An overview of the Elastic Stack geospatial capabilities

Jorge Sanz and Thomas Neirynck
jorge.sanz@elastic.co - thomas@elastic.co
https://ela.st/foss4g-2021

2021-10-01 - FOSS4G 2021 Buenos Aires

Agenda

What is the Elastic Stack

Quick intro to the different components of the Elastic portfolio of products

Ingesting geospatial data

Different approaches to upload geodata into your cluster

Search and aggregate

Core processes to analyze geospatial data

Visualize

Render geospatial data at scale

elastic

The Elastic Stack

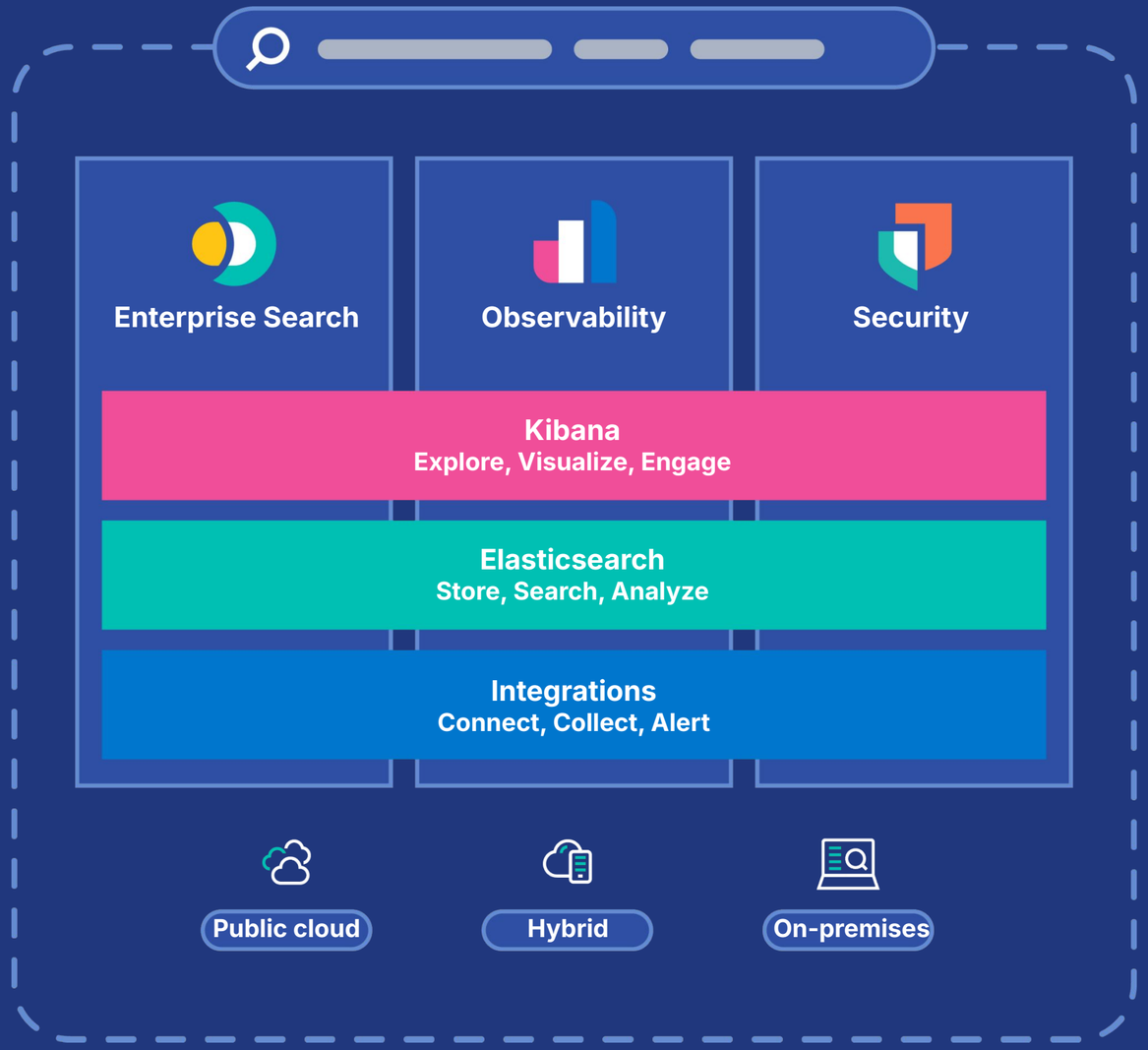
elastic

We build search solutions on a single stack

- Enterprise Search
- Observability
- Security

elastic

The Elastic Search Platform



elastic

Speed

Find matches in milliseconds within structured and unstructured datasets

Scale

Scale massively and horizontally across hundreds of systems

Relevance

Generate highly relevant results and actionable insights from data

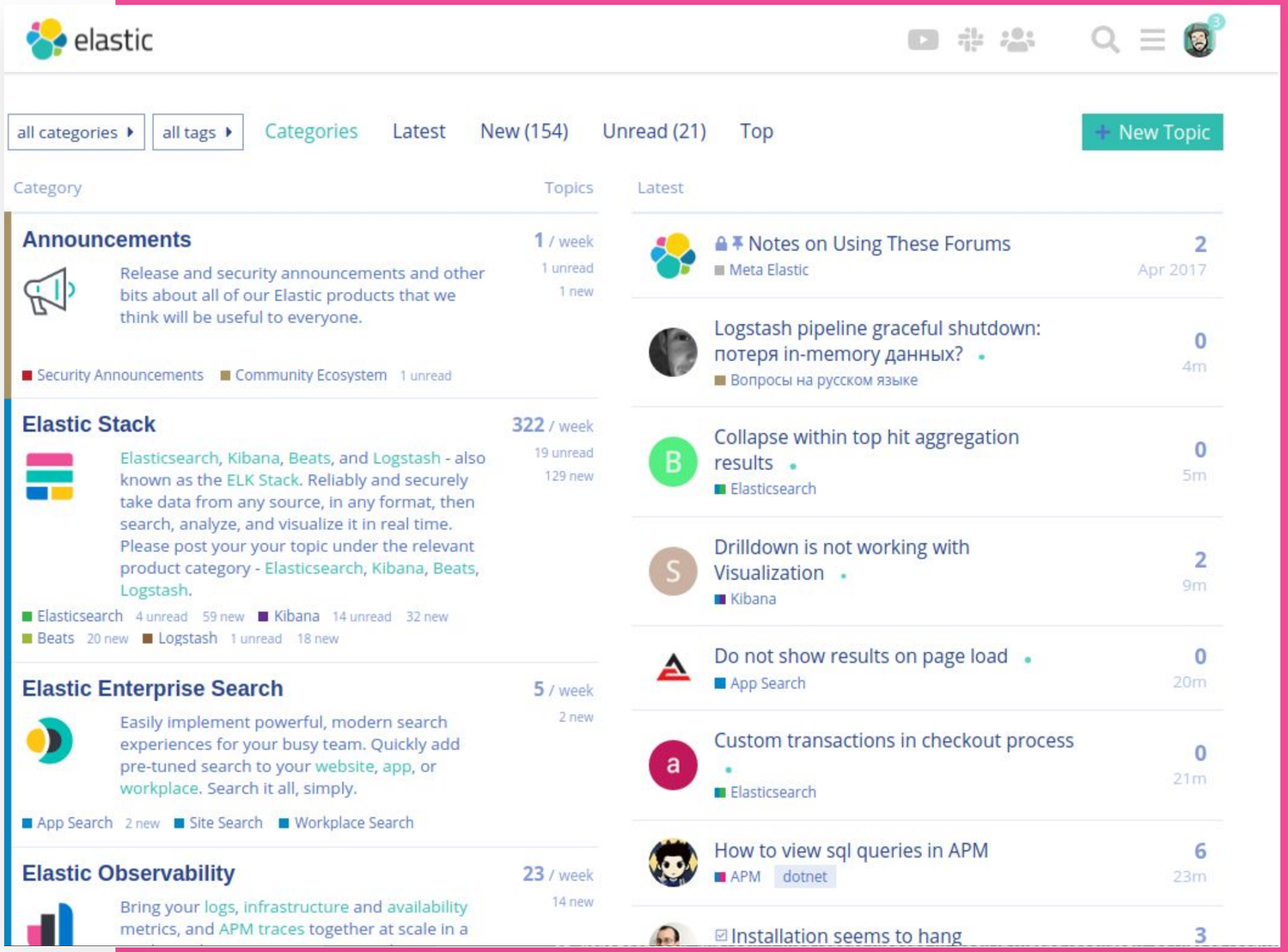
elastic

Community

https://github.com/elastic

https://ela.st/slack

https://discuss.elastic.co



Ingesting data into Elasticsearch

Using Elastic stack and third party products to upload geospatial datasets

elastic

Beats

