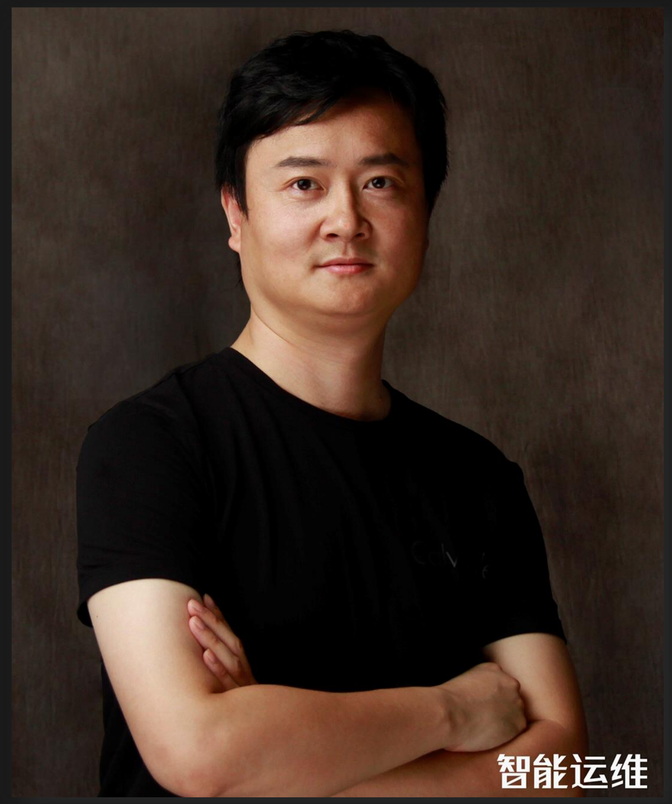
混合云全景可观测技术架构探索与实践

王肇刚（梓弋）

阿里云-基础产品事业部-混合云平台



阿里云高级技术专家
王肇刚（花名：梓弋)

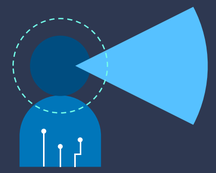
个人介绍

王肇刚（花名：梓弋），阿里云基础产品事件部混合云全景监控平台团队（前阿里集团监控平台Sunfire团队）及混合云云+应用一体化运维项目负责人。在智能监控、运维领域工作多年，一直在努力通过产品化、智能化的方式提升监控、运维的效率和能力。

内容提要

- 混合云场景下落地可观测能力的技术挑战
- 面向混合云客户的企业级监控平台技术架构探索
- 混合云可观测实战案例

从监控到可观测



监控

通过采集、分析和使用特定信息来观察判断系统状态。

.vs.

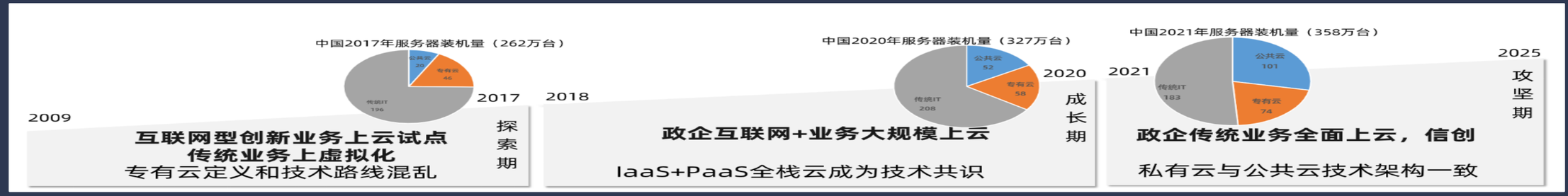
可观测

通过分析系统主动暴露或生成的数据理解和推演出系统的状态。

被动施加	从外挂式监控到内置式监控 →	主动透出
关注具体指标和现象	从孤立、割裂的指标、事件到全景、全栈化的态势感知和关联分析 分析 →	关注上下文和背后原因
关注报警和概况	从报警响应到故障全声明周期的问题排查、处置和长期优化 →	关注根因和处置方案

混合云客户运维可观测需求概览

混合云行业增长趋势明显，目前处于攻坚期



受监控（可观测）产品自身技术演进趋势影响

全栈监控	众多的NPM和APM厂商进入ITIM(IT基础设施监控）领域，监控产品供应商之间的界限越发模糊	注重分析	监控工具更多地关注数据采集（收集）和展示，并提供数据分析功能来突出产品的差异化能力。

受客户IT技术形态演进趋势影响

混合架构	监控领域的客户更多关注在混合基础架构（多云、异构网络、物联网）等领域的监控能力	成本运营	运维人员希望通过一起使用ITIM工具和云原生监控能力，来达成（成本）优化的目标。

混合云客户对可观测能力的三大需求

全栈可观测
- 指标、链路、日志三位一体

全景可观测
- 业务、应用、云实例、云平台、基础设施五位一体

服务于故障处理全声明周期的可观测能力
- 故障发现、定级、定界、处置一体化联动

如何在复杂技术栈下落地全栈可观测

应用架构差异

- 传统应用架构
- 微服务架构

技术栈差异

- 开发语言差异
- 技术框架差异

研发模式差异

- 自研应用
- ISV研发应用

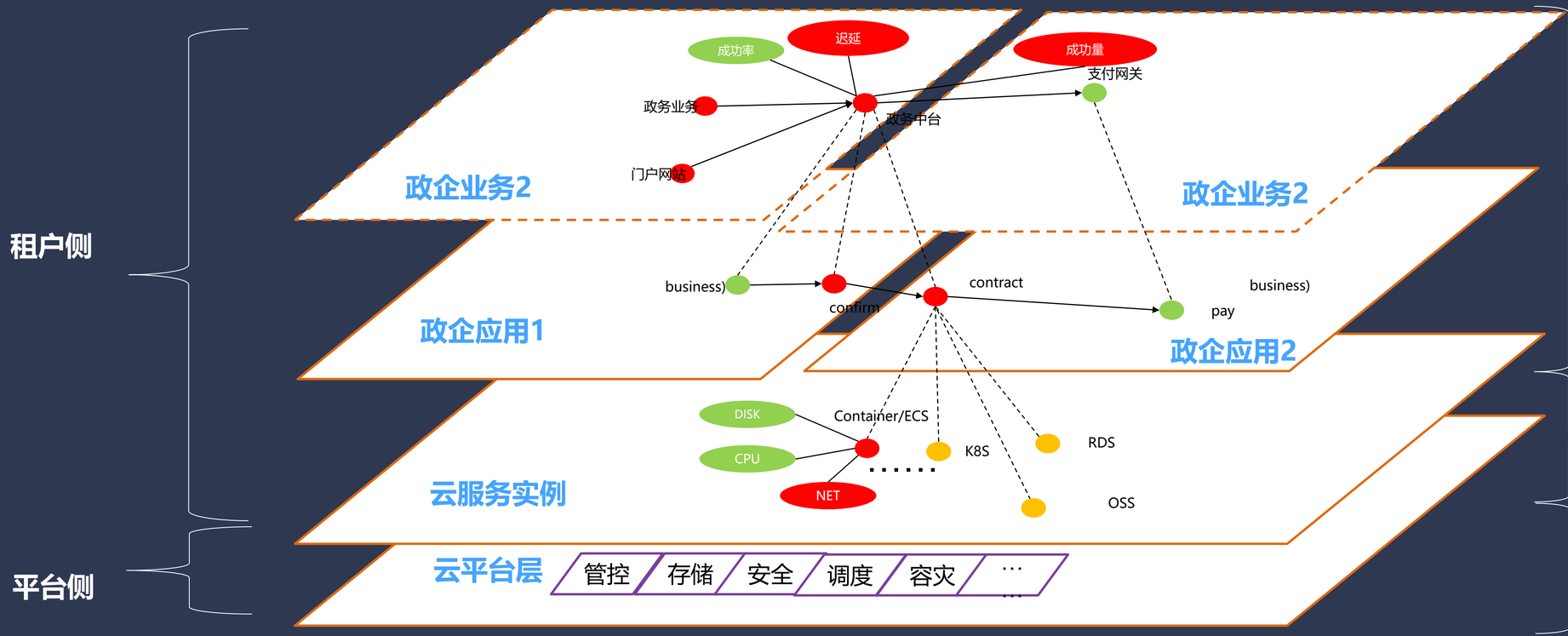
运维模式差异

- 基于物理机/虚拟机运维
- 基于容器/K8S运维

技术环境差异导致全栈监控落地困难

- 如何进行监控元数据获取或同步?
- 如何兼容监控视角和数据采集、聚合维度的差异?
- 如何满足应用性能监控的多语言支持?
- 如何观测应用自身的资源和状态?
- 如何基于应用暴露业务指标?
- 如何让监控、运维同步联动能力?
- 如何在混合架构下度量和管控资源?

