



GOTC 2023 全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

OPEN SOURCE, INTO THE FUTURE

eBPF 专场

使用 eBPF 实现零侵扰的云原生可观测性

向阳 2023年05月28日



可观测性
挑战

DeepFlow
方案

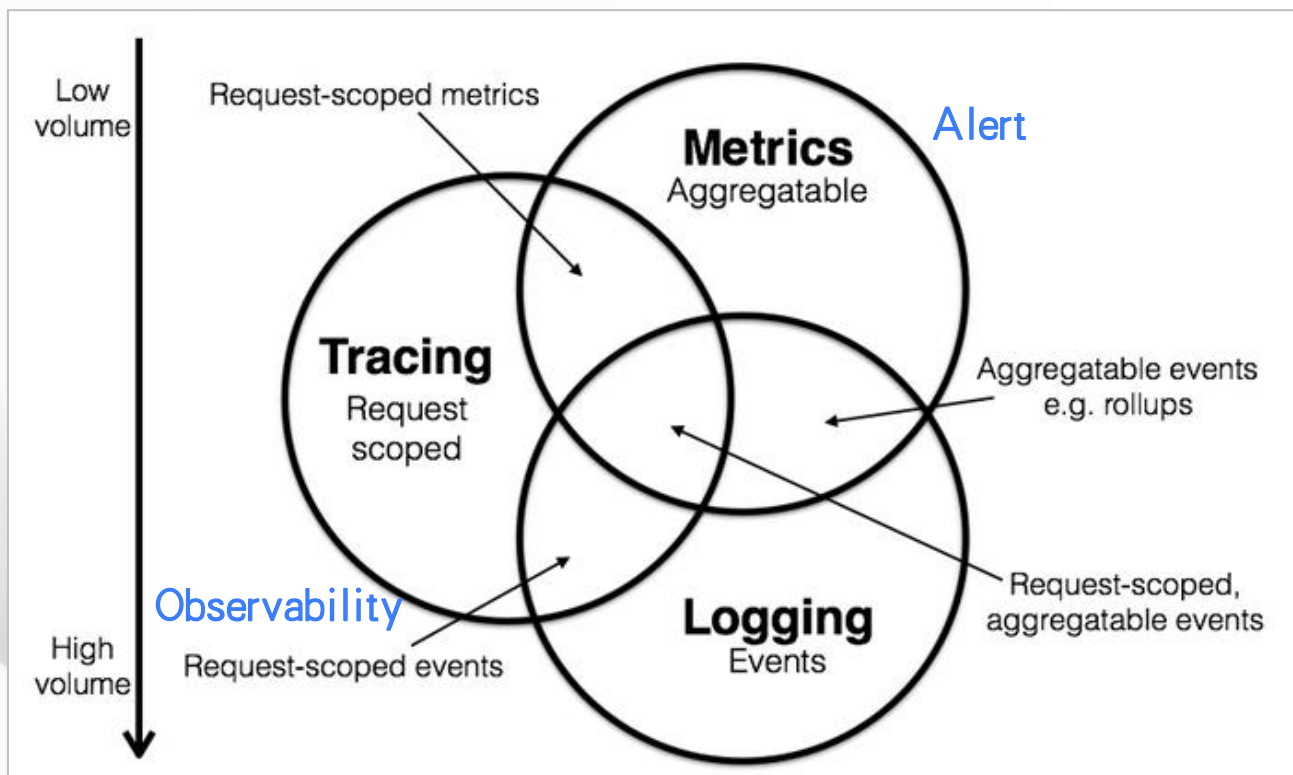
DeepFlow
用例

可观测性的灵魂，在于 Distributed Tracing

Peter Bourgon [Home](#) [About](#) [Talks](#) [Articles](#) [Blog](#)

Metrics, tracing, and logging

2017 02 21



[Tracing](#) [Metrics](#) [Ask Miss O1ly](#)

Ask Miss O1ly: To Metric or to Trace?

By [Martin Thwaites](#) | Last modified on May 9, 2023

Honeycomb.io

You lose **dimensionality** as the more metric labels you add, the harder it is to maintain the backend as metrics systems aren't built for it. Because of that, engineers tend to stick with some basic labels.

You lose the **granularity** of seeing what individual requests or processing contexts are doing as they aggregate over time.

You're not, however, using them to understand **your system and users**: you're using them to understand the infrastructure that supports them.

分布式追踪 15 年，插桩的 15 年

2023

2008



COMMUNITY



Apache SkyWalking

Application performance monitor tool for distributed systems, especially designed for microservices, cloud native and container-based (Kubernetes) architectures.

Dapper, a Large-Scale Distributed Systems Tracing Infrastructure

Benjamin H. Sigelman, Luiz André Barroso, Mike Burrows, Pat Stephenson,
Manoj Plakal, Donald Beaver, Saul Jaspan, Chandan Shanbhag

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

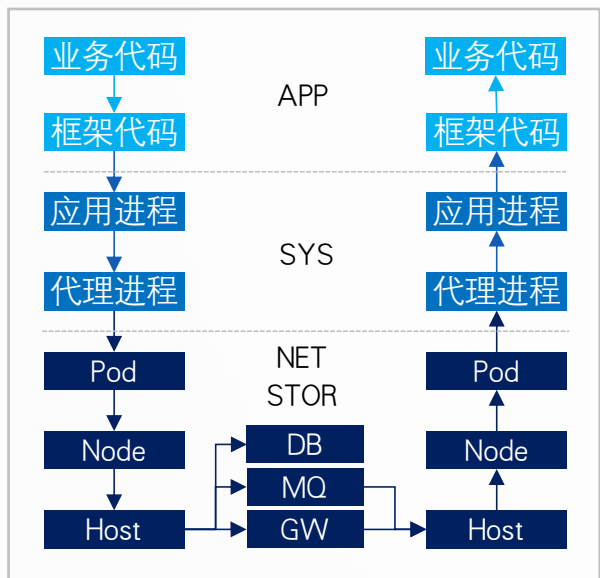
DeepFlow[®]

云原生时代，越来越多服务无法插桩

GOTC

#路径= N^2

连接微服务的基础设施
路径越来越长、多



插桩方案存在多少盲点？



服务
N



路径
 N^2



#服务=N

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

单个服务越来越简单，服务发布越来越快速
通用逻辑逐渐卸载至基础设施，开发语言和框架越来越自由

DeepFlow[®]

► eBPF 的内核可编程性是否能为可观测性带来变革



Zero Code Change
Zero Config Change
Zero Restart



Any Language
Full Stack

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

DeepFlow[®]

但事情并没有这么完美：内核和应用之间的 Gap



Zero Code Change
Zero Config Change
Zero Restart



eBPF

Any Language
Full Stack

内核可编程性



应用可观测性

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

DeepFlow[®]



