AIOps Challenge

2023 CCF国际AIOps挑战赛决赛
暨“大模型时代的AIOps”研讨会

基于大模型和多AGENT协同的自主决策、自动修复运维方案

轻舟已过万重山

陈东辉、陈润、陈子扬、郭广宇、胡彬、吕颖、谢志鹏、袁俊、张楠、张曦

华为技术有限公司

主办单位：中国计算机学会（CCF）、清华大学、中国建设银行股份有限公司、南开大学

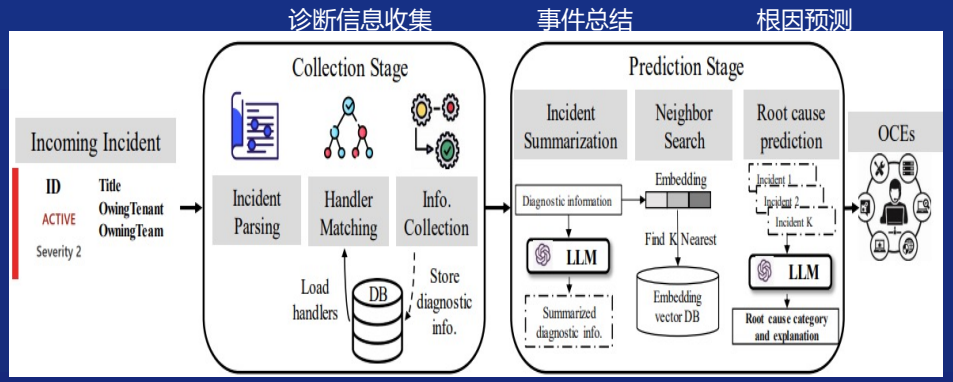
承办单位：中国计算机学会互联网专委会、清华大学计算机科学与技术系、中国建设银行股份有限公司运营数据中心、南开大学软件学院、北京必示科技有限公司

赞助单位：华为技术有限公司、国网宁夏电力有限公司电力科学研究院、软通动力信息技术（集团）股份有限公司

大模型给AIOps带来新机遇

AIOps Challenge

云厂商利用大模型对运维事故进行根因定位并给出故障缓解措施建议



7成以上运维人员对LLM的分析结果满意(>3分)[2]

Score	# of incident owners	In percent (%) of total
5	2	7.41
4	**9**	**33.33**
3	8	29.63
2	6	22.22
1	2	7.41

合计
70+%

针对运维领域海量知识快速获取、辅助诊断和故障分析能力

- 将LLM较为广泛的知识储备（横向能力）与运维领域专业知识（运维垂域）相结合
- 运维工具繁多，利用LLM+多智能体协同使能运维自动驾驶

针对多模态数据进行快速高效准确的异常检测能力

- 多损益、多类型数据：Log、Metric、业务黄金指标、Trace
- 数据异常不等于故障发生，识别可能造成运维故障的数据异常波动

针对多源复杂部署的运维数据进行快速根因定位能力

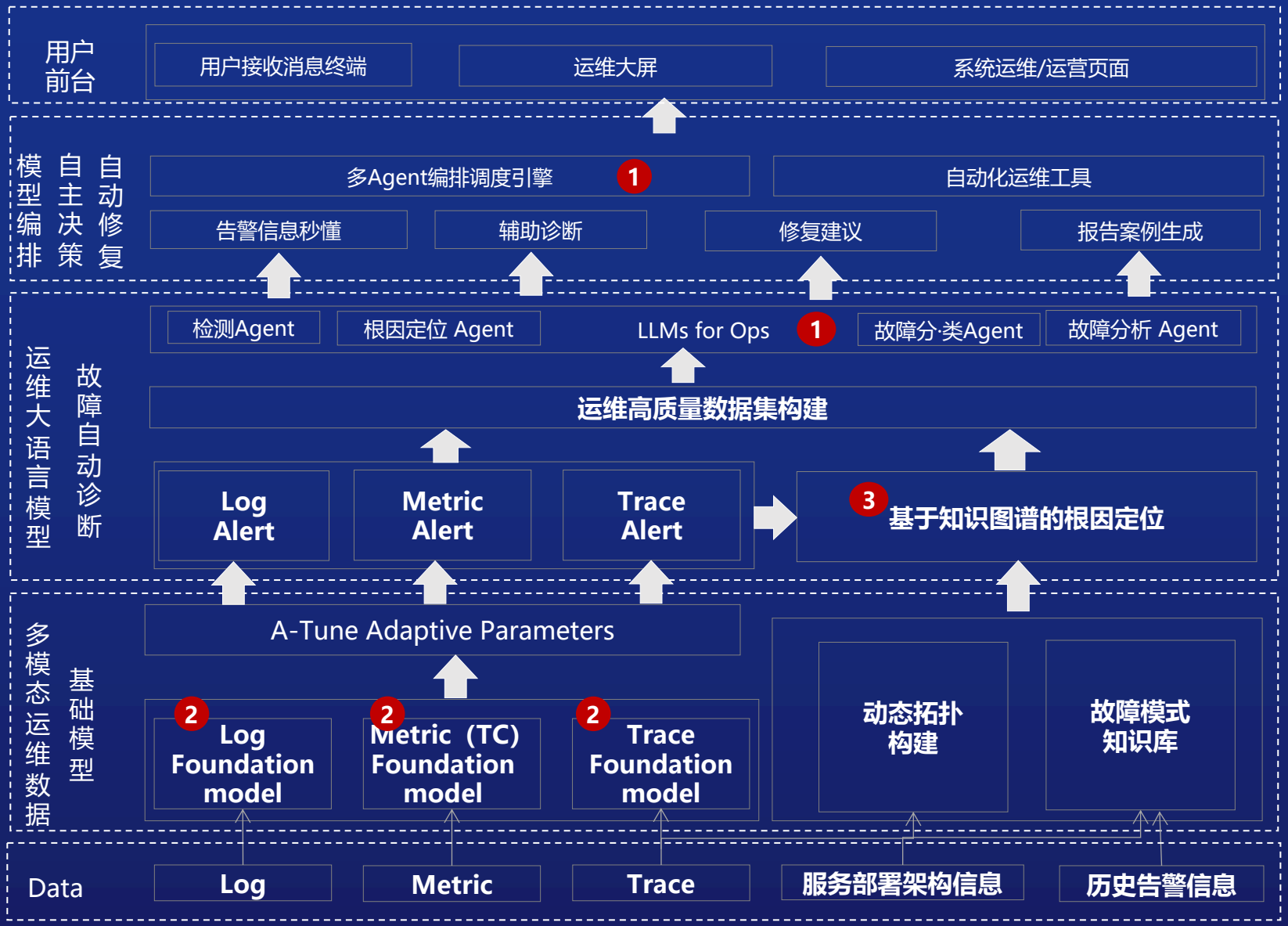
- 告警繁多、区分故障传播节点和故障根因节点
- 历史排障经验积累数字化，形成知识库

