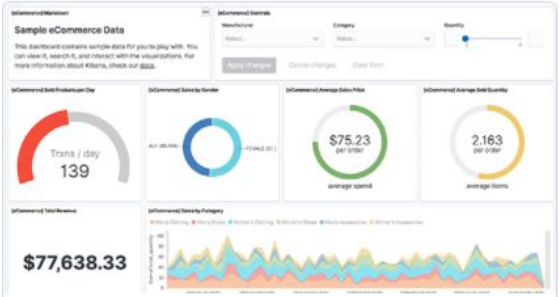
```
POST /custom_analyzer_index/my_type
{
"product": "Learning Elastic Stack 7"
}

POST /custom_analyzer_index/my_type
{
"product": "Mastering Elasticsearch"
}
```

Analizadores personalizados

```
GET /custom_analyzer_index2/_search
{
  "query": {
    "match": {
      "product": "Elastic"
    }
  }
}
```

Analizadores personalizados

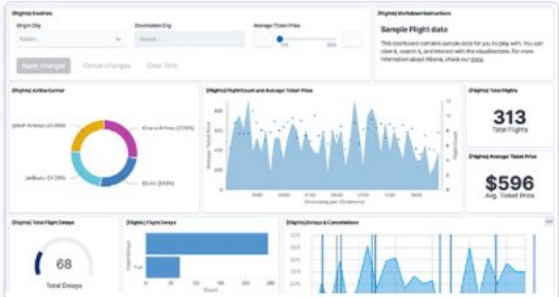
```
PUT my-index-tokenizer_max_length
{
  "settings": {
    "analysis": {
      "analyzer": {
        "my_analyzer": {
          "tokenizer": "my_tokenizer"
        }
      },
      "tokenizer": {
        "my_tokenizer": {
          "type": "standard",
          "max_token_length": 5
        }
      }
    }
  }
}
```



Sample eCommerce orders

Sample data, visualizations, and dashboards for tracking eCommerce orders.

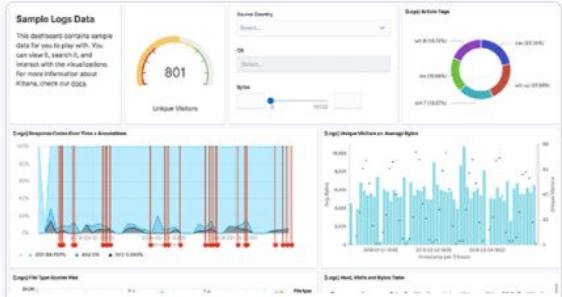
Add data



Sample flight data

Sample data, visualizations, and dashboards for monitoring flight routes.

Add data



Sample web logs

Sample data, visualizations, and dashboards for monitoring web logs.

Add data

Elasticsearch

Dos tipos de búsquedas

- **Queries**
  - ¿Qué vuelos tienen **origen** en “**Amsterdam**”?
  - ¿Qué vuelos se **retrasaron 60 minutos o más**?
- **Agregaciones**
  - ¿Cuáles son los **top 3** aeropuertos **origen**?
  - ¿Qué **retraso** tienen en **media** los vuelos en los **2 top** aeropuertos **origen**?

Elasticsearch

Búsquedas - Queries

Query DSL (Domain Specific Language) basado en JSON para definir queries

¿Qué vuelos tienen **origen** en **"Amsterdam"**?

```
GET kibana_sample_data_flights/_search
{
  "size": 1,
  "query": {
    "match": {
      "origin.text": "amsterdam"
    }
  }
}
```


```
{
  "took" : 1,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 1,
      "relation" : "eq"
    },
    "max_score" : 0.2876821,
    "hits" : [
      {
        "_index" : "flights",
        "_type" : "_doc",
        "_id" : "1",
        "_score" : 0.2876821,
        "_source" : {
          "FlightNum" : "652J76O",
          "Origin" : "Amsterdam Airport Schiphol",
          "Dest" : "Jorge Chavez International Airport",
          "timestamp" : "Mar 8, 2020 @ 21:08:25.000",
          "FlightDelayMin" : 225,
          "Cancelled" : false
        }
      }
    ]
  }
}
```


¿Qué vuelos se **retrasaron 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
      }
    }
  }
}
```


```
{
  "hits" : {
    "hits" : [
      {
        "_source" : {
          "FlightNum" : "EAYQW69",
          "FlightDelayMin" : 180
        }
      },
      {
        "_source" : {
          "FlightNum" : "EVARI8I",
          "FlightDelayMin" : 300
        }
      },
      {
        "_source" : {
          "FlightNum" : "RBFKZBX",
          "FlightDelayMin" : 120
        }
      },
      {
        "_source" : {
          "FlightNum" : "R43CELD",
          "FlightDelayMin" : 300
        }
      },
      {
        "_source" : {
          "FlightNum" : "1TJWK8F",
          "FlightDelayMin" : 90
        }
      },
```


elastic

Búsquedas

```
# Vuelos con origen en Amsterdam
GET kibana_sample_data_flights/_search
{
  "query": {
    "match": {
      "origin.text": "amsterdam"
    }
  }
}
```

Búsquedas

```
# Vuelos que se retrasan 60 minutos o mas
GET kibana_sample_data_flights/_search?filter_path=hits.hits._source
{
  "size": 10,
  "_source": ["FlightDelayMin","FlightNum"],
  "query": {
    "range": {
      "FlightDelayMin": {
        "gte": 60
      }
    }
  }
}
```

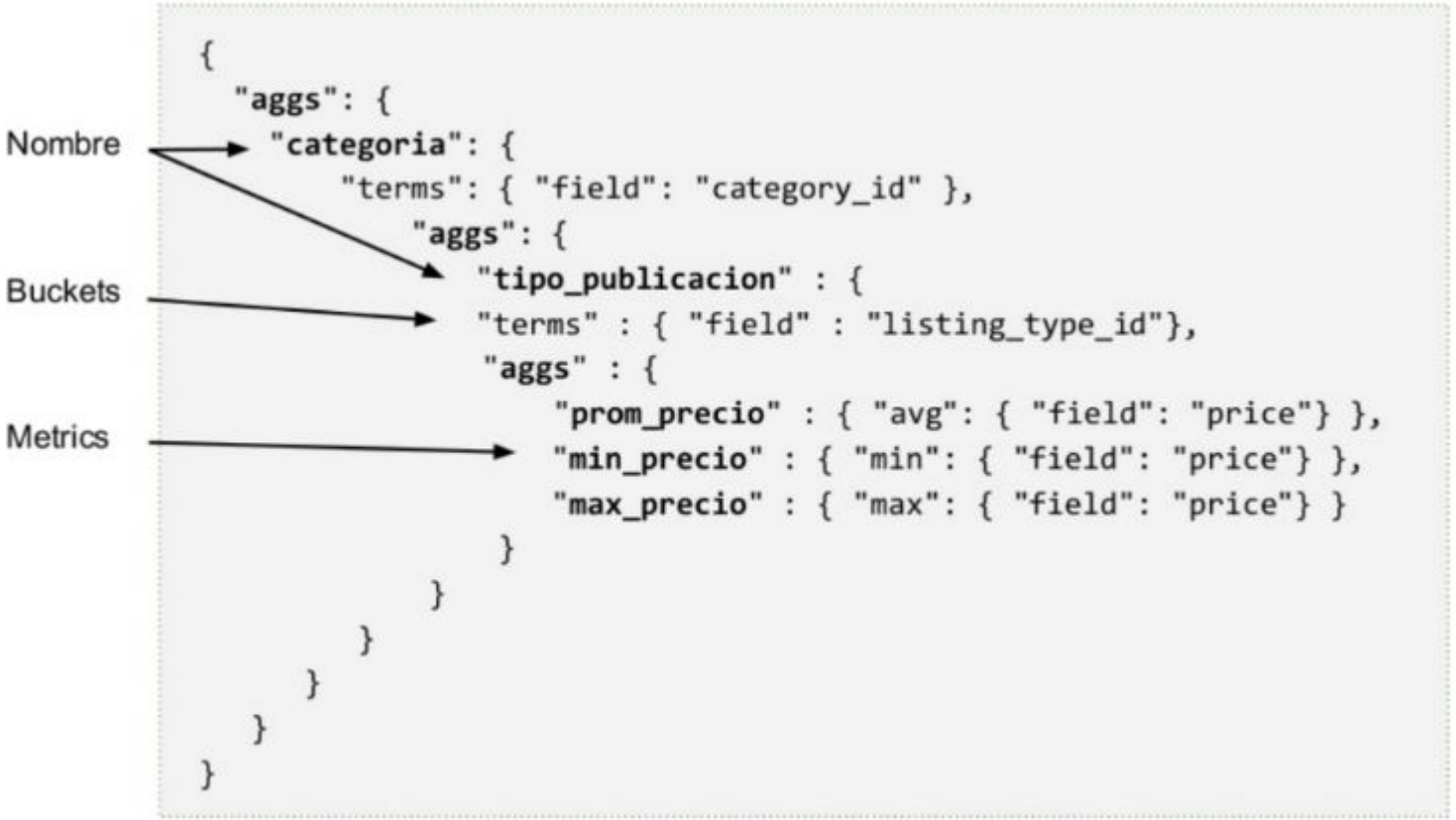
AGREGACIONES

- Buckets
- Agregaciones por métricas
- Agregaciones buckets
- Agregaciones estadísticas
- Agregaciones anidadas
- Agregaciones por rango
- Histogramas

Buckets

```
"aggs”: {
    “name_of_aggregation”: {
      “type_of_aggregation”: {
        “field”: “document_field_name”
}
```

Buckets


Nombre
Buckets
Metrics
{
"aggs": {
"categoria": {
"terms": { "field": "category_id" },
"aggs": {
"tipo_publicacion" : {
"terms" : { "field" : "listing_type_id"},
"aggs" : {
"prom_precio" : { "avg": { "field": "price"} },
"min_precio" : { "min": { "field": "price"} },
"max_precio" : { "max": { "field": "price"} }
}
}
}
}
}
}

Agregaciones por métricas

SELECT avg(column) FROM table;

Esta consulta calcula la puntuación media para una determinada columna.

```
"aggs" : {
"<name_of_aggregation>" : {
"avg" : {
"field" : "column"
}
}
}
```

Elasticsearch

Búsquedas - 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
      }
    }
  }
}
```


```
{
  "took" : 43,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 10000,
      "relation" : "gte"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "top_aeropuertos" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 12262,
      "buckets" : [
        {
          "key" : "Mariscal Sucre International Airport",
          "doc_count" : 285
        },
        {
          "key" : "Ministro Pistarini International Airport",
          "doc_count" : 258
        },
        {
          "key" : "El Dorado International Airport",
          "doc_count" : 254
        }
      ]
    }
  }
}
```


¿Qué **retraso** tienen en **media** los vuelos en los **2 top** 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"
          }
        }
      }
    }
  }
}
```


```
{
  "took" : 44,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 10000,
      "relation" : "gte"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "top_aeropuertos_origen" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 12516,
      "buckets" : [
        {
          "key" : "Mariscal Sucre International Airport",
          "doc_count" : 285,
          "media_retraso" : {
            "value" : 51.63157894736842
          }
        },
        {
          "key" : "Ministro Pistarini International Airport",
          "doc_count" : 258,
          "media_retraso" : {
            "value" : 45.0
          }
        }
      ]
    }
  }
}
```


elastic

Agregaciones por métricas

```
GET kibana_sample_data_flights/_search
{
 "aggregations": {
    "km_average": {
      "avg": {
        "field": "DistanceKilometers"
      }
    }
  },
  "size": 0
}
```

Agregaciones buckets

```
SELECT column, count(*) FROM table GROUP BY column;
```

Esta consulta divide la tabla por los diferentes valores de la columna y devuelve un recuento de documentos dentro de cada valor de la columna.

Agregaciones buckets

```
GET kibana_sample_data_flights/_search?size=0
{
  "query": {
    "match_all": {}
  },
  "aggregations": {
    "OriginCityName": {
      "terms": {
        "field": "OriginCityName",
        "size": 10
      }
    }
  }
}
```

Agregaciones buckets

```
"aggregations": {
  "aggregation_name": {
    "buckets": [
     {
       "key": value,
       "doc_count": value
     }
     ]
  }
}
```

Agregaciones por término

```
GET kibana_sample_data_flights/_search
{
  "aggs": {
    "by_Origin": {
      "terms": {
        "field": "Origin"
      }
    }
  },
   "size": 0
}
```

Agregaciones estadísticas

```
GET kibana_sample_data_flights/_search
{
 "aggregations": {
    "km_stats": {
      "stats": {
        "field": "DistanceKilometers"
      }
    }
  },
  "size": 0
}
```

Agregaciones estadísticas

```
GET kibana_sample_data_flights/_search
{
 "aggregations": {
    "km_extended_stats": {
      "extended_stats": {
        "field": "DistanceKilometers"
      }
    }
  },
  "size": 0
}
```

Agregación cardinalidad

```
GET kibana_sample_data_flights/_search
{
 "aggregations": {
    "unique_origin": {
      "cardinality": {
        "field": "Origin"
      }
    }
  },
  "size": 0
}
```

Histogramas

```
GET kibana_sample_data_flights/_search
{
  "aggs": {
    "by_km": {
      "histogram": {
        "field": "DistanceKilometers",
        "interval": 5000
      }
    }
  },
   "size": 0
}
```

Agregaciones por rango

```
GET kibana_sample_data_flights/_search
{
  "aggs": {
    "by_km": {
      "range": {
        "field": "DistanceKilometers",
        "ranges": [
          { "to": 1000 },
          { "from": 1000, "to": 5000 },
          { "from": 5000 }
        ]
      }
    },
    "size": 0
}
```

Agregaciones por rango

```
GET kibana_sample_data_flights/_search
{
  "aggs": {
    "by_km": {
      "range": {
        "field": "DistanceKilometers",
        "ranges": [
          { "key": "Upto 1000 km","to": 1000 },
          { "key": "From 1000 to 5000 km","from": 1000, "to": 5000 },
          { "key": "More than 5000 km","from": 5000 }
        ]
      }
    }
  },
   "size": 0
}
```

Agregaciones por geolocalización

```
GET kibana_sample_data_flights/_search
{
  "aggs": {
    "within_radius": {
      "geo_distance": {
        "field": "OriginLocation",
        "origin": {"lat": 50.033333,"lon": 8.570556},
        "ranges": [{"from": 1000,"to": 150000}]
      }
    }
  },
   "size": 0
}
```

Agregaciones por filtro

```
GET kibana_sample_data_flights/_search
{
  "aggs": {
    "filter_origin": {
      "filter": {
        "term": {
          "Origin": "Frankfurt am Main Airport"
        }
      }
    }
  },
  "size": 0
}
```

Agregación ip_range

```
GET /kibana_sample_data_logs/_search?size=0
{
  "aggs": {
    "ip_ranges": {
      "ip_range": {
        "field": "clientip",
        "ranges": [
          {
            "from": "223.87.60.0"
          },
          {
            "to": "223.87.60.255"
          }
        ]
      }
    }
  }
}
```

Agregaciones anidadas

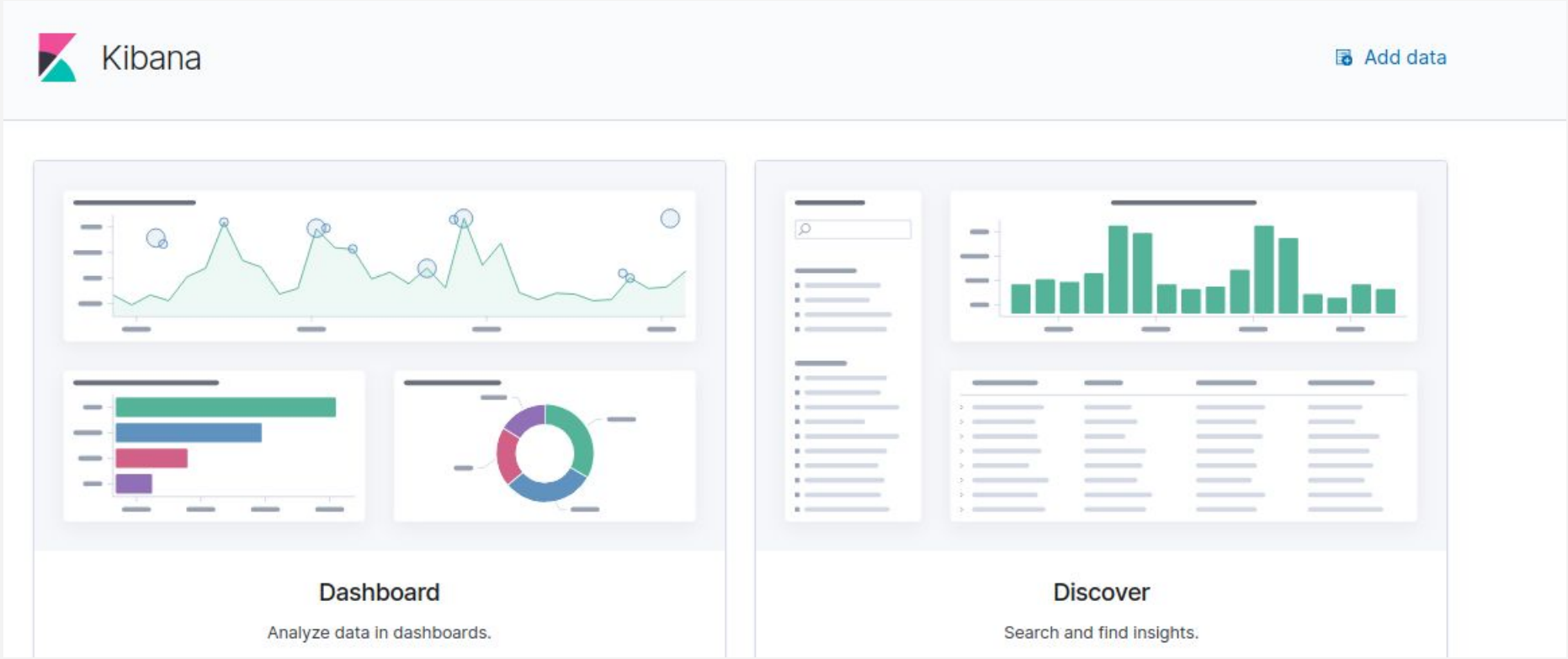
```
PUT nested_aggregation
{
  "mappings": {
    "properties": {
      "employee": {
        "type": "nested",
          "properties" : {
            "first_name" : { "type" : "text" },
           "last_name" : { "type" : "text" },
           "salary" : { "type" : "double" }
        }}}
}}
```

