



Elastic Stack 7.x 最新技术更新

朱杰 Jerry
Elastic 架构师



2019.4.10

2019.5.20

2019.6.26

2019.8.1

2019.10.2



2019.12.3

Elasticsearch 新功能

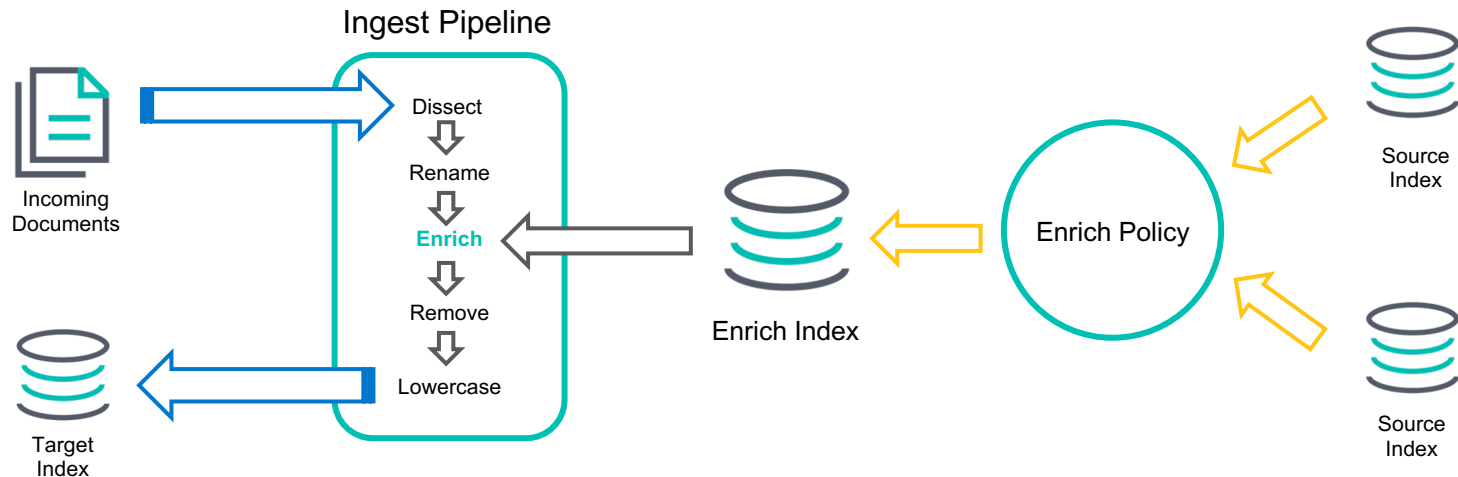
查询&分析

运维&架构

新的数据增强pipeline

7.5

工作原理



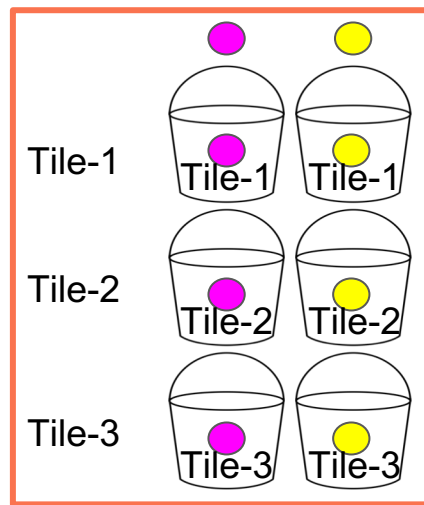
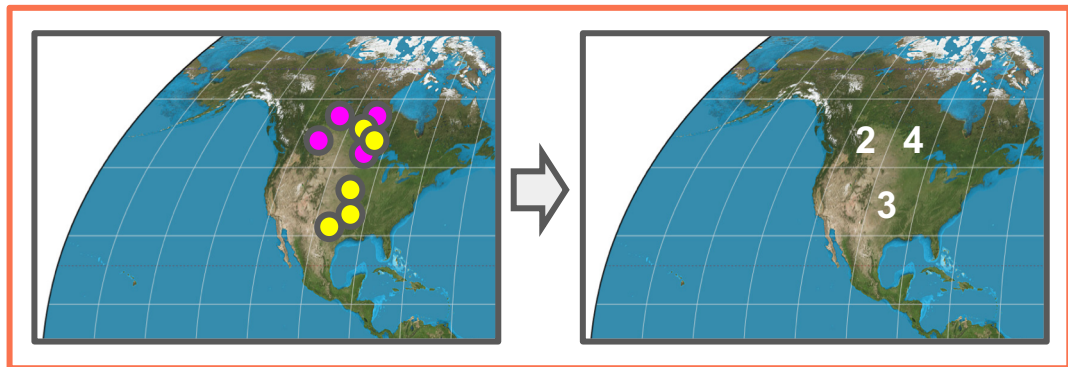
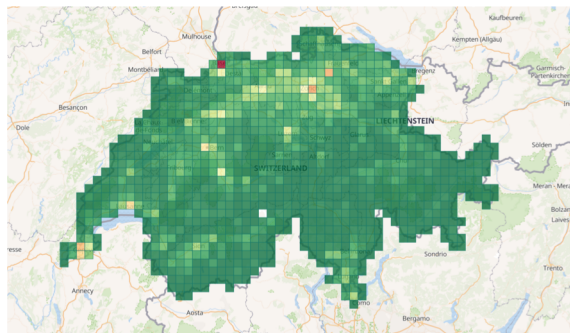
- 用户界面会后续推出
- 现在支持两种Policy Types: Match 和 Geo Match

高精度Geotile分析

7.5

可以在 GeoTile Grid上面使用 Composite Aggregation

- 结果集的遍历更加节省内存
 - 在地图中更细粒度的分片颜色
 - 机器学习可以使用更多数量的分片: $100 \times 100 = 10,000$
- 可以结合 geo 数据 和 非geo 数据
 - 在地图上的每个分片内每种颜色多少个点



SQL中支持Shape查询

7.5

在SQL中可以使用你自己定义的坐标系

```
POST ogc/_bulk
{"index":{"_id": "101"}}
{"ogc_type":"lakes", "fid": 101, "name": "BLUE LAKE", "shore": "POLYGON ((52 18, 66 23, 73 9, 48 6, 52 18), (59 18, 67 18, 67 13, 59 13, 59 18))"}
{"index":{"_id": "102"}}
{"ogc_type":"road_segments", "fid": 102, "name": "Route 5", "num_lanes": 2, "centerline": "LINESTRING (0 18, 10 21, 16 23, 28 26, 44 31)"}
{"index":{"_id": "103"}}
```



```
POST _sql?format=txt
{
  "query": "SELECT address, position, footprint from ogc where ogc_type='buildings'"
}
```

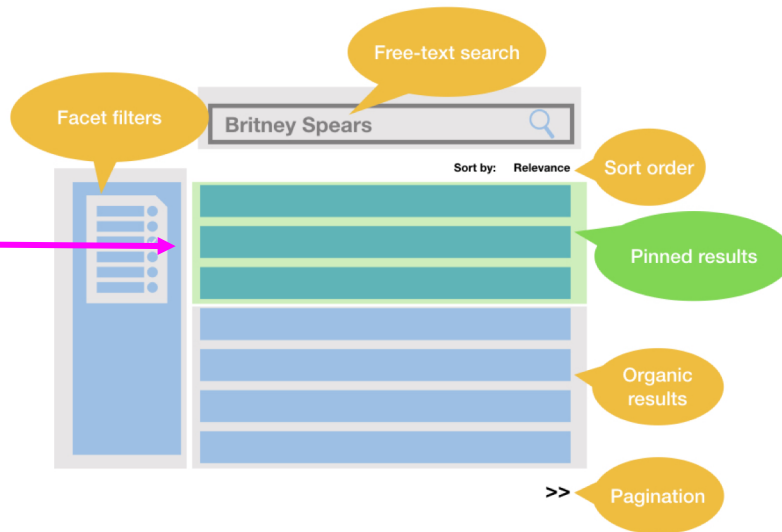


1	address	IST_X(position)	IST_Y(position)
2	-----+-----+-----		
3	123 Main Street	152.0	130.0
4	215 Main Street	164.0	133.0
5			

Results Pinning

7.4

```
GET /_search
{
  "query": {
    "pinned" : {
      "ids" : ["1", "4", "100"],
      "organic" : {
        "match":{
          "description": "iphone"
        }
      }
    }
  }
}
```



地理位置功能增强

7.4

新的Geo-Shape 数据类型

- Elasticsearch 7.4 以前用户需要使用地球坐标系来表达地理位置信息
- 新的 shape datatype 让用户可以使用他们自己的 x,y 坐标系统来存储和查询地理位置
- 这个特性对于非WGS84坐标系统, 室内定位, 虚拟世界 (游戏) 和其它应用非常有用

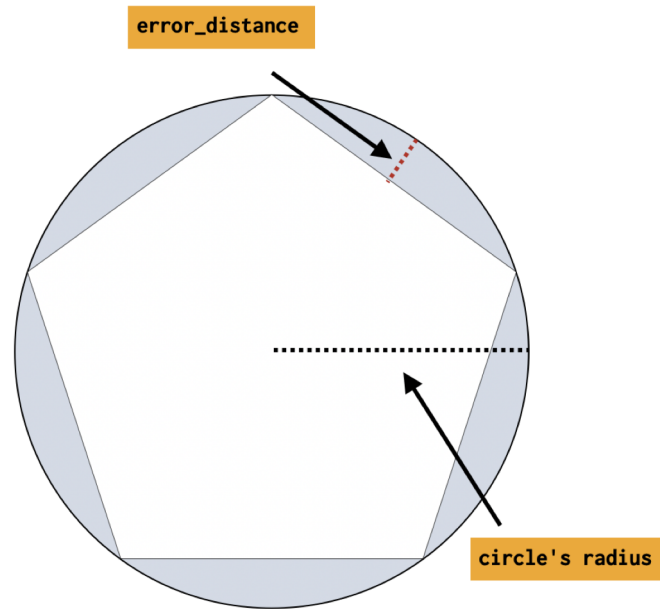


地理位置功能增强

7.4

Circle Ingest Processor

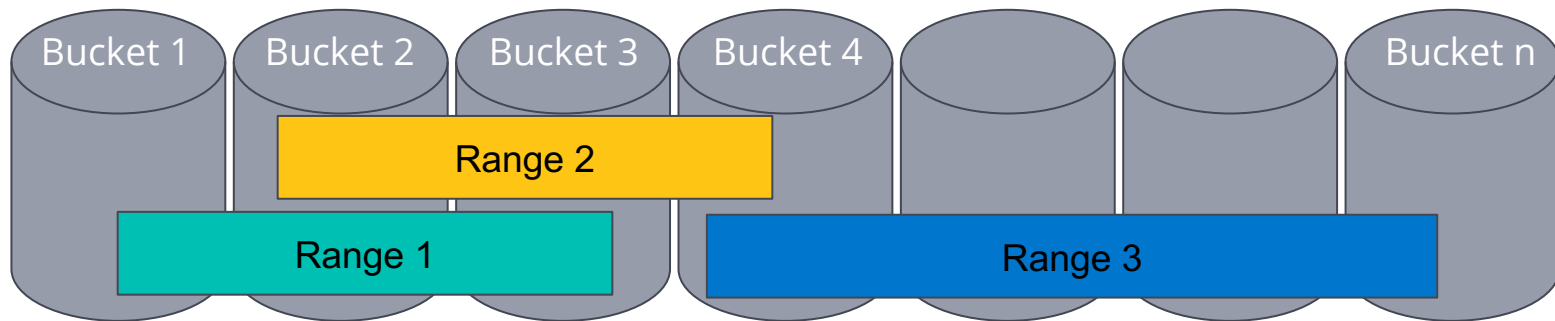
- 新的 circle ingest processor 将圆形表达成跟它近似的多边形
- 这是一个简单和高效的方法在BKD树里面索引圆形，有助于产生更小、更快、更精确的索引
- 可以通过圆形边缘到多边形的边缘的最大距离error_distance来控制拟合的精度



