uberMetrics

Start

Produkt

Chat

uberMetrics

Sprint 4

Amine Mastouri, Branimir Galic, Malte Müller, Rodney Mitchell, Alan Meyer, Jakob Mausbach

Sprintplan in Absprache mit uberMetrics

Sprint 2: Lasttester Alpha, Theorie begonnen

→ mindestens: Daten schreiben und updaten, Dummy-Queries stellen (evtl. vereinfacht), Daten gleichverteilt

Sprint 3: Lasttester Release, Theorie ausgearbeitet und aufbereitet

→ Ausarbeitungen an uber-Entwickler, muss konkret von Nutzen sein beim Konfigurieren der uber-Cluster

Sprint 4: Optimieren

→ Queries: Dokumenten- und Indexstruktur, sowie Umsetzung der Queries mit Elastic Search optimieren.

Sprint 4

Sprint 4: Optimieren und fertigstellen

Ausgearbeitete Theorie mit Quellenangaben ✓

Dokumentation zum Lasttester ✓

Lasttest mit Ergebnissen beschreiben ✓

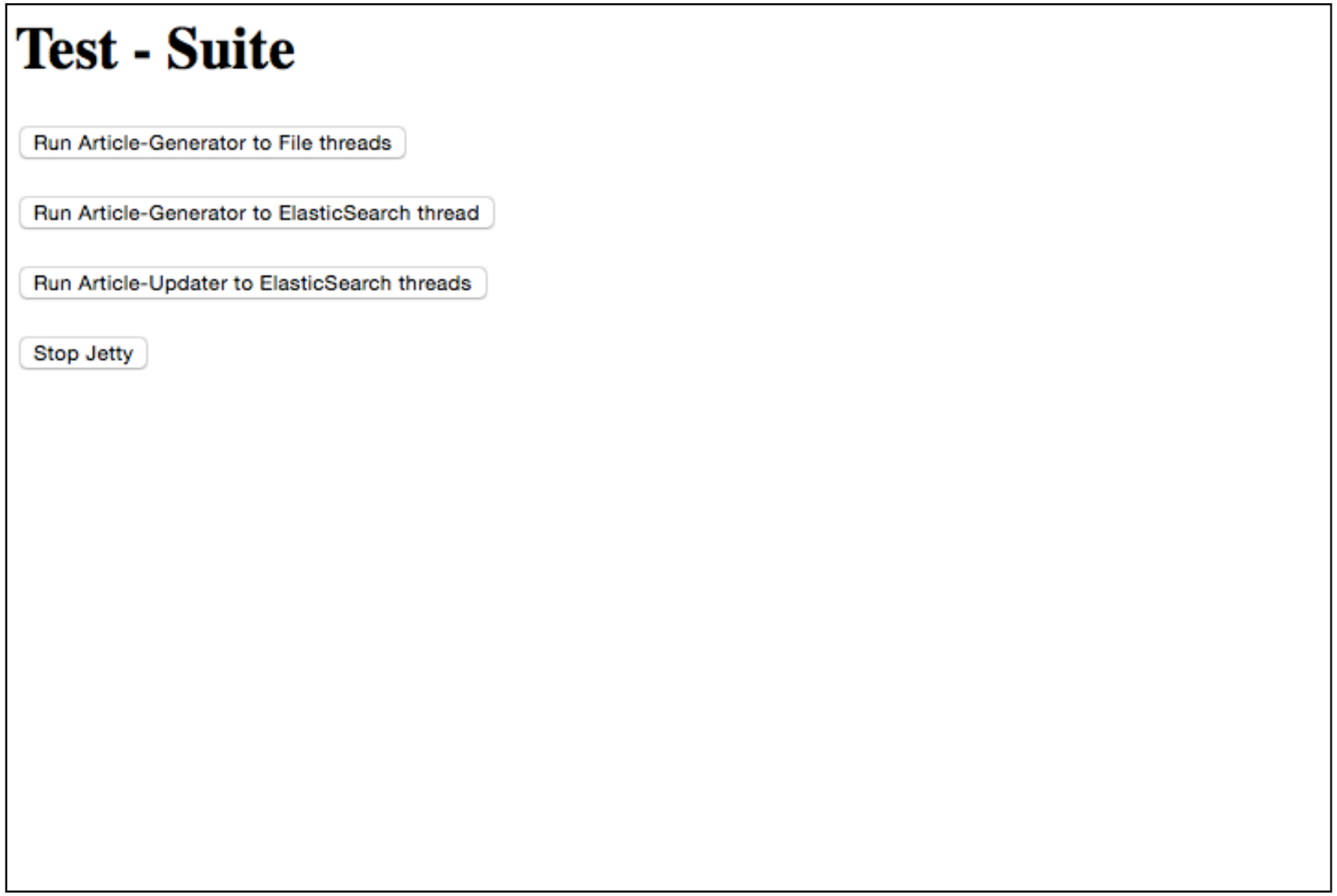
Queries optimieren ✓

uberMetrics

Ergebnisse

Eigene Web-GUI

Bedienungsoberfläche auf Basis von JavaServerPages, zugänglich über den Port 8080