Lightweight data shippers

FileBeat
CSVs

MetricBeat
System metrics

PacketBeat
Network Data

WinLogBeat
Window Events

HeartBeat
Uptime Monitoring

AuditBeat
Audit Data

FunctionBeat
Serverless Shipper

GeoIP

Plus, more than 70 community Beats and growing...

build passing

Earthquakebeat

Welcome to Earthquakebeat.

Earthquakebeat is a beat which periodically pulls data from USGS earthquake API. There are 2 api calls done eacr `Period` which request new and updated earthquakes.

`New` earthquakes call will request data in GeoJSON format and use attribute `starttime` set to `Now-Period`. That meas beat will pull data from past X Period of time you define. Example

```
https://earthquake.usgs.gov/fdsnws/event/1/query?format=geojson&starttime=2019-08-13T09%3A18%3A18
```

`Updated` earthquakes does the same, except is uses attribute `updatedafter` to pull last updated data. Example:

```
https://earthquake.usgs.gov/fdsnws/event/1/query?format=geojson&starttime=2019-08-13T09%3A18%3A18
```

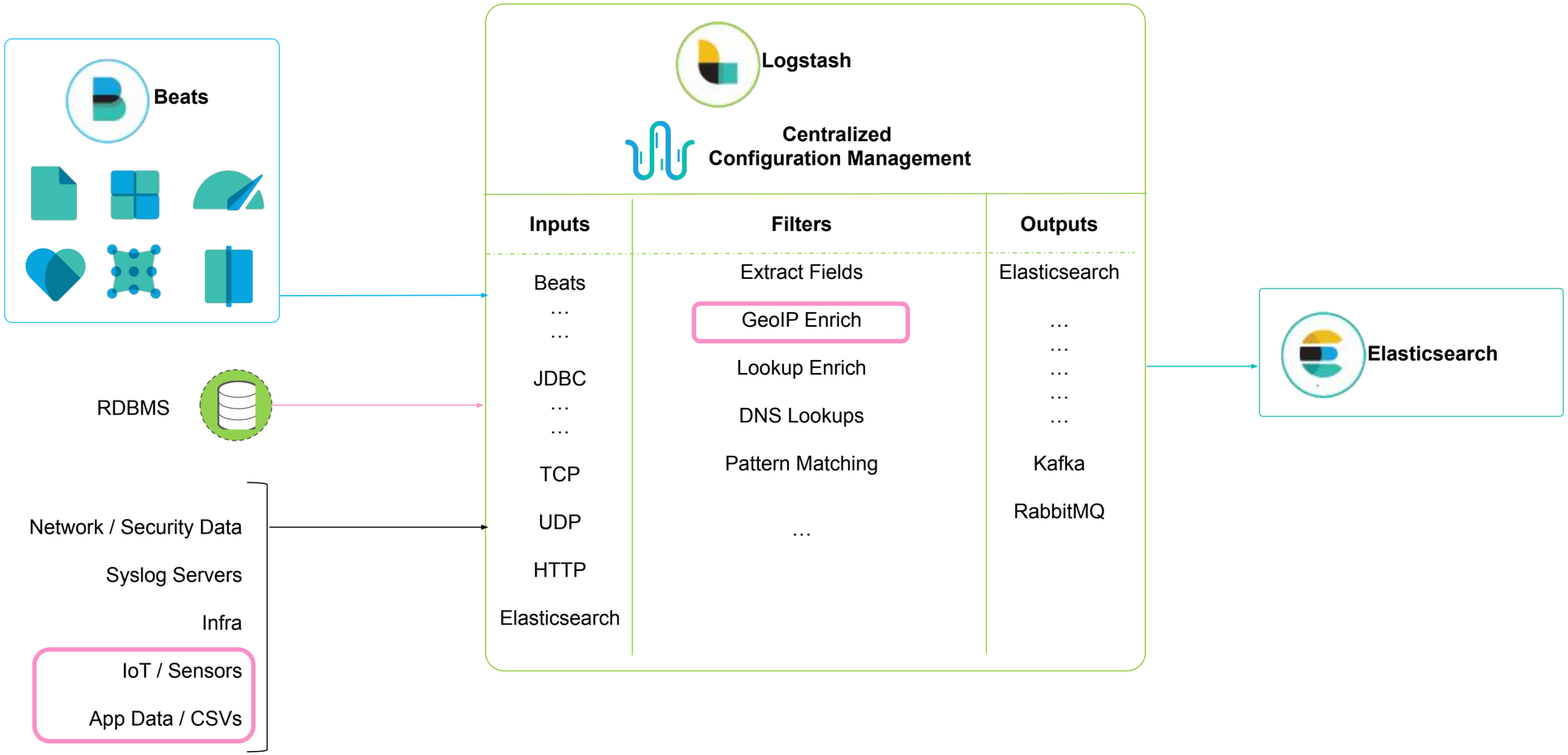
All other attributes are default and earthquakes from all over the world are being pulled.

Note: Beat preserve earthquake original ID to not to duplicate data in index.