```
GET /nested_aggregation/_search
{
  "aggs": {
    "nested_aggregation" : {
            "nested": {
              "path": "employee"
            },
    "aggs": {
      "avg_salary": {
        "avg": {
          "field": "employee.salary"
        }
      }
    }
}}}
```

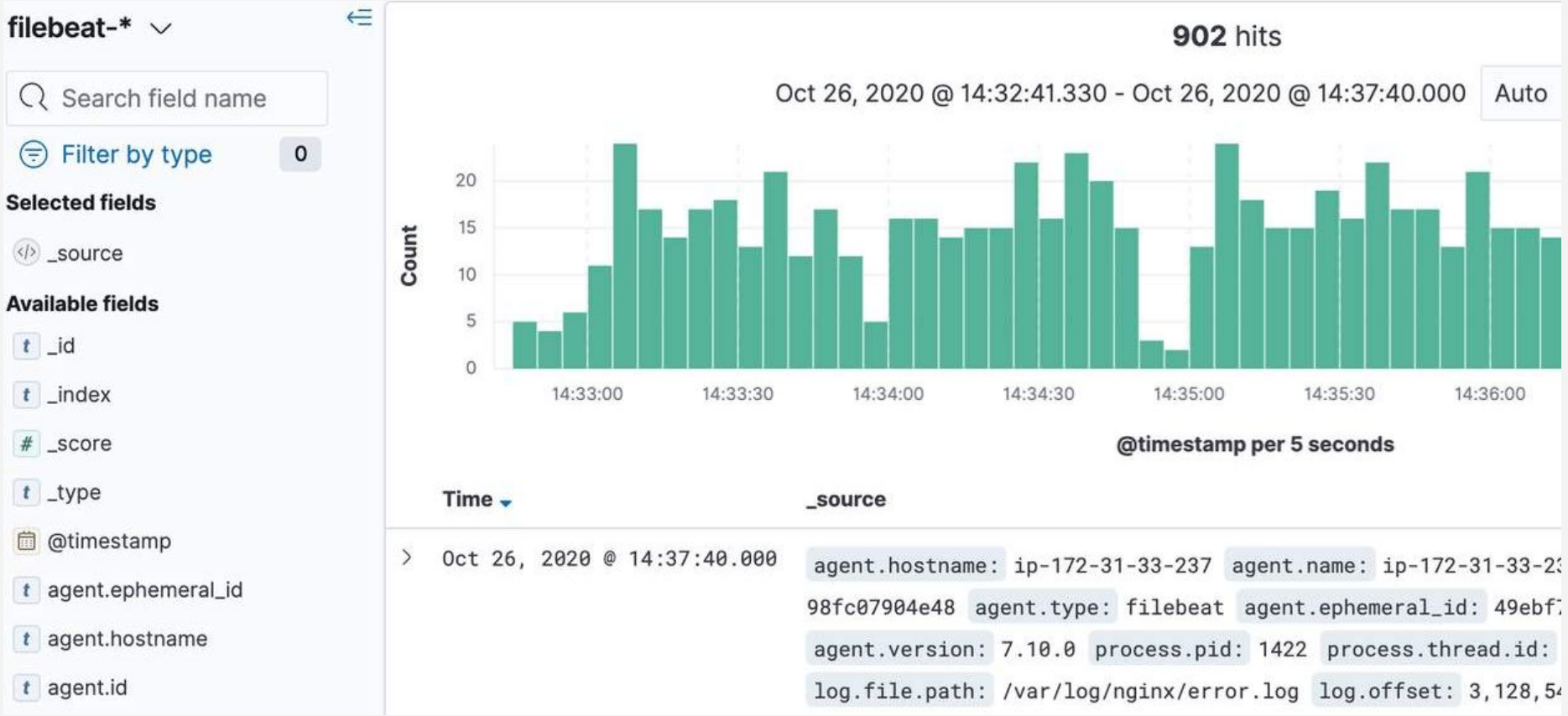
KIBANA

- Análisis con Kibana
- Configuración y administración
- Kibana Discover (Análisis de Logs)
- Visualizaciones: Histograma
- Visualizaciones: Pie
- Visualizaciones: Gauge
- Visualizaciones: Mapas
- Visualizaciones: Controles
- Visualizaciones: Metric
- Visualizaciones: Tablas de datos
- Otras visualizaciones
- Kibana Dashboards

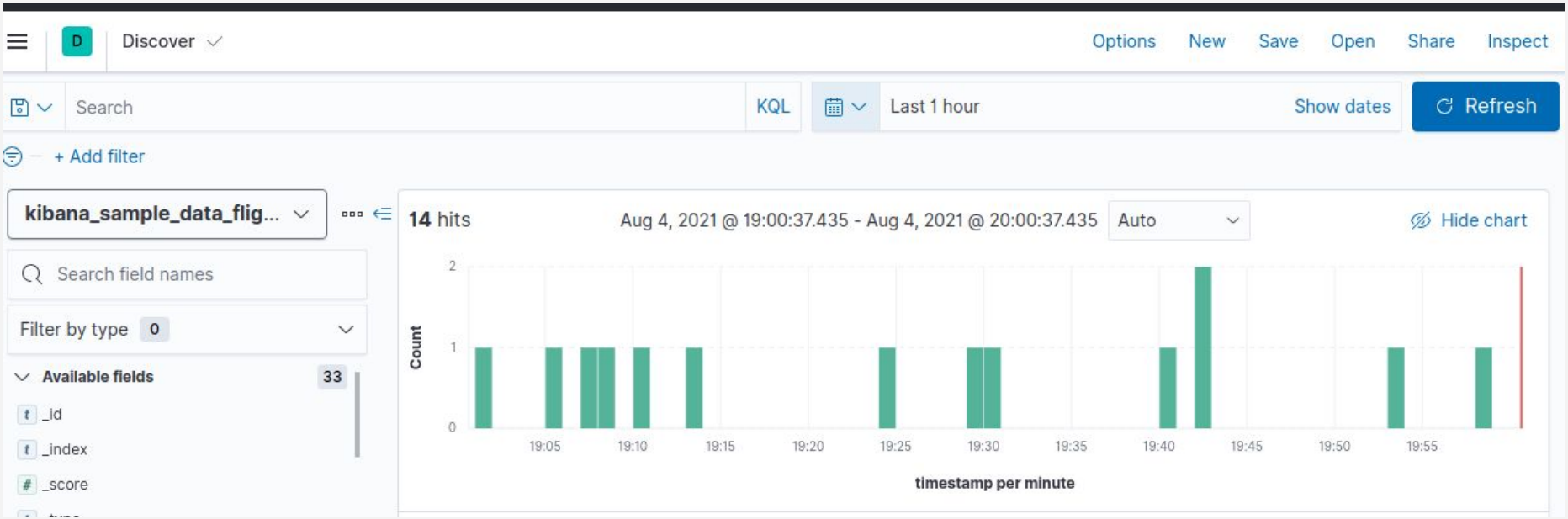
Análisis con Kibana



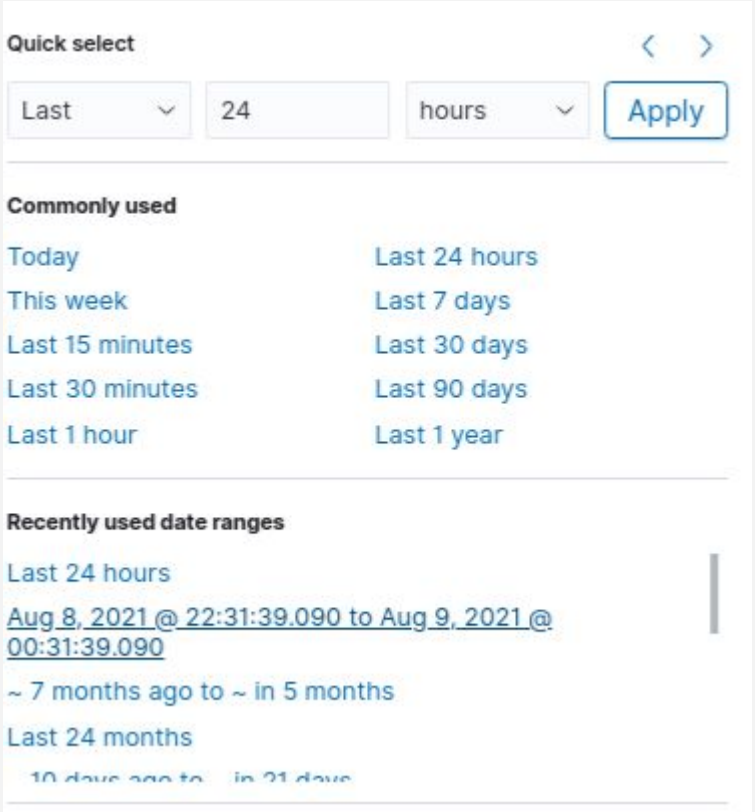
Kibana Discover (Análisis de Logs)



Kibana Discover (Análisis de Logs)

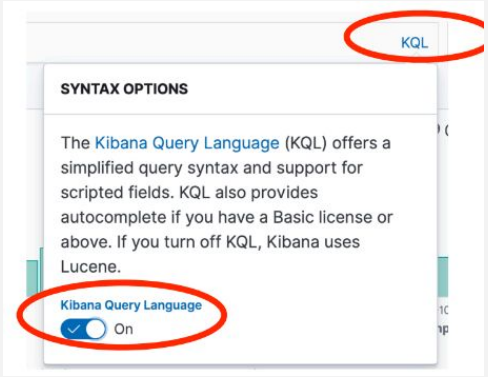


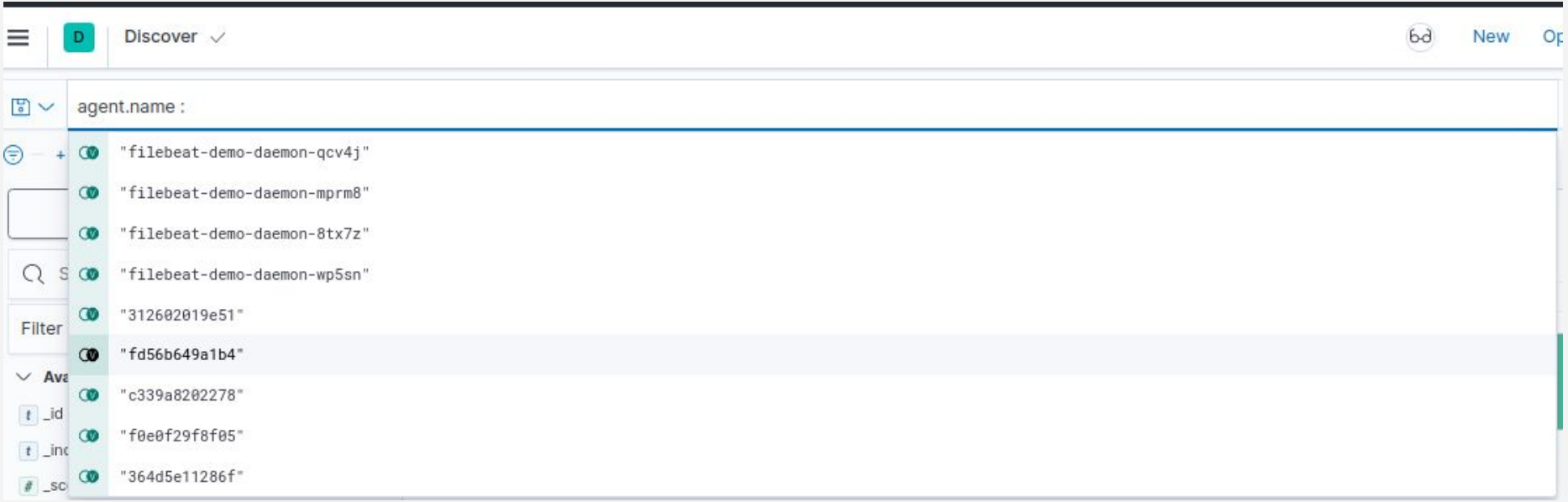
Kibana Discover (Análisis de Logs)



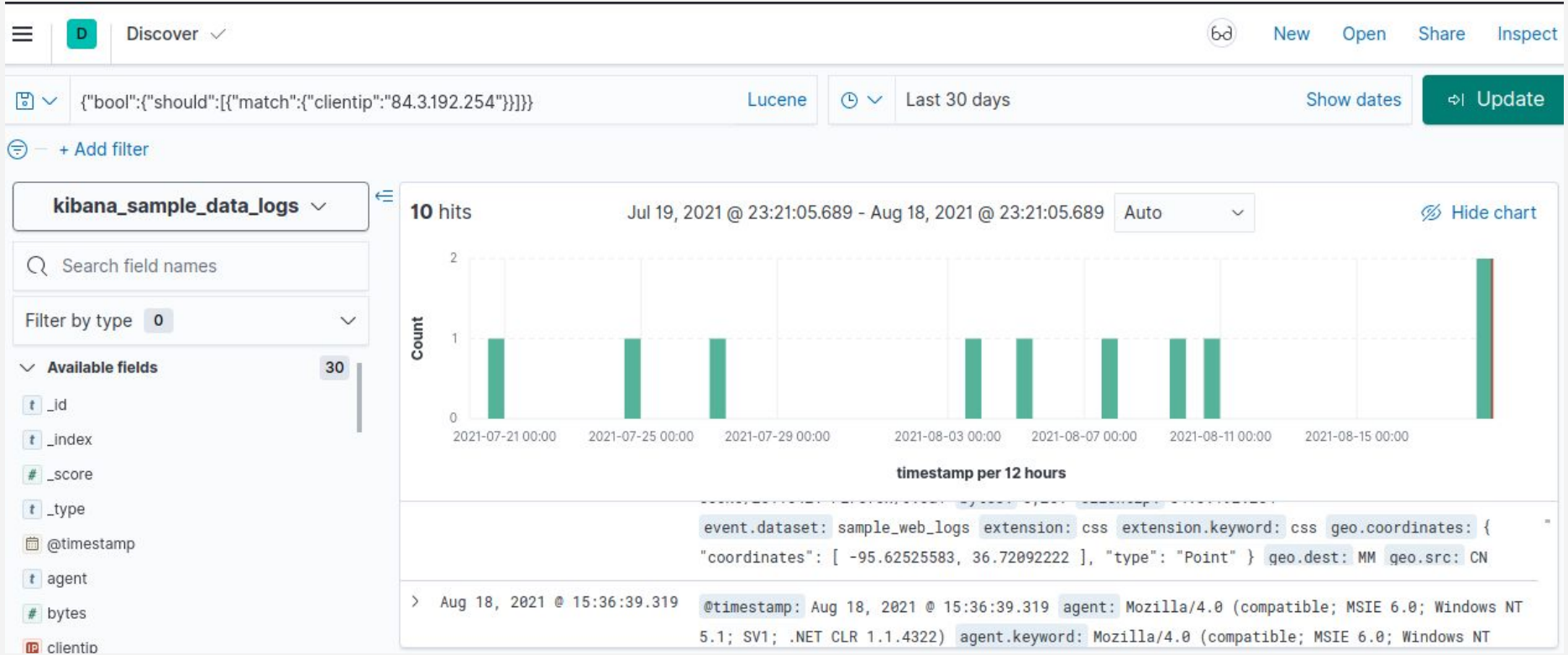
Kibana Discover (Análisis de Logs)

Kibana Discover (Análisis de Logs)

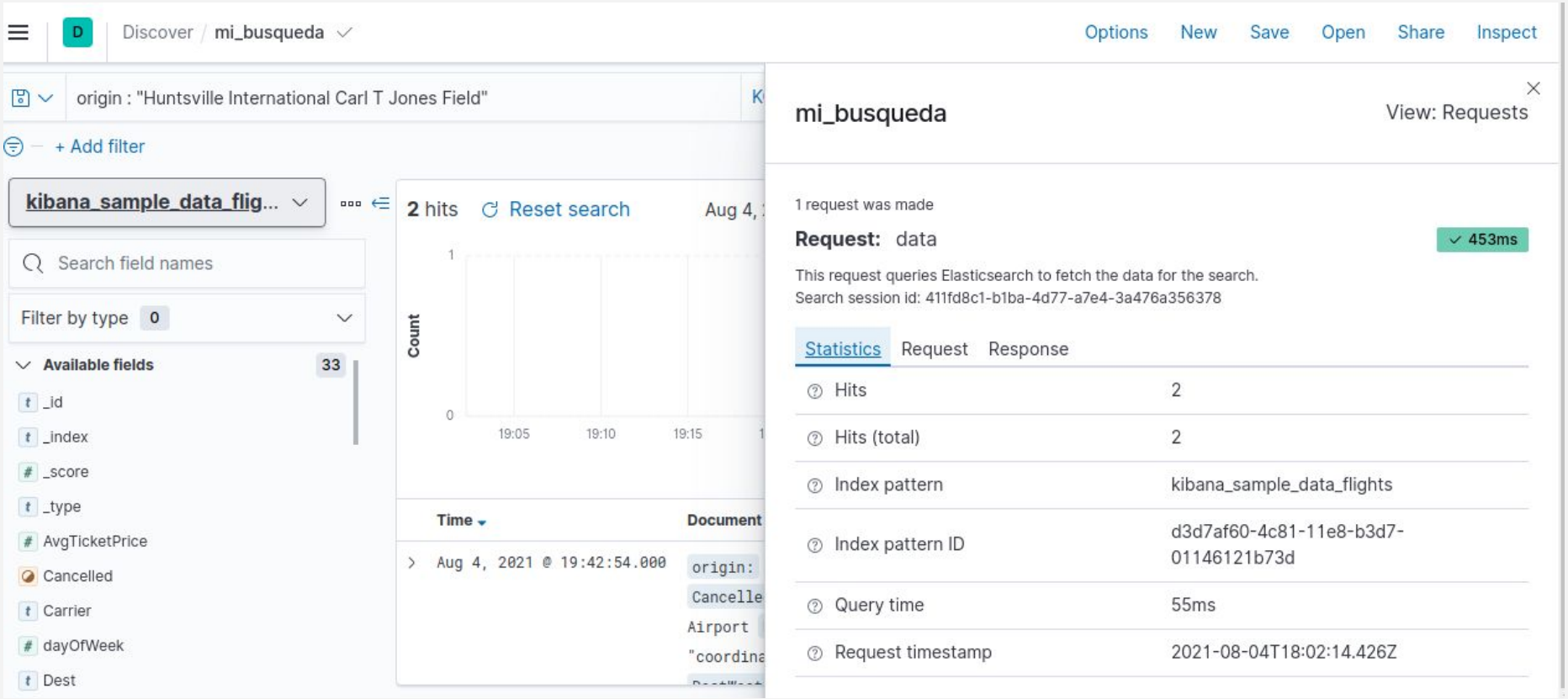




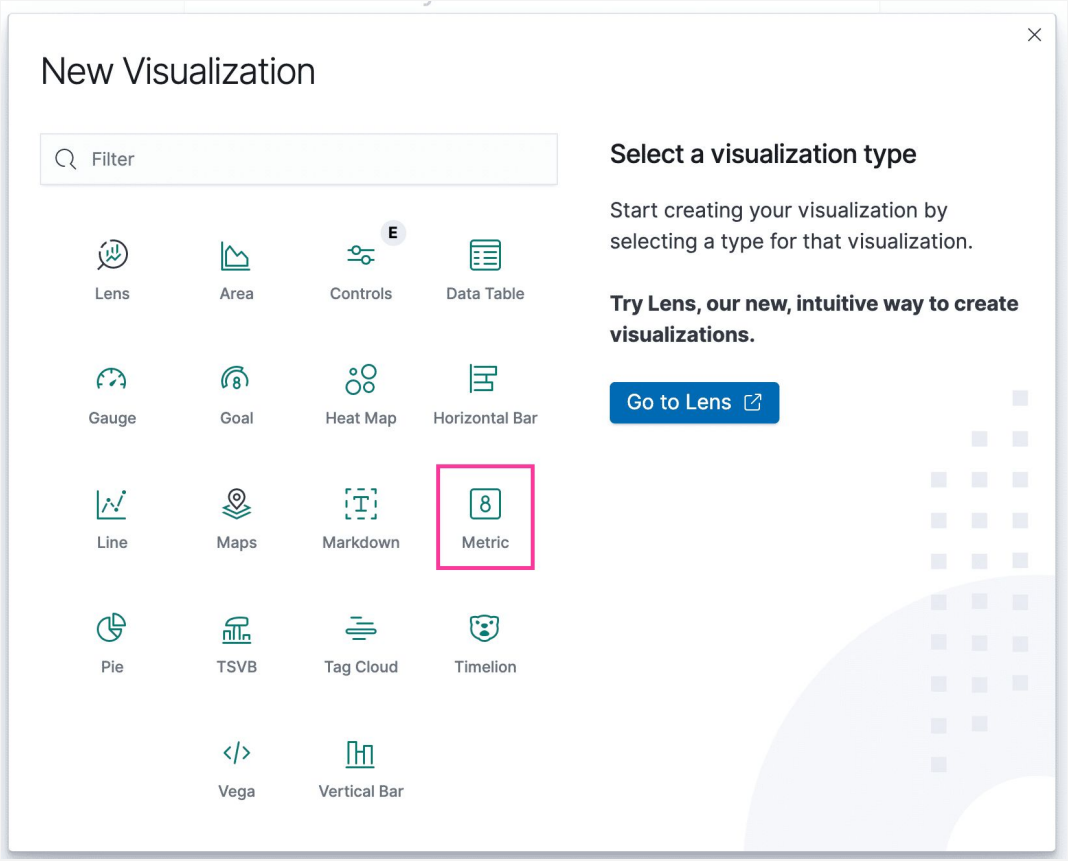
Kibana Discover (Análisis de Logs)



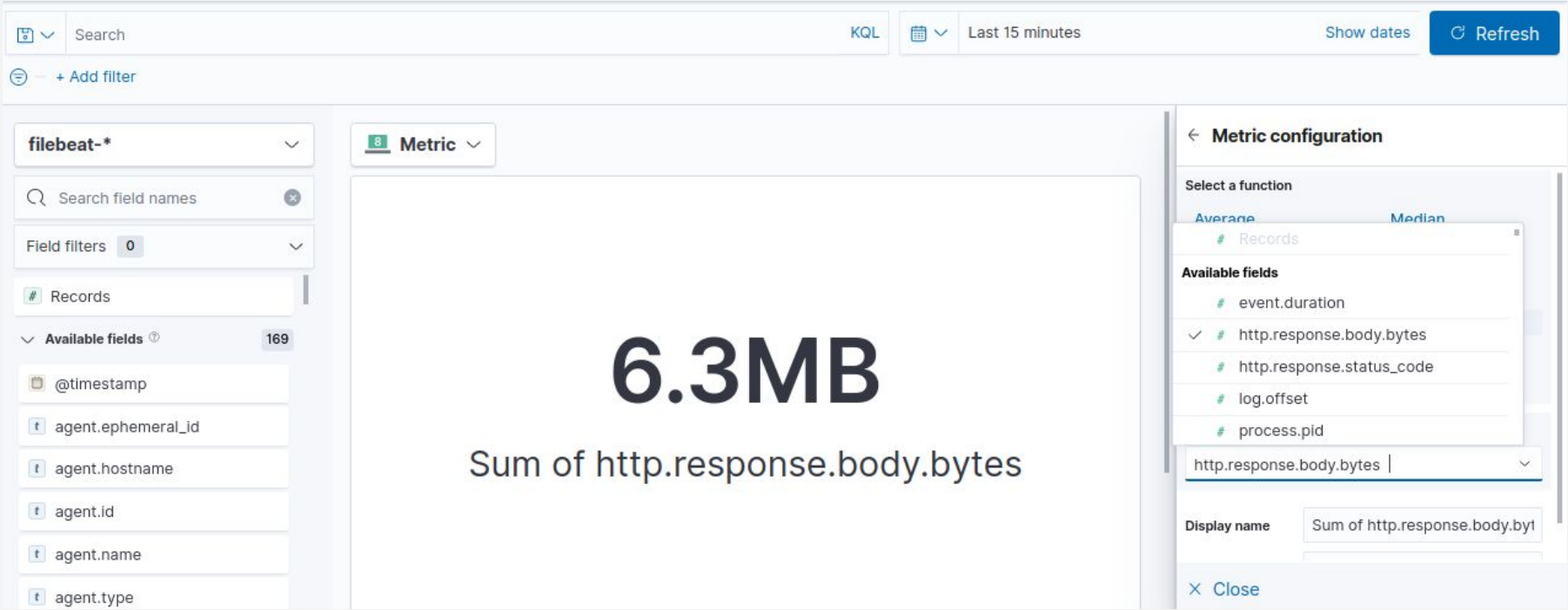
Kibana Discover (Análisis de Logs)



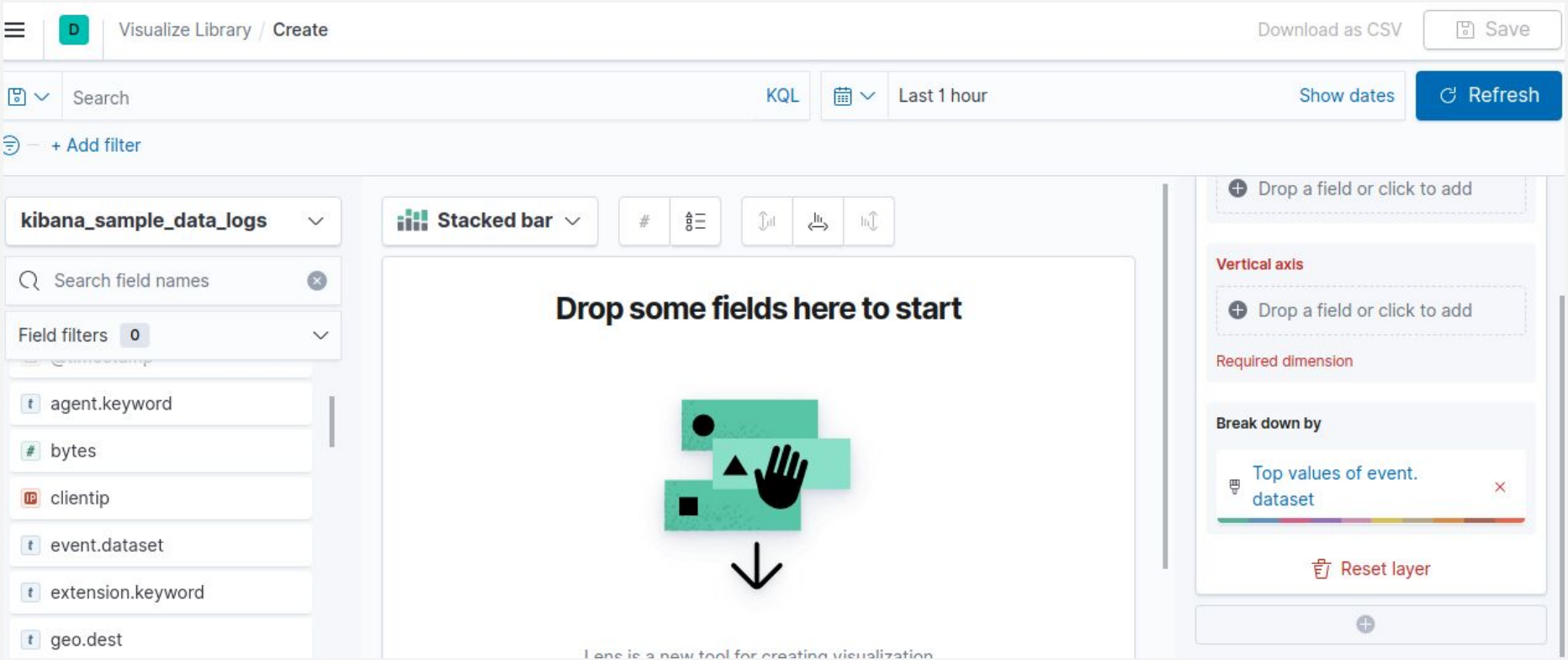
Kibana Visualize



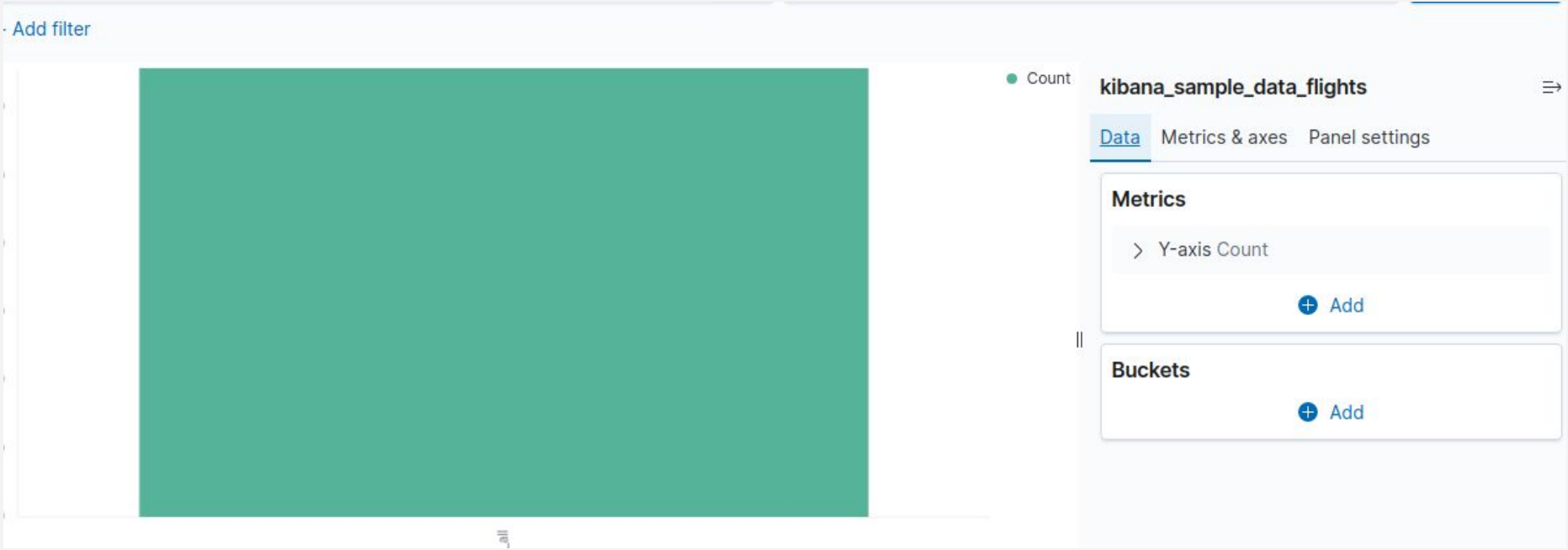
Kibana visualizaciones



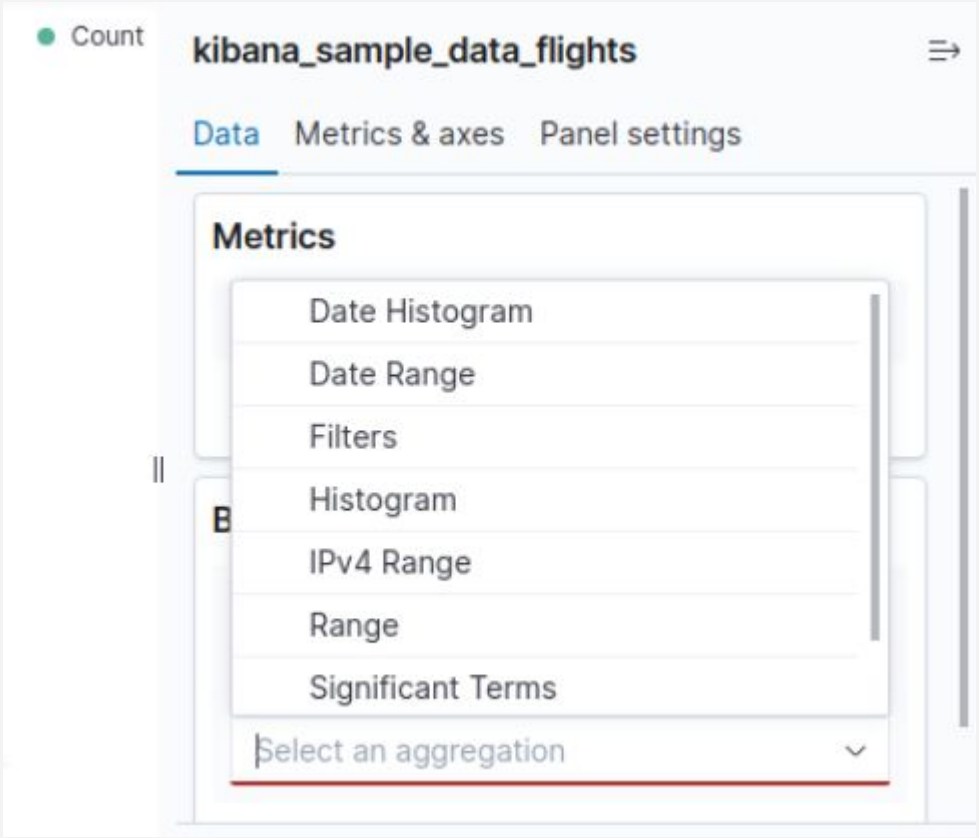
Kibana lens


Visualize Library / Create
Download as CSV
