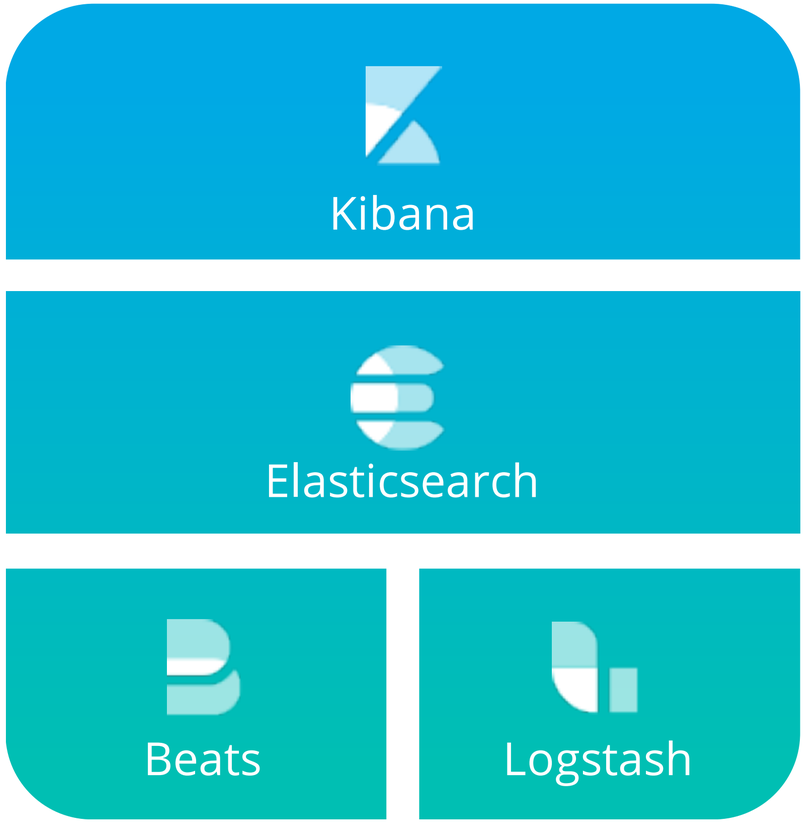
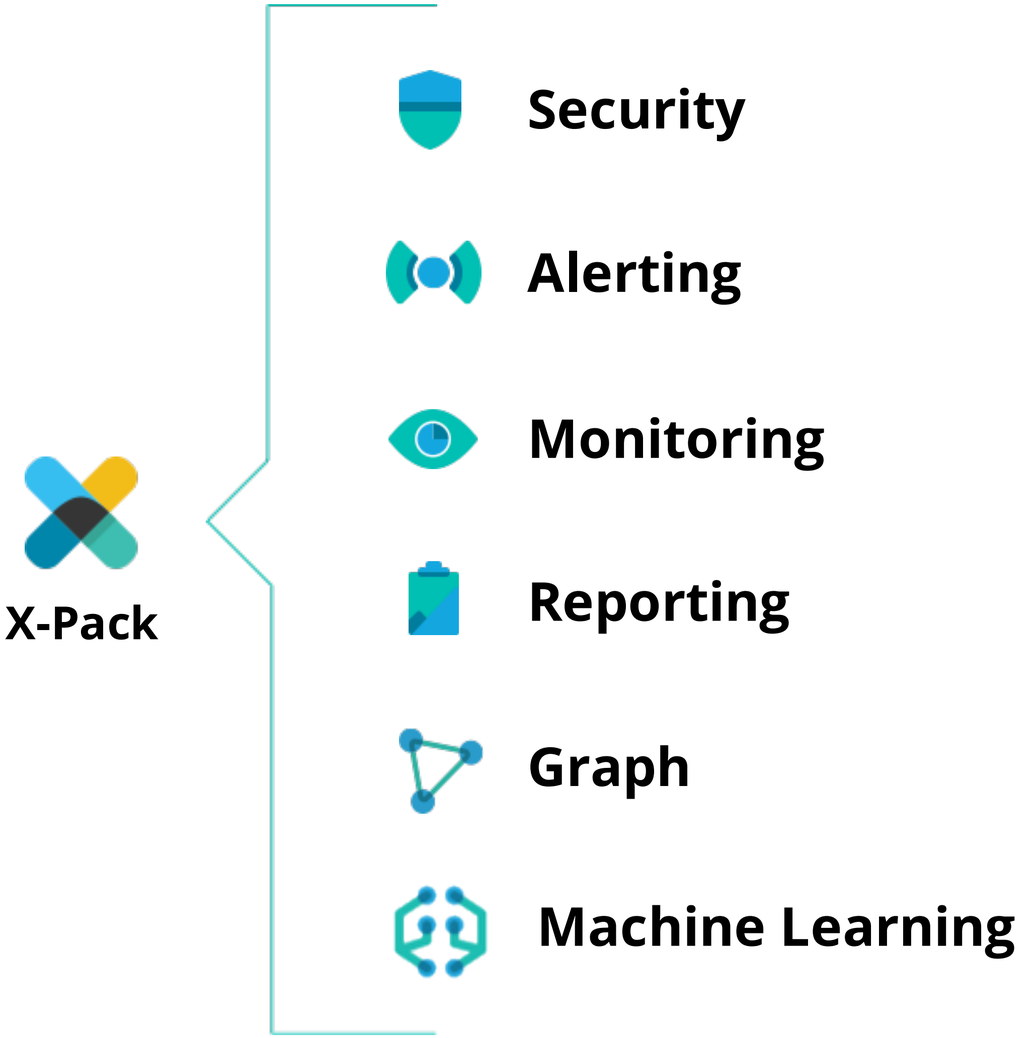
The State of Elastic Stack

What's new and what we are working on

Medcl

Elastic Stack & X-Pack



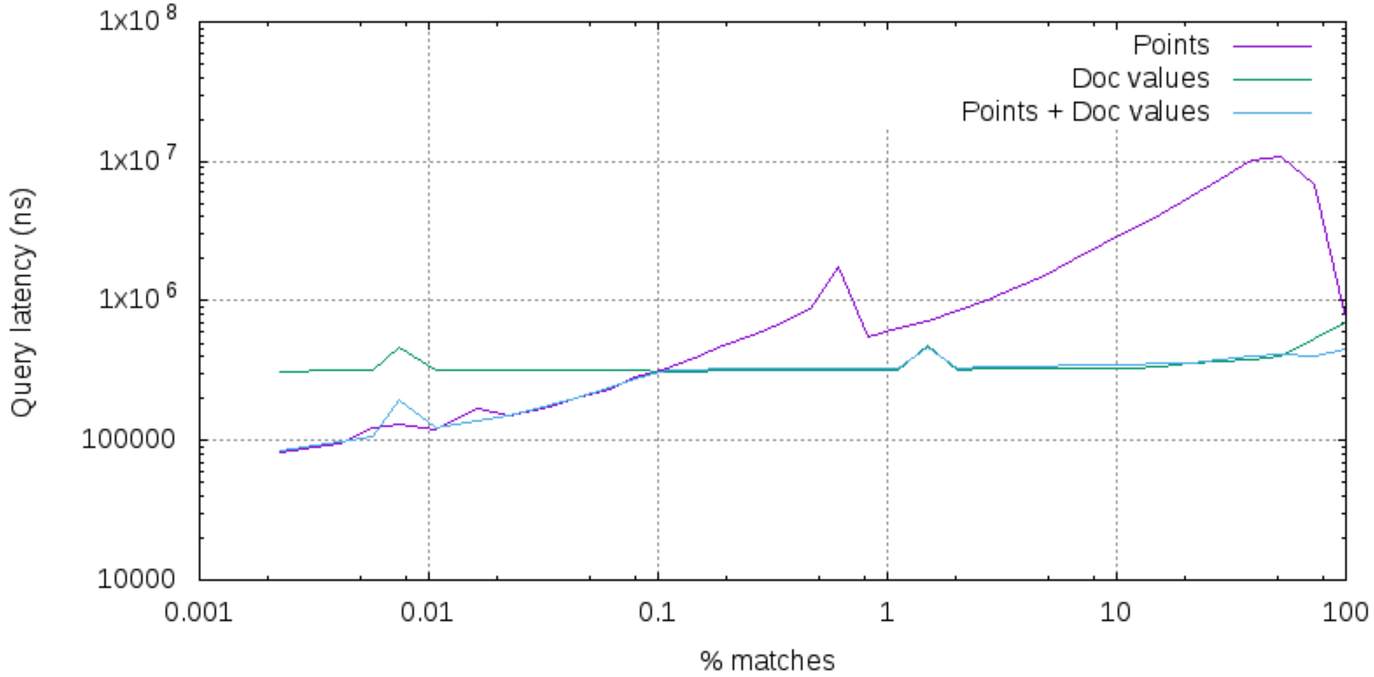


elasticsearch

Optimized Query Execution

Same queries, just faster.

- `range`
  - automatically chooses the more efficient of two query modes
  - Additional materials: blog
- `nested`
  - Nested mappings / queries get a speed boost
  - Writeup of how/why: 23079



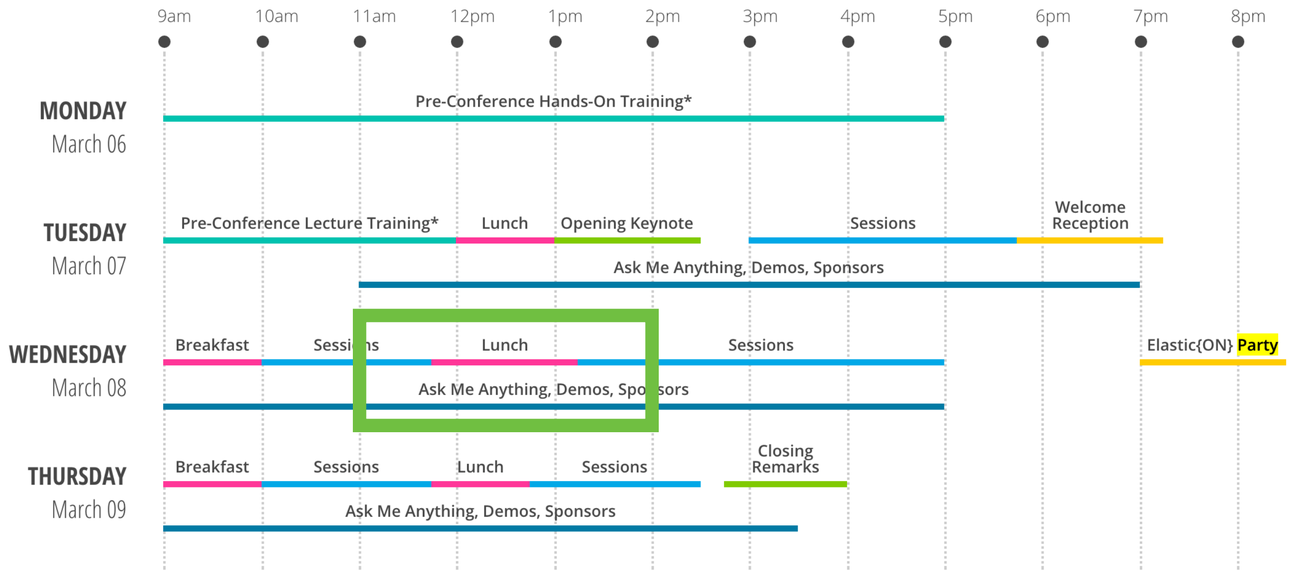
Query latency for 0.1% term and range

Numeric & Date Range Fields (5.2)

Mapping Improvements

- New types for date/number ranges (5.2)
  (date_range, int_range, float_range)

What's happening Wednesday 11am - 2pm



Terms Aggregation Partitioning (5.2)

Returning ALL the Terms, in Manageable Chunks

- frequent request
- return all responses from a terms aggs
- Terms can now be broken into partitions and partitions are returned by number

```
{
   "size": 0,
   "aggs": {
      "expired_sessions": {
         "terms": {
            "field": "account_id",
            "include": {
               "partition": 0,
               "num_partitions": 20
            },
            "size": 10000,
            "order": {
               "last_access": "asc"
            }
         },
         "aggs": {
            "last_access": {
               "max": {
                  "field": "access_date"
               }
            }
         }
      }
   }
}
```

Cross Cluster Search (5.3)

Tribe node is dead. Long live Cross-cluster search.

- Minimal viable solution to supersede tribe
- Addresses many of the challenges with tribe node
- Reduces the problem domain to query execution
- Cluster related information is reduced to a namespace

Field Collapsing (5.3)

One method to rule them all…

- Simple (almost) no setup!
- Great for query-time group/category de-dup

```
GET /twitter/tweet/_search
{
    "query": {
        "match": {
            "message": "elasticsearch"
        }
    },
    "collapse" : {
        "field" : "user", ❶
        "inner_hits": {
            "name": "last_tweets", ❷
            "size": 5, ❸
            "sort": [{ "date": "asc" }] ❹
        },
        "max_concurrent_group_searches": 4 ❺
    },
    "sort": ["likes"]
}
```

Elasticsearch Keystore

If you like it, you should put it in a keystore.

