



# GALA-金融行业IT系统亚健康诊断利器

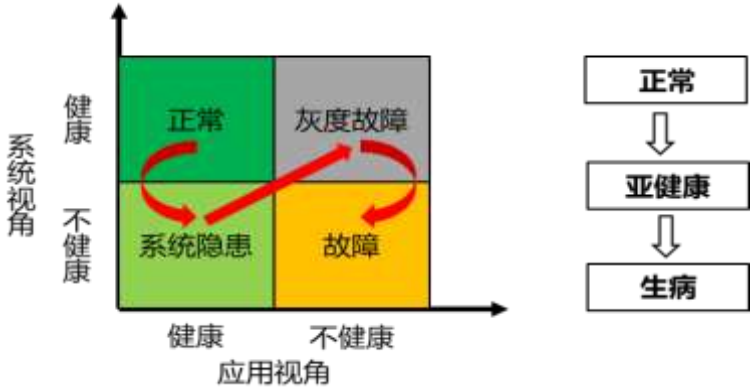
# GALA项目背景介绍



## 理论前提

**灰度故障（系统亚健康）正在成为云规模系统的致命弱点**，操作系统智能运维可以利用系统可观测性优势来实现更快的灰度故障定界、定位和自愈

【HotOS 17】 Gray Failure: The Achilles' Heel of Cloud-Scale Systems



## 业界现状与趋势

传统智能运维方案通常在应用SLI感知到故障之后再进行根因下钻，在业务负载均衡等高可用设计场景下，无法感知底层系统灰度故障这类隐患，**自底向上的操作系统智能运维存在机会窗**，弥补传统智能运维短板



## OS系统自治愿景：

- **1min定界**：确定故障位置，识别故障域（物理机/虚拟机/容器/进程粒度），指导自愈流程根据故障域选择自愈粒度；
- **3min定位**：确定故障根因，对应9大类故障（6大类软件故障、3大类硬件故障），指导自愈流程选取自愈策略；
- **5min自愈**：根据故障位置和故障根因，执行预训练的自愈策略（配置变更、快速重启、故障隔离、冷/热迁移等），完成故障快速恢复自闭环；



## 技术特征

- **智能观察**：基于AI的基础设施精准观察。
- **平台化**：基础软件的观测数据通过生态接口开放。
- **涵盖广泛的基础设施**：涵盖openEuler系列的任何类型的基础设施。

## 解决方案

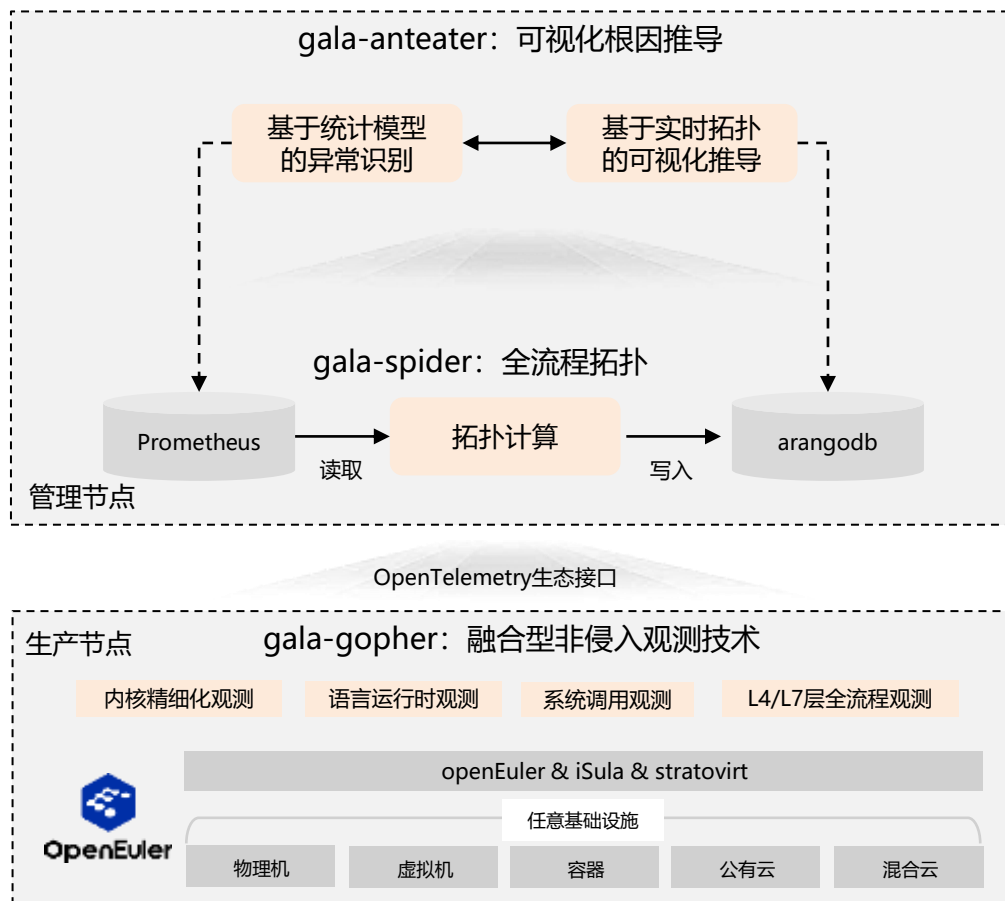
- **基础设施监控**：透明基础设施，简化运维。
- **应用性能监控**：在线识别和优化应用性能瓶颈。
- **应用安全**：实时评估和修复安全漏洞。
- **自动化**：持续和自动跟踪业务流程。

