Elasticsearch，不仅仅是搜索

曾勇 - Elastic

You know， for search!

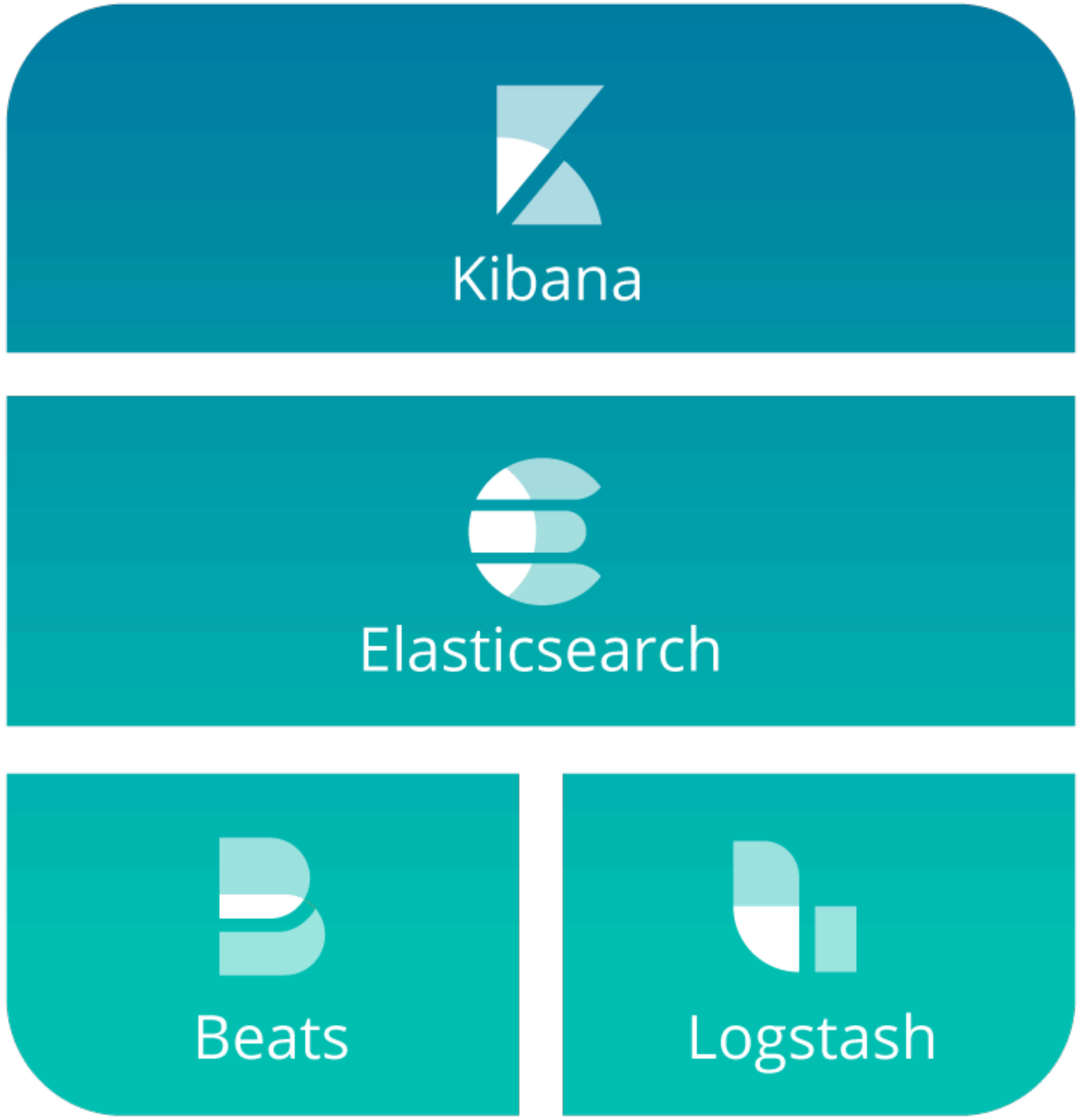
You know， for logging!

Elasticsearch

Logstash

Kibana

Elastic Stack

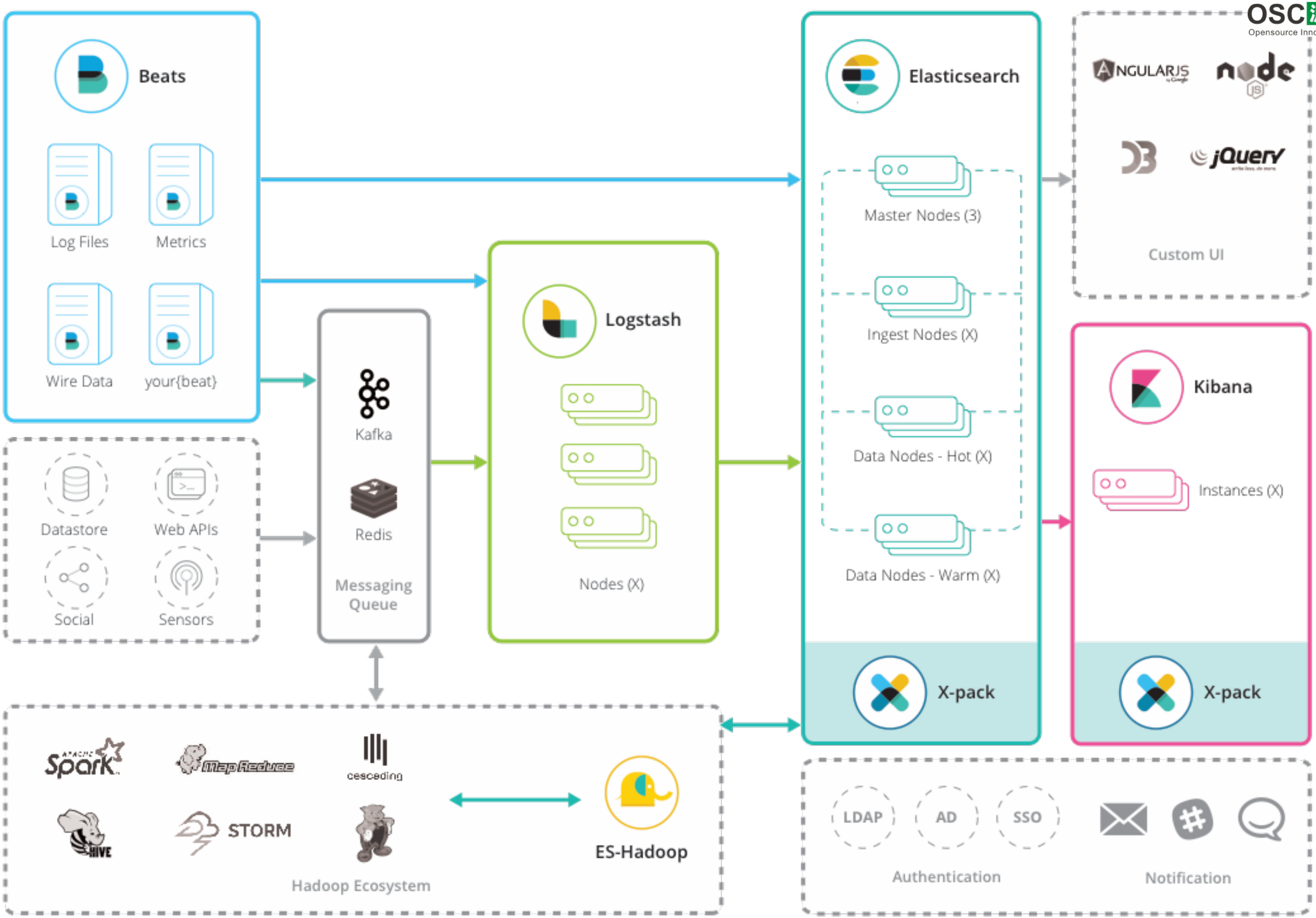


Kibana

Elasticsearch

Beats

Logstash


Beats
Log Files
Metrics
Wire Data
your{beat}
Datastore
Web APIs
Social
Sensors
Kafka
Redis
Messaging Queue
Logstash
Nodes (X)
Elasticsearch
Master Nodes (3)
Ingest Nodes (X)
Data Nodes - Hot (X)
Data Nodes - Warm (X)
X-pack
ANGULARJS
node
D3
jQuery
Custom UI
Kibana
Instances (X)
X-pack
Spark
MapReduce
cascading
HIVE
STORM
Hadoop Ecosystem
ES-Hadoop
LDAP
AD
SSO
Authentication
Notification

You know, …

- You know, for public sentiment analysis！
- You know, for marketing analysis！
- You know, for OLAP analysis！
- You know, for geo analysis!
- You know, for security!
- You know for APM/NPM?
- 。。。

