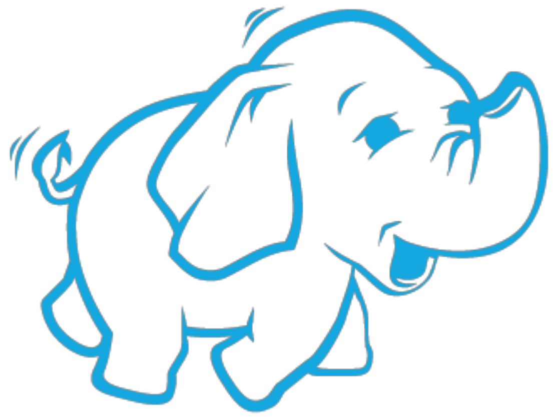
elastic

Elastic Stack & Machine Learning

Medcl - Elastic

About

- Medcl
  - http://github.com/medcl
  - https://elasticsearch.cn/people/medcl
- Elastic
  - http://github.com/elastic
  - https://www.elastic.co



高度伸缩，不容易实时，**总体拥有成本高**

键/值 存储，无模式,缺乏分析能力

结构化数据，复杂join，不支持非结构化数据

定制或
专有系统

单一用例，不是为支持多个用例而构建

有很多伟大的工具存在，但是还不够优雅。

水平伸缩

数据实时可用

灵活的数据模型

快速的查询执行

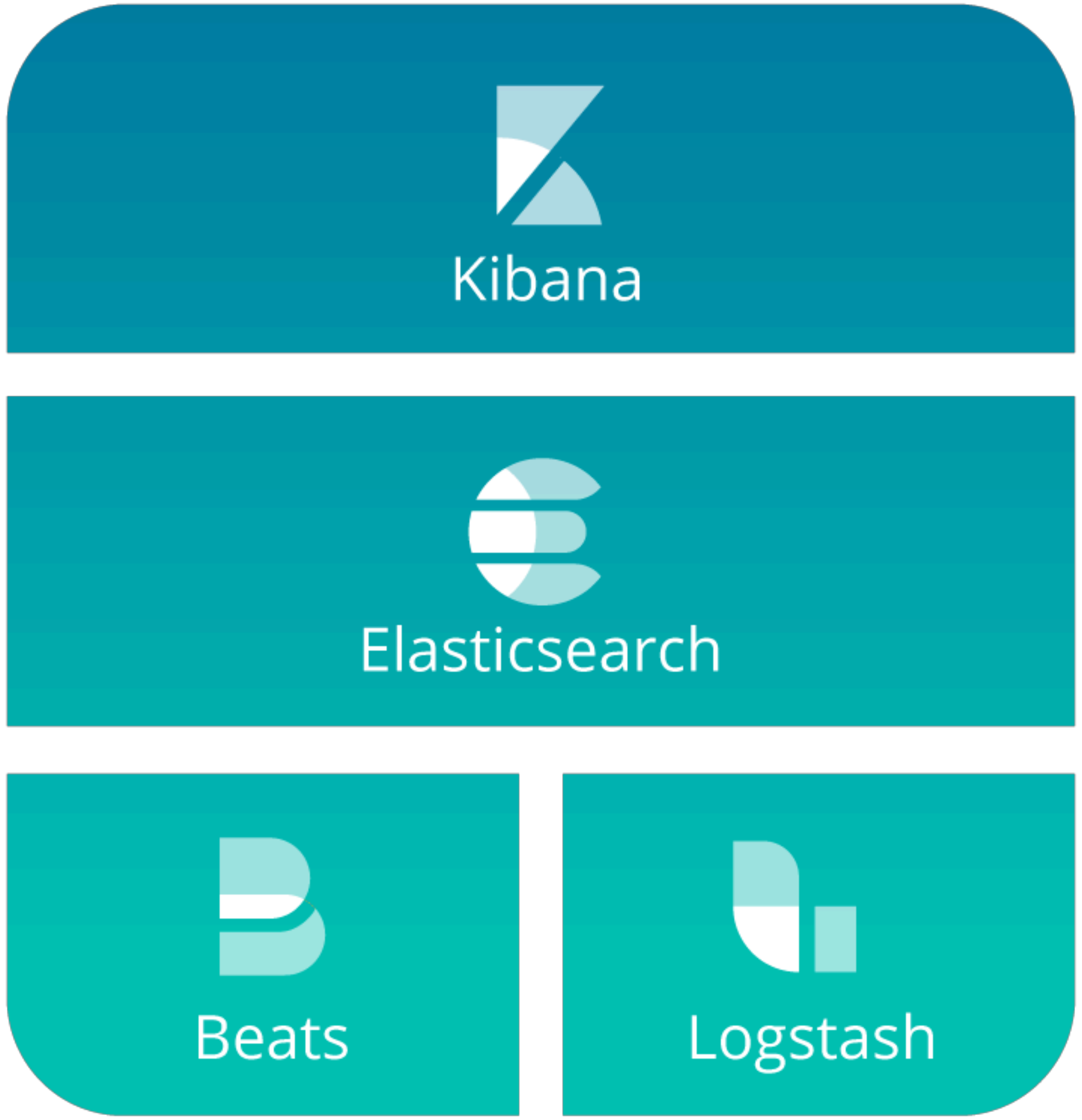
精致的查询语言

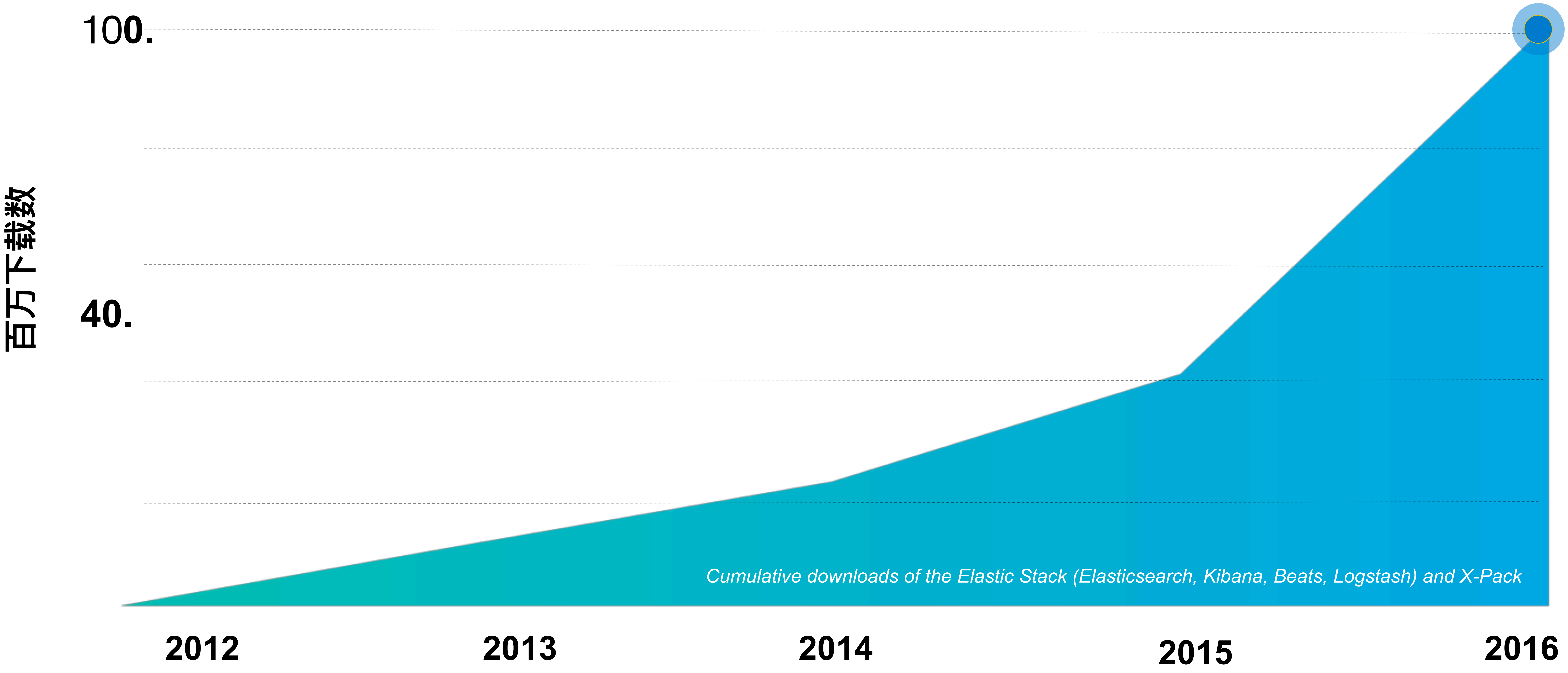
无模式

当今开发者们的要求

Elastic Stack

100% 开源
没有“企业版本”
5.0 全面升级





Cumulative downloads of the Elastic Stack (Elasticsearch, Kibana, Beats, Logstash) and X-Pack