备注：  
• gala-ops系列组件隶属A-Ops解决方案，并提供A-Ops内关键基础能力（包括数据采集、拓扑感知、智能化分析等能力）

# GALA: 全栈/全链路可观测+拓扑感知/可视化推导, 实现灰度故障分钟级诊断



## 监控盲区“清零” + 织网行动 + AI推导



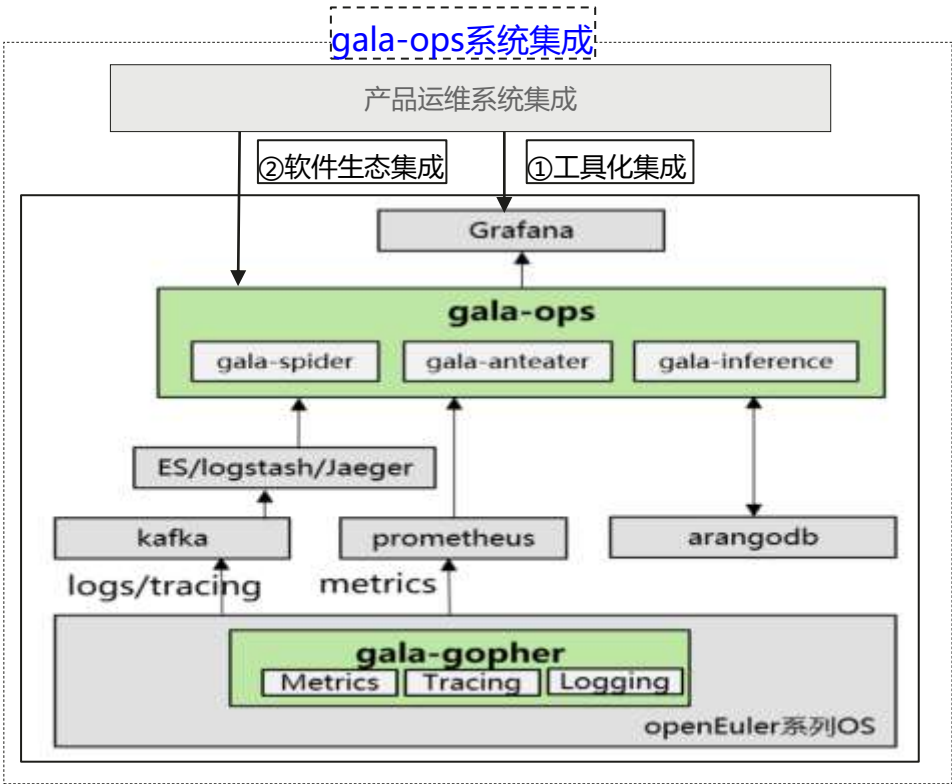
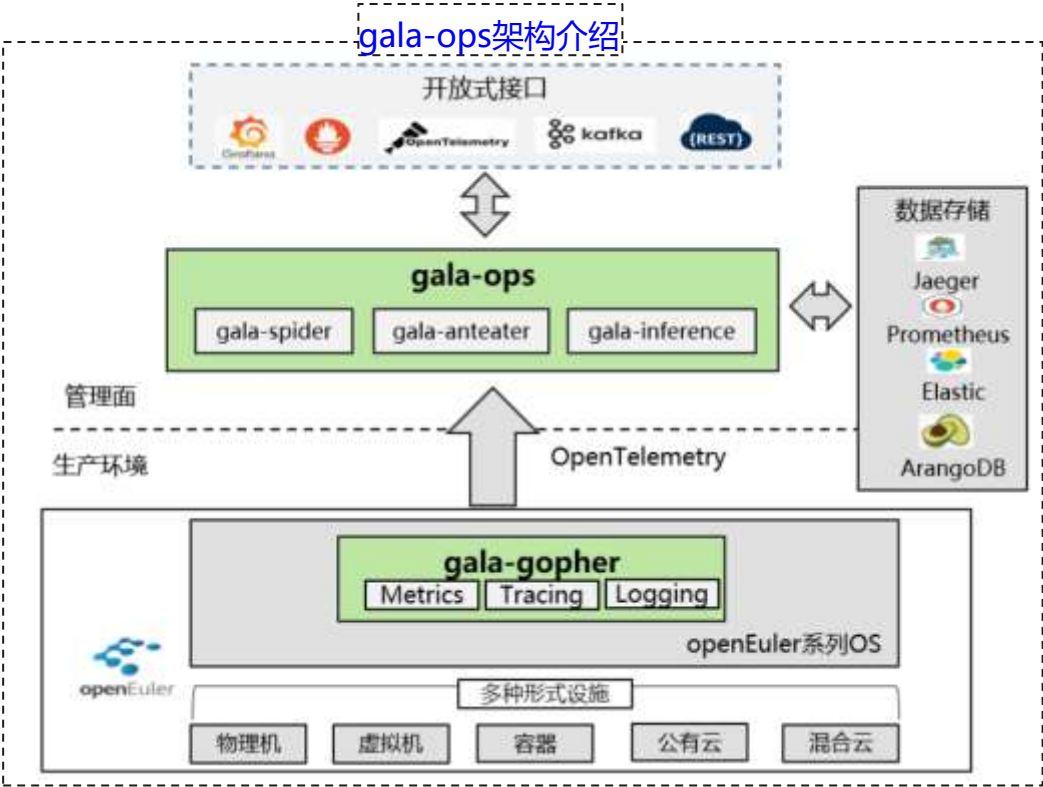
业务效果: 基础IT系统灰度故障诊断时长达到**分钟级**。

软件开销: 观测**底噪**<0.5%, 应用性能**干扰**<1%。

## 关键技术:

- 融合型非侵入观测技术: 融合eBPF、Java agent等不同观测技术优点, 实现多语言、全软件栈的观测能力。
- 全流程拓扑: 基于时序化数据 (L4/L7层流量等), 实时计算生成时序化拓扑结构, 动态展现业务集群拓扑变化。
- 可视化根因定位: 统计推理模型结合全流程拓扑, 实现可视化&分钟级的问题根因诊断。

