elasticsearch

Table Of Content

Introduction

Big data refers to data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many fields offer greater statistical power, while data with higher complexity may lead to a higher false discovery rate.

150+

Data that is successfully created for a purpose

250+

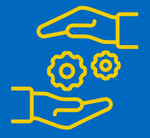
Data stored and owned by us

It is referred to as a data set that is too large and considered too complex for traditional data processing software.

Big data relies on three concepts:

Speed

Variety



Volume

Big Data Nowadays

Big Data Nowadays consists of predictive analytics, user behavior analysis, or specific other advanced data analysis methods.

Speed

Speed in data collection increases

Processing

Data processing becomes more effective

Big Data Includes

Big data includes data sets of sizes beyond the capabilities of software tools commonly used to capture, curate, manage, and process data in a relatively fast time.

Big data includes:

Unstructured

Raw data that has not been processed at all

Semi Structured

Daw data that has been organized by category

Structured Data

Raw data that has been organized for a purpose

What Is Elastic Search?

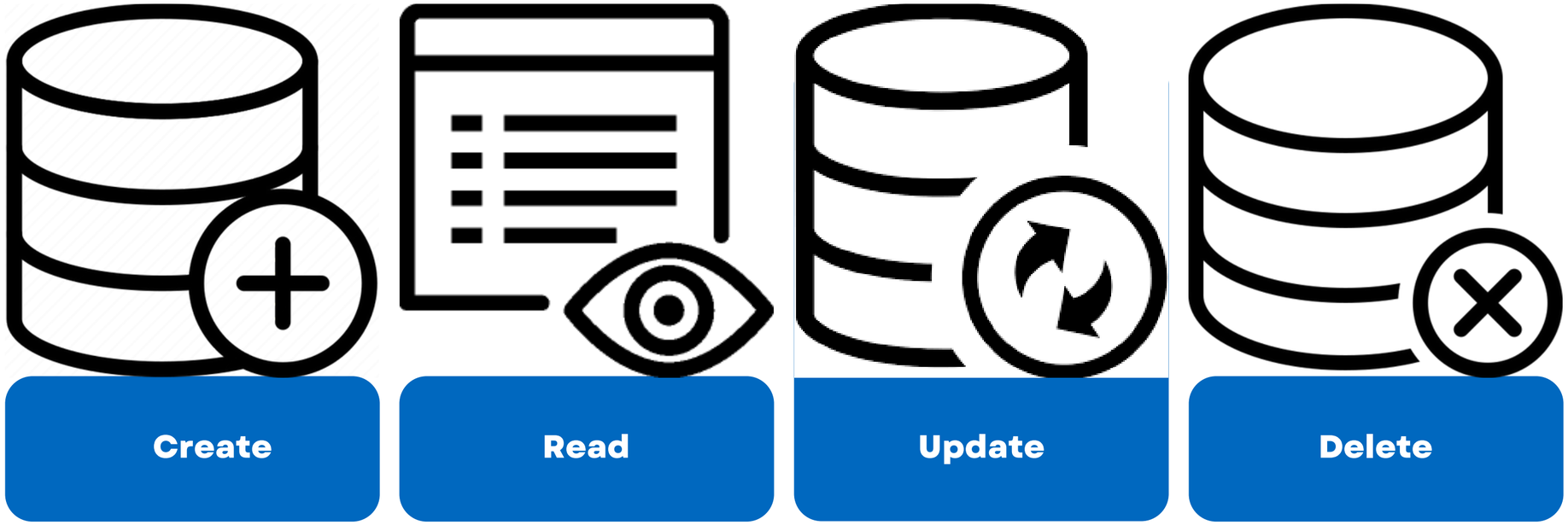
Elasticsearch is a distributed search and analytics engine built on Apache Lucene. Since its release in 2010, Elasticsearch has quickly become the most popular search engine and is commonly used for log analytics, full-text search, security intelligence, business analytics, and operational intelligence use cases.

What Is Kibana?

Kibana is a data visualization and exploration tool used for log and time-series analytics, application monitoring, and operational intelligence use cases. It offers powerful and easy-to-use features such as histograms, line graphs, pie charts, heat maps, and built-in geospatial support.

CRUD

In computer programming, create, read, update, and delete are the four basic operations of persistent storage. CRUD is also sometimes used to describe user interface conventions that facilitate viewing, searching, and changing information using computer-based forms and reports.



Document

A document is a unitary element, in JSON format, stored in ElasticSearch.

Node

A node is an instance of ElasticSearch, a node belongs to a cluster.

