DeepFlow®

容器化微服务可观测性方案

云杉网络

目录

1. 前言 2

2. 容器化微服务面临的监控运维挑战 3

监控对象数量增大 3

监控路径动态变化 3

信息维度多面 4

传统监控难以触达 4

3. 监控与可观测性 5

零侵扰 5

多维度 6

实时性 6

4. DeepFlow®容器化微服务可观测性方案 7

运行环境 8

兼容现有监控工具 9

5. 方案优势 11

分钟级就绪 11

零侵扰采集 11

海量数据处理 13

多维度信息关联 13

6. 方案价值 14

分钟级定位故障边界 14

大幅提升排障效率 17

微服务可用性指标 18

7. 总结 19

1. 前言

应用软件架构经过单体架构、分布式架构，面向服务架构（SOA）、微服务架构发展，在云计算、容器、网络等技术的支撑下，更多企业的新一代应用都开始选择使用微服务架构。

微服务是以业务功能为主的软件设计架构，每一个服务都具有自主运行的业务功能，并使用轻量级机制进行通信，通过一个或多个微服务组成应用程序。围绕业务功能创建的服务依靠全自动部署机制独立部署运行，具备避免集中式管理，可以选择不同的编程语言编写等优势。

容器技术在 LXC (Linux Container)、cgroups、namespaces 技术基础上发展成型，是更轻量化的虚拟技术。随着 Kubernets 的发展解决了大规模容器集群的编排和管理问题，将被拆分解耦的业务系统再统一管理起来让服务可以自由融合、彼此协作为业务的发展提供高效的基础能力。

2. 容器化微服务面临的监控运维挑战

容器化微服务在具备面向业务功能、降低开发难度、增加容错度、服务间松耦合、弹性扩展、提高生产效率等优势同时，也面临服务发现、限流、权限、版本管理等挑战，服务监控也是其中挑战之一。

监控对象数量增大

传统监控以单体应用为粒度，结合计算、存储、网络等基础设施监控进行运维保障。但在容器化微服务架构下，监控粒度细致到容器 POD 或微服务 API 级别，使得监控对象的数量相比单体应用呈指数级增长。

监控路径动态变化

基于微服务架构的应用动态性是监控面临的一个重要挑战。应用系统由很多个微服务组成，同时运行于容器环境中，服务通过相互调用、自身处理产生出复杂的行为。微服务并不是静态的，而是随着业务需求不断变化，改变调用关系，处理过程以及优化路径。

信息维度多面

保障容器化微服务，首要基础就要有能力全面“度量刻画”微服务，这其中涉及到应用、系统、基础设施等相关资源属性和性能指标等，这是一张多维关联的知识图谱。以一个异常的服务访问路径为例，涉及到的指标包括错误率，调用延时等。仅仅发现异常还不足以解决问题，因此会继续产生一系列的疑问：

- 路径上的异常服务有哪几个？
- 服务运行在哪个节点上，哪个区域中？
- 访问路径在异常前有无变化？
- 服务历史中的调用关系有哪些？
- 异常可能影响的范围？

回答这些疑问，任何单一维度的排查都是管中窥豹。必须通过关联服务、容器平台、网络流量、平台事件等数据信息，构建多维度的知识图谱，在解决问题的过程中方可成竹在胸。

传统监控难以触达

在云环境中，由于虚拟化技术应用，传统监控工具难以触达池化后的各类资源，导致“黑盒”化严重。此外，传统监控工具从数据开放性、架

构扩展性、监控粒度上都难以满足云原生应用的需求。新兴可观测技术，旨在通过系统的外部监测数据（指标、日志、追踪等）实时分析系统的内部状态（吞吐、错误、时延等），彻底解决云内监控的“黑盒”问题。

3. 监控与可观测性

随着业务系统不断上云，容器、微服务、持续发布等云原生技术被广泛采用，从而为 IT 系统的监控带来了全新挑战。为保障云原生应用的稳定性，可观测技术被越来越多的企业所采用。针对于 IT 系统，尤其是面向云原生应用，可观测技术应包含如下需求：

零侵扰

传统 APM/NPM 等工具，要么需要应用程序中打桩插码，要么需要基础设施中分光镜像，均会对 IT 系统进行侵扰。可观测技术使用外部数据做分析，因此采用零侵扰的方式获取监控数据，无需打桩插码、分光镜像，而是通过开放系统架构直接获取监控数据。零侵扰的另一方面是要求低功耗，不能因为采集数据而影响应用或基础设施性能，通常采集点功耗不能超过业务功耗的 1%。

