

# ELK Stack

Elasticsearch

Logstash

Beats

Kibana







Somkiat Puisungnoen



Somkiat

Home



Somkiat Puisungnoen

Update Info **1**

View Activity Log **10+**



Timeline

About

Friends 3,138

Photos

More ▾



When did you work at Opendream?



22 Pending Items



Intro

Software Craftsmanship



Software Practitioner at สยามชำนาญกิจ 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 · 🌐 ▾

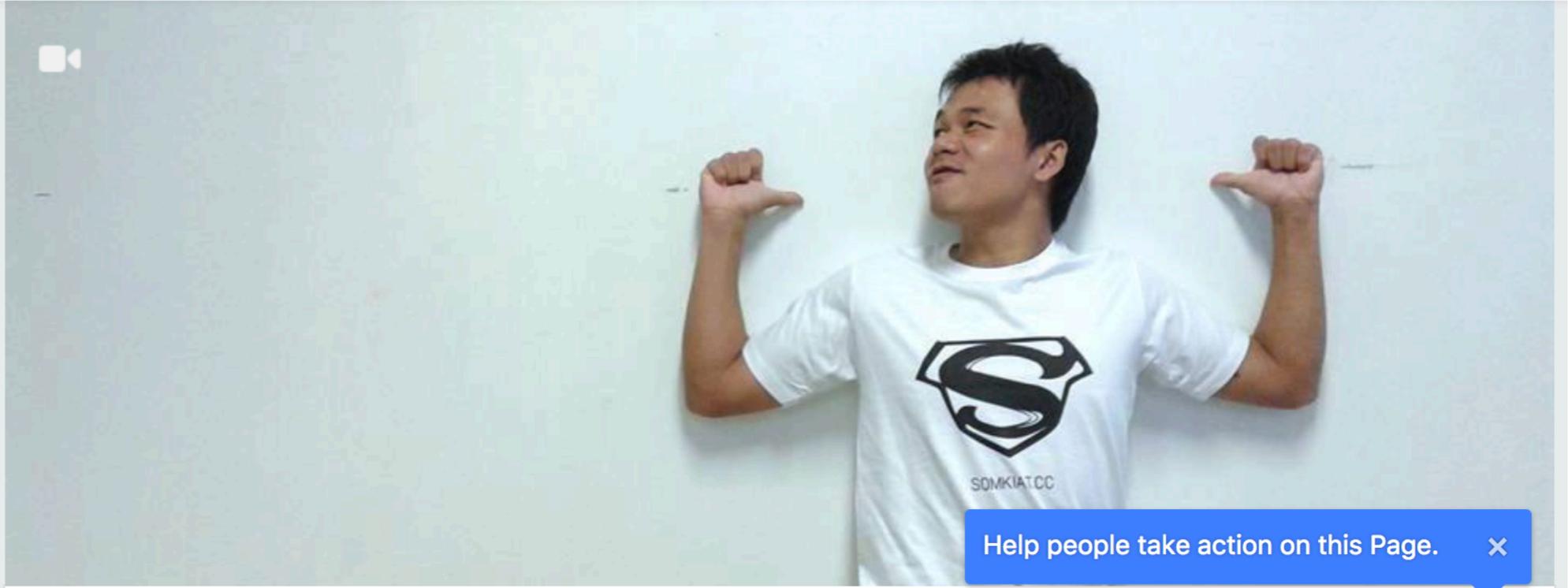
Java and Bigdata





somkiat.cc  
@somkiat.cc

- Home
- Posts
- Videos
- Photos



Help people take action on this Page. x

Liked Following Share

+ Add a Button



**[https://github.com/up1/course\\_elk](https://github.com/up1/course_elk)**



# Agenda

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



# Agenda

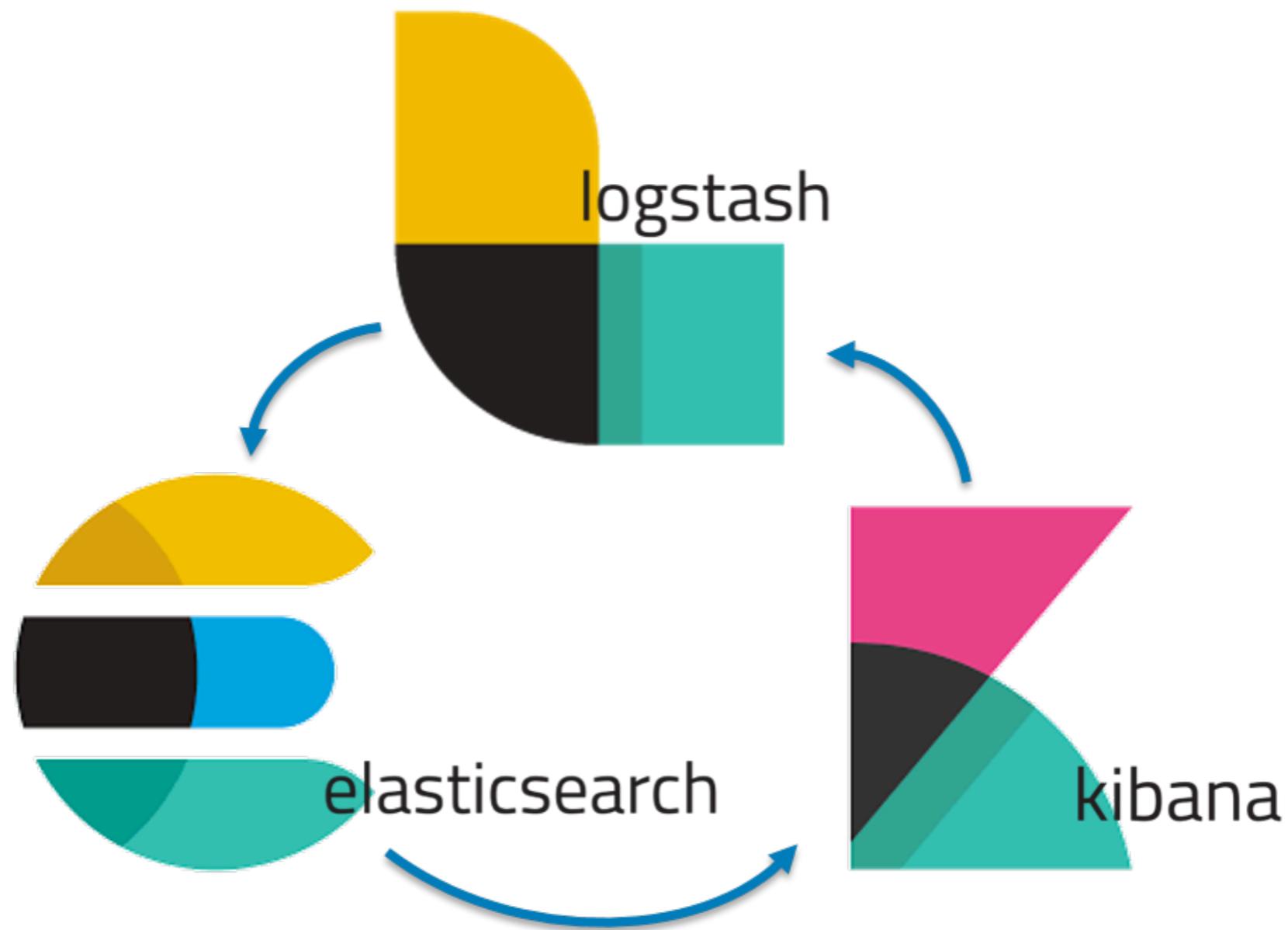
- Working with Kibana
- Useful features
  - Auto-suggestion
  - ngram algorithm
- Clustering management
- Design for scaling
- Working with Logstash
- Machine Learning with ELK



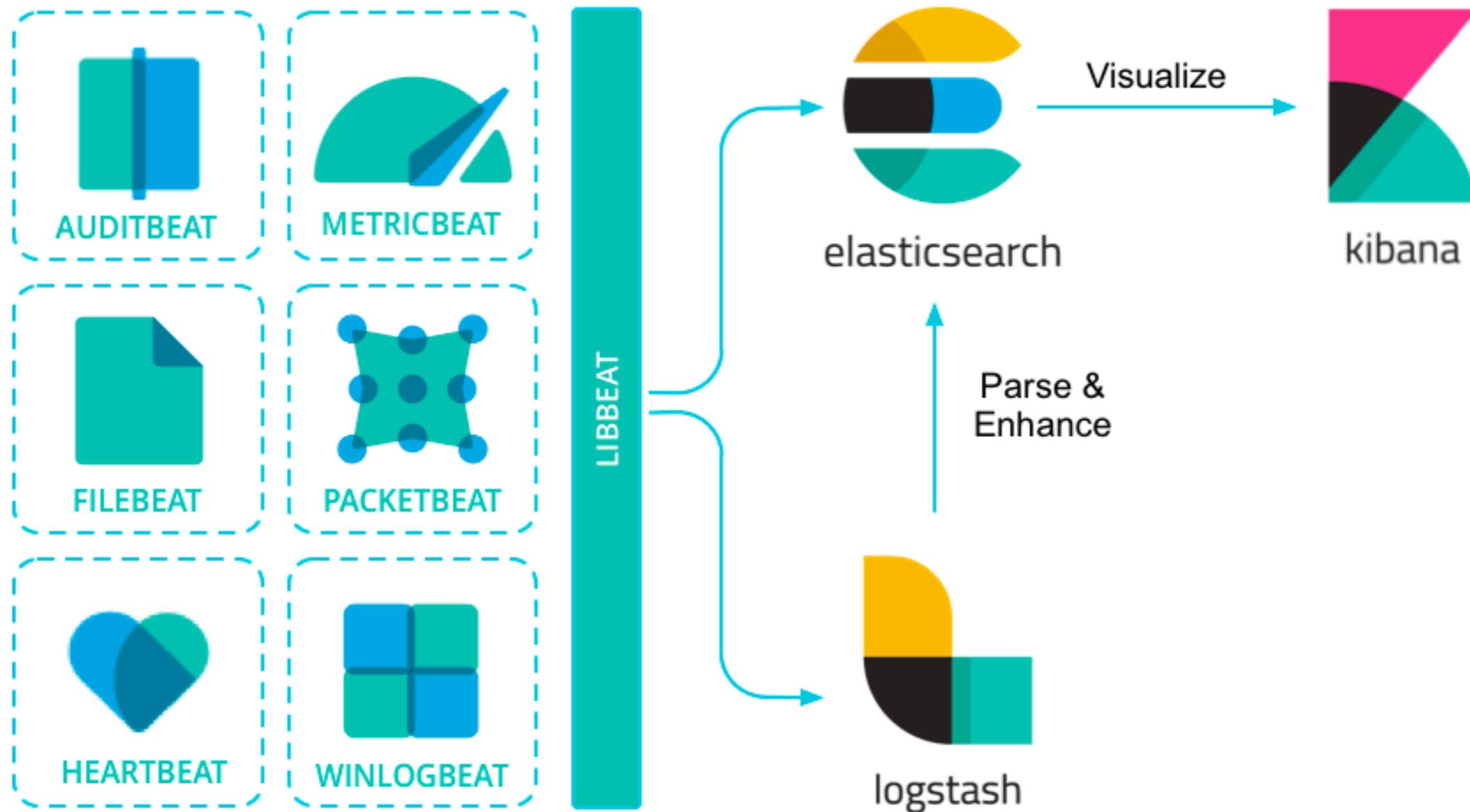
# Agenda

- Working with Prometheus
- Working with Grafana
- Workshop





# Beat



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



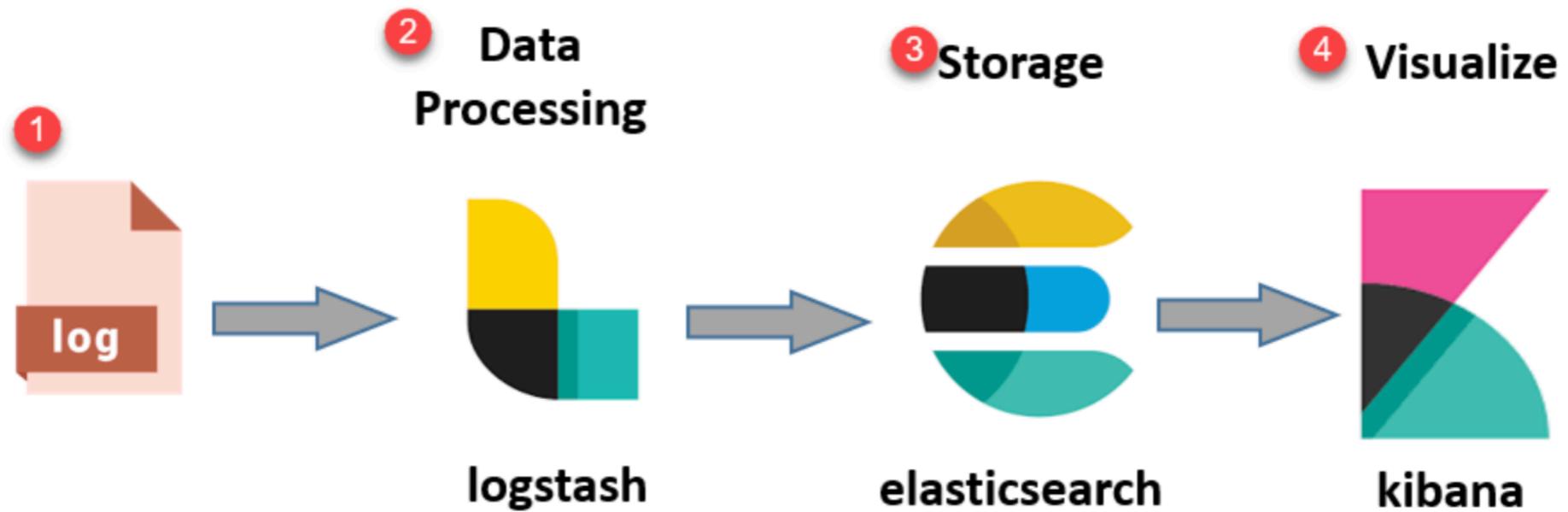
# Beat

Purpose	Library
Audit data	Auditbeat
Log files	Filebeat
Cloud data	Functionbeat
Availability	Heartbeat
Metrics	Metricbeat
Network traffic	Packetbeat
Windows event logs	Winlogbeat

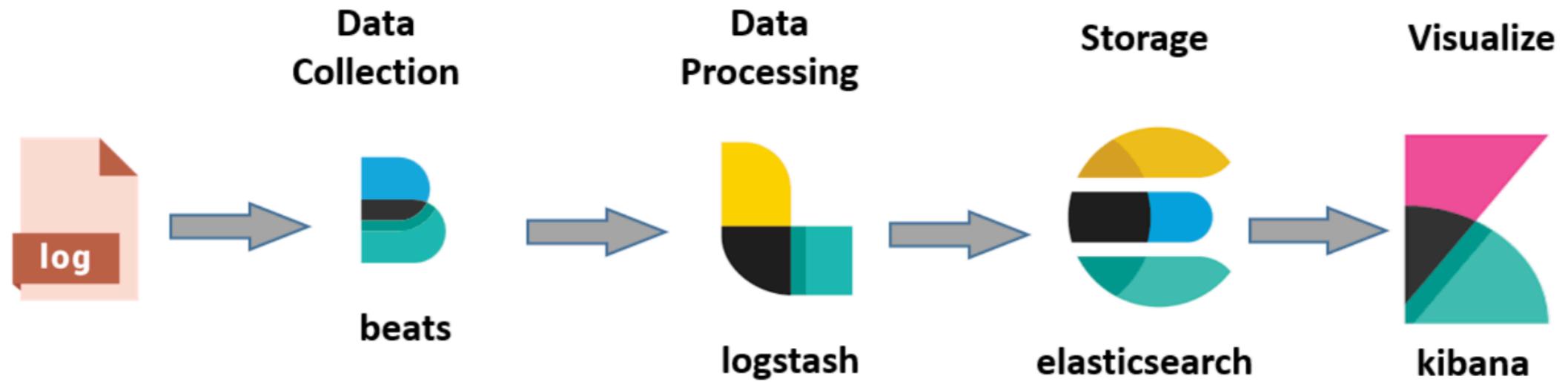
<https://www.elastic.co/guide/en/beats/libbeat/current/beats-reference.html>



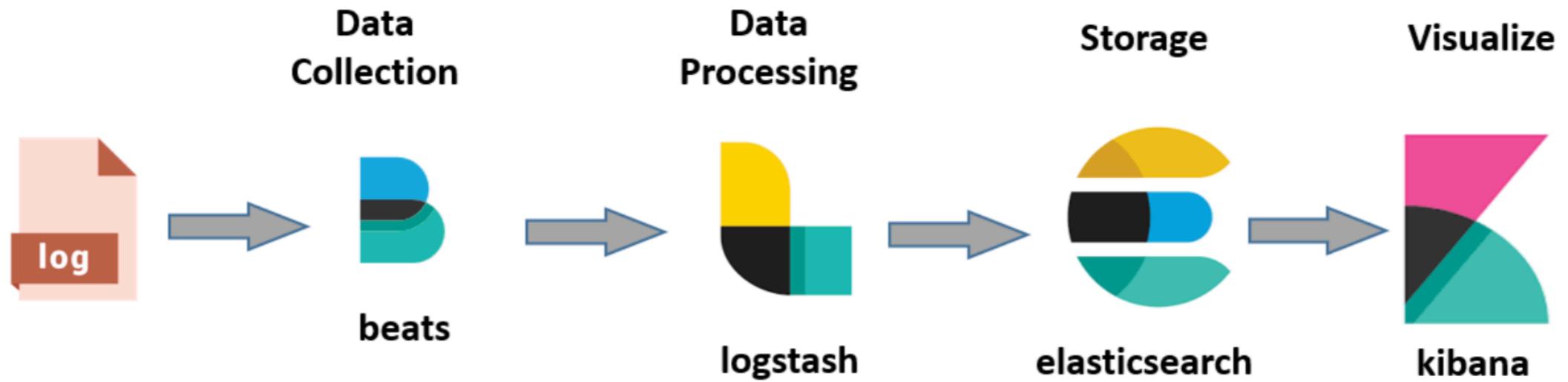
# ELK



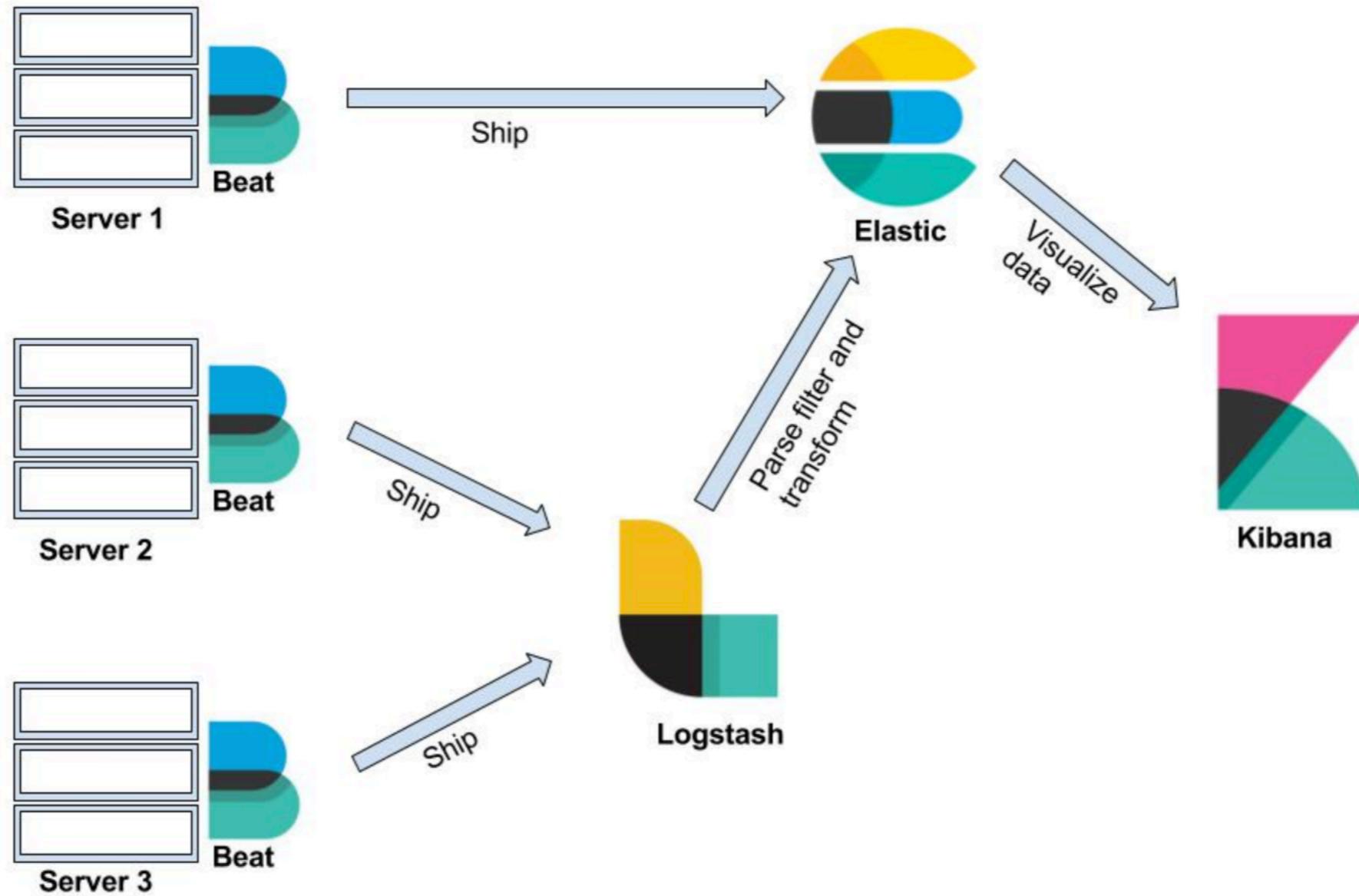
# ELK + Beats



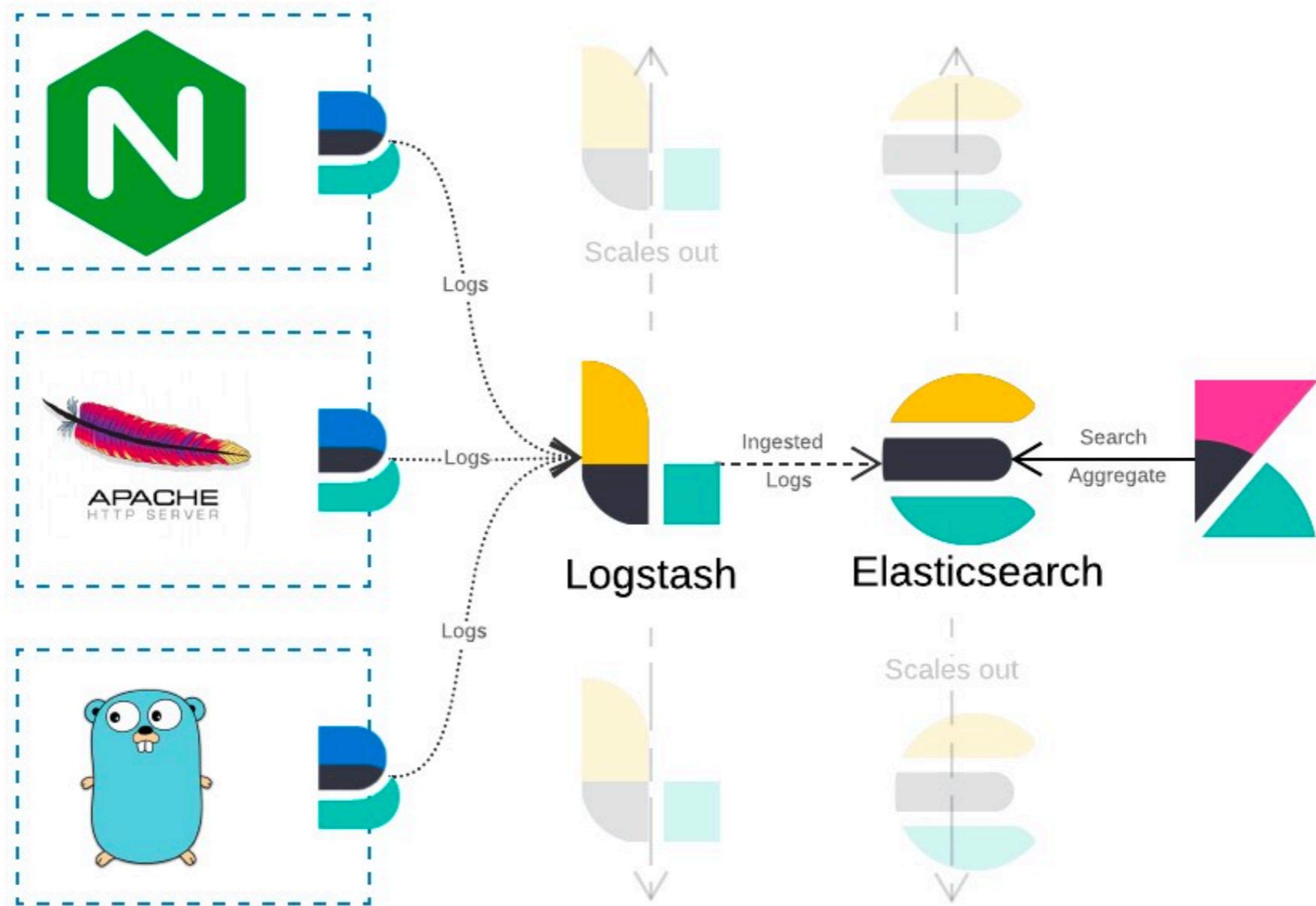
# ELK + Beats



# ELK + Beats



# ELK + Beats



# Elasticsearch

<https://www.elastic.co/elasticsearch/>



# Elasticsearch

Search  
Analytic  
Real-time  
Distributed  
Scalability



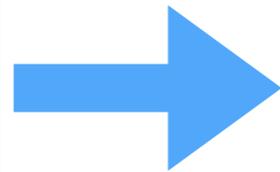
# Distributed Search Engine

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

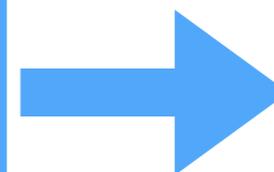


# Distributed Search Engine

1999



2004



2010



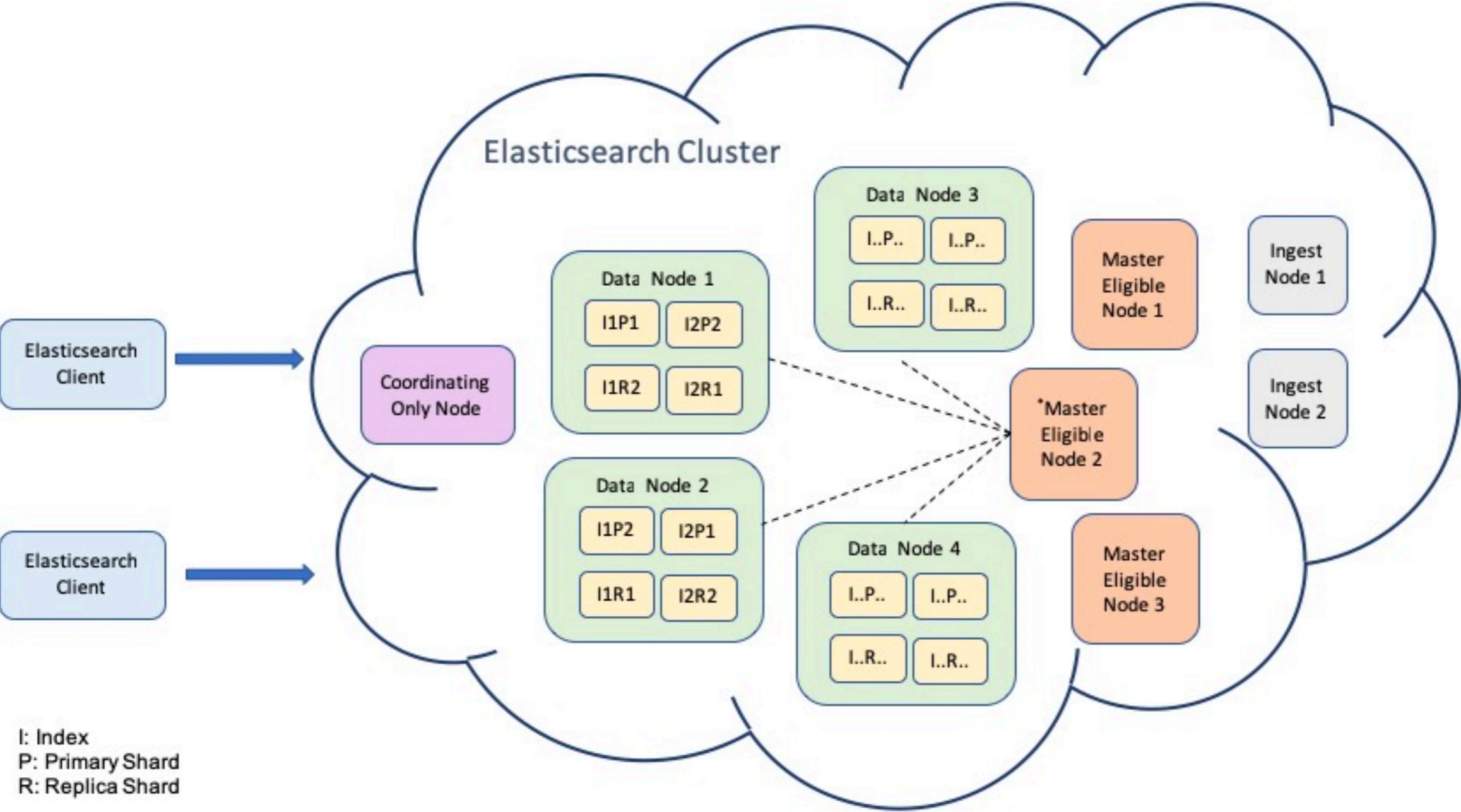
Compass



elasticsearch



# Elasticsearch Cluster



# Apache Lucene

The screenshot shows the Apache Lucene website homepage. At the top left is the Lucene logo, a stylized feather with the word 'Lucene' written in a cursive font. To the right is a search bar with the text 'Search with Apache So' and a dropdown menu labeled '@ select provider'. Below the search bar are three tabs: 'CORE (JAVA)', 'SOLR', and 'PYLUCENE'. The main banner features the text 'Ultra-fast Search Library and Server' on the left, navigation arrows in the center, and the Lucene and Solr logos on the right. Below the banner is a dark blue bar with the text 'Apache Lucene and Solr set the standard for search and indexing performance'. The main content area has a heading 'Welcome to Apache Lucene' followed by a paragraph: 'The Apache Lucene™ project develops open-source search software, including:'. Below this are three bullet points: '• 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.', and '• PyLucene is a Python port of the Core project.' To the right of the text are two 'DOWNLOAD' buttons. The top one is green and labeled 'Apache Lucene 7.5.0'. The bottom one is orange and labeled 'Apache Solr 7.5.0'. At the bottom right of the content area is a 'Projects' link.

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



# Document based

JSON (JavaScript Object Notation)

**Dynamic Schema (Schema-less)**

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>



# Ranking from DB Engine (2018)

350 systems in ranking, June 2019

Rank			DBMS	Database Model	Score		
Jun 2019	May 2019	Jun 2018			Jun 2019	May 2019	Jun 2018
1.	1.	1.	Oracle +	Relational, Multi-model <i>i</i>	1299.21	+13.67	-12.04
2.	2.	2.	MySQL +	Relational, Multi-model <i>i</i>	1223.63	+4.67	-10.06
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model <i>i</i>	1087.76	+15.57	+0.03
4.	4.	4.	PostgreSQL +	Relational, Multi-model <i>i</i>	476.62	-2.27	+65.95
5.	5.	5.	MongoDB +	Document	403.90	-4.17	+60.12
6.	6.	6.	IBM Db2 +	Relational, Multi-model <i>i</i>	172.20	-2.24	-13.44
7.	7.	↑ 8.	Elasticsearch +	Search engine, Multi-model <i>i</i>	148.82	+0.20	+17.78
8.	8.	↓ 7.	Redis +	Key-value, Multi-model <i>i</i>	146.13	-2.28	+9.83
9.	9.	9.	Microsoft Access	Relational	141.01	-2.77	+10.02
10.	10.	10.	Cassandra +	Wide column	125.18	-0.54	+5.97

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



# Use cases

Security/log analytics

Marketing

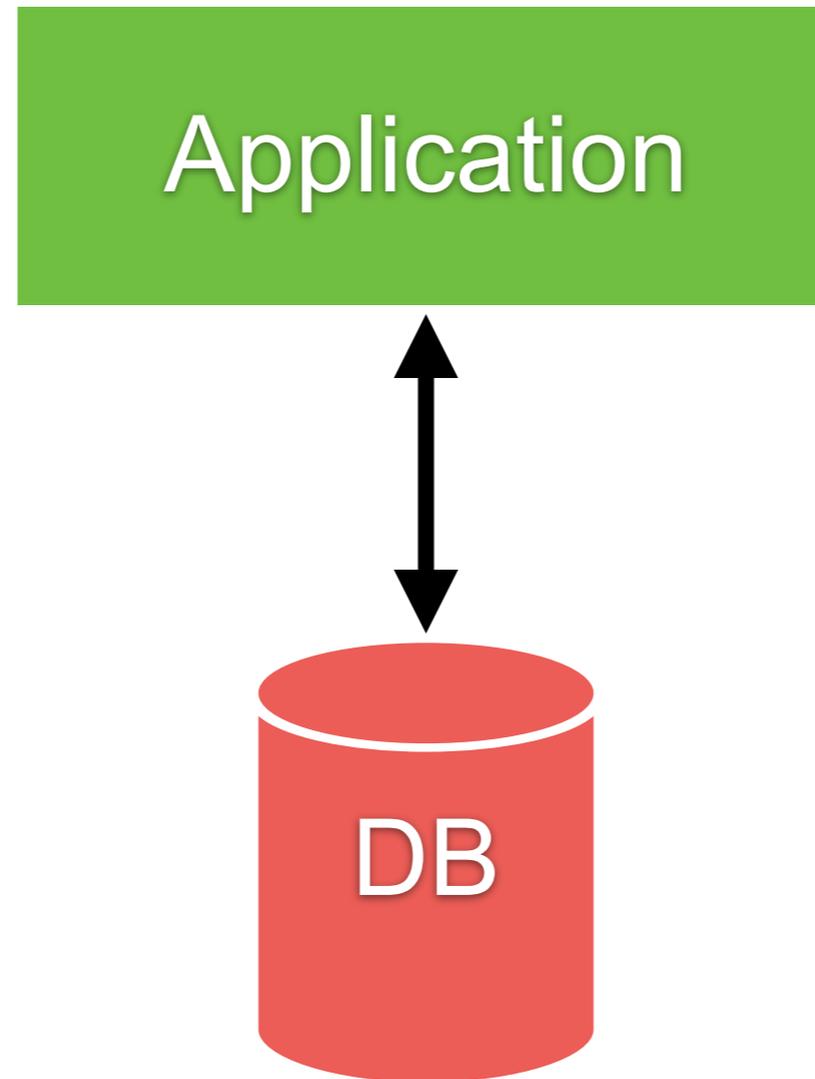
Operations

Searching data



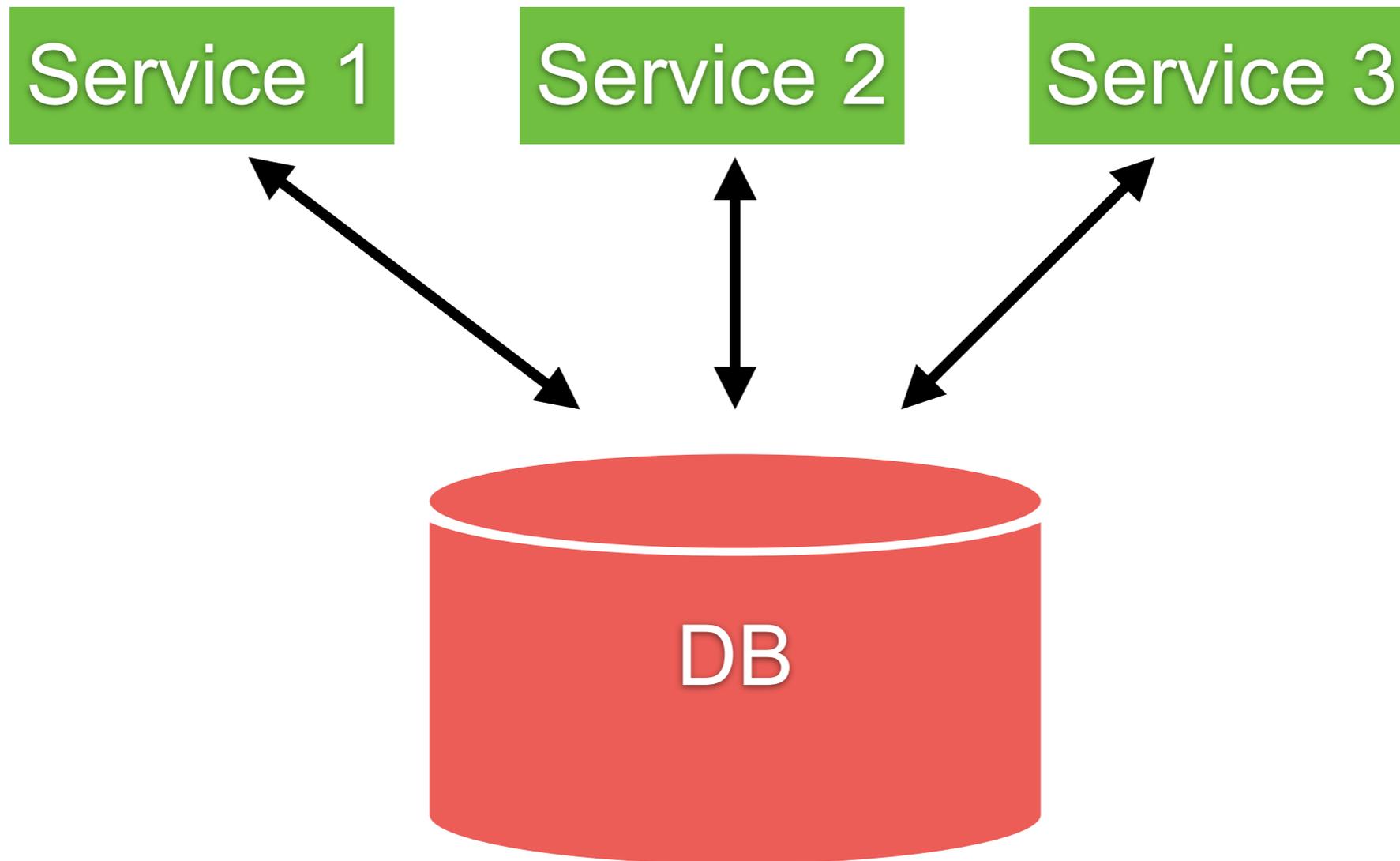
# Problem ?

Single/Centralize database



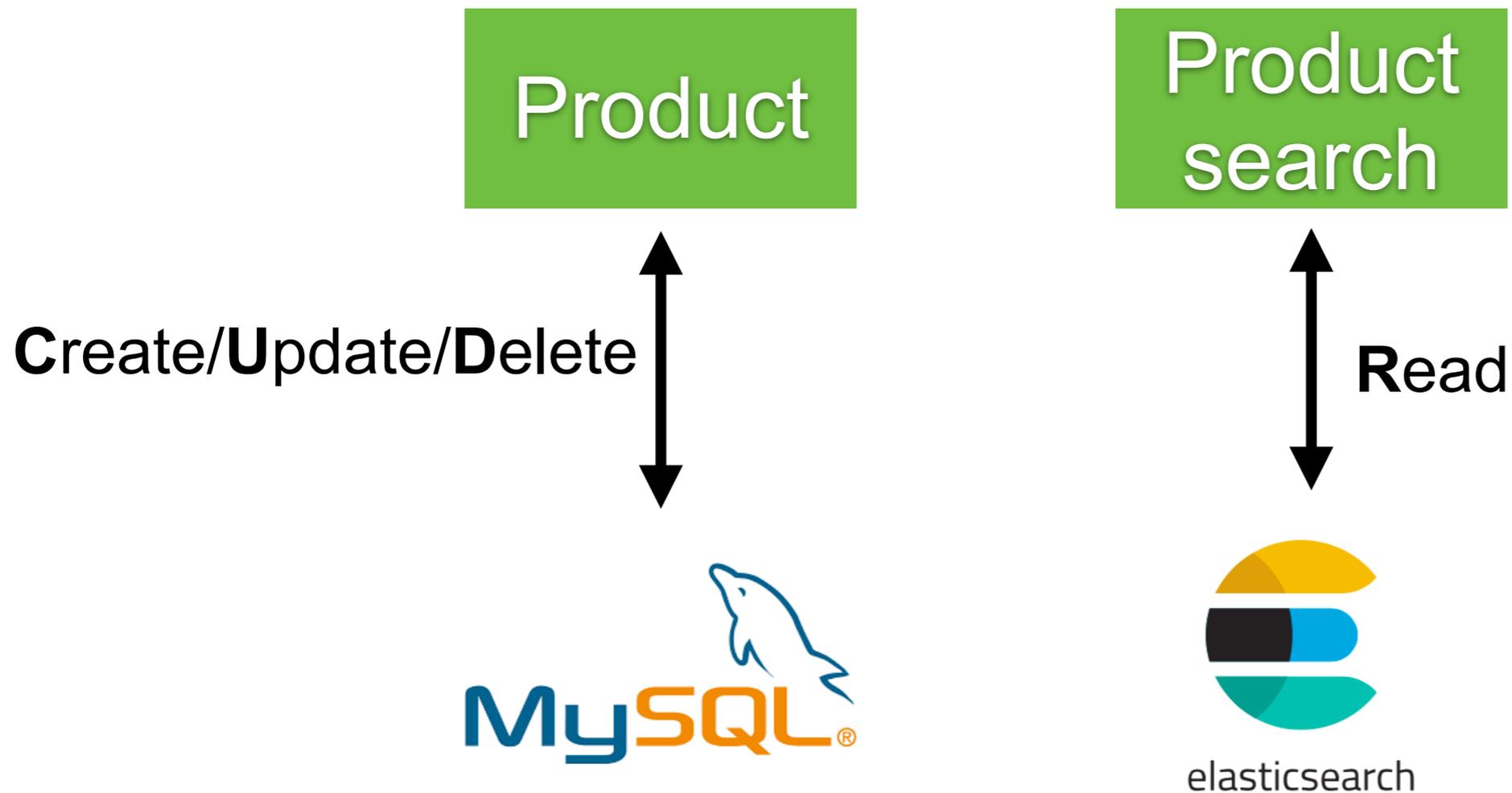
# Problem ?

