@somkiat

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

About me

สมเกียรติ ปุ๋ยสูงเนิน

สยามชำนาญกิจ

SPRINT3R

somkiat.cc

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Agenda

Introduction

Speed by example

- Search
- Aggregation
- Operating System and Hardware

Search system

```
SELECT

FROM TABLE

WHERE TEXT LIKE '%SHIT%'
```

Report/analytic system

SELECT

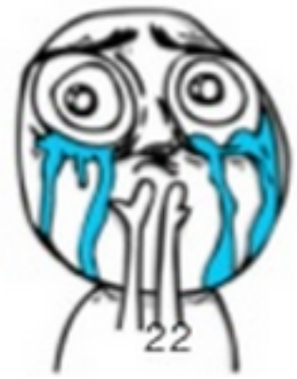
FROM many Table

GROUP BY some column

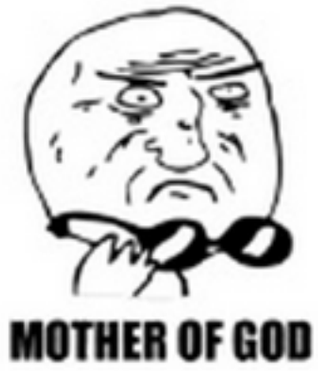
Saaaaaaaaaddddddddddd

1k records

500k records



20m records



History

Created by **Shay Banon**

First version 0.4 in Feb 2010

- Rewrite from Compass project
- Add scalability

Current version 2.1.1

Elasticsearch

an open source, distributed, scalable,

highly availability, document-oriented, RESTful

full text search engine

with near real-time search and analytics

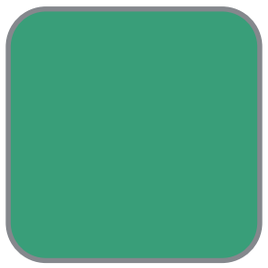
Elasticsearch

an **open source**, distributed, scalable,
highly availability, document-oriented, RESTful
full text search engine
with near real-time search and analytics

Apache 2.0 Licence

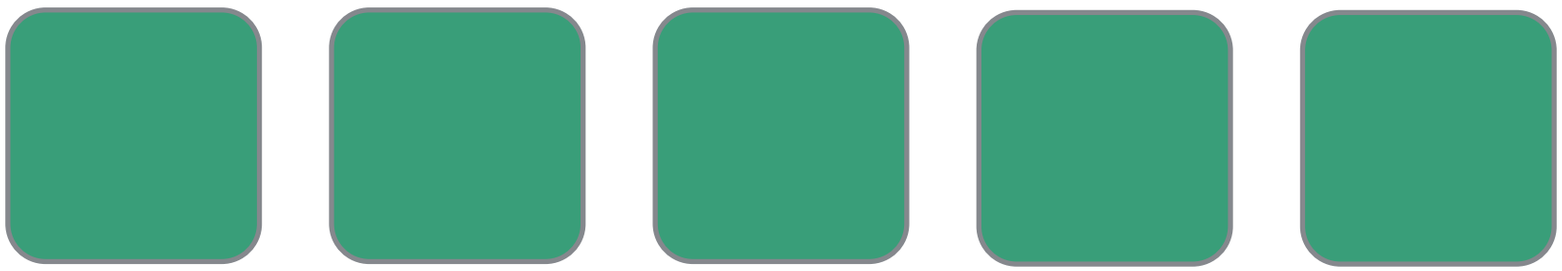
Elasticsearch

an open source, **distributed, scalable,**
highly availability, document-oriented, RESTful
full text search engine
with near real-time search and analytics



Elasticsearch

an open source, **distributed, scalable,**
highly availability, document-oriented, RESTful
full text search engine
with near real-time search and analytics



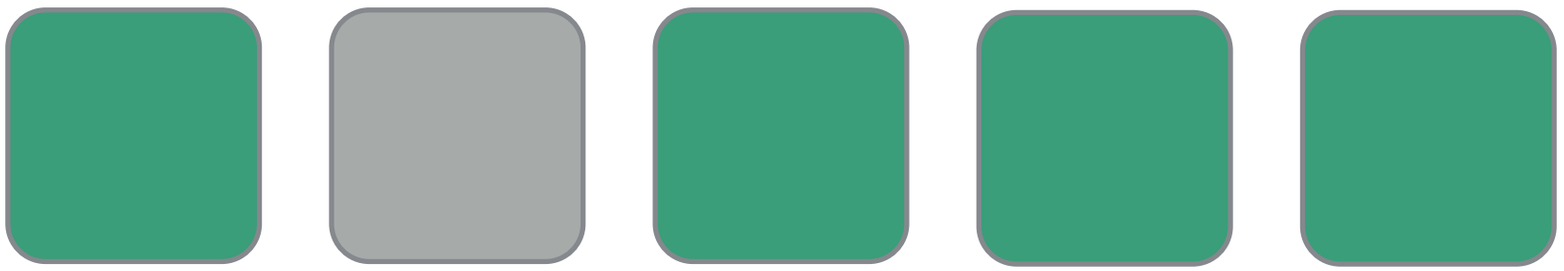
Elasticsearch

an open source, distributed, scalable,

highly availability, document-oriented, RESTful

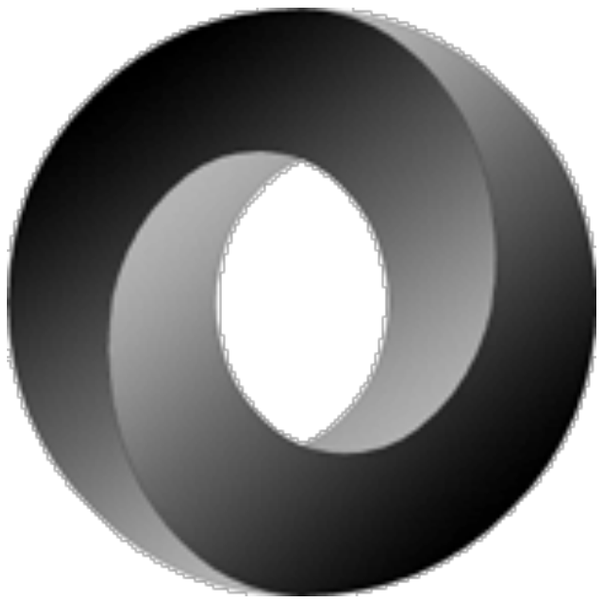
full text search engine

with near real-time search and analytics



Elasticsearch

an open source, distributed, scalable,
highly availability, **document-oriented**, RESTful
full text search engine
with near real-time search and analytics



```
{
    "id" : 1,
    "name" : "somkiat",
    "family_name" : "puisungnoen",
    "website" :  "somkiat.cc"
}
```

Elasticsearch

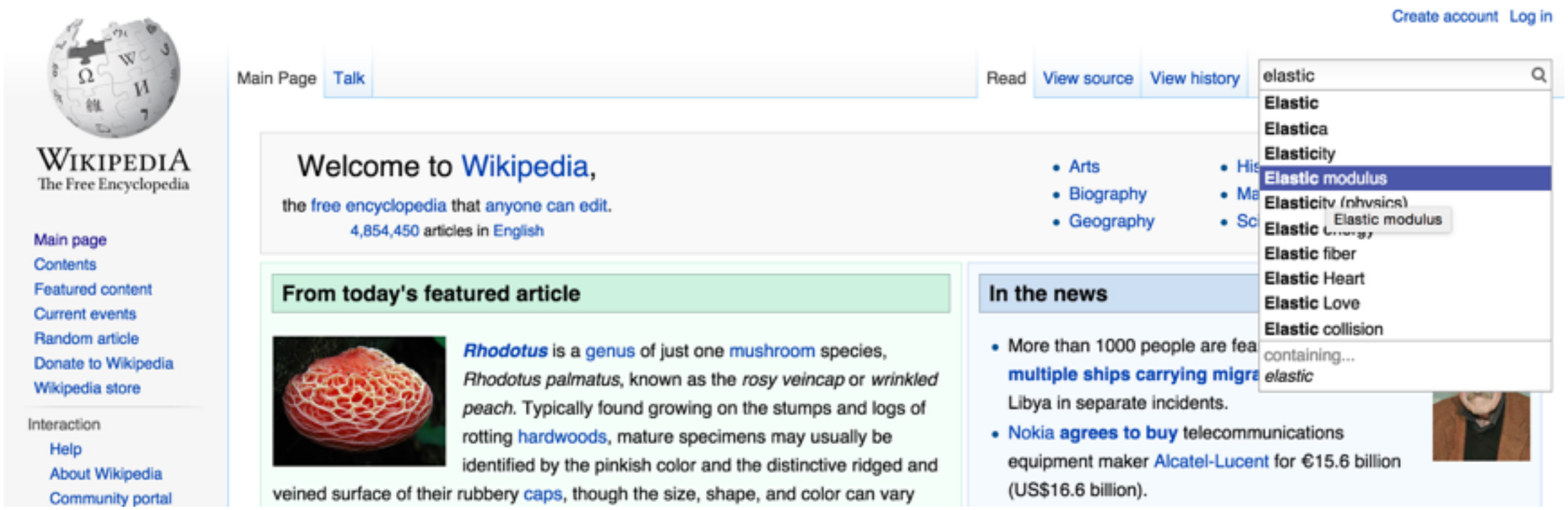
an open source, distributed, scalable,
highly availability, document-oriented, **RESTful**
full text search engine
with near real-time search and analytics