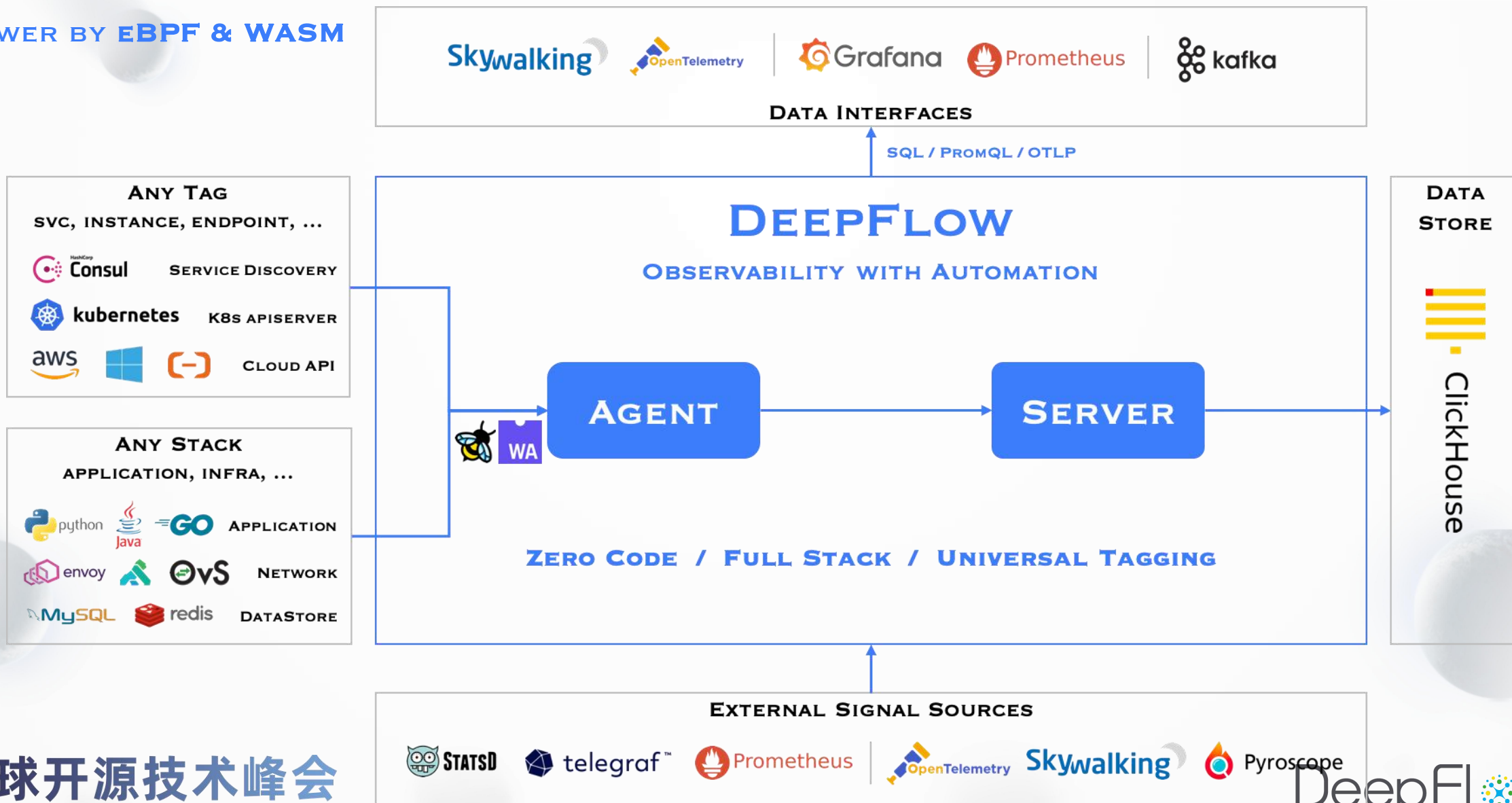
可观测性
挑战

DeepFlow
方案

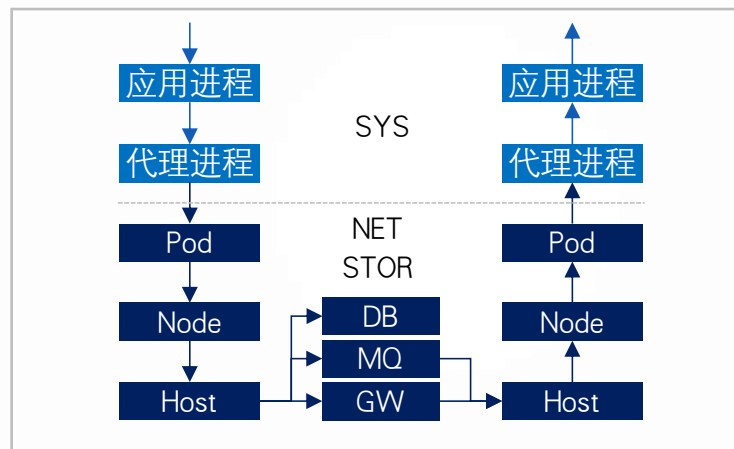
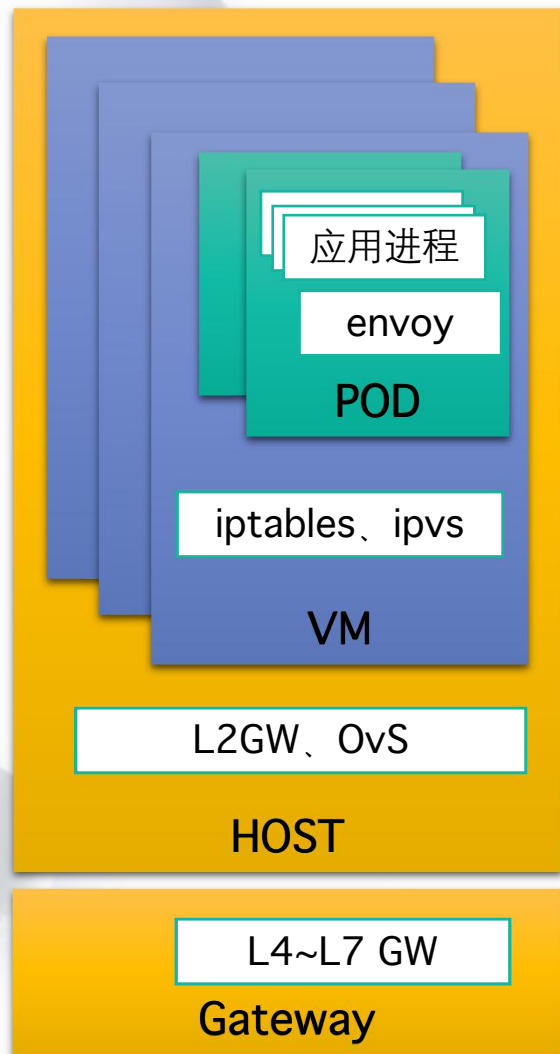
DeepFlow
用例

DeepFlow: 零插桩、全覆盖、全关联的可观测性

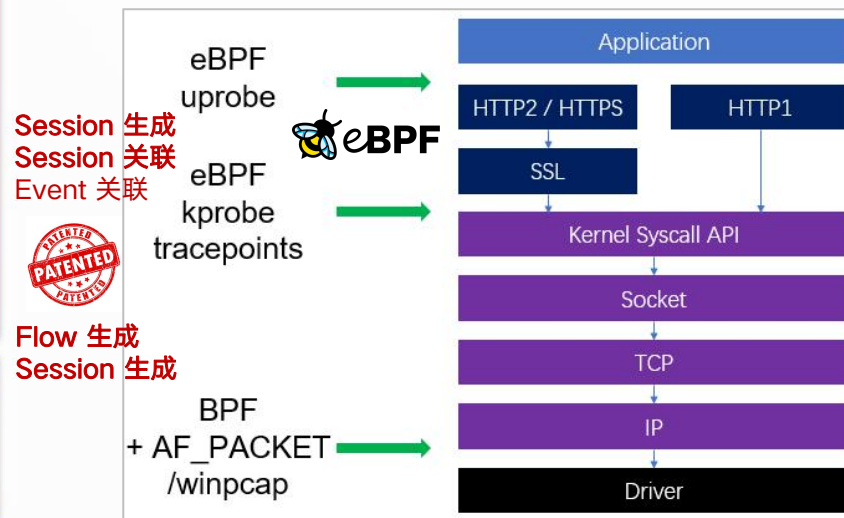
POWER BY EBPf & Wasm



DeepFlow 的技术制高点：零插桩 Zero Code、全覆盖 Full Stack



Distributed Tracing, Service Map, Request Log



CNCF CLOUD NATIVE Landscape
eBPF Project Landscape



全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

DeepFlow[®]

DeepFlow 的技术制高点：全关联 Universal Tagging



标准 Tag: 开销 10x 降低

资源池	网络资源	容器资源	Application
区域	VPC	容器集群	ServiceName
可用区	子网	容器节点	FunctionName
云平台	CIDR	命名空间	Endpoint
租户	IP地址	容器服务	TraceId
云资源	NATGW	Ingress	SpanId
宿主机	ALB	Workload	RequestId
云服务器	...	POD	...

