

故障定位智能体的建设挑战和实践经验

向阳 云杉网络

主办单位：中国计算机学会（CCF）

承办单位：中国计算机学会互联网专委会、中国科学院计算机网络信息中心、中国移动研究院、清华大学

协办单位：华为2012实验室、阿里云、中兴通讯、中国移动九天团队、南开大学、西安电子科技大学、清华大学计算机科学与技术系、神州灵云

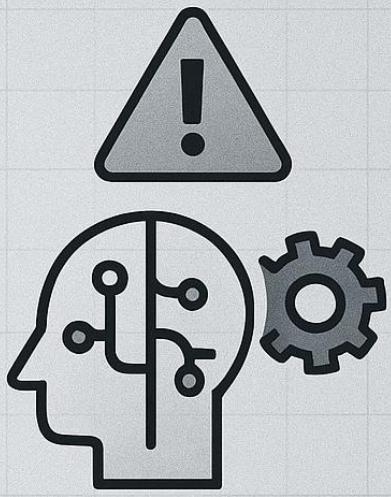
目录 CONTENTS

- AIOps 故障定位的挑战和困境
- 解决全栈数据采集和关联的挑战
- 解决告警收敛和责任关联的挑战
- 解决持续学习和量身定制的挑战
- DeepFlow 智能体的探索和经验

传统的AIOps产品遇到的两大难题



可观测性数据
质量不高
(观测盲点
观测数据孤岛)



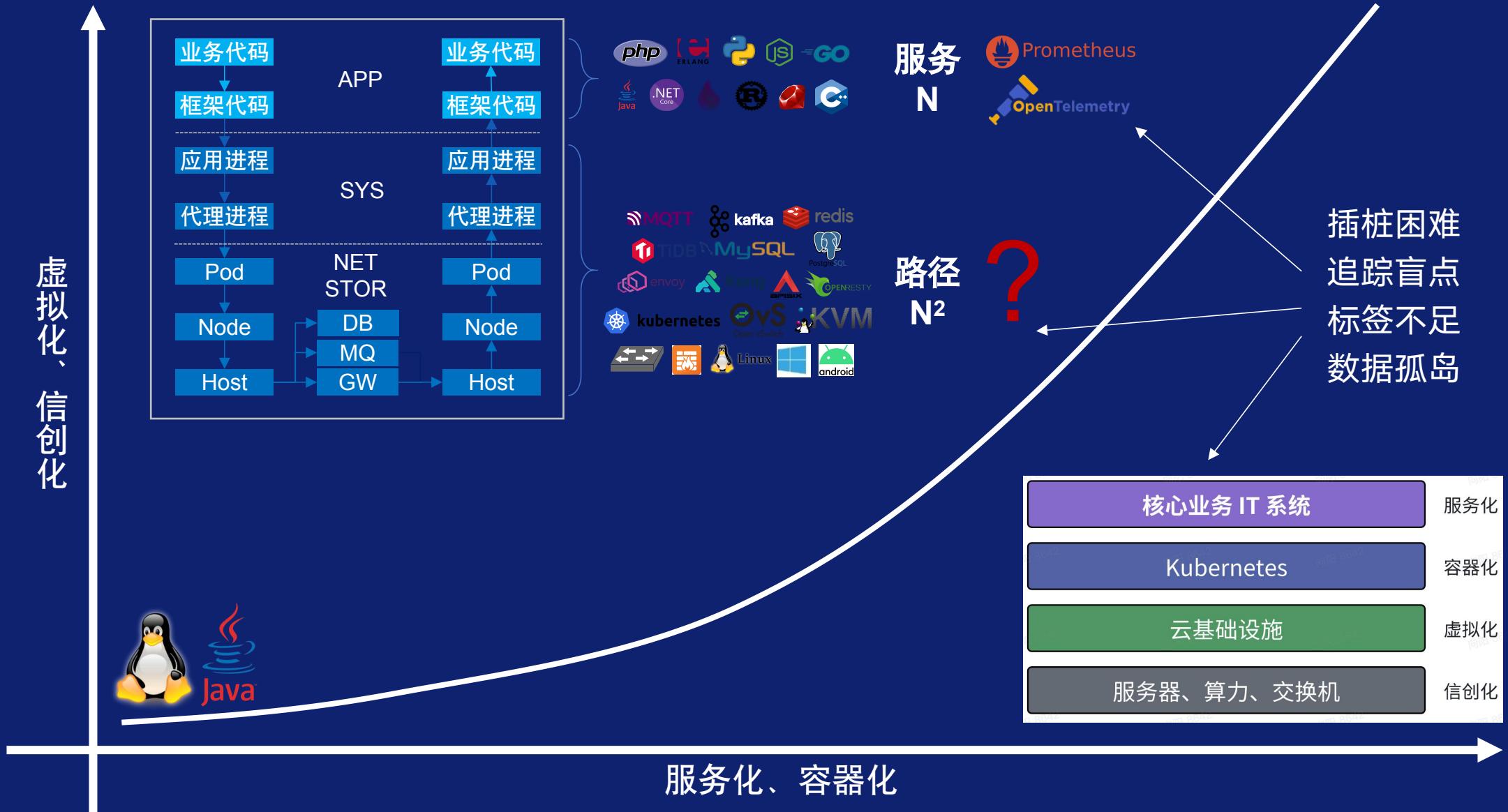
问题分析的
智能不足
(依靠机器学习、小模型)
和基于规则的专家经验)
缺乏智能

