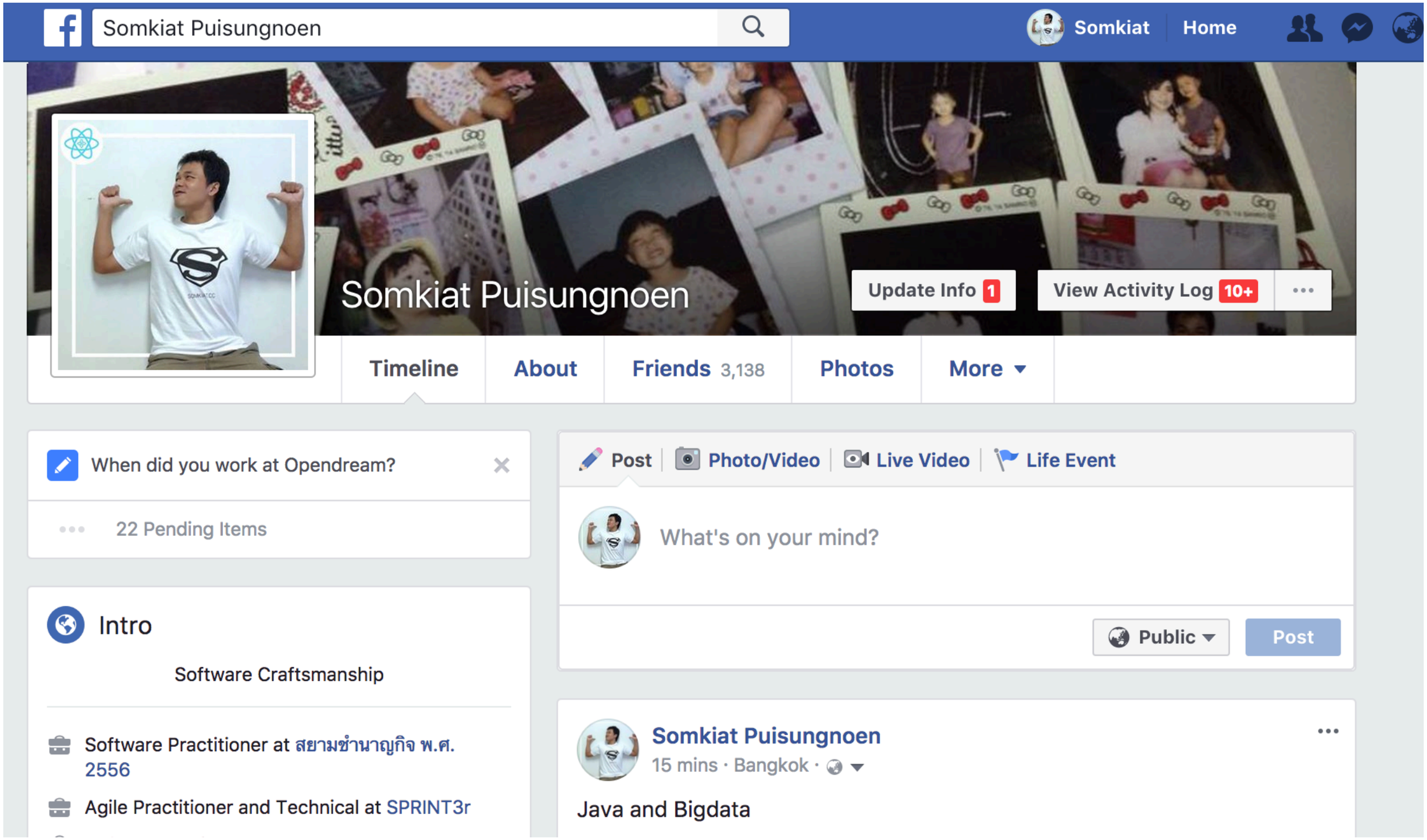
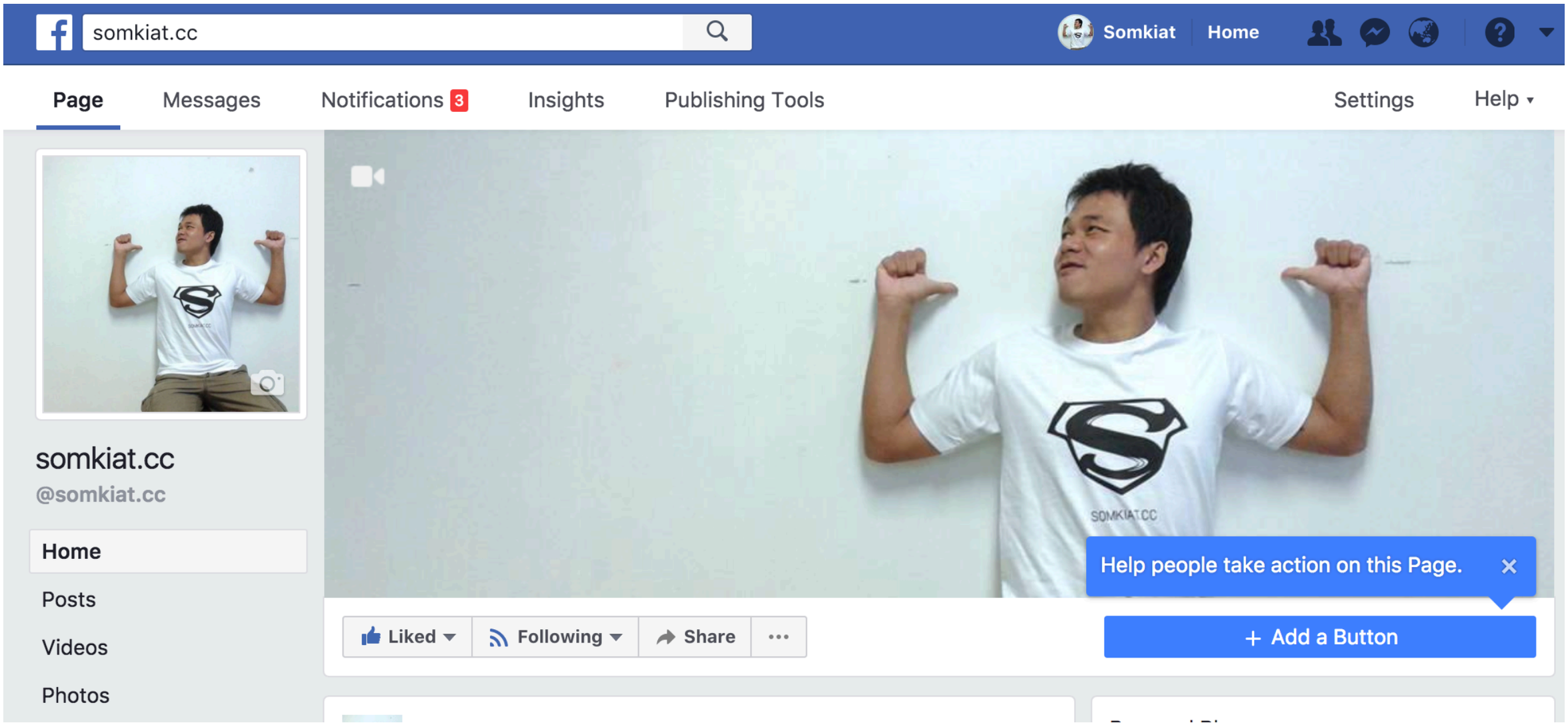
สยามชำนาญกิจ
SIAM CHAMNAN KIT

ELK Stack



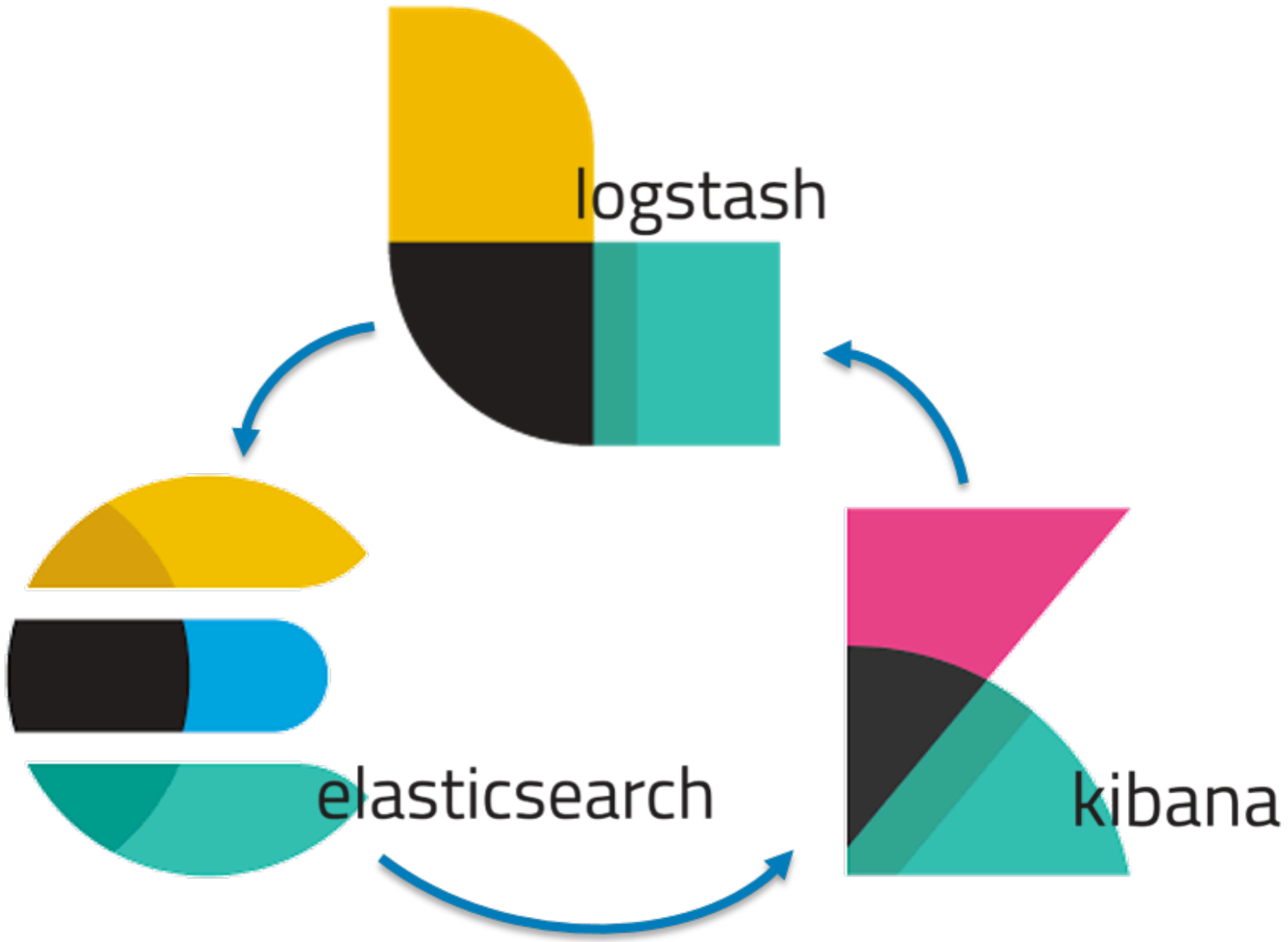


Agenda

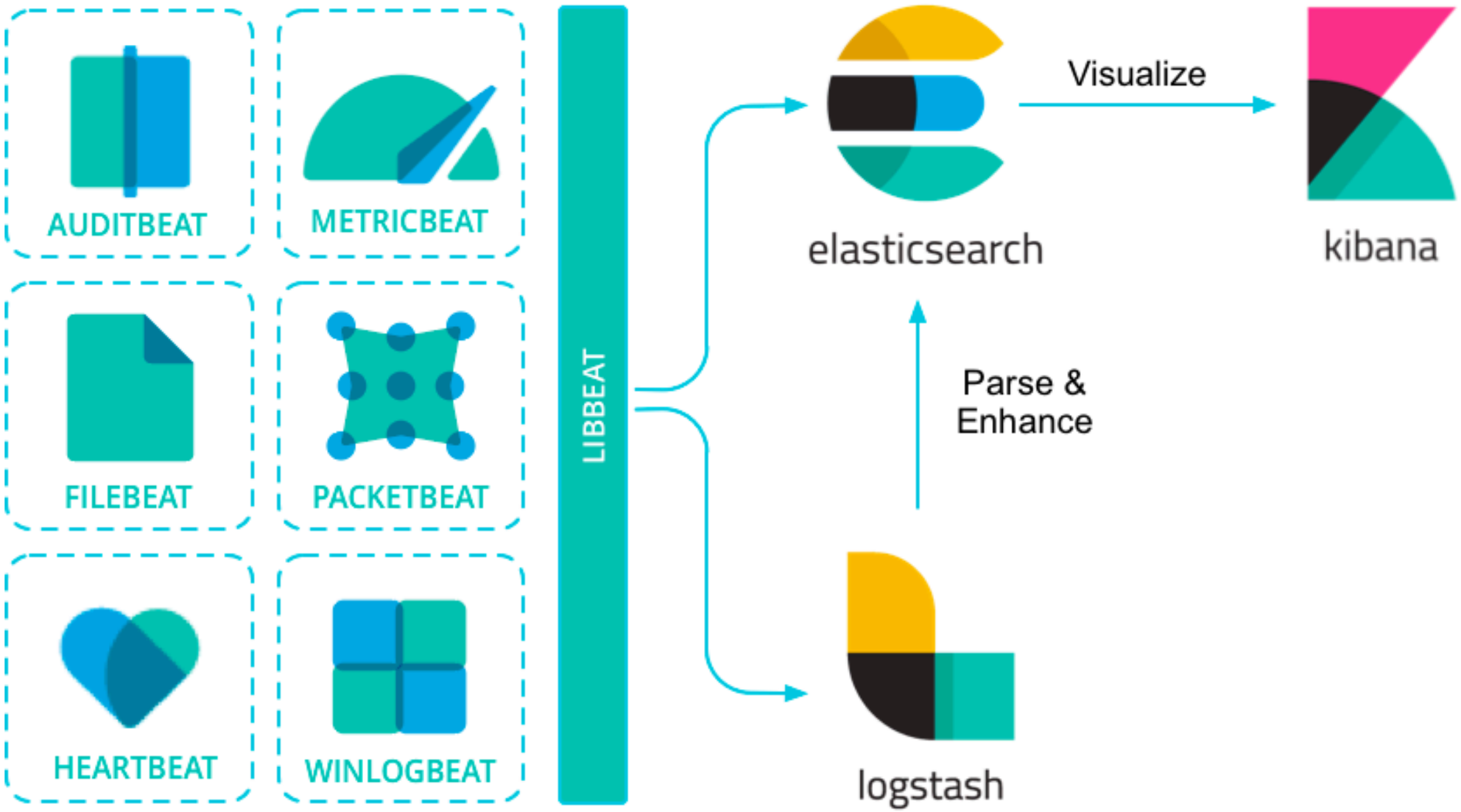
- ELK stack
- Introduction to Elasticsearch
- CRUD (Create, Read, Update, Delete)
- Search DSL (Domain Specific Language)
- Analyzer
- Mapping
- Aggregation

Agenda

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



Beat



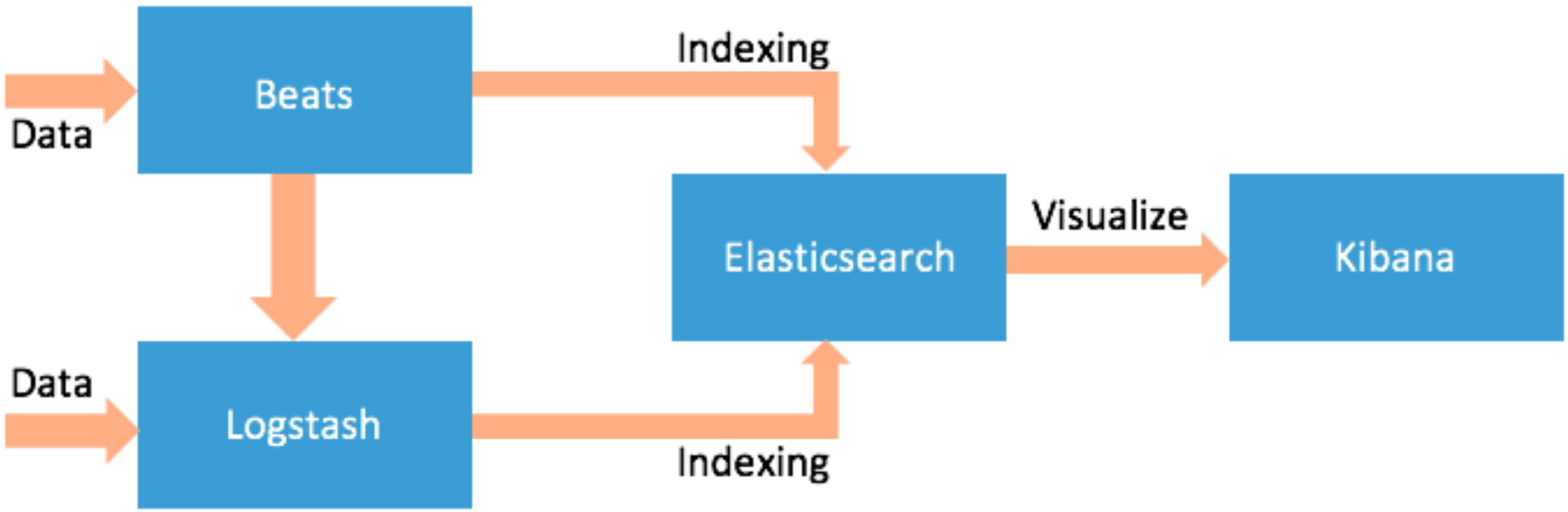
https://www.elastic.co/guide/en/beats/libbeat/current/index.html

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 stack



Elasticsearch ?

Elasticsearch

Search

Analytic

Real-time

Distributed

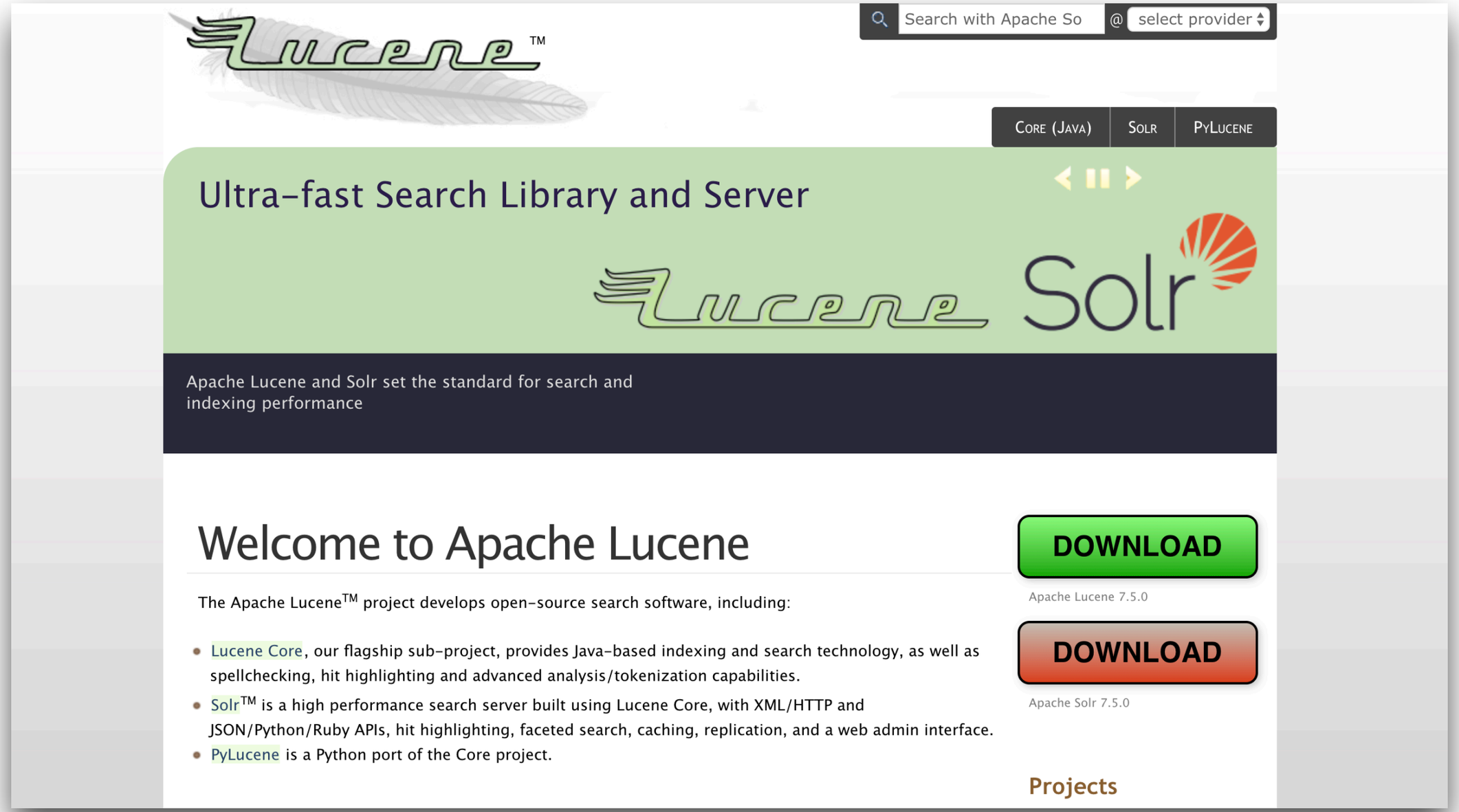
Scalability

Distributed Search Engine

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

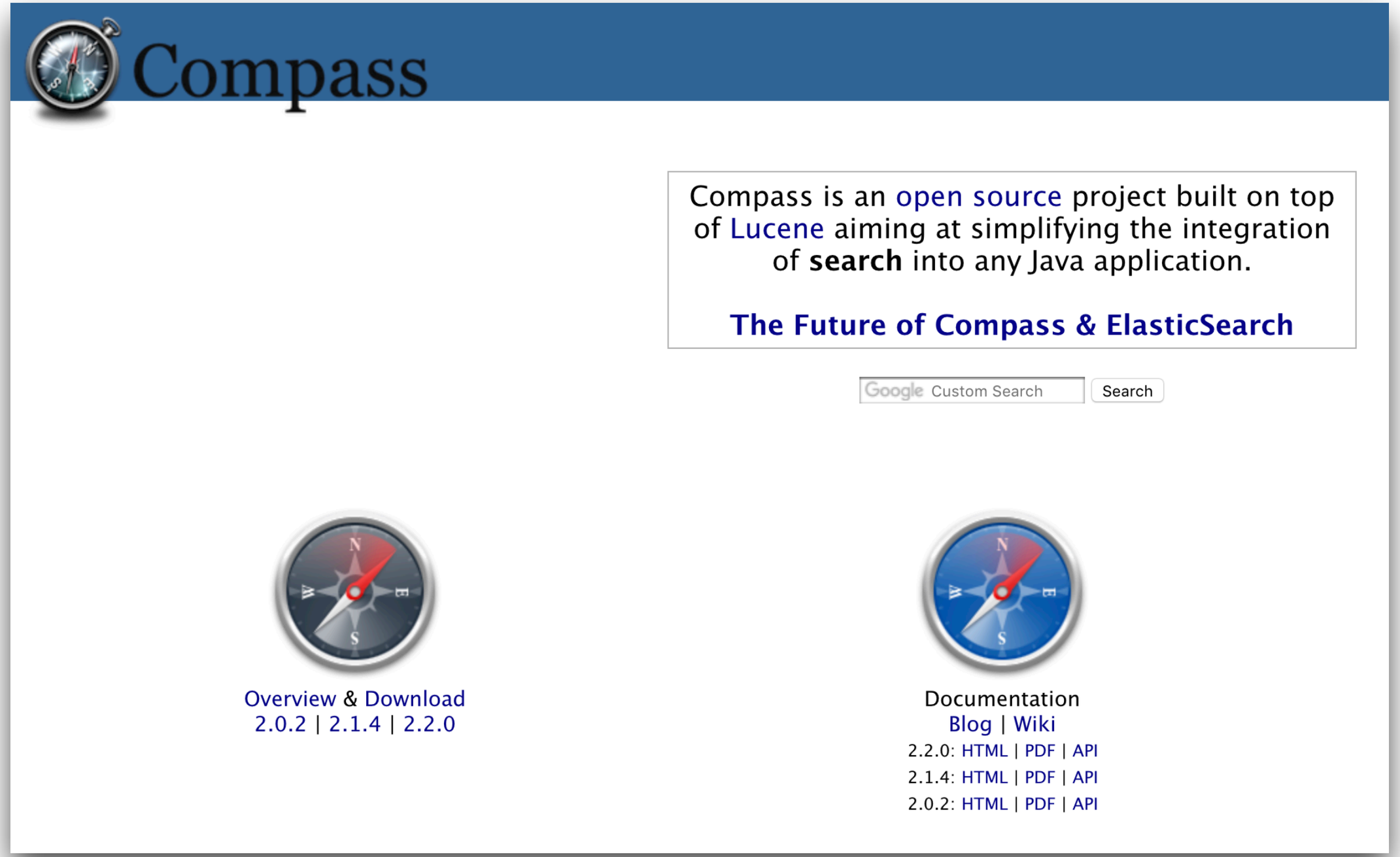
Distributed Search Engine

Apache Lucene



http://lucene.apache.org/

Compass



http://www.compass-project.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 Jun 2019	Rank May 2019	Rank Jun 2018	DBMS	Database Model	Score Jun 2019	Score May 2019	Score Jun 2018
1.	1.	1.	Oracle	Relational, Multi-model	1299.21	+13.67	-12.04
2.	2.	2.	MySQL	Relational, Multi-model	1223.63	+4.67	-10.06
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1087.76	+15.57	+0.03
4.	4.	4.	PostgreSQL	Relational, Multi-model	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	172.20	-2.24	-13.44
7.	7.	↑ 8.	Elasticsearch	Search engine, Multi-model	148.82	+0.20	+17.78
8.	8.	↓ 7.	Redis	Key-value, Multi-model	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

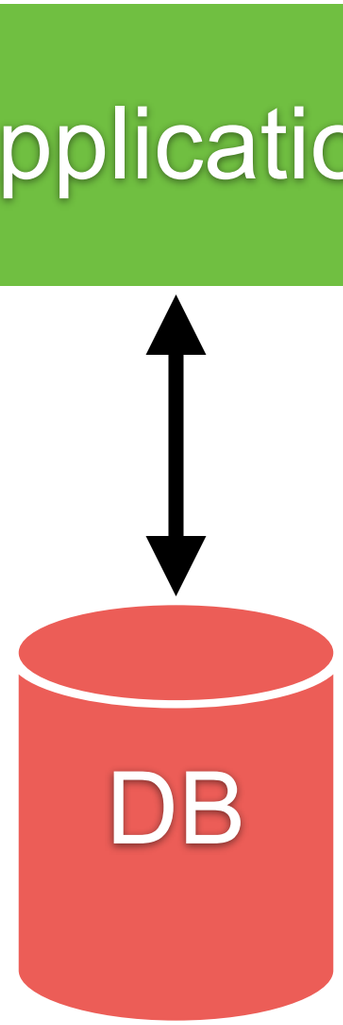
https://db-engines.com/en/ranking

Use cases

Security/log analytics
Marketing
Operations
Searching data

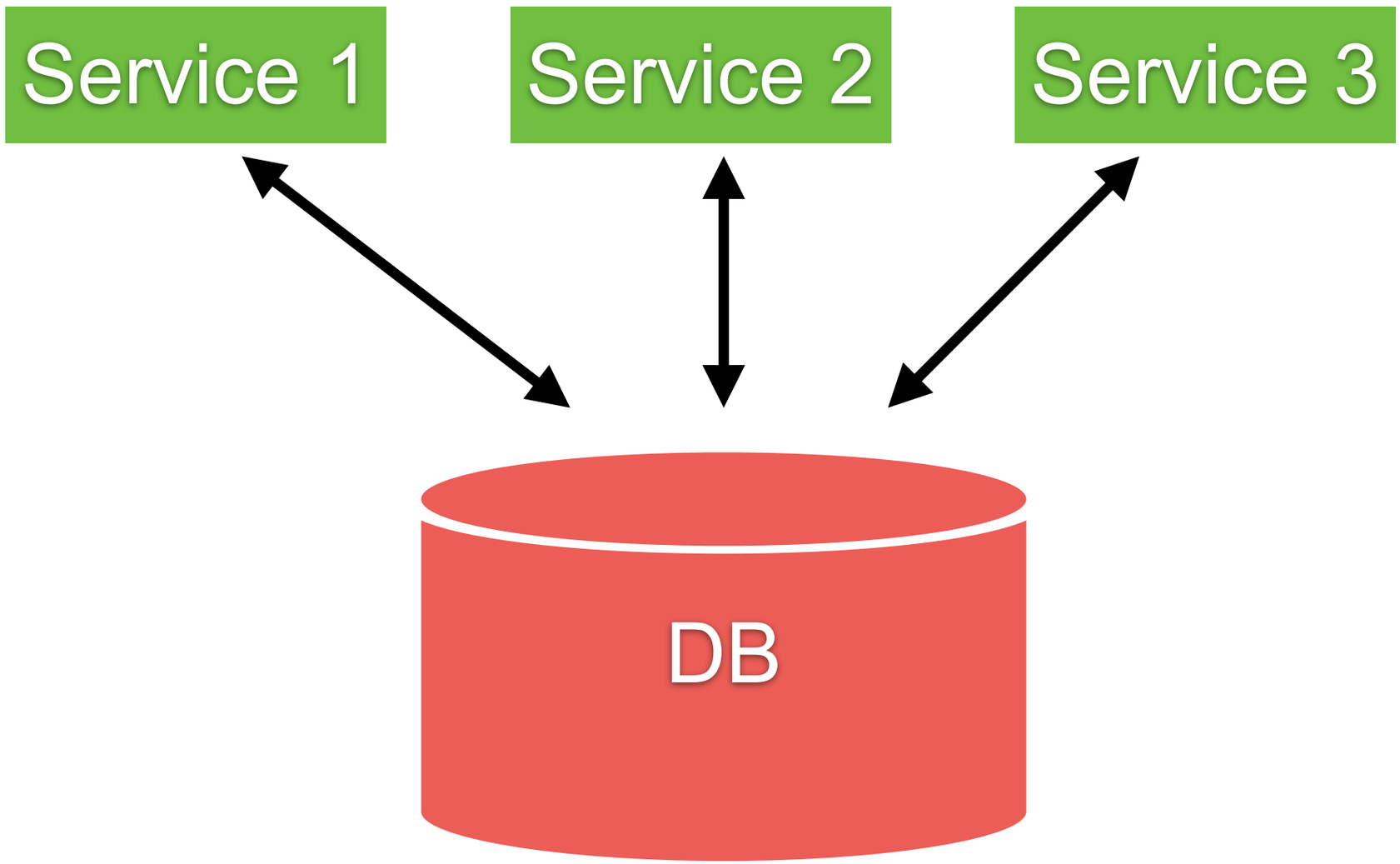
Problem ?

Single/Centralize database



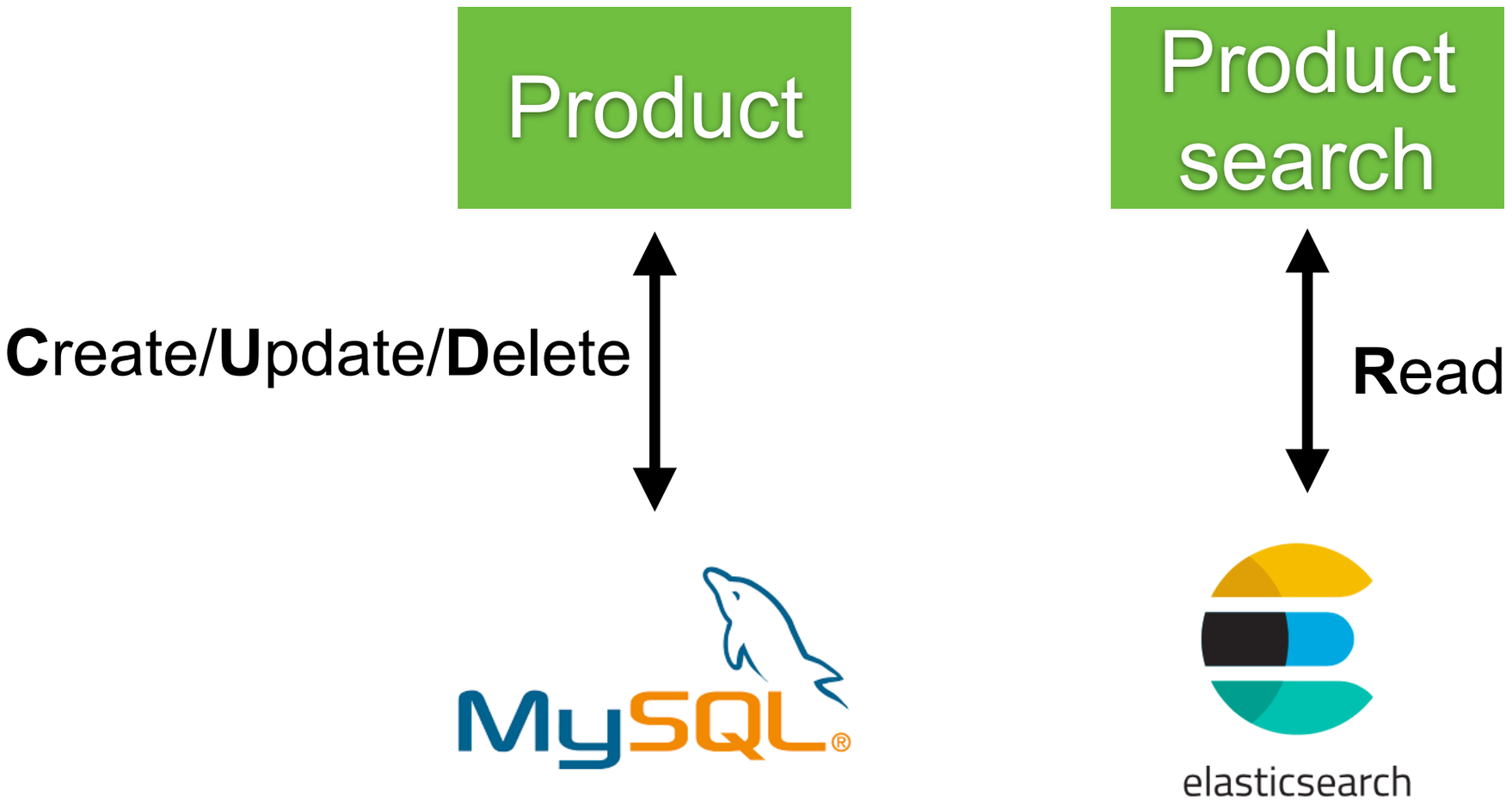
Problem ?

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
ansport] [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

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

X-Pack ?