Elastic Cause Awards

@Elastic{ON}17

Ministério Público do Estado de Goiás

CNCFLORA Centro Nacional de Conservação da Flora

Microsoft

MEDECINS SANS FRONTIERES

VALCRI

uc3m

The Nature Conservancy

ISTresearch

NoSchoolViolence.org

cloudpainter

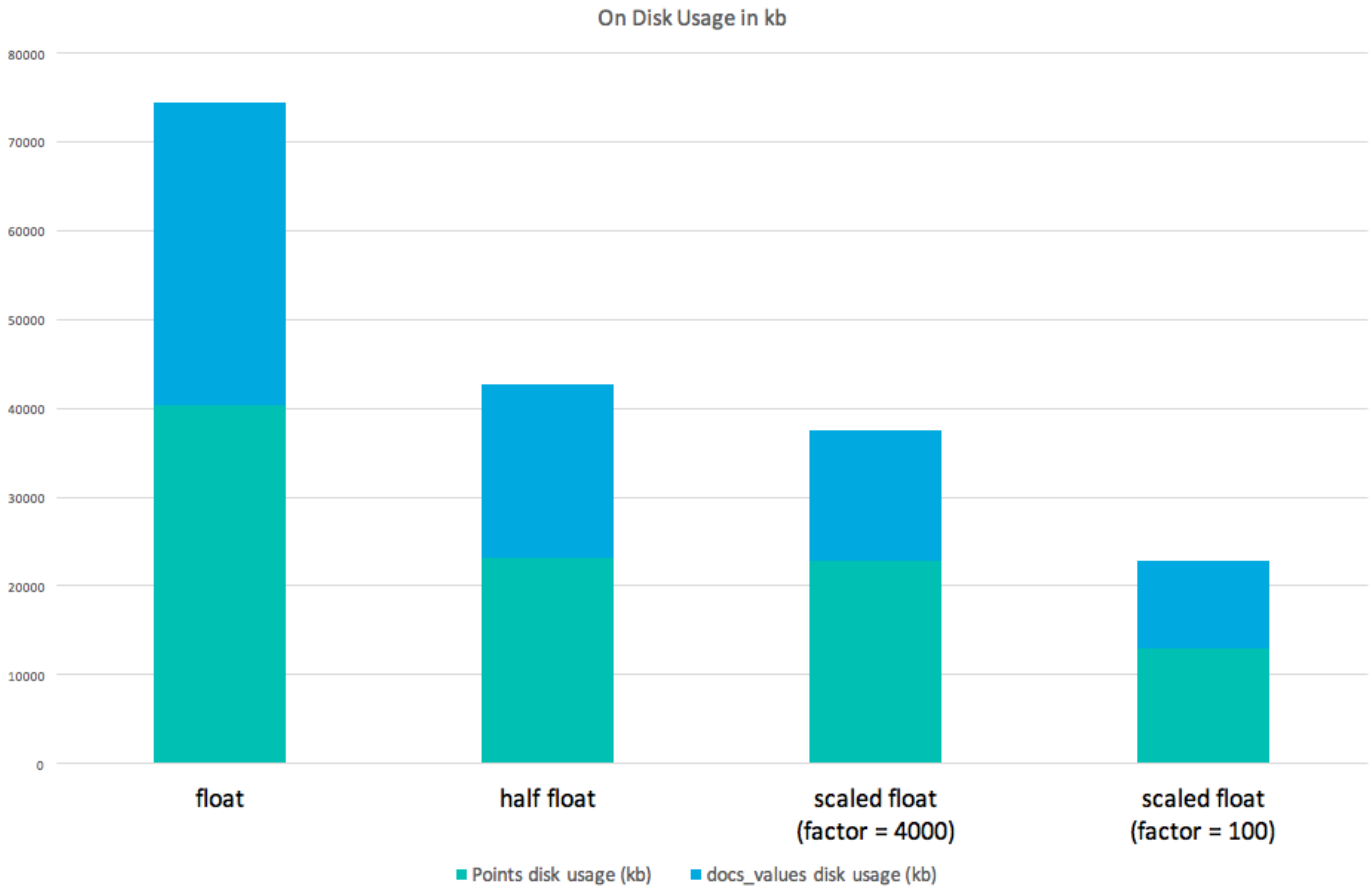
eHealth AFRICA GLOBAL HEALTH INFORMATICS

Elasticsearch 从 5.0 到 6.0

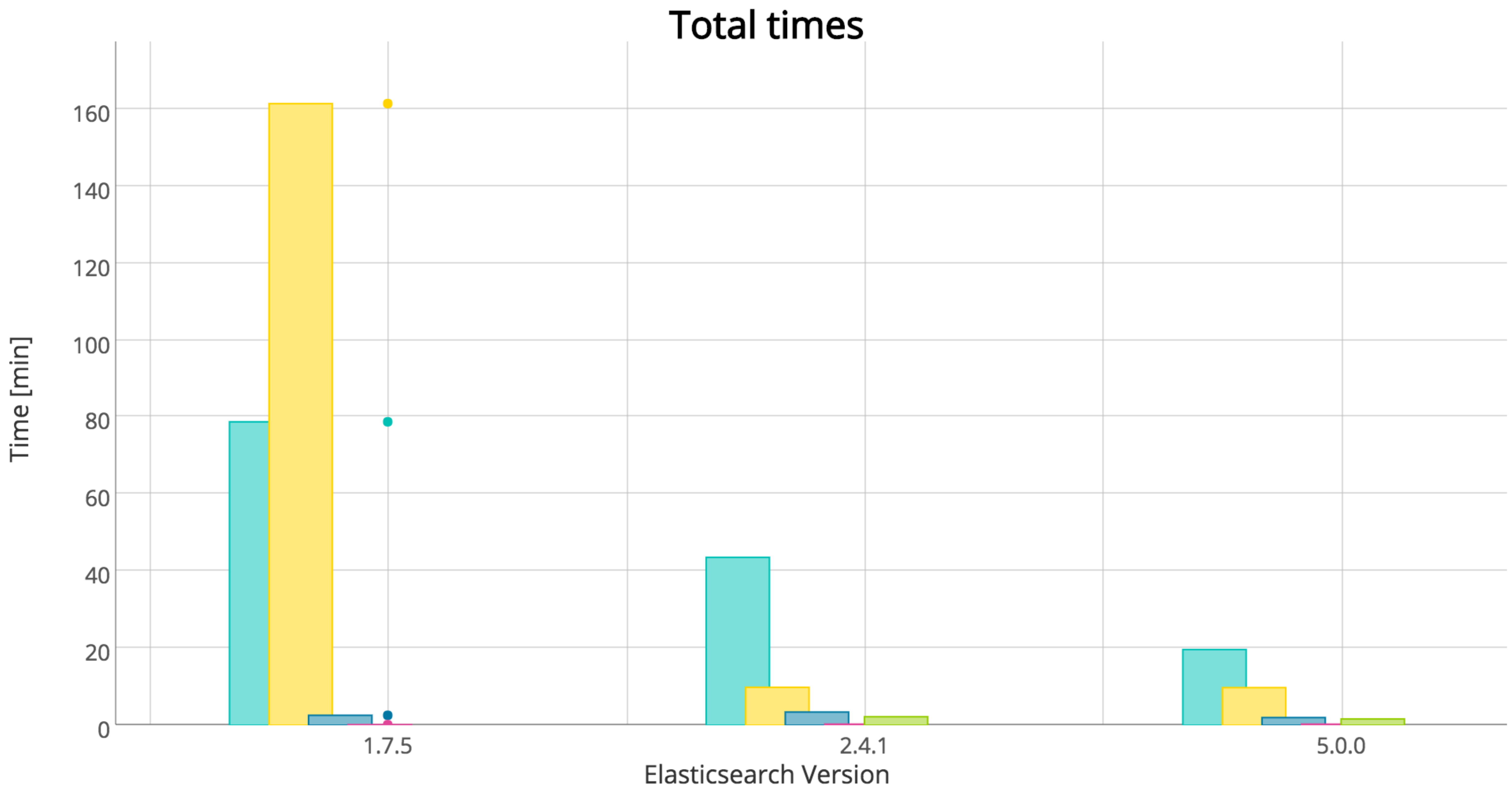
Better Support for Numbers

Faster & reduced memory/disk for many use cases

- BKD trees
- Lower heap usage
- IPv6 support
- Scaled / Half float



Improved Indexing Time Performance



1.7.5:

Indexing time (min): 78.8

Merge time (min): 161.44

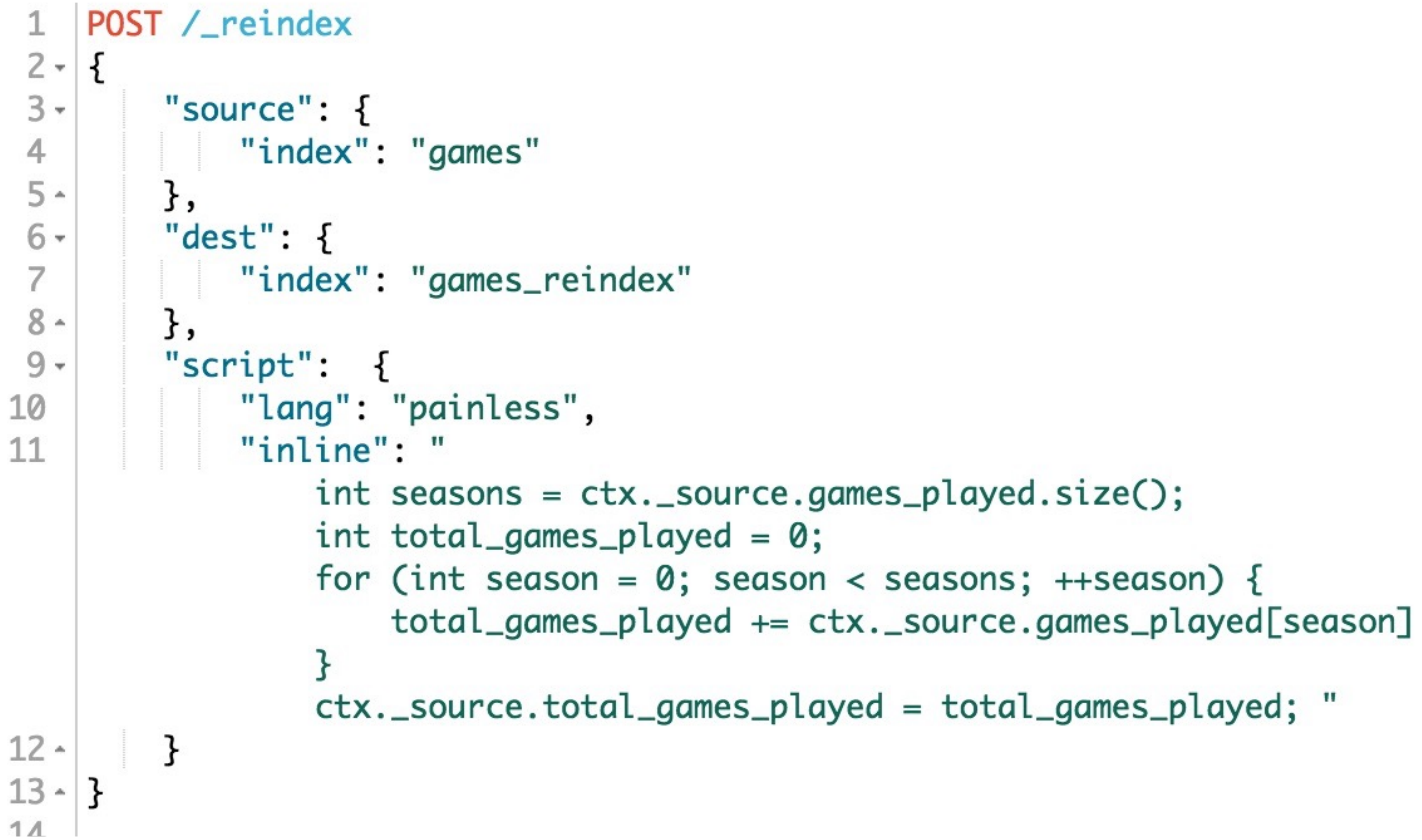
Refresh time (min): 2.58

Flush time (min): 0.15

Fast, Safe Scripting Language

Say “Heya” to Painless

- Secure and production-safe
- Significantly faster than Groovy
- Familiar syntax
- Can be used in various places:
    - ingest node pipeline, function scoring more

