Elasticsearch 和 Kibana 在 Hulu视频的应用实践

倪顺 @ Hulu
2017.5.21, shun.ni@hulu.com

关于我

2016年加入Hulu，围绕提高用户播放视频体验，从事QoS（Quality of Service）数据相关工作。

主要内容包括数据集成、分析、可视化、自动监控等。



Hulu

- 高质量视频内容
- 丰富客户端支持
  - 网站
  - 手机
  - 电视盒子
  - Roku，Xbox



“吉祥物”



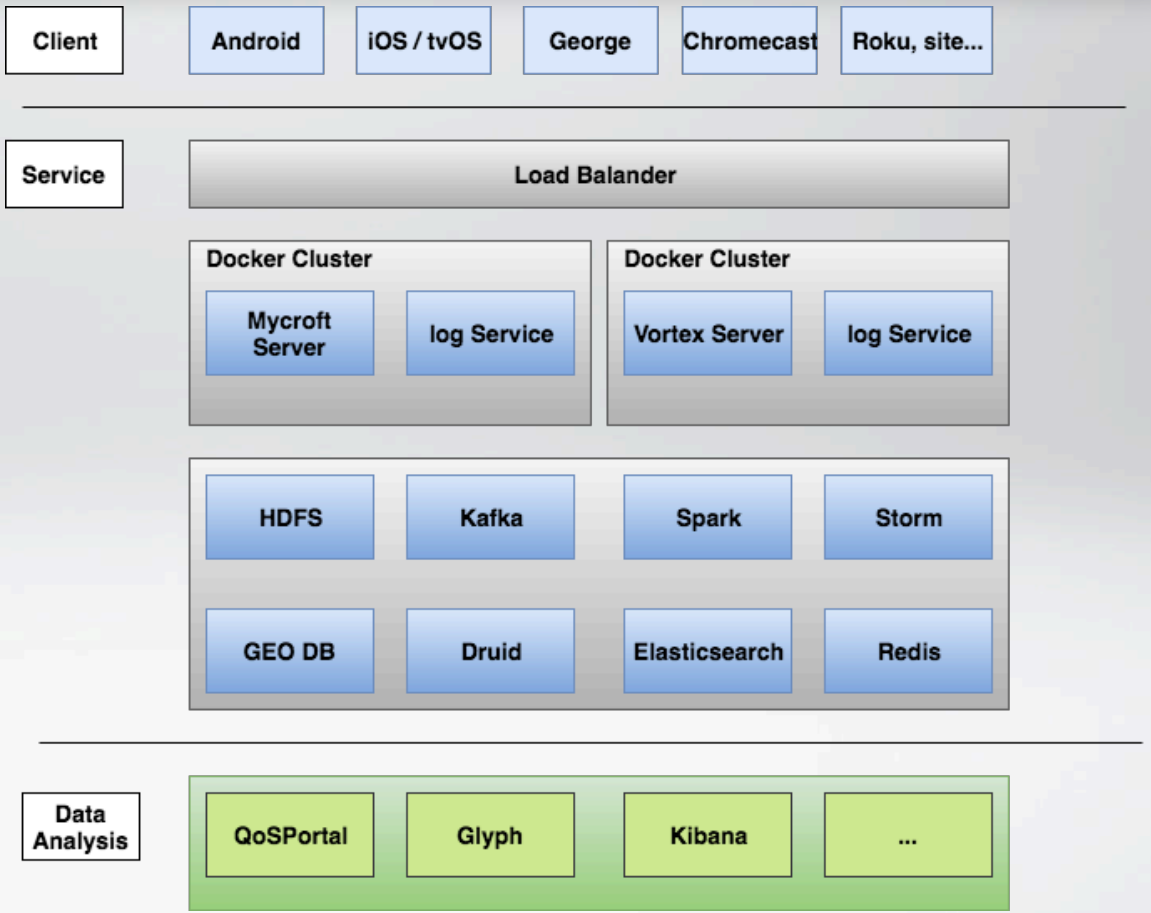
hulu

主要内容

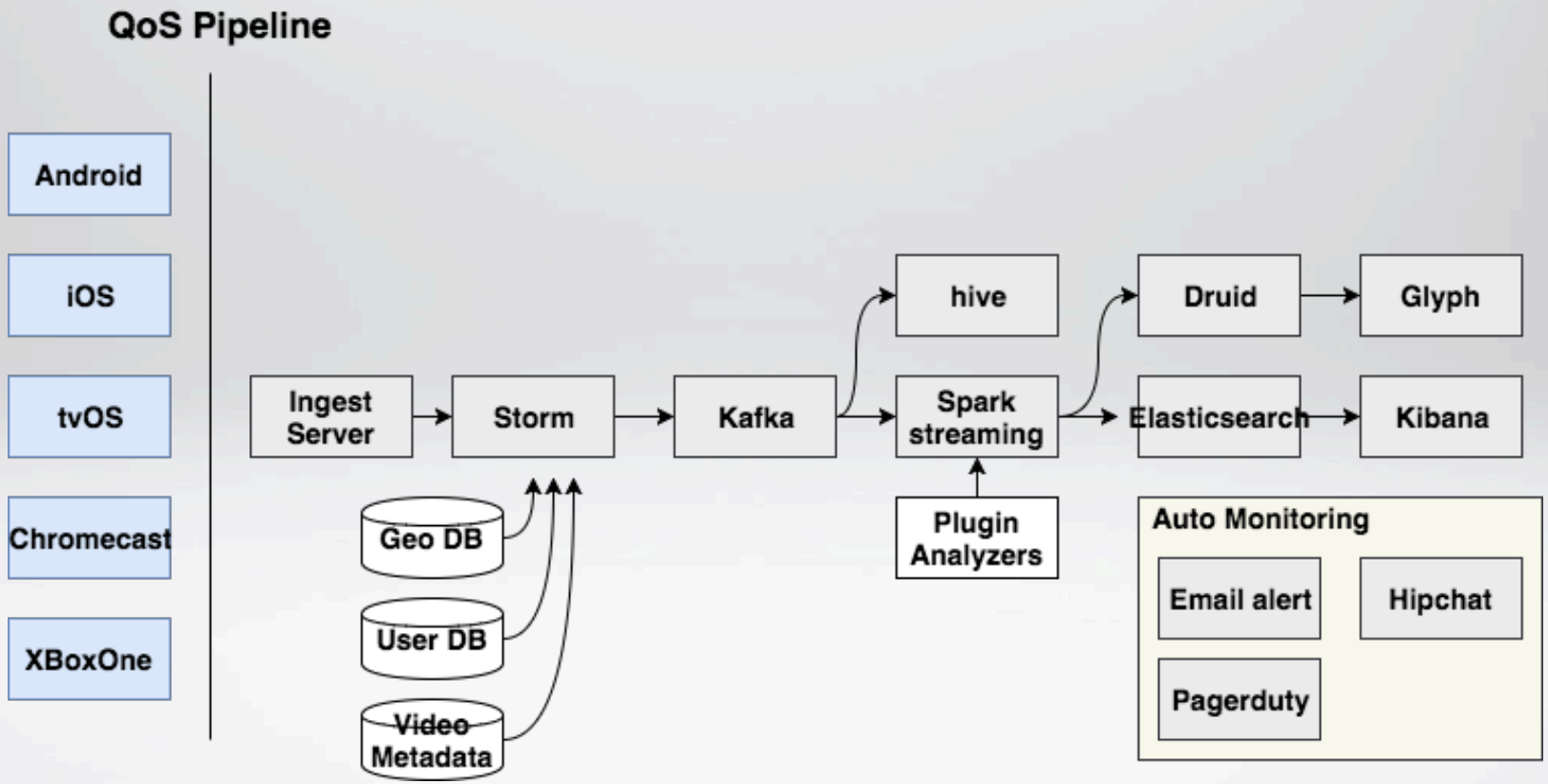
- 技术架构
- 应用实践
  - 视频播放质量报表
  - AB test 支持
  - Kibana可视化定制
- 经验心得
- Q & A