# GALA系统介绍：易安装，开放式架构（集成与被集成），生态标准化接口

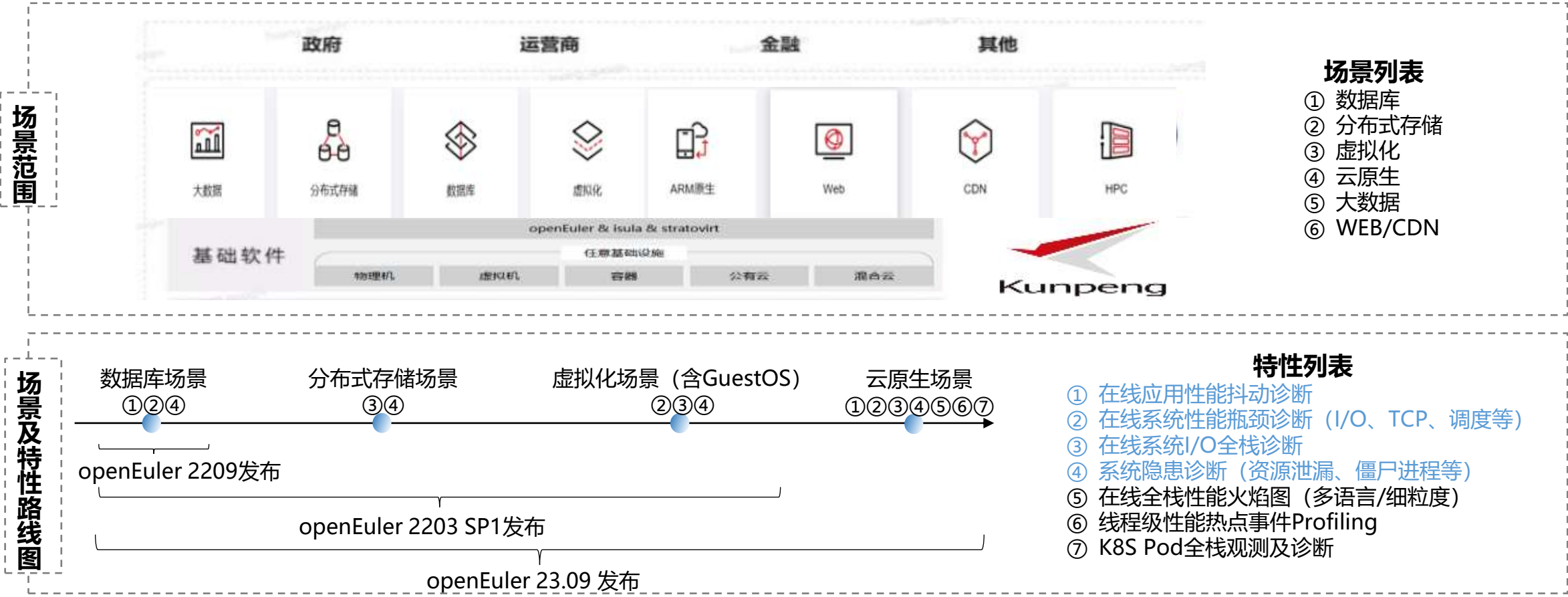


- **架构说明：** gala-ops是C/S架构，整个架构由gala-gopher、gala-ops两个软件组成。前者提供数据采集能力，后者提供数据分析能力，数据存储依赖openTelemetry行业标准标准（Prometheus、Jaeger、Elastic等），UI呈现由Grafana提供。
- **软件定位：** 作为openEuler A-Ops解决方案的一部分，致力于包括云原生场景在内的系统/应用行为、性能观测&诊断能力，[随openEuler LTS/SP版本发布](#)，并由下游OSV厂商集成（比如麒麟OS、华为云HCE）。
- **系统集成：** gala-gopher推荐由OSV厂商集成，gala-ops推荐由产品运维系统集成。
- **安装说明：** 请参考社区[指导手册](#)。

# GALA业务路线图：参考鲲鹏定义六大场景+行业客户推广计划，阶段性落地场景



- 概况：作为openEuler开源项目，提供全面社区文档支撑。
- 已发布版本：openEuler 22.03 SP1。
- 业务定位：解决政企客户IT系统内基础软件层（尤其是OS）引发的业务**灰度故障**问题。
- 已发布特性：**在线应用性能诊断（面向数据库场景）**，**系统性能诊断（TCP、I/O）**，**系统I/O全栈诊断（面向分布式存储）**，**系统隐患诊断（资源泄漏、僵尸进程等）**。
- 23年计划发布特性：**在线全栈性能火焰图（多语言/细粒度）**，**线程级性能热点事件Profiling**，**K8S Pod全栈观测及诊断**。



# GALA技术路线图：立足操作系统提供全栈观测能力



- 基本思路：基于非侵入观测技术（eBPF + Java agent等）技术立足操作系统，提供包括云原生在内场景的可观测能力，作为被集成系统，广泛的被集成进OSV、ISV厂商运维系统。
- 能力特征：围绕**系统行为可观测**、**业务全流程观测**、**应用行为可观测**、**性能观测**四个能力方向，构建基础设施监控、应用性能监控、应用安全、自动化及监控四大解决方案。
- **[1]系统行为可观测**：提供系统关键事件、性能事件、错误信息观测能力，包括Linux、Container、K8S基础软件范围；提供集群动态拓扑能力，包括Service Map、集群资源状态、应用视角软件栈拓扑等。
- **[2]业务全流程观测**：提供云原生业务全流程跟踪能力，包括分布式业务流Trace，DNS访问过程，云原生网络监控等能力，实现应用、基础设施之间的网络问题定界能力。
- **[3]应用行为观测**：提供应用视角下钻式观测能力（包括应用粒度的CPU&内存、TCP/IP数据，I/O数据等），应用性能观测能力（覆盖云原生常见协议gRPC/HTTP/RPC等），基础中间件观测能力（包括redis、kafka、DNS、PG等）。
- **[4]性能观测**：提供全栈性能热点分析能力（包括CPU、内存、I/O、Lock等），应用微基准性能分析能力（包括网络访问、文件操作、锁操作等系统性能事件），资源占用区位分析能力（包括CPU、内存等）。





# 重点场景介绍：云上线上服务性能抖动在线诊断



常见问题:

- 监控失真：如图1所示
- 监控误差：如图2所示

备注：以Redis举例描述问题，但常见线上服务都存在相似问题。

云上服务性能抖动在线诊断技术方案:

- ① 应用性能SLI监控 (如图3)
- ② 应用性能微基准根因分析
- ③ 应用性能通用监控能力

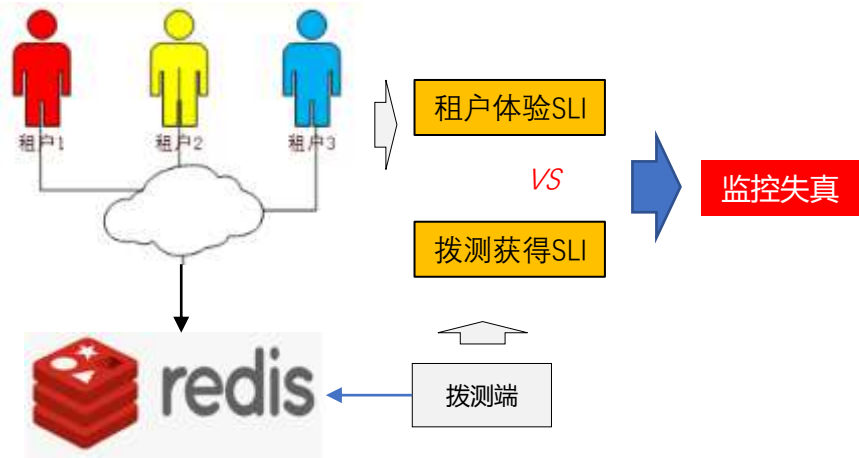


图1 DCS服务拨测系统

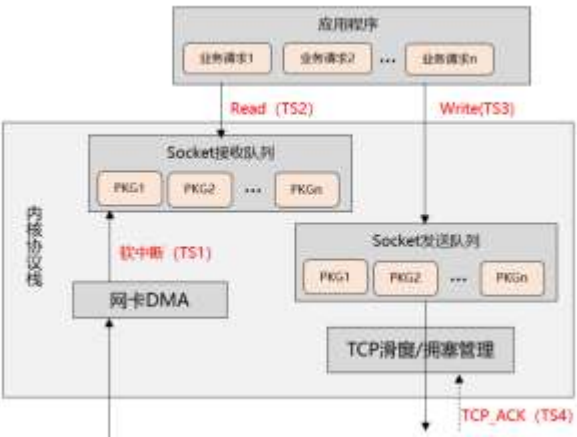


图3 应用性能SLI监控

- 应用时延性能SLI: TS4 - TS1
- 应用处理时延: TS3 - TS2
- 应用调度时延: TS2 - TS1 [2]
- 发送方向TCP时延: TS4 - TS3

