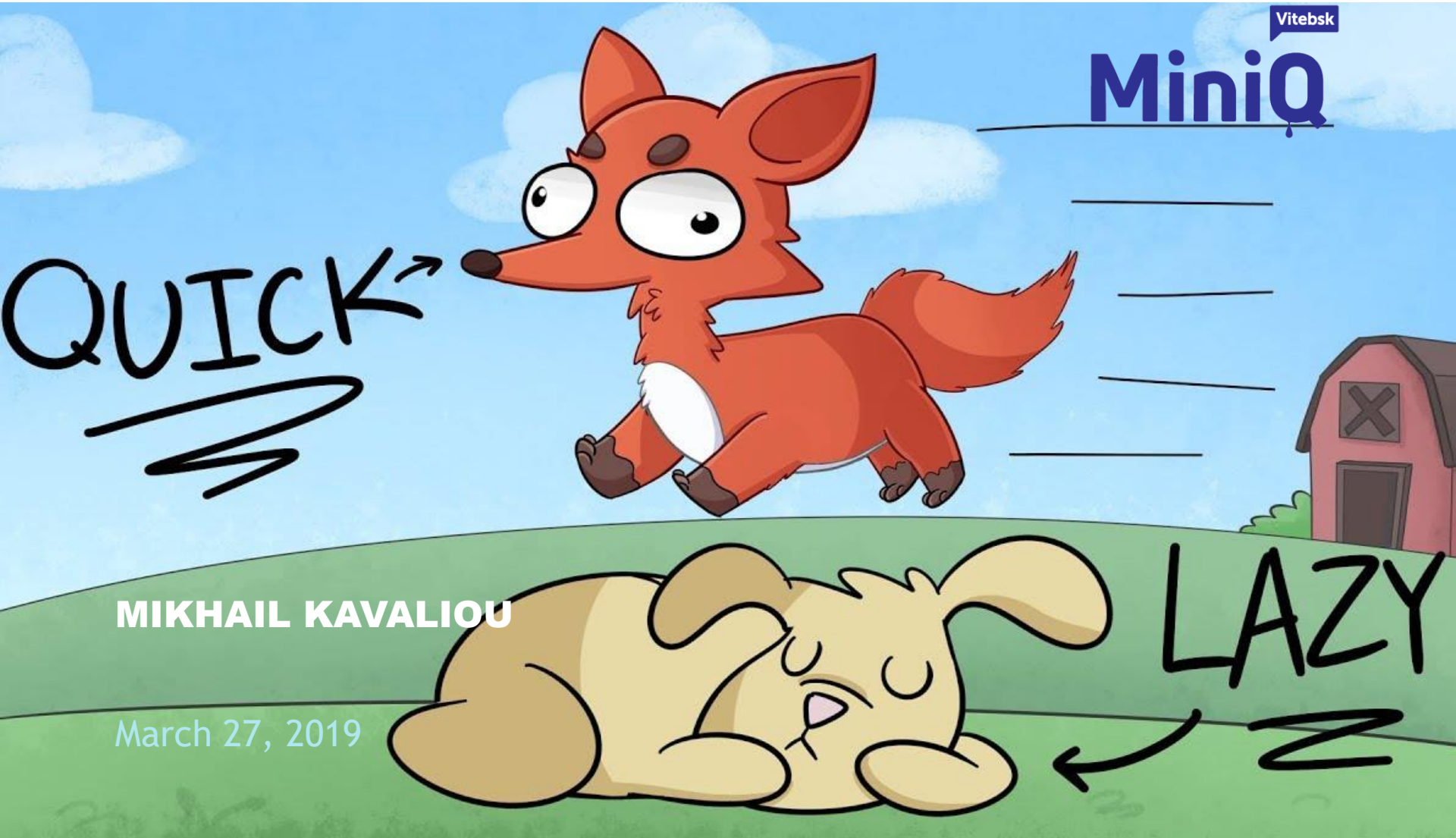
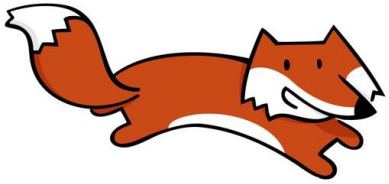
Elasticsearch
Производительность моей мечты