Save
Search
KQL
Last 1 hour
Show dates
Refresh
+ Add filter
kibana_sample_data_logs
Search field names
Field filters 0
agent.keyword
bytes
clientip
event.dataset
extension.keyword
geo.dest
Stacked bar
Drop some fields here to start
Drop a field or click to add
Vertical axis
Drop a field or click to add
Required dimension
Break down by
Top values of event.dataset
Reset layer

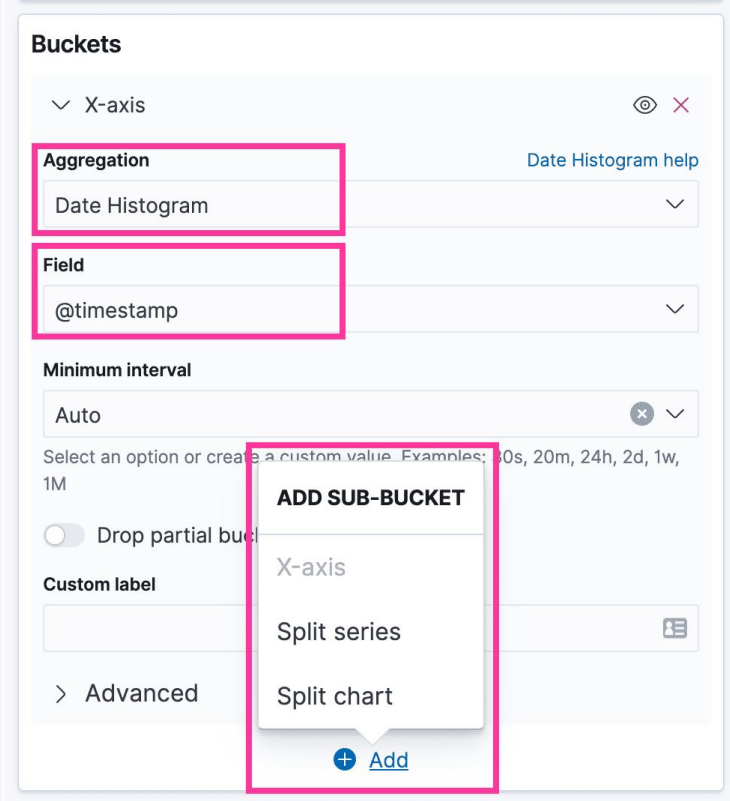
Métricas y agregaciones



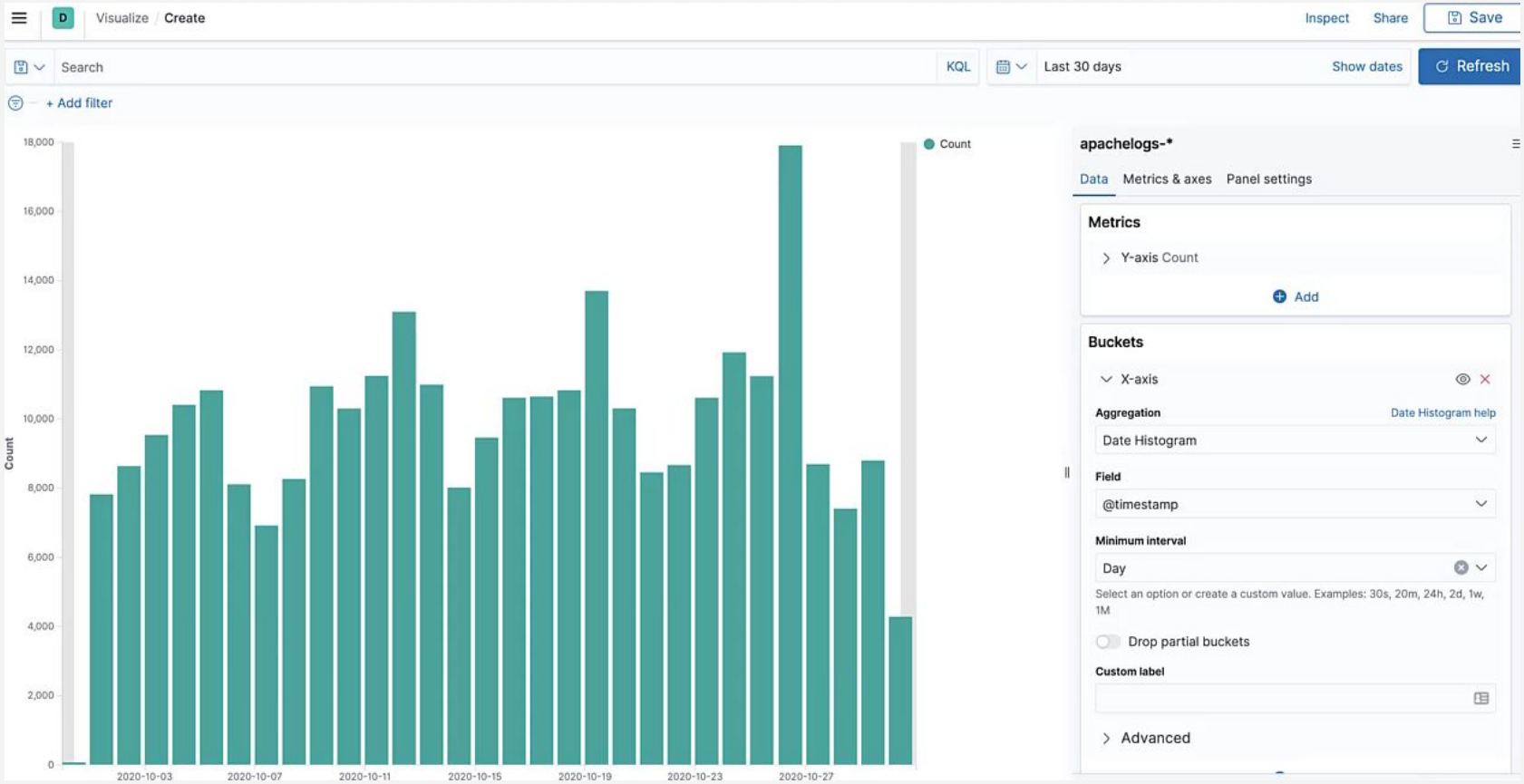
Buckets


Count
kibana_sample_data_flights
Data
Metrics & axes
Panel settings
Metrics
Date Histogram
Date Range
Filters
Histogram
IPv4 Range
Range
Significant Terms
Select an aggregation

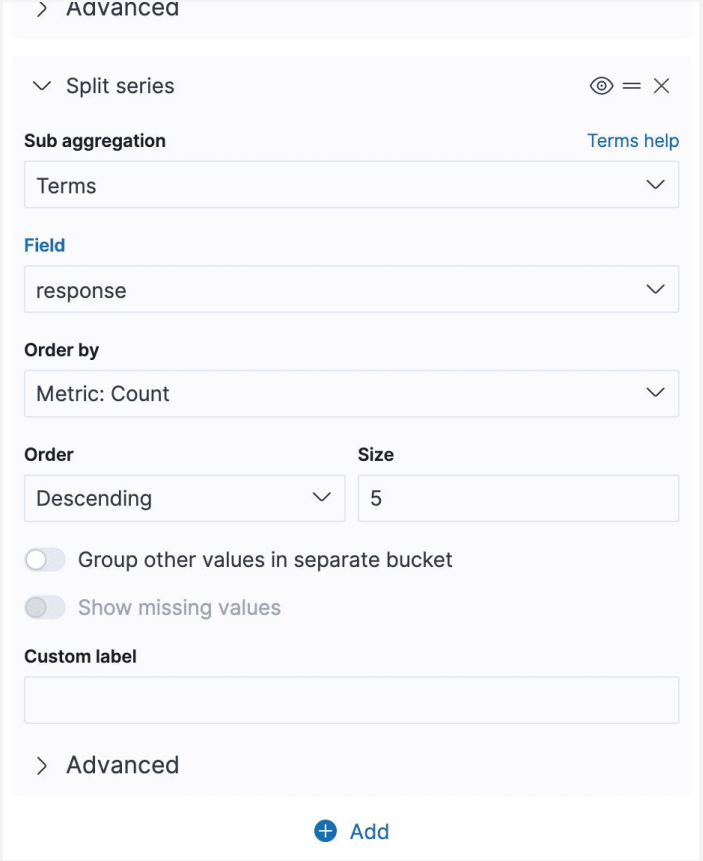
Buckets


Buckets
X-axis
Aggregation
Date Histogram help
Date Histogram
Field
@timestamp
Minimum interval
Auto
Select an option or create a custom value. Examples: 30s, 20m, 24h, 2d, 1w, 1M
Drop partial buc
Custom label
Advanced
ADD SUB-BUCKET
X-axis
Split series
Split chart
Add

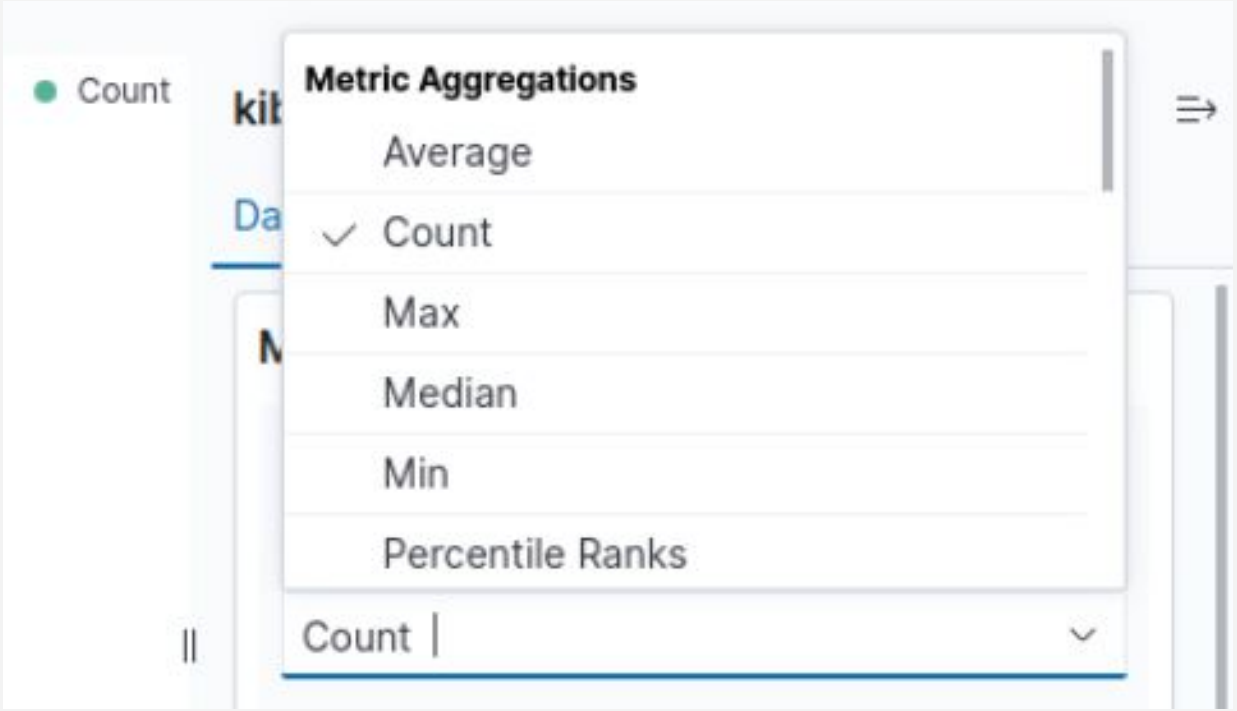
Buckets


Visualize / Create
Inspect
Share
Save
Search
KQL
Last 30 days
Show dates
Refresh
+ Add filter
Count
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
2020-10-03
2020-10-07
2020-10-11
2020-10-15
2020-10-19
2020-10-23
2020-10-27
apachelogs-*
Data
Metrics & axes
Panel settings
Metrics
Y-axis Count
Add
Buckets
X-axis
Aggregation
Date Histogram help
Date Histogram
Field
@timestamp
Minimum interval
Day