OpenRCA

Leaderboard

Method Name	Model	Org.	Correct ↓	Partial	Date
RCA-Agent	Claude 3.5 Sonnet		11.34%	17.31%	2025/1/23
RCA-Agent	GPT-4o		8.96%	17.91%	2025/1/23
Prompting (Oracle)	Gemini 1.5 Pro		7.16%	23.58%	2025/1/23
Prompting (Balanced)	Gemini 1.5 Pro		6.27%	24.18%	2025/1/23
Prompting (Oracle)	GPT-4o		6.27%	15.82%	2025/1/23

挑战 #1：缺失全栈观测数据

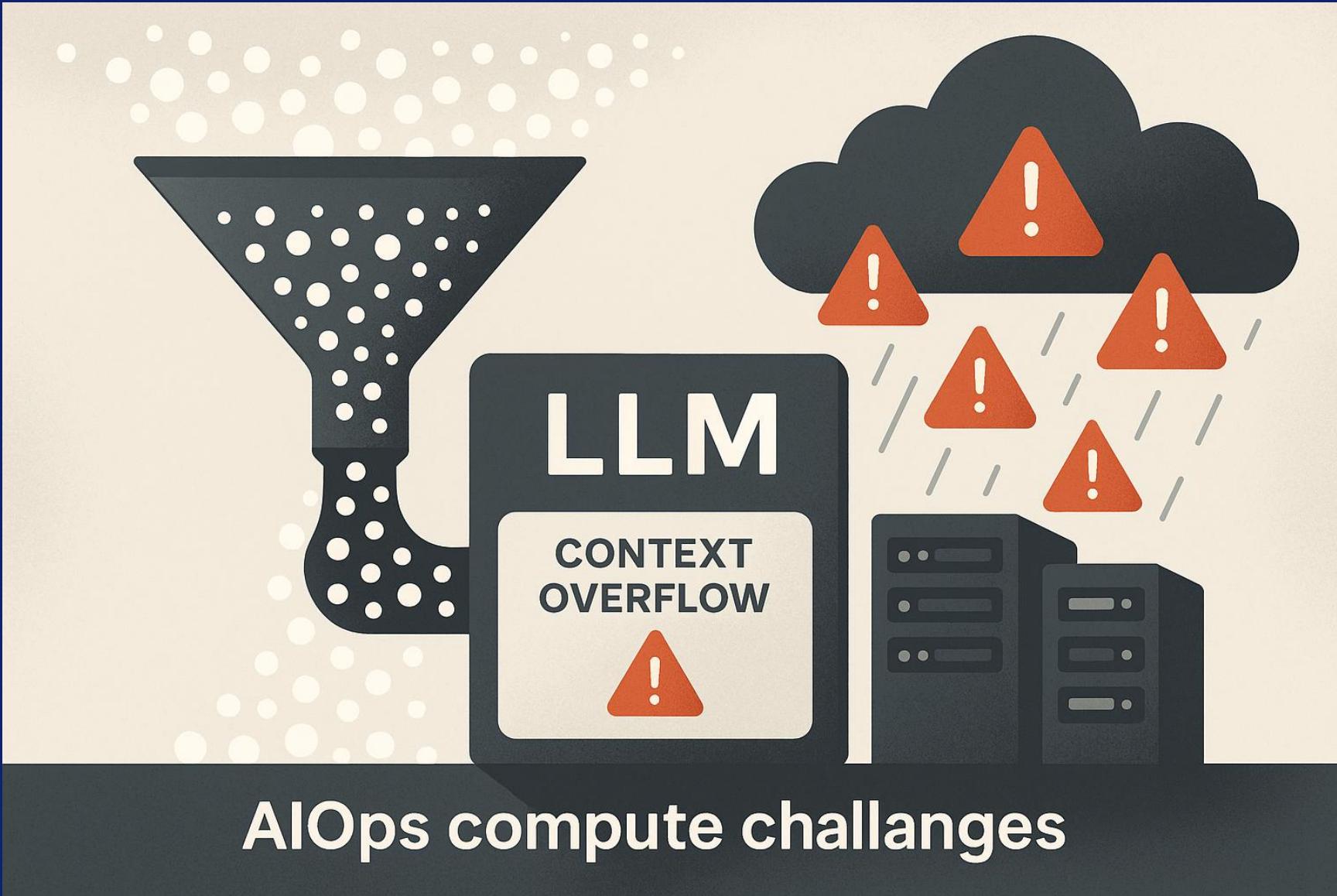


挑战 #2：对算力的需求

AIOps | 2025 CCF国际AIOps挑战
赛 25 CCF International AIOps Challenge

海量观测数据 · 上下文长度

基础设施故障 · 告警风暴