如何在割裂的运维体系下落地全景可观测



获取拓扑困难

- 业务和业务之间的横向拓扑
- 业务和应用之间的纵向拓扑
- 应用与应用之间的横向拓扑
- 应用与云产品实例（中间件、DB）之间的纵向拓扑
- 云产品实例和云平台组件之间的纵向拓扑

?

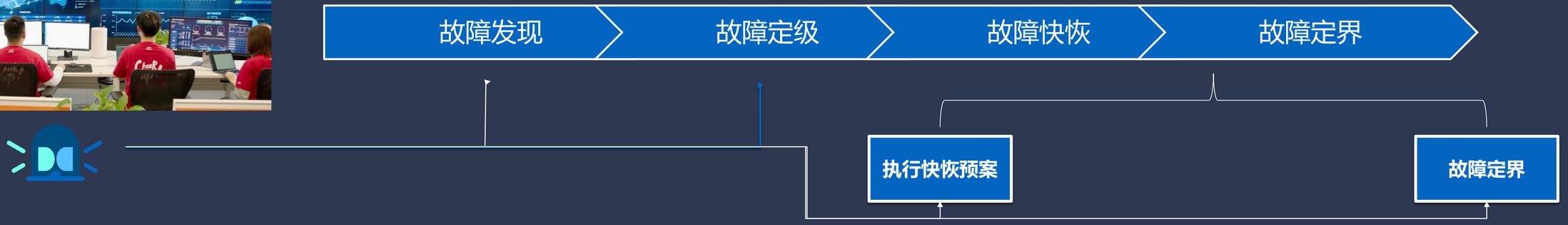
割裂层之一：应用运维和平台运维之间的割裂

割裂层之二：平台运营和平台运维之间的割裂

割裂层之三：监控报警和应急处置之间的割裂

割裂层之四：不同的垂直应用系统之间的割裂

如何让监控报警更好地服务于故障定界和处置



告警服务于故障发现

告警服务于故障定级

告警服务于故障快恢

告警服务于故障定界

- 报警风暴掩盖关键业务告警
- 故障定级难以综合技术容灾能力和业务影响
- 告警和快恢入口割裂，快恢决策依赖人工判断
- 针对不同监控对象的告警杂乱发送，无法结构化地服务于故障定界