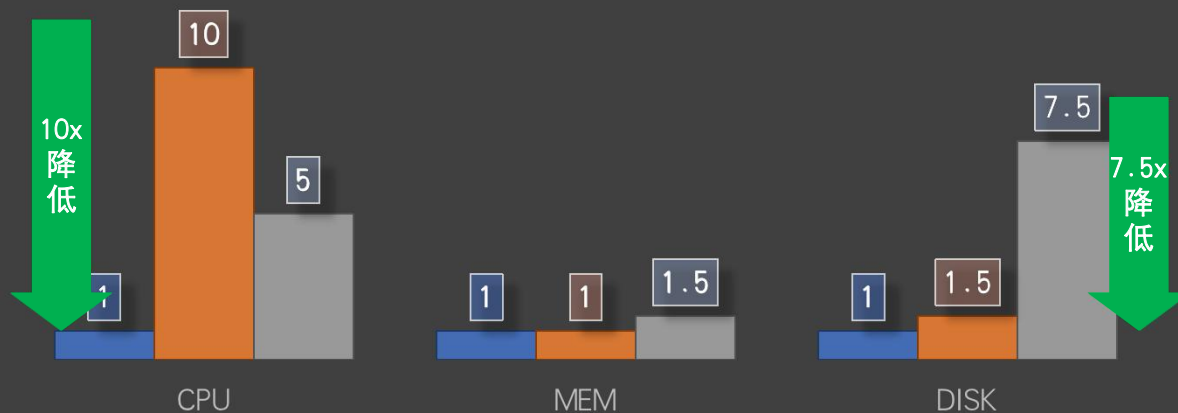
自定义 Tag: 零开销

K8s labels	Annotations *
app	biz/terminalType
version	cidcd/deploymentId
env	...
owner	...
stage	OS ENV *
commitId	MODULE_NAME
...	...



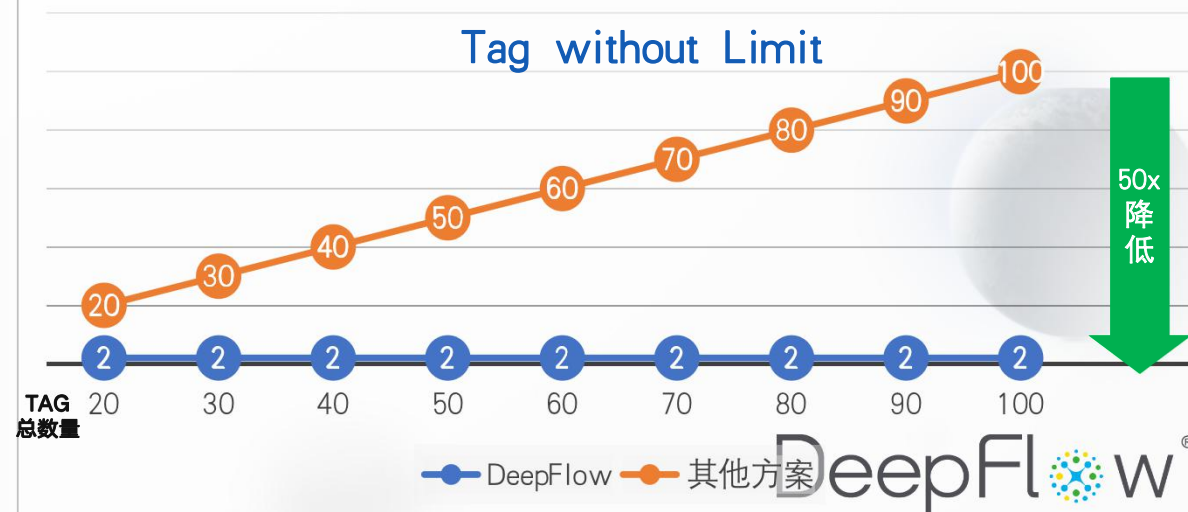
资源消耗对比（标准 TAG）

■ DeepFlow+ClickHouse ■ ClickHouse (LowCard) ■ ClickHouse (String)



随机生成一组长度为 16 字符串标签，Cardinality 为 5000，持续极限速率写入。

资源消耗 vs. # (标准 + 自定义 Tag)



假设标准 TAG 固定 20 个，自定义 TAG 从 0 个增长到 80 个。

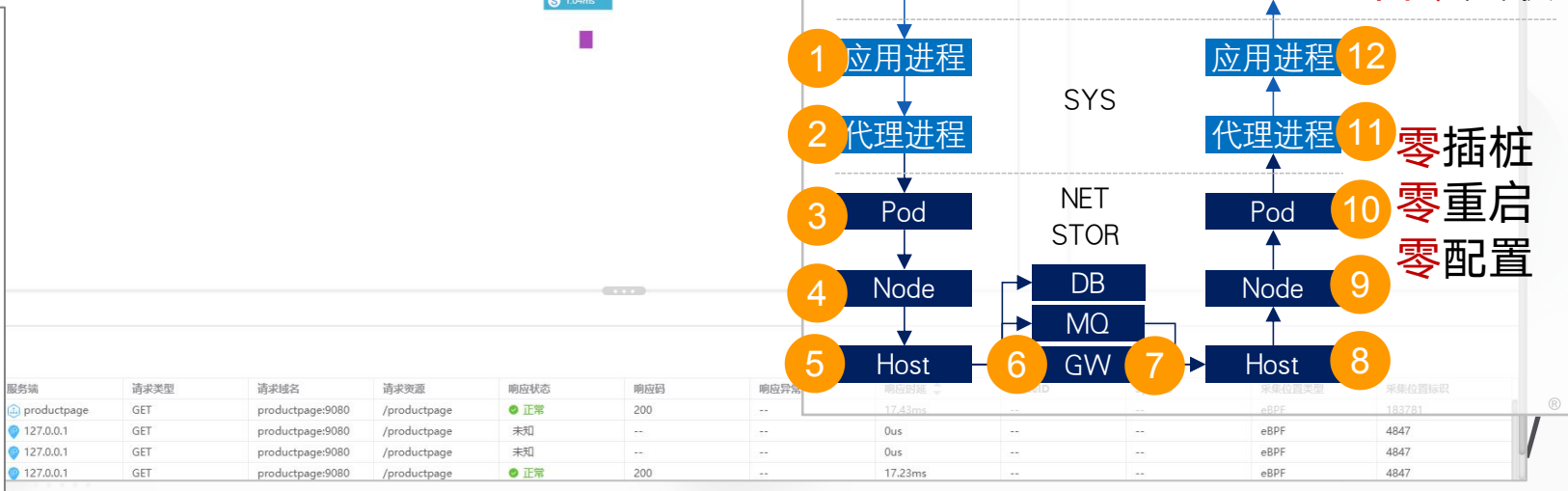
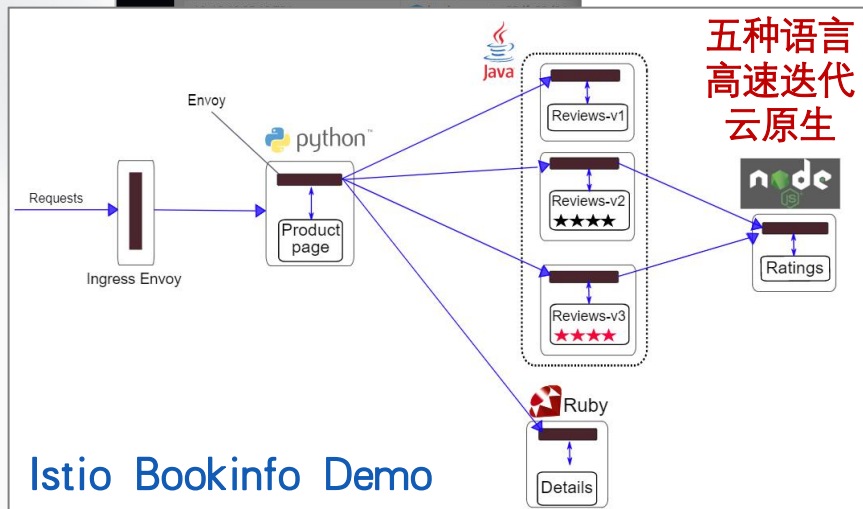
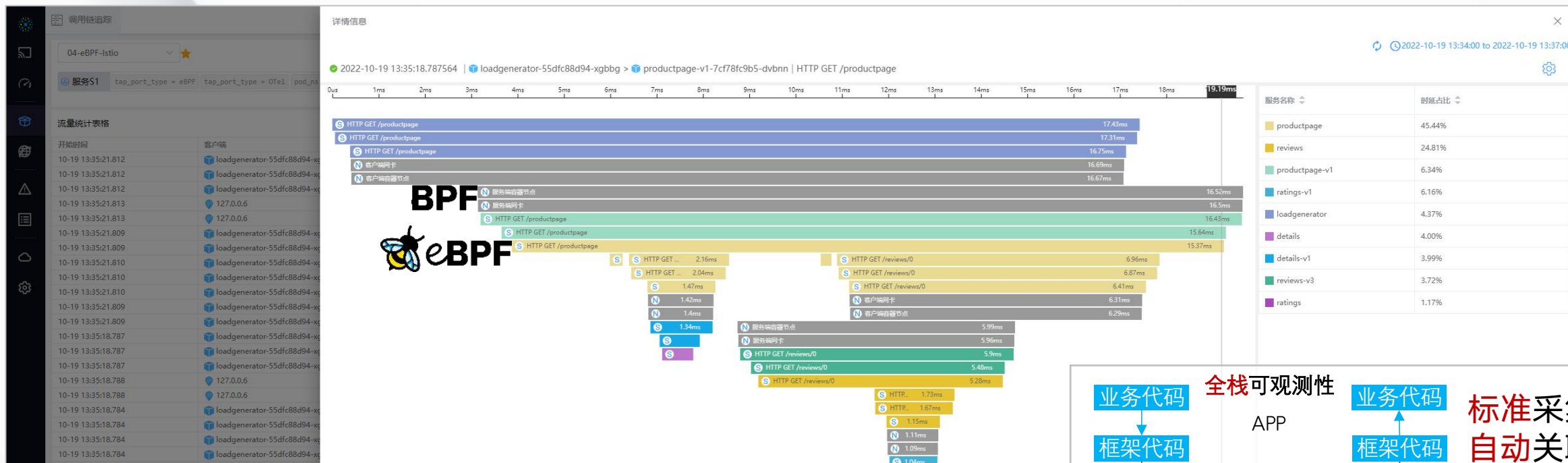