参考：
[1] Microsoft: Empowering Practical Root Cause Analysis by Large Language Models for Cloud Incidents
[2] Recommending Root-Cause and Mitigation Steps for Cloud Incidents using Large Language Models

HUAWEI

基于大模型和多Agent协同自主决策、自动修复运维方案

AIOps Challenge



关键技术

1 基于多Agent协同的编排调度

①梳理异常检测告警信息和运维故障模式，形成运维领域高质量数据集，使用finetuning+外挂向量数据库的方案，使得LLM具备运维领域故障分析定位能力；

②使用主管LLM对异常问题与子领域Agent进行桥接，多个子领域Agent协同工作，提升运维效率

2 更全能的多模态数据异常检测基础模型

①针对成百上千条metric序列采取一定维度聚合的方法，化繁为简，一旦检测出异常，再细化分析关键指标的关键表现；

②针对半结构化和非结构化日志采取针对性的解析方式：前者注重模板提取，后使用uADR/sADR进行异常检测，后者注重语义理解，使用BigLog预训练模型结合Deep SVDD+SAD进行异常检测

③针对结构化明显的trace数据，抽取分离调用关系并定义节点关键数据，通过转化时间序列有效识别异常节点，并通过拓扑关系分析帮助推测可能根因节点及故障传播方向

3 基于知识图谱的根因定位

①利用DBSCAN从时间维度聚类，形成异常事件

②利用Trace数据实时生成拓扑结构图

③根据故障模式知识库分析故障传播链路

HUAWEI

多模态异常检测-Metric基础模型

AIOps Challenge

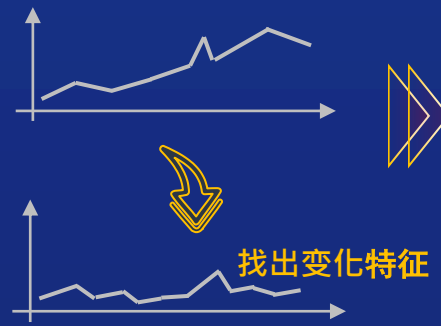
Metric异常检测

① 按采样频率分组

- 采样频率1
- 采样频率2
- 采样频率3
- ...

