AIOps对监控报警架构的挑战

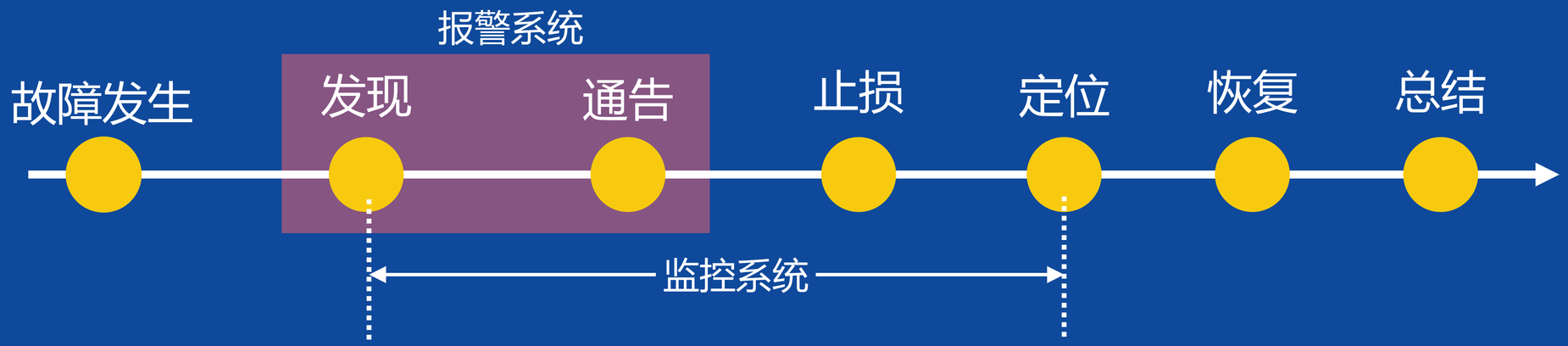
周伟 范月林
百度基础架构部

TABLE OF

CONTENTS 大纲

- 背景介绍
- 报警系统业务模型
- 异常判断子系统
- 事件管理子系统
- 通告发送子系统
- 总结

百度Noah报警系统



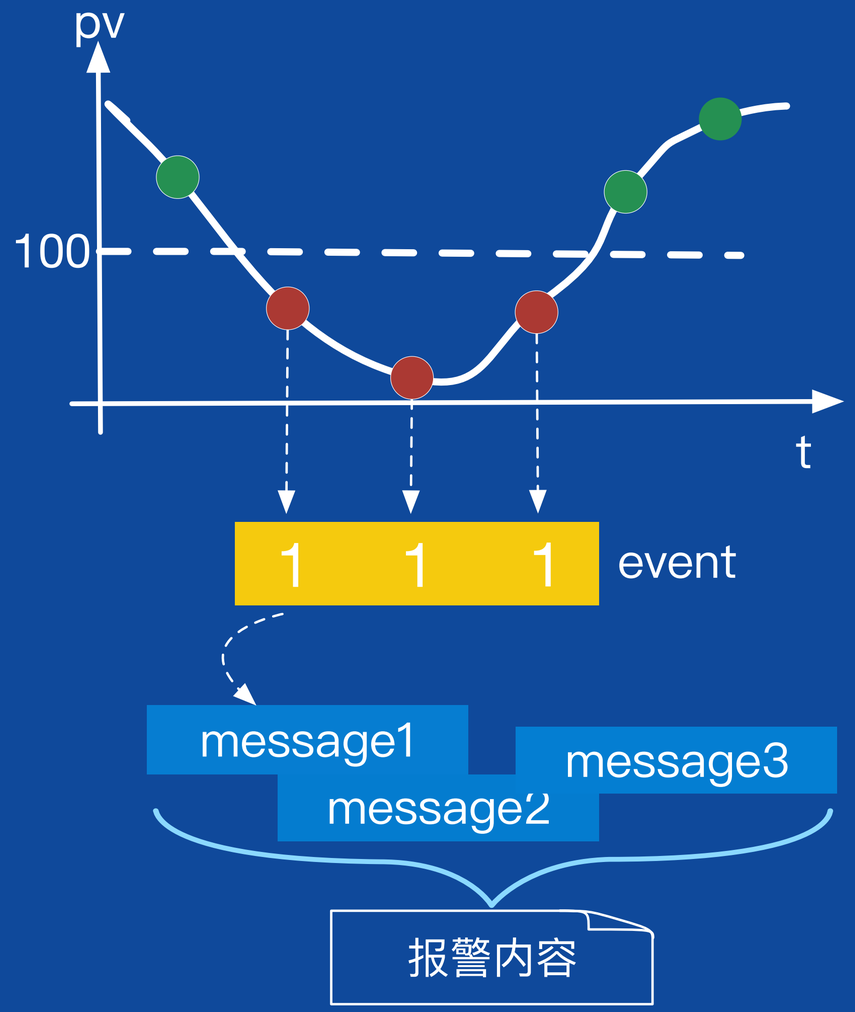
支撑多种上层业务

- 内部监控平台
- 公有云监控服务BCM
- 私有云运维监控产品NoahEE
- 百度AIOps智能运维产品

面临巨大架构挑战

- 数据量：五千万数据点/秒
- 报警配置：百万级别
- 报警事件量：千万/每天
- 报警时效性：秒级

使命：精准告警



遇到的问题：

1. 四则运算无法满足业务所有需求（准确率<50%）
2. 核心告警经常被遗漏，故障不能及时发现和处理
3. 机房故障等原因会造成报警风暴，淹没核心告警

真实的需求：

1. 升级异常判断算法，提高准确率
2. 报警分级/报警认领/逐级通告，防止告警遗漏
3. 报警合并，抑制报警风暴

需求1：落地AIOps算法的挑战

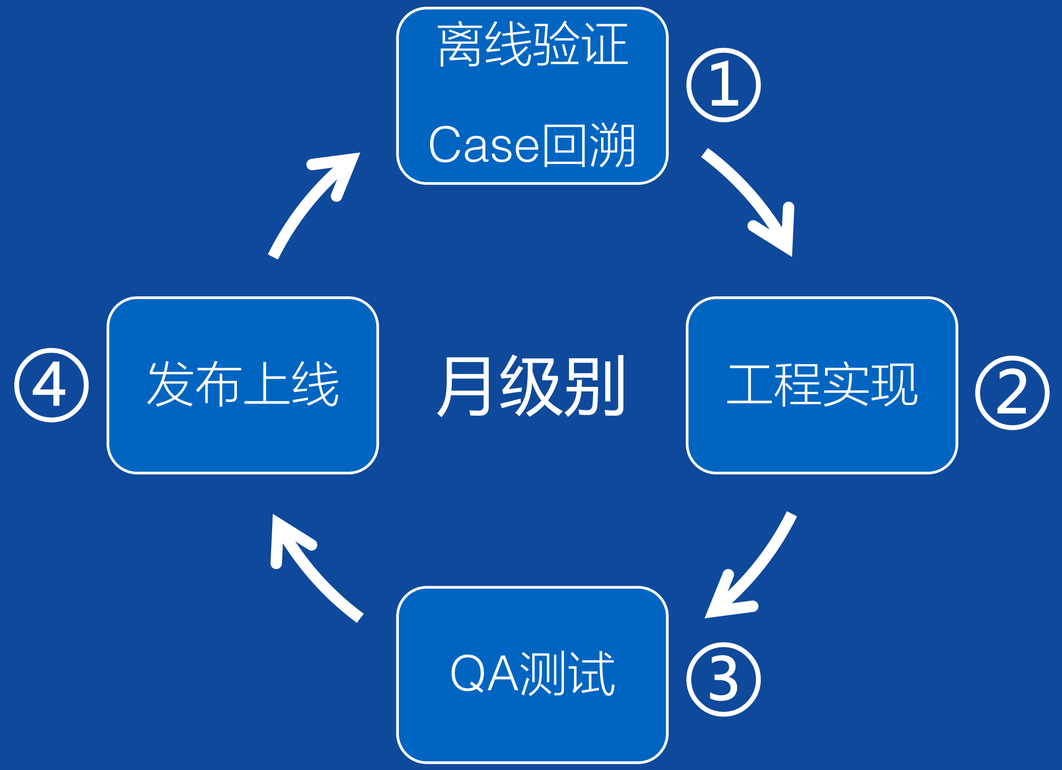
场景：PV流量指标的突升突降检测

$$z-score=\frac{y-f(x)}{\sqrt{f(x)}}\sim N(0,1)$$

其中f(x)依赖根据历史数据训练得到的模型

AIOps算法落地难点：

- 算法不固定，不断迭代
- 强依赖模型，模型变更频繁
- 算法资源消耗千差万别



需求2：报警管理的挑战

场景	需求
网络抖动导致误报警	防抖动过滤
一线值班人未能及时响应	重复提醒，报警升级
故障处理过程中仍继续收到报警提醒	报警认领
磁盘容量满需人工清理日志	报警驱动日志自动清理

如何能提供一个标准的报警模型，应对繁琐的需求？

需求3：报警风暴的挑战

场景	现象	报警量
机器故障	机器层面报警 –> 实例层面报警 –> 上游应用报警	10倍
应用故障	应用下所有实例告警 –> 上游应用告警	100倍
机房故障	造成机器的存活性、域名的存活性、业务请求流量突降等异常	1000倍