MIKHAIL KAVALIOU

March 27, 2019

A minute for me



Mikhail Kavaliou

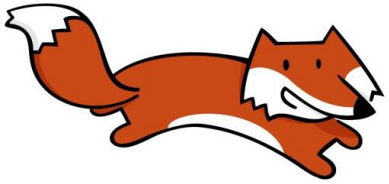
Senior Java Developer

6+ years in Java

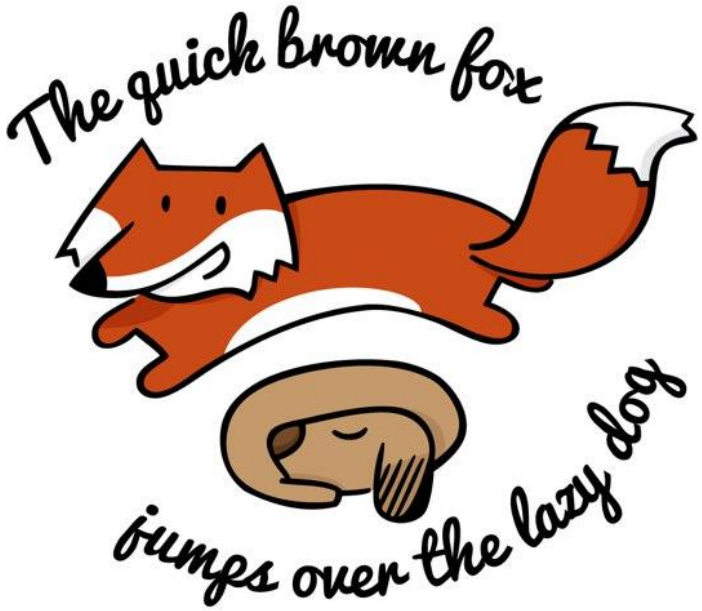
3+ years in EPAM

2+ years in ES

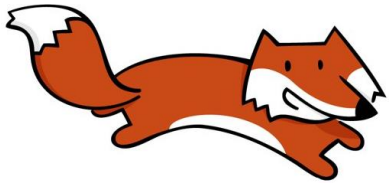
A minute for agenda



1. A minute for ES
2. A minute for stats
3. A minute for examples
4. A minute for questions



A minute for terms



NODE

INDEX

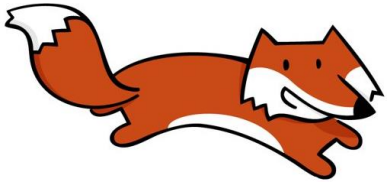
SHARD

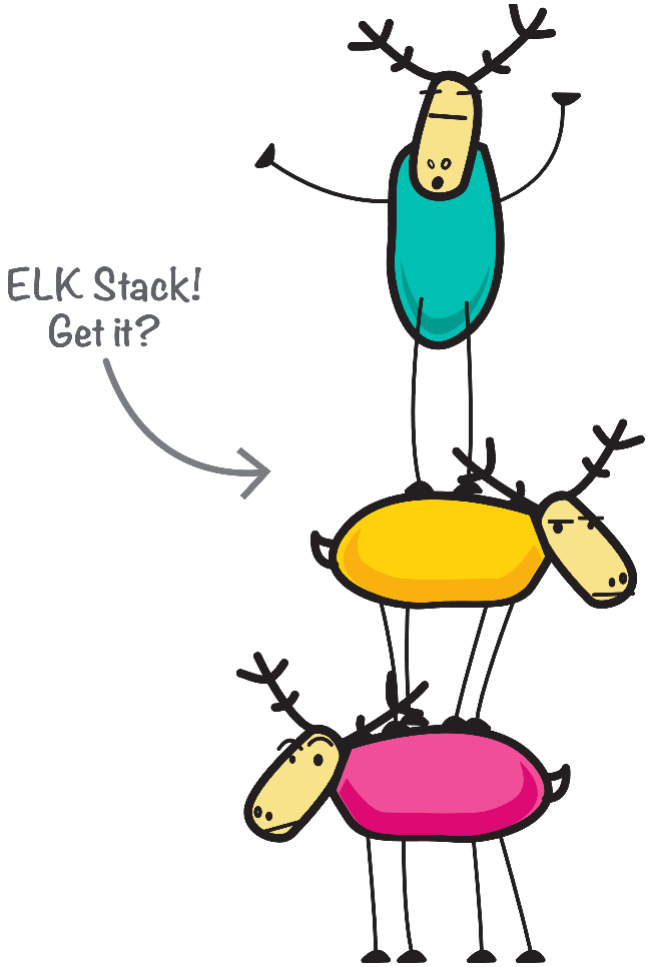
SCHEME (MAPPING)

