Scale Out

5.6.1

How much data can Elasticsearch index handle ?

<= 2.1 billion documents

<= 274 billion distinct term

https://lucene.apache.org/core/4_9_0/core/org/apache/lucene/codecs/lucene49/package-summary.html#Limitations

Add nodes to cluster

Benefits

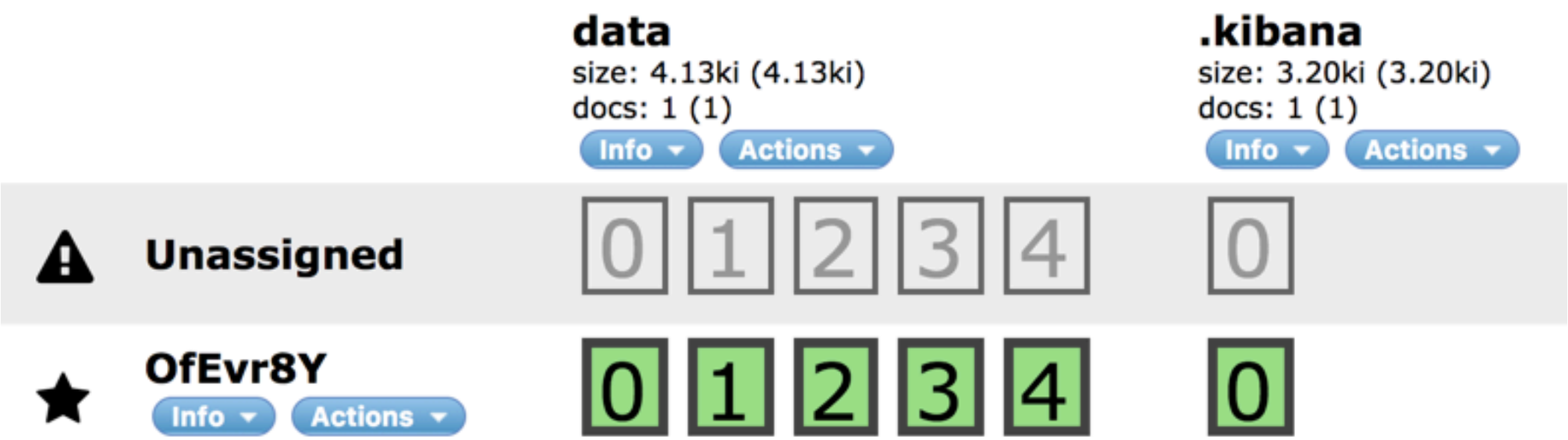
High availability
Increase performance

Cluster Health

http://localhost:9200/_cluster/health


```
{
    cluster_name: "elasticsearch",
    status: "yellow",
    timed_out: false,
    number_of_nodes: 1,
    number_of_data_nodes: 1,
    active_primary_shards: 6,
    active_shards: 6,
    relocating_shards: 0,
    initializing_shards: 0,
    unassigned_shards: 6,
    delayed_unassigned_shards: 6,
    number_of_pending_tasks: 0,
    number_of_in_flight_fetch: 0,
    task_max_waiting_in_queue_millis: 0,
    active_shards_percent_as_number: 50
}
```

Cluster Health



Single Node

$./bin/elasticsearch

Single Node

Change configuration in elasticsearc.yml

node.max_local_storage_nodes: 2

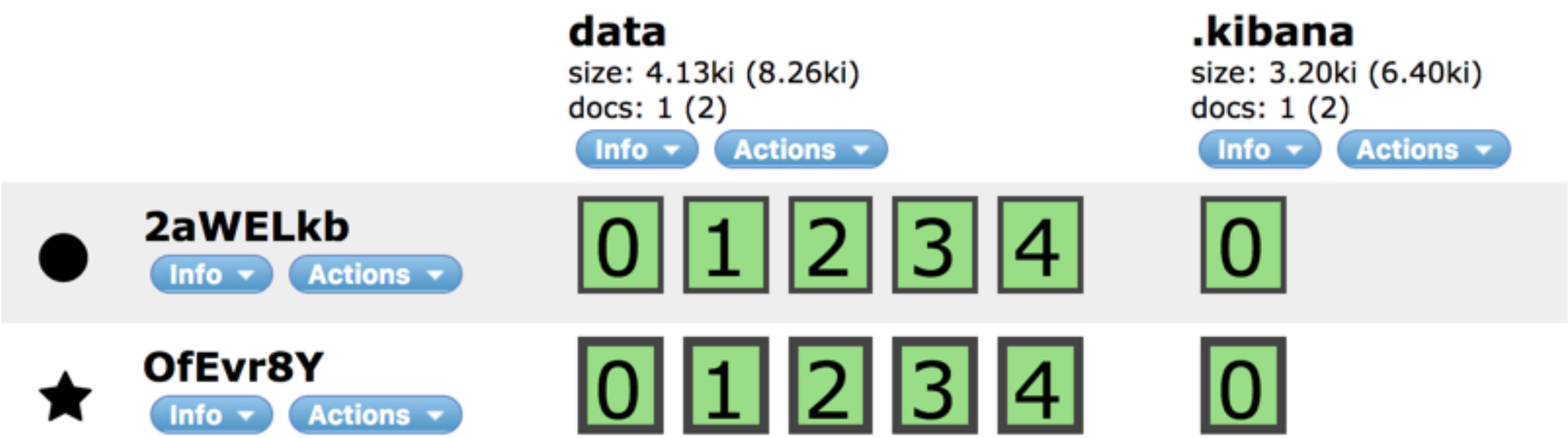
https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-node.html

Cluster Health

http://localhost:9200/_cluster/health

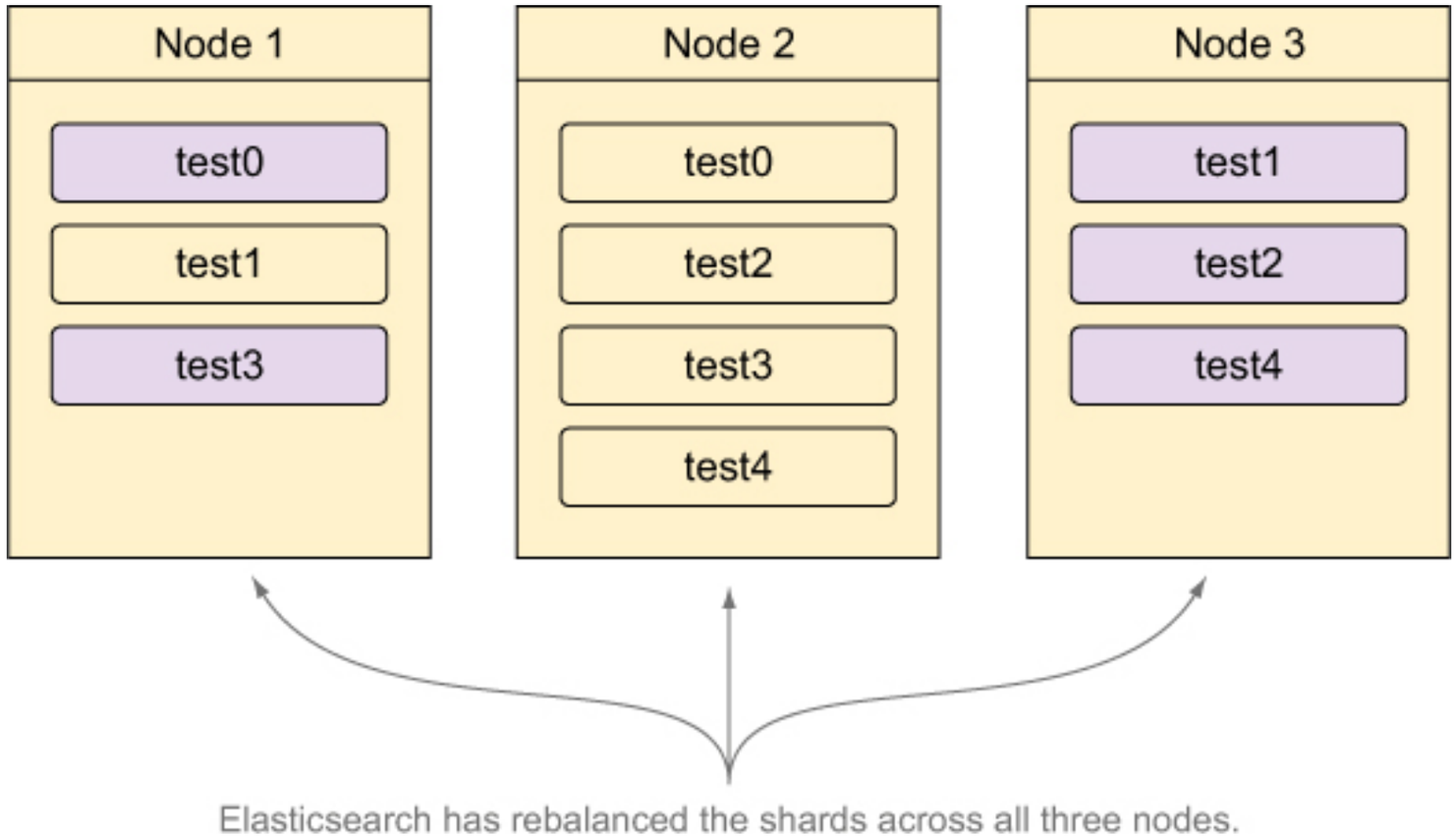

```
{
    cluster_name: "elasticsearch",
    status: "green",
    timed_out: false,
    number_of_nodes: 2,
    number_of_data_nodes: 2,
    active_primary_shards: 6,
    active_shards: 12,
    relocating_shards: 0,
    initializing_shards: 0,
    unassigned_shards: 0,
    delayed_unassigned_shards: 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
}
```

Cluster Health



Add more node !!

$./bin/elasticsearch



Start new node with config

$./bin/elasticsearch \
-Epath.conf=/Users/somkiat/node01

```
cluster.name: your_cluster_name
node.name: node02
node.master: false
node.data: true
node.ingest: false
path.data: /config/path/node02
```

