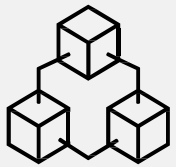
Modern Data Architectures

Elasticsearch

HO GENT

1

What is Elasticsearch



2

What is Elasticsearch / Elastic Search

Key elements:

- Distributed document store
- Search and analytics engine
- Has the ability to be schema-less
- First released in 2010 and rebranded to Elastic Stack

Goals:

- Near real-time search and analytics
- Structured & Unstructured data

3

What is Elasticsearch

Instead of storing information as rows of columnar data, Elasticsearch stores complex data structures that have been serialized as **JSON documents**.

When you have multiple Elasticsearch nodes in a cluster, **stored documents** are **distributed** across the cluster and can be accessed immediately from any node.

Elasticsearch uses a data structure called an **inverted index** that supports very fast full-text searches. An inverted index lists every unique word that appears in any document and identifies all of the documents each word occurs in.

Source: https://www.elastic.co/guide/en/elasticsearch/reference/current/elasticsearch-intro.html & https://www.elastic.co/guide/en/elasticsearch/reference/current/documents-indices.html

What is Elasticsearch

An index can be thought of as an optimized collection of documents and each document is a collection of fields, which are the key-value pairs that contain your data.

By default, Elasticsearch indexes all data in every field and each indexed field has a dedicated, optimized data structure.

For example, text fields are stored in inverted indices, and numeric and geo fields are stored in **BKD trees**. The ability to use the per-field data structures to assemble and return search results is what makes Elasticsearch so fast.

Source: https://www.elastic.co/guide/en/elasticsearch/reference/current/elasticsearch-intro.html & https://www.elastic.co/guide/en/elasticsearch/reference/current/documents-indices.html

5

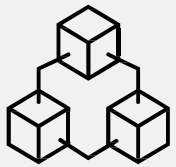
What is Elasticsearch

*In the **CAP theorem**, Elasticsearch falls into the **AP** category of the CAP theorem, which prioritizes availability and partition tolerance over consistency. In other words, in the event of a network partition or node failure, Elasticsearch will prioritize returning results quickly and ensuring that the system remains available over providing strong consistency guarantees. However, Elasticsearch still ensures **eventual consistency**, meaning that updates to data are eventually propagated to all nodes in the cluster.*

Source: https://medium.com/@TechTim42/elastic-search-and-open-search-a-brief-history-of-the-license-war-8f474743e2ff