CLUSTER

AGGREGATION

A minute for ES



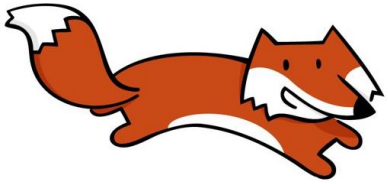


E Elasticsearch

L Logstash

K Kibana

A minute for stats

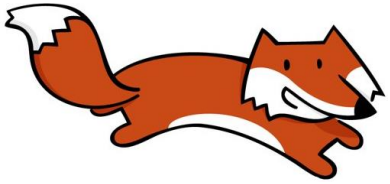


9-10M calls per day ~ 100-120 RPS

Page Results

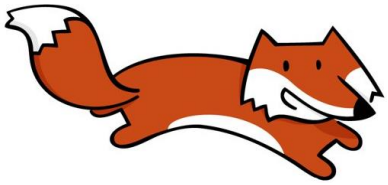
Page Type	Pass/Fail	Time to First Byte (TTFB)	Expected TTFB
inv	✓	24	187.0
inv	✓	8	154.0
inv	✓	29	200.0
inv	✓	14	200.0
inv	✓	17	143.0
inv	✓	45	200.0
inv	✓	46	200.0

Schemaless item



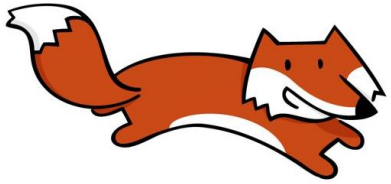

```
{
    "_index": "schemaless-backup",
    "_type": "item",
    "_id": "1",
    "_version": 3,
    "_seq_no": 19875,
    "_primary_term": 2,
    "found": true,
    "_source": {
        "stringField": "stringValue",
        "numericField": 42,
        "doubleField": 42.1,
        "booleanField": true,
        "dateField": "2019-03-15T15:53:00+03:00",
        "nullField": null
    }
}
```

Schemaless schema



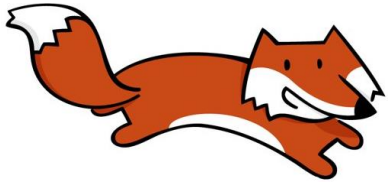
```
    "mappings": {
        "item": {
            "properties": {
                "booleanField": {
                    "type": "boolean"
                },
                "dateField": {
                    "type": "date"
                },
                "doubleField": {
                    "type": "float"
                },
                "numericField": {
                    "type": "long"
                },
                "stringField": {
                    "type": "text",
                    "fields": {
                        "keyword": {
                            "type": "keyword",
                            "ignore_above": 256
                        }
                    }
                }
            }
        }
    },
```

Schema settings



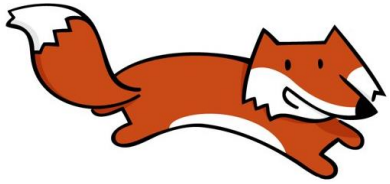
```
{
   "settings":{
      //"aliases": {
      //   "index-alias": {}
      //},
      "index":{
         "refresh_interval":"60s"
      },
      "search": {
          "slowlog": {
              "threshold": {
                  "fetch": {
                      "warn": "300ms"
                  },
                  "query": {
                      "warn": "1500ms"
                  }
              }
          }
      }
   },
   "mappings":{
```