Cluster

A cluster is made up of one or more ElasticSearch nodes that are connected to each other and share the same name. A cluster has a single master node. In case of failure of the master node, a new master node is elected among the remaining nodes.

Index

An index is a logical grouping of a set of documents. An index is made up of shards. All documents belong to an index.

Type

A type is a subset of an index that is used to group documents. Similar to indexes, types are used to configure document storage. Every document belongs to a type.
⚠ Types are deprecated since version 6.x. Previously it was possible to have several types within the same index. For indexes created since version 6, only one type is allowed per index. For Elasticsearch versions < 6, prefer _doc as the type name so that the URLs are compatible with version 7.x.

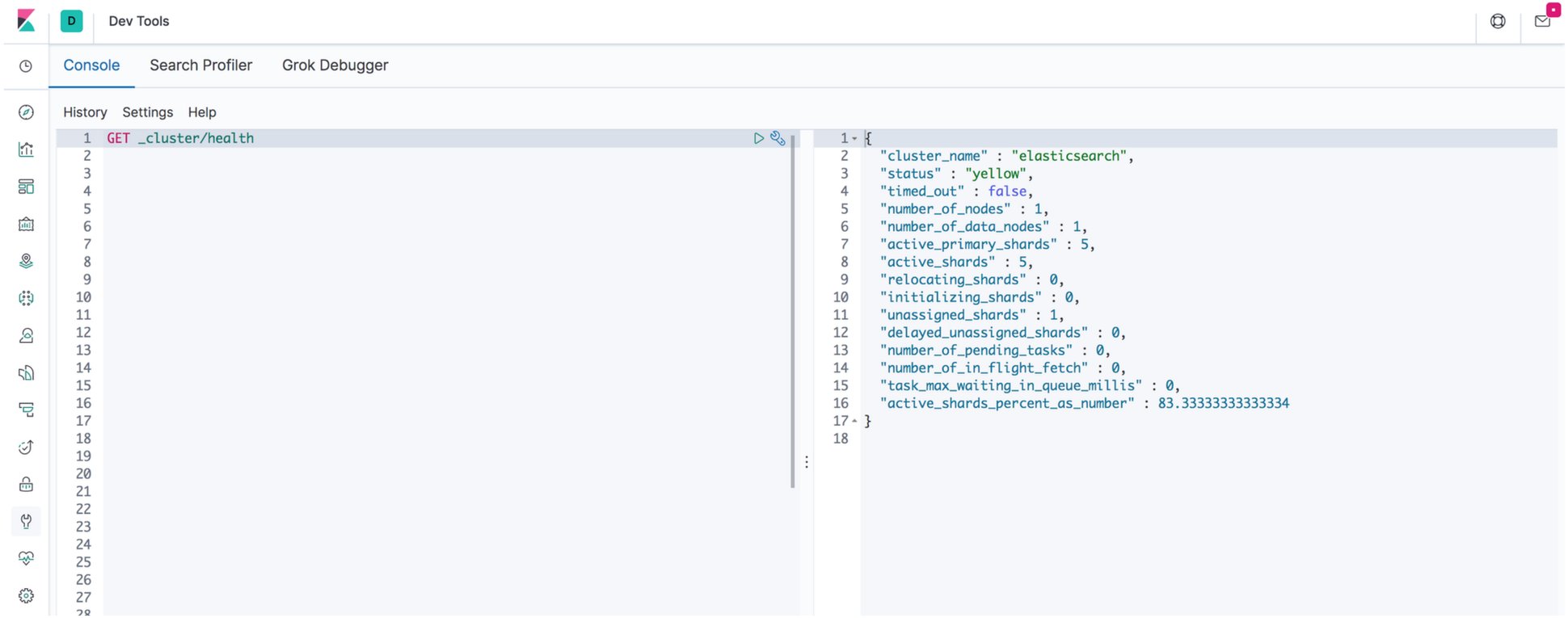
Shard

A shard is a fragment of an index. These are the shards which make it possible to partition the indexes on several nodes. Thus, an index can be partitioned on as many nodes as this index has shards. The default number of shards is 1 (it was 5 in versions of elasticsearch <= 6).

Replicate

A replicate is a full copy of the data in an index. Replicas help improve cluster fault tolerance and data durability. A replica has as many shards as the original index. The default number of replicas is 1.