技术架构

- 用户播放数据收集
  - Start, End
  - Pause, Resume
  - Rebuffer
  - Seek
  - Bitrate
  - …
- 数据预处理
- 数据分析可视化



技术架构 – QoS Pipeline



主要内容

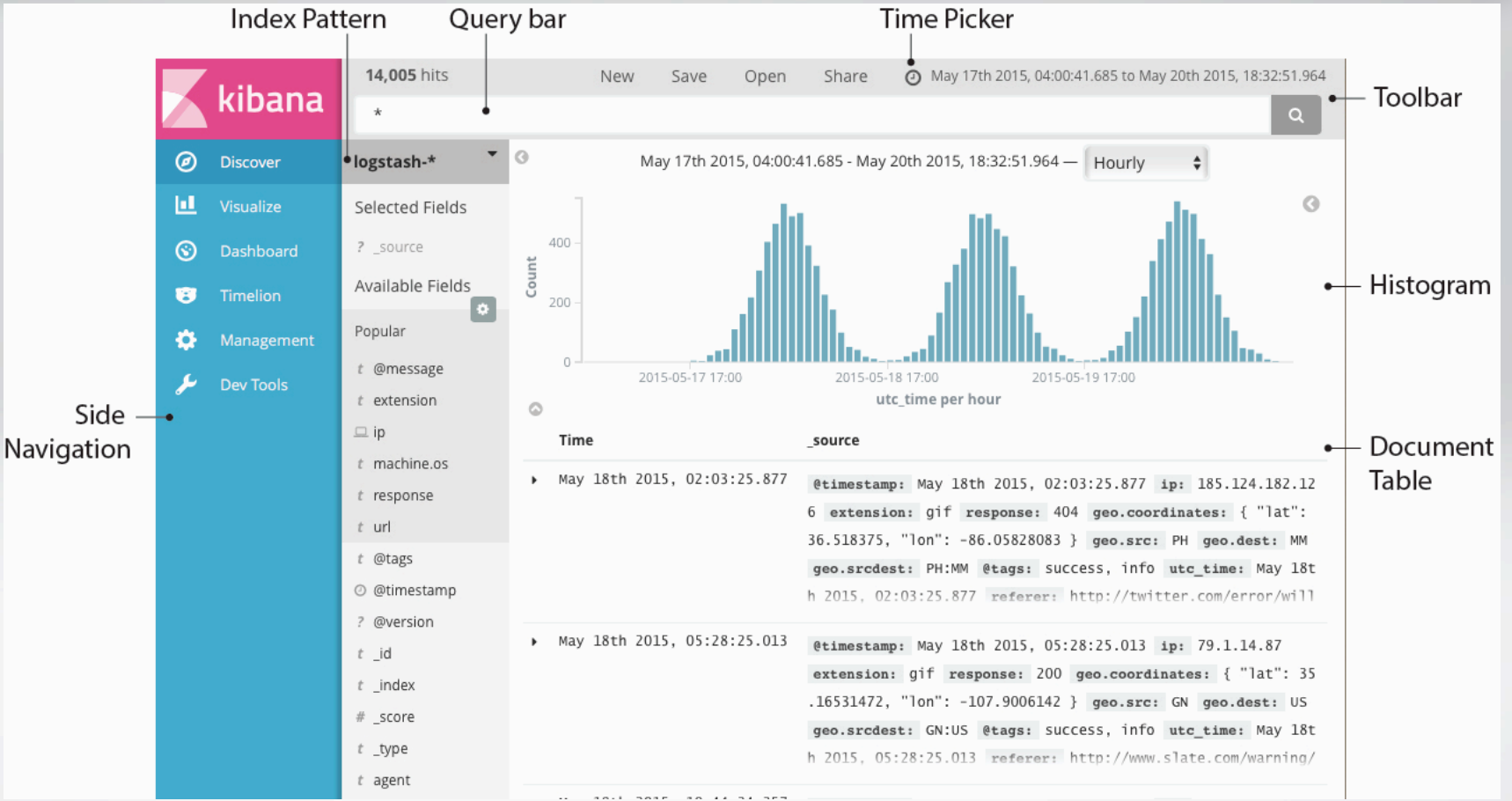
- 技术架构
- 应用实践
  - 视频播放质量报表
  - AB test 支持
  - Kibana可视化定制
- 经验心得
- Q & A