新的聚合

7.4

在Range类型字段上计算直方图



- Bucket 1 == 1
- Bucket 2 == 2
- Bucket 3 == 2
- Bucket 4 == 2
- Bucket 5..n == 1
- 在range字段上进行聚合计算使得统计有相互交错的范围数量更加简单
- 比如，每个电话有开始和结束时间，统计每天每分钟有多少电话进行中？

新的聚合

7.4

累积的基数聚合

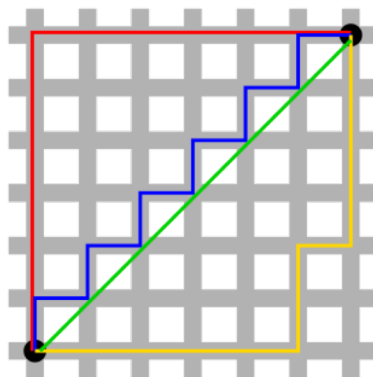
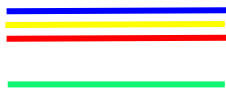
- 可以比较容易地计算一段时间内新增的事件
- 比如，这个新的聚合使用户可以计算给定一段时间独立用户访问数，而不需要更多的过滤和去重的步骤

新的向量距离函数

7.4

通过向量增强相关性排序

- 在很多场景下向量相似度被用来衡量文档的相似性
- 可以是向量文档或者一些算法可以将图片和文字表达为向量
- Elasticsearch 7.4, 我们引入了两种新的向量相似度度量:
 - Manhattan 距离
 - Euclidean 距离
- 提供预定义的Painless functions, 他们可以在script_score查询中和其它查询元素一起使用



Flattened field

7.3

将多个字段索引在一个扁平化的JSON里面 – 减少 mapping

很多任意类型的字段, 比如:

- Beats
- HTTP headers
- 电商平台(多个数据源)

Keywords 搜索 (不需要全文搜索)

单一的扁平 JSON 字段代替多个字段

减少了mapping和索引状态的大小

```
container.labels.annotation_io_kubernetes_container_ports
container.labels.annotation_io_kubernetes_container_restartCount
container.labels.annotation_io_kubernetes_container_terminationMessagePath
container.labels.annotation_io_kubernetes_container_terminationMessagePolicy
container.labels.annotation_io_kubernetes_pod_terminationGracePeriod
container.labels.io_confluent_docker
container.labels.io_confluent_docker_build_number
container.labels.io_confluent_docker_git_id
container.labels.io_kubernetes_container_logpath
container.labels.io_kubernetes_container_name
container.labels.io_kubernetes_docker_type
container.labels.io_kubernetes_pod_name
container.labels.io_kubernetes_pod_namespace
container.labels.io_kubernetes_pod_uid
container.labels.io_kubernetes_sandbox_id
```

稀有 terms 聚合

7.3

精确地找出最稀有的情况

对SIEM应用非常重要, 因为 稀有==可疑

Visible in Kibana (sort ascending)

我们克服了很多工程上的挑战



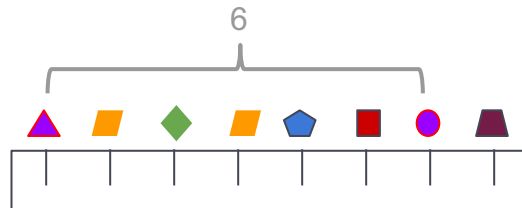
在Interval 查询中使用通配符

7.3

两个 terms间的距离 – 专利和法律 搜索

Interval 查询是一种新的span:

- 简单 == 更加经济
- 不需要重复实现analyzer逻辑



现在通过 增加 * 和 ? 通配符

Biolog* captures Biology and Biological

Optimi?ation captures Optimization and Optimisation

鼓励用户使用Interval查询来代替Span查询

distance_feature 查询

7.2

很多时候, 时间最近的数据或者地理位置较近的数据相关性更好

distance_feature 让你可以基于时间和距离提升排名

简单易用, 内部实现还是复杂的:

- 使用 top-k 快速查询
- 支持归一化 w/ saturation

```
GET items/_search
{
  "query": {
    "bool": {
      "must": {
        "match": {
          "name": "chocolate"
        }
      },
      "should": {
        "distance_feature": {
          "field": "production_date",
          "pivot": "7d",
          "origin": "now"
        }
      }
    }
  }
}
```

Data Frame Transform

7.2

将一个既有的index转换聚合成另外一个index

比如:

- 汇总用户行为
- 创建实体索引

机器学习app中有一个新的向导配置页面

现在通过批处理作业来实现

Source index filebeat-nginx-2019.02.05 showing 5 of 40 fields

@timestamp ↑	agent.type	http.response.body.bytes	http.response.status_co...	source.ip
January 9th 2019, 07:27:50	filebeat	410	200	35.224.108.130
January 9th 2019, 07:28:09	filebeat	410	200	35.224.108.130
January 9th 2019, 07:28:50	filebeat	410	200	35.224.108.130
January 9th 2019, 07:28:53	filebeat	410	200	35.224.28.11
January 9th 2019, 07:29:28	filebeat	319	200	185.246.208.82

Rows per page: 5

Data frame pivot preview

source.ip ↑	http.response.body.bytes.avg
1.2.241.97	194
1.214.221.2	194
14.102.189.231	194
14.139.184.212	166.4
18.228.119.52	171

Rows per page: 5

SQL 增强

7.2

7.2新的SQL功能:

- 支持地理位置查询
- median absolute deviation 函数
- 支持CASE WHEN/ELSE/END

```
SELECT count(*) AS count,  
       CASE WHEN NVL(languages, 0) = 0 THEN 'zero'  
            WHEN languages = 1 THEN 'one'  
            WHEN languages = 2 THEN 'bilingual'  
            WHEN languages = 3 THEN 'trilingual'  
            ELSE 'multilingual'  
       END as lang_skills  
FROM employees  
GROUP BY lang_skills  
ORDER BY lang_skills;
```

快速的 Top Hits 查询

7.0

基于新的Block-MAX WAND 算法

对于应用全文搜索场景的相关性排序场景
有很大的好处

不是在所有场景都起作用:

- 不能使用在聚合上
- 无法在Kibana里使用
- 无法用在需要精确hit count的场景下



Elasticsearch 新功能

查询&分析

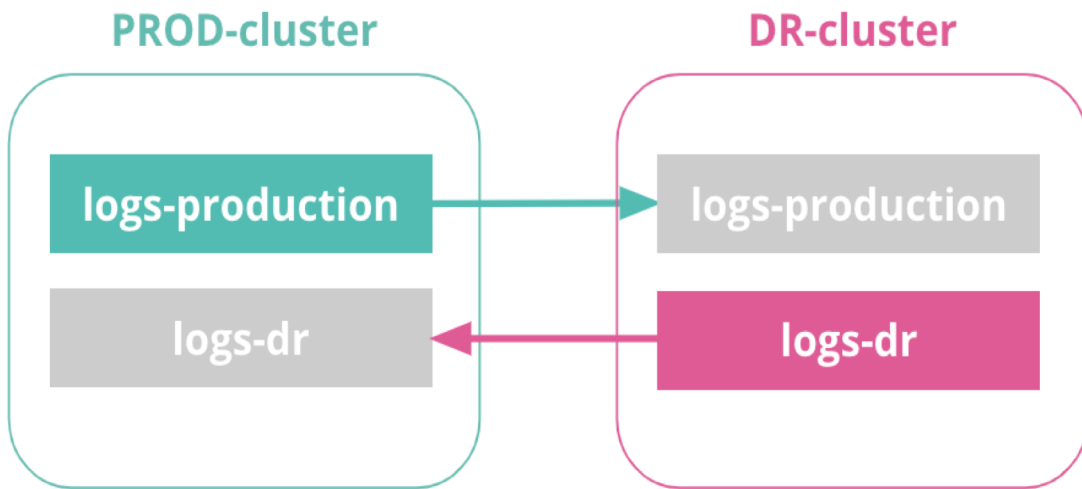
运维&架构

Cross-cluster Replication (CCR) 增强

7.5

新的更新改进了双向复制的CCR架构

- CCR 双向复制
 - 非常流行的 CCR 架构
 - 避免手动故障恢复
- Pause / Resume API



自动取消查询

7.4

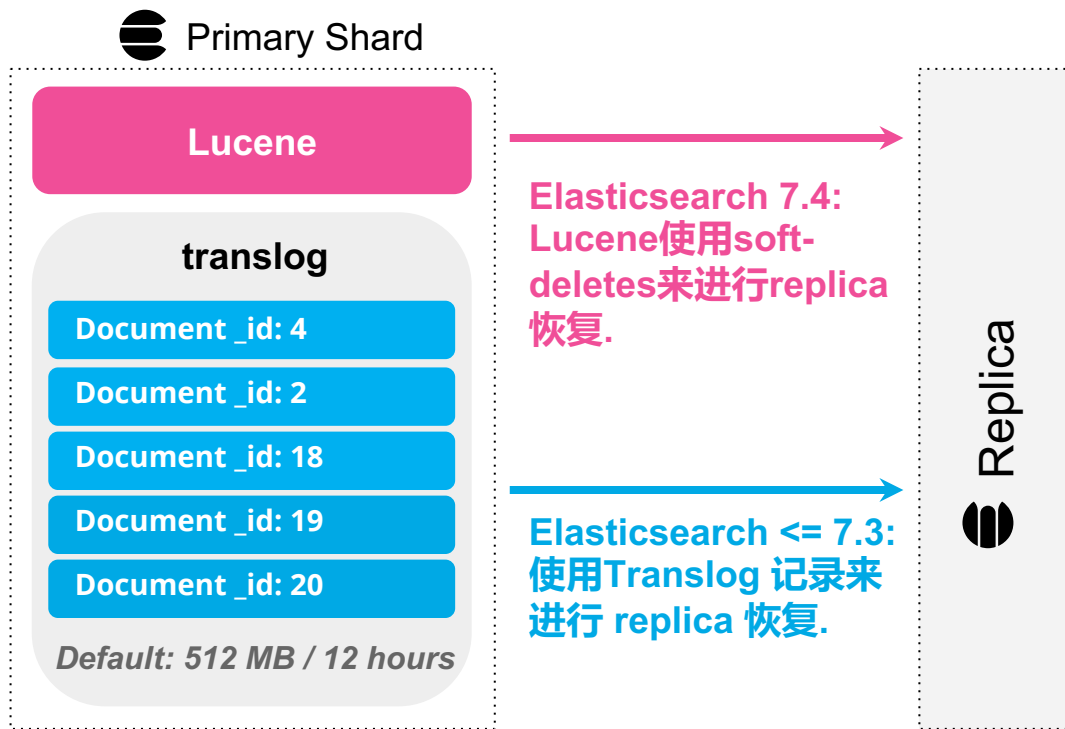
避免在不需要的查询上浪费资源

- 在 Elasticsearch 7.4以前, 如果用户想结束一个客户端发起的查询, 必须通过 Task management API来完成
- 在Elasticsearch 7.4, 在客户端连接关闭的时候可以自动取消通过_search 接口正在执行的查询
- 可以避免浪费资源在不需要的查询上, 比如:
 - 自动取消错误执行的长时间查询或者用户离开了某个页面
 - 自动取消某些运行错误的本地测试

更小的 Translog = 节省存储空间

7.4

- shard恢复已经不需要 Translog 记录
 - 过去:** 为了恢复需要每个 shard 要保存缺省 512MB 的 Translog
 - 现在:** Lucene 的 Soft-deletes 可以在 shard 恢复的时候被用来追踪文档的改变
- 对某些用户来讲可以节省大量存储空间



Snapshot UIs

7.2

新的snapshot UI具有的功能:

- 注册 snapshots 源 (支持多种插件)
- 查看 repositories and snapshots

后续会有更多的UI的改进

Repositories / Add repository

Register repository

Repository name
A unique name for the repository.