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

Zen Discovery

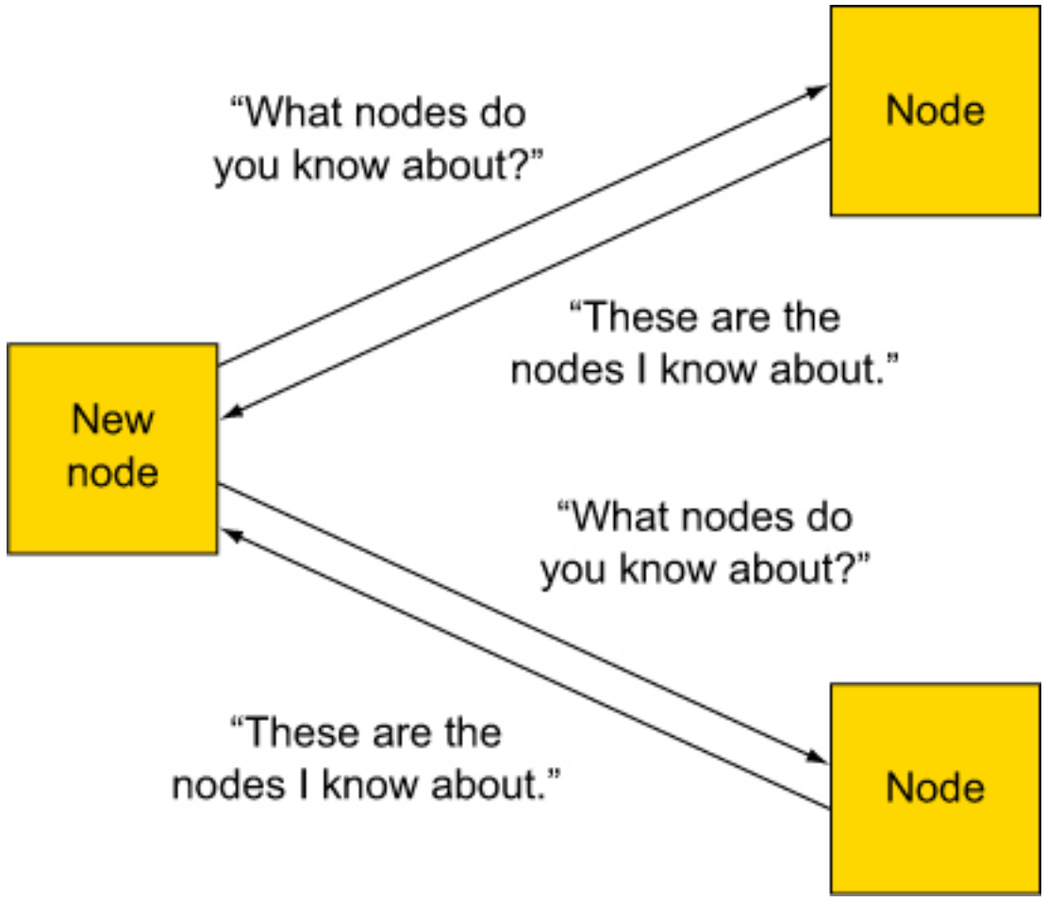
Discovery method

Unicast discovery (default)

Multicast discovery (plugin)

Unicast discovery

Use a list of hosts for Elasticsearch

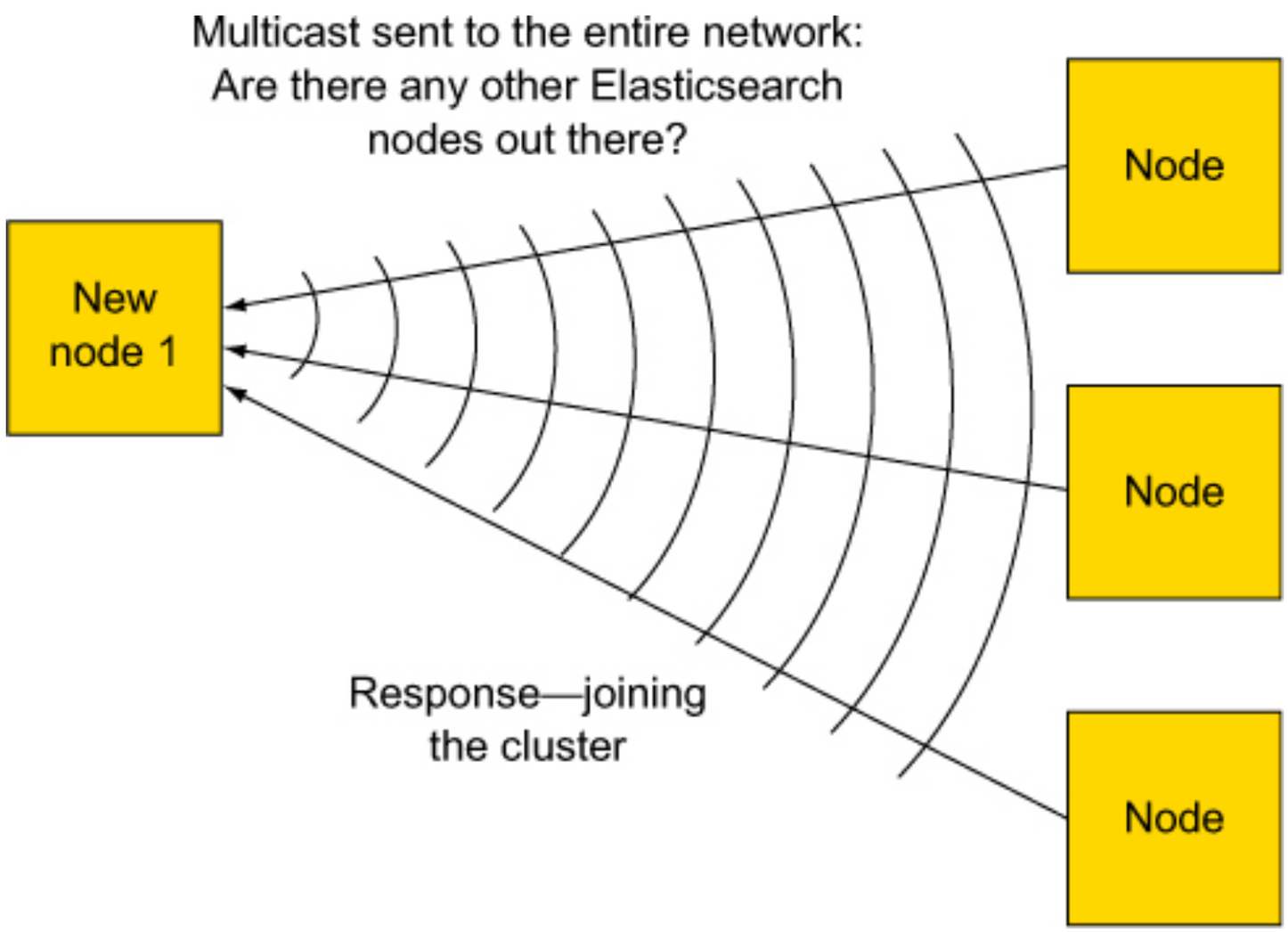


Unicast discovery

Change config in elasticsearch

discovery.zen.ping.unicast.hosts: ["10.0.0.3", "10.0.0.4:9300", "10.0.0.5[9300-9400]"]

Multicast discovery



Master election

https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-discovery-zen.html#master-election

Master node

Manage the state of cluster
Settings
State of shards, indices and nodes

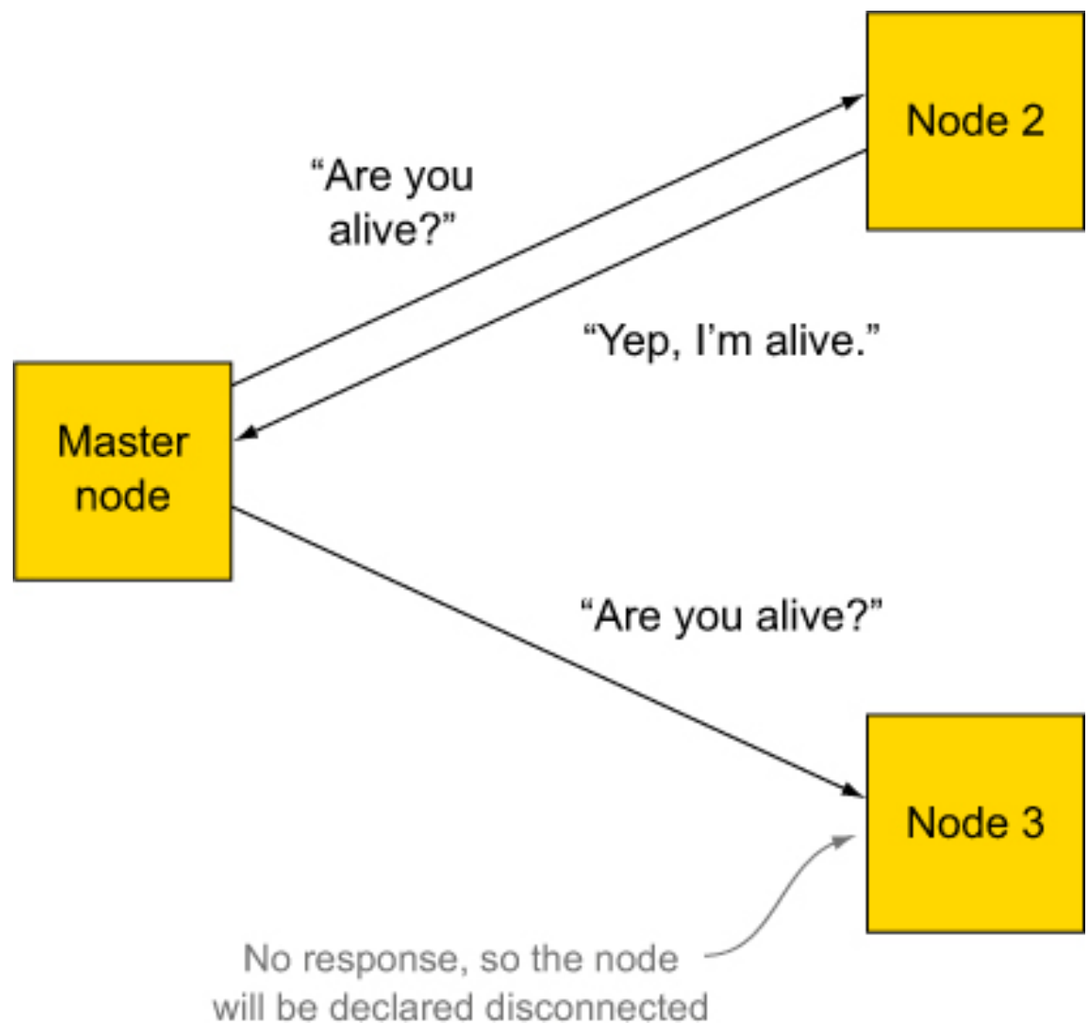
node.master: true

Master election

discovery.zen.ping_timeout: 3s
discovery.zen.join_timeout: 20

Fault detection

Default is 1 seconds



Fault detection

discovery.zen.fd.ping_interval: 1s
discovery.zen.fd.ping_timeout: 30s
discovery.zen.fd.ping_retries: 3