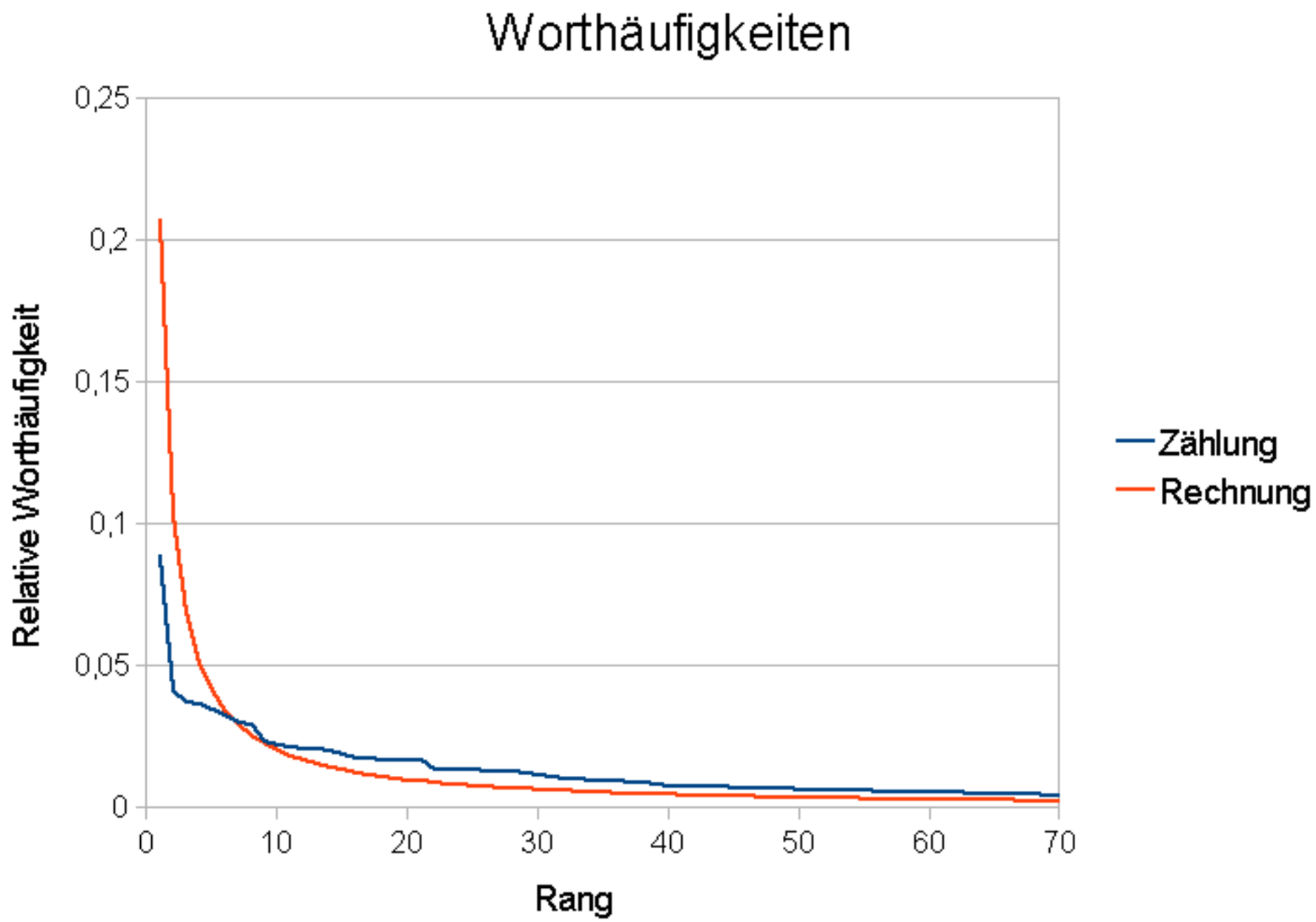


Softwarekonfiguration über config-Datei

Sprache und Anzahl der zu generierenden Daten

```
#indexingClient Properties
numberOfLoop=10000000
language=DE
#   deutsch = DE
#   englisch = EN
```

Datengenerator für Artikel mit korrekter Zipf'scher Verteilung



Threaded Generierung von Artikel in .json-Dateien

Neue Datei ab selbst definierte Dateigröße

```
Generated thread135000.json... duration : 17569 ms
Generated thread437000.json... duration : 17635 ms
Generated thread336000.json... duration : 17649 ms
Generated thread237000.json... duration : 17665 ms
Generated thread136000.json... duration : 17754 ms
Generated thread438000.json... duration : 17840 ms
Generated thread238000.json... duration : 17872 ms
Generated thread337000.json... duration : 17893 ms
Generated thread137000.json... duration : 17990 ms
Generated thread439000.json... duration : 18076 ms
Generated thread239000.json... duration : 18104 ms
Generated thread338000.json... duration : 18174 ms
Generated thread138000.json... duration : 18203 ms
```

