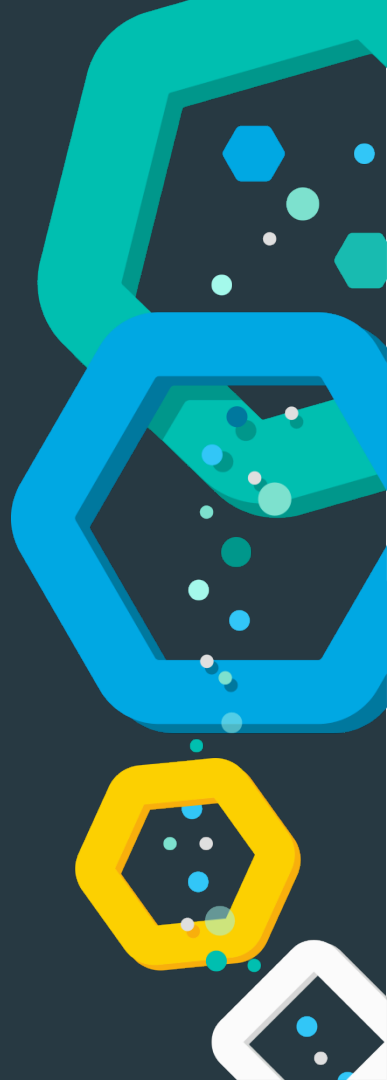
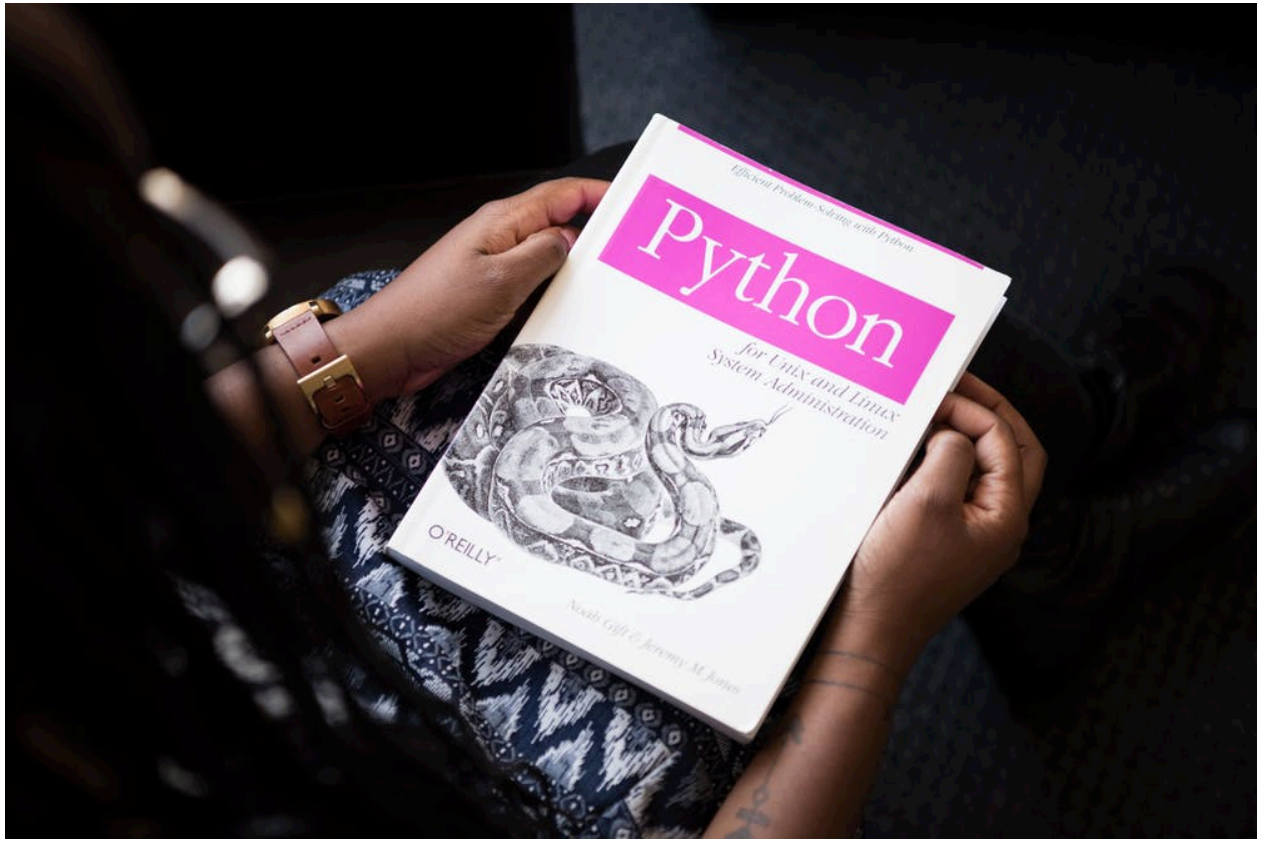
elastic

