

Elasticsearch在阿里云的实践分享

广州MEETUP

阿里云 赵弘扬



阿里云Elasticsearch



Kibana



Elasticsearch



Logstash



Beats



Security

Alerting

Monitoring

Graph

Reporting

Machine Learning



阿里云
aliyun.com

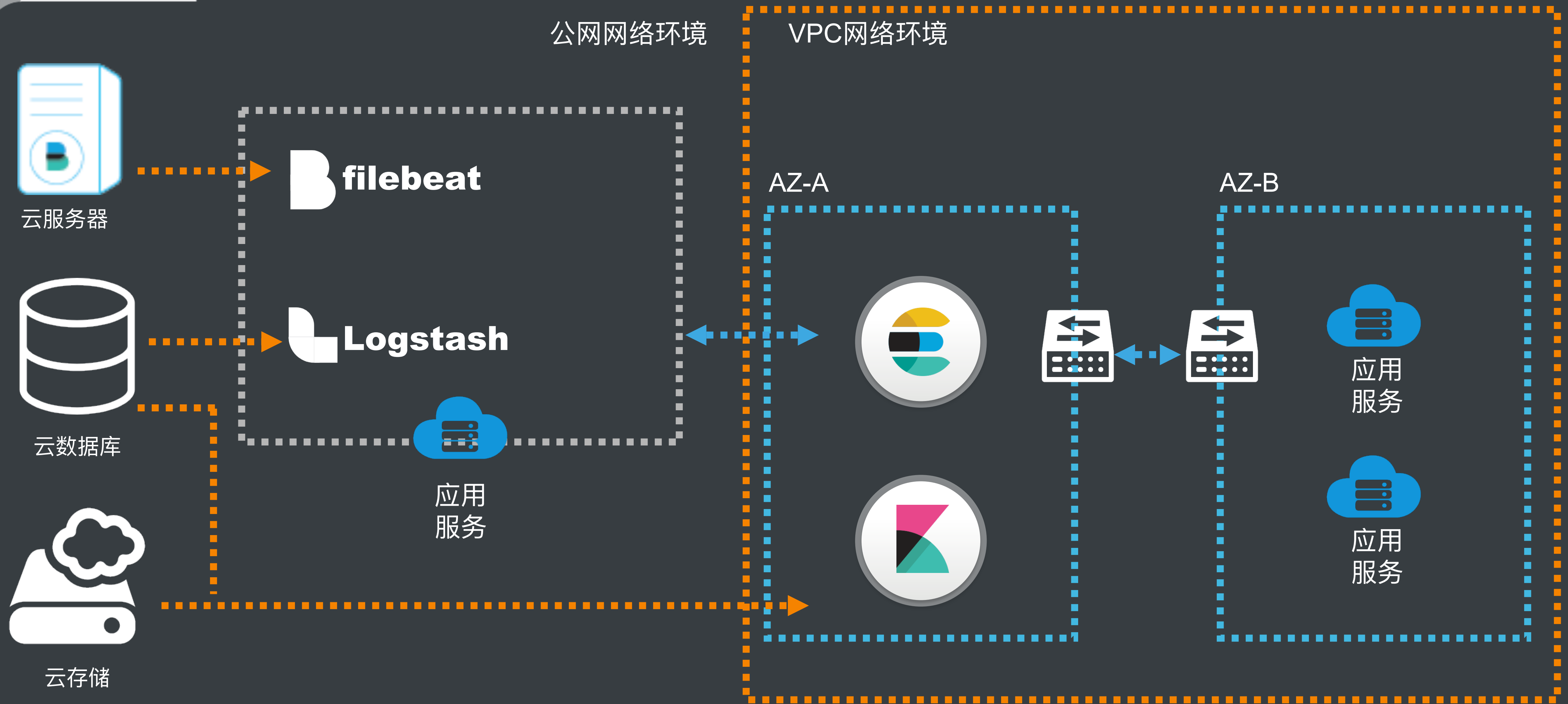
阿里云Elasticsearch功能优势

云上生态

集群安全

弹性扩容

阿里云Elasticsearch



举个“栗子”——图书馆的搜索需求

图书馆Q是一家大型图书馆，图书馆藏书众多，纸质图书600多万册，电子图书7000多万册，总数有八千多万册，这些图书之前都是人工检索维护的，现在需要做一个系统来存储管理这些图书信息。

