exensio
intelligente informationssysteme

Umsetzung von BI-Lösungen mit Unterstützung einer Suchmaschine

6. Workshop Open Source Business Intelligence

05.03.2015

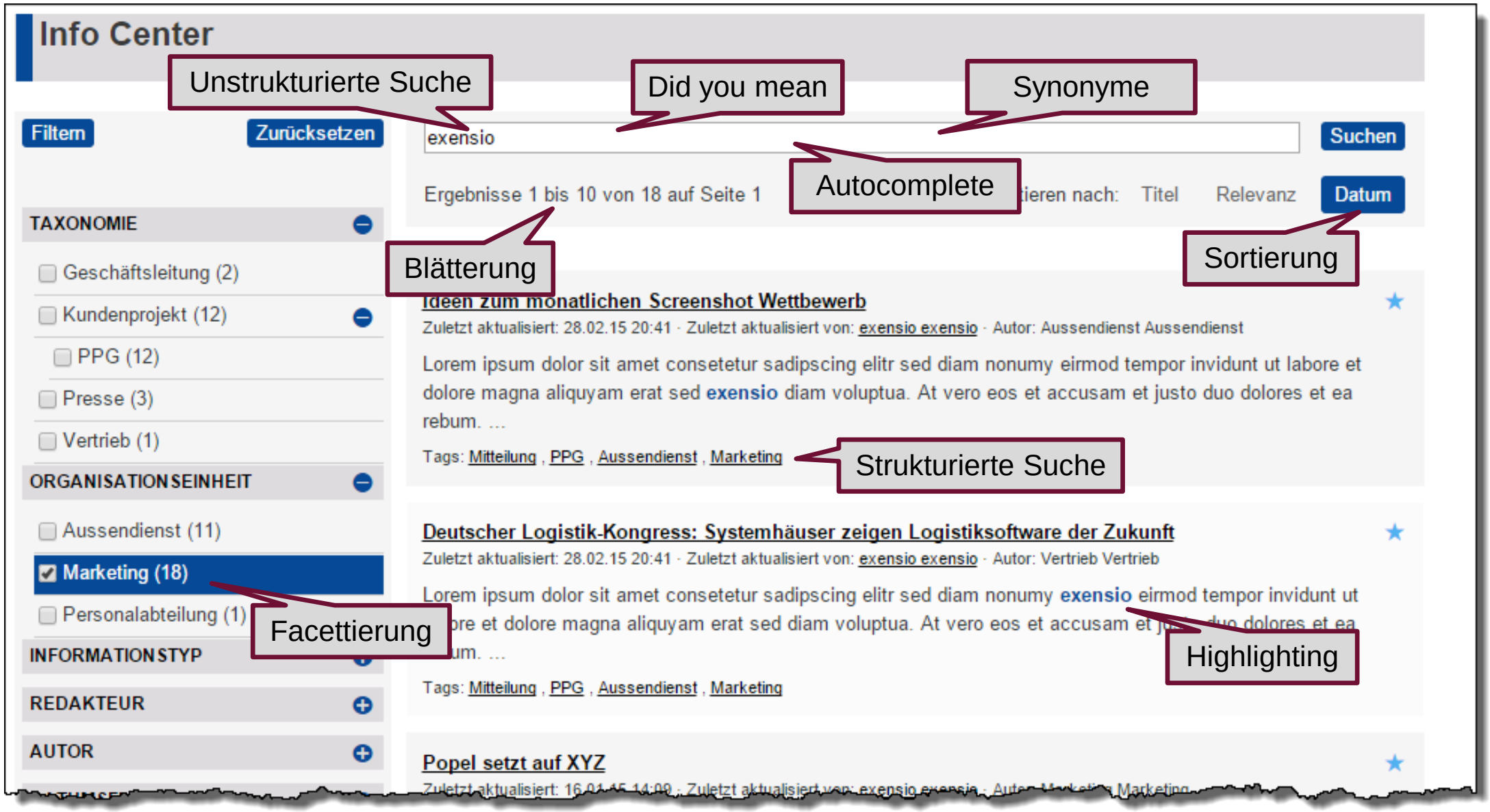
Tobias Kraft, exensio GmbH

Agenda

- Suchmaschinen
- Elasticsearch
- BI-Stack mit Elasticsearch
- Umsätze Pharma

exensio
intelligente informationssysteme

Funktionen einer Suche



exensio
intelligente informationssysteme

Für Suchmaschinen gibt es viele Einsatzszenarien

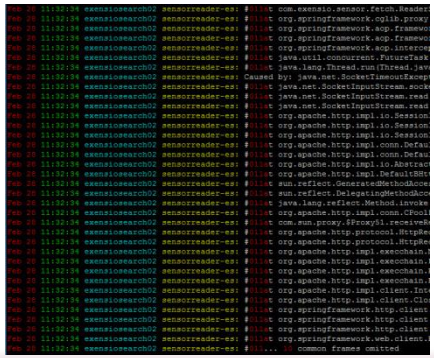
Speicher

Volltextsuche

Cache

Geo-Suche

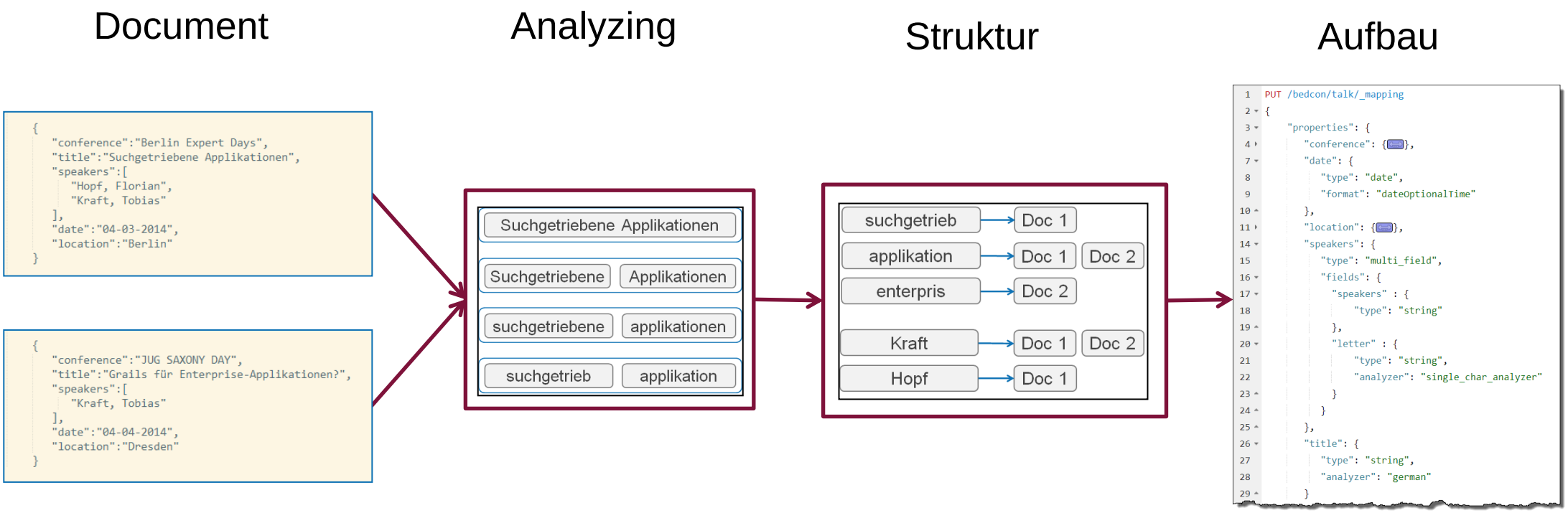
Logfile-Analyse



Analytics

exensio
intelligente informationssysteme

Speichern von Daten in einer Suchmaschine



exensio
intelligente informationssysteme

Elasticsearch im Überblick