多维度

要保障云原生应用稳定运行，可观测技术必须包含多维度数据分析能力。具体来说，要将应用的 API、容器、主机、网络等监控数据进行全栈关联分析。传统的 APM 工具，可以定位代码层问题，却无法追踪容器或主机网络服务引起的故障。而传统的 NPM 工具，又不能关联应用的 TraceID 从而追踪穿越 NAT、LB 等网元的流量。因此，多维度的全栈数据分析，是可观测平台的第二个需求。

实时性

云原生应用的动态性要求可观测平台必须具备实时性。如果应用的升级/扩容在分钟级完成，那么监控系统就必须提供秒级的反馈能力。注意，这里的反馈需要对海量指标/追踪/日志数据进行查找分析，因此对可观测平台的海量数据实时处理能力提出了极高要求。

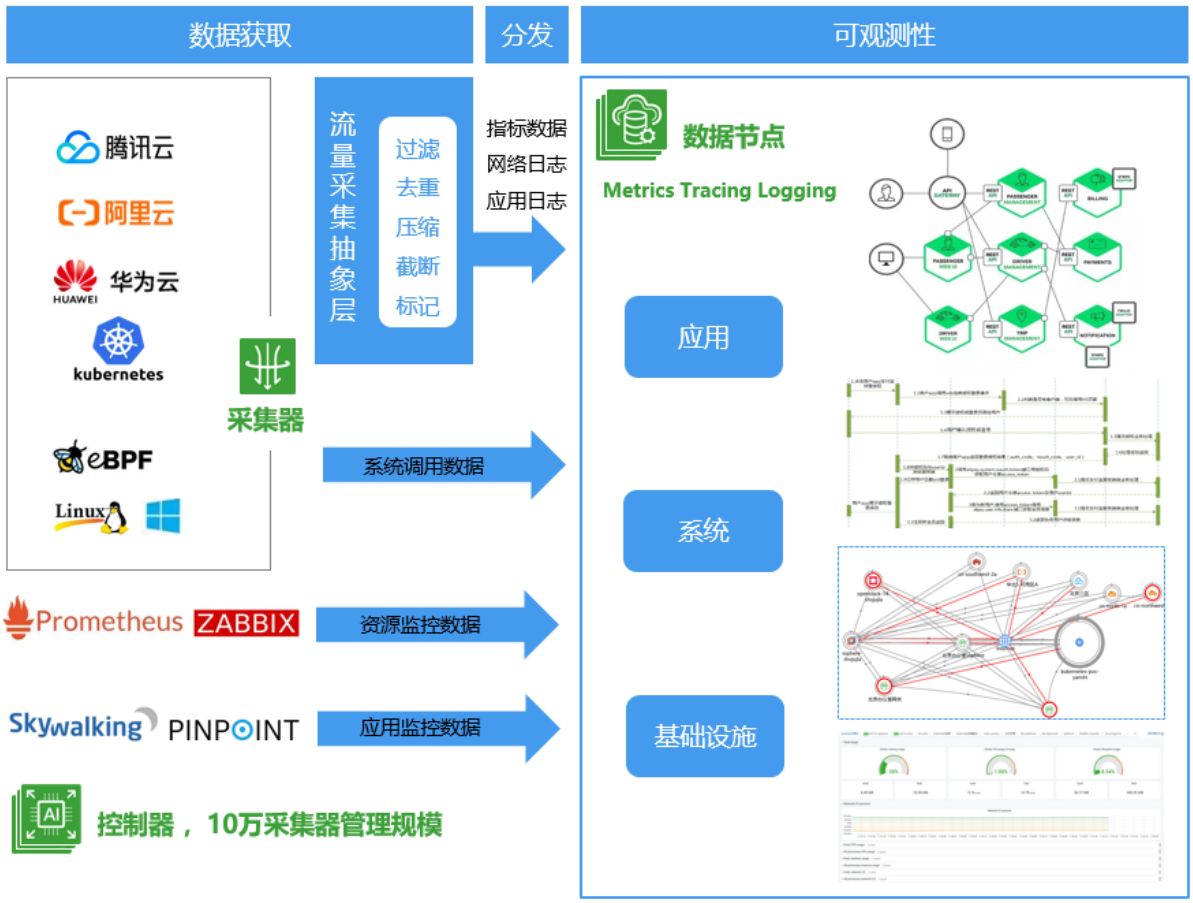
在 Google Cloud 的定义中，“可观测性是可帮助团队有效调试其系统的工具或技术解决方案，并有能力探索未知或事先未定义的属性和模式。”可观测性并不是通过简单使用一个工具所能具备，是需要根据企业

组织、业务应用、基础设施以及已有的监控体系的需求现状，明确阶段目标，伴随着业务发展逐步建立，是一个持续发展的过程。

本文通过云杉网络与客户的共同实践，基于云原生应用发展的现状，通过 DeepFlow ®与客户的应用、网络、基础设施整合，综合分析各类指标、日志以及追踪数据，形成一站式的容器化微服务可观测性方案。

