Elasticsearch,

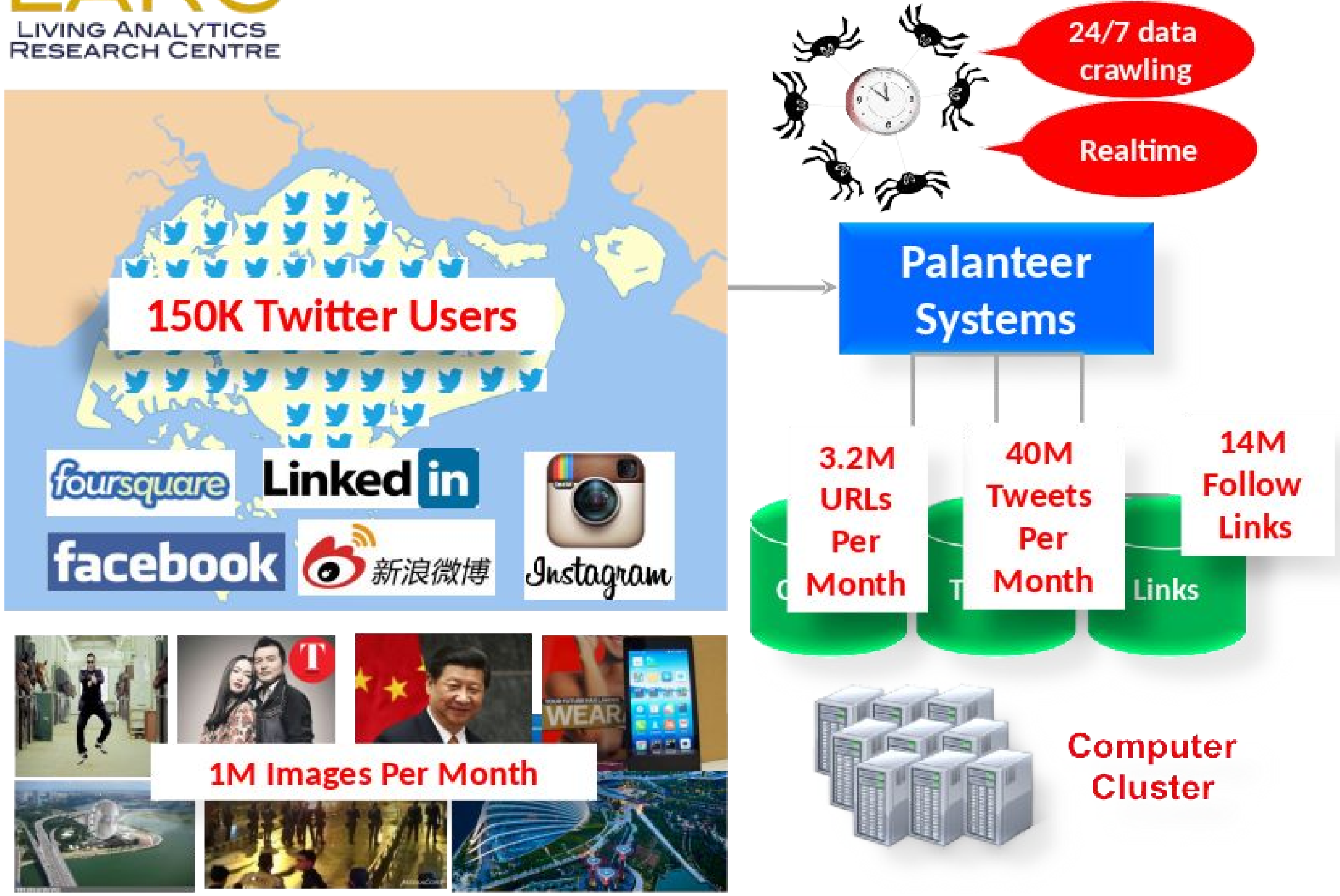
You Know For Search! And More!

Medcl, 曾勇 (Zeng Yong)

Philips Kokoh Prasetyo
Casey Vu
Arinto Murdopo

LARC
Living Analytics Research Centre

LARC
LIVING ANALYTICS RESEARCH CENTRE



Outline

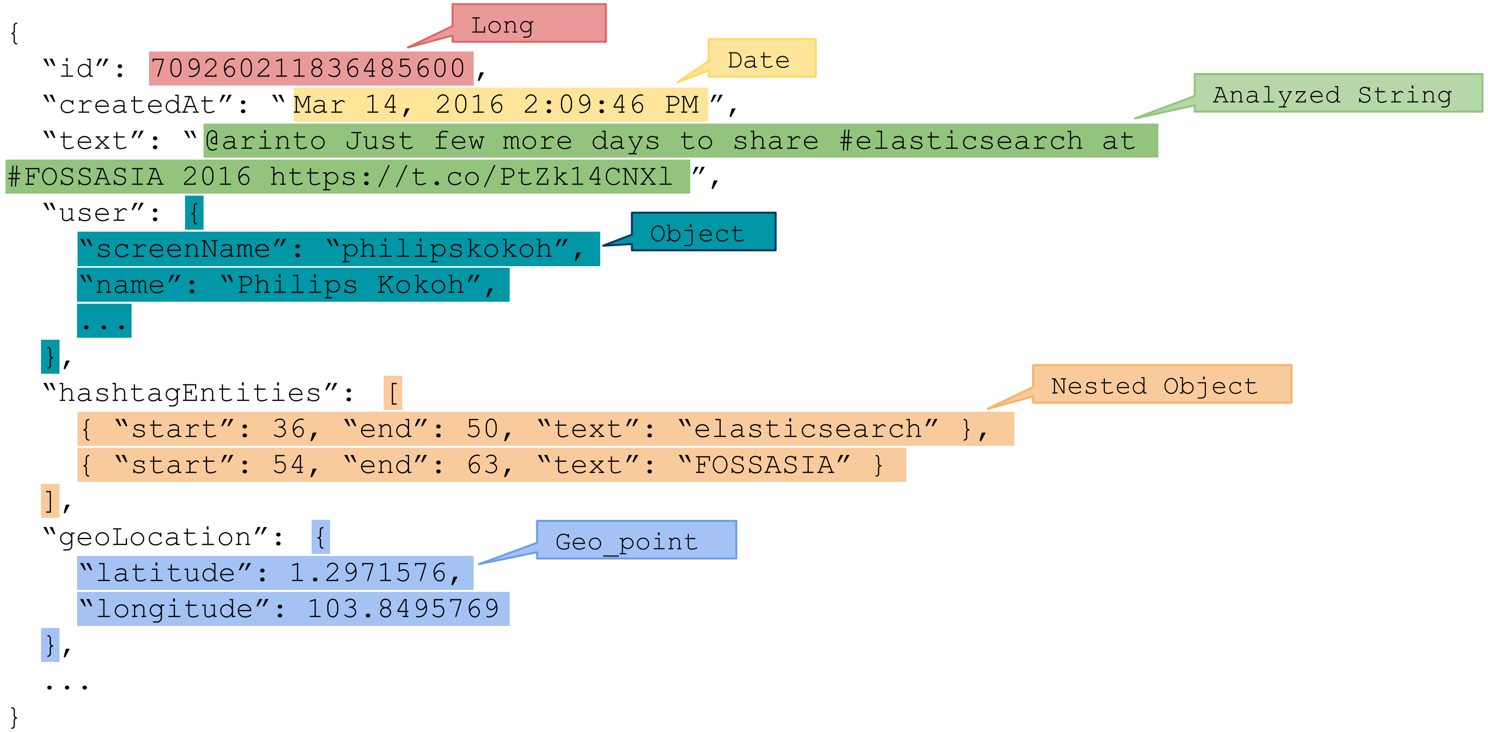
- Real time Search (Philips)
- Aggregation & Analytics (Casey)
- Lesson Learned @LARC (Arinto)

Real-time Search

Real-time Search

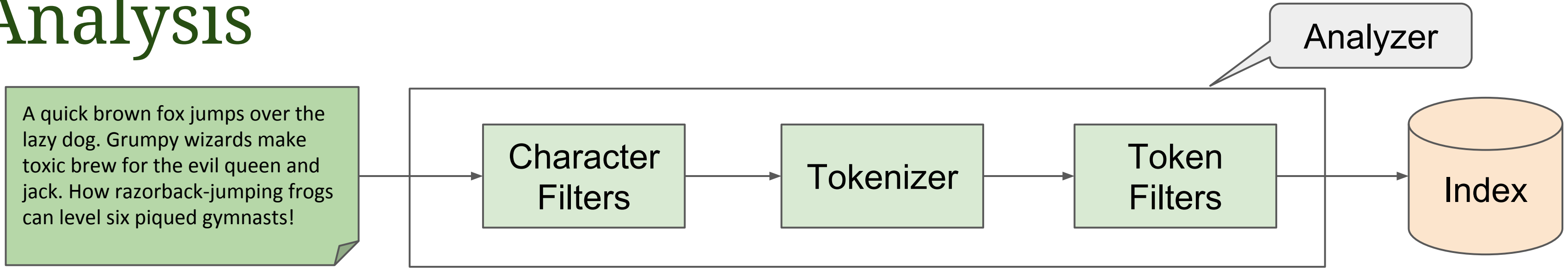
- Indexing
  - Mapping
    - How a document, and the **fields** it contains are *stored* and *indexed*
      - Simple type: string, date, long, double, boolean, ip
      - Hierarchical: object, nested object
      - Specialized type: geo_point, geo_shape, completion
  - Analysis
    - How **full text** is *processed* to make it *searchable*
      - Standard Analyzer, Simple Analyzer, Language Analyzer, Snowball Analyzer
- Searching
  - Query DSL
    - Flexible and powerful query language used by Elasticsearch

Mapping



```
{
  “id”: 709260211836485600,
  “createdAt”: “Mar 14, 2016 2:09:46 PM”,
  “text”: “@arinto Just few more days to share #elasticsearch at
#FOSSASIA 2016 https://t.co/PtZk14CNXl”,
  “user”: {
    “screenName”: “philipskokoh”,
    “name”: “Philips Kokoh”,
    ...
  },
  “hashtagEntities”: [
    { “start”: 36, “end”: 50, “text”: “elasticsearch” },
    { “start”: 54, “end”: 63, “text”: “FOSSASIA” }
  ],
  “geoLocation”: {
    “latitude”: 1.2971576,
    “longitude”: 103.8495769
  },
  ...
}
```

Analysis


A quick brown fox jumps over the lazy dog. Grumpy wizards make toxic brew for the evil queen and jack. How razorback-jumping frogs can level six piqued gymnasts!
Character Filters
Tokenizer
Token Filters
Analyzer
Index

Built-in Analyzer

“Set the shape to semi-transparent by calling set_trans(5)”

- Standard Analyzer
  - [set, the, shape, to, semi, transparent, by, calling, set_trans, 5]
- Simple Analyzer
  - [set, the, shape, to, semi, transparent, by, calling, set, trans]
- Whitespace Analyzer
  - [Set, the, shape, to, semi-transparent, by, calling, set_trans(5)]
- Language Analyzer, e.g. english, french, spanish, arabic, …
  - English analyzer: [set, shape, semi, transpar, call, set_tran, 5]

Searching

Query DSL

- Based on JSON to define queries :)
- Behavior:
    - Query Context
        - Answering: “How well this document match this query clause?”
        - Return: **_score** → **relevance score**
    - Filter Context
        - Answering: “Does this document match this query clause?”
        - Return: boolean **yes** or **no**

Let’s Search!!

