AIOps与ClickHouse的深度碰撞

2021年03月 高鹏

目录

目录

ClickHouse的引入

about me

MySQL DBA => 数据分析

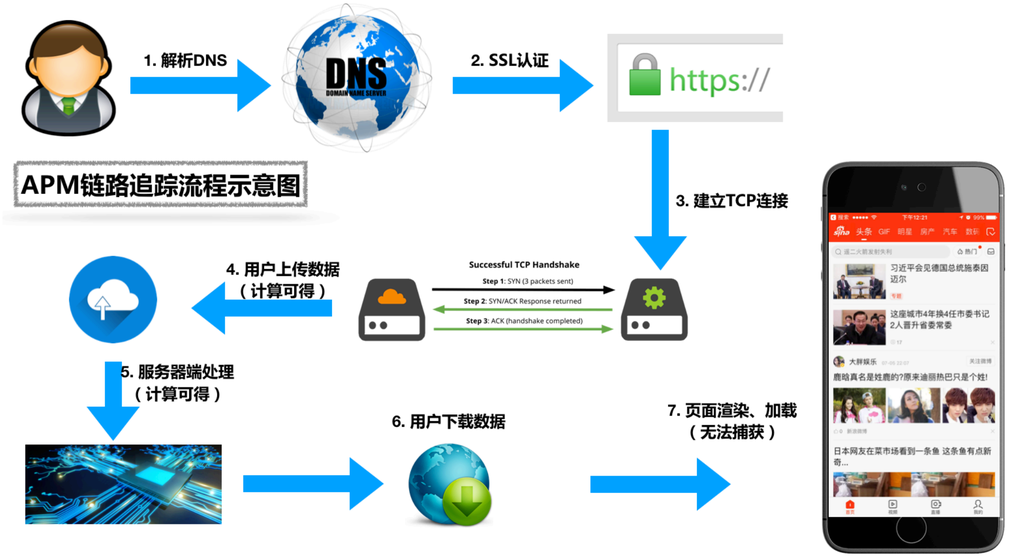
到底有没有好的OLAP产品?

