Log all the things!

elastic

Honza Král
@honzakral

Logs?

Events!

Log lines

Twitter feed

Invoices

Metrics

Why?

What happened last Tuesday?

Grep?

Multiple machines

Multiple logs

Analysis/Discovery

Time period

Time? Time?! Time!

apache	`[23/Jan/2014:17:11:55 +0000]`
unix timestamp	`1390994740`
log4j	`[2014-01-29 12:28:25,470]`
postfix.log	`Feb 3 20:37:35`
ISO 8601	`2009-01-01T12:00:00+01:00`

Correlate events

Web Server logs VS Load Balancer

- see immediately that caching is off
- static files leaking to gunicorn

Web Server VS Database

500s VS Deploys

- new version has a bug

Traffic VS Ad Campaigns

Ideal state

Central storage

Even for data from different systems

Enriched data

IP -> location, hostname

URL -> author, product, category

Search

user:honza status:404

Analysis

Visualisations for easy pattern discovery

Centralised Logging

Steps

Collect data

Parse data

Enrich data

Store data

Search and aggregate

Visualize data

Elastic Stack

Steps in Elastic Stack

Collect data

Parse data

Enrich data

Store data

Search and aggregate

Visualize data

beats

logstash

elasticsearch

kibana

Steps in Elastic Stack

Collect data

Parse data

Enrich data

Store data

Search and aggregate

Visualize data

beats

logstash

elasticsearch

kibana

beats

beats

```
metricbeat:
  modules:
    - module: redis
      metricsets: ["info"]
      hosts: ["host1"]
      period: 1s
      enabled: true
    - module: apache
      metricsets: ["info"]
      hosts: ["host1"]
      period: 30s
      enabled: true
```

```
filebeat:
  prospectors:
    - paths:
        - "logs/acc
      document_type: access
      multiline:
        pattern: ^#
        negate: true
        match: after
```

```
protocols:
  http:
    ports: [80, 8000]

  mysql:
    ports: [3306]

  redis:
    ports: [6379]

  pgsql:
    ports: [5432]

  thrift:
    ports: [9090]
```

```
output:
  logstash:
    hosts: ["localhost:5044"]
```

logstash

Inputs

Monitoring	collectd, graphite, ganglia, snmptrap, zenoss
Datastores	elasticsearch, redis, sqlite, s3
Queues	kafka, rabbitmq, zeromq
Logging	beats, eventlog, gelf, log4j, relp, syslog, varnish log
Platforms	drupal_dblog, gemfire, heroku, sqs, s3, twitter
Local	exec, generator, file, stdin, pipe, unix
Protocol	imap, irc, stomp, tcp, udp, websocket, wmi, xmpp

Filters

aggregate alter anonymize collate csv cidr clone cipher checksum date dns drop elasticsearch extractnumbers environment elapsed fingerprint geoip grok i18n json json_encode kv mutate metrics multiline metaevent prune punct ruby range syslog_pri sleep split throttle translate uuid urldecode useragent xml zeromq …

Outputs

Store	elasticsearch, gemfire, mongodb, redis, riak, rabbitmq, solr
Monitoring	ganglia, graphite, graphtastic, nagios, opentsdb, statsd, zabbix
Notification	email, hipchat, irc, pagerduty, sns
Protocol	gelf, http, lumberjack, metriccatcher, stomp, tcp, udp, websocket, xmpp
External service	google big query, google cloud storage, jira, loggly, riemann, s3, sqs, syslog, datadog
External monitoring	boundary, circonus, cloudwatch, librato
Local	csv, dots, exec, file, pipe, stdout, null