Elastic Stack Extension
Security
Monitoring
Alerting
Reporting
Machine Learning

Licence

	FREE		GOLD	PLATINUM
	OPEN SOURCE	BASIC		
	Download		Request Info	Request Info
ELASTIC STACK				
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

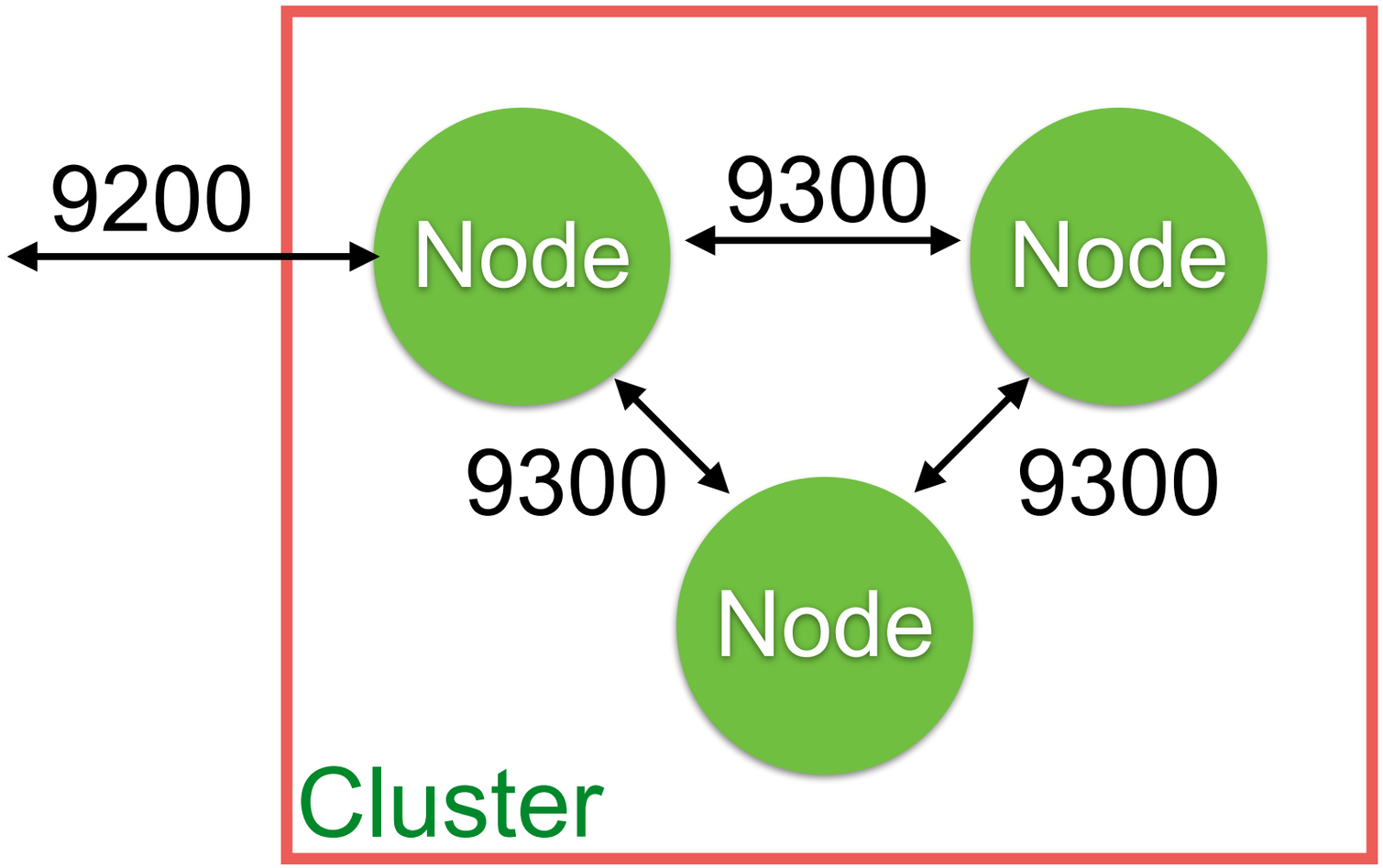
Hello Elasticsearch

http://localhost:**9200**/


```
{
    name: "Somkiats-MacBook-Pro",
    cluster_name: "elasticsearch",
    cluster_uuid: "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"
}
```


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


```
{
  "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

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
```