- Suchmaschine unter Apache 2 Open Source License
- Erstes Release 2010

- Basiert auf Java
- Basiert auf Lucene

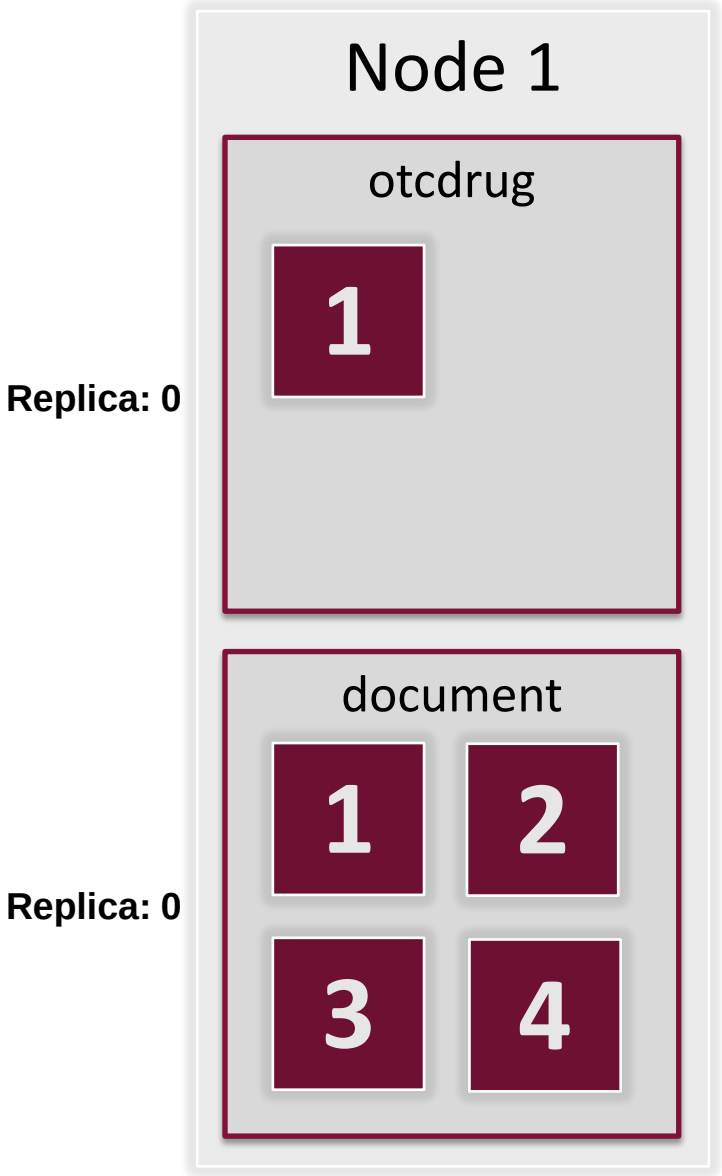
- JSON-API
- Schemalos
- Plugins

- Runterladen und loslegen
- Im Trend

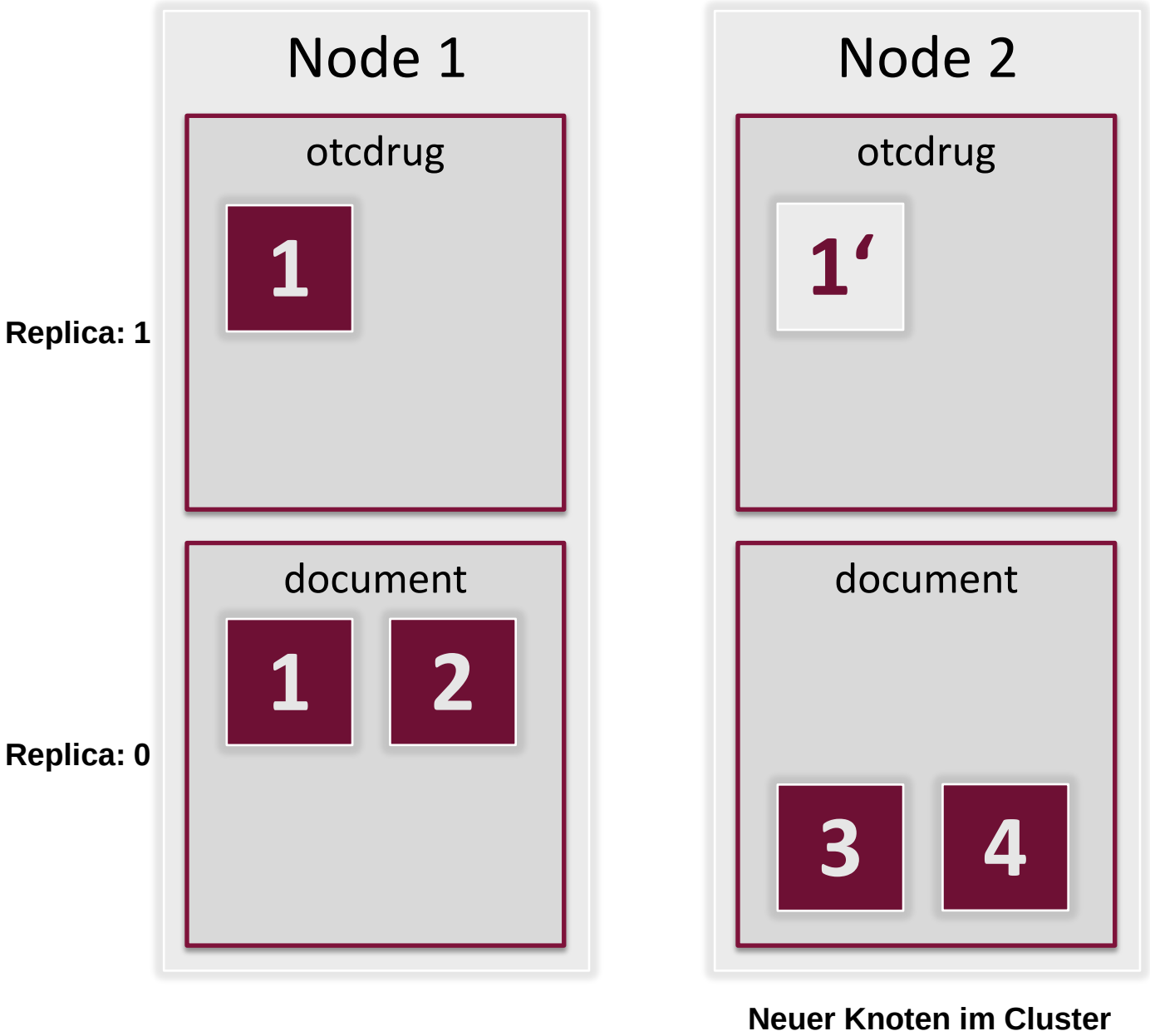
```
GE
{
}
```

```
  "hits": {
    "total": 7124,
    "max_score": 0.66425383,
    "hits": [
      {
        "_index": "factmarket",
        "_type": "jdbc",
        "_id": "AUvW12UnPHK6qcGfww5-",
        "_score": 0.66425383,
        "_source": {
          "id": 42181,
          "version": 0,
          "aut_idem": null,
          "date_created": "2013-01-16T00:52
          "gross_sales": null,
          "last_updated": "2013-01-16T00:52
          "launch_revenues0": null,
          "launch_revenues1": null,
          "market_gen_launches0": null
```

