

What's new in the Elastic Stack?

Alexander Reelsen
alex@elastic.co
@spinscale

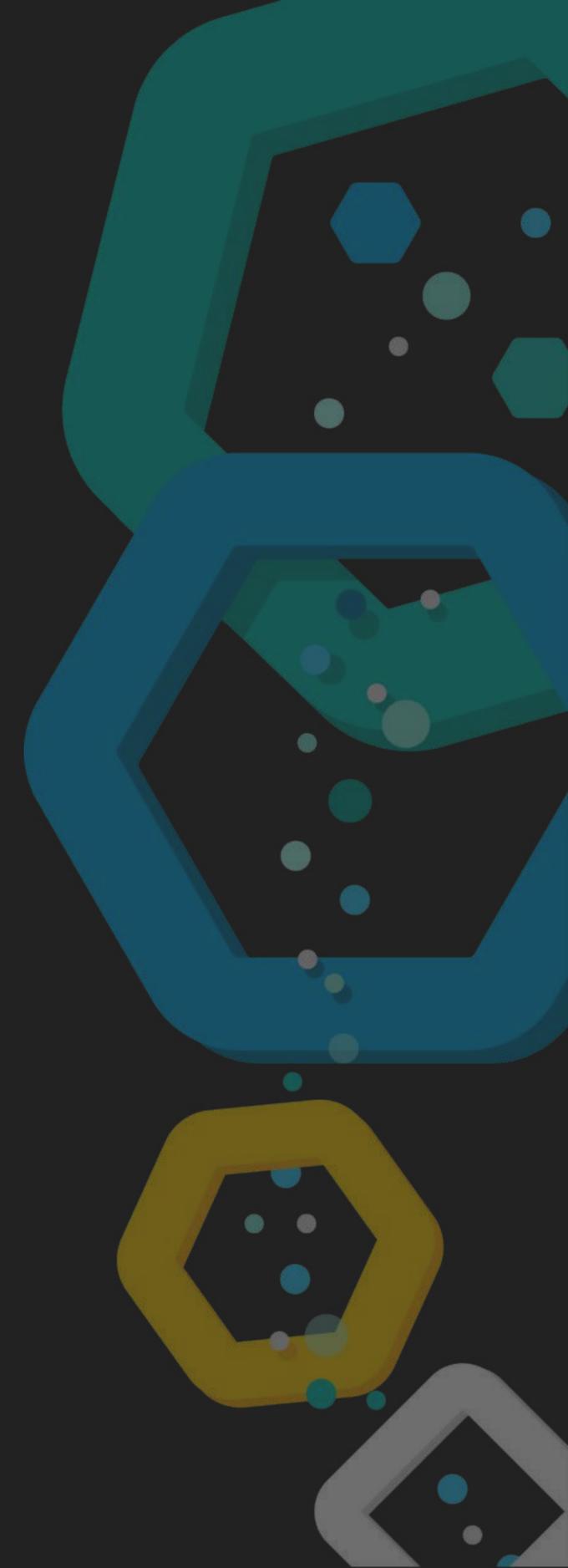


Agenda

- ▶ What's new in 6.x?
- ▶ What's new in 7.x?
- ▶ Q & A



What's new in 6.x?



Elasticsearch 6.x

6.0

- Zero downtime upgrades
- Cross cluster search
- Sequence id based recoveries
- Index sorting
- range based datatypes

6.1

- Index splitting

6.2

- Rank evaluation API

6.3

- Rollup
- Java 10 support

6.4

- Reloadable secure settings
- Field Aliases
- Korean analyzer

6.5

- G1GC support, Java 11
- Minimal snapshots (50% less)

6.6

- Frozen indices
- BKD backed geoshapes

6.7

- CCR
- SQL
- ILM
- Upgrade Assistant

6.8

- ECK (Elastic for Kubernetes)
- Move security features into basic

Elasticsearch 6.7 - CCR

- 🌸 Cross Cluster Replication + UI
- 🌸 Replicate data across data centers
- 🌸 Leader index requires soft deletes to be set
- 🌸 Follower index configures cluster and leader index
- 🌸 Follower index can also be a pattern

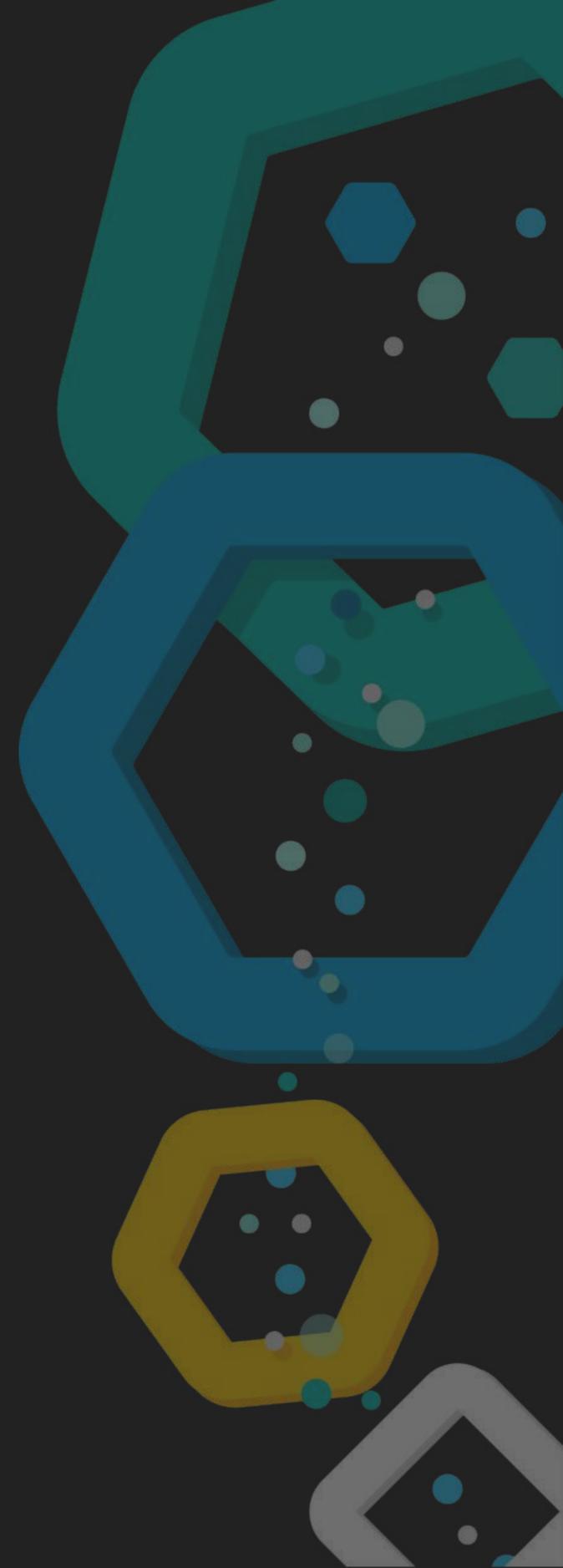
Elasticsearch 6.7 - SQL

 REST API

 CLI

 JDBC

 ODBC



Elasticsearch 6.7 - ILM

```
PUT _ilm/policy/full_policy
{
  "policy": {
    "phases": {
      "hot": {
        "actions": {
          "rollover": {
            "max_age": "7d",
            "max_size": "50G"
          }
        }
      },
      "warm": {
        "min_age": "30d",
        "actions": {
          "forcemerge": {
            "max_num_segments": 1
          },
          "shrink": {
            "number_of_shards": 1
          },
          "allocate": {
            "number_of_replicas": 2
          }
        }
      }
    }
  },

