己实现

(!) Caution

之前已经实现用户注册,把用户的腾讯云密钥保存在数据库中的功能,还有声纹识别同步更新.env 文件中腾讯云密钥

智能体的创建参考https://github.com/business-science/ai-data-science-team

场景

每天凌晨1点服务端自动调用编排器,实现以下流程:

读取.env文件中的腾讯云密钥—在COS桶下载前—天的日志—三类日志交由对应的智能体进行数据分析和向量化—API调用LLM对向量数据库进行分析—得到两个文件,分别是LLM由向量数据库得到的分析结论和LLM总结的文件

```
xiaozhi-esp32-server > analysis results > 2025-08-07 > 🗉 llm summary.txt
    ### 1. **镜像拉取失败 (Image Pull Error) **
     - **涉及对象**: 多个 Pod (如 `config-app`, `robot-service`, `mysql` 等)
    - **原因**:
 8
     - `BackOff`: 拉取镜像失败, Kubernetes 正在重试。
                                                       从日志获取到的原因+错误信息
 Q
     - `ImagePullBackOff`: 持续拉取失败, 导致 Pod 启动失败。
10
    - **具体错误信息**:
11
       - 镜像地址: `my-registry/config-app:v1.2.0`、`registry.example.com/robot-service:v1.0.0`
     - 错误详情: 镜像无法访问、超时或不存在。
12
13
    - **可能原因**:
14
     - 私有镜像仓库配置错误。
                                                    AI推测的原因
     - 镜像标签错误或镜像未上传。
15
16
     - 网络问题导致无法访问镜像仓库。
17
18
19
20 ### 2. **调度失败 (Failed Scheduling) **
    - **涉及对象**: 多个系统组件和应用 Pod (如 `node-problem-detector`, `mysql`, `nginx` 等)
21
      - **资源不足**: 节点内存不足(`Insufficient memory`)。
23
    - **节点亲和性限制**: Pod 无法调度到满足 `NodeAffinity` 的节点。
25
     - **持久卷未绑定**: Pod 依赖的 PVC 未绑定 PV。
26
    - **影响**: Pod 无法启动, 服务不可用。
28
29
30 ### 3. **自动扩缩容失败**
31
    - **涉及对象**:
      - `HorizontalPodAutoscaler` (如 `tke-kube-state-metrics`, `hpa-robot-service`)
33
    - 无法获取指标数据(如 CPU、内存使用率)。
34
```

早上10点,实现以下两套流程:

△ Warning

要做到"一机多用",就需要保存不同的腾讯云密钥;可能会出现小智第一天的主人和第二天的主人不是同一个人,这样的话凌晨1点执行的任务对第二天的主人无效,提出以下两种方案

情况1 (MCP已实现): 用户知道自己连续两天使用这个机器人

用户打招呼"你好,小智"唤醒小智—用户说开始播报日志—小智默认播报前一天的日志—读取前一天的 LLM总结文件并开始播报

设计提示词

∨ 1. 各类日志异常情况概述:

- 业务日志的异常情况
- 事件日志的异常情况
- 审计日志的异常情况
- 2. 重要发现和风险提示
- 3. 需要立即处理的问题
- 4. 可能的解决方案

最后询问用户: 是否需要针对某类日志进行更详细的分析? 我可以:

- 深入分析单个类型的日志
- 关联分析多个类型的日志以找出问题根源
- 提供具体的解决方案和回滚操作建议
- ∨ 请告诉我您想重点分析哪个方面。""",

```
[DDLE, 1. **镜像拉不到**
LDDLE, 好几个服务(像 `config-app`、`robot-service`)都因为镜像拉取失败起不来,可能是仓库配置错了或者网络有问题
[DDLE, 2. **调度失败**
[DDLE, 有些 Pod 因为内存不够或者亲和性设置问题,根本没法调度到节点上
[DDLE, 3. **自动扩容出问题**
[DDLE, 像是 `hpa-robot-service` 这种自动扩容的组件拿不到指标数据,扩不了容,怕是要挂喔
[DDLE, 4. **集群扩容也挂了**
[DDLE, 因为最大节点数到了上限,新 Pod 根本没地儿去~
IDDLE, `mysql-pv-claim` 这个存储声明因为 `StorageClass` 不存在, 绑不上 Persistent Volume
[DDLE, 6. **监控组件抽风**
[DDLE, `tke-monitor-agent` 访问不到监控服务, 监控数据全没了
[DDLE, 总结一下就是
LDDLE, 镜像拉不到、资源不够、调度失败、监控掉线...建议爷爷优先处理镜像和调度问题,不然服务都起不来啦
IDDLE, 要不要我帮你深入分析某一类问题呀?
IDDLE, 比如专门看镜像问题或者调度问题?
AST, None
7-8393-4fld-b671-49f83b665377", "type":"listen", "state":"start", "mode":"auto"}
```

