

# Introduction into Elasticsearch

## & Spring Data Elasticsearch

Alexander Reelsen

Community Advocate

[alex@elastic.co](mailto:alex@elastic.co) | [@spinscale](https://twitter.com/spinscale)



elastic

# TOC

- Why do you need a search engine in your app?
- Introduction into Elasticsearch
- Introduction into Spring Data Elasticsearch
- Demo
- Running Elasticsearch: Scaling your cluster
- Next steps

# Why do you need a search engine?

... or any data store

# Speed, Scale & Relevance



# Speed

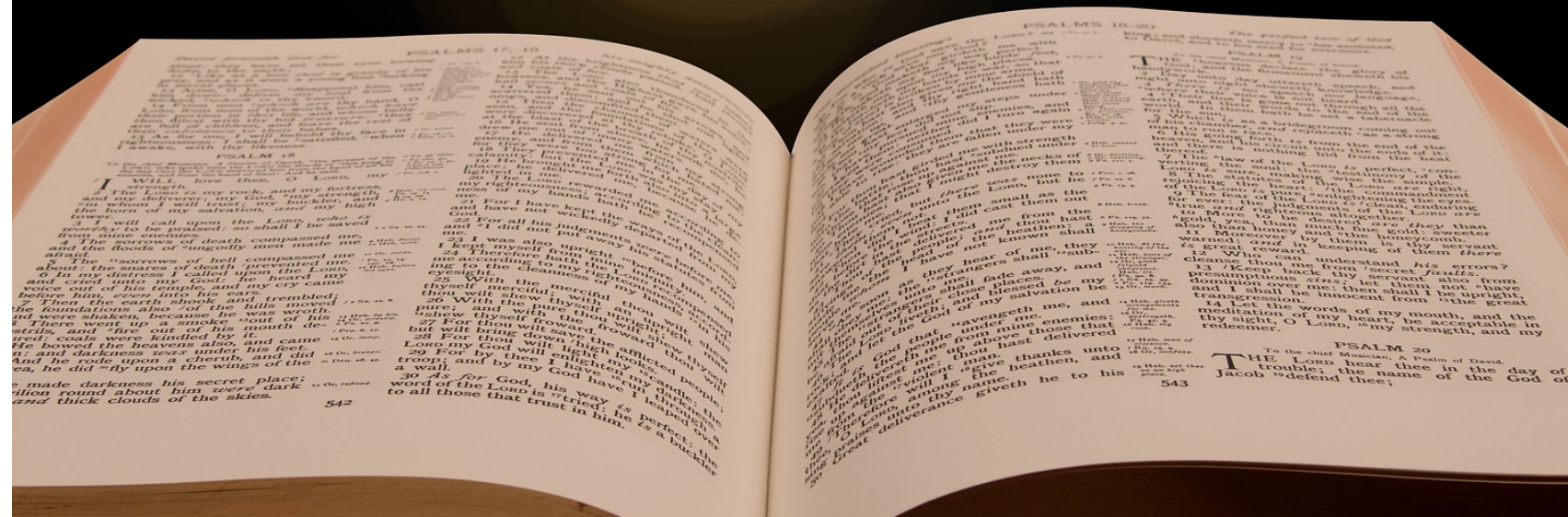




**Scale**



# Relevance

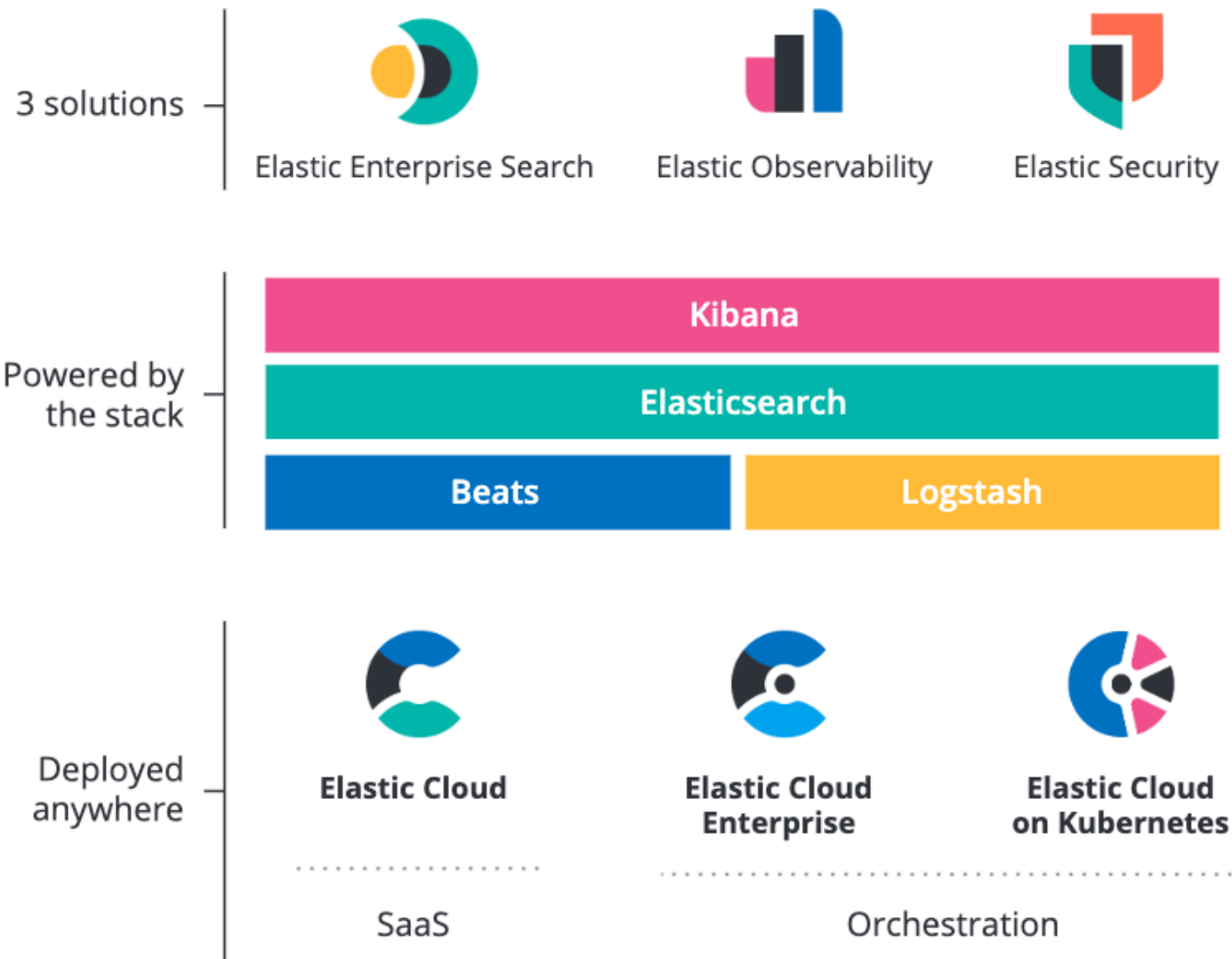


## ... and much more

- NRT: Searching & Indexing
- Read scalability & write scalability
- Resiliency
- Operational simplicity & monitoring capabilities
- Developer experience
- Infrastructure integration
- Team experience
- Use-Cases: Observability, Workplace Search, Security, Product Search, Wikipedia



