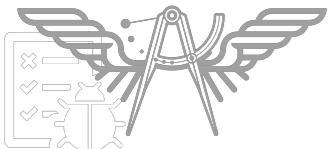


ELK Stack





Somkiat Puisungnoen

Somkiat Puisungnoen

Update Info 1 View Activity Log 10+ ...

Timeline About Friends 3,138 Photos More

When did you work at Opendream? X

... 22 Pending Items

Intro

Software Craftsmanship

Software Practitioner at สยามชัมนาณกิจ พ.ศ. 2556

Agile Practitioner and Technical at SPRINT3r

Post Photo/Video Live Video Life Event

What's on your mind?

Public Post

Somkiat Puisungnoen 15 mins · Bangkok · ⚙️

Java and Bigdata



Page

Messages

Notifications 3

Insights

Publishing Tools

Settings

Help ▾



somkiat.cc

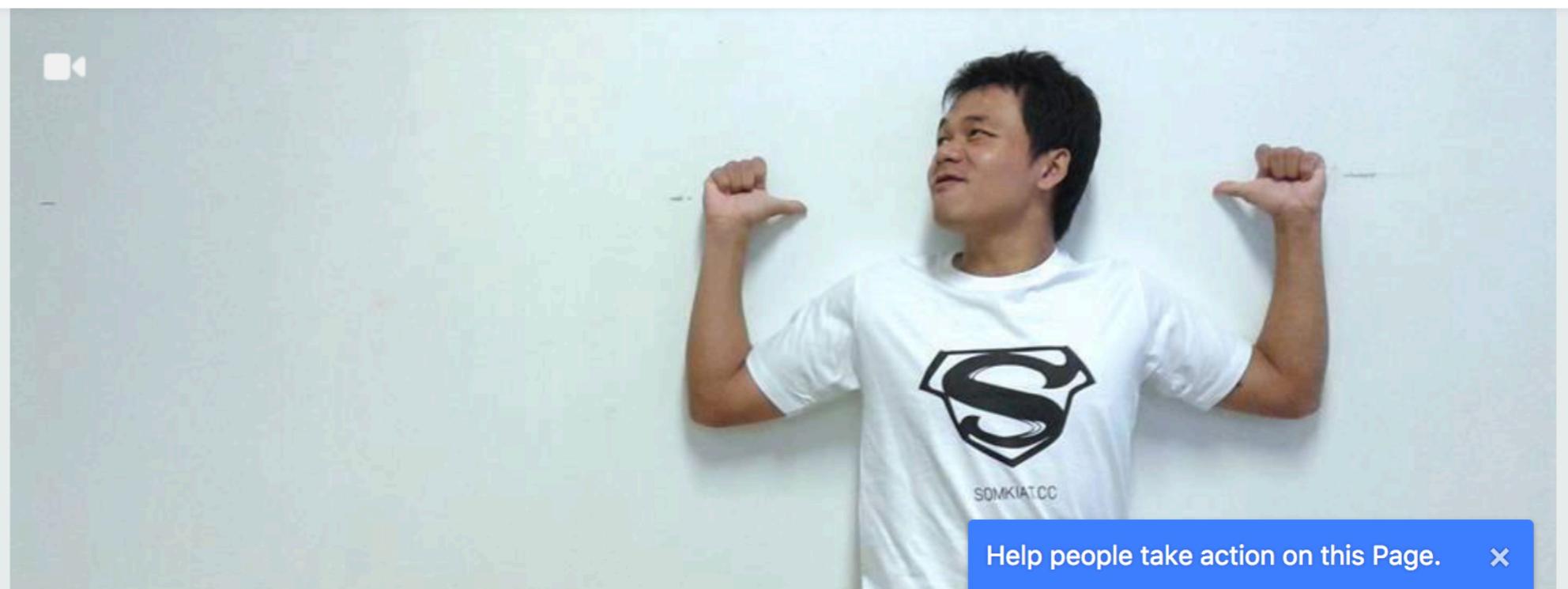
@somkiat.cc

Home

Posts

Videos

Photos



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Agenda

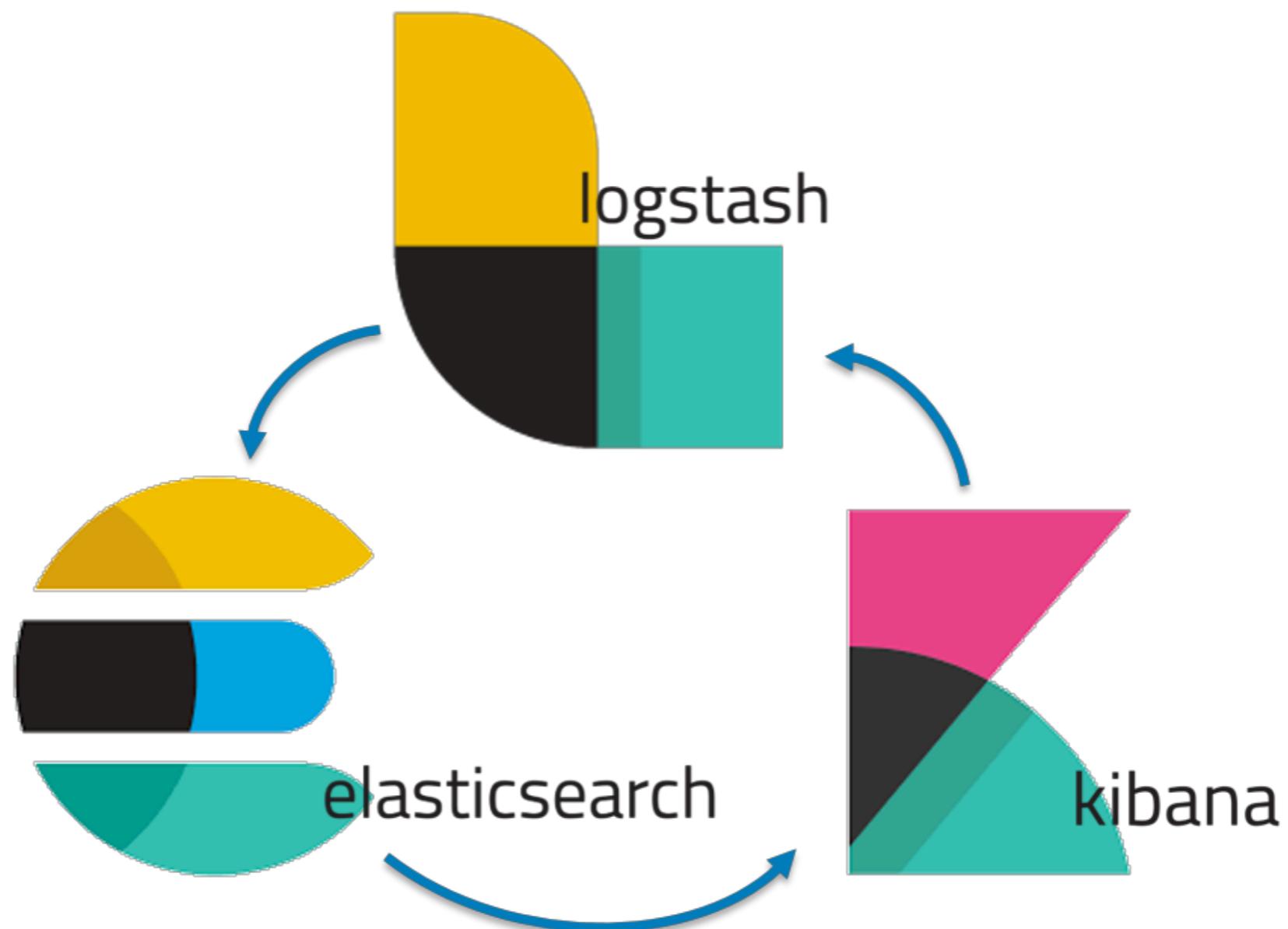
- ELK stack
- Introduction to Elasticsearch
- CRUD (Create, Read, Update, Delete)
- Search DSL (Domain Specific Language)
- Analyzer
- Mapping
- Aggregation



Agenda

- Working with Kibana
- Useful features
 - Auto-suggestion
 - ngram algorithm
- Clustering management
- Design for scaling
- Working with Logstash





Elasticsearch ?



Elasticsearch

Search
Analytic
Real-time
Distributed



Distributed Search Engine

Open Source
Document-based
Based on Apache Lucene
JSON over HTTP



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Document based

JSON (JavaScript Object Notation)

Dynamic Schema

Some relationship (nested, parent/child)



StackOverflow Question

```
{  
  "items": [  
    {  
      "owner": {  
        "reputation": 13,  
        "user_id": 9796344,  
        "user_type": "registered",  
        "profile_image": "",  
        "display_name": "Cherry",  
        "link": "https://stackoverflow.com/users/9796344/cherry"  
      },  
      "score": 0,  
      "last_activity_date": 1528986761,  
      "creation_date": 1528986761,  
      "post_type": "question",  
      "post_id": 50859951,  
      "link": "https://stackoverflow.com/q/50859951"  
    }  
  ],  
  "has_more": false,  
  "quota_max": 10000,  
  "quota_remaining": 9986  
}
```

<https://api.stackexchange.com/docs/posts-by-ids>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Ranking from DB Engine (2018)

343 systems in ranking, June 2018

Rank			DBMS	Database Model	Score		
Jun 2018	May 2018	Jun 2017			Jun 2018	May 2018	Jun 2017
1.	1.	1.	Oracle	Relational DBMS	1311.25	+20.84	-40.51
2.	2.	2.	MySQL	Relational DBMS	1233.69	+10.35	-111.62
3.	3.	3.	Microsoft SQL Server	Relational DBMS	1087.73	+1.89	-111.23
4.	4.	4.	PostgreSQL	Relational DBMS	410.67	+9.77	+42.13
5.	5.	5.	MongoDB	Document store	343.79	+1.67	+8.79
6.	6.	6.	DB2	Relational DBMS	185.64	+0.03	-1.86
7.	7.	↑ 9.	Redis	Key-value store	136.30	+0.95	+17.42
8.	↑ 9.	↑ 11.	Elasticsearch	Search engine	131.04	+0.60	+19.48
9.	↓ 8.	↓ 7.	Microsoft Access	Relational DBMS	130.99	-2.12	+4.44
10.	10.	↓ 8.	Cassandra	Wide column store	119.21	+1.38	-4.91
11.	11.	↓ 10.	SQLite	Relational DBMS	114.26	-1.19	-2.44

<https://db-engines.com/en/ranking>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Let's start



Installation

Elasticsearch
Kibana



Elasticsearch

Required Java 8

JDK 1.8.0_131+

Need \$JAVA_HOME



Start Elasticsearch

./bin/elasticsearch

```
[0g8-71W] loaded module [reindex]
[0g8-71W] loaded module [repository-url]
[0g8-71W] loaded module [transport-netty4]
[0g8-71W] loaded module [tribe]
[0g8-71W] no plugins loaded
[0g8-71W] using discovery type [zen]
initialized
[0g8-71W] starting ...
[0g8-71W] publish_address {127.0.0.1:9300},
[0g8-71W] recovered [0] indices into cluster_state
transport] [0g8-71W] publish_address {127.0.0.1:9200},
```



Default of Memory

1 GB !!! (Java need more memory)

```
 ] [DW5j42N] JVM arguments [-Xms1g, -Xmx1g, -  
ction=75, -XX:+UseCMSInitiatingOccupancyOnly, -XX:  
Dfile.encoding=UTF-8, -Djna.nosys=true, -XX:-Omit  
.netty.noKeySetOptimization=true, -Dio.netty.recy  
led=false, -Dlog4j2.disable.jmx=true, -Djava.io.t  
T/elasticsearch.G4kbTLZn, -XX:+HeapDumpOnOutOfMem  
s_err_pid%p.log, -Xlog:gc*,gc+age=trace,safepoint  
ize=64m, -Djava.locale.providers=COMPAT, -XX:UseA
```



Config of JVM

`$ES_HOME/config/jvm.options`

```
# Xms represents the initial size of total heap space  
# Xmx represents the maximum size of total heap space  
  
-Xms1g  
-Xmx1g
```



Default plugins

```
[o.e.p.PluginsService      ] [DW5j42N] loaded module [aggs-matrix-stats]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [analysis-common]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [ingest-common]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [lang-expression]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [lang-mustache]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [lang-painless]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [mapper-extras]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [parent-join]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [percolator]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [rank-eval]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [reindex]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [repository-url]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [transport-netty4]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [tribe]
```



Install X-Pack by default

```
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-core]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-deprecation]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-graph]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-logstash]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-ml]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-monitoring]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-rollup]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-security]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-sql]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-upgrade]
[o.e.p.PluginsService      ] [DW5j42N] loaded module [x-pack-watcher]
[o.e.p.PluginsService      ] [DW5j42N] no plugins loaded
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/installing-xpack-es.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

X-Pack ?

Elastic Stack Extension
Security
Monitoring
Alerting
Reporting
Machine Learning



Hello Elasticsearch

<http://localhost:9200/>

```
{  
  "name": "DW5j42N",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "boIZeF6MSHyxZ2owIG66rg",  
  "version": {  
    "number": "6.4.2",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "04711c2",  
    "build_date": "2018-09-26T13:34:09.098244Z",  
    "build_snapshot": false,  
    "lucene_version": "7.4.0",  
    "minimum_wire_compatibility_version": "5.6.0",  
    "minimum_index_compatibility_version": "5.0.0"  
  },  
  "tagline": "You Know, for Search"  
}
```



```
{  
  "name": "DW5j42N",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "boIZeF6MSHyxZ2owIG66rg",  
  "version": {  
    "number": "6.4.2",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "04711c2",  
    "build_date": "2018-09-26T13:34:09.098244Z",  
    "build_snapshot": false,  
    "lucene_version": "7.4.0",  
    "minimum_wire_compatibility_version": "5.6.0",  
    "minimum_index_compatibility_version": "5.0.0"  
  },  
  "tagline": "You Know, for Search"  
}
```



```
{  
  "name": "DW5j42N",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "boIZeF  
  "version": {  
    "number": "6.4.2",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "04711c2",  
    "build_date": "2018-09-26T13:34:09.098244Z",  
    "build_snapshot": false,  
    "lucene_version": "7.4.0",  
    "minimum_wire_compatibility_version": "5.6.0",  
    "minimum_index_compatibility_version": "5.0.0"  
  },  
  "tagline": "You Know, for Search"  
}
```

Name of node and cluster



Name of node and cluster

\$ES_HOME/config/elasticsearch.yml

```
# ----- Cluster -----  
#  
# Use a descriptive name for your cluster:  
#  
cluster.name: my-application  
#  
# ----- Node -----  
#  
# Use a descriptive name for the node:  
#  
node.name: node-1  
#  
# Add custom attributes to the node:  
#  
#node.attr.rack: r1  
#
```



Change in Elasticsearch 7.x

Default name = Hostname

<https://www.elastic.co/guide/en/elasticsearch/reference/master/breaking-changes-7.0.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Try to change and restart !!!