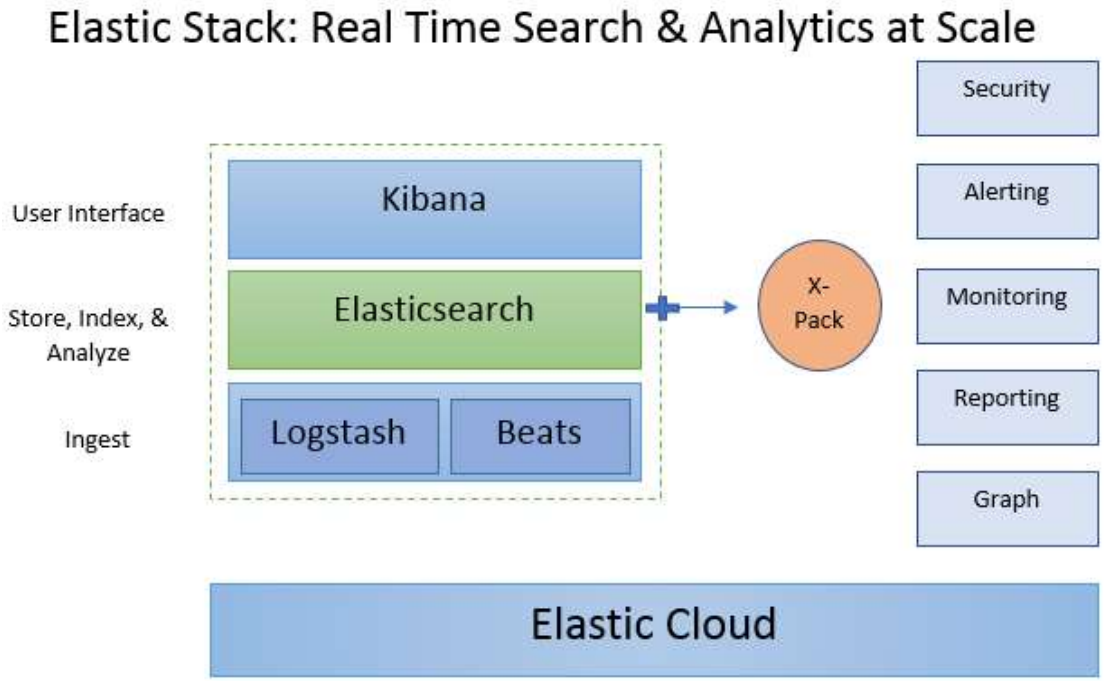
6

Key components of Elasticsearch



Components of Elasticsearch

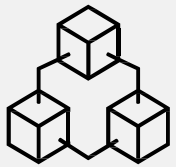
Multiple components of what is often referred to “ELK”-stack



- E(lasticsearch)
- L(ogstash)
- K(ibana)

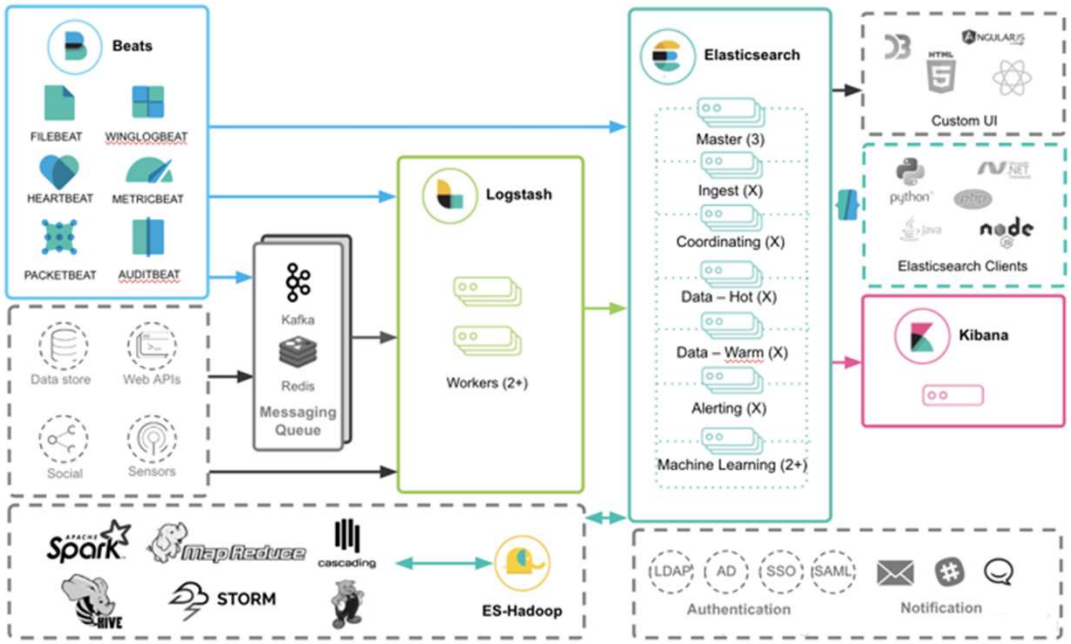
8

Often as part of a pipeline



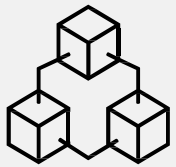
9

Combining technologies



Source: https://www.alibabacloud.com/blog/getting-started-with-beats_597070

Quick start



Quick start

Add a single document:

```
POST books/_doc {"name": "Snow Crash", "author": "Neal Stephenson",
"release_date": "1992-06-01", "page_count": 470}
```