Name
my_backup

Repository type
Elasticsearch supports file system and read-only URL repositories. Additional types require plugins. [Learn more about plugins.](#)


Shared file system
[Learn more](#)
✓ Selected


Read-only URL
[Learn more](#)
Select


Azure
[Learn more](#)
Select


Google Cloud Storage
[Learn more](#)
Select


Hadoop HDFS
[Learn more](#)
Select


AWS S3
[Learn more](#)
Select

Source-only snapshots
Creates source-only snapshots that take up to 50% less space. [Learn more about source-only repositories.](#)

☒ Source-only snapshots

Next >

Snapshot & Restore UI

7.3

Snapshot 删除 和 恢复

- 从snapshot恢复索引
- 在恢复状态页追踪恢复的进度
- 删除不需要的snapshots

Snapshot and Restore

Use repositories to store and recover backups of your Elasticsearch indices and clusters.

[Snapshot and Restore docs](#)

Snapshots Repositories Restore Status

Refresh data every 30 seconds

Index	Status	Last activity	Shards completed	Shards in progress
restored_logstash_0	In progress	now	0	1

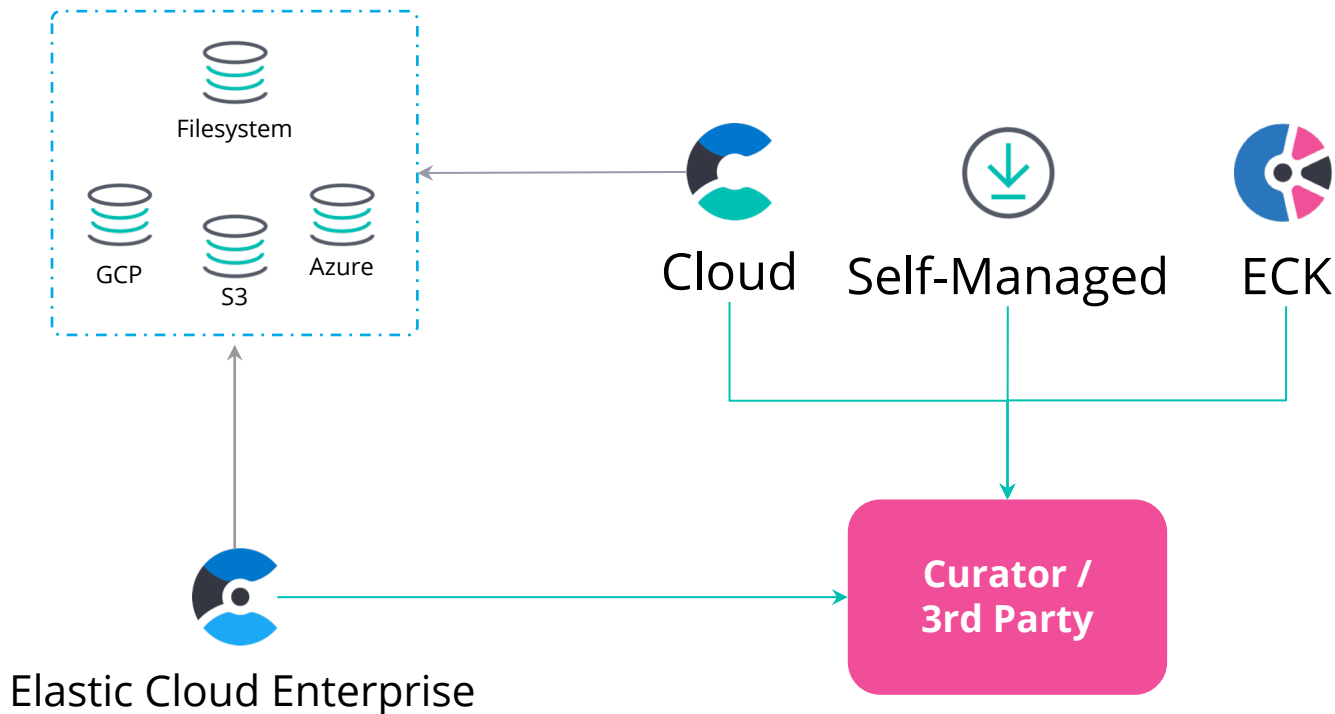
ID	Stage	Start time	End time	Duration	Repository	Snapshot	Version	Target host	Target node	Bytes	Files
0 P	INDEX	03 July 2019 14:10:44		6 seconds	my-backups	snapshot-3	8.0.0	127.0.0.1	Jens-MacBook-Pro.local	0.0%	4.3%
restored_logstash_1	In progress						0		1		
restored_sample_data_logs	Complete						1		0		
restored_sample_data_flights	Complete						1		0		
restored_sample_data_ecom...	Complete						1		0		

Rows per page: 20

Snapshot 生命周期管理(过去)

7.4

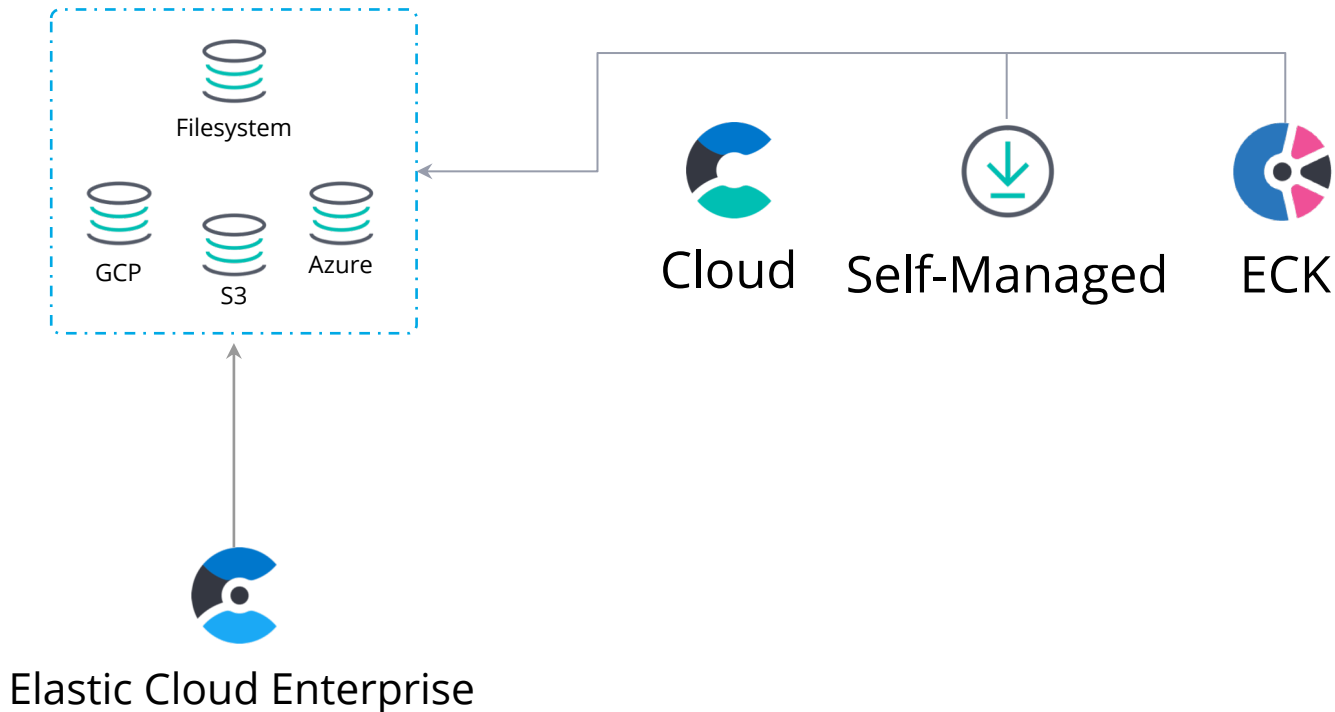
过去处理snapshot的方式



Snapshot 生命周期管理(现在)

7.4

不再需要Curator和第三方工具



Snapshot 生命周期管理留存规则设置

7.5

新的API

Create policy

✓

✓

3

4

LogisticsSnapshot settingsSnapshot retentionReview

Snapshot retention (optional)

[Snapshot retention docs](#)

Expiration

The time to wait before deleting snapshots.

Delete after

30

days

Snapshots to retain

The minimum and maximum number of snapshots to store in your cluster.

Minimum count

5

Maximum count

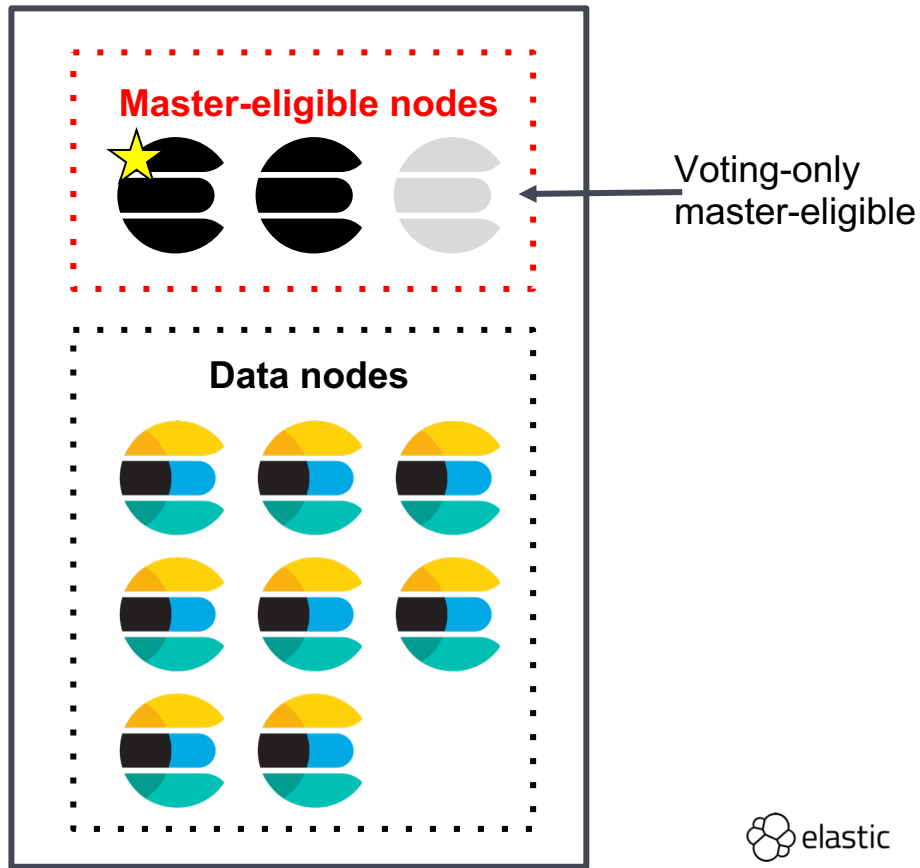
20

[< Back](#)[Next >](#)[Cancel](#)

Voting-only Master eligible节点

7.3

- 使用 ES 7.0 新的集群协调 (快速 master 选举, 不需要minimum_master_nodes, ...)
 - 开启 voting-only master-eligible 节点部署
- 在master选举的时候仅仅投票
 - 不会成为master
- 需要更小的硬件配置
 - 网络
 - 可用性
- 在部署大量集群的时候可以节省成本