```
{  
  "name": "DW5j42N",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "boIZeF6MSHyxZ2owIG66rg",  
  "version": {  
    "number": "6.4.2",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "04711c2",  
    "build_date": "2018-09-26T13:34:09.098244Z",  
    "build_snapshot": false,  
    "lucene_version": "7.4.0",  
    "minimum_wire_compatibility_version": "5.6.0",  
    "minimum_index_compatibility_version": "5.0.0"  
  },  
  "tagline": "You Know, for Search"  
}
```



Apache Lucene

The screenshot shows the official Apache Lucene website. At the top, there's a navigation bar with a search bar containing "Search with Apache So" and a dropdown menu "select provider". Below the search bar are three tabs: "CORE (JAVA)", "SOLR", and "PYLUCENE". The main content area features the Apache Lucene logo (a green feather) and the Solr logo (a red sunburst icon). A banner below the logos reads "Ultra-fast Search Library and Server". A sub-banner below the main banner says "Apache Lucene and Solr set the standard for search and indexing performance". On the left side of the main content area, there's a "Welcome to Apache Lucene" section with a list of what the project develops. On the right side, there are two "DOWNLOAD" buttons: one for "Apache Lucene 7.5.0" (green button) and one for "Apache Solr 7.5.0" (orange button). Below these buttons is a "Projects" link.

Ultra-fast Search Library and Server

Apache Lucene and Solr set the standard for search and indexing performance

Welcome to Apache Lucene

The Apache Lucene™ project develops open-source search software, including:

- [Lucene Core](#), our flagship sub-project, provides Java-based indexing and search technology, as well as spellchecking, hit highlighting and advanced analysis/tokenization capabilities.
- [Solr™](#) is a high performance search server built using Lucene Core, with XML/HTTP and JSON/Python/Ruby APIs, hit highlighting, faceted search, caching, replication, and a web admin interface.
- [PyLucene](#) is a Python port of the Core project.

[DOWNLOAD](#)

Apache Lucene 7.5.0

[DOWNLOAD](#)

Apache Solr 7.5.0

[Projects](#)

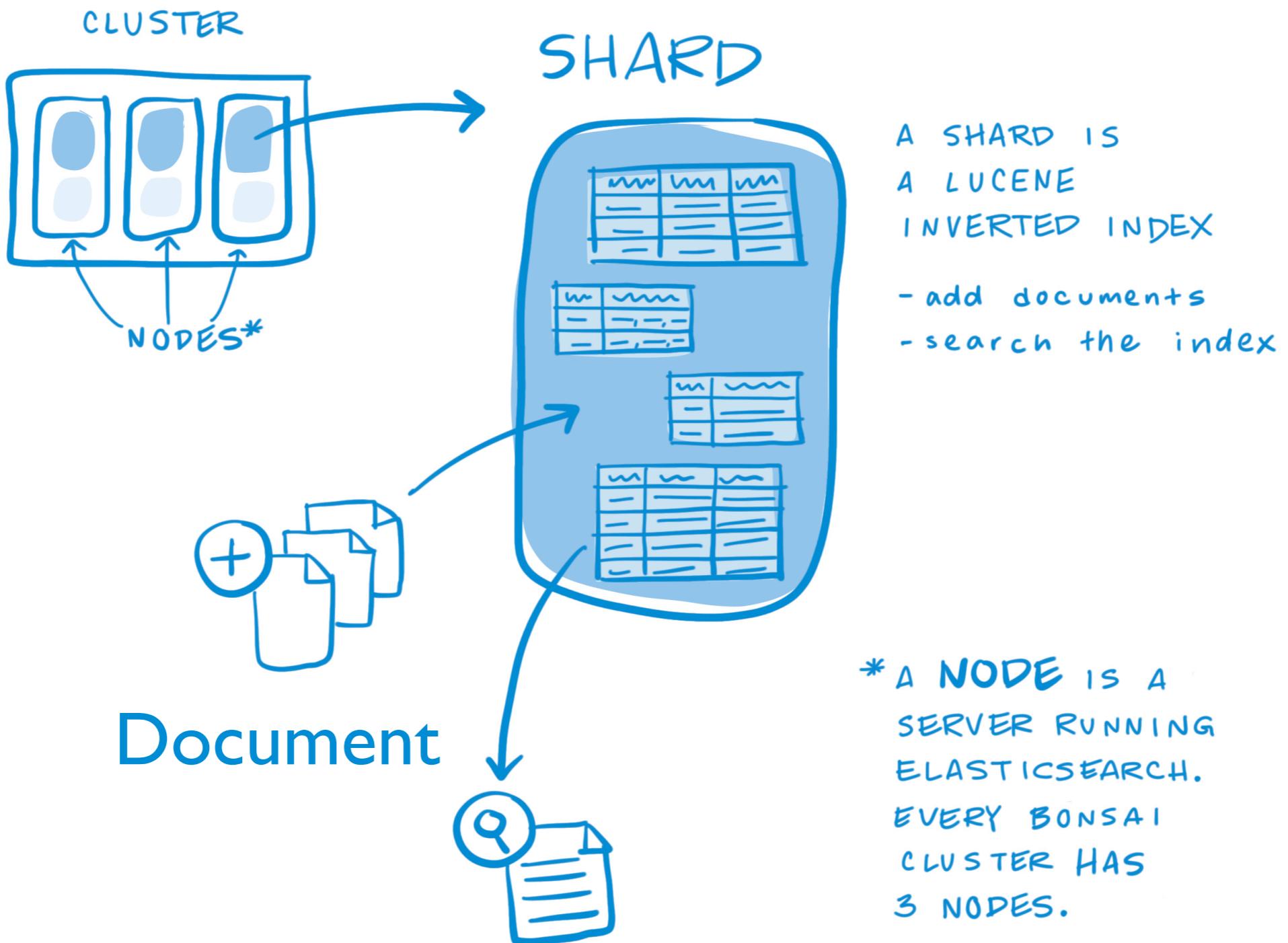
<http://lucene.apache.org/>



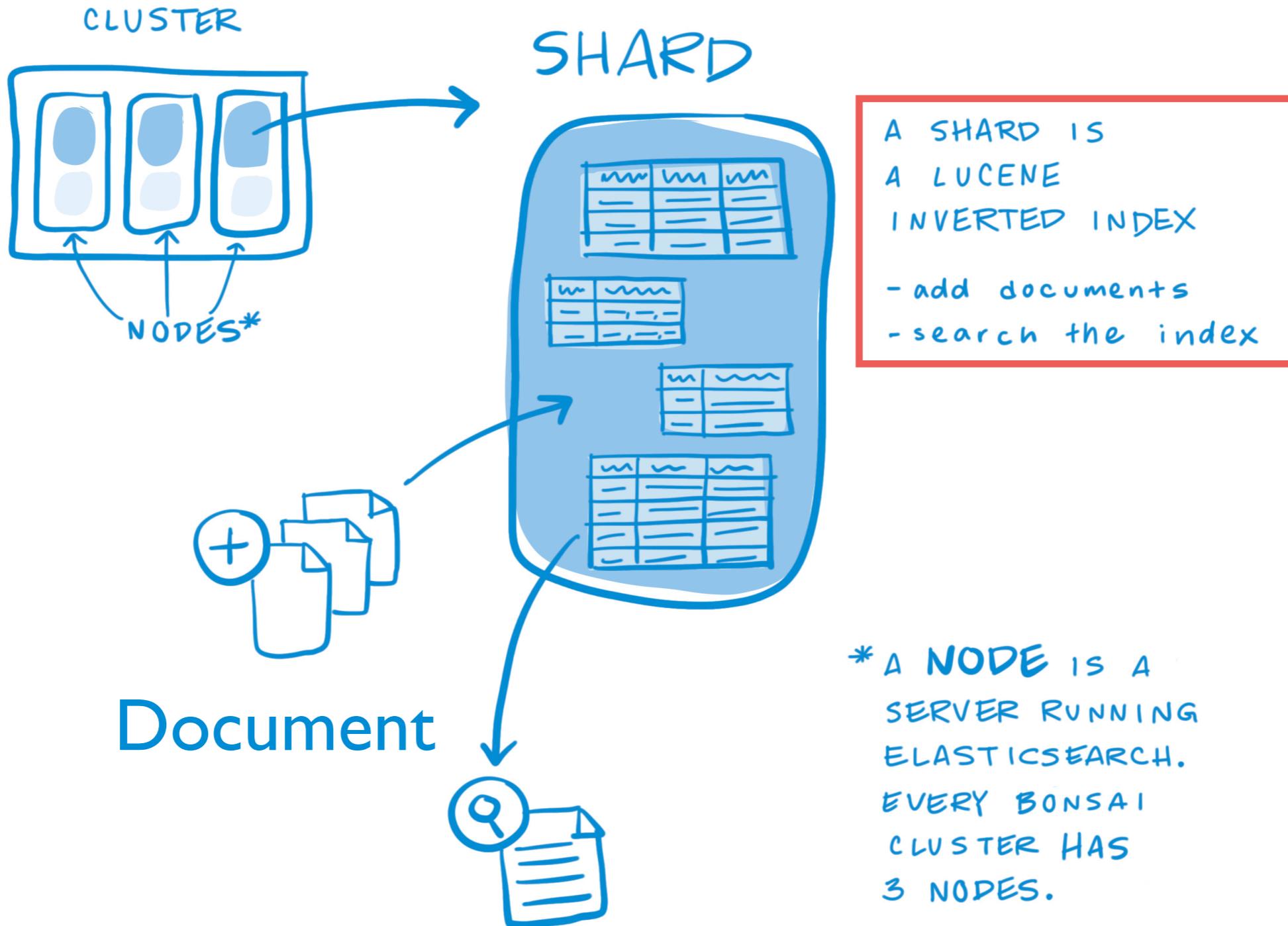
ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Basic concepts

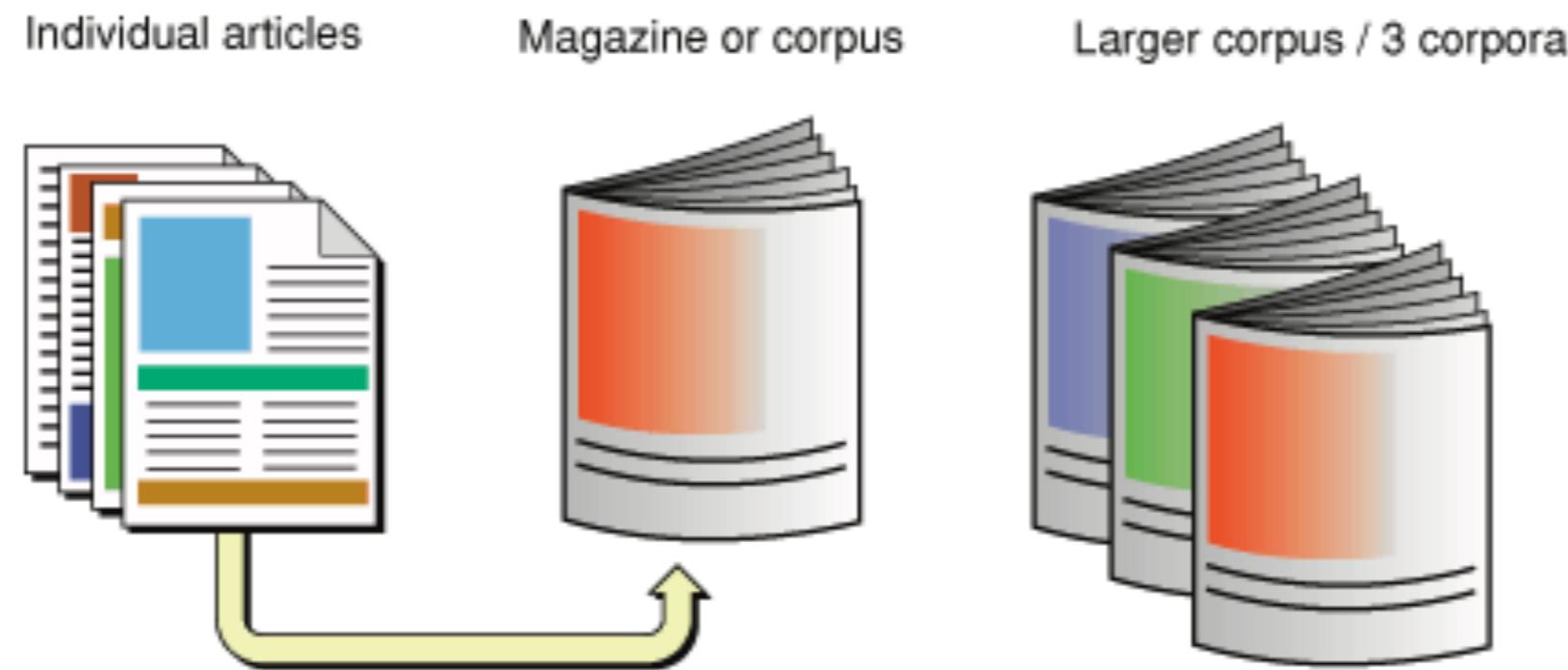


Basic concepts



Inverted Index

Corpus is a collection of documents



https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html#/apple_ref/doc/uid/TP40002843-TPXREF101

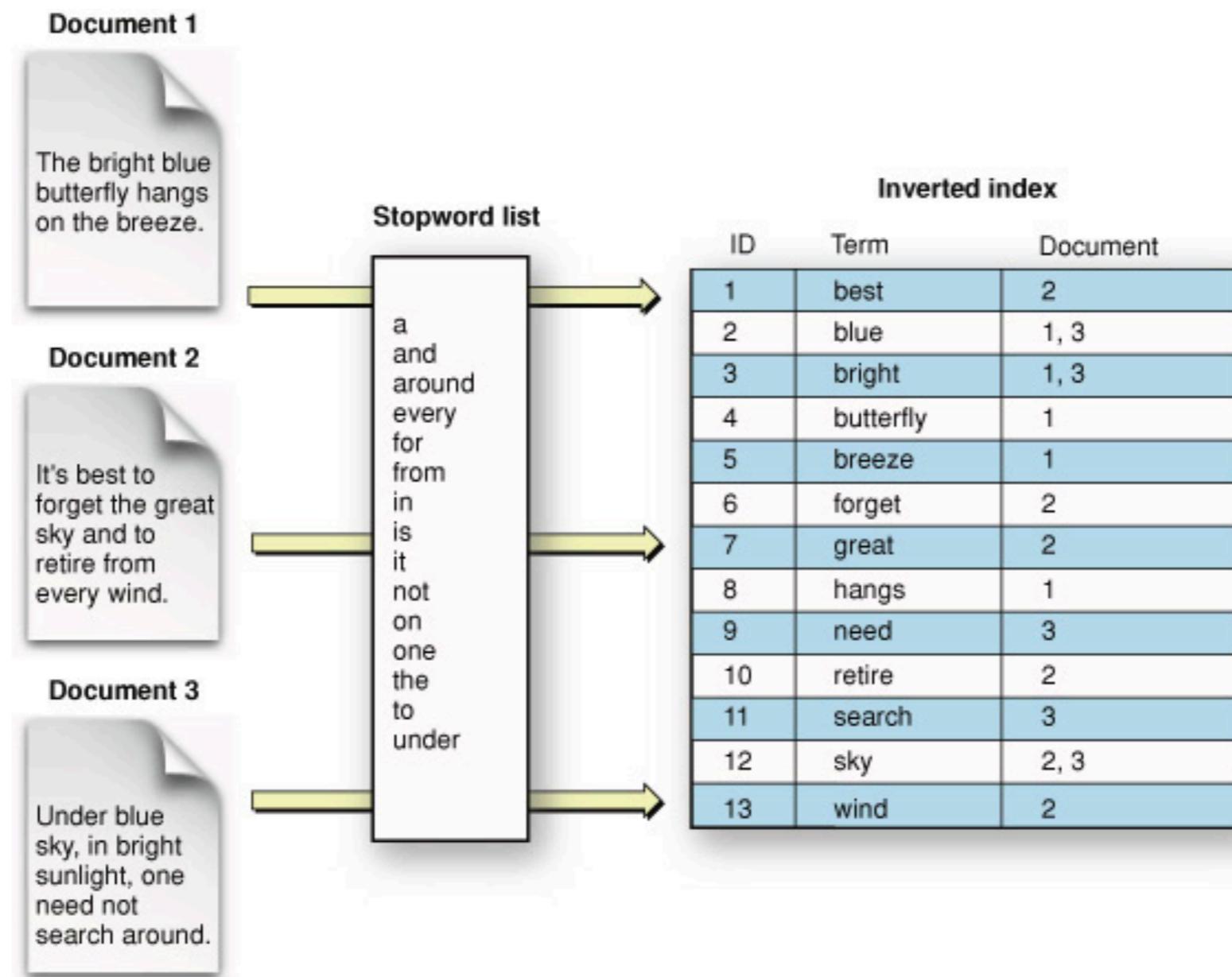


ELK Stack

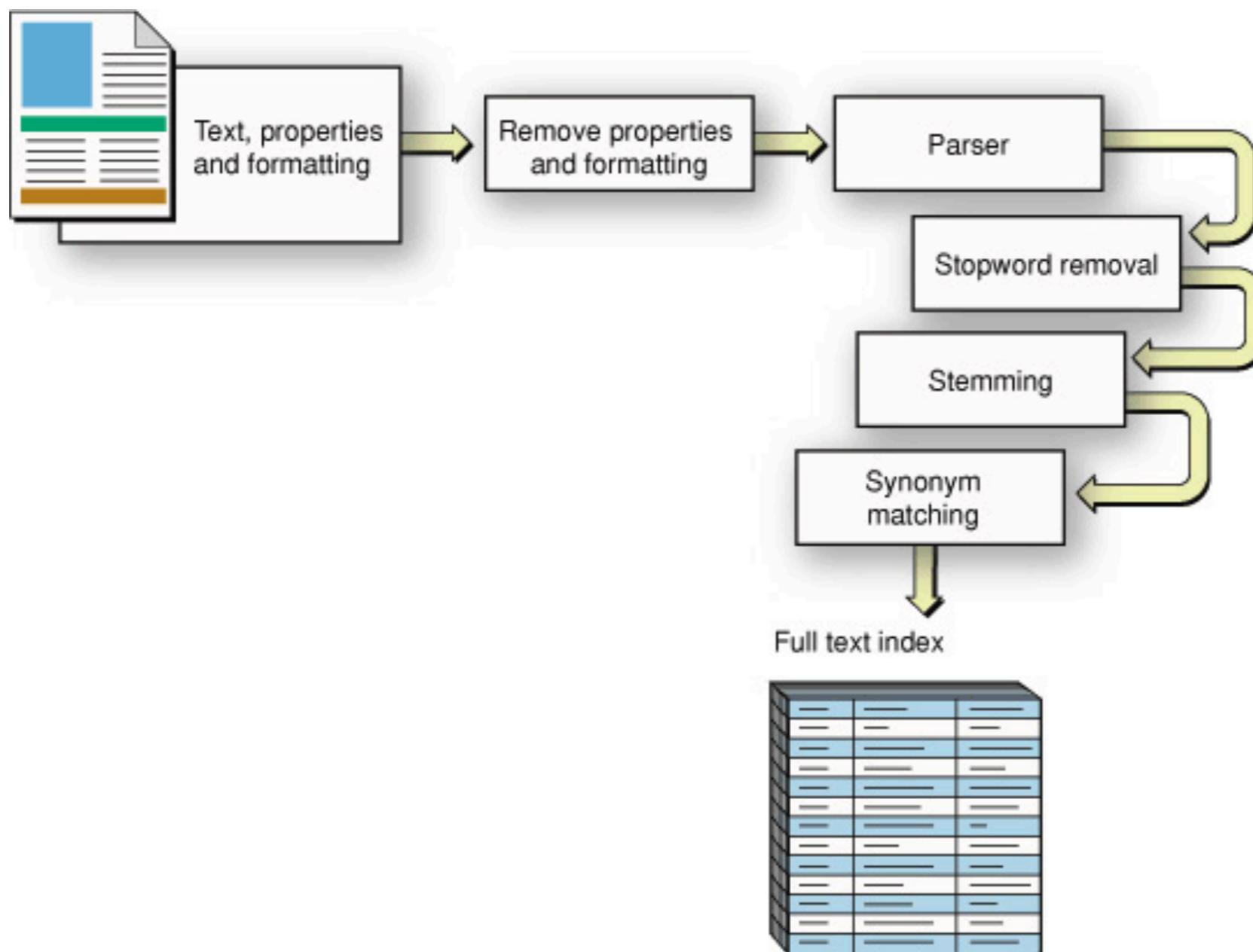
© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

Inverted Index

Try to construct index



Text extraction