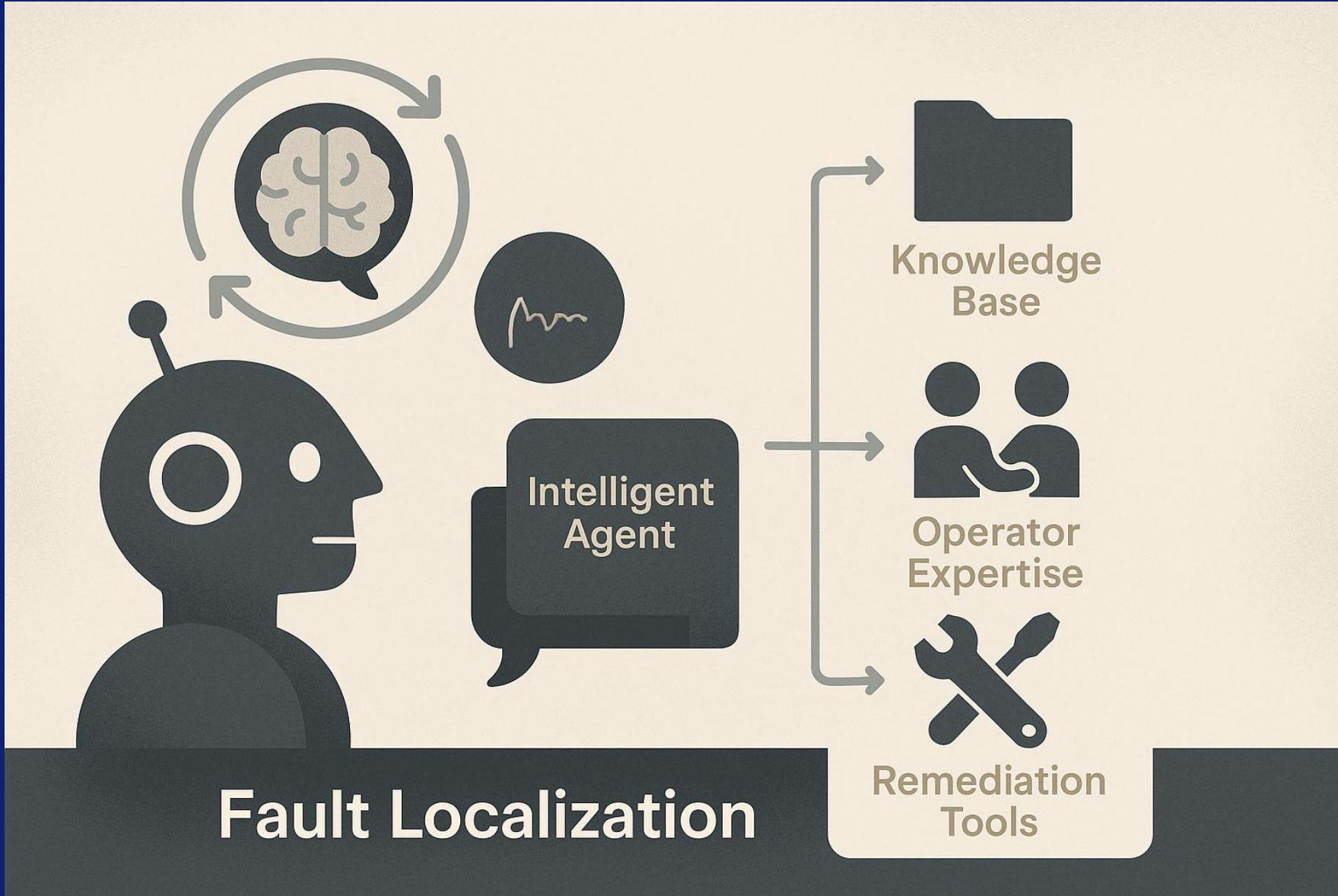


挑战 #3：为企业提供服务

AIOps | 2025 CCF国际AIOps挑战
赛 25 CCF International AIOps Challenge

私有化部署 · 持续学习

私有数据 · 为企业量身定制



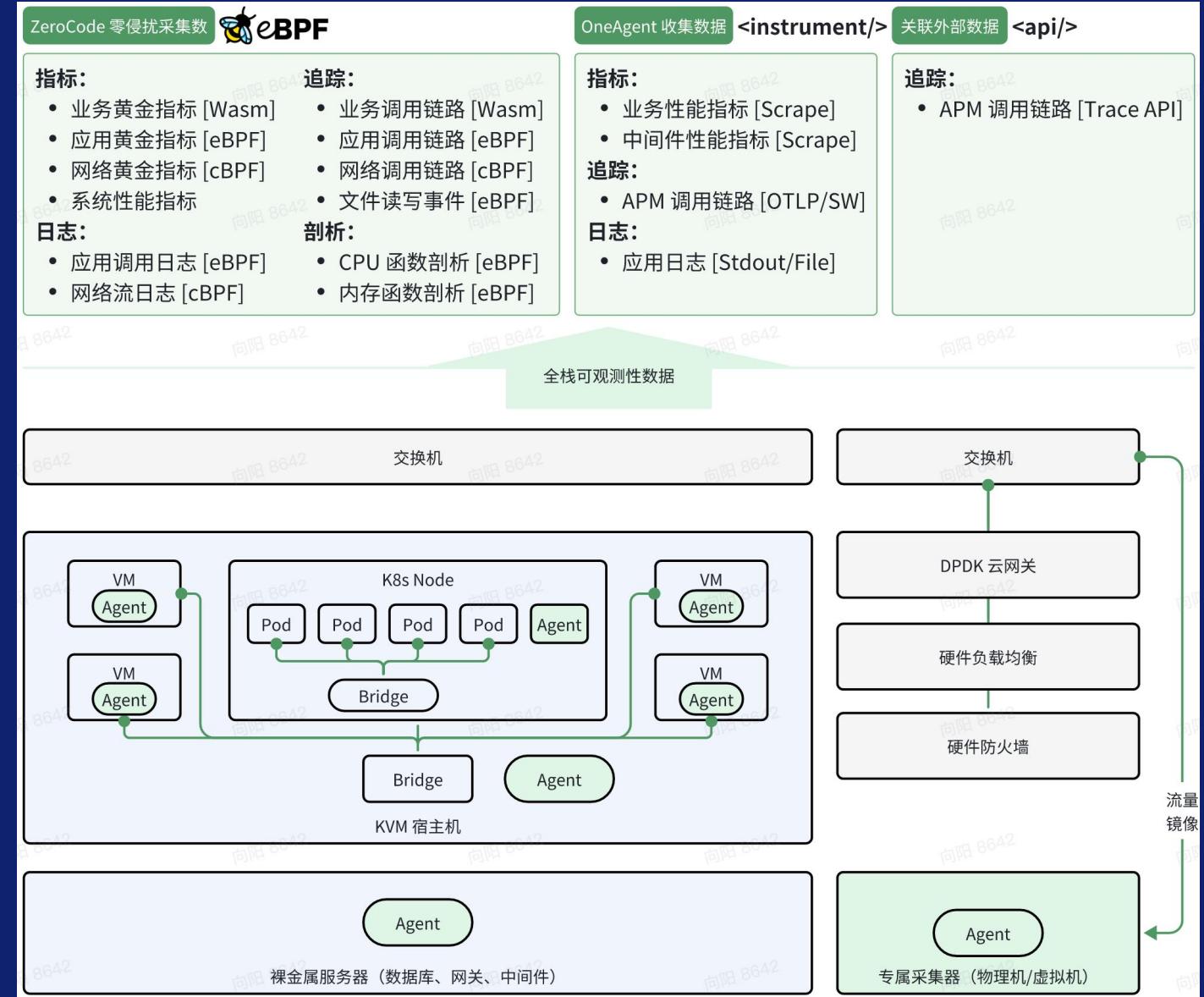
解决数据的挑战：全栈可观测数据的采集和统一标注

AIOps | 2025 CCF国际AIOps挑战
赛 25 CCF International AIOps Challenge



零侵扰采集
统一标注

- 零侵扰：业务无需插码、无需重启
- 全栈追踪：业务、应用、系统、网络
- 智能标注：云资源、容器服务、CMDB



解决算力的挑战：使用多模型的感知、以业务为中心

AIOps | 2025 CCF国际AIOps挑战
赛 25 CCF International AIOps Challenge

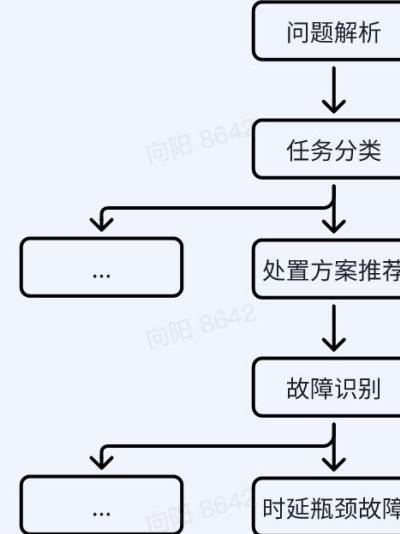


DeepFlow 诊断智能体：工作流自动规划、智能推理