4. DeepFlow ®容器化微服务可观测性方案

DeepFlow ®提供适用于容器化微服务的可观测性，解决云原生应用诊断难的核心痛点。通过对全局微服务（Universal Services）间的通信访问、系统调用、平台环境等数据进行深度分析，提供监控告警、故障定位及风险排查，保障业务在云原生环境中的稳定、高效运行。



图：方案示意图

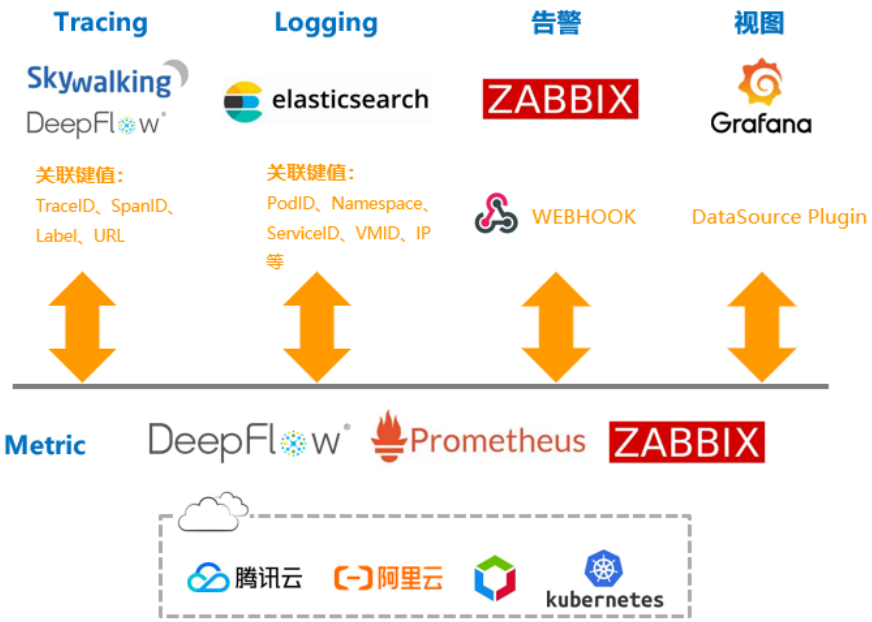
运行环境

在公有云环境中，例如由腾讯云、阿里云等的云服务商提供的 IaaS 服务基础上，可以直接选择容器服务或者通过开源 Kubernetes 软件搭建容器平台，部署运行各类容器化微服务。DeepFlow ®通过零侵扰采集技术，获取各类运维监控数据，如基础设施资源配置、网络流量、API 调用