RESTful API

GET PUT POST DELETE

Elasticsearch

an open source, distributed, scalable,
highly availability, document-oriented, RESTful
full text search engine
with near real-time search and analytics



SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Elasticsearch

an open source, distributed, scalable,
highly availability, document-oriented, RESTful
full text search engine
with **near real-time search and analytics**

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Elasticsearch

High-Availability

Plug-ins

Lucene

Distributed

Scalability

RESTFul

JSON

API

open-source

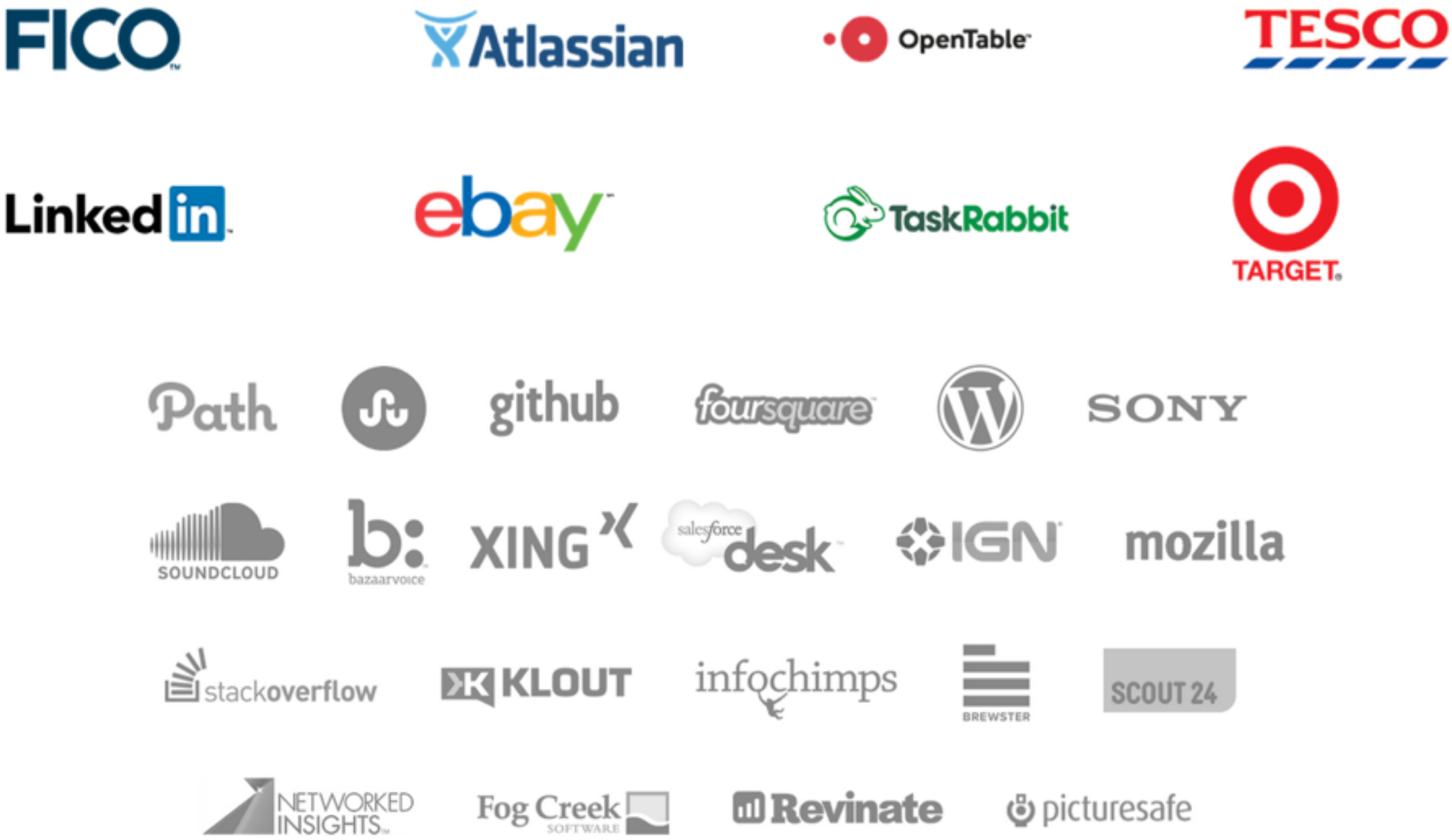
realtime, search and analytics engine

documentation

document store

JAVA

ใครใช้บ้าง ?


FICO
Atlassian
OpenTable
TESCO
LinkedIn
ebay
TaskRabbit
TARGET
Path
github
foursquare
SONY
SOUNDCLOUD
bazaarvoice
XING
salesforce desk
IGN
mozilla
stackoverflow
KLOUT
infochimps
BREWSTER
SCOUT 24
NETWORKED INSIGHTS
Fog Creek SOFTWARE
Revinate
picturesafe

ใครใช้บ้าง ?

- Search repositories, users, issues, pull request
- Search sourcecode 130 พันล้านบรรทัด
- Track alerts, events และ logs

ใครใช้บ้าง ?

- ใช้ Full text seach + geolocation
- ใช้ feature More-like-this ในการหาคำถาม และ คำตอบ

Clients

- Java
- PHP
- Ruby
- Python
- JavaScript
- NodeJS
- Go
- Scala
- .Net
- Clojure
- Erlang
- R

http://www.elasticsearch.org/guide/en/elasticsearch/client/community/current/clients.html

สวัสดี Elasticsearch

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

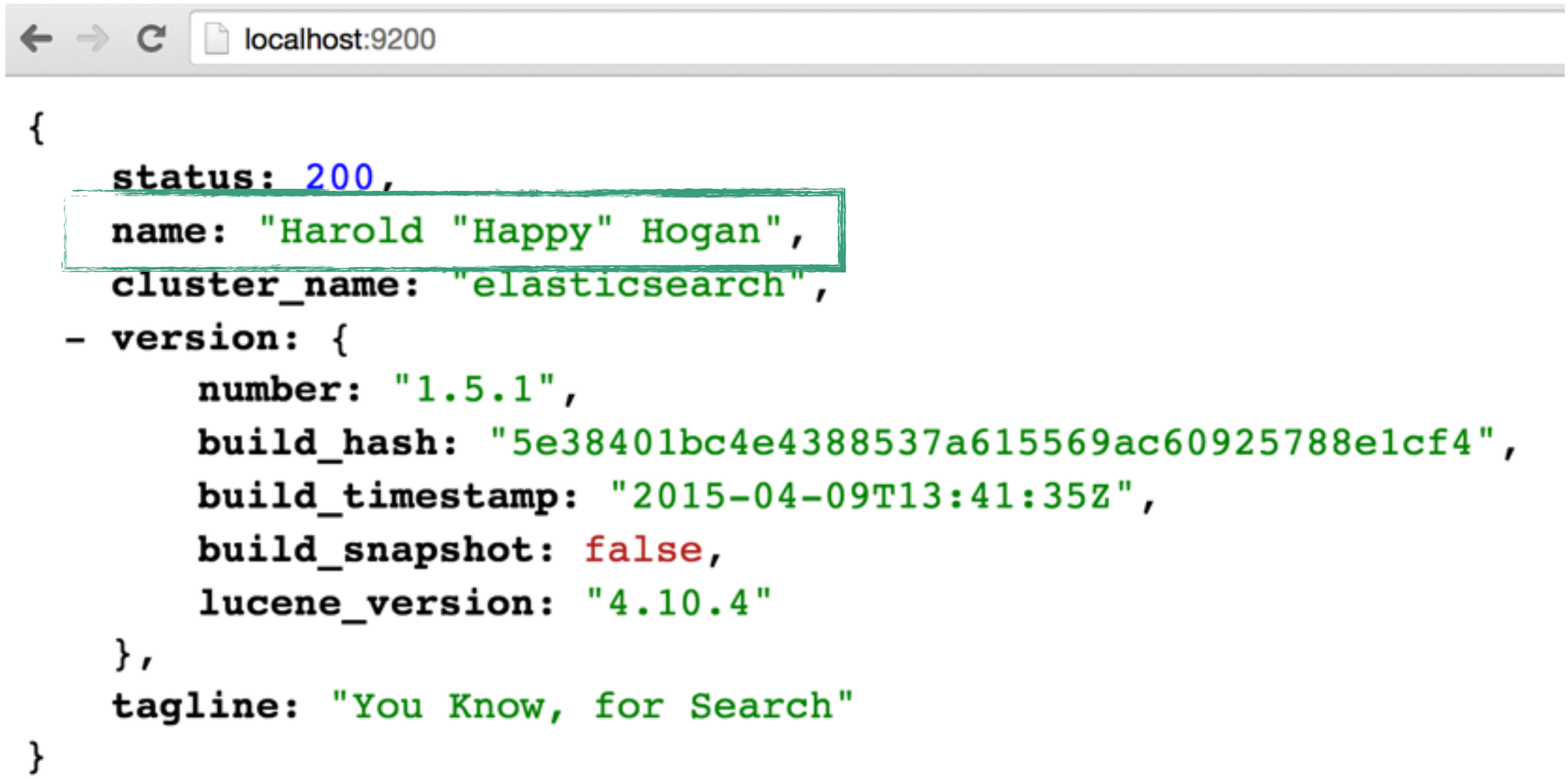
สวัสดี Elasticsearch

```
$ wget  https://download.elastic.co/elasticsearch/
elasticsearch/elasticsearch-1.5.1.zip

$ unzip  elasticsearch-1.5.1.zip
$ cd elasticsearch-1.5.1

$ bin/elasticsearch
```

