

Introduction into Elasticsearch

& Spring Data Elasticsearch

Alexander Reelsen

Community Advocate

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



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

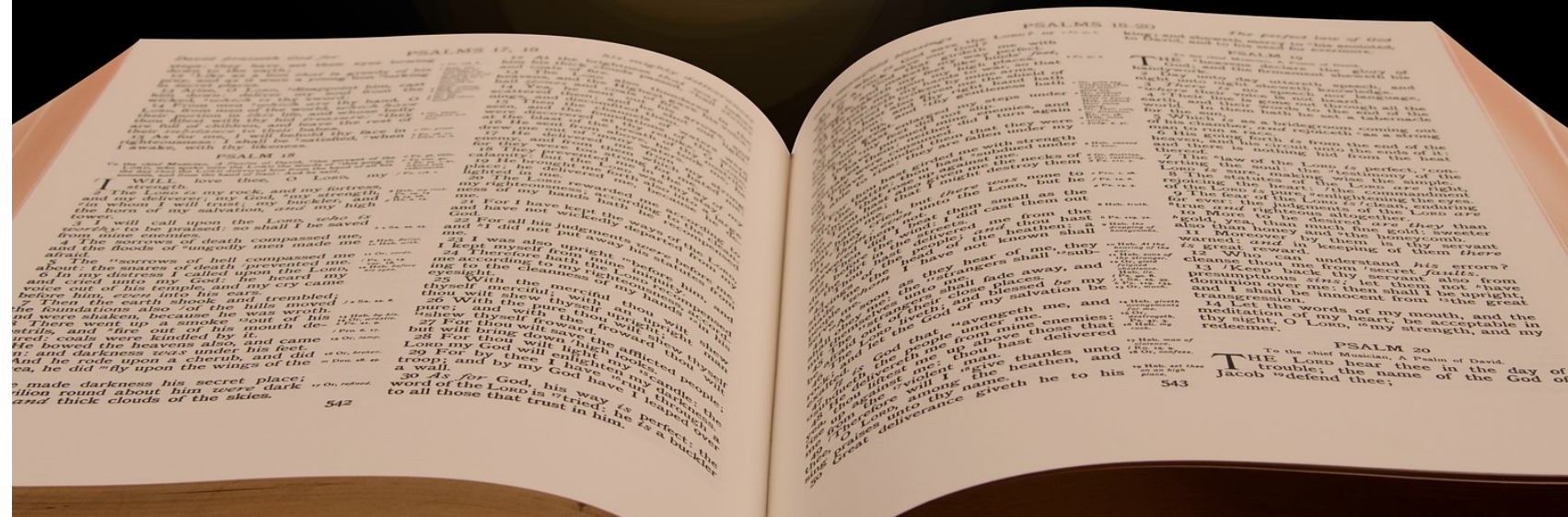
Speed



Scale



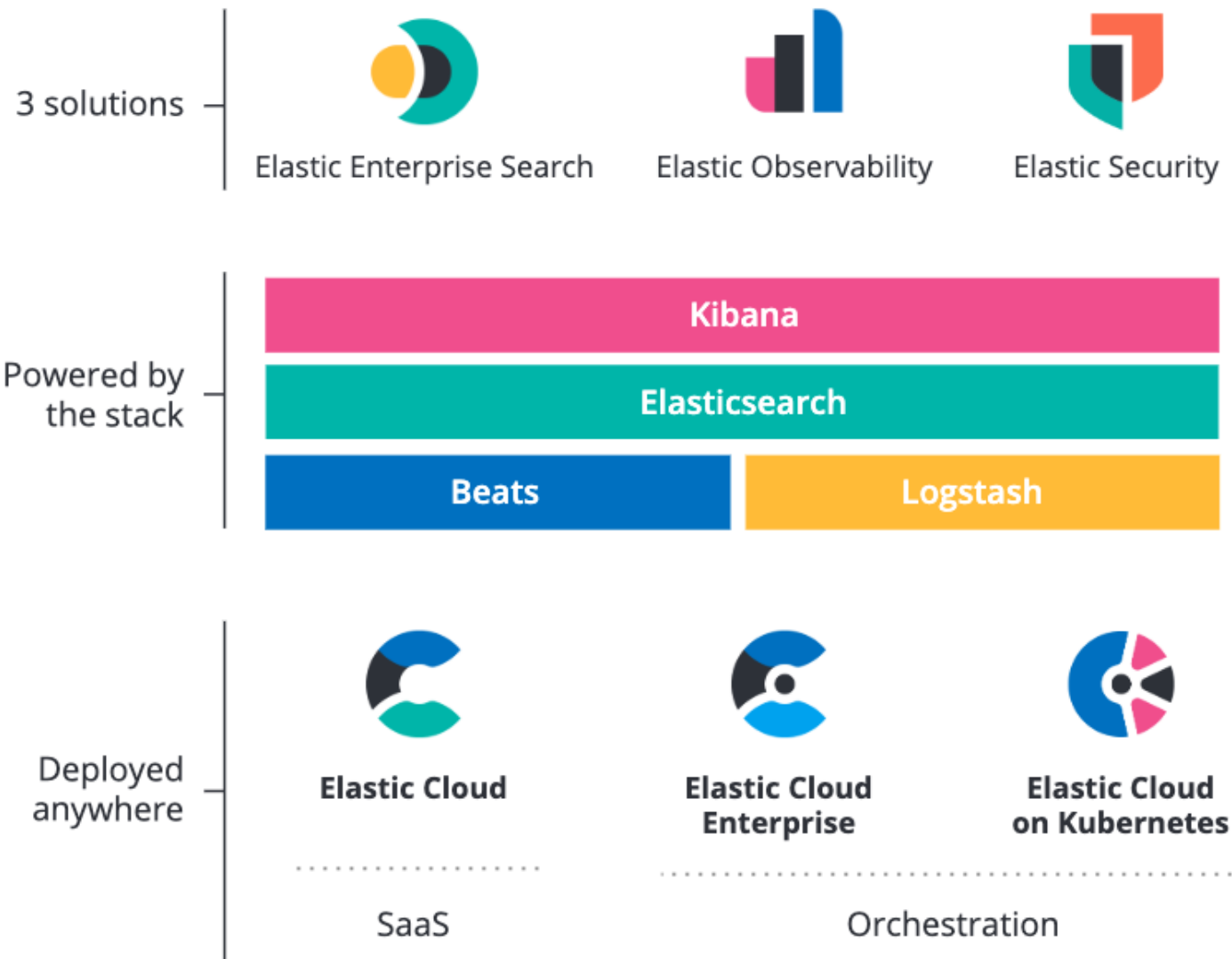
Relevance



... 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



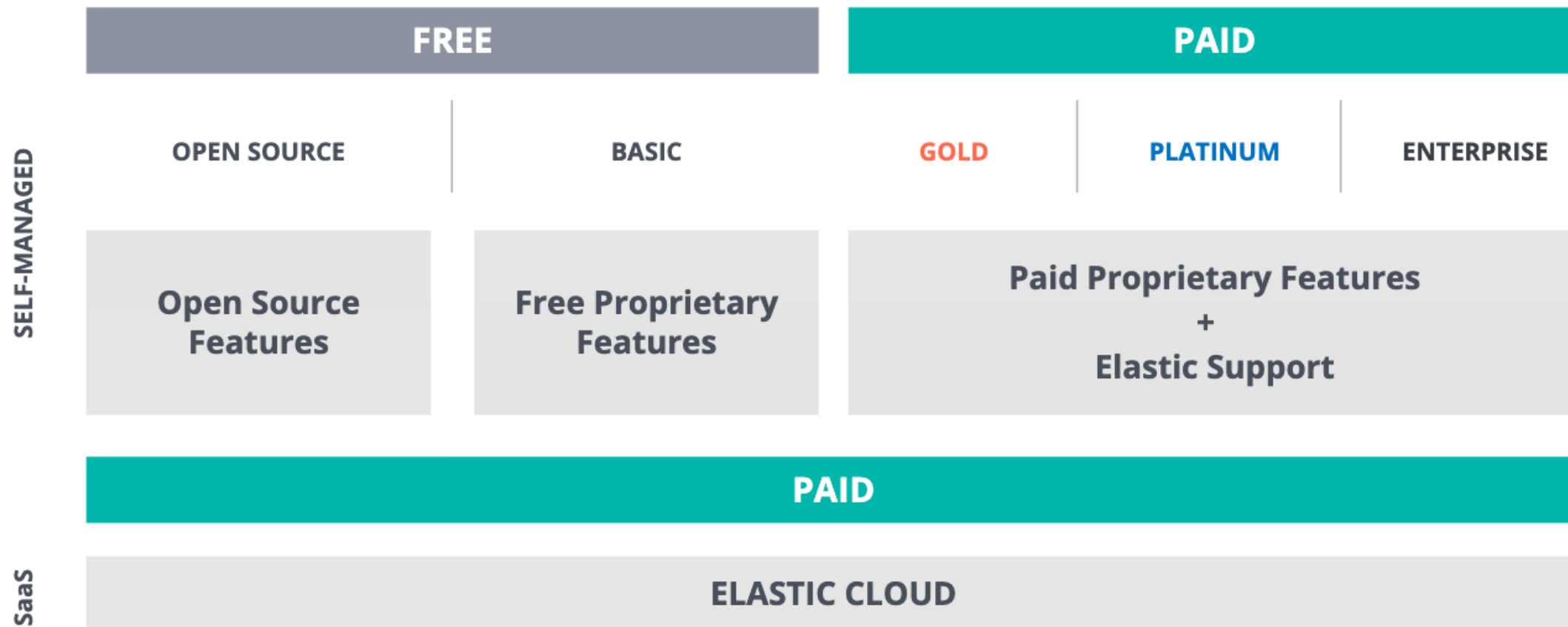
**Elastic Cloud
Enterprise**



**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-7.9.1-darwin-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.9.1-linux-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.9.1-windows-x86_64.zip

tar xzf elasticsearch-7.9.1-darwin-x86_64.tar.gz
cd elasticsearch-7.9.1

./bin/elasticsearch
```

```
wget https://artifacts.elastic.co/downloads/kibana/kibana-7.9.1-darwin-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/kibana/kibana-7.9.1-linux-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/kibana/kibana-7.9.1-windows-x86_64.zip

