Elastic Stack

Arhitekture Sistema Velikih Skupova Podataka

Student
Stefan Aleksić
E2-42-2022

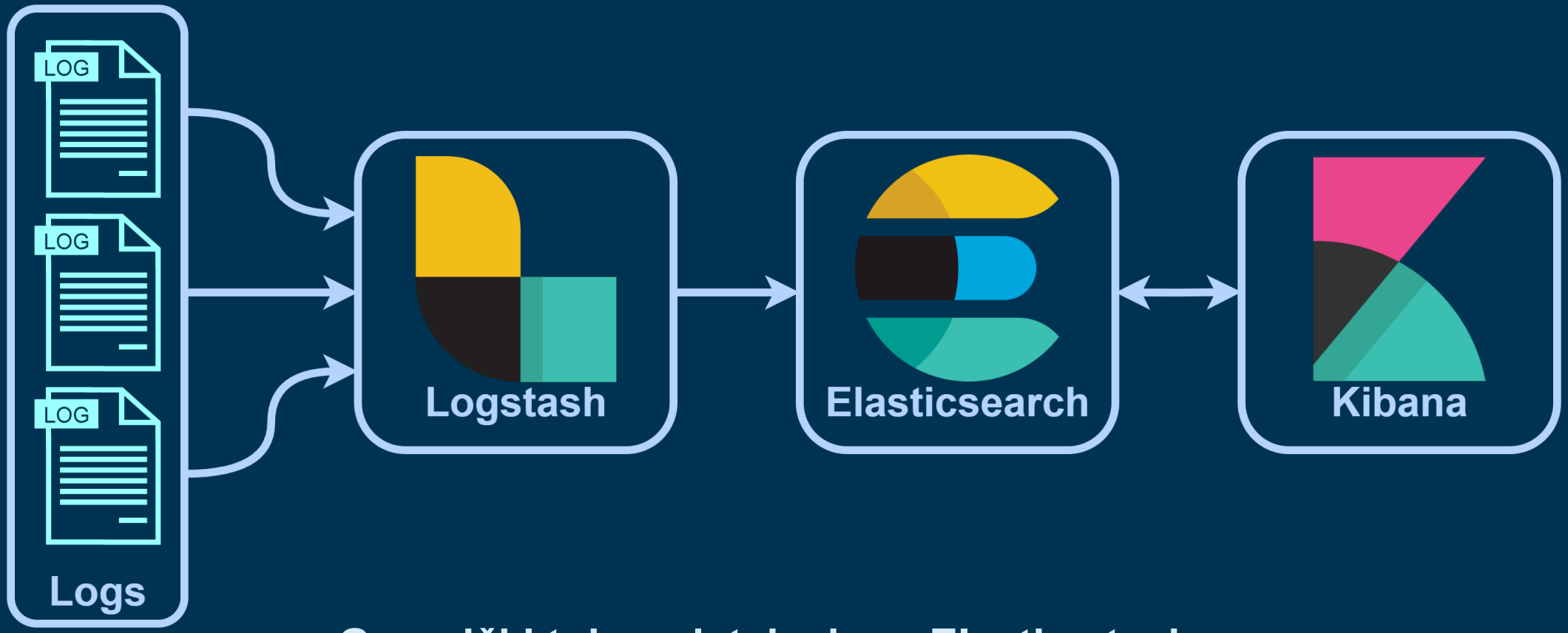
Agenda

- Uvod
- Komponente Elastic Stack-a
- Elasticsearch
- Logstash
- Kibana
- Primer korišćenja
- Zaključak

Uvod

- Elastic stack predstavlja alat otvorenog koda kompanije Elastic
- Funkcionalnosti podrazumevaju:
  - Prikupljanje podataka u bilo kom formatu,
  - Obradu podataka,
  - Obogaćivanje podataka,
  - Skladištenje podataka u formi dokumenata (JSON-a),
  - Pretraživanje podataka po sadržaju (eng. full-text search),
  - Analizu podataka i
  - Vizuelizaciju podataka.
- Akronim **ELK** identifikuje ključne komponente ovog steka:
  - **E**lasticsearch, **L**ogstash i **K**ibana.

