download

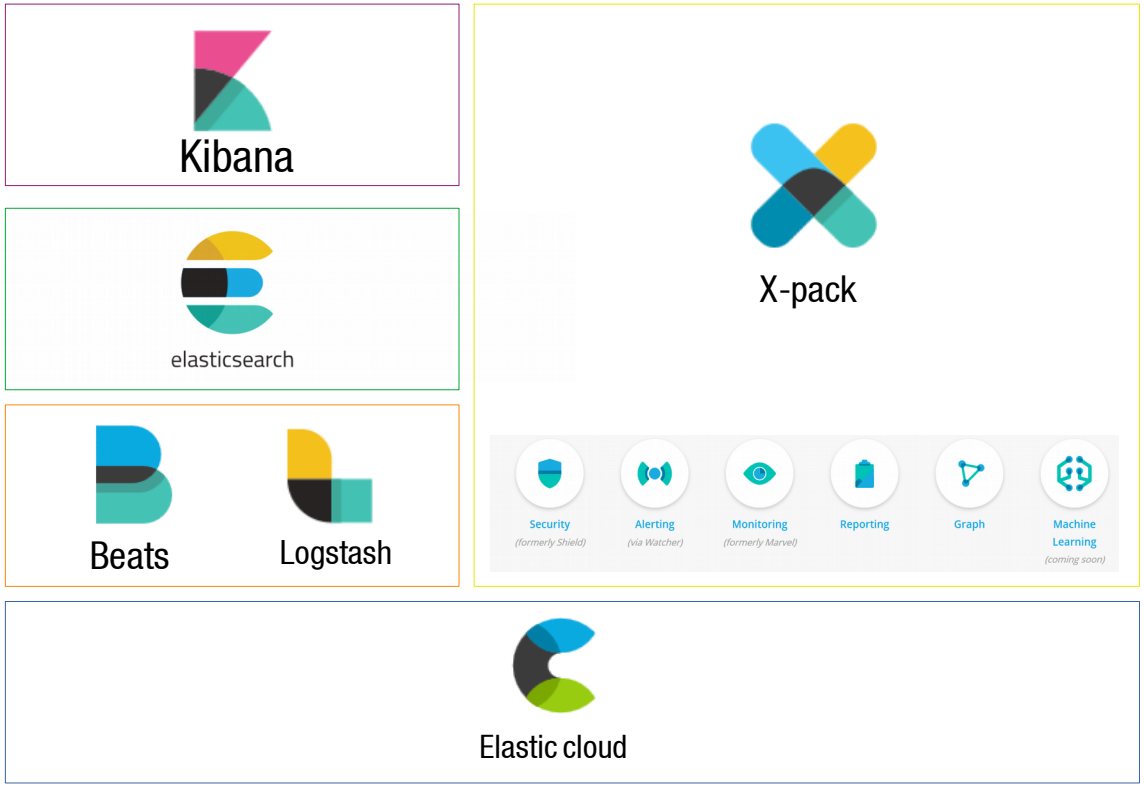
DATA ANALYSIS WITH THE ELASTIC STACK

Mišo Mijatović
Security DevOps @ Sorint.SEC

elastic

SORINT.SEC SORINT.lab

ELASTIC STACK



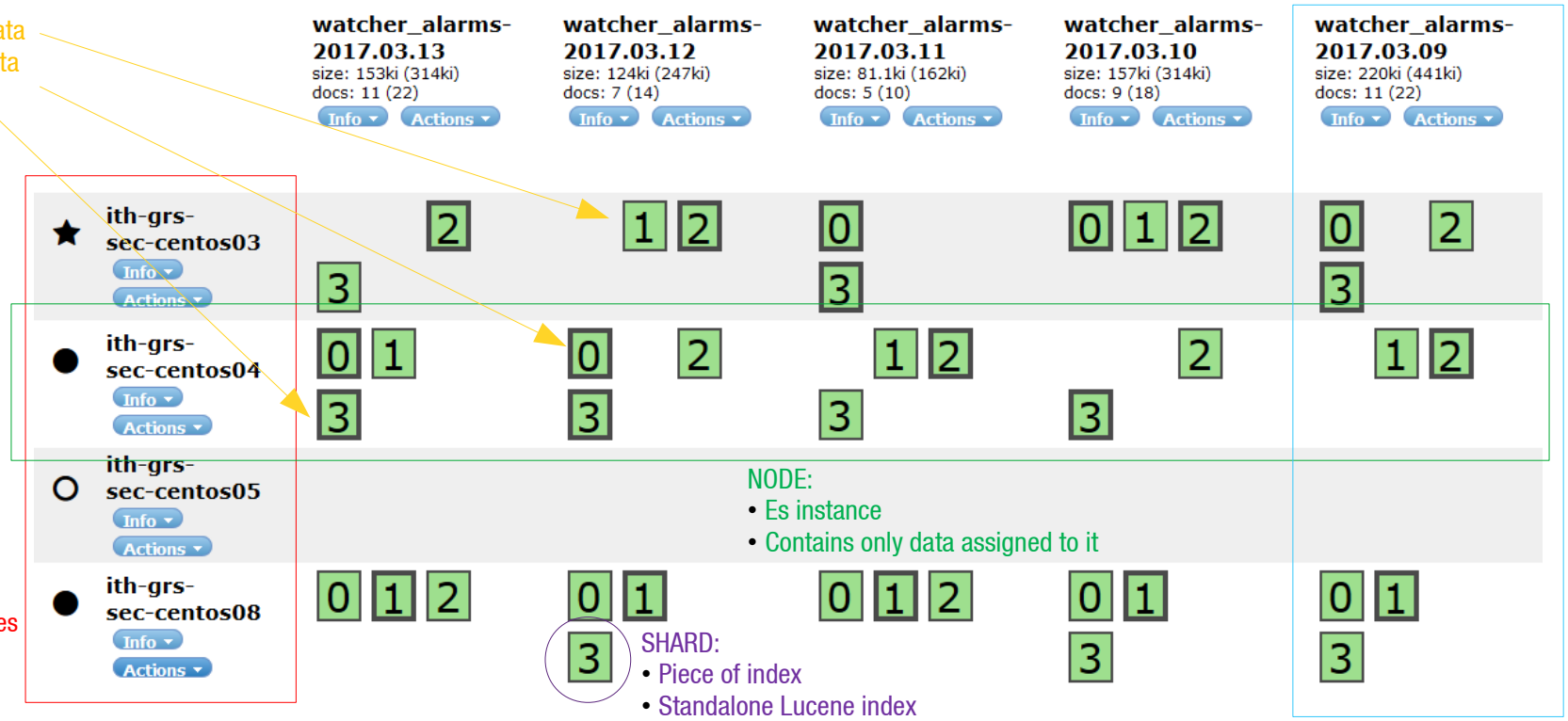
ELASTICSEARCH: WHAT

- Distributed data storage & information retrieval platform
  - Search engine / aggregation engine / analytics
  - Suggestions / percolation / highlighting / geo
  - Document store
  - Horizontally distributed
  - Highly available
  - Real time & near real time
- Open Source
  - Apache License
  - Proprietary plugins available for security & messaging
- Simple RESTful API
- Fast document retrieval
- Versatile - Many popular use cases
  - Log data analysis
  - Event data
  - Full text search
  - Analytics & Aggregations
  - Data visualization w/ Kibana
  - Alerting & classification w/ Percolator
  - Suggestion engine

ELASTICSEARCH: HOW

DOCUMENTS
- JSON objects
- Contains metadata
- Contains raw data

INDEX:
- Container of data
- Document -> indexed
- Each node hosts indices' shards



NODE:
- Es instance
- Contains only data assigned to it

CLUSTER:
- Horizontal scale
- Multiple es instances
- Distributed
- Ports 9300 9200

SHARD:
- Piece of index
- Standalone Lucene index
- Either primary or replica