科技	Adobe	hp	salesforce	Microsoft	CISCO	DELL
金融	SwissLife	Scottrade	Goldman Sachs	HSBC	E*TRADE	ZURICH
电信	verizon	COX COMMUNICATIONS	orange	ROGERS	Alcatel·Lucent	Sprint
消费	Audi	UBER	GROUPON	Columbia Sportswear Company	DHL	ebay

企业客户遍布每个行业

Elasticsearch

分布式，
可伸缩，
弹性

为水平伸缩而设计
具备高可用

开发者友好

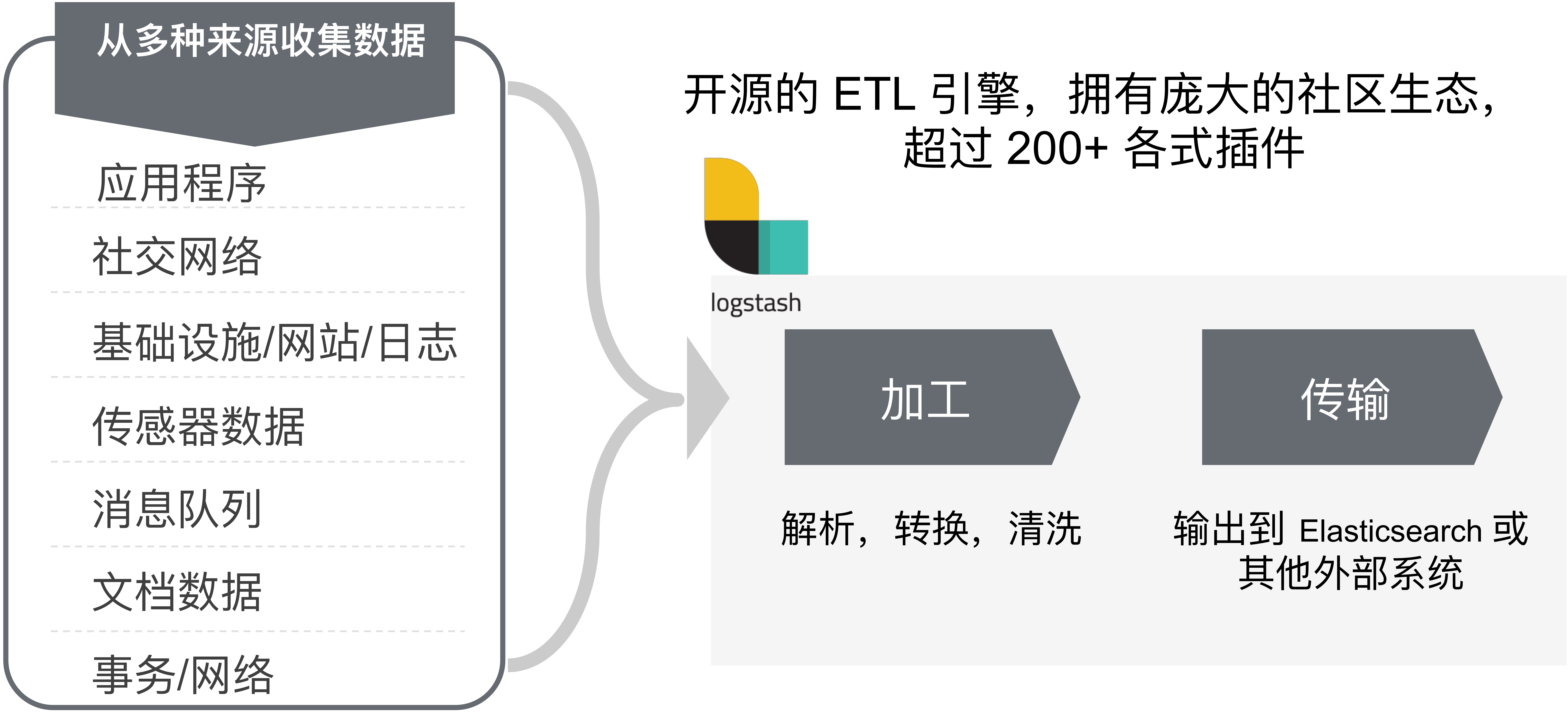
- API优先，RESTful
- 无模式
- 原生 JSON
- 多种客户端SDK
- Java/.NET/PHP/Python, etc

