滴滴出行海量数据场景下的智能监控与故障定位实践

李培龙

滴滴
滴滴一下 美好出行

2017.12

极客时间

重拾极客精神·提升技术认知

下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新

扫一扫下载极客时间App

主办方 Geekbang. 极客邦科技 InfoQ

AiCon

全球人工智能与机器学习技术大会

助力人工智能落地

2018.1.13 - 1.14　北京国际会议中心

扫描关注大会官网

个人简介

- 2015年加入滴滴
  - 质量架构部负责人
  - 主要工作
    - ✓ 分布式调用链追踪系统、问题定位系统
    - ✓ 日志服务平台、智能异常检测系统
    - ✓ 全链路压测平台
- 之前供职于百度
  - 监控与问题定位系统
  - 性能测试与分级发布平台

背景

✦ 海量指标的产生

- ✧ 微服务化&云化：监控指标量级提升约100倍
- ✧ 指标维度增加：组合爆炸
- ✧ 单机平均指标：约10000

✦ 关键技术挑战

- ✧ 计算与存储
- ✧ 异常检测
- ✧ 故障定位

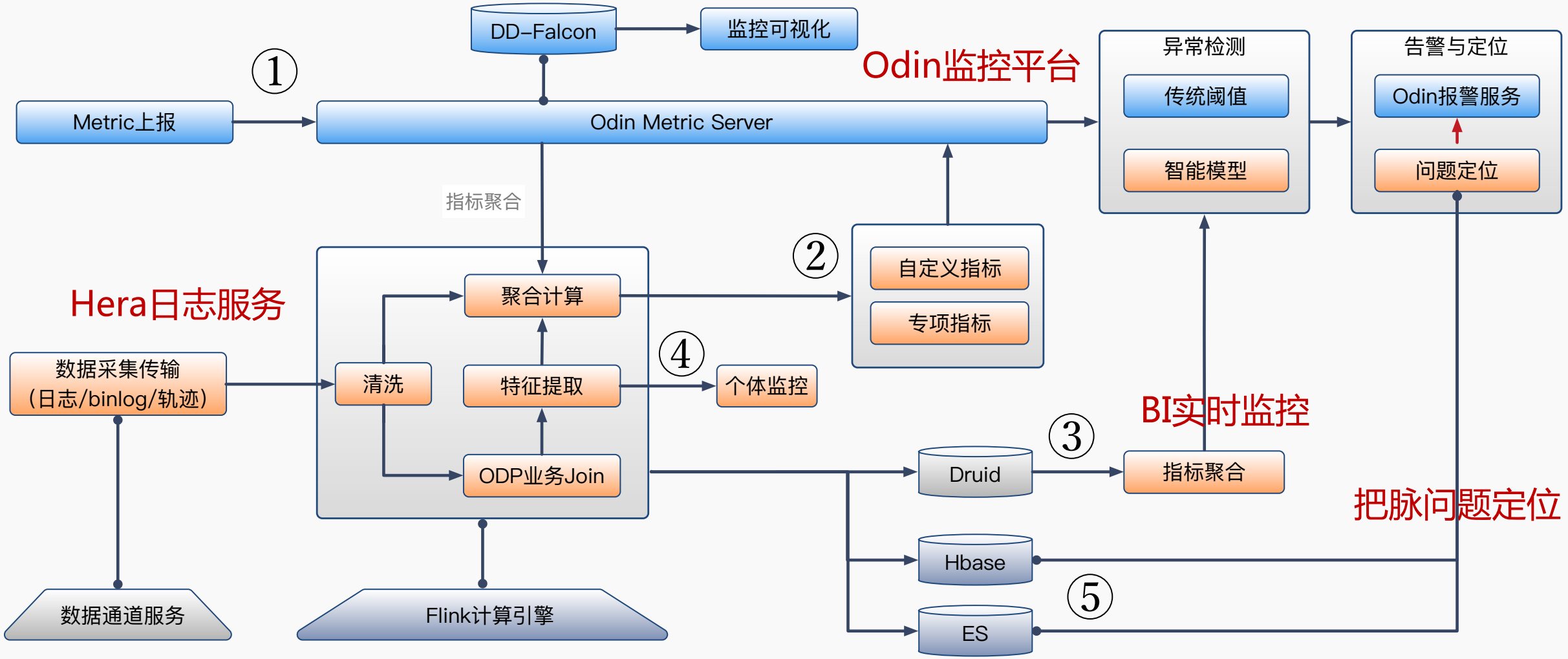
内容提纲

一、监控架构

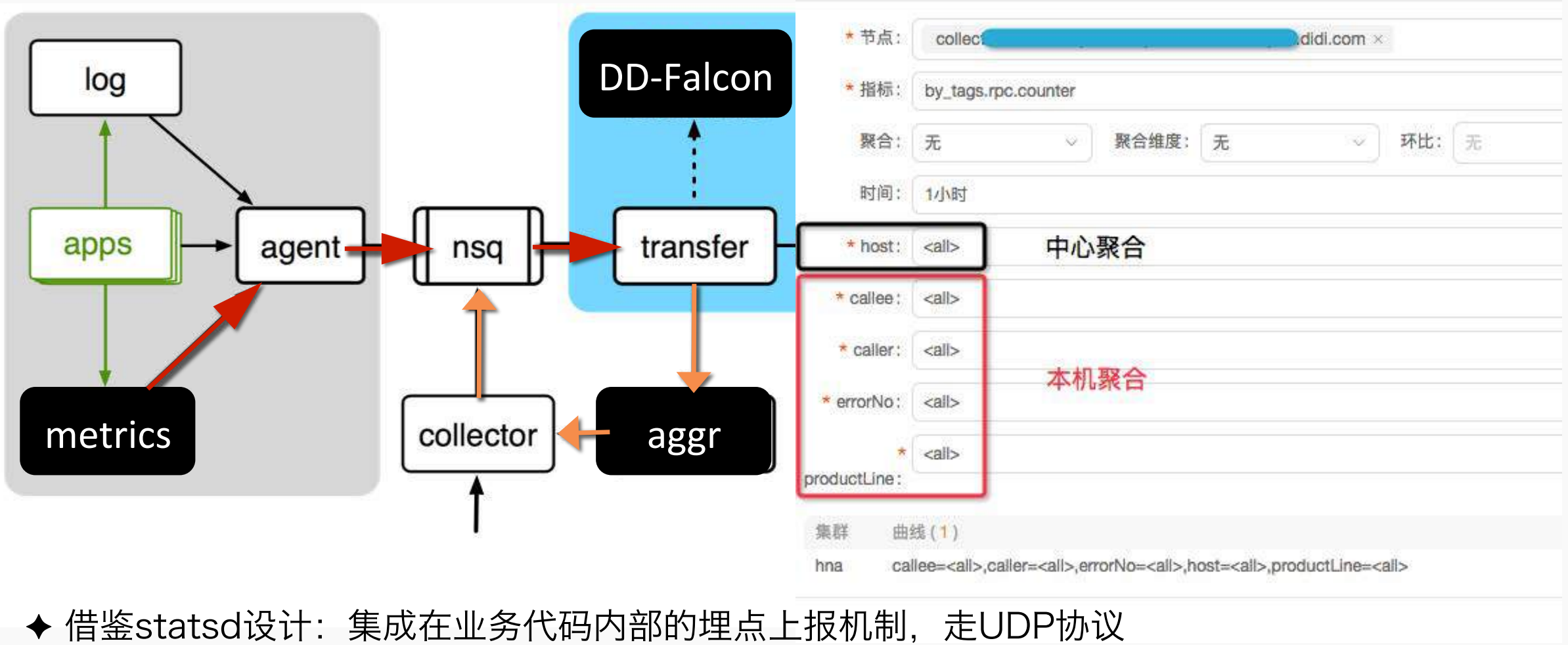
二、异常检测

三、快速定位

滴滴-监控系统概览

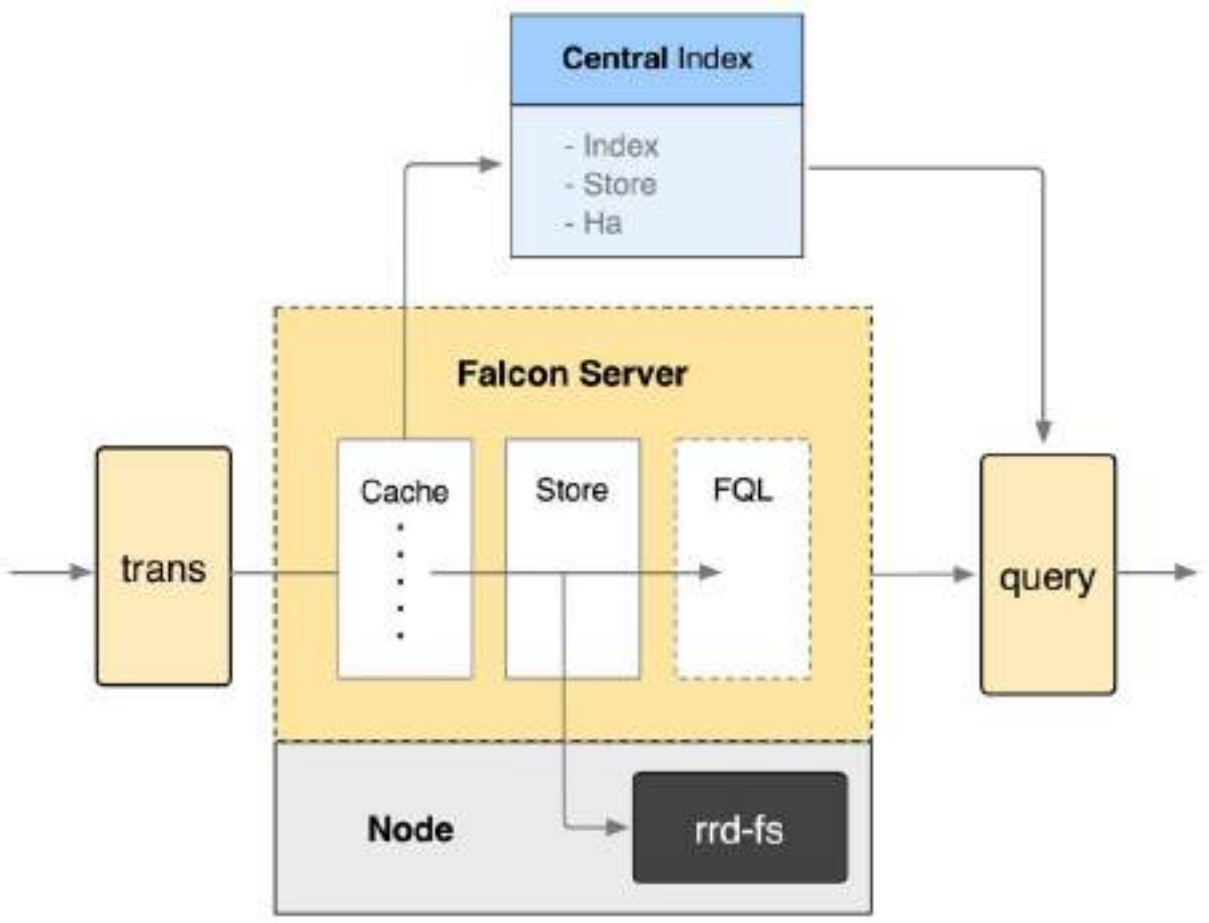

DD-Falcon
监控可视化
①
Odin监控平台
异常检测
传统阈值
智能模型
告警与定位
Odin报警服务
问题定位
Metric上报
Odin Metric Server
指标聚合
②
自定义指标
专项指标
Hera日志服务
聚合计算
数据采集传输
(日志/binlog/轨迹)
清洗
特征提取
④
个体监控
BI实时监控
③
ODP业务Join
Druid
指标聚合
把脉问题定位
Hbase
⑤
ES
数据通道服务
Flink计算引擎

