Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

Elasticsearch presentation

University Stuttgart — IMS
Course: Texttechnology
Superviser: Andre Blessing
Students: Brandon Sorensen, Haywood Shannon, Johannes Krämer, King DeLaney
WS 2018/19

Elasticsearch

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

- it's a database

Elasticsearch

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

- it's a database
- it's also a search engine (optimized for text)

Elasticsearch

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

- it's a database
- it's also a search engine (optimized for text)
  - boolean search
  - phrase search / proximity search
  - wildcard queries
  - term weighting
  - aggregation

Elasticsearch

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

- it's a database
- it's also a search engine (optimized for text)
  - boolean search
  - phrase search / proximity search
  - wildcard queries
  - term weighting
  - aggregation
- implemented in java, on top of lucene
- incredibly fast
- RESTful (libraries are available e.g. for Python)
- no-sql $\rightarrow$ schema-less $\rightarrow$ example

The Pokemon Example

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
{
  "name": "Charmander",
  "type": "Fire",
  "height": "0.6",
  "region": "Kanto",
  "generation": "1"
}
```

(a) Charmander pokedex entry

The Pokemon Example

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
{
  "name": "Charmander",
  "type": "Fire",
  "height": "0.6",
  "region": "Kanto",
  "generation": "1"
}
```

(a) Charmander pokedex entry

The Pokemon Example

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
{
  "name": "Zigzagoon",
  "pokedex number": "#263",
  "type": "Normal",
  "height": "0.4",
  "region": "Hoenn",
  "generation": "3",
  "stats":
  {
    "HP": "38",
    "Attack": "30",
    "Defense": "31",
    "Sp.Atk": "30",
    "Sp.Def": "31",
    "Speed": "60",
    "Total": "240"
  }
}
```

(a) Zigzagoon pokedex entry

The Pokemon Example

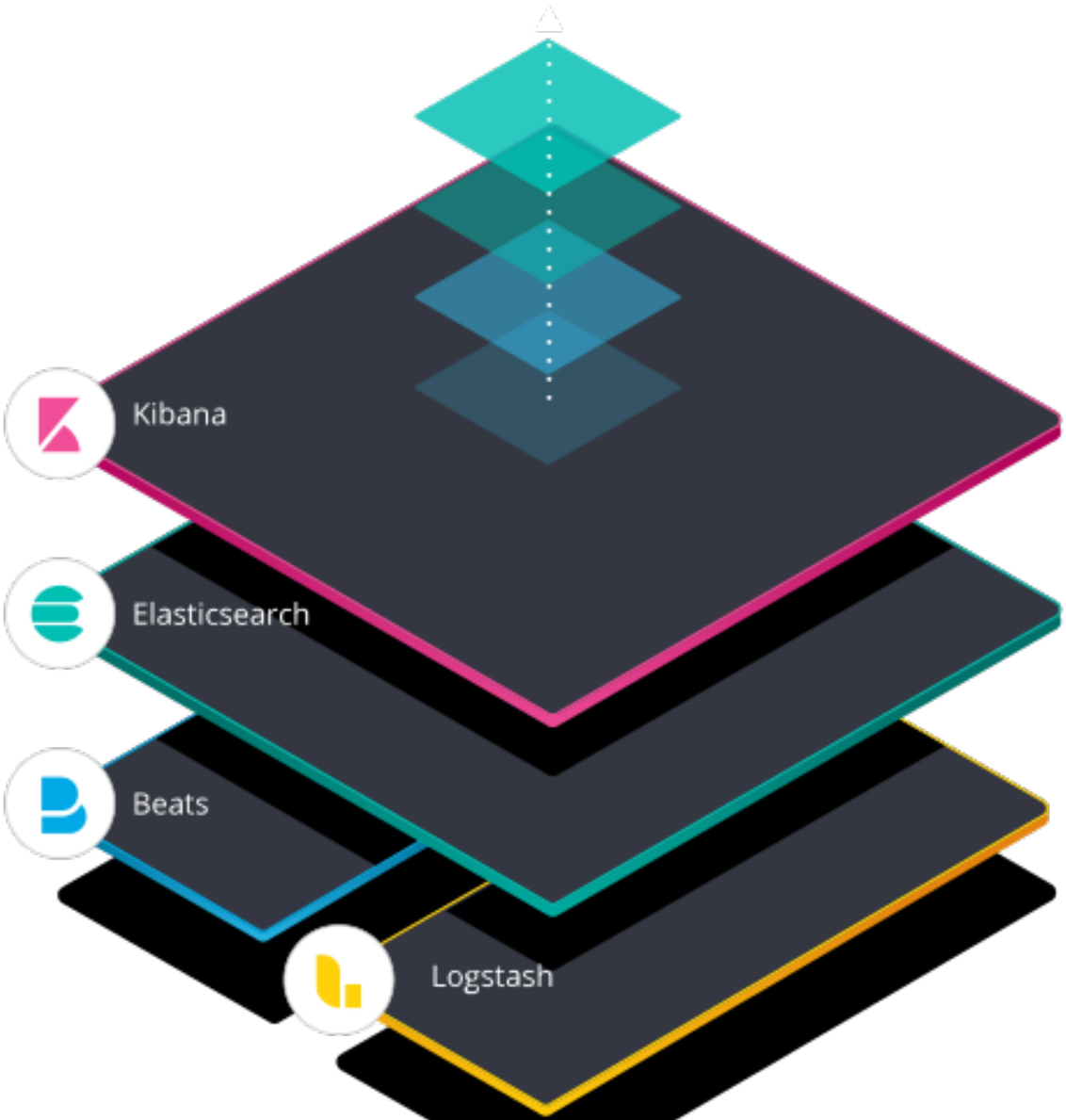
Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
{
  "name": "Zigzagoon",
  "pokedex number": "#263",
  "type": "Normal",
  "height": "0.4",
  "region": "Hoenn",
  "generation": "3",
  "stats":
  {
    "HP": "38",
    "Attack": "30",
    "Defense": "31",
    "Sp.Atk": "30",
    "Sp.Def": "31",
    "Speed": "60",
    "Total": "240"
  }
}
```