Threaded Generierung von Artikel in ElasticSearch

Zugang via Marvel - Sense

```
GET _search
{
  "query": {
    "match_all": {}
  }
}

POST /uebermetric/article
{

}

DELETE /uebermetric/article/
{

}

GET /uebermetric/article/_search
{
  "query": {
    "match_all": {}
  }
}

GET /_cluster/nodes?pretty=true

GET /uebermetric/article/_search
{
  "query": {
    "match": {
      "data.resultId": 500
    }
  }
}
```


```
            }
          ]
        }
      },
      {
        "_index": "uebermetric",
        "_type": "article",
        "_id": "AUpOCUxIXEiImU8j3tHv",
        "_score": 1,
        "_source": {
          "data": [
            {
              "accountId": 3,
              "searchId": 3,
              "resultId": 9272,
              "domainId": 2,
              "published": "Fri Jul 30 23:29:45 CEST 2010",
              "url": "http://drei-geschmack.de/und/fr/die/der.html",
              "title": "Unterwerken sorgeberechtigter Runddraht Daumenrichtungen durchdürften ",
              "description": "Schusswunden Haferkerne Standardvordrucken Schweißprozess Jackpot Zollformalität Siegesaltar Wettkampfdisziplin hergetrieben Ausbildungsformats Flurabstand Scheitelpunkte romanisierter Platinfamilien Malversation Seegeschützfeuers Zeilenbeginn Militäruntersuchung Brandspore Bundesratsbeamtin Taschengeldzahlung Kirchfassade Kriegsgemetzel Binnentransportaufkommens Datenausgabeleitungen Verkehrsmeldungen Ladegeräte geholfen Geräteteilen lippenblütlerartigem ",
              "content": "warum übrig heiß sei für sich Masse Haltung Schauspieler wieviel dringend ein das als indes erstenmal der Umstand laufen damit die Sie der der einem Kinder Wachstum den des ",
              "category": 2,
              "sentiment": 0,
              "languageCode": 3,
              "mediaSegment": 6,
              "isRead": false,
              "isHidden": false,
              "inlinkCount": 0
            }
          ]
        }
      },
      {
        "_index": "uebermetric",
        "_type": "article",
        "_id": "AUpOCUxXXEiImU8j3tHz",
        "_score": 1,
        "_source": {
          "data": [
```

Threaded updaten von zufälligen Artikeln in der gestarteten Elastic Search Instanz

Zugang via Marvel - Sense

```
1  GET _search
2  {
3    "query": {
4      "match_all": {}
5    }
6  }
7
8  POST /uebermetric/article
9  {
10
11 }
12
13 DELETE /uebermetric/article/
14 {
15
16 }
17
18 GET /uebermetric/article/_search
19 {
20   "query": {
21     "match_all": {}
22   }
23 }
24
25 GET /_cluster/nodes?pretty=true
26
27 GET /uebermetric/article/_search
28 {
29   "query": {
30     "match": {
31       "data.resultId": 500
32     }
33   }
34 }
```

```
94          ]
95        }
96      },
97      {
98        "_index": "uebermetric",
99        "_type": "article",
100       "_id": "AUpOCUxIXEiImU8j3tHv",
101       "_score": 1,
102       "_source": {
103         "data": [
104           {
105             "accountId": 3,
106             "searchId": 3,
107             "resultId": 9272,
108             "domainId": 2,
109             "published": "Fri Jul 30 23:29:45 CEST 2010",
110             "url": "http://drei-geschmack.de/und/fr/die/der.html",
111             "title": "Unterwerken sorgeberechtigter Runddraht Daumenrichtungen durchdürften ",
112             "description": "Schusswunden Haferkerne Standardvordrucken Schweißprozess Jackpot Zollformalität
Siegesaltar Wettkampfdisziplin hergetrieben Ausbildungsformats Flurabstand Scheitelpunkte romanisierter Platinfamilien
Malversation Seegeschützfeuers Zeilenbeginn Militäruntersuchung Brandspore Bundesratsbeamtin Taschengeldzahlung Kirchfassa
de Kriegsgemetzel Binnentransportaufkommens Datenausgabeleitungen Verkehrsmeldungen Ladegeräte geholfen Geräteteilen
lippenblütlerartigem ",
113             "content": "warum übrig heiß sei für sich Masse Haltung Schauspieler wieviel dringend ein das als
indes erstenmal der Umstand laufen damit die Sie der der einem Kinder Wachstum den des ",
114             "category": 2,
115             "sentiment": 0,
116             "languageCode": 3,
117             "mediaSegment": 6,
118             "isRead": false,
119             "isHidden": false,
120             "inlinkCount": 0
121           }
122         ]
123       }
124     },
125     {
126       "_index": "uebermetric",
127       "_type": "article",
128       "_id": "AUpOCUxXXEiImU8j3tHz",
129       "_score": 1,
130       "_source": {
131         "data": [
```

Softwaredokumentation / Bedienungsanleitung

BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

uberMetrics

Software-Dokumentation

Im Modul Enterprise Datamanagement
des Studienganges Medieninformatik Master
der Beuth Hochschule für Technik Berlin

