ELK Stack

ElasticSearch, Logstash, Kibana, FileBeat.

Andranik Barseghyan

https://github.com/Andranik93/Presentation

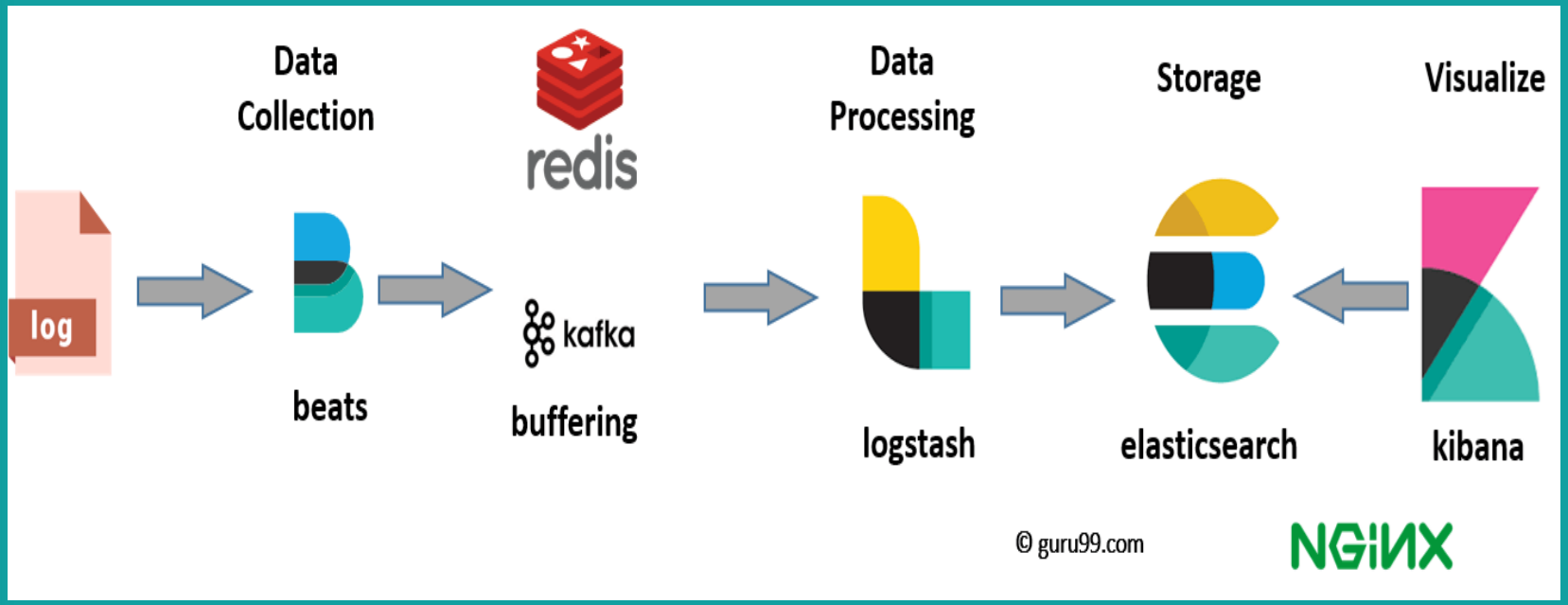
ELK Stack

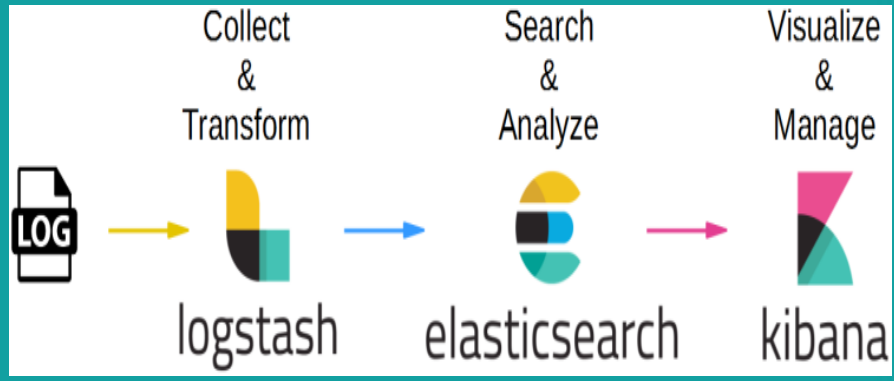
Data Collection

Data Aggregation & Processing

Indexing & storage

Analysis & visualization



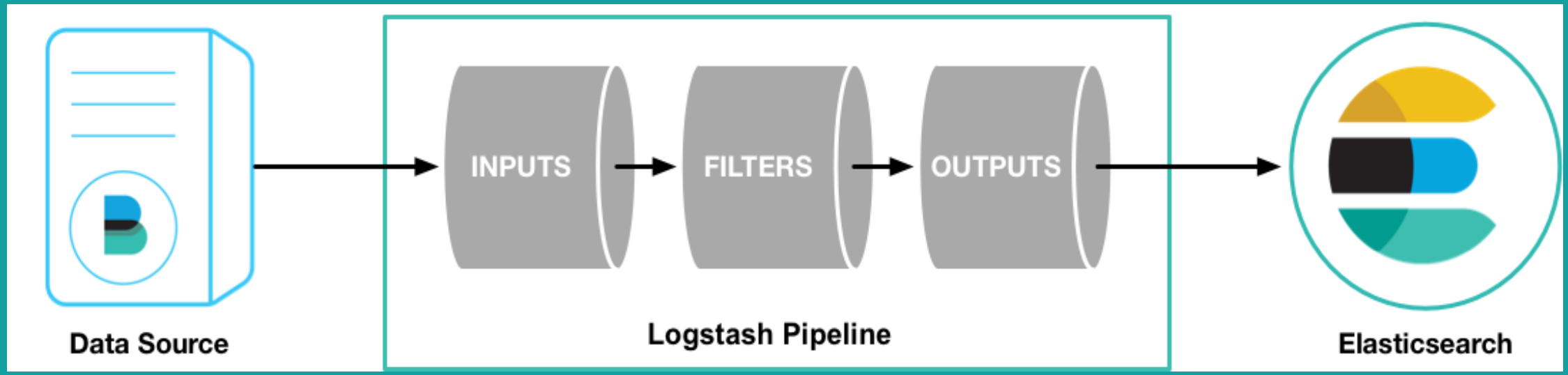


beats

- Beats is the platform for single-purpose data shippers. They send data from hundreds or thousands of machines and systems to Logstash or Elasticsearch.

logstash

- Logstash is an open source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously, transforms it, and then sends it to your favorite “stash.” (Ours is Elasticsearch, naturally.).
- Writen on JRuby.



What is **ElasticSearch**

- ElasticSearch is a free and open source distributed inverted index created by shay banon.
- Build on top of Apache Lucene
- Lucene is a most popular java-based full text search index implementation.
- First public release version v0.4 in February 2010.
- Developed in Java, so inherently cross-plateform.