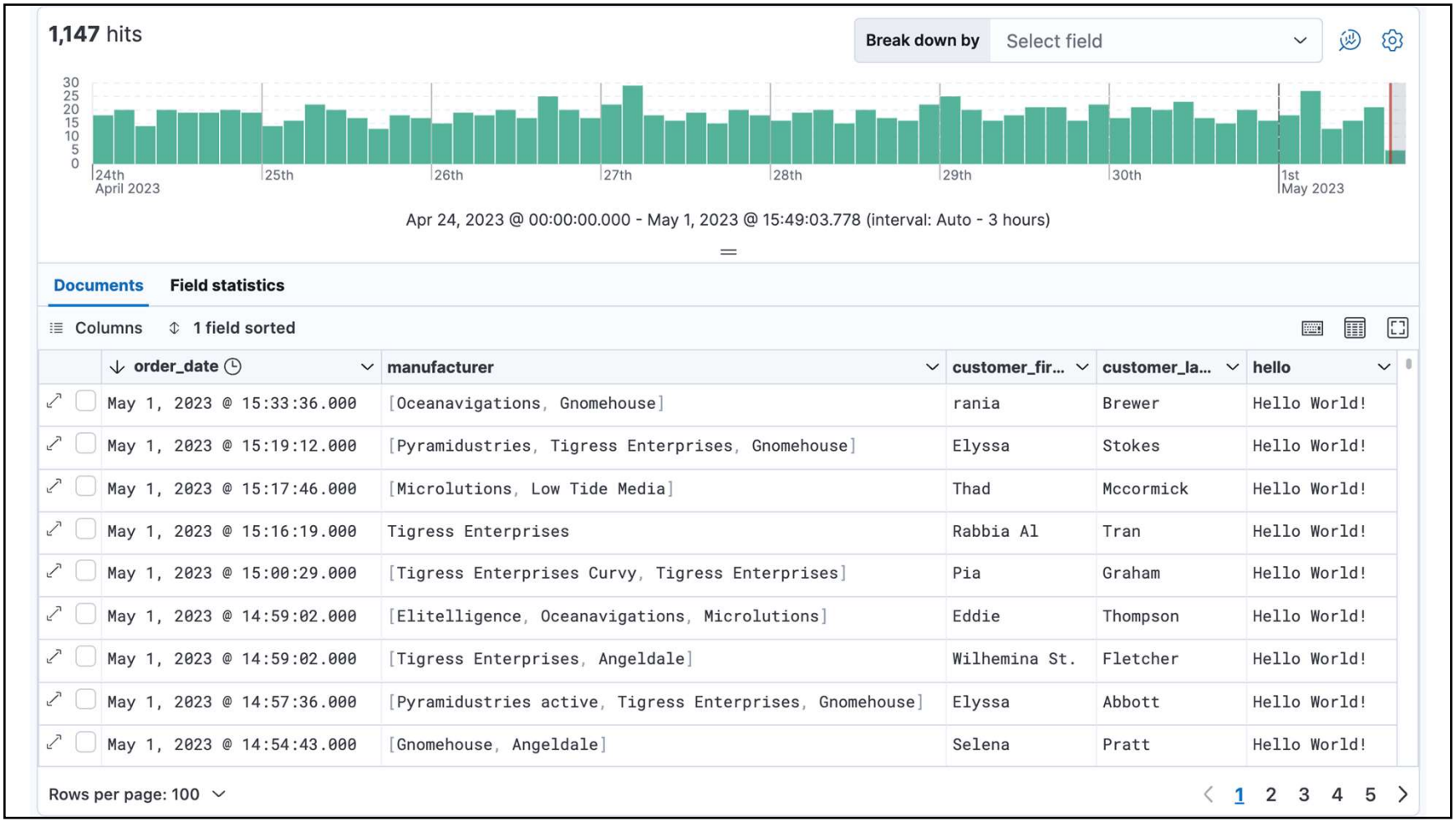
Search all documents:

```
GET books/_search
```