内容提要

- 混合云场景下落地可观测能力的技术挑战
- 面向混合云客户的企业级监控平台技术架构探索
- 混合云可观测实战案例

阿里云混合云可观测产品功能架构

场景化监控能力

一体化定级 | 应急指挥 | 故障自愈 | 异常检测 | 告警根因分析 | 容量预测 | 日志故障排查 | 运营KPI看板 | 用户体验分析

事件中心

事件集成 | 事件检索 | 事件通知 | 事件处理 | 策略管理 | 事件定级 | 智能诊断场景 | 诊断编排 | 诊断报告

应急盯屏 | 应急预案 | 应急快恢 | 应急业务视图 | 应急应用视图 | 应急协同 | 运维能力集成 | IM通知集成 | 故障演练集成

业务监控

监控全景视图 | 秒级监控 | SLS集成

业务大盘 | 日志监控 | Kafka集成

业务链路 | 黄金指标监控 | SQL集成

Dashboard | 智能基线检测 | Trace集成

多维下钻 | 异常场景检测 | 生态集成

应用监控

应用全局拓扑 | API监控 | 线程池监控

应用健康画像 | 远程调用监控 | 容器&经典应用

应用Trace分析 | NoSQL调用监控 | 代码级诊断

应用核心链路 | 数据库调用监控 | Prometheus集成

异常场景检测 | JAVA异常监控 | OTEL生态集成

云实例监控（应用视角）

应用视角云资源 | 计算类云实例 | 阿里公有云集成

云实例健康画像 | 网络类云实例 | 阿里专有云集成

应用云实例拓扑 | 存储类云实例 | 异构云集成

云实例关系挖掘 | 数据库类云实例 | RPC框架集成

报警模板注入 | 中间件类云实例 | 生态集成

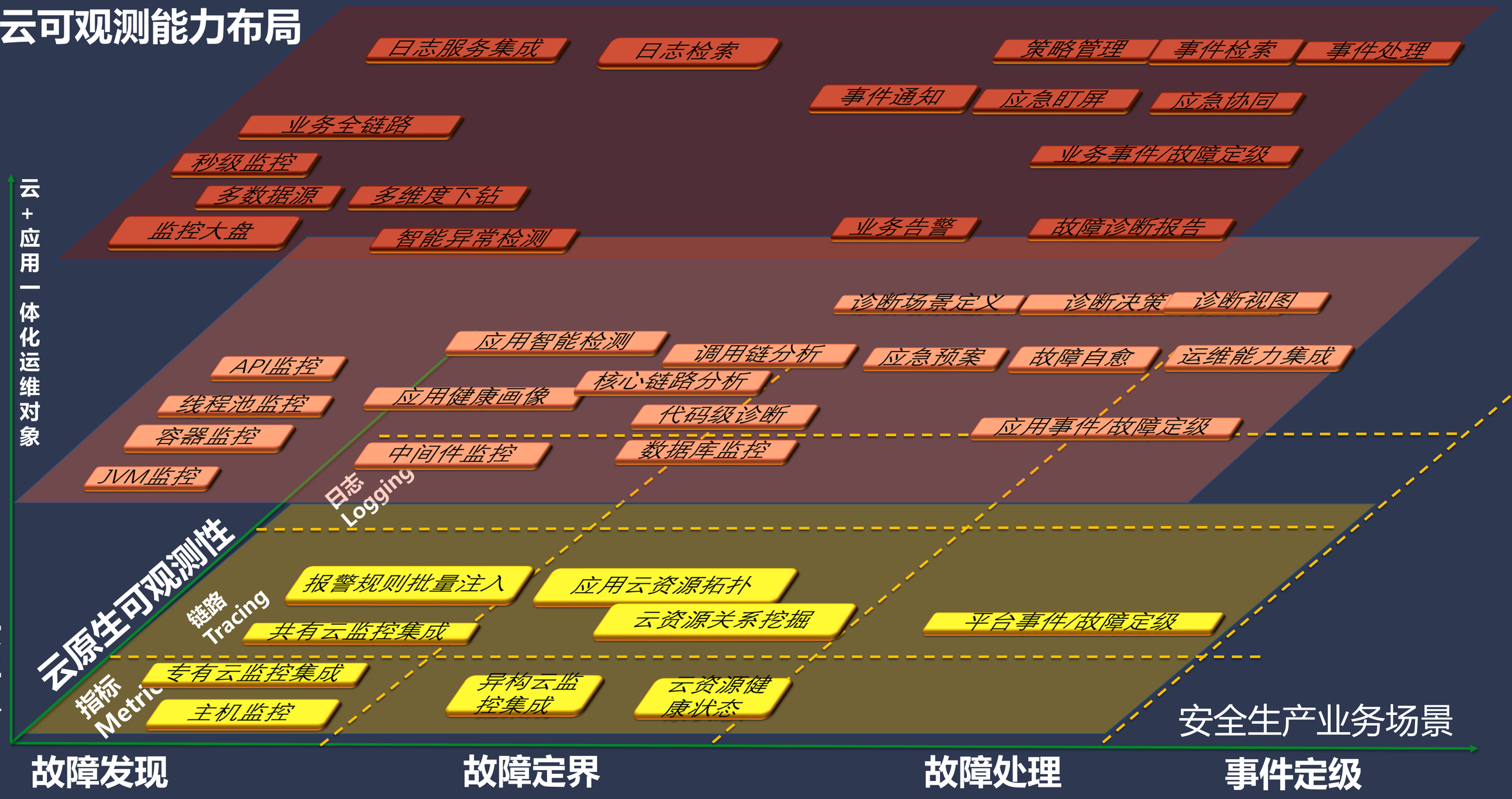
统一采控平台

统一采控框架 | K8S集成 | 国产化 | 小型化 | 调度计算引擎 | 自动化部署 | 元数据管理 | 统一存储管控 | 代理管理

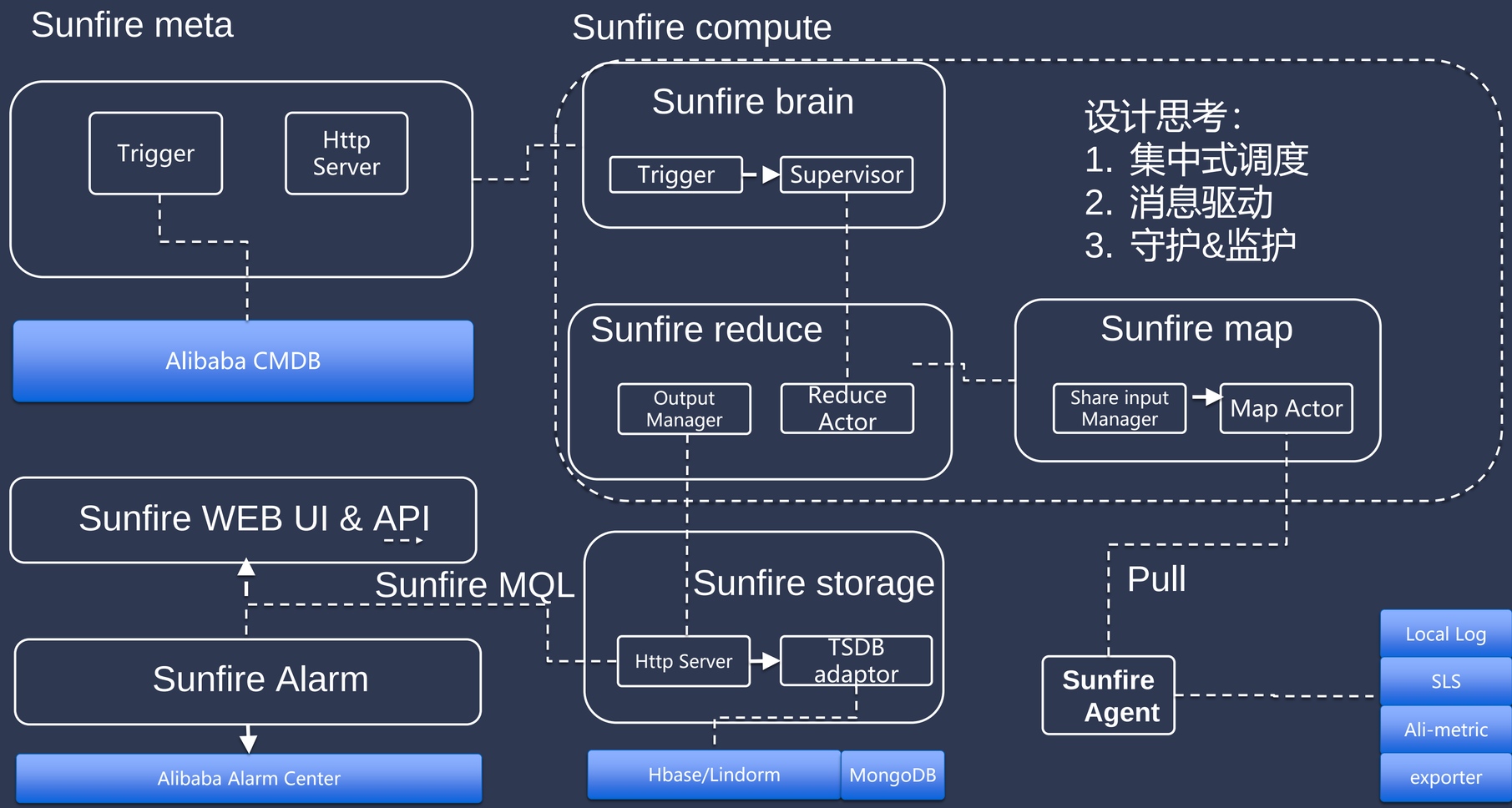
监控对象

业务系统 | 应用系统 | 数据库 | 中间件 | 虚拟机&容器 | 物理服务器 | 网络设备

混合云可观测能力布局

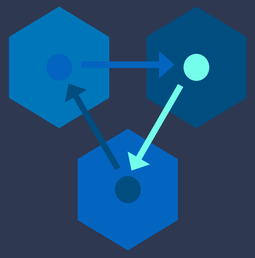

日志服务集成
日志检索
策略管理
事件检索
事件处理
事件通知
应急盯屏
应急协同
业务全链路
秒级监控
业务事件/故障定级
多数据源
多维度下钻
监控大盘
智能异常检测
业务告警
故障诊断报告
诊断场景定义
诊断决策
诊断视图
应用智能检测
调用链分析
应急预案
故障自愈
运维能力集成
API监控
核心链路分析
线程池监控
应用健康画像
代码级诊断
应用事件/故障定级
容器监控
中间件监控
数据库监控
JVM监控
日志 Logging
云原生可观测性
报警规则批量注入
应用云资源拓扑
链路 Tracing
云资源关系挖掘
平台事件/故障定级
共有云监控集成
专有云监控集成
异构云监控集成
云资源健康状态
指标 Metric
主机监控
业务
应用
云平台
云+应用一体化运维对象
安全生产业务场景
故障发现
故障定界
故障处理
事件定级

混合云可观测架构实现路径
起点：阿里集团监控平台（Sunfire）技术架构



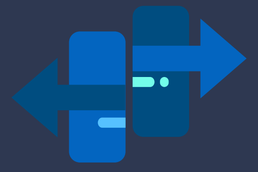
双十一期间百万级别容器日志采集规模、复杂汇聚规则、
计算核心业务指标，不超过**4.7秒**的数据迟延