应用实践－视频播放质量报表

- Kibana basic features warm up
  - 支持明细查询
  - 支持 Visualization：Pie chart, Line chart, histogram, tile map, etc.
  - 支持 Dashboard：多个 visualization 的集合

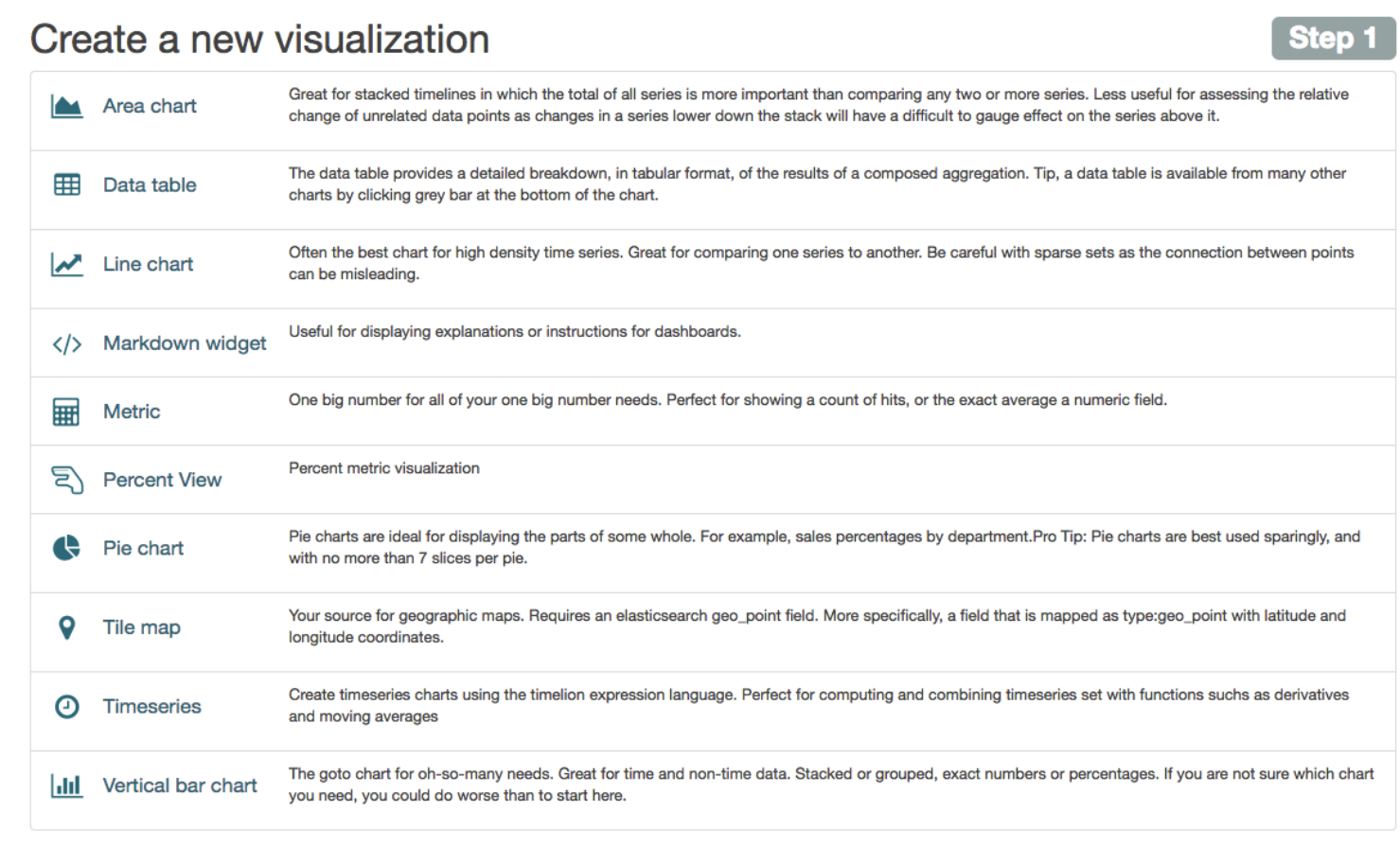
应用实践－视频播放质量报表

Discover



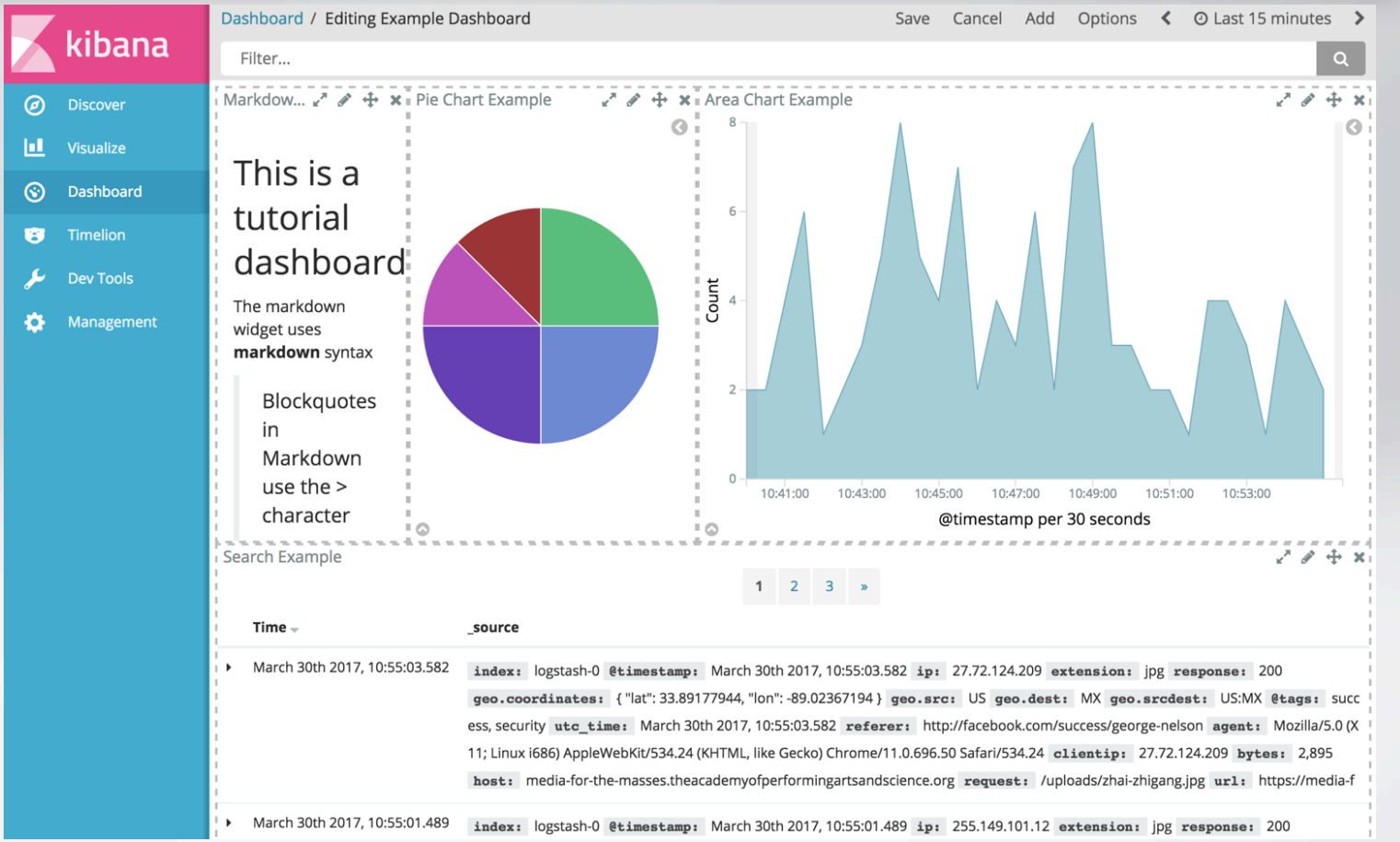
应用实践－视频播放质量报表

Visualization



应用实践－视频播放质量报表

Dashboard



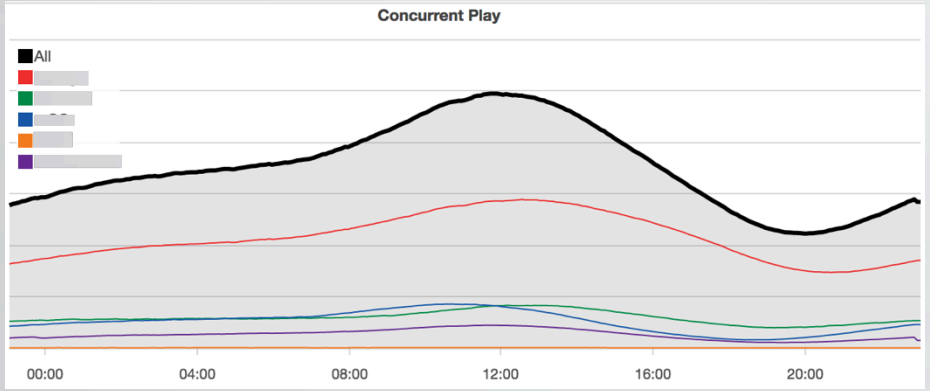
应用实践－视频播放质量报表