(a) Zigzagoon pokedex entry

ELK Stack

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion



Logstash Python vs. Conf

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

1. Logstash is a preprocessing tool

```
def transformdate(d):
    date,time = d.split()
    month,day,year = date.split('/')
    if len(month) == 1:
        month = '0' + month
    if len(day) == 1:
        day = '0' + day
    hour,minute = time.split(':')
    if int(hour) == 24:
        hour = '00'
    return f'{year}-{month}-{day}T{hour}:{minute}:01+00:00'
```

indexit.py

Logstash Python vs. Conf

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
filter {
  csv {
    separator => ","
    columns => ["datetime", "city", "state", "country", "shape", "duration (seconds)", "duration (hours/min)", "comments", "date posted", "latitude", "longitude"]
    skip_header => true
  }
```

ufoindex.conf

Logstash Python vs. Conf

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
mutate {
  add_field => {"location" => "%{latitude}, %{longitude}"}
  remove_field => ["date posted", "latitude", "longitude", "duration (hours/min)"]
  gsub => [
    "country", "^us$", "America",
    "country", "^ca$", "Canada",
    "country", "^gb$", "Great Britain",
    "country", "^au$", "Australia",
    "country", "^de$", "Germany",
    "datetime", "24:00", "23:59"
  ]
  convert => {
    "duration (seconds)" => "integer"
  }
}
```

ufoindex.conf

Logstash Python vs. Conf

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

```
  date {
    match => ["datetime", "MM/dd/yyyy HH:mm", "M/d/
    yyyy HH:mm", "MM/d/yyyy HH:mm", "M/dd/yyyy HH:mm
    "]
  }
}
output {
  stdout { codec => rubydebug }
  elasticsearch {
    hosts => [ "127.0.0.1" ]
    index => "ufoindex"
    document_type => "seightings"
  }
}
```

ufoindex.conf

Kibana

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

Kibana demo

Conclusion

Elasticsearch
Schema-less example
The real power
Logstash
Kibana
Conclusion

1. don't confuse it with a NLP tool, it's not
2. not recommended for PROD $\rightarrow$ not stable
3. fast changes (e.g. deprecations) with releases $\rightarrow$ careful with old tutorials
4. remarkable $\rightarrow$ get insights into data in no time