Kibana's DevTools in Elastic Search to write commands



http://localhost:5601/app/dev_tools#/console

Create

```
curl -XPOST 'http://localhost:9200/heroes/_doc/ironman' -H 'Content-Type: application/json' -d '{
        "firstName" : "Tony",
        "lastName" : "Stark"
}'
```

Read

```
curl -XGET 'http://localhost:9200/heroes/_doc/ironman'
```

Update

```
curl -XPOST 'http://localhost:9200/heroes/_doc/ironman' -H 'Content-Type: application/json' -d '{
        "firstName" : "Tony",
        "lastName" : "Stark"
}'
```

Delete

```
curl -XDELETE 'http://localhost:9200/heroes/_doc/ironman'
```

Exists

```
curl --head 'http://localhost:9200/heroes/_doc/ironman'
```

- 200 indicates that the document exists
- 404 indicates that the document does not exist

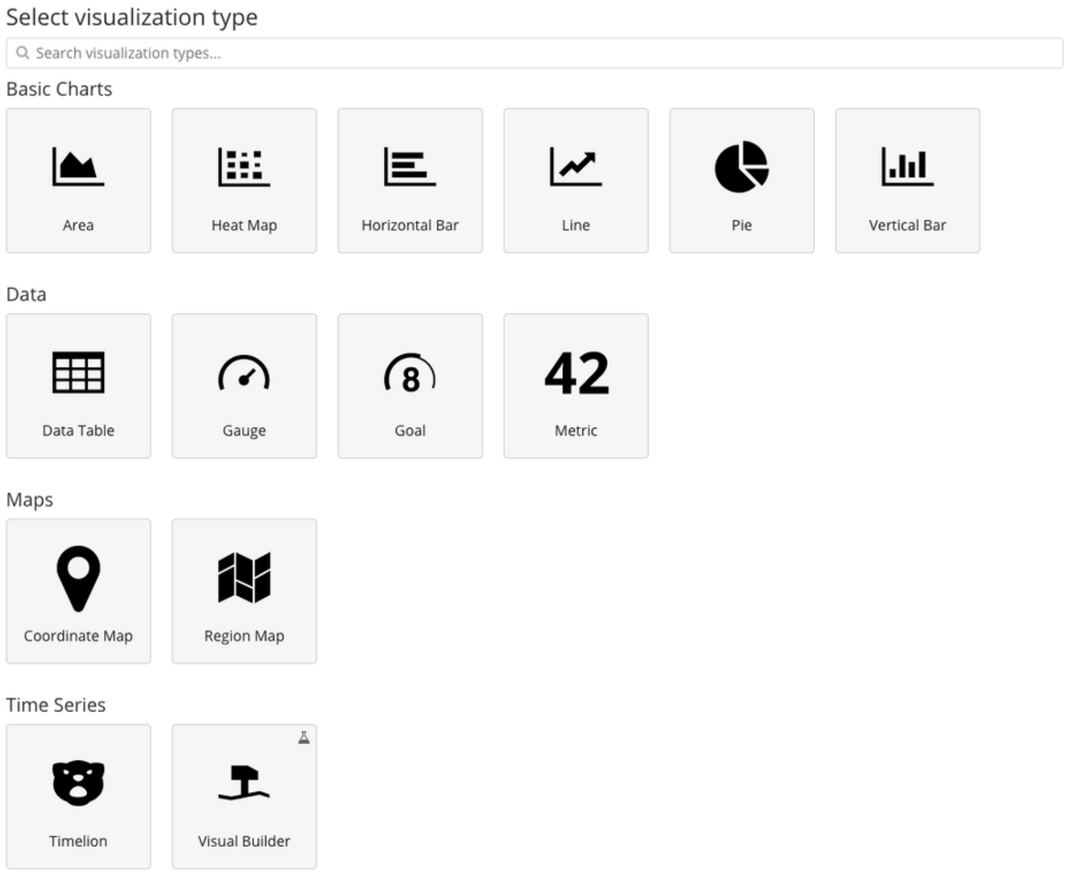
Visualization

You can create visualizations and build dashboards in Kibana to understand your data and share information.