Retrieve tweet containing “elasticsearch” **or** “fossasia”
published before today by philipskokoh without geoLocation

```
{
  “query”: {
    “bool”: {
      “must”: [
        {“match”: { “text”: “elasticsearch fossasia” }}
      ],
      “filter“: [
        {“term”: { “user.screenName”: “philipskokoh” }},
        {“range”: { “createdAt”: {“lt”: “now/d” }}}
      ],
      “must_not”: [
        {“exists”: { “field”: “geoLocation” }}
      ]
    }
  }
}
```

Let’s Search!!

Retrieve tweet containing “elasticsearch” **or** “fossasia” published before today by philipskokoh without geoLocation

```
{
  “query”: {
    “bool”: {
      “must”: [
        {“match”: { “text”: “elasticsearch fossasia” }}
      ],
      “filter“: [
        {“term”: { “user.screenName”: “philipskokoh” }},
        {“range”: { “createdAt”: {“lt”: “now/d” }}}
      ],
      “must_not”: [
        {“exists”: { “field”: “geoLocation” }}
      ]
    }
  }
}
```

```
{
  "took": 22, "timed_out": false,
  "_shards": {
    "total": 42, "successful": 42, "failed": 0
  },
  "hits": {
    "total": 1, "max_score": 4.0619926,
    "hits": [
      {
        "_index": "plr_sg_tweet_201603",
        "_type": "tweet",
        "_id": "707403325390520320",
        "_score": 4.0619926,
        "_source": {
          ...
          "createdAt": "Mar 9, 2016 11:11:10 AM",
          "id": 707403325390520300,
          "text": "I will be giving a workshop at
#FOSSASIA 2016 titled: Elasticsearch: You know, for
search! and more! https://t.co/FRCQlQdHhH\nCome,
join us!□",
          "user": {
            ...
            "screenName": "philipskokoh",
            "lang": "en",
            "name": "Philips Kokoh",
          },
          "retweetCount": 2,
        }
      }
    ]
  }
}
```

Geo Distance Range Query

Retrieve tweets containing fossasia published before today within 2km from Science Centre

```
{
  "query": {
    "bool": {
      "must": [
        {"match": { "text": "mrt" }}
      ],
      "filter": [
        {"range": { "createdAt": {"lt": "now/d" }}},
        {"geo_distance_range": {
          "gt": "0km", "lt": "1km",
          "geoLocation": {
            "lat": 1.332906,
            "lon": 103.736110
          }
        }}
      ]
    }
  }
}
```

Geo Distance Range Query

Retrieve tweets containing fossasia published before today within 2km from Science Centre

```
{
  "query": {
    "bool": {
      "must": [
        {"match": { "text": "mrt" }}
      ],
      "filter": [
        {"range": { "createdAt": {"lt": "now/d" }}},
        {"geo_distance_range": {
          "gt": "0km", "lt": "1km",
          "geoLocation": {
            "lat": 1.332906,
            "lon": 103.736110
          }
        }}
      ]
    }
  }
}
```

```
{
  "took": 50, "timed_out": false,
  "_shards": {
    "total": 42, "successful": 42, "failed": 0
  },
  "hits": {
    "total": 974, "max_score": 3.6536937,
    "hits": [
      { ... },
      {
        "_index": "plr_sg_tweet_201602",
        "_type": "tweet",
        "_id": "700461563812192256",
        "_score": 3.646007,
        "_source": {
          ...
          "createdAt": "Feb 19, 2016 7:27:05 AM",
          "id": 700461563812192300,
          "text": "Mrt slow dao (@ Jurong East MRT
Interchange (NS1/EW24) - @smrt_singapore in
Singapore) https://t.co/K1av2dk7GI",
          "geoLocation": {
            "lat": 1.33378498,
            "lon": 103.74183655
          },
          "user": {
            "id": 252470398, ...
          }
        }
      },
      ...
    ]
  }
}
```

Geo Bounding Box Query

Retrieve tweets containing singapore published inside Marina Bay area

```
{
  “query”: {
    “bool”: {
      “must”: [
        {“match”: { “text”: “singapore” }}
      ],
      “filter”: [
        {“geo_bounding_box”: {
          “geoLocation”: {
            “top_left”: [103.852311, 1.289884],
            “bottom_right”: [103.860465, 1.279158]
          }
        }}
      ]
    }
  }
}
```

Accept GeoJSON format!

```
{
  "took": 8, "timed_out": false,
  "_shards": {
    "total": 42, "successful": 42, "failed": 0
  },
  "hits": {
    "total": 5758, "max_score": 4.6547956,
    "hits": [
      {
        "_index": "plr_sg_tweet_201602",
        "_type": "tweet",
        "_id": "696276978584801280",
        "_score": 4.6547956,
        "_source": {
          ...
          "text": "In #Singapore",
          "geoLocation": {
            "lat": 1.28902587,
            "lon": 103.85594832
          },
          "user": { ... },
          ...
        }
      },
      ...
    ]
  }
}
```

Nested Object

Retrieve tweets that starts with fossasia hashtag

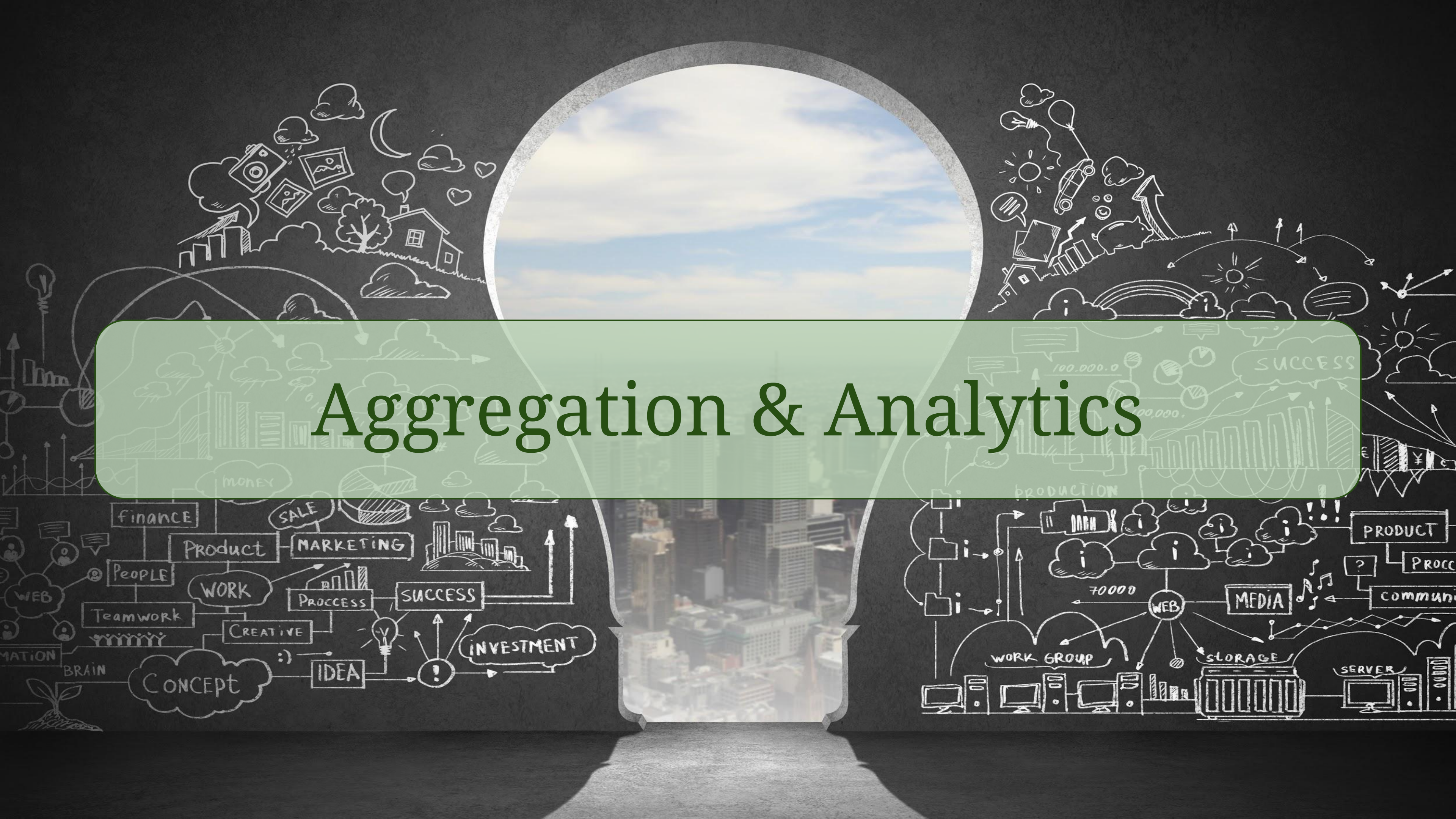
```
{
  “query”: {
    “nested”: {
      “path”: “hashtagEntities”,
      “query”: {
        “bool”: {
          “must”: [
            {“match”: { “hashtagEntities.text”: “fossasia”}}
          ],
          “filter“: [
            {“range”: { “hashtagEntities.start”: { “lt”: 1 
}}}
          ]
        }
      }
    }
  }
}
```

Nested Object

Retrieve tweets that starts with fossasia hashtag

```
{
  “query”: {
    “nested”: {
      “path”: “hashtagEntities”,
      “query”: {
        “bool”: {
          “must”: [
            {“match”: { “hashtagEntities.text”: “fossasia”}}
          ],
          “filter“: [
            {“range”: { “hashtagEntities.start”: { “lt”: 1
}}}
          ]
        }
      }
    }
  }
}
```

```
{
  "took": 6, "timed_out": false,
  "_shards": {
    "total": 42, "successful": 42, "failed": 0
  },
  "hits": {
    "total": 3, "max_score": 16.199848,
    "hits": [
      { "_score": 16.199848, ...
        "_source": {
          ...
          "hashtagEntities": [
            { "start": 0, "end": 9,
              "text": "FOSSASIA" }
          ],
          "text": "#FOSSASIA #GoogleCodeIn #GCI
speeding up. 150+ students currently working on
tasks! Great you are joining @hpdang
@mariobehling @mohitkanwal"
        }
      },
      { "_score": 15.867135, ...
        "_source": {
          ...
          "hashtagEntities": [
            { "start": 0, "end": 9,
              "text": "FOSSASIA" },
          ],
          "text": "#FOSSASIA 2016 is keen to
get more students to attend. Learn coding n
tech. Happy to share more details https://t.
co/1bRmvZVrOP #edsg"
        }
      }, { ... }
    ]
  }
}
```



Aggregation & Analytics

Aggregation and Analytics

Types of aggregations (that we often use):

- Terms Aggregation:
  - Bucketing documents based on numeric/textual content
- Date Histogram Aggregation:
  - Bucketing documents based on date/time value
- Geo Distance Aggregation
  - Bucketing documents based on distance from an origin location

Combined Aggregations

Combined Aggregations & Queries

Terms Aggregation

“What are the popular platforms Twitter users use?”