Why ElasticSearch

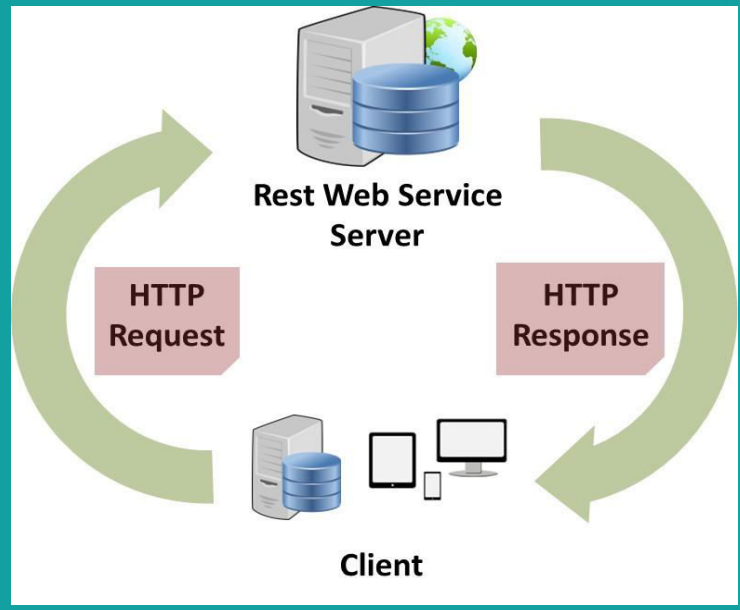
- ❑ Easy to scale (Distributed)
- ❑ Everything is one JSON call away
- ❑ Unleashed power of Lucene under the hood
- ❑ Excellent Query DS
- ❑ Multi-tenancy
- ❑ Support for advanced search features (Full Text)
- ❑ Configurable and Extensible
- ❑ Document Oriented
- ❑ Schema free

Easy to Scale (Distributed)

- Elasticsearch allows you to start small, but will grow with your business. It is built to scale horizontally out of the box. As you need more capacity, just add more nodes, and let the cluster reorganize itself to take advantage of the extra hardware
- One server can hold one or more parts of one or more indexes, and whenever new nodes are introduced to the cluster they are just being added to the party. Every such index, or part of it, is called a shard, and Elasticsearch shards can be moved around the cluster very easily.

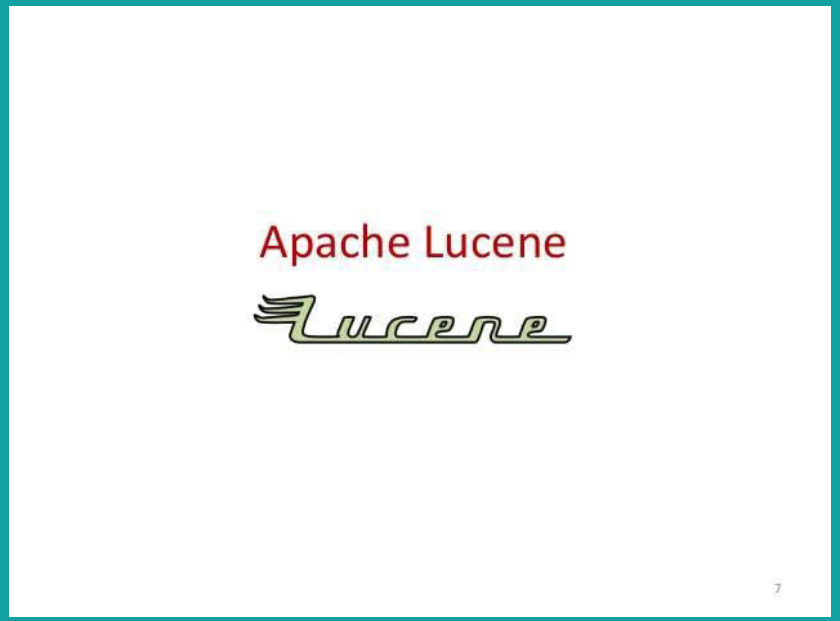
RESTful API

- Elasticsearch is API driven. Almost any action can be performed using a simple RESTful API using JSON over HTTP. An API already exists in the language of your choice.
- Responses are always in JSON, which is both machine and human readable.



Build on top of Apache Lucene

- Apache Lucene is a high performance, full-featured Information Retrieval library, written in Java. Elasticsearch uses Lucene internally to build its state of the art distributed search and analytics capabilities.
- Since Lucene is a stable, proven technology, and continuously being added with more features and best practices, having Lucene as the underlying engine that powers Elasticsearch.



Excellent Query DSL

- The REST API exposes a very complex and capable query DSL, that is very easy to use. Every query is just a JSON object that can practically contain any type of query, or even several of them combined.
- Using filtered queries, with some queries expressed as Lucene filters, helps leverage caching and thus speed up common queries, or complex queries with parts that can be reused.
- Faceting, another very common search feature, is just something that upon-request is accompanied to search results, and then is ready for you to use.