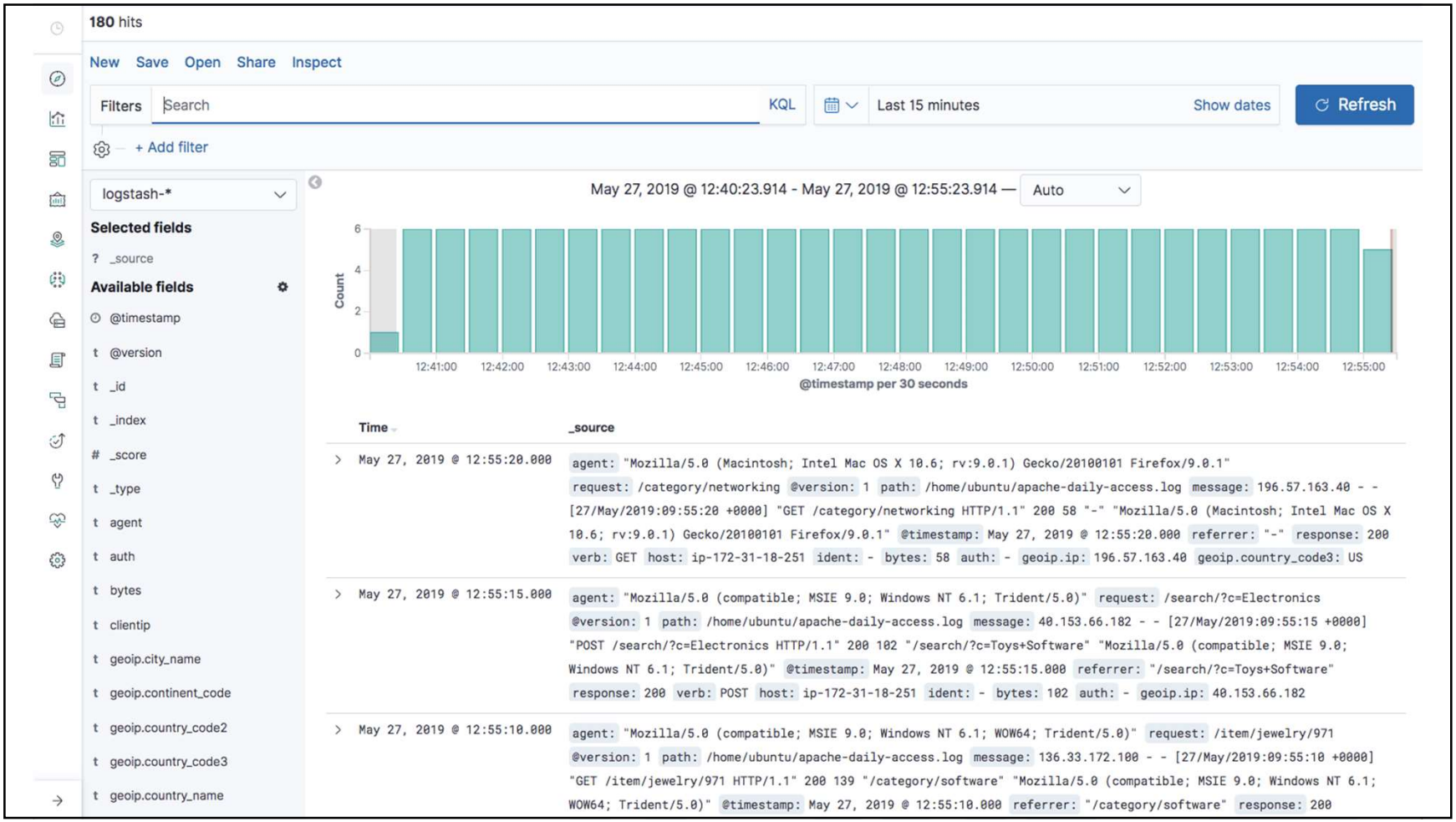
Match query:

```
GET books/_search
{
  "query": {
    "match": {
      "name": "brave"
    }
  }
}
```

→ Kibana

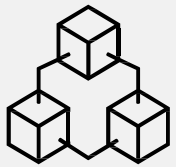

1,147 hits
Break down by
Select field
30
25
20
15
10
5
0
24th
April 2023
25th
26th
27th
28th
29th
30th
1st
May 2023
Apr 24, 2023 @ 00:00:00.000 - May 1, 2023 @ 15:49:03.778 (interval: Auto - 3 hours)
Documents
Field statistics
Columns
1 field sorted
order_date
manufacturer
customer_fir...
customer_la...
hello
May 1, 2023 @ 15:33:36.000
[Oceanavigations, Gnomehouse]
rania
Brewer
Hello World!
May 1, 2023 @ 15:19:12.000
[Pyramidustries, Tigress Enterprises, Gnomehouse]
Elyssa
Stokes
Hello World!
May 1, 2023 @ 15:17:46.000
[Microlutions, Low Tide Media]
Thad
Mccormick
Hello World!
May 1, 2023 @ 15:16:19.000
Tigress Enterprises
Rabbia Al
Tran
Hello World!
May 1, 2023 @ 15:00:29.000
[Tigress Enterprises Curvy, Tigress Enterprises]
Pia
Graham
Hello World!
May 1, 2023 @ 14:59:02.000
[Elitelligence, Oceanavigations, Microlutions]
Eddie
Thompson
Hello World!
May 1, 2023 @ 14:59:02.000
[Tigress Enterprises, Angeldale]
Wilhemina St.
Fletcher
Hello World!
May 1, 2023 @ 14:57:36.000
[Pyramidustries active, Tigress Enterprises, Gnomehouse]
Elyssa
Abbott
Hello World!
May 1, 2023 @ 14:54:43.000
[Gnomehouse, Angeldale]
Selena
Pratt
Hello World!
Rows per page: 100
1 2 3 4 5