```
{  
  "name": "DW5j42N",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "boIZeF6MSHyxZ2owIG66rg",  
  "version": {  
    "number": "6.4.2",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "04711c2",  
    "build_date": "2018-09-26T13:34:09.098244Z",  
    "build_snapshot": false,  
    "lucene_version": "7.4.0",  
    "minimum_wire_compatibility_version": "5.6.0",  
    "minimum_index_compatibility_version": "5.0.0"  
  },  
  "tagline": "You Know, for Search"  
}
```

DSL version

Index version



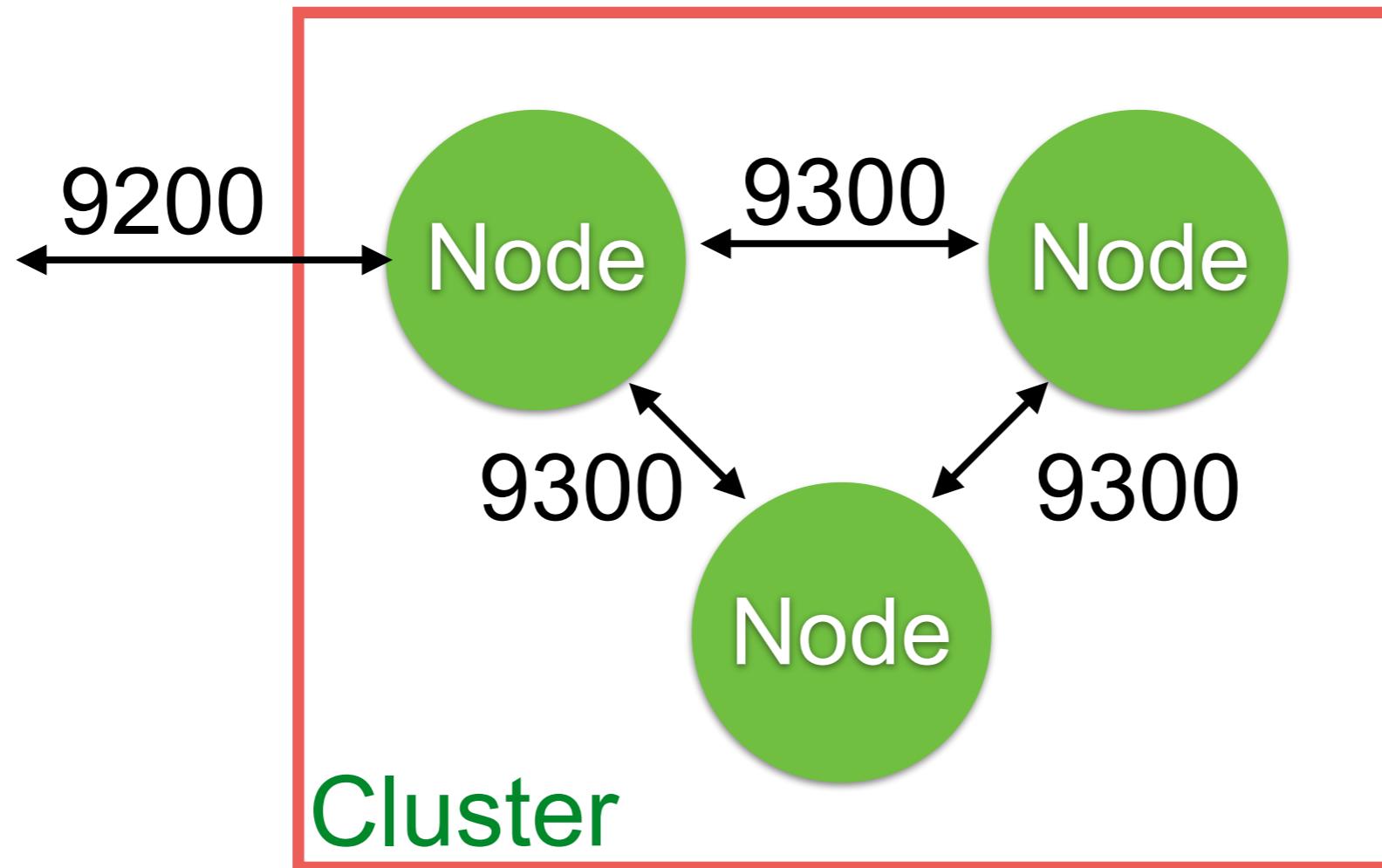
```
{  
  "name": "DW5j42N",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "boIZeF6MSHyxZ2owIG66rg",  
  "version": {  
    "number": "6.4.2",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "04711c2",  
    "build_date": "2018-09-26T13:34:09.098244Z",  
    "build_snapshot": false,  
    "lucene_version": "7.4.0",  
    "minimum_wire_compatibility_version": "5.6.0",  
    "minimum_index_compatibility_version": "5.0.0"  
  },  
  "tagline": "You Know, for Search"  
}
```

Index version



Ports of Elasticsearch

RESTful API with JSON Over HTTP (9200)
Java API (9300)



Health of cluster

http://localhost:9200/_cluster/health

```
{  
  "cluster_name": "elasticsearch",  
  "status": "green",  
  "timed_out": false,  
  "number_of_nodes": 1,  
  "number_of_data_nodes": 1,  
  "active_primary_shards": 0,  
  "active_shards": 0,  
  "relocating_shards": 0,  
  "initializing_shards": 0,  
  "unassigned_shards": 0,  
  "delayed_unassigned_shards": 0,  
  "number_of_pending_tasks": 0,  
  "number_of_in_flight_fetch": 0,  
  "task_max_waiting_in_queue_millis": 0,  
  "active_shards_percent_as_number": 100.0  
}
```



Health of cluster

Status	Meaning
Green	All shards are allocated
Yellow	Primary shard is allocated, but replicas are not
Red	Shard not allocated in the cluster



cat APIs

`http://localhost:9200/_cat`

```
=^.^=
/_cat/allocation
/_cat/shards
/_cat/shards/{index}
/_cat/master
/_cat/nodes
/_cat/tasks
/_cat/indices
/_cat/indices/{index}
/_cat/segments
/_cat/segments/{index}
/_cat/count
/_cat/count/{index}
/_cat/recovery
/_cat/recovery/{index}
/_cat/health
/_cat/pending_tasks
/_cat/aliases
/_cat/aliases/{alias}
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/cat.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

cat APIs

`http://localhost:9200/_cat/nodes?v`

ip	heap.percent	ram.percent	cpu	load_1m	load_5m	load_15m	node.role	master	name
127.0.0.1	20	100	7	1.98			mdi	*	DW5j42N



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Start Kibana

```
[status][plugin:xpack_main@6.4.2] Status changed from yellow to green - Ready
[status][plugin:searchprofiler@6.4.2] Status changed from yellow to green - Ready
[status][plugin:ml@6.4.2] Status changed from yellow to green - Ready
[status][plugin:tilemap@6.4.2] Status changed from yellow to green - Ready
[status][plugin:watcher@6.4.2] Status changed from yellow to green - Ready
[status][plugin:index_management@6.4.2] Status changed from yellow to green - Ready

[status][plugin:graph@6.4.2] Status changed from yellow to green - Ready
[status][plugin:grokdebugger@6.4.2] Status changed from yellow to green - Ready
[status][plugin:logstash@6.4.2] Status changed from yellow to green - Ready
[status][plugin:reporting@6.4.2] Status changed from yellow to green - Ready
[kibana-monitoring][monitoring-ui] Starting monitoring stats collection
[status][plugin:security@6.4.2] Status changed from yellow to green - Ready
[license][xpack] Imported license information from Elasticsearch for the [monitor]
tatus: active
[listening][server][http] Server running at http://localhost:5601
```



Hello Kibana

<http://localhost:5601/>

The screenshot shows the Kibana home page with a sidebar on the left containing icons for APM, Logging, Metrics, Security analytics, Dashboard, Discover, Timelion, Visualize, Console, Index Patterns, and Saved Objects.

Add Data to Kibana

Use these solutions to quickly turn your data into pre-built dashboards and monitoring systems.

APM
APM automatically collects in-depth performance metrics and errors from inside your applications.
[Add APM](#)

Logging
Ingest logs from popular data sources and easily visualize in preconfigured dashboards.
[Add log data](#)

Metrics
Collect metrics from the operating system and services running on your servers.
[Add metric data](#)

Security analytics
Centralize security events for interactive investigation in ready-to-go visualizations.
[Add security events](#)

Visualize and Explore Data

Dashboard
Display and share a collection of visualizations and saved searches.

Discover
Interactively explore your data by querying and filtering raw documents.

Timelion
Use an expression language to analyze time series data

Visualize
Create visualizations and aggregate data stores in your

Manage and Administer the Elastic Stack

Console
Skip cURL and use this JSON interface to work with your data directly.

Index Patterns
Manage the index patterns that help retrieve your data from Elasticsearch.

Saved Objects
Import, export, and manage your saved searches,

Data already in Elasticsearch?
[Set up index patterns](#)



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Using Dev Tools

The screenshot shows the Kibana interface with a sidebar on the left and a main content area on the right.

Left Sidebar:

- Kibana logo
- Discover
- Visualize
- Dashboard
- Timelion
- APM
- Dev Tools** (highlighted with a red box)
- Monitoring
- Management

Main Content Area:

Dev Tools

Welcome to Console

Quick intro to the UI

The Console UI is split into two panes: an editor pane (left) and a response pane (right). Use the editor to type requests and submit them in the response pane on the right side.

Console understands requests in a compact format, similar to cURL:

```
1 # index a doc
2 PUT index/type/1
3 {
4   "body": "here"
5 }
6
7 # and get it ...
8 GET index/type/1
```

While typing a request, Console will make suggestions which you can then accept by hitting Enter/Tab. These suggestions are made based on types.

A few quick tips, while I have your attention

- Submit requests to ES using the green triangle button.
- Use the wrench menu for other useful things.
- You can paste requests in cURL format and they will be translated to the Console syntax.
- You can resize the editor and output panes by dragging the separator between them.
- Study the keyboard shortcuts under the Help button. Good stuff in there!

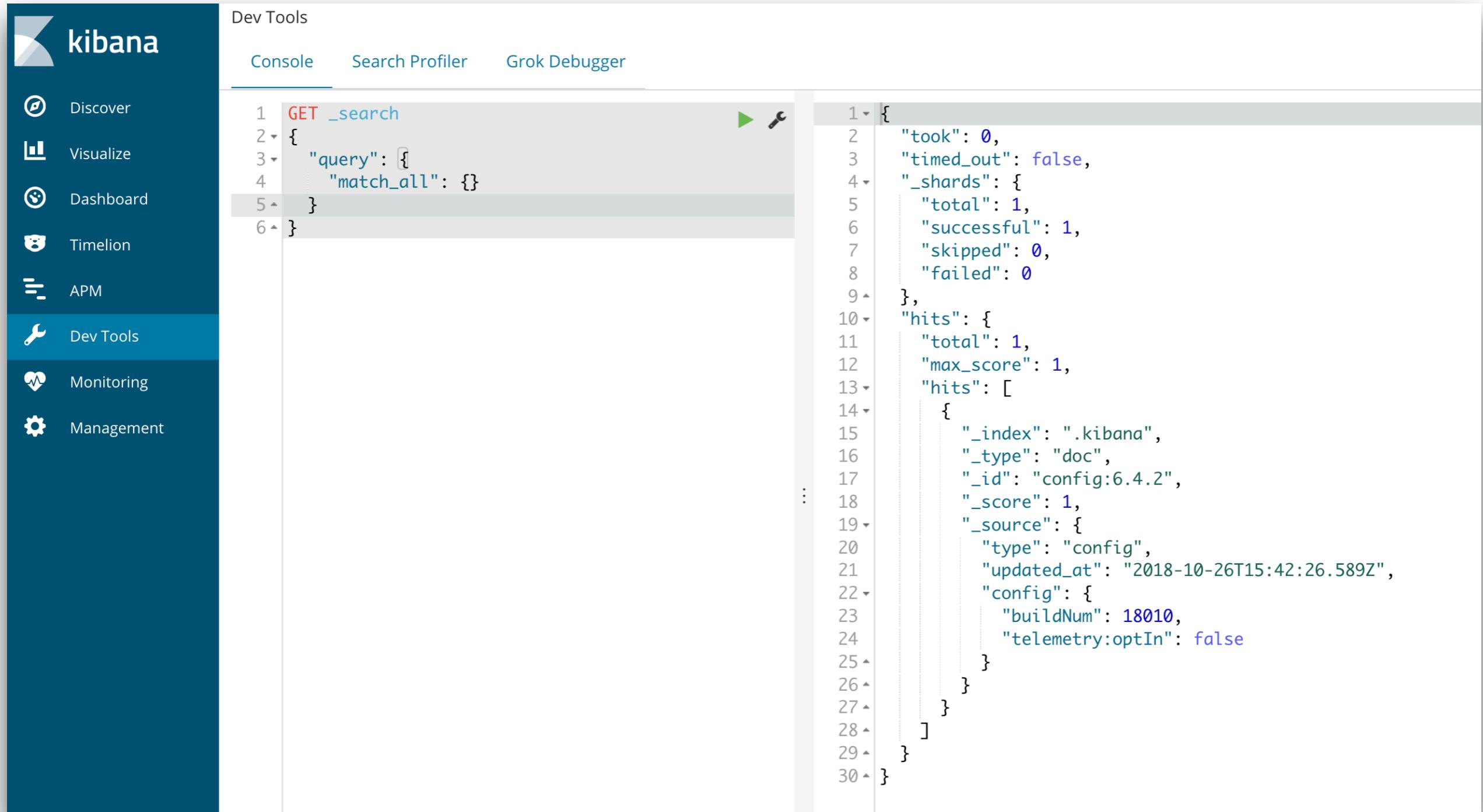
[Get to work](#)



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Ready to start



The screenshot shows the Kibana Dev Tools interface. On the left is a sidebar with icons for Discover, Visualize, Dashboard, Timelion, APM, Dev Tools (which is selected), Monitoring, and Management. The main area is titled "Dev Tools" and contains three tabs: Console, Search Profiler, and Grok Debugger. The "Console" tab is active, displaying a code editor with a GET _search request and its JSON response. The request is:

```
1 GET _search
2 {
3   "query": {
4     "match_all": {}
5   }
6 }
```

The response is:

```
1 {
2   "took": 0,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": 1,
12    "max_score": 1,
13    "hits": [
14      {
15        "_index": ".kibana",
16        "_type": "doc",
17        "_id": "config:6.4.2",
18        "_score": 1,
19        "_source": {
20          "type": "config",
21          "updated_at": "2018-10-26T15:42:26.589Z",
22          "config": {
23            "buildNum": 18010,
24            "telemetry:optIn": false
25          }
26        }
27      }
28    ]
29  }
30 }
```



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

CRUD with Elasticsearch

02-crud/book_document.json



CRUD with Elasticsearch

Create document

Read document

Update document

Delete document



Create a document

PUT /store/book/1

```
{  
  "title": "Elasticsearch: The Definitive Guide",  
  "author_name": [  
    "Clinton Gormley",  
    "Zachary Tong"  
,  
  "tag": [  
    "search",  
    "computer"  
,  
  "isbn-13": "978-1449358549",  
  "isbn-10": "1449358543",  
  "price": 44.3,  
  "page": 724,  
}
```



Create document

PUT **/store/book/1**

Index name

Type name

Document ID



Compare with RDBMS

Database

Table

Row

Column

Index

Type*

Document

Field

* Only 1 type per index



Change in Elasticsearch 7.x

of shard of index change from 5 to 1

#! Deprecation: the default number of shards will change from [5] to [1] in 7.0.0; if you wish to continue using the default of [5] shards, you must manage this on the create index request or with an index template

```
{  
  "_index": "store1",  
  "_type": "book",  
  "_id": "2",  
  "_version": 1,  
  "result": "created",  
  "_shards": {  
    "total": 5,  
    "successful": 5,  
    "failed": 0  
  }  
}
```



Read document

GET /store/book/1

```
{  
  "_index": "store",  
  "_type": "book",  
  "_id": "1",  
  "_version": 1,  
  "found": true,  
  "_source": {  
    "title": "Elasticsearch: The Definitive Guide",  
    "author_name": [  
      "Clinton Gormley",  
      "Zachary Tong"  
    ],  
    "tag": [  
      "search",  
      "computer"  
    ]  
  }  
}
```

Information of document



Update document

Whole document
Partial document



Update whole document

PUT /store/book/123

```
{  
  "title": "Update",  
  "author_name": [  
    "user1",  
    "user2"  
  ],  
  "tag": [  
    "update",  
    "book"  
  ]  
}
```



Update partial document

POST /store/book/123/_update

```
{  
  "doc": {  
    "title": "partial update",  
    "tag": [  
      "test",  
      "computer"  
    ],  
    "views": 0  
  }  
}
```



Delete document

DELETE /store/book/1

```
{  
  "_index": "store",  
  "_type": "book",  
  "_id": "1",  
  "_version": 2,  
  "result": "deleted",  
  "_shards": {  
    "total": 2,  
    "successful": 1,  
    "failed": 0  
  },  
  "_seq_no": 1,  
  "_primary_term": 1  
}
```



More features

Update by query

Delete by query

Partial update document



Workshop

02-crud/book_document.json



Bulk API

03-bulk/book_bulk.json

<https://www.elastic.co/guide/en/elasticsearch/reference/current/docs-bulk.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

