

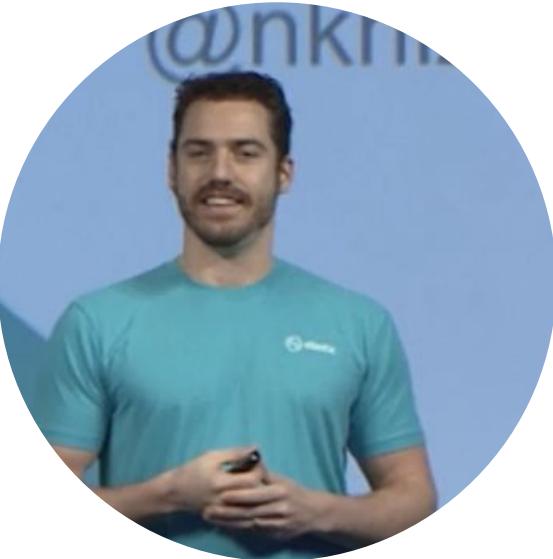


Geo Capabilities in Elasticsearch

Nicholas Knize
@nknize

May 3, 2018





Nicholas Knize
Elasticsearch & Apache Lucene Geo Software Guy
Elastic

Housekeeping & Logistics

- Slides and recording will be available following the webinar
- Chat via IRC #elastic-webinar
 - #elastic-webinar @ Freenode
 - Click "Join the Chat" link, create an IRC account
- Please select high resolution in the YouTube video player



Elastic Stack

100% open source
No enterprise edition





X-Pack

Single install
Extensions for the Elastic Stack
Subscription pricing



Security



Alerting



Monitoring



Reporting



Graph



Machine Learning



Elastic Cloud

Hosted Elasticsearch & Kibana
Get started with a 14-day free trial
at www.elastic.co/cloud

The screenshot shows the Elastic Cloud web interface. At the top right are 'Account' and 'Sign out' buttons. Below the header is a navigation bar with 'cloud' (with a logo), 'Clusters', 'Plugins', and 'Help'. The main area has two sections: 'Summary' and 'Cluster Size'.
Summary: Shows configuration details:

- Region: US East (N. Virginia)
- Memory: 1 GB
- Storage: 24 GB
- SSD: Yes
- High availability: No
- Hourly rate: \$0.0612
- Monthly rate: \$45

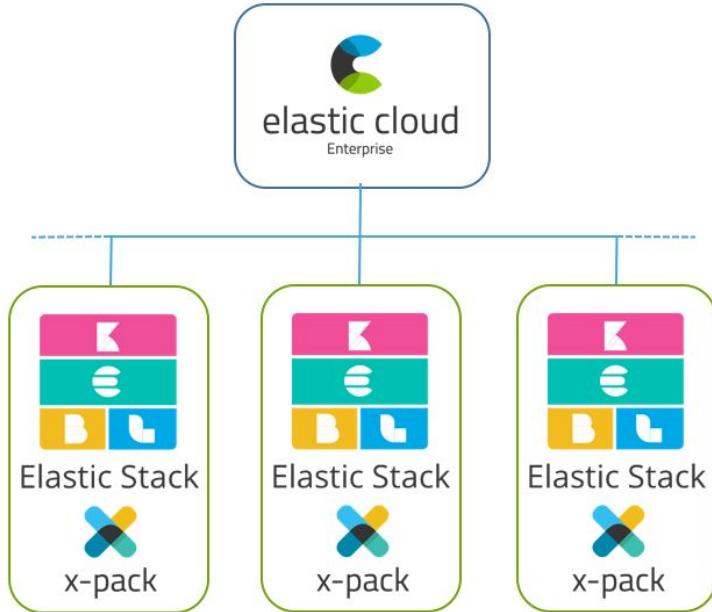
A large blue 'Create' button is at the bottom.
Cluster Size: A slider allows selecting cluster size from 1 to 6144 units, with 24 selected. It includes a note: "Choose a cluster size. Cluster size can be changed later without downtime." A legend indicates blue for Memory and light blue for Storage. A checked checkbox says "SSD — Selected for improved storage performance." A link "Need a larger cluster? Contact us." is at the bottom right.

Available in AWS Marketplace, add it to your AWS bill
Choose your preferred Cloud provider, AWS or GCP
X-pack features and ticket based support



Elastic Cloud Enterprise

Provision and manage multiple Elastic Stack environments; Expose logging as a service to your entire organization



Geo Stuff!

Topics

Geospatial Indexing, Search, and Visualization

- 1 Geo Field Mappings
- 2 Geo Indexing, Search, and Lucene Data Structures
- 3 Geo Aggregations
- 4 Kibana / Elastic Maps Service

Mappings

Geo Field Types

geo_point mapping

define

```
PUT crime/incidents/_mapping
{
  "properties" : {
    "location" : {
      "type" : "geo_point",
      "ignore_malformed" : true,
    }
  }
}
```

geo_point mapping

insert

```
POST crime/incidents
{
  "location" : { "lat" : 41.12, "lon" : -71.34 }
}
```

```
POST crime/incidents
{
  "location" : "41.12, -71.34"
}
```

```
POST crime/incidents
{
  "location" : [[-71.34, 41.12], [-71.32, 41.21]]
}
```

geo_shape mapping

define

```
PUT police/precincts/_mapping
{
  "properties" : {
    "coverage" : {
      "type" : "geo_shape",
      "ignore_malformed" : false,
      "tree" : "quadtree",
      "precision" : "5m",
      "distance_error_pct" : 0.025,
      "orientation" : "ccw",
      "points_only" : false
    }
  }
}
```

geo_shape mapping

insert

```
POST police/precincts/
{
  "coverage" : {
    "type" : "polygon",
    "coordinates" : [
      [-73.9762134, 40.7538588],
      [-73.9742356, 40.7526327],
      [-73.9656733, 40.7516774],
      [-73.9763236, 40.7521246],
      [-73.9723788, 40.7516733],
      [-73.9732423, 40.7523556],
      [-73.9762134, 40.7538588]
    ]
  }
}
```

geo_shape mapping

insert

- Shapes are parsed using OGC and ISO standards definitions
 - OGC Simple Feature Access
 - ISO Geographic information — Spatial Schema (19107:2003)
- Supports the following geo_shape types
 - Point, MultiPoint
 - LineString, MultiLineString
 - Polygon (with holes), MultiPolygon (with holes)
 - Envelope (bbox)

geo_point mapping

Pre 5.0

`geohash`

Should the geo-point also be indexed as a geohash in the `.geohash` sub-field?
Defaults to `false`, unless `geohash_prefix` is `true`.

`geohash_precision` The maximum length of the geohash to use for the `geohash` and `geohash_prefix` options.

`geohash_prefix` Should the geo-point also be indexed as a geohash plus all its prefixes? Defaults to `false`.

`ignore_malformed` If `true`, malformed geo-points are ignored. If `false` (default), malformed geo-points throw an exception and reject the whole document.

`lat_lon` Should the geo-point also be indexed as `.lat` and `.lon` sub-fields? Accepts `true` and `false` (default).

`precision_step` Controls the number of extra terms that are indexed for each lat/lon point. Defaults to `16`. Ignored if `lat_lon` is `false`.

geo_point mapping

5.0+

geohash	Should the geo-point also be indexed as a geohash in the .geohash sub-field? Defaults to false, unless geohash_prefix is true.
geohash_precision	The maximum length of the geohash to use for the geohash and geohash_prefix options.
geohash_prefix	Should the geo-point also be indexed as a geohash plus all its prefixes? Defaults to false.

`ignore_malformed` If true, malformed geo-points are ignored. If false (default), malformed geo-points throw an exception and reject the whole document.

lat_lon	Should the geo-point also be indexed as .lat and .lon sub-fields? Accepts true and false (default).
precision_step	Controls the number of extra terms that are indexed for each lat/lon point. Defaults to 16. Ignored if lat_lon is false.

geo_shape mapping

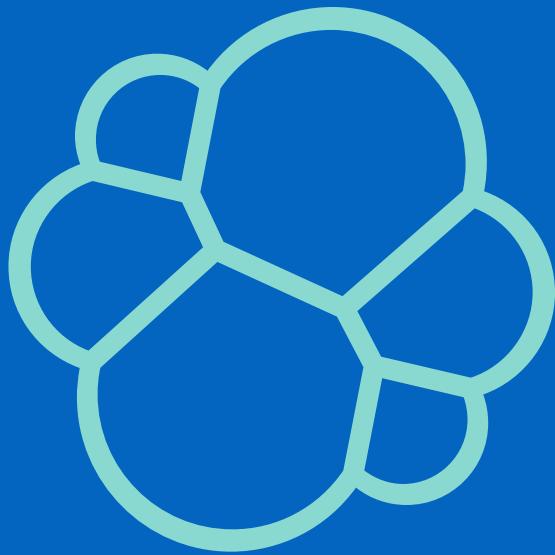
current

Option	Description	Default
tree	Name of the PrefixTree implementation to be used: geohash for GeohashPrefixTree and quadtree for QuadPrefixTree.	geohash
precision	This parameter specifies the desired precision of the index. The value should be a number followed by an optional distance unit. Valid distance units include: in, inch, yd, yard, mi, miles, km, kilometers, m, meters, cm, centimeters, mm, millimeters.	meters
tree_levels	Maximum number of layers to be used by the PrefixTree. Defaults to the default value of the chosen PrefixTree implementation.	50m
strategy	The strategy parameter defines the approach for how to represent shapes at indexing and search time.	recursive
distance_error_pct	Used as a hint to the PrefixTree about how precise it should be.	0.025
orientation	Optionally define how to interpret vertex order for polygons / multipolygons. The default orientation (counterclockwise) complies with the OGC standard.	ccw
points_only	Setting this option to true configures the geo_shape field type for point shapes only (NOTE: Multi-Points are not yet supported).	false