In production cluster

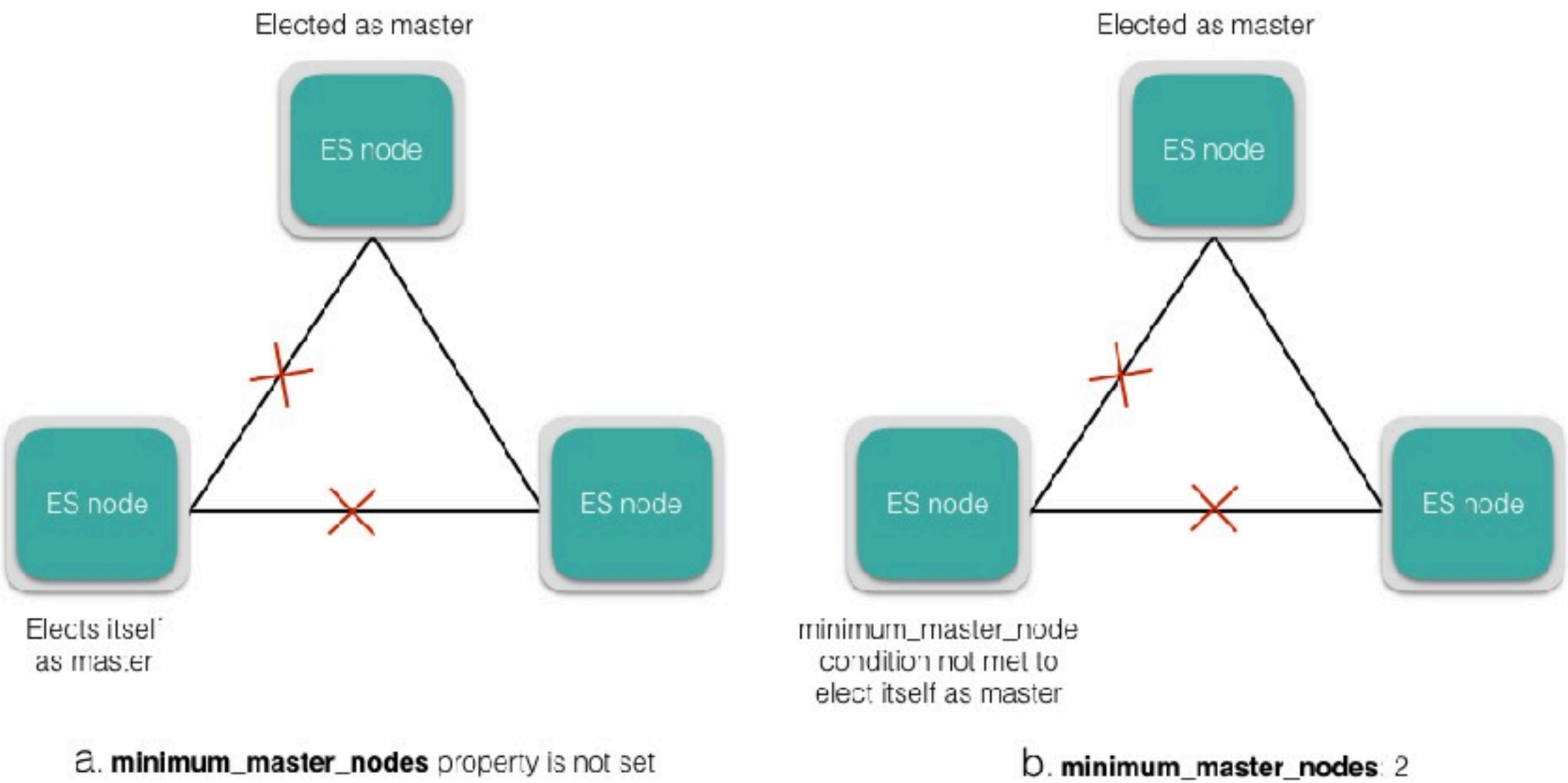
Need to set the minimum number of master node

(number of nodes / 2) + 1

> 1 to prevent the split brain problem

discovery.zen.minimum_master_nodes: 2

Split brain


Elected as master
ES node
ES node
ES node
Elects itself as master
Elected as master
ES node
ES node
ES node
minimum_master_node condition not met to elect itself as master
a. **minimum_master_nodes** property is not set
b. **minimum_master_nodes**: 2

Get information of nodes in cluster

http://localhost:9200/_cluster/state/master_node,nodes


```
{
  cluster_name: "elasticsearch",
  master_node: "OfEvr8Y6TXGVM93q3z5uyg",
- nodes: {
    - 2aWELkbBRdWIC4fIKToFRw: {
          name: "2aWELkb",
          ephemeral_id: "p_-TNriIRk6MIvIHLf3ylw",
          transport_address: "127.0.0.1:9301",
          attributes: { }
      },
    - OfEvr8Y6TXGVM93q3z5uyg: {
          name: "OfEvr8Y",
          ephemeral_id: "5mGoqywDRWiIR-xCllgm4g",
          transport_address: "127.0.0.1:9300",
          attributes: { }
      }
  }
}
```

Get information of nodes in cluster

http://localhost:9200/_cluster/state/master_node,nodes

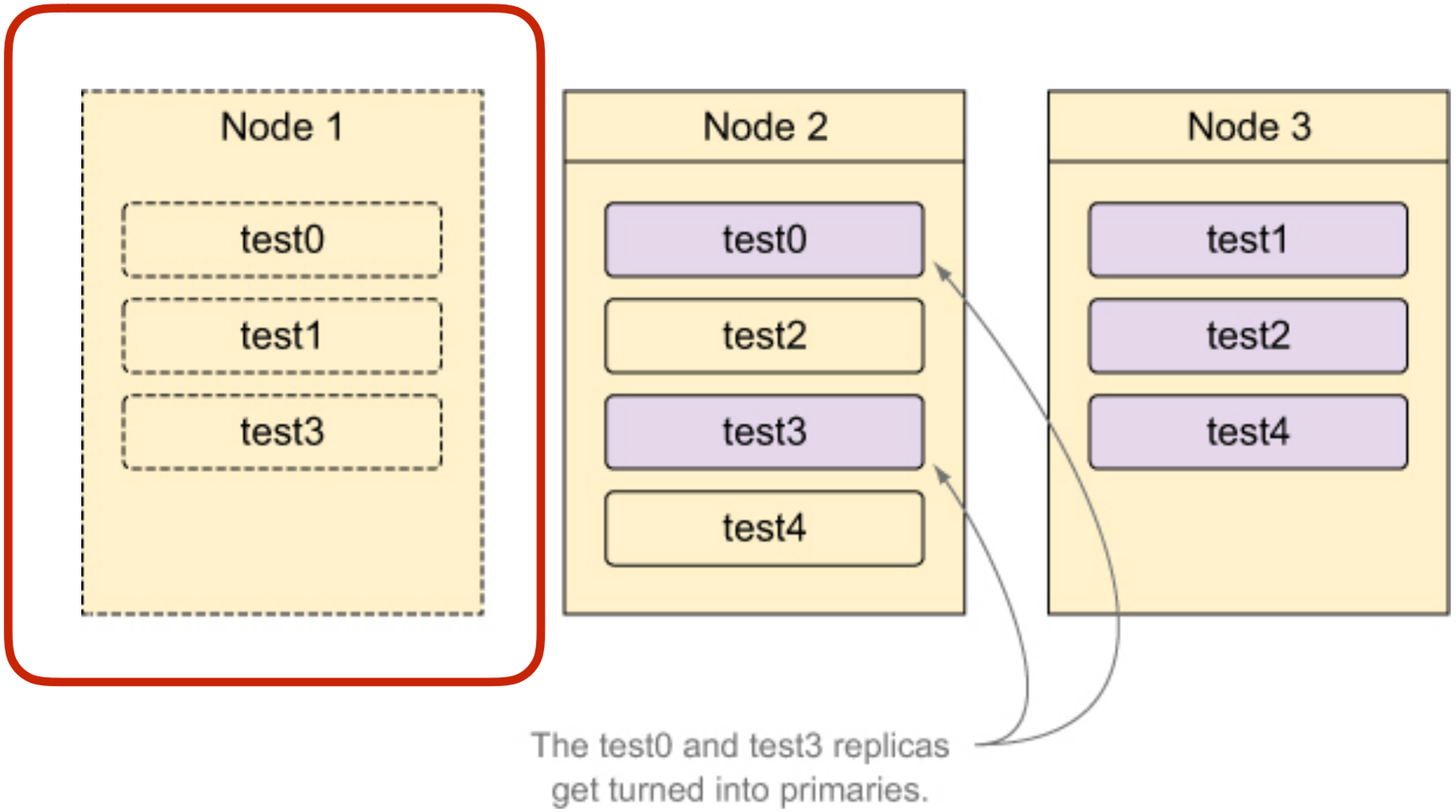

```
{
    cluster_name: "elasticsearch",
    master_node: "OfEvr8Y6TXGVM93q3z5uyg",
  - nodes: {
      - 2aWELkbBRdWIC4fIKToFRw: {
            name: "2aWELkb",
            ephemeral_id: "p_-TNriIRk6MIvIHLf3ylw",
            transport_address: "127.0.0.1:9301",
            attributes: { }
        },
      - OfEvr8Y6TXGVM93q3z5uyg: {
            name: "OfEvr8Y",
            ephemeral_id: "5mGoqywDRWiIR-xCllgm4g",
            transport_address: "127.0.0.1:9300",
            attributes: { }
        }
    }
}
```

Remove nodes from cluster

https://www.elastic.co/guide/en/elasticsearch/reference/5.6/modules-cluster.html

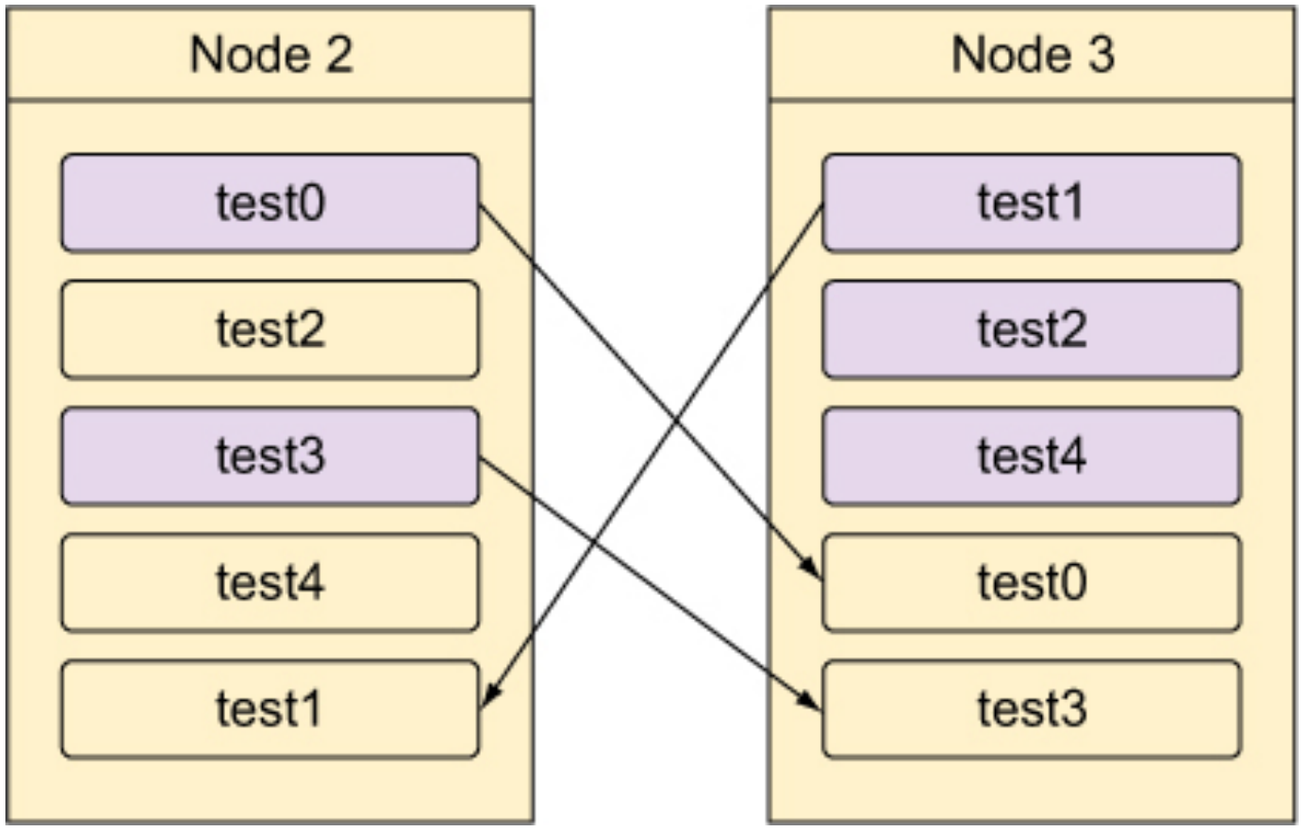
What happen when a node drop ?