Schema mapping



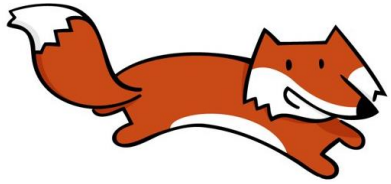
```
"mappings":{
   "item":{
      // "date_detection": false
      // "dynamic_date_formats": ["MM/dd/yyyy"]
      // "numeric_detection": true
      "properties":{
         // "dynamic": true
         "dateField": {
            "type":"date"
            // "doc_values": false
            // "index": false
         },
         "textField": {
             "type":"text",
             // "norms": false
             // "eager_global_ordinals": false - terms aggs
             "fields": {
                 "cityName": {
                     "type": "keyword"
                 }
             }
         },
         "objectField": {
             "enabled": false // no parsing at all, faster indexing
         },
         "doubleField":{
            "type":"double"
         },
         "numericField":{
            "type":"long"
         },
         "stringField":{
            "type":"keyword"
         },
         "extraFields1":{
```

Schema mapping



```
    "mappings": {
        "item": {
            "properties": {
                "dateField": {
                    "type": "date"
                },
                "doubleField": {
                    "type": "double"
                },
                "extraFields1": {
                    "type": "keyword"
                },
                "numericField": {
                    "type": "long"
                },
                "objectField": {
                    "type": "object",
                    "enabled": false
                },
                "stringField": {
                    "type": "keyword"
                },
                "textField": {
                    "type": "text",
                    "fields": {
                        "cityName": {
                            "type": "keyword"
                        }
                    }
                }
            }
        }
    },
```

Schema refresh interval

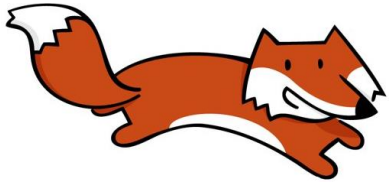


Refresh interval = 60 seconds:
Took 37.7823662 seconds.