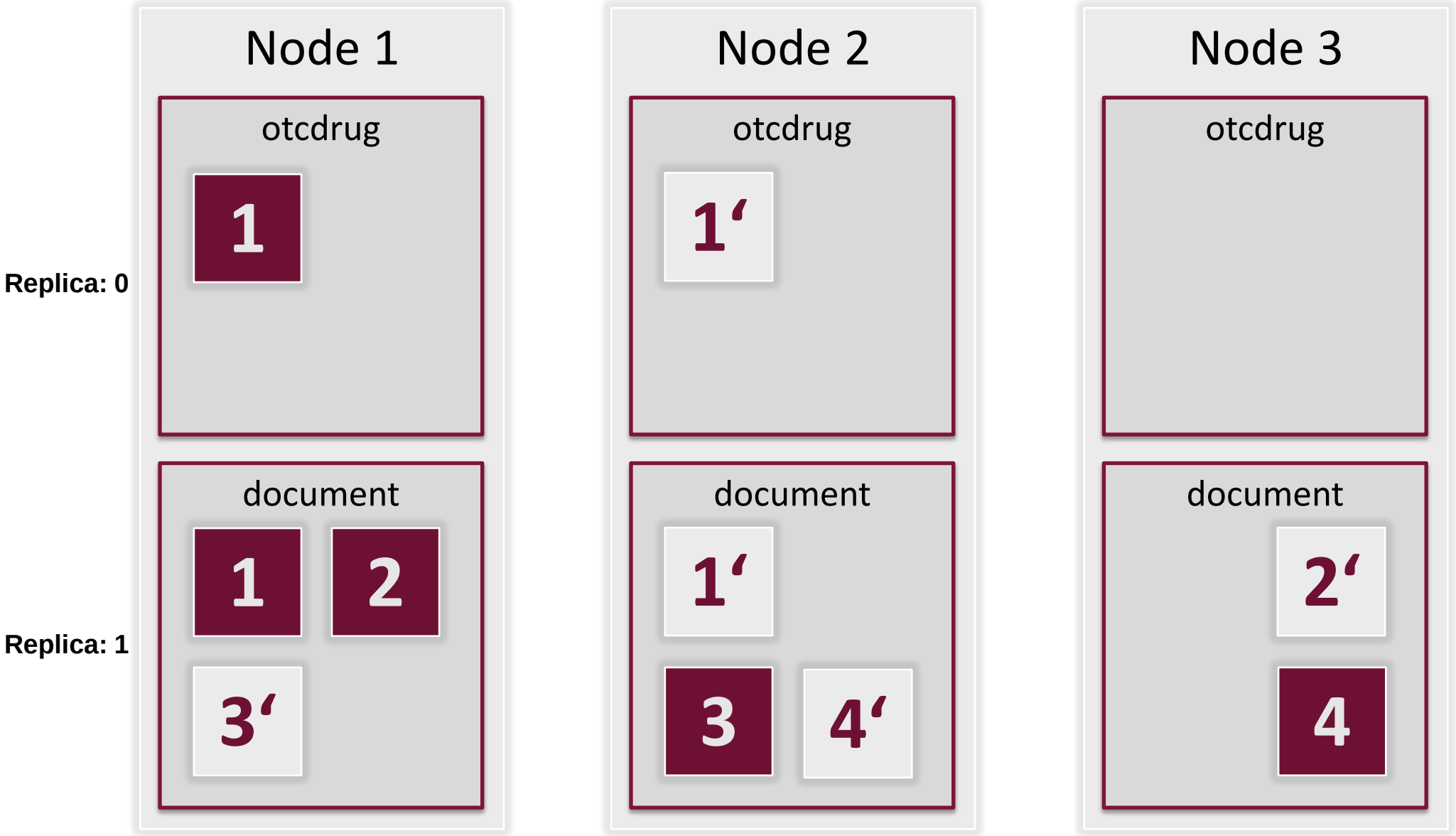
Große Datenmengen über Shards verwalten



Große Datenmengen über Shards verwalten

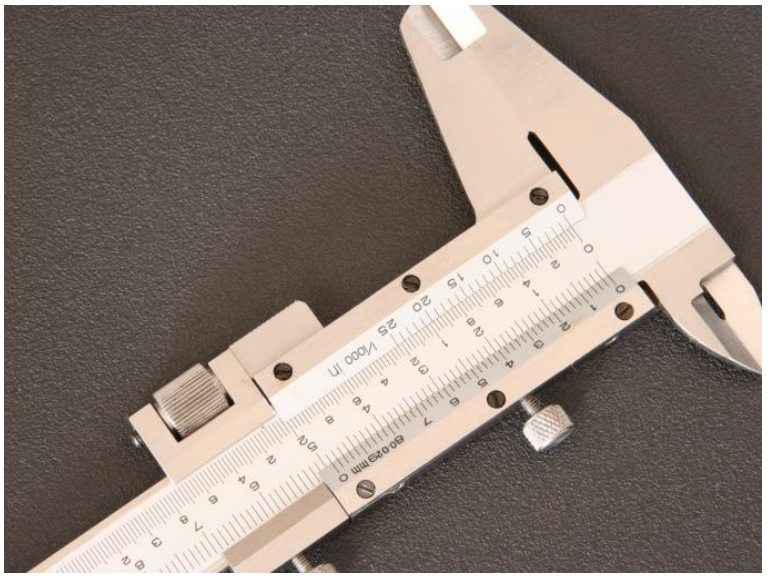


Große Datenmengen über Shards verwalten



Aggregationen - Buckets und Metrics

http://mrg.bz/lQNZFq



http://mrg.bz/Nn57cJ

Buckets

- Terme
- Ranges
- Histograme
- Geo-Distanz

Metrics

- Anzahl
- Summe
- Min / Max / Average
- Varianz
- Perzentile

Aggregationen für Analysen

```
POST /events/_search
{
  "aggs" : {
    "categories" : { "terms" : { "field" : "category" } },
    "indications" : { "terms" : { "field" : "indication" } },
    "locations" : { "terms" : { "field" : "location" } },
    "countries" : { "terms" : { "field" : "country" } },
    "created_range": {
        "date_range": {
            "field": "created",
            "format": "yyyy-MM-dd'T'HH:mm:ss.SSSZ",
            "ranges": [
                { "from": "now-1M/M" },
                { "from": "now-3M/M" },
                { "from": "now-6M/M" }
            ]
        }
    }
  }
}
```

Aggregationen

```
POST /even
{
  "aggs" :
    "categ
    "indic
    "locat
    "count
    "creat
        "d
        }
    }
  }
}
```

```
  "aggregations": {
    "locations": {
      "buckets": [
        {
          "key": "Berlin",
          "doc_count": 626
        },
        {
          "key": "Basel",
          "doc_count": 93
        },
        .....
      ]
    },
    "countries": {
      "buckets": [
        {
          "key": "Germany",
          "doc_count": 671
        },
        {
          "key": "Switzerland",
          "doc_count": 107
        },
```

exensio
intelligente informationssysteme

BI mit Elasticsearch

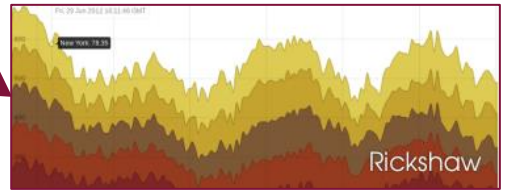
ETL | **Speicherung / Berechnungen** | **Analyse**