需求如下：

- 图书总量目前八千多万册，考虑到未来二十年的增长，需要系统能支持一亿的存储量。
- 图书信息很重要，不能接受丢失发生。
- 图书的名字和作者名字需要支持模糊搜索。
- 每本书的属性最多有一百多个，且不固定，不同类型的图书的属性列差异较大。且未来可能会新增属性列。

图书馆的搜索需求

需要面对的几个问题：

存储成本

极高的数据可靠性要求

全量数据的存储成本控制

数据类目属性列较多且不固定

查询效率

固定属性模糊查询

精确查询

快速查询的结果返回

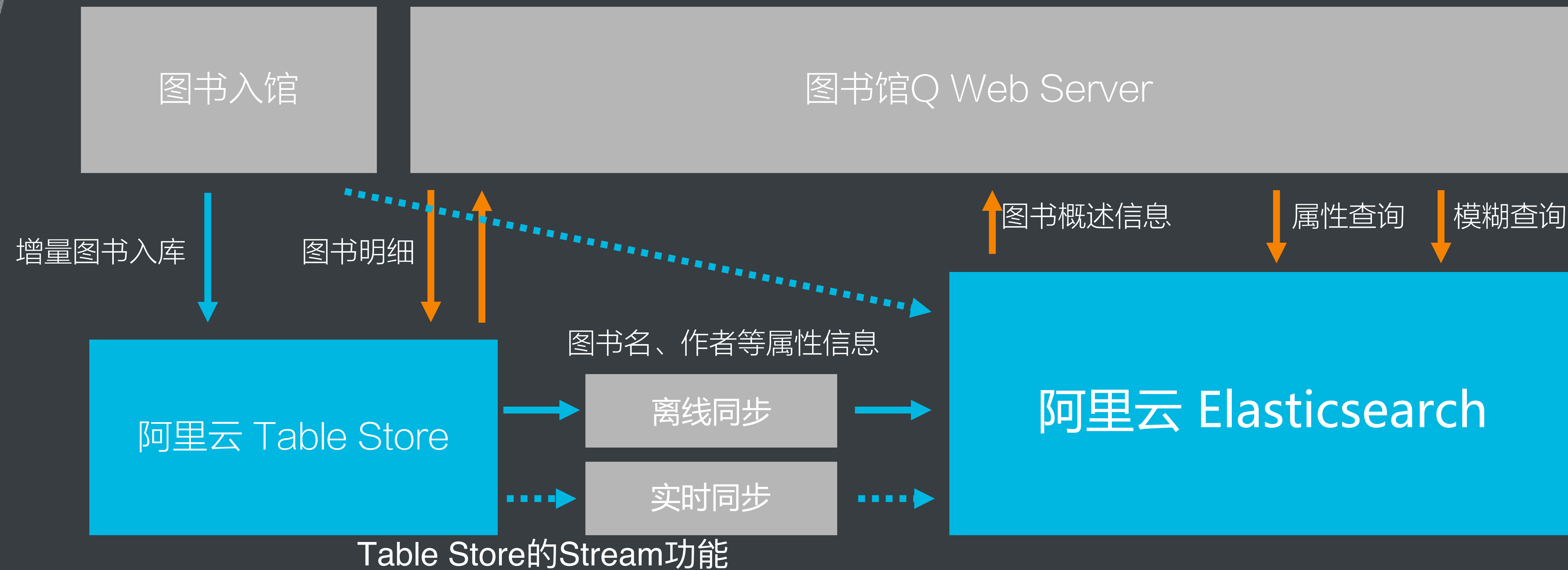
增量更新

增量数据更新

减少数据不一致风险

图书馆的搜索需求

解决方案及架构



再举个“栗子”——图书馆的日志分析需求

图书馆Q是每天接待上万名的访问用户，并且随着信息化的发展，也推出了远程借书、阅读等服务，随着访问形式的增长，图书馆Q希望自己可以对线上的客户访问分布做一次分析，获取自己用户的地域分布、访问时间、阅读兴趣等信息。