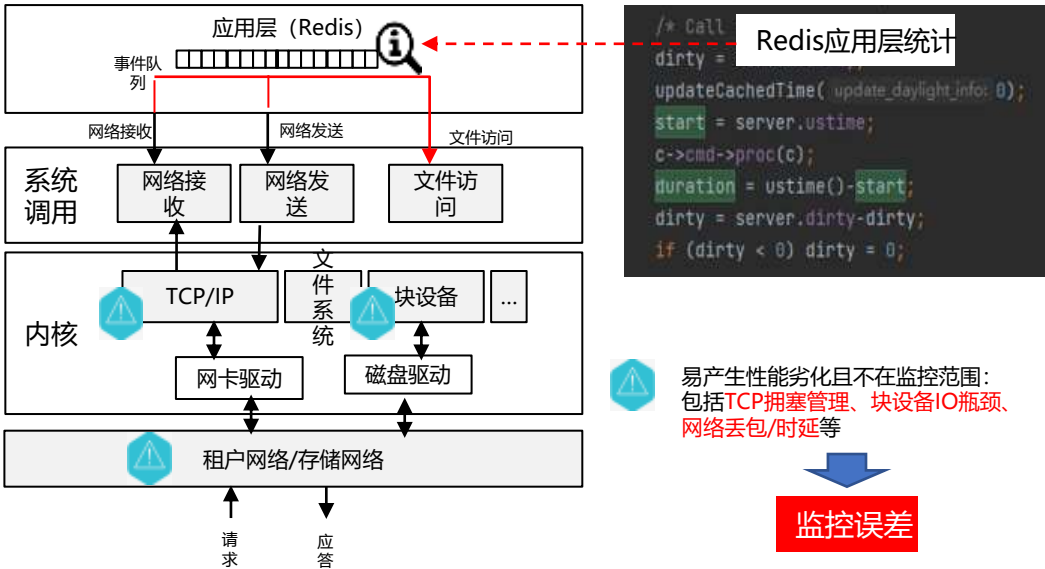


图2 应用自身监控

# 重点场景介绍：数据库全栈可观测&诊断

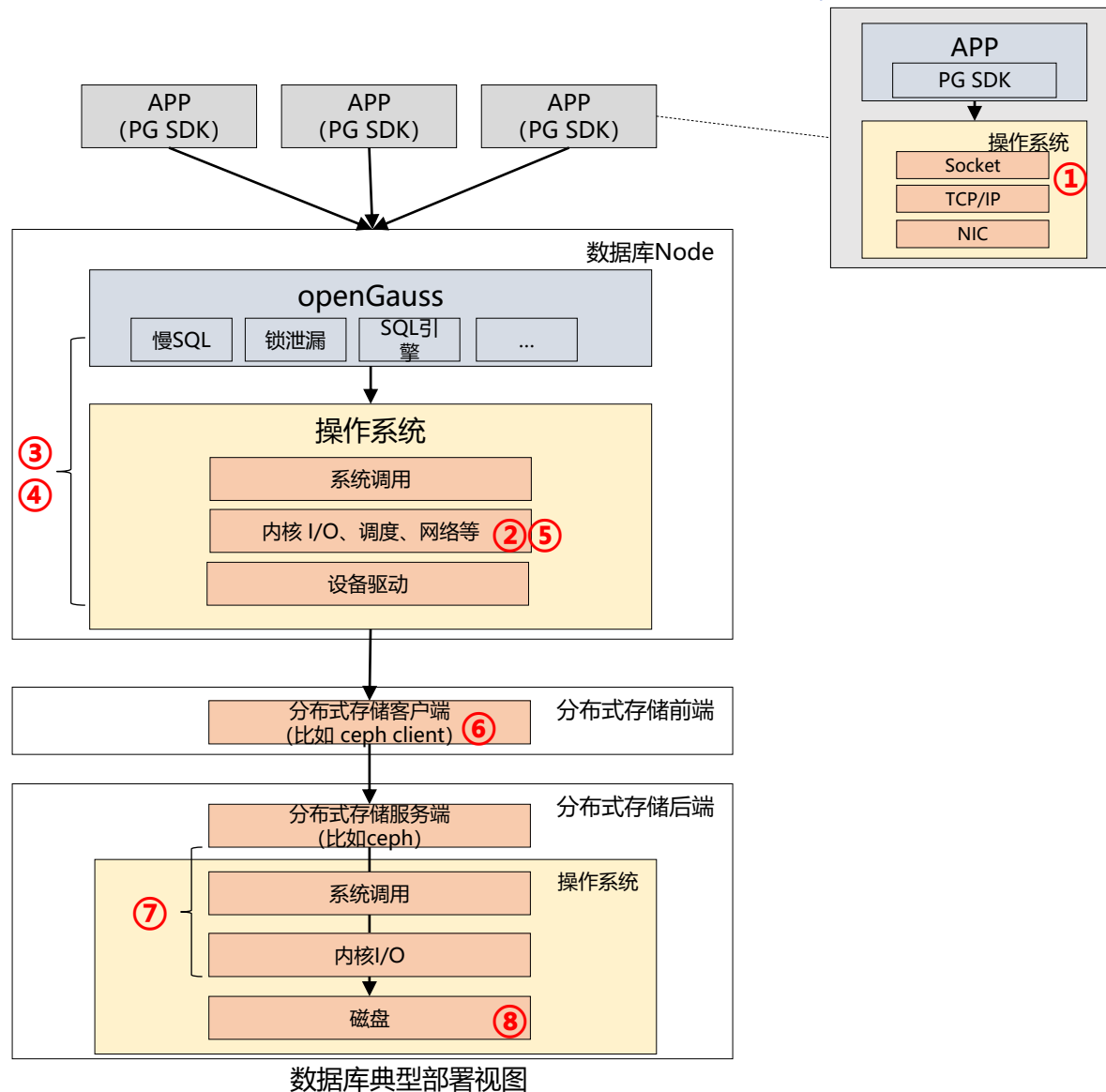


## 数据库运维面临的常见问题：

- 磁盘硬件故障
- 磁盘I/O性能抖动
- 网络I/O性能抖动
- 分布式存储I/O性能抖动
- 数据库锁泄漏
- SQL使用不当（慢SQL）
- 数据库自身算法欠优化产生性能瓶颈

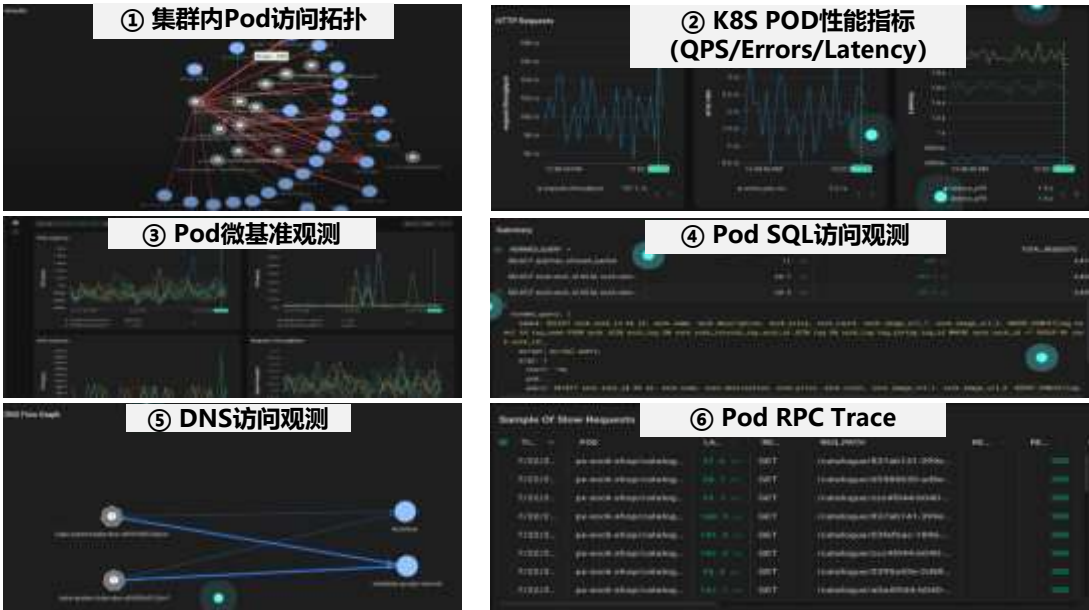
## 可观测技术方案：

- ① SQL吞吐量时延观测（应用L7层协议观测）
- ② 数据库进程TCP时延、错误观测
- ③ 数据库进程全栈性能Profiling
- ④ 数据库线程级慢SQL分析
- ⑤ 数据库进程、Block I/O时延、错误观测
- ⑥ 分布式存储客户端TCP时延观测
- ⑦ 分布式存储服务端进程、Block I/O时延、错误观测
- ⑧ 磁盘硬件故障观测