Logstash

rivers

Implementierungen für
- JDBC
- CSV
- …

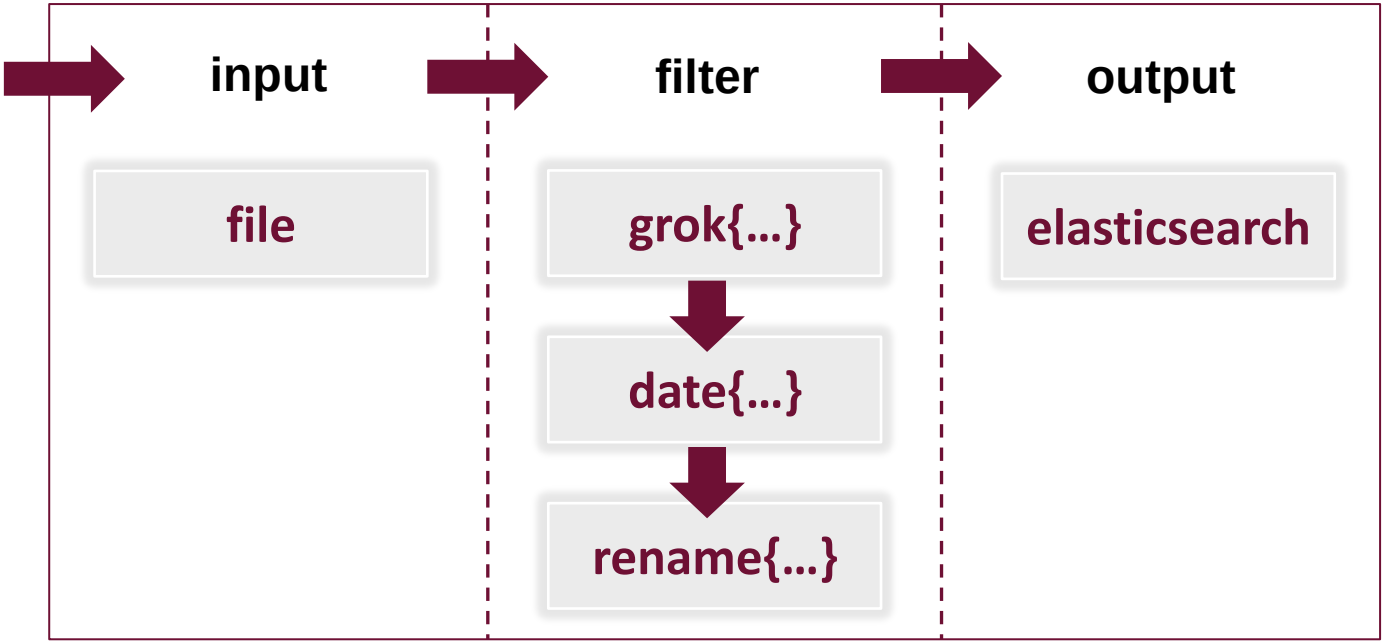
Eigene Loader mit
ES-Client (Bsp. SpringBoot)



Eigene Visualisierung

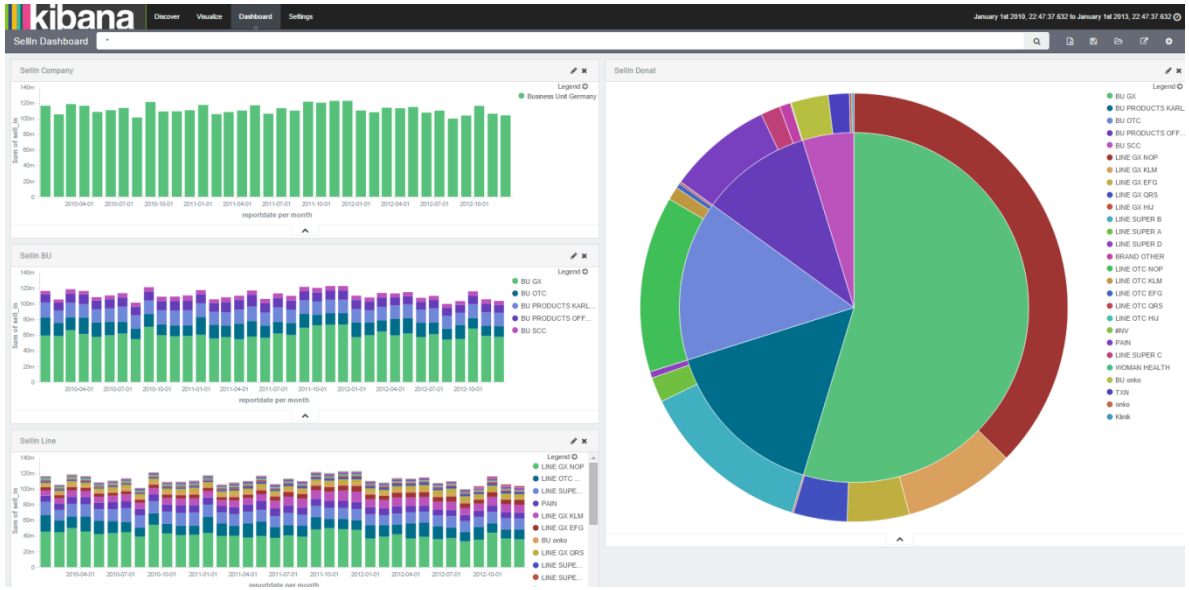
Datentransport mit Logstash

- Event Processing Engine
- Optimiert für Log-Dateien
- Pipeline-Prinzip
  - Input (50+)
  - Filter (60+)
  - Output (75+)