复制关闭的索引

7.2

关闭的索引:

- 除了维护元数据以外几乎不占资源
- 无法进行读写操作

尽可能使用索引冻结

- 可以像关闭的索引一样最小化资源利用，并且可以被搜索

Closed Index

Frozen Index

Replicated

(≥ 7.2)

**Minimal
resource
utilization**

Searchable

新的集群协调机制

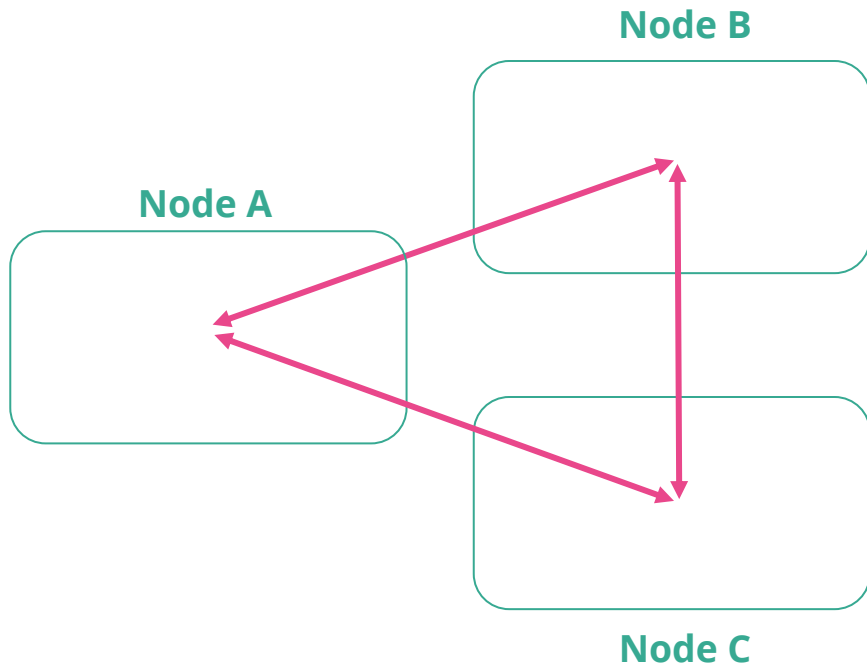
7.0

重建了集群协调层代码，为未来更多功能打下基础

构建在强健的理论基础上，经过了大量密集测试

带来的收益

不再需要 `minimum_master_nodes` 设置
秒级以下的master选举时间
快速的移除僵尸节点



Cross-Cluster Search (CCS) 改进

7.0

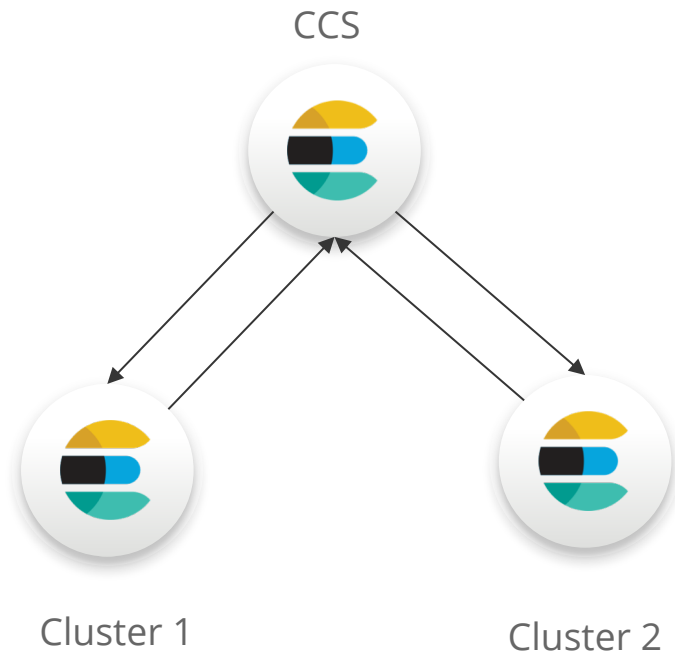
引入了新的执行模式
(`ccs_minimize_roundtrips`) 对广域网环境
下的跨集群查询来优化

以前:

每一个远端集群的每个节点都会有数据返回，
导致大量的小的数据请求在网际传播

现在:

每个远程集群只会从协调节点返回一个数据
响应，极大减少了在网络中的数据往返开销



Kibana 新功能

全新的外观和体验优化

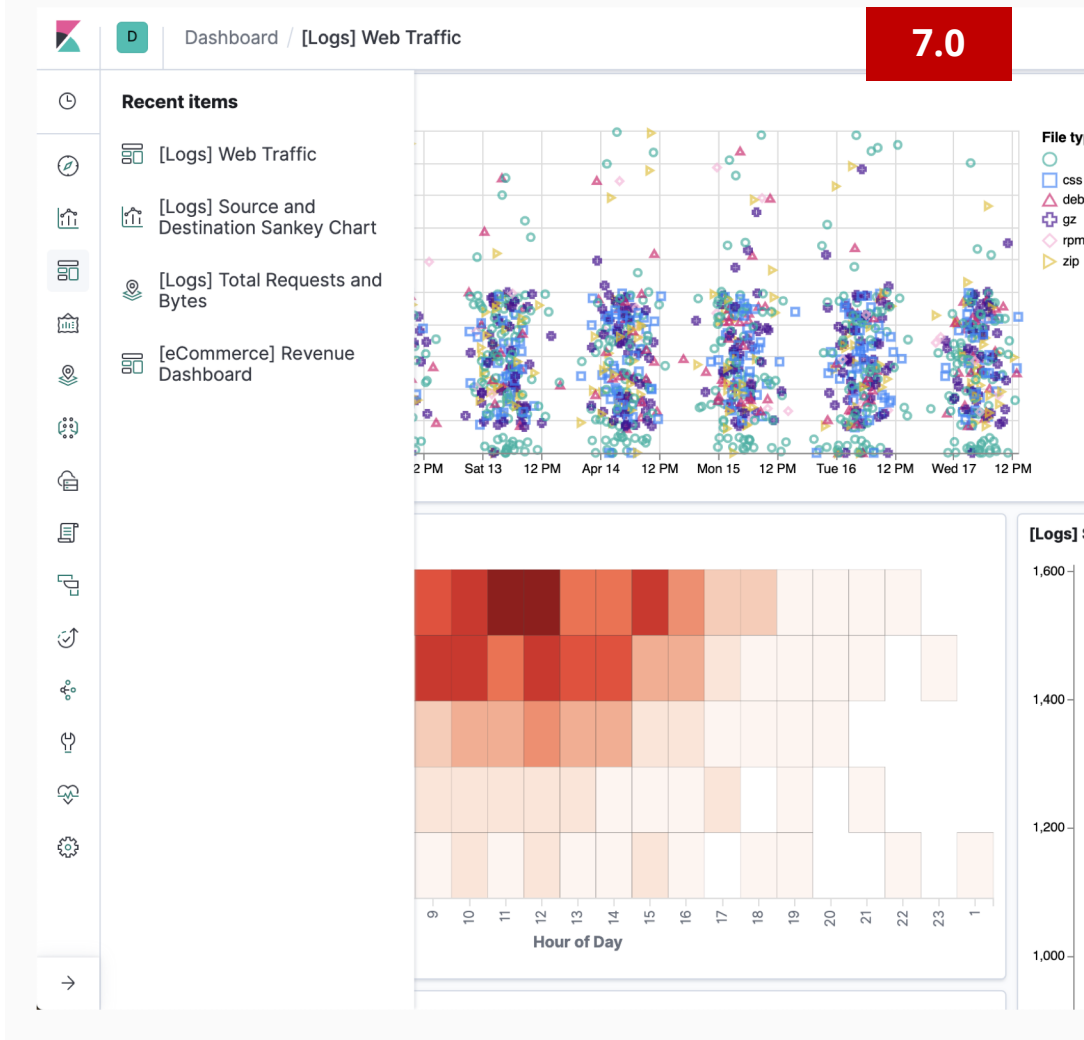
新的全局导航菜单

全面的深色界面模式

响应式的dashboard

增强的时间选择器

缺省使用Kibana Query Language KQL



全新的外观和体验优化

新的全局导航菜单

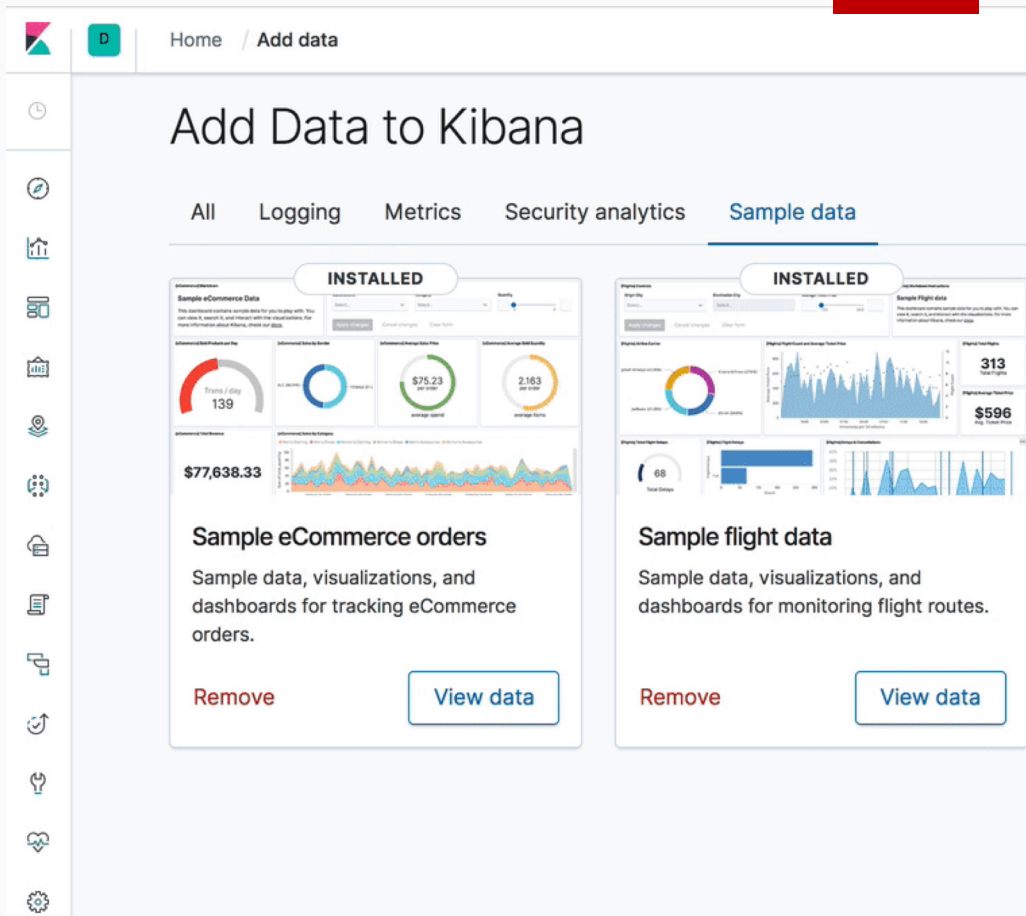
全面的深色界面模式

响应式的dashboard

增强的时间选择器

缺省使用Kibana Query Language KQL

7.0



全新的外观和体验优化

7.0

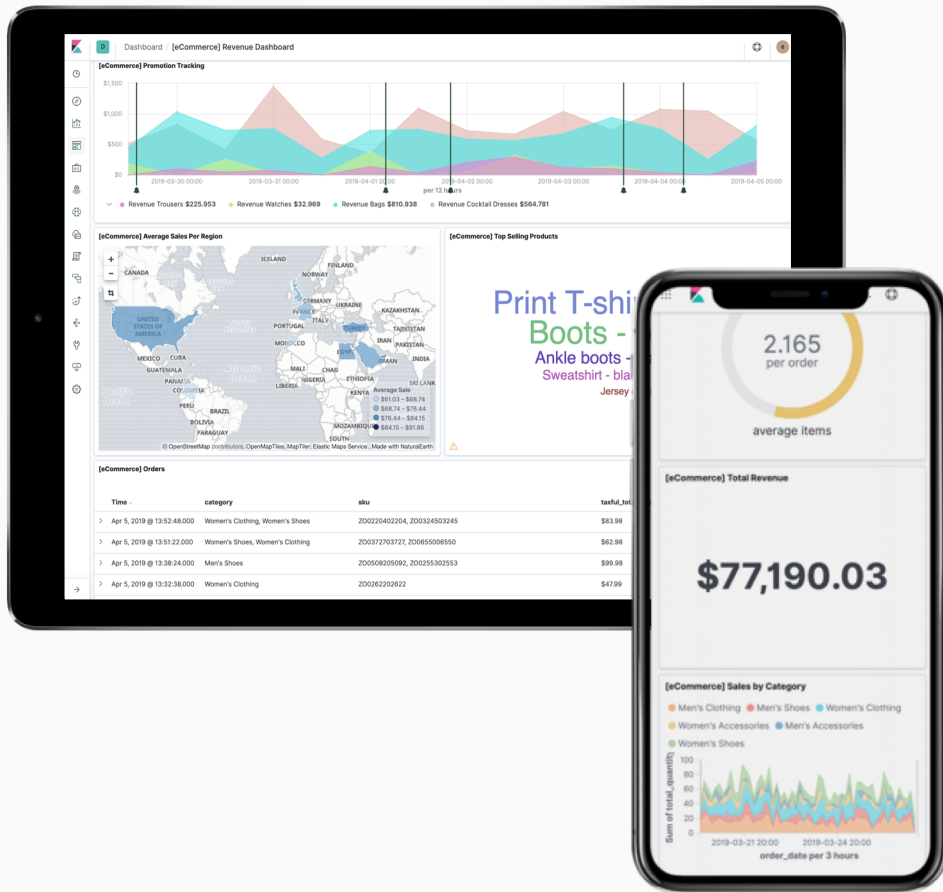
新的全局导航菜单

全面的深色界面模式

响应式的dashboard

增强的时间选择器

缺省使用Kibana Query Language KQL



全新的外观和体验优化

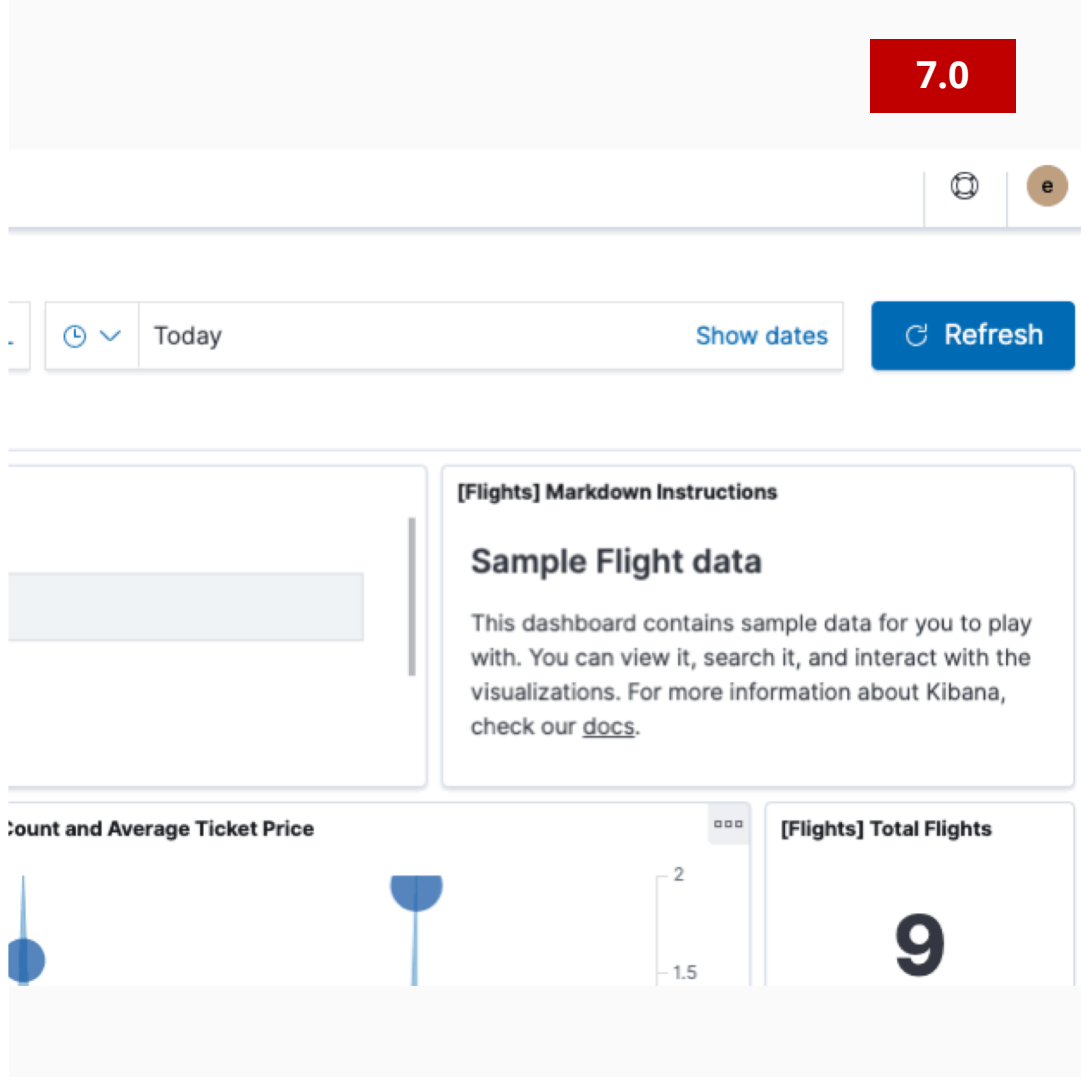
新的全局导航菜单

全面的深色界面模式

响应式的dashboard

增强的时间选择器

缺省使用Kibana Query Language KQL



全新的外观和体验优化

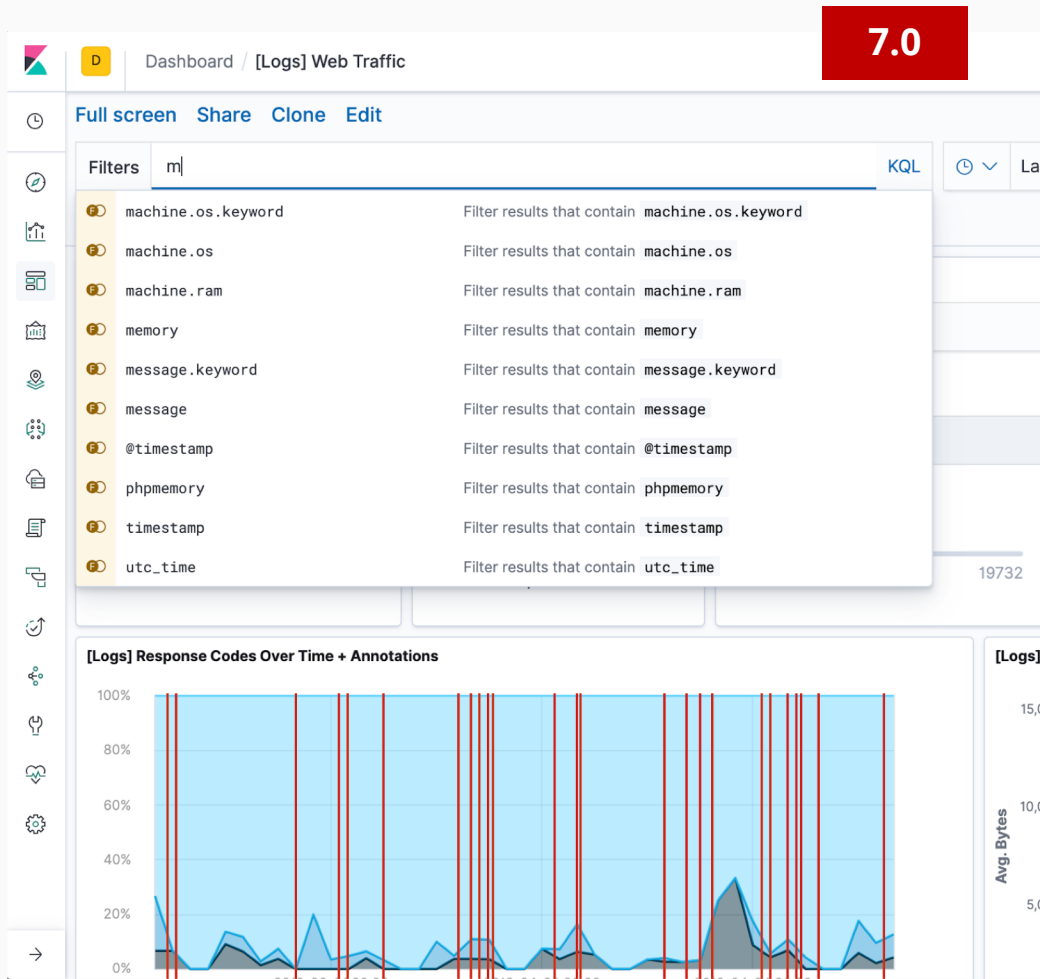
新的全局导航菜单

全面的深色界面模式

响应式的dashboard

增强的时间选择器

缺省使用Kibana Query Language KQL



Kibana Lens

7.5

Search KQL Last 7 days

+ Add filter

kibana_sample_data_logs

Search field names

Filter by type 0

Records

Individual fields

- agent.keyword
- bytes
- clientip
- event.dataset
- extension.keyword
- geo.dest
- geo.src
- geo.srcdest
- host.keyword
- index.keyword
- ip
- machine.os.keyword

Drop some fields here to start



Lens is a new tool for creating visualizations **BETA**

[Make requests and give feedback](#)

数据驱动

一切从数据开始

寻找合适的字段:

- 可以按名字搜索
- 按照类型过滤
- 快速的字段值预览
- 字段统计

filebeat-*

Search field names

Filter by type

0

event.severity

event.start

fileset.name.keyword

1

host.architecture.keyword

host.containerized.keyword

host.hostname.keyword

host.id.keyword

host.name.keyword

host.os.codename.keyword

host.os.family.keyword

host.os.kernel.keyword

host.os.name.keyword

host.os.platform.keyword

host.os.version.keyword

TOP VALUES

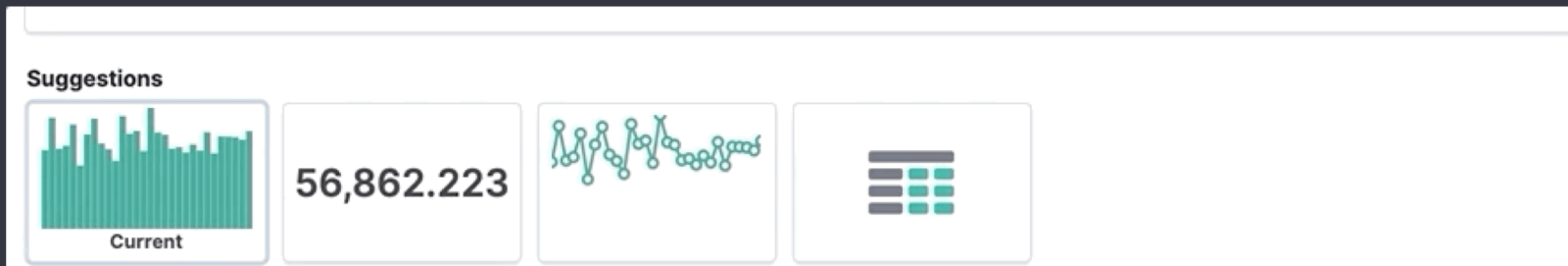
notice	17%
http	16%
auth	16%
ssl	14%
eve	13%
dns	12%
connection	12%