需求如下：

- 访问日志和用户行为分别以日志和结构化数据的形式存储在服务器和关系型数据库里，每天大概有上万条的记录增加。
- 所有的数据之间由用户的ID作为唯一标志，涉及多表Join。
- 需要获取地域、阅读类目、访问活跃时间等数据信息，并可视化呈现。
- 分析结果需要按日更新。

图书馆的日志分析需求

需要面对的几个问题：

离线数据处理

多个数据源

数据需要ETL后才能可视化分析

涉及到多表join

分析

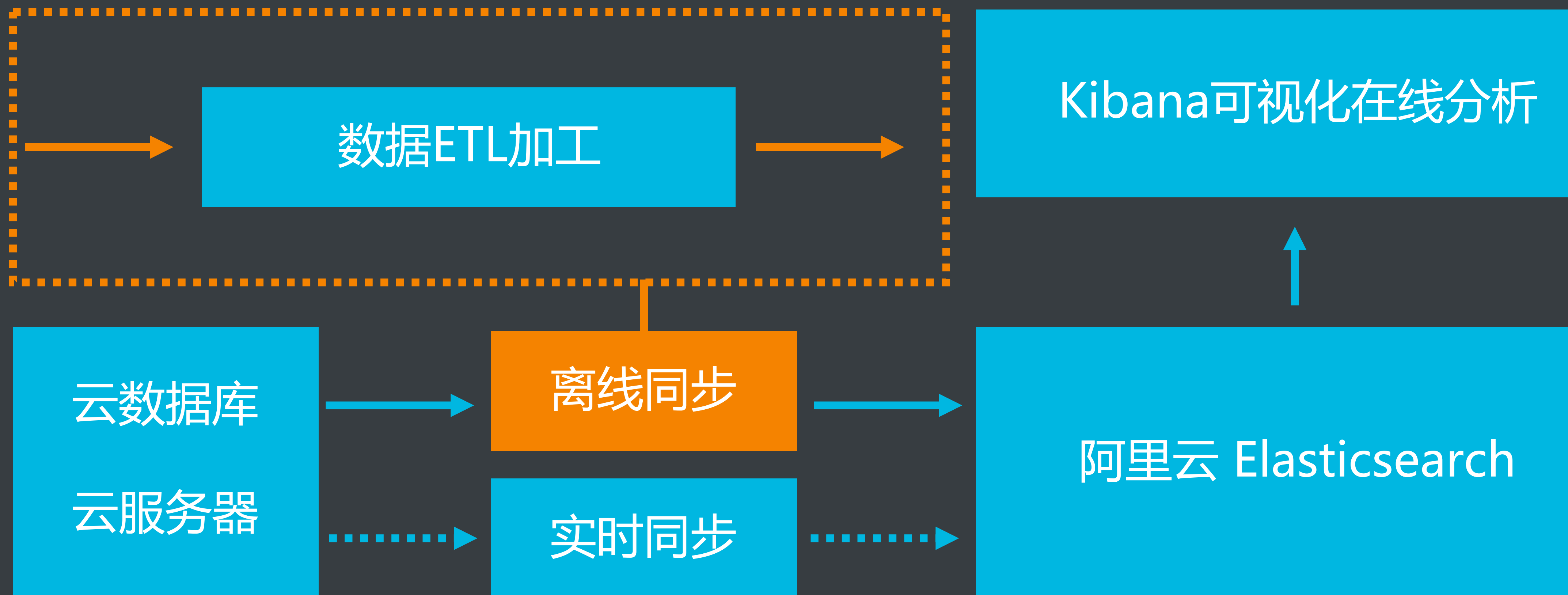
需要ms级返回全量分析结果