13


180 hits
New Save Open Share Inspect
Filters Search KQL Last 15 minutes Show dates Refresh
+ Add filter
logstash-*
Selected fields
? _source
Available fields
@timestamp
t @version
t _id
t _index
_score
t _type
t agent
t auth
t bytes
t clientip
t geoip.city_name
t geoip.continent_code
t geoip.country_code2
t geoip.country_code3
t geoip.country_name
May 27, 2019 @ 12:40:23.914 - May 27, 2019 @ 12:55:23.914 — Auto
Count
6
4
2
0
12:41:00 12:42:00 12:43:00 12:44:00 12:45:00 12:46:00 12:47:00 12:48:00 12:49:00 12:50:00 12:51:00 12:52:00 12:53:00 12:54:00 12:55:00
@timestamp per 30 seconds
Time
_source
May 27, 2019 @ 12:55:20.000
agent: "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:9.0.1) Gecko/20100101 Firefox/9.0.1" request: /category/networking @version: 1 path: /home/ubuntu/apache-daily-access.log message: 196.57.163.40 - - [27/May/2019:09:55:20 +0000] "GET /category/networking HTTP/1.1" 200 58 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:9.0.1) Gecko/20100101 Firefox/9.0.1" @timestamp: May 27, 2019 @ 12:55:20.000 referrer: "-" response: 200 verb: GET host: ip-172-31-18-251 ident: - bytes: 58 auth: - geoip.ip: 196.57.163.40 geoip.country_code3: US
May 27, 2019 @ 12:55:15.000
agent: "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0)" request: /search/?c=Electronics @version: 1 path: /home/ubuntu/apache-daily-access.log message: 40.153.66.182 - - [27/May/2019:09:55:15 +0000] "POST /search/?c=Electronics HTTP/1.1" 200 102 "/search/?c=Toys+Software" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; Trident/5.0)" @timestamp: May 27, 2019 @ 12:55:15.000 referrer: "/search/?c=Toys+Software" response: 200 verb: POST host: ip-172-31-18-251 ident: - bytes: 102 auth: - geoip.ip: 40.153.66.182
May 27, 2019 @ 12:55:10.000
agent: "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)" request: /item/jewelry/971 @version: 1 path: /home/ubuntu/apache-daily-access.log message: 136.33.172.100 - - [27/May/2019:09:55:10 +0000] "GET /item/jewelry/971 HTTP/1.1" 200 139 "/category/software" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)" @timestamp: May 27, 2019 @ 12:55:10.000 referrer: "/category/software" response: 200

14

Use cases



15

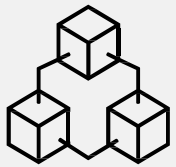
Use cases: “searching”

Whenever search operations, retrieval and reporting are crucial.

- Logging in applications
- E-commerce
- Security events – SIEM
  - Security Information and Event Management
  - “ELK”-stack
  - Opensearch / Wazuh
    - https://wazuh.com/

16

Elasticsearch vs Opensearch



17

Elasticsearch vs Open Search

Interesting war over “licenses”

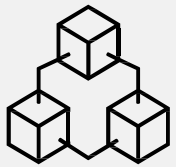
- 2014: Apache License (version 2.0) – permissive open source license
- 2018: “Elastic license” – restrictions added
- 2021: SSPL, similar to MongoDB (2019) → restrictions aimed to cloud providers

AWS is one of the largest users of Elastic Search and has contributed significantly to its development. In response to Elastic NV’s license change, **AWS announced that it would be forking Elastic Search and creating its own version of the software, called OpenSearch**