https://github.com/radoondas/earthquakebeat

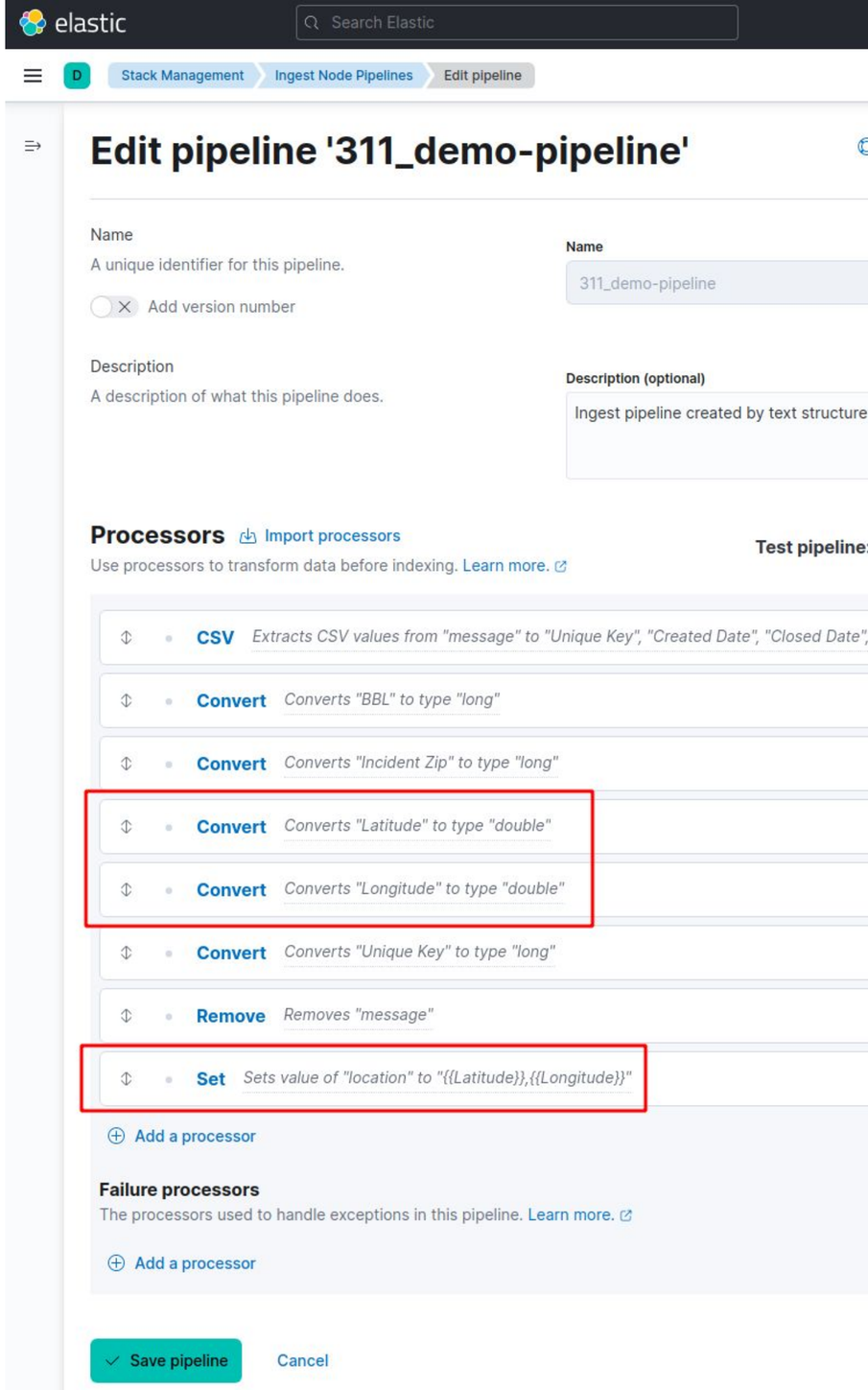
Logstash

Normalize and Enrich Data before Indexing



Ingest pipelines

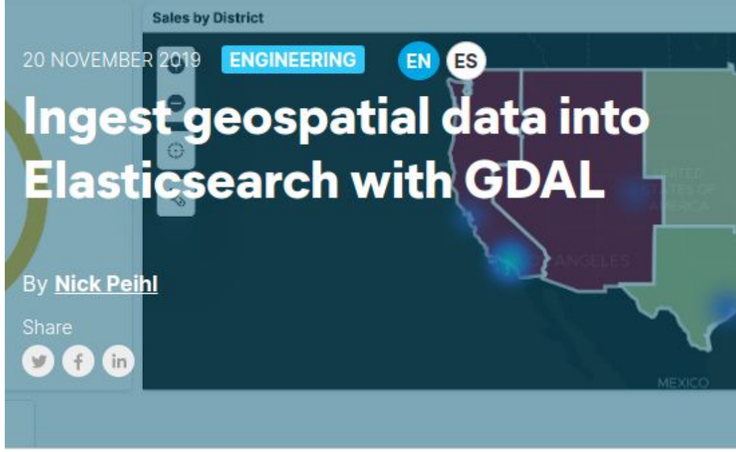
- Transform data at ingest time
- They can run in dedicated nodes 📖
- Processors 📖
  - convert, set/remove fields, ...
  - **geoip**, **enrich**
- Blog: How to map custom boundaries in Kibana with reverse geocoding
- Convenient UI in Kibana Stack Management



Ingest with ogr2ogr

https://gdal.org/drivers/vector/elasticsearch.html

- `ogr2ogr` can read and write into Elasticsearch
- Support for custom mapping definitions
- Blog posts:
  - How to ingest geospatial data into Elasticsearch with GDAL
  - Import OSM data into Elasticsearch with ogr2ogr and Docker



Have you used **Elastic Maps** in Kibana yet? I am very excited about mu
layer support. Heat maps, vector layers from the Elastic Maps Service,
individual documents all in the same interface! What a fantastic way to
and visualize your data.

But what about geospatial data that's not in Elasticsearch? Maybe you
overlay a shapefile of regional sales territories with sales aggregations.
you have a CSV file of distribution center locations, and you want to ge
data into Elasticsearch, but configuring Filebeat or Logstash is not idea
ingesting static datasets. Well, we have the perfect solution for you: GD

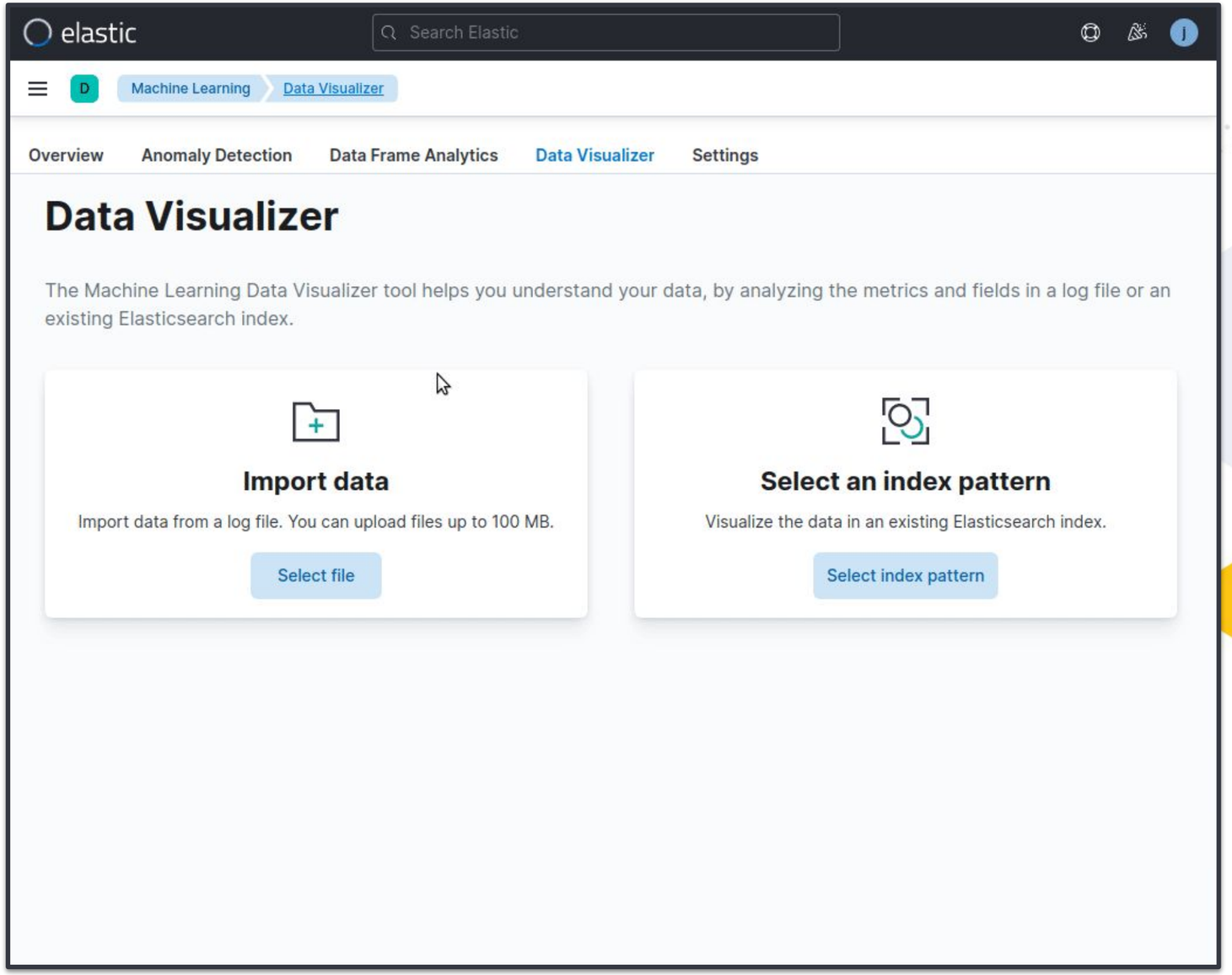
GDAL (Geospatial Data Abstraction Library) contains command line too
can convert geospatial data between over 75 different geospatial file fo
including **Elasticsearch**. GDAL can be **compiled from source** or **install**
package managers. GDAL can also be installed via **Homebrew OSGeo**
(ex. `brew tap osgeo/osgeo4mac && brew install osgeo-gdal`). Note, yo
have GDAL v3.1 or later to ingest data into Elasticsearch 7.x.