- Sensitive settings should not be protected by filesystem permissions only.
- Commands feel familiar:
  - bin/elasticsearch-keystore create
  - bin/elasticsearch-keystore list
  - bin/elasticsearch-keystore add the.setting.name.to.set
  - bin/elasticsearch-keystore remove the.setting.name.to.remove
- Just the framework/start: sensitive settings to be pulled in

Elasticsearch 6.0 is coming

- Remove Type

  Sparse Doc Values

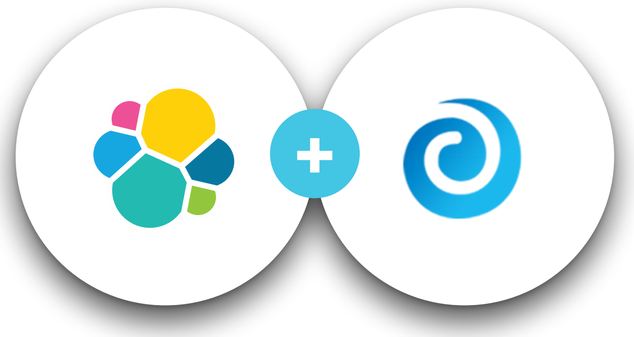
  Index Sorting

  Sequence Numbers

  Rolling Upgrades

  …

Elastic Stack 6.0.0-alpha2 Released



Machine Learning

UNSUPERVISED MACHINE LEARNING

- Automatically detect anomalies
- Advanced correlation and categorization
- Identify root cause(s)
- Expose early warning signs

NEW USE CASES

- Analyze time series data
- Expand security, IT Ops, fraud, finance, and many more use cases
- Currently beta; building a more native integration into the Elastic Stack

elastic

Job: it_ops-kpi and 2 others

Top Influencers

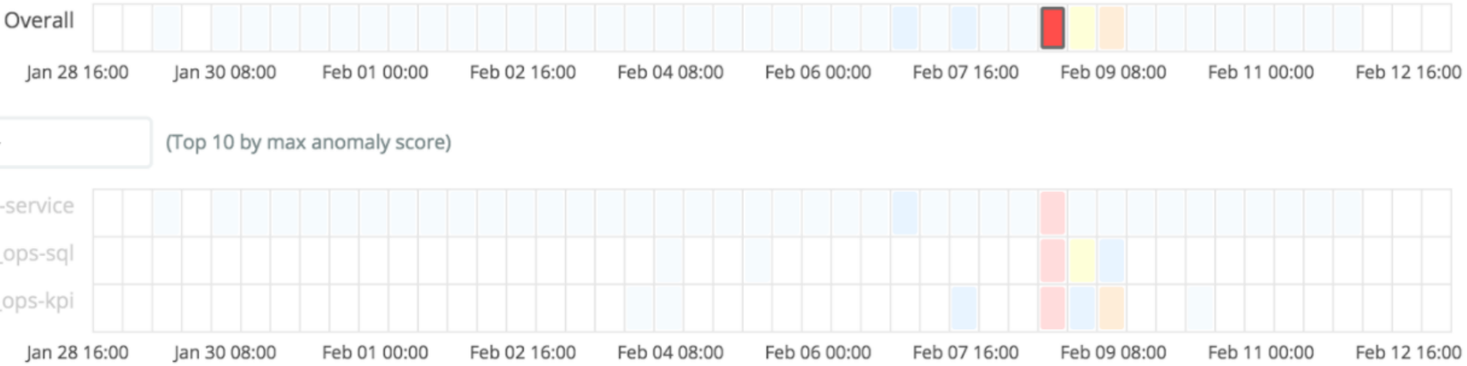
service

inventory-us-east-1-34	94	97
auth-us-west-1-1e	7	51
test-srv-02	5	29
elasticsearch-22	3	16
elasticsearch-77	2	11
payment-srv-21	2	6
payment-srv-11	1	5
backup-srv-13	1	9
test-srv-01	1	7
inventory-us-west-1-4e	1	7

hostname

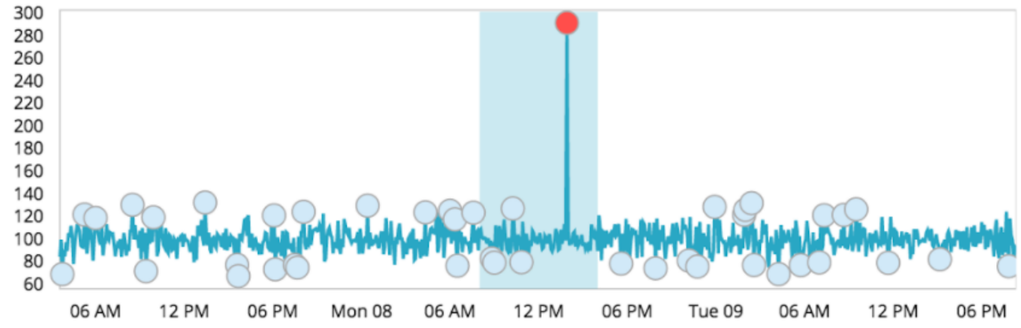
MSSQL-0783E4076	94	1237

Anomaly timeline

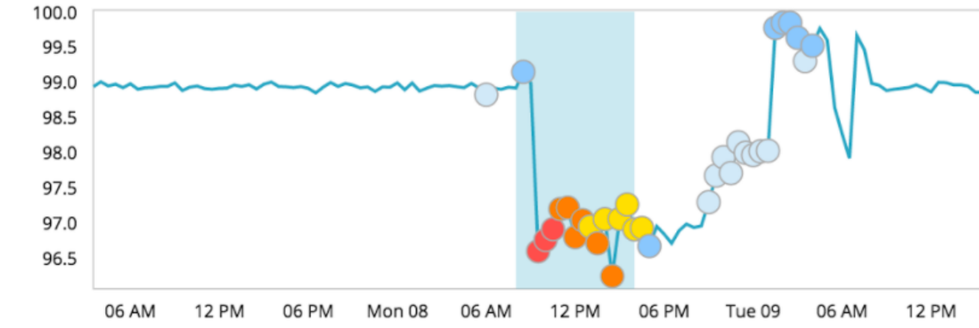


Anomalies

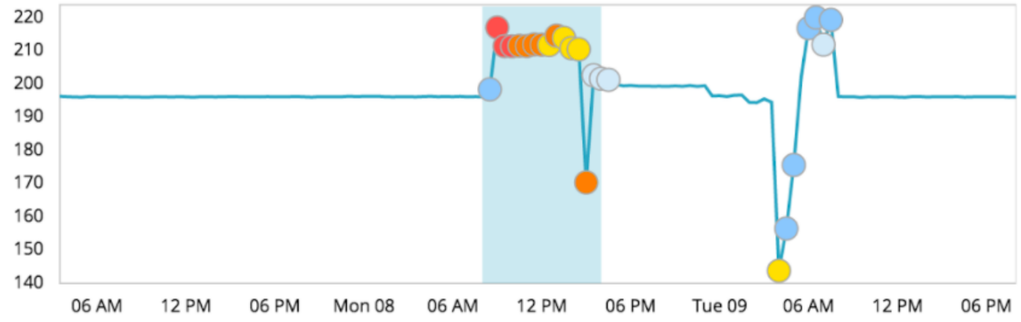
mean responsetime service inventory-us-east-1-34



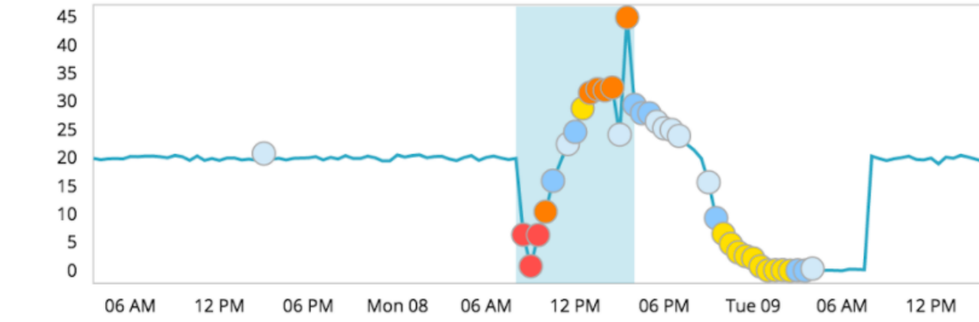
mean SQLServer_Buffer_Manager_Buffer_cache_hit_ratio hostname MSSQL-0783E4076



mean SQLServer_General_Statistics_User_Connections hostname MSSQL-0783E4076



mean SQLServer_SQL_Statistics_Batch_Requests_sec hostname MSSQL-0783E4076



Elasticsearch-SQL

Elasticsearch-SQL Coming soon!

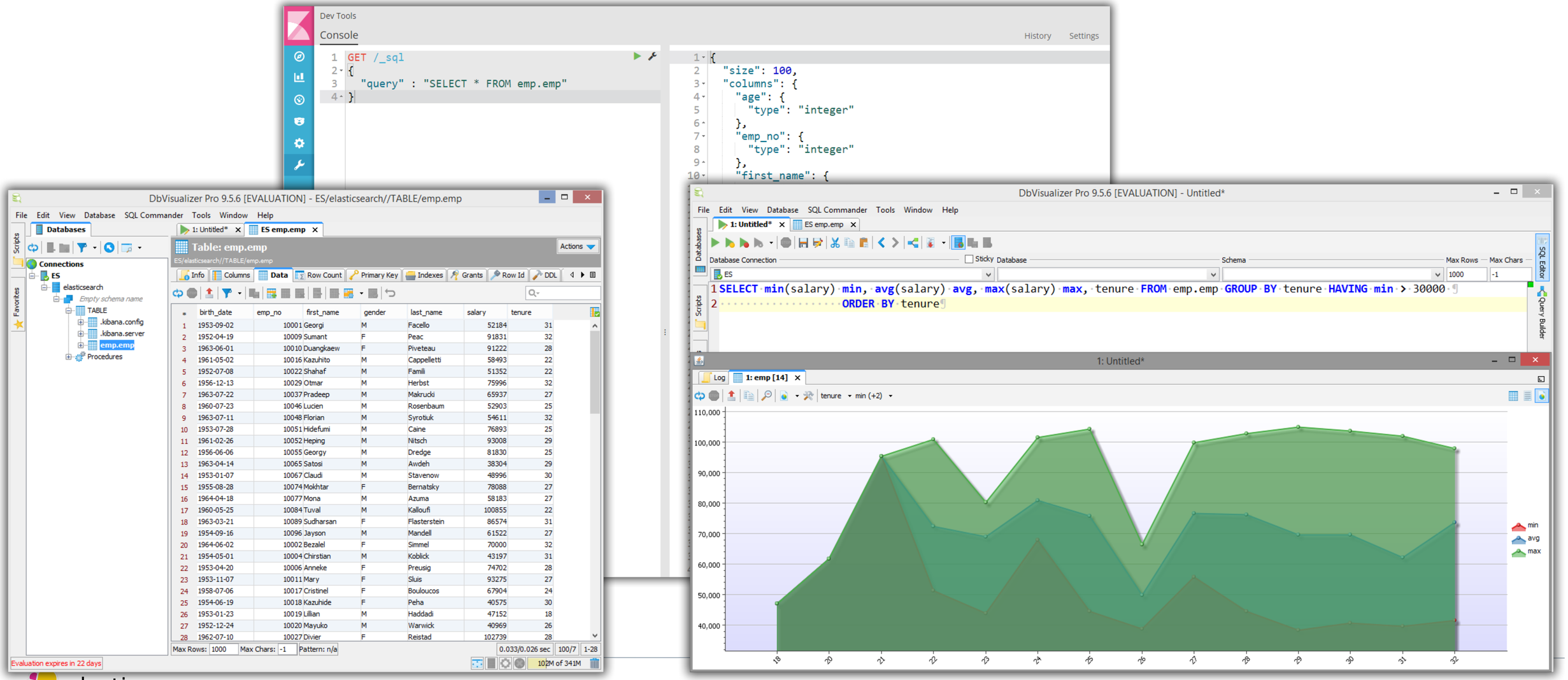
CLI

- OS independent
- Quick diagnostics and sanity checks
- Admin focused
- Optimized for efficiency

JDBC

- Dedicated client (driver) and server component
- JDBC 4.2/Java 8 (downgrade possible)
- Supports `java.sql` and `javax.sql` APIs
- Pays attention to details
  - Timeouts (connect vs read vs network)
  - Logging
- Light, without dependencies