当线上的MySQL出现了慢查询……

ClickHouse的引入 穷而思变

问题：

如何实现百维、亿级、秒级数据分析



新浪新闻APP APM管理

ClickHouse的引入　传统手艺

Hadoop

Spark：计算引擎　　链路复杂

Hive：数据仓库　　速度奇慢

MySQL：数据加速　　出错难修

ELK

架构简单

DSL反人类

资源永不够

非本职工作

ClickHouse的引入　传统手艺

Hadoop

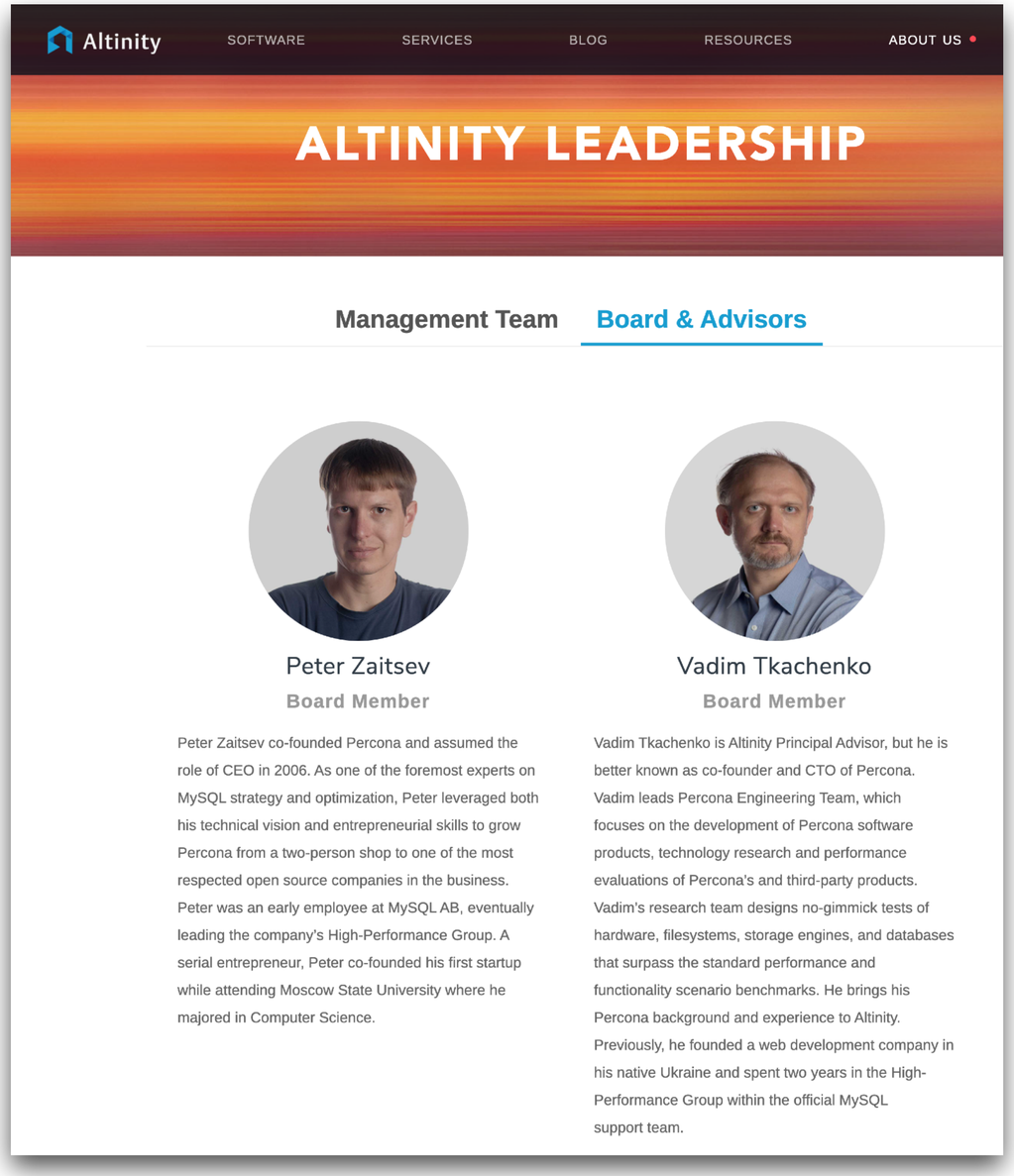
ELK

人力成本高
机器成本高
技术难度大
速度还慢

ClickHouse的引入 ClickHouse的结缘

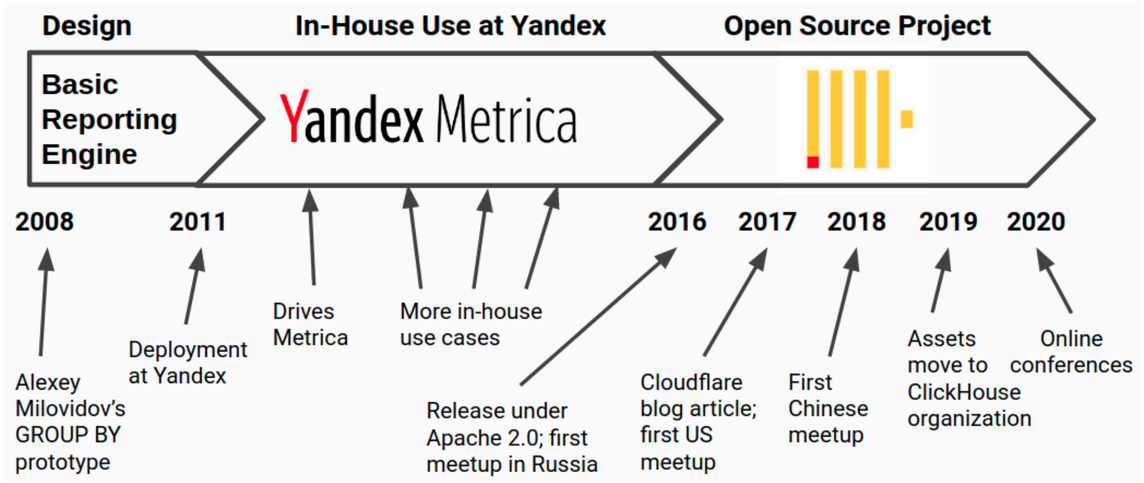
Percona对ClickHouse的推广

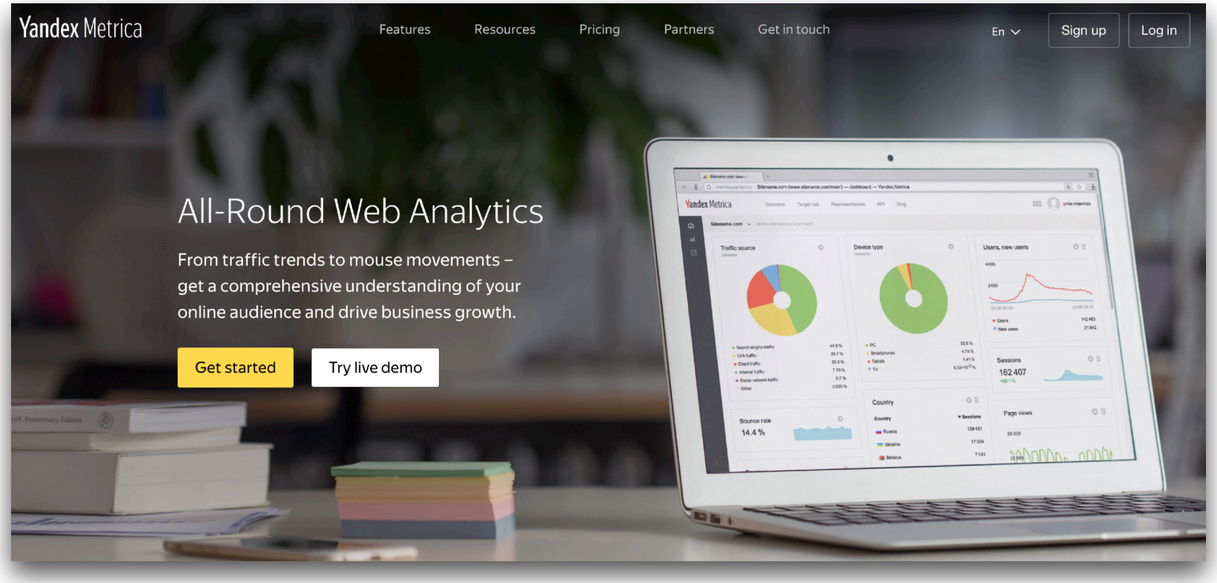
Percona与Altinity的关系



ClickHouse的引入 ClickHouse的结缘

ClickHouse的发展历程





ClickHouse的引入 ClickHouse的结缘

- 单机写入峰值100w
- 数据压缩3-10倍
- 查询性能碾压开源、商业产品

```
:) select count(*) from apm_msg_all;
```

```
:)
```

Compare: ClickHouse | Vertica | Vertica (x3) | Vertica (x6) | InfiniDB | MonetDB | Infobright | Hive | MySQL | MemSQL | Greenplum(x2) | Greenplum

Dataset size: 10 mln. | 100 mln. | 1 bn.

Run number: first (cold cache) | second | third

Relative query processing time (lower is better):

Full results:

Query	ClickHouse (1.1.53960)		InfiniDB (Enterprise 3.6.23)		Hive (0.11, ORC File)		Greenplum (4.3.9.1)	
SELECT count() FROM hits	(0.297 s.)	(0.071 s.)	x6.97 (2.070 s.)	x30.91 (0.340 s.)	x372.65 (110.676 s.)	x9557.27 (105.130 s.)	x13.84 (4.110 s.)	x392.73 (4.320 s.)
SELECT count() FROM hits WHERE AdvEngine	(0.470 s.)	(0.000 s.)	x4.39 (2.060 s.)	x30.00 (0.300 s.)	x319.05 (55.195 s.)	x3643.50 (36.435 s.)	x14.80 (2.560 s.)	x253.00 (2.530 s.)
SELECT sum(AdvEngineID), count(), avg(Resol	(0.274 s.)	(0.068 s.)	x5.29 (1.450 s.)	x18.92 (1.230 s.)	x145.95 (39.991 s.)	x540.66 (35.143 s.)	x18.18 (4.980 s.)	x79.08 (5.140 s.)
SELECT sum(UserID) FROM hits	(0.562 s.)	(0.046 s.)	—	—	x80.55 (44.465 s.)	x741.98 (34.131 s.)	x12.03 (6.640 s.)	x130.22 (5.990 s.)
SELECT uniq(UserID) FROM hits	(0.480 s.)	(0.057 s.)	x8.62 (4.180 s.)	x40.10 (3.890 s.)	x228.23 (110.690 s.)	x1092.30 (105.953 s.)	x12.08 (5.860 s.)	x58.97 (5.720 s.)
SELECT uniq(SearchPhrase) FROM hits	(0.625 s.)	(0.216 s.)	x42.11 (26.320 s.)	x133.03 (28.070 s.)	x108.99 (68.119 s.)	x307.26 (64.831 s.)	x10.74 (6.710 s.)	x32.04 (6.760 s.)
SELECT min(EventDate), max(EventDate) FROI	(0.184 s.)	(0.04 s.)	x8.29 (1.360 s.)	x22.13 (1.040 s.)	x230.54 (37.809 s.)	x702.57 (33.021 s.)	x25.91 (4.250 s.)	x93.40 (4.390 s.)
SELECT AdvEngineID, count() FROM hits WHE	(0.059 s.)	(0.009 s.)	x10.00 (0.580 s.)	x32.00 (0.320 s.)	x960.50 (53.788 s.)	x5126.10 (51.261 s.)	x80.89 (4.530 s.)	x423.00 (4.230 s.)
SELECT RegionID, uniq(UserID) AS u FROM hit	(0.985 s.)	(0.427 s.)	x5.32 (2.140 s.)	x10.63 (4.540 s.)	x90.56 (87.479 s.)	x199.21 (85.062 s.)	x9.44 (9.120 s.)	x21.62 (9.230 s.)
SELECT RegionID, sum(AdvEngineID), count() /	(0.970 s.)	(0.309 s.)	x8.07 (7.830 s.)	x16.10 (8.180 s.)	x109.87 (106.577 s.)	x202.52 (102.879 s.)	x15.23 (14.770 s.)	x30.28 (15.380 s.)
SELECT MobilePhoneModel, uniq(UserID) AS u	(0.620 s.)	(0.196 s.)	x3.16 (1.960 s.)	x8.33 (1.400 s.)	x97.42 (60.400 s.)	x318.44 (53.498 s.)	x5.19 (3.220 s.)	x17.08 (2.870 s.)
SELECT MobilePhone, MobilePhoneModel, uniq	(0.665 s.)	(0.197 s.)	x2.67 (1.750 s.)	x8.13 (1.520 s.)	x93.55 (61.275 s.)	x287.16 (53.698 s.)	x6.73 (4.410 s.)	x22.94 (4.290 s.)