Connecting to Elasticsearch

Once you've installed GDAL, open your command line or terminal windo
try connecting to your Elasticsearch cluster using the `ogrinfo` tool. We
the URL with "ES:" to tell GDAL to use the Elasticsearch driver.

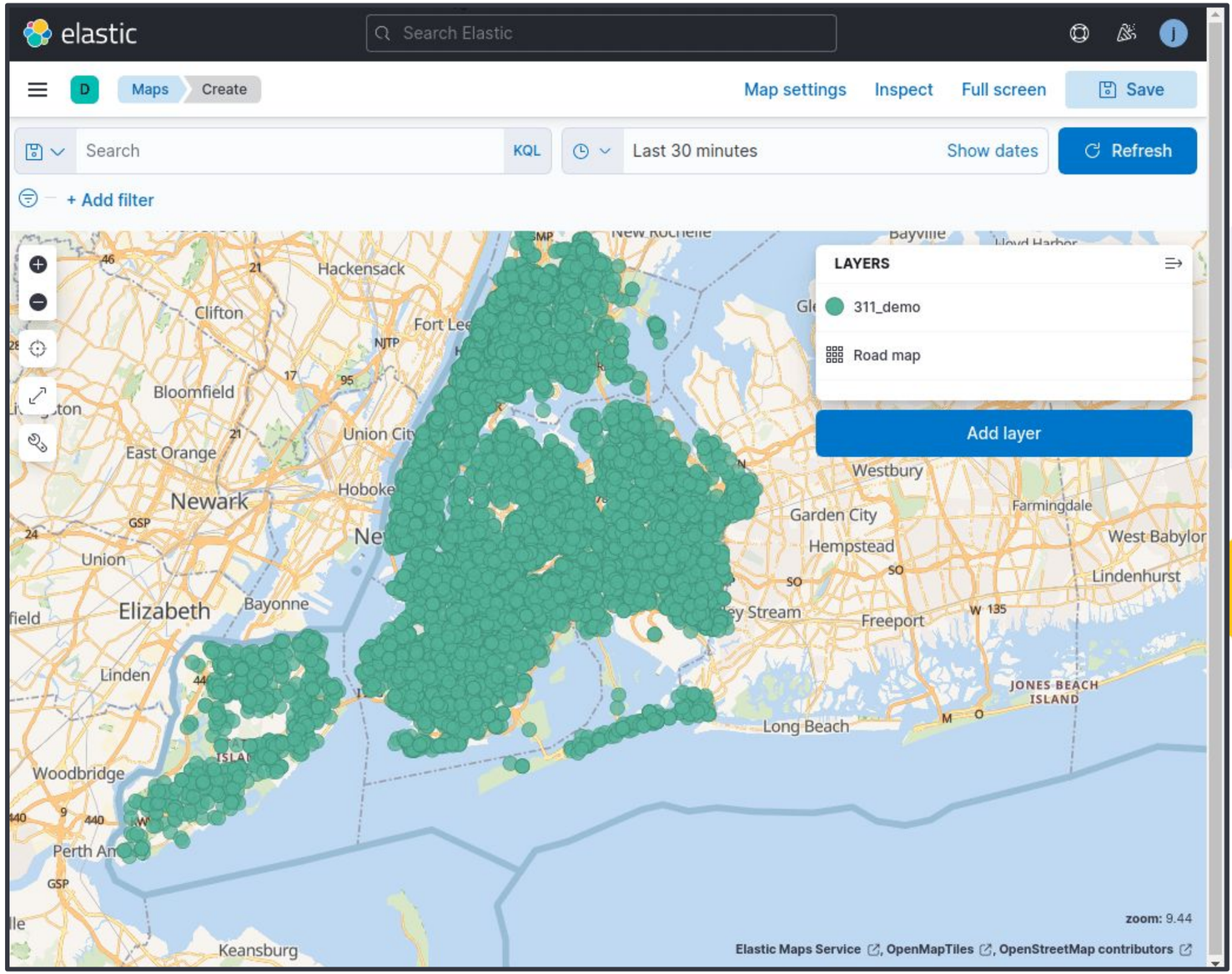
```
ogrinfo ES:http://localhost:9200
```