geo_shape mapping

7.0+

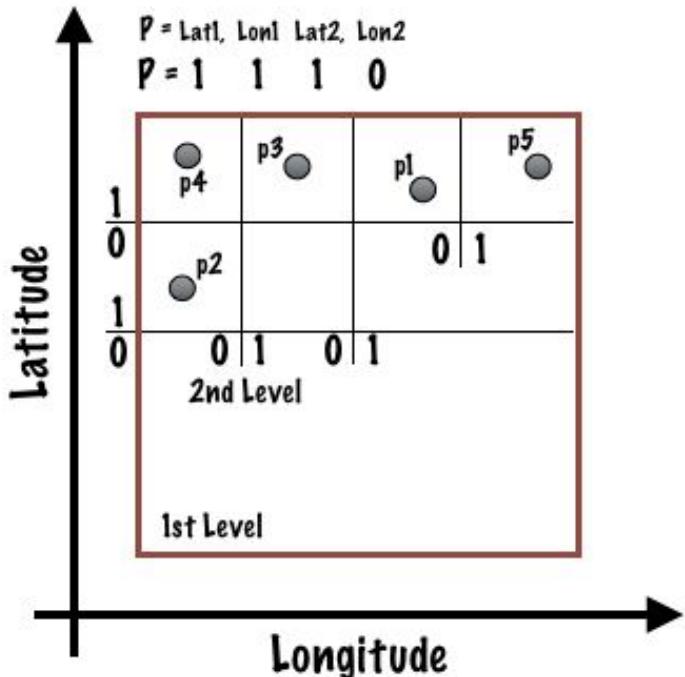
Option	Description	Default
tree	Name of the PrefixTree implementation to be used: geohash for GeohashPrefixTree and quadtree for QuadPrefixTree.	geohash
precision	This parameter specifies the desired precision of the index. The value should be a number followed by an optional distance unit. Valid distance units include: in, inch, yd, yard, mi, miles, km, kilometers, m, meters, cm, centimeters, mm, millimeters.	meters
tree_levels	Maximum number of layers to be used by the PrefixTree. Defaults to the default value of the chosen PrefixTree implementation.	50m
strategy	The strategy parameter defines the approach for how to represent shapes at indexing and search time.	recursive
distance_error_pct	Used as a hint to the PrefixTree about how precise it should be.	0.025
orientation	Optionally define how to interpret vertex order for polygons / multipolygons. The default orientation (counterclockwise) complies with the OGC standard.	ccw
points_only	Setting this option to <code>true</code> configures the <code>geo_shape</code> field type for point shapes only (NOTE: Multi-Points are not yet supported).	false



Geo Indexing

geo_point indexing

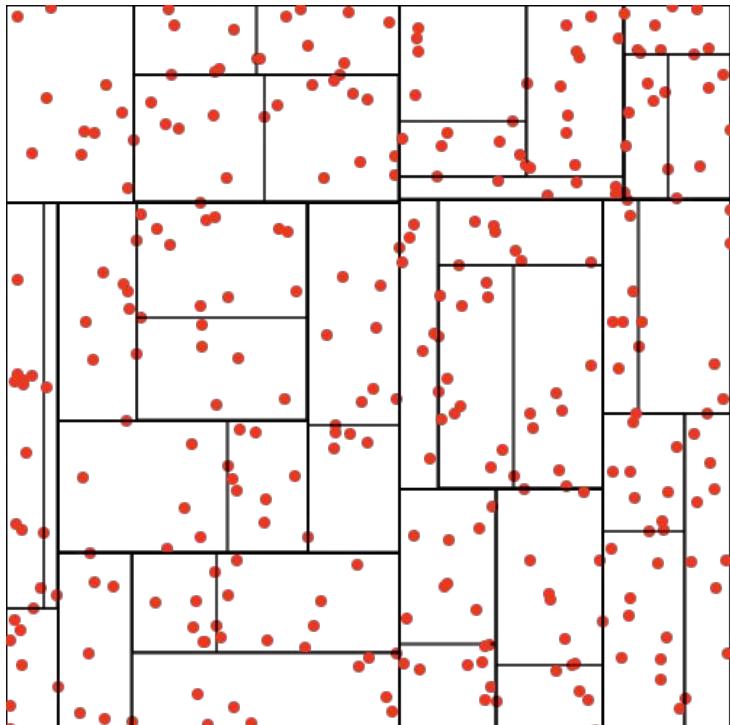
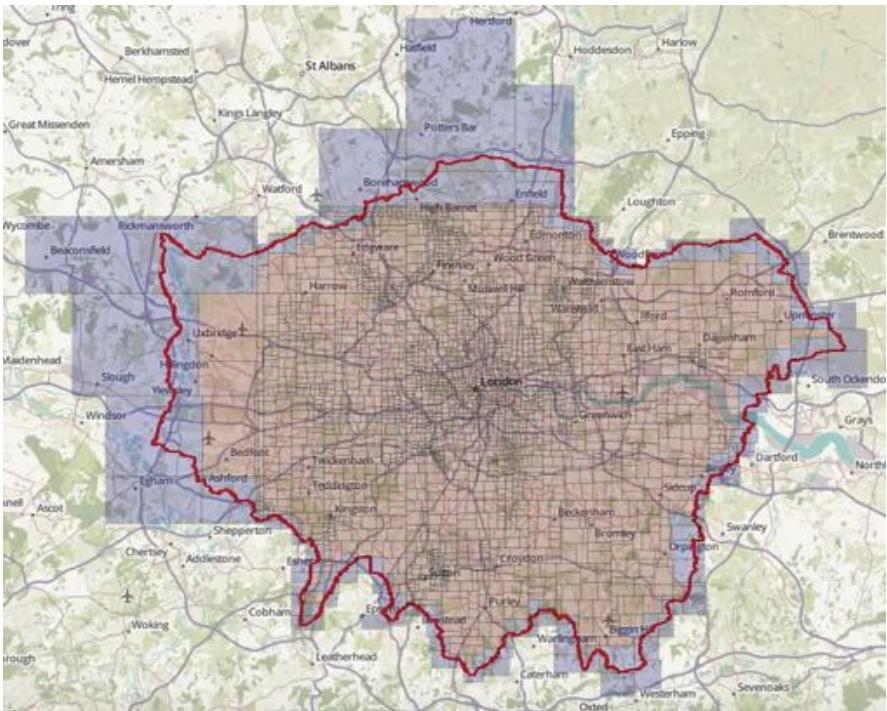
2.x term/postings encoding



term	postings (doc ids)
1	1, 2, 3, 4, 5
10	1, 2, 4
11	3, 5
100	1
101	2, 4
111	3, 5
1000	2
1010	4
1011	3
1110	3
1111	5

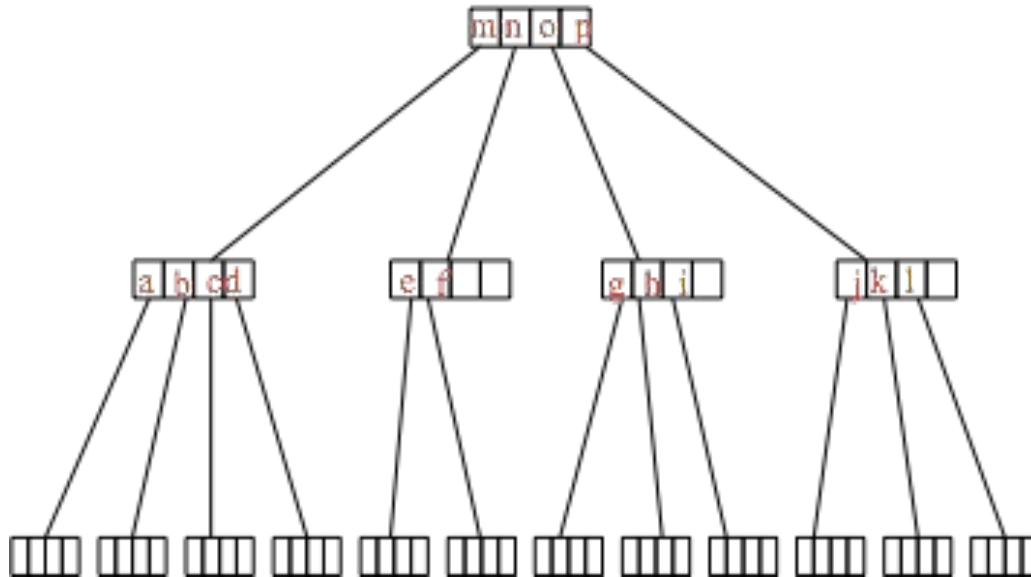
geo_point indexing

5.0 - “points” data structure - (Bkd-tree)