Terms Aggregation

Constant keywords

Name of the aggregation

```
{
"aggs": {
   "mostPopularSource": {
      "terms": {
         "field": "source",
         "size": 3
      }
    }
  }
}
```

Type of the aggregation

The targeted field to perform aggregation on

Limit the size of the result buckets

Terms Aggregation

```
{
"aggs": {
    "mostPopularSource": {
        "terms": {
            "field": "source",
            "size": 3
          }
       }
    }
}
```

```
{ ...
"aggregations": {
   "mostPopularSource": {
      "doc_count_error_upper_bound": 87696 ,
      "sum_other_doc_count": 12907898 ,
      "buckets": [
        {
            "key": "Twitter for iPhone" ,
            "Doc_count": 27928770
        },
        {
            "key": "Twitter for Android" ,
            "Doc_count": 21327691
        },
        {
            "key": "Twitter Web Client" ,
            "Doc_count": 6243422
        }
      ]
   }
}
```

Doc counts are approximate (-> upper bound on doc_count error for each term)

The sum of doc counts for buckets not in the response

Term: the bucket's keyword

The doc counts for this bucket

Date Histogram Aggregation

“Number of tweets collected each month?”

Date Histogram Aggregation

```
{
"aggs": {
   "numberOfTweetsByMonth": {
      "date_histogram": {
         "field": "createdAt",
         "interval": "month"
        }
      }
    }
}
```

Type of the aggregation

The targeted field to perform aggregation on

Define the interval to "bucket" the count

Date Histogram Aggregation

```
{
"aggs": {
    "numberOfTweetsByMonth": {
        "date_histogram": {
            "field": "createdAt",
            "interval": "month"
          }
        }
    }
}
```

→

```
{ ...
"aggregations": {
    "numberOfTweetsByMonth": {
       "buckets": [
          {
             "key_as_string":
                "Jan 1, 2016 12:0:0 AM",
             "key": 1451606400000
             "doc_count": 28067435
          },
          {
             "key_as_string":
                "Feb 1, 2016 12:0:0 AM",
             "key": 1454284800000
             "doc_count": 25912385
          },
          {
             "key_as_string":
                "Mar 1, 2016 12:0:0 AM",
             "key": 1456790400000
             "doc_count":  14427961
          }, …}}}
```

We have 28 millions tweets that was tweeted (createdAt) in Jan 2016

Terms + Date Histogram Aggregation Combined

“What is the platform Twitter users use?”

“On each day”

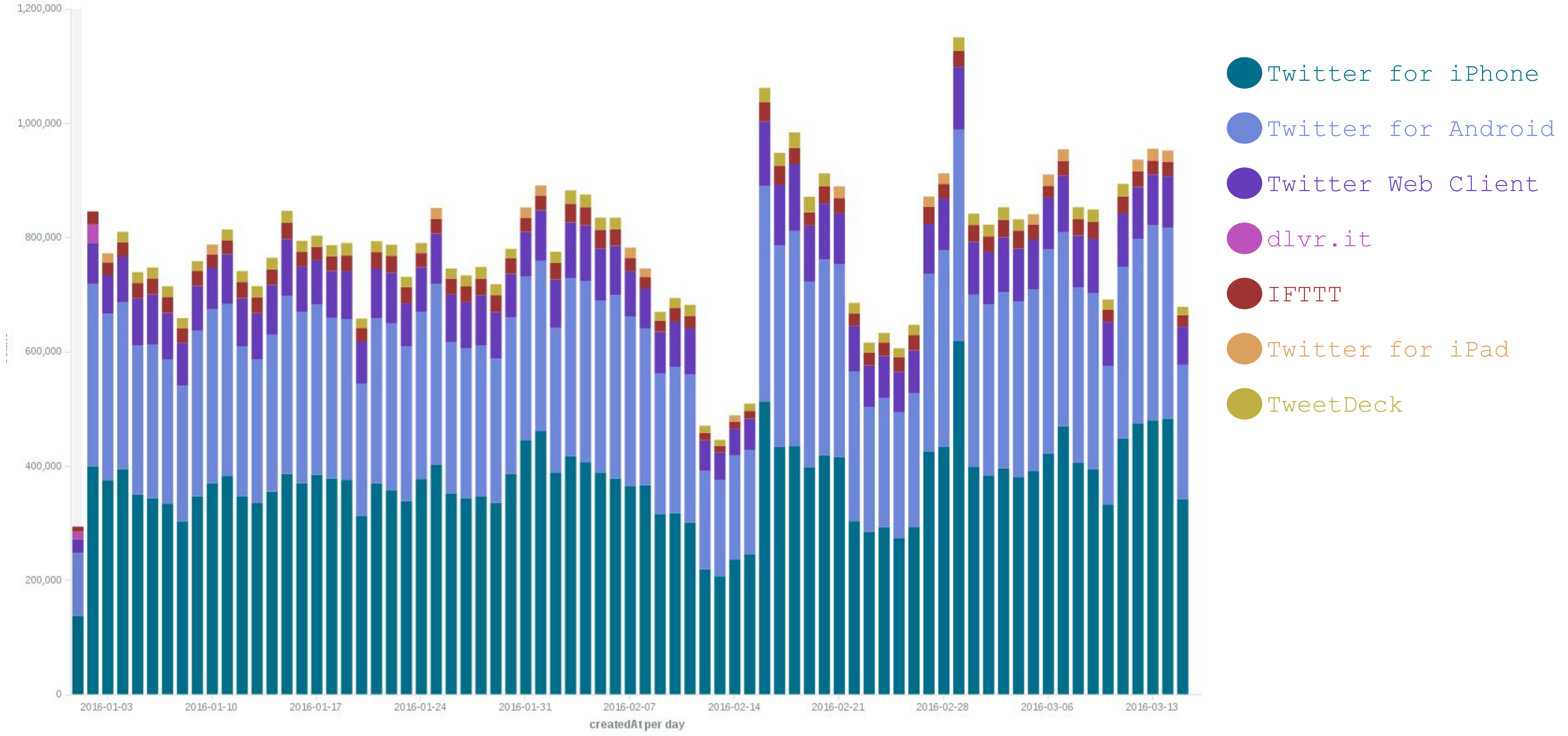
Terms + Date Histogram Aggregation Combined

```
{
  "aggs": {
    "numberOfTweetsByDay": {
      "date_histogram": {
        "field": "createdAt",
        "interval": "day"
      },
      "aggs": {
        "mostPopularSource": {
          "terms": {
            "field": "source"
          }
        }
      }
}}}
```

Aggregation

Sub-Aggregation

Terms + Date Histogram Aggregation Combined



Geo Distance Aggregations

```
{
“aggs”: {
   “numberOfTweetsByRadius”: {
      “geo_distance”: {
         “field”: “geoLocation”,
         “origin”: “1.3,103.8”,
         “unit”: “km”,
         “ranges”: [
            { “to”: 1 },
            { “from”: 1, “to”: 3},
            { “from”: 3 }
         ]
      }
   }
}
}
```

The targeted field

Type of the aggregation

Distances are computed from this origin

The radius’ unit

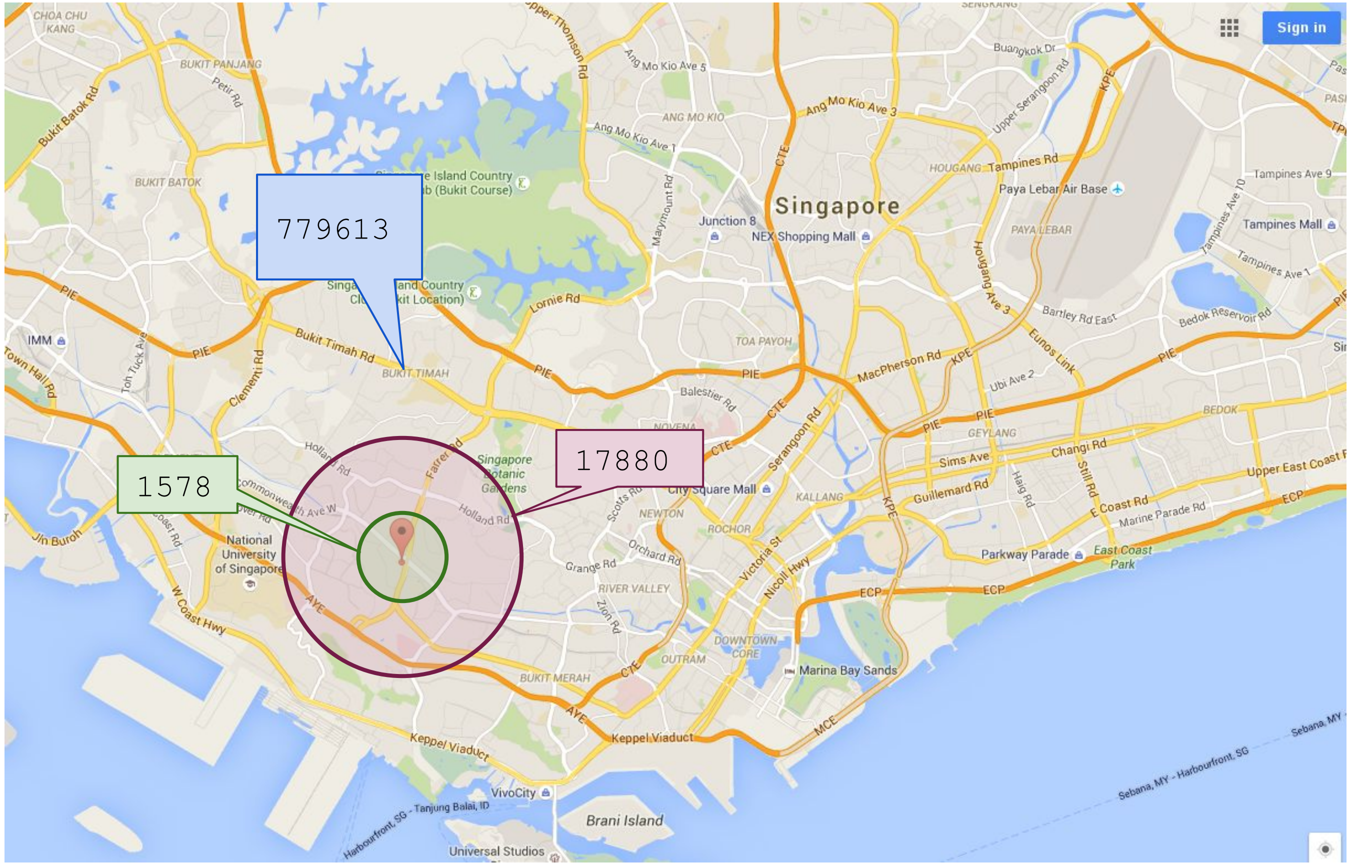
Define the buckets

Geo Distance Aggregations

```
{
“aggs”: {
   “numberOfTweetsByRadius”: {
      “geo_distance”: {
         “field”: “geoLocation”,
         “origin”: “1.3,103.8”,
         “unit”: “km”,
         “ranges”: [
            { “to”: 1 },
            { “from”: 1, “to”: 3},
            { “from”: 3 }
         ]
        }
      }
   }
}
```

```
{ ...
“aggregations”: {
   “numberOfTweetsByRadius”: {
      “buckets”: [
         {
         “key”: “*-1.0”,
         “from”: 0, “from_as_string”:“0.0”,
         “to”: 1, “to_as_string”: “1.0”
         “doc_count”: 1578
         },
          {
          “key”: “1.0-3.0”,
         “from”: 1, “from_as_string”:“1.0”,
         “to”: 3, “to_as_string”: “3.0”,
         “doc_count”: 17880
         },
          {
          “key”: “3.0-*”,
         “from”: 3, “from_as_string”:”3.0”
         “doc_count”: 779613
         }
       ]
}}}
```