Select an option or create a custom value. Examples: 30s, 20m, 24h, 2d, 1w, 1M
Drop partial buckets
Custom label
Advanced

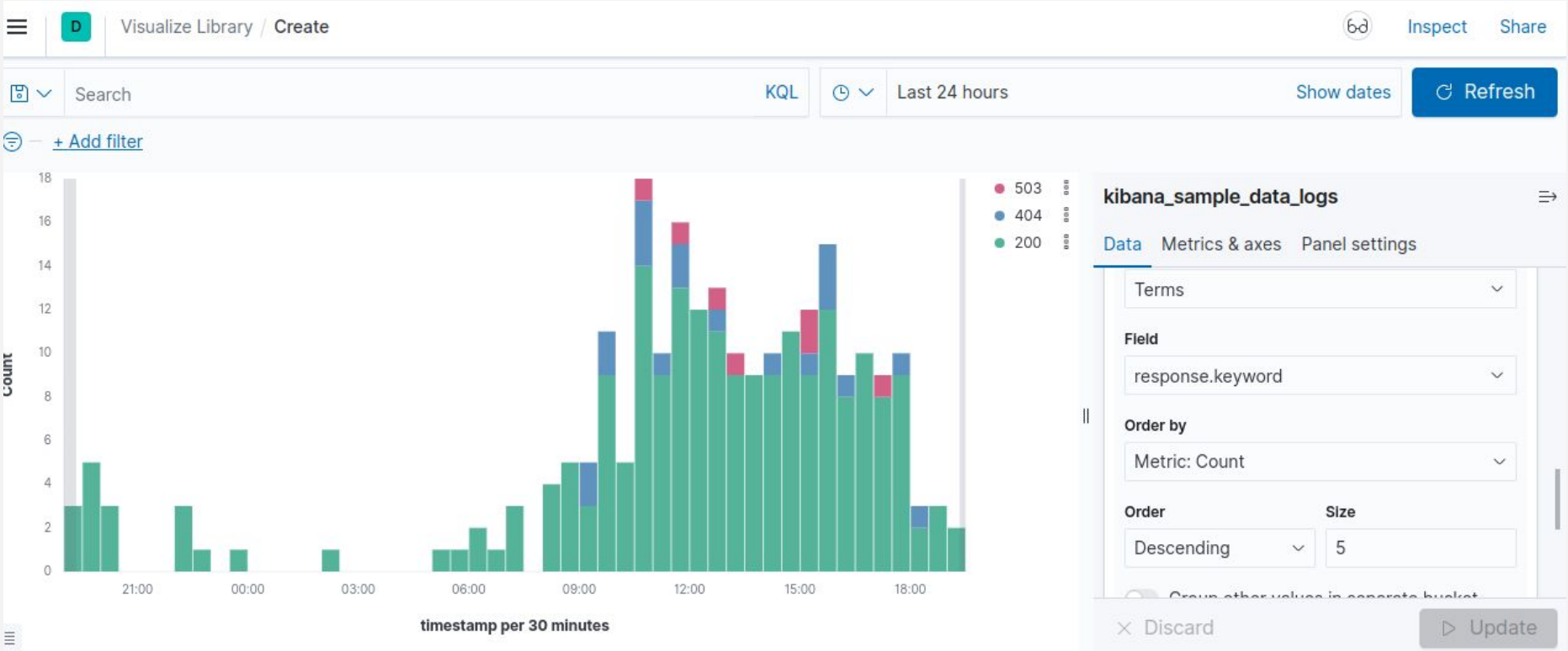
Buckets


Advanced
Split series
Sub aggregation
Terms help
Terms
Field
response
Order by
Metric: Count
Order
Size
Descending
5
Group other values in separate bucket
Show missing values
Custom label
Advanced
Add

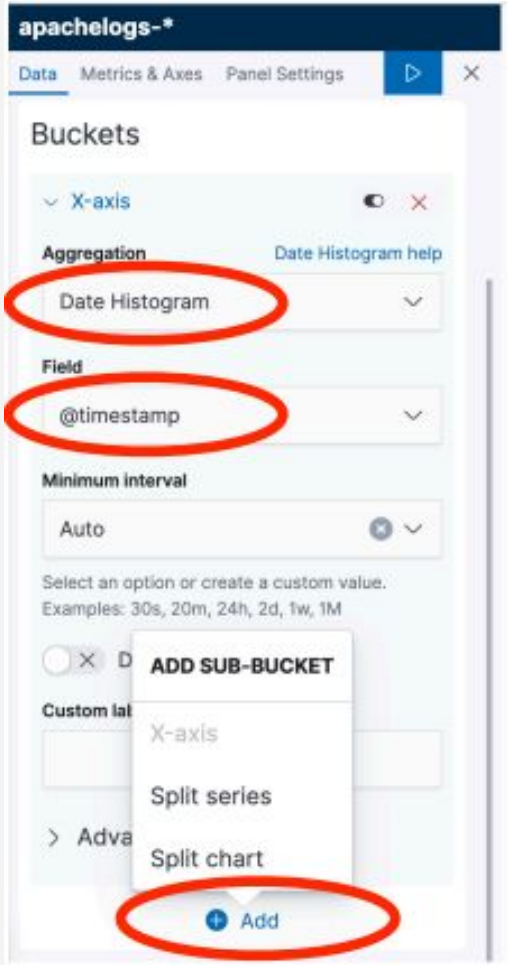
Métricas

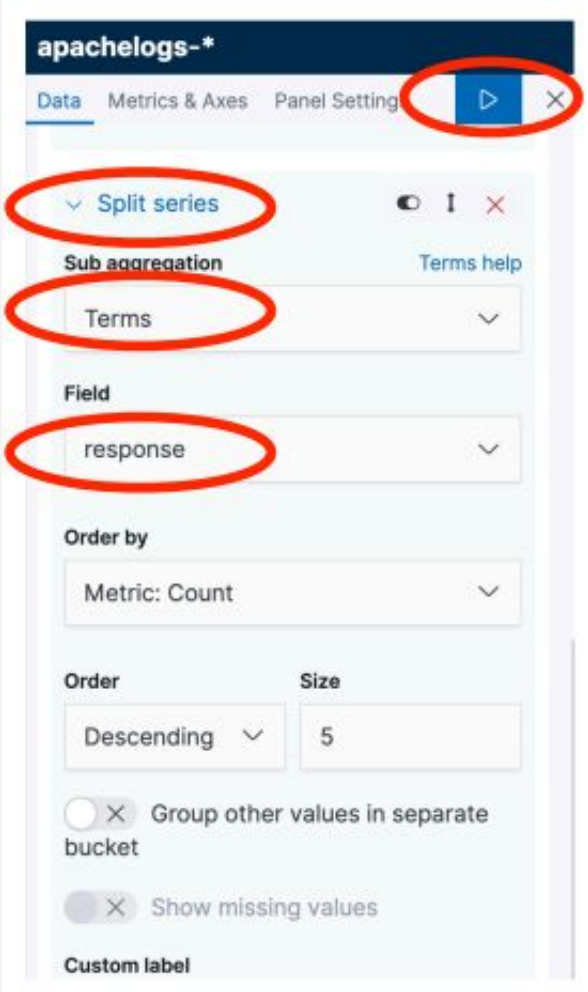

Count
Metric Aggregations
Average
Count
Max
Median
Min
Percentile Ranks
Count

Buckets

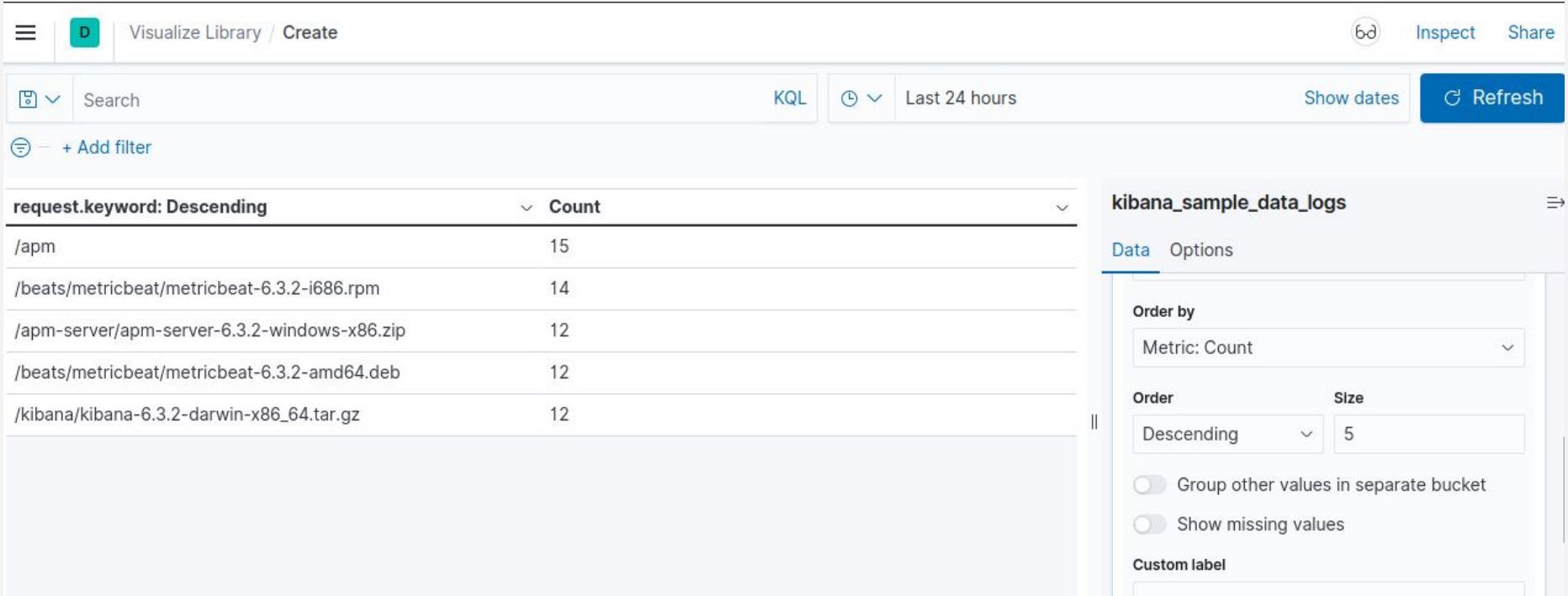

Visualize Library / Create
Inspect
Share
Search
KQL
Last 24 hours
Show dates
Refresh
+ Add filter
503
404
200
Count
timestamp per 30 minutes
kibana_sample_data_logs
Data
Metrics & axes
Panel settings
Terms
Field
response.keyword
Order by
Metric: Count
Order
Descending
Size
5
Discard
Update

Buckets

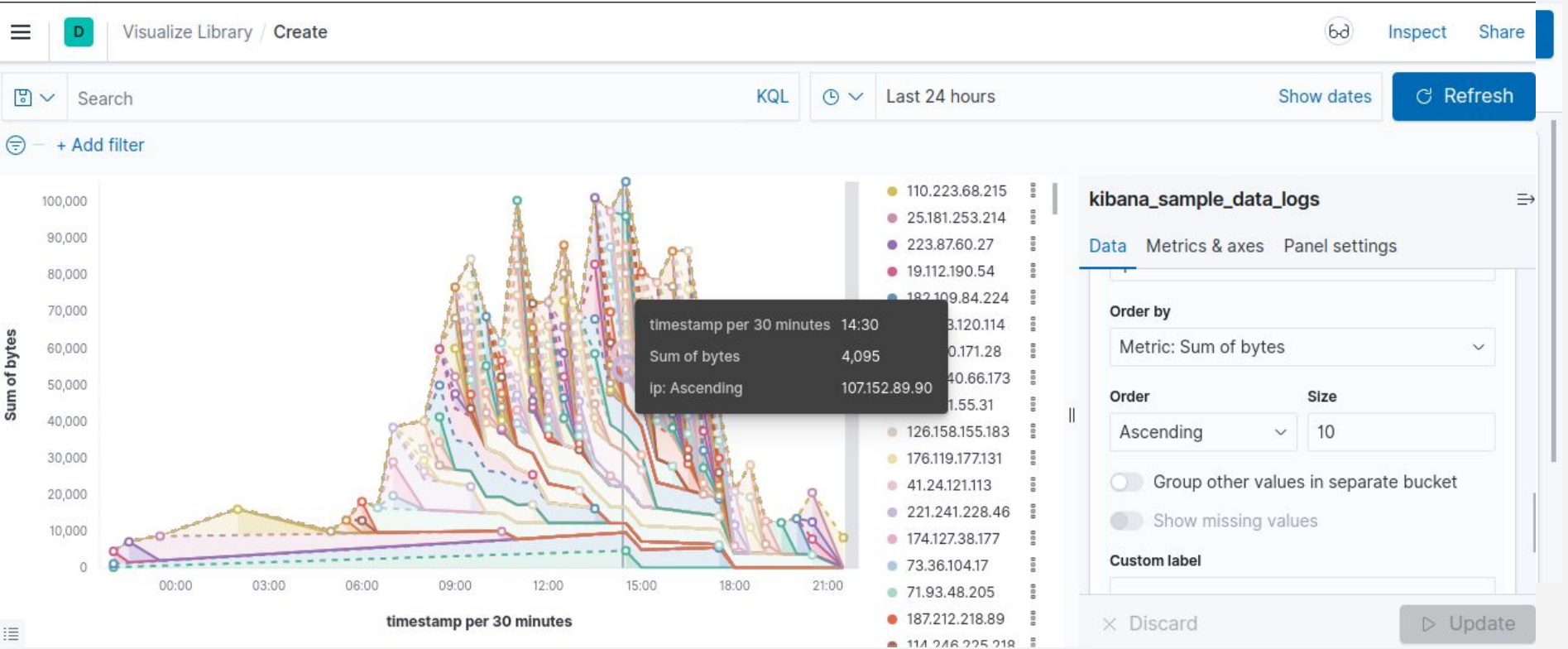

apachelogs-*
Data
Metrics & Axes
Panel Settings
Buckets
X-axis
Aggregation
Date Histogram help
Date Histogram
Field
@timestamp
Minimum interval
Auto
Select an option or create a custom value.
Examples: 30s, 20m, 24h, 2d, 1w, 1M
ADD SUB-BUCKET
X-axis
Split series
Split chart
Add


apachelogs-*
Data
Metrics & Axes
Panel Setting
Split series
Sub aggregation
Terms help
Terms
Field
response
Order by
Metric: Count
Order
Size
Descending
5
Group other values in separate bucket
Show missing values
Custom label

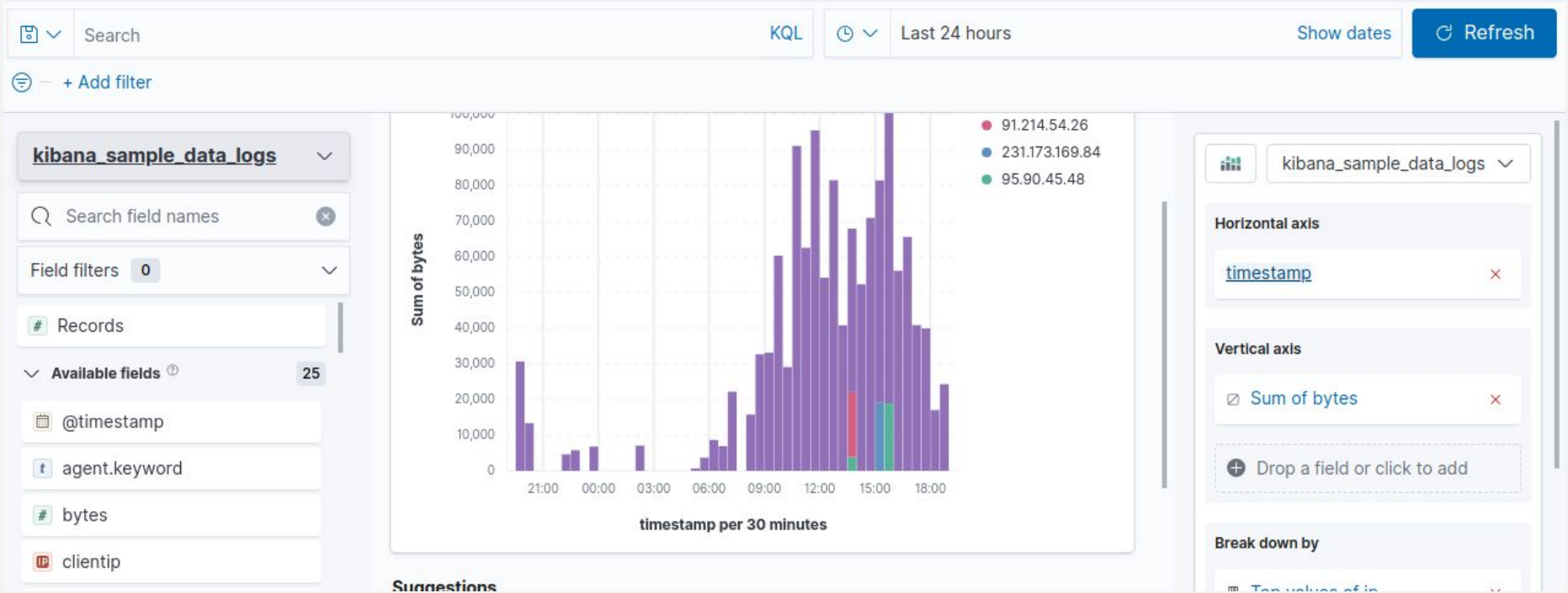
Buckets


D
Visualize Library / Create