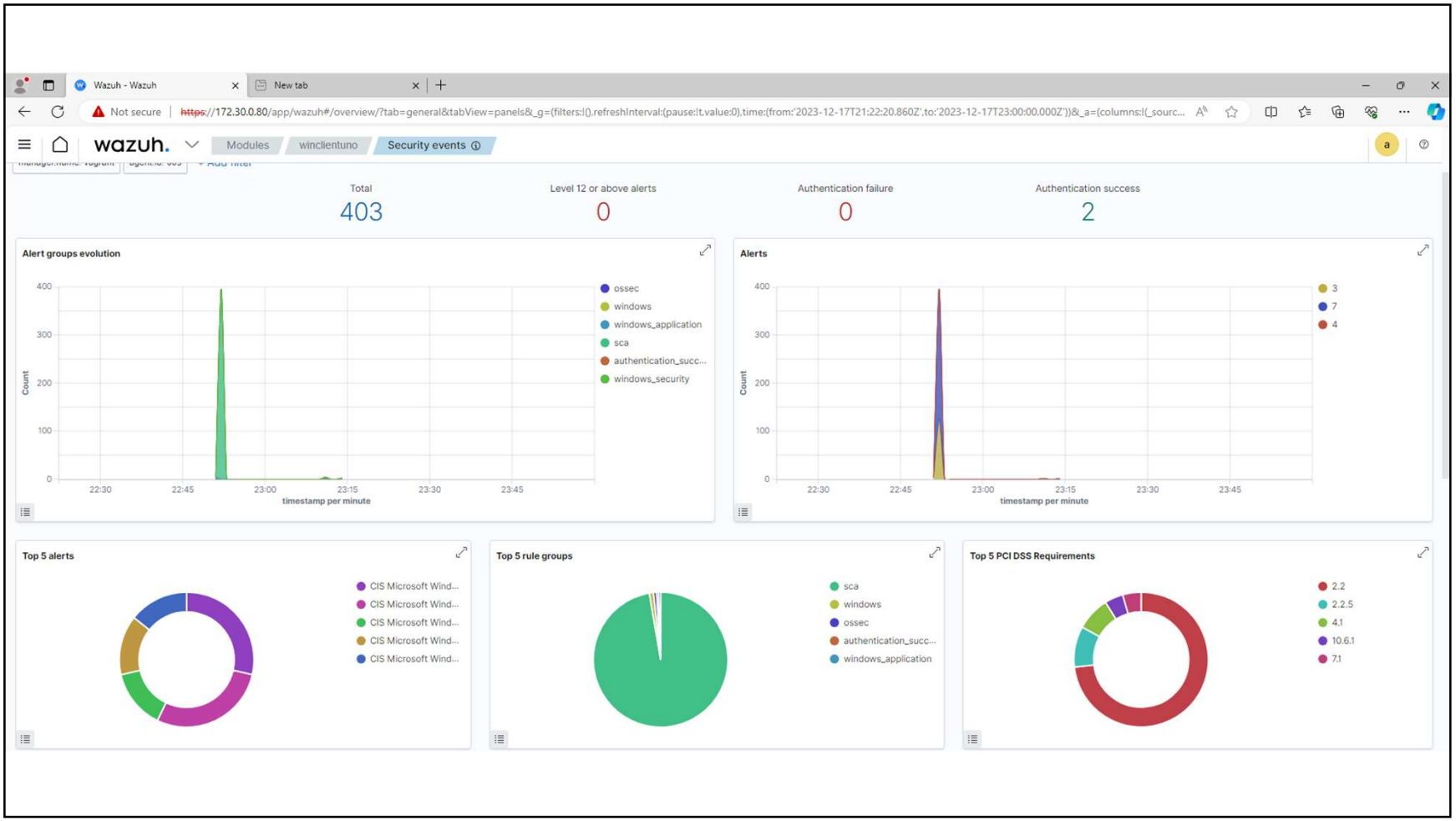
More information: https://medium.com/@TechTim42/elastic-search-and-open-search-a-brief-history-of-the-license-war-8f474743e2ff