สวัสดี Elasticsearch



สวัสดี Elasticsearch

SPRINT3R

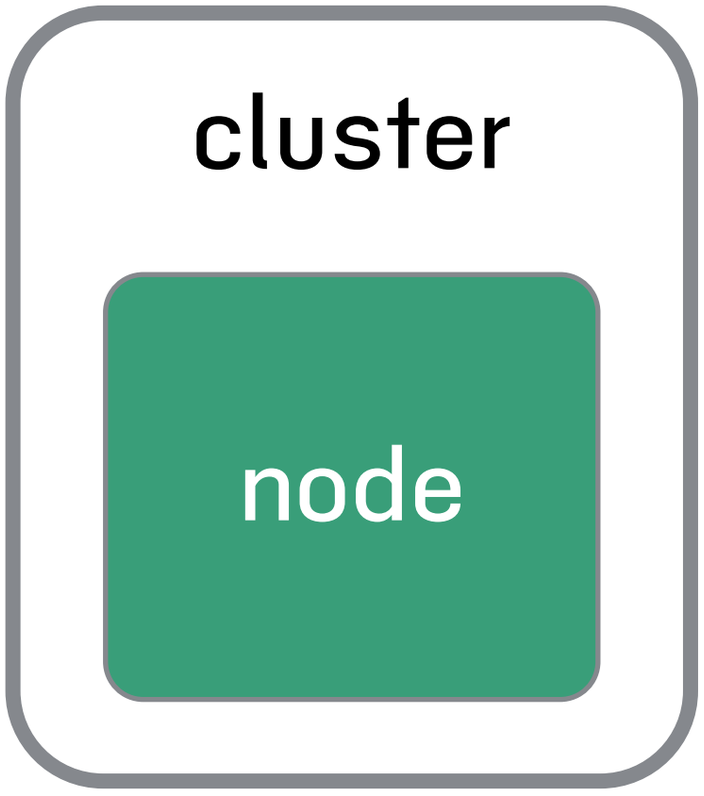
Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Scaling

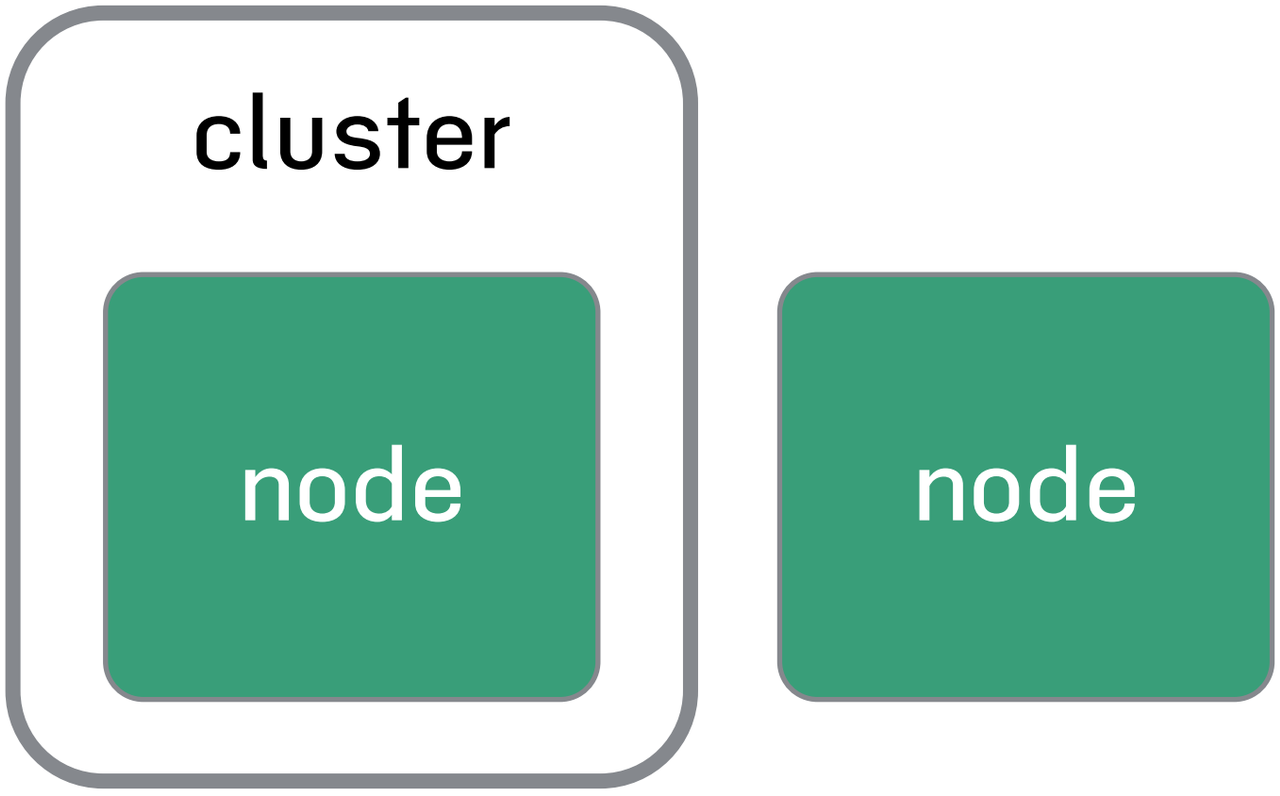
SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