监控集群自身规模（节点数量）过万~
分布在阿里集团多个数据中心，生产突袭验证
下的全局高可用



高效
基于消息的异步调度

稳定
租户隔离的分布式部署



准确
拉模式下的
数据齐全度保障

混合云可观测架构实现路径
起源：阿里集团全局故障应急背景下的监控方案

淘宝交易创建量

阿里云ECS宕机数

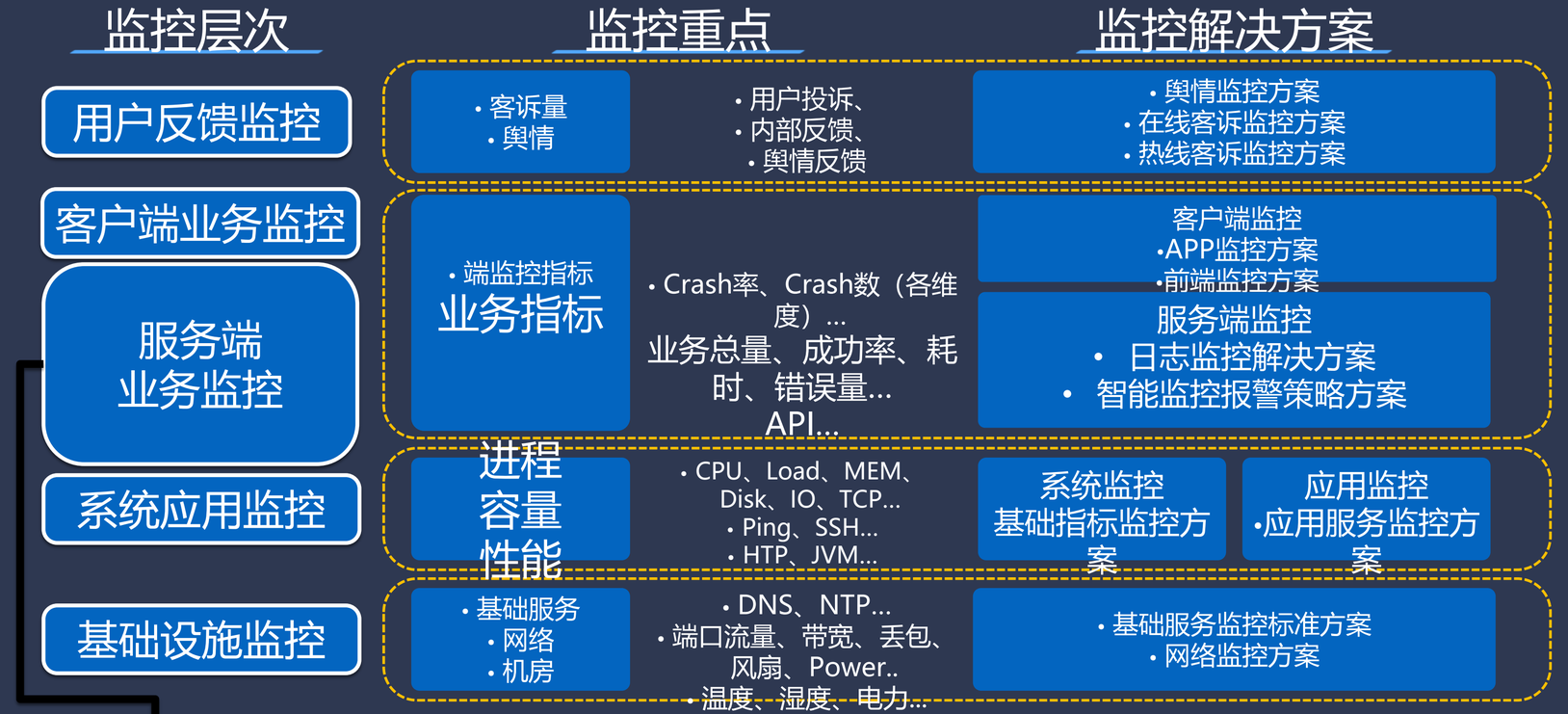
钉钉文本消息量

优酷视频全国播放量

集团故障应急由业务监控而非系统/应用监控触发

直接根据业务影响面和影响程度进行实时故障级别判定和指挥调度

不影响业务的系统/应用报警不触发全局的故障应急调度



阿里集团故障应急流程

混合云可观测架构实现路径—阿里集团监控平台转型之痛

直面Sunfire转型之痛

大规模监控计算调度和在混合云现有客户场景下并非刚需。

客户侧数据迟延较大，秒级监控几无用武之地。

客户普遍缺失业务监控的理念

客户侧技术栈不统一、部署环境复杂多变。