Ingest with Kibana: CSV file upload



elastic

Ingest with Kibana: GeoJSON upload



- Docs 📖
- Tutorial 📖

elastic

Store, search, aggregate

Data types, Elastic query DSL,
geospatial aggregations

elastic

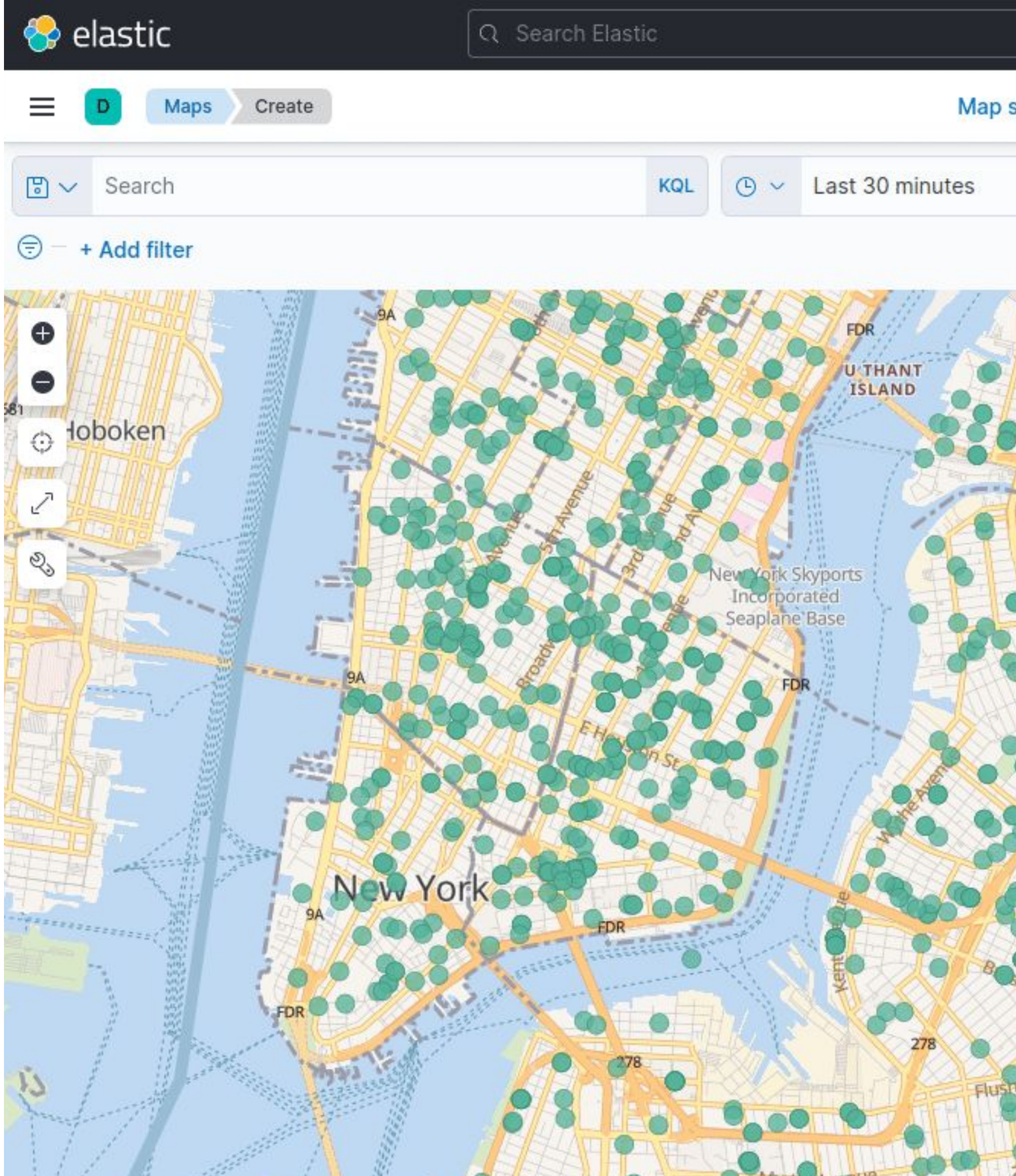
Elasticsearch geospatial data types

- `geo_point` 📖
  - A single pair of latitude and longitude **coordinates**
  - Can be inserted as an object, WKT, array, geohash
- `geo_shape` 📖
  - Supports any **lat/lon** geometry type, incl. envelope and circle
  - Inserted with GeoJSON or WKT notation
- `shape` 📖
  - Supports any **cartesian** geometry type
  - Inserted with GeoJSON or WKT notation

Elasticsearch queries

Filter documents with geospatial relationships

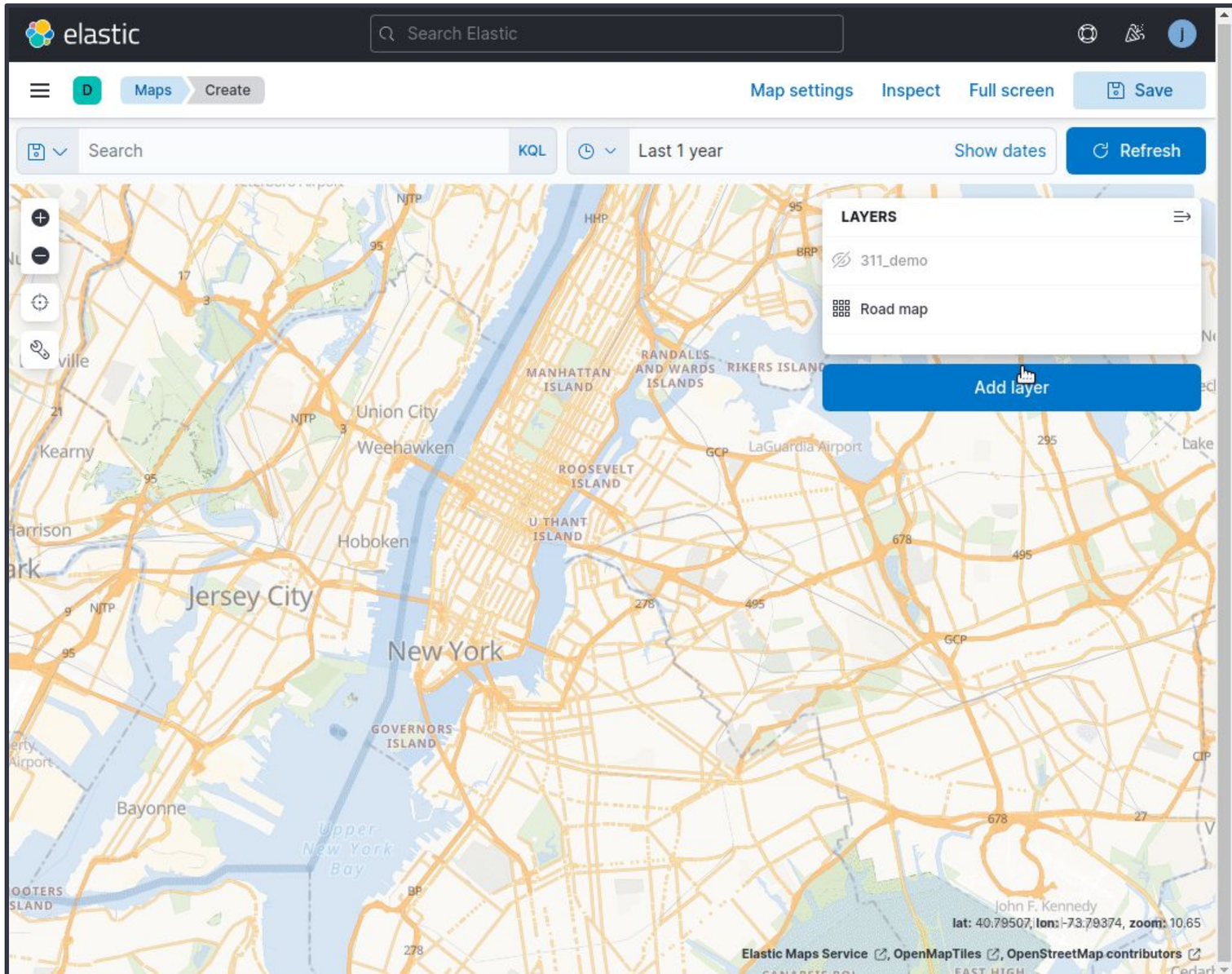
- Bounding box
- Point and radius
- Polygon
- An indexed geo_shape