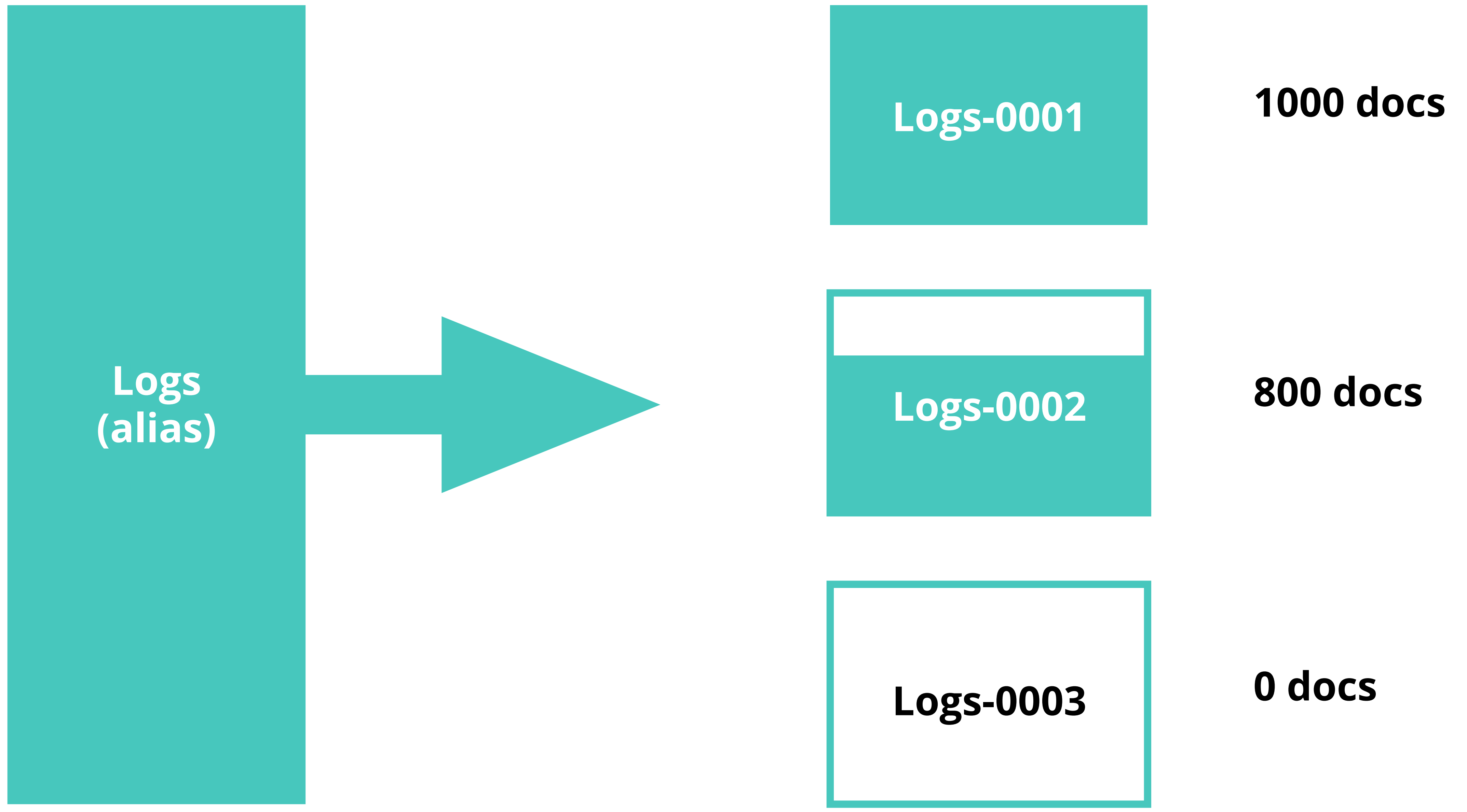


```
POST /_reindex
{
    "source": {
        "index": "games"
    },
    "dest": {
        "index": "games_reindex"
    },
    "script":  {
        "lang": "painless",
        "inline": "
            int seasons = ctx._source.games_played.size();
            int total_games_played = 0;
            for (int season = 0; season < seasons; ++season) {
                total_games_played += ctx._source.games_played[season]
            }
            ctx._source.total_games_played = total_games_played; "
    }
}
```

