阿里云服务器智能异常调度系统及LLM OPS构建与实践

郭红科　阿里云

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 上海站
K+全球软件研发行业创新峰会
时间：2024.06.21-22

K+峰会 敦煌站
K+思考周®研习社
时间：2024.10.17-19

K+峰会 香港站
K+思考周®研习社
时间：2024.11.10-12

K+峰会详情

AiDD峰会

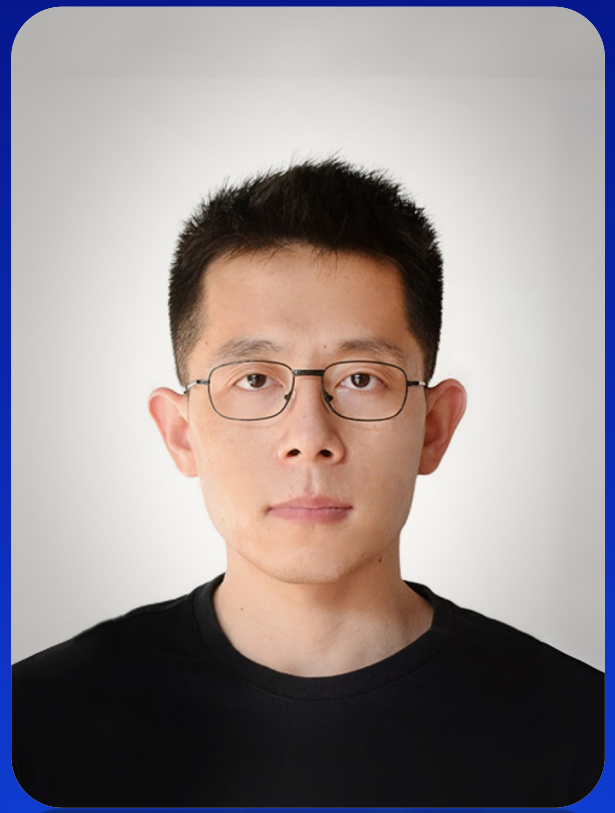
AiDD峰会 上海站
AI+研发数字峰会
时间：2024.05.17-18

AiDD峰会 北京站
AI+研发数字峰会
时间：2024.08.16-17

AiDD峰会 深圳站
AI+研发数字峰会
时间：2024.11.08-09

AiDD峰会详情

演讲嘉宾



郭红科

阿里云 高级开发工程师

毕业于大连理工，一直从事AIOps领域相关工作，专注于日志异常检测、指标异常检测和根因定位等多个方向。

21年加入阿里云ECS异常调度，致力于探究并实现人工智能技术在云计算场景下的创新应用，具有在ECS变更拦截、实时批量风险检测以及ECS性能诊断等关键场景中实现有效解决方案的实战经验。

目录
CONTENTS

PART 01

阿里云智能异常调度体系介绍

异常调度复杂性

X86计算

g8 通用型	c8 计算型	r8 内存型	hfc7 高主频计算型	i4 本地SSD型	d3 大数据型
g7 通用型	c7 计算型	r7 内存型	hfg7 高主频通用型		d2 大数据型
		re7 内存增强			
sn2ne 通用网络增强	sn1ne 计算网络增强	se1ne 内存网络增强	hfr7 高主频内存型	i3 本地SSD型	d1ne 大数据型