Drop/Stop/Remove node from cluster



replica = 1

After remove node 1



replica = 1

Remove node before shutdown

By _name, _ip, _host

```
PUT _cluster/settings
{
  "transient" : {
    "cluster.routing.allocation.exclude._ip" : "10.0.0.1"
  }
}
```

https://www.elastic.co/guide/en/elasticsearch/reference/current/allocation-filtering.html

Monitoring for bottlenecks

Elasticsearch provides APIs

Memory consumption
Node membership
Shard distribution
I/O performance

Check cluster health

http://localhost:9200/_cluster/health


```
{
    cluster_name: "elasticsearch",
    status: "green",
    timed_out: false,
    number_of_nodes: 2,
    number_of_data_nodes: 2,
    active_primary_shards: 6,
    active_shards: 12,
    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
}
```

Status mean Cluster performance

Green

Yellow

Red

Green

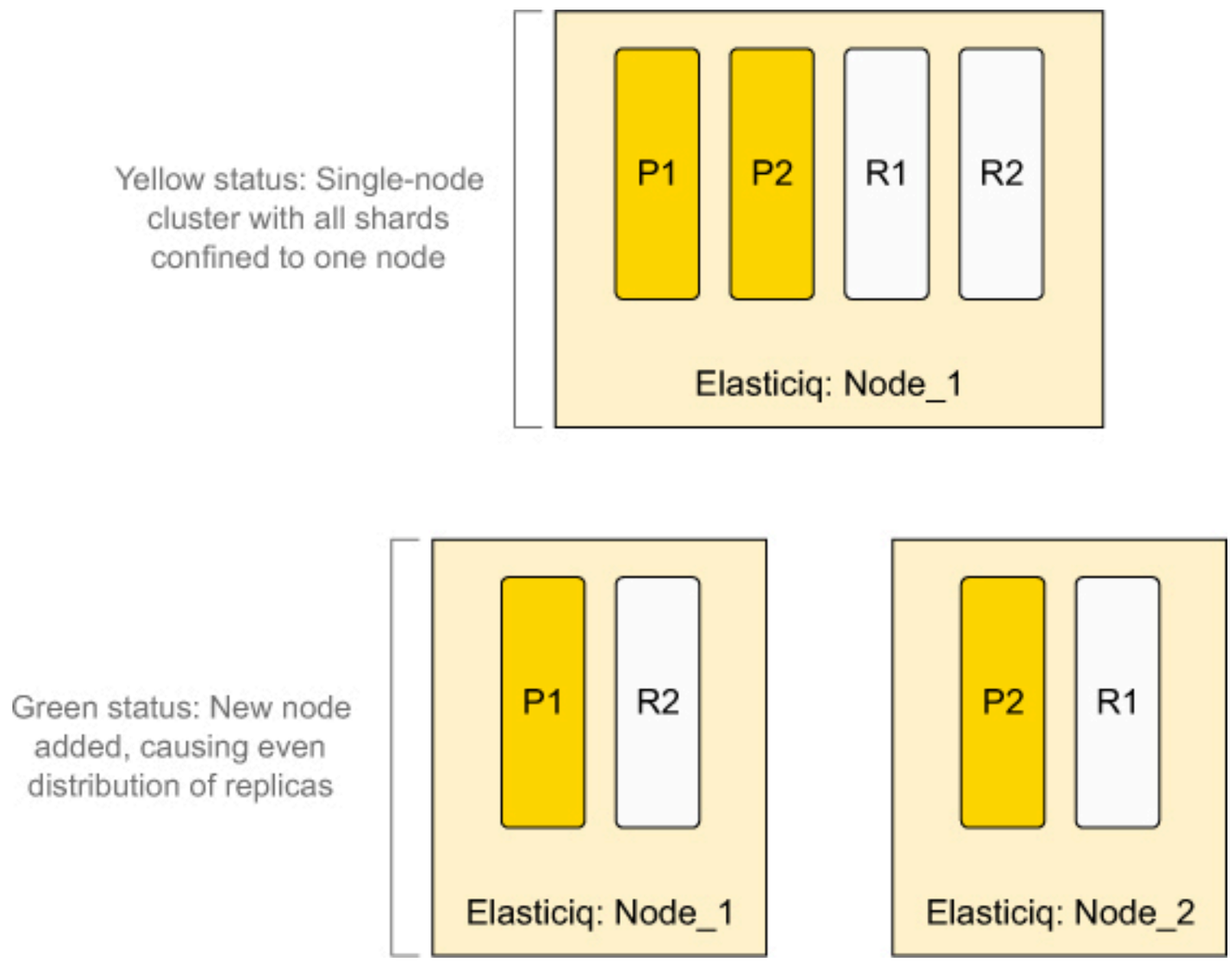
Primary and replica shards are fully functional and distributed

Yellow

Missing replica shards
Cluster unstable
Lead to data loss
Some nodes aren’t initialized

Yellow

Making node accessible



Red

Critical state of cluster
Primary shard in the cluster not found
Prohibiting indexing operation
Lead to inconsistency query result
Some nodes aren missing in cluster

Check cluster health

http://localhost:9200/_cluster/health


```
{
    cluster_name: "elasticsearch",
    status: "green",
    timed_out: false,
    number_of_nodes: 2,
    number_of_data_nodes: 2,
    active_primary_shards: 6,
    active_shards: 12,
    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
}
```

Relocating_shards

Moving shards of data across the cluster
to improve balance and failover

Occur when add new node, restart/remove node

Initializing_shards

Occur when created a new index
or restarted a node

Unassigned_shards

Occur when unassigned replica

Check cluster health

Task APIs (_tasks)

```
{
    cluster_name: "elasticsearch",
    status: "green",
    timed_out: false,
    number_of_nodes: 2,
    number_of_data_nodes: 2,
    active_primary_shards: 6,
    active_shards: 12,
    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
}
```

https://www.elastic.co/guide/en/elasticsearch/reference/current/cluster-pending.html

Slow log

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

Slow log

Slow search log (query/fetch)
Slow index log

Disable by default !!

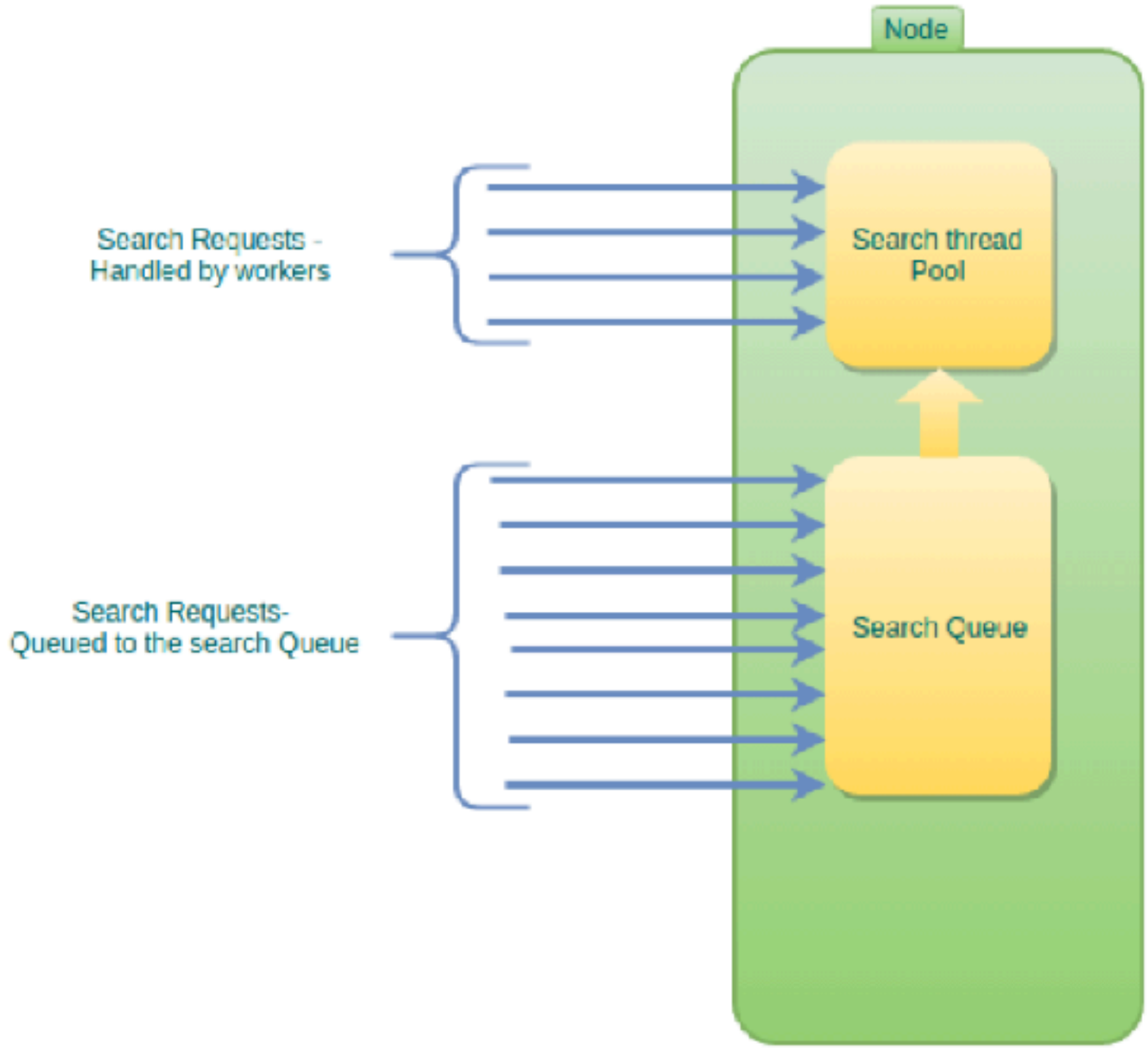
Enable Slow log

```
PUT /sample/_settings
{
  "index.search.slowlog.threshold.query.warn": "0s",
  "index.search.slowlog.threshold.fetch.warn": "0s",
  "index.indexing.slowlog.threshold.index.warn": "0s"
}
```