… …

急需补全的能力

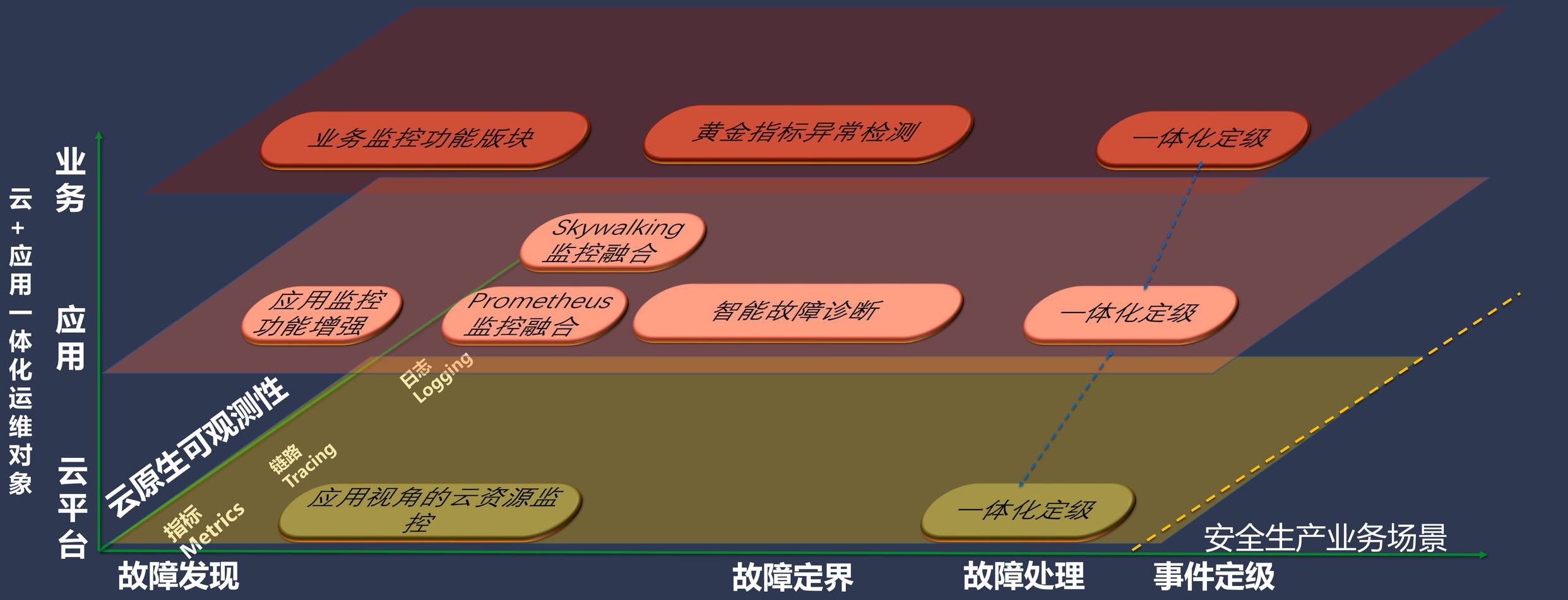
客户侧专有云资源严格规划，小型化瘦身和部署能力增强是当务之急

需要兼容全栈监控能力，增加链路监控和日志监控能力。

集成和兼容客户侧多样监控数据源和监控工具报警事件的能力。

… …

混合云可观测架构演进路线图



业务监控向三层全景监控架构演进

实时计算架构和Prometheus架构兼容

指标监控和开源链路监控系统集成

智能化框架融合和演进

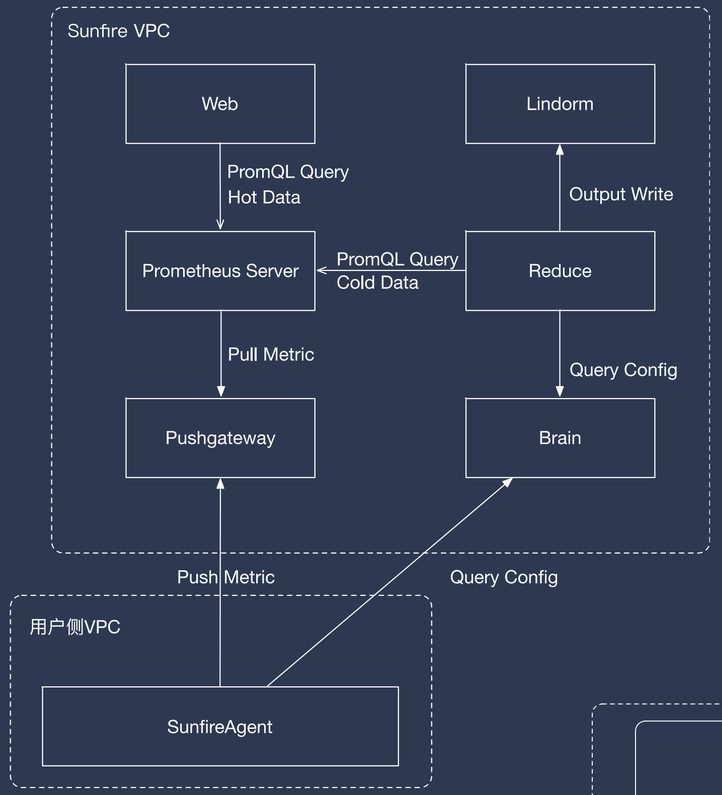
面向一体化定级的事件中心架构演进

实时计算架构和Prometheus架构融合

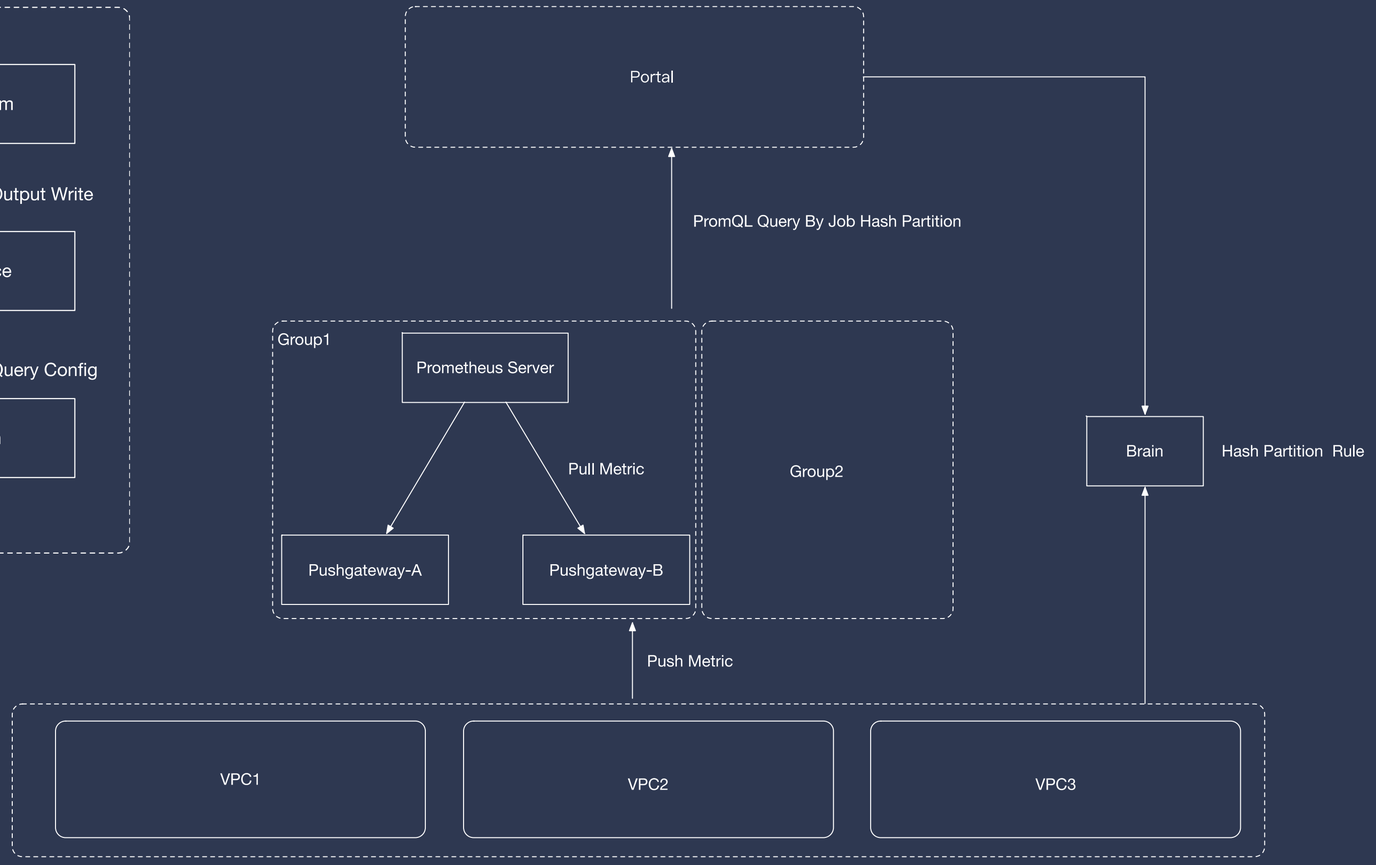
架构演进要点

- 将Promethues原生架构的计算能力和Sunfire计算、存储进行有机结合