异构计算

gn7 GPU	f5 FPGA
gn6 GPU	f3 FPGA

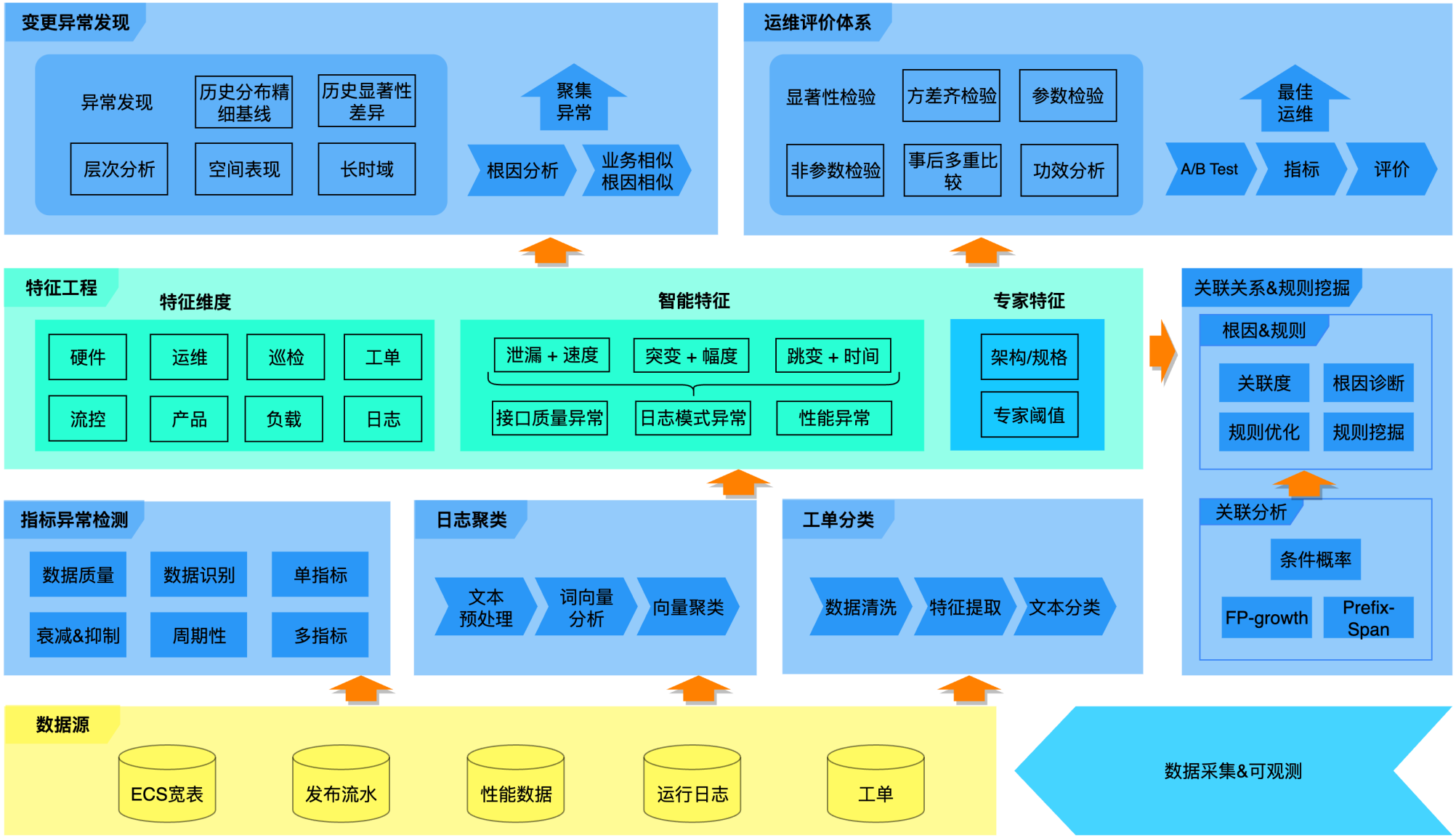
ARM计算

g8 通用型	c8 计算型	r8 内存型
g6 通用型	c6 计算型	

裸金属&高性能计算

- EBM 弹性裸金属（神龙）
- SCC 超级计算集群

智能体系架构图


变更异常发现
异常发现
历史分布精细基线
历史显著性差异
层次分析
空间表现
长时域
聚集异常
根因分析
业务相似根因相似
运维评价体系
显著性检验
方差齐检验
参数检验
非参数检验
事后多重比较
功效分析
最佳运维
A/B Test
指标
评价
特征工程
特征维度
硬件
运维
巡检
工单
流控
产品
负载
日志
智能特征
泄漏 + 速度
突变 + 幅度
跳变 + 时间
接口质量异常
日志模式异常
性能异常
专家特征
架构/规格
专家阈值
关联关系&规则挖掘
根因&规则
关联度
根因诊断
规则优化
规则挖掘
关联分析
条件概率
FP-growth
Prefix-Span
指标异常检测
数据质量
数据识别
单指标
衰减&抑制
周期性
多指标
日志聚类
文本预处理
词向量分析
向量聚类
工单分类
数据清洗
特征提取
文本分类
数据源
ECS宽表
发布流水
性能数据
运行日志
工单
数据采集&可观测

智能体系落地方法论

- 面向平台 → 锦上添花
  - 指标异常检测
  - 日志模式聚类
  - 关联关系发掘
  - ……

- 面向业务 → 雪中送炭
  - 变更拦截
  - 工单分类
  - 性能诊断
  - ……

PART 02

大模型时代对AIOps行业的革新影响

AIOps? MLOps? LLMOps?