Simplified Architecture

Automatic time-based index management

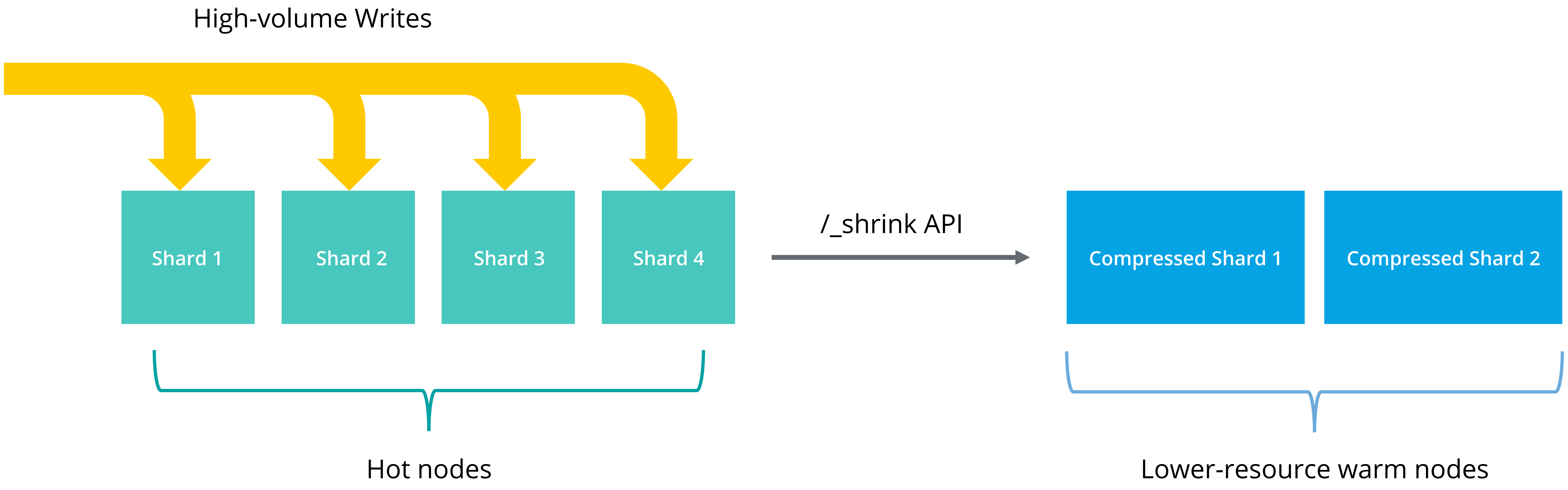
- **Rollover APIs**



Simplified Architecture

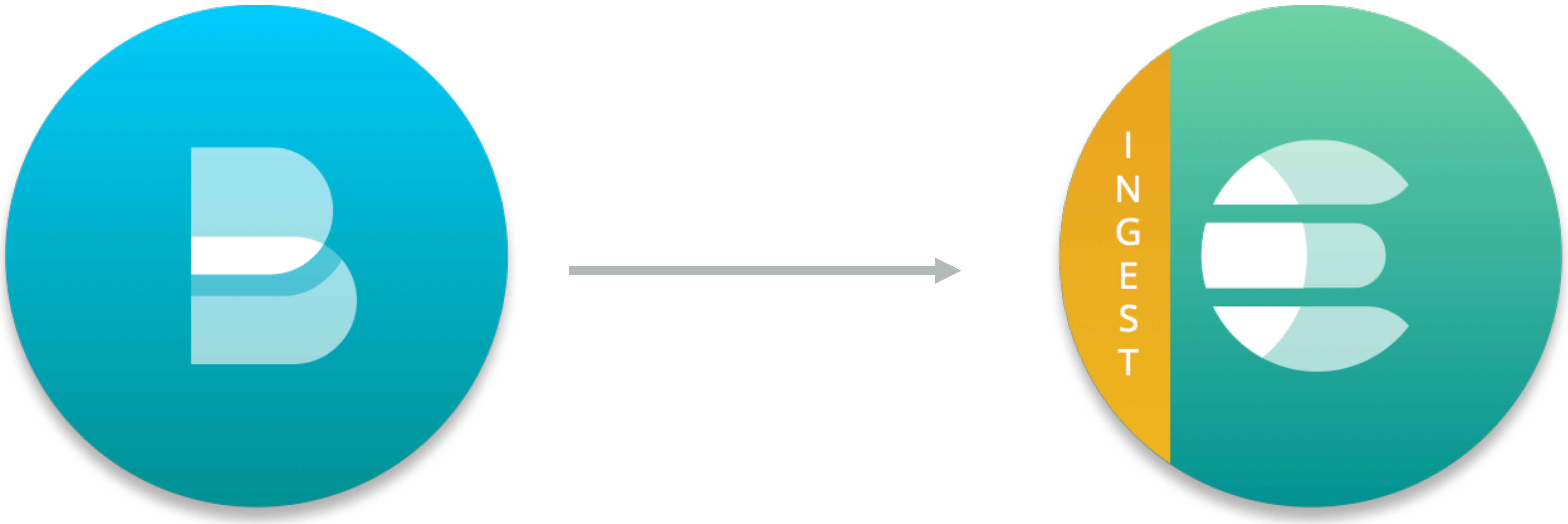
Automatic time-based index management

- **Shrink APIs**



Say Heya to Ingest Node

Process incoming data directly in Elasticsearch

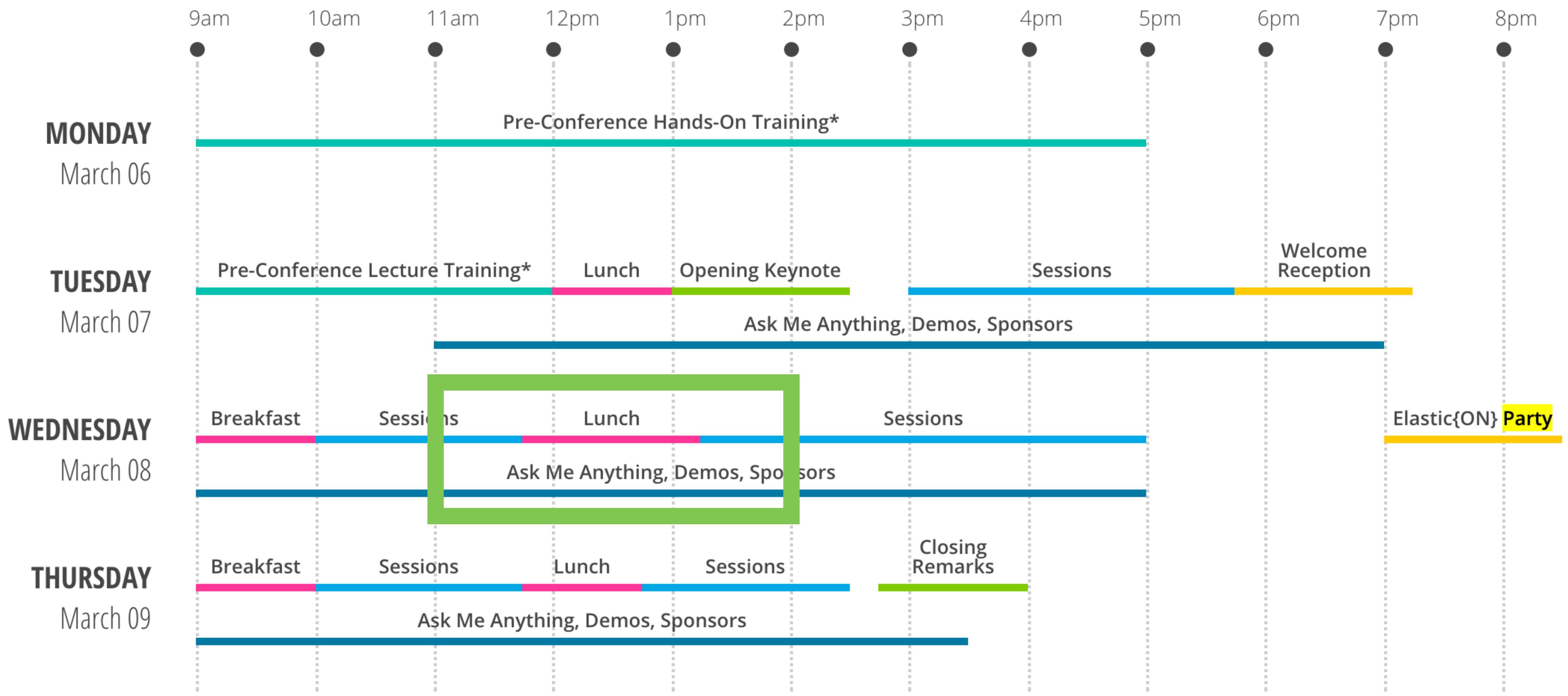


Numeric & Date Range Fields

Mapping Improvements

- New types for date/number ranges (5.2)
  (date_range, int_range, float_range)

What's happening Wednesday 11am - 2pm



9am 10am 11am 12pm 1pm 2pm 3pm 4pm 5pm 6pm 7pm 8pm

MONDAY
March 06
Pre-Conference Hands-On Training*

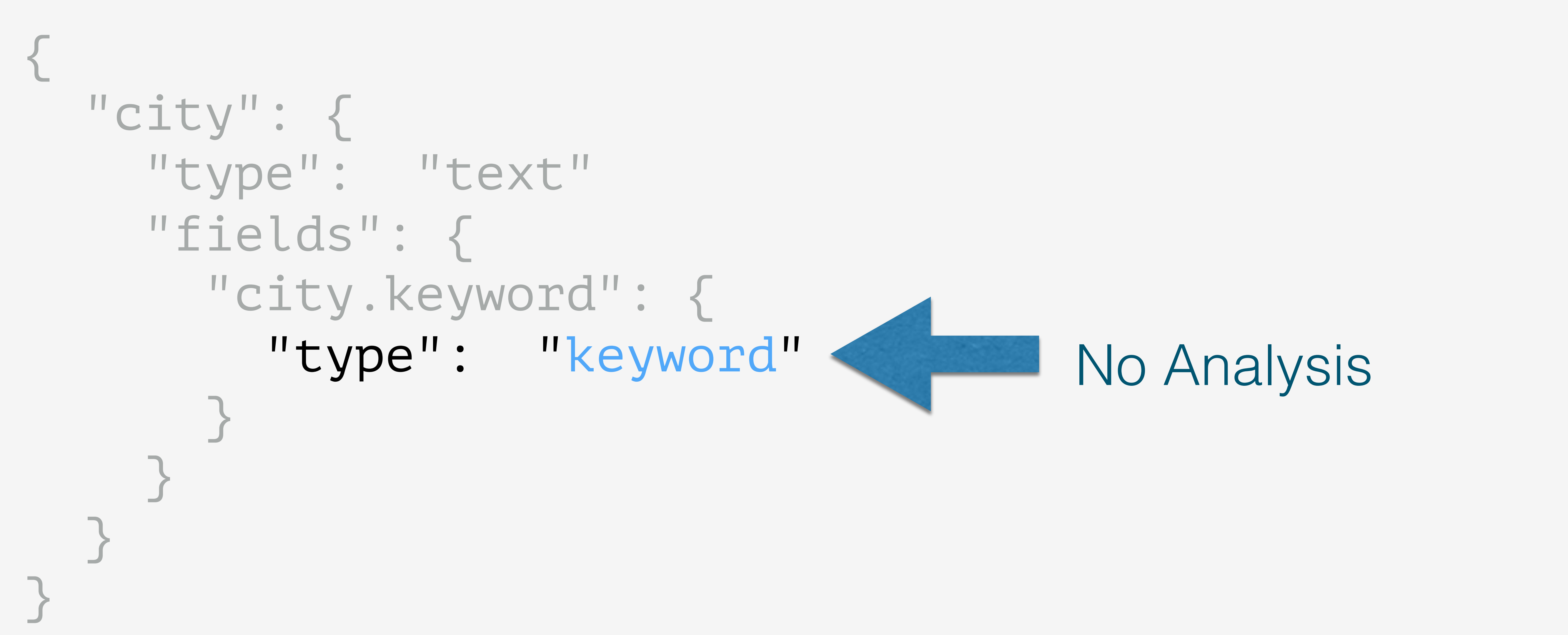
TUESDAY
March 07
Pre-Conference Lecture Training* · Lunch · Opening Keynote · Sessions · Welcome Reception
Ask Me Anything, Demos, Sponsors

WEDNESDAY
March 08
Breakfast · Sessions · Lunch · Sessions · Elastic{ON} Party
Ask Me Anything, Demos, Sponsors

THURSDAY
March 09
Breakfast · Sessions · Lunch · Sessions · Closing Remarks
Ask Me Anything, Demos, Sponsors

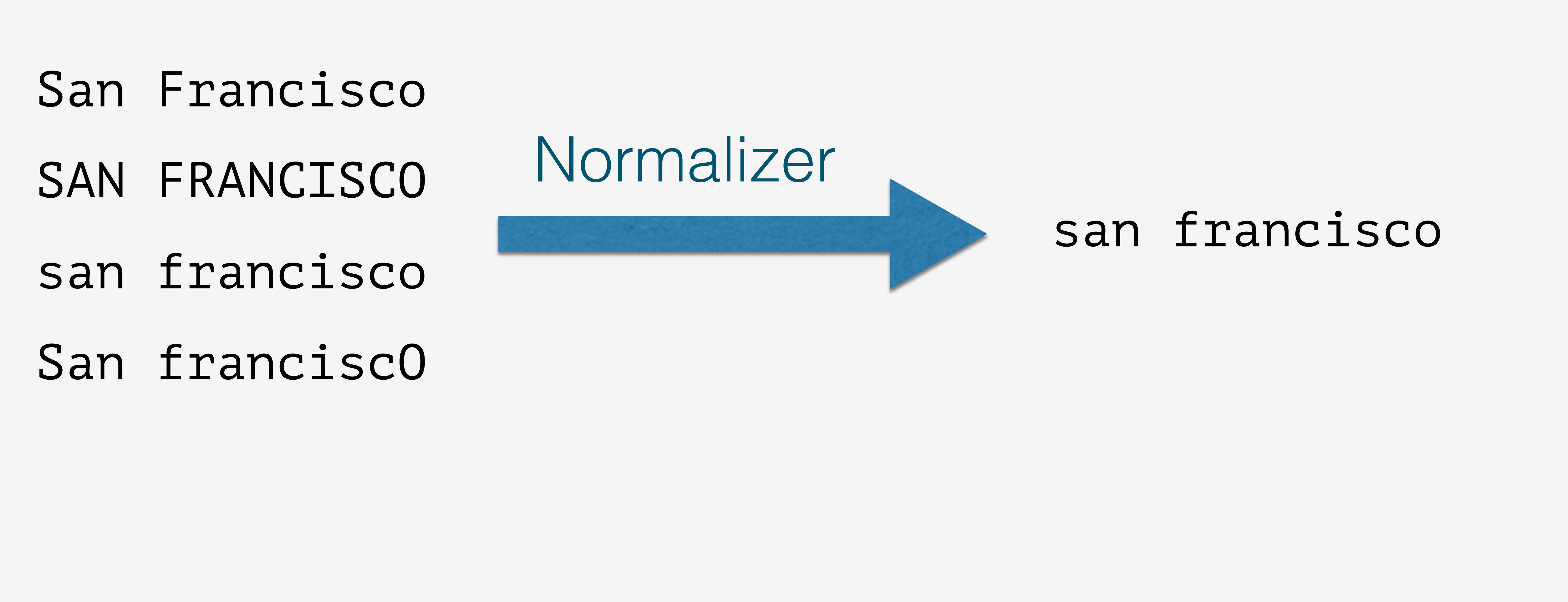
Keyword Normalizer

Mapping Improvements



```
{
  "city": {
    "type":  "text"
    "fields": {
      "city.keyword": {
        "type":  "keyword"
      }
    }
  }
}
```

No Analysis



San Francisco

SAN FRANCISCO

san francisco

San franciscO

Normalizer

san francisco

Terms Aggregation Partitioning

Returning ALL the Terms, in Manageable Chunks

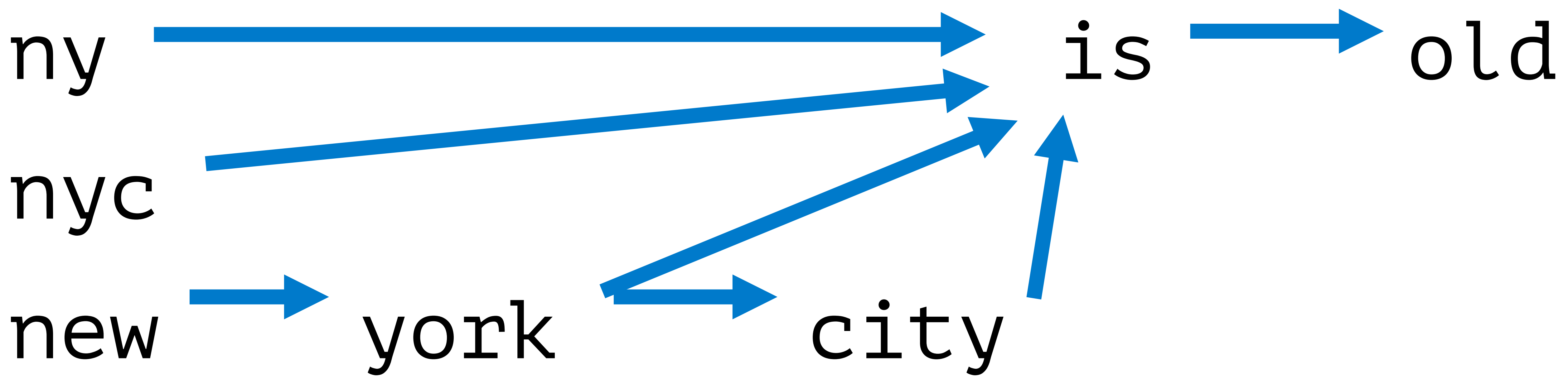
- frequent request
- return all responses from a terms aggs
- Terms can now be broken into partitions and partitions are returned by number

```
{
   "size": 0,
   "aggs": {
      "expired_sessions": {
         "terms": {
            "field": "account_id",
            "include": {
               "partition": 0,
               "num_partitions": 20
            },
            "size": 10000,
            "order": {
               "last_access": "asc"
            }
         },
         "aggs": {
            "last_access": {
               "max": {
                  "field": "access_date"
               }
            }
         }
      }
   }
}
```