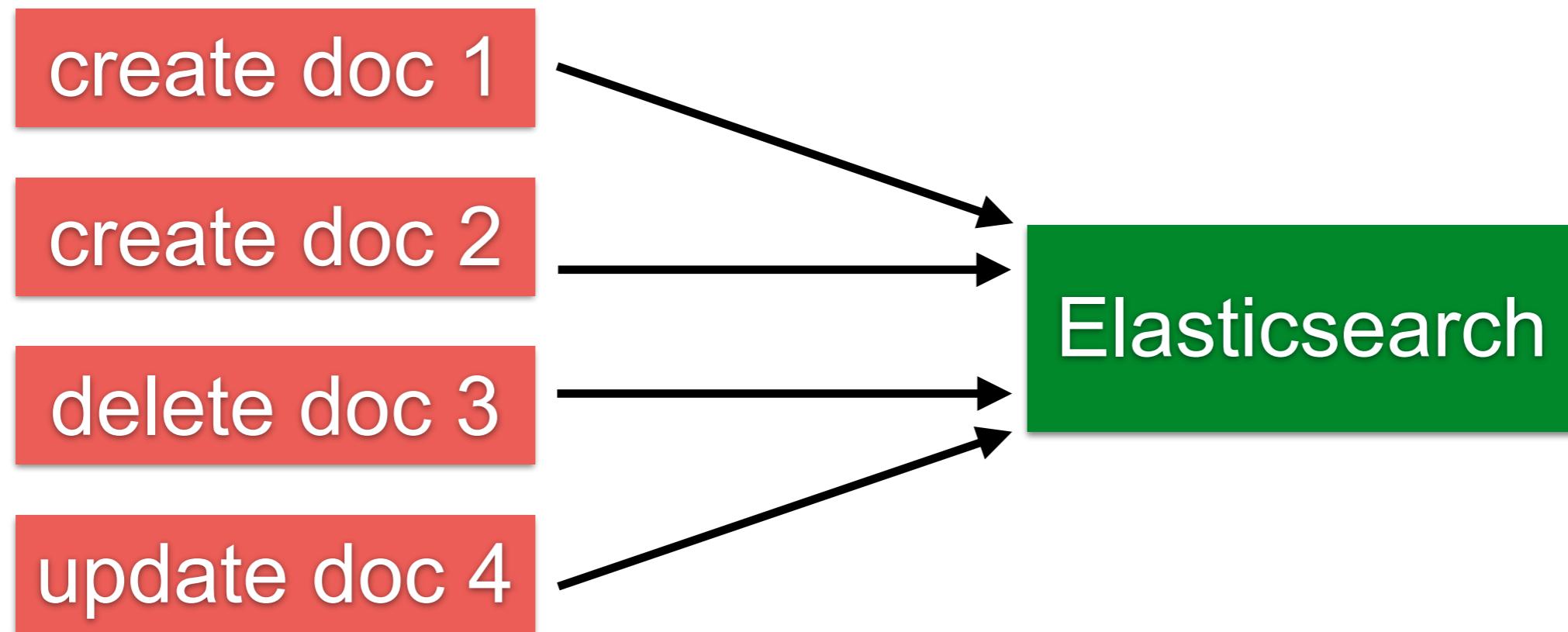
Bulk API

Perform many index/delete operation in single API call

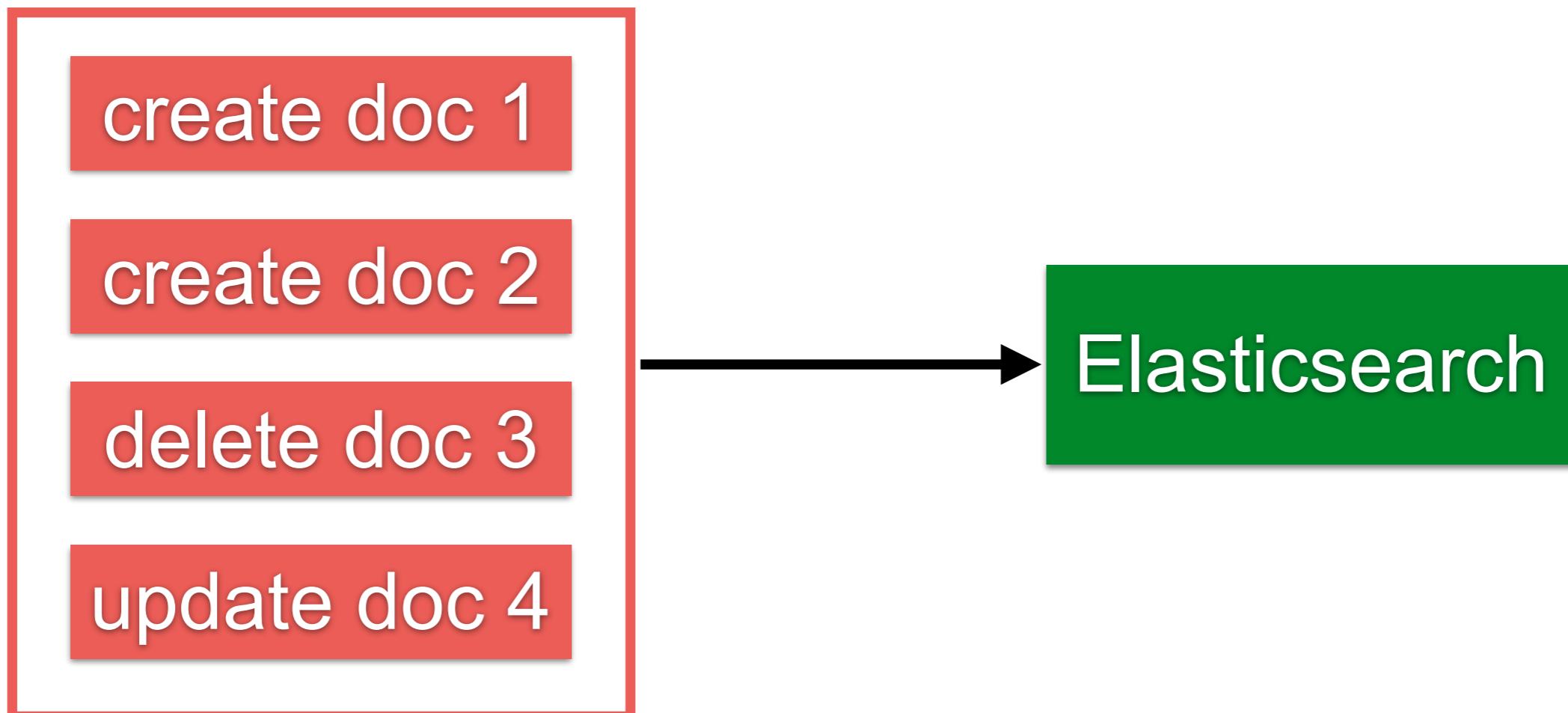
Increase indexing speed



Without Bulk API



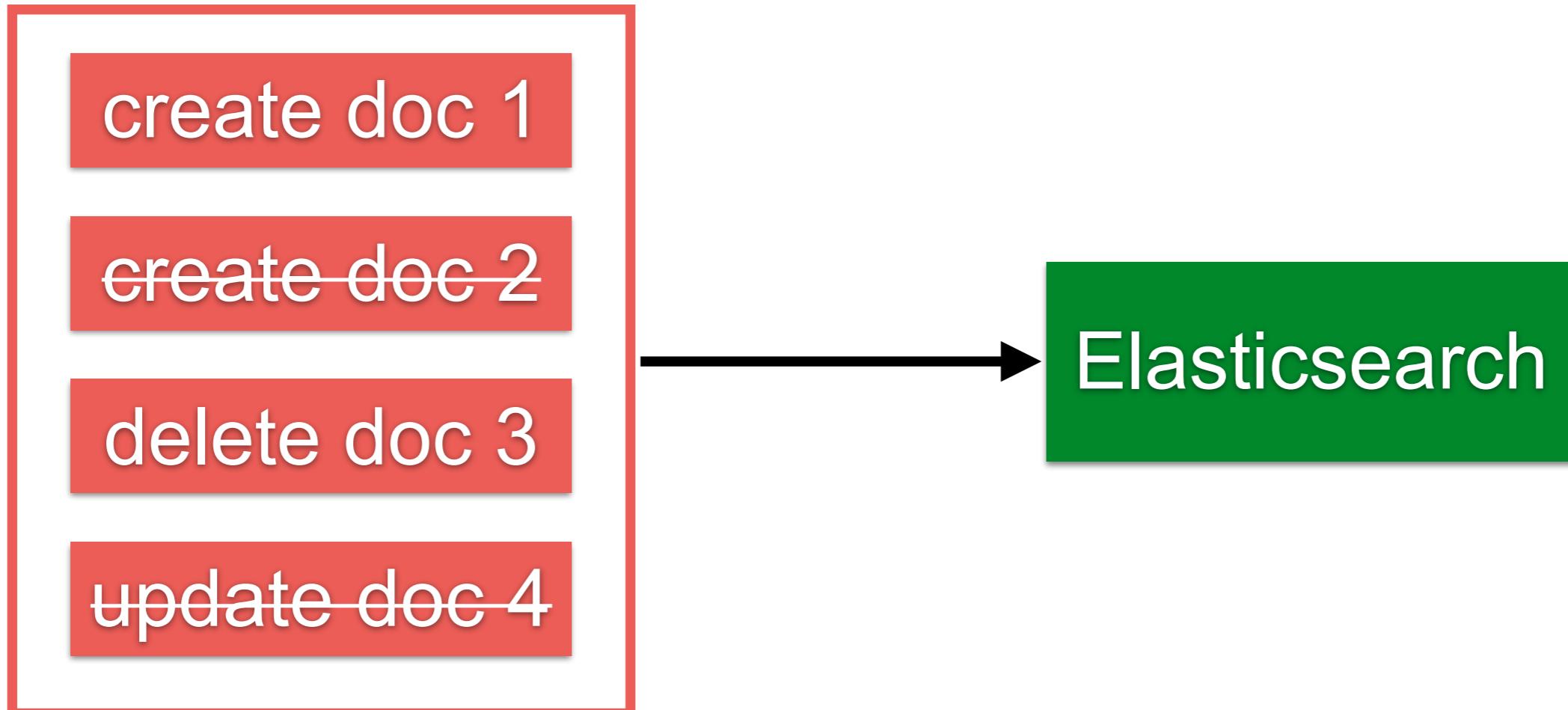
With Bulk API



Store in memory 5-15 MB



No transaction in bulk api



Create a document

POST /store/book/_bulk

```
{"create":{"_id":"1001"}  
{"title":"new book 1000","description":"my new book"}}
```



Response from Bulk API

```
{  
  "took": 89,  
  "errors": false,  
  "items": [  
    {  
      "create": {  
        "_index": "store",  
        "_type": "book",  
        "_id": "1001",  
        "_version": 1,  
        "result": "created",  
        "_shards": {  
          "total": 2,  
          "successful": 1,  
          "failed": 0  
        },  
        "_seq_no": 0,  
        "_primary_term": 1,  
        "status": 201  
      }  
    }  
  ]  
}
```

Time in milliseconds

HTTP Status 201 = Created



Search API

04-search/search_api.json



Query DSL

04-search/book_search.json



Query DSL

Domain Specific Language for query data
Flexible query language
Based on JSON format

<https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl.html>

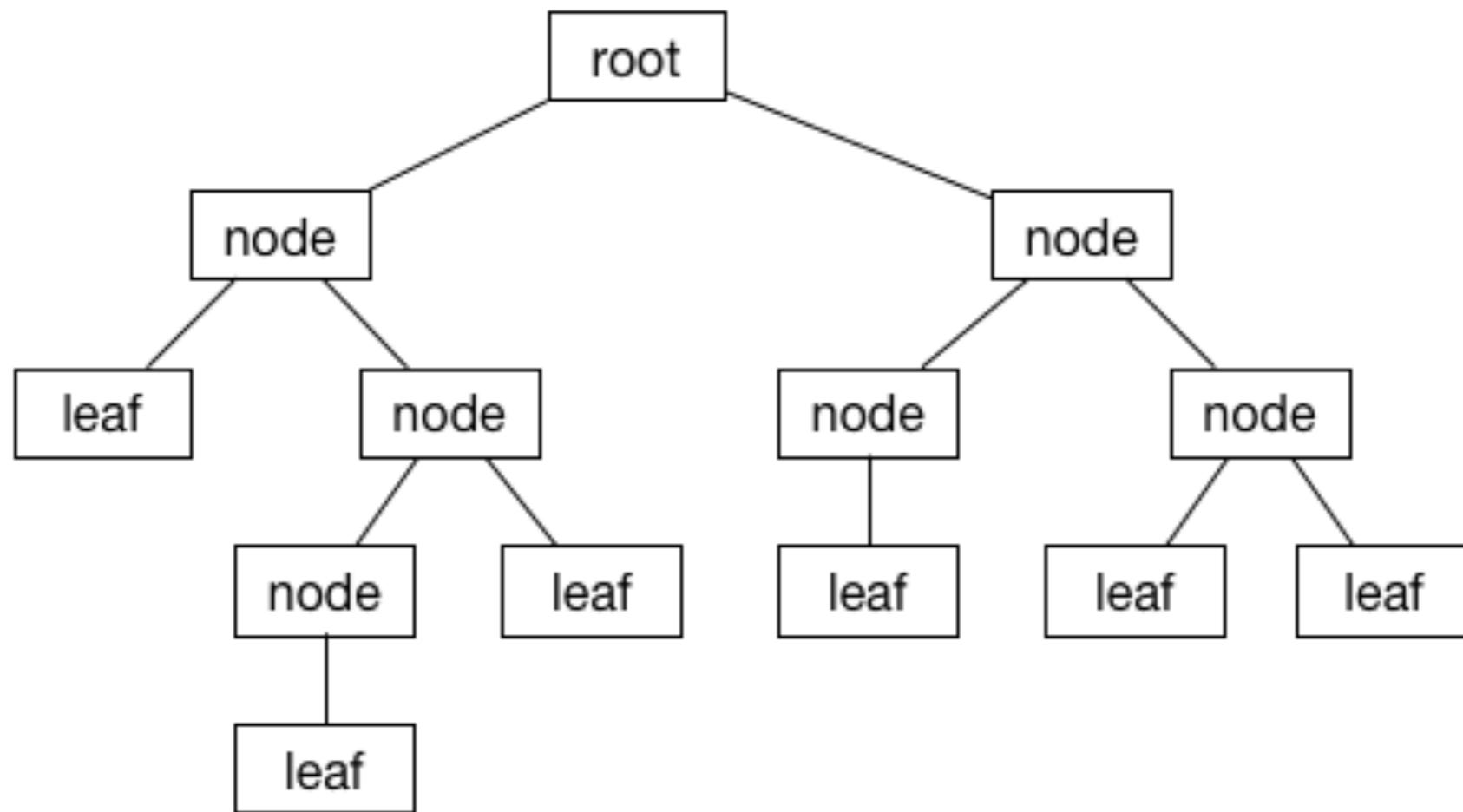


ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

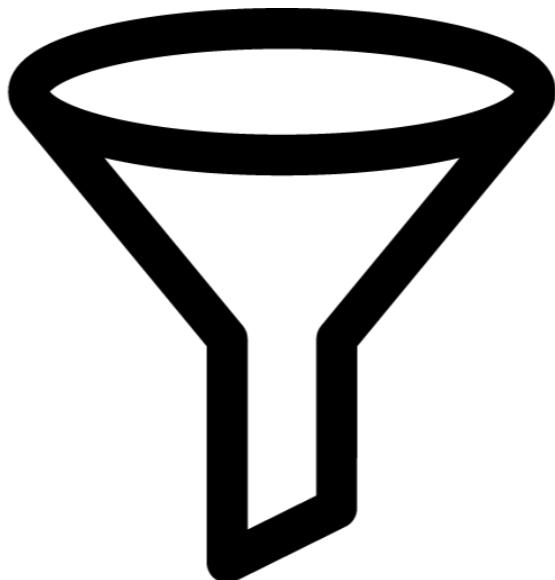
Query DSL

1. Leaf query clause
2. Compound query clause



Query DSL

Query (unstructured data)
Filter (structured data)



Query



ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

Query DSL

Query	Filter
Relevance	Boolean, yes/no
Full text search	Exact values
Not cached	Cached
Slower	Faster

Filter first, then query remaining documents



Query DSL

Full text query
Term level query
Compound query
Joining query
Geo query
Specialized query
Span query



Leaf query clause

GET /store/book/_search

```
{  
  "query": {  
    "match_all": {}  
  }  
}
```



Compound query clause

GET /store/book/_search

```
{  
  "query": {  
    "bool": {  
      "must": [{}],  
      "should": [{}],  
      "must_not": [{}]  
    }  
  }  
}
```



Use case

amazon

All ▾ elasticsearch

New to Amazon? Click here to learn more

Deliver to Thailand

Departments ▾ Your Amazon.com Today's Deals Gift Cards Sell

EN ▾ Hello. Sign in Account & Lists ▾ Orders Cart 0

1-16 of 119 results for "elasticsearch"

Show results for

Books

- Computers & Technology
- Data Processing
- Web Development & Design
- Online Internet Searching
- Databases & Big Data
- ▼ See more

Kindle Store

- Computers & Technology
- Business Software
- Search Engines
- Application Development
- Computer Databases
- ▼ See more
- ▼ See All 8 Departments

Refine by

Book Language

- English

Book Format

- Paperback

Packt

SPONSORED BY PACKT PUBLISHING

Complete Database Solutions with PostgreSQL

Shop now ▾

SQL Server 2017 Administrator's Guide

By Waheed Ahmad and Imanuele Pollicino

PostgreSQL 9.6 High Performance

By Waheed Ahmad, Gregory Smith

SQL Server 2017 Administrators Guide

3 ★★★★★ prime

PostgreSQL 9.6 High Performance: Optimize your...

1 ★★★★★ prime

Advertisement

Sponsored ⓘ

Learning Elastic Stack 6.0: A beginner's guide to distributed search, analytics, and visualization using Elasticsearch, Logstash and Kibana

Dec 22, 2017

by Pranav Shukla and Sharath Kumar M N

Eligible for Shipping to Thailand

Get to grips with the new features introduced in Elastic Stack 6.0, and deliver end-to-end real-time distributed data processing solutions.

Paperback

\$34⁹⁹

In Stock

★★★★★ 7

Previous Page 1 2 3 ... 8 Next Page



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Query

amazon

All ▾ elasticsearch

Departments ▾ Your Amazon.com Today's Deals Gift Cards Sell

New to Amazon? EN Hello. Sign Account Lists Orders Cart

Sort by Featured

Filter

Books

- Computers & Technology
- Data Processing
- Web Development & Design
- Online Internet Searching
- Databases & Big Data
- ▼ See more

Kindle Store

- Computers & Technology
- Business Software
- Search Engines
- Application Development
- Computer Databases
- ▼ See more
- ▼ See All 8 Departments

Refine by

Book Language

- English

Book Format

- Paperback

1-16 of 119 results for "elasticsearch"

Packt

SPONSORED BY PACKT PUBLISHING

Complete Database Solutions with PostgreSQL

Shop now ›

Learning Elastic Stack 6.0

A beginner's guide to distributed search, analytics, and visualization using Elasticsearch, Logstash, and Kibana

By Pranav Shukla and Sharath Kumar M N

Eligible for Shipping to Thailand

Get to grips with the new features introduced in Elastic Stack 6.0, and deliver end-to-end real-time distributed data processing solutions.

Paperback \$

Paging

Previous Page 1 2 3 ... 8 Next Page

Sorting

SQL Server 2017 Administrator's Guide

PostgreSQL 9.6 High Performance

PostgreSQL 9.6 High Performance: Optimize your...

