Introduction into Elasticsearch

& Spring Data Elasticsearch

Alexander Reelsen
Community Advocate
alex@elastic.co | @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

elastic

Scale

elastic

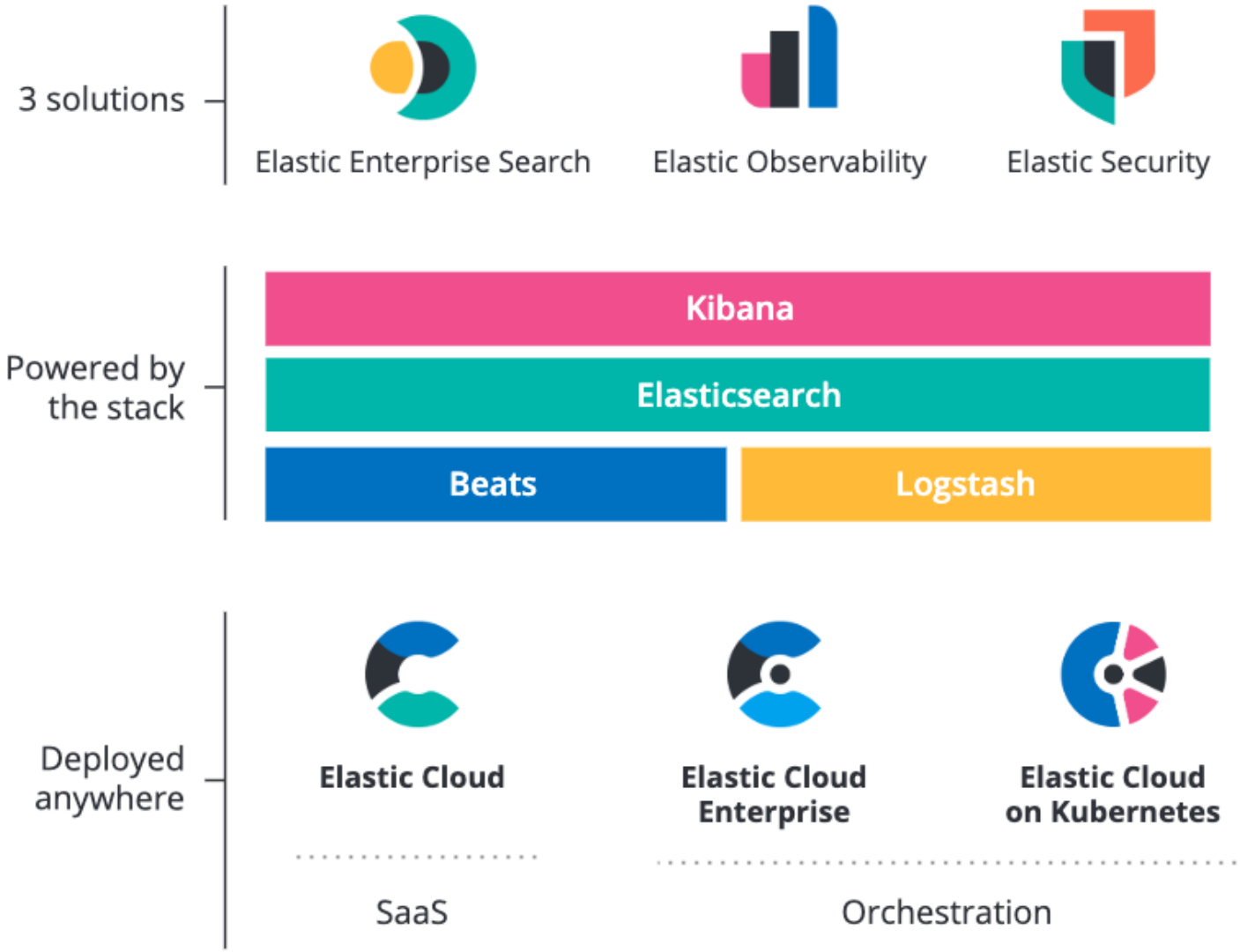
Relevance

elastic

... 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



elastic

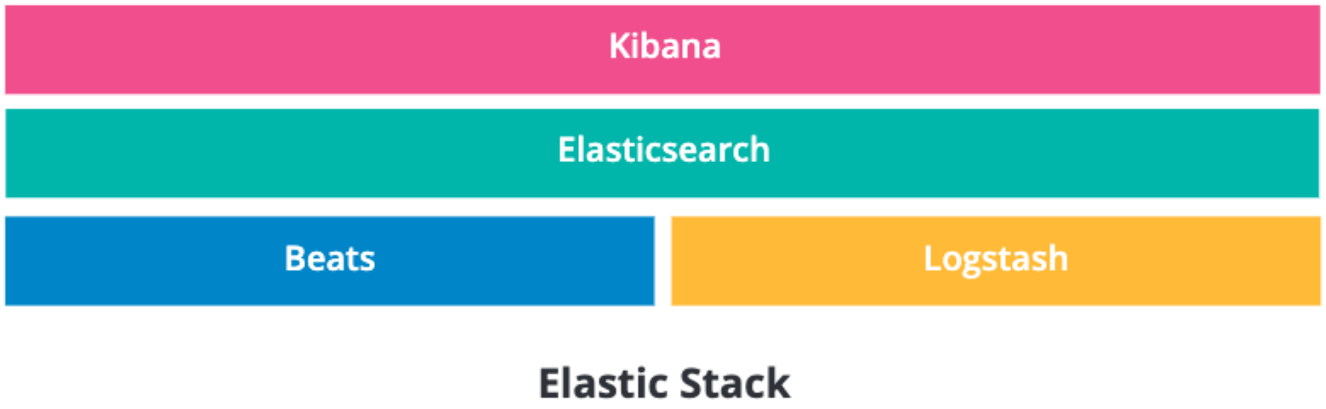
Solutions