需要可视化数据展现

定时结果更新

图书馆的日志分析需求

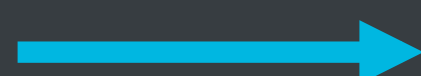
解决方案及架构



图书馆的日志分析需求

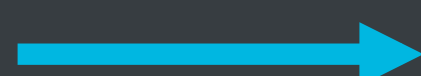
浏览行为

服务器



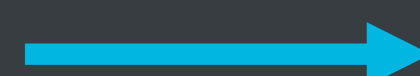
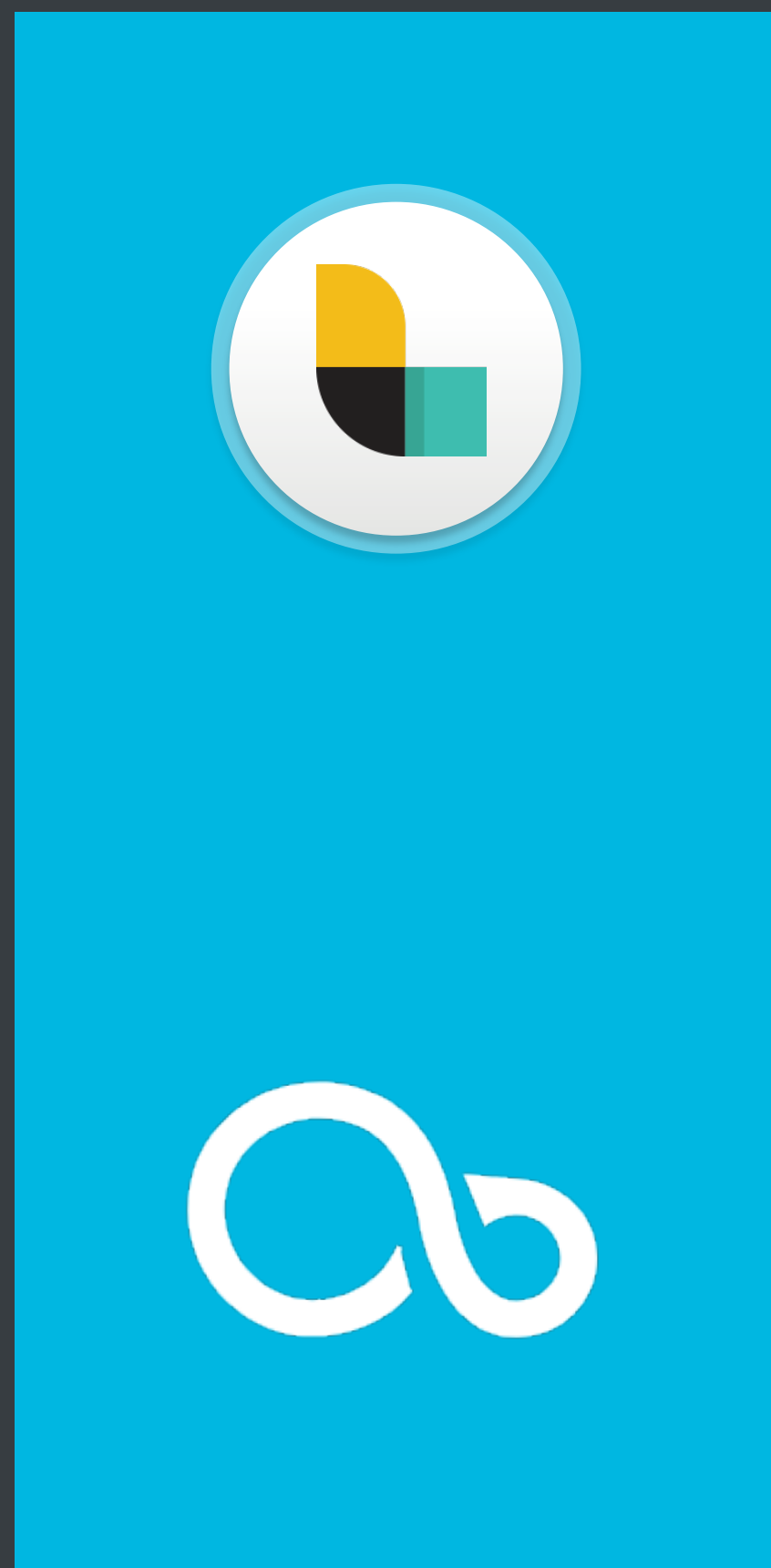
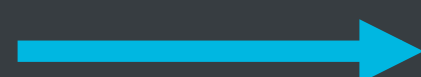
借阅记录

