UPRAVLJANJE DIGITALNIM DOKUMENTIMA

Teorijske osnove:
Elasticsearch, ELK stack, MinIO

Ivan Mršulja, ivan.mrsulja@uns.ac.rs

Šta je Elasticsearch?

Distribuirani Sistem Pretrage

Open-Source

Omogućava:

- Fleksibilnost (napredni query jezik, fuzzy pretraga, boolean pretraga, relevance scoring, indeks i tip...)
- Skalabilnost (Cluster, Node, Shard,)
- Brzu pretragu (Apache Lucene ispod haube, invertovani indeks, TF-IDF)

Word	TF		IDF	TF*IDF	
	A	B		A	B
The	1/7	1/7	log(2/2) = 0	0	0
Car	1/7	0	log(2/1) = 0.3	0.043	0
Truck	0	1/7	log(2/1) = 0.3	0	0.043
Is	1/7	1/7	log(2/2) = 0	0	0
Driven	1/7	1/7	log(2/2) = 0	0	0
On	1/7	1/7	log(2/2) = 0	0	0
The	1/7	1/7	log(2/2) = 0	0	0
Road	1/7	0	log(2/1) = 0.3	0.043	0
Highway	0	1/7	log(2/1) = 0.3	0	0.043

$$IDF(t,D)=\log(N/df(t))$$

***N* -** ukupan broj dokumenata u kolekciji

df(*t*) - broj dokumenata koji sadrže termin ***t***

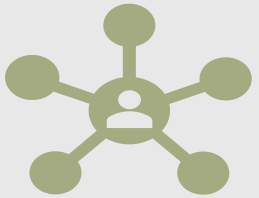
TF-IDF podsjetnik

- A: The car is driven on the road.
- B: The truck is driven on the highway.

Elasticsearch – osnovni pojmovi

Klaster (Cluster)**:**

Klaster se sastoji od jednog ili više čvorova koji dele isto ime klastera. Svaki klaster ima jednog master čvora koji se automatski bira od strane klastera i može biti zamenjen ako trenutni master čvor "umre".



Čvor (Node):

Čvor je pokrenuta instanca Elasticsearch-a koja pripada klasteru. Više čvorova može biti pokrenuto na jednom serveru u svrhu testiranja, ali obično biste trebali imati jedan čvor po serveru. Prilikom pokretanja, čvor će koristiti unicast (ili multicast, ako je navedeno) da otkrije postojanje klastera sa istim imenom klastera i pokušaće da se pridruži tom klasteru.

Elasticsearch – osnovni pojmovi

- Shard:
  - Shard je pojedinačna instanca *Lucene*-a. To je *low level* radna jedinica kojom automatski upravlja *Elasticsearch*.
  - *Elasticsearch* distribuira shardove među svim čvorovima u klasteru i može automatski premestiti shardove sa jednog čvora na drugi u slučaju umiranja čvora ili dodavanja novih čvorova.
  - Indeks je zapravo logički *namespace* koji pokazuje na primarne i replika *shard*-ove. Može imati jedan ili više primarnih i nula ili više replika shard-ova.

Šta je ELK stack?

Elasticsearch

- Skladišti i omogućava pretragu strukturisanih i nestrukturisanih podataka.

Logstash:

- Alat za prikupljanje, transformaciju i slanje log podataka (*Elasticsearch*-u).

Kibana:

- Pruža interaktivni interfejs za vizualizaciju i analizu podataka.

Zašto ELK Stack?

Omogućava praćenje i analizu podataka u realnom vremenu.

Kibana pruža bogate mogućnosti vizualizacije podataka kroz grafikone, dijagrame i *dashboard*-ove.

Elasticsearch efikasno skladišti i omogućava brzu pretragu log podataka.

Primjer Upotrebe ELK Stack-a

- **Praćenje Performansi Aplikacija**: Analiza logova kako bi se identifikovali problemi u performansama.
- **Bezbednosna Analiza**: Praćenje i analiza sigurnosnih događaja radi otkrivanja nepravilnosti.
- **Ops Monitoring**: Praćenje stanja servera, resursa i aplikacija za održavanje stabilnosti sistema.

MinIO

- *Open-source* objektni sistem za skladištenje podataka u *cloud*-u (*object storage*)
- Visoka dostupnost i otpornost na greške
- Arhitektura sa distribuiranim čvorovima omogućava (lako) horizontalno skaliranje
- Koristi se kao lokalni ili distribuirani sistem skladištenja podataka

- Kompatibilnost sa standardnim *Amazon S3* interfejsom