

可观测性+AI的实践和未来

华明 北京快猫星云科技有限公司联合创始人&CTO



◆ 大纲



- ① 愿景
- ② 挑战
- ③ 实践
- ④ 思考



①

愿景





可观测性夙愿

根因定位自动化、智能化



愿景

不久就会到来
的未来

智能化增强观测

- ✓ 代码由人类主导编写，AI辅助
- ✓ 代码质量主要由人类把关
- ✓ 代码和系统由人类发布

Copilot

- 可观测性智能体 - AI Agent

或快或慢也会
到来的未来

智能化闭环观测

- ✓ 代码由AI编写
- ✓ 代码质量由AI把关
- ✓ 代码和系统由AI发布

Autopilot

- 已经没有了人类的位置

愿景

Copilot: AI分析的等级越高，对数据的依赖也越大，但效果也越好

L1: 单点

信息翻译



eg.
解读一段日志 / 告警信息

L2: 多点

要点分析



eg.
针对一批告警分析关键告警

L3: 线

流程加速



eg.
针对固定的workflow分析

L4: 面/立体

自主分析



eg.
根据所掌握的信息自主分析



②

挑战



国外企业都在为智能化解决什么障碍？



自动化的理解生产系统 – 构建知识图谱

Resolve AI automatically maps and keeps up-to-date a complete **knowledge graph** of any environment, without needing any upfront training or user input. Resolve AI understands infrastructure, application, and source code entities, and all changes. It understands which signals and knowledge within tools (e.g. dashboards, alerts, KPIs, etc) are the most relevant for any entity and situation. It builds this understanding by automatically inspecting available content in these tools, learning from its interactions with users, and by using the tools as described next.

训练AI使用工具

- **Adapt to Organizational Conventions:** Each organization—and even different teams—use different sets of tools and have unique conventions for data (e.g., names of metrics, labels on logs), which the AI must comprehend and navigate.
- **Perform Complex Actions:** Without understanding tool-specific setup and capabilities, it cannot page on-callers, apply scaling actions, or implement configuration changes.
- **Handle the scale and live data:** These tools have large volumes of constantly churning data. Trying to brute force RAG all the data from them is a non-starter.

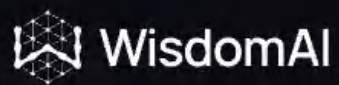
规划器+不同的Agents

Resolve AI Agentic Intelligence platform consists of several agents that have specialized (and composable) capabilities needed to accomplish production engineering tasks. Each agent brings together deep domain intelligence and extensive training about how to accomplish tasks with external tools. A planner takes a task and orchestrates execution across all these agents until the desired goal is achieved. The planner leverages state of the art techniques to balance quality, explainability, and latency. All agents are able to leverage learnings from prior interactions and create new learnings going forward, to enable a high coverage over time even for novel incidents. The learnings are decomposed in a way that they can be reused in a variety of different situations, not just specific to the one in which they were acquired. This is a very challenging system to build as it needs a really large and diverse set of agents operating on multi-modal and dynamic data.

可交互的动态用户界面

Traditional UIs, such as dashboard-centered or chat interfaces, are too inflexible and lack the depth needed for complex tasks. This led us to create an on-the-fly UI, specifically designed to adapt to various tasks and incidents, providing the right tools and insights exactly when and where they're needed. Think of generated dashboards with many different components, built on-the-fly, for the task at hand.

国外企业都在为智能化解决什么障碍？



核心创新是构建一个 Knowledge Fabric

Our core innovation is the Knowledge Fabric - a dynamic, interconnected map of your data ecosystem woven with essential business context. While highly automated, this fabric is continuously refined by your domain experts, the very people who understand your business best, adding the vital institutional knowledge AI alone can't replicate.

Sitting atop this intelligent fabric are our AI Agents. They're designed to think and act - to understand natural language questions, intelligently navigate the data landscape using the Knowledge Fabric as their guide, and automate the heavy lifting of data analysis to surface the insights you need.

更多的 Agents 解决不同的智能化难题

At the heart of WisdomAI are our specialized agents, each tackling a specific insights challenge:

Knowledge Curation Agent This agent is like the master learner for your business, quickly grasping your unique data language and context. Crucially, it learns and is continuously shaped by your team's expertise, incorporating vital human understanding and oversight. This ensures the Knowledge Fabric becomes a true, living reflection of how your business runs, capturing terminology, KPIs, and data relationships.

Instant Answers Agent This agent delivers accessibility, transforming complex business questions asked in plain language by anyone into accurate, immediate insights using the Knowledge Fabric. No technical skills required.