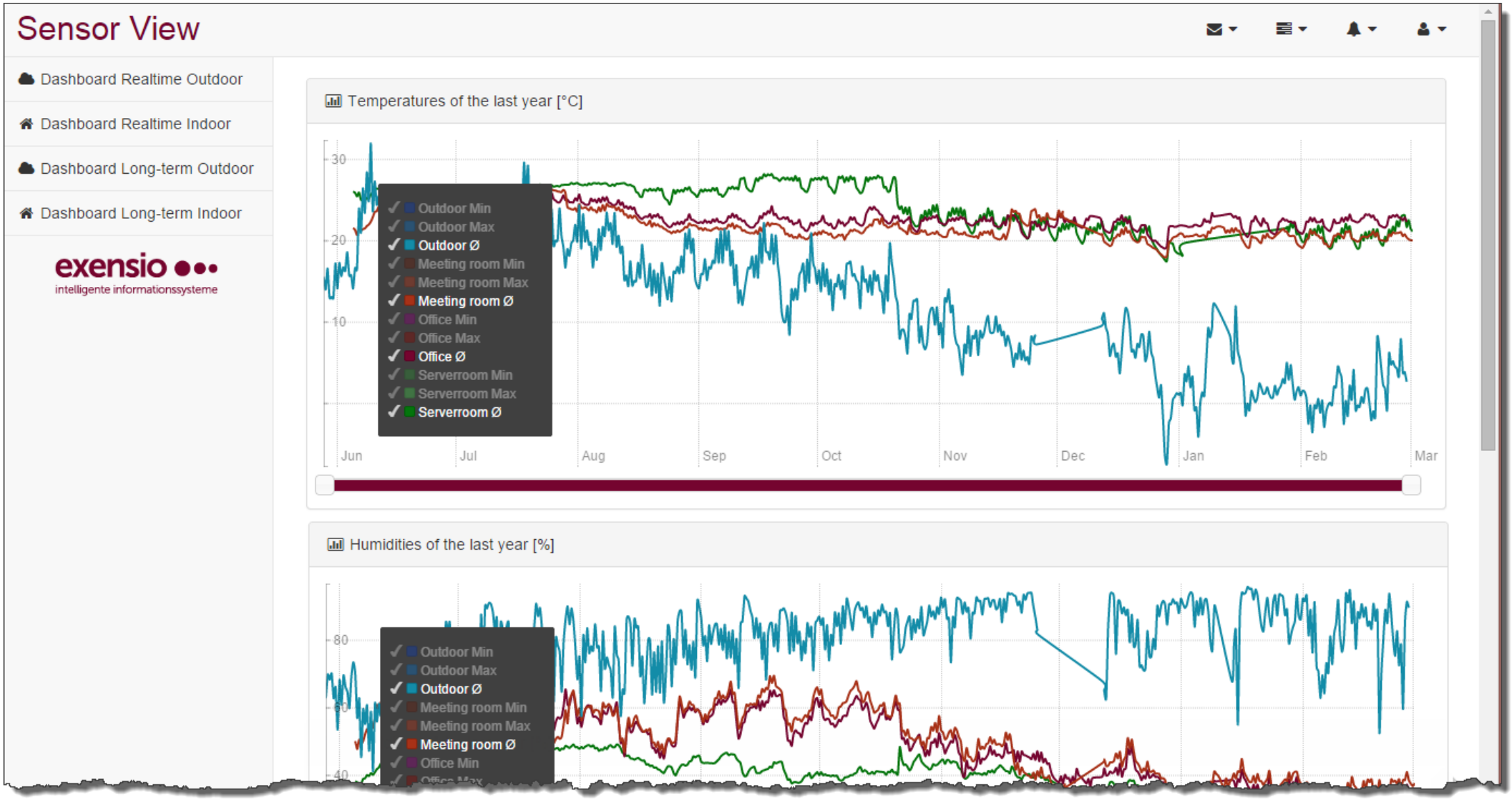
Visualisierungen mit Kibana

- Aktuelles Release: Kibana 4
- Browserbasierte Visualisierung von Daten
  - Abfragen über JSON an ES
- Aufbereitung über
  - Discover
  - Visualize
  - Dashboards



exensio
intelligente informationssysteme

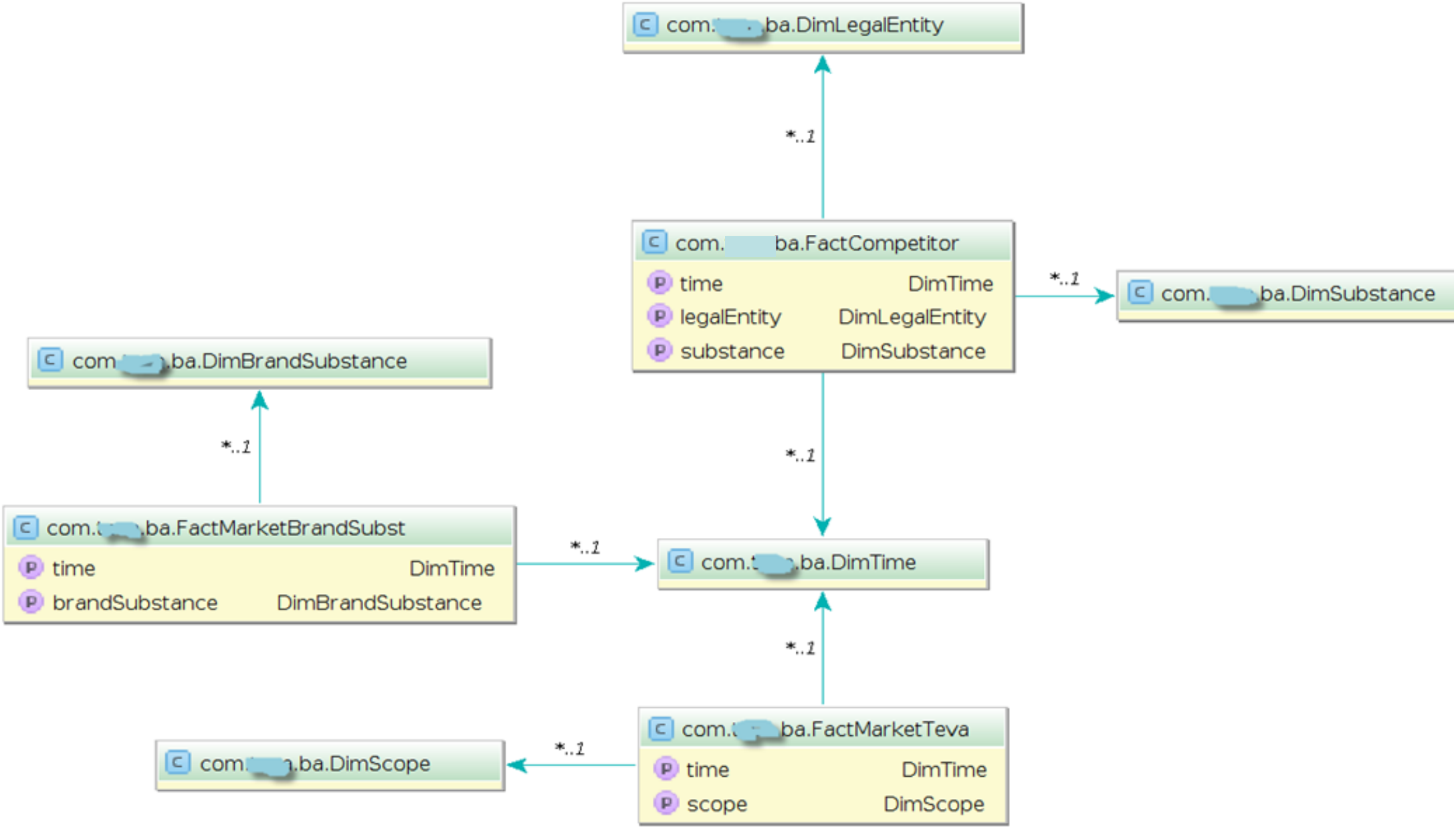
Eigene Analyse-Komponenten sind einfach zu erstellen