```

```
    "cold": {
      "min_age": "60d",
      "actions": {
        "allocate": {
          "require": {
            "type": "cold"
          }
        }
      }
    },
    "delete": {
      "min_age": "90d",
      "actions": {
        "delete": {}
      }
    }
  }
}
```

Kibana 6.7

 Maps

 Uptime

 Canvas

 Infrastructure is GA

 Logs is GA

 Localization

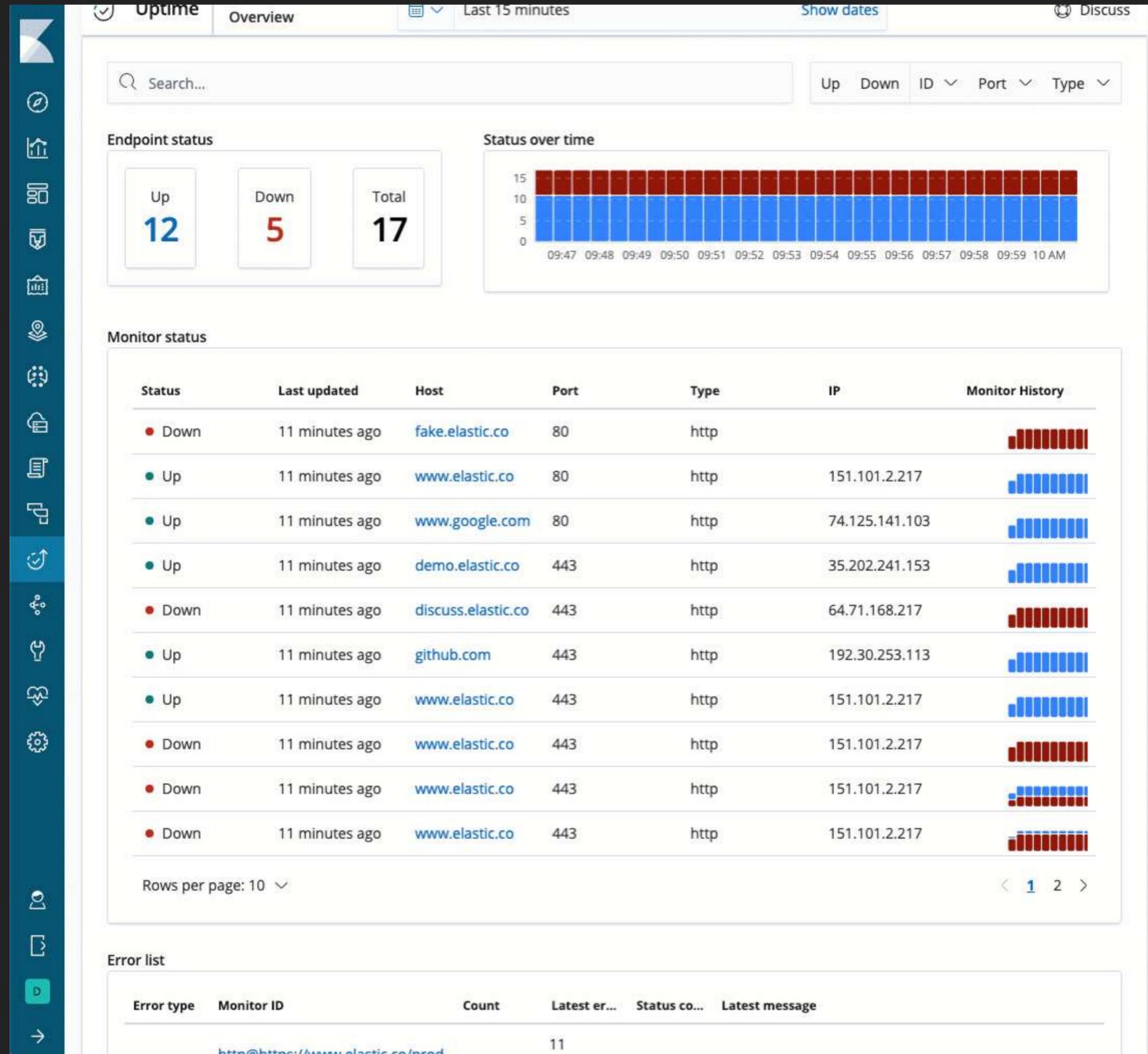