情况2 (MCP快实现): 用户知道自己和前一天的人不是同一个

这就意味着用户需要重新调用编排器

用户发出调用编排器的指令—声纹识别同步更新.env文件里面的腾讯云密钥—使用这一套新密钥开始流程—播报

缺点

- 1. 日志数量太大,考虑优化算法,提升速度,如果用于情况1的话问题不大,但是情况2的话会浪费用户时间
- 2. 目前还是只能用户说"开始播报"以后才开始播报

待实现

小智播报好日志后会询问**是否需要针对某类日志进行更详细的分析?**

自动接收告警

上面说的是每天早上的情况,还有一个场景是

腾讯云平台推送告警的时候,小智机器人同时需要对告警内容进行分析,必要的时候拉取日志进行原因 +措施分析

现在还在研究让小智支持远程MCP唤醒的实现方案,即实时接收告警

Important

腾讯云自身就会推送告警信息到微信,小智的作用是对日志中的异常进行分析+对平台推送的告警信息进行分析吗?

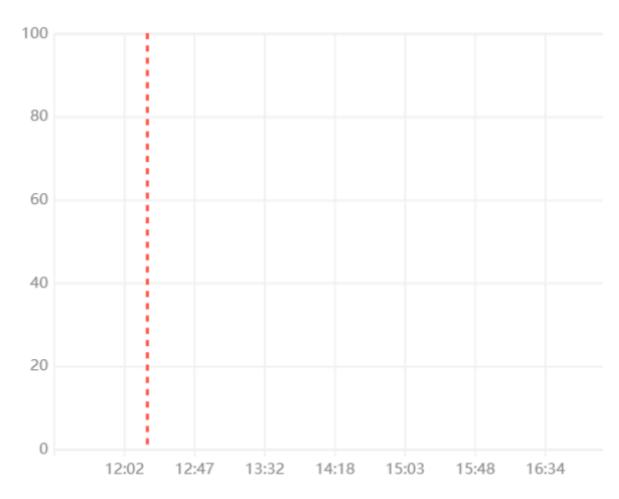
集群告警(云产品监控)



集群ID:cls-l9taan9w | CPU总配置 >0核

CPU总配置 >0核

**** 触发时间 **** 恢复时间



告警详情

告警级别 -

当前数值 3.859核(CPU总配置)

项目|地域 - | 南京

触发时间 2025-07-29 12:17:00



场景demo1:业务日志入手

Important

上次会议说到拉取三类日志有机地串联,现在的效果是用编排器直接把三类日志聚集起来直接分析 出原因,不太像上次会议说到的LLM自己判断下载哪类日志然后分析

如果用户说开始对业务日志进行更详细的分析,以业务日志为出发点

- 设计"告警驱动溯源"场景:
 - 。 当**业务日志**检测到报错或异常时,自动提取关键字段(如时间、Pod/容器名、错误码等)。
 - 。 以这些关键信息为线索, 自动, 在集群**事件日志**中检索相关资源调度、重启、异常事件。
 - 若事件日志中发现异常,再进一步在集群审计日志中查找对应时间段、相关用户或组件的操作记录,分析是否有配置变更、权限操作等。
- 涉及的难点:
 - 溯源: 在大量数据中检索与异常相关的关键词、日志
- 目标: 小智播报异常原因+修复方案+询问是否回滚
- 目前思路
 - 。 结合MCP与编排器, 在场景demo1的情况下调用MCP, 执行编排器
 - 在向量数据库中进行检索并对检索进行增强,做到能快速返回错误原因
- 下一步计划
 - 。 上面的demo实现完了以后再开始实现从事件日志、审计日志出发的demo

计划

上面的功能实现以后实现小智接受告警后会主动拉取日志并分析,得到告警原因+措施