可观测性
挑战

DeepFlow
方案

DeepFlow
用例

DeepFlow: AutoTracing via eBPF

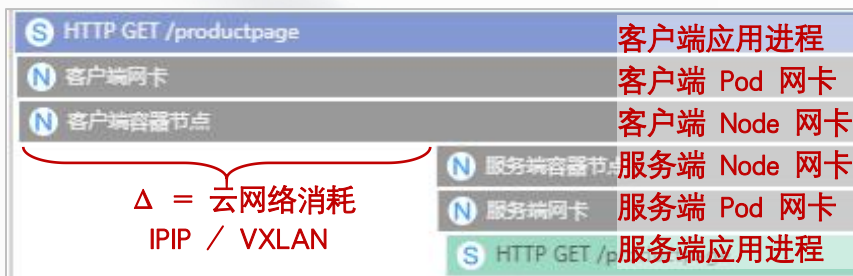


Distributed Tracing

Service Map

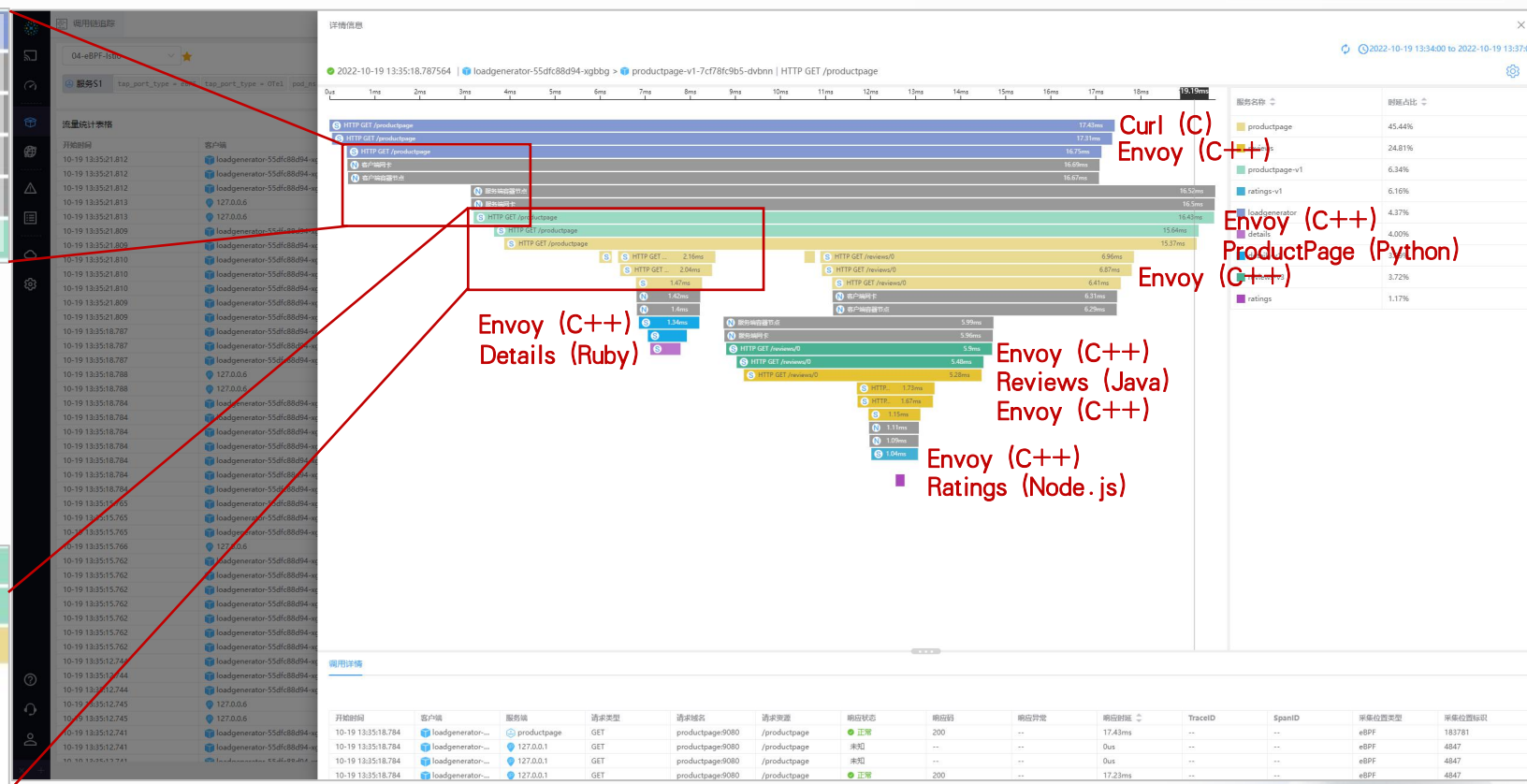
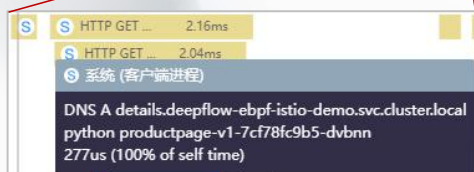
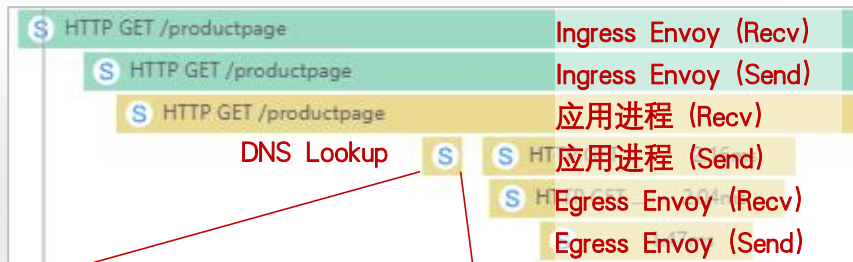
Request Log