18

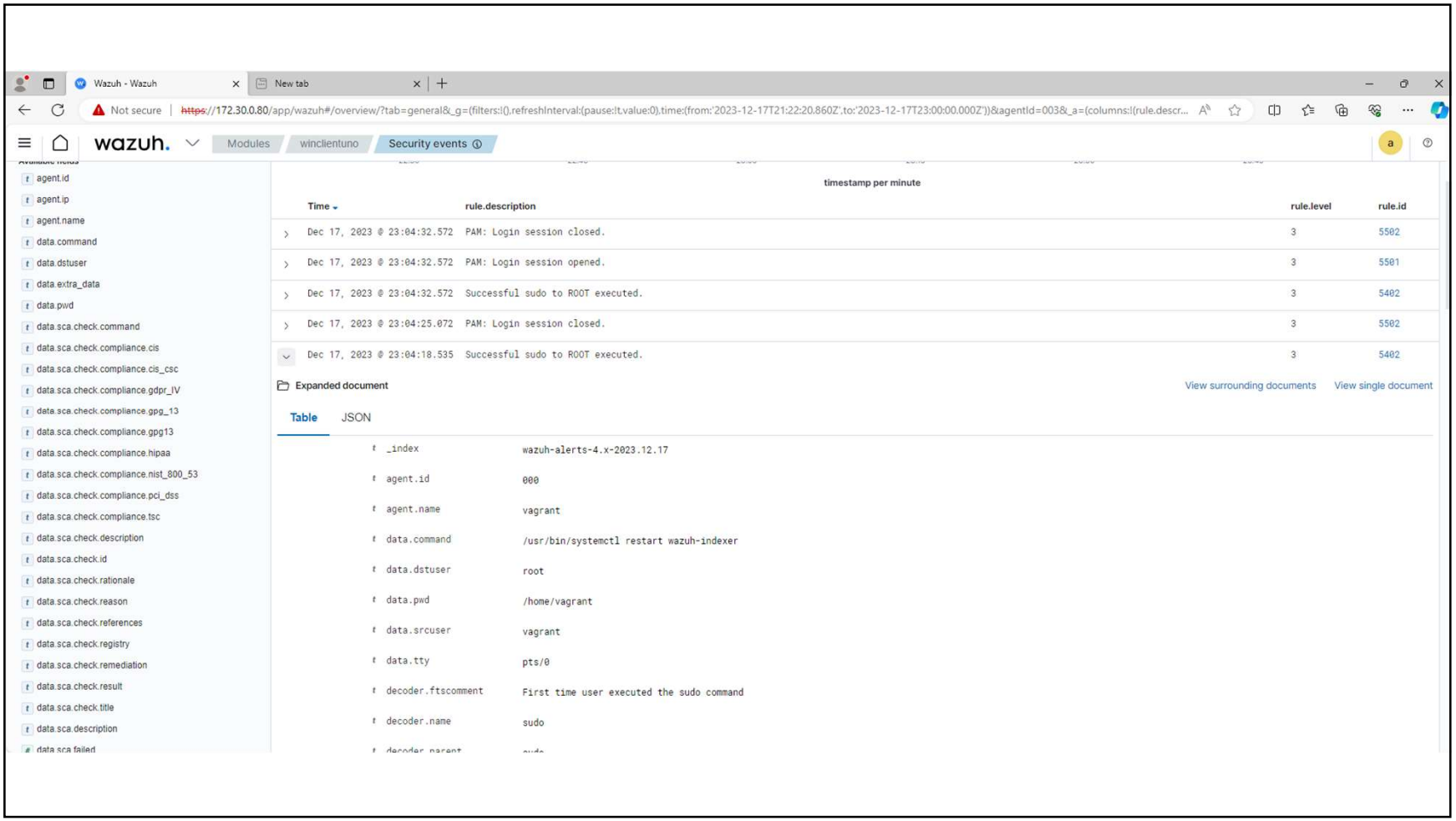
Demo



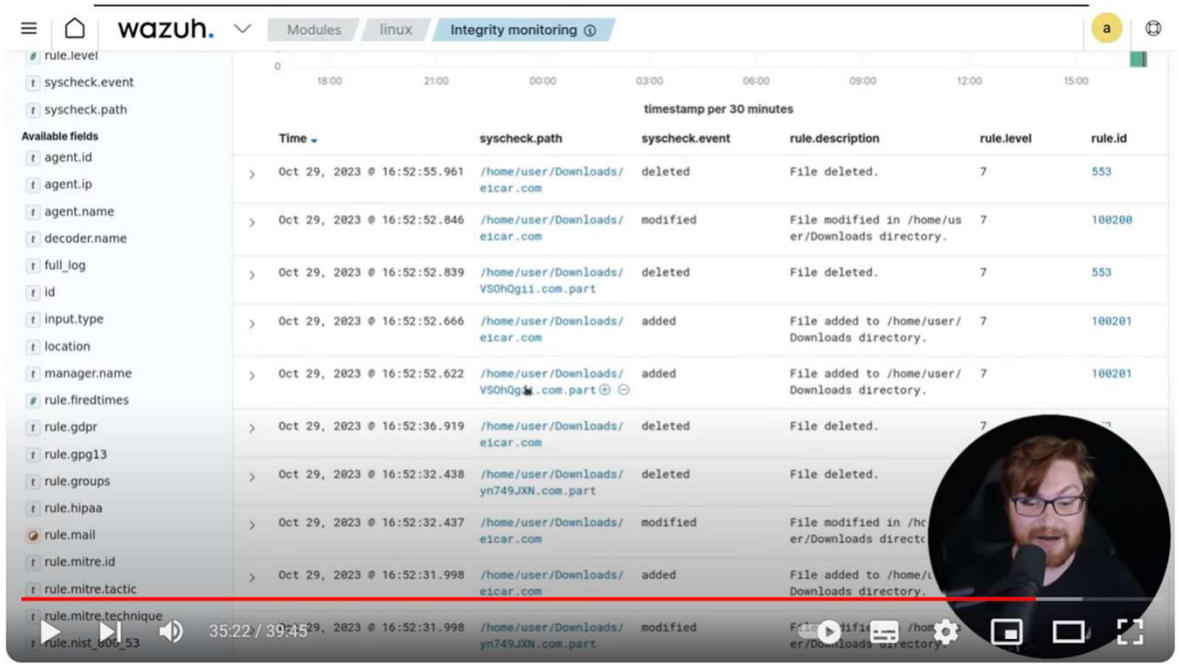
19


Wazuh - Wazuh
New tab
Not secure
https://172.30.0.80/app/wazuh#/overview/?tab=general&tabView=panels&_g=(filters:!(),refreshInterval:(pause:!t,value:0),time:(from:'2023-12-17T21:22:20.860Z',to:'2023-12-17T23:00:00.000Z'))&_a=(columns:!(_sourc...
wazuh.
Modules
winclientuno
Security events
Total
403
Level 12 or above alerts
0
Authentication failure
0
Authentication success
2
Alert groups evolution
Count
timestamp per minute
ossec
windows
windows_application
sca
authentication_succ...
windows_security
Alerts
3
7
4
Top 5 alerts
CIS Microsoft Wind...
Top 5 rule groups
sca
windows
ossec
authentication_succ...
windows_application
Top 5 PCI DSS Requirements
2.2
2.2.5
4.1
10.6.1
7.1

20


Wazuh - Wazuh
New tab
Not secure
https://172.30.0.80/app/wazuh#/overview/?tab=general&_g=(filters:!(),refreshInterval:(pause:!t,value:0),time:(from:'2023-12-17T21:22:20.860Z',to:'2023-12-17T23:00:00.000Z'))&agentId=003&_a=(columns:!(rule.descr...
wazuh.
Modules
winclientuno
Security events
agent.id
agent.ip
agent.name
data.command
data.dstuser
data.extra_data
data.pwd
data.sca.check.command
data.sca.check.compliance.cis
data.sca.check.compliance.cis_csc