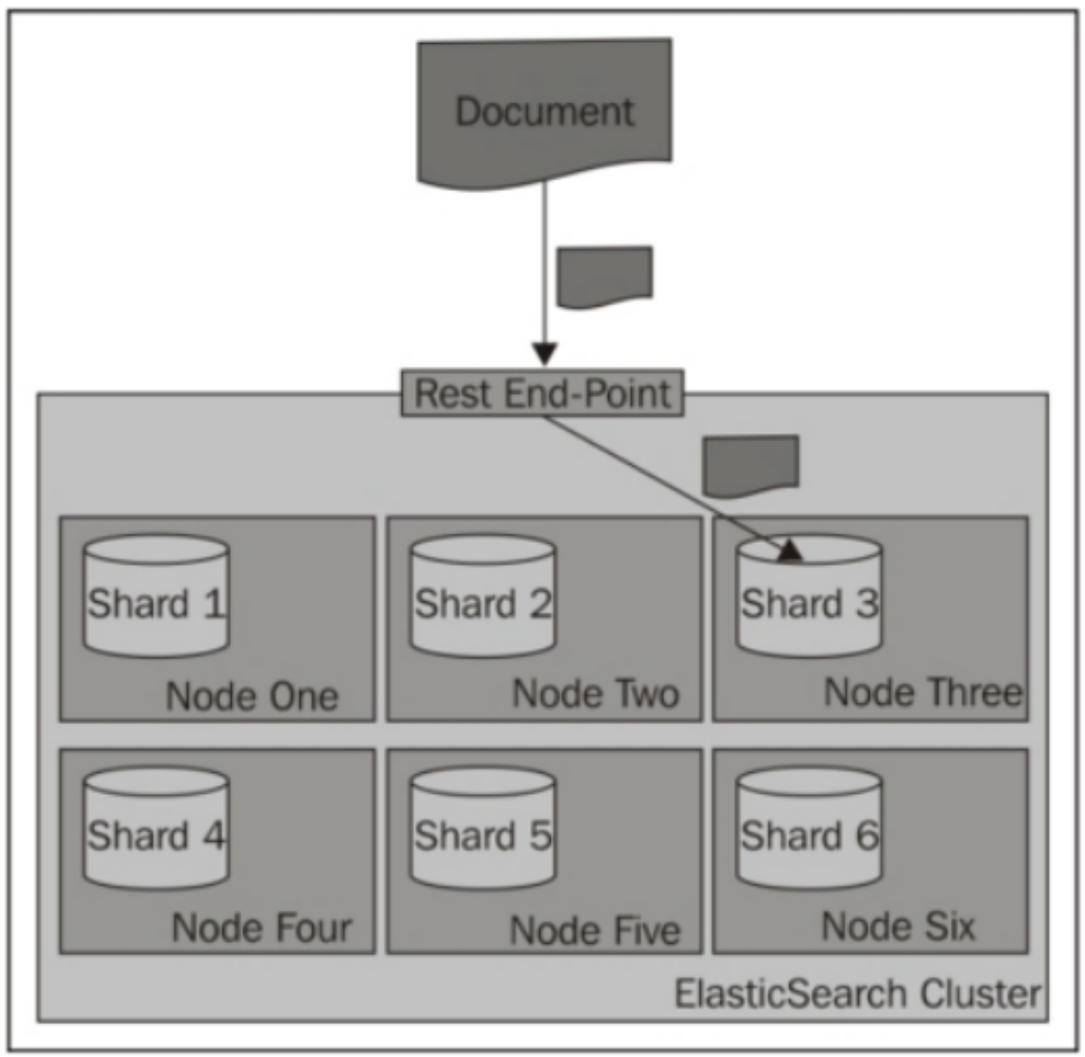
Vorgelegt von Malte Müller, Branimir Galic, Rodney Mitchell, Amine Mastouri, Alan Meyer und Jakob Mausbach

Wintersemester 2014

Prof. Dr. Alexander Löser

Ausgearbeitete Theorie

Indexing



Allgemeines

- während einer Index Operation wird von ElasticSearch auf den benötigten werden soll
- ElasticSearch bildet daraufhin einen Hashwert und legt das Dokument in e

Kibana Dashboard

Eigenes Kibana Dashboard in dem die Ergebnisse des Lasttests visualisiert werden können

- Events in einer Zeitspanne
- Durchschnittsgeschwindigkeit der Queries
- Minimum Maximum Response der Queries

http://148.251.140.15:9200/_plugin/kibana/#/dashboard/elasticsearch/Logstash%20Search

Marvel Dashboard

Marvel Dashboard gibt detaillierte Informationen über die einzelnen Nodes wieder

- Echtzeitanalyse von Metriken
- Cluster-Status als Übersicht
- Visualisierung der Cluster-Events

http://148.251.140.15:9200/_plugin/marvel/kibana/index.html#/dashboard/file/marvel.overview.json

Kopf Dashboard

Kopf Dashboard ermöglicht Zugang zu zahlreichen Funktionen der elasticsearch API über Web-GUI

Gewährt Überblick über alle:

Cluster, Indizies, und Shards

http://148.251.140.15:9200/_plugin/kopf/#!/cluster

Query 2

```
SELECT Date(Published), COUNT(*) FROM Results

WHERE AccountId = 157

AND SearchID IN (190,...)

AND Published BETWEEN ... AND ...

GROUP BY Date
```

Query Optimierung - Datum

```
"aggs": {
    "Dates": {
        "terms": {
            "field" : "published"
        }
    }
  },
```

```
"aggs": {
    "Dates": {
        "date_histogram": {
            "field" : "published",
            "interval" : "day",
            "format" : "dd.MM.yyyy"
        }
    }
  },
```