Inspect
Share
Search
KQL
Last 24 hours
Show dates
Refresh
+ Add filter
request.keyword: Descending
Count
/apm
15
/beats/metricbeat/metricbeat-6.3.2-i686.rpm
14
/apm-server/apm-server-6.3.2-windows-x86.zip
12
/beats/metricbeat/metricbeat-6.3.2-amd64.deb
12
/kibana/kibana-6.3.2-darwin-x86_64.tar.gz
12
kibana_sample_data_logs
Data
Options
Order by
Metric: Count
Order
Size
Descending
5
Group other values in separate bucket
Show missing values
Custom label

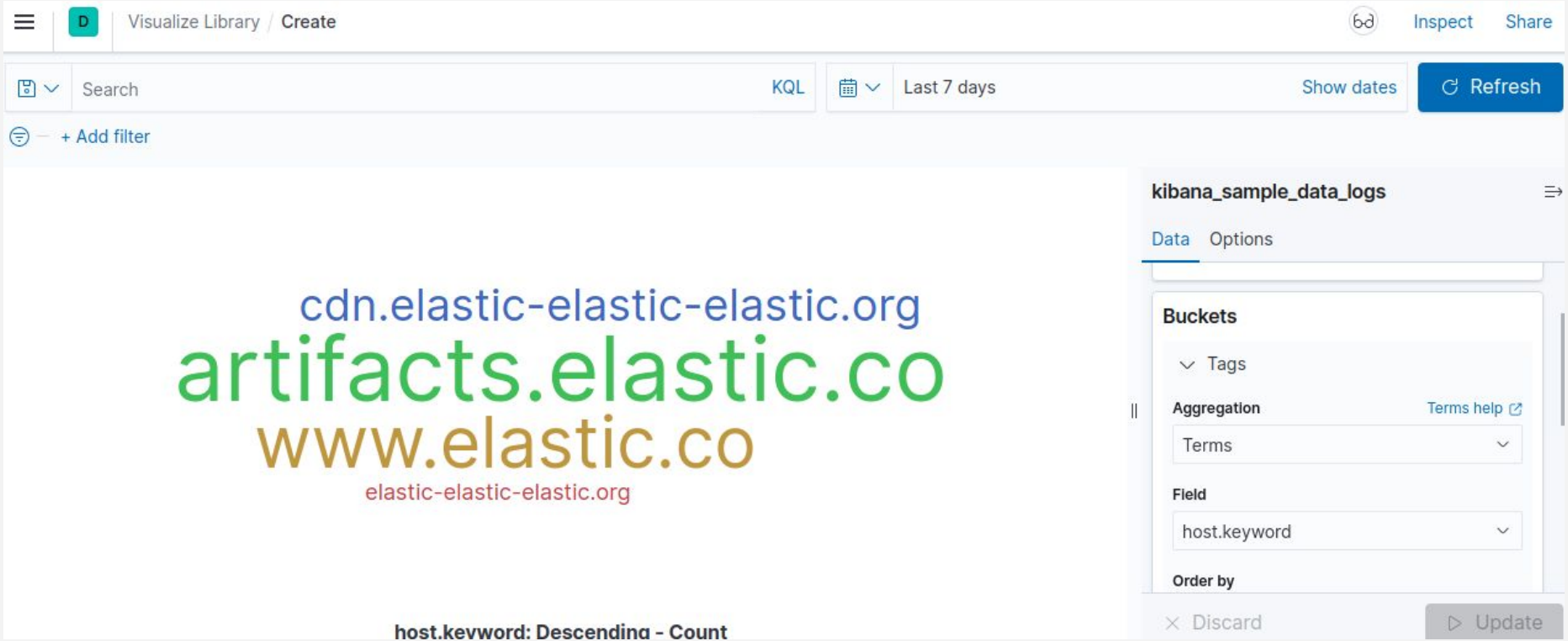
Buckets


Visualize Library / Create
Inspect
Share
Search
KQL
Last 24 hours
Show dates
Refresh
+ Add filter
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
Sum of bytes
00:00
03:00
06:00
09:00
12:00
15:00
18:00
21:00
timestamp per 30 minutes
timestamp per 30 minutes 14:30
Sum of bytes 4,095
ip: Ascending 107.152.89.90
110.223.68.215
25.181.253.214
223.87.60.27
19.112.190.54
126.158.155.183
176.119.177.131
41.24.121.113
221.241.228.46
174.127.38.177
73.36.104.17
71.93.48.205
187.212.218.89
kibana_sample_data_logs
Data
Metrics & axes
Panel settings
Order by
Metric: Sum of bytes
Order
Ascending
Size
10
Group other values in separate bucket
Show missing values
Custom label
Discard
Update

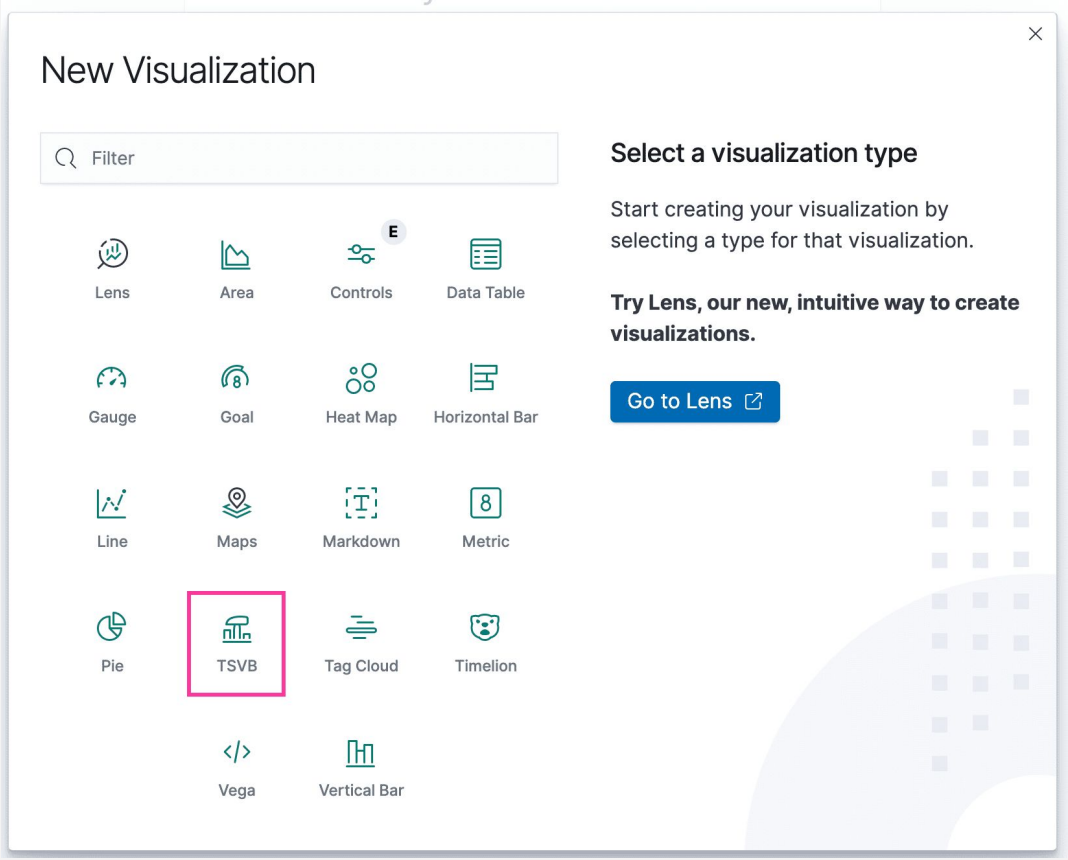
Buckets


Search
KQL
Last 24 hours
Show dates
Refresh
+ Add filter
kibana_sample_data_logs
Search field names
Field filters 0
Records
Available fields 25
@timestamp
agent.keyword
bytes
clientip
91.214.54.26
231.173.169.84
95.90.45.48
Sum of bytes