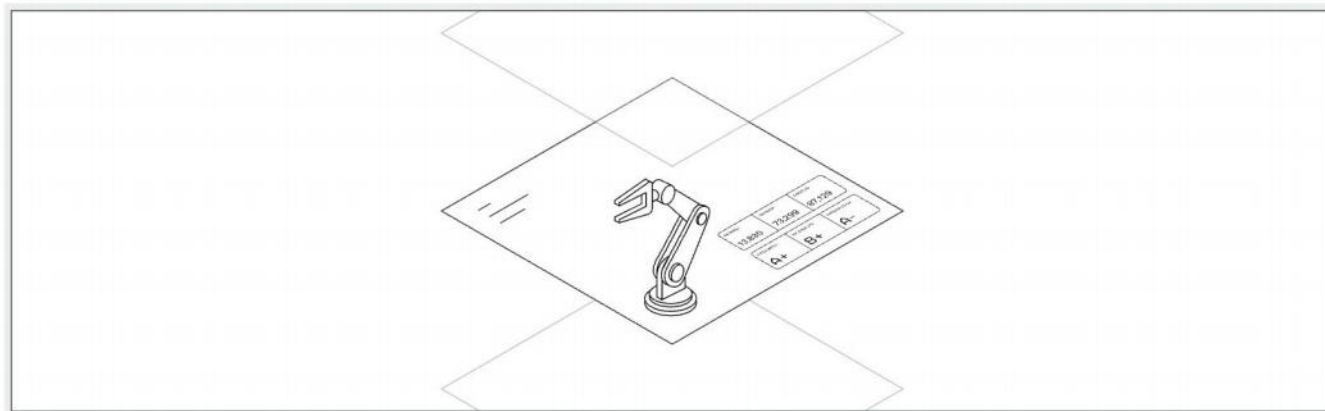
Proactive Insights Agent Perhaps the most revolutionary, this agent doesn't wait for questions; it constantly monitors your data, surfacing valuable trends and opportunities you might not even be looking for yet. Think of early warnings or identifying new revenue segments.



国外企业都在为智能化解决什么障碍？



本体论 - ontology



Ontology building

The Palantir **Ontology** is an operational layer for the organization. The Ontology sits on top of the digital assets integrated into the Palantir platform (datasets, virtual tables, and models) and connects them to their real-world counterparts, ranging from physical assets like plants, equipment, and products to concepts like customer orders or financial transactions. In many settings, the Ontology serves as a digital twin of the organization, containing both the semantic elements (objects, properties, links) and kinetic elements (actions, functions, dynamic security) needed to enable use cases of all types.

挑战



- ① AI如何理解你的系统
- ② AI如何查询你的数据
- ③ 与AI互动共同走向最终答案



挑战

AI如何理解你的系统？

为AI提供企业IT环境上下文

服务元信息、关联信息 – 知识图谱



挑战

AI如何查询你的数据？

为AI提供数据查询能力

打通观测数据源通道，无障碍获取观测数据



◆ 挑战

与AI互动共同走向最终答案

互动、修正

知识库、可动态输入/交互的Chat-UI



③

实践





Flashcat

基于开源夜莺（Nightingale）实现的**统一可观测性产品**，包括完整的指标、日志、链路、事件等维度的观测能力。

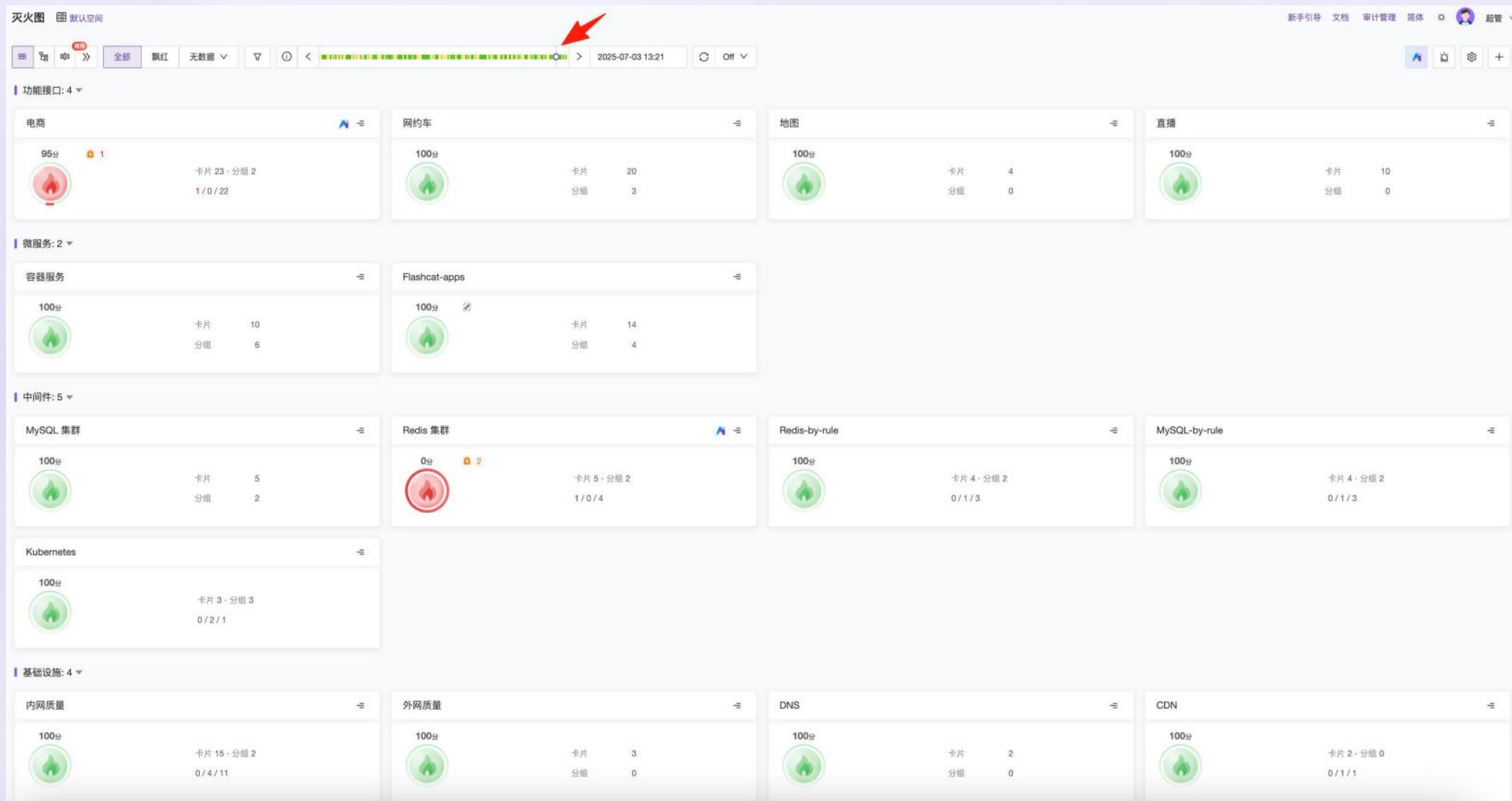
面向稳定性保障场景进行了增强，帮助用户快速发现故障，定位故障，**是企业保障服务稳定运行的支撑平台**。