Query 1

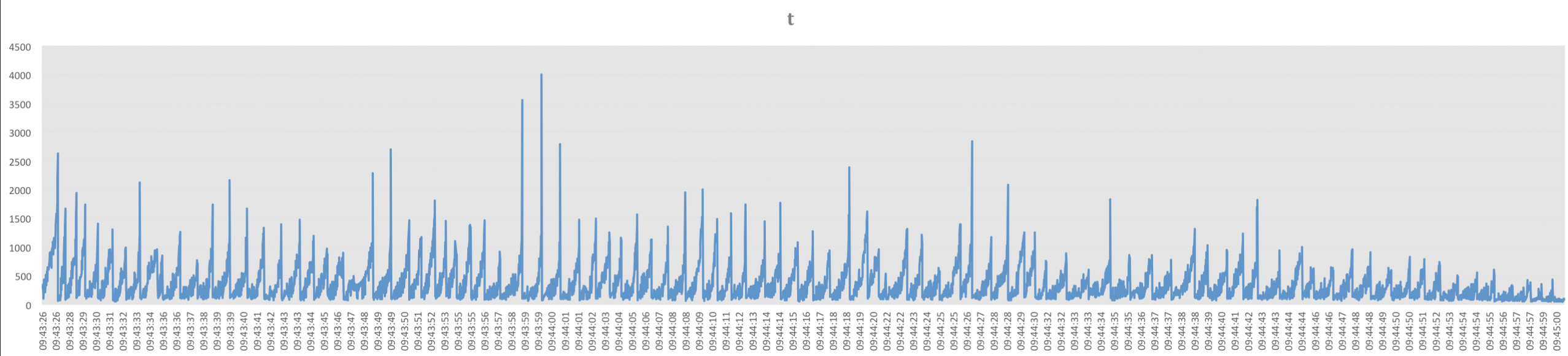
```
curl -XGET 'http://localhost:9200/ubermetrics/results/_search?pretty' -d '{
  "size" : 100,
  "sort" : [{"published" : {"order" : "desc"}}],
  "filter": {
    "and": [
      {
        "range": {
          "date": {
            "gt": "2008-01-01",
            "lt": "2014-01-01"
          }
        }
      },
      {
        "term": {
          "accountid": 2
        }
      },
      {
        "terms": {
          "searchid": [
            32
          ]
        }
      }
    ]
  }
}
```