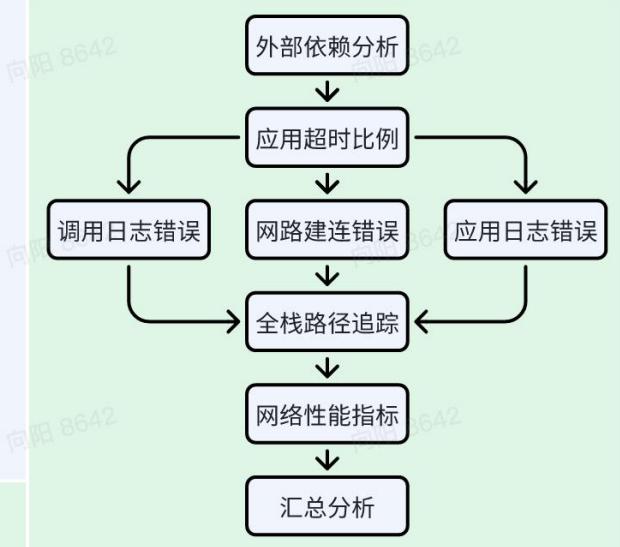
AIOps | 2025 CCF国际AIOps挑战
赛 25 CCF International AIOps Challenge



Planning / 规划



Reasoning / 推理



Reasoning / 推理



- 基于先验知识自动规划
- 基于实时感知动态推理
- 支持深度定制

DeepFlow 诊断智能体：全栈观测数据的感知、推理

AIOps | 2025 CCF国际AIOps挑战赛
2025 CCF International AIOps Challenge



DeepFlow 诊断智能体：处置方案的推理、学习、预测

AIOps | 2025 CCF国际AIOps挑战赛
25 CCF International AIOps Challenge



处置方案推荐

关键信息提取

方案推荐

1.APP-LIMIT-002-启用API限流策略 (可能性: 80%)

- 匹配理由: 电商系统/shop/full-test API在09:00-09:10响应慢, 该预案通过限流应对请求量突增, 缓解API时延瓶颈。

展开▼

2.SYS-PARAM

- 匹配理

3.CONT-ROL

- 匹配理

根因分析报告

电商系统 /shop/full-test 在 02:00-02:10 时段响应慢的根因是 svc-order 服务对下游 db-demo 数据库依赖时, 数据库连接数耗尽 (MySQL 1040 错误), 导致大量请求无法及时获得数据库连接, 进而引发整体响应时延显著升高。svc-order 服务本身资源与并发处理能力充足, 无资源瓶颈或网络异常。

展开▼

详情

方案推荐

1.DB-CONN-LIMIT-002-修改数据库最大连接数 (可能性: 82%)

- 匹配理由: 电商系统 /shop/full-test 响应慢因 db-demo 数据库连接数耗尽导致, 该方案直接提升最大连接数, 解决连接耗尽问题。
- 执行工具: sandbox-mysql-连接数修改
- 工具参数: cluster:db-demo namespace:default pod:svc-order-7c6fcfd4ddd-twbs workload:svc-order