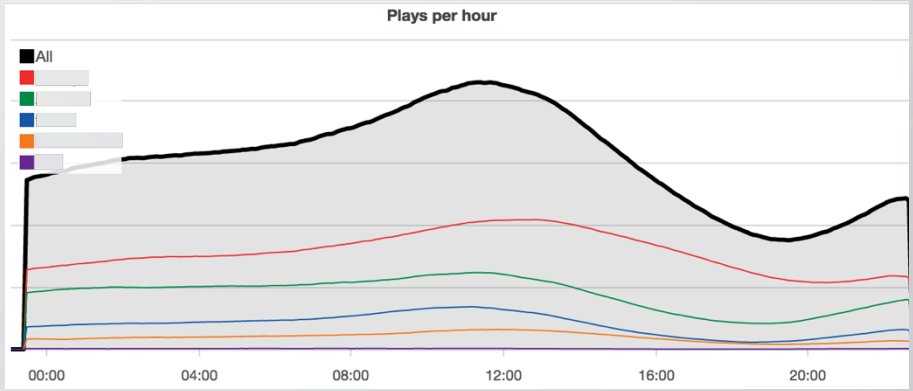
- 问题：播放日志有了，我们怎么利用?
  - Each record with 100 ～ 200 fields

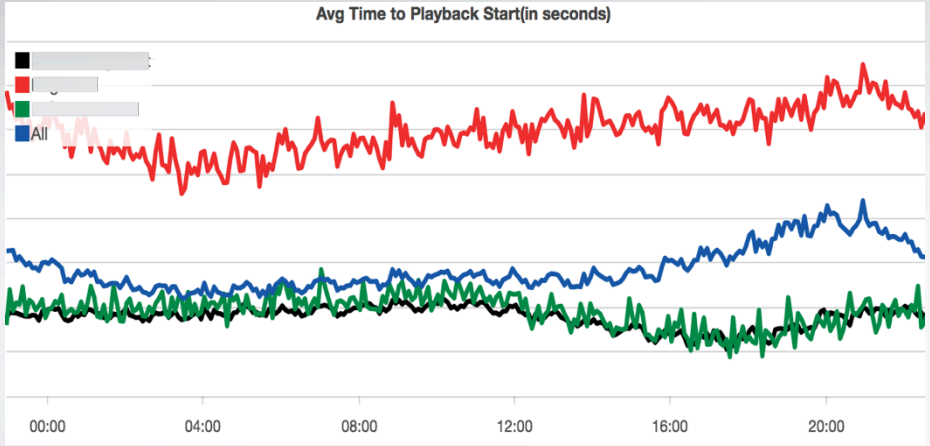

```
{
  "app": "hulu",
  "bitrate": 3200,
  "bitrate_switch_count": 2,
  "content_id": "1043295",
  "client": "xboxone",
  "time_to_playback_start": 2312,
  "geo": [0.4073030090332031E2, -0.814533462524414E2],
  "timestamp": "May 18th 2017, 17:33:57.963"
}
```