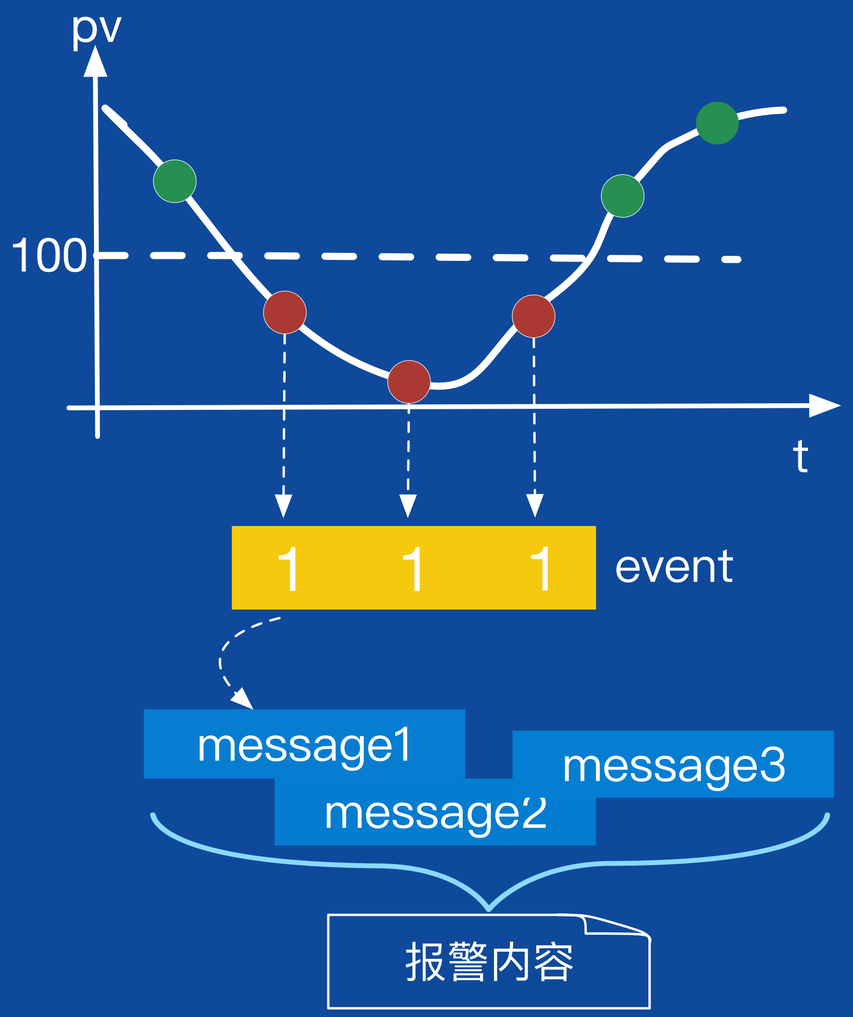
如何将运维工程师从报警风暴中解救出来？

TABLE OF

CONTENTS 大纲

- 背景介绍
- **报警系统业务模型**
- 异常判断子系统
- 事件管理子系统
- 通告发送子系统
- 总结

报警系统业务模型


pv
100
t
1
1
1
event
message1
message3
message2
报警内容

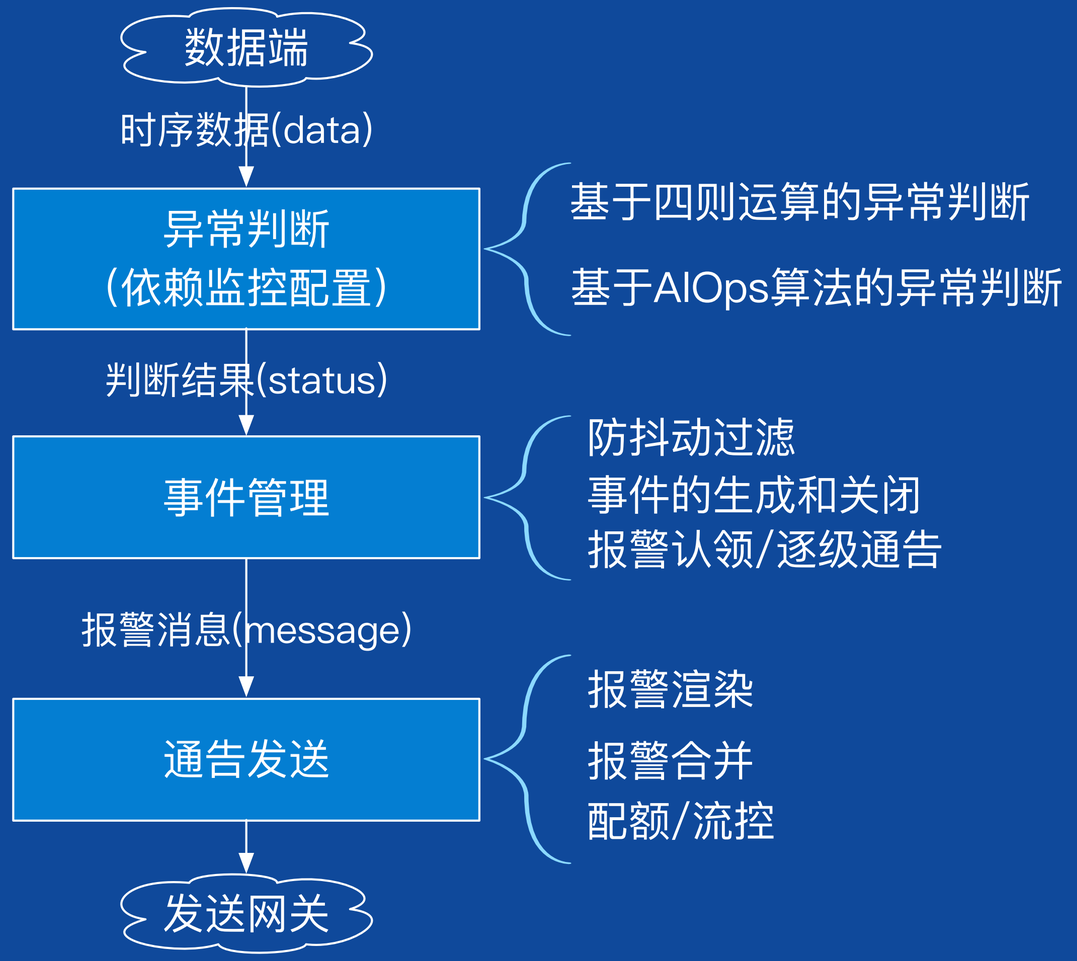

数据端
时序数据(data)
异常判断
(依赖监控配置)
基于四则运算的异常判断
基于AIOps算法的异常判断
判断结果(status)
事件管理
防抖动过滤
事件的生成和关闭
报警认领/逐级通告
报警消息(message)
通告发送
报警渲染