3 solutions powered by 1 stack

Elastic Enterprise Search

Elastic Observability

Elastic Security

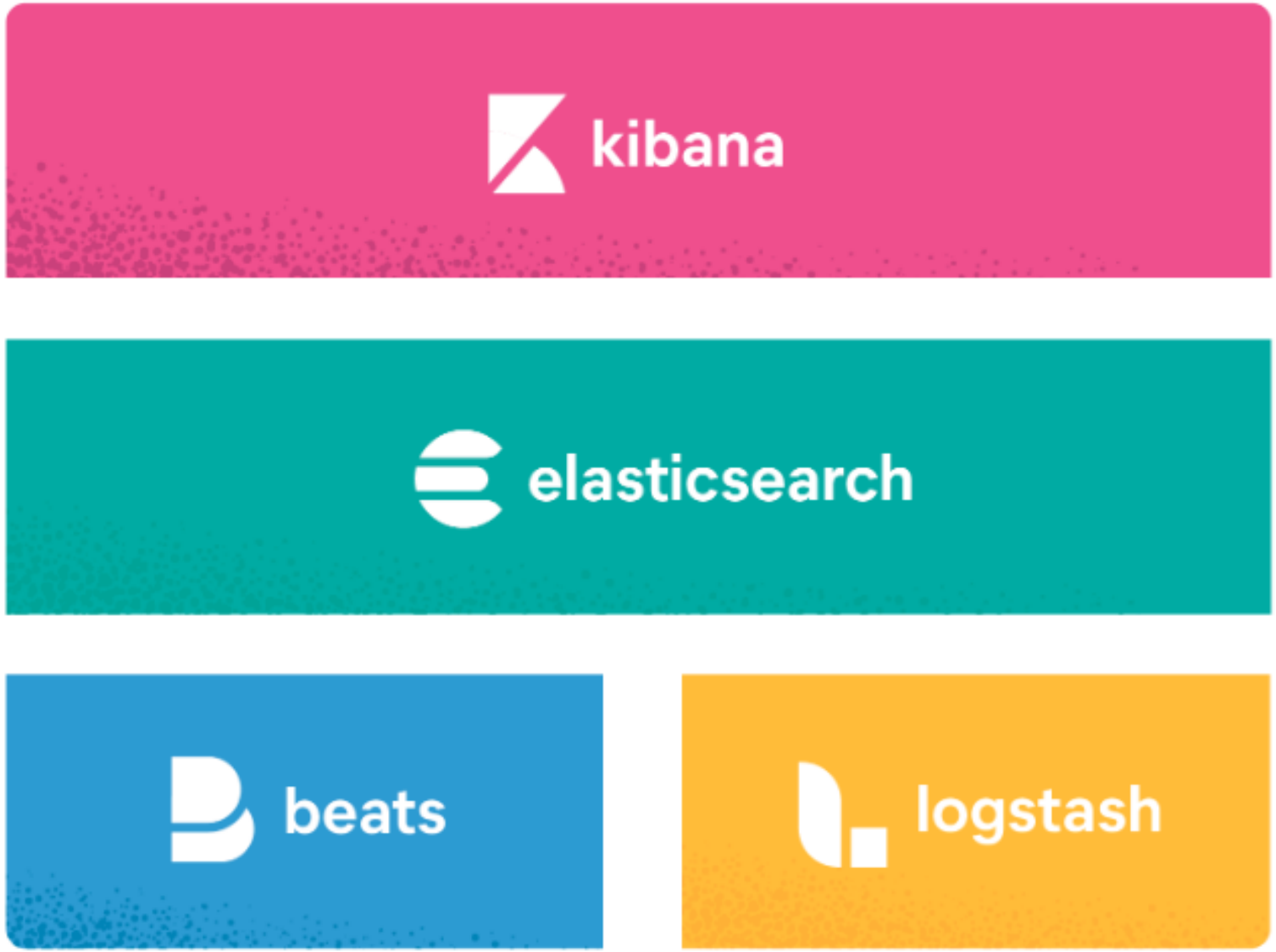


Elastic Stack

elastic

Elastic Stack

building blocks



elastic

Deployment options

Elastic Cloud

Elastic Cloud Enterprise

Elastic Cloud on Kubernetes

SaaS

Orchestration

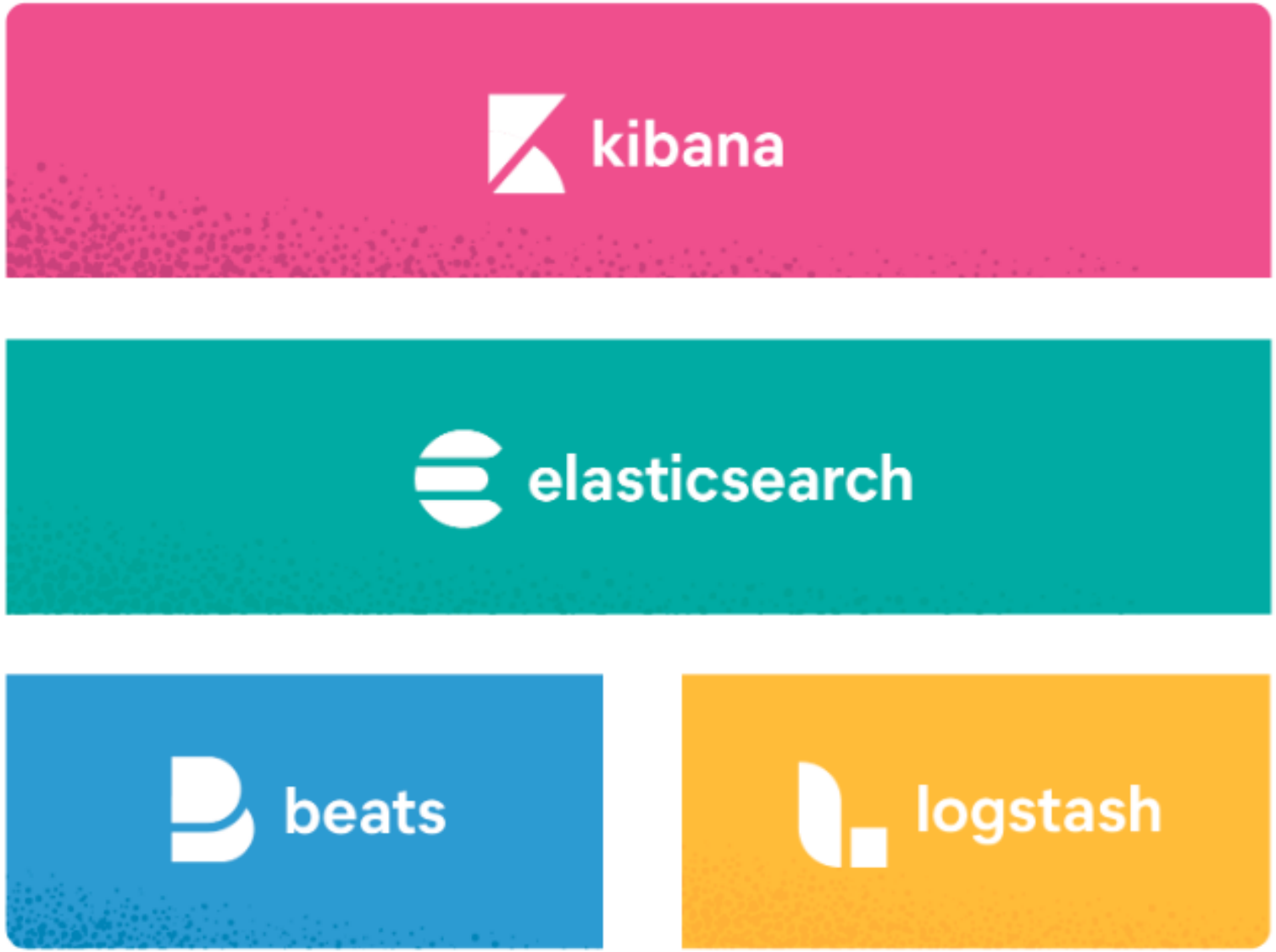
elastic

Licensing

elastic

Elastic Stack