data.sca.check.compliance.gdpr_IV
data.sca.check.compliance.gpg_13
data.sca.check.compliance.gpg13
data.sca.check.compliance.hipaa
data.sca.check.compliance.nist_800_53
data.sca.check.compliance.pci_dss
data.sca.check.compliance.tsc
data.sca.check.description
data.sca.check.id
data.sca.check.rationale
data.sca.check.reason
data.sca.check.references
data.sca.check.registry
data.sca.check.remediation
data.sca.check.result
data.sca.check.title
data.sca.description
data.sca.failed
timestamp per minute
Time
rule.description
rule.level
rule.id
Dec 17, 2023 @ 23:04:32.572 PAM: Login session closed. 3 5502
Dec 17, 2023 @ 23:04:32.572 PAM: Login session opened. 3 5501
Dec 17, 2023 @ 23:04:32.572 Successful sudo to ROOT executed. 3 5402
Dec 17, 2023 @ 23:04:25.072 PAM: Login session closed. 3 5502
Dec 17, 2023 @ 23:04:18.535 Successful sudo to ROOT executed. 3 5402
Expanded document
View surrounding documents
View single document
Table
JSON
_index wazuh-alerts-4.x-2023.12.17
agent.id 000
agent.name vagrant
data.command /usr/bin/systemctl restart wazuh-indexer
data.dstuser root
data.pwd /home/vagrant
data.srcuser vagrant
data.tty pts/0
decoder.ftscomment First time user executed the sudo command
decoder.name sudo

21



https://www.youtube.com/watch?v=i68atPbB8uQ

22