Single/Centralize database



# Separate data for read and write

For example MySQL to write, Elasticsearch to search



# Let's start



# Installation

Elasticsearch  
Kibana



# Install Elasticsearch



# Elasticsearch

Required Java 8  
JDK and Open JDK  
Need \$JAVA\_HOME



# JAVA\_HOME

`$echo %JAVA_HOME% //For Windows`

`$echo $JAVA_HOME // for Linux/Mac`



# 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},
```



# Configuration files

elasticsearch.yml

jvm.options

log4j2.properties

Default : `$ES_HOME/config`

Custom config path : `$ES_PATH_CONF`





# 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>



# X-Pack ?

Elastic Stack Extension  
Security  
Monitoring  
Alerting  
Reporting  
Machine Learning



# Licence

	FREE		GOLD	PLATINUM
	OPEN SOURCE	BASIC		
	<a href="#">Download</a>		<a href="#">Request Info</a>	<a href="#">Request Info</a>
<b>ELASTIC STACK</b>				
Elasticsearch				
✓ Scalability & Resiliency	✓	✓	✓	✓
✓ Query & Analytics	✓	✓	✓	✓
✓ Data Enrichment	✓	✓	✓	✓
✓ Management & Tooling	✓	✓	✓	✓
✓ Security			✓	✓
✓ Alerting			✓	✓
✓ Machine Learning				✓
Kibana				
✓ Explore & Visualize	✓	✓	✓	✓
✓ Stack Management & Tooling	✓	✓	✓	✓
✓ Stack Monitoring		✓	✓	✓

<https://www.elastic.co/subscriptions>



# Hello Elasticsearch

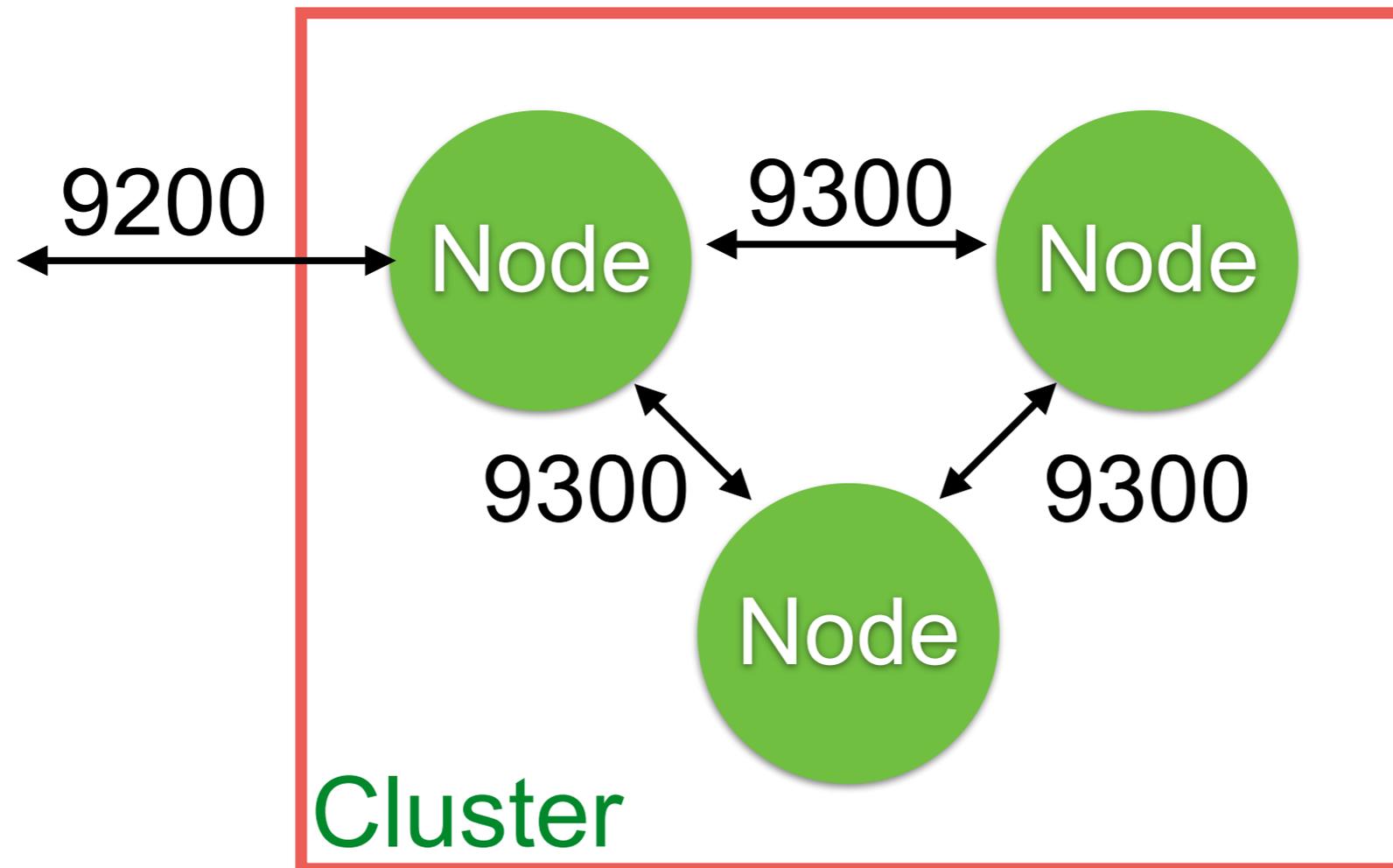
<http://localhost:9200/>

```
{  
  "name": "Somkiats-MacBook-Pro",  
  "cluster_name": "elasticsearch",  
  "cluster_uuid": "AmWXLi6DRFOWuZbZEi9FCw",  
  "version": {  
    "number": "7.14.0",  
    "build_flavor": "default",  
    "build_type": "tar",  
    "build_hash": "dd5a0a2acaa2045ff9624f3729fc8a6f40835aa1",  
    "build_date": "2021-07-29T20:49:32.864135063Z",  
    "build_snapshot": false,  
    "lucene_version": "8.9.0",  
    "minimum_wire_compatibility_version": "6.8.0",  
    "minimum_index_compatibility_version": "6.0.0-beta1"  
  },  
  "tagline": "You Know, for Search"  
}
```



# Ports of Elasticsearch

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



## Name of node and cluster

```
{
  name: "Somkiats-MacBook-Pro",
  cluster_name: "elasticsearch",
  cluster_uuid: "gglAIg0NRHyn4AefFR61aw",
- version: {
  number: "7.1.1",
  build_flavor: "default",
  build_type: "tar",
  build_hash: "7a013de",
  build_date: "2019-05-23T14:04:00.380842Z",
  build_snapshot: false,
  lucene_version: "8.0.0",
  minimum_wire_compatibility_version: "6.8.0",
  minimum_index_compatibility_version: "6.0.0-beta1"
},
  tagline: "You Know, for Search"
}
```



# 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>



# Try to change and restart !!!



```
{
  name: "Somkiats-MacBook-Pro",
  cluster_name: "elasticsearch",
  cluster_uuid: "gglAIg0NRHyn4AefFR61aw",
- version: {
    number: "7.1.1",
    build_flavor: "default",
    build_type: "tar",
    build_hash: "7a013de",
    build_date: "2019-05-23T14:04:00.380842Z",
    build_snapshot: false,
    lucene_version: "8.0.0",
    minimum_wire_compatibility_version: "6.8.0",
    minimum_index_compatibility_version: "6.0.0-beta1"
  },
  tagline: "You Know, for Search"
}
```



# Compatibility of DSL and Index



```
{
  name: "Somkiats-MacBook-Pro",
  cluster_name: "elasticsearch",
  cluster_uuid: "gglAIg0NRHyn4AefFR61aw",
- version: {
  number: "7.1.1",
  build_flavor: "default",
  build_type: "tar",
  build_hash: "7a013de",
  build_date: "2019-05-23T14:04:00.380842Z",
  build_snapshot: false,
  lucene_version: "8.0.0",
  minimum_wire_compatibility_version: "6.8.0",
  minimum_index_compatibility_version: "6.0.0-beta1"
},
  tagline: "You Know, for Search"
}
```

DSL version

Index version



```
{
  name: "Somkiats-MacBook-Pro",
  cluster_name: "elasticsearch",
  cluster_uuid: "gglAIg0NRHyn4AefFR61aw",
- version: {
    number: "7.1.1",
    build_flavor: "default",
    build_type: "tar",
    build_hash: "7a013de",
    build_date: "2019-05-23T14:04:00.380842Z",
    build_snapshot: false,
    lucene_version: "8.0.0",
    minimum_wire_compatibility_version: "6.8.0",
    minimum_index_compatibility_version: "6.0.0-beta1"
  },
  tagline: "You Know, for Search"
}
```

Index version



# Health of cluster

[http://localhost:9200/\\_cluster/health](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](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>



# cat APIs

[http://localhost:9200/\\_cat/nodes?v](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

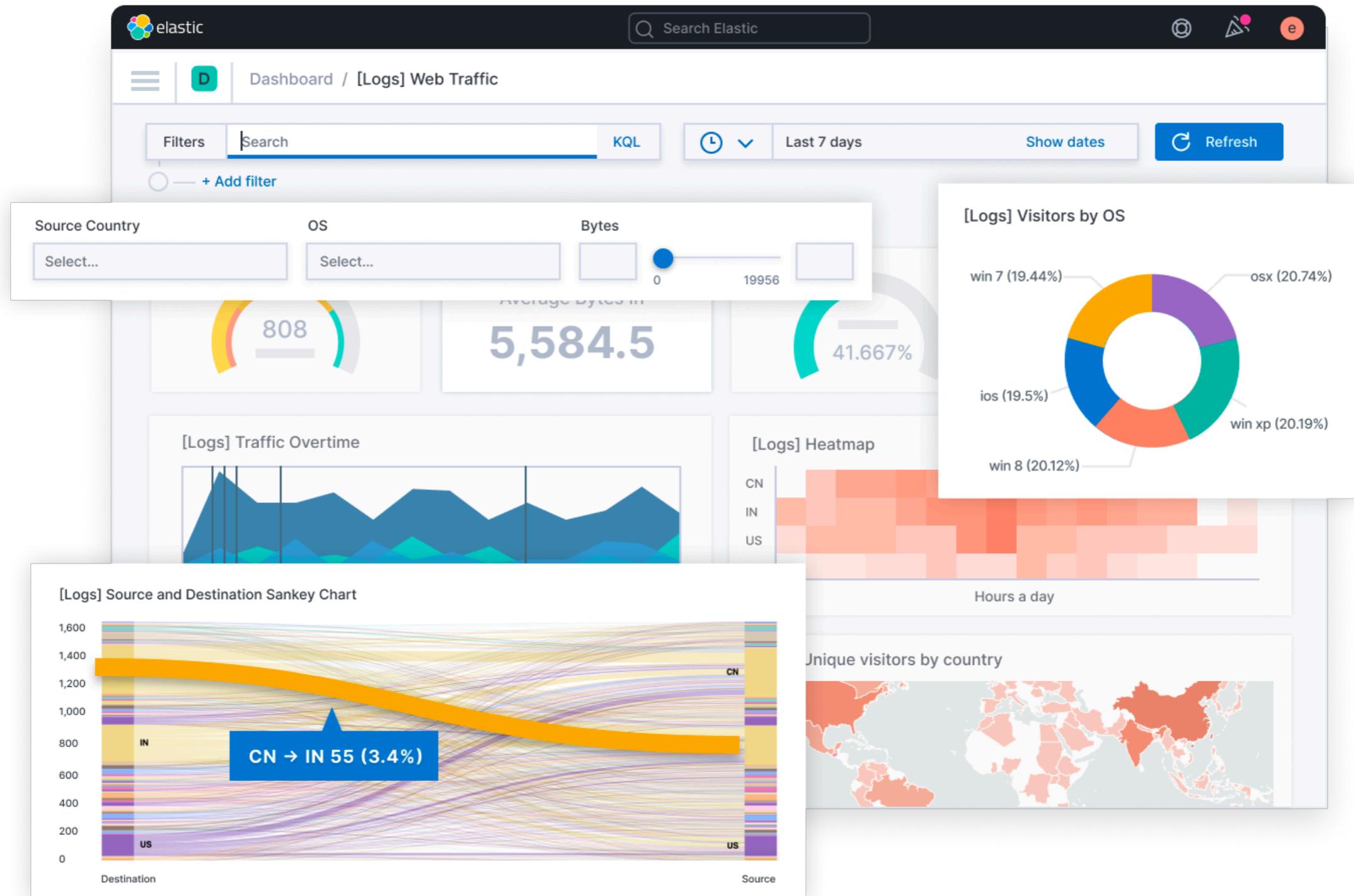


# Kibana

<https://www.elastic.co/kibana/>



# Kibana



# 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 - Rea

[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/

**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](#)

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

**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,



# Using Dev Tools

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 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 by types.

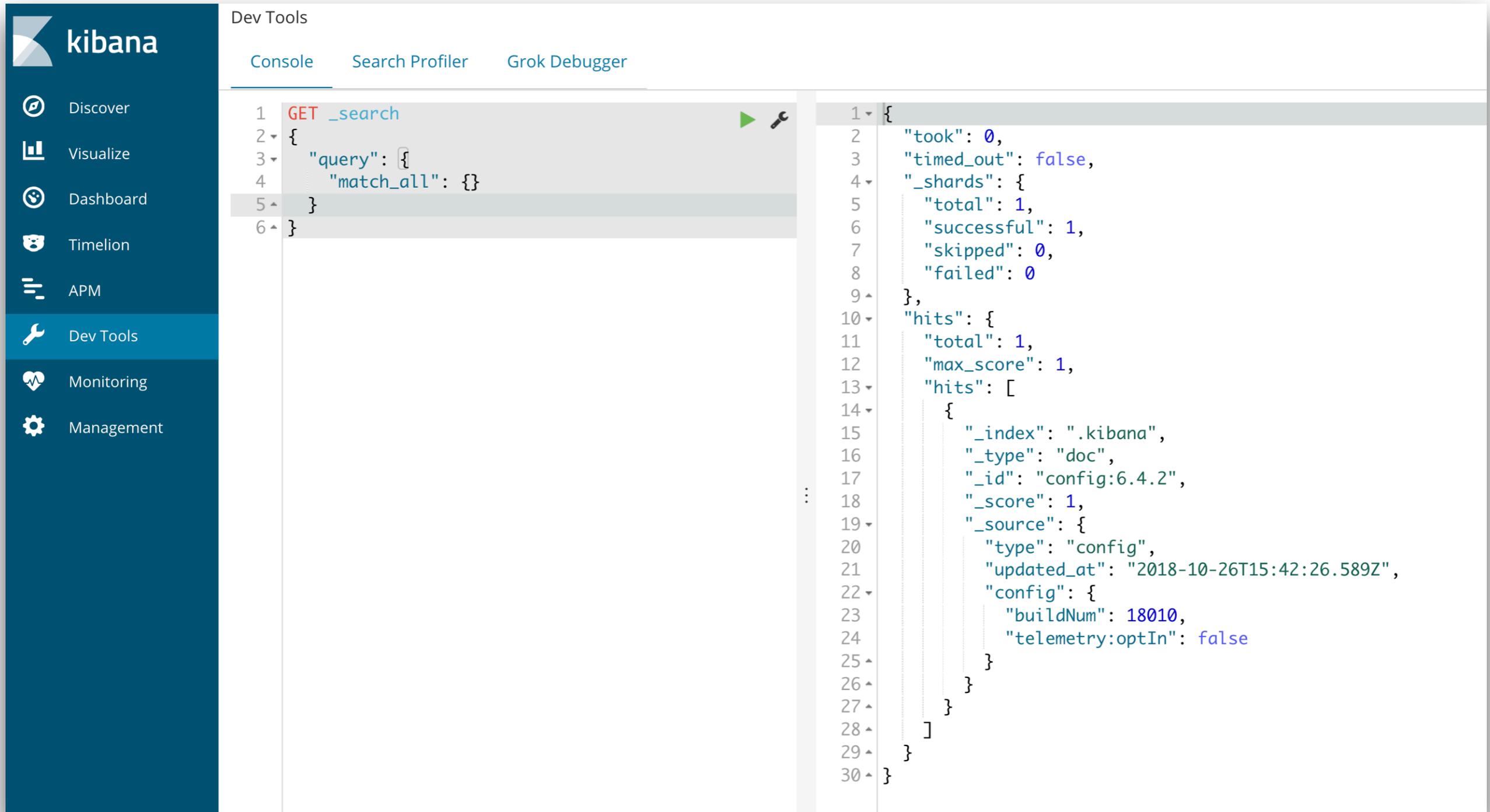
### A few quick tips, while I have your attention

- 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](#)



# Ready to start



The screenshot displays the Kibana Dev Tools interface. On the left is a dark blue sidebar with the Kibana logo and navigation links: Discover, Visualize, Dashboard, Timelion, APM, Dev Tools (highlighted), Monitoring, and Management. The main area is titled 'Dev Tools' and has three tabs: Console, Search Profiler, and Grok Debugger. The Console tab is active, showing a REST client interface with a request and its response.

**Request:**

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

**Response:**