- 通过扩展设计，提升了Prometheus的高可用能力。

Map-Reduce融合Prometheus

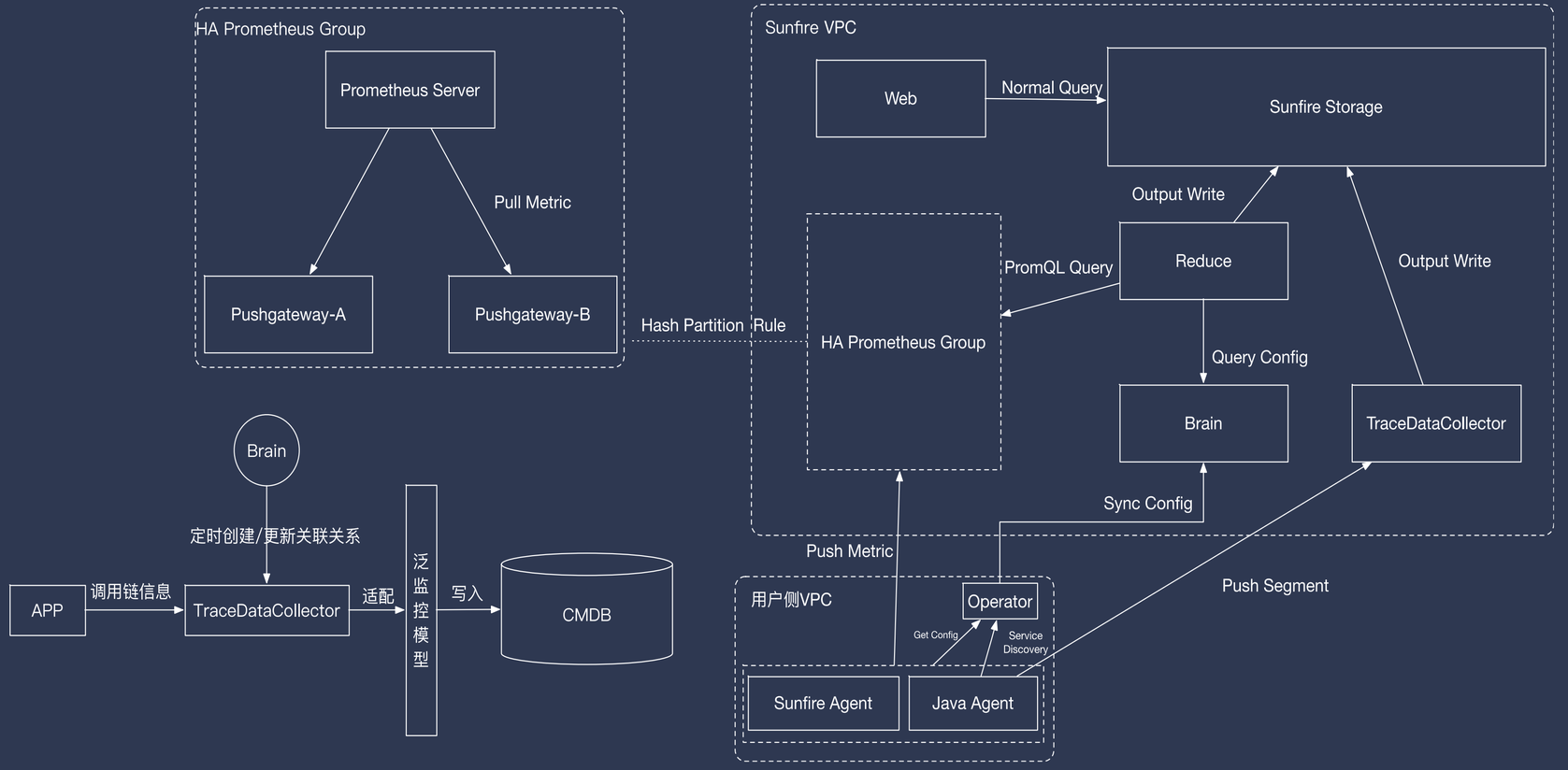


HA Prometheus Group



指标监控和开源链路监控系统（Skywalking）集成

Sunfire * Prometheus * SkyWalking -> 云原生可观测性

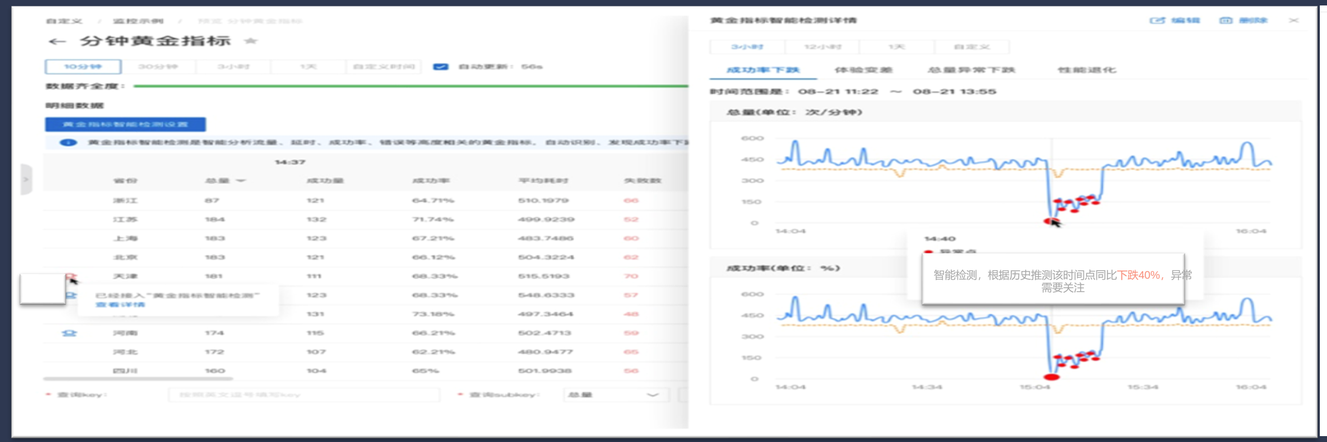


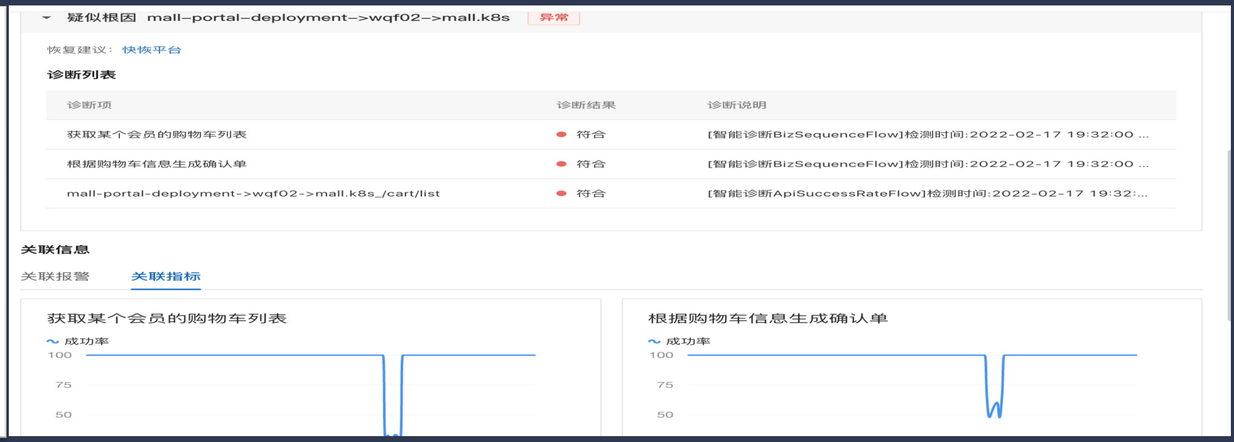
架构演进要点
调用链信息与应用性能监控指标联动
不需要在JavaAgent端加指定参数，能做到服务自发现

智能化框架融合和演进—算法功能演进

算法功能演进

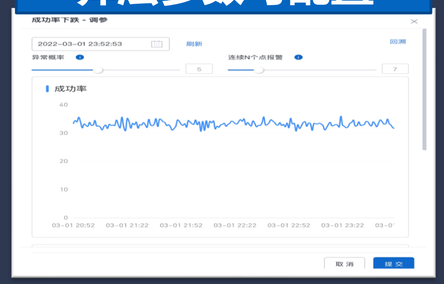
从智能基线到黄金指标异常检测，再到智能诊断、智能配置推荐、智能.. ...