ClickHouse的收益

某APM产品示例

某APM指标数据

ClickHouse的收益

某APM产品示例

某后端服务质量监控

目录

ClickHouse的收益

成本

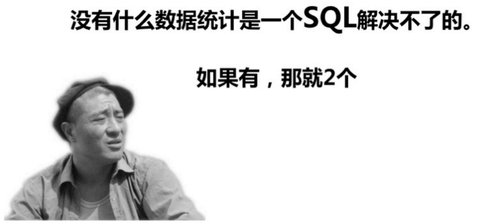
- 存储成本：ES的1/30
- 计算成本：利用率极高

效率

- 查询速度：远超Hive/ES
- 开发效率：使用SQL，替代Spark逻辑，数据链路变短

用户体验

- SQL
- 开发：彻底解决日志存放难题，随便打，读写都不会成为瓶颈



ClickHouse的实践

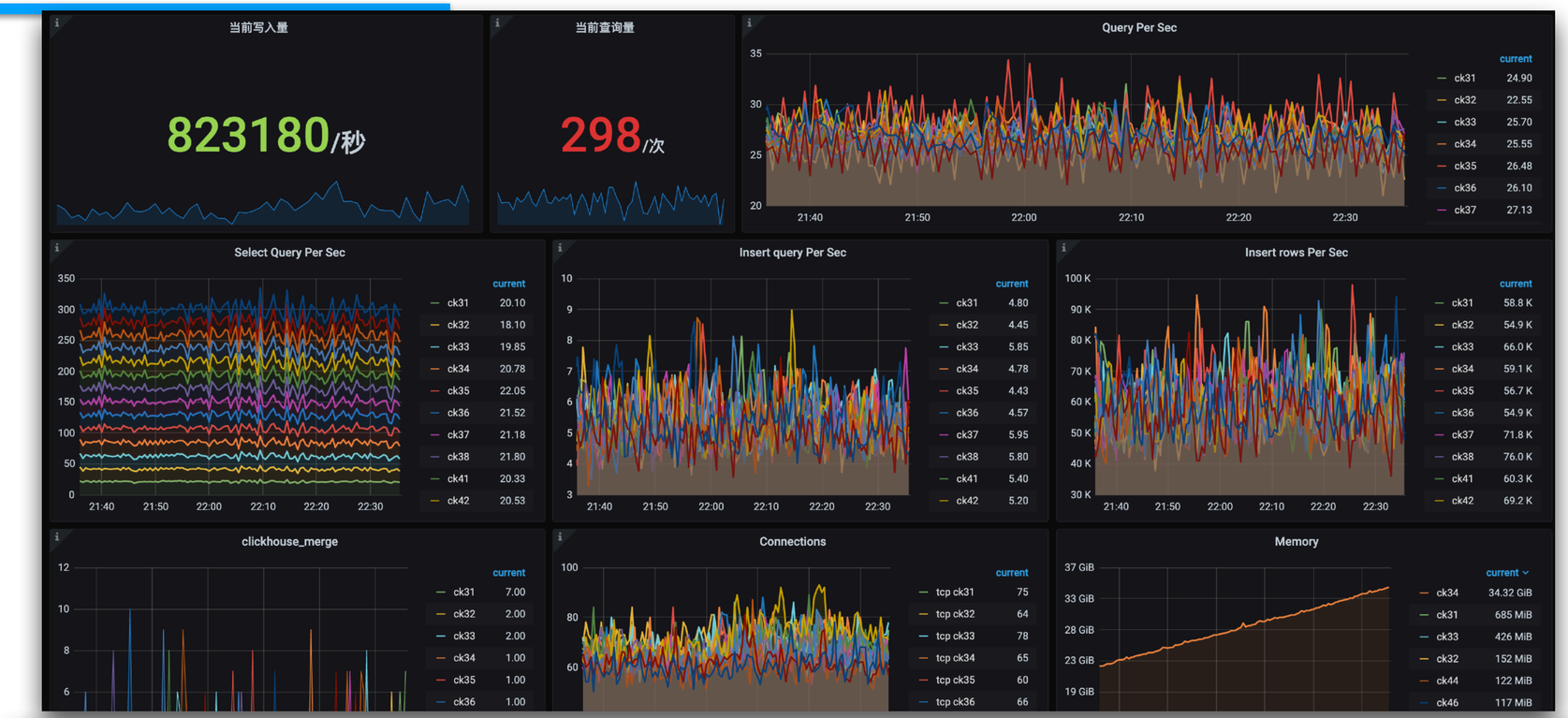
20台机器

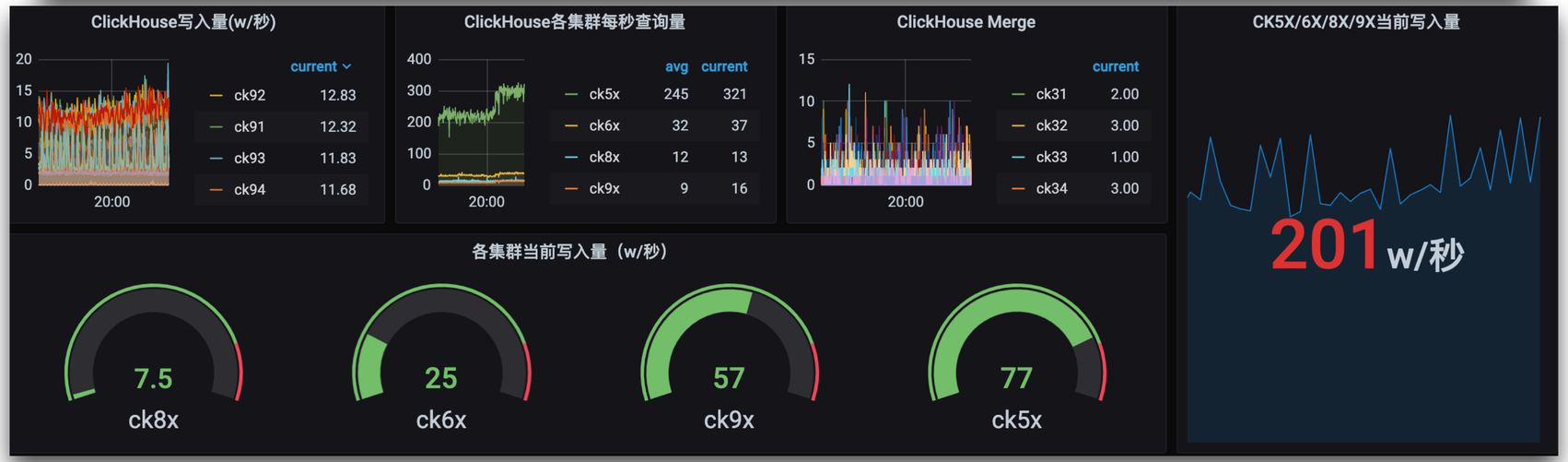
1300亿/天

3kw Query /天

搞清定位

用其所长





ClickHouse的实践 分析MySQL慢查询

clicktail

```
SELECT
    normalized_query,
    count(*) AS c,
    round(avg(query_time), 4) AS latency,
    round(quantile(0.99)(query_time), 4) AS latency_p99,
    round((latency * c) / (max(_time) - min(_time)), 4) AS load
FROM mysql_slow_log
GROUP BY normalized_query
HAVING c > 1
ORDER BY load DESC
LIMIT 10
```

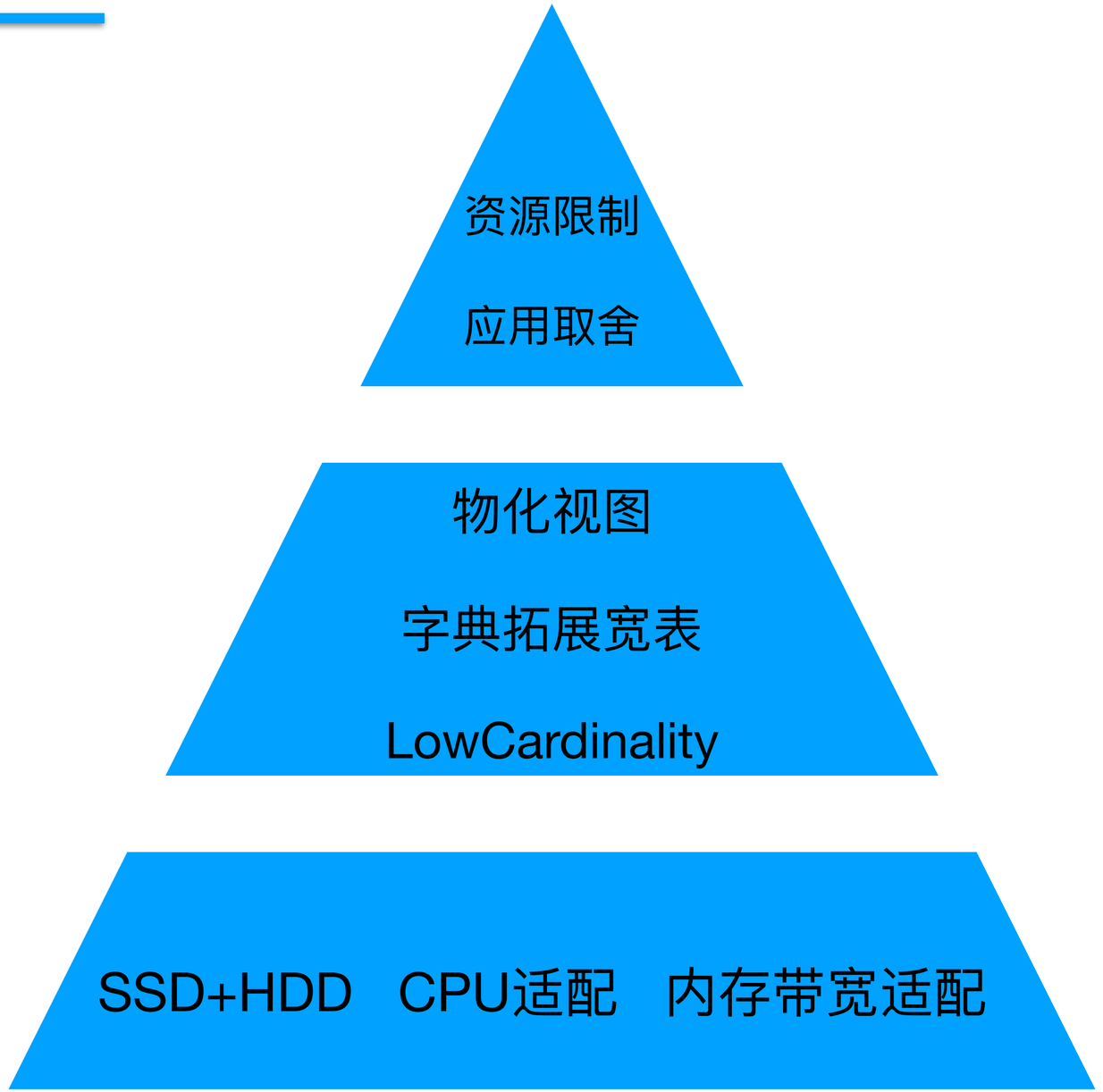
normalized_query	c	latency	latency_p99	load
commit	25341210	0.0207	0.0743	1.8352
update sbtest1 set c = ? where id = ?	30909259	0.0067	0.1341	0.7245
update sbtest1 set k = k + ? where id = ?	34156671	0.0059	0.1134	0.7051
delete from sbtest1 where id = ?	27972323	0.0067	0.1301	0.6557
select c from sbtest1 where id = ?	341676766	0.0001	0.0002	0.1195
select distinct c from sbtest1 where id between ? and ? order by c	34157704	0.0005	0.0007	0.0598
select c from sbtest1 where id between ? and ? order by c	34173442	0.0003	0.0006	0.0359
select c from sbtest1 where id between ? and ?	34153845	0.0003	0.0005	0.0358
select sum(k) from sbtest1 where id between ? and ?	34165923	0.0002	0.0004	0.0239
insert into sbtest1(id,k,c,pad) values (?, ?, ?, ?)	25344830	0.0002	0.0006	0.0177

```
10 rows in set. Elapsed: 8.415 sec. Processed 656.44 million rows, 35.36 GB (78.00 million rows/s., 4.20 GB/s.)
```