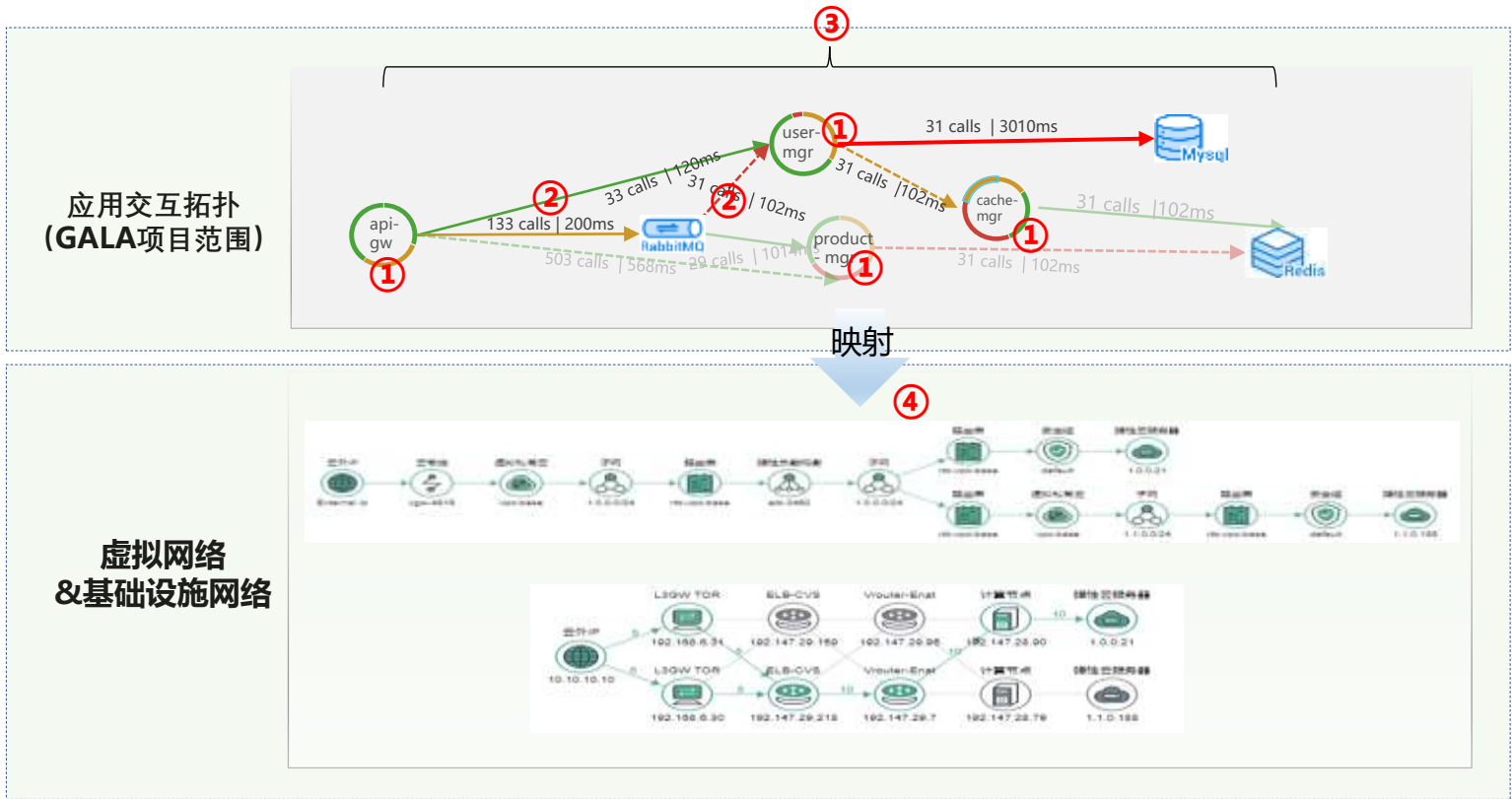




- **K8S运维面临的常见问题：**
  - 应用/资源监控割裂
  - 缺乏应用视角全栈监控能力
  - 应用与K8S基础设施之间定界困难
  - 应用与中间件（比如数据库）定界困难
- **K8S运维问题技术方案：**
  - ① 集群内Pod访问拓扑
  - ② Pod性能指标观测（QPS/Latency/Errors）
  - ③ Pod微基准观测能力（包括TCP/IO/内存等）
  - ④ Pod SQL访问观测（时延、吞吐量、慢SQL Trace）
  - ⑤ Pod DNS访问观测
  - ⑥ Pod RPC Trace



- 混合云常见问题：
  - 应用、网络监控割裂，导致应用/网络定界困难
  - 应用流trace无法覆盖网络基础设施
- 全栈云一体化运维技术方案：
  - ① 应用L7层流量时延、吞吐量、错误率观测
  - ② 应用L4/7层业务流拓扑
  - ③ 应用全流程非侵入式跟踪（见附录）
  - ④ 虚拟&物理基础设施网络拓扑与应用拓扑映射



# 特性介绍：系统可观测&亚健康巡检



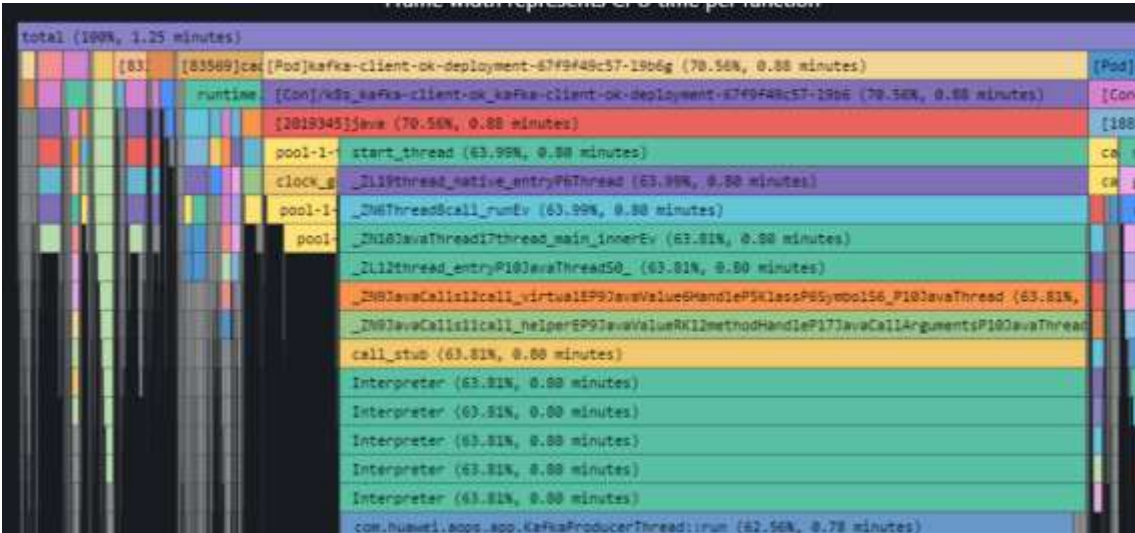
- 观测范围：[300+Metrics](#)（内核、容器、运行时、基础中间件等）；巡检范围：[50+系统异常事件](#)（TCP、Socket、I/O、disk、nic...）。