elasticsearch

Distributed Search Engine

Open Source

Document-based

Based on Lucene

JSON over HTTP

Data Management

Cluster

Collection of Nodes

Index

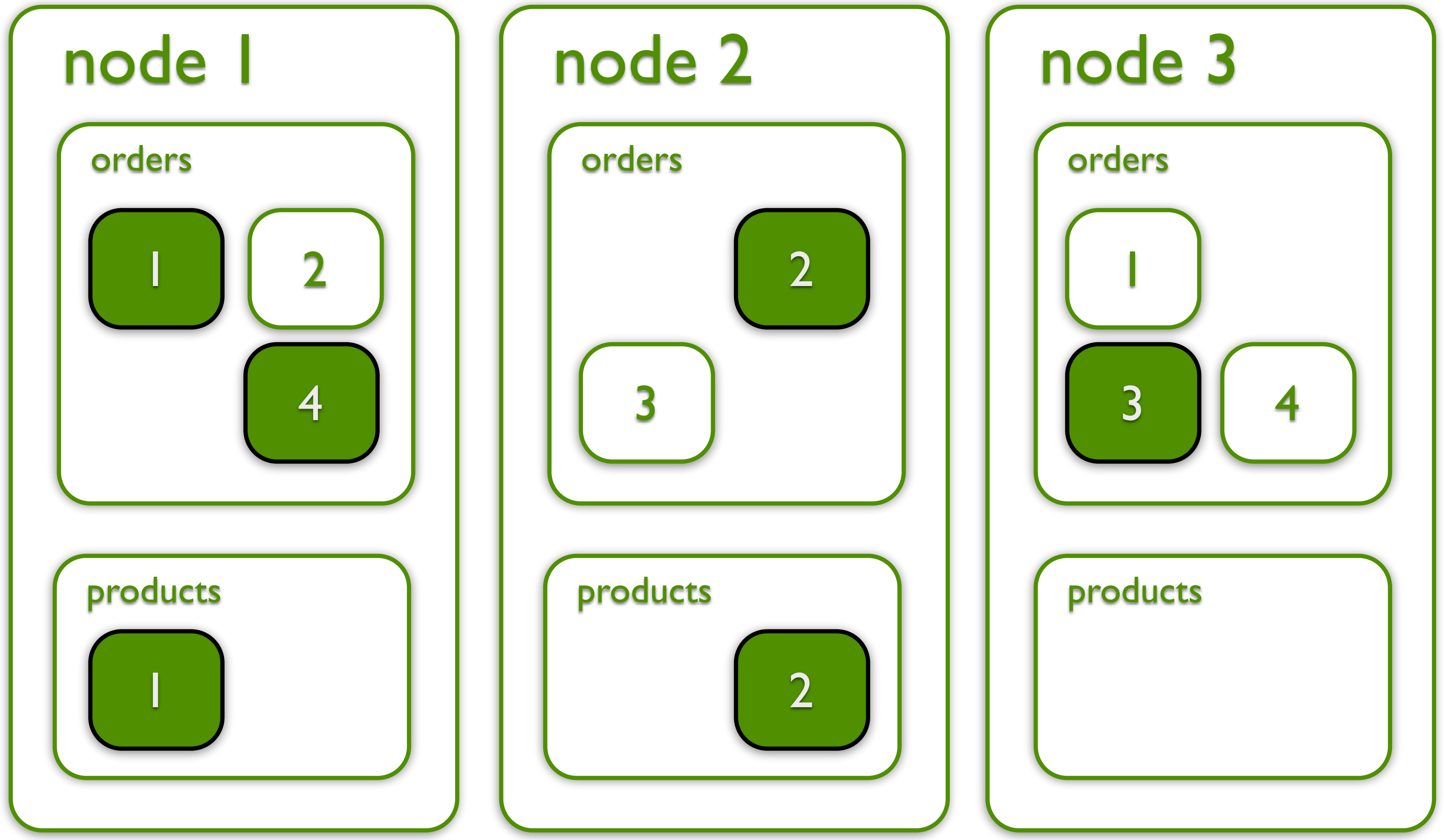
Collection of Shards

Shard

Unit of scale

Distributed across cluster

Primary and replica



Time based data flow

Current	replicas to speed up search on stronger boxes
Week old	snapshot keep only 1 replica
Month old	move to weaker boxes
2 months	close the indices
3 months	delete

kibana

Discover
Visualize
Dashboard
Settings
Last 36h
Site Status
warning
Tags over time
250
200
150
100
50
0
Count of documents
03:00
06:00
09:00
12:00
15:00
18:00
21:00
00:00
03:00
06:00
09:00
12:00
@timestamp per 30 min
● info
● success
● security
● warning
● error
● login
Top 10 counties by authorization status
Sum of file sizes
3M
2M
1M
0
Sum of bytes
1000
3000
5000
7000
9000
bytes
Tagged countries by auth status
CN: geo.src
IN: geo.src
US: geo.src
ID: geo.src
PK: geo.src
BR: geo.src
NG: geo.src
RU: geo.src
BD: geo.src
JP: geo.src
GB: geo.src
EG: geo.src