tar xzf kibana-7.9.1-darwin-x86_64.tar.gz
cd kibana-7.9.1
./bin/kibana
```

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

Click Dev-Tools

Samples in Kibana

Samples in Github



The screenshot shows the Elastic Dev Tools interface. At the top, there is a navigation bar with a logo, a 'D' icon, and a 'Home' label. Below this is a sidebar menu with the following items: 'Recently viewed', 'Discover', 'Visualize', 'Dashboard', 'Canvas', 'Maps', 'Machine Learning', 'Metrics', 'Logs', 'APM', 'Uptime', 'SIEM', 'Dev Tools' (highlighted), 'Stack Monitoring', and 'Management'. The main content area is divided into two sections. The top section is titled 'lastic Stack' and contains the text 'ge data helps us manage and improve our products and services'. The bottom section is titled 'Kibana' and contains the text 'o quickly turn your data into pre-built dashboards and monitoring'. Below the 'Kibana' section, there are two cards. The left card is titled 'APM' and contains the text 'ly collects in- ce metrics and sive your ions.' Below this card is a button labeled 'APM'. The right card is titled 'Logging' and contains the text 'Ingest logs from popular data sources and easily visualize in preconfigured dashboards.' Below this card is a 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

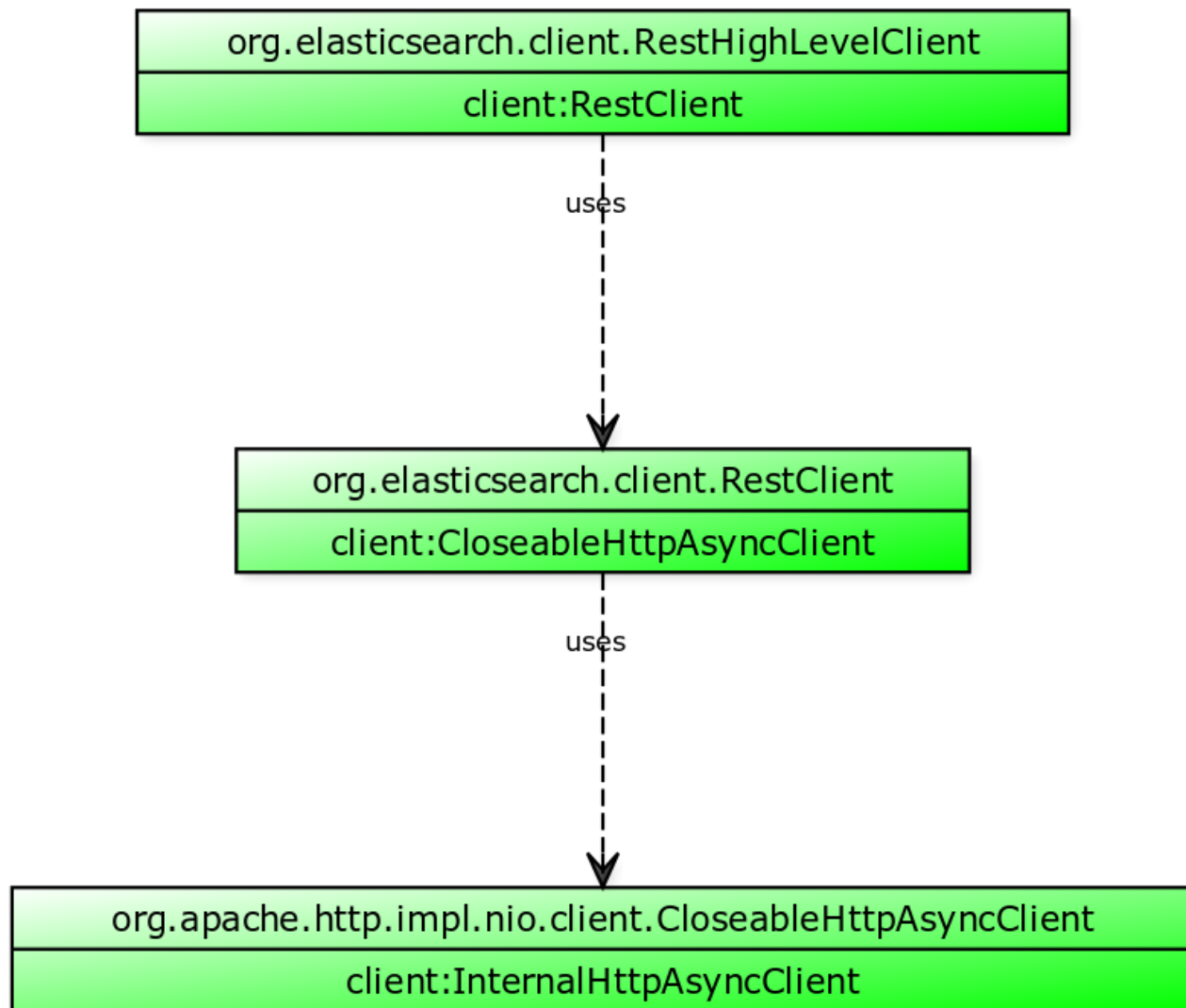
Introduction into Spring Data Elasticsearch

- 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

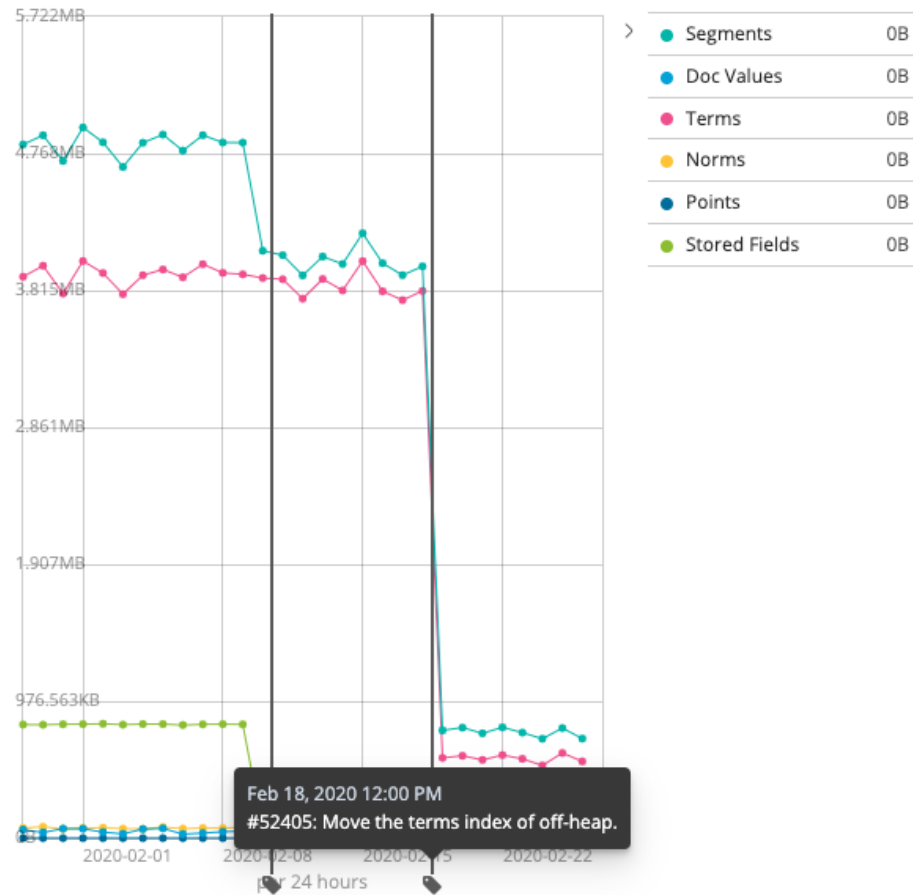
Default

Date range for nightly

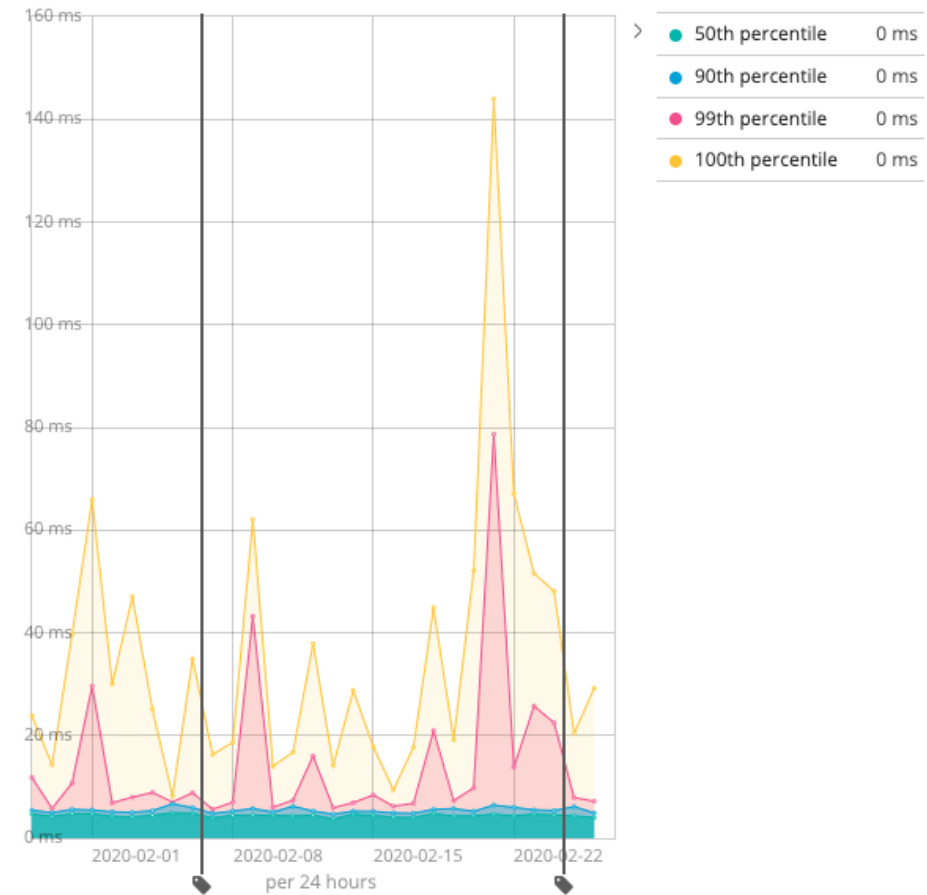
Last 6 months

2020-02-01 2020-02-08 2020-02-15 2020-02-22
per 24 hours2020-02-01 2020-02-08 2020-02-15 2020-02-22
per 24 hours

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 | [@spinscale](https://twitter.com/spinscale)



elastic

Elastic Cloud






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

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

Elastic Cloud pricing

Pricing for our suite of SaaS offerings, which make it easy to deploy, operate, and scale Elastic products in the cloud.

 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	See pricing	Start free trial
 App Search Service Build a fast, relevant, search experience for your custom application in just a few minutes.	AS LOW AS \$49/month	See pricing	Start free trial
 Site Search Service Everything you need to deliver a powerful search experience for your website — without the learning curve.	AS LOW AS \$79/month	See pricing	Start free trial