100% of 179 documents

更加智能的方式可视化

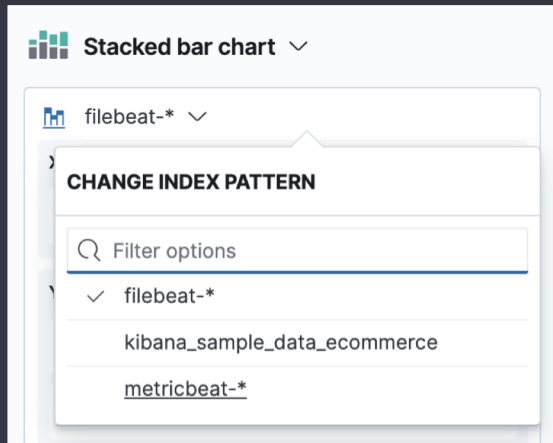
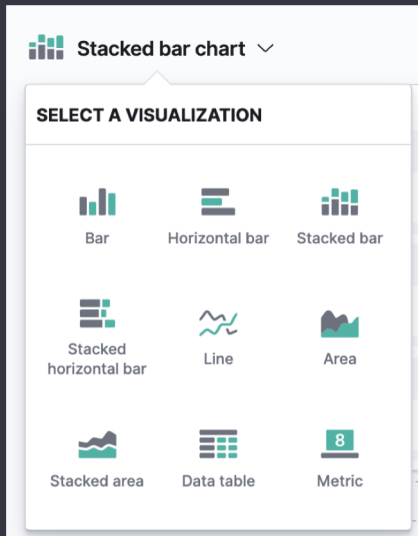
智能建议图表

- 根据选中的字段来建议合适的聚合
- 根据数据来推荐图表类型
- 数据建模
- 快捷方式-基于最常用的用例来智能的提供缺省值



标准的并且可复用

- 运行用户以不同的方式浏览数据
- 可以修改index patterns 和 queries
- 可以用不同的图表类型查看数据
- 在字段间快速切换



Kibana 管理: 在不同的Space间拷贝保存的对象

使得跨Space管理对象更简单

- 可以将对象从一个Space拷贝到多个目标Spaces
 - 以前必须要导出/导入才能跨Space拷贝对象
- 拷贝的对象是克隆了一份新的对象而不是链接
- 如果拷贝的是嵌套对象，所有嵌套的对象也会被拷贝
- 需要对目的Space具有写权限



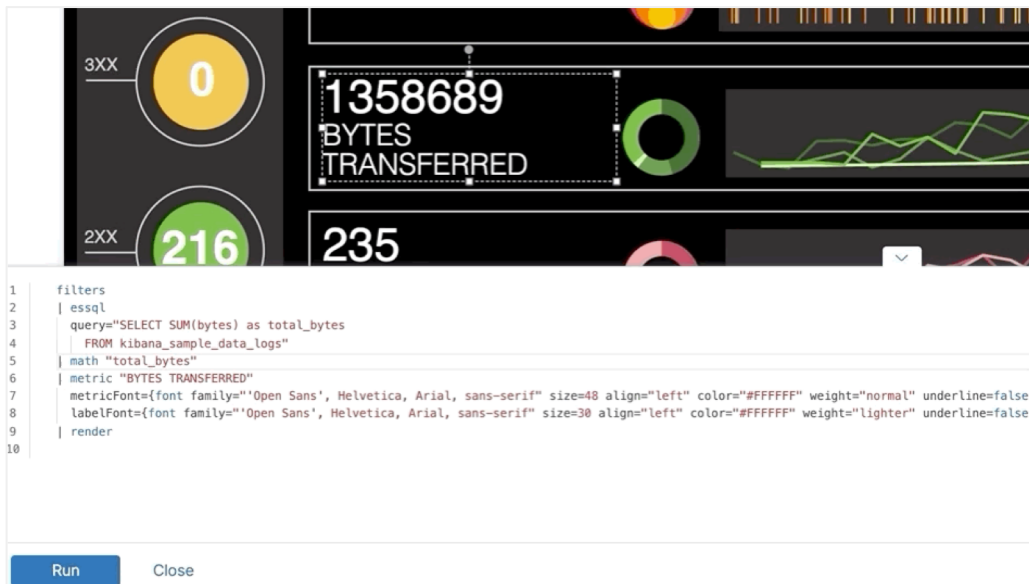
Copy to Space

Kibana 分析: Canvas

7.4

使用 Monaco 编辑器改进Canvas表达式编辑器体验

- darkmode
- 最小化视图
- 语法高亮
- 鼠标悬停提供上下文帮助
- 新的全屏模式
- 改进文档体验

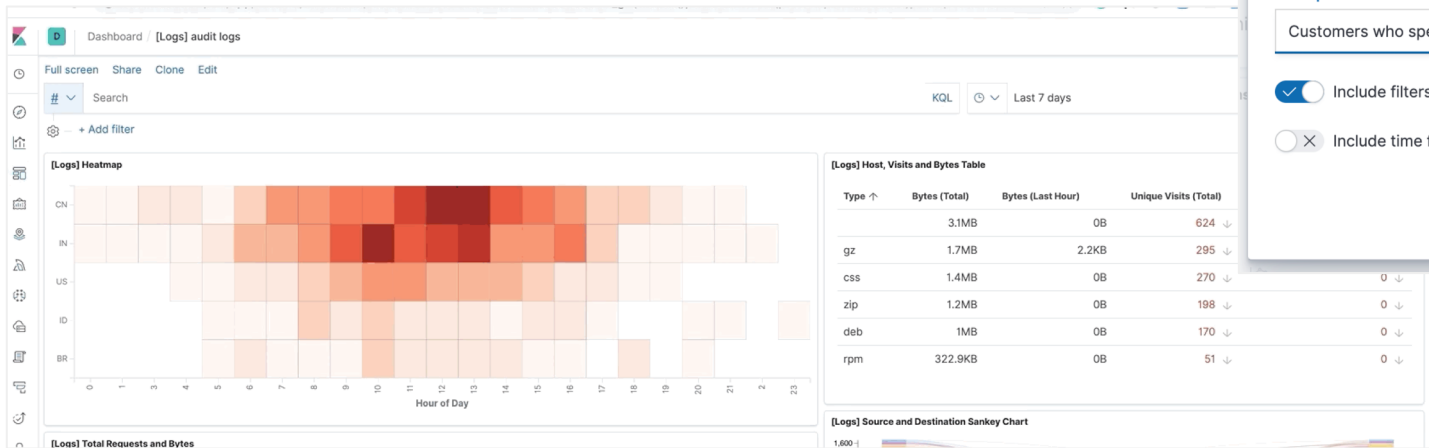


Kibana 分析: 保存 查询

7.4

保存, 应用, 分享 和复用Kibana 查询

- 结合了过滤器, 搜索框, 时间选择器
- 在Kibana dashboard, 可视化, 和日志浏览器间复用查询
- 分享和协作: 保证每个人在分析时候都使用一样的查询条件



The 'Save query' dialog box is shown, allowing users to save their current query and filters. It includes fields for 'Name' and 'Description', and checkboxes for 'Include filters' and 'Include time filter'.

Save query

Save query text and filters that you want to use again.

Name

Strategic accounts in Europe

Name is required. Name cannot contain leading or trailing whitespace. Name must be unique.

Description

Customers who spend 200K+ with dedicated SA

☒ Include filters

☐ Include time filter

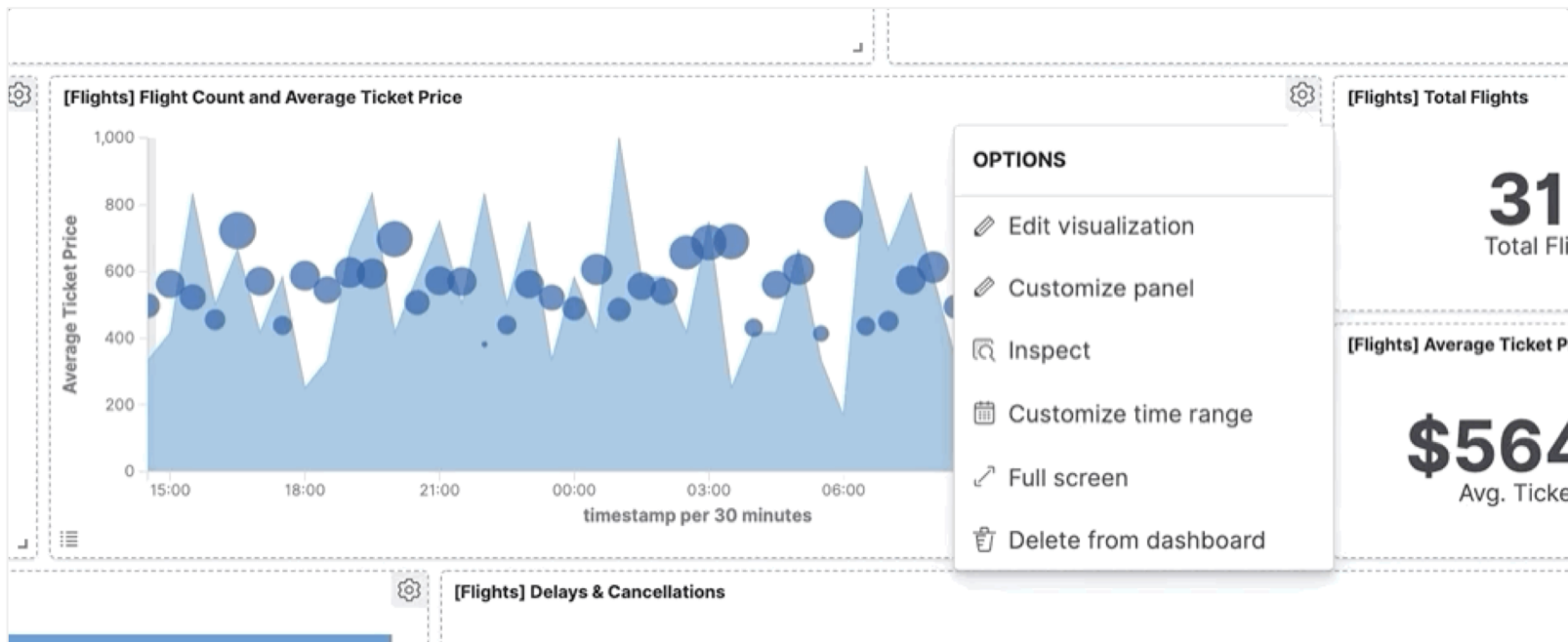
Cancel Save

Kibana 分析: 每个面板独立的时间范围

7.4

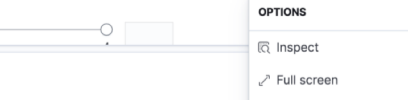
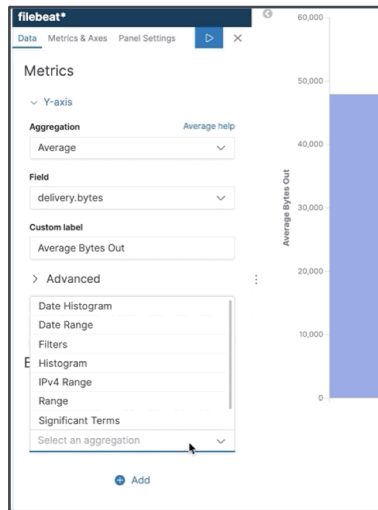
在一个dashboard上查看和比较多个时间段的图表