报警合并
配额/流控
发送网关

TABLE OF

CONTENTS 大纲

- 背景介绍
- 报警系统业务模型
- **异常判断子系统**
- 事件管理子系统
- 通告发送子系统
- 总结

异常判断的需求

目标：迭代周期从月级别减少为天级别

① 线上和线下保持一致环境

- 相同的运行界面，保证输入和输出的数据模型一致
- 不需要改写算法，保证线上/线下相同运行逻辑

② 小流量测试

- 真实流量测试，验证线上/线下一致
- 性能和资源评估，验证稳定性

③ 算法和模型分离，提高模型迭代效率

策略运行平台

策略运行平台

离线环境

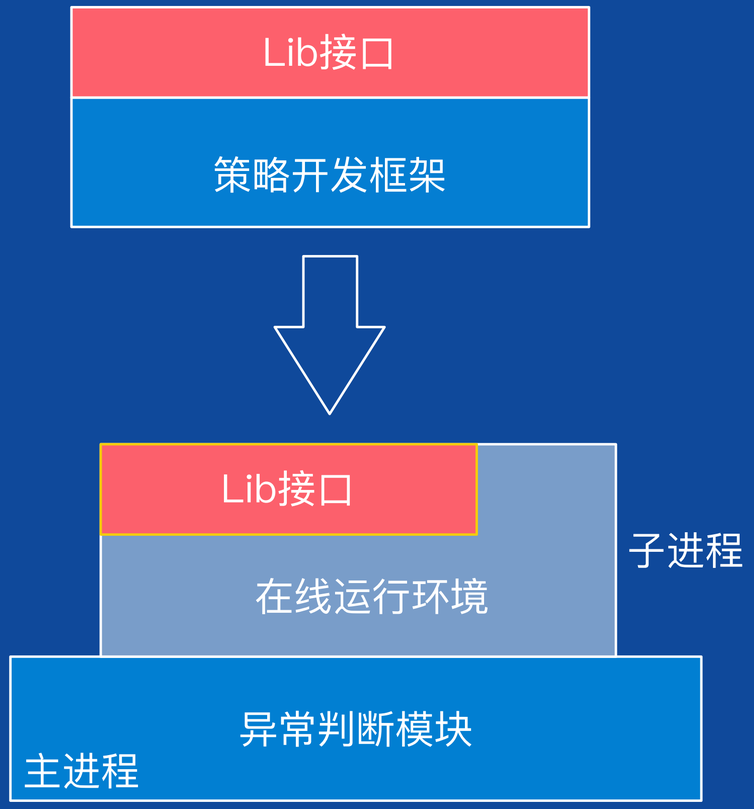
- 提供一致的算法开发接口，保证在线上和线下的运行环境一致
- 支持离线回溯Case，进行算法评估