原来机器学习还可以这样玩！

8月24日
北京



你心目中的机器学习是什么样子?



elastic

你可能正在用.....

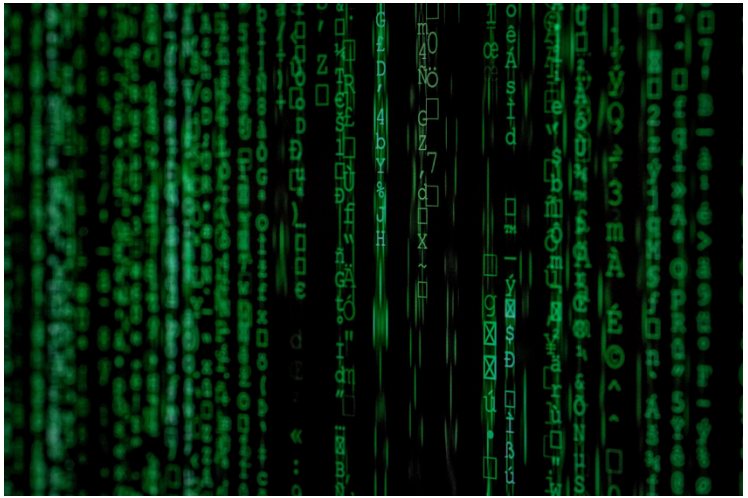
1. Pandas 是一个开放源码、BSD许可的库，为Python编程语言提供高性能、易于使用的数据结构和数据分析工具。

2. Scikit-learn是专门面向机器学习的Python开源框架，它实现了各种成熟的算法，并且易于安装与使用。

3. Jupyter Notebook 是一个 Web 应用程序，便于创建和共享文学化程序文档，支持实时代码、数学方程、可视化和 Markdown，其用途包括数据清理和转换、数值模拟、统计建模、机器学习等等。

更早的时候你可能在用.....

过去发生.....

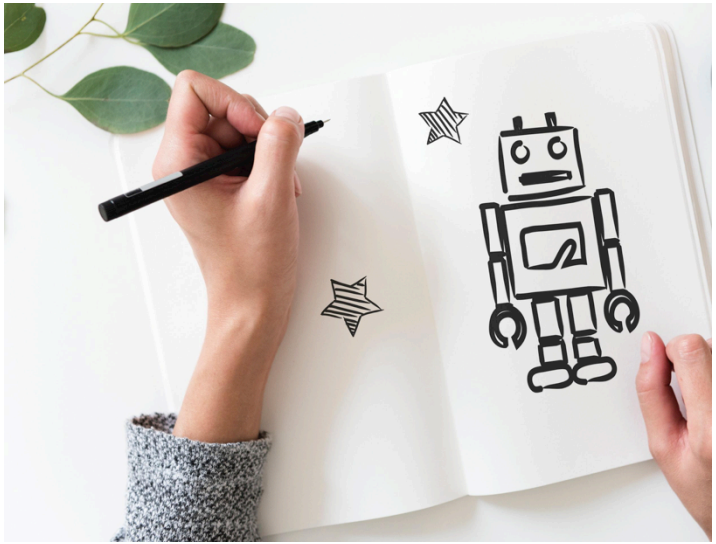


数据

计算能力

elastic

正在发生.....