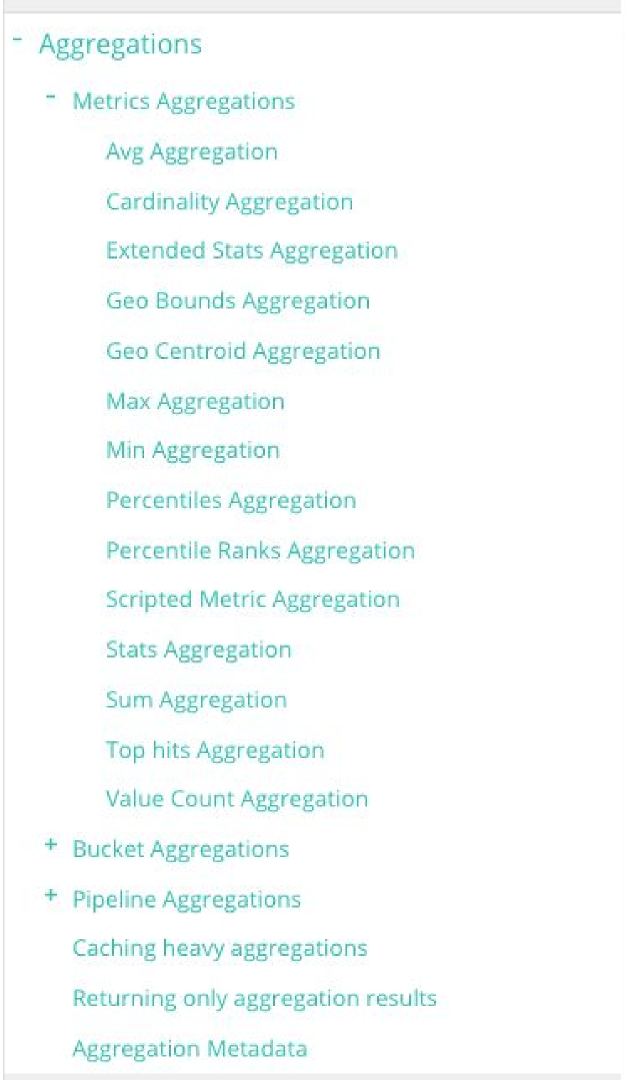
Geo Distance Aggregations



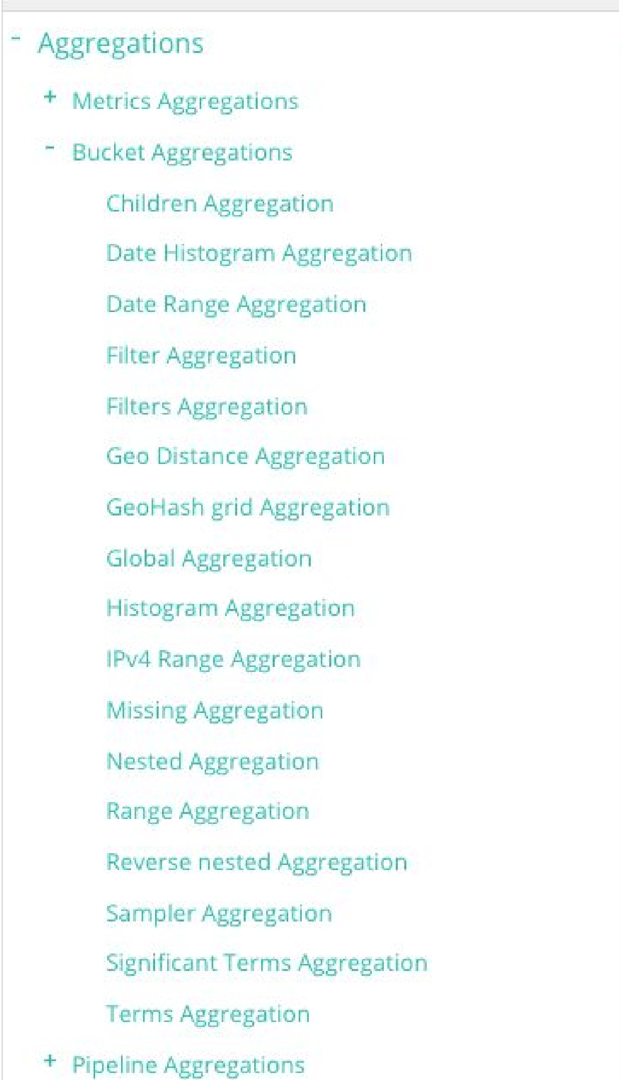
And many other types of aggregations

Refer to:

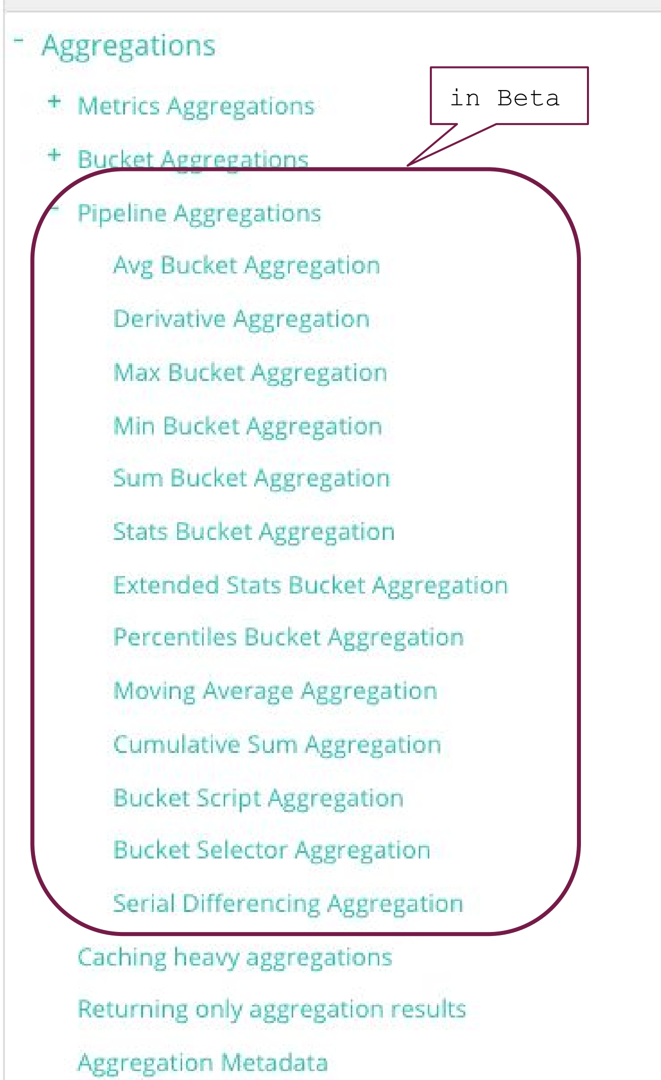
https://www.elastic.co/guide/en/elasticsearch/reference/current/search-aggregations.html



- − Aggregations
  - − Metrics Aggregations
    - Avg Aggregation
    - Cardinality Aggregation
    - Extended Stats Aggregation
    - Geo Bounds Aggregation
    - Geo Centroid Aggregation
    - Max Aggregation
    - Min Aggregation
    - Percentiles Aggregation
    - Percentile Ranks Aggregation
    - Scripted Metric Aggregation
    - Stats Aggregation
    - Sum Aggregation
    - Top hits Aggregation
    - Value Count Aggregation
  - + Bucket Aggregations
  - + Pipeline Aggregations
  - Caching heavy aggregations
  - Returning only aggregation results
  - Aggregation Metadata



- − Aggregations
  - + Metrics Aggregations
  - − Bucket Aggregations
    - Children Aggregation
    - Date Histogram Aggregation
    - Date Range Aggregation
    - Filter Aggregation
    - Filters Aggregation
    - Geo Distance Aggregation
    - GeoHash grid Aggregation
    - Global Aggregation
    - Histogram Aggregation
    - IPv4 Range Aggregation
    - Missing Aggregation
    - Nested Aggregation
    - Range Aggregation
    - Reverse nested Aggregation
    - Sampler Aggregation
    - Significant Terms Aggregation
    - Terms Aggregation
  - + Pipeline Aggregations



in Beta

- − Aggregations
  - + Metrics Aggregations
  - + Bucket Aggregations
  - − Pipeline Aggregations
    - Avg Bucket Aggregation
    - Derivative Aggregation
    - Max Bucket Aggregation
    - Min Bucket Aggregation
    - Sum Bucket Aggregation
    - Stats Bucket Aggregation
    - Extended Stats Bucket Aggregation
    - Percentiles Bucket Aggregation
    - Moving Average Aggregation
    - Cumulative Sum Aggregation
    - Bucket Script Aggregation
    - Bucket Selector Aggregation
    - Serial Differencing Aggregation
  - Caching heavy aggregations
  - Returning only aggregation results
  - Aggregation Metadata