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
21:00
00:00
03:00
06:00
09:00
12:00
15:00
18:00
timestamp per 30 minutes
kibana_sample_data_logs
Horizontal axis
timestamp
Vertical axis
Sum of bytes
Drop a field or click to add
Break down by
Suggestions

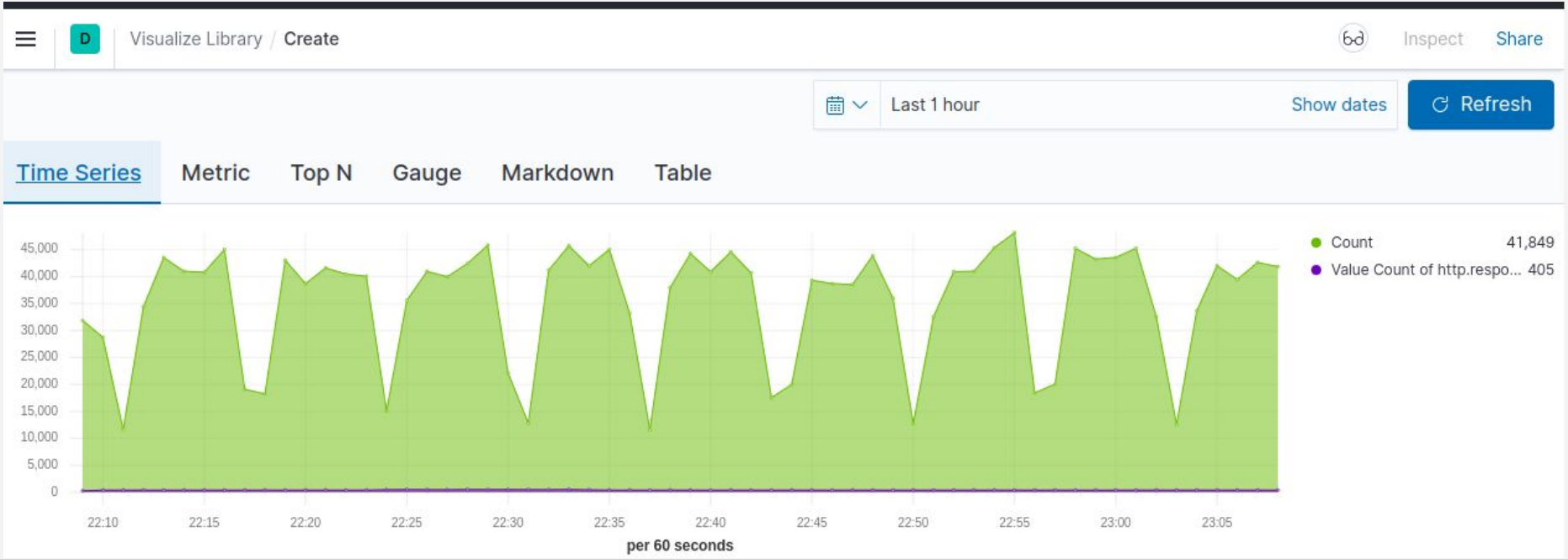
Buckets


Visualize Library / Create
Inspect
Share
Search
KQL
Last 7 days
Show dates
Refresh
+ Add filter
cdn.elastic-elastic-elastic.org
artifacts.elastic.co
www.elastic.co
elastic-elastic-elastic.org
host.keyword: Descending - Count
kibana_sample_data_logs
Data
Options
Buckets
Tags
Aggregation
Terms help
Terms
Field
host.keyword
Order by
Discard
Update

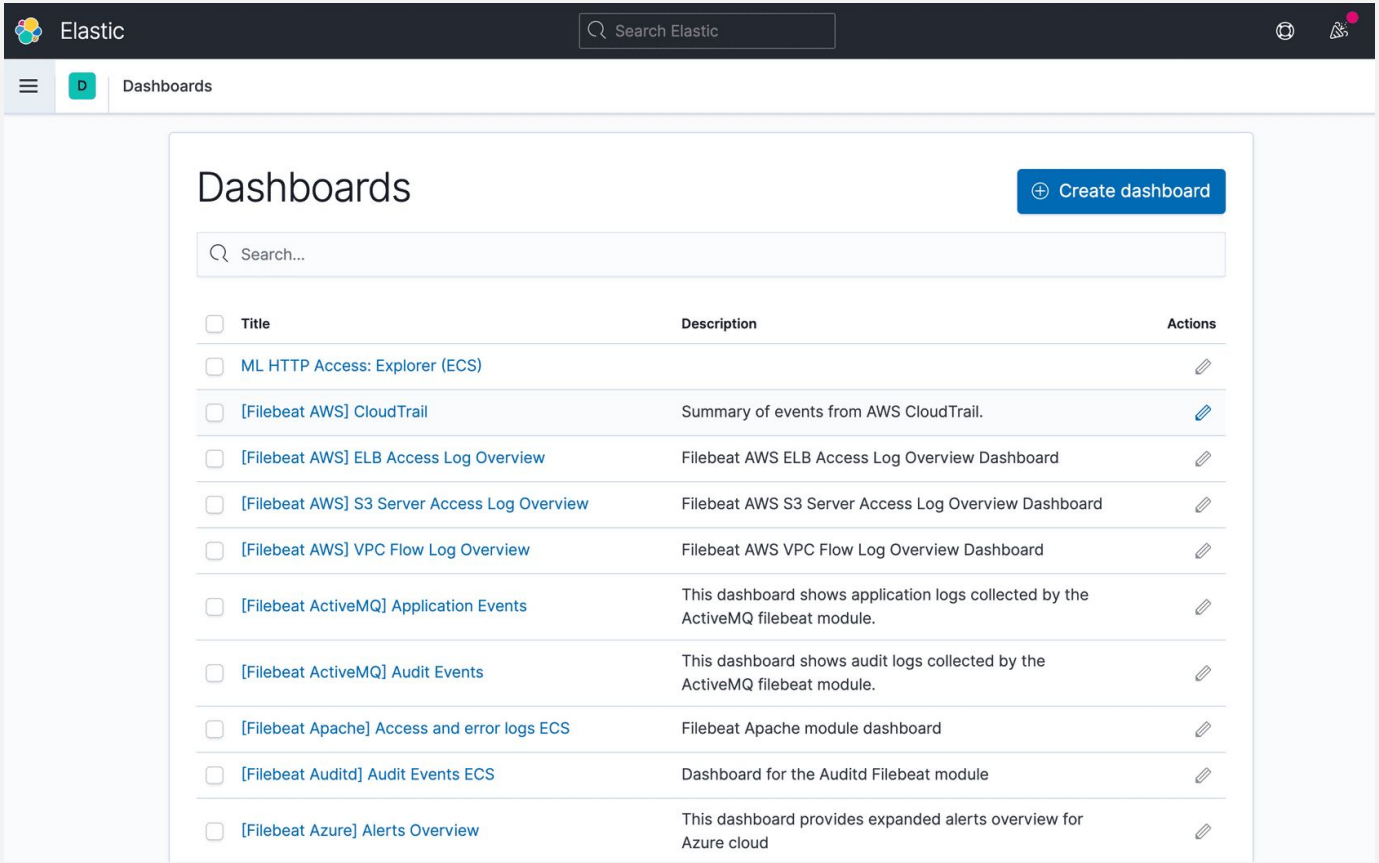
Kibana time series



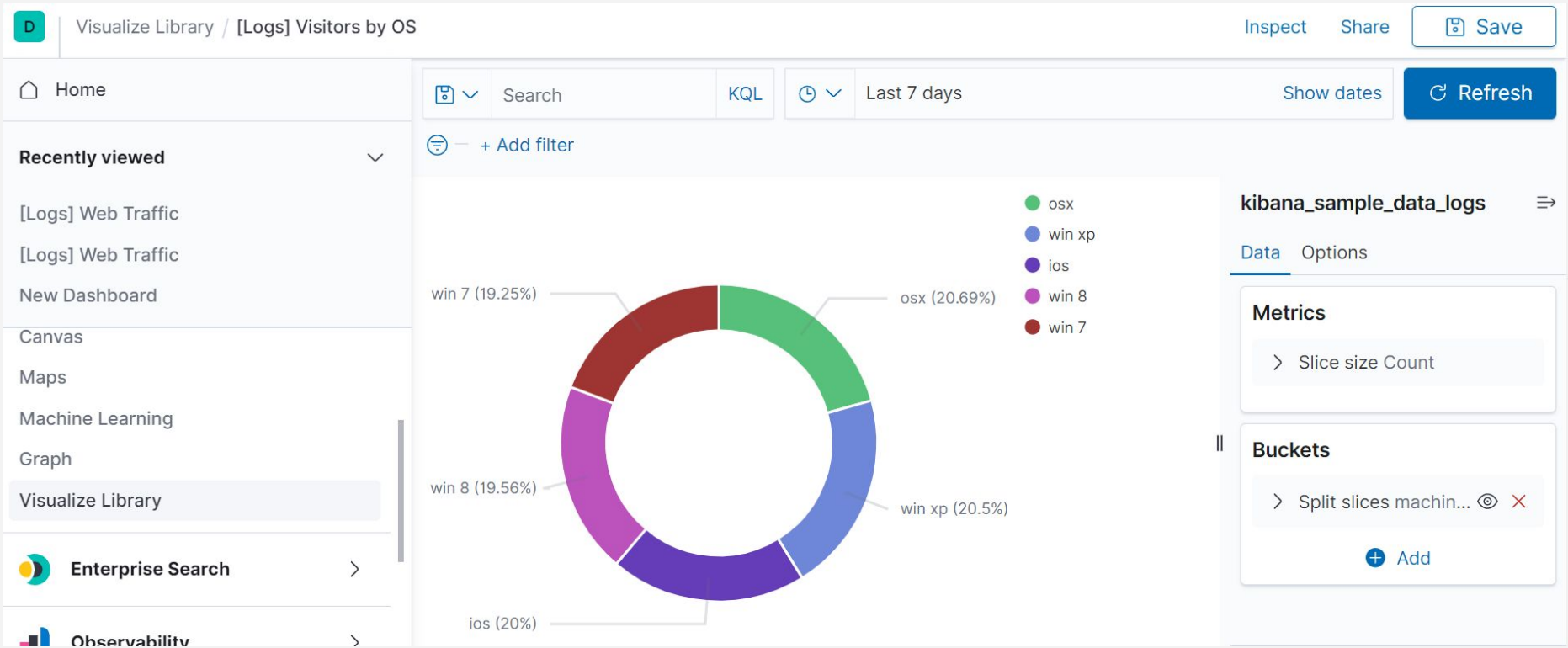
Kibana time series



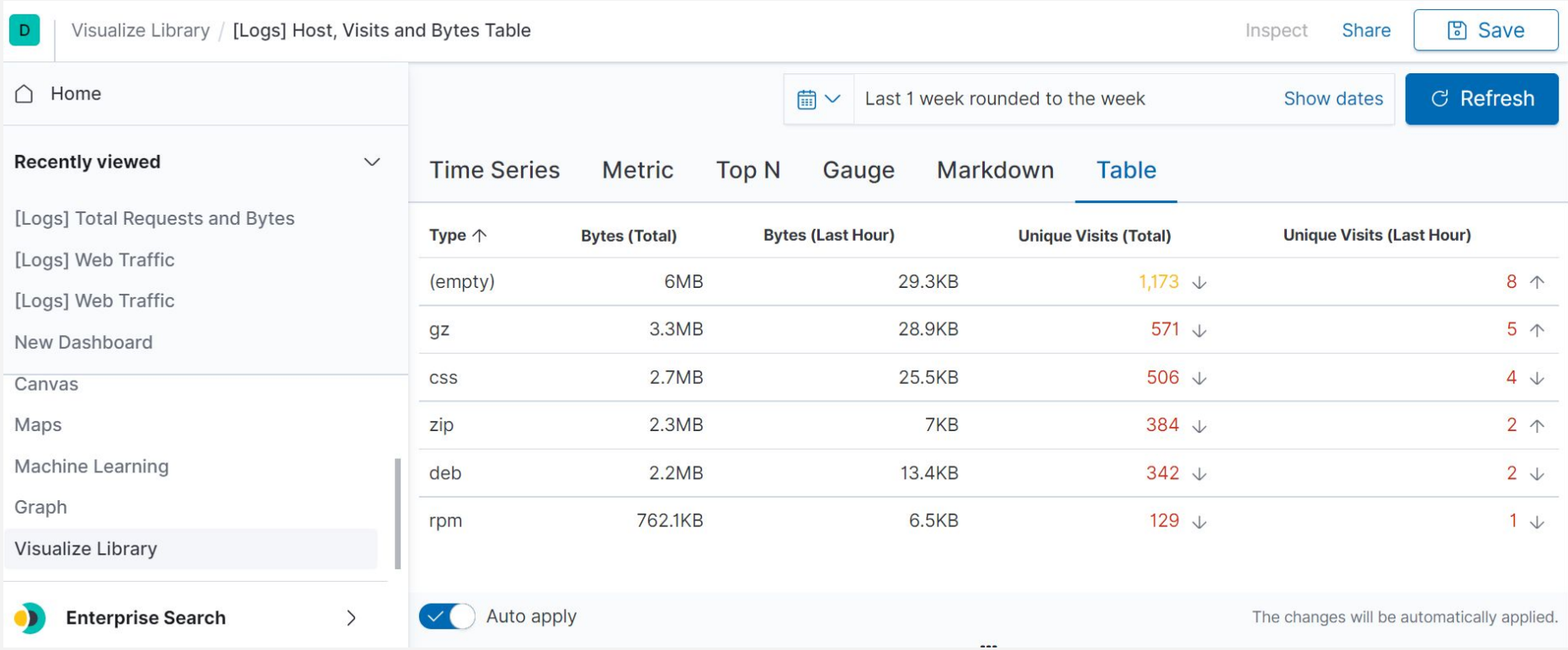
Kibana Dashboards



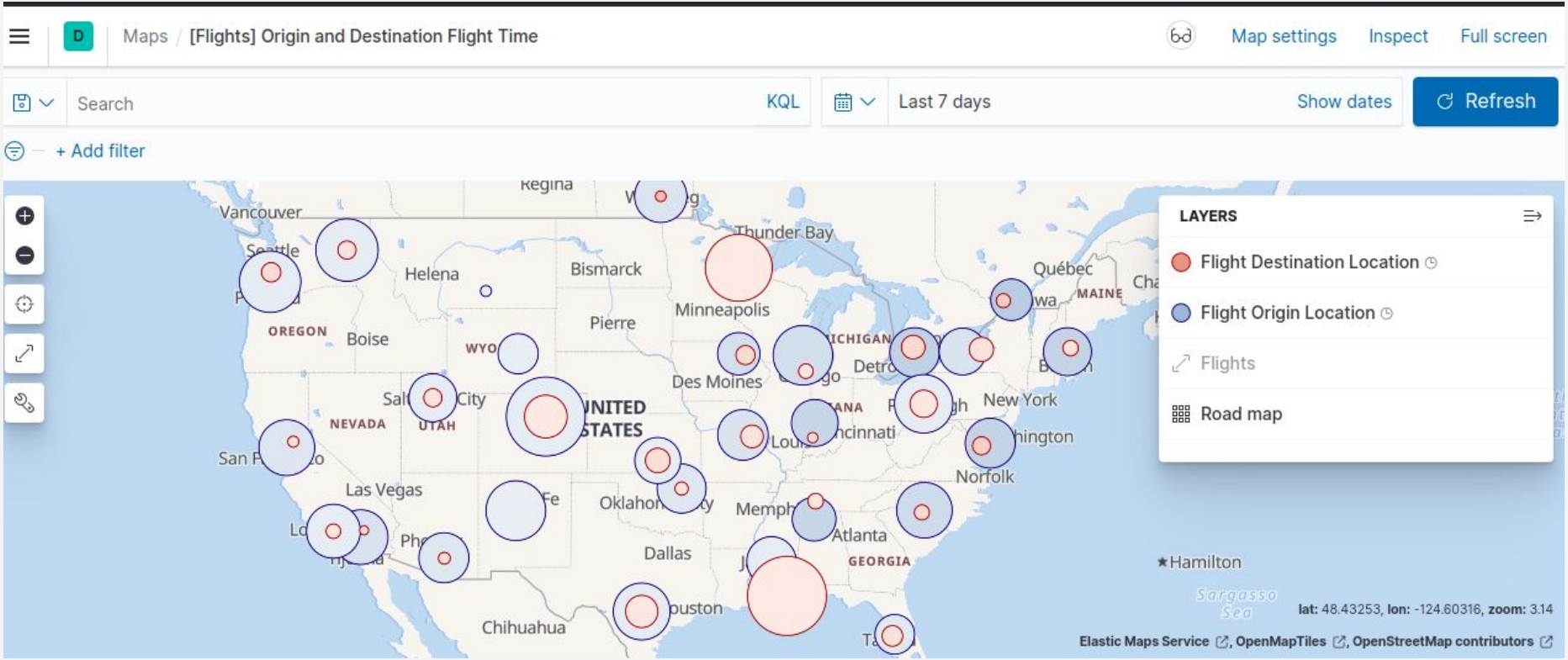
Visualizaciones: Pie



Visualizaciones: Tablas de datos

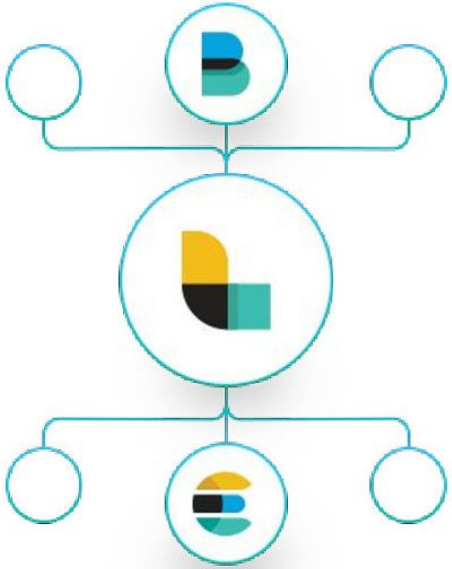


Visualizaciones: maps



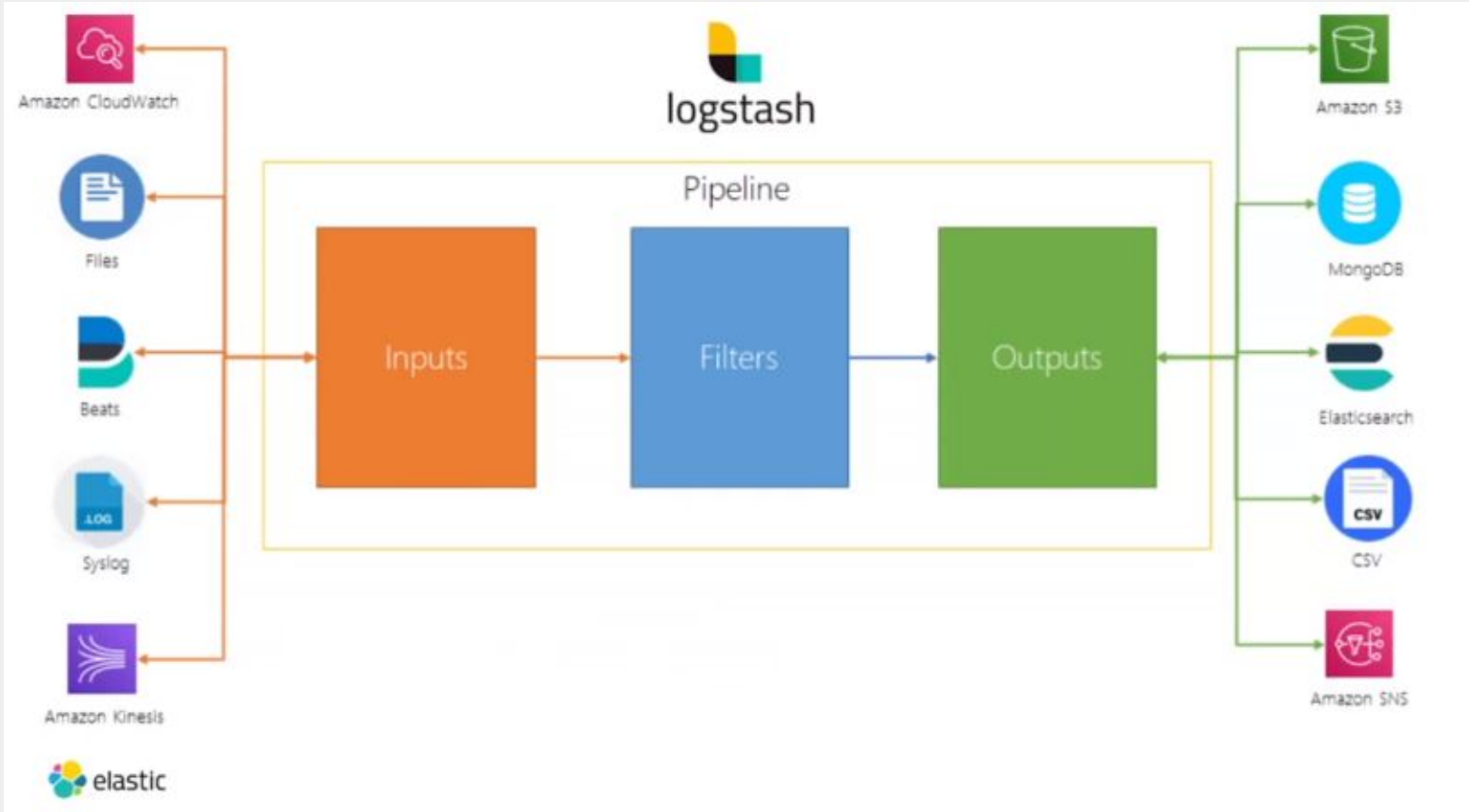
Logstash

ETL para Elasticsearch

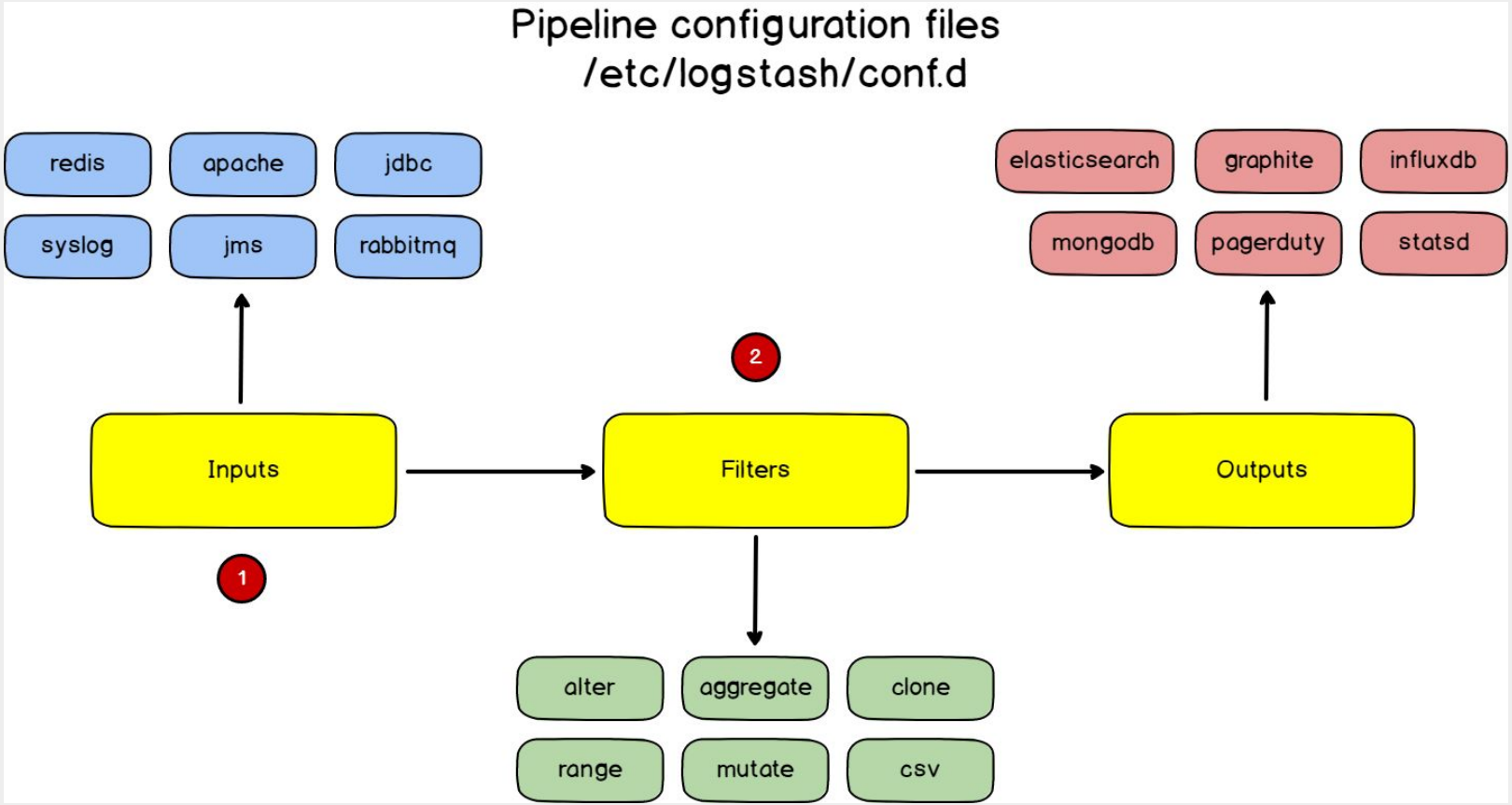


Ingestar datos de distintos tipos y fuentes	Parsear y transformar dinámicamente datos	Transportar datos a varias salidas
Securizar y encriptar entradas de datos	Construye tus propias pipelines	Gran variedad de plugins

Diseño de Logstash



Arquitectura de Logstash



Instalación de Logstash

sudo apt-get install logstash

sudo systemctl start logstash