DeepFlow: AutoTracing 解决的问题



业务开发团队
服务网格团队
容器运维团队
云运维团队

无需插桩
快速定界



1. 零插桩: 且无需向 HTTP 头注入 TraceID 或 SpanID
 2. 全链路: 4 个调用、38 个 Span, 分为 24 eBPF Span + 14 BPF Span
 3. 多语言: Java、Python、Ruby、Node.js 及 C/C++ (curl/envoy)
 4. 全栈: 追踪两个微服务之间的网络路径, 从 Pod 到 Node 到 KVM, IPIP、VXLAN、...
 5. 全栈: 追踪微服务内从 Envoy Ingress → 服务 → DNS → Envoy Egress 全过程
- 案例: 某互联网客户, 使用 DeepFlow 5 分钟内定位客户端慢服务端不慢的经典扯皮问题。

全球开源技术

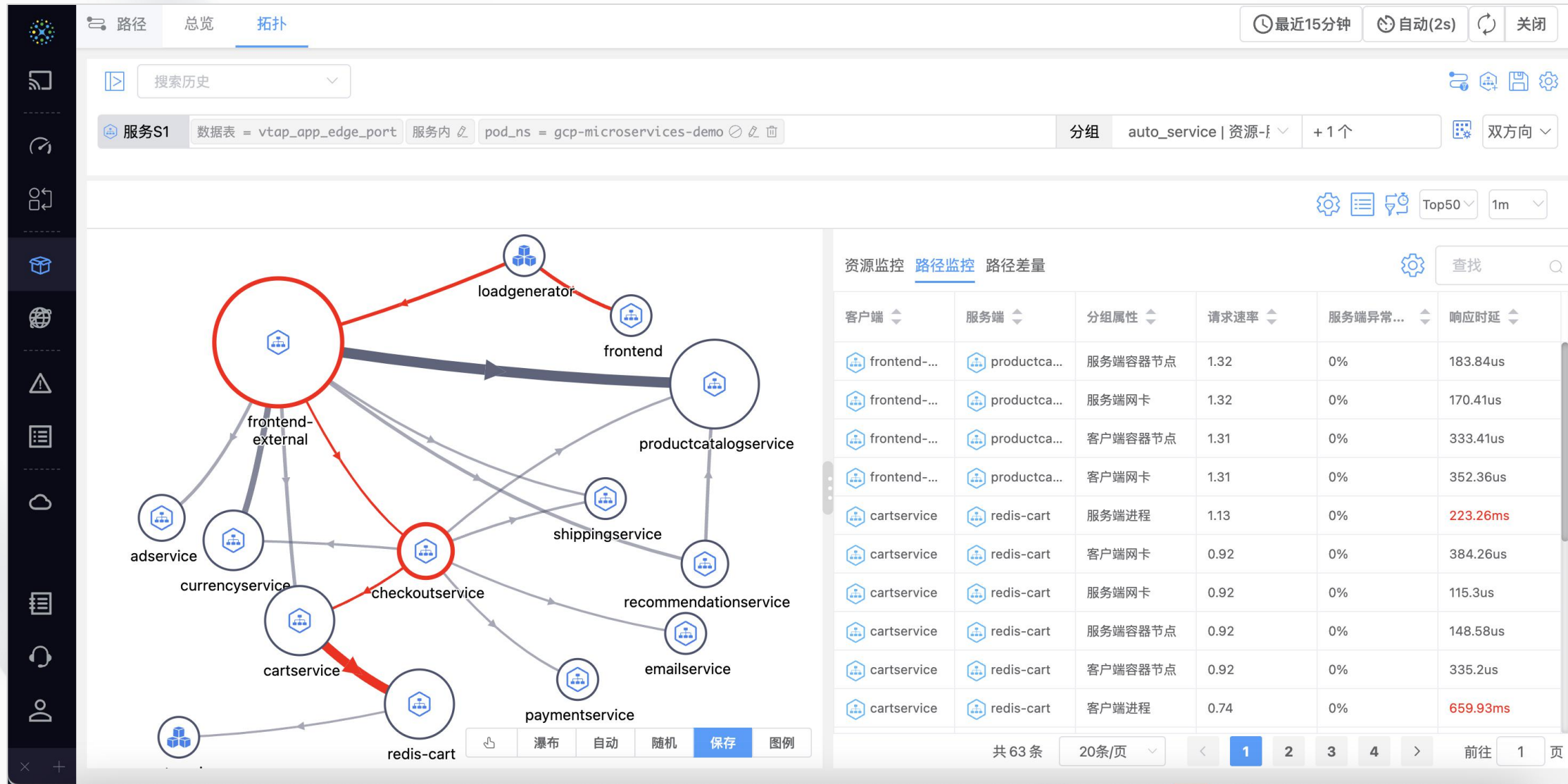
THE GLOBAL OPENSOURCE TECHNOLOGY COMMUNITY

Distributed Tracing

Service Map

Request Log

DeepFlow: Service Map via eBPF



全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

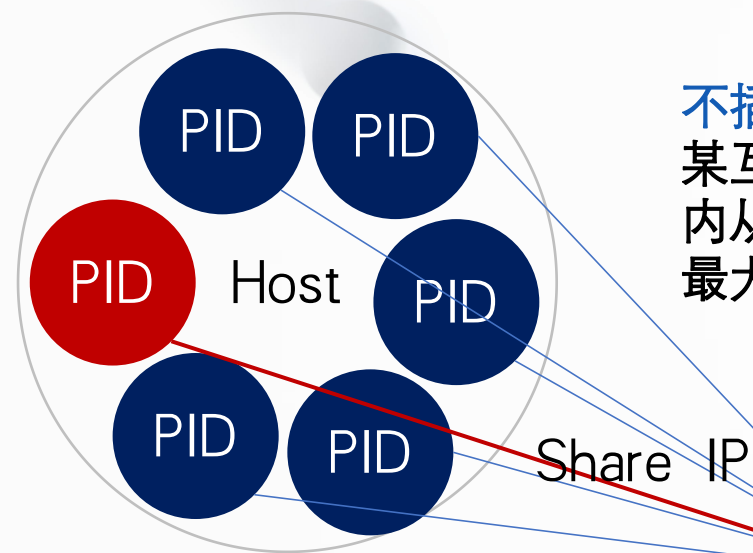
DeepFlow[®]