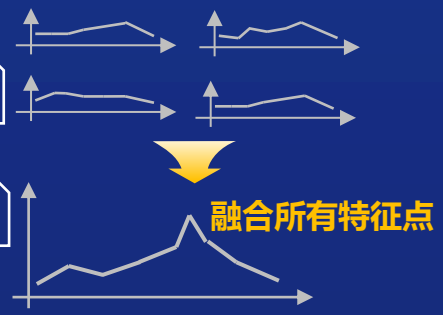
② 提取指标变化特征

- **原始指标曲线**
- 指标标准化
- 窗口差分取绝对值
- **指标变化特征曲线**



③ 同类型指标时序融合

- **指标变化特征曲线**
- **融合曲线**



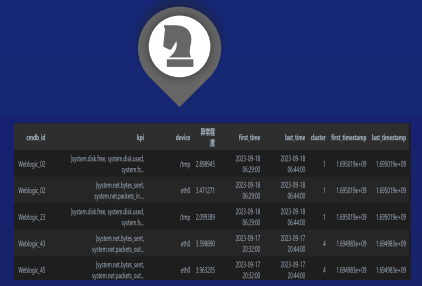
④ 异常检测与时序聚类

- 融合曲线
- **多个异常检测器**
- 针对单个融合指标的DBSCAN时序聚类
- 异常时间戳整体DBSCAN时序聚类
- **系统异常发现！**

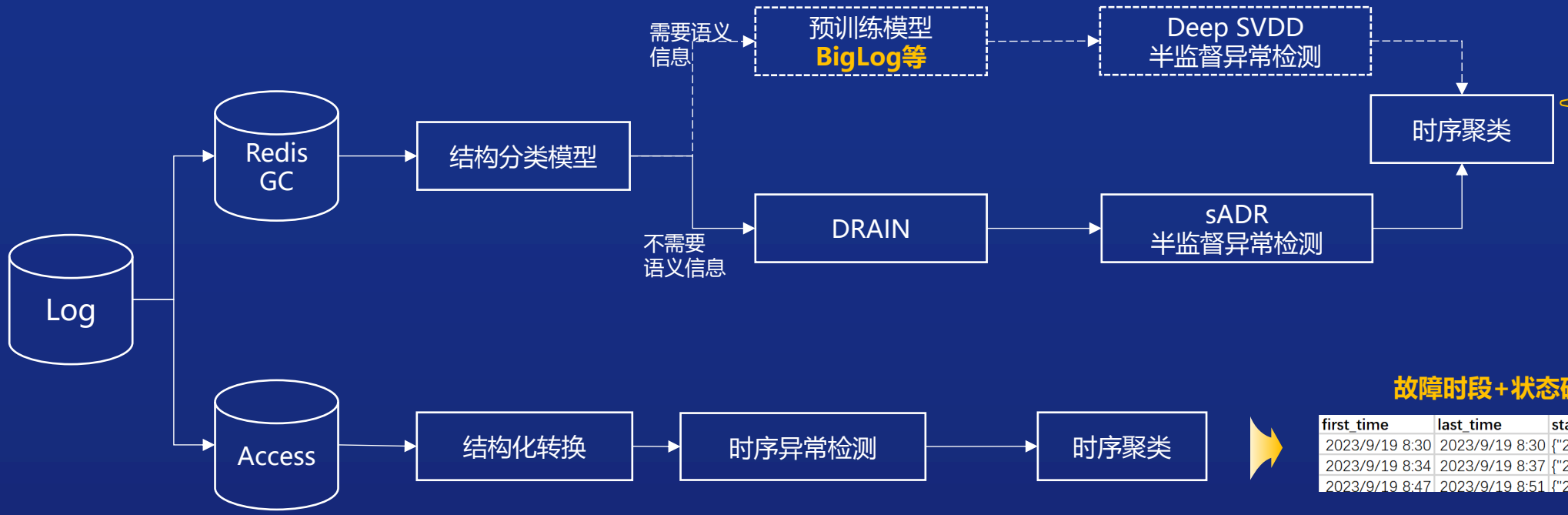
LLM主管Agent**决策**

⑤ 检测Agent

- 根因单元的异常对应类型全部指标变化特征曲线
- 多个异常检测器
- **异常指标**



多模态异常检测-Log基础模型



故障时段+关键异常日志文本信息

first_time	last_time	log_info
2023/9/18 6:37	2023/9/18 6:38	["<*> <*> [GC (Metadata GC Threshold) [PSYoungGen: <*> <*> <*> secs] [Times:
2023/9/18 6:51	2023/9/18 6:53	["<*> <*> [Full GC (Metadata GC Threshol [PSYoungGen: <*> [ParOldGen: <*> <*> [Metaspace: <*> <*> secs] [Times: <*> <* secs]", "<*> <*> [GC (Metadata GC Thresh [PSYoungGen: <*> <*> <*> secs] [Times: <*> secs]", "<*> <*> [GC (System.gc()) [PSYoungGen: <*> <*> <*> secs] [Times:

故障时段+状态码信息

first_time	last_time	statucode_count
2023/9/19 8:30	2023/9/19 8:30	{"200": 26}
2023/9/19 8:34	2023/9/19 8:37	{"200": 61, "500": 2}
2023/9/19 8:47	2023/9/19 8:51	{"200": 32, "500": 11}

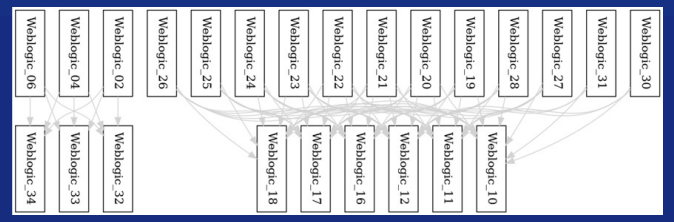
- ✓ **日志类型全覆盖**：自动识别日志结构，对结构化、半结构化和非结构化日志自适应匹配时序异常检测、基于日志解析提取模板的半监督异常检测和基于预训练语言模型BigLog理解语义的半监督检测算法
- ✓ **高效率在线检测**：可对流式日志数据实时检测并实时更新模型

多模态异常检测-Trace基础模型

AIOps Challenge

① 拓扑图构建

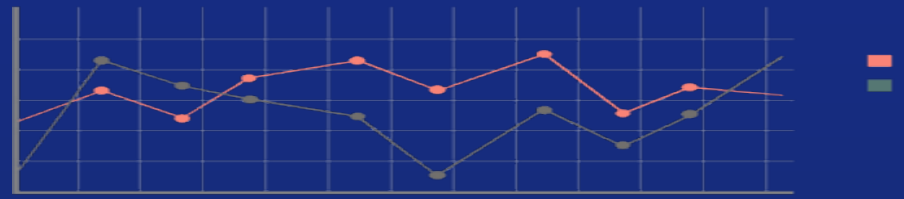
TraceId	cmdb_id	ParentSpan	…
e4670ad5630a1096874ffff6efaa6c28	Weblogic_10	Weblogic_27	…
76b0f51edd1753898209a29920a75062	Weblogic_11	Weblogic_19	…
76b0f51edd1753898209a29920a75062	Weblogic_16	-	…