https://www.percona.com/blog/2018/02/28/analyze-raw-mysql-query-logs-clickhouse/

ClickHouse的实践 as TSDB

自底向上的优化



目录

AIOps与ClickHouse的碰撞

运维眼里的传统运维

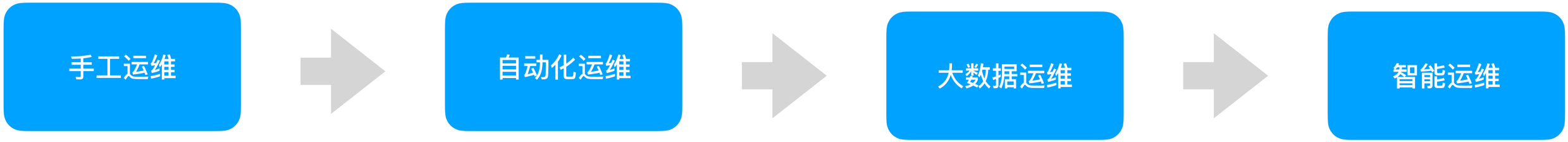
外行眼中的AI运维

AIOps与ClickHouse的碰撞

“AIOps是Gartner（高德纳，IT咨询公司）在2016年提出的概念，AIOps即 Artificial Intelligence for IT Operations，智能运维，将人工智能应用于运维领域，基于已有的运维数据（日志、监控信息、应用信息等），通过机器学习的方式来进一步解决自动化运维没办法解决的问题”

AIOps与ClickHouse的碰撞

AIOps与ClickHouse的碰撞



AIOps与ClickHouse的碰撞

智能运维

大数据运维

数据化运营

问题分析

精细化运营

成本优化

服务自治

问题关联

问题发现

问题定位

AIOps与ClickHouse的碰撞

数据存储

- 海量数据存储
- 吞吐量10W以上
- 成本可控

数据检索

- 秒级查询
- 快速聚合
- 功能丰富

数据观察

- 图表丰富
- 可定制化
- 可交互

异常报警

- 灵活
- 通用
- 准确

异常分析

- 根因分析
- 关联分析

AIOps与ClickHouse的碰撞

数据存储

数据检索

数据观察

异常报警

异常分析

- Isolation Forest
- K-means
- DBScan
- DTW
- STL

- 关联规则
- 全链路系统

目录

AIOps的落地

落地的思路

工具的定位

Data science produces insights 数据科学产生洞见

Machine learning produces predictions 机器学习做出预测

Artificial intelligence produces actions 人工智能生成行为

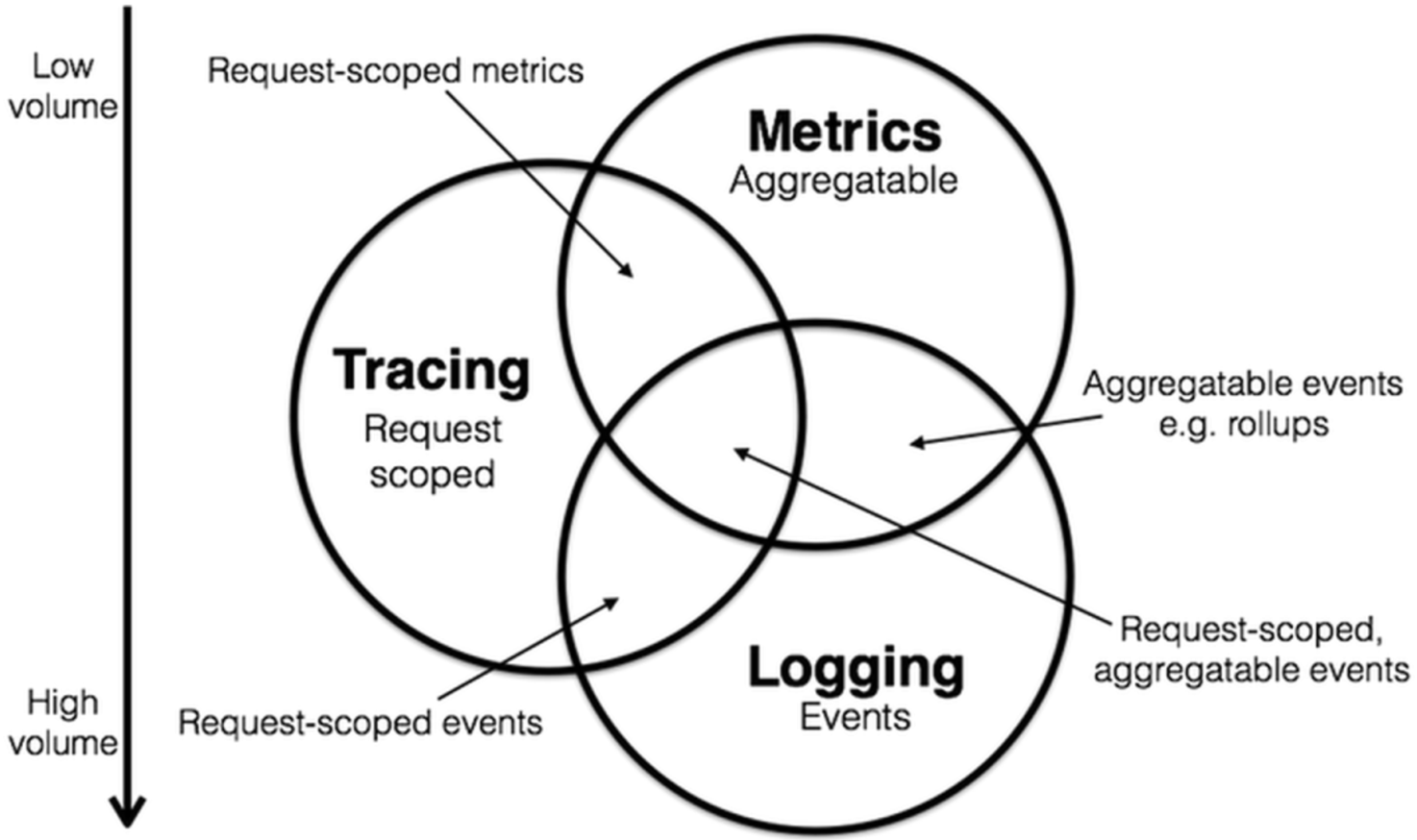
What's the difference between data science, machine learning, and artificial intelligence?

http://varianceexplained.org/r/ds-ml-ai/

AIOps的落地

落地的思路

软件的“可观测”性



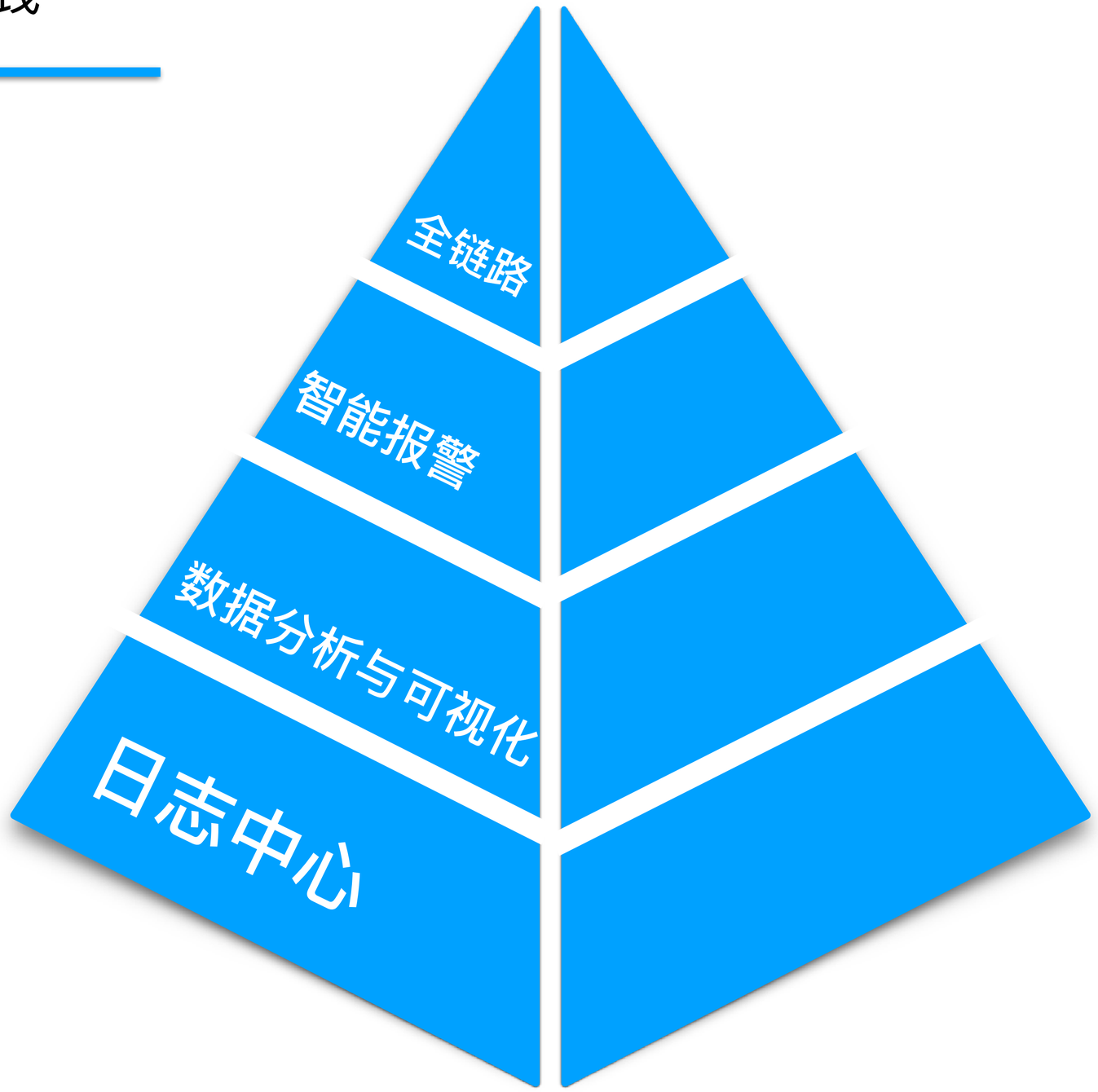
Peter Bourgon对于Metrics, tracing, and logging的分析