① Metric通路



- 借鉴statsd设计：集成在业务代码内部的埋点上报机制，走UDP协议
- 本机agent聚合：10s粒度聚合，以及维度聚合
- Server端中心聚合：机器粒度聚合

① Metric通路：DD-Falcon时序数据存储



实时降采样

- rrdtool, 写入时 即完成降采样(平衡读写能力)
- 提高 长时间跨度 时的读效率

冷热分离

- 索引与数据分离, 分级索引, 优化索引查询
- 缓存10分钟最新数据, 优化即时查询

数据清洗

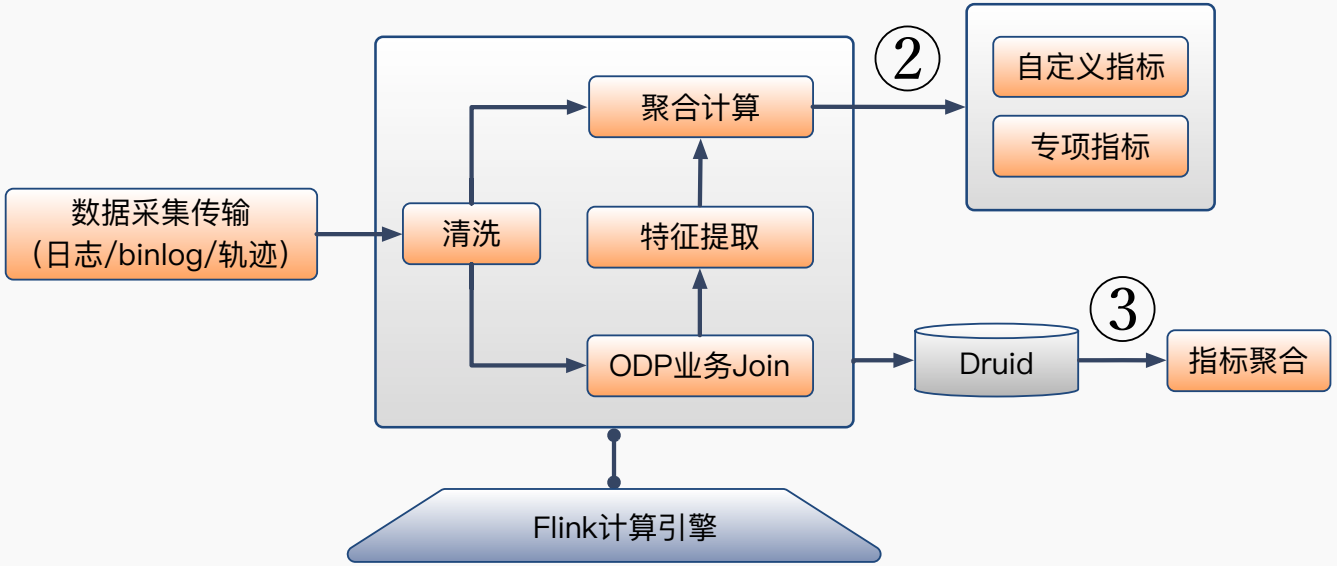
- 通过容量控制, 兜底
- 通过多维度自动检测, 主动发现、过滤非ts数据

磁盘读写优化等(由Open-Falcon提供)

②+③： 日志计算通路

基于流式计算的指标聚合

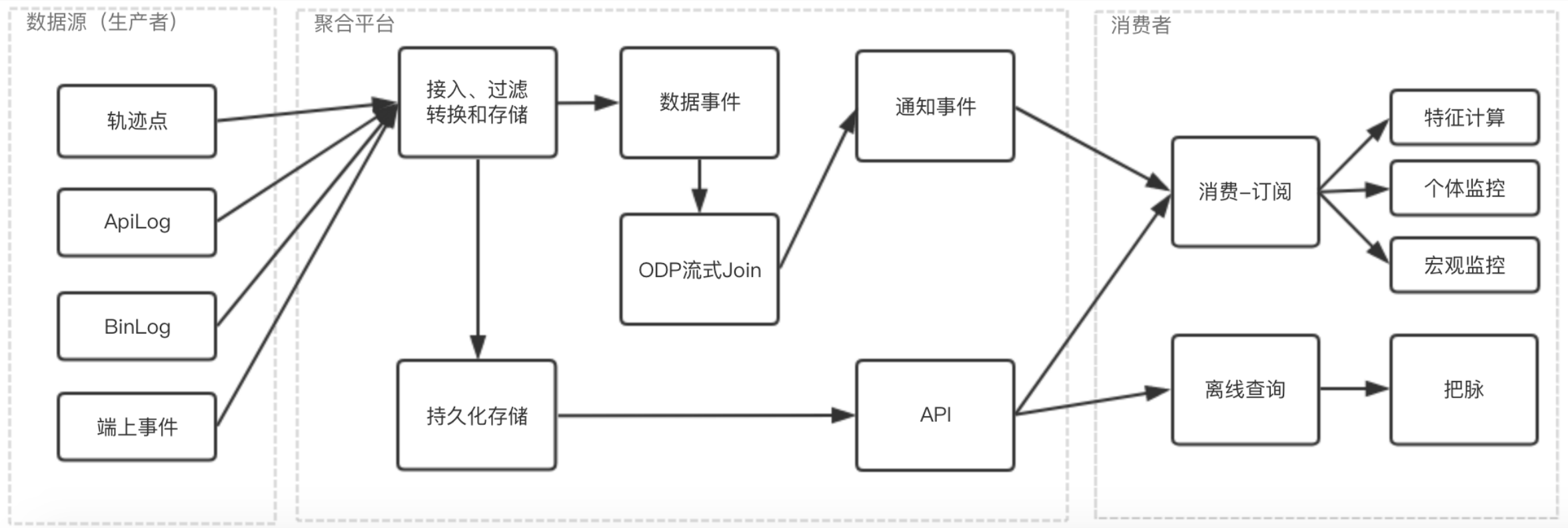
- 日志在Flink中完成ETL、Join、聚合，仅存聚合指标
- 提供类SQL的流式计算配置化服务



基于Druid存储的指标聚合

- 原始数据在Flink完成ETL、Join
- 原始指标数据存入Druid
- 借助Druid的预聚合及计算能力实现监控指标聚合

④ ODP数据Join介绍



- 接入数据在存储后转换为数据事件，参与流式Join生成通知事件
- 实时：订阅通知事件触发特征查询和特征计算、监控
- 离线：把脉问题定位-离线数据使用

二、异常检测：背景

- 海量指标的驱动
  - 迫使改变传统的人工配置模式，探索模型算法
  - 无监督学习，降低标注成本
- 问题定义
  - 核心指标：高准召率，基于标注训练或者人工精细化调参
  - 非核心指标：低成本接入，中准召率，无标注训练，冷启动，基于反馈自动调整
- 模型算法
  - 预测+异常判定

二、异常检测：我们经历的几个阶段

1. 人工配置

2. 单模型
（一阳指）

3. 多模型
（六脉神剑）

4. 通用模型
（独孤九剑）

阶段2（一阳指）：单模型—三阶指数平滑

- **预测**：三阶指数平滑（Holt-Winters）
  - 适用于有趋势和周期性的时序指标
  - 模型参数：α/β/γ，截距/斜率/周期平滑系数
  - 参数确定：
    - ✓ 人工配置
    - ✓ 自动训练：排除异常点→最大化拟合度

- **异常判定**：
  - 明确上下界：预测值±δ
  - 固定阈值
  - 历史周期点的指数平滑
  - 滑动窗口的偏差标准差

- Prediction:
  $\hat{y}_{t+1} = a_t + b_t + c_{t+1-m}$
- Baseline ("intercept"):
  $a_{t+1} = \alpha(y_{t+1} - c_{t+1-m}) + (1-\alpha)(a_t + b_t)$
- Linear Trend ("slope"):
  $b_{t+1} = \beta(a_{t+1} - a_t) + (1-\beta)b_t$
- Seasonal Trend:
  $c_{t+1} = \gamma(y_{t+1} - a_{t+1}) + (1-\gamma)c_{t+1-m}$

In addition to prediction, compute a measure of deviation for each time point: d_t

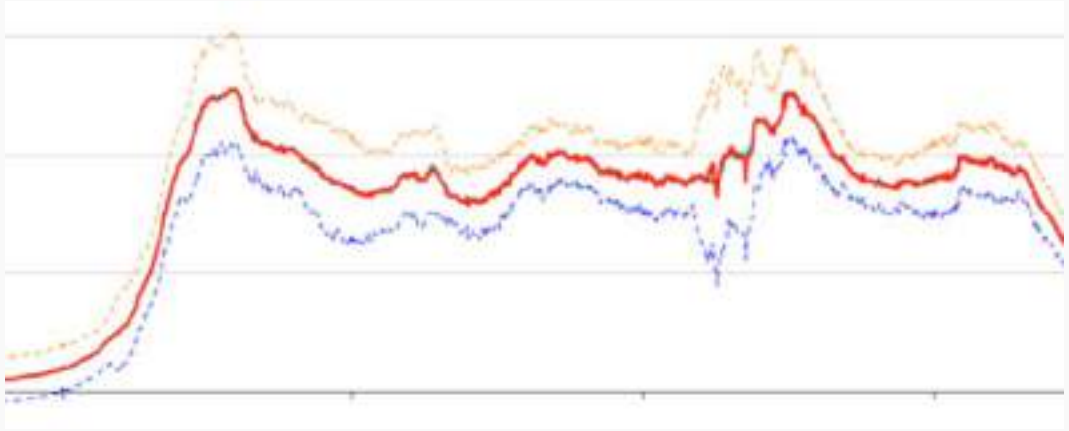
Use weighted average absolute deviation updated via exponential smoothing:

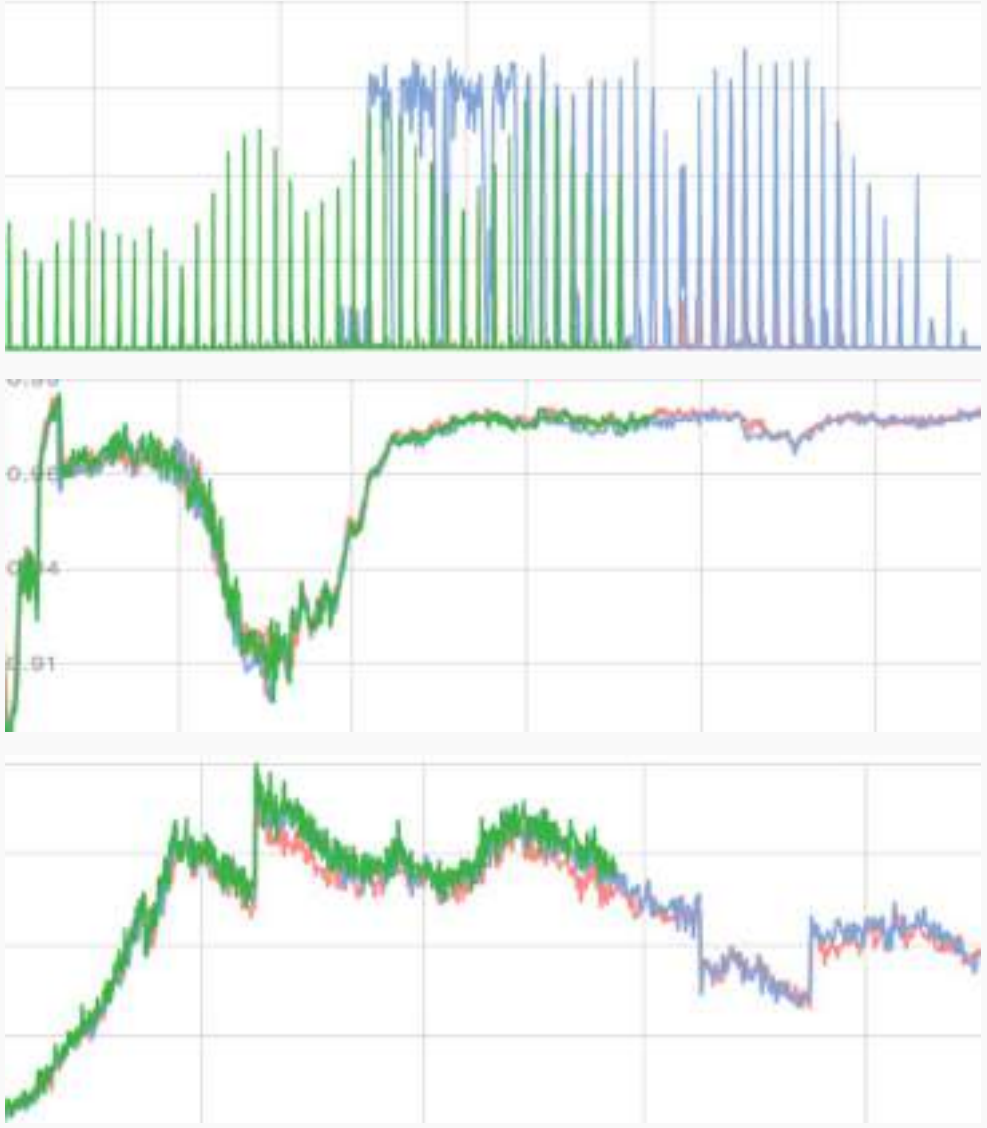
$$d_{t+1} = \gamma \cdot |y_{t+1} - \hat{y}_{t+1}| + (1-\gamma)d_{t+1-m}$$

Confidence bands: a collection of confidence intervals of the form:

$$(\hat{y}_{t+1} - \delta \cdot d_{t+1-m}, \hat{y}_{t+1} + \delta \cdot d_{t+1-m})$$

阶段2（一阳指）：单模型—三阶指数平滑

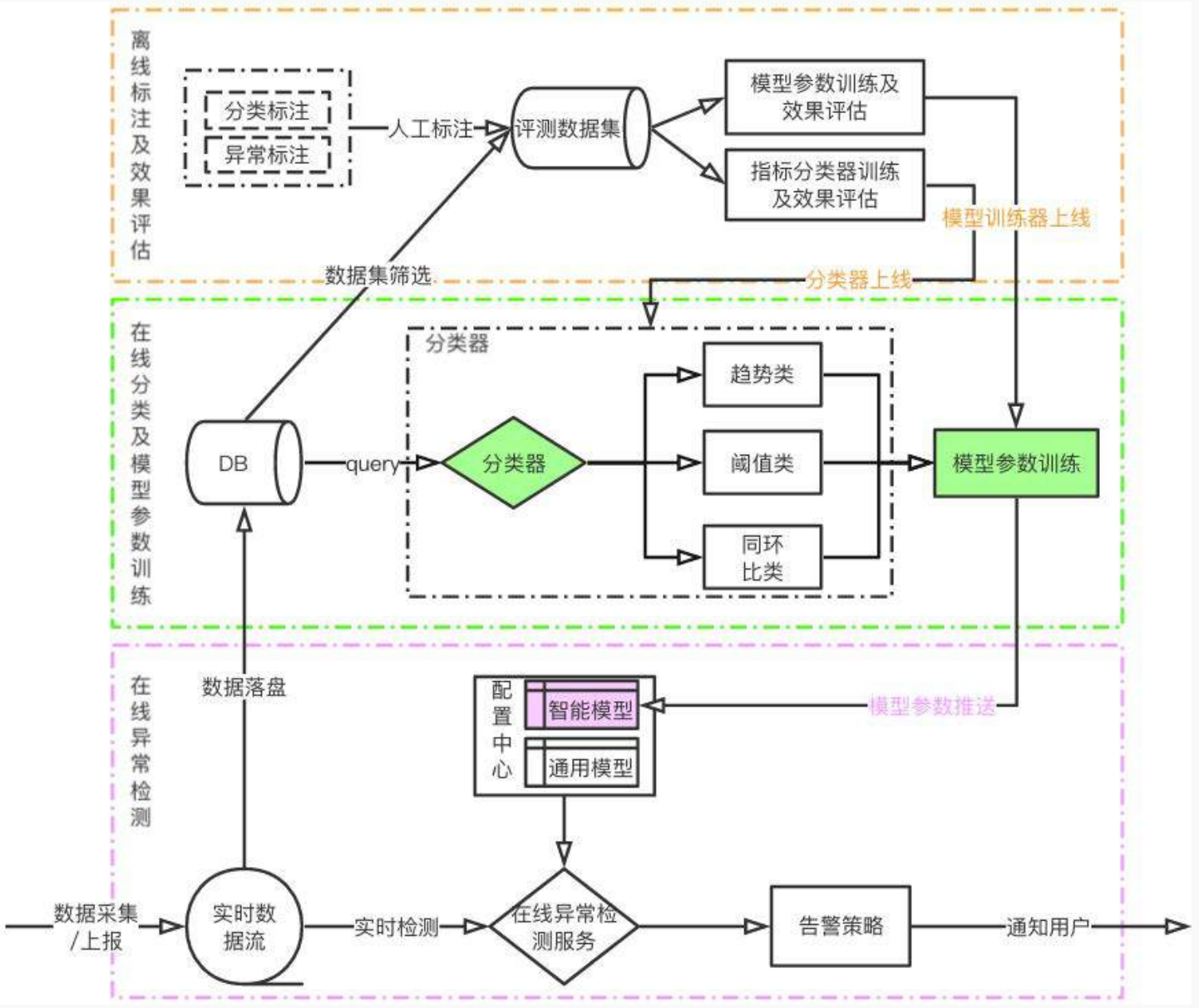




- 当前应用情况
  - 滴滴核心业务指标：百级别
  - 准召率90%+
- 适用场景及局限
  - 适用于稳定且有周期的指标
  - 指标需连续且无突增突降
  - 接入效率偏低

阶段3（六脉神剑）：多模型，分而治之

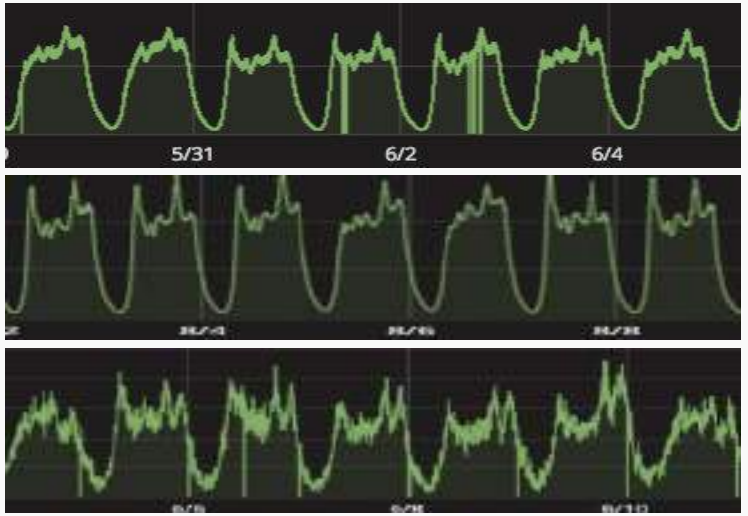
- ✦ 实现思路
  - ✧ 根据指标特征自动寻找合适模型
  - ✧ 自动选择模型参数
  - ✧ 目前支持类别
    - ✓ 阈值类/同环比/趋势类
- ✦ 当前应用及效果
  - ✧ 应用于线上万级别指标
  - ✧ 召回线上问题50+
  - ✧ 准确率约60%
  - ✧ 召回率约70%



阶段3（六脉神剑）：分类

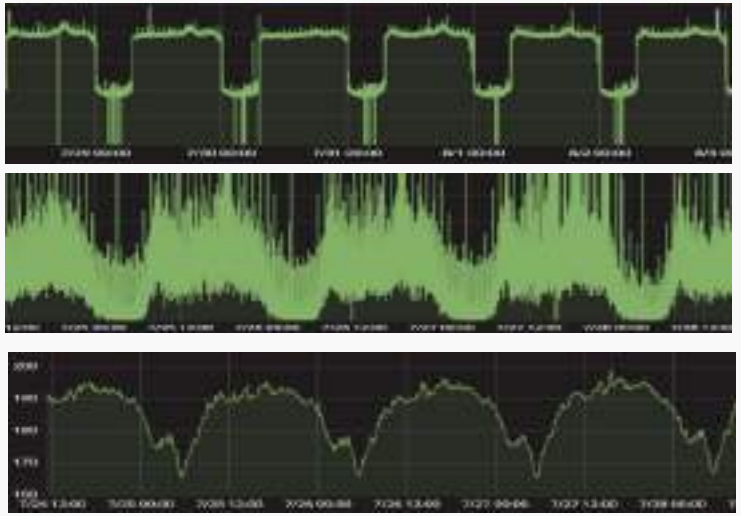
✦ 趋势类

- ✧ 多周期性
- ✧ 趋势性
- ✧ 高稳定，波动小
- ✧ 平滑，无突增突降



✦ 同环比类

- ✧ 有周期性
- ✧ 中低稳定，波动大
- ✧ 不平滑，有突增突降



✦ 动态阈值类

- ✧ 数值分布集中
- ✧ 成功率、时延等指标

阶段3（六脉神剑）：模型参数训练

动态阈值模型：

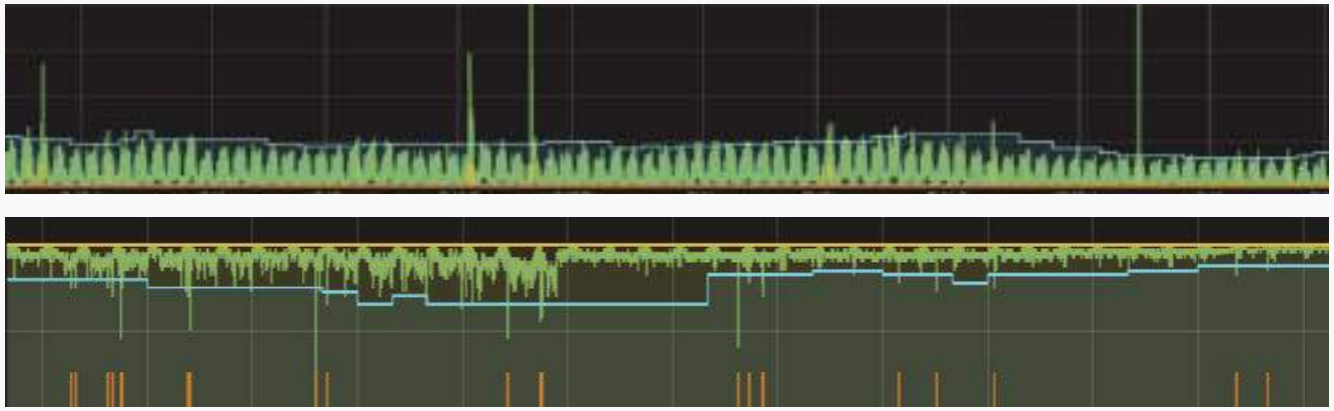
$$Split\ by\ Days:\ D_i,\ i \in [0, \ldots, N)$$

$$TopK: tk_i = top(sorted(D_i), K)$$

$$BotK: bk_i = bottom(sorted(D_i), K)$$

$$UpperLimit = max(LOFmax(tk_i))$$

$$LowerLimit = min(LOFmin(bk_i))$$



加权同环比模型：

$$f(x) = \omega_1 \cdot x_1 + \omega_2 \cdot x_2 + \cdots + \omega_7 \cdot x_7 + \delta$$