	定义	关键	代表
AIOps	AIOps是结合**大数据**和**机器学习**技术，去**自动化**IT运维过程，包括事件关联、异常检测和因果关系确定	AI for Ops	阿里云、必示
MLOps	MLOps是设计、构建、启用和支持在生产中高效部署ML模型的过程和实践，以持续改进业务活动	Ops for ML	阿里云PAI、魔搭社区、Hugging Face
LLMOps	LLMOps的意思是面向LLM的MLOps	Ops for LLM	阿里云PAI 魔搭社区、Hugging Face

AIOps

LLM for Ops

LLM OPS的行业的可能性

道、法、术、**器**、势

通用模型

- 通义千问
- 百川智能
- Meta Llama

领域模型

- LogPatternLLM
- TimeSeriesLLM
- EcsRcLLM
- ……

日志模式提取 Prompt

请对输入的log message进行模式提取，综合考虑日志文本，保留日志中的频繁信息，使用{placeholder}形式替换模式中的变量，
Log messages:
{your messages}

Pattern results:[输出日志模式]

指标异常检测 Prompt

请对给出一段时序序列，序列等距排列，请分析序列，找出其中可能的异常波动，波动的类型有突增、趋势上升等，请给出判断结果和异常趋势开始位置，下面是一些例子：
<例子>
序列：[1,2,3,5,6,7,8]
结果：趋势上升，0
序列：[1,2,3,2,2,3,9]
结果：突增，6
</例子>

序列：{series}
结果：[判断结果]