实时搜索
与分析

- 实时数据聚合
- 地理位置
- 全文检索
- 支持结构化和非结构化数据

Logstash

开源的 ETL 引擎，拥有庞大的社区生态，
超过 200+ 各式插件



从多种来源收集数据

- 应用程序
- 社交网络
- 基础设施/网站/日志
- 传感器数据
- 消息队列
- 文档数据
- 事务/网络

logstash

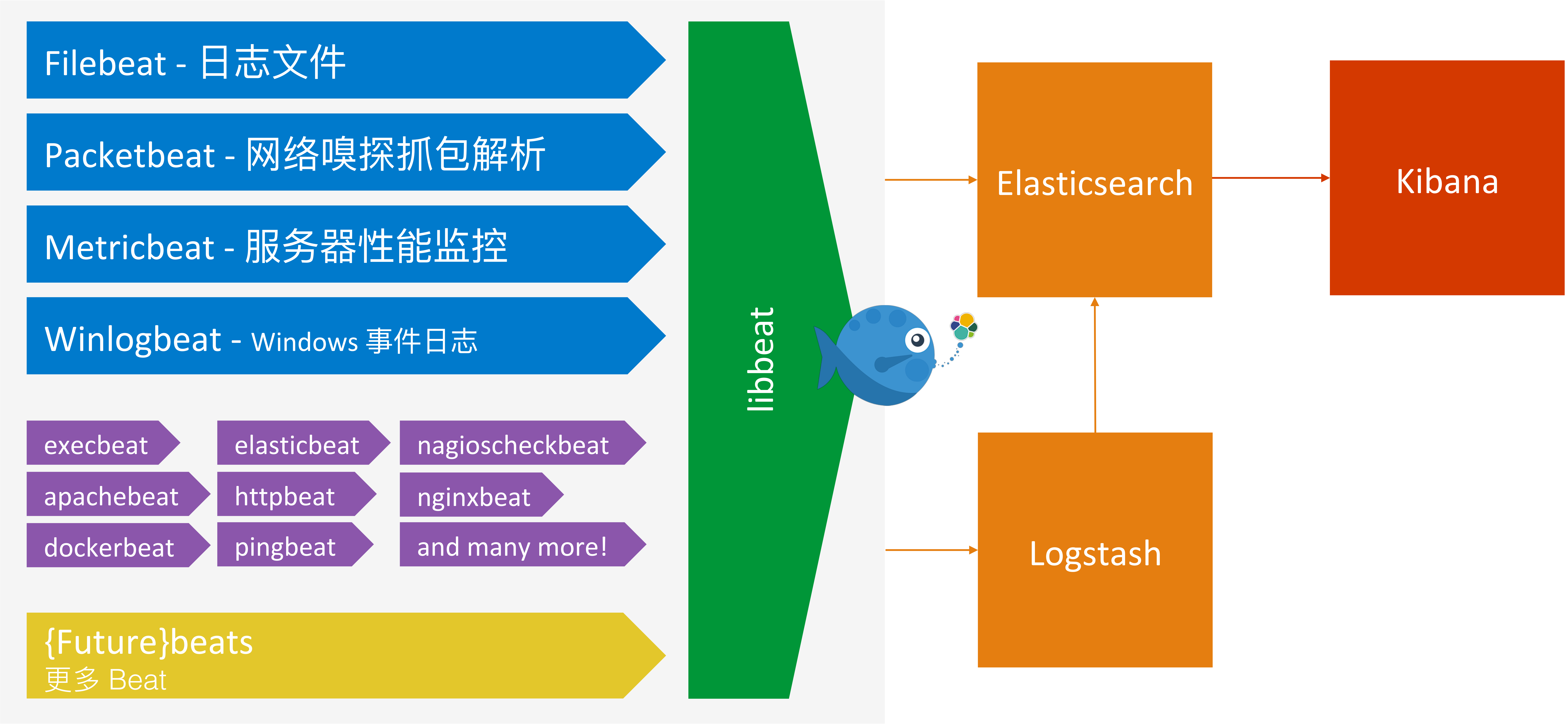
加工

解析，转换，清洗

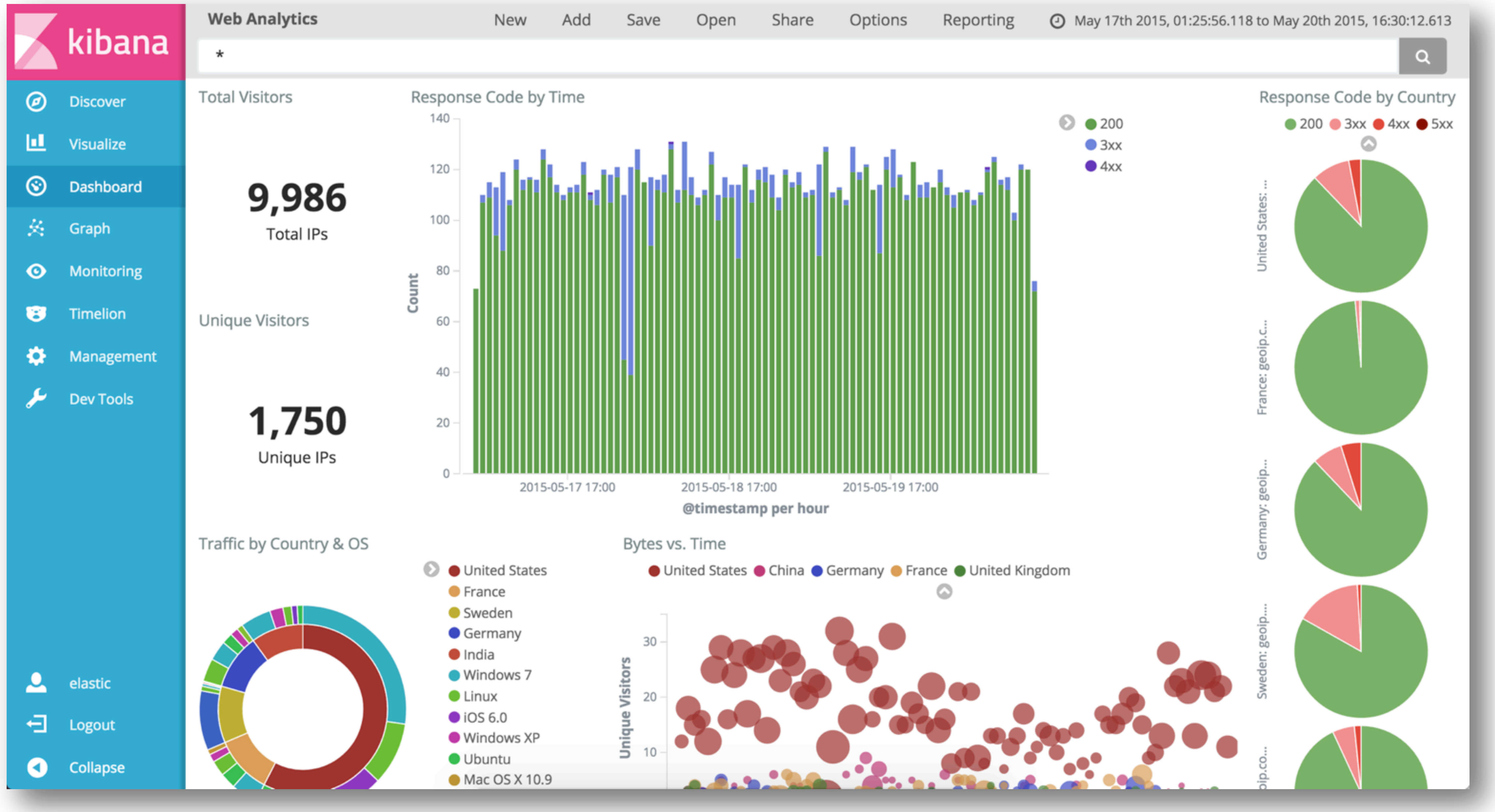
传输

输出到 Elasticsearch 或
其他外部系统

Beats

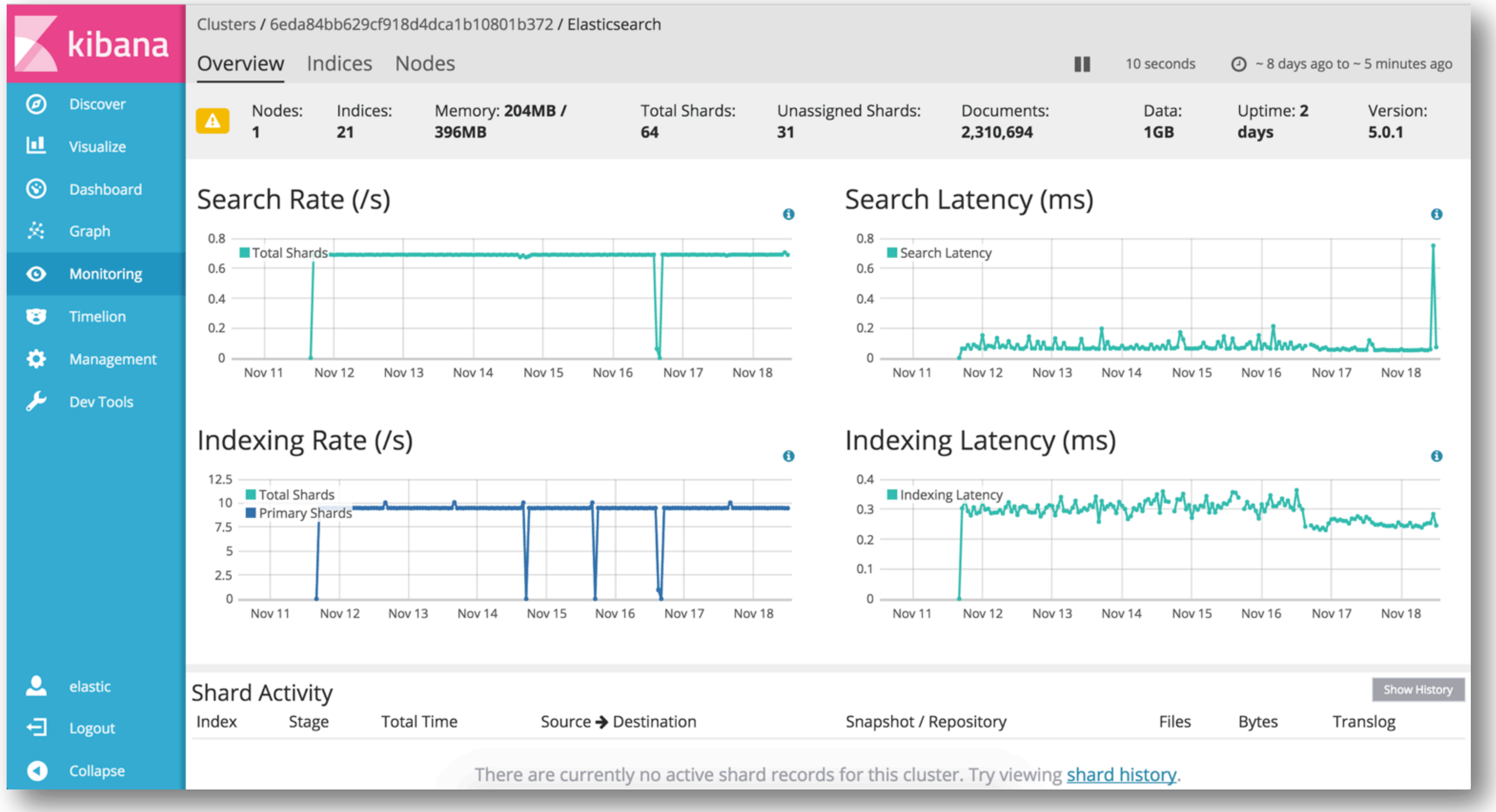

Filebeat - 日志文件
Packetbeat - 网络嗅探抓包解析
Metricbeat - 服务器性能监控
Winlogbeat - Windows 事件日志
execbeat
elasticbeat
nagioscheckbeat
apachebeat
httpbeat
nginxbeat
dockerbeat
pingbeat
and many more!
{Future}beats
更多 Beat
libbeat
Elasticsearch
Kibana
Logstash

Kibana



9,986
Total IPs

1,750
Unique IPs



Nodes: 1 | Indices: 21 | Memory: 204MB / 396MB | Total Shards: 64 | Unassigned Shards: 31 | Documents: 2,310,694 | Data: 1GB | Uptime: 2 days | Version: 5.0.1

Shard Activity

Index	Stage	Total Time	Source → Destination	Snapshot / Repository	Files	Bytes	Translog

There are currently no active shard records for this cluster. Try viewing shard history.