类型	指标	采集粒度 精度：>5s	标签信息	约束&范围	计划	OS版本
网络	TCP丢包/重传/乱序、建链/侦听失败、OOM、慢建链、时延	进程	进程ID、进度名、TCP五元组、容器ID、Pod Name	范围：IPv4、IPv6	23/Q1	• openEuler 22.03 SP1
	协议栈丢包（ARP/路由/Contrack）	容器/Node	丢包IP/Port信息			
	虚拟化Vring丢包	设备	VM ID			
进程	进程I/O操作分段统计、系统调用失败、BIO操作失败、I/O huang失败等	进程	进程ID、进程名、容器ID、Pod Name	范围：C/C++语言进程、JAVA进程、GO语言进程		
I/O	Block时延统计	块对象	Disk Name、Block Name	驱动范围：SCSI、virtblk、NVME、ISCSI		
	Block驱动层时延统计					
	Block设备时延统计					
内存	OOM、UCE	Node	进程ID、进程名		23/Q3	• openEuler 22.03 SP1 • openEuler 23.09
磁盘	介质错误、磁盘满	磁盘	Disk Name			
网卡	网卡丢包/错误/拥塞，TC丢包、队列长度	NIC	NIC Name			
Pod/Container	Cpu/Network TxRx/IO WR Bytes/vm/rss	Pod/容器	Pod Name、Container、Pod IP	容器引擎：Dockerd、Containerd	23/Q3	
应用L4/L7层观测	RPC L4/L7TxRx统计	Pod/容器	进程ID、进程名、endpoint IP、Service IP、容器ID、Pod name、Pod IP	• RPC协议范围：HTTP1.X、PostgreSQL、MySQL、Redis、Kafka、 <a href="#">HTTP2.0 (24/Q1)</a> 、 <a href="#">MongoDB (24/Q1)</a> 、 <a href="#">RocketMQ (24/Q1)</a> • 语言范围：C/C++、Go、Java • 加解密支持：非JSSE场景23Q3提供，JSSE场景24Q1提供（时间待定）		
	RPC吞吐量/时延/错误率					
	RPC跟踪（比如慢SQL）				23/Q4	• openEuler 22.03 SP1 • openEuler 23.09
火焰图	OnCPU、OffCPU、Mem性能火焰图	Node/Pod/C ontainer/进程	进程ID/名、容器ID、Pod Name	• 语言范围：C/C++、Go、Java • 能力范围：OnCPU、Mem火焰图 23Q3提供，24Q1提供OffCPU火焰图 • JDK版本：>= 1.8，包括oracle JDK、open JDK。 • 提供按需启动、设置监控范围、多实例能力	23/Q2	
线程级性能统计	系统性能热点事件统计（比如文件操作、网络访问、锁操作等）	线程		热点事件范围：文件I/O操作、进程操作（比如fork、sleep、pause等）、内存操作、网络访问、日志操作等。		
语言运行时	JVM OOM、JVM Heap、JVM缓存、JVM GC统计、JVM线程死锁、JVM内存占用、JVM对象统计	进程/线程	进程ID/名	语言范围：java，（golang规划中） JDK版本：>= 1.8，包括oracle JDK、open JDK。	23/Q2	• openEuler 22.03 SP1 • openEuler 23.09

# 特性介绍：系统性能profiling



- **当前问题：** 传统性能工具（比如perf、jvm等）存在数据采集精细度不够；全栈能力不足；语言强相关；底噪大等问题。
- **技术方案：** 通过eBPF + perf event + 系统调用事件提供性能火焰图、线程级性能Profiling。技术特征：**多种语言（C/C++/GO/Rust/JAVA），高精度（10ms），精细化，低底噪。**
- **使用场景：**
  - 锁资源泄漏导致应用挂死；
  - 应用线程之间资源（比如句柄/锁等）抢占导致性能下降；
  - 应用内存泄漏；
  - 系统sys cpu冲高；
  - 系统资源不足导致应用性能下降；
  - 应用处理逻辑（比如SQL引擎、RPC序列化操作等）性能瓶颈分析；
  - 应用异常系统调用操作（比如频繁写日志）导致性能下降。
  - ...