Elasticsearch-SQL



elastic

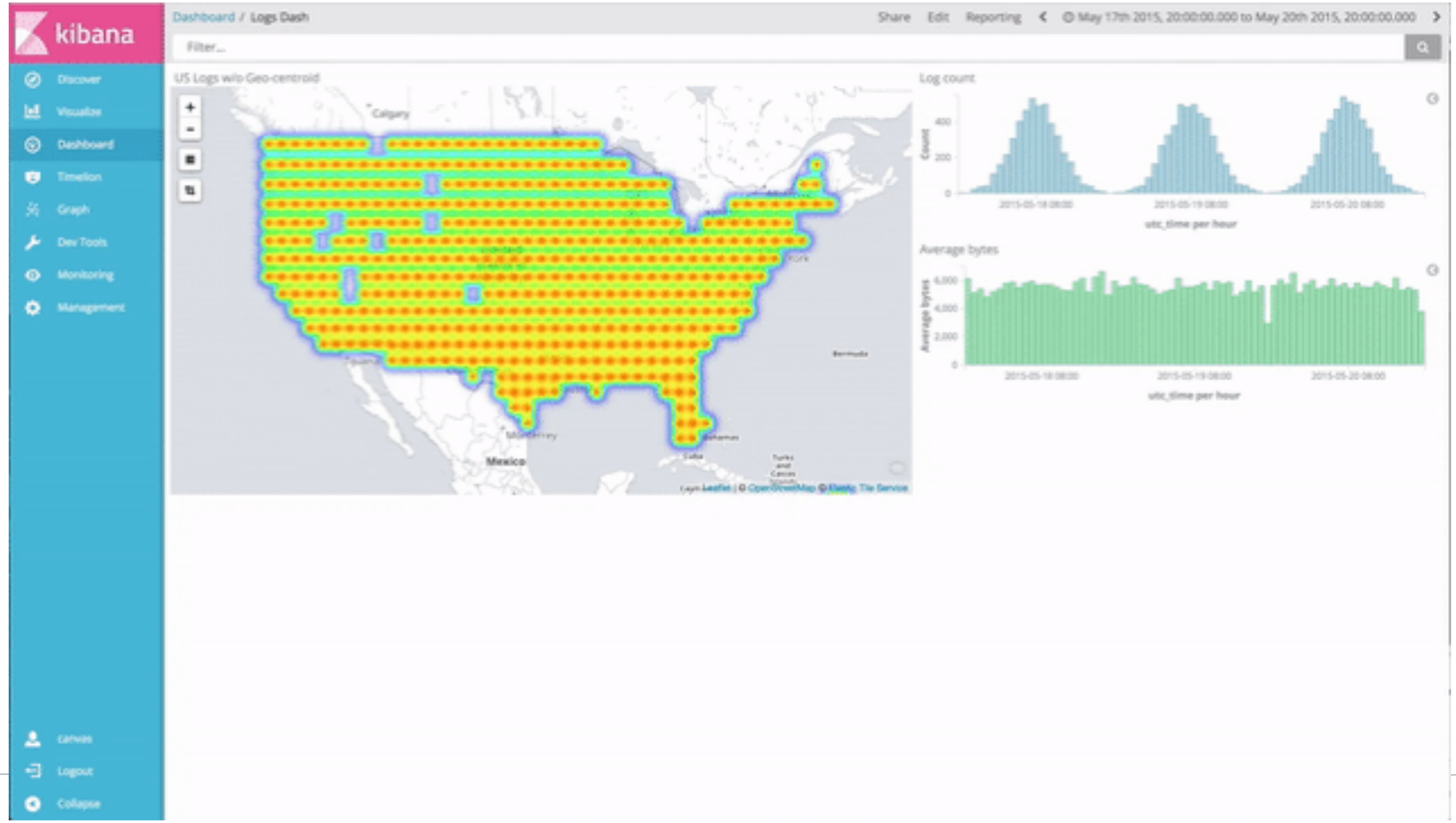
kibana

Kibana 5.4!

- **20** Kibana enhancements with 5.4
- **10** Visualization customizations
  - Combo chart (2), horizontal bar chart (2), multiple y-axis (4), easily switch between viz (2)
- **5** Pipeline agg support
  - Derivatives (5)
- **2** Timepicker enhancements
  - “to” field in relative, start / end of day
- **1** Dashboard enhancement
  - Allow duplicate dashboard names
- **1** Reporting enhancement
  - Report should reflect visualization query when defined

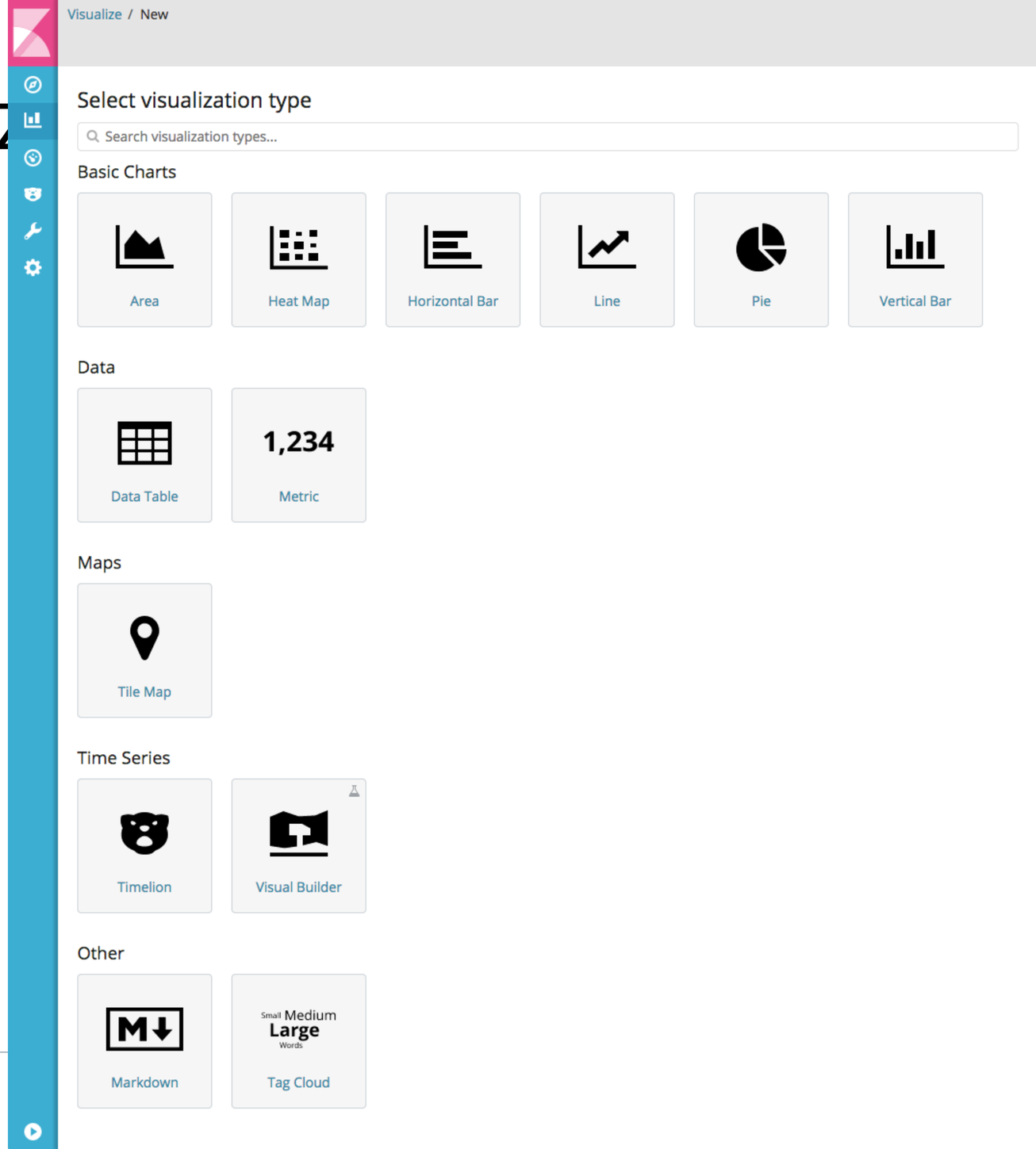
Dashboard View & Edit Mode

Gone are the days of accidental panel resizing



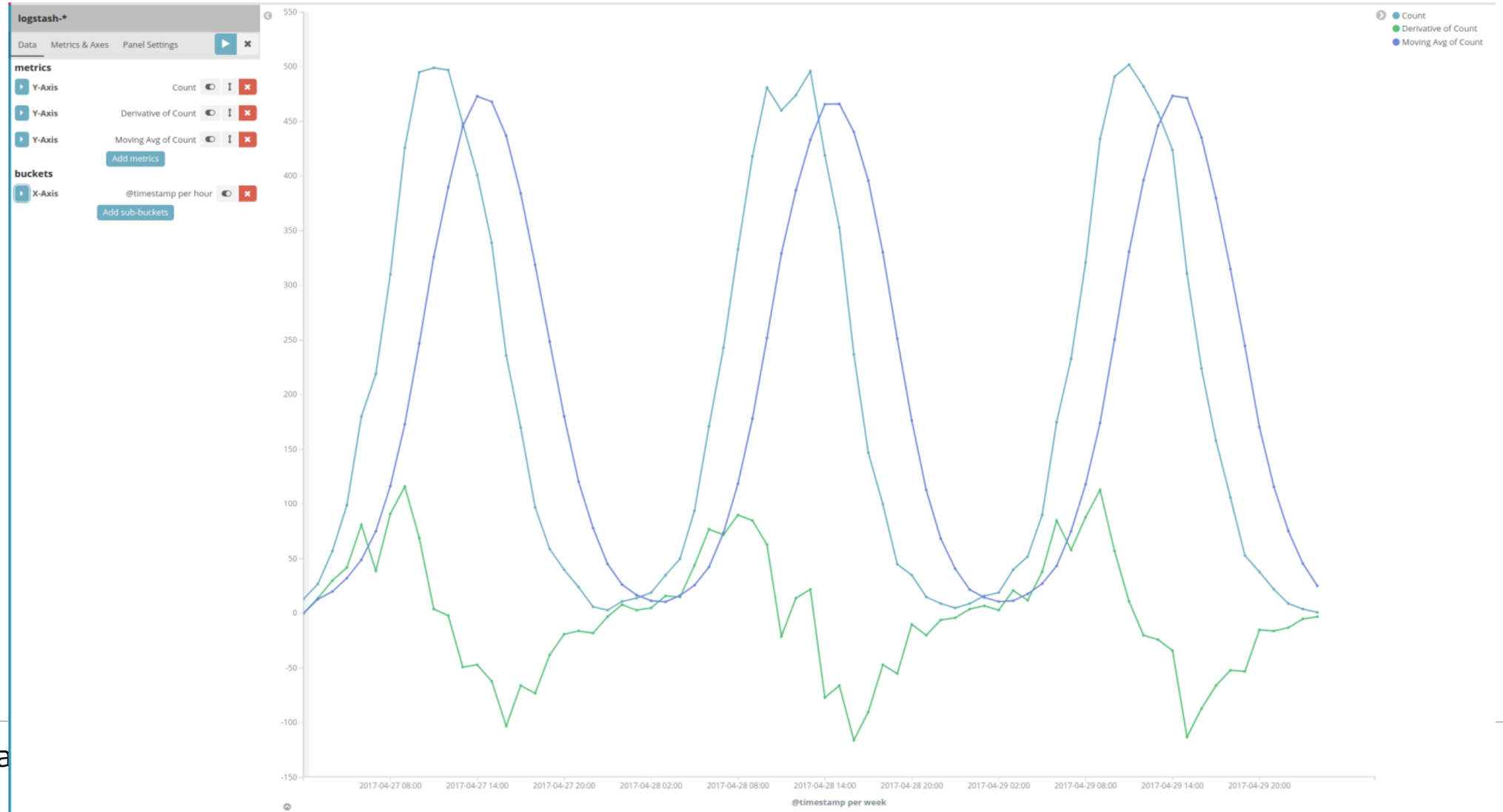
Create Visualization Wiz

- Creating a new visualization can be daunting
- Group visualization types into buckets with icons for each visualization type



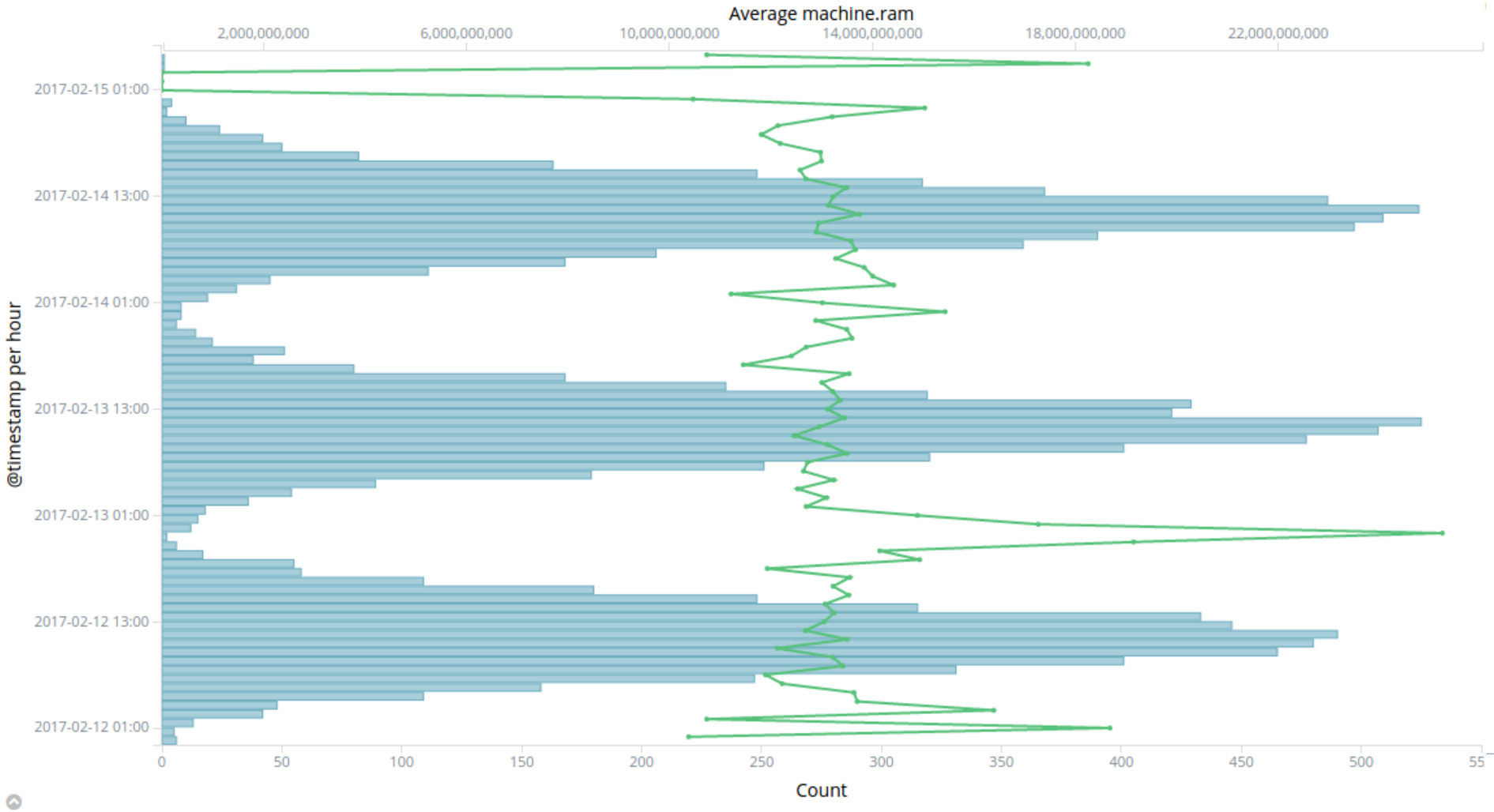
Pipeline Aggs in Kibana!