数据探索与可视化

- 即席发现与搜索
- 交互式图表与仪表盘
- 图关联与报表生成

管理

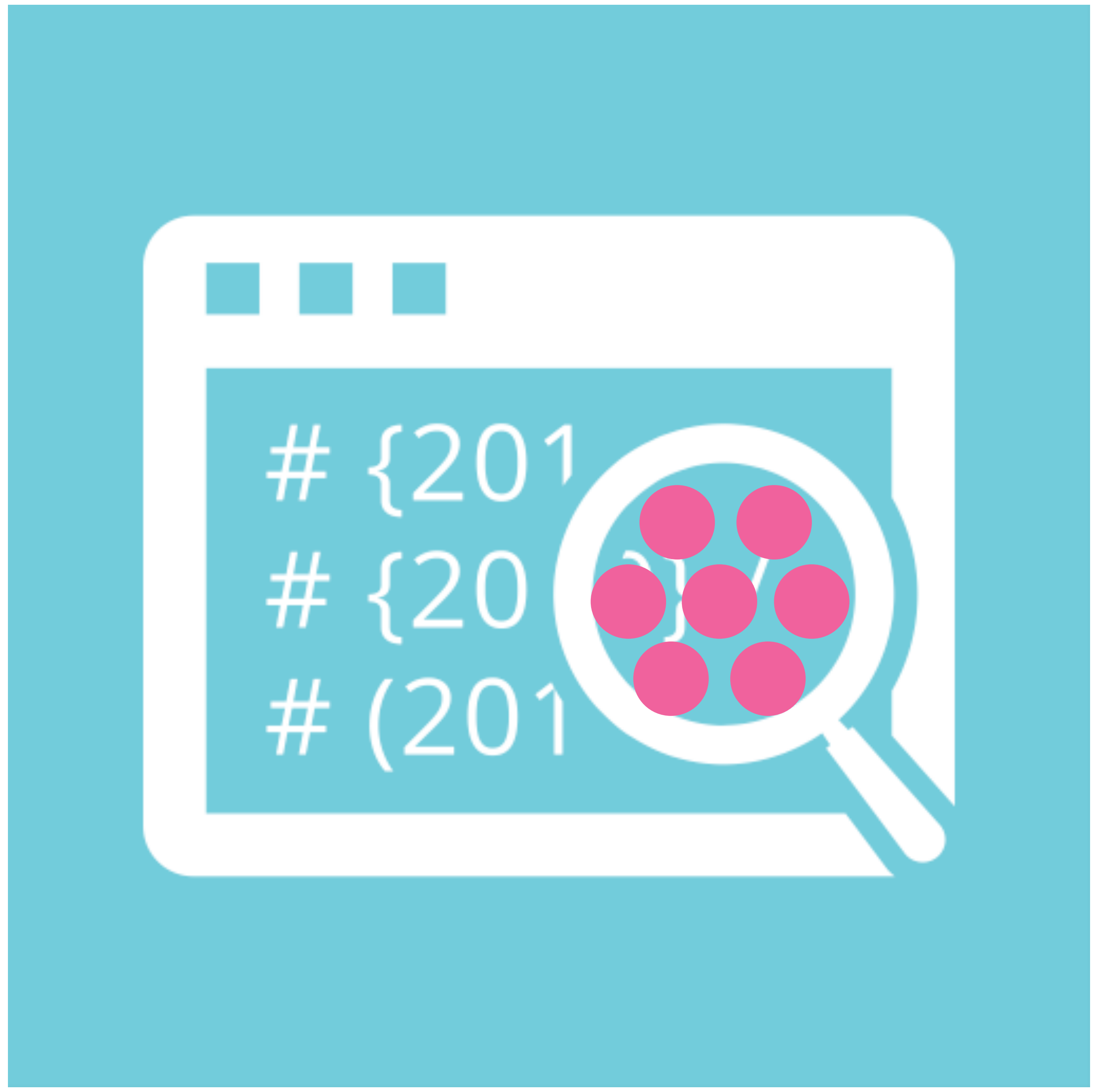
- 设置管理
- 开发工具
- 管理，监控

X-Pack

- Security
- Alerting
- Monitoring
- Reporting
- Graph
- Machine Learning

Machine Learning

Extracting useful, valuable information is hard



Search

Aggregations

Visualization

Machine Learning

Machine Learning is:

- a broad umbrella term for a variety of techniques and technologies
- a (often misused) marketing buzzword
- Many applications
  - Image recognition
  - Language translation
  - Recommendation
  - Anomaly Detection

Machine Learning[1]

Algorithms and methods for data driven prediction, decision making, and modelling

Supervised Learning

Prediction based on examples of correct behavior

Unsupervised Learning

No explicit target, only data, goal to model/discover

Semi-supervised Learning

Supplement limited annotations with unsupervised learning

Transfer Learning

How to apply what you have learned from A to B

Active Learning

Learn to query the examples actually needed for learning

Reinforcement Learning

Learning to act, not just predict; goal to optimize the consequences of actions

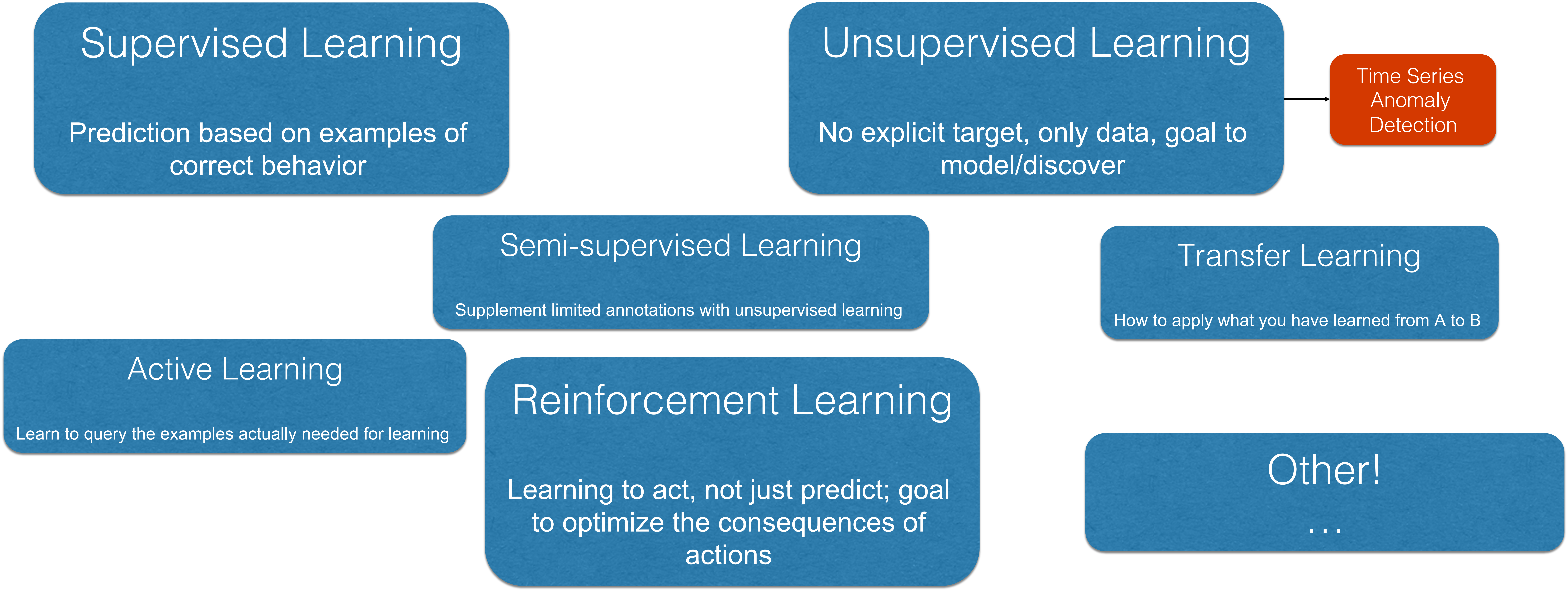
Other!