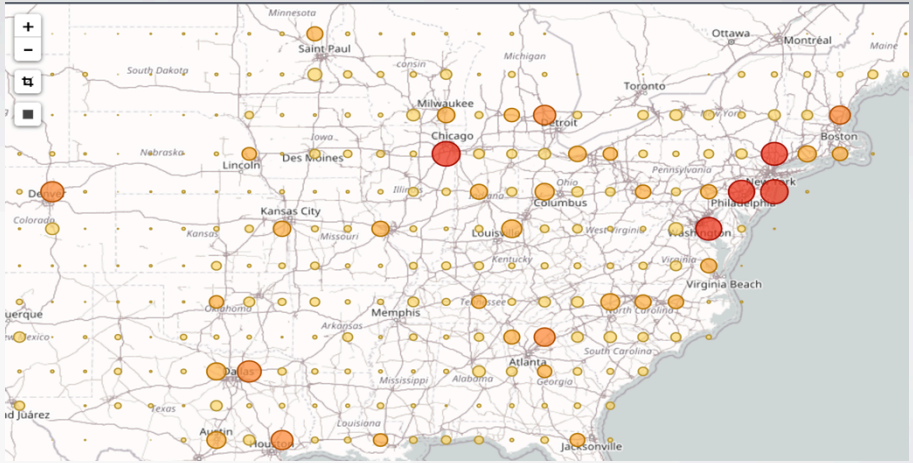

- 比较关心的指标
  - 播放量
  - 播放启动时间
  - 视频卡顿率
  - 播放码率
- 形成 QoS dashboard

应用实践－视频播放质量报表



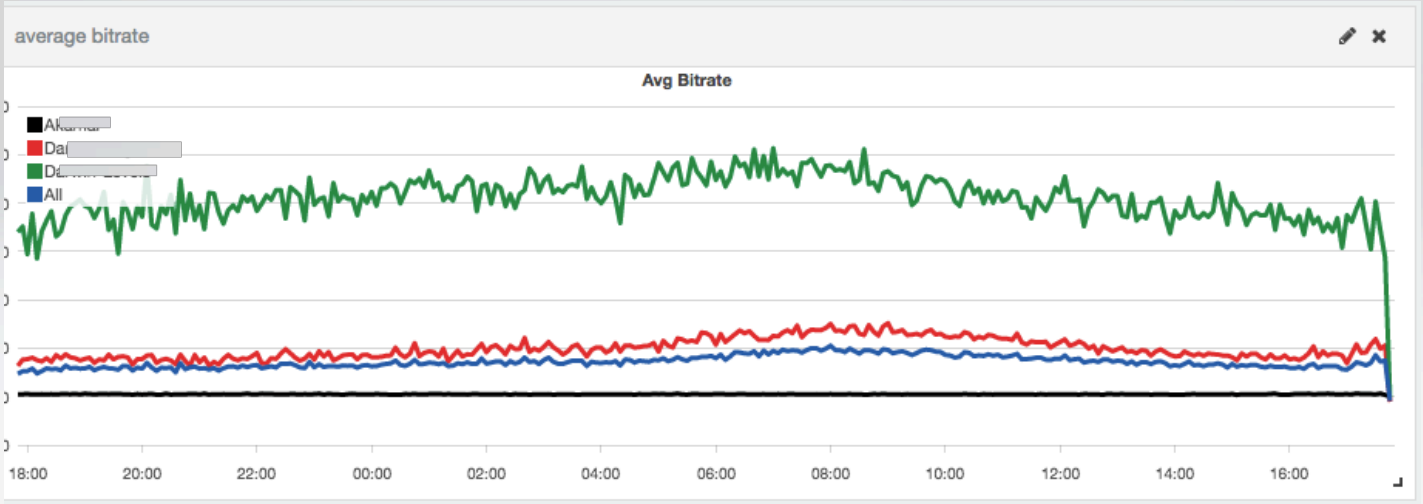






应用实践－视频播放质量报表

- 除此之外，支持Ad hoc的 query & dashboard
  - Drill down by CDN, ISP, Content, etc.
  - 即时地比较各个CDN provider的质量



主要内容

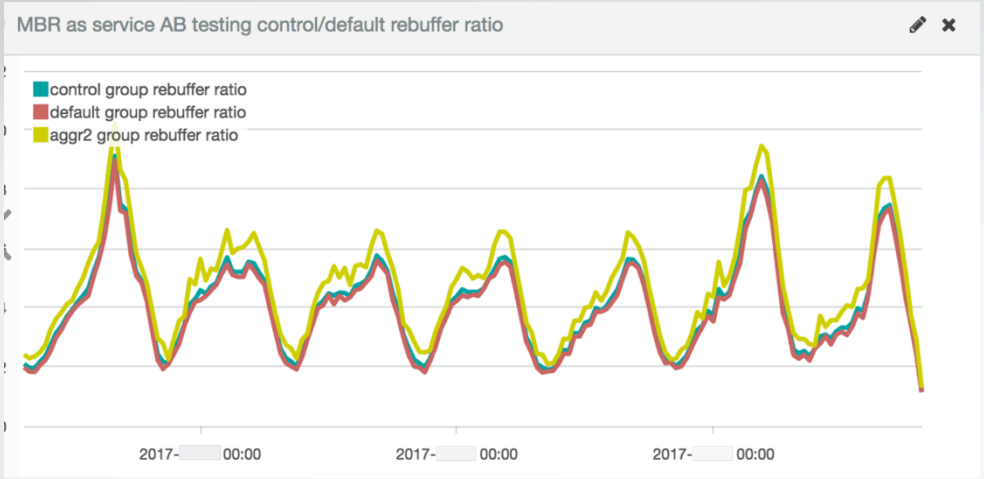
- 技术架构
- 应用实践
  - 视频播放质量报表
  - AB test 支持
  - Kibana可视化定制
- 经验心得
- Q & A