自动

从洞察到行动

elastic

下一波颠覆性技术 - 增强型分析

Gartner

Gartner确认2019年十大数据与分析技术趋势之一增强性分析：增强型分析利用机器学习（ML）与人工智能改变分析内容的开发、消费与共享方式。

- 到2020年，增强型分析将成为分析与商业智能、数据科学与机器学习平台以及嵌入式分析新购的主要驱动力；
- 增强型分析能够鉴别隐藏的模式，同时消除人为的偏见；增强型分析和自动洞察将最终嵌入到企业的应用之中；

“By 2020, more than 40% of data science tasks will be ***automated***”

Source：https://www.gartner.com/en/newsroom/press-releases/2019-02-18-gartner-identifies-top-10-data-and-analytics-technolo

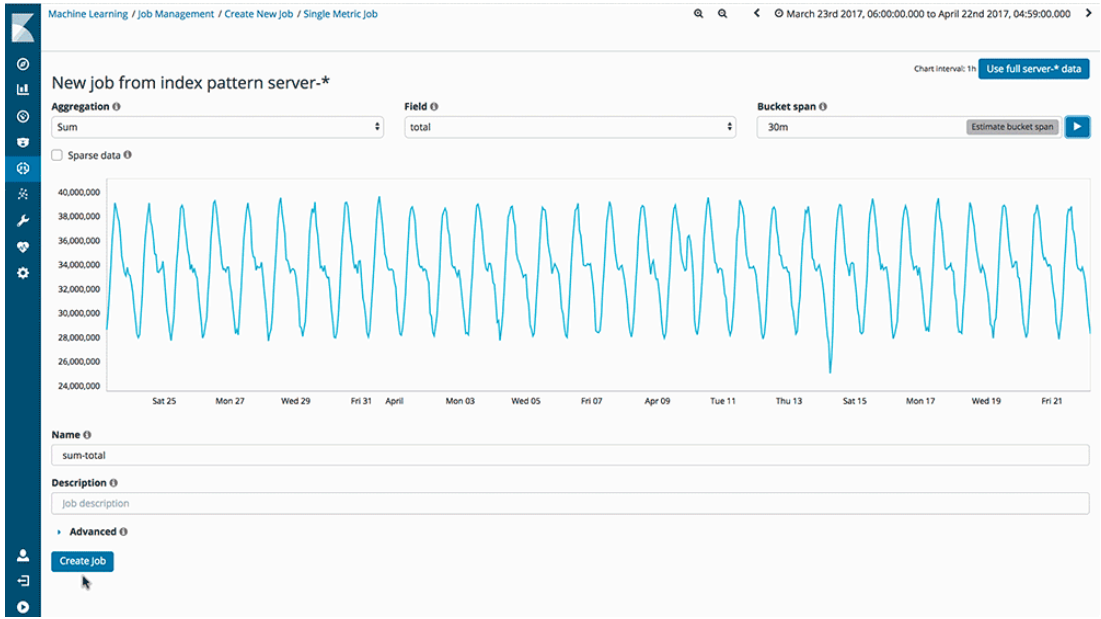
elastic

Elastic 机器学习

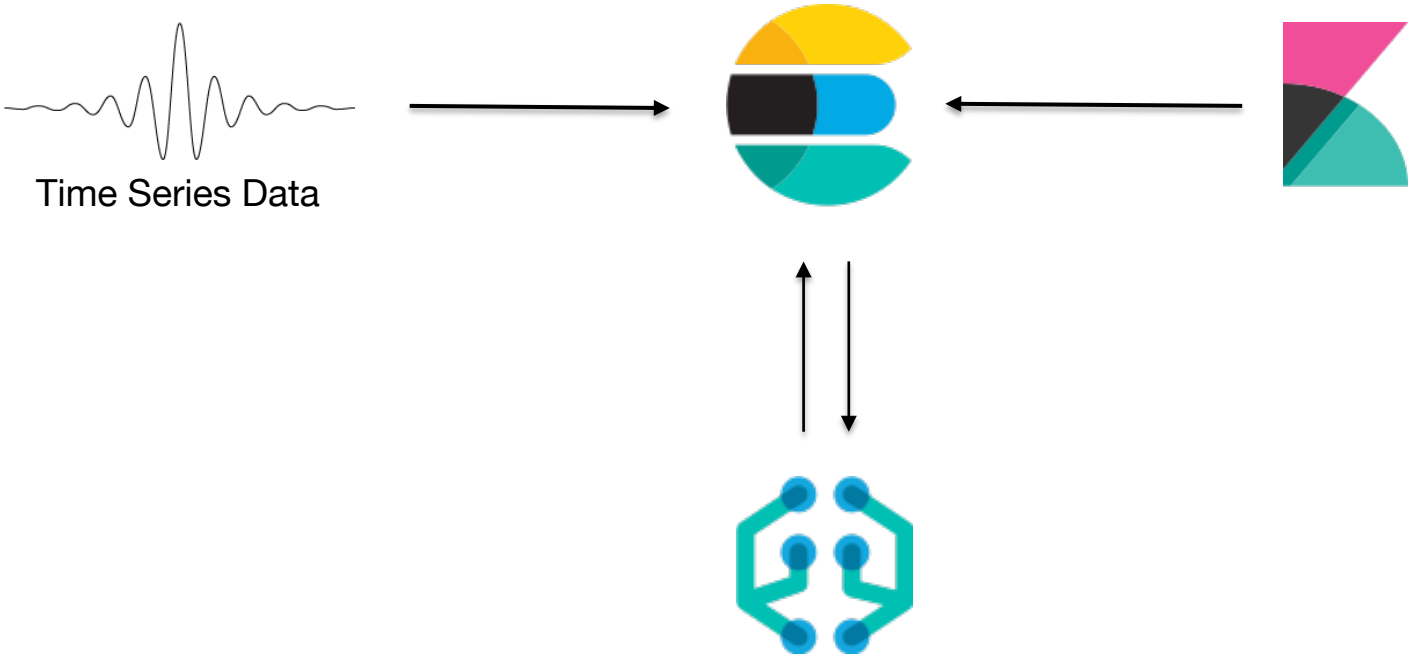
- 互动的
- 自动的机器学习任务
- 异常侦测
- 行为分析
- 预测

而不是:

- 通用的机器学习
- 魔法



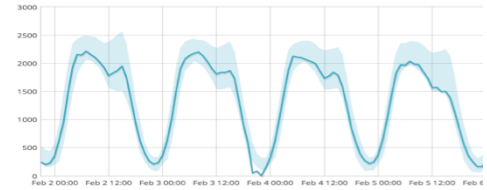
Elastic 机器学习工作流



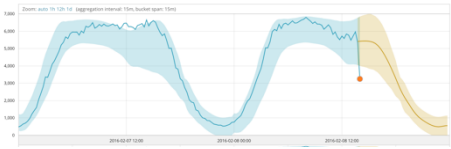
elastic

Elastic 机器学习工作流

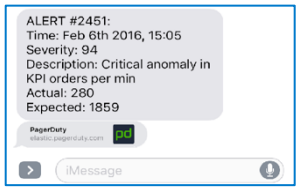
学习



预测



操作

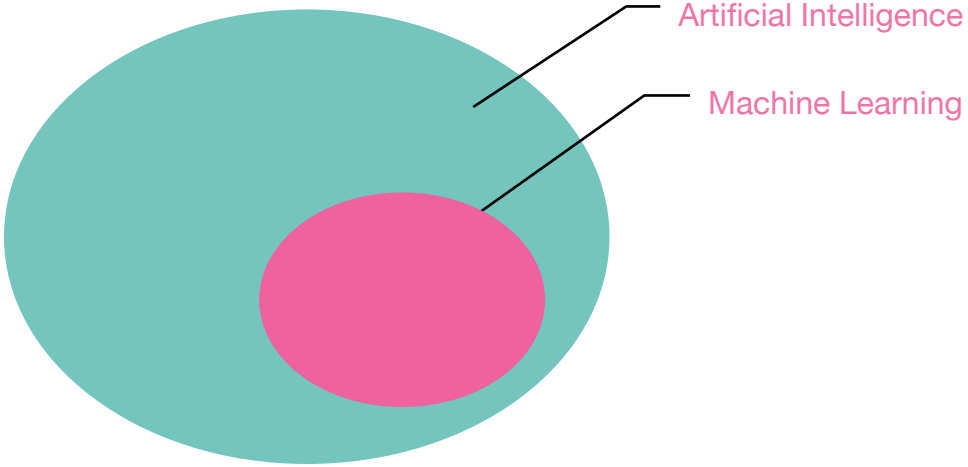


- 从数据中学习行为模式
- 自动生成模型，判断当前和未来的行为
- 考虑周期和季节性因素
- 对异常进行打分，按照严重程度告警
- 预测未来的行为
- 对历史和实时数据进行操作型实时分析
- 易于使用，提供可视化界面
- 提供Restful API