# Product Overview



# 3 solutions powered by 1 stack

## Solutions



Elastic Enterprise Search



Elastic Observability



Elastic Security



Elastic Stack

# Elastic Stack

building blocks



# Deployment options



**Elastic Cloud**

SaaS



**Elastic Cloud  
Enterprise**

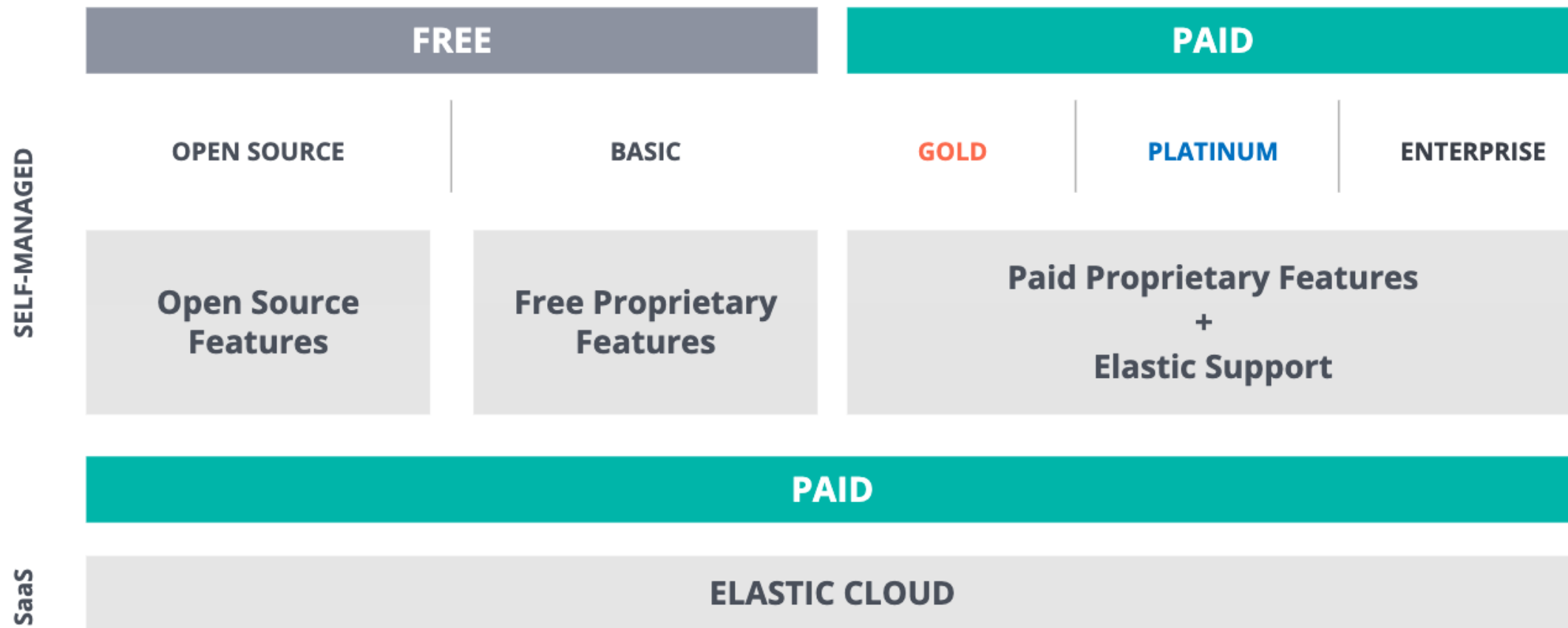


**Elastic Cloud on  
Kubernetes**

Orchestration



# Licensing



# Elastic Stack

building blocks



# Elasticsearch in 10 seconds

- Search Engine (FTS, Analytics, Geo), near real-time
- Distributed, scalable, highly available, resilient
- Interface: HTTP & JSON
- Heart of the Elastic Stack (Kibana, Logstash, Beats)

# Installation & Start

```
# https://www.elastic.co/downloads/elasticsearch
wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.8.0-darwin-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.8.0-linux-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.8.0-windows-x86_64.zip

tar xzf elasticsearch-7.8.0-darwin-x86_64.tar.gz
cd elasticsearch-7.8.0

./bin/elasticsearch
```

```
wget https://artifacts.elastic.co/downloads/kibana/kibana-7.8.0-darwin-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/kibana/kibana-7.8.0-linux-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/kibana/kibana-7.8.0-windows-x86_64.zip

tar xzf kibana-7.8.0-darwin-x86_64.tar.gz
cd kibana-7.8.0
./bin/kibana
```

Point your browser to <http://localhost:5601/>

# Click Dev-Tools

Samples in Kibana

Samples in Github



The image shows a screenshot of the Elastic DevTools interface. On the left is a vertical navigation menu with the following items: Home, Recently viewed, Discover, Visualize, Dashboard, Canvas, Maps, Machine Learning, Metrics, Logs, APM, Uptime, SIEM, Dev Tools (highlighted), Stack Monitoring, and Management. On the right, a card for "Logging" is visible, featuring the Elastic logo and the text: "Ingest logs from popular data sources and easily visualize in preconfigured dashboards." Below the text is a blue button labeled "Add log data".



D

Dev Tools



Console

Search Profiler

Grok Debugger



History

Settings

Help



1 GET /

2

3 GET \_cat/indices



1 # GET /

2 {

3 "name" : "rhincodon",

4 "cluster\_name" : "elasticsearch",

5 "cluster\_uuid" : "fQGQJn\_oQgu5ou0Z9WNDHg",

6 "version" : {

7 "number" : "7.5.0",

8 "build\_flavor" : "default",

9 "build\_type" : "tar",

10 "build\_hash" : "e9ccaed468e2fac2275a3761849cbee64b39519f",

11 "build\_date" : "2019-11-26T01:06:52.518245Z",

12 "build\_snapshot" : false,

13 "lucene\_version" : "8.3.0",

14 "minimum\_wire\_compatibility\_version" : "6.8.0",

15 "minimum\_index\_compatibility\_version" : "6.0.0-beta1"

16 },

17 "tagline" : "You Know, for Search"