Advertisement



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Aggregation API

05-aggregation/book_aggregation.json

<https://www.elastic.co/guide/en/elasticsearch/reference/current/search-aggregations.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

```
SELECT count(1), sum(price)  
FROM some_table  
GROUP BY some_column
```



Aggregation Types

Bucketing
Metric
Matrix
Pipeline



Structure

```
"aggregations" : {  
    "<aggregation_name>" : {  
        "<aggregation_type>" : {  
            <aggregation_body>  
        }  
        [ , "meta" : { [ <meta_data_body> ] } ]?  
        [ , "aggregations" : { [ <sub_aggregation> ]+ } ]?  
    }  
    [ , "<aggregation_name_2>" : { ... } ]*  
}
```



Count by category

GET /store/book/_search

```
{  
  "aggs": {  
    "all_book_title": {  
      "terms": {  
        "field": "category.keyword"  
      }  
    }  
  }  
}
```



Count by category

GET /store/book/_search

```
{  
  "aggs": {  
    "all_book_title": {  
      "terms": {  
        "field": "category.keyword"  
      }  
    }  
  }  
}
```

Aggregation name



Count by category

GET /store/book/_search

```
{  
  "aggs": {  
    "all_book_title": {  
      "terms": { Aggregation type  
        "field": "category.keyword"  
      }  
    }  
  }  
}
```



Result of aggregation

```
{  
  "hits": {  
    "total": 5,  
    "max_score": 1,  
    "hits": [  
      {  
        "_source": {  
          "title": "The Logstash Book"  
        }  
      },  
      {  
        "_source": {  
          "title": "Elasticsearch Server: Second Edition"  
        }  
      }  
    ]  
  }  
}
```

Search result



Result of aggregation

```
"aggregations": {  
    "all_book_title": {  
        "doc_count_error_upper_bound": 0,  
        "sum_other_doc_count": 0,  
        "buckets": [          Aggregation result  
            {  
                "key": "Computer & Technology",  
                "doc_count": 5  
            },  
            {  
                "key": "Online Searching",  
                "doc_count": 3  
            },  
            {  
                "key": "Java Programming",  
                "doc_count": 2  
            }  
        ]  
    }  
}
```



Show only aggregation result

GET /store/book/_search

```
{  
  "size": 0, Set search result size = 0  
  "aggs": {  
    "all_book_title": {  
      "terms": {  
        "field": "category.keyword"  
      }  
    }  
  }  
}
```



Range of price

GET /store/book/_search

```
{  
  "size": 0,  
  "aggs": {  
    "price_range": {  
      "range": {  
        "field": "price",  
        "ranges": [  
          { "from": 0, "to": 10 },  
          { "from": 11, "to": 20 },  
          { "from": 21, "to": 50 }  
        ]  
      }  
    }  
  }  
}
```



Result of aggregation

```
"buckets": [
  {
    "key": "0.0-10.0",
    "from": 0,
    "to": 10,
    "doc_count": 1
  },
  {
    "key": "11.0-20.0",
    "from": 11,
    "to": 20,
    "doc_count": 0
  },
  {
    "key": "21.0-50.0",
    "from": 21,
    "to": 50,
    "doc_count": 3
  }
]
```



Range of price and ordering

GET /store/book/_search

```
{  
  "size": 0,  
  "aggs": {  
    "price_range": {  
      "range": {  
        "field": "price",  
        "ranges": [  
          { "from": 0, "to": 10 },  
          { "from": 11, "to": 20 },  
          { "from": 21, "to": 50 }  
        ]  
      }  
    }  
  }  
}
```



Workshop aggregation with car

05-aggregation/car.json



Try by yourself

Best seller by color

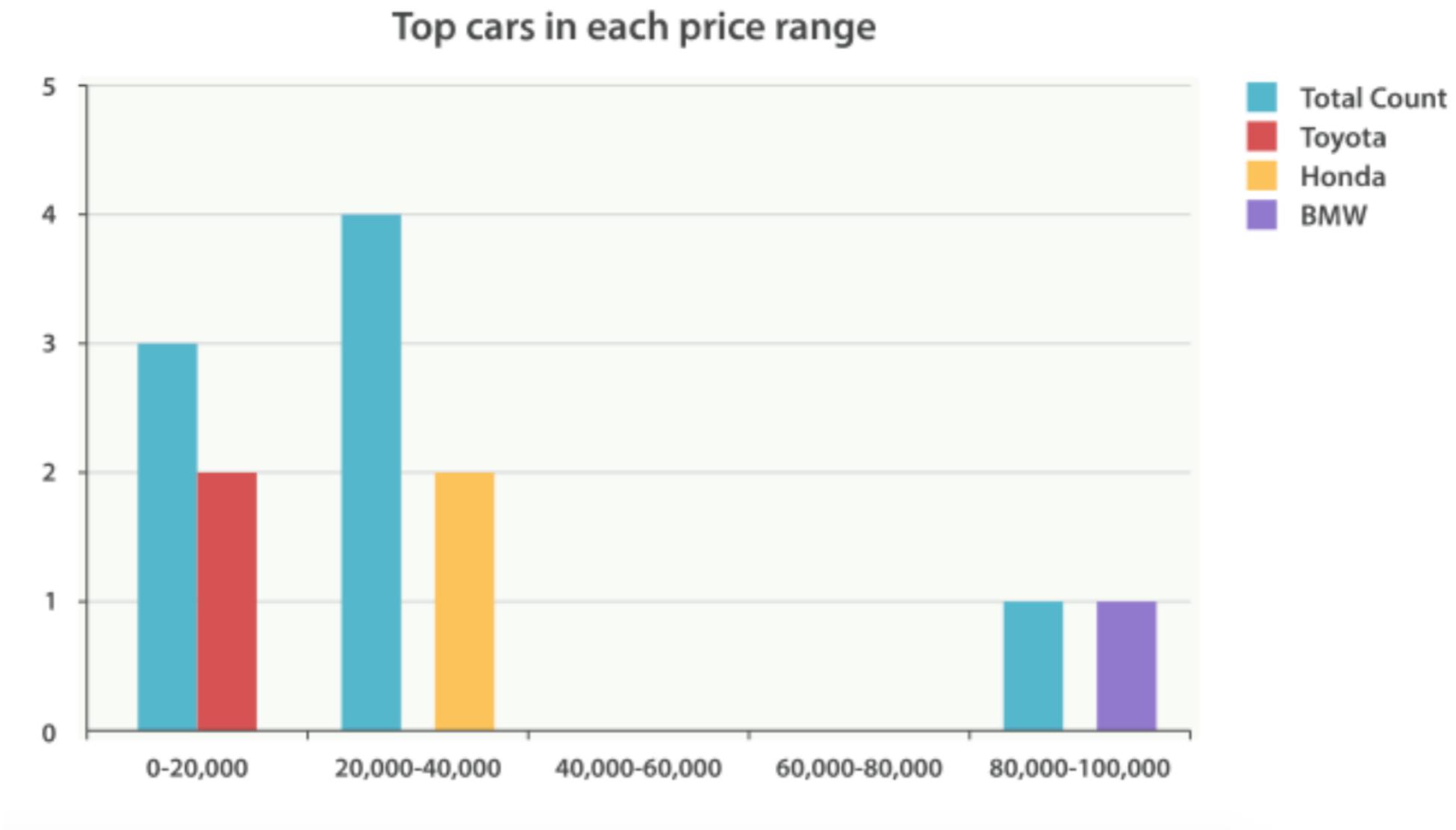
Statistic of best seller by color

Detail of car in each color

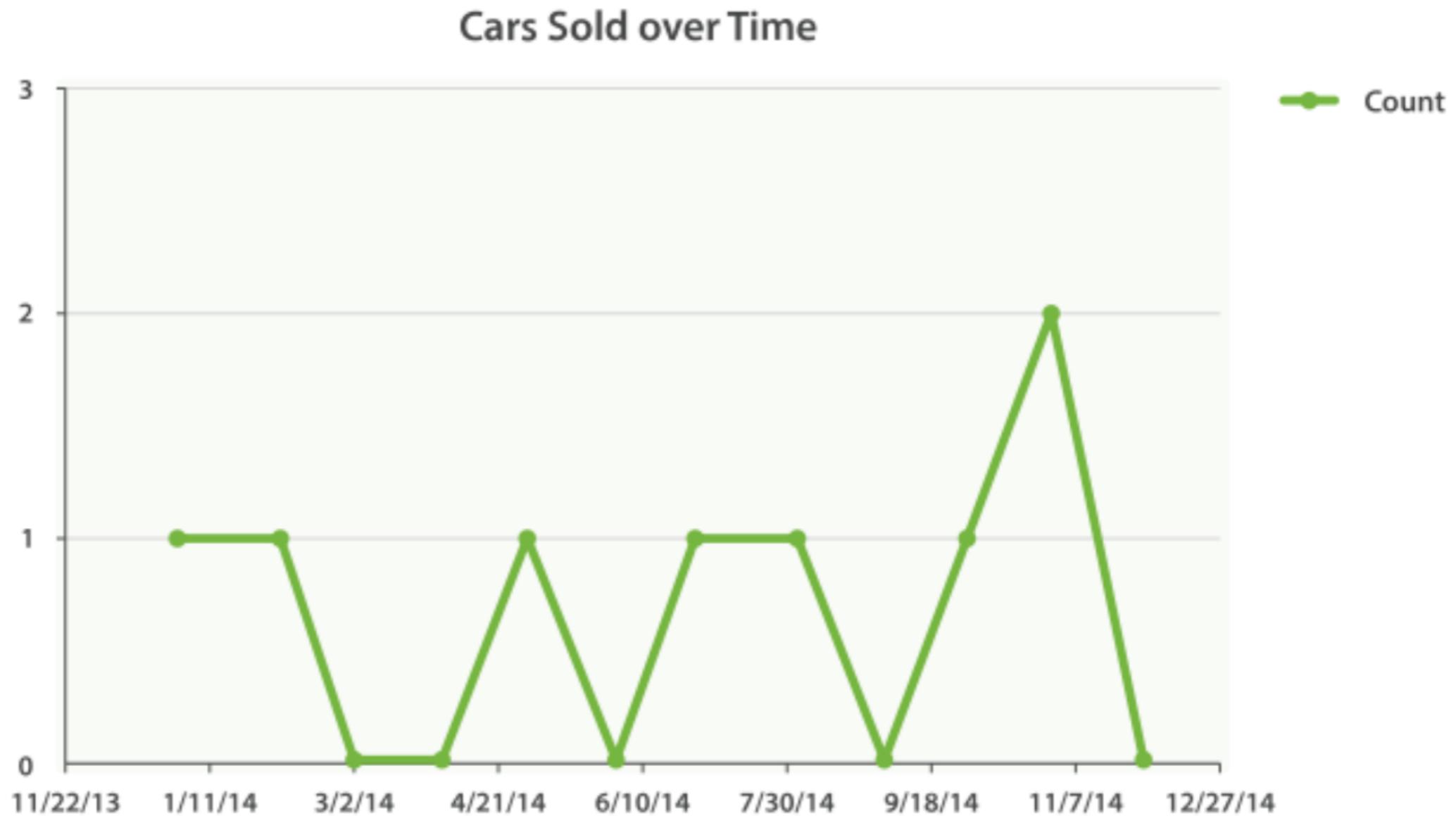
Min/max of price by make



Top cars in each price range ?



Cars sold over Time ?



Mapping

<https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Mapping type

Meta-fields

Field or properties



Meta-field

Metadata of document
`_index, _type, _id, _source`



Field or properties

List of fields or properties of document



Mapping/Schema of document

GET /store/_mapping/book

```
"mappings": {  
    "book": {  
        "properties": {  
            "author name": {  
                "type": "text",  
                "fields": {  
                    "keyword": {  
                        "type": "keyword",  
                        "ignore_above": 256  
                    }  
                }  
            }  
        }  
    }  
}
```



Mapping/Schema of document

GET /store/_mapping/book

```
"mappings": {  
    "book": {  
        "page": {  
            "type": "long"  
        },  
        "price": {  
            "type": "float"  
        },  
        "published_date": {  
            "type": "date"  
        }  
    }  
}
```



Field Datatypes

text	date
keyword	ip
long	boolean
double	completion
geo_point	geo_shape
array	object
nested	binary

<https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-types.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Field Datatypes

text	date
keyword	ip
long	boolean
double	completion
geo_point	geo_shape
array	object
nested	binary

<https://www.elastic.co/guide/en/elasticsearch/reference/current/mapping-types.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Array

No data type **array** in Elasticsearch

["name", "title"]	array of string
[1, 2, 3]	array of integer
[{"name": "up1", "age": 30}]	array of object



Mapping configuration

Maximum number of fields = 1,000

Maximum depth of fields = 20

Maximum depth of nested fields = 50



Dynamic mapping

Fields and mapping types not need to defined before being used



Analyzer

07-analyzer

<https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

More tools



Elasticsearch Head

 **ElasticSearch Head**
offered by travistx

★★★★★ (75) | [Developer Tools](#) | 45,312 users

[OVERVIEW](#) | [REVIEWS](#) | [SUPPORT](#) | [RELATED](#)

ElasticSearch http://192.168.7.8:9200/ Connect Rick cluster health: yellow (6, 18)

Overview Browser Structured Query Any Request Info Status Nodes Stats Cluster Nodes Cluster State Cluster Health

Cluster Overview New Index

	cu_docs	bnil	cu_msg	anvil
Leon	size: 180Gb (540Gb) docs: 995131 (995131)	size: 80kb (480kb) docs: 90 (90)	size: 313Gb (1.56Tb) docs: 10047450 (10140915)	index: close
Pris	Info Actions	Info Actions	Info Actions	Info Actions
Rick	0 1	0 1	0 1 2 3 4	0 1 2 3 4
Rachel	1 2	0 1	0 1 2 3 4	0 1 2 3 4
Zhora	1 2	0 1	0 1 2 3 4	0 1 2 3 4
Roy	0 2	0 1	0 1 2 3 4	0 1 2 3 4
Unassigned	0 0	0 1	0 1 2 3 4	0 1 2 3 4

Compatible with your device

ElasticSearch Head
Chrome Extension containing the excellent ElasticSearch Head application.

[Website](#) | [Report Abuse](#)

Additional Information
Version: 0.1.3
Updated: December 4, 2017
Size: 434KiB
Language: English (United States)



<https://github.com/mobz/elasticsearch-head>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Elasticsearch Dump



<https://github.com/taskrabbit/elasticsearch-dump>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Kibana



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

110

Geo Location

08-geo-location/sample_geo.json



Geo Location Type

Geo-point
Geo-shape



Geo-point

Must pre-define in mapping of index

```
PUT /my_map
{
  "mappings": {
    "city": {
      "properties": {
        "name": {
          "type": "text"
        },
        "location": {
          "type": "geo_point"
        }
      }
    }
  }
}
```



Geo-point Format

Geo-point as object

Geo-point as string

Geo-point as array

Geo-point as geohash

<https://www.elastic.co/guide/en/elasticsearch/reference/current/geo-point.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Geo-point Format

Type	Format
Object	lat = lon =
String	lat, lon
Array	[lon, lat] ** GeoJSON **

<https://en.wikipedia.org/wiki/GeoJSON>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Geohash converter

Geohash Converter

Simple and fast conversion from geohash to latitude/longitude and from latitude/longitude to geohash.

GeoHash

Lat, Lng

Precision

<http://geohash.co/>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Geo-point query

Geo-bounding-box

Geo-distance

Geo-polygon

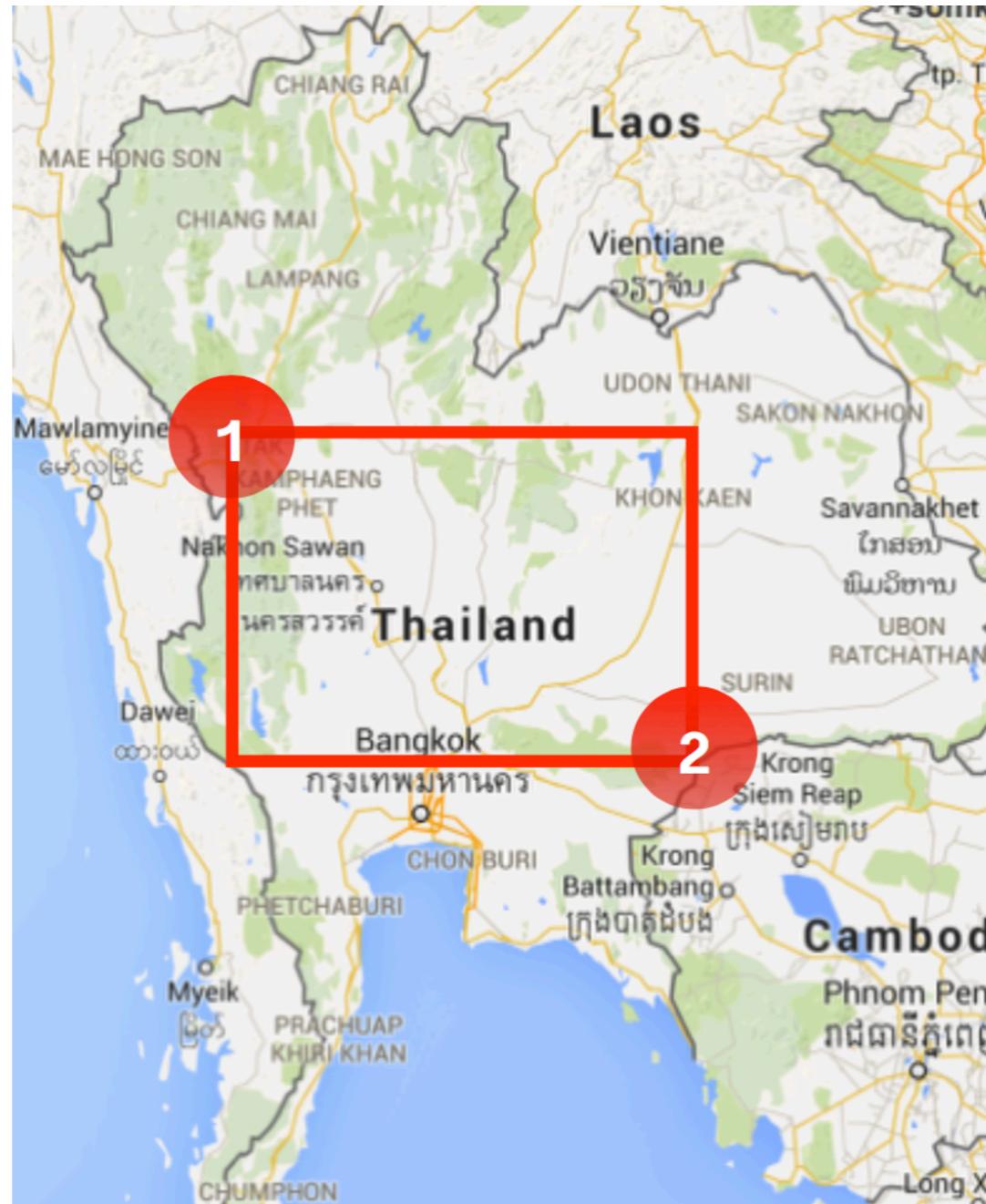
<https://www.elastic.co/guide/en/elasticsearch/reference/current/geo-queries.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Bounding Box



<https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-geo-bounding-box-query.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Geo Distance