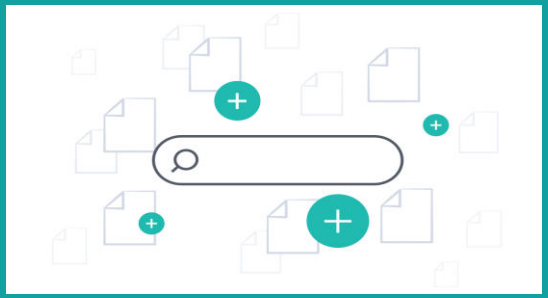
Multi-tenancy

- You can host multiple indexes on one Elasticsearch installation - node or cluster. Each index can have multiple "types", which are essentially completely different indexes.
- The nice thing is you can query multiple types and multiple indexes with one simple query. This opens quite a lot of options.

curl -X GET “127.0.0.1:9200”

Support for advanced search features (Full Text)

- Elasticsearch uses Lucene under the covers to provide the most powerful full
- text search capabilities available in any open source product.
- Search comes with multi-language support, a powerful query language, support for geolocation, context aware did-you-mean suggestions, autocomplete and search snippets.
- script support in filters and scorers



Configurable and Extensibl

- Many of Elasticsearch configurations can be changed while Elasticsearch is running, but some will require a restart (and in some cases reindexing). Most configurations can be changed using the REST API too.
- Elasticsearch has several extension points - namely site plugins (let you serve static content from ES - like monitoring javascript apps), rivers (for feeding data into Elasticsearch), and plugins that let you add modules or components within Elasticsearch itself. This allows you to switch almost every part of Elasticsearch if so you choose, fairly easily.
- If you need to create additional REST endpoints to your Elasticsearch cluster, that is easily done as well.

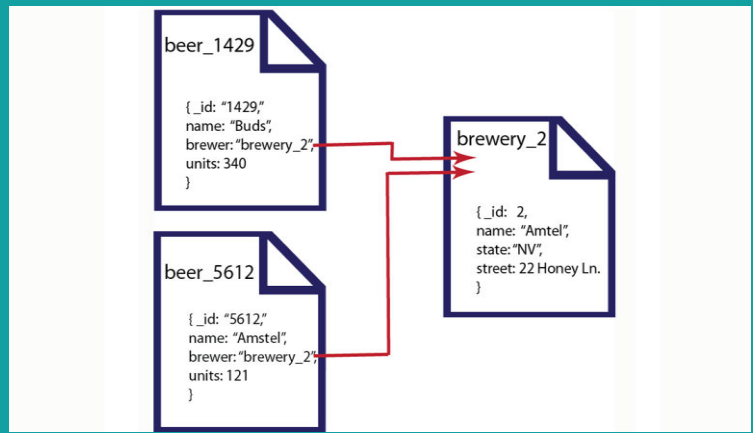
Document Oriented

- Store complex real world entities in Elasticsearch as structured JSON documents. All fields are indexed by default, and all the indices can be used in a single query, to return results at breath taking speed.

Per-operation Persistence

- Elasticsearch puts your data safety first. Document changes are recorded in transaction logs on multiple nodes in the cluster to minimize the chance of any data loss

- Stemer
- Soundex
- Ngramn



Schema free

- Elasticsearch allows you to get started easily. Toss it a JSON document and it will try to detect the data structure, index the data and make it searchable. Later, apply your domain specific knowledge of your data to customize how your data is indexed.

Conflict management

- Optimistic version control can be used where needed to ensure that data is never lost due to conflicting changes from multiple processes.

Active community

- The community, other than creating nice tools and plugins, is very helpful and supporting. The overall vibe is really great, and this is an important metric of any OSS project
- There are also some books currently being written by community members, and many blog posts around the net sharing experiences and knowledge

Basic Concepts

Cluster

A cluster consists of one or more nodes which share the same cluster name. Each clu
has a single master node which is chosen automatically by the cluster and which can be
replaced if the current master node fails.