- 可以在一个dashboard上显示不同的时间区间



一些小的改进

-
- Stacked bar chart showing the number of people in the 'In the car' category across different times of day on 2019-07-25. The chart shows a peak of 18 people at 09:00. A red box in the top right corner displays the total count of 7.3.
- | Time | Category | Count |
|------------------|-------------|-------|
| 2019-07-24 21:00 | Light Blue | 9 |
| | Dark Blue | 1 |
| | Light Green | 1 |
| 2019-07-25 03:00 | Light Blue | 6 |
| | Dark Blue | 2 |
| | Light Green | 1 |
| 2019-07-25 09:00 | Light Blue | 7 |
| | Dark Blue | 3 |
| | Light Green | 1 |
| 2019-07-25 15:00 | Light Blue | 8 |
| | Dark Blue | 2 |
| | Light Green | 1 |



The screenshot shows a web application with a table. A context menu is open over the first row of the table. The menu has a title 'OPTIONS' and three items: 'Inspect', 'Full screen', and 'Download CSV'. The table has columns for a unique identifier, a price, and a count. The first row has the value 'ZO0596805968', a price of '\$26.98', and a count of '2'.

	taxful_t	
ZO0596805968	\$26.98	2
ZO0488604886	\$45.98	2
ZO0087200872	\$46.98	2
ZO0481204812, ZO0398403984, ZO0282402824	\$258.96	4
ZO0610206102, ZO0514405144, ZO0586505865	\$97.96	4

功能菜单控制

隐藏或者限制Kibana界面里面应用或者功能的访问

为每个角色设置功能访问列表

在特定的Space为每个角色设置功能访问列表

设置Space设定来隐藏或者显示某些功能或者App

权限等级:

- 不能访问
- 只读
- 全部

7.2

Customize feature display (18 / 19 features visible) [hide](#)

Control which features are visible in this space.

The feature is hidden in the UI, but is not disabled.

Want to secure access? Go to [Roles](#).

Feature	Show? (change all)
 Discover	☒
 Visualize	☒
 Dashboard	☒
 Dev Tools	☐ 
 Advanced Settings	☒
 Index Pattern Management	☒
 Saved Objects Management	☒
 Timelion	☒
 Graph	☒
 Stack Monitoring	☒
 Machine Learning	☒
 APM	☒