等，并通过后台实时数仓，对多维度数据进行实时分析处理，满足不同场景下的微服务可观测能力。

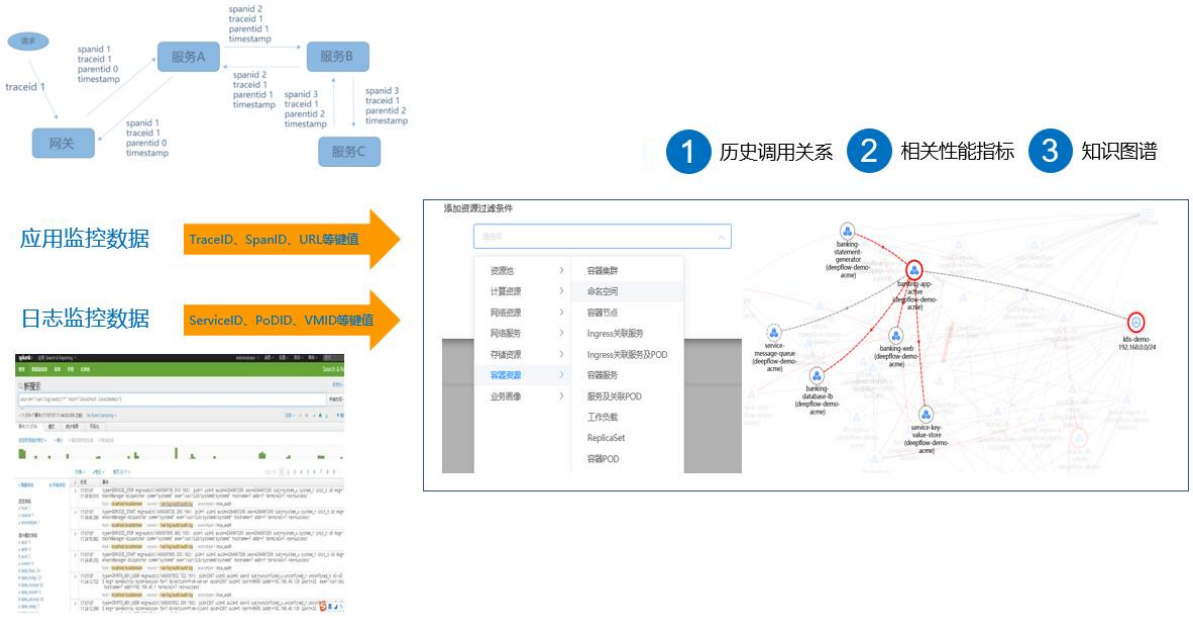
兼容现有监控工具

在监控体系发展以及客户的实际使用中，存在各类监控工具，如资源监控 ZABBIX、容器监控 Prometheus、日志监控 Elastic Search、应用追踪 Skywalking 等等。DeepFlow ®凭借零侵扰、低消耗、高性能的流量采集技术，可以为现有监控工具补充难以获取的丰富指标数据，结合客户已有的环境及工具平台，达到监控数据关联，运维流程联动，满足微服务保障的告警、定位、排障以及预测量化服务质量的要求。



图：兼容现有监控工具

以 DeepFlow ®做为基础数据平台，结合 Prometheus、ZABBIX、云监控数据构成对应用、系统、云基础设施的指标数据基础，通过开放数据对接告警平台、认证平台以及 Grafana 等数据可视化平台，向微服务开发部门、业务部门展现关键指标数据。通过 TraceID、SpanID、URL、资源 ID 等关键值关联追踪数据，在应用出现异常或者调优过程中，以高效对接历史指标趋势划定资源范围，使特定应用请求过程与指标钻取统计不再割裂。实现对云上微服务的全方位可观测能力。



图：指标数据关联分析

5. 方案优势

分钟级就绪

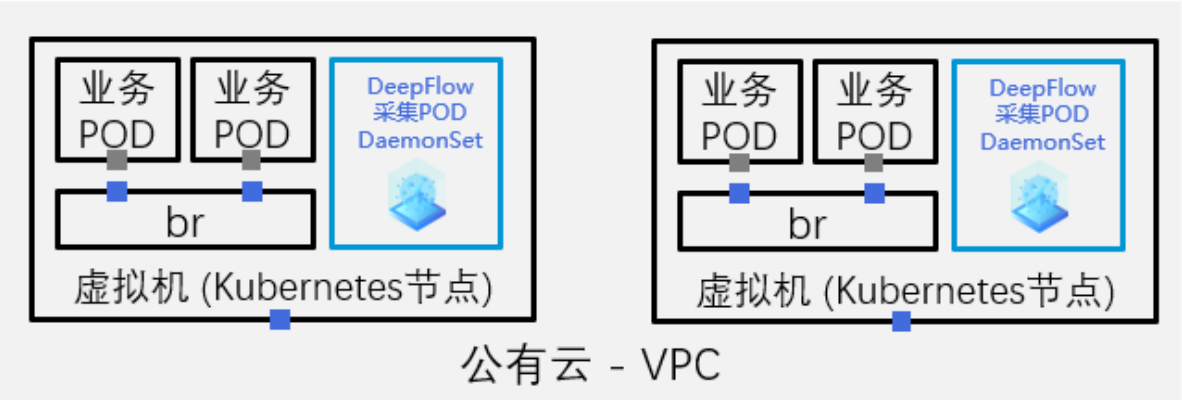
即使是万级容器节点规模，通过 yaml 配置文件，分钟级实现采集器部署应用，就绪对 POD、服务以及系统数据获取能力。避免侵入、配置业务 POD，完全实现云原生规模化部署安装，具备对每一采集器粒度工作状态的监控。