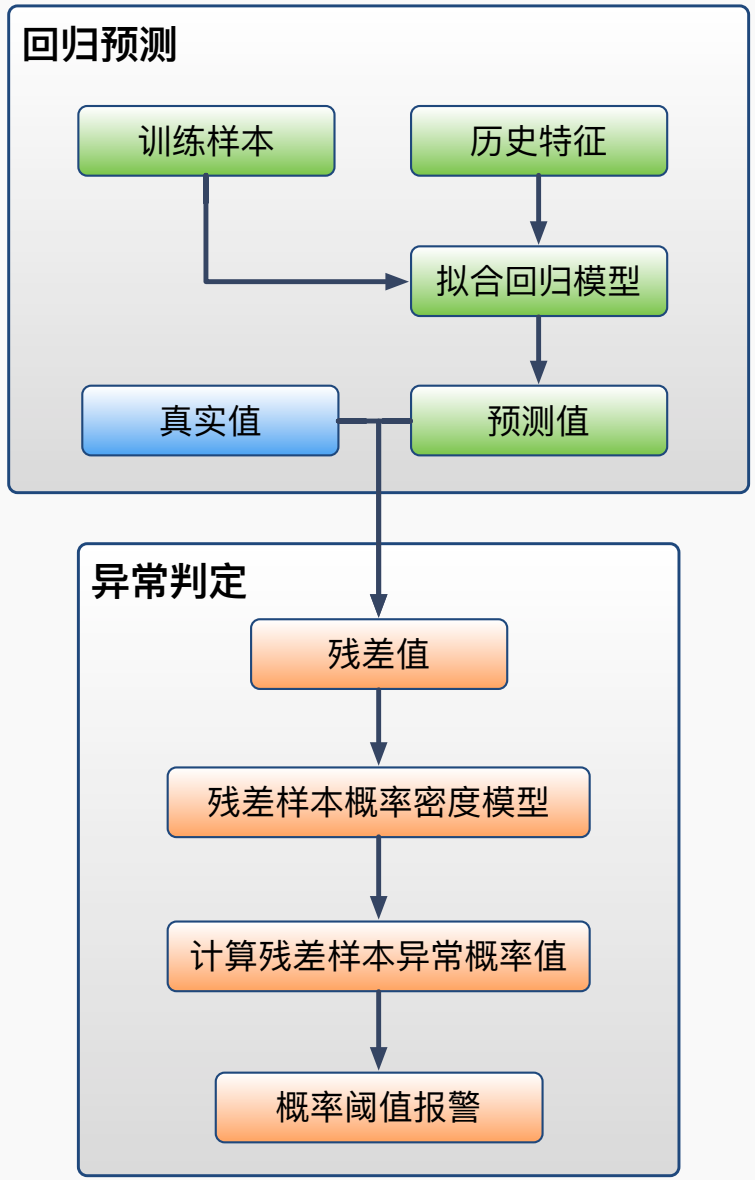
$$RidgeRegression + Standard$$

阶段3（六脉神剑）：分类模式的缺陷

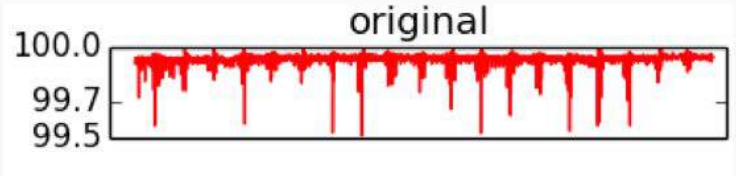
- 分类算法：合理性与准确性
  - 分类边缘指标与模型的适配性差
  - 分类覆盖不全：10%无法分类
- 模型选择及参数训练
  - 无标注场景下，参数训练较困难
  - 新模型研发成本高，周期长
- 算法架构
  - 不够灵活，落地略困难

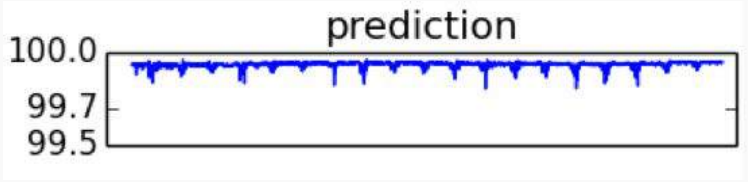
阶段4（独孤九剑）：Canary算法---普适性探索

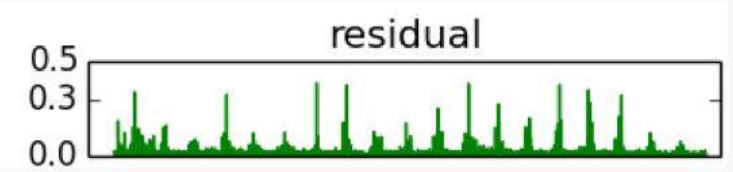
- ✦ 核心思路
  - ✧ 回到“预测+异常判定”的基本思路
  - ✧ 寻找普适性的回归预测模型，弥补HW缺陷
    - ✓ 特征的全面性
  - ✧ 异常判定：基于残差的概率密度建模
    - ✓ 默认阈值的选择
    - ✓ 实时标注反馈机制

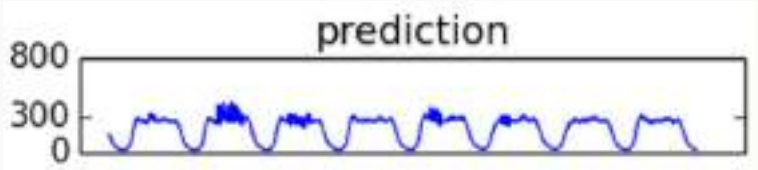


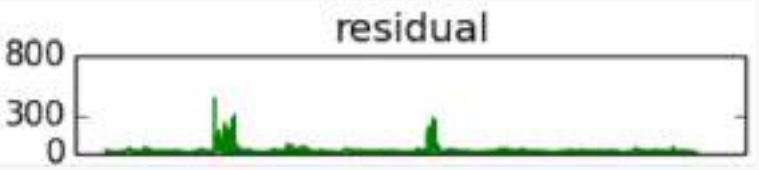
阶段4（独孤九剑）：Canary算法探索

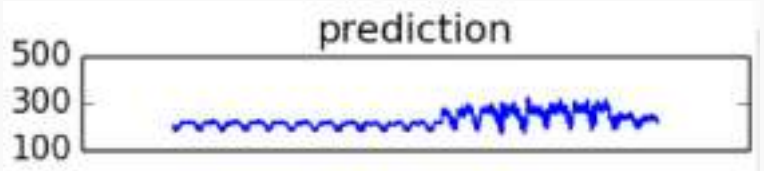


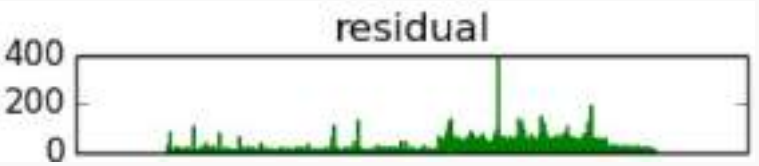






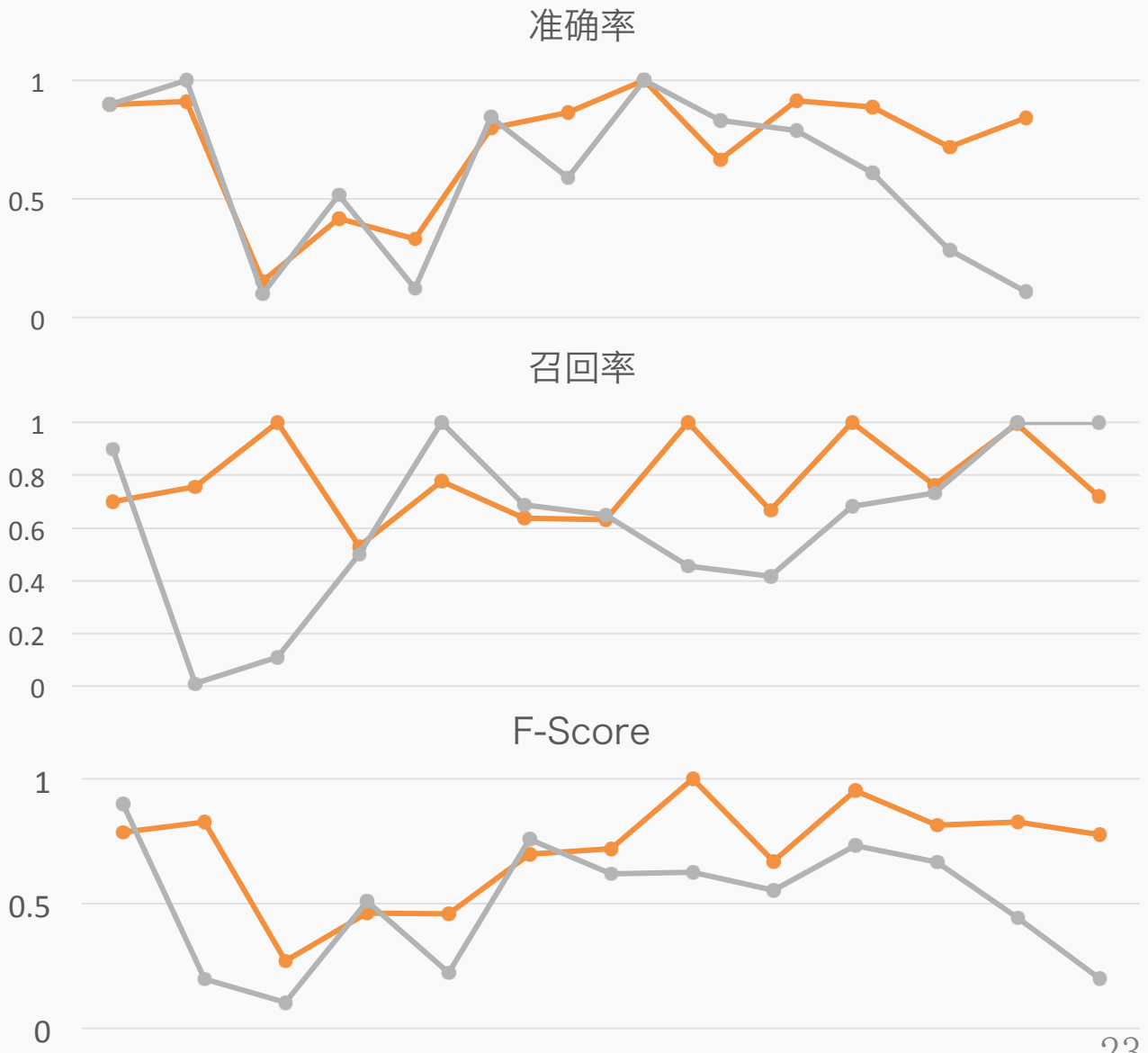






效果对比：分类算法 vs Canary

- ✦ 分类算法
  - ✧ 准确率：60%
  - ✧ 召回率：68.6%
  - ✧ F-Score：58.5%
- ✦ Canary算法
  - ✧ 准确率：72.3%
  - ✧ 召回率：78.3%
  - ✧ F-Score：71.3%