② 时间序列构建

各节点真正耗时 = 当前节点耗时 - 子节点总耗时

对每个二元组 (cmdb_id, kind) → Timestamp: RealDuration



③ 异常检测算法

时间序列异常检测器

节点耗时序列 → 异常记录 → **基于知识消除误报** → 过滤后的异常记录 → 异常事件

时序聚类

- ✓ **高效：** 将Trace数据转化为时间序列，形成一套时序异常检测算法+不同参数的轻量化方案，秒级异常检测
- ✓ **实时：** 通过实时拓扑关系分析帮助推测可能根因节点及故障传播方向

HUAWEI

多Agent协同框架：基于企业组织管理方法的高效协同诊断

AIOps Challenge

Single-Agent增强

zero-shot → one-shot → few-shot

Chain of Thought → Tree of Thought → Graph of Thought

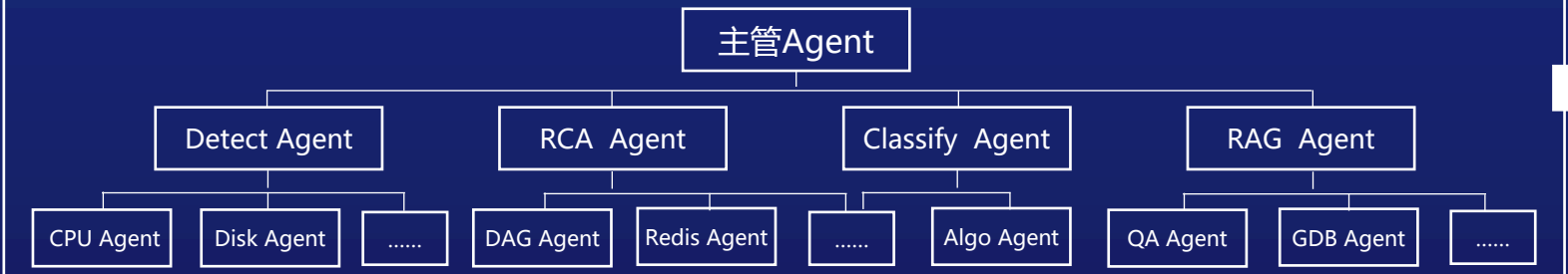
ReAct → Self-ask → Plan-and-execute

Multi-Agent协同

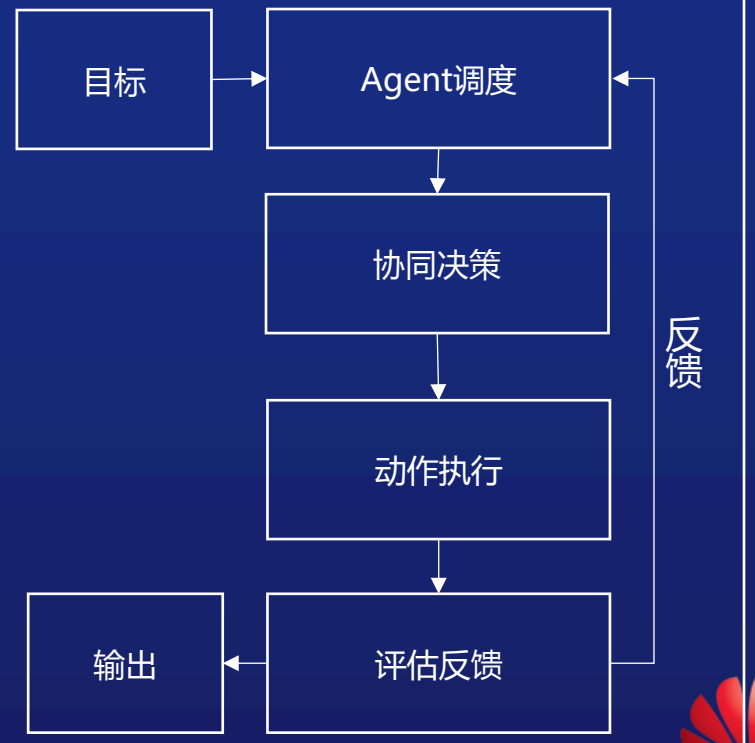
6W2H　KPI树　PDCA　高效会议管理方法

运维SRE War Room

Agents组织架构



多Agent协同诊断



HUAWEI

多Agent协同故障诊断

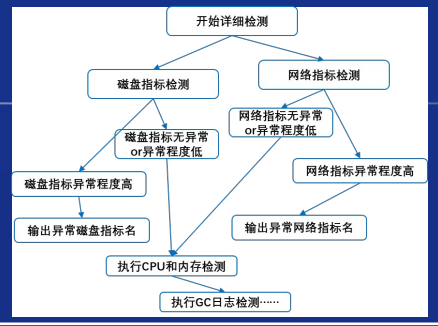
AIOps Challenge

运维知识库（企业内源）

知识图谱
Trace拓扑结构

系统异常发现!