零侵扰采集

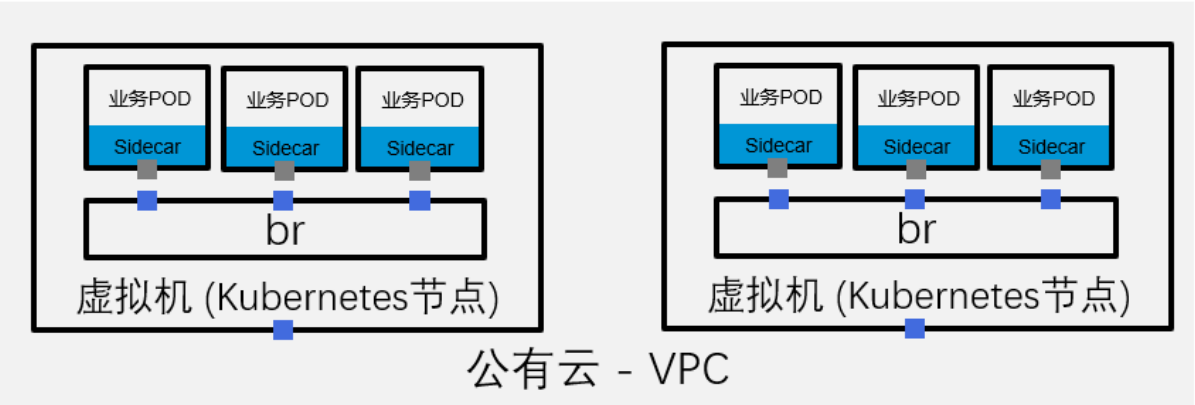
微服务间性能指标数据采集包含网络流量、函数调用、平台资源等监控数据的收集及处理。在 Kubernetes 中，Pod 是能够创建、调度、和管理的最小部署单元，是一组容器的集合，同一个 Pod 里的容器共享同一个网络命名空间， IP 地址及端口。获取 Pod 间的网络流量信息，通过命名空间、工作负载、服务等维度进行聚合，可以最大程度自动化地绘制服务间访问性能及访问依赖。但规模化获取 Pod 间网络流量并非易事，需要考虑对生产面的侵入、依赖程度、管理复杂度以及可扩展性。

微服务间性能数据获取主要有 DaemonSet、Sidecar 两种方案，Sidecar 方案中对生产流量有侵入，并非一种完全的旁路方式，同时，每一个 Pod 中都要启动容器，管理及资源开销都比较大，在整体方案中不

做推荐。以 DeamonSet 方式运行采集器，如下图所示，只在 Kubernetes Node 粒度就可采集所有 Pod 流量并做预处理获取各类指标数据。对于容器节点 Linux 内核版本大于 4.14 的环境增加 eBPF 数据处理能力，将大幅增加数据维度及提升数据收集能力，包括网络流量、应用性能、内核追踪，流量控制等维度，对实现可观测性提供了强有力的支持。



图：DaemonSet 方式获取监控数据



图：Sidecar 方式获取监控数据

海量数据处理

实现可观测性涉及各类型海量监控数据，在 DeepFlow ®平台中核心收集网络流量、应用性能、内核追踪、调用关系等类型数据，分布式处理存储数据后，支持以多种维度进行检索查询。通过运行在容器 Node 上的采集器，实现分布式系统对海量运维数据的处理，并严格约束每个采集器的资源使用上限。

DeepFlow ®提供强大的指标数据能力，针对不同维度的访问关系对端到端性能等提供 12 类 110 种指标量，与算子结合可产生 2000 多个指标量统计值。涵盖网络层、传输层、应用层、系统层等各类型数据，尤其在微服务中使用的 HTTP、DNS 等协议的异常、时延和超时。

多维度信息关联

与云紧密结合并且具备云原生应用的特性，是实现可观测性的重要条件。做为被监控观测的对象，云、容器以及云原生应用不应该被孤立看待，需要各类型层面的监控数据有效整体联动。比如网络流量不再仅仅由 IP、端口等来标识，能对应到虚拟机、容器 POD、Service 等维度；应用中的异常，通过 TraceID、资源 ID 等关联至网络性能指标；微服务的调用关系及性能定义 SLI。可观测性方案支持多云环境，自动获取并绘制知