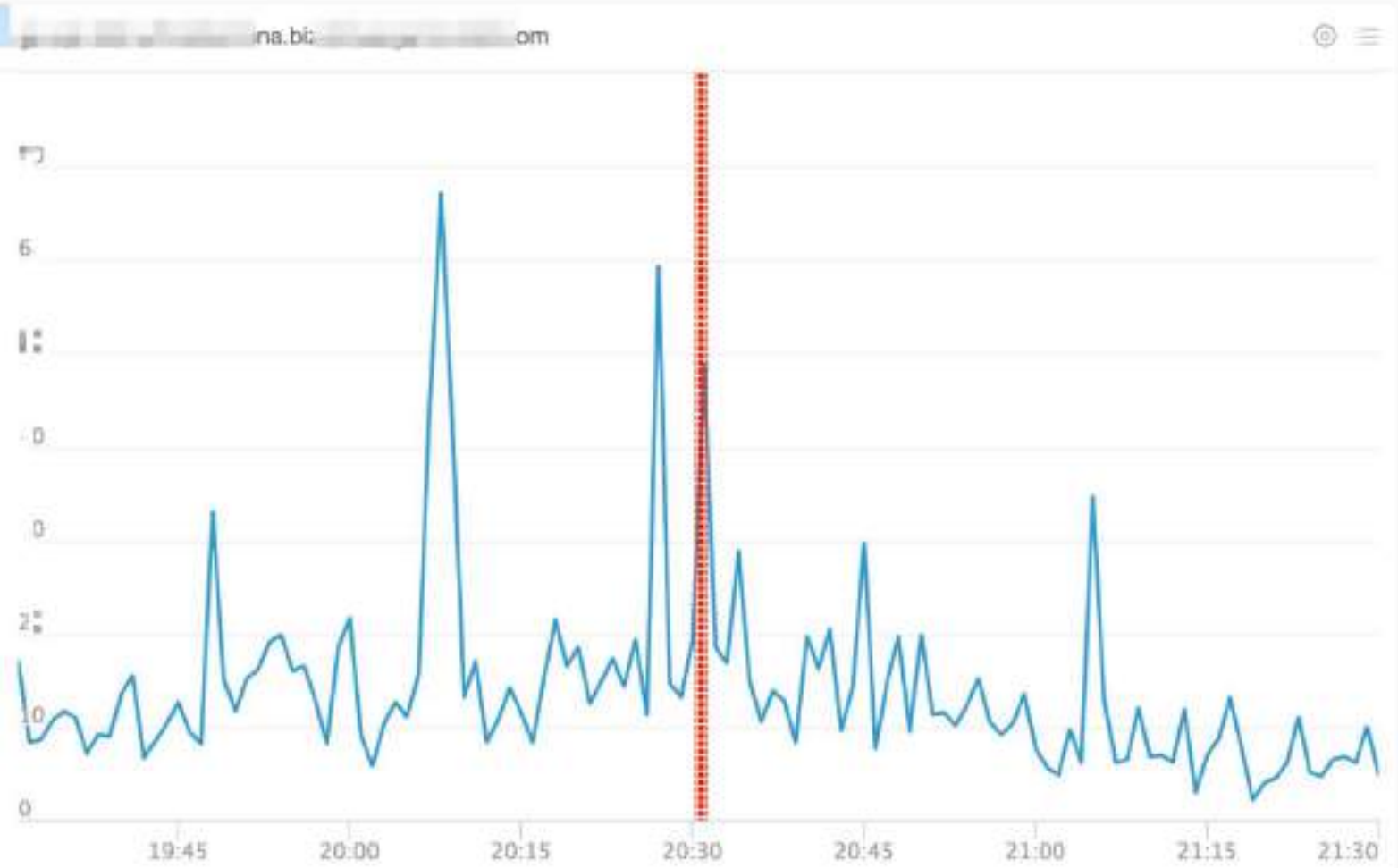


三、快速定位

- 定位案例
- 定位技术方案

案例一：特定errorcode报警

odin报警现场



策略名称：wflog-alert-fail-through

报警状态：P3 故障

通知结果：已通知 报警接收组 sms-gs-api

发生时间：2017-[redacted] 20:31:20

节点：[redacted]

主机：[redacted]

指标：[redacted]pi-didi.wf, errno=5000012,level=WARNING

表达式：[redacted]didi.wf,#5}>=22; tags: errno=5000012, level=WARNING

现场值：

2017-[redacted] 20:30:20	34
2017-[redacted] 20:30:40	99
2017-[redacted] 20:30:50	25
2017-[redacted] 20:31:00	32
2017-09-14 20:31:20	28

案例一：特定errorcode报警---日志详情及Trace关联

把脉告警分析

抽样日志　调用链路　原始日志　请求拓扑

codeline line=/home/ proot/ \/app hp +12 clas

errmsg errno:<ERRNO>|errmsg:|detail:cache lock faile|key

errno 5000012

rawlog [WARNING][2017- 20:30:29.516+0800][line /src/util/Log +12 class=util\] _undef traceId=6

446982859ba7665cabd82d80e6 32164 /drive er | controller=driver/d

_msg=|errno:5000012|errmsg:|detail:cache lock faile|key TER_ORDE 95728473270_10|ret:false|Trace: |

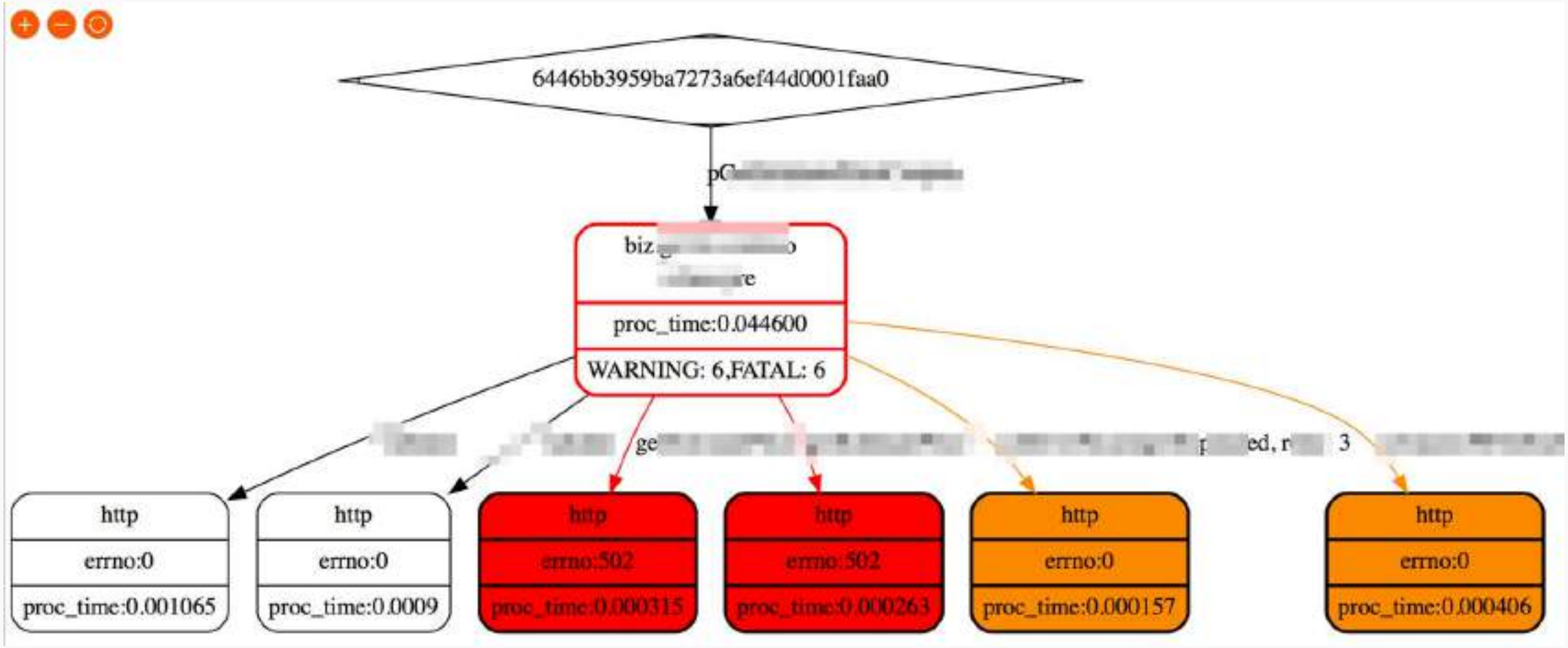
traceId 6446982859ba7665cabd82d80e61bfb0