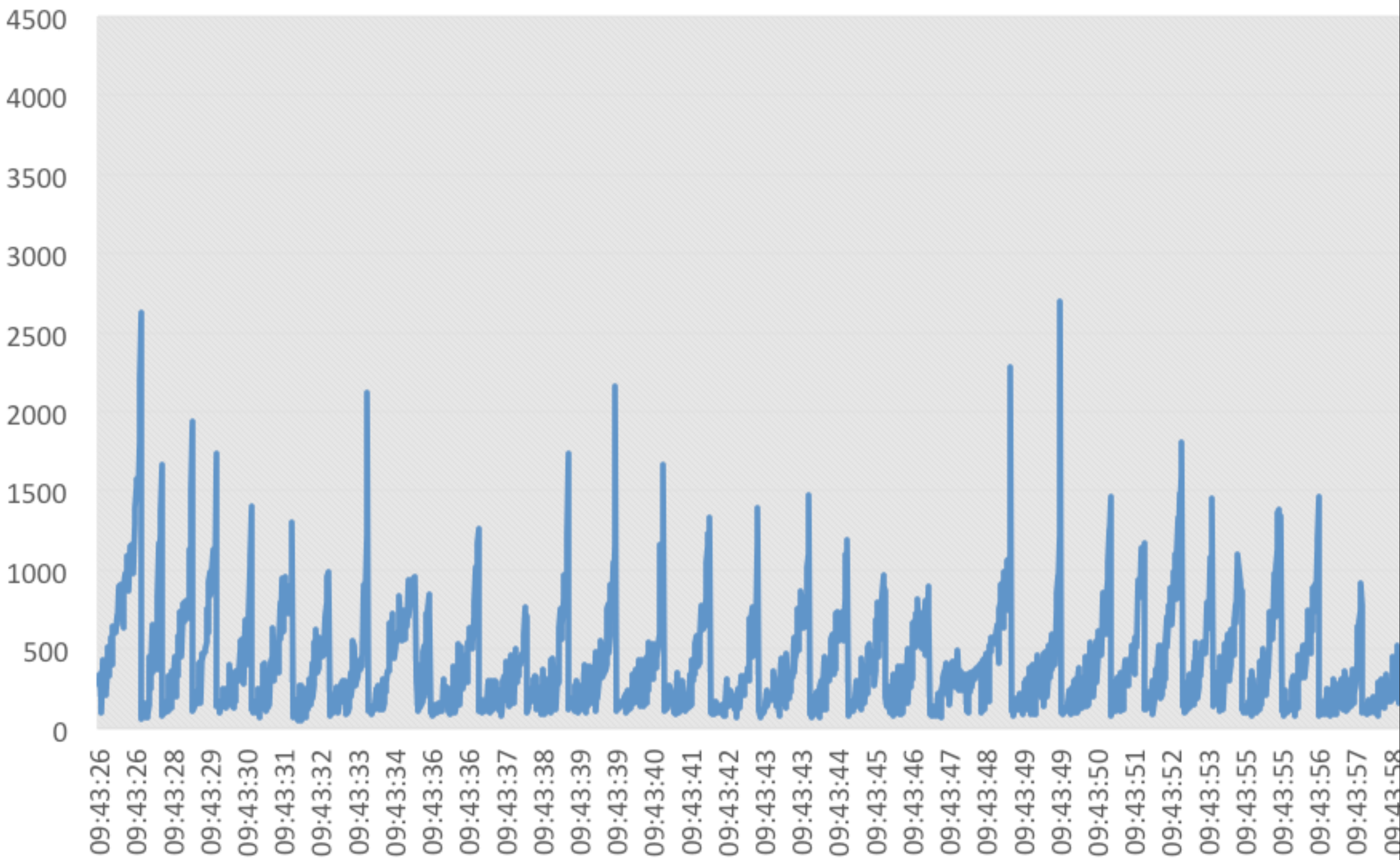
Query 2

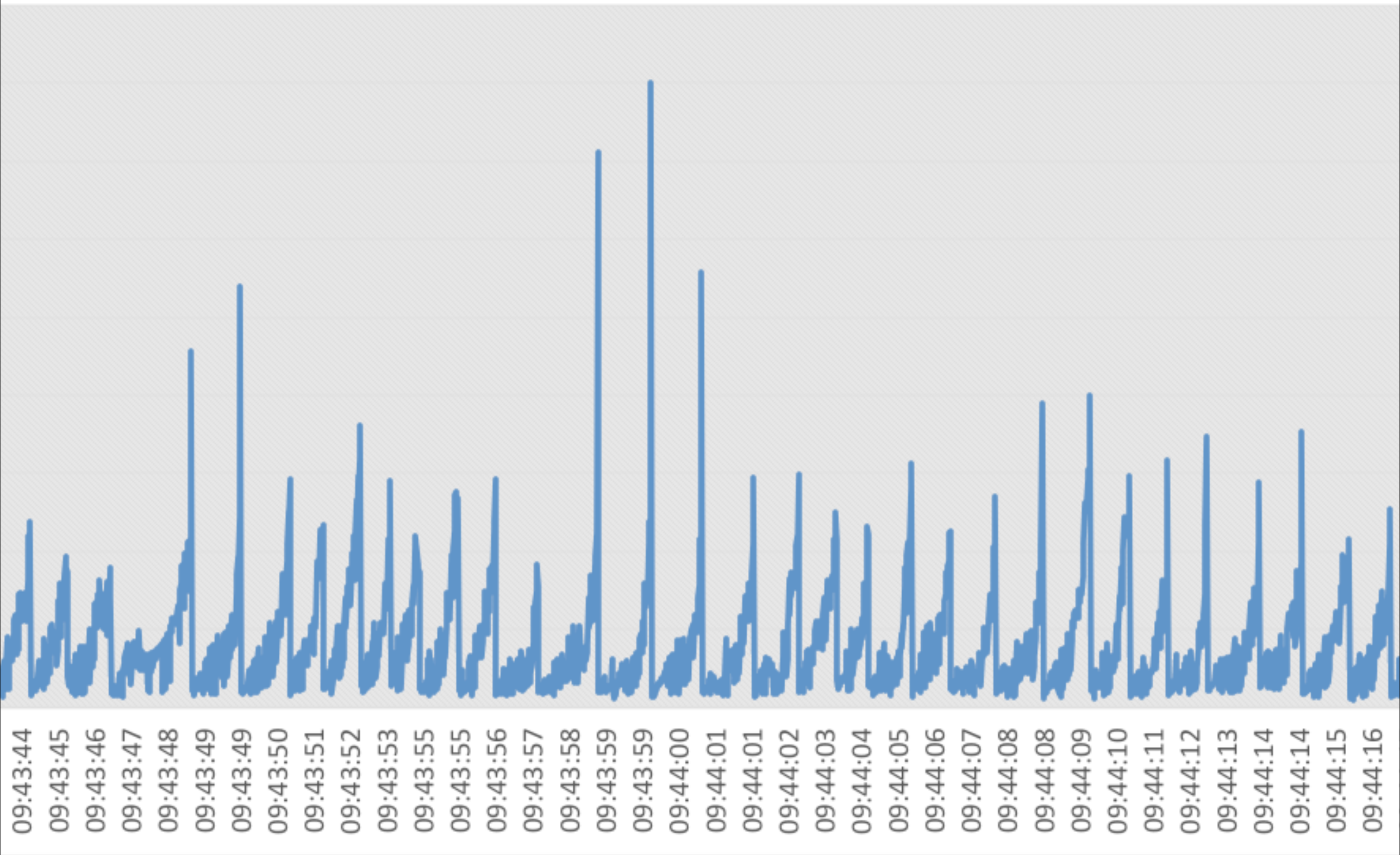
```
curl -XGET 'http://localhost:9200/ubermetrics/results/_search?search_type=count' -d '{
    "aggs": {
        "dates": {
            "date_histogram": {
                "field" : "published",
                "interval" : "day",
                "format" : "dd.MM.yyyy"
            }
        }
    },
    "filter": {
        "and": [
        {
            "range": {
                "published": {
                    "gt": "02.01.2008",
                    "lt": "02.10.2014"
                }
            }
        },
        {
            "term": {
                "accountId": 2
            }
        },
        {
            "terms": {
                "searchId": [
                    32

                ]
            }
        }
    ]
  }
}
```