 Upgrade Assistant



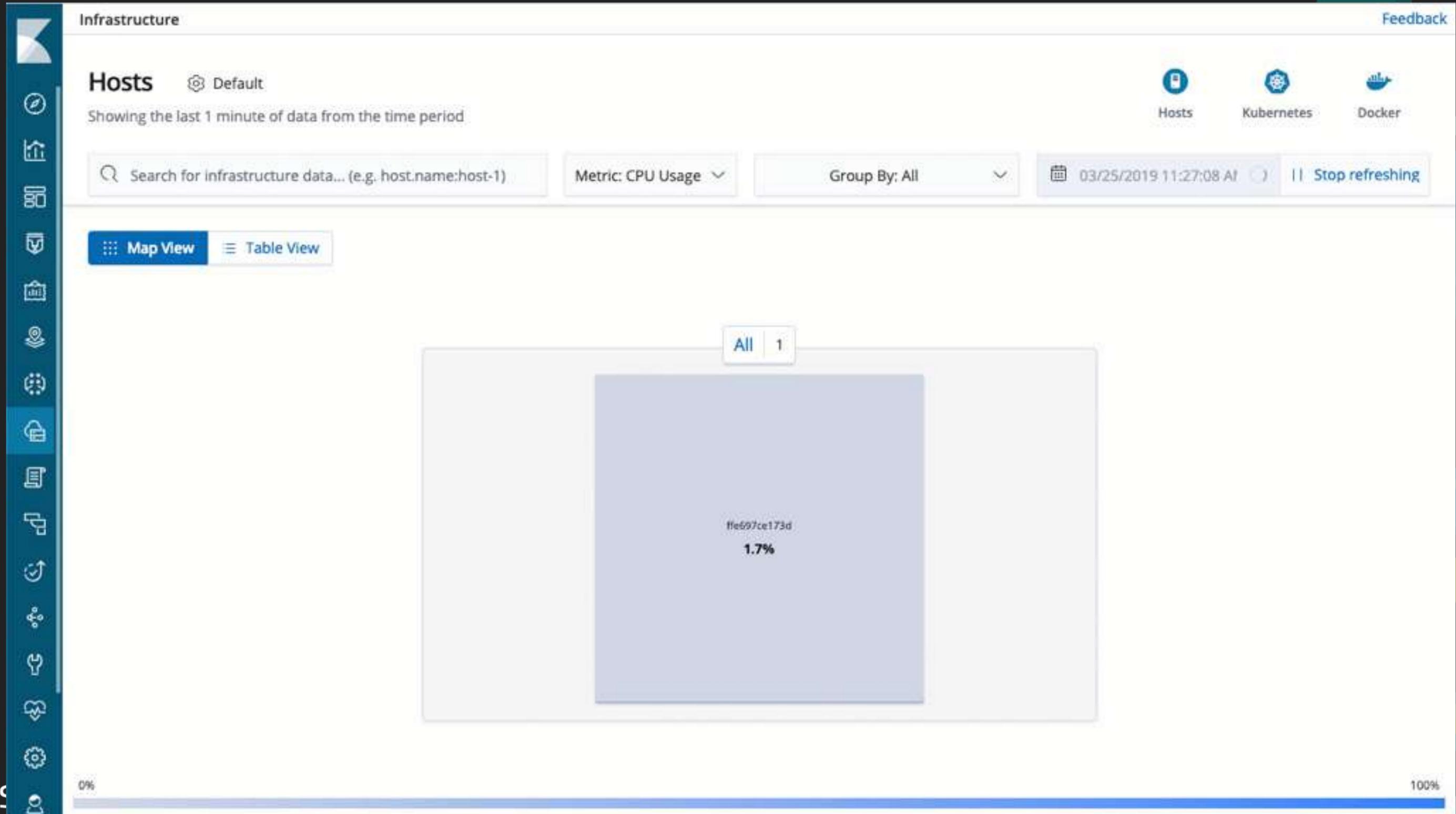
Kibana 6.7 - Maps



Kibana 6.7 - Uptime

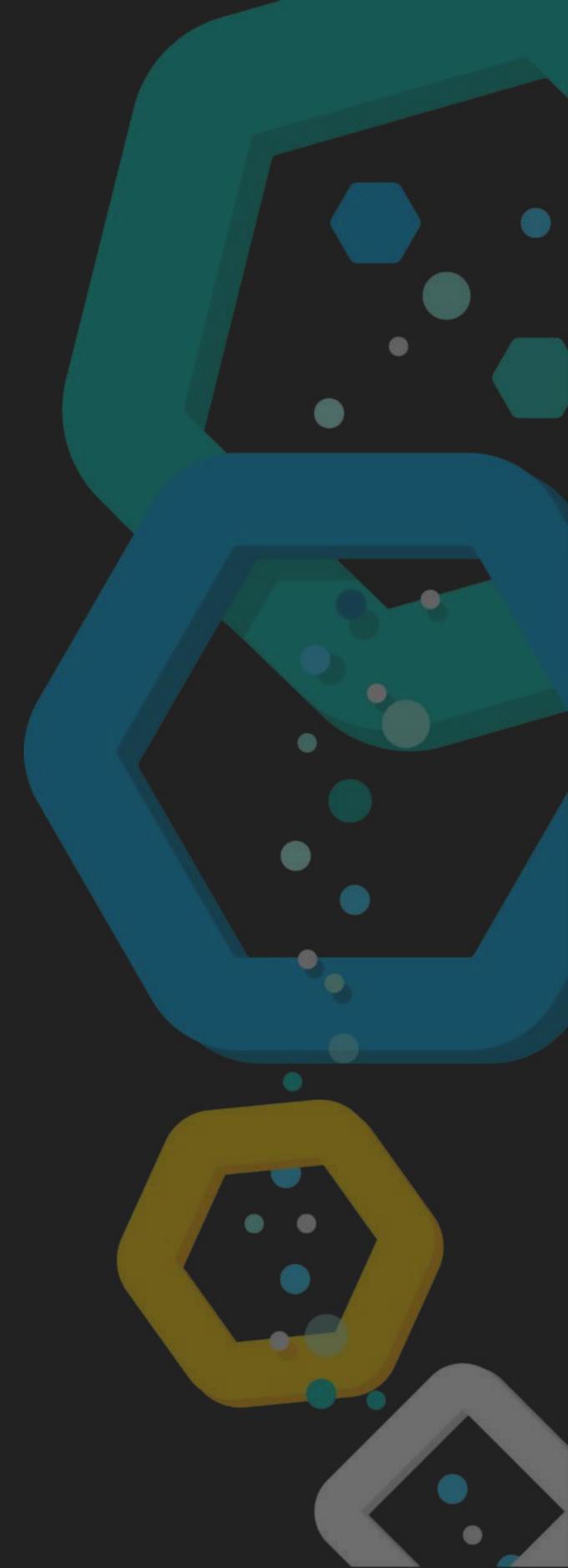


Kibana 6.7 - Infrastructure UI

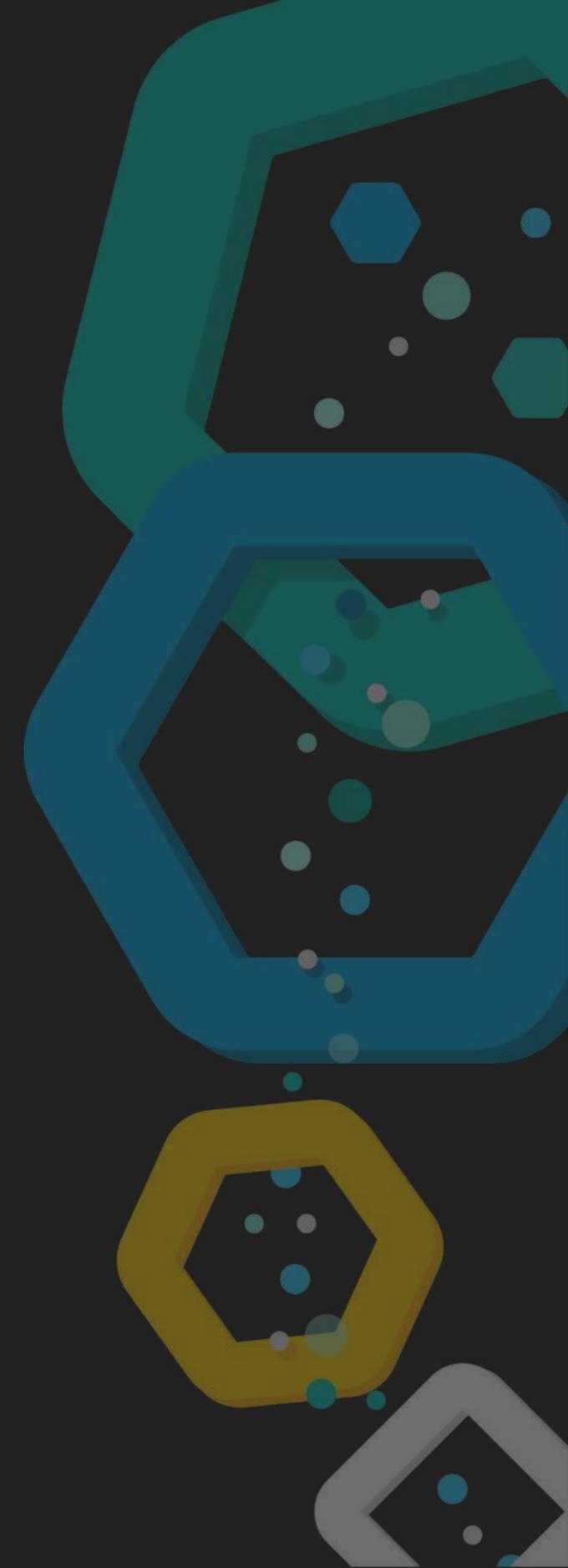


Elasticsearch 6.8 - Security & ECK

- 🌸 Native & file realm now free
- 🌸 TLS now free
- 🌸 Elastic Cloud on Kubernetes (Operator)



What's new in 7.0?