近线环境

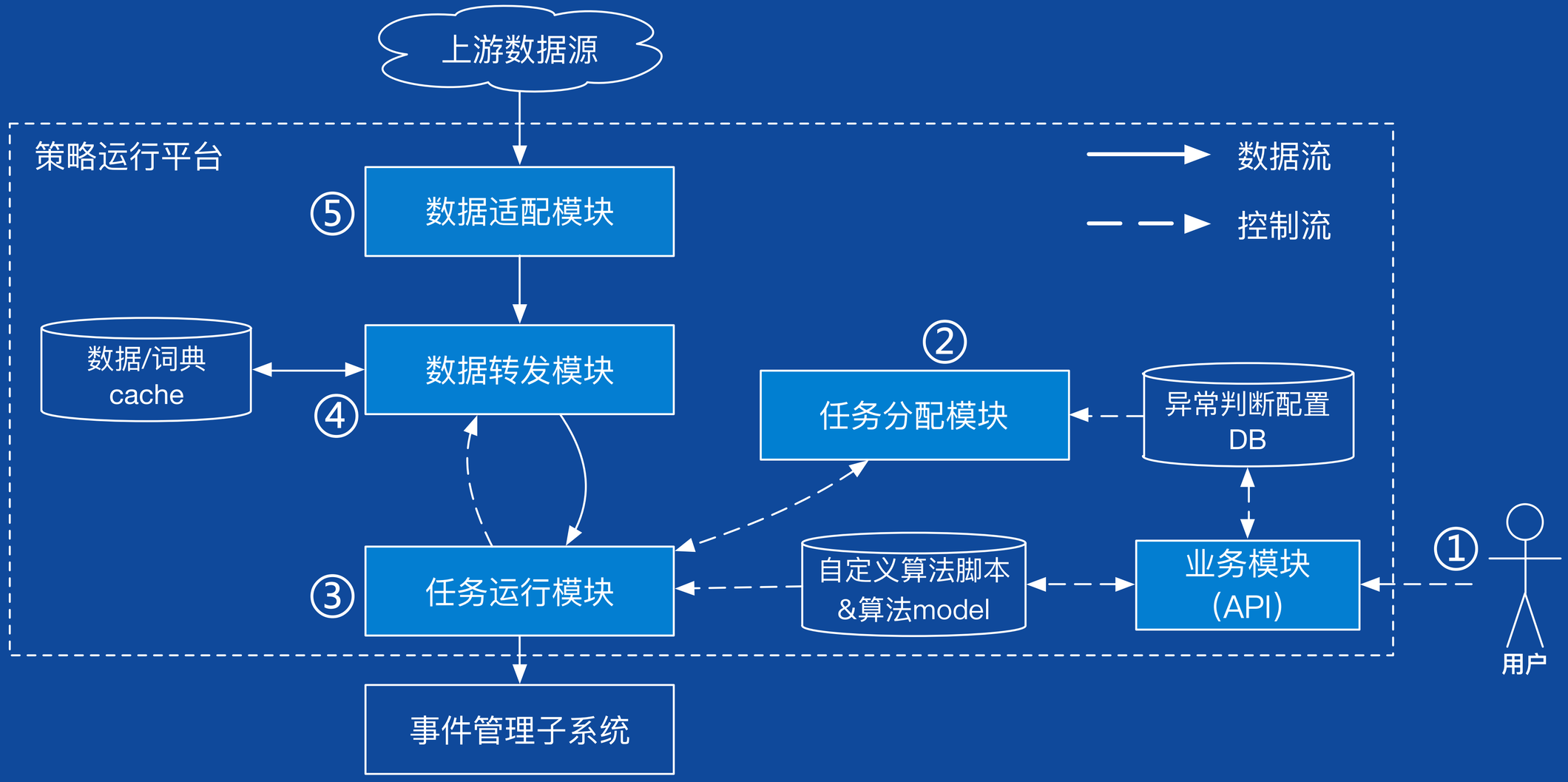
- 提供小规模数据的运行环境，能够快速验证算法在线上的真实运行效果
- 评估资源消耗和稳定性

在线环境

- 提供稳定可靠的算法运行环境
- 异常判断结果发到事件管理子系统



策略运行平台架构图


上游数据源
策略运行平台
数据流
控制流
⑤
数据适配模块
数据/词典
cache
数据转发模块
④
②
任务分配模块
异常判断配置
DB
③
任务运行模块
自定义算法脚本
&算法model
业务模块
(API)
①
用户
事件管理子系统