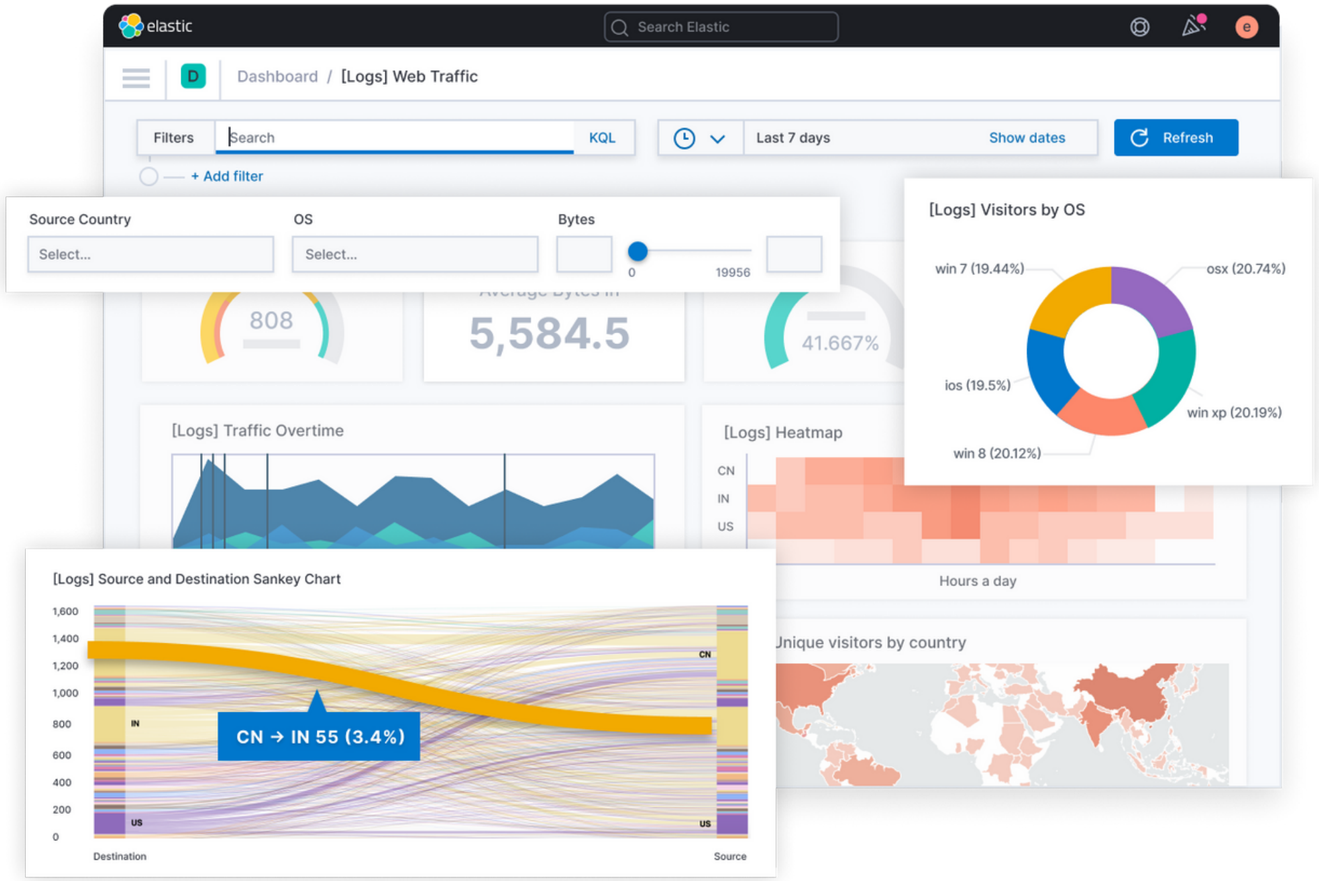
1. Open the Kibana main menu, then select Dashboard.
2. Select Create a Dashboard > Create visualization.
3. Drag and drop fields to create a visualization and then select Save and return. For example, to create a bar chart that shows the average balance by state:
    a. Drag the balance and state.keyword fields onto the workspace.
    b. From the Visualization type dropdown, select Bar horizontal.
    c. In the layer pane, select Top 5 values of state.keyword, edit the Name of the vertical axis, increase the number of states that are shown, then select Close.
    d. In the layer pane, select Median of balance, change the function to Average, edit the Name of the horizontal axis, then select Close.
4. Create more visualizations or select Save and return to save the dashboard.

Kibana Visualization Types

Kibana's visualization options are rich and varied and one of the reasons why users love the ELK Stack. Sometimes though, too much choice can actually complicate matters.



Visualization Examples



Conclusion

ElasticSearch is an excellent program but is not easy to setup specially for complex situations. In all the challenges presented on this article the main key is “documenting yourself” on how things works inside ElasticSearch and how to make things work as you expect.

About US

Arous Achraf

achraflarousse@gmail.com

Jmal Ahmed

ahmed.jemel2639@gmail.com

Thank You