Distributed Tracing

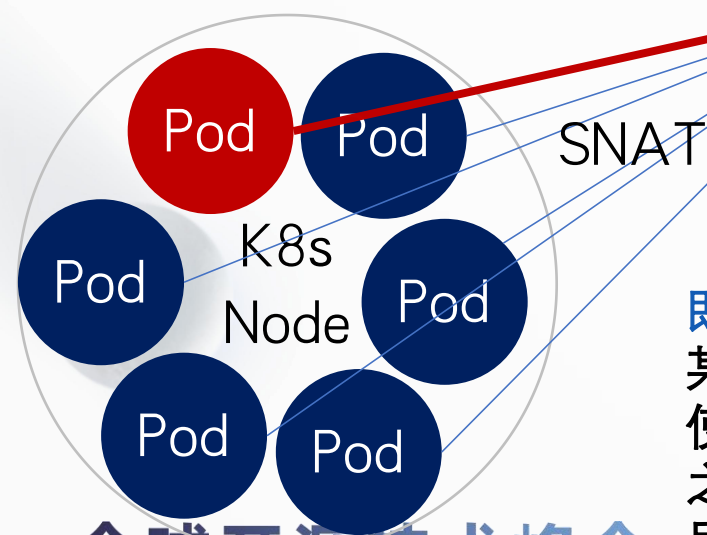
Service Map

Request Log

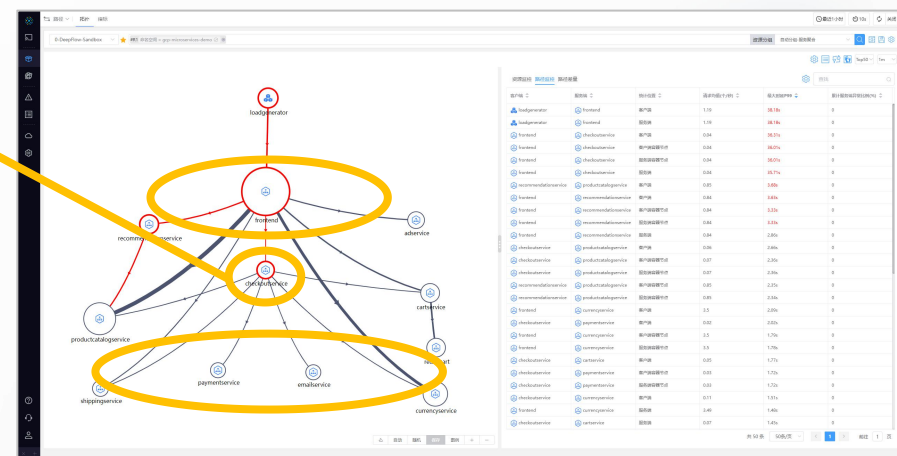
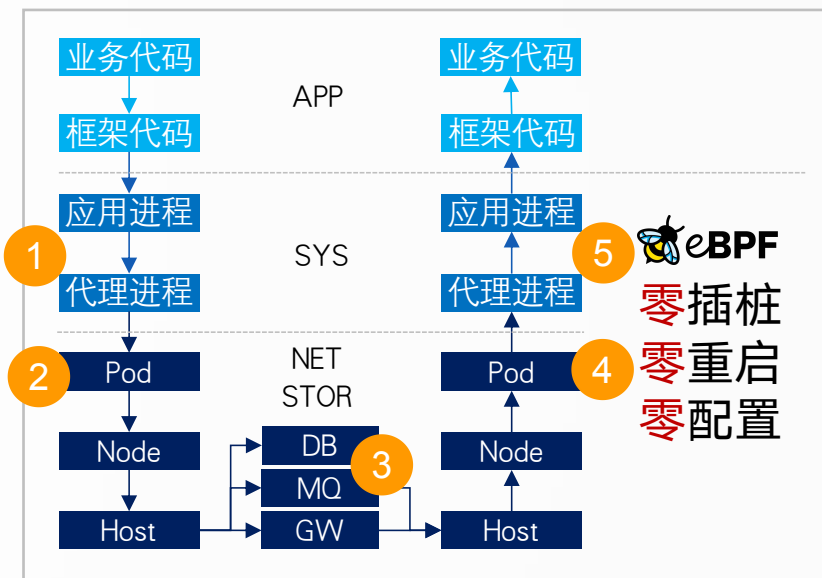
DeepFlow: Universal Service Map 解决的问题 / 1



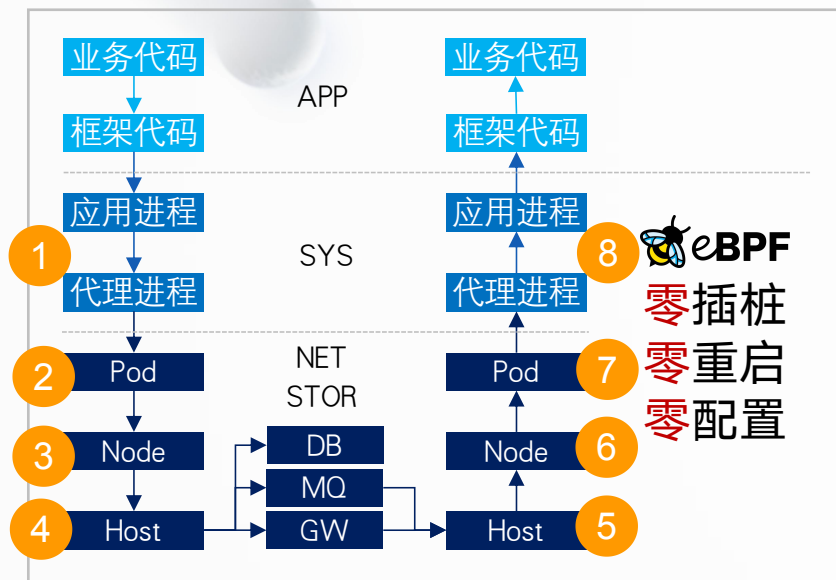
不插桩，你真的知道谁在访问你吗？
某互联网客户，使用 DeepFlow 5 分钟内从**数万个** Pod 中定位 RDS 访问量最大的 Pod、服务、团队。



即便插桩，你真的知道你在访问谁吗？
某银行客户，信用卡核心业务上线受阻，使用 DeepFlow 5 分钟内发现两个服务之间 **API 网关是性能瓶颈**，检查配置后发现缓存设置不合理。



DeepFlow: Universal Service Map 解决的问题 / 2



某互联网客户，使用 DeepFlow 5 分钟定位服务间 K8s 网络瓶颈。

某银行客户，使用 DeepFlow 5 分钟定位 ARP 导致的 POD 无法上线。

某系统软件客户，使用 DeepFlow 5 分钟定位客户端未及时收包导致 gRPC 超时。

从云基础设施到云原生应用的全栈性能指标：

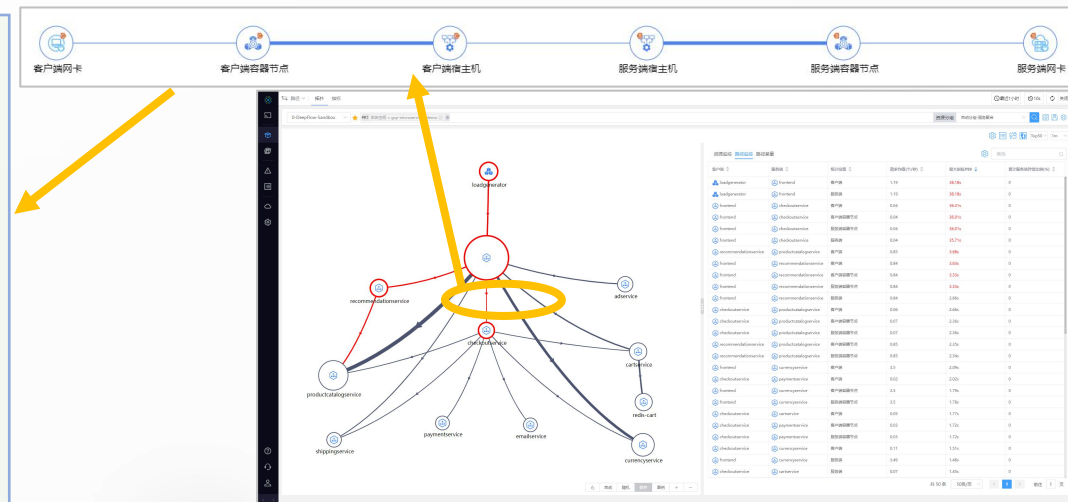
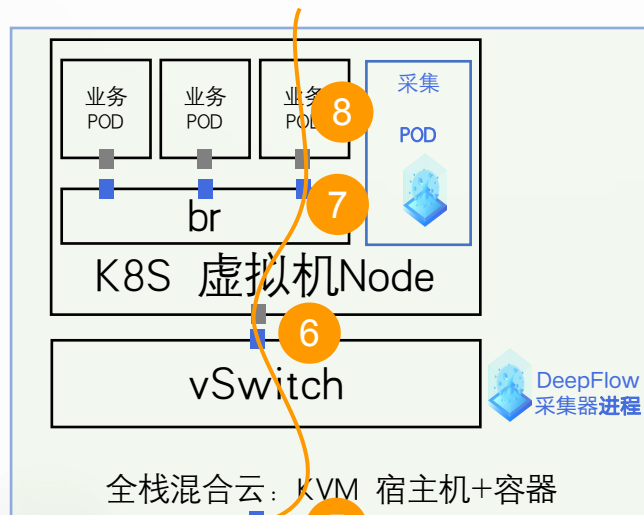
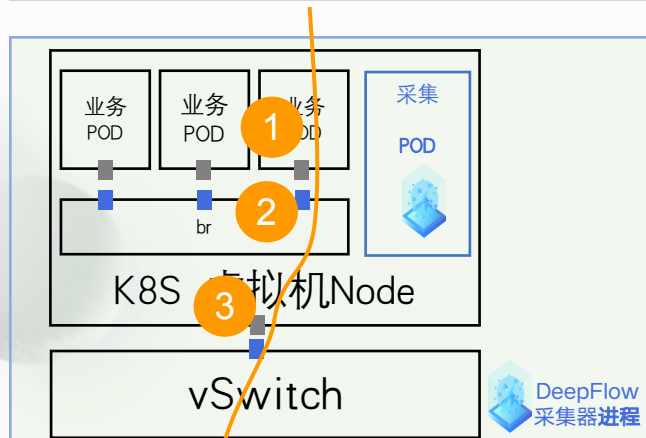
应用性能：吞吐、时延、异常（RED）

应用协议：HTTP(S)，RPC，SQL、MQ、DNS、...