kibana
Discover
Visualize
Dashboard
Settings
Last 2 years
*
usagov
Data
Options
Greenland
Barents Sea
Siberia
ARCTIC CIRCLE
ASIA
NORTH AMERICA
North Atlantic Ocean
EUROPE
Gobi
TROPIC OF CANCER
Sahara
Arabian Peninsula
AFRICA
EQUATOR
Amazon Basin
Congo Basin
Indian Ocean
Indonesia
TROPIC OF CAPRICORN
SOUTH AMERICA
South Atlantic Ocean
AUSTRA
South Pacific Ocean

view options

Map type: Heatmap

Radius: 8

Blur: 7

Maximum zoom: 10

Minimum opacity: 0.16

☑ Normalize data for heatmap intesity

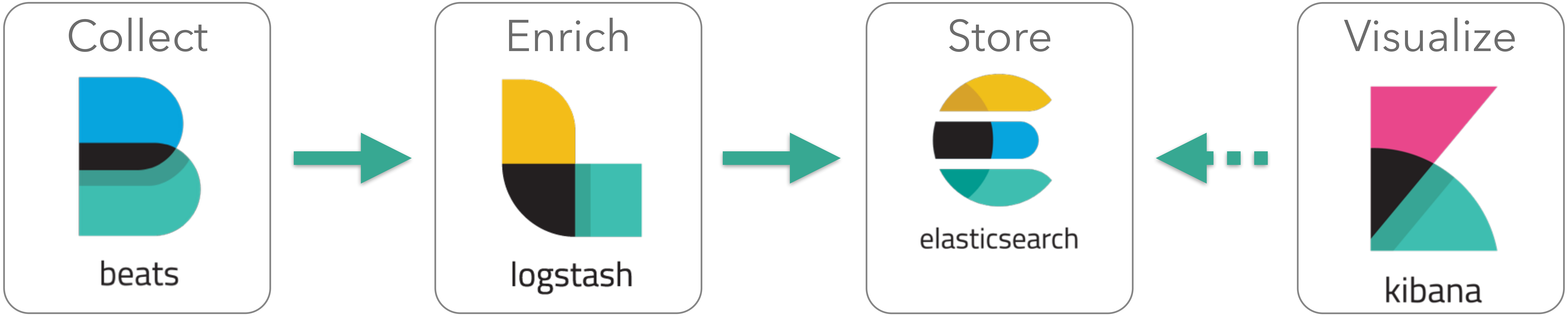
☑ Show Tooltip

☑ Desaturate map tiles

kibana
Discover
Visualize
Dashboard
Settings
October 1st 2013, 00:00:00.000 to December 31st 2014, 00:00:00.000
*
bytes: "400KB to 600KB"
user_agent.agent: "IE 8.0.0"
Actions ▸
usagov
Selected Fields
bytes
destination
Available Fields
@timestamp
_id
_index
_type
creation_time
geo.city
geo.coordinates
geo.country_code
geo.region
referrer
request
timezone
user
user_agent.agent
user_agent.agent_version
Everything
10,139 hits
October 1st 2013, 00:00:00.000 - December 31st 2014, 00:00:00.000 — by week
Count
400
300
200
100
0
2013-10-31
2013-11-30
2013-12-31
2014-01-31
2014-03-31
2014-04-30
2014-05-31
2014-06-30
2014-07-31
2014-08-31
2014-09-30
2014-10-31
2014-11-30
@timestamp per week

Time ▾	bytes	destination
▸ December 30th 2014, 23:56:37.218	576.685KB	http://www.ncbi.nlm.nih.gov/pubmed/17636814
▸ December 30th 2014, 23:35:54.218	411.201KB	http://www.ssd.noaa.gov/PS/TROP/TCFP/data/current/mtloopfpr17.gif
▸ December 30th 2014, 22:59:19.218	598.749KB	http://www.army.mil/rotc
▸ December 30th 2014, 22:06:15.218	486.421KB	http://www.usajobs.gov/
▸ December 30th 2014, 22:05:36.218	448.31KB	http://portal.hud.gov/portal/page/portal/HUD/program_offices/cpo
▸ December 30th 2014, 21:59:52.218	411.201KB	http://portal.hud.gov/hudportal/HUD?src=/program_offices/administration/dds
▸ December 30th 2014, 21:40:22.218	400.169KB	http://www.hud.gov/katrina/citizens.cfm

Architecture



Logging and Python

Enhance your logs

Track metrics

- execution time
- query time
- # of queries

Include metadata

- user_id
- content

Log as JSON

Structlog

Add structured info

Track info through services

Log to file

Add filebeat to read the file

Thanks!

elastic

Honza Král
@honzakral