elastic

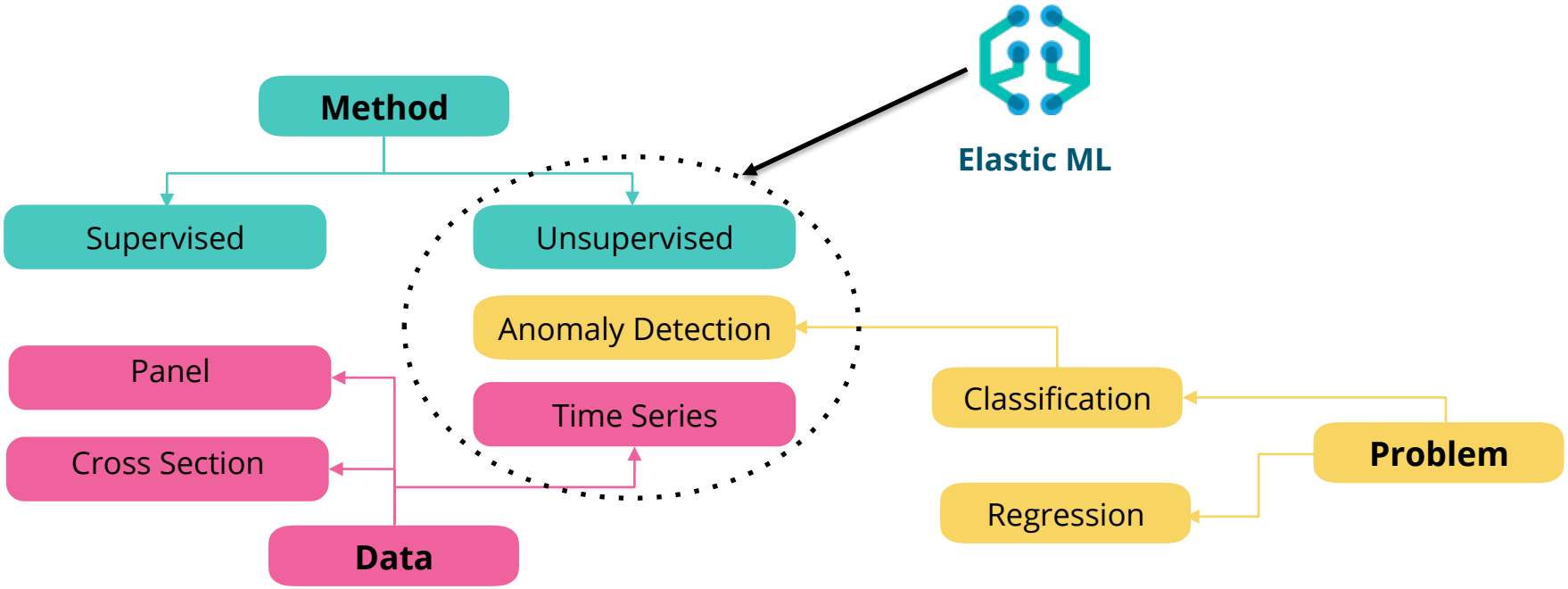
Elastic 机器学习能力范围

- 从数据中学
- 使用统计学和概率
- 无需编写程序



elastic

Elastic机器学习能力范围



elastic

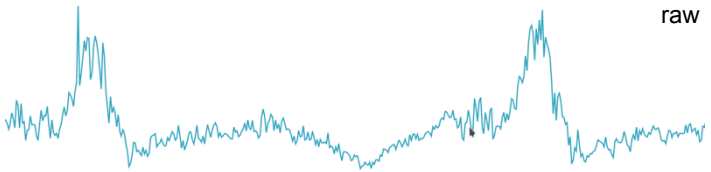
Elastic 机器学习原理

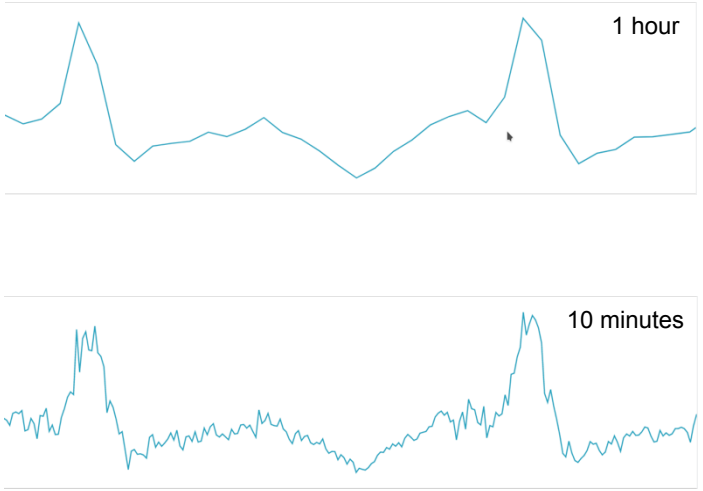
黑盒子里面发生了什么?