识图谱，关联应用性能数据、日志数据等各类数据源，快速具备在云端的云原生应用的可观测性能力。

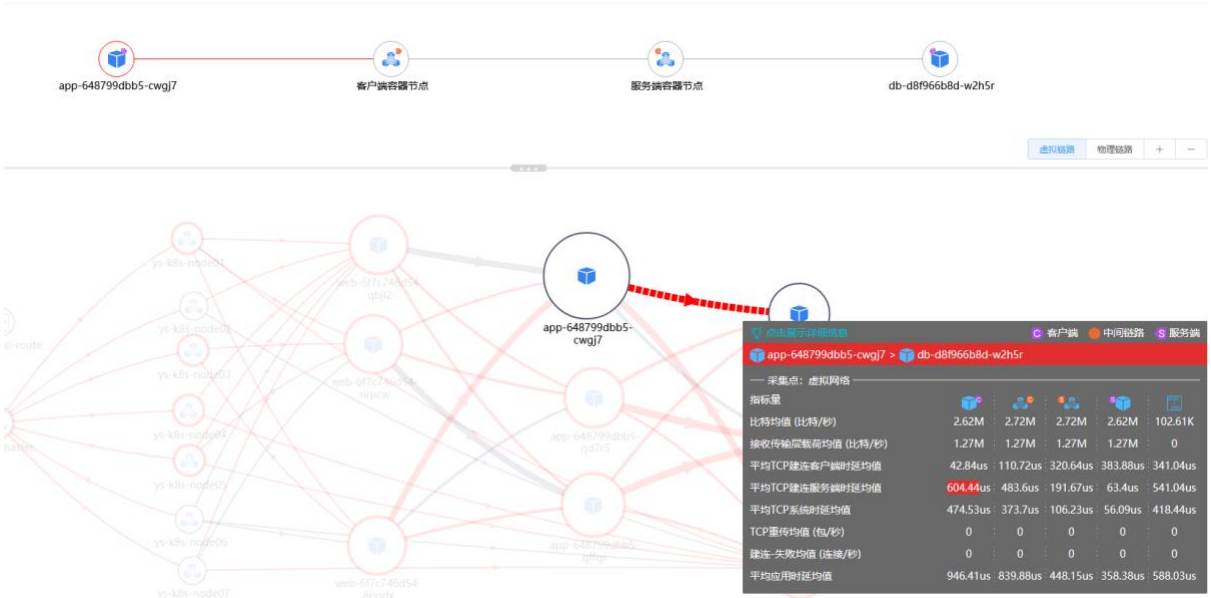
监控平台本身也要符合云原生架构特性，应对监控对象弹性伸缩以及海量数据存储及成本。

6. 方案价值

分钟级定位故障边界

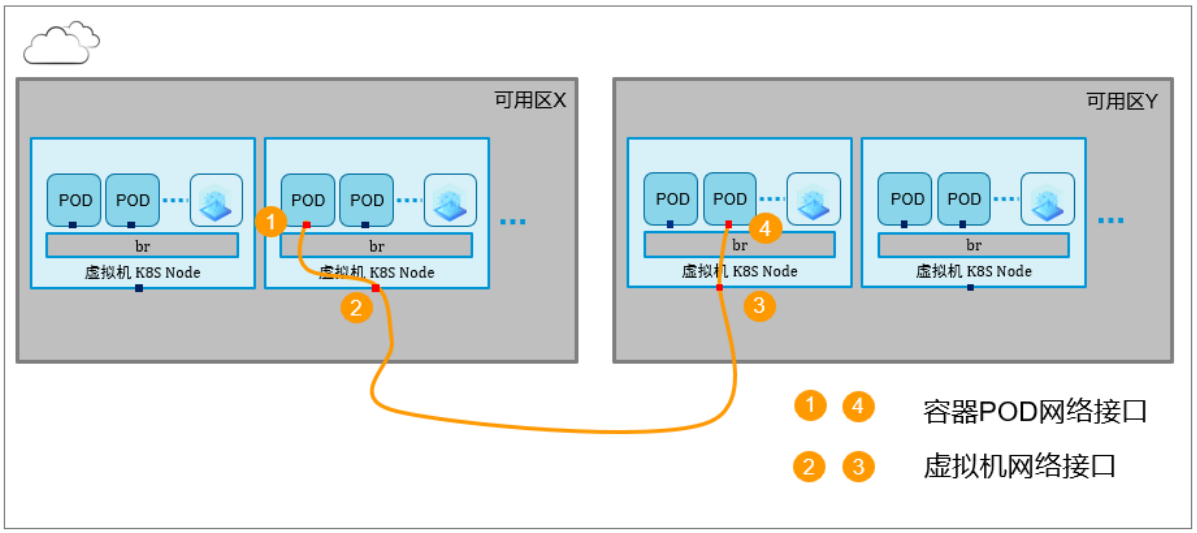
基于容器化微服务的云原生应用出现故障时，快速明确问题边界是解决问题的第一步。基于知识图谱、微服务调用链、全栈追踪等功能组合，快速检索到异常单元所关联涉及到的其他维度信息和影响范围；直观展示系统、容器、虚拟主机全栈性能指标锁定性能洼地等。