We've all wanted it, here it is!



Visualization Enhancements

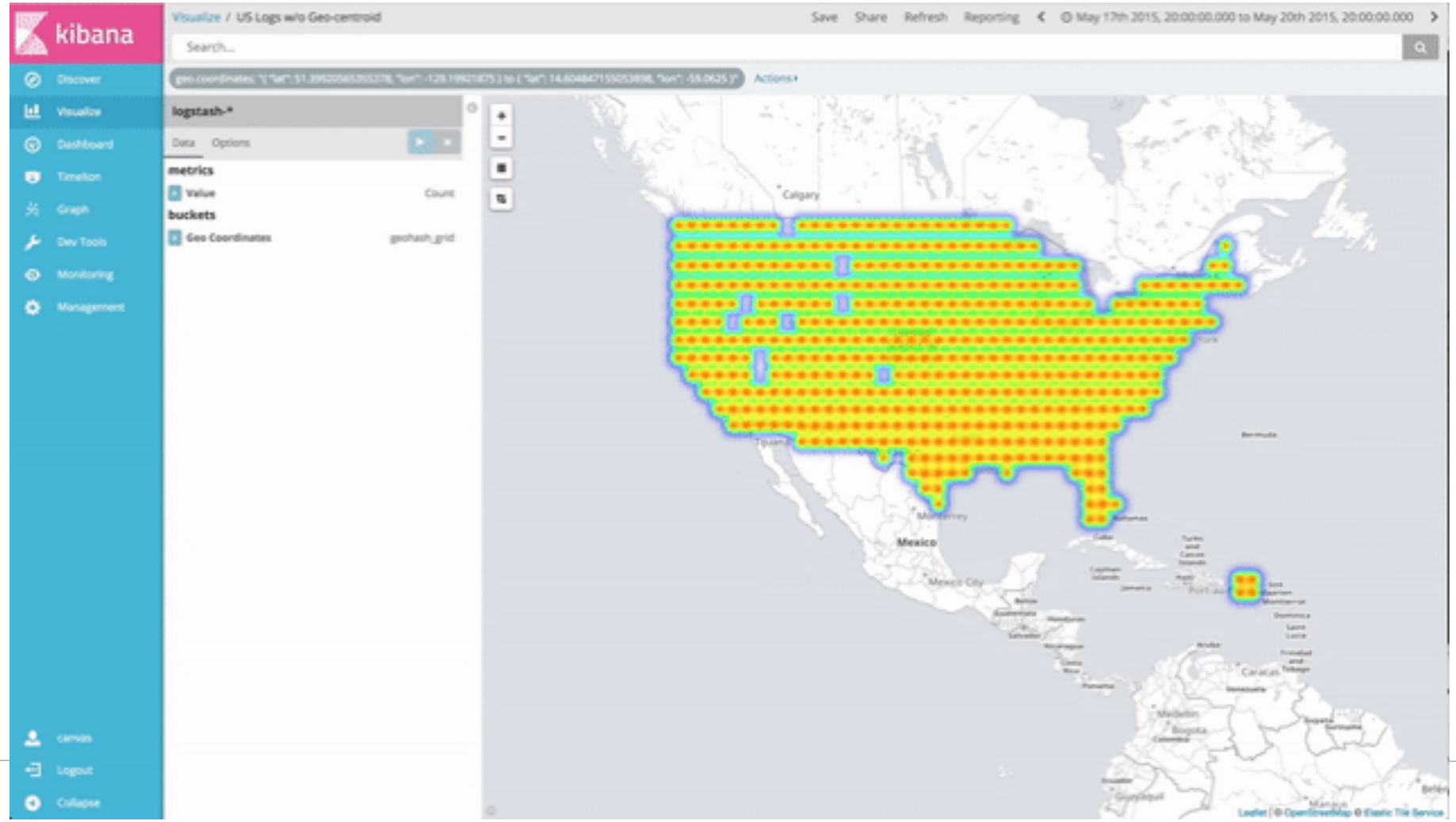
Combination charts, horizontal charts, multiple y-axis & more!



elastic

Geo-Centroid Support

More natural looking maps. Less awkward boxes.



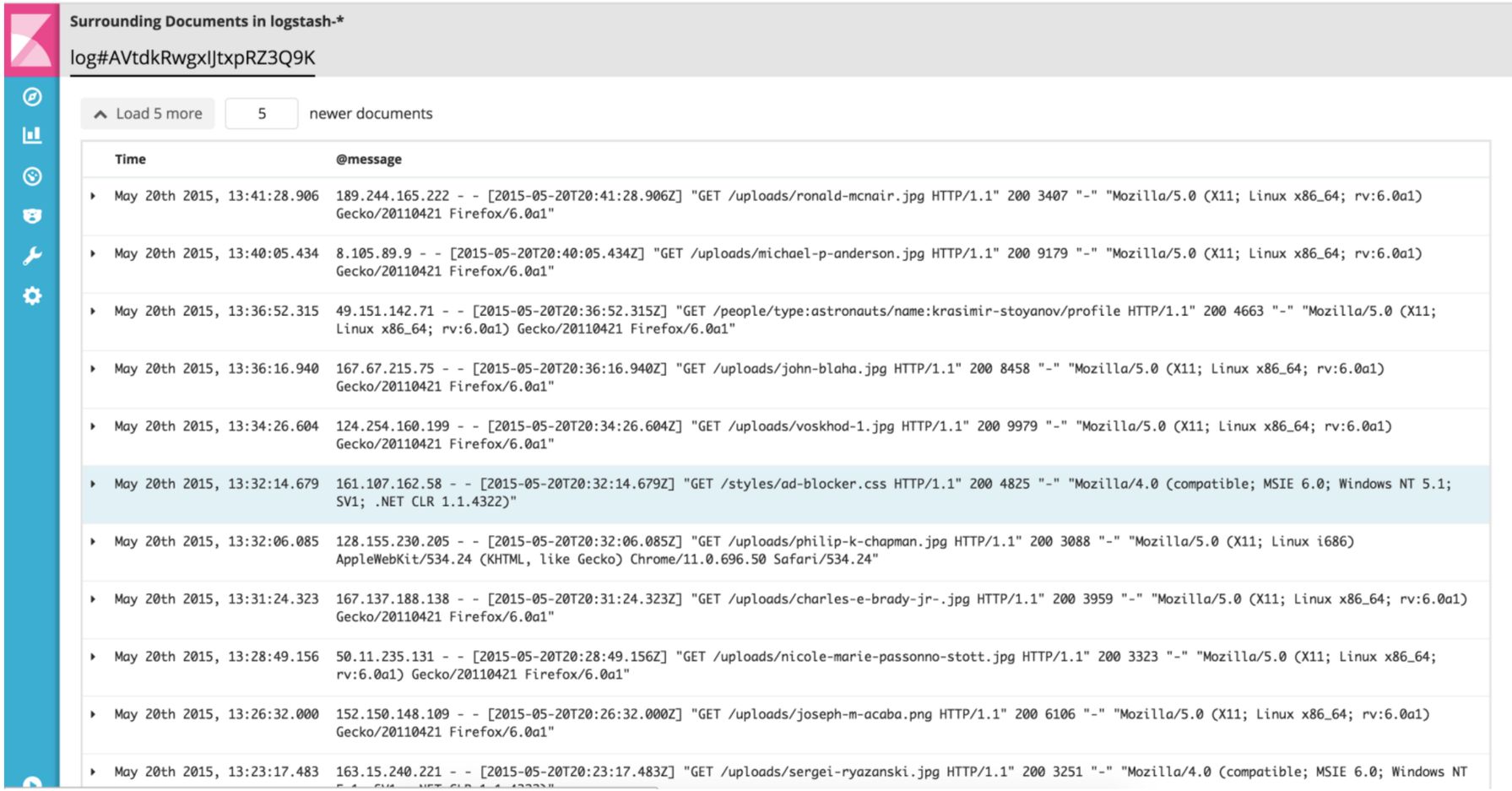
Event Context

The Needle in the Haystack

Filter a number of events

Understand ‘before’ and ‘after’ events

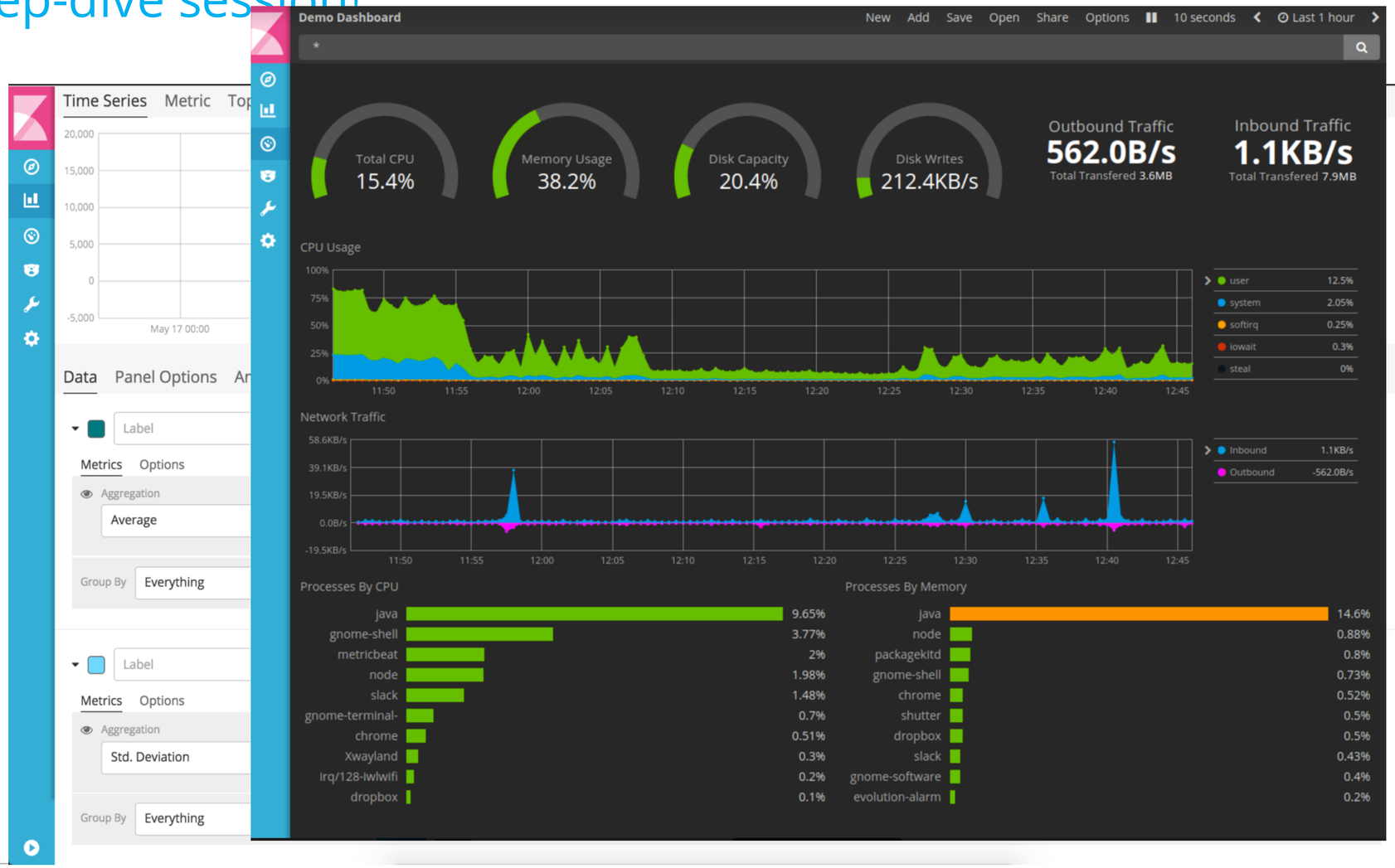
Enter from Discover through ‘View surrounding documents’ link



Time Series Visual Builder

More to come, an entire deep-dive session!