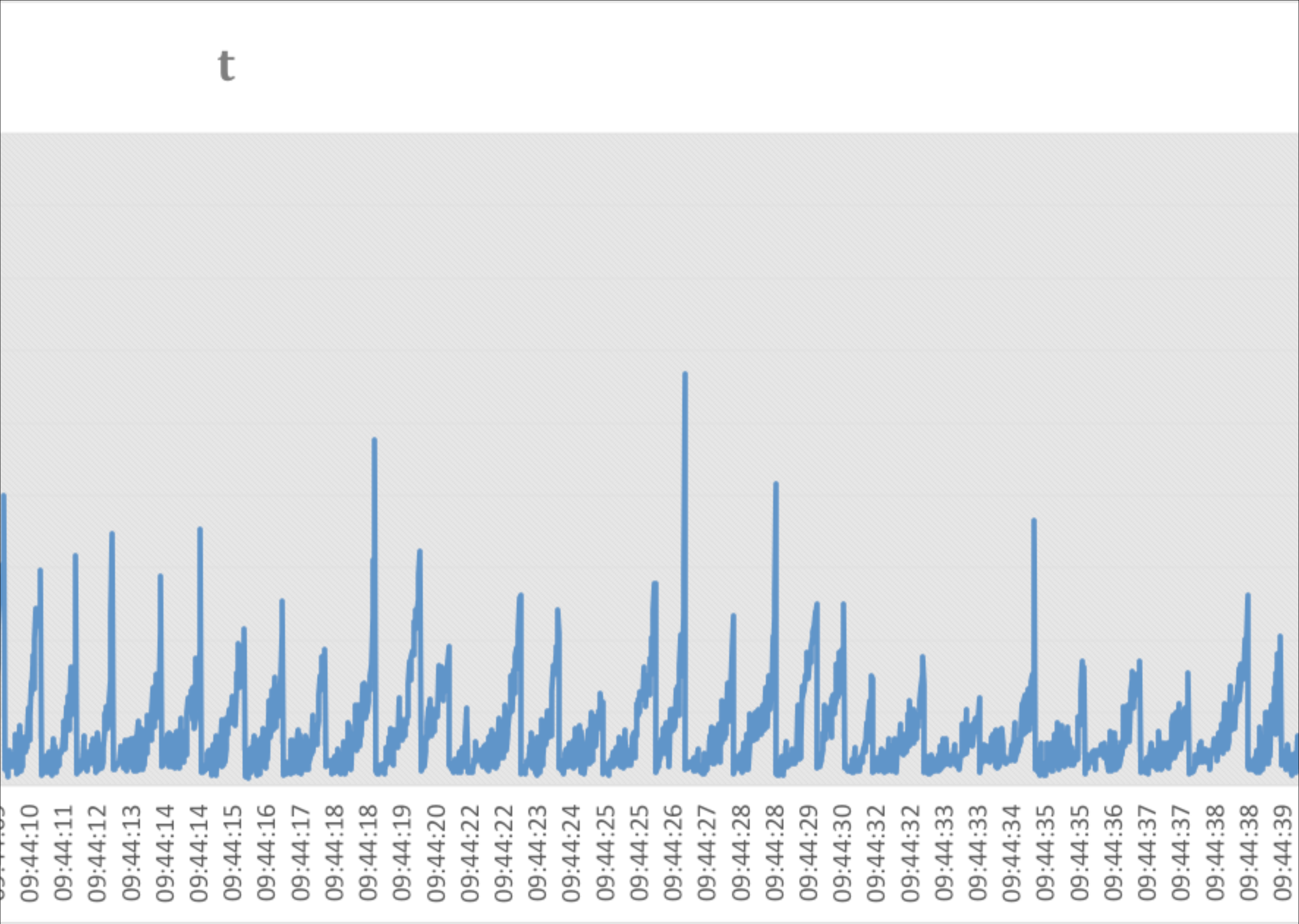
Query 2 - Ausführungsgeschwindigkeit

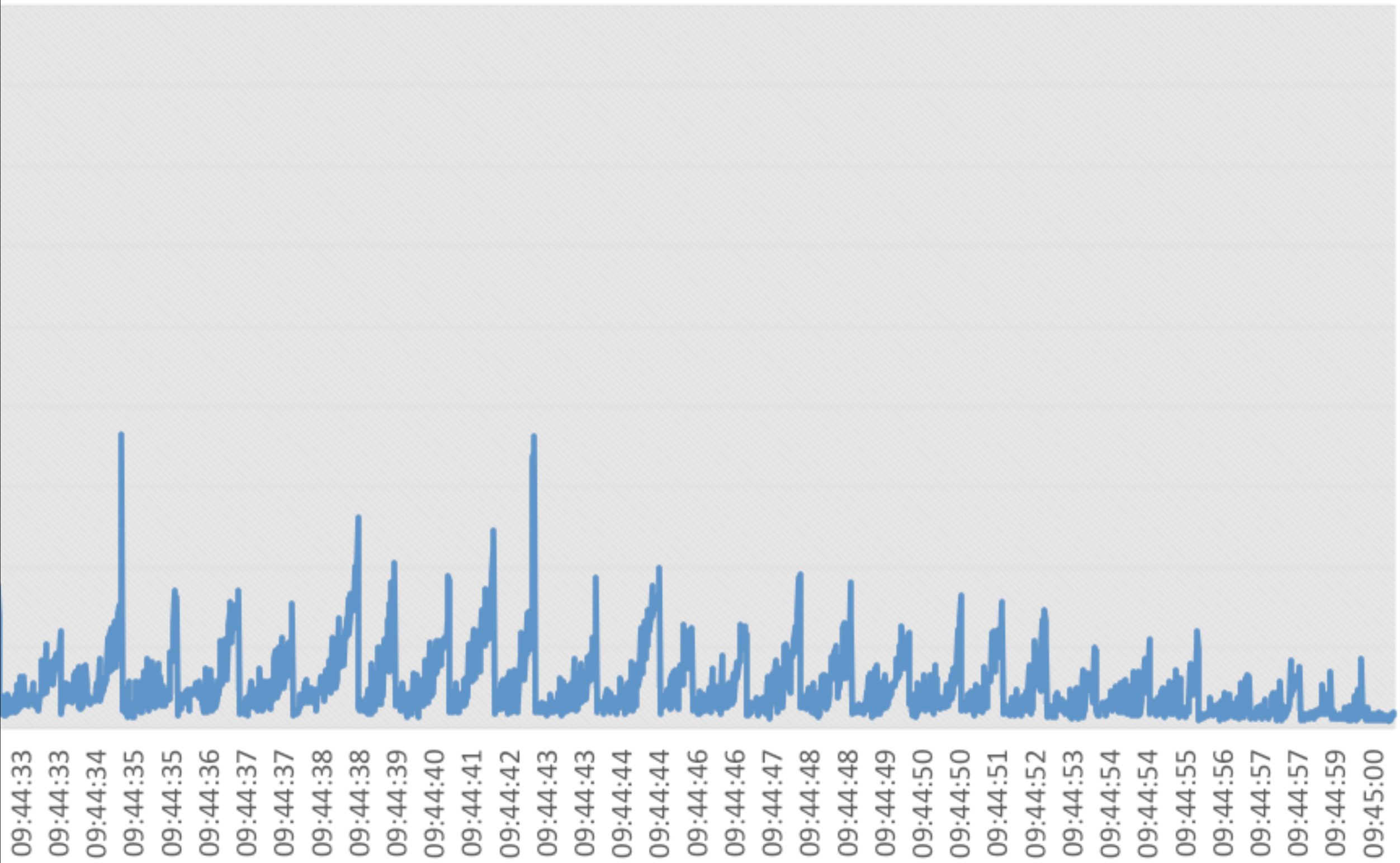
3 Nodes
3 Primary Shards
6 Replica Nodes
50 Threads
100 Wiederholungen
10 mio. Dokumente




4500
4000
3500
3000
2500
2000
1500
1000
500
0
09:43:26
09:43:26
09:43:28
09:43:29
09:43:30
09:43:31
09:43:32
09:43:33
09:43:34
09:43:36
09:43:36
09:43:37
09:43:38
09:43:39
09:43:39
09:43:40
09:43:41
09:43:42
09:43:43
09:43:43
09:43:44
09:43:45
09:43:46
09:43:47
09:43:48
09:43:49
09:43:49
09:43:50
09:43:51
09:43:52
09:43:53
09:43:55
09:43:55
09:43:56
09:43:57


t
09:43:44
09:43:45
09:43:46
09:43:47
09:43:48
09:43:49
09:43:49
09:43:50
09:43:51
09:43:52
09:43:53
09:43:55
09:43:55
09:43:56
09:43:57
09:43:58
09:43:59
09:43:59
09:44:00
09:44:01
09:44:01
09:44:02
09:44:03
09:44:04
09:44:05
09:44:06
09:44:07
09:44:08
09:44:08
09:44:09
09:44:10
09:44:11
09:44:12
09:44:13
09:44:14
09:44:14
09:44:15
09:44:16


t
09:44:10
09:44:11
09:44:12
09:44:13
09:44:14
09:44:14
09:44:15
09:44:16
09:44:17
09:44:18
09:44:18
09:44:19
09:44:20
09:44:22
09:44:22
09:44:23
09:44:24
09:44:25
09:44:25
09:44:26
09:44:27
09:44:28
09:44:28
09:44:29
09:44:30
09:44:32
09:44:32
09:44:33