18 }

19

20

21 # GET \_cat/indices

22 green open .kibana\_task\_manager\_1 nVdc4g8NRi0mshOWPq63zQ 1 0 2 6 42

.9kb 42.9kb

23 green open .apm-agent-configuration uyFzuj-nS76soUaGN3MYSQ 1 0 0 0

230b 230b

24 green open .kibana\_1 JRM24T5aScmZ5fhYxzRCpg 1 0 4 0 16

.3kb 16.3kb

25

# Introduction into Spring Data Elasticsearch

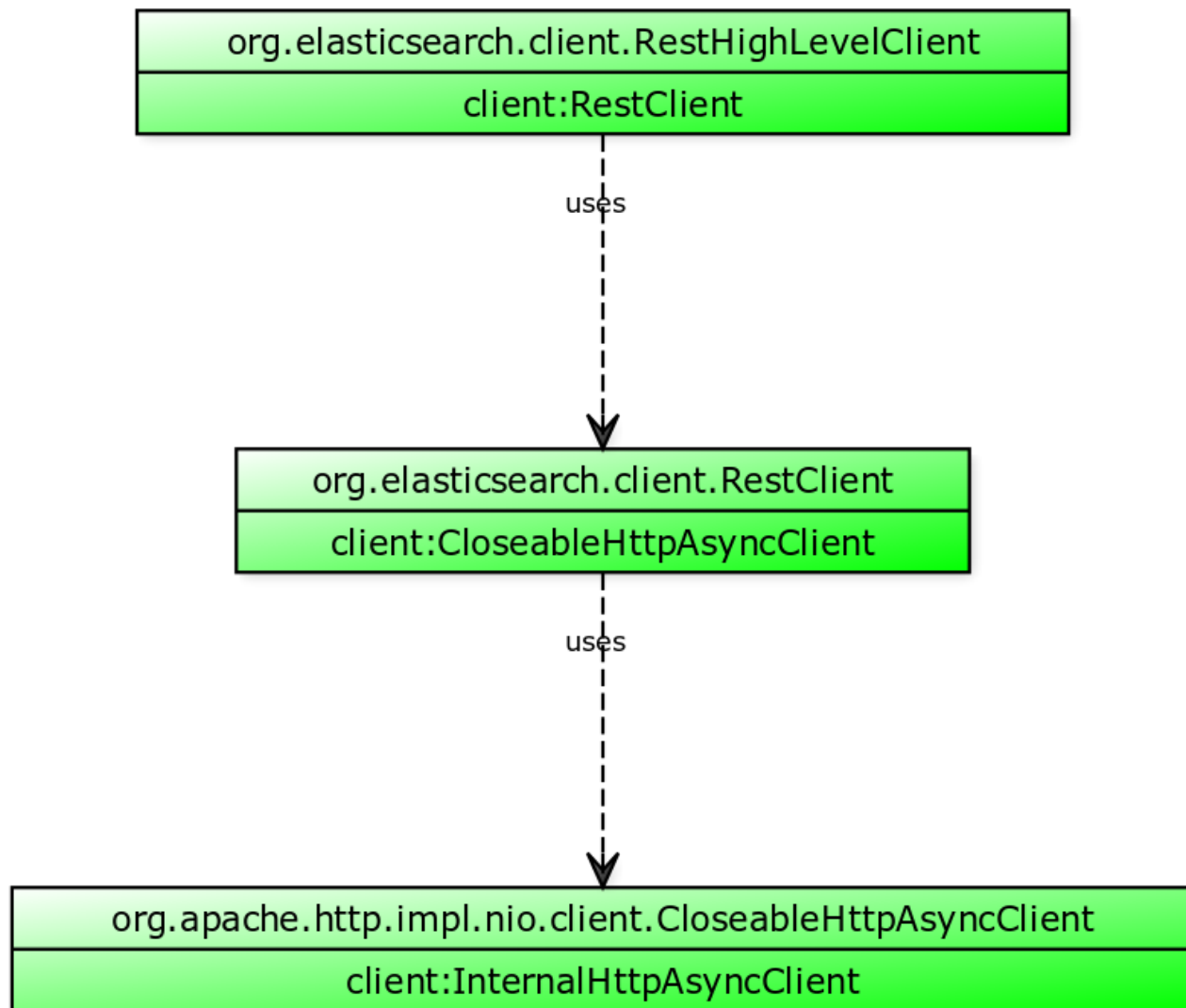
- Community maintained Spring Data Extension
- Reactive extension
- Make sure to use major version 4 (based on Elasticsearch 7.x), default in Spring Boot 2.3
- Uses the Elasticsearch REST Client

# Elasticsearch REST client

- Depends on the Elasticsearch core project
- Based on Apache HTTP Client (works on java 8), might want to consider shading
- Supports synchronous calls & cancellable async calls
- Threadsafe
- `RestClient`
- `RestHighLevelClient`



## Elasticsearch REST client architecture



# Spring Data Elasticsearch - Basics

- `ElasticsearchTemplate` & `ElasticsearchRestTemplate`
- `MappingElasticsearchConverter`
- `CrudRepository`
- Auditing, Entity Callbacks, efficient scroll searching

# Spring Data Elasticsearch - Entities

```
@Document(indexName = "persons", shards = 1, createIndex = false)
public class Person {
    @Id
    private String id;

    private String name;

    @Email
    @Field(type = FieldType.Keyword)
    private String email;

    @Field(name="created_at", type = FieldType.Date, format = DateFormat.date_time)
    private Date createdAt;

    @Size(max=500)
    @Pattern(regexp = "https?://.*", message = "must start with http:// or https://")
    @URL
    @Field(type = FieldType.Keyword)
    private String url;

    private List<Person> friends; // creates an array
    private Point location; // maps to geo_point
}
```

# Spring Data Elasticsearch - Repositories

```
import org.springframework.data.elasticsearch.repository.ElasticsearchRepository;  
  
public interface UserProfileRepository extends ElasticsearchRepository<UserProfile, String> {  
  
}
```

- Dynamic finders like `findByEmail(String email)`
- **Attention:** Inefficient queries like `findByDescriptionEndingWith()`

# Spring Data Elasticsearch - Searching

```
final BoolQueryBuilder qb = QueryBuilders.boolQuery()
    // somewhat stable randomization to make sure users get an arbitrary document
    .must(QueryBuilders.scriptScoreQuery(QueryBuilders.matchAllQuery(), new Script("randomScore(1000, 'created_at')")))
    // only consider contributions that are not yet approved
    .filter(QueryBuilders.termQuery("state", Contribution.State.CREATED.name()))
    // ensure only contributions from the same region are shown
    .filter(QueryBuilders.termQuery("region", profile.getRegion()))
    // only consider languages spoken by the user as well as english
    .filter(QueryBuilders.termsQuery("language", languages))
    // exclude documents that were created by this user
    .mustNot(QueryBuilders.termQuery("submitted_by.email", profile.getEmail()))
    // exclude documents that were already voted on
    .mustNot(QueryBuilders.termQuery("comments.submitted_by.email", profile.getEmail()));
Query query = new NativeSearchQuery(qb).setPageable(PageRequest.of(0, 1));

final SearchHit<Contribution> result = elasticsearchRestTemplate.searchOne(query, Contribution.class);
```