https://peter.bourgon.org/blog/2017/02/21/metrics-tracing-and-logging.html

AIOps的落地

落地的实践

构建之路



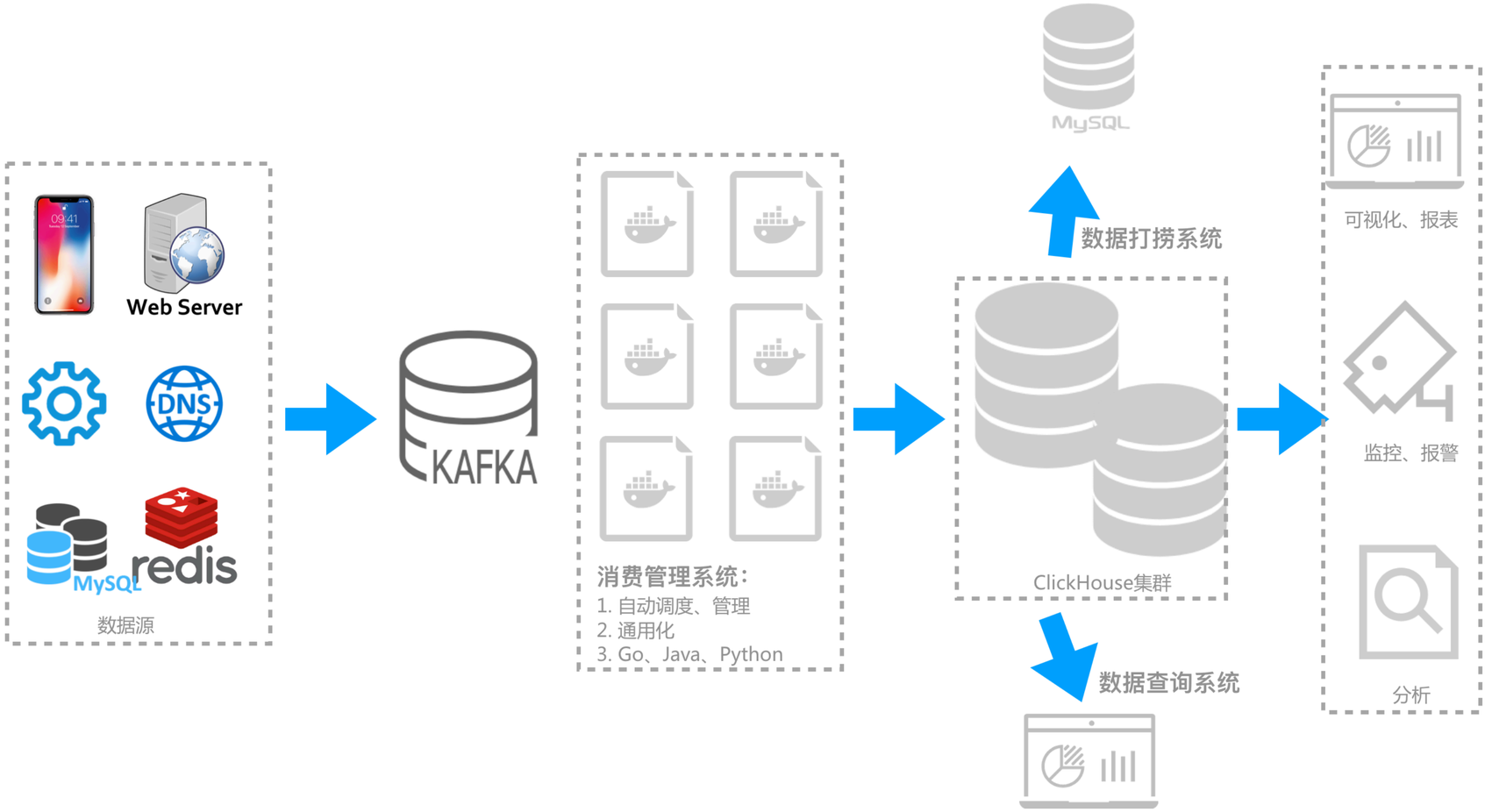
AIOps的落地

日志中心

数据接入

存储

分析



ClickHouse数据架构

AIOps的落地

日志中心

对用户友好的数据查询工具

AIOps的落地

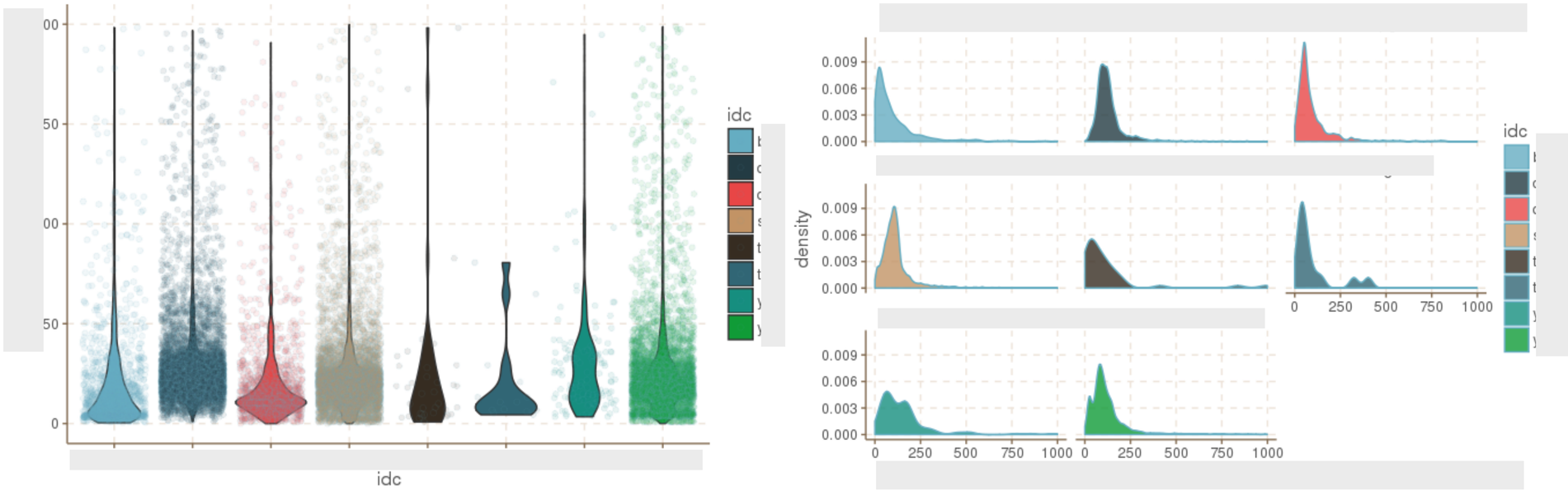
日志中心

对用户友好的数据查询工具

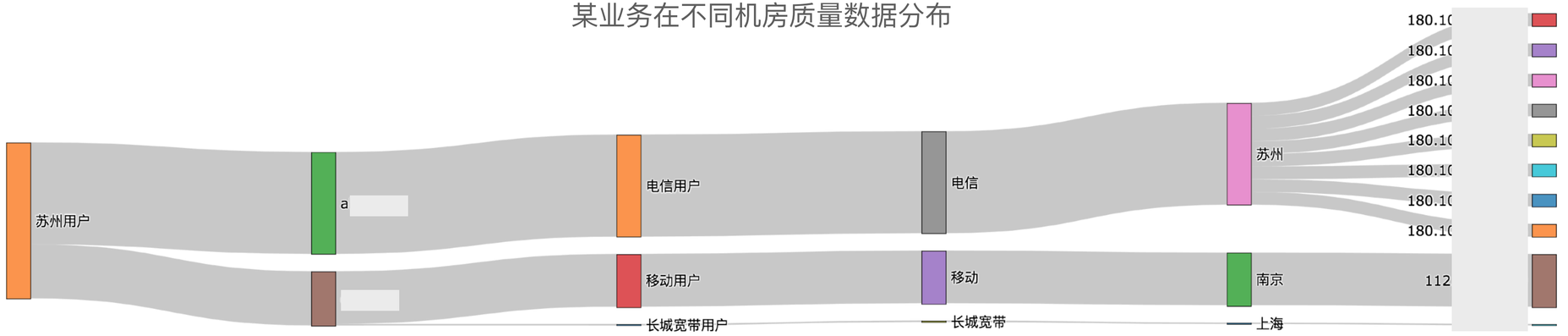
AIOps的落地

数据分析与可视化