```
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 }
```



# Elasticsearch architecture



# Basic concepts

Cluster

Node

Shard

Replica

Gateway

Index

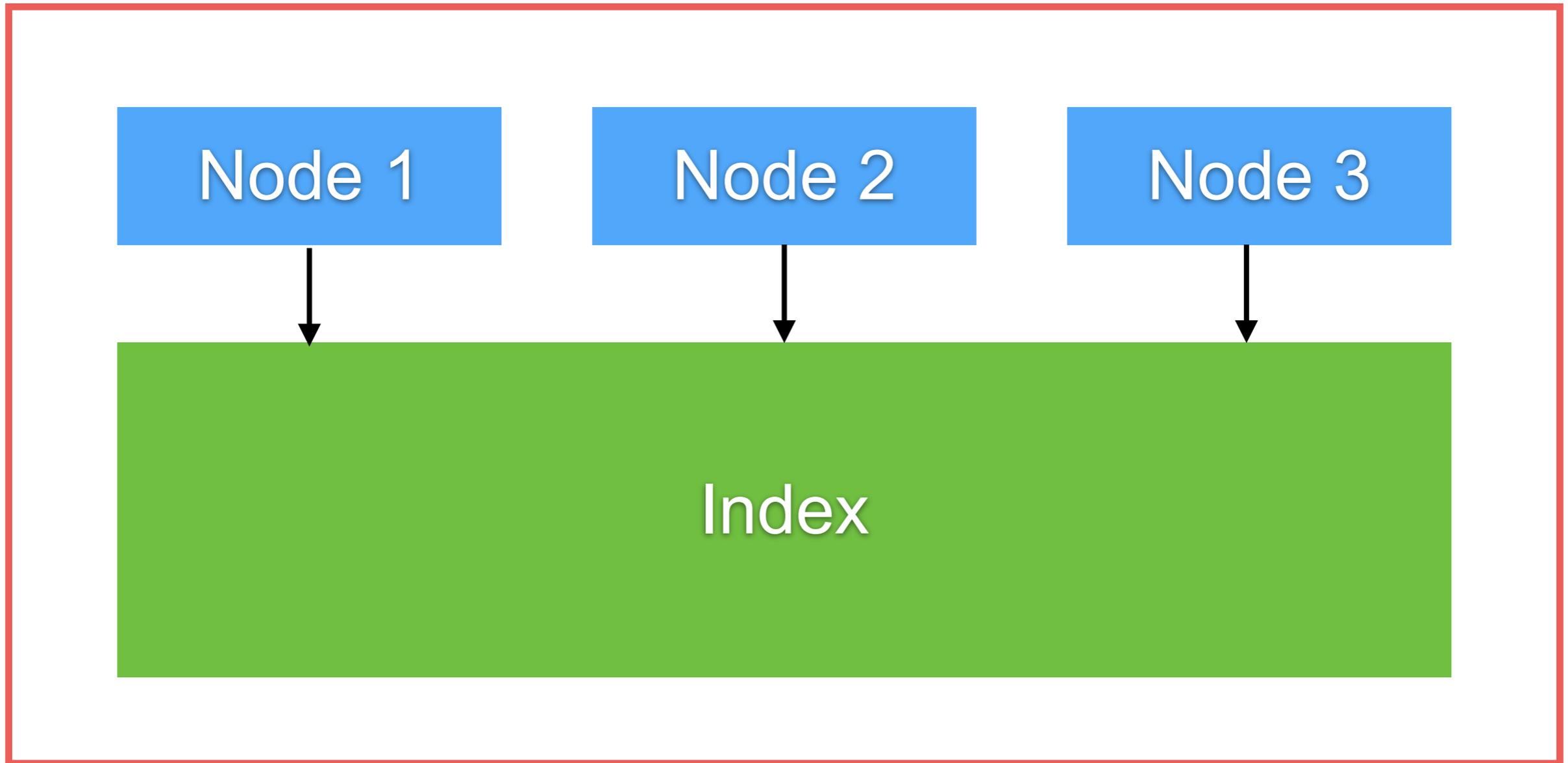
Document

Type

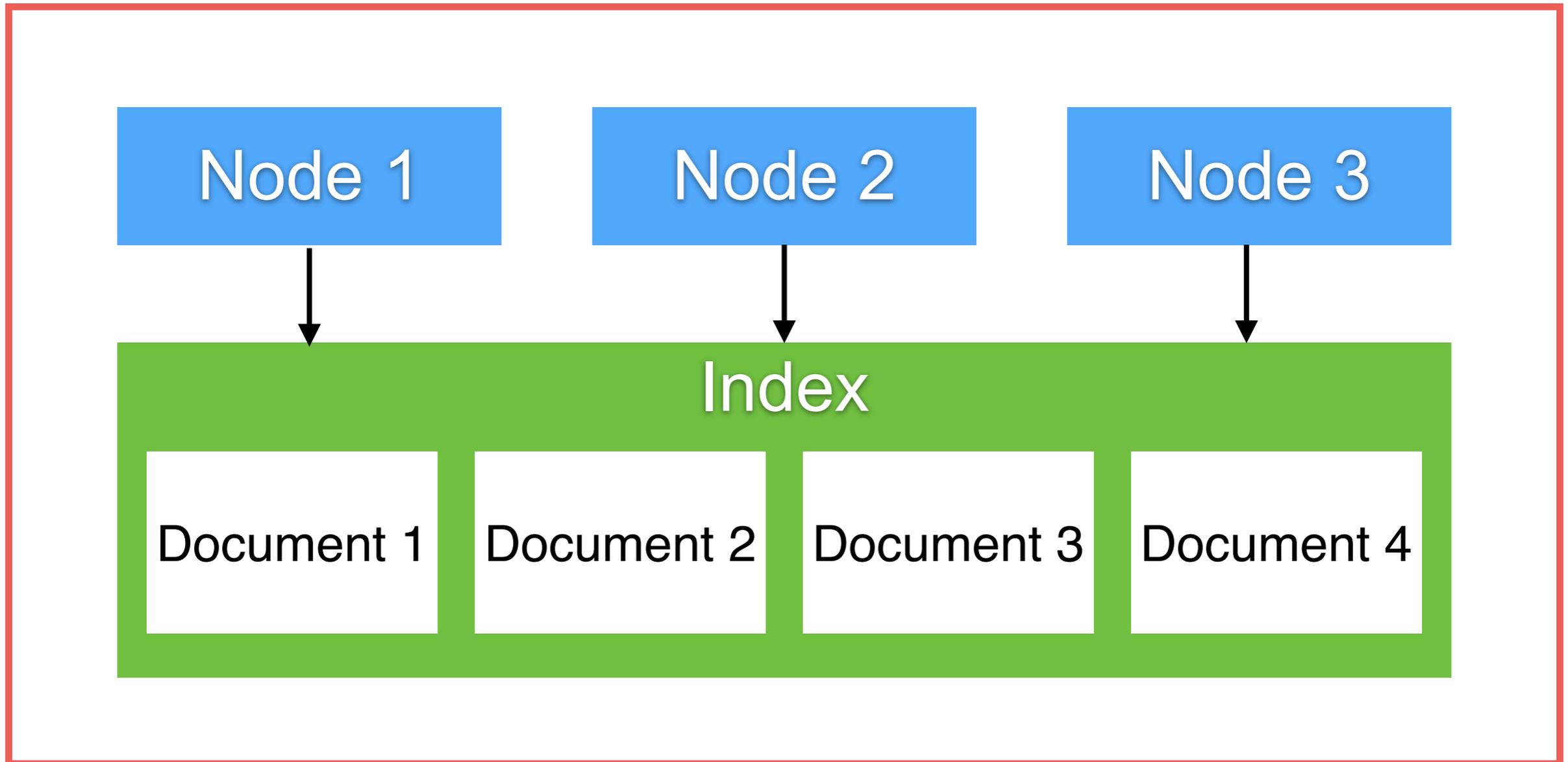
Mapping



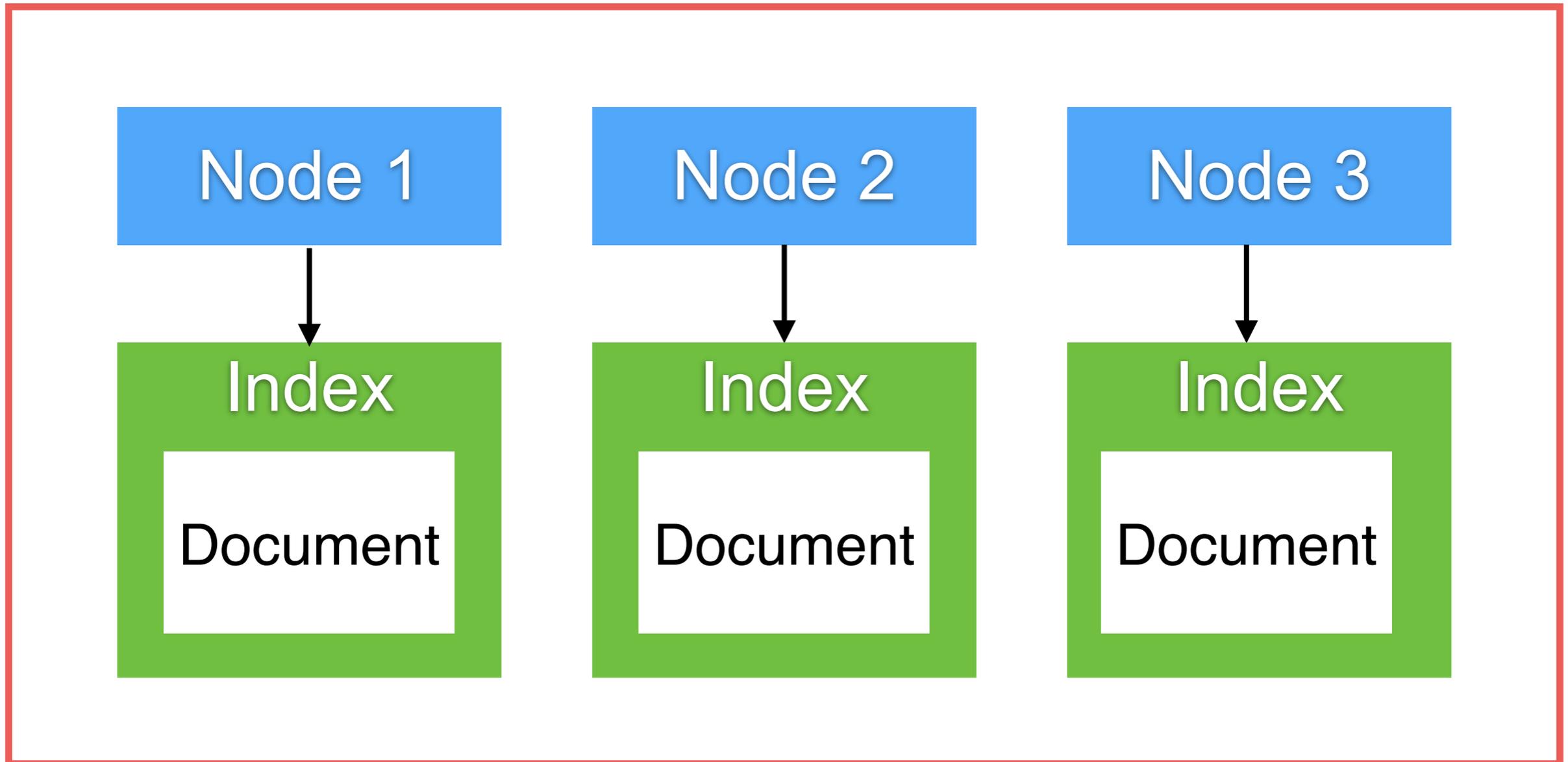
# Cluster



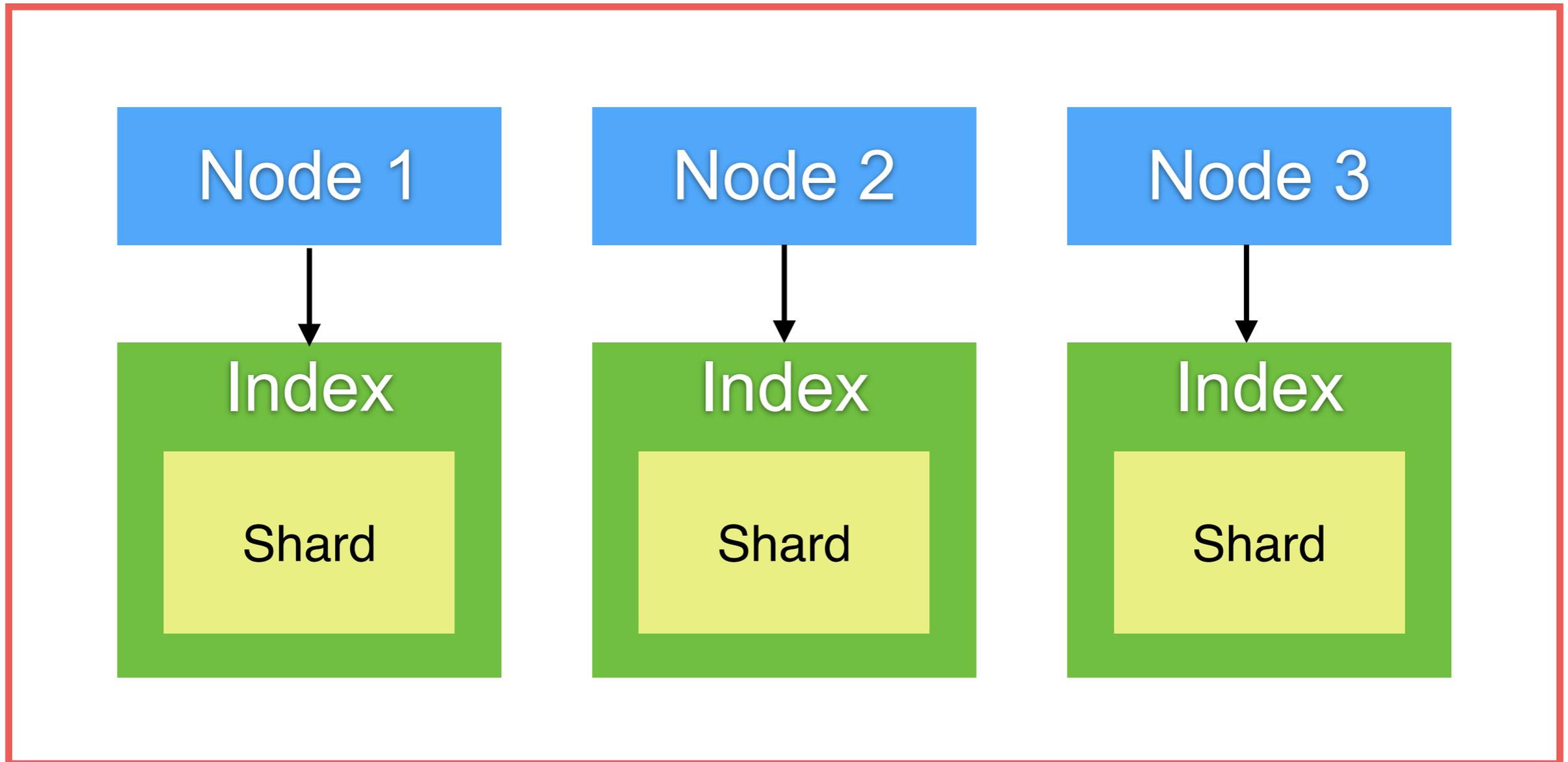
# Documents !!



# Distributed database



# Design for scale (Shard)



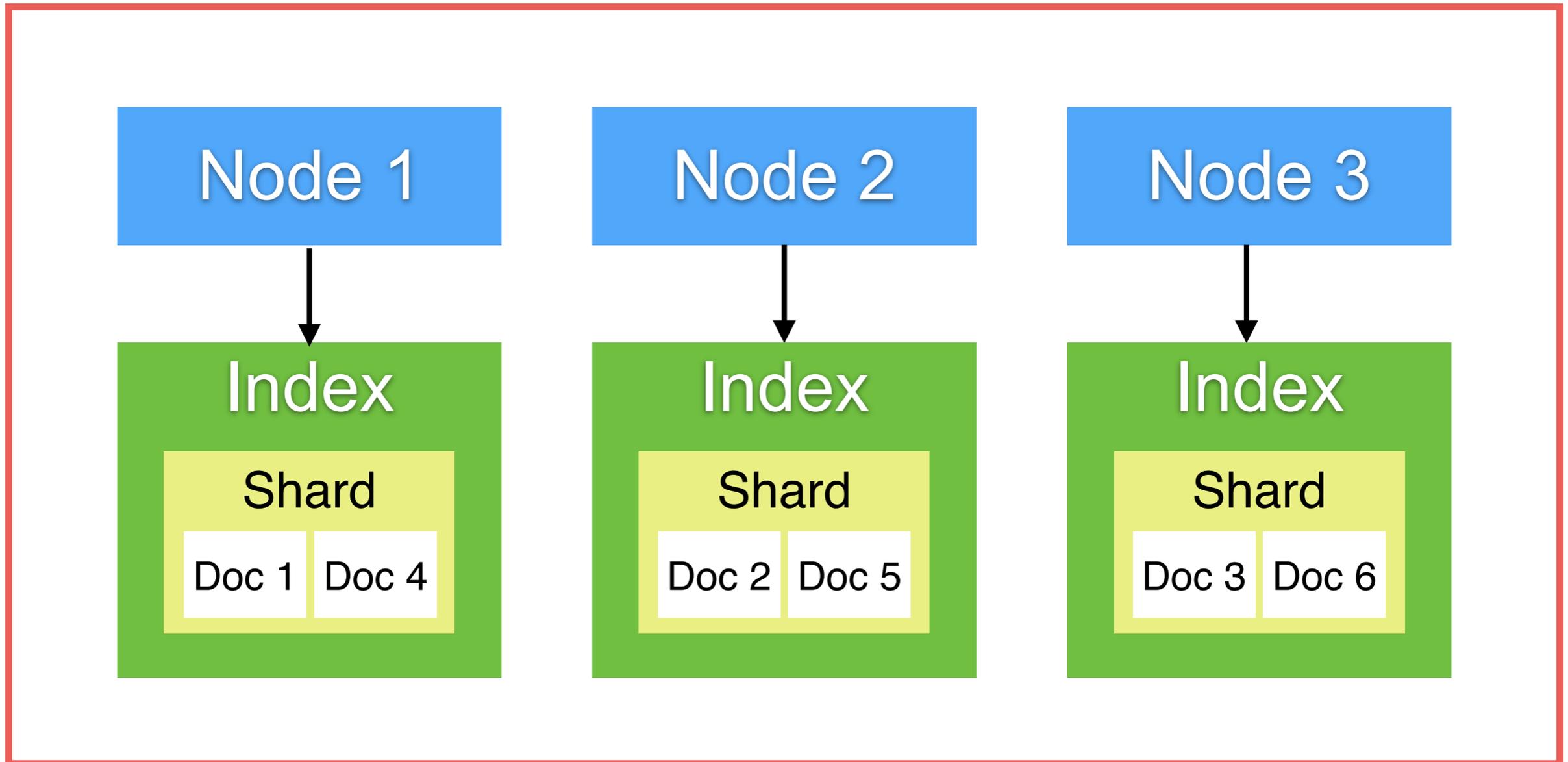
# Index is split into shards

Each shard may be on a different node in cluster

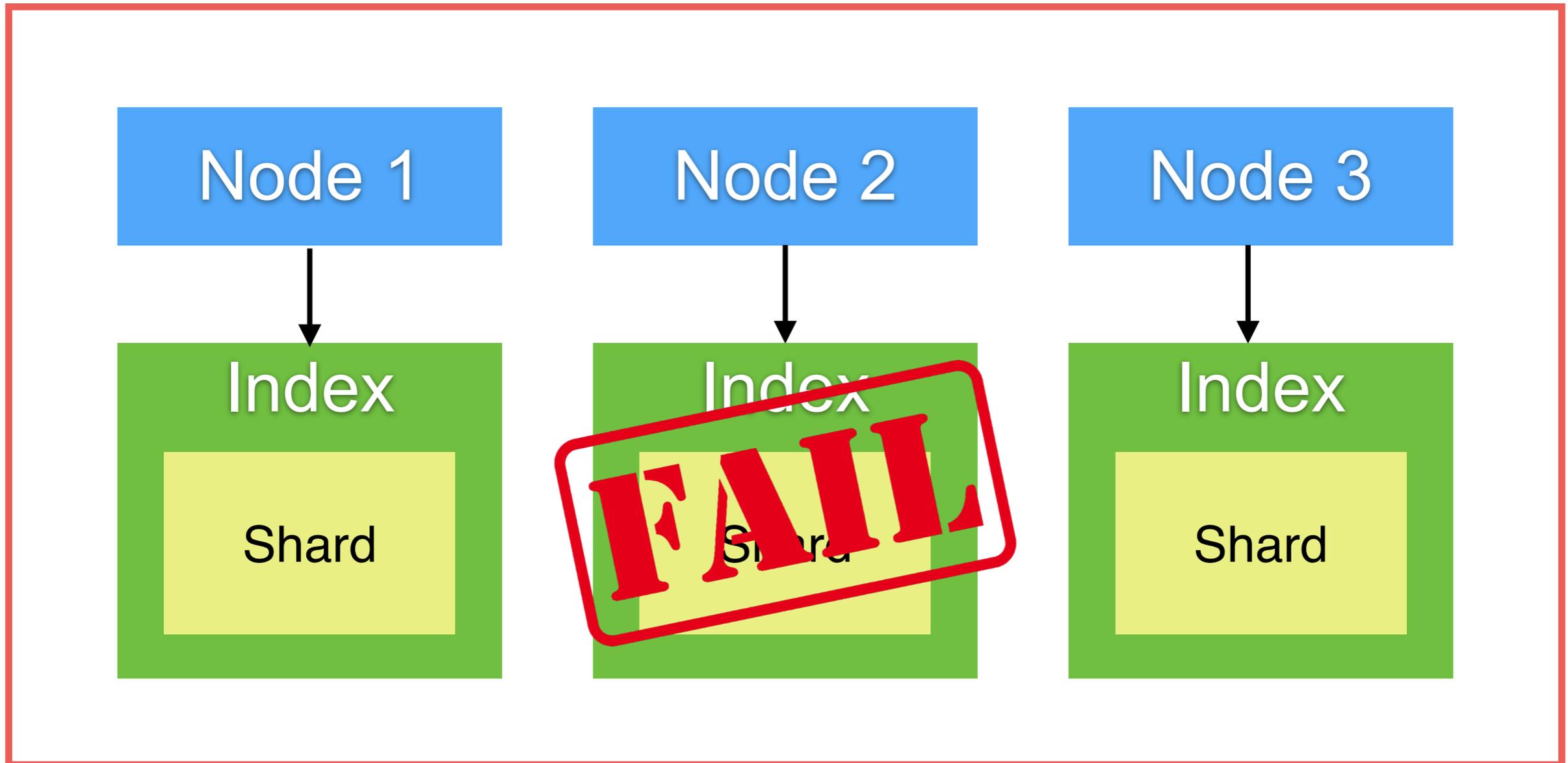
Every shard is a self-contained *Lucene index*



# Documents are hashed



# Design for fail (Replica)



# Replica = copy of shards

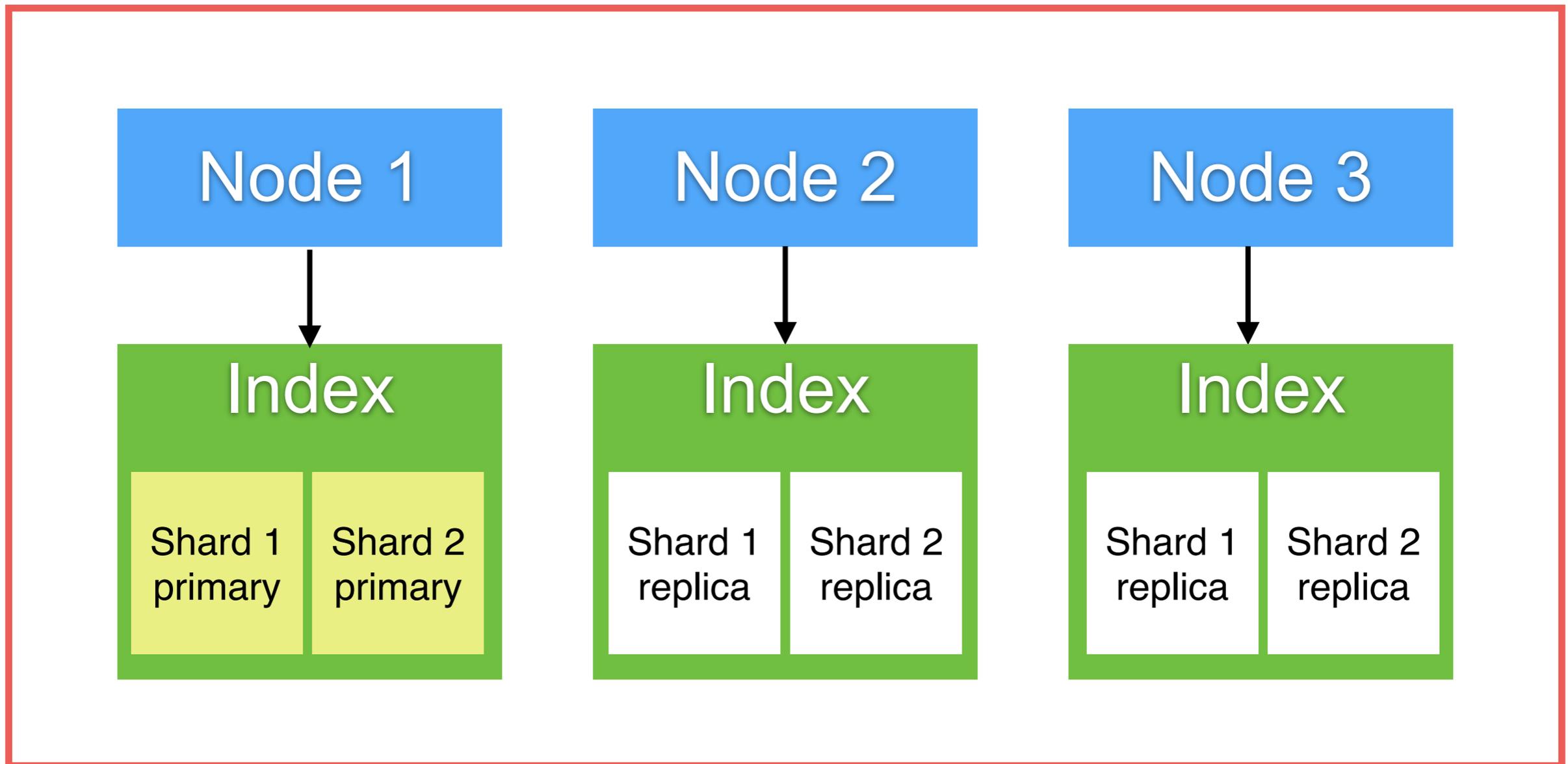
1 replica = 1 Primary + 1 Replica

2 replicas = 1 Primary + 2 Replicas

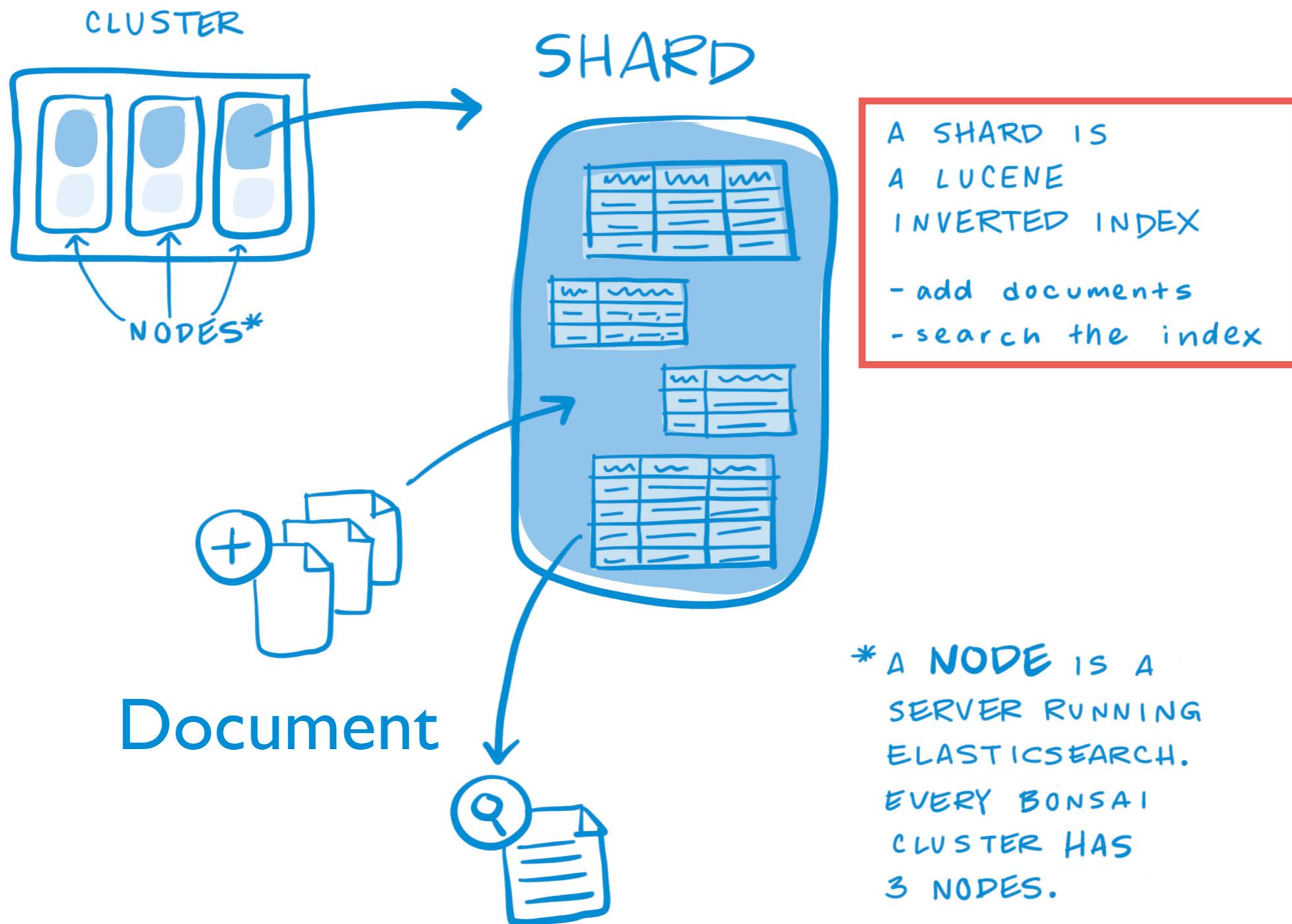
3 replicas = 1 Primary + 3 Replicas



# Replica = 2



# Basic concepts

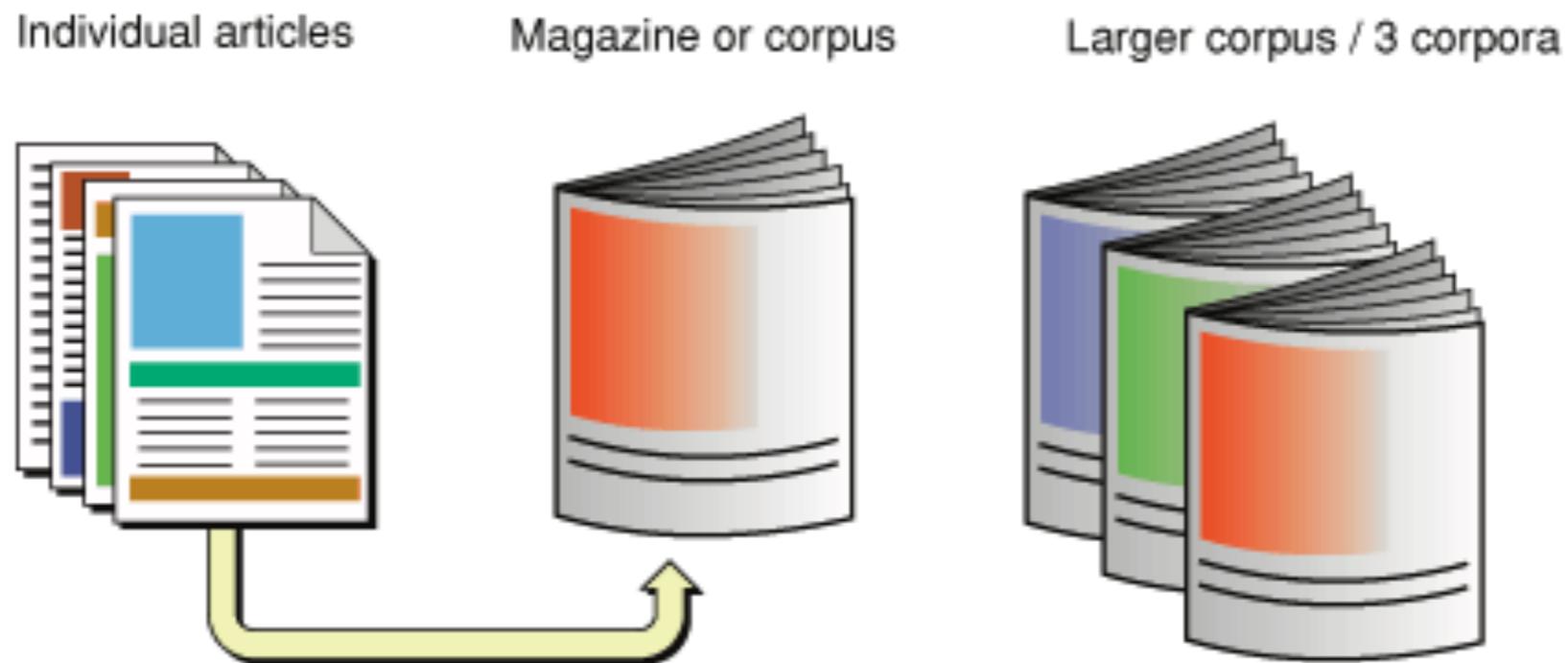


# Inverted Index



# Inverted Index

Corpus is a collection of documents

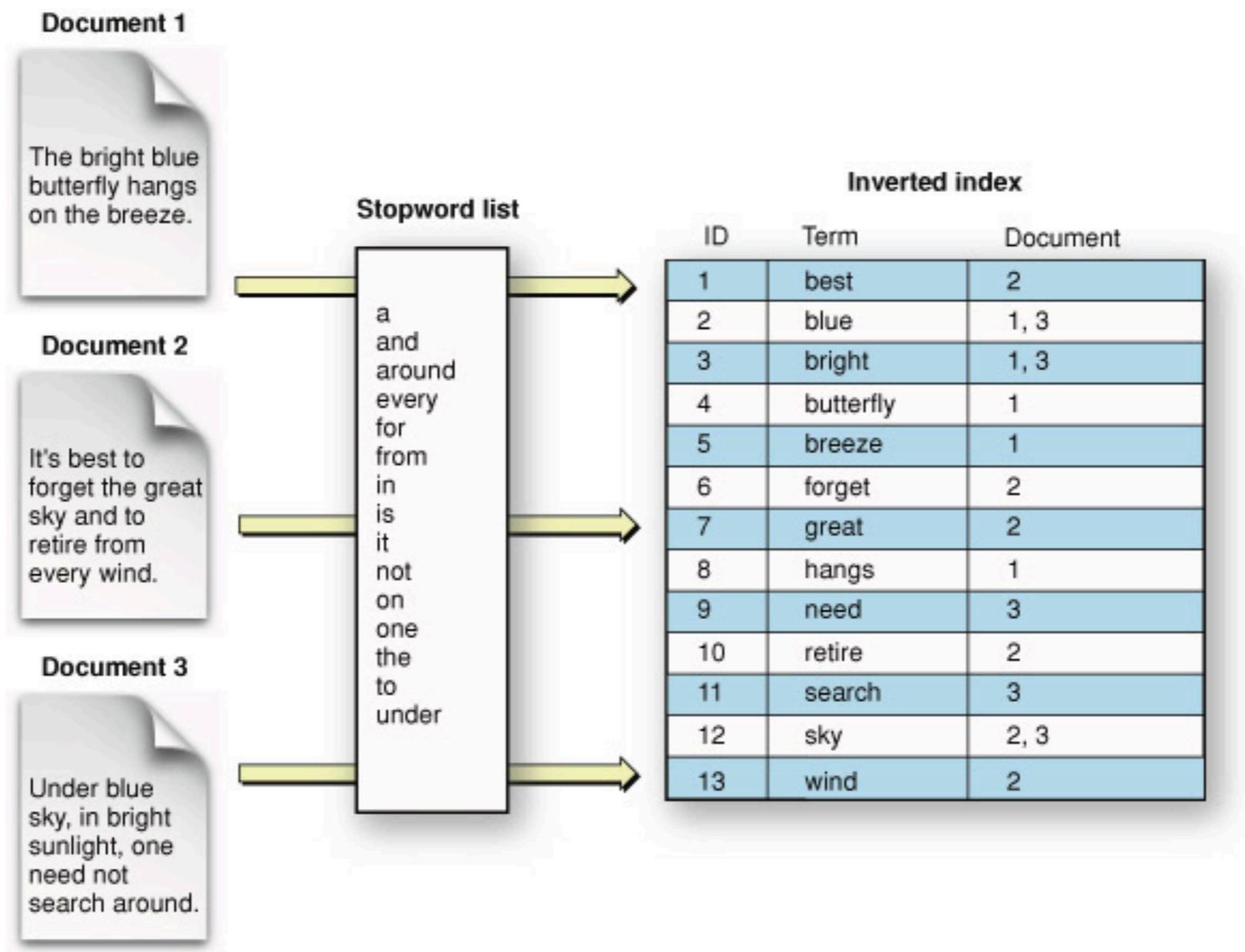


[https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit\\_basics/searchKit\\_basics.html#//apple\\_ref/doc/uid/TP40002843-TPXREF101](https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html#//apple_ref/doc/uid/TP40002843-TPXREF101)



# Inverted Index

## Try to construct index



# Apache Lucene

Elasticsearch

Apache Lucene

JVM

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



# Basic concept

Apache Lucene writes all the information to a structure called **“Inverted Index”**



# Basic concept

Document

Field

Term

Token



# Example

Document no.	Data
1	Elasticsearch Server
2	Mastering Elasticsearch Second Edition
3	Apache Solr Cookbook Third Edition



# Token

Token	Document no.
Elasticsearch	1
Elasticsearch	2
Server	1
Mastering	2
Second	2
Edition	2
Edition	3
Apache	3
Solr	3
Cookbook	3
Third	3



# Term

Token	Count	Document no.
elasticsearch	2	1,2
server	1	1
mastering	1	2
second	1	2
edition	2	2,3
apache	1	3
solr	1	3
cookbook	1	3
third	1	3



# Lucene inverted index

Token	Count	Document no.
elasticsearch	2	1,2
server	1	1
mastering	1	2
second	1	2
edition	2	2,3
apache	1	3
solr	1	3
cookbook	1	3
third	1	3



# Lucene inverted index

Write-once and read-many-times structure

Called “**Segment**”

Can't be delete (just marked to deleted)



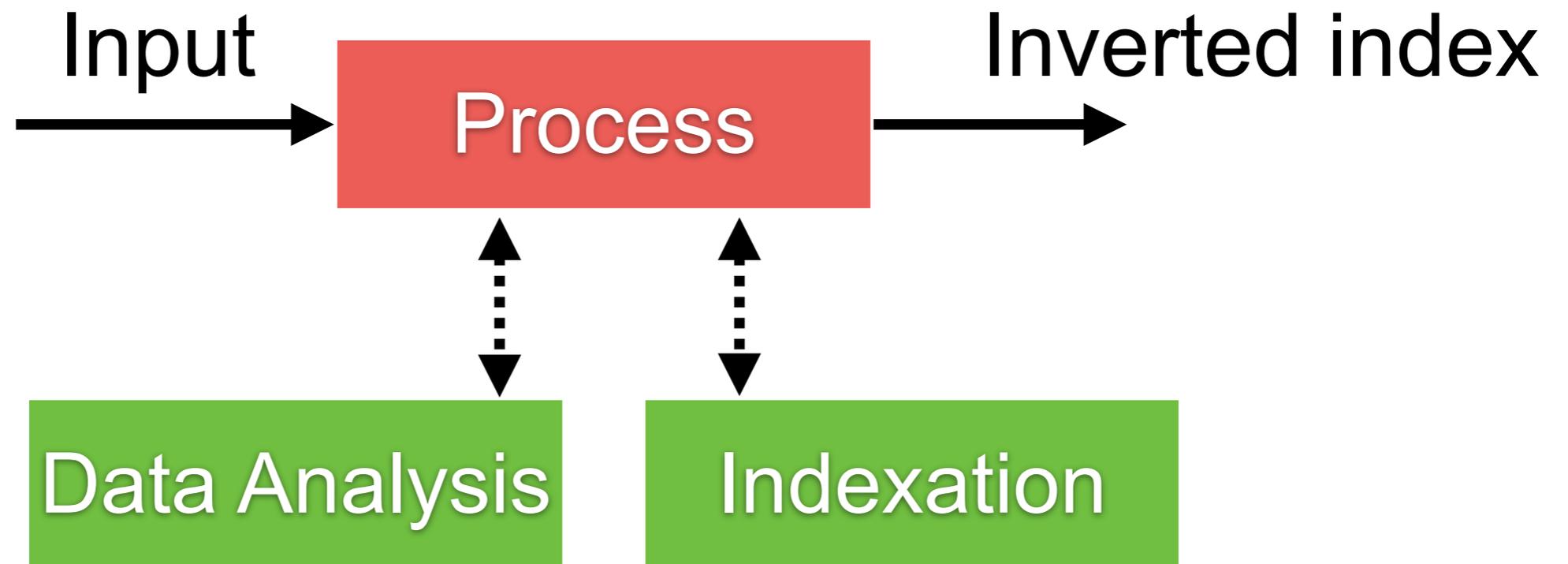
# Input data analysis

Write-once and read-many-times structure

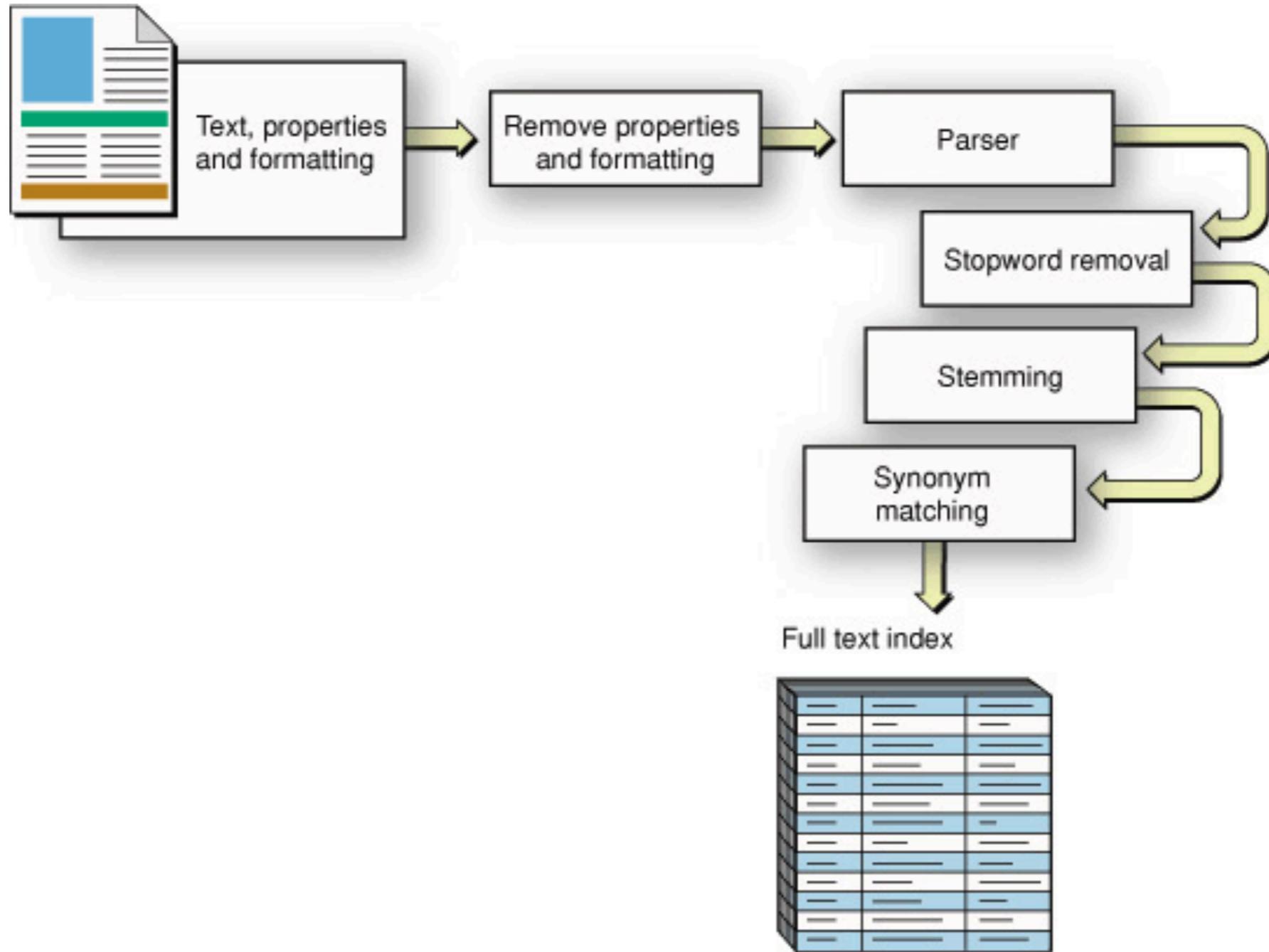


# Input data analysis

Write-once and read-many-times structure



# Process ?



# CRUD with Elasticsearch

02-crud/book\_document.json



# CRUD with Elasticsearch

**Create document**

**Read document**

**Update document**

**Delete document**



# Compare with RDBMS

Database

Table

Row

Column

Index

Type\*

Document

Field

*\* Only 1 type per index*



# 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



# 1 Type per Index

## Removal of mapping types



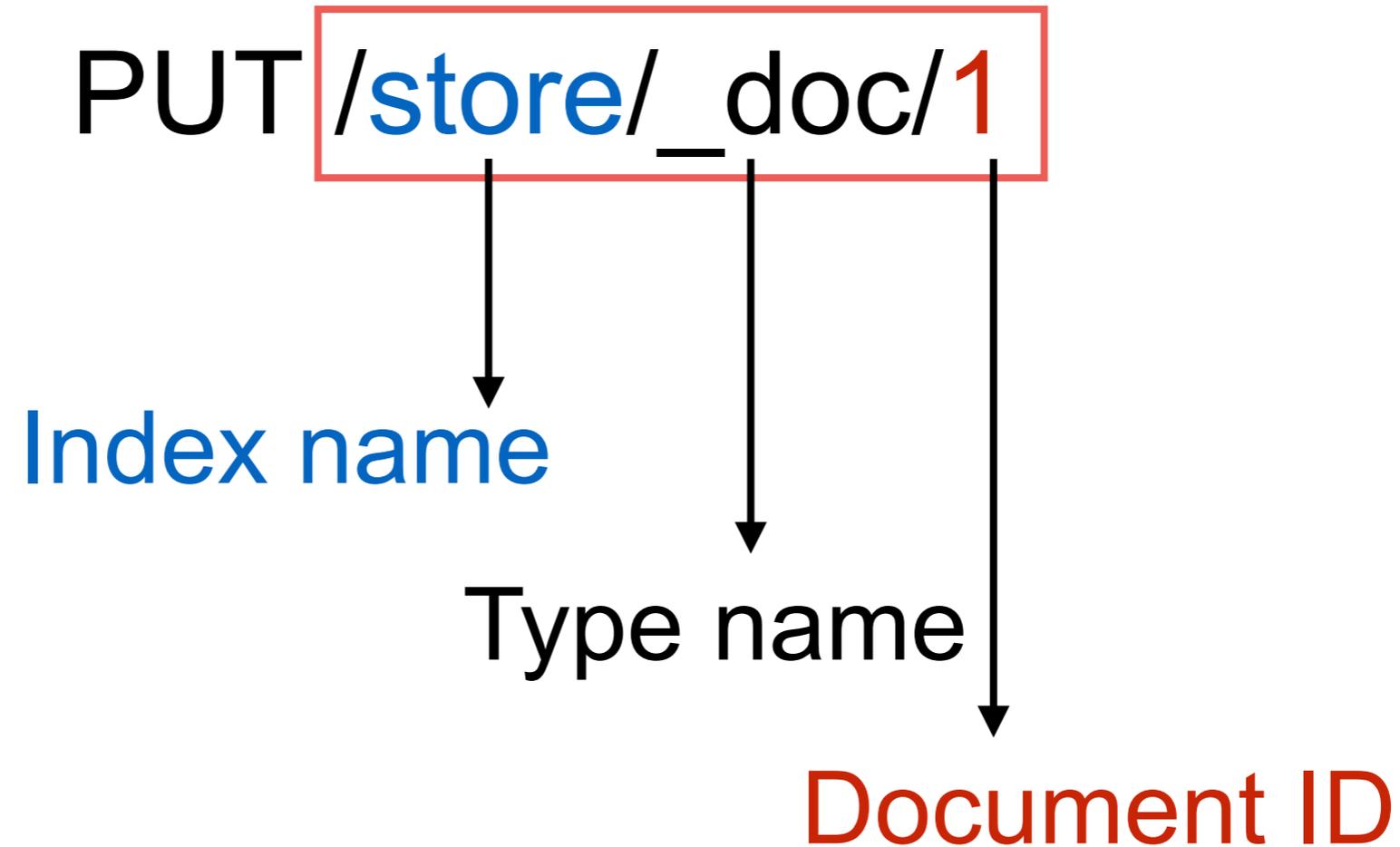
**IMPORTANT**

Indices created in Elasticsearch 6.0.0 or later may only contain a single [mapping type](#). Indices created in 5.x with multiple mapping types will continue to function as before in Elasticsearch 6.x. Mapping types will be completely removed in Elasticsearch 7.0.0.

<https://www.elastic.co/guide/en/elasticsearch/reference/6.5/removal-of-types.html>



# Create document (1 type)



# 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": {
```



# Read document

GET /store/book/1

{

```
"_index": "store",  
"_type": "book",  
"_id": "1",  
"_version": 1,  
"found": true,
```

Information of document

```
"_source": {  
  "title": "Elasticsearch: The Definitive Guide",  
  "author_name": [  
    "Clinton Gormley",  
    "Zachary Tong"  
  ],  
  "tag": [  
    "search",  
    "computer"  
  ]  
}
```

}



# Update document

Whole document

Partial document



# Update whole document

PUT /store/\_doc/123

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



# Update partial document

POST /store/\_update/123

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



# Delete document

```
DELETE /store/_doc/1
```

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

*Not delete document !!*  
*Marked deleted only*



# Delete index (delete from disk)

```
DELETE /store
```



# More features

Update by query

Delete by query

Partial update document



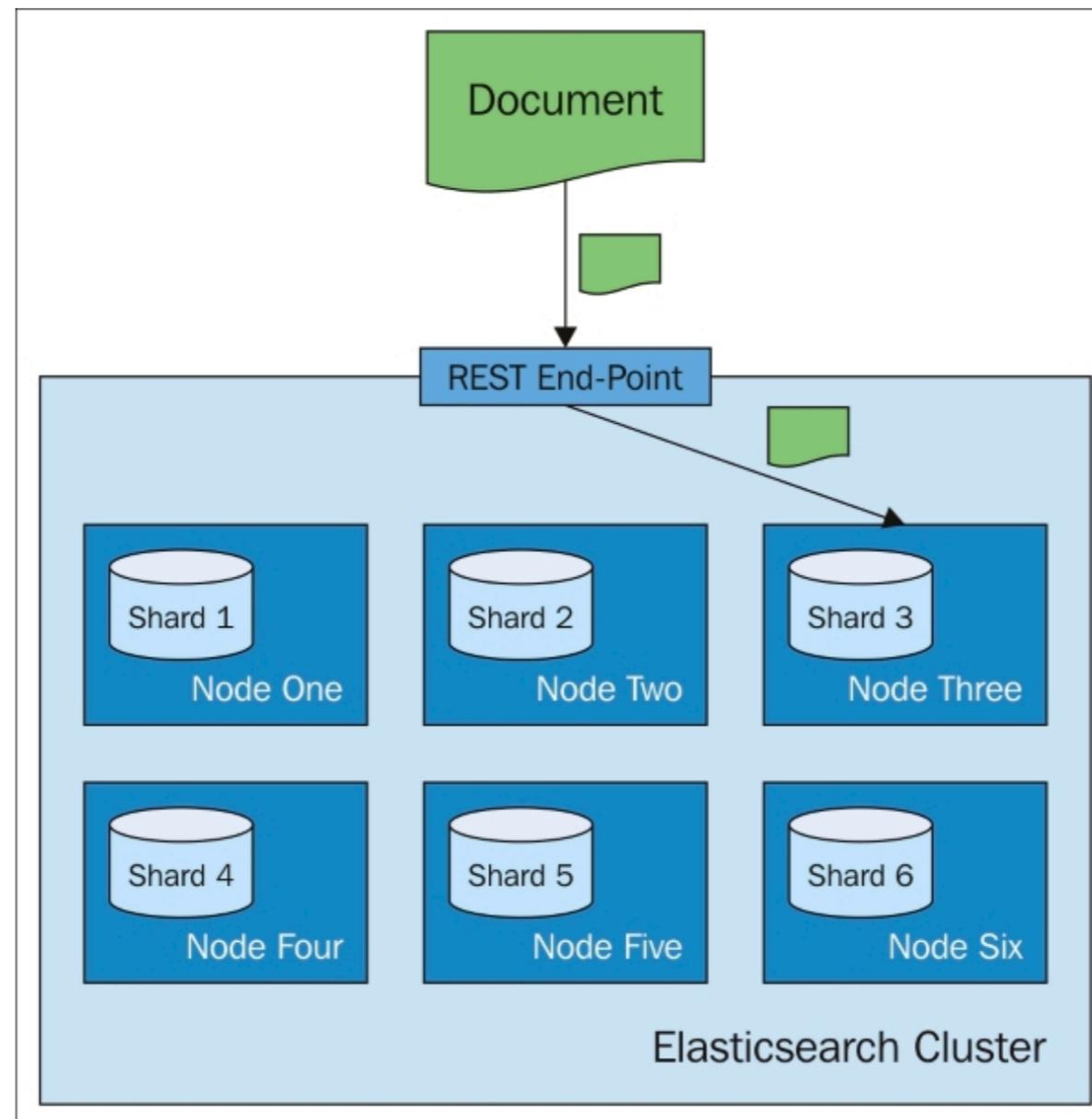
# Routing

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



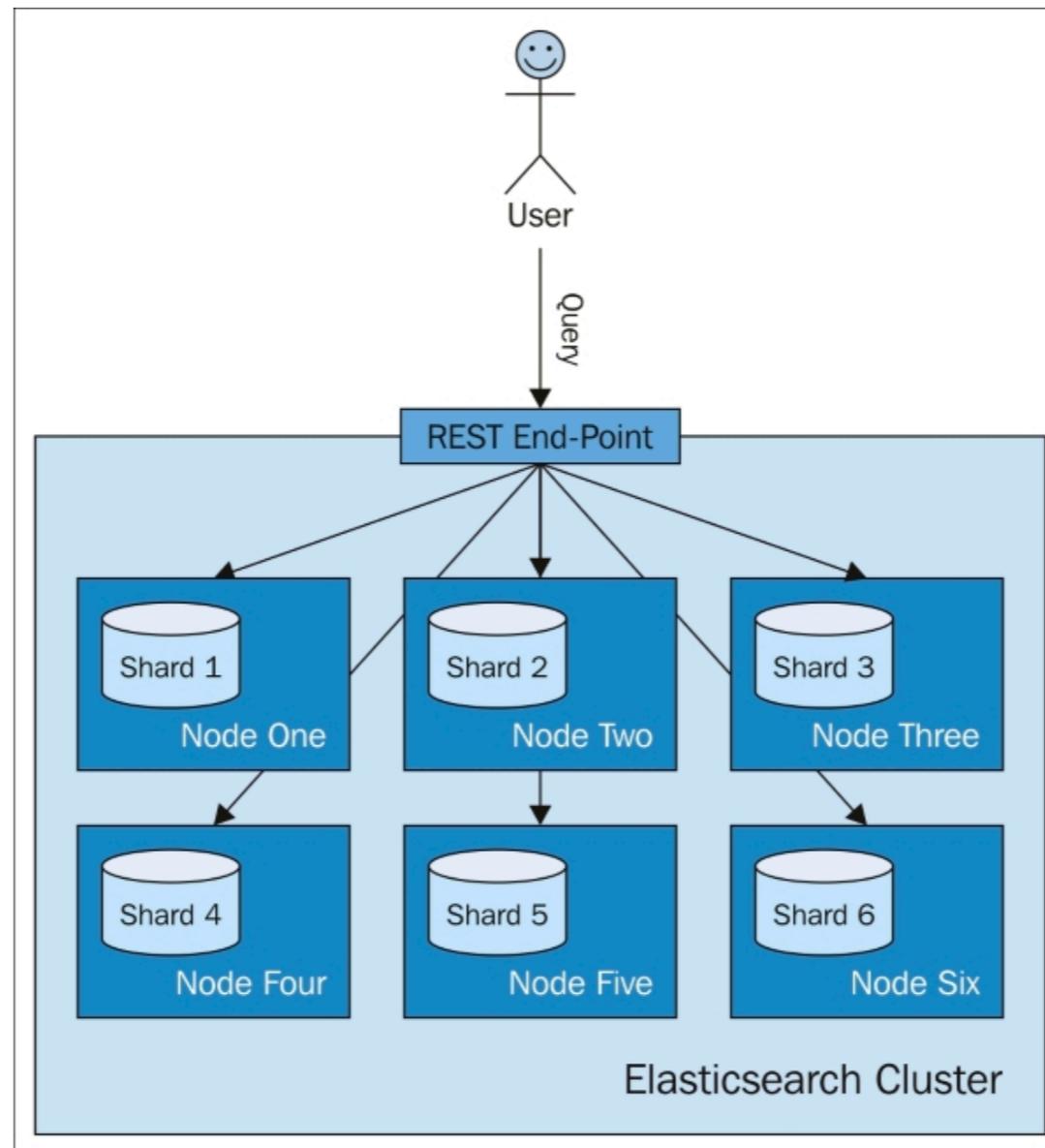
# Default indexing

ES calculate the hash value of the doc id



# Default searching/query

Query all the shards to get data  
(depend on search type)



# Custom routing

Routing field

Routing to index partition



# Routing field

```
shard_num =  
hash(_routing) % num_primary_shards
```



# Routing to index partition

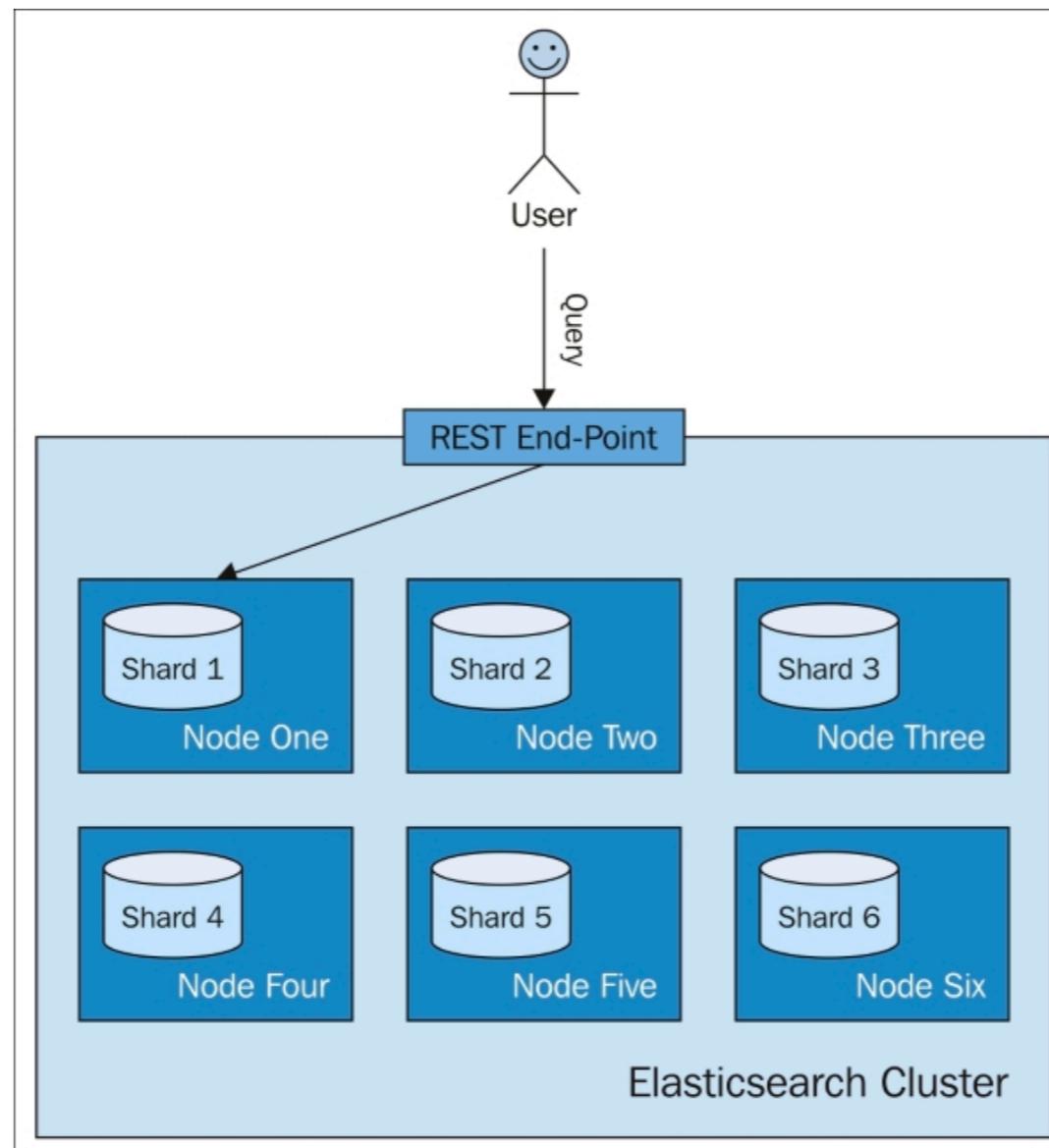
$$\begin{aligned} \text{shard\_num} \\ = \\ (\text{hash}(\_routing) + \text{hash}(\_id) \% \text{routing\_partition\_size}) \\ \% \\ \text{num\_primary\_shards} \end{aligned}$$