关键设计：状态恢复

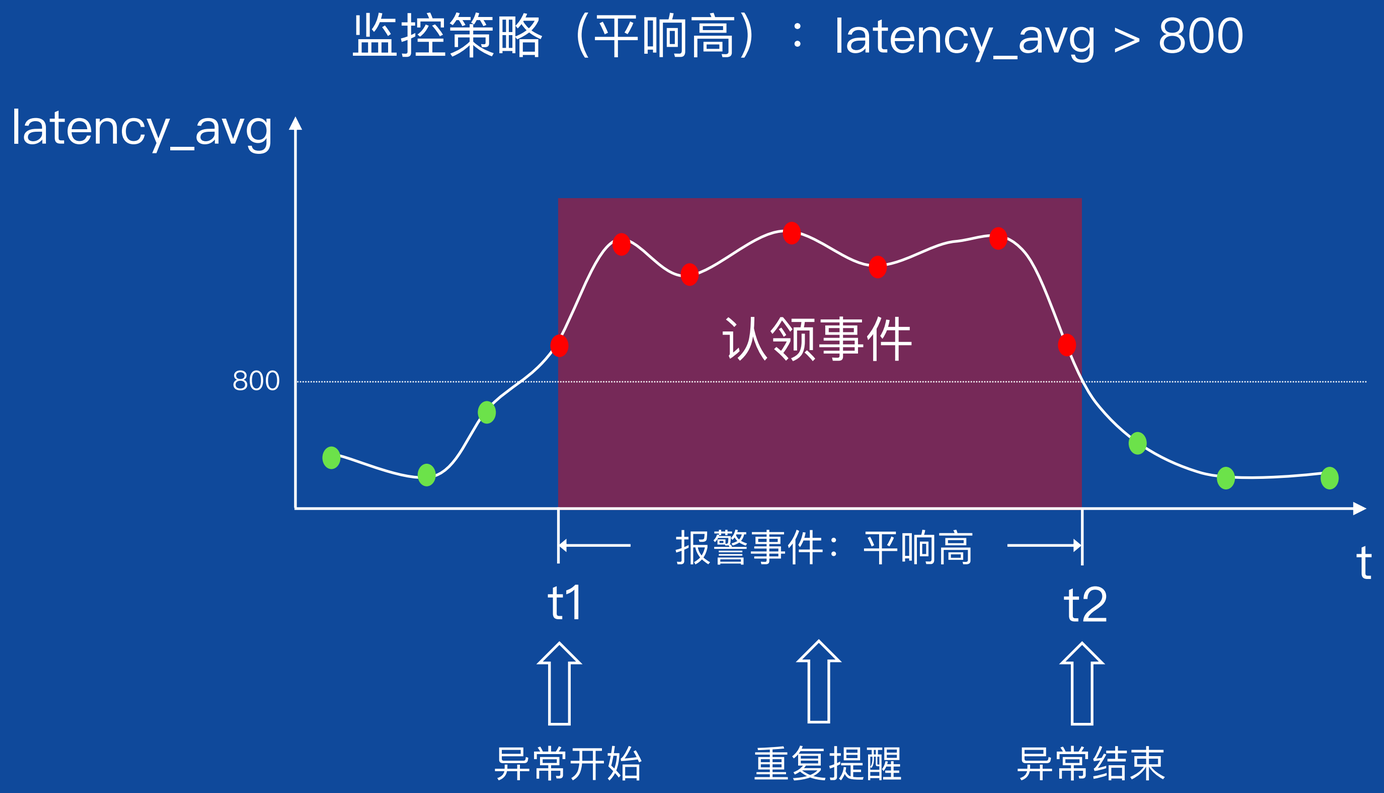
- 问题：Failover时，如何快速恢复算法的运行状态
- 分析：谁应该负责状态的恢复
  - 平台方：需要恢复全部状态，成本高，稳定性差
  - 开发方：增加用户开发成本
- 方案：将近期的数据重新运行，自动恢复运行状态
  - AIOps算法只依赖数据和模型
  - AIOps算法运行速度很快