▶ 执行

2.DB-CONN-CONFIG-001-调整数据库连接参数 (可能性: 82%)

- 匹配理由: db-demo 数据库连接管理不合理引发 /shop/full-test 响应慢, 优化连接参数可提升连接效率, 缓解连接耗尽。

▶ 执行

3.DB-CONN-POOL-004-优化数据库连接池配置 (可能性: 81%)

- 匹配理由: /shop/full-test 响应慢由连接池配置与并发负载不匹配引起, 优化连接池配置可提升并发承载力。

▶ 执行

评估 & 反思

Evaluation & Reflexion

▶ 执行

▶ 执行

精准

DeepFlow 智能体典型使用场景

AIOps | 2025 CCF国际AIOps挑战
赛 25 CCF International AIOps Challenge

消除隐患，优化 MTBF

趋势分析

电商业务 持续中

Redis 协议统计到大量的服务端错误

业务是否受影响 否 影响服务总数 5 异常请求总数 1547

「修复建议：检查 Redis 用户名密码配置，并确保用户未被禁用」

电商业务 已结束

第三方支付服务稳定性分析

业务是否受影响 是 影响服务总数 1 异常请求总数 42

「修复建议：实现智能重试机制，针对检测到的RST报文立即触发重试，并设置最多3次」

不间断巡检

实现了"预防为主"的理念

快速处置，降低 MTTR

相关参数：

- 业务名称: 电商系统
- 对像类型: service
- 对像名称: Frontend
- 时间范围: 10:38 - 11:39

第二步：根因分析

问题分析流程

工具调用结果

应用故障：时延超过 1s 的情况，发生在 11:06 - 11:10 分之间。

热点列表：/shop/cart/checkout, /product, /cart 存在响应时延超过 1s 的情况。

请输入问题描述

分钟级诊断

解决了"快速止血"的问题

提升资源和算力利用率

第三步：分析报告

尊敬的领导：

“知己知彼，百战不殆。”—《孙子兵法》

我们向您汇报关于电商系统的容量评估的分析结果。作为公司的核心业务系统，确保其具备足够的容量以应对业务增长和流量高峰至关重要，当前的带宽状况及未来扩展需求。

根据全面的系统分析，我们可以为您提供以下电商系统的并发承载能力的详细评估：

• 当前系统状况

- 业务逻辑：高峰期平均 QPS 为 1500，峰值 QPS 达 1850
- 系统瓶颈：frontend 服务 P95 峰值时延 350ms，错误率仅 0.2%
- 关键系统：recommendationservice 服务 P95 时延超过 1s

• 关键发现

- 热点接口调用：
 - /hipstershop.RecommendationService/ListRecommendations 接口
 - 请逐步提高吞吐量 2.31GB
 - 内存类型为高带宽/低延迟型
 - 这一发现极为关键，表明系统瓶颈不仅限于计算能力不足，更是架构设计不合理
- 资源利用情况：
 - 计算资源：集群 CPU(32%) 和内存 (35%) 使用率较低
 - 网络带宽：带宽利用率仅 45.6%，网络状况存在瓶颈
 - Redis 资源：CPU 使用率 18.9%，并发量 4.32K，表现良好
- 并发承载能力分析：
 - 当前承载下：
 - 可支持约 3000-3700 QPS (当前峰值的 1.6-2 倍)
 - 主要受限于 recommendationservice 服务的静态资源处理能力
 - 实施优化后：
 - 将缓存体内容迁移至 CDN 后，系统可支持 8000-10000 QPS (当前峰值的 4-5 倍)
 - 再增加 recommendationservice 服务数，可进一步提升至 15000 QPS 以上

电商系统当前的瓶颈主要在于 recommendationservice 服务处理大量静态资源内容的架构设计不合理，而非资源不足。通过将静态资源由增加服务器核数，系统并发承载能力可提升 5-8 倍，足以支撑业务长期发展。若不进行任何优化，系统在 QPS 达到约 3000 时将降级；实施建议的优化后，系统将轻松支持 10000+ QPS 的并发访问。

请输入问题描述

一句话问数

满足了"及时决策"的需求





THANKS

主办单位：中国计算机学会（CCF）

承办单位：中国计算机学会互联网专委会、中国科学院计算机网络信息中心、中国移动研究院、清华大学

协办单位：华为2012实验室、阿里云、中兴通讯、中国移动九天团队、南开大学、西安电子科技大学、清华大学计算机科学与技术系、神州灵云