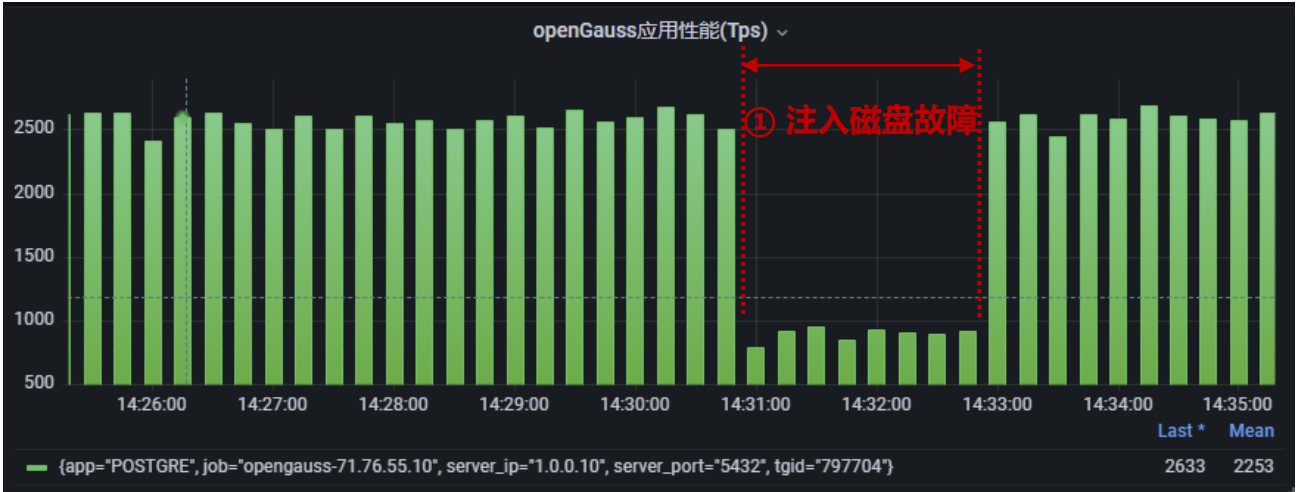


性能火焰图



线程级性能profiling

# 成果演示：openGauss应用性能灰度故障诊断

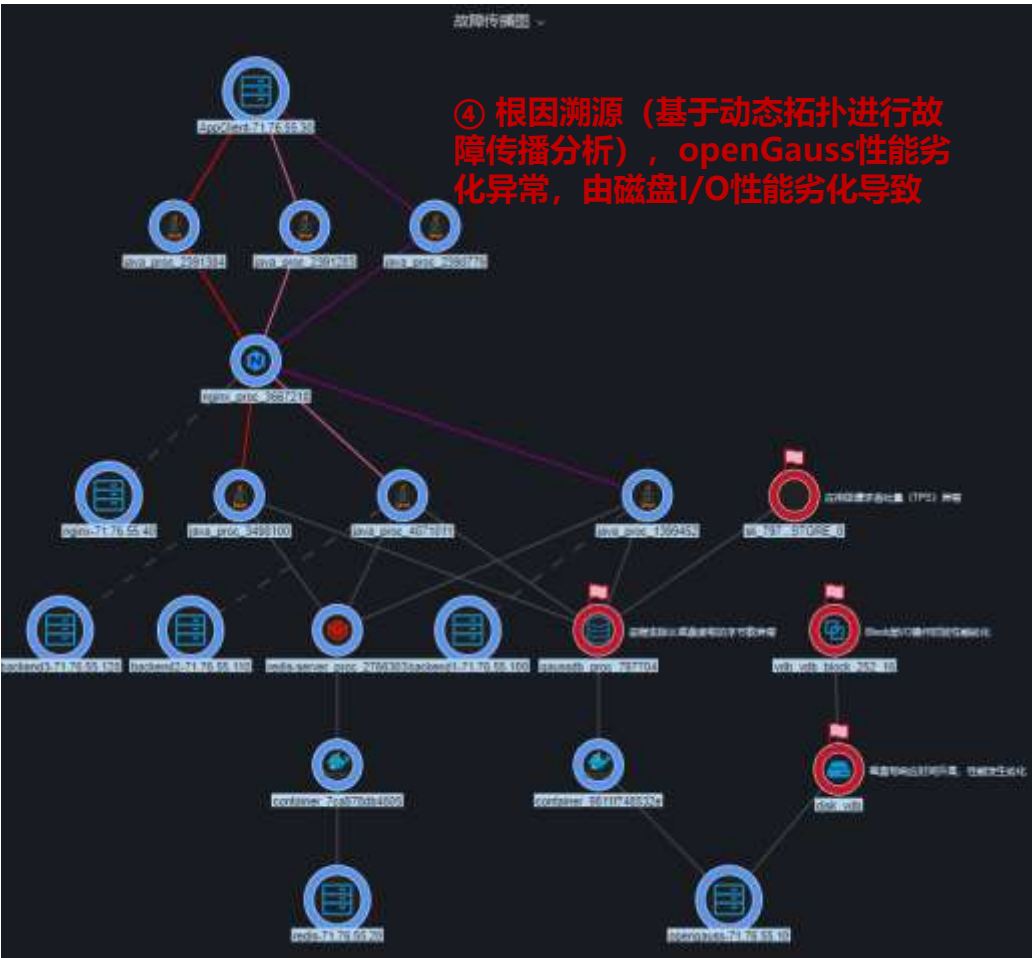


故障发现 ~

@timestamp ↓	description	EventId	② 分钟级发现应用性能劣化故障 (TPS异常)
2023-04-18 14:33:18	应用级请求吞吐量 (TPS) 异常	799598214_e7fca4b8-3c85-4842-8db8-3a90b4d11b34_sli_797704_16389_POSTGRE_0	

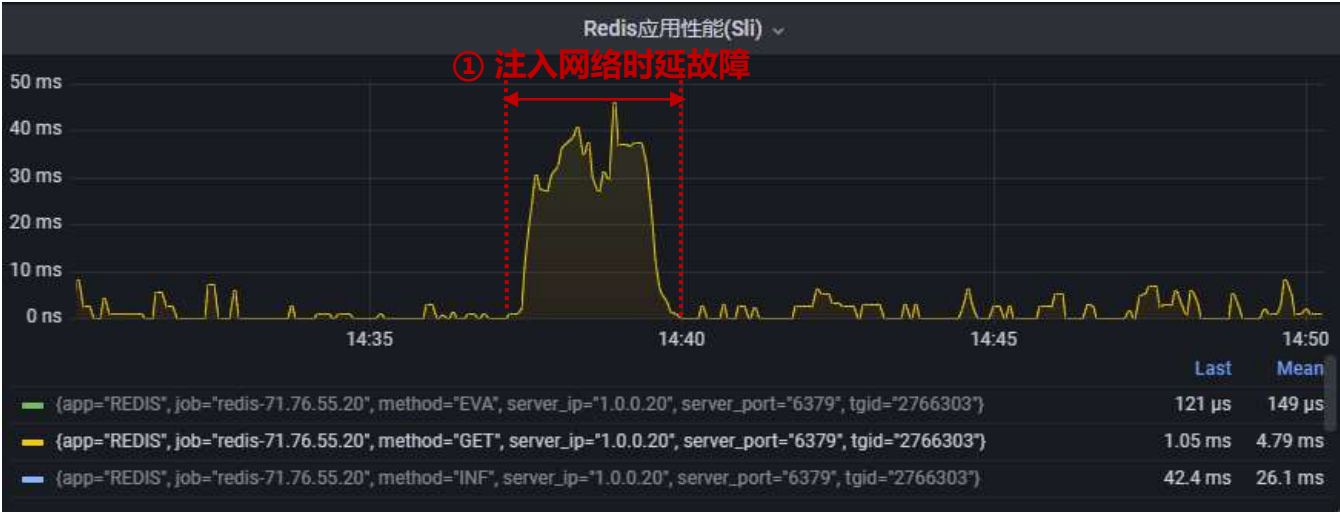
根因定位 ~

@timestamp ↓	desc	③ 分钟级定位故障根因 (进程IO异常)
2023-04-18 14:33:18	应用级请求吞吐量 (TPS) 异常, 前 1 个根因是: 1. 进程实际从磁盘读取的字节数异常;	