Synonym Graph Token Filter

Search & Aggregation Improvements

- Improved querying for multi-word synonyms `SynonymGraphFilter`



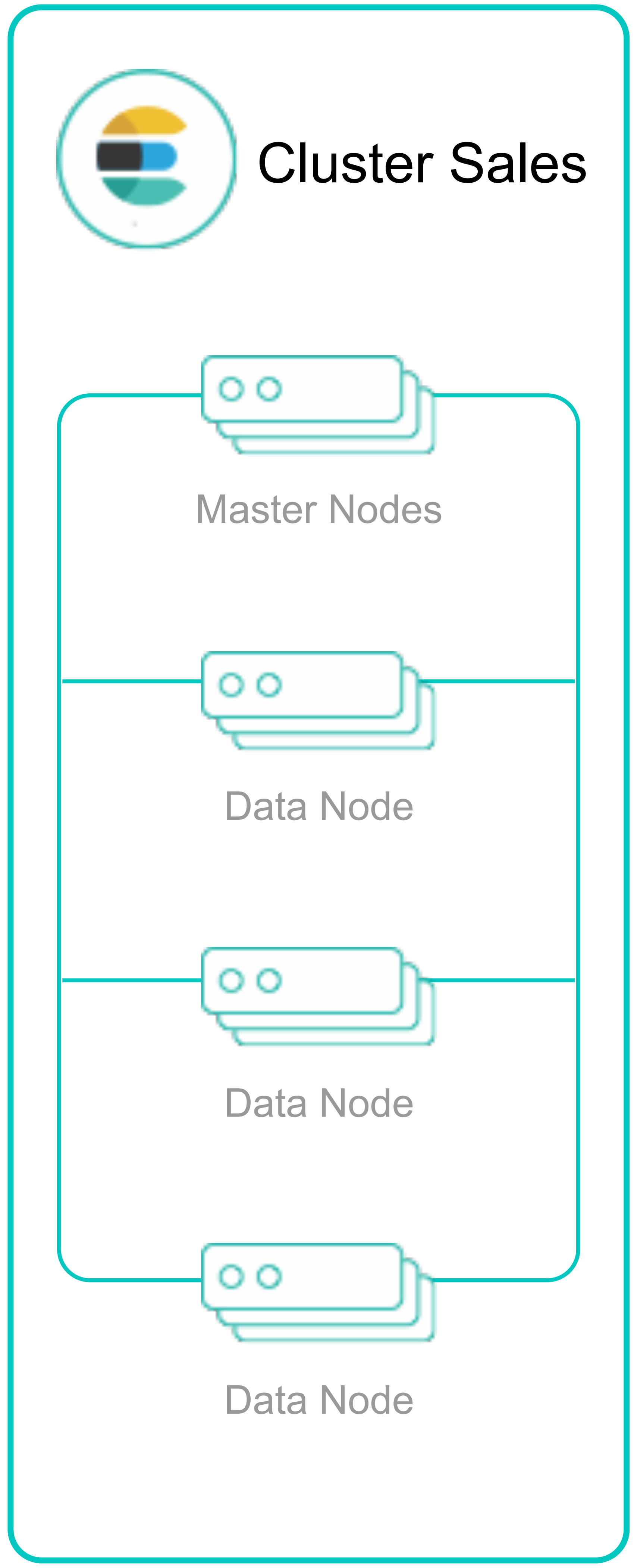
Cluster Allocation Explain API

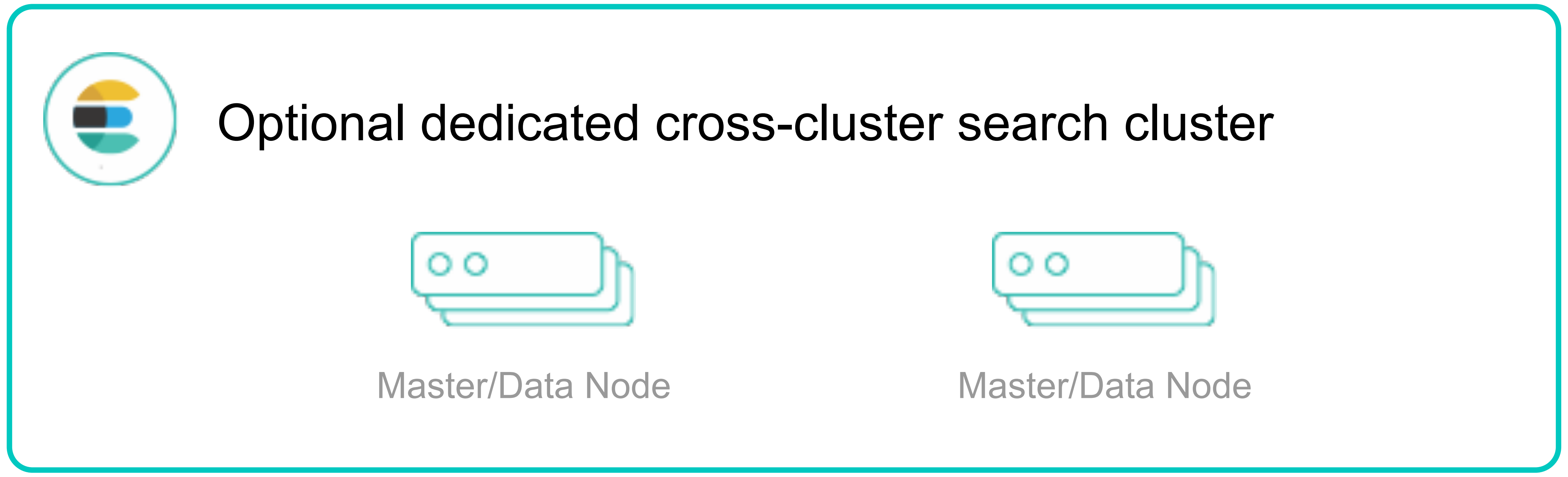
Operational Optimizations - Understand Shard Allocation

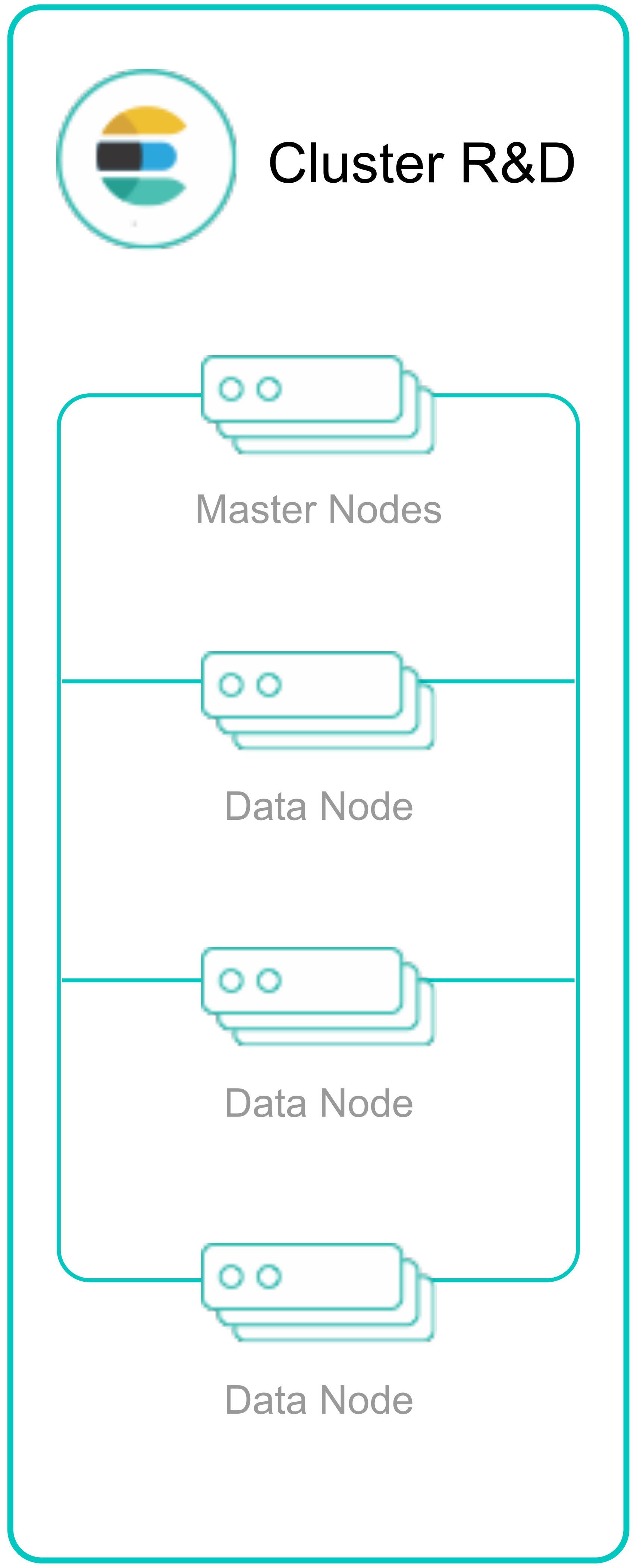
```
/_cluster/allocation/explain
```

- Diagnose unassigned shards
- clear human readable descriptions when things fail