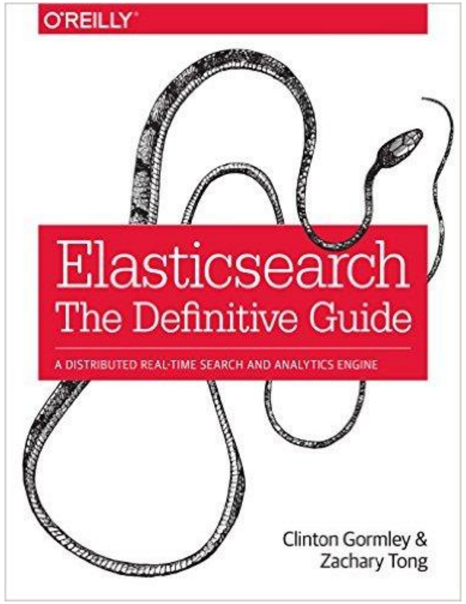
Lesson Learned

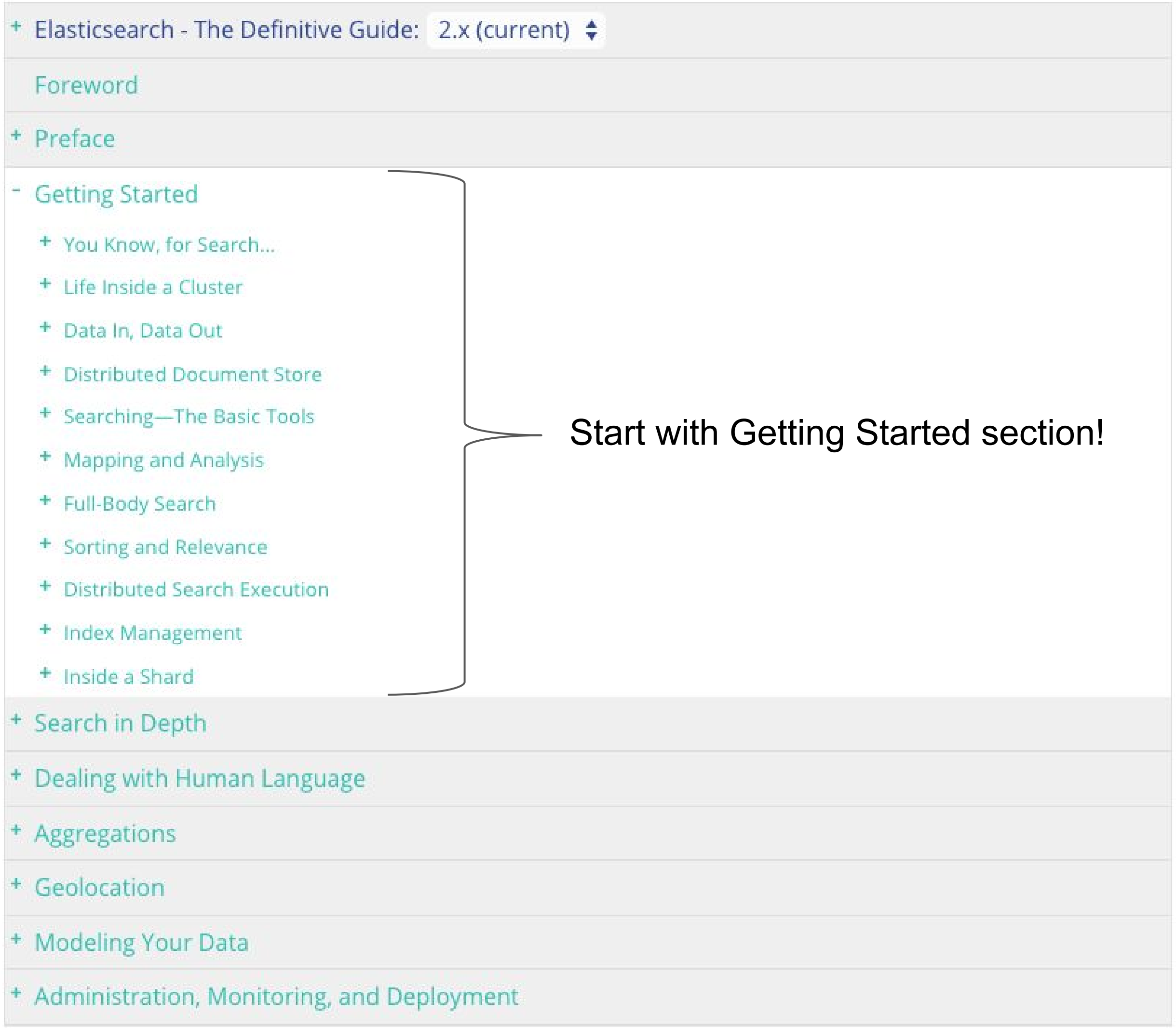
ES @ LARC: Lesson Learned

0.19.x -> 0.90.10 -> 1.7.0 -> 2.0.0

Read

Elasticsearch the Definitive Guide (online)





Start with Getting Started section!

Define the correct mapping

Elasticsearch dynamic mapping, OK for production?

```
{
  “id”: 709260211836485600,
  “createdAt”: “ Mar 14, 2016 2:09:46 PM ”,
  “text”: “ @arinto Just few more days to
      share #elasticsearch at
      #FOSSASIA 2016 https://t.co/PtZk14CNXl ”,
  “geoLocation”: {
    “latitude”: 1.2971576,
    “longitude”: 103.8495769
  },
  “favorites”: [{
    “screenName”: “arinto”,
    “origin”: “Indonesia”},
    {
    “screenName”: “casey”,
    “origin”: “Vietnam”}]
  }
}
```

string instead of date

double instead of geo_point

No relation between fields!
Searching for
(screenName == arinto && origin == vietnam)
will return both data

Define the correct mapping

```
{
  “id”: 709260211836485600,
  “createdAt”: “Mar 14, 2016 2:09:46 PM”,
  “text”: “@arinto Just few more days to
      share #elasticsearch at
      #FOSSASIA 2016 https://t.co/PtZk14CNXl”,
  “geoLocation”: {
    “latitude”: 1.2971576,
    “longitude”: 103.8495769
  },
  “favorites”: [{
    “screenName”: “arinto”,
    “origin”: “Indonesia”},
    {
    “screenName”: “casey”,
    “origin”: “Vietnam”}]
  }
}
```

```
{
//rest of the mapping
  "id": {
    "type": "long" },
  "createdAt": {
    "format": "MMM d, y h:m:s a",
    "type": "date"},
  "text": {
    "type": "string" },
  "geoLocation": {
    "type": "geo_point" },
  "favorites": {
    "type": "nested"
    "properties": {…}}
//..rest of the mapping
}
```

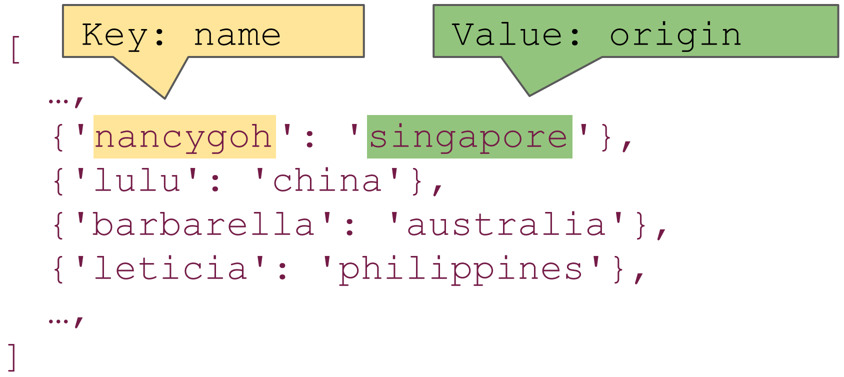
Elasticsearch dynamic mapping, not OK for production.

Define the correct mapping! Check the docs to learn more about mapping

KV store in Elasticsearch

Key-value store in Elasticsearch

- Field name as the key
- 10 or 100 keys are okay..
- What if you have million of keys?
- Does it scale?



```
[
  …,
  {'nancygoh': 'singapore'},
  {'lulu': 'china'},
  {'barbarella': 'australia'},
  {'leticia': 'philippines'},
  …,
]
```

KV store - Mapping Explosion!

- Dynamically add new fields in a mapping is an **expensive** operation
  - Lock the index, add new fields, and propagate index structure changes
- Halt the cluster!

KV store - Solution

Correct mapping:

```
{
  "mydata": {
    "mappings": {
      "kv": {
        "dynamic": "strict",
        "properties": {
          "key": {
            "type": "string",
             "index": "not_analyzed" },
          "value": {
            "type": "nested",
            "properties": {.....} //detail hidden
          }
        }
      }
    }
  }
}
```

Sample indexed data:

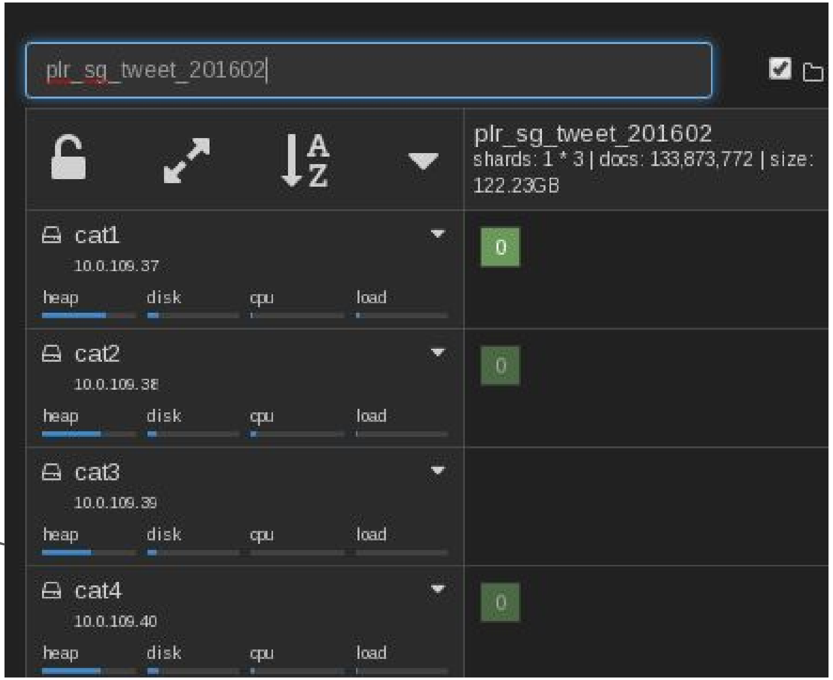
```
[{
     "key": "nancygoh",
     "value": { "origin": "singapore " }
}, {
     "key": "lulu",
     "value": { "origin": "china" }
}, {
     "key": "barbarella",
     "value": { "origin": "australia" }
}, {
     "key": "leticiabongnino",
     "value": { "origin": "philippines" }
}]
```

Keep the number of fields under control!