抽样日志　调用链路　原始日志　请求拓扑

提示：点击+/-进行折叠，点击每行任意地方查看详细信息。以下日志已自动转换编码格式　反查已开启

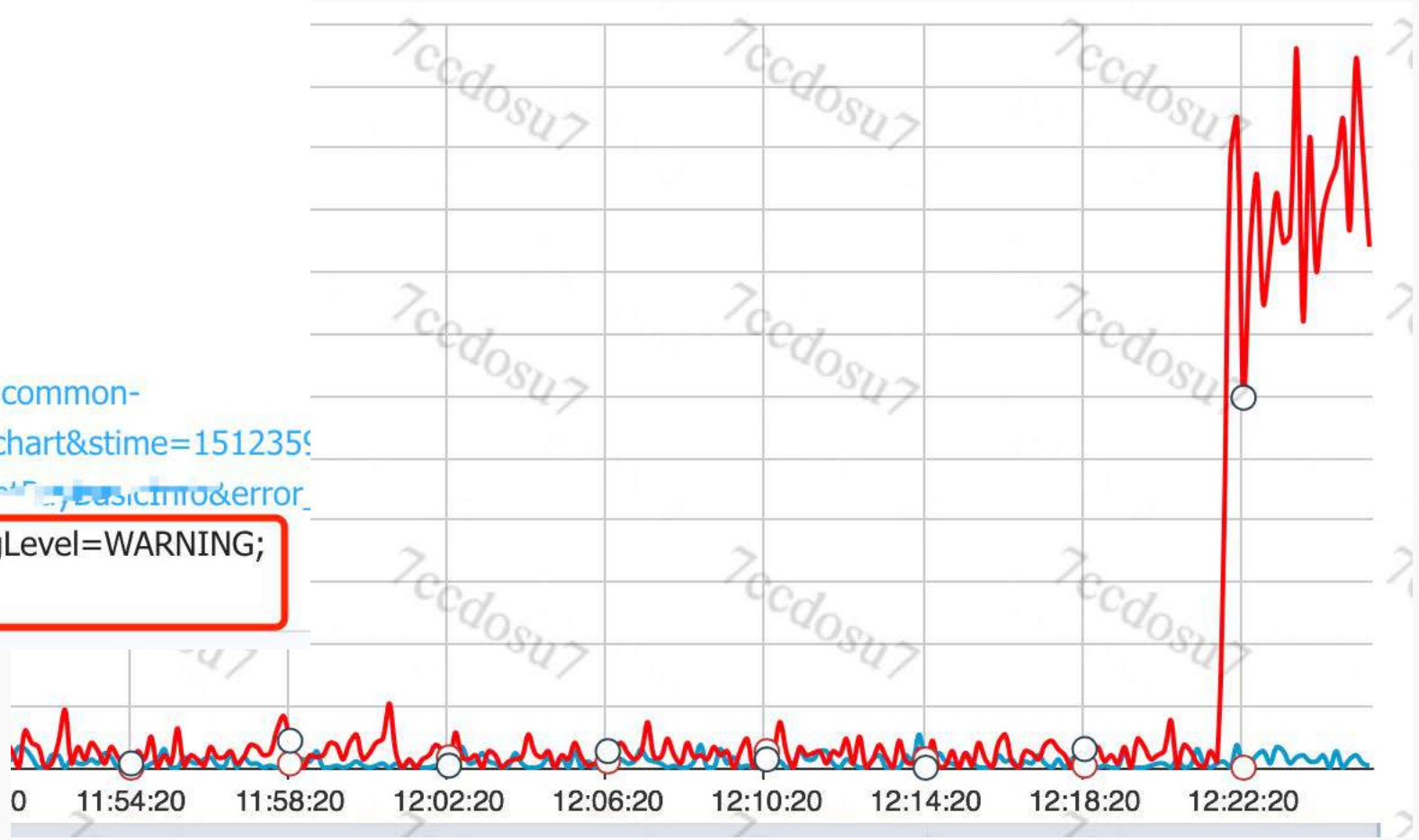
模块	类型	状态	服务/方法	时间线
— p 2 n n	pG	W F	/g coupon	44.60ms
rt.pu om...	http	✓	val	1.06ms
pre om...	http	✓	val	0.90ms
lat,...	http	✗	isC n	0.41ms
mm	http	✓	isD d	0.28ms

案例一：特定errorcode报警---调用拓扑

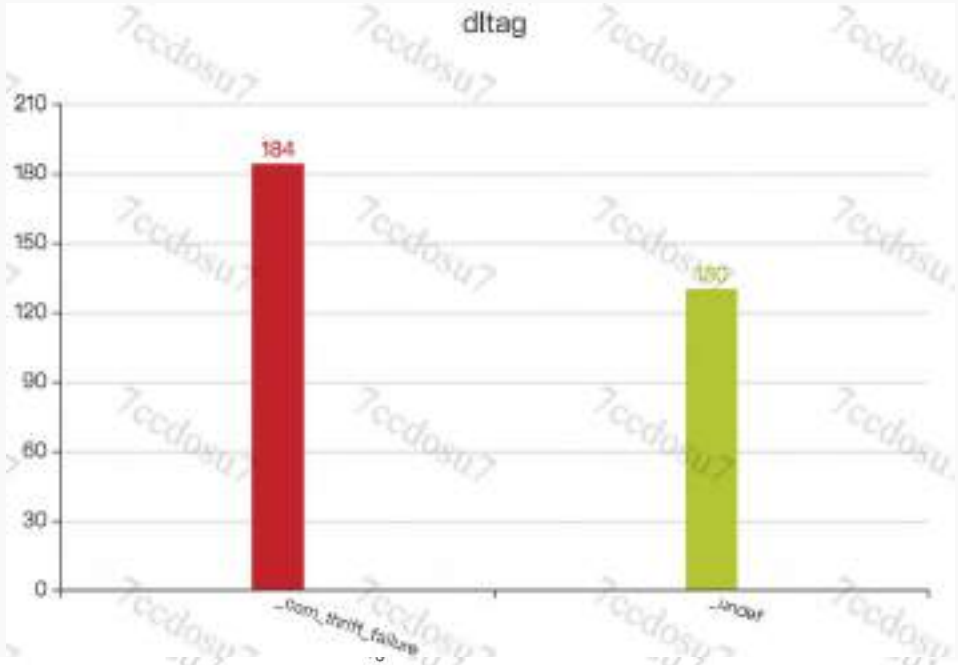


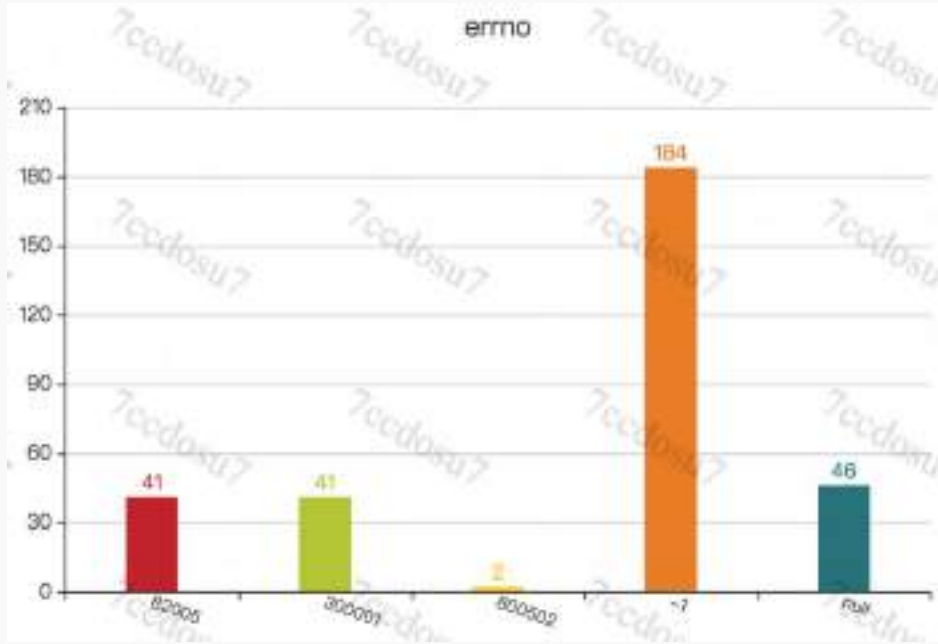
案例二：趋势类指标报警

告警项: 错误日志统计
应用: con
NS: machin com
指标类型: uri
指标: /g Info
当前值: 254.0
预估值: 8
阈值倍率: 15.00
时间: 2017 T12:23:20
详情: http:// /service?tagapp=common-
&tagp=inter_real&tagc=inter_errorchart&stime=151235
000000000&error_uri=/g &error_
突增维度: dltag=_com_thrift_failure; errno=-1; logLevel=WARNING;
错误信息: TSocket read 0 bytes



案例二：趋势类指标报警---成分分析





错误上报详情　错误分析统计

rawLog

创建时间	日志级别	TraceId	Errno	Dltag	Uri	Count	原始日志
2017-09-14 14:37:30:00	WARNING	0aaa3cab59ba23a600004f2c26e92b38		_undef	/gulf y/v1/ er	1	点击查看
2017-09-14 14:37:30:00	WARNING	64469f3559ba23a71f70158a10ff8502	28	_com_http_failure	/gulf	1	点击查看
2017-09-14 14:37:30:00	WARNING	74cc3c4cab374f8b9c9ba53e09a2c8ca	81103	_undef	/g ce	1	点击收起

[WARNING][2017-09-14T14:37:30. 0+0800][line=/home ebroo fstream/ap p +29 class= ontrol::P _unde

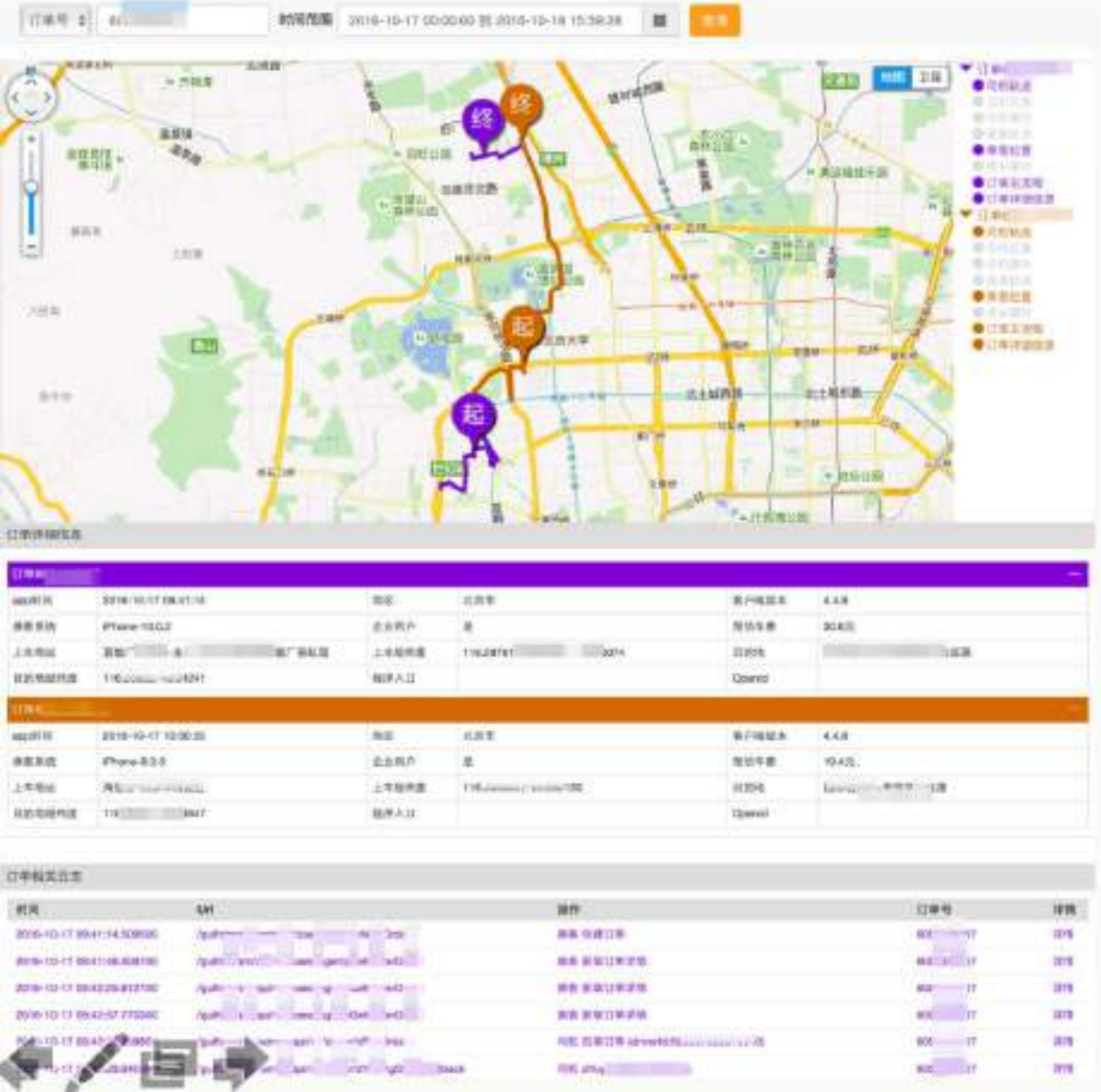
||traceid=74cc3c4cab374f8b9c9ba53e09a2c8ca||uri=/g /1/not 3665||logid=17552234060 =0||||module=c ier||controller= ce||_msg=errno:8