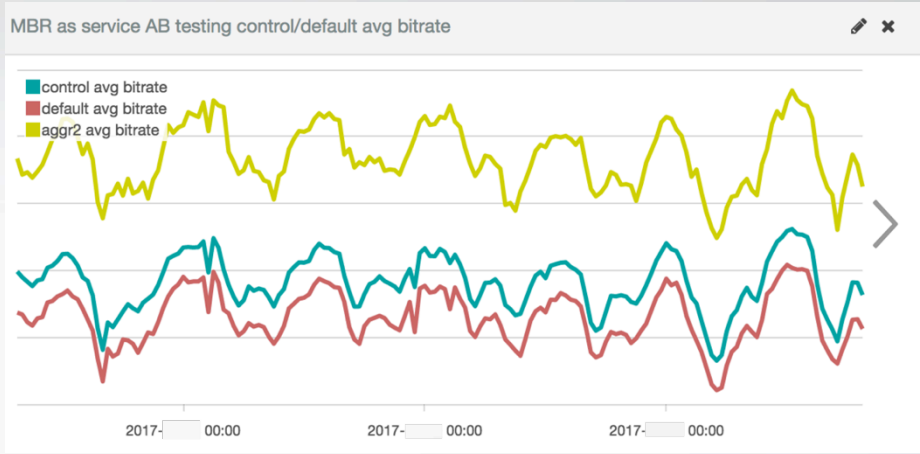
应用实践－AB Test 支持

- 问题：如何验证新的播放策略，以减少卡顿，并且提高码率?
- Elasticsearch＋Kibana对数据灵活支持
  - 设置对照组和实验组
  - 收集目标数据
  - 可视化对比实验结果
- 优势
  - 开发人员完全掌握原始数据，可自由可视化数据，减少对数据团队的依赖

应用实践－AB Test 支持

- 具体场景：观看高清视频是否可以提高用户留存率?
- 实验组设置
  - 对照组
  - 实验组：激进地使用高码率的片源





主要内容

- 技术架构
- 应用实践
  - 视频播放质量报表
  - AB test 支持
  - Kibana可视化定制
- 经验心得
- Q & A

应用实践－kibana可视化定制

- Kibana可视化：饼图，线图，柱状图，...
- 然而美中不足：
  - 无法对多个查询进行数学运算后再可视化
  - 在一张图中显示多个的查询结果
  - 操作简洁，可是不够灵活，如支持moving average平滑化

应用实践－kibana可视化定制 - Timelion

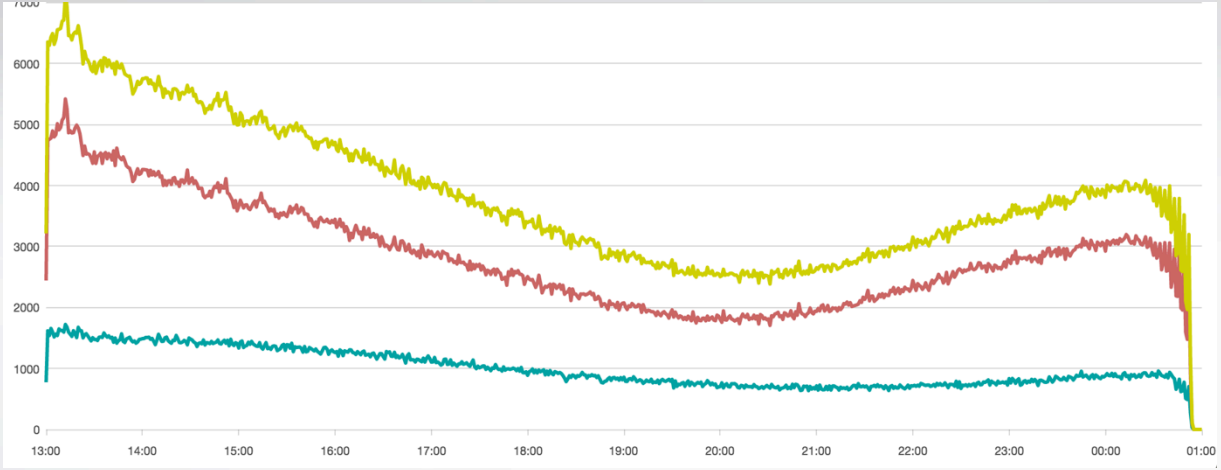
- Open source kibana plugin: timelion
- 功能：
  - 支持数学运算：加、减、乘、除、abs、movaverage、min、max...
  - 定制画图：控制线粗、颜色、填充
  - 基于 jQuery flot，易于自主开发定制新功能

应用实践－kibana可视化定制 - Timelion

Timelion add

add two data series
两个数列对应点相加

.es(q='query1'),
.es(q='query2'),
.es(q='query1').add(.es(q='query2'))



应用实践－kibana可视化定制 - Timelion

Timelion divide

divide two data series of the same length. 两个数列对应点相除

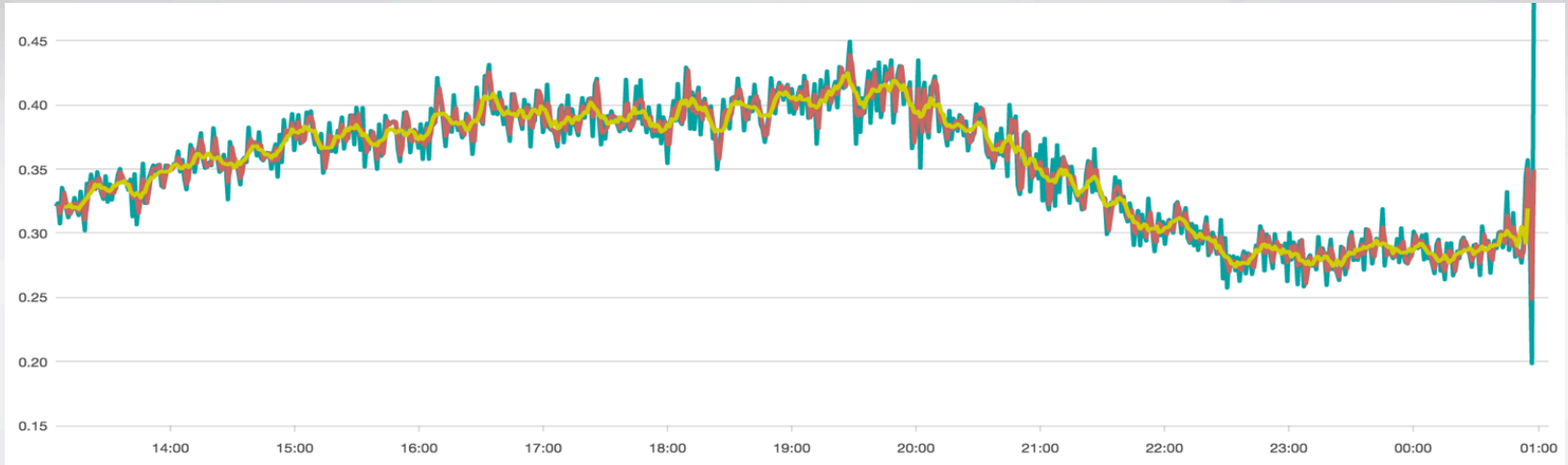
.es(q='query1'),
.es(q='query2'),
.es(q='query1').divide(.es(q='query2'))