Canvas 增强

7.2

自定义组件

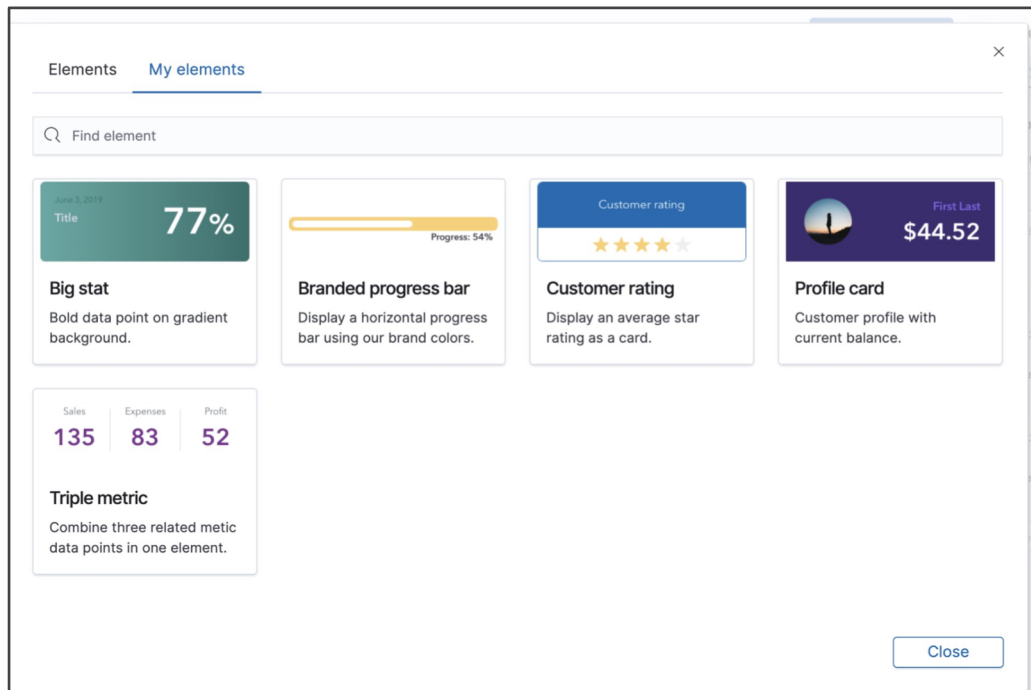
把组件保存为自定义组件，可以在不同的工作簿间复用，可以在团队内分享

过滤器组

将过滤器应用在一组元素上面

大屏模式

全屏自动播放



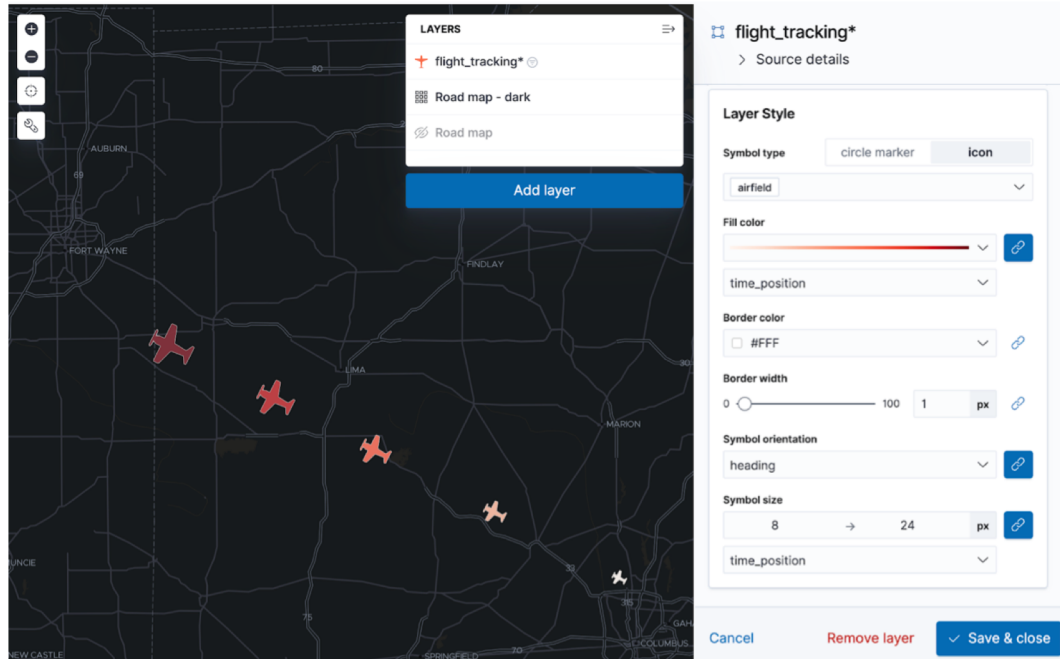
Kibana Map 新功能

更好的渲染时序数据

7.5

将图形的颜色和大小与时间字段联系起来

- 对于时间序列追踪非常有用
- 追踪最新的车辆，包裹，人的位置
- 以前只能根据数字字段来进行不同的渲染

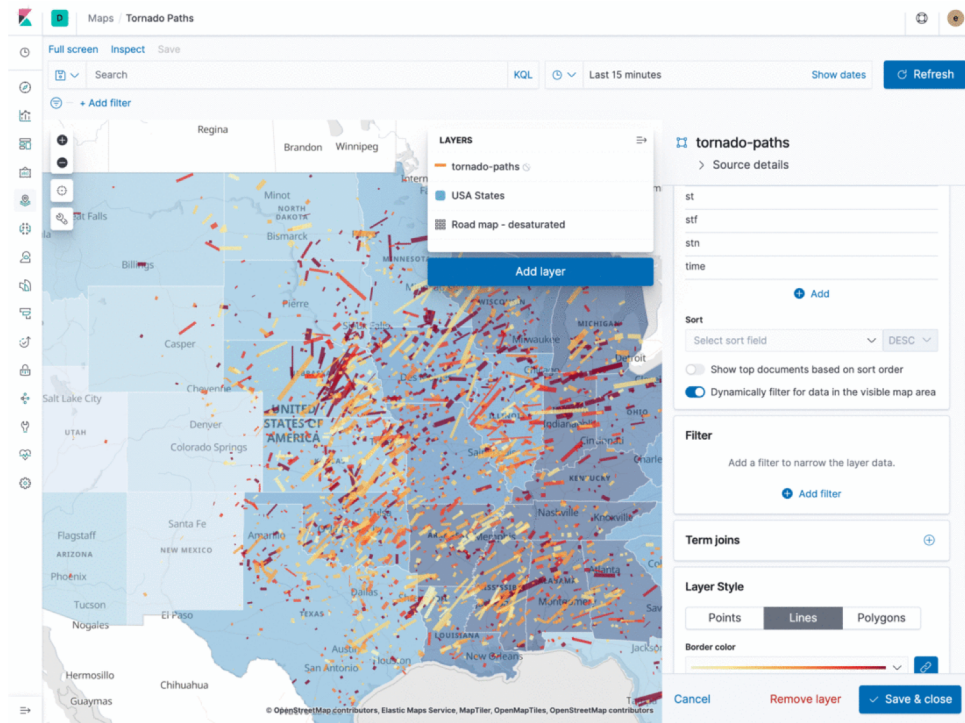


按不同的层来排序文档

7.5

确保最重要的数据在最上面

- 结合根据时间渲染的特性，给予时序数据的展示更多的控制
- 更加一致的数据可视化

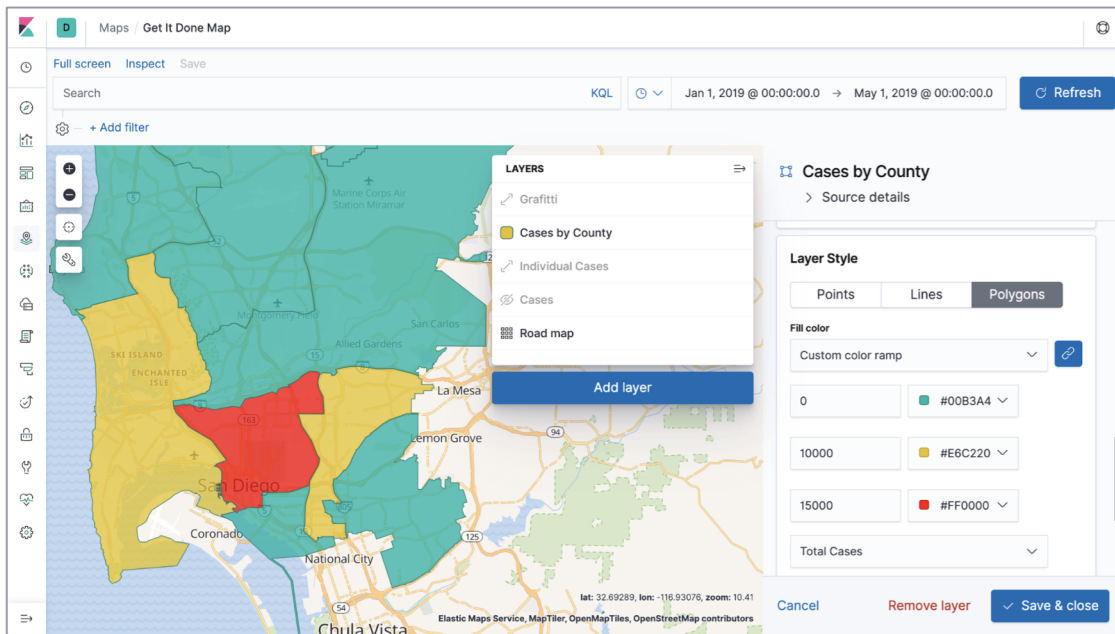


通过静态值设置自定义的色彩梯度

7.4

一个呼声很高的功能

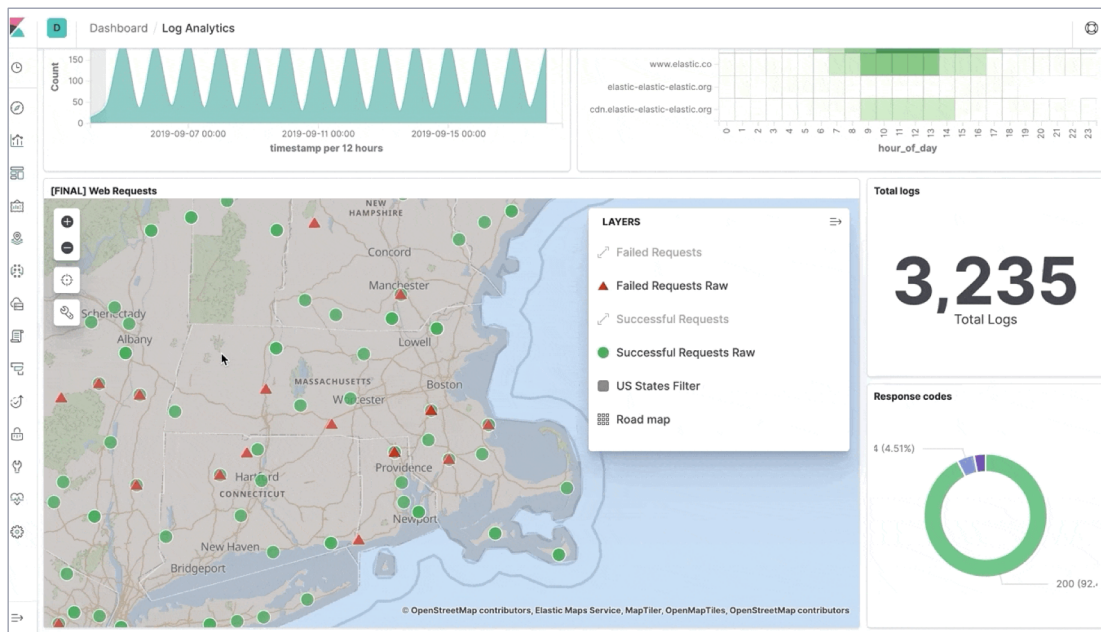
- 定义静态值和色彩来为不同的数值创建色彩梯度
- 可以应用在地图上的形状和点上
- 对于用户了解他们的数据并且有静态范围的场景非常有用



点击矢量形状进行过滤

7.4

- 在地图上创建可以用于过滤的地理区域
- 可以使用来自于Elastic Maps Service的形状或者GeoJSON的自定义形状
- 对于根据区域边界进行快速过滤非常有用

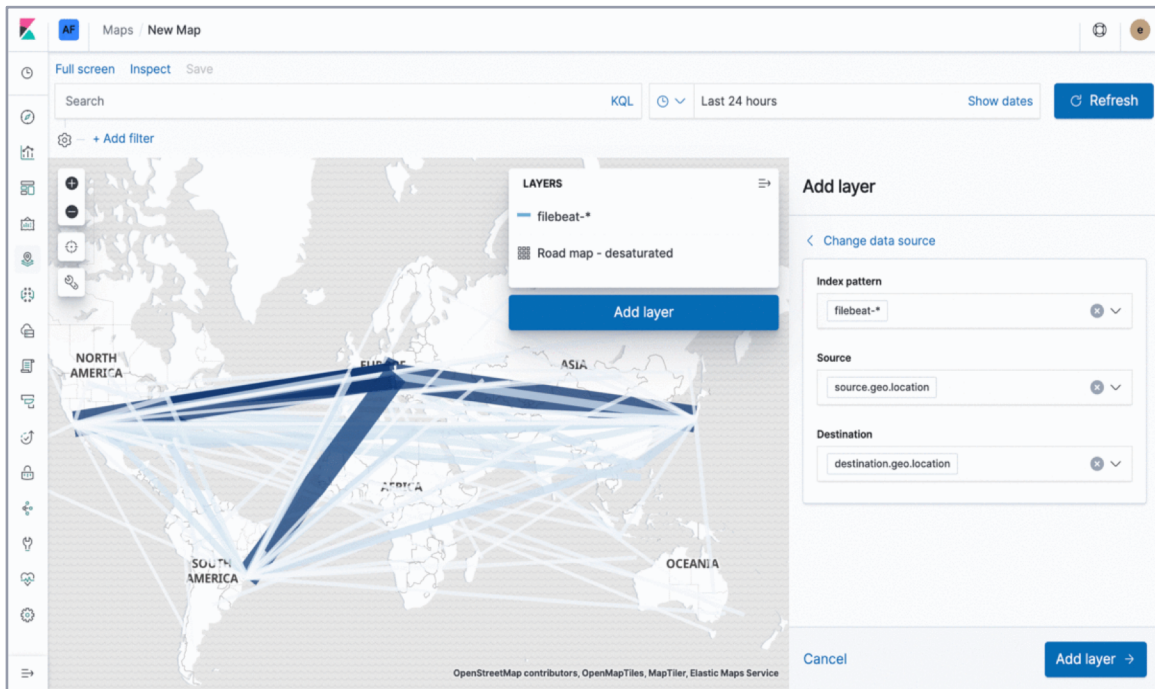


比较方便地查看两点间的路径

7.4

制作自定义的点对点地图

- 一个新的数据源来显示两点之间的路径
- 和其它形状一样的风格设定
- 被用作SIEM新的网络地图

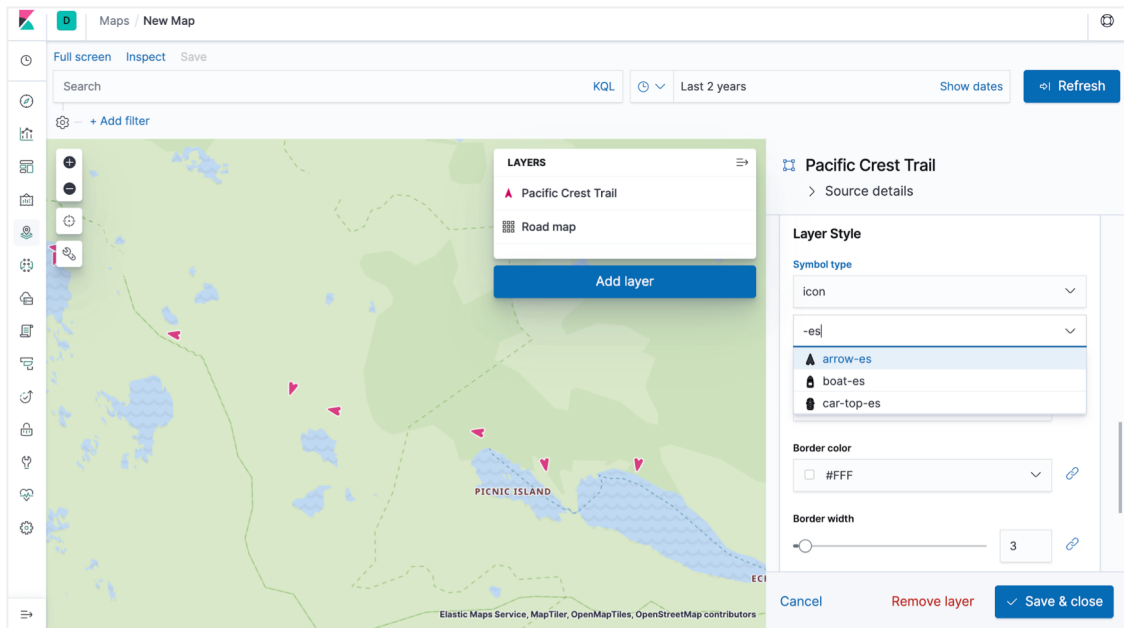


方向图标

7.4

箭头, 船, 汽车 更多图标

- 新的图标和动态的方向
- 可以自定义图标的边缘
- 对于追踪轨迹等应用场景非常有用

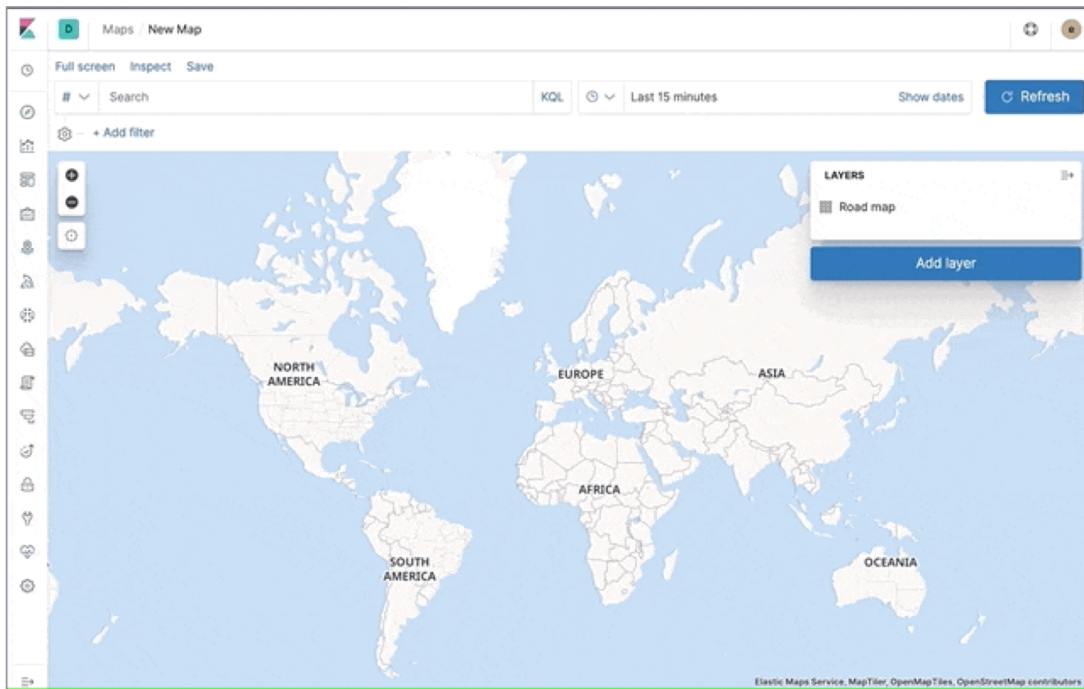


矢量分片

7.4

一个更平滑更现代的方式来创建基础图层

- 7.4 引入了矢量基础图层
- 使得浏览和缩放地图的时候更加平滑，锐利和轻量
- 大多数现代的地图应用都是使用矢量分片



Maps 增强

7.2

- 地图可以嵌入Dashboard
- 可以按层来过滤
- 在Elastic Maps 服务里面支持深色基础图层