Node

A node is a running instance of elasticsearch which belongs to a cluster. Multiple nodes can be started on a single server for testing purposes, but usually you should have one node per server.
At startup, a node will use unicast (or multicast, if specified) to discover an existing cluster with the same cluster name and will try to join that cluster.

Index

An index is like a 'database' in a relational database. It has a mapping which defines multiple types.
An index is a logical namespace which maps to one or more primary shards and can have zero or more replica shards.

Type

A type is like a 'table' in a relational database. Each type has a list of fields that can be specified for documents of that type. The mapping defines how each field in the document is analyzed

Document

A document is a JSON document which is stored in elasticsearch. It is like a row in a table in a relational database. Each document is stored in an index and has a type and an id.
A document is a JSON object (also known in other languages as a hash / hashmap / associative array) which contains zero or more fields, or key-value pairs. The original JSON document that is indexed will be stored in the _source field, which is returned by default when getting or searching for a document.

Field

A document contains a list of fields, or key-value pairs. The value can be a simple (scalar) value (eg a string, integer,
date), or a nested structure like an array or an object. A field is similar to a column in a table in a relational database.
The mapping for each field has a field 'type' (not to be confused with document type) which indicates the type of data that can be stored in that field, eg integer, string, object. The mapping also allows you to define (amongst other things) how the value for a field should be analyzed.

Mapping

A mapping is like a 'schema definition' in a relational database. Each index has a mapping, which defines each type within the index, plus a number of index-wide settings. A mapping can either be defined explicitly, or it will be generated automaticaly when a document is indexed

Shard

A shard is a single Lucene instance. It is a low-level "worker" unit which is managed automatically by elasticsearch. An index is a logical namespace which points to primary and replica shards.
Elasticsearch distributes shards amongst all nodes in the cluster, and can move shards automatically from one node to another in the case of node failure, or the addition of new nodes.

Running...

http://localhost:9200/?pretty

Response :

```
{
    "status" : 200,
    "name" : “elasticsearch",
    "version" : { "number" : "1.1.1",
    "build_hash" :
        "f1585f096d3f3985e73456debdc1a0745f512bbc",
        "build_timestamp" : "2014-04-16T14:27:12Z",
    "build_snapshot" : false,
    "lucene_version" : "4.7"
    },
    "tagline" : "You Know, for Search"
}
```