系统性能：新建连接、活跃连接、建连异常、...

网络时延：建连时延、系统时延、数据时延、...

网络性能：吞吐、重传、零窗、传输层载荷、...



全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

Distributed Tracing

Service Map

Request Log

从应用到基础设施全栈

响应状态、响应码、响应异常、响应结果

TraceId、SpanId、XRequestId

… // Wasm、Lua* 插件式扩展能

// WASM 2D 插值 测试 性能为

HTTP: HTTP1、HTTP2、HTTPS

RPC: Dubbo、gRPC、SOFARPC、

DB: MySQL、PostgreSQL、Redis

MO: Kafka, MOTT

Infra: DNS, TCP, UDP, ...

... // 持续迭代中

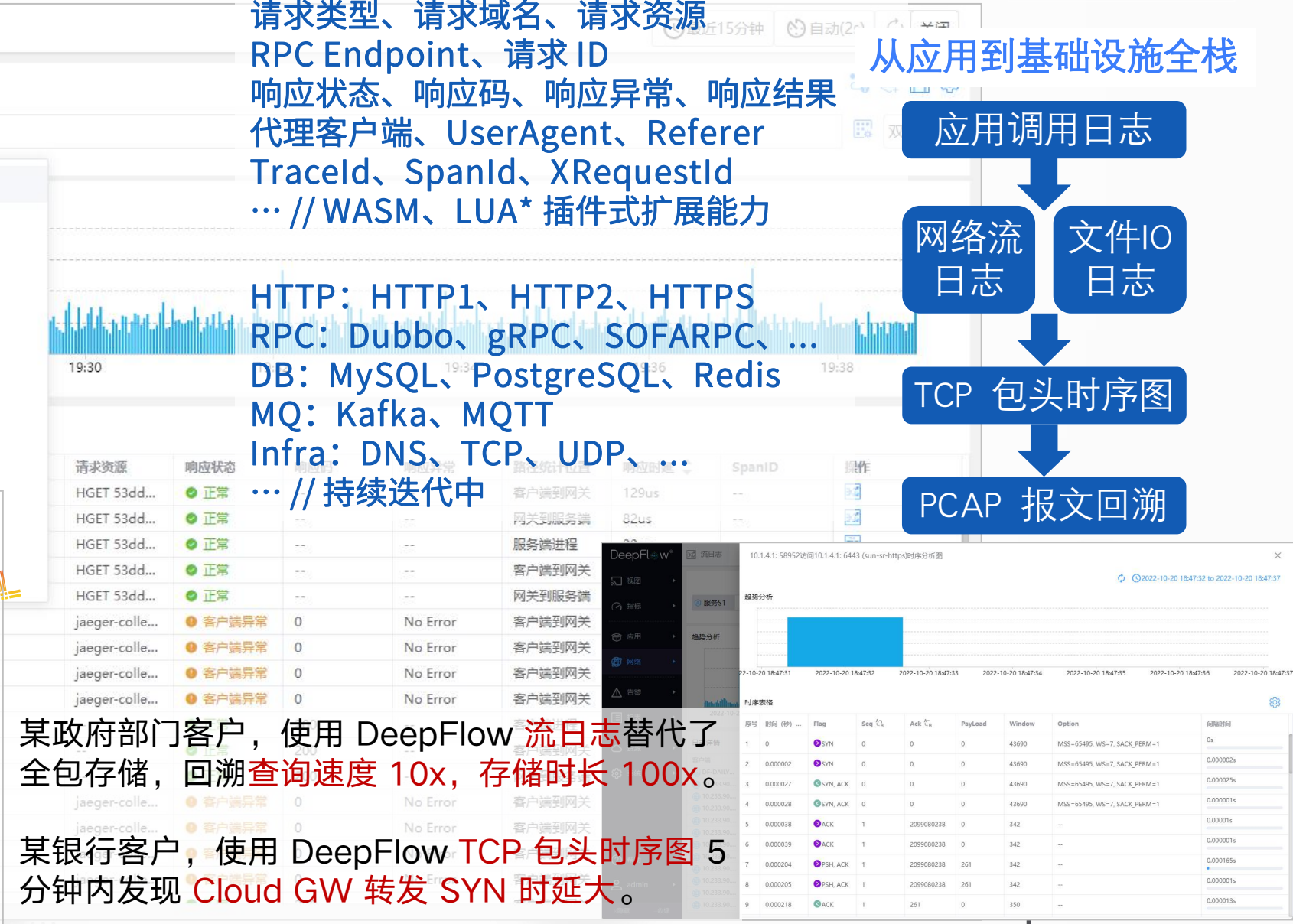
应用调用日志

网络流 目录

文件IO 日志

TCP 包头时序图

PCAP 报文回溯



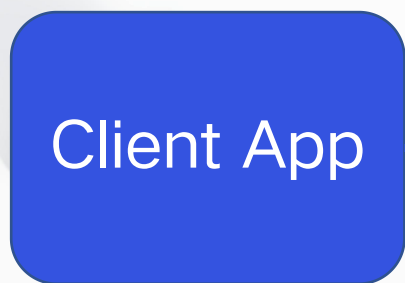
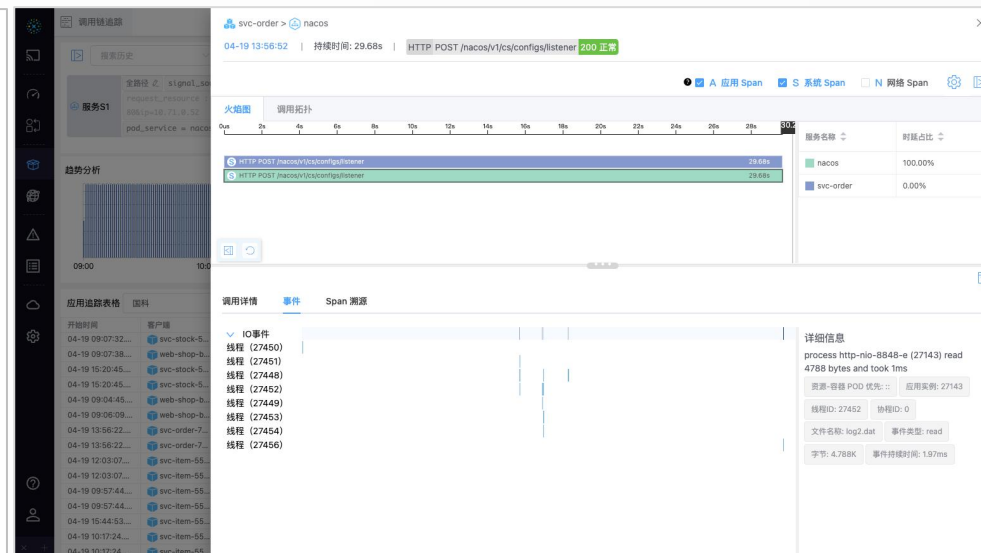
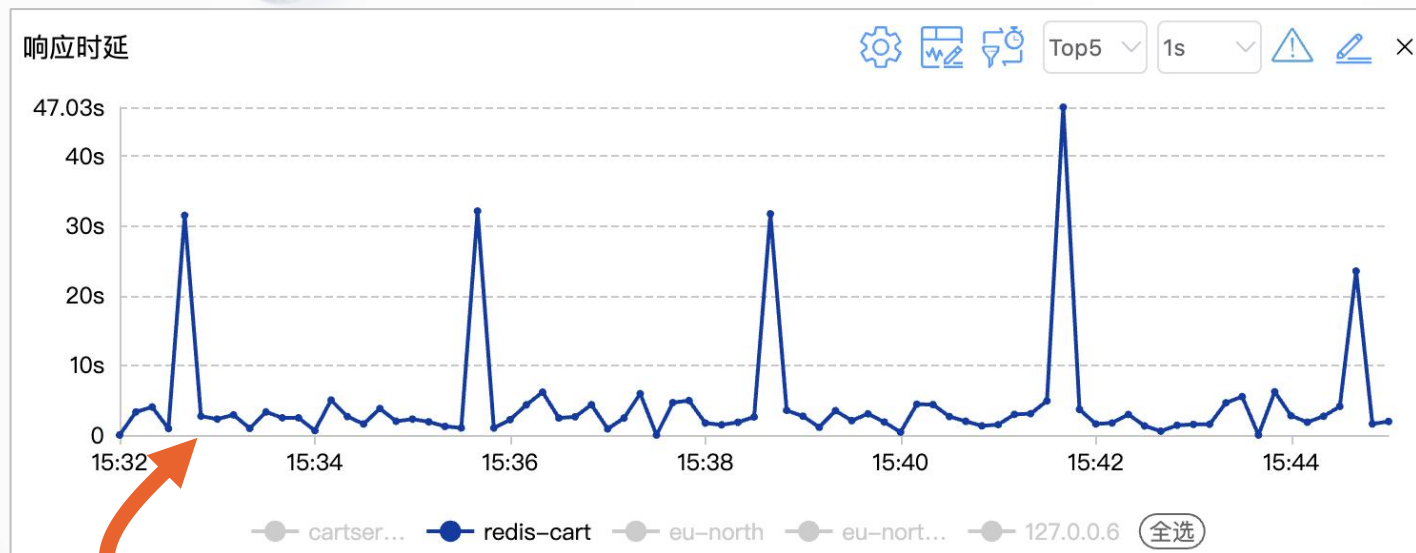
某政府部门客户，使用 DeepFlow 流日志替代了全包存储，回溯查询速度 10x，存储时长 100x。

某银行客户，使用 DeepFlow TCP 包头时序图 5 分钟内发现 Cloud GW 转发 SYN 时延大。

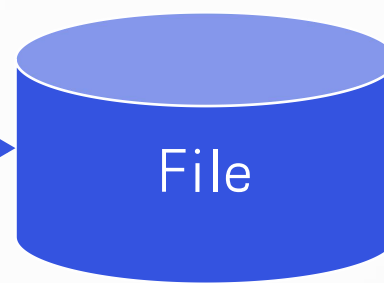
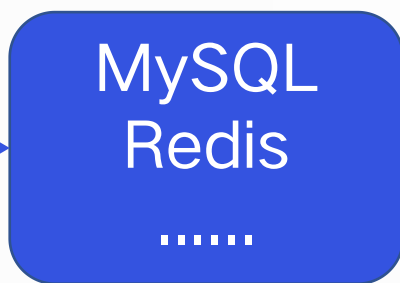
DeepFlow: Request Log 与 IO/MEM Event 关联



Request Log 生命周期内的 File IO 事件