building blocks



elastic

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

./bin/elasticsearch
```

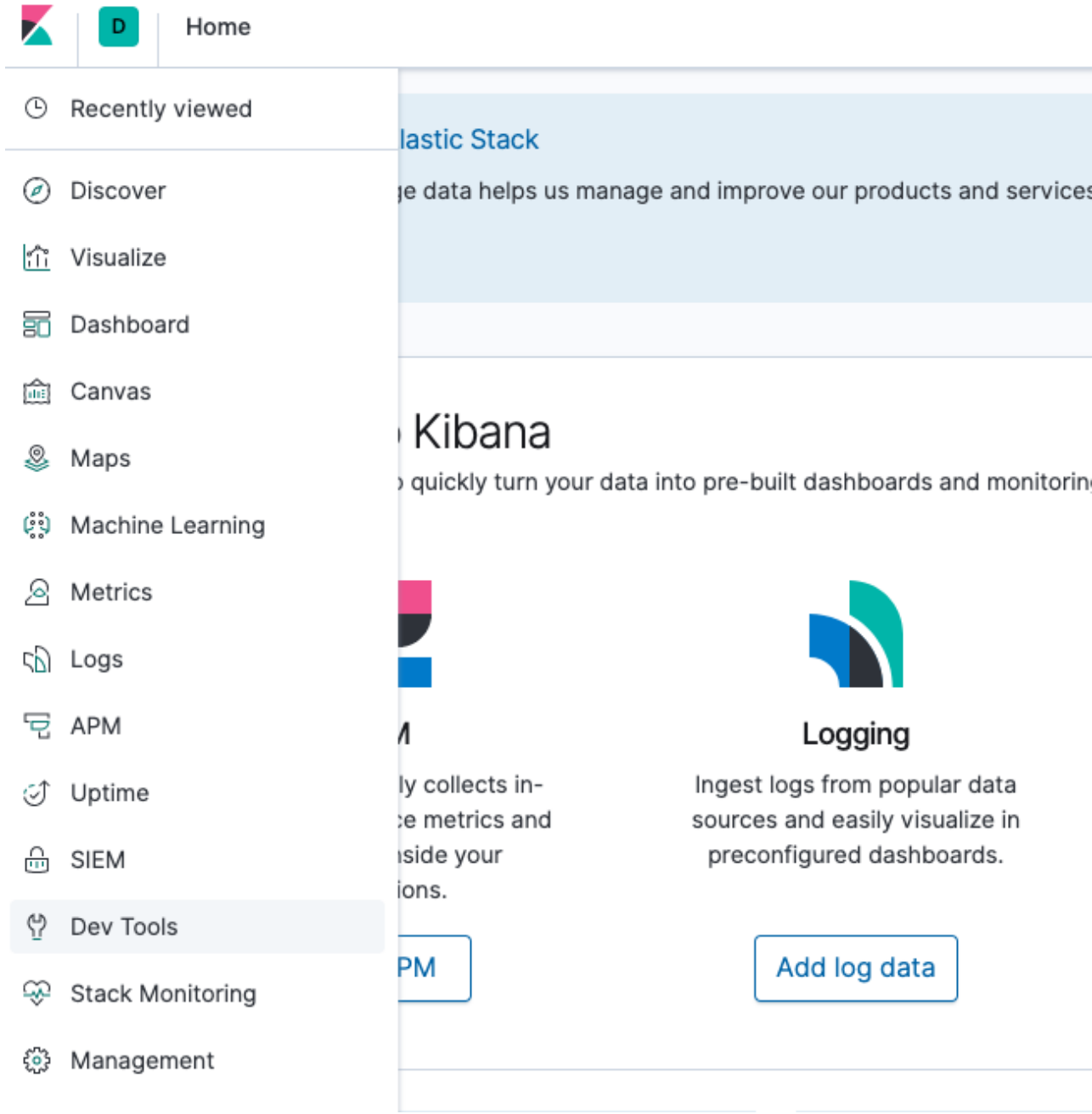
```
wget https://artifacts.elastic.co/downloads/kibana/kibana-7.9.1-darwin-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/kibana/kibana-7.9.1-linux-x86_64.tar.gz
# wget https://artifacts.elastic.co/downloads/kibana/kibana-7.9.1-windows-x86_64.zip