Default refresh interval (real time):
Took 45.6564286 seconds.
+ 20 %

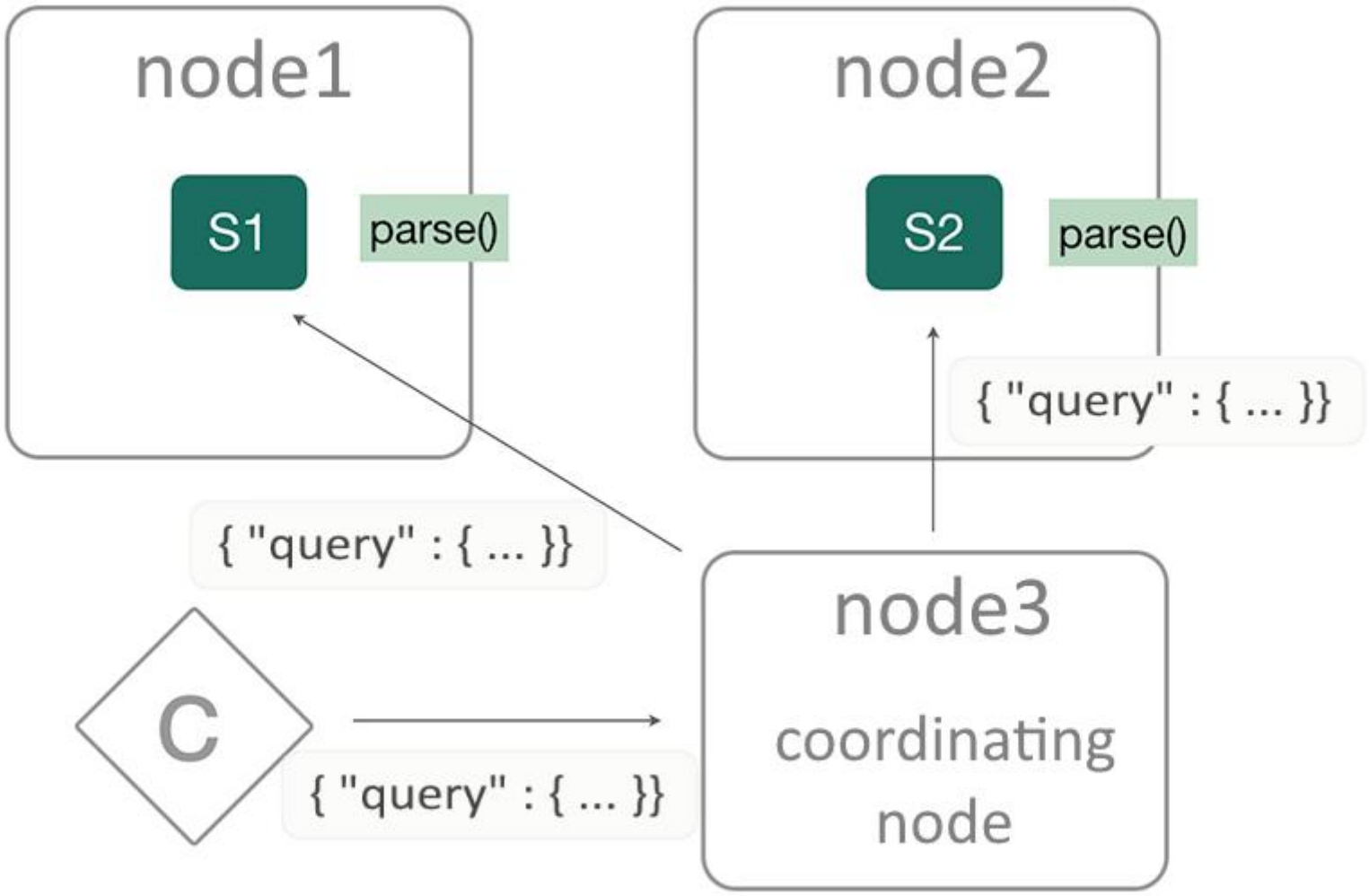
Данные могут быть недоступны для чтения в течение заданного периода refresh interval – **eventual consistency**