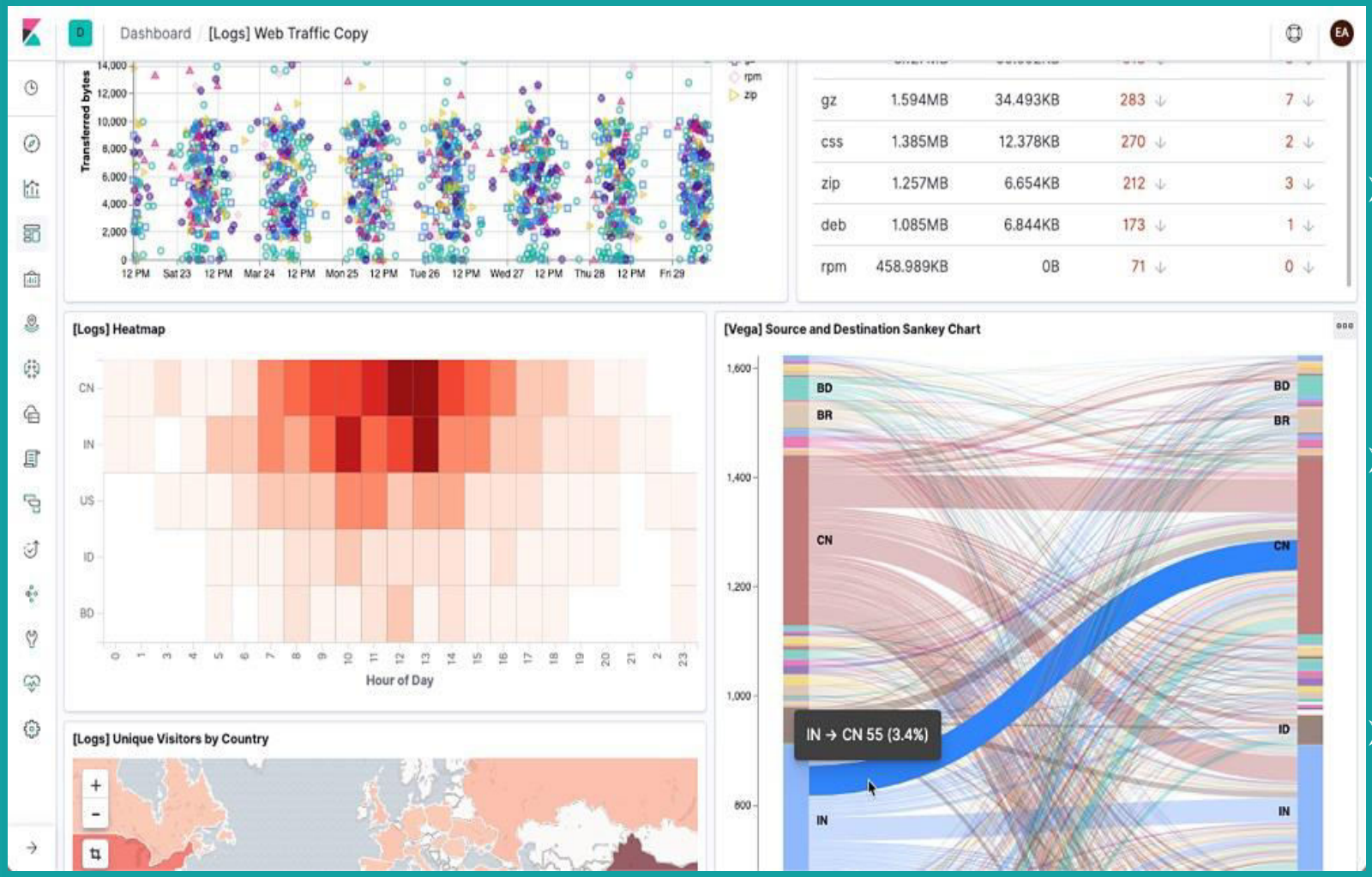
Request :

```
$ curl -XPUT "http://localhost:9200/test-data/cities/21"  -d '{
    "rank": 21,
    "city": "Yerevan",
    "state": "Armenia",
    "population2010": 617594,
    "land_area": 48.277, "location": {
    "lat": 42.332,
    "lon": 71.0202 },
    "abbreviation": "AM"
    }'
```

Response :

```
{"ok":true,"_index":"test-data","_type":"cities","_id":"21","_version":1}
```