某产品数据分析样例



某业务在不同机房质量数据分布

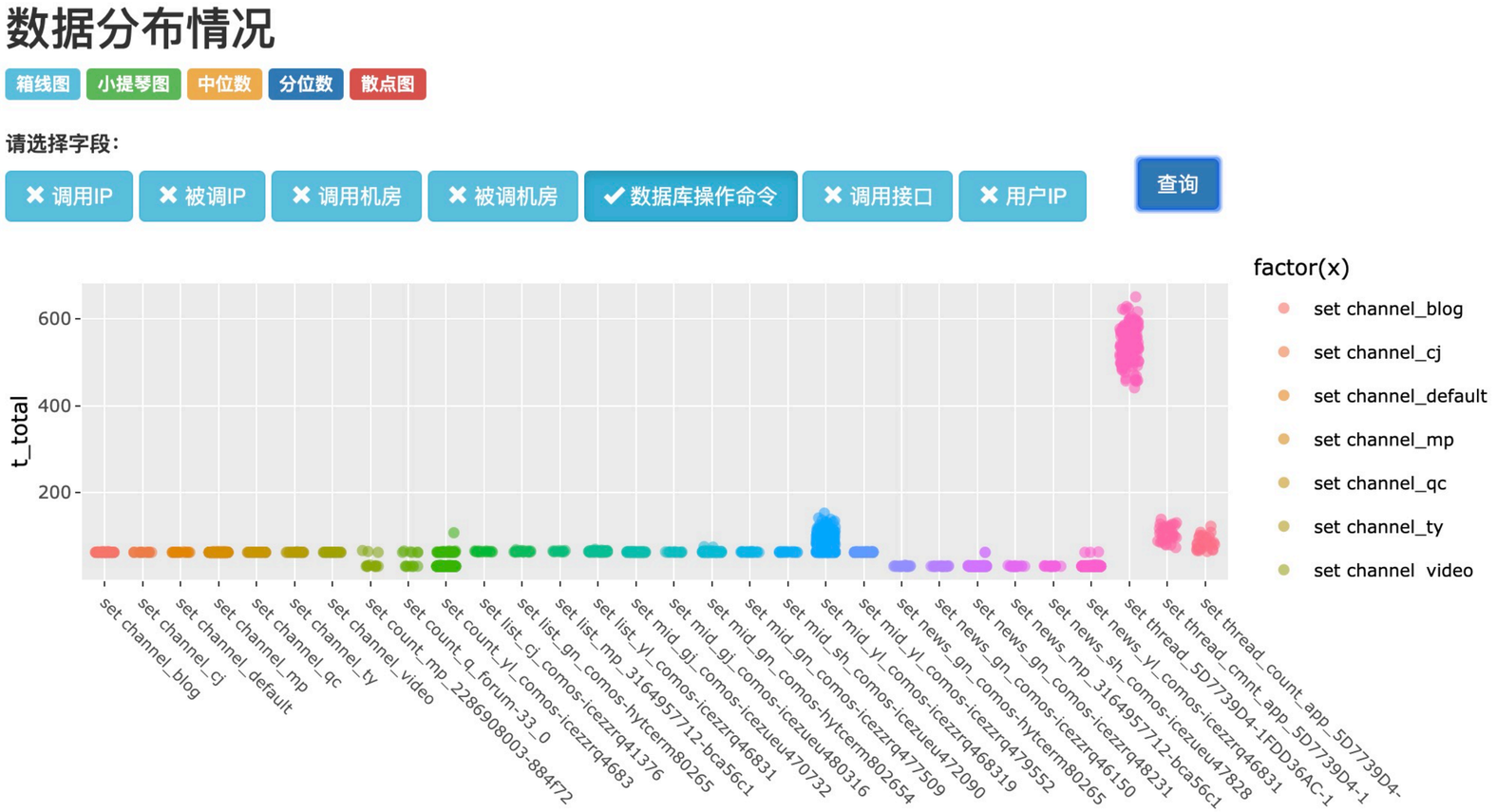


某地用户CDN流量分布

AIOps的落地

数据分析与可视化

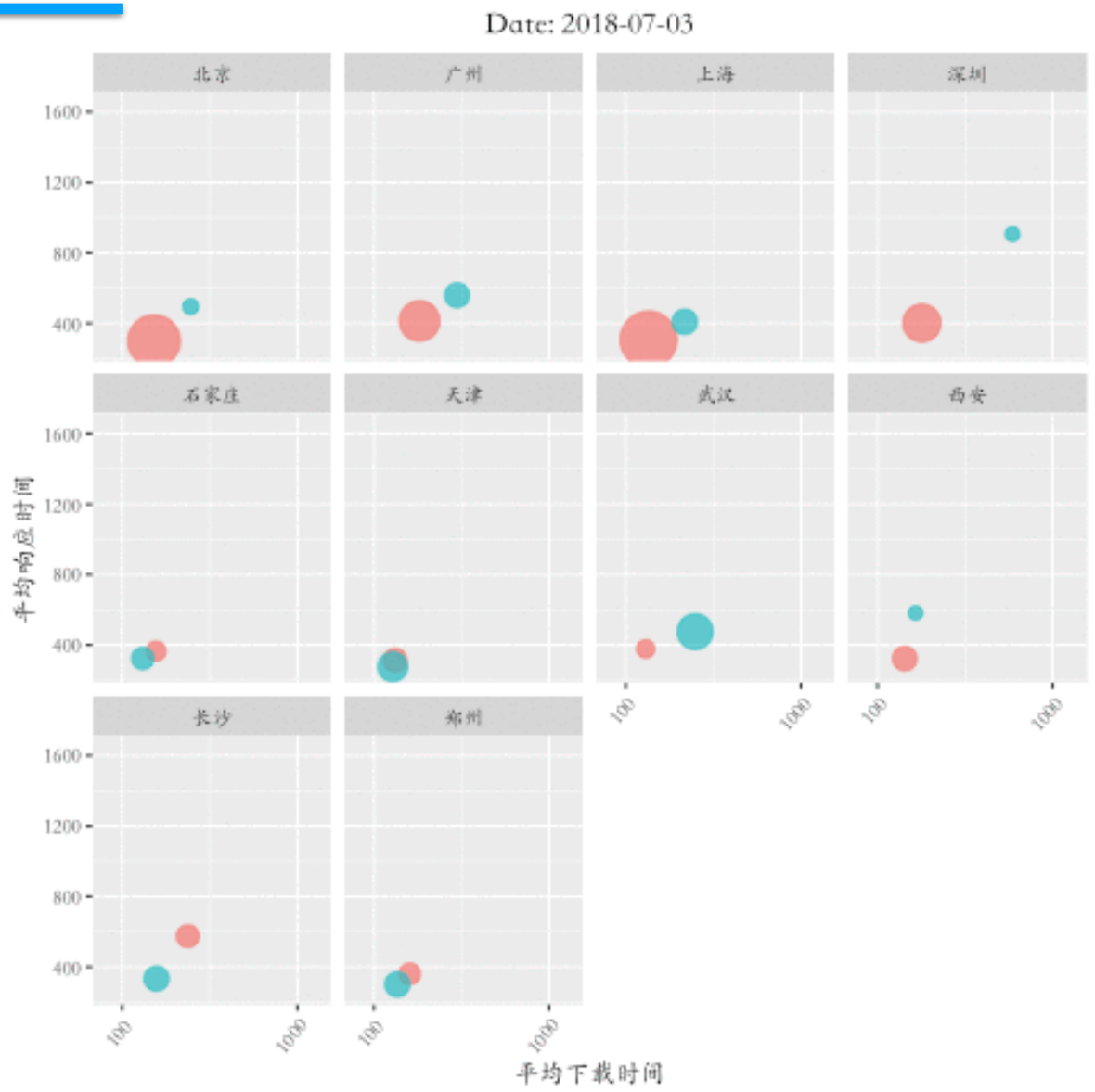
某mc请求响应时间分布图



AIOps的落地

数据分析与可视化

某APM产品示例



某CDN质量数据 gif动图

AIOps的落地

交互式分析工具

plotly | Dash

客户端APM数据回捞

选择观察对象、APP、时间，依次回捞用户的性能日志和错误日志

chose:

高鹏-新闻

选择用户:

新闻客户端

选择日期:

2019-09-09

起始时间:

00 00 00

结束时间:

23 59 59

查询

访问日志总量:	访问错误总量:	移动端错误日...	HTTP错误日志...
524	0	0	0

访问慢请求占...	慢请求中不知...	慢请求中下载...
3.82%	2.86%	0%

下列信息完全一样，表格中不再显示：

```
用户LocalDNS :  2408:8000:1010:1::8,2408:8000:1010:2::8,123.123.123.123
定位信息 :  none_none
上传耗时(ms) :  0
服务器响应耗时(ms) :  0
其他时间(ms) :  0
手机信息 :  ios_apple_iphonex_7.21.5_1.6.7
```

Show 30 entries Searc

表：用户正常访问（含慢请求）详情（各字段均可过滤、筛选）

时间戳	用户信息	访问信息	响应时间(ms)	DNS解析(ms)	TCP建联(ms)	SSL认证(ms)	下载时间(ms)	请求大小(KB)	Url
	All	All							All
2019-09-09 09:59:34	北京_114.242.249.60_联通_mobile	newsapi.sina.cn_123.125.29.208_北京联通	6154	6046	38	0	2.6	0.1	http://newsapi.sina.cn/?resource=activity/common&accessToken=2.0(iphonex&from=6000093012&idfa=9B9A2591-Aiphonex__SinaNews__7.21.5__iphone__12.4.
2019-09-09 09:59:35	北京_114.242.249.60_联通_mobile	newsapi.sina.cn_123.125.29.208_北京联通	6126.5	7	28	0	1.2	0.1	http://newsapi.sina.cn/?resource=register&accessToken=2.00w22dxBeiphonex&from=6000093012&idfa=9B9A2591-Aiphonex__SinaNews__7.21.5__iphone__12.4.

新闻客户端图片访问情况

选择覆盖率范围

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

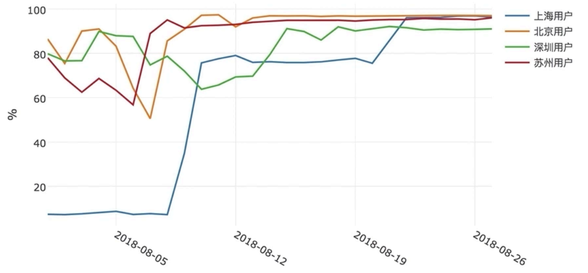
选择日访问量范围

0M 2M 4M 6M 8M 10M 12M 14M 16M 18M 20M

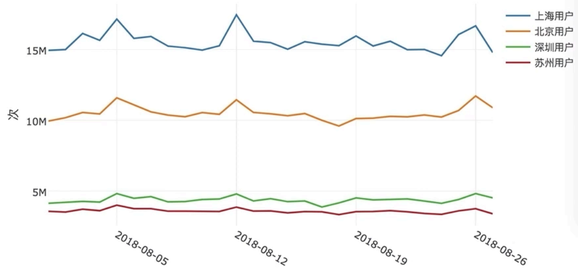
选择日期

2018-08-01 2018-08-02 2018-08-03 2018-08-04 2018-08-05 2018-08-06 2018-08-07 2018-08-08 2018-08-09 2018-08-10 2018-08-11 2018-08-12 2018-08-13 2018-08-14 2018-08-15 2018-08-16 2018-08-17 2018-08-18 2018-08-19 2018-08-20 2018-08-21 2018-08-22 2018-08-23 2018-08-24 2018-08-25 2018-08-26 2018-08-27

本地化比例走势



本地访问量走势



AIOps的落地

Why智能报警

传统监控的困境：

1. 数据太深：同一维度，不同业务，差异巨大，导致添加报警繁琐无比

2. 数据太个性：不同时间范围，波动巨大，导致整体阈值无法“一刀切”

3. 数据太宽：维度众多，出了问题不知道是什么导致的，导致报警只是“吹哨”

AIOps的落地

Why智能报警

脏值过滤
周期判定

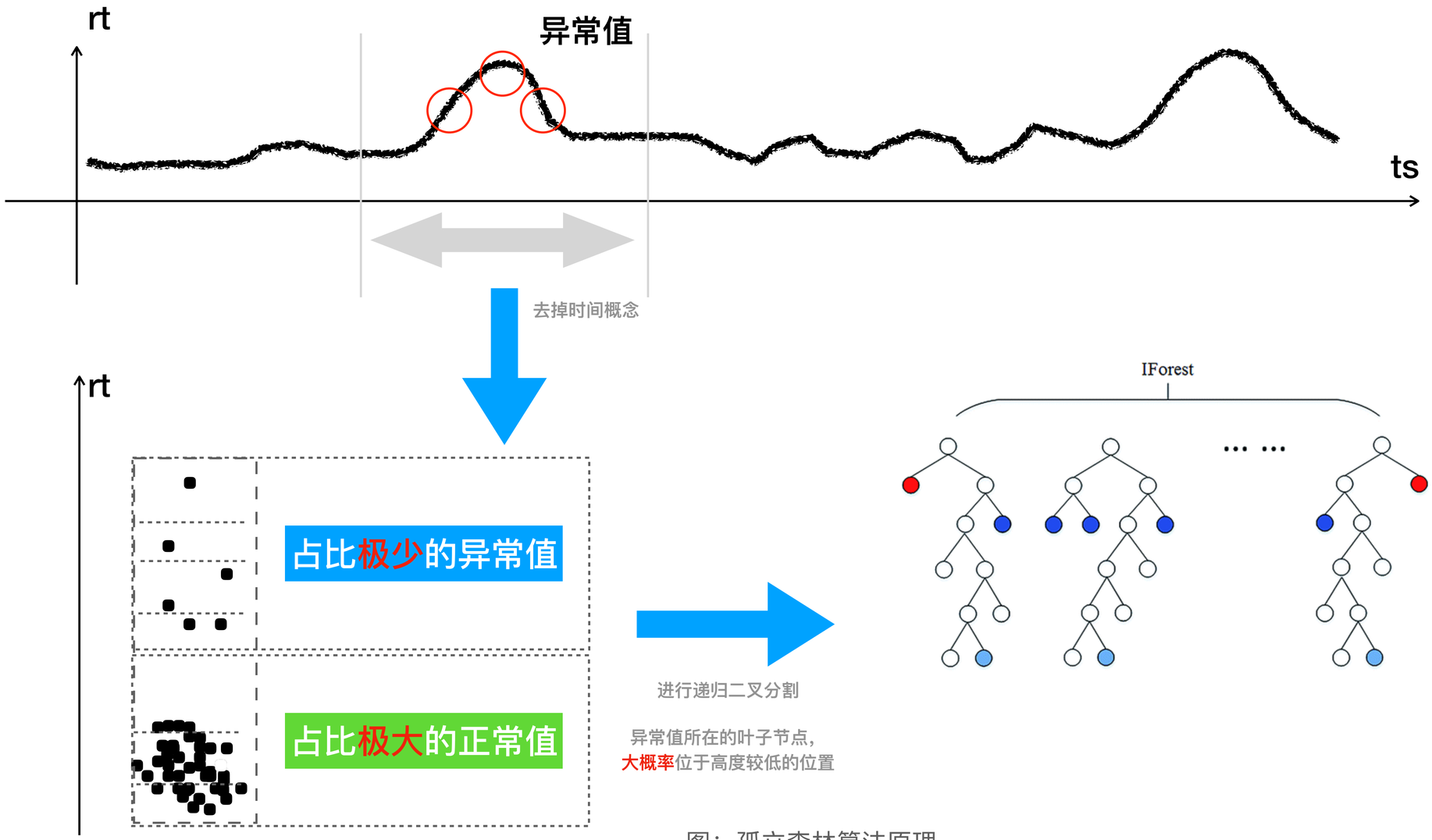
模型训练
模型应用

特征提取
根因聚类

AIOps的落地

异常检测

孤立森林算法原理



图：孤立森林算法原理

AIOps的落地

异常检测

不同业务

阈值个性化

告警通知

2018年8月27日 星期一

来自「新浪数据分析平台」的报警

告警内容：新浪数据分析平台:Dpool响应时间异常告警-
[redacted]a.com.cn域名3s以上占比异常
基准值:0.47%
最近3分钟均值:41.45%

告警发生时间：2018-08-27 20:04:09

请尽快处理

告警通知

2018年8月27日 星期一

来自「新浪数据分析平台」的报警

告警内容：新浪数据分析平台:Dpool响应时间异常告警-
[redacted]nce.sina.com.cn域名3s以上占比异常
基准值:0.06%
最近3分钟均值:8.12%