LLM主管Agent**决策**



检测Agent

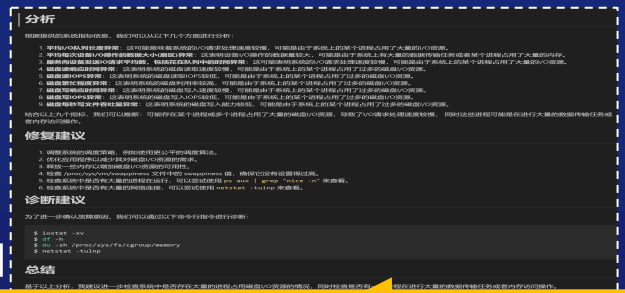
根因定位Agent

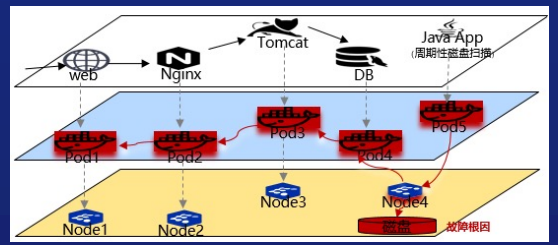
异常通知:
2023/9/20 6:05检测到节点Weblogic_16 eth0=\、eth0= vda出现异常
Weblogic_17 eth0=\、eth0= vda出现异常

主管Agent:节点Weblogic_16Weblogic_17是web节点，需要关注!
请检测Agent执行详细检测，
请根因定位Agent根据详细检测结果查找根因
请故障分析Agent给出故障分析报告返回给我

检测Agent执行磁盘模块检测...有异常，异常程度高，已获得高优先级检测结论，无需进一步检测CPU内存指标，
输出详细检测结果
节点Weblogic_16 eth0=\ ['system.cpu.i_dle']指标异常、异常程度0.849 、异常开始时间2023/9/20 6:05
节点Weblogic_16 eth0=vda ['system.io.avg_q_sz', 'system.io.avg_rq_sz', 'system.io.await', 'system.io.r_await', 'system.io.r_s', 'system.io.util', 'system.io.w_await', 'system.io.w_s', 'system.io.wkb_s']指标异常、异常程度1.049、异常开始时间2023/9/20 6:05
节点Weblogic_17 eth0=\ ['system.cpu.i_dle', 'system.cpu.pct_usage', 'system.load.1', 'system.load.15', 'system.load.5',

根因定位Agent执行根因定位分析中...
根因为Weblogic_16 eth0=vda





提供故障爆炸半径，为恢复策略提供依据

故障分类Agent:已知根因节点异常信息为节点Weblogic_16 eth0=\ ['system.cpu.i_dle']指标异常、异常程度0.849 、异常开始时间2023/9/20 6:05
节点Weblogic_16 eth0=vda ['system.io.avg_q_sz', 'system.io.avg_rq_sz', 'system.io.await', 'system.io.r_await', 'system.io.r_s', 'system.io.util', 'system.io.w_await', 'system.io.w_s', 'system.io.wkb_s']
分析故障类别
输出：故障类别为 磁盘故障

调用运维工具

执行修复建议或进一步诊断

故障分析Agent

故障分类Agent

运维知识库（企业内源）

HUAWEI

方案创新性、通用性及实用性

AIOps Challenge

创新

多Agent协同
完成复杂运维任务

- **基于企业组织管理方法的多Agent协同框架**，复杂运维任务处理更高效
- 多Agent协同完成运维主流程：异常检测->根因定位->故障分类->故障分析->修复建议

通用

解决运维领域
共性问题

- 多模态异常检测基础模型，包含**Trace、Metric、Log数据全方面处理能力，开箱即用**
- 框架与算法不依赖具体特定应用场景，结合大模型实现较强的泛化能力

实用

故障恢复优先
各模块松耦合

- 故障诊断报告体现**可解释的故障爆炸半径**，为实际生产运维故障快速恢复提供有力依据
- 各模块松耦合可插拔，可以全面应用于各类场景故障快恢需求，已在公司内部多场景落地

HUAWEI

参考文献

1. Zhang S, Pan Z, Liu H, et al. Efficient and Robust Trace Anomaly Detection for Large-Scale Microservice Systems. ISSRE, 2023.
2. Li D, Zhang S, Sun Y, et al. An Empirical Analysis of Anomaly Detection Methods for Multivariate Time Series. ISSRE, 2023.
3. Wang Z, Liu Z, Zhang Y, et al. RCAgent: Cloud Root Cause Analysis by Autonomous Agents with Tool-Augmented Large Language Models. arXiv, 2023.
4. Jin P, Zhang S, Ma M, et al. Assess and Summarize: Improve Outage Understanding with Large Language Models. ESEC/FSE, 2023.
5. Chen Y, Xie H, Ma M, et al. Empowering Practical Root Cause Analysis by Large Language Models for Cloud Incidents. arXiv, 2023.
6. Zhou X, Li G, Sun Z, et al. D-Bot: Database Diagnosis System using Large Language Models. arXiv, 2023.
7. Zhou X, Li G, Liu Z. Llm as dba. arXiv, 2023.
8. Wen Q, Gao J, Song X, et al. RobustSTL: A robust seasonal-trend decomposition algorithm for long time series. AAAI, 2019.
9. Liu Y, Tao S, Meng W, et al. LogPrompt: Prompt Engineering Towards Zero-Shot and Interpretable Log Analysis. arXiv, 2023.
10. Tao S, Liu Y, Meng W, et al. Biglog: Unsupervised large-scale pre-training for a unified log representation. IWQoS, 2023.
11. Ma L, Yang W, Xu B, et al. KnowLog: Knowledge Enhanced Pre-trained Language Model for Log Understanding. ICSE, 2023.
12. Zhong Z, Fan Q, Zhang J, et al. A Survey of Time Series Anomaly Detection Methods in the AIOps Domain. arXiv, 2023.
13. Wu H, Hu T, Liu Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis. ICLR, 2023.
14. Yu G, Chen P, Li P, et al. Logreducer: Identify and reduce log hotspots in kernel on the fly. ICSE, 2023.