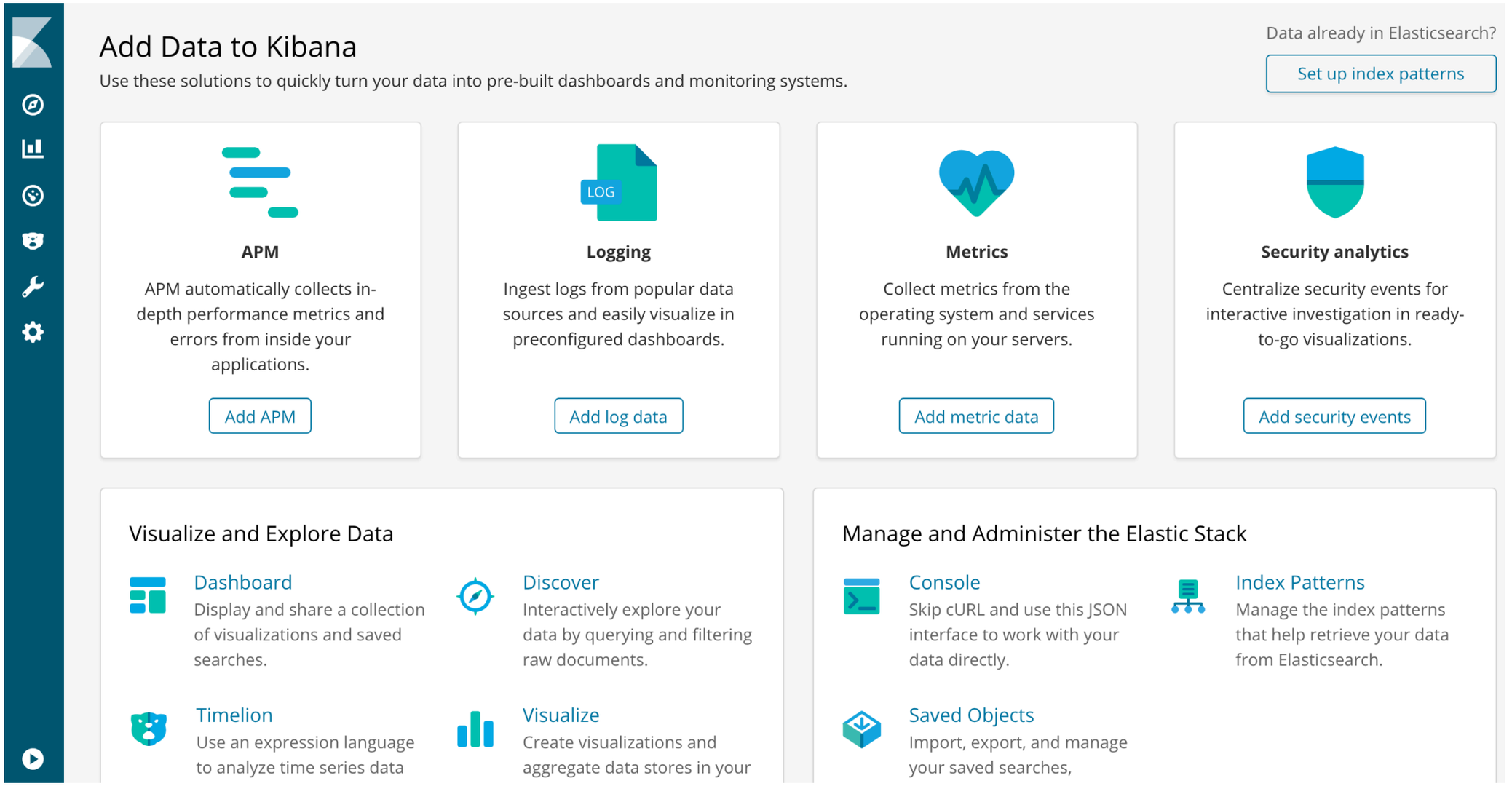
Install Kibana

Start Kibana

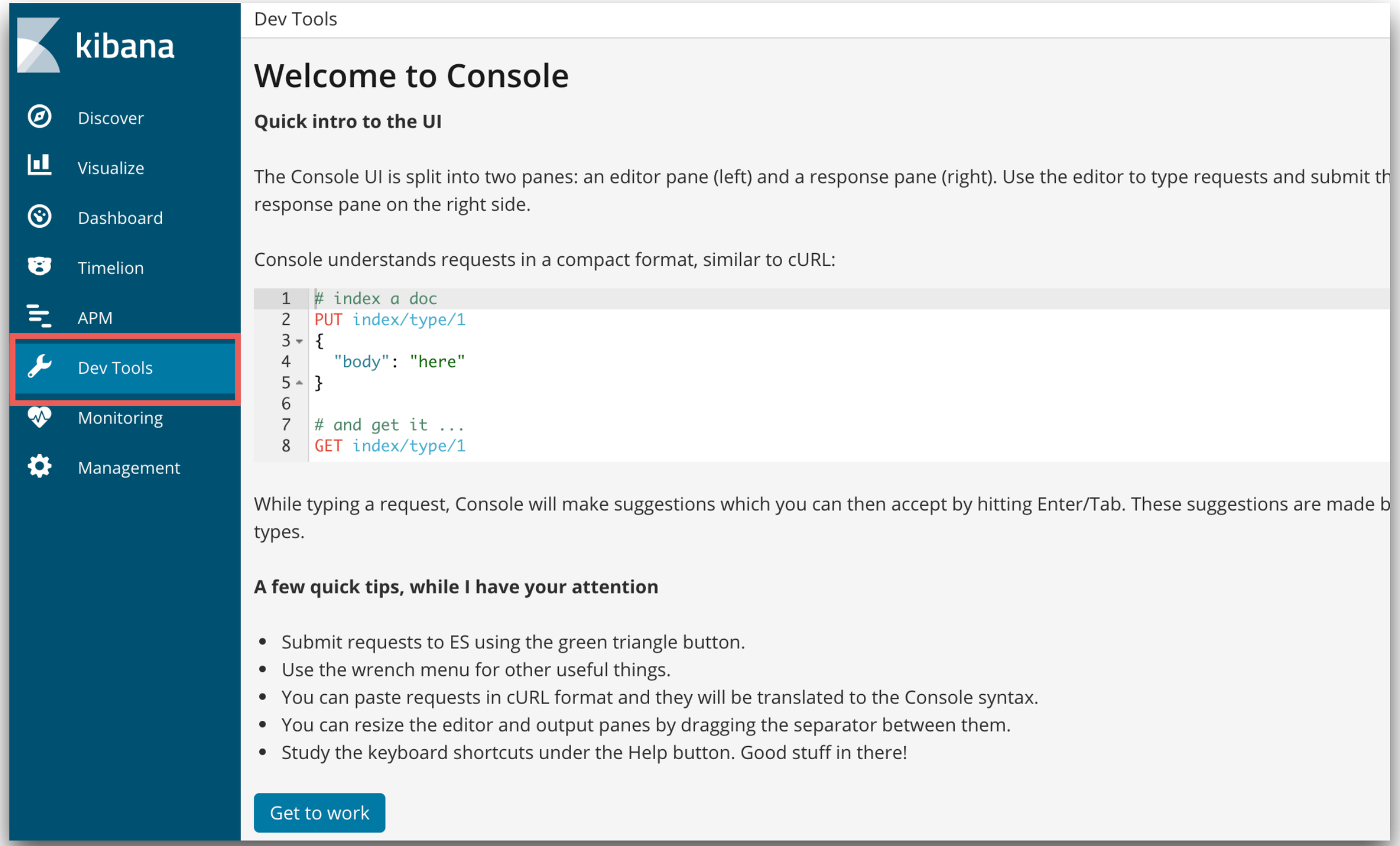
```
[status][plugin:xpack_main@6.4.2] Status changed from yellow to green - Ready
[status][plugin:searchprofiler@6.4.2] Status changed from yellow to green - Ready
[status][plugin:ml@6.4.2] Status changed from yellow to green - Ready
[status][plugin:tilemap@6.4.2] Status changed from yellow to green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

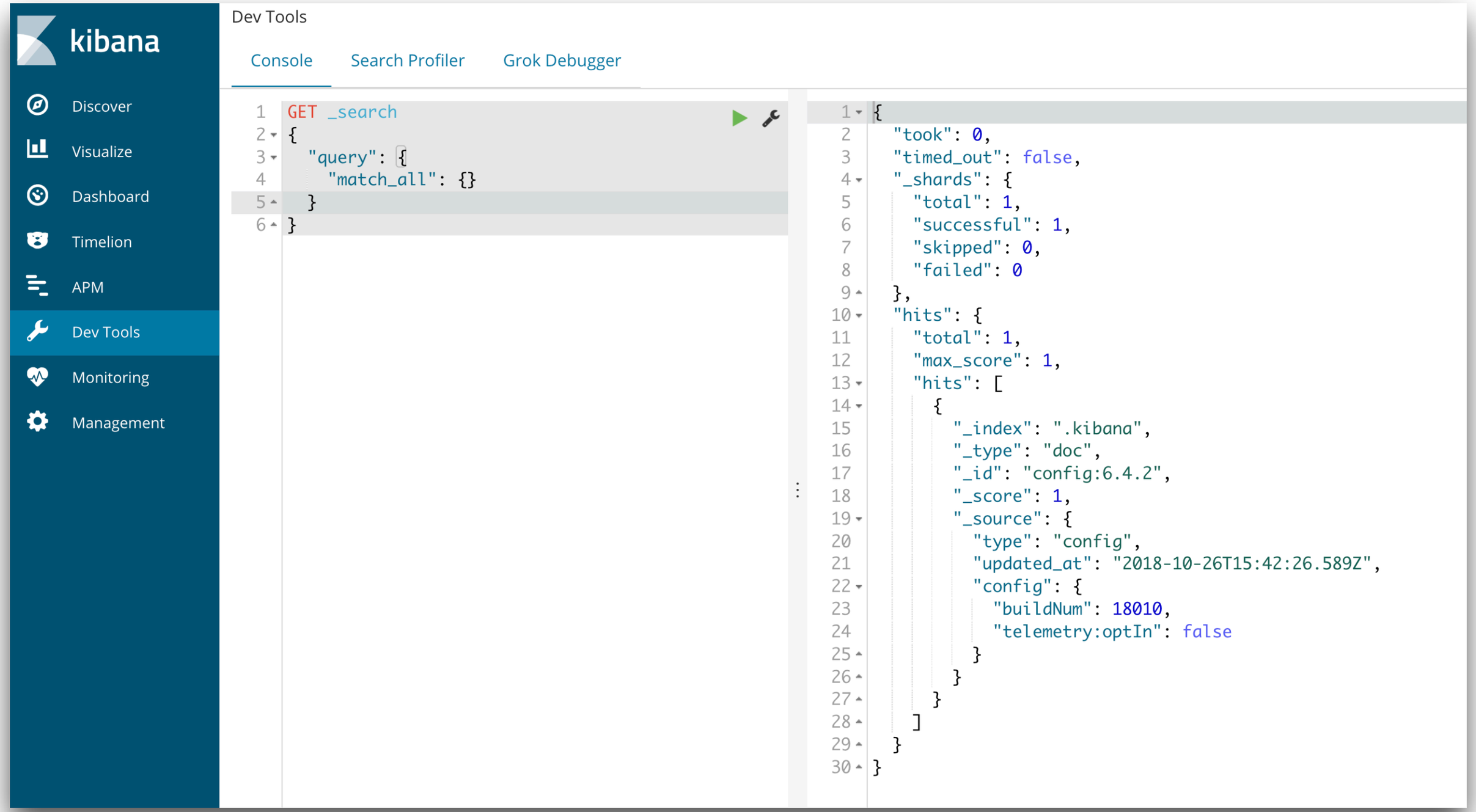
http://localhost:**5601**/



Using Dev Tools



Ready to start



Elasticsearch architecture

Basic concepts

Cluster

Node

Shard

Replica

Gateway

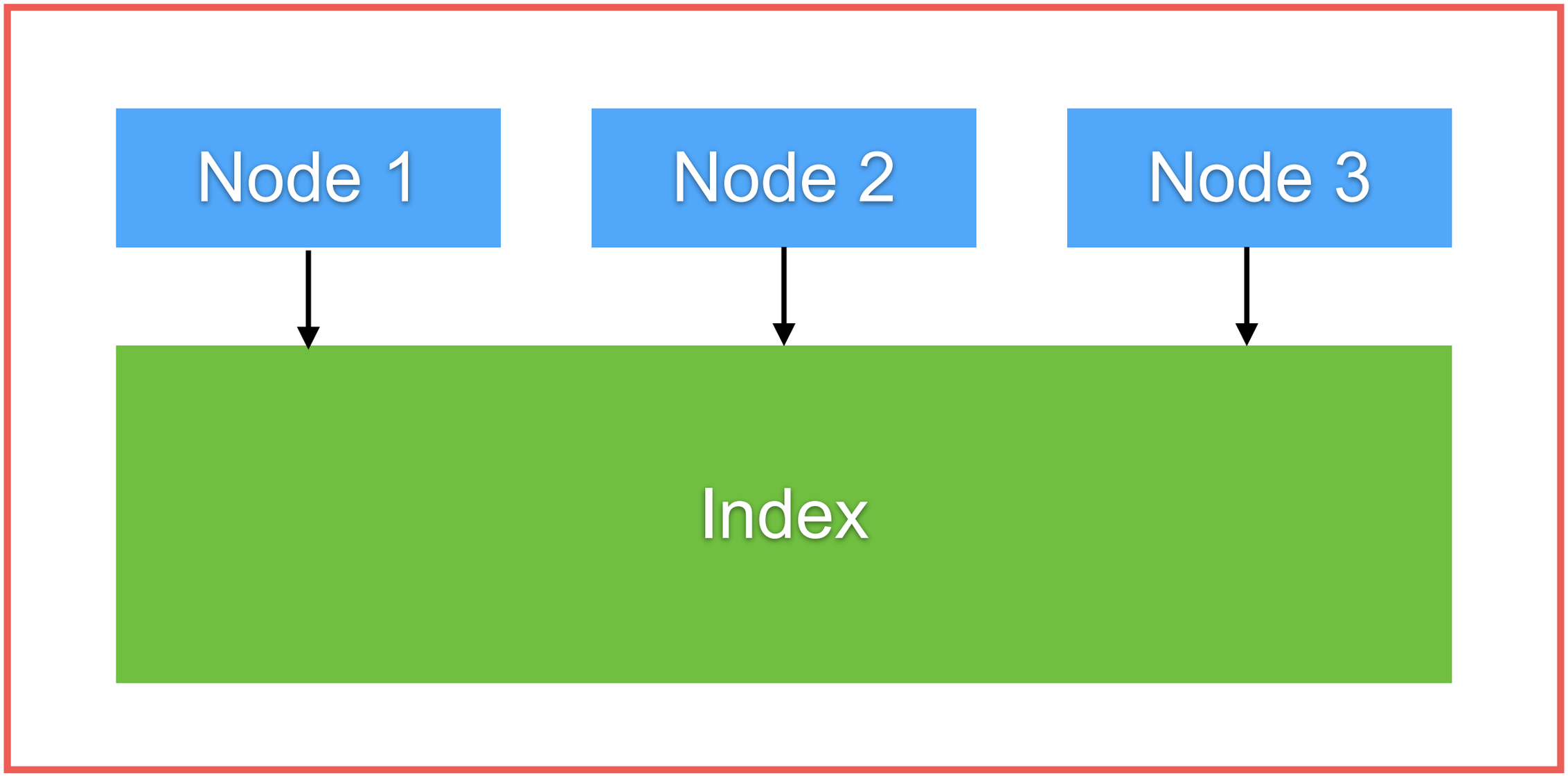
Index

Document

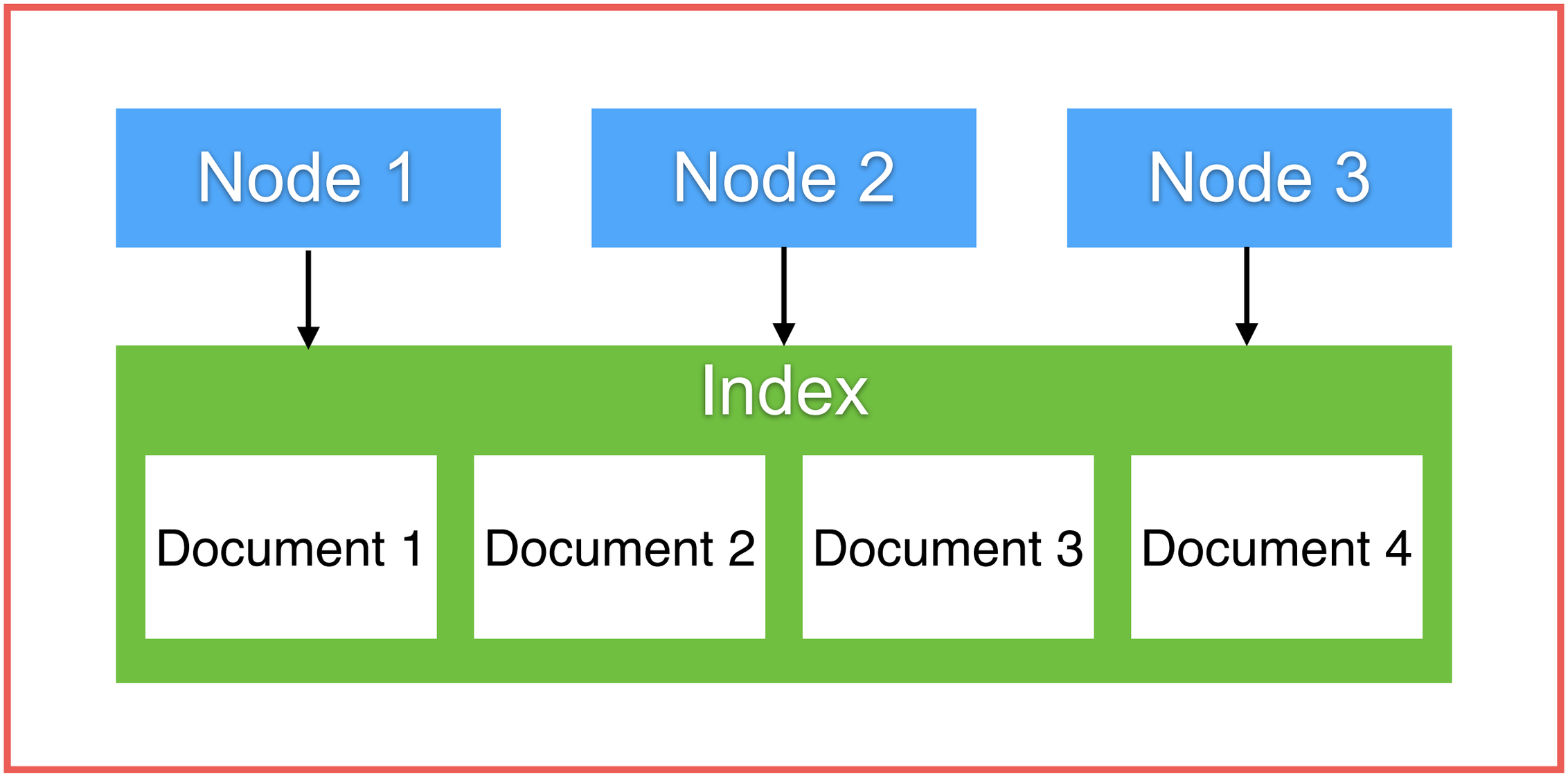
Type

Mapping

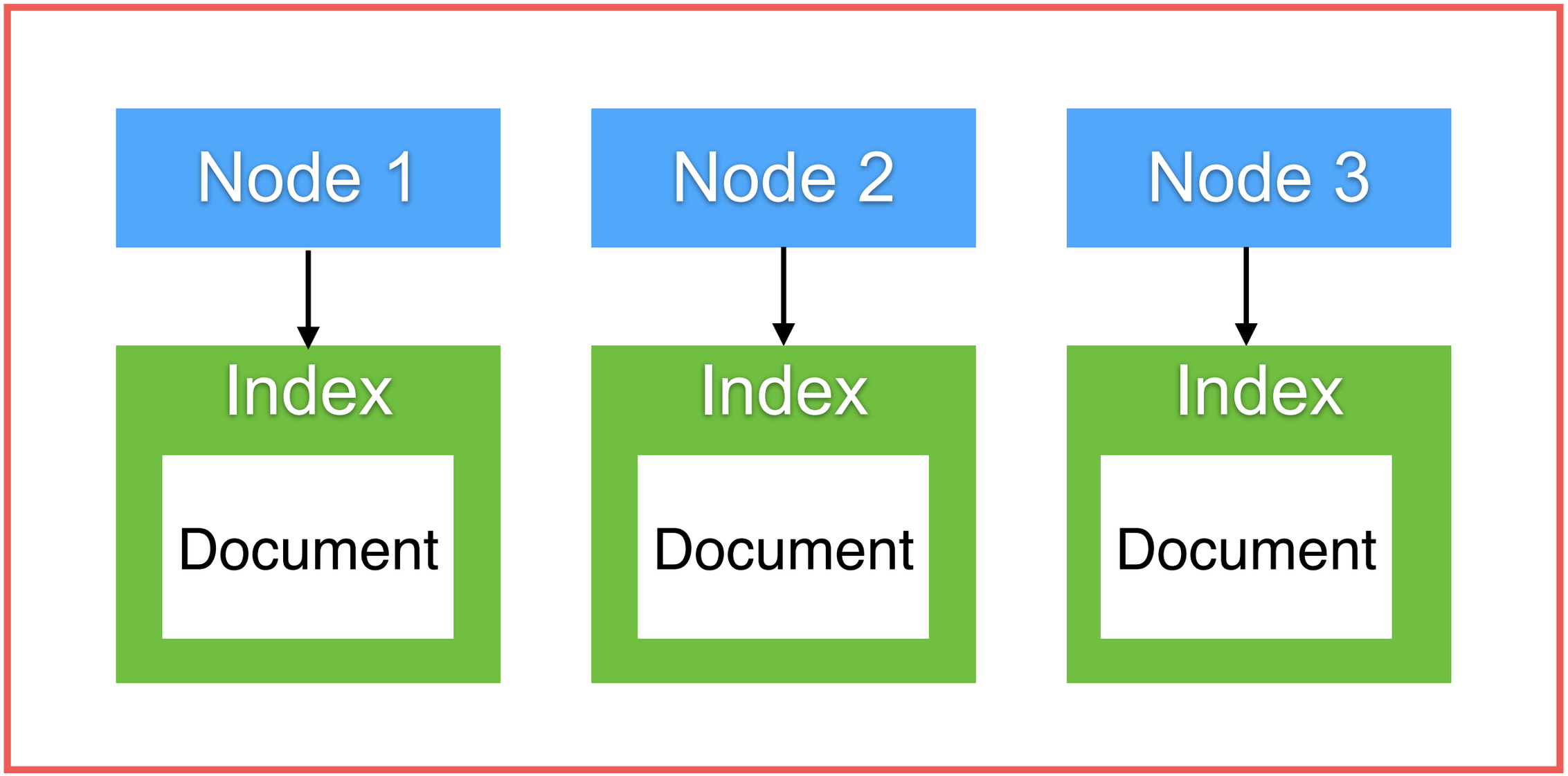
Cluster



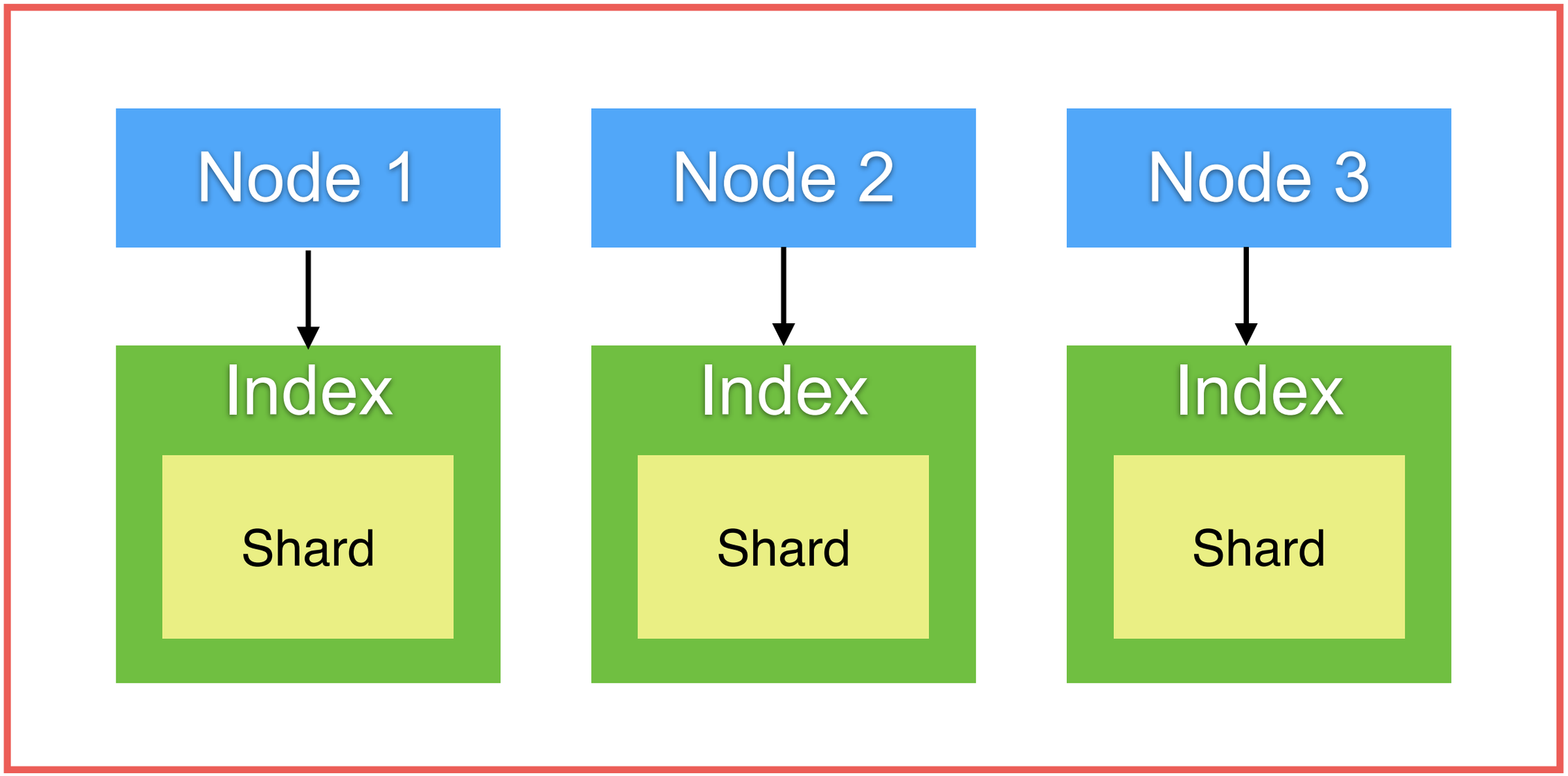
Documents !!



Distributed database



Design for scale (Shard)

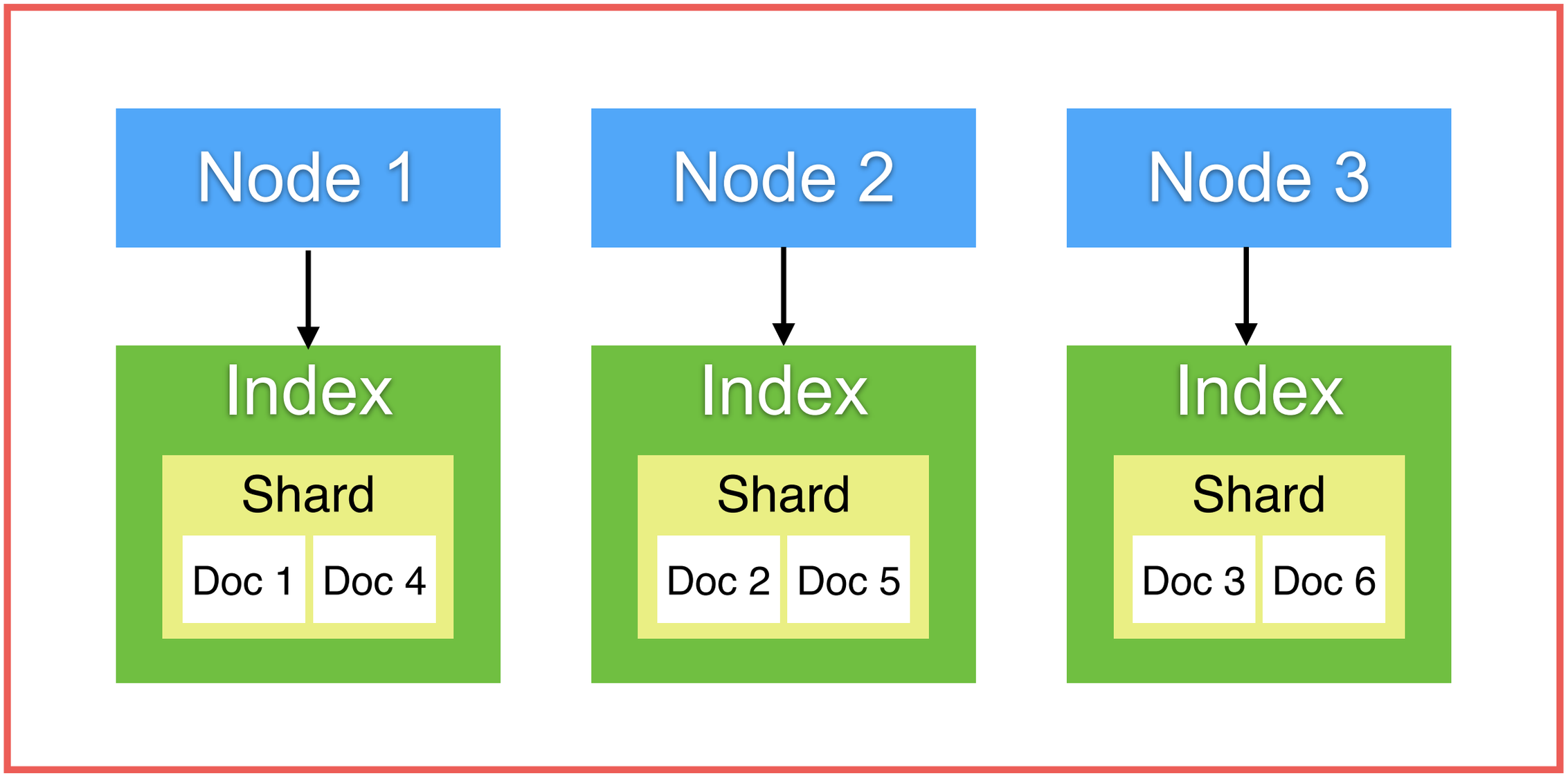


Index is split into shards

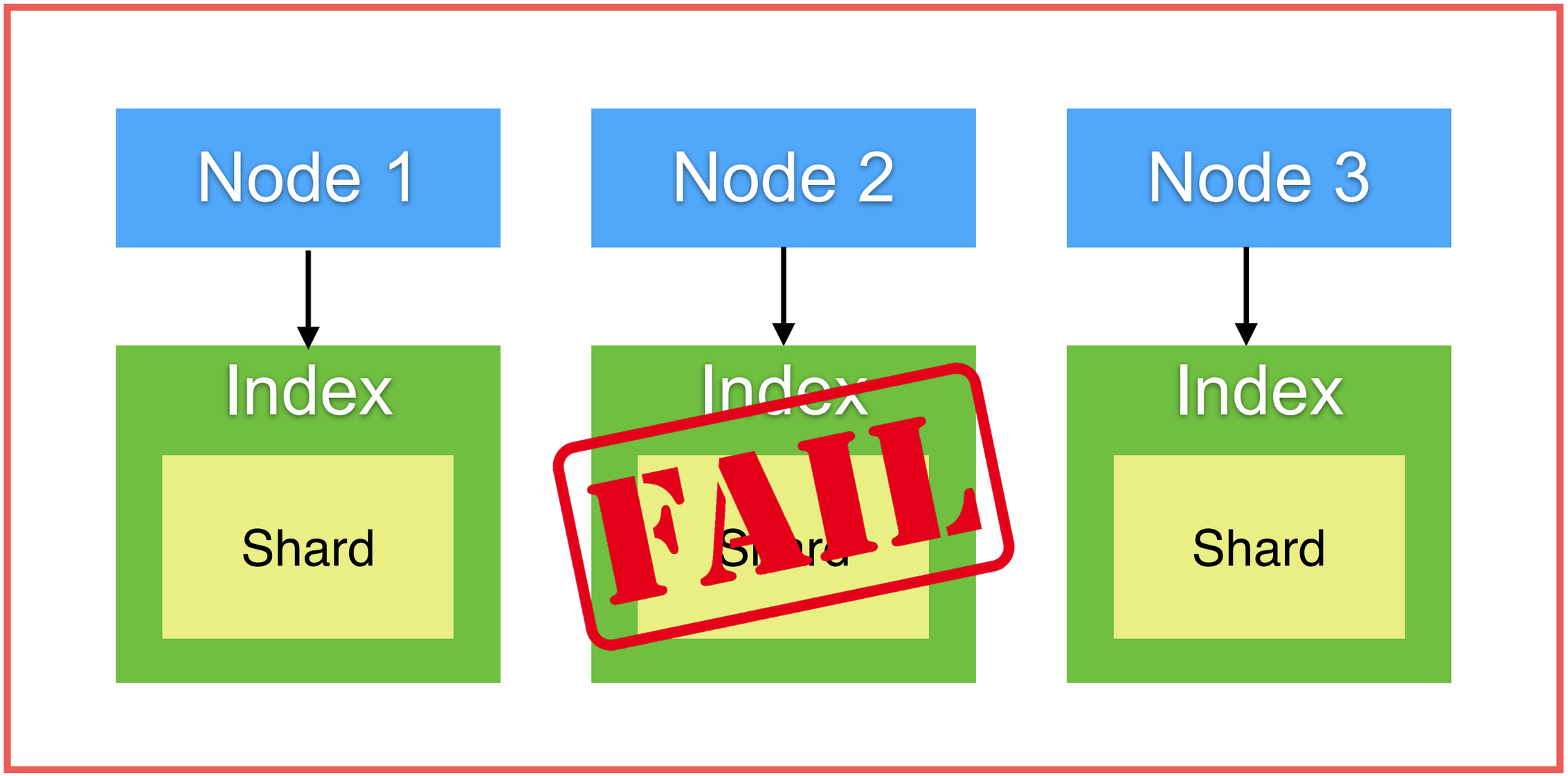
Each shard may be on a different node in cluster

Every shard is a self-contained ***Lucene index***

Documents are hashed



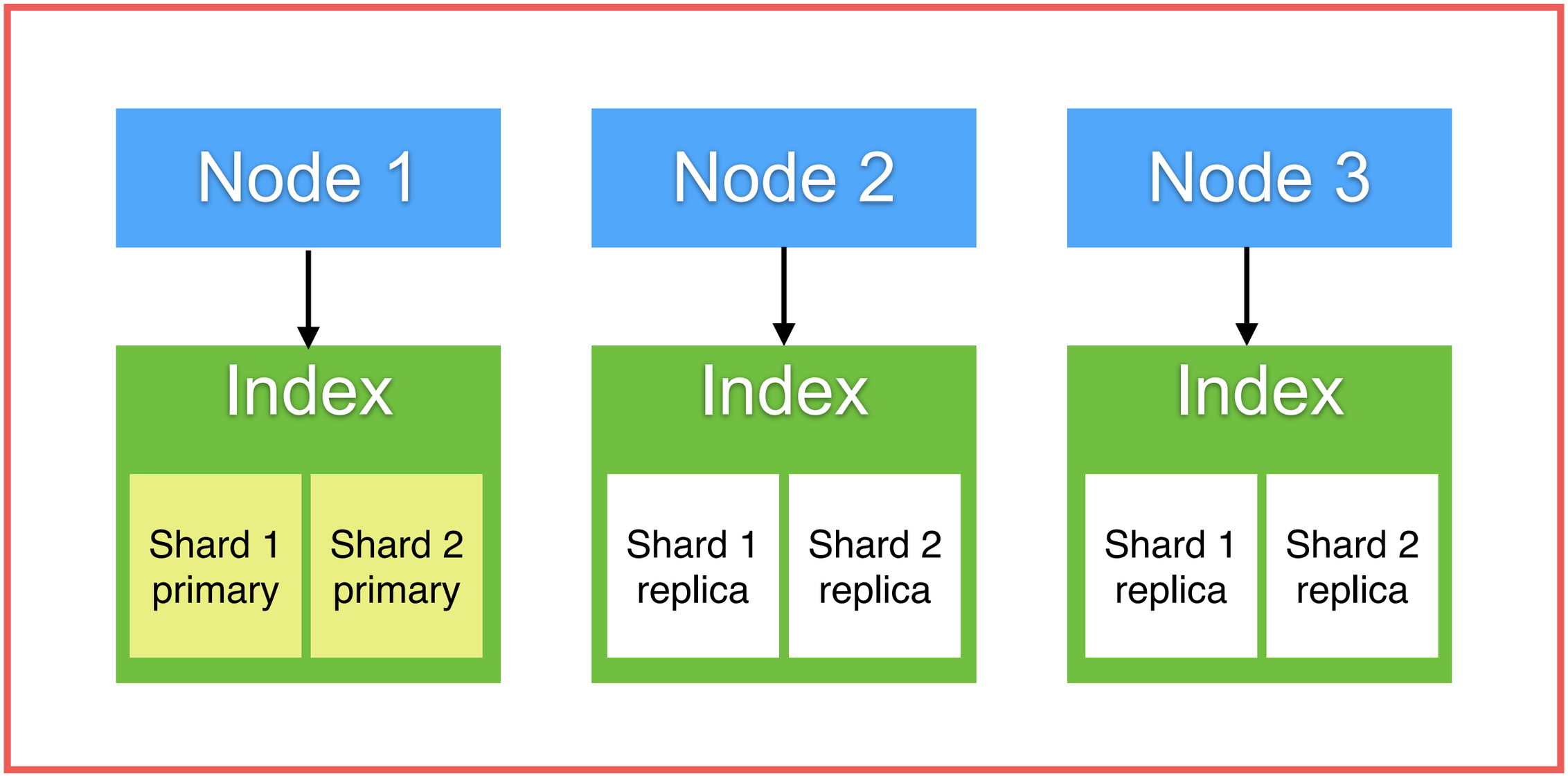
Design for fail (Replica)



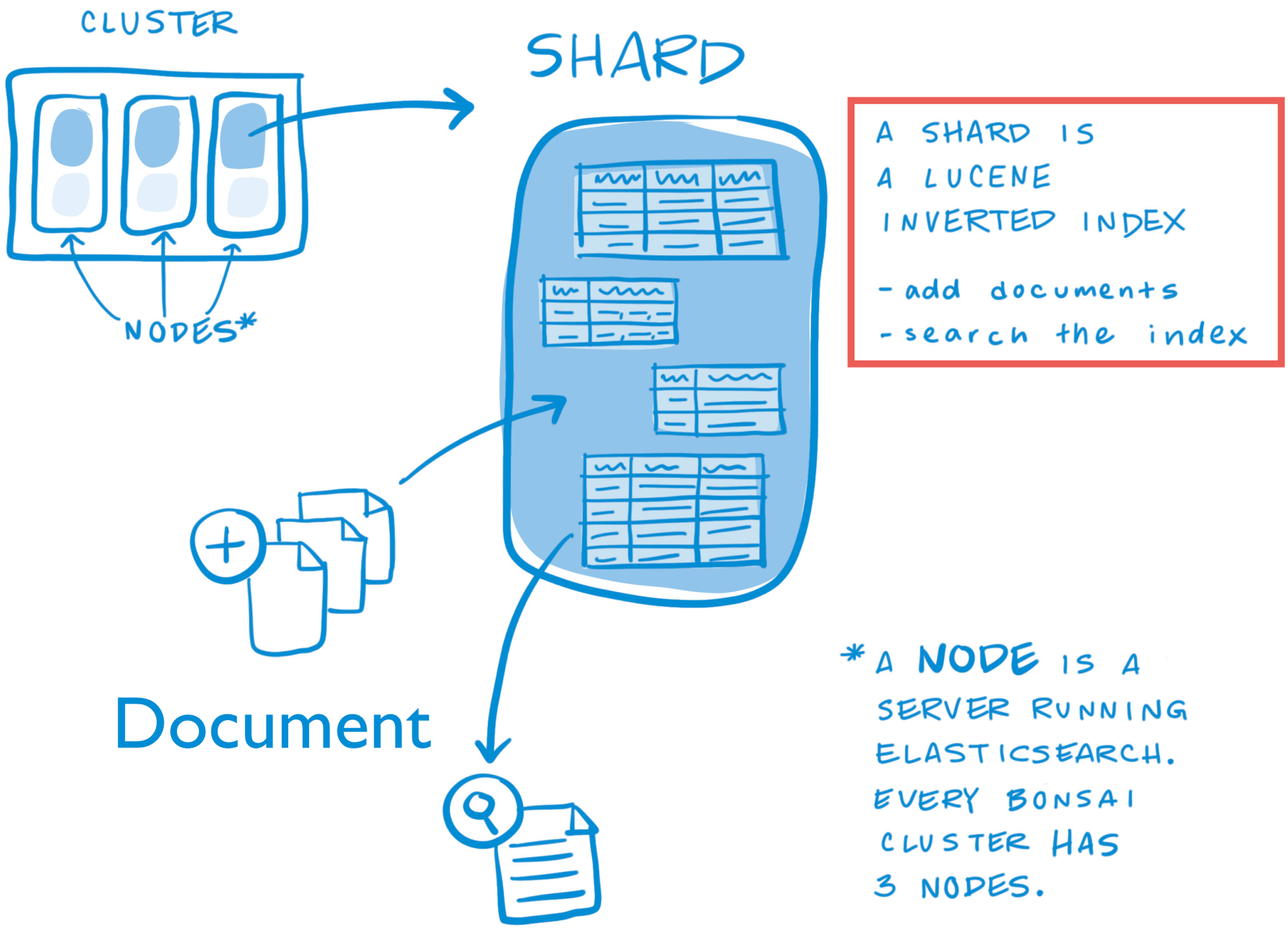
Repica = copy of shards

1 replica = 1 Primary + 1 Replica
2 replicas = 1 Primary + 2 Replicas
3 replicas = 1 Primary + 3 Replicas

Replica = 2



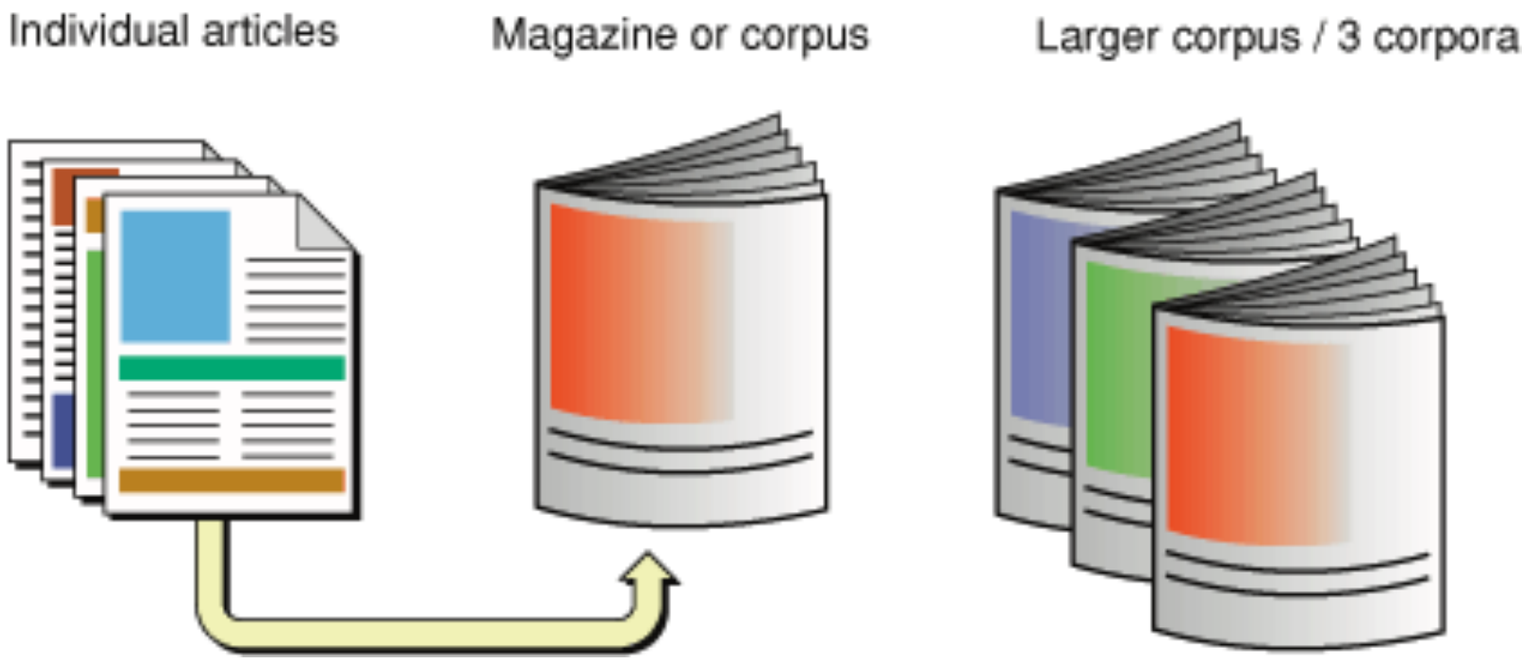
Basic concepts



Inverted Index

Inverted Index

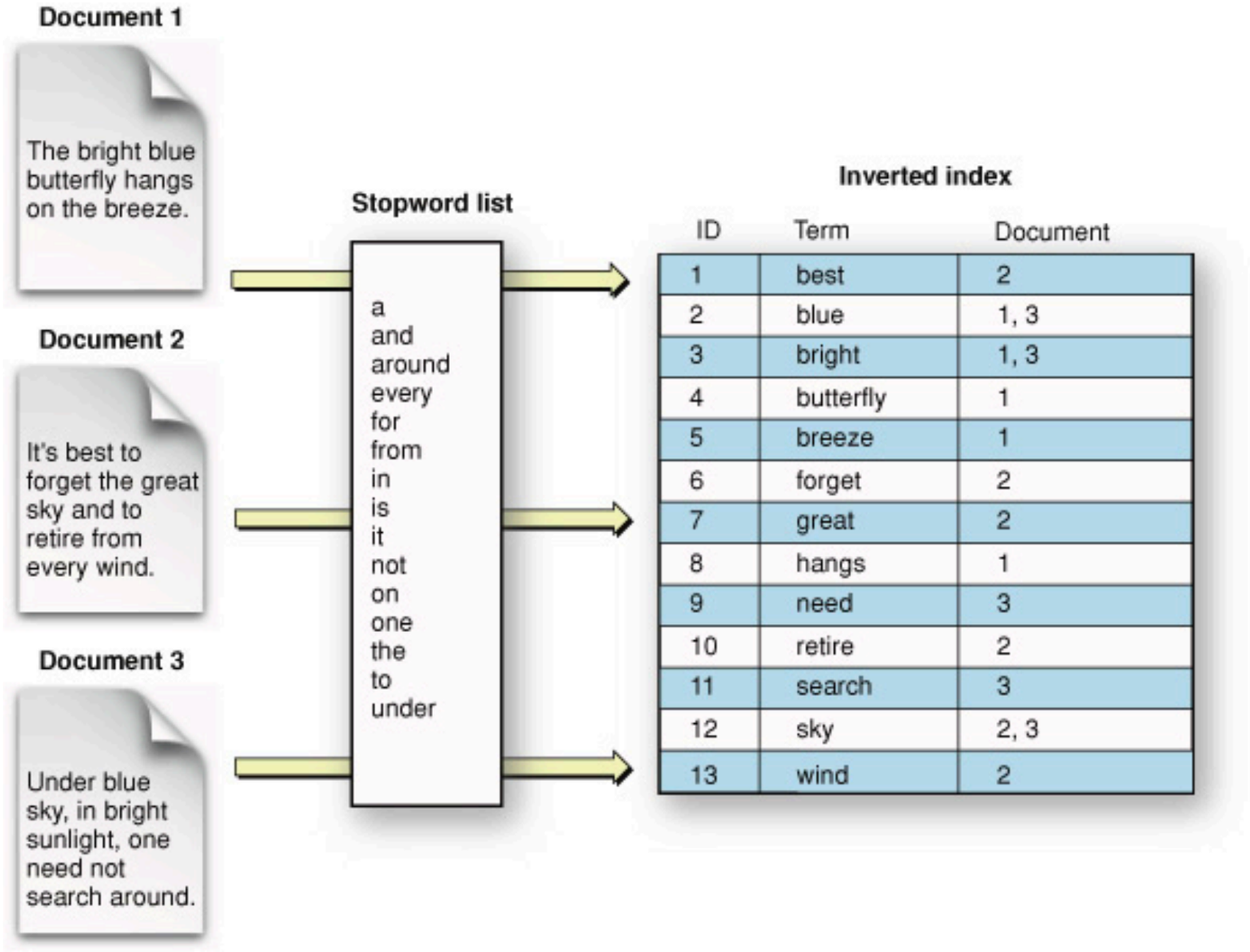
Corpus is a collection of documents



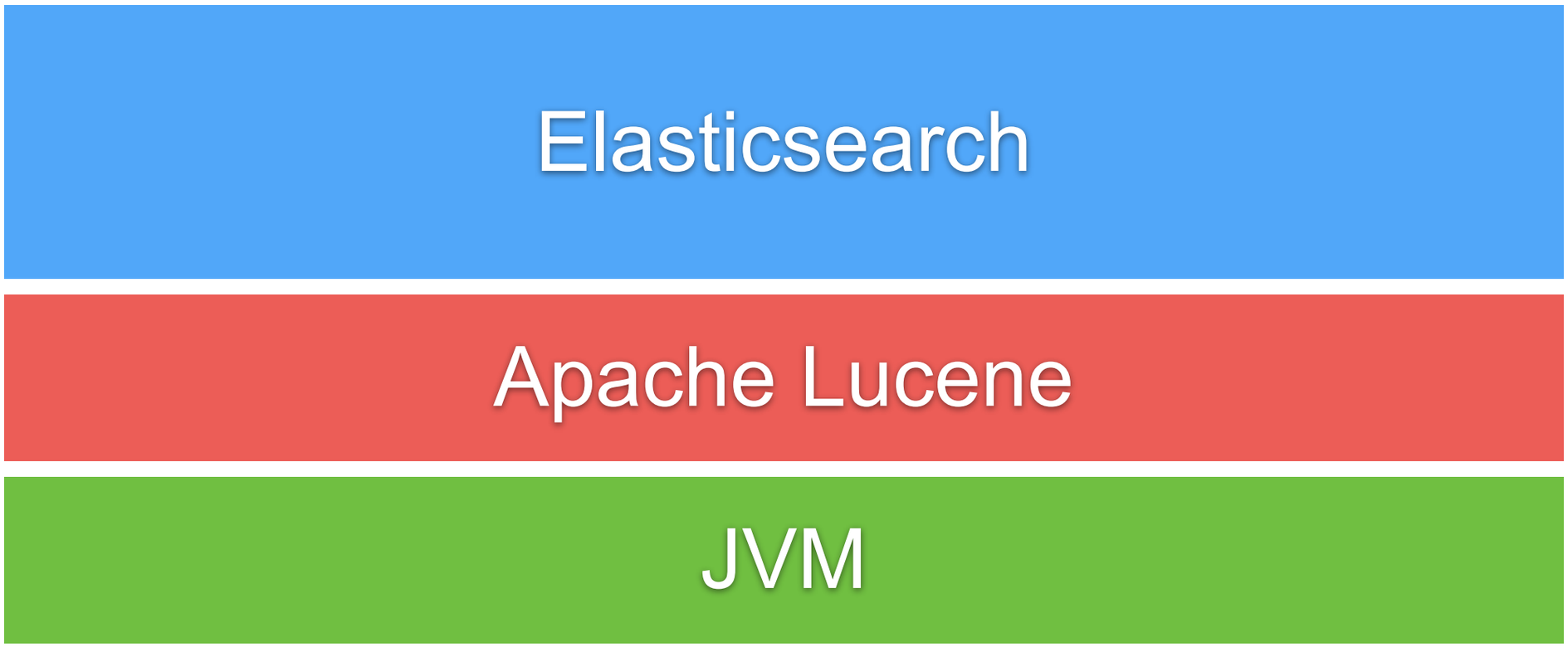
https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html#//apple_ref/doc/uid/TP40002843-TPXREF101

Inverted Index

Try to construct index



Apache Lucene



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**”

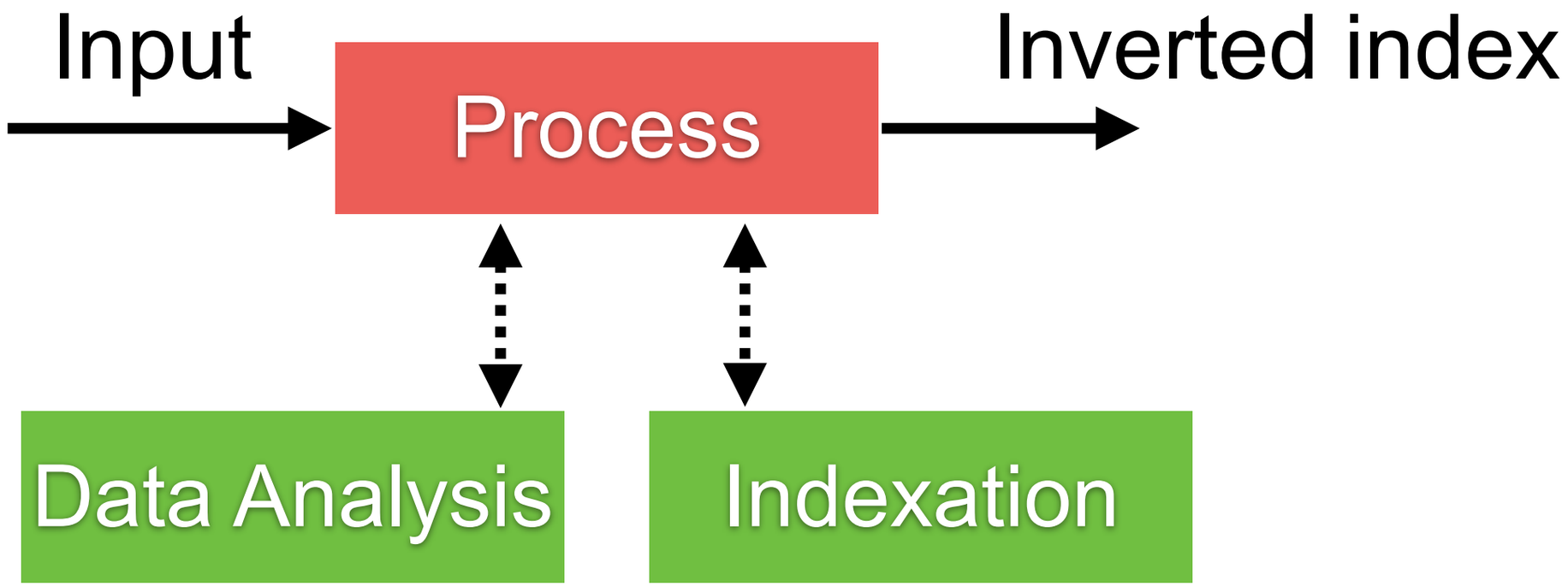
Can’t be delete (just marked to deleted)

Input data analysis

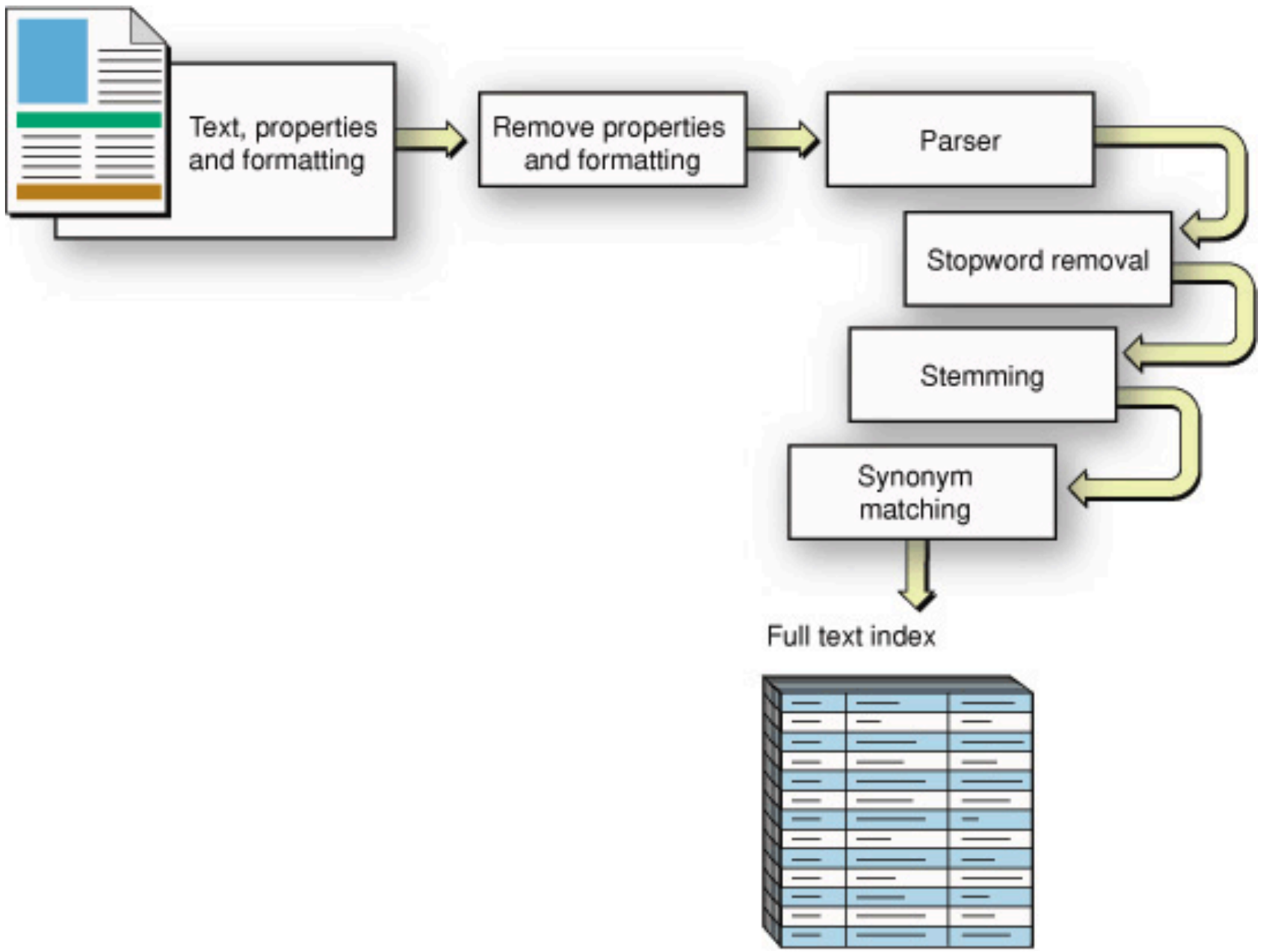
Write-once and read-many-times structure

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

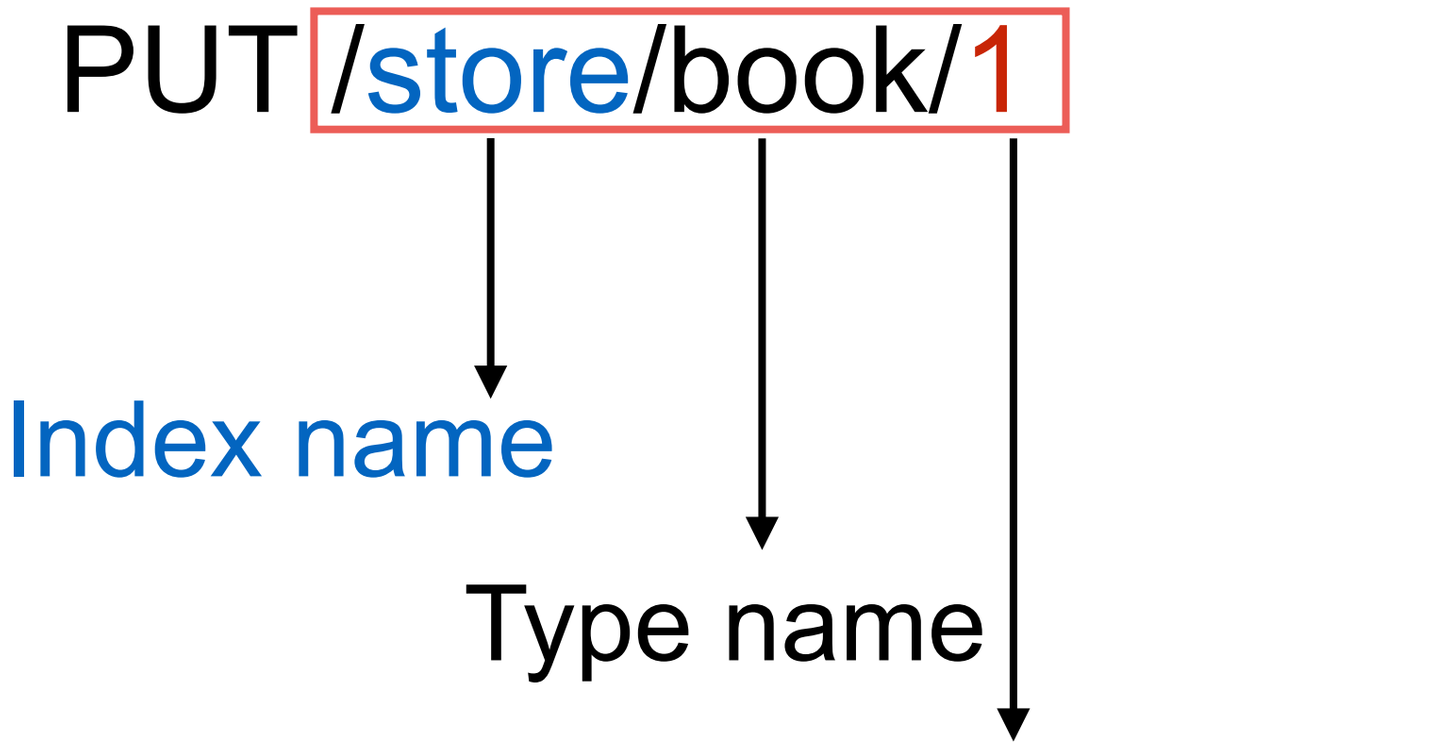
** Only 1 type per index*

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



1 Type per Index

Removal of mapping types

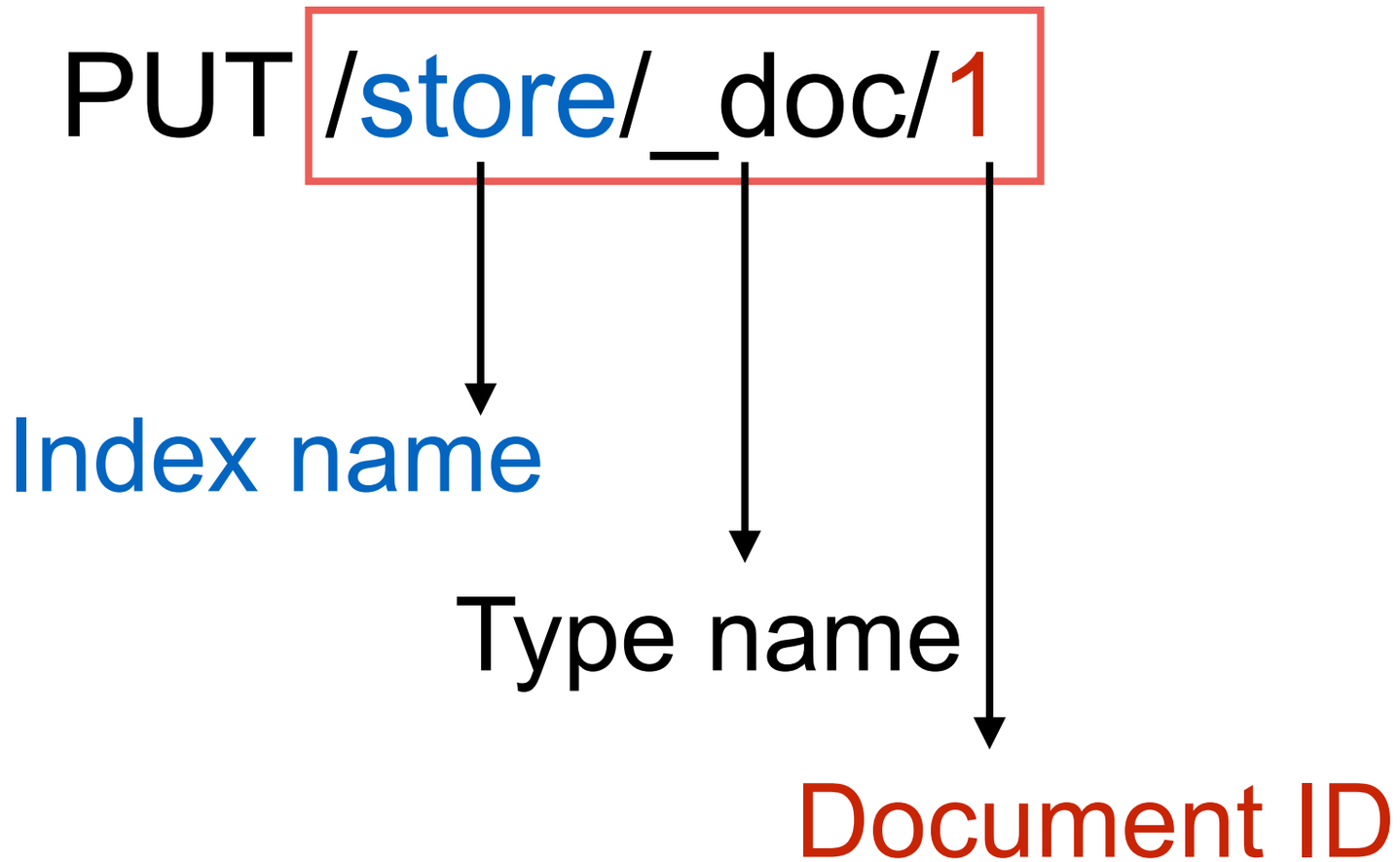
edit

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,
  "_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/_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

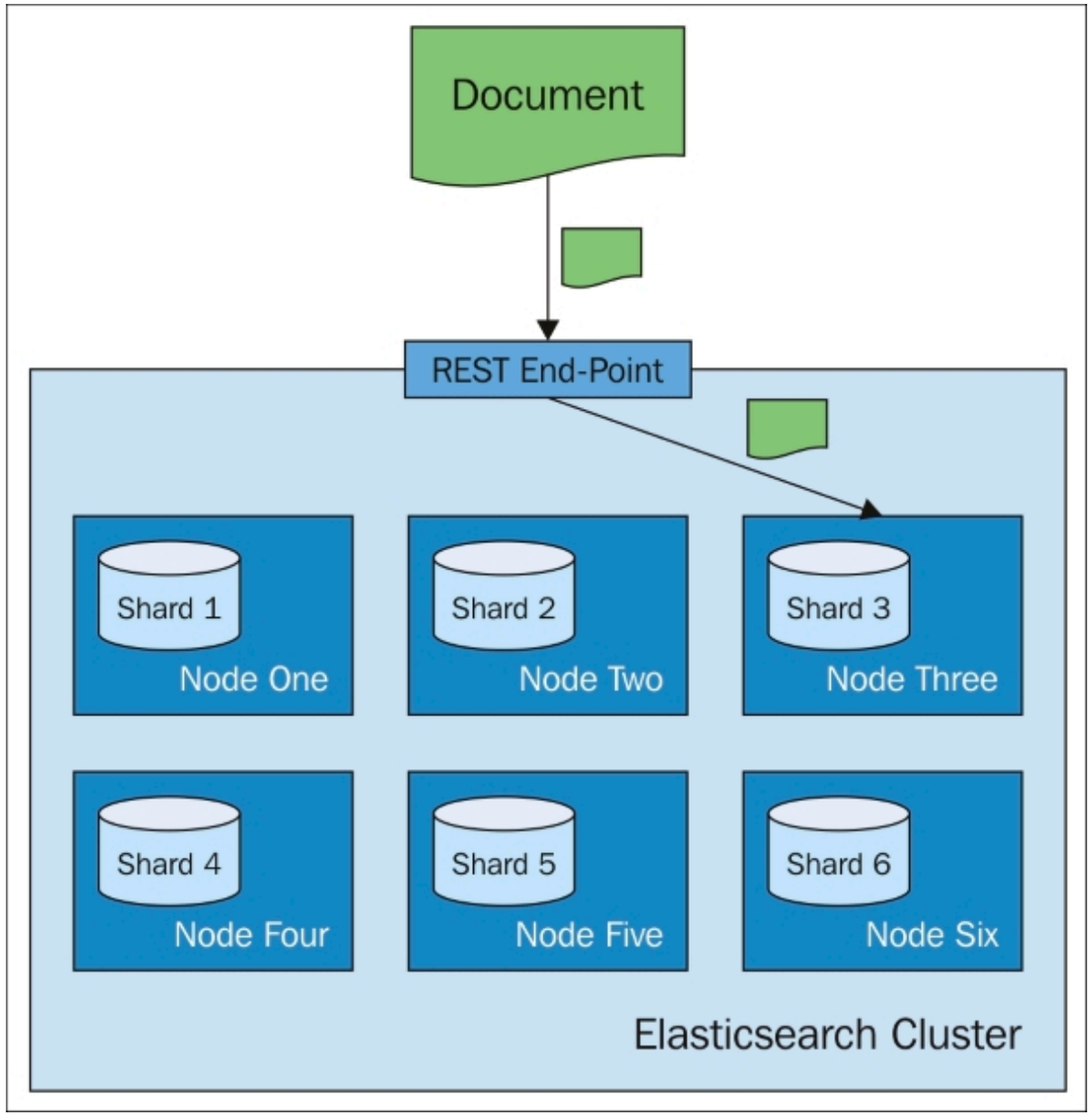
Update by query
Delete by query
Partial update document

Routing

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

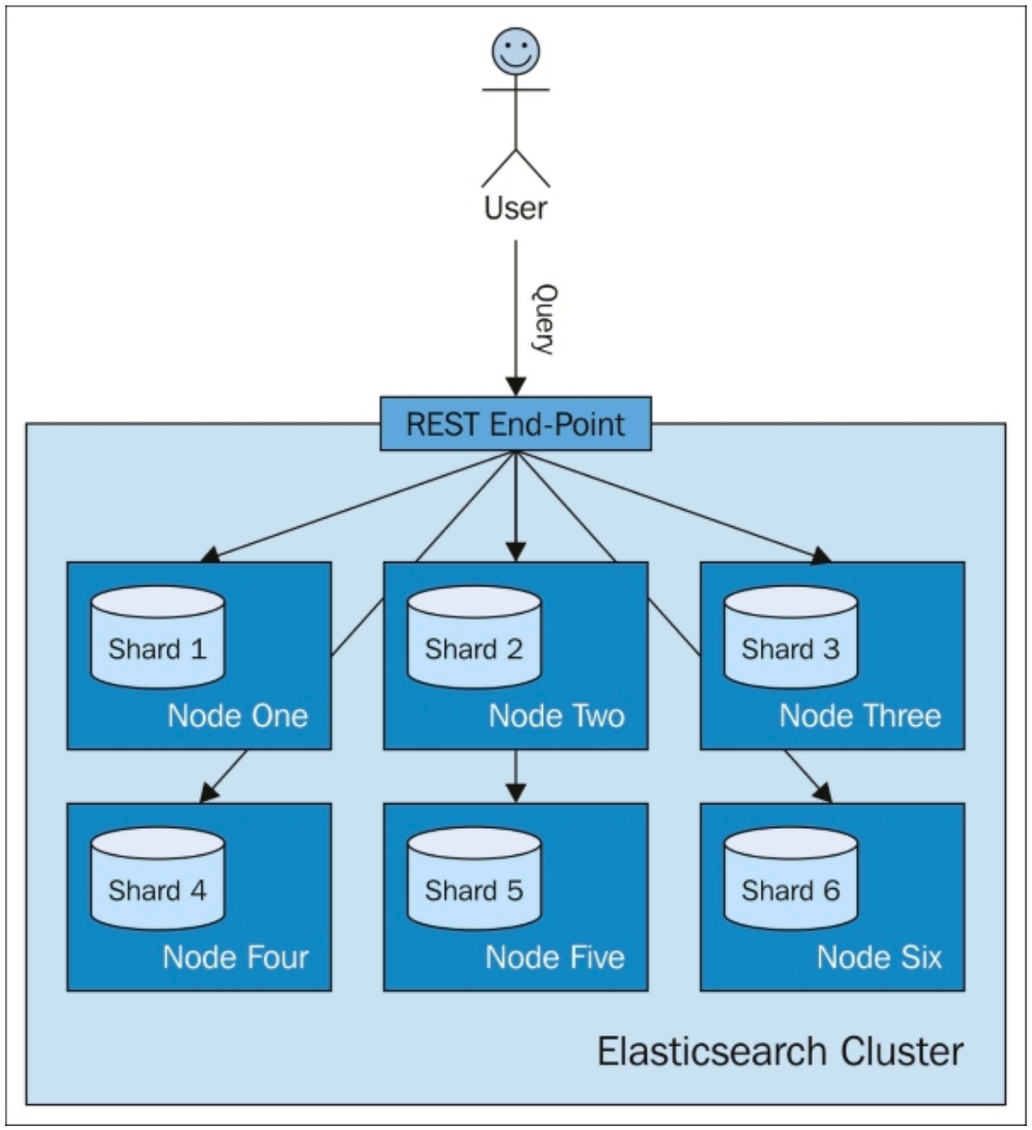
Default indexing

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

shard_num

=

(hash(**_routing**) + hash(**_id**) % **routing_partition_size**)

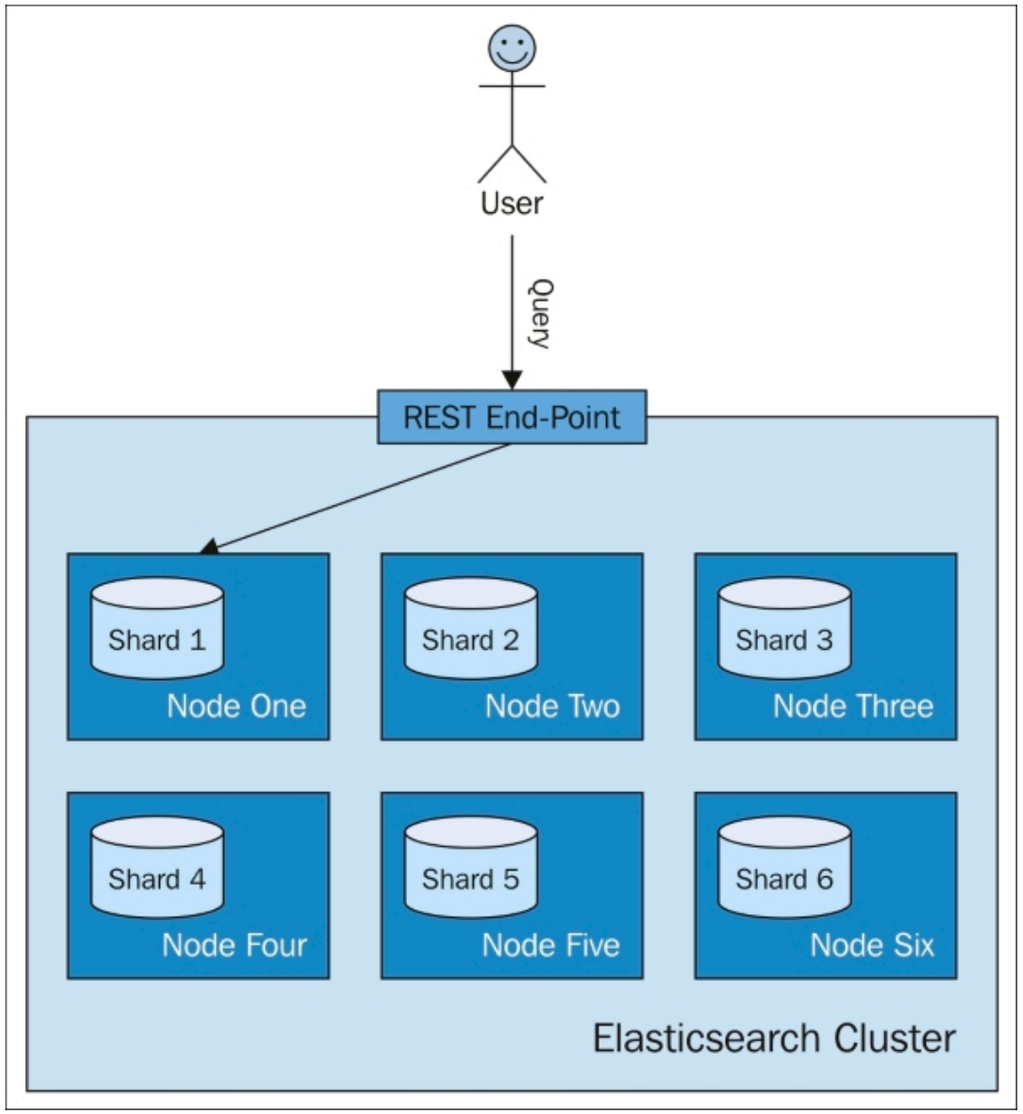
%

num_primary_shards

routing_partition_size = 1 (default)

Custom routing

ES will send our query to a single shard



Bulk API

03-bulk/book_bulk.json

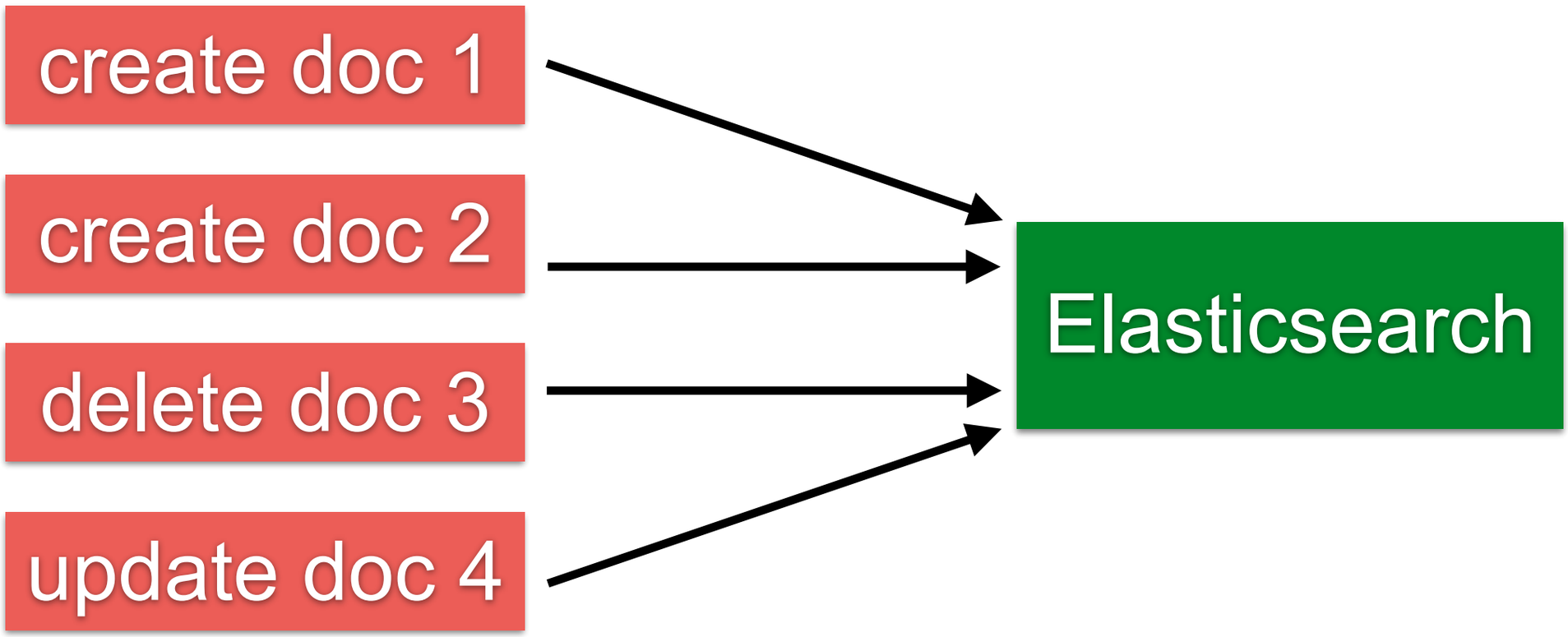
https://www.elastic.co/guide/en/elasticsearch/reference/current/docs-bulk.html

Bulk API

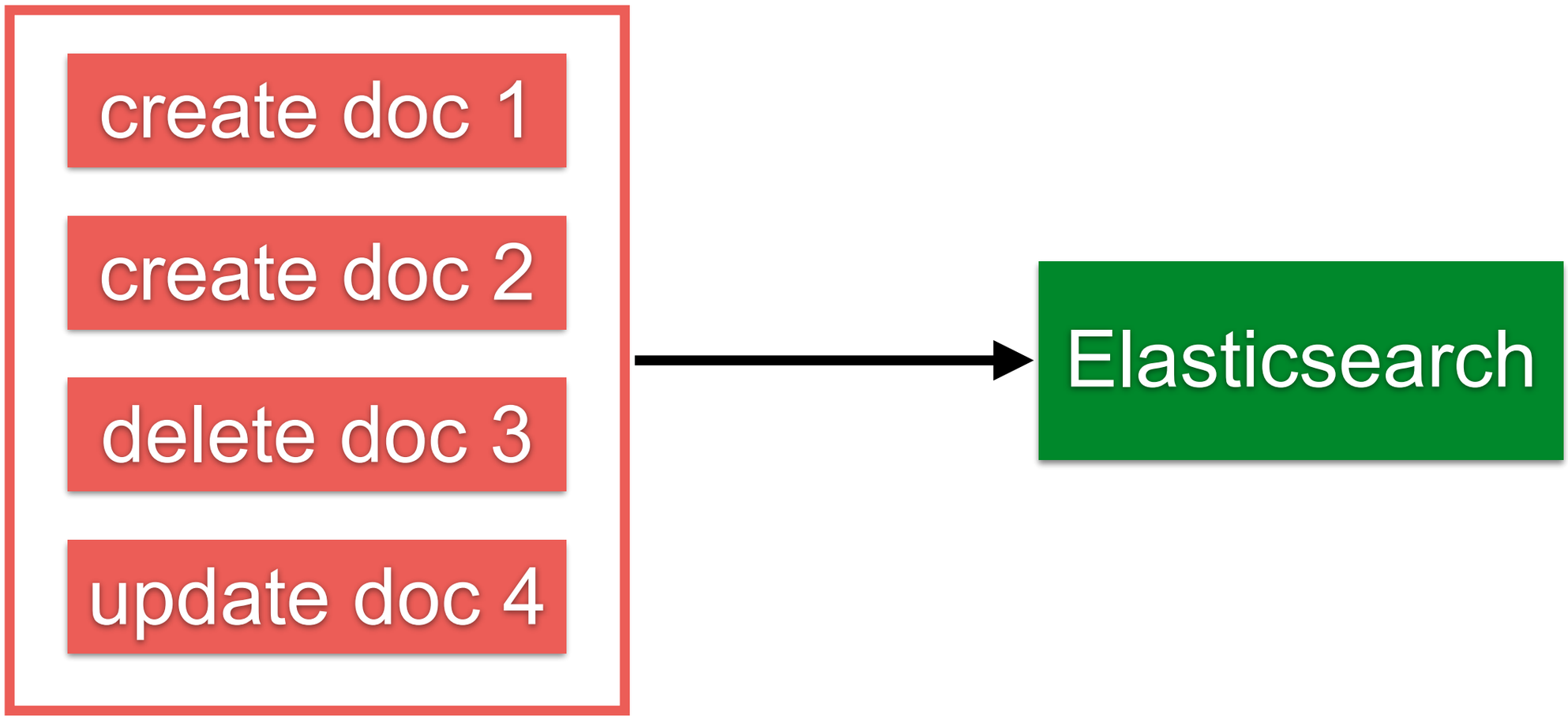
Perform many index/delete operation in single API call

Increase indexing speed

Without Bulk API

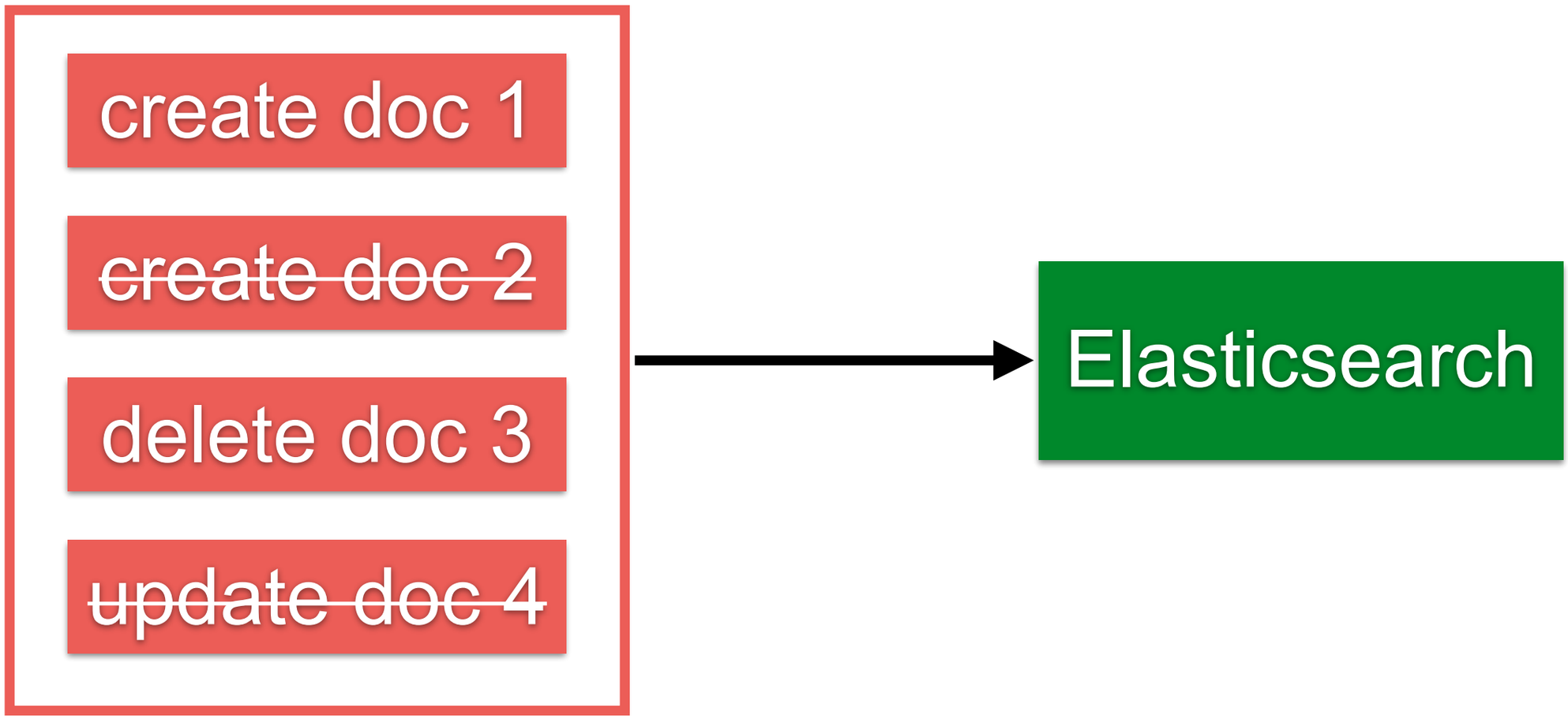


With Bulk API



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

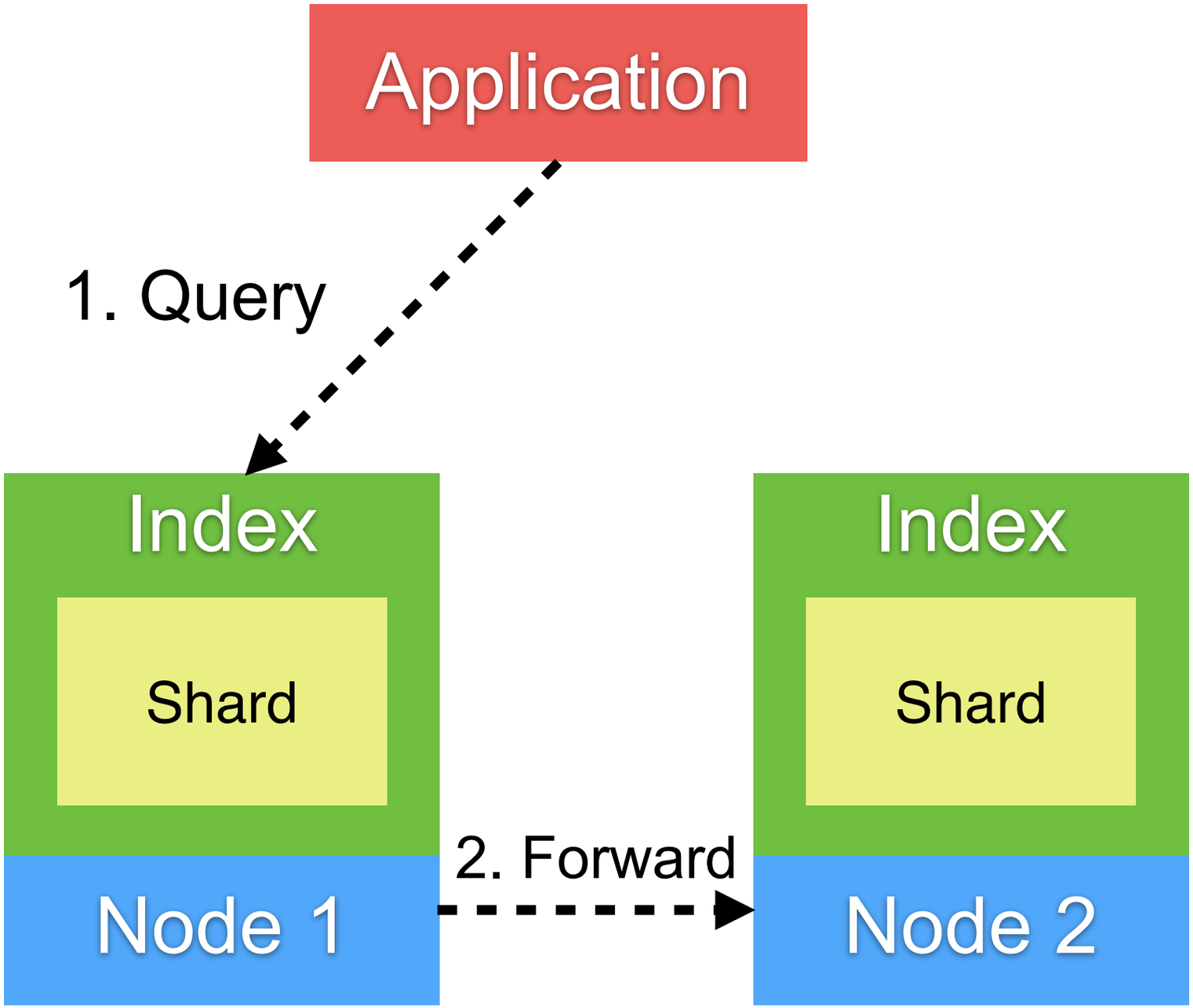
HTTP Status 201 = Created

Searching / Query

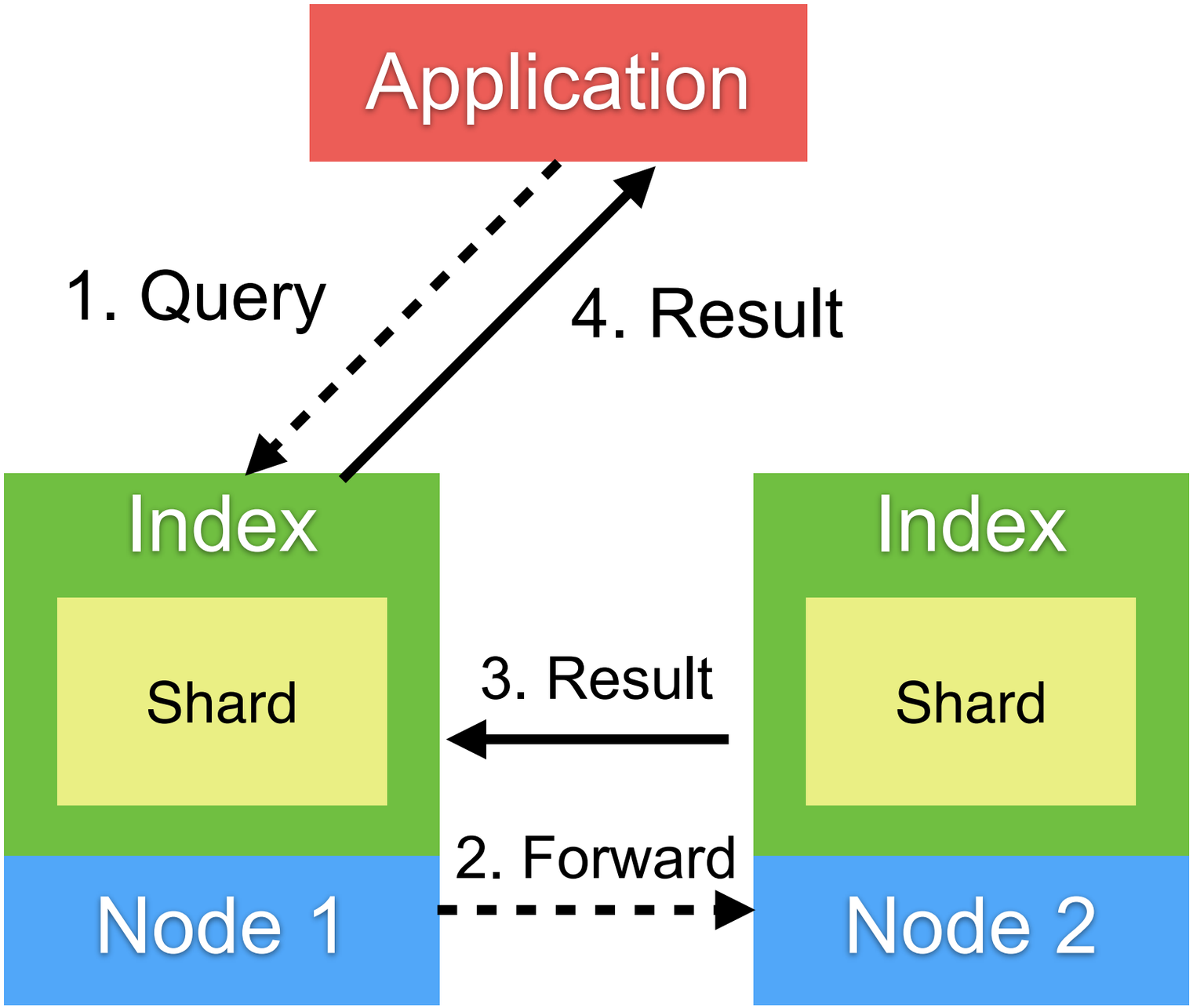
Searching

Scattering phase
Gathering phase

Scattering phase



Gathering phase



Query DSL

04-search/book_search.json

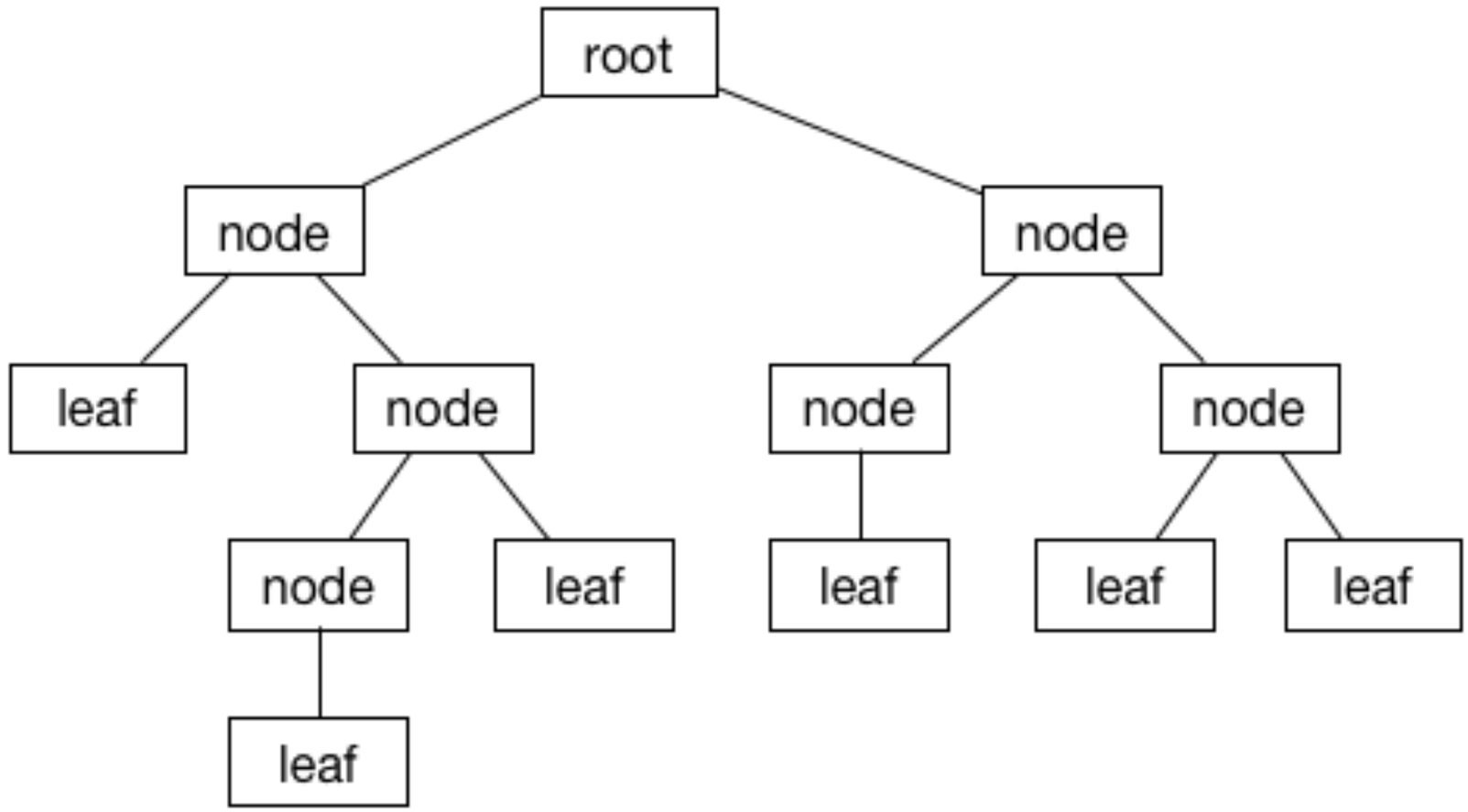
Query DSL

Domain **S**pecific **L**anguage for query data
Flexible query language
Based on JSON format

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

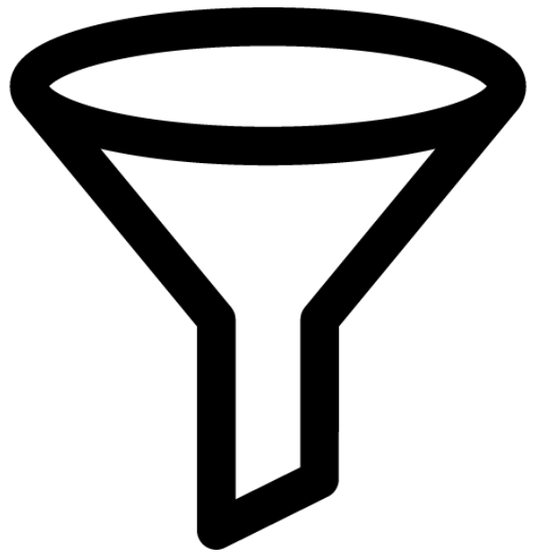
Query DSL

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": [{}]
      }
   }
}
```