09:44:33
09:44:34
09:44:35
09:44:35
09:44:36
09:44:37
09:44:37
09:44:38
09:44:38
09:44:39


09:44:33
09:44:33
09:44:34
09:44:35
09:44:35
09:44:36
09:44:37
09:44:37
09:44:38
09:44:38
09:44:39
09:44:40
09:44:41
09:44:42
09:44:43
09:44:43
09:44:44
09:44:44
09:44:46
09:44:46
09:44:47
09:44:48
09:44:48
09:44:49
09:44:50
09:44:50
09:44:51
09:44:52
09:44:53
09:44:54
09:44:54
09:44:55
09:44:56
09:44:57
09:44:57
09:44:59
09:45:00

Query 1 - Ausführungsgeschwindigkeit

3 Nodes
3 Primary Shards
6 Replica Nodes
50 Threads
100 Wiederholungen
10 mio. Dokumente

Min. Wert: 47 ms
Max. Wert: 224 ms
Mittelwert: 143 ms

Mögliches Vorgehen für zukünftiges Team

Texte aus anderen Quellen zum Testen mit einbeziehen: Projekt Gutenberg, Korpora, Wikipedia....

Erweitern der config.properties Datei um weitere Attribute

Erweiterung des ArticleUpdater Threads

Lessons learned

Vorzeitig ausgiebig mit der Theorie von Elasticsearch beschäftigen, da diese sehr umfangreich ist

Übersetzen von Queries ist nicht trivial

Viele Optimierungsmöglichkeiten und Einstellungen --> Für Einsteiger sehr komplex & unüberschaubar

Arbeiten in großen Gruppen erfordert viel Zeit und Koordination

Fragen?