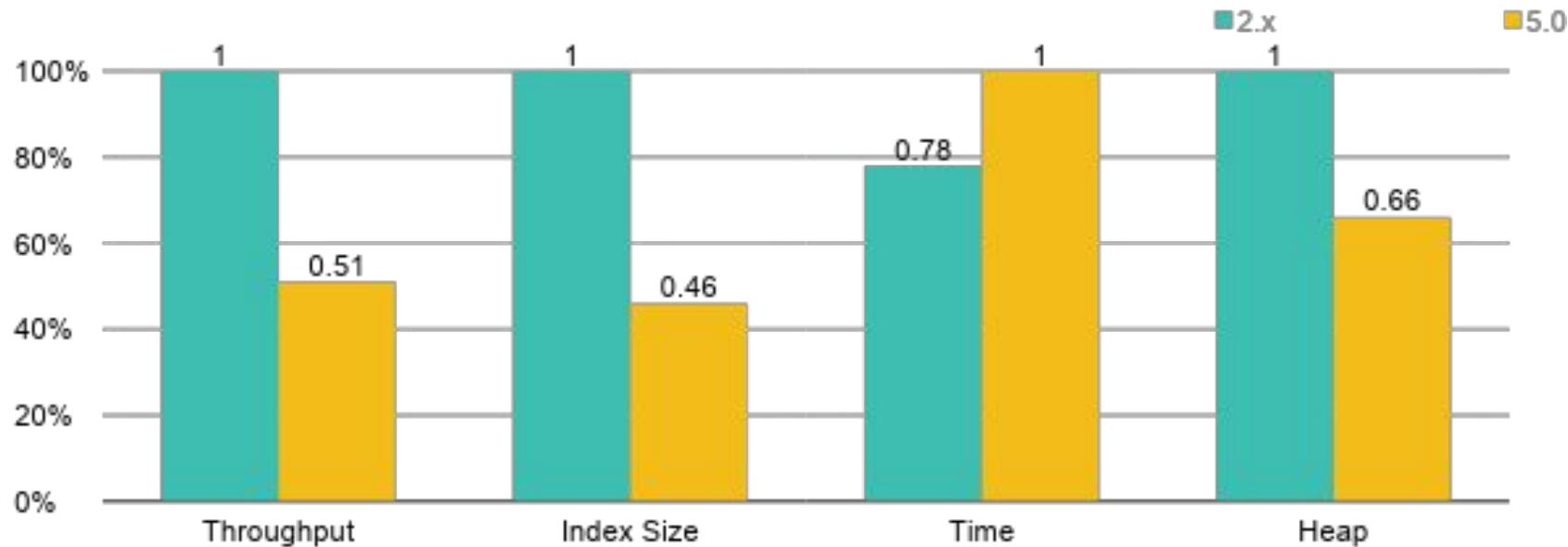
geo_point indexing

5.0 - “points” data structure - (Bkd-tree)



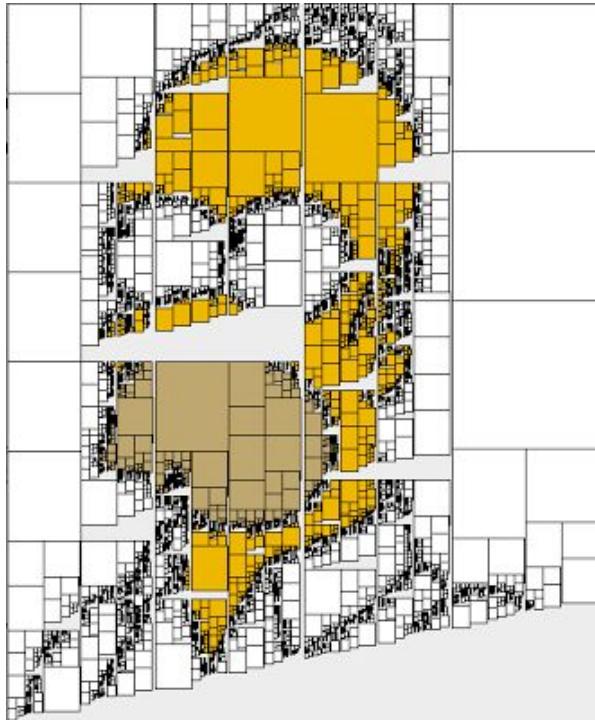
geo_point indexing

performance improvements



geo_shape indexing

current - terms/postings encoding

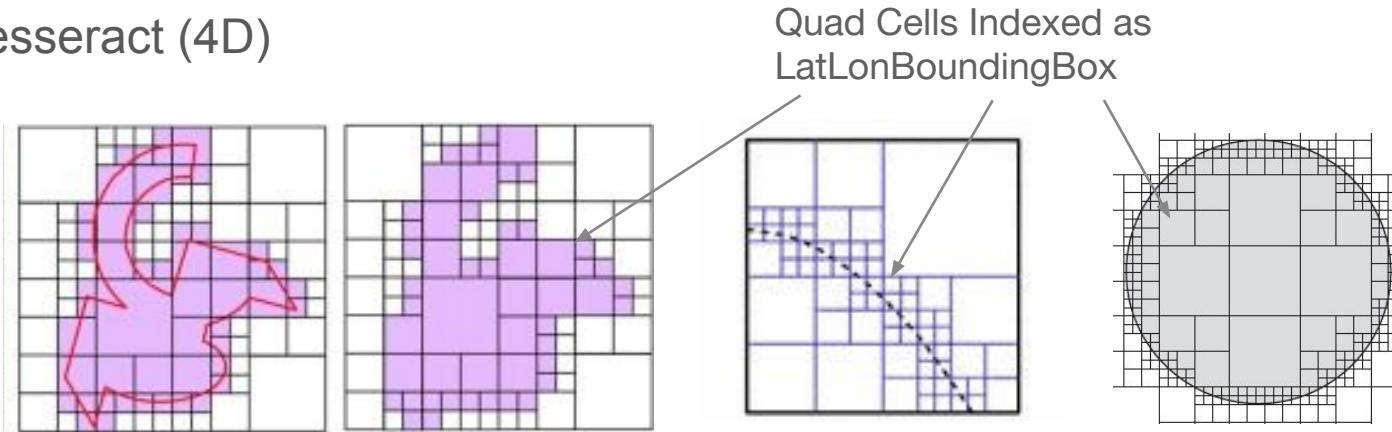


- Max tree_levels == 32 (2 bits / cell)
- distance_error_pct
 - “slop” factor to manage transient memory usage
 - % of the diagonal distance (degrees) of the shape
 - Default == 0 if precision set (2.0)
- points_only
 - optimization for points only shape index
 - short-circuits recursion

geo_shape indexing

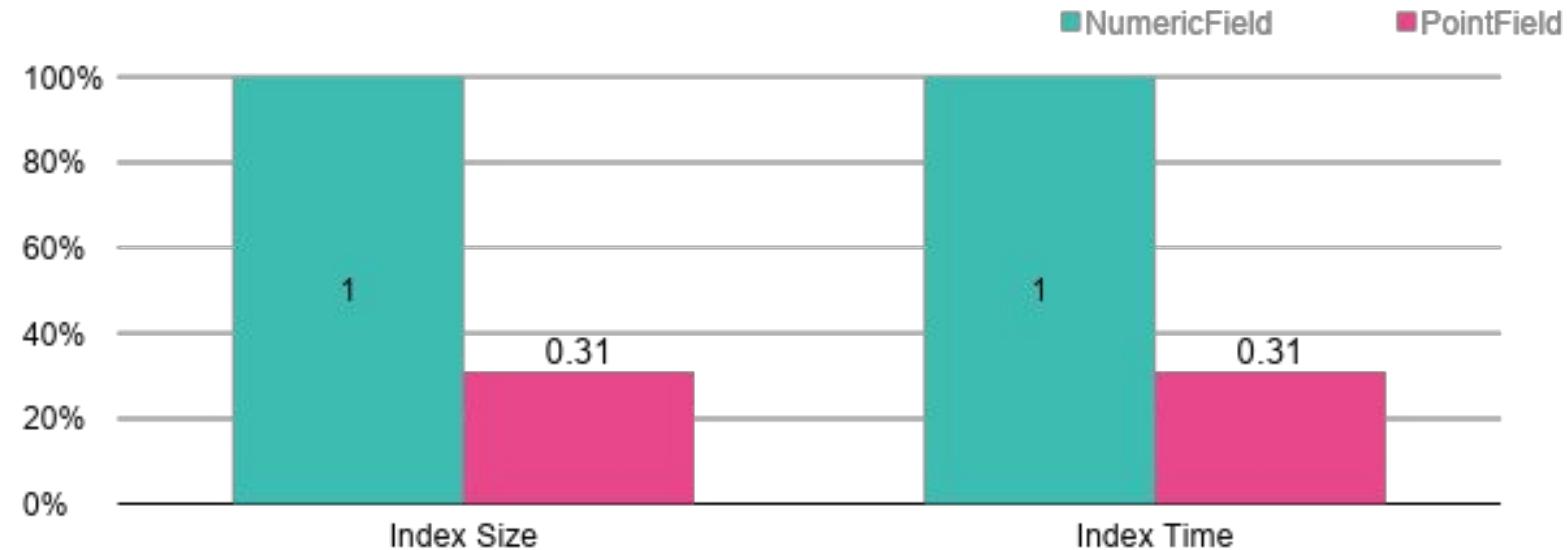
7.0+ - “ranges” encoding (Bkd-tree)