*routing\_partition\_size = 1 (default)*



# Custom routing

ES will send our query to a single shard



# Bulk API

03-bulk/book\_bulk.json

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



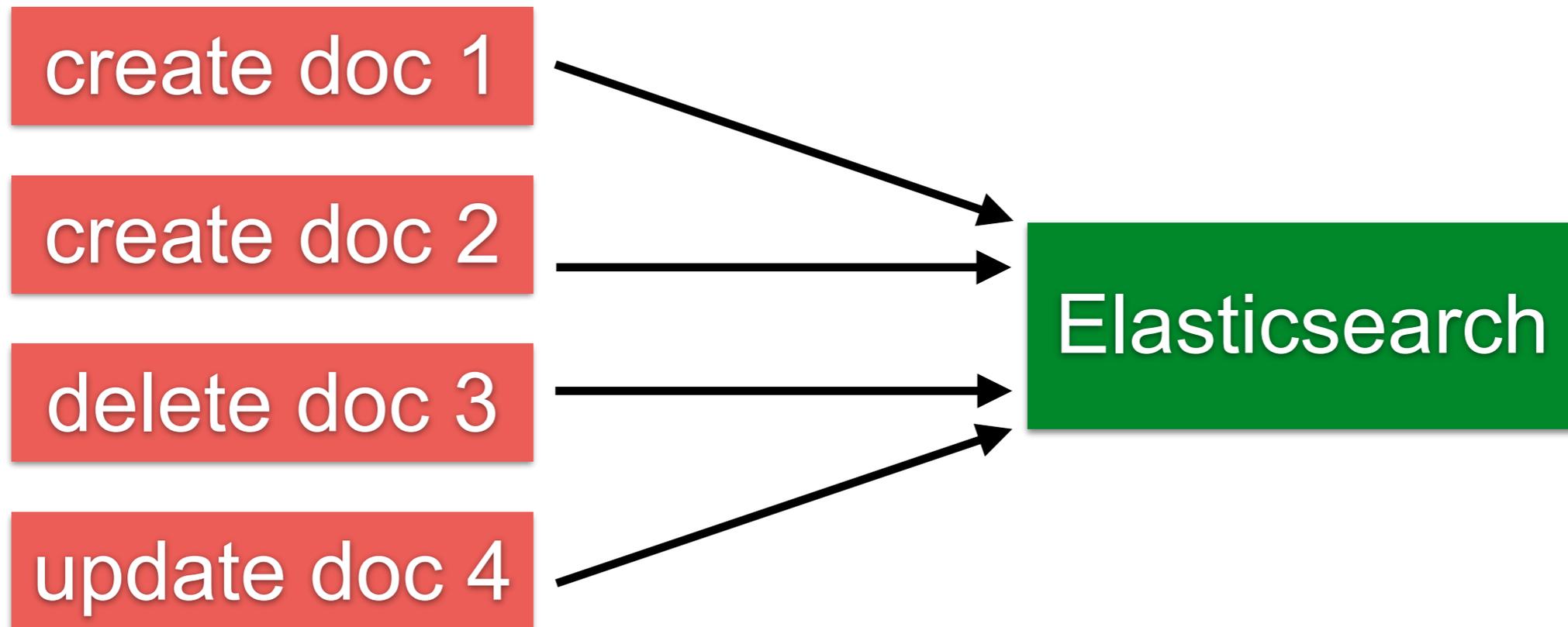
# Bulk API

Perform many index/delete operation in  
single API call

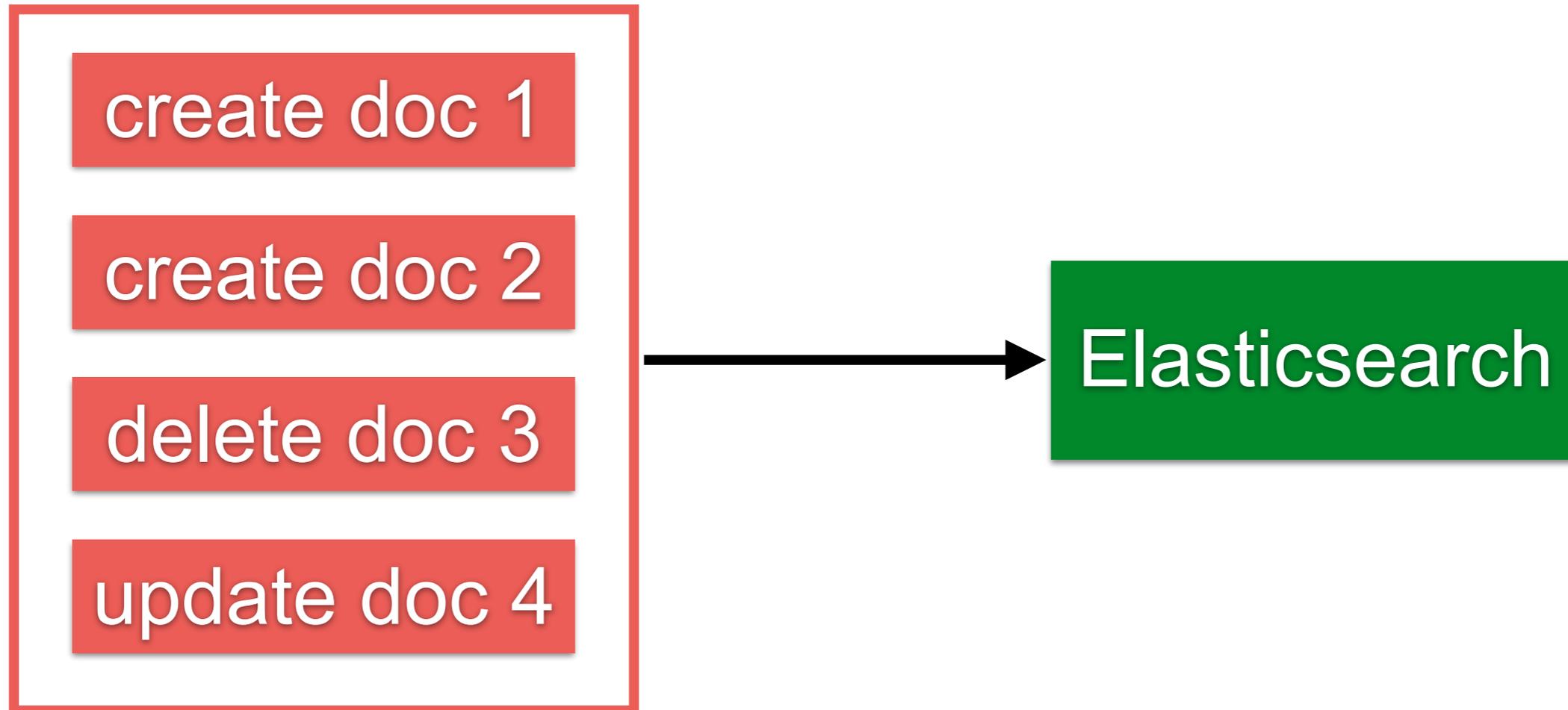
Increase indexing speed



# Without Bulk API



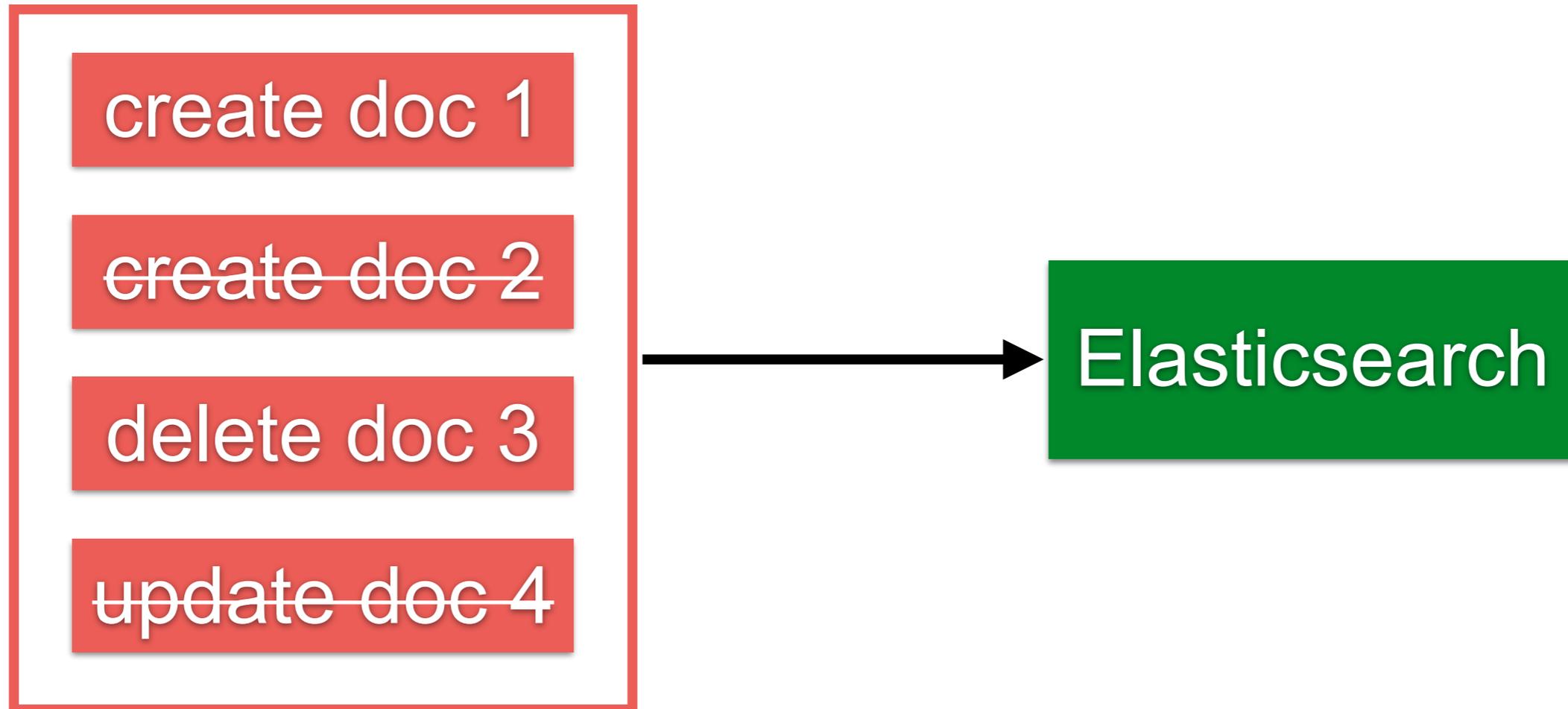
# With Bulk API



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



# Searching / Query



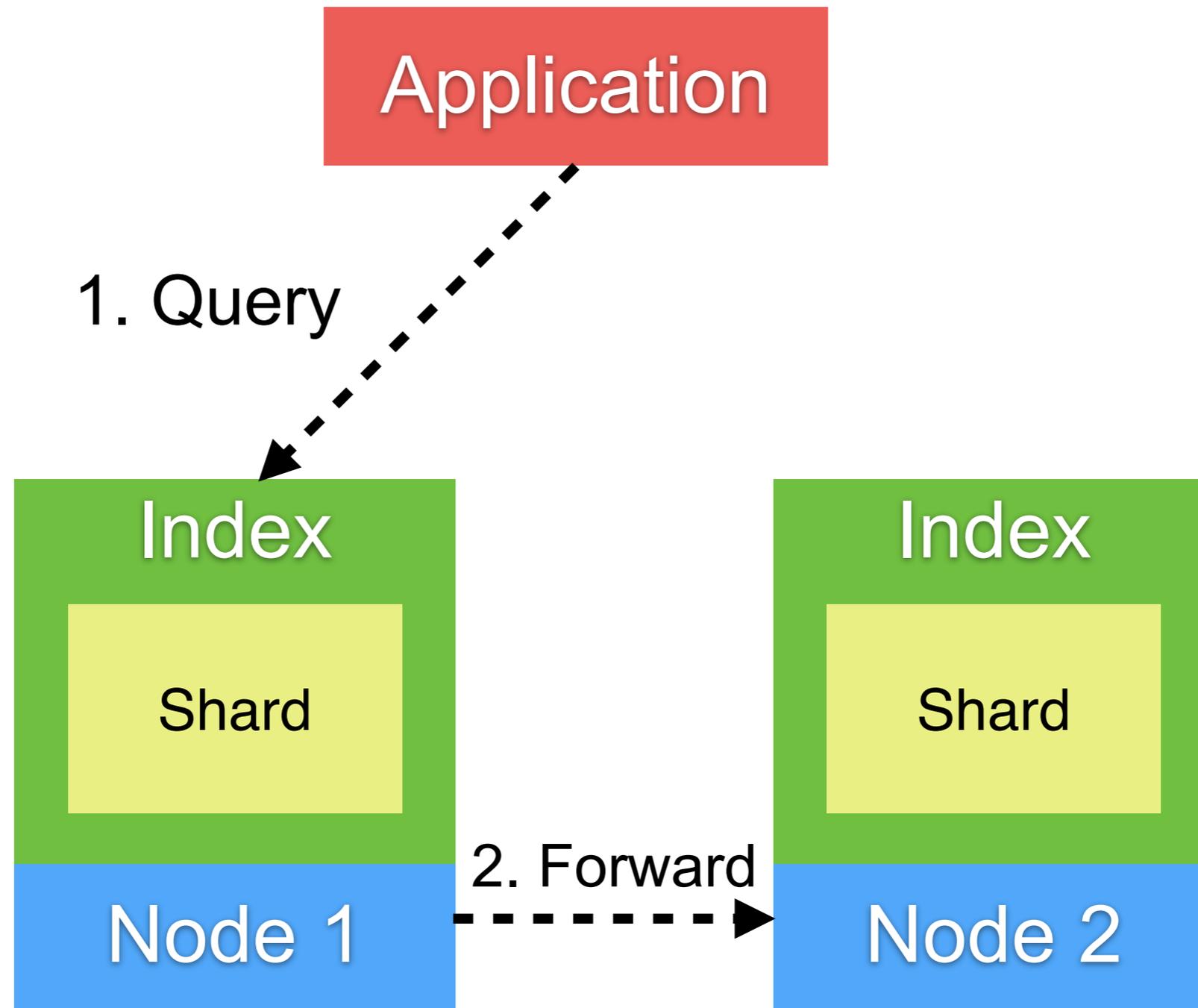
# Searching

Scattering phase

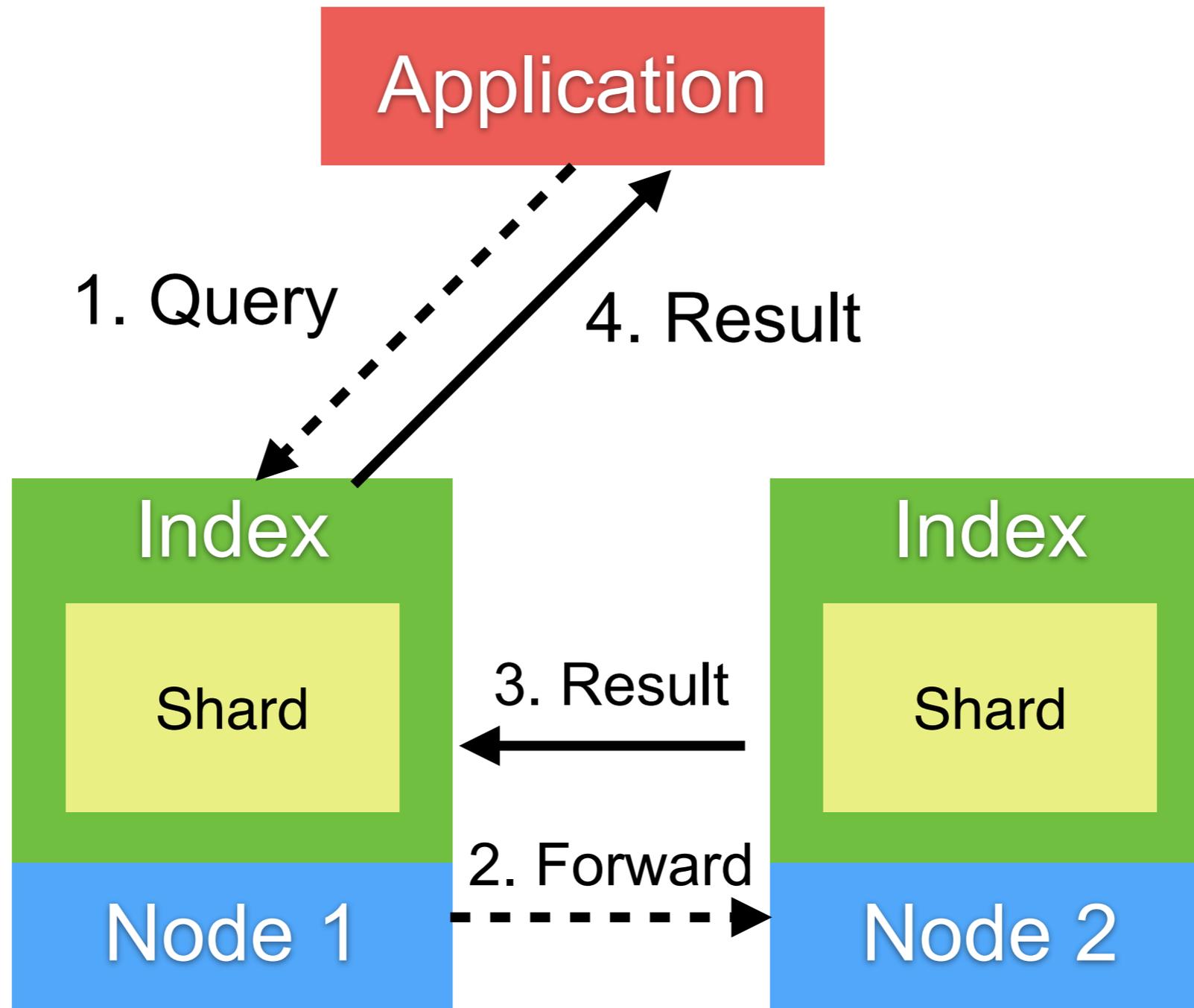
Gathering phase



# Scattering phase



# Gathering phase



# Query DSL

04-search/book\_search.json



# Query DSL

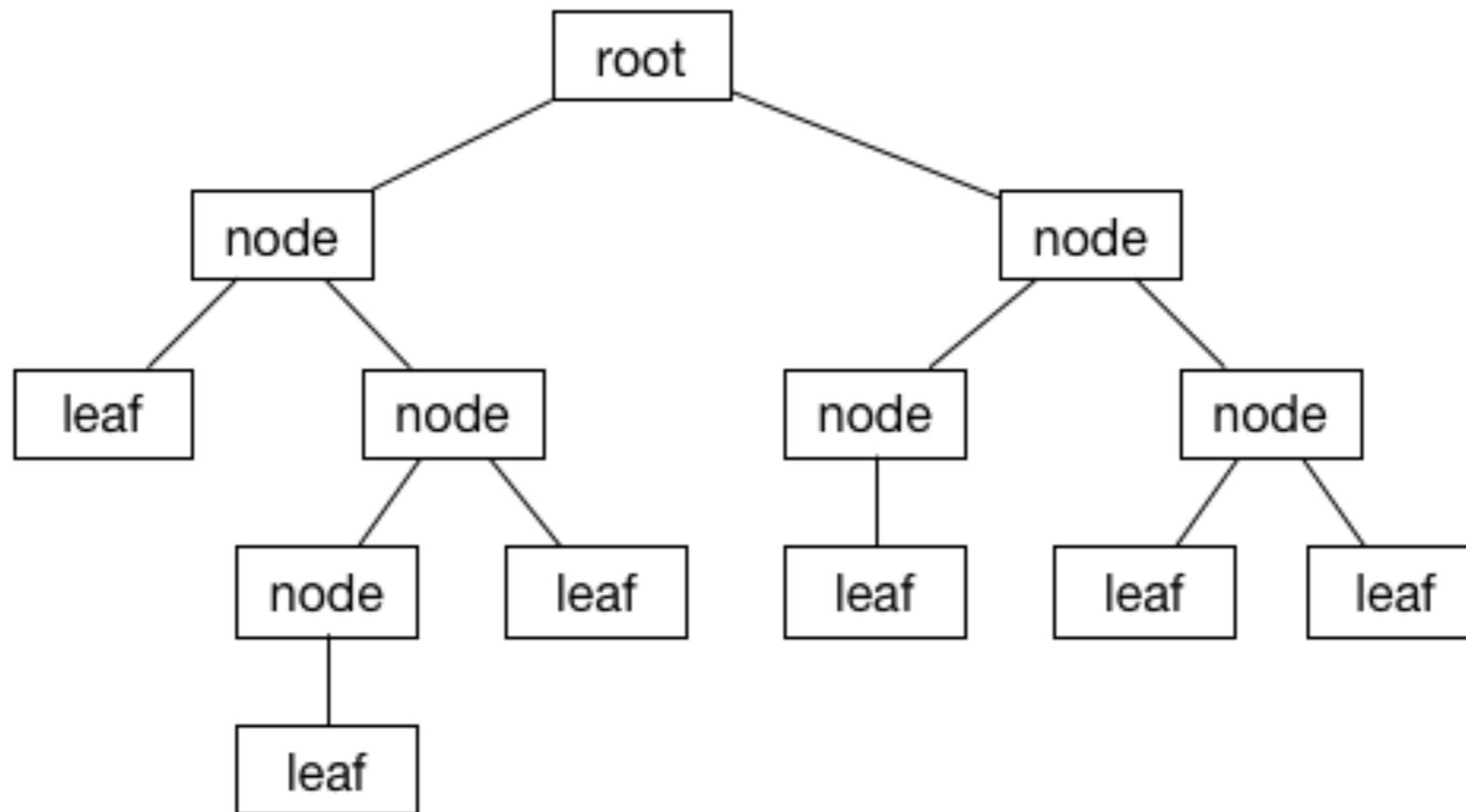
**Domain Specific Language** for query data  
**Flexible query language**  
**Based on JSON format**

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



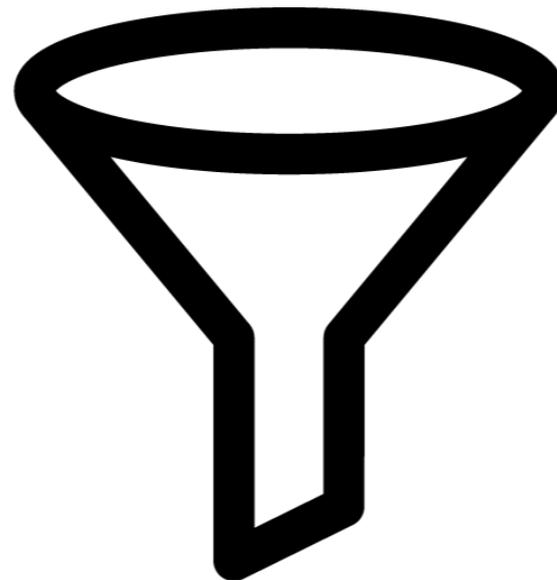
# Query DSL

1. Leaf query clause
2. Compound query clause



# Query DSL

Query (unstructured data)  
Filter (structured data)



Query



# 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": [{}],  
      "filter": [{}]  
    }  
  }  
}
```



# Workshop

amazon   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

1-16 of 119 results for "elasticsearch"

Sort by

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



SPONSORED BY PACKT PUBLISHING

## Complete Database Solutions with PostgreSQL

[Shop now >](#)



SQL Server 2017 Administrators Guide

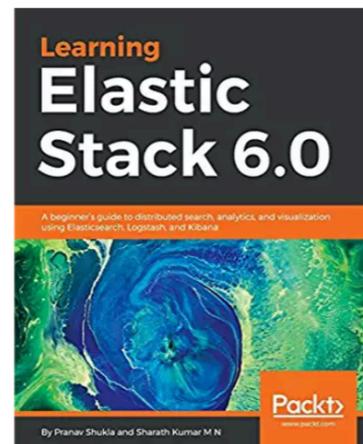
★★★★☆ 3  
✓prime



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

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.

★★★★★ 7

Paperback

\$34.99

In Stock

[Previous Page](#)

[1](#) [2](#) [3](#) ... [8](#)

[Next Page](#)



# Query

amazon

Deliver to Thailand Departments Your Amazon.com Today's Deals Gift Cards Sell

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

# Sorting

Sort by

1-16 of 119 results for "elasticsearch"

# 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



SPONSORED BY PACKT PUBLISHING

## Complete Database Solutions with PostgreSQL

Shop now >



SQL Server 2017 Administrators Guide

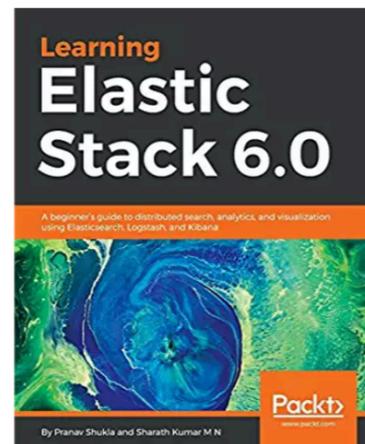
★★★★☆ 3  
prime



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

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.

★★★★★ 7

Paperback

# Paging

Previous Page 1 2 3 ... 8 Next Page



# Aggregation API

05-aggregation/book\_aggregation.json

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



```
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/\_search

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



# Count by category

GET /store/\_search

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



# Count by category

```
GET /store/_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/\_search

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



# Range of price

GET /store/\_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/\_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

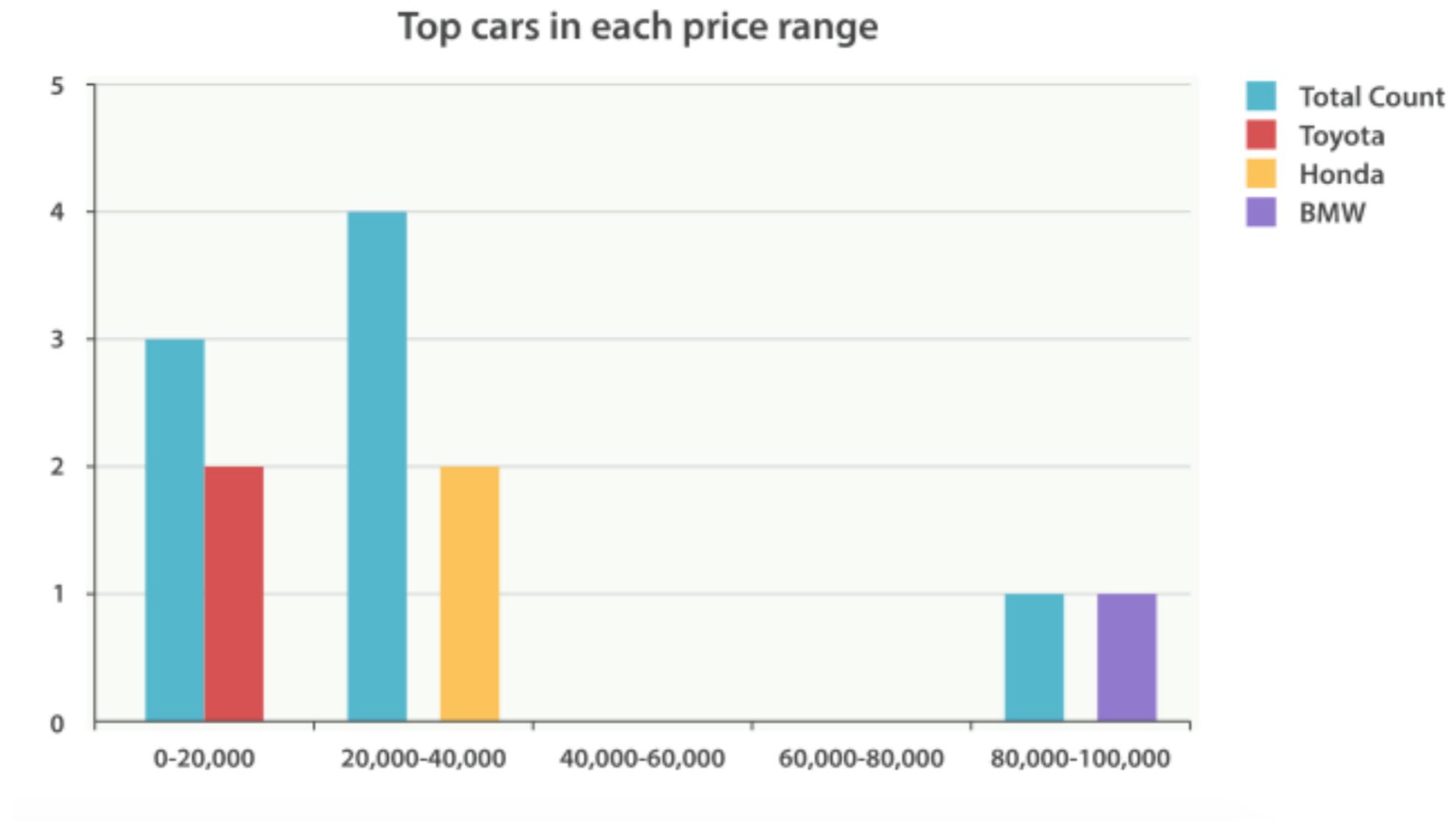
Statistic of best seller by color

Detail of car in each color

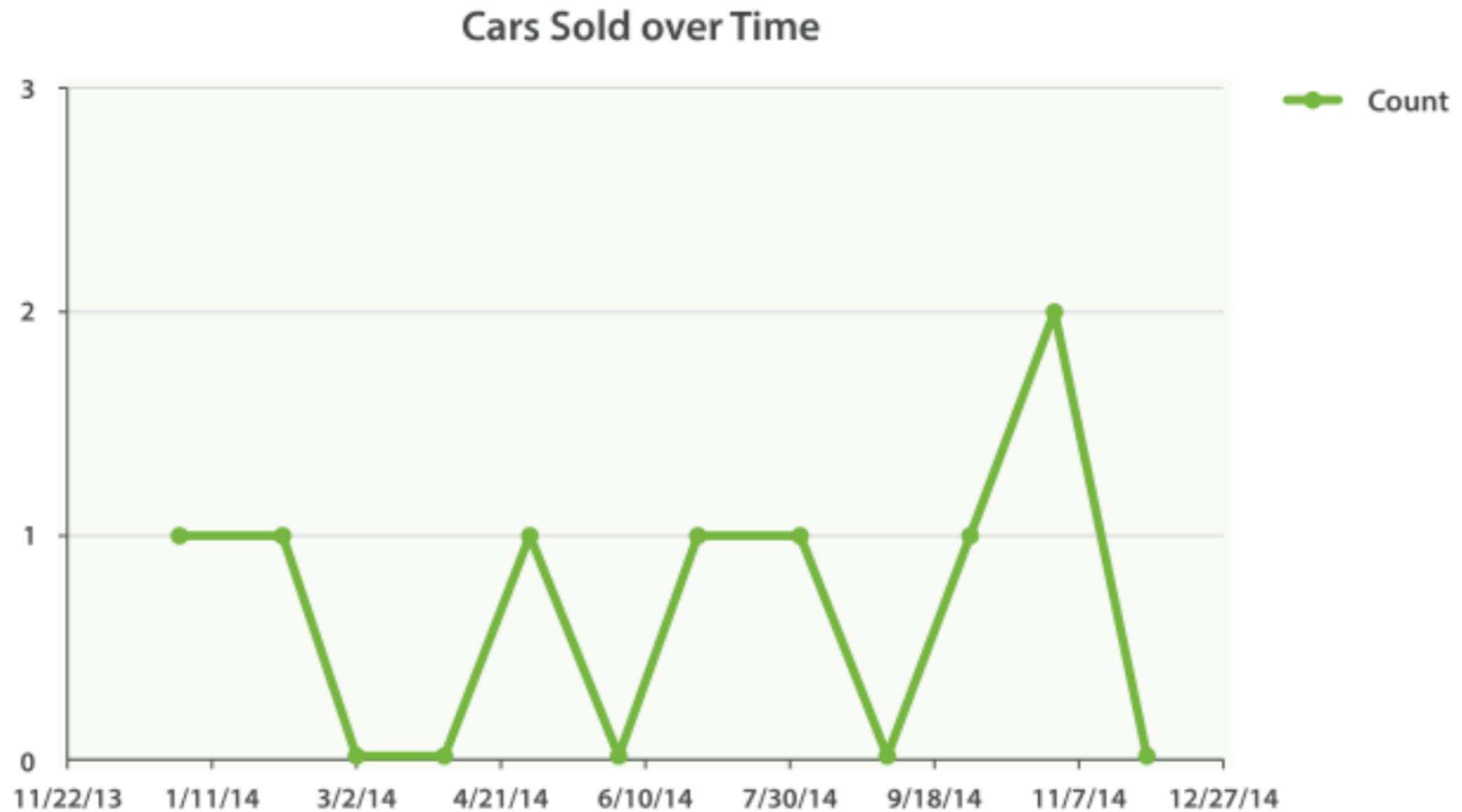
Min/max of price by make



# Top cars in each price range ?



# Cars sold over Time ?



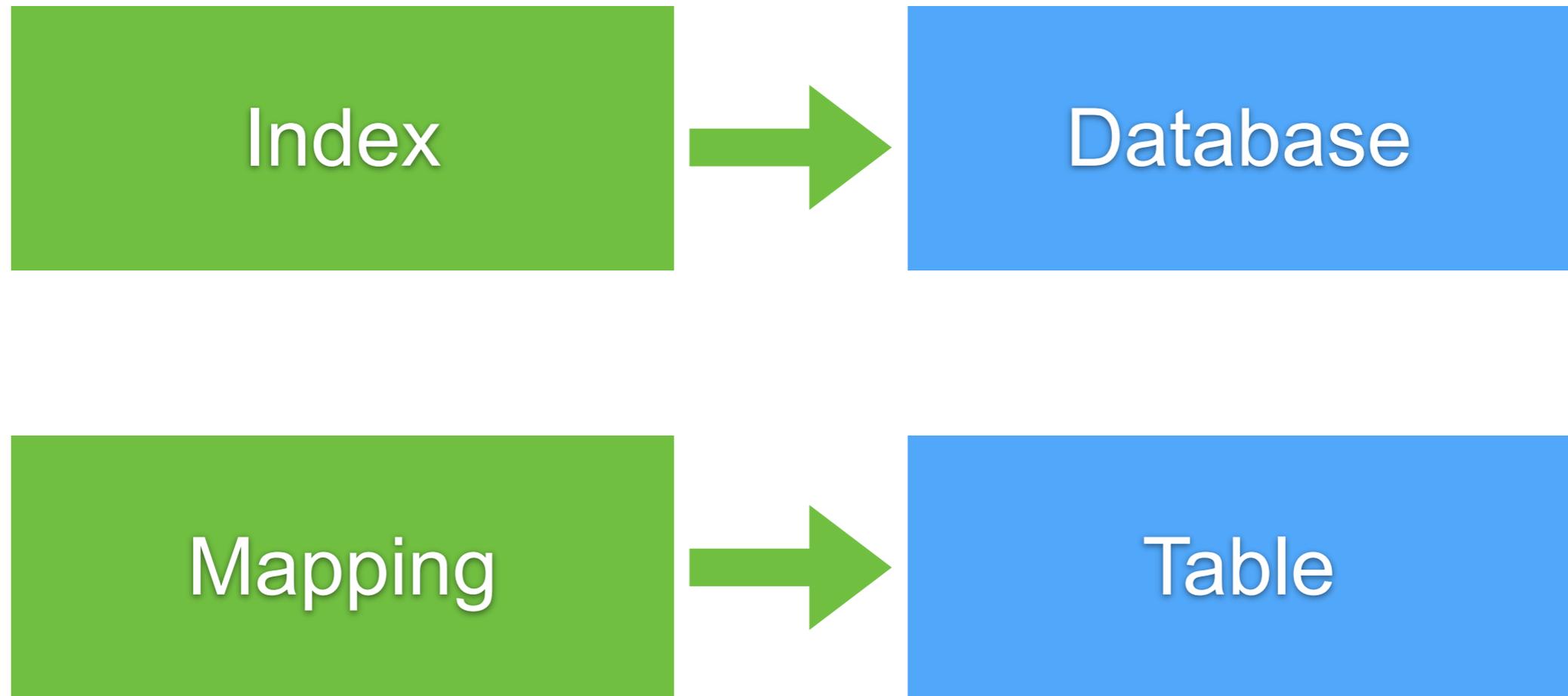
# Mapping

**(Structure of document)**

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



# Explicit Mapping



# 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

Core

Complex

Geo

Specialized



# Field Datatypes

<b>text</b>	<b>match_only_text</b>
<b>keyword</b>	ip
long	boolean
double	completion
geo_point	geo_shape
array	object
nested	binary
date	

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



# Field Datatypes

text	date
keyword	ip
long	boolean
double	completion
geo_point	geo_shape
<b>array</b>	<b>object</b>
<b>nested</b>	binary

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



# 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/Explicit mapping

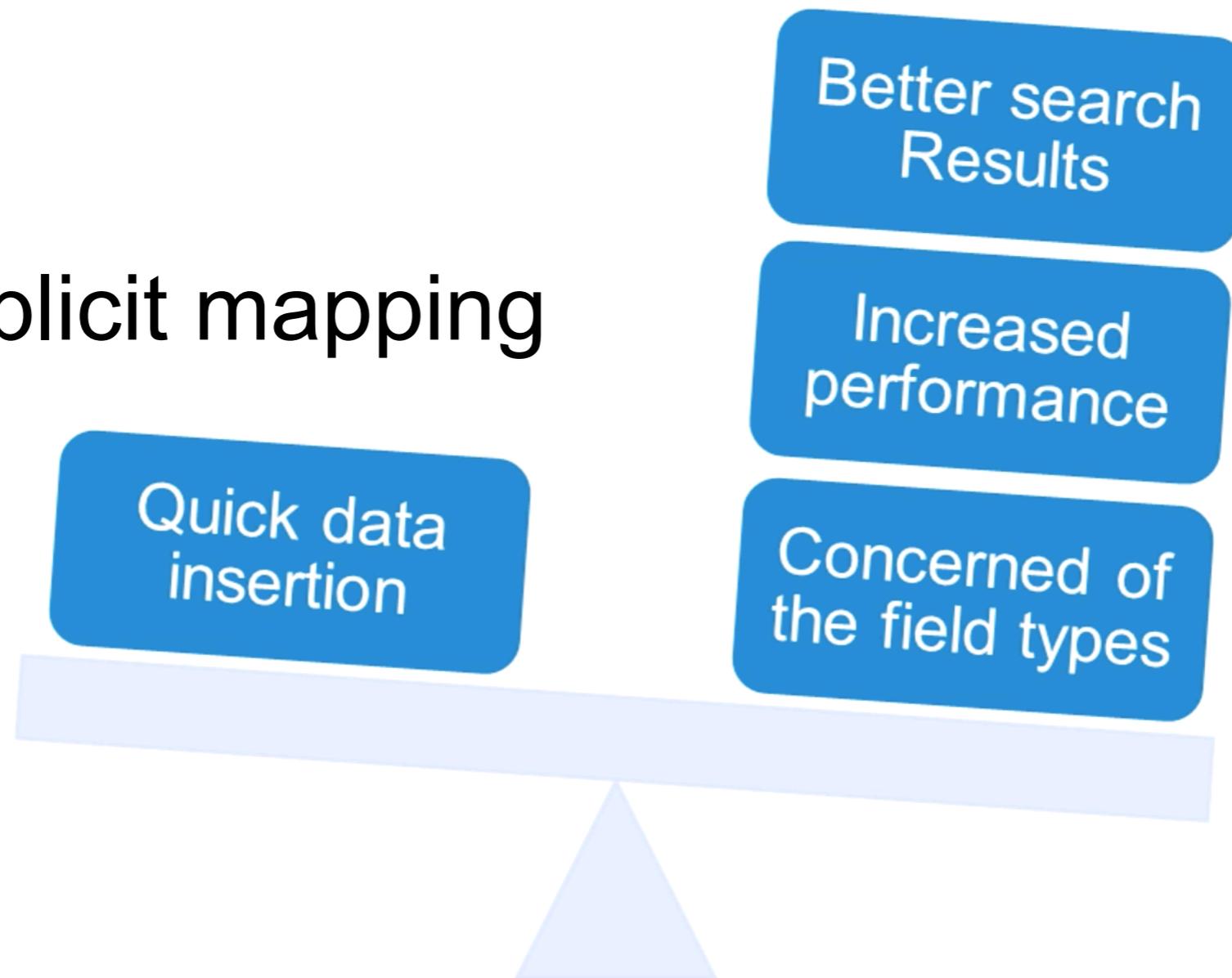
Fields and mapping types not need to defined before being used



# Mapping

Manual mapping

Explicit mapping



# Custom Mapping in ES 7

Don't specify the type name !!