_all field is disabled (ES 6.0+)

`_all` field

Deprecated in 6.0.0.

`_all` may no longer be enabled for indices created in 6.0+, use a custom field and the mapping `copy_to` parameter

Space issue
Analyzer issue
Highlighting issue

https://www.elastic.co/guide/en/elasticsearch/reference/6.5/mapping-all-field.html

Workshop

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 | 0 Cart

1-16 of 119 results for "elasticsearch"

Sort by Featured

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 Administrators Guide

3

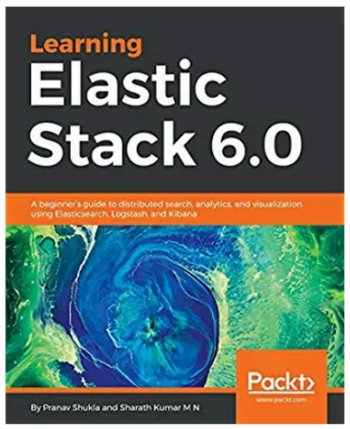
prime

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.

7

Paperback

$34.99

In Stock

Previous Page 1 2 3 ... 8 Next Page

Query

Sorting

Filter

Paging

amazon

All elasticsearch

New to Amazon?

Deliver to Thailand

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

EN Hello. Sign Account & Lists Orders Cart

1-16 of 119 results for "elasticsearch"

Sort by Featured

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

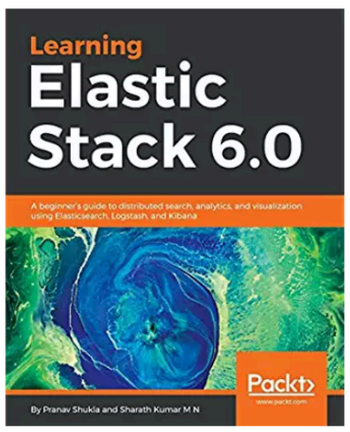
prime

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 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.

7

Paperback

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

Aggregation name

Count by category

GET /store/_search

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

"terms": { Aggregation type

Result of aggregation

Search result


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


Result of aggregation

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

Aggregation result

Show only aggregation result

GET /store/_search

```
{
  "size": 0,
  "aggs": {
    "all_book_title": {
      "terms": {
        "field": "category.keyword"
      }
    }
  }
}
```