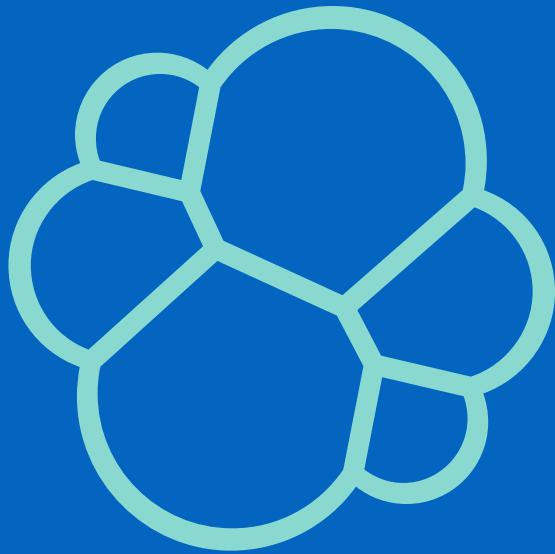
- Dimensional Shapes represented using Minimum Bounding Ranges (MBR)
 - Ranges (1D) - **Available from 5.1+ for numerics, dates, and IP (v4 and v6)**
 - Rectangles (2D) - LatLonBoundingBox **Available in Lucene 7.1+**
 - Cubes (3D)
 - Tesseract (4D)



geo_shape indexing

performance - 1D Numerics

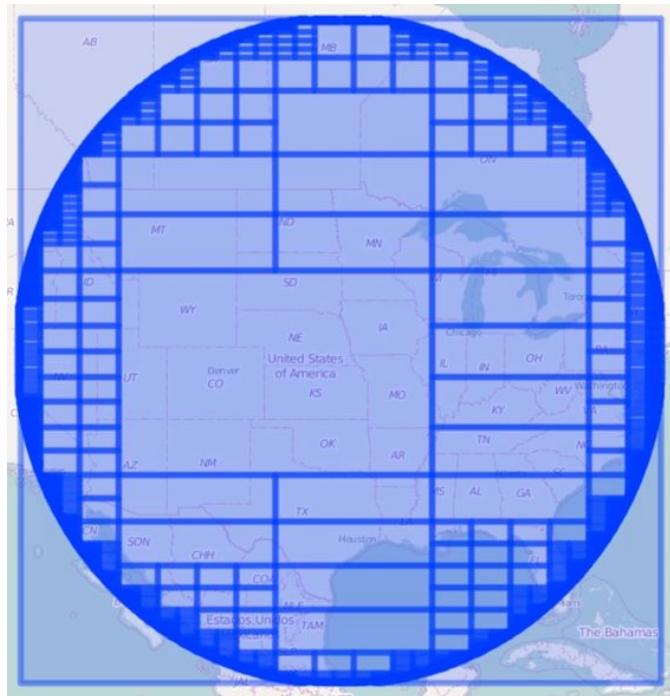




Geo Search

geo_point search

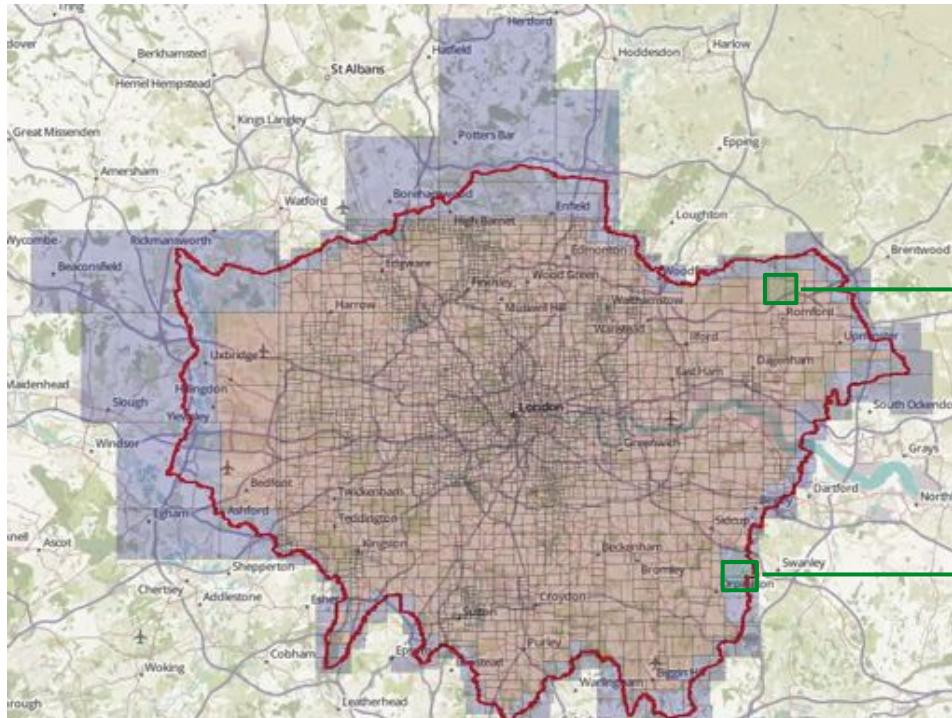
Pre 5.0 - terms/postings encoding



- Spatial Queries
 - BoundingBox, Distance, DistanceRange, Polygon
- PRECISION_STEP controls number of query terms (must match with index)
- TwoPhaseFilter
 - Delays boundary confirmation so other query (filters, conjunctions) can pre-filter

geo_point search

5.0+ - “points” encoding (Bkd-tree)

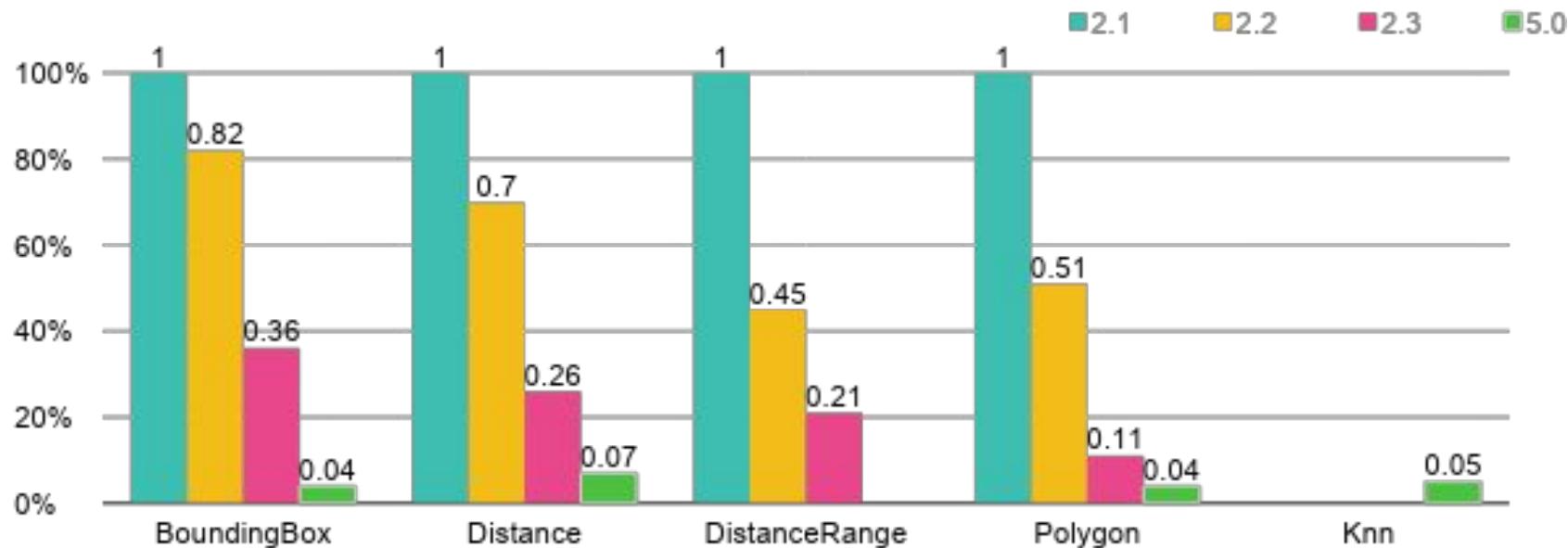


1 Leaf cell is fully within polygon (salmon) - return all docs

2 Leaf cell crosses the boundary (gray) - two-phase check

geo_point search

5.0+ - performance improvements



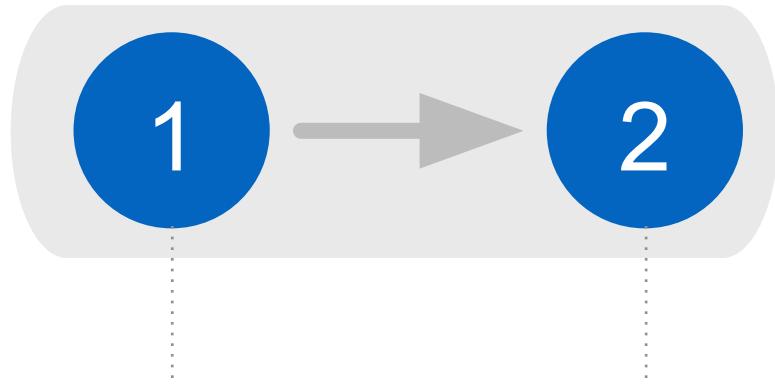
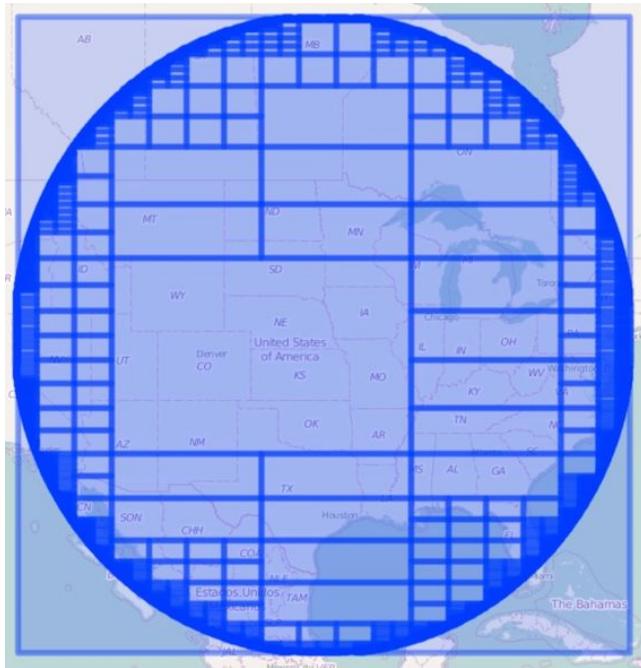
geo_shape search