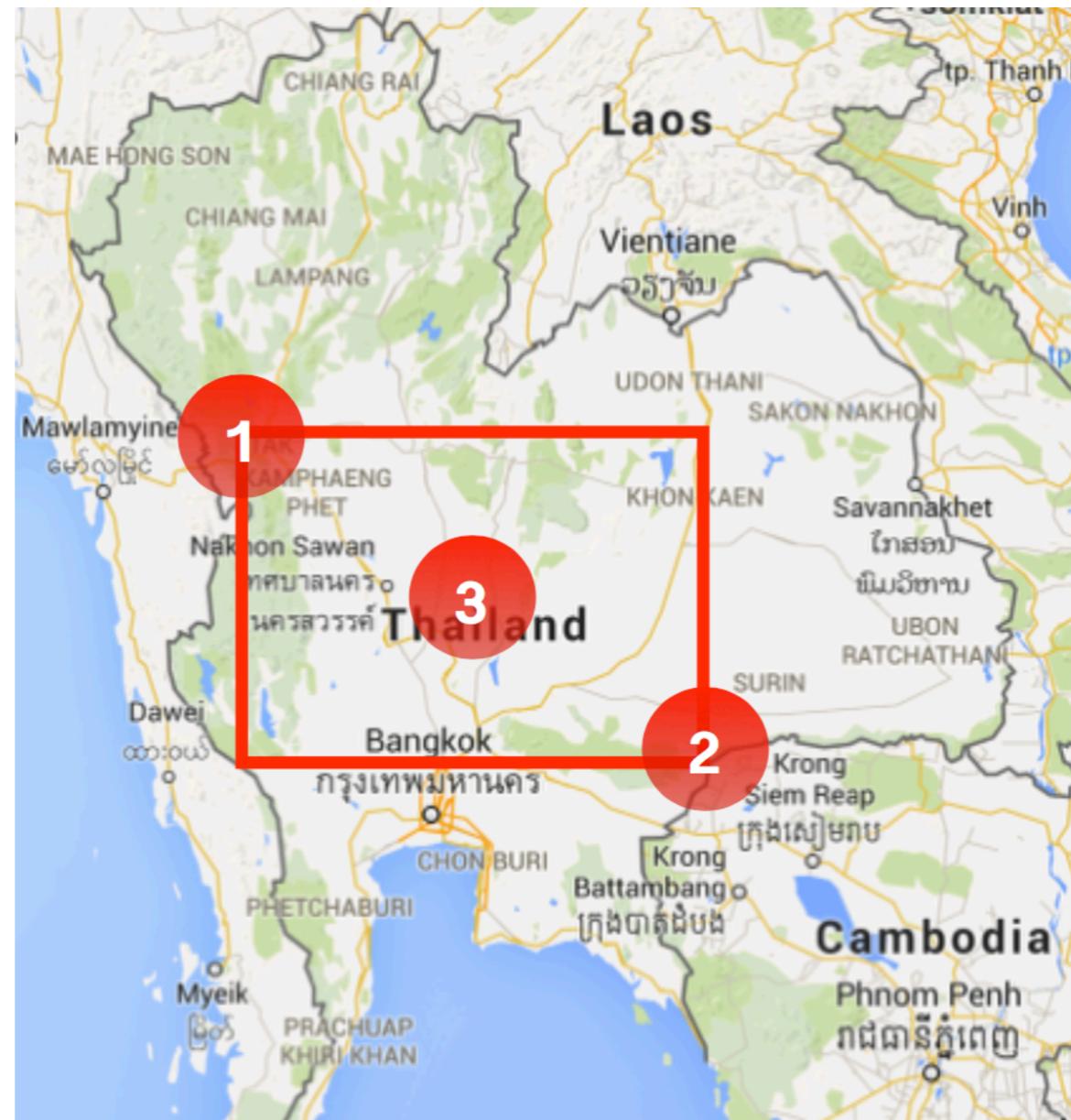
<https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-geo-distance-query.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Try to ordering result



Explain and Profiling your query



2 ways

Explain API
Profile API



Explain API

GET /my_map/_search

```
{  
  "explain": true,  
  "query": {  
    "bool": {
```

<https://www.elastic.co/guide/en/elasticsearch/reference/6.3/search-explain.html>



ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

123

Profile API

Debugging tool

Add overhead to search execution

Output is verbose and depend on internal operation

<https://www.elastic.co/guide/en/elasticsearch/reference/6.3/search-profile.html>



ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

Profile API

GET /my_map/_search

```
{  
  "profile": true,  
  "query": {  
    "bool": {
```



Working with Data

<https://www.elastic.co/guide/en/kibana/current/tutorial-load-dataset.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Working with Data

\$elasticsearch-plugin install **ingest-geoip**

<https://www.elastic.co/guide/en/elasticsearch/plugins/current/ingest-geoip.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

GeolP with Elasticsearch

geoip/instruction.json



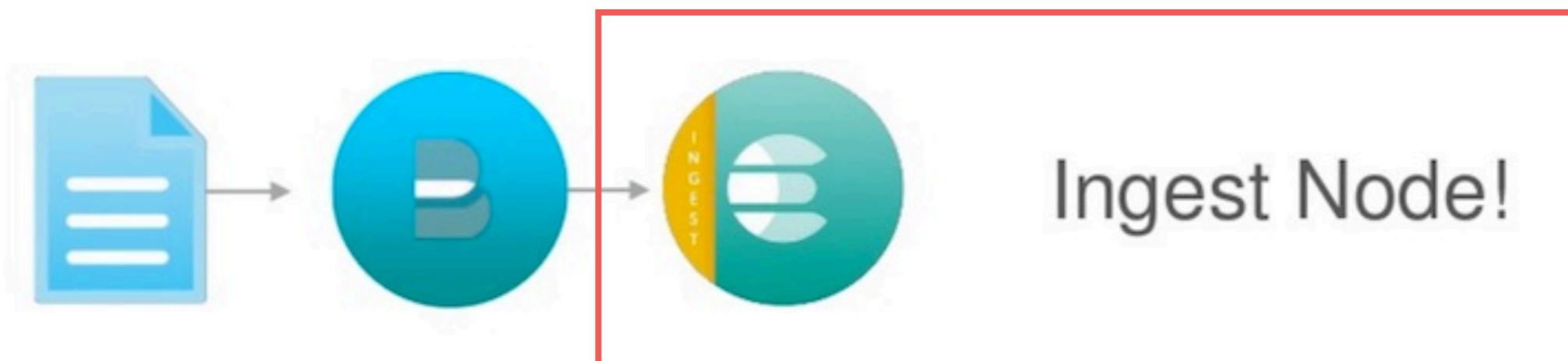
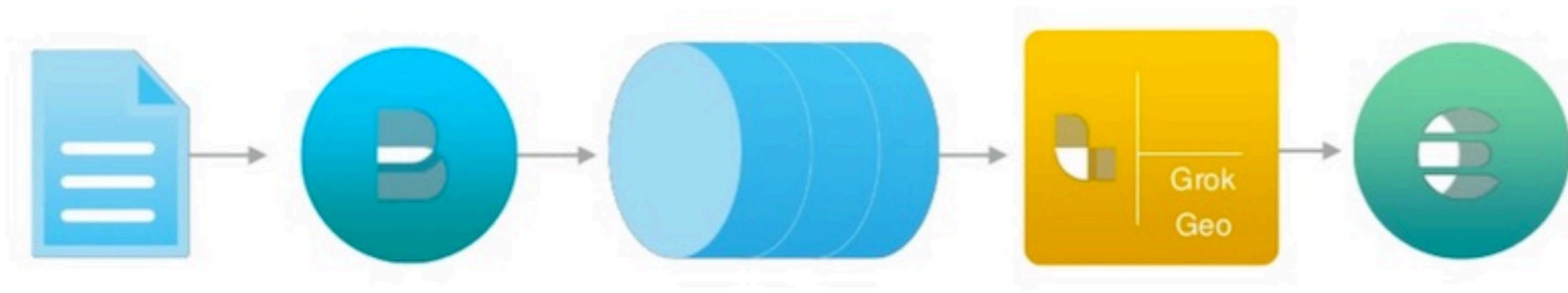
Sample Data

```
{"index":{"_index":"logstash-2015.05.18","_type":"log"}},  
{"@timestamp":"2015-05-18T09:03:25.877Z","ip":"185.124.182.12"},  
{"index":{"_index":"logstash-2015.05.18","_type":"log"}},  
{"@timestamp":"2015-05-18T12:28:25.013Z","ip":"79.1.14.87","e"},  
{"index":{"_index":"logstash-2015.05.18","_type":"log"}},  
{"@timestamp":"2015-05-18T17:44:34.357Z","ip":"178.209.1.7"},  
{"index":{"_index":"logstash-2015.05.18","_type":"log"}},  
{"@timestamp":"2015-05-18T13:04:18.120Z","ip":"118.140.92.127"},  
{"index":{"_index":"logstash-2015.05.18","_type":"log"}},  
{"@timestamp":"2015-05-18T11:37:40.653Z","ip":"235.154.34.221"},  
{"index":{"_index":"logstash-2015.05.18","_type":"log"}},  
{"@timestamp":"2015-05-18T08:46:07.025Z","ip":"228.216.38.41"}
```



Working with Ingest

Pre-process document before actual indexing



<https://www.elastic.co/guide/en/elasticsearch/reference/current/ingest.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Install plugin

```
$elasticsearch-plugin install ingest-geoip
```

<https://www.elastic.co/guide/en/elasticsearch/plugins/current/ingest-geoip.html>



ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

Working with Logstash

<https://www.elastic.co/guide/en/logstash/current/index.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Logstash



Input
Filter
Output



Design your input first !!

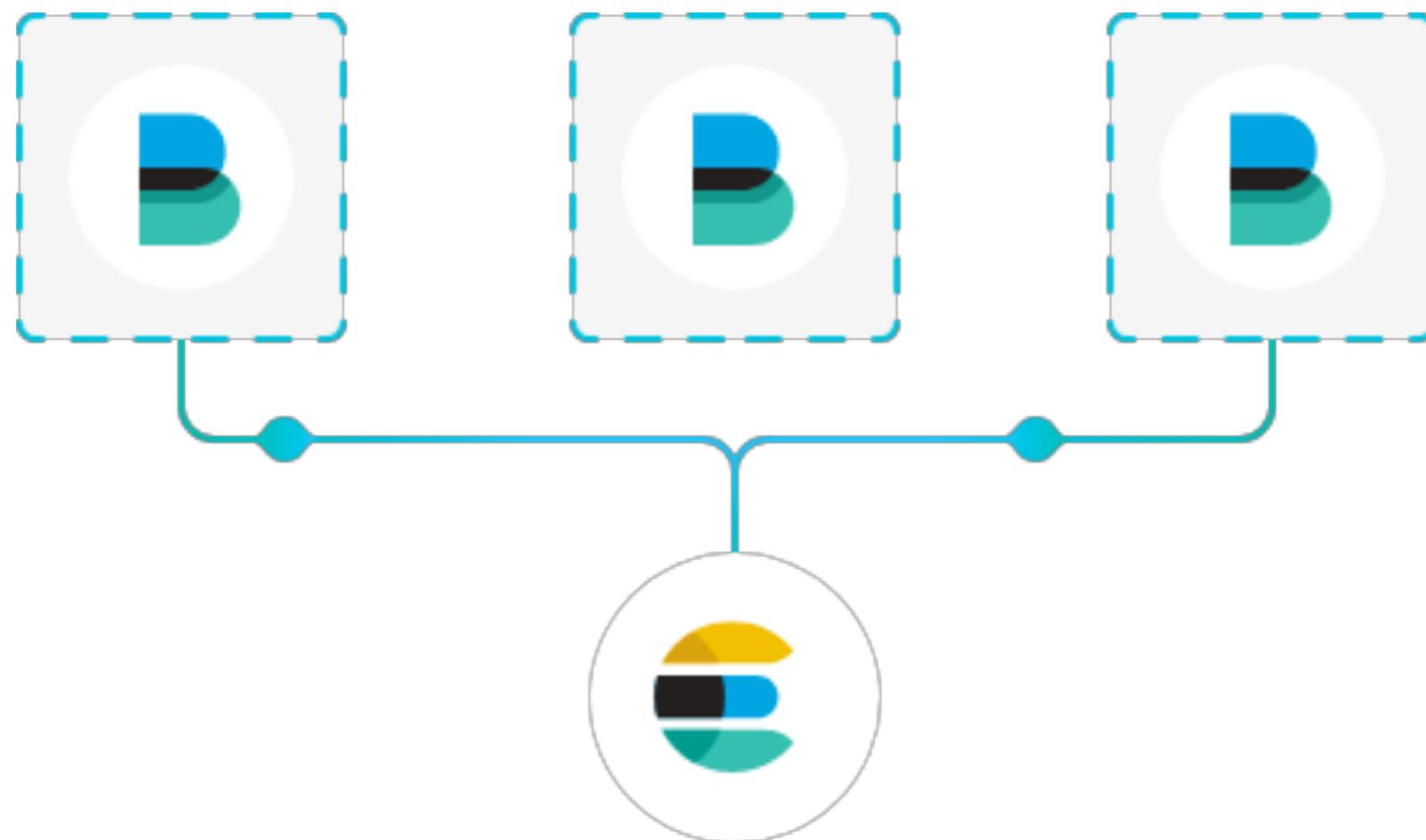


Use beats is better

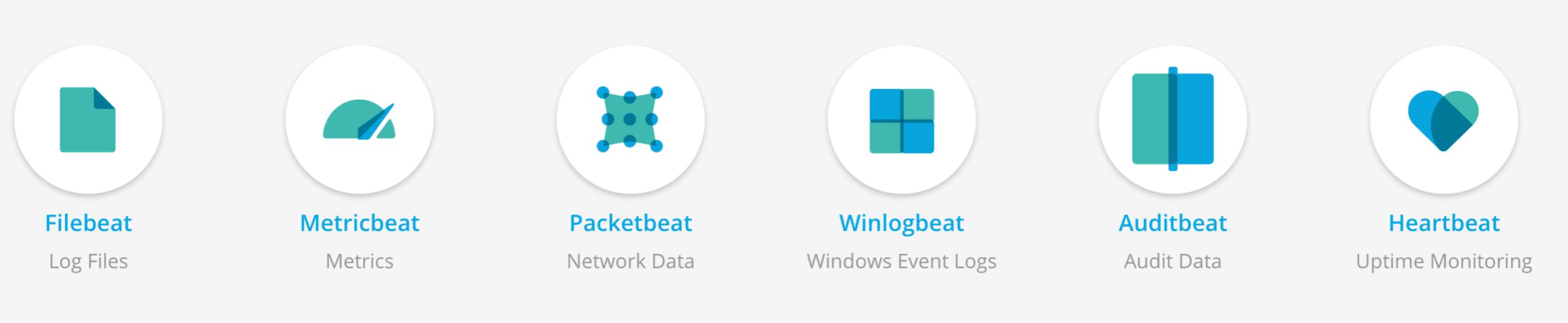
<https://www.elastic.co/products/beats>



Beat



Beat



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

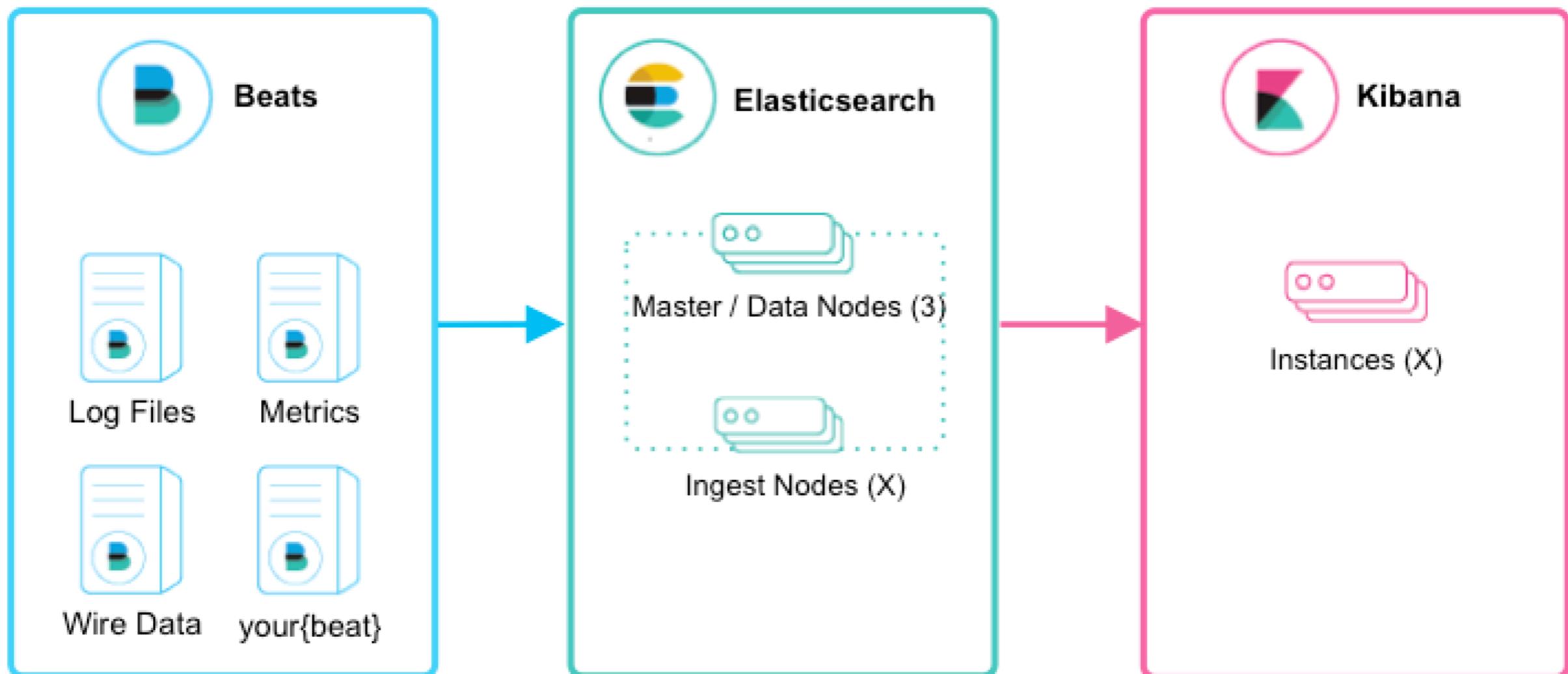
Example

```
$filebeat -e -c beat.yml -d "publish"
```

<https://www.elastic.co/guide/en/beats/filebeat/current/index.html>



Beat and Logstash



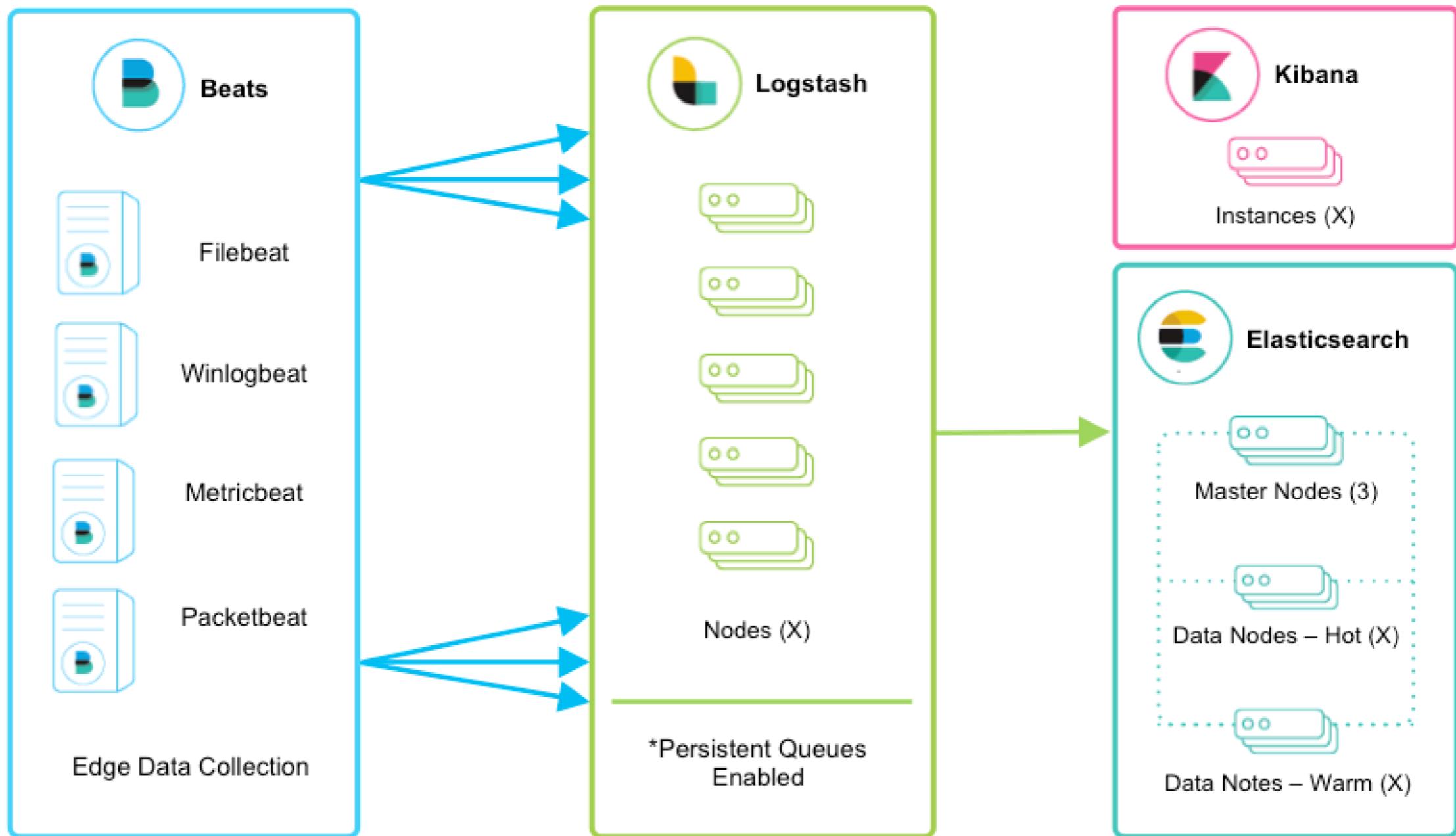
<https://www.elastic.co/guide/en/logstash/current/deploying-and-scaling.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Scaling