Set search result size = 0

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 }
        ]
      }
    }
  }
}
```

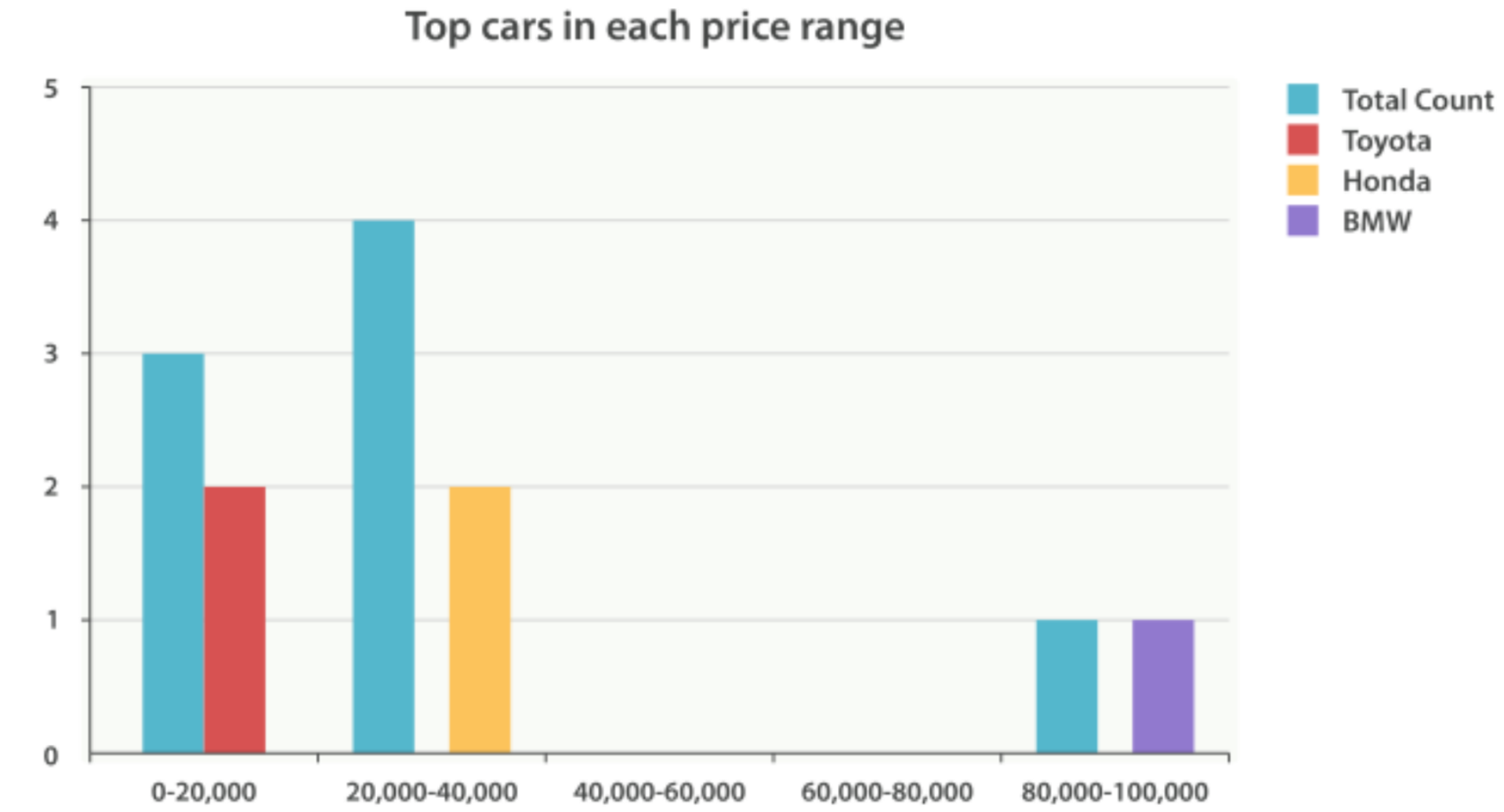
Workshop aggregation with car

05-aggregation/car.json

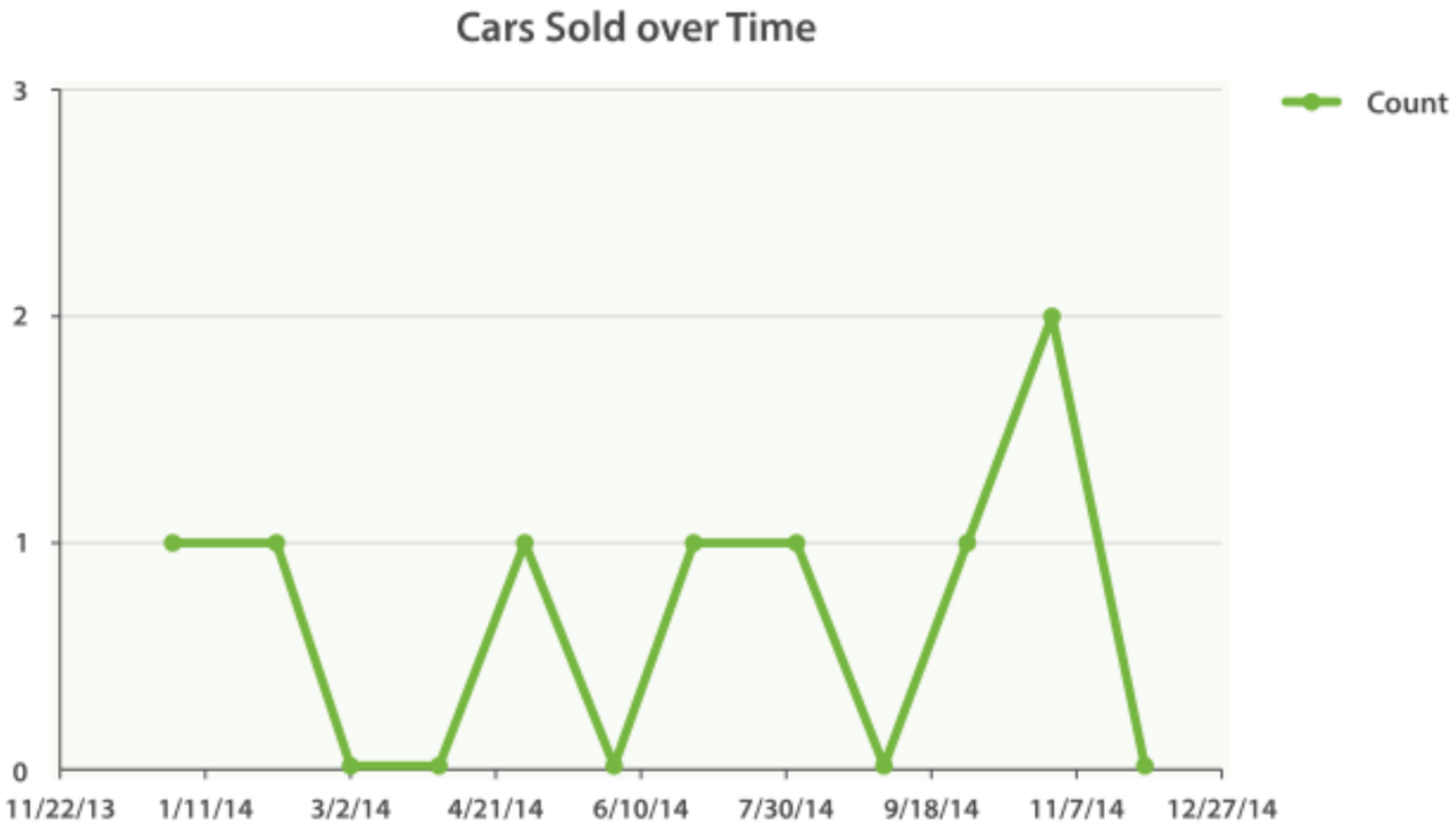
Try by yourself

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 ?



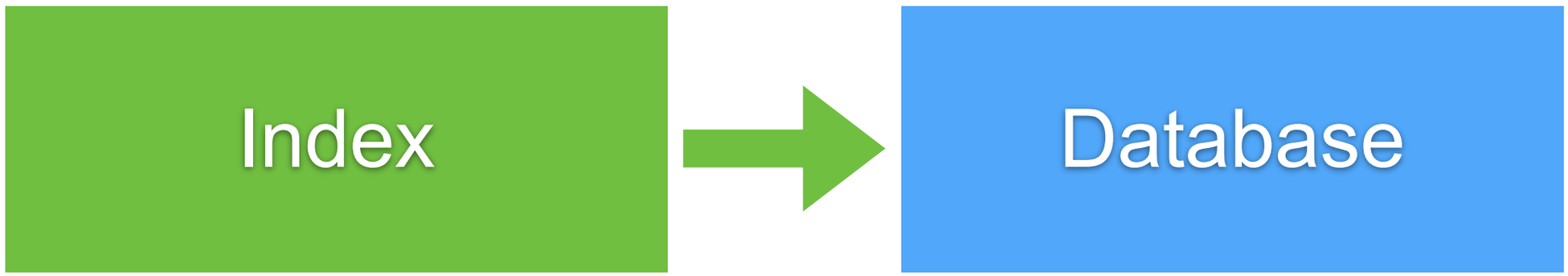
Cars sold over Time ?

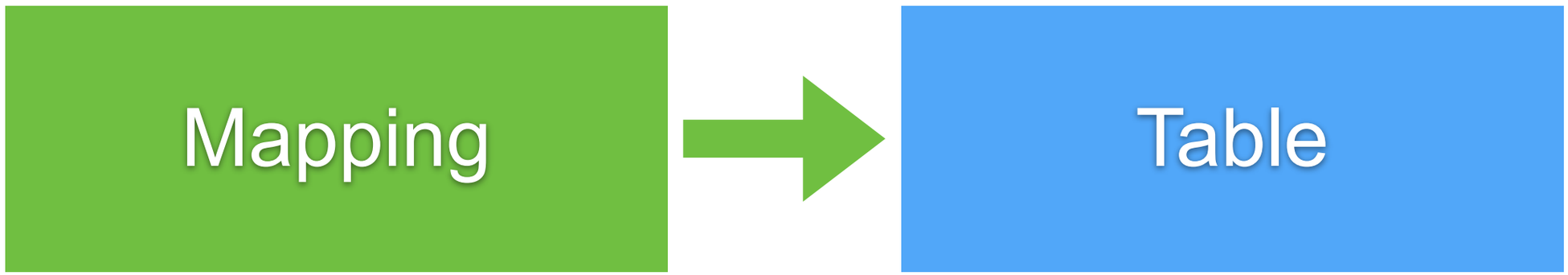


Mapping

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

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

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

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

Better search Results

Increased performance

Concerned of the field types

Explicit mapping

Quick data insertion

Custom Mapping in ES 7

Don’t specify the type name !!

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

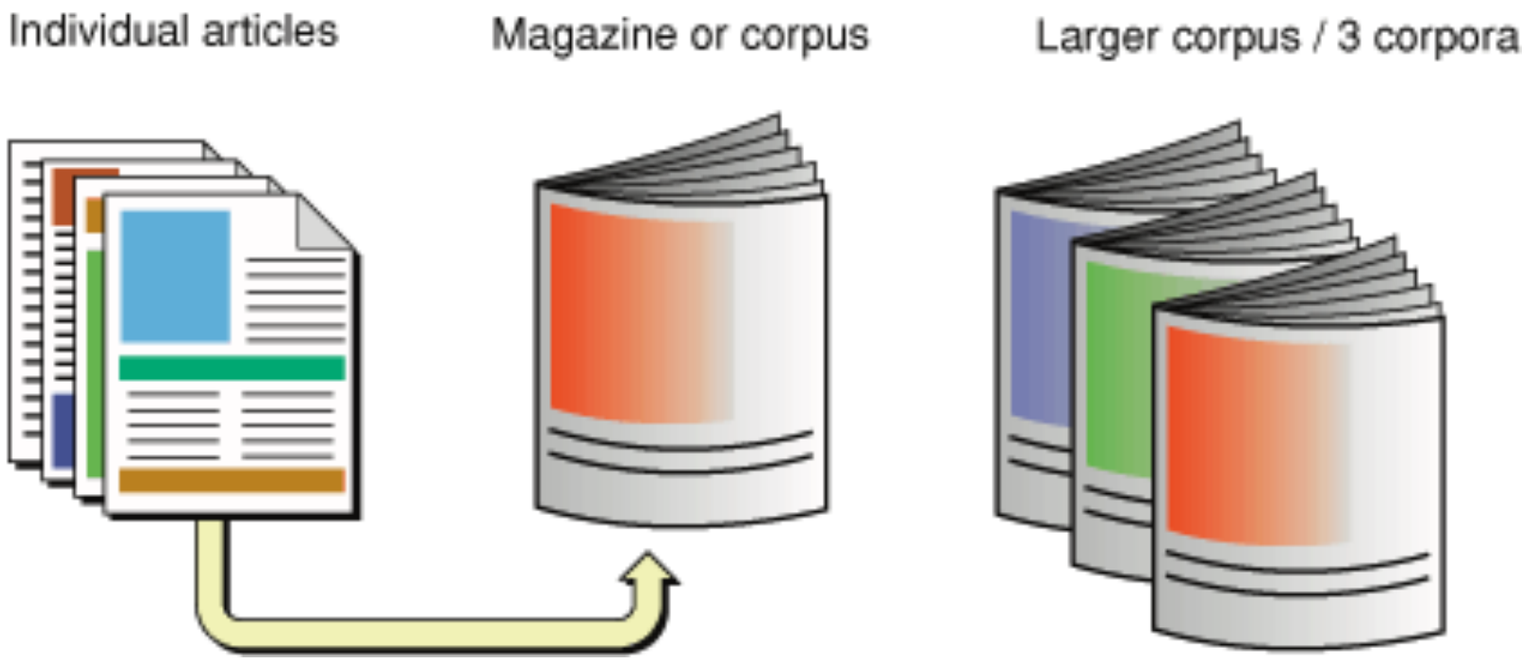
Analyzer

07-analyzer

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

Inverted Index

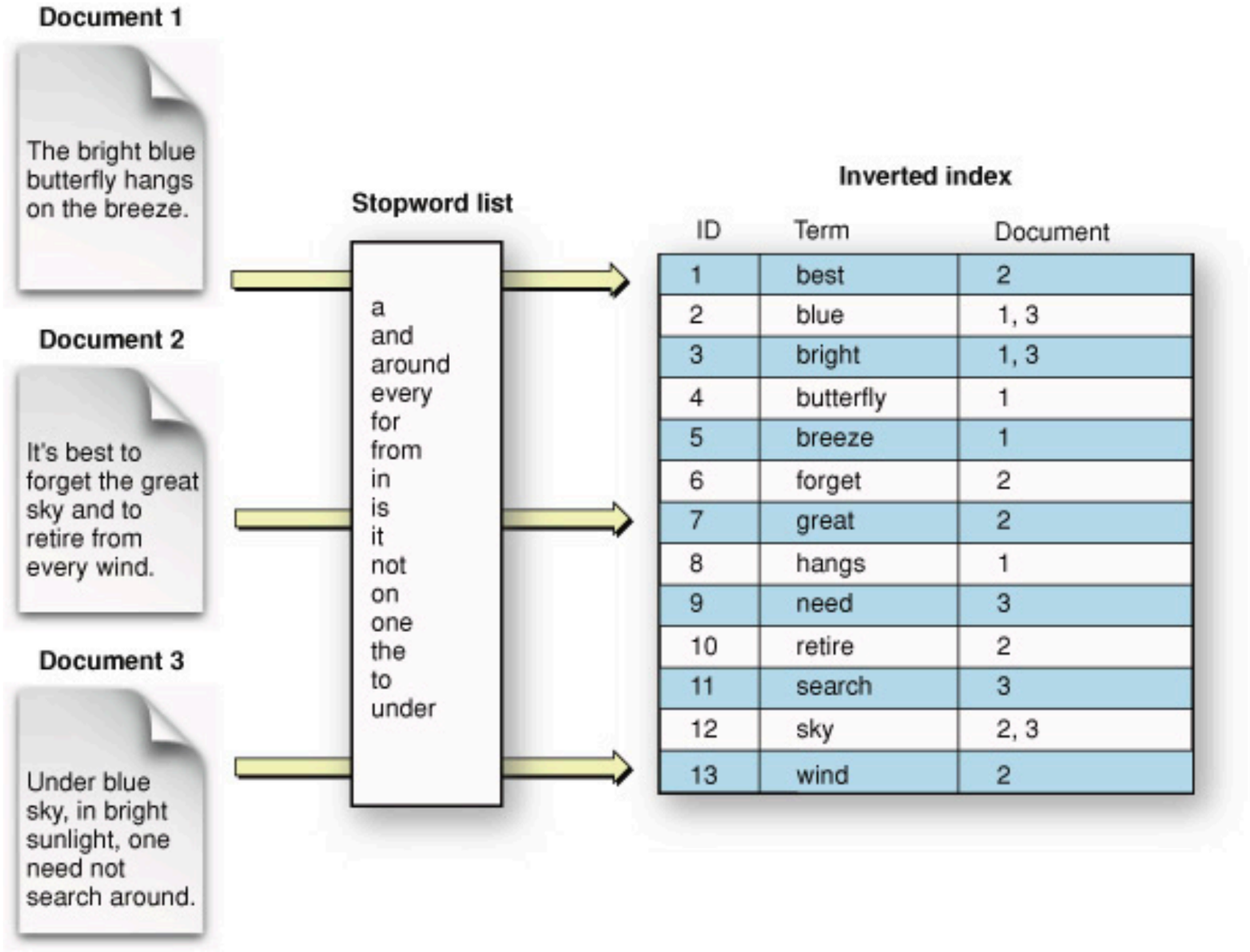
Corpus is a collection of documents



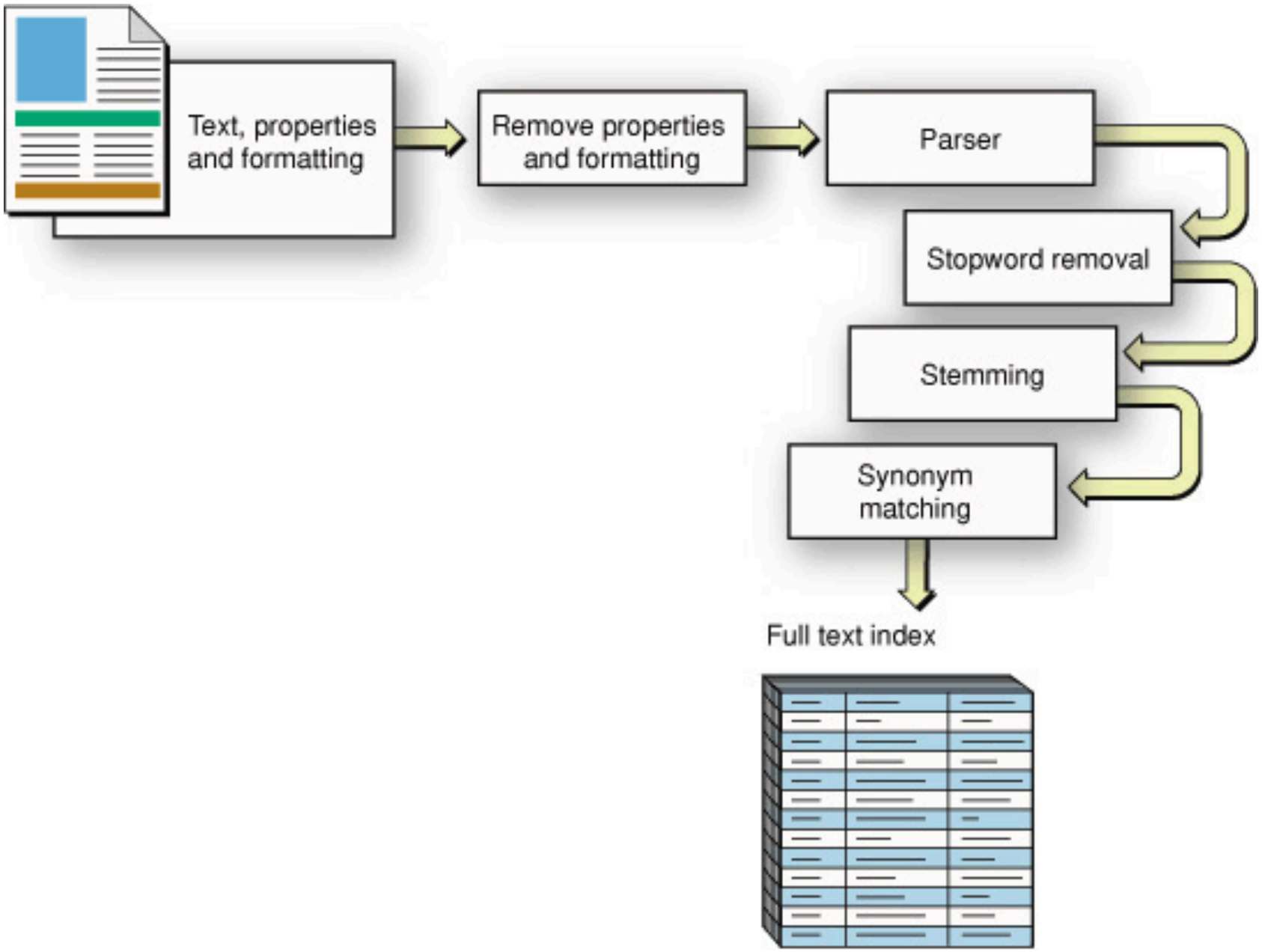
https://developer.apple.com/library/archive/documentation/UserExperience/Conceptual/SearchKitConcepts/searchKit_basics/searchKit_basics.html#//apple_ref/doc/uid/TP40002843-TPXREF101

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":     "สวัสดีประเทศไทย"
}
```

Tokenizer and filter

```
POST _analyze
{
  "tokenizer": "standard",
  "filter":  [ "lowercase", "asciifolding" ],
  "text":      "Is this déja 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 tokentizer

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

N-Gram

Terms as a sequence of n words

search

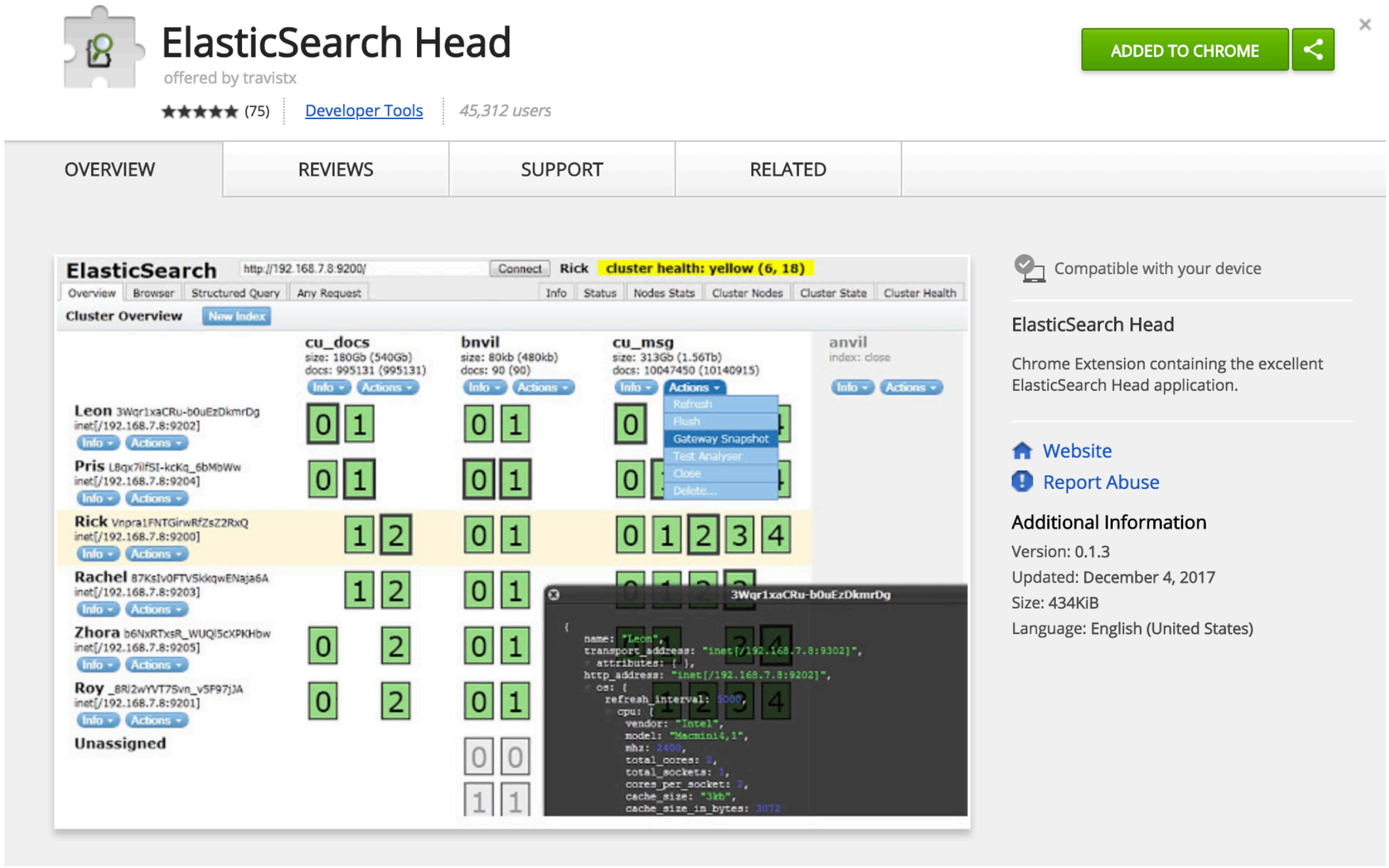
Unigram	s,e,a,r,c,h
Bigram	se, ea, ar,rc, ch
Trigram	sea, ear, arc, rch
4-gram	sear, earc, arch
5-gram	searc, earch

Workshop

ngram/ngram.json

More tools

Elasticsearch Head



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

Elasticsearch Dump

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

Make Logs

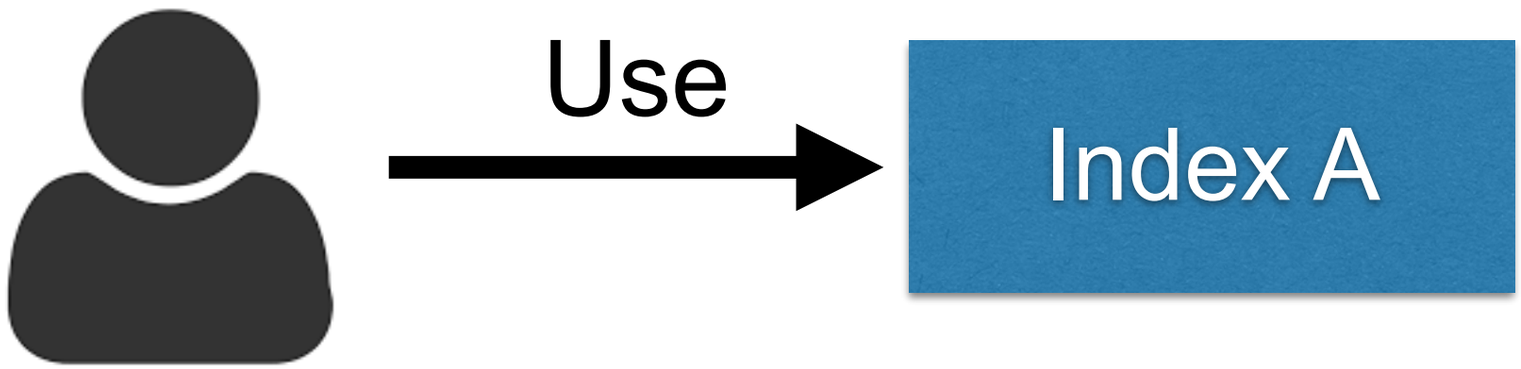
Simple generator used to push fake HTTP traffic logs into elasticsearch

npm install -g @elastic/makelogs

https://github.com/elastic/makelogs

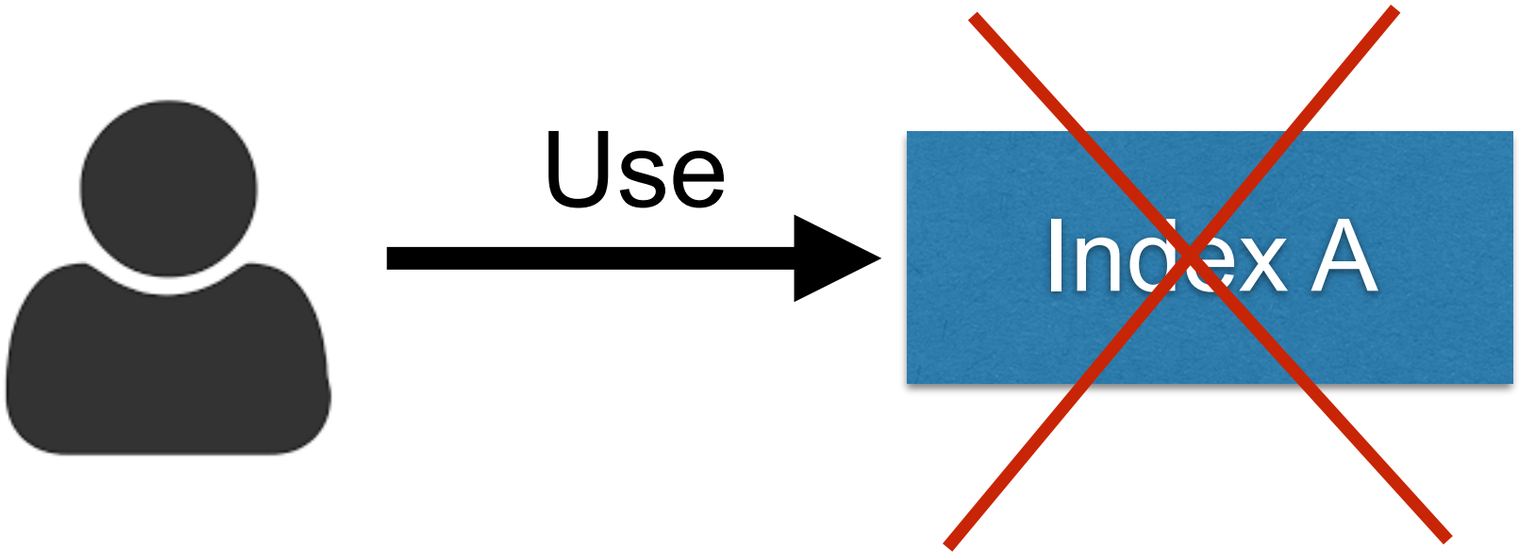
Alias Index

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

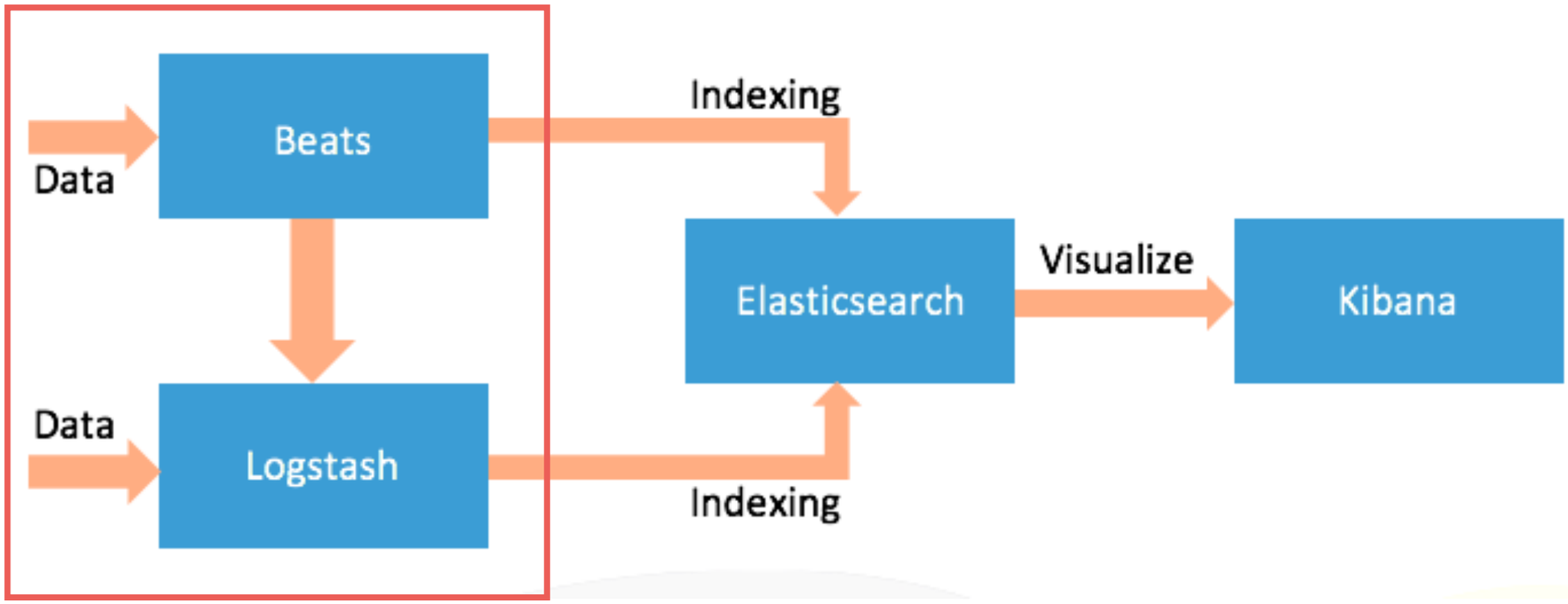
Using Alias Index

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

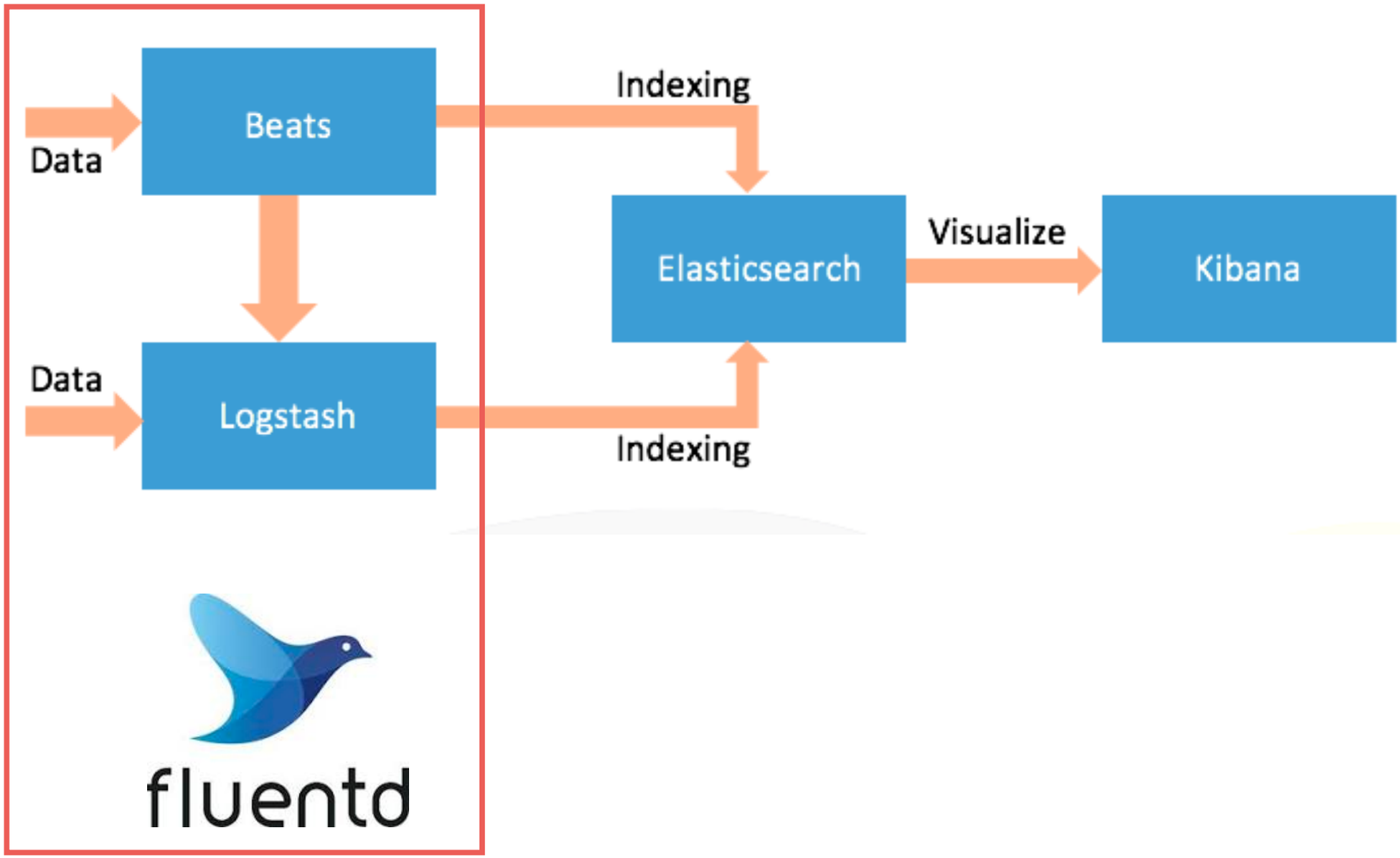
Using Alias Index

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

ELK stack



EFK stack

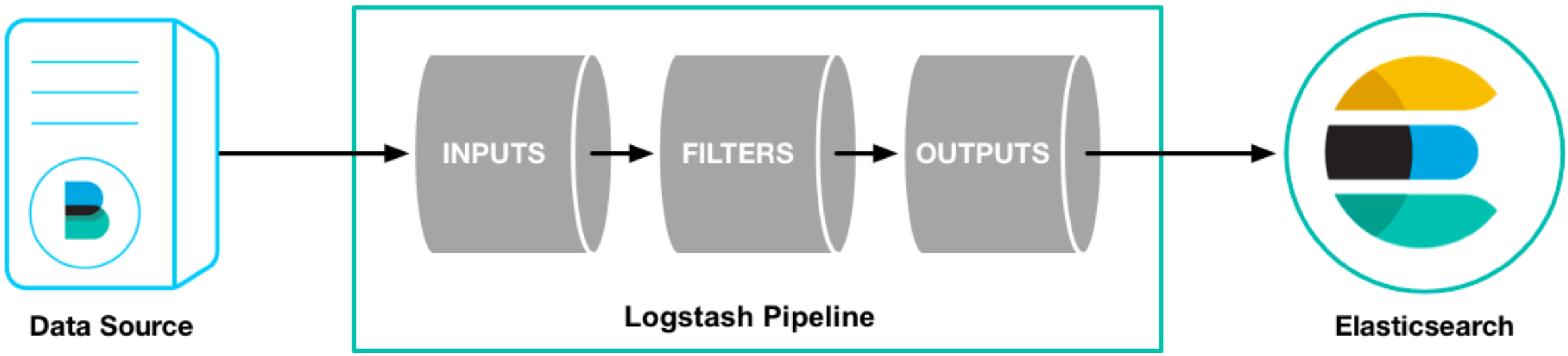


Working with Logstash

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

Standard Input

Standard Output

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

File system

Standard Output

Elasticsearch

workshop/logstash-beat-fluentd/demo.conf

try.conf

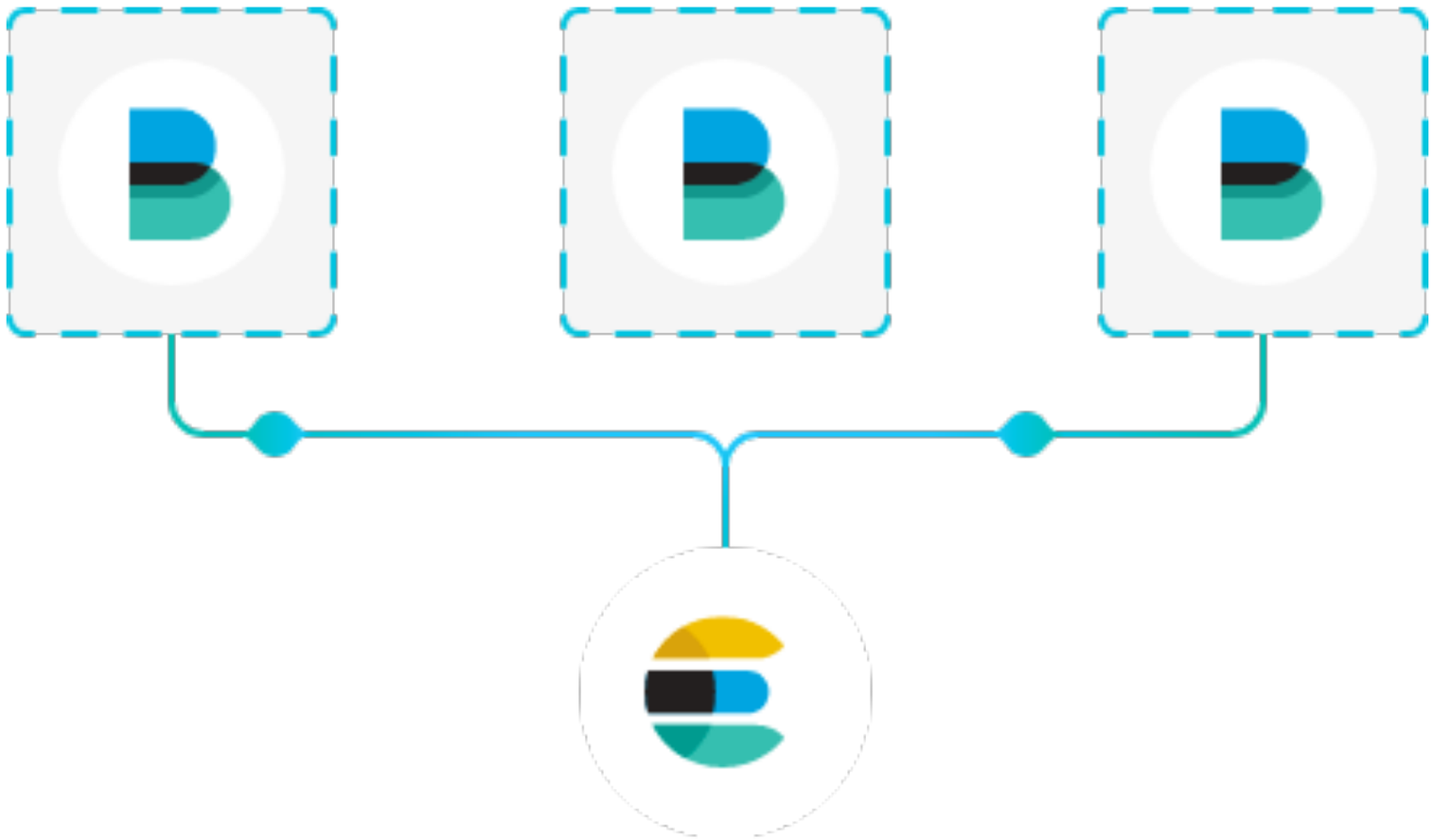
```
input {
    stdin{}
}