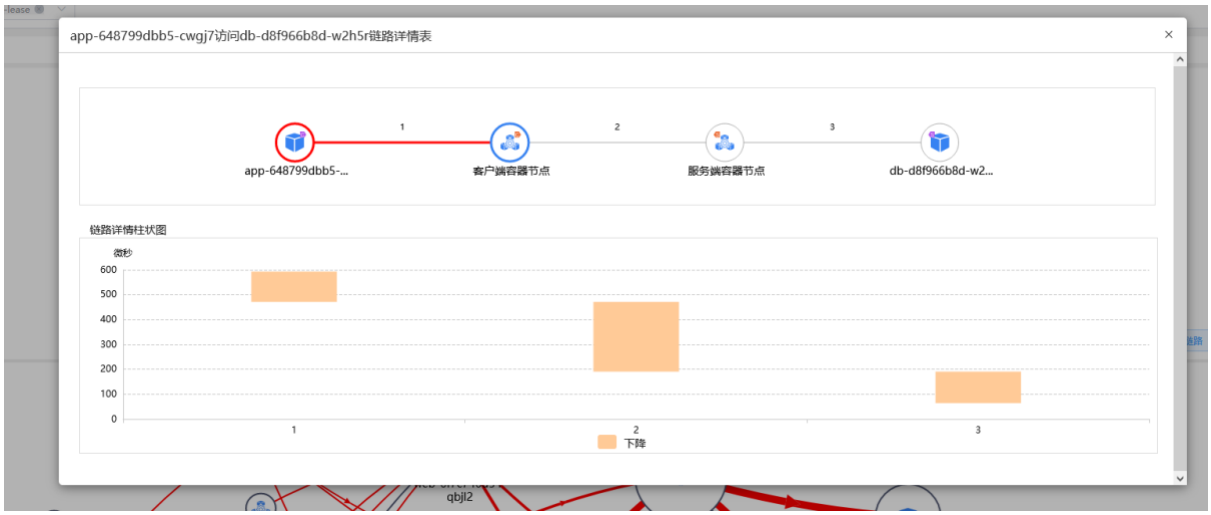
在以下服务间访问示例图中，通过告警或监控视图，快速发现两个容器服务间的 TCP 建连延时异常。



图：服务间访问示例

在容器化微服务的拓扑视图中，清楚地发现两个服务间 TCP 建连延时较大，但这个异常现象，并不像传统物理环境中直观简单地可以判断问题边界所在。在云环境中，故障边界涉及应用代码、容器节点、云主机、网络服务等一系列技术栈，因此全栈端到端性能监控是应对新环境运维排障挑战的必备能力。以全栈视角，分解展开以上 TCP 建连过程，可以剖析访问端发起 TCP 建立连接所经过的 POD 接口、虚拟机接口、乃至网关等链路。





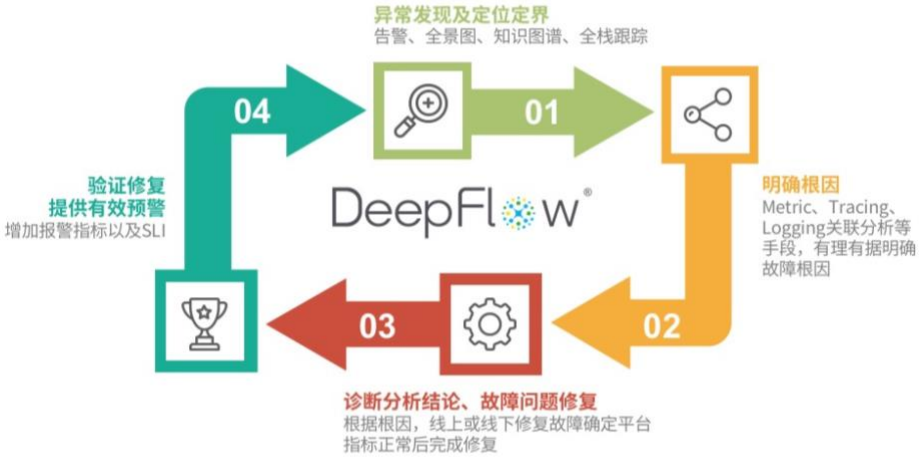
图：服务调用全栈跟踪示意图

全栈跟踪针对云中服务间的调用访问，将虚拟化所实现的逻辑通信进行逐步展开，清晰便捷展示每段的网络状态、性能，结合知识图谱及丰富的指标数据，快速定位性能异常的问题范围边界。以上所述 TCP 建连延时为例，可以快速确定异常边界在客户端容器节点之内。

在没有 DeepFlow ®全栈端到端性能监控的情况下，排查服务访问调用的性能将是一个复杂繁琐并且冗长无头绪的过程，同时对于一线运维人员要求所掌握的技术栈较多，综合能力强，很可能耽误宝贵的运维窗口时间。

大幅提升排障效率

排障过程并不仅是找到故障根因并修复，而是一整套运维保障操作闭环。容器环境叠加微服务架构使得排障更加复杂，需要避免以单次异常故障为视角，避免以单一数据为依据判断根因进行排障工作。