# Spring Data Elasticsearch - Count

```
private boolean canSubmitMoreContributions(String email) {  
    final BoolQueryBuilder qb = QueryBuilders.boolQuery()  
        .filter(QueryBuilders.termQuery("submitted_by.email", email))  
        .filter(QueryBuilders.rangeQuery("created_at").gte("now-1d"));  
    final long recentlySubmittedCount = elasticsearchRestTemplate.count(new NativeSearchQuery(qb), Contribution.class);  
    return recentlySubmittedCount <= 10;  
}
```

# Spring Data Elasticsearch - Count

```
public interface ContributionRepository extends ElasticsearchRepository<UserProfile, String> {  
    @Query("{\"bool\": { \"must\" : [ { \"term\" : { \"submitted_by.email\": \"?0\" } }, { \"range\" : { \"created_at\" : { \"gte\" : \"?1\" } } } ] } }")  
    long countRecentContributions(String email, String date);  
}
```

# Spring Data Elasticsearch - Aggregations

```
// filter by approved
final BoolQueryBuilder qb = QueryBuilders.boolQuery()
    .filter(QueryBuilders.termQuery("state", Contribution.State.APPROVED.name()))
    .filter(QueryBuilders.termQuery("region", region.name()));

final NativeSearchQuery query = new NativeSearchQuery(qb);
// aggregate on username, get top 10, sum up score
query.addAggregation(AggregationBuilders.terms("by_user").field("submitted_by.email").size(40)
    .subAggregation(AggregationBuilders.sum("total_score").field("score")))
    // make sure we get the full name of the last contribution
    .subAggregation(AggregationBuilders.topHits("by_name").size(1).sort(SortBuilders.fieldSort("created_at")
        .order(SortOrder.DESC)).fetchSource("submitted_by.full_name", "")));
query.setPageable(Pageable.unpaged());
final SearchHits<Contribution> hits = elasticsearchRestTemplate.search(query, Contribution.class);

// returns an Elasticsearch class
final Aggregations aggregations = hits.getAggregations();
```



# Demo

# Running Elasticsearch: Scaling your cluster

- **Do not overshard:** Single shard can easily contain 20-50GB
- Let the filesystem cache get to work
- Performance test, on your data! Use `rally`
- Hint: `Capacity Planning Webinar`