关系型数据库



读书笔记

MongoDB



阿里云 Elasticsearch

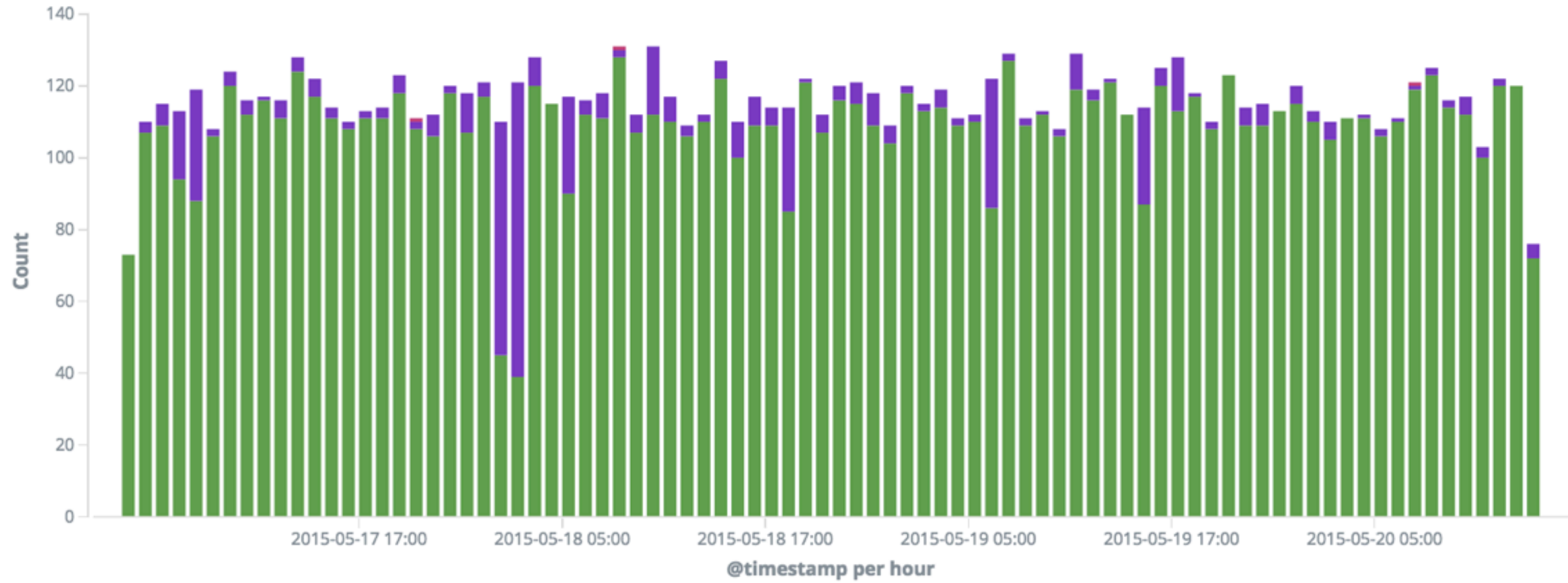
Total Visitors

9,986
Total IPs

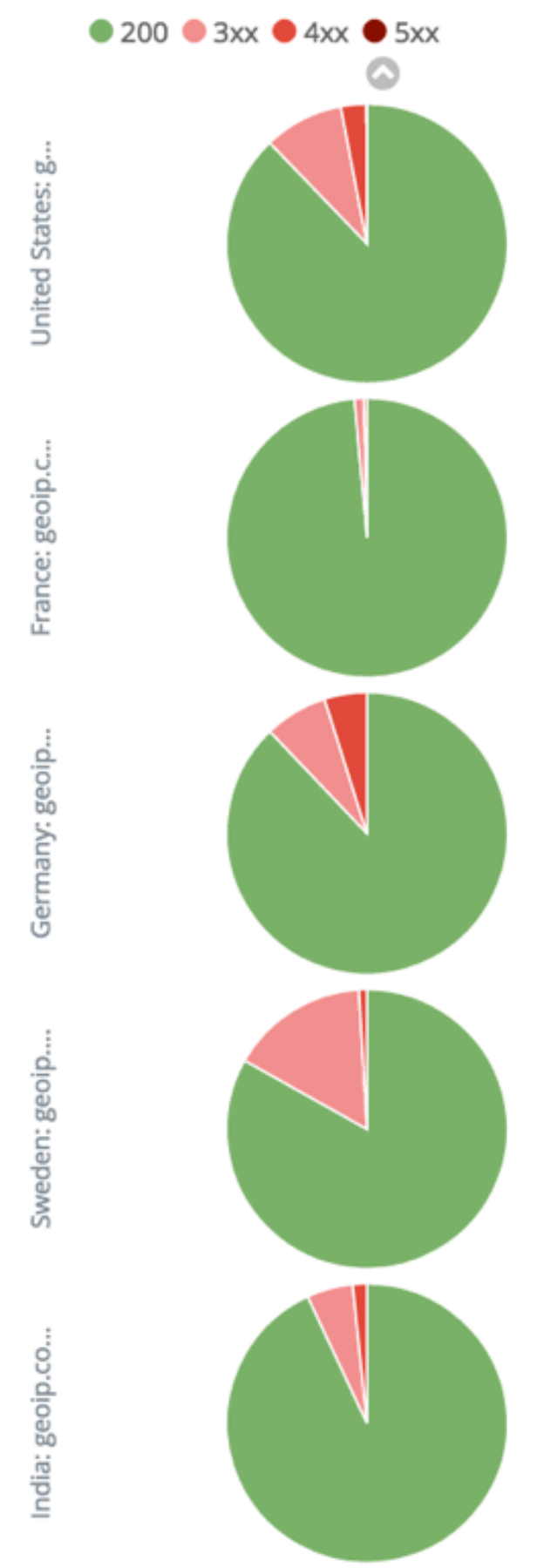
Unique Visitors

1,750
Unique IPs

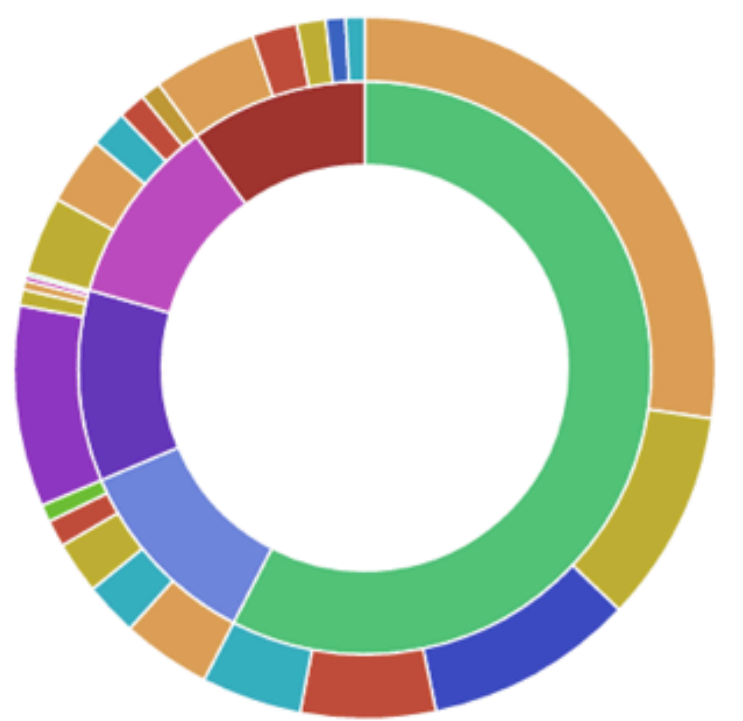
Response Code by Time



Response Code by Country