Darstellung der JSON-Rückgabewerte von ES auf Basis von JavaScript

exensio
intelligente informationssysteme

Marktanalysen Pharmamarkt – Starschema



exensio
intelligente informationssysteme

Marktanalysen Pharmamarkt – Suchmaschine

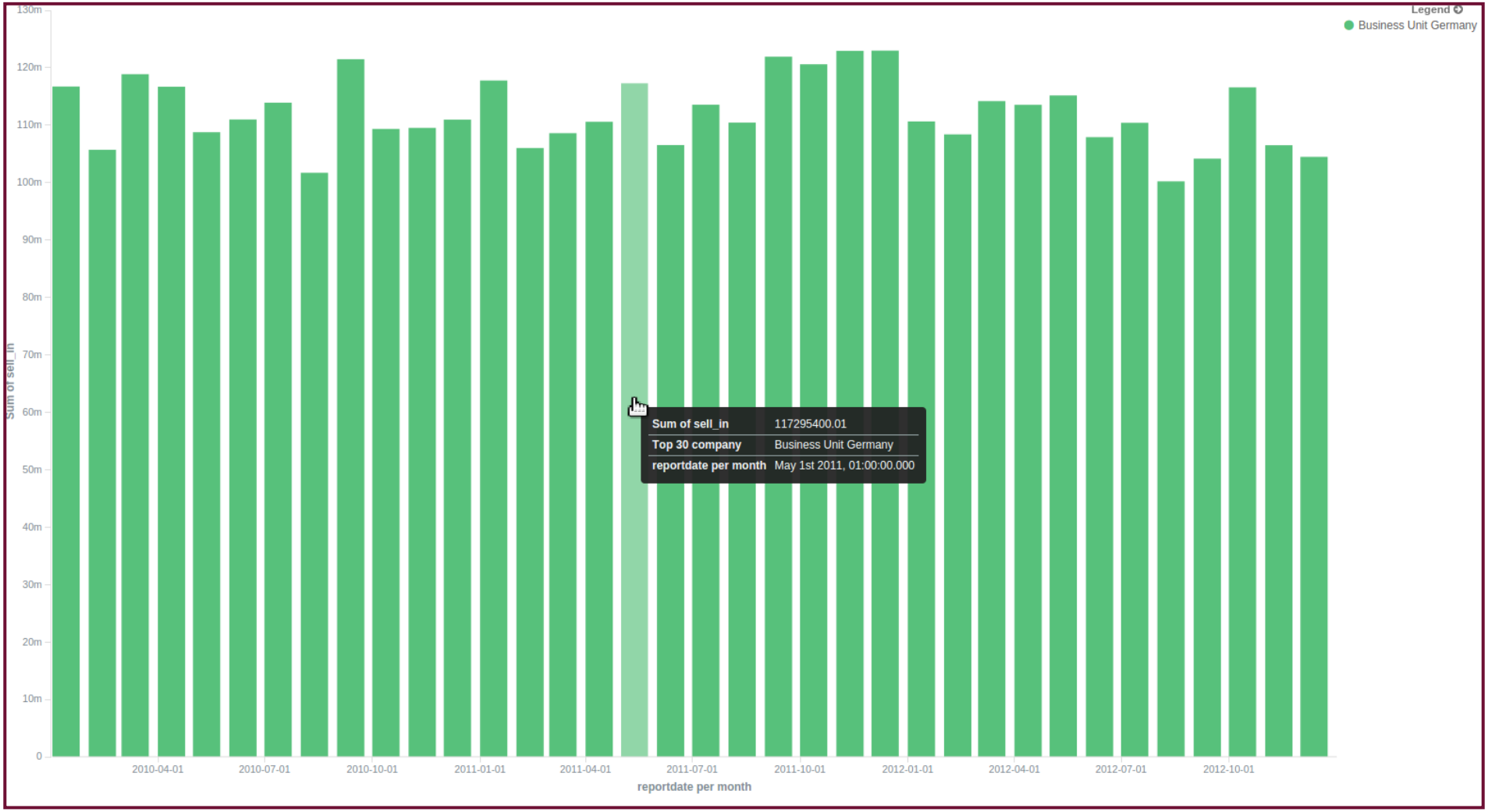
```
{
   "factmarket": {
      "mappings": {
         "market": {
            "properties": {
               "aut_idem": {
                  "type": "double"
               },
               "bu": {
                  "type": "string",
                  "index": "not_analyzed"
               },
               "company": {
                  "type": "string",
                  "index": "not_analyzed"
               },
               "reportdate": {
                  "type": "date",
                  "format": "dateOptionalTime"
               },
               "id": {
                  "type": "long"
               },
               "level": {
                  "type": "long"
               },
               "line": {
                  "type": "string",
                  "index": "not_analyzed"
               },
               "market_gen_launches0": {
                  "type": "double"
               },
               "market_gen_launches1": {
                  "type": "double"
               },
               "market_launches0": {
                  "type": "double"
               },
               "market_launches1": {
                  "type": "double"
               },
               "market_sell_out": {
                  "type": "double"
               },
               "market_vo_aut_idem": {
```

exensio
intelligente informationssysteme

Marktanalysen Pharmamarkt – Suchmaschine

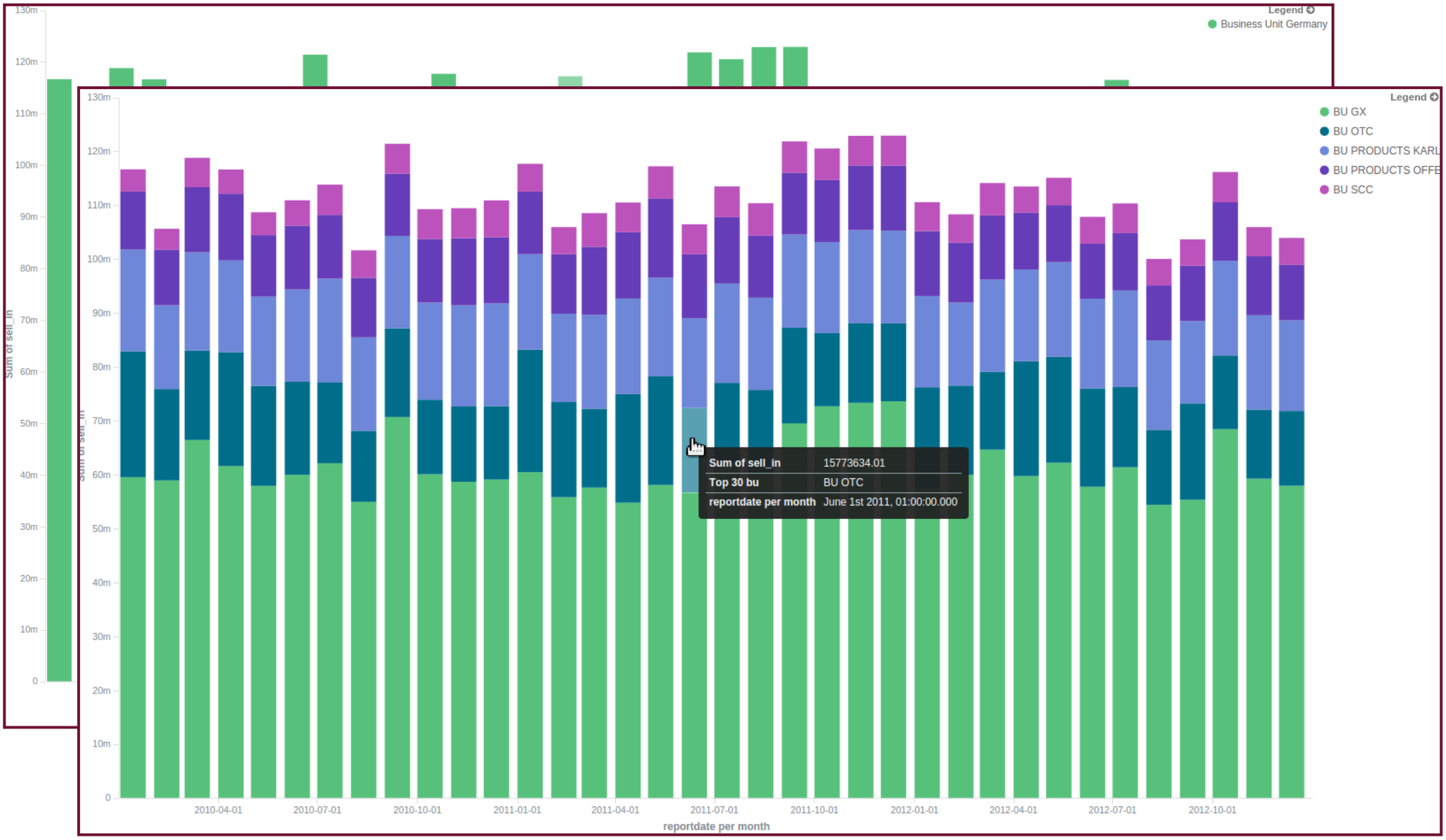
exensio
intelligente informationssysteme