Schema sizes

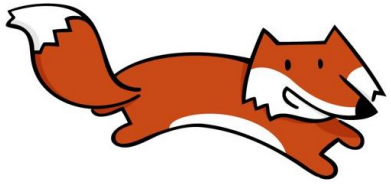


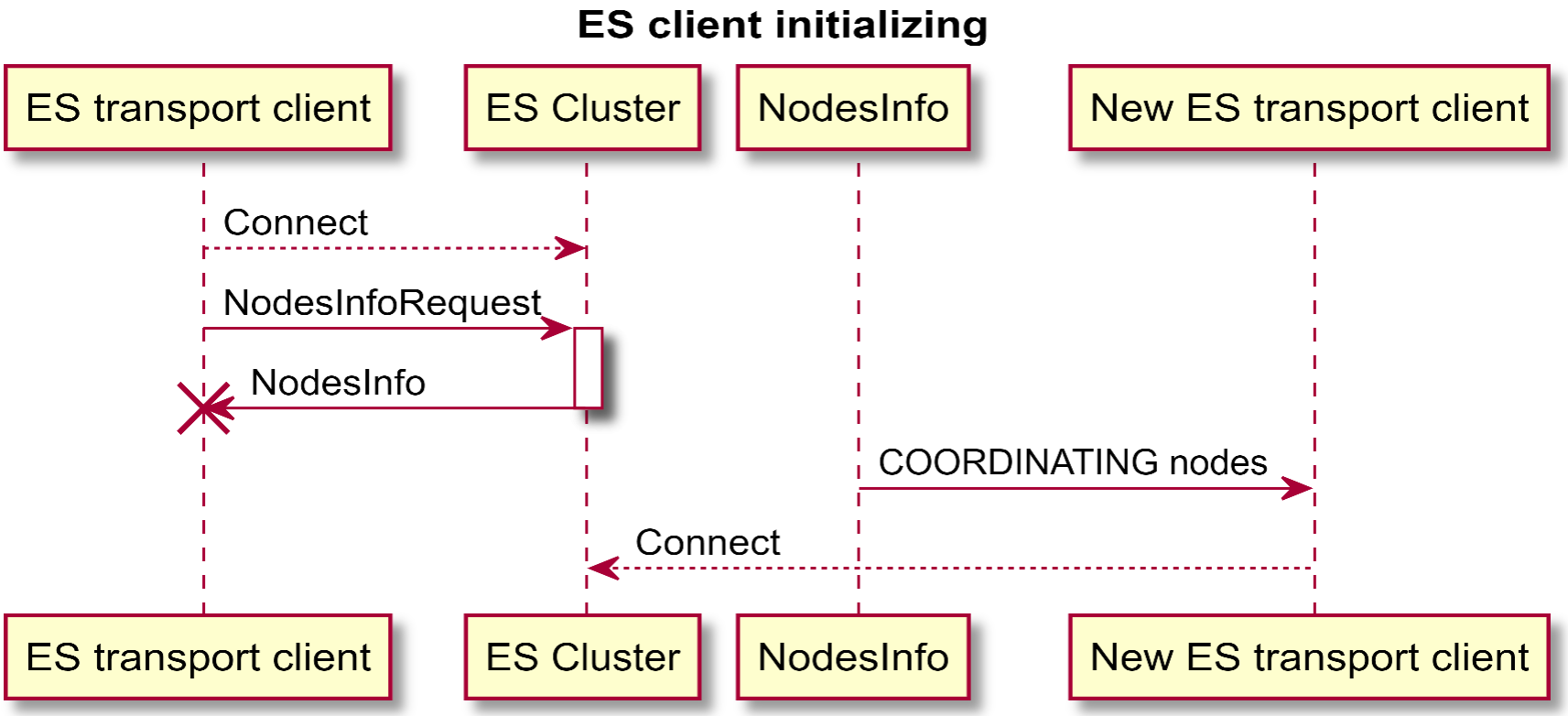

```
252        "schemaless-backup": {
253            "uuid": "eTGnxEr5SZybhitM5VZ7vg",
254            "primaries": {
255                "docs": {
256                    "count": 100000,
257                    "deleted": 1
258                },
259                "store": {
260                    "size_in_bytes": 29392345
261                },
497        "schemaful-backup": {
498            "uuid": "F6QxZd8PSlOV3RSH3zkZRA",
499            "primaries": {
500                "docs": {
501                    "count": 100000,
502                    "deleted": 0
503                },
504                "store": {
505                    "size_in_bytes": 24973230
506                },
```

Coordinating nodes

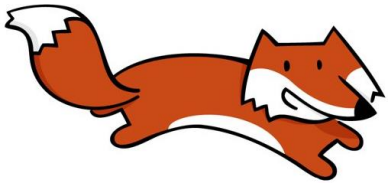


Coordinating nodes





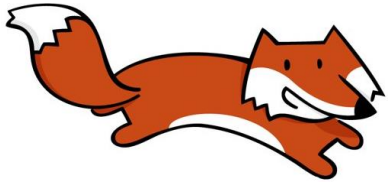
Timeouts



```
private SearchRequestBuilder
        getLimitedWithTimeout(TransportClient client) {

    return client.prepareSearch("schemaful").setTypes("item")
            .setTimeout(TimeValue.timeValueMillis(100))
            .setQuery(QueryBuilders.matchAllQuery())
            .setSize(1);
}
```

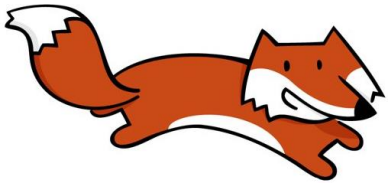
Fields projection



```
private SearchRequestBuilder
        getProjected(TransportClient client) {

    return client.prepareSearch("schemaful").setTypes("item")
            .setQuery(QueryBuilders.matchAllQuery())
            .setSource(new SearchSourceBuilder()
                    .fetchSource(new String[]{"numericField"},
                                 new String[]{})
            );
}
```