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

Point your browser to `http://localhost:5601/`

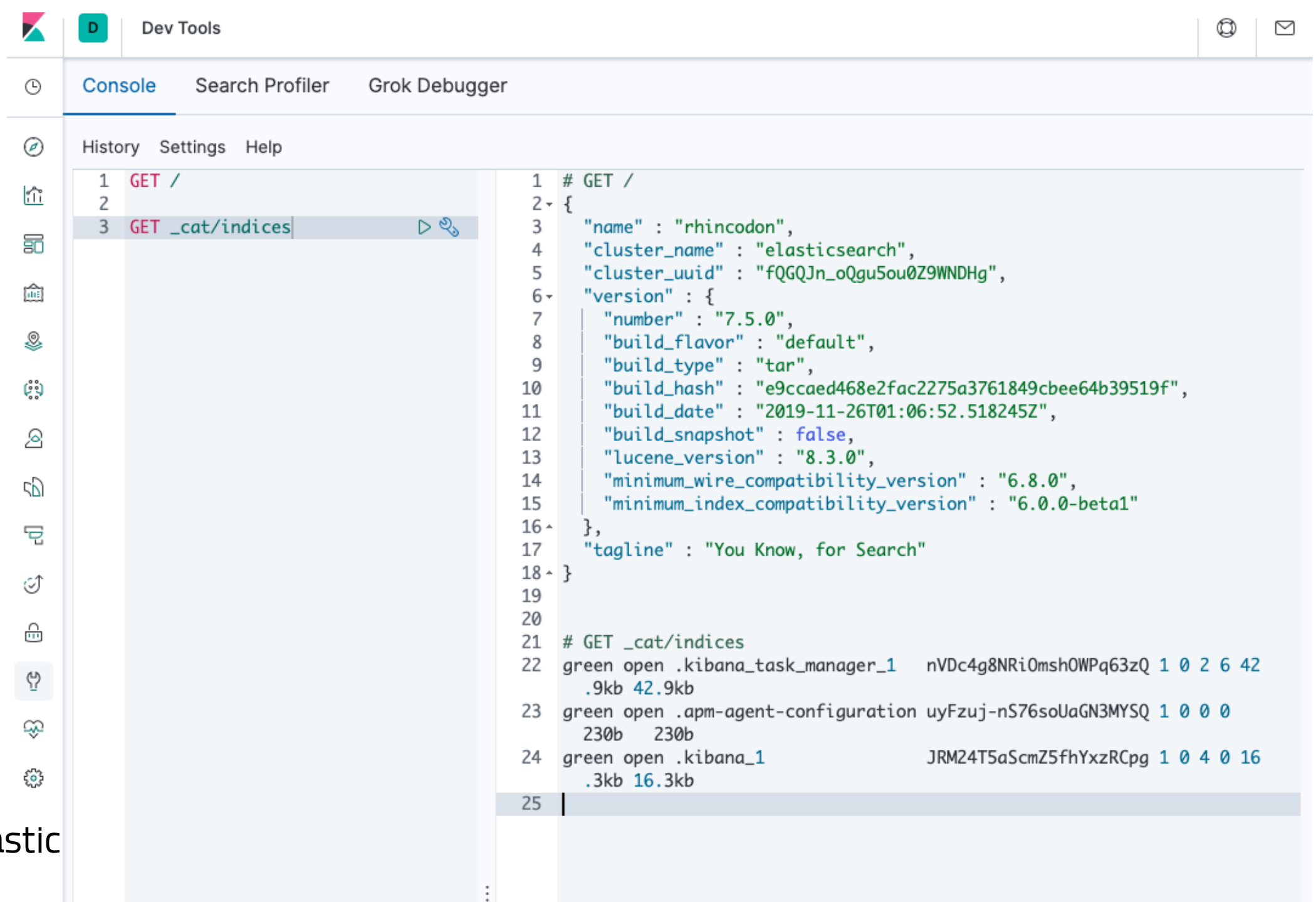
Click Dev-Tools

Samples in Kibana

Samples in Github



elastic


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 nVDc4g8NRiOmshOWPq63zQ 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
elastic

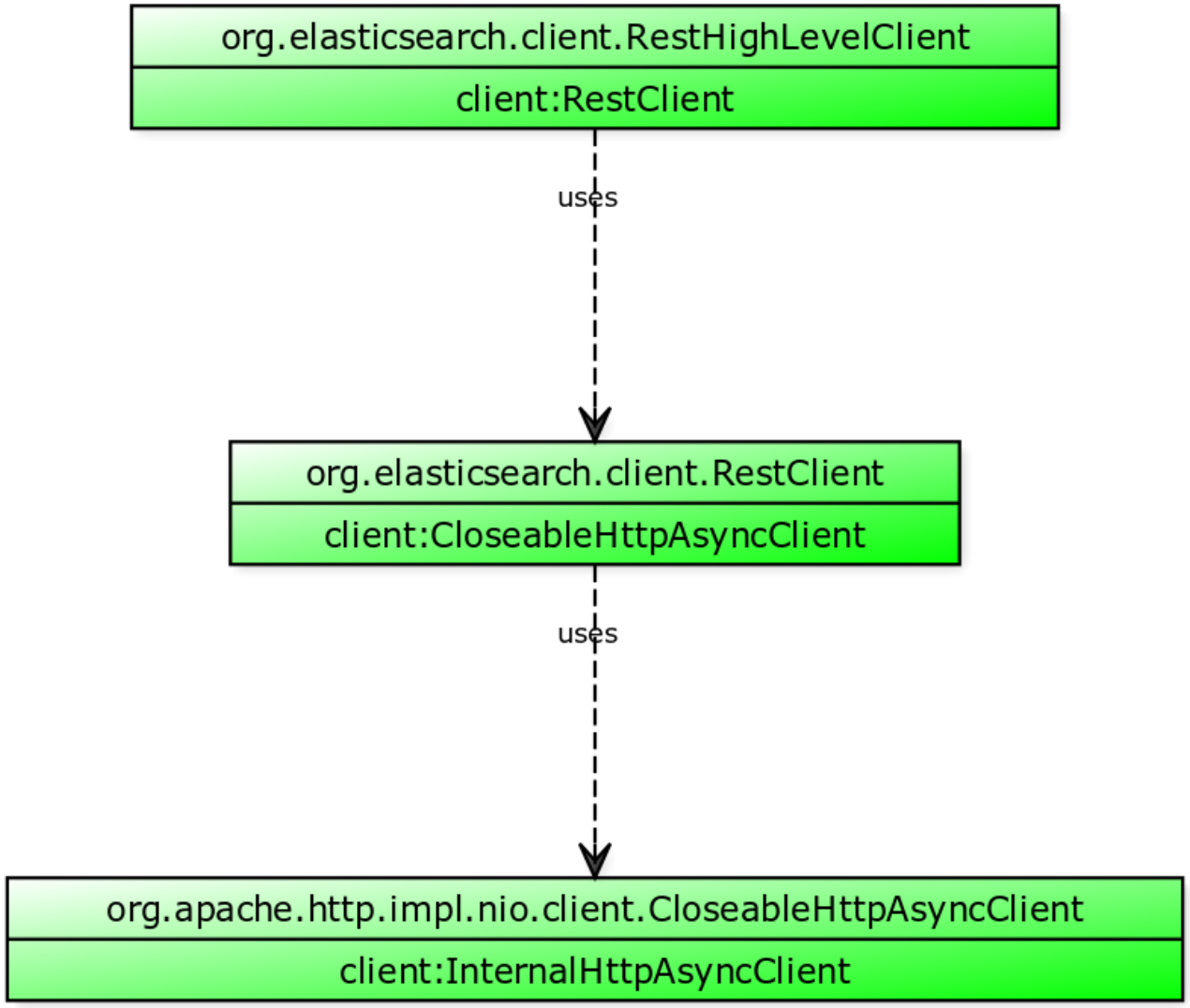
Introduction into Spring Data Elasticsearch