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.

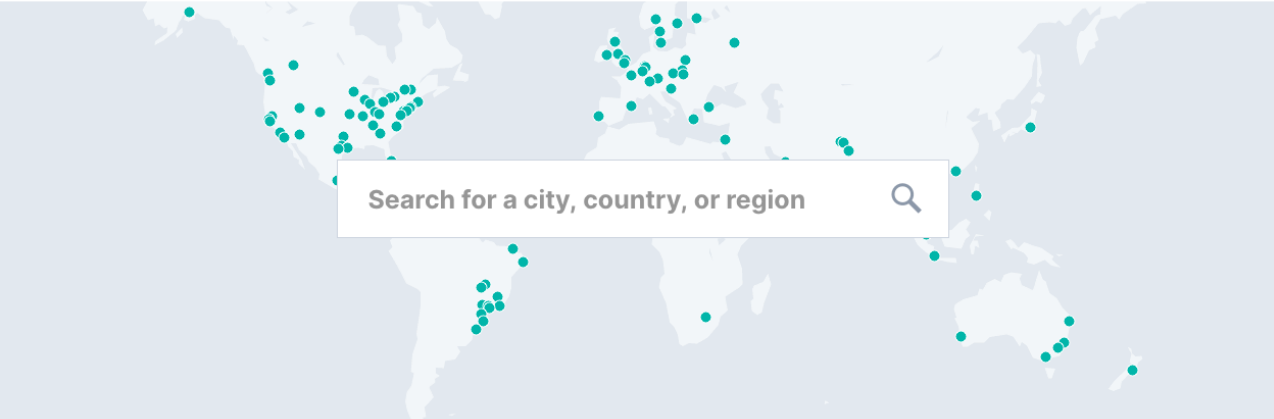
FREE		Gold	Platinum	Enterprise
Open Source Apache 2.0: Now and always.	Basic The forever-free plan.	More features. Dedicated support.	Advanced functionality. Around the clock support.	Stack orchestration and endpoint protection by default.
Feature highlights include:	Everything in Open Source plus:	Everything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

Discuss Forum

<https://discuss.elastic.co>



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



Community & Meetups

<https://community.elastic.co>



Explore by region

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

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

Elastic YouTube Community

<https://ela.st/yt-community>



A screenshot of the YouTube channel page for the "Official Elastic Community". The page features a blue header with the Elastic logo and the word "Community". Below the header, there is a channel profile section with the channel name "Official Elastic Community", 31 subscribers, and a "SUBSCRIBED" button. Navigation tabs for "HOME", "VIDEOS", "PLAYLISTS", "CHANNELS", "DISCUSSION", and "ABOUT" are visible. The main content area is divided into two sections: "Elastic Observability" and "Elastic Security". Each section includes a "PLAY ALL" button and a description of the videos. The "Elastic Observability" section lists four videos: "Monitoring and Preventing Threats as Employees Transition from Office to Home with Salt Lake County" (53:11), "How I Monitor my Home with Elastic Canvas" (23:42), "Elastic Machine Learning + Homechoice's Elastic Observability Journey" (1:05:40), and "Capgemini: Advanced Threat Hunting & Monitoring with Elastic APM" (51:58). The "Elastic Security" section lists four videos: "SIEM 101: What, Why & How of Information Security - May 5, 2020 Elastic Meetup" (58:46), "Upgrade Your Attack Model: Stopping Fileless Attacks..." (37:20), "Elastic SIEM: Part 1 Getting Started to Investigating..." (1:01:23), and "Elastic SIEM: Part 2 Getting Started to Investigating..." (52:26). Each video card includes a thumbnail, title, channel name, and view/like information.

Thanks for listening

Q & A

Alexander Reelsen

Community Advocate

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



elastic