2023 CCF国际AIOps挑战赛决赛暨“大模型时代的AIOps”研讨会

THANKS

大模型给AIOps带来新机遇

AIOps Challenge

针对运维领域海量知识快速获取、辅助诊断和故障分析能力

- 将LLM较为广泛的知识储备（横向能力）与运维领域专业知识（运维垂域）相结合
- 运维工具繁多，利用LLM+多智能体协同使能运维自动驾驶

针对多模态数据进行快速高效准确的异常检测能力

- 多损益、多类型数据：Log、Metric、业务黄金指标、Trace
- 数据异常不等于故障发生，识别可能造成运维故障的数据异常波动

针对多源复杂部署的运维数据进行快速根因定位能力

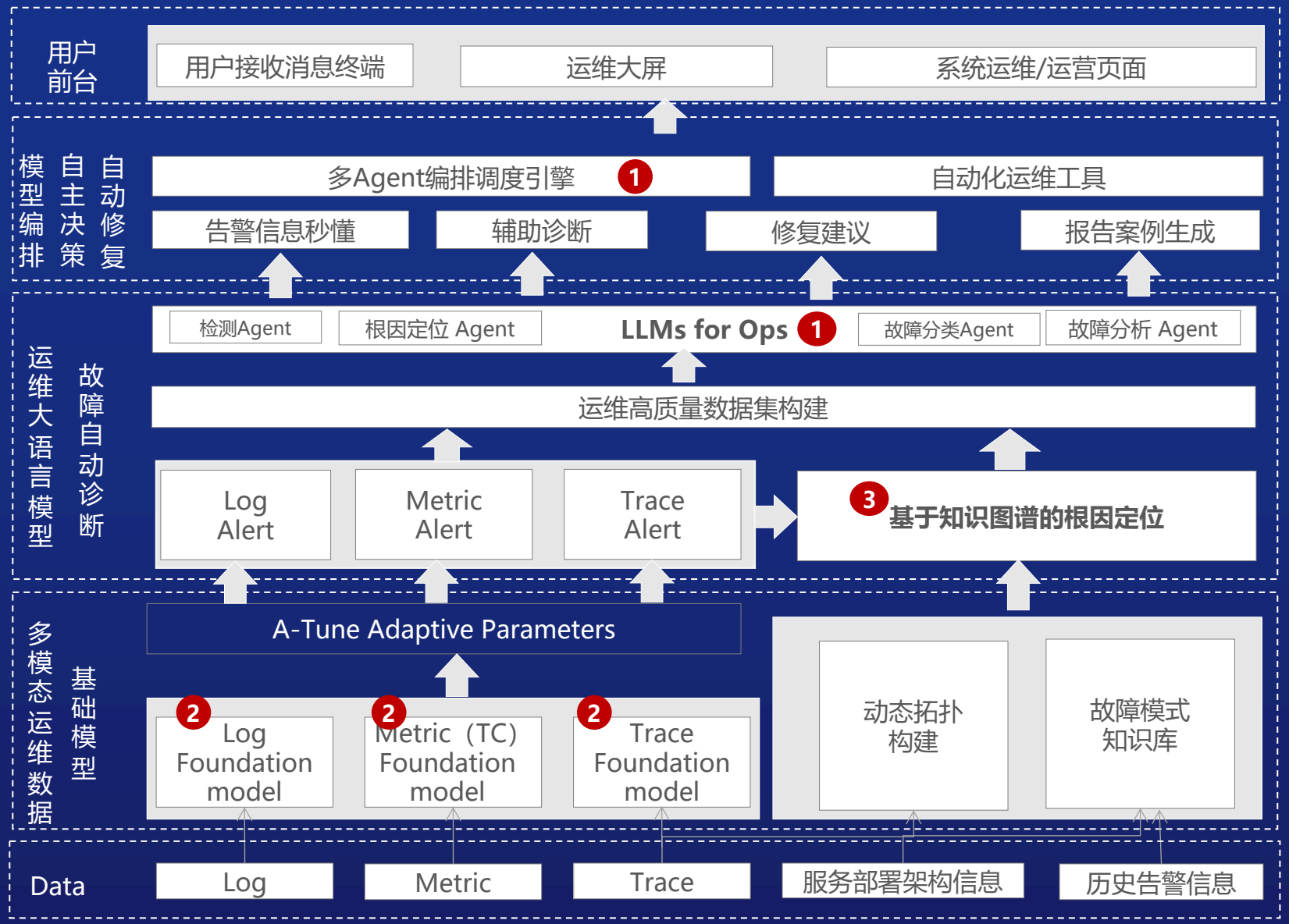
- 告警繁多、区分故障传播节点和故障根因节点
- 历史排障经验积累数字化，形成知识库