Flashcat 具备**增强AI分析能力**，加速故障的智能分析定位，已取得了实质进展，未来将在智能化方向继续大力投入。



实践：基于海量观测数据构建IT的知识图谱

AI如何理解你的系统 - 灭火图

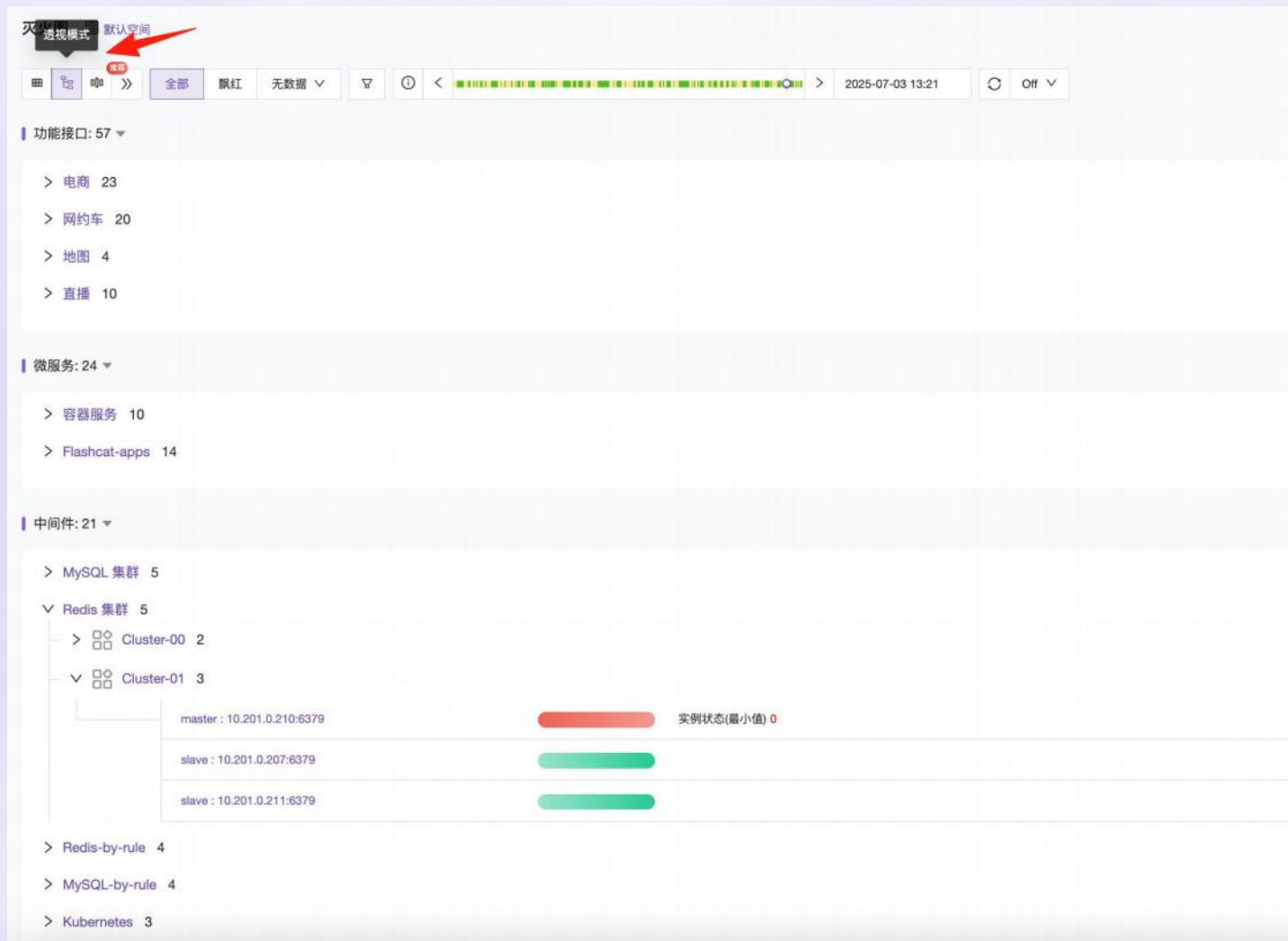




实践：完整的IT服务和状态观测

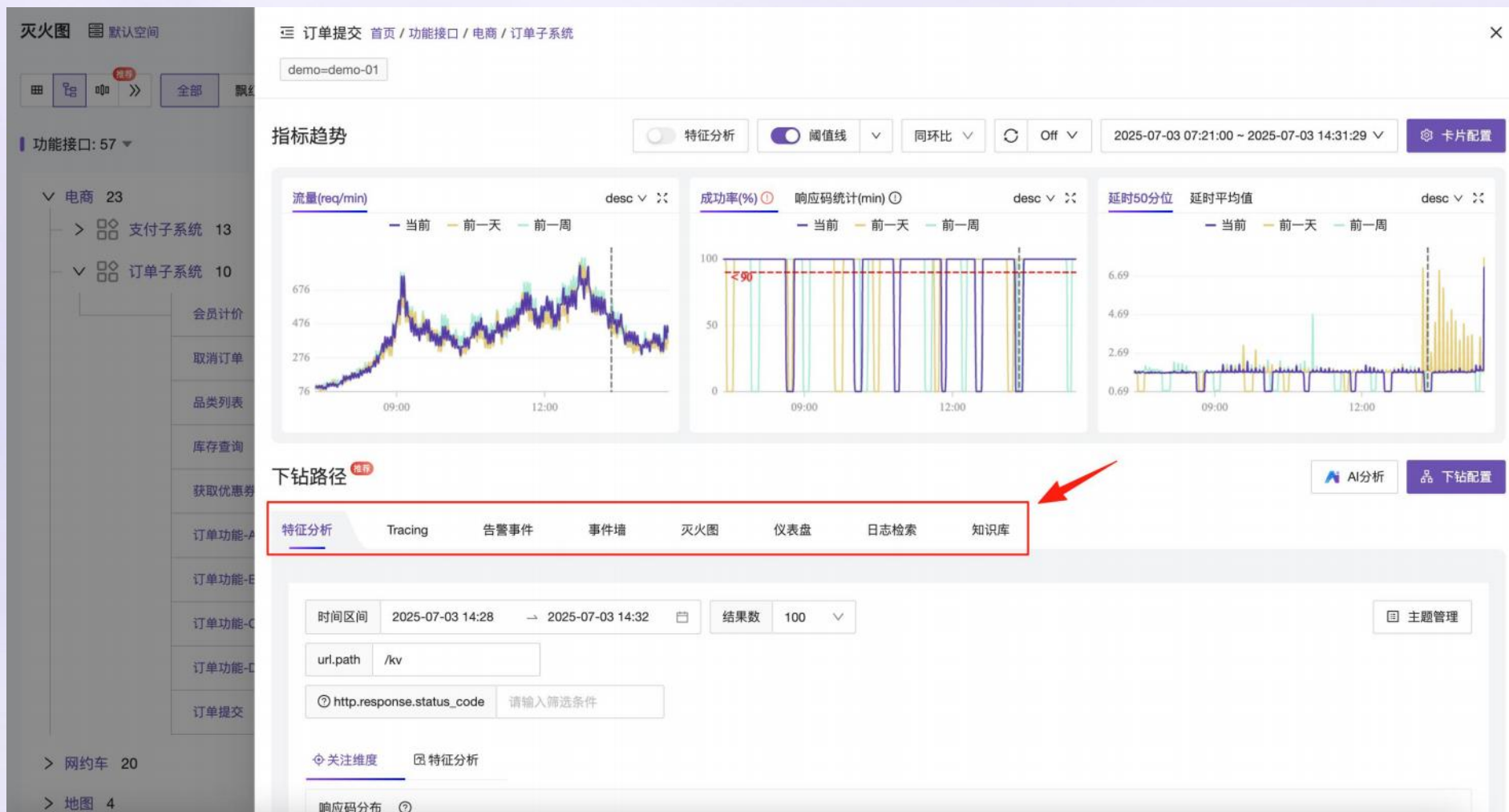


AI如何理解你的系统 - 灭火图



实践：IT对象和观测数据的关联信息

AI如何理解你的系统 – 灭火图



实践：集成市面上所有常见的开源和公有云观测数据源

AI如何查询你的数据 - 数据集成

The image displays three screenshots of a data integration dashboard, categorized into three main sections: 指标数据源 (Metric Data Sources), 日志数据源 (Log Data Sources), and 事件数据源 (Event Data Sources).

指标数据源 (Metric Data Sources): This section lists various metric data sources for integration. The sources include InfluxDB, Oracle, Prometheus Like, MySQL, Zabbix, PostgreSQL, ClickHouse, SQLServer, JSON API, and MongoDB. Each source has a corresponding icon and a '添加' (Add) button.

日志数据源 (Log Data Sources): This section lists various log data sources for integration. The sources include kafka, Elasticsearch, ClickHouse, 腾讯云CLS (Tencent Cloud CLS), 阿里云SLS (Alibaba Cloud SLS), OpenSearch, Loki, and Doris. Each source has a corresponding icon and a '添加' (Add) button.