- 了解基本的概念
- 什么是模型? 有多少模型?
- 从检测到推测

Elastic 机器学习原理

数据按照数据桶来进行聚合

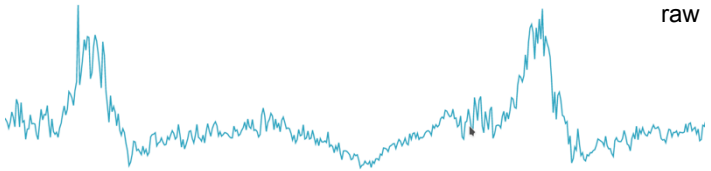


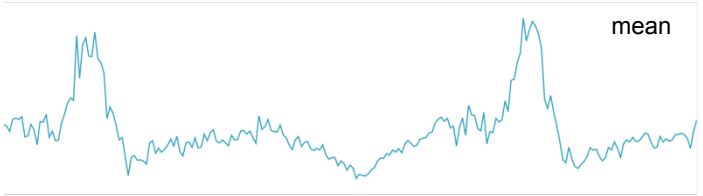


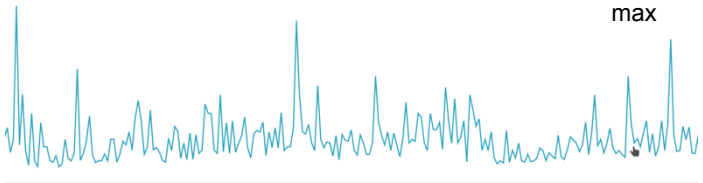
elastic

Elastic 机器学习原理

函数定义了数据如何转换







elastic

Elastic 机器学习原理

使用到的算法

- 时序数据分解算法
- 相关分析
- 贝叶斯分布
- 聚类
- 分类

贝叶斯

过去发生的概率，满足某种概率分布，例如泊松分布

后验概率 = 标准相似度*先验概率

从统计学的角度，来找到相似度

7.3之后，数据帧来了..

更强的数据转换能力

更强的OLAP特性

elastic

Elastic机器学习的五大分析

基于自身历史行为的模型分析，通俗的说“自己和自己比”来甄别异常

基于种群内行为的模型分析，通俗的说“自己和种群中的其他个体比”来甄别异常

从海量数据中找出出现频率非常低的数据，例如从大量的错误信息中找出罕见的错误代码

透过对字符文本类的数据的聚类分组分析，无需理解字符文本含义，而定位到异常的信息

7.3新特性，以距离或者密度为基准的分类分析

为什么选择Elastic的机器学习

- IT系统过于复杂难以人工分析
- 人的限制：经验和人力
- 自动化过程提升效率
- 可以分析更多的数据
- 动态告警，而不是基于规则和阀值的静态告警
- 主导探测异常行为
- 找到相关的可能的原因