```
PUT project/_mapping
{
  "properties": {
    "user_id": {
      "type": "keyword"
    },
    "image_name": {
      "type": "keyword"
    }
  }
}
```



# Analyzer

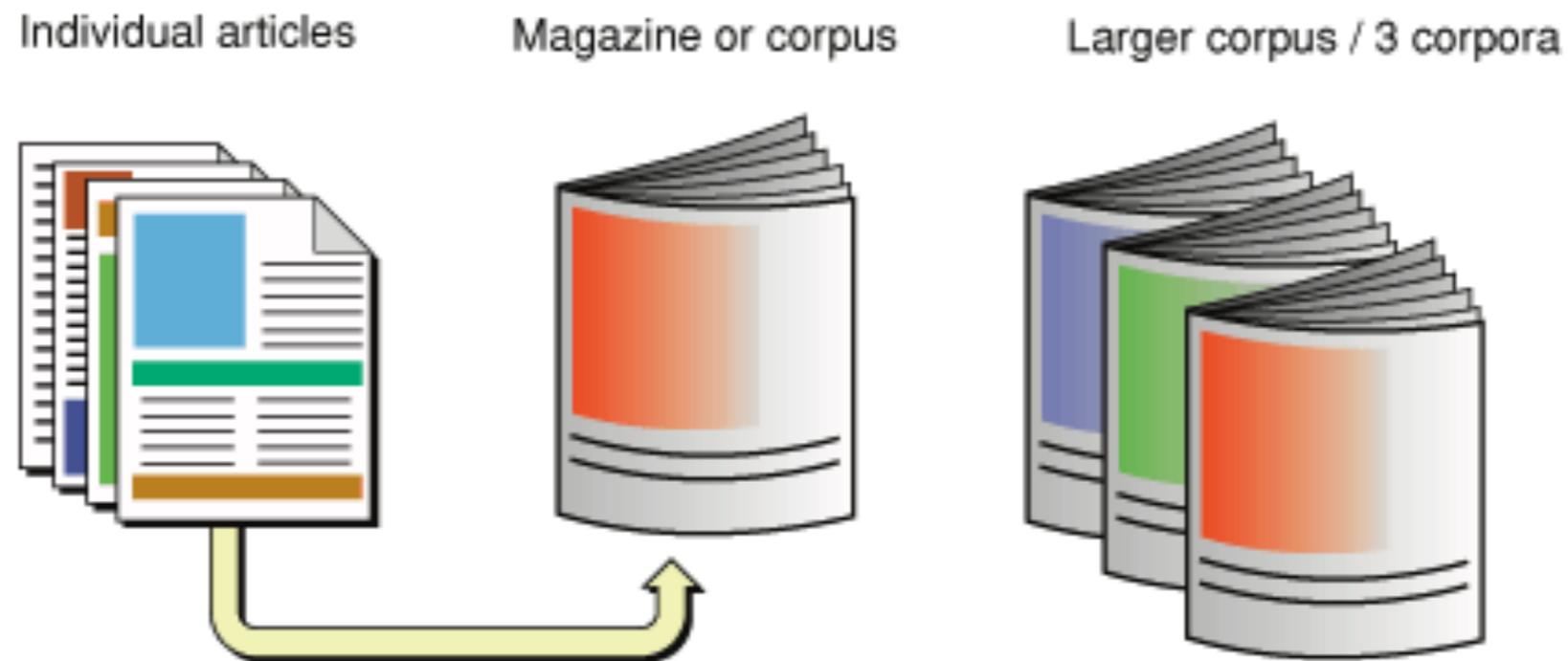
## 07-analyzer

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



# Inverted Index

Corpus is a collection of documents

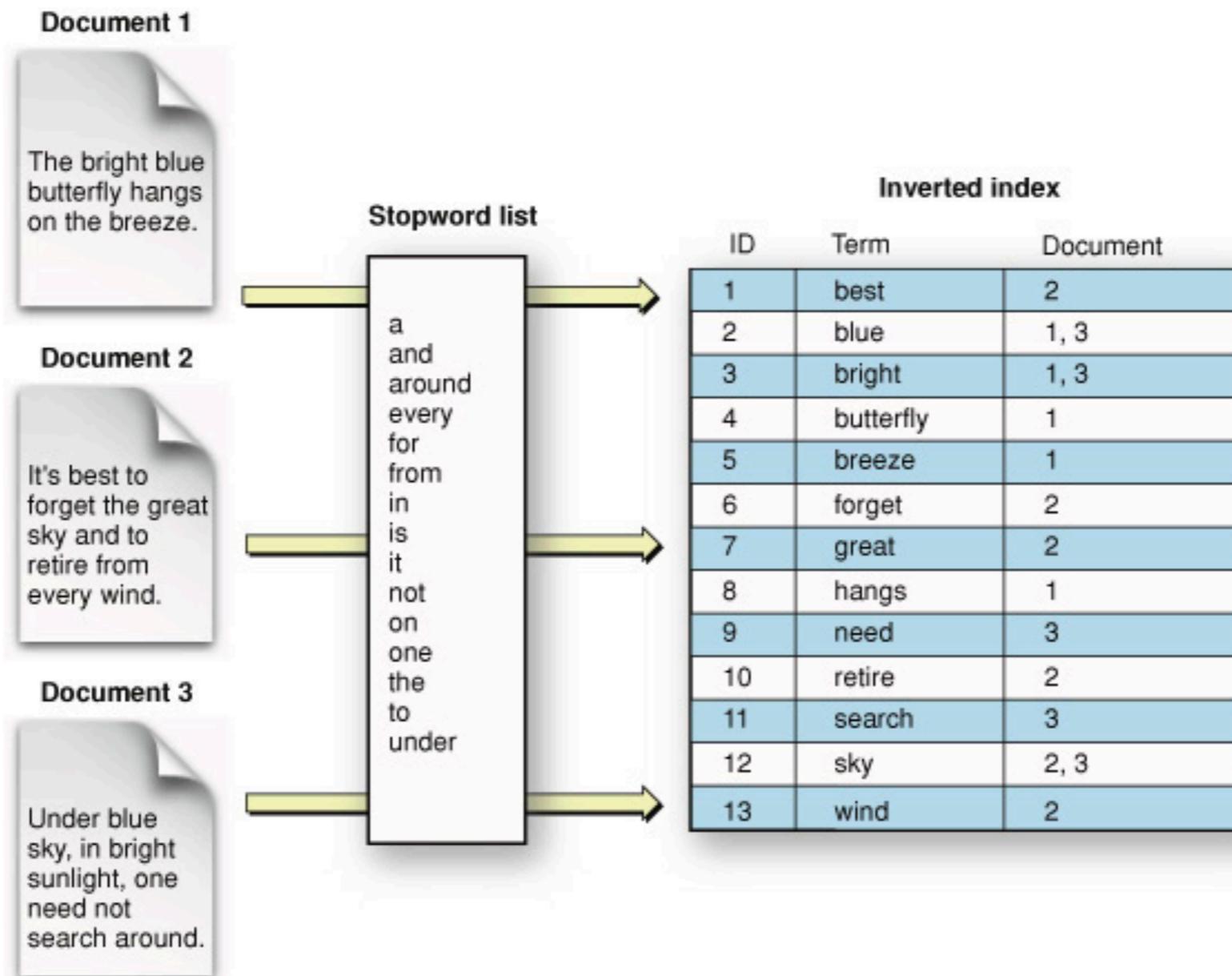


[https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit\\_basics/searchKit\\_basics.html#//apple\\_ref/doc/uid/TP40002843-TPXREF101](https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html#//apple_ref/doc/uid/TP40002843-TPXREF101)

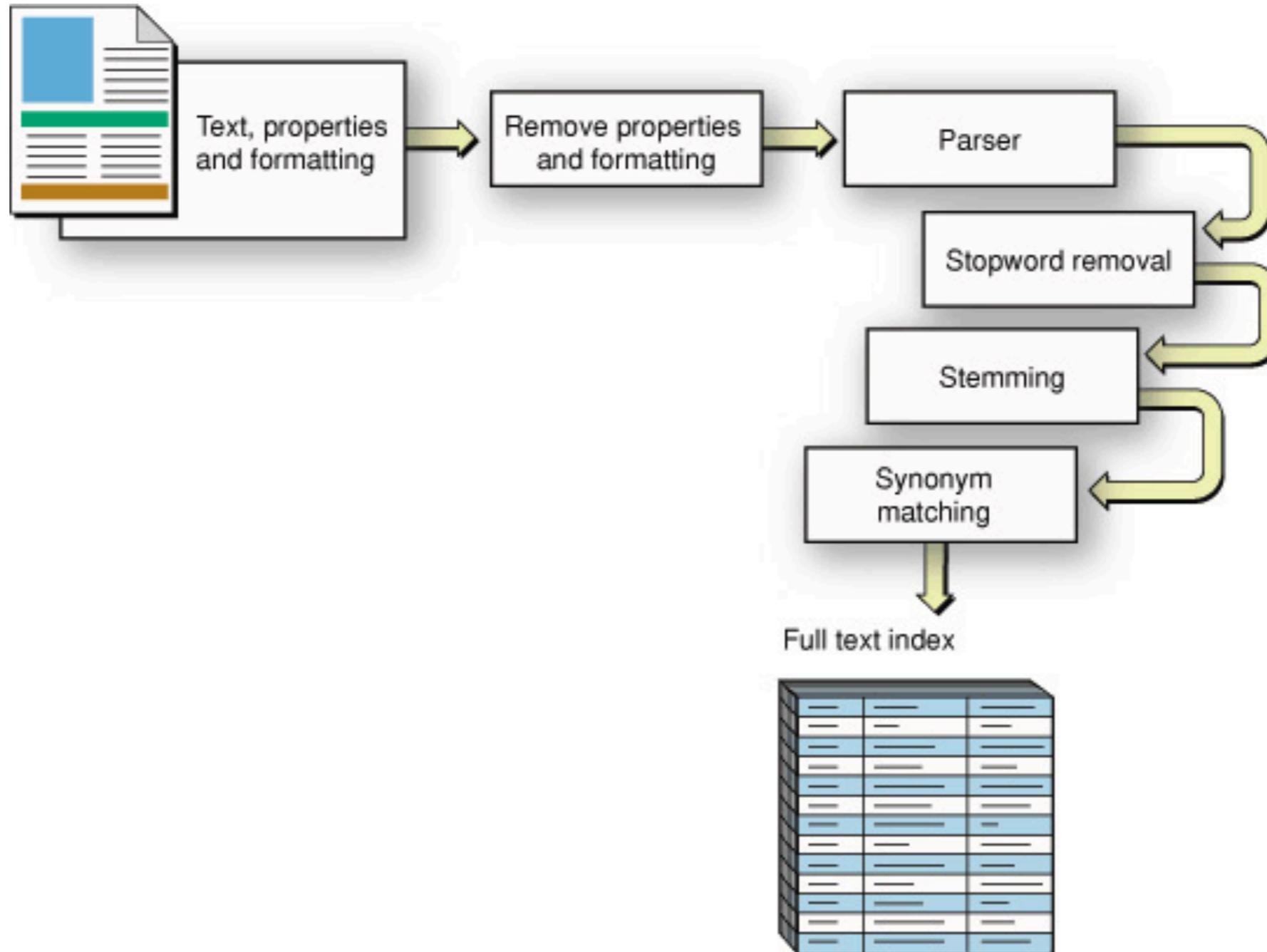


# Inverted Index

## Try to construct index



# Text extraction



# Analyzer in ES

Analyzer  
Tokenizer  
Filter



# Testing analyzer !!

Very important



# Testing analyzer !!

```
POST _analyze
{
  "analyzer": "whitespace",
  "text": "The quick brown fox."
}
```



# Default analyzer !!

```
POST _analyze
{
  "text": "The quick brown fox."
}
```



# Thai analyzer !!

```
POST _analyze
{
  "analyzer": "thai",
  "text": "สวัสดิ์ประเทศไทย"
}
```



# Tokenizer and filter

```
POST _analyze
{
  "tokenizer": "standard",
  "filter": [ "lowercase", "asciifolding" ],
  "text": "Is this déjà vu?"
}
```



# Analyze by index

```
POST my_index/_analyze
{
  "analyzer": "your_analyzer",
  "text": "your text"
}
```



# Analyze by field

```
POST my_index/_analyze
{
  "field": "my_text",
  "text": "your text"
}
```



# Working with Suggester

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



# Suggesters in ES

Term  
Phrase  
Completion  
Context



# Basic knowledge

N-gram tokenizer

Edge-ngram tokenizer

<https://www.elastic.co/guide/en/elasticsearch/reference/7.1/analysis-ngram-tokenizer.html>



# N-Gram

Terms as a sequence of n words

search

Unigram	s,e,a,r,c,h
Bigram	se, ea, ar,rc, ch
Trigram	sea, ear, arc, rch
4-gram	sear, earc, arch
5-gram	searc, earch



# Workshop

## ngram/ngram.json



# More tools



# Elasticsearch Head



## ElasticSearch Head

offered by travistx

★★★★★ (75)

[Developer Tools](#)

45,312 users

ADDED TO CHROME



OVERVIEW

REVIEWS

SUPPORT

RELATED

The screenshot shows the ElasticSearch Head interface. At the top, it displays the cluster name 'Rick' and its health status 'cluster health: yellow (6, 18)'. Below this, there are navigation tabs for 'Overview', 'Browser', 'Structured Query', and 'Any Request'. The main area is titled 'Cluster Overview' and shows a table of nodes with their names, IP addresses, and status indicators. A modal window is open over the 'Rick' node, displaying its configuration details.

Node Name	IP Address	cu_docs	bnvil	cu_msg	anvil
Leon	192.168.7.8:9202	0 1	0 1	0	index: close
Pris	192.168.7.8:9204	0 1	0 1	0	
<b>Rick</b>	192.168.7.8:9200	1 2	0 1	0 1 2 3 4	
Rachel	192.168.7.8:9203	1 2	0 1	0 1 2 3 4	
Zhora	192.168.7.8:9205	0 2	0 1	0 1 2 3 4	
Roy	192.168.7.8:9201	0 2	0 1	0 1 2 3 4	
Unassigned			0 0		

```
{
  name: "Leon",
  transport_address: "inet[/192.168.7.8:9302]",
  attributes: {},
  http_address: "inet[/192.168.7.8:9202]",
  os: {
    refresh_interval: 5000,
    cpu: {
      vendor: "Intel",
      model: "Macmini4,1",
      mhz: 2400,
      total_cores: 2,
      total_sockets: 1,
      cores_per_socket: 2,
      cache_size: "3kb",
      cache_size_in_bytes: 3072
    }
  }
}
```

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>



# Elasticsearch Dump



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



# Make Logs

Simple generator used to push fake HTTP traffic logs into elasticsearch

```
npm install -g @elastic/makelogs
```

<https://github.com/elastic/makelogs>



# 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>



# Geo-point Format

Type	Format
Object	lat = lon =
String	lat, lon
Array	[lon, lat] <b>** GeoJSON **</b>

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



# 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/>



# Geo-point query

Geo-bounding-box

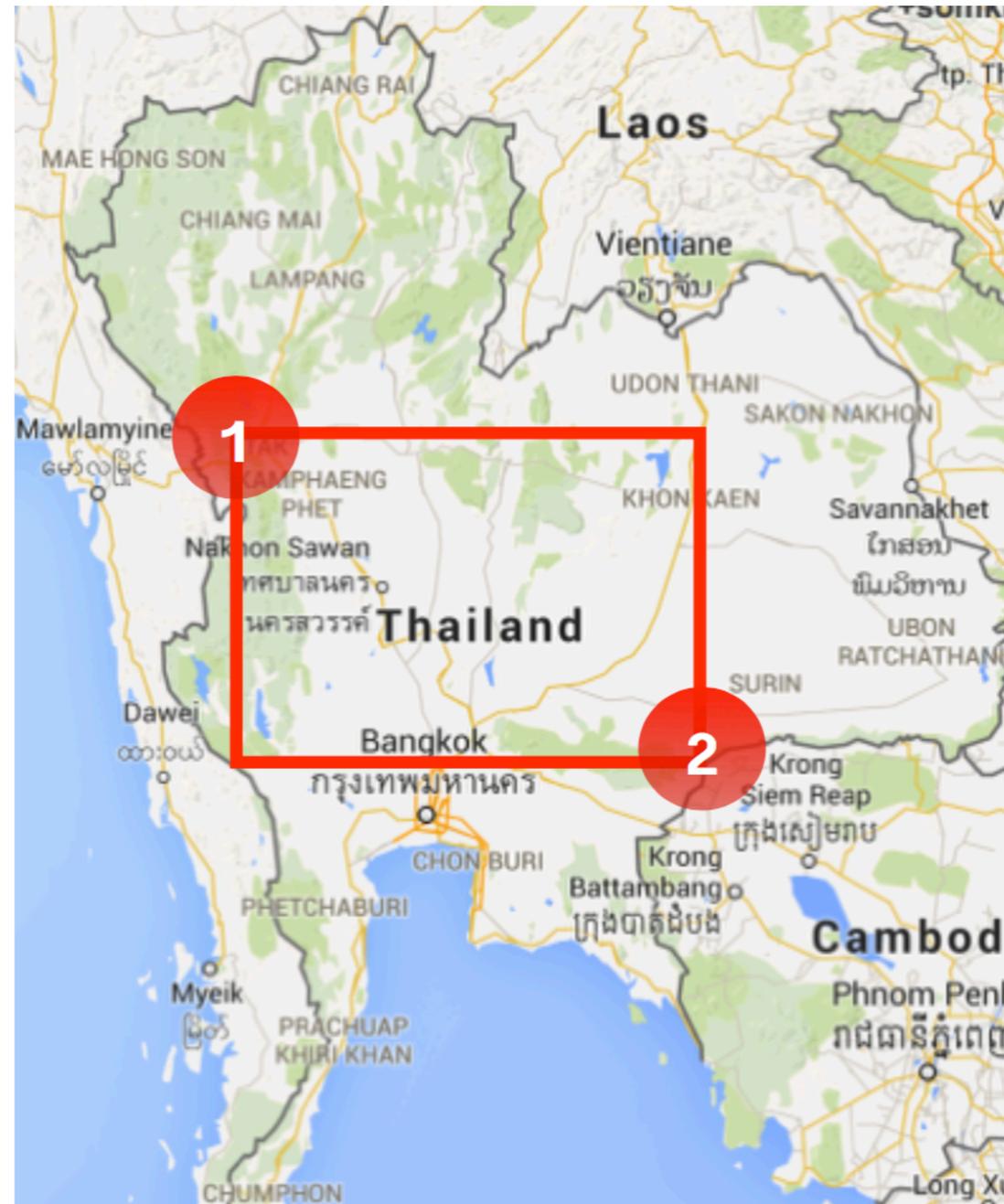
Geo-distance

Geo-polygon

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



# Bounding Box



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



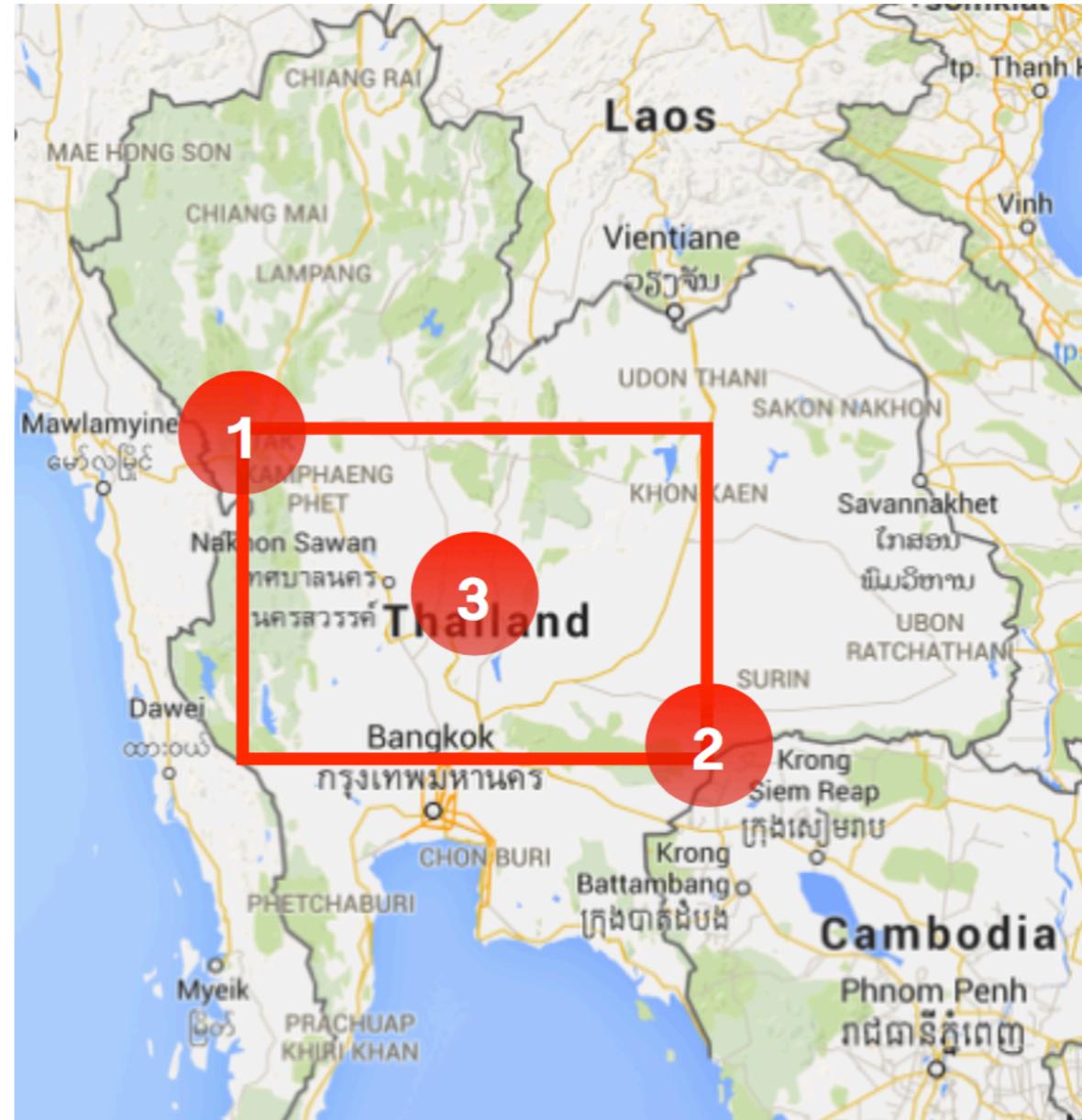
# Geo Distance



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



# 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>



# 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>



# 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>



# Working with Data

`$elasticsearch-plugin install ingest-geoip`

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



# GeoIP with Elasticsearch

geoip/instruction.json



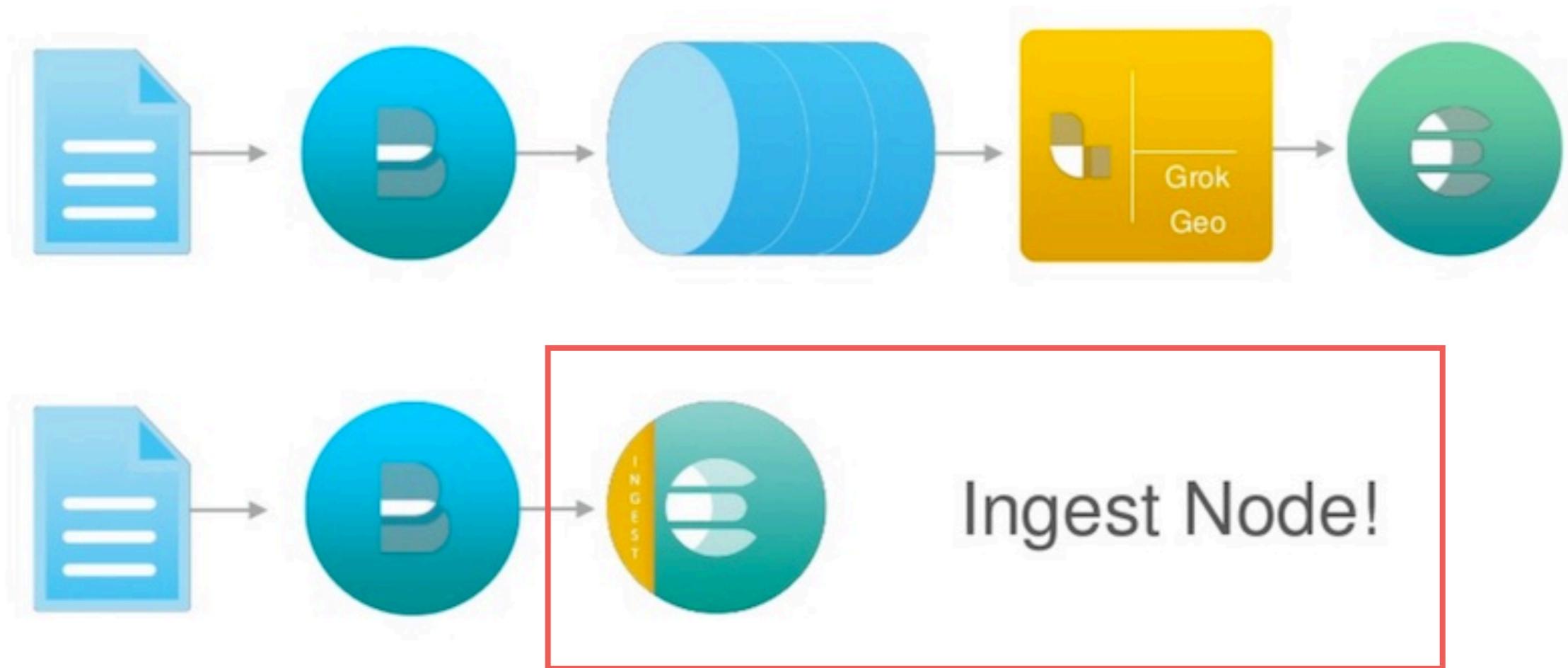
# Sample Data

```
{ "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>



# Install plugin

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

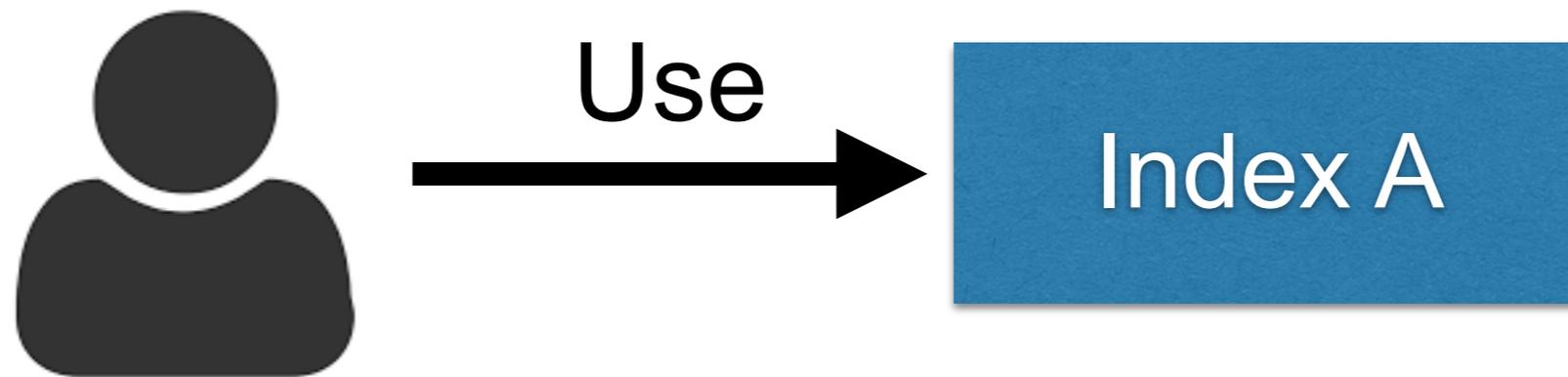
<https://www.elastic.co/guide/en/elasticsearch/plugins/current/ingest-geoip.html>



# Alias Index



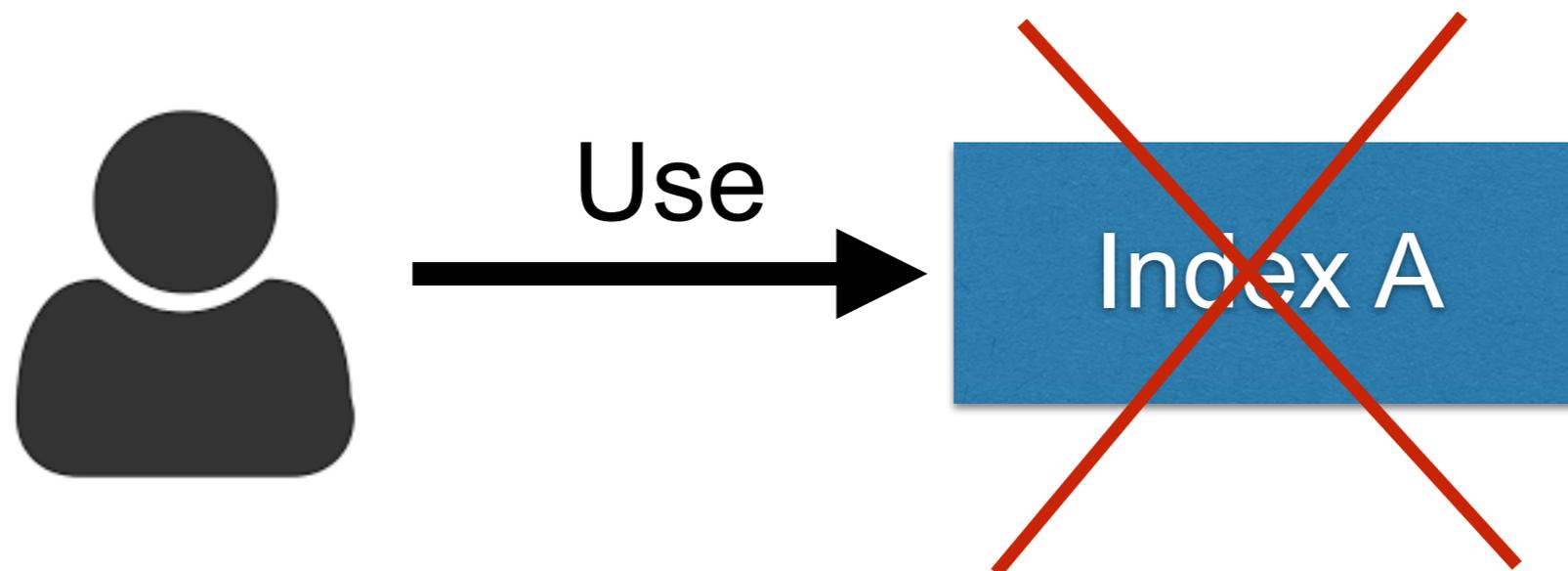
# Common usage



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



# Problem ?



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



# Using Alias Index



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



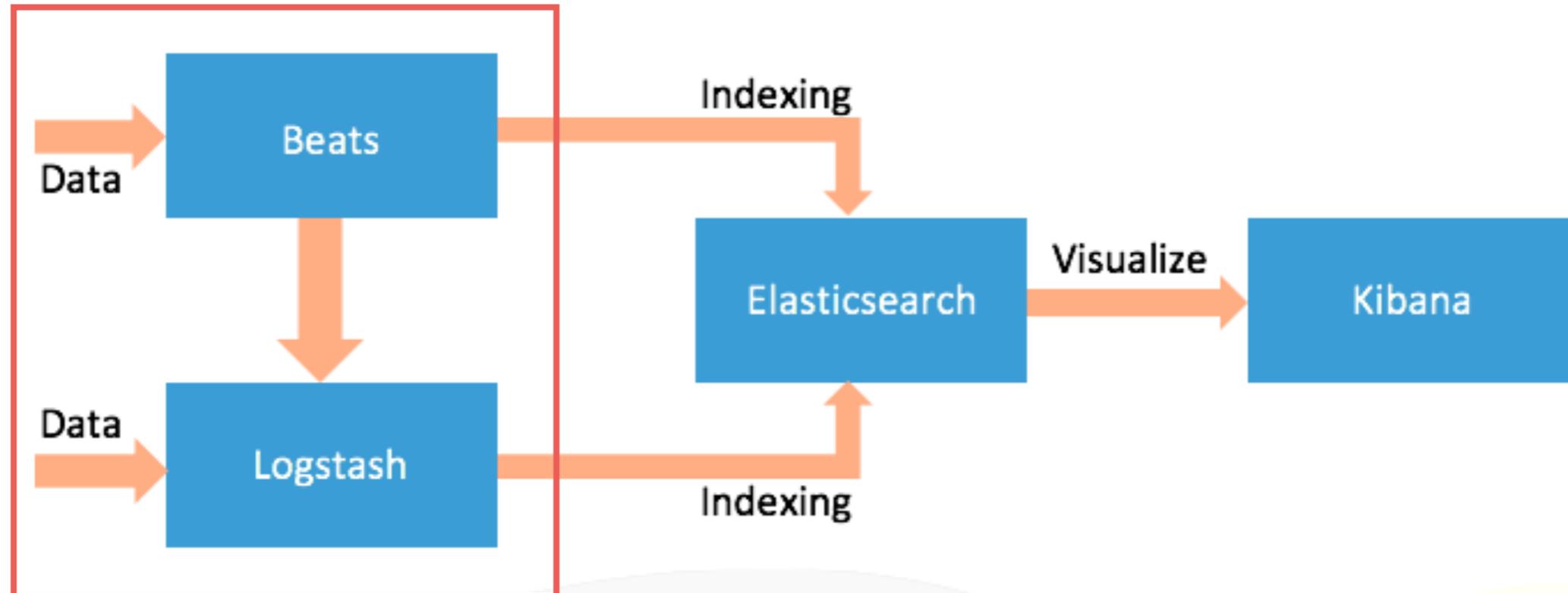
# Using Alias Index



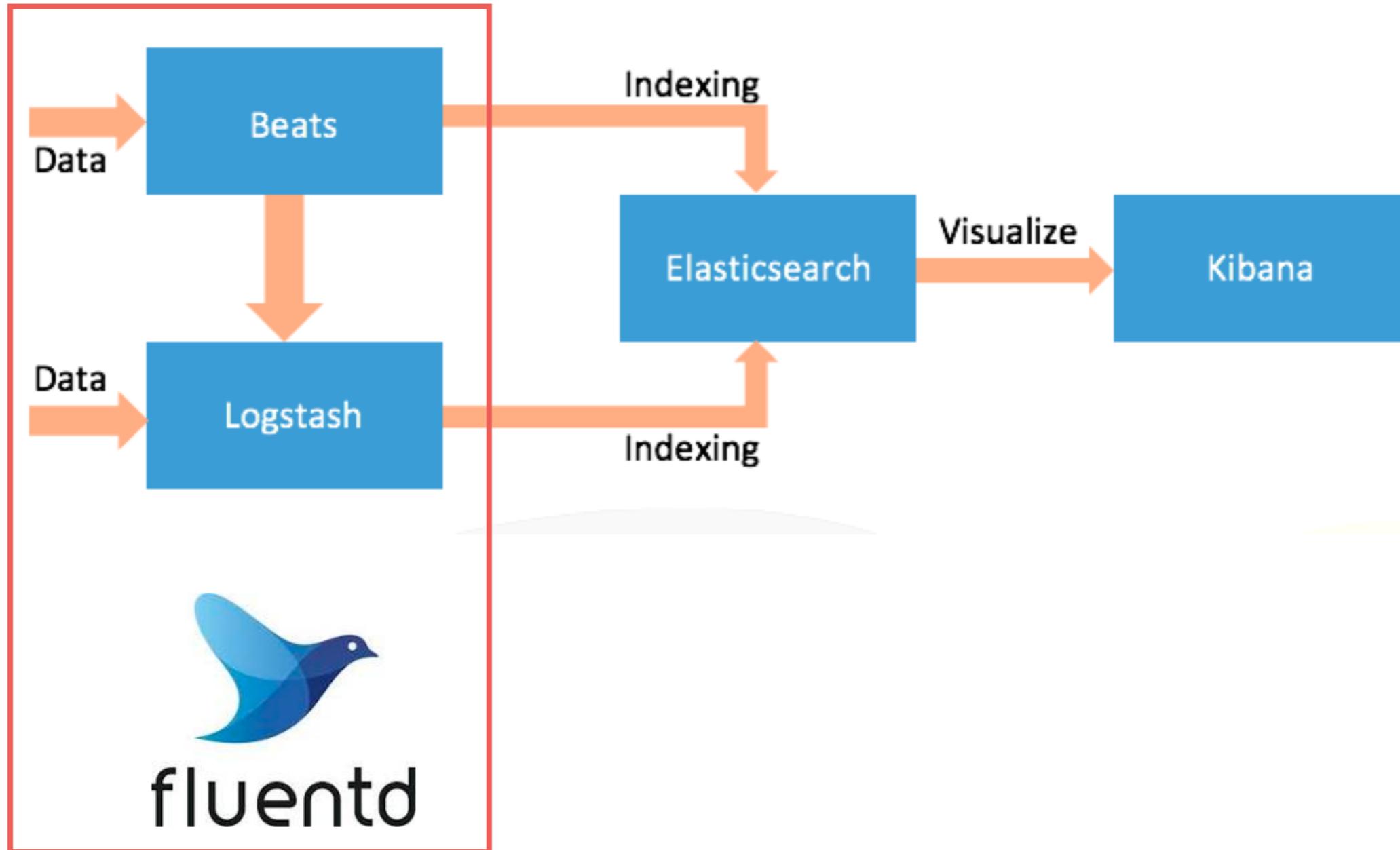
<https://www.elastic.co/guide/en/elasticsearch/reference/current/indices-aliases.html>



# ELK stack



# EFK stack



# Working with Logstash

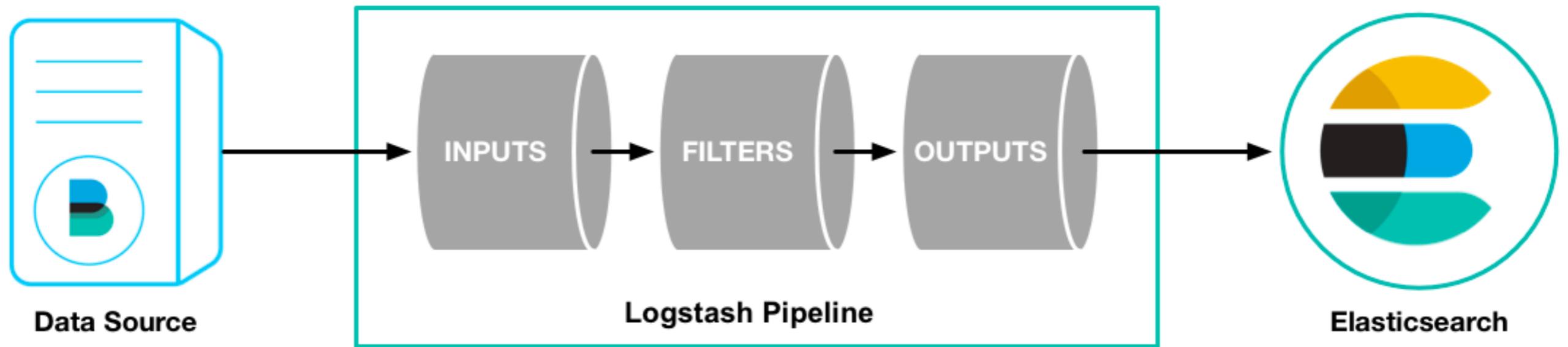
<https://www.elastic.co/guide/en/logstash/current/index.html>



# Logstash



# Logstash



# Example



Standard Input

Standard Output



# sample.conf

```
input {  
  stdin{}  
}
```

```
output {  
  stdout {  
    codec => rubydebug  
  }  
}
```



# Example

\$logstash -f sample.conf

```
hello world
{
  "message" => "hello world",
  "@timestamp" => 2019-06-20T06:01:30.048Z,
  "@version" => "1",
  "host" => "Somkiats-MacBook-Pro"
}
```



# Change output to ES



Standard Input

Standard Output  
Elasticsearch

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



# sample.conf

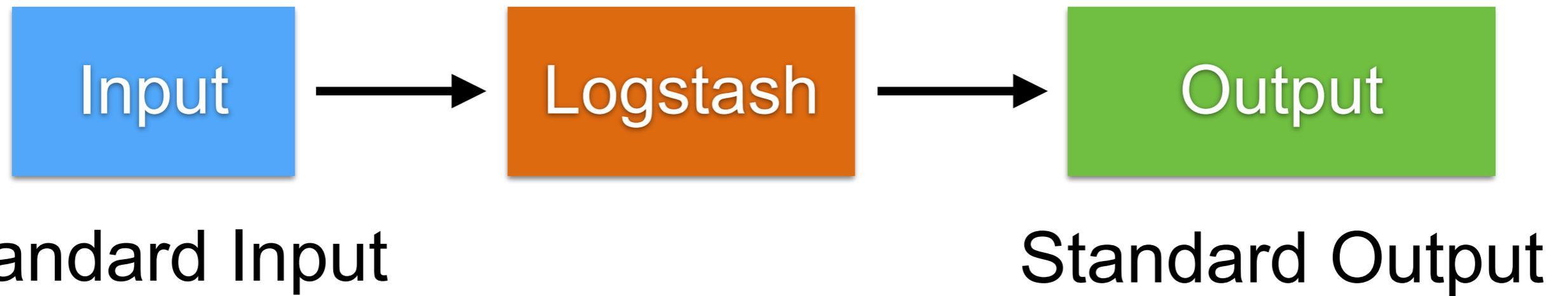
```
input {  
  stdin{}  
}
```

```
output {  
  stdout {  
    codec => rubydebug  
  }  
}
```

```
elasticsearch {  
}  
}
```



# Working with Filter



<https://www.elastic.co/guide/en/logstash/current/filter-plugins.html>



# sample.conf

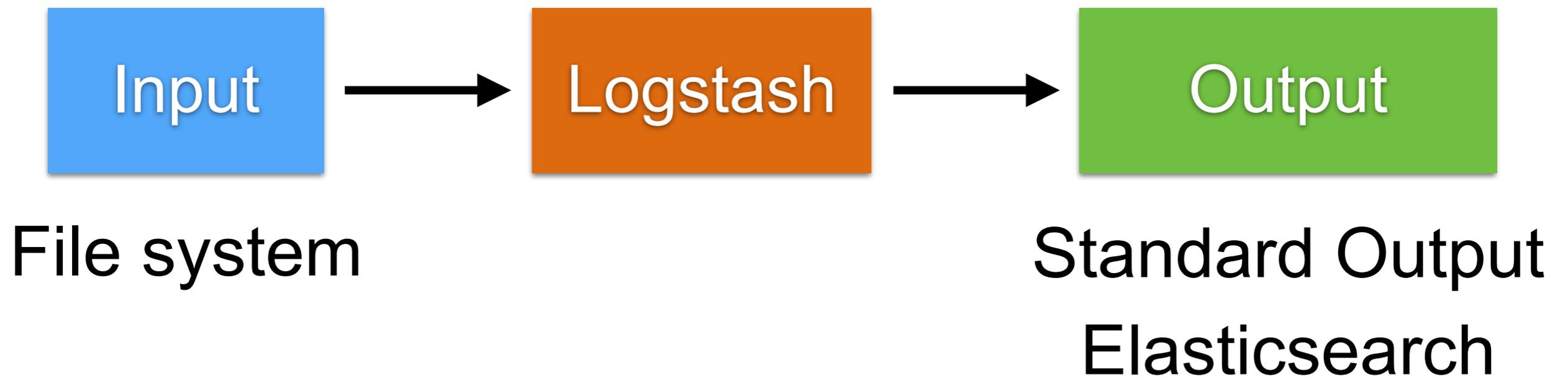
```
input {  
  stdin{}  
}
```

```
filter {  
  grok {  
    match => { "message" => "%{WORD:firstname} %  
{WORD:lastname}" }  
  }  
}
```

```
output {  
  stdout {  
    codec => rubydebug  
  }  
}
```