# Results

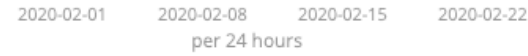
Show release charts

Distribution flavor for nightly

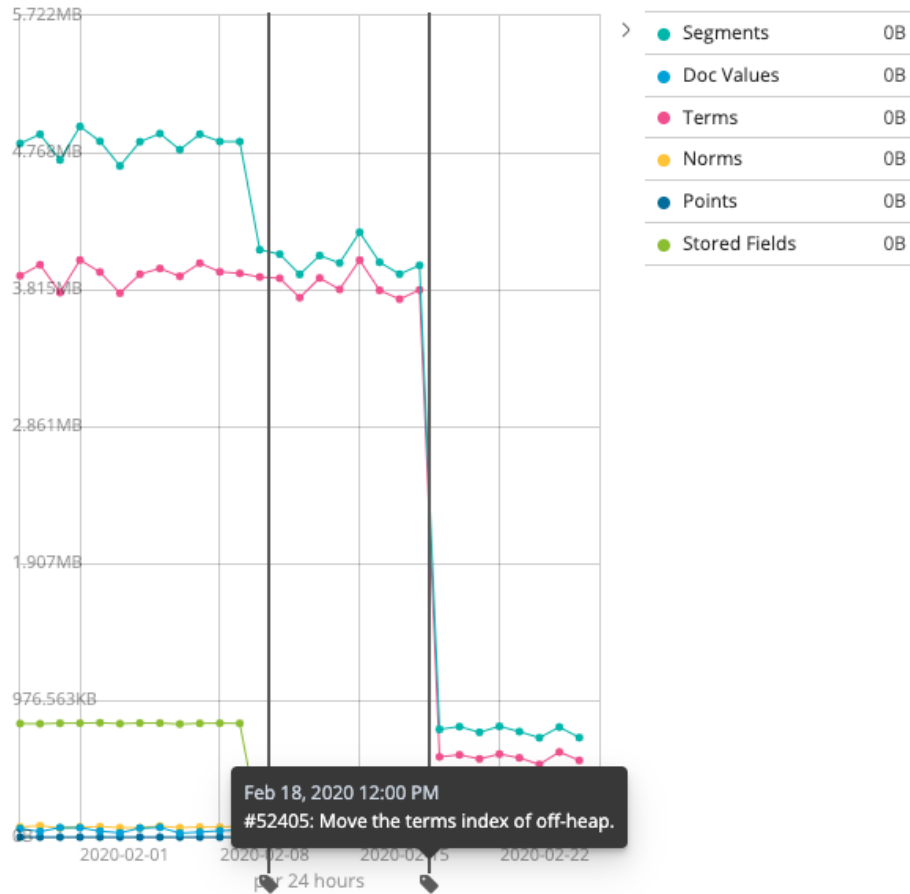
Default

Date range for nightly

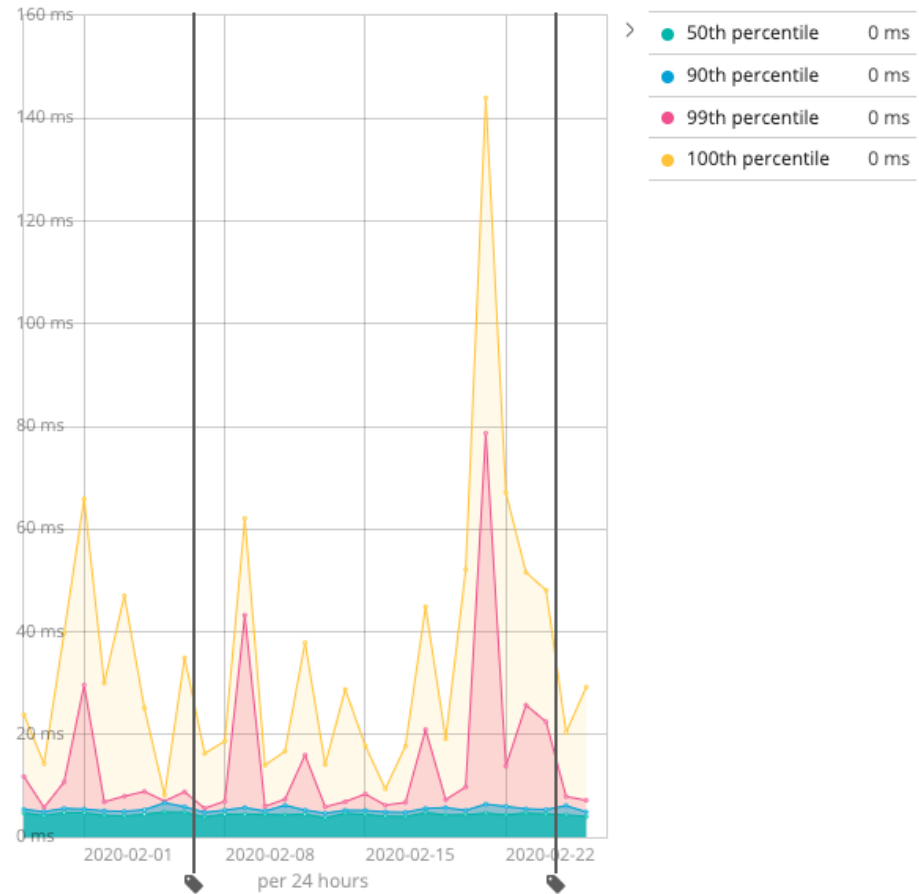
Last 6 months



nightly-basic-geonames-add-defaults-segment-memory



nightly-basic-geonames-add-defaults-index-stats-latency



# Compute Resources

- Storage: SSDs for hot data, HDDs for warm/cold, avoid NAS
- Memory: JVM heap + OS cache
- Compute: Thread pool scaling based on CPU count
- Network: The faster the better (**careful** cloud providers with burst rates)

# Next steps

- Improve your search: Learn about mappings and queries
- Improve your model
- Figure out expected throughput
- Use aliases, always!

# Summary

- Search is never done!
- Use the reference documentation
- Ask your users about expectations, do not guess!
- Testing: TestContainers

# Resources

- [spinscale/link-rating](#)
- [Qovery](#)
- [Spring Data Elasticsearch Documentation](#)
- [Elasticsearch Java REST Client Documentation](#)
- [Elasticsearch Nightly Benchmarks](#)

**Thanks for listening**

**Q & A**

Alexander Reelsen

Community Advocate

[alex@elastic.co](mailto:alex@elastic.co) | [@spinscale](https://twitter.com/spinscale)



**elastic**



# Elastic Cloud



The screenshot shows the Elastic Cloud pricing page. At the top, there is a navigation bar with the Elastic logo, links for Products, Learn, Company, and Pricing, and buttons for Contact, Try Free, and Login. Below the navigation bar, there are three tabs: SAAS (Elastic Cloud), STANDALONE (Elastic on-prem), and ORCHESTRATION (Elastic on-prem). The main heading is "Elastic Cloud pricing", followed by a sub-heading: "Pricing for our suite of SaaS offerings, which make it easy to deploy, operate, and scale Elastic products in the cloud." The page features three service cards: Elasticsearch Service (\$16/month), App Search Service (\$49/month), and Site Search Service (\$79/month). Each card includes a description, a price, a "See pricing" link, and a "Start free trial" button.