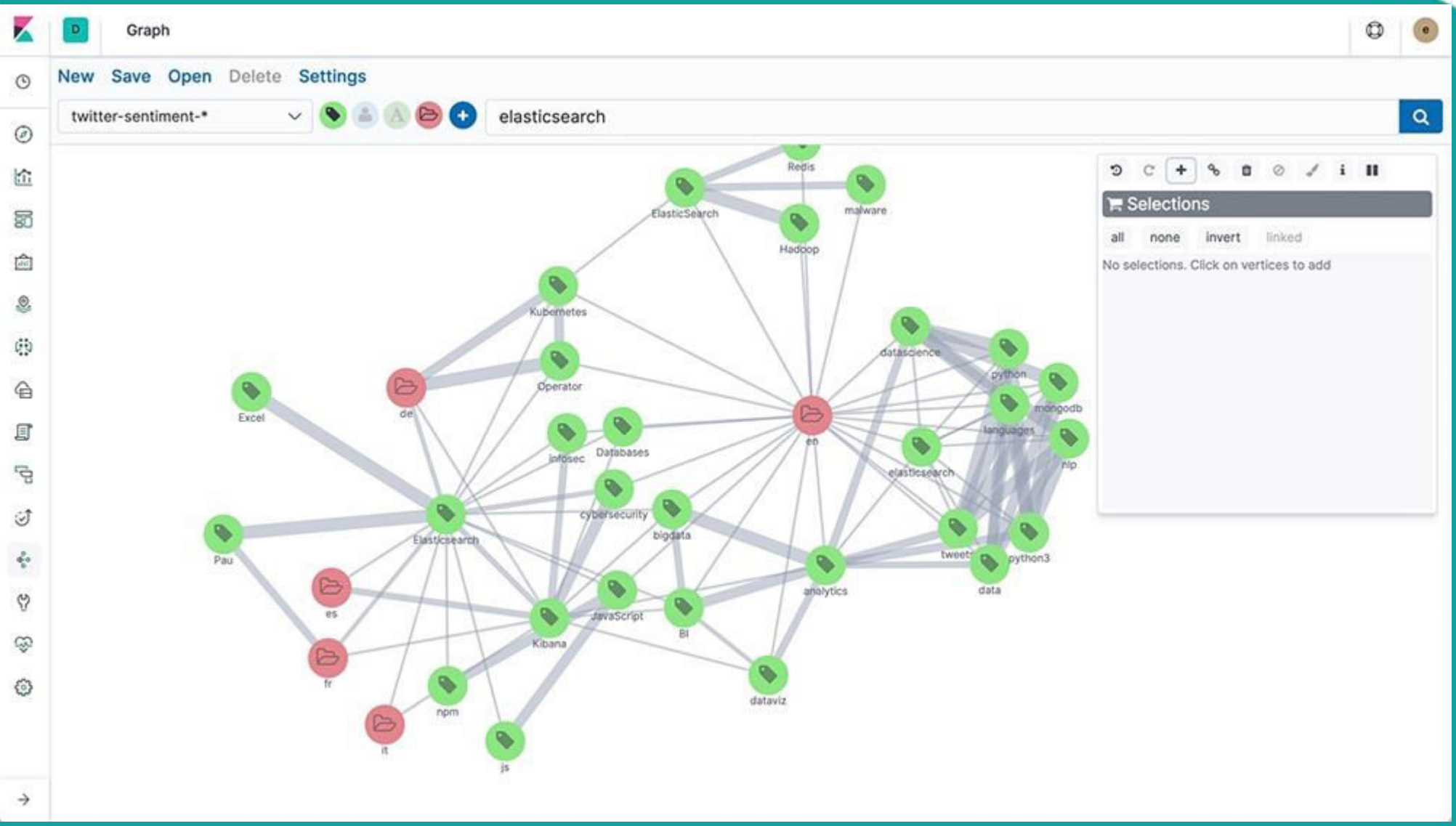
kibana



- Kibana lets you visualize your Elasticsearch data and navigate the Elastic Stack, so you can do anything from learning why you're getting paged at 2:00 a.m. to understanding the impact rain might have on your quarterly numbers.
- Kibana gives you the freedom to select the way you give shape to your data. And you don't always have to know what you're looking for. With its interactive visualizations, start with one question and see where it leads you.
- NodeJS + JS client for elasticsearch can visualize data from elasticsearch. Web GUI.

Put Geo Data on Any Map

Analyze Relationships with Graph



IDS
IPS

IDS

Intrusion detection system (IDS), is a device or software that monitors a network or systems for malicious activity or policy violations. Any malicious activity or violation is typically reported either to an administrator or collected centrally using a security information and event management (SIEM) system. A SIEM system combines outputs from multiple sources, and uses alarm filtering techniques to distinguish malicious activity from false alarms.

IPS

Intrusion Prevention System (IPS), is a tool that is used to sniff out malicious activity occurring over a network and/or system. Intrusion prevention systems can also be referred to as intrusion detection and prevention systems (IDPS). Intrusion prevention systems function by finding malicious activity, recording and reporting information about the malicious activity, and trying to block/stop the activity from occurring.

Get Creative with Canvas

THANKS

https://www.elastic.co/