```
][logstash.outputs.elasticsearch][main] ES Output version determined
][logstash.outputs.elasticsearch][main] Detected a 6.x and above clus
][logstash.outputs.elasticsearch][main] New Elasticsearch output {:cl
][logstash.javapipeline    ][main] Starting pipeline {:pipeline_id=>"
][logstash.javapipeline    ][main] Pipeline Java execution initializa
][logstash.inputs.beats    ][main] Beats inputs: Starting input liste
][logstash.javapipeline    ][main] Pipeline started {"pipeline.id"=>"
][logstash.agent           ] Pipelines running {:count=>1, :running_p
][org.logstash.beats.Server][main][0e3b303d763998eaa8de60184a2986df70
][logstash.agent           ] Successfully started Logstash API endpoi
```

```
bin/logstash -e 'input { stdin { } } output { stdout {} }'
```

```
bin/logstash --verbose -f sample.conf
```

bin/logstash -e 'input { stdin { } } output { stdout { codec => rubydebug} }'

```
{
"message" => " Hello Logstash",
"@timestamp" => "2021-01-01T23:48:05.335Z",
"@version" => "1",
"host" => "127.0.0.1"
}
```

Input Stdin (stdin_simple.conf)

```
input {
    stdin { }
}
output {
    stdout {
        codec => rubydebug
    }
}
```

Input Stdin

```
{
"message" => " Hello logstash",
"@timestamp" => "2021-09-20T23:48:05.335Z",
"@version" => "1",
"host" => "host"
}
```

Input Stdin (stdin_json.conf)

```
input {
  stdin{}
}

output {
  stdout{
     codec=>json_lines
  }
}
```

Input file

```
input {
 file {
    path => ["/home/logstash/testdata.log"]
    sincedb_path => "/dev/null"
    start_position => "beginning"
   }
}
```

Input file (file_json.conf)

```
input {
    file {
        path => "/home/linux/Descargas/elk/logstash/test.json"
        start_position => "beginning"
        codec => "json"
    }
}

filter {

}

output {
    stdout { codec => rubydebug }
}
```

Input file (test_logs.conf)

```
input {
  file {
    path => "/home/linux/Descargas/elk/logstash/logs_apm.log"
    type => "logs"
    sincedb_path => "/dev/null"
    start_position => "beginning"
  }
}

filter{}

output {
    stdout{
        codec => rubydebug
    }

}
```

Input beats (logstash-beats-sample.conf)

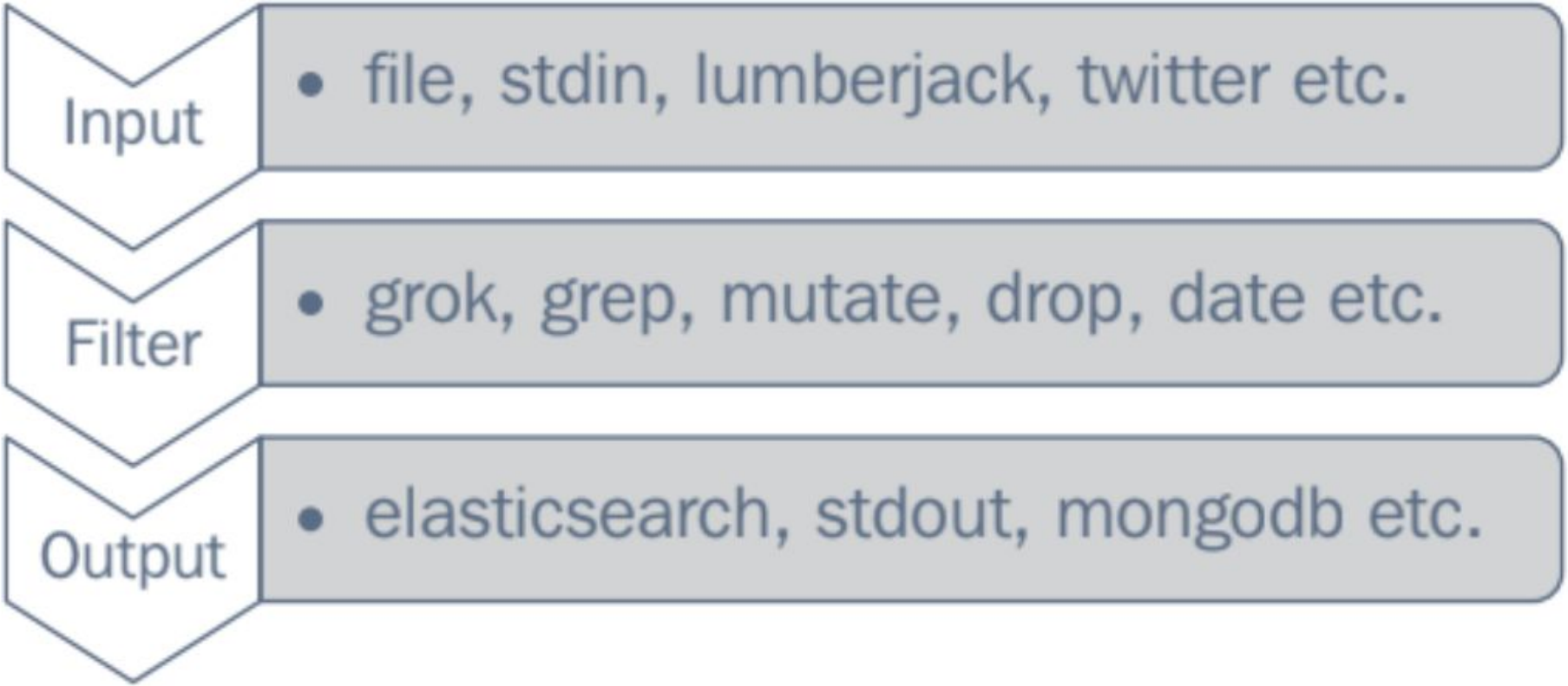
```
input {
  beats {
    port => 5044
  }
}

output {
  elasticsearch {
    hosts => ["http://localhost:9200"]
    index =>
"%{[@metadata][beat]}-%{[@metadata][version]}-%{+YYYY.MM.dd}
"
  }
}
```

Logstash plugins

coralogix-resources Update twitter.conf	`4bc316c` on 17 Jul	**63** commits
dead-letter-queue	Update dlq.conf	3 months ago
exec	exec modifications	4 months ago
generator	adding a few more configs	4 months ago
heartbeat	updating index heartbeat epoch	4 months ago
http-input	external request	3 months ago
http-poller	test	3 months ago
twitter	Update twitter.conf	3 months ago
unix	edit conf socket	4 months ago
websockets	restructuring files and folders	4 months ago

Twitter plugin (twitter-input.conf)