Thread pool of cluster

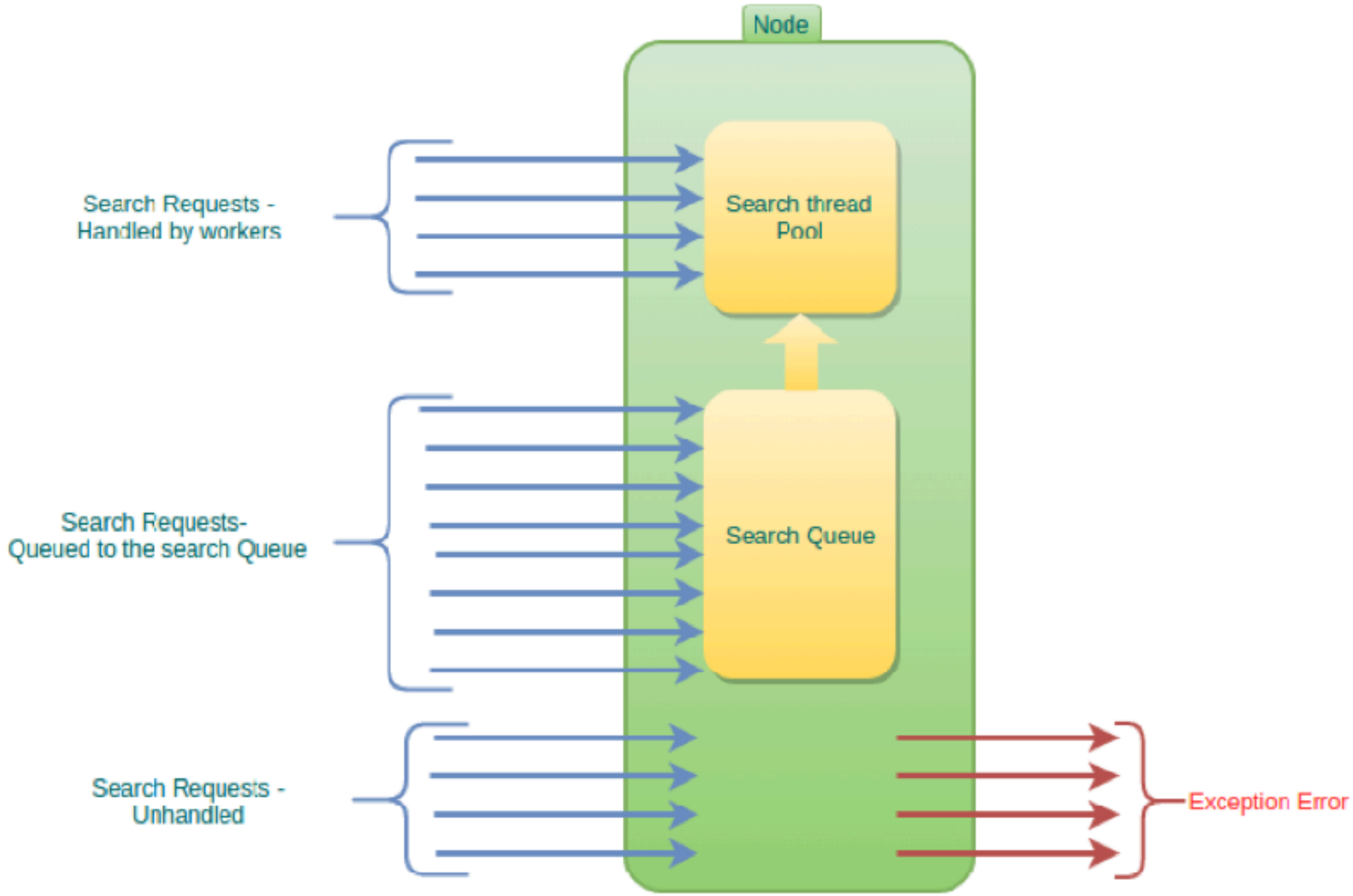
https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-threadpool.html

Thread pool



https://qbox.io/blog/thread-pools-elasticsearch-search-request-errors

Request exceed limit !!



https://qbox.io/blog/thread-pools-elasticsearch-search-request-errors

Solution

Increase the size of thread pool
Increase the size of search pool
Increase the nodes and replicas

Using _cat API

_cat APIs provides

Diagnostic and debugging tool
Print data in a more human-readable

http://localhost:9200/_cat/..?v

_cat APIs provides

```
=^.^=
/_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}
/_cat/thread_pool
/_cat/thread_pool/{thread_pools}
/_cat/plugins
/_cat/fielddata
/_cat/fielddata/{fields}
/_cat/nodeattrs
/_cat/repositories
/_cat/snapshots/{repository}
/_cat/templates
```

_cat APIs provides

Name	Description
allocation	Number of shards allocated to each node
count	Number on of documents
health	Health of the the cluster
indices	Information of indices
master	Current elected master node
nodes	Information of all nodes in the cluster
recovery	Status of ongoing shard recoveries in the cluster
shards	Display count, size and names of shards in the cluster

Planning to scaling strategies

Strategies on production

Over-sharding
Split data between indices and shards
Maximize throughput

Over-shading

Create a larger number of shards for an index
Each shard requires a number of file descriptor
Memory overhead

Number of sharing ?

No perfect shard-to-index ratio for all use case

Elasticsearch picks a good default = 5

Memory overhead

Split data in to indices

Maximize throughput

Maximize throughput ?

Indexing throughput
Search faster

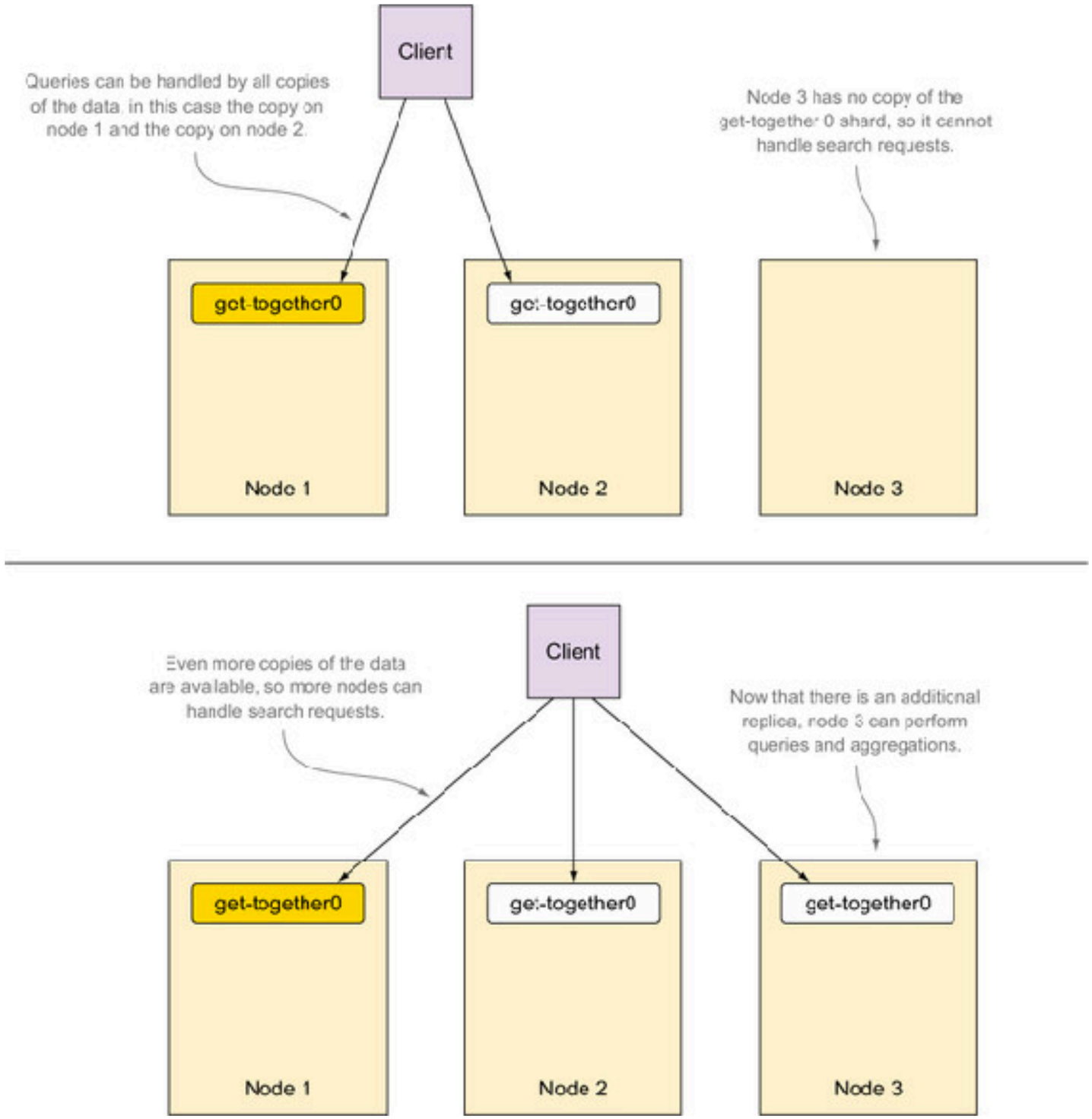
Receive many of data ?

How to indexing data as fast as possible ?

Receive many of data ?

Temporary reduce the number of replica
(= 0 if you’re OK with the risk)

Replica for search



More query ?

add new node with

node.master: false
node.data: false

Alias and custom routing

Default Shard of document

shard_num = hash(id) % num_primary_shards

https://en.wikipedia.org/wiki/MurmurHash

Default Routing

shard_num = hash(_routing) % num_primary_shards

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

Using _routing

PUT my_index/my_type/1?**routing=user1**

```
{
  "title": "This is a document”
}
```

GET my_index/my_type/1?**routing=user1**

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

Improve performance

Use cases ?

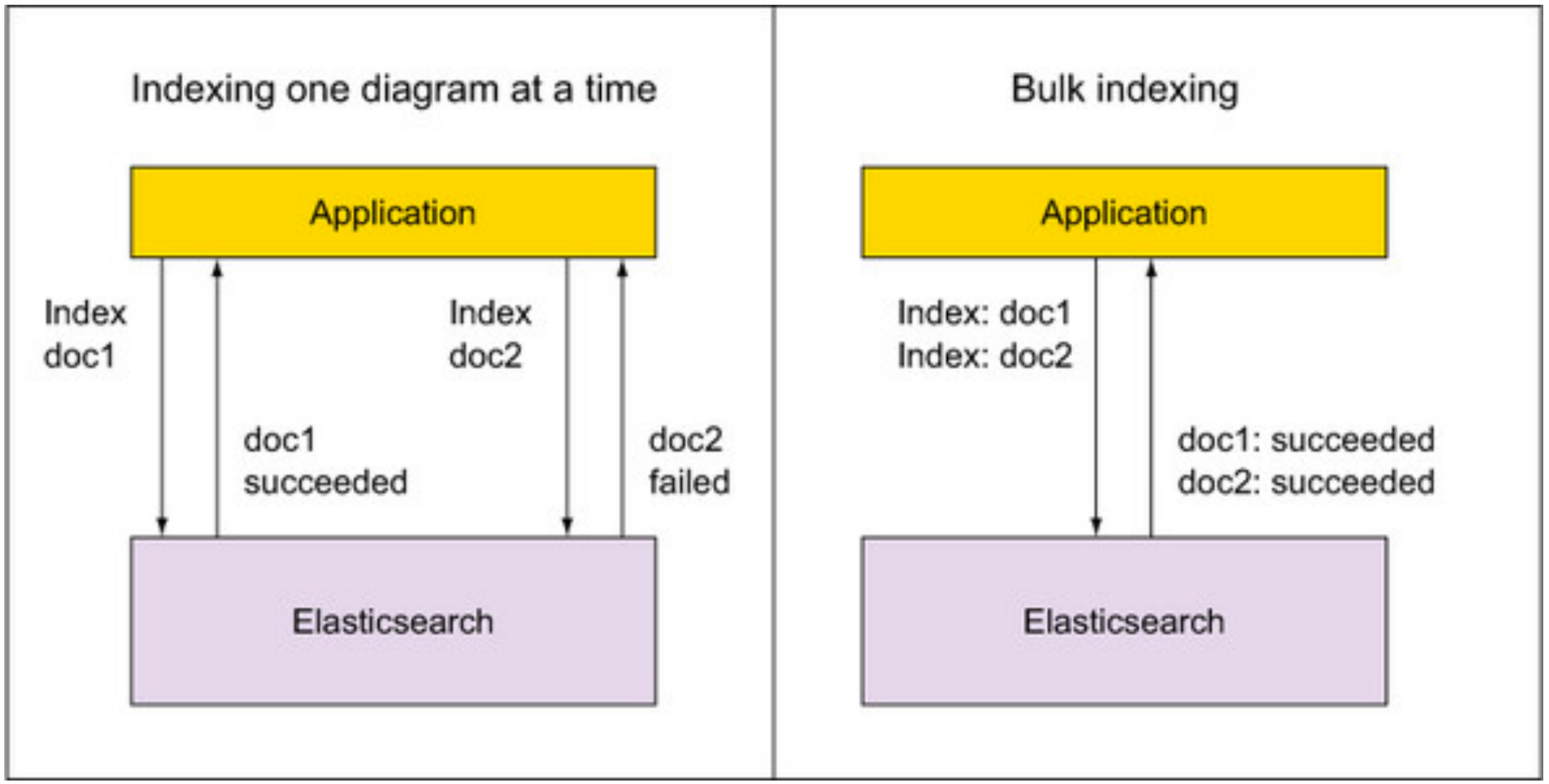
Application complexity
Indexing speed for search speed
Memory