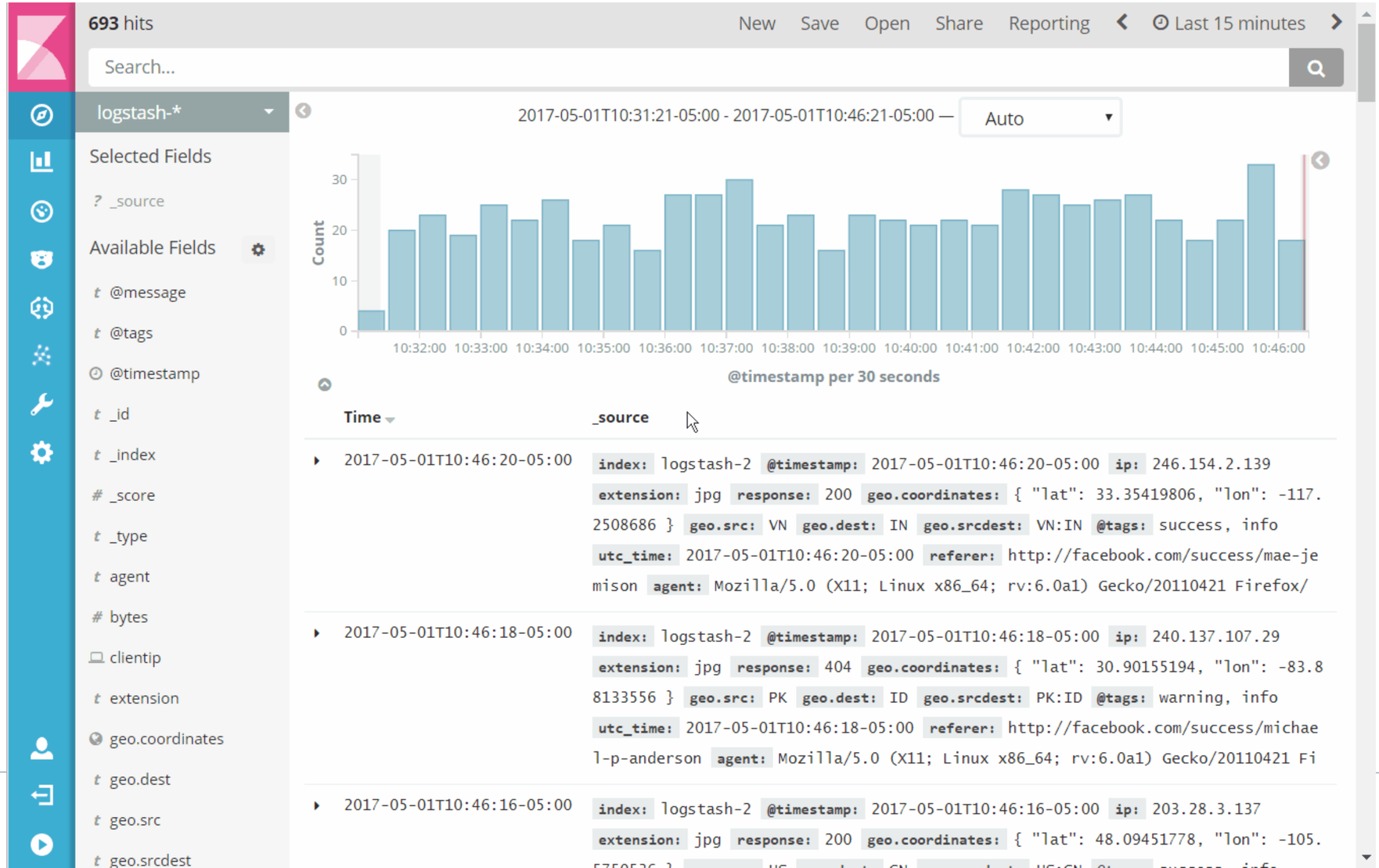
- A curated UI - just for time series data with features such as…
- Chart multiple indices
- Pipeline aggregations
- Complex calculations
- Conditional formatting
- Annotations
- Series offset



elastic

Watcher UI

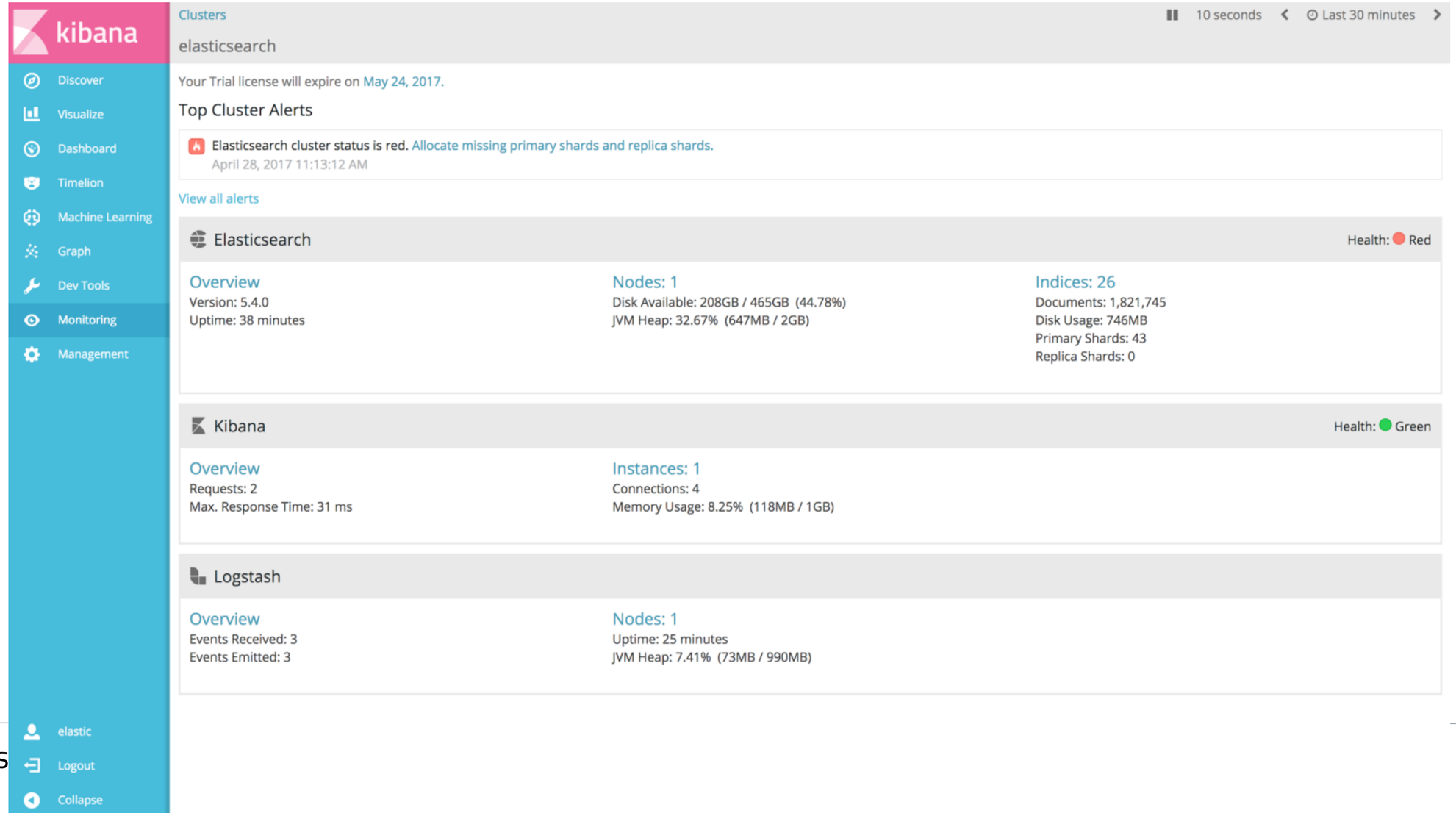
Phase 1 - View, manage, and create Watches



elastic

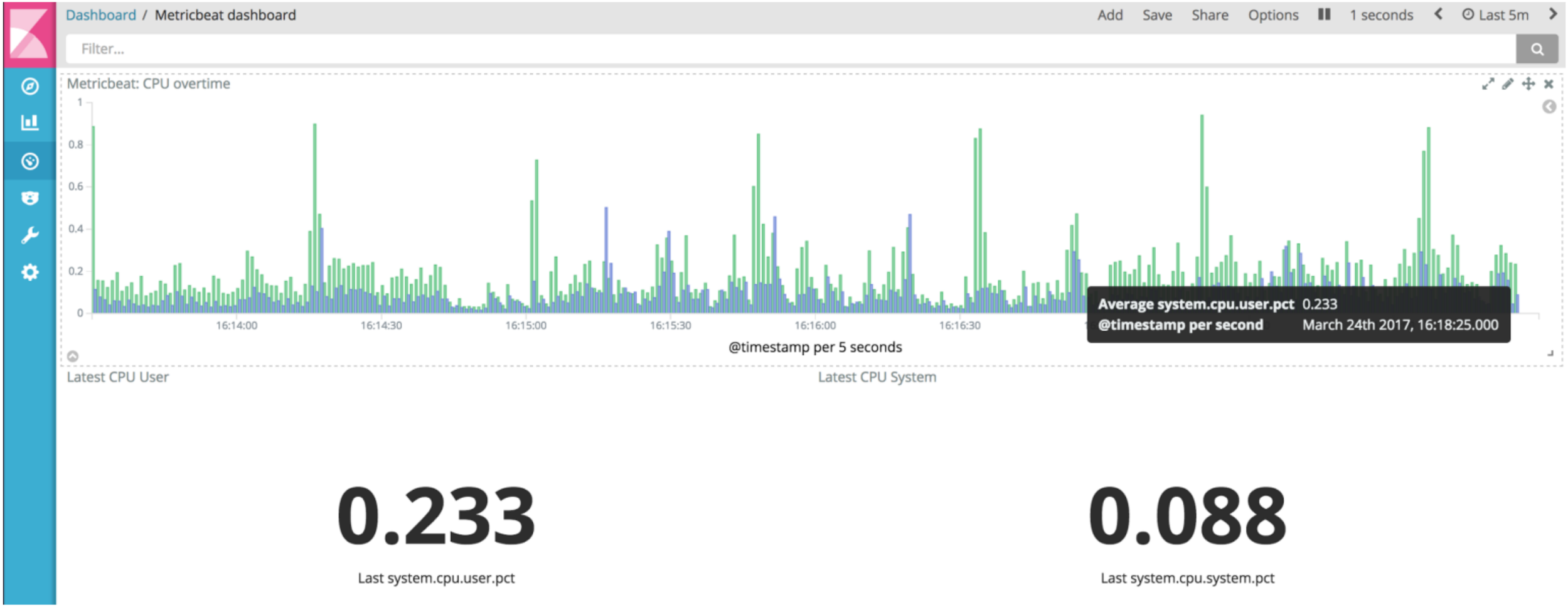
Cluster Alerts

Built-in Watches to help diagnose cluster issues quickly



New Analytics

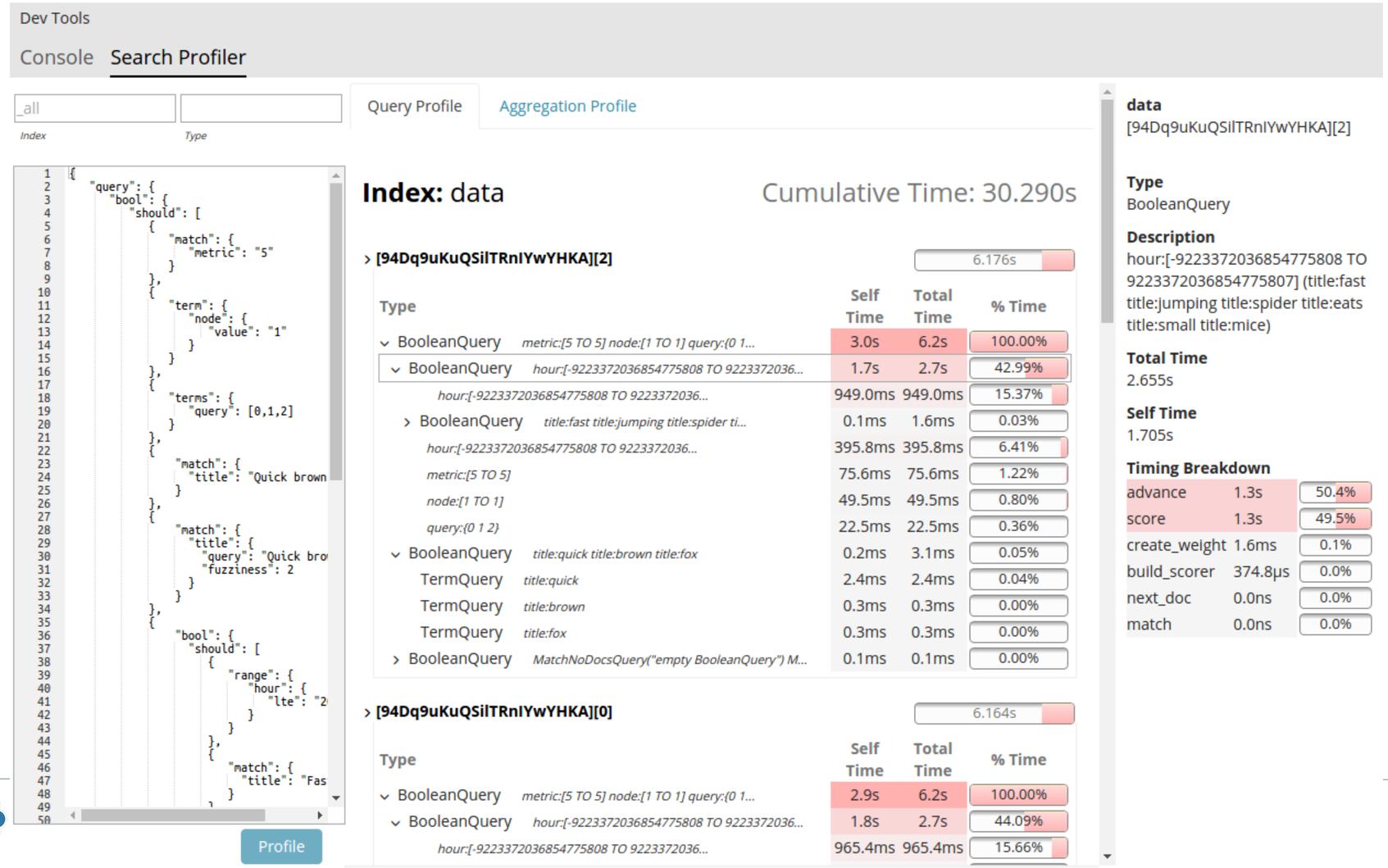
Top Hits Aggregation (5.3) - Visualize the ‘latest’ metric



