www.ebpftravel.com

SysOM可观测体系建设

毛文安,《深入理解eBPF与可观测性》作者

龙蜥社区eBPF技术SIG 负责人



第三届 eBPF开发者大会

www.ebpftravel.com

SysOM架构介绍



SysOM整体架构

SysOM: 一站式智能运维平台, 中心端采用微服务的方式处理 来自多个region的Profiling 和监控数据;

SysAK: 系统工具集, 处理来自中心端的请求和数据预处理, 进行诊断结果汇总和Profiling、 监控数据上传;

Coolbpf: eBPF开发框架和功能库,提供持续剖析数据采样,网络、IO、内存、调度及GPU指标等功能。





Coolbpf 整体架构



TOPP CPU&GPU业务痛点解析

- 1、进程性能瓶颈导致业务异常
- 2、偶发抖动,根因难以回溯
- 3、定位周期长,业务潜在问题仍然存在
- 4、CPU还是GPU问题难以定界
- 5、算子夯、GPU驱动夯、NCCL夯如何检测

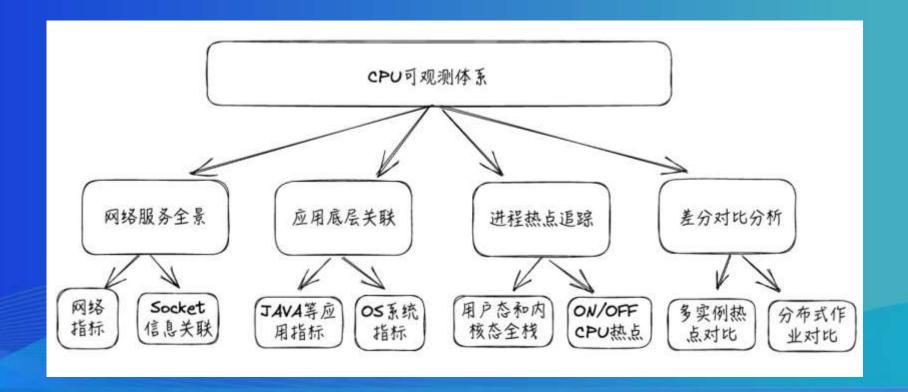


第三届 eBPF开发者大会

www.ebpftravel.com

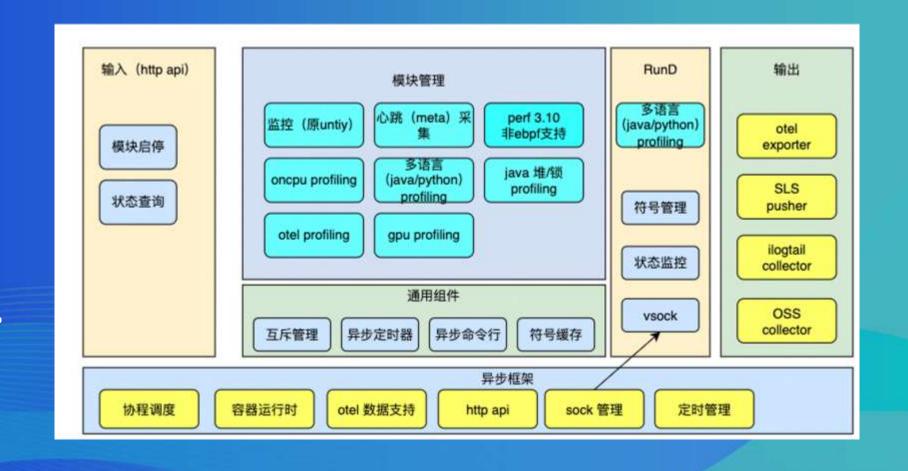
Cpu的可观测体系

で CPU侧的观测能力

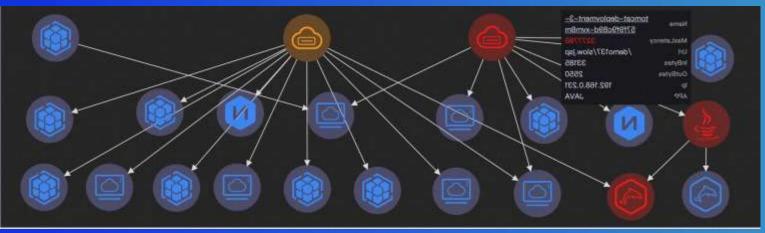


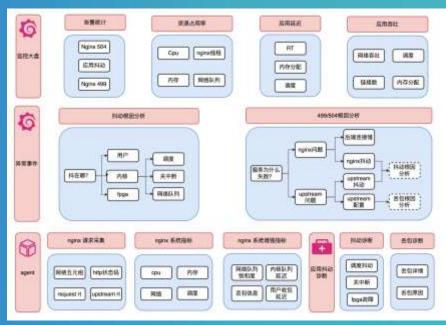
で CPU侧的持续剖析livetrace

livetrace是SysOM里面进行持续剖析的主要模块,其在节点端的架构如图所示,它采用异步机制构建了一套高度模块化的插件管理框架。



www.ebpftravel.com





- 异常事件详情

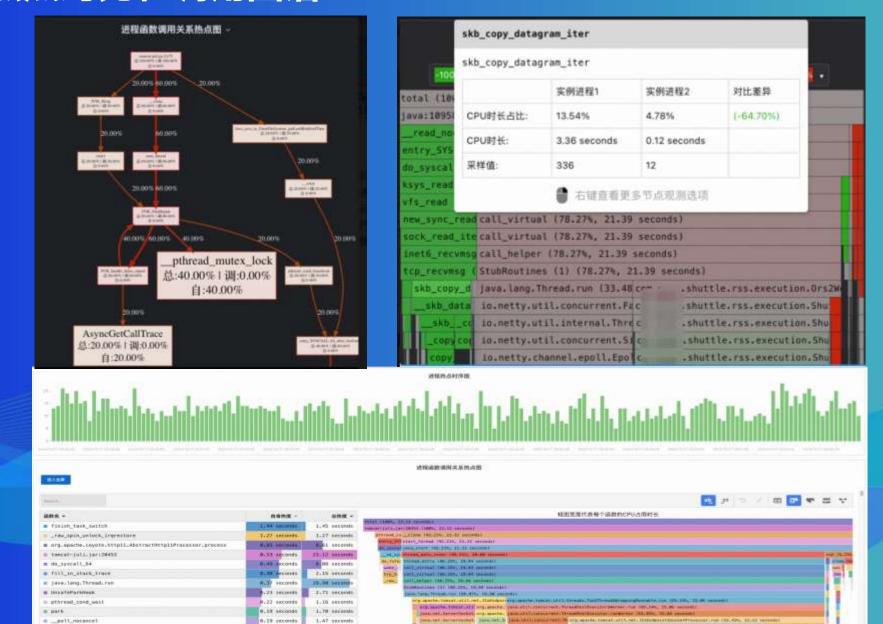
					异常请求详情			
time	remoteAddr	upetreamAddr	request	requestTime	upstreamResponseTime	matur	Peason	applicationDiagnosis
27/11/2023 16:27:28	101.200.145.40	192.168.0.76:8000	GET /polls/ HTTP/1.1	67 ms	67 ms	200	未发现网络丢包和重传	应用到动物圈