Elastic stack



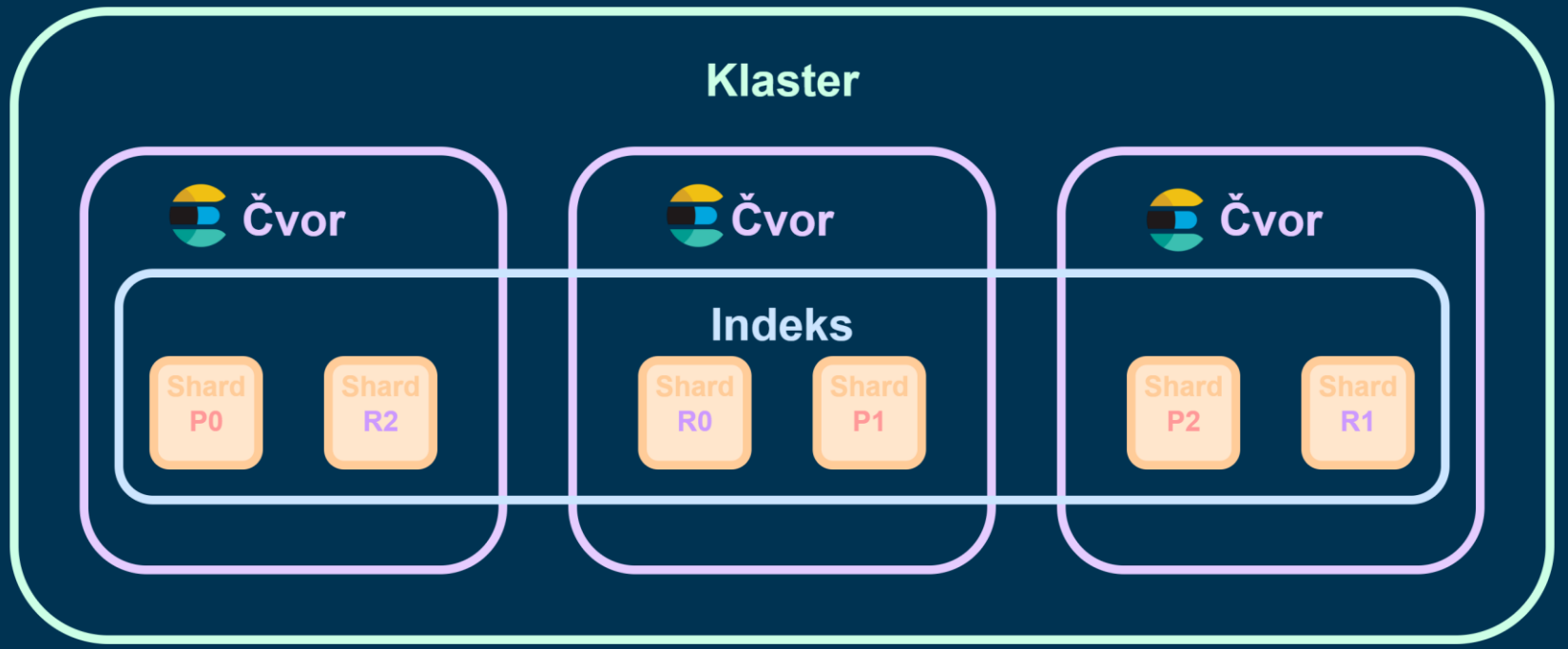
Generički tok podataka kroz Elastic stack

Elasticsearch

- Besplatan, distribuirani alat otvorenog koda koji služi za skladištenje, pretragu i analizu (*eng. search and analytics engine*), raznih tipova podataka, uključujući tekstualne, numeričke, geo-prostorne, strukturne i nestrukturne podatke.
- Zasnovan na *Apache Lucene* biblioteci
- Podaci se skladište kao *JSON* dokumenti, koji pripadaju jednom indeksu, dok je uz pomoć distribuiranog modela, indekse moguće podeliti na manje komponente, odnosno krhotine (*eng. shard*), koje mogu biti rasprostranjene kroz veći broj čvorova (*eng. node*).
- *REST API* za manipulaciju podataka.