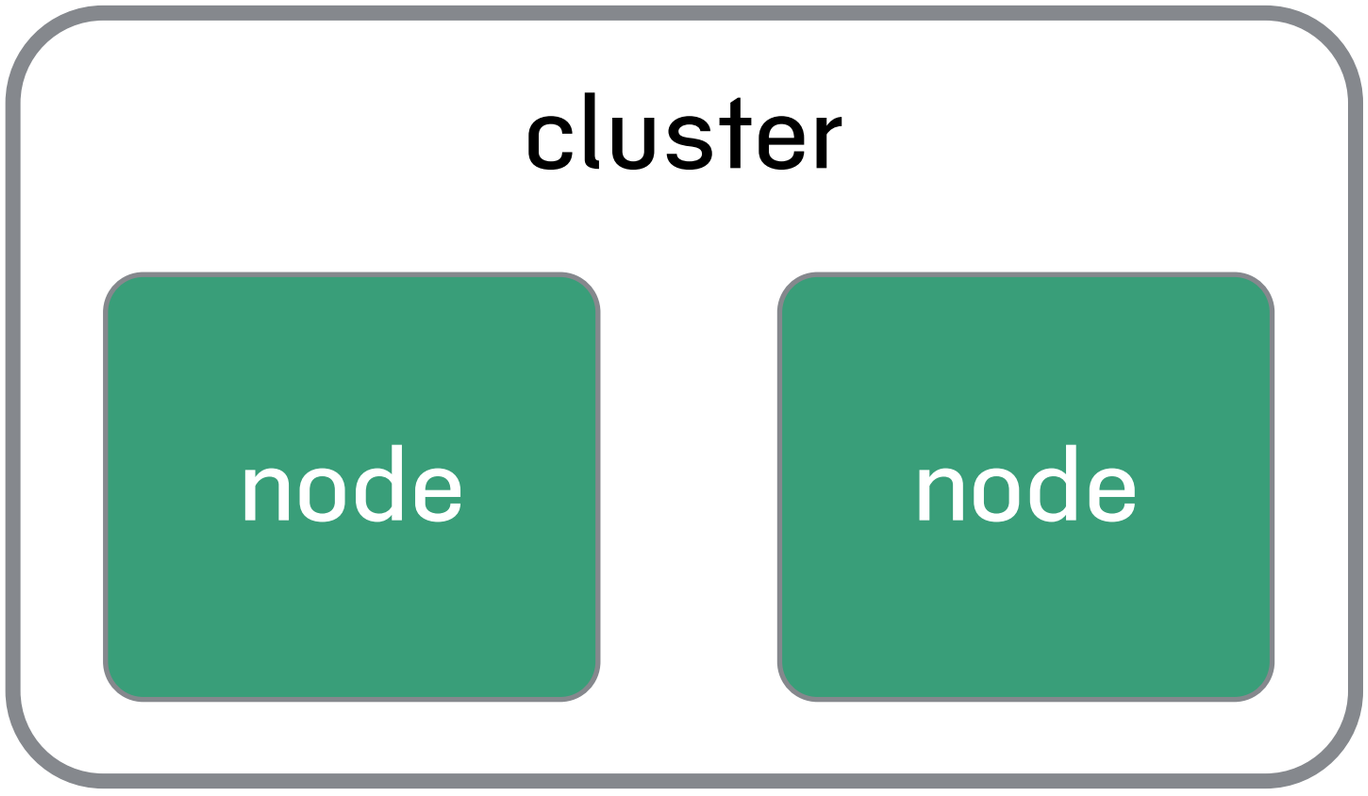
Cluster :: กลุ่มของ node



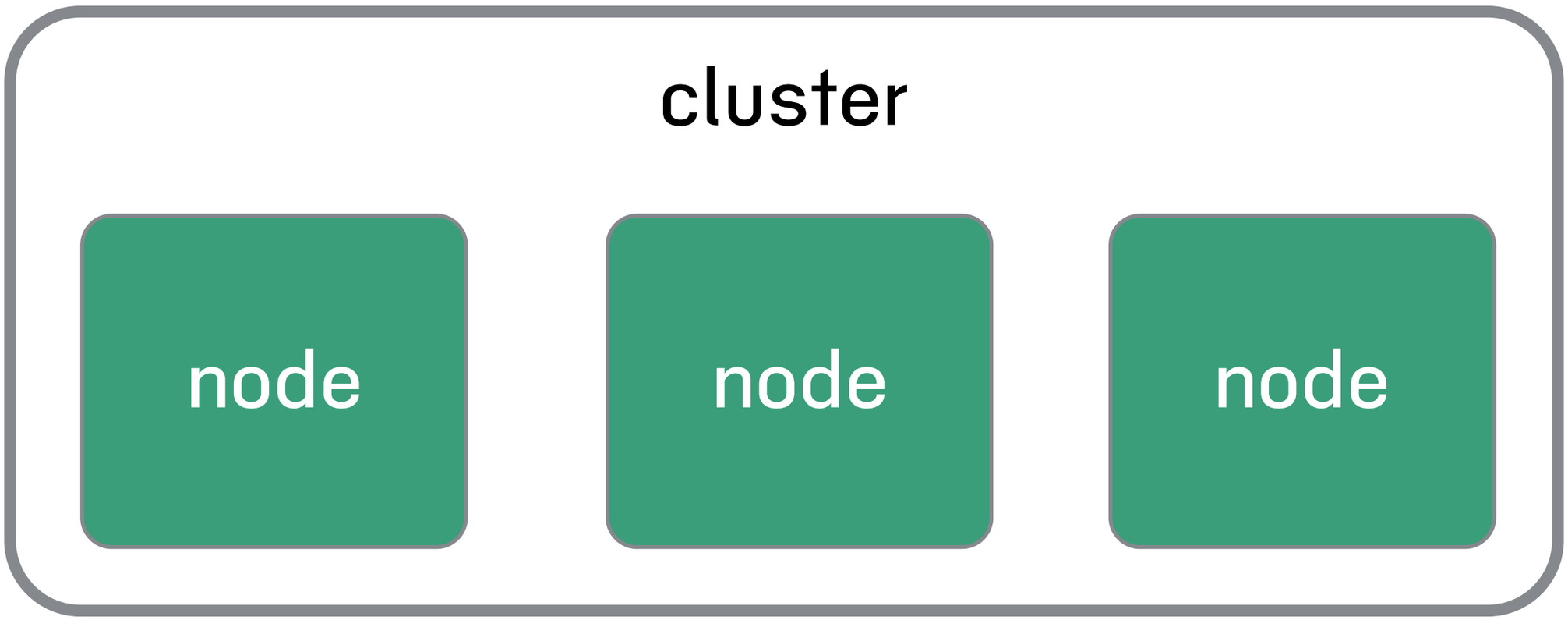
Cluster :: กลุ่มของ node



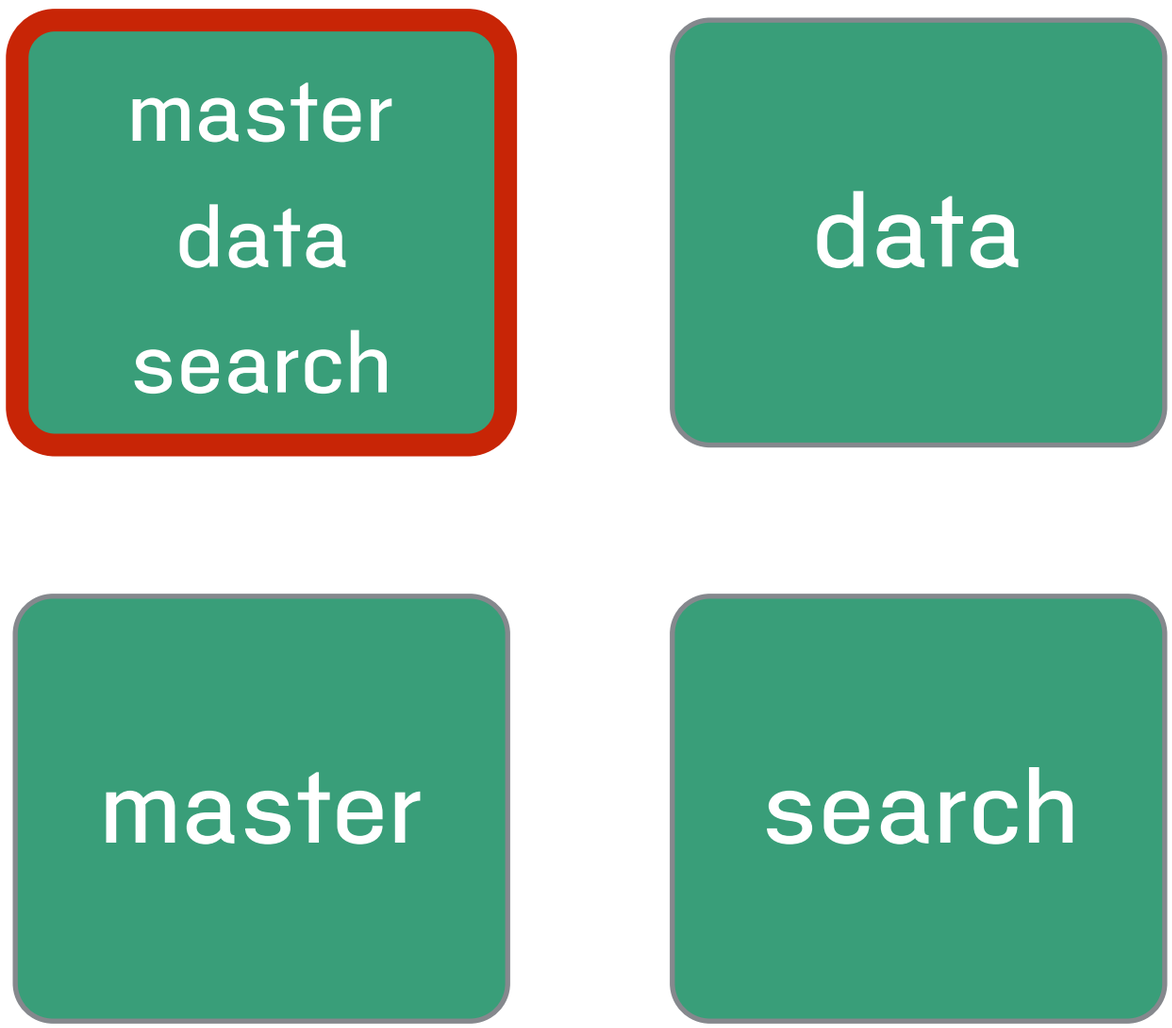
Cluster :: กลุ่มของ node



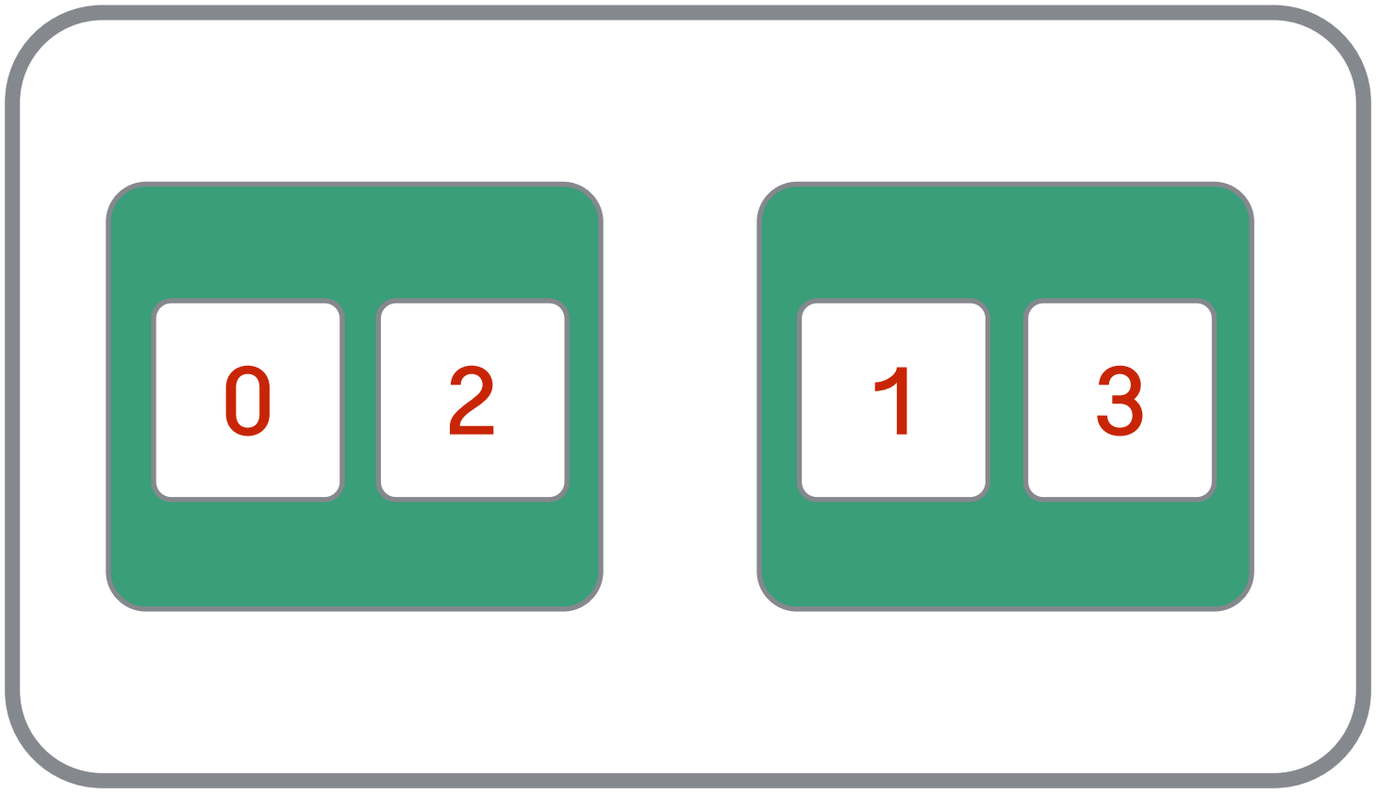
Cluster :: กลุ่มของ node



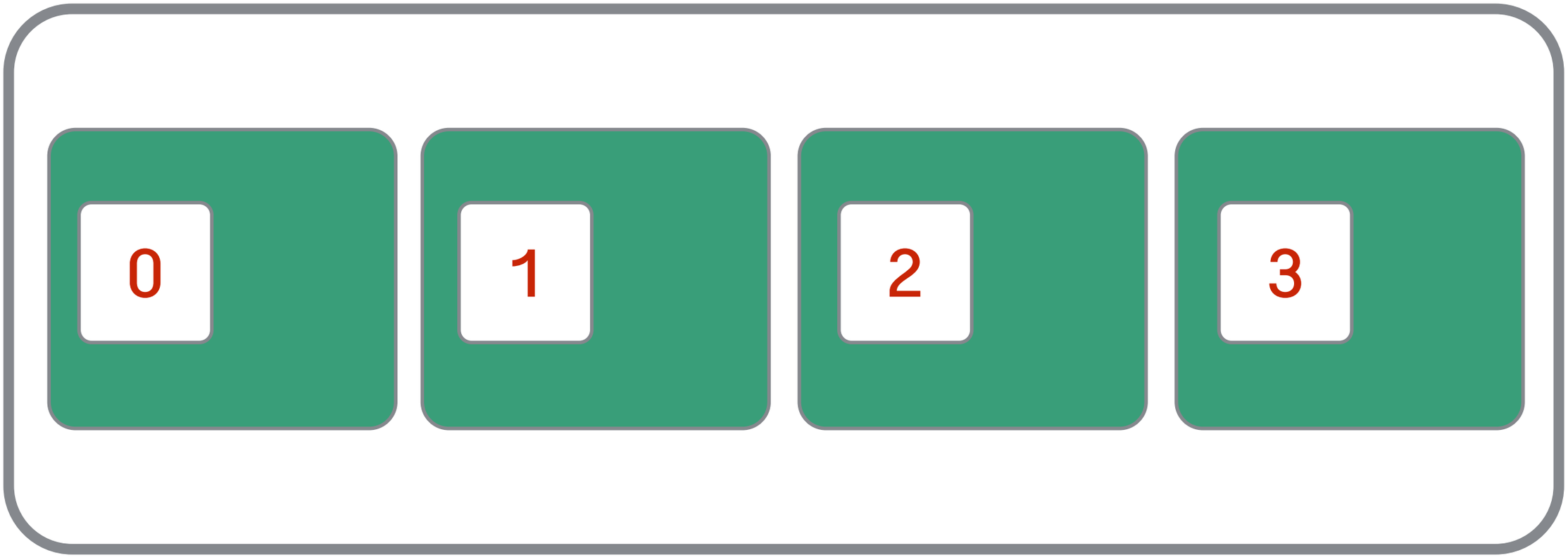
Node :: มี 4 ชนิด



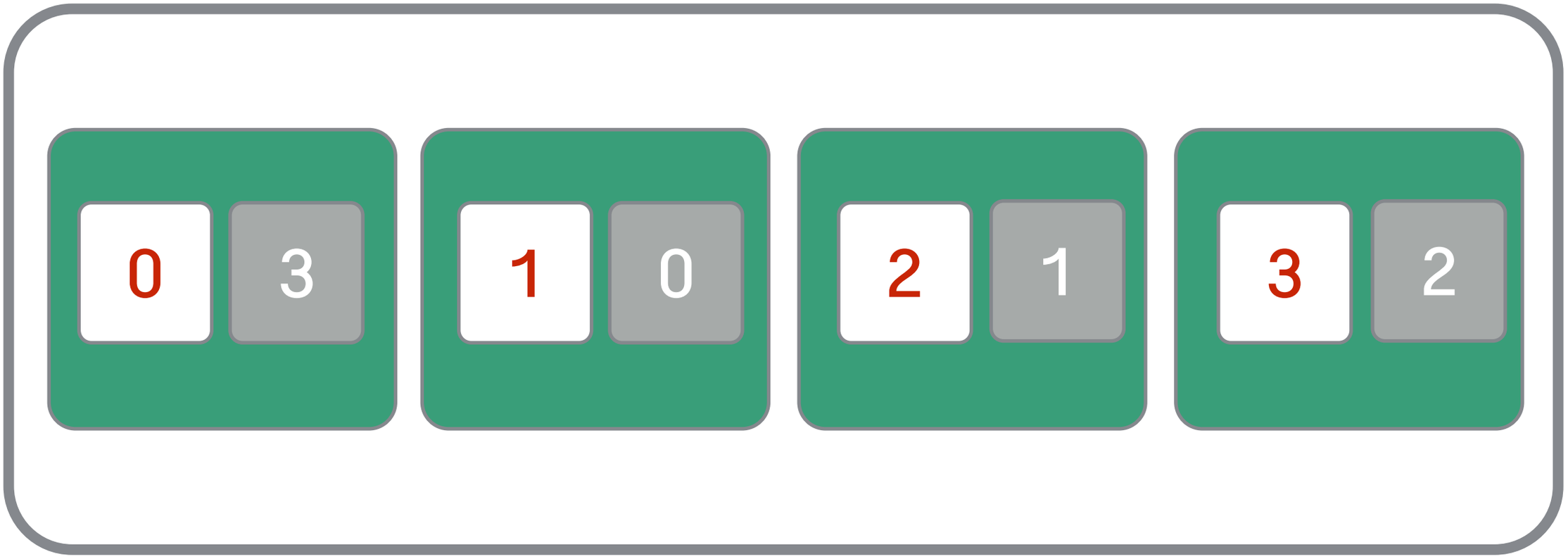
Shard :: unit of scale



Shard :: unit of scale



Replication



Search

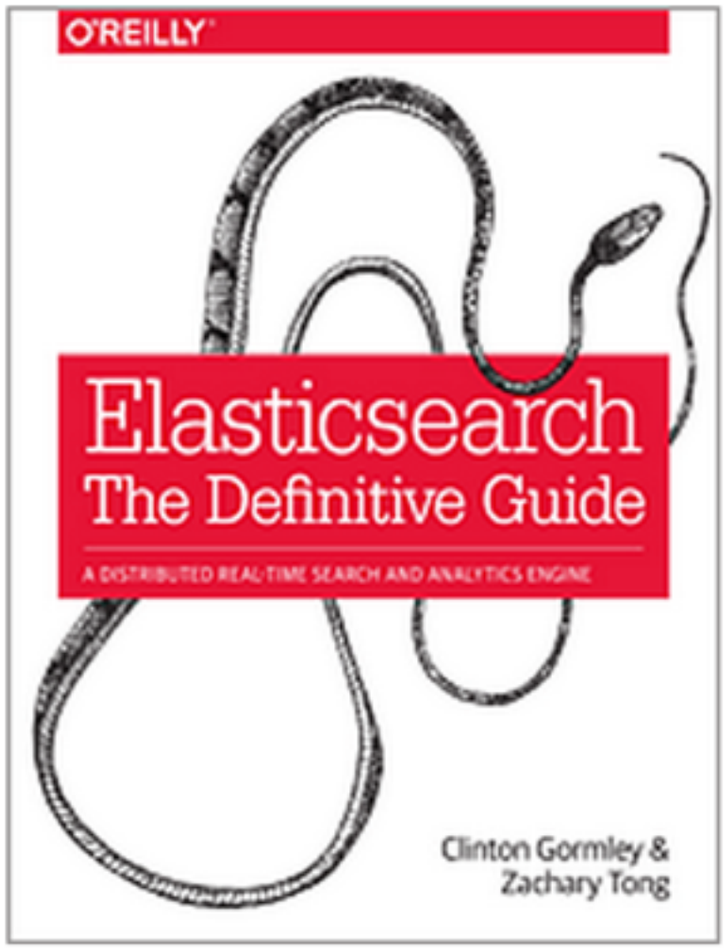
SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

RDBMS vs Elasticsearch

DATABASE	TABLE	ROW	COLUMN
INDEX	TYPE	DOCUMENT	FIELD

CRUD



Larger Cover

Elasticsearch: The Definitive Guide

A distributed real-time search and analytics engine

By Clinton Gormley, Zachary Tong

Publisher: O'Reilly Media
Final Release Date: January 2015
Pages: 724

4.4

Read **10** Reviews | Write a Review

Whether you need full-text search or real-time analytics of structured data—o
Elasticsearch distributed search engine is an ideal way to put your data to wo
guide not only shows you how to search, analyze, and explore data with Elast
also helps you deal with the complexities of human...