LLM OPS的行业的可能性

道、法、**术**、器、势

RAG框架

- **Naive RAG**

  朴素的RAG

- **Advanced RAG**

  pre：索引（meta + index）

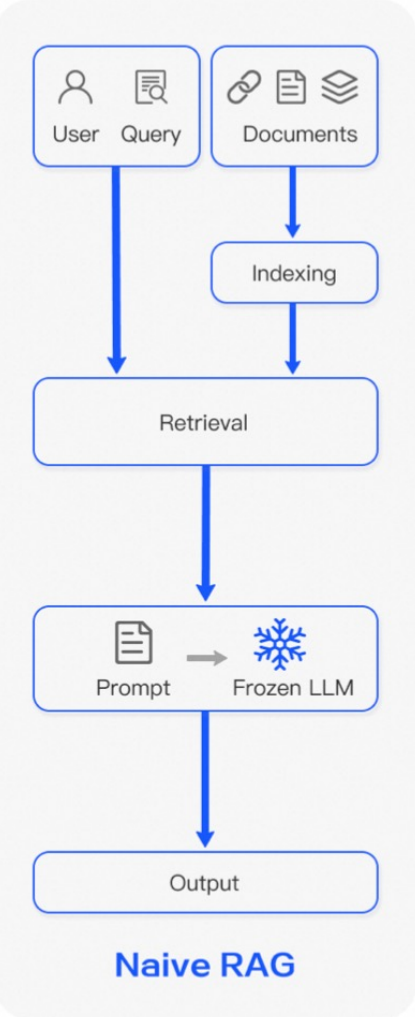
  检索（rewrite、hierarchical）

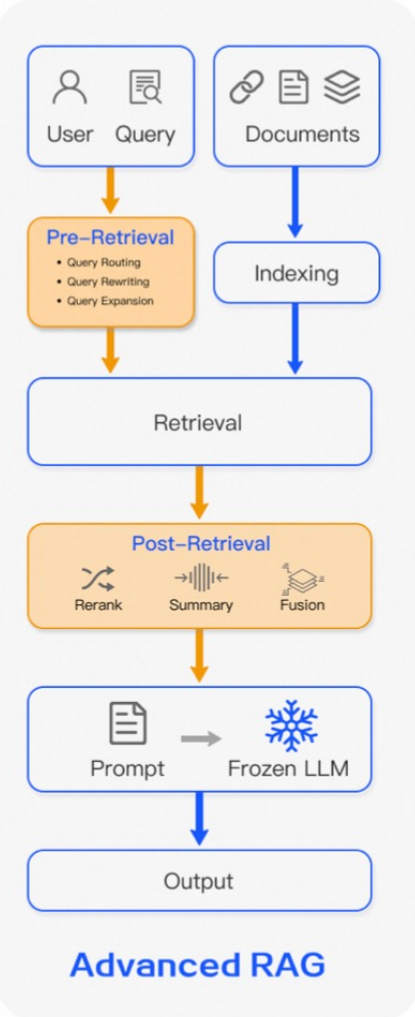
  混合检索

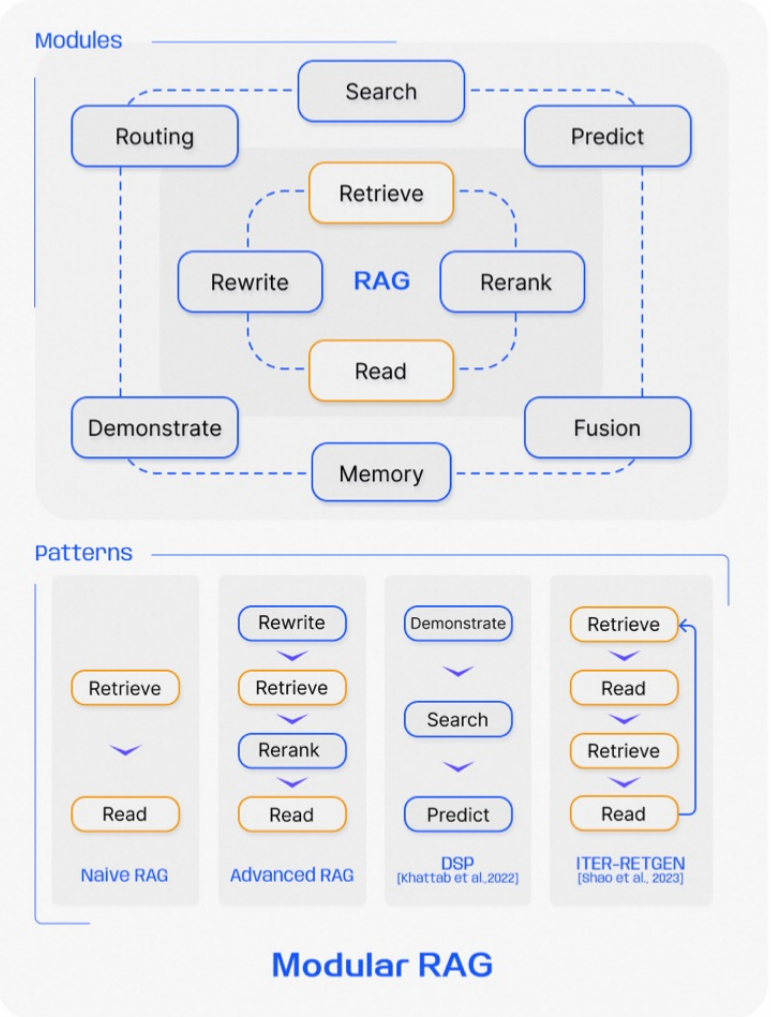
  post：re-rank、compression

- **Modular RAG**

  模块 + 模式：灵活、按需