Elasticsearch bucket aggregations

Bin documents based on their location into categories

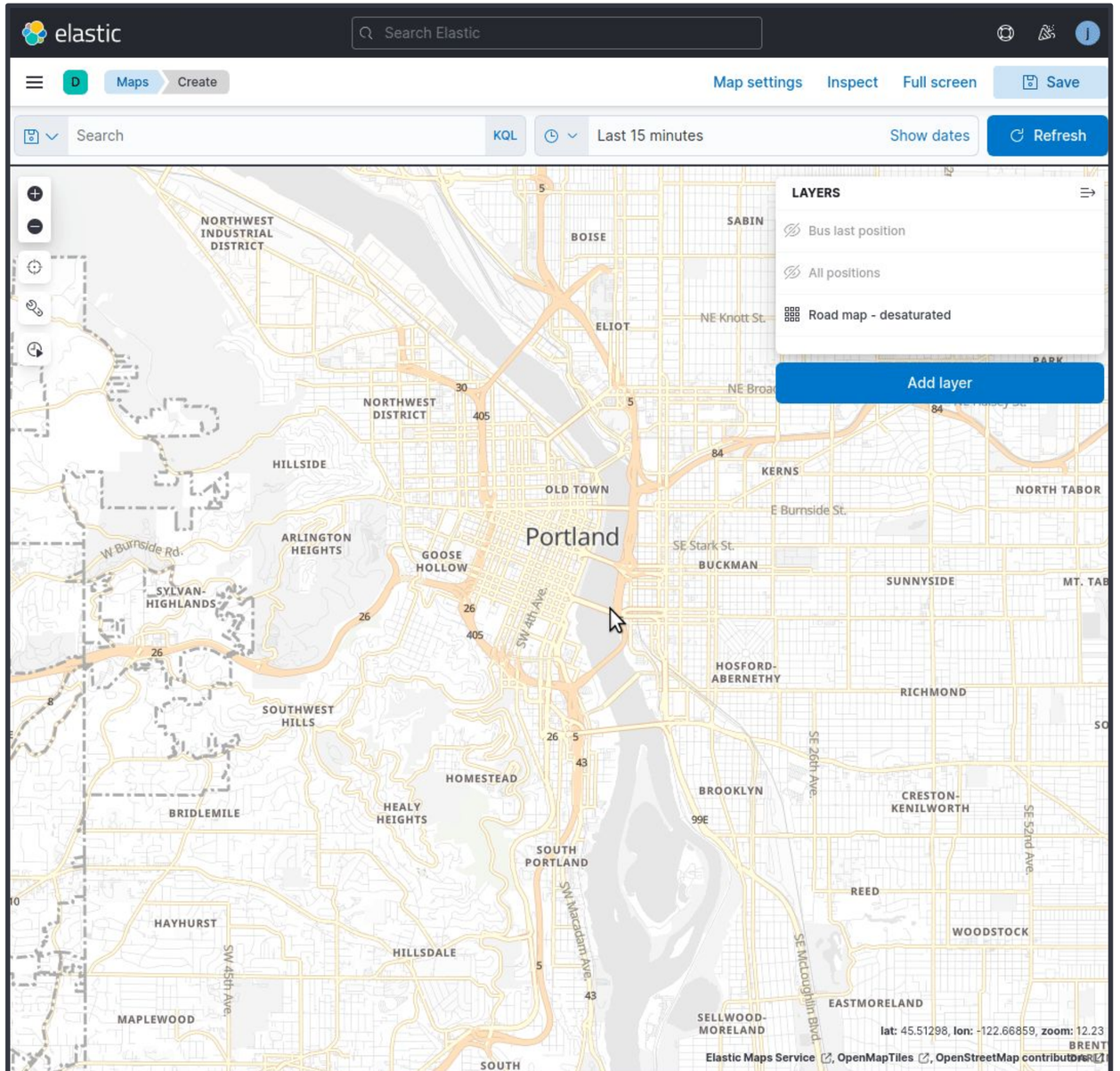
- Distance (rings) 📖
- Hash 📖
- Geotile 📖



Elasticsearch metric aggregations

Compute geospatial metrics derived from aggregating documents

- Centroid 📖
- Bounds 📖
- Geoline 📖



Visualization strategies

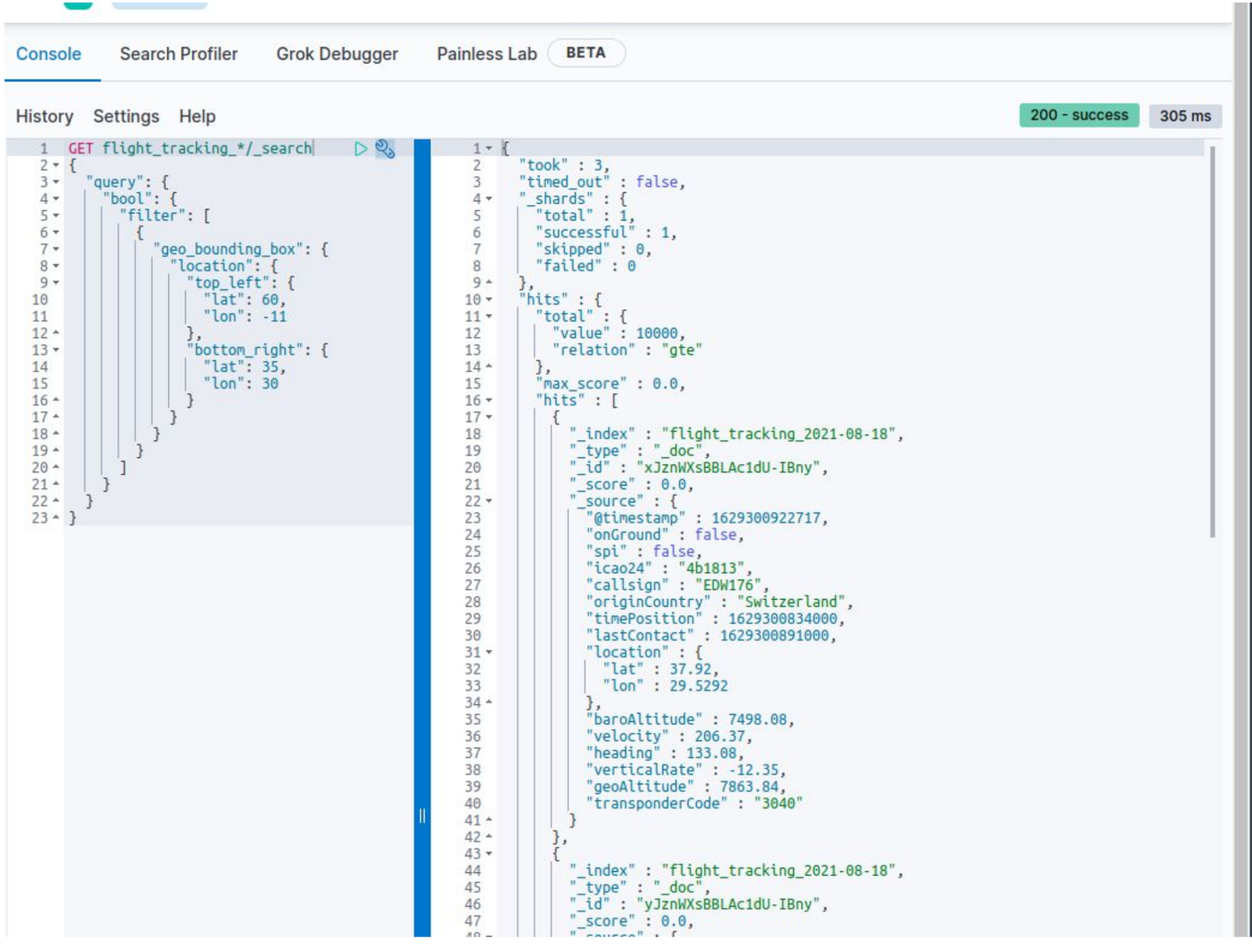
Render geospatial data with Kibana
or your own application