Shards

- No magic number
  - However.. you must determine when you create the index
  - Estimate data growth
  - Prepare for reindex

kopf plugin


plr_sg_tweet_201602

plr_sg_tweet_201602
shards: 1 * 3 | docs: 133,873,772 | size: 122.23GB

cat1
10.0.109.37
heap disk cpu load
0

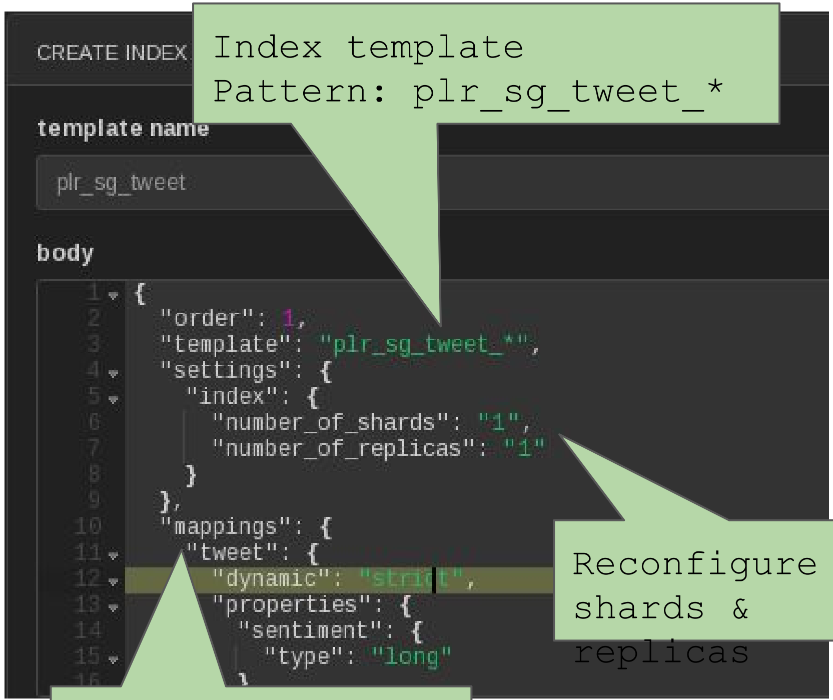
cat2
10.0.109.38
heap disk cpu load
0

cat3
10.0.109.39
heap disk cpu load

cat4
10.0.109.40
heap disk cpu load
0

Index template & rolling index

Rolling(time-based) index



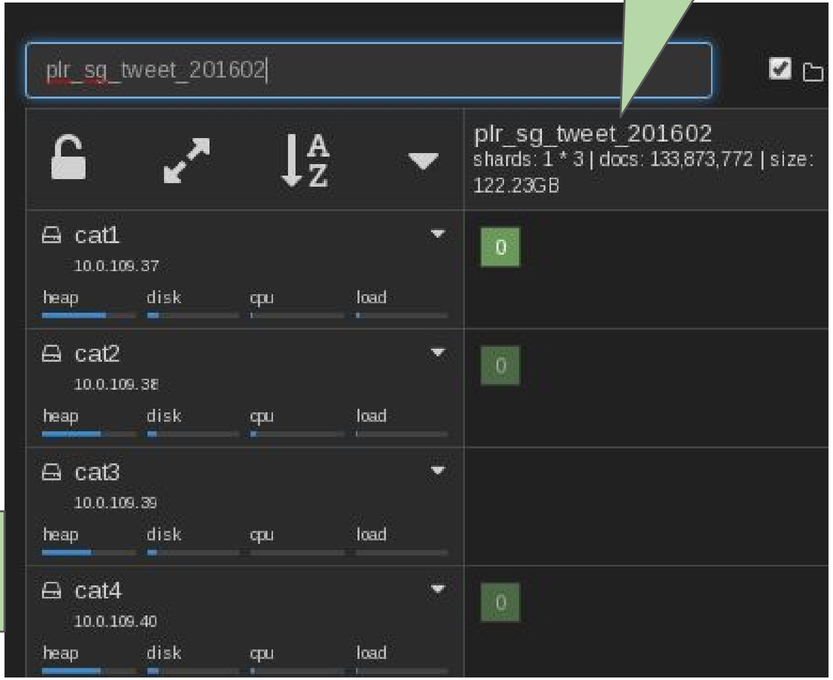
plr_sg_tweet

```
{
  "order": 1,
  "template": "plr_sg_tweet_*",
  "settings": {
    "index": {
      "number_of_shards": "1",
      "number_of_replicas": "1"
    }
  },
  "mappings": {
    "tweet": {
      "dynamic": "strict",
      "properties": {
        "sentiment": {
          "type": "long"
        }
```

Index template
Pattern: plr_sg_tweet_*

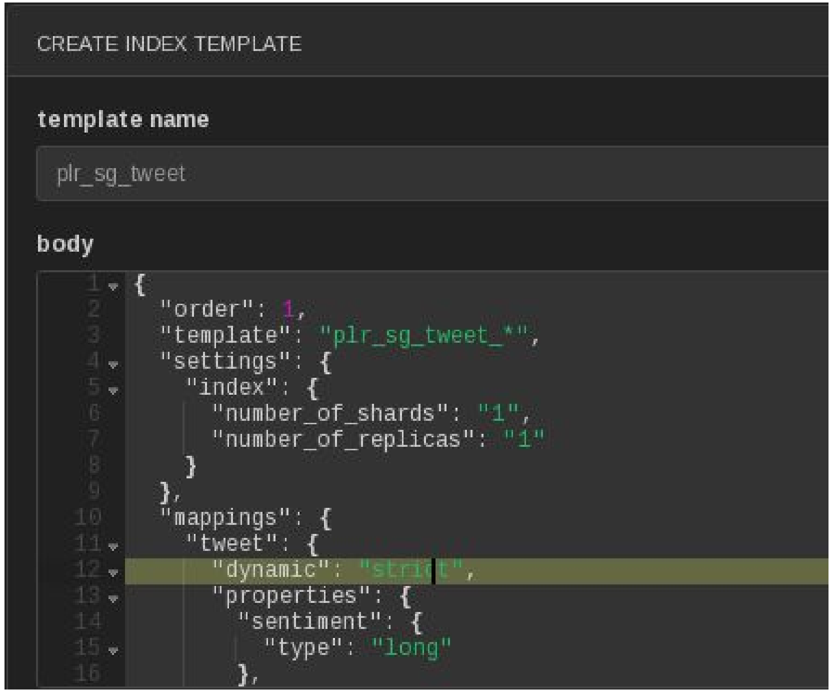
Reconfigure shards & replicas

Define the correct mapping



plr_sg_tweet_201602
shards: 1 * 3 | docs: 133,873,772 | size: 122.23GB

Problematic shard



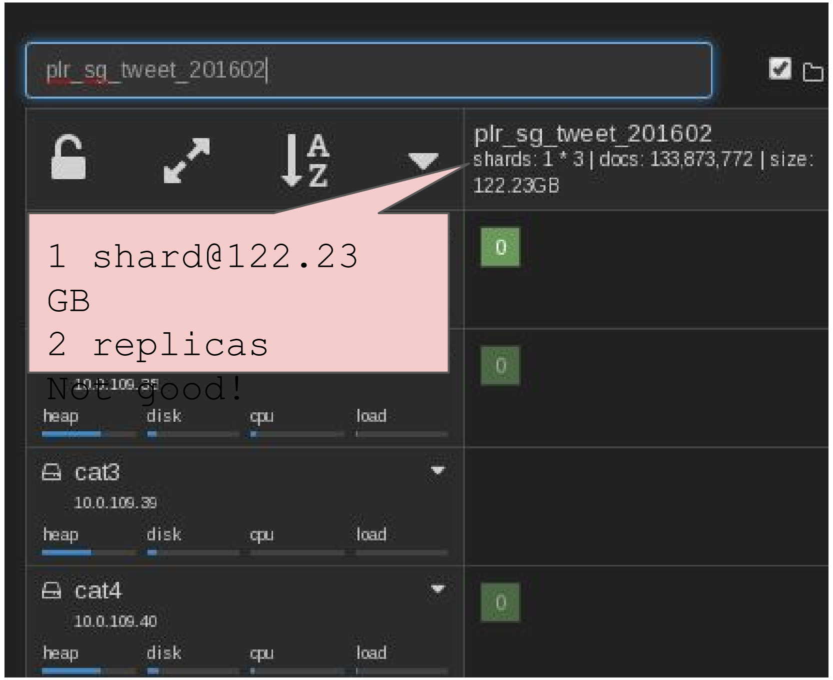
CREATE INDEX TEMPLATE

template name

plr_sg_tweet

body

```
{
  "order": 1,
  "template": "plr_sg_tweet_*",
  "settings": {
    "index": {
      "number_of_shards": "1",
      "number_of_replicas": "1"
    }
  },
  "mappings": {
    "tweet": {
      "dynamic": "strict",
      "properties": {
        "sentiment": {
          "type": "long"
        },
```

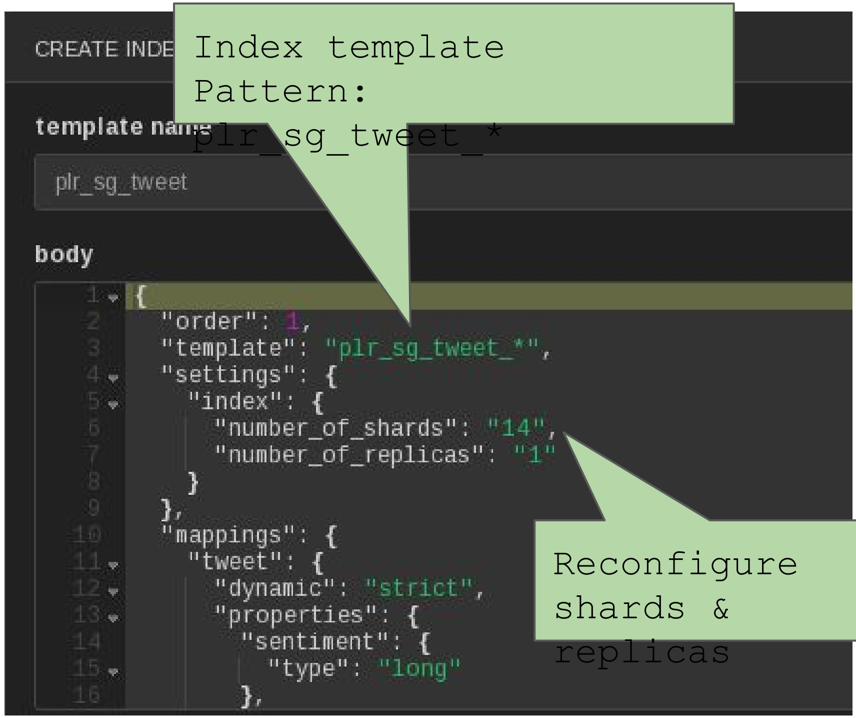


plr_sg_tweet_201602

plr_sg_tweet_201602
shards: 1 * 3 | docs: 133,873,772 | size: 122.23GB

1 shard@122.23 GB
2 replicas
Not good!