- 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 approvaed
  .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);

}
```

elastic

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

Elasticsearch Benchmarks | Geonames | Geopoint | Percolator | PMC | NYC taxis | Nested | HTTP Logs | noaa

Results

Show release charts

Distribution flavor for nightly

Default

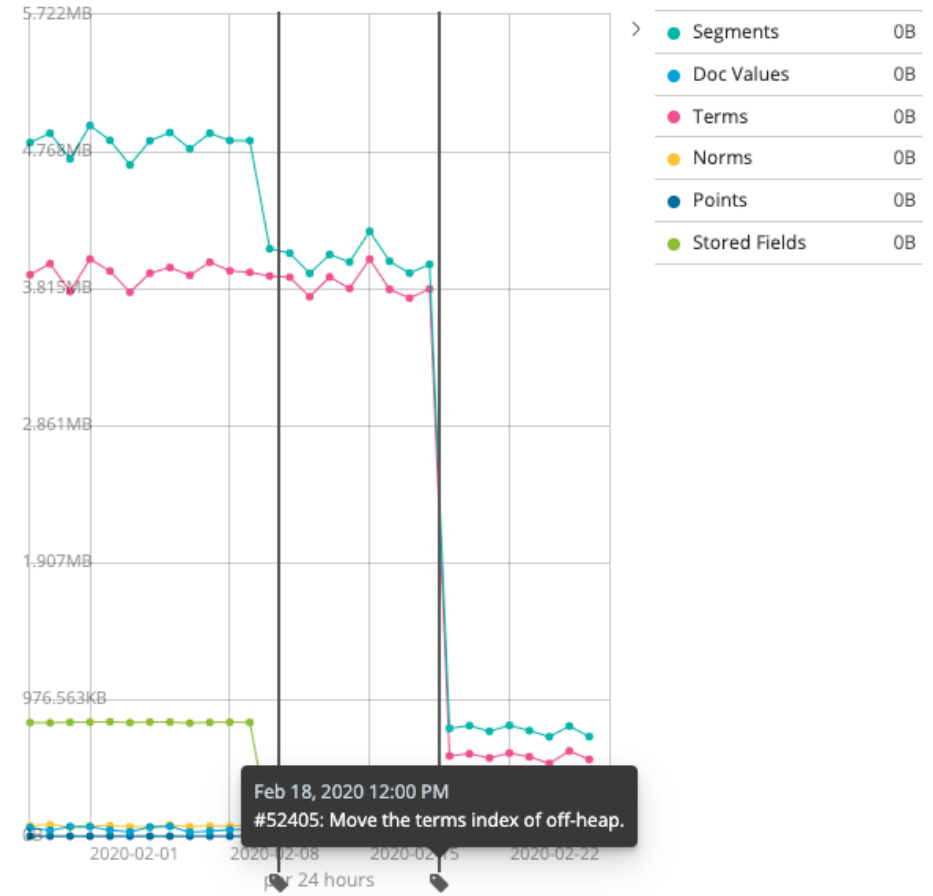
Date range for nightly

Last 6 months

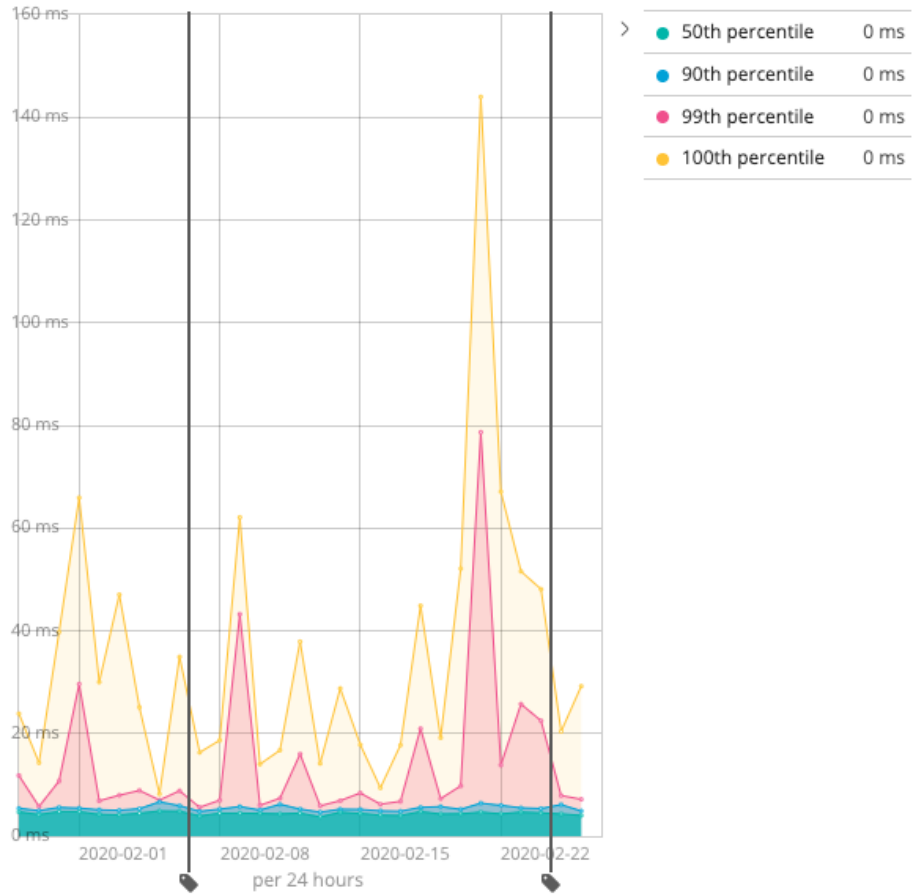
2020-02-01 2020-02-08 2020-02-15 2020-02-22

per 24 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

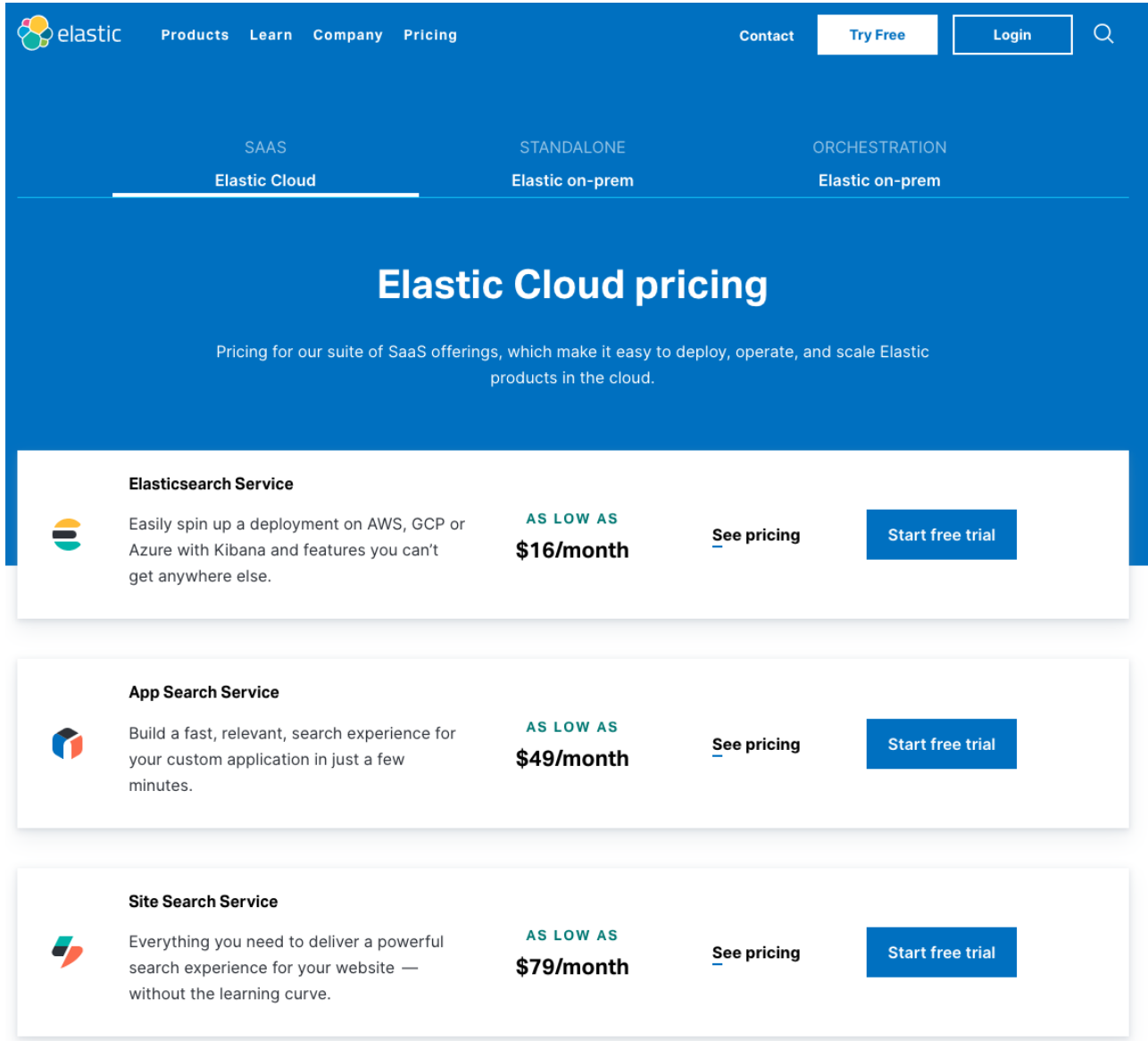