Kibana 7.0

- 🌸 Elastic UI Library
- 🌸 KQL by default (+ autocomplete)
- 🌸 Responsive dashboards



Kibana 7.0

Home / Add data

Add Data to Kibana

All Logging Metrics Security analytics **Sample data**

INSTALLED

Sample eCommerce Data

This dashboard contains sample data for you to play with. You can view it, search it, and interact with the visualizations. For more information about Kibana, check out [Kibana 7.0](#).

Advanced Order Products per Day: **Trans / day 139**

Advanced Order by Brand: **\$75.23 Average Order**

Advanced Order by Price: **2,163 Average Price**

Advanced Order by Status: **\$77,638.33**

Sample eCommerce orders

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

[Remove](#) [View data](#)

INSTALLED

Sample Flight data

This dashboard contains sample data for you to play with. You can view it, search it, and interact with the visualizations. For more information about Kibana, check out [Kibana 7.0](#).

Flight by Status: **313 Total Flights**

Flight by Price: **\$596 Avg. Total Price**

Flight by Route: **68 Total Routes**

Sample flight data

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

[Remove](#) [View data](#)

INSTALLED

Sample Logs Data

This dashboard contains sample data for you to play with. You can view it, search it, and interact with the visualizations. For more information about Kibana, check out [Kibana 7.0](#).

Log by Status: **801 Total Errors**

Log by Type: **Log Total Price**

Sample web logs

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

[Remove](#) [View data](#)

Kibana 7.0

Discover - Kibana

localhost:5601/app/kibana#/discover?_g=(filters:!(),refreshInterval:(pause:!t,value:0),time:(from:now%2Fw,to:now%2Fw))&a=(columns:!(_source),filters:!(),index:ff959d...

Discover

1,043 hits

New Save Open Share Inspect

Filters > Search... (e.g. status:200 AND extension:PHP) Options This week Show dates Refresh

+ Add filter

kibana_sample_data_...

Feb 10, 2019 @ 00:00:00.000 - Feb 16, 2019 @ 23:59:59.999 — Millisecond Scaled to hour

Count

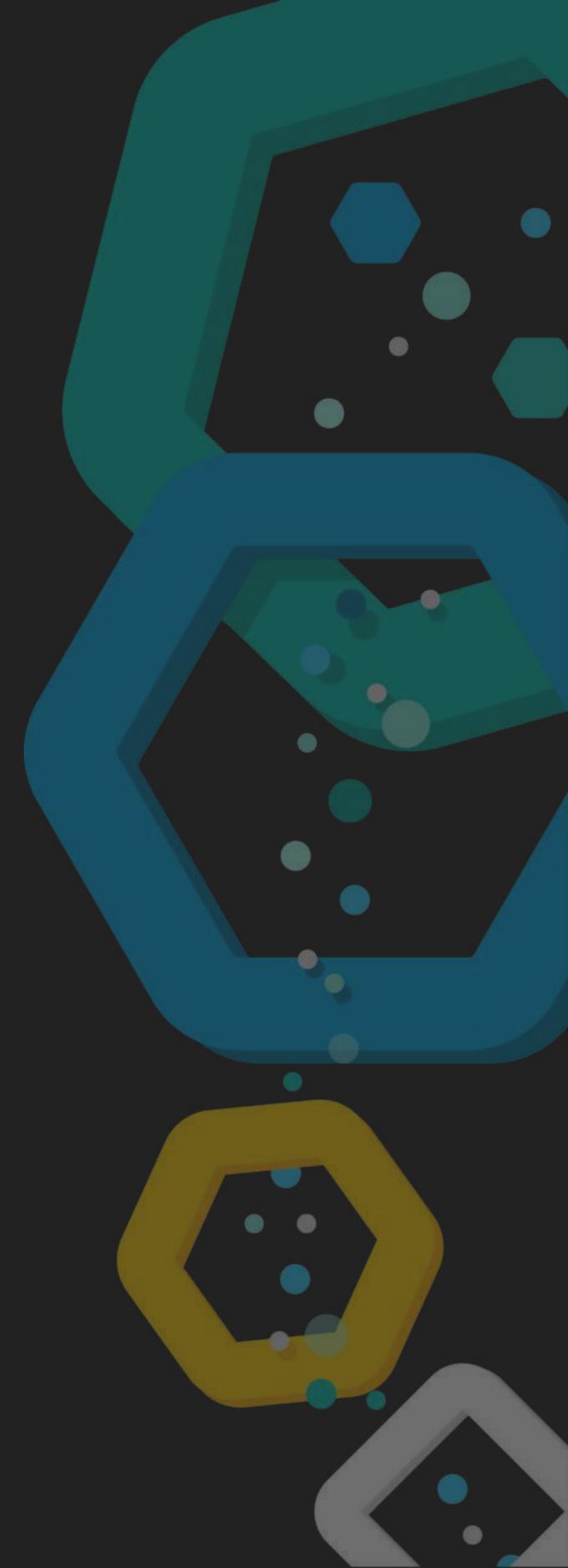
order_date per hour

Time	_source
> Feb 16, 2019 @ 23:58:05.000	category: Women's Clothing currency: EUR customer_first_name: Stephanie customer_full_name: Stephanie Figueroa customer_gender: FEMALE customer_id: 6 customer_last_name: Figueroa customer_phone: day_of_week: Sunday day_of_week_i: 6 email: stephanie@figueroa-family.zzz manufacturer: Pyramidustries active, Sphererecords order_date: Feb 16, 2019 @ 23:58:05.000 order_id: 573826 products: { "base_price": 10.99, "discount_percentage": 0, "quantity": 1, "manufacturer": "Pyramidustries active", "tax_amount": 0, "product_id": 21654, "category": "Women's Clothing", "sku":
> Feb 16, 2019 @ 23:52:19.000	category: Women's Clothing currency: EUR customer_first_name: Diane customer_full_name: Diane Parker customer_gender: FEMALE customer_id: 22 customer_last_name: Parker customer_phone: day_of_week: Sunday day_of_week_i: 6 email: diane@parker-family.zzz manufacturer: Pyramidustries, Oceanavigations order_date: Feb 16, 2019 @ 23:52:19.000 order_id: 573821 products: { "base_price": 20.99, "discount_percentage": 0, "quantity": 1, "manufacturer": "Pyramidustries", "tax_amount": 0, "product_id": 17147, "category": "Women's Clothing", "sku":
> Feb 16, 2019 @ 23:45:07.000	category: Men's Clothing currency: EUR customer_first_name: Abdulraheem Al customer_full_name: Abdulraheem Al Potter customer_gender: MALE customer_id: 33 customer_last_name: Potter customer_phone: day_of_week: Sunday day_of_week_i: 6 email: abdulraheem al@potter-family.zzz manufacturer: Microlutions, Low Tide Media order_date: Feb 16, 2019 @ 23:45:07.000 order_id: 573814 products: { "base_price": 14.99, "discount_percentage": 0, "quantity": 1,