ELASTICSEARCH: MORE INFO

- Shards replication
- Cluster status: green yellow red
- Node types: master, data, client, …
- Mappings & Templates
- Aliases

REST APIs POST, GET, PUT, …

_cat _cluster _search _script _analyze
_template _mapping _alias
_settings _stats _nodes _segments

And many more…

- Documents are JSON objects
- _source contains the raw data
- Every document also contains metadata
- Elasticsearch Datatypes
  - String - "Hello!"
  - Number - 42
  - Geopoint / Geoshape - [52.3751349, 4.8694318]
  - Bool - true | false
  - Date - 1979-01-10 06:33:03
  - IPAddress - 54.149.131.206
  - Object - {"person": {"name":"Joe Smith", "age":37} }

TEXT ANALYSIS & INVERTED INDEX

1. The quick brown fox jumps over the lazy dog
2. Fast jumping spiders

Tokenizer →

2	Fast
1	The
1	brown
1	dog
1	fox
2	jumping
1	jumps
1	lazy
1	over
1	quick
2	spiders
1	the

Stopwords →

2	Fast
1	brown
1	dog
1	fox
2	jumping
1	jumps
1	lazy
1	over
1	quick
2	spiders

Lowercasing →

1	brown
1	dog
1	fast
1	fox
2	jumping
1	jumps
1	lazy
1	over
1	quick
2	spiders

Stemming →

1	brown
1	dog
1	fast
1	fox
1,2	jump
1	lazy
1	over
1	quick
2	spiders

Synonyms →

1	brown
1	dog
1,2	fast
1	fox
1,2	jump
1	lazy
1	over
1,2	quick
2	spiders

...and more

Process pre indexing but also prior to searching!