Thanks for listening

Q & A

Alexander Reelsen
Community Advocate
alex@elastic.co | @spinscale

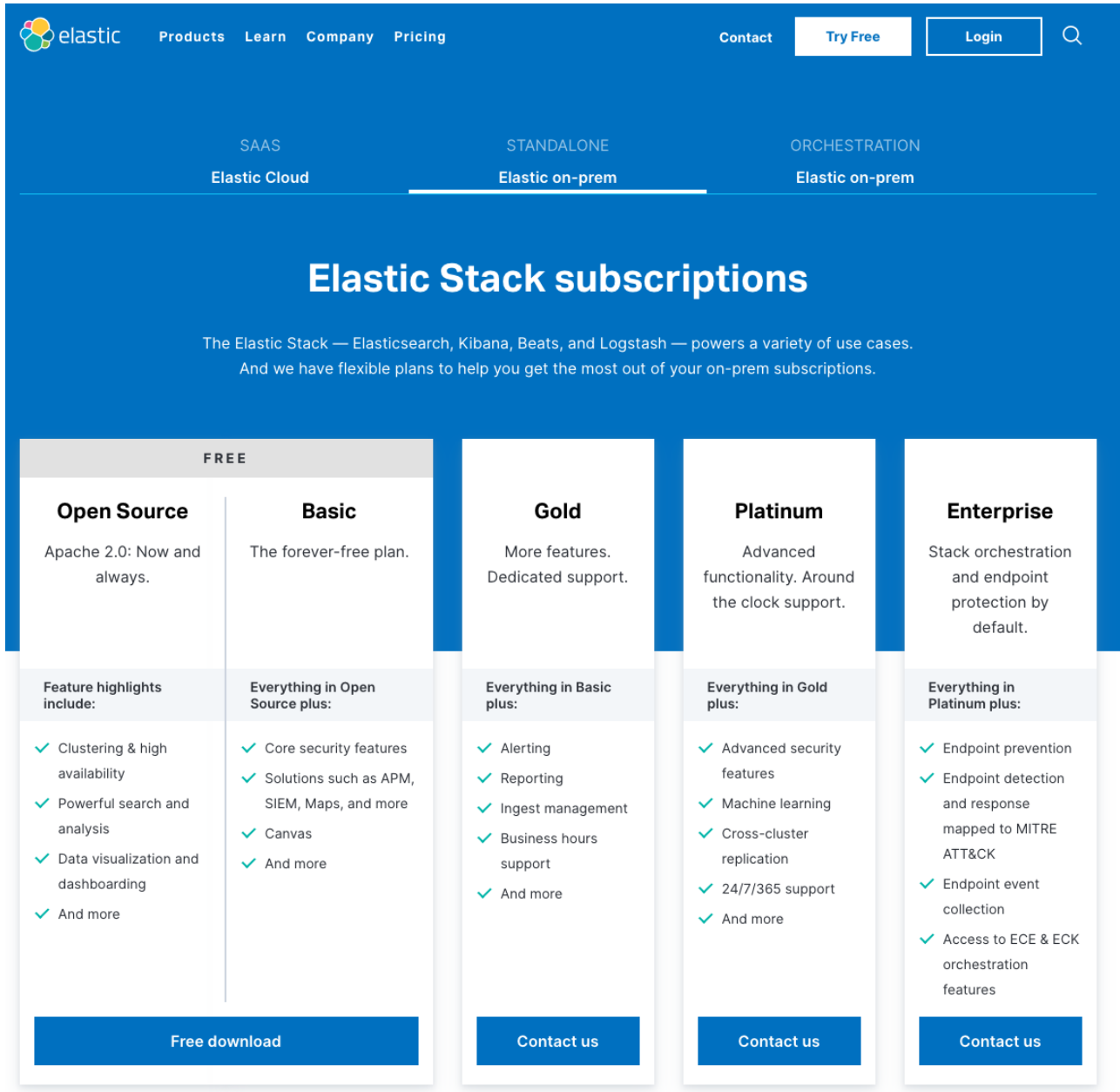
elastic

Elastic Cloud



elastic

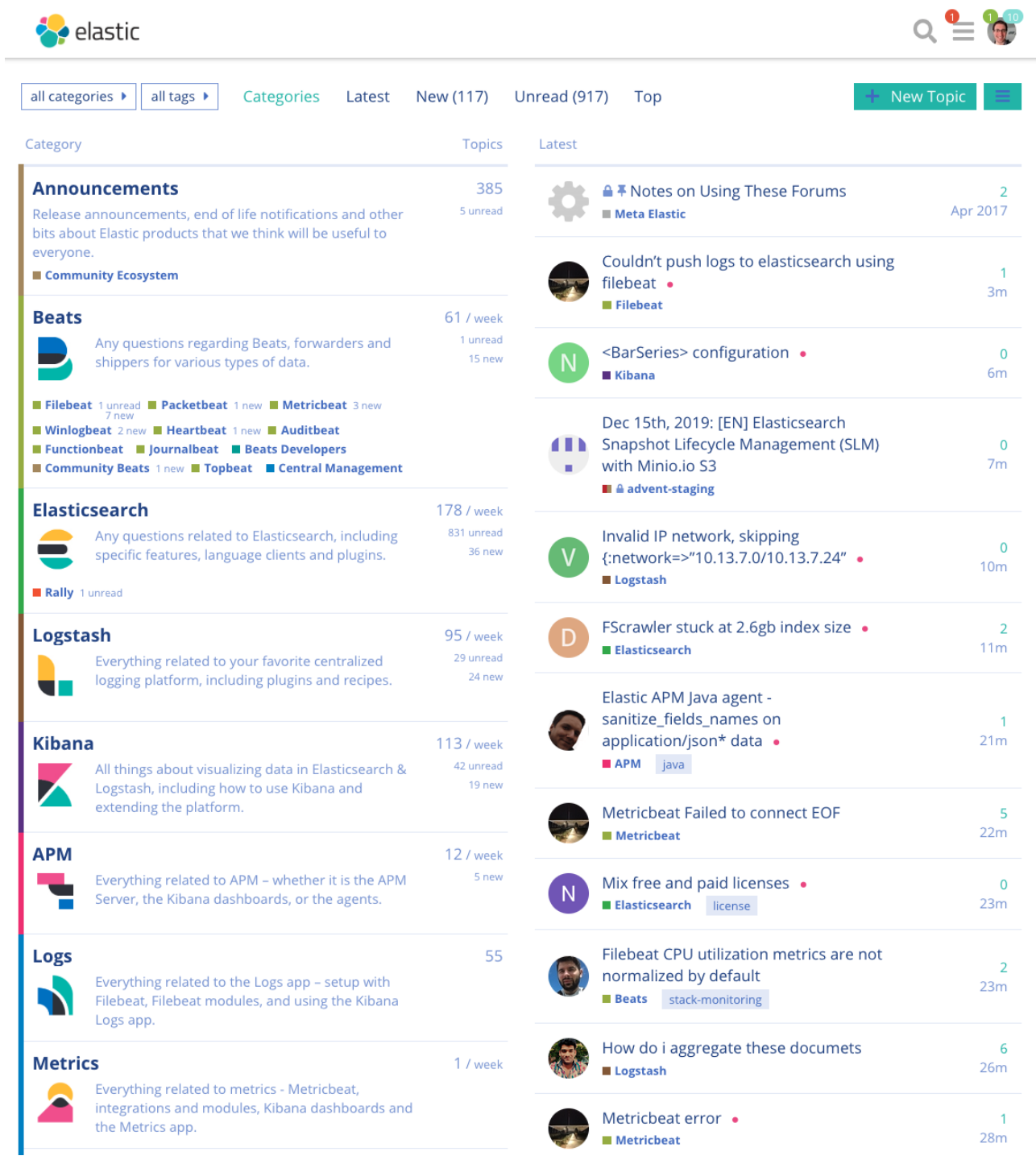
Elastic Support Subscriptions



elastic

Discuss Forum

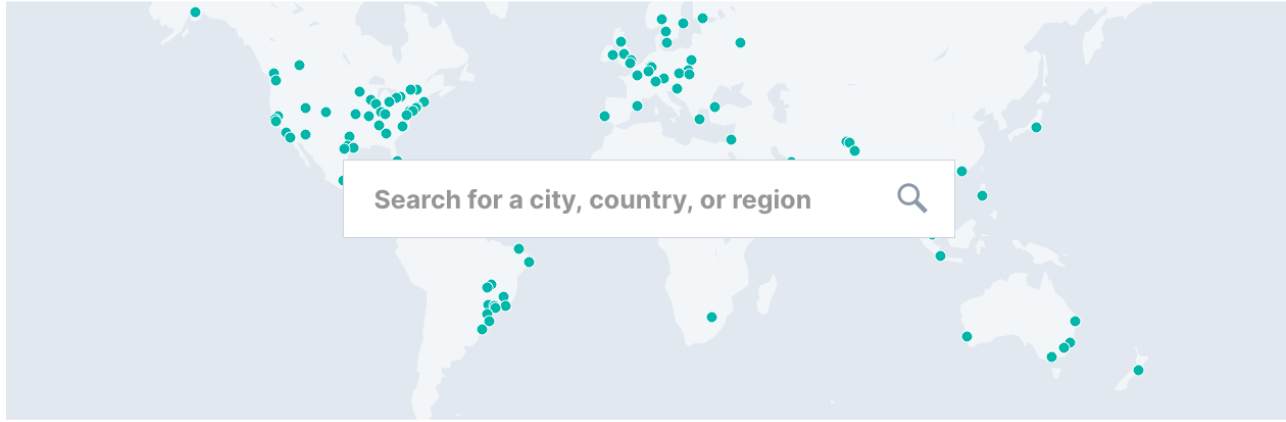
https://discuss.elastic.co