Modify index template



CREATE INDE

template name

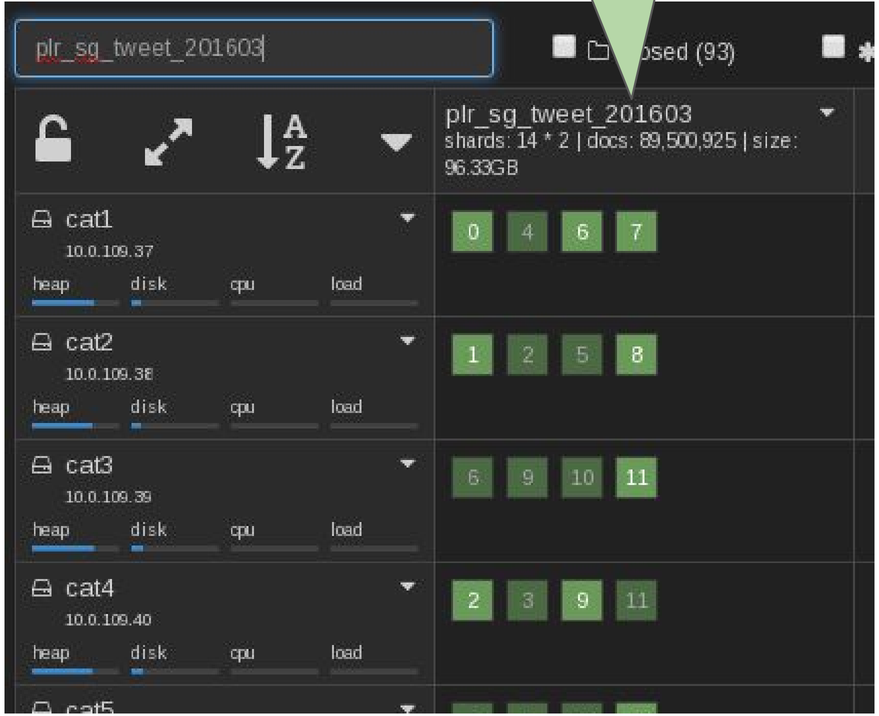
plr_sg_tweet

body

```
{
  "order": 1,
  "template": "plr_sg_tweet_*",
  "settings": {
    "index": {
      "number_of_shards": "14",
      "number_of_replicas": "1"
    }
  },
  "mappings": {
    "tweet": {
      "dynamic": "strict",
      "properties": {
        "sentiment": {
          "type": "long"
        },
```

Index template Pattern: plr_sg_tweet_*

Reconfigure shards & replicas



plr_sg_tweet_201603

osed (93)

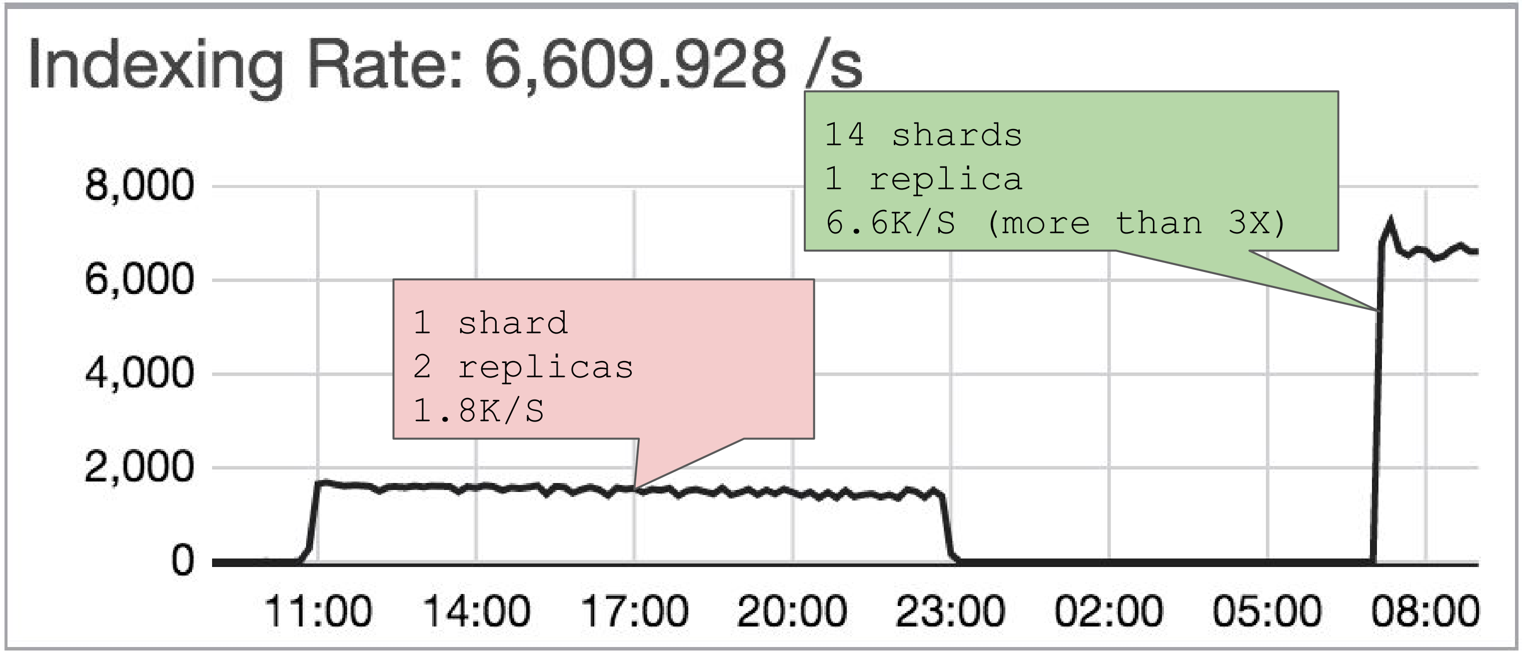
plr_sg_tweet_201603
shards: 14 * 2 | docs: 89,500,925 | size: 96.33GB

cat1 10.0.109.37 heap disk cpu load	0 4 6 7
cat2 10.0.109.38 heap disk cpu load	1 2 5 8
cat3 10.0.109.39 heap disk cpu load	6 9 10 11
cat4 10.0.109.40 heap disk cpu load	2 3 9 11
cat5	

201603
14 shards
1 replica

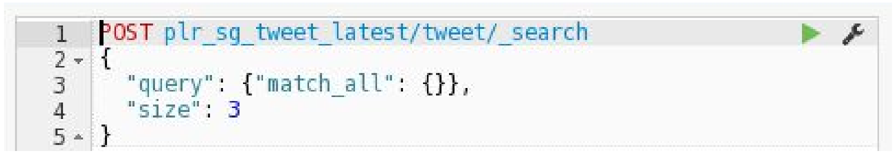
Indexing performance

Tweet bulk-loading indexing performance

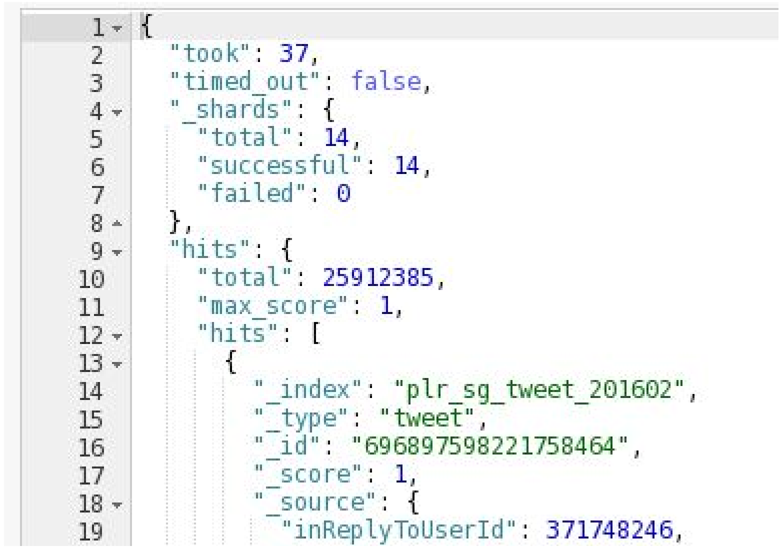


Alias

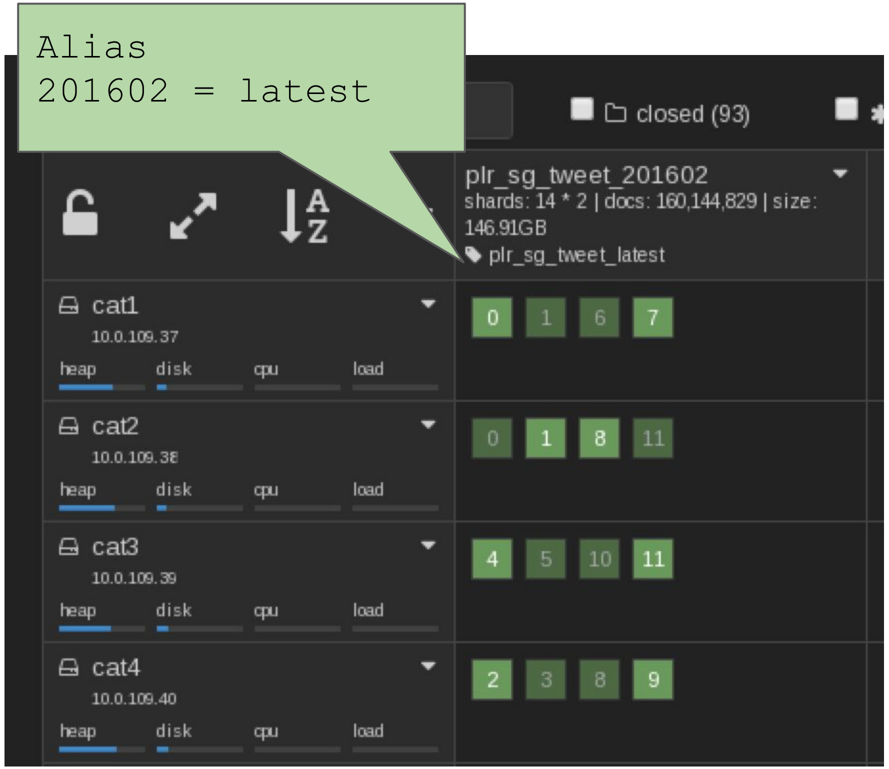
No change in application code:



```
POST plr_sg_tweet_latest/tweet/_search
{
  "query": {"match_all": {}},
  "size": 3
}
```



```
{
  "took": 37,
  "timed_out": false,
  "_shards": {
    "total": 14,
    "successful": 14,
    "failed": 0
  },
  "hits": {
    "total": 25912385,
    "max_score": 1,
    "hits": [
      {
        "_index": "plr_sg_tweet_201602",
        "_type": "tweet",
        "_id": "696897598221758464",
        "_score": 1,
        "_source": {
          "inReplyToUserId": 371748246,
```



```
Alias
201602 = latest
```

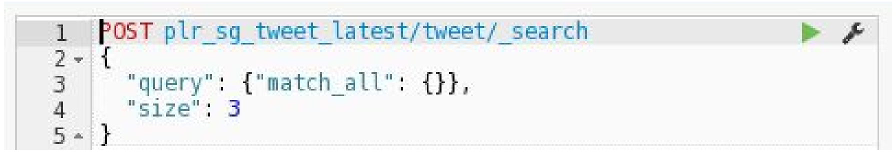
closed (93)

plr_sg_tweet_201602
shards: 14 * 2 | docs: 160,144,829 | size: 146.91GB
plr_sg_tweet_latest

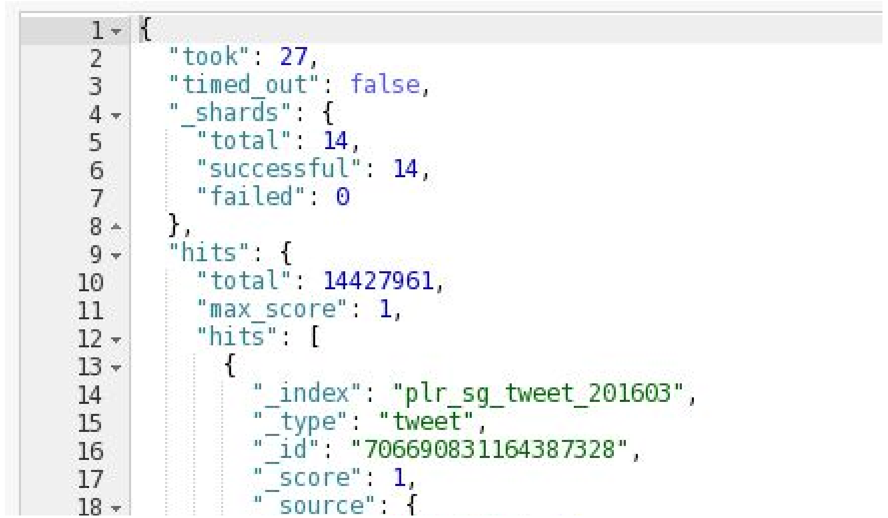
cat1	10.0.109.37	0	1	6	7
cat2	10.0.109.38	0	1	8	11
cat3	10.0.109.39	4	5	10	11
cat4	10.0.109.40	2	3	8	9