elastic

Elasticsearch JSON output

Getting data out to visualize

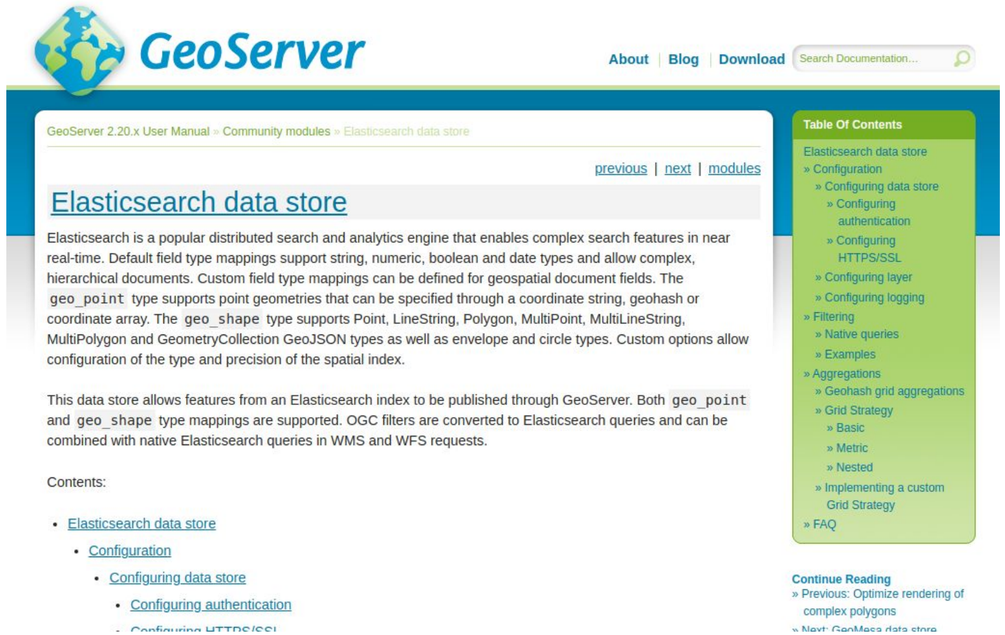
- **JSON** is the default output format
- `_search` limits to 10K docs
- Pagination needs a “point in time” to freeze the search context



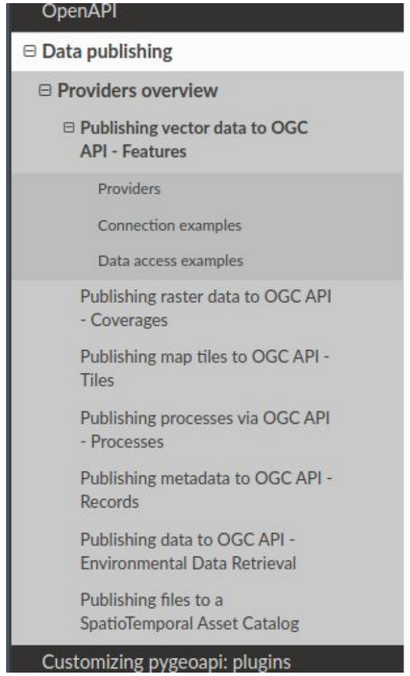
OGC servers and Elasticsearch

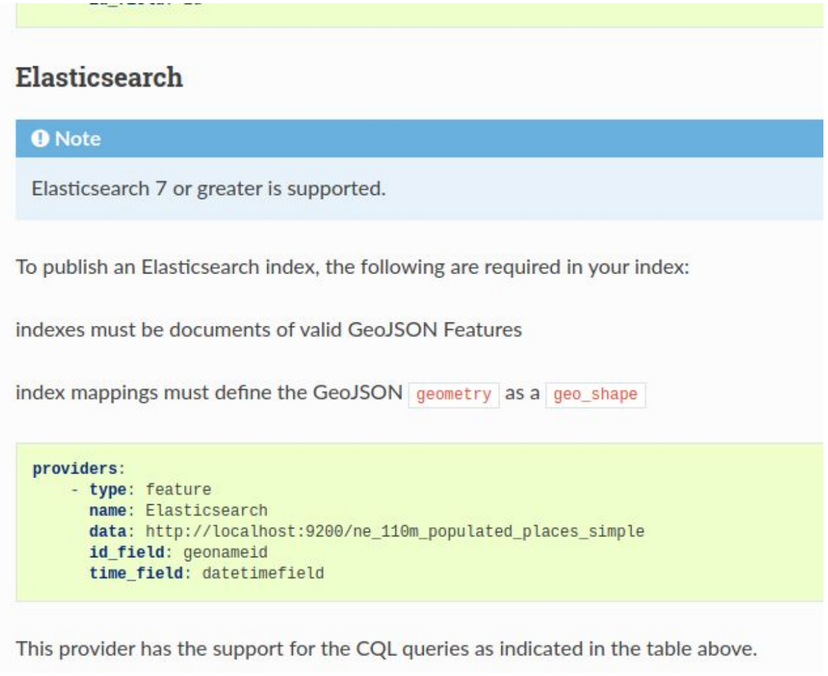
Expose Elasticsearch indices as OGC services