DrillDown des SellIn von Firma, Business Unit und Linie



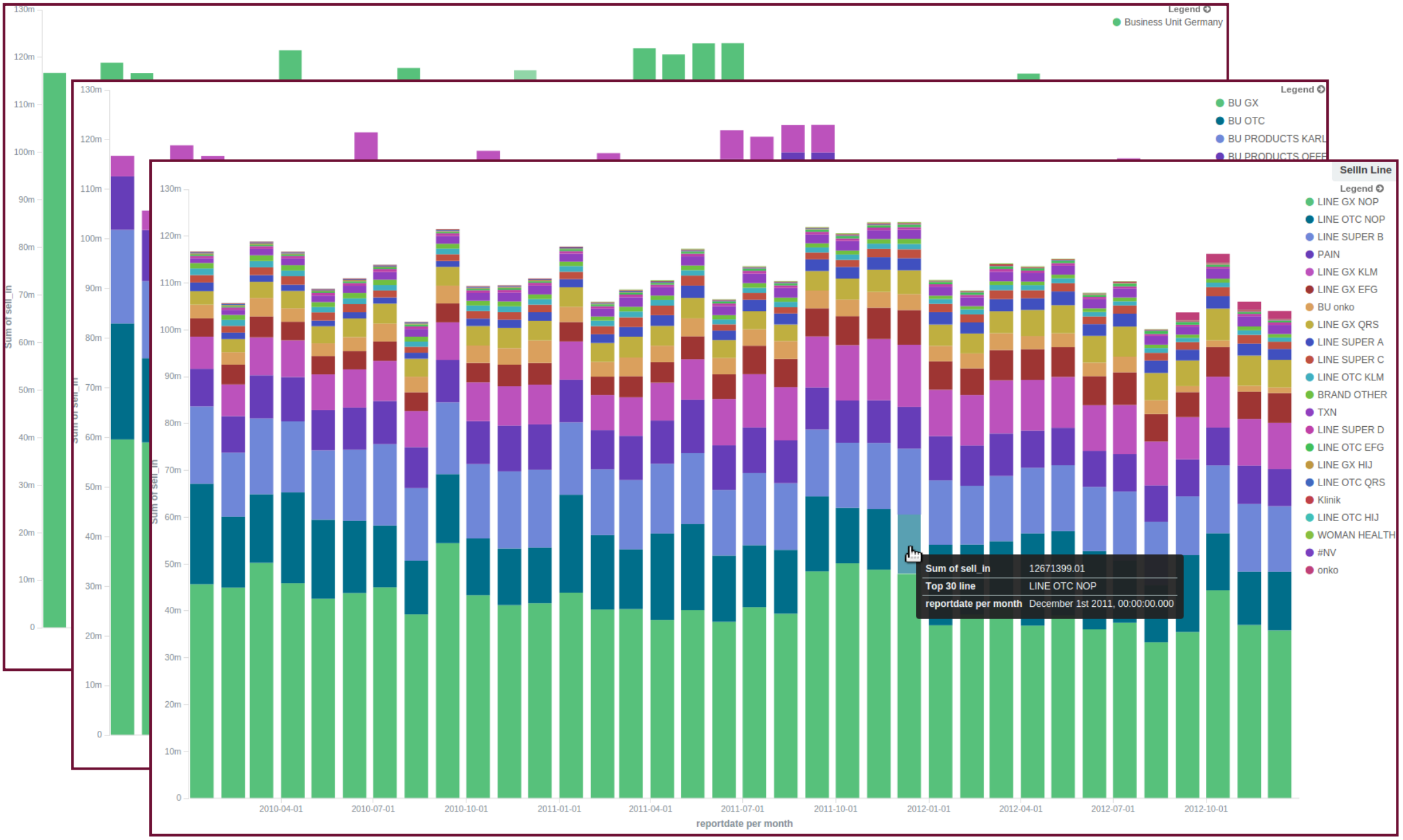
exensio
intelligente informationssysteme

DrillDown des SellIn von Firma, Business Unit und Linie



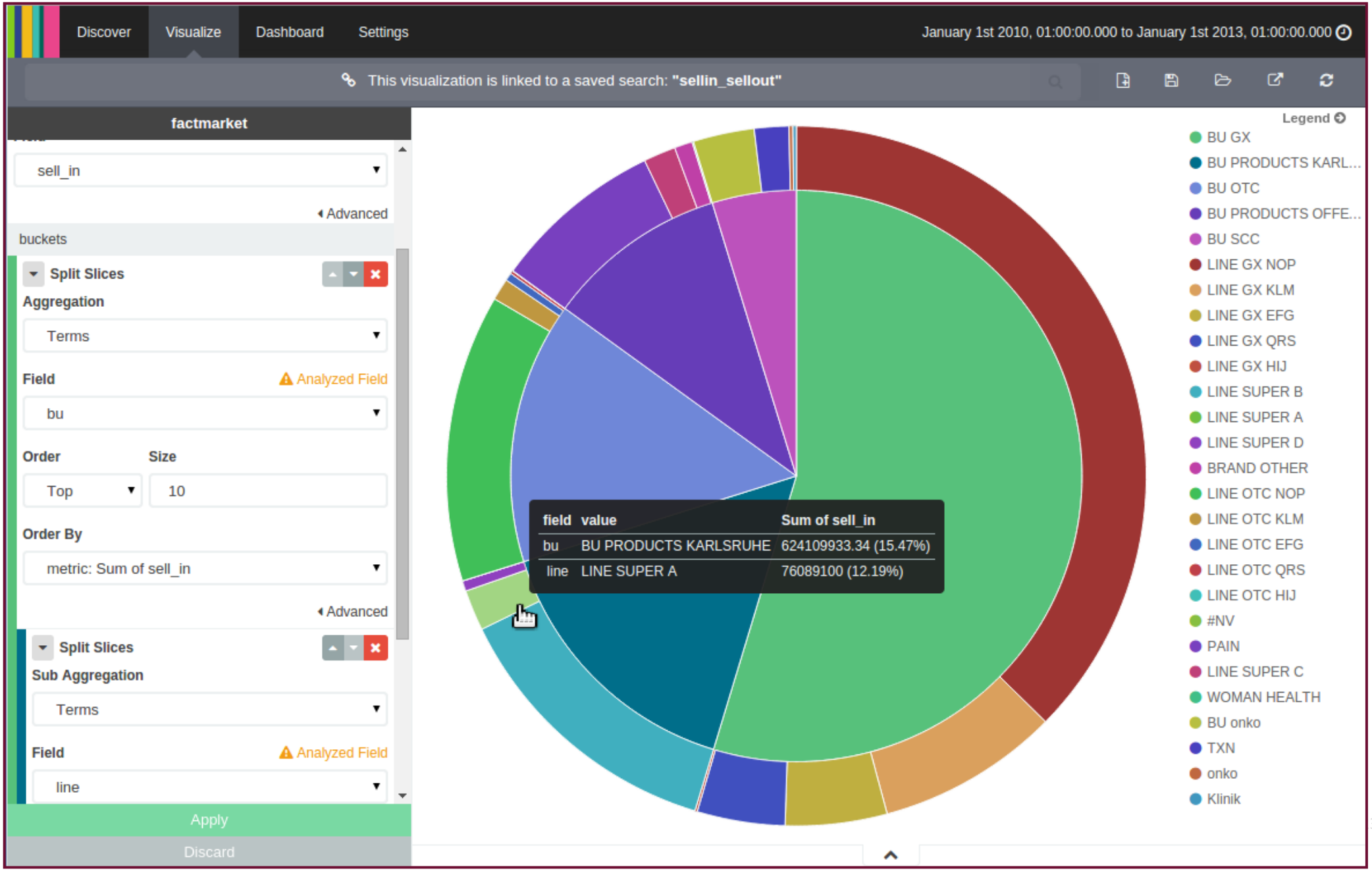
exensio
intelligente informationssysteme