Elasticsearch

- Osnovni elementi alata
  - Klaster
  - Čvor
  - Indeks
  - Shard
    - Primarni
    - Replika
  - Analizator

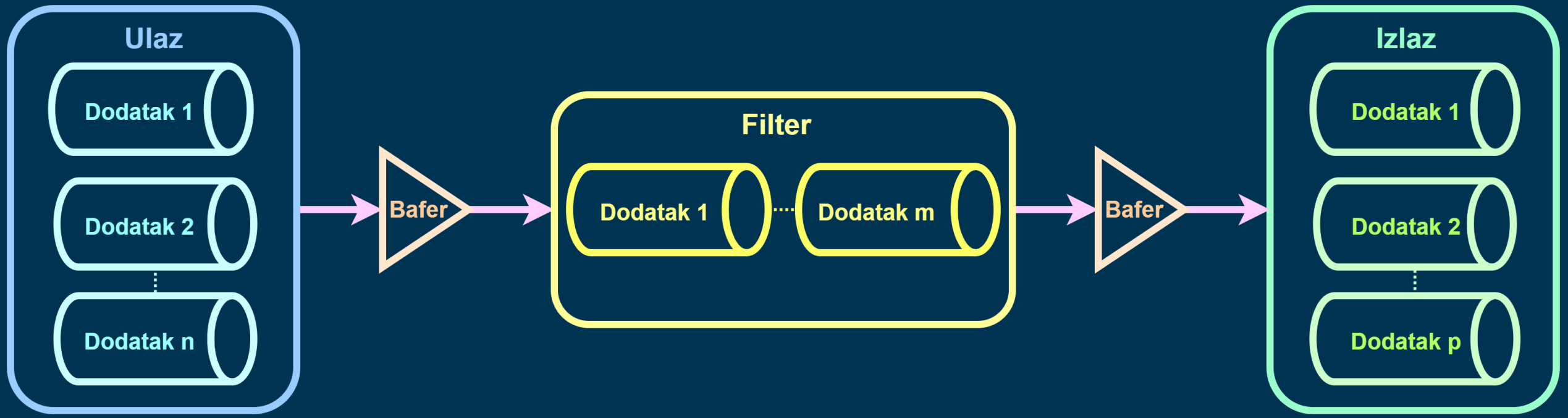


Arhitektura

Logstash

- Alat otvorenog koda za prikupljanje podataka sa mogućnostima nadovezivanja (*eng. pipelining*) u realnom vremenu.
- Mogućnost dinamičkog objedinjavanja podataka iz različitih izvora, njihova integracija, normalizacija i na kraju dostavljanje podataka na unapred definisana odredišta.
- Mogućnost prečišćavanja podataka za različite slučajeve upotrebe, naprednu analizu ili vizuelizaciju.
- Bilo koja vrsta događaja može biti prosleđena alatu, koji uz pomoć dodataka odgovarajućih etapa vrši prijem, transformaciju i dostavljanje podataka.
- Sam proces obrade je podržan velikim brojem često korišćenih kodera, a pritom je alat dizajniran za obradu velike količine podataka u realnom vremenu.
- Razijen u programskom jeziku *JRuby* i izvršava se na javinoj virtuelnoj mašini (*eng. Java Virtual Machine*), skraćeno *JVM*.