```
input{
    twitter{
        consumer_key => "90Bg8UeX8uLOLnx3leGS4Z3cB"
        consumer_secret => "Pzkh169mTzPXKX0FKypxLhSQGFhLerPMTA0596mYtCeCMupEXV"
        oauth_token => "201832916-39XbHLtdlinPZPVq4lkbuwgjVs0HyoHgAnqLE1QP"
        oauth_token_secret => "Q6P2clvCpKhjklKnpKeXc95eC9B5SnmgY1bLRy9lqgcQY"
        keywords=>["ELK","ElasticSearch","Logstash","Kibana"]
    }
}
output{
    file{
        path=>"tweets.txt"
    }
    stdout{
        codec => rubydebug
    }
}
```


Input
• file, stdin, lumberjack, twitter etc.
Filter
• grok, grep, mutate, drop, date etc.
Output
• elasticsearch, stdout, mongodb etc.

LOGSTASH FILTERS

- Filter: Grok. Procesamiento básico de los logs
- Filter: Grok. Procesamiento avanzado de logs
- Filter Grok. Usos avanzados
- Filter: Mutate
- Filter: Date
- Filter: Translate
- Filter: GeoIp
- Filter: Ruby

Filter: Grok. Procesamiento básico de los logs

% {PATTERN: FIELDNAME}

%{NUMBER:bytes_transferred}

Filter: Grok. Procesamiento básico de los logs

54.3.245.1 GET /index.html 14562 0.056

%{IP:client_ip} %{WORD: request_method }
%{URIPATHPARAM:uri_path}
%{NUMBER:bytes_transferred}
%{NUMBER:duration}

Filter: Grok. Procesamiento básico de los logs

```
55.3.244.1 GET /index.html 15824 0.043
```

```
input {
  file {
    path => "/var/log/http.log"
  }
}
filter {
  grok {
    match => { "message" => "%{IP:client} %{WORD:method} %
{URIPATHPARAM:request} %{NUMBER:bytes} %{NUMBER:duration}" }
  }
}
```

Filter: Grok. Procesamiento básico de los logs

```
filter{
grok{
match => { "message"
=>"%{IP:client_ip} %{WORD:request_
method} %{URIPATHPARAM:uri_path}
%{NUMBER:bytes_transferred}
%{NUMBER:duration}"}
}
}
```

- client_ip : 54.3.245.1
- request_method : GET
- uri_path :/index.html
- bytes_transferred :14562
- duration :0.056

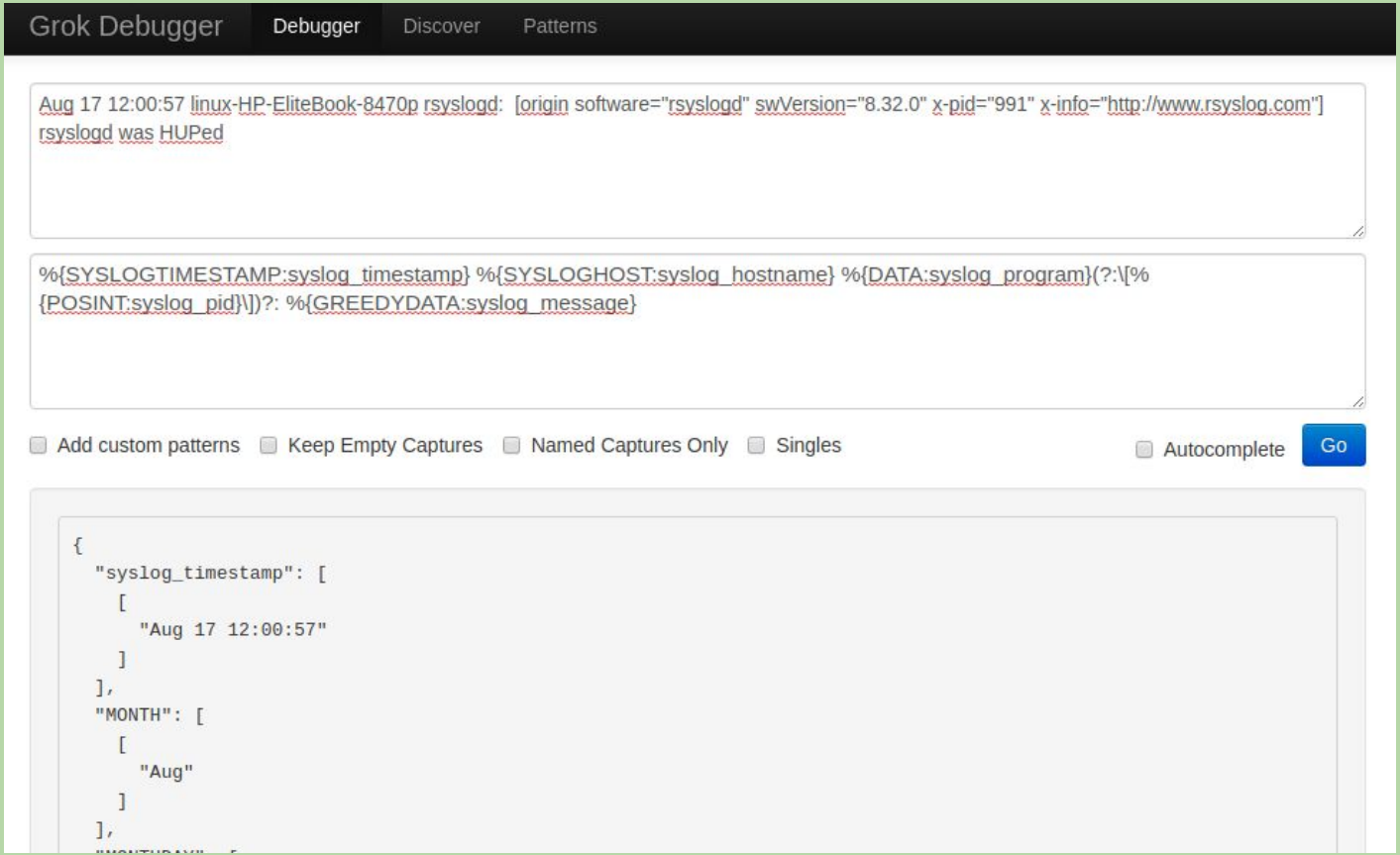
Filter: Grok. Procesamiento avanzado de logs

```
filter {
  if [type] == "syslog" {
    grok {
      match => { "message" => "%{SYSLOGTIMESTAMP:syslog_timestamp}
%{SYSLOGHOST:syslog_hostname}
%{DATA:syslog_program}(?:\[%{POSINT:syslog_pid}\])?:
%{GREEDYDATA:syslog_message}" }
      add_field => [ "received_at", "%{@timestamp}" ]
      add_field => [ "received_from", "%{host}" ]
    }
    syslog_pri { }
    date {
      match => [ "syslog_timestamp", "MMM  d HH:mm:ss", "MMM dd HH:mm:ss" ]
    }
  }
}
```

Filter: Grok. Procesamiento avanzado de los logs

- https://github.com/logstash-plugins/logstash-patterns-core/tree/master/patterns
- http://grokdebug.herokuapp.com/
- http://grokconstructor.appspot.com/

Filter: Grok. Procesamiento avanzado de los logs



Filter Mutate

```
filter {
mutate {
convert => # hash of field and data type (optional)
join => # hash of fields to be joined (optional)
lowercase => # array of fields to be converted (optional)
merge => # hash of fields to be merged (optional)
rename => # hash of original and rename field (optional)
replace => # hash of fields to replaced with (optional)
split => # hash of fields to be split (optional)
strip => # array of fields (optional)
uppercase => # array of fields (optional)
}
}
```

Filter Mutate (logstash_mutate.conf)

```
input {
    stdin { }
}
filter {
    mutate {
        uppercase => [ "message" ]
    }
}
output {
    stdout {
        codec => rubydebug
    }
}
```

Filter Mutate (file_json_filter.conf)

```
filter {
  mutate {
    remove_field => [ "@version" ]
    add_field => { "tipoUsuario" => "cliente" }
    gsub => ["surname", " - ", ""]
  }
}
```

Filter Date

```
date{
match => ["date_field", "yyyy-MM-dd"]
target => "@timestamp"
}
```

```
match => ["date_field", "MMM dd YYY HH: mm: ss", "ISO8601", "MMddYYYY", "MMM d AAAA HH: mm: ss"]
```

Filter translate (logstash_translate.conf)

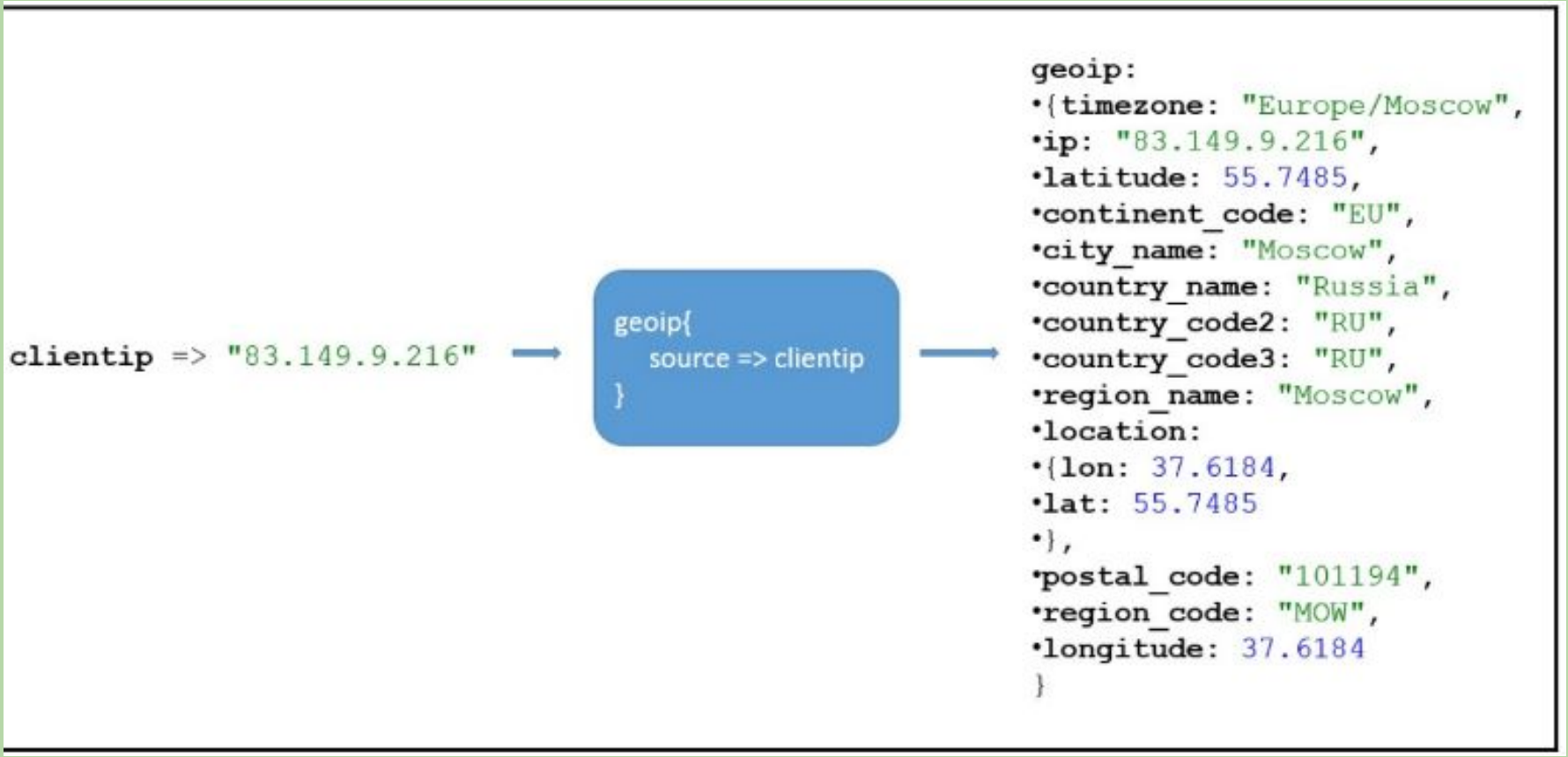
```
input {
  # The generator creates an input event
  generator {
    lines => [
     '{"my_msg": "testing1234", "lookup_id": "1234"}'
    ]
    count => 1
    codec =>  "json"
  }
}

filter {
  # Enrich the event using the lookup_id
  translate {
    field => "[lookup_id]"
    destination => "[enrichment_data]"
    fallback => "not_found"
    dictionary => {
      "1234" => "1234 found in the dictionary"
      "5678" => "5678 found  in the dictionary"
    }
  }
}

output {
  stdout { codec =>  "rubydebug" }
}
```

Filter geoip (logstash_geoip.conf)

```
$ echo "logstash 19.1.193.230 $(date --iso-8601=seconds)" >> ~/input.txt
```

```
input {
  file {
    path => "/home/linux/Descargas/elk/logstash/input_geoip.txt"
    start_position => "beginning"
    sincedb_path => "/dev/null"
  }
}
filter {
  grok {
    match => { "message" => "%{WORD:name} %{IP:ip} %{TIMESTAMP_ISO8601:date}" }
    remove_field => [ "message", "path", "@version", "host" ]
  }
  geoip {
    source => "ip"
  }
}
output {
  stdout {
        codec => rubydebug
    }
}
```

Filter geoip



Filter geoip

```
{
    "@timestamp" => 2021-09-11T19:57:45.958Z,
            "ip" => "19.1.193.230",
          "date" => "2021-09-11T21:13:17+02:00",
         "geoip" => {
             "longitude" => -97.822,
              "location" => {
            "lon" => -97.822,
            "lat" => 37.751
        },
              "timezone" => "America/Chicago",
         "country_code2" => "US",
        "continent_code" => "NA",
                    "ip" => "19.1.193.230",
         "country_code3" => "US",
          "country_name" => "United States",
              "latitude" => 37.751
    },
          "name" => "logstash"
}
```

Filter ruby (logstash_ruby_filter.conf)

```
input { # ... }
filter {
  ruby {
    code => 'size = event.get("message").size;
             event.set("message_size", size)'
  }
}
output { # ... }
```

Filter csv

```
filter{
  csv {
   columns => ["date","open_price","close_price"]
   path => "/path/to/file.csv",
   separator => ","
   }
}
```

Filter csv (read_csv.conf)

```
input {
    file {
        path => "/home/linux/Descargas/elk/logstash/file.csv"
        start_position => "beginning"
        sincedb_path => "/dev/null"
    }
}

filter {
  csv {
      separator => ","
      skip_header => "true"
      columns =>
["id","timestamp","paymentType","name","gender","ip_address","purpose","country","age"]
  }
}
output {
    stdout { codec => rubydebug }
}
```

LOGSTASH OUTPUTS

- Output: Stdout
- Output: Elasticsearch
- Otros plugins de salida
- Combinando configuraciones
- Monitorización de Logstash
- Configuración avanzada
- Múltiples pipelines
- Uso de pipelines

Output: Stdout

```
output {
  stdout {
    codec => rubydebug
    workers => 2
   }
}
```

Output: Elasticsearch

```
output {
elasticsearch {
hosts => ["localhost:9200"]
index => "logstash"
user => “elastic”
password => “elastic”
doc_as_upsert => true
}
}
```

Otros plugins de salida

```
output {
file {
        create_if_deleted => true
        file_mode => 777
        filename_failure => "failedpath_file"
        flush_interval => 0
        path => "/home/linux/Descargas/elk/logstash/file_output.txt"
    }
}
```

Monitorización de Logstash

```
GET /logstash-*/_search
{
  "query": {
    "match_all": {}
  },
  "sort": {
    "@timestamp": {
      "order": "desc"
    }
  }
}
```

Monitorización de Logstash

curl -X GET http://localhost:9600?pretty
curl -X GET http://localhost:9600/_node/?pretty
curl -X GET http://localhost:9600/_node/os?pretty
curl -X GET http://localhost:9600/_node/jvm?pretty
curl -X GET http://localhost:9600/_node/pipeline?pretty
curl -X GET http://localhost:9600/_node/plugins?pretty
curl -X GET http://localhost:9600/_node/stats?pretty
curl -X GET http://localhost:9600/_node/stats/jvm?pretty
curl -X GET http://localhost:9600/_node/stats/process?pretty
curl -X GET http://localhost:9600/_node/stats/pipeline?pretty
curl -X GET http://localhost:9600/_node/hot_threads?pretty

Configuración avanzada

```
$ bin/logstash –w 12
$ bin/logstash –-pipeline.workers 12
```

```
$ bin/logstash -b 50
$ bin/logstash --pipeline.batch.size 50
```

Configuración avanzada

```
$ bin/logstash -l PATH
$ bin/logstash --path.logs <PATH>
```

$ bin/logstash --log.level <LEVEL>

- fatal
- error
- warn
- info
- debug
- trace

Pipelines

```
{
"description" : "...",
"processors" : [ ... ]
}
```

```
http://localhost:9200/_ingest/pipeline
```