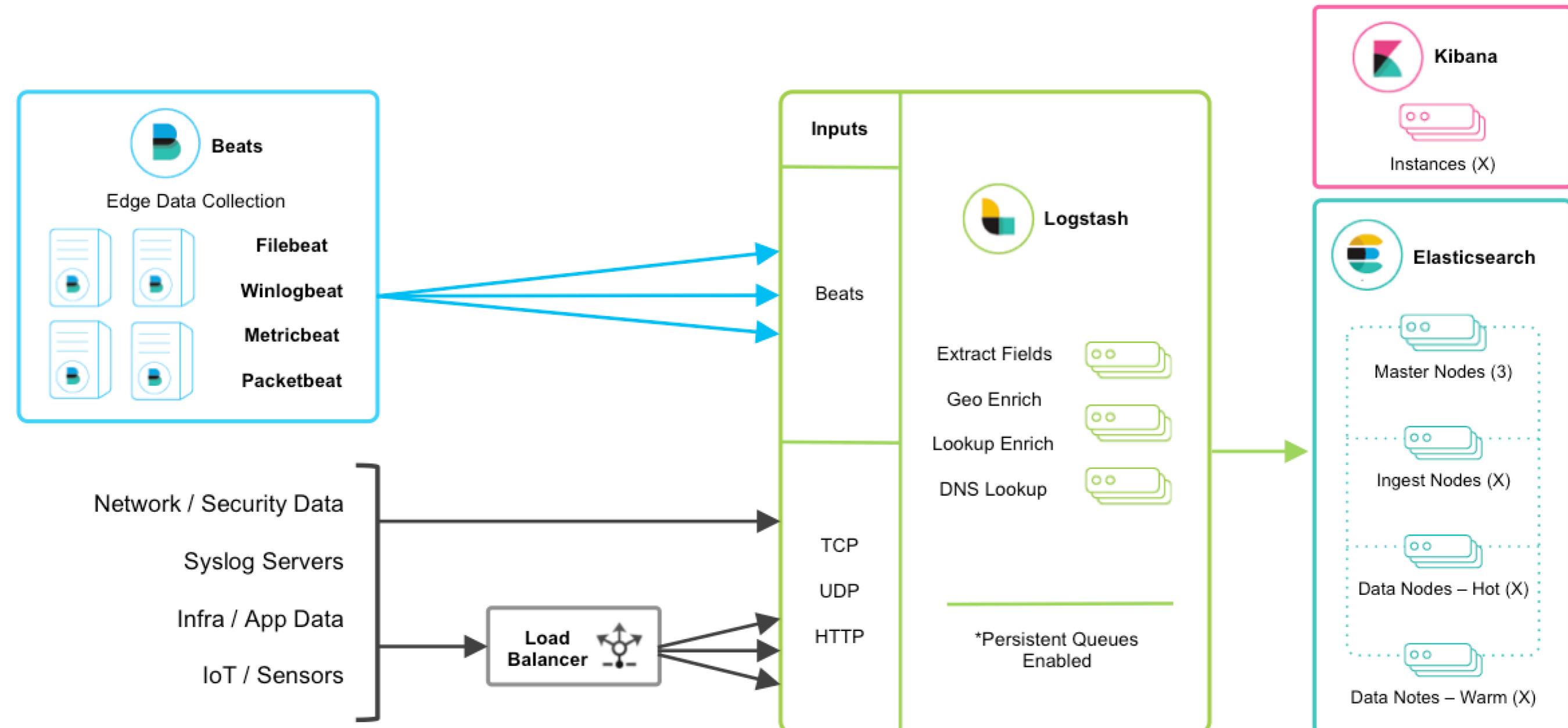
<https://www.elastic.co/guide/en/logstash/current/deploying-and-scaling.html>



ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

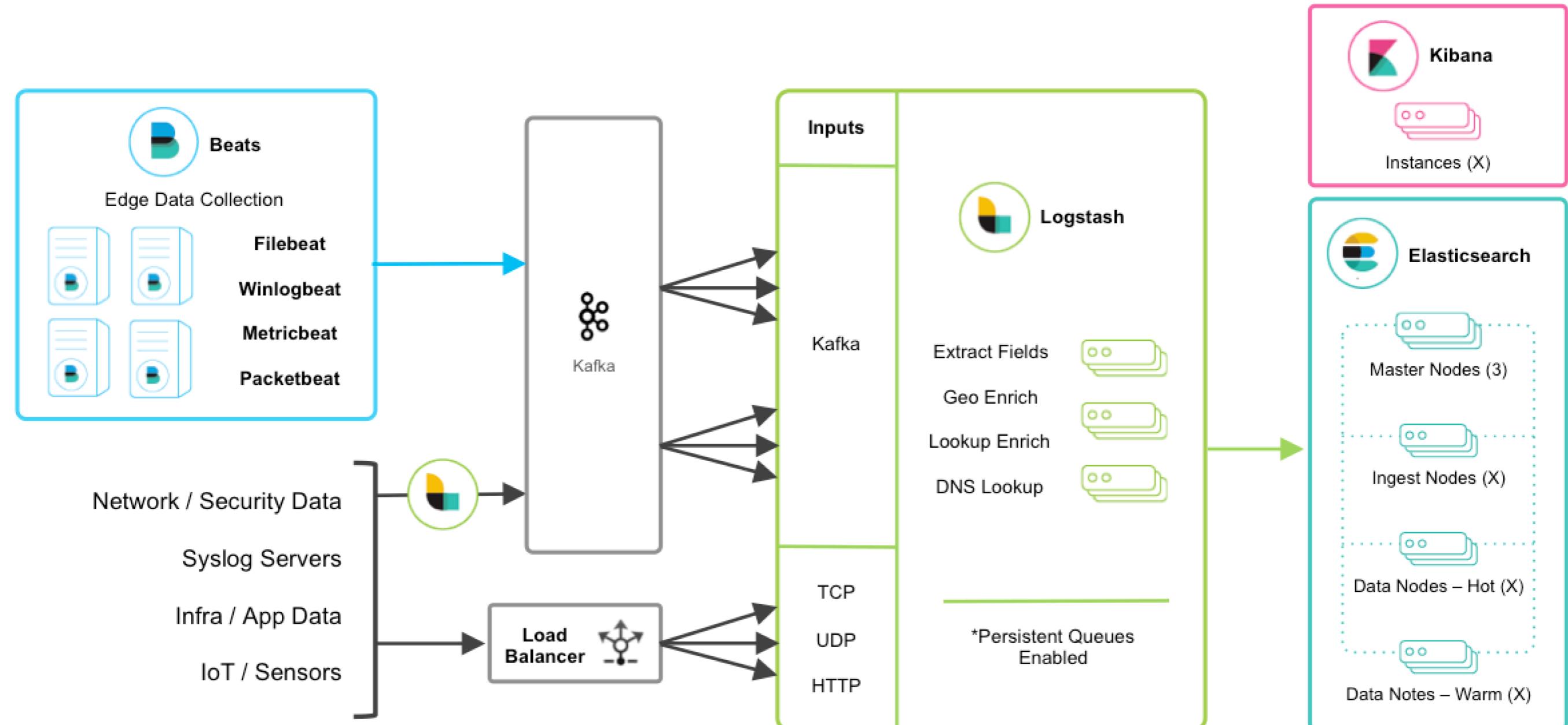
More data sources



<https://www.elastic.co/guide/en/logstash/current/deploying-and-scaling.html>



Use messaging Queue



<https://www.elastic.co/blog/logstash-persistent-queue>



ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

Design for Failure

09-cluster



Elasticsearch Nodes

Node Type	Description
Master	Control the cluster
Data	Keep/store data
HTTP/Query	Run your query
Coordinating	Smart Load Balancer
Ingest	Pre-processing documents before indexing

<https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-node.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

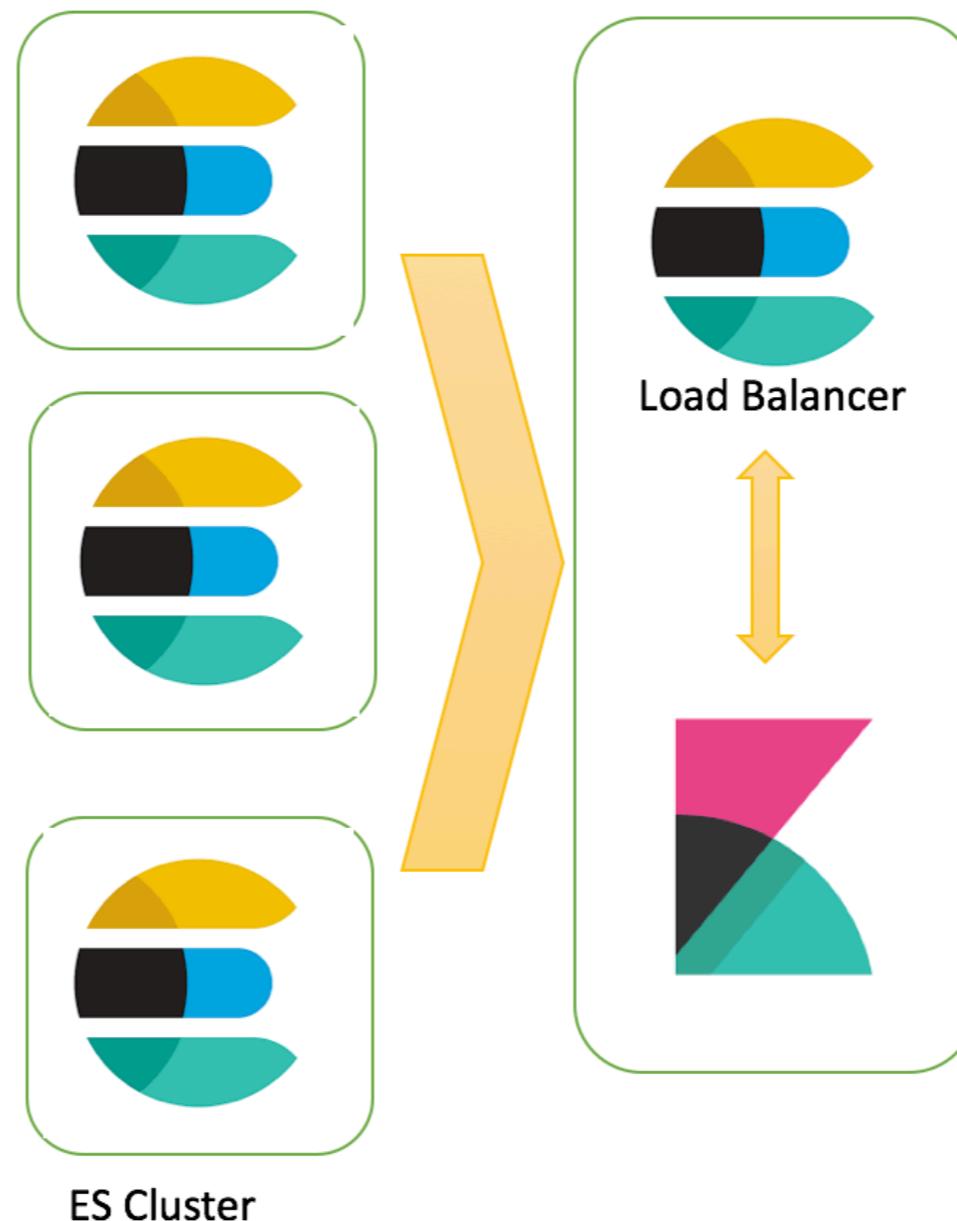
Elasticsearch Nodes

← → ⌂ ⓘ Not Secure | 35.240.161.188:9200/_cat/nodes?v&h=ip,name,node.role,master,heap.percent,ram.percent

ip	name	node.role	master	heap.percent	ram.percent
10.148.0.2	master	m	*	17	33
10.148.0.4	query	-	-	10	63
10.148.0.5	coordinator	-	-	9	78
10.148.0.3	data	d	-	13	63



Elasticsearch Nodes



<https://www.elastic.co/guide/en/kibana/current/production.html#load-balancing>



ELK Stack

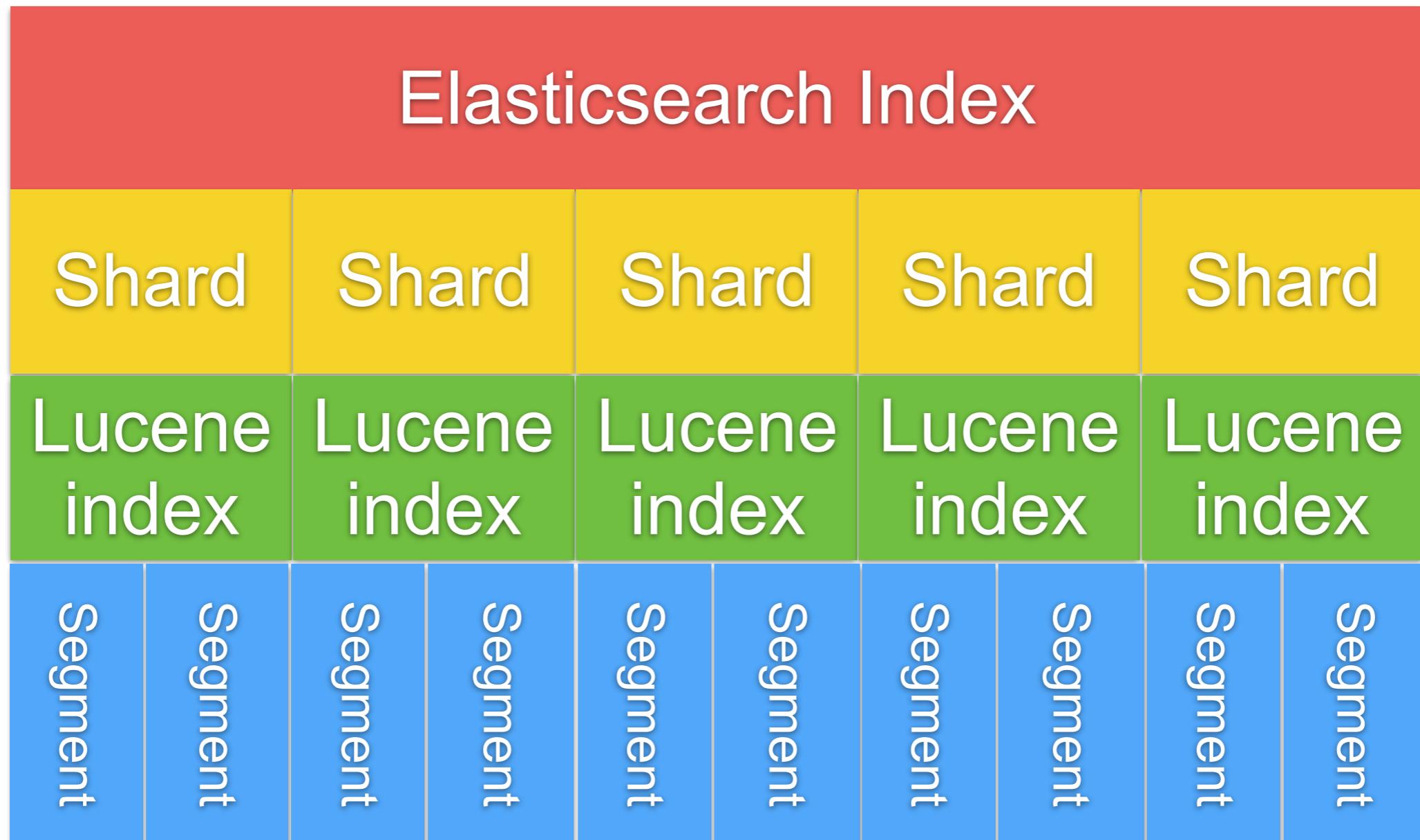
© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Apache Lucene