Service	Description	Price	Action
Elasticsearch Service	Easily spin up a deployment on AWS, GCP or Azure with Kibana and features you can't get anywhere else.	AS LOW AS \$16/month	<a href="#">See pricing</a> <a href="#">Start free trial</a>
App Search Service	Build a fast, relevant, search experience for your custom application in just a few minutes.	AS LOW AS \$49/month	<a href="#">See pricing</a> <a href="#">Start free trial</a>
Site Search Service	Everything you need to deliver a powerful search experience for your website — without the learning curve.	AS LOW AS \$79/month	<a href="#">See pricing</a> <a href="#">Start free trial</a>

# Elastic Support Subscriptions



elastic [Products](#) [Learn](#) [Company](#) [Pricing](#) [Contact](#) [Try Free](#) [Login](#) [🔍](#)

SAAS **Elastic Cloud** **STANDALONE Elastic on-prem** ORCHESTRATION **Elastic on-prem**

## Elastic Stack subscriptions

The Elastic Stack — Elasticsearch, Kibana, Beats, and Logstash — powers a variety of use cases. And we have flexible plans to help you get the most out of your on-prem subscriptions.

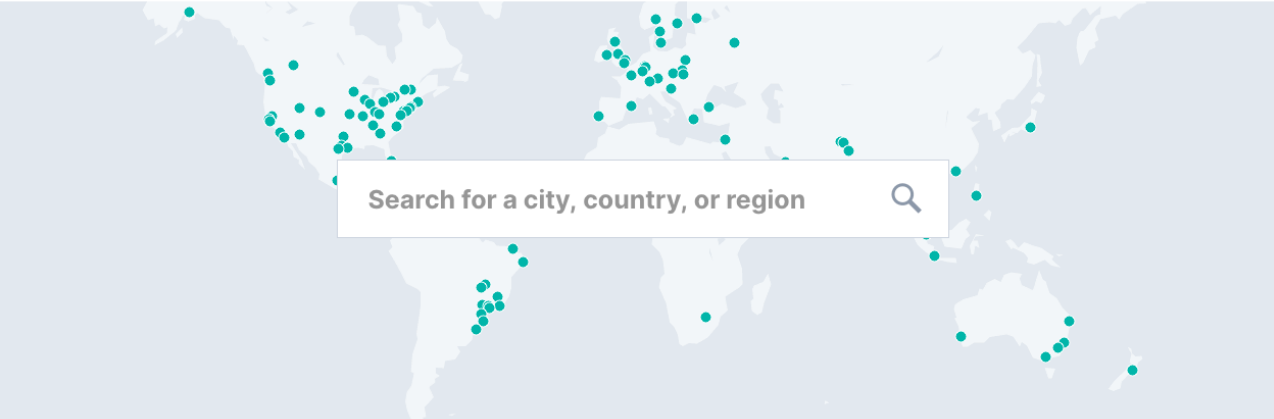
FREE		Gold	Platinum	Enterprise
<b>Open Source</b> Apache 2.0: Now and always.	<b>Basic</b> The forever-free plan.	More features. Dedicated support.	Advanced functionality. Around the clock support.	Stack orchestration and endpoint protection by default.
<b>Feature highlights include:</b>	<b>Everything in Open Source plus:</b>	<b>Everything in Basic plus:</b>	<b>Everything in Gold plus:</b>	<b>Everything in Platinum plus:</b>
<ul style="list-style-type: none"><li>✓ Clustering &amp; high availability</li><li>✓ Powerful search and analysis</li><li>✓ Data visualization and dashboarding</li><li>✓ And more</li></ul>	<ul style="list-style-type: none"><li>✓ Core security features</li><li>✓ Solutions such as APM, SIEM, Maps, and more</li><li>✓ Canvas</li><li>✓ And more</li></ul>	<ul style="list-style-type: none"><li>✓ Alerting</li><li>✓ Reporting</li><li>✓ Ingest management</li><li>✓ Business hours support</li><li>✓ And more</li></ul>	<ul style="list-style-type: none"><li>✓ Advanced security features</li><li>✓ Machine learning</li><li>✓ Cross-cluster replication</li><li>✓ 24/7/365 support</li><li>✓ And more</li></ul>	<ul style="list-style-type: none"><li>✓ Endpoint prevention</li><li>✓ Endpoint detection and response mapped to MITRE ATT&amp;CK</li><li>✓ Endpoint event collection</li><li>✓ Access to ECE &amp; ECK orchestration features</li></ul>
<a href="#">Free download</a>		<a href="#">Contact us</a>	<a href="#">Contact us</a>	<a href="#">Contact us</a>