TABLE OF

CONTENTS 大纲

- 背景介绍
- 报警系统业务模型
- 异常判断子系统
- **事件管理子系统**
- 通告发送子系统
- 总结

什么是报警事件

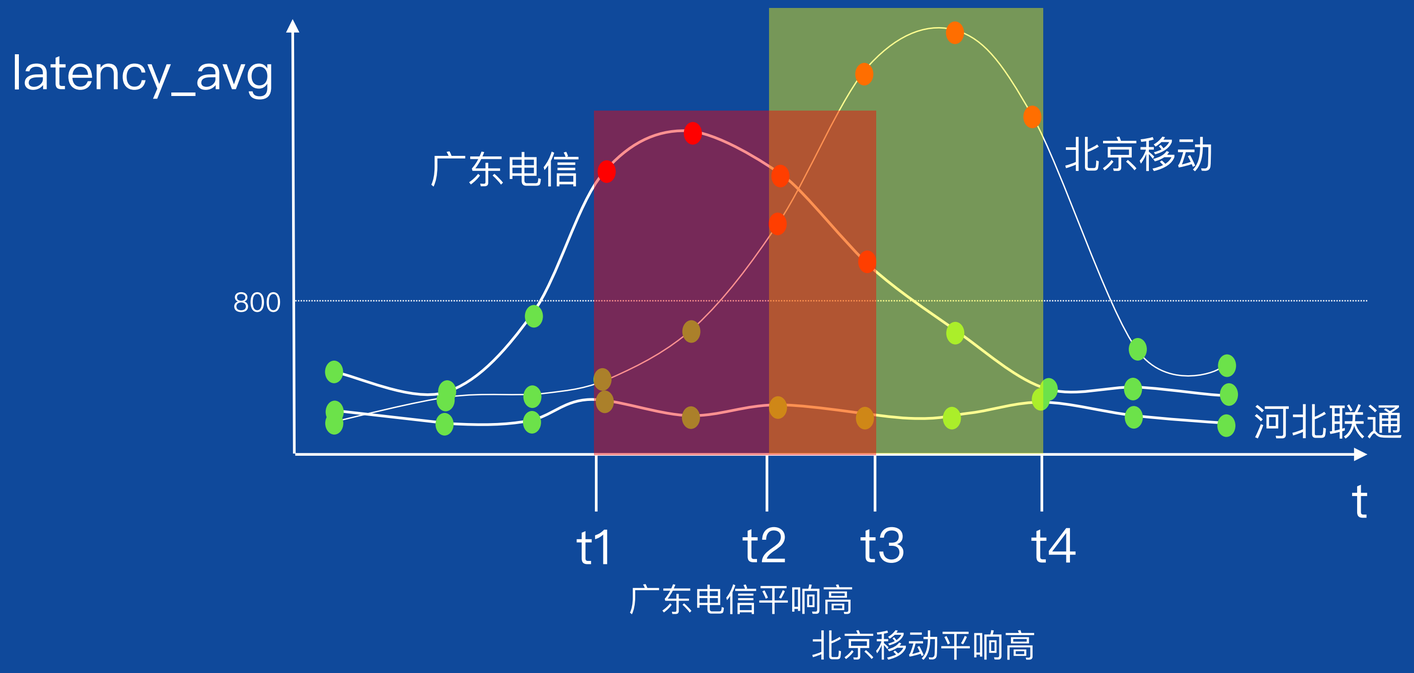


事件基本特征：

- 持续性
  - 异常时间、恢复时间
- 事件与通知 1:N
  - 报警通知、重复提醒、恢复通知等
- 认领对象–事件

多维度报警事件

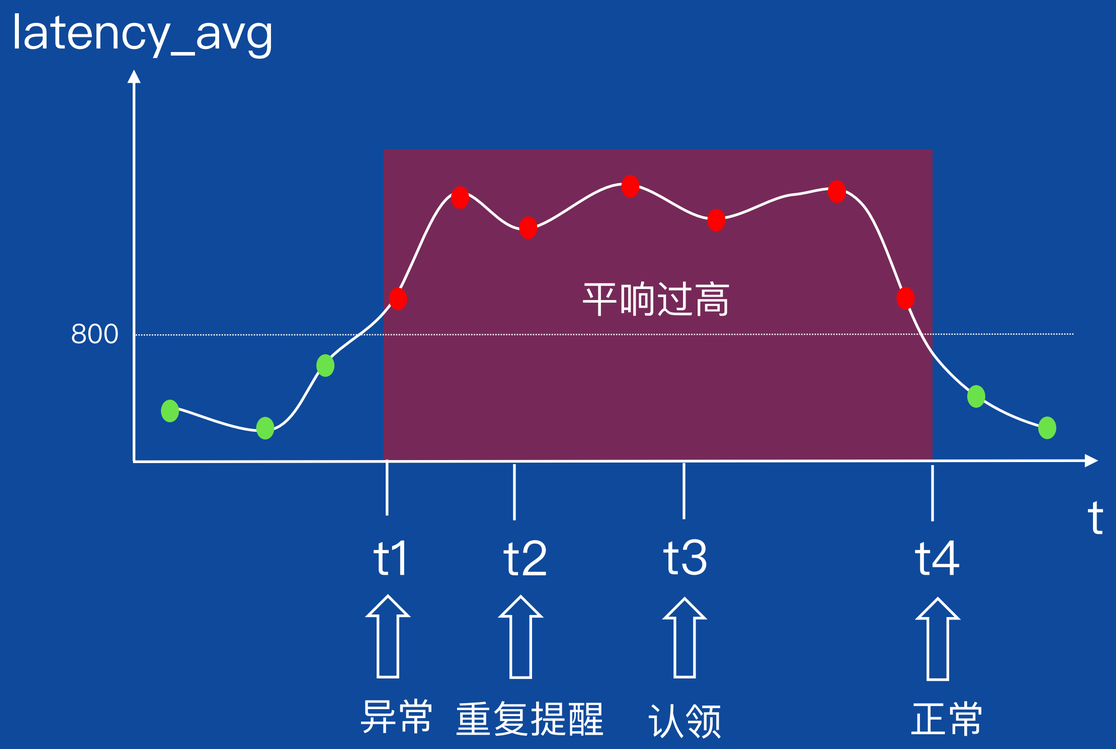
监控策略（分省份分运营商平响高）：latency_avg {isp=*, province=*} > 800