Logstash

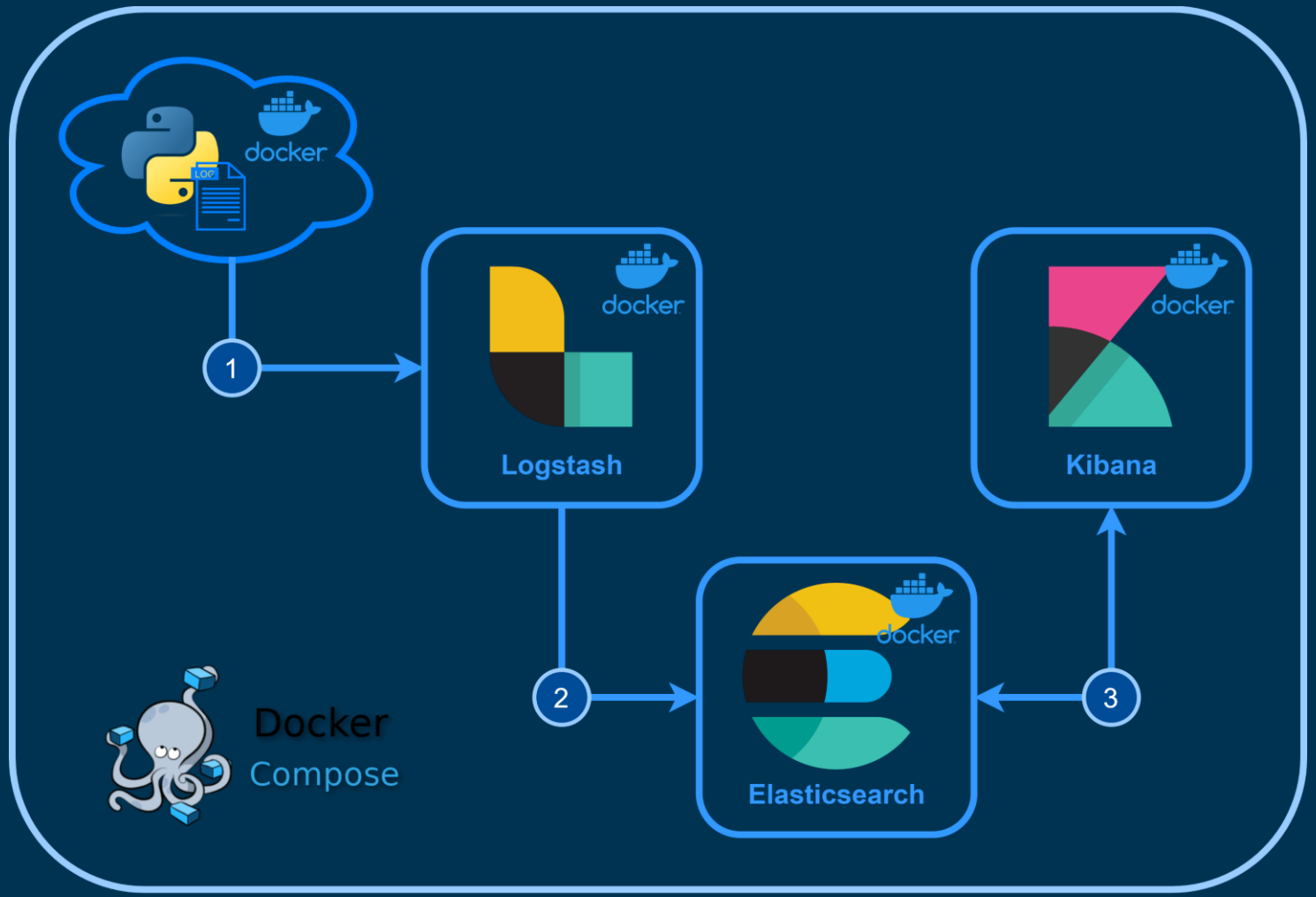


Tok podataka (*eng. pipeline*)

Kibana

- Aplikacija otvorenog koda koja pruža pregledan korisnički interfejs ka Elastik steku, uz mogućnosti pretraživanja, vizuelizacije i analize podataka indeksiranih *Elasticsearch*-om.
- Omogućava nadgledanje, upravljanje i održavanje bezbednosti klastera u ELK steku.
- Alat za kreiranje raznovrsnih grafikona (*eng. charts*), poput: poluga, pita, tabela, histograma, mapa i slično.
- Komandna tabla (*eng. dashboard*) kombinuje ove elemente na jedan pano i time omogućava analitički pregled podataka u realnom vremenu za razne slučajeve korišćenja, poput: analize podataka evidencije, nadgledanja metričkih podataka infrastrukture i kontejnera, nadgledanja performansi aplikacija (*eng. Application performance monitoring, APM*), vizuelizacije geo-prostornih podataka, analize sigurnosnih podataka, analize podataka biznis logike.