# Discuss Forum

<https://discuss.elastic.co>



Category	Topics	Latest
<b>Announcements</b> Release announcements, end of life notifications and other bits about Elastic products that we think will be useful to everyone. <b>Community Ecosystem</b>	385 5 unread	<b>Notes on Using These Forums</b> <span>2</span> <b>Meta Elastic</b> <span>Apr 17</span>
<b>Beats</b> Any questions regarding Beats, forwarders and shippers for various types of data.	61 / week 1 unread 15 new	<b>Couldn't push logs to elasticsearch using filebeat</b> <span>1</span> <b>Filebeat</b> <span>3m</span>
<b>Elasticsearch</b> Any questions related to Elasticsearch, including specific features, language clients and plugins. <b>Rally</b> 1 unread	178 / week 831 unread 36 new	<b>&lt;BarSeries&gt; configuration</b> <span>0</span> <b>Kibana</b> <span>6m</span>
<b>Logstash</b> Everything related to your favorite centralized logging platform, including plugins and recipes.	95 / week 29 unread 24 new	<b>Dec 15th, 2019: [EN] Elasticsearch Snapshot Lifecycle Management (SLM) with Minio.io S3</b> <span>0</span> <b>advent-staging</b> <span>7m</span>
<b>Kibana</b> All things about visualizing data in Elasticsearch & Logstash, including how to use Kibana and extending the platform.	113 / week 42 unread 19 new	<b>Invalid IP network, skipping {network=&gt;"10.13.7.0/10.13.7.24"</b> <span>0</span> <b>Logstash</b> <span>10m</span>
<b>APM</b> Everything related to APM - whether it is the APM Server, the Kibana dashboards, or the agents.	12 / week 5 new	<b>FScrawler stuck at 2.6gb index size</b> <span>2</span> <b>Elasticsearch</b> <span>11m</span>
<b>Logs</b> Everything related to the Logs app - setup with Filebeat, Filebeat modules, and using the Kibana Logs app.	55	<b>Elastic APM Java agent - sanitize_fields_names on application/json* data</b> <span>1</span> <b>APM</b> <b>java</b> <span>21m</span>
<b>Metrics</b> Everything related to metrics - Metricbeat, integrations and modules, Kibana dashboards and the Metrics app.	1 / week	<b>Metricbeat Failed to connect EOF</b> <span>5</span> <b>Metricbeat</b> <span>22m</span>
		<b>Mix free and paid licenses</b> <span>0</span> <b>Elasticsearch</b> <b>license</b> <span>23m</span>
		<b>Filebeat CPU utilization metrics are not normalized by default</b> <span>2</span> <b>Beats</b> <b>stack-monitoring</b> <span>23m</span>
		<b>How do i aggregate these documets</b> <span>6</span> <b>Logstash</b> <span>26m</span>
		<b>Metricbeat error</b> <span>1</span> <b>Metricbeat</b> <span>28m</span>



# Community & Meetups

<https://community.elastic.co>



### Explore by region

Asia Pacific and Japan | **Europe, Middle East and Africa** | North and South America | Virtual

<b>ELASTIC - BARCELONA</b> Spain 🇪🇸	<b>ELASTIC - COPENHAGEN</b> Denmark 🇩🇰	<b>ELASTIC - GOTEBORG</b> Sweden 🇸🇪	<b>ELASTIC - SCOTLAND</b> United Kingdom 🇬🇧
<b>ELASTIC - STOCKHOLM</b> Sweden 🇸🇪	<b>ELASTIC - TEL AVIV</b> Israel 🇮🇱	<b>ELASTIC - TURKEY</b> Turkey 🇹🇷	<b>ELASTIC BONN USER GROUP</b> Germany 🇩🇪
<b>ELASTIC CAMBRIDGE &amp; EAST ANGLIA USER GROUP</b> United Kingdom 🇬🇧	<b>ELASTIC DUBAI USER GROUP</b> United Arab Emirates 🇦🇪	<b>ELASTIC FR</b> France 🇫🇷	<b>ELASTIC GREECE</b> Greece 🇬🇷
<b>ELASTIC HELSINKI</b> Finland 🇫🇮	<b>ELASTIC KRAKOW USER GROUP</b> Poland 🇵🇱	<b>ELASTIC LONDON USER GROUP</b> United Kingdom 🇬🇧	<b>ELASTIC LUXEMBOURG USER GROUP</b> Luxembourg 🇱🇺
<b>ELASTIC MANCHESTER USER GROUP</b> United Kingdom 🇬🇧	<b>ELASTIC MOSCOW</b> Russian Federation 🇷🇺	<b>ELASTIC NIGERIA</b> Nigeria 🇳🇮	<b>ELASTIC OSLO USER GROUP</b> Norway 🇳🇴
<b>ELASTIC PORTUGAL</b> Portugal 🇵🇹	<b>ELASTIC RHEINRUHR</b> Germany 🇩🇪	<b>ELASTIC SLOVAK USER GROUP</b> Slovakia 🇸🇰	<b>ELASTIC USER GROUP - CZ</b> Czech Republic 🇨🇪
<b>ELASTIC USER GROUP - DUBLIN</b> Ireland 🇮🇪	<b>ELASTIC USER GROUP ABIDJAN</b> Côte d'Ivoire 🇨🇮	<b>ELASTIC WARSAW USER GROUP</b> Poland 🇵🇱	<b>ELASTIC ZAGREB</b> Croatia 🇭🇷
<b>ELASTICSEARCH - SOUTH AFRICA</b> South Africa 🇿🇦	<b>ELASTICSEARCH SWITZERLAND</b> Switzerland 🇨🇭	<b>ELASTICSEARCH USER GROUP PAKISTAN</b> Pakistan 🇵🇰	<b>SEARCH MEETUP MUNICH</b> Germany 🇩🇪

**Thanks for listening**

**Q & A**

Alexander Reelsen

Community Advocate

[alex@elastic.co](mailto:alex@elastic.co) | [@spinscale](https://twitter.com/spinscale)



**elastic**