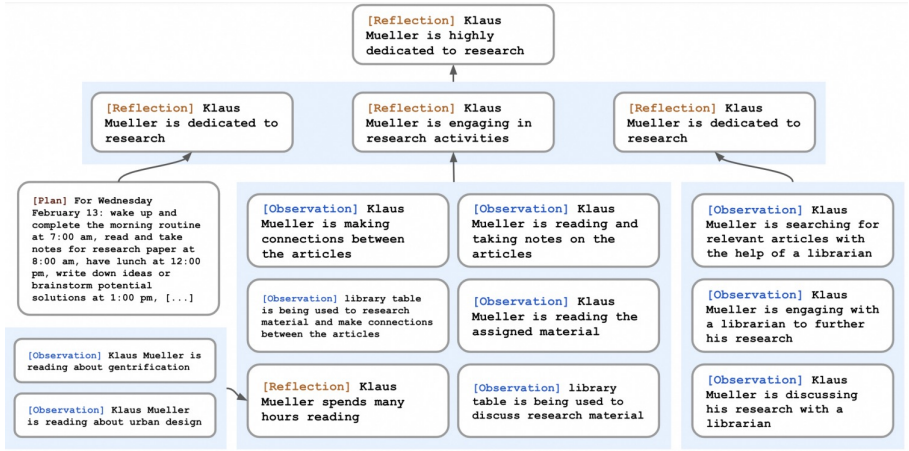


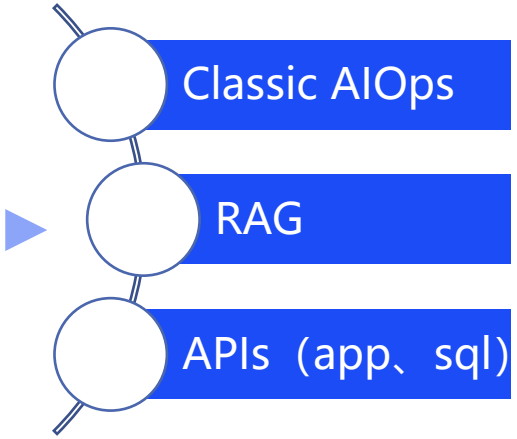


图片来源：[Retrieval-Augmented Generation for Large Language Models: A Survey]

LLM OPS的行业的可能性

道、法、**术**、器、势

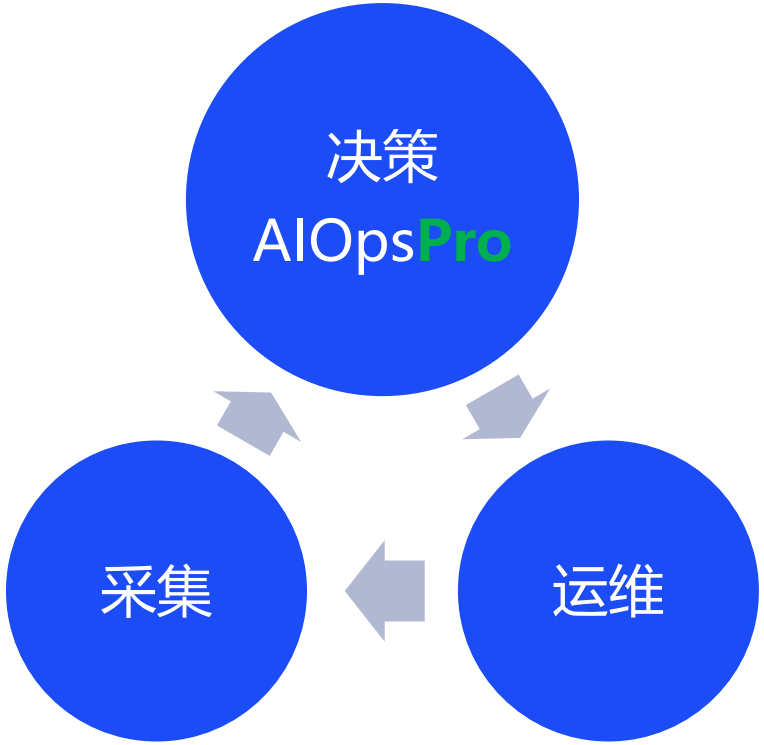




Agent = LLM + memory + planning + tools

- 分解
  - Chain of thoughts
  - Tree of thoughts
- 自省
  - ReAct
  - Chain of Hindsight