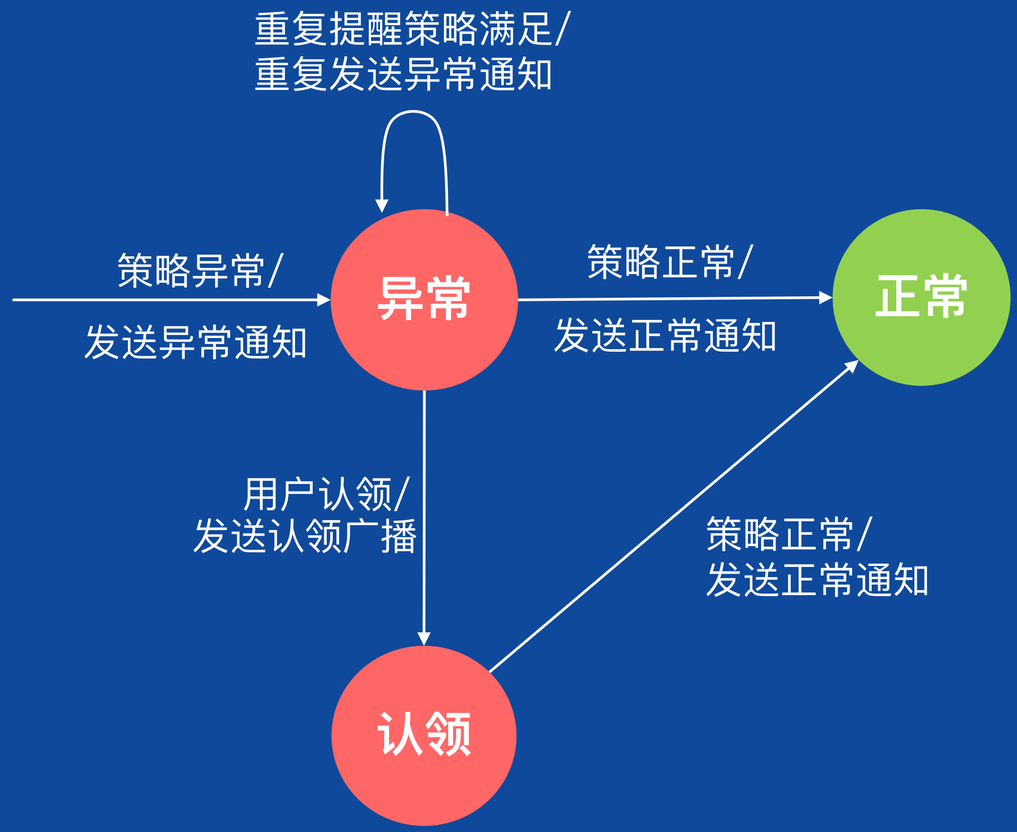


事件基本特征：

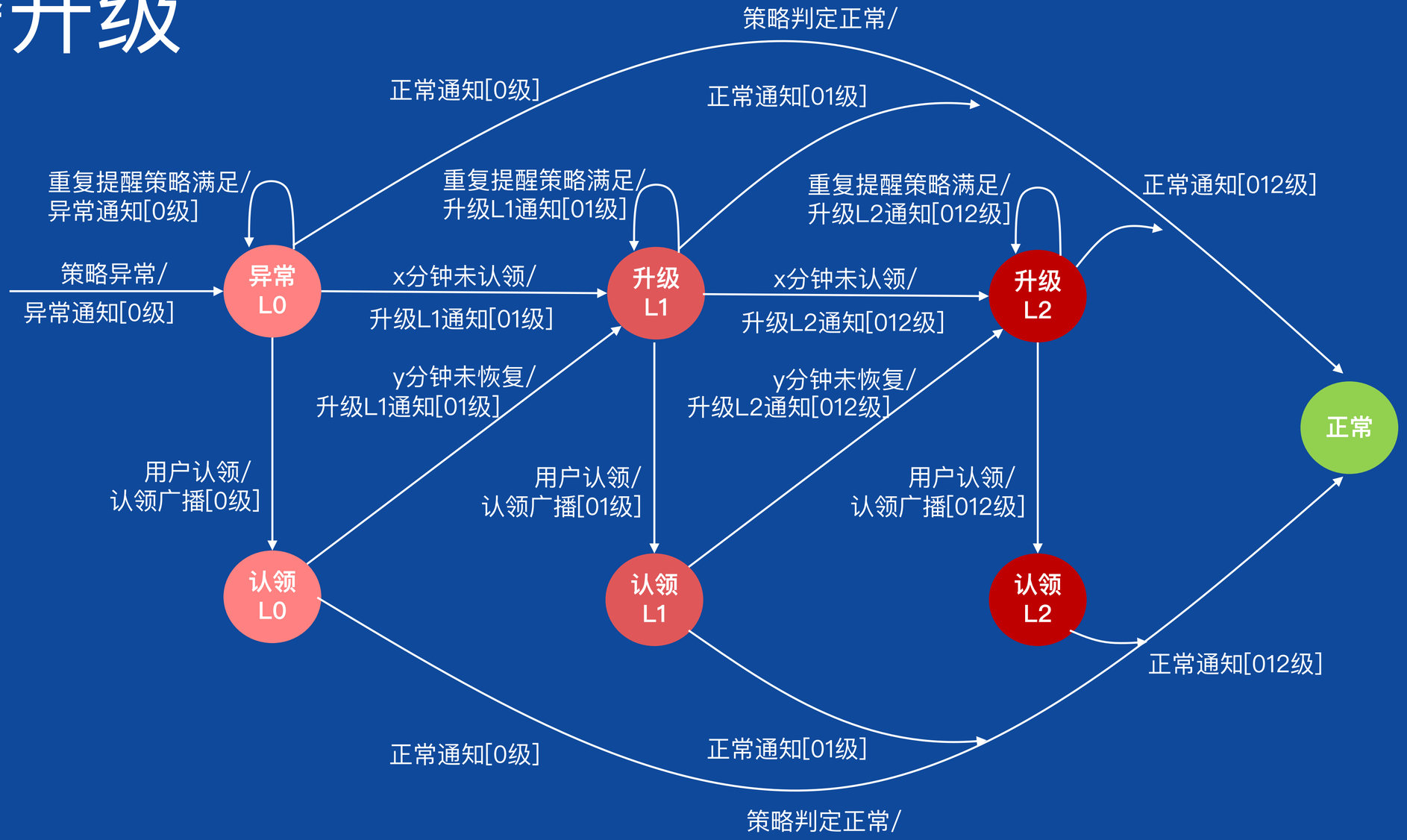
- 持续性
  - 异常时间、恢复时间
- 事件与通知 1:N
  - 报警通知、重复提醒、恢复通知等
- 认领对象–事件
- 策略与事件 1:N
  - 策略+维度

报警事件状态


latency_avg
800
平响过高
t
t1
t2
t3
t4
异常
重复提醒
认领
正常


重复提醒策略满足/
重复发送异常通知
策略异常/
发送异常通知
异常
策略正常/
发送正常通知
正常
用户认领/
发送认领广播
策略正常/
发送正常通知
认领

报警升级


策略判定正常/
正常通知[0级]
正常通知[01级]
正常通知[012级]
重复提醒策略满足/
异常通知[0级]
重复提醒策略满足/
升级L1通知[01级]
重复提醒策略满足/
升级L2通知[012级]
策略异常/
异常通知[0级]
异常
L0
x分钟未认领/
升级L1通知[01级]
升级
L1
x分钟未认领/
升级L2通知[012级]
升级
L2
正常
y分钟未恢复/
升级L1通知[01级]
y分钟未恢复/
升级L2通知[012级]
用户认领/
认领广播[0级]
用户认领/
认领广播[01级]
用户认领/
认领广播[012级]
认领
L0
认领
L1
认领
L2
正常通知[012级]
正常通知[0级]
正常通知[01级]
策略判定正常/

事件状态机引擎

- 模型抽象
  - 基本元素：状态、条件、动作
  - 多种需求一个模型
- 可扩展性强
  - 状态机描述文件，自由定义状态机行为
- 引擎式运行
  - 实例化事件，event=new StateMachine（config）
- 运维成本低
  - 逻辑清晰，表达力强，避免大量 if else