. . .

[1] Machine Learning Overview, Tommi Jaakkola, MIT

Machine Learning[1]

Algorithms and methods for data driven prediction, decision making, and modelling



Supervised Learning

Prediction based on examples of correct behavior

Unsupervised Learning

No explicit target, only data, goal to model/discover

Time Series Anomaly Detection

Semi-supervised Learning

Supplement limited annotations with unsupervised learning

Transfer Learning

How to apply what you have learned from A to B

Active Learning

Learn to query the examples actually needed for learning

Reinforcement Learning

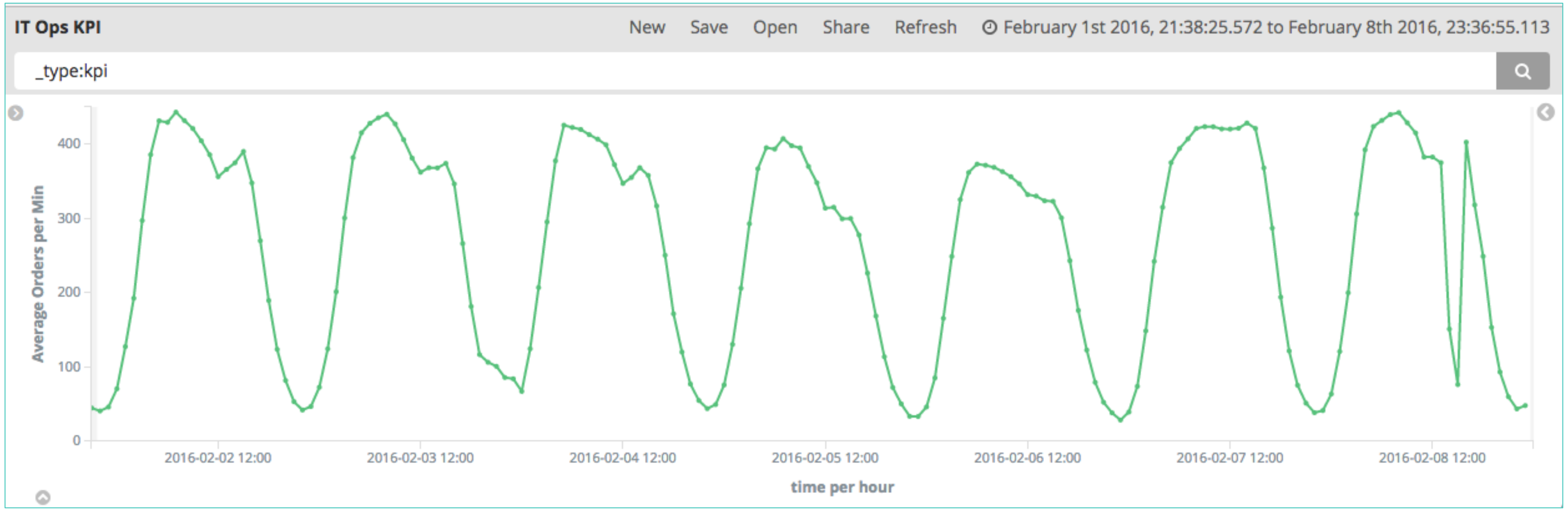
Learning to act, not just predict; goal to optimize the consequences of actions

Other!

. . .

[1] Machine Learning Overview, Tommi Jaakkola, MIT

Has my order rate dropped significantly?