Grouping request

Bulk indexing, updating and deleting
Multiple search and get

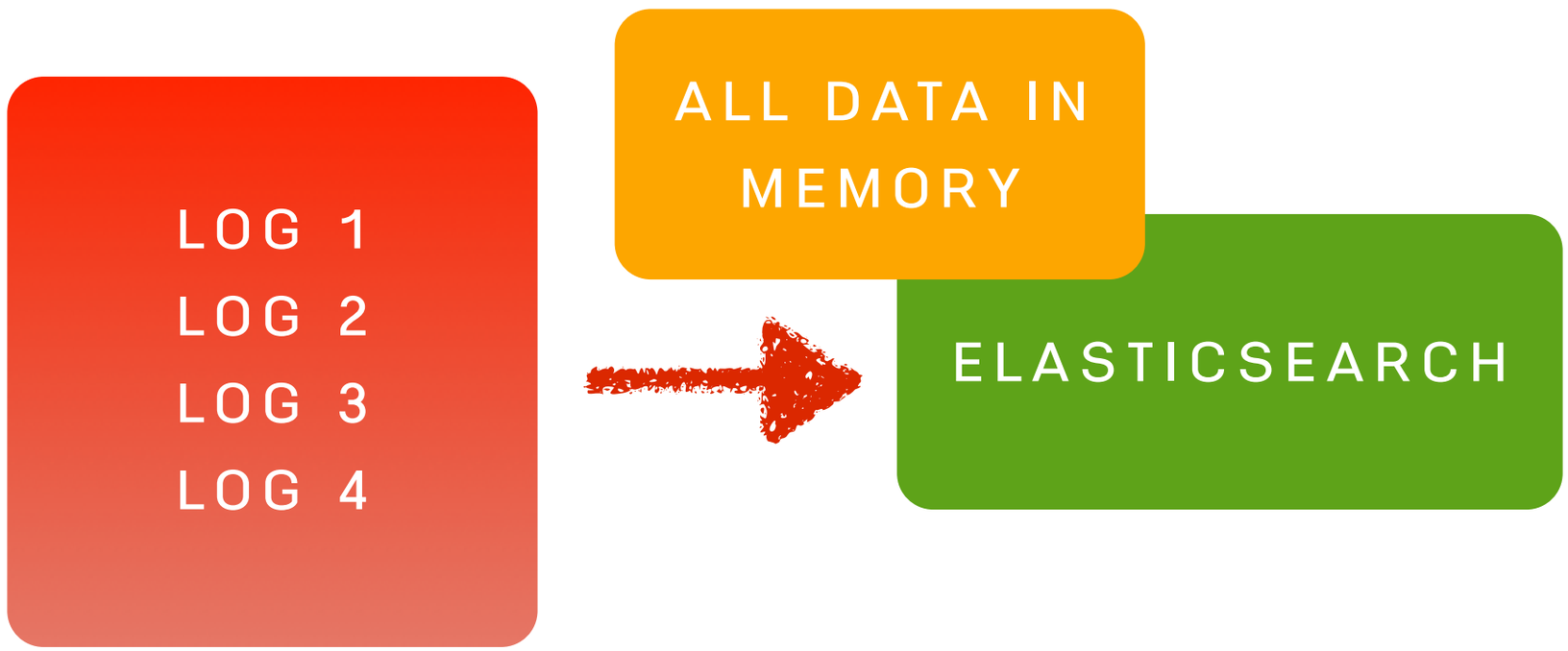
Bulk API

improve speed 20-40%



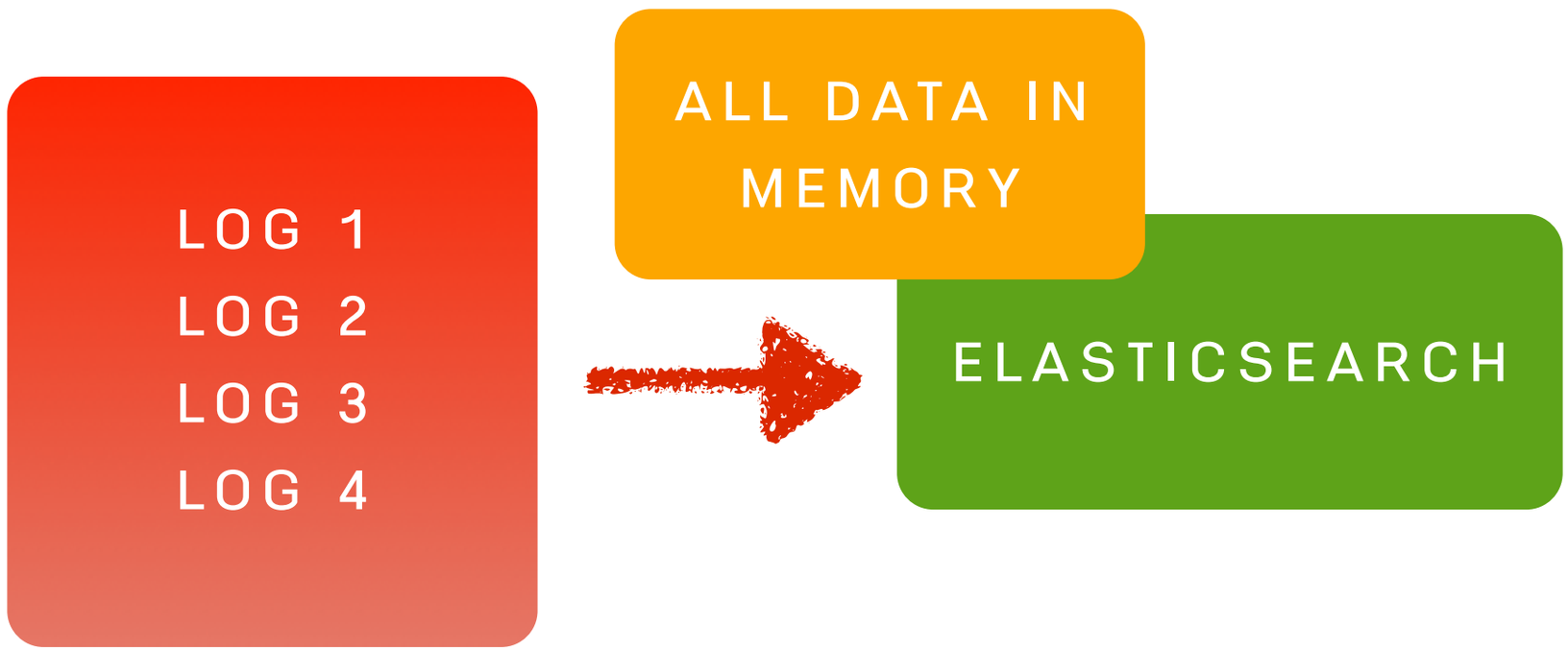
Bulk API

Size of data ?