Full description

CRUD :: Create

PUT <index name>/<type>/<id>

```
{
        ... your data ..
}
```

CRUD :: Create

```
PUT shop/book/1
{
   "name" : "Elasticsearch: The Definitive Guide",
   "authors" : [ "Clinton Gormley", "Zachary Tong" ],
   "pages": 724,
   "published_date": "2015/01/31"
}
```

CRUD :: Read

```
GET shop/book/1
```

```
GET shop/book/_search
```

```
GET shop/book/_search?q=elasticsearch
```

CRUD :: Delete

```
DELETE shop/book/1
```

Searching with DSL

GET shop/book/_search

Searching with DSL

- **Full text query**
- **Term level query**
- **Compound query**
- Geo query
- Joining query

Full text query

```
GET shop/book/_search
{
  "query": {
    "match": {
      "title": "elasticsearch book"
    }
  }
}
```

Search response in JSON


```
"hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
        {
            "_index": "shop",
            "_type": "book",
            "_id": "1",
            "_score": 0.15342641,
            "_source": {
                "name": "Elasticsearch: The Definitive Guide",
                "authors": [
                    "Clinton Gormley",
                    "Zachary Tong"
                ],
                "pages": 724,
                "published_date": "2015/01/31"
            }
        }
    ]
}
```

Search from all fields

```
GET shop/book/_search
{
  "query": {
    "match": {
      "_all": "elasticsearch book"
    }
  }
}
```

Reduce search response in JSON


```
"hits": {
    "total": 1,
    "max_score": 0.15342641,
    "hits": [
        {
            "_index": "shop",
            "_type": "book",
            "_id": "1",
            "_score": 0.15342641,
            "_source": {
                "pages": 724,
                "name": "Elasticsearch: The Definitive Guide"
            }
        }
    ]
}
```

Reduce search response in JSON

```
GET shop/book/_search
{
    "_source": ["name", "pages"],
    ...
}
```

Full text query

```
GET shop/book/_search
{
  "query": {
    "match": {
      "title": {
        "query": "elasticsearch book",
        "operator": "or"
      }
    }
  }
}
```

Full text query

```
GET shop/book/_search
{
  "query": {
    "query_string": {
     "default_field": "title",
     "query": "elasticsearch OR book"
    }
  }
}
```

Term level query

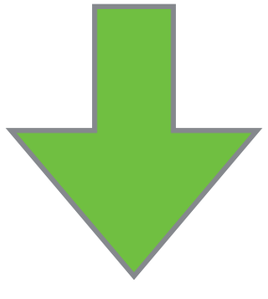
```
GET shop/book/_search
{
  "query": {
    "term": {
      "title": {
        "value": "elasticsearch"
      }
    }
  }
}
```

Term level query

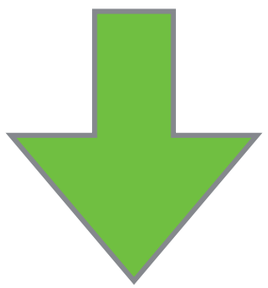
```
GET shop/book/_search
{
  "query": {
    "term": {
      "title": {
        "value": "elasticsearch book"
      }
    }
  }
}
```

Term vs Match

{ “match” : { “title” : “elasticsearch” } }



title:elasticsearch



{ “term” : { “title” : “elasticsearch” } }

Term vs Match

{ “match” : { “title” : “elasticsearch book” } }

title:elasticsearch OR title:book

```
{ “bool” : {
    “should” : [
        { “term” :  { “title” : “elasticsearch” } },
        { “term” :  { “title” : “book” } }
    ]
}}
```

Compound query

```
GET shop/book/_search
{
  "query": {
    "bool": {
      "must": [],
      "should": [],
      “must_not": [],
}}}
```

Compound query

```
GET shop/book/_search
{
  "query": {
    "bool": {
      "should": [{
    "term": {
      "title": {
        "value": "elasticsearch"
      }
    }
  }
]}}}
```

Compound query with long tail

```
GET shop/book/_search
{
  "query": {
    "bool": {
      "should": [
        { "term": { "title" : "elasticsearch" } },
        { "term": { "title" : "book" } },
        { "term": { "title" : "sales" } }
      ],
      "minimum_should_match" : 2
    }
  }
}
```

Compound query with filter

```
GET shop/book/_search
{
  "query": {
    "bool": {
      "filter": {
          "term": { "title" : "book" }
      },
      "should": [
        { "term": { "title" : "elasticsearch" } },
        { "term": { "title" : "book" } }
      ],
      "minimum_should_match" : "1"
    }
  }
}
```