# 成果演示：Redis应用性能灰度故障诊断

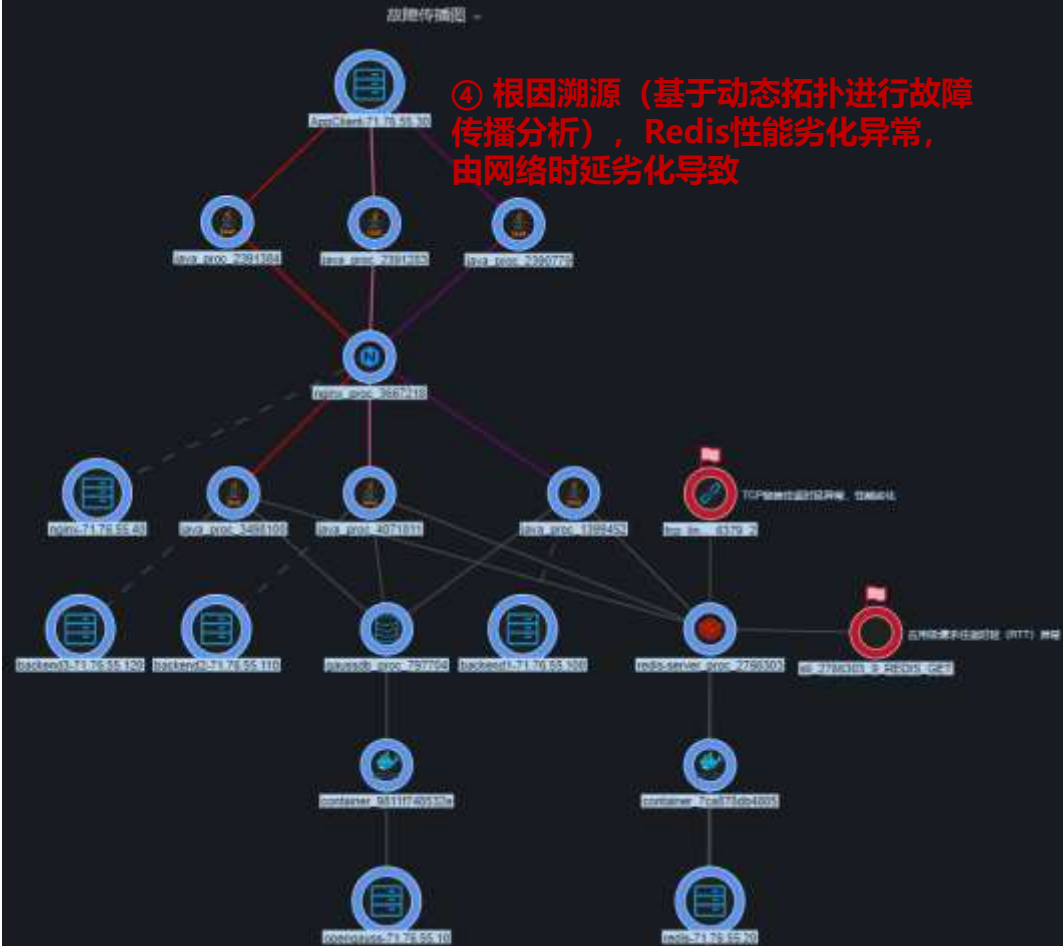


故障发现 ▾

@timestamp ▾	description	EventId
2023-04-18 14:39:18	应用级请求往返时延 (RTT) 异常	681799958213_2939e4ae-9c51-439d-9232-348853ec7061_sli_2766303_9_REDIS_GET

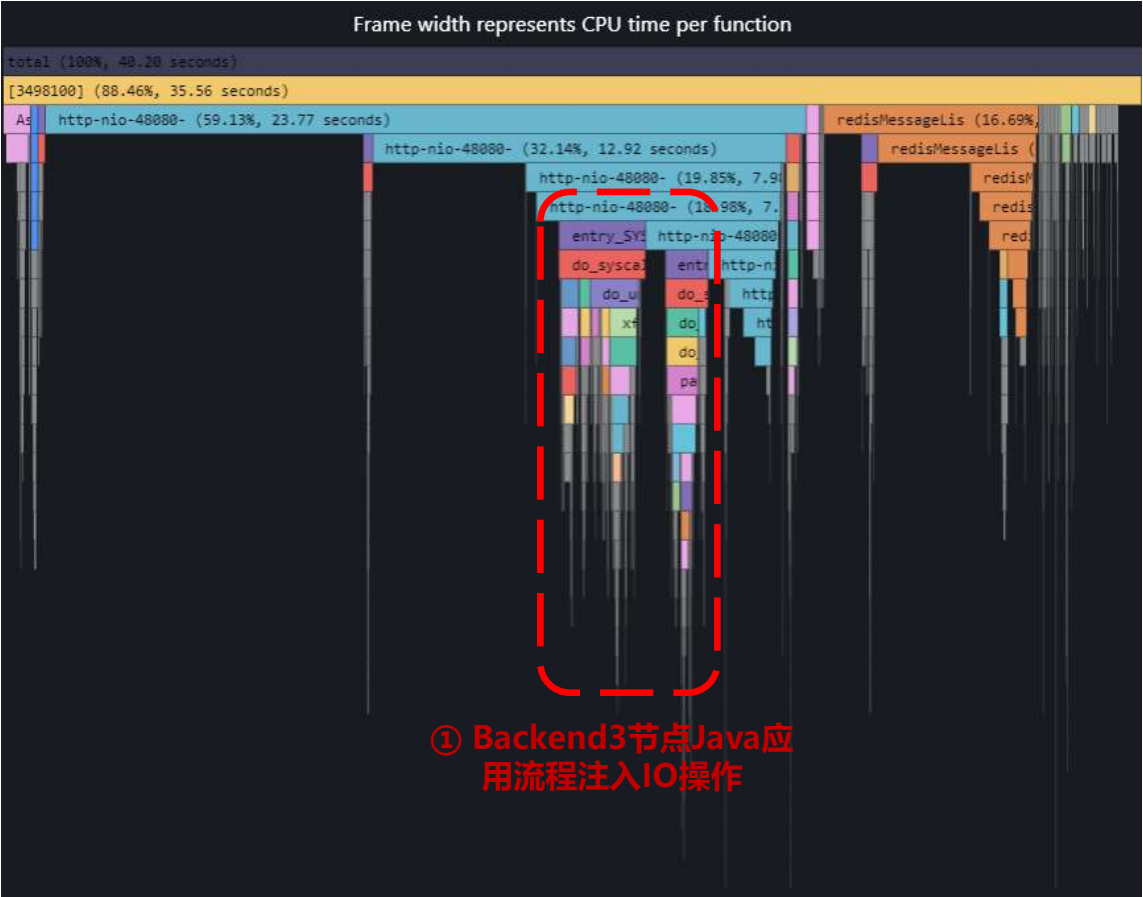
根因定位 ▾

@timestamp ▾	desc
2023-04-18 14:39:18	应用级请求往返时延 (RTT) 异常, 前 3 个根因是: 1. TCP链接往返时延异常, 性能劣化; 2. 对端应用性能或网络拥塞导致TCP发送窗...





# 成果演示：在线火焰图分析Java应用性能热点



# 欢迎加入gala-ops社区交流



项目总体介绍: <https://gitee.com/openeuler/gala-docs>

往期软文:

- 数据库场景在线应用性能诊断案例: <https://forum.openeuler.org/t/topic/827>
- gala-ops安装介绍: <https://forum.openeuler.org/t/topic/355>
- 基于性能火焰图进行性能诊断案例介绍: <https://forum.openeuler.org/t/topic/443>
- 在线性能火焰图(on-CPU)功能介绍及底噪测试: <https://forum.openeuler.org/t/topic/464>
- 在线性能火焰图(off-CPU)功能介绍: <https://forum.openeuler.org/t/topic/1312>

合作客户:



华为云



群聊: gala-ops(隶属 A-Ops)客  
户交流群



该二维码7天内(6月7日前)有效,重新进入将更新