事件数据源 (Event Data Sources): This section lists various event data sources for integration. The sources include 自定义事件 (Custom Event), Jira, Kubernetes, Jenkins, 自定义事件 (Custom Event), Prometheus, Zabbix, Nightingale, Open-Falcon, and 腾讯云监控 CM (Tencent Cloud Monitoring CM). Each source has a corresponding icon and a '添加' (Add) button.

实践：企业、行业知识库自定义

与AI互动共同走向最终答案

知识库

固定输入：

- 企业知识库
- 产品知识库
- 历史结论反馈

知识库管理

当前对话 + ↺ 📌 ✕

名称	更新时间	更新人	操作
知识库	2025-01-01 10:00	root	编辑 删除
知识库	2025-01-01 10:00	root	编辑 删除
知识库	2025-01-01 10:00	root	编辑 删除

共 5227 条 < 1 2 3 4 5 ... 100 > 30条/页 ▾

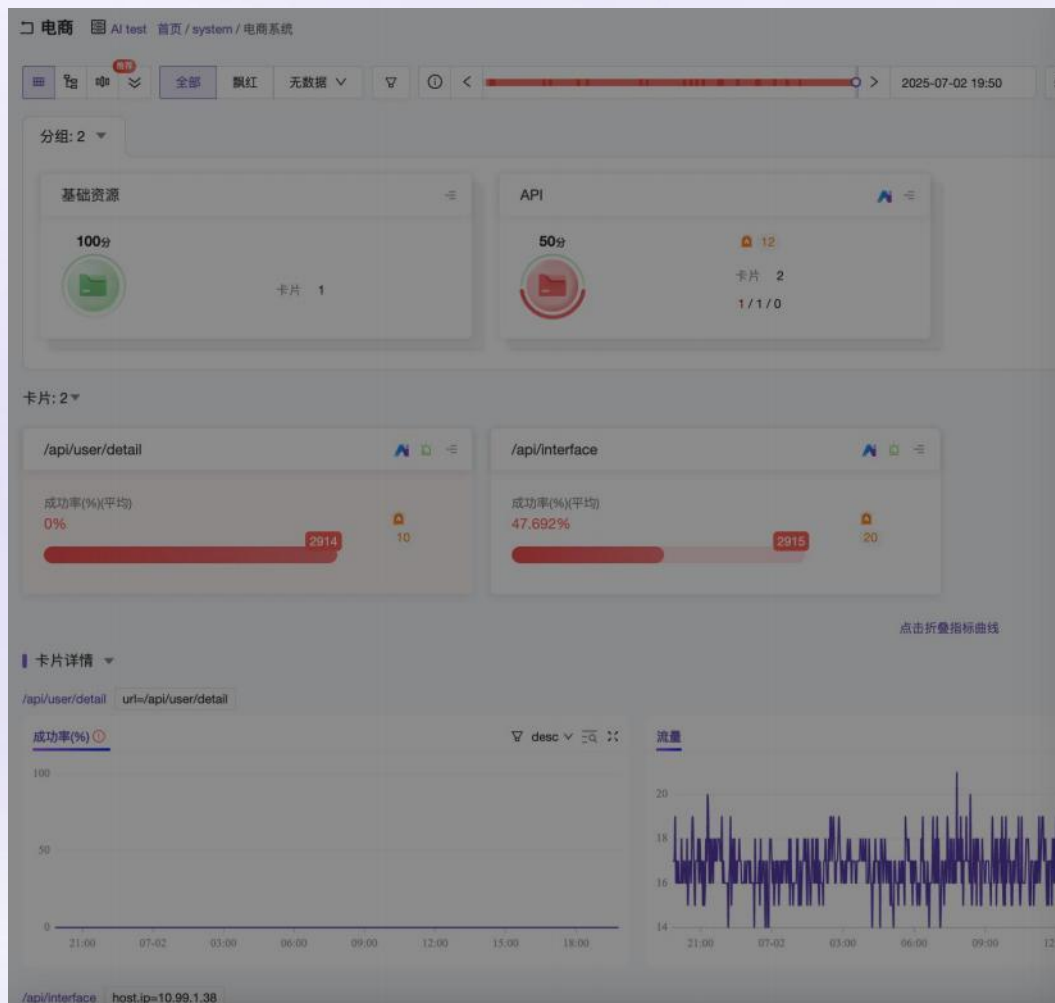
实践：互动对话框，动态输入

与AI互动共同走向最终答案

Chat-UI