# Workshop



`workshop/logstash-beat-fluentd/demo.conf`



# try.conf

```
input {  
  stdin{}  
}
```

```
output {  
  stdout {  
    codec => rubydebug  
  }  
}
```

```
elasticsearch {  
}  
}
```



# Design your input first !!

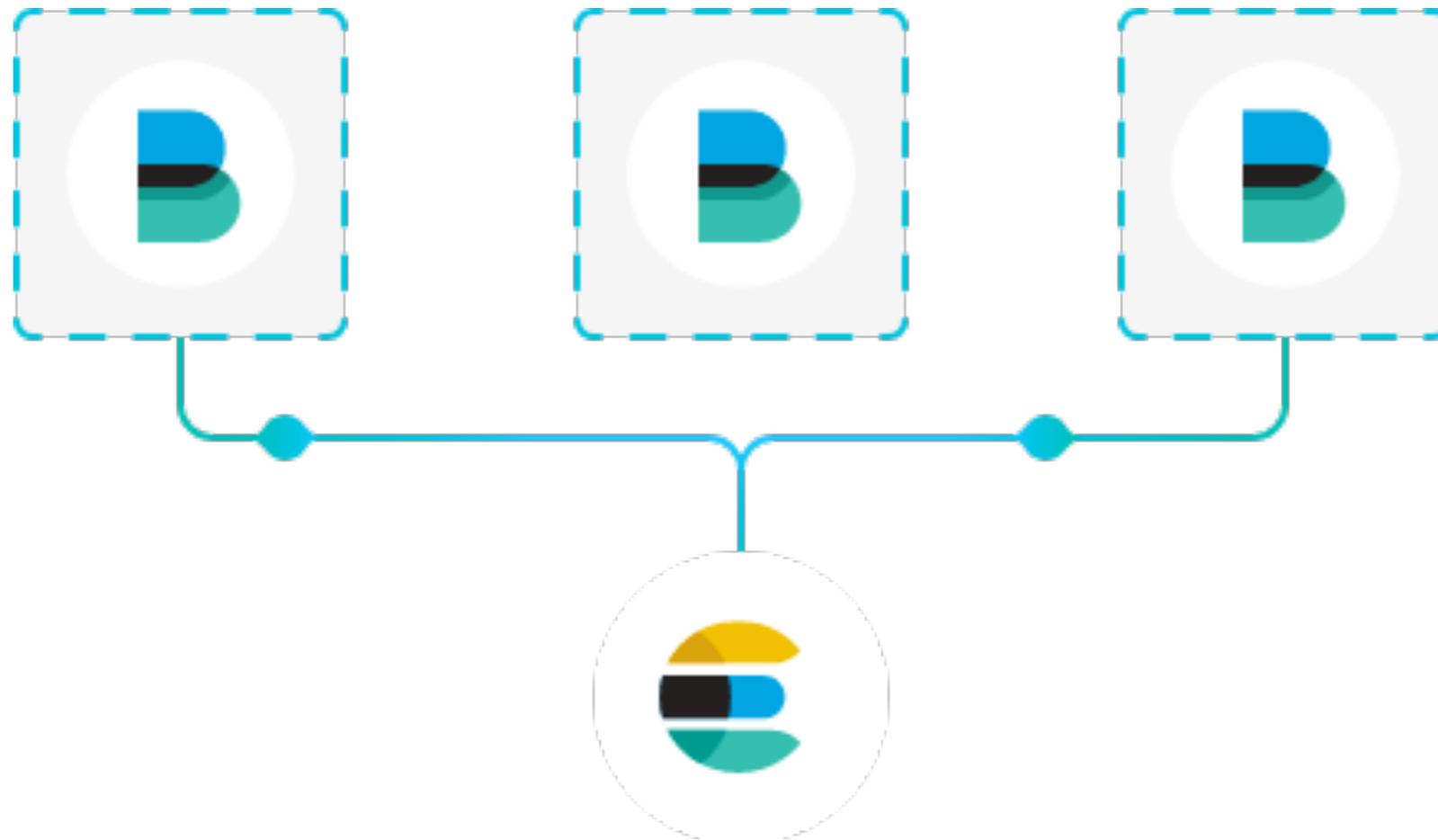


# Use beats is better

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



# Beat



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



# Beat



**Filebeat**

Log Files



**Metricbeat**

Metrics



**Packetbeat**

Network Data



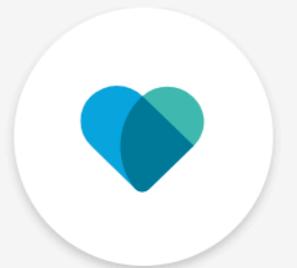
**Winlogbeat**

Windows Event Logs



**Auditbeat**

Audit Data

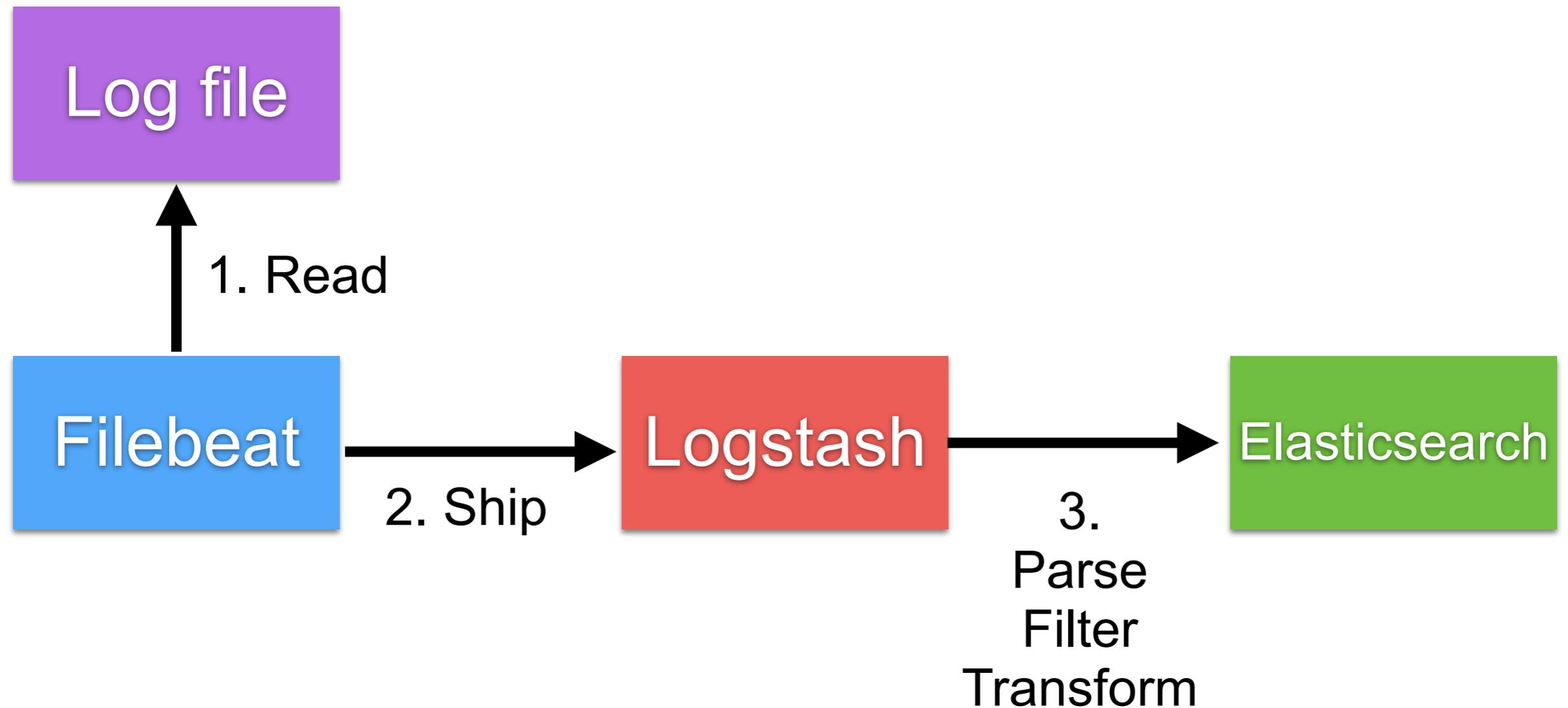


**Heartbeat**

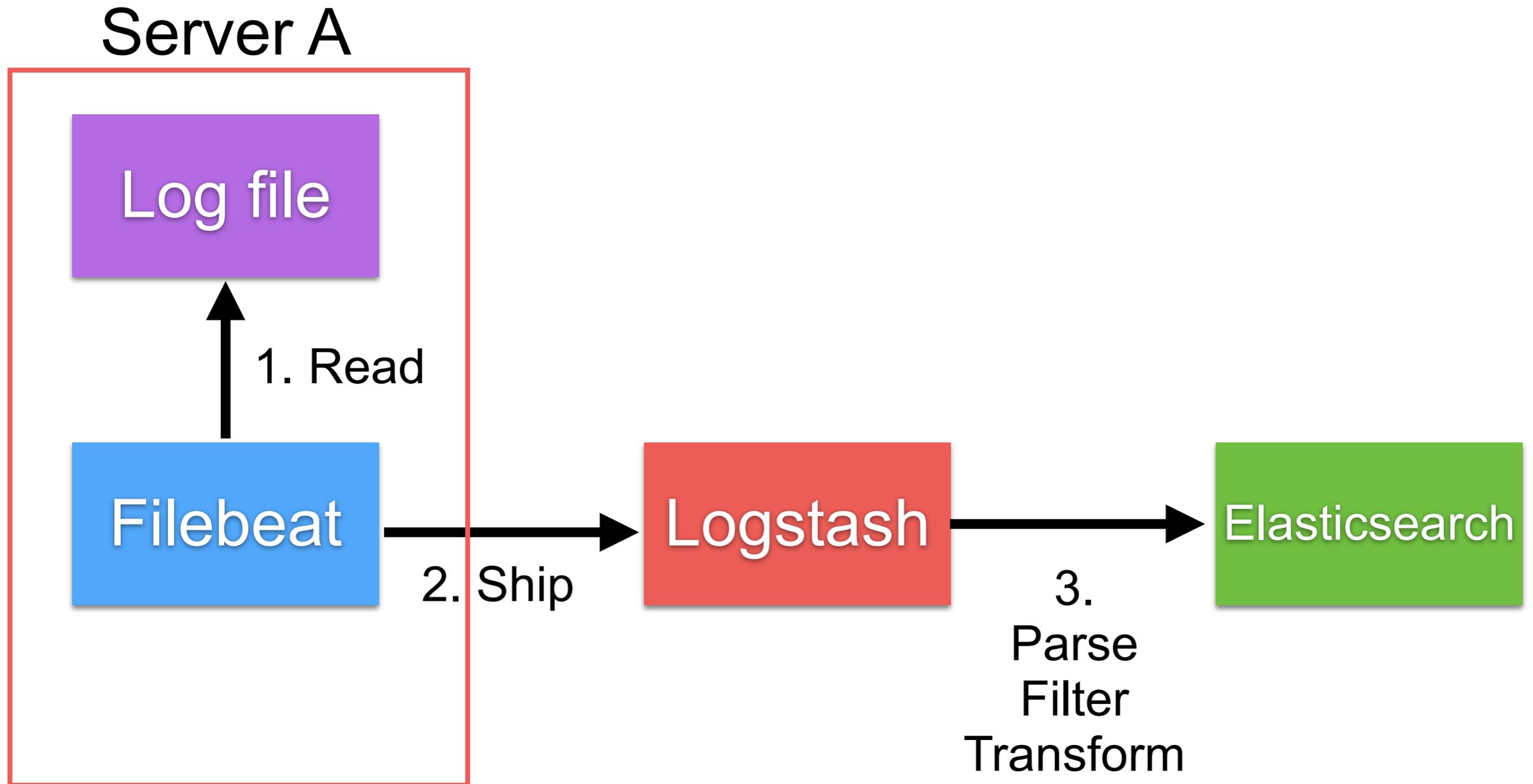
Uptime Monitoring



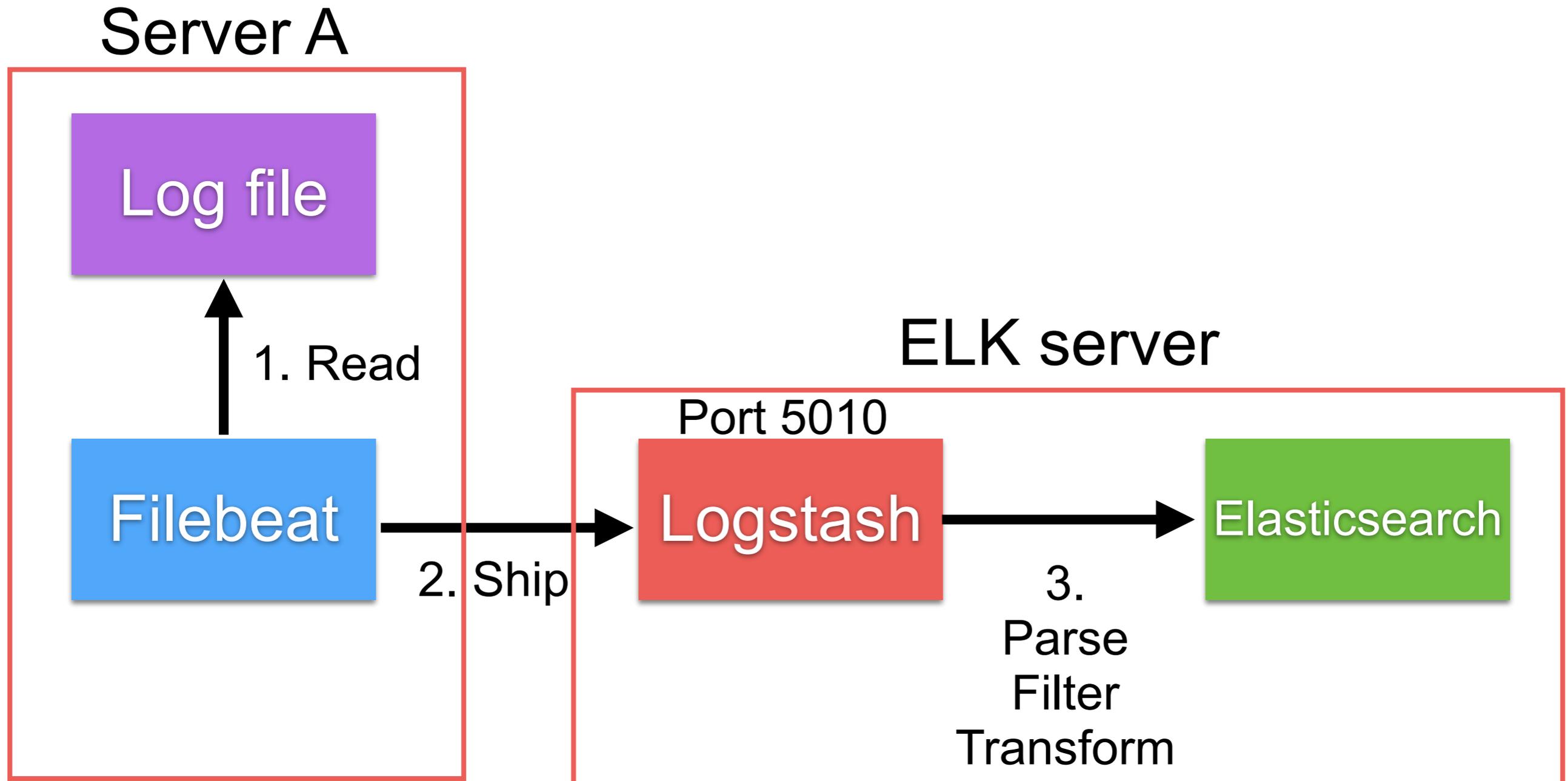
# Example of filebeat



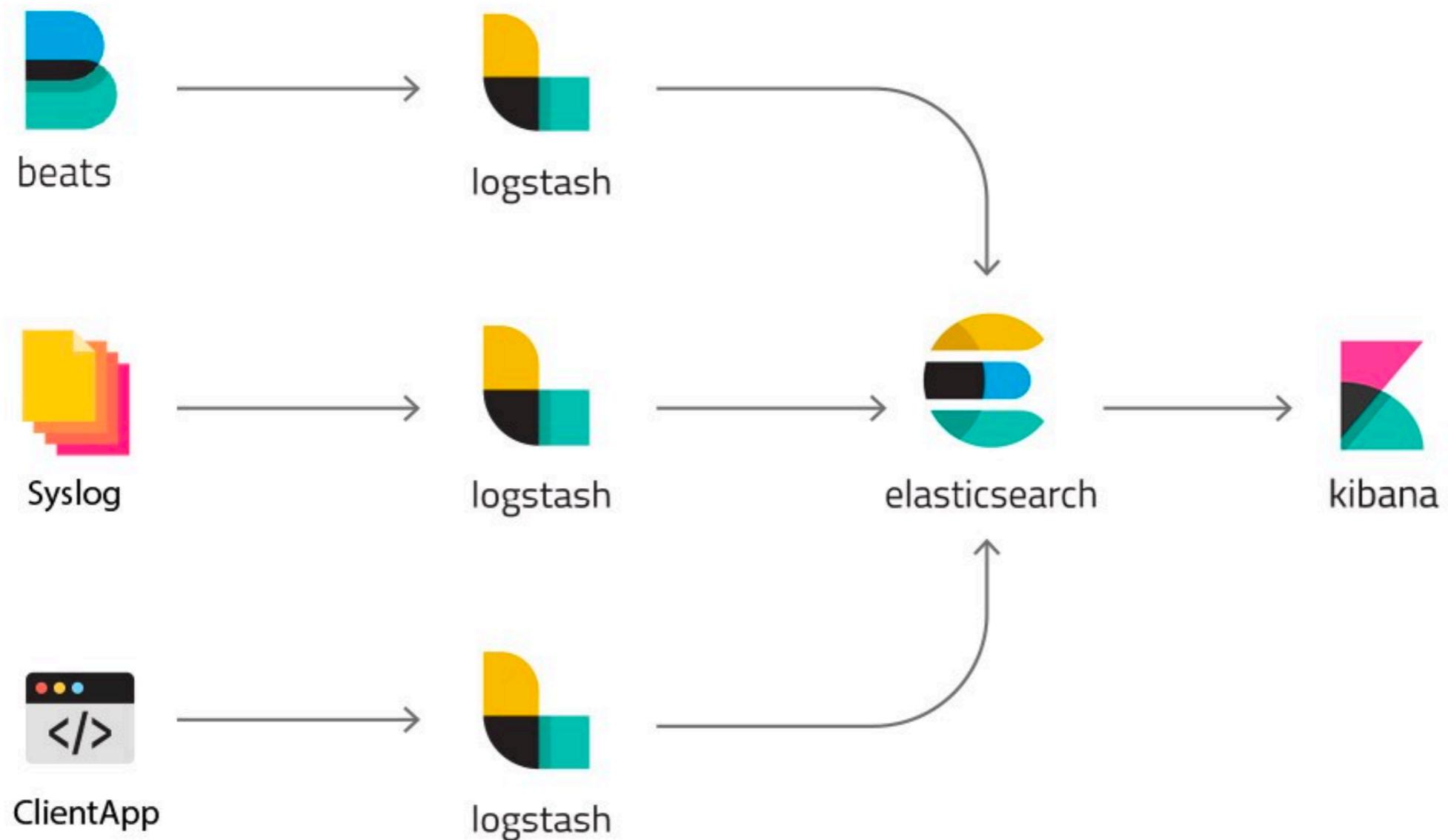
# Example of filebeat



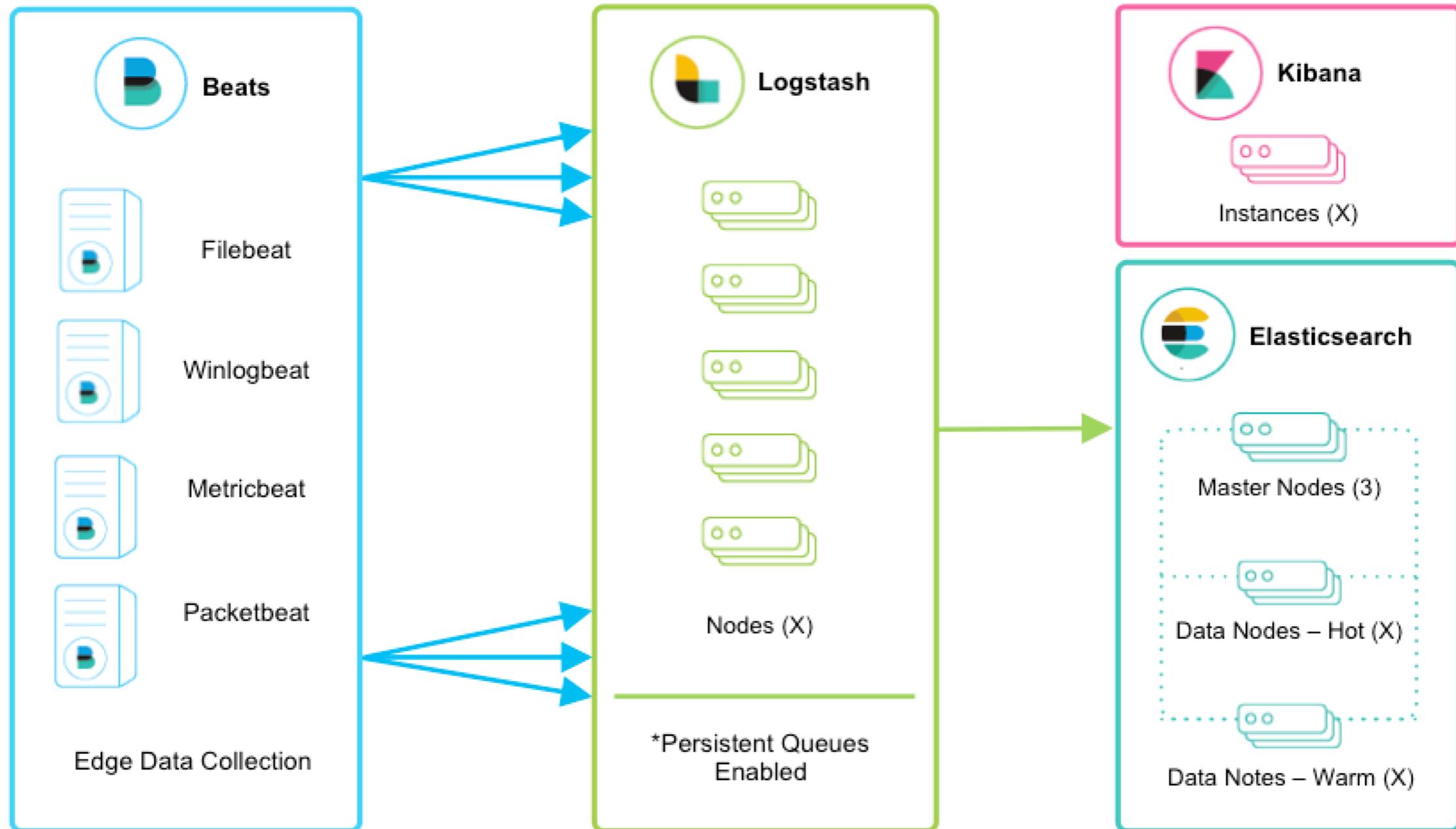
# Example of filebeat



# Beat and Logstash



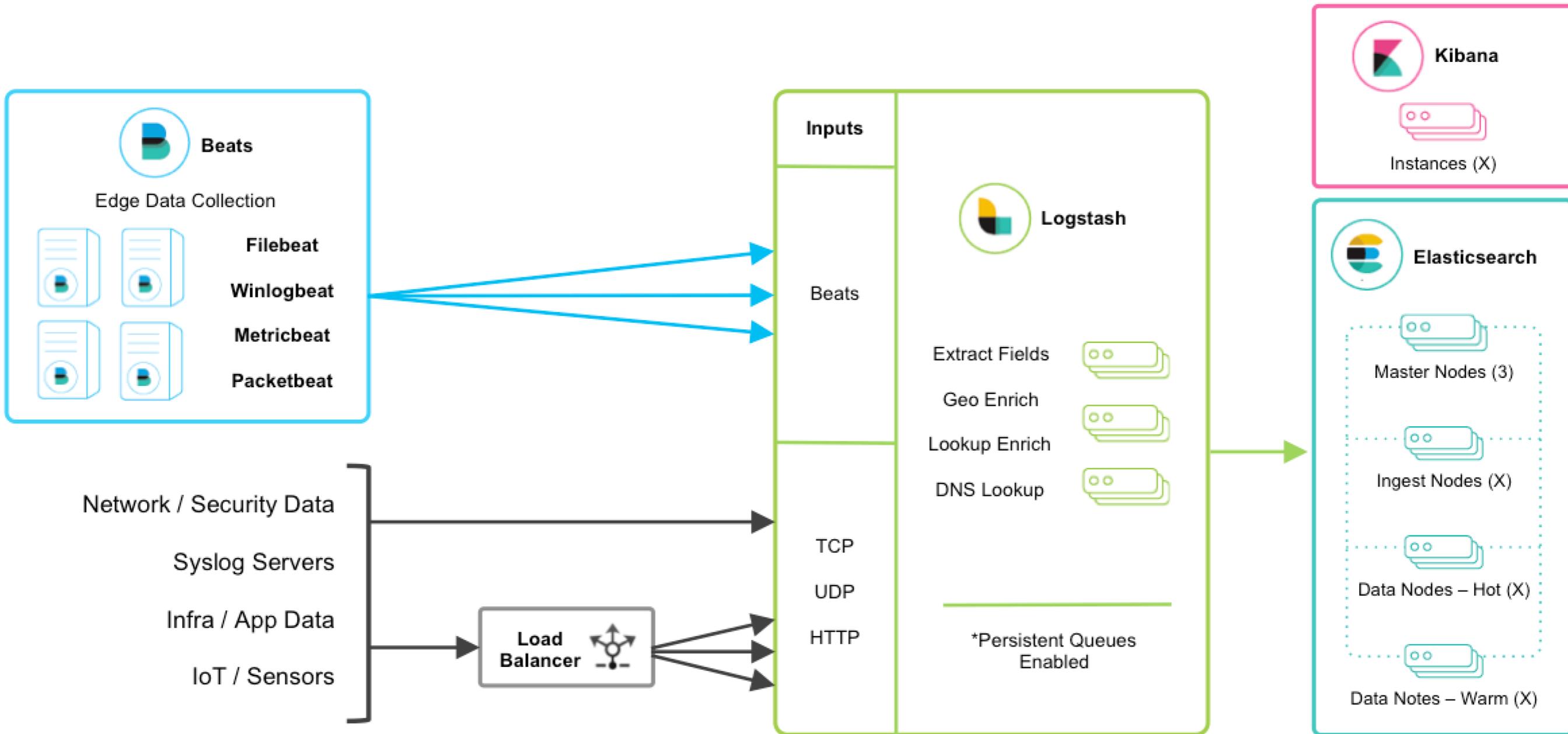
# Scaling



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



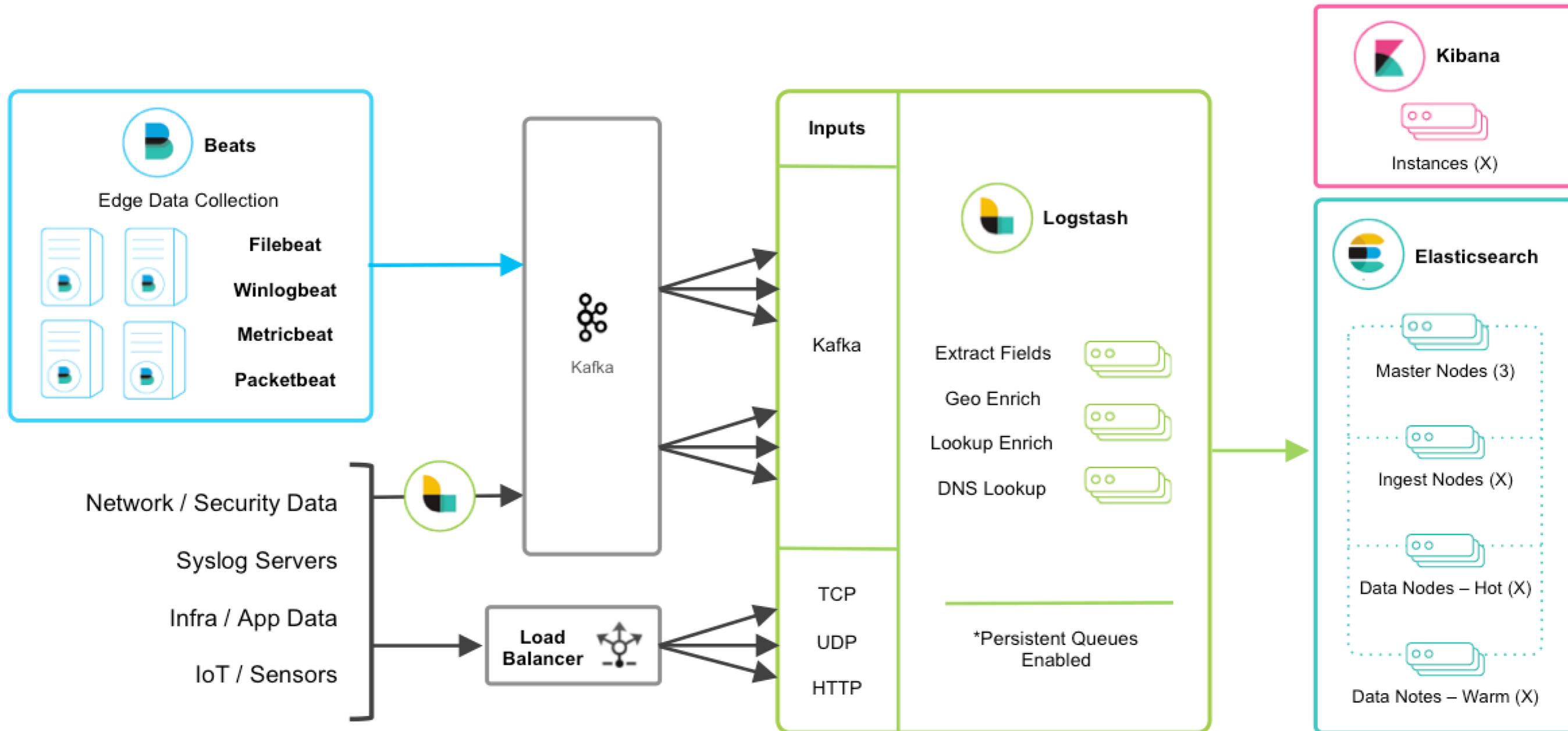
# More data sources



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



# Use messaging Queue



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

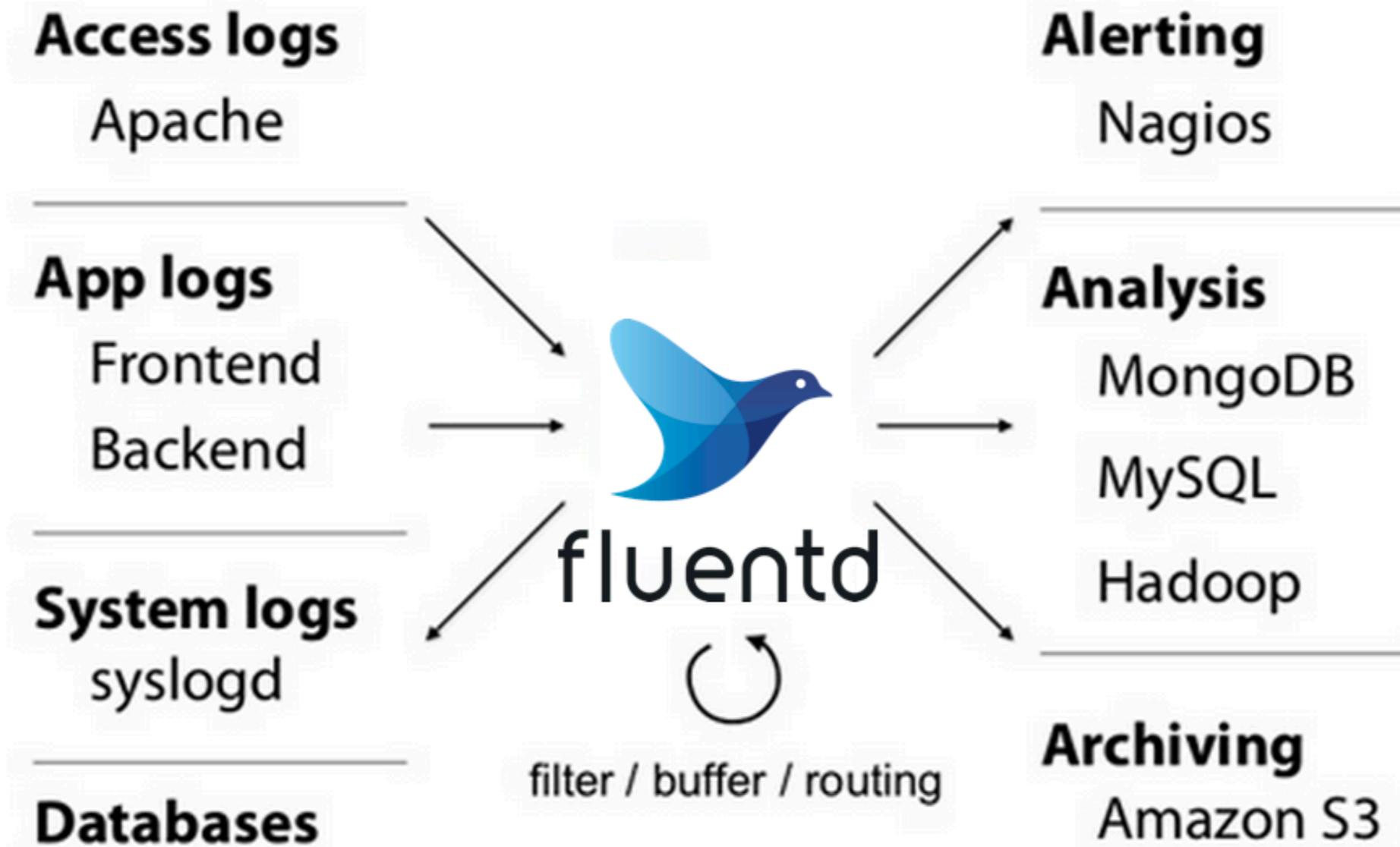


# Working with Fluentd

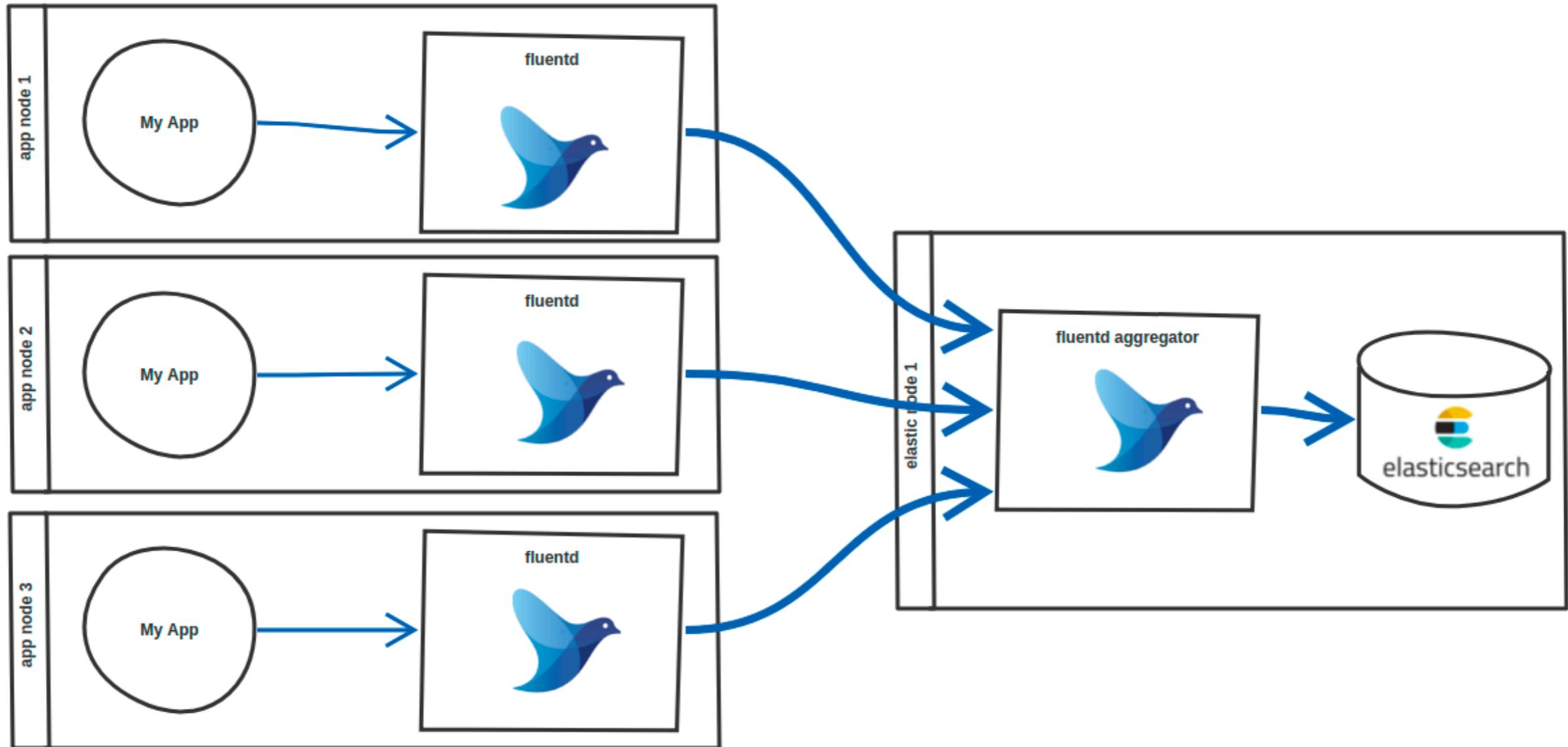
<https://www.fluentd.org/>



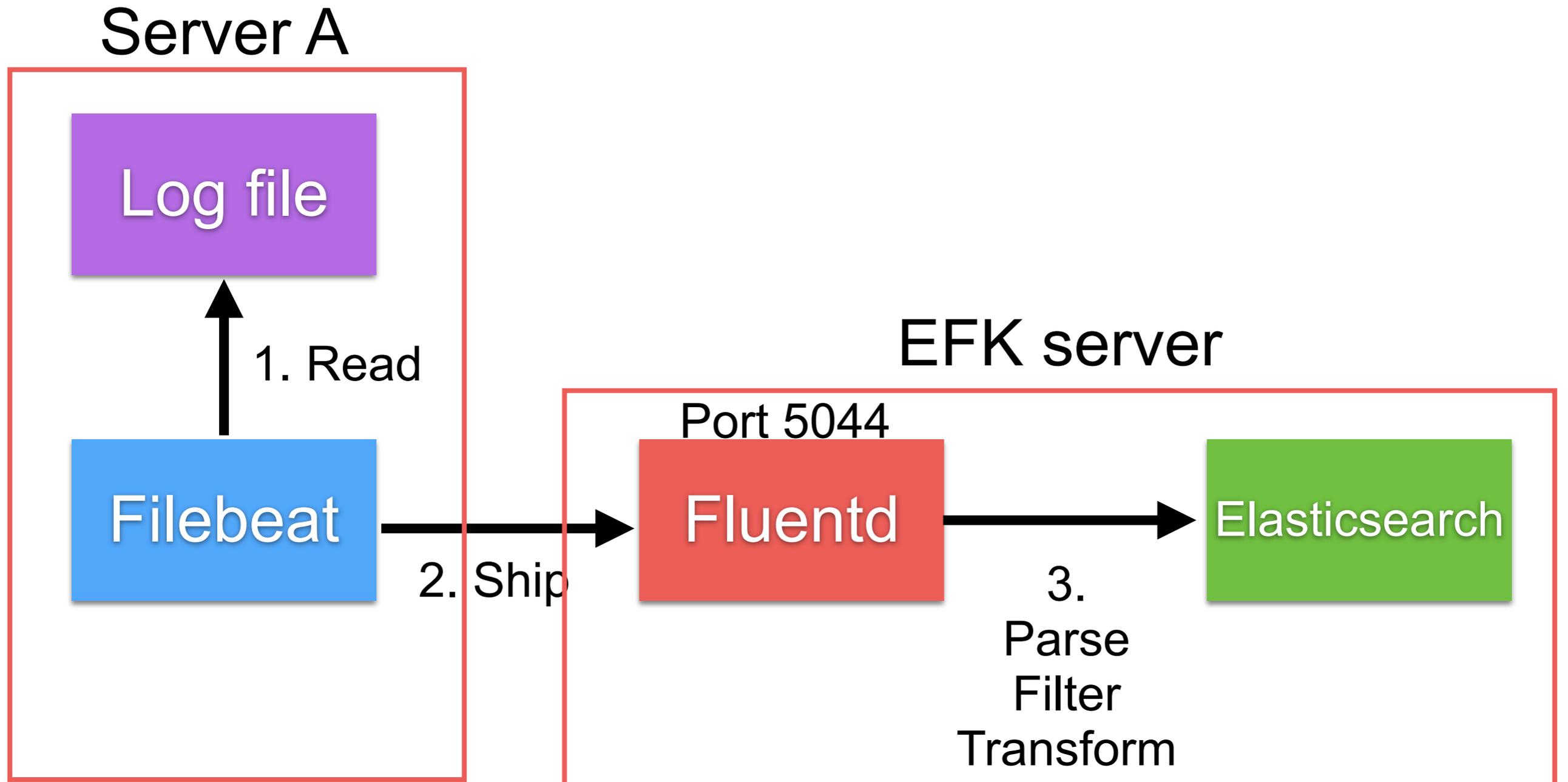
# Fluentd



# EFK stack



# Example of fluentd



# Fluentd regex editor

**Fluentular** fluentd v1.5.2

a Fluentd regular expression editor

</> Regular Expression

🗨 Test String

🕒 Custom Time Format (See also ruby document; [strptime](#))