elastic

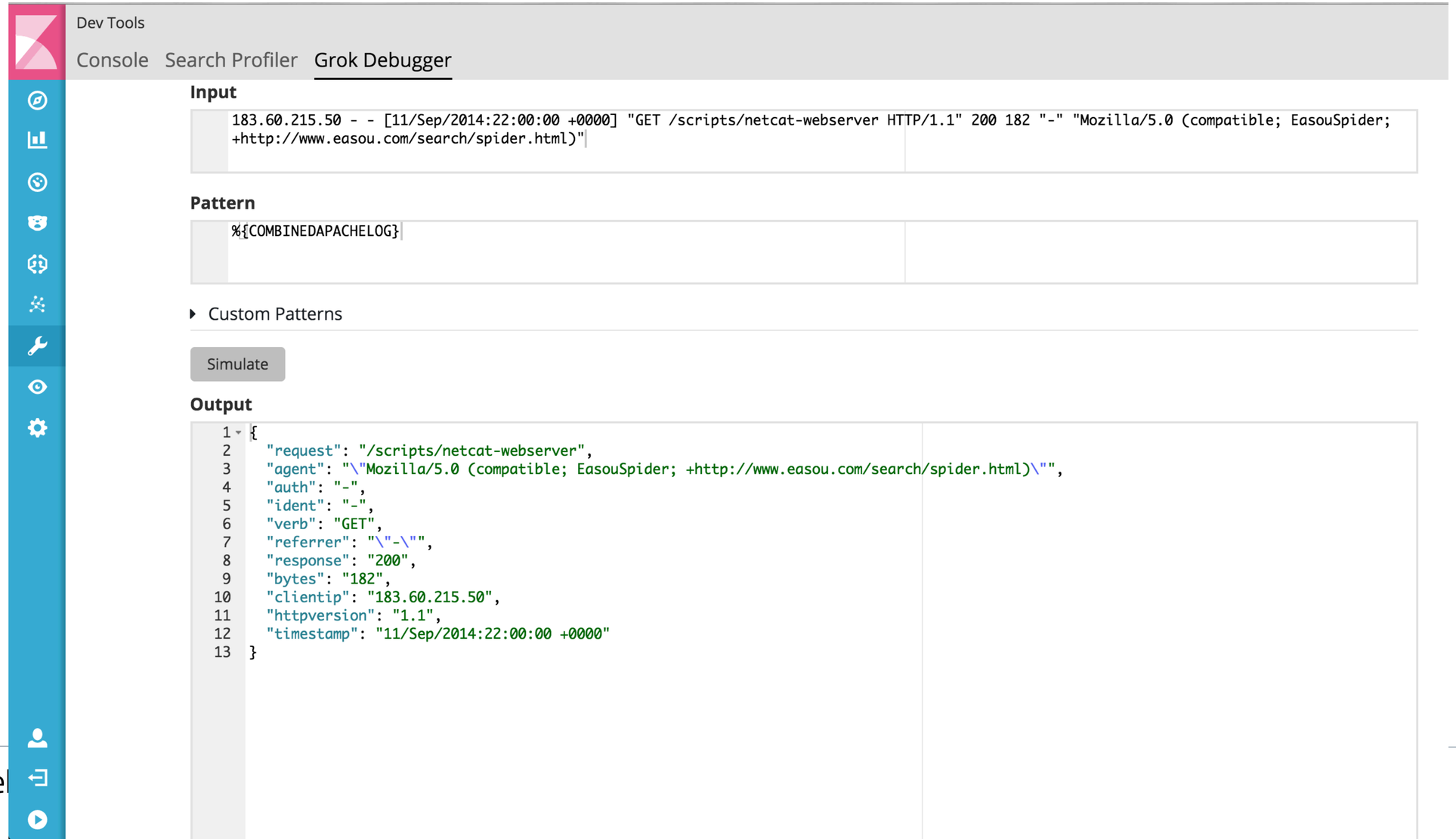
Profile your Search Queries

Search Profiler (5.1) - Detect and visualize bottlenecks in your query



* requires X-Pack (Basic)

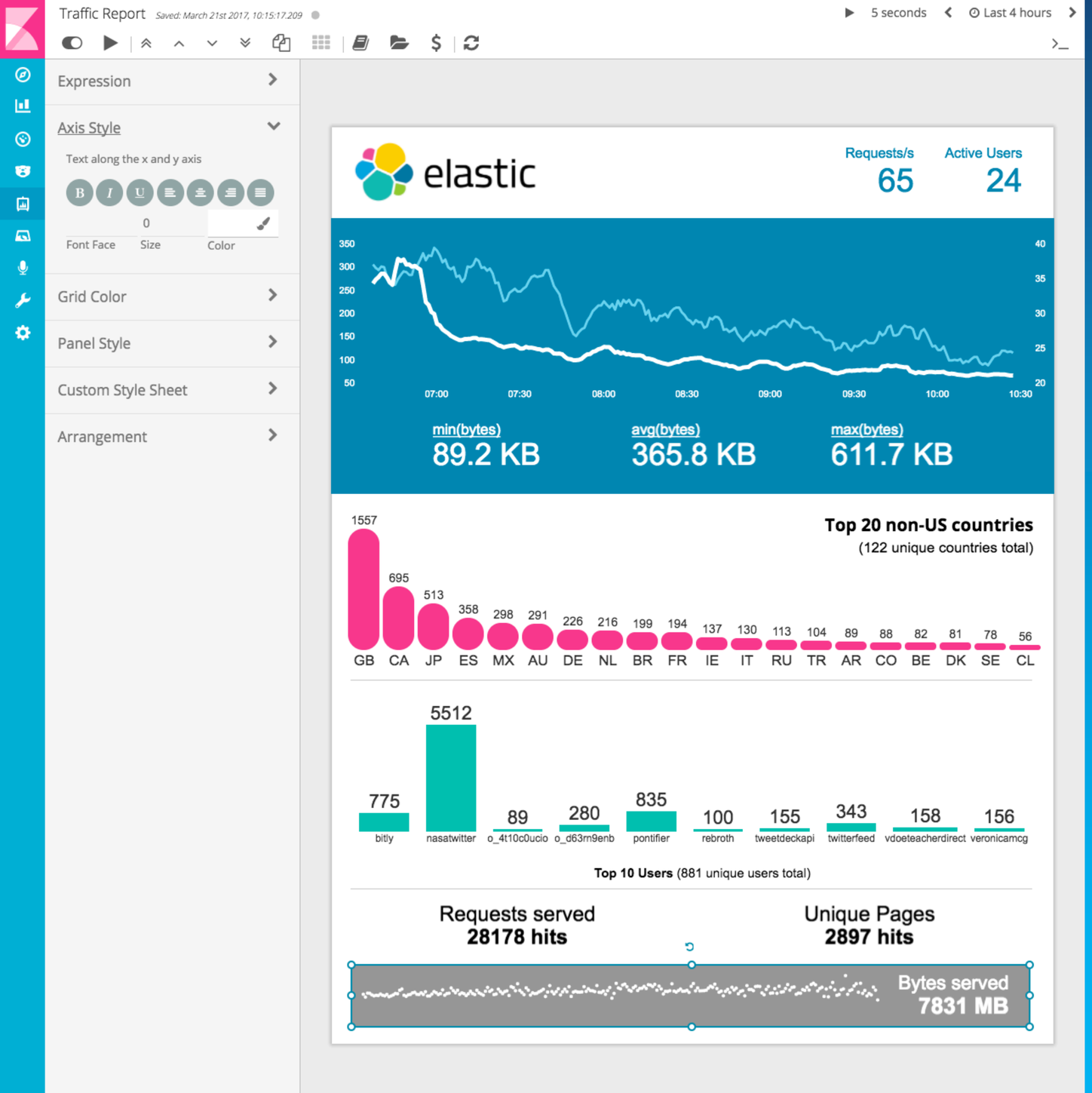
Grok Debugger



Internationalization Support

I18N, phase 1 complete (5.2)

- Adheres to browser preference for language
- Translations as plugins
- Thanks IBM!



Kibana Canvas

- New visualization application on top of Elasticsearch data
- Use Case:
  - live infographics
  - presentations with live data feeds
  - highly customized reports
- Currently, in the prototyping phase
- Release date: TBD

logstash

Logstash 5.4 Features

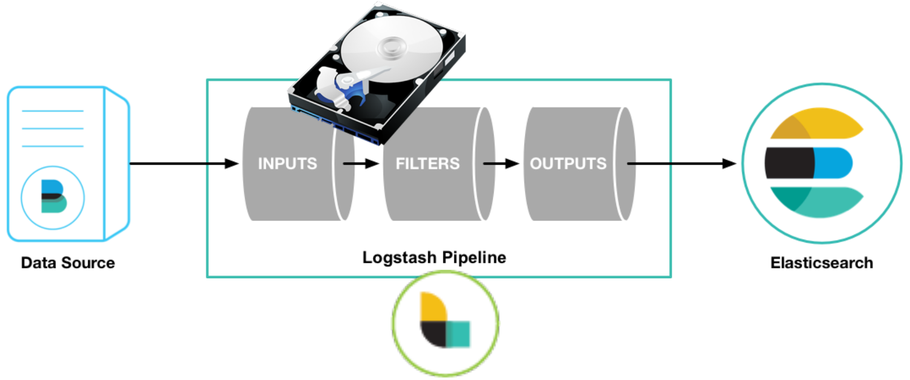
1. **Persistent Queues GA** 

1. **GeoIP2 ISP & ASN Lookup Support**
   - Lookup ISP and ASN data with the Maxmind GeoIP2 ISP Database (separate license)
   - GeoIP filter v4.1.1

1. **Log4j2 Support** (Log4j input plugin deprecated) - two options:
   - Log4j2 SocketAppender sends JSON to Logstash TCP input
   - Log4j2 logs to file for Filebeat to collect

Persistent Queues

Features



Disk-Based Queuing - GA in 5.4

Resiliency
- Durability across node failures
- **At-least-once** delivery guarantees

Adaptive buffering
- External queuing layer no longer required to absorb throughput

Control max disk usage

Limited impact on performance

Monitoring UI integration
- Queue type and queue lag
- *Future: disk usage and disk IO*

Opt-in feature

Persistent Queues

Resiliency

- Make sure `**queue.checkpoint.writes**` = **1** for at-least-once delivery guarantees

- PQs are resilient across node failures. For protection against disk-level corruption or failure, disk redundancy is recommended:
  - **On-Prem** – RAID
  - **AWS** – EBS (i.e. t2, m4, c4, p2)
  - **GCP** – Persistent Disk
  - **Azure** – Premium / Standard Managed Disks (ZRS or higher for multi-DC redundancy)

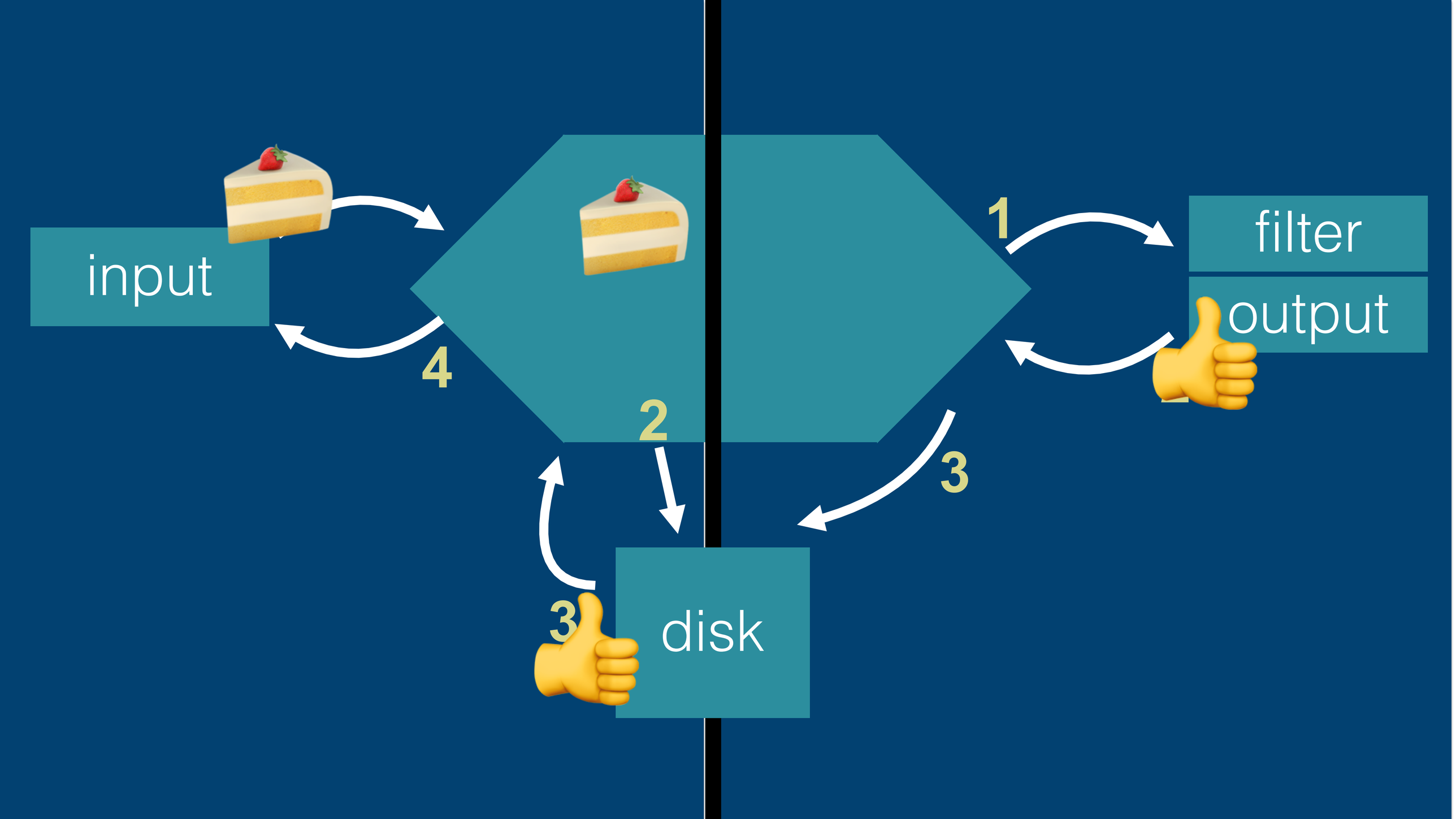
- Shared network filesystems (CEPH, GlusterFS, NFS, etc.) are **not** recommended.