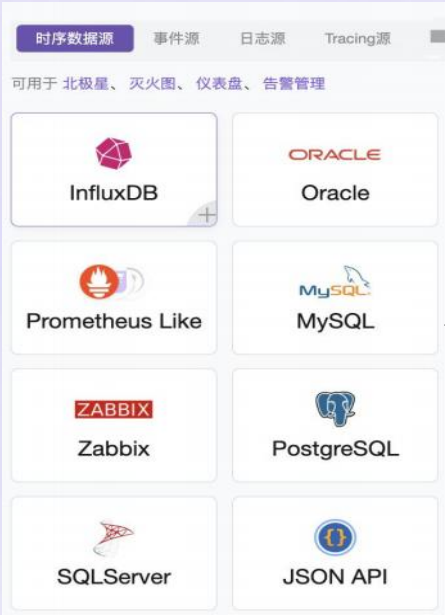
动态输入：

- 交互Prompt
- 问题修正
- 输出格式指定





灭火图：完整的服务信息



数据集成、查询通道

理解你的系统

Agentic 框架



查询你的数据



企业大模型

模型通用能力



Flashcat AI 分析引擎



智能分析结论

模型个性化能力



知识库

Flashcat 内置模型

- 向量模型
- 小型LLM
- 机器学习模型
- 通用规则、算法

企业知识库
产品知识库
历史结论反馈

Chat-UI

静态输入
动态交互



实践

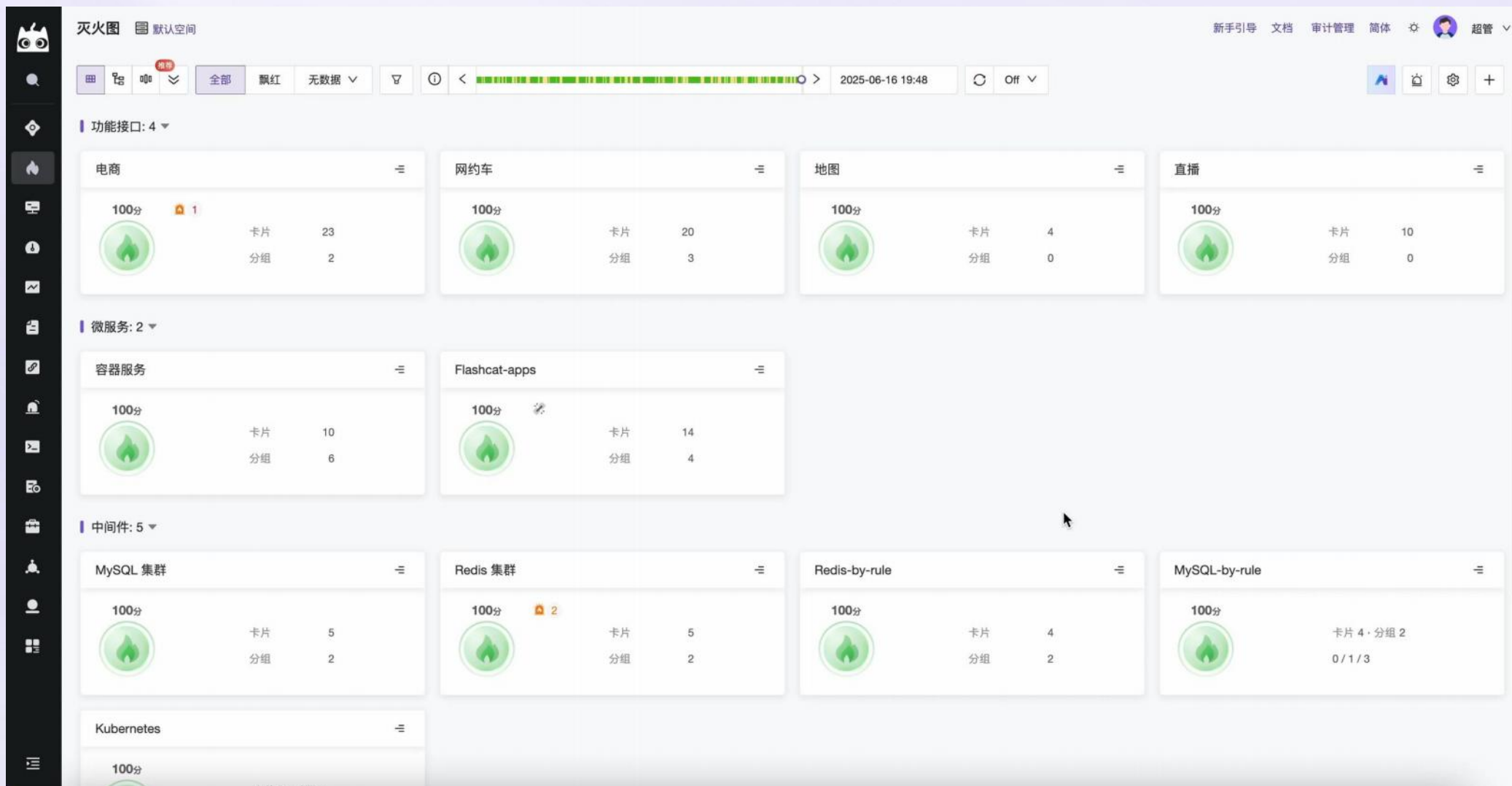
灭火图标识异常

AI从各维度并行分析

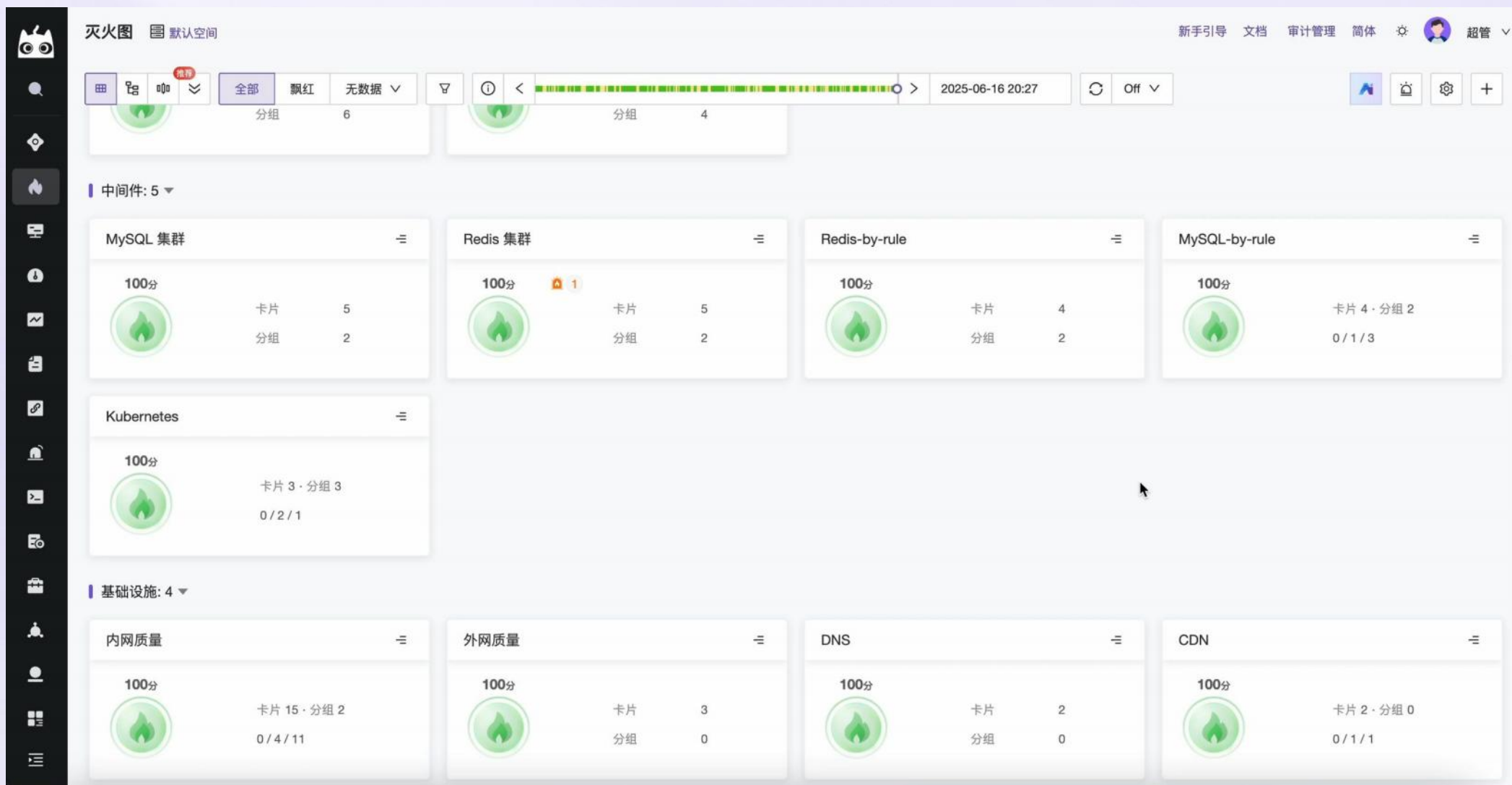
综合得出了准确的根因



实践：基于全局信息智能分析



实践：从异常点下钻深度分析



实践：对话模式，动态输入指令、补充信息

灭火图 AI test

新手引导 文档 审计管理 AI全局分析 Root 账号

全部 飘红 无数据

system: 2

出行服务

33分 30 99+ 卡片 15 · 分组 6 10 / 0 / 5

电商

40分 42 卡片 5 · 分组 2 3 / 0 / 2



Flashduty：一站式告警响应平台：降噪、值班、分派、升级、触达。

已具备 告警信息的AI智能分析功能，AI智能问答助手的内测也在进行中。

实践

Flashduty AI增强：AI Summary

Issue Error: 网络请求超时, 请稍后重试

Info

待处理

IDB59EBB

7小时54分

Flasheye Frontend

认领

故障概览

关联告警

时间线

历史变更

问题修复

历史故障

故障描述

AI总结

网络请求超时, 请稍后重试

故障标签 共11项

展示中文对照

session_count

28

error_count

101

env

demo

suspected_cause_source

auto

error_type

Error

展开更多

关联告警 0 / 2

告警标题

发生时间

application_name

env

1

Issue Error: 网络请求超时, 请稍后重试

6月30日 11:28:35

云眼前端 (demo/ 9000)

demo

2

Issue Error: record not found

6月30日 15:21:57

云眼前端 (demo/ 9000)

demo

AI总结

✕

概述

云眼前端 (demo/9000) 环境同时出现“记录未找到”和“网络请求超时”两类错误, 涉及代码逻辑和网络连接问题。

影响

1. 服务: Flashcat 服务

2. 环境: demo 环境

3. 用户影响: 1 次会话出现“记录未找到”错误, 28 次会话出现“网络请求超时”错误 (累计 101 次错误)

4. 潜在根因:

- 代码无效对象访问 (code.invalid_object_access)
- 网络请求延迟或连接问题 (network.error)

建议

1. 调试无效对象访问: 检查代码中数据查询逻辑, 确保对象存在性验证。

2. 优化网络请求: 分析超时请求的目标服务状态, 调整超时阈值或重试机制。

3. 监控强化: 针对高频网络错误接口增加实时监控, 快速定位延迟源。

取消

确定

实践

Flashduty AI增强：向量相似度匹配

故障列表

集成数据

处理进度 ▾ 最近

☐

新奇

#8C47F3

待处理

☐

#8AD49B

Issue

待处理

☐

#869D7C

Issue

待处理

☐

#827561

Issue

待处理

☐

#8087AB

Issue

待处理

☐

#808436

Issue

待处理

故障列表

集成数据

分派策略

降噪配置

告警聚合 ?

智能聚合

自动智能聚合，当新触发告警与活跃故障的内容高度相似时，将告警合入故障。

编辑

聚合窗口

未开启，将持续合入新告警，直到故障被关闭。

风暴提醒

未开启，不进行风暴提醒。

参与计算字段

系统将根据以下字段的内容计算告警相似程度。

标题

描述

labels.service

labels.resource

您好，
问题吧

可以解

如何调

我没

Flashduty MCP 服务

中文 | English

Flashduty MCP 服务是一个 [模型上下文协议 \(MCP\)](#) 服务，它提供了与 Flashduty API 的无缝集成，为开发人员和工具提供了高级的故障管理和自动化功能。

感谢 [github-mcp-server](#)，本仓库完全基于 github-mcp-server 的架构和模式设计。

使用场景

- 自动化 Flashduty 工作流和流程。
- 从 Flashduty 提取和分析数据。
- 构建与 Flashduty 交互的 AI 驱动的工具和应用程序。

远程 Flashduty MCP 服务

Cursor [Install Server](#)

远程 Flashduty MCP 服务提供了与 Flashduty 集成的最简单方法。如果您的 MCP 主机不支持远程 MCP 服务，您可以改用[本地版本](#)的 [Flashduty MCP 服务](#)。

先决条件

1. 支持最新 MCP 规范和远程服务的 MCP 主机，例如 [Cursor](#)。
2. 来自您 Flashduty 账户的 Flashduty APP 密钥。

安装

以 Cursor 为例：

```
{
  "mcpServers": {
    "flashduty": {
      "url": "https://api.flashcat.cloud/mcp/?toolsets=flashduty_incidents,flashduty_teams&read_only=true",
      "authorization_token": "Bearer <your_flashduty_app_key>"
    }
  }
}
```

Load older messages

查看第一条故障的时间线

Thought for 4s

好的，我来为您查询第一条故障（ID: 686616c5cf0919676e61446c）的时间线。

> Called flashduty_get_incident_timeline ✓

好的，这是故障 686616c5cf0919676e61446c 的时间线：

- 时间: 2025-07-22 17:01:05
 - 事件: 创建新故障
 - 详情: 故障标题 “内部HTTP下游依赖出错 / victormetrics - customers-vpn-01”，严重性为 “Info”。
- 时间: 2025-07-22 17:01:06
 - 事件: 分派故障
 - 详情: 故障被自动分派给 “Default” 规则。
- 时间: 2025-07-22 17:01:07
 - 事件: 发送通知
 - 详情: 系统向相关人员发送了通知。
- 时间: 2025-07-22 17:03:50
 - 事件: 解决故障
 - 详情: 故障已解决。

修改故障的描述为：系统负载过高，继续扩容。

Thought for 12s

好的，请问您想修改哪个故障的描述？是刚才我们查看的这个（ID: 686616c5cf0919676e61446c）吗？

是的

Add Context

Plan, search, build anything

Agent #1 · gemini-2.5-pro-preview-06-05

④

思考





数据治理是否有捷径？

数据治理难，是否可能绕过数据建设和治理，直接实现观测智能化？

- ✓ Garbage in garbage out
- ✓ Data for AI 和 AI for data
- ✓ 马粪问题的解决不在马车时代，数据治理难题也不会在人写代码的时代彻底解决
- ✓ 数据将是企业最重要的资产，数据治理可能是未来为数不多的高价值工作



数据孤岛和存量数据难题

推到重来
全部重采

适合:

- ✓ 从零开始的新服务
- ✓ 重采成本不高的服务

同步转换
集中存储

适合:

- ✓ 无法直接采集的数据

集成打通
随用随查

适合:

- ✓ 存量数据已很完备的环境





大模型持续迭代，产品门槛在哪里？

一定时间内，哪些能力是大模型升级也替代不了的

- ✓ 元数据和服务知识图谱的建设
- ✓ 企业和行业的知识库、流程和know-how
- ✓ 以上信息面向AI的输出方式，AI-Ready 建设



总 结

- 两大阶段：
 - Copilot
 - Autopilot
- 三大挑战：
 - AI如何理解你的系统
 - AI如何查询你的数据
 - 与AI互动互动，共同走向最终答案
- 三个思考：
 - 数据治理是否有捷径
 - 数据孤岛和存量数据难题
 - 大模型持续迭代，产品门槛在哪里



Thanks

谢谢