capabilities

- Supports the following geo_shape types
 - Point, MultiPoint
 - LineString, MultiLineString
 - Polygon (with holes), MultiPolygon (with holes)
 - Envelope (bbox)
- Supports relational queries
 - INTERSECTS, DISJOINT, WITHIN, CONTAINS

geo_shape search

current - terms/postings encoding

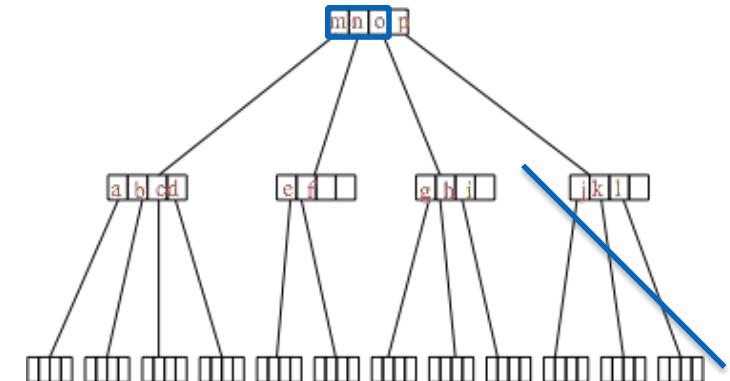
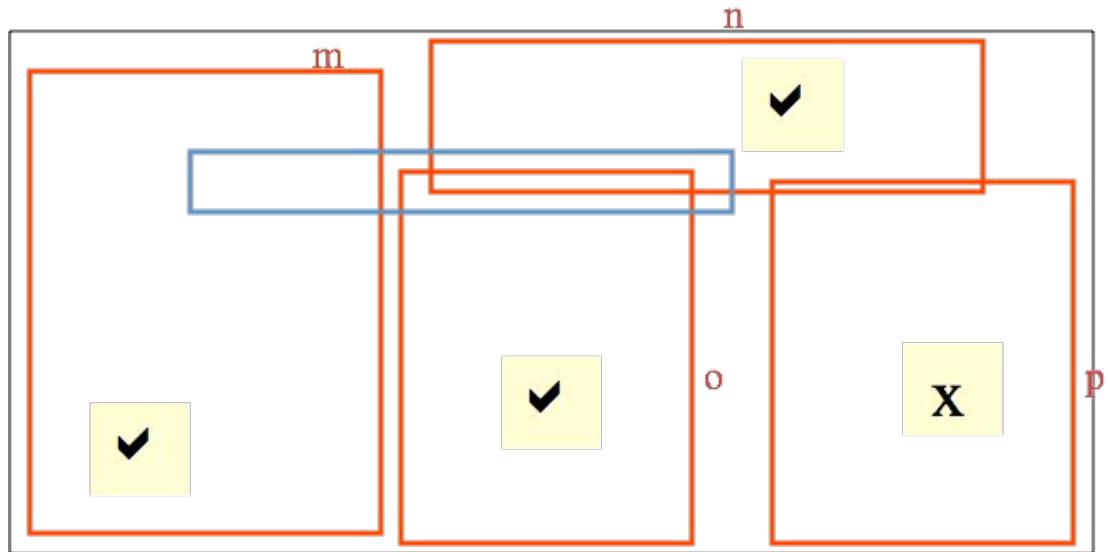


Recursively Traverse
Query terms

Collect DocIDs from
Postings based on
requested relation

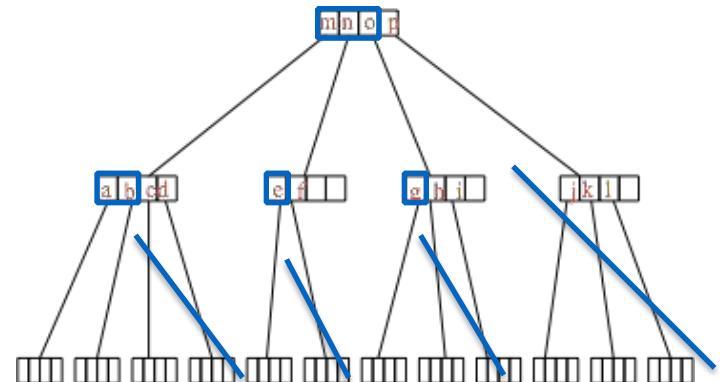
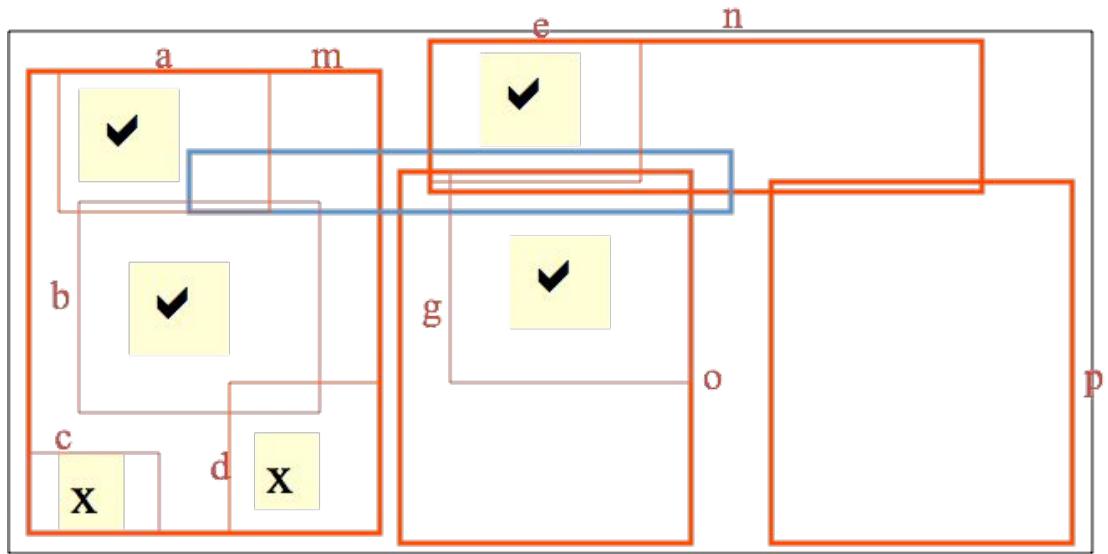
geo_shape search

7.0+ - “points” encoding (B-kd Tree)