- Inputs that do not support acknowledgements (i.e. TCP, UDP) can still lose data.

Persistent Queues

Things To Know

- Understand your hardware
  - Know your available disk capacity needs
  - Monitoring UI can help monitor PQs
  - SSDs are **not** required
- PQs should be enabled when ingesting from Kafka
  - Resilient transport from Kafka to ES
  - PQs mitigate the need for expensive reprocessing on recovery
  - Recommend default 1GB PQ max disk size
- For ephemeral storage, multi-DC queue replication, or centralized data replay:
  - Add or leverage an HA Kafka cluster


input
filter
output
disk
1
2
3
4
3

Centralised Management

- Elasticsearch as a remote config store
- Manage configurations via UI
- Group multiple Logstash under roles
- Simple alternative to puppet, chef

elastic

Offline Plugin Management (5.2)

Air-gapped Networks and Offline Environments

Prepare and Pack Plugins on Staging Box

```
$ bin/logstash-plugin prepare-offline-pack logstash-filter-* logstash-input-beats
```

Move Offline Pack to Offline Boxes

- Default pack location: `/LOGSTASH_HOME/logstash-offline-plugins-5.2.0.zip`
- Change pack location using `--output /path/to/pack` parameter

Install or Update Plugins

```
$ bin/logstash-plugin install file:///path/to/logstash-offline-plugins-{logstash_version}.zip
```

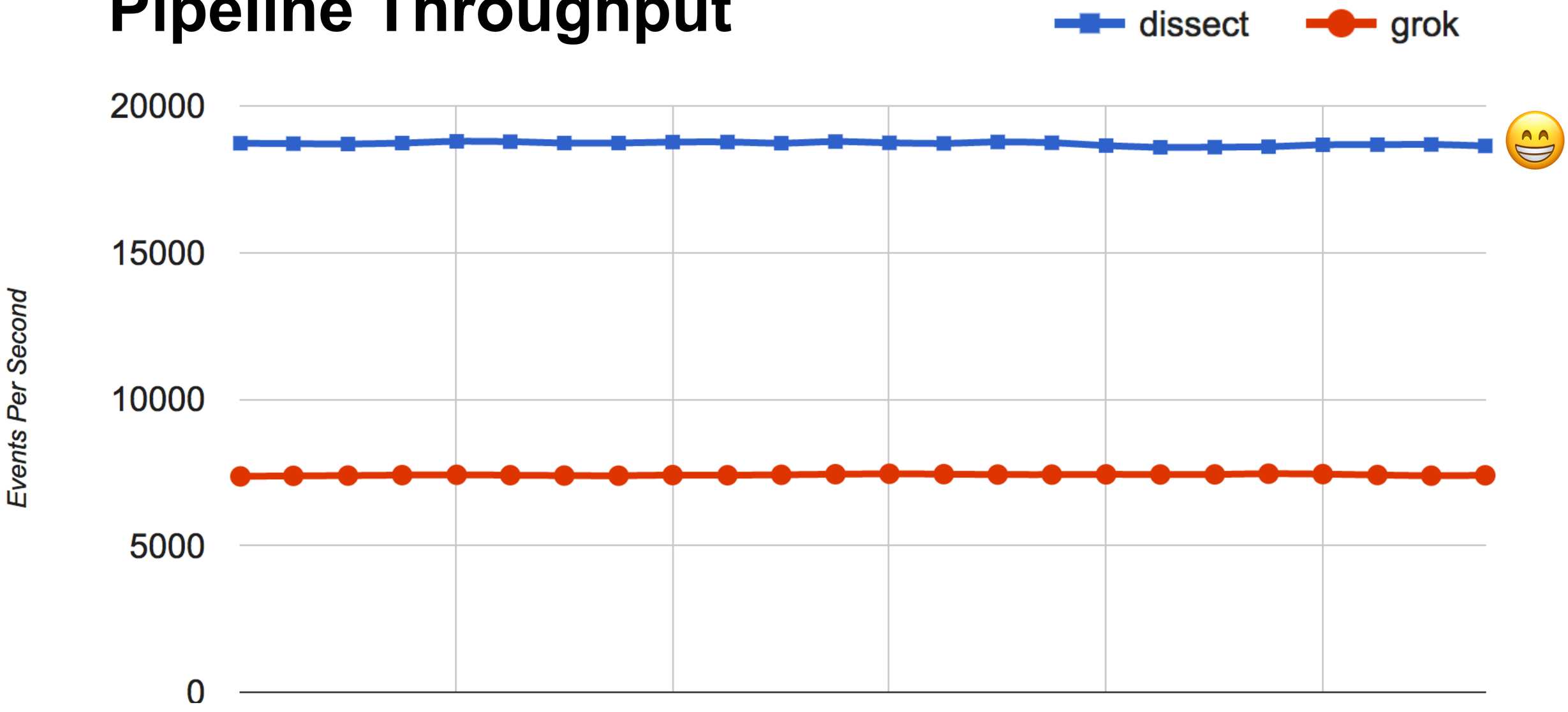
Database Lookup Enrichment

JDBC_streaming filter (5.3)

- Enrich Logstash events with DB data (streaming joins)
  - Executes JDBC lookup queries per event (add one or more fields)

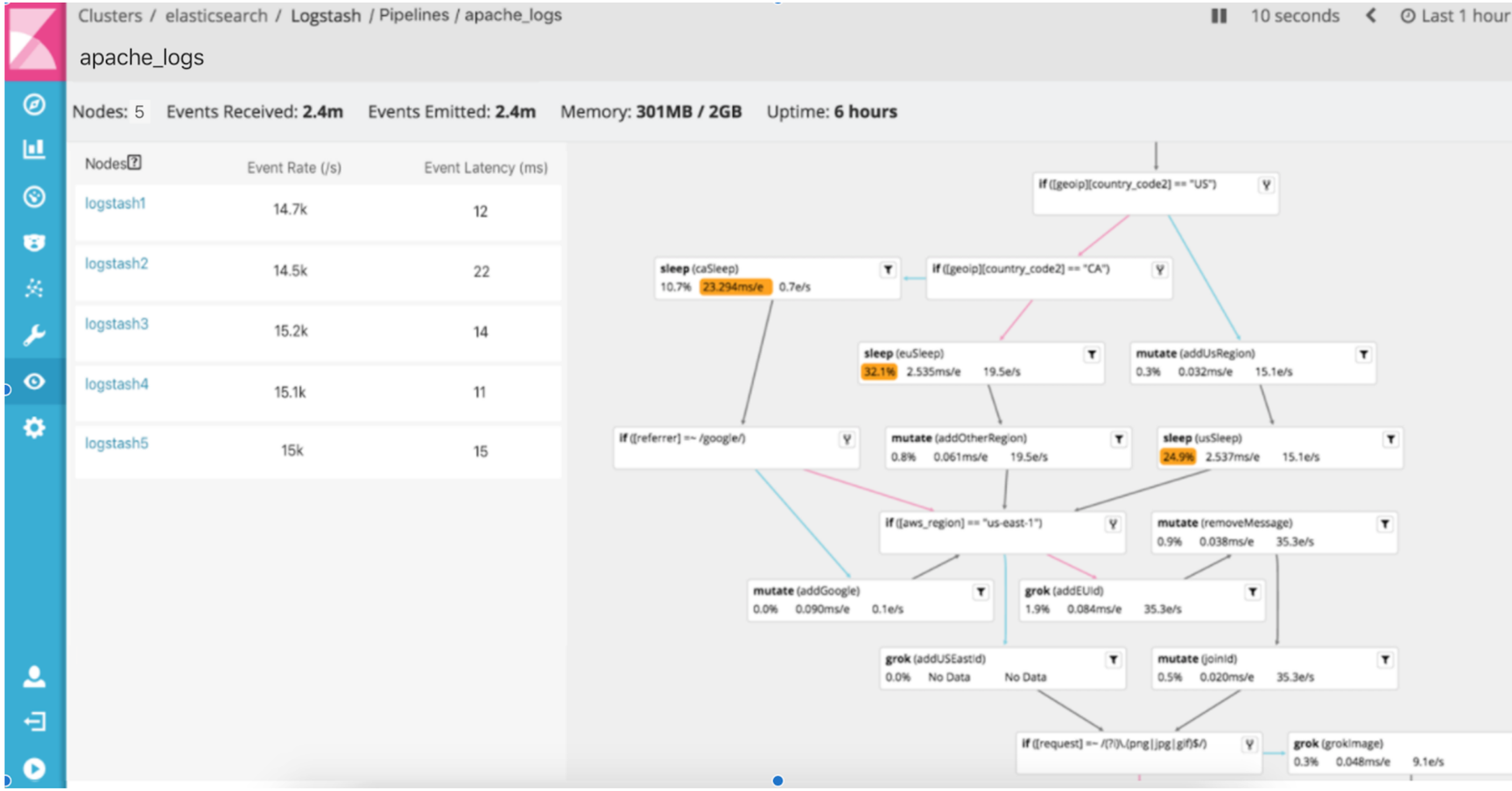
```
jdbc_streaming {
  jdbc_connection_string => "jdbc:mysql://localhost:3306/mydatabase"
  statement => "select * from PRODUCTS.FRUITS WHERE SKU = :sku"
  parameters => { "sku" => "sku_code"}
}
```

Pipeline Throughput



More

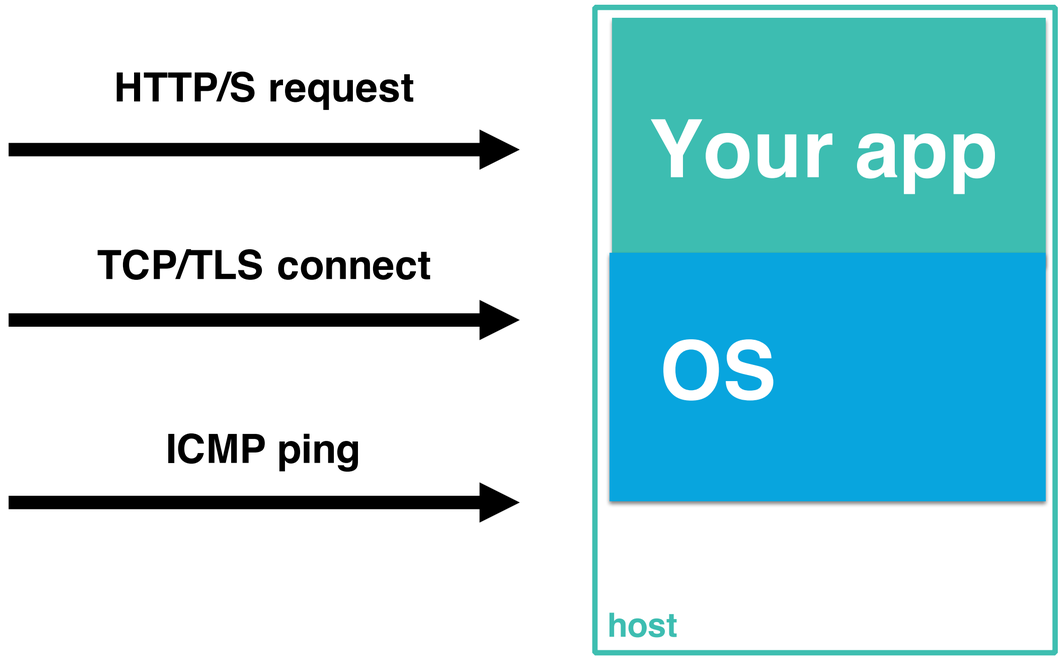
- Logstash modules（5.5）
- Dead letter queue
- Multi Logstash pipelines
- Logstash Intermediate Representation（LIR）



beats

New Beat: Heartbeat (beta in 5.2)