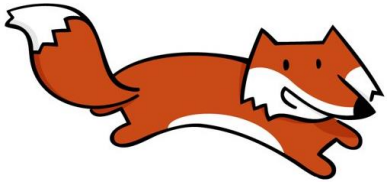
Filter VS must



```
private SearchRequestBuilder
        getFiltered(TransportClient client) {

    return client.prepareSearch("schemaful").setTypes("item")
            .setQuery(QueryBuilders.boolQuery()
        .filter(QueryBuilders.termQuery("field", "value"))
        // .must(QueryBuilders.termQuery("field", "value"))
            );
}
```

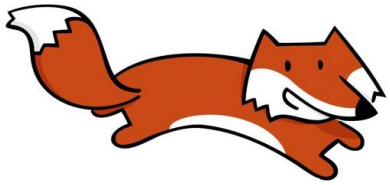
Zero result aggregations



```
private SearchRequestBuilder
        getAggsOnly(TransportClient client) {

    return client.prepareSearch("schemaful").setTypes("item")
            .addAggregation(AggregationBuilders.terms("agg1")
                                               .field("fieldName1"))
            .addAggregation(AggregationBuilders.terms("agg2")
                                               .field("fieldName2"))
            .addAggregation(AggregationBuilders.terms("agg3")
                                               .field("fieldName3"))
            .setQuery(QueryBuilders.matchAllQuery())
            .setSize(0);
}
```

Multi search



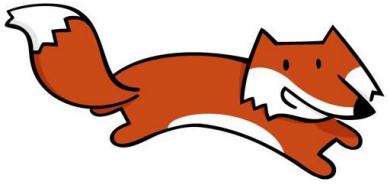
```
TransportClient client = new PreBuiltTransportClient(Settings.EMPTY)
        .addTransportAddress(new TransportAddress(
                        InetAddress.getByName("localhost"), 9300));

MultiSearchRequestBuilder multiSearchRequest = client
    .prepareMultiSearch()
        .add(getLimitedWithTimeout(client))
        .add(getAggsOnly(client))
        .add(getFiltered(client))
        .add(getProjected(client));

MultiSearchResponse response = multiSearchRequest.get();

for (MultiSearchResponse.Item item : response.getResponses()) {
    System.out.println(item.getResponse());
}
```

Routings



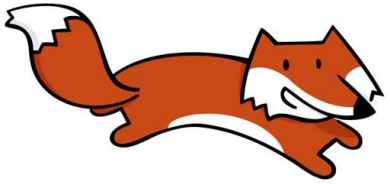
Операции поиска с использованием routing –
2 шарды по 50k записей
Total hits: 50000

Операции поиска без routing
Total hits: 100000

Эффект: поиск на 50%+ быстрее

Поиск только по id становится недоступен,
параметр routing обязателен.

A minute for summary



Правильный mapping: ускорение индексации

Timeout: максимальное время выполнения

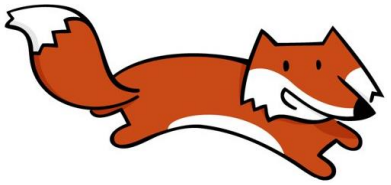
Сокращение накладных расходов:

- Projection
- Aggregations
- Filter vs must
- Multisearch

Coordinating nodes: повышение устойчивости к нагрузке

Routing: точечный поиск

A minute for docs



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

2. https://www.elastic.co/blog

3. https://github.com/MikhailKavaliou/miniqEsPerformanceOfMyDreamExamples

4. RESTER REST Plugin:
    1. https://addons.mozilla.org/en-US/firefox/addon/rester/
    2. https://chrome.google.com/webstore/detail/rester/eejfoncpjfgmeleakejdcanedmefagga?hl=en

A minute for questions???