Query vs Filter

Query vs Filter

bool filter -> True | False

bool query -> _score

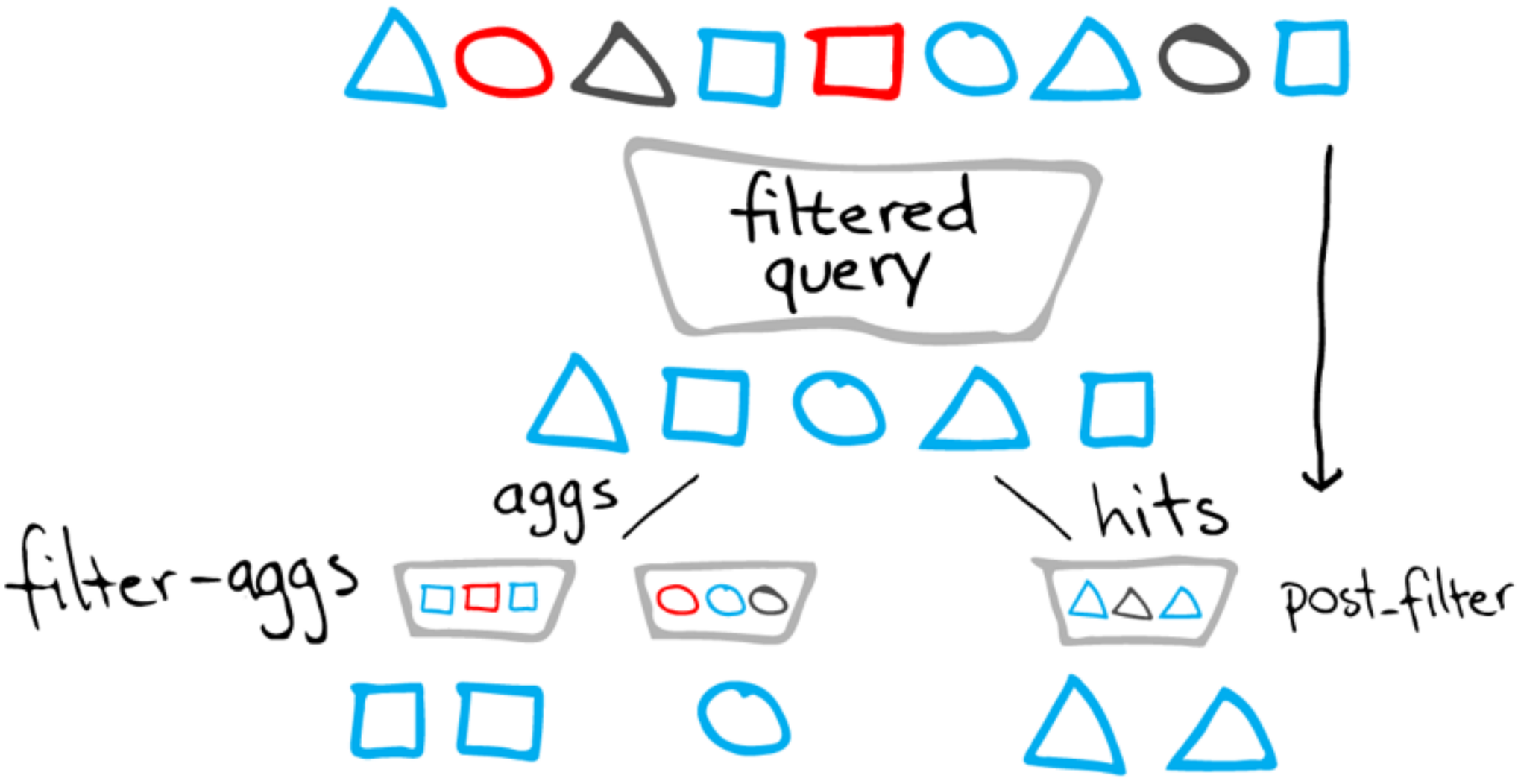
Aggregations

analytic your data
explore your data

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

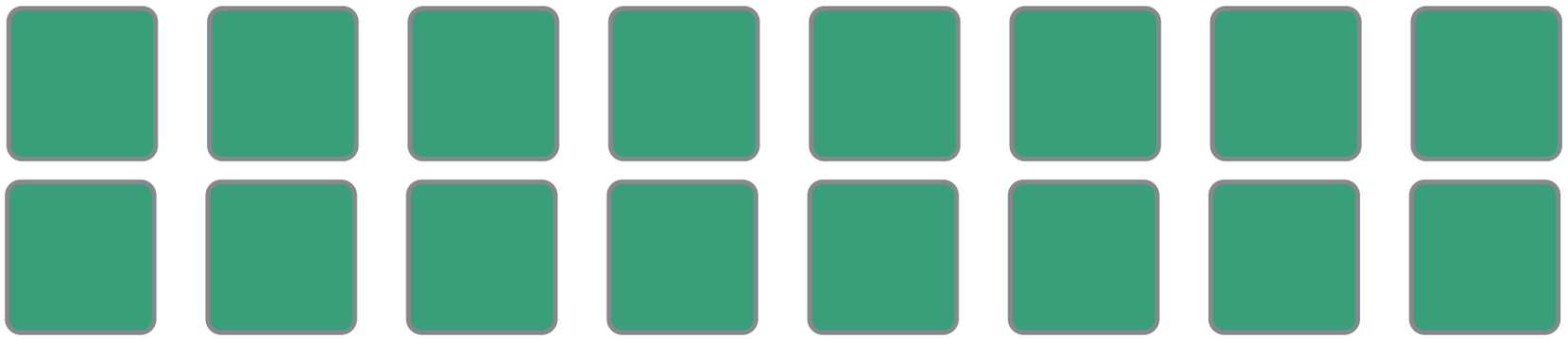
High level



Buckets and **Metrics**

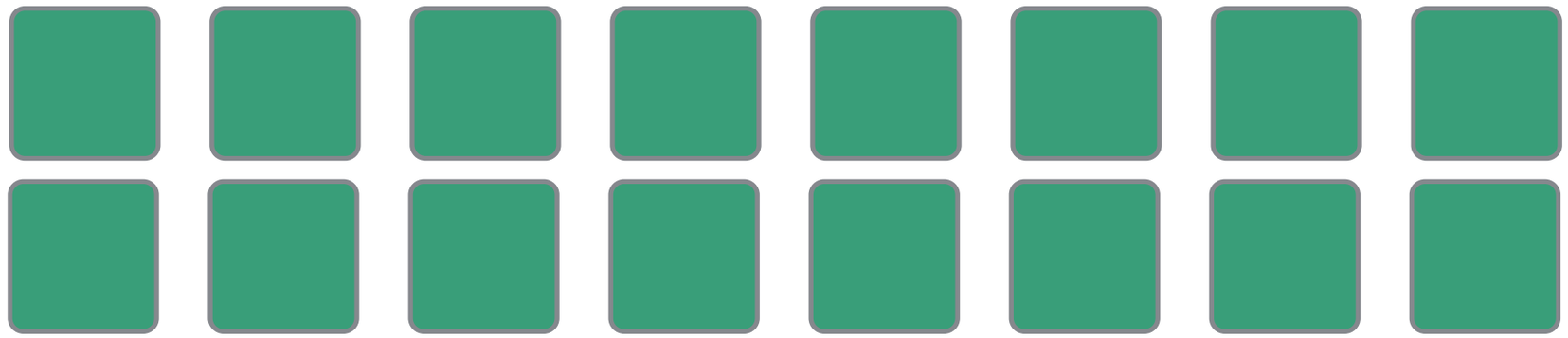
Aggregations

Documents

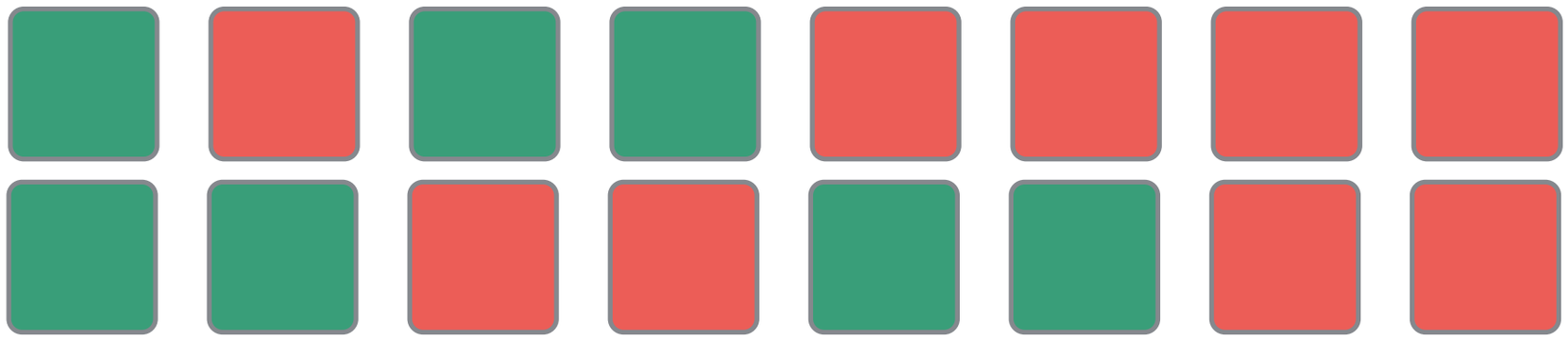


Aggregations

Documents



Query/Filter

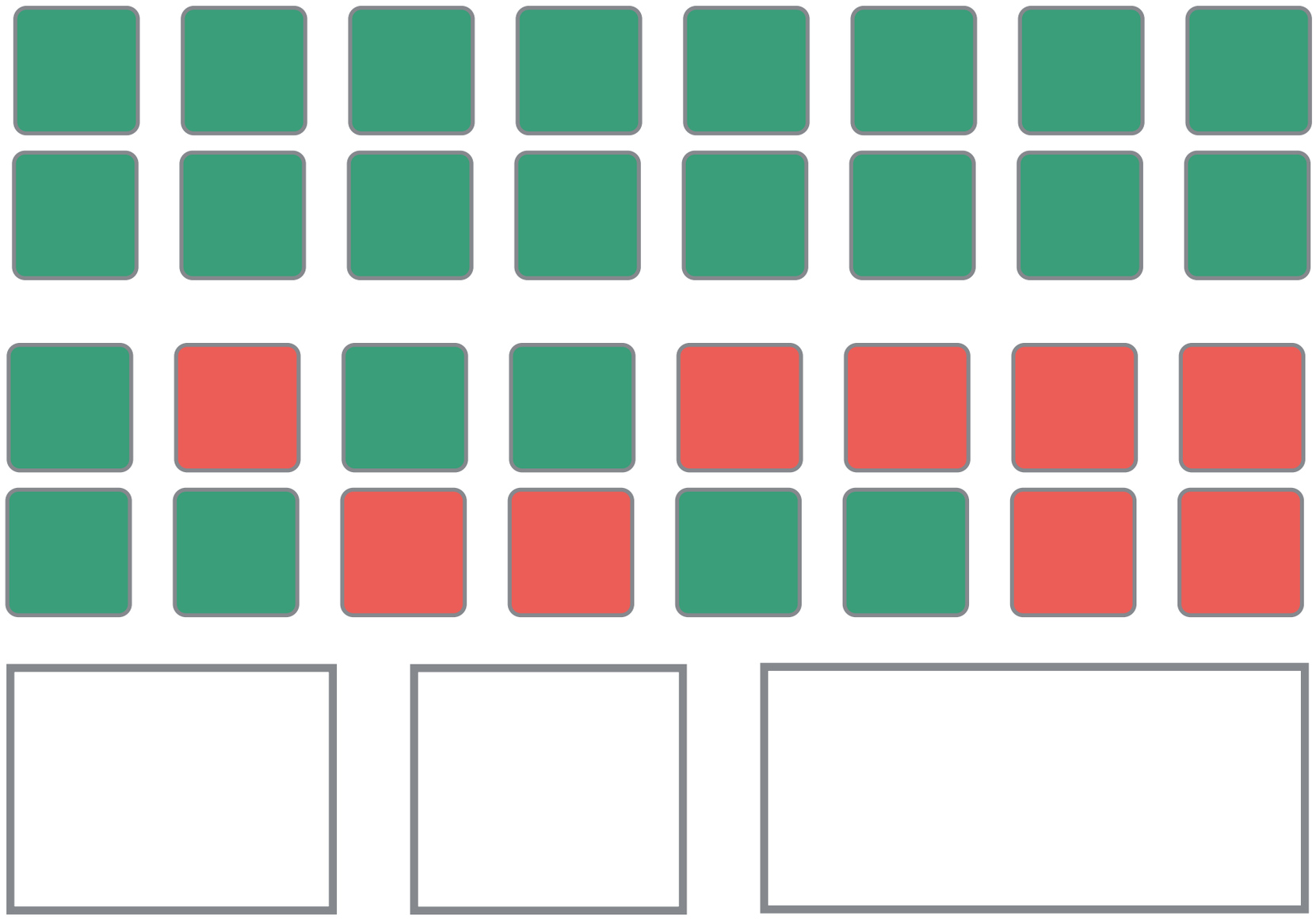


Aggregations

Documents

Query/Filter

Buckets



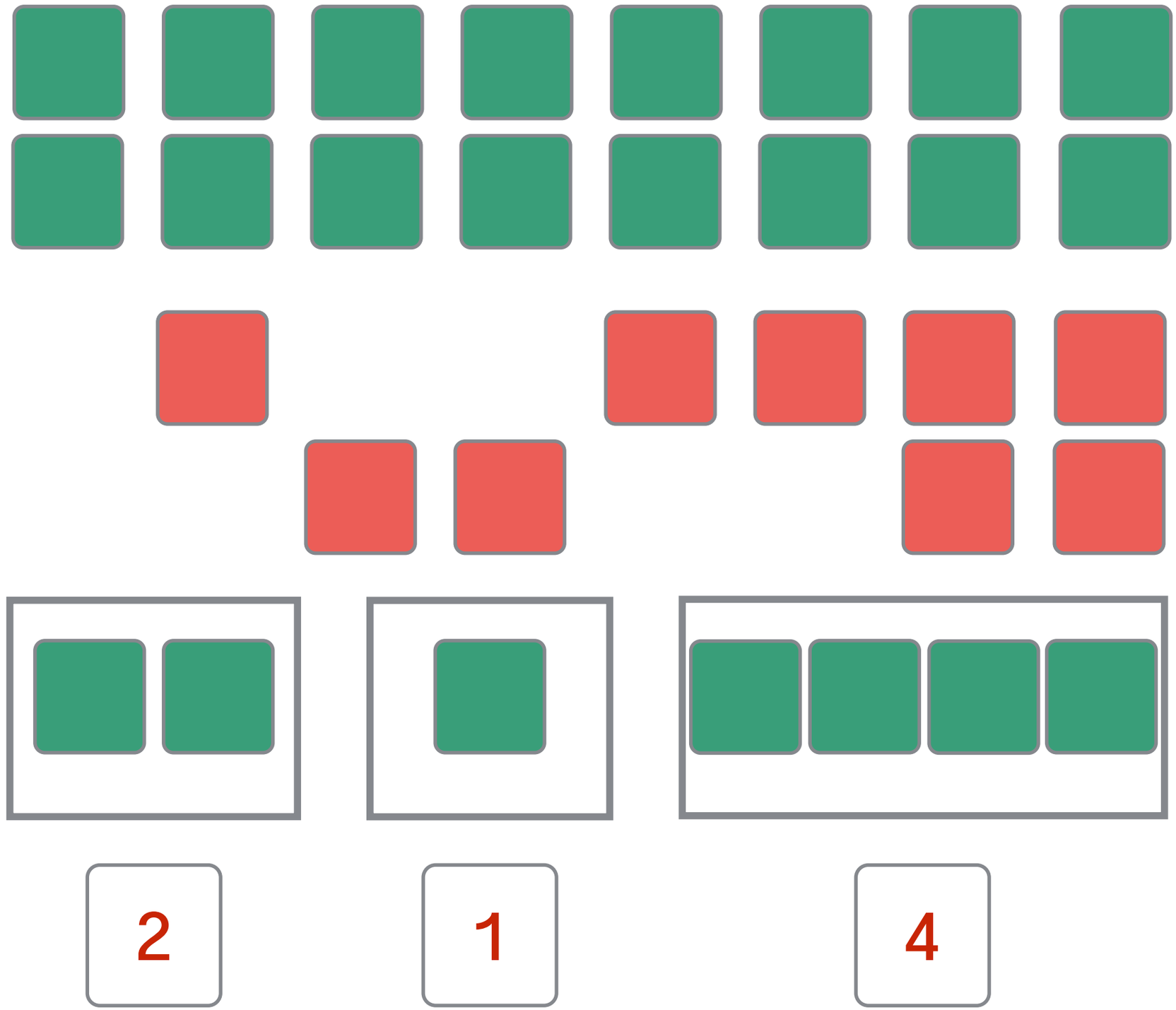
Aggregations

Documents

Query/Filter

Buckets

Metrics



Structure of Aggregation

ใช้ชื่อย่อว่า aggs

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

Aggregations Types

min

max

avg

sum

terms

significant terms

filter

filters

missing

range

date range

histogram

date histogram

geo

Aggregations

ผู้แต่งแต่ละคนมีหนังสือกี่เล่ม ?

Aggregations

```
GET /store/book/_search
{
  "aggs": {
    "all_book_title": {
      "terms": {
        "field": “author_name"
      }
    }
  }
}
```

aggregation name

aggregation type

aggregation body

Result


```
"aggregations": {
    "all_book_title": {
        "doc_count_error_upper_bound": 0,
        "sum_other_doc_count": 0,
        "buckets": [
            {
                "key": "computer",
                "doc_count": 5
            },
            {
                "key": "technology",
                "doc_count": 5
            },
            {
                "key": "online",
                "doc_count": 3
            },
            {
                "key": "searching",
                "doc_count": 3
            },
            {
                "key": "java",
                "doc_count": 2
            },
            {
                "key": "programming",
                "doc_count": 2
            }
        ]
    }
}
```


ยาวมากๆ ลดได้ไหม ?

Aggregations

```
GET shop/book/_search
{
  "size": 0,
 "aggs": {
  "book_by_author": {
   "terms": {
    "field": "author_name"
   }
  }}}
```

Results :: ปัญหาคืออะไร ?


```
"aggregations": {
    "book_by_author": {
        "doc_count_error_upper_bound": 0,
        "sum_other_doc_count": 0,
        "buckets": [
            {
                "key": "clinton",
                "doc_count": 1
            },
            {
                "key": "gormley",