elastic

IT运维：日志和监控

告警
归因分析

安全分析

威胁狩猎

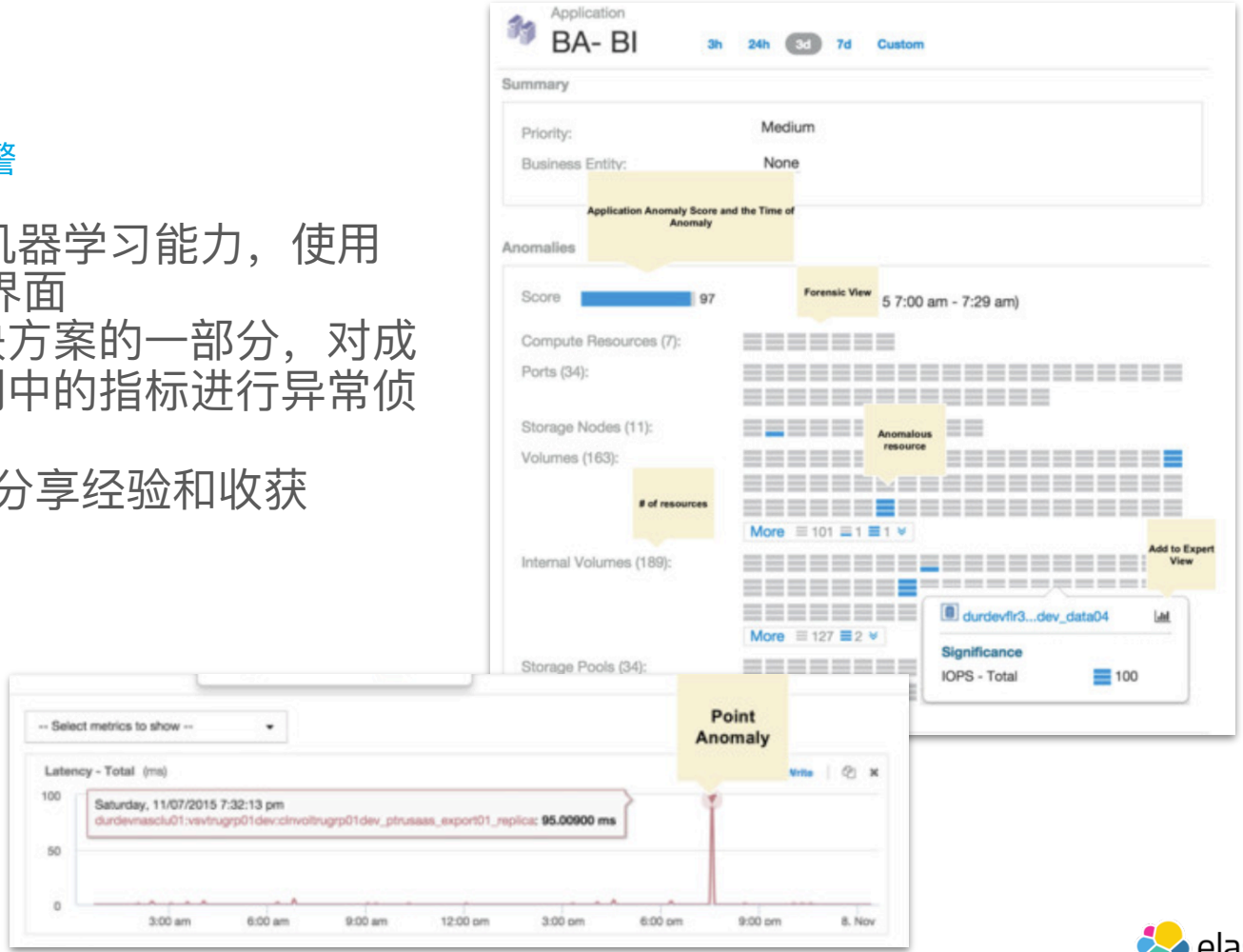
商业分析

趋势分析
KPI分析

NetApp

基于大量数据的运维预警

- OEM使用Elastic机器学习能力，使用NetApp的定制化界面
- 作为存储管理解决方案的一部分，对成百上千的存储阵列中的指标进行异常侦测
- Elastic{ON} 2018分享经验和收获



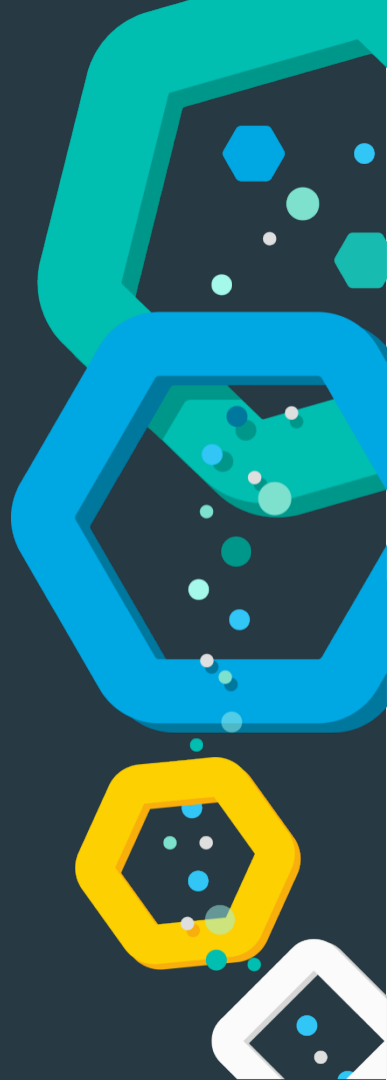
elastic

商业分析

- 跟踪客户在试用期间的转换率
- 发现的严重异常的事件发到Slack的交流组供进一步分析
- Elastic{ON} 2018 分享经验和收获

elastic

Let‘s Demo



机器学习“玩”起来

可用的公开数据：

1. https://github.com/elastic/examples/tree/master/Machine%20Learning
2. Machine Learning with the Elastic Stack　试验数据
3. https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq

感谢有你，一路同行！

Elastic社区活动的机器学习的演讲
2019年8月

elastic

elastic
中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区
https://elasticsearch.cn/