elastic

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

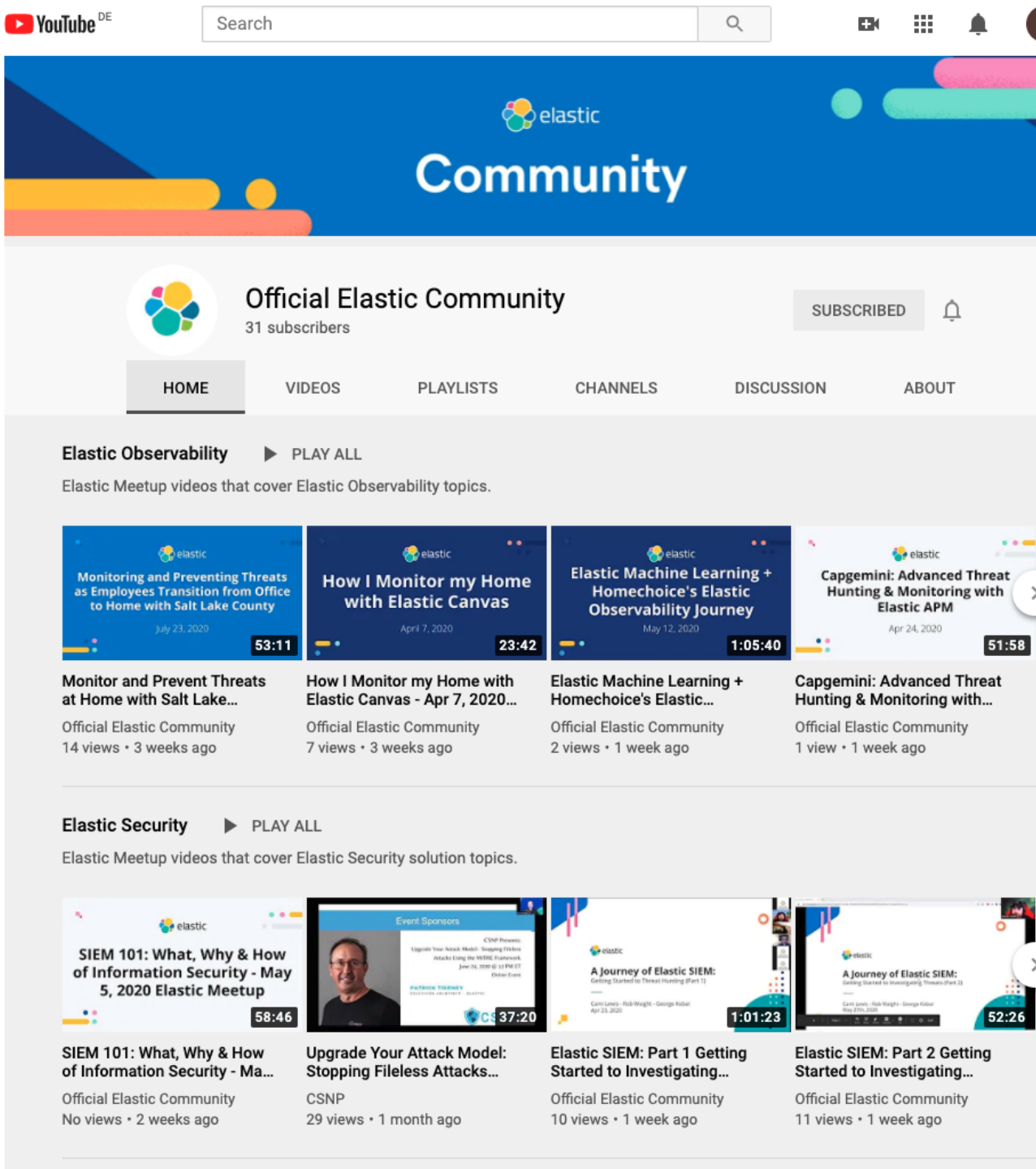
SEARCH MEETUP MUNICH
Germany

elastic

Elastic YouTube Community

https://ela.st/yt-community

elastic



Thanks for listening

Q & A

Alexander Reelsen
Community Advocate
alex@elastic.co | @spinscale

elastic