Ingest 7.0

- 🌸 Beats: ECS/ILM integration
- 🌸 Filebeat: zeek, santa, netflow support, encodings
- 🌸 Auditbeat: system module
- 🌸 Metricbeat
 - 🌸 Elasticsearch, Logstash & Kibana modules
 - 🌸 NATS, MSSQL, EC2, CouchDB
- 🌸 Logstash
 - 🌸 Native Java Plugins
 - 🌸 Java execution engine on by default



Stack 7.0

 ECS

 ES-Hadoop

 Kerberos Integration

 Java 8 required

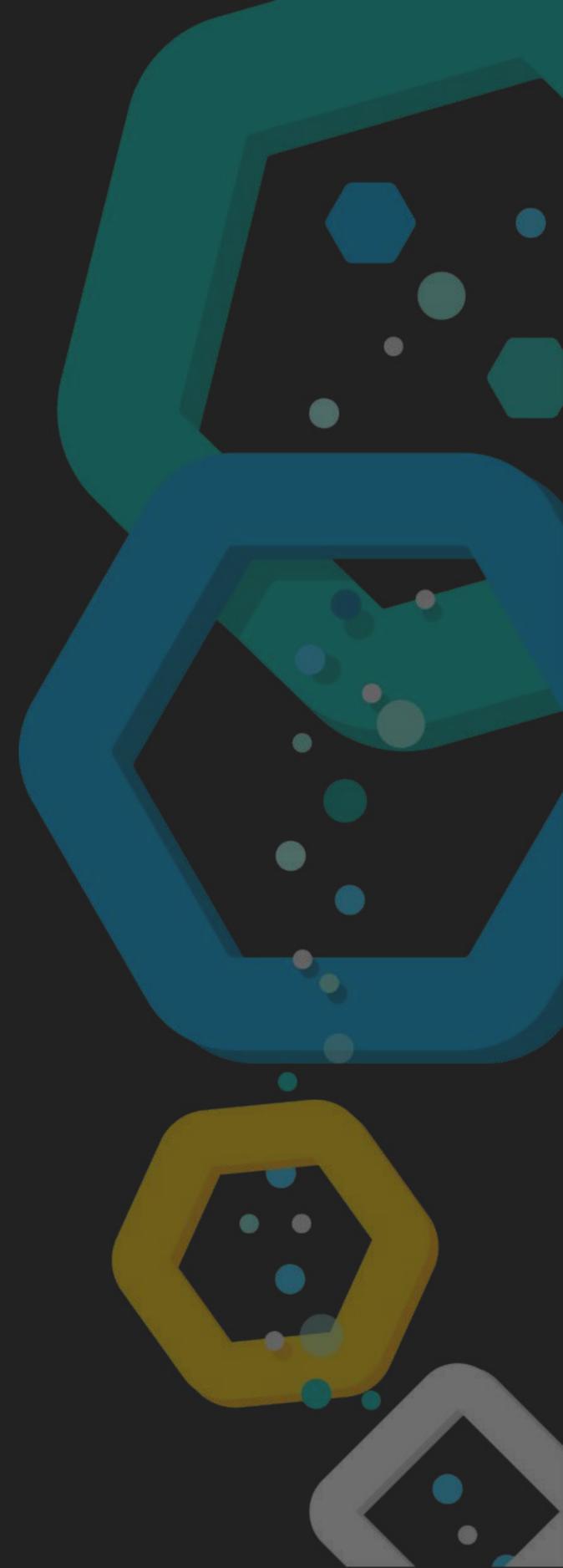
 Cascading support removed

 Clients

 Rewritten JavaScript client

 New Go Client

 Java: High Level REST Client



Elasticsearch 7.0

🌸 Mapping types

- 🌸 date_nanos
- 🌸 rank_feature
- 🌸 rank_features
- 🌸 dense_vector
- 🌸 sparse_vector

🌸 Queries

- 🌸 intervals query
- 🌸 script_score query (supercedes function_score)
- 🌸 rank_feature query

🌸 Searches

- 🌸 faster top-k retrieval
- 🌸 adaptive replica selection enabled by default
- 🌸 No refresh on idle shards (faster indexing)

🌸 Others

- 🌸 Rewritten cluster coordination
- 🌸 Lucene 8
- 🌸 High Level REST client
- 🌸 Docker part of the build
- 🌸 Single shard index by default
- 🌸 Rewritten memory circuit breaker
- 🌸 Type is optional now
- 🌸 TLS 1.3
- 🌸 Ships with OpenJDK