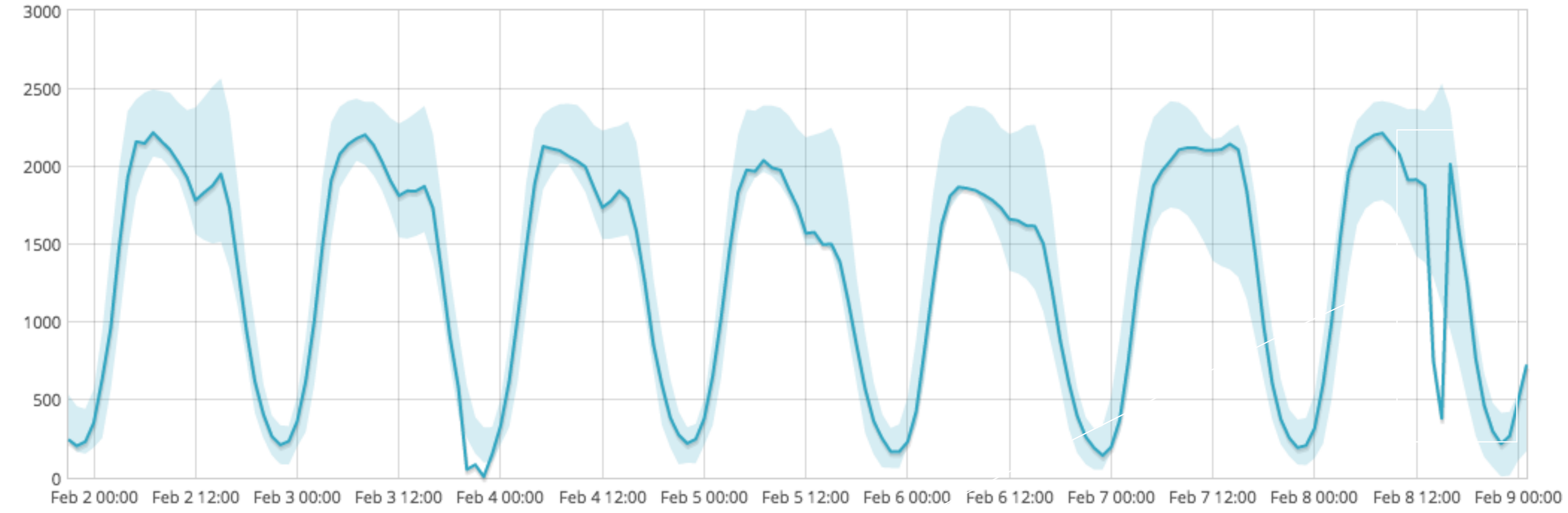
IT Ops KPI

February 1st 2016, 21:38:25.572 to February 8th 2016, 23:36:55.113

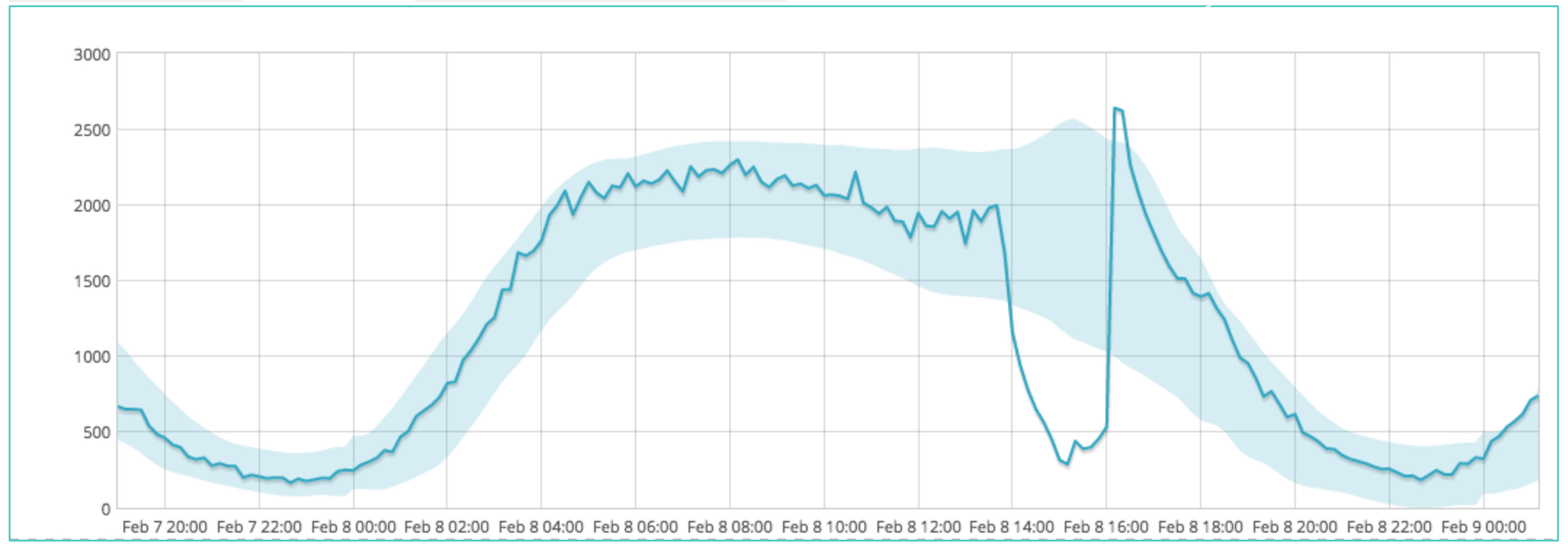
_type:kpi

Has my order rate dropped significantly?

- Learn models from past behaviour (training, modelling)

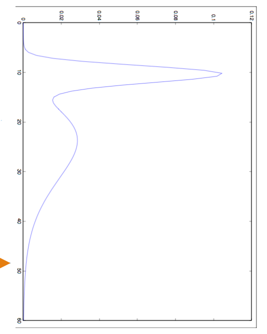


- Use models to predict future behaviour (prediction)

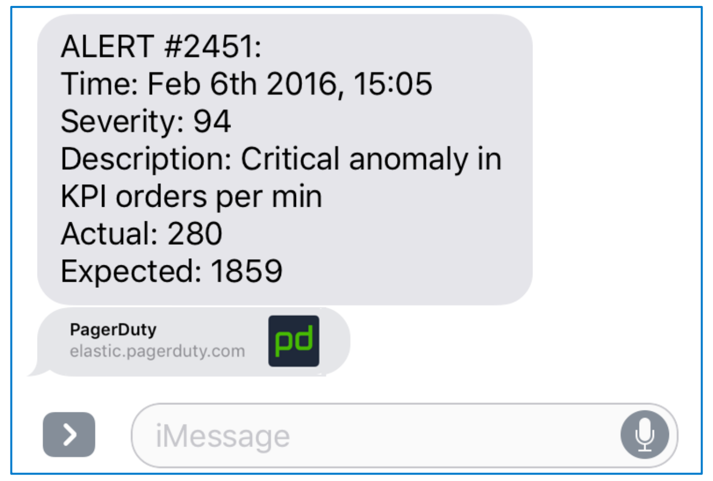


Actual value @ 15:05 = 280

Probability = 0.0000174025



- Use predictions to make decisions



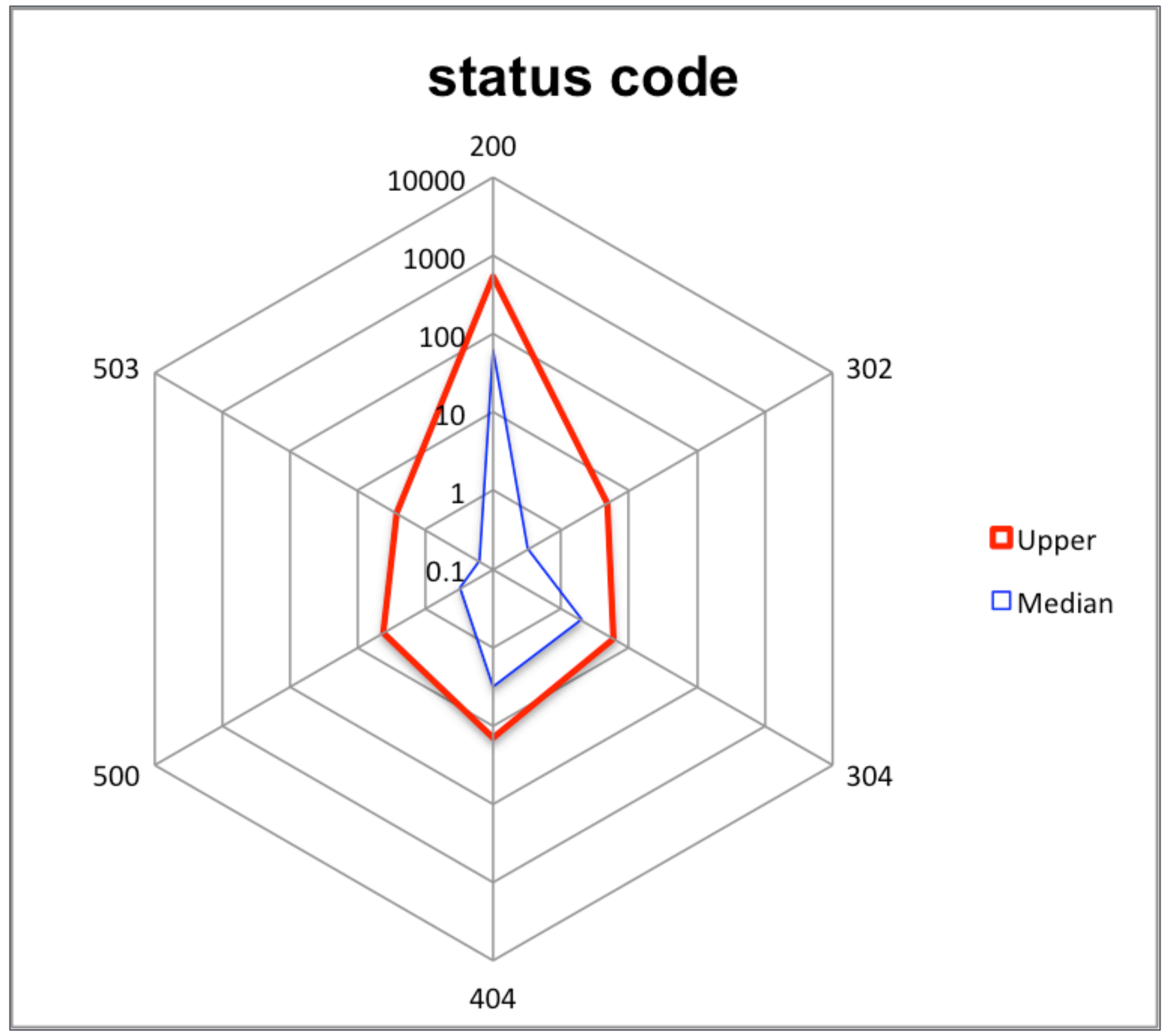
ALERT #2451:
Time: Feb 6th 2016, 15:05
Severity: 94
Description: Critical anomaly in KPI orders per min
Actual: 280
Expected: 1859