                "doc_count": 1
            },
            {
                "key": "tong",
                "doc_count": 1
            },
            {
                "key": "zachary",
                "doc_count": 1
            }
        ]
    }
}
```

Results :: ปัญหาคืออะไร ?


```
"aggregations": {
    "book_by_author": {
        "doc_count_error_upper_bound": 0,
        "sum_other_doc_count": 0,
        "buckets": [
            {
                "key": "clinton",
                "doc_count": 1
            },
            {
                "key": "gormley",
                "doc_count": 1
            },
            {
                "key": "tong",
                "doc_count": 1
            },
            {
                "key": "zachary",
                "doc_count": 1
            }
        ]
    }
}
```

ปัญหาจาก Inverted index

Term	Document ID
clinton	1
gormley	1
tong	1
zachary	1

สิ่งที่ต้องการจาก Inverted index

Term	Document ID
Clinton Gormley	1
Zachary Tong	1

กำหนด mapping

```
PUT shop
{
    "mappings": {
        "book": {
            "properties": {
                "author_name": {
                    "index": "not_analyzed",
                    "type": "string"
                }
            }
        }
    }
}
```

Results


```
"aggregations": {
    "book_by_author": {
        "doc_count_error_upper_bound": 0,
        "sum_other_doc_count": 0,
        "buckets": [
            {
                "key": "Clinton Gormley",
                "doc_count": 1
            },
            {
                "key": "Zachary Tong",
                "doc_count": 1
            }
        ]
    }
}
```

Workshop with aggregations

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Operating System & Hardware

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Elasticsearch use

- **CPU**

  Indexing, searching

- **Memory**

  Indexing, searching, merging

- **I/O**

  Aggregation, indices

- **Network**

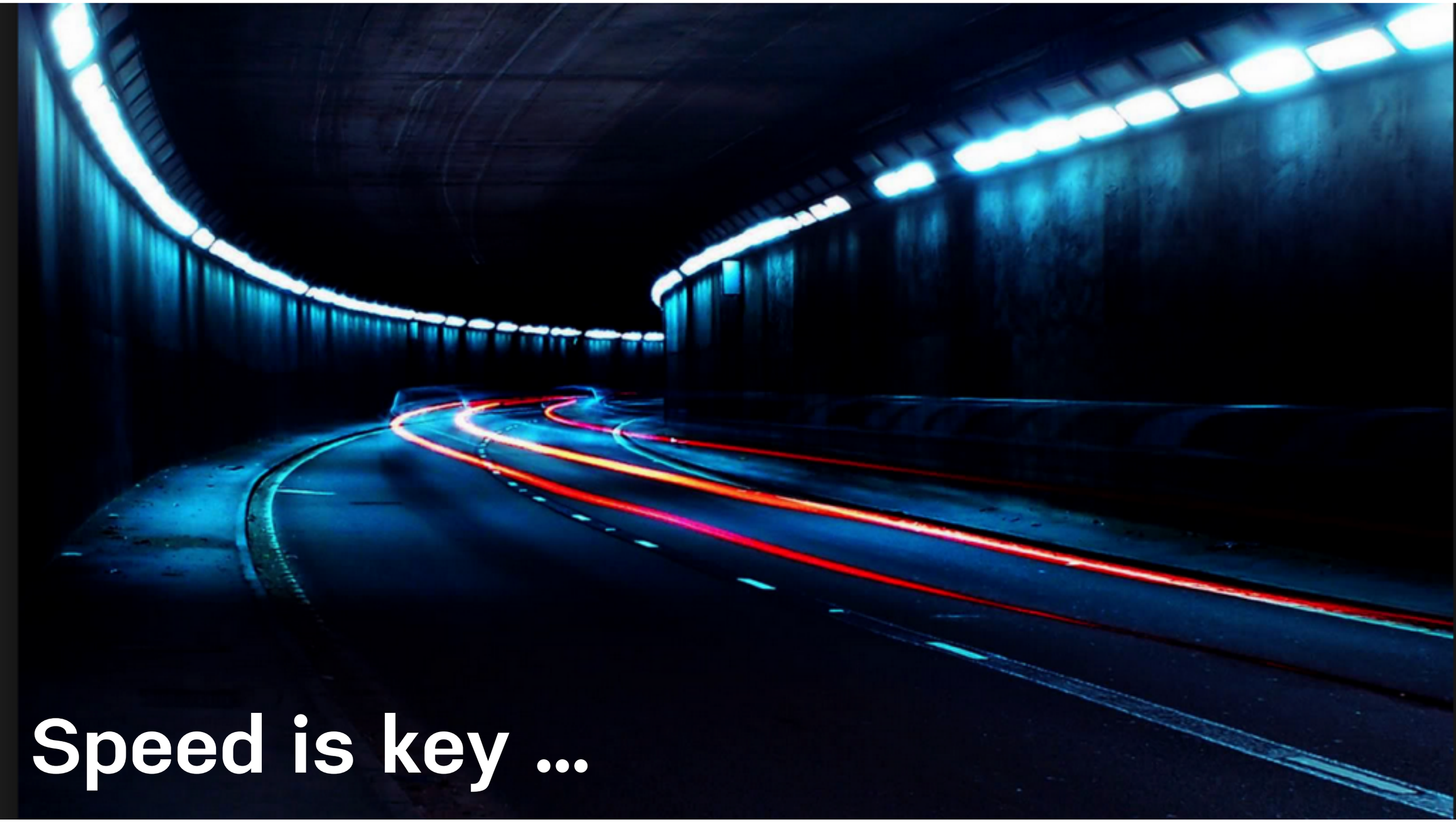
  Relocation, snapshot & restore

Summary

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Summary



SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Summary

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Summary

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance

Thank you and Questions

SPRINT3R

Siam Chamnankit Co., Ltd., Odd-e (Thailand) Co., Ltd. and Alliance