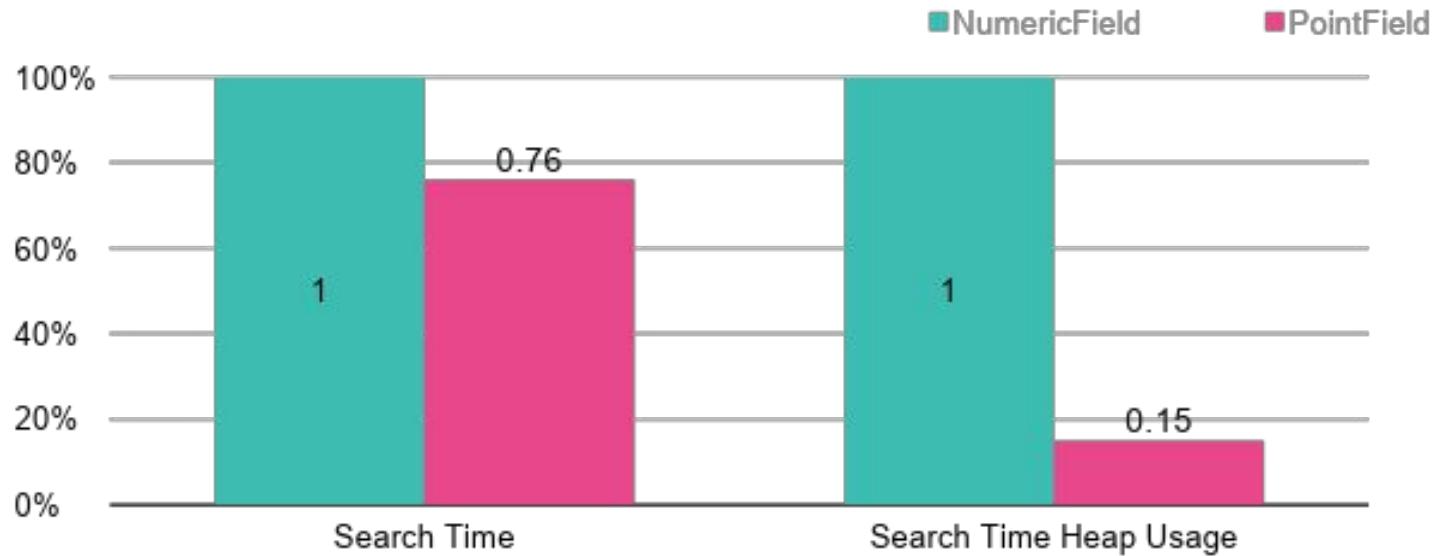
geo_shape search

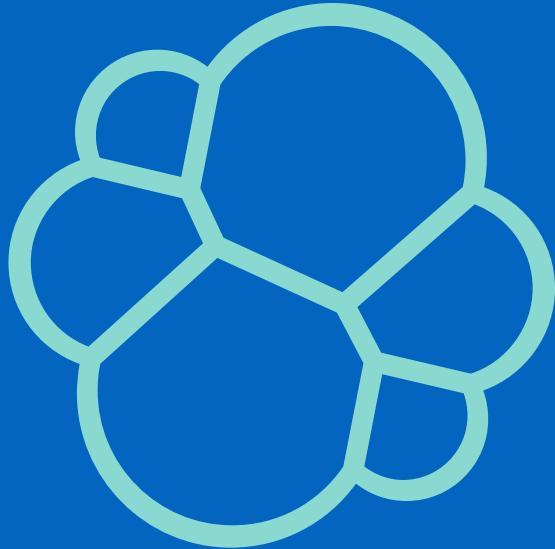
7.0+ - “points” encoding (B-kd Tree)



geo_shape search

1D numeric range performance



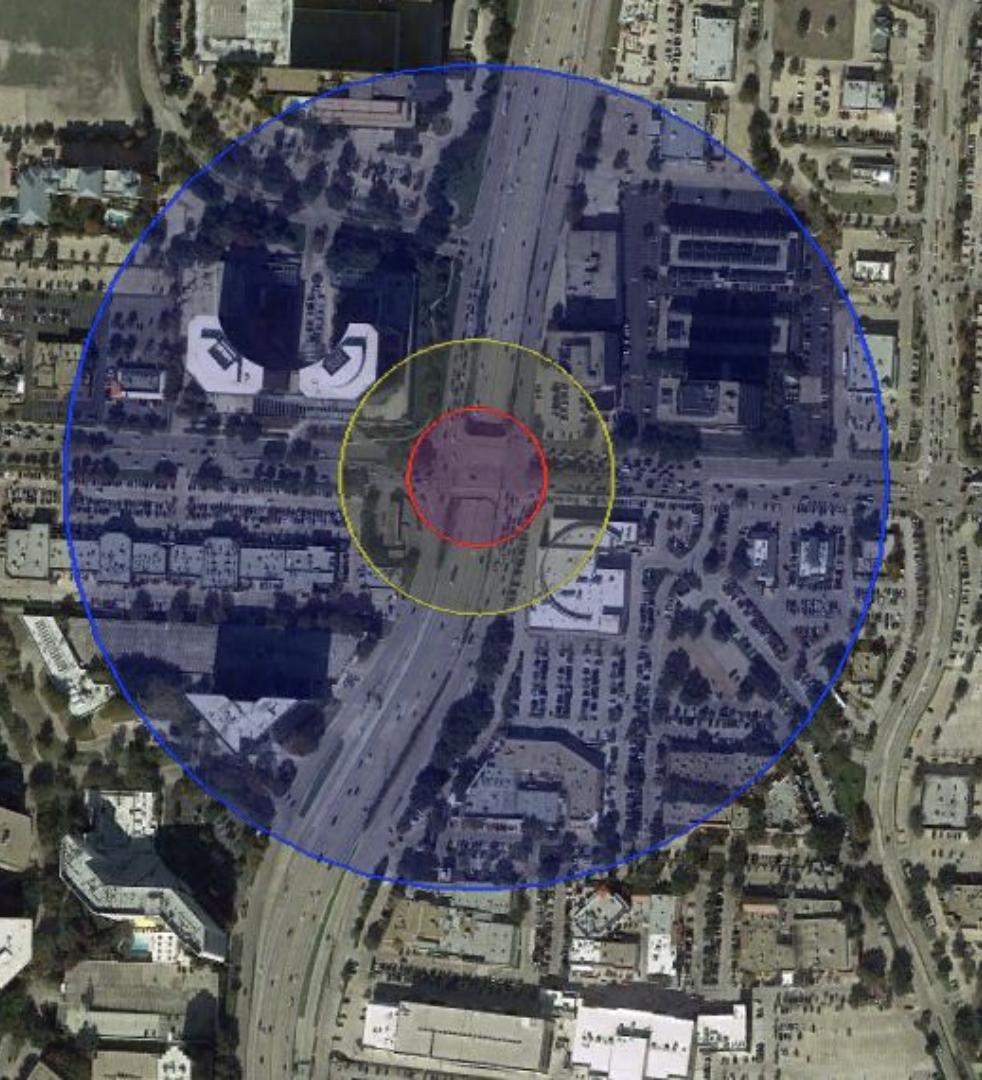


Geo Aggregations

GeoDistance Agg

```
{  
  "aggs" : {  
    "sf_rings" : {  
      "geo_distance" : {  
        "field" : "location",  
        "origin" : [32.95,  
                   -96.82],  
        "ranges" : [  
          { "to" : 50 },  
          { "from" : 50,  
            "to" : 100 },  
          { "from" : 100,  
            "to" : 300 }  
        ]  
      }  
    }  
  }  
}
```

GeoDistance Agg



GeoGrid Agg

```
{  
  "aggs": {  
    "crime_cells": {  
      "geohash_grid": {  
        "field": "location",  
        "precision": 8  
      }  
    }  
  }  
}
```

GeoGrid Agg



GeoCentroid Agg

```
"query" : {  
  "match" : {  
    "crime" : "burglary"  
  }  
},  
"aggs" : {  
  "towns" : {  
    "terms" : { "field" : "town" },  
    "aggs" : {  
      "centroid" : {  
        "geo_centroid" : {  
          "field" : "location"  
        }  
      }  
    }  
  }  
}
```

GeoCentroid Agg



GeoCentroid Agg



Geo Aggregations

more available, and coming soon...

- matrix_stats - (Matrix Aggs) plugin
 - kurtosis/skewness
 - variance-covariance matrix
 - pearson's product correlation matrix
- geo_stats - Future?
 - Moran's I - measuring spatial auto-correlation
 - Getis-Ord - spatial hot spot analysis

Kibana / Elastic Maps Service

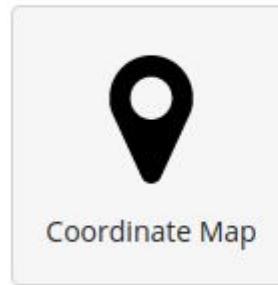
Kibana Visualizations

Out-of-the-box visualizations for geodata in Elasticsearch

2 types

- Coordinate Maps
- Region Maps

Maps



Coordinate Map

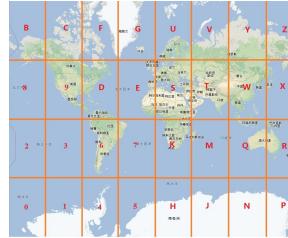


Region Map

Visualize is built on top of the Elasticsearch aggregations

Coordinate Map Visualization

Shows result of **geohash_grid** aggregations.