[Parse](#)

**Example (Apache)**

Regular expression:

```
^(?<host>[ ^ ]*) [ ^ ]* (?<user>[ ^ ]*) \[(?<time>[ ^\ ]*)\] "(?<method>\S+)(?: +(?<path>[ ^ ]*+\S*)?" (?<code>[ ^ ]*) (?<size>[ ^ ]*)(?: "(?<referer>[ ^\" ]*)" "(?<agent>[ ^\" ]*)" )?&#36;
```

Time Format:

```
%d/%b/%Y:%H:%M:%S %z
```

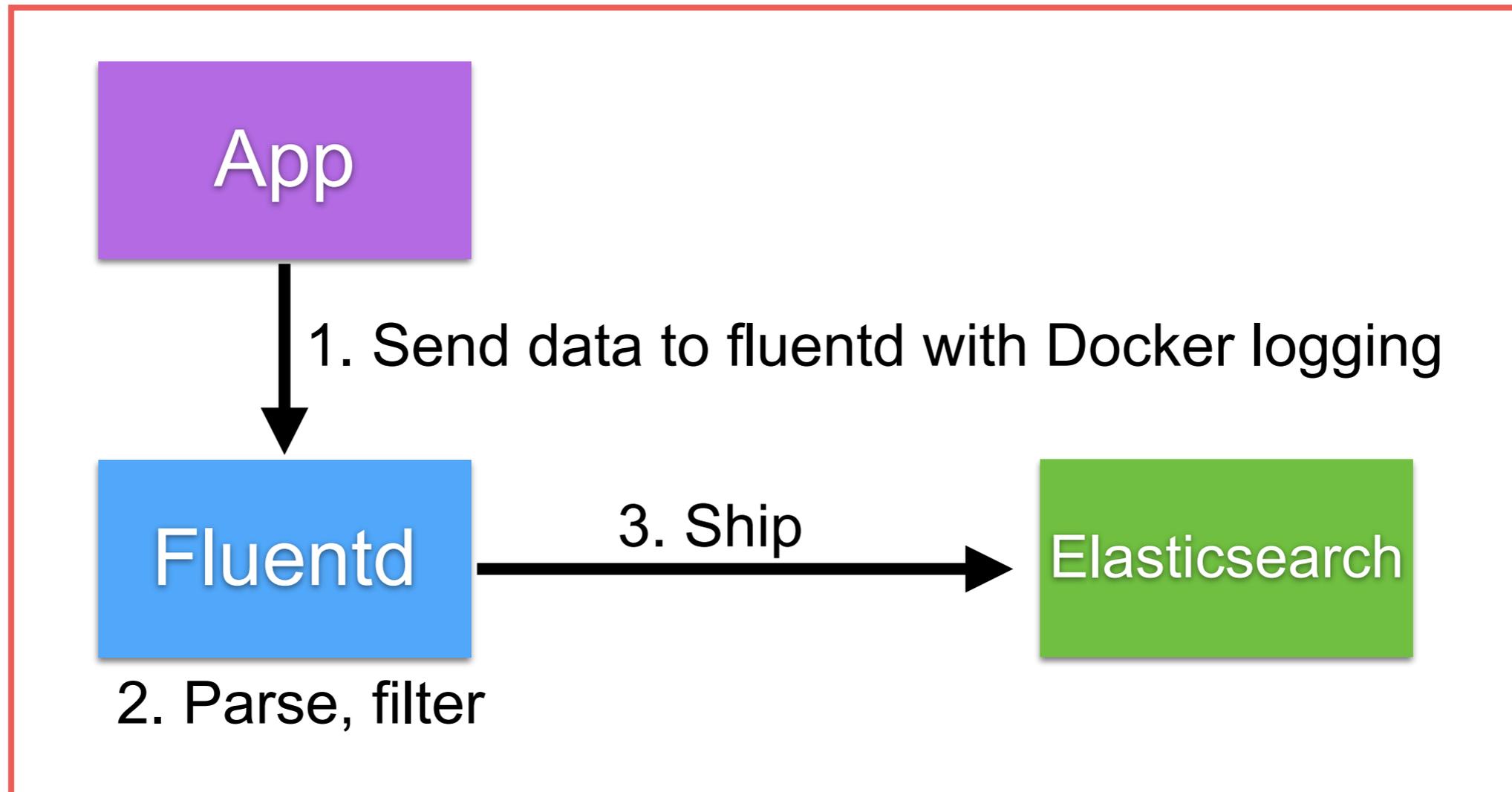
<http://fluentular.herokuapp.com/>



# Fluentd with Docker

## Using docker-compose

Host



# Design for Failure

09-cluster



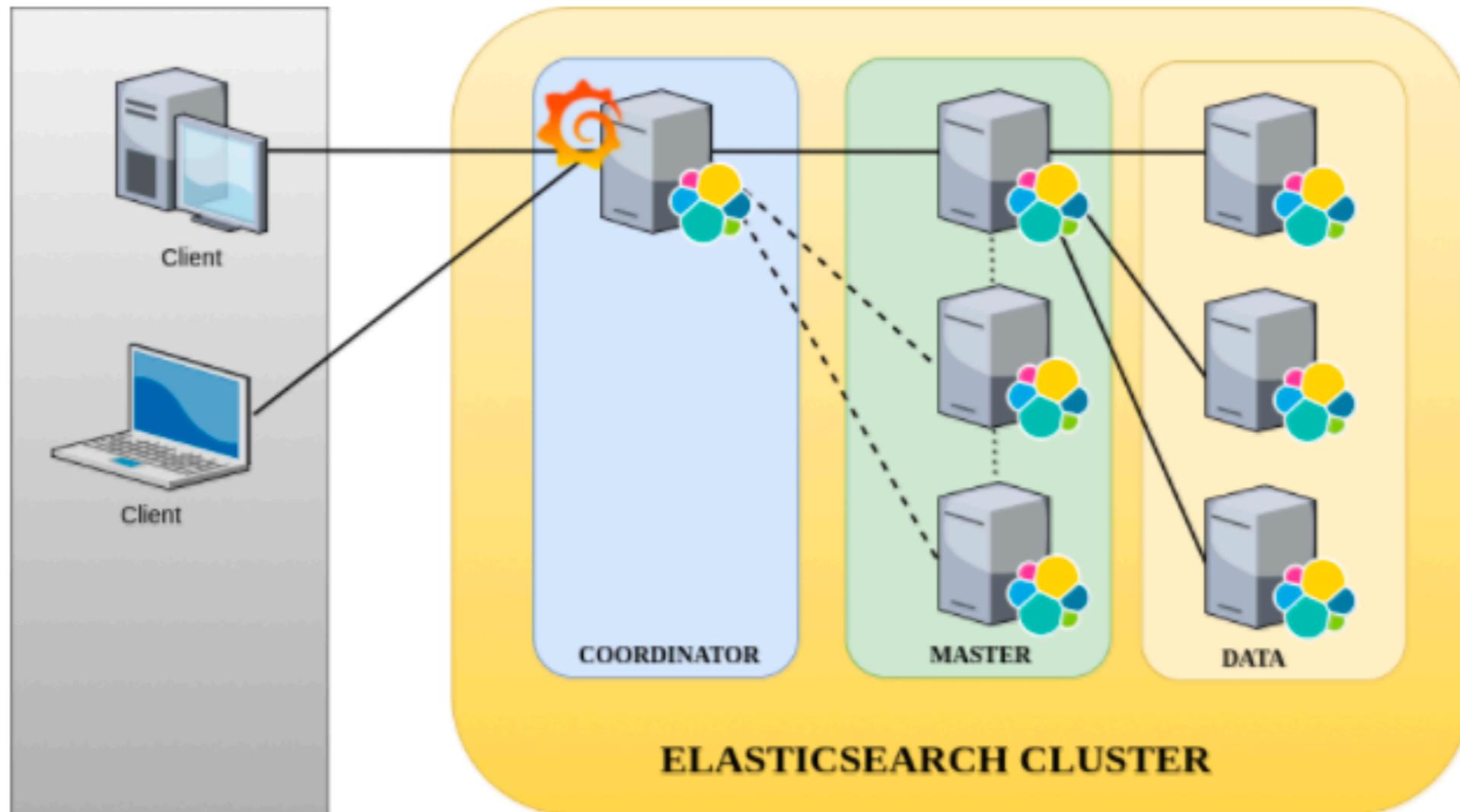
# Elasticsearch Nodes

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
Machine Learning	Required subscription !!

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



# Elasticsearch Cluster



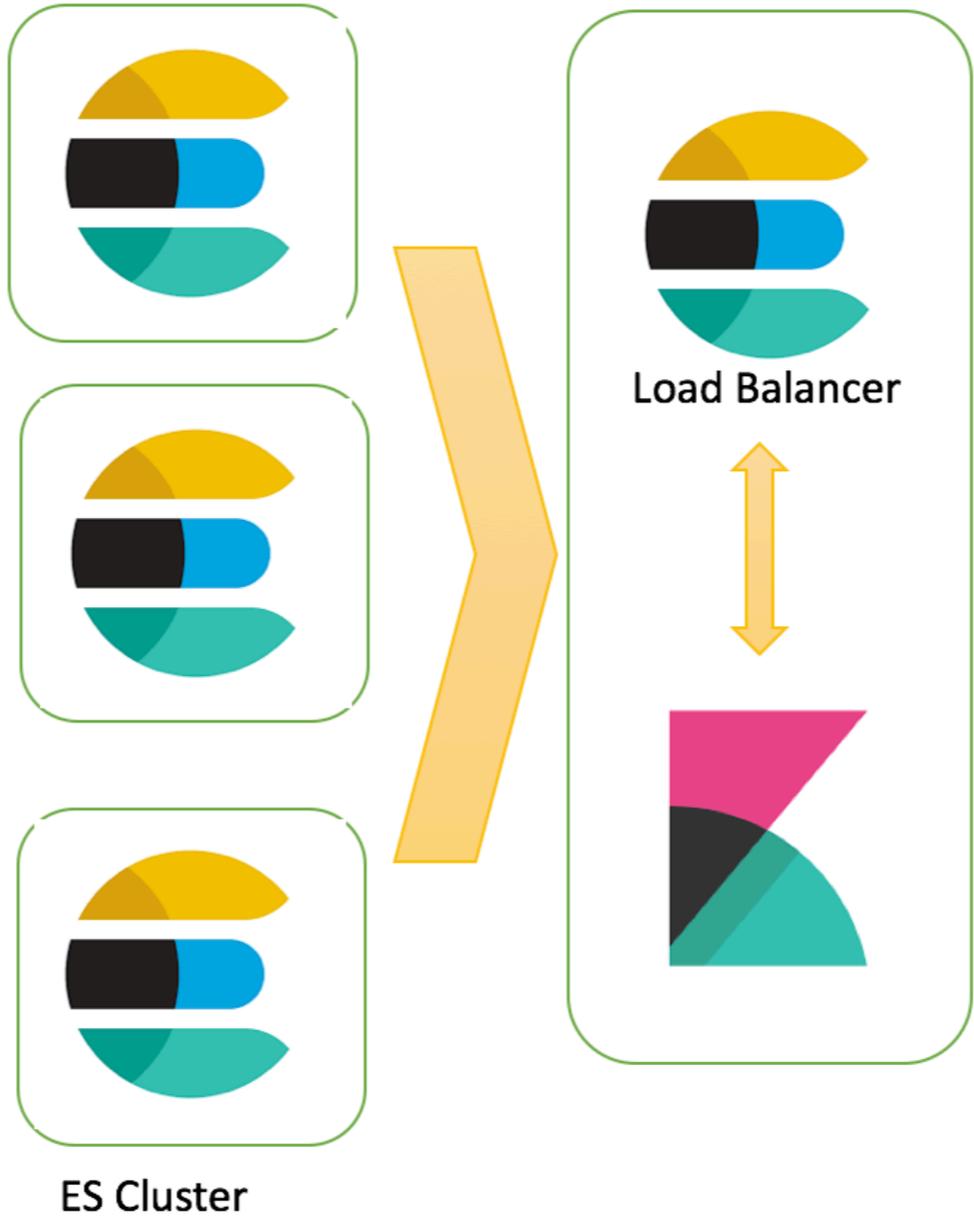
# Elasticsearch Nodes

← → ↻ ⓘ 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>



# Elasticsearch Nodes

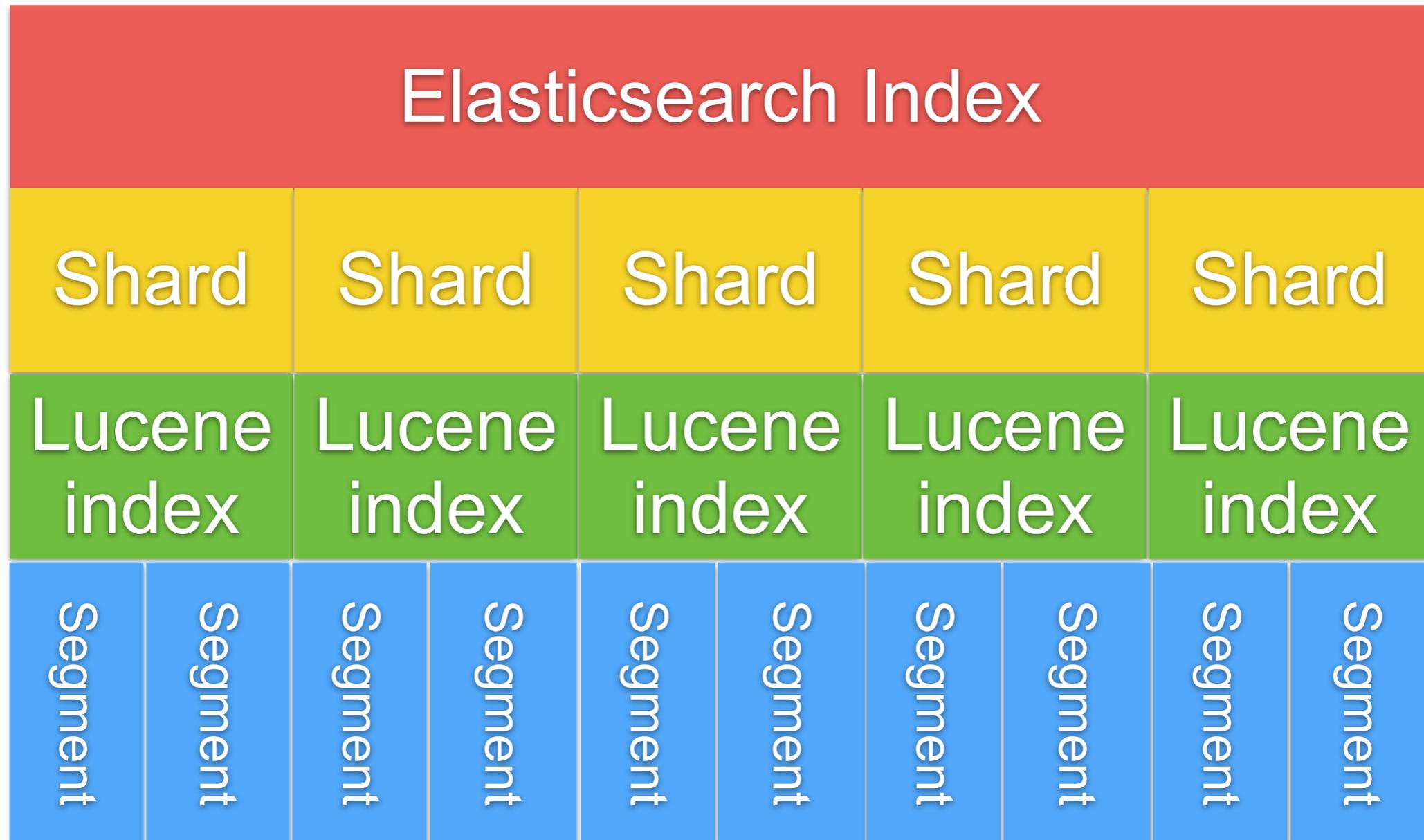


# Apache Lucene

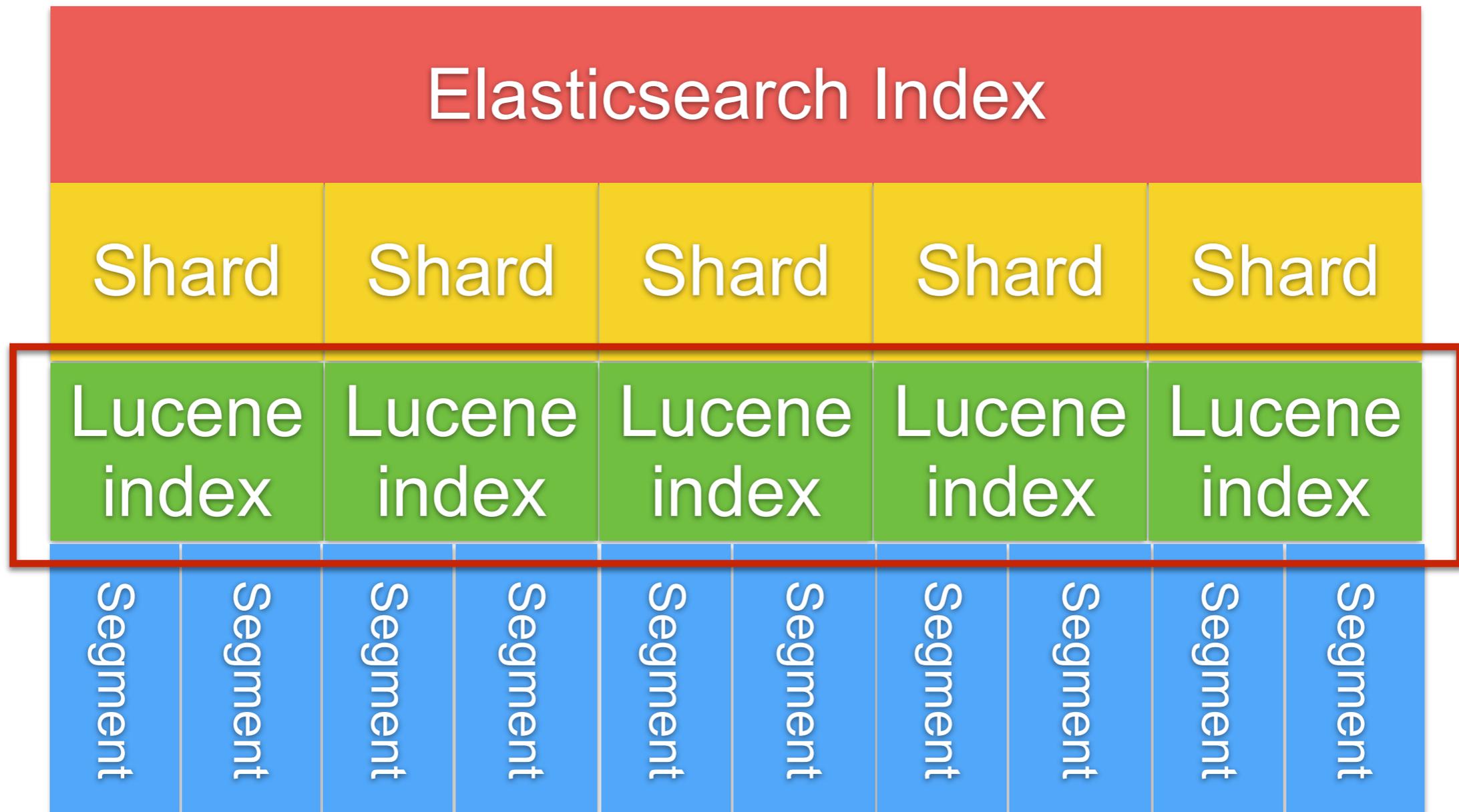
<http://lucene.apache.org/>



# Apache Lucene



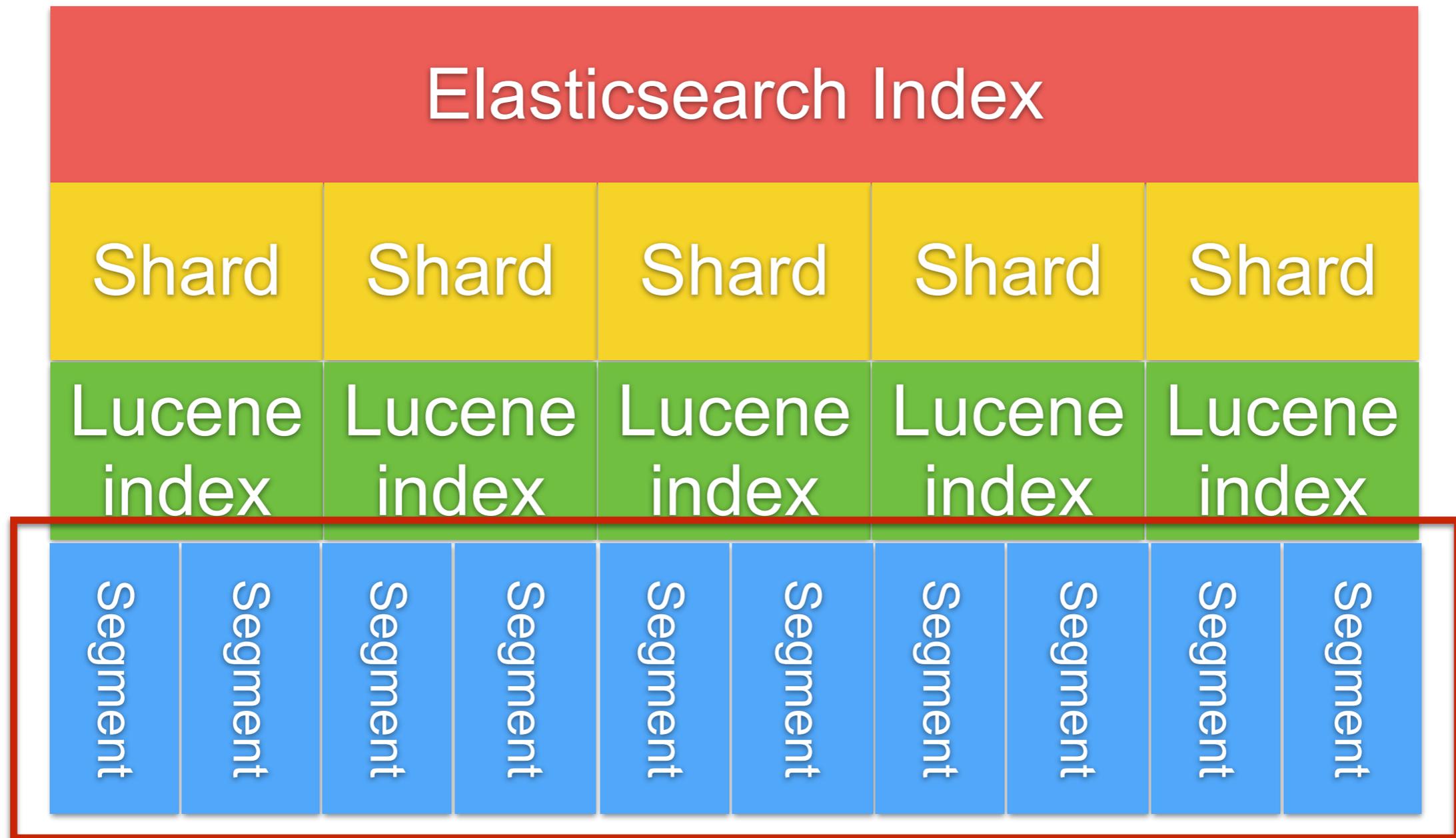
# Apache Lucene



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



# Apache Lucene



**Segments are immutable**



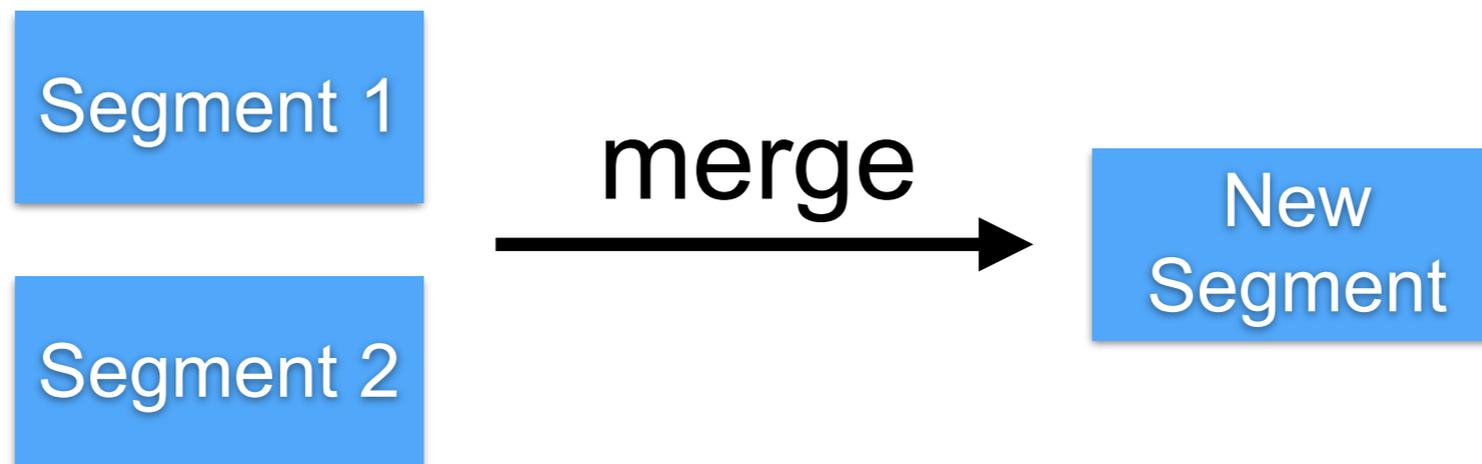
# Segment

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	$(\# \text{ of document} / 5M) + 1$



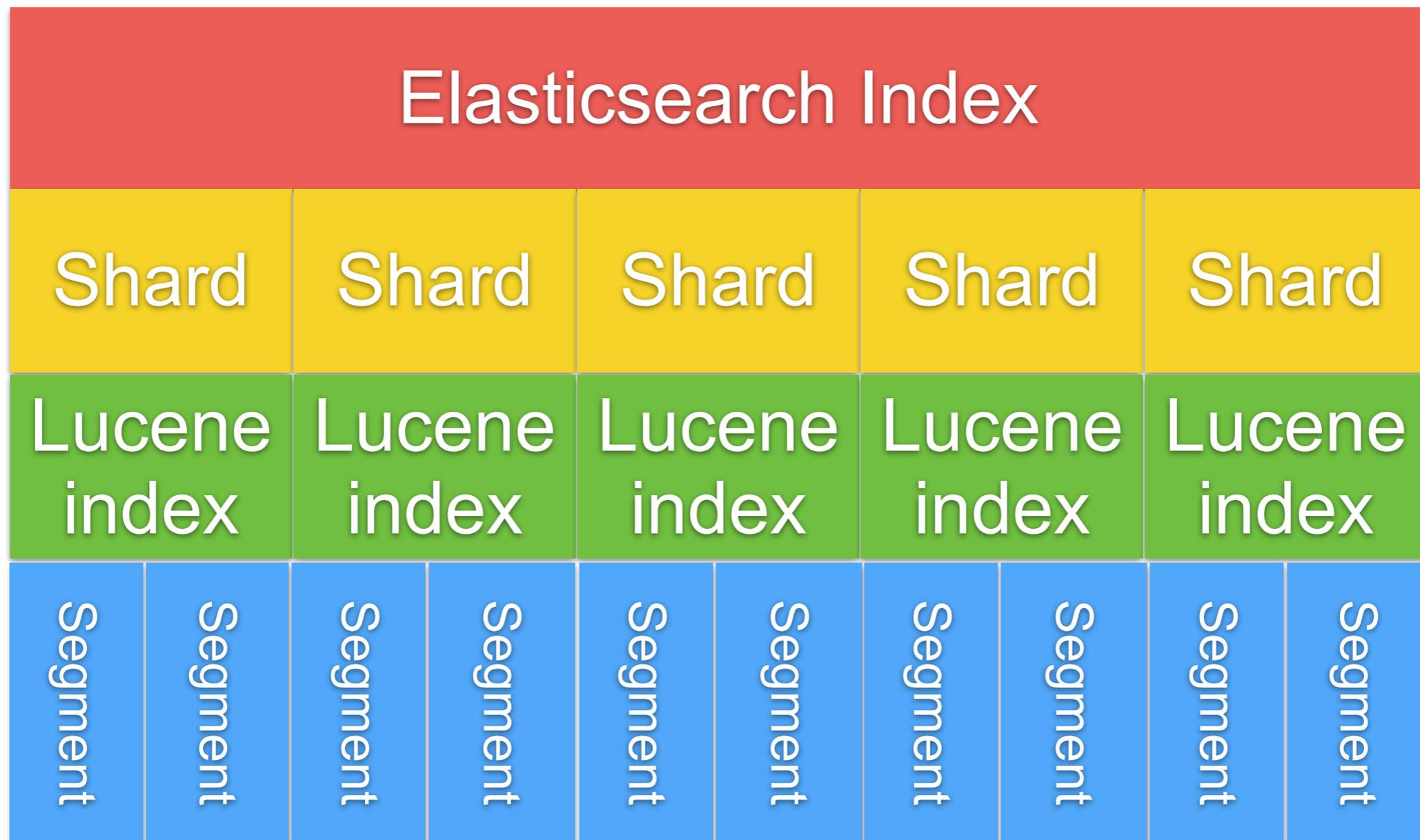
# Sharding

**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 !!



Need to config file descriptor

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



**Don't create more shard than  
you need !!**



# Replication

Prevent data loss

Default = 1

$$\# \text{ nodes} = [(\text{primary} + \# \text{ replication}) / 2 ] + 1$$



# 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>



# 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>



# 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

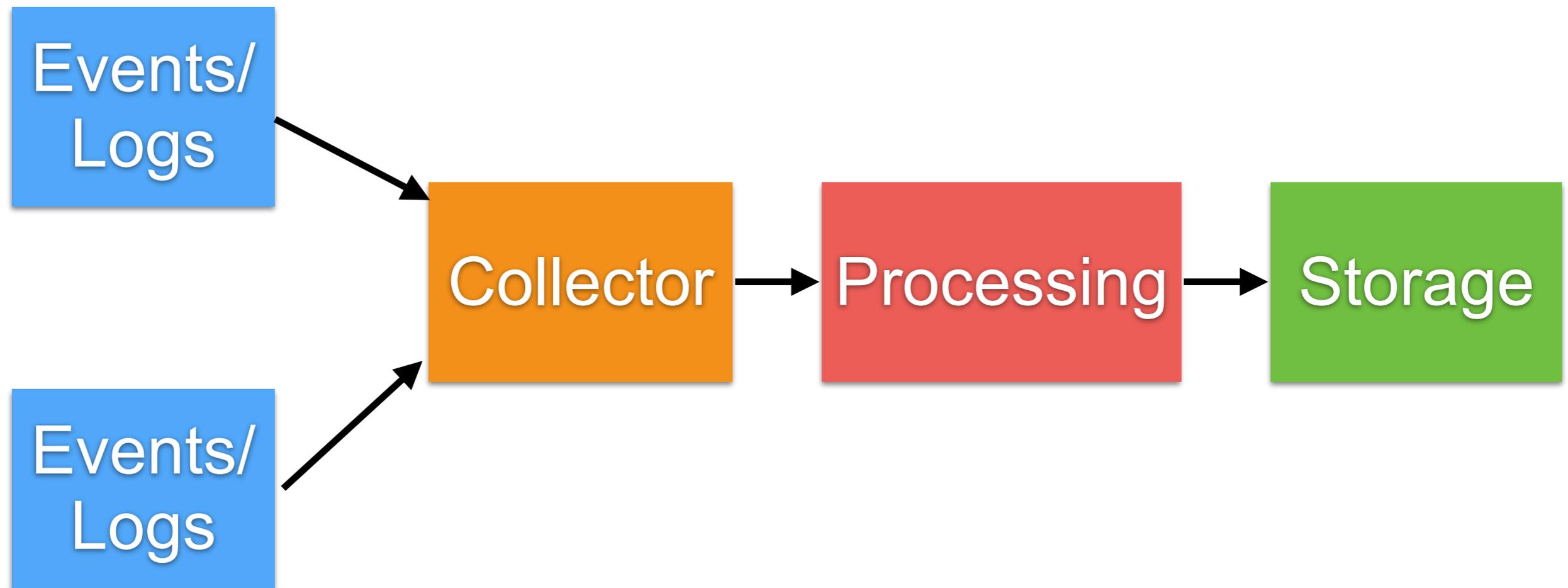
<https://www.elastic.co/guide/en/elasticsearch/reference/current/query-cache.html>



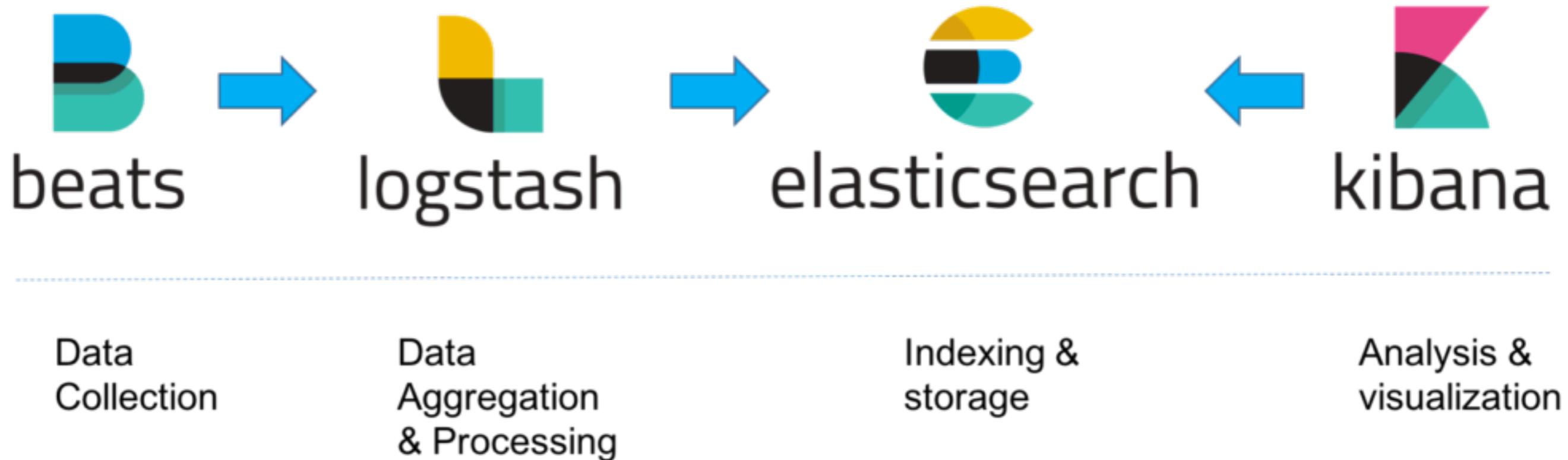
# Use cases



# Event or Logging from Servers



# Event or Logging from Servers



# Event or Logging from Servers



Data  
Collection

Buffering

Data  
Aggregation  
& Processing

Indexing &  
storage

Analysis &  
visualization



# Monitoring



# Collect data from ?

Elasticsearch nodes

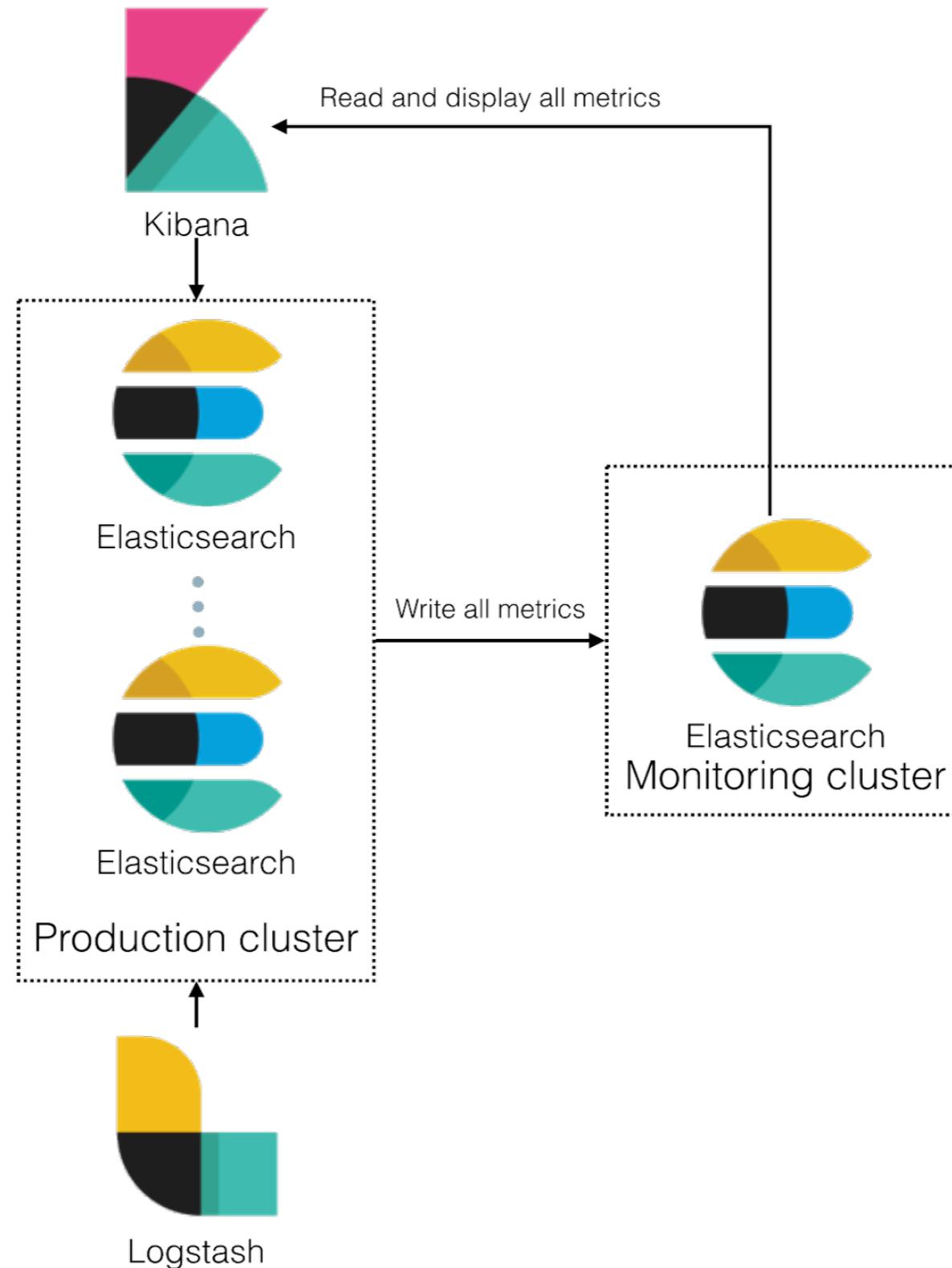
Logstash nodes

Kibana instances

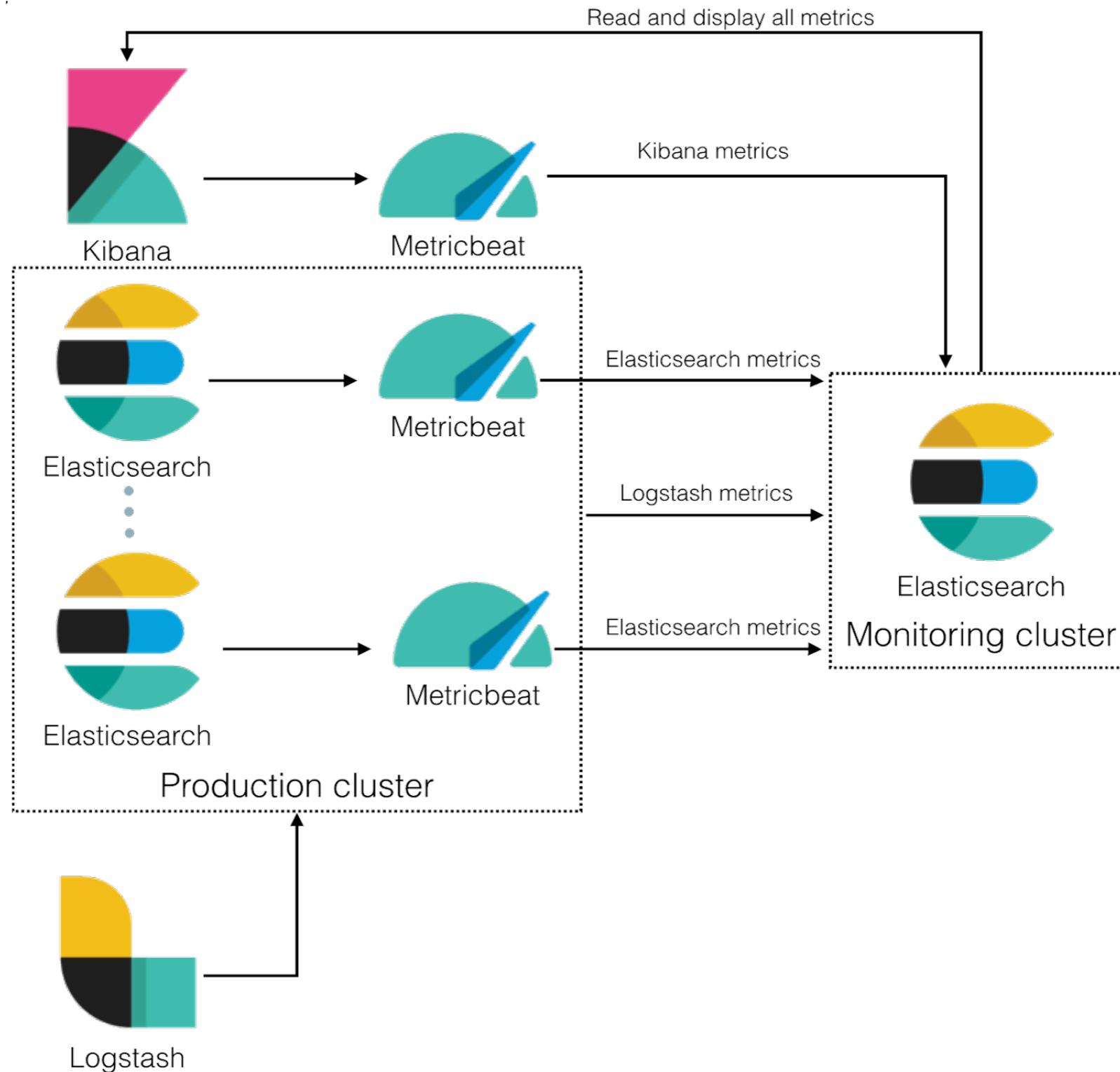
<https://www.elastic.co/guide/en/elastic-stack-overview/6.5/xpack-monitoring.html>