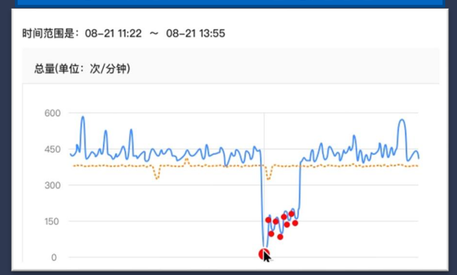


算法产品化能力迭代

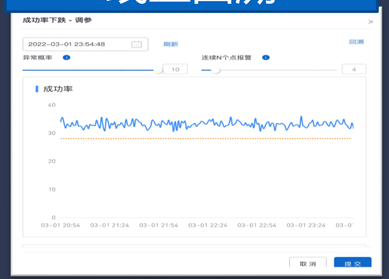
算法参数可配置



检测边界可视化



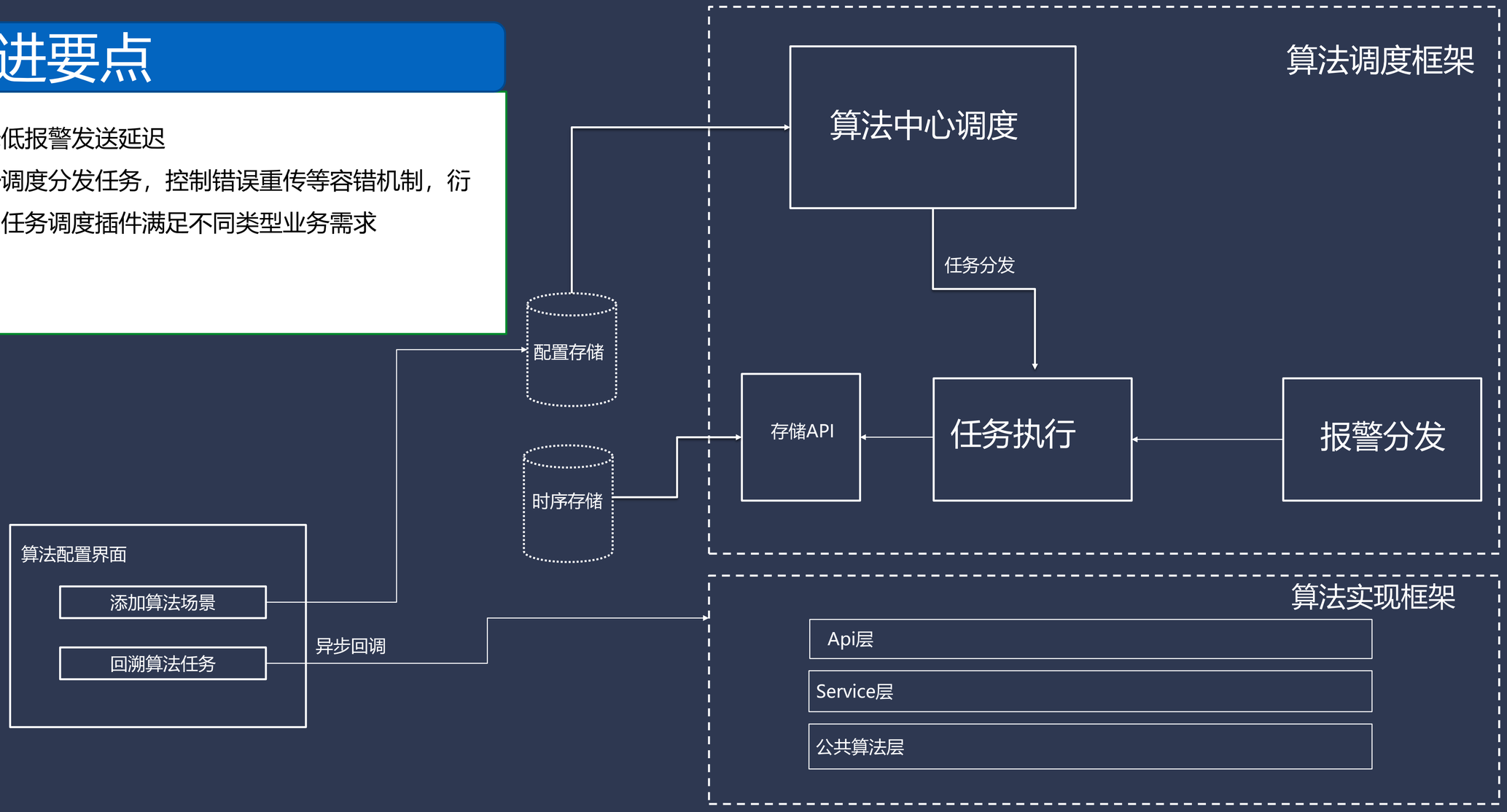
线上回溯



智能化框架融合和演进—算法工程架构演进

架构演进要点

- 存算一体化，降低报警发送延迟
- 统一调度：统一调度分发任务，控制错误重传等容错机制，衍生出不同类型的任务调度插件满足不同类型业务需求



面向云+应用一体化运维的事件中心功能布局

一体化定级解决方案

事件盯屏 | 应用视图 | 业务视图 | 故障定级

统一事件中心解决方案

异构事件接入 | 事件聚合规则 | 事件抑制配置 | 报警收敛报告

事件中心产品功能

事件接入
- K8s事件接入
- 应用层告警事件接入
- 云平台层告警事件接入
- 异构事件配置化接入
- 统一推送API
- 统一拉取SPI

事件去重
- 事件去重（无状态）
- 事件去重（有状态）

事件丰富
- 元数据丰富
- 自定义内容丰富
- 业务树丰富

事件抑制
- 按时间抑制
- 抖动抑制
- 变更抑制

事件关联
- 比例关联
- 就高关联

事件聚合
- 自定义聚合
- 维度智能聚合
- 历史关联聚合

事件处置
- 事件认领
- 相似事件推荐
- 事件诊断
- 事件处理记录

事件订阅
- 通知策略
- 接收人管理
- 逐级通告
- 接收账号管理
- 通知组管理
- 值班组管理

事件通知
- 模板渲染
- 钉钉群通知
- 渠道管理
- Webhook推送

事件数据分析
- 报警数据多维分析
- 客户成功分析
- 预警事件挖掘
- 报警收敛优化报告
- 数据导出
- 报警日报

企业级能力

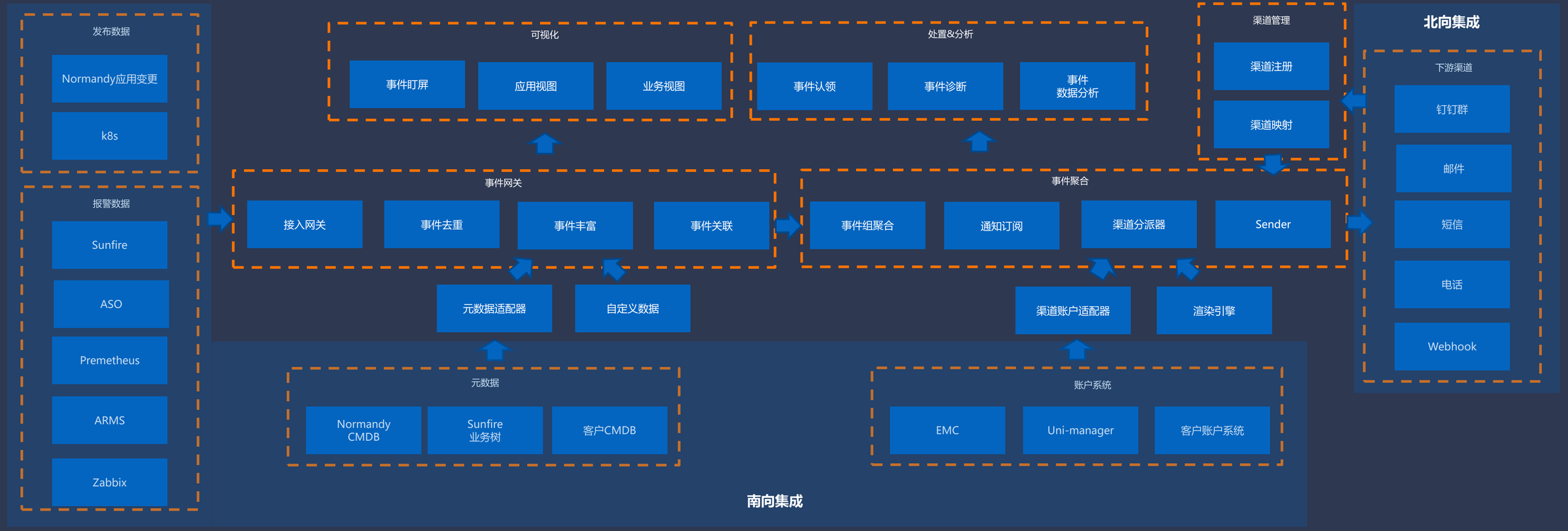
统一认证 | 权限 | 租户管理 | 自监控 | 容量管理 | 事件数据流

面向云+应用一体化运维的事件中心架构设计

架构演进要点

- 通过开放性设计和事件模型抽象，提供丰富地南向、北向集成能力。
- 结合阿里云专有云的部署方案，支持各种场景下的容灾能力。

技术架构

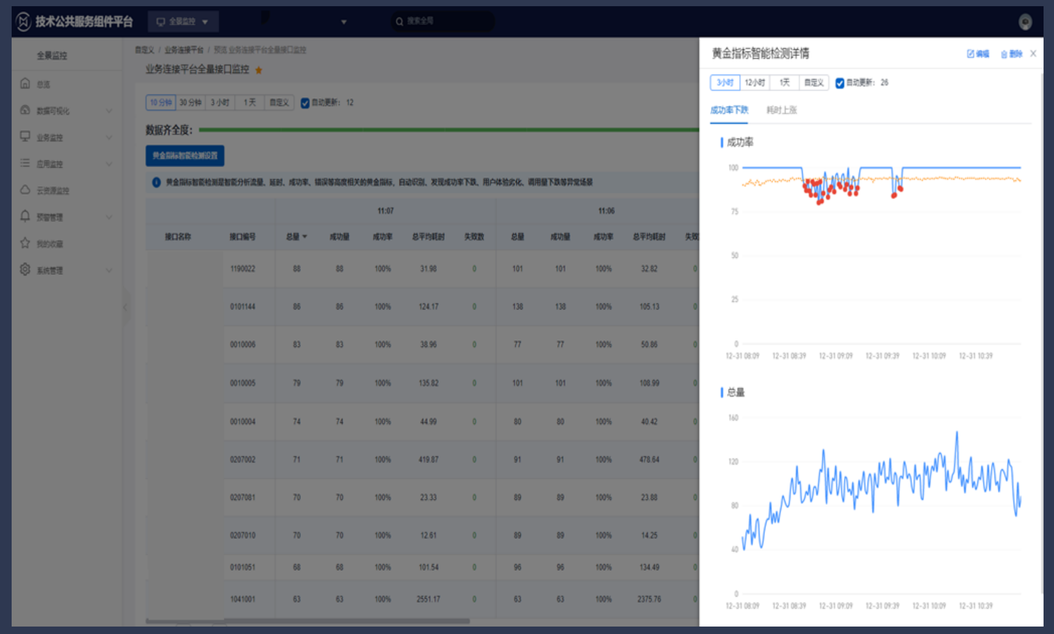


内容提要

- 混合云场景下落地可观测能力的技术挑战
- 面向混合云客户的企业级监控平台技术架构探索
- 混合云可观测实战案例

某大型能源企业监控最佳实践案例

通过建立一套总部与省侧两级监控的全景监控体系，实现从SaaS、PaaS 、IaaS 层的全面覆盖。通过对IT各层面的信息的采集以及监控报警规则的定义，不仅实现对每层监控的快速发现，快速告警，同时为监控数据分析提供全面的数据支撑。