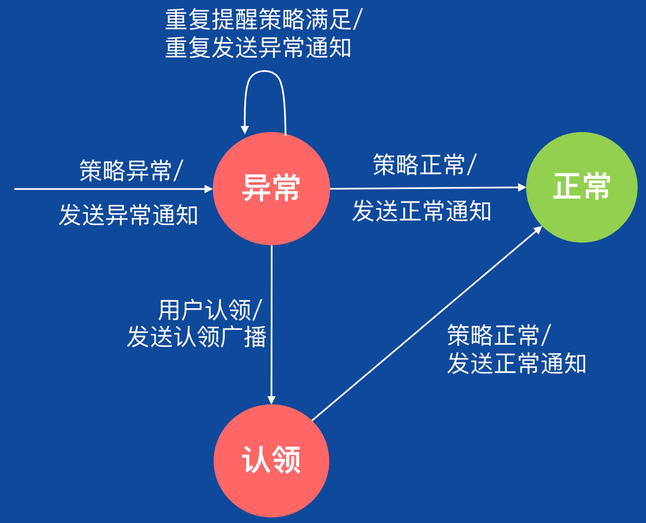


TABLE OF

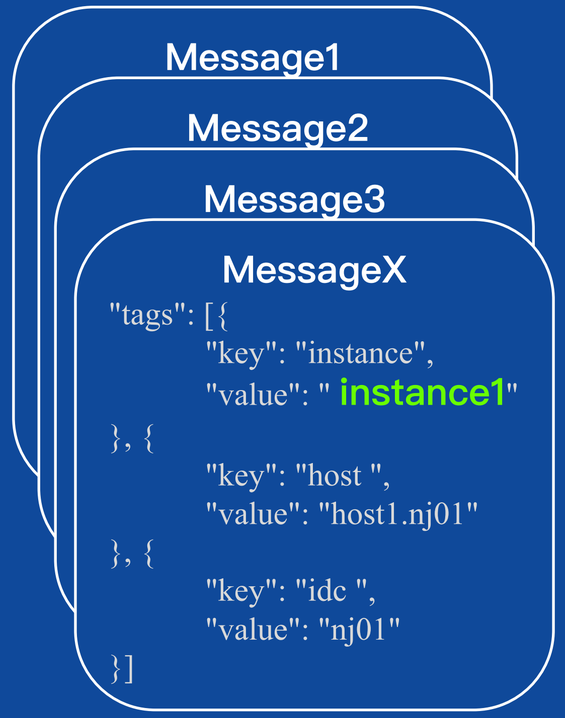
CONTENTS 大纲

- 背景介绍
- 报警系统业务模型
- 异常判断子系统
- 事件管理子系统
- **通告发送子系统**
- 总结

报警合并

- 按部署架构合并
  - 相同实例、服务、集群
  - 相同机器、机房

合并前

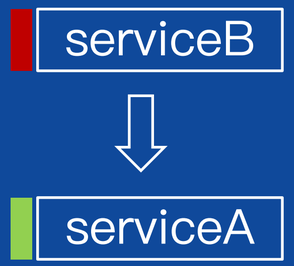


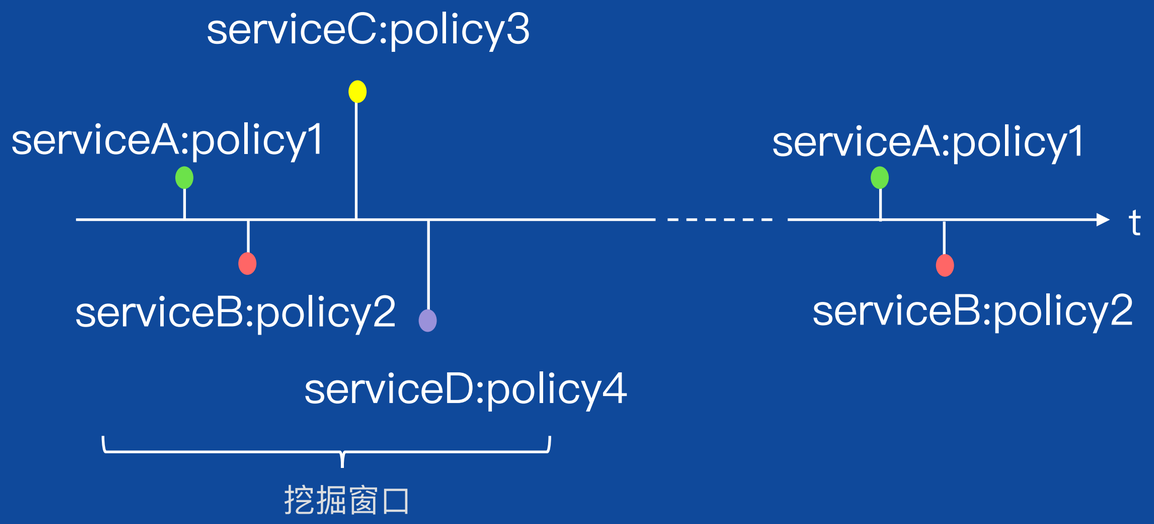
合并后

- **报警等级**
  - 严重
- **策略信息**
  - service.a.all:instance:FATAL
  - 实例类型报警
- **异常实例个数**
  - 100
- **异常实例列表**
  - 0.opr-zty5-000-cc.A.bjdc
  - 1.opr-zty5-000-cc.A.bjdc
  - ...
- **时间**
  - 2019-11-02 16:49:36
- **报警详情链接**
  - http://dwz.cn/...

报警合并

- 跨部署架构的合并
  - 上下游关系
  - 离线策略挖掘





策略关联关系表

serviceA:policy1 → serviceB:policy2

合并机制

报警源

合并缓存

报警消息 Message Info	触发时刻 Fire Time	最大等待时长 Linger Time
A:Policy1	5	20
A:Policy2	10	20
B:Policy3	15	40
A:Policy3	20	20
C:Policy4	25	60

合并后
Package

- A:Policy1
- A:Policy2
- A:Policy3

总结

- 关键指标
  - 异常检测准确率**90**%、召回率**99**%
  - 报警时效性**2**秒（99分位值）
  - 报警短信量削减**85**%
- 报警能力
  - 无数据报警
  - 报警升级、认领、回调
  - 报警合并、流控
- 商业化产品
  - 公有云监控产品BCM
  - 私有云运维产品NoahEE
  - 百度AIOps智能运维产品

AIOps智能运维公众号

架构师成长路径指南

扫码查看

持续提升 I 初级	技术进阶 I 中级	能力拓展 I 高级
邱岳的产品手记	许式伟的架构课	微服务架构实战 160 讲
微服务架构核心 20 讲	从 0 开始学微服务	Linux 性能优化实战
MySQL 实战 45 讲	技术管理实战 36 讲	左耳听风
从 0 开始学架构	Elasticsearch 核心技术与实战	Spring Boot 与 Kubernetes 云原生微服务实践

CHINA · BEIJING

THANKS

Global
Architect Summit