output {
    stdout {
        codec => rubydebug
    }

    elasticsearch {
    }
}
```

Design your input first !!

Use beats is better

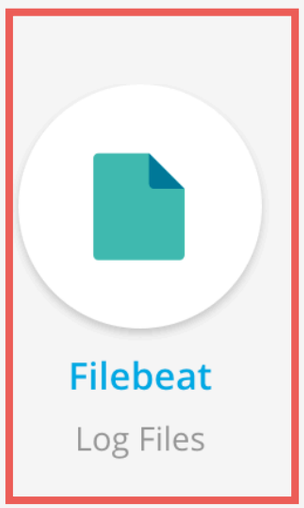
https://www.elastic.co/products/beats

Beat



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

Beat



Filebeat
Log Files

Metricbeat
Metrics

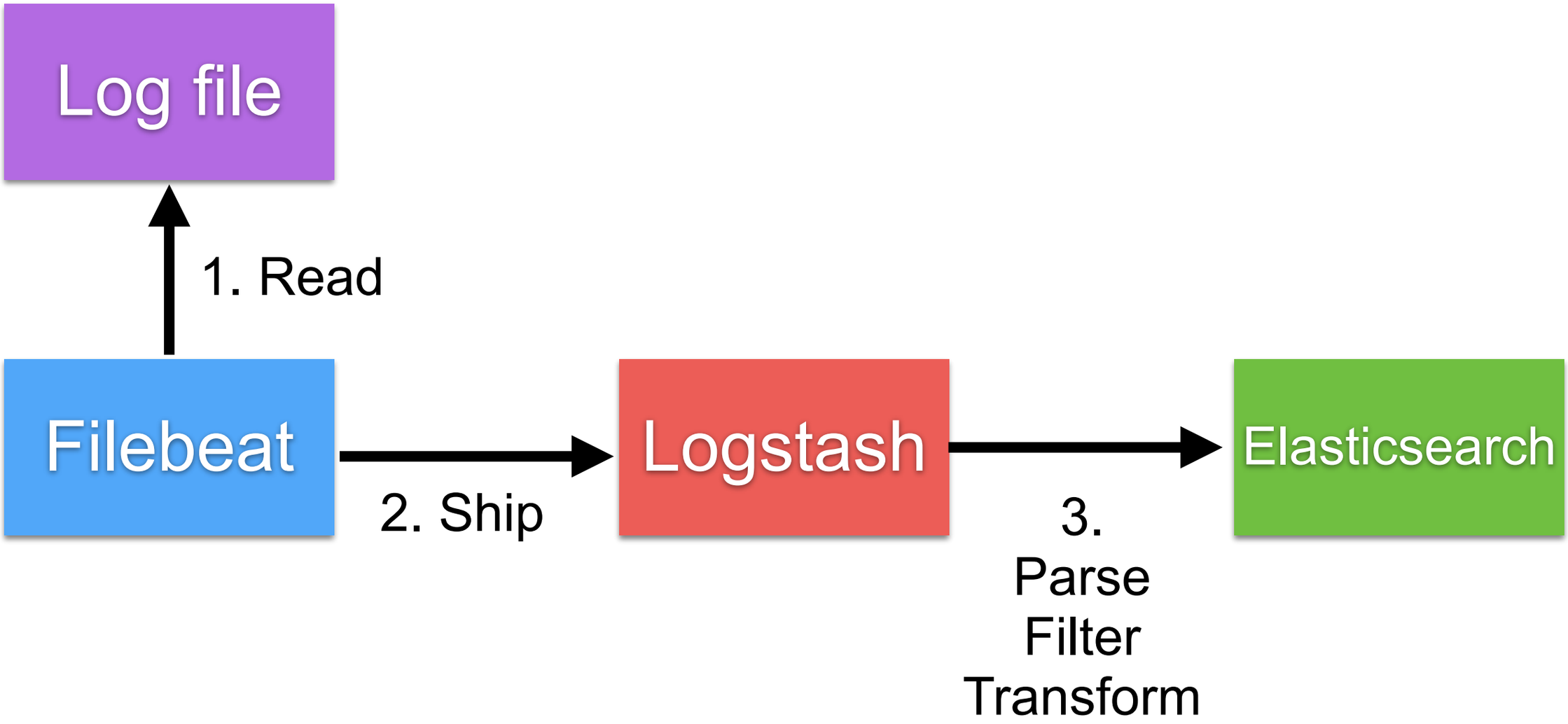
Packetbeat
Network Data

Winlogbeat
Windows Event Logs

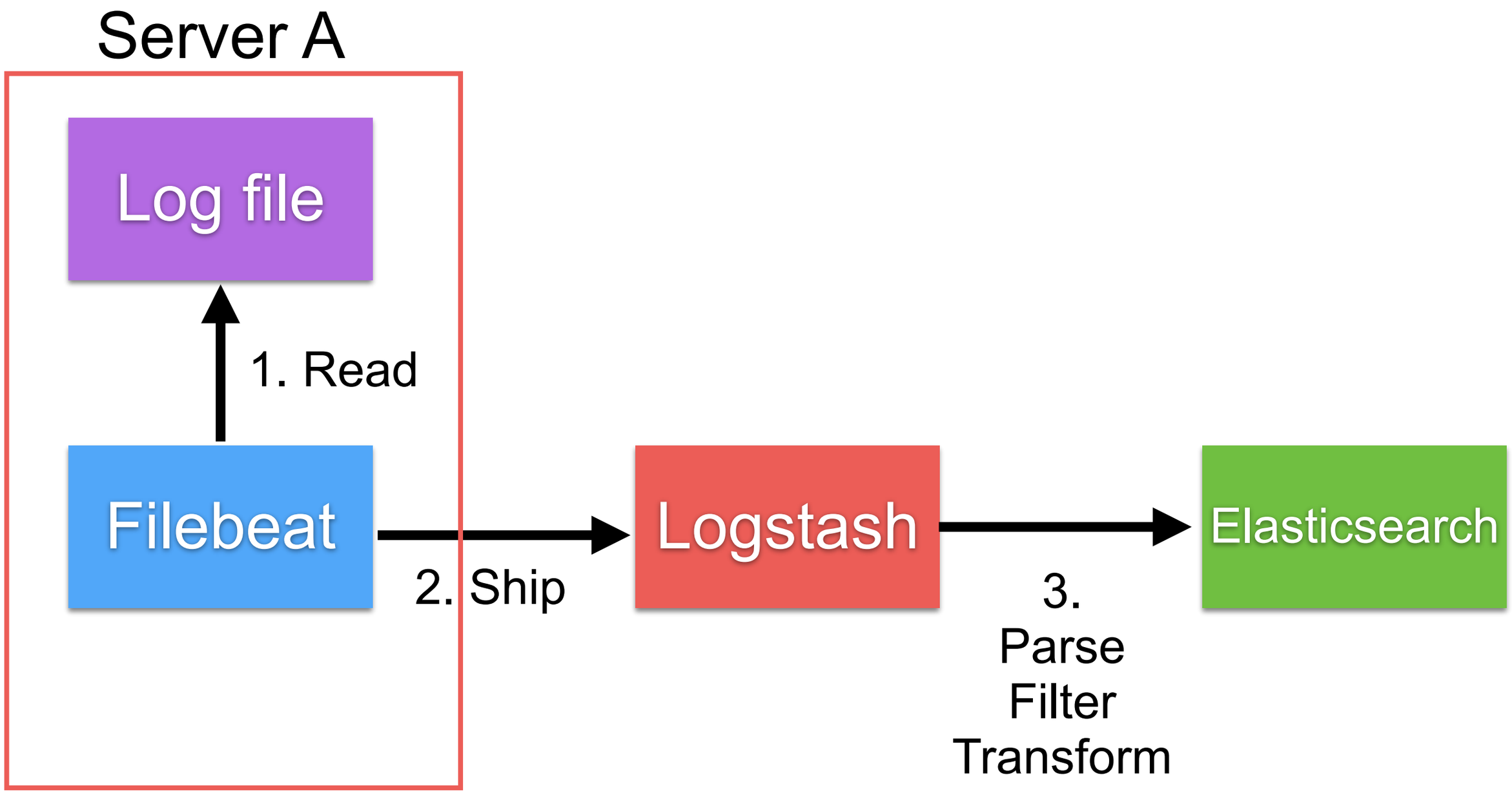
Auditbeat
Audit Data

Heartbeat
Uptime Monitoring

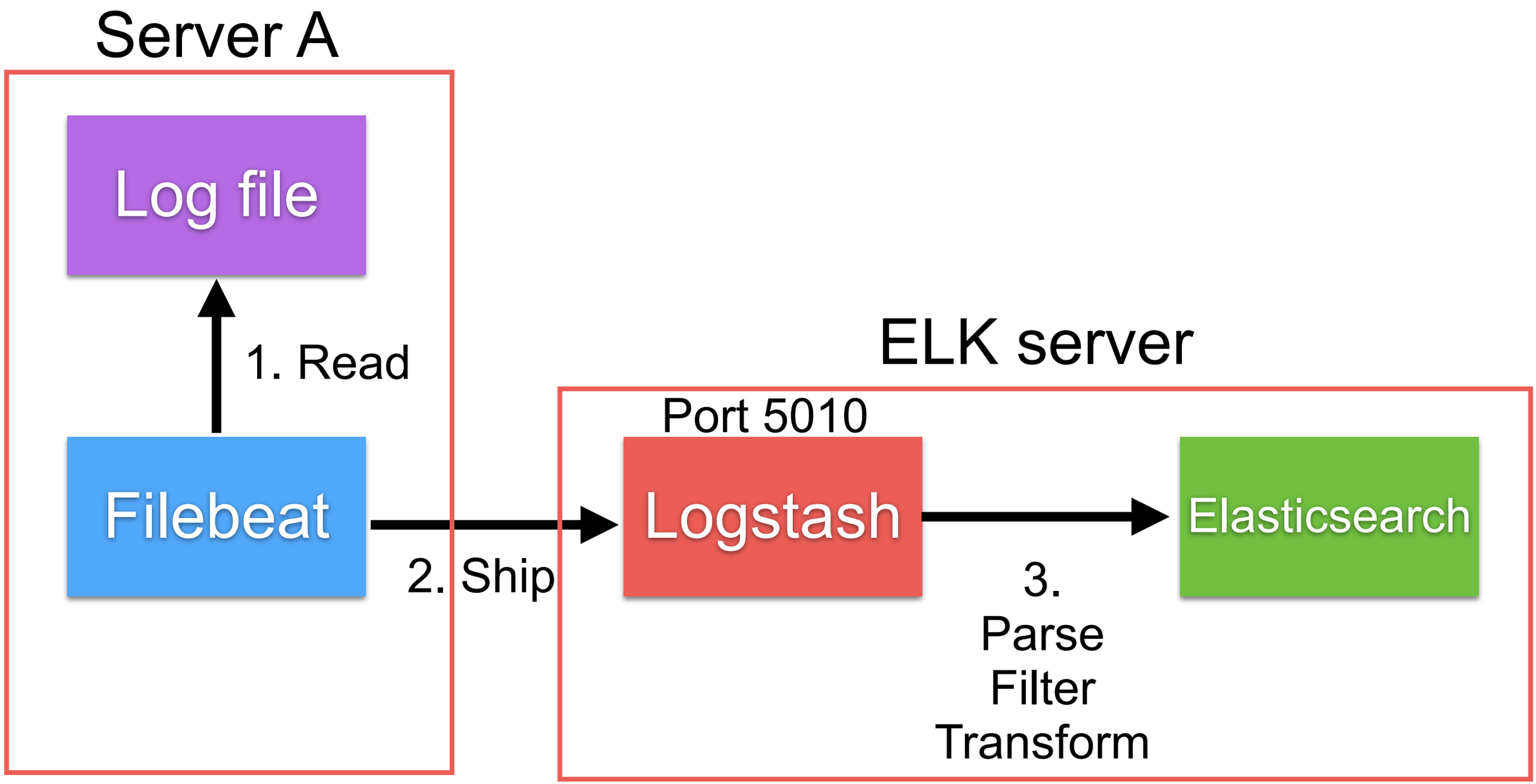
Example of filebeat



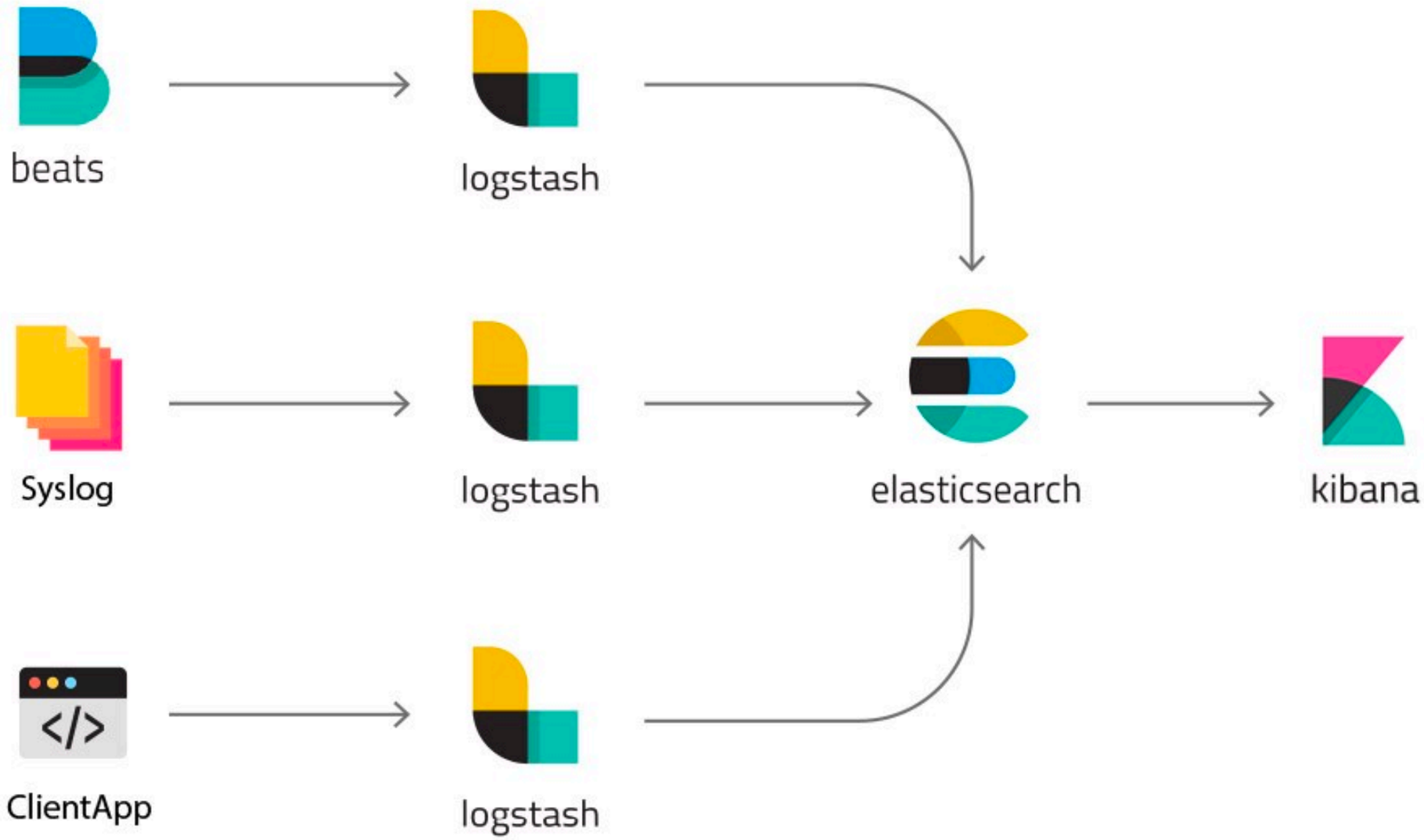
Example of filebeat



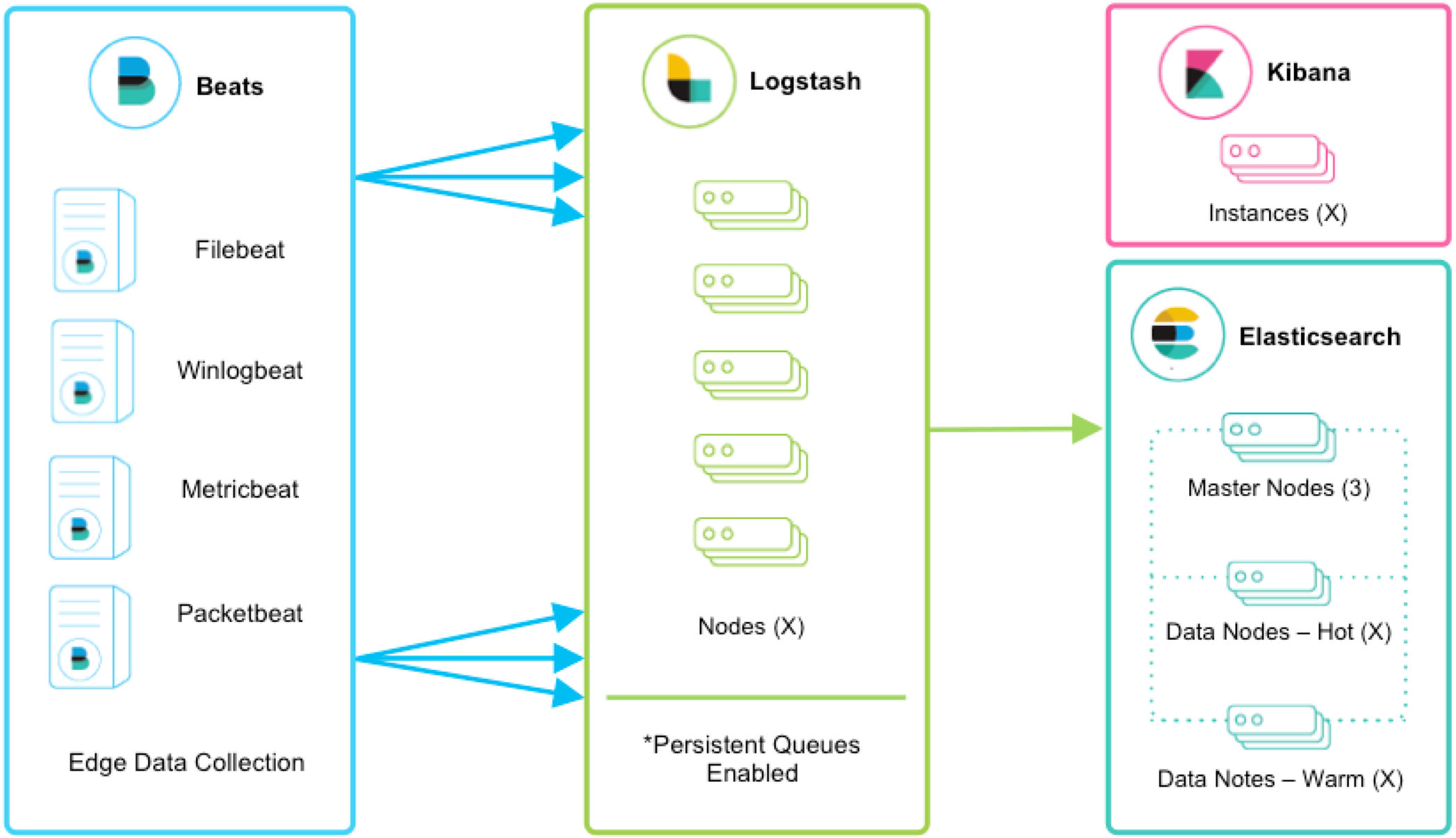
Example of filebeat



Beat and Logstash

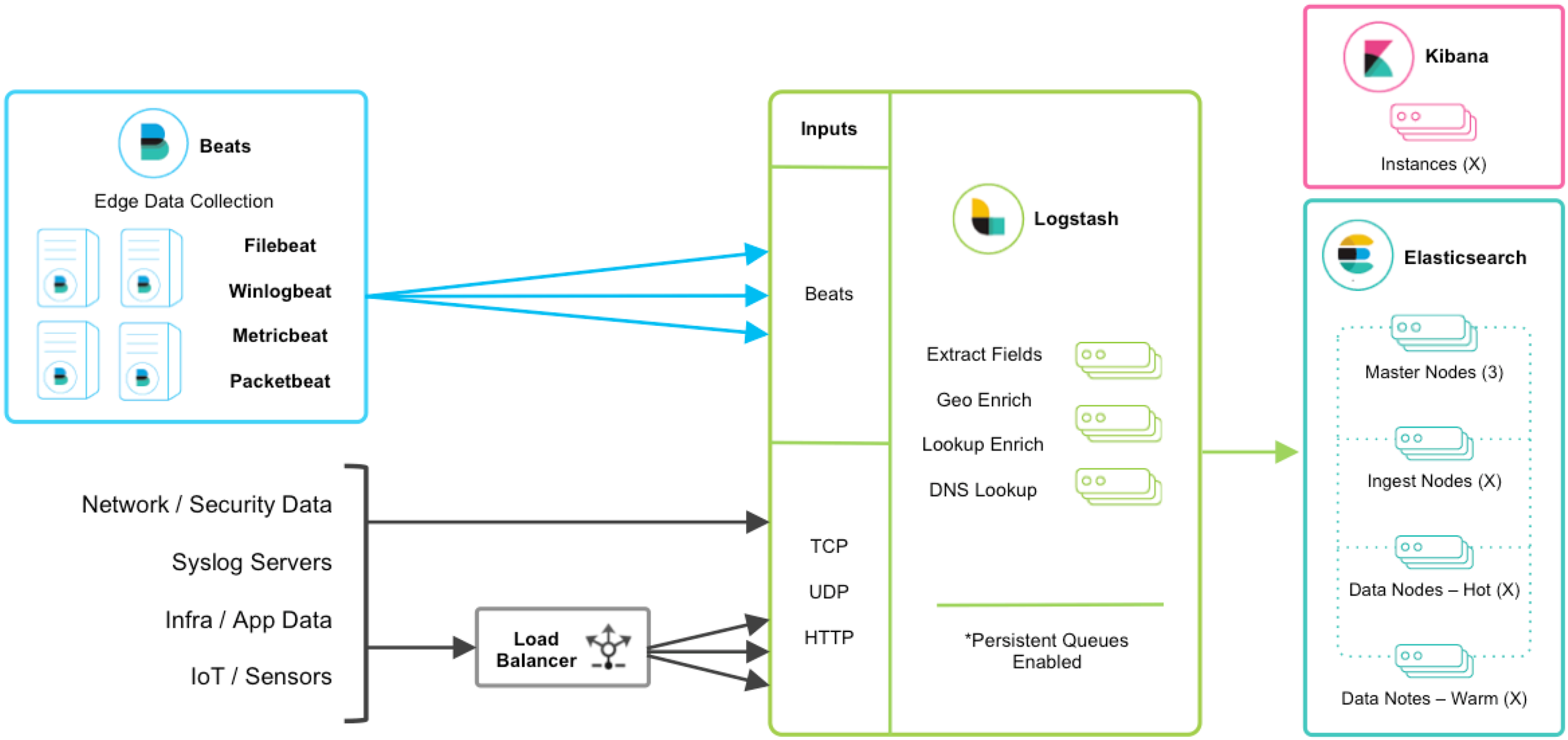


Scaling



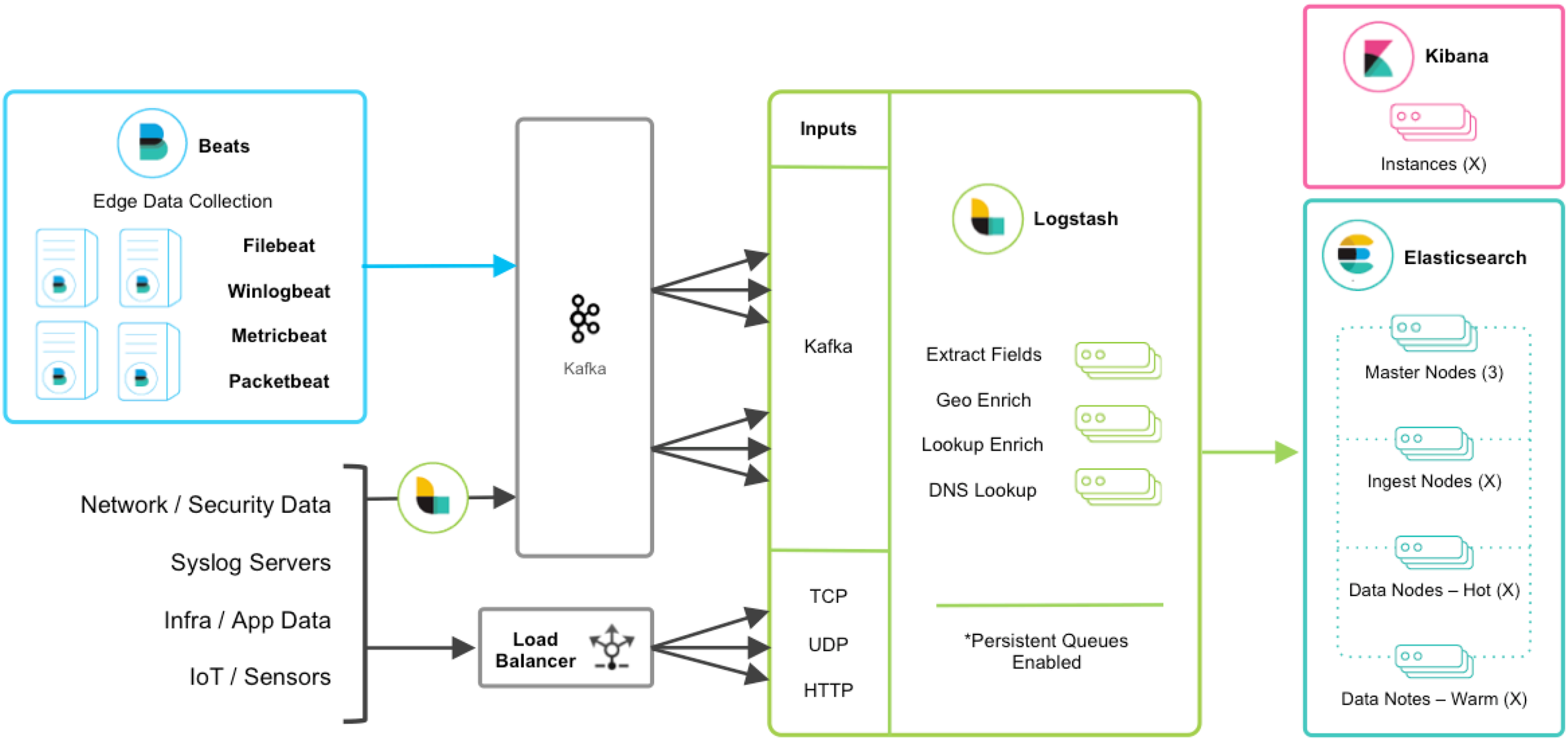
https://www.elastic.co/guide/en/logstash/current/deploying-and-scaling.html

More data sources



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

Use messaging Queue

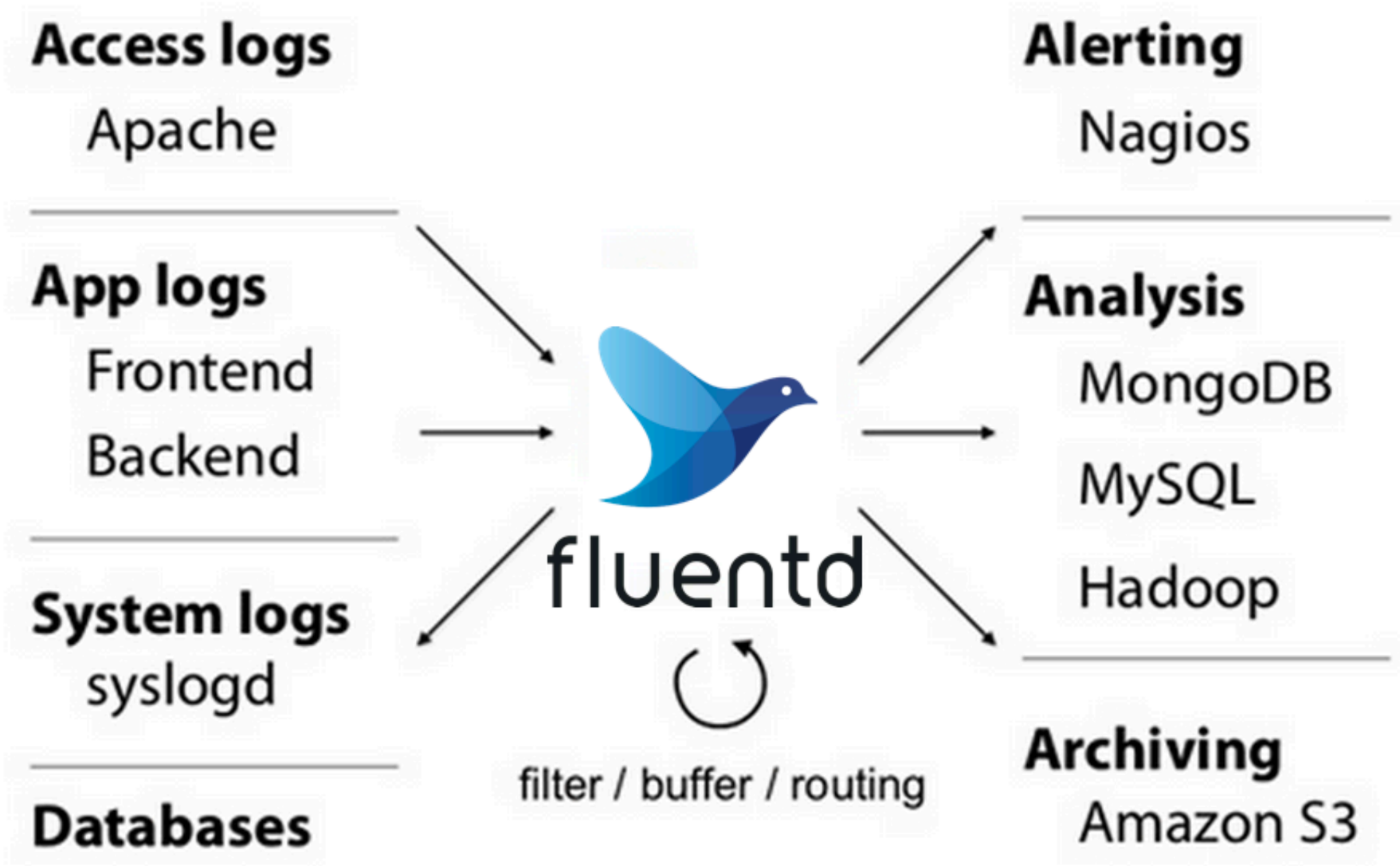


https://www.elastic.co/blog/logstash-persistent-queue

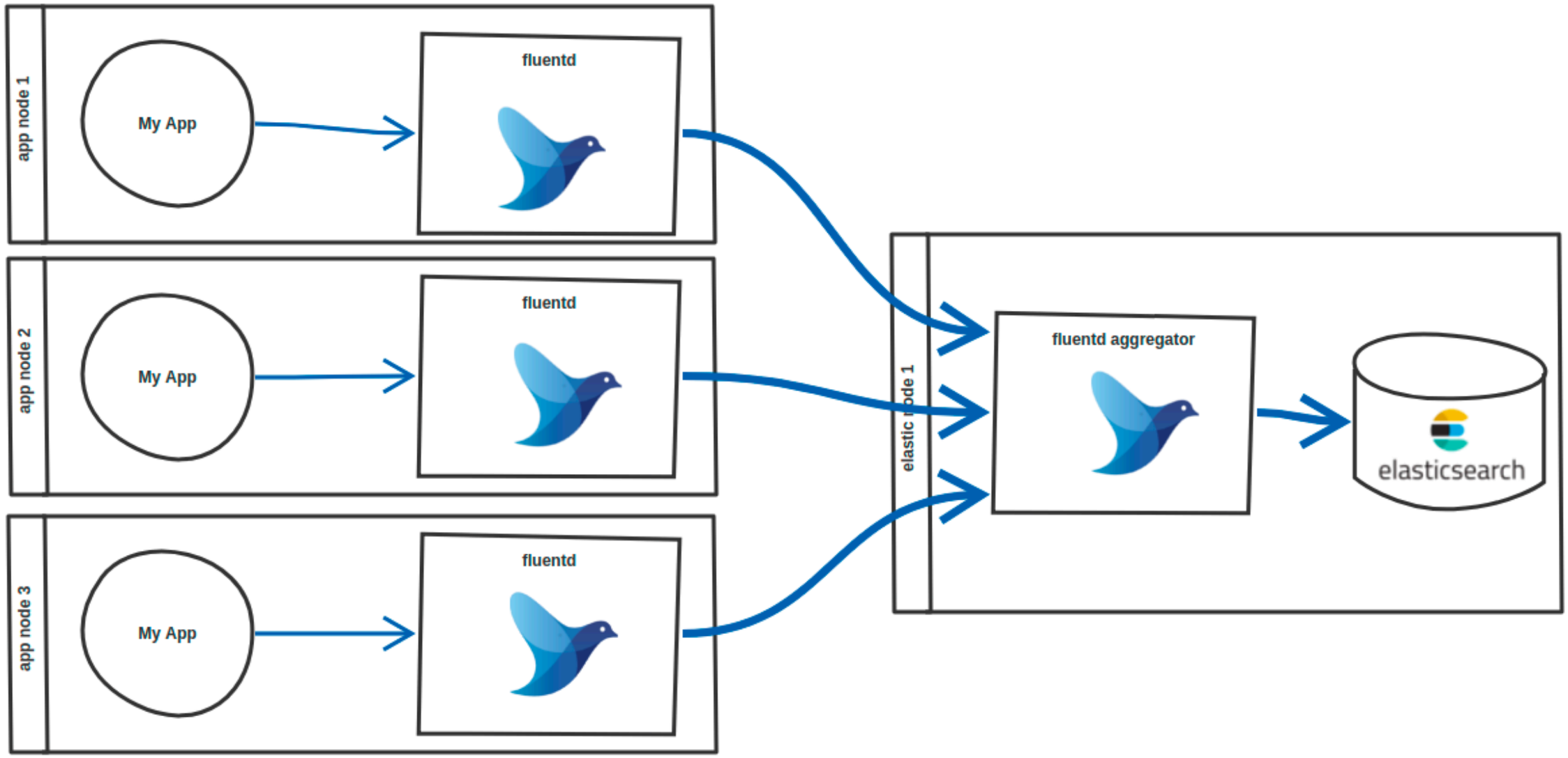
Working with Fluentd

https://www.fluentd.org/

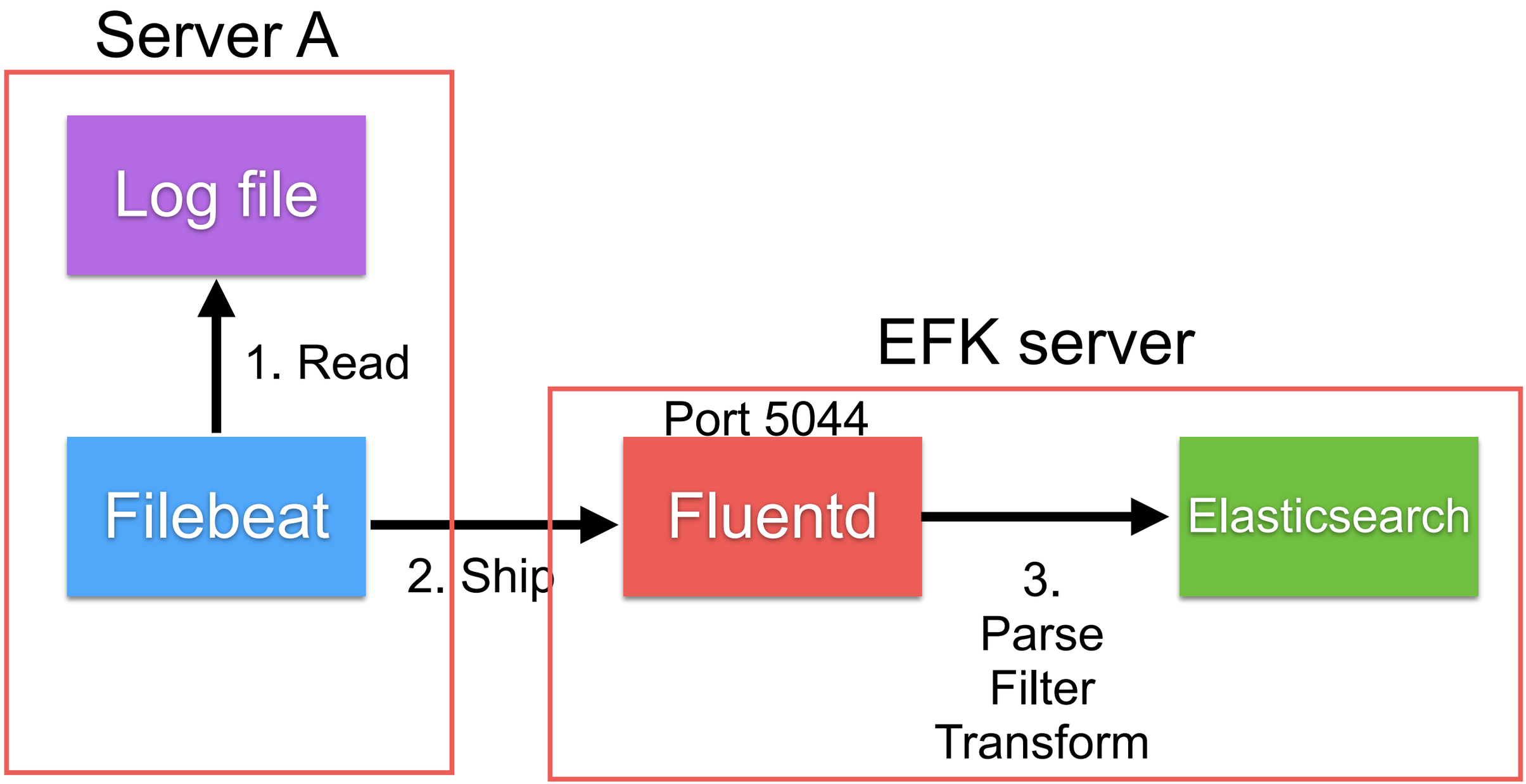
Fluentd



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>[^\"]*)")?$
```

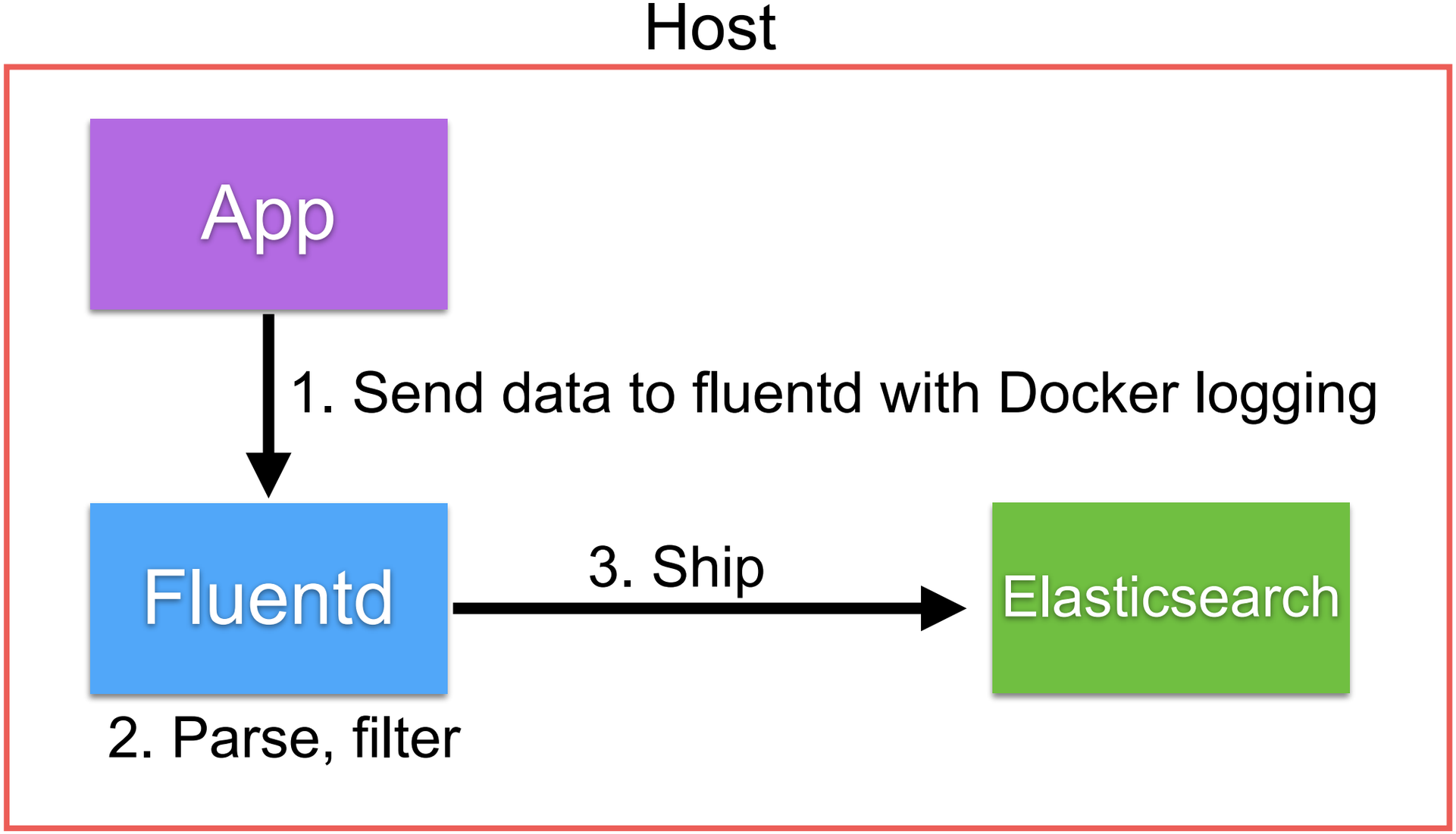
Time Format:

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

http://fluentular.herokuapp.com/

Fluentd with Docker

Using docker-compose



Design for Failure

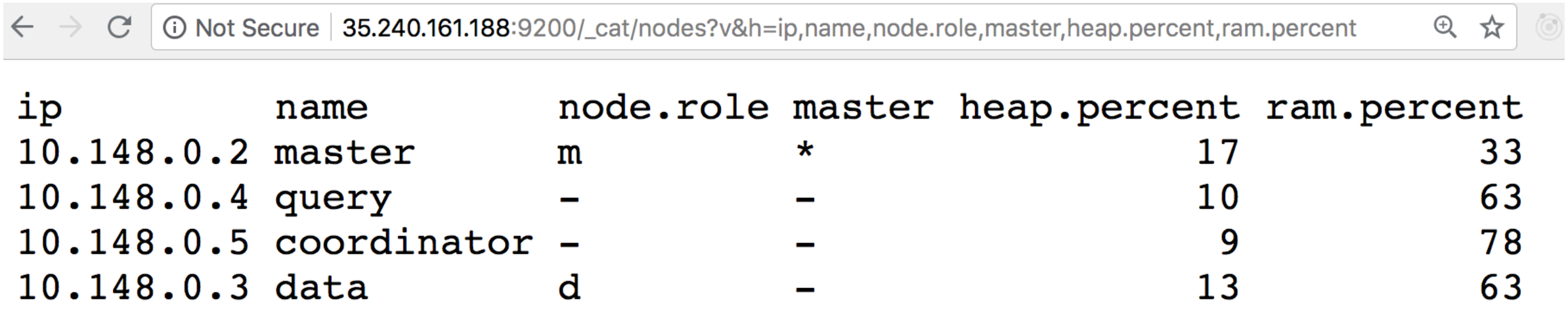
09-cluster

Elasticsearch Nodes

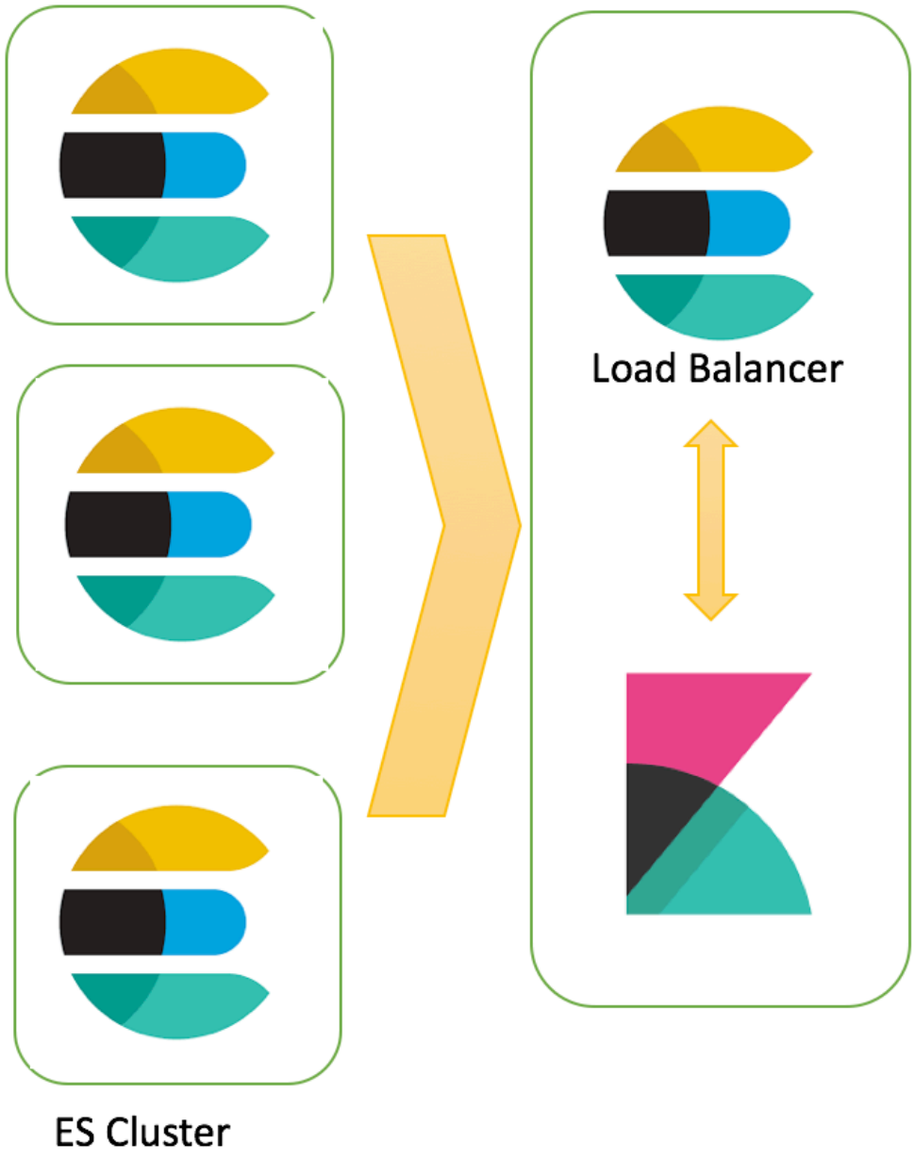
Node Type	Description
Master	Control the cluster
Data	Keep/store data
HTTP/Query	Run your query
Coordinating	Smart Load Balancer
Ingest	Pre-processing documents before indexing
Machine Learning	Required subscription !!

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

Elasticsearch Nodes



Elasticsearch Nodes



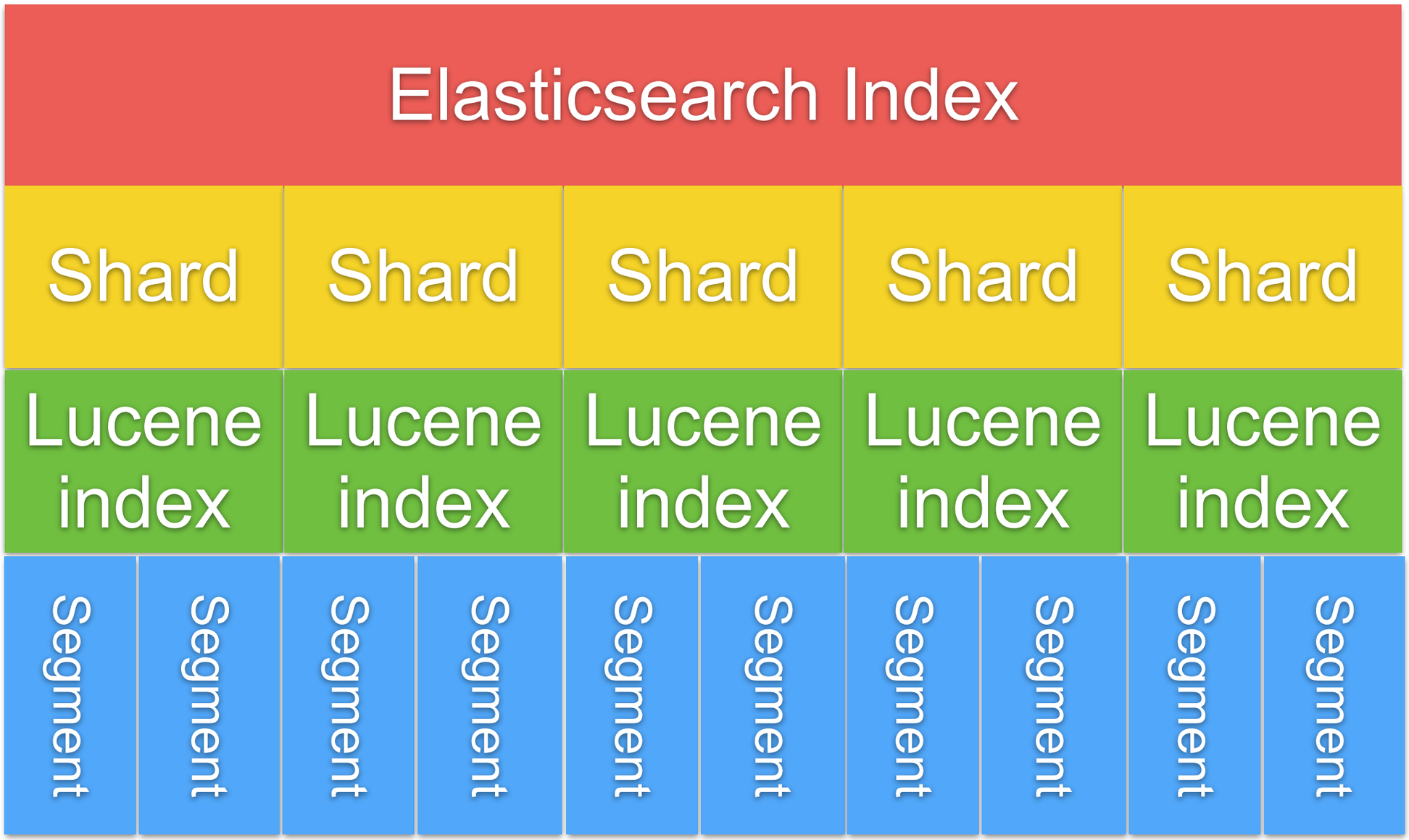
https://www.elastic.co/guide/en/kibana/current/production.html#load-balancing

Elasticsearch Nodes

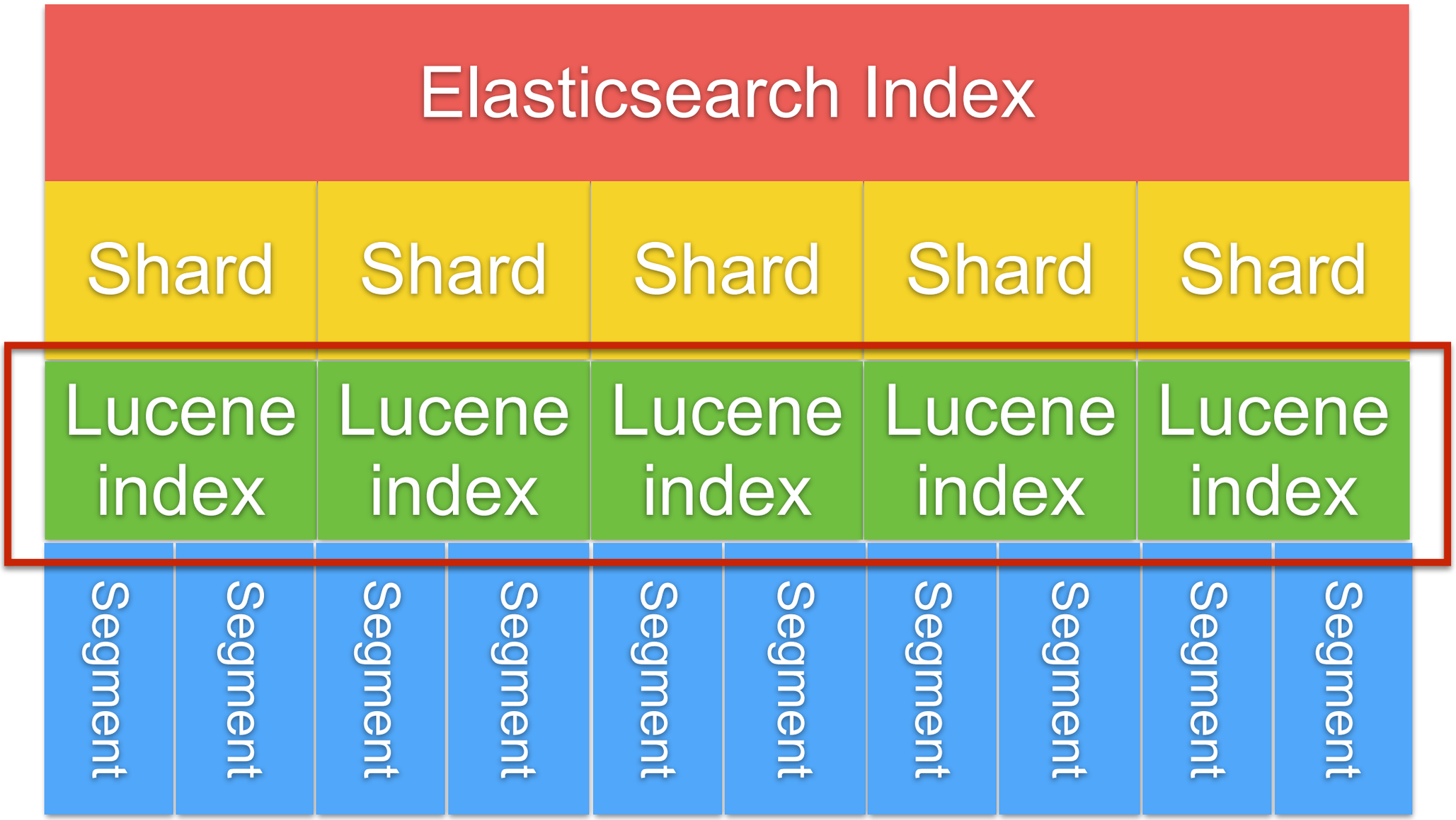
Apache Lucene

http://lucene.apache.org/

Apache Lucene

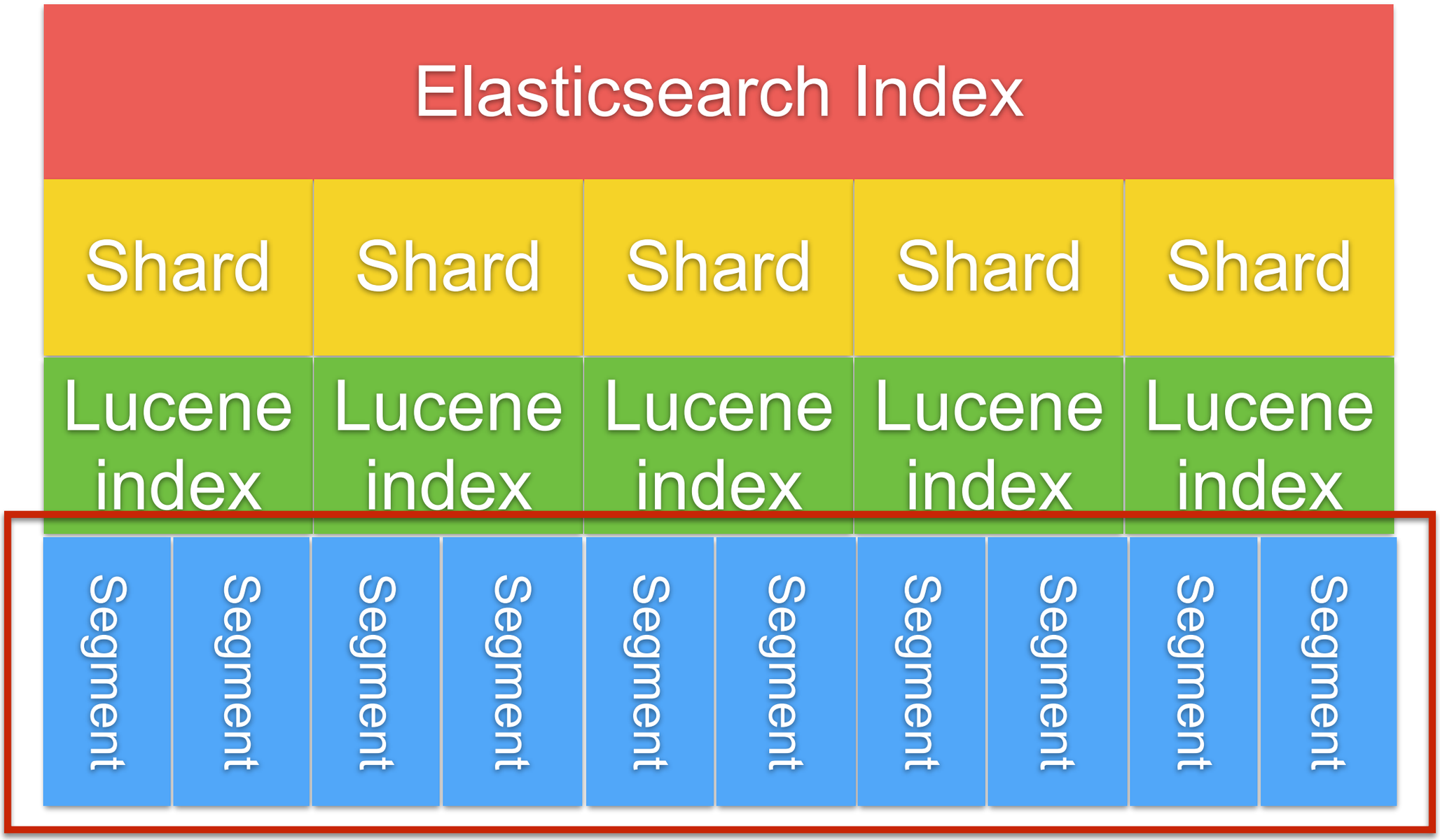


Apache Lucene



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

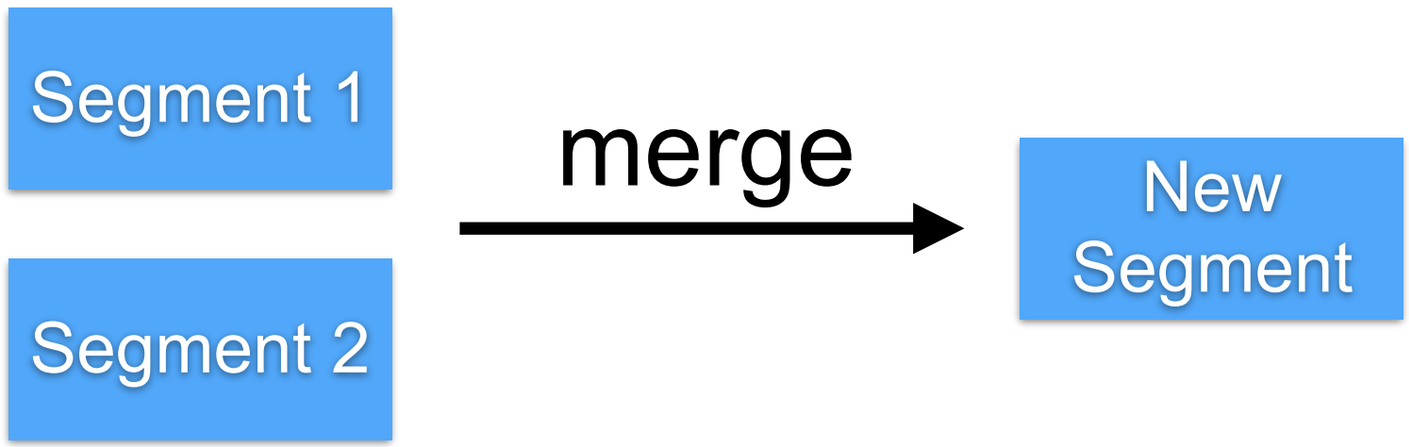
Apache Lucene



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

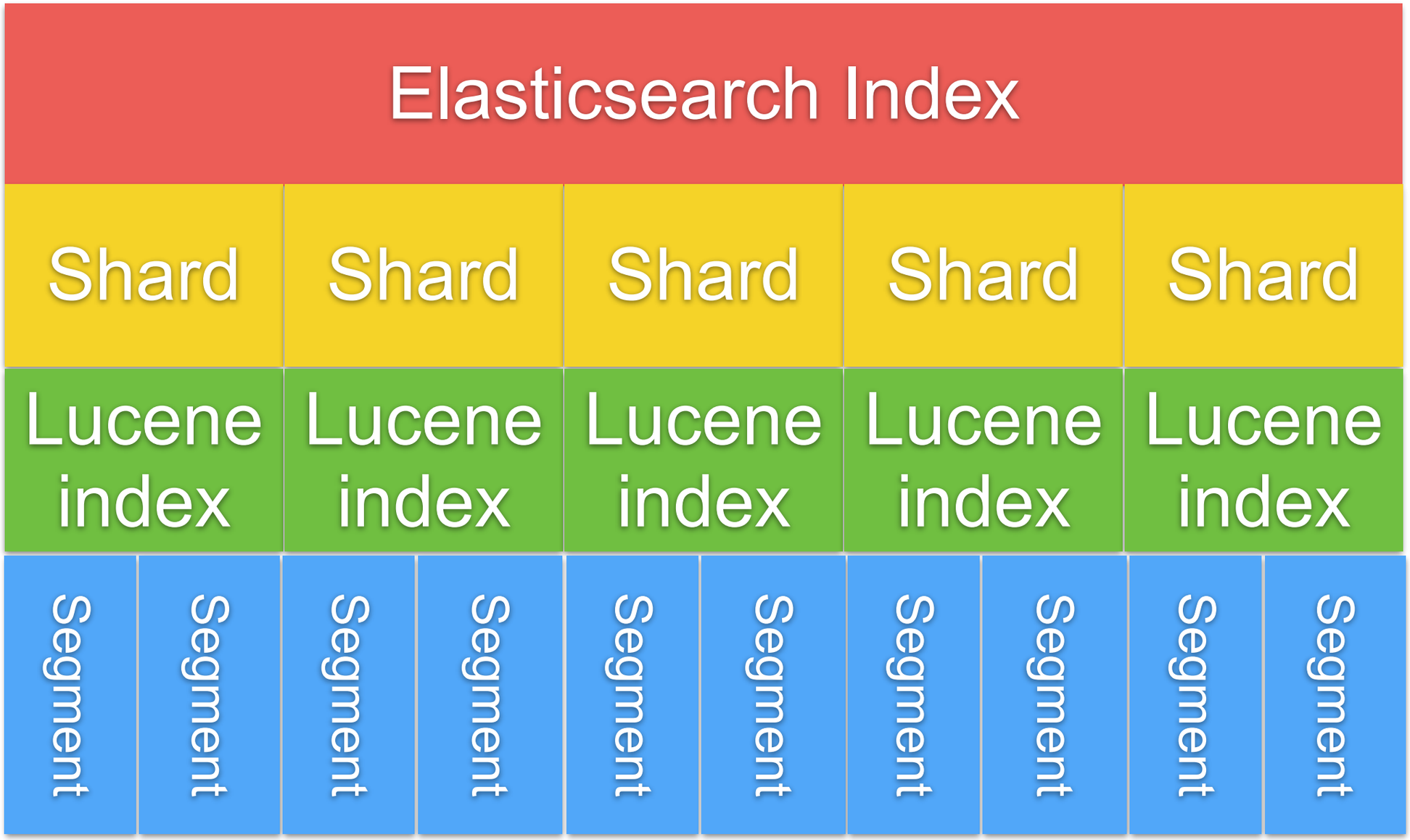
Data Size	# of shard
< 3M	1
>3M <5M	2
>5M	(# of document / 5M) + 1

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

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

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

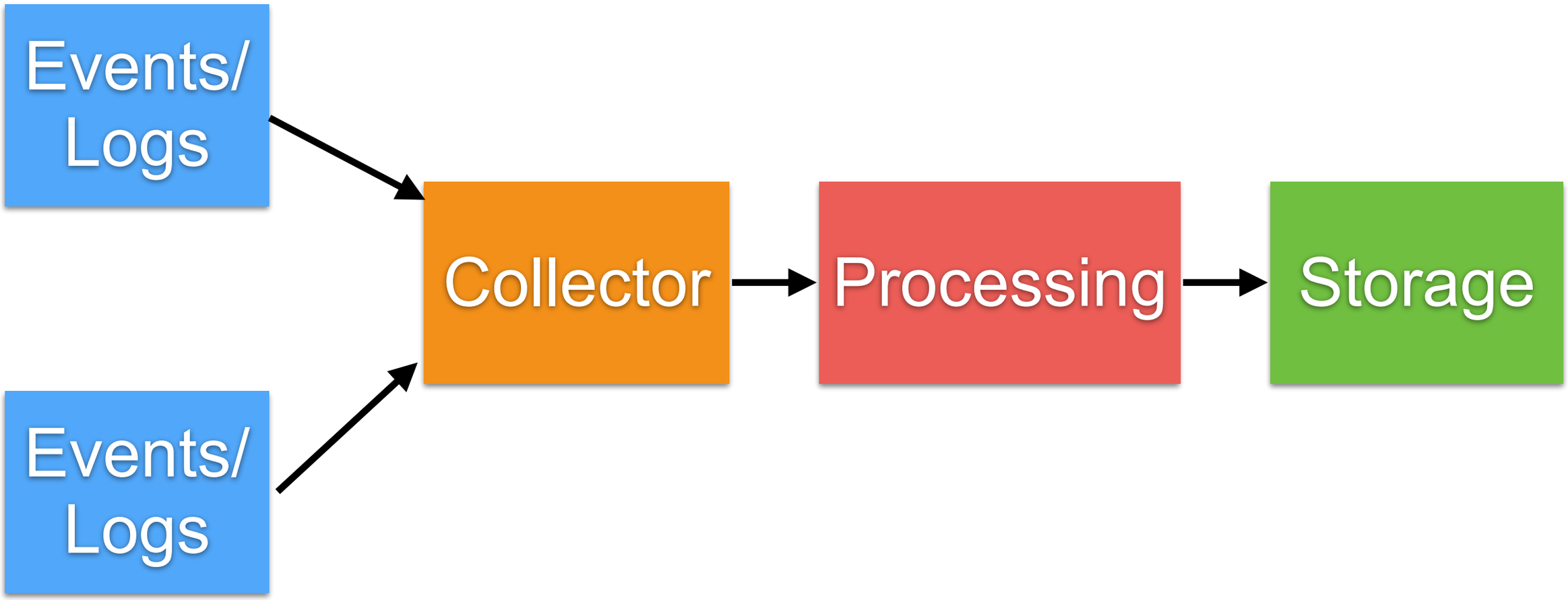
Query data

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

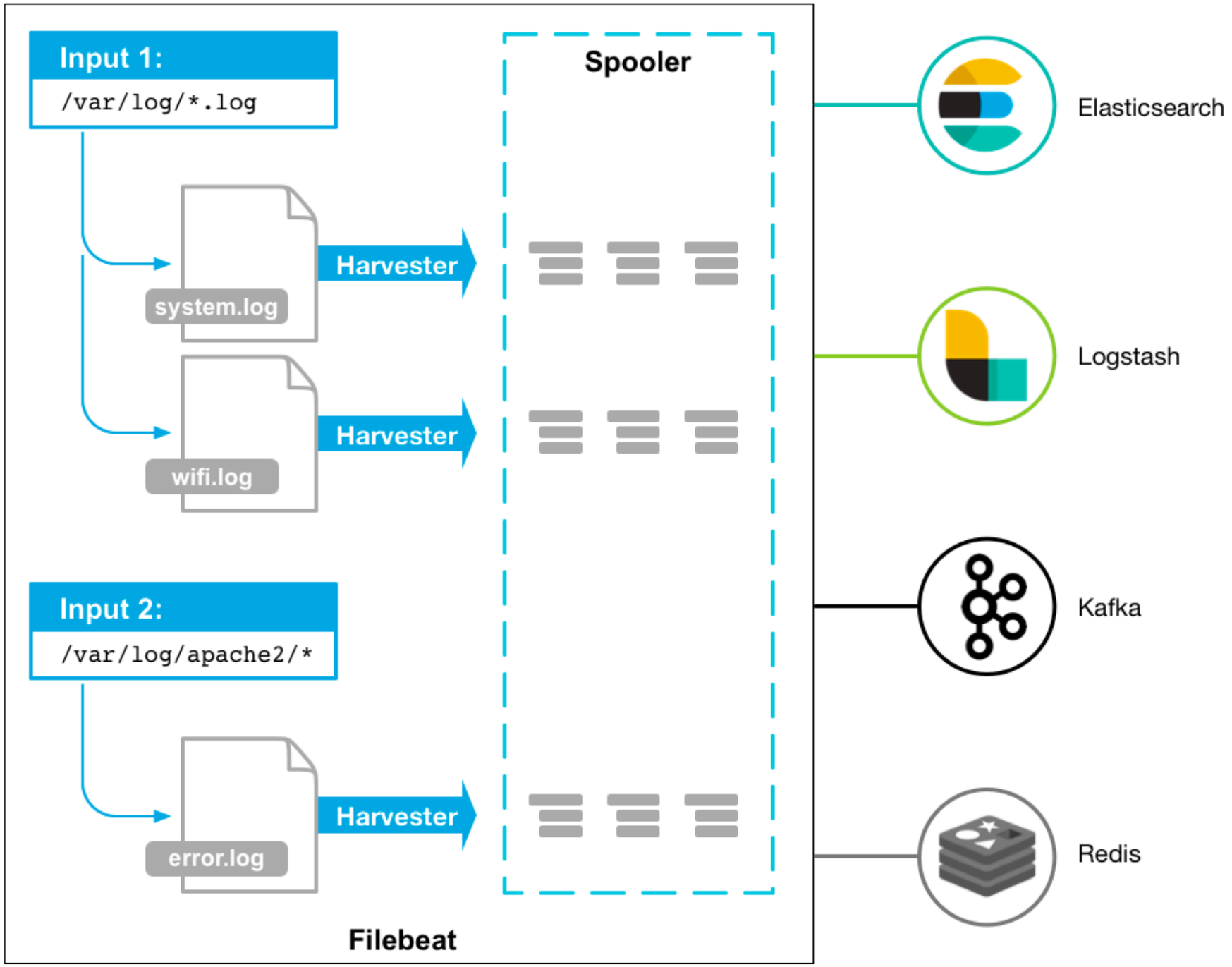
Event or Logging from Servers



Event or Logging from Servers

Data Collection	Data Aggregation & Processing	Indexing & storage	Analysis & visualization

Working with Filebeat



Event or Logging from Servers

RabbitMQ

NGINX

Data Collection	Buffering	Data Aggregation & Processing	Indexing & storage	Analysis & visualization

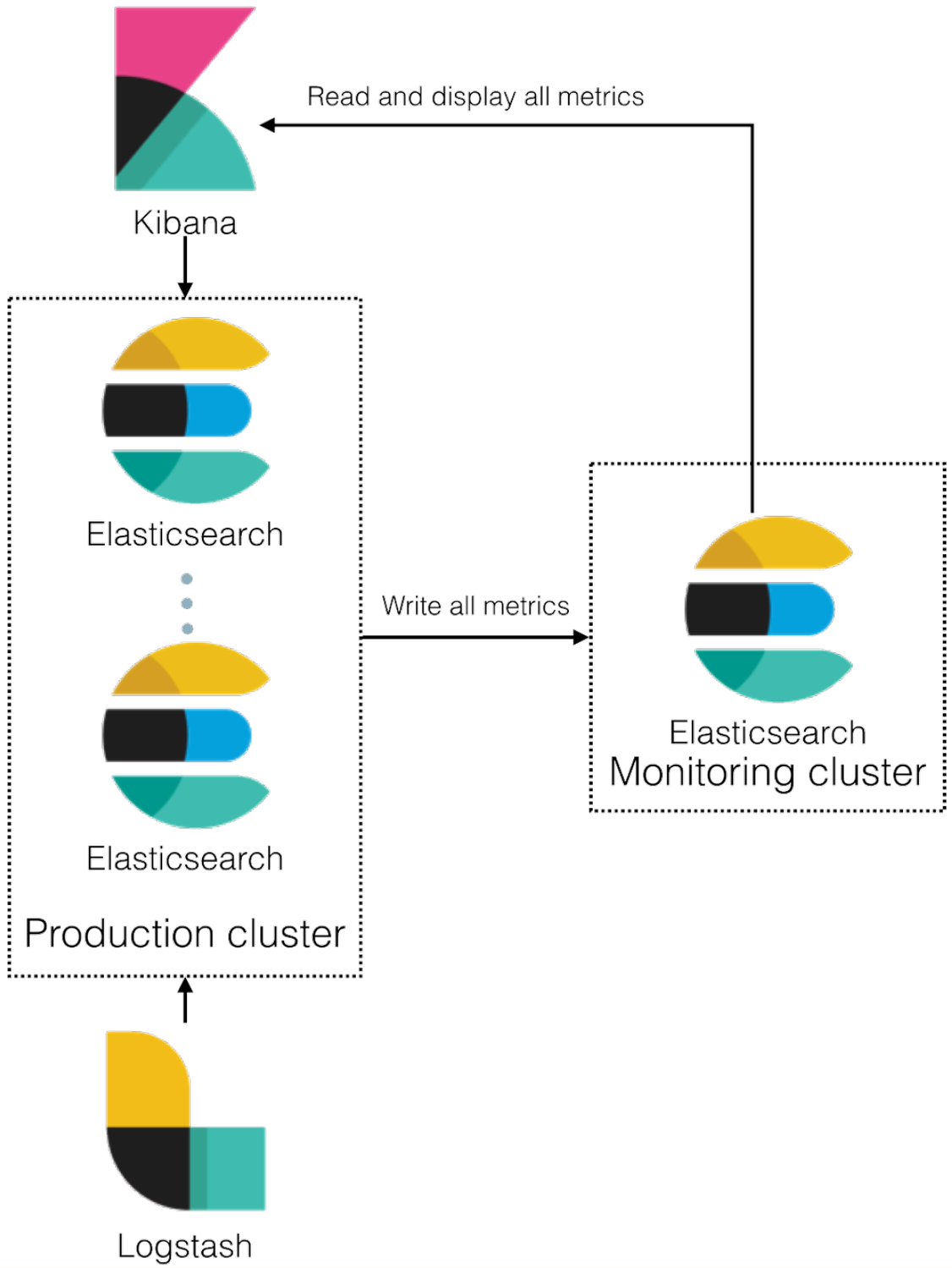
Monitoring

Collect data from ?