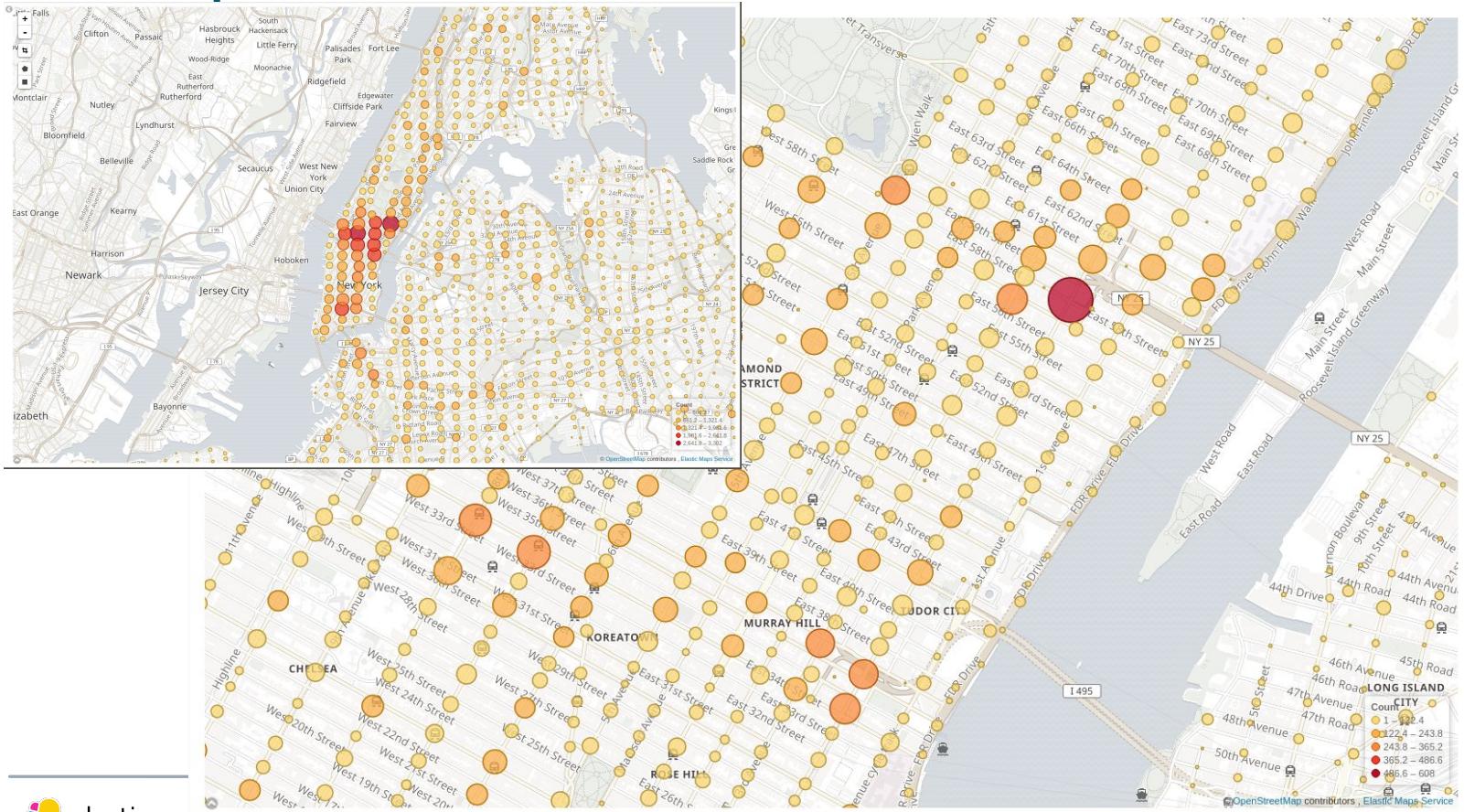
Shows summary of all documents that belong to a single cell.

Put location of “summarized” point in the “geo-centroid” (weighted middle). This gives a better approximate location.

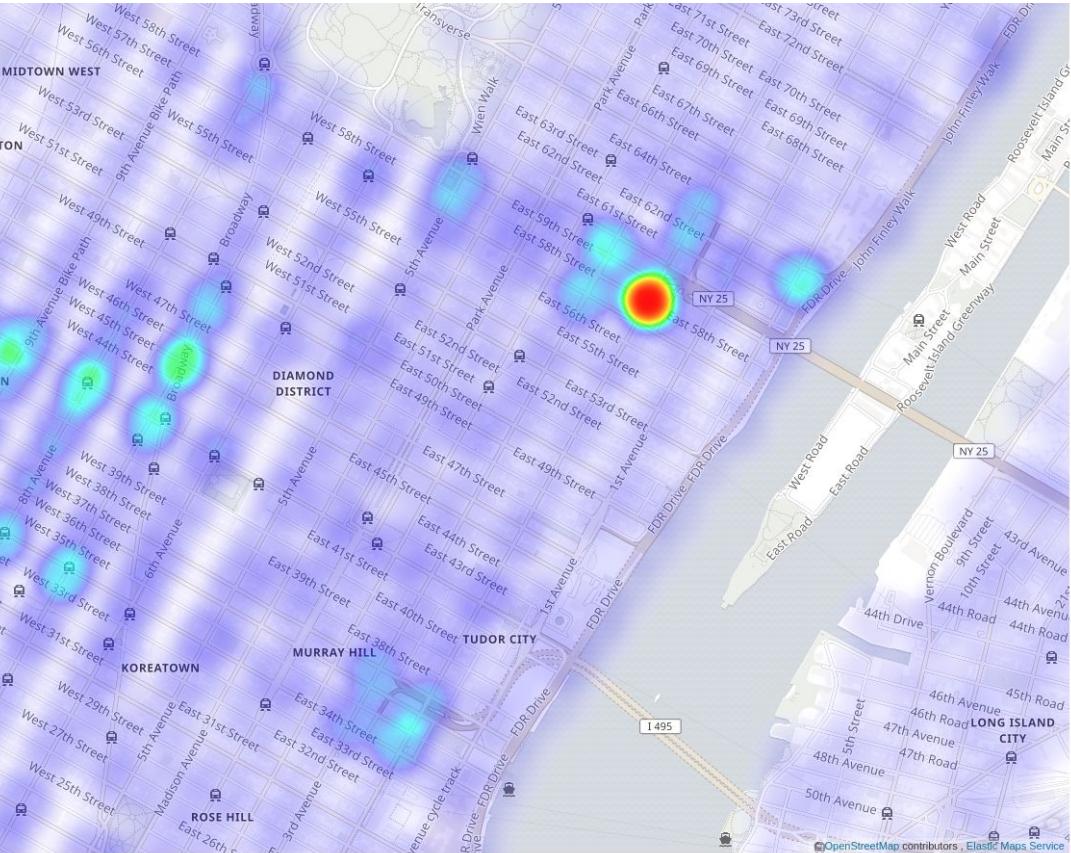
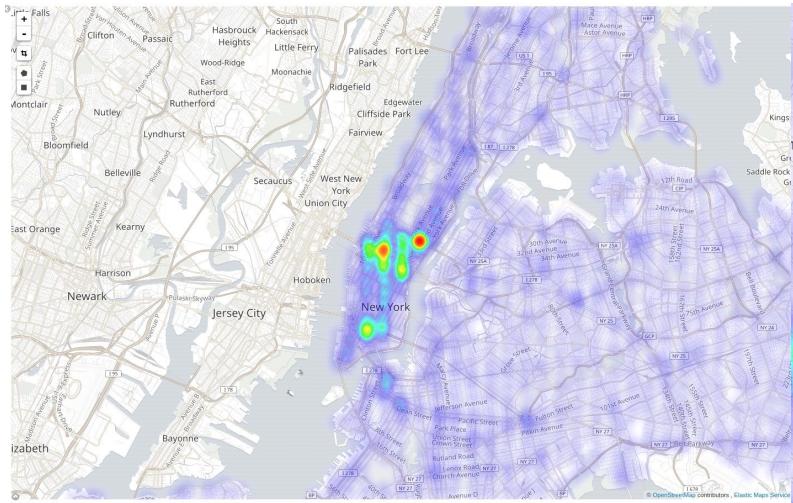
The more zoomed in, the more precise the location.

Different marker-styles (bubbles, heatmap)

Example 1



Example 2



Region Maps

“Choropleth maps”

Thematic maps: color intensity correspond to magnitude of metric

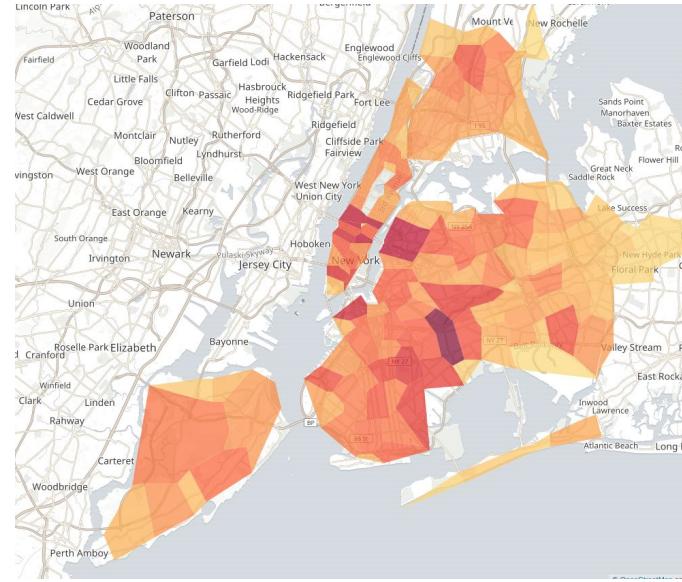
Shows result of **terms aggregations**.

“Client-side” join between the result of term aggregation and a reference shape layer.