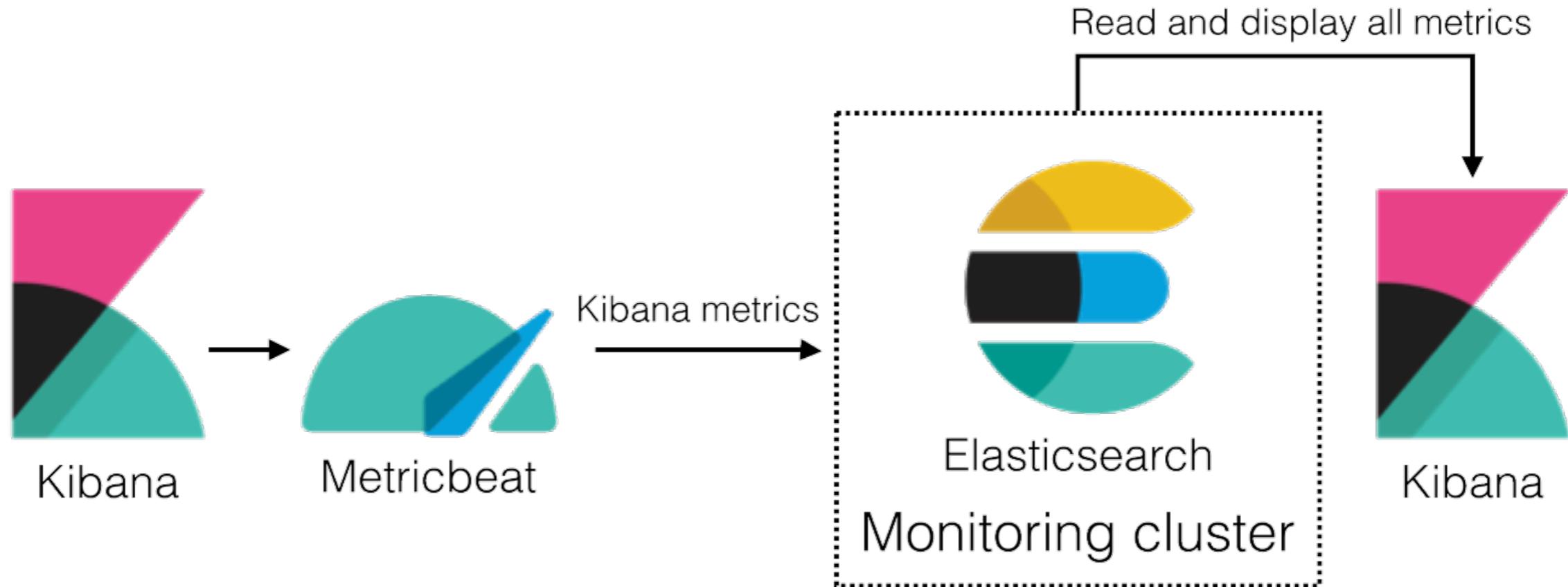
# Recommend architecture



# Elasticsearch 6.4 + (beta)



# Try to separate kibana



# Metrics

workshop/prometheus-grafana



# Sample Architecture



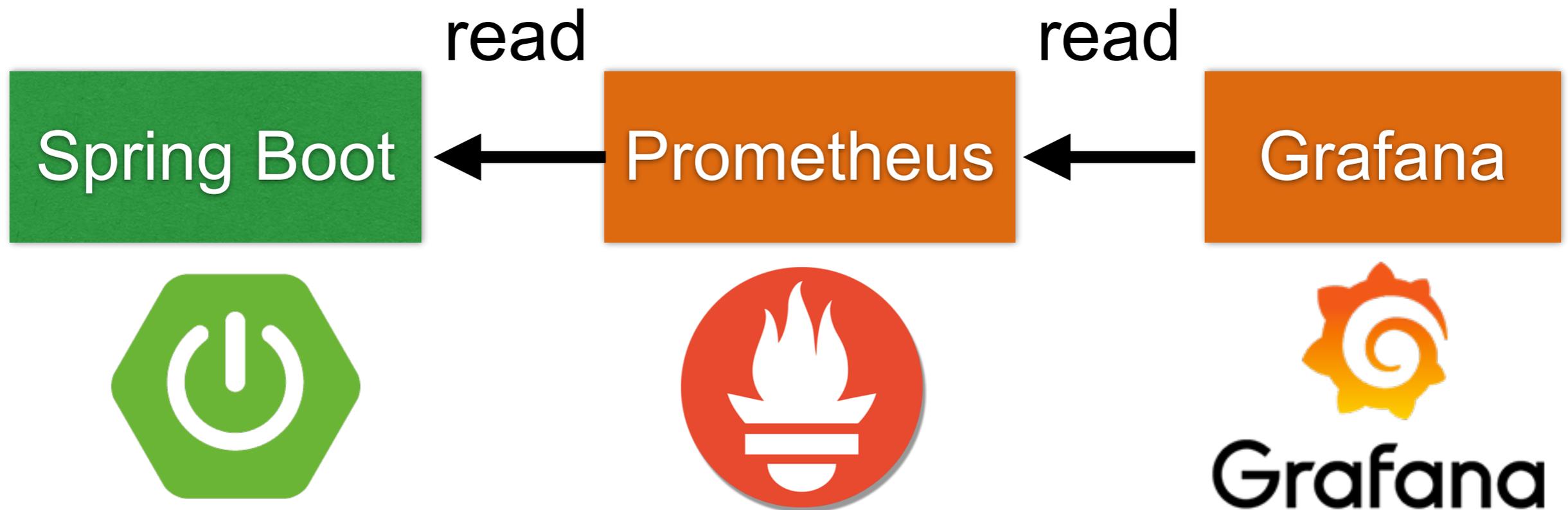
+



+



# Sample Architecture



# Metric in Spring Boot

Spring Boot **Actuator** for Spring Boot 1.x

**MicroMeter** for Spring Boot 2.0



# Spring Boot Actuator (1)

Add library to pom.xml

```
<dependency>  
  <groupId>org.springframework.boot</groupId>  
  <artifactId>spring-boot-starter-actuator</artifactId>  
</dependency>
```



# Spring Boot Actuator (2)

Enabled endpoint in application.properties

```
info.app.name=Toy Store
```

```
info.app.description=This is my first spring boot application
```

```
info.app.version=1.0.0
```

```
management.endpoints.web.exposure.include=health,info,metrics,httptrace
```



# Spring Boot Actuator (3)

List of endpoints = /actuator/

localhost:8080/actuator/

```
{
  - _links: {
    - self: {
      href: "http://localhost:8080/actuator",
      templated: false
    },
    - health: {
      href: "http://localhost:8080/actuator/health",
      templated: false
    },
    - info: {
      href: "http://localhost:8080/actuator/info",
      templated: false
    },
    - metrics-requiredMetricName: {
      href: "http://localhost:8080/actuator/metrics/{requiredMetricName}",
      templated: true
    },
    - metrics: {
      href: "http://localhost:8080/actuator/metrics",
      templated: false
    },
    - httptrace: {
      href: "http://localhost:8080/actuator/httptrace",
      templated: false
    }
  }
}
```



# Spring Boot Actuator (4)

Info endpoint = /actuator/info

```
localhost:8080/actuator/info
```

```
{  
  - app: {  
    name: "Toy Store",  
    description: "This is my first spring boot application",  
    version: "1.0.0"  
  }  
}
```



# Spring Boot Actuator (5)

Info endpoint = /actuator/info

← → ↻ ⓘ localhost:8080/actuator/info

```
{  
  - app: {  
    name: "Toy Store",  
    description: "This is my first spring boot application",  
    version: "1.0.0"  
  }  
}
```



# Spring Boot Actuator (6)

Info endpoint = /actuator/httptrace

localhost:8080/actuator/httptrace

```
{
  - traces: [
    - {
      timestamp: "2018-03-06T13:33:02.800Z",
      principal: null,
      session: null,
      - request: {
        method: "GET",
        uri: "http://localhost:8080/prometheus",
        - headers: {
          - host: [
            "localhost:8080"
          ],
          - user-agent: [
            "Prometheus/2.0.0"
          ],
          - accept: [
            "text/plain;version=0.0.4;q=1,*/*;q=0.1"
          ],
          - accept-encoding: [
            "gzip"
          ],
          - x-prometheus-scrape-timeout-seconds: [
            "5.000000"
          ]
        },
        remoteAddress: null
      },
    },
  ],
}
```



# Spring Boot Actuator (7)

List of metrics endpoint = /actuator/metrics

```
← → ↻ ⓘ localhost:8080/actuator/metrics
```

```
{  
  - names: [  
    "jvm.buffer.memory.used",  
    "jvm.memory.used",  
    "jvm.gc.memory.allocated",  
    "jvm.memory.committed",  
    "http.server.requests",  
    "jdbc.connections.min",  
    "tomcat.sessions.created",  
    "tomcat.sessions.expired",  
    "hikaricp.connections.usage",  
    "tomcat.global.request.max",  
    "tomcat.global.error",  
    "jvm.gc.max.data.size",  
    "logback.events",  
    "system.cpu.count",  
    "jvm.memory.max",  
    "jdbc.connections.active",  
    "jvm.buffer.total.capacity",  
    "jvm.buffer.count",  
    "process.files.max",  
    "jvm.threads.daemon",
```



# Spring Boot Actuator (8)

/actuator/metrics/http.server.requests

localhost:8080/actuator/metrics/http.server.requests

```
{
  name: "http.server.requests",
  - measurements: [
    - {
      statistic: "COUNT",
      value: 269
    },
    - {
      statistic: "TOTAL_TIME",
      value: 1.1072010200000002
    },
    - {
      statistic: "MAX",
      value: 0.04373569
    }
  ],
  - availableTags: [
    - {
      tag: "exception",
      - values: [
        "None"
      ]
    },
    - {
      tag: "method",
      - values: [
        "GET"
      ]
    }
  ],
}
```



# Spring Boot 2.0 with MicroMeter



**MICROMETER**  
application monitoring

[Documentation](#) [Github](#) [Twitter](#) [Slack](#)



## Vendor-neutral application metrics facade

Micrometer provides a simple facade over the instrumentation clients for the most popular monitoring systems, allowing you to instrument your JVM-based application code without vendor lock-in. Think SLF4J, but for metrics.

```
nds)
rtLatency, statistic=count
  Min : 184.786m
  Last : 198.255m
  Cnt : 180.000

requests/second)
rtLatency, statistic=count
  Min : 13.124k
  Last : 63.218k
  Cnt : 180.000
```

<https://micrometer.io/>



# Service metric for Prometheus



# Enable Prometheus (1)

Add library to pom.xml

```
<dependency>  
  <groupId>io.micrometer</groupId>  
  <artifactId>micrometer-registry-prometheus</artifactId>  
  <version>1.0.1</version>  
</dependency>
```



# Enable Prometheus (2)

Enabled endpoint in application.properties

```
management.endpoints.web.exposure.include  
=.....,prometheus
```



# Enable Prometheus (3)

New endpoint = actuator/prometheus

localhost:8080/actuator/prometheus

```
# HELP jvm_memory_used_bytes The amount of used memory
# TYPE jvm_memory_used_bytes gauge
jvm_memory_used_bytes{area="nonheap",id="Code Cache",} 1.49056E7
jvm_memory_used_bytes{area="nonheap",id="Metaspace",} 5.6766712E7
jvm_memory_used_bytes{area="nonheap",id="Compressed Class Space",} 7617096.0
jvm_memory_used_bytes{area="heap",id="PS Eden Space",} 1.7135864E7
jvm_memory_used_bytes{area="heap",id="PS Survivor Space",} 1.6235192E7
jvm_memory_used_bytes{area="heap",id="PS Old Gen",} 2.1936456E7
# HELP hikaricp_connections_idle Idle connections
# TYPE hikaricp_connections_idle gauge
hikaricp_connections_idle{pool="HikariPool-1",} NaN
# HELP tomcat_threads_config_max
# TYPE tomcat_threads_config_max gauge
tomcat_threads_config_max{name="http-nio-8080",} 200.0
# HELP tomcat_servlet_error_total
# TYPE tomcat_servlet_error_total counter
tomcat_servlet_error_total{name="default",} 0.0
# HELP jvm_threads_peak The peak live thread count since the Java virtual machine start
# TYPE jvm_threads_peak gauge
jvm_threads_peak 28.0
# HELP hikaricp_connections_pending Pending threads
# TYPE hikaricp_connections_pending gauge
hikaricp_connections_pending{pool="HikariPool-1",} NaN
# HELP system_cpu_count The number of processors available to the Java virtual machine
```

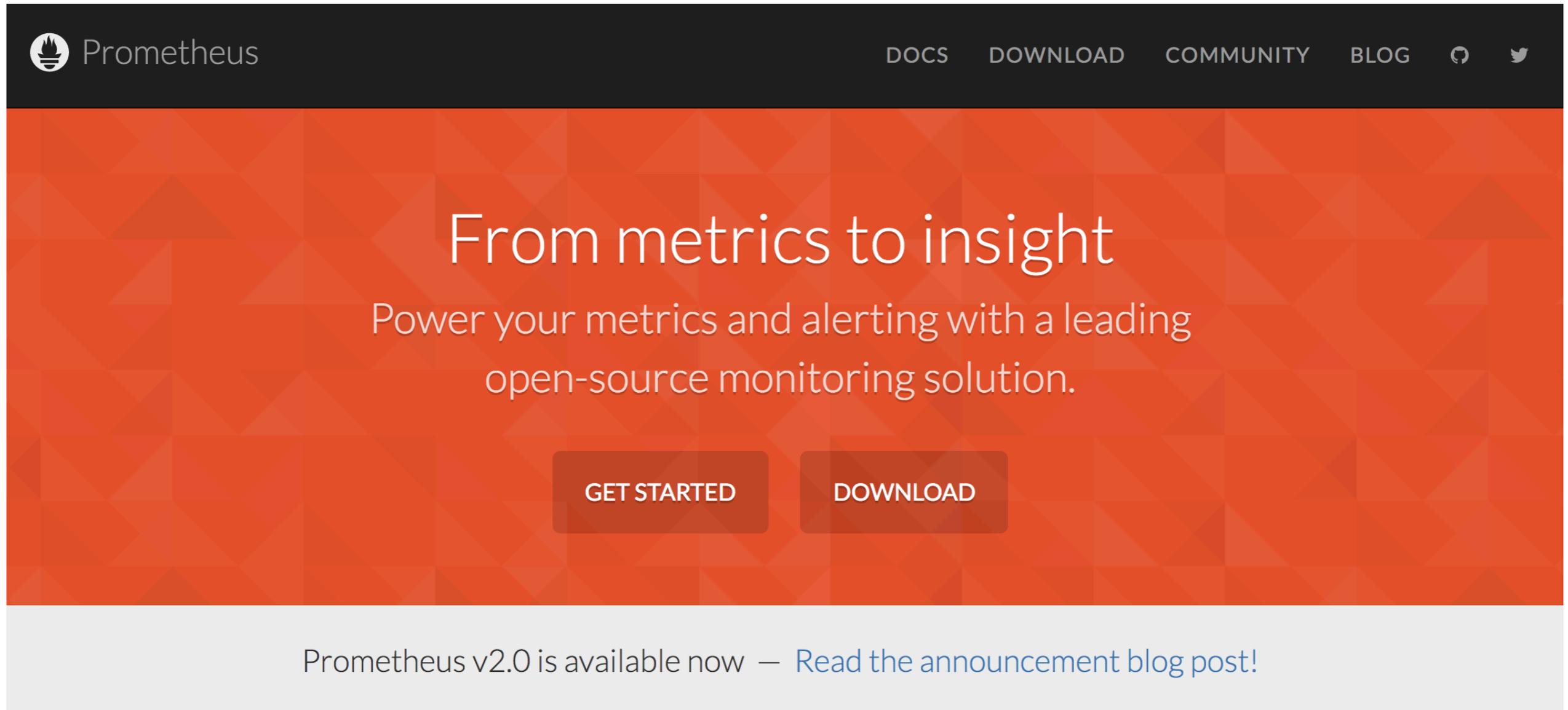


# Keep data in Prometheus

<https://prometheus.io/>



# Prometheus



The screenshot shows the Prometheus website homepage. At the top left is the Prometheus logo (a flame in a circle) and the word "Prometheus". To the right are navigation links: "DOCS", "DOWNLOAD", "COMMUNITY", and "BLOG", followed by social media icons for GitHub and Twitter. The main content area has an orange background with a geometric pattern. It features the headline "From metrics to insight" and the sub-headline "Power your metrics and alerting with a leading open-source monitoring solution." Below this are two buttons: "GET STARTED" and "DOWNLOAD". At the bottom of the orange section, a light gray banner contains the text "Prometheus v2.0 is available now – [Read the announcement blog post!](#)".

<https://prometheus.io/>



# Prometheus

PUBLIC | AUTOMATED BUILD

[prom/prometheus](#) ☆

Last pushed: 17 hours ago

[Repo Info](#) [Tags](#) [Dockerfile](#) [Build Details](#)

## Short Description

Short description is empty for this repo.

## Full Description

Prometheus build passing

circleci passing

container ready

docker pulls 63M

Visit [prometheus.io](#) for the full documentation, examples and guides.

Prometheus is a systems and service monitoring system. It collects metrics

## Docker Pull Command

```
docker pull prom/prometheus
```

## Owner



prom

## Source Repository

[prometheus/prometheus](#)

<https://hub.docker.com/r/prom/prometheus/>



# Create container of Prometheus

```
$docker container run --rm
```

```
-p 9090:9090
```

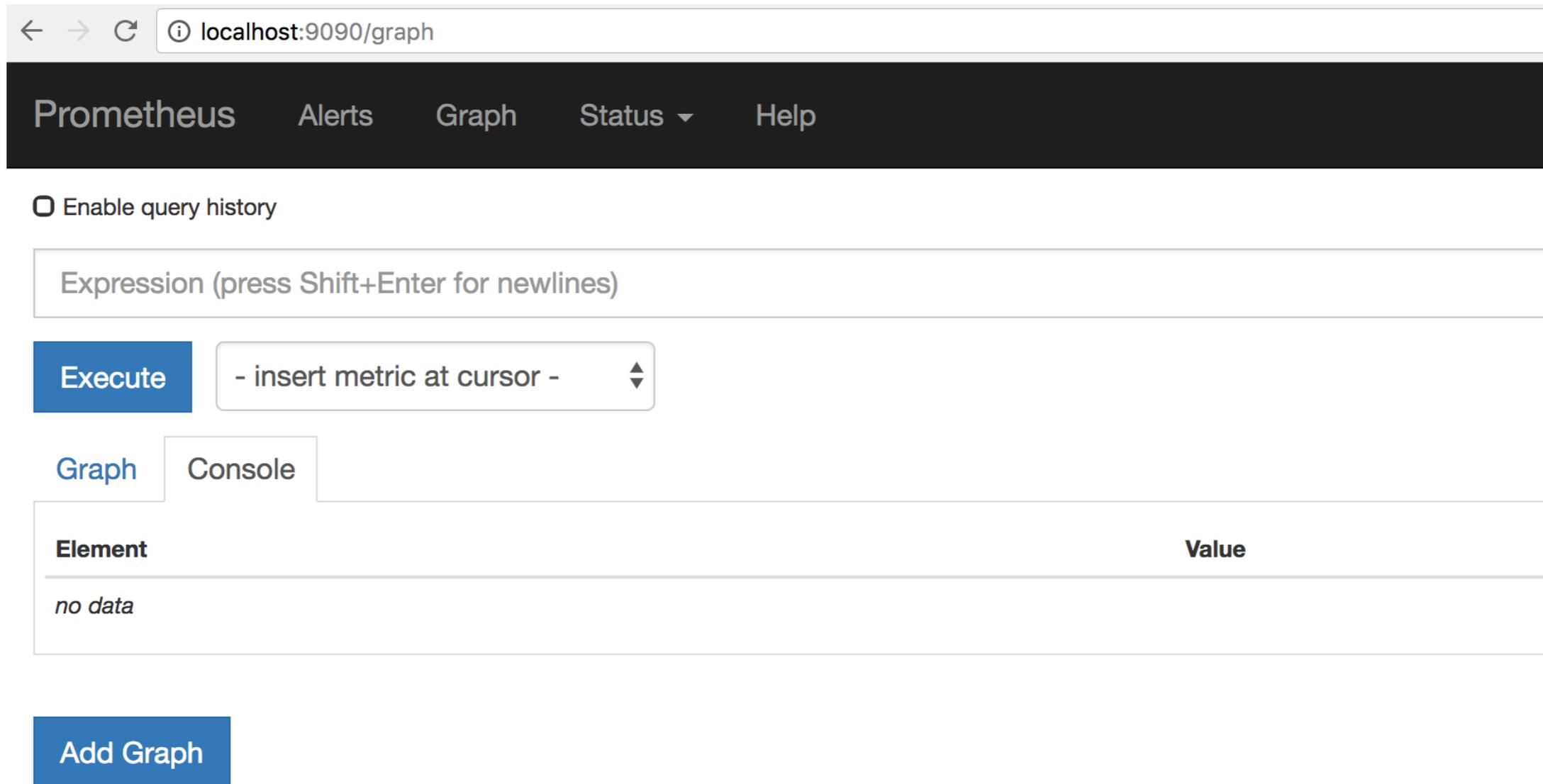
```
-v $(pwd)/prometheus.yml:/etc/prometheus/  
prometheus.yml
```

```
--name monitor prom/prometheus
```



# Check Data in Prometheus

http://localhost:9090/



The screenshot shows the Prometheus web interface. At the top, there is a navigation bar with links for Prometheus, Alerts, Graph, Status, and Help. Below the navigation bar, there is a checkbox for "Enable query history". The main area contains a text input field for the query expression, a blue "Execute" button, and a dropdown menu for inserting metrics. Below the input field, there are two tabs: "Graph" and "Console". The "Console" tab is active, showing a table with two columns: "Element" and "Value". The table contains a single row with the text "no data". At the bottom left, there is a blue "Add Graph" button.

localhost:9090/graph

Prometheus Alerts Graph Status Help

Enable query history

Expression (press Shift+Enter for newlines)

Execute - insert metric at cursor -

Graph Console

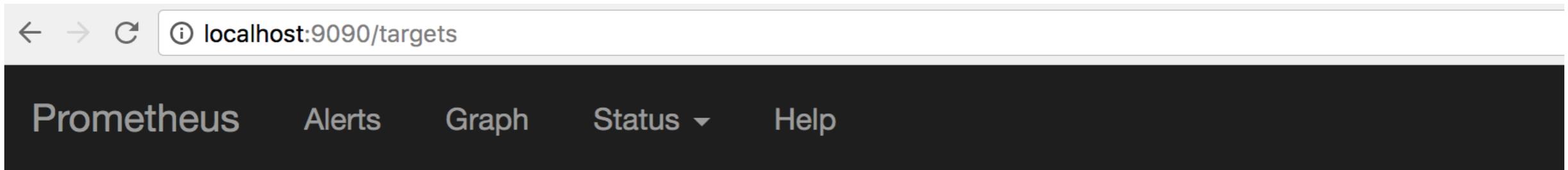
Element	Value
no data	

Add Graph



# Check Target in Prometheus

Status -> Targets



## Targets

Only unhealthy jobs

spring-boot (1/1 up) [show less](#)

Endpoint	State	Labels	Last Scrape	Error
<a href="http://10.10.99.59:8080/actuator/prometheus">http://10.10.99.59:8080/actuator/prometheus</a>	UP	instance="10.10.99.59:8080"	2.355s ago	

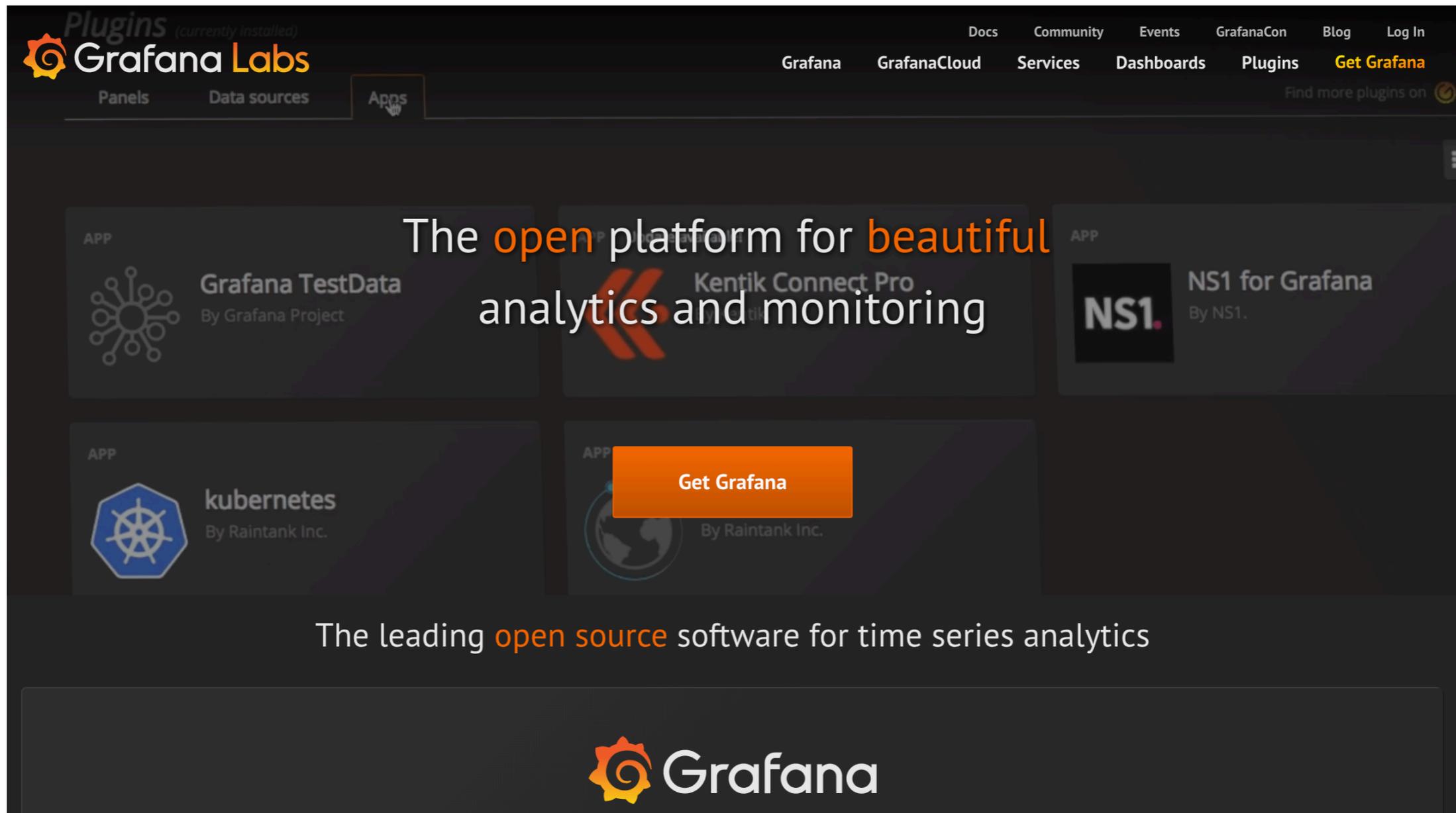


# Show data in Grafana

<https://grafana.com/>



# Grafana



The screenshot shows the Grafana Labs website interface. At the top, there's a navigation bar with 'Grafana Labs' logo and 'Plugins (currently installed)' text. Below the navigation, there are tabs for 'Panels', 'Data sources', and 'Apps'. The main content area displays several plugin cards: 'Grafana TestData', 'Kentik Connect Pro', 'NS1 for Grafana', and 'kubernetes'. A prominent orange button labeled 'Get Grafana' is overlaid on the 'kubernetes' card. The central text overlay reads: 'The open platform for beautiful analytics and monitoring'. At the bottom of the screenshot, it says 'The leading open source software for time series analytics' and features the Grafana logo.

The **open** platform for **beautiful** analytics and monitoring

Get Grafana

The leading **open source** software for time series analytics

 Grafana

<https://grafana.com/>



# Grafana

PUBLIC REPOSITORY

[grafana/grafana](#) ☆

Last pushed: 25 minutes ago

Repo Info [Tags](#)

## Short Description

The official Grafana docker container

## Full Description

### Grafana Docker image

This project builds a Docker image with the latest master build of Grafana.

### Running your Grafana container

Start your container binding the external port `3000`.

```
docker run -d --name=grafana -p 3000:3000 grafana/grafana
```

## Docker Pull Command

```
docker pull grafana/grafana
```

## Owner



grafana

<https://hub.docker.com/r/grafana/grafana/>



# Create container of Grafana

\$docker container run

--name=grafana

-p 3000:3000 grafana/grafana



# Grafana Dashboard

<https://grafana.com/dashboards/4701>

All dashboards » [JVM \(Micrometer\)](#)



## JVM (Micrometer) by mweirauch

DASHBOARD

Dashboard for Micrometer instrumented applications (Java, Spring Boot)  
Last updated: 21 days ago

Downloads: 74

Overview Revisions



A dashboard for [Micrometer](#) instrumented applications (Java, Spring Boot).

### Features

- JVM memory
- Process memory (provided by [micrometer-jvm-extras](#))
- CPU-Usage, Load, Threads, File Descriptors, Log Events
- JVM Memory Pools (Heap, Non-Heap)
- Garbage Collection

Get this dashboard:

4701

[Copy ID to Clipboard](#)

[Download JSON](#)  
[How do I import this dashboard?](#)

Dependencies:

-  GRAFANA 4.6.3
-  GRAPH



# Take to your home

Always improve, always practice



# Machine Learning



# Machine Learning

Dataset increase in size and complexity

**Human effort is limited !!**

Infrastructure problem

Cyber attacks

Business issues

<https://www.elastic.co/guide/en/kibana/current/xpack-ml.html>



# Machine Learning in ES

Supervised  
Unsupervised



# Machine Learning in ES

**Supervised**  
**Classification**  
**Regression**



# Machine Learning in ES

**Unsupervised**  
Outlier detection  
Anatomy detection



# Machine Learning in ES

For basic licence (30 days trial)

## Data Visualizer

The Machine Learning Data Visualizer tool helps you understand your data, by analyzing the metrics and fields in a log file or an existing Elasticsearch index.

EXPERIMENTAL



### Import data

Import data from a log file. You can upload files up to 100 MB.

Upload file



### Select an index pattern

Visualize the data in an existing Elasticsearch index.

Select index

### Start trial

To experience the full Machine Learning features that a [Platinum subscription](#) offers, start a 30-day trial.

Start trial



# Continuous Learning Process

Data collection  
Feature engineering  
Training model  
Evaluate model  
Deploy on production  
Monitoring  
Continuous Improvement



# Let's workshop machine-learning/instruction.md



# 1. Data collection



# Add data to ES

## Using Add sample data

### 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



#### SIEM

Centralize security events for interactive investigation in ready-to-go visualizations.

Add security events

#### Add sample data

Load a data set and a Kibana dashboard

#### Upload data from log file

Import a CSV, NDJSON, or log file

#### Use Elasticsearch data

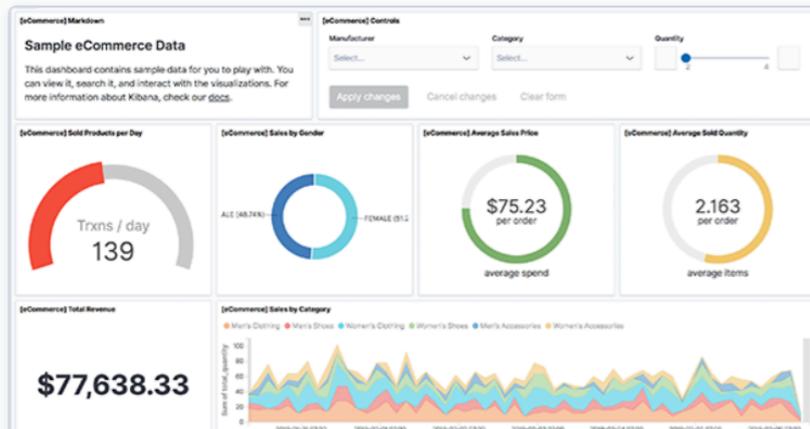
Connect to your Elasticsearch index



# Add data to ES

## Add Data to Kibana

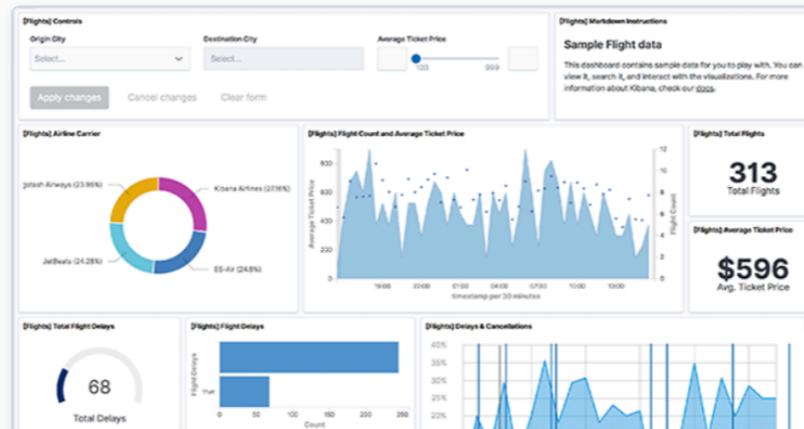
All Logging Metrics SIEM Sample data



### Sample eCommerce orders

Sample data, visualizations, and dashboards for tracking eCommerce orders.

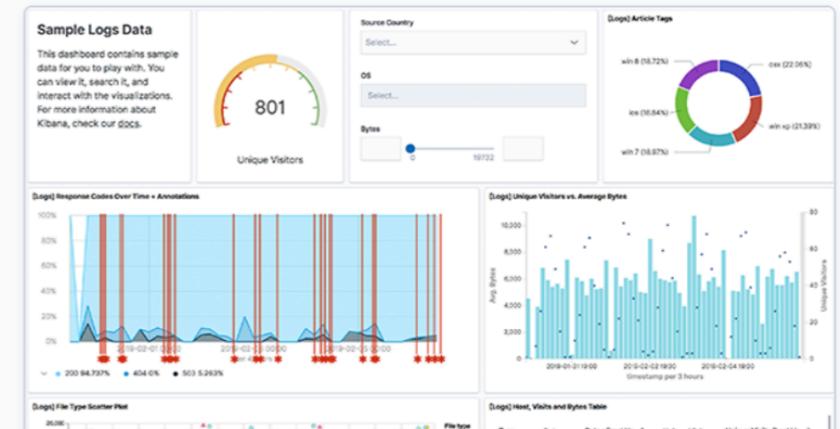
Add data



### Sample flight data

Sample data, visualizations, and dashboards for monitoring flight routes.

Add data



### Sample web logs

Sample data, visualizations, and dashboards for monitoring web logs.

Add data



# Query sample data

GET \_cat/indices?v

health	status	index	uuid	pri	rep	docs.count
		.size				
		pri.store.size				
green	open	.kibana_task_manager_1	JeYrzD1dTeeBgp34aFT2Fg	1	0	2
		.4kb				
		51.4kb				
green	open	.apm-agent-configuration	V9bxRNDIRm-QIT88FKU1Gw	1	0	0
		283b				
		283b				
green	open	.kibana_1	GMiyZBhbTyaomAA6JfCQ6A	1	0	72
		.6kb				
		107.6kb				
green	open	kibana_sample_data_flights	I-f1LueyTk22RIHu9TIyRQ	1	0	13059
		6mb				
		6mb				



# Query sample data

GET kibana\_sample\_data\_flights/\_search

```
{
  "took" : 0,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 10000,
      "relation" : "gte"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "kibana_sample_data_flights",
        "_type" : "_doc",
        "_id" : "r8fXHG8B8iN3UstZlXDN",
        "_score" : 1.0,
```

10,000 != 13,059 ?



# Query sample data

GET kibana\_sample\_data\_flights/\_search?  
scroll=1m

```
{
  "_scroll_id" : "DXF1ZXJ5QW5kRmV0Y2gBAAAAAAAAEucL
    0ZfdlFsYkpBQQ==",
  "took" : 0,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 13059,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
```

<https://www.elastic.co/guide/en/elasticsearch/reference/current/search-request-body.html#request-body-search-scroll>



# 2. Training model with ES



# Create data frame

Choose source and destination

Choose features/fields

Choose model to analyse

```
"analysis": {  
  "classification": {  
    outlier_detection API "FlightDelay",  
    regression API  
    "training_percent": 10  
  }  
},
```

<https://www.elastic.co/guide/en/elastic-stack-overview/7.5/dfa-classification.html>



# Data frame analytic

Data frame analytics type	Learning type	Evaluation type
outlier detection	unsupervised	binary soft classification
regression	supervised	regression
classification	supervised	classification

<https://www.elastic.co/guide/en/elastic-stack-overview/7.5/ml-dfa-overview.html>



# Start job to create model

POST `_ml/data_frame/analytics/model-flight-delay-classification/_start`

```
"data_frame_analytics" : [  
  {  
    "id" : "model-flight-delay-classification",  
    "state" : "analyzing",  
    "progress" : [  
      {  
        "phase" : "reindexing",  
        "progress_percent" : 100  
      },  
      {  
        "phase" : "loading_data",  
        "progress_percent" : 100  
      },  
      {  
        "phase" : "analyzing",  
        "progress_percent" : 46  
      },  
      {  
        "phase" : "writing_results",  
        "progress_percent" : 0  
      }  
    ]  
  }  
],
```



# See result

GET df-flight-delayed/\_search

```
"ml" : {
  "top_classes" : [
    {
      "class_probability" : 0
        .6113160839068559,
      "class_name" : "true"
    },
    {
      "class_probability" : 0
        .3886839160931441,
      "class_name" : "false"
    }
  ],
  "FlightDelay_prediction" : "true",
  "is_training" : true
}
```



# 3. Evaluate your model



# See result

POST `_ml/data_frame/_evaluate`

```
"classification" : {
  "multiclass_confusion_matrix" : {
    "confusion_matrix" : [
      {
        "actual_class" : "false",
        "actual_class_doc_count" : 8822,
        "predicted_classes" : [
          {
            "predicted_class" : "false",
            "count" : 7660
          },
          {
            "predicted_class" : "true",
            "count" : 1162
          }
        ]
      },
      {
        "other_predicted_class_doc_count" : 0
      }
    ]
  }
}
```



# Continuous Improvement



# Experiment features in ES !!



# Limitation of data frame

Cross cluster search is not supported

Delete data frame job does not delete index !!

Data frame can't be updated

Memory limitation

Missing fields are skipped

<https://www.elastic.co/guide/en/elastic-stack-overview/7.5/ml-dfa-limitations.html>