GeoServer



pygeoapi

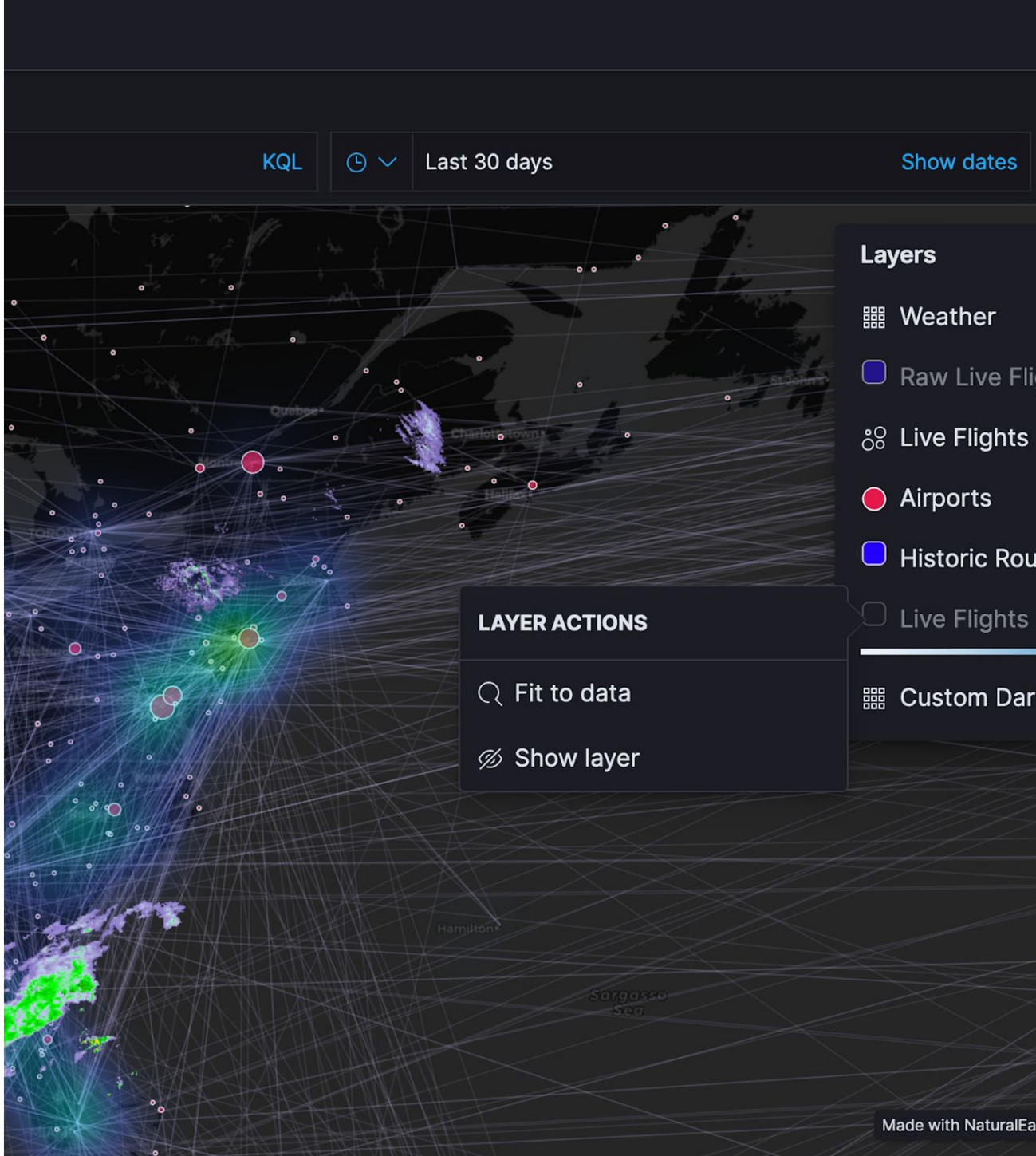




elastic

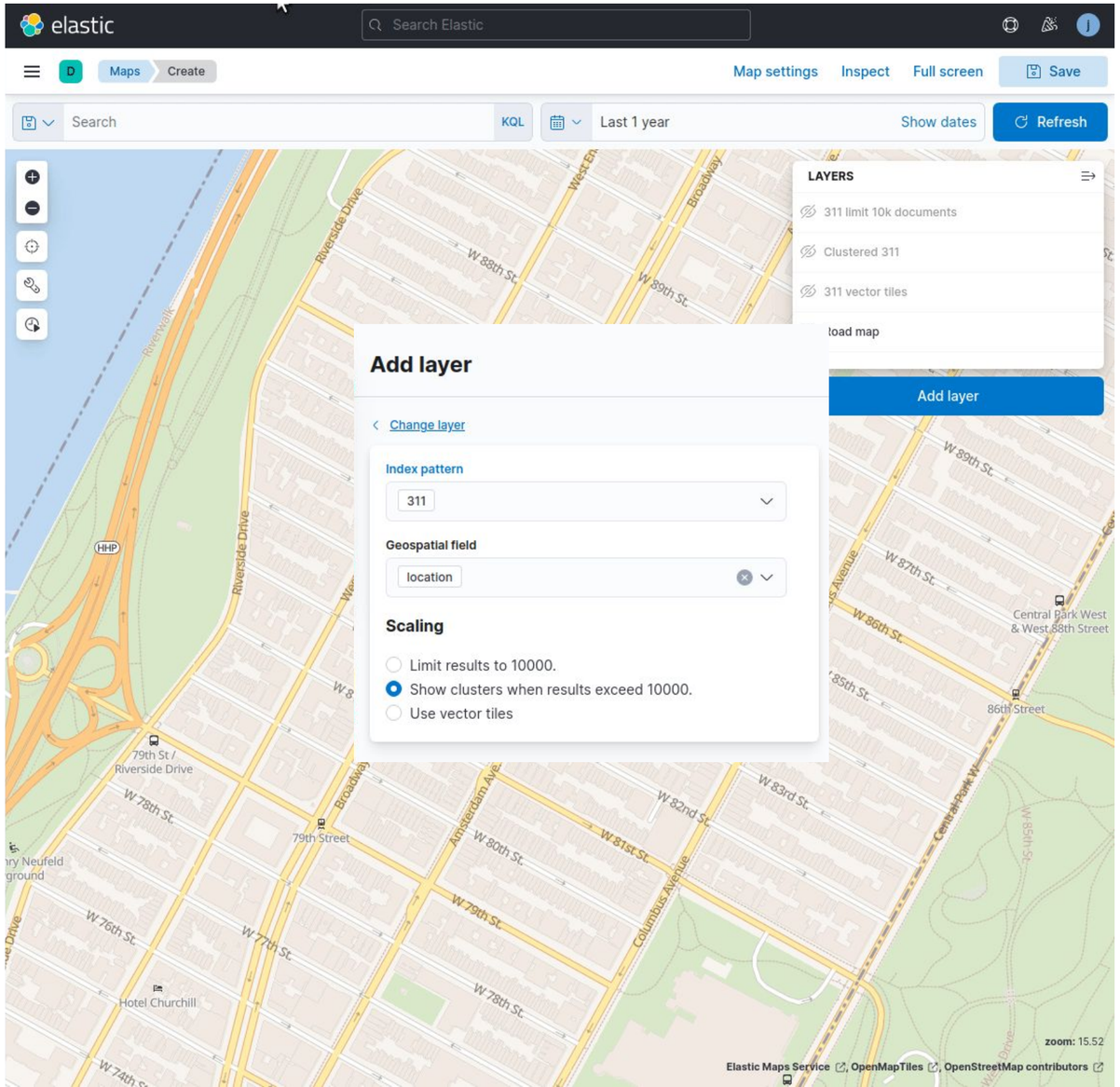
Elastic Maps

- Friendly user experience
- MapboxGL/MapLibre
- `geo_point` and `geo_shape`
- Aggregations: heat map, clustering, grids, geoline
- Data driven styling
- Tools for drawing, filtering, measuring
- Used alone or in dashboards or Canvas workdpads
- The map component for other Kibana applications



Elastic Maps rendering strategies

- Single request with up to 10K documents by map extent
- Automatic clustering when >10K documents
- Vector tiles with up to 10K documents per tile (better caching)



New features

elastic

Vector Tiles

A new output format for Elasticsearch

- New endpoint: `_mvt` (7.15)
- Output in protobuffer format
- To be consumed by a middleware to configure filters, aggregations, etc.
- Will replace current vector tiles implementation in Kibana Elastic Maps

elastic

Runtime fields and geo

Elasticsearch is not anymore only a “schema on write” database

- Ability to define new fields at query time
- New in **7.14**: get centroid, height, and width of `geo_shape` fields with Painless scripting 📖
- More geospatial functionality will come

elastic

Wrap up

- **Download** the stack or start a **Elastic Cloud** 14 days free trial from https://www.elastic.co/downloads
- Use any of the **ingest** tools to upload your own data
- **Explore** and **visualize** with Elastic Maps and Kibana
- Share your feedback and questions at discuss.elastic.co

elastic

Thanks!

Jorge Sanz and Thomas Neirynck
jorge.sanz@elastic.co - thomas@elastic.co
https://ela.st/foss4g-2021
https://ela.st/elasticon21-geo-kibana

2021-10-01 - FOSS4G 2021 Buenos Aires