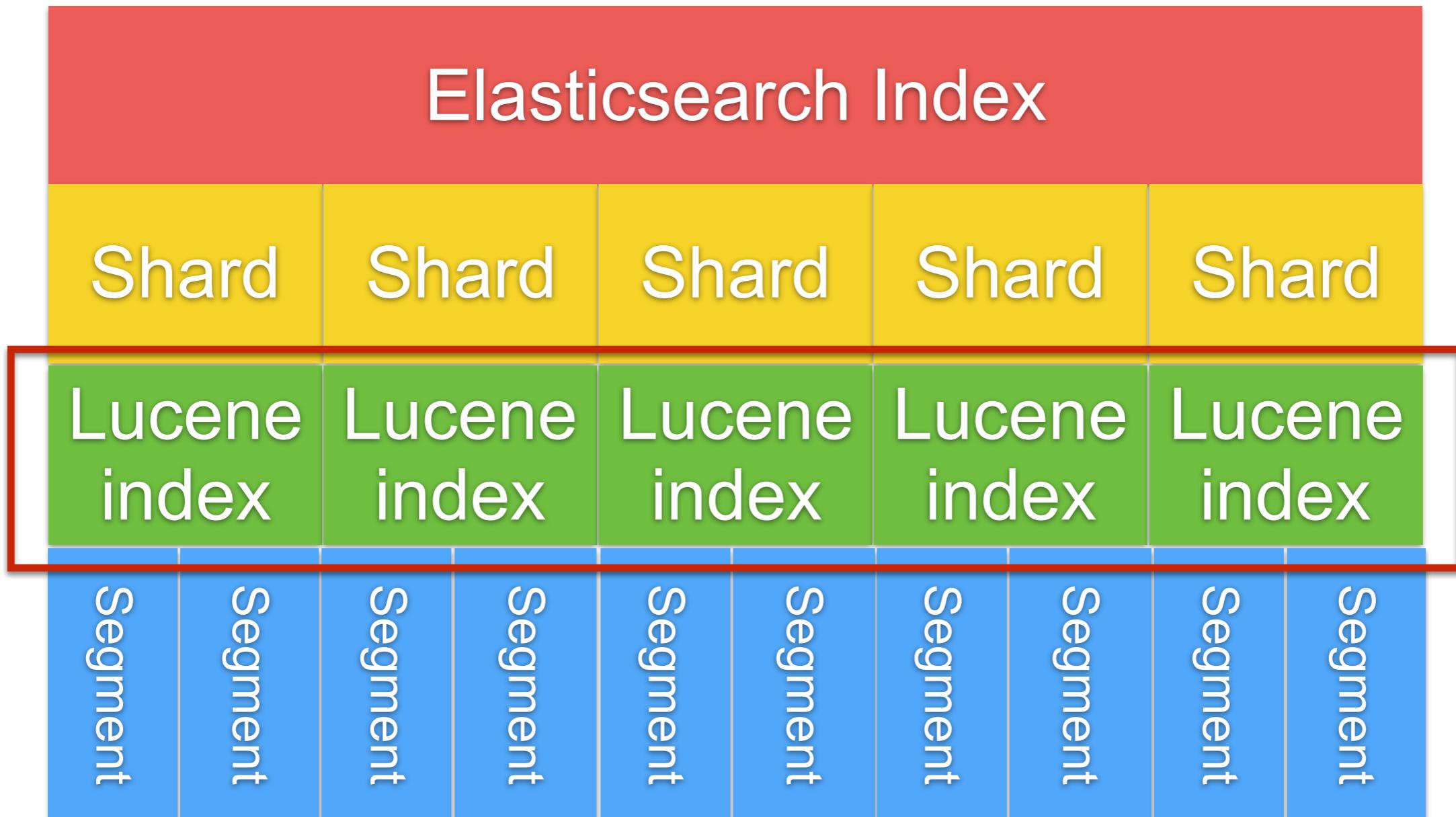
<http://lucene.apache.org/>



Apache Lucene



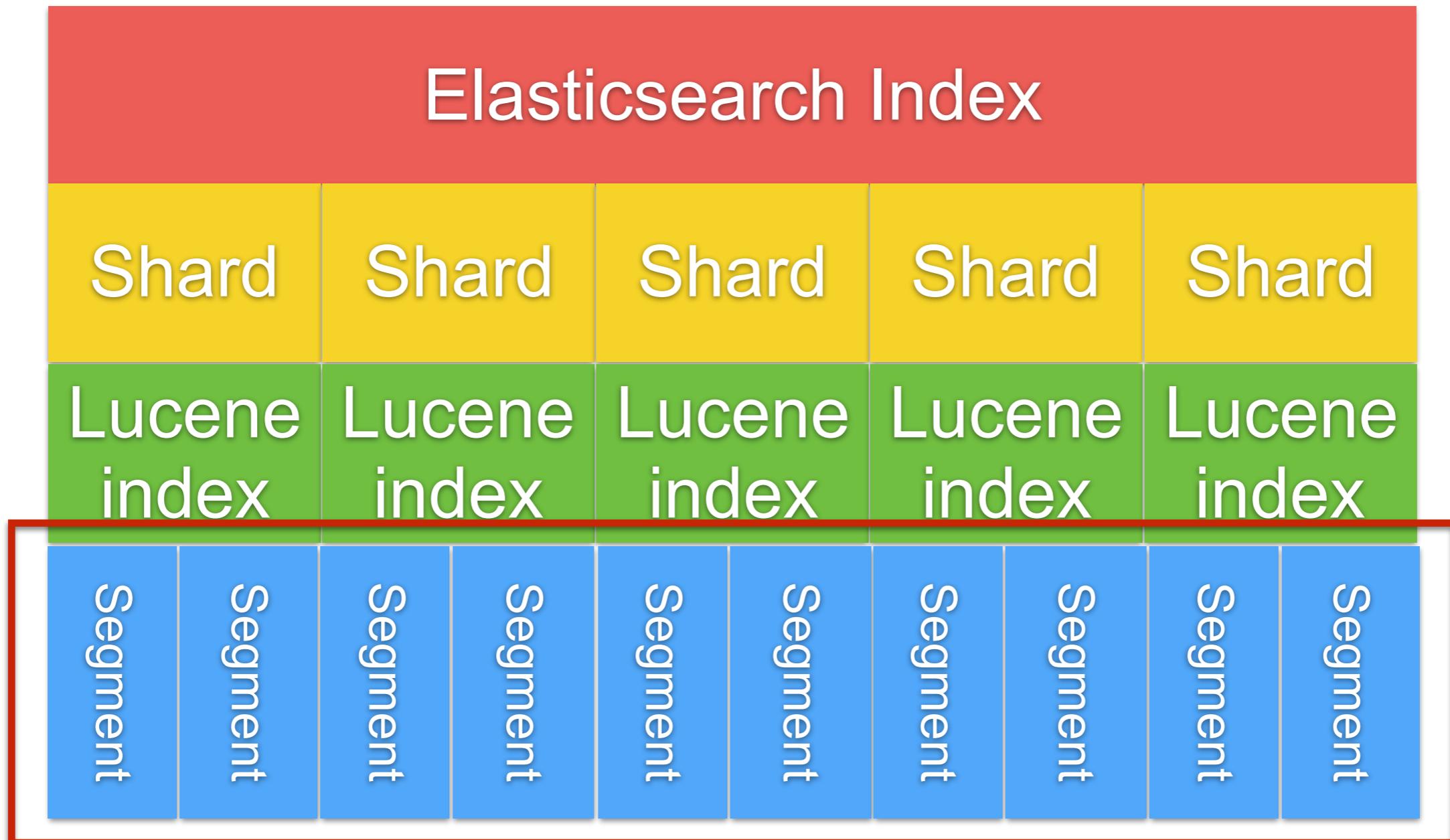
Apache Lucene



Max # of document of Lucene index = 2,147,483,519



Apache Lucene



Segments are immutable

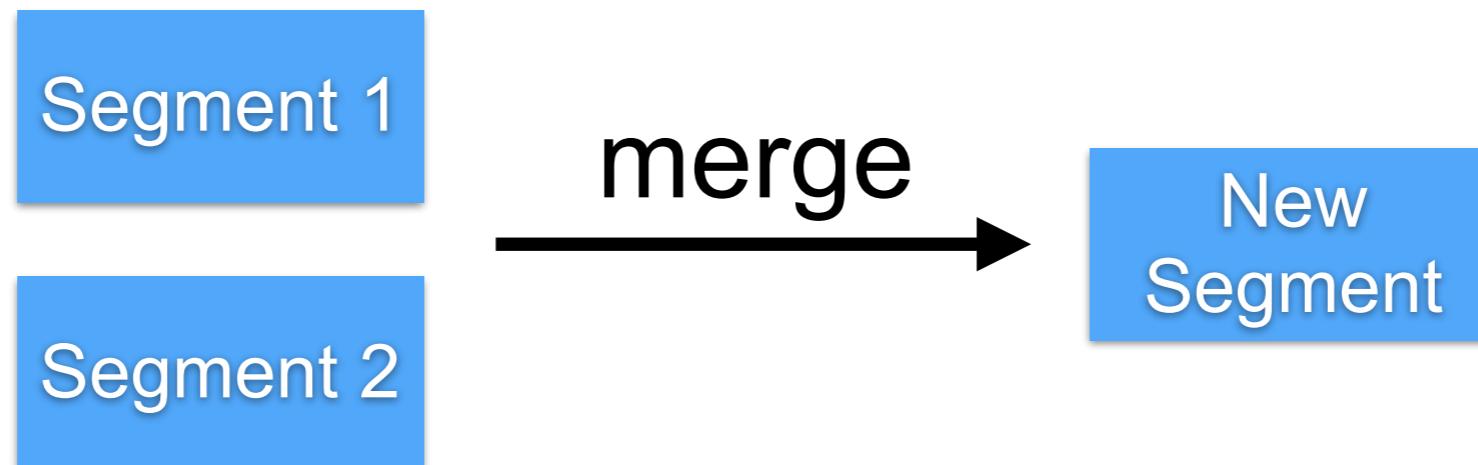


Segment

More shards, more segments

Documents are never delete !!

Lucene segment **merge** use more CPU/IO
Segments are immutable



Hardware



Hardware

CPU
Memory
Network
Storage



Memory

Enable bootstrap.memorylock

Disable all swap files

Change `ES_HEAP_SIZE` (default 1G)

<https://www.elastic.co/guide/en/elasticsearch/reference/current/setup-configuration-memory.html>



Design your index



Design your index

Sharding
Replication



Sharding

Elasticsearch divides the data in **logical** parts
of sharding is define when index created



How many shard ?



Need to know your size of data

Data Size	# of shard
< 3M	1
>3M <5M	2
>5M	(# of document / 5M) + 1



Sharding

Small shards on multiple nodes make the cluster recovery faster

Small shards on a lot of nodes solve memory mgt problem when query on large data



More shard, more Segment !!

Elasticsearch Index				
Shard	Shard	Shard	Shard	Shard
Lucene index	Lucene index	Lucene index	Lucene index	Lucene index
Segment	Segment	Segment	Segment	Segment

Need to config file descriptor

<https://www.elastic.co/guide/en/elasticsearch/reference/current/system-config.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Don't create more shard than you need !!



Replication

Prevent data loss

Default = 1

```
# nodes = [(primary + # replication) /2 ] + 1
```



Problems with scaling

CPU consumption
Load average
Request rate
Search latency



Slow log

```
PUT /myindex/_settings
```

```
{  
  "index.search.slowlog.threshold.query.warn: 1s",  
  "index.search.slowlog.threshold.query.info: 500ms",  
  "index.search.slowlog.threshold.query.debug: 1500ms",  
  "index.search.slowlog.threshold.query.trace: 300ms",  
  "index.search.slowlog.threshold.fetch.warn: 500ms",  
  "index.search.slowlog.threshold.fetch.info: 400ms",  
  "index.search.slowlog.threshold.fetch.debug: 300ms",  
  "index.search.slowlog.threshold.fetch.trace: 200ms"  
}
```

If can't optimize then add more resources or rewrite

<https://www.elastic.co/guide/en/elasticsearch/reference/current/index-modules-slowlog.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

Indexing Data



Indexing data

Must be define data schema for your need
Default mapping == more cost (Memory/Disk)
Default for data is “text” + “keyword”
Understand analyzer and tokenizer
Use auto generated IDs if possible



Indexing data

Prefer bulk indexing

Change refresh interval

Time based index for log data

<https://www.elastic.co/guide/en/elasticsearch/reference/current/indices-update-settings.html>



ELK Stack

© 2017 - 2018 Siam Chamnankit Company Limited. All rights reserved.

For Large data

Increase refresh interval
Decrease replica number

```
PUT /logstash-2015.05.20/_settings
{
  "index" : {
    "refresh_interval" : "-1",
    "number_of_replicas" : 0
  }
}
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/indices-update-settings.html>



Query Data



Query data

Use filters as much as possible

Use scan and scroll for dumping large data

Node query cache

Shard query cache

Retrieve only necessary fields

<https://www.elastic.co/guide/en/elasticsearch/reference/current/query-cache.html>



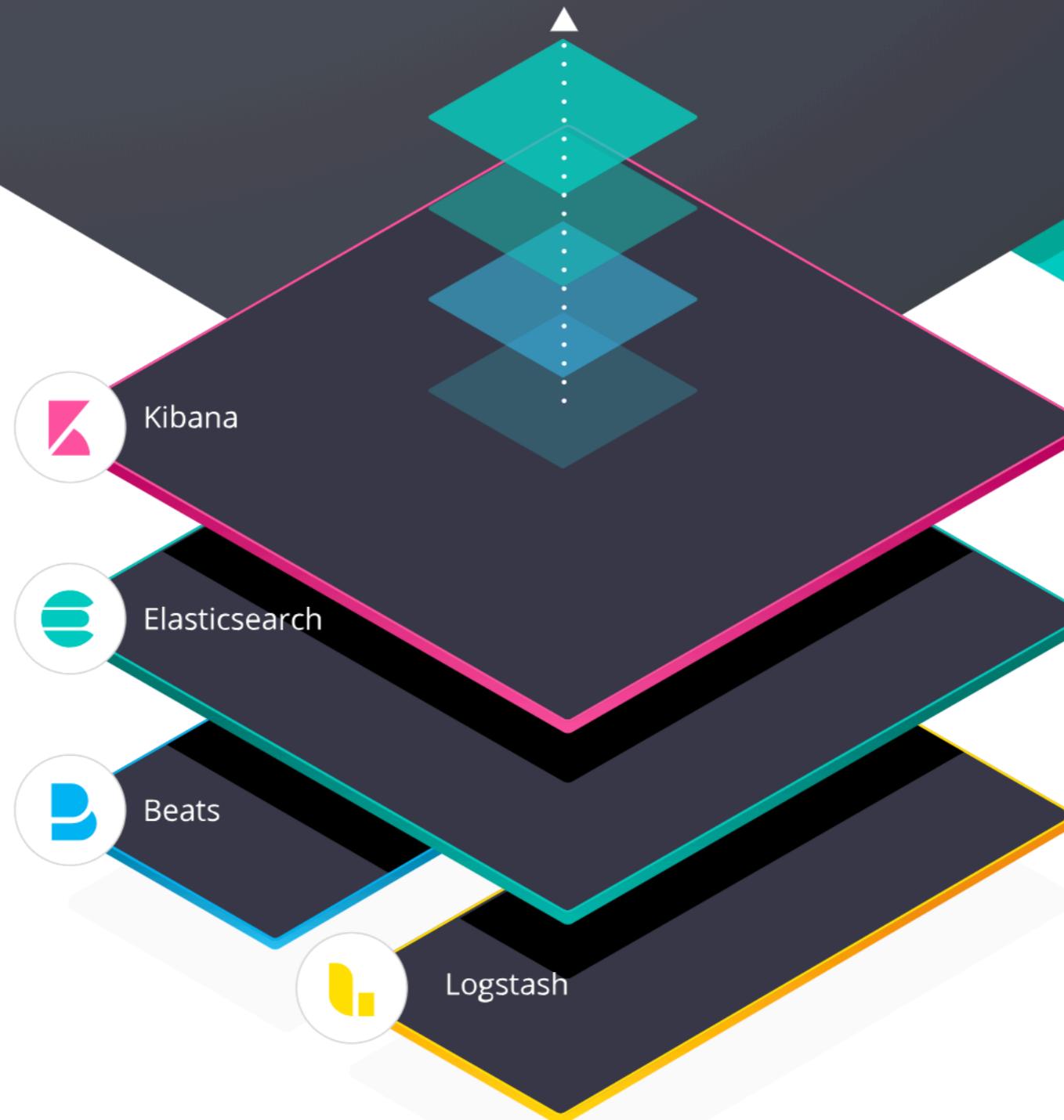
ELK Stack

© 2017 - 2018 Siam Chamnkit Company Limited. All rights reserved.

Use cases



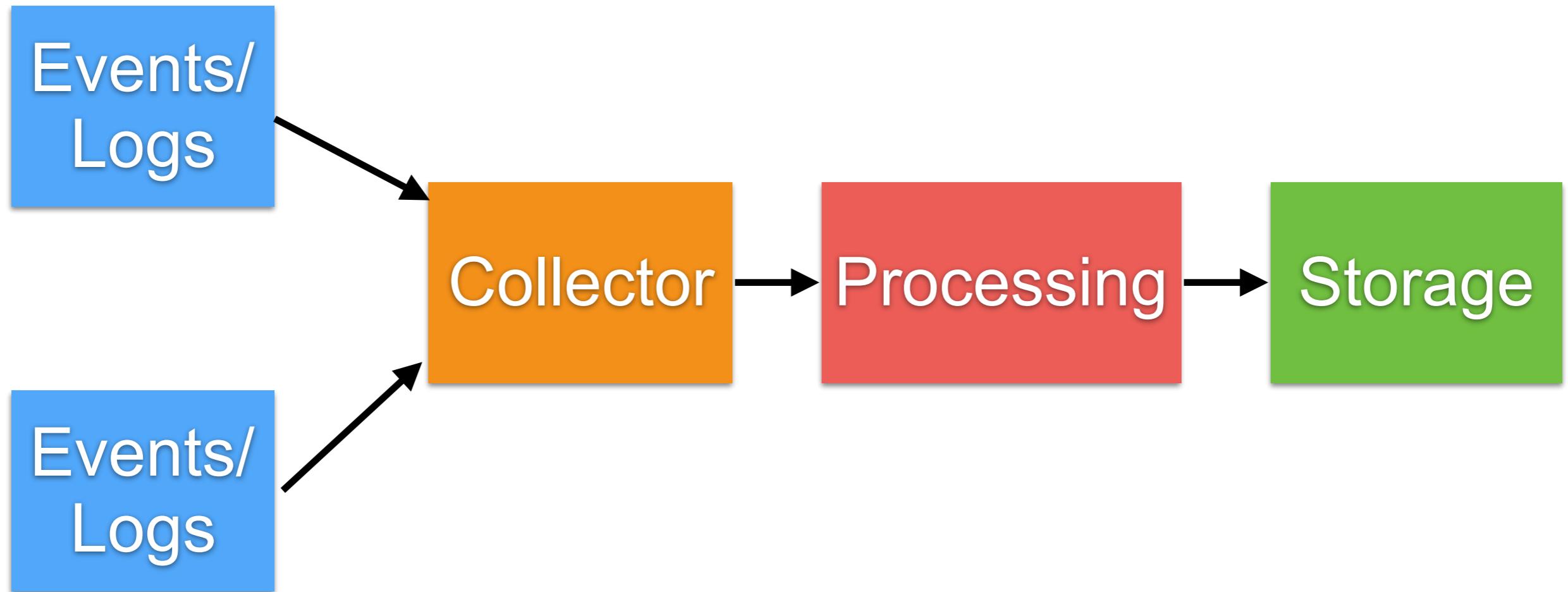
ELK stack



<https://www.elastic.co/elk-stack>



Event or Logging from Servers



Event or Logging from Servers



Data
Collection

Data
Aggregation
& Processing

Indexing &
storage

Analysis &
visualization



Event or Logging from Servers