Elasticsearch 7.0 - Rewritten cluster coordination

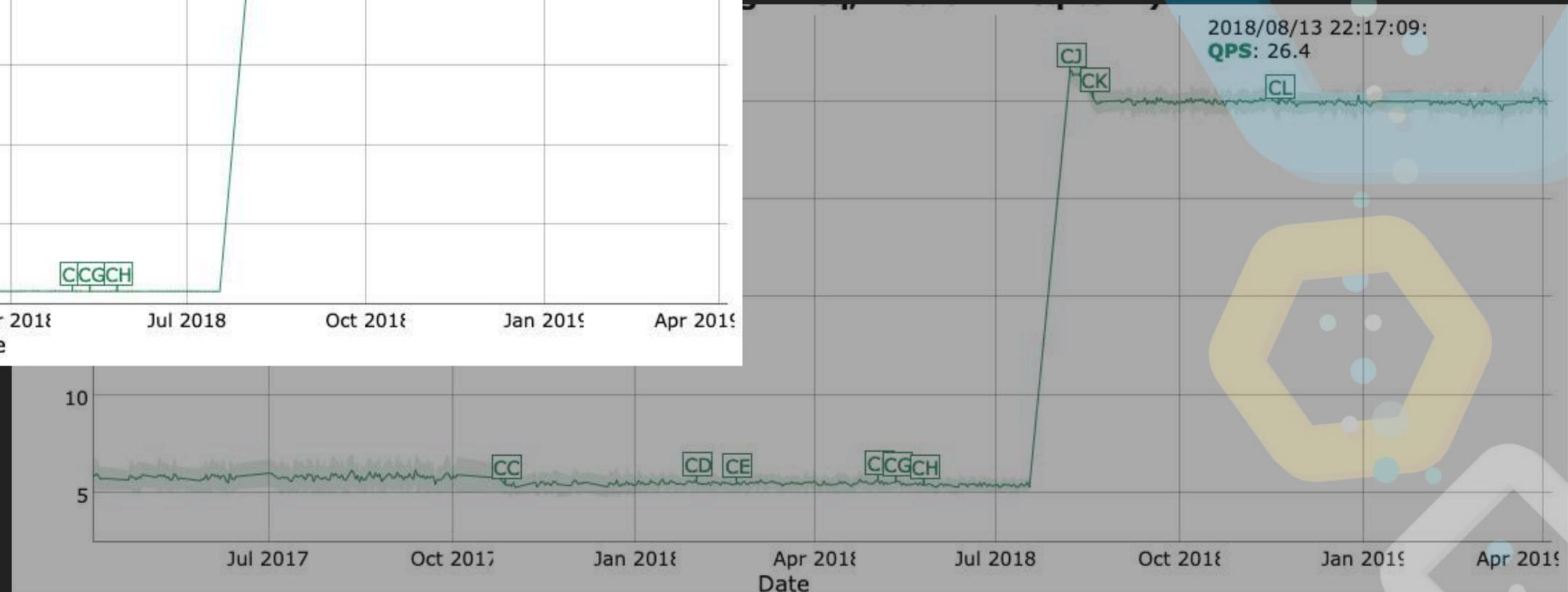
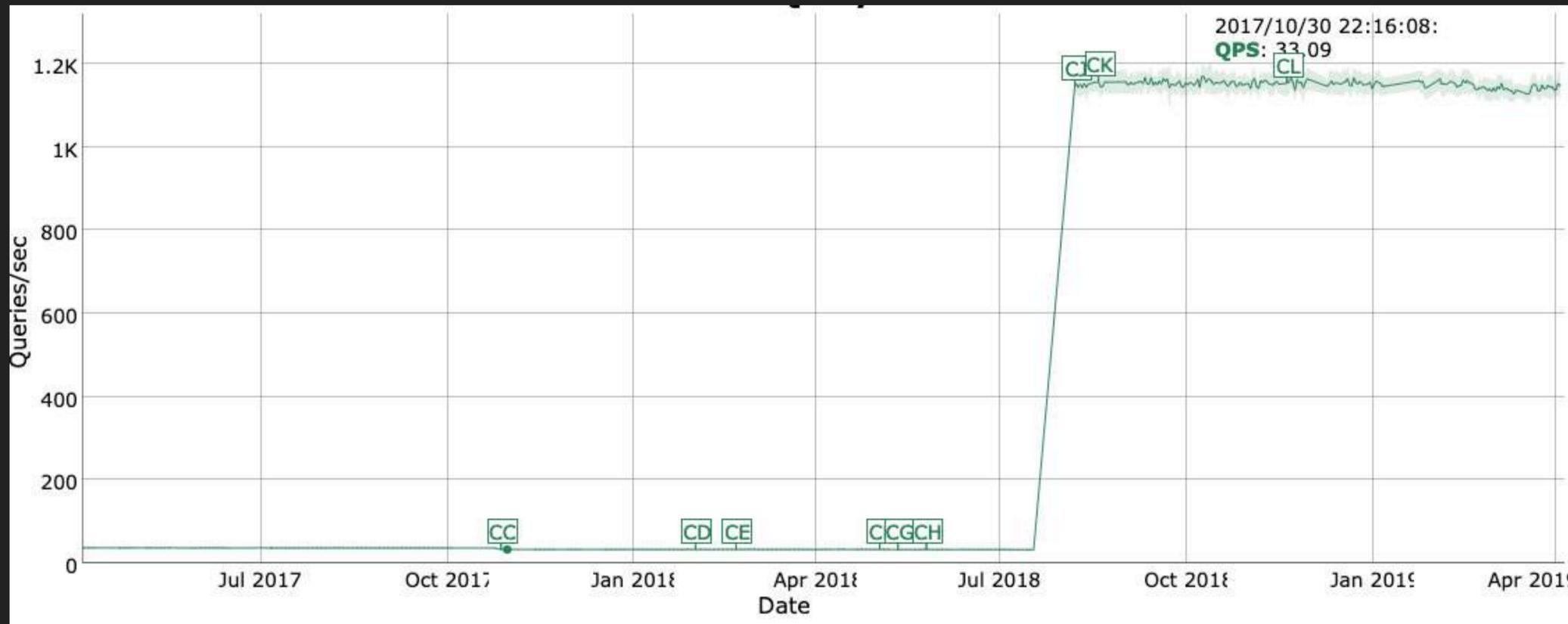
- 🌸 Gone: `discovery.zen.minimum_master_nodes`
- 🌸 Sub-second master election
- 🌸 Simplifying growing/shrinking of cluster
- 🌸 Cluster bootstrapping/Voting configuration
- 🌸 Rolling upgrades from 6 to 7 work
- 🌸 Formal verification via TLA+

Elasticsearch 7.0 - Faster top-k retrieval

- 🌸 While querying, exclude documents that cannot make it into the top hits
- 🌸 Search: Elasticsearch OR Kibana
- 🌸 Term 1: Elasticsearch (max score 5.0)
- 🌸 Term 2: Kibana (max score 3.0)
- 🌸 If first k results all have a score > 3.0 , then documents only containing Kibana can be ignored
- 🌸 Number of potential candidates is reduced while running

Elasticsearch 7.0 - Faster top-k retrieval

- 🌸 Scores may no longer be negative
- 🌸 Total hits are not counted by default



Elasticsearch - Adaptive Replica Selection

- 🌸 Problem: Coordinating node round robins requests between data nodes
- 🌸 Underperforming node harms the whole cluster
- 🌸 Adaptive replica selection
 - 🌸 Response time of previous requests
 - 🌸 Search execution time of the data node
 - 🌸 Queue size of the search threadpool on the data node

Elasticsearch - Rank feature

- 🌸 New `rank_feature` type
- 🌸 New `rank_feature` query
- 🌸 Index numbers that can be used to boost queries
- 🌸 Modifies the scoring formula to in-/decrease score based on the value of the document
- 🌸 Query functions: `saturation`, `logarithm`, `sigmoid`

Elasticsearch - Rank feature

```
PUT test
{
  "mappings": {
    "properties": {
      "pagerank": {
        "type": "rank_feature"
      },
      "url_length": {
        "type": "rank_feature",
        "positive_score_impact": false
      }
    }
  }
}
```

Elasticsearch - Rank feature

```
PUT test/_doc/1
```

```
{  
  "url": "http://en.wikipedia.org/wiki/2016_Summer_Olympics",  
  "content": "Rio 2016",  
  "pagerank": 50.3,  
  "url_length": 42,  
}
```

```
PUT test/_doc/2
```

```
{  
  "url": "http://en.wikipedia.org/wiki/2016_Brazilian_Grand_Prix",  
  "content": "Formula One motor race held on 13 November 2016 at the Autódromo José Carlos Pace in São Paulo, Brazil",  
  "pagerank": 50.3,  
  "url_length": 47,  
}
```

```
PUT test/_doc/3
```

```
{  
  "url": "http://en.wikipedia.org/wiki/Deadpool_(film)",  
  "content": "Deadpool is a 2016 American superhero film",  
  "pagerank": 50.3,  
  "url_length": 37,  
}
```