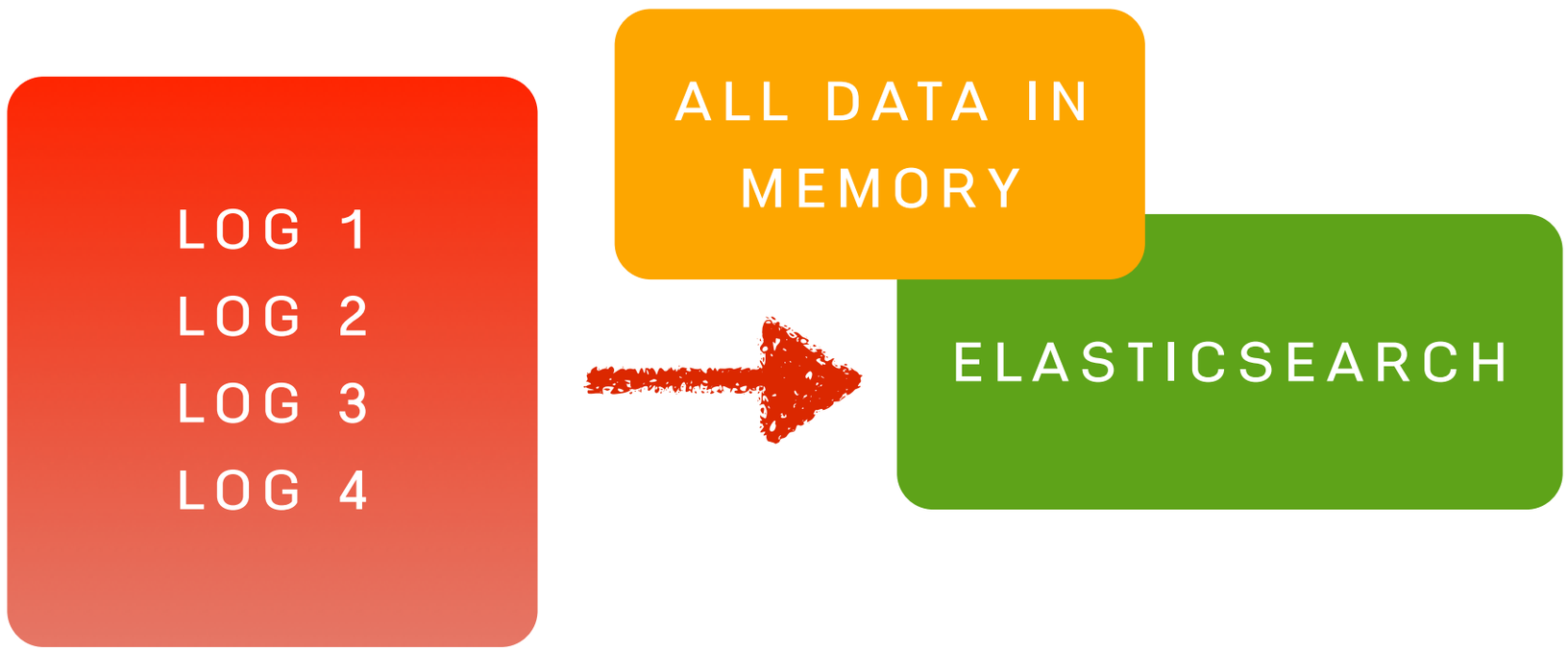
Bulk API

Depend on your application and use cases



Bulk API

Recommend is 5- 15 MB



Multiple search and get

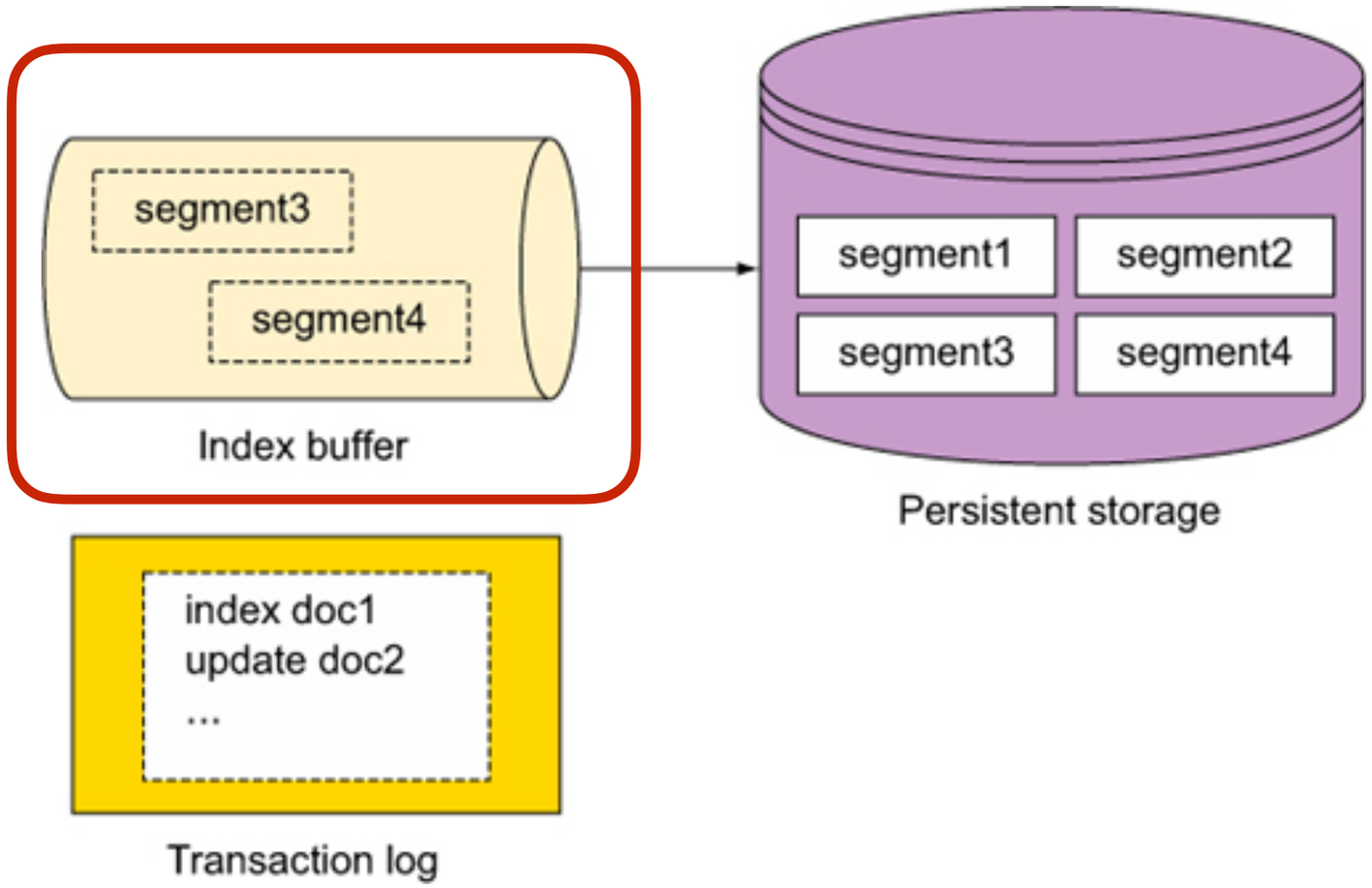
Reduce network latency
Better for search and get data

https://www.elastic.co/guide/en/elasticsearch/reference/current/search-multi-search.html