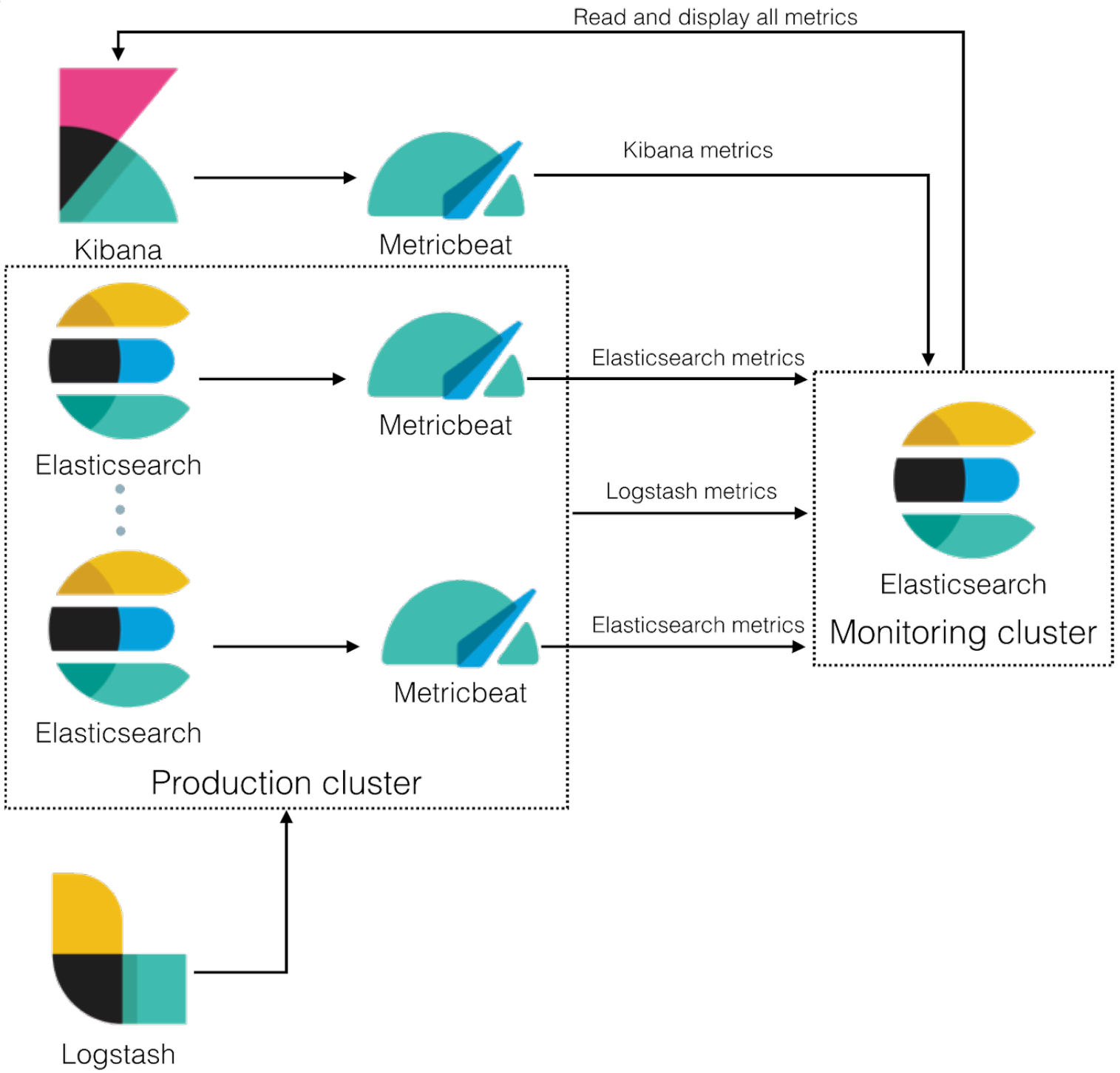
Elasticsearch nodes
Logstash nodes
Kibana instances

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

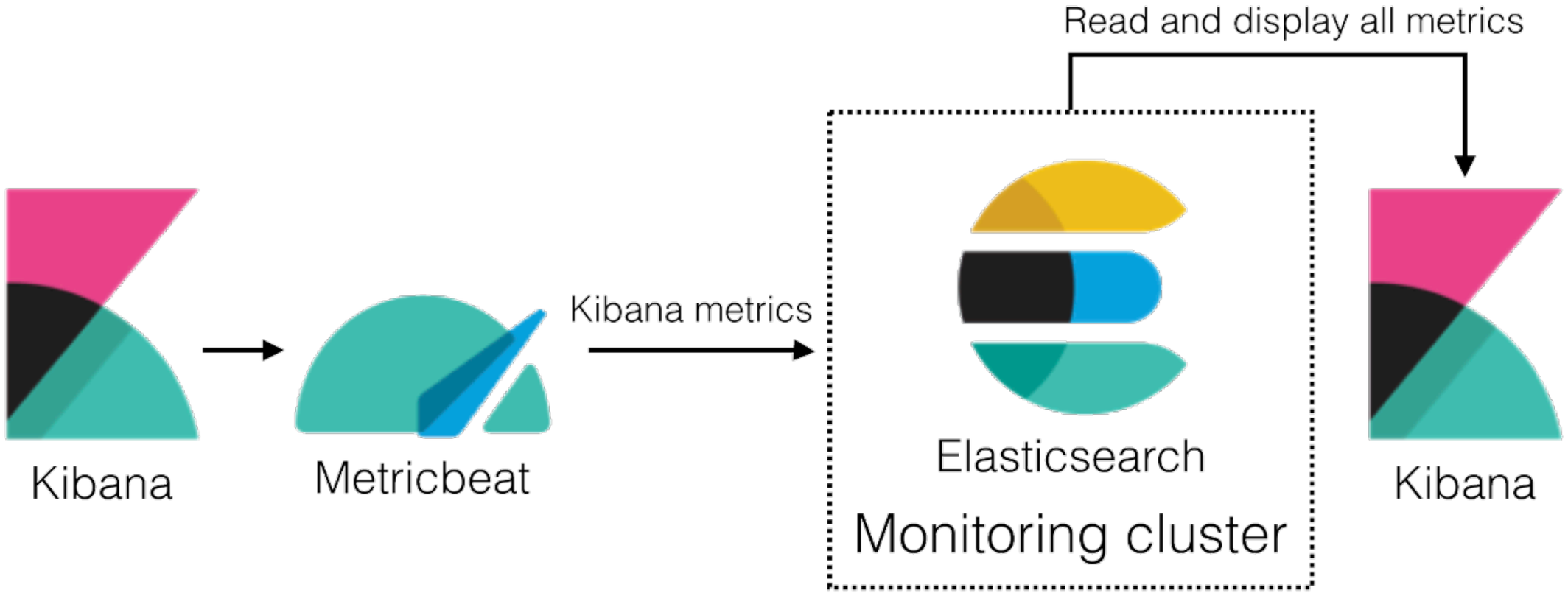
Recommend architecture



Elasticsearch 6.4 + (beta)



Try to separate kibana

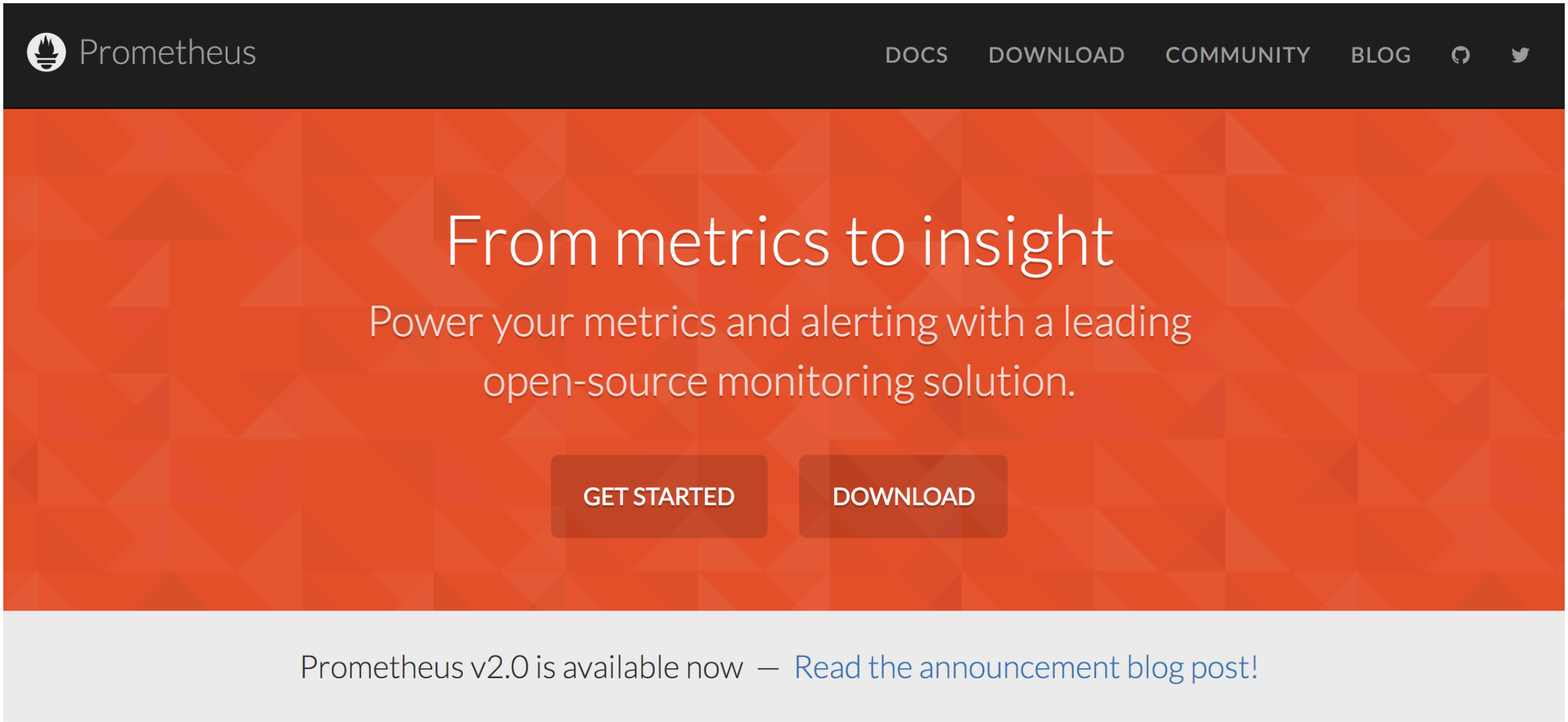


Agenda

- Introduction to Prometheus
- Introduction to Grafana
- Install and configuration
- Working with Prometheus plugins
- Working with Grafana
  - Authentication
  - Datasource
  - Visualization and dashboard
  - Alert system

Prometheus

Prometheus



https://prometheus.io/

Prometheus

Metrics monitoring system

Gather **fine grained data** at frequent interval

Make them useful by **label**

Analyse then to understand what is going on

Prometheus

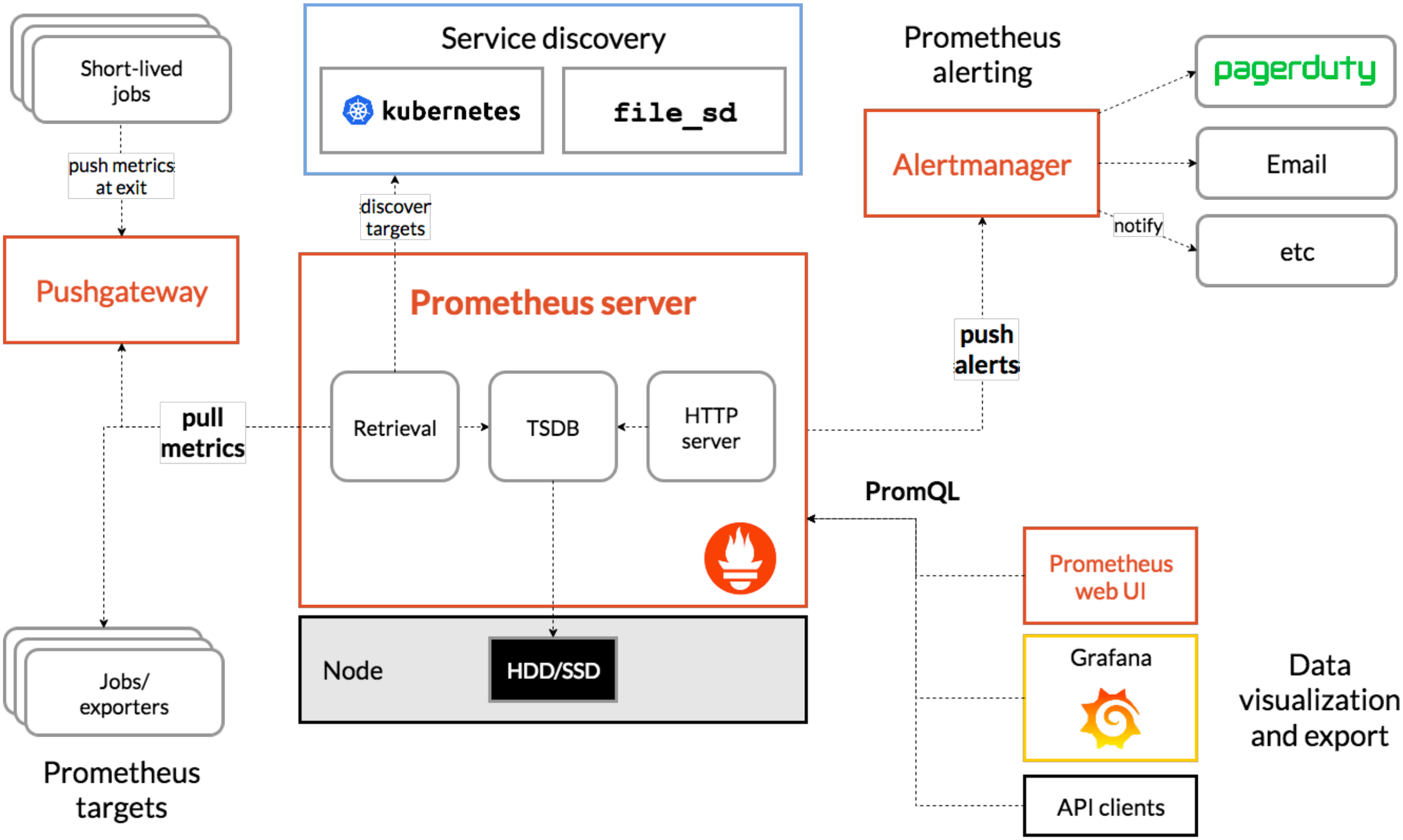
Easy to config, deploy and maintain
Designed in multiple services
Container ready
Dynamic configuration

Prometheus

Open source
Develop with Go language
Support for multiple OS
Time series database
Many **exporters**
Alerting

https://prometheus.io/docs/instrumenting/exporters/

What Prometheus do ?



Prometheus exporters

Expose metrics with and HTTP API
Binding available for many language
Exporters o not save data

https://prometheus.io/docs/instrumenting/exporters/

Prometheus exporters

Node exporter
Blackbox exporter
Mysql exporter

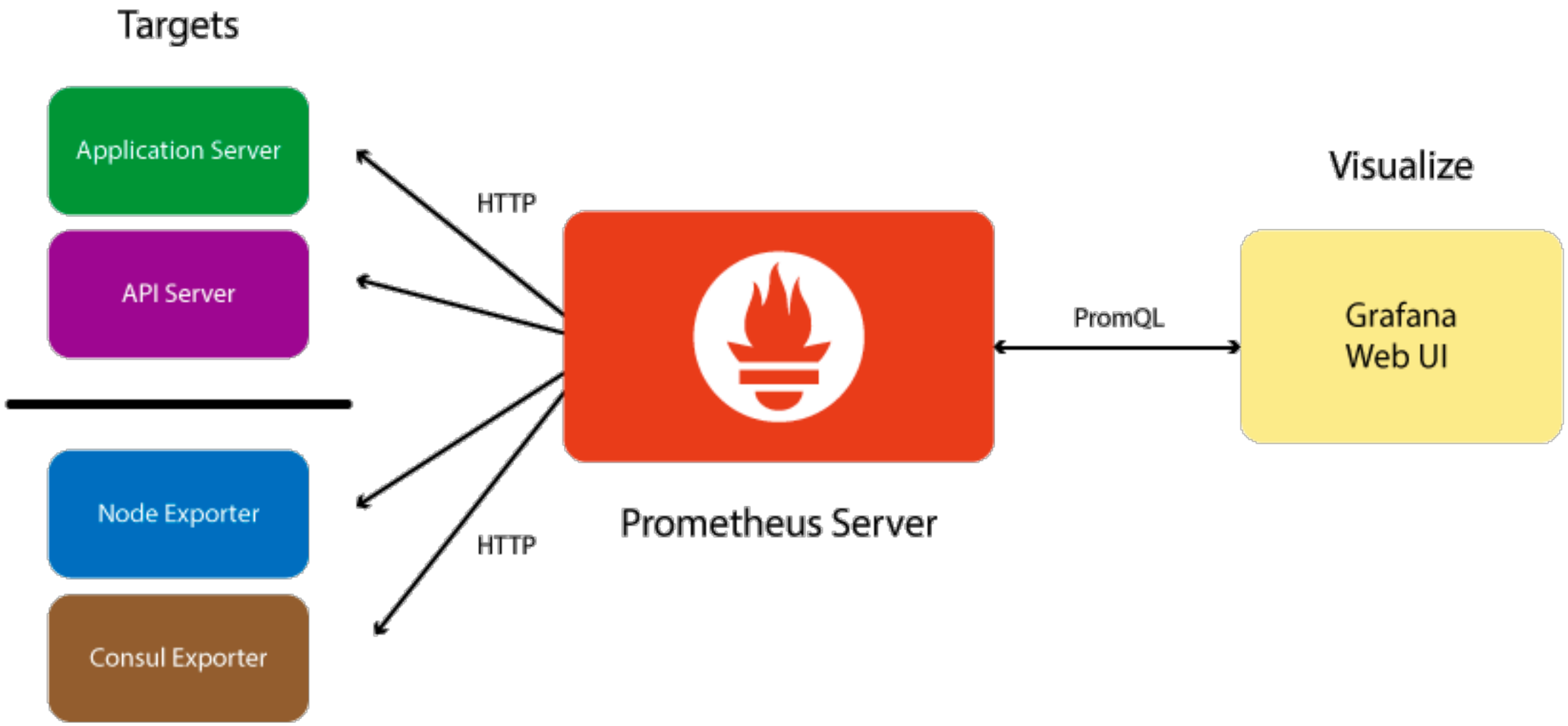
https://prometheus.io/docs/instrumenting/exporters/

Alerting

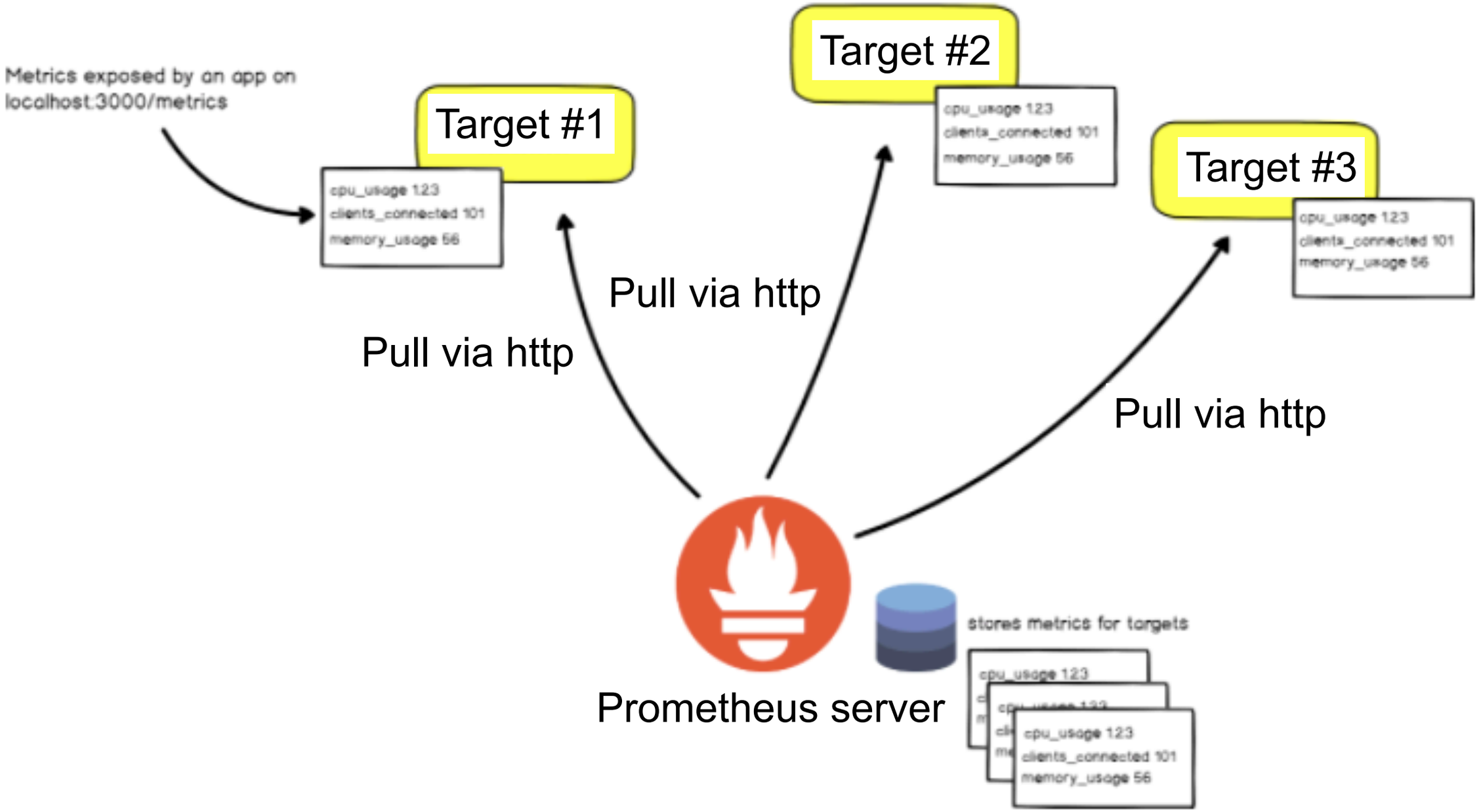
Prometheus has recording rules
Record frequent queries or alerting
Alert base on any **PromQL** queries

https://prometheus.io/docs/alerting/latest/overview/

What Prometheus do ?



What Prometheus do ?



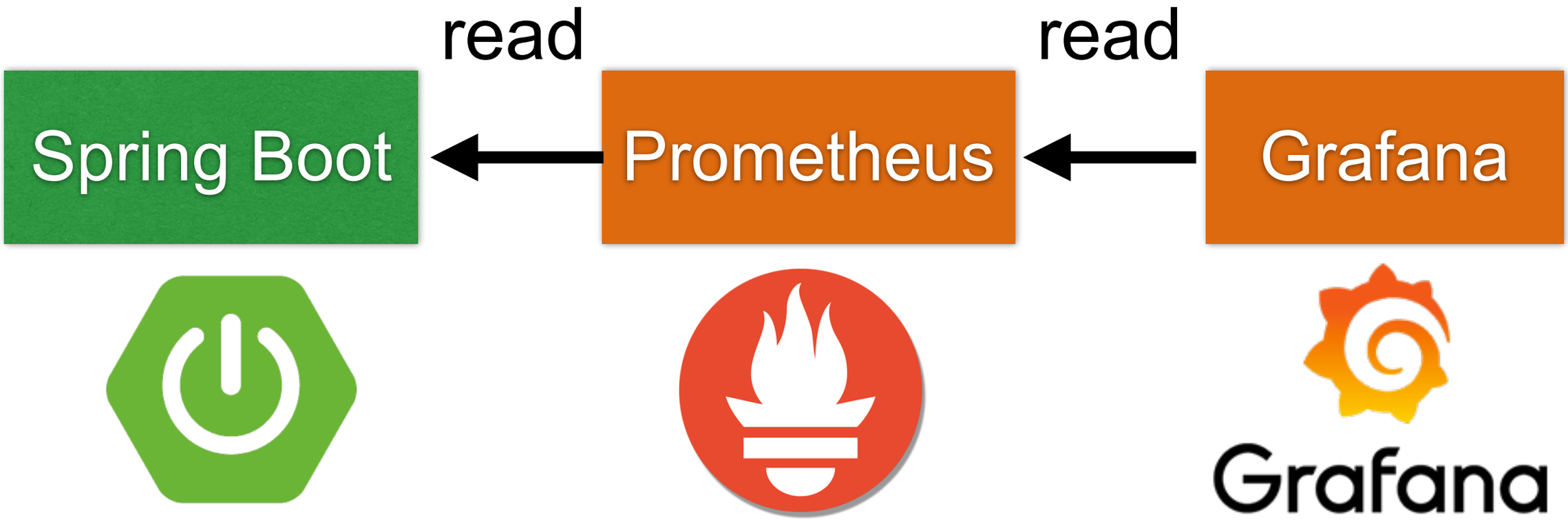
Example

https://prometheus.io/docs/instrumenting/exporters/

Workshop

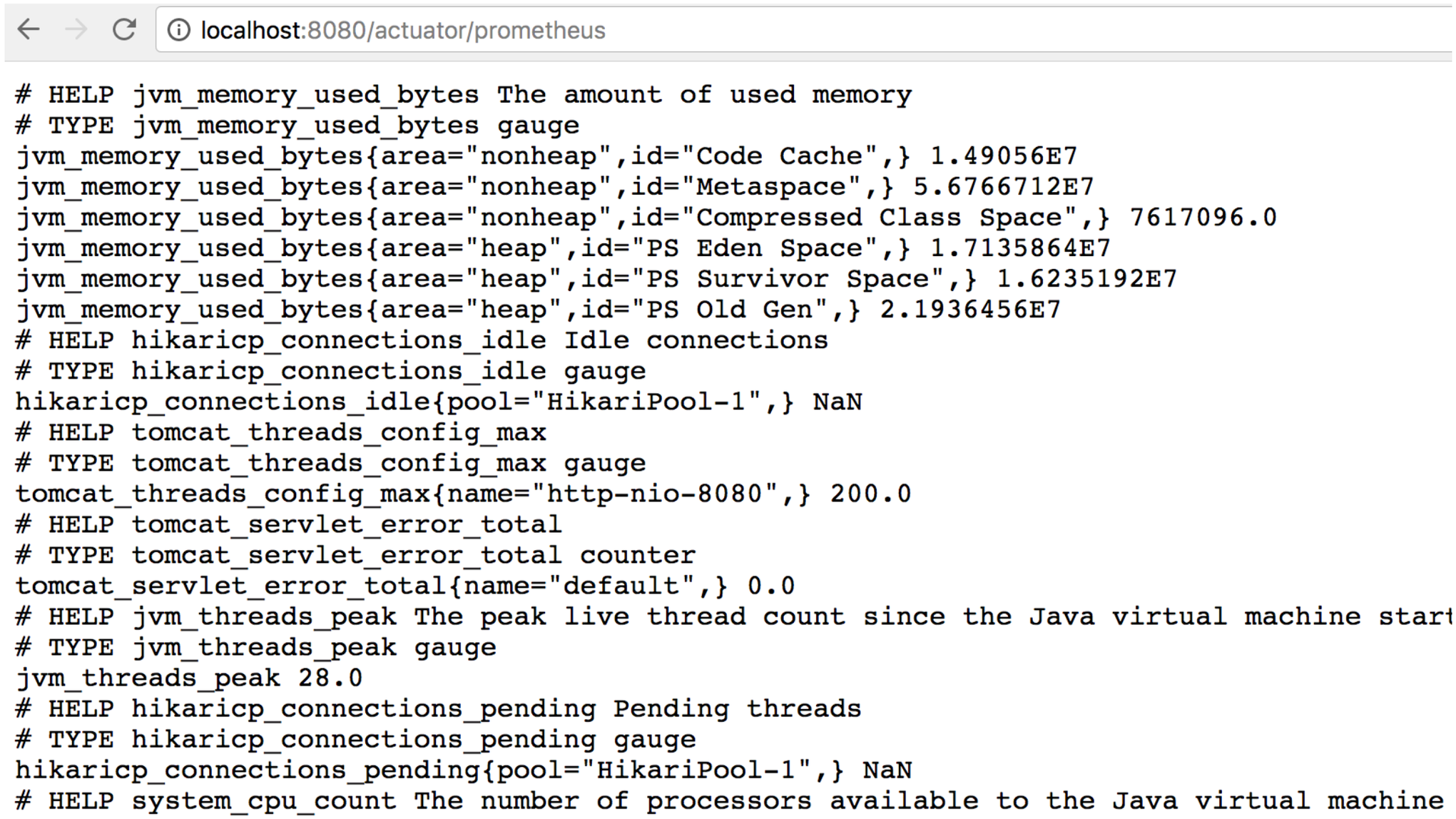
workshop/spring-prometheus-grafana

Sample Architecture



Enable Prometheus

New endpoint = actuator/prometheus



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.

Docker Pull Command

```
docker pull prom/prometheus
```

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

Owner

prom

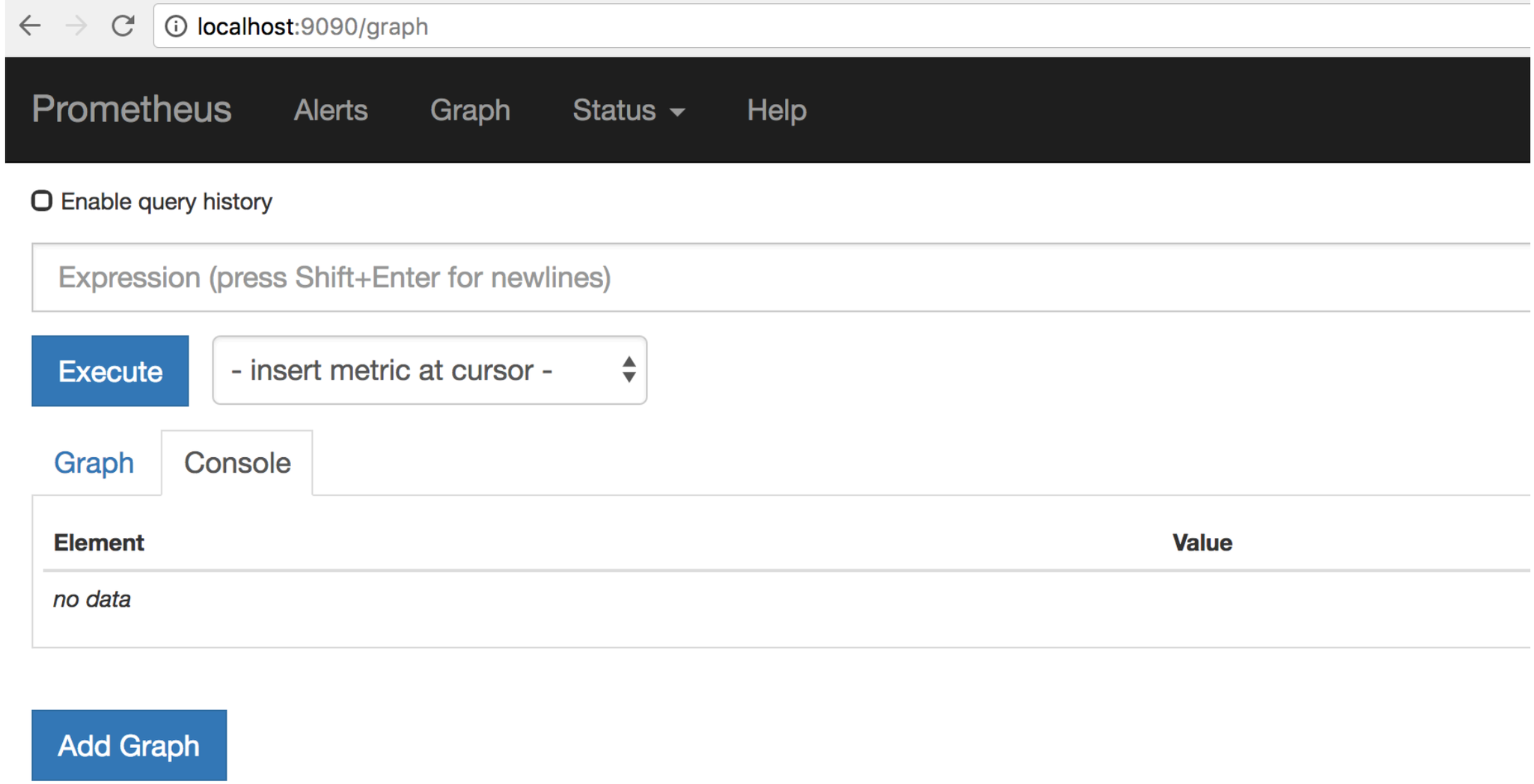
Source Repository

prometheus/prometheus

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

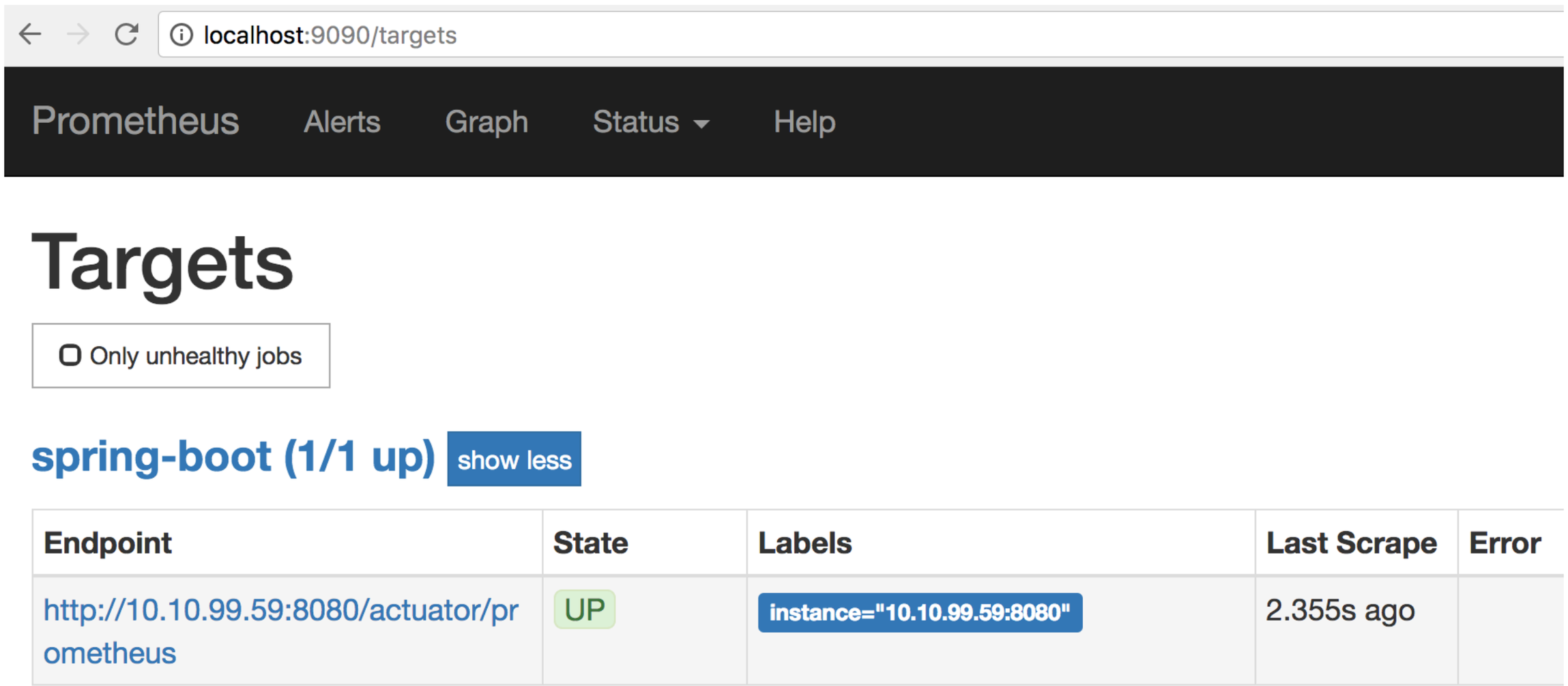
Check Data in Prometheus

http://localhost:9090/



Check Target in Prometheus

Status -> Targets



Show data in Grafana

https://grafana.com/

Grafana

https://grafana.com/

Grafana

Open source
Specialized in visualization
Plugable
Multiple datasources
Has an API/ Web app

Grafana

Fork from Kibana 3
Develop by JavaScript

Grafana

PUBLIC REPOSITORY

grafana/grafana ☆

Last pushed: 25 minutes ago

Repo Info Tags

Short Description

The official Grafana docker container

Docker Pull Command

```
docker pull grafana/grafana
```

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
```

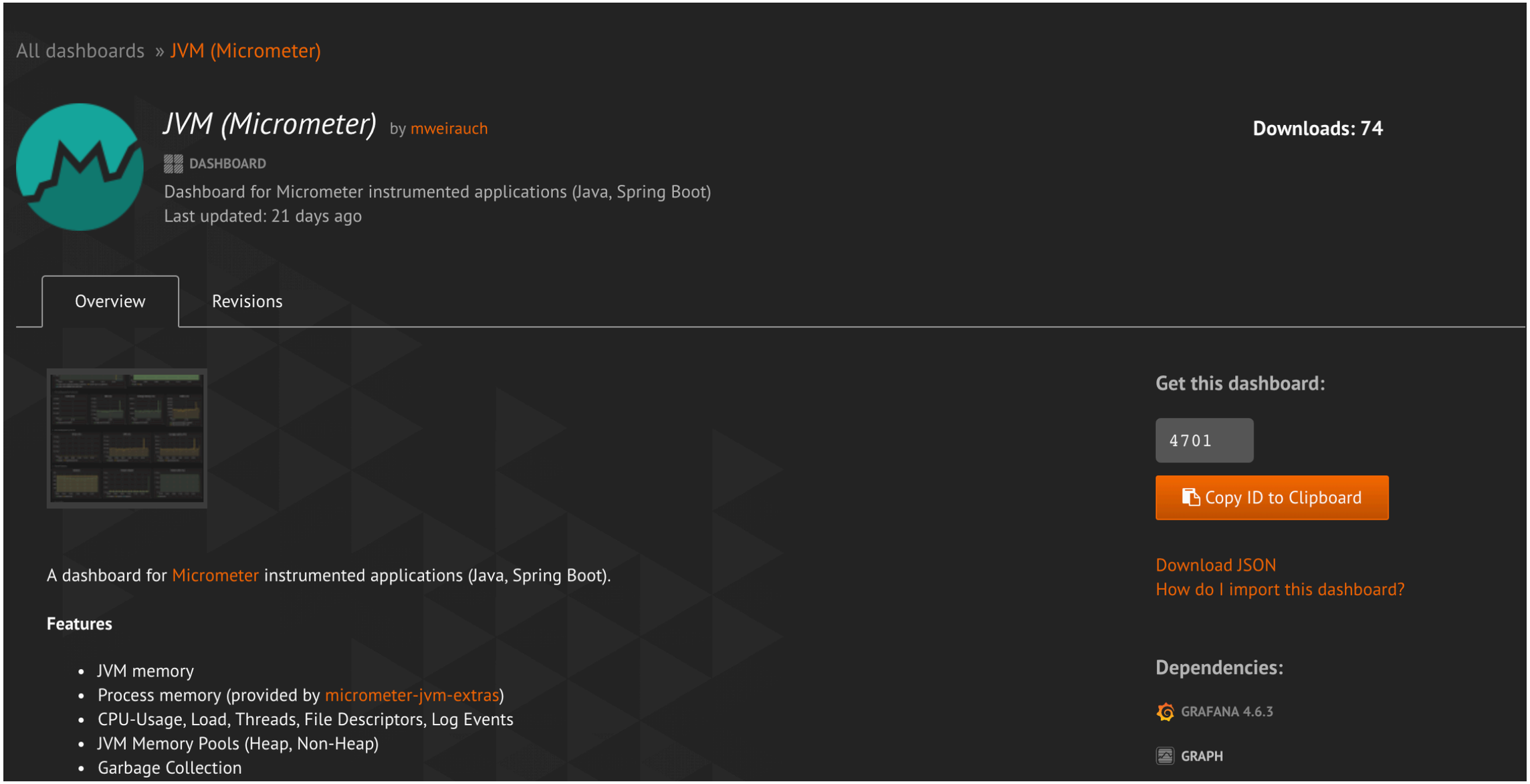
Owner

grafana

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

Grafana

https://grafana.com/dashboards/4701



Workshop

https://grafana.com/