

belangrijke_concepten



context

dit document beschrijft enkele belangrijke_concepten van ELK stack

veld (field)

- kleinste data eenheid in ELK
- volledig aanpasbaar, vb: auteur, datum, samenvatting, ...
- elk veld heeft 1 **data type**: string, number, date, boolean, geolocation, ...)
- elk veld bevat 1 blokje data

multi-fields

- dienen op meer dan 1 manier geïndexeerd te worden om beter doorzoekbaar te zijn
- zo kan data geïndexeerd worden als string en keyword

meta-fields

- bevatten metadata van een document

documenten

- JSON objecten in ES index
- basis eenheid van storage, vergelijkbaar met een rij in een databank
- bevat velden die uit een key/value paar bestaan
 1. key: naam van het veld
 2. value: waarde (string, keyword, number, ...)
- bevat tevens metadata, eigen aan het document

mapping

- vergelijkbaar met een schema voor een databank.
- bevat de verschillende types die in een index zitten

- definieert de velden voor documenten van een specifiek type en hoe deze moeten geïndexeerd worden
- voorbeeld:

```
# Example
curl -XPUT localhost:9200/example -d '{
  "mappings": {
    "mytype": {
      "properties": {
        "name": {
          "type": "string"
        },
        "age": {
          "type": "long"
        }
      }
    }
  }
}'
```

Index

- logische partitie met documenten, vergelijkbaar met een databank
- je kan meerdere indexen hebben (vb 1 voor produkten, 1 voor klanten, ...)
- net als in een boek, is een boek een mapping tussen een woord en de plaats waar het voorkomt in het boek:

Tomato chutney.....	page 645
Tomato soup.....	page 23, 78
Umami burger.....	page 378

- search engines werken op een vergelijkbare manier: zoekresultaten geven je weer waar een woord/zin voorkomt.
- eerste stap in aanmaken van index, is **document parsing/processing**:
 - documenten worden doorzocht
 - lijst wordt opgemaakt van woorden (**terms**) die in het doc voorkomen
 - reference (mapping) wordt gemaakt die wijst naar de documenten waarin de term voorkomt.
- deze indexen worden op disk bewaard, zodat die later gebruikt kunnen worden bij zoekactie.

tokenization

- proces waarbij betekenis wordt gezocht in raw data ne deze wordt opgedeeld in betekenisvolle chunks
- vergelijkbaar met hoe we tekst opdelen in zinnen en zinnen in woorden.
- er bestaan verschillende tokenizers (camelcase, URL, path) die tekst herkennen.

shards

- is een Lucene index
- belangrijkste oorzaak van ES crashes, is te grote indexes
- shards splisten indexen op in delen, die je over verschillende servers kan benaderen.

replicas

- kopie van een index shard, gebruikt om een restore te doen
- kunnen ook gebruikt worden als read-only om performantie te verbeteren

analyzers

- zetten data om in terms, tijdens indexering
- Standard Analyzer en Simple Analyzer
- aangevuld met filters

meer info

voeg hier linken toe naar verdere uitleg

[elk stack](#)

From:

<https://louslab.be/> - **Lou's lab**

Permanent link:

https://louslab.be/doku.php?id=elk_stack:belangrijke_concepten

Last update: **2024/11/16 18:14**