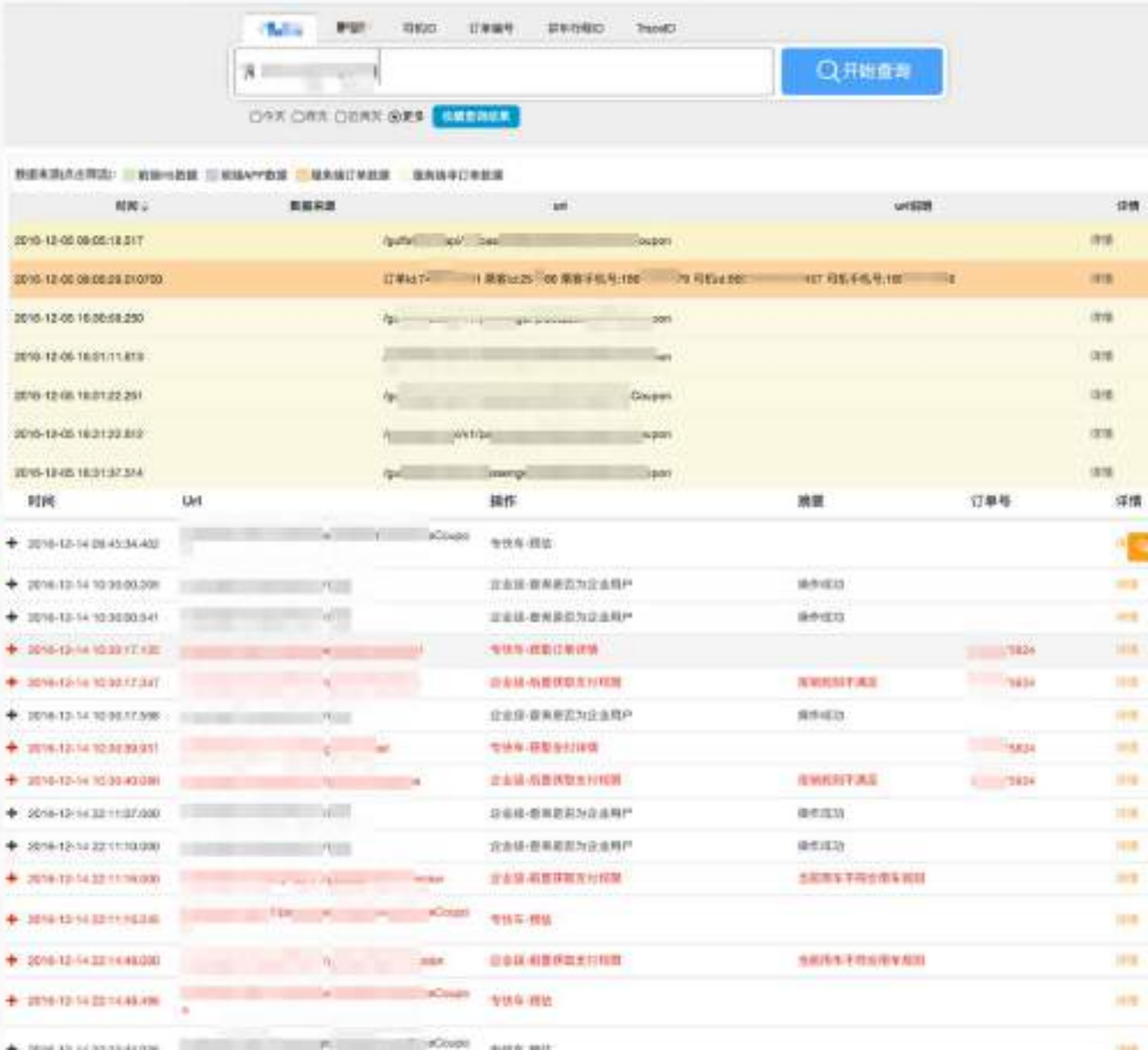
案例三：性能报警


实时抽样TraceID
接口性能分析
错误日志分析
接口调用分析
URI筛选
起止时间
刷新
刷新频率
10秒
自动刷新
响应时间趋势图
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
响应时间：4281ms
上报时间：2017-01-02 20:39:50
点击查看trace详情
响应时间：632ms
上报时间：2017-01-01 20:39:50
点击查看trace详情
20:30:40
20:34:40
20:38:40
20:42:40
20:46:40
20:50:40

案例三：性能报警---链路瓶颈分析

案例四：业务问题定位



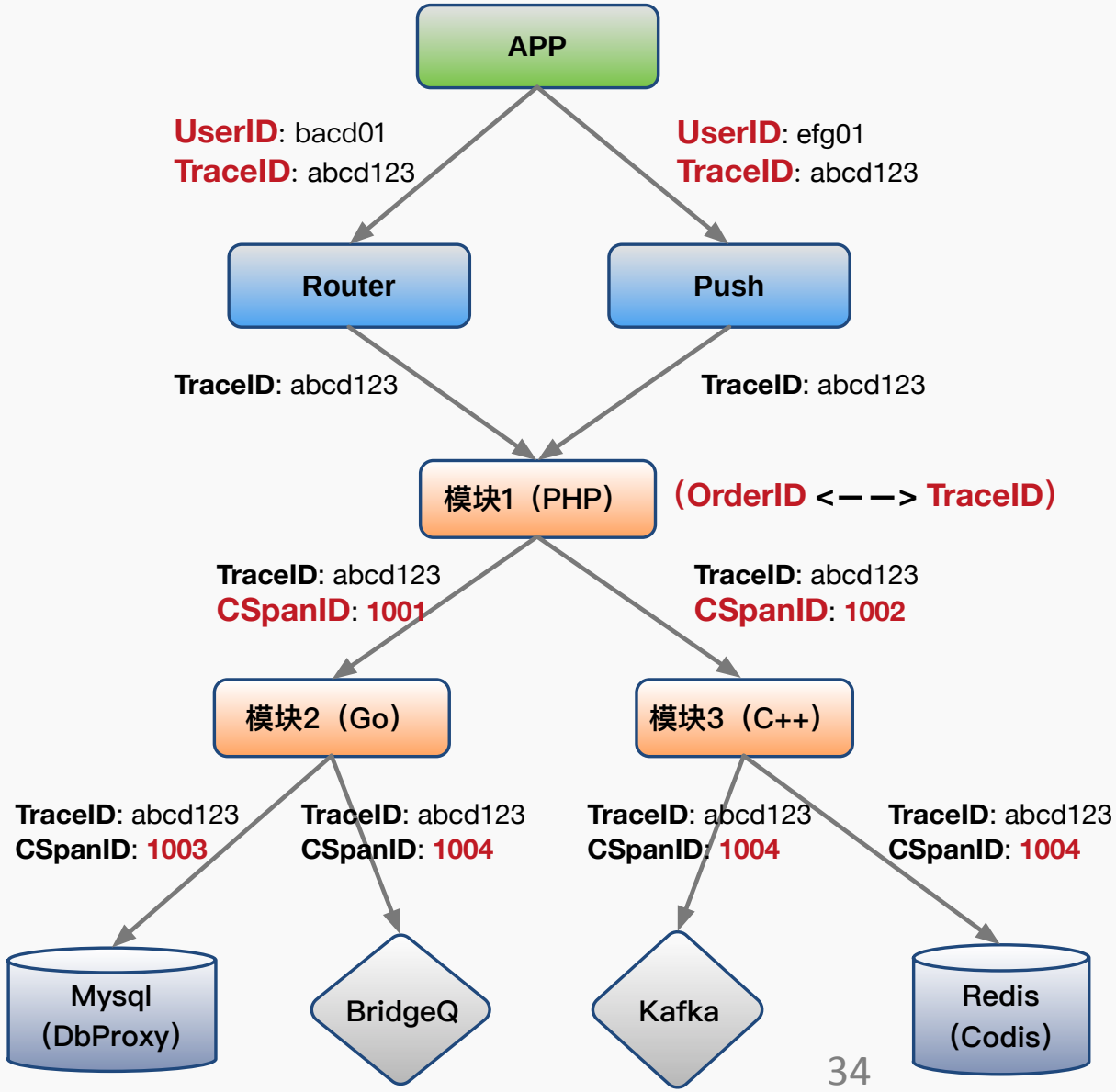


定位技术方案

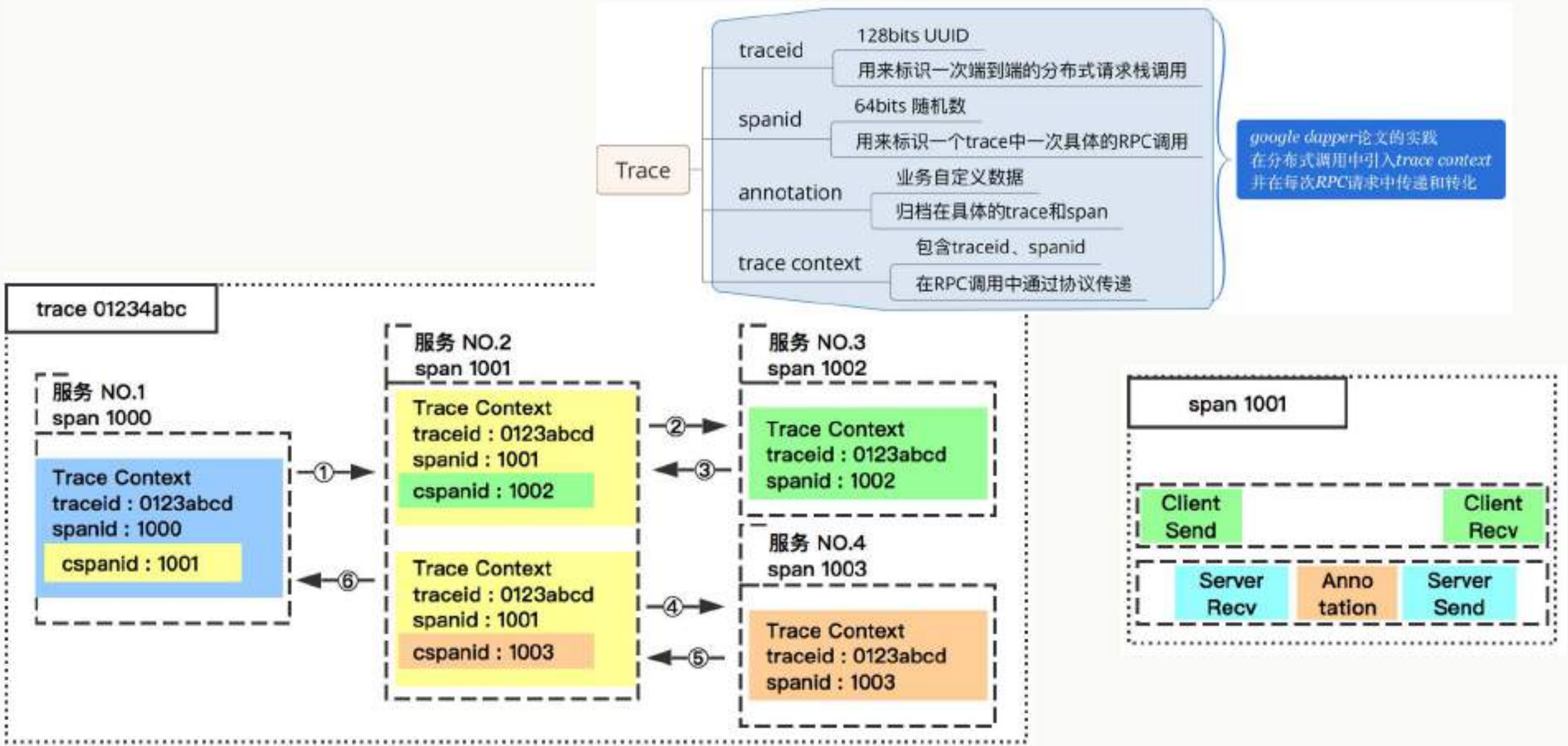
- ✦ 链路追踪与还原
  - ✧ 用户、订单、请求、调用
- ✦ 海量日志治理
  - ✧ 标准化、云端化、关联分析