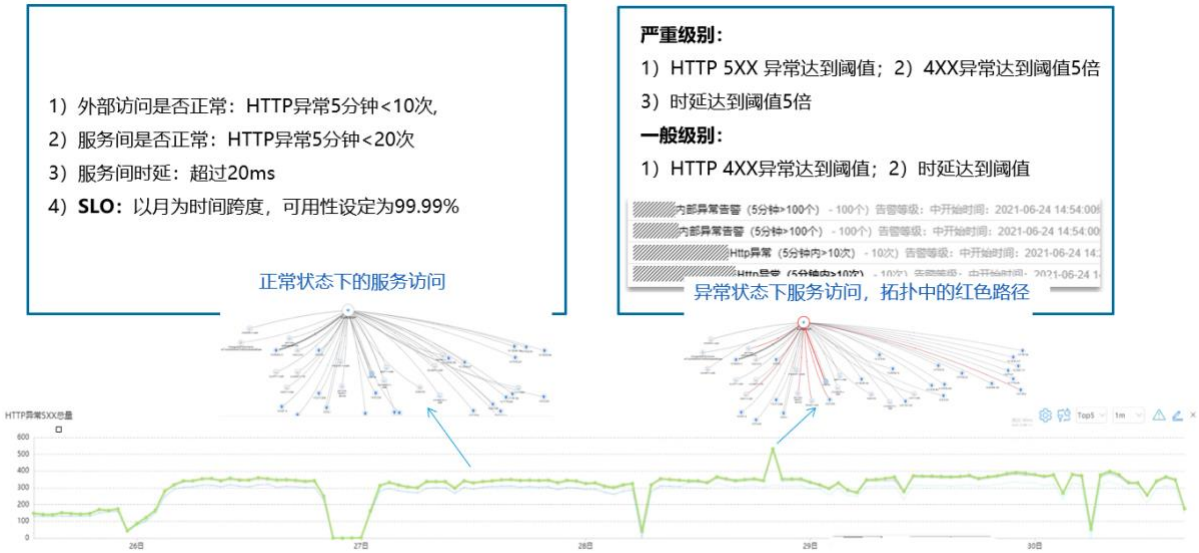
图：运维保障操作闭环

DeepFlow ®平台的排障过程可分为四个阶段，定界、根因、修复以及验证预防，从每一个阶段着手，有效提升排障效率。在第一阶段，平

台能数量级地提升发现异常后定位根因边界范围的效率，尤其广泛面临的性能问题，通过分段解析性能消耗，达到分钟级定位问题边界。在明确根因阶段，由于故障复杂度及场景类型多样，同时需要专家参与，很难量化根因确定的时长周期，但有效地将应用、容器平台、系统调用等运维数据进行关联，且对比指标、跟踪以及日志特征，能降低发现根因的复杂度，从而提升效率。修复阶段是基于根因的准确性进行的，一些棘手深入的问题，虽然修复不复杂，但隐蔽且不易复现，如果不能明确故障特征，用数据及指标说明根因，很可能会导致表面规避故障，不定期出现后造成更严重损失或重复排障。所以在这个阶段可通过历史视图、系统运行表现、修复验证等指标来确认从而提升效率。修复验证后并不是排障最后一步，根据故障的指标特征有效设置告警并加入到 SLI 中，达到长期修复验证效果，同时也避免其他区域服务、新上线服务可能出现相同的问题。

微服务可用性指标

应用微服务化后，衡量判断众多微服务的质量以及可用性是一个繁琐的问题。涉及到不同开发团队，设定具体指标，周期性的记录和评估，发现性能洼地及热点等等。这些工作都是要建立在数据积累的基础之上，DeepFlow ®平台也是基于此来进行对微服务各维度的画像评估。



图：SLO 指标量化

在以上的示例中，应用中涉及到几十个微服务，通过运行的历史数据统计分析，以“每 5 分钟 WEB 前端微服务的 HTTP 异常总数”为指标进行量化，在一个月的运行周期中，能实时监控业务是否达到 99.99%的可用性要求（Service Level Object），并分析出潜在影响可用性的各种原因。

7. 总结

DeepFlow ®容器化微服务可观测性方案，面向公有云 K8s、容器环境，利用 eBPF 等新技术的零侵入特性实现对网络、系统、应用的全栈黄金指标的采集，对服务调用链以及 Service mesh、iptables/ipvs、NAT

的逐跳链路追踪，对服务访问的零采样全留存，并结合云资源知识图谱和变更事件数据，搭建立体化的微服务可观测平台。

了解更多信息

专业的售前技术支持及商务合作，协助您选择最合适的解决方案

详询：400-9696-121

网址：www.yunshan.net

北京云杉世纪网络科技有限公司

北京市海淀区成府路 28 号优盛大厦 A 座 1209

版权所有 © 2021 Yunshan Networks 保留所有权利。 本资料中的文字内容和产品相关图片未经北京云杉世纪网络科技有限公司书面许可禁止擅自摘抄、复制部分和全部内容，并不能以任何形式传播。