Primer korišćenja



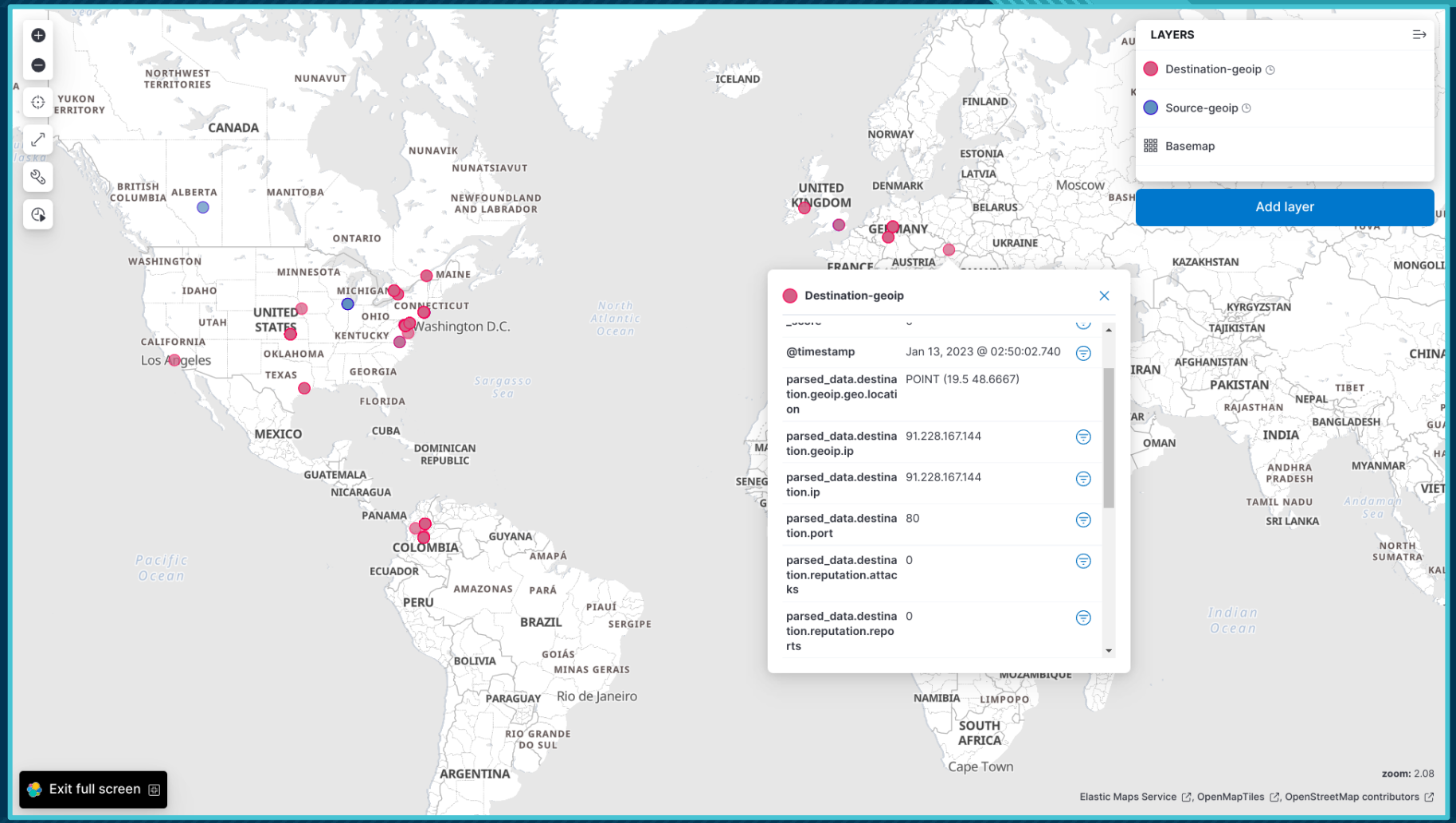
Arhitektura sistema za analizu podataka mrežnog saobraćaja