告警发生时间：2018-08-27 20:05:05

请尽快处理

从业务层面，就天然存在差异
左边的，平时有0.5%的异常
右边的，平时只有0.06%的异常

AIOps的落地

根因分析

- **传统方案**
  - 报警只负责“值”，不管“因”
  - 人工排查原因，效率低下
  - 举例：新闻APP图片404数量过万，是什么原因?
    - A. 某个CDN运营商出了问题
    - B. 某个地区运营商出了问题
    - C. 某个客户端版本出了Bug
    - D. 某个渠道用户出了问题
    - E. 某个类型的域名出了问题
    - F. 某个CDN运营商的某个IP出了问题
    - G. ……
  - 如果有100个维度，那么复杂度是：

$$C_{100}^{1} + C_{100}^{2} + C_{100}^{3} + \cdots\cdots$$

- **根因分析算法原理**
  - 关联规则算法：啤酒与尿布

根因分析Apriori算法示意图

AIOps的落地

根因分析

报警原因，直截了当

新浪数据分析平台:BIP_ALERT-新闻客户端服务端5XX错误监控 Value: 5364.0(次/min)
数据特征如下:
<domain=newsapi.sina.cn> 占比88.22%
<httpCode=504> 占比74.38%
<domain=newsapi.sina.cn,httpCode=504> 占比65.83%
<serverip=web009.newsapp.[illegible].sinanode.com> 占比41.35%
<domain=newsapi.sina.cn,serverip=w[illegible]ewsapp.msina.dbl.sinanode.com> 占比37.3%
<httpCode=504,serverip=web009.newsapp.msina.dbl.sinanode.com> 占比30.98%

time: 2018-09-02 15:29:12【新浪】

Value: 1838.0(次/min)
数据特征如下:
[_http_code=404] 占比98.1%
[CDN_ISP=edge] 占比44.02%
[CDN_ISP=edge,_http_code=404] 占比43.14%
[CDN_ISP=alicdn] 占比42.55%
[CDN_ISP=alicdn,_http_code=404] 占比42.49%
[_http_code=404,_request_url=http://l.sinaimg.cn/n/front/384/w230h154/20180716/N8vH-hfkffak7381394.jpg/w345h231t0l50q75z1apl.webp] 占比34.0%

time: 2018-07-20 09:02:27

AIOps的落地

全链路分析

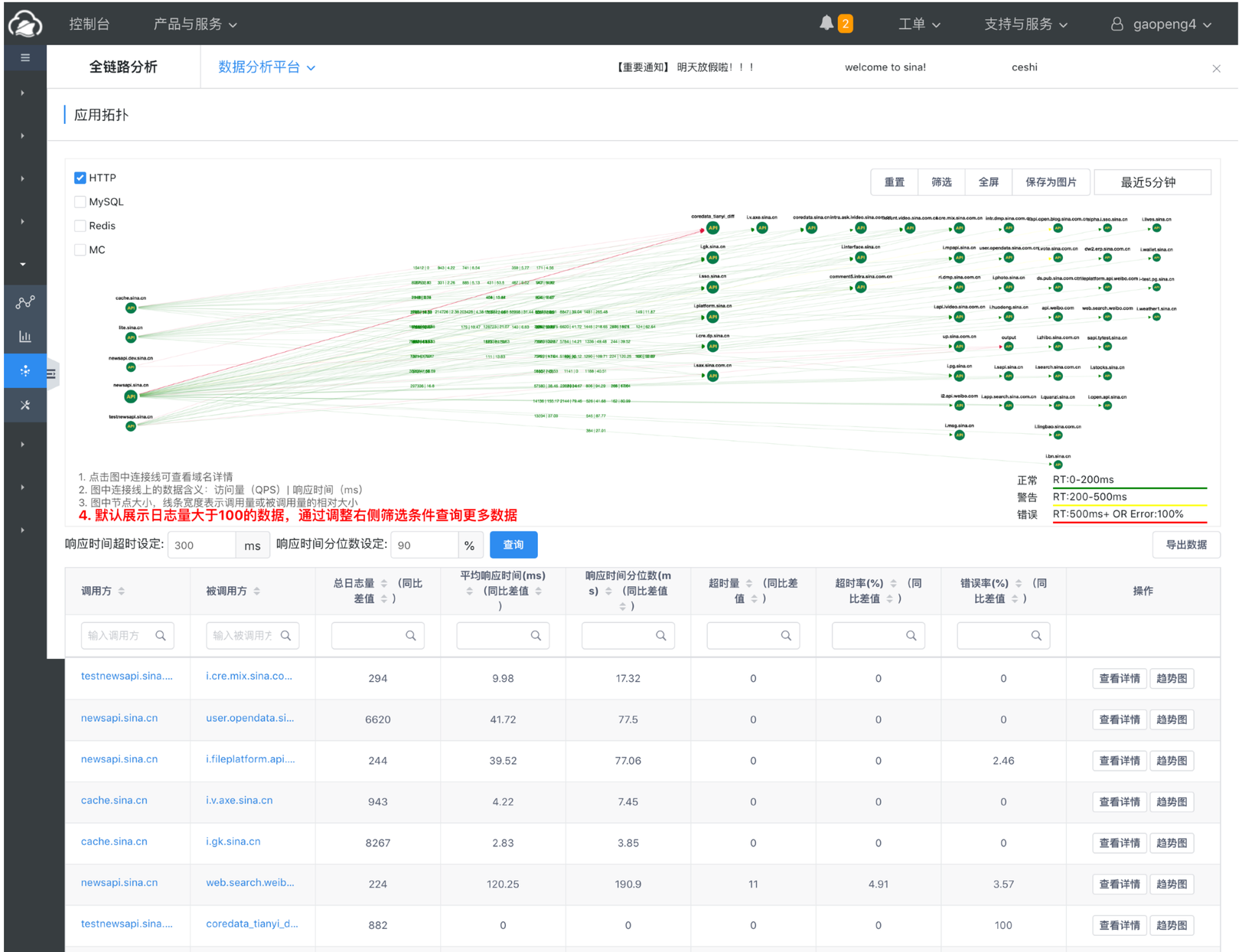
单一局部监控

无法形成体系

AIOps的落地

全链路分析

业务拓扑



AIOps的落地

全链路分析

多维分析

调用方、被调用方
可自定义超时条件、分位数条件

AIOps的落地

全链路分析

报警关联分析

ID	310723676
产品线	新浪数据分析平台
Object	dpool重点域名超时（3s）占比监控<domain=tousu.sina.com.cn>
Subject	: 当前值: 29.91 % (1min内) 基准值: 0.05 % (1min内) 当前值是基准值的546.69倍 报警时间: 2021-03-17 06:48:00. 已持续触发 <5> 分钟 数据信息见详情
正文	报警定制信息: 请求次数: 24444 报警特征信息: [idc= 占比 99.99% [upstream_ip= 6:80] 占比 37.21% [ha_ip=172.1] 占比 33.9% [ha_ip=172.1] 占比 33.23% [ha_ip=172.1] 占比 32.86% [api=/api/company/received_complaints] 占比 29.3%
时间	2021-03-17 06:49:43

ID	310739654
产品线	新浪数据分析平台
Object	全链路-tousu项目Mysql响应时间监控-Dpool<调用关系=tousu.sina.com.cn调用s34 na.com.cn沙溪>
Subject	: 当前值: 3731.3 ms (1min内) 阈值: 1000.00 ms (1min内) 当前值是阈值的3.73倍 报警时间: 2021-03-17 07:02:00. 已持续触发 <15> 分钟 数据信息见详情
正文	报警定制信息: 请求次数: 171 报警特征信息: [被调接口=mysql://s346 m.cn:3460/tousu/utf8] 占比 100.0% [调用域名=tousu.sina.com.cn] 占比 100.0% [调用机房=] 占比 100.0% [调用接口=http://tousu.sina.com.cn/api/company/received_complaints] 占比 83.78% [资源请求命令=SELECT uid FRO WHERE `is_del` = ? AND `p_uid` = ?] 占比 66.41%
时间	2021-03-17 07:03:59

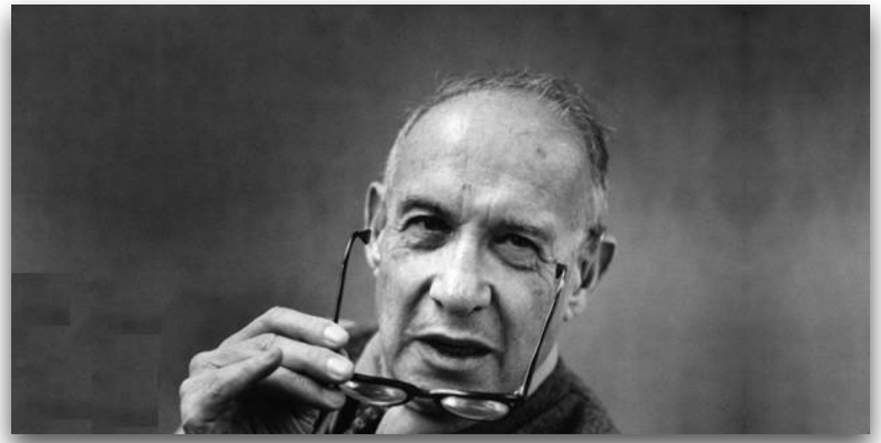
目录

探讨

1. ClickHouse：RDBMS or Engine
2. ClickHouse存储计算分离

1. AIOps的前提是BigdataOps
2. AIOps的后置是自动化、标准化
3. 数学上的异常，以实际情况看来，可能在容忍范围内
4. 算法存在不可解释的情况，难以向业务说明
5. 大样本下实现链路串联，依旧比较困难

“If You Can’t Measure It, You Can’t Improve It”



彼得·德鲁克（Peter F. Drucker，1909.11.19~2005.11.11），现代管理学之父