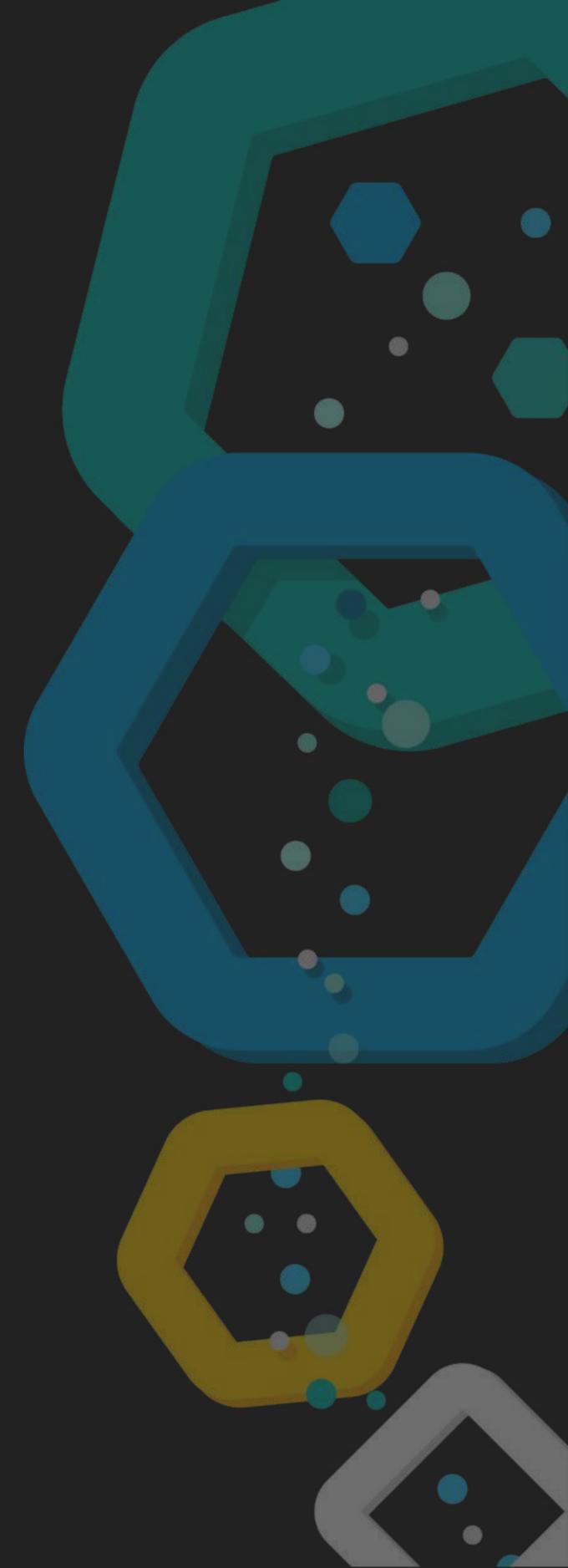
Elasticsearch - Rank feature

```
GET test/_search
```

```
{
  "query": {
    "bool": {
      "must": [ { "match": { "content": "2016" } } ],
      "should": [
        { "rank_feature": { "field": "pagerank" } },
        { "rank_feature": { "field": "url_length", "boost": 0.1 } }
      ]
    }
  }
}
```

Elasticsearch - Rank features

- 🌸 New rank_features type
- 🌸 Key/Value pairs instead of single values



Elasticsearch - Rank feature

```
PUT test/_doc/1
{
  "url": "http://en.wikipedia.org/wiki/2016_Summer_Olympics",
  "content": "Rio 2016",
  "pagerank": 50.3,
  "url_length": 42,
  "topics": {
    "sports": 50,
    "brazil": 30
  }
}
```

Elasticsearch - Rank feature

```
GET test/_search
```

```
{
  "query": {
    "bool": {
      "must": [ { "match": { "content": "2016" } } ],
      "should": [
        { "rank_feature": { "field": "pagerank" } },
        { "rank_feature": { "field": "url_length", "boost": 0.1 } },
        { "rank_feature": { "field": "topics.sports", "boost": 0.4 } }
      ]
    }
  }
}
```

Elasticsearch - Rank feature

```
GET test/_search
```

```
{  
  "query": {  
    "rank_feature": {  
      "field": "pagerank",  
      "saturation": {  
        "pivot": 8  
      }  
    }  
  }  
}
```

Elasticsearch - Rank feature Limitations

- 🌸 Field values must be single-valued and positive
- 🌸 rank_feature fields do not support querying, sorting or aggregating
- 🌸 Field values are not exact (relative error of about 0.4%)
- 🌸 Uses top-k faster retrieval mechanism for speed (hit count!)

Elasticsearch - script_score query

- replaces the `function_score` query

- full painless scripting

- predefined functions:

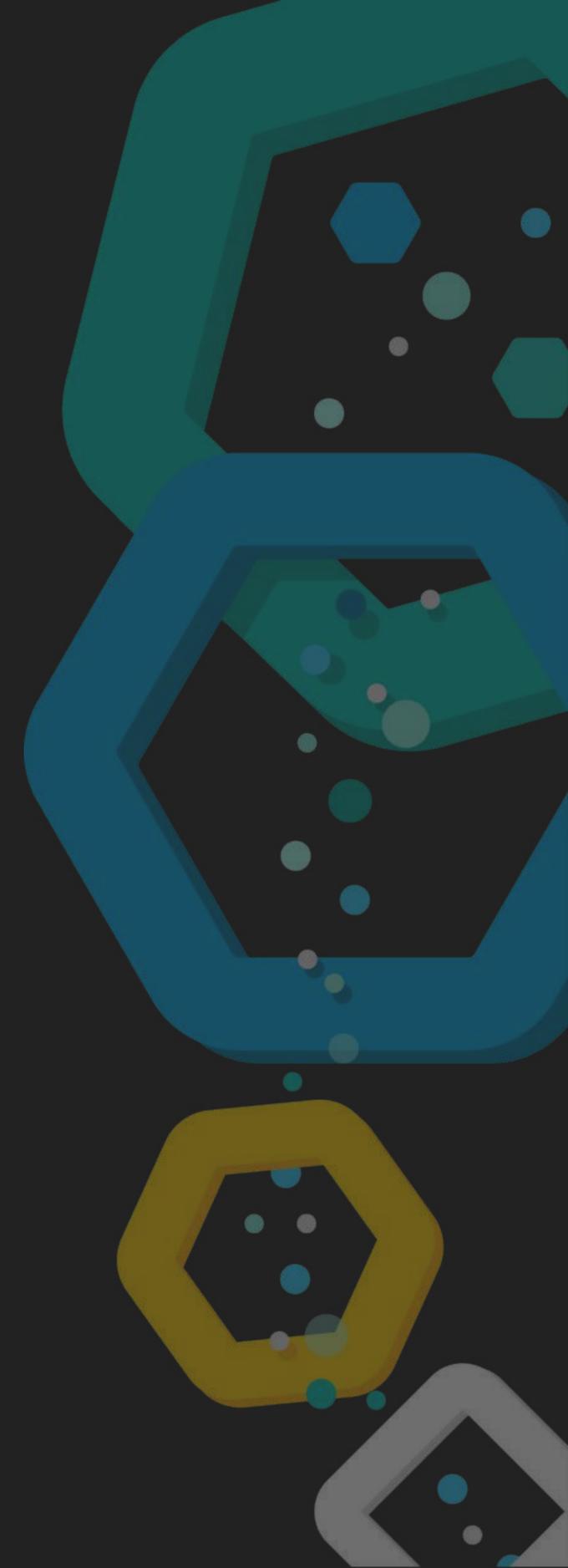
`saturation`, `sigmoid`, `randomScore`,

`decay(Numeric|Date|Geo)(Linear|Exp|Gauss)`

Elasticsearch - Nanosecond support

- 🌸 new datatype: `date_nanos`
- 🌸 stores nanoseconds since the epoch (reduced range!)
- 🌸 internally: moved from Joda-Time to java time
- 🌸 Aggregations: millisecond resolution!
- 🌸 Beware: Upgrade path from 6.x!