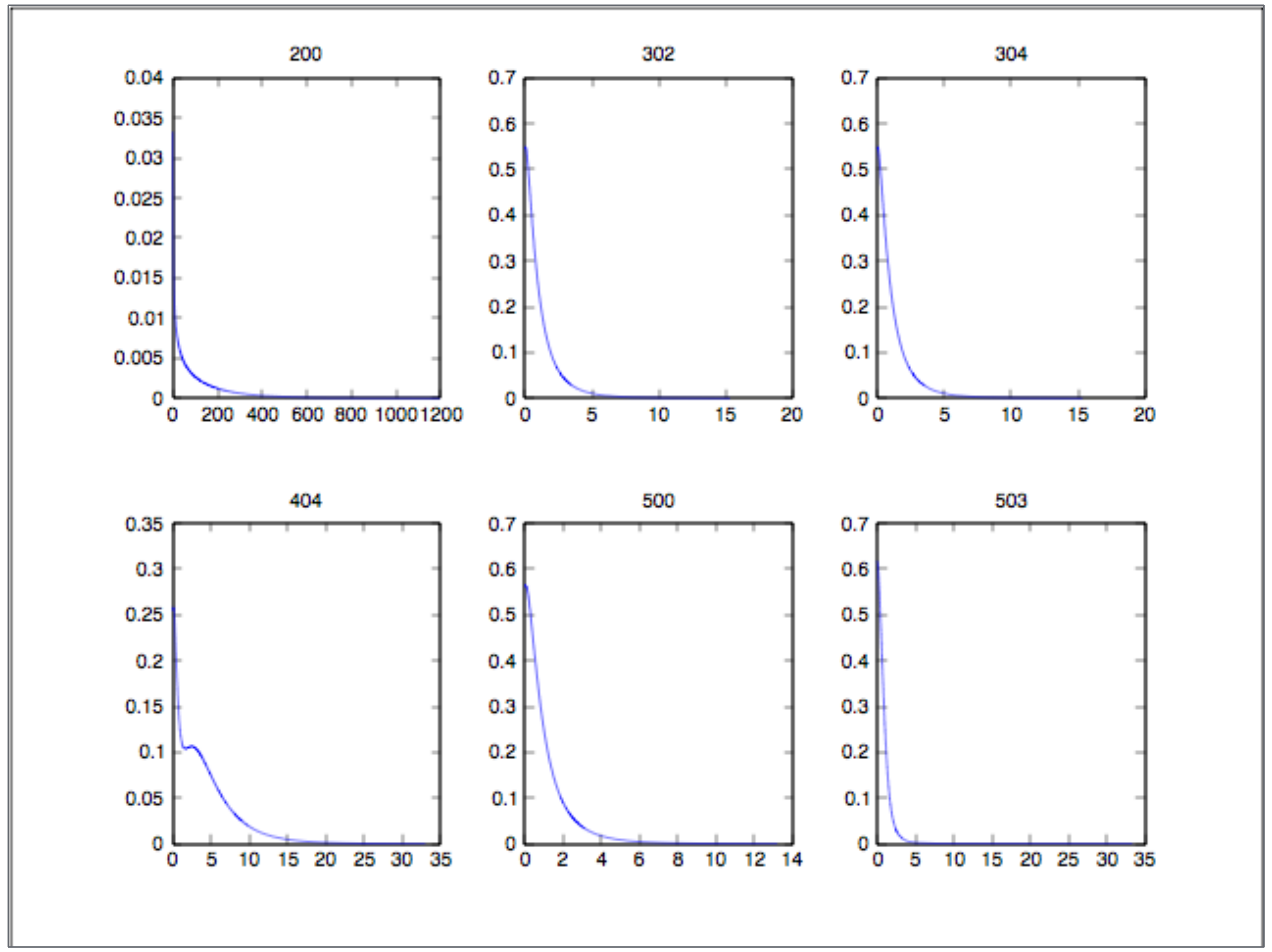
PagerDuty
elastic.pagerduty.com

Entity Profiling

```
10.12.211.69 - - [01/Jan/2016:00:07:21 +0000] "GET /css/ccc_style.jsp HTTP/1.1" 200 19196 "https://www.prelertstation.com/" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5"
10.12.211.69 - - [01/Jan/2016:00:07:22 +0000] "GET /js/openWin.js HTTP/1.1" 200 2272 "https://www.prelertstation.com/" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5"
10.12.211.69 - - [01/Jan/2016:00:07:22 +0000] "GET /css/themes/ HTTP/1.1" 404 988 "https://www.prelertstation.com/" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5"
```

- Create ‘profile’ of status code responses for a typical client:

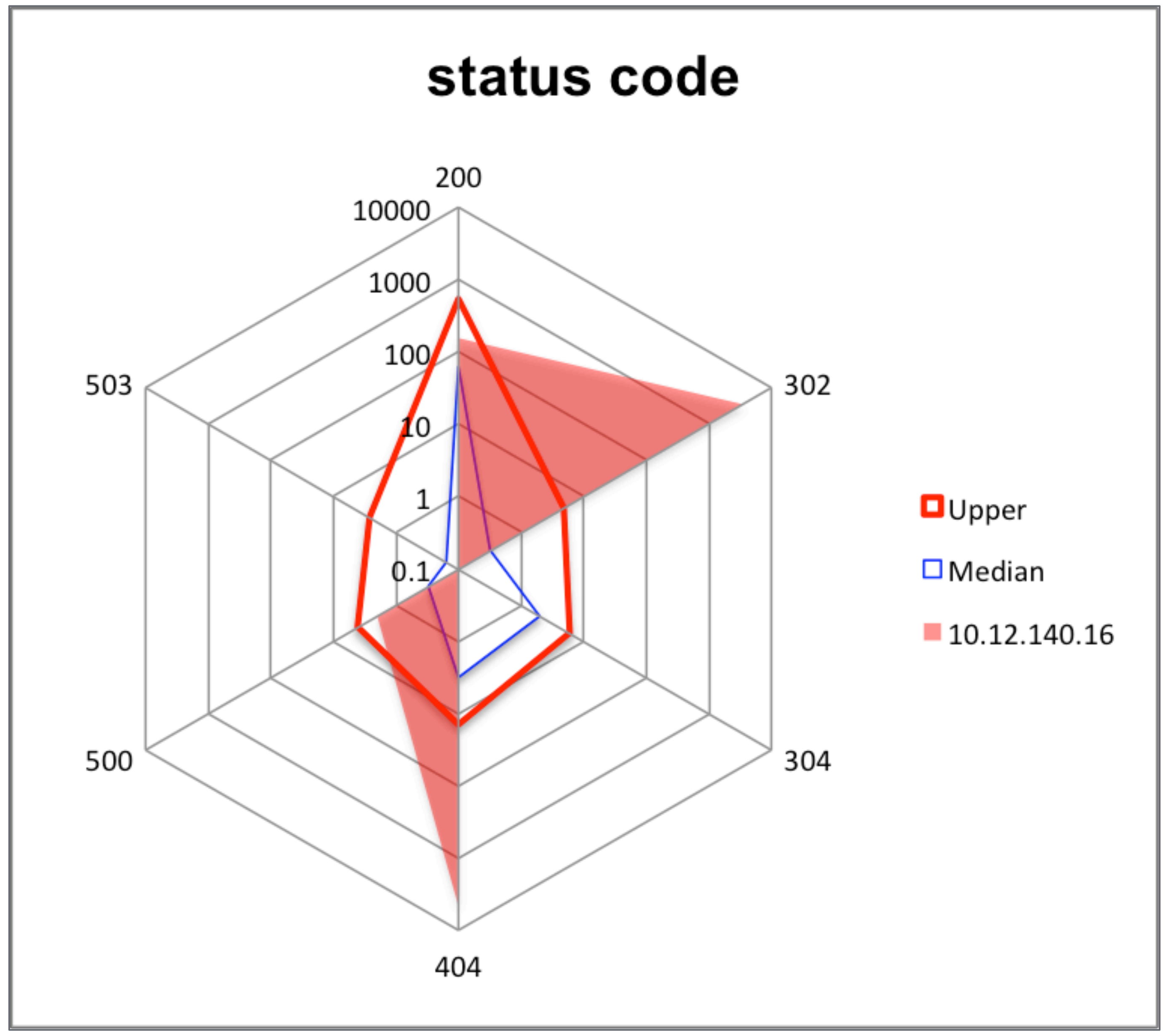


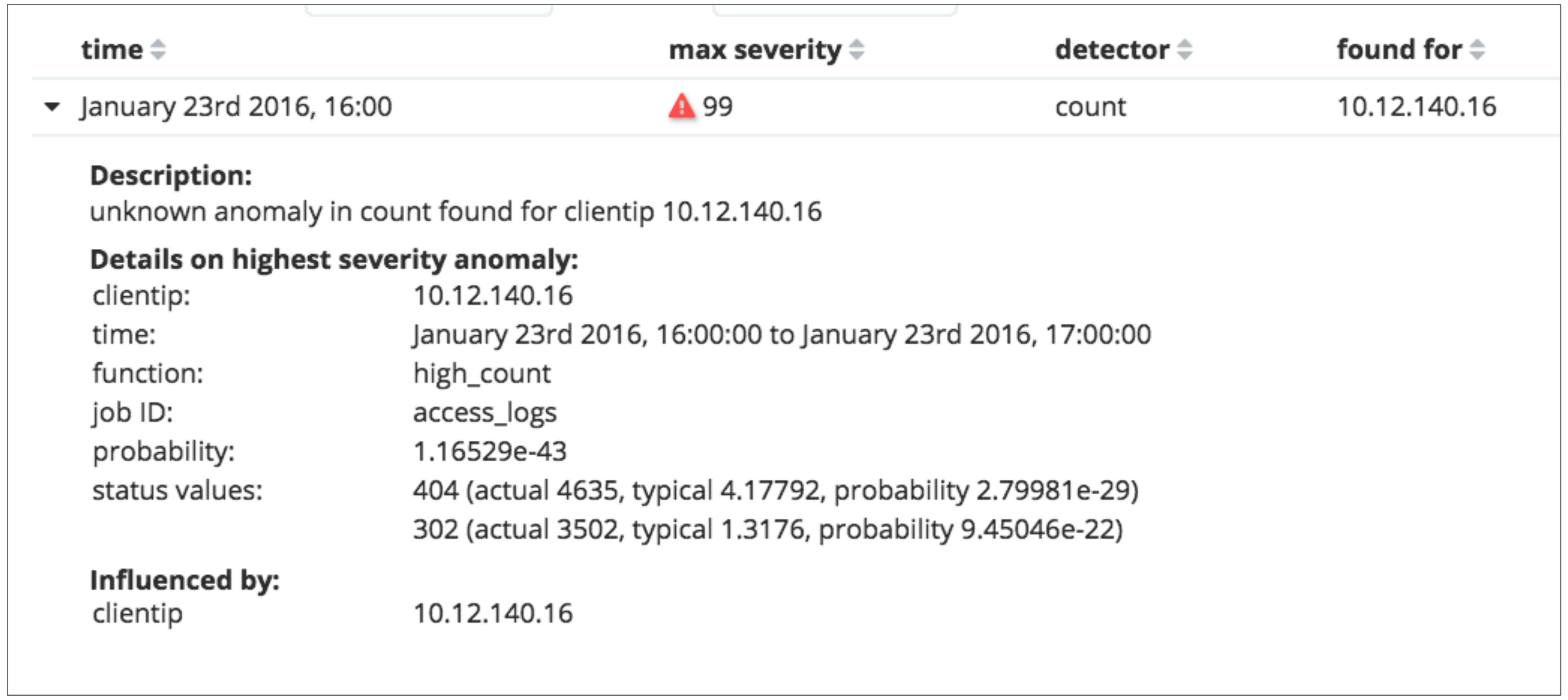


Entity Profiling

```
10.12.211.69 - - [01/Jan/2016:00:07:21 +0000] "GET /css/ccc_style.jsp HTTP/1.1" 200 19196 "https://www.prelertstation.com/" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5"
10.12.211.69 - - [01/Jan/2016:00:07:22 +0000] "GET /js/openWin.js HTTP/1.1" 200 2272 "https://www.prelertstation.com/" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5"
10.12.211.69 - - [01/Jan/2016:00:07:22 +0000] "GET /css/themes/ HTTP/1.1" 404 988 "https://www.prelertstation.com/" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.5) Gecko/2008120122 Firefox/3.0.5"
```

- Create ‘profile’ of status code responses for a typical client:





time	max severity	detector	found for
January 23rd 2016, 16:00	⚠ 99	count	10.12.140.16

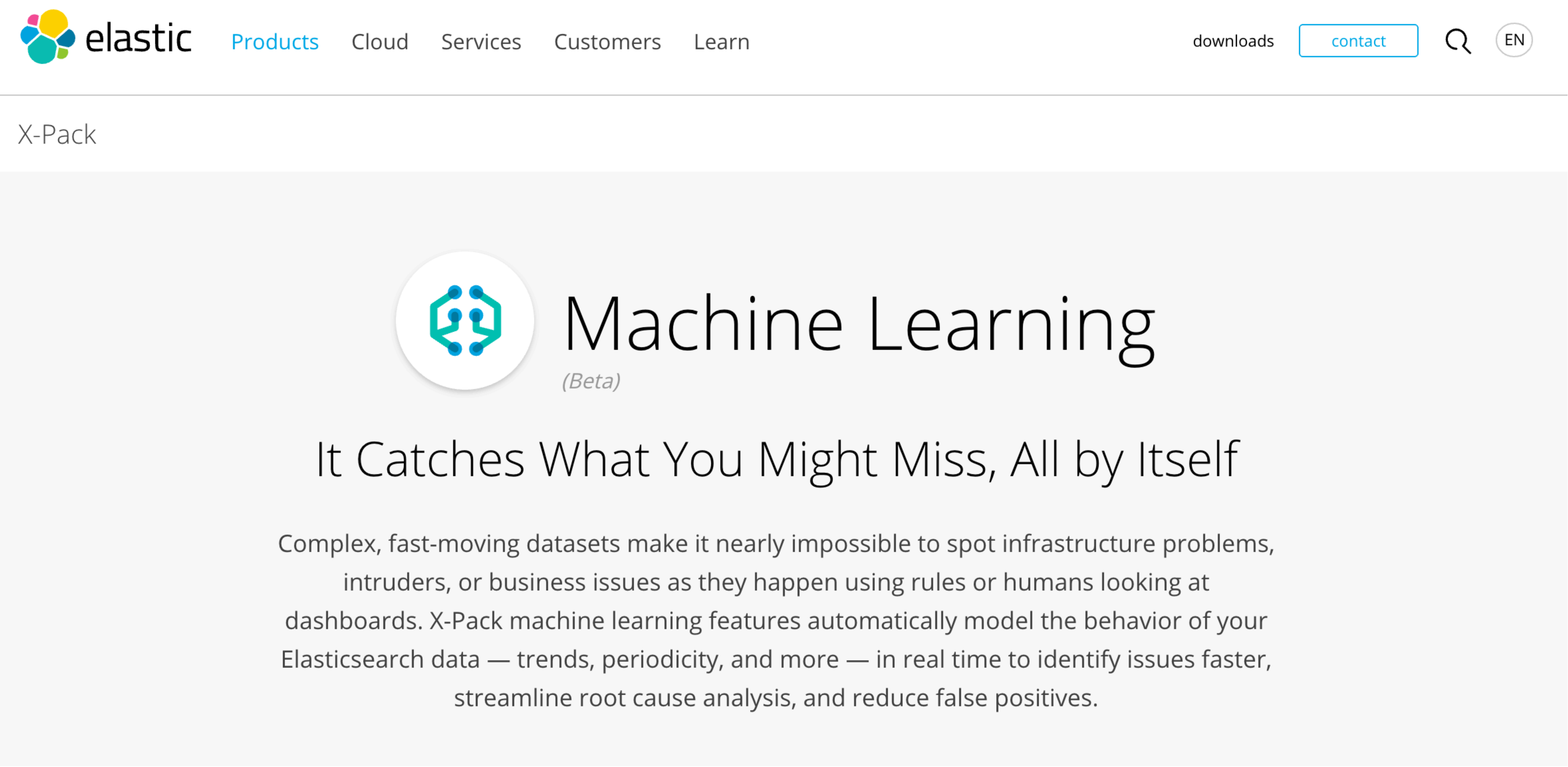
Description:
unknown anomaly in count found for clientip 10.12.140.16

Details on highest severity anomaly:
clientip: 10.12.140.16
time: January 23rd 2016, 16:00:00 to January 23rd 2016, 17:00:00
function: high_count
job ID: access_logs
probability: 1.16529e-43
status values: 404 (actual 4635, typical 4.17792, probability 2.79981e-29)
302 (actual 3502, typical 1.3176, probability 9.45046e-22)

Influenced by:
clientip 10.12.140.16

Demo!

Try it yourself!



X-Pack

Machine Learning

(Beta)

It Catches What You Might Miss, All by Itself

Complex, fast-moving datasets make it nearly impossible to spot infrastructure problems, intruders, or business issues as they happen using rules or humans looking at dashboards. X-Pack machine learning features automatically model the behavior of your Elasticsearch data — trends, periodicity, and more — in real time to identify issues faster, streamline root cause analysis, and reduce false positives.

https://www.elastic.co/products/x-pack/machine-learning elastic

www.elastic.co