应用实践－kibana可视化定制 - Timelion

Timelion movingaverage

平滑线图，多点均线图
.es(q='query1'),
.es(q='query2'),
.es(q='query1').divide(.es(q='query2')) (blue)
.es(q='query1').divide(.es(q='query2')).movingaverage(window=2) (red)
.es(q='query1').divide(.es(q='query2')).movingaverage(window=8) (yellow)



应用实践－kibana可视化定制 - Customize

- 定制Kibana
  - 使用OpenStreetMap & 高德地图替代默认地图，支持18级别放大



应用实践－kibana可视化定制 - Customize

- 定制Timelion
  - 支持 divideseries：多个数列对应除以多个数列
  - 支持柱状图显示
  - 支持Percentile查询

应用实践－kibana可视化定制 - Customize

- Divideseries


```
{
  "app": "hulu",
  "bitrate": 3200,
  "bitrate_switch_count": 2,
  "content_id": "1043295",
  "client": "xboxone",
  "rebuffer_count": 1,
  "time_to_playback_start": 2312,
  "geo": [0.4073030090332031E2, -0.814533462524414E2],
  "timestamp": "May 18th 2017, 17:33:57.963"
}
```

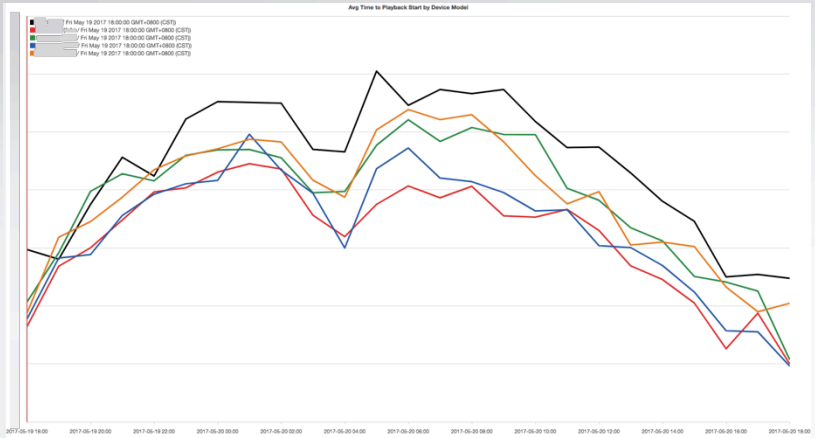

- 查询所有client上的rebuffer ratio
  - .es(q='rebuffer_count:>0').divide(.es(q='*'))
- 查询 xboxone 平台上的 rebuffer_ratio
  - .es(q='rebuffer_count:>0 AND client:xboxbox').divide(.es(q='client:xboxone'))
- 查询 Android、iOS、tvOS、XboxOne、Chromecast上的rebuffer_ratio
  - .es(q='rebuffer_count:>0 AND client:Android').divide(.es(q='client:Android'))
  - .es(q='rebuffer_count:>0 AND client:iOS').divide(.es(q='client:iOS'))
  - ...

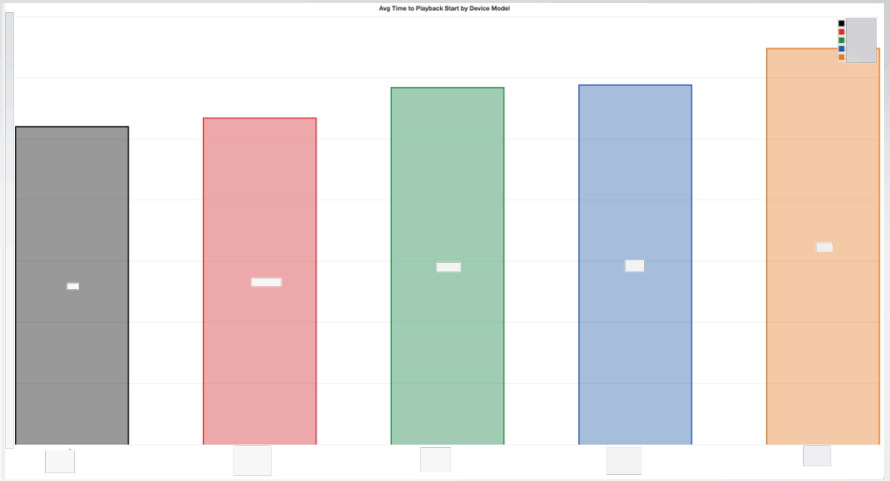
Try divideseries, 一行解决

- 查询 Android、iOS、tvOS、XboxOne、Chromecast上的rebuffer_ratio
  - .es(q='rebuffer_count:>0', split='cilent:5').divideseries(.es(q='*', split='client:5'))