What's new in 7.1?



Elasticsearch 7.1

- 🌸 security features moved into basic
- 🌸 ECK (Elastic Cloud on Kubernetes/K8s)

What's new in 7.2?



Elasticsearch 7.2

- 🌸 `search_as_you_type` mapping
- 🌸 `distance_feature` query
- 🌸 Replication of closed indices
- 🌸 `dense/sparse_vector` datatype



Elasticsearch - vector datatypes

- 🌸 `dense_vector`: stores dense vectors of float values, supplied as an array
- 🌸 `sparse_vector`: stores sparse vectors of float values, supplied as map
- 🌸 Use-Case: User centric recommendation based on past decisions

Elasticsearch - vector datatypes

```
PUT my_index
{
  "mappings": {
    "properties": {
      "sparse": {
        "type": "sparse_vector"
      },
      "dense": {
        "type": "dense_vector"
      },
      "my_text" : {
        "type" : "keyword"
      }
    }
  }
}
```

Elasticsearch - vector datatypes

```
PUT my_index/_doc/1
```

```
{  
  "my_text" : "text1",  
  "dense" : [0.5, 10, 6]  
  "sparse" : {"1": 0.5, "5": -0.5, "100": 1}  
}
```

```
PUT my_index/_doc/2
```

```
{  
  "my_text" : "text2",  
  "dense" : [-0.5, 10, 10, 4]  
  "sparse" : {"103": 0.5, "4": -0.5, "5": 1, "11" : 1.2}  
}
```

Elasticsearch - script_score query

🌸 sparse/dense functions:

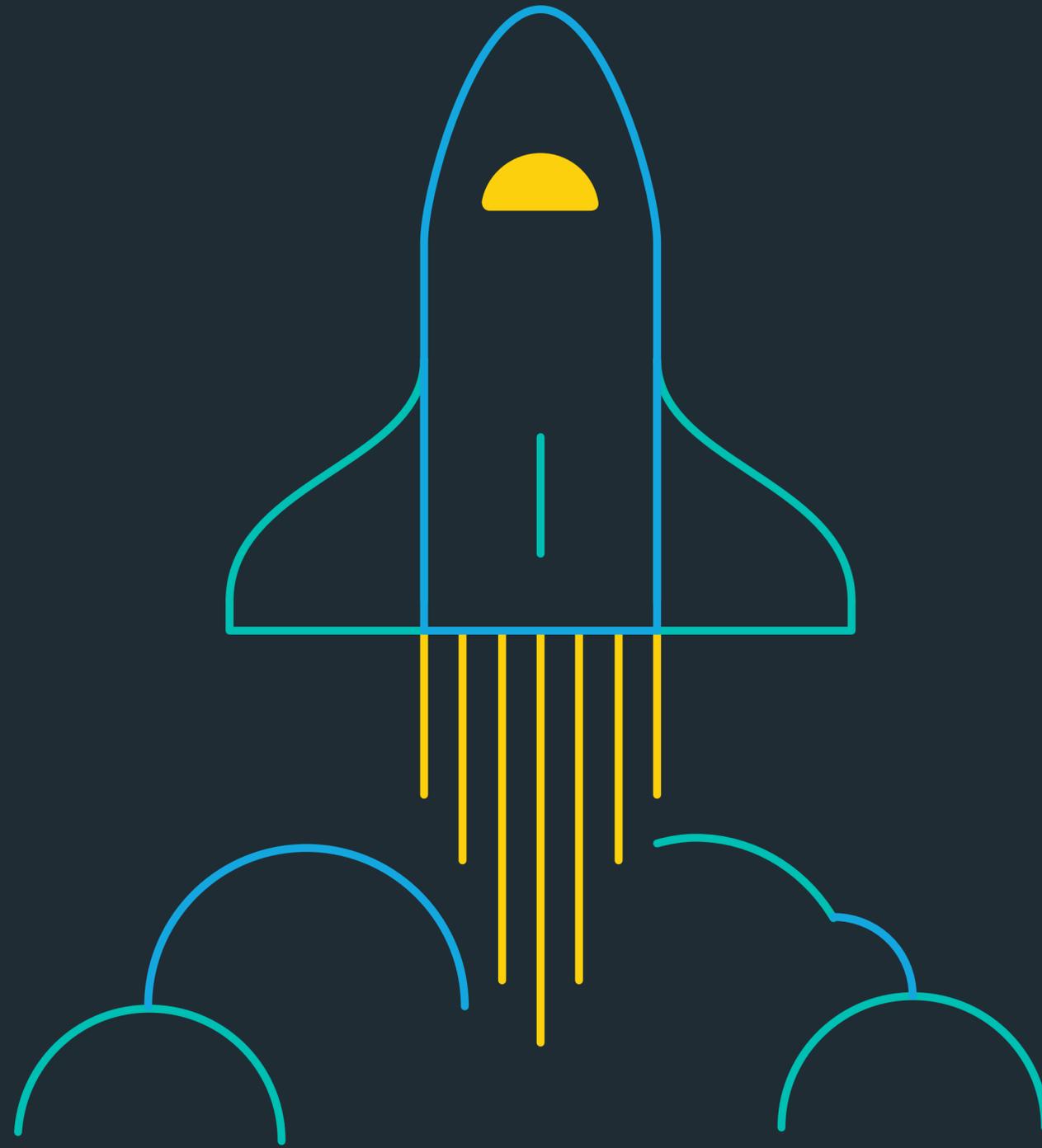
```
cosineSimilarity(Sparse)
```

```
dotProduct(Sparse)
```



Discussion

... ask all the things!



Links

Elasticsearch

-  <https://www.elastic.co/blog/easier-relevance-tuning-elasticsearch-7-0>
-  <https://www.elastic.co/blog/faster-retrieval-of-top-hits-in-elasticsearch-with-block-max-wand>
-  <https://www.elastic.co/blog/creating-frozen-indices-with-the-elasticsearch-freeze-index-api>
-  <https://www.elastic.co/blog/follow-the-leader-an-introduction-to-cross-cluster-replication-in-elasticsearch>
-  <https://www.elastic.co/blog/moving-from-types-to-typeless-apis-in-elasticsearch-7-0>
-  <https://www.elastic.co/blog/improving-node-resiliency-with-the-real-memory-circuit-breaker>
-  <https://www.elastic.co/blog/a-new-era-for-cluster-coordination-in-elasticsearch>
-  <https://www.elastic.co/elasticon/conf/2018/sf/reliable-by-design-applying-formal-methods-to-distributed-systems>
-  <https://github.com/elastic/elasticsearch-formal-models>
-  C3: <https://www.usenix.org/system/files/conference/nsdi15/nsdi15-paper-suresh.pdf>

Beats

-  <https://www.elastic.co/blog/introducing-auditbeat-system-module>

Links

-  <https://www.elastic.co/blog/security-for-elasticsearch-is-now-free>
-  <https://www.elastic.co/blog/introducing-elastic-cloud-on-kubernetes-the-elasticsearch-operator-and-beyond>