Alias

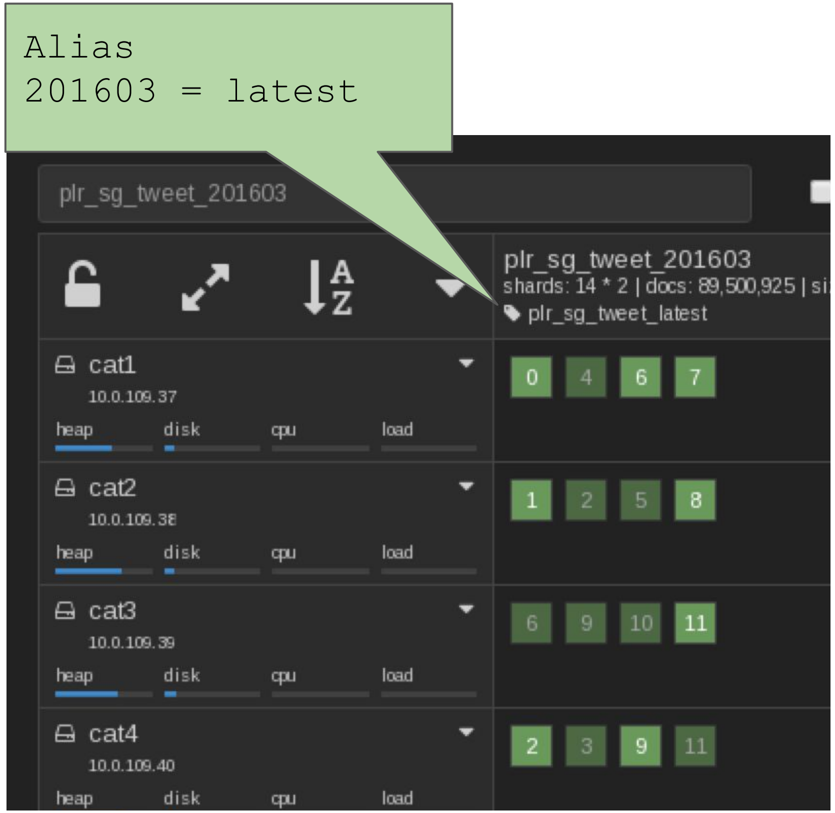
No change in application code:



```
POST plr_sg_tweet_latest/tweet/_search
{
  "query": {"match_all": {}},
  "size": 3
}
```



```
{
  "took": 27,
  "timed_out": false,
  "_shards": {
    "total": 14,
    "successful": 14,
    "failed": 0
  },
  "hits": {
    "total": 14427961,
    "max_score": 1,
    "hits": [
      {
        "_index": "plr_sg_tweet_201603",
        "_type": "tweet",
        "_id": "706690831164387328",
        "_score": 1,
        "_source": {
```



Alias
201603 = latest

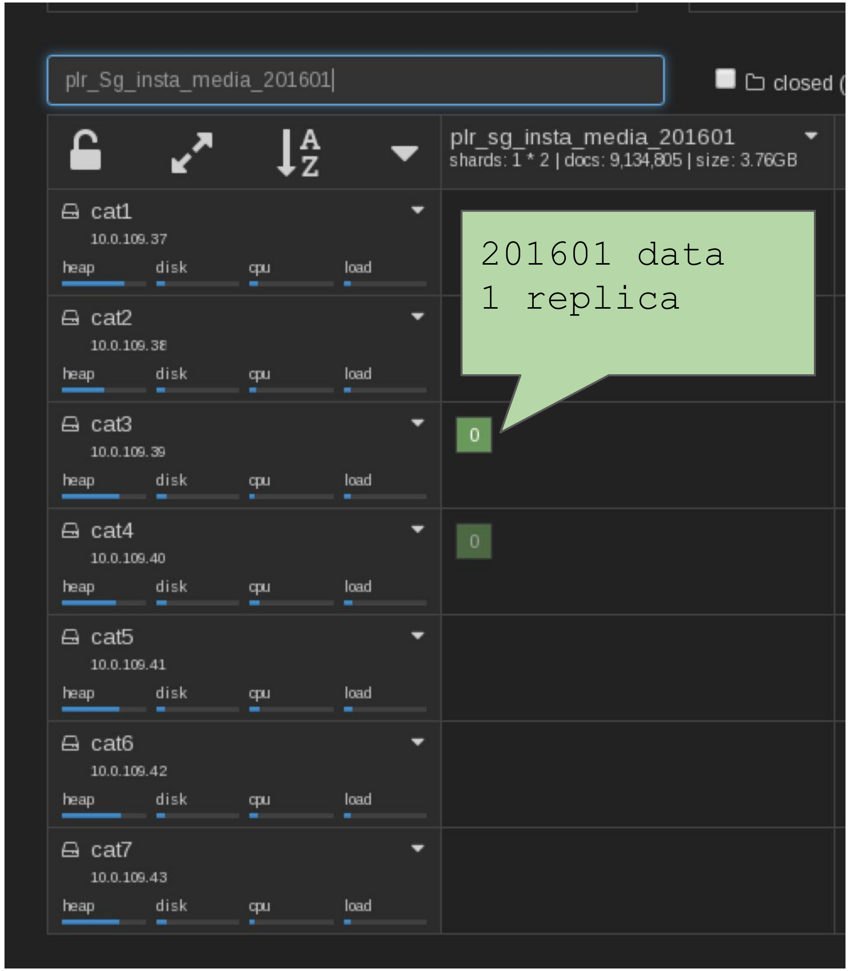
plr_sg_tweet_201603

plr_sg_tweet_201603
shards: 14 * 2 | docs: 89,500,925 | si
plr_sg_tweet_latest

cat1 10.0.109.37 heap disk cpu load	0 4 6 7
cat2 10.0.109.38 heap disk cpu load	1 2 5 8
cat3 10.0.109.39 heap disk cpu load	6 9 10 11
cat4 10.0.109.40 heap disk cpu load	2 3 9 11

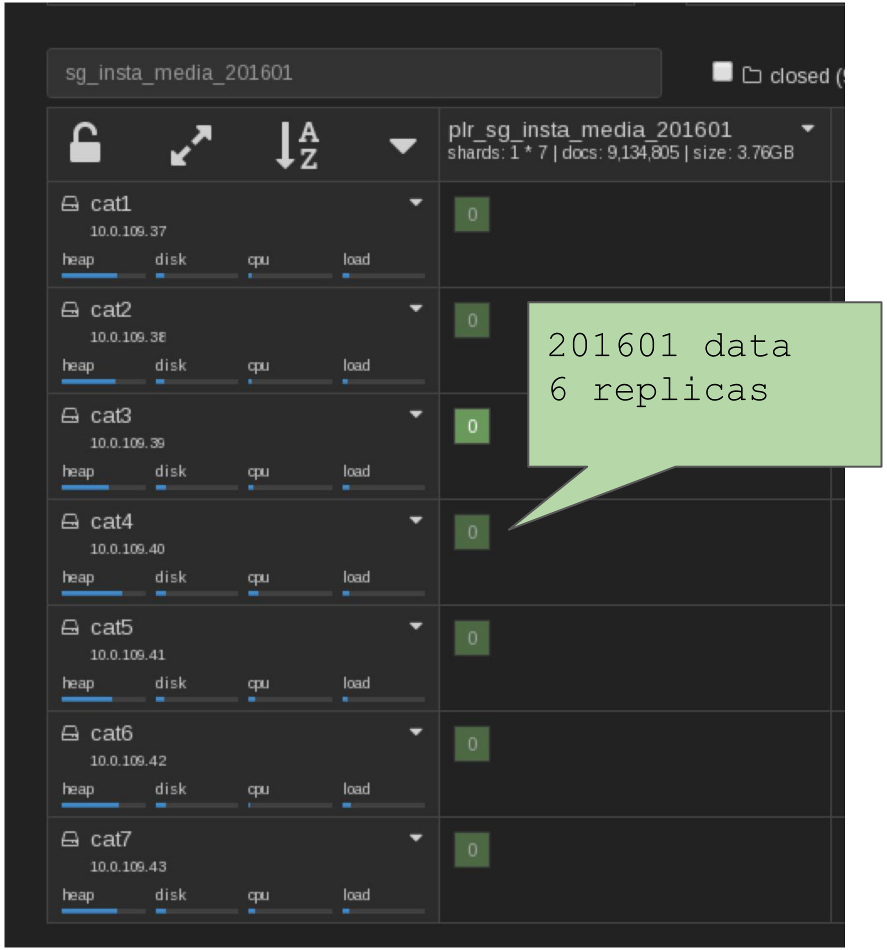
Replicas

- You can change it on the fly
- Start with 1 for fault tolerance
- What happen when the read request rate is very high?


plr_Sg_insta_media_201601
closed
plr_sg_insta_media_201601
shards: 1 * 2 | docs: 9,134,805 | size: 3.76GB
cat1 10.0.109.37 heap disk cpu load
cat2 10.0.109.38 heap disk cpu load
cat3 10.0.109.39 heap disk cpu load
cat4 10.0.109.40 heap disk cpu load
cat5 10.0.109.41 heap disk cpu load
cat6 10.0.109.42 heap disk cpu load
cat7 10.0.109.43 heap disk cpu load
0
0
201601 data
1 replica

Replicas

- Increase accordingly to
  - balance load
  - increasing the availability
  - scale up read requests



sg_insta_media_201601

closed (

plr_sg_insta_media_201601
shards: 1 * 7 | docs: 9,134,805 | size: 3.76GB

```
cat1
10.0.109.37
heap   disk   cpu   load          0

cat2
10.0.109.38
heap   disk   cpu   load          0

cat3
10.0.109.39
heap   disk   cpu   load          0

cat4
10.0.109.40
heap   disk   cpu   load          0

cat5
10.0.109.41
heap   disk   cpu   load          0

cat6
10.0.109.42
heap   disk   cpu   load          0

cat7
10.0.109.43
heap   disk   cpu   load          0
```

201601 data
6 replicas

Lesson Learned

- Read the book
- Define correct mapping
- Index template & rolling indices
- Use `alias`
- Scale up using replicas

Q&A

Elastic Training in Singapore:

- Core Elasticsearch: Operations 25 April 2016
- Core Elasticsearch: Developer 26-27 April 2016

training.elastic.co

LARC
LIVING ANALYTICS RESEARCH CENTRE

We're Hiring!!!

bit.ly/larchiring