Request
SQL / ...



File IO
#bytes, delay

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

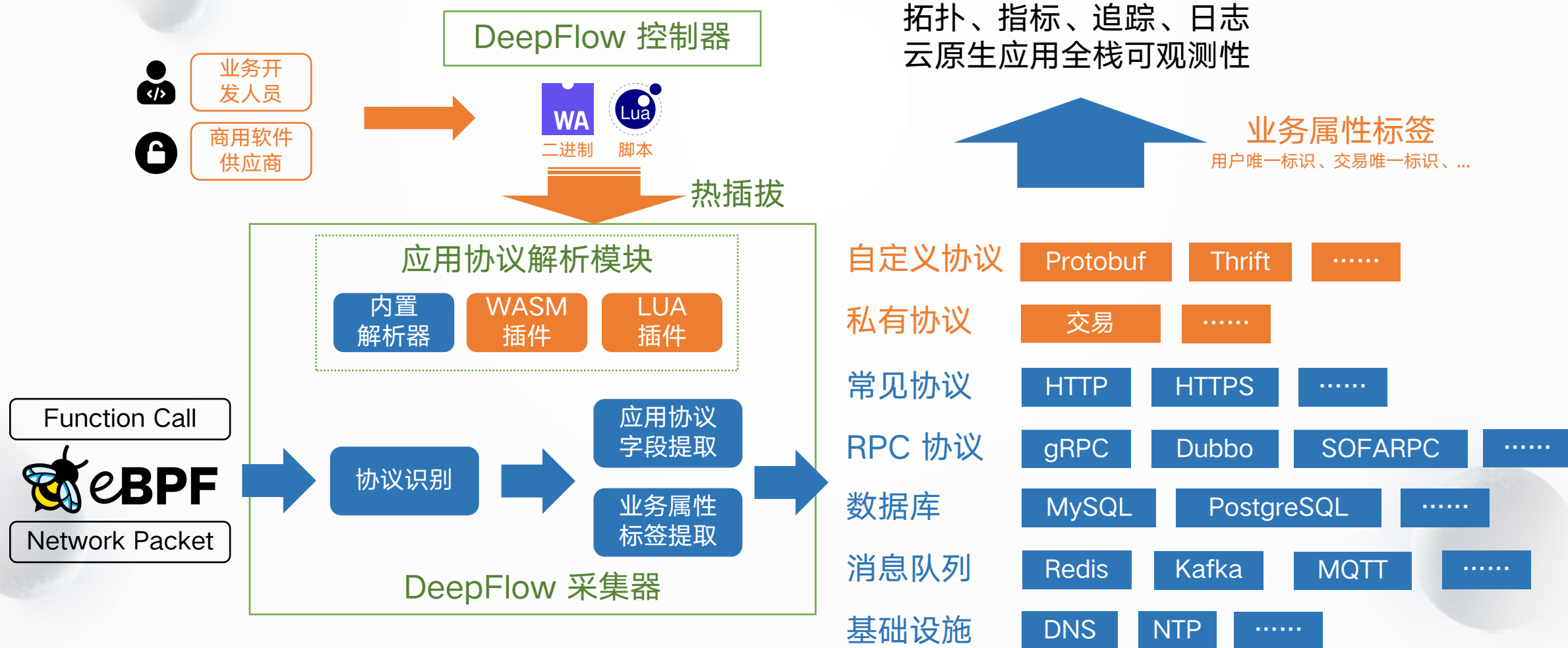
DeepFlow[®]

Distributed Tracing

Service Map

Request Log

DeepFlow: eBPF + WASM, 私有协议解析、业务字段提取



全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

DeepFlow[®]

Distributed Tracing

Service Map

Request Log

THANKS! 欢迎关注 DeepFlow

GOTC

2023.06.10 @ 深圳

蓝鲸 × DeepFlow

可观测性 Meetup

基于 eBPF 的可观测性实践

🕒 2023/6/10 下午13:00 ~ 18:00
📍 深圳·深南大道10000号
腾讯大厦二楼报告厅

分享嘉宾



刘文平 高级研发工程师
腾讯



向阳 DeepFlow 研发VP
云杉网络



谭剑 监控系统高级工程师
小米



刘梦馨 专家工程师
灵雀云



DeepFlow
社区微信群

2023.09.10 @ New York

清华大学 × YunShan



SIGCOMM 2023 New York City

中国原创

可观测性平台 DeepFlow

<< 入选 SIGCOMM 2023 >>

《Network-Centric Distributed Tracing with DeepFlow:
Troubleshooting Your Microservices in Zero Code》

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

DeepFlow[®]