- Polygons/Multipolygons (simple feature)
- Documents in elasticsearch need to have field that matches a property of the layer

Region Maps

Request traffic

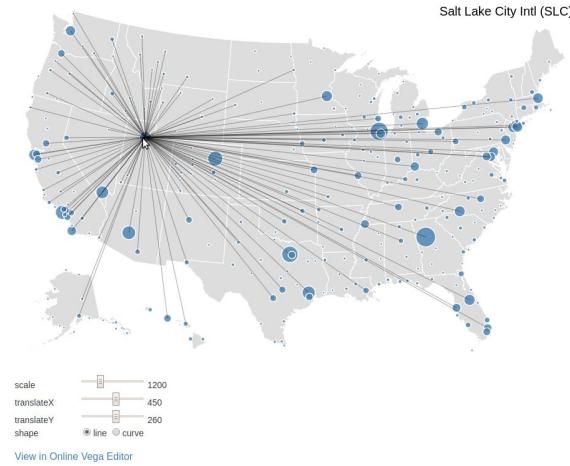


Vega

Experimental feature

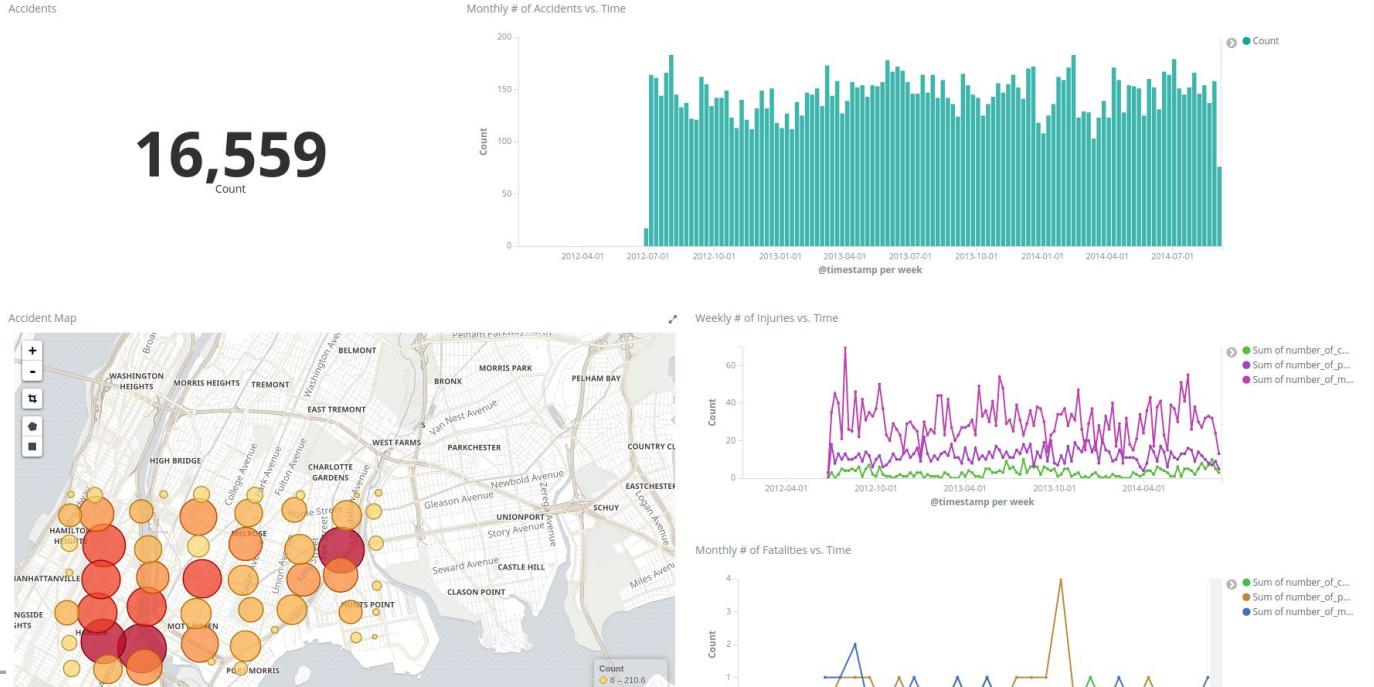
Vega/VegaLite is a domain language in JSON to create visualizations.

Vega has support for geographic projection.



Dashboard integration

- Use map for spatial filtering of data ...
- ... and have other filters applied to your map



Elastic Maps Service

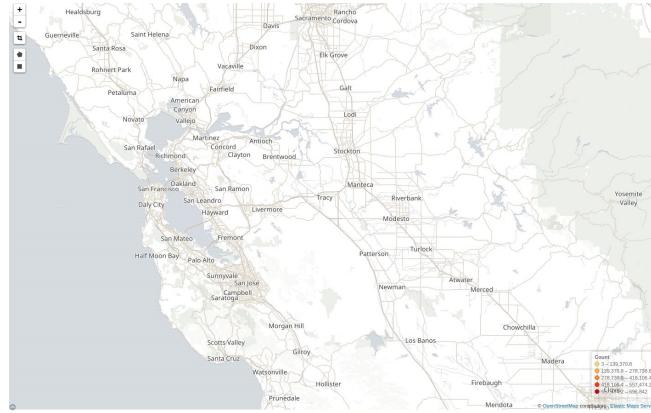
Elastic Maps Service

Reference basemapping and reference data service hosted by Elastic.

“Getting started” experience for mapping.

(1) World base map

- Base for Coordinate Map, Region Map



(2) Shape layers

- World countries, US States, Germany States, Canada Provinces, USA zip-codes
- Number of identifier fields (name in one or more languages, and ISO-identifiers)

Integrating Custom Maps

Custom base maps

- (1) Configure global base-map in kibana.yml by using Tile Map Service URL

tilemap.url: https://tiles.elastic.co/v2/default/{z}/{x}/{y}

- (2) Configure visualization-specific base-map using WMS (web map service)

- Requires 3rd party geo-service
 - Geoserverb
 - ArcGIS Server
 - MapServer
 -

WMS is an OGC standard for map image services. For more information, go [here](#).

WMS url*

WMS layers*

WMS version*

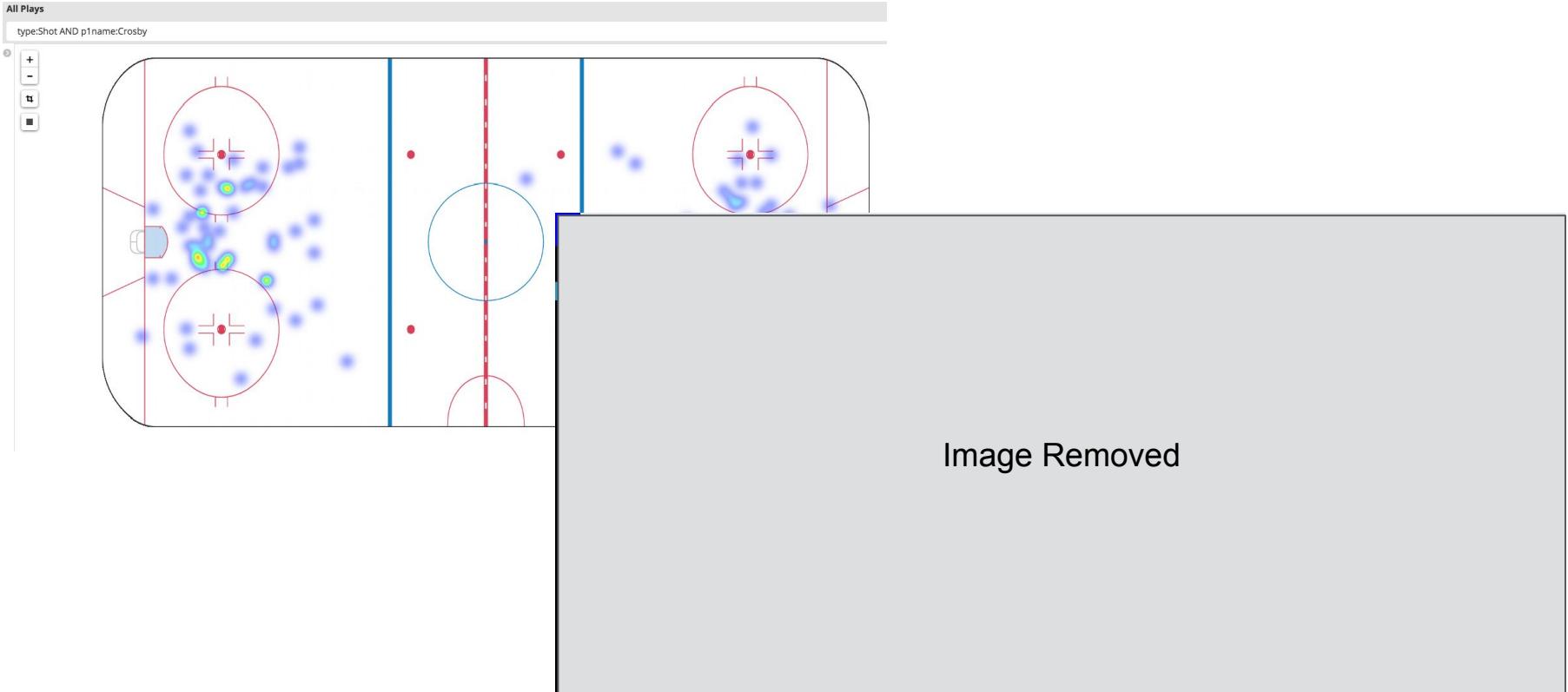
WMS format*

WMS attribution

WMS styles*

* if this parameter is incorrect, maps will fail to load.

Custom maps examples



Custom shape layers

- geojson/topojson
- Configure in kibana.yml -> available in region maps UI

```
regionmap:  
  includeElasticMapsService: false  
  layers:  
    - name: "Departments of France"  
      url: "http://my.cors.enabled.server.org/france_departements.geojson"  
      attribution: "INRAP"  
      fields:  
        - name: "department"  
          description: "Full department name"  
        - name: "INSEE"  
          description: "INSEE numeric identifier"
```

- Use any web-server
 - Make sure is CORS enabled so Kibana can download the data (!)

Useful blog posts

- customization
 - <https://www.elastic.co/blog/kibana-and-a-custom-tile-server-for-nhl-data>
 - <https://www.elastic.co/blog/custom-region-maps-in-kibana-6-0>

Questions?





Thank You

- **Web** : www.elastic.co
- **Products** : <https://www.elastic.co/products>
- **Forums** : <https://discuss.elastic.co/>
- **Community** : <https://www.elastic.co/community/meetups>
- **Twitter** : @elastic