	运维能力	是否主要	权重
1	**多模态异常检测（log、metric、trace）**	**是**	**0.5**
2	**多Agent协同故障诊断框架**	**是**	**0.5**
3	根因定位能力	否	0.1
4	基于LLM的故障分类	否	0.1
5	基于LLM的故障分析报告与修复建议	是	0.1

HUAWEI

基于大模型和多Agent协同自主决策、自动修复运维方案

AIOps Challenge



关键技术

1 具备运维领域知识经验的LLM

①梳理异常检测告警信息和运维故障模式，形成运维领域高质量数据集，使用finetuning+外挂向量数据库的方案，使得LLM具备运维领域故障分析定位能力；

②使用主管LLM对异常问题与子领域Agent进行桥接，多个子领域Agent协同工作，提升运维效率

2 更全能的多模态数据异常检测基础模型

①针对成百上千条metric序列采取一定维度聚合的方法，化繁为简，一旦检测出异常，再细化分析关键指标的关键表现；

②针对半结构化和非结构化日志采取针对性的解析方式：前者注重模板提取，后使用uADR/sADR进行异常检测，后者注重语义理解，使用BigLog预训练模型结合Deep SVDD+SAD进行异常检测

③针对结构化明显的trace数据，抽取分离调用关系并定义节点关键数据，通过转化时间序列有效识别异常节点，并通过拓扑关系分析帮助推测可能根因节点及故障传播方向

3 基于知识图谱的根因定位

①利用DBSCAN从时间维度聚类，形成异常事件

②利用Trace数据实时生成拓扑结构图

③根据故障模式知识库分析故障传播链路

HUAWEI

方案创新性、通用性及实用性

AIOps Challenge

创新

多Agent协同
完成复杂运维任务

- 基于LLM实现故障分类、故障分析报告及修复建议
- 多Agent协同完成运维主流程：异常检测->根因定位->故障分类->故障分析->修复建议

通用

解决运维领域
共性问题

- 多模态异常检测基础模型，近开箱即用，少量参数即可完成参数调优
- 基于公域运维语料、知识库，完成运维领域的结构化知识问答

实用

各模块松耦合

- 松耦合的LLM底座，低成本替换、可插拔
- 各模块可插拔，既可以解决运维故障感知问题也可以全面应用于故障定位、排障、修复

HUAWEI

多Agent协同框架：基于专业组织管理方法的高效协同诊断

AIOps Challenge

Agent组织与管理方法

MBO（目标管理） 明确每个Agent目标，提高协作效率。	CW（协同作战） 多个Agent共享资源、信息和知识。
SM（自我管理） Agent自我调整、自我学习和自我修复	KM（知识管理） 通过有效知识共享和创新，提升智能水平
SM（战略管理） 对Agent进行长期规划和发展，持续演进	IM（激励管理） 激发Agent的积极性和创造力