27/11/2023 16:25:52	101,200,145,40	192.168.0.76:8000	GET /polis/ HTTP/1.1	68 ms	67 ms	200	未发现网络舌包和重传	ACCURATE REPORT
27/11/2023 16:23:21	101.200.145.40	192.168.0.76:8000	GET /poils/ HTTP/1.1	30 ms	29 ms	200	未发现网络丢包和重传	应胜抖劫診靈
27/11/2023 16:23:09	101,200,145,40	192.168.0.76.8000	GET /polls HTTP/1.1	11 ms	11 ms	301	未发现网络丢包和重传	252 THE 2 4 SECTION 1885
27/11/2023 16:22:44	101.200.145.40	192.168.0.76:8000	GET /polls HTTP/1.1	11 ms	11 ms	301	未发现网络丢包和重传	应胜抖劫诊匪
27/11/2023 16:15:48	101,200,145,40	192.168.0.76:8000	GET /polls HTTP/1.1	34 ms	34 ms	301	未发现网络丢包和重传	拉斯科斯斯
27/11/2023 16:14:33	101.200.145.40	192.168.0.76:8000	GET /polls/ HTTP/1.1	12 ms	12 ms	200	未发现网络丢包和重传	应用抖动诊重
27/11/2023 16:13:19	101,200,145,40	192.168.0.76:8000	GET /polis/ HTTP/1.1	48 ms	48 ms	200	未发现网络丢包和重传	拉斯科斯斯斯
27/11/2023 16:13:10	101.200.145.40	192.168.0.76:8000	GET /polls/ HTTP/1.1	72 ms	72 ms	200	未发现网络丢包和重传	应用抖动诊断
27/11/2023 16:12:36	101,200,145,40	192.168.0.76:8000	GET /polls HTTP/1.1	84 ms	83 ms	301	未发现网络丢包和重传	成期34款股 值
					12 10 10 10 100			

m think Park

m call fortget

m Happ

www.ebpftravel.com





第三届 eBPF开发者大会

www.ebpftravel.com

GPU可观测体系

THE AI Infra可观测

围绕可观测的四大支柱进行:

监控指标: DCGM和CPU&GPU侧信息

链路追踪: 推理服务全链路追踪

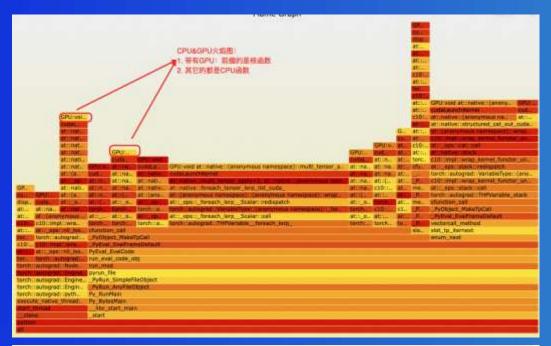
日志告警: 基于xid等异常信息构建告警

Profiling: uBPF等方式构建持续剖析能力





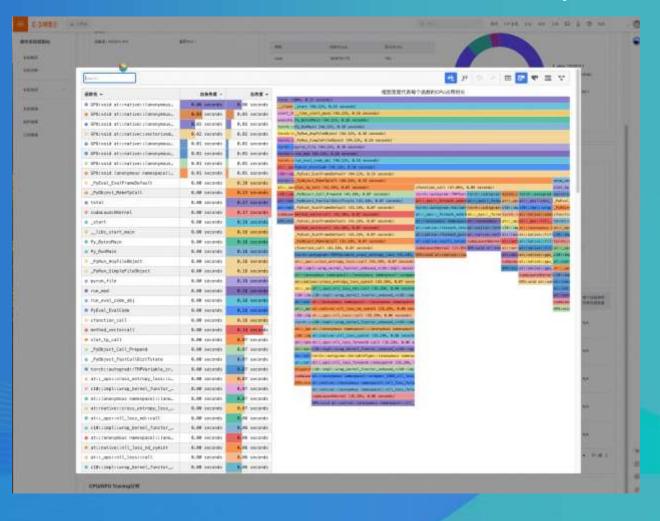
動ear 函数级别的火焰图



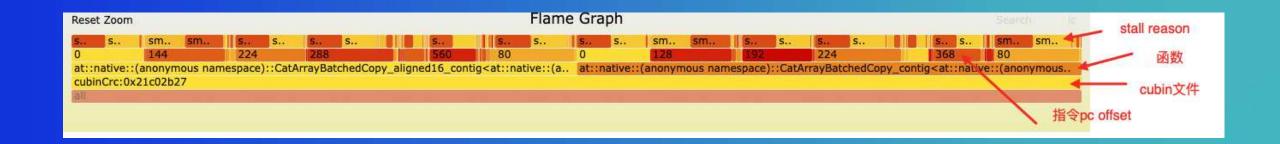


第三届eBPF开发者大会

www.ebpftravel.com



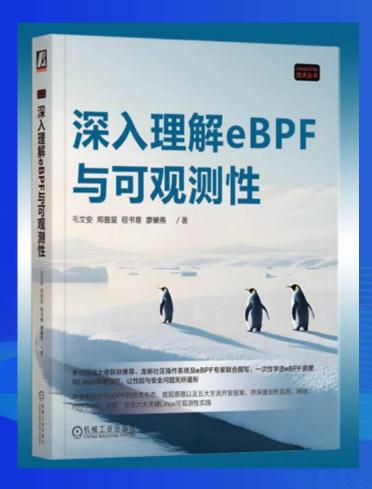
简elife 指令级别的火焰图



限制:

- 1. cupti限制:并行kernel没有correlation id,无法和cpu栈进行关联
- 2. cuda binary不携带文件和debug信息,无法获取源文件及偏移





关注右侧微信号,邀请拉入"酷玩BPF学习交流群"。在群里参与抽奖免费赢取《深入理解 eBPF 与可观测性》书籍



欢迎关注"酷玩BPF" 公众号 一起来探索eBPF技术



酷玩BPF