LOGSTASH & BEATS

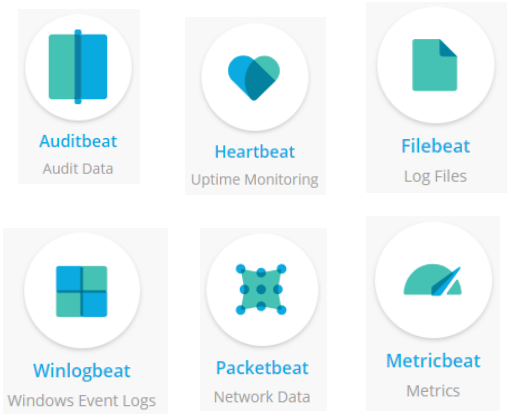
LOGSTASH

- Data shipper
  - Collect
  - Parse
  - Transform
- Has many plugins out of the box - pluggable
- Processing pipeline

BEATS

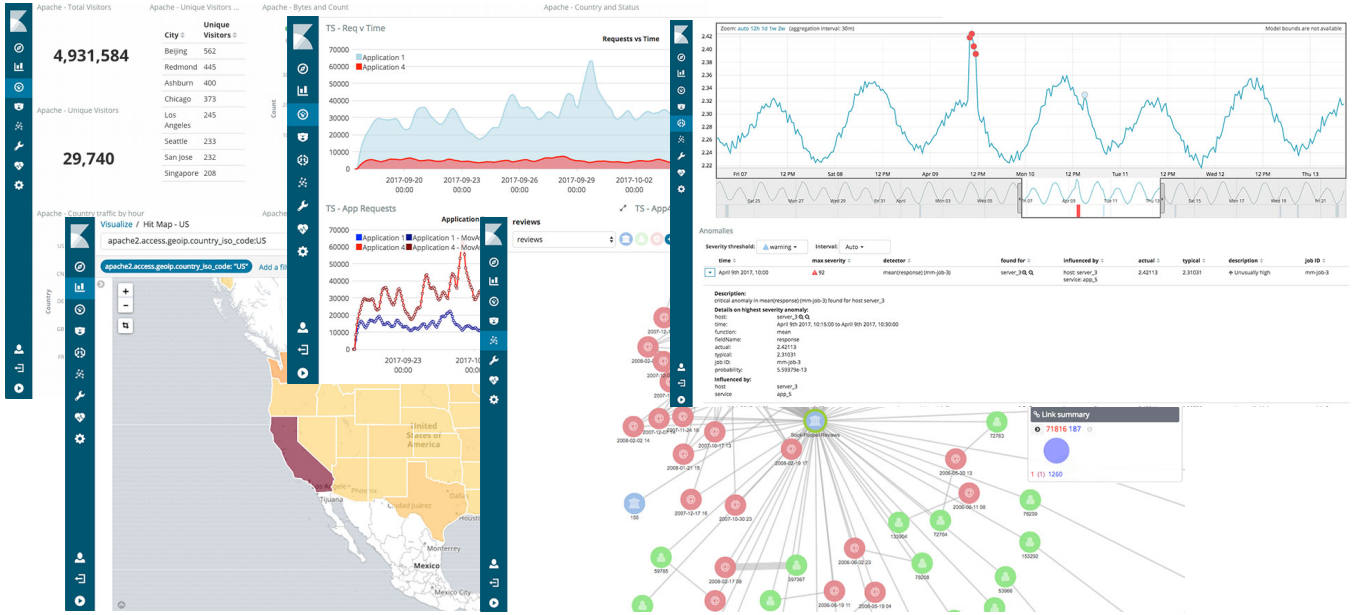
- Lightweight data shipper
- Single purpose
- Modules

aggregate, csv, date, DNS, Elasticsearch geoip, json, grok, kv, mutate, urldecode, …. many more



KIBANA

The data visualization framework
Let’s you visualize your Elasticsearch data and navigate the Elastic Stack



Flexible
Pluggable
Dynamic
Powerful
Search!

download
elastic
SORINTsec SORINTlab

DEMO
download
elastic
SORINTsec
SORINTlab

CONTACTS & MORE

ELASTIC

https://www.elastic.co

https://github.com/elastic

https://discuss.elastic.co/

Mišo Mijatović

mmijatovic@sorint.it

misomij@gmail.com

https://github.com/MisoMadao

Dataset by

kaggle

https://www.kaggle.com/misomij

DEMO HERE:

https://github.com/MisoMadao/download2018

download

Thanks!

mmijatovic@sorint.it

In collaborazione con:

BERGAMO CITTÀ DEI MILLE
COMUNE DI BERGAMO

SORINT SEC

SORINT lab