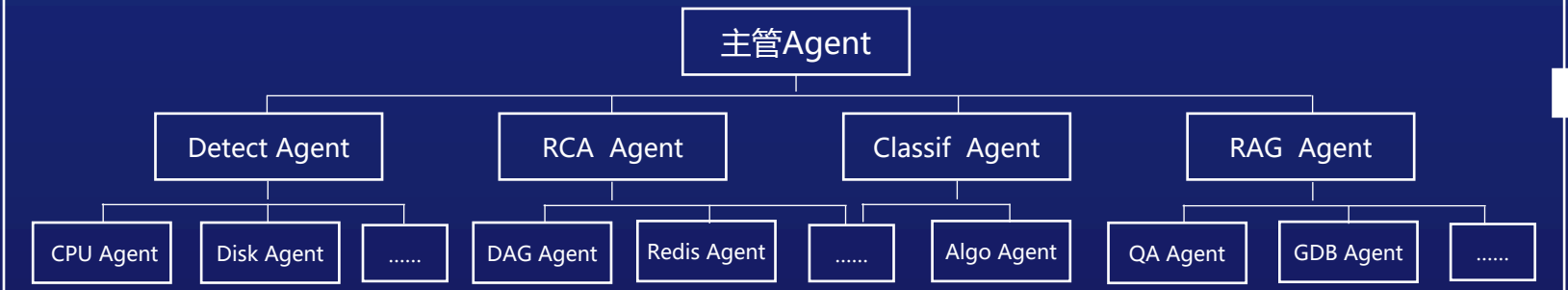
多Agent协同机制

Agent调度机制
根据当前任务描述信息，动态选择一组领域最相关的Agent，避免资源浪费和无关Agent的错误引导

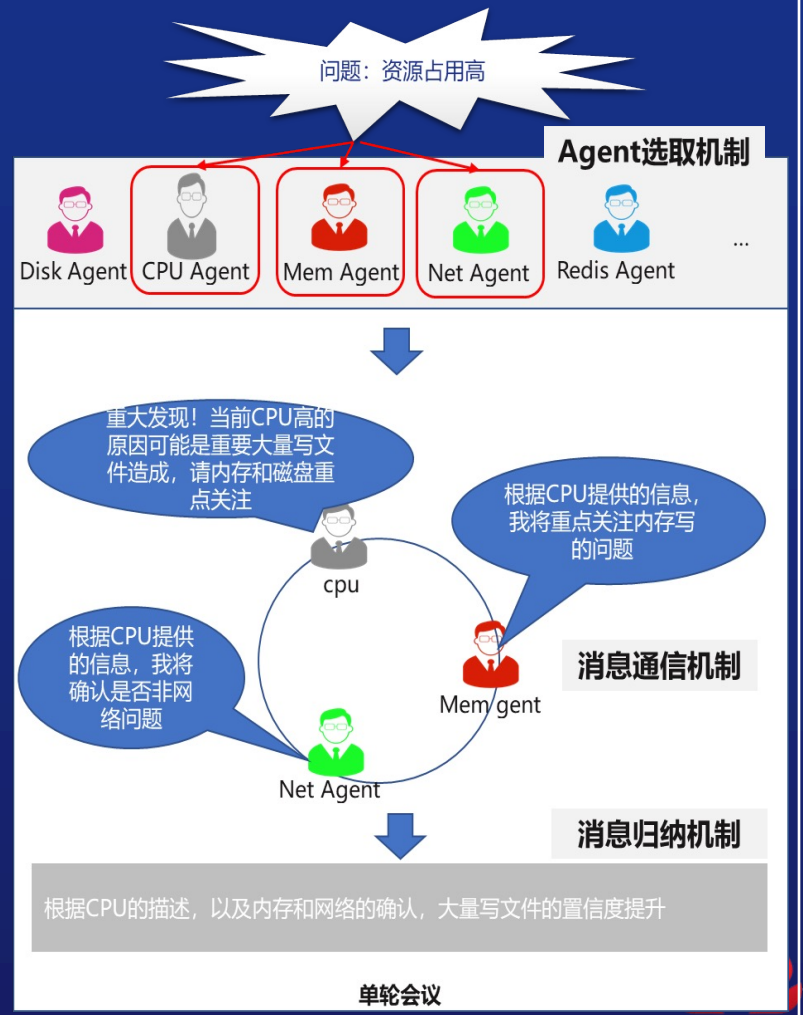
消息通信机制
多个Agent通过异步通信方式（发布-订阅）共享实时进展，减少冗余工作，及时调整分析方向

消息归纳机制
提炼总结协同内容，去除干扰信息，纠正错误路线，归纳得出当前结论

AIOps Agents组织架构



多Agent协同诊断



多Agent协同故障诊断

AIOps Challenge

运维知识库（企业内源）

Trace模块 | Web日志模块

网络模块 | Redis日志模块

xx模块 | CPU模块

知识图谱
Trace拓扑结构

系统异常
发现!

LLM主管
Agent**决策**

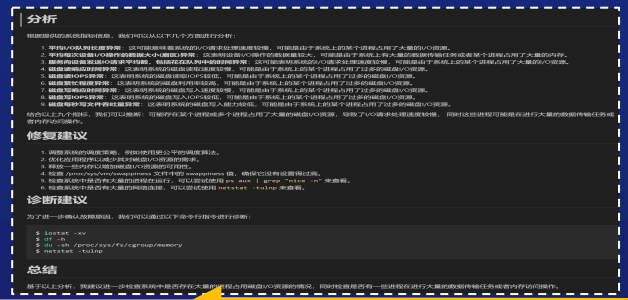
检测Agent

根因定位Agent

异常通知:
2023/9/20 6:05检测到节点Weblogic_16 eth0=\、eth0= vda出现异常
Weblogic_17 eth0=\、eth0= vda出现异常

主管Agent:节点Weblogic_16Weblogic_17是web节点，需要关注!
请检测Agent执行详细检测，
请根因定位Agent根据详细检测结果查找根因
请故障分析Agent给出故障分析报告返回给我

检测Agent执行磁盘模块检测...有异常，异常程度高，已获得高优先级检测结论，无需进一步检测CPU内存指标，
输出详细检测结果
节点Weblogic_16 eth0=\ ['system.cpu.i_dle']指标异常、异常程度0.849 、异常开始时间2023/9/20 6:05
节点Weblogic_16 eth0=vda ['system.io.avg_q_sz', 'system.io.avg_rq_sz', 'system.io.await', 'system.io.r_await', 'system.io.r_s', 'system.io.util', 'system.io.w_await', 'system.io.w_s', 'system.io.wkb_s']指标异常、异常程度1.049、异常开始时间2023/9/20 6:05
节点Weblogic_17 eth0=\ ['system.cpu.i_dle', 'system.cpu.pct_usage', 'system.load.1', 'system.load.15', 'system.load.5',



根因定位Agent执行根因定位分析中...
根因为Weblogic_16 eth0=vda

故障分类Agent:已知根因节点异常信息为节点Weblogic_16 eth0=\ ['system.cpu.i_dle']指标异常、异常程度0.849 、异常开始时间2023/9/20 6:05
节点Weblogic_16 eth0=vda ['system.io.avg_q_sz', 'system.io.avg_rq_sz', 'system.io.await', 'system.io.r_await', 'system.io.r_s', 'system.io.util', 'system.io.w_await', 'system.io.w_s', 'system.io.wkb_s']
分析故障类别
输出：故障类别为 磁盘故障

调用运维工具

执行修复建议或进一步诊断

故障分析Agent

故障分类Agent

运维知识库（企业内源）

HUAWEI