对AIOps的革新影响



个人观点：基于LLM的Ops是AIOps的加强版，并不是颠覆，主要体现在Ops的器和术上，让我们的检测工具更多样更锋利，可以让决策的过程更丝滑。

- 新的检测方法

  兼容不同场景，有可能真正实现all in one

- 更智能的统筹决策

  简化知识->code的转化过程，更鲁棒的决策能力

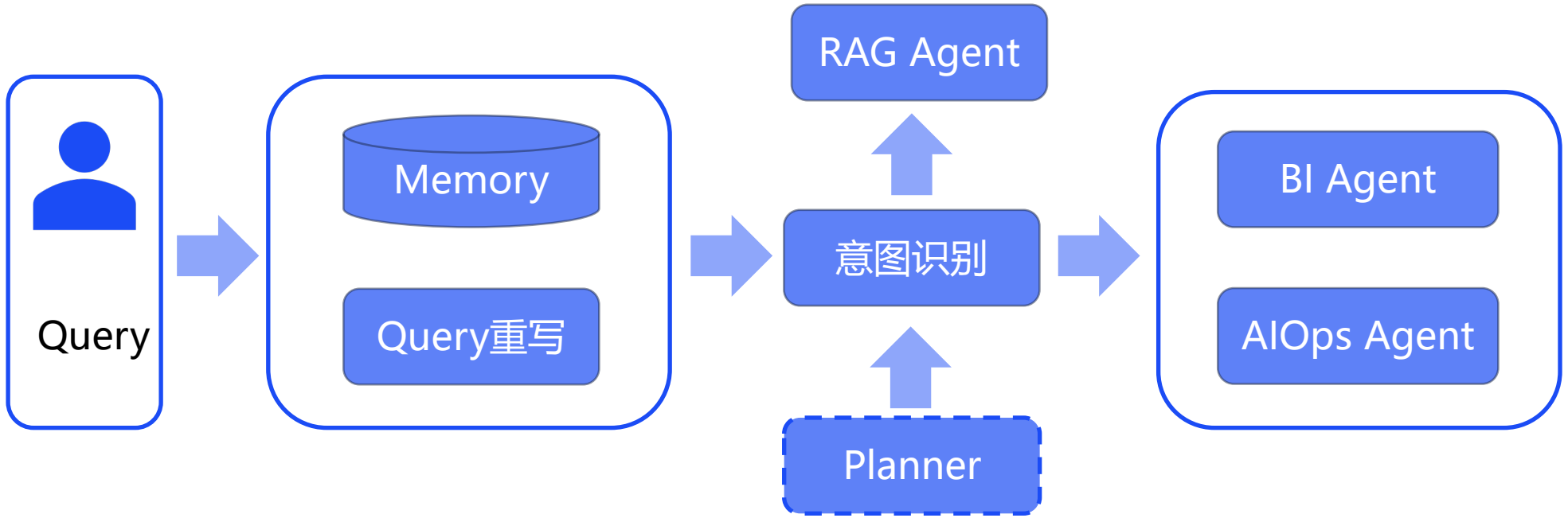
成本 速度 幻觉

PART 03

ECS智能运维在LLM OPS下的创新实践

ECS智能运维在LLM OPS下的创新实践

我们的场景主要是ECS问题排查，包括值班、日常运维



LLM OPS – Memory & RAG

数据库选型：向量 + 全文检索

数据库选型	优势
AnalyticDB PostgreSQL 版	和使用PostgreSQL一样，会有部分能力增强
云原生内存数据库Tair	TairVector + TairSearch，适合多租户场景
智能开放搜索OpenSearch	功能强大，入门门槛较高

Memory管理：向量 + summery

- 一次对话存储<human, ai, context> 三方信息
- 滑动窗口summary：总结信息在后记忆中的权重更大

{“content”: ”1、[IP] 在 2024-03-10 10:26:05 至 2024-03-12 10:26:05 发起根因诊断，
诊断信息如下：机器发起升级，运维重启宕机，链接 https://llmops.aliyun-inc \n 2、[IP]出现宿主机单Socket 打满可能会出现降频行为 \n3、
台机器[IP]云盘达到bps上限, “additional_kwargs”: {“message_ids”: [170507, 170510, 170513]}, "type": "system"}

- 短期记忆：获取topN
- 长期记忆：向量召回 + 权重排序

LLM OPS – Memory & RAG

数据库选型：向量 + 全文检索

数据库选型	优势
AnalyticDB PostgreSQL 版	和使用PostgreSQL一样，会有部分能力增强
云原生内存数据库Tair	TairVector + TairSearch，适合多租户场景
智能开放搜索OpenSearch	功能强大，入门门槛较高

RAG

- 知识库整理&文本chunk
  - 高质量私有知识QA对
  - 基于阿里云文档智能的高效文本分析，支持OCR
- 向量索引：bge-large-zh-v1.5
- 全文索引：
  - 停用词库、关键词库
  - 词法分析：Zhparser
- 多路召回&文档压缩：
  - 召回：BM25、RRF
  - LLM Compression ， 需要考虑tokens、性能

LLM OPS – query重写 & 意图识别

重写的必要性：错别字、语义顺序、上下文实体

Query重写 Prompt

请跟据Hunam和AI的对话历史对新问题进行重写，使得新问题的主体和意图更加明确，重写请遵循如下原则：
1、如果新问题是历史对话的延续则根据历史对问题进行重写
2、如果发现新问题与历史对话没有关联则直接对新问题进行简单的错别字和语法纠错，并以此作为重写结果
3、请使用“重写后的问题：”开头输出输出重写后的问题，直接返回重写结果，不要进行询问

对话历史:
Human: ECS云服务器的定义是什么?
AI:ECS云服务器是一种安全可靠、弹性可伸缩的云计算服务，助您降低 IT 成本，提升运维效率。

新问题:怎么买一台?