Cross-Cluster search



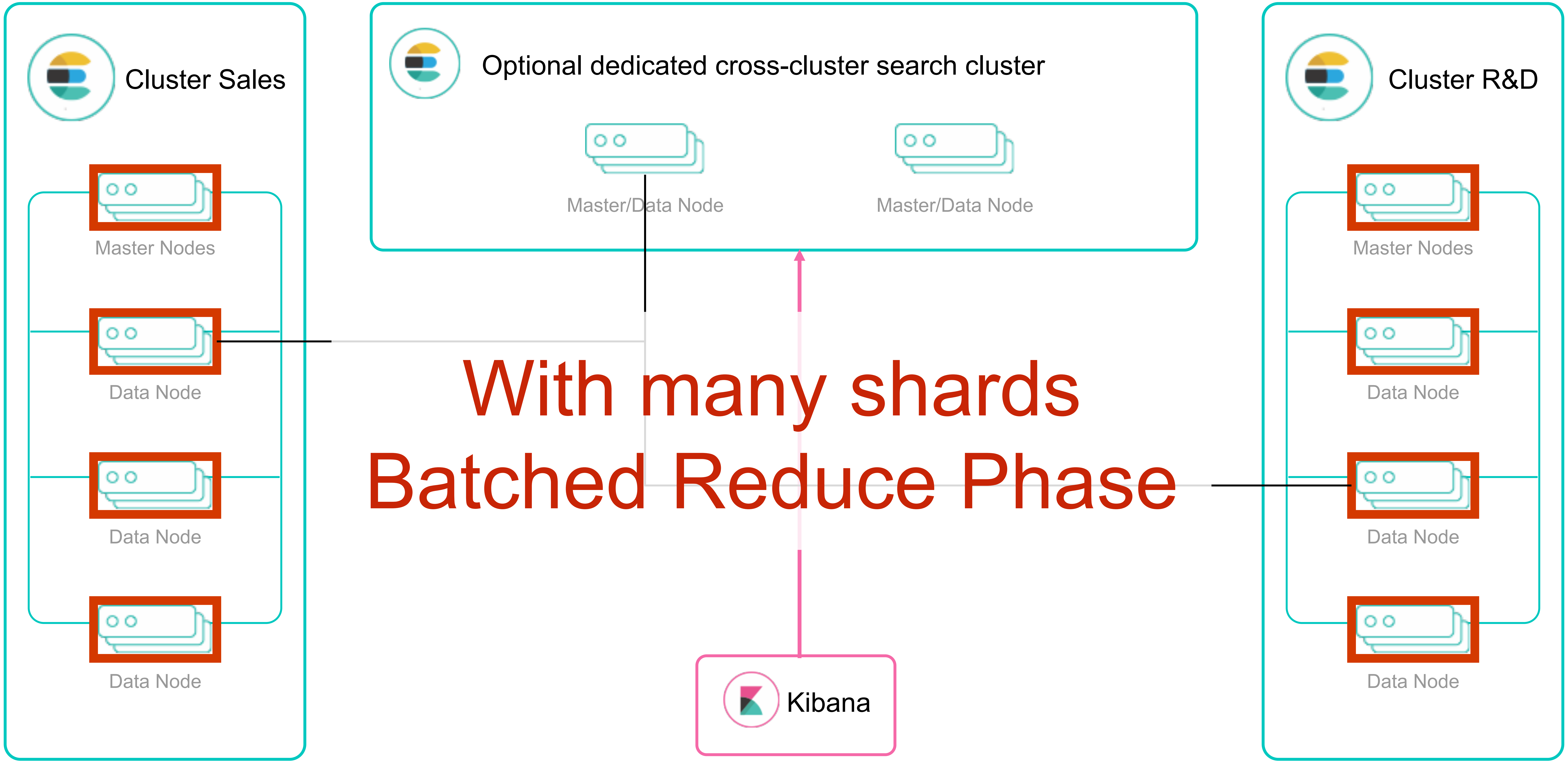




```
PUT _cluster/settings
{
  "transient": {
    "search.remote": {
      "sales.seeds":   "10.0.0.1:9300",
      “r_and_d.seeds”: "10.1.0.1:9300"
    }
  }
}
```

Dynamic settings

Cross-Cluster search



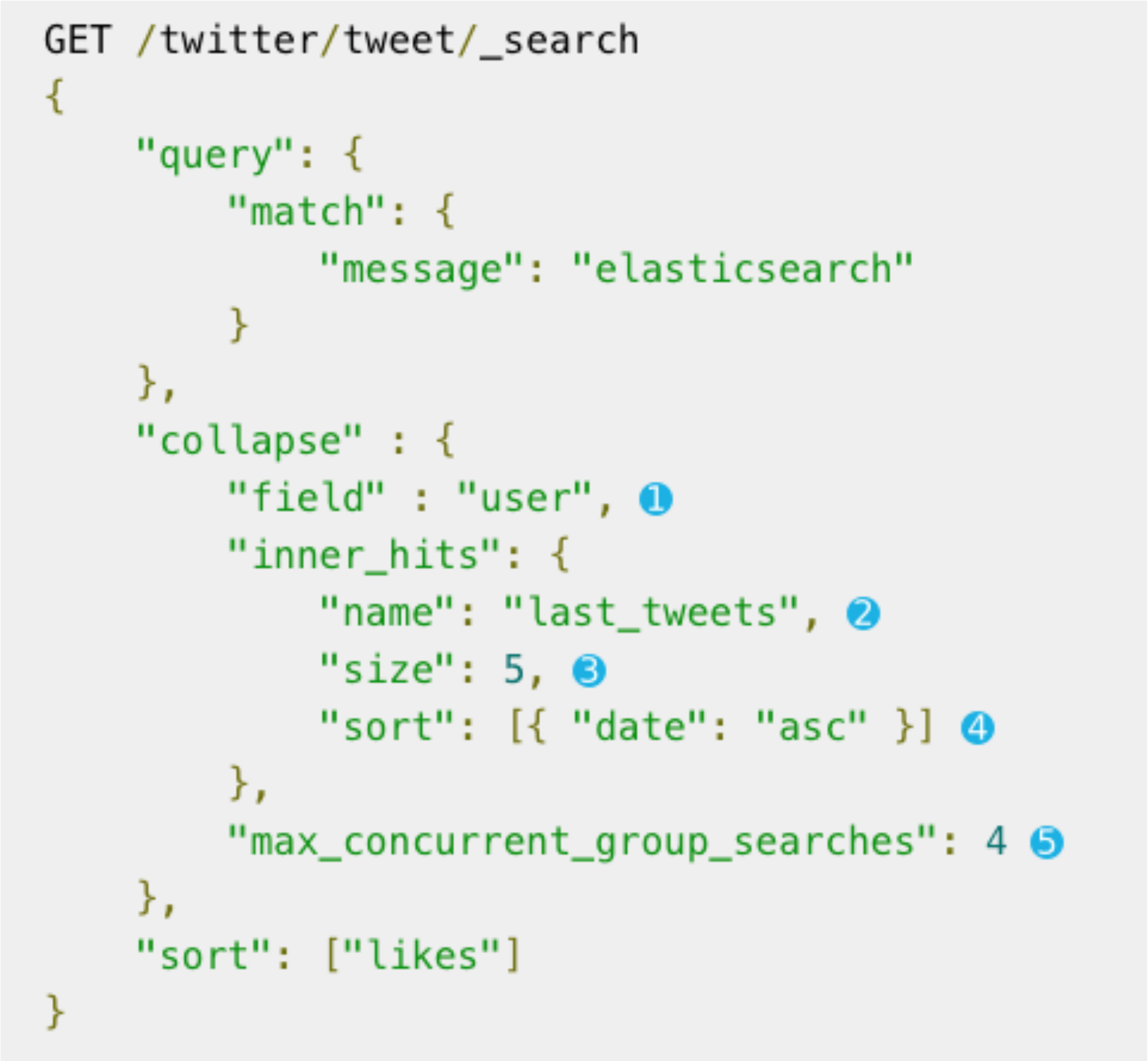
With many shards

Batched Reduce Phase

Field Collapsing

One method to rule them all...

- Simple (almost) no setup!
- Great for query-time group/category de-dup



```
GET /twitter/tweet/_search
{
    "query": {
        "match": {
            "message": "elasticsearch"
        }
    },
    "collapse" : {
        "field" : "user", ❶
        "inner_hits": {
            "name": "last_tweets", ❷
            "size": 5, ❸
            "sort": [{ "date": "asc" }] ❹
        },
        "max_concurrent_group_searches": 4 ❺
    },
    "sort": ["likes"]
}
```

Elasticsearch Keystore

If you like it, you should put it in a keystore.

- Sensitive settings should not be protected by filesystem permissions only.
- Commands feel familiar:
  - bin/elasticsearch-keystore create
  - bin/elasticsearch-keystore list
  - bin/elasticsearch-keystore add the.setting.name.to.set
  - bin/elasticsearch-keystore remove the.setting.name.to.remove
- Just the framework/start: sensitive settings to be pulled in

And many more …

- Batched reduction of search results
- Smarter query caching
- Faster geo, range, and nested queries
- Unified highlighter
- Cancellable searches
- More Painless improvements
- Index partitioning/routing
- Adjacency matrix