Optimizing the handling of Lucene segment

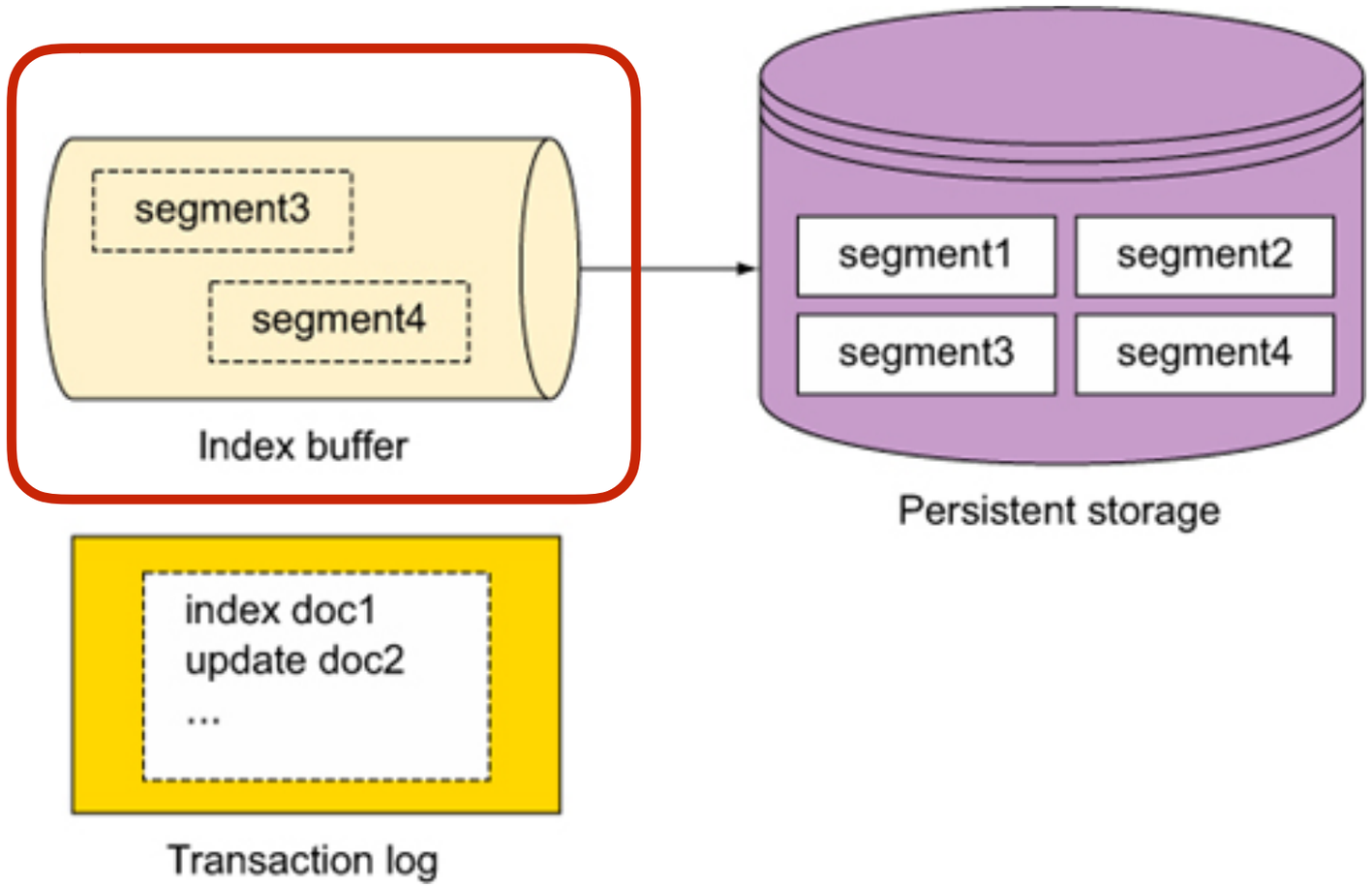
Refresh

You can search a newly indexed data



Refresh

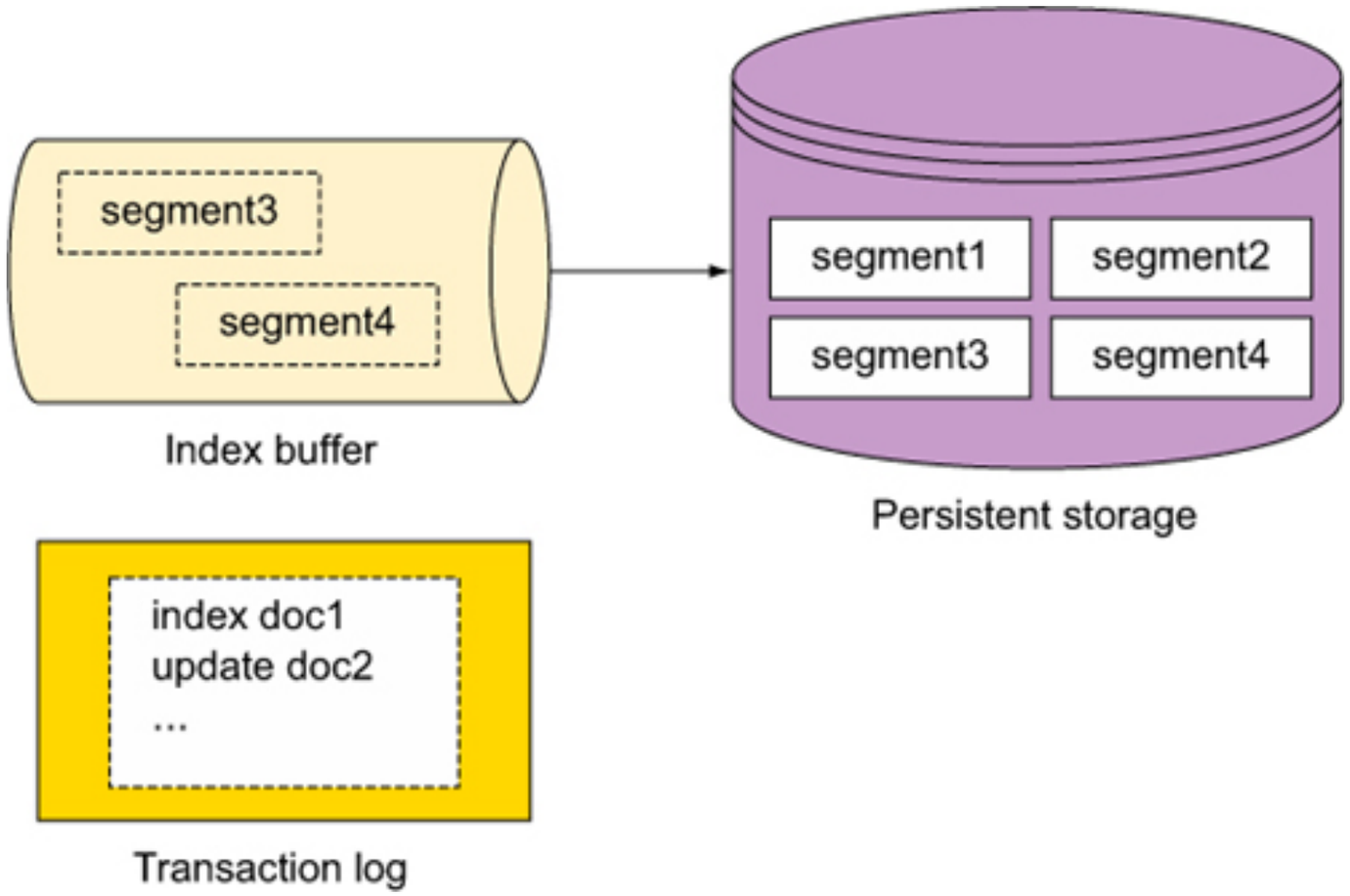
refresh_interval and _refresh



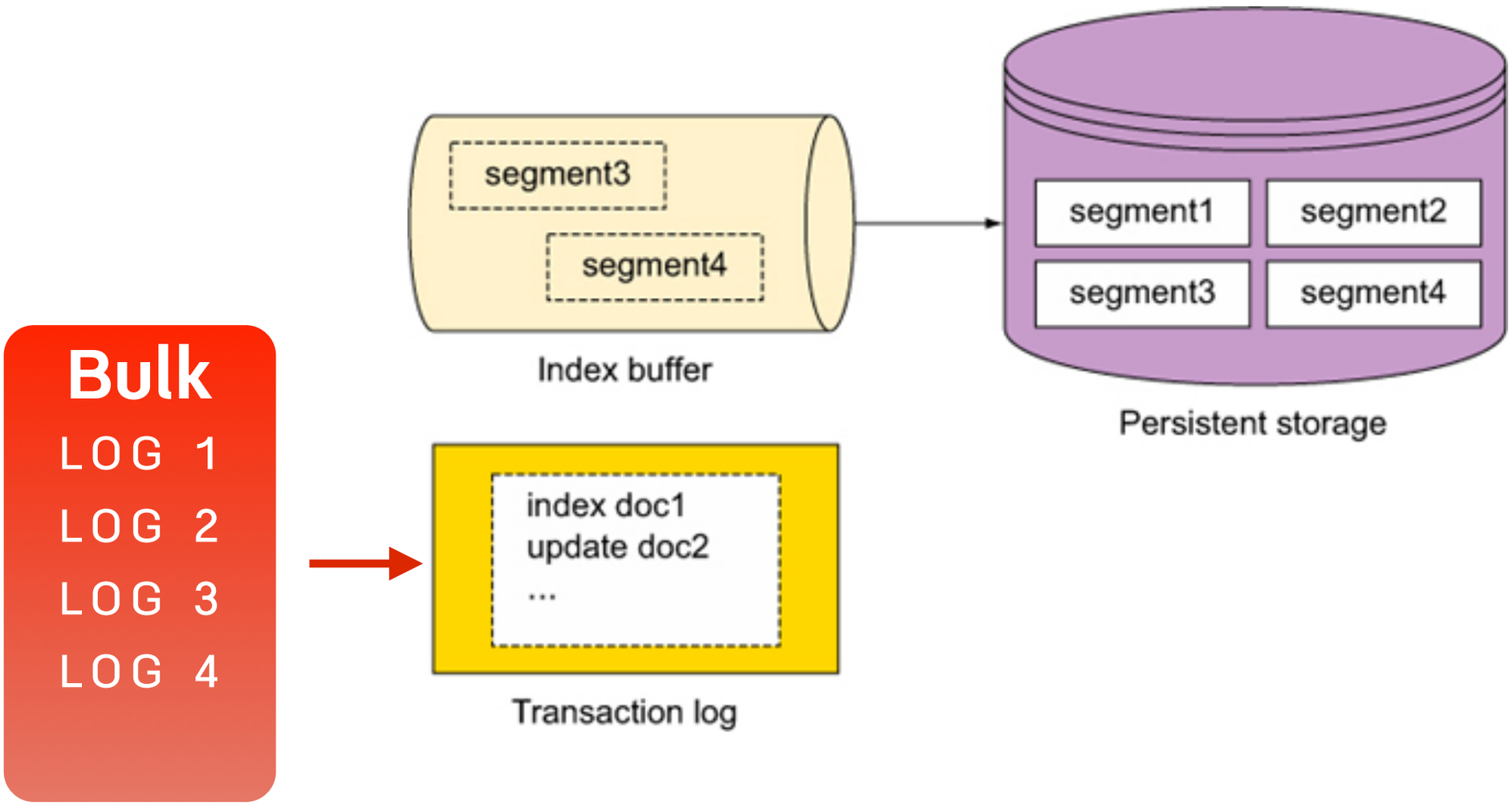
Refresh every second !!

Some caches will be invalidated
Slow down search
Slow down indexing
Reopen some process that need by itself !!

Memory to Disk ?

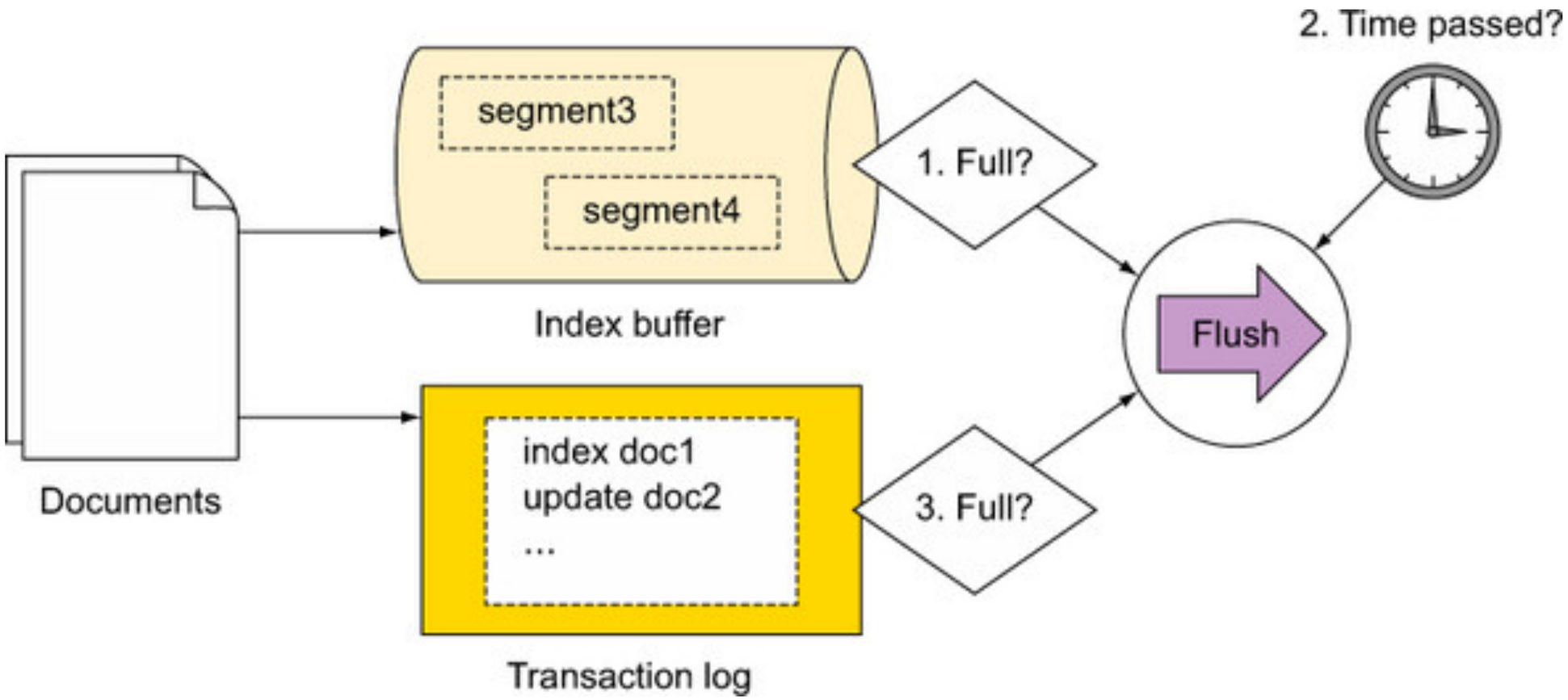


Memory to Disk ?



Flush is triggered

Memory buffer is full
Time passed
Transaction log hit a size threshold



Index buffer size

indices.memory.index_buffer_size: 10%

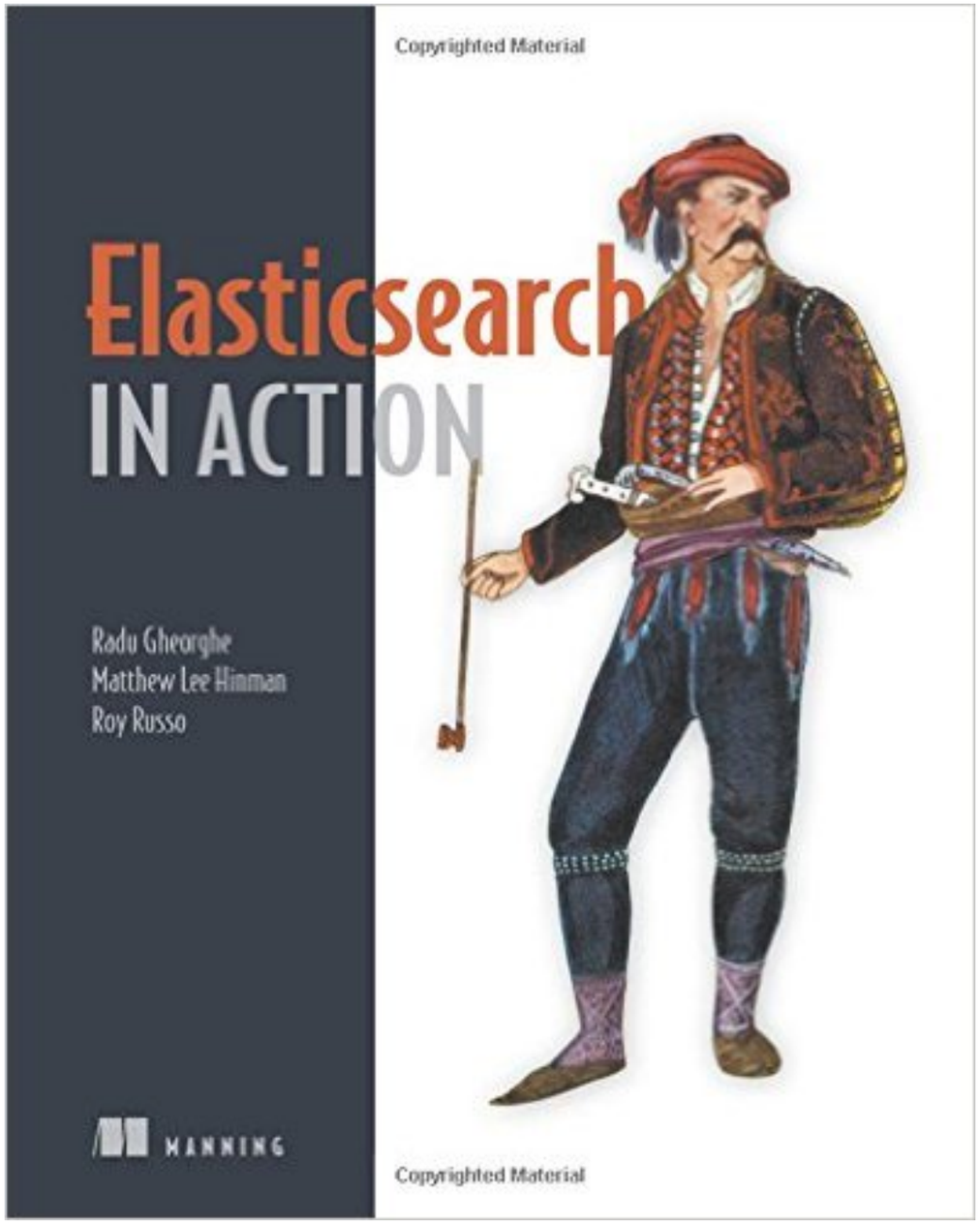
https://www.elastic.co/guide/en/elasticsearch/reference/current/indexing-buffer.html

Transaction log threshold size

index.translog.durability: request
index.translog.sync_interval: 5s

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

Resources


Copyrighted Material
Elasticsearch
IN ACTION
Radu Gheorghe
Matthew Lee Hinman
Roy Russo
MANNING
Copyrighted Material