一周告警数据汇总

某党建类项目全站监控最佳实践案例

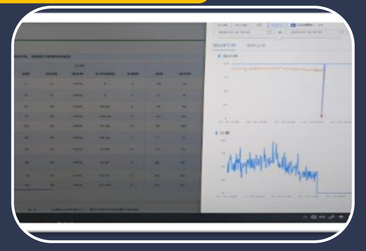
整体背景

该项目是国家级重点项目，对系统稳定性要求极高。同时整体业务板块多、规模大，集群超2W节点，入口总QPS峰值超过80W

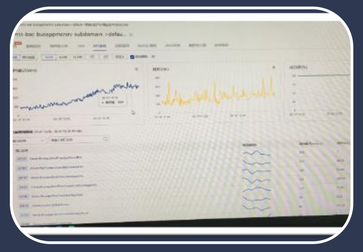
客户痛点

- 监控系统分散，无法全链路排查
- 监控能力部分缺失，出现断层
- 多套系统运行、维护成本高昂
- 现有监控系统自动化能力不足

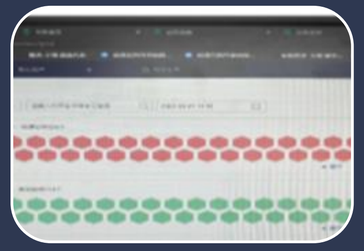
解决方案



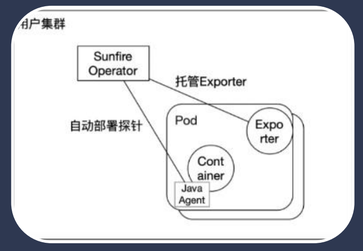
业务监控快速覆盖+智能异常检测 快速发现故障



深度链路监控能力，沉淀海量链路监控数据。



应用及云资源监控功能，定位云服务潜在异常



自动发现+自动部署，解决大规模监控覆盖问题 。

- 统一的**监控实体领域模型**，自动化的监控接入流程。
- 业务、应用、云资源**三层全链路监控**，一站式解决业务连续性保障需求。
- 自主可控的**组件化监控架构**，快速扩展私有协议监控，采用链路性能提升90%。

业务价值

- 依托自动化部署，上线1个月已接入**上万**监控指标，被监控节点达数千个，每天链路数据超过**数百万条**。
- 从业务角度出发，**建设了整套业务、应用、云资源监控体系**，确保问题能第一时间发现。
- 通过业务监控告警，结合链路排查和SQL监控，已成功发现1次线上故障，并快速锁定为SQL语句错误。

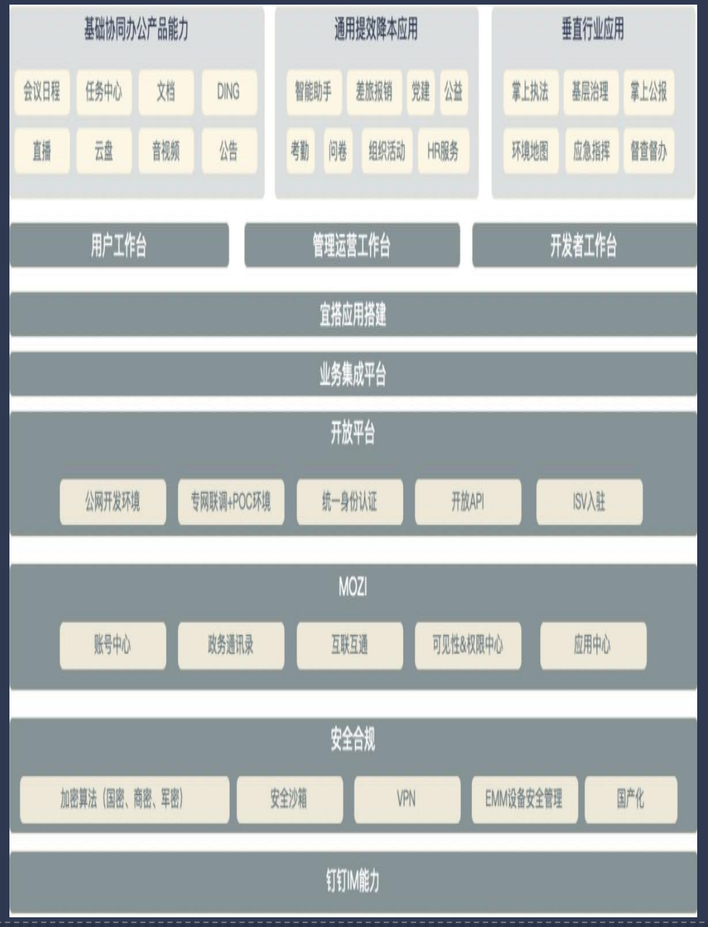
某省份政务钉、政务中台全景监控案例

客户痛点

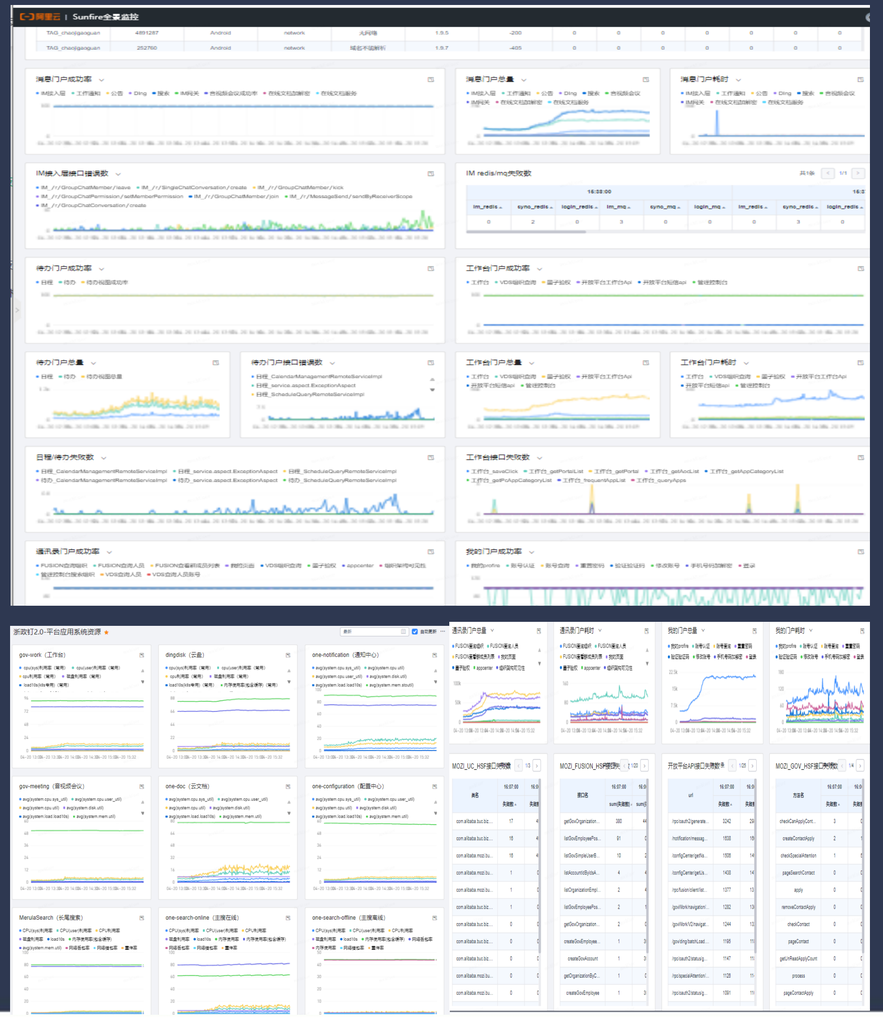
- 客户对于稳定性要求高
- 问题需快速发现及处置
- 需持续改进形成闭环
- 全链路考核管理

解决方案

业务梳理和监控覆盖



持续监控



业务价值

监控规模

- 数百个应用
- 数千个实例
- 数万个指标

- 覆盖政务钉和政务中台100%核心业务模块，实现业务、应用、资源的全面监控。

故障处理

对接阿里云故障管理平台和服务，监控发现率超90%，业务故障5分钟快速响应，重大故障15分钟快速恢复。

想一想，我该如何把这些技术应用在工作实践中？

THANKS