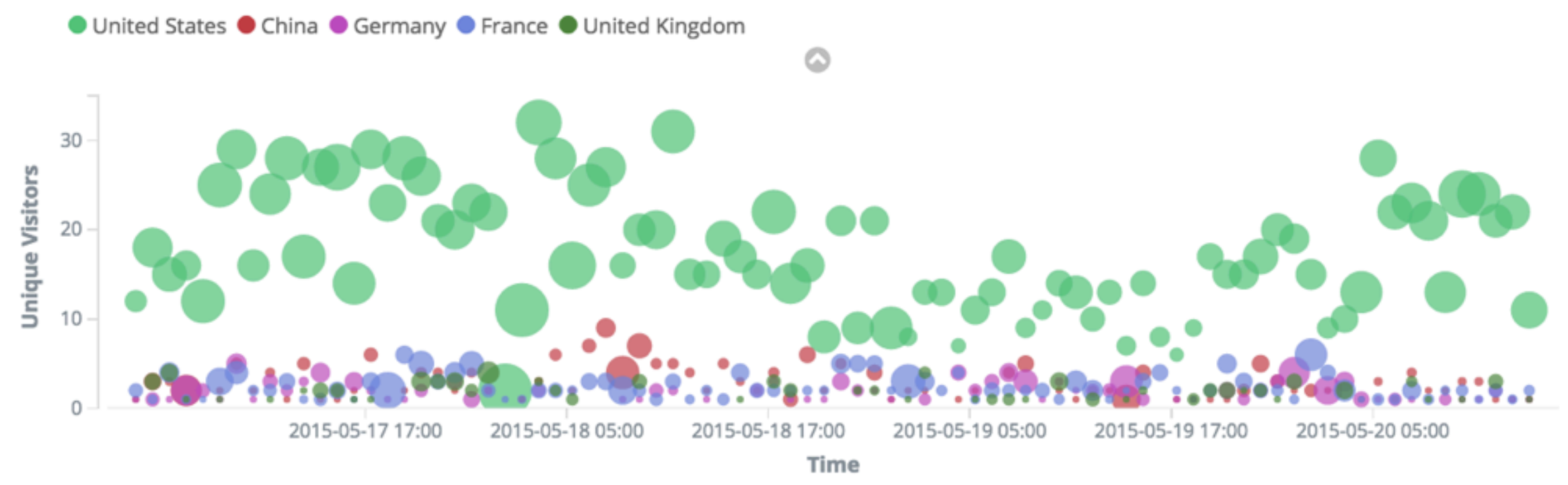


Traffic by Country & OS



- United States
- France
- Sweden
- Germany
- India
- Windows 7
- Linux
- iOS 6.0
- Windows XP
- Ubuntu
- Mac OS X 10.9
- Mac OS X 10.9.1
- Mac OS X 10.8
- Mac OS X 10.7
- Windows 8.1

Bytes vs. Time



Unique Visits by City

City	Unique Visitors	Total Visitors
Beijing	114	192
Redmond	38	75
Morganton	28	108
Chicago	26	53
Ashburn	24	146
San Jose	23	46
San Francisco	22	236

Traffic vs. Location



急需人才 求志同道合者加入我们



深圳



北京、杭州

联系我：

zhywd9104@163.com

为了无法计算的价值 |  阿里云