应用实践－kibana可视化定制 - Customize

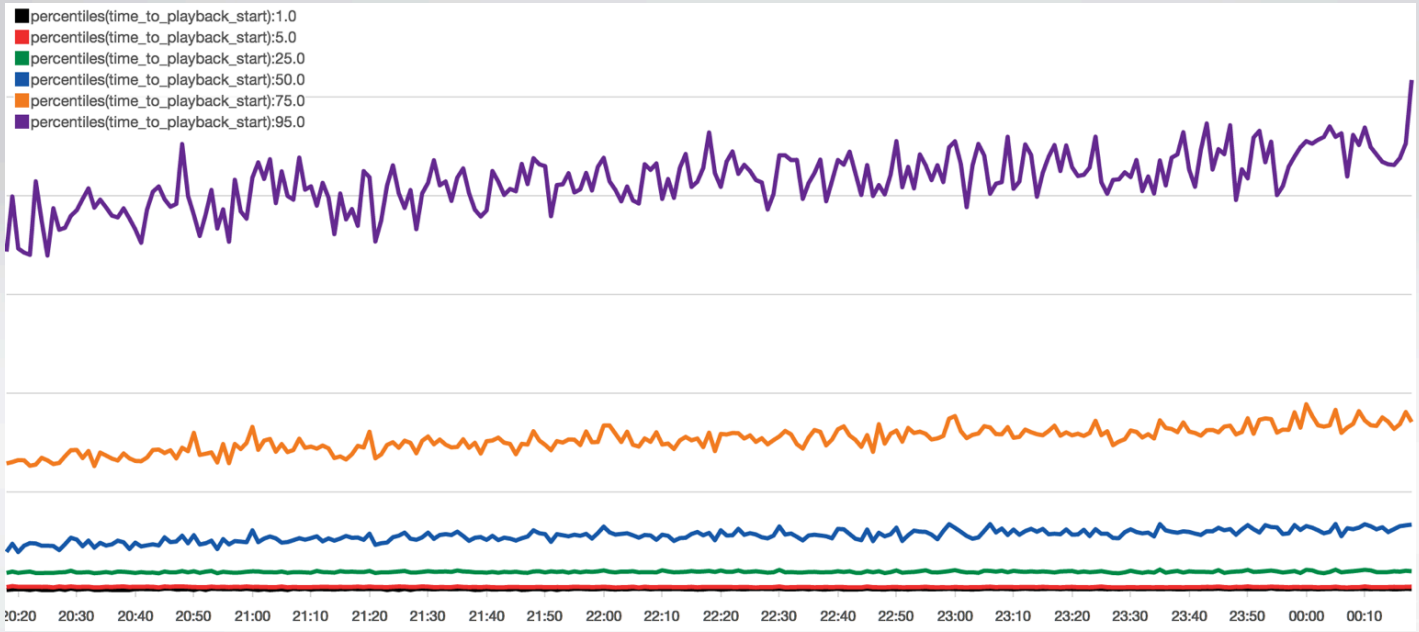
- Timelion柱状图显示
  - Timelion是基于jQuery flot画图
  - 支持柱状图显示
  - 线图到柱状图的值域映射：avg, sum, etc.





应用实践－kibana可视化定制 - Customize

- Timelion percentile
  - 查询百分位图：1,5,25,50,75,95
  - .es(q='*', split='client:6', metric='percentiles:time_to_playback_start:1,5,25,50,75,99')



主要内容

- 技术架构
- 应用实践
  - 视频播放质量报表
  - AB test 支持
  - Kibana可视化定制
- 经验心得
- Q & A

经验心得

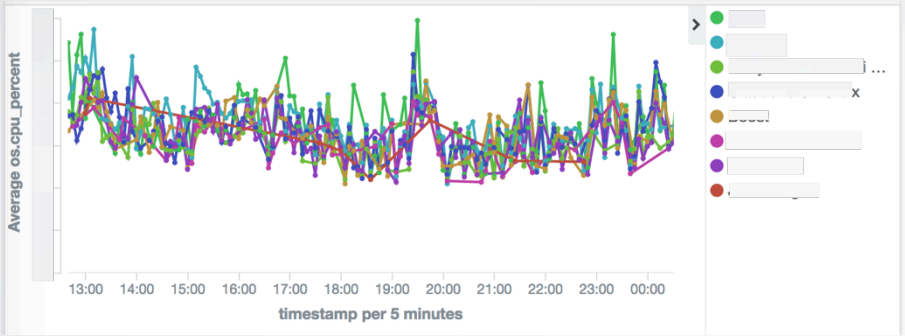
- 应用 Elasticsearch 和 Kibana 遇到的坑和经验
  - 针对时间序列类的数据：Index命名实例
    - [domain]_[team]_[app]_[version]_[date]
    - Hulu_player_qos_**v01**_2017-03-03, Hulu_player_qos_**v02**_2017-03-03
    - 解决问题：日内更改index mapping，业务需求改变频繁
      - 对所有 text 字段，增加 field_name.raw 字段 (not analyzed text field)
  - 某字段类型常变化
    - 为避免与现有mapping冲突，采用备用字段
    - Myfieldname + myfieldname_v01
    - 解决问题：已存文档的content_id字段映射为int，新文档的content_id是

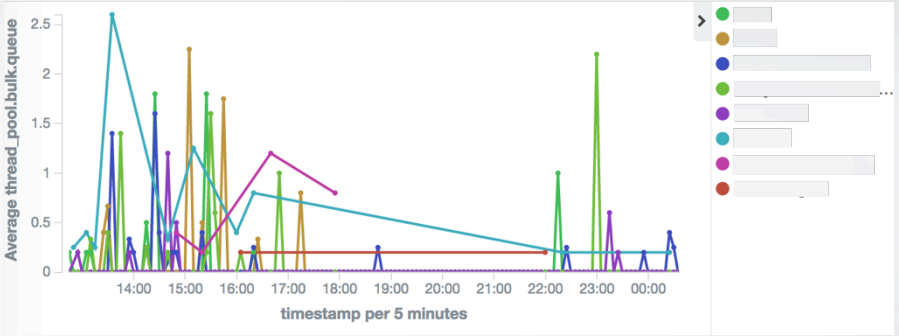
经验心得

- 应用 Elasticsearch 和 Kibana 遇到的坑和经验
  - Shard数量
    - fixed
    - 1 ~ 3 shards per datanode per index
    - Replication shard
  - 控制数据存入速度
    - Google guava rate limiter

经验心得

- 应用 Elasticsearc 和 Kibana 遇到的坑和经验
  - 自己动手监控Elasticsearch cluster的状态
    - 定期查询 /_nodes/stats
    - CPU、内存、插入队列等待长度、文档总数、shard总数、relocate的shard数
    - ES cluster历史状态全知道





未来计划

- 未来计划
  - 扩展Timelion
    - 整合更多数据源
    - 支持用户编写自定义计算逻辑
  - 增强Kibana
    - 扩展Kibana本身的Visualization，支持灵活计算
    - 导出Dashboard数据
  - Smart Alert

Q & A



Thanks