链路追踪：用户，订单，请求

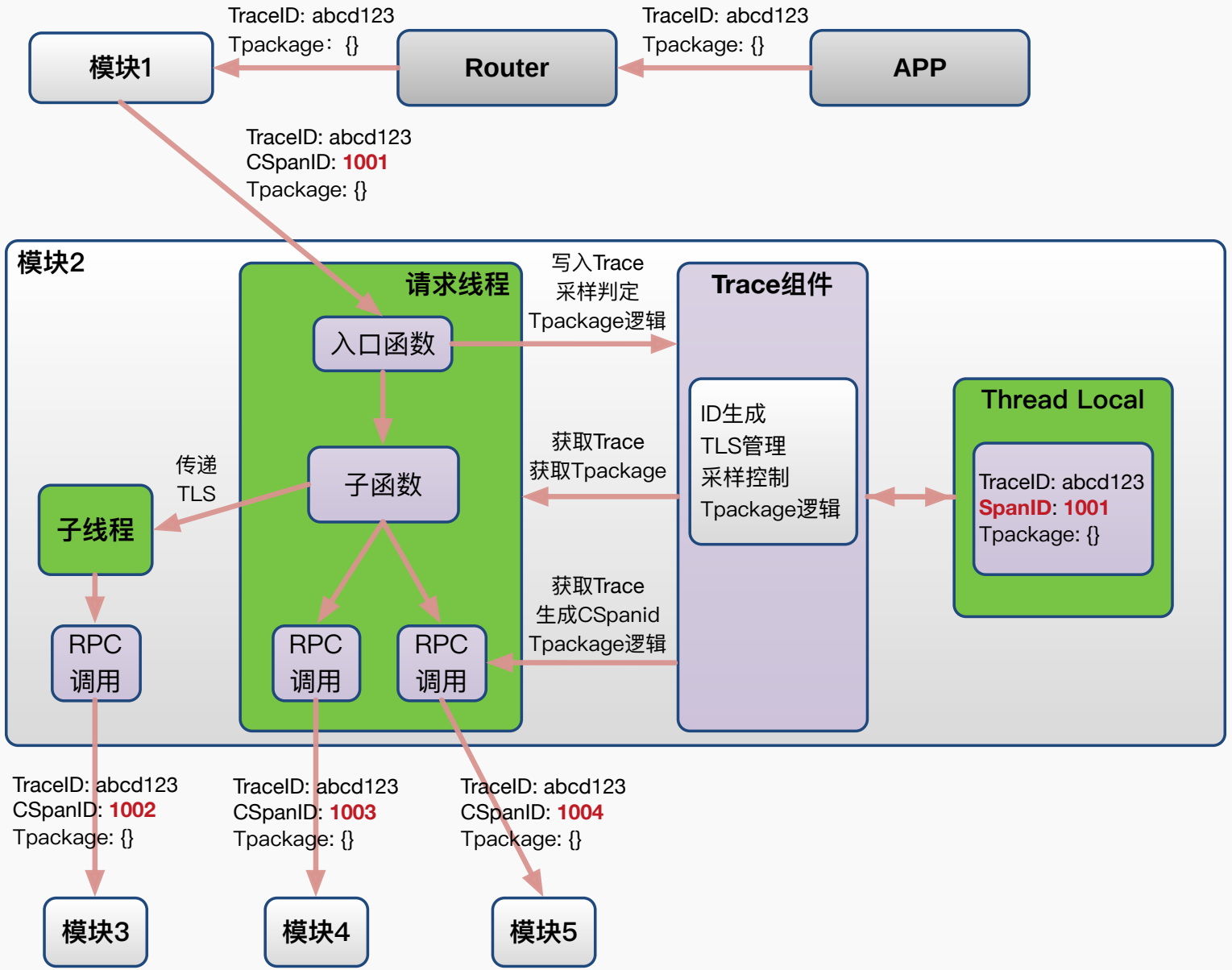
- 请求链路
  - TraceID透传
  - 标识唯一一次请求
- 用户链路
  - APP透传UserID到接入层
- 订单链路
  - API层：OrderID关联TraceID
  - 司乘数据关联



链路追踪：调用链


Trace
traceid
128bits UUID
用来标识一次端到端的分布式请求栈调用
spanid
64bits 随机数
用来标识一个trace中一次具体的RPC调用
annotation
业务自定义数据
归档在具体的trace和span
trace context
包含traceid、spanid
在RPC调用中通过协议传递
google dapper论文的实践
在分布式调用中引入trace context
并在每次RPC请求中传递和转化
trace 01234abc
服务 NO.1
span 1000
Trace Context
traceid : 0123abcd
spanid : 1000
cspanid : 1001
服务 NO.2
span 1001
Trace Context
traceid : 0123abcd
spanid : 1001
cspanid : 1002
Trace Context
traceid : 0123abcd
spanid : 1001
cspanid : 1003
服务 NO.3
span 1002
Trace Context
traceid : 0123abcd
spanid : 1002
服务 NO.4
span 1003
Trace Context
traceid : 0123abcd
spanid : 1003
①
②
③
④
⑤
⑥
span 1001
Client Send
Client Recv
Server Recv
Anno tation
Server Send

链路追踪：内部机制



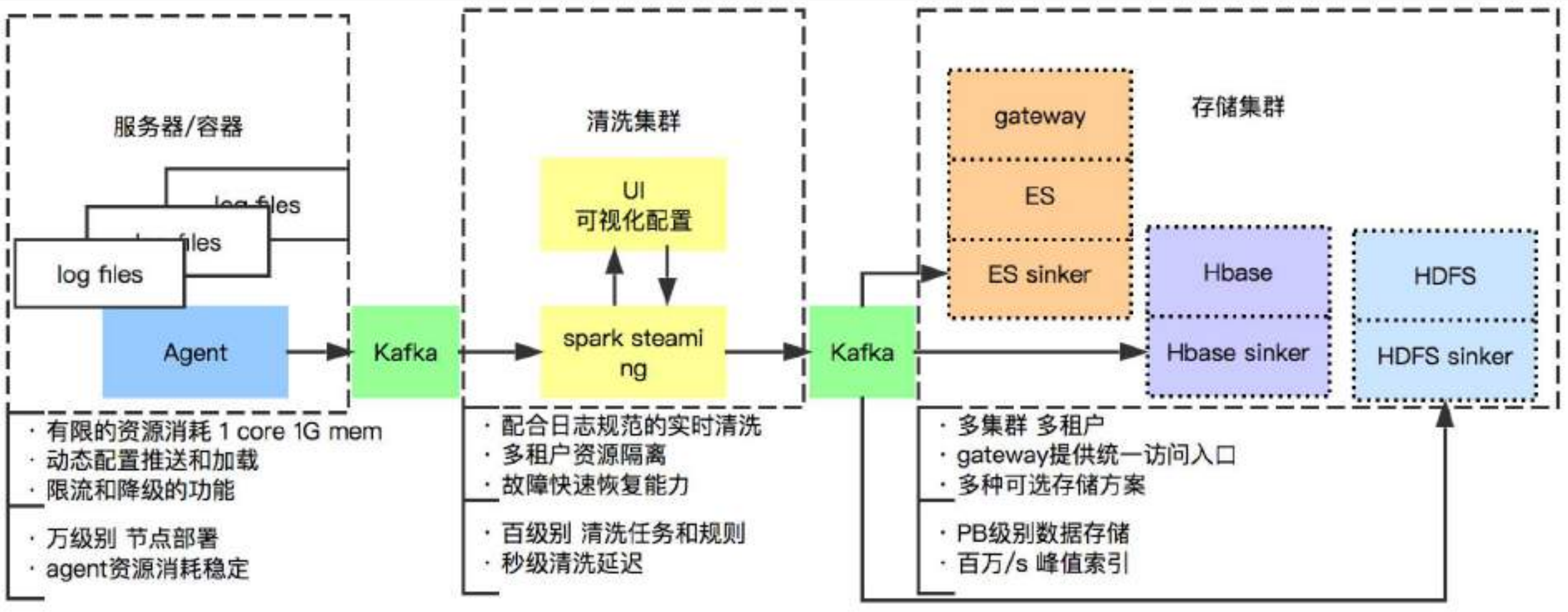
- ✦ 内部透传方案设计
  - ✧ 低（无）业务侵入：TLS
- ✦ 挑战：异构链路
  - ✧ 多语言支持：php/go/java/c++
  - ✧ 多协议支持：http/thrift
  - ✧ 存储及队列：kafka/BQ/mysql
- ✦ 数据透传服务Tpackage
  - ✧ 全链路压测：压测流量标识
  - ✧ 分城市发布：城市标识
- ✦ 采样机制引入
  - ✧ 细粒度日志/临时排查

日志治理：标准化


日志规范
业务日志打印格式、位置、内容的约定
引入schema概念
配套日志数据流式处理系统
日志组件
提供5种语言的日志打印组件，兼容现有大部分模块代码
引入trace机制，自动记录trace context关键内容
日志规范落地
业务层面不需要额外关注 trace context


[NOTICE][2017-03-16T21:18:41.473+0800][line=]
固定日志头：等级、时间戳、代码行
_com_request_in
tag：描述日志打印时机和内容
‖traceid=0a61355fe111f8bfed17419410310303‖sampling=0
‖spanid=248b0b7136fe0b77‖order_id=1001
内容：k=v的格式 ‖分隔

日志治理：云端化数据架构



THANK YOU

滴滴
滴滴一下 美好出行

北京嘀嘀无限科技发展有限公司

李培龙

北京 海淀

扫一扫上面的二维码图案，加我微信