重写后的问题:

用户原问题：怎么买一台?

被改写为：如何购买ECS云服务器?

可有效提高召回和意图识别准召！！

意图识别的必要性：tools、agent的定向和编排

意图识别 Prompt

工具集中有一个特殊的工具：默认工具。当无法判断合适的工具进行匹配时，使用默认工具。根据用户输入，挑选一个最合适的工具。

工具集：
qa:知识库问答，值班问题答疑等问答
diagnose:对单个或批量的ECS实例或宿主机进行诊断，查看诊断信息、异常信息、性能问题、根因定位等。
order_analyze:根据给出的工单链接或id，总结工单内容并给出处置建议
默认工具：qa

用户输入：
帮我诊断下这台nc：1.2.3.4

json格式输出：

选中diagnose，调用aiops agent相关能力

多轮对话中的query是被改写之后的

LLM OPS –Agent

Agent：tools调用 + tools编排

ReAct Prompt

```
Answer the following questions as best you can. You have access to the following tools:

health_check：check the status of vm
migrate: migrate the vm from nc to another
...

Use the following format:

Question: the input question you must answer
Thought: you should always think about what to do
Action: the action to take, should be one of [{tool_names}]
Action Input: the input to the action
Observation: the result of the action
... (this Thought/Action/Action Input/Observation can repeat N times)
Thought: I now know the final answer
Final Answer: the final answer to the original input question

Begin!

Question: {input}
Thought:{agent_scratchpad}
```

我们的尝试之后的效果其实并不理想

- 如何让agent不要陷入一个问题的循环？
- 如何让大模型在思考的过程中准确命中已知的解决方案

可以尝试的方案：

- 把RAG作为tools中的一环，通过prompt让agent自行判断如何调用RAG
- 把RAG从思考链中单列出来，外挂在观察阶段，这样每次思考的结果也可以到知识库或者记忆中寻找解决方案

LLM OPS –结果

问答&工单检索

自然语言诊断

LLM OPS –结果

运维工具调用

一键智能诊断

PART 04
总结&展望

总结&展望

- Fine-tuning 还是 prompt？背靠大树好乘凉
- 智能体如何更智能？如何让agent有全局视野
- AIOps基础建设要跟上：tools本身也需要治理？巧妇难为无米之炊
- 如何做到知识共享？没有或者少FT的情况下如何及时更新知识？1对多对话怎么做？
- 如何确认知识边界？做好取舍
- 大模型是万能的吗？More is different？情绪价值？
- 多模态...

附录参考文献

【1】Retrieval-Augmented Generation for Large Language Models: A Survey
【2】https://lilianweng.github.io/posts/2023-06-23-agent/
【3】Generative Agents: Interactive Simulacra of Human Behavior
【4】https://huggingface.co/spaces/mteb/leaderboard
【5】Seven Failure Points When Engineering a Retrieval Augmented Generation System
【6】Are Emergent Abilities of Large Language Models a Mirage?
【7】ReAct: Synergizing Reasoning and Acting in Language Models
【8】Automatic Root Cause Analysis via Large Language Models for Cloud Incidents

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

KEYLINK ing

K+峰会

K+峰会 上海站
K+全球软件研发行业创新峰会
时间：2024.06.21-22

K+峰会 敦煌站
K+思考周®研习社
时间：2024.10.17-19

K+峰会 香港站
K+思考周®研习社
时间：2024.11.10-12

K+峰会详情

AiDD峰会

AiDD峰会 上海站
AI+研发数字峰会
时间：2024.05.17-18

AiDD峰会 北京站
AI+研发数字峰会
时间：2024.08.16-17

AiDD峰会 深圳站
AI+研发数字峰会
时间：2024.11.08-09

AiDD峰会详情

AiDD

THANKS