Primer korišćenja

```
214.1.211.251 - - [15/Apr/2011:09:40:17 -0700] "GET /global.asa HTTP/1.0" 404 315 "-" "-"
214.1.211.251 - - [15/Apr/2011:09:40:17 -0700] "GET /~root HTTP/1.0" 404 310 "-" "-"
214.1.211.251 - - [15/Apr/2011:09:40:18 -0700] "GET /~apache HTTP/1.0" 404 312 "-" "-"
219.167.17.173 - - [17/Apr/2011:17:55:40 -0700] "POST /sony/mmr HTTP/1.1" 200 130 "-" "PS
218.41.54.67 - - [17/Apr/2011:18:20:18 -0700] "POST /sony/mmr HTTP/1.1" 200 130 "-" "PS3A
10.132.93.114 - - [18/Apr/2011:11:05:39 -0700] "POST /sony/mmr HTTP/1.1" 200 61 "-" "Ledi
10.132.93.114 - - [18/Apr/2011:11:07:07 -0700] "POST /sony/mmr HTTP/1.1" 200 61 "-" "Ledi
10.132.93.114 - - [18/Apr/2011:11:13:52 -0700] "POST /sony/mmr HTTP/1.1" 200 61 "-" "Ledi
218.41.54.67 - - [20/Apr/2011:17:42:37 -0700] "POST /sony/mmr HTTP/1.1" 200 100 "-" "PS3A
60.34.131.229 - - [20/Apr/2011:18:22:32 -0700] "POST /sony/mmr HTTP/1.1" 200 100 "-" "PS3
202.213.251.245 - - [21/Apr/2011:21:16:45 -0700] "POST /sony/mmr HTTP/1.1" 200 100 "-" "P
202.213.251.245 - - [21/Apr/2011:21:24:43 -0700] "POST /sony/mmr HTTP/1.1" 200 100 "-" "P
178.202.110.92 - - [22/Apr/2011:18:59:05 -0700] "GET / HTTP/1.1" 200 315 "-" "Mozilla/5.0
178.202.110.92 - - [22/Apr/2011:18:59:05 -0700] "GET /favicon.ico HTTP/1.1" 404 333 "-" "
178.202.110.92 - - [22/Apr/2011:18:59:05 -0700] "GET /favicon.ico HTTP/1.1" 404 333 "-" "
178.202.110.92 - - [22/Apr/2011:18:59:07 -0700] "GET /access-navigator-media HTTP/1.1" 20
178.202.110.92 - - [22/Apr/2011:19:05:00 -0700] "GET /admin/cdr/counter.txt HTTP/1.1" 404
178.202.110.92 - - [22/Apr/2011:19:05:41 -0700] "GET //help/readme.nsf?OpenAbout HTTP/1.1
178.202.110.92 - - [22/Apr/2011:19:05:54 -0700] "GET /catinfo?A HTTP/1.1" 404 329 "-" "Mo
178.202.110.92 - - [22/Apr/2011:19:06:08 -0700] "GET /errors-navigator-media HTTP/1.1" 20
178.202.110.92 - - [22/Apr/2011:19:27:04 -0700] "GET / HTTP/1.1" 200 315 "-" "Mozilla/5.0
```

Generički prikaz podataka mrežnog saobraćaja

Vizuelizacija podataka



Pogled geo-lokacije izvornih i odredišnih adresa internet protokola

Zaključak

- Elasticsearch-ov model podataka je zaista fleksibilan i intuitivan.
- Logstash pruža neverovatnu fleksibilnost pri obrađivanju podataka u realnom vremenu.
- Kibana, kao alat za vizuelizaciju ima veliki broj mogućnosti
- ELK stack je veoma stabilan alat. Tokom realizacije sistema, konfiguracija je bila vrlo jednostavna, a rušenje i ponovno podizanje kontejnera nije pravilo nikakve probleme u međusobnoj komunikaciji između komponenata.
- Kada se poredi sa tehnologijama poput Apache Spark-a ili Hadoop-a, mogu se naći zajedničke karakteristike. To je, u suštini, rezultat toga da svaki radni okvir želi da pruži način za obradu skupa velikih podataka, čime se granice raznih tehnologija zamagljuju. Ipak svaki alat služi određenoj svrsi i moramo da izaberemo ono što najbolje odgovara datim zahtevima.
- Ako jednostavno želimo da lociramo dokumente prema ključnoj reči i izvršimo jednostavnu analizu, onda ELK stack može da posluži kao sjajan alat za to. Ako imamo ogromnu količinu podataka za koje je potreban širok spektar različitih tipova složene obrade i analize, onda Hadoop pruža najširi spektar alata i najveću fleksibilnost.