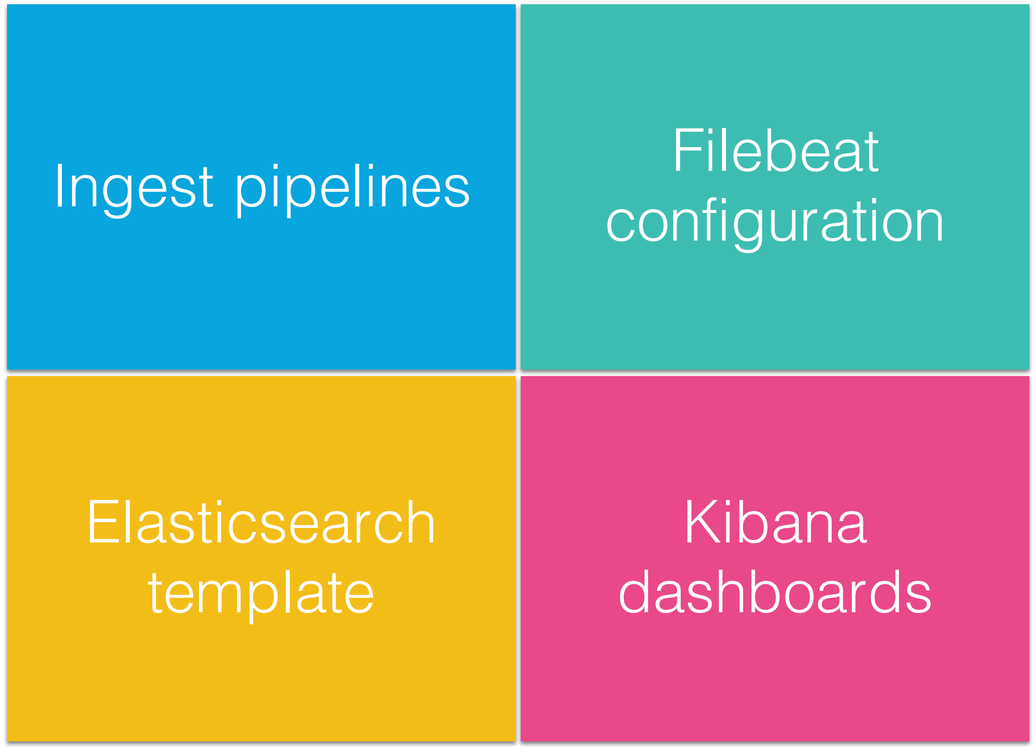
- Ping all the things
- Gather round trip metrics
- Many to many
- Ping IPs behind load balancers



Filebeat Modules (5.3)

They are like Metricbeat modules…only different

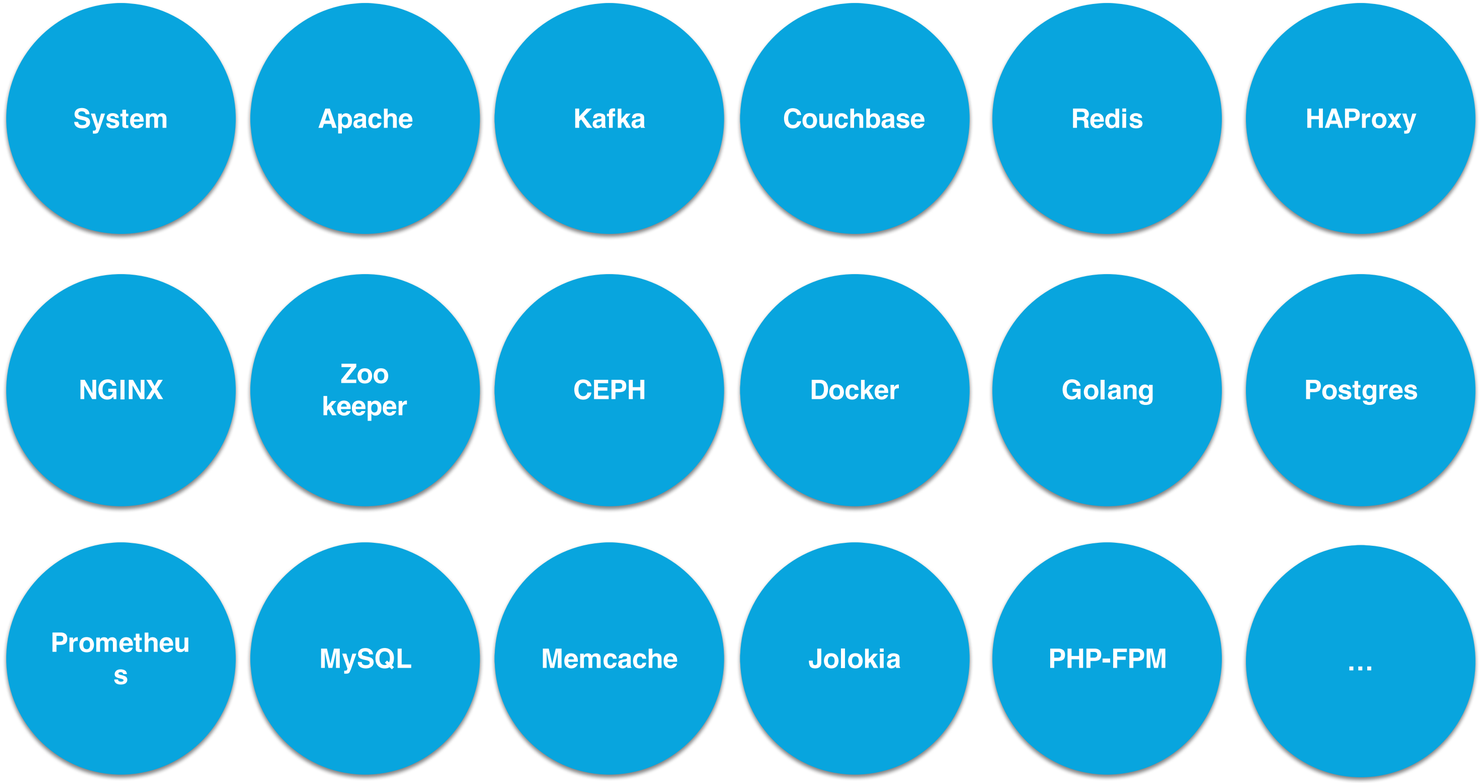
- Because simple things should be simple
- Prepackaged configs for common log formats
- Get to a dashboard in minutes
- First release includes Apache, Nginx, MySQL, system modules. More to come.



More Modules in Metricbeat

Modules are growing

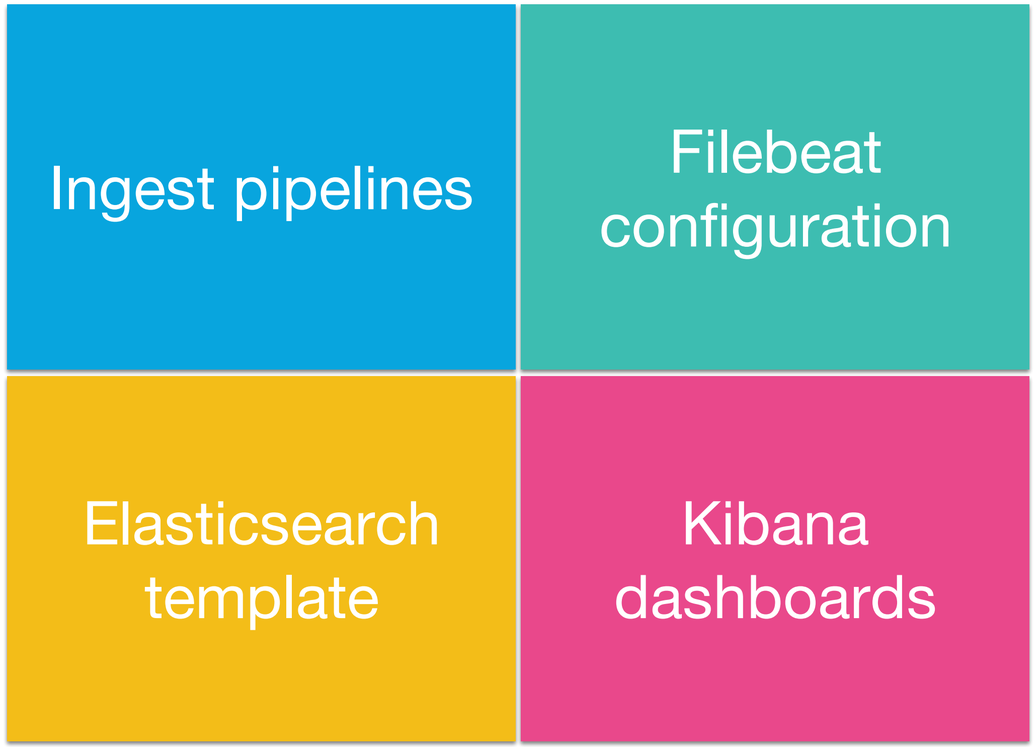
- Docker
- Kubernetes
- Kafka
- Prometheus
- Elasticsearch
- Kibana
- vSphere
- RabbitMQ
- … …



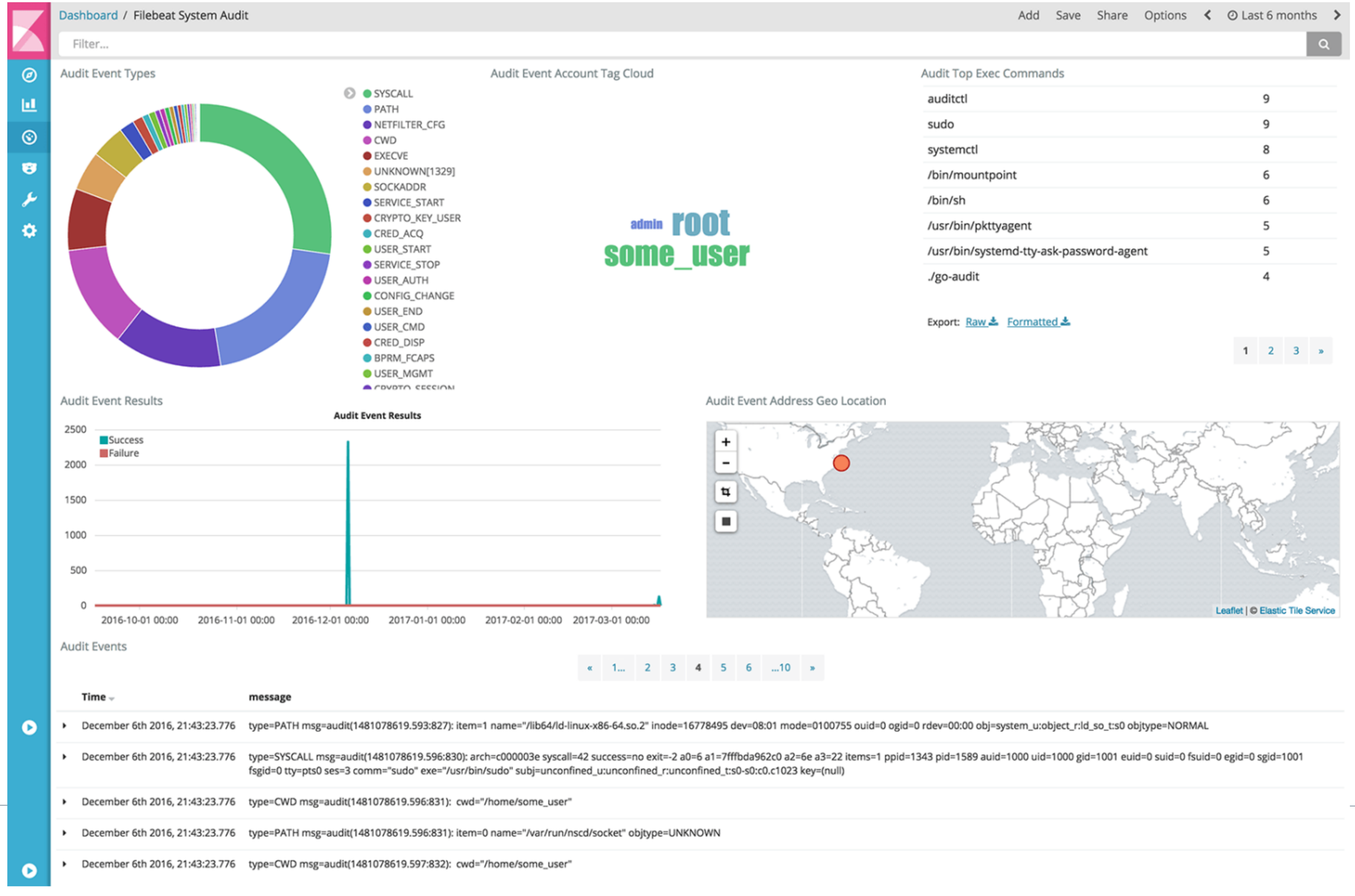
JMX Monitoring

Modules make the common tasks easy.

- Metricbeat 5.4 Jolokia Module
- Jolokia provides REST-like access to JMX with JSON over HTTP.



Linux auditd Filebeat Module



elastic

Linux system authentication logs

elastic

Elastic & Community

- 中文权威指南已上线！
- References 翻译已启动！
- 欢迎加入！
- QQ群： 109764489

elastic

Thanks !