DrillDown des SellIn von Firma, Business Unit und Linie



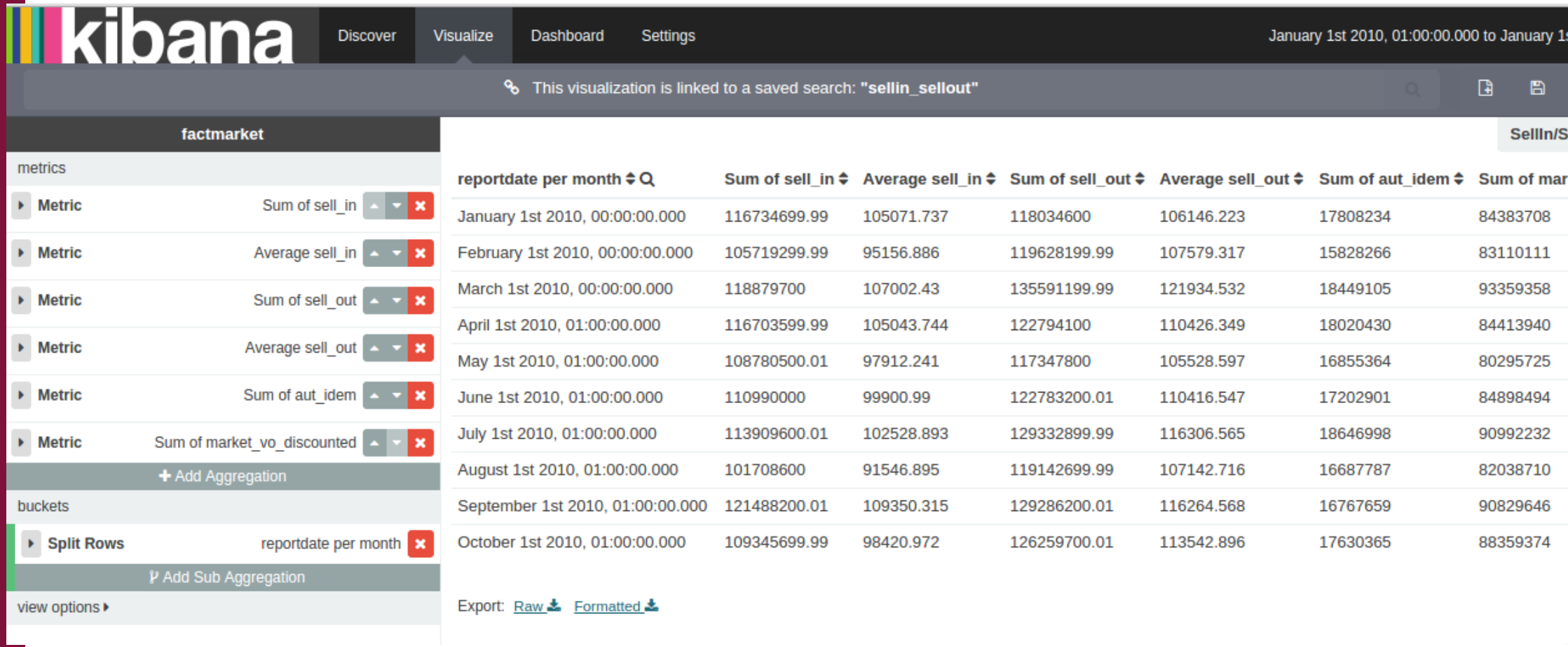
exensio
intelligente informationssysteme

Donat-Diagramm für SellIn auf BU-und Linien-Ebene



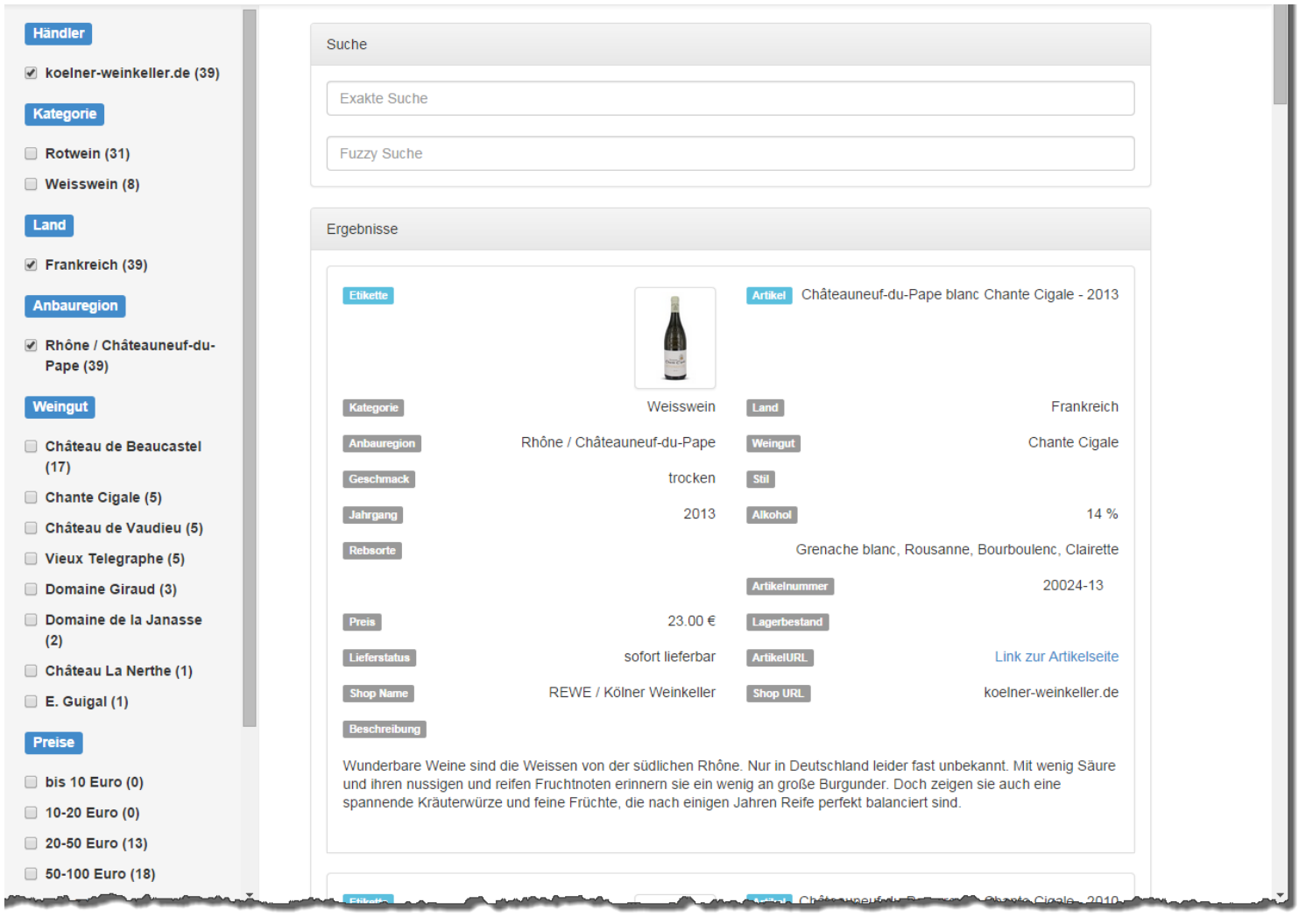
exensio
intelligente informationssysteme

Visualisierung von Aggregationen mit einer Tabelle



Competitive Intelligence als weiteres Einsatzszenario

Beispiel Shop/Handel: Mapping bspw. über den Weinnamen, falls keine eindeutige ID wie EAN Code vorhanden ist



Fazit

- Aggregationen ersetzen Dimensionen im Star-Schema
- Near Realtime
- Leichtgewichtig
- Kostengünstige Alternative
- Einfach zu integrieren und erweitern
- Kombinierbar mit Vorteilen von Suchmaschinen
  - Matching
  - Unstrukturierte Daten

exensio
intelligente informationssysteme

Fragen?

Vielen Dank!

http://blog.exensio.de

@tokraft

Partner:

elasticsearch.

ORACLE PARTNERNETWORK