Elasticsearch 6.0 is coming

- Remove Type

  Sparse Doc Values

  Index Sorting

  Sequence Numbers

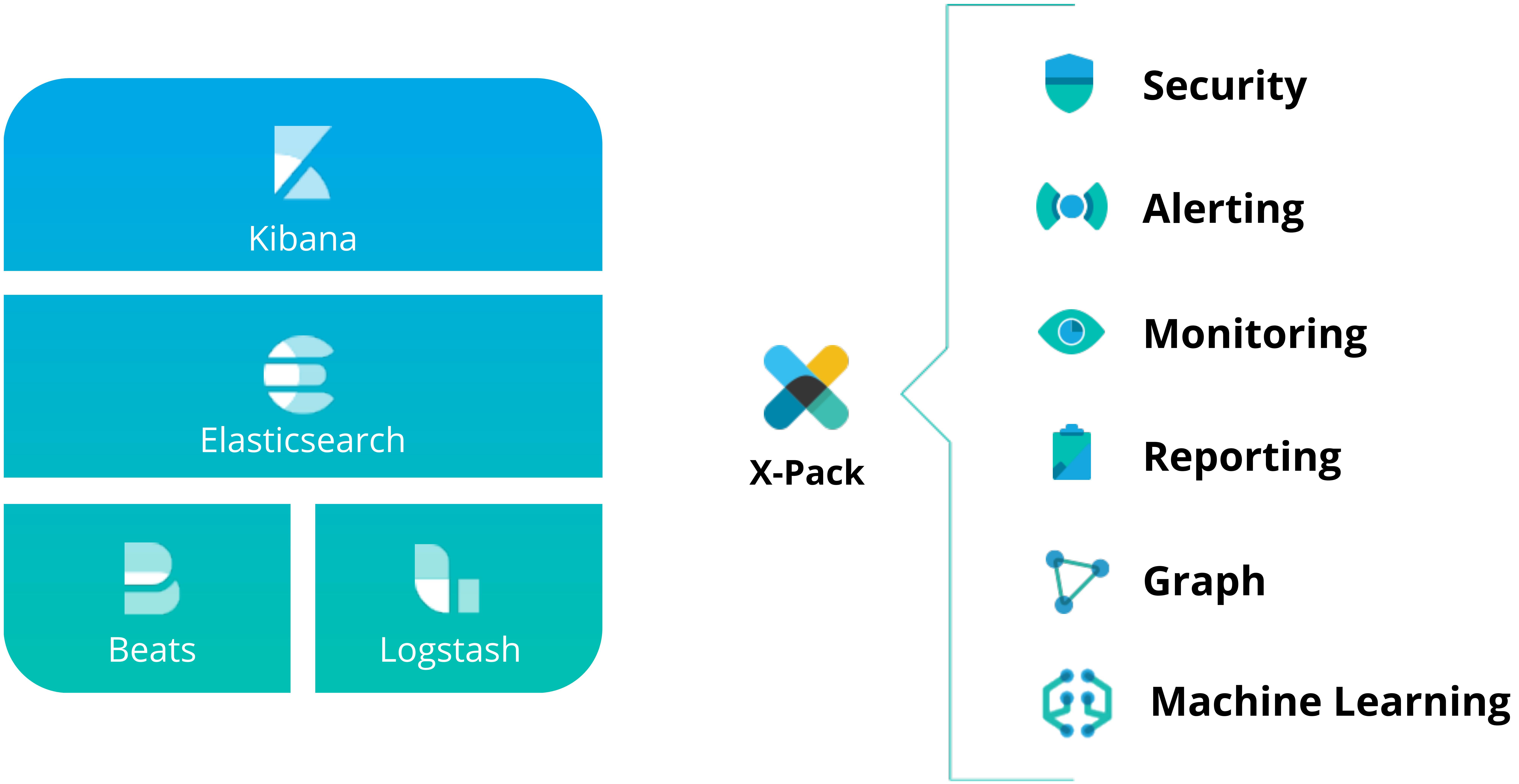
  Rolling Upgrades

  …

BOLDLY TAKING THE ELASTIC STACK WHERE NO STACK HAS GONE BEFORE
elastic
PIONEERS
5.0

2017.05.09
Elastic Stack 6.0.0-alpha1 Released

X-Pack



Kibana

Elasticsearch

Beats

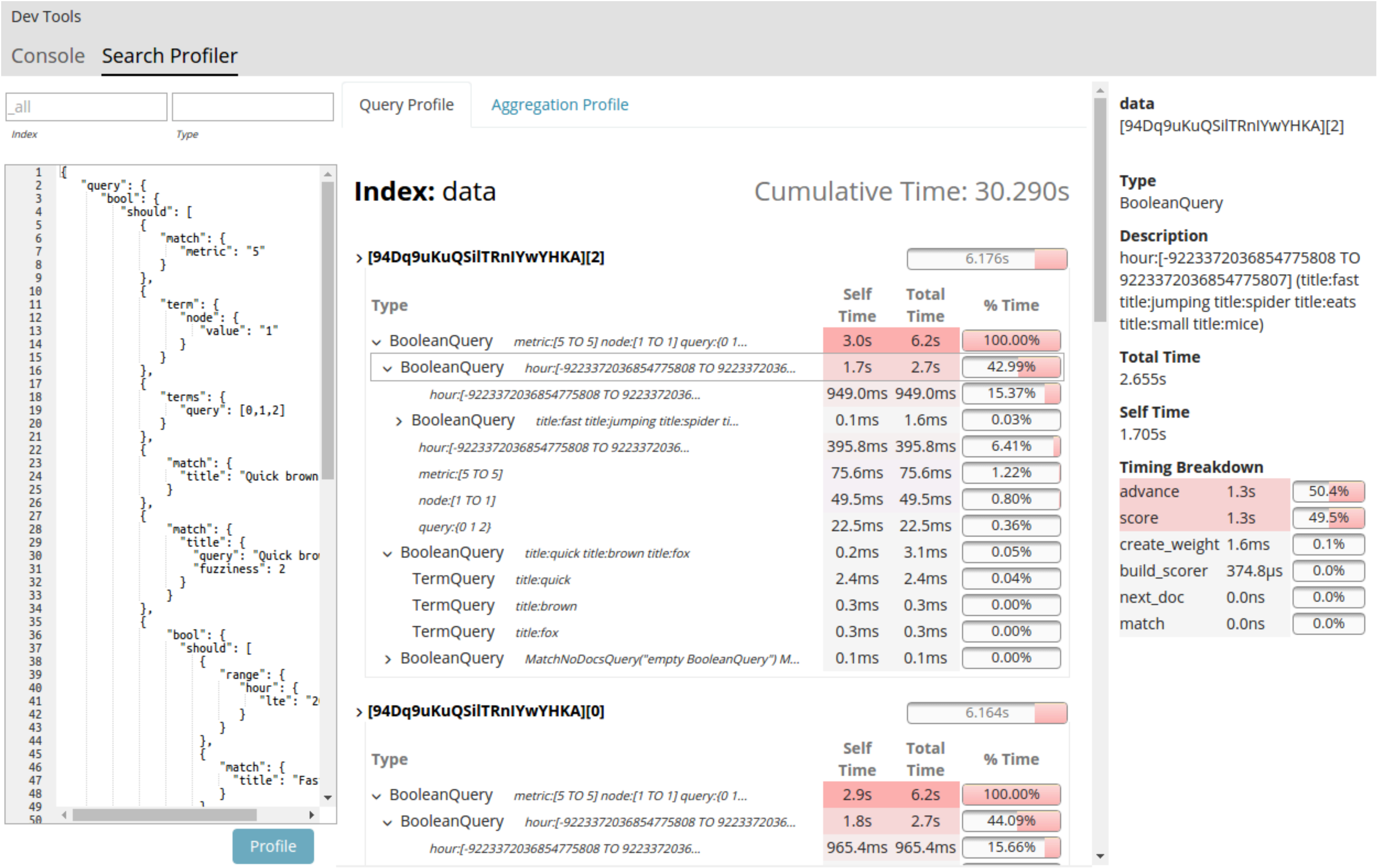
Logstash

X-Pack

- Security
- Alerting
- Monitoring
- Reporting
- Graph
- Machine Learning

Profile your Search Queries

Search Profiler (5.1) - Detect and visualize bottlenecks in your query



```
{
    "query": {
        "bool": {
            "should": [
                {
                    "match": {
                        "metric": "5"
                    }
                },
                {
                    "term": {
                        "node": {
                            "value": "1"
                        }
                    }
                },
                {
                    "terms": {
                        "query": [0,1,2]
                    }
                },
                {
                     "match": {
                        "title": "Quick brown
                     }
                },
                {
                     "match": {
                        "title": {
                          "query": "Quick bro
                          "fuzziness": 2
                        }
                     }
                },
                {
                     "bool": {
                        "should": [
                           {
                              "range": {
                                 "hour": {
                                    "lte": "2
                                 }
                              }
                           },
                           {
                              "match": {
                                 "title": "Fas
                              }
```

Index: data　　Cumulative Time: 30.290s

› [94Dq9uKuQSilTRnIYwYHKA][2]　6.176s

Type		Self Time	Total Time	% Time
⌄ BooleanQuery	*metric:[5 TO 5] node:[1 TO 1] query:{0 1...*	3.0s	6.2s	100.00%
⌄ BooleanQuery	*hour:[-9223372036854775808 TO 9223372036...*	1.7s	2.7s	42.99%
	hour:[-9223372036854775808 TO 9223372036...	949.0ms	949.0ms	15.37%
› BooleanQuery	*title:fast title:jumping title:spider ti...*	0.1ms	1.6ms	0.03%
	hour:[-9223372036854775808 TO 9223372036...	395.8ms	395.8ms	6.41%
	metric:[5 TO 5]	75.6ms	75.6ms	1.22%
	node:[1 TO 1]	49.5ms	49.5ms	0.80%
	query:{0 1 2}	22.5ms	22.5ms	0.36%
⌄ BooleanQuery	*title:quick title:brown title:fox*	0.2ms	3.1ms	0.05%
TermQuery	*title:quick*	2.4ms	2.4ms	0.04%
TermQuery	*title:brown*	0.3ms	0.3ms	0.00%
TermQuery	*title:fox*	0.3ms	0.3ms	0.00%
› BooleanQuery	*MatchNoDocsQuery("empty BooleanQuery") M...*	0.1ms	0.1ms	0.00%

› [94Dq9uKuQSilTRnIYwYHKA][0]　6.164s

Type		Self Time	Total Time	% Time
⌄ BooleanQuery	*metric:[5 TO 5] node:[1 TO 1] query:{0 1...*	2.9s	6.2s	100.00%
⌄ BooleanQuery	*hour:[-9223372036854775808 TO 9223372036...*	1.8s	2.7s	44.09%
	hour:[-9223372036854775808 TO 9223372036...	965.4ms	965.4ms	15.66%

data
[94Dq9uKuQSilTRnIYwYHKA][2]

Type
BooleanQuery

Description
hour:[-9223372036854775808 TO 9223372036854775807] (title:fast title:jumping title:spider title:eats title:small title:mice)

Total Time
2.655s

Self Time
1.705s

Timing Breakdown

advance	1.3s	50.4%
score	1.3s	49.5%
create_weight	1.6ms	0.1%
build_scorer	374.8µs	0.0%
next_doc	0.0ns	0.0%
match	0.0ns	0.0%

* requires X-Pack (Basic)

Machine Learning

UNSUPERVISED MACHINE LEARNING

- Automatically detect anomalies
- Advanced correlation and categorization
- Identify root cause(s)
- Expose early warning signs

NEW USE CASES

- Analyze time series data
- Expand security, IT Ops, fraud, finance, and many more use cases
- Currently beta; building a more native integration into the Elastic Stack

Jobs
Summary view
Explorer
Time series
Connections
Support
~ a year ago to ~ a year
OSC源创会
Opensource Innovation Meetup
IT大咖说
知识分享平台
Job

it_ops-kpi and 2 others

Top Influencers

service

inventory-us-east-1-34	94	97
auth-us-west-1-1e	7	51
test-srv-02	5	29
elasticsearch-22	3	16
elasticsearch-77	2	11
payment-srv-21	2	6
payment-srv-11	1	5
backup-srv-13	1	9
test-srv-01	1	7
inventory-us-west-1-4e	1	7

hostname

MSSQL-0783E4076	94	1237

Anomaly timeline

Overall
Jan 28 16:00
Jan 30 08:00
Feb 01 00:00
Feb 02 16:00
Feb 04 08:00
Feb 06 00:00
Feb 07 16:00
Feb 09 08:00
Feb 11 00:00
Feb 12 16:00

View by: job ID (Top 10 by max anomaly score)

it_ops-service
it_ops-sql
it_ops-kpi
Jan 28 16:00
Jan 30 08:00
Feb 01 00:00
Feb 02 16:00
Feb 04 08:00
Feb 06 00:00
Feb 07 16:00
Feb 09 08:00
Feb 11 00:00
Feb 12 16:00

Anomalies

mean responsetime service inventory-us-east-1-34

mean SQLServer_Buffer_Manager_Buffer_cache_hit_ratio hostname MSSQL-0783E4076

mean SQLServer_General_Statistics_User_Connections hostname MSSQL-0783E4076

mean SQLServer_SQL_Statistics_Batch_Requests_sec hostname MSSQL-0783E4076

Elasticsearch-SQL Coming soon!

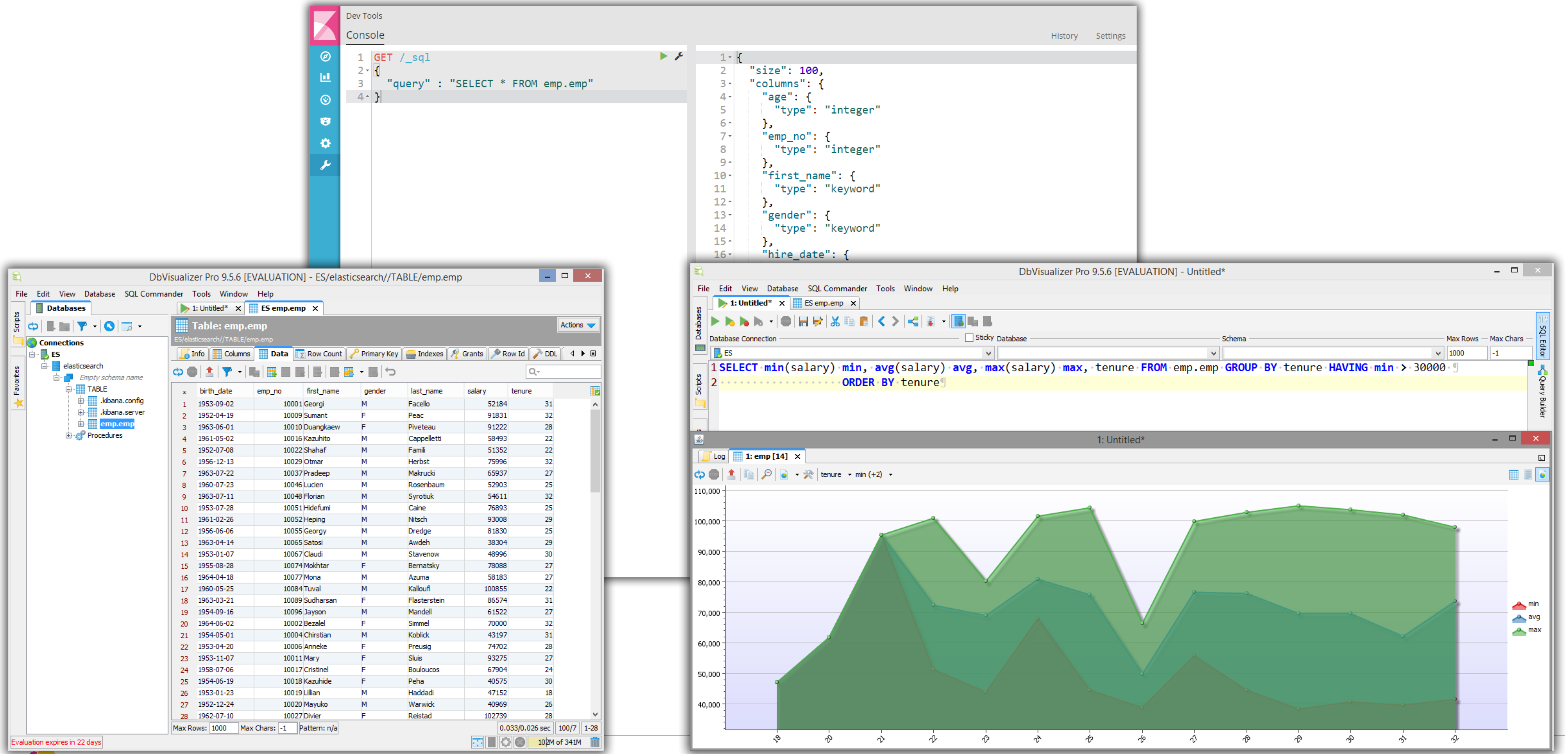
CLI

- OS independent
- Quick diagnostics and sanity checks
- Admin focused
- Optimized for efficiency

JDBC

- Dedicated client (driver) and server component
- JDBC 4.2/Java 8 (downgrade possible)
- Supports `java.sql` and `javax.sql` APIs
- Pays attention to details
    - Timeouts (connect vs read vs network)
    - Logging
- Light, without dependencies

SQL



```
GET /_sql
{
  "query" : "SELECT * FROM emp.emp"
}
```

```
{
  "size": 100,
  "columns": {
    "age": {
      "type": "integer"
    },
    "emp_no": {
      "type": "integer"
    },
    "first_name": {
      "type": "keyword"
    },
    "gender": {
      "type": "keyword"
    },
    "hire_date": {
```

	birth_date	emp_no	first_name	gender	last_name	salary	tenure
1	1953-09-02	10001	Georgi	M	Facello	52184	31
2	1952-04-19	10009	Sumant	F	Peac	91831	32
3	1963-06-01	10010	Duangkaew	F	Piveteau	91222	28
4	1961-05-02	10016	Kazuhito	M	Cappelletti	58493	22
5	1952-07-08	10022	Shahaf	M	Famili	51352	22
6	1956-12-13	10029	Otmar	M	Herbst	75996	32
7	1963-07-22	10037	Pradeep	M	Makrucki	65937	27
8	1960-07-23	10046	Lucien	M	Rosenbaum	52903	25
9	1963-07-11	10048	Florian	M	Syrotiuk	54611	32
10	1953-07-28	10051	Hidefumi	M	Caine	76893	25
11	1961-02-26	10052	Heping	M	Nitsch	93008	29
12	1956-06-06	10055	Georgy	M	Dredge	81830	25
13	1963-04-14	10065	Satosi	M	Awdeh	38304	29
14	1953-01-07	10067	Claudi	M	Stavenow	48996	30
15	1955-08-28	10074	Mokhtar	F	Bernatsky	78088	27
16	1964-04-18	10077	Mona	M	Azuma	58183	27
17	1960-05-25	10084	Tuval	M	Kalloufi	100855	22
18	1963-03-21	10089	Sudharsan	F	Flasterstein	86574	31
19	1954-09-16	10096	Jayson	M	Mandell	61522	27
20	1964-06-02	10002	Bezalel	F	Simmel	70000	32
21	1954-05-01	10004	Chirstian	M	Koblick	43197	31
22	1953-04-20	10006	Anneke	F	Preusig	74702	28
23	1953-11-07	10011	Mary	F	Sluis	93275	27
24	1958-07-06	10017	Cristinel	F	Bouloucos	67904	24
25	1954-06-19	10018	Kazuhide	F	Peha	40575	30
26	1953-01-23	10019	Lillian	M	Haddadi	47152	18
27	1952-12-24	10020	Mayuko	M	Warwick	40969	26
28	1962-07-10	10027	Divier	F	Reistad	102739	28

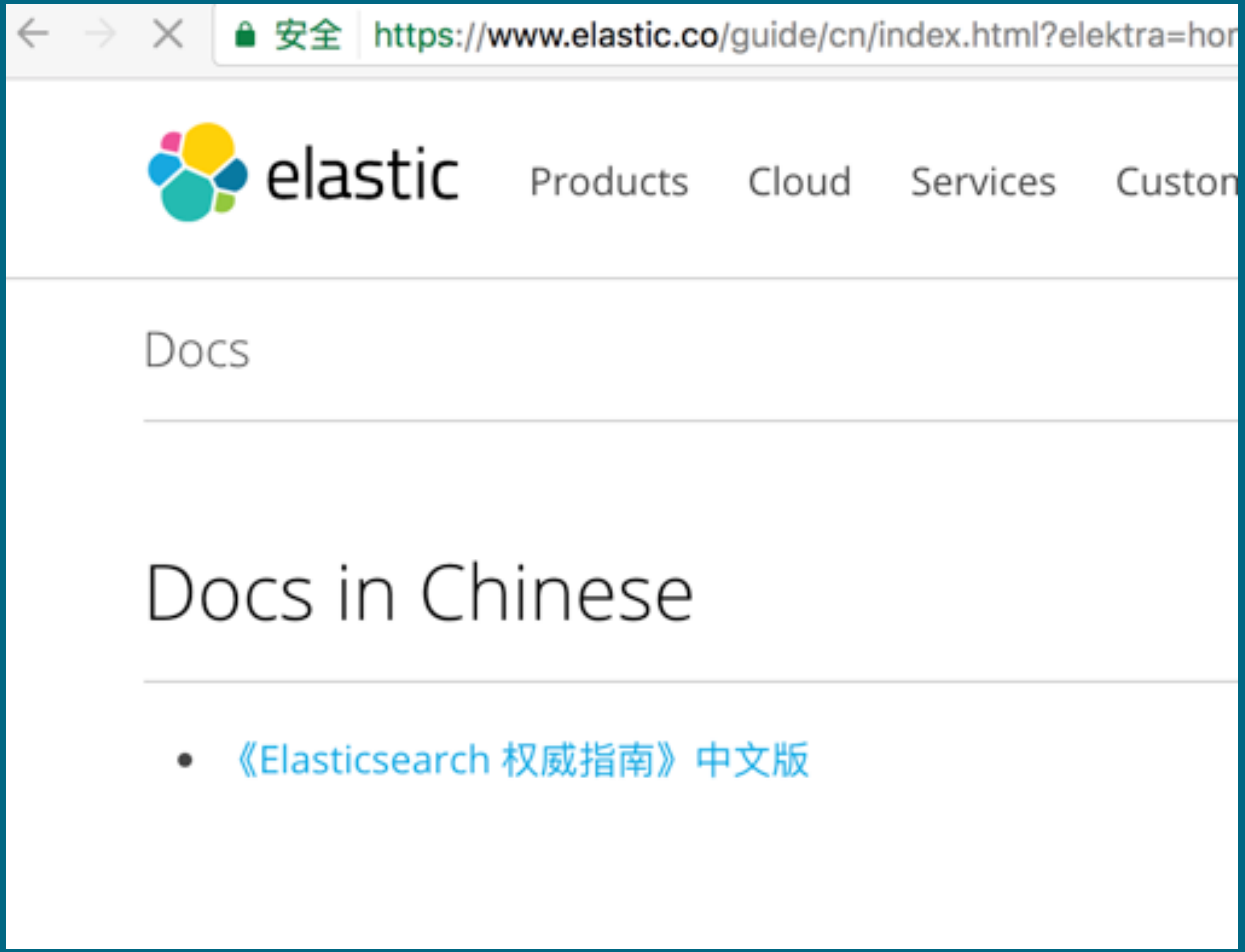
Max Rows: 1000 Max Chars: -1 Pattern: n/a 0.033/0.026 sec 100/7 1-28

102M of 341M

```
SELECT min(salary) min, avg(salary) avg, max(salary) max, tenure FROM emp.emp GROUP BY tenure HAVING min > 30000
                  ORDER BY tenure
```

Elastic & Community

- 上海 Meetup
  - https://elasticsearch.cn/article/163
- 中文权威指南已上线！



Docs

Docs in Chinese